Chemistry & Biology, Vol. 9, 253–264, February, 2002, ©2002 Elsevier Science Ltd. All rights reserved. PII S1074-5521(02)00098-4

Selection of v-Abl Tyrosine Kinase Substrate Sequences from Randomized Peptide and Cellular Proteomic Libraries Using mRNA Display

Thomas P. Cujec,[1] Patricia F. Medeiros, Phil Hammond, Cecil Rise, and Brent L. Kreider
Phylos
128 Spring Street
Lexington, Massachusetts, 02421

Summary

Methodologies for rapidly identifying cellular protein interactions resulting in posttranslational modification of one of the partners are lacking. Here, we select for substrates of the v-abl tyrosine kinase from two protein display libraries in which the protein is covalently linked to its encoding mRNA. Successive selection cycles from a randomized peptide library identified a consensus sequence closely matching that previously reported for the v-abl tyrosine kinase. Selections from a proteomic library derived from cellular mRNA identified several novel targets of v-abl, including a new member of a class of SH2 domain-containing adaptor proteins. Upon modification, several of the substrates obtained in these selections were found to be effective inhibitors of v-abl kinase activity in vitro. These experiments establish a novel method for identifying the substrates of tyrosine kinases from synthetic and cellular protein libraries.

Introduction

Phosphorylation of proteins on their serine, threonine, and tyrosine residues is one of the most commonly occurring posttranslational modifications in eukaryotic cells. Cellular phosphorylation cascades allow for the amplification of extracellular signals following changes in environmental conditions via the ability of phosphorylated activators to modulate the expression of numerous genes. Because these reactions are rapidly reversible, they are important for the regulation of many cellular functions including signal transduction, cell division, and proliferation. The specificity of these interactions is determined by a number of parameters, including cellular colocalization of the enzyme and its substrate, secondary interactions between the enzyme and the substrate, as well as the substrate specificity of the kinase catalytic domain [1–3]. Identification of kinase substrate sequences is a critical step toward understanding the biology of these kinases and provides an important venue for the development of kinase inhibitory reagents.

Numerous in vivo and in vitro protein display technologies have developed in an attempt to decipher the myriad of biological pathways that occur in a cell [4, 5]. Strategies involving in vivo display of protein domains generally involve making proteomic libraries derived from cellular mRNA and then selecting for desired interactions either within the organism itself (yeast two-hybrid and its derivatives) [6–8] or following cell surface expression (phage and *Eschericia coli* display) [9–12]. Although these systems are capable of detecting protein-protein interactions, their ability to identify proteins targeted for posttranslational modification has not been demonstrated. Numerous in vitro protein display systems have also been developed [13, 14]; however, in most cases, the absence of a robust linkage between genotype and phenotype limits their applicability in studying enzymatic modifications. Although methodologies for the identification of kinase substrates has generally lagged behind those aimed at elucidating protein-protein interactions, several approaches have been employed. These include phosphorylation of proteins immobilized on a nitrocellulose filter, following induction of a phage expression library [15]; mutation of the kinase's ATP binding site, allowing it to accept N^6-modified ATP analogs that can be used as substrate tags [16]; and the selection of phosphorylated targets from synthetic peptide libraries [17]. In this latter approach, a random peptide library is phosphorylated in vitro by the kinase of interest, the substrates are recovered on a metal-chelating column, and the mixture of proteins is sequenced.

mRNA-protein fusion libraries (mRNA display), in which proteins are linked to their encoding mRNA via the *O*-methyl tyrosine portion of puromycin, offer a powerful approach to screening large protein libraries. In this system, rabbit reticulocyte lysates are used to produce proteins in vitro and covalently attach them to their encoding mRNA. Because the link between the protein and its genotype is covalent, the protein moiety can be selected under robust conditions and its genetic material can be amplified by PCR in an entirely in vitro system [18, 19]. Repeated rounds of selection from a library of mRNA-protein fusion molecules allows for the rapid enrichment of proteins having a desired phenotype. Since library synthesis and selection occurs entirely in vitro, it is possible to minimize the loss of genetic information normally associated with the subcloning and transformation steps of in vivo protein display systems (see [20] for a review). Recently, "mRNA display" has been used to select for high-affinity streptavidin binding aptamers ($<$5 nM) from a nonconstrained peptide library [21]. The in vitro nature of the mRNA-protein selection technique makes it ideally suited for the identification of protein target sequences undergoing specific posttranslational modifications and allows a variety of different selection conditions to be examined. Here, we describe the identification of optimal target sequences for the v-abl tyrosine kinase from a random peptide library as well as from a proteomic library derived from the mRNA of human bone marrow cells.

The nonreceptor tyrosine kinase v-abl is encoded by the Abelson murine leukemia virus and is a potent oncogene in mice. The closely related human proto-oncogene c-abl contains the src-homology (SH) regions 1, 2, and 3, which confer phosphorylation activity (SH1), ability to bind phosphorylated tyrosines (SH2), and kinase inhibitory functions (SH3) to the protein [22, 23]. A

[1]Correspondence: tcujec@phylos.com

reciprocal translocation between *c-abl* (chromosome 9) and the *bcr* gene (chromosome 22) results in the potent oncogene BCR-abl, which is implicated in the pathogenesis of 95% of chronic myelogenous leukemias and 10% of acute lymphocytic leukemias [24, 25]. Although numerous putative substrates of the abl kinases have been identified, including those involved in cell architecture, transcription, and numerous signal transduction pathways, the mechanism of action of these kinases remains unclear (see [23, 26, 27] for compilation). A minimal v-abl consensus sequence (I/V/L-Y-X_n-P/F, where n = 2 or 3) has been previously identified from a peptide library randomized at eight amino acids, but the effects of larger sequence motifs on substrate selection is unknown [28].

As a step toward understanding the abl kinase response pathway, we applied a selection-based technology to identify optimal substrate sequences of the potent oncogene v-abl from a large randomized peptide library ($\sim 10^{12}$–10^{13}). In addition, we identified cellular targets of v-abl from a proteomic mRNA-protein library derived from the mRNA of human bone marrow cells. The approach we describe here offers a powerful new strategy for identifying novel substrates of tyrosine kinases and has led to the rapid identification of lead candidates for kinase inhibitors.

Results

Model System

Substrate sequences surrounding a phosphorylated residue are important in determining protein kinase specificity. Consequently, in order to progress toward a greater understanding of the biology of v-abl, we sought to identify kinase substrates from a randomized peptide library and a cellular proteomic library displayed as mRNA-protein fusions. In a proof of concept experiment, a control mRNA-protein fusion molecule containing an optimized v-abl phosphorylation site (EAIY AAPFAKKK, New England Biolabs) was synthesized. Fusions were purified from the reverse transcription reaction by oligo d(T) chromatography, phosphorylated with recombinant v-abl, and immunoprecipitated with the α4G10 antibodies. Approximately 3% of the control mRNA-peptide fusion molecules were immunoprecipitated by the α4G10 antibody following phosphorylation by v-abl (data not shown). Immunoprecipitation was specific for the phosphotyrosine antibody and absolutely dependent upon recombinant v-abl. Together these results demonstrated that v-abl is functional at substrate concentrations (50–100 nM) significantly below the K_m of the enzyme (50 μM) and that the translational lysates do not contain significant endogenous v-abl activity. We reasoned, therefore, that repeated rounds of phosphorylation of a mRNA-protein fusion library by v-abl, followed by selection of tyrosine-phosphorylated fusions with the α4G10 antibody, could be used to identify substrate targets of the v-abl oncogene.

Enrichment of mRNA-Peptide Fusions Containing Phosphorylated Tyrosines

To identify peptide substrates of the v-abl tyrosine kinase, we synthesized an mRNA-peptide fusion library in which an invariant tyrosine was flanked by five randomized amino acids (GCGGX_5-Y-X_5GCG). In addition, the fixed flanking regions contained two cysteine residues to provide the opportunity for the random region to be presented as a constrained loop. The mRNA-peptide fusions were phosphorylated by v-abl in vitro, immunoprecipitated using the phosphotyrosine-specific α4G10 antibody and the attached genetic information amplified by PCR (Figure 1).

We anticipated that the starting mRNA-peptide library (25 pmol) should contain approximately one copy of each possible sequence (20^{10} members). In order to minimize the loss of genetic diversity, the starting library was not precleared in the first round of the selection, and approximately 1.5% of the input mRNA-peptide fusion molecules were eluted (E1 + E2) from the α4G10 antibody in round one (Figure 2A). In subsequent rounds, the library was precleared by incubating it with α4G10/protein A Sepharose beads prior to phosphorylation with v-abl in order to remove nonspecific binders and any protein domains having tyrosines fortuitously phosphorylated during the in vitro translation reaction. In rounds two and three, the elutions did not yield detectable radioactive counts, although PCR yielded genetic material that was taken to the next round. In contrast, approximately 6.5% of the input mRNA-peptide fusion molecules were eluted (E1 + E2) from the α4G10 antibody in round four, and this increased to almost 20% in round six. PCR analysis demonstrated that the last wash of each round was essentially devoid of genetic material (data not shown).

In order to confirm the enrichment of v-abl substrates, the DNAs from the starting library and the post-round six pool were expressed as free peptides and tested for phosphorylation by v-abl (Figure 2B). No phosphorylation of the starting peptide pool was detected, presumably due to the low abundance of v-abl target sequences (lane 3). In contrast, phosphorylation of the post-round six peptide pool was readily observed (lane 4), demonstrating the enrichment of target sequences. A peptide containing a single tyrosine residue in the FLAG epitope (DYKDDDDK) was not phosphorylated by v-abl (lane 2), demonstrating that phosphorylation of the test peptides was sequence specific, while phosphorylation of a control fusion molecule containing a known v-abl recognition motif was readily detected (lane 1). Together these results demonstrate that substrates of v-abl were enriched from the mRNA-peptide library following six rounds of selection and PCR amplification of genetic material.

Sequence Analysis of the Selected Pool

In total, 69 clones from the post-round six pool were sequenced. The invariant tyrosine was missing in 2 of the 69 clones. In one of these cases, a novel tyrosine was introduced upstream of the invariant residue, while, in the other case, there were no tyrosine residues. Of the remaining 67 cones, 46 matched the I/L/V-Y-$X_{1\text{-}5}$-P/F consensus motif with respect to the invariant tyrosine (Figure 3A). The majority of these clones had a proline or phenylalanine residue four (22/46) or five (23/46) amino acids downstream of the invariant tyrosine.

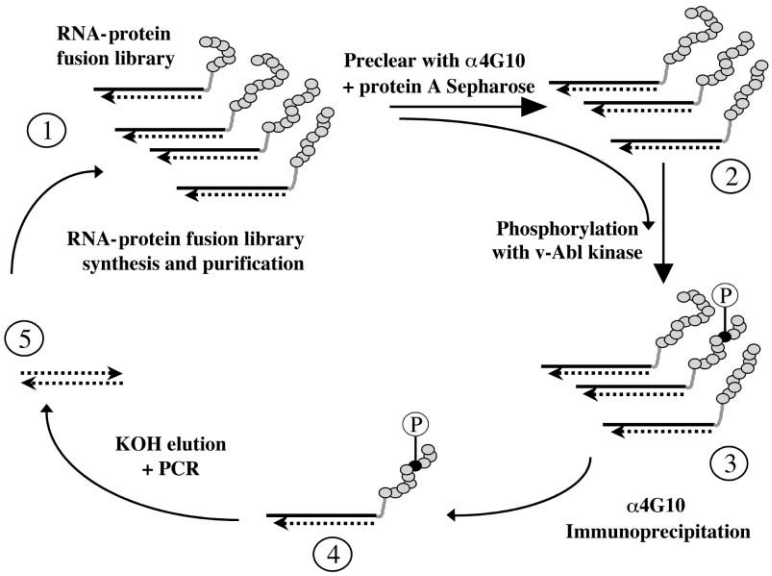


Figure 1. A Schematic Drawing of the Selection Procedure

In the first round of the selection, purified mRNA-protein fusion molecules were phosphorylated by v-abl and subjected to immunoprecipitation with the α4G10 antibody. The mRNA-protein fusion molecules were eluted from the antibody using KOH and the cDNAs amplified by PCR. In subsequent rounds, the mRNA-protein pool was precleared with the α4G10/protein A Sepharose beads prior to the kinase reaction to remove any nonspecific binders, as well as mRNA-protein fusions undergoing nonspecific tyrosine phosphorylation during in vitro translation.

In most clones (66/68), a second tyrosine was introduced into the random region either upstream (62) or downstream (4) of the fixed residue (Figure 3B). Of the 21 clones in which the invariant tyrosine was not part of a consensus sequence, 11 contained a second tyrosine that did match the I/L/V-Y-X_{1-5}-P/F consensus. In the majority of clones, an upstream tyrosine was introduced either three (13/62) or four (45/62) residues upstream of the fixed tyrosine. Only 3/13 clones containing the tyrosine (y) at position n = −3 relative to the fixed residue (Y) (y-X-I/L/V-Y) conformed to the consensus phosphorylation sequence. This result is not surprising given the requirement for a hydrophobic residue upstream of the invariant tyrosine. In contrast, if the tyrosine (y) was present at n = −4, almost half the clones matched the consensus sequence (23/45). All of the clones containing an additional tyrosine immediately upstream of the fixed residue (n = −1) or at positions n = −2 or n = −5 matched the v-abl kinase consensus motif with respect to either the novel or invariant tyrosine. Only 11 clones were either missing a tyrosine or contained one that did not match an obvious phosphorylation consensus motif (Figure 3C).

In summary, 58/69 clones contained a tyrosine within the I/L/V-Y-X_{1-5}-P/F consensus motif. The v-abl kinase phosphorylated all of the clones (12/12) representing the various classes of consensus motifs that were tested (data not shown). These results demonstrate the enrichment of v-abl substrates from the randomized peptide library and suggest that v-abl can potentially phosphorylate a wider range of peptide substrates than previously thought.

Selection of Cellular Proteins Phosphorylated by v-abl

Peptide substrates of v-abl were successfully selected from a randomized mRNA-peptide fusion library following six rounds of selection. However, our results demonstrated that v-abl can phosphorylate a broad consensus sequence, making it difficult to identify potential cellular phosphorylation targets. In order to identify cellular targets of v-abl, we screened a proteomic mRNA-protein library derived from human bone marrow cells. A cellular mRNA-protein fusion library was constructed by randomly priming cellular mRNA to create a mixture of full-length and partial cDNA fragments. A random priming approach was used in order to minimize the presence of translational stop codons that prevent mRNA-protein fusion formation. Following in vitro translation and fusion, a proteomic library was generated in which most cellular proteins were expected to be represented in sizes ranging from epitope size to full length. Cellular targets of v-abl were selected from this library following phosphorylation by v-abl and immunoprecipitation with the α4G10 antibody. Repeated rounds of selection produced an enriched pool of v-abl cellular substrates.

Enrichment of mRNA-Protein Fusions Containing Phosphorylated Tyrosines

Approximately 1 pmol of the starting mRNA-protein library was used in the first round of the proteomic selection. In order to minimize the loss of genetic diversity, the preclear step was again omitted in the first round of the selection. The first two rounds of the selection did not yield detectable radioactive counts in the elutions; although, PCR yielded genetic material that was taken to the next round (data not shown). In contrast, approximately 2% of the input mRNA-protein fusion molecules were eluted (E1 + E2) from the α4G10 antibody in round three, and this increased to 6% in rounds four and five. The percentage of input mRNA-protein fusion molecules eluted in rounds four and five was significantly greater than that in wash five (W5) and was similar to that recovered using the optimal v-abl target (see above). In addition, PCR analysis demonstrated that the last wash of round five (W5) was devoid of mRNA-protein fusion molecules (data not shown). Together these results demonstrate that substrates of v-abl were enriched from the cellular mRNA-protein library following five rounds of selection and PCR amplification of genetic material.

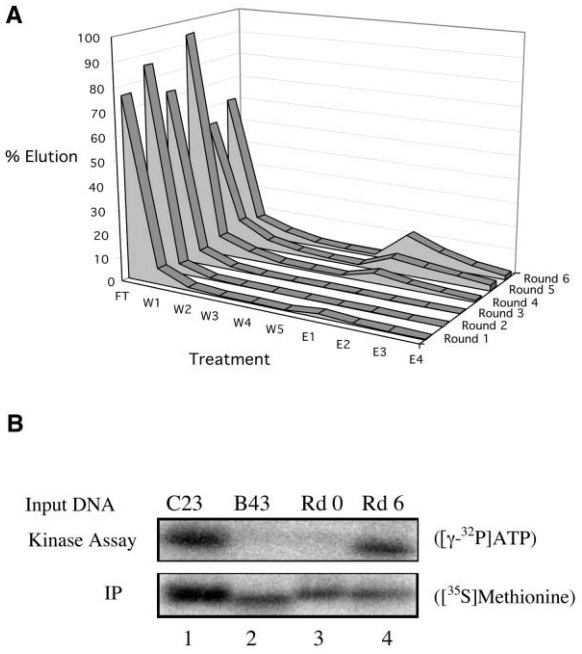


Figure 2. Enrichment of v-abl Substrates from a Randomized Peptide Library

(A) Elution profiles of a randomized peptide library displayed as mRNA-peptide fusions following six rounds of v-abl phosphorylation and α4G10 immunoprecipitation. The percent elution was based on the input of mRNA-peptide fusion molecules labeled with [^{35}S]methionine. In round one, the library was not precleared prior to phosphorylation with v-abl.

(B) Enrichment of v-abl substrates from a randomized peptide library following six rounds of selection. DNA from the starting library (lane 3) and the post-round six elution (lane 4) was transcribed by T7 polymerase, translated in the absence of radiolabeled methionine, and immunoprecipitated with αFLAG protein A Sepharose beads. In the kinase reactions, immunoprecipitates were phosphorylated with v-abl in the presence of [γ-^{32}P]ATP, washed extensively, and then resolved on a 4%–20% Tris-glycine gel. As a control for immunoprecipitation, parallel in vitro translation reactions were done in the presence of [^{35}S]methionine. The B43 clone is a negative control for the kinase reaction (lane 2) and contains a single tyrosine residue in the FLAG epitope (DYKDDDDK), while the C23 clone serves as the positive control (lane 1) and contains the known v-abl phosphorylation sequence (IYAAP).

Analysis of Selected Pools

Since the starting mRNA-protein fusion library consists of the full repertoire of proteins expressed in human bone marrow cells, the enrichment of desired clones can be rapidly ascertained by hybridization to human cDNA microarrays. As a first step in the analysis of the selected pools, DNA from rounds four and five of the selection were hybridized to cDNA microarrays containing >18,000 EST clones, and the hybridization intensities were compared to those obtained with the starting (round zero) library (Figure 4A). The filter contained four known v-abl substrates, namely, Crk1 [29, 30], the regulatory subunit (p85) of phosphoinositol-3-kinase [31], the Arg/Abl-interacting protein ArgBP2 [32], and the JAK1 tyrosine kinase [33, 34], as well as the large subunit of RNA polymerase II, a previously reported target of c-abl [35]. All of these targets were enriched during the selection procedure (Figure 4B). Numerous other ESTs containing potential v-abl phosphorylation sites were also detected in the hybridization analysis. In most cases, these correspond to proteins known to interact with v-abl, or proteins that are involved in cellular processes affected by v-abl [23, 26, 27].

The potential v-abl phosphorylation sites identified in the filter analysis were very similar to the v-abl consensus sequence identified from the selection of a randomized peptide library. Nevertheless, to examine the nature of the enriched pools more precisely, the pools from rounds four and five of the selection were subcloned, and the resulting clones were sequenced. Sequence analysis of 102 clones indicated that the v-abl target clones could be divided into the following three major groups: clones fitting the kinase consensus sequence I/L/V-Y-X$_{1-5}$-P/F (64/102), those lacking a downstream proline/phenylalanine residue (14/102), and those lacking an upstream hydrophobic residue (14/102). Only a small fraction of the clones (10/102) had more than one difference from the consensus sequence or no tyrosine. To determine whether members of the three classes could serve as v-abl substrates, representative clones were expressed individually as free peptides and tested for v-abl phosphorylation in vitro (data not shown). As summarized in Figure 4C, clones representing all three groups were phosphorylated. Only 3 of the 13 clones tested in the individual kinase assays were not phosphorylated by v-abl. These sequences closely matched the v-abl consensus sequence, and failure to detect phosphorylation might be due to the close proximity of the tyrosines to the FLAG epitope and the antibody used to immunoprecipitate the free proteins. Together these results demonstrate that cellular substrates of v-abl can be selected from a cellular mRNA-protein fusion library and that the repertoire of potential v-abl targets is much broader than indicated by previous studies.

Homology searches using the sequenced clones identified 13 known genes and 3 predicted ORFs. One of the known genes was identified as actin, previously published as a v-abl-interacting protein [36, 37]. The translation elongation factor eIF-4AIII was found twice, and the selected clone is phosphorylated by v-abl in vitro. Nebulin was found 31 times, and cytochrome C oxidase subunit 3 was found 25 times. A total of 33 clones showed no homology to any genes in the GenBank database. Efforts are underway to characterize a subset of these targets and establishe their significance in v-abl function.

Many of the nebulin and cytochrome C oxidase clones contained different 5′ or 3′ termini, indicating that the clones arose from distinct priming events during construction of the library. This suggests that, as expected, overlapping protein domains are present in the mRNA-protein fusion library. Nebulin is closely associated with actin and may represent a novel substrate for v-abl given its high representation in the selected pools and its phosphorylation by v-abl in vitro. In contrast, the segment of cytochrome C oxidase selected in this study is embedded inside the core of the full-length protein and may not be accessible to v-abl in vivo.

Shg Is a Novel SH2 Domain-Containing Adaptor Protein

Sequence analysis indicated that one of the fragments (TC26) obtained in round five of the selection (Rd5-26,

A

In Reference to Fixed Tyrosine: GCGG-xxxxx**Y**xxxxx-GCG [46/67]
I/V/L-**Y**-x-P/F [7/46]
I/V/L-**Y**-xx-P/F [5/46]
I/V/L-**Y**-xxx-P/F [22/46]
I/V/L-**Y**-xxxx-P/F [23/46]

B

Upstream Tyrosines: GCGG-$\mathbf{Y}_{1-5}$-*Y*-xxxxx-GCG [62/68]		Downstream Tyrosines: GCGG-xxxxx-*Y*-$\mathbf{Y}_{1-5}$-GCG [4/68]	
N=1: xxxx-**Y**-*Y*-xxxxx [3/62]	*[3/3]*	N=1: xxxxx-*Y*-**Y**-xxxx [1/4]	*[0/4]*
N=2: xxx-**Y**-x-*Y*-xxxxx [2/62]	*[0/2]*	N=2: xxxxx-*Y*-x-**Y**-xxx [2/4]	*[0/4]*
N=3: xx-**Y**-xx-*Y*-xxxxx [13/62]	*[3/13]*	N=3: xxxxx-*Y*-xx-**Y**-xx [1/4]	*[0/4]*
N=4: x-**Y**-xxx-*Y*-xxxxx [45/62]	*[23/45]*	N=4: xxxxx-*Y*-xxx-**Y**-x [0/4]	*[0/4]*
N=5: -**Y**-xxxx-*Y*-xxxxx [5/62]	*[0/5]*	N=5: xxxxx-*Y*-xxxx-**Y**- [1/4]	*[0/4]*

C

Sequences not Conforming to the Consensus:
Containing fixed tyrosine: [10/69]
Sequences lacking tyrosine: [1/69]

Figure 3. Summary of the Sequences Obtained from the Randomized Peptide Selection Following Six Rounds of Selection

(A) Phosphorylation consensus sequences with respect to the fixed tyrosine residue. Amino acids immediately surrounding the randomized sequences are shown. The number of sequences in each category is indicated within the brackets. A total of 11 sequences fall into more than one category.

(B) Phosphorylation consensus sequences with respect to additional tyrosine residues introduced within the randomized region by the selection. The number of clones in each category is indicated directly adjacent to the consensus sequence. The number of sequences within each category in which the novel tyrosine is part of the I/L/V-Y-X_{1-5}-P/F consensus sequence is italicized. Six sequences fall into more than one category.

(C) Sequences either lacking a tyrosine or an obvious v-abl phosphorylation motif.

Figure 4C) had homology to a class of SH2 domain (Src homology 2)-containing adaptor proteins (Figure 5A). Two members of this family, Shc and Shd, are phosphorylated by v-abl in vivo [38, 39]. TC26 contains four potential v-abl phosphorylation sites. Experiments using mutagenized peptides demonstrated that only the first tyrosine is a substrate for v-abl (Figures 5B and 5C). Because this family of proteins appears to be important for abl function, a PCR-based strategy was used to obtain sequence information for the full-length *shg* clone (Figure 5D). The Shg protein has a predicted molecular weight of 54 kDa (495 aa) and is relatively basic (pI = 9.2). Similar to other members of the family, Shg contains an SH2 domain near its carboxyl terminus (393–477). The amino termini of all known family members are highly divergent; however, Shg is unique in that it contains an N-terminal myristylation site, which could serve to anchor this cytosolic protein to the cell membrane. Experiments to confirm the phosphorylation of Shg by v-abl in vivo and to elucidate the biological function of this interesting protein are in progress.

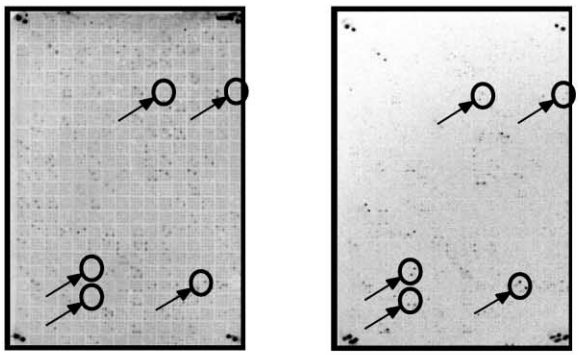
Inhibition of v-abl Kinase Activity

In addition to the information gained by the identification of novel kinase targets, common sequence motifs across selected targets allow for a focused interrogation of potential active site inhibitors. The activity of several serine/threonine kinases has been inhibited using kinase pseudosubstrates in which the target serine has been mutated to an alanine [40–42]. In an analogous manner, the tyrosine kinase activity of ZAP-70 was inhibited by a pseudosubstrate containing a tyrosine-to-phenylalanine substitution mutation [43]. Consequently, we sought to determine whether v-abl substrates selected from the random peptide and proteomic libraries could inhibit kinase activity if the target tyrosine was mutated to phenylalanine. One of the mutagenized peptides (m-M16) inhibited v-abl activity at 50 μM (Figures 6A and 6B), while three additional peptides (mRd6-21, m-B47, and mT13) were effective inhibitors at concentrations of 500 μM. In marked contrast, a phenylalanine substitution in an optimized v-abl target sequence derived from selections of a randomized peptide library (v-abl-mut) [28] was less effective as a kinase inhibitor, further underlying the observation that distal amino acids are important determinants of catalytic specificity. A subset of control peptides (r series) containing the same amino acids as the test peptides, but in a scrambled order, did not have a significant effect on v-abl kinase activity.

Discussion

A randomized peptide library and a cellular protein library derived from the mRNA of human bone marrow cells were displayed as mRNA-protein fusions. Substrates of v-abl were selected from these libraries following successive rounds of enrichment. Selections from the randomized peptide library identified a consensus sequence for v-abl phosphorylation, which closely matched that reported previously. Selections from the proteomic library derived from cellular mRNA identified numerous proteins previously reported to be substrates of v-abl in vivo. In addition, several novel potential targets of v-abl were also selected, including a previously

A



Round 0 Round 4

B

Clone	GenBank Accession #	Potential v-Abl Phosphorylation Sites
Crk1	H75531	PGPYAQPSV
Phosphoinositol-3- kinase, regulatory subunit (p85 subunit)	H63824	EPYNLYGSL
Arg/Abl-interacting protein ArgBP2	BAA34497	Unknown
RNA polymerase II (RPB1)	W23577	SYSPTSPSYSP
Protein-tyrosine kinase JAK1	A39577	Unknown
Retinoblastoma-like 1 (p107)	R21904	IYISP
RNA polymerase II (RPB2)	T89038	LYNGF
SLAP (src-like adaptor protein)	R52378	WYYISP
Grb-2 associated binding protein	N68193	NYVP
MAPKAP kinase	W68281	LYDSP
GAP-associated tyrosine phosphoprotein p62 (SAM68)	R66397	AYKP
Stress-activated protein kinase 10 (JNK3)	R39221	VYKVP
SWI/SNF related actin dependent regulator of chromatin	AAC50694	AYRNF
Filamin A (actin-binding protein-280)	H53989	VYGFEYYPMVP

C

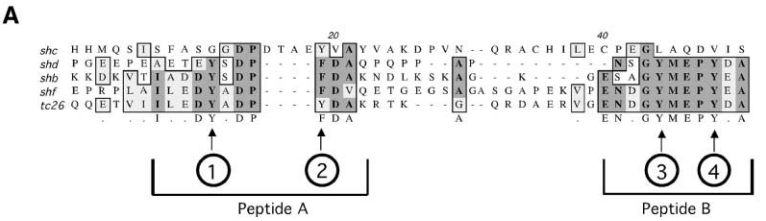
Clone	Group	Name	Insert Sequence	Phosph. by v-Abl
Rd5-13a	2	Actin	GP**IY**E**GYAQP**HAIMRLDLAA	NT
Rd4-M7	1	Nebulin	ADDED**Y**EEVVE**YY**TEEV**VYEEVP**GETITKI	Yes
Rd4-B47	1	Cytochrome C oxidase subunit 3	NSE**Y**FESPFTISDG**IYGSTF**FVATGFHGR	Yes
Rd4-B32	1	KIAA0673	GGET**Y**TIGLQFAPSQRVGEEEIL**IYTP**	Yes
Rd4-T13	1	Unknown	TMTETSFR**VYALYHSLP**TT	Yes
Rd4-M16	1	Unknown	NGGRASGSQGRLEMLPPSA**IYLFF**SFE	Yes
Rd5-22	1	Unknown	RP**IY**SSSK**Y**LLG**LYFVP**GTVLDTGH	Yes
Rd4-T22	1	Unknown	**Y**IL**LYDLGP**	No
Rd4-T29	1	Unknown	DQNEGAFTE**LYFYYTP**	No
Rd5-16	1	Unknown	HAYNFIPGCMRCF**LYMP**H	No
Rd4-T5	2	Myeloid Cell Nucl. Differ. Antigen	**AYGLF**MLQKKSVHKKNTIYE	Yes
Rd4-B41	2	Hematopoietic specific gene	DDLL**LY**INARELRIT**Y**GQGSGTA**VY**NNNM **Y**ANM**Y**NTGK	Yes
Rd4-T23	3	Translation initiation factor eIF-4AIII	LVLDEADEMLNKGFKEQ**IYDVYRYLPP**A	Yes
Rd5-26	3	Shb = Src Homology	QQETVIILE**DYADPY**DAKRTKGQRDAERV GENDG**YMEPYD**A	Yes

Figure 4. Analysis of Clones Obtained from Round Four or Five of the Selection for Cellular Targets of v-abl

(A) Hybridization of radiolabeled DNA from round zero or round four to a cDNA microarray filter. Hybridization of the pools to one of the filter's six grids is shown. Examples of increased signals following hybridization of the round four pool relative to the round zero pool are indicated.
(B) Clones identified by the DNA microarray analysis as targets of v-abl kinase. The first five clones are published targets of v-abl. A subset of ESTs showing increased signals following hybridization of the round four DNA pool that contains potential v-abl phosphorylation sites is shown.
(C) A subset of genes identified by sequence analysis following round four (Rd4) or five (Rd5) of the v-abl selection. Parent gene identities were determined based on BLAST homology searches. Relevant tyrosine residues and surrounding amino acids are underlined and in bold type. The first clone has been reported to bind to v-abl.

unidentified protein that is homologus to a family of SH2 domain adaptor proteins.

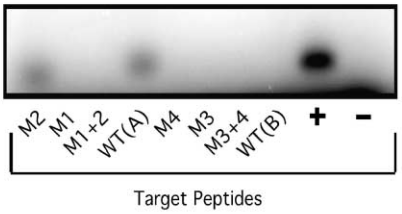
Identification of kinase substrates from synthetic random peptide libraries allows for the survey of all possible sequence combinations having the desired phenotype. This allows for structure-function analysis of the kinase active site and the efficient development of kinase inhibitors. Selections from a randomized peptide library



B

Peptide Sequence	Name
ILEDYADPYDAK	WT (A)
ILED**F**ADP**F**DAK	M 1+2
ILED**F**ADPYDAK	M 1
ILEDYADP**F**DAK	M 2
ENDGYMEPYIA	WT(B)
ENDG**F**MEP**F**IA	M 3+4
ENDG**F**MEPYIA	M 3
ENDGYMEP**F**IA	M 4

C



D

```
                                                            60
MQWSPTPGAS ACLGWASSLA CSTAPTLLGR AGRGPLMAAK WFKEFPLNLK TVSERAKPGG
                                                           120
GGGKLRKNSE AGGAGPGPGK GRKNSAAELG SGRAGVGPKD SRLSRDSLQG LIQAAAGKGR
                                                           180
KNSRATEEEP HRGATKSSGC STYINRLIKV DTQEKNGKSN YPSSSSSSSS SSSSASSSPS
                                                           240
SLGPELDKGK IIKQQETVII LEDYADPYDA KRTKGQRDAE RVGENDGYME PYDAQQMITE
                                                           300
IRRRGSKDPL VKALQLLDSP CEPADGGLKS ETLAKRRSSK DLLGKPPQLY DTPYEPAEGG
                                                           360
PRAEGKARPP DSRLPENDER PAAEYEQPWE WKKEQIVRAL SVQFEGAERP SFREETVRQH
                                                           420
HRQKSWTQKI LKPALSDHSE GEKVDPGLPL EKQPWYHGAI SRAEAESRLQ PCKEAGYLVR
                                                           480
NSESGNSRYS IALKTSQGCV HIIVAQTKDN KYTLNQTSAV FDSIPEVVHY YSNEKLPFKG

AEHMTLLYPV HSKLH*
```

Figure 5. Sequence Analysis of Shg, a Novel Member of a Class of SH2 Domain-Containing Adaptor Proteins

(A) Sequence comparison between TC26 and other members of the SH2 domain adaptor protein family for which full-length sequences are published. Potential v-abl phosphorylation sites are indicated.

(B) Peptides used to map the tyrosine residue phosphorylated by v-abl.

(C) v-abl phosphorylates the first tyrosine of the TC26 clone. Phosphorylation reactions were done in the presence of recombinant v-abl and trace amounts of [γ-^{32}P]ATP in the absence (−) of kinase substrate, using a commercially available v-abl target peptide (+) as the substrate or one of the peptides shown in (B) as the substrate.

(D) The amino acid sequence of the full-length Shg clone. The original sequence obtained following five rounds of selection is boxed, with the tyrosine substrate of v-abl underlined and in bold. A putative myristylation site and an SH2 domain are indicated by a dashed and solid underline, respectively. The asterisk denotes the stop codon.

displayed as mRNA-peptide fusions identified I/L/V-Y-$X_{1\text{-}5}$-P/F as a target sequence for v-abl phosphorylation. Although n = 4 or 5 were preferred, more proximal positioning of the P/F residues was tolerated. This sequence is similar to the I/L/V-Y-$X_{3\text{-}4}$-P/F consensus identified previously. Differences in the scaffold sequences surrounding the randomized region, the larger size of the library assayed (10^{13} members), and the incorporation of all possible amino acids in the synthesis of the randomized segments may account for slight differences in the consensus sequences. In addition, the use of PCR to amplify genetic material encoding the desired phenotype allowed for a sensitive determination of substrate specificity. Interestingly, many of the selected peptides contained more than one tyrosine. In some cases, the novel tyrosine was not part of the consensus sequence, and, therefore, its presence may not be significant. However, in other cases, dual phosphorylation of the tyrosine residues may have aided in selection of the target. The presence of cysteines flanking the randomized region allows for the possibility that the randomized peptides can be exhibited as part of a constrained surface loop during the selection. Although impossible to rule out, we believe that this is rather unlikely given the presence of DTT in the v-abl kinase reaction and the absence of cysteine residues in the selected clones.

Most clones selected as v-abl targets from the cellular mRNA-protein library contained sequences closely matching the consensus sequence identified from the randomized peptide library. Cellular mRNA-protein libraries offer several advantages for the identification of novel kinase-substrate interactions. First, since they are derived from cellular mRNA, they reflect the cell's proteome and therefore consist of a pool of biologically relevant targets. Second, since the libraries are of cellular origin, clones enriched during the selection can be efficiently analyzed by hybridization to cDNA microarray filters analogous to the methods used to quantify cellular transcript levels. Third, the relatively limited number of mRNAs in a cell, coupled with the large size of the library, allows for multiple representation of each protein in the starting library. And, finally, because a random priming approach is used during cDNA synthesis of the mRNA-protein fusion library, minimal interacting domains of interest can be mapped from overlapping clones enriched during the selection.

Numerous substrates identified from the cellular mRNA-protein selection are closely associated with proteins previously identified as ligands or substrates of v-abl. For example, nebulin is a skeletal protein closely associated with actin [37, 44]. Significantly, nebulin also contains an SH3 domain, a motif present in other abl-interacting proteins such as Grb-2, Nck, and Crk v-abl

A

Name	Peptides Tested	Isoelectric Point	Inhibition	
			500 μM	50 μM
v-abl(mut)	EALFAAPFAKKK	9.5	+	–
mRd6-21	IFIAFFSYGSFKK	9.6	++	–
mRd6-38	LIFSLFAWPFLKKK	10.1	+	–
mRd6-45	IHASLFFVWPFKKK	10.1	+	–
m-T5	AFGLFMLQKKS	9.8	+	–
m-T23	QIFDVFRYLPPAKK	9.6	–	–
m-B47	SDGIFGSTFFVKK	8.7	++	–
m-M7	EEVVFEEVPG	3.25	–	–
m-M16	PSAIFLFFSFEKK	8.7	++	+
m-T13	SFRVFALFHSLPTT	8.8	++	–
m-22	LLGLFFVPGTKK	9.6	+	–
m-TC26a	ILEDFADPFDAK	3.5	–	–
m-TC26a-1	ILEDFADPYDAK	3.2	–	–
m-TC26b	ENDGYMEPYIA	3.4	–	–
m-TC26b-1	ENDGFMEPFIA	3.4	–	–
m-TC26b-2	ENDGYMEPFIA	3.2	–	–
m-TC26b-3	ENDGFMEPYIA	3.2	–	–
r-Abl	EKPFKAAALAFK	9.5	–	–
r-B47	VSDGFSGFITFKK	8.75	–	–
r-M7	EVEGPVEEFV	3.25	–	–

B

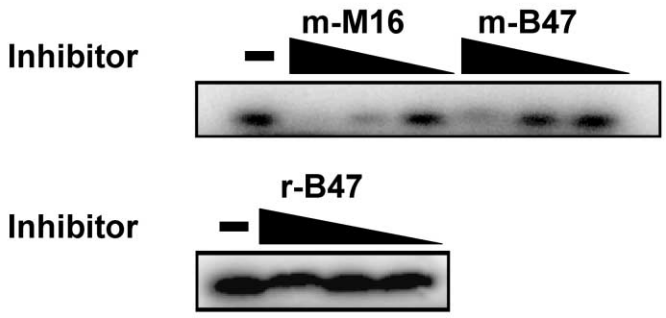


Figure 6. Tyrosine-to-Phenylalanine Mutations Can Convert Cellular Targets of v-abl to Kinase Inhibitors

(A) Sequences of the phenylalanine derivatives tested for inhibition of v-abl kinase activity. Mutagenized test peptides (m series) as well as the scrambled controls (r series) are shown. Relative v-abl activity was scored as: 90%–100% (−), 50%–90% (+), 10%–50% (++). A subset of control peptides (r series) containing the same amino acids as the test peptides, but in a scrambled order, did not have a significant effect on v-abl kinase activity.

(B) Examples of kinase inhibition assays. Mutagenized peptides were added to kinase reactions containing an optimized v-abl substrate at a 100:1, 10:1, or 1:1 molar ratio relative to the target peptide. Kinase reactions in the absence of an inhibitory peptide (−) were done for comparison.

[29]. The high representation of nebulin in the selected pools, and the ability of v-abl to phosphorylate nebulin in vitro, suggests that this protein may play a biological role in v-abl function. Interestingly, Filamin A [45] and another actin binding protein associated with the SWI/SNF complex [46] were also enriched in these experiments. In addition, we also selected for the Grb2-associated binding protein Gab1 [47]. Grb2 has previously been reported to interact with v-abl, c-abl, and BCR-abl [23]. It may be significant that Gab1 can serve as a docking site for several SH2-containing proteins and that it contains numerous possible tyrosine phosphorylation sites.

C-abl is both a nuclear and cytosolic protein and has been shown to bind to DNA and to phosphorylate the carboxyl-terminal domain (CTD) of the large subunit (RPB1) of RNA polymerase II [35]. In contrast, v-abl is found only in the cytoplasm, and therefore nuclear proteins are unlikely to serve as biological substrates. Nevertheless, v-abl and c-abl share significant regions of homology. Therefore, the selection of the RPB2 subunit of RNA polymerase II and the translation factor eIF-4AIII in our experiments suggests that these proteins might also serve as substrates of c-abl. In addition to substrate sequences at the site of catalysis, kinase specificity is also determined by cellular localization and the interaction of the kinase with adaptor proteins or domains [2]. All these factors must be considered when evaluating the biological relevance of protein domains identified during the selection.

We have obtained the full-length sequence of Shg, a novel member of a class of SH2 domain-containing adaptor proteins that consists of the proteins Shb, Shc, Shd, She, and Shf. Shb functions in T cell receptor signaling and binds to the SH3 domain of Grb2 via its proline-rich motif [48, 49], while the SH2 domain of Shc is believed to mediate the function of the epidermal growth factor receptor (EGFR) function through its interaction with an unidentified protein [50]. Shd and She were originally identified in a yeast two-hybrid screen using the c-abl kinase domain as the bait [38], while Shf was cloned by searching the GenBank database for proteins homologous to Shb [51]. Although Shc and Shd are substrates of v-abl in vivo, the roles they play in the

abl signaling pathways are unclear. The presence of a serine-rich domain and a myristylation site in Shg suggests a unique role for this protein in abl function.

In addition to the residues immediately surrounding the target tyrosine, more distal amino acids clearly play a role in the presentation of the target to the kinase active site. Consequently, a subset of v-abl peptides and cellular targets identified in this study were modified by a tyrosine-to-phenylalanine mutation and tested for their ability to inhibit v-abl kinase activity in vitro. Four of the 13 peptides tested inhibited v-abl kinase activity. The ability of these peptides to inhibit kinase activity in vivo is being tested.

Together these experiments demonstrate that tyrosine kinase substrate sequences can be selected from synthetic mRNA-protein fusion libraries as well as from proteomic libraries derived from cellular mRNA. As a universal tool for proteomic applications, the cellular mRNA-protein fusion libraries offer a tremendous resource. The robustness of the mRNA-protein covalent bond, and the fact that library production and selection occur entirely in vitro, makes this approach ideally suited for multiple purposes. This includes the identification of protein-protein [52] and protein-drug (M. McPherson et al., personal communication) interactions in addition to substrates undergoing posttranslational modifications. When used in conjunction with protein structural information, protein localization data, and gene expression profiling, this technology will provide a critical element for the postgenomic analysis of the proteome.

Significance

Tyrosine kinases play a critical role in numerous biological processes including the response of cells to changing environmental conditions. Identification of kinase substrates is a critical step toward understanding the cellular function of kinases. Here, we employ a display technology in which proteins are covalently linked to their encoding mRNA via a puromycin linker (mRNA-protein fusion) to identify v-abl tyrosine kinase substrates from a large (10^{12}–10^{13}) synthetic peptide library. In addition, a library in which the proteome of human bone marrow cells is represented as mRNA-protein fusions was synthesized and screened for cellular targets of v-abl. Several known targets of v-abl were identified, as well as novel targets including a new member of a class of SH2 domain-containing adaptor proteins. The identification of v-abl targets from both libraries led to the rapid development of peptides that inhibit v-abl kinase activity in vitro. Proteomic mRNA-protein libraries provide an important tool for the identification of cellular protein-protein interactions. In addition to determining the substrates of tyrosine kinases, cellular libraries can also be used to identify a variety of other enzyme-substrate interactions, since the in vitro nature of the selection cycle allows for reaction conditions to be optimized and eliminates nonspecific protein modification(s) during propagation in the host cell. Furthermore, the in vitro construction of the libraries minimizes the loss of genetic information normally associated with conventional subcloning and transformation of libraries into host cells. In addition to querying enzyme-substrate interactions, cellular mRNA-protein libraries have also been used recently to identify protein-protein and protein-drug interactions. When used in conjunction with protein structural studies and transcript profiling, cellular mRNA-protein libraries will be an important tool in deciphering the network of cellular protein interactions.

Experimental Procedures

Construction of mRNA-Protein Fusion Libraries

Randomized Peptide Library

Approximately 1.2 nmol of the gel-purified Abl-2NNS oligo (5′-GGCGAGGAGGGATGTGGAGGANNS$_5$TACNNS$_5$GGATGTGGAGAC TACAAGGACGAG-3′), encoding the amino acid sequence 5′-GEEG CGGX$_5$YX$_5$GCGDYKDE-3′, was amplified by PCR (3.5 ml) using the f-Abl-NNS primer (5′-TAATACGACTCACTATAGGGACAATTACTAT TTACAATTACAATGGTGAGCAAGGGCGAGGAGGGATGTGGAGG-3′) and the r-Abl-NNS primer (5′-AGCTTTTGGCTCGTCCTTGTAGTC TCC-3′). The forward primer was designed to incorporate sequences required for in vitro transcription and translation, while the reverse primer was used to add sequences needed for RNA-DNA ligation and fusion production. The library was amplified through five cycles of PCR in a microtiter plate containing Ready-To-Go PCR beads (Amersham Pharmacia) at the following conditions: 95°C for 1 min, 65°C for 2 min, and 72°C for 1 min, followed by a 5-min extension step at 72°C. PCR products were extracted with phenol/chloroform/iso-amyl alcohol, concentrated by ethanol precipitation, and transcribed using the T7 polymerase MEGAshortscript in vitro transcription kit (Ambion). Following treatment with DNase, the RNA was phenol extracted, purified on a NAP 25 column (Amersham Pharmacia), and concentrated by isopropanol precipitation. Purified RNA was ligated to a DNA-puromycin linker (p-dA$_{28}$CCPu) using the SPL-PKA oligo (5′-TTTTTTTTTTNAGCTTTTGGCTCGTC-3′) as a splint between the 3′ terminus of the RNA and the 5′ terminus of the puromycin linker. The RNA, DNA-puromycin linker, and splint (35 nmol each) were heated to 70°C for 3 min and incubated at room temperature prior to the addition of 10× ligation buffer and T4 ligase (Promega). The ligation mixture was resolved on a polyacrylamide-urea denaturing gel (NOVEX), and the ligated product was eluted (200 mM NaCl, 10 mM Tris [pH 7.4], 1 mM EDTA). Following ethanol precipitation, 1 nmol ligated RNA was added to a 1.8 ml Retic lysate IVT translation reaction (Ambion) containing rabbit reticulocyte lysate and [^{35}S]methionine. Following incubation at 30°C for 30 min, mRNA-protein fusion was promoted by the addition of KCl and Mg_2Cl (final concentrations of 500 mM and 150 mM, respectively) and incubation at 25°C for 1 hr. mRNA-protein fusion molecules were purified from the in vitro translation mix by oligo d(T) chromatography by virtue of the poly A sequence in the DNA-puromycin linker. Briefly, the translation-fusion mix was diluted 10-fold in cold (4°C) oligo d(T) binding buffer (100 mM Tris-HCl [pH 8.0], 1 M NaCl, 0.25% Triton X-100, 10 mM EDTA) and bound to equilibrated oligo d(T) resin (Amersham Pharmacia) in a batch format. The oligo d(T) slurry was transferred to a column, and the bed was washed extensively with binding buffer lacking EDTA (4°C). mRNA-protein fusions were eluted in water (25°C) and quantified by scintillation counting. cDNA synthesis from the fusion library (50 pmol) was done using the r-Abl-NNS primer and SUPERSCRIPT II RNase H$^-$ Reverse Transcriptase (GIBCO-BRL Life Technologies) as suggested by the manufacturer. Approximately 25 pmol of the starting library was used in the first round of the selection.

Cellular mRNA-Protein Library

Poly-A$^+$ mRNA from human bone marrow (Clontech) was primed using the R-HBM1 oligonucleotide 5′-GCCTTATCGTCATCGTCCT TGTAGTCGAAACTAGAN$_9$-3′, and cDNA was synthesized using SUPERSCRIPT II RT (Promega). After RNase H treatment, unextended primer was removed by purification over an S-300 (Amersham Pharmacia) size exclusion column equilibrated in TE (10 mM Tris-HCl [pH 8.0], 1 mM EDTA). Second strand cDNA synthesis by

the Klenow fragment of *E. coli* DNA polymerase was primed using the F-HBM2 oligonucleotide 5′-GGACAATTACTATTTACAATTA CAATGN$_9$-3′. Unextended primers were again removed using a S-300 column, and the cDNA products were amplified by PCR using the primer pairs 5′-TAATACGACTCACTATAGGGACAATTACTATT TACAATT-3′ (T7TMVUTR) and 5′-AGAAGATGCGCGATCGTCATCG TCCTTGTAGTC-3′ (FLAGRASS). Taq polymerase (Promega) was added to PCR reactions (1.2 ml) after an initial 5-min denaturation step at 95°C. An annealing temperature of 44°C (2 min) was used in the first four rounds of the PCR reaction, followed by 65°C in the subsequent rounds (20–30 cycles). Denaturation and extension steps were done at 95°C (1 min) and 72°C (2 min), respectively, followed by a final 10-min extension step at 72°C. PCR products were concentrated using the QIAquick PCR purification kit (Qiagen) and fractionated on an S-300 Sephadex column. DNA from the first two fractions (500 μl) was ethanol precipitated. RNA (~5 nmol) was synthesized using the T7 polymerase MEGAscript in vitro transcription kit (Ambion) according to the manufacturer's suggestions. RNA (~3 nmol) was ligated to 4.5 nmol of the p-dA$_{28}$CCPu puromycin linker using the biotinylated SPLINTRASS oligo (4.5 nmol, 5′-B2B1GCAACGACCAACTTTTTTTTTN-3′) as the splint. The RNA, DNA-puromycin linker, and splint were heated to 80°C for 10 min and then cooled to 20°C (0.1°C/min) prior to the addition of 10× ligation buffer and T4 ligase (Promega). Reactions were incubated at 20°C overnight, and then diluted to 1 ml in PBS buffer. Neutra-Avidin beads (Pierce) were added (600 μl) to the mixture and incubated for 1 hr at 30°C. The beads were washed three times in prewarmed PBS (30°C). Ligated RNA was dissociated from the biotinylated splint by resuspending the Neutra-Avidin beads in an equal volume of water and incubating at 45°C for 15 min (repeated twice). Approximately 200 pmol ligated RNA was recovered in a typical reaction. Following ethanol precipitation of the RNA, fusion production and cDNA synthesis were done as described for the random peptide library, except that the RT-RASS oligo (5′-TTTTTTAGAA GATGCGCGATCGTCA-3′) was used as the primer. The selection was initiated with approximately 1 pmol of the library.

Selection of mRNA-Protein Fusions Containing Phosphorylated Tyrosines

Kinase substrates were selected from the random peptide and cellular mRNA-protein fusion libraries using the same strategy. Typically, PCR reactions (1.2 ml) were done using 1/10 the reaction volume of eluted DNA from the previous round of selection. Following synthesis of the mRNA-protein library, a typical round of selection involved preclearing the library with α4G10/protein A Sepharose beads, phosphorylation of substrates with v-abl, immunoprecipitation of phosphotyrosines with the α4G10/protein A Sepharose beads, and recovery of the attached genetic information by PCR. Briefly, following cDNA synthesis (see Construction of mRNA-Protein Fusion Libraries), mRNA-protein libraries were diluted 10-fold into α4G10 binding buffer (50 mM Tris-HCl [pH 7.6], 150 mM NaCl, 1% NP 40 and 1 mM EGTA, 10% glycerol, 1 mM Na_3VO_4, and 1 mM NaF) and incubated with 20 μg α4G10 antibody (Upstate Biotechnology) at 4°C for 2 hr. Preequilibrated protein A Sepharose beads (400 μl, 50% slurry; Amersham Pharmacia) were added to the immunoprecipitation buffer, and incubation continued for 1 hr. The nonbound void fraction was diluted 2-fold in oligo d(T) binding buffer and purified on an oligo d(T) column as described previously. Peak fractions (10–20 pmol) were added directly to kinase reactions (3 hr at 30°C) containing recombinant v-abl (New England Biolabs) and 100 μM ATP under conditions described by the manufacturer. Phosphorylated targets were immunoprecipitated using the α4G10 antibody and protein A Sepharose beads, and the beads were washed five times with the immunoprecipitation buffer. Bound complexes and encoding cDNAs were eluted with 0.1 N KOH (4× 100 μl) and neutralized with 1/10 volume of a 1 M Tris-HCl (pH 8.0) and acetic acid solution prior to PCR. RNA-protein fusion molecules were quantified by scintillation counting in the presence of scintillation fluid. In round one of the randomized peptide and cellular selections, the preclear step was omitted and the mRNA-protein fusion libraries were added to kinase reactions after the reverse transcription step and purification on an oligo d(T) column. Immunoprecipitation and recovery of cDNA from the bound complexes was as described. Exceptionally, [γ-^{32}P]ATP was added to the kinase reaction in round one of the cellular selection to facilitate the detection of the fusions.

Analysis of cDNA Microarrays

Pool PCR products were labeled by random priming using the Redi-prime DNA labeling system (Amersham Pharmacia) and [α-^{33}P]dCTP (2000–4000 Ci/mmol). Unincorporated [α-^{33}P]dCTP was removed by chromatography, and labeled DNA was hybridized to the filters (GenomeSystems) as described by the manufacturer. Hybridization signals were visualized by phosphorimager analysis, and the relative intensities were determined using the manufacturer's Genome Discovery Software (GDS).

Kinase Assays

Phosphorylation of the mRNA-protein fusion library by v-abl was done as described above. To test individual clones for phosphorylation by v-abl, unligated RNA (~50 pmol) was in vitro translated in the absence of [^{35}S]methionine. FLAG-tagged peptides were immunoprecipitated with M2-FLAG agarose beads (Sigma) and equilibrated in v-abl kinase buffer. Phosphorylation reactions were done in the presence of [γ-^{32}P]ATP essentially as suggested by the manufacturer. Free [γ-^{32}P]ATP was removed by washing the beads in kinase buffer. Phosphorylated products were resolved on a 4%–12% BIS-Tris gel with MES as the running buffer and visualized by the STORM 860 phosphorimaging system (Molecular Dynamics). To confirm the immunoprecipitation of the desired peptides, parallel immunoprecipitations were done using peptides labeled with [^{35}S]methionine. Kinase inhibition assays contained 5 μM of the commercial v-abl substrate (EAIYAAPFAKKK; New England Biolabs) and increasing concentrations (5 μM, 50 μM, or 500 μM) of the test peptides. Peptides mRd6-21, mRd6-38, mRd6-45, mRd4-T5, Rd4-T23, Rd4-B47, rRd4-B47, Rd4-M16, and the R5-TC26b series were dissolved in a DMF (50%)-water (50%) solution (500 μM). The remaining peptides were soluble in water.

Cloning of Full-Length Shg

A nested PCR strategy using gene-specific (TC26) and flanking adaptor primers was used to amplify the 3′ and partial 5′ *shg* sequences from double-stranded cDNA prepared from human bone marrow mRNA (Marathon-Ready cDNA, Clontech). Additional 5′ *shg* sequences were obtained by 5′ RACE (GIBCO-BRL Life Technologies) essentially as described by the manufacturer, except that cDNA synthesis was done at 55°C and the terminal transferase reaction was done in the presence of dATP. Nested gene-specific primers and appropriate flanking primers (3′ RACE adaptor and AUAP primers, GIBCO-BRL Life Technologies) were used to PCR the terminal 5′ sequences of *shg*. PCR analysis using shg-specific flanking primers yielded a single 1.49-kbp band that hybridized to a transcript ~1.5 kbp in size following hybridization to Northern blots (data not shown).

Sequence Analysis

PCR products were cloned into pCR2.1-TOPO vector and transformed into TOP10 cells as recommended by the manufacturer (Invitrogen). DNA was prepared using the Qiaprep Spin miniprep kit (Qiagen), and inserts were sequenced (DNA Sequencing Core Facility, Massachusetts General Hospital, Boston, MA). BLAST and ClustalW[60] were used to search the GenBank database and align protein sequences, respectively.

Acknowledgments

We wish to thank Laird Bloom, Stephen Hale, Markus Kurz, Peter Loshe, and Rick Wagner for helpful discussions and review of the manuscript.

Received: June 19, 2001
Revised: November 22, 2001
Accepted: November 26, 2001

References

1. Acuto, O., and Cantrell, D. (2000). T cell activation and the cytoskeleton. Annu. Rev. Immunol. *18*, 165–184.
2. Hunter, T. (2000). Signaling—2000 and beyond. Cell *100*, 113–127.
3. Pawson, T., and Scott, J.D. (1997). Signaling through scaffold, anchoring, and adaptor proteins. Science *278*, 2075–2080.
4. Mendelsohn, A.R., and Brent, R. (1999). Protein interaction methods—toward an endgame. Science *284*, 1948–1950.
5. Phizicky, E.M., and Fields, S. (1995). Protein-protein interactions: methods for detection and analysis. Microbiol. Rev. *59*, 94–123.
6. Brent, R., and Finley, R.L., Jr. (1997). Understanding gene and allele function with two-hybrid methods. Annu. Rev. Genet. *31*, 663–704.
7. Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. Nature *340*, 245–246.
8. Vidal, M., and Legrain, P. (1999). Yeast forward and reverse 'n'-hybrid systems. Nucleic Acids Res. *27*, 919–929.
9. Hu, J.C., Kornacker, M.G., and Hochschild, A. (2000). *Escherichia coli* one- and two-hybrid systems for the analysis and identification of protein-protein interactions. Methods *20*, 80–94.
10. Johnsson, K., and Ge, L. (1999). Phage display of combinatorial peptide and protein libraries and their applications in biology and chemistry. Curr. Top. Microbiol. Immunol. *243*, 87–105.
11. Smith, G.P., and Petrenko, V.A. (1997). Phage display. Chem. Rev. *97*, 391–410.
12. Wittrup, K.D. (1999). Phage on display. Trends Biotechnol. *17*, 423–424.
13. Schaffitzel, C., Hanes, J., Jermutus, L., and Pluckthun, A. (1999). Ribosome display: an in vitro method for selection and evolution of antibodies from libraries. J. Immunol. Methods *231*, 119–135.
14. Shusta, E.V., VanAntwerp, J., and Wittrup, K.D. (1999). Biosynthetic polypeptide libraries. Curr. Opin. Biotechnol. *10*, 117–122.
15. Obata, T., Yaffe, M.B., Leparc, G.G., Piro, E.T., Maegawa, H., Kashiwagi, A., Kikkawa, R., and Cantley, L.C. (2000). Peptide and protein library screening defines optimal substrate motifs for AKT/PKB. J. Biol. Chem. *275*, 36108–36115.
16. Habelhah, H., Shah, K., Huang, L., Burlingame, A.L., Shokat, K.M., and Ronai, Z. (2001). Identification of a new JNK substrate using ATP pocket mutant ADK and a corresponding ATP analogue. J. Biol. Chem. *276*, 18090–18095.
17. Songyang, Z., Blechner, S., Hoagland, N., Hoekstra, M.F., Piwnica-Worms, H., and Cantley, L.C. (1994). Use of an oriented peptide library to determine the optimal substrates of protein kinases. Curr. Biol. *4*, 973–982.
18. Roberts, R.W., and Szostak, J.W. (1997). RNA-peptide fusions for the in vitro selection of peptides and proteins. Proc. Natl. Acad. Sci. USA *94*, 12297–12302.
19. Roberts, R.W. (1999). Totally in vitro protein selection using mRNA-protein fusions and ribosome display. Curr. Opin. Chem. Biol. *3*, 268–273.
20. Gold, L. (2001). mRNA display: diversity matters during in vitro selection. Proc. Natl. Acad. Sci. USA *98*, 4825–4826.
21. Wilson, D.S., Keefe, A.D., and Szostak, J.W. (2001). The use of mRNA display to select high-affinity protein-binding peptides. Proc. Natl. Acad. Sci. USA *98*, 3750–3755.
22. Rosenberg, N., and Witte, O.N. (1988). The viral and cellular forms of the Abelson (abl) oncogene. Adv. Virus Res. *35*, 39–81.
23. Zou, X., and Calame, K. (1999). Signaling pathways activated by oncogenic forms of Abl tyrosine kinase. J. Biol. Chem. *274*, 18141–18144.
24. Cortes, J.E., Talpaz, M., and Kantarjian, H. (1996). Chronic myelogenous leukemia: a review. Am. J. Med. *100*, 555–570.
25. Gishizky, M.L. (1996). Molecular mechanisms of Bcr-Abl-induced oncogenesis. Cytokines Mol. Ther. *2*, 251–261.
26. Laneuville, P. (1995). Abl tyrosine protein kinase. Semin. Immunol. *7*, 255–266.
27. Pendergast, A.M. (1996). Nuclear tyrosine kinases: from Abl to WEE1. Curr. Opin. Cell Biol. *8*, 174–181.
28. Songyang, Z., Carraway, K.L., 3rd, Eck, M.J., Harrison, S.C., Feldman, R.A., Mohammadi, M., Schlessinger, J., Hubbard, S.R., Smith, D.P., Eng, C., et al. (1995). Catalytic specificity of protein-tyrosine kinases is critical for selective signalling. Nature *373*, 536–539.
29. Feller, S.M., Ren, R., Hanafusa, H., and Baltimore, D. (1994). SH2 and SH3 domains as molecular adhesives: the interactions of Crk and Abl. Trends Biochem. Sci. *19*, 453–458.
30. Ren, R., Ye, Z.S., and Baltimore, D. (1994). Abl protein-tyrosine kinase selects the Crk adapter as a substrate using SH3-binding sites. Genes Dev. *8*, 783–795.
31. Varticovski, L., Daley, G.Q., Jackson, P., Baltimore, D., and Cantley, L.C. (1991). Activation of phosphatidylinositol 3-kinase in cells expressing abl oncogene variants. Mol. Cell. Biol. *11*, 1107–1113.
32. Wang, B., Golemis, E.A., and Kruh, G.D. (1997). ArgBP2, a multiple Src homology 3 domain-containing, Arg/Abl-interacting protein, is phosphorylated in v-Abl-transformed cells and localized in stress fibers and cardiocyte Z-disks. J. Biol. Chem. *272*, 17542–17550.
33. Danial, N.N., Losman, J.A., Lu, T., Yip, N., Krishnan, K., Krolewski, J., Goff, S.P., Wang, J.Y., and Rothman, P.B. (1998). Direct interaction of Jak1 and v-Abl is required for v-Abl-induced activation of STATs and proliferation. Mol. Cell. Biol. *18*, 6795–6804.
34. Danial, N.N., and Rothman, P. (2000). JAK-STAT signaling activated by Abl oncogenes. Oncogene *19*, 2523–2531.
35. Baskaran, R., Dahmus, M.E., and Wang, J.Y. (1993). Tyrosine phosphorylation of mammalian RNA polymerase II carboxyl-terminal domain. Proc. Natl. Acad. Sci. USA *90*, 11167–11171.
36. Van Etten, R.A., Jackson, P.K., Baltimore, D., Sanders, M.C., Matsudaira, P.T., and Janmey, P.A. (1994). The COOH terminus of the c-Abl tyrosine kinase contains distinct F- and G-actin binding domains with bundling activity. J. Cell Biol. *124*, 325–340.
37. Wang, J.Y. (1993). Abl tyrosine kinase in signal transduction and cell-cycle regulation. Curr. Opin. Genet. Dev. *3*, 35–43.
38. Oda, T., Kujovich, J., Reis, M., Newman, B., and Druker, B.J. (1997). Identification and characterization of two novel SH2 domain-containing proteins from a yeast two hybrid screen with the ABL tyrosine kinase. Oncogene *15*, 1255–1262.
39. Raffel, G.D., Parmar, K., and Rosenberg, N. (1996). In vivo association of v-Abl with Shc mediated by a non-phosphotyrosine-dependent SH2 interaction. J. Biol. Chem. *271*, 4640–4645.
40. Cujec, T.P., Okamoto, H., Fujinaga, K., Meyer, J., Chamberlin, H., Morgan, D.O., and Peterlin, B.M. (1997). The HIV transactivator TAT binds to the CDK-activating kinase and activates the phosphorylation of the carboxy-terminal domain of RNA polymerase II. Genes Dev. *11*, 2645–2657.
41. Okamoto, H., Cujec, T.P., Peterlin, B.M., and Okamoto, T. (2000). HIV-1 replication is inhibited by a pseudo-substrate peptide that blocks Tat transactivation. Virology *270*, 337–344.
42. Poteet-Smith, C.E., Shabb, J.B., Francis, S.H., and Corbin, J.D. (1997). Identification of critical determinants for autoinhibition in the pseudosubstrate region of type I alpha cAMP-dependent protein kinase. J. Biol. Chem. *272*, 379–388.
43. Nishikawa, K., Sawasdikosol, S., Fruman, D.A., Lai, J., Songyang, Z., Burakoff, S.J., Yaffe, M.B., and Cantley, L.C. (2000). A peptide library approach identifies a specific inhibitor for the ZAP-70 protein tyrosine kinase. Mol. Cell *6*, 969–974.
44. Wang, K., Knipfer, M., Huang, Q.Q., van Heerden, A., Hsu, L.C., Gutierrez, G., Quian, X.L., and Stedman, H. (1996). Human skeletal muscle nebulin sequence encodes a blueprint for thin filament architecture. Sequence motifs and affinity profiles of tandem repeats and terminal SH3. J. Biol. Chem. *271*, 4304–4314.
45. Mejean, C., Lebart, M.C., Boyer, M., Roustan, C., and Benyamin, Y. (1992). Localization and identification of actin structures involved in the filamin-actin interaction. Eur. J. Biochem. *209*, 555–562.
46. Wang, W., Xue, Y., Zhou, S., Kuo, A., Cairns, B.R., and Crabtree, G.R. (1996). Diversity and specialization of mammalian SWI/SNF complexes. Genes Dev. *10*, 2117–2130.
47. Holgado-Madruga, M., Emlet, D.R., Moscatello, D.K., Godwin, A.K., and Wong, A.J. (1996). A Grb2-associated docking protein in EGF- and insulin-receptor signalling. Nature *379*, 560–564.
48. Welsh, M., Mares, J., Karlsson, T., Lavergne, C., Breant, B., and

Claesson-Welsh, L. (1994). Shb is a ubiquitously expressed Src homology 2 protein. Oncogene *9*, 19–27.

49. Welsh, M., Songyang, Z., Frantz, J.D., Trub, T., Reedquist, K.A., Karlsson, T., Miyazaki, M., Cantley, L.C., Band, H., and Shoelson, S.E. (1998). Stimulation through the T cell receptor leads to interactions between SHB and several signaling proteins. Oncogene *16*, 891–901.
50. O'Bryan, J.P., Lambert, Q.T., and Der, C.J. (1998). The src homology 2 and phosphotyrosine binding domains of the ShcC adaptor protein function as inhibitors of mitogenic signaling by the epidermal growth factor receptor. J. Biol. Chem. *273*, 20431–20437.
51. Lindholm, C.K., Frantz, J.D., Shoelson, S.E., and Welsh, M. (2000). Shf, a Shb-like adapter protein, is involved in PDGF-alpha-receptor regulation of apoptosis. Biochem. Biophys. Res. Commun. *278*, 537–543.
52. Hammond, P.W., Alpin, J., Rise, C.E., Wright, M.C., and Kreider, B.L. (2001). In vitro selection and characterization of Bcl-XL binding proteins from a mix of tissue-specific mRNA display libraries. J. Biol. Chem. *276*, 20898–20906.