Protein Structure-Based Combinatorial Chemistry: Discovery of Non-Peptide Binding Elements to Src SH3 Domain

Andrew P. Combs,† Tarun M. Kapoor,† Sibo Feng, James K. Chen, Lygia F. Daudé-Snow, and Stuart L. Schreiber*

Howard Hughes Medical Institute
Department of Chemistry, Harvard University
12 Oxford Street, Cambridge, Massachusetts 02138

Received November 7, 1995

Small molecule ligands can be used to cause a conditional loss or gain of function of their protein receptors and, therefore, can be viewed as equivalents of conditional alleles.[1−3] In order to extend their use, methods to identify such ligands de novo are required. We have previously reported the use of biased combinatorial libraries to discover peptide ligands to proteins,[4] and the coupled use of combinatorial chemistry and structural biology to understand the nature of protein−ligand interactions.[5−7] More recently, we have been exploring whether the knowledge of protein structure can facilitate the design of monomers and linking elements leading to vast numbers of potential ligands targeted to a particular protein. We now report a first illustration of this strategy resulting in the discovery of novel and specific ligands containing non-peptide structural elements.

Structural investigations of SH3−peptide complexes have revealed that SH3 domains bind peptide ligands in either of two orientations (classes I and II; these differ in the directionality of the backbone amides[5,6]) involving the three pockets depicted in Figure 1.[4−7] We designed a library of ligands predisposed to adopt the class I orientation by attaching a common low-affinity (K_d > 1 mM) biasing sequence PLPPLP (P = Pro, L = Leu) to a solid support. This sequence was expected to fill the two pockets (labeled 1 and 2) that bind Leu-Pro dipeptides. Furthermore, structural analyses show that the N-terminal proline should be positioned to orient elements attached to its pyrrolidine nitrogen into the third pocket (labeled 3), which is lined by the nSrc and RT loops common to all SH3 domains and is the primary determinant of ligand specificity.[8] We synthesized an encoded[9] combinatorial library derived from 32 monomers incorporated during three consecutive cycles of split-and-pool synthesis[10,11] following the synthesis of the common PLPPLP sequence (synthesized in the C to N direction) and terminating with one of 32 capping reagents (Figure 2). We purposefully incorporated an encoded blank ("skip-codon")[12] during monomer and cap incorporation in order to increase library diversity significantly by creating sublibraries with deletions at any one or more of the three monomer and one cap sites.

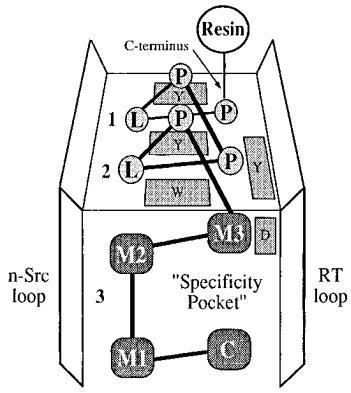


Figure 1. Schematic of the strategy used for the rational design of a library of ligands that direct non-peptide binding elements into the specificity pocket of SH3 proteins. W, D, and Y are one-letter amino acid codes for tryptophan, aspartic acid, and tyrosine, respectively.

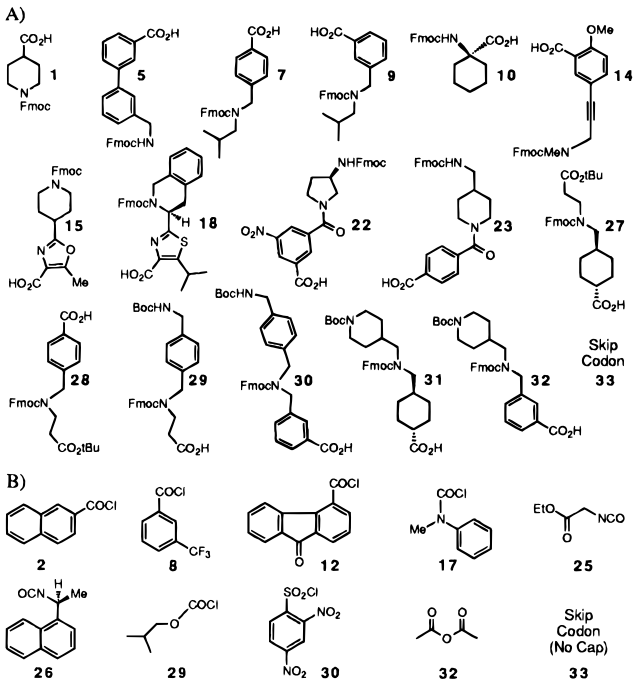


Figure 2. (A) Sample structures of 33 monomers (M1, M2, and M3) used in library synthesis (standard FMOC chemistry). (B) Sample structures of 33 caps (C) used in library synthesis as sequence-terminating reagents. A complete listing of monomers and capping agents can be obtained in the supporting information.

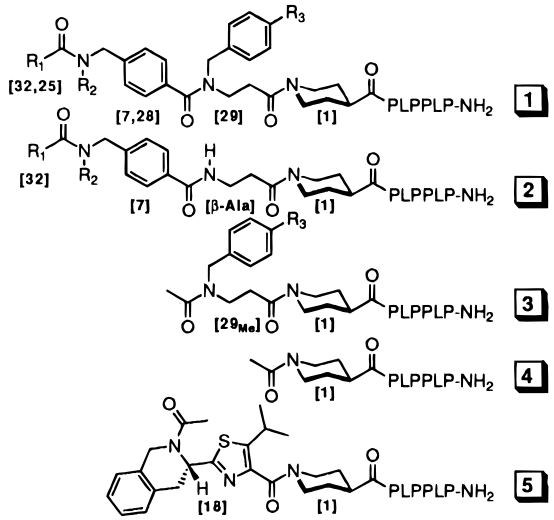
A sensitive binding assay for the SH3 domain from the protein tyrosine kinase Src was developed using N-terminally biotinylated Src SH3 (cf. ref 4) and streptavidin-alkaline phosphatase as a colorimetric reporter.[13] A three-stage screen allowed for the removal of false positives and ensured the selection of the darkest beads from the library.[14] Screening a biased library containing ∼1.1 million discrete compounds identified 15 beads containing ligands specific for the Src SH3 domain. Decoding revealed two consensus sequences (Table 1). Position 3 (M3) was decoded as monomer 1 (see Figure 2A) in 12 of 15 beads. Position 2 (M2) was occupied by either monomer 29 or monomer 18 in 13 of 15 beads. The monomers at the M1 site

* Address correspondence to this author.
† These authors contributed equally to this research.

(1) Brown, E. J.; Albers, M. W.; Shin, T. B.; Ichikawa, K.; Keith, C. T.; Lane, W. S.; Schreiber, S. L. *Nature* **1994**, *369*, 756−768.
(2) Spencer, D. M.; Wandless, T. J.; Schreiber, S. L.; Crabtree, G. R. *Science* **1993**, *262*, 1019−1024.
(3) Pruschy, M. N.; Spencer, D. M.; Kapoor, T. M.; Miyake, H.; Crabtree, G. R.; Schreiber, S. L. *Chem. Biol.* **1994**, *1*, 163−172.
(4) Chen, J. K.; Lane, W. S.; Brauer, A. W.; Tanaka, A.; Schreiber, S. L. *J. Am. Chem. Soc.* **1993**, *115*, 12591−12592.
(5) Yu, H.; Chen, J. K.; Feng, S.; Dalgarno, D. C.; Brauer, A. W.; Schreiber, S. L. *Cell* **1994**, *76*, 933−945.
(6) Feng, S.; Chen, J. K.; Yu, H.; Simon, J. A.; Schreiber, S. L. *Science* **1994**, *266*, 1241−1247.
(7) Chen, J. K.; Schreiber, S. L. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 953−969.
(8) Feng, S.; Kasahara, C.; Rickles, R. J.; Schreiber, S. L. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 12408−12415.
(9) Ohlmeyer, M. H. J.; Swanson, R. N.; Dillard, L. W.; Reader, J. C.; Asouline, G.; Kobayashi, R.; Wigler, M.; Still, W. C. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 10922−10926.
(10) Furka, Á.; Sebestyén, F.; Asgedom, M.; Dibó, G. *Int. J. Pept. Protein Res.* **1991**, *37*, 487−493.
(11) Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. *Nature* **1991**, *354*, 82−84.
(12) This procedure generates a complete representation of the sublibraries: C-M-M-M-PLPPLP, C-M-M-PLPPLP, C-M-PLPPLP, C-PLPPLP.

(13) The library was initially incubated with the preformed complex of N-terminally biotinylated (via semicarbazide linkage) Src SH3 domain and streptavidin-alkaline phosphatase (SAAP) and then washed extensively. Treatment of the library with substrate for alkaline phosphatase gave varying degrees of color deposition on the beads. The darkest beads (∼350) were physically removed with a syringe, destained, and then reassayed with SAAP alone in order to eliminate false positives. None of the beads were identified as SAAP binders. A final high-stringency assay of all ∼350 beads was performed with lower concentrations of Src SH3−SAAP in order to identify the highest affinity ligands.

0002-7863/96/1518-0287$12.00/0



Ligand	Structure	R_1	R_2	R_3	K_d (μM)[a]
32-7-29-1-PLPPLP-NH_2	**1A**	CH_3		CH_2NH_2	3.4
25-28-29-1-PLPPLP-NH_2	**1B**	EtO_2C 		CH_2NH_2	6.6
32-7-29_{Me}[b]-1-PLPPLP-NH_2	**1C**	CH_3		CH_3	6.2
32-7-β-Ala-1-PLPPLP-NH_2	**2**	CH_3			80
Ac-29_{Me}-1-PLPPLP-NH_2	**3**			CH_3	15
Ac-1-PLPPLP-NH_2	**4**				220
Ac-PLPPLP-NH_2					>1000
Ac-18-1-PLPPLP-NH_2	**5**				11

[a]Dissociation constants (K_d) measured by fluorescence perturbation assays.[4] [b]29_{Me} = variant of monomer 29 with R_3=CH_3.

Figure 3. Structures and dissociation constants (K_d's) of ligands.

Table 1. Decoding Data Obtained from the 15 Darkest Beads

Beads	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
C	32	8	25	26	29	2	25	?	?	?	?	12	25	2	8
M1	7	7	28	28	23	9	31	10	10	10	10	30	32	23	22
M2	29	29	29	29	29	29	29	18	18	18	18	9	29	29	31
M3	1	1	1	1	1	1	1	1	1	1	1	1	27	29	29

?= No peaks corresponding to encoding tags observed in GC-ECD trace for this position. Different shading highlights two consensus sequences.

of sequences containing M2 = 29 are less conserved, although they maintain a similar aromatic backbone; no preference was observed at the C site. The sequences containing M2 = 18 are further differentiated from those of M2 = 29 by monomer 18's inability to couple to the sterically encumbered monomer 10. This was suggested by the absence of tags at the C sites and was confirmed by resynthesis and isolation of the single truncated sequence Ac-18-1-PLPPLP-NH_2.

Ligands 1A (bead 1), 1B, and 5 (derived from beads 3−4 and 8−11, respectively) were resynthesized on solid phase, cleaved, and individually analyzed for their ability to bind to the Src SH3 domain by a fluorescence perturbation assay (Figure 3).[4] 1A, 1B, and 5 were shown to be within 1 order of magnitude of the highest affinity SH3 ligands known (K_d = 3.4 μM, K_d = 6.6 μM, K_d = 11 μM, respectively). Truncated ligands (1C, 2, 3, 4) were synthesized, and their affinities for Src SH3 domain were determined in order to evaluate the contributions of individual monomer elements. The K_d's of 1C and 2 demonstrate that the primary amine functionality of monomer 29 is not essential for binding, while the remaining portion of the benzyl substituent is necessary. The importance of monomer 29 was further verified upon analysis of truncation ligands 3 and 4. Ligand 3 shows only a small decrease in affinity, while truncation ligand 4 binds with >70-fold less affinity than ligand 1A. To investigate the specificity of these ligands for Src SH3 domain relative to a homologous SH3 domain, we measured the binding affinity of ligand 1A to the SH3 domain in the p85 component of PI3K. These studies revealed that 1A is the most selective ligand known (52-fold) for Src SH3 domain over PI3K SH3 domain (K_d = 170 μM).

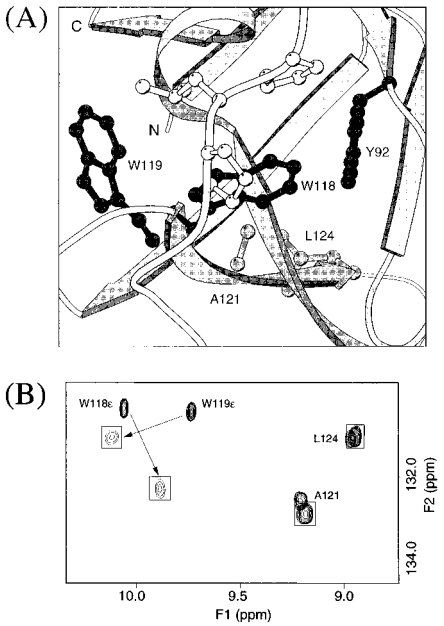


Figure 4. (A) Ribbon drawing of Src SH3 domain complexed to peptide ligand (Ac-VSLARRPLPPLP-NH_2) solved by NMR spectroscopy[8] and illustrating the positioning of W118 and W119 in the specificity pocket. (B) NMR titration of ligand 1A with ^{15}N-labeled Src SH3 domain (1:1 complex) showing large chemical shift perturbations (boxed) of the indolic NH's of W118 and W119 in the specificity pocket compared to SH3 domain alone. Note that the resonances for residues not in the binding pocket are relatively unperturbed.

Preliminary structure determinations of the Src SH3−ligand complexes involving ligands 1A, 1B, and 5 using uniformly ^{13}C/^{15}N labeled Src SH3 and multidimensional NMR confirm that the structure-based library design strategy provided ligands that bind with the anticipated orientation, and that place the non-peptide elements within the specificity pocket (pocket 3, Figure 1) of the SH3 domain. For example, when the chemical shift perturbations of ^{15}N−H cross peaks in the Src SH3−1A complex (relative to Src SH3 alone) were determined in an HSQC experiment, only the NH cross peaks belonging to the residues at the binding site showed large changes in chemical shift and intensity (Figure 4). These perturbations include those of the indolic NH's of tryptophans 118 and 119, both of which reside in the specificity pocket.

We have demonstrated that protein structure-based combinatorial chemistry is an effective method for the identification of novel ligands containing non-peptide binding elements, in this case directed to the specificity pocket of SH3 domains. The creation of new libraries using structural variants of monomers 1, 29, and 7 should allow the discovery of ligands with increased binding affinity and selectivity, and with the ability to penetrate cells so that the intracellular function of their protein targets can be ascertained. The strategy used in this study is expected to be applicable to the discovery of ligands to proteins in general.

Acknowledgment. We thank Dr. Benjamin Miller for assisting in the development of the screening protocol. J.K.C. was supported by fellowships from the NSF and The American Chemical Society Division of Organic Chemistry. A.P.C. and L.F.D.-S. were supported by the Howard Hughes Medical Institute (HHMI). S.L.S. is an Investigator at the HHMI.

Supporting Information Available: Descriptions of syntheses, screening, and reagents (6 pages). This material is contained in many libraries on microfiche, immediately follows this article in the microfilm version of the journal, can be ordered from the ACS, and can be downloaded from the Internet; see any current masthead page for ordering information and Internet access instructions.

(14) Madden, D.; Krchnák, W.; Lebl, M. *Perspect. Drug Discovery Des.* **1994**, *2*, 269−285.

JA953750V