Biased Combinatorial Libraries: Novel Ligands for the SH3 Domain of Phosphatidylinositol 3-Kinase

James K. Chen,† William S. Lane,‡ Andrew W. Brauer,§ Akito Tanaka,† and Stuart L. Schreiber*,†

Department of Chemistry, Harvard University Cambridge, Massachusetts 02138
Microchemistry Facility, Harvard University Cambridge, Massachusetts 02138
ARIAD Pharmaceuticals, Inc. Cambridge, Massachusetts 02139

Received September 27, 1993

Several strategies for ligand discovery involving large collections (*libraries*) of ligands prepared by combinatorial syntheses have recently been reported.[1] These frequently involve the preparation of short peptides that are selected on the basis of their ability to bind to a target receptor. We now report a modification of this strategy that uses the general ligand-binding characteristics of a receptor (Figure 1). By incorporating a structural motif that favors receptor-binding into a large collection of peptides, we have generated a biased combinatorial library that is effective in yielding specific ligands to the Src homology 3 (SH3) domain from phosphatidylinositol 3-kinase (PI3K). SH3 domains are small (55–70 amino acids) receptor domains that are found in many intracellular signaling proteins.[2,3] The specific role of the PI3K SH3 domain in signal transduction pathways, however, is not well understood; no natural ligands have yet been identified. The discovery of peptide ligands for this domain is an important step toward understanding its biological function and molecular specificity.

Our initial investigations of the PI3K SH3 domain utilized a combinatorial peptide library on beads in analogy to that described by Lam *et al.*[1e] Libraries synthesized by this technique are comprised of millions of peptides, each covalently attached to a unique bead. Beads containing high-affinity ligands are then identified by their ability to interact with a fluorescently tagged receptor. In our first experiments the SH3 domain of PI3K was used to probe two different random libraries: one containing two million hexapeptides and another containing two million cyclic heptapeptides.[4] Both screenings failed to yield suitable ligands. Although the amino acid sequences of positive beads followed general trends, no clear consensus sequence was observed and the binding affinities of these ligands for the SH3 domain were weak (>100 μM).[5]

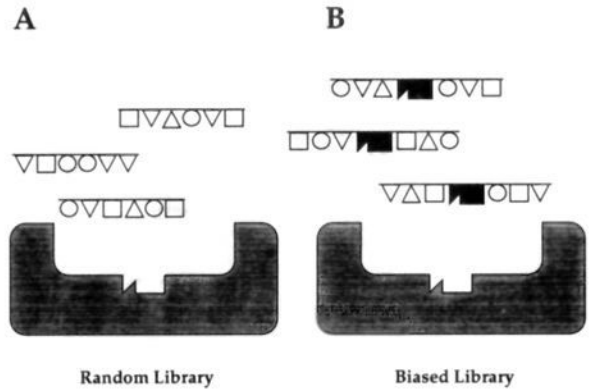

Figure 1. A random library (A) (see text). A biased library (B) can be created for receptors that have predetermined ligand-binding preferences. The biasing element tends to increase the average affinity of the ligands to the target receptor and to orient the ligands in a uniform way.

Disappointed by these results, we attempted to prepare a library of peptide ligands that are intrinsically biased to bind SH3 domains. Recent investigations have demonstrated that SH3 domains bind to proline-rich peptide motifs. The SH3 domain of tyrosine kinase Abl binds specifically to the peptide APTMPPPLPP, a fragment of the SH3-binding protein 3BP1.[6] The two SH3 domains of the adaptor protein Grb2 bind to proline-rich regions such as VPPPVPPRRR in the guanine nucleotide exchange factor Sos1.[7] These and several other peptide ligands of SH3 domains have in common a PPXP motif. On the basis of this observation, we constructed a combinatorial peptide library of the form XXXPPXPXX, where X represents any amino acid other than cysteine.

Approximately two million peptides[8] with this structural bias were synthesized on poly(dimethylacrylamide) beads using a "split and pool" procedure.[9] A spacer (β-alanine ϵ-aminocaproic acid ethylenediamine) was also included to minimize steric interactions of the SH3 domain with the solid support. Next, the PI3K SH3 domain was prepared with a fluorescein tag placed strategically at its N-terminus, which is on the face of the domain opposite that of its ligand-binding site (Figure 2).[3c] This was accomplished by treating the bacterially-expressed, recombinant SH3 domain with sodium periodate, thereby converting its N-terminal serine into an N-terminal glyoxalamide.[10] Condensation with fluorescein-5-thiosemicarbazide provided the SH3 domain with a fluorescein reporter linked to its N-terminus via a thiosemicarbazone moiety (RNHCSNHN=CHCONH–SH3, R = fluorescein). The fluorescein-conjugated PI3K SH3 was mixed with

† Department of Chemistry, Harvard University.
‡ Microchemistry Facility, Harvard University.
§ ARIAD Pharmaceuticals, Inc.

(1) (a) Geysen, H. M.; Meloen, R. H.; Barteling, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 3998. (b) Parmley, S. F.; Smith, G. P. *Gene* **1988**, *73*, 305. (c) Scott, J. K.; Smith, G. P.; Dower, W. J. In *Advances in Gene Technology: The Molecular Biology of Immune Diseases and the Immune Response*; Streilein, J. W., Ed.; IRL: Oxford, 1990; p 224. (d) Fodor, S. P. A.; Read, J. L.; Pirrung, M. C.; Stryer, L.; Lu, A.; Solas, D. *Science* **1991**, *251*, 767. (e) Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. *Nature* **1991**, *354*, 82. (f) Houghten, R. A.; Pinilla, C.; Blondelle, S. E.; Appel, J. R.; Dooley, C. T.; Cuervo, J. H. *Nature* **1991**, *354*, 84. (g) Cull, M. G.; Miller, J. F.; Schatz, P. J. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 1865. (h) Brenner, S.; Lerner, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5381. (i) Kerr, J. M.; Banville, S. C.; Zuckermann, R. N. *J. Am. Chem. Soc.* **1993**, *115*, 2529.

(2) For reviews, see: (a) Musacchio, A.; Gibson, T.; Lehto, V.-P.; Sarastes, M. *FEBS Lett.* **1992**, *307*, 55. (b) Pawson, T.; Schlessinger, J. *Curr. Biol.* **1993**, *3*, 434.

(3) For structures of SH3 domains, see the following. Spectrin: (a) Musacchio, A.; Noble, M.; Pauptit, R.; Wierenga, R. K.; Saraste, M. *Nature* **1992**, *359*, 851. Src: (b) Yu, H.; Rosen, M. K.; Shin, T. B.; Seidel-Dugan, C.; Brugge, J. S.; Schreiber, S. L. *Science* **1992**, *258*, 1665. PI3K: (c) Koyama, S.; Yu, H.; Dalgarno, D.; Shin, T. B.; Zydowsky, L. D.; Schreiber, S. L. *Cell* **1993**, *72*, 945. (d) Booker, G. W.; Gout, I.; Downing, A. K.; Driscoll, P. C.; Boyd, J.; Waterfield, M. D.; Campbell, I. D. *Cell* **1993**, *73*, 813. PLCγ: (e) Kohda, D.; Hatanaka, H.; Odaka, M.; Mandiyan, V.; Ullrich, A.; Schlessinger, J.; Inagaki, F. *Cell* **1993**, *72*, 953. Fyn: (f) Noble, M.; Musacchio, A.; Saraste, M.; Courtneidge, S. A.; Wierenga, R. K. *EMBO J.* **1993**, *12*, 2617.

(4) The general structures of the hexapeptide and cyclic heptapeptide libraries were XXXXXX and CXXXXXC, respectively, where X represents any amino acid other than cysteine. The heptapeptide library was cyclized by disulfide bond formation.

(5) Botfield, M.; Chen, J. K. Unpublished Results. Dissociation constants were measured by the Biosensor System (BIAcore).

(6) (a) Cicchetti, P.; Mayer, B. J.; Thiel, G.; Baltimore, D. *Science* **1992**, *257*, 803. (b) Ren, R.; Mayer, B. J.; Cicchetti, P.; Baltimore, D. *Science* **1993**, *259*, 1157.

(7) (a) Buday, L.; Downward, J. *Cell* **1993**, *73*, 611. (b) Egan, S. E.; Giddings, B. W.; Brooks, M. W.; Buday, L.; Sizeland, A. M.; Weinberg, R. A. *Nature* **1993**, *363*, 45. (c) Rozakis-Adcock, M.; Fernley, R.; Wade, J.; Pawson, T.; Bowtell, D. *Nature* **1993**, *363*, 83. (d) Li, N.; Batzer, A.; Daly, R.; Yajnik, V.; Skolnik, E.; Chardin, P.; Bar-Sagi, D.; Margolis, B.; Schlessinger, J. *Nature* **1993**, *363*, 85. (e) Gale, N. W.; Kaplan, S.; Lowenstein, E. J.; Schlessinger, J.; Bar-Sagi, D. *Nature* **1993**, *363*, 88.

(8) Although there are 47 million possible (non-cysteine-containing) peptides with the motif XXXPPXPXX, our experiments used a random subset of these peptides (ca. 2 million) for operational simplicity.

(9) (a) Furka, Á.; Sebestyén, F.; Asgedom, M.; Dibó, G. *Abstr. 14th Int. Congr. Biochem.*; Prague, Czechoslovakia, 1988; Vol. 5, p 47. (b) Furka, Á.; Sebestyén, F.; Asgedom, M.; Dibó, G. *Int. J. Pept. Protein Res.* **1991**, *37*, 487.

(10) Geoghegan, K. F.; Stroh, J. G. *Bioconjugate Chem.* **1992**, *3*, 138.

0002-7863/93/1515-12591$04.00/0

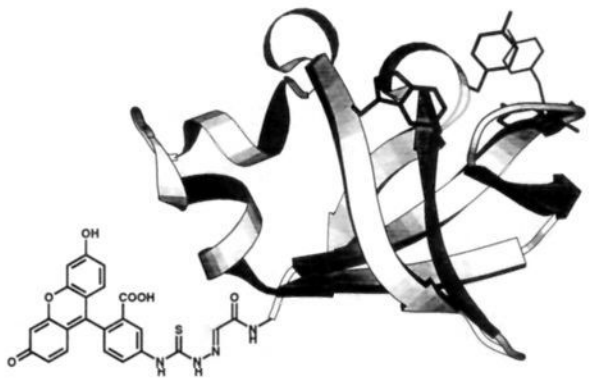

Figure 2. Ribbon drawing of the N-terminally derivatized PI3K SH3 domain that was used to screen the biased combinatorial library. Residues that form the putative binding site are shown in thick solid lines. (Figure generated using the coordinates of the SH3 domain determined by 2D and 3D NMR[3c] and the MOLSCRIPT program.[13])

Table I. Amino Acid Sequences of Bead-Bound Peptides That Bind to the SH3 Domain of PI3K

class I		class II
RKLPPRPSK	**RMLPPKPRV**	VWKPP**LPKR**
RKLPPRPAF	**RKLPPKPKW**	LNKPP**LPKR**
RRLPPRPRR	**RPLPPHPRR**	NRKPP**LPAR**
RYLPPRPMY	**RALPPHPRF**	PWHPP**LPLR**
RLLPPRPTF	**RKLPPLPKA**	
RDLPPRPAA	**RPLPPAPWK**	
PAL**PPRPHS**		

the library, and the beads were washed repeatedly and then analyzed by fluorescence microscopy. Seventy beads were fluorescent, and 17 of the brightest beads were isolated and sequenced by the automated Edman degradation technique.

In contrast to the standard combinatorial libraries that we had first examined, the biased library yielded peptide sequences (Table I) with several conserved elements. These sequences appear to fall into two classes: one particularly striking group comprising 13 sequences and a second group comprising four sequences. With only one exception, the sequences in the first group contain an RXL motif immediately preceding the biasing element. Eleven of these sequences contain a basic residue (R, K, or H) at the site between the prolines (X of PPXP), with arginine being the predominant amino acid at this position. The remaining sites are not highly conserved, although there appears to be a general tendency for basic residues. Thus, the overall consensus for this first class of ligands is RXLPPRPXX, where X tends to be basic. This high frequency of basic residues is consistent with earlier predictions that were based on the structure of the PI3K SH3 domain.[3c]

It is likely that these 13 peptides bind to the SH3 receptor site with a common orientation. The positioning of the RXL motif next to the biasing element is highly conserved, and the arginine and leucine residues probably make specific contacts with structural elements in the SH3 domain. Peptides from the second class of sequences, however, deviate significantly from the pattern

(11) Li *et al.* (see ref 7d) reported that a peptide containing the PPVPPR motif binds to Grb2 with a K_d of 25 nM when measured with the BIAcore system. According to our fluorescence binding assays, however, this interaction is probably much weaker. The N-terminal SH3 domain of Grb2 binds to the peptide VPPPVPPRRR with much lower affinity (see text), and the C-terminal SH3 domain binds with a K_d of 39 μM; the entire Grb2 protein binds with affinity comparable to that of the individual SH3 domains (Chen, J. K.; Simon, J. A.; Schreiber, S. L. Unpublished results).

Table II. Affinities of Synthetic Peptides to the SH3 Domain of PI3K

peptide	K_d, μM
RKLPPRPSK	8.7
RKLPPRPAF	11
RYLPPRPMY	13
RKLPPLPKA	18
RALPPHPRF	30
LNKPPLPKR	13
composite:	
RKLPPRPRR	7.6 ± 0.2[a]

[a] K_d value is the mean and standard deviation of three separate determinations, each derived from fluorescence measurements at nine different ligand concentrations.

mentioned above; these four peptides contain neither the arginine nor the leucine of the RXL motif. It is interesting that the two classes of sequences share an LPXR motif. Thus, the second class might be binding in a different frame from the first, although it is possible that binding occurs in a unique manner, such as in the reverse orientation.

In spite of these differences, both groups of peptides appear to have comparable affinities for the SH3 domain of PI3K. Six representative peptides were chemically synthesized, and their affinities for the native SH3 domain were evaluated by fluorescence spectroscopy. The measured dissociation constants (K_d's) range from 8.7 to 30 μM (Table II), and a composite peptide RKLPPRPRR binds to the PI3K SH3 domain with a K_d of 7.6 ± 0.2 μM. These affinities are among the highest reported for SH3 ligands. The strongest SH3–ligand interaction measured to date is the binding of VPPPVPPRRR to the N-terminal SH3 domain of Grb2. The dissociation constant of this complex obtained by fluorescence spectroscopy is 5.6 μM.[11] Other reported SH3–ligand affinities are much weaker. For example, a fragment of the motor protein dynamin was reported to bind to the PI3K SH3 domain with a K_d of 300 μM,[3d] and the 3BP1-derived peptide APTMPPPLPP binds to the Abl SH3 domain with a K_d of 37 μM.[12] These data demonstrate that the biased combinatorial library yields peptides with affinities that are at least comparable to those of peptides derived from natural ligands.

In summary, we report the conception and application of a biased combinatorial library to molecular investigations of SH3 domains. These results illustrate the utility of biased combinatorial libraries for ligand discovery in systems where there is some general knowledge of the ligand-binding characteristics of the receptor. Since this is a common circumstance, we anticipate that the method should be widely applicable.

Acknowledgment. This work was supported by the National Institute for General Medical Sciences (GM-44993). J.K.C. gratefully acknowledges a National Science Foundation predoctoral fellowship. We thank T. B. Shin for helpful discussions and K. F. Geoghegan for providing experimental details for the N-terminal derivatization of peptides and proteins. We are especially grateful to R. J. Rickles for sharing the results of his independent studies of biased libraries prior to publication.

Supplementary Material Available: General experimental procedures and representative fluorescence titration data (4 pages). This material is contained in many libraries on microfiche, immediately follows this article in the microfilm version of the journal, and can be ordered from the ACS; see any current masthead page for ordering information.

(12) Gomez, F. A.; Chen, J. K.; Tanaka, A.; Schreiber, S. L.; Whitesides, G. M. *J. Org. Chem.*, submitted. This dissociation constant was measured by fluorescence spectroscopy.

(13) Kraulis, P. J. *J. Appl. Crystallogr.* **1991**, *24*, 946.