THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 277, No. 24, Issue of June 14, pp. 21666–21674, 2002

Printed in U.S.A.

In Vitro Evolution of Recognition Specificity Mediated by SH3 Domains Reveals Target Recognition Rules*

Received for publication, October 10, 2001, and in revised form, February 14, 2002
Published, JBC Papers in Press, April 19, 2002, DOI 10.1074/jbc.M109788200

Simona Panni, Luciana Dente‡, and Gianni Cesareni§

From the Department of Biology, University of Rome Tor Vergata, Via della Ricerca Scientifica, Rome 00133, Italy

We have designed a repertoire of 10^7 different SH3 domains by grafting the residues that are represented in the binding surfaces of natural SH3 domains onto the scaffold of the human Abl-SH3 domain. This phage-displayed library was screened by affinity selection for SH3 domains that bind to the synthetic peptides, APTYPPPLPP and LSSRPLPTLPSP, which are peptide ligands for the human Abl or Src SH3 domains, respectively. By characterizing the isolates, we have observed that as few as two or three amino acid substitutions lead to dramatic changes in recognition specificity. We propose that the ability to shift recognition specificity with a small number of amino acid replacements is an important evolutionary characteristic of protein binding modules. Furthermore, we have used the information obtained by these *in vitro* evolution experiments to generate a scoring matrix that evaluates the probability that any SH3 domain binds to the peptide ligands for the Abl and Src SH3 domains. A table of predictions for the 28 SH3 domains of baker's yeast is presented.

Protein interaction inside the cell is often mediated by families of protein modules that occur in proteins of very different function (1). Each module is specialized in recognizing specific features of the protein surface: for example, SH3 domains bind to peptides that fold into a poly-proline helix, SH2 domains have affinity for peptides containing phosphorylated tyrosines, and PDZ domains recognize carboxyl-terminal peptides (1–4). Within each type of module, molecular recognition is modulated by changing the chemical characteristics of the domain surface, which in turn determines the preference for different contexts of the common target structural theme (for instance the poly-proline helix for SH3 domains).

Some reports have suggested the existence of a molecular recognition code for protein interaction modules (5). Although everyone accepts that, ultimately, protein recognition must be based on the fundamental laws of physics and chemistry, simple rules like "residue A at position *x* in the domain calls for residue B at position *y* in the ligand" have been proposed (6–10). Most of these rules have had limited success in accurately predicting protein-protein interactions. The prospect of being able to crack some sort of recognition code relies on the assumption that the solutions to the problem of binding a specific domain are concentrated in a small cluster in the sequence/structure space. Alternatively, the problem of finding the *consensus* ligand for any receptor domain would change into the more difficult problem of finding "all" the *consensus* sequences that share the potential of binding to that receptor.

Another intriguing issue is why relatively few scaffolds have been selected during evolution for the purpose of maintaining a rather complex protein interaction network. Apparently, once a solution for binding to a specific surface feature is found, that solution is explored extensively to find new binding specificities. An advantageous characteristic of a protein recognition domain would be the ability to modify recognition specificity by a limited number of changes of its primary sequence without extensive structural rearrangements and without the need to explore long evolutionary pathways encompassing non-functional states. This would permit, by accumulation of a small number of mutations, the growth of a functionally large natural repertoire, wherefrom the selection of new binding specificity could be possible.

To obtain evidence of this postulated malleability of protein recognition modules and of the elusive recognition code, we chose to focus on SH3 domains, because they represent the most numerous family of protein interaction modules in eukaryotic genomes (11) and because the structures of several SH3 domains, both isolated and in complex with their targets, have been determined (reviewed in Refs. 12–14). SH3 domains bind to their targets by accommodating a peptide segment, which is folded into a poly-proline II helix, into a binding cleft formed by three molecular pockets of their surface. Two of the pockets are hydrophobic and host the P*XX*P motif, which is considered the signature of SH3 ligands, whereas the third is negatively charged and determines the specificity and the orientation of the ligand by hosting a positively charged residue that either precedes or follows the P*XX*P motif (15–17). Peptides characterized by the R*XX*P*XX*P motif bind in an orientation called "class I" orientation, whereas peptides displaying the P*XX*P*X*R motif bind in the opposite orientation and are termed "class II." A third class of ligands, exemplified by peptide ligands of the Abl SH3 domain, do not contain a positively charged residue, are characterized by the consensus P*X@XX*P*XX*P (@ = aromatic residue), and bind in the class I orientation (2, 16, 18). Although SH3 domains share ~30% amino acid similarity, and a common fold, the ability to predict the peptide recognition specificity of any given SH3 domain has been challenging.

Over the past few years, several groups have used combinatorial peptide libraries to characterize the recognition specificity of protein interaction modules (reviewed in Ref. 19). We have generated an SH3 repertoire by modulating the chemical characteristics of the ligand binding surface of a specific SH3 scaffold. We have then used this repertoire to characterize potential evolutionary pathways that would change the domain

* This work was supported by grants from Associazione Italiana per la Ricerca sul Cancro, Target project biotechnology from CNR and the European Union Biotechnology project. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Present address: Dept. of Physiology and Biochemistry, University of Pisa, Via Carducci, 56010 Pisa, Italy.

§ To whom correspondence should be addressed. Tel.: 39-06-7259-4315; Fax: 39-06-2023-500; E-mail: cesareni@uniroma2.it.

recognition specificity. Finally, we have exploited the information obtained from the characterization of SH3 domains that bind to two specific peptides to produce a position specific scoring matrix and to develop an algorithm that permits to infer the molecular recognition properties of SH3 domains.

EXPERIMENTAL PROCEDURES

Phage, Plasmids, and Fusion Proteins—The Lambda vector, used for the display of the SH3 repertoire, was derived from λpRH825 (20) by deletion of an *Xba*I fragment. This deletion removes one of the two *Lox*P sites and the entire plasmid DNA. Like the original vector, this derivative has a second copy of the *D* gene, but includes at its 3′-end two sites, *Spe*I and *Not*I, that can be used for the insertion of DNA fragments (21) and is more stable (*i.e.* the second *D* gene is not deleted after several growth cycles). The constructions of the GST-SH3[1] fusion plasmids expressing yeast *SH3* domains have been described elsewhere (22).

Library Construction—To assemble a hybrid sequence encoding an Abl SH3 domain with a discrete, degenerate codons, we designed a set of partially overlapping oligonucleotides, whose sequences are reported here: R287, CCACCCACGAATTCAACCTGTTCGTT; R286, GCTCTGYWCGACTWCGTTGCTVNSRNSVVSRVSRMSCTGTCCATCACCAAAGGT; R288, GAAAAACTGCGTGTTCTGGGTTACRVCCACAACGGTRRSTGGTGCGAAGCTCAG; R277, GTTCCGTCCAMCTWCATCACCCCGGTGGATCCTCTCCGCC; R278, AGCAACGWAGTCGWRCAGAGCAACGAACAGGTTGAATTCGTGGGTGG; R279, GTAACCCAGAACACGCAGTTTTTCACCTTTGGTGATGGACAG; R280, TGATGWAGKTGGACGGAACSHMACCCTGACCGTTTTTGGTCTGAGCTTCGCACCA; and R289, GGCGGAGAGGATCCACCGGGG.

Sequence symbols are according to IUPAC format: W = A, T; R = A, G; M = A, C; Y = C, T; N = G, A, T, C; S = C, G; H = A, C, T; K = G, T; V = A, C, G. The oligonucleotides were phosphorylated at their 5′ termini with T4 polynucleotide kinase and the hybrid gene assembled by raising the temperature to 94 °C and then slowly cooling down to 37 °C. The partially assembled gene was amplified by polymerase chain reaction using primers R314 (CCACCGACACTAGTAATGACCCCAACCTTTTCG) and R344 (GCGCATGCGCGGCCGCGACTGTTCACCGGGGTGATGWA), which contain *Spe*I and a *Not*I sites, respectively. After amplification, the DNA was digested with *Spe*I and *Not*I restriction enzymes and purified using the QIAquick-spin PCR purification kit (Qiagen). Two micrograms of *Spe*I/*Not*I-digested L14 DNA were ligated to 10–30 ng of purified insert. The ligation mixture was packaged using a Lambda packaging kit (Amersham Biosciences, Inc.) and plated with *Escherichia coli* BB4 on Luria broth (LB) plates, containing 10 mM $MgCl_2$ and 0.2% maltose. After 8 h at 37 °C, phage particles were recovered by adding 10 ml/plate of SM (10 mM Tris-HCl, 0.2 M NaCl, 5 mM $MgCl_2$) and eluted for 3 h at 4 °C. The complexity of the library (number of independent clones) was 10^7 plaque forming units (pfu), and the titer of the eluted phages was about 10^{10} pfu/ml.

Affinity Selection—Affinity selections with poly-proline peptides were performed in microtiter plates (Nunc) coated overnight a 4 °C with 5 μg/ml streptavidin (Sigma) in 100 μl of PBS (10 mM Na_2HPO_4/KH_2PO_4 (pH 7.2), 150 mM NaCl). In these conditions ~0.4 μg of streptavidin remain bound to the plastic well. The coated plates were washed ten times with PBS-0.05% Tween 20 and incubated for 30 min at 25 °C with biotinylated peptides (10 μM in PBS). Plates were washed again and blocked for 1 h at 25 °C with 4% bovine serum albumin (BSA) in PBS. About 10^9 phage particles from the library (*i.e.* ~100 library equivalents) were added to each well and incubated at 4 °C overnight. After five washes with PBS-0.05% Tween 20, the selected phages were recovered by adding 100 μl of BB4 cells in 10 mM $MgSO_4$ and incubating for 30 min at 37 °C. The infected cells were then plated with additional indicator bacteria and top agar and grown overnight at 37 °C, and phage were eluted as described above. After titrating the number of phage particles in the phage suspension, the selection cycle was carried out for two more times.

Plaque Assay—Phage plaques from each selection round were transferred onto nitrocellulose membranes by overlaying the membrane onto the top agar of the Petri plate and by incubating for 4 h at 37 °C. Filters were blocked for 2 h at room temperature in PBS containing 4% BSA. Biotinylated peptides were bound to streptavidin-alkaline phosphatase (Sigma Chemical Co.) for 30 min at room temperature and incubated with filters at 4 °C overnight. After five washes with PBS-0.05% Tween, positive plaques were revealed by a colorimetric reaction using nitro blue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate as substrates (Sigma).

ELISA—ELISA assays were performed in microtiter plates (Nunc), coated with biotinylated peptides as described above. Positive plaques were collected and used to prepare plate lysates. Phages were eluted from a 6-cm plate with 2 ml of SM buffer and precipitated with 1 volume of 20% polyethylene glycol-2.5 M NaCl for 1 h on ice. After centrifugation, the pellet was resuspended in 100 μl of SM buffer and 10 μl of suspension, containing about 10^7 phages in 100 μl of PBS-4% BSA, was added to each microtiter well. After 10 washes with PBS-0.05% Tween 20, plates were incubated with an anti-lambda polyclonal antibody for 1 h at room temperature and then with an anti-rabbit alkaline phosphatase-conjugated antibody (Sigma). Retention of the phage particles in the microtiter plate wells was revealed by adding 100 μl/well of a 1 μg/ml solution of *p*-nitrophenyl phosphate in 50 mM $NaHCO_3$, pH 9.6, 20 mM $MgCl_2$.

To compare the binding strength of different GST-SH3 proteins, microtiter wells, which had been coated with biotinylated peptides as described above, were incubated with serial dilutions of fusion proteins in PBS-4% BSA for 2 h at 4 °C. After 10 washes, a goat polyclonal anti-GST serum (Amersham Biosciences, Inc.) was added in PBS-4% BSA for 1 h at room temperature, followed by incubation with anti-goat Ig antibodies conjugated to alkaline phosphatase (Sigma). The reaction was revealed as described above. Data were fitted using a single class of an equivalent binding site equation.

Site-specific Mutagenesis—Mutant Abl-SH3 coding sequences were assembled from two overlapping DNA fragments obtained by PCR amplification with pairs of complementary primers, each carrying the mutated sequence, and two primers that prime from the 5′- and 3′-ends of the wild type SH3 domain sequence. This latter pair of primers, which were common to all the mutagenesis experiments, contain *Bam*HI and *EcoR*I restriction sites, respectively, for directional cloning into the pGEX2T vector; their sequences are: R596,GGCAGCTAGGATCCAATGACCCCAACC, and R597, GGACGAGTGAATTCCACTGTTGACTGG. The mutagenic primers used in the first PCR amplification step were: R599, AGTTATGCTTAGGTCGTTATCTCCACTGGC and R598, AGTGGAGATAACGACCTAAGCATAACTAAAG for T79D; R618, GTCTTAGGCTATTGTCACAATGGGGAATG and R619, ACACCATTCCCCATTGTGACAATAGCCTAAGAC for N94C; R620, CAAAAATGGCCAAGGCTATGTCCCAAGCAAC and R621, GTAGTTGCTTGGGACATAGCCTTGGCCATT for W110Y; R764, CAAAAATGGCCAAGGCTTGGTCCCAAGCAAC and R765, GTAGTTGCTTGGGACCAAGCCTTGGCCT for W110V; R766, CAAAAATGGCCAAGGCGTGGTCCCAAGCAAC and R767, GTAGTTGCTTGGGACCACGCCTTGGCCT for W110L.

RESULTS

Design of an SH3 Repertoire—As a scaffold for our general repertoire, we chose a well-characterized SH3 domain, the one of the human protein kinase Abl, whose structure in complex with the peptide ligand APTMPPPLPP (1Abo) has been determined at high resolution by x-ray crystallography (23). Inspection of the three-dimensional structure of the complex permitted the identification of the residues within the SH3 domain that make contact with the ligand and were, therefore, likely to be involved in determining its recognition specificity. To identify the positions to diversify within this scaffold, we aligned 560 different eukaryotic SH3 domains, whose sequence was available in the PFAM data base (24) when we started the project, and determined the frequency of occurrence of each amino acid residue at each position that potentially contact its ligand. Finally we designed a hybrid gene that encodes the Abl SH3 domain but has a degenerate sequence in the codons for the contact residues. The extent of the degeneracy at each position is a compromise between the desirability of obtaining a repertoire that includes all the residues that are found at those positions in natural SH3 domains and the necessity of maintaining the size of the repertoire within the limits set by the transformation efficiency of *E. coli*. Although the theoretical complexity of the repertoire is ~2×10^9, complete coverage was not possible because of the design limitations imposed by

[1] The abbreviations used are: GST, glutathione *S*-transferase; pfu, plaque-forming units; PBS, phosphate-buffered saline; BSA, bovine serum albumin; ELISA, enzyme-linked immunosorbent assay.

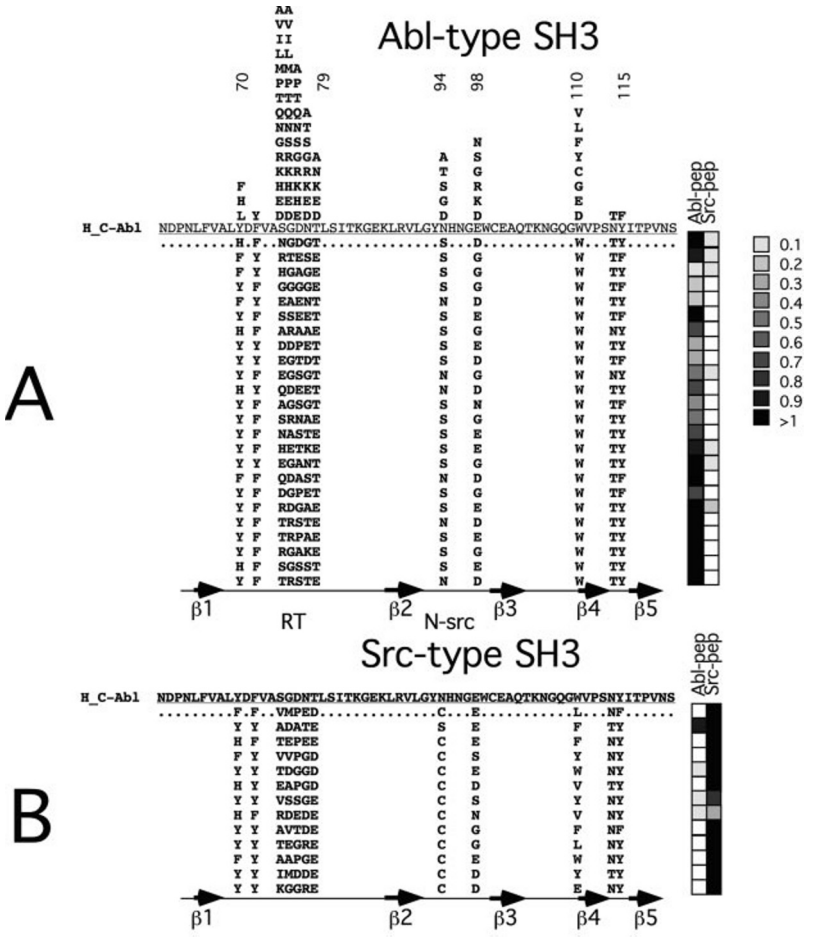


FIG. 1. **Primary structure of the SH3 domains that bind to two types of proline rich peptides.** *A*, *underlined* is the amino acid sequence of the SH3 domain of the human Abl tyrosine kinase. Above this sequence, we list the residues that were allowed in each corresponding position in the design of the degenerate Abl-SH3 coding region. *Numbers* refer to residue positions in the Abl protein. Below the Abl-SH3 sequence, we report the residues that were observed in the degenerate positions in the SH3 domains, which were selected for binding to the APTYPPPLPP peptide (Abl-pep). *B*, as in *A*, but the reported sequences correspond to SH3 domains that were selected for binding to the LSSRPLPTLPSP peptide (Src-pep). The selected domains were tested, in an ELISA format, for binding to biotinylated Abl-pep or to Src-pep that had been immobilized on streptavidin-coated microtiter plate wells. The strength of binding in the wells is depicted on the *right* with a *gray intensity scale*. The *arrow diagram* in the *lower part* represents the five-stranded β structure of SH3 domains.

the genetic code; however, the repertoire includes at least 90% of the natural variability at each contact position.

The library of variant Abl SH3 domains was generated in bacteriophage lambda (25). The coding region for the domain was fused to the carboxyl terminus of the D capsid protein, which tolerates inserts and exists at ~400 copies on each virus particle. Ten million recombinants were created, representing ~1% of the theoretical size of the repertoire. By sequencing random isolates, we confirmed that 60% of the clones contain in-frame fusions of the D and SH3 domain coding regions and that the SH3 domains display random combinations of the allowed residues at the degenerate positions.

Selection of SH3 Domains That Bind to Two Different Peptides—Because the SH3 domains in the repertoire are expressed on the surface of bacteriophage lambda, ligands for any kind of poly-proline peptide can be screened by a plaque lift assay or selected by panning with a target peptide (26). We first asked whether, by these techniques, we could recover from the library SH3 domains that would bind to the peptide APTYPPPLPP (Abl-pep), which is a high affinity ligand of the Abl-SH3 domain (27). Although the percentage of clones in the library that binds to Abl-pep is less than 0.1%, after one or two panning cycles, ~1 and 25% of the clones display an SH3 domain that binds with apparent affinity that is comparable to the one of the wild type SH3 domain (data not shown). Binding is specific, because the vast majority of the selected clones do not bind to a peptide ligand (LSSRPLPTLPSP; Src-pep) for the human Src SH3 domain (Fig. 1*A*) (18, 28).

Comparison of the residues that are present or missing at specific degenerate positions of the domains selected after two panning cycles (Fig. 1*A*) permits one to identify the requirements for binding of an SH3 domain to the Abl-pep probe. Some of the degenerate positions did not show any residue preference within the limits of the variability engineered in the repertoire; for instance, at positions 70 and 115, tyrosine and phenylalanine are equally acceptable. By contrast, at position 110, of the 9 residues tested at this position, only tryptophan is found in the SH3 domains selected with the Abl-pep. Interestingly, the threonine present at position 79, which flanks the P-3 peptide binding site and was previously identified as a key residue in determining the preference for a tyrosine instead of a positively charged residue at P-3 (16), can be replaced by a glutamate (but not aspartate) in the SH3 domain scaffold without destroying binding.

Next, we asked whether the SH3 repertoire contained domains that would bind to an unrelated peptide that is normally recognized by a different class of SH3 domains. For this purpose we selected the peptide sequence, LSSRPLPTLPSP (Src-pep), which efficiently binds to SH3 domains of the Src family (*i.e.* Src, Yes, Fyn). The SH3 domains of these protein-tyrosine kinases are only 30% similar in primary sequence to the SH3 domain of Abl and contain an n-Src loop that is one residue shorter than in Abl. After three rounds of affinity selection, we could identify SH3 domains that bind to Src-pep, although at a much lower frequency than observed with the Abl-pep. Sequence analysis of the isolates revealed that all, but one, had a Cys residue at position 94. Interestingly, this residue was not included in the original repertoire design, because it is not present in any natural SH3 domain, and thus this codon likely resulted from errors in oligonucleotide synthesis.

The residues that are enriched with respect to the unselected repertoire in the two panning experiments can be compared in Fig. 2 where the font size is proportional to the enrichment factor for each residue in each degenerate position. Residues

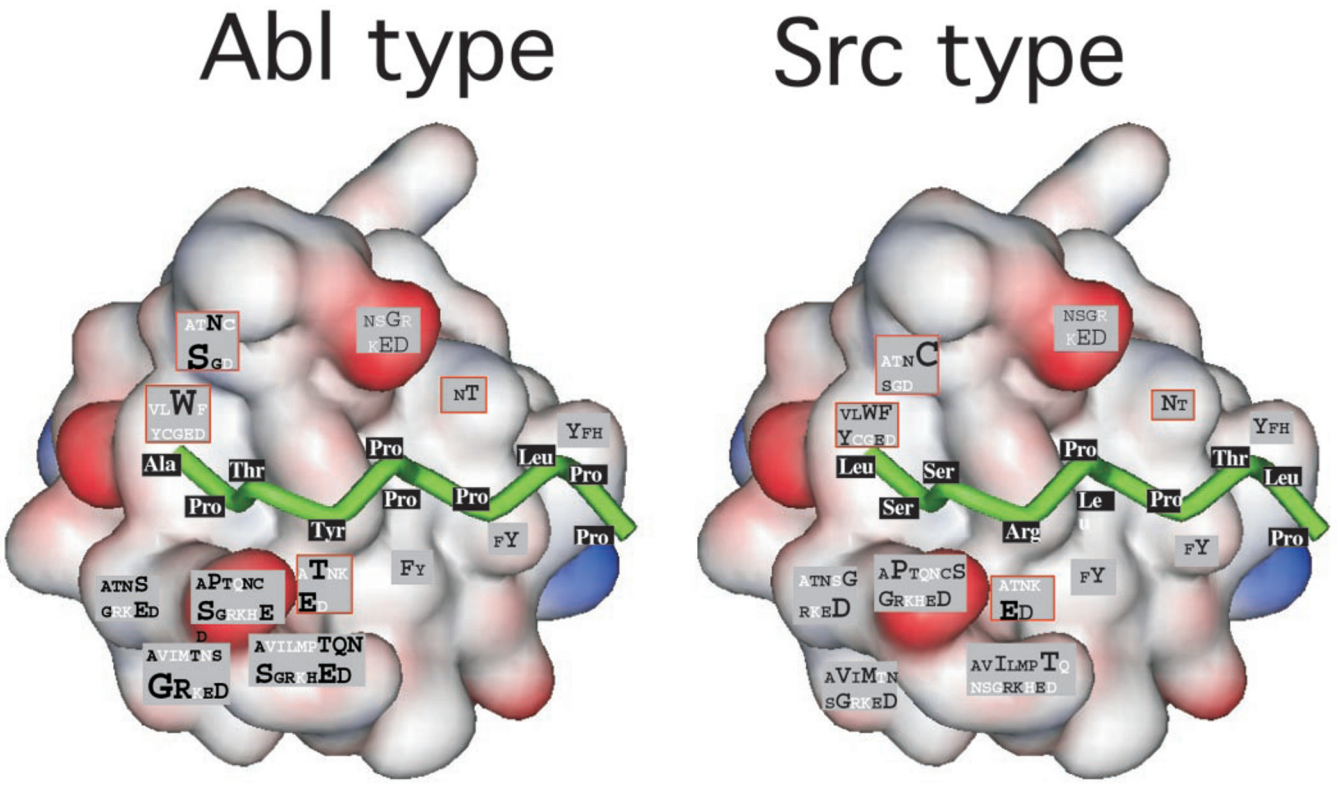


FIG. 2. **Schematic representation of the relative abundance of the residues observed at each degenerate position of the SH3 repertoire in the two selection experiments.** In each *gray box*, corresponding to the twelve degenerate positions of the SH3 repertoire, the *font size* is proportional to the relative frequency of the specific residue in the selected *versus* the unselected repertoire. Residues that were never observed in the sample of selected domains, which were characterized in this work, although present in the repertoire design, are in *reverse color*. The amino acid frequencies were calculated from the data in Fig. 2, and the corresponding *boxes* are depicted on a surface representation (*colored for charge*) of the Abl SH3 domain. Positions that showed a marked difference in the two experiments are framed in *red*.

that, although included in the original repertoire design, are never found in the selected domains are in *reverse color*. This comparison reveals more or less subtle differences whose rationalization is not always straightforward. The residues forming the two hydrophobic pockets that host the P*XX*P motif do not display a high selectivity when tested with the two peptides used in this experiment. At positions 70 and 115 both Src-pep and Abl-pep ligands preferentially display a Tyr. However, also the other residues (His and Phe) allowed by the repertoire design are accepted, irrespective of the ligand considered, albeit at a lower frequency. At position 98 both peptides prefer negative residues and do not tolerate positive ones. Finally, a certain degree of selectivity is determined by the identity of the amino acids at position 114 in the 3/10 helix preceding the β5 strand and in residue 72, which participates in the formation of the central pocket. At position 114, Thr favors the interaction with the Abl-pep whereas an Asn determines the preference for the Src-pep. Note that the wild type Abl-SH3 domain has an Asn at this position. By contrast, at position 72, the Abl and Src peptides favor Phe and Tyr, respectively.

However, the most striking differences, emerging from the comparison of the amino acid sequences of the two families of the selected domains, involve residues that flank the third hydrophilic pocket, which hosts the amino side of class I peptides; namely, residue 110 that is only Trp in Abl-pep ligands and residue 94 that is a Cys in the vast majority of the SH3 domains selected for binding to the Src-pep. Finally, as already pointed out, position 79 is a major determinant for the selection of the residue at position P-3 in the poly proline peptide with Thr favoring hydrophobic and aromatic residues and Asp determining the preference for a positively charged residue.

Characterization of the Molecular Determinants of Peptide Recognition Specificity—From the two selection experiments described above, we have identified positions 79, 94, and 110 within the scaffold of the Abl SH3 domain to be the major determinants that discriminate between binding to the Abl and Src peptide ligands. However, this approach cannot exclude that other residues or combination of residues may play an important, albeit less prominent, role. To ascertain how few of these residues should one change to reverse the Abl SH3 ligand specificity, we constructed a series of site-directed mutations by changing residues 79, 94, and 110 in the wild type Abl SH3 domain. (For simplicity, these mutant domains will be referred to by three-letter code such that the three letters correspond to the residues at position 79, 94, and 110, respectively.) The mutant domains were expressed as fusions to the glutathione *S*-transferase (GST) protein, and their ability to bind different peptides was evaluated by ELISA. As seen in Fig. 3*A*, as few as two amino acid replacements, either at positions 79 and 94 or 94 and 110, are sufficient to switch the peptide ligand preference of the Abl SH3 domain. By changing the three residues into TCL or TCV the binding propensity is completely reversed and the apparent affinity of the two mutant peptides becomes comparable to that of an SH3 domain of Src (Fig. 3*B*).

The characterization of the domains selected from the SH3 repertoire for their ability to bind to Abl-pep or Src-pep was not sufficient to establish a correlation between specific residues in the SH3 binding surface and the preference for specific residues at a given position within the ligand peptide. To this possibility, we have performed a complementary set of experiments by selecting peptide ligands from a phage-displayed combinatorial 9-mer peptide library using as baits GST fusions

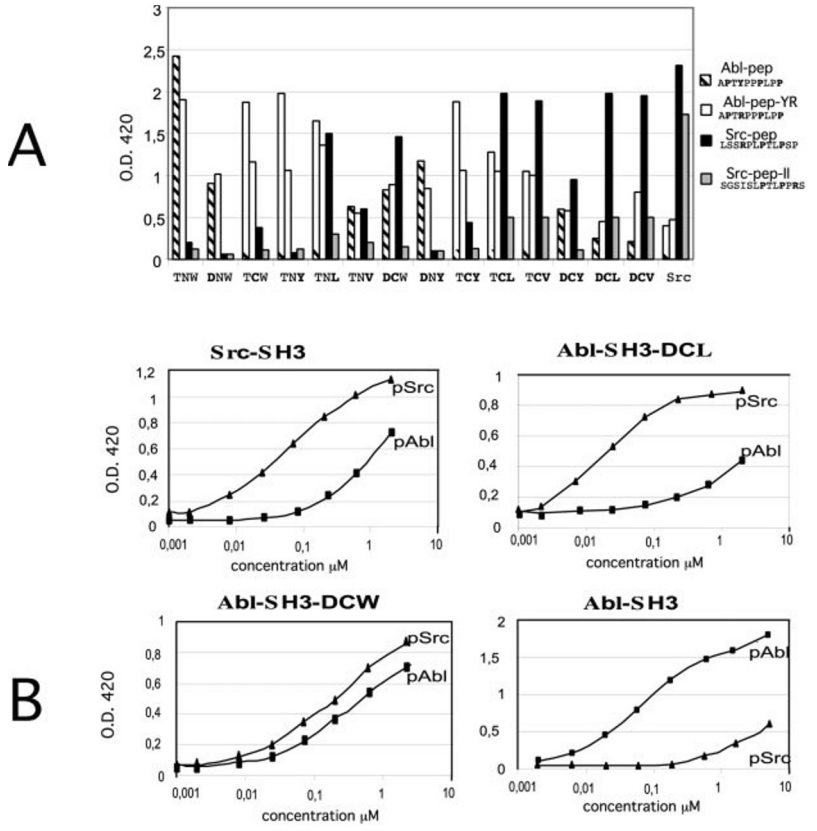


FIG. 3. **Solid phase assay.** *A*, four different biotinylated peptides (10 μM) were adsorbed to a microtiter wells, which had been previously coated with 5 μg/ml streptavidin. 0.3 μg of a mutant Abl-SH3 domain (fused to GST) was added to each well, and its binding was monitored by probing with polyclonal anti-GST goat serum and an anti-goat alkaline-phosphatase-conjugated antibody. The amino acid sequences of the peptides used in the assay are reported to the *right* of the histogram. The mutant domains are indicated with a three-letter code according to the residues present at position 79, 94, and 110. Abl-pep-YR is an Abl-pep derivative in which the Tyr at position −3 has been substituted by an Arg. Src-pep II, a typical class II peptide, is used as a control in this experiment. *B*, binding of four different SH3 domains was probed, as in *A*, at different domain concentrations. In this experiment the wells were coated with a peptide concentration of 0.1 μM. Each data point is the average of two independent measurements that did not differ by more than 10%.

to the wild type Abl-SH3 domain and five variants carrying one two or three mutations at positions 79, 94, and 110. The six GST fusions all select peptides containing the P*XX*P consensus, thereby confirming that the changes that we have introduced do not affect the preference for the typical SH3 recognition motif (Fig. 4). The wild type SH3 domain of Abl, which has residues TNW at these three positions, prefers peptide ligands that have an aromatic side chain at P-3 and a Pro or a Phe at P-5. Although, the DNW mutant did not discriminate between the Abl-pep and a peptide derivative in which the Tyr was changed into an Arg (Fig. 3*A*) from the peptide selection experiment, it is clear that replacement of the Thr residue with an Asp at position 79 in the Abl SH3 domain scaffold favors the selection of peptide ligands with an Arg at P-3. This preference is a characteristic of all the domains that have Asp at 79. The TCW mutant (*i.e.* Cys in place of Asn at position 94) has a dramatically altered ligand specificity and selects peptides with Arg at P-5 and a Leu at P-1. In the double mutant, DCW, the selected peptides match the specificity of the Src SH3 domain, with Arg at P-3 and Leu at P-1 in the consensus. Finally, by substituting W110 with a smaller residue in the triple mutants DCY and DCV, the selected peptides also prefer Pro at P-2. In conclusion, this set of experiments has permitted to identify a strong correlation between the Cys at position 94 in the SH3 domain and the Leu at position P-1 in the ligand peptide. Furthermore, the substitution of the Thr at 79 with an Asp shifts the preference for peptides that have a hydrophobic side chain at P-3 to peptides that have an Arg at the same position. Collectively, these results contribute to characterize a number of potential evolutionary pathways leading, via three single point mutations, from an SH3 domain that binds to APTYPPPLPP and not to LSSRPLPTLPSP, to a second SH3 domain with opposite specificity (Fig. 5).

The SH3 Profile Method to Infer the Recognition Specificity of the Yeast SH3 Domains—We have recently shown that the *consensus* ligand peptides obtained by panning synthetic peptide libraries with SH3 domains can be used to develop position-specific scoring matrices that have a high predictive value in the identification of the physiological partners of SH3 containing proteins (22). The results presented in this manuscript provide complementary information in that they permit the identification of a *consensus* "SH3 binding surface" for recognition of a specific poly-proline peptide.

Thus we have used the information contained in the multiple sequence alignment of Fig. 1 to define, for any peptide ligand, a position-specific surface profile that characterizes SH3 domains that bind to that peptide. In the hypothesis that these profiles are scaffold-independent, they could be used to infer the recognition specificity of any uncharacterized SH3 domain.

We define the peptide-specific SH3 profile as a 12×20 matrix, where the twelve rows represent the twelve degenerate positions in our repertoire and the twenty columns the twenty amino acids. Each position of the matrix contains the ratio between the frequencies of the corresponding residue at that surface position in the selected and unselected repertoires. A peptide-specific score can then be assigned to any given SH3 by

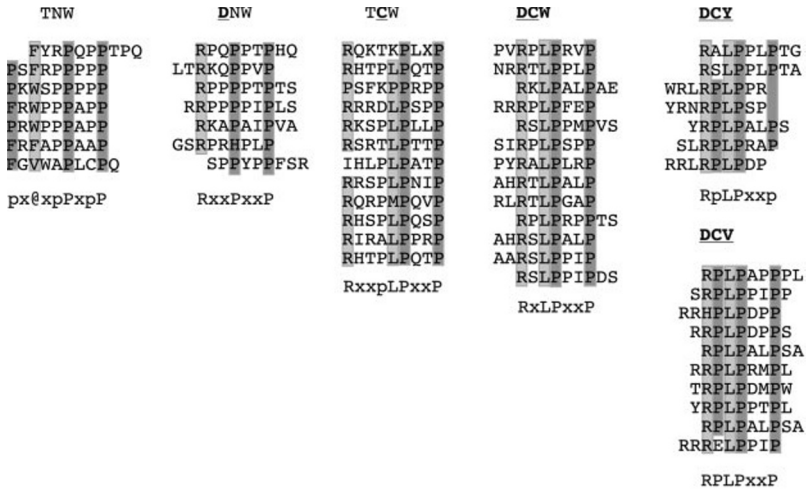
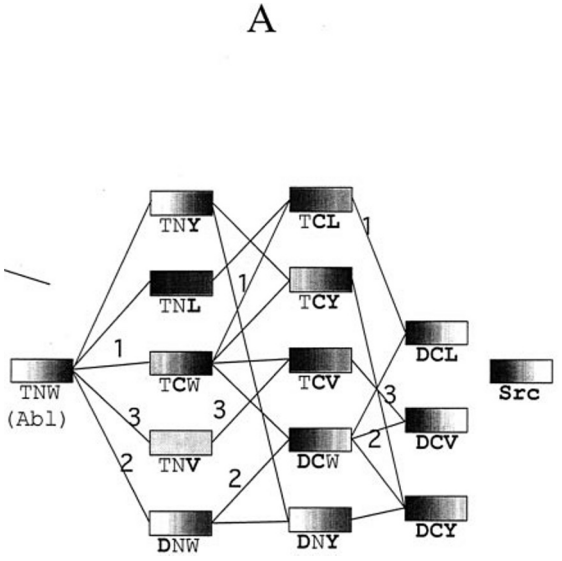


FIG. 4. **Preferred ligands of the different SH3 domain mutants.** Six SH3 domains were used to select by affinity a combinatorial peptide library of 9-mers displayed on the surface of the filamentous phage M13. After three rounds of selection, the binding of the selected clones was confirmed by ELISA and then sequenced to determine the amino acid sequence of the displayed peptides. The peptides are aligned, with the consensus sequence reported *below* each alignment. Whenever a residue is conserved in more than 90% of the peptides, it is indicated with a *capital letter* in the consensus, whereas residues that are conserved in more than 50% of the clones are in *small letters*. @ stands for an aromatic residue.



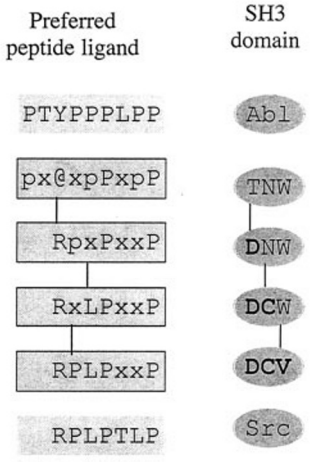


FIG. 5. **Schematic representation of potential evolutionary pathways.** *A*, each mutant Abl SH3 domain is represented as a *rectangle*. The domains that are linked by a single mutational event are joined by a *line*. The apparent affinity of each domain for Abl-pep and Src-pep, as deduced from the experiment in Fig. 4, is represented in a *gray intensity scale* on the *right* and *left side* of each *rectangle*, respectively. Representative evolutionary pathways, described in the main text, are labeled with *numbers*. *B*, details of the pathway labeled *2*, representing the coevolution of the SH3 specificity pocket and the sequence of the preferred ligand.

adding, for each of the twelve SH3 positions, the figure in the profile corresponding to the residue that is present at that position in the query SH3.

We have applied our scoring profile on the entire set of SH3 domains present in the baker's yeast, *Saccharomyces cerevisiae*, by ranking the 29 SH3 domains of this organism according to the probability that they would bind to the Abl-pep or to Src-pep (*white bars* in Fig. 6). To test these predictions, we overexpressed 24 of the domains as GST fusion proteins and examined binding to the peptides in an ELISA. Interestingly, the SH3 profile method correctly ranks in the top three positions the three SH3 domains, Abl-SH3, Myo5-SH3, and Myo3-SH3, that efficiently bind to the Abl-pep. On the other hand, although the Src-pep profile correctly ranked, among the six most probable ligands, five yeast domains that experimentally were found to bind to the Src-pep, the SH3 domain of Abp1 scored as a false positive. Furthermore the SH3 domain of the protein H_cSrc is not predicted among the best ligands. This is possibly a consequence of the oversimplification of the method that requires the sequences of the SH3 domains to be aligned also in regions that corresponds to loops of different length. The n-src loop of the SH3 domain of kinases of the Src family is one residue shorter than the one in Abl. Notwithstanding this limitation, by using the profiles derived from the screening of the present repertoire, we have correctly identified the yeast SH3 domains that bind to Abl-pep and to Src-pep, and we have only erroneously included Abp1 among the yeast domains that bind to Src-pep. Similar results were obtained by using a Hidden Markov Model approach (not shown). To evaluate the statistical significance of the prediction, we have generated 100 random profiles by simulating 100 different panning experiments. We have then used these profiles to rank the yeast SH3 domains, and we have compared the predictions with those obtained with the experimentally derived profiles. While the experimental profile ranks four SH3 ligands in the top four positions in the case of Src-pep and three in the top three for Abl-pep (Fig. 6), on average the mock profiles rank 1.3 (Src-pep) and 0.3 (Abl-pep) experimentally verified ligands in the top four positions, the best performance being positions three and one, respectively.

Thus, despite billions of years of evolution, the same function (binding to Abl-pep or to Src-pep) is associated to protein surfaces that can be identified, with sufficient confidence, by a profile derived by an *in vitro* evolution experiment of our synthetic repertoire.

DISCUSSION

A Repertoire of SH3 Binding Surfaces That Is Similar to the Natural Repertoire—We have designed and assembled a repertoire of SH3 domains by grafting different combinations of the residues observed in the ligand binding pockets of natural domains onto the scaffold of the Abl-SH3 domain. This repertoire was used to search for elements that bind to several poly-proline peptides (some of these selections have not been

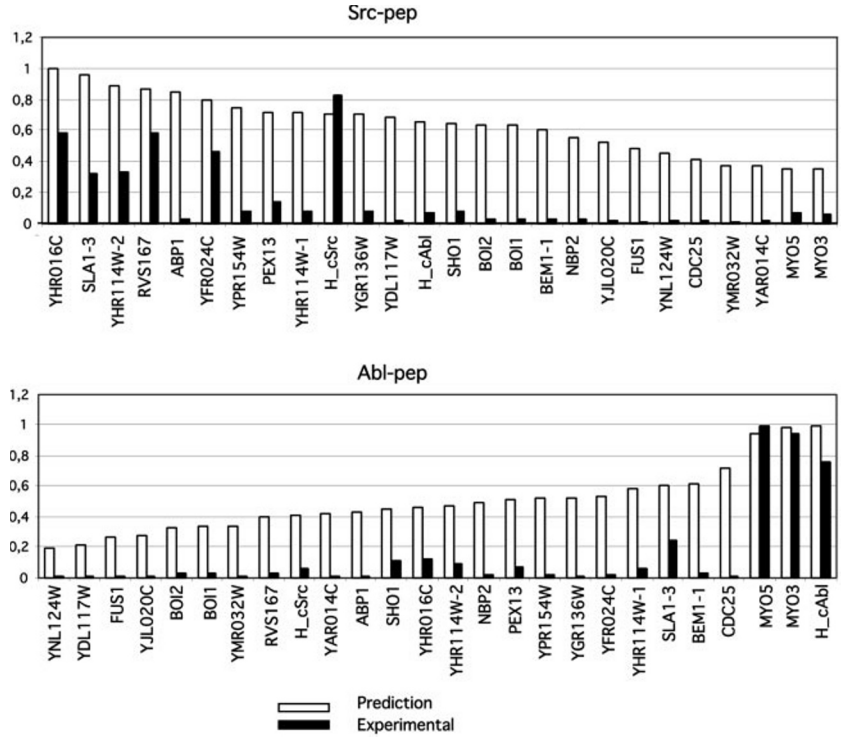


FIG. 6. **The SH3 profile method.** Two different profiles were constructed for the peptides APTYPPPLPP (Abl-pep) and LSSRPLPTLPSP (Src-pep), respectively. A profile is a peptide-specific 12×20 matrix that contains, for each of the twelve degenerate positions in the SH3 repertoire, the frequency of occurrence of the twenty amino acids at that specific position in the pool of SH3 domains, which were selected for binding to that peptide. In the case of the Abl-pep probe, we have used the data in Fig. 1*A*. For Src-pep, to increase the statistical significance of the approach, we have added the sequence of six domains that have been selected with the related peptide LSSRPLPTAPSP to the sequences of the Src-pep-specific domains in Fig. 1*B* (not shown). The profiles have been used to evaluate the propensity of each of 24 SH3 domains present in the *S. cerevisiae* proteome to bind either Abl-pep or Src-pep. This was obtained by adding up, for each SH3 degenerate position, the frequency corresponding to the amino acid present at that position in the SH3 domain under scrutiny. The scores, for each domain, are reported in an arbitrary scale (*white bars*). The *filled bars* represent the optical density (OD) values obtained in an ELISA carried out as described in Fig. 4.

described in this report). Similar, although more limited repertoires, were used to select PDZ (9), SH3 (28), and WW (10) domains that would bind more tightly to specific peptide ligands or to new peptide ligand sequences. Our repertoire, by contrast, was designed with the aim of representing the entire binding potential of natural SH3 domains. This goal was obtained by building a degenerate Abl-SH3 gene that would encode in the 12 positions involved in target recognition most of the residues that are found, at the corresponding location, in natural SH3 domains. Thus most residue combinations that would result in unfolded molecules or non-functional binding surfaces are not contained in the repertoire. Furthermore, our synthetic repertoire of 10^7 different molecules represents most of the sequence and structure space clustered around the ensemble of natural SH3 domains. As a consequence, most of the SH3 binding surfaces that have been explored by natural evolution are likely to be represented. In fact, by panning with a variety of peptides containing the P*XX*P SH3 binding motif, we have always been able to find ligands (unpublished experiments). It would be interesting to ask whether such a repertoire also contains elements that would bind to less typical SH3 targets as recently described for several natural SH3 domains (29–31).

Evolutionary Pathways—In vitro screening and selection of peptide repertoires is a powerful tool for answering fundamental questions regarding the evolution of protein properties (32–34). To be able to identify potential evolutionary pathways of recognition specificity, we have chosen two distantly related poly-proline peptides that bind SH3 domains of the Abl and Src families with very different binding constants. By screening the SH3 repertoire with the two distinct peptide ligands, we have characterized SH3 domains representing possible intermediate steps in short evolutionary pathways that, with as few as two or three amino acid substitutions on the surface of the domain can switch the preference of the SH3 domain between the ligands (Fig. 5). These results provide a clear and well-defined example of how different specificities can evolve in protein interaction.

In one scenario (Fig. 5*A*, *pathway 1*), we envision that the specificity of the TNW SH3 domain, which efficiently binds to Abl-pep but not to Src-pep, can be first mutated to TCW without substantially changing its recognition properties. A second substitution, TCL, then causes a sudden shift in ligand preference, which is refined by a third amino acid substitution (DCL). In a second scenario (Fig. 5*A*, *pathway 3*), the shift from Abl-pep to Src-pep binding is first mediated by a mutation (TNV) that leads to a non-functional domain; the TNV domain weakly binds to both peptides and fails to select peptides from our combinatorial peptide libraries (data not shown). Finally, in a third scenario, *pathway 2* can be considered an example of co-evolution of the two binding partners, because each domain along the pathway has a different preferred peptide and one can conceive that the selection of SH3 domain variants is driven, at each step, by corresponding changes of the poly-proline peptide partner or *vice versa* (Fig. 5*B*). Thus, the evolution of different SH3 domain specificities can be driven by very simple evolutionary pathways involving only a few amino acid changes in the SH3 domain. We suggest that this property characterizes small domains involved in protein recognition and that it represents one of the main reasons for their success during evolution and for fixation in the proteome.

Binding Specificity—By selecting the preferred ligands of many SH3-Abl variants, we have determined some "soft" rules that hint to the existence of a recognition code, which, however,

is highly degenerate and context-dependent. For instance, Asp at position 79 of the SH3 domain selects for peptides that display Arg at P-3, but Src-pep with an Arg at P-3 binds to SH3 domains irrespective of whether they have Asp or Glu at that position. On the other hand, Abl-pep with a Tyr at P-3 tolerates both Thr and Glu (but not Asp) at position 79. One highly conserved residue in the SH3 domains that is selected from the repertoire for binding to Abl-pep is the Trp at 110, which is involved, together with Trp-99, in the formation of a hydrophobic pocket that hosts the Pro at P-5 in the target peptide.

Also striking is the preference for Cys at position 94 in SH3 domains that bind to Src-pep. By selecting the preferred ligands of SH3-Abl variants that contain Cys-94, we have observed a correlation between the presence of this residue in the SH3 domain and Leu at P-1 in the ligand peptide. We have demonstrated that Cys-94 forms a disulfide bridge with Cys-100 in the Abl SH3 scaffold (not shown), most likely resulting in a distortion of the n-Src loop and possibly in the formation of a larger, and hydrophobic, cavity that can host the Leu at P-1 in the peptide. The correlation between Cys-94 in the SH3 domain and Leu at P-1 in the peptide is scaffold-dependent and cannot be used to infer the recognition specificity of a different SH3 domain.

Some other preferences have been detected by our experiments: For instance most SH3 domains that bind to Src-pep have an Asn at position 114, whereas Abl-pep enriches for a Thr at that position. Similarly, Phe or Tyr at position 72 are preferentially found in domains that bind to Abl-pep and Src-pep, respectively.

Inferring Recognition Specificity—Several reports have convincingly shown that the consensus ligand obtained by panning peptide repertoires with protein binding modules have a high predictive value when used, as templates in computer searches, to identify the natural partners of the domains (for a review see Ref. 19). Our approach addresses a complementary question and permits one to identify the characteristics of the domain binding surface for any given target peptide (in this specific case a poly-proline peptide). In the work reported here, we have asked whether the "binding information" obtained by screening artificial domain repertoires could be used to identify, in a proteome, those SH3 domains that are likely to bind to a specific poly-proline peptide. Because many protein interactions are mediated by small protein recognition modules, the development of a reliable predictive algorithm would permit one to infer a large fraction of the interaction network with a limited set of experiments. A difficulty in this approach may arise from the implicit assumption that the chemical characteristic of the binding surfaces are largely scaffold-independent and that the information extracted from Fig. 2 can be compared with any SH3 domain, whatever the characteristics of its scaffold.

Despite the correlations uncovered by our experiments, however, it has not been possible to find a simple set of rules that would permit one to establish whether any given SH3 domain in our repertoire would bind to Abl-pep or Src-pep. Nevertheless, we asked whether a position-specific scoring matrix, based on the amino acid frequencies observed in the domains selected from the SH3 artificial repertoire, could be used to rank natural domains according to the probability that they would bind to the poly-proline peptide used in the selection experiment. We have named this scoring matrix a "peptide-specific profile." This is a rather naive approach, because it neglects the contribution to binding of the residues that do not make direct contact with the target peptide but still may have an influence on binding affinity. We have tested this simple approach by asking which of the 29 SH3 domains in the yeast proteome have the potential to bind Src-pep and Abl-pep, and we have compared the prediction with the experimental results obtained by ELISA. The comparison shows that the Abl-pep profile performs satisfactorily, because the two yeast SH3 domains (Myo5-SH3 and Myo5-SH3), which are experimentally found to bind to Abl-pep, obtained the highest score. Also the prediction obtained by the Src-pep profile successfully identified the yeast SH3 domains (Rvs167, Yhr016, Yfr024, Yhr114_2 and Sla1_3) that bind to Src-pep. However, the human Src-SH3 domain itself obtains a score that is worse than the yeast Abp1 domain, which does not bind to Src-pep in ELISA. False negatives, like Src-SH3, could represent a "solution" to the binding problem that is different from the one that is prominent in our artificial repertoire. As mentioned above, both false negative and false positives could arise, because the underlying scaffold has a stronger influence than assumed in our approach. We have recently solved the three-dimensional structure of the Abp1 SH3 domain and, by site-directed mutagenesis, we have been able to prove that the Glu residue at position 69 has a strong influence on ligand preference, because an Abp1 SH3 mutant having a Leu at that position, differently from wild type, binds efficiently to Src-pep (35). Position 69 was not randomized in our repertoire, and, as a consequence, it does not contribute to the profile score. Because this residue does not make contact with the ligand, its influence on recognition specificity represents a clear example of a context effect in ligand binding.

The influence of the scaffold on target binding is particularly evident in the preference for Cys at position 94 in the Src-pep selection. Cys-94 forms a disulfide bridge with a second Cys that is present in the Abl scaffold. Because the formation of the disulfide bridge is essential for binding to Src-pep (not shown), it is unlikely that the presence of a Cys at that position in most other natural scaffolds, which do not have a Cys corresponding to Cys-100, would favor binding to Src-pep. Another difficulty arises from the uncertainty in aligning loops of different length. We suspect that our prediction for Src-pep binding peptides would have performed better if we had designed our peptide repertoire with an n-Src loop one amino acid shorter, as observed in most SH3 domains of the Src kinase family. Despite all these limitations, we have been able to show that position-specific scoring matrices, calculated from the frequency of occurrence of residues at 12 positions of the binding surface of a combinatorial repertoires, are rather powerful tools to infer the binding preference of SH3 domains.

This Repertoire Can Be Used as a Source of Perturbagens—The 10^7 structures that form our SH3 domain repertoire represent a unique source of molecules that can be used to search for affinity reagents to any target protein. Several authors have assembled combinatorial peptide repertoires displayed on different scaffolds and showed that these repertoires can be used to select ligands (perturbagens) that, by binding to intracellular protein targets, interfere with physiological pathways (36–39). This is a powerful approach to dissect cellular pathways. It is unlikely, however, that repertoires of this size and structure will be comprehensive and represent a source of ligands for any protein. An approach that is alternative to a single large and general repertoire would consist in the assembly of a collection of specialized repertoires each aimed at a more restricted region of the structural space. For instance, a repertoire that aims at peptides with a free carboxyl terminus (based on the PDZ scaffold) (9) or a second one that aims at peptides phosphorylated in Tyr (based on an SH2 scaffold) (40). We have shown that our SH3 repertoire contains elements that can bind to a variety of poly-proline peptides. The SH3 domain repertoire is directed to the many proteins (~50% of the yeast proteins) that display on their surface peptides that contain the

P*XX*P signature of an SH3 ligand. Because often these peptides are used in signal transduction pathways to assemble signaling complexes via interaction with WW or SH3 domains, it is anticipated that a sizeable number of ligands targeted to these structures will disturb functional pathways.

Acknowledgment—We thank B. Kay for comments and help in improving the manuscript.

REFERENCES

1. Pawson, T., and Scott, J. D. (1997) *Science* **278,** 2075–2080
2. Ren, R., Mayer, B. J., Cicchetti, P., and Baltimore, D. (1993) *Science* **259,** 1157–1161
3. Moran, M. F., Koch, C. A., Anderson, D., Ellis, C., England, L., Martin, G. S., and Pawson, T. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 8622–8626
4. Fanning, A. S., and Anderson, J. M. (1999) *Curr. Opin. Cell Biol.* **11,** 432–439
5. Sudol, M. (1998) *Oncogene* **17,** 1469–1474
6. Songyang, Z., Gish, G., Mbamalu, G., Pawson, T., and Cantley, L. C. (1995) *J. Biol. Chem.* **270,** 26029–26032
7. Lee, C. H., Leung, B., Lemmon, M. A., Zheng, J., Cowburn, D., Kuriyan, J., and Saksela, K. (1995) *EMBO J.* **14,** 5006–5015
8. Stricker, N. L., Christopherson, K. S., Yi, B. A., Schatz, P. J., Raab, R. W., Dawes, G., Bassett, D. E., Jr., Bredt, D. S., and Li, M. (1997) *Nat. Biotechnol.* **15,** 336–342
9. Schneider, S., Buchert, M., Georgiev, O., Catimel, B., Halford, M., Stacker, S. A., Baechi, T., Moelling, K., and Hovens, C. M. (1999) *Nat. Biotechnol.* **17,** 170–175
10. Kasanov, J., Pirozzi, G., Uveges, A. J., and Kay, B. K. (2001) *Chem. Biol.* **8,** 231–241
11. Ponting, C. P., Schultz, J., Milpetz, F., and Bork, P. (1999) *Nucleic Acids Res.* **27,** 229–232
12. Kay, B. K., Williamson, M. P., and Sudol, M. (2000) *FASEB J.* **14,** 231–241
13. Mayer, B. J. (2001) *J. Cell Sci.* **114,** 1253–1263
14. Musacchio, A. (2002) in *How SH3 Domains Recognize Proline*. *Advances in Protein Chemistry* (Janin, J., and Wodak, S., eds) p. 61, Academic Press, New York
15. Lim, W. A., Richards, F. M., and Fox, R. O. (1994) *Nature* **372,** 375–379
16. Musacchio, A., Saraste, M., and Wilmanns, M. (1994) *Nat. Struct. Biol.* **1,** 546–551
17. Feng, S., Chen, J. K., Yu, H., Simon, J. A., and Schreiber, S. L. (1994) *Science* **266,** 1241–1247
18. Rickles, R. J., Botfield, M. C., Weng, Z., Taylor, J. A., Green, O. M., Brugge, J. S., and Zoller, M. J. (1994) *EMBO J.* **13,** 5598–5604
19. Kay, B. K., Kasanov, J., Knight, S., and Kurakin, A. (2000) *FEBS Lett.* **480,** 55–62
20. Santi, E., Capone, S., Mennuni, C., Lahm, A., Tramontano, A., Luzzago, A., and Nicosia, A. (2000) *J. Mol. Biol.* **296,** 497–508
21. Castagnoli, L., Zucconi, A., Quondam, M., Rossi, M., Vaccaro, P., Panni, S., Paoluzi, S., Santonico, E., Dente, L., and Cesareni, G. (2001) *Comb. Chem. High Throughput Screen* **4,** 121–133
22. Tong, A. H., Drees, B., Nardelli, G., Bader, G. D., Brannetti, B., Castagnoli, L., Evangelista, M., Ferracuti, S., Nelson, B., Paoluzi, S., Quondam, M., Zucconi, A., Hogue, C. W., Fields, S., Boone, C., and Cesareni, G. (2002) *Science* **295,** 321–324
23. Pisabarro, M. T., Serrano, L., and Wilmanns, M. (1998) *J. Mol. Biol.* **281,** 513–521
24. Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L., and Sonnhammer, E. L. (2000) *Nucleic Acids Res.* **28,** 263–266
25. Smith, G. P. (1985) *Science* **228,** 1315–1317
26. Zucconi, A., Dente, L., Santonico, E., Castagnoli, L., and Cesareni, G. (2001) *J. Mol. Biol.* **307,** 1329–1339
27. Pisabarro, M. T., and Serrano, L. (1996) *Biochemistry* **35,** 10634–10640
28. Sparks, A. B., Adey, N. B., Quilliam, L. A., Thorn, J. M., and Kay, B. K. (1995) *Methods Enzymol.* **255,** 498–509
29. Mongiovì, A. M., Romano, P. R., Panni, S., Mendoza, M., Wong, W. T., Musacchio, A., Cesareni, G., and Paolo Di Fiore, P. (1999) *EMBO J.* **18**, 5300–5309
30. Kang, H., Freund, C., Duke-Cohan, J. S., Musacchio, A., Wagner, G., and Rudd, C. E. (2000) *EMBO J.* **19,** 2889–2899
31. Kato, M., Miyazawa, K., and Kitamura, N. (2000) *J. Biol. Chem.* **275,** 37481–37487
32. Hoffmuller, U., Knaute, T., Hahn, M., Hohne, W., Schneider-Mergener, J., and Kramer, A. (2000) *EMBO J.* **19,** 4866–4874
33. Altamirano, M. M., Blackburn, J. M., Aguayo, C., and Fersht, A. R. (2000) *Nature* **403,** 617–622
34. Keefe, A. D., and Szostak, J. W. (2001) *Nature* **410,** 715–718
35. Fazi, B., Cope, M. J., Douangamath, A., Ferracuti, S., Schirwitz, K., Zucconi, A., Drubin, D. G., Wilmanns, M., Cesareni, G., and Castagnoli, L. (2002) *J. Biol. Chem.* **277,** 5290–5298
36. Geyer, C. R., Colman-Lerner, A., and Brent, R. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 8567–8572
37. Blum, J. H., Dove, S. L., Hochschild, A., and Mekalanos, J. J. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 2241–2246
38. Xu, X., Leo, C., Jang, Y., Chan, E., Padilla, D., Huang, B. C., Lin, T., Gururaja, T., Hitoshi, Y., Lorens, J. B., Anderson, D. C., Sikic, B., Luo, Y., Payan, D. G., and Nolan, G. P. (2001) *Nat. Genet.* **27,** 23–29
39. Norman, T. C., Smith, D. L., Sorger, P. K., Drees, B. L., O'Rourke, S. M., Hughes, T. R., Roberts, C. J., Friend, S. H., Fields, S., and Murray, A. W. (1999) *Science* **285,** 591–595
40. Malabarba, M. G., Milia, E., Faretta, M., Zamponi, R., Pelicci, P. G., and Di Fiore, P. P. (2001) *Oncogene* **20,** 5186–5194