Studying Receptor−Ligand Interactions Using Encoded Amino Acid Scanning†

Julio A. Camarero, Brenda Ayers, and Tom W. Muir*

Laboratory of Synthetic Protein Chemistry, The Rockefeller University, 1230 York Avenue, New York, New York 10021

Received December 19, 1997; Revised Manuscript Received March 26, 1998

ABSTRACT: A novel technique is described that allows the synthesis, functional analysis, and quantitative readout of defined arrays of polypeptide analogues in aqueous solution. Key to this approach is the use of a simple encoding−decoding system in which a unique Fmoc-amino acid tag is covalently attached to the C terminus of each member of a molecular array through a selectively cleavable bond. These tags can be cleanly removed from the molecules they encode, allowing single-step characterization and quantification of the entire mixture by HPLC. The utility of this technique is illustrated through the preparation of an array of proline-rich sequences based on the exchange factor C3G, one of the natural ligands of the N-terminal SH3 domain from the proto-oncogene, c-Crk. The array was designed to systematically modify those residues within the C3G peptide ligand thought to make key interactions with the c-Crk SH3 domain. Using competition binding experiments, it was possible to determine the relative ED_{50} values for the entire array of molecules simultaneously. These studies revealed that in order to maintain optimal binding to the SH3 domain, the P-3 side chain of the ligand must be positively charged and the P-0 side chain must be hydrophobic and extend beyond the γ-carbon. The excellent correlation between these relative ED_{50} values and a series of relative K_d values determined from individual peptides suggests that this approach may be useful in determining, in a parallel fashion, the relative biological activities of arrays of polypeptides.

The relationship between peptide or protein structure and function is of significant interest to biological chemists. Systematic investigation into the molecular basis of peptide/protein function typically involves altering the chemical structure of the molecule, followed by evaluation of the effect of this modification on structure and biological activity. Site-specifically modified polypeptides can be generated either by recombinant DNA-based techniques (*1*) or, in the case of smaller systems, by chemical synthesis (*2*). However, the time and effort required to modify the entire covalent structure of a peptide/protein using either of these systematic techniques is often prohibitive for many researchers. Recently, the development of combinatorial library approaches has provided a rapid means for creating chemical diversity in biopolymers (*3*). Several techniques for the chemical synthesis of combinatorial peptide libraries have been developed, most notably, spatially addressable parallel synthesis strategies (*4*−*7*) and split-resin strategies (*8*−*10*). Furthermore, the use of combinatorial oligonucleotide synthesis in conjuction with expression in bacteria (*11*) or on the surface of phage (*12*) has provided a powerful and complemetary method of generating large peptide and protein libraries.

While these synthetic and biosynthetic combinatorial strategies are suited for the identification of the new "lead" compounds, they are considerably less useful in studying, in any rigorous manner, the structure−activity relationships in peptides and proteins. The latter endeavor usually requires information on the relative effects on structure or function of performing a series of defined modifications over a region of the molecule (*13*). Extracting such data from a large combinatorial library is impossible due to the complexity of the mixture. Thus, the screening/selection steps these strategies depend on identify only the "fittest" members of the library, with no information being obtained on the remainder of the mixture.

The recently described protein signature analysis approach attempts to combine the synthetic advantages associated with combinatorial techniques with the detailed information made available through the systematic modification of peptide or protein structure (*14*, *15*). Using a modified split-resin procedure, an array of polypeptide analogues is prepared in a single synthesis, such that each member of the array is modified at a unique and defined position in the sequence (*15*). Following a functional selection step, the molecular composition of the resulting active and inactive pools is determined in a single operation using a built-in chemical decoding system in combination with a mass spectrometric readout. Protein signature analysis provides a means of systematically mapping protein−protein interactions and has been used to study the molecular contacts between an SH3[1] domain and its cognate proline-rich ligand (*14*).

Despite the potential of protein signature analysis for studying structure−activity relationships in peptides and proteins, a significant constraint is placed on the technique by its reliance on mass spectrometry for the readout step. This is due to the qualitative nature of mass spectrometric

† This research was supported by the Rockefeller University, a Ministerio de Educación y Ciencia of Spain Postdoctoral Fellowship (J.A.C.), the Pew Scholars Program in the Biomedical Sciences (T.W.M.), and the U.S. National Institutes of Health (GM55843-01, T.W.M).

* Address correspondence to this author at the Laboratory of Synthetic Protein Chemistry, The Rockefeller University, Box 223, 1230 York Ave., New York, NY10021 [telephone (212) 327-7368; fax (212) 327-7358; e-mail muirt@rockvax.rockefeller.edu].

S0006-2960(97)03122-X CCC: $15.00
Published on Web 04/29/1998

analysis of polypeptides (*16*), which leads to an inability to make quantitative comparisons between individual members of a mixture with respect to their relative biological activities. Since this type of information is often crucial to understanding structure−activity relationships, protein signature analysis is limited in the type of questions it can address.

In the present work, we describe a novel technique (termed "encoded amino acid scanning") that allows the quantitative readout of defined arrays of polypeptide analogues in solution (Figure 1). The first step in this approach involves the solid-phase synthesis of an encoded array of modified polypeptides. Each member of the synthetic array contains a defined modification, the position and chemical nature of which are encoded by a unique Fmoc-amino acid tag linked to the C terminus of the molecule through a thioester bond. Following cleavage from the solid support, the soluble array of encoded peptides is subjected to functional selection, giving rise to pools of various activities. The final step involves the identification and quantification of the compounds present in the different pools. This is achieved in a single operation by chemoselectively removing the tags from the pool of polypeptides and then passing the entire cleavage mixture (peptide + tags) down a reversed-phase HPLC column. Here we take advantage of two essential features of the Fmoc-amino acid tags, namely, each tag has a characteristic retention time allowing the entire mixture to be cleanly resolved, and the presence of the Fmoc group on the amino acids allows us to monitor at a wavelength (300 nm) unique to the tags, thereby greatly simplifying the readout. Moreover, because there is a precise one-to-one ratio between tag and peptide analogue, it is possible to quantify the molar composition of the mixture.

The utility of the encoded amino acid scanning method is illustrated through the preparation of an encoded library of peptides designed to explore the binding interaction between the N-terminal SH3 domain from the proto-oncogene c-Crk and its polyproline ligand derived from the exchange factor C3G (*17*).

MATERIALS AND METHODS

Reagents and General Methods. Boc-amino acids were obtained from Novabiochem (San Diego, CA) or Bachem (Torrance, CA). HBTU, Pam-resins, and 4-methylbenzhydrylamine-resin were obtained from Richelieu Biotechnologies (Montreal, Canada), Applied Biosystems (Foster City, CA), and Peninsula Laboratories (Belmont, CA), respectively. *N*,*N*-Dimethylformamide and HPLC grade acetonitrile was purchased from Fisher. Trifluoroacetic acid was purchased from Halocarbon (River Edge, NJ). HF was purchased from Matheson Gas. All other reagents were obtained form Aldrich Chemical Co.

RP-HPLC. Analytical HPLC was performed on a Hewlett-Packard 1100 series instrument with diode array detection.

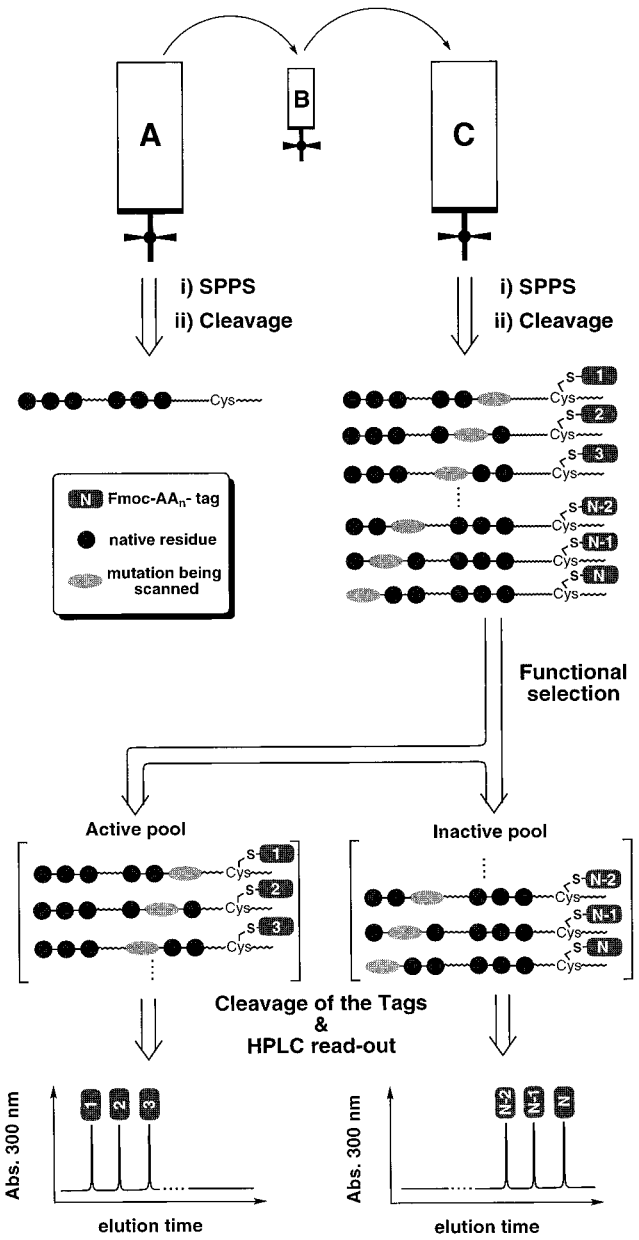


FIGURE 1: Principle of the encoded amino acid scanning approach. Step 1: Total chemical synthesis is used to generate an encoded array of peptide molecules derived from a single parent sequence. Step 2: The array is subjected to functional selection resulting in the generation of pools of various activities. Step 3: Each pool is chemically decoded by selectively removing the Fmoc-amino acid tags. The composition of the mixture is then read-out using reverse-phase HPLC.

Analytical runs were performed on a reversed-phase Vydac C_{18} column (5 μm, 4.6 × 150 mm) at a flow rate of 1 mL/min. Preparative HPLC was performed on a Waters DeltaPrep 4000 system fitted with a Waters 486 tunable absorbance detector using a Vydac C_{18} column (15−20 μm, 50 × 250 mm) at a flow rate of 50 mL/min. All runs used linear gradients of 0.1% aqueous TFA (solvent A) versus 90% acetonitrile plus 0.1% TFA (solvent B).

Mass Spectrometry. Electrospray mass spectrometry (ESMS) was routinely applied to all synthetic peptides and components of reaction mixtures. ESMS was performed on a Sciex API-100 single-quadrupole electrospray mass spectrometer. Calculated masses were obtained using the pro-

[1] Abbreviations: Boc, *tert*-butyloxycarbonyl; DBU, 1,8-diazabicyclo-[5.4.0]undec-7-ene; DIEA, diisopropylethylamine; DMF, dimethylformamide; EDTA, ethylenediaminetetraacetic acid; ESMS, electrospray mass spectrometry; Fm, 9-fluorenylmethyl; Fmoc, 9-fluorenylmethoxycarbonyl; HBTU, 2-[1*H*-benzotriazolyl]-1,1,3,3-tetramethyluronium hexafluorophosphate; (RP)-HPLC, (reversed-phase) high-performance liquid chromatography; Pam, phenylacetamidomethyl; SH3, Src homology 3; SPPS, stepwise solid-phase peptide synthesis; TFA, trifluoroacetic acid; Tos, tosyl (4-methylbenezensulfonyl). Standard IUPAC single- and triple-letter codes for amino acids are used throughout.

gram MacProMass (Sunil Vemuri and Terry Lee, City of Hope, Duarte, CA).

Solid-Phase Peptide Synthesis. All polypeptides were manually synthesized according to the in-situ neutralization/HBTU activation protocol for Boc SPPS (*18*). Peptide ligands were assembled on 4-methylbenzhydrylamine-resin, and the Crk-N SH3 domain was synthesized on a Boc-Arg-(Tos)-OCH_2-Pam resin. Global deprotection and cleavage from the support were achieved by treatment with HF containing 4% v/v *p*-cresol, for 1 h at 0 °C. Following removal of the HF, crude peptide products were precipitated and washed with anhydrous cold Et_2O before being dissolved in degassed aqueous acetonitrile (10−50%) and lyophilized. Polypeptides were purified by preparative HPLC, and, in all cases, polypeptide composition and purity were confirmed by ESMS and analytical HPLC.

Synthesis of the Encoded Peptide Library. The procedure makes use of three reaction vessels (A, B, and C). For the C3G library, the sequence KRXC(Fm)X (where X = Ahx and Fm = 9-fluorenylmethyl) was first synthesized in reaction vessel A on a 0.3 mmol scale. At this point the N^{α}-Boc-protected peptide-resin was suspended in 5 mL of DMF, and 6 × 333 μL aliquots (6 × 20 μmol) were transferred to six small reaction vessels (B_1, B_2, . . ., B_6), where they were treated in parallel as follows. The Cys-(Fm) was deprotected with 20% piperidine, 2% DBU, and 2% 2-mercaptoethanol in DMF (3 × 5 min.). For each reaction vessel, B_n, a unique Fmoc-AA_n-OH tag (0.5 mmol, activated in situ with HBTU and DIEA in DMF) was coupled to the free thiol group on the peptide-resin. The resins in reaction vessels A and B_n were then treated with TFA (2 × 1 min) to remove the N^{α}-Boc group, and the natural residue (vessel A) and the mutated residue (vessel B_n) were attached. The resins present in reaction vessels B_n were then transferred to reaction vessel C, and the synthesis was continued, in parallel, in reaction vessels A and C until residue Pro^6. At this point the resin in vessel A was again aliquoted, tagged, and analogued as described above. Finally, the remaining residues were added to the peptide-resins in reaction vessel C, after which the encoded peptide library was deprotected and cleaved from the resin with HF.

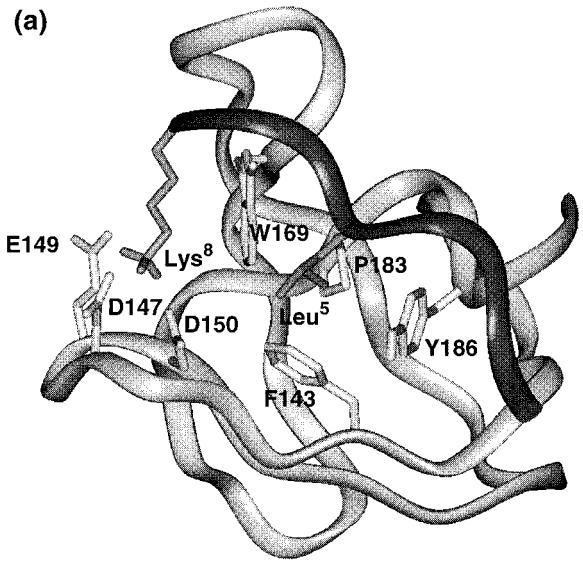
Cleavage and HPLC Analysis of the Fmoc-Amino Acid Tags. The array of C3G peptides was chemically decoded as follows. The lyophilized and desalted array was dissolved in 30 mM $Hg(OAc)_2$, 30% MeCN, and 5% HOAc buffer (150 μL) for 18−20 h. The cleavage mixture (100 μL) was then analyzed by analytical RP-HPLC, monitoring at 300 nm and using the following linear gradient system: 32% B isochratic over 2 min, 32−42% B over 45 min, 42−47% B over 3 min, and 47−50% B over 20 min. Quantification of the different Fmoc-amino acid tags was carried out by integrating the area of the corresponding peaks at 300 nm.

Synthesis of the Crk-N Affinity Column. The polypeptide H-Cys-Ahx-[c-Crk(134−191)]-OH was synthesized and purified using the general procedures described above. The covalent structure of the purified product was characterized by electrospray mass spectrometry [observed mass = 7178.8 ± 2.8 Da, predicted (average) = 7178.1 Da]. H-Cys-Ahx-[c-Crk(134−191)]-OH (≈1 mg) was dissolved in 50% MeCN in buffer A (300 μL) and mixed with Sulfolink resin (Pierce) preswollen in 50 mM Tris·HCl and 5 mM EDTA buffer at pH 8 (400 μL). The reaction was left with occasional shaking for 2 h. Unreacted iodoalkyl groups were capped by treatment with 2% 2-mercaptoethanol in the previous buffer for 1 h. The functional substitution of the column (i.e., the maximum amount of ligand that the column is able to bind) was determined by HPLC to be 50 μM ± 10%.

Affinity Selection of the C3G Peptide Library. The C3G peptide library (8 nmol) was dissolved in 20 mM sodium phosphate and 50 mM NaCl buffer at pH 7.2 (100 μL) and loaded onto a 0.2 mL Crk-N-column pre-equilibrated with the same buffer. After 1 h, the nonspecifically bound material was washed off the column with 20 mM sodium phosphate and 150 mM NaCl buffer at pH 7.2 (1.6 mL). The bound members of the library were then eluted with 50% MeCN in buffer A (800 μL). Both the bound and nonbound fractions were desalted using a disposable C_{18} column (Sep-Pak, Waters) and lyophilized. The amino acid tags were then cleaved and analyzed by HPLC. The competition elution experiments were carried out basically as above, with the only exception that the C3G peptide library (8 nmol) was loaded in the presence of increasing amounts of wild-type peptide ligand (0.1, 1, 3, 5, 9, 18, 27, 54, 81, 181, 243, 486, and 1140 nmol), incubated for 1 h, and then washed/eluted as before. ED_{50} values for the competitive elution experiments were extracted from nonlinear regression curves generated using the program PRISM (GRAPHPAD Software Inc., San Diego, CA).

Fluorescence-Based Ligand Binding Assays. The equilibrium dissociation constants for the c-Crk SH3-ligand interactions were measured using a fluorescence-based titration assay. Experiments were conducted at 25 °C in a stirred 1 cm path length cell using a Spex Fluorolog-3 spectrofluorometer. Excitation was at 298 nm with a 5 nm slit, and the fluorescence emission was monitored at 347 nm through a 5 nm slit. In all cases, the protein concentration was kept at 0.1 μM in a buffer containing 20 mM sodium phosphate and 150 mM NaCl at pH 7.2. The dissociation constants were determined by changes in the fluorescence of the protein solution upon addition of the corresponding peptide ligand at defined concentrations; calculations were made assuming formation of a 1:1 complex (*19*).

Molecular Modeling. Molecular minimizations were performed on an Indigo 2 Silicon Graphics workstation using the Insight II program. All calculations and minimizations employed a cvff force field. The starting structure was built by adding the necessary hydrogen atoms at pH 7.0 to the crystal structure of the c-Crk SH3−N domain complexed with the C3G derived peptide [entry 1cka in the Brookhaven Protein Data Bank (*20*)]. The side chains of residues Asp^{147}, Glu^{149}, and Asp^{150} in the SH3 domain were kept in the carboxylate form to allow the interaction with the ϵ-NH_3^+ group of the Lys^8 residue of the C3G peptide. Mutations in position 5 of the ligand were carried out by replacing the original Leu residue with the corresponding residues. Minimized structures were generated by steepest descent minimization (1000 iterations) in a vacuum using a distance-dependent dielectric of 4*Rij*. The stabilization energies were calculated by comparing the energy of the corresponding complex with the energies of the SH3 domain and the ligand alone. The energies were calculated with a cutoff of 100 Å using van der Waals and electrostatic energy parameters.



(b)

Name	Sequence										S-Tag		Fmoc-AA-OH tag
wt	P	P	P	A	**Leu**	P	P	**Lys**	K	R	X	C	X	Ser
8a	-	-	-	-	**-**	-	-	**Orn**	-	-	-	-	-	Asp
8b	-	-	-	-	**-**	-	-	**Gln**	-	-	-	-	-	Gly
8c	-	-	-	-	**-**	-	-	**Arg**	-	-	-	-	-	Ala
8d	-	-	-	-	**-**	-	-	**Cit**	-	-	-	-	-	Aib
8e	-	-	-	-	**-**	-	-	**Nle**	-	-	-	-	-	Abu
5a	-	-	-	-	**Ala**	-	-	**-**	-	-	-	-	-	Val
5b	-	-	-	-	**Abu**	-	-	**-**	-	-	-	-	-	Nva
5c	-	-	-	-	**Val**	-	-	**-**	-	-	-	-	-	Phg
5d	-	-	-	-	**Nva**	-	-	**-**	-	-	-	-	-	Ile
5e	-	-	-	-	**Ile**	-	-	**-**	-	-	-	-	-	Leu
5f	-	-	-	-	**Nle**	-	-	**-**	-	-	-	-	-	Phe
5g	-	-	-	-	**Cha**	-	-	**-**	-	-	-	-	-	HPhe
5h	-	-	-	-	**Phe**	-	-	**-**	-	-	-	-	-	Phe(pCl)

FIGURE 2: Design of the C3G peptide array: (a) crystal structure of the Crk-N domain complexed with the C3G peptide (*20*) showing the key intermolecular contacts; (b) members of the C3G peptide array and the corresponding Fmoc-AA-OH tags.

RESULTS

Probing a Ligand−Receptor Interaction Using Encoded Amino Acid Scanning. The N-terminal Src homology type 3 (SH3) domain of the adaptor protein c-Crk (herein referred to as Crk-N) binds to a short proline-rich peptide sequence derived from the GDP−GTP exchange factor C3G (*17*). The molecular basis of the Crk-N/C3G interaction has been studied both by X-ray crystallography (*20*) and by protein signature analysis (*14*). These studies suggest that the P-0 and P-3 residues within the peptide ligand make key interactions with the SH3 domain (Figure 2a). In the present work, we were interested in using the encoded amino acid scanning technique to study, in a quantitative fashion, the relative contributions of the P-0 and P-3 ligand side chains to overall binding affinity.

A focused array of C3G peptides was designed in which the residues corresponding to positions P-0 and P-3 were systematically replaced by a series of natural and unnatural amino acids (Figure 2b). In position P-3, the natural lysine residue was replaced by arginine, ornithine, citrulline, glutamine, and norleucine. The Arg and Orn mutations were designed to explore the effect of altering the geometry of the electrostatic/H-bonding interactions in the P-3 pocket. It was hoped that mutation to the uncharged residues Cit and Gln would allow the electrostatic versus H-bonding contributions to binding to be resolved. Finally, mutation to the isosteric Nle was expected to provide information on whether the hydrophobic interaction with Trp^{169} makes any significant contribution to the binding energy. Eight mutations were made at the P-0 position with the naturally occurring leucine residue being replaced by alanine, 2-aminobutyric acid, valine, norvaline, isoleucine, norleucine, cyclohexylalanine, and phenylalanine. These mutations were chosen to explore the effect on the binding interaction of systematically varying the length and/or the branching pattern of the hydrophobic P-0 side chain. In addition to the 13 mutant peptides described above, a positive control, namely, the wild-type peptide, was also included in the library.

Selection of Fmoc-Amino Acid Tags. As previously stated, our encoding−decoding system makes use of Fmoc-amino acid tags that are attached to the C terminus of the peptide analogue through a cleavable thioester bond. There are a large number of commercially available Fmoc-amino acids, both natural and unnatural, many of which are suitable for our purposes. The amino acid tags used in the present work were chosen on the basis of the following criteria. First, their side-chain functionalities should not interfere with the thioester linkage (Cys, His, and Glu/Asp residues were excluded for this reason). Second, any side-chain functionalities must be suitably protected [e.g., Fmoc-Ser(Bzl)-OH] to avoid unwanted branching during the Boc-based SPPS. Finally, the entire mixture of tags must be well resolved by analytical RP-HPLC. The set of Fmoc-amino acids used to encode our C3G library is listed in Figure 2b, and as shown by Figure 3b, this entire set of tags was resolved by RP-HPLC.

Synthesis of the C3G Peptide Array. The chemical synthesis of the 14-member C3G encoded peptide array was carried out using a variation on the previously described split-resin method (*15*). Importantly, a Cys residue was placed at the C terminus of each molecule to allow the attachment of the corresponding Fmoc-amino acid tag through an alkyl thioester bond.[2] Two 6-aminohexanoic acid (Ahx) spacer residues were also introduced on either side of this Cys residue to minimize potential interactions between the tag and the binding site, as well as to facilitate rapid tag introduction during the solid-phase synthesis.[3] The sulfhydryl group of this C-terminal Cys residue was protected with the base-labile 9-fluorenylmethyl (Fm) group. This allowed the orthogonal introduction of the Fmoc-amino acid tags through the desired thioester linkage at the appropriate point in the synthesis.

Preparation of the encoded peptide array involved initiating peptide synthesis within a standard SPPS reaction vessel (vessel A). At the appropriate point in the chain assembly, n equimolar aliquots of peptide-resin were removed from vessel A (where n = the number of mutations being made at a given position) and each aliquot was transferred to a

[2] Alkyl thioesters are stable to both the chain assembly and cleavage conditions used in the Boc-solid-phase peptide synthesis conditions (*21*).

[3] Preliminary studies indicated that when the Cys(Fm) residue is placed at the immediate C terminus, the deprotection of the Fm group is slow, giving rise to incomplete S-acylations. This problem was solved by inserting an aminocaproate group between the Cys and the resin.

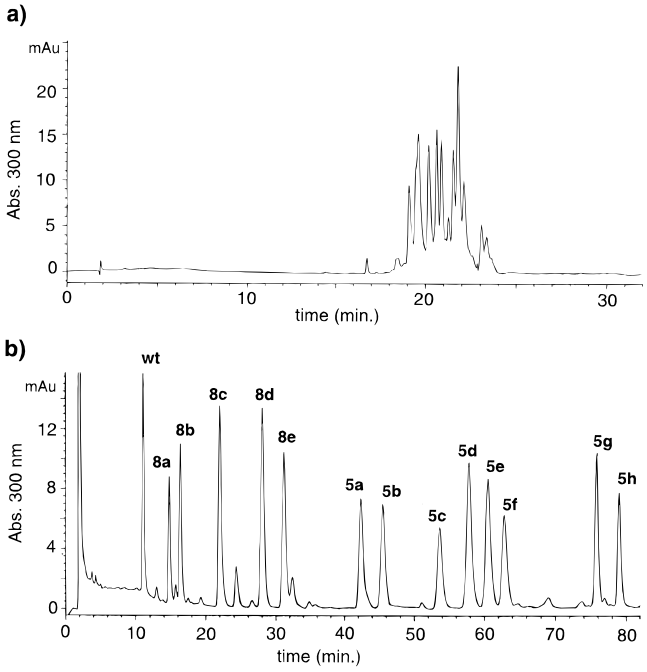


FIGURE 3: Readout of the C3G peptide array: (a) RP-HPLC chromatogram (gradient 0−73% buffer B over 30 min) of the crude encoded peptide array following HF treatment; (b) RP-HPLC chromatogram of the same peptide array chemically decoded with $Hg(OAc)_2$ (see Materials and Methods for gradient). Each peak corresponds to a different Fmoc-AA-OH tag and has been labeled with C3G peptide that it specifically encodes (see Figure 2b). Such a readout is straightforward since the theoretical retention time of each tag is known.

separate reaction vessel, B_n. Here, both the tag and the modification it encodes were introduced. The resulting modified, tagged peptide-resin aliquots were then combined and transferred to vessel C, where the remainder of the amino acids in the sequence were attached. Upon completion of the chain assembly, the desired encoded array was present as a mixture within vessel C, while vessel A contained peptide-resin corresponding to the underivatized parent sequence. Note that because of the split-resin protocol employed, the material in vessel A is a useful internal control for the quality of the synthesis. In the present study, cleavage and analysis of the peptide generated in vessel A gave a single component as indicated by HPLC and ESMS analysis.

Readout of the Peptide Array. The encoded C3G library was cleaved from the solid support by treatment with HF. Analysis of the resulting peptide array by analytical HPLC afforded, as expected, a complex chromatogram with several overlapping peaks (Figure 3a). Each HPLC peak was individually analyzed by ESMS and chemically decoded by treatment with $Hg(OAc)_2$. These studies confirmed that each of the expected tagged peptides was present in the array and indicated that the mixture was essentially free of peptide contaminants (e.g., deletion, termination, or alkylation products).

Straightforward compositional readout of the mixture was achieved by chemically cleaving the Fmoc-amino acid tags from the peptides and then subjecting the entire cleavage mixture to analytical HPLC (Figure 3b). In practice, this chemical decoding step was performed by treating a small aliquot of the library with $Hg(OAc)_2$ at pH $\approx$4.[4] The known avidity of mercury for sulfur atoms (*22*) renders the thioester linkage sufficiently electrophilic to be cleanly hydrolyzed at this pH. Significantly, these conditions also ensured that the Fmoc group was not removed from the tags, thereby allowing 300 nm detection to be employed during the HPLC analysis of the cleavage mixture (note, only the tags are visible at this wavelength). The composition of the decoded mixture of Fmoc-amino acids could then be simply readout from this HPLC chromatogram since the retention time of each Fmoc-amino acid was already known.

Functional Selection of the C3G Array. To facilitate functional selection of the C3G peptide library, a Crk-N affinity column was prepared. This involved chemoselectively linking a synthetic Crk-N polypeptide (residues 134−191 of c-Crk with the additional residues Cys-Ahx at the N terminus[5]) to an iodoalkyl−agarose support. Preliminary results indicated that these Crk-N-agarose beads were able to bind either the tagged or untagged wild-type peptide in a specific manner.[6] Moreover, a series of solution-based studies involving a spectroscopic ligand binding assay indicated that the dissociation constant (K_d) for the Crk-N/C3G interaction was essentially unaltered by the presence of an Fmoc-AA-OH tag at the C terminus of the ligand (PPPALPPKKRXYX, $K_d = 0.44 \pm 0.01$ μM; PPPALPPKKRXC[Fmoc-Ala]X, $K_d = 0.63 \pm 0.07$ μM, where X = aminohexanoic acid).

Functional selection of the 14-member array of C3G analogues was initially carried out by loading the mixture onto the Crk-N affinity column. The column was then extensively washed to remove nonspecifically bound components, and the remaining specifically bound C3G peptides were eluted by denaturing the SH3 domain. Comparison of the HPLC profiles obtained from the chemically decoded nonspecifically (Figure 4a) and specifically (Figure 4b) bound pools revealed a large difference in the tolerance of positions P-0 and P-3 to systematic mutation. A positively charged side-chain group appeared to be absolutely required in the P-3 position since mutation to anything other than Arg or Orn resulted in a significant loss of binding. In contrast, the binding activity of the ligand appeared to be much less affected by mutations in the P-0 position, since any hydrophobic side-chain structure was accepted provided it extended beyond the β-carbon (only the Ala mutation resulted in an inactive ligand).

Although useful in identifying gross differences in activity, the above selection procedure is not well suited to the analysis of more subtle changes in binding affinity that may be present across the peptide array. Moreover, it is conceivable that peptides which appear active in the above procedure are in fact interacting with a region of the receptor other than the binding site (i.e., they are binding nonspecifically). To address both of these issues, a series of competition binding studies were performed in which increasing concentrations of wild-type peptide were incubated with the peptide mixture during the column loading step. The

[4] Preliminary results indicated that quantitative cleavage of Cys-S-thioesters by $Hg(OAc)_2$ is achieved in 18 h as monitored by HPLC.

[5] The Cys residue was added to the N terminus to chemoselectively attach the protein domain to a Sulfolink (iodo-*n*-hexyl−agarose) support through a thioether bond. The Ahx residue was used as a spacer.

[6] No specific interactions were observed when the wild-type C3G peptides were passed over a control affinity column (2-mercaptoethanol-capped Sulfolink matrix).

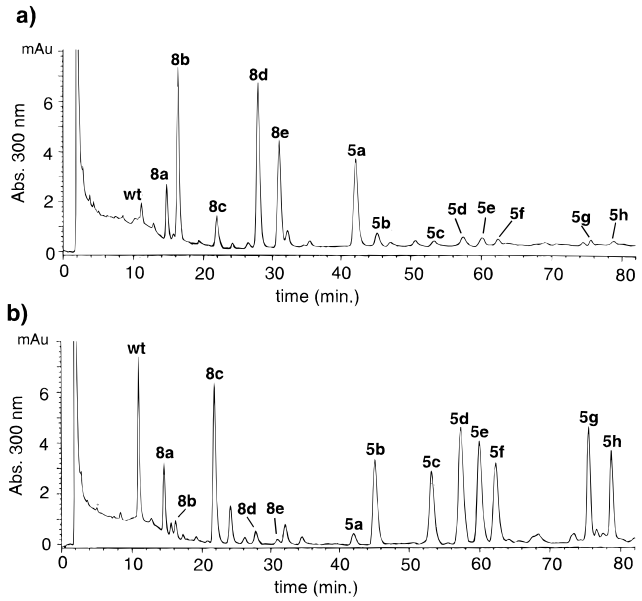


FIGURE 4: Functional selection of the C3G peptide library on a Crk-N affinity column. Shown are the RP-HPLC chromatograms of decoded nonspecifically (a) and specifically (b) bound fractions.

specifically and nonspecifically bound pools were then separated using the same washing and elution protocols as before. The amount of each mutant peptide competed from the column by a given concentration of wild-type ligand was then determined by simply integrating the decoded peaks in the appropriate HPLC chromatogram. Analysis of these data revealed that all of the active C3G peptides were competitively desorbed from the affinity column by the wild-type ligand (Figure 5). Each desorption data set obeyed the classical sigmoidal profile for competitive desorption, allowing a parameter related to ED_{50}[7] to be calculated for each of the active C3G peptides (Figure 6).

Consistent with our initial findings, the absence of a positively charged side chain in the P-3 position (Lys^8) resulted in a complete loss of binding activity. Thus, relative ED_{50} values could be determined only for the Arg and Orn mutants, revealing losses of binding affinity of 8- and 13-fold, respectively, compared to the wild-type ligand. Of the mutations in the P-0 position (Leu^5), only the Ala substitution appeared to wipe out binding to the Crk-N column, again consistent with our initial observations. Next in line was the Abu mutation, which had a ≈10-fold reduction in its relative ED_{50} value compared to wild-type. The remaining P-0 mutations resulted in only small to moderate reductions (2−5-fold) in binding affinity.

It should be stressed that the relative ED_{50} values shown in Figure 6 were all determined simultaneously and that this type of quantitation is to our knowledge unprecedented using a library of polypeptide molecules. Because of this, it was important to confirm, through additional experiments, that these relative binding activities were representative. Several members of the C3G peptide array were individually synthesized, and in each case the dissociation constant for the Crk-N interaction was determined using an established

[7] ED_{50} is defined as the amount of unlabeled ligand yielding 50% displacement of the labeled ligand from the column. The ratio of two ED_{50} values provides a rough measure of the relative binding affinities of two different ligands (*23*).

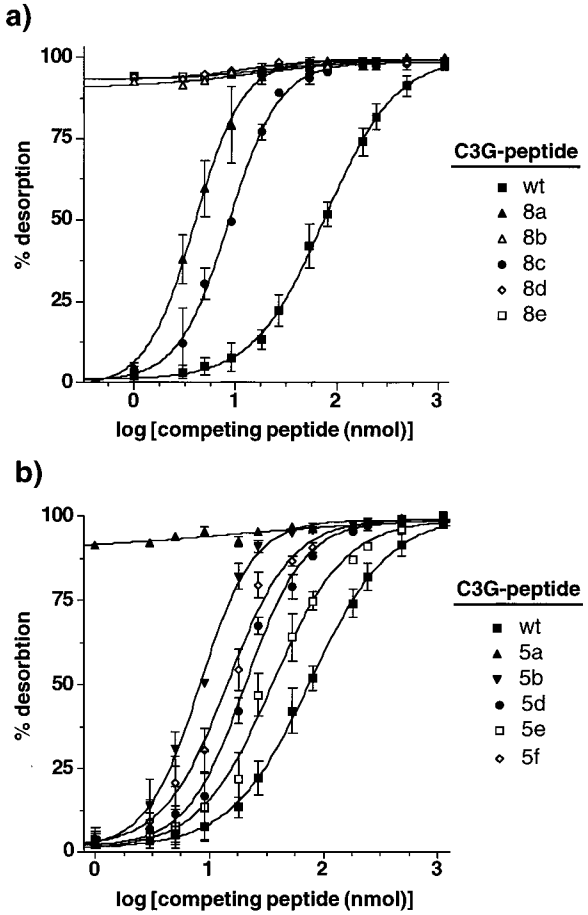


FIGURE 5: Competitive desorption of the C3G peptide library from the Crk-N-affinity column using increasing concentrations of wild-type peptide ligand. Shown are the elution profiles for the members of the library mutated in the P-3 position (a) and P-0 position (b). For reasons of clarity, the desorption curves for 5c, 5g, and 5h are not shown; however, all follow the expected sigmoidal profile. Data points were obtained by integrating the HPLC peaks corresponding to the various Fmoc-AA-OH tags at a given concentration of competing agent from two independent experiments.

fluorescence-based ligand binding assay (*19*). As can be seen from Figure 6, there is an extremely good correlation between the relative K_d values determined using individual compounds and the relative ED_{50} values obtained from the competition binding experiments on the intact library. Thus, we conclude that the quantitative binding data obtained on the library are a reliable measure of the relative effects of the mutations on binding affinity.

DISCUSSION

A novel chemical strategy, termed encoded amino acid scanning, is described that allows the quantitative readout of defined arrays of peptides in solution. As with the previously reported protein signature analysis technique (*14*, *15*), this new approach is designed to explore the structure−activity relationships in peptides and small protein domains. Encoded amino acid scanning overcomes the two principle limitations of protein signature analysis, which are the incorporation of non-amide bonds within the polypeptide sequence being studied and the use of a mass spectrometric readout. The former of these is required to facilitate site-specific chemical cleavage but leads to the unavoidable (and often undesirable) backbone engineering of the system.

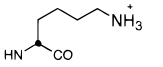
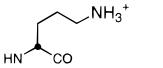
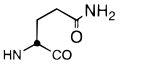
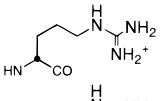
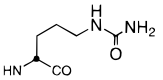
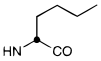
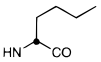
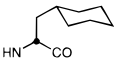
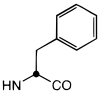
P-3 POSITION				P-0 POSITION			
RESIDUE	STRUCTURE	$ED_{50,rel}$	$K_{d,rel}$	RESIDUE	STRUCTURE	$ED_{50,rel}$	$K_{d,rel}$
Lysine (Lys)		1.0	1.0	Leucine (Leu)		1.0	1.0
Ornitine (Orn)		13 ± 2	14.5 ± 0.7	Alanine (Ala)		nb	20 ± 1
Glutamine (Gln)		nb	39 ± 3	2-Aminobutyric Acid (Abu)		8 ± 1	7.5 ± 0.4
Arginine (Arg)		8 ± 1	7.5 ± 0.3	Norvaline (Nva)		3.6 ± 0.7	3.2 ± 0.2
Citrulline (Cit)		nb	nd	Norleucine (Nle)		5 ± 1	nd
Norleucine (Nle)		nb	nd	Valine (Val)		2.1 ± 0.4	3.2 ± 0.1
				Isoleucine (Ile)		2.3 ± 0.5	3.9 ± 0.3
				Cyclohexyl-alanine (Cha)		2.4 ± 0.6	nd
				Phenylalanine (Phe)		1.4 ± 0.5	nd

FIGURE 6: Relative affinities of the members of the C3G library for the c-Crk SH3 domain. Both $ED_{50,rel}$ and $K_{d,rel}$ were calculated relative to the wild-type of peptide. The ED_{50} values were calculated by competitive elution from the c-Crk SH3 affinity column with wild-type peptide from two independent experiments as described under Materials and Methods. The K_d values were determined by changes in the fluorescence of the protein solution upon addition of the corresponding peptide ligand at defined concentrations as described under Materials and Methods (nb, did not bind to column under conditions used; nd, not determined).

Second, mass spectrometry is a qualitative analytical technique (*16*), which means that the molar composition of a polypeptide mixture cannot be reliably extracted from a mass spectrum of that mixture. The use of quantitative ESMS for screening mixtures of enzyme inhibitors has recently been reported (*24*). However, this study involved the analysis of small molecules giving rise to only a single charge state. The situation is considerably more complex in a polypeptide library due to both the degeneracy in peptide masses and the presence of multiple charge states for each species. Consequently, quantitative comparisons cannot be made on the relative biological activities of a polypeptide library using mass spectrometry.

Encoded amino acid scanning does not require the introduction of non-amide linkages within the sequence of interest, nor does it involve the use of mass spectrometry in the readout step. Instead, each member of a defined peptide array is encoded with a unique Fmoc-amino acid tag attached to the C terminus of the molecule through a selectively cleavable bond. Following functional selection, the tags are removed en masse and the entire mixture is analyzed by analytical HPLC with 300 nm UV detection. Since there is a precise one-to-one relationship between the tag and the molecule that it encodes, and the ϵ_{300nm} is the same for each tag, it is possible to quantify the mixture. An additional advantage of our technique versus other encoding systems (*14*, *15*, *25*−*31*) is that it involves simple peptide chemistry and all of the reagents used, including the tags, are commercially available. Furthermore, the readout step relies on the use of analytical HPLC, which is available to many researchers.

As a model system for the encoded amino acid scanning approach, we studied the interaction between the N-terminal SH3 domain of the proto-oncogene c-Crk and a proline-rich peptide ligand. The cellular adaptor protein, c-Crk, contains an N-terminal SH2 domain followed by two SH3 domains (*32*). Studies on c-Crk have shown that it can bind to several cellular proteins including the GDP−GTP exchange factors SOS and C3G (*17*). The binding of c-Crk to SOS and C3G suggests that the Ras signaling pathway may be implicated in Crk transformation. Both of these exchange factors bind to the N-terminal SH3 domain of c-Crk via proline-rich sequences (*17*). Furthermore, isolated poly-Pro peptides derived from these proteins are able to bind Crk-N with high affinity (*20*), with the C3G/Crk-N interaction being one of the tightest and most specific reported for an SH3−ligand system ($K_d \approx 1-2\ \mu M$).

The crystal structure of the Crk-N/C3G peptide complex has been solved (*20*), revealing that the peptide binds to the SH3 domain in the so-called reversed [or minus (*33*)] orientation. Residues at the P-0 and P-3 positions within the peptide (Leu^5 and Lys^8, respectively, in the sequence) appear to make the key interactions that determine the orientation of the ligand with respect to the receptor (Figure

2a). The side chains of both residues are located within specific binding pockets on the surface on the SH3 domain. Specifically, the aliphatic portion of the Lys-8 side chain packs against the face of Trp^{169} in the protein, and the positively charged ϵ-amino group coordinates to three side-chain carboxylates from residues Asp^{147}, Glu^{149}, and Asp^{150}, all of which are located in the RT-loop of Crk-N. In contrast to the negatively charged P-3 pocket, the side chain of Leu-5 sits in the hydrophobic P-0 cleft defined mainly by residues Phe^{143}, Pro^{183}, Tyr^{186}, and Trp^{169} in the SH3 domain.

We were intrigued by the potential effects on function of systematically modifying the chemical structure of the P-0 and P-3 residues within the ligand. A competition binding experiment was performed in which the encoded library was incubated with the immobilized receptor in the presence of increasing concentrations of wild-type ligand. The quantitative readout unique to our approach allowed the amount of each tagged component displaced from the column by a given concentration of competing untagged wild-type ligand to be determined. Interpretation of these competition data is complicated by the fact that an individual component could be competed from the column both by the competing agent and by the other members of the peptide library. However, this self-competition within the library should be roughly constant during the experiment (since identical conditions were used throughout), allowing the desorption specific to the competing agent to be resolved. Indeed, the data sets obtained from this experiment were observed to follow the expected profile for competitive desorption (Figure 5). Analysis of each curve allowed the wild-type ED_{50}/mutant ED_{50} ratio to be determined (Figure 6). The excellent correlation between these relative ED_{50} values and the relative K_d values, measured using individual peptides, suggests that this competition binding experiment is a useful way of determining the relative biological activities of an entire array of molecules in a parallel fashion.

The Crk-N domain is known to interact ≈10-fold more tightly with a C3G-derived peptide than with the corresponding sequence from SOS (*20*). Insights into the structural basis of this selectivity have emerged from the cocrystal structures of the two complexes (*20*). In the Crk-N/C3G complex the ϵ-NH_3^+ group of Lys^8 (P-3) is tightly coordinated to three carboxylate side chains in the SH3 domain, the interaction involving both electrostatic and hydrogen bonding (Figure 2a). In contrast, the SOS-derived peptide contains an Arg at the P-3 position, which leads to a suboptimal conformation for H-bonding to the three carboxylates. Consistent with these structural studies, our results indicate that replacement of Lys^8 by either Arg (peptide 8c) or Orn (peptide 8b) results in significant decreases in binding affinity of 8- and 13-fold, respectively. The present study also reveals that a positively charged P-3 side chain is mandatory for binding to Crk-N, since neither glutamine nor citrulline, both of which are able to donate hydrogen-bonds, was able to bind the Crk-N affinity column.

The hydrophobic P-0 cleft of Crk-N appears to be relatively promiscuous in the range of side-chain structures it accepts. However, our studies do indicate that systematic truncation beyond the δ-carbon of the P-0 side chain leads to an incremental loss of binding. This is supported by the fact that when Leu^5 is replaced by alanine, 2-aminobutyric acid, and norvaline, the binding is completely lost, reduced ≈8-fold, and reduced ≈3.5-fold, respectively. Extension of the P-0 side chain beyond the γ-carbon appears to be the principal determinant for binding since altering the branching pattern of the side chain did not significantly destabilize the complex (Val, Nva, Ile, and Nle were only 2 and 5 times less active than wild-type). Indeed, not even the introduction of bulky phenylalanine or cyclohexyalanine residues at this position resulted in a significant decrease in affinity compared to the wild-type peptide. Consistent with these binding data, molecular modeling of the receptor−ligand complex indicated that the P-0 hydrophobic cleft can accommodate a range of branched structures without the introduction of unfavorable steric interactions. For example, the structures of the minimized complexes containing Phe and Cha at the P-0 position of the ligand had rms deviations of only ≈0.2−0.3 Å from the wild-type complex, and the corresponding stabilization energies were similar to that obtained in the wild-type complex. In agreement with our results, Kay et al. have recently used a phage-displayed library approach to show that position P-0 in the C3G ligand can tolerate hydrophobic residues such as Ile, Val, and Phe without significant loss in activity (*34*).

In contast to combinatorial library techniques, which serve as effective broad screening tools, the encoded peptide scanning strategy is specifically designed to allow peptide structure−activity relationships to be studied in detail. In combination with competition binding experiments, our technique can be used to determine the relative biological activities of an entire array of molecules simultaneously. This feature distinguishes the approach from all existing library-based techniques and makes the two strategies (combinatorial peptide library and focused peptide array) complementary in the type of information they provide. It is also worth noting that dozens of chemically suitable Fmoc-amino acid tags are commercially available. This, combined with the resolving power of current reversed-phase columns, means that moderately sized peptide arrays can be prepared and analyzed. For example, an array of peptides containing 30 members has recently been generated in our laboratory (J. A. Camarero et al., unpublished data).

Recent years have seen the development of several spatially addressable parallel synthesis approaches that allow the preparation of large numbers of peptides tethered to solid supports (*4*−*7*). Such approaches have proven useful for mapping short epitopes of 6−8 amino acids in length; however, synthetic limitations have prevented their application to longer peptide sequences (*4*−*7*). The encoded amino acid scanning approach relies on optimized Boc-SPPS protocols throughout and, so, allows these larger peptide systems to be studied (*2*, *14*, *18*). Indeed, we have recently applied the approach to the successful generation of an encoded array of 36-mers derived from region 4 of the *Escherichia coli* σ^{70} factor (J. A. Camarero et al., unpublished data). It is also important to note that the use of an encoded soluble mixture means that all postsynthetic operations (i.e., cleavage, workup, and functional analysis) need only be performed once. In the absence of such encoding, these manipulations must be carried out multiple times, thereby greatly increasing the time and effort associated with a project. Thus, we believe that the soluble encoded peptide array strategy described in the present work will complement the use of conventional multiple-peptide synthesis.

ACKNOWLEDGMENT

We thank Graham Cotton for useful discussions and Joanna Pavel for help in the early part of this work.

REFERENCES

1. Smith, M. (1994) *Angew. Chem., Int. Ed. Engl. 33*, 1214.
2. Muir, T. W., Dawson, P. E., and Kent, S. B.(1997) *Methods Enzymol. 289*, 266−298.
3. Janda, K. D. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 10779−10785.
4. Geysen, H. M., Meloen, R. H., and Bartelig, S. J. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 3998−4002.
5. Houghten, R. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 5131−5134.
6. Fodor, S. P., Read, J. L., Pirrung, M. C., Stryer, L., Lu, A. T., and Solas, D. (1991) *Science 251*, 767−773.
7. Frank, R. (1992) *Tetrahedron 48*, 9217.
8. Furka, A., Sebestyen, F., Asgedom, M., and Dibo, G. (1991) *Int. J. Pept. Protein Res. 37*, 487−493.
9. Houghten, R. A., Pinilla, C., Blondelle, S. E., Appel, J. R., Dooley, C. T., and Cuervo, J. H. (1991) *Nature 354*, 85−86.
10. Lam, K. S., Salmo, S. E., Hersh, E. M., Hruby, V. J., Kazmiersky, W. M. and Knapp, R. J. (1991) *Nature 354*, 82−84.
11. Reidhaar-Olson, J. F., and Sauer, R. T. (1988) *Science 241*, 53−57.
12. Smith, G. P., and Petrenko, V. A. (1997) *Chem. Rev. 97*, 391−410.
13. Cunningham, B. C., and Wells, J. A. (1989) *Science 244*, 1081−1085.
14. Muir, T. W., Dawson, P. E., Fitzgerald, M. C., and Kent, S. B. H. (1996) *Chem. Biol. 3*, 817−825.
15. Dawson, P. E., Fitzgerald, M. C., Muir, T. W., and Kent, S. B. H. (1997) *J. Am. Chem. Soc. 119*, 7917−1927.
16. Cohen, S. L., and Chait, B. T. (1996) *Anal. Chem. 68*, 31−37.
17. Knudsen, B. S., Feller, S. M., and Hanafusa, H. (1994) *J. Biol. Chem. 269*, 32781−32787.
18. Schnölzer, M., Alewood, P., Jones, A., Alewood, D., and Kent, S. B. H. (1992) *Int. J. Pept. Protein Res. 40*, 180−193.
19. Viguera, A. R., Arrondo, J. L. R., Musacchio, A., Saraste, M., and Serrano (1994) *Biochemistry 33*, 10925−10933.
20. Wu, X., Knudsen, B., Feller, S. M., Zheng, J., Sali, A., Cowburn, D., Hanafusa, H., and Kuriyan, J. (1995) *Structure 3*, 215−226.
21. Hojo, H., and Aimoto, S. (1991) *Bull. Chem. Soc. Jpn. 64*, 111−117.
22. Andreu, D., Albericio, F., Sole, N. A., Munson, M. C., Ferrer, M., and Barany, G. (1994) in *Methods in Molecular Biology, Vol. 35: Peptide Synthesis Protocols* (Pennington, M. W., and Dunn, B. N., Eds.) pp 91−169, Humana Press, Totowa, NJ.
23. Munson, P. J., and Rodbard, D. (1980) *Anal. Biochem. 107*, 220- 239.
24. Wu, J., Takayama, S., Wong, C. H., and Siuzdak, G. (1997) *Chem. Biol. 4*, 653−657.
25. Brenner, S., and Lerner, R. A. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 5381−5383.
26. Kerr, J. M., Banville, S. C., and Zuckermann, R. N. (1993) *J. Am. Chem. Soc. 115*, 2529−2531.
27. Ohlmeyer, M. H., Swanson, R. N., Dillard, L. W., Reader, J. C., Asouline, G., Kobayashi, R., Wigler, M., and Still, W. C. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 10922−10926.
28. Moran, E. J., Sarshar, S., Cargill, J. F., Shahbaz, M. M., Lio, A., Mjalli, A. M. M., and Armstrong, R. W. (1995) *J. Am. Chem. Soc. 117*, 10787−10788.
29. Nicolaou, K. C., Xiao, X.-Y., and Nova, M. P. (1995) *Angew. Chem., Int. Ed. Engl. 34*, 2289−2291.
30. Geysen, H. M., Wagner, C. D., Bodnar, W. M., Markworth, C. J., Parke, G. J., Schoenen, F. J., Wagner, D. S., and Kinder, D. S. (1996) *Chem. Biol. 3*, 679−688.
31. Ni, Z.-J., Maclean, D., Holmes, C. P., Murphy, M. M., Ruhland, B., Jacobs, J. W., Gordon, E. M., and Gallop, M. A. (1996) *J. Med. Chem. 39*, 1601−1608.
32. Feller, S. M., Ren, R., Hanafusa, H., and Baltimore, D. (1994) *Trends Biochem. Sci. 19*, 453−459.
33. Yu, H., Chen, J. K., Feng, S., Dalgarno, D. C., Brauer, A. W., and Schreiber, S. L. (1994) *Cell 76*, 933−945.
34. Sparks, A. B., Rider, J. E., Hoffman, N. G., Fowlkes, D. M., Quilliam, L. A., and Kay, B. K. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 1540−1544.

BI973122L