doi:10.1006/jmbi.2000.3670 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2000) 298, 313–328

JMB

SH3-SPOT: An Algorithm to Predict Preferred Ligands to Different Members of the SH3 Gene Family

Barbara Brannetti, Allegra Via, Gianluca Cestra, Gianni Cesareni and Manuela Helmer Citterich*

Centro di Bioinformatica Molecolare, Department of Biology, University of Rome ''Tor Vergata'', via della Ricerca Scientifica 00133, Rome, Italy

*Corresponding author

We have developed a procedure to predict the peptide binding specificity of an SH3 domain from its sequence. The procedure utilizes information extracted from position-specific contacts derived from six SH3/peptide or SH3/protein complexes of known structure. The framework of SH3/peptide contacts defined on the structure of the complexes is used to build a residue-residue interaction database derived from ligands obtained by panning peptide libraries displayed on filamentous phage.

The SH3-specific interaction database is a multidimensional array containing frequencies of position-specific contacts. As input, SH3-SPOT requires the sequence of an SH3 domain and of a query decapeptide ligand. The array, that we call the SH3-specific matrix, is then used to evaluate the probability that the peptide would bind the given SH3 domain. This procedure is fast enough to be applied to the entire protein sequence database.

Panning experiments were performed to search putative specific ligands of different SH3 domains in a database of decapeptides, or in a database of protein sequences. The procedure ranked some of the natural partners of interaction of a number of SH3 domains among the best ligands of the $\sim 5.6 \times 10^9$ different decapeptides in the SWISSPROT database. We expect the predictive power of the method to increase with the enrichment of the SH3-specific matrix by interaction data derived from new complex structures or from the characterization of new ligands. The procedure was developed using the SH3 domain family as test case but its application can easily be extended to other families of protein domains (such as, SH2, MHC, EH, PDZ, etc.).

Keywords: SH3 domain; prediction; specificity; protein/peptide interaction; bioinformatics

Introduction

The formation of macromolecular complexes in the cell is often mediated by relatively small protein recognition modules which are found repeatedly in several protein structures. Each domain family (SH3, SH2, EH, PDZ, WW, etc.) binds to peptides displaying specific sequence and structural characteristics. For instance, SH2 binds to peptides containing phosphotyrosine residues while PDZ prefers peptides with a free carboxy terminus. Within each domain family, specificity is modulated by varying the chemical properties of the module binding surface, which in turn determines a preference for peptides within a range of variability on the common structural theme.

Among these module families, SH3 is the best represented in the proteins characterized so far. A ψBLAST search (Altschul *et al*., 1997) in the SWISSPROT database reveals more than 250 hits. These proteins are involved in a variety of cellular processes as diverse as signal transduction, cytoskeletal organization, membrane traffic or organelle assembly (Stahl *et al*., 1988; Trahey *et al*., 1988; Sahr *et al*, 1990; Otsu *et al*., 1991; Musacchio *et al*., 1994a).

All SH3 domains fold into a compact structure formed by two antiparallel, four-stranded β-sheets connected by loops of different length (Musacchio

Abbreviations used: SH3, Src homology 3.

e-mail address of the corresponding author: citterich@uniroma2.it

0022-2836/00/020313–16 $35.00/0

et al., 1992). The two β-sheets interact at a right-angle to form a modified β-barrel.

The solution of the structure of several SH3 domains complexed with peptide ligands and the analysis of preferred SH3 ligands, as determined by screening peptide repertoires, has led us to propose a general model of peptide recognition mediated by SH3 domains.

According to this model, SH3 ligands are characterized by a PxxP motif and recognize the SH3 binding surface in a polyproline helix type II conformation in one of two opposite orientations. The two proline residues of the core motif occupy two hydrophobic pockets, formed by aromatic residues that are conserved in most SH3 domains. The third binding pocket is lined by negative residues and can host a positively charged side-chain. Ligand orientation depends on the position of the positive residue in the target peptide. Peptides that bind in a type I orientation conform to the consensus Rx**LP**x**ZP** (where Z is normally a hydrophobic or an Arg residue as in PI3K binding peptides), while peptides characterized by the **#P**x**#P**xR consensus (type II, where # is a hydrophobic residue) bind in the opposite orientation. The residues in bold occupy the sites in the hydrophobic pocket that are normally referred to as P_{-1}, P_0 P_{+2} and P_{+3} (Figure 1), while the Arg residue occupies site P_{-3}. Different SH3 domains display different preferences for more or less subtle variations of the common motif on the polyproline peptide. A notable exception to this model is represented by the SH3 of Abl which prefers peptides that have an aromatic residue instead of the positively charged residue at P_{-3}. These peptides are positioned in the SH3 binding pocket in a type I orientation. More recently, a growing number of SH3 domains have been reported to bind preferentially to peptides that cannot be easily reconciled with this model (Cestra *et al.*, 1999; Mongiovì *et al.*, 1999).

SH3 protein recognition can also be modulated by contacts with other target regions that may be located far from the polyproline peptide (Lee *et al.*, 1996). Furthermore, *in vivo* other factors like protein localization in a specific cell compartment and increase of local concentration, mediated by other protein interactions, are very important in determining whether a specific SH3 will ever encounter its ''*in vitro* predicted'' partner. These latter considerations are particularly important in the case of SH3 domains that bind to their partners with relatively low affinities (ranging from 100 nM to 10 μM).

To summarize, *in vivo* partner selection, mediated by a specific protein recognition module, is influenced by three factors: (1) peptide binding specificity; (2) stabilization by extra contacts; and (3) local partner concentration.

Theoretical methods have been applied to the analysis of SH3 domains specificity such as: computer simulation of protein-protein interaction (Pisabarro & Serrano, 1996; Pisabarro *et al.*, 1998) or protein docking programs (unpublished results).

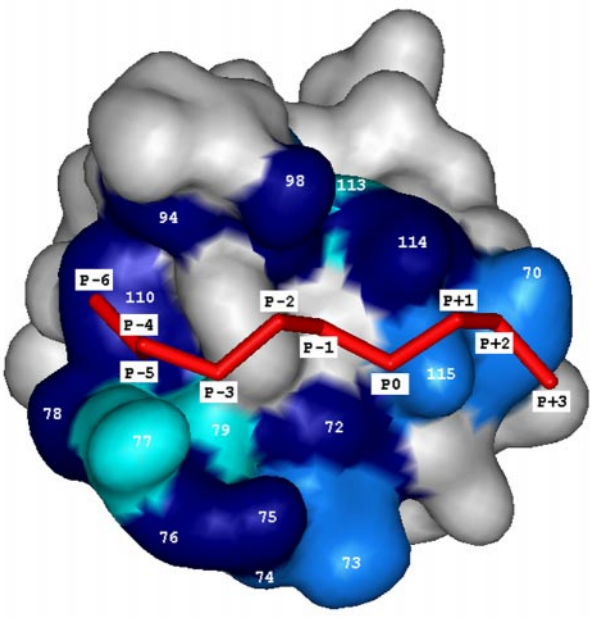

Figure 1. SH3 domain and peptide-contacting positions. The Abl SH3 domain and peptide complex was used to describe a generic SH3/peptide contact. The SH3 domain is represented with a solvent-accessible surface generated with the WebLab Pro program, M.S.I. The trace of the peptide is in red. The positions identified in the SH3 multiple alignment are reported in white on the SH3 surface and the positions from P_{-6} to P_{+3} are reported in white boxes in the corresponding positions of the peptide. The SH3 residues which were shown to correlate with the preference for a particular ligand residue, as discussed in Table 2, are in color; cyan and blue regions determine recognition specificity only of class I and class II peptides, respectively. Residues that were found to be important for recognition specificity in both ligand classes are in dark blue.

These methods can be applied only when the 3D structures of the protein moieties are known, or, alternatively, when reliable molecular models can be obtained.

We set as our goal the development of an algorithm (SPOT: **S**pecificity **P**rediction **O**f **T**arget) that could predict peptide recognition specificity of any newly discovered SH3 domain from its sequence. The information extracted from the structure of SH3/peptide complexes and from the analysis of multiple alignments is used to organize the experimental data obtained by panning peptide repertoires displayed on filamentous phage capsids.

The results of the panning experiments are organized in contact matrices which describe, for each SH3/peptide contact position, the frequency of occurrence of any specific pair of residues in the SH3 and in the ligand peptide. The matrices are then used to evaluate the probability of interaction between different SH3 domains and a given peptide (Figure 2).

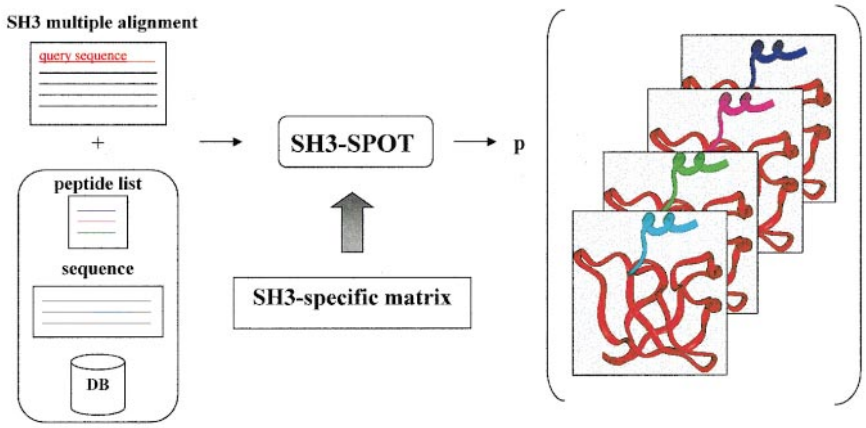

Figure 2. Scheme of the SH3-SPOT procedure. The procedure takes as input a multiple alignment of SH3 sequences (the sequence of the query SH3 is in red) and the database of decapeptides or proteins to be evaluated for affinity to the query SH3 sequence. An SH3-specific matrix is built from contact data extracted from SH3/peptide complexes of known structure and from phage display data (Table 1), as described in Methods. The SH3-specific matrix associates to each SH3/peptide contact position the frequency of occurrence of any specific pair of residues extracted from the database of experimental contacts. The matrix is then used to evaluate the probability of interaction between different SH3 domains and a given peptide.

SPOT can be applied to any family of protein recognition domains for which both structural information about at least one protein-peptide complex and binding data between different domain-peptide sequences are available.

Results

Class I and class II contact matrices

In order to construct contact matrices we identified the positions, on the SH3 binding surface, that are sufficiently close to the ligand and may therefore contribute to binding specificity. These are likely to vary in the different SH3s and may depend on the specific ligand considered. To avoid missing residues that may turn out to be important in defining ligand preference in SH3 domains of unknown structure, we have used a somewhat loose definition of residue contact (see Methods). Two residues are defined as being in contact when, in any of the six complexes of known crystallographic structure, the shortest distance between their atoms is less than the sum of their van der Waals radii +3 Å. Twenty-seven positions of the multiple alignment of SH3 sequences shown in Figure 3(a) are in contact with at least one residue of a ligand according to this criterion.

Thus, an SH3/ligand contact map is a 27×10 matrix in which most elements are empty, while those corresponding to residues that are in ''contact'' contain information about the frequency of specific amino acid to amino acid contact in experimentally determined ligands.

Although C^{α} atoms of class I and class II peptides occupy similar positions in the SH3 binding pockets, we have considered the two types of complexes separately. This approach generated more accurate predictions than a similar approach in which the two types of complexes were considered together. Overall the two contact maps are similar. However, some differences stand out as shown in Figure 3(b) in which ligand residues are named according to Lim *et al.* (1994) and SH3 residues according to the numbering in the Abl SH3 (Musacchio *et al.*, 1994b): (1) only class I peptides are found to interact with the SH3 domain by making contacts at positions P_{-6} and P_{-5} in the SH3/peptide complexes of known structure; (2) positions 77, 84, 93, 95, 100, and 101 of the SH3 multiple alignment make contact only with class I peptides; (3) positions 73, 97 and 116 of the multiple alignment make contact only with class II ligands.

Some of these differences may be due to the relatively small size of the database of SH3/peptide complexes available at the moment. This is particularly true for the absence of hits at positions P_{-5} and P_{-6} in the class II contact map. In three out of the four class II structures considered, in fact, the peptides in the complex are shorter and do not include residues at positions P_{-5} and P_{-6}. In the 1efn structure (where the SH3 domain of the Hck kinase is complexed with the HIV Nef protein) the residues of the Nef protein corresponding to the positions P_{-5} and P_{-6} are not in contact with the SH3 residues. Since panning of peptides displayed by phage capsids reveals that recognition specificity of class II peptides may be modulated by the residues present at P_{-5} and P_{-6}, we decided to consider contacts between P_{-5} and SH3 positions 77, 78, 99 and 110, as well as contacts between P_{-6} and SH3 positions 99 and 110 to the class II array, as explained in Methods.

Analysis of the frequencies of specific contacts

The class I and class II SH3-specific matrices (Figure 3(b)) were calculated from the experimen-

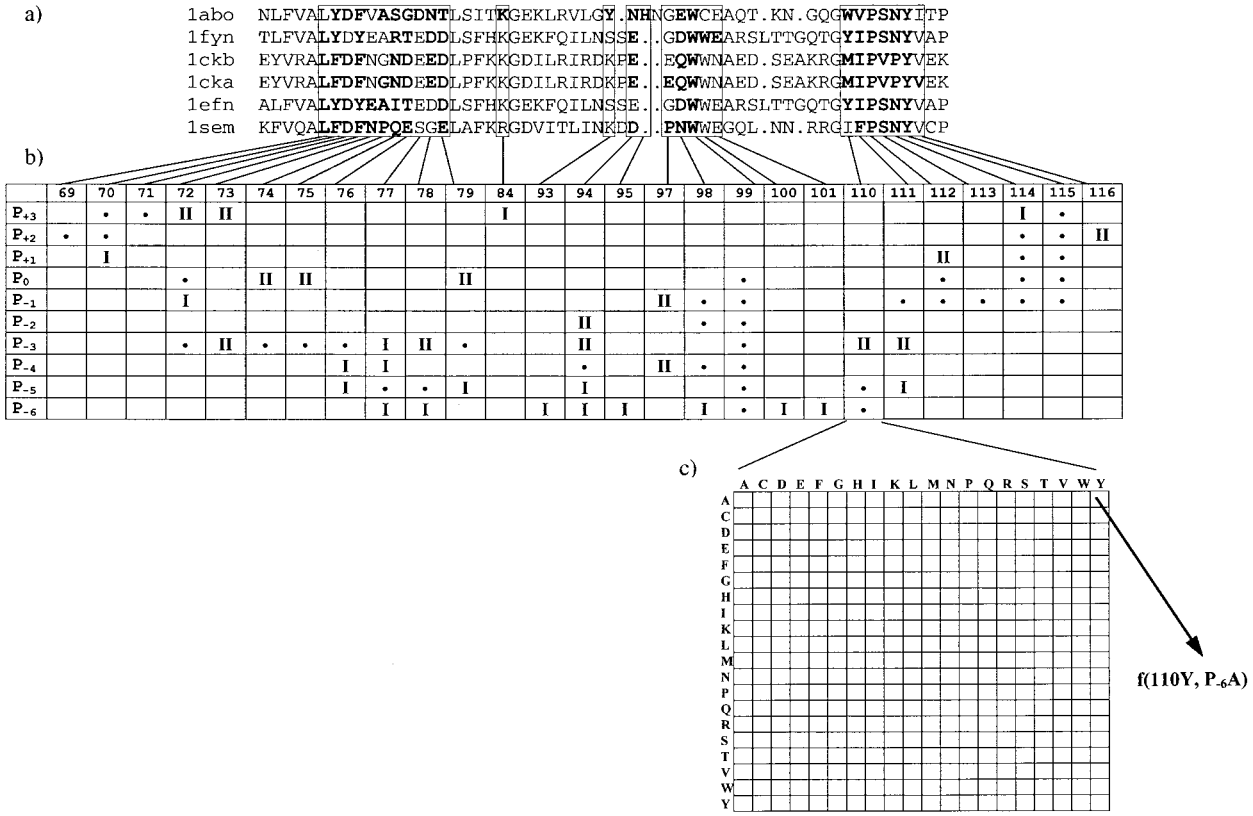

Figure 3. Construction of SH3-specific matrices. (a) Multiple alignment of the SH3 domain sequences belonging to PDB complexes. The SH3 residues in contact with residues of the peptide are shown in bold. The 27 positions of the multiple alignment of SH3 sequences containing at least one SH3 contacting residue are boxed. (b) The columns of the table describe the 27 contacting position of the SH3 multiple alignment and are identified with the residue numbering of the Abl SH3 structure (Musacchio *et al*., 1994b); the rows refer to the peptide positions with the P_{+3} ,..., P_{-6} nomenclature as defined by Lim *et al*. (1994). The open elements of the table identify SH3 and peptide positions which are in contact (see Methods). The elements corresponding to SH3 and peptide positions which are not in contact are shown in grey. Contacts which are common to class I and class II complexes are marked with a dot. Contacts which are specific of the class I and class II complexes are marked with I and II. (c) Each one of the white elements of the SH3/peptide contact map shown in (b) contains a 20 × 20 matrix where we calculate the frequency of interaction of all the possible residue/residue pairs in the defined SH3/peptide position. The matrix in (c) describes the interaction between SH3 position 110 and peptide position P_{-6}. In the AY element, we tabulate the frequency of occurrence of Tyr in position 110 of SH3 sequences and of Ala in the P_{-6} position of the peptide ligands.

tally determined frequency of occurrence of amino acid *x* and *y* at any specific element of the contact map. These data were obtained from published and unpublished experiments of phage display and include 319 ligand peptides selected from 25 different SH3 domains, as described in Methods (Table 1). After normalization the maximum value, relative to a residue pair in a specific contact position is equal to 1. Due to the relatively small amount of contact information available, some of the figures in the two matrices have different degrees of statistical significance. In other words, a high figure at position A*x*A*y* can be either derived from a single observation of a ligand containing amino acid *x* in contact with amino acid *y* on the SH3 binding surface or alternatively it can be obtained from several SH3 and ligands always showing *x* and *y* at those positions. More experimental data are needed to eliminate this inconsistency.

Inspection of the values in the SH3-specific matrix reveals some correlation between residues in the SH3 and in the ligand that might turn out to be important in determining recognition specificity. The most significant are reported in Table 2.

Extracting contact information from the SH3-specific matrices

The use of interaction matrices, such as the one that we have described, to predict domain-peptide binding is equivalent to describing the interaction between two proteins as the sum of independent interactions between their contacting residues.

In order to assess the reliability and limits of this assumption and the predictive power of the SH3-

Table 1. Database of SH3 domain complexes and number of selected peptides.

PDB code	SH3 protein name	Number of class I peptides	Number of class II peptides	SH3 source	Reference
A.					
1abo	Abl	1	-	*Mus musculus*	Musacchio *et al.* (1994b)
1cka, 1ckb	c-Crk	-	2	*Mus musculus*	Wu *et al.* (1995)
1efn	fyn mutant R96I Tyrosine kinase	-	1	*Homo sapiens*	Lee *et al.* (1996)
1fyn	Fyn	1	-	*Homo sapiens*	Musacchio *et al.* (1994b)
1sem	Sem-5	-	1	*Mus musculus*	Lim *et al.* (1994)
B.					
1abo	Abl Tyrosine kinase	30	-	*Mus musculus*	Sparks *et al.* (1996); Rickles *et al.* (1994)
1ad5	Hck	7	-	*Homo sapiens*	Cesareni *et al.* (unpublished)
-	Amphiphysin	-	12	*Homo sapiens*	Cestra *et al.* (1999)
-	Cortactin	-	11	*Mus musculus*	Sparks *et al.* (1996)
-	c-Crk	-	20	*Gallus gallus*	Sparks *et al.* (1996)
-	Endophilin1	3	-	*Homo sapiens*	Cestra *et al.* (1999)
-	Endophilin2	3	-	*Homo sapiens*	Cestra *et al.* (1999)
-	Endophilin3	9	-	*Homo sapiens*	Cestra *et al.* (1999)
1fmk	c-Src	56	10	*Homo sapiens*	Sparks *et al.* (1996); Rickles *et al.* (1994, 1995)
1fyn	Fyn	13	-	*Homo sapiens*	Rickles *et al.* (1994, 1995)
1gbq	Grb2	19	13	*Mus musculus*	Sparks *et al.* (1996)
-	Itk	5	-	*Homo sapiens*	Bunnell *et al.* (1996)
-	Lyn	13	-	*Homo sapiens*	Rickles *et al.* (1994, 1995)
1ycs	P53bp2	-	13	*Homo sapiens*	Sparks *et al.* (1996)
1pht	Pi3	24	4	*Homo sapiens*	Rickles *et al.* (1994, 1995)
2hsp	Plc-γ	-	11	*Homo sapiens*	Sparks *et al.* (1996)
-	myo3	8	-	*S. cerevisiase*	Unpublished results
-	myo5	13	-	*S. cerevisiase*	Unpublished results
1shg	Spectrin	5	-	*Homo sapiens*	Unpublished results
-	Yes	11	-	*Homo sapiens*	Sparks *et al.* (1996); Rickles *et al.* (1995)
	TOTAL	221	98		

specific matrices, we decided to: (1) evaluate the correlation between experimental and SPOT-predicted results on a defined dataset (see Methods); (2) compare the performance of different matrices obtained: (i) with all the available interaction data; and (ii) selectively removing data about single SH3 domains and performing predictions on SH3 domains which were or were not included in the experimental data utilized to calculate the matrices.

We want to assess whether the addition of new data to the matrices affects their performance in prediction and whether we can obtain a reliable prediction for an SH3 domain of unknown specificity.

As a measure of success, we decided to use the correlation between the peptide sequences found in the phage display data (input) and the best scoring peptides (output) obtained using the different SH3-specific matrices. Each peptide list (either the input or the output list) can be described as a 10×20 table (Figure 4(a) and (b)). The ten columns describe the ten ligand positions while in the rows we report the frequency with which the specific amino acid residues are observed in the ligand peptides at that position.

The correlation between the input and output 10×20 tables is calculated as described in Methods and shown in Figure 4(c). In order to evaluate the contribution of the different SH3 interaction data, we built different SH3-specific matrices (Table 3 and Figure 5) derived from the whole dataset of SH3/peptide interactions and from a dataset deprived of the interaction data derived from different SH3 domains. Each prediction was evaluated by calculating the correlation between the experimental and predicted results (see Table 3 and Figure 5).

The test performed with the Src (1fmk) SH3 domain as query sequence, using interaction data derived from experiments performed with different members of the Src family, shows that: when no interaction data about other SH3 domains of the Src family contribute to the matrix, the correlation between experimental and SPOT-predicted results is low ($\sim$0.6); the addition of an increasing number of experimental results obtained with SH3s of the Src family improves the correlation to 0.97. Interestingly a high correlation is also observed when the Src SH3 experimental data are not included in the training dataset. Similar results were obtained with the other SH3 domains of the Src family, as shown in Table 3 and Figure 5(a).

The Abl SH3 domain binds to polyproline peptides that are different in sequence from the classical RxxPxxP Src-type peptide. However, we could not detect a negative interference between Abl interaction data and data obtained with other SH3

a)

Consensus input

	P_{-6}	P_{-5}	P_{-4}	P_{-3}	P_{-2}	P_{-1}	P_0	P_{+1}	P_{+2}	P_{+3}
A	.26	.00	.13	.00	.16	.10	.00	.13	.03	.00
C	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
D	.00	.00	.03	.00	.00	.00	.00	.03	.00	.00
E	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
F	.00	.00	.00	.10	.00	.00	.00	.00	.00	.00
G	.10	.00	.03	.00	.03	.00	.00	.00	.00	.00
H	.00	.00	.10	.00	.03	.00	.00	.03	.00	.00
I	.00	.00	.00	.00	.00	.00	.00	.13	.00	.00
K	.00	.00	.06	.03	.00	.00	.00	.03	.00	.00
L	.00	.00	.03	.00	.03	.03	.00	.03	.03	.00
M	.00	.00	.00	.03	.00	.00	.00	.00	.00	.00
N	.00	.00	.03	.00	.00	.00	.00	.00	.00	.00
P	.45	.87	.13	.00	.45	.74	1.00	.42	.26	1.00
Q	.00	.00	.03	.00	.00	.00	.00	.00	.00	.00
R	.03	.03	.06	.06	.00	.00	.00	.00	.03	.00
S	.00	.00	.13	.00	.16	.00	.00	.00	.00	.00
T	.00	.00	.13	.00	.03	.00	.00	.00	.00	.00
V	.00	.00	.00	.00	.00	.03	.00	.00	.19	.00
W	.00	.00	.03	.19	.00	.00	.00	.00	.00	.00
Y	.00	.00	.00	.45	.00	.00	.00	.00	.00	.00

b)

Consensus output

	P_{-6}	P_{-5}	P_{-4}	P_{-3}	P_{-2}	P_{-1}	P_0	P_{+1}	P_{+2}	P_{+3}
A	.23	.00	.17	.00	.13	.03	.00	.13	.07	.00
C	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
D	.00	.00	.07	.00	.00	.00	.00	.00	.00	.00
E	.00	.00	.00	.00	.00	.00	.00	.03	.00	.00
F	.00	.00	.03	.07	.00	.00	.00	.00	.00	.00
G	.07	.00	.03	.00	.07	.00	.00	.00	.00	.00
H	.00	.00	.10	.00	.03	.00	.00	.03	.00	.00
I	.00	.00	.00	.00	.00	.00	.00	.00	.10	.00
K	.00	.00	.03	.00	.00	.00	.00	.03	.00	.00
L	.03	.00	.00	.00	.00	.13	.00	.03	.07	.00
M	.00	.00	.00	.07	.00	.00	.00	.00	.00	.00
N	.00	.00	.03	.03	.00	.00	.00	.00	.00	.00
P	.57	1.00	.13	.00	.53	.83	1.00	.37	.20	1.00
Q	.00	.00	.03	.03	.00	.00	.00	.00	.00	.00
R	.03	.00	.03	.10	.00	.00	.00	.00	.03	.00
S	.00	.00	.17	.00	.20	.00	.00	.00	.00	.00
T	.00	.00	.13	.00	.03	.00	.00	.00	.00	.00
V	.00	.00	.00	.00	.00	.00	.00	.03	.20	.00
W	.00	.00	.03	.20	.00	.00	.00	.00	.00	.00
Y	.00	.00	.00	.53	.00	.00	.00	.00	.00	.00

c)

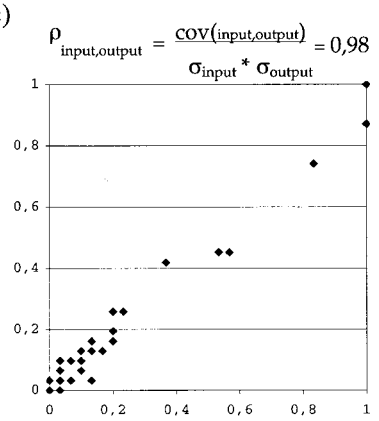

Figure 4. Correlation between experimental and SPOT-predicted results with Abl SH3 used as query sequence. (a) Table describing the frequency of occurrence of the different peptide residues in the ten positions studied. In this example, the peptide list of the Abl peptides selected by the phage display panning technique is described; it is easy to see that the Abl peptides always display a proline residue in the P_0 and P_{+3} positions of the peptide, corresponding to the two proline residues in the canonic PxxP consensus sequence, respectively. There is a strong preference for a Pro residue in position P_{-1} and P_{-2}. A preference for aromatic residues is displayed at position P_{-3} where other SH3 domains usually bind an Arg residue in the class I orientation (Lim *et al.*, 1994). (b) Frequencies of the different residues in the ten positions of the 30 best scoring peptides SPOT-selected from the Abl SH3 domain in the database of 319 peptides. (c) A comparison can be made between the table describing the input peptide list, shown in (a), and the table obtained from the 30 best scoring peptides of the different test performed, shown in (b). The correlation ρ between the input and output tables is calculated according to the formula shown and as described in Methods.

domains of the Src family (see Table 3 and Figure 5(a)). The inclusion of up to six different sets of SH3/peptide binding data in the training dataset did not decrease the correlation between predicted and experimental results obtained with Abl SH3.

Similar tests were performed with the SH3 domains of the Abl family (Figure 5(b)).

In summary: (1) feeding the SH3-specific matrix experimental data obtained with different SH3 domains does not decrease the prediction performance on single SH3 test cases; (2) the recognition specificity of a given SH3 can be successfully predicted if the SH3-specific matrix was derived from experimental data that include ligands of an SH3 domain that is related (belongs to the same family) to the query SH3 (Figure 5(a) and (b)).

Thus the approximation of considering the interaction between an SH3 domain and its ligand peptide as the sum of independent interactions between their contacting residues provides consistent results.

Database search results

SH3-SPOT was used to search the SWISSPROT (release 37) database of protein sequences with the Abl, Src, Hck, Fyn and Grb SH3 domains. The SWISSPROT release that we have used contains approximately 80,000 proteins (about 5.6×10^9 decapeptides) that were scanned first as class I and then as class II peptides.

SH3-SPOT evaluates the ''affinity'' of an SH3 for any decapeptide and does not provide information about the potential interaction with a protein containing the same peptide. In the context of the protein the peptide might be unavailable for interaction either because it is hidden inside the protein or because it is folded in a conformation different from the polyproline-type II helix. Furthermore, other contacts with distal parts of the protein might be important in establishing complex formation. Nevertheless, it is interesting to note that some of the known natural partners of the SH3 domains studied were ranked in the very first positions of the score (as shown in Table 4), namely: (1) among the ten best scoring decapeptide ligands for the Abl SH3 domain, one peptide from the 3BP-2 protein was selected (Cicchetti *et al.*, 1992); (2) the HIV Nef protein that is a characterized ligand of the Hck (=1ad5) SH3 domain is ranked as one of the best ligands for Hck by SH3-SPOT (Saksela *et al.*, 1995); (3) in the third position of the rank list obtained with the Src SH3 domain, one finds the voltage-gated potassium channel protein kv1.5. The SH3 domain of Src was observed to mediate the interaction with this protein (Holmes *et al.*, 1996). (4) The best predicted ligand for the Grb2 SH3 domain is the Sos protein, a known natural ligand (Vidal *et al.*, 1998). Among the top positions one also finds the proline-rich region of dynamin 1, another known ligand of Grb (Ando *et al.*, 1994; Vidal *et al.*, 1998); the Sos protein scores also in the first positions of the Src and fyn SH3 domains (Park *et al.*, 1998). (5) The cbl proto-oncogene is a natural partner of interaction of the Hck (Howlett *et al.*, 1999), Grb (Park *et al.*, 1998) and Fyn (Hunter *et al.*, 1999) SH3 domains. This protein is found among the highest ranking

Table 2. Preferences for specific SH3 residue/peptide residue interactions

Peptide position	Class I[a]	Class II
$\mathbf{P}_{-5}$	High values are reported for Y or P preferences related to residues found in positions 76, 77, 78 and 110 of the multiple alignment.	A significant preference for Y in this position is found for 110F
$\mathbf{P}_{-4}$	A few SH3s display a preference for a P and they have 76N, or 77Q or 94D. Significant is the correlation between 98N and a P	Generally an x; with 97G there is a significant preference for P in this position; with 98Q a D is preferred
$\mathbf{P}_{-3}$ (R/K)	75R, 76T, 79D are correlated with a preference for an R	72Y, 73E, 73 K, 75Q, 76N, 78D, 94D and 111F are generally related to a preference for an R; 75N and 76D are (although less significantly) related to a preference for a K
$\mathbf{P}_{-2}$	Most SH3 domains do not select any specific residue at this ligand position; 98N correlates with a preference for R	x of the consensus; in two different SH3s a preference for P correlates with 98N
$\mathbf{P}_{-1}$	72F, 98N, 111F or 111V, 113T, 114A indicate a preference for a P; 111I significantly prefers an L	P of the Pxx**P** motif
$\mathbf{P}_{0}$	It corresponds to the first P in the **P**xxP conserved core motif. Practically only P is tolerated	Uusually an x; single SH3 domains select L (74 K, 75E or 114T), V (75I) or R (115F) residues in this position
$\mathbf{P}_{+1}$	This is usually an x of the consensus; however with 114A the P has a high score	In a single SH3, a T in the 114th position of the alignment is always contacting a P
$\mathbf{P}_{+2}$	-	P of the **P**xxP motif
$\mathbf{P}_{+3}$	P of the Pxx**P** motif	Usually an x; with 73N the P is significantly preferred; only associated with 70W and 115L a preference for R is detected

[a] Position-specific contact frequency between residues is considered when higher than 0.7 and when reported for more than one SH3 domain associated with a list of binding peptides. x represents any of the 20 amino acid residues.

positions of the lists of predicted peptide ligands for all of these domains.

Spectrin and Drk SH3 domains are structurally characterized, but little experimental information is known about their ligands. Spectrin (Spc) is a membrane skeletal protein that plays an important role in the establishment and maintenance of the cell shape and polarity. The Drk protein is an adaptor protein which displays strong homology to mammalian Grb2 and *Caenorhabditis elegans* SEM-5 and is composed of one SH2 and two SH3 domains. Experimental evidence of *in vitro* interaction is described between Spc SH3 domain and e3B1 kinase (Ziemnicka-Kotula *et al.*, 1998) and between Drk SH3 domains and the Sos tail (Raabe *et al.*, 1995). SH3-SPOT prediction (see Table 4) confirms the interaction between Drk SH3 domain and the Sos protein, while the e3B1 kinase score is low in the Spc SH3 domain specificity score.

We would like to emphasize that the only natural target included in the SH3-specific matrices used for this test is the Nef RPQVPLRPMT sequence, associated to the 1efn SH3 sequence (fyn mutant R96I), as shown in Table 1.

Table 3. Correlation between experimental and SPOT-predicted results when using different matrices.

Query SH3	−5	−4	−3	−2	−1	All
1fmk	0.5996	0.7454	0.8762	0.9613	0.9640	0.9698
1ad5	0.5785	0.7250	0.8386	0.8574	0.8384	0.8828
1fyn	0.6588	0.8068	0.9122	0.9228	0.9098	0.9229
lyn	0.5387	0.6317	0.7857	0.8429	0.8570	0.9245
yes	0.5449	0.5843	0.7714	0.9253	0.9279	0.9469
1abo	0.9705	0.9742	0.9742	0.9764	0.9761	0.9761
1abo	-	0.5548	0.5739	0.9603	0.9603	0.9761
Myo3	-	0.5324	0.6623	0.9432	0.9449	0.9508
Myo5	-	0.5283	0.9169	0.9269	0.9310	0.9337
Spc	-	0.5811	0.6133	0.7680	0.7696	0.8219
1fmk	-	0.9694	0.9768	0.9688	0.9698	0.9698

Table 4. SWISSPROT results

SH3 domain and preferred ligand[a]	Decapeptide selected in SWISSPROT	SH3-SPOT score	SWISSPROT identification code[b]	Short description of the SWISSPROT selected protein
abo	PPAYSPPPPP	.9649	EXTN_TOBAC	cell wall hydroxyproline-rich glycoprotein(common tobacco)
	PPTYLPPPPP	.9603	EXTN_TOBAC	"
PPPYPPPPPP 24.48	PPAYPPPPVP	.9575	**3BP2_MOUSE,3BP2_HUMAN**	**SH3 domain-binding protein 3BP-2 (human and mouse)**
AFSR L ALR	PPVYSPPPPP	.9550	EXTN_TOBAC	cell wall hydroxyproline-rich glycoprotein(common tobacco)
GRLW V	PPPRPPPPPP	.9548	ATI2_HSV1F	alpha trans-inducing factor 77 kd (herpes simplex virus)
RYFK W	PPPRPPPPLP	.9514	VNUA_PRVKA	probable nuclear antigen (herpesviridae)
TF I	PPSRPPPPPP	.9501	KYK1_DICDI	non-receptor tyrosine kinase spore lysis a (slime mold)
A	PPPRLPPPPP	.9359	JUND_CHICK	transcription factor jun-d (chicken)
	PPGYGPPPGP	.9198	Y0A8_MYCTU	hypothetical 33.1 kd protein (Mycobacterium tuberculosis)
	PPPQPPPPPP	.9158	KNA3_ARATH	knotted-like homeobox protein 3 (arabidopsis)
	PPPQPPPPPP	.9158	HD_RAT	huntingtin (rat)
	PPPSPPPPPP	.9149	SSGP_VOLCA	sulfated surface glycoprotein 185 (green algae)
	PPPSPPPPPP	.9149	EBN2_EBV	ebna-2 nuclear protein (ebv)
	PPPSPPPPPP	.9149	ACRO_HUMAN	acrosin precurson (human)
	PPPHPPPPPP	.9147	Y029_HUMAN	hypothetical protein KIAA0029 (human)
ad5 cl.I	RSSRPLPPLP	.9480	G33_RAT	gene 33 polypeptide (rat)
	SSARPLPPPP	.9399	VPG_PLRVW,VPG_PLRVR,VPG_PLRV1,VPG_PLRV	putative genome-linked protein precursor (vpg)
RRFRPLPPPP 20.01	CRPRPLPPLP	.9342	FGR_FSVGR	tyr-protein kin. transforming protein fgr (feline sarc virus)
PPAK P ALR	TTFRPLPPPP	.9241	EXON_HSV11	alkaline exonuclease (hsv)
SSEV V	PGVRPLPPLP	.9220	*CIK5_HUMAN*	potassium channel protein kv1.5 (human)
LYIH W	SRARNLPLPP	.9126	ROK_HUMAN	heterogeneous nuclear ribonucleoprotein k (human)
IT I	PTLRDLPPPP	.9126	**CBL_MOUSE,CBL_HUMAN**	**proto-oncogene c-cbl (mouse,human)**
AV	NRKRPLPPTP	.9071	HT16_HYDAT	tyrosine-protein kinase htk16 (hydra vulgaris)
A	SGVRPLPPLP	.9056	*CIK5_HUMAN*	potassium channel protein kv1.5 (human)
ad5 cl.II	RPPVPPRPPR	.9809	HSS2_HUMAN	heparin sulfate n-deacetylase/n-sulfotransferase (mouse)
	RPPVPPRPQR	.9797	HSS2_MOUSE	heparin sulfate n-deacetylase/n-sulfotransferase (human)
RPPVPPRPST 28.02	PPPVPPRRRP	.9578	*SOS1_MOUSE,HUMAN*	son of sevenless 1 (human, mouse)
P QLALKRM	PPPVPLRPPE	.9555	*SOS2_HUMAN,MOUSE*	son of sevenless 2 (human, mouse)
K AR K	RPPVPPRGPL	.9529	SP49_HUMAN	spliceosome associated protein 49 (human)
A I A	RPQVPLRPMT	.9506	**NEF_HV1A2 (14 variants)**	**hiv-1 nef protein (human virus)**
P	RPPVPHRQPI	.9324	ZEP1_MOUSE	alpha a-crystallin-binding protein i (mouse)
	APPVPSRPGA	.9237	*DYN1_HUMAN,MOUSE,RAT*	dynamin-1 (human,mouse,rat)
	GPPVPPRQST	.9215	*SOS1_MOUSE,HUMAN*	son of sevenless 1 (human, mouse)
	APPVPFRPGP	.9201	*DYN3_RAT*	dynamin 3 (rat)
	KPQVPLRPMT	.9165	NEF_HV1BN (3 variants)	hiv-1 nef protein
	RPLVPDRPSI	.9154	V120_VZVD	capsid assembly protein 21 (varicella-zoster virus)
	PPQVPSRPNR	.9153	*DYN1_HUMAN,MOUSE,RAT*	dynamin-1 (human,mouse,rat)
	RPQVPKRPF	.9153	*DYN1_CAEEL*	dynamin-1 (caenorhabditis elegans)
	RPKVPLRAMT	.9141	NEF_SIVMK,SIVML	hiv-1 nef protein
fmk cl.I	PPPRPLPPRP	.9628	ACRO_HUMAN	acrosin precurson (human)
	RSSRPLPPLP	.9478	G33_RAT	gene 33 polypeptide (rat)
PPPRPLPPPP 20.84	PGVRPLPPLP	.9420	**CIK5_HUMAN**	**potassium channel protein kv1.5 (human)**
LSSK P ALR	APPRLLPPLP	.9407	GLI1_HUMAN	zinc finger protein gli1 (human)
SRG V	PTLRDLPPPP	.9390	CBL_MOUSE,HUMAN	proto-oncogene c-cbl (mouse,human)
RYR W	SSARPLPPPP	.9376	VPG_PLRVW,PLRVR,PLRV1,PLRV	putative genome-linked protein precursor (Potato leafroll virus)
ILL I	PLLRFLPPLP	.9333	GAG_FRSFV	gag polyprotein (friend spleen focus-forming virus)
AT	SPRRPLPPAP	.9323	CHS2_EMENI	chitin synthase 2 (aspergillus nidulans)
FA	PPRRSLPSPP	.9262	NO20_MEDTR	early nodulin 20 precursor (N-20)(Medicago truncatula)
TG	CRPRPLPPLP	.9247	FGR_FSVGR	tyr-protein kin. transforming protein fgr (feline sarc virus)
F	PPSRPLPADP	.9234	FGD1_MOUSE,HUMAN	putative RHO/RAC guanine nucleotide exchange factor (mouse,human)
V	PGPRPLPWPP	.9226	BAT2_HUMAN	large proline-rich protein bat2 (human)
	TTFRPLPPPP	.9215	EXON_HSV11	alkaline exonuclease (herpes simplex virus)
	LKLRSLPPLP	.9204	IPA4_SHIFL	65.4 kd antigen (shigella flexneri)
	LIRRVLPPPP	.9196		protein ul88 (human cytomegalovirus)

Table 4 (*continued*)

			UL88_HCMVA	
fmk cl.II	RPPVPPRPPR	.9771	HSS2_HUMAN	heparin sulfate n-deacetylase/n-sulfotransferase (human)
	RPPVPPRPQR	.9700	HSS2_MOUSE	heparin sulfate n-deacetylase/n-sulfotransferase (mouse)
RPPVPPRPGT 29.35	PPPVPLRPPE	.9522	**SOS2_HUMAN,MOUSE**	**son of sevenless 2 (human, mouse)**
P QLALKRS	RPQVPLRPMT	.9484	*NEF_HV1A2 (14 variants)*	hiv-1 nef protein (human virus)
K AP SPGM	APPVPSRPGA	.9273	*DYN1_HUMAN,MOUSE,RAT*	dynamin-1 (human,mouse,rat)
A I N	RPPVPPRGPL	.9226	SP49_HUMAN	spliceosome associated protein 49 (human)
G P	APPVPFRPGP	.9213	*DYN3_RAT*	dynamin 3 (rat)
Q	KPQVPLRPMT	.9158	NEF_HV1BN (3 variants)	hiv-1 nef protein
R	PPPVPPRRRP	.9131	SOS1_MOUSE,HUMAN	son of sevenless 1 (mouse,human)
	PPPLPPRPPH	.9124	UL20_HCMVA	hypothetical protein ul20 precursor (human cytomegalovirus)
	PPPLPPRPRQ	.9112	NG3_DROME	new-glue protein 3 precursor (drosophila melanogaster)
	RPLVPDRPSI	.9081	V120_VZVD	capsid assembly protein 21 (varicella-zoster virus)
	PPQVPSRPNR	.9070	*DYN1_HUMAN,MOUSE,RAT*	dynamin-1 (rat,mouse,human)
	PPPLPPRPAA	.9054	LMBV_CHICK	laminin beta-1 chain variant (chicken)
	RPQVPKRPF	.9011	DYN1_CAEEL	dynamin-1 (caenorhabditis elegans)
fyn cl.I	RSSRPLPPLP	.9708	G33_RAT	gene 33 polypeptide (rat)
	PPPRPLPPRP	.9616	ACRO_HUMAN	acrosin precurson (human)
PPPRPLPPPP 20.30	APPRLLPPLP	.9475	GLI1_HUMAN	zinc finger protein gli1 (human)
RYSK P ALR	SSARPLPPPP	.9430	VPG_PLRVW,PLRVR,PLRV1,PLRV	putative genome-linked protein precursor (vpg)
SSGY V	PGVRPLPPLP	.9429	*CIK5_HUMAN*	potassium channel protein kv1.5 (human)
LLR W	PTLRDLPPPP	.9364	**CBL_MOUSE,HUMAN**	**proto-oncogene c-cbl (mouse,human)**
ARL I	PLLRFLPPLP	.9359	GAG_FRSFV	gag polyprotein (friend spleen focus-forming virus)
IF	SPRRPLPPAP	.9339	CHS2_EMENI	chitin synthase 2 (aspergillus nidulans)
FT	CRPRPLPPLP	.9263	FGR_FSVGR	tyr-protein kin. transforming protein fgr (feline sarc virus)
HG	PPRRSLPSPP	.9240	NO20_MEDTR	early nodulin 20 precursor (N-20)(Medicago truncatula)
TA	PGPRPLPWPP	.9230	BAT2_HUMAN	large proline-rich protein BAT2 (human)
H	SGVRPLPPLP	.9227	*CIK5_HUMAN*	potassium channel protein kv1.5 (human)
	TTFRPLPPPP	.9226	EXON_HSV11	alkaline exonuclease (herpes simplex virus)
	PPSRPLPADP	.9211	FGD1_MOUSE,HUMAN	putative RHO/RAC guanine nucleotide exchange factor (mouse,human)
	RSRRQLPPVP	.9179	CCT1_RAT	nickel-sensitive t-type calcium channel alpha-1 subunit (rat)
fyn cl.II	RPPVPPRPPR	.9750	HSS2_HUMAN	heparin sulfate n-deacetylase/n-sulfotransferase (mouse)
	RPPVPPRPQR	.9686	HSS2_MOUSE	heparin sulfate n-deacetylase/n-sulfotransferase (human)
RPPVPPRPMT 29.57	RPQVPLRPMT	.9528	*NEF_HV1A2 (14 variants)*	hiv-1 nef protein (human virus)
P QLALKRS	PPPVPLRPPE	.9503	**SOS2_HUMAN,MOUSE**	**son of sevenless 2 (human, mouse)**
K AP SPGG	RPQVPTRPMT	.9369	NEF_SIVCZ	hiv-1 nef protein
A R P	RPPVPPRGPL	.9209	SP49_HUMAN	spliceosome associated protein 49 (human)
I	KPQVPLRPMT	.9205	NEF_HV1BN (3 variants)	hiv-1 nef protein
G	APPVPSRPGA	.9151	*DYN1_HUMAN,MOUSE,RAT*	dynamin-1 (rat,mouse,human)
	PPPVPPRRRP	.9096	*SOS1_MOUSE,HUMAN*	son of sevenless 1 (human, mouse)
	APPVPFRPGP	.9091	*DYN3_RAT*	dynamin 3 (rat)
	PPPLPPRPPH	.9078	UL20_HCMVA	hypothetical protein ul20 precursor (human cytomegalovirus)
	RPLVPDRPSI	.9070	V120_VZVD	capsid assembly protein 21 (varicella-zoster virus)
	PPPLPPRPRQ	.9046	NG3_DROME	new-glue protein 3 precursor (drosophila melanogaster)
	PPQVPSRPNR	.9040	*DYN1_HUMAN,MOUSE,RAT*	dynamin-1 (rat,mouse,human)
	PPPLPPRPAA	.9027	LMBV_CHICK	laminin beta-1 chain variant (chicken)
grb cl.I	PPPRPPPPLP	.9097	VNUA_PRVKA	probable nuclear antigen (pseudorabiae virus)
	APPRLLPPLP	.9014	GLI1_HUMAN	zinc finger protein gli1 (human)
RFPRRLPPLP 19.65	RLPRALPRLP	.8948	CYGD_HUMAN	retinal guanylyl cyclase 1 precursor (human)
YPDK P APR	SPPSRLPPLP	.8875	UL29_HCMVA	hypothetical protein ul29 (cytomegalovirus)
PYYY V	PPPRPLPPRP	.8853	ACRO_HUMAN	acrosin precurson (human)
KWRD I	PPPRPPPPPP	.8841	ATI2_HSV1F	alpha trans-inducing factor 77 kd (herpes simplex virus)
SLSS W	KFPRSLPFLP	.8767	CPT7_MOUSE	cytochrome p450 xviia1 (mouse)
RGA T	RSSRPLPPLP	.8765	G33_RAT	gene 33 polypeptide (rat)
K F	PPPRLPPPPP	.8760	JUND_CHICK	transcription factor jun-d (chicken)
S Q	RLPRPLPAVP	.8734	ANAG_HUMAN	alpha-n-acetylglucosaminidase precursor (human)
	CRPRPLPPLP	.8704	FGR_FSVGR	tyr-protein kin. Transforming protein fgr (feline sarc virus)
	RPPPPLPPPP	.8673	VNUA_PRVKA	probable nuclear antigen (pseudorabiae virus)
	PPPPPLPPLP	.8648	FASL_HUMAN	fas antigen ligand (apoptosis antigen ligand)(APTL)(human)
	RPPRPLPVAP	.8648	P85A_RAT,MOUSE,HUMAN,BOVIN	phosphatidylinositol 3-kinase regulatory alpha subunit
	KYPRSLPSLP	.8570	CPT7_SHEEP,BOVIN	cytochrome p450 xviia1 (bovine, sheep, pig)

Table 4 (*continued*)

grb cl.II	PPPVPPRRRP	.9714	**SOS1_MOUSE,HUMAN**	**son of sevenless 1 (human, mouse)**
	RPPVPPRPPR	.9507	HSS2_HUMAN	heparin sulfate n-deacetylase/n-sulfotransferase (human)
PPPVPPRPS 28.86	PPPVPLRPPE	.9496	**SOS2_HUMAN,MOUSE**	**son of sevenless 2 (human, mouse)**
R QLTSKQP	RPPVPPRPQR	.9386	HSS2_MOUSE	heparin sulfate n-deacetylase/n-sulfotransferase (mouse)
K AR TRR	GPPVPPRQST	.9320	**SOS1_MOUSE,HUMAN**	**son of sevenless 2 (human, mouse)**
A I SG	PPPLPPRPPH	.9255	UL20_HCMVA	hypothetical protein ul20 (cytomegalovirus)
G P EM	RPPVPPRGPL	.9255	SP49_HUMAN	spliceosome associated protein 49
N	PPPLPPRPRQ	.9237	NG3_DROME	new-glue protein 3 (fruit fly)
	APPVPPRQNS	.9226	**SOS2_HUMAN,MOUSE**	**son of sevenless 2 (human, mouse)**
	PPCVPPRDSL	.9185	CD19_HUMAN	differentiation antigen cd19 (human B cells)
	PPPLPPRPAA	.9178	LMBV_CHICK	laminin beta-1 chain variant (chicken)
	PPKVPPREPL	.9162	G33_RAT	gene 33 polypeptide (rat)
	PPAVPPRAPI	.9138	MAPA_HUMAN,RAT	microtubule-associated protein 1a (human, rat)
	PPQVPSRPNR	.9062	**DYN1_HUMAN,MOUSE,RAT**	**dynamin-1 (rat,mouse,human)**
	APPVPSRPGA	.9045	**DYN1_HUMAN,MOUSE,RAT**	"
	PPQVPSRPTR	.9016	**DYN3_RAT**	**dynamin-3 (rat)**
	PPPLPPREAP	.9006	ENV_GALV	env polyprotein precursor (gibbon ape leukemia virus)
	LPPVPPRLDL	.8994	**CBL_HUMAN,MOUSE**	**proto-oncogene c-cbl (human, mouse)**
	PPPLPPRRKE	.8991	**SOS_DROME**	**son of sevenless protein (fruit fly)**
	PPLVPPRGEV	.8977	GPK1_DROME	G protein-coupled receptor kinase 1 (fruit fly)
	PPRVPSRPRR	.8972	UL43_HSV11	membrane protein ul43 (herpessimplex virus)
	APPVPFRPGP	.8961	**DYN3_RAT**	**dynamin-3 (rat)**
	PPPRPPRPPR	.8954	GLE_CHLRE	autolysin precursor (chlamydomonas reinhardtii)
	RPPVPHRQPI	.8953	ZEP1_MOUSE	alpha a-crystallin-binding protein I (mouse)
	PPFVPPRYLR	.8940	PEN3_ADE02,ADE05	penton base protein (human adenovirus type 2, type 5)
drk n-term cl.I	APPRLLPPLP	.9409	GLI1_HUMAN	zinc finger protein gli1 (human)
	KFPRSLPFLP	.9402	CPT7_MOUSE	cytochrome p450 xviia1 (mouse)
RFPR LPPLP 16.45	CRPRPLPPLP	.9287	FGR_FSVGR	tyr-protein kin. transforming protein fgr (feline sarc virus)
PPDK P APR	PPPRPPPPLP	.9249	VNUA_PRVKA	probable nuclear antigen (herpesviridae)
RYY V	GFPYGLPPLP	.9243	Y054_HUMAN	hypothetical protein KIAA0054 (human)
W D	RLPRALPRLP	.9178	CYGD_HUMAN	retinal guanylyl cyclase 1 precursor (human)
Y S	GRPRSLPALP	.9064	CPT7_CHICK	cytochrome p450 xviia1 (chicken)
L A	PPPRPLPPRP	.8981	ACRO_HUMAN	acrosin precurson (human)
K F	PPPRPPPPPP	.8938	ATI2_HSV1F	alpha trans-inducing factor 77 kd (herpes simplex virus)
S Q	PPPRLPPPPP	.8938	JUND_CHICK	transcription factor jun-d (chicken)
	KYPRSLPSLP	.8936	CPT7_SHEEP,BOVIN	cytochrome p450 xviia1 (bovine, sheep)
	KYPRSLPVLP	.8929	CPT7_PIG	cytochrome p450 xviia1 (pig)
	KLPRSLPSLP	.8922	CPT7_RAT	cytochrome p450 xviia1 (rat)
	PRPRGPPPPP	.8909	S145_HUMAN	spliceosome associated protein 145
	PLPRALPPEP	.8887	VENV_MCV2	major envelope protein (P43K)
drk n-term cl.II	PPPVPLRPPE	.9520	**SOS2_HUMAN,MOUSE**	**son of sevenless 2 (human, mouse)**
	PPPVPPRRRP	.9472	**SOS1_MOUSE,HUMAN**	**son of sevenless 1 (human, mouse)**
PPPVPPRPSW 25.45	RPPVPPRPPR	.9415	HSS2_HUMAN	heparin sulfate n-deacetylase/n-sulfotransferase (human)
RFQL KRG	RPPVPPRPQR	.9370	HSS2_MOUSE	heparin sulfate n-deacetylase/n-sulfotransferase (mouse)
K AR T M	APPVPSRPGA	.9154	*DYN1_HUMAN,MOUSE,RAT*	dynamin-1 (rat,mouse,human)
A I A	GPPVPPRQST	.9139	**SOS1_MOUSE,HUMAN**	**son of sevenless 1 (human, mouse)**
P	APPVPFRPGP	.9099	*DYN3_RAT*	dynamin-3 (rat)
	PPQVPSRPTR	.9096	*DYN3_RAT*	dynamin-3 (rat)
	RPPVPPRGPL	.9089	SP49_HUMAN	spliceosome associated protein 49
	PPQVPSRPNR	.9084	*DYN1_HUMAN,MOUSE,RAT*	dynamin-1 (rat,mouse,human)
	PPPLPPRPAA	.9083	LMBV_CHICK	laminin beta-1 chain variant (chicken)
	PPPLPPRPPH	.9061	UL20_HCMVA	hypothetical protein ul20 (cytomegalovirus)
	PPPLPPRPRQ	.9037	NG3_DROME	new-glue protein 3 (fruit fly)
	PPCVPPRDSL	.9011	CD19_HUMAN	differentiation antigen cd19 (human B cells)
	VPPVPVRPRG	.8980	RPC1_TRYBB	DNA-directed RNA polymerase III largest subunit

Table 4 (*continued*)

drk c-term cl.I **PFPYRPPPLP** 20.41 RRSK L PR AP R R K P V W W	LFPYQPPPPP	.9393	CEBA_HUMAN,MOUSE,RAT	CCAAT/enhancer binding protein alpha (human,mouse,rat)
	PPPRPPPPLP	.9216	VNUA_PRVKA	Probable nuclear antigen (herpesviridae)
	AFPPPPPPPP	.9206	CC12_SCHPO	Cell division control protein 12 (S.pombe)
	PPPRPPPPPP	.9192	ATI2_HSV1F	alpha trans-inducing factor 77 kd (herpes simplex virus)
	SFPPPPPPPP	.9173	POL2_TRSVR	RNA2 polyprotein (tomato ringspot virus)
	PRPRGPPPPP	.9145	S145_HUMAN	spliceosome associated protein 145
	PPPRLPPPPP	.9115	JUND_CHICK	transcription factor jun-d (chicken)
	AFPSPPPPPP	.9101	ZYX_CHICK	zyxin (chicken)
	LFPPAPPPPP	.9093	FRE4_HUMAN	forkhead-related transcription factor 4
	GFPPLPPPPP	.9058	GTPA_BOVIN	GTPASE-activating protein
	SPPPRPPPPP	.9025	ATI2_HSV1F	alpha trans-inducing factor 77 kd (herpes simplex virus)
	SPPPRPPPPP	.9025	VINC_CAEEL	vinculin (C.elegans)
	PFPSPPPPAP	.9010	MKK2_MOUSE	MAP kinase-activated protein kinase 2
	PPPPPPPPLP	.8969	CLC1_RAT	Chloride channel protein, skeletal muscle
	PPPPPPPPLP	.8969	FASL_HUMAN	fas antigen ligand (apoptosis antigen ligand)(APTL)(human)
drk c-term cl.II **PPPVPPRRX** 28.47 R QL KP K AI K A R Q	PPPVPPRRRP	.9956	**SOS1_MOUSE,HUMAN**	**son of sevenless 1 (human, mouse)**
	PPPLPPRPAA	.9694	LMBV_CHICK	laminin beta-1 chain variant (chicken)
	RPPVPPRPPR	.9679	HSS2_HUMAN	heparin sulfate n-deacetylase/n-sulfotransferase (human)
	PPPLPPRRKE	.9666	**SOS_DROME**	**son of sevenless (fruit fly)**
	RPPVPPRPQR	.9665	HSS2_MOUSE	heparin sulfate n-deacetylase/n-sulfotransferase (mouse)
	PPPLPPRPPH	.9665	UL20_HCMVA	hypothetical protein ul20 (cytomegalovirus)
	PPPLPPRKKF	.9657	**SOS2_HUMAN,MOUSE**	**son of sevenless 2 (human, mouse)**
	PPPLPPRPRQ	.9650	NG3_DROME	new-glue protein 3 (fruit fly)
	PPPLPPREAP	.9581	ENV_GALV	env polyprotein precursor (gibbon ape leukemia virus)
	PPPVPLRPPE	.9533	**SOS2_HUMAN,MOUSE**	**son of sevenless 2 (human, mouse)**
	RPPVPPRGPL	.9531	SP49_HUMAN	spliceosome associated protein 49
	PPPLPPRATP	.9525	PRGR_HUMAN	progesterone receptor
	PPPLPPRDYP	.9492	UL25_HCMVA	hypothetical protein UL25 (human cytomegalovirus)
	APPVPPRQNS	.9451	**SOS2_HUMAN,MOUSE**	**son of sevenless 2 (human, mouse)**
	GPPVPPRQST	.9386	**SOS1_MOUSE,HUMAN**	**son of sevenless 1 (mouse,human)**
spc cl.I **PPSRPPPPLP** 19.82 RSLP L APR ADFH V SYPK I FFGY W GR F G	PPSRPPPPPP	.9963	KYK1_DICDI	non-receptor tyrosine kinase spore lysis a (slime mold)
	PPSPPPPPPP	.9836	ACRO_HUMAN	acrosin precurson (human)
	PPSPPPPPPP	.9836	EBN2_EBV	ebna-2 nuclear protein (ebv)
	PPSPPPPPPP	.9836	FORM_CHICK	formin (limb deformity protein)
	PPSPPPPPPP	.9836	SSGP_VOLCA	sulfated surface glycoprotein 185 (volvox carteri)
	PPPRPPPPLP	.9741	VNUA_PRVKA	probable nuclear antigen (pseudorabiae virus)
	PPPRPPPPPP	.9704	ATI2_HSV1F	alpha trans-inducing factor 77 kd (herpes simplex virus)
	PPLPPPPPPP	.9619	EBN2_EBV	ebna-2 nuclear protein (ebv)
	PPLPPPPPPP	.9619	FASL_HUMAN	fas antigen ligand (apoptosis antigen ligand)(APTL)(human)
	PPLPPPPPPP	.9619	VNUA_PRVKA	probable nuclear antigen (pseudorabiae virus)
	PPPPPPPPLP	.9614	CLC1_RAT	Chloride channel protein, skeletal muscle
	PPPPPPPPLP	.9614	FASL_HUMAN	fas antigen ligand (apoptosis antigen ligand)(APTL)(human)
	PPPPPPPPLP	.9614	FOR4_MOUSE	formin 4 (limb deformity protein)
	PPPPPPPPLP	.9614	FORM_MOUSE	formin (limb deformity protein)
	PPPPPPPPLP	.9614	NDPP_MOUSE	NPC derived proline rich protein 1 (NDPP-1)
spc cl.II **RPPVPLRPP** 18.25 K QL PK S P AI A R G	RPPVPPRPPR	.9960	HSS2_HUMAN	heparin sulfate n-deacetylase/n-sulfotransferase (human)
	RPPVPPRPQR	.9833	HSS2_MOUSE	heparin sulfate n-deacetylase/n-sulfotransferase (mouse)
	PPPVPLRPPE	.9744	*SOS2_HUMAN,MOUSE*	son of sevenless 2 (human, mouse)
	RPPVPPRGPL	.9722	SP49_HUMAN	spliceosome associated protein 49
	RPPVPHRQPI	.9515	ZEP1_MOUSE	alpha a-crystallin-binding protein I (mouse)
	PPPVPPRRRP	.9449	*SOS1_MOUSE,HUMAN*	son of sevenless 2 (mouse, human)
	RPQVPLRPMT	.9419	*NEF_HV1A2 (14 variants)*	hiv-1 nef protein
	PPPLPPRPPH	.9250	UL20_HCMVA	hypothetical protein ul20 (cytomegalovirus)
	KPPVPGRSDS	.9248	NIA2_ARATH	nitrate reductase 2 (arabidopsis thaliana)
	RPLVPDRPSI	.9247	V120_VZVD	capsid assembly protein 21 (varicella-zoster virus)
	RPQVPKRPF	.9221	DYN1_CAEEL	dynamin-1 (caenorhabditis elegans)
	PPPLPPRPRQ	.9208	NG3_DROME	new-glue protein 3 precursor (drosophila melanogaster)
	RPEVPLRRGQ	.9184	SERA_MOUSE	D-3-phosphoglycerate dehydrogenase
	RPEVPLRRGQ	.9184	SERA_RAT	D-3-phosphoglycerate dehydrogenase
	KPQVPLRPMT	.9168	NEF_HV1BN (3 variants)	hiv-1 nef protein

[a] Name of the SH3 domain used for the panning experiment; the sequence of the preferred ligand is reported in bold and is calculated as described in Methods. Other residues with decreasing scores are reported under the preferred residue. The score corresponding to the preferred ligand is reported and is used to normalize the score values reported in the SH3-SPOT scores.

[b] The known natural partners of interaction are reported in bold; proteins which are known to interact with SH3 other than the query SH3 sequence are underlined.

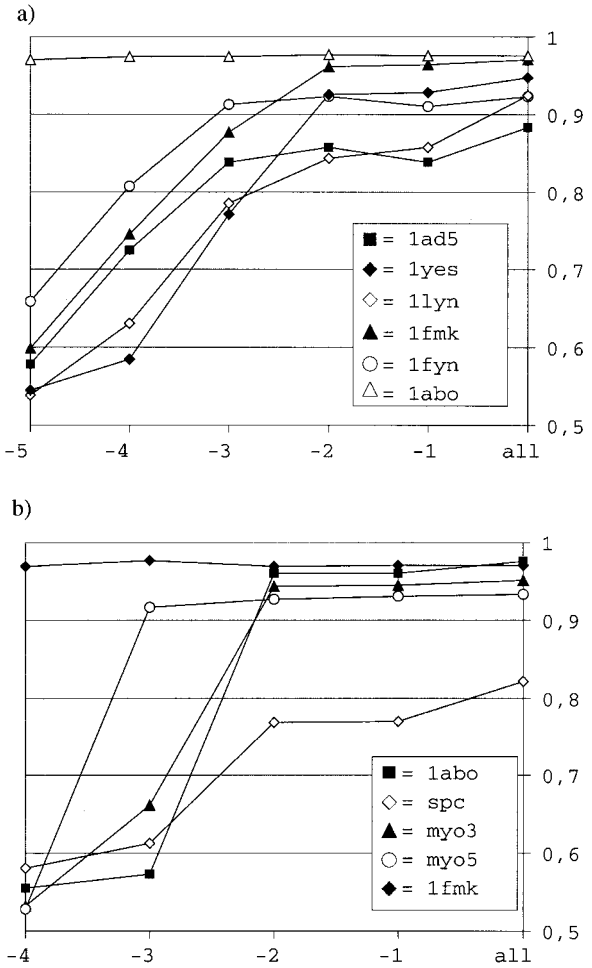

Figure 5. Correlation plots obtained with different SH3-specific matrices (see Methods). Different SH3-specific matrices were used in prediction tests on the SH3 sequences of members of the Src family (1fmk, 1ad5, 1fyn, 1lyn and 1yes) and of the Abl family (1abo, myo3, myo5 and spectrin). The different SH3 query domains were launched in SPOT prediction tests with SH3-specific matrices built with an increasing number of SH3 interaction data. In each test, the interaction data were added to the matrices according to increasing sequence similarity to the query SH3 domain, as evaluated with a FASTA comparison (see Methods). (a) Correlation tests performed with SH3 domains of the Src family; (b) correlation tests performed with the Abl SH3 domain family.

Discussion

The SH3-SPOT procedure is a new method to predict specificity of peptide recognition mediated by SH3 domains. Information about the SH3/peptide contact positions are derived from six SH3 peptide complexes of known structure. The contact framework defined on the complexes of known structure is filled with contact data (virtual contacts) derived from SH3/peptide binding data obtained with the phage display technique. Information about position-specific SH3-residue/peptide-residue contacts is stored in an SH3-specific matrix, where the frequency of occurrence of any pair of residues in the SH3 and in the ligand peptide is tabulated. The SH3-specific matrix is then used to evaluate the probability of interaction between SH3 domains and peptides (see Figure 2).

We analyzed the class I and class II SH3-specific matrices derived from our database of SH3/peptide contact data (Table 1) and report differences between the two binding modes (Figures 3(b) and 1). Some of these differences may be due to the relatively small size of the database of structures of SH3/peptide complexes. Nevertheless, analysis of the frequency of occurrence of the residue-residue contacts in the different SH3 and bound-peptide positions highlights the presence of SH3 positions which are consistently related to preferences for specific residues in the peptide ligand (Figure 1 and Table 2).

The procedure is based on the assumption that the interaction between two proteins can be described, in a first approximation, as the sum of independent interactions between their contacting residues.

Test experiments were performed to support this rationale: a number of different SH3-specific matrices, containing interaction data extracted from an increasing number of SH3/peptide complexes, were used to predict the peptide recognition specificity of SH3 domains that were not included in the training set. No negative interference was detected between data coming from SH3 domains of different families; a positive effect on the correlation between experimental and predicted results was obtained with the addition of interaction data coming from SH3 domains of the same family (see Figure 5(a) and (b)).

In the test case of the Src and Abl SH3 families, we observed a good correlation between experimental and predicted values even when the specific SH3 considered had not been included in the training set. Despite the relatively large number of experimental data considered, our SH3 specific matrix is still largely incomplete. It is likely that the SH3 domains characterized by phage display represent a biased selection within the SH3 domain family. The SH3 family multiple sequence alignment contains 632 members (Bateman *et al.*, 1999). The SH3 domains of known specificity are not evenly interspersed among the branches of the corresponding similarity tree. We estimate that the matrix that we have assembled, in each single contact position does not contain information for approximately 25% of the known SH3 domains. The results that we have presented indicate that we can confidently predict recognition specificity of SH3 domains belonging to the two families (Src and Abl) represented in our experimental data set. However, distantly related SH3 are likely to be wrongly predicted by SPOT.

We are currently collecting more SH3/peptide data obtained from SH3 domains belonging to different families. We expect correlation between

experimental and predicted results to increase together with the increase of contact information.

In principle, the SH3-SPOT procedure scores decapeptide sequences and cannot evaluate the potential interaction between an SH3 domain and the entire protein. It is interesting to note that, nevertheless, some of the known natural partners of the SH3 domains studied were ranked in the very first positions of the SPOT rank list (as shown in Table 4). A few experimentally determined interactions were correctly predicted: (1) the Abl SH3 domain and the 3BP-2 protein; (2) the Hck SH3 domain with the proto-oncogene c-cbl and the HIV Nef protein; (3) the src SH3 domain with the proto-oncogene c-cbl, the potassium channel protein kv1.5 and the Son of sevenless 1 and 2 proteins; (4) the fyn SH3 domain with the proto-oncogene c-cbl and the Son of sevenless 1 and 2 proteins; (5) the grb SH3 domain and the Son of sevenless 1 and 2 proteins, dynamin 1 and 3 proteins, and c-cbl. It is also possible that some interactions that were predicted from SPOT have not yet been detected in experimental studies.

The evaluation method used by SPOT differs from a simple sequence search in a protein database performed with an experimentally determined consensus. In fact each SH3/peptide pair is evaluated with scores related to the observed frequencies of residue-residue interaction and does not depend on any residue substitution matrix. Furthermore SPOT is designed to predict the preferred ligand of SH3 domains whose specificity has not been determined experimentally.

Finally, the SPOT method, i.e. the construction of a structure-specific matrix and the use of domain/ligand position specific contacts to evaluate the probability of interaction, can easily be extended to other families of protein domains where at least one crystal structure of the protein-ligand complex is known and interaction data obtained with combinatorial methods are available.

Methods

Database of SH3 complex structures and phage display data

We used the Protein Data Bank (Bernstein *et al.*, 1977) coordinates of the following SH3/peptide or SH3/protein complexes: 1abo, 1cka, 1ckb, 1efn, 1fyn, 1sem (Table 1A). These complexes represent the six SH3 domain/peptide structures solved by X-ray diffraction. In these complexes the bound peptide is always in the standard polyproline II conformation. We did not consider the p53 tumor suppressor bound to the 53bp2 SH3 domain (PDB code: 1ycs), where the peptide bound into the SH3 pocket displays an unusual non-polyproline II conformation. Our database of SH3/peptide interaction includes 319 ligand peptides selected from 25 different SH3 domains (Table 1B).

Multiple alignment of SH3 sequences

We produced a multiple alignment (Figure 3(a)) of the SH3 sequences belonging to the SH3/peptide complexes of known structure. The PILEUP program of the Wisconsin Package software (Genetics Computer Group (GCG), 1999) was initially used to align the sequences of the SH3 domains whose structure, complexed with a ligand peptide, has been determined by X-ray crystallography. The alignment was then structurally refined in the SH3 binding regions using the profile3D software (De Rinaldis *et al.*, 1998). Thereby we generated a 3D multiple superposition of SH3 domains, based on the optimization of surface similarity. We then aligned in the same column of the multiple sequence alignment the residues situated in the same element of the 3D motif grid of the SH3 surface multiple alignment.

Definition of contacts and virtual contacts in the SH3/peptide complexes

Two residues were defined in contact if, in any of the six complexes of known crystallographic structure, the smallest distance between their atoms is less than the sum of the van der Waals radii +3 Å. The SH3 residues in contact with residues of the peptide are shown in bold in Figure 3(a): 27 positions of the multiple alignment of SH3 sequences shown in Figure 3(a) are in contact with at least one residue of a ligand according to this criterion.

In order to access and exploit the SH3/peptide interaction data obtained by phage display, we defined two SH3 and peptide residues as being in virtual contact if: (1) the SH3 sequence can be aligned to the multiple alignment shown in Figure 3(a); (2) the peptide sequence can be aligned to at least one of the peptides of the SH3/peptide complexes of known structure; and (3) the SH3 and the peptide were experimentally shown to interact.

The SH3 and peptide residues contact positions are identified by the SH3 and peptide multiple sequence alignments, respectively.

The definition of virtual contact is supported by the theoretical basis of homology modelling. An SH3 sequence which can be aligned to a multiple alignment of SH3 domains of known structure can be homology-modeled with high reliability, depending on its sequence similarity to the template structure (Sali, 1995; Srinivasan *et al.*, 1996). A polyproline sequence which has been shown to interact with an SH3 domain in a phage display experiment is very likely to assume a polyproline II structure. It is therefore reasonable to assume that the contact pattern identified on the crystal structures of a number of complexes is conserved in the homologous sequences and can be extended to the phage display interaction data.

Virtual contacts between SH3 domains and peptide positions P_{-5} and P_{-6} in the class II orientation have been assigned as shown in Figure 3(b). The choice was driven by the observation that the contacts common to class I and class II orientations are always defined, in the peptide positions ranging from P_{+3} to P_{-4}, by high numbers of atomic contacts in the complex crystal structures. Correlation tests performed with matrices built with this rationale can predict class II orientation specificity also for peptide positions P_{-5} and P_{-6} and perform better than the corresponding matrices defined only with contacts detected in the crystal structures (i.e. lacking class II contacts in the P_{-5} and P_{-6} peptide positions). Moreover,

this modification allowed us to afford the prediction of specificity for a few SH3 domains of the Src family where the two positions discussed are known to be important in modulating peptide recognition, as described in Results.

Construction of an SH3-specific matrix of frequencies of residue-residue contacts

In a complex of known structure, the contacts between residues in specific positions of the SH3 sequences are easily identified, as already explained. We may therefore consider a contact table (Figure 3(b)) where different columns identify the 27 SH3 contact positions and different rows identify the ten peptide positions. Some elements of this table are "closed" (in grey in Figure 3(b)) because they represent SH3 and peptide positions which are too far to be in contact. Other elements, representing observed SH3/peptide contact positions, are "open" (in white in Figure 3(b)).

Each open position of the 27×10 table corresponds to a 20×20 matrix describing all the possible residue/residue contacting pairs found in the SH3/peptide complexes (Figure 3(c)). In the different positions of this $27 \times 10 \times 20 \times 20$ interaction scheme, we calculated the frequencies of occurrence of the observed residue/residue pair in experiments in which SH3 ligands were obtained by searching combinatorial peptide libraries displayed on filamentous phage capsid. We called this scheme of observed interactions between SH3 domains and peptides "SH3-specific matrix".

Evaluation of probabilities of interaction between an SH3 sequence and a decapeptide, a protein sequence, a database of protein sequences

In order to evaluate the probability of interaction between an SH3 sequence and a ten residue peptide, we perform the following procedure: (1) the query SH3 sequence is manually aligned to the multiple alignment shown in Figure 3(a) and the 27 SH3 contacting positions are assigned; (2) the SH3-specific matrix is accessed and the sum of the frequencies corresponding to the residues of the chosen SH3 and peptide is scored.

To evaluate the probability of interaction between an SH3 domain and a protein sequence or a database of protein sequences, the procedure scans the protein sequence or sequences in ten residue long peptides at one residue steps. A simple computer program may also access the matrix in order to calculate, for each SH3 domain, the best ranking residues in the different positions of the peptide sequence (Table 4). The peptide composed of the best ranking residues is the best ranking peptide and it is used to normalize the scores of each single SH3 domains.

SH3 domains can bind to polyproline II peptides in two different orientations

SH3 domains may bind to peptides in polyproline II conformation in two different and opposite orientations (Lim *et al.*, 1994, Yu *et al.*, 1994), identified as class I and class II orientations. We analysed the contact schemes related to the different orientations and describe in the Results section the differences and similarities between the two SH3-specific contact schemes. The SH3-SPOT procedure scores each decapeptide in both orientations, using class I and class II SH3-specific matrices.

Correlation between expected and observed results

In order to test SH3-SPOT predictive power, we built different SH3-specific matrices, calculated either using all the experimental data about preferred SH3 ligands or a subset of these.

We selected five SH3 domains of the src family (1ad5, 1fmk, 1fyn, 1lyn and 1yes) and, for each single SH3 sequence, we ran a FASTA sequence comparison *versus* the sequences of the other four members of the family. Different SH3-specific matrices were built for each query SH3 domain, obtained by adding the interaction data related to the other members of the family according to an increasing sequence similarity to the query SH3 domain. For instance, when using the Src SH3 sequence (1fmk) as query SH3 domain, we built a -5 matrix obtained from all the SH3/peptide interaction data except those related to the 1lyn, 1ad5, 1fyn, 1yes and 1fmk SH3 domains. A -4 matrix was built adding to the -5 matrix the SH3 interaction data related to the SH3 domain displaying the lowest degree of sequence similarity to the the query sequence (1lyn). The -3, -2 and -1 matrices were built by adding the interaction data related to 1ad5, 1fyn and 1yes to the preceding matrices. The order was chosen according to the increasing sequence identity with the 1fmk SH3 domain.

The same protocol, *mutatis mutandis*, was used for the other members of the Src and Abl families. The different matrices were then used to rank the peptides present in our peptide database, composed of a grand total of about 319 peptides (Table 3 and Figure 5).

The correlation between the peptide sequences found in the phage display data and the best scoring peptides obtained using the different SH3-specific matrices was calculated according to the formula:

$$\rho_{X,Y} = \frac{\mathrm{cov}(X, Y)}{\sigma_X^* \sigma_Y}$$

(see Figure 4(c)), where X is the ensemble of the input values (Figure 4(a)), Y is the ensemble of the output values (Figure 4b), cov(X,Y) is X and Y covariance, σ_x and σ_Y are the standard deviations of the X and Y values, respectively.

Acknowledgments

The authors are grateful to Dr A. M. Lesk for helpful advice and discussion and for critically reading the manuscript. We are grateful to Luisa Montecchi Palazzi for valuable help. This work was supported by C.N.R. (target project Biotechnology) to M.H.C. and G.C.; the EU Biotechnology Program Framework 4 and AIRC. We would like to thank the members of the Molecular Genetics lab for providing experimental results before publication.

References

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389-3402.

Ando, A., Yonezawa, K., Gout, I., Nakata, T., Ueda, H., Hara, K., Kitamura, Y., Noda, Y., Takenawa, T.,

Hirokawa, N., *et al*. (1994). A complex of GRB2-dynamin binds to tyrosine-phosphorylated insulin receptor substrate-1 after insulin treatment. *EMBO J.* **13**, 3033-3038.

Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Finn, R. D. & Sonnhammer, E. L. (1999). Pfam 3.1: 1313 multiple alignments match the majority of proteins. *Nucl. Acids Res.* **27**, 260-262.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535-542.

Bunnell, S. C., Henry, P. A., Kolluri, R., Kirchhausen, T., Rickles, R. J. & Berg, L. J. (1996). Identification of Itk/Tsk Src homology 3 domain ligands. *J. Biol. Chem.* **271**, 25646-25656.

Cestra, G., Castagnoli, L., Dente, L., Minenkova, O., Petrelli, A., Migone, N., Hoffmuller, U., Schneider-Mergener, J. & Cesareni, G. (1999). The SH3 domain of endophilin and amphiphysin bind to the proline rich region of synaptojanin at distinct sites that display an unconventional binding specificity. *J. Biol. Chem.* **274**, 32001-32007.

Cicchetti, P., Mayer, B. J., Thiel, 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.

De Rinaldis, M., Ausiello, G., Cesareni, G. & Helmer-Citterich, M. (1998). Three-dimensional profiles: a new tool to identify protein surface similarities. *J. Mol. Biol.* **284**, 1211-1221.

Genetics Computer Group (GCG) (1999). *Wisconsin Package Version 9.0*, , Madison, WI.

Holmes, T. C., Fadool, D. A., Ren, R. & Levitan, I. B. (1996). Association of Src tyrosine kinase with a human potassium channel mediated by SH3 domain. *Science,* **274**, 2089-2091.

Howlett, C. J., Bisson, S. A., Resek, M. E., Tigley, A. W. & Robbins, S. M. (1999). The proto-oncogene p120 (Cbl) is a downstream substrate of the Hck protein-tyrosine kinase. *Biochem. Biophys. Res. Commun.* **257**, 129-138.

Hunter, S., Burton, E. A., Wu, S. C. & Anderson, S. M. (1999). Fyn associates with Cbl and phosphorylates tyrosine 731 in Cbl, a binding site for phosphatidylinositol3-kinase. *J. Biol. Chem.* **274**, 2097-2106.

Lee, C. H., Saksela, K., Mirza, U. A., Chait, B. T. & Kuriyan, J. (1996). Crystal structure of the conserved core of HIV-1 Nef complexed with a Src family SH3 domain. *Cell,* **85**, 931-942.

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.

Mongiovì, A. M., Romano, P. R., Panni, S., Mendoza, M., Wong, W. T., Musacchio, A., Cesareni, G. & Di Fiore, P. P. (1999). A novel peptide- SH3 interaction. *EMBO J.* **18**, 5300-5309.

Musacchio, A., Noble, M., Pauptit, R., Wierenga, R. & Saraste, M. (1992). Crystal structure of a Src-homology 3 (SH3) domain. *Nature,* **359**, 851-855.

Musacchio, A., Wilmanns, M. & Saraste, M. (1994a). Structure and function of the SH3 domain. *Prog. Biophys. Mol. Biol.* **61**, 283-97.

Musacchio, A., Saraste, M. & Wilmanns, M. (1994b). High-resolution crystal structures of tyrosine kinase SH3 domains complexed with proline-rich peptides. *Nature Struct. Biol.* **1**, 546-551.

Otsu, M., Hiles, I., Gout, I., Fry, M. J., Ruiz-Larrea, F., Panayotou, G., Thompson, A., Dhand, R., Hsuan, J., Totty, N., Smith, A. D., Morgan, S. J., Courtneidge, S. A., Parker, P. J. & Waterfield, M. D. (1991). Characterization of two 85 kd proteins that associate with receptor tyrosine kinase, middle-T/pp60$^{c\text{-}src}$ complexes, and PI3k-kinase. *Cell,* **65**, 91-104.

Park, R. K., Kyono, W. T., Liu, Y. & Durden, D. L. (1998). CBL_GRB2 interaction in myeloid immunoreceptor tyrosine activation motif signaling. *J. Immunol.* **160**, 5018-5027.

Pisabarro, M. T. & Serrano, L. (1996). Rational design of specific high-affinity peptide ligands for the Abl-SH3 domain. *Biochemistry,* **35**, 10634-10640.

Pisabarro, M. T., Serrano, L. & Wilmanns, M. (1998). Crystal structure of the Abl-SH3 domain complexed with a designed high-affinity peptide ligand: implications for SH3-ligand interactions. *J. Mol. Biol.* **281**, 513-521.

Raabe, T., Olivier, J. P., Dickson, B., Liu, X., Gish, G. D., Pawson, T. & Hafen, E. (1995). Biochemical and genetic analysis of the Drk SH2/SH3 adaptor protein of Drosophila. *EMBO J.* **14**, 2509-2518.

Rickles, R. J., Botfield, M. C., Weng, Z., Taylor, J. A., Green, O. M., Brugge, J. S. & Zoller, M. J. (1994). Identification of Src, Fyn, Lyn, PI3k and Abl SH3 domain ligands using phage display libraries. *EMBO J.* **13**, 5598-5604.

Rickles, R. J., Botfield, M. C., Zhou, X. M., Henry, P. A., Brugge, J. S. & Zoller, M. J. (1995). Phage display selection of ligand residues important for Src homology 3 domain binding specificity. *Proc. Natl Acad. Sci. USA,* **92**, 10909-10913.

Sahr, K. E., Laurila, P., Kotula, L., Scarpa, A. L., Coupal, E., Leto, T. L., Linnenbach, A. J., Winckelmann, J. C., Speicher, D. W., Marchesi, V. T., Curtis, P. J. & Forget, B. J. (1990). The complete cDNA and polypeptide sequences of human erythroid alpha-spectrin. *J. Biol. Chem.* **265**, 4434-4443.

Saksela, K., Cheng, G. & Baltimore, D. (1995). Proline-rich (PxxP) motifs in HIV-1 Nef bind to SH3 domains of a subset of Src kinases and are required for the enhanced growth of Nef+ viruses but not for down-regulation of CD4. *EMBO J.* **14**, 484-491.

Sali, A. (1995). Modeling mutations and homologous proteins. *Curr. Opin. Biotechnol.* **4**, 437-451.

Sparks, A. B., Rider, J. E., Hoffman, N. G., Fowlkes, D. M., Quillam, L. A. & Kay, B. K. (1996). Distinct ligand preferences of Src homology 3 domain from Src, Yes, Abl, Cortactin, p53bp2, Plc-γ, Crk and Grb2. *Proc. Natl Acad. Sci. USA,* **93**, 1540-1544.

Srinivasan, N., Guruprasad, K. & Blundell, T. L. (1996). Comparative modelling of proteins. In *Protein Structure Prediction, A Practical Approach* (Sternberg, M. J. E., Rickwood, D. & Hames, B. D., eds), Oxford University Press, Oxford.

Stahl, M. L., Ferenz, C. R., Kelleher, K. L., Kriz, R. W. & Knopf, J. L. (1988). Sequence similarity of phospholipase C with the non-catalytic region of src. *Nature,* **332**, 269-272.

Trahey, M., Wong, G., Halenbeck, R., Rubinfeld, B., Martin, G. A., Ladner, M., Long, C. M., Crosier, W. J., Koths, K. & McCormick, F. (1988). Molecular cloning of two types of GAP complementary DNA from human placenta. *Science,* **242**, 1697-1700.

Vidal, M., Montiel, J. L., Cussac, D., Cornille, F., Duchesne, M., Parker, F., Tocqué, B., Roques, B. P. & Garbay, C. (1998). Differential interactions of the

growth factor receptor-bound protein 2 N-SH3 domain with son of sevenless and dynamin. Potential role in the Ras-dependent signaling pathway. *J. Biol. Chem.* **273**, 5343-5348.

Wu, X., Knudsen, B., Feller, S. M., Zheng, J., Sali, A., Cowburn, D., Hanafusa, H. & Kuryian, J. (1995). Structural basis for the specific interaction of lysine-containing proline-rich peptides with the N-terminal SH3 domain of c-Crk. *Structure,* **3**, 215-226.

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 the SH3 domains. *Cell,* **76**, 933-945.

Ziemnicka-Kotula, D., Xu, J., Gu, H., Potempska, A., Kim, K. S., Jenkins, E. C., Trenkner, E. & Kotula, L. (1998). Identification of a candidate human spectrin Src homology 3 domain-binding protein suggests a general mechanism of association of tyrosine kinases with the spectrin-based membrane skeleton. *J. Biol. Chem.* **273**, 13681-13692.

Edited by J. Thornton

(Received 23 September 1999; received in revised form 7 February 2000; accepted 10 February 2000)