Protein Science (2000), *9*:2170–2180. Cambridge University Press. Printed in the USA.

The identification of conserved interactions within the SH3 domain by alignment of sequences and structures

STEFAN M. LARSON[1] AND ALAN R. DAVIDSON[1,2]

[1]Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada
[2]Department of Biochemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada

(RECEIVED June 12, 2000; FINAL REVISION September 1, 2000; ACCEPTED September 1, 2000)

Abstract

The SH3 domain, comprised of approximately 60 residues, is found within a wide variety of proteins, and is a mediator of protein–protein interactions. Due to the large number of SH3 domain sequences and structures in the databases, this domain provides one of the best available systems for the examination of sequence and structural conservation within a protein family. In this study, a large and diverse alignment of SH3 domain sequences was constructed, and the pattern of conservation within this alignment was compared to conserved structural features, as deduced from analysis of eighteen different SH3 domain structures. Seventeen SH3 domain structures solved in the presence of bound peptide were also examined to identify positions that are consistently most important in mediating the peptide-binding function of this domain. Although residues at the two most conserved positions in the alignment are directly involved in peptide binding, residues at most other conserved positions play structural roles, such as stabilizing turns or comprising the hydrophobic core. Surprisingly, several highly conserved side-chain to main-chain hydrogen bonds were observed in the functionally crucial RT-Src loop between residues with little direct involvement in peptide binding. These hydrogen bonds may be important for maintaining this region in the precise conformation necessary for specific peptide recognition. In addition, a previously unrecognized yet highly conserved β-bulge was identified in the second β-strand of the domain, which appears to provide a necessary kink in this strand, allowing it to hydrogen bond to both sheets comprising the fold.

Keywords: β-bulge; β-turn; hydrophobic core; peptide-binding; protein folding; sequence alignment; SH3 domain; structure comparison

The rapidly expanding protein sequence and structural databases provide an invaluable source of data for the elucidation of the principles governing the folding and function of proteins. In particular, the analysis of conservation patterns observed in sequence alignments of homologous proteins can lead to the identification of the key sequence features defining a protein fold. When sequence comparisons can be coupled to analysis of the functional properties and three-dimensional structures of the proteins in question, it is possible to gain a clear understanding of why particular residues are seen at certain positions, and what role these residues play in mediating function or stability. To this end, the sequences and structures comprising several large protein families, including the globins, immunoglobulins, and serine proteases, have been examined in detail (Bashford et al., 1987; Lesk & Fordham, 1996; Chothia et al., 1998). These studies demonstrated that these families possess a common "core" structure, and that high conservation is seen at positions directly involved in function or deeply buried in the three-dimensional structure. Sequence alignments have also proved useful for improving secondary structure prediction (Benner et al., 1994; Rost & Sander, 1994) and in the prediction of the effects of site-directed mutations (Steipe et al., 1994; Maxwell & Davidson, 1998).

In this study, we have undertaken a comprehensive analysis of the sequence and structural variation seen in the SH3 domain. First recognized as a noncatalytic homology region in protein kinases related to Src (Mayer et al., 1988), SH3 domains have now been identified in more than 350 different proteins in organisms ranging from yeast to humans. SH3 domains are found in kinases, lipases, GTPases, adaptor proteins, structural proteins, and others, and these proteins act in diverse processes including signal transduction, cell cycle regulation, and actin organization. SH3 domains mediate specific protein–protein interactions by binding to PXXP-containing sequence motifs in target proteins (for reviews, see Pawson, 1995; Dalgarno et al., 1997). SH3 domains generally bind their targets with affinities between 1 and 50 μM, and individual domains display distinct specificities (Alexandropoulos et al., 1995; Rickles et al., 1995; Sparks et al., 1996).

Reprint requests to: Alan R. Davidson, Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada; e-mail: alan.davidson@utoronto.ca.

Comprised of approximately 60 residues, the SH3 domain fold is composed of five β-strands arranged into two sheets packed at right angles (Fig. 1). The first sheet is formed by β-strands *a*, *e*, and the first half of *b*, while the second is formed by β-strands *c*, *d*, and the second half of *b*. A kink in β-strand *b* allows it to participate in both sheets. β-Strands *a* and *b* are separated by the long RT-Src loop, which possesses an irregular antiparallel structure. The shorter N-Src and Distal loops are found between β-strands *b* and *c*, and *c* and *d*, respectively. The four residues N-terminal to β-strand *b* form a type II β-turn, and the three residues separating β-strands *d* and *e* are generally found in a 3_{10}-helical conformation. Peptide binding by the SH3 domain is mediated by a surface region rich in aromatic residues (Fig. 1), and by various polar residues located in the RT-Src and N-Src loops.

The SH3 domain provides an excellent system for the examination of sequence and structural conservation. Hundreds of very diverse SH3 domain sequences are available for analysis, providing a broad sampling of sequences that are consistent with the SH3 domain fold. In addition, 44 structures of 19 different SH3 domains are available in the structural database. Due to its small size and its amenability to in vitro analysis, the SH3 domain has been a popular subject for protein folding studies (Viguera et al., 1994; Chen et al., 1996; Grantcharova & Baker, 1997; Plaxco et al., 1998; Filimonov et al., 1999) and mutagenesis studies where effects on peptide binding, thermodynamic stability, and folding kinetics have been measured (Lim & Richards, 1994; Weng et al., 1995; Grantcharova et al., 1998; Martinez et al., 1998; Maxwell & Davidson, 1998). This wealth of experimental data aids us in attempting to interpret conservation patterns observed in the SH3 domain sequence alignment.

In the work described here, we have constructed and analyzed a large alignment of SH3 domain sequences present in the database. We have also aligned the structures of 18 different SH3 domains, allowing characterization of variation in the SH3 domain at both the sequence and structural level. By combining sequence and structural analysis, it has been possible to interpret the pattern of conservation seen in the SH3 domain and identify the interactions in this fold that are crucial for its stability and function.

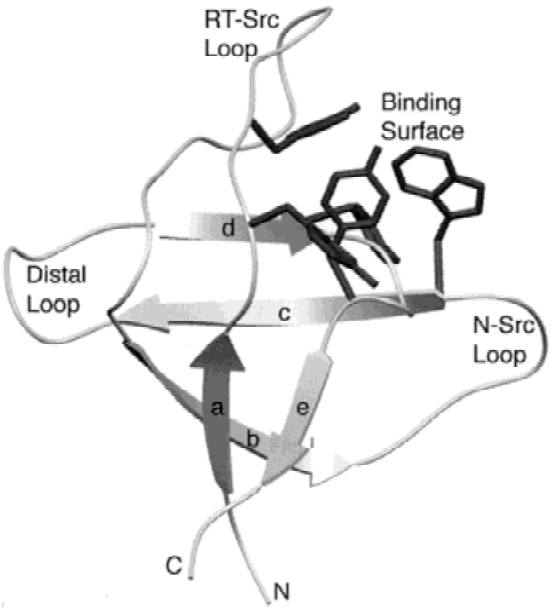

Fig. 1. Ribbon diagram of the SH3 domain from the Fyn tyrosine kinase (1shf). The β-strands are labeled a–e, and the loops are designated. Aromatic side chains involved in peptide binding are shown.

Results and discussion

Construction of the SH3 domain sequence alignment

To ensure a complete and accurate SH3 domain alignment for analysis, an iterative routine of PSI-BLAST searches and progressive alignments was used to align all available SH3 domain sequences (see Materials and methods). Our final alignment contained 266 nonredundant (no two sequences are more than 90% identical) sequences (see Supplementary material in the Electronic Appendix for the complete alignment). Variability in the length of SH3 domains resulted predominantly from insertions or deletions in the N-Src and Distal loop regions: 96 sequences differed from the typical SH3 domain in the N-Src loop region, with 25 sequences one residue shorter, 34 sequences one residue longer, and the rest having longer inserts of variable length. The observed lengths of the Distal loop fell mostly into three groups with lengths of two residues (101 sequences), three residues (50 sequences), and four residues (63 sequences), with the others (51 sequences) having longer inserts. Considerably less variability was seen in the RT-Src loop where only 39 sequences had insertions or deletions, and only four sequences had insertions in the type II β-turn following the RT-Src loop. The gaps in our alignment, which always occurred between secondary structure elements, had to be positioned somewhat arbitrarily due to the high sequence diversity in these regions. Thus, in our analyses, any positions that were gapped in more than one-third of the sequences were not included. The remaining positions were numbered 1 to 60 using the numbering scheme shown in Figure 2.

The overall diversity of the sequences in our SH3 domain alignment was assessed by performing an all vs. all pairwise identity analysis. The pairwise identity of every sequence compared with every other sequence was determined, and the frequency of occurrence of each pairwise identity was plotted (Fig. 3). It can be seen that there is a close to normal distribution of percent identities and that most fall in the range between 15 and 45%. The average pairwise identity of all the sequences in the alignment is 27%. These data show that the SH3 domain alignment is very diverse. Despite this diversity, however, certain more closely related subgroups of sequences were still found to be overrepresented in the alignment, causing a bias in residue frequencies toward these sequences. For this reason, a weighting algorithm (Henikoff & Henikoff, 1994) was used to downweight overrepresented sequence subgroups when residue frequencies at each position in the alignment were calculated (see Supplementary material in the Electronic Appendix for the complete set of positional residue frequencies).

The frequency of occurrence of each amino acid at each position in our alignment was used to assess the degree of conservation at each position in the alignment using the Shannon informational entropy calculation (Shenkin et al., 1991). In this work, we express the entropy value in an exponentiated form, which we call "positional entropy" where a positional entropy of *n* is equivalent to the diversity of *n* residues occurring at the position with a frequency of $1/n$. For example, if a position has a positional entropy value of 8, its variability is equivalent to a situation where eight different

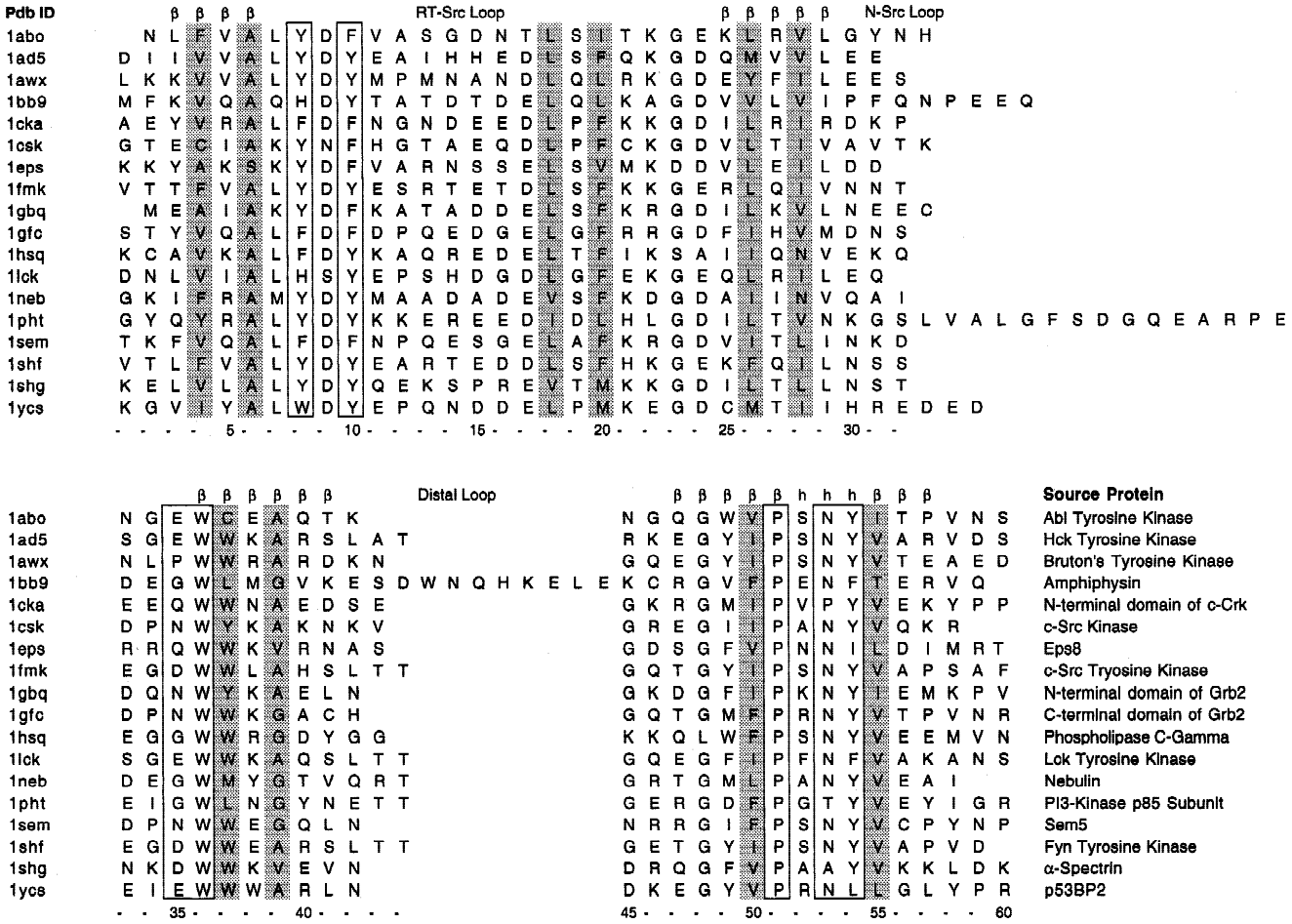

Fig. 2. SH3 domain alignment. An alignment of the sequences of the 18 SH3 domains with solved structures that were used in the structure comparison performed in this work is shown. The SH3 domain consensus sequence and numbering system derived in this work are also shown. The PDB accession and parent protein are shown for each sequence. Conserved hydrophobic core positions are shaded, and the most conserved positions involved in peptide binding are boxed. This small alignment is representative of the complete SH3 domain alignment in the positions of gap regions. Numbers are only given to positions that were gapped in less than one-third of sequences in the 266 sequence alignment.

residues occur at the position, all with equivalent frequencies of 12.5% (1/8). A position that is completely conserved (i.e., only one residue appears there) will have a positional entropy of 1. Conservation of hydrophobicity at each position in the alignment was assessed by multiplying the residue frequencies by the appropriate hydrophobicity value as quantitated by water to octanol transfer free energy values (Fauchère & Pliska, 1983). The conservation data from the sequence alignment are summarized in Figure 4A–C. In the rest of this work, positions in the SH3 domain will be designated by their number in the alignment, and the residue occurring most often at that position will be placed in parentheses after the number.

SH3 domain structural alignment

The protein structural database contains the structures of 19 different SH3 domains, many of which have been solved multiple times. For our analysis, we chose the highest resolution X-ray

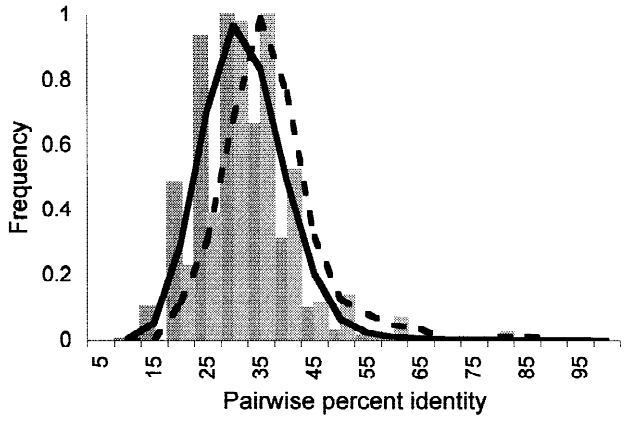

Fig. 3. Distribution of pairwise identities in the whole SH3 domain alignment (dark bars, solid line) and in the 18 sequences of the SH3 domain with solved structures (light bars, dotted line).

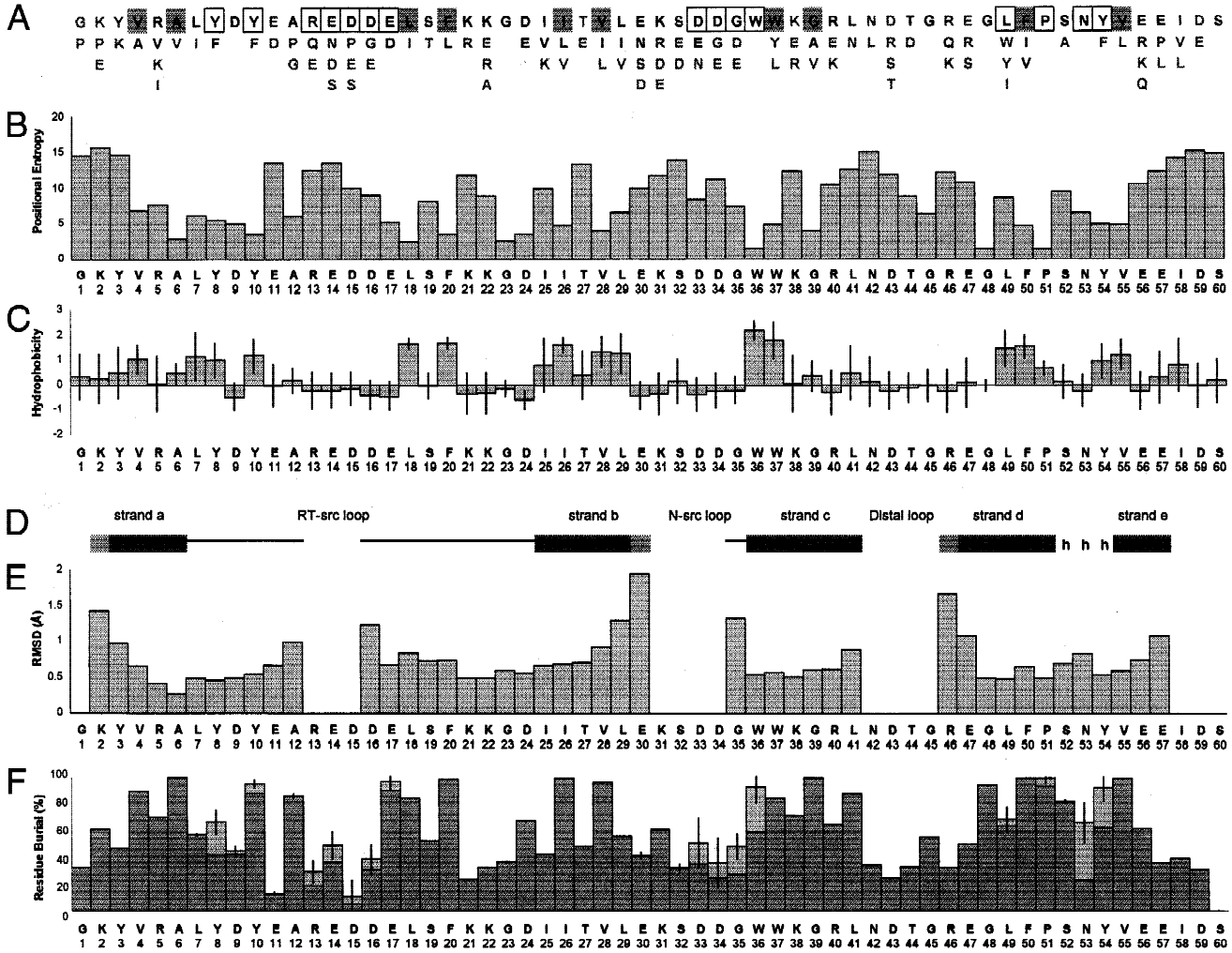

Fig. 4. Summary of sequence and structure conservation in the SH3 domain family. **A:** The consensus SH3 domain sequence. Other residues that appear frequently are listed below the consensus sequence. Residues occurring at a frequency between 10 and 20% are gray, and residues occurring at a frequency greater than 20% are black. Hydrophobic core positions are shaded and residues involved in peptide binding are boxed. **B:** Sequence conservation is displayed by graphing the positional entropy for each position in the SH3 domain alignment (values range from 1–20 where 1 represents complete conservation). **C:** SH3 domain hydrophobicity profile. Mean (gray bars) and standard deviations (black error bars) for hydrophobicity at each position were calculated based on the weighted residue frequencies from the 266 sequence alignment. **D:** Conservation of secondary structure in 18 different SH3 domains. Black, positions in β-strands in more than 15 structures; gray, positions in β-strands in more than 10 structures; lines, loops; h, 3_{10} helix. **E:** Structural conservation at each position in the SH3 domain. Loop positions with average pairwise RMSD values greater than 2 Å were not included in the structural core and are not shown here. **F:** Residue burial of each position in the SH3 domain averaged from 18 structures. Also shown is the average increase in percent residue burial upon ligand binding, based on 17 SH3 domain–ligand complex structures. Dark gray, no ligand; light gray, with bound peptide. Error bars represent one standard deviation for increase in residue burial upon ligand binding.

crystal structure of each domain wherever possible. The NMR structure of the SH3 domain from the Itk tyrosine kinase was excluded because we found that its structure showed relatively large deviations from those of any other SH3 domain structure. By examining various measures of structural quality, such as the WHATCHECK (Hooft et al., 1996) report associated with the Protein Data Bank (PDB) file, we concluded that the deviations displayed by the Itk SH3 domain structure were most likely due to the low resolution of the structure determination. Figure 2 shows the amino acid sequences and PDB identifiers of the 18 structures used for structural alignment. Fortuitously, these 18 sequences are almost as diverse as the 266 sequences in our whole alignment with the average pairwise identity of the sequences of the known structures being ~30%, while the average pairwise identity in the whole sequence alignment is ~27%. An all vs. all pairwise identity analysis performed, as described above, on the 18 sequences of solved SH3 domains (Fig. 2) demonstrated that the distribution of these sequences is also similar to the full alignment (Fig. 3). This similarity allows for valid comparisons between the variation seen in the sequence alignment and the variations in properties of the 18 structures.

After renumbering the residues in each structure file according to the alignment numbering, an all vs. all set of pairwise structural alignments was carried out. No pairs of corresponding residues

outside of gap regions (according to the sequence alignment) were misaligned in any of the structural alignments, indicating that our sequence alignment is structurally valid. The distance between the C_α atoms of corresponding residues in each pair of structures was measured, and the average pairwise root-mean-square deviation (pairwise RMSD) of aligned residues at each position in all the structures was tabulated (Fig. 4E). This information was used to determine the regions of the domain that are most conserved in each structure. This "structural core," defined as the region where the average pairwise RMSD of aligned residues was less than 2 Å, included most of the SH3 domain except residues 13–15 in the RT-Src loop, residues 31–34 in the N-Src loop, residues 42–45 in the Distal loop, and residues at the N- and C-termini of the domain. The positions of the secondary structural elements in each structure as defined by the program DSSP (Kabsch & Sander, 1983) were found to be highly conserved (Fig. 4D).

The overall pairwise RMSD for every pair of SH3 domain structures was calculated using only the structural core positions. These pairwise RMSD values ranged from 0.47 to 1.72 Å, averaging 0.98 ± 0.27 Å. The average pairwise RMSD between each structure and the other 17 structures was used as a measure of how closely each structure resembles a "typical" SH3 domain structure. Most structures have an average pairwise RMSD between 0.8 and 1.0 Å. Two structures, the SH3 domain from amphiphysin (1bb9) and the SH3 domain from PI3 kinase (1pht), are distinguished by large loops (Fig. 2). Strikingly, the structural core of each of these domains remains unaffected even with these large insertions (average pairwise RMSD of 0.91 and 0.98 Å, respectively). The structure of the SH3 domain of Eps8 was solved as a novel strand-swapped dimer, but the structural core of the hybrid monomeric domain has an average pairwise RMSD of only 0.88 Å. Only the structures of Btk (1awx), C-terminal Grb2 (1gfc), PLC-γ (1hsq), and nebulin (1neb) SH3 domains, all solved by NMR, have average pairwise RMSD values over 1.0 Å. The larger deviations among structures solved by NMR is most likely a result of the lower resolution of these structures compared to those solved by X-ray crystallography. The average pairwise RMSD between all the structures solved by X-ray crystallography is only 0.78 ± 0.12 Å. Overall, the core structure of the 18 solved SH3 domains examined here is remarkably conserved given that sequences of these domains are quite diverse.

Side-chain burial in SH3 domains with and without peptide ligand

To characterize the environment of each position in SH3 domain structures, the side-chain surface solvent accessibility of each position was averaged in the 18 structures under examination. The results are expressed as percent residue burial (Fig. 4F). To quantitate the importance of each residue in ligand binding, the difference in residue burial with and without ligand was calculated for each position in 17 SH3 domain structures (these are not all the same structures discussed above, see Materials and methods) that were solved in the presence of a target peptide. The average changes in surface accessibility of each position in the SH3 domain upon ligand binding, with standard deviations, are also shown in Fig. 4F. It should be noted that in our comparisons of SH3 domain structures solved with and without target peptide, no significant differences in the overall conformation of the SH3 domain in the presence of ligand were seen. Thus, the changes in side-chain accessibility seen in the ligand-bound structures are due only to the direct interaction between these side chains and ligand.

The roles of conserved positions in the SH3 domain

Table 1 lists the 30 most conserved positions in the SH3 domain alignment as evaluated by their positional entropies. In the following sections, we discuss the roles of these positions in the structure and function of the SH3 domain.

Residues involved in target peptide binding

From the analysis of side-chain burial described above, 15 positions were identified as playing a significant role in target peptide binding. Among these positions, 51(P) and 36(W) are the two most conserved positions in the SH3 domain alignment and five more, 8(Y), 10(Y), 35(G), 53(N), and 54(Y), possess positional entropy values below 7.5 (Table 1). At some positions, the increase in residue burial in the presence of bound ligand varies widely from structure to structure. However, there is considerably less

Table 1. *The 30 most conserved positions in the SH3 domain*

Rank	Position	Consensus residue	Positional entropy	Role[a]
1	51	P	1.3	Binding
2	36	W	1.3	Binding
3	48	G	1.4	Structural
4	18	L	2.4	Core
5	23	G	2.5	β-Turn
6	6	A	2.8	Core
7	24	D	3.4	β-Turn
8	10	Y	3.4	Binding
9	20	F	3.5	Core
10	39	G	3.8	Core
11	28	V	3.9	Core
12	50	F	4.5	Core
13	26	I	4.7	Core
14	37	W	4.7	Core
15	55	V	4.7	Core
16	54	Y	4.9	Binding
17	9	D	5.0	RT-Src
18	17	E	5.2	RT-Src
19	8	Y	5.4	Binding
20	12	A	6.0	RT-Src
21	7	L	6.1	Structural
22	45	G	6.2	β-Turn
23	53	N	6.5	Binding
24	29	L	6.6	β-Bulge
25	4	V	6.9	Core
26	35	G	7.3	Binding
27	5	R	7.6	?
28	19	S	8.1	RT-Src
29	33	D	8.4	Binding
30	49	L	8.6	Binding

[a]These are roles determined for these positions as discussed in the text. "Binding" refers to positions involved in peptide binding, "core" are hydrophobic core positions, "RT-Src" are positions involved in conserved interactions in the RT-Src loop, positions designated "structural" appear to be important in determining the domain structure, but their role is not well defined.

variability in the increase in burial seen at the seven most conserved positions involved in ligand interaction, as illustrated by the relatively low standard deviations in the average increases in residue burial seen at these positions (Fig. 4F). Thus, the most conserved positions that contact ligand are seen to be consistently important in ligand binding in all the solved structures. On the other hand, positions that contact ligand that are not highly conserved in the sequence alignment [13(R), 14(E), 15(D), 16(E), 33(D), 34(D), and 49(L)] also vary a great deal more in their degree of burial upon ligand binding in different structures. The sequence and structural variation at these positions suggest that they may play a greater role in defining target specificity, as their importance in the binding reaction is dependent on the combination of SH3 domain and target peptide under examination. This point is illustrated in Figure 5A, which shows an SH3 domain structure and an overlay of SH3 domain target peptides in their bound conformations. The residues with positional entropy values below 7.5 (red) contact the peptides in a region where they all adopt a similar polyproline type II helical conformation. Conversely, the less conserved residues (blue) contact the target peptides in a region where the bound peptide conformations diverge. The almost total conservation seen at the 51(P) and 36(W), and extremely high conservation at the 10(Y) position may be explained by their location at the very center of the region contacted by peptide (Fig. 5A).

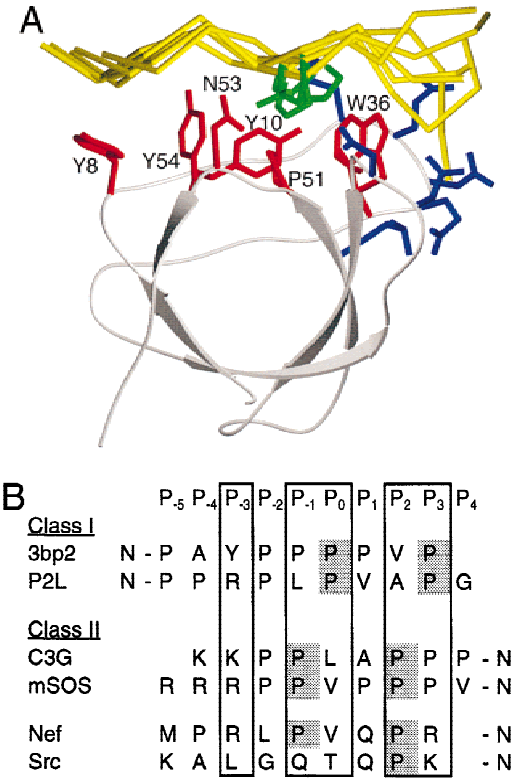

Fig. 5. A: Structure of the SH3 domain with bound peptides. The structure of the Fyn SH3 domain (1shf) is shown with structurally overlaid bound peptide structures from six different SH3 domain–ligand complexes [1fyn (3bp2), 1azg (P2L), 1cka (C3G), 1gbq (mSOS), 1efn (Nef), 1fmk (whole Src tyrosine kinase)]. Red side chains (residue identities are those found in the Fyn SH3 domain) are at highly conserved positions contacting ligand and blue side chains are at positions that contact ligand in most structures, but display less conservation in the sequence alignment. The ligand structures are shown in yellow, and side chains at the P_{-1} position are shown in green. **B:** Names and sequences of the structurally aligned bound peptides shown in part A. The sequences are aligned according to their structural superposition. Some sequences are shown in the C- to N-terminal direction left to right because they bind to SH3 domain in the opposite orientation to the others (i.e., class II binding). The designation of positions in the peptide sequences are according to Lim et al. (1994).

SH3 domain ligands are pseudosymmetrical and are bound in both NH_2 to COOH (Class I) and a COOH to NH_2 (Class II) orientations (Feng et al., 1994; Lim et al., 1994). Structural alignment of target peptide sequences (Fig. 5B) shows that the binding mode is the same whether Class I or Class II targets are bound. Interestingly, the conserved PXXP sequence motif seen in most SH3 domain ligands does not align structurally in the Class I and Class II ligands (Lim et al., 1994). It can also be seen that the peptide bound by the Src SH3 domain in the structure of the complete Src kinase (1fmk) (Xu et al., 1997) overlays well with the other SH3 domain ligands, even though it contains only one proline residue. Further discussion of specific details of the interaction of SH3 domains and target peptides can be found in other publications (Feng et al., 1994; Lim et al., 1994; Lee et al., 1996; Lim, 1996; Pisabarro & Serrano, 1996; Arold et al., 1997; Dalgarno et al., 1997; Sicheri et al., 1997; Xu et al., 1997).

The hydrophobic core

The hydrophobic core of a protein is comprised of a group of hydrophobic side chains that are predominantly inaccessible to solvent due to their close packing in the interior of the protein. Because the hydrophobic core is critical for protein stability (for review, see Dill, 1990), hydrophobic core positions in an alignment are generally highly conserved and are mostly occupied by hydrophobic residues (Bashford et al., 1987; Chothia et al., 1998). In the SH3 domain alignment, there are nine positions [4(V), 10(Y), 18(L), 20(F), 26(I), 28(V), 37(W), 50(F) and 55(V)] that are on average >85% buried and have an average hydrophobicity of >1 (Fig. 4C). Analysis of the SH3 domain structures indicates that all of these positions except position 10(Y) are indeed hydrophobic core residues. Position 10(Y) is not considered to be a hydrophobic core residue because it contacts ligand and is not packed closely with most of the other core residues. Although they do not have high average hydrophobicities, the 6(A) and 39(G) positions are designated as hydrophobic core positions. They are 100% buried in all SH3 domain structures, they are almost never occupied by polar residues, and their side chains make extensive contacts with other hydrophobic core side chains. Although the 41(L) position is highly buried and hydrophobic in a number of SH3 domains, it cannot be considered as part of the conserved hydrophobic core of the domain because it is one of the least conserved positions in our alignment and is often occupied by polar residues. Furthermore, a variety of amino acid substitutions at this position in the Fyn SH3 domain was found to have only small effects on stability and function (Maxwell & Davidson, 1998).

The 10 positions comprising the hydrophobic core are highly conserved in our alignment, with 7 of 10 possessing positional entropy values between 3 and 5 (Table 1). Two core positions, 6(A) and 18(L), show considerably greater conservation, each displaying the consensus residue ~75% of the time and possessing positional entropy values below 3. Because these two positions flank the functionally crucial RT-Src loop, their high degree of conservation may be a result of a requirement to maintain this

region in exactly the correct conformation for ligand binding. The 4(V) position is considerably less conserved than the other core positions with a positional entropy value of 7. The less stringent sequence requirements at this position may be due to its location at the edge of the core on the opposite side of the domain from the peptide binding surface. Examination of the residue frequencies at the core positions shows that each has defined preferences for particular hydrophobic residues. For example, Phe is seen at the 20(F) position 55% of the time with Ile seen only 7% of the time, while at the 26(I) position, Ile is seen 36% of the time and Phe is seen with a frequency of only 6%. In an extensive mutagenesis study performed in our laboratory on the hydrophobic core of the Fyn SH3 domain, we have found that substituting positions with amino acids not often seen in the alignment generally results in destabilization of the domain (A.A. Di Nardo, S.M. Larson, & A.R. Davidson, unpubl. results). Thus, the conservation seen at core positions appears to reflect constraints on the ways in which hydrophobic residues can stably pack into the SH3 domain core.

Other conserved hydrophobic positions

The 7(L) and 29(L) positions are both striking in that the average hydrophobicity of residues occurring at these positions is similar to hydrophobic core positions and their positional entropy is relatively low, yet neither position is highly buried in either bound or unbound structures (Fig. 4C,F). Interestingly, the ϕ/ψ angles at these two positions lie in the right-handed α-helical region of the Ramachandran plot in most of the SH3 domain structures (Fig. 6A). Both the 7(L) and 29(L) residues lie at positions where an extended polypeptide chain changes direction. At the 7(L) position, β-strand *a* ends and the change in direction defines the beginning of the RT-Src loop. The direction change seen at the 29(L) position allows β-strand *b* to participate in hydrogen bond interactions with both β-sheets comprising the SH3 domain.

Although the 7(L) position is clearly playing a conserved structural role in that it always lies at a point where the chain changes direction and its amide hydrogen is always hydrogen bonded to the carbonyl oxygen of 54(Y), the high preference for Leu at this position is not easily explained. The Leu side chain is invariably packed against the backbone atoms of the 55(V) and 56(E) and could sometimes make stabilizing interactions with the side chain at the 56(E) position. The Leu side chain is also packed quite closely to the 8(Y) side chain and may be important for maintaining the proper conformation of this side chain for peptide binding. In general, the 7(L) position is not directly involved in peptide binding, but it might play a direct functional role in some cases where the SH3 domain is binding to a whole protein rather than just an isolated small peptide as in most of the structures solved to date.

Conservation at the 29(L) position is most likely due to this position occupying position 1 of a classic β-bulge in most SH3 domain structures (Chan et al., 1993). The defining features of the β-bulge in the SH3 domain are that the NH atoms of both the 29(L) and 30(E) positions hydrogen bond to the CO atom of the 38(K) position and the CO atom of the 30(E) position hydrogen bonds to the NH of the 38(K) position (Fig. 7). The right-handed helical ϕ/ψ angles generally seen at the 29(L) position (Fig. 6A) are characteristic of position 1 in the classic β-bulge. In addition, Leu, Ile, and Val are observed most commonly at this position in classic β-bulges (Chan et al., 1993), and these are the same three residues that are seen most commonly at the 29(L) position (Fig. 4A). The reason that these particular hydrophobic residues are favorable in the β-bulge is not known, but the SH3 domain could provide a system for investigating this question. It is notable that three out of the five SH3 domain structures that do not display the classic β-bulge have large inserts in their N-Src loops

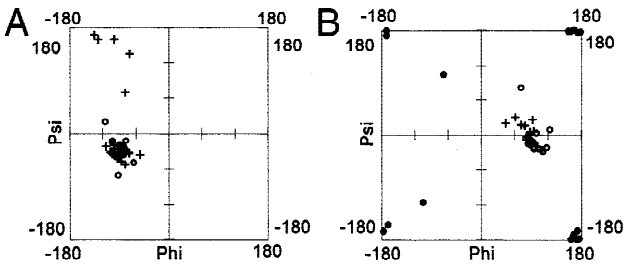

Fig. 6. Phi and psi backbone torsion angles observed in 18 different SH3 domain structures at (**A**) the 7(L) position (circles) and 29(L) position (crosses), and (**B**) the 23(G) (open circles), 45(G) (crosses), and 48(G) (filled circles) positions.

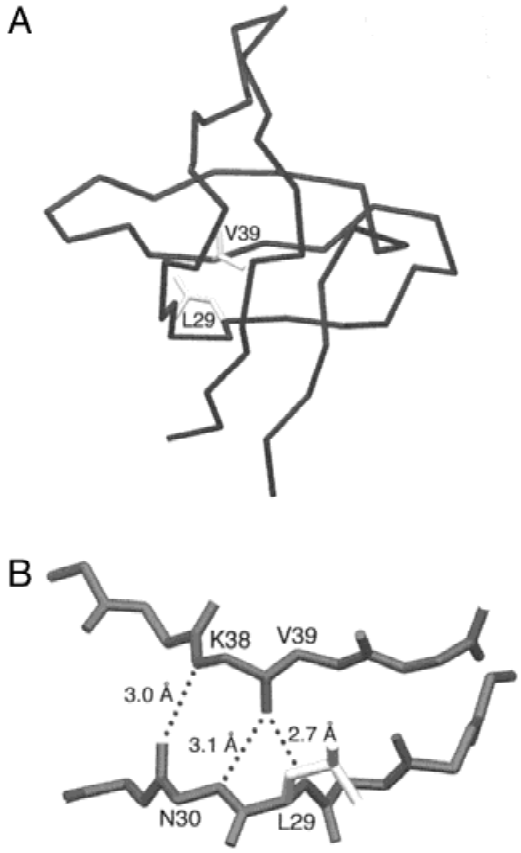

Fig. 7. Structure of the conserved β-bulge at the 29(L) position. **A:** The position of the 29(L) residue in the structure of the spectrin SH3 domain (1shg) is shown. The orientation of the molecular here is similar to that in Figure 1. The sharp change in direction of the polypeptide chain caused by the β-bulge can be clearly seen. **B:** A close-up view of the β-bulge region shows the characteristic hydrogen bonding interactions. For this view, the molecule was rotated approximately 180° around the y-axis compared to the view in **A**. Residue identities are those found in this particular domain, but numbering is according to that used in this work.

(amphiphysin and PI3-kinase) or are dimeric (Eps8) with the domain swap occurring at the N-Src loop.

Conserved glycine residues and turns

Three positions, 23(G), 45(G), and 48(G), are highly conserved, and are usually occupied by Gly. The 23(G) position is at the $i + 2$ position of the type II β-turn preceding β-strand *b*. Gly is seen at the $i + 2$ position of most type II β-turns because this position of the turn requires the backbone to adopt a left-handed helical conformation in which Gly is most favorable (Hutchinson & Thornton, 1994). The 23(G) position does occupy the left-handed helical region of the Ramachandran plot in 17 out of the 18 SH3 domain structures, all of which possess the conserved type II β-turn (Fig. 6B). This turn is almost invariably stabilized by a hydrogen bond between an Asp or Glu side chain at the 24(D) position and the NH of the 21(K) position (Fig. 8). The very high degree of conservation seen at the 23(G) and 24(D) positions (Table 1) suggests that this turn must be very important in the formation or maintenance of the SH3 domain structure. Mutagenesis studies have shown that some substitutions at the 24(D) position of the Fyn SH3 domain are very destabilizing (Maxwell & Davidson, 1998), and that both positions are important in folding kinetics (Grantcharova et al., 1998). However, a number of substitutions at position 24(D) with residues rarely observed in SH3 domains (e.g., Lys, Thr, and His) had little effect on in vitro stability or function (Maxwell & Davidson, 1998), suggesting that this position may have some important in vivo role, which is not reflected in in vitro assays. The consensus sequence and structure in the region surrounding 23(G) and 24(D) positions match that of a conserved local protein motif known as the "diverging type II β-turn" (Yi et al., 1998), which is consistent with the high frequency of occurrence of Asp and Glu the 24(D) position.

The 45(G) position, when occupied by Gly, also invariably adopts the left-handed helical conformation (Fig. 6B). This position lies in the Distal loop, which adopts a β-turn conformation in 17 of the 18 SH3 domain structures. Unlike the type II β-turn region mentioned above, a variety of β-turn types are seen at the Distal loop. Ten out of the 11 SH3 domain structures with Distal loop lengths of 3, 4, or greater display a type I β-turn at this position with 45(G) at the $i + 3$ position, while structures with Distal loop lengths of 2 possess types I, II and II′ β-turns with 45(G) at the $i + 2$ position in the latter two types. In the spectrin SH3 domain, the 45(G) position is occupied by an Asp residue. Substituting this position with Gly resulted in a considerable stabilization of the domain (Martinez et al., 1998). Interestingly, the Gly substitution actually changed the conformation of the β-turn from a type II′ in the wild-type structure to a type I′ in the mutant. Although the distal loop region of SH3 domains almost invariably adopts a β-turn conformation, there is considerable flexibility in the type of β-turn that is observed. In addition, although Gly residues at position 45 are aligned in our sequence alignment, they do not always align structurally in that the Gly residues can occupy the $i + 2$ or $i + 3$ positions of the β-turn, depending on the type of turn adopted in a particular structure.

The 48(G) position is particularly striking because it is one of the most conserved positions in the SH3 domain, yet it lies in the middle of a β-strand where Gly residues are generally disfavored (Minor & Kim, 1994; Smith et al., 1994). It has been postulated that this position requires Gly because any side chain here would clash with residues in the RT-Src loop, which is close by Borchert et al. (1994). This explanation may have some validity, but the 48(G) position also always adopts unusual ϕ/ψ angles that are more energetically favorable for Gly residues (Fig. 6B), suggesting that there are some highly conserved features of the SH3 fold causing an unusual backbone conformation at this position. We have found that amino acid substitutions at the 48(G) position in the Fyn SH3 domain greatly reduce stability and peptide binding activity (A.A. Di Nardo & A.R. Davidson, unpubl. results), underscoring the importance of this position in SH3 domain structure and function. Further studies will be required to elucidate the structural explanation for the importance of the 48(G) position.

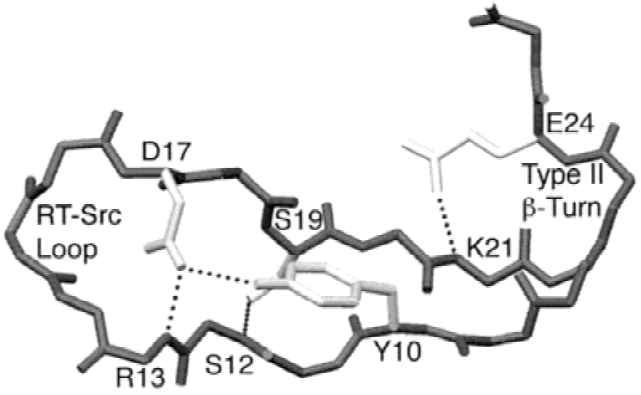

Fig. 8. Conserved side chain–main chain hydrogen bonds in the RT-Src loop and the type II β-turn. This region in the structure of the Src SH3 domain (1fmk) is shown. Residue identities are those found in this particular domain, but numbering is according to that used in this work.

Conserved interactions in the RT-Src loop

The conformation of the RT-Src loop appears to be stabilized by a number of highly conserved interactions. The side chain at the 9(D) position, which is occupied by Asp in 60% of SH3 domains, makes a hydrogen bond to the NH atom of the 22(K) position in all SH3 domain structures (Fig. 8). This hydrogen bond is seen even when the 9(D) position is occupied by Asn or Ser as it is in the Csk and Lck structures, respectively. The side chain of the 17(D) position, which is Asp or Glu 75% of the time, is hydrogen bonded to the NH atom of the 13(R) position in all SH3 domain structures. When Tyr occupies the 10(Y) position, an additional hydrogen bond is formed between Asp or Glu at the 17(D) position and the Tyr side chain (Fig. 8). Although the 19(S) position is less conserved than the above-mentioned positions, whenever it is Ser or Thr (as it is greater than 50% of the time in the alignment and in nine structures), its side chain forms a hydrogen bond with the NH atom of the 12(A) position. An Asp residue at the 19(S) position in the structure of the PI3-kinase SH3 domain also makes the same hydrogen bond. The requirement for the formation of a side chain to backbone NH hydrogen bond to position 12(A) when the 19(S) position is occupied by Ser or Thr is underscored by our finding through covariation analysis (Larson et al., 2000) that Ser and Thr are almost never found at the 19(S) position in combination with Pro at the 12(A) position. The 12(A) position, another conserved position in the RT-Src loop, is greater than 75% buried on average and is occupied by amino acids with small, neutral side chains (Pro, Ala, Gly, or Ser) greater than 80% of the time. The

12(A) position is in close contact with the side chains of the 10(Y) and 17(D) positions and is likely involved in maintaining their proper conformations. Larger residues at the 12(A) position would result in disruption of the conserved hydrogen bond between the 17(D) position and the backbone (discussed above), and would change the orientation of the 10(Y) side chain.

The large number of conserved interactions involving RT-Src loop positions underscores the importance of this region in mediating the peptide binding activity of the SH3 domain. Although the side chains of many conserved residues in this region make minimal amounts of contact with peptide, the intramolecular hydrogen bonds and packing interactions made by these residues may help to maintain the RT-Src loop in an optimal conformation for peptide binding. Alteration of these interactions could have significant effects on the affinity or specificity of the binding reaction.

The role of conserved residues in protein folding

A number of recent studies have examined sequence alignments to identify conserved positions that are crucial for the folding reaction of the protein in question (Michnick & Shakhnovich, 1998; Mirny et al., 1998; Ptitsyn, 1998). A high degree of conservation at certain positions whose roles are not obvious from structural examination could be explained if they were important for the folding reaction. For example, the 48(G) position is the third most conserved position in the SH3 domain alignment, yet it does not directly contact target peptide, nor does it make obvious stabilizing interactions in the domain; thus, one might postulate that it could be crucial for the folding of the SH3 domain. However, mutations at this position, though destabilizing, have been shown to have no effect on the folding rate of the Src or Fyn SH3 domains (Grantcharova et al., 1998; A.A. Di Nardo & A.R. Davidson, unpubl. obs.). On the other hand, the type II β-turn, which is highly conserved in the sequence alignment [i.e., the 23(G) and 24(D) positions] and in SH3 domain structures, has been shown to form early in the SH3 domain folding reaction (Grantcharova et al., 1998). Interestingly, this region of the Src SH3 domain (residues 20–26) has been shown to have a high propensity to form the native turn structure even as an isolated peptide (Yi et al., 1998). These data indicate that this conserved region may indeed form part of the folding nucleus. Substitution of residues in the distal loop region of both the Src and Spectrin SH3 domains have a large effect on protein folding kinetics (Grantcharova et al., 1998; Martinez et al., 1998), but both the sequence and structural conservation seen in this region is relatively low [except for the 45(G) position]. Overall, it does not appear that residues crucial for folding kinetics of the SH3 domain are particularly highly conserved. A lack of correlation between the effects of amino acid substitutions in the Spectrin and Src SH3 domains on folding kinetics (as quantitated by Φ-value analysis) and amino acid conservation has been clearly demonstrated in a study specifically examining this issue (Plaxco et al., 2000).

Conclusions

When a single structure of a protein or protein–ligand complex is examined, evaluation of the energetic or functional importance of any observed atomic interaction can be difficult. However, examination of structures of many homologues in combination with a careful sequence alignment analysis can provide much more definitive information. In this respect, the SH3 domain provides one of the best available subjects for investigation owing to the large numbers of SH3 domain sequences and structures available in the databases. In this study, by performing a thorough alignment and analysis of a large number of SH3 domain sequences and structures, we have been able to assign definite roles to most of the residues occupying the 30 most conserved positions in the domain (Table 1).

The high degree of sequence and structural conservation at certain positions in the SH3 domain indicates interactions likely to be crucial that were hitherto unrecognized as such. Some examples are the highly conserved side-chain to main-chain hydrogen bonds in the RT-Src loop and the β-bulge involving the 29(L) position. Future experiments can be directed at understanding how these conserved interactions and others identified in this work affect the stability or activity of the domain. In our own laboratory, the data accumulated in this study have already proven to be extremely useful in the design and interpretation of experiments. In studies on positions 24(D) and 41(L) of the Fyn SH3 domain and on the hydrophobic core of the same domain, we have found a good correlation between the stability of site directed mutants and the frequency with which the residue substituted occurs at the given position in our alignment (Maxwell & Davidson, 1998; A.A. Di Nardo, S.M. Larson, & A.R. Davidson, unpubl. obs.). In addition, we used this sequence alignment data to design amino acid substitutions in an SH3 domain from the yeast Abp1 protein that increased its T_m by 30 °C (Rath & Davidson, 2000).

This study has shown that thorough analysis of sequence and structural alignments can contribute a great deal to the understanding of the roles of conserved positions in the stability and activity of a protein. As the sequence and structural databases grow, more domains will have as many structures and sequences associated with them as does the SH3 domain. Future sequence and structural alignment studies on these domains will provide a crucial source of information for the design and interpretation of experiments in many fields of biology.

Materials and methods

Alignment assembly

An initial target sequence (fyn_chick) was used to initiate a PSI-BLAST search (two iterations, default values) of the nonredundant database compilation nrdb (current as of March 1999; Altschul et al., 1997). The set of homologous sequences was retrieved and the stretch of residues aligned to the initial target domain (i.e., the SH3 domain) was extracted from each protein sequence. Any redundant sequences in this set of homologues were removed and the remaining sequences were aligned using ClustalW(1.7) (Thompson et al., 1994). Secondary structure information from the fyn_chick file (SWISS-PROT) was used to set gap masks. The resulting alignment was then manually refined to arrange the alignment into large, ungapped blocks, with gap insertion favored in the loop regions. The exact boundaries for each block (i.e., the exact alignment of loop residues) were somewhat arbitrary. However, as the loop regions were not a focus of this study, any measure of uncertainty due to this arbitrary element in the alignment of loop residues is not significant in the final results. Once the initial alignment was constructed, an all vs. all comparison was used to calculate the average percent identity of each sequence compared to the rest of the alignment. Using this similarity measure as a guide, new target sequences were selected and used to initiate new

rounds of sequence gathering and alignment. This iterative process was continued until the result of each homology search produced no new SH3 domain sequences. Sequences with average identities in the range of 15–25%, in the twilight zone of sequence identity (Vogt et al., 1995), were carefully examined for SH3-like sequence properties before inclusion into the alignment [e.g., correct hydrophobic core positions, 36(W)/37(W/hydrophobic) signature, 23(G)/45(G)/48(G), 51(P)]. The new sequences at each round were added to the alignment, first by using ClustalW, and then by manually refining to minimize gap sizes. The final alignment was again checked thoroughly for redundancy and misalignment.

Our initial SH3 domain alignment contained 370 sequences, but to decrease sequence bias a smaller 266 sequence alignment that had been reduced to 90% redundancy (i.e., no two sequences are more than 90% identical; see Holm & Sander, 1998) was used in all analyses. Comparisons of weighted residue frequencies from the 370 sequence and 266 sequence alignments showed an average absolute change of only 0.35%, indicating that no information was lost in this reduction. The complete set of 266 sequences used in this study is included in Supplementary material in the Electronic Appendix.

Sequence weighting

The sequence alignment was weighted by the Henikoff algorithm, which downweights sequences according to how similar they are in comparison to the rest of the alignment (Henikoff & Henikoff, 1994). These sequence weights were used to calculate weighted residue frequencies at each position in the alignment. The weighted residue frequencies were used in all analyses to reduce the effects of sequence bias.

Residue conservation

To measure the level of conservation at each position in the alignment, the Shannon informational entropy at each position was calculated (Shenkin et al., 1991).

For position i, with residues $r = (A, C, D, \ldots, V, W, Y)$ occurring at frequencies $p_i(r)$, the entropy $H(i)$ is defined as

$$H(i) = -\sum_{r=A}^{Y} p_i(r)\ln(p_i(r)).$$

A more intuitive measure, positional entropy, is expressed as

$$N(i) = e^{H(i)}.$$

Structure comparison

Eighteen solved structures of SH3 domains were selected for alignment and analysis (in total, 19 different SH3 domain structures have been solved; the SH3 domain from Itk (1awj) was excluded from our analysis because the available NMR structure was low resolution). Of the SH3 domain structures that have been solved a number of times, the highest resolution X-ray crystal structure was used in each case. The structure of Eps8 (1aoj) was solved as a strand-swapped dimer (Kishan et al., 1997); for our analyses, a hybrid monomeric structure file (1eps) was created by incorporating the first half of chain A (K1–R33) and the second half of chain B (R33–T60). The structures were aligned using StrAlign (Akutsu & Tashimo, 1996) and renumbered according to the sequence alignment numbering to allow comparison of corresponding residues between structures. For each structure, the distance between side-chain centers-of-mass of each residue pair was calculated. The center-of-mass of each side chain was calculated by averaging the vectors from the alpha carbon to each atom of the side chain (for glycine, the C_α coordinates were used). Secondary structure profiles for each structure were calculated using DSSP (Kabsch & Sander, 1983).

To calculate residue burial in the absence and presence of peptide or protein ligand, a different subset of structures was used. Fourteen peptide-bound SH3 domain structures (1abo, 1azg, 1cka, 1ckb, 1fyn, 1gbq, 1nlo, 1nlp, 1prm, 1qwe, 1qwf, 1rlq, 1sem, 3gbq), one protein-bound SH3 domain structure (1efn), and two complete kinase structures (1ad5, 1fmk) were analyzed. Percent burial for each residue was calculated using the Environments program (Bowie et al., 1991) for both the complete structure (i.e., with peptide, protein, or the whole kinase) and the SH3 domain alone (these structure files were created by simply deleting all atoms except those in the SH3 domain).

Supplementary material in the Electronic Appendix

The electronic supplementary material consists of one file, named sh3.ali, containing all 266 sequences included in the SH3 domain alignment analyzed in this work. The other file, named sh3.frq, contains the weighted frequencies of occurrence of each amino acid at each position in the alignment. These files are saved as tab delimited text.

Acknowledgments

We thank Kevin Plaxco, Lydia Gregoret, and Karen Maxwell for helpful comments on the manuscript. This work was supported by a grant from the Medical Research Council of Canada. S.M.L. was supported by an Ontario Graduate Student Scholarship.

References

Akutsu T, Tashimo H. 1996. Protein structure comparison using representation by line segment sequences. *Pac Symp Biocomput* 25–40.

Alexandropoulos K, Cheng G, Baltimore D. 1995. Proline-rich sequences that bind to Src homology 3 domains with individual specificities. *Proc Natl Acad Sci USA 92*:3110–3114.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res 25*:3389–3402.

Arold S, Franken P, Strub MP, Hoh F, Benichou S, Benarous R, Dumas C. 1997. The crystal structure of HIV-1 Nef protein bound to the Fyn kinase SH3 domain suggests a role for this complex in altered T cell receptor signaling. *Structure 5*:1361–1372.

Bashford D, Chothia C, Lesk AM. 1987. Determinants of a protein fold. Unique features of the globin amino acid sequences. *J Mol Biol 196*:199–216.

Benner SA, Jenny TF, Cohen MA, Gonnet GH. 1994. Predicting the conformation of proteins from sequences. Progress and future progress. *Adv Enzyme Regul 34*:269–353.

Borchert TV, Mathieu M, Zeelen JP, Courtneidge SA, Wierenga RK. 1994. The crystal structure of human CskSH3: Structural diversity near the RT-Src and n-Src loop. *FEBS Lett 341*:79–85.

Bowie JU, Luthy R, Eisenberg D. 1991. A method to identify protein sequences that fold into a known three-dimensional structure. *Science 253*:164–170.

Chan AW, Hutchinson EG, Harris D, Thornton JM. 1993. Identification, classification, and analysis of beta-bulges in proteins. *Protein Sci 2*:1574–1590.

Chen YJ, Lin SC, Tzeng SR, Patel HV, Lyu PC, Cheng JW. 1996. Stability and folding of the SH3 domain of Bruton's tyrosine kinase. *Proteins 26*:465–471.

Chothia C, Gelfand I, Kister A. 1998. Structural determinants in the sequences of immunoglobulin variable domain. *J Mol Biol 278*:457–479.

Dalgarno DC, Botfield MC, Rickles RJ. 1997. SH3 domains and drug design: Ligands, structure, and biological function. *Biopolymers 43*:383–400.
Dill KA. 1990. Dominant forces in protein folding. *Biochemistry 29*:7133–7155.
Fauchère J-L, Pliska V. 1983. Hydrophobic parameters π of amino-acid side chains from the partitioning of N-acetyl-amino-acid amides. *Eur J Med Chem 18*:369–375.
Feng S, Chen JK, Yu H, Simon JA, Schreiber SL. 1994. Two binding orientations for peptides to the Src SH3 domain: Development of a general model for SH3-ligand interactions. *Science 266*:1241–1247.
Filimonov VV, Azuaga AI, Viguera AR, Serrano L, Mateo PL. 1999. A thermodynamic analysis of a family of small globular proteins: SH3 domains. *Biophys Chem 77*:195–208.
Grantcharova VP, Baker D. 1997. Folding dynamics of the src SH3 domain. *Biochemistry 36*:15685–15692.
Grantcharova VP, Riddle DS, Santiago JV, Baker D. 1998. Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. *Nat Struct Biol 5*:714–720.
Henikoff S, Henikoff JG. 1994. Position-based sequence weights. *J Mol Biol 243*:574–578.
Holm L, Sander C. 1998. Removing near-neighbour redundancy from large protein sequence collections. *Bioinformatics 14*:423–429.
Hooft RW, Vriend G, Sander C, Abola EE. 1996. Errors in protein structures. *Nature 381*:272.
Hutchinson EG, Thornton JM. 1994. A revised set of potentials for beta-turn formation in proteins. *Protein Sci 3*:2207–2216.
Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers 22*:2577–2637.
Kishan KV, Scita G, Wong WT, Di Fiore PP, Newcomer ME. 1997. The SH3 domain of Eps8 exists as a novel intertwined dimer. *Nat Struct Biol 4*:739–743.
Larson SM, Di Nardo AA, Davidson AR. 2000. Analysis of covariation in an SH3 domain sequence alignment: Applications in tertiary contact prediction and the design of compensating hydrophobic core substitutions. *J Mol Biol 303*:433–446.
Lee CH, Saksela K, Mirza UA, Chait BT, Kuriyan J. 1996. Crystal structure of the conserved core of HIV-1 Nef complexed with a Src family SH3 domain. *Cell 85*:931–942.
Lesk AM, Fordham WD. 1996. Conservation and variability in the structures of serine proteinases of the chymotrypsin family. *J Mol Biol 258*:501–537.
Lim WA. 1996. Reading between the lines: SH3 recognition of an intact protein. *Structure 4*:657–659.
Lim WA, Richards FM. 1994. Critical residues in an SH3 domain from Sem-5 suggest a mechanism for proline-rich peptide recognition. *Nat Struct Biol 1*:221–225.
Lim WA, Richards FM, Fox RO. 1994. Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains. *Nature 372*:375–379.
Martinez JC, Pisabarro MT, Serrano L. 1998. Obligatory steps in protein folding and the conformational diversity of the transition state. *Nat Struct Biol 5*:721–729.
Maxwell KL, Davidson AR. 1998. Mutagenesis of a buried polar interaction in an SH3 domain: Sequence conservation provides the best prediction of stability effects. *Biochemistry 37*:16172–16182.
Mayer BJ, Hamaguchi M, Hanafusa H. 1988. A novel viral oncogene with structural similarity to phospholipase C. *Nature 332*:272–275.
Michnick SW, Shakhnovich E. 1998. A strategy for detecting the conservation of folding-nucleus residues in protein superfamilies. *Fold Des 3*:239–251.
Minor D Jr, Kim PS. 1994. Measurement of the beta-sheet-forming propensities of amino acids. *Nature 367*:660–663.
Mirny LA, Abkevich VI, Shakhnovich EI. 1998. How evolution makes proteins fold quickly. *Proc Natl Acad Sci USA 95*:4976–4981.
Pawson T. 1995. Protein modules and signalling networks. *Nature 373*:573–580.
Pisabarro MT, Serrano L. 1996. Rational design of specific high-affinity peptide ligands for the Abl-SH3 domain. *Biochemistry 35*:10634–10640.
Plaxco KW, Guijarro JI, Morton CJ, Pitkeathly M, Campbell ID, Dobson CM. 1998. The folding kinetics and thermodynamics of the Fyn-SH3 domain. *Biochemistry 37*:2529–2537.
Plaxco KW, Larson SM, Ruczinski I, Riddle DS, Thayer EC, Buchwitz B, Davidson AR, Baker D. 2000. Evolutionary conservation in protein folding kinetics. *J Mol Biol 298*:303–312.
Ptitsyn OB. 1998. Protein folding and protein evolution: Common folding nucleus in different subfamilies of c-type cytochromes? *J Mol Biol 278*: 655–666.
Rath A, Davidson AR. 2000. The design of a hyperstable mutant of the Abplp SH3 domain by sequence alignment analysis. 2000. *Protein Sci 9*. Forthcoming.
Rickles RJ, Botfield MC, Zhou XM, Henry PA, Brugge JS, Zoller MJ. 1995. Phage display selection of ligand residues important for Src homology 3 domain binding specificity. *Proc Natl Acad Sci USA 92*:10909–10913.
Rost B, Sander C. 1994. Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins 19*:55–72.
Shenkin PS, Erman B, Mastrandrea LD. 1991. Information-theoretical entropy as a measure of sequence variability. *Proteins 11*:297–313.
Sicheri F, Moarefi I, Kuriyan J. 1997. Crystal structure of the Src family tyrosine kinase Hck. *Nature 385*:602–609.
Smith CK, Withka JM, Regan L. 1994. A thermodynamic scale for the beta-sheet forming tendencies of the amino acids. *Biochemistry 33*:5510–5517.
Sparks AB, Rider JE, Hoffman NG, Fowlkes DM, Quillam LA, Kay BK. 1996. Distinct ligand preferences of Src homology 3 domains from Src, Yes, Abl, Cortactin, p53bp2, PLCgamma, Crk, and Grb2. *Proc Natl Acad Sci USA 93*:1540–1544.
Steipe B, Schiller B, Pluckthun A, Steinbacher S. 1994. Sequence statistics reliably predict stabilizing mutations in a protein domain. *J Mol Biol 240*:188–192.
Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res 22*:4673–4680.
Viguera AR, Arrondo JL, Musacchio A, Saraste M, Serrano L. 1994. Characterization of the interaction of natural proline-rich peptides with five different SH3 domains. *Biochemistry 33*:10925–10933.
Vogt G, Etzold T, Argos P. 1995. An assessment of amino acid exchange matrices in aligning protein sequences: The twilight zone revisited. *J Mol Biol 249*:816–831.
Weng Z, Rickles RJ, Feng S, Richard S, Shaw AS, Schreiber SL, Brugge JS. 1995. Structure–function analysis of SH3 domains: SH3 binding specificity altered by single amino acid substitutions. *Mol Cell Biol 15*:5627–5634.
Xu W, Harrison SC, Eck MJ. 1997. Three-dimensional structure of the tyrosine kinase c-Src. *Nature 385*:595–602.
Yi Q, Bystroff C, Rajagopal P, Klevit RE, Baker D. 1998. Prediction and structural characterization of an independently folding substructure in the src SH3 domain. *J Mol Biol 283*:293–300.