Methods for the Chemical Synthesis and Readout of Self-Encoded Arrays of Polypeptide Analogues

Philip E. Dawson,* Michael C. Fitzgerald, Tom W. Muir,† and Stephen B. H. Kent‡

Contribution from The Scripps Research Institute, La Jolla, California 92037

Received September 5, 1996[⊗]

Abstract: The synthesis of defined arrays of polypeptide analogues in conjunction with a simple self-encoded chemical readout system provides a powerful method for the systematic investigation of the relationship between polypeptide molecular structure and function. A novel solid-phase synthesis procedure was used to prepare arrays of polypeptide analogues in which a specific modification was systematically incorporated into a unique position in a peptide sequence. The synthesis was carried out in such a way that the resulting arrays contained a defined family of modified peptides, with each peptide molecule containing only a single specific modification. The array of polypeptide analogues was self-encoded in a positional fashion by incorporating a selectively cleavable bond into the analogue structure. Following cleavage of the polypeptide analogue array, analysis of the resulting peptide fragments by MALDI mass spectrometry defined, in a single step, the presence and identity of each peptide analogue in the mixture. The feasibility of this approach was demonstrated by the synthesis and mass spectrometric readout of an array of nine analogues of the 58-residue polypeptide chain of the cCrk N-terminal SH3 domain, before and after folding and affinity selection.

Introduction

An important current challenge in biomedical research is to understand and control the potent activities of peptides and of polypeptides that have folded to form the defined three-dimensional structures of functional protein molecules. One of the major strategies for determining the relationship between the chemical structure of a polypeptide and its biological activity is to systematically alter the covalent structure and observe the consequent effects on function. While the study of peptides has been dominated by chemical synthesis, in recent years, the study of proteins has been dominated by the use of site-directed mutagenesis.[1] This powerful recombinant DNA-based protein engineering technique has given many insights into the contribution of individual side chains to protein function. In particular, "alanine scanning",[2] an extension of this technique, has been used to identify specific amino acid side chains involved in ligand binding interactions.

Advances in the total synthesis of polypeptides have recently opened the world of proteins to direct application of the tools of organic chemistry. By using total chemical synthesis, a variety of protein analogues has been synthesized and important insights into the mechanism of action of enzymes and other proteins have been obtained.[3−8] Typically, individual protein analogues are prepared by chemical synthesis, and each analogue is then assayed to determine the effect of each modification on function. Although structure−function relationships in proteins can be studied using individual analogues prepared by either recombinant or chemical techniques, development of a profile of effects across the whole protein molecule is hindered by the time and effort required to generate and analyze multiple protein analogues.[9]

An alternative, currently popular method of studying polypeptides and proteins is through combinatorial chemistry. The use of combinatorial oligonucleotide synthesis in conjunction with expression in bacteria[10] or on phage[11] has provided a powerful method for studying large numbers of peptide and protein analogues. These techniques allow diverse pools of expressed polypeptide sequences to be probed for a desired function. The combinatorial chemistry approach has also had a major impact on the chemical synthesis of peptide analogues and has contributed greatly to the search for new biologically active peptides.[12] "Multiple peptide synthesis" allows peptides to be synthesized simultaneously;[13] individual peptide products are spatially separated and can be analyzed either attached to a solid

* To whom correspondence should be addressed: The Scripps Research Institute, 10550 North Torrey Pines Road, Mail Drop CVN6, La Jolla, CA 92037; email dawson@scripps.edu.

† Present address: The Rockefeller University, Box 223, New York, NY 10021.

‡ Present address: Gryphon Sciences, Suite 90, 250 E Grand Ave., South San Francisco, CA 94080.

⊗ Abstract published in *Advance ACS Abstracts,* July 15, 1997.

(1) Smith, M. *Angew. Chem., Int. Ed. Engl.* **1994**, *33*, 1214−1220.

(2) Wells, J. A. *Methods Enzymol.***1991**, *202*, 390−411.

(3) (a) Schnölzer, M.; Kent, S. B. H. *Science* **1992**, *256*, 221−225. (b) Jackson, D. Y.; Burnier, J.; Quan, C.; Stanley, M.; Tom, J.; Wells, J. A.; *Science* **1994**, *266*, 243−247. (c) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. *Science* **1994**, *266*, 776−779. (d) Canne, L. E.; Ferre-D'Amare, A. R.; Burley, S. K.; Kent, S. B. H. *J. Am. Chem. Soc.* **1995**, *117*, 2998−3007.

(4) Baca, M.; Alewood, P. F.; Kent, S. B. H. *Prot. Sci.* **1993**, *2*, 1085−1091.

(5) Rajarathnam, K.; Sykes, B. D.; Kay, C. M.; Dewald, B.; Geiser, T. Baggiolini, M.; Clark-Lewis, I.; *Science* **1994**, *264*, 90−92.

(6) Baca, M.; Muir, T. W.; Schnolzer, M.; Kent, S. B. H. *J. Am. Chem. Soc.* **1995**, *117*, 1881−1887.

(7) (a) Milton, R. C.; Milton, S. C.; Kent, S. B. H. *Science* **1992**, *256*, 1445−1448. (b)Zawadzke, L. E.; Berg, J. M. *J. Am. Chem. Soc.* **1992**, *114*, 4002−4003. (c) Fitzgerald, M. C.; Chernushevich, I.; Standing, K. G.; Kent, S. B. H.; Whitman, C. P. *J. Am. Chem. Soc.* **1995**, *117*, 11075−11080. (d) Schumaacher, T. N. M.; Mayr, L. M.; Minor, D. L.; Milhollen, M. A.; Burgess, M. W.; Kim, P. S. *Science* **1996**, *271*, 1854−1857.

(8) (a) Baca, M.; Alewood, P. F.; Kent, S. B. H. *Protein Sci.* **1993**, *2*, 1085−1091. (b) Baca, M.; Kent, S. B. H. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 11638−11642.

(9) Matthews, B. W. *Annu. Rev. Biochem.* **1993**, *62*, 139−160.

(10) (a) Reidhaar-Olsen, J. F.; Sauer, R. T. *Science* **1988**, *241*, 53−57 (b) Gregoret, L. M.; Sauer, R. T.; *Proc. Natl. Acad. Sci. U.S.A.* **1993**. *90*. 4246−4250.

(11) (a) Scott, J. K.; Smith G. P. *Science* **1990**, *249*, 386-390. (b) Lowman, H. B.; Wells, J. A. *J. Mol. Biol.* **1993**, *243*, 564−578.

(12) (a) For review, see: Thompson, L. A.; Ellman, J. A. *Chem. Rev.* **1996**, *96*, 555−600. (b) Gordon, E. M.; Barrett, R. W.; Dower, W. J.; Fodor, S. P. A.; Gallop, M. A. *J. Med. Chem.* **1994**, *37*, 1385−1401. (c) Scott, J. K.; Craig, L. *Curr. Opin. Biotechnol.* **1994**, *5*, 40−48.

(13) (a) Geysen, H. M.; Meloen, R. H.; Bartelig, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 3998−4001. (b) Houghten, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 5131−5134.

S0002-7863(96)03122-8 CCC: $14.00

support or in solution. In addition, established "split synthesis"[14] procedures allow for the rapid generation of huge numbers of peptide sequences in a single mixture through the repetition of a simple divide, couple, and recombine process.

The compositional diversity made possible by the phage display and split resin peptide synthesis approaches are advantageous for the discovery of new "lead" compounds since, in principle, all possible structural variants can be explored for the desired activity and only the few active oligomers of interest need to be individually identified.[12] However, where information about a complete set of functional and nonfunctional components is desired over many positions in a polypeptide sequence, such libraries are unnecessarily complex and may have limited utility.

A more systematic investigation of the molecular basis of peptide and protein function requires a different type of molecular diversity. Instead of a polypeptide mixture of high compositional diversity, it would be useful to construct an array of polypeptides which differ from each other in a precise and defined manner. In principle, one way to access this population would be as a subfraction of a large, fully combinatorial library. For example, such an array of analogues could consist of all polypeptides which differ from a target sequence by a single amino acid substitution at each position in a peptide sequence (cf. "alanine scans"). By removing this population from the context of a complex, fully combinatorial mixture of polypeptides, handling and analysis would be greatly simplified and a more useful profile of the effects of substituting the amino acid throughout the polypeptide chain would be obtained. Current split resin methods[14] do not allow for this type of control over the composition of a polypeptide library.

To apply chemistry to the study of polypeptide function, it would be useful to combine the valuable information gained by systematic modification of individual polypeptide analogues with the synthetic advantages of combinatorial methods. Here we describe a new strategy and associated experimental techniques for the chemical synthesis of defined arrays of polypeptide analogues together with a novel "self-encoding" principle for the direct readout of the composition of polypeptide analogue arrays. These techniques have been demonstrated in a model peptide system. We also describe a strategy and techniques for the functional separation of arrays of polypeptide analogues, after folding to form the corresponding protein structure, in this case the cCrk SH3 domain. The novel chemical synthesis and direct compositional readout techniques described in this paper form the basis of a systematic approach to the dissection of structure−activity relationships in peptides and proteins.

Results

Strategy for the Preparation of a Defined Array of Polypeptide Analogues. A polypeptide analogue array for use with the proposed self-encoding scheme should have two critical features. First, the components of the array must be approximately equimolar. Second, to avoid ambiguities, the array should consist only of polypeptides containing a *single* chemical modification per molecule, at each of a defined number of positions in the sequence. A straightforward procedure for synthesizing an array of this type has been developed and is schematically represented in Figure 1. For simplicity, the procedure will be illustrated for a hypothetical array of peptides consisting of substitutions of a single amino acid analogue at each of 10 consecutive positions in the amino acid sequence of the parent peptide.

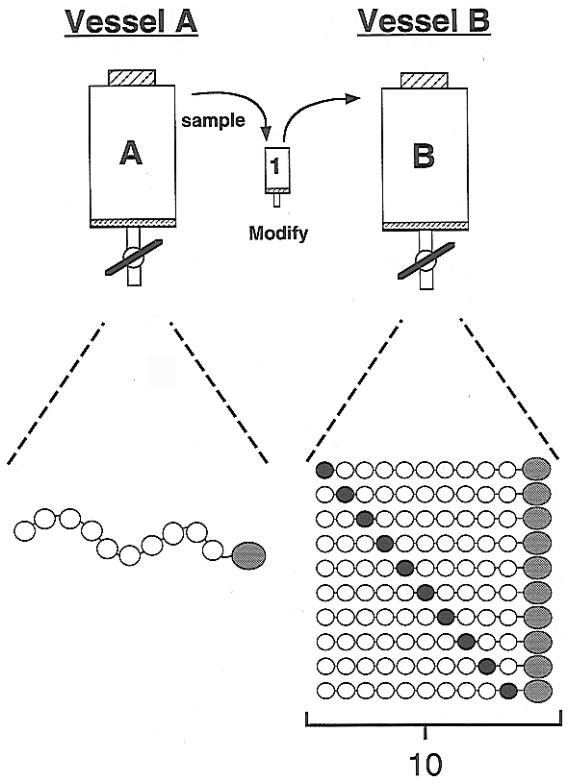


Figure 1. Basic strategy for the synthesis of defined arrays of peptide analogues. The general approach is to have two main reaction vessels, one for unmodified peptide resin, **A**, and the other for modified peptide resin, **B**. Standard stepwise solid-phase peptide synthesis of the parent amino acid sequence is performed in vessels **A** and **B**. Modifications to the sequence are made in a single auxilliary vessel, **1**. At the beginning of each step in which introduction of an analogue structure is desired, a sample of peptide resin is transferred from **A** to **1**, where it is modified and then transferred from **1** to **B** after completion of that cycle of synthesis in both **A** and **B**.

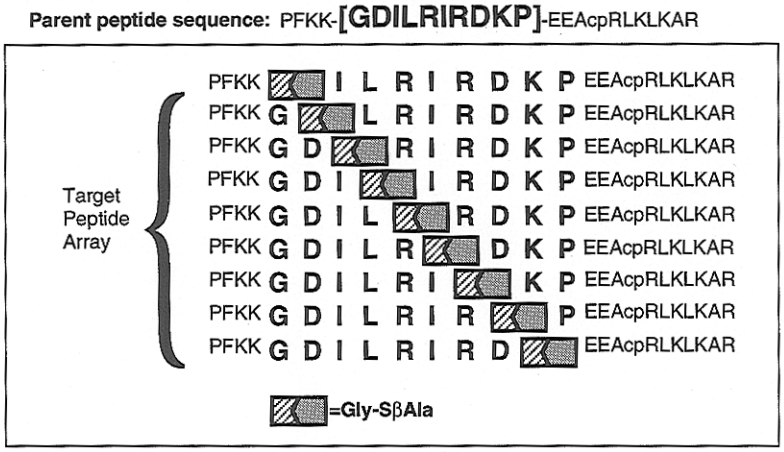
Two manual solid-phase peptide synthesis (SPPS) reaction vessels, **A** and **B**, and a small fritted funnel, **1**, are used to manipulate the peptide resin. The synthesis begins with 10 units of peptide resin in vessel **A**. After deprotection of the α-amino group, one unit of peptide resin is removed from **A** and added to the auxilliary vessel **1**. The first amino acid is then coupled to the nine units of peptide resin in **A** and the analogue moiety to the one unit peptide resin sample in **1**. After the coupling step, the analogue-modified peptide resin from **1** is transferred to **B**.

To initiate the next cycle of synthesis, the peptide resins in vessels **A** and **B** are deprotected. Then another unit of peptide resin is removed from **A** and transferred to the now empty **1**. The next amino acid in the sequence of the parent peptide is added in activated form to both **A** and **B**, while the analogue moiety is reacted with the new peptide resin sample in **1**. After completion of this cycle, the modified peptide resin in **1** is added to **B**. The synthesis continues in this manner for the requisite 10 cycles.

Throughout the synthesis, vessel **A** contains only unmodified peptide resin. Vessel **B** contains the single-site modified peptide resins, and vessel **1** contains the current sample of peptide resin which is being modified with the analogue unit. All chemical steps carried out in vessels **A** and **B** are identical, adding the amino acids of the unmodified sequence. At the end of 10 cycles, all the resin in vessel **A** has been transferred into vessel

(14) (a) Furka, A.; Sebestye, F.; Asedom, M.; Dibo, G. *Int. J. Pept. Protein Res.* **1991**, *37*, 487−494. (b) Lam, K.; Salmo, S.; Hersh, E.; Hruby, V.; Kazmiershi, W.; Knapp, R. *Nature* **1991**, *354*, 82−84.

A.



B.

Cycle#	A to 1	Neutralize 1	Couple AA_n A, B	+BocGly-SβAla to 1	DMF Wash A, B, 1'	1' to B	+TFA A, B	1 to 1'	DMF Wash A, B
1) Pro^{165}	10mL -1mL	√	√	√	√		√	√	√
2) Lys^{164}	9mL -1mL	√	√ √	√	√ √	√	√ √	√	√ √
3) Asp^{163}	8mL -1mL	√	√ √	√	√ √ √	√	√ √	√	√ √
etc.									
9) Asp^{157}	2mL -1mL	√	√ √	√	√ √ √	√	√ √	√	√ √
10) Gly^{156}			√ √		√ √ √	√	√ √		√ √

Figure 2. (A) Target composition of the nine-member array of peptide analogues. The sequence PFKKGDILRIRDKPEE was derived from residues 152−167 of the murine cCrk SH3 domain and the C-terminal AcpRLKLKAR sequence was used to facilitate analysis by MALDI mass spectrometry.[15] (B) Synthetic operations required for the synthesis of a peptide array consisting of nine overlapping dipeptide analogues over a 10-amino acid sequence. The synthesis was performed in a single day.

B which now contains the desired array of peptide analogues in resin-bound form (Figure 1).

Synthesis of a Defined Peptide Array. A peptide array consisting of defined analogues of a 10-amino acid sequence, GDILRIRDKP, within a 24-residue polypeptide of sequence PFKK-GDILRIRDKP-EEAcpRLKLKAR was chosen as a target to demonstrate the approach following the procedure described above. The target array is shown in Figure 2A and consists of overlapping dipeptide analogues spanning the region of interest. To facilitate characterization by mass spectrometry, the array was synthesized on resin bearing the sequence EEAcpRLKLKAR, where Acp is ϵ-aminocaproic acid.[15] The sequence PFKK was added to demonstrate the ability to modify peptides in the middle of a sequence and to facilitate the characterization of N-terminal fragments by mass spectrometry. The dipeptide analogue moiety corresponding to $-NHCH_2COSCH_2CH_2CO-$ (Gly-SβAla) was introduced as Boc-Gly-SβAla. Since the analogue moiety was incorporated as a *dipeptide*, a modification was made to the synthetic procedure outlined above and shown in Figure 1. To keep the synthetic operations being performed on the peptides in vessels **A** and **B** in register, the sample being derivatized in **1** was held out for *two* cycles before transfer to vessel **B**. To accommodate this modification, a second auxilliary vessel **1′** was added.

In practice, the peptide resin sample from vessel **A** was added to the auxilliary vessel in position **1**, where the dipeptide analogue coupling was initiated. After one cycle, the auxilliary vessel was moved to position **1′**, where the dipeptide analogue coupling continued during a second cycle of chain elongation in vessels **A** and **B**. The analogue-containing sample of peptide resin was then washed with DMF and transferred to vessel **B**. The synthetic steps for this synthesis are outlined in Figure 2B. After substituting dipeptide analogues for nine consecutive dipeptide sequences spanning a region of 10 amino acids, four additional amino acids, PFKK, were coupled to the array of peptide resins in vessel **B** to complete the target sequence.

Characterization of the Peptide Array. The peptide array described above contains a mixture of nine peptides, all 24 residues in length, each differing only in the *position* of a Gly-SβAla dipeptide substitution within the target sequence. As expected, the analytical HPLC of this array is quite complex, with many overlapping peaks (Figure 3A). The MALDI mass spectrum is also poorly resolved since the peptides in the array have a high redundancy in their molecular weights (Figure 3B). Thus the sequence -LRIRD- contains the dipeptides LR, RI, and IR, each of which have a molecular weight of 269 Da, and RD, which has a molecular weight of 271 Da. When substituted with Gly-SβAla (145 Da) in the peptide arrays, each of these substitutions would result in a peptide analogue with a molecular weight of 125 ± 1 Da below that of the unmodified sequence (PFKK-GDILRIRDKP-EEAcpRLKLKARr·amide, MW 2920 Da). The resulting MALDI mass spectrum of this peptide array would be expected to have a large peak around 2795 Da, representing the sum of four different peptide components (see Figure 3B).

(15) Zhao, Y; Muir, T. W.; Kent, S. B. H.; Tischer, E.; Scardina, J. M.; Chait, B. T. *Proc. Nat. Acad. Sci. U.S.A.* **1996**, *93*, 4020−4024..

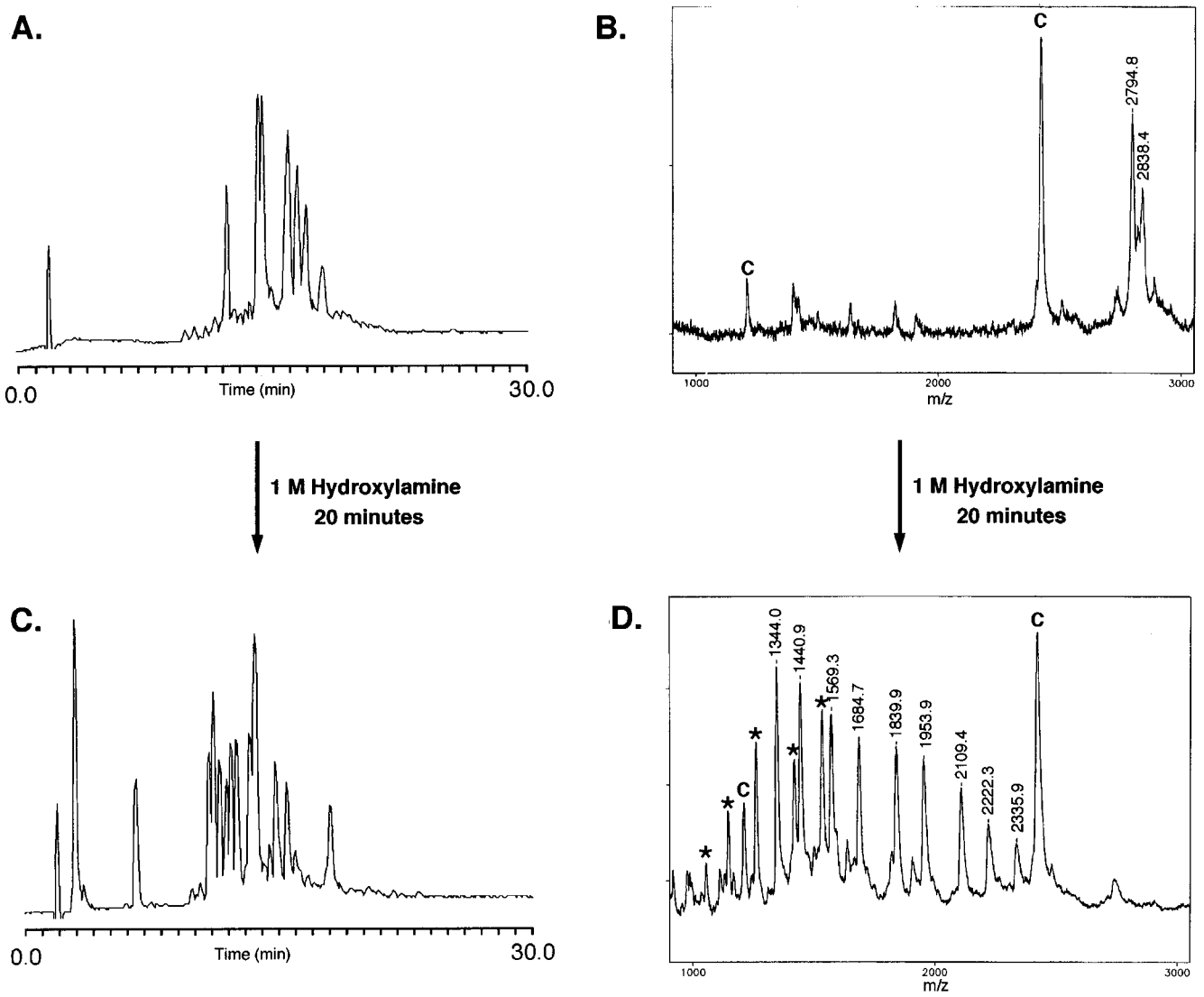


Figure 3. Analysis of a nine-component array of peptide analogues. (A**)** Analytical HPLC of crude full length product (gradient, 20−50% buffer B over 30 min). (B**)** MALDI mass spectrum of crude full-length product. Unlabeled peaks at lower mass are termination byproducts from the synthesis, and peaks labeled with C are from an internal calibrant. (C**)** Analytical HPLC of hydroxylamine-cleaved HPLC product on the same gradient. (D**)** MALDI readout of hydroxylamine-cleaved peptide array (peaks with * are N-terminal-containing fragments; unlabeled peaks are C-terminal containing fragments).

"Self-Encoded" Peptide Arrays. Neither analytical HPLC nor MALDI mass spectrometry are adequate for identification of components present in the array of peptide analogues. The distinguishing feature of the components in this peptide array is the *unique position* in the sequence of the modification. One approach to the unambiguous identification of the peptide components is to incorporate a selectively cleavable bond in the analogue unit. Cleavage of this bond in an analogue peptide molecule would result in two peptide fragments whose lengths, measured as mass, would identify the analogue peptide and define the *position* of the analogue unit in the peptide from which they derived. Such a chemical cleavage site would have to be stable to normal handling (folding and assay) conditions, while permitting selective cleavage on demand.

Synthesis and Characterization of a Peptide Containing a Cleavable Thioester Backbone. To demonstrate the synthesis and properties of a peptide containing a cleavable bond, the peptide LYRA(Gly-SβAla)-YGGFL·amide, was synthesized by stepwise SPPS using in situ neutralization coupling protocols. The thioester-containing dipeptide analogue, Boc-Gly-SβAla was activated as an HOBt ester and then coupled to preneutralized NH_2-YGGFL-(4-Me)benzhydrylamine) resin. Following deprotection and cleavage from the peptide resin, the stability and chemical cleavage characteristics of the model thioester-containing peptide was determined.

Thioester bonds within peptide sequences have been found to be stable at neutral pH.[3a,6,16] To test for stability to base hydrolysis, the peptide was dissolved at pH 9.0 in 200 μL of 100 mM Tris, 1 M Gn·HCl, vortexed vigorously for 10 s and left at 23 °C for 30 min. Surprisingly, no hydrolysis was observed under these conditions. Addition of 20 μL of 1 M NaOH (pH ∼13) gave complete hydrolysis after just 10 min as monitored by HPLC and electrospray mass spectrometry. In contrast to their stability to hydrolysis, thioesters have been shown to be very labile to hydroxylamine at neutral pH.[17] The thioester peptide was completely cleaved into LYRAG-NHOH and $HSCH_2CH_2CO$-YGGFL·amide when dissolved in 1 M NH_2OH, 200 mM NH_4HCO_3 (pH 6.0) for 30 min.[18]

Readout of the Nine-Component Peptide Analogue Array of the Parent Sequence PFKK-GDILRIRDKP-

(16) Canne, L; Bark, S. J.; Kent, S. B. H.; *J. Am. Chem. Soc*. **1996**, *118*, 5891−5896.

(17) Bruice, T. C.; Fedor, L. R. *J. Am. Chem. Soc.* **1964**, *86*, 4886−4897.

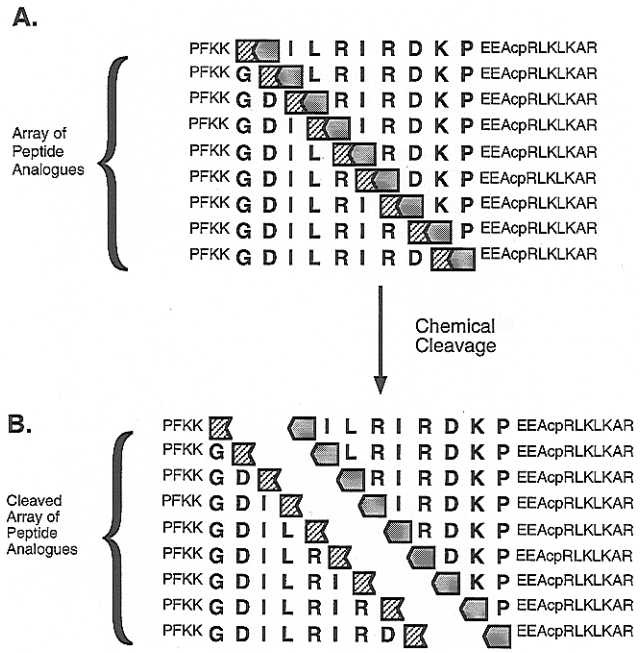


Figure 4. Schematic representation of the cleavage of the nine-component array of peptide analogues. (A**)** Full-length array of nine-peptide analogues. (B**)** Cleaved array of peptide analogues. The mixture consists of 18 peptides corresponding to nine N-terminal fragments and nine C-terminal fragments.

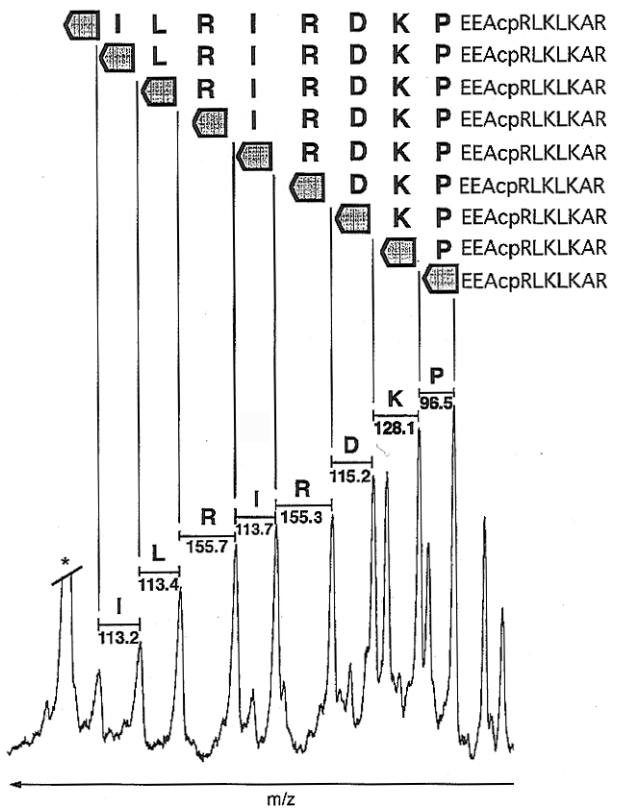


Figure 5. Relation of the MALDI spectrum to the peptide C-terminal fragment array. The horizontal mass scale spectrum has been inverted to align it with the standard N-to-C terminal orientation of the peptide sequence. The peaks corresponding to the nine C-terminal peptide fragments are clearly resolved and can be assigned sequentially. In addition to the position of the peak in the mass spectrum, the mass difference between adjacent peaks identifies the individual amino acids in the peptide sequence that has been subjected to analoging. The starred peak corresponds to the calibrant.

EEAcpRLKLKAR·amide. Synthesis of this array has been described above. Each member of this peptide array contains a Gly-SβAla replacement at one of the nine possible dipeptide positions within the 10 amino acid sequence [GDILRIRDKP] as shown in Figure 2A. To facilitate unambiguous identification of the components in the array, the thioester readout chemistry was incorporated into the dipeptide analogue. The thioester bond in the Gly-SβAla dipeptide analogue introduces a unique cleavage site into each member of the peptide array. Chemical cleavage of the peptide analogue array was expected to produce 18 peptide fragments as shown in Figure 4.

The peptide analogue array was cleaved by treatment with 1 M NH_2OH, 200 mM NH_4HCO_3 (pH 6.0) for 20 min. The resulting peptide fragments were then analyzed by HPLC and by MALDI mass spectrometry. As with the uncleaved peptide array (Figure 3A), the components of the cleaved peptide array still give rise to a complicated and essentially uninformative HPLC chromatogram (Figure 3C). By contrast, the MALDI spectrum of the cleaved peptide array (Figure 3D) provides a very straightforward characterization of the peptide array. As shown in Figure 5, the masses of the nine C-terminal fragments from the cleaved array bear a unique relationship to one another[19] and served to unambiguously identify the position of the analogue unit in each component of the original peptide array before cleavage.

The MALDI mass spectrometric readout characterized the peptide array in several ways. Cleavage of the analogue unit, at different positions throughout the array, produces two peptide fragments for each cleavage point. The resulting two families have either the N- or C-terminus of the parent peptide in common (Figure 4). The experimental identification of a member of the C-terminal family of peptides by MALDI can be directly related to the presence of each full length peptide analogue in the parent array. In addition to the C-terminal family, five members of the N-terminal family of peptides were also observed. The starred peaks on mass spectrum in Figure 3D correspond to these N-terminal peptides. In practice, the identification of either the N-terminal or C-terminal peptide fragments would serve to unambiguously characterize the peptide analogue array.

Another characterization of the peptide array can be obtained by looking at the difference in masses between the peaks on the mass spectra. As shown in Figure 5, the nine C-terminal peptide fragments are all of different lengths, and the mass differences between neighboring fragments correspond to the masses of individual amino acid residues. By correlation of these mass differences to the mass of individual amino acids, the sequence through which the analogue unit was substituted can be confirmed. This type of analysis has been previously used to sequence native peptides by "protein ladder sequencing".[20]

Application to a Protein System: Strategy for the Synthesis of Polypeptide Analogue Arrays and Readout after Functional Selection. Using the methods developed above on the model peptide system, a family of nine analogues was prepared to investigate the effect of an extensive covalent modification of the 58-amino acid residue cCrk SH3 polypeptide

(18) Note: Further studies have demonstrated that thioesters can be completely cleaved at concentrations of NH_2OH as low as 10 mM in <30 min (data not shown).

(19) Note: Signals corresponding to five of the nine N-terminal peptides were also observed. The masses that were not resolved were obscured by matrix ions in the region below 1000 Da in the spectrum of the peptide mixture.

(20) Chait, B. T.; Wang, R.; Beavis, R. C.; Kent, S. B. H. *Science* **1993**, *262*, 89−92.

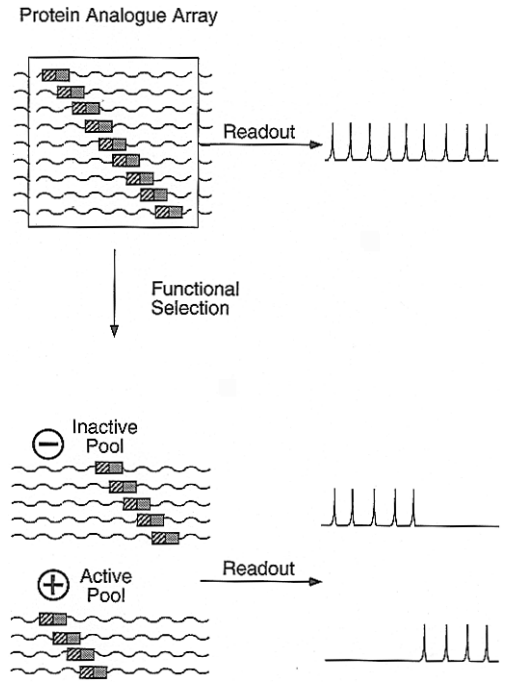


Figure 6. Integrated strategy for the chemical synthesis, functional separation, and analysis of a self-encoded array of protein analogues. The array of protein analogues is prepared by total chemical synthesis in a single procedure. Each analogue unit contains a selectively cleavable bond. Site-specific cleavage yields fragments that identify each protein component and define the position of the analogue unit within the polypeptide chain. This decoding procedure is applied to the parent array of analogues and to the active and inactive pools after separation based on function.

chain on its ability to fold and bind its specific ligand. The principle of this application of the method is shown in Figure 6. Each member of the synthetic protein array contained a single dipeptide analogue unit, $-NHCH_2COSCH_2CH_2CO-$ (Gly-SβAla), replacing pairs of adjacent amino acids at unique positions in the native sequence (see Figure 2). The mixture of analogue polypeptide chains was folded and assayed for binding to a specific ligand, a short proline-rich synthetic peptide derived from the sequence of the guanine nucleotide exchange factor C3G.[21] After the affinity selection, the binding and nonbinding pools of protein analogues were cleaved selectively at the thio ester bond contained in each analogue unit, and the composition of each pool was read out using MALDI mass spectrometry. By comparing the readout patterns obtained before and after a functional separation, a profile is obtained relating the effects of the analogue structure to its position in the polypeptide chain. Accumulation and interpretation of such qualitative profiles of protein structure−function relationships "protein signatures" will give insights into the chemical basis of peptide and protein function.

Chemical Synthesis of the Native 58-Residue SH3 Polypeptide. The polypeptide chain of the murine cCrk N-terminal SH3 domain, corresponding to residues 134−191 of the full cCrk signaling protein,[21] was assembled by highly optimized, stepwise solid-phase peptide synthesis using machine-assisted in situ neutralization protocols for *tert*-butoxycarbonyl (Boc) chemistry.[22] After deprotection and cleavage from the resin support, the crude polypeptide was purified by semipreparative reversed-phase HPLC and lyophilized using procedures previously described.[23] The purified product was characterized by analytical HPLC and by electrospray mass spectrometry.[24] The results are shown in Figure 7.

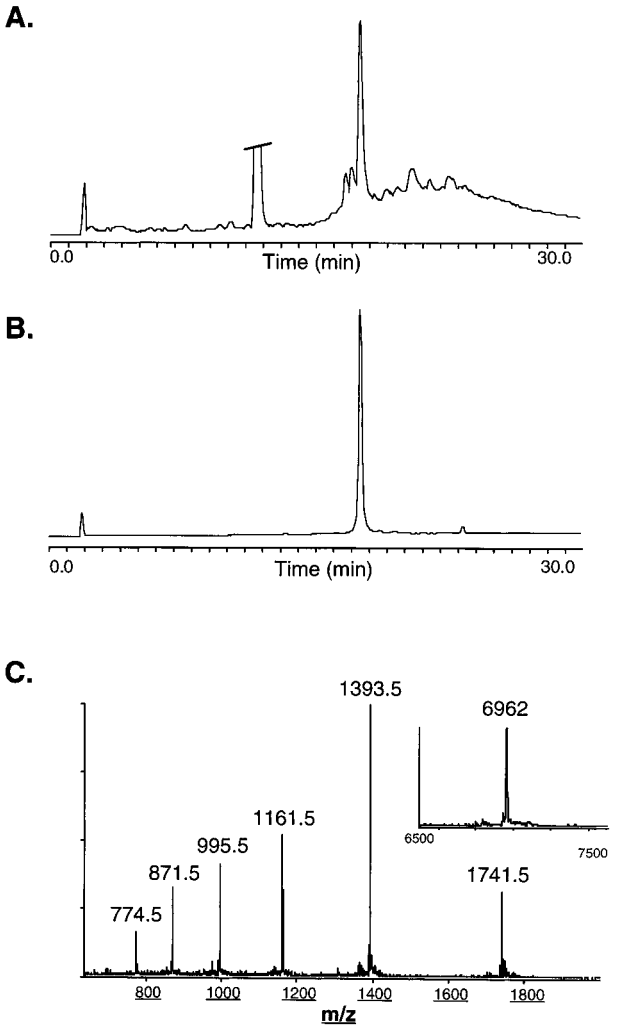


Figure 7. Characterization of the purified synthetic murine cCrk 134−191, N-terminal SH3 domain. (A) Analytical HPLC of the total crude peptide products from HF cleavage. (B) Analytical HPLC of the purified product on a gradient of 20%−50%B over 40 min. (C) Electrospray mass spectra of the purified product. Inset spectrum is reconstructed to a single charge state from the raw data below. Calculated average mass for $C_{313}H_{470}N_{84}O_{95}S_1$, 6961.8 Da; observed mass 6962 ± 1 Da.

Functional Characterization of the Synthetic SH3 Domain. The N-terminal cCrk SH3 domain was folded under these conditions: 0.2 mg of the purified 58-residue polypeptide in 600 μL of 20 mM HEPES and 50 mM NaCl (pH 7.3) at room temperature for 15 min. The folded protein was structurally characterized by NMR and by crystallization and X-ray diffraction studies.[25] The resulting synthetic protein domain was then assayed for its affinity for two different proline-rich peptides. Binding of this SH3 domain to its cognate peptide ligand buries a tryptophan side chain of the protein found in the binding pocket. This change in solvent exposure leads to

(21) Knudsen, B. S.; Feller, S. M.; Hanafusa, H. *J. Biol. Chem.* **1994,** *269*, 32781−32787.

(22) Schnölzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B. H. *Int. J. Pept. Protein Res.* **1992**, *40*, 180−193.

(23) Clark-Lewis, I.; Kent, S. B. H. In *The Use of HPLC in Receptor Biochemistry*; Venter, J. C., Harrison, L. C., Eds.; Alan R. Liss, Inc: New York, 1989; Chapter 3.

(24) Schnolzer, M.; Jones, A.; Alewood, P. F.; Kent, S. B. H. *Anal. Biochem.* **1992**. *204*. 335−343.

(25) The folded protein was charcterized by two-dimensional NOESY ^{1}H NMR spectroscopy and was identical with the recombinantly derived material as well as a synthetic sample which had been folded by gradual dialysis (Dawson, P.; Muir, T. Manuscript in preparation). In addition, the synthetic protein solution used for NMR analysis spontaneously formed highy diffracting crystals.

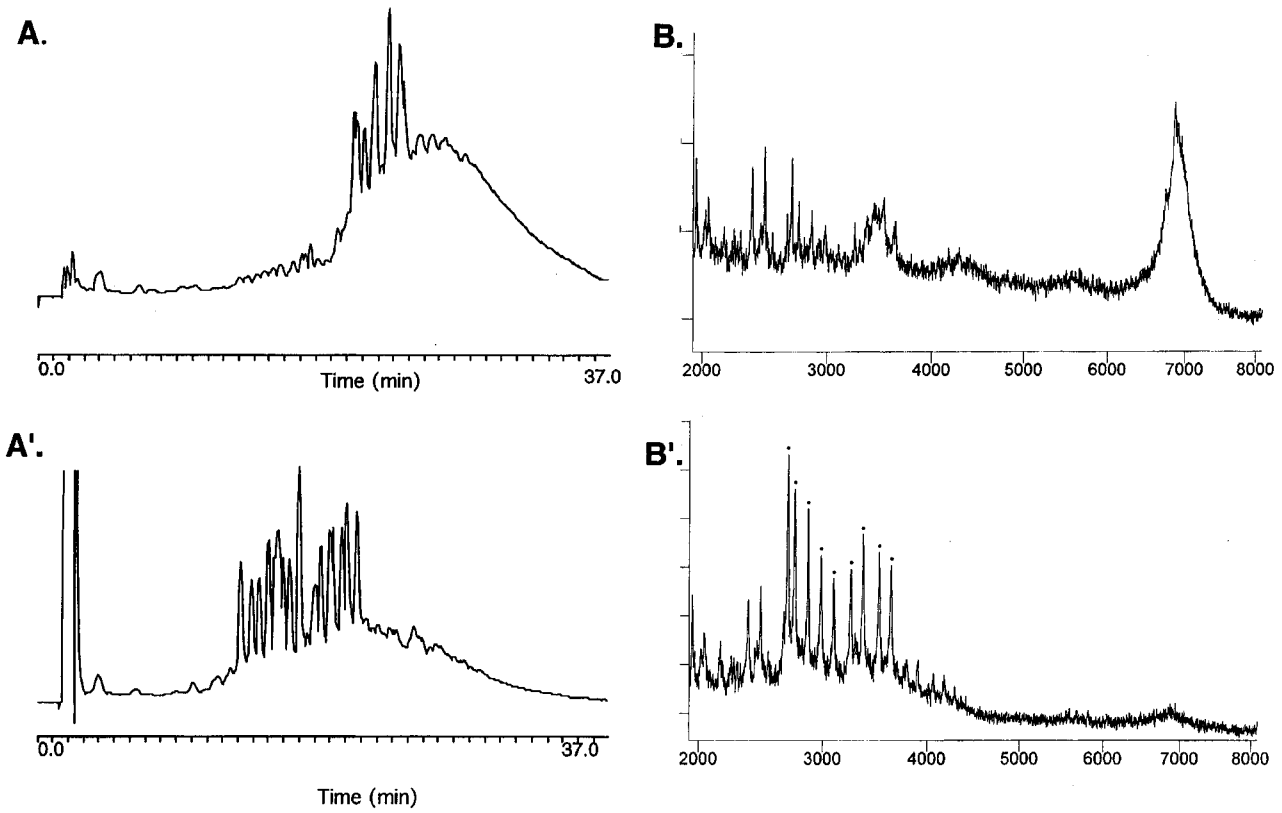


Figure 8. Characterization of the parent array of synthetic analogues of the cCrk SH3 domain. (A**)** Reverse-phase HPLC analysis of the crude nine-component protein array. (B) MALDI mass spectrum of the same crude product mixture. The protein array contained predominantly full-length polypeptide products. The presence of lower molecular weight species in the MALDI spectrum result from termination reactions during chemical synthesis. (A′) Treatment of the array with hydroxylamine produces the cleaved peptide products. Reverse-phase HPLC of the mixture after chemical cleavage with NH_2OH showed near-complete resolution of the 18 peptide fragments generated. (B′) MALDI mass spectrometry of the same cleaved mixture showed the characteristic patterns of cleavage fragments. The peaks marked with • unambiguously identified the protein components present in the original mixture. The order of these peaks in the mass spectrum identifies the corresponding analogue in the parent array and defines the position of the analogue unit in the polypeptide sequence.

an increase in the fluorescence intensity of the tryptophan side chain which can be monitored as a function of increasing ligand concentration.[26] The K_d's for the peptides C3G (PPALPPKKR• amide) and a peptide designed for attachment to an affinity column, acetyl-CWAcp-C3G, were found to be 1.8 and 2.4 μM, respectively.[27] The C3G affinity of 1.8 μM for the C3G-derived peptide is comparable to the affinity reported for a recombinantly derived SH3 domain (1.9 μM).[28]

Chemical Synthesis of an Array of Analogues of the 58-Residue Polypeptide Chain. The target array consisted of nine polypeptides, each containing a single $-NHCH_2COSCH_2CH_2-CO-$ (Gly-SβAla) substitution at one of the nine possible dipeptide units within the 10-amino acid sequence defined by residues 156−165 within the cCrk sequence 134−191. The analogue substitutions covered overlapping dipeptide sequences through the region -GDRILRIDKP-, corresponding to residues 156−165 of the cCrk sequence. This sequence is the same as that subjected to analoging in the model peptide studies (Figure 2).

The polypeptide analogue array was synthesized using a combination of manual and machine-assisted protocols. The amino acids corresponding to cCrk(166−191) and cCrk(134−155) were added using a machine-assisted protocol for Boc solid-phase chemistry.[29] Following synthesis of the sequence cCrk(166−191), the peptide resin was removed from the peptide synthesizer and placed in a manual-synthesis reaction vessel. Synthesis of cCrk(156−165) with concomitant introduction of the analogue unit at each position was performed manually as previously described above for a model peptide system using a modified split-resin procedure. Before addition of each residue (157−165), a sample of peptide resin was removed from reaction vessel **A**, modified with a dipeptide analogue held aside for two synthetic cycles, and then transferred to a second reaction vessel **B**. Identical synthetic operations building up the native amino acid sequence were carried out in reaction vessels **A** and **B**. In this manner, an array of nine resin-bound polypeptides was created as a single mixture, containing consecutive overlapping Gly-SβAla substitutions. Chain elongation of the nine-component mixture was continued through the sequence cCrk(134−155) using machine-assisted synthetic cycles, to complete the full length 58-residue target polypeptide sequence. The mixture of full-length analogue-containing polypeptide resins was subjected to HF cleavage and simultaneous side chain deprotection to give a crude lyophilized mixture of nine analogues of the 58-amino acid polypeptide chain of the N-terminal c-Crk SH3 domain.

The members of the polypeptide array were folded by dissolving 0.5 mg of crude peptide product in 200 μL of 20 mM HEPES and 50 mM NaCl (pH 7.3) at room temperature.

(26) Feng, S.; Chen, J. K.; Yu, H.; Simon, J. A.; Schreiber, S. A. *Science* **1994,** *266,* 1241−1245

(27) Due to the polyproline nature of the ligands, fluorescence measurements were taken after 12 h incubation of the protein with the peptide ligand to allow equillibration of the multiple cis and trans isomers. K_d's were extracted from the binding isotherms as previously described (Viguera et al. *Biochemistry* **1994**, *33*, 10925−10933).

(28) Wu, X.; Knudsen, B.; Feller, S. M. Zheng, J.; Sali, A.; Cowburn, D.; Hanafusa, H.; Kuriyan, J. *Structure* **1995**, *3*, 215−226.

(29) Schnölzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B. H. *Int. J. Pept. Protein Res.* **1992**, *40,* 180−193.

After 1 h, the protein mixture was analyzed by HPLC and MALDI mass spectrometry. By HPLC **(**Figure 8B), the nine protein components were partially resolved against a background of synthetic byproducts. MALDI mass spectrometry **(**Figure 8A′) showed an unresolved mixture of full-length polypeptide chains centered around 6950 Da; minor amounts of terminated components formed as byproducts in the chain assembly were also present. The presence of full-length protein analogues indicates that the thioester-containing polypeptide chains are stable under these handling conditions. However, neither HPLC nor MALDI-MS was able to define the composition of the array of protein analogues.

Readout of the Composition of the Parent Array of Protein Analogues. To characterize the protein analogue array, the polypeptide chains were specifically cleaved with hydroxylamine through nucleophilic attack at the thioester bond. The resulting peptide fragments were analyzed by both HPLC and MALDI-TOF (Figure 8A′,B′). The treatment with hydroxylamine specifically cleaved each analogue-containing polypeptide chain at the site of modification, resulting in a mixture of peptide fragments. As described in the model system, the cleavage of the protein array produces a mixture of peptide fragments, each with a different number of amino acids in the peptide chain. As shown in Figure 8A′, reverse-phase HPLC analysis of the cleaved mixture yields complicated sets of partly unresolved peaks. In addition, the order of the peaks bears no direct relationship to the position of the analogue unit in the parent polypeptide chain. Analysis of the cleaved mixture by MALDI mass spectrometry, on the other hand, produces a series of well-resolved peaks, the relative positions and masses of which are directly related to the position in which the analogue unit was placed in each full-length polypeptide chain and which serve to identify the analogue molecules present in the array. The peptide ladder shown in Figure 6B′ corresponded to all nine of the expected N-terminal peptide fragments resulting from cleavage of the nine protein analogues and unambiguously characterized the protein array.

Since the cleavable bond was placed within the 58-residue SH3 polypeptide chain, the cleavage of each analogue must produce both N- and C-terminal peptide fragments. In the MALDI spectrum, however, only the N-terminal peptide fragments were observed with high intensity. The analysis of peptide mixtures by MALDI mass spectrometry can vary depending on the choice of matrix and solvent composition.[30] This phenomena did not compromise our experimental results since the MALDI readout system relied on the comparison of mass spectra before and after selection. In addition, detection of only one of the two peptide fragments from each protein was required for identification of the parent protein analogue. In fact, interpretation can be simplified when only fragments corresponding to one end of the polypeptide chain are observed. (If necessary, this simplification can be systematically achieved by fractionation based on suitable end-labeling of the array of intact synthetic polypeptide chains during synthesis.)

Affinity Chromatography of the Array of cCrk SH3 Protein Analogues. The lyophilized crude mixture of analogues of the 58-residue polypeptide chain was folded by dissolving 1.5 mg in 600 μL of 20 mM HEPES and 50 mM NaCl (pH 7.3). The dissolved protein array was then applied to an affinity column loaded with the C3G-derived sequence acetyl-C-Acp-PPALPPKKR·amide and left to bind for 6 h. The column was then washed with 0.5 M NaCl and 1 M NaCl, 100 mM sodium phosphate (pH 7.0) to remove nonspecific binding proteins from the column. Specifically bound protein analogues were then eluted with 1 M hydroxylamine and 200 mM NH_4OH (pH 6.0). Amounts of eluted peptide were monitored by UV absorbance at 280 nm. This procedure was used for both the C3G peptide column and for a control column loaded with acetyl-C-AcpYGGFL·amide. The hydroxylamine cleavage of the thioester bond resulted in the elution of peptide fragments corresponding to the proteins which were able to bind to the affinity column under these conditions. Specific binding was found only for the C3G peptide column.

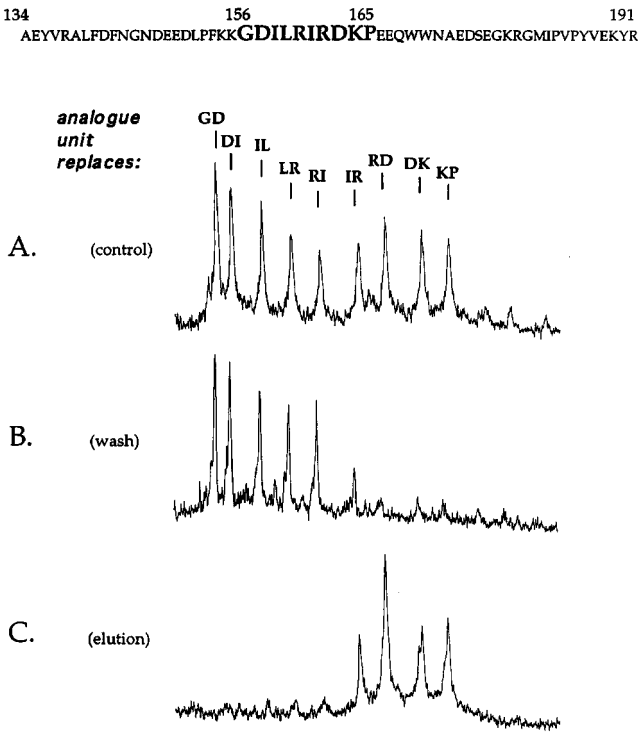


Figure 9. Readout of the composition of the parent nine-component cCrk SH3 domain array and the binding and nonbinding pools. The cCrk SH3 domain array of protein analogues was folded in assay buffer and added to a C3G peptide affinity column. Column fractions were treated with hydroxylamine and then analyzed by MALDI mass spectrometry. (A, control) Composition of the parent array of cCrk SH3 domain analogues. (B, wash) Nonbinding cCrk SH3 domain analogues eluted in the 0.5 M NaCl wash. (C, elution) Specifically bound cCrk SH3 domain analogues eluted with hydroxylamine. MALDI peaks are marked by the single-letter code for the dipeptide that had been was substituted with Gly-SβAla. Note: only the N-terminal peptide fractions are observed in the three MALDI spectra (see main text).

Readout of the Composition of the Arrays of Protein Analogues. To demonstrate the utility of the self-encoding principle, the composition of the binding and nonbinding fractions obtained from affinity chromatography of the array of SH3 analogues was determined by chemical cleavage and MALDI MS. After a desalting step, fractions from the parent array of folded protein analogues and the 0.5 M NaCl wash from the affinity chromatography of the array were cleaved separately with hydroxylamine. MALDI mass spectra were obtained for the peptide fragments generated from the (parent) protein array, for the salt wash (nonbinding), and for the peptide fragments generated by the hydroxylamine elution of the specifically bound analogues. The results are shown in Figure 9. Nine components were present in the array of protein analogues that was added to the C3G column as one mixture **(**Figure 9A**)**. Of these nine components, five did not bind significantly to the affinity column and were present only in the 0.5 M NaCl wash fractions **(**Figure 9B**)**. Three protein analogues, however, were able to bind to the C3G peptide on the column and were eluted only after cleavage with hy-

(30) Cohen, S. L.; Chait, B. T. *Anal. Chem.* **1996**, *68*, 31−37.

droxylamine **(**Figure 9C**)**. One of the protein analogues can be identified in both the wash and elution spectra, indicating intermediate folding and/or binding properties for this analogue. Thus, one-step MALDI-MS readout following chemical cleavage could be used to unequivocally determine the composition of three different protein analogue arrays. These experiments demonstrate the utility of the self-encoded polypeptide library approach to the study of analogue mixtures.

Discussion

To elucidate the molecular basis of polypeptide function, a typical approach has been to systematically alter the covalent structure and to monitor the consequent effect on function. Traditionally, each peptide analogue has been separately synthesized, purified, and the biological activity characterized, on an individual basis. The data from a number of these experiments can be pooled and interpreted to give a profile of structure−function relationships. In this work, we outline an approach to streamline such studies. Instead of synthesizing individual polypeptide analogues, a defined array is synthesized as a single mixture. This polypeptide array can be assayed as a pool, after which individual peptide components can be identified through a novel self-encoding scheme. Using this approach, the comparative properties of the individual components can be determined for a given assay/function.

Polypeptide arrays were synthesized using a modified "split-resin" procedure which, unlike previous procedures, resulted in a defined array of components with only the desired complexity. A key aspect of this approach is the ability to prepare a defined subset of a fully combinatorial mixture. By synthesizing such a subset, arrays of manageable size can be prepared that contain defined modifications at a large number of positions in the polypeptide sequence. For example, using the approach described here, five different amino acid analogues could be substituted at each position of a 10 amino acid sequence so that there is only one modification in each peptide molecule. The resulting array would result in a mixture of 50 peptides (5 analogue structures × 10 positions). However, using the standard split synthesis approach,[14] the same 50 peptides would be only a small fraction of a library of $\sim 10^7$ (5^{10}) peptides. Mixtures of limited, defined diversity may be of greater utility for systematically defining structure−activity relationships

Once diversity has been generated, and a selection performed, a method is needed to identify individual components. Decoding polypeptide mixtures has in the past presented a substantial challenge. Currently, most encoding strategies involve a molecular tag which can be read by sensitive analytical techniques[31] or by amplification.[32] Many of these techniques, however, rely on the assay of molecules still attached to a solid support through the isolation and analysis of individual beads. To avoid the necessity for a solid support, the encoding must be associated with the peptide analogue at the molecular level.[32]

Incorporation of a chemically cleavable bond at specific sites within each polypeptide analogue provides an example of an alternative simple and practical encoding scheme at the molecular level. The components of a polypeptide analogue array can be decoded through chemical cleavage and one-step mass spectrometric readout. Detection of the resulting peptide fragments unambiguously defines the presence or absence of a given analogue molecule in the selected population. As demonstrated in the analysis of the nine-member peptide array, matrix-assisted laser-desorption ionization (MALDI) mass spectrometry is well suited for the decoding of linear arrays of peptides. This use of mass spectrometry to decode mixtures of peptide analogues is analogous to the use of gel electrophoresis to separate nucleotides by length during DNA sequencing and analysis.[33] The high resolution and sensitivity (<1 pmol/component)[34] of MALDI mass spectrometry allows the characterization of even small quantities of the entire peptide array.

In this work, we have demonstrated the feasibility of this approach through the synthesis of a polypeptide array which was then characterized by MALDI mass spectrometry following chemical cleavage. A nine-component peptide array of the parent sequence, PFKK-GDILRIRDKP-EEAcpRLKLKAR· amide, was synthesized in which a single dipeptide analogue, Gly-SβAla, was introduced into consecutive positions through the sequence -GDILRIRDKP- (Figure 2). This array was self-encoded with a chemically cleavable thioester bond which was incorporated into the analogue unit. The peptide components were then identified by cleaving the thioester bond in each peptide with hydroxylamine, followed by MALDI mass spectrometry. The resulting series of peaks on the mass spectrum unambiguously identified the presence of all nine peptide analogues in the peptide array.

The work reported in this paper also demonstrates the application of combinatorial synthetic chemistry techniques to a protein target. Straightforward synthetic access to protein arrays containing a dipeptide analogue Gly-SβAla has been demonstrated in the context of residues 156−165 of the cCrk N-terminal SH3 domain (residues 134−191) of cCrk. The latent readout functionality has been shown to be stable to conditions of protein folding and ligand binding, yet is cleavable by brief treatment with 1 M hydroxylamine. A selection for binding activity based on the use of a C3G-derived synthetic peptide for affinity chromatography has been developed and was used to demonstrate the functional separation of the array of protein analogues. Finally, chemical cleavage of synthetically introduced latent cleavage sites followed by MALDI-TOF mass spectroscopy has been used to read out the composition of "self-encoded" pools of protein analogues.

The power of a chemical synthesis approach to the study of polypeptides and proteins lies in the straightforward access it provides to a wide range of variations in molecular structure. Many types of chemical modifications can be introduced using the approach outlined here. We have already applied this approach to study the effects of adding an extra methylene group into the backbone of a protein.[35] Further studies could investigate the ability of the protein to tolerate restrictions in backbone conformation; aminoisobutyric acid (α,α-dimethylglycine) residues are known to restrict Ramachandran space to α-helical conformations,[36] while insertion of β-turn mimics into the polypeptide chain can provide a test for such secondary

(31) (a) Needels, M. C.; Jone, D. G.; Tate, E. H.; Heinkel, G. L.; Koxhersperger, L. M.; Dower, W. J.; Barrett, R. W.; Gallop, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 10700−10704. (b)Kerr, J. M.; Banville, S. C.; Zuckermann, R. N. *J. Am. Chem. Soc.* **1993**, *115*, 2529−2531. (c) Nikolaiv, V.; Stierandova, A.; Krchnak, V.; Seligmann, B.; Lam, K. S.; Salmon, S. E.; Lebl, M. *Pept. Res.* **1993**, *6*, 161−170. (d) Ohlmeyer, M. H. J.; Swanson, R. N.; Dillar, 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.

(32) (a) Brenner, S.; Lerner, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5381−5383. (b) Nielsen, J.; Brenner, S.; Janda, K. D. *J. Am. Chem. Soc*. **1993**, *115*, 9812−9813.

(33) A similar cleavage and separation by size readout was used for nucleic acids in the following: (a) Pan, T.; Gutell, R. R.; Uhlenbeck, O. C. *Science* **1991**, *254*, 1361−1364. (b) Hayashibara, K. C.; Verdine, G. L. *J. Am. Chem. Soc.* **1991**, *113* (3), 5104−5106.

(34) Chait, B. T.; Kent, S. B. H. *Science* **1992**, *257*, 1885−1894.

(35) Muir, T. W.; Dawson, P. E.; Fitzgerald, M. C.; Kent, S. B. H. *Chem. Biol.* **1996,** *3*, 817−825.

(36) Marshall, G. R.; Bosshard, H. E. *Circ. Res.* **1972**, *Suppl. II 30 and 31*, 143−150.

structural features. The modifications possible using the tools of molecular biology are good for monitoring side chain interactions but, with the exception of glycine and proline, do little to probe the conformational properties of the peptide backbone. The use of chemical synthesis allows for experiments which probe the tolerance of cis rather than trans peptide bond replacements and the ability of the peptide backbone to explore "D" Ramachandran space. Such experiments, may give insight into the molecular characteristics of the peptide backbone. Using the chemical techniques described in this paper, the full range of modifications that have been used to elucidate the structure−function relationships of peptides can now be applied to systematically study the world of proteins.

Conclusion

The combination of the synthesis of an array of peptides corresponding to a defined subset of a fully combinatorial mixture with a self-encoding strategy results in an information-rich approach to the elucidation of the structure−activity relationship of peptides, polypeptides, and proteins. The power of this approach is illustrated when the polypeptide array is subjected to a functional selection. Since *all* members of the polypeptide array can be observed in a single readout step, this approach can generate information on both the positively and negatively selected members of the array. The ability to synthesize multiple polypeptide analogues in a single procedure, followed by functional characterization of the entire array, should give greater insight into the molecular basis of polypeptide function.[15]

The techniques described should be generally applicable to proteins accessible by chemical synthesis.[37] With the introduction of modular chemical ligation techniques,[3a,6] individual protein domains, as well as series of these domains, can be investigated. Since these protein domains are the basic units of protein function, the systematic generation of arrays of such protein analogues will allow the tools of organic chemistry to be used for the systematic elucidation of the molecular basis of protein function.

Experimental Methods

General Methods. Analytical HPLC was performed on a Hewlett-Packard 1050 system with 214 nm detection using a Vydac C18 column (5 μm, 4.6 × 150 mm) at a flow rate of 1 mL/min. Runs used either a linear 0%−67%B gradient over 30 min or a linear 25%−45%B gradient over 30 min (where buffer A was 0.1% TFA in H_2O and buffer B was 90% acetonitrile, 10% H_2O, and 0.09% TFA). Semipreparative HPLC was performed on a Rainin HPXL dual-pump system with detection on a Dynamax UV detector using a Vydac C18 column (10 μm, 10 × 250 mm) at 3 mL/min. Electrospray mass spectrometric analysis of all synthetic peptides was performed on a Sciex API-III triple-quadrupole electrospray mass spectrometer. Calculated masses were obtained using the program MacProMass (Sunil Vemuri and Terry Lee, City of Hope, Duarte, CA).

Matrix-Assisted Laser Desorption Ionization Mass Spectrometry (MALDI). Mass spectra of the peptide arrays were recorded using a Vestec Model VT 2000 laser desorption, linear time-of-flight mass spectrometer. Samples were desorbed/ionized using the focused output of a 355 nm frequency-tripled Lumonics Model HY 400 Nd:YAG laser (Lumonics, Kanata, ON, Canada). Ions were accelerated through a dual-stage source to a total potential of 30 keV and detected by a 20-stage focused mesh All spectra were acquired in the positive ion mode and summed over 50 laser pulses. Time-to-mass conversion was accomplished by internal calibration using the $[M + H]^+$ and $[M + 2H]^{2+}$ ion signals from a standard peptide (MW 2419.1 Da). Samples were prepared by dissolving the crude peptide array in 1:1 acetonitrile:H_2O, 0.1% TFA to a concentration of 1−10 μM per peptide component. A 2 μL portion of this solution was mixed with 5 μL of a saturated solution of 2,5-dihydroxybenzoic acid (DHB) in the same solvent. Ultimately, 2 μL of this mixture containing ∼1−10 pmol of each peptide component was added to a stainless steel probe tip (3.14 mm^2) and the solvent allowed to evaporate under ambient conditions.

Mass spectra of the cCrk protein array were recorded using a prototype laser desorption, linear time-of-flight mass spectrometer from Ciphergen Biosystems (Palo Alto, CA). The peptide samples were ionized using 337 nm radiation output from a nitrogen laser (Laser Science, Inc., Newton MA). Samples were prepared by adding a 2 μL aliquot of the 1.5 mL desalted affinity column fraction to 5 μL of a saturated solution of α-cyano-4-hydroxycinnamic acid in 50% acetonitrile in water, 0.1% TFA. From this mixture, 2 μL, containing ∼1−10 pmol of each peptide component, was added to a stainless steel probe tip (3.14 mm^2) and the solvent allowed to evaporate slowly under ambient conditions. All spectra were acquired in the positive ion mode and summed over 20−50 laser pulses. Time-to-mass conversion was accomplished by internal calibration using the $[M + H]^+$ signals from the largest and smallest peptide components in each array.

Solid-Phase Peptide Synthesis. Except where noted, all peptides were synthesized using either manual or machine-assisted in situ neutralization/HBTU activation protocols for Boc-solid-phase chemistry as previously described.[29] Machine-assisted synthesis was perfomed using a modified Applied Biosystems 430A peptide synthesizer. Model peptides were synthesized on (4-Me)benzhydrylamine−copoly(styrene−1% DVB) resin (Peninsula Laboratories, 0.93 mmol/g), and the cCrk SH3 domain was synthesized using Boc-Arg(Tos)-OCH_2-Pam resin (Applied Biosystems, 0.59 mmol/g). Following synthesis, the N^{α}Boc group was removed and the peptide cleaved from the resin with concomitant removal of side chain protecting groups by treatment for 1 h at 0 °C with anhydrous HF containing 5% anisole or 5% *p*-cresol as a scavenger. After evaporation of the HF, the crude peptide was precipitated and washed with cold diethyl ether, dissolved in 1:1 acetonitrile:H_2O, containing 0.1% TFA, filtered to remove the resin, and lyophilized.

Synthesis of the Nine-Component Peptide Analogue Array of the Parent Sequence PFKK-[GDILRIRDKP]-EEAcpRLKLKAR. The sequence EEAcpRLKLKAR was synthesized in reaction vessel **A** (Figure 1) on 0.1 mmol of MBHA resin. Onto this sequence, an array of nine Gly-SβAla-substituted peptide analogues was synthesized through the sequence GDILRIRDKP using the protocol described below. Boc-Gly-SCH_2CH_2COOH (0.25 mmol, 66 mg) was preactivated for 1 h with DIC (0.25 mmol, 39 μL) and HOBt (0.25 mmol, 34 mg) in 600 μL DMF (∼0.4 M) and used for five consecutive cycles (125 μL/cycle), after which a second 0.25 mmol of the dipeptide analogue was activated under the same conditions and used for the remaining four cycles.

Cycle 1. First, the N^{α}-Boc deprotected peptide resin (0.10 mmol) was suspended in 10 mL of DMF. One milliliter (∼0.01 mmol) of the suspension was removed and added to a small fritted funnel. This sample was then neutralized for 1 min with 10% DIEA in DMF, drained, placed in position **1**, and reacted with the activated thio ester dipeptide analogue (125 μL, 0.05 mmol). During neutralization of the sample, the first subsequent activated amino acid, Boc-proline, was coupled to the resin in vessel **A** using manual in situ neutralization synthetic cycles with HBTU as the activating agent. After 20 min of coupling, the peptide resin in vessel **A** was washed with DMF, treated with TFA, and washed again with DMF. The first peptide resin sample was then moved to position **1′**, where dipeptide coupling was allowed to continue.

Cycle 2. The deprotected peptide resin in vessel **A** was suspended in 9 mL of DMF and 1 mL (∼0.01 mmol) was transferred to a second small fritted funnel and placed in position **1**. After neutralization of this sample, 125 μL (0.05 mmol) of the activated dipeptide was added to the sample which was placed in the now open position **1**. At the same time, activated Boc-Lys was added to vessel **A**. After the lysine coupling in vessel **A** was complete, the first (dipeptide analogue) peptide resin sample was transferred from position **1′** to reaction vessel **B**. The

(37) Muir, T. W.; Kent, S. B. H. *Curr. Opin. Biotechnol.* **1993**, *4*, 420−427.

peptide resins in **A** and **B** were then deprotected with TFA and washed, and finally, the sample in position **1** was moved to position **1′**.

This procedure was continued for a total of 10 cycles of chain elongation, with the final cycle skipping the removal of resin from vessel **A**. Finally, the sequence PFKK was added to the peptide resin in vessel **B**, by stepwise SPPS. Following deprotection and cleavage from the resin, the lyophilized peptide analogue array was analyzed by analytical reverse-phase HPLC, electrospray mass spectrometry, and MALDI mass spectrometry.

Synthesis of the Nine-Component Analogue Array of cCrk (134−191). The target array of analogue polypeptide chains is shown in Figure 3. The sequence corresponding to cCrk (166−191) was synthesized on a 0.2 mmol of Boc-Arg(Tos)-OCH_2-Pam resin (0.59 mmol/g) using machine-assisted synthetic cycles. The array of polypeptide analogues was manually synthesized on ∼0.05 mmol (250 mg) of this peptide resin by a modified split-resin procedure previously described for the peptide array. Boc-Gly-SβAla-OH (0.25 mmol, 66 mg) was preactivated for 1 h with DIC (0.25 mmol, 38 μL) and HOBt (0.25 mmol) in DMF: total volume 650 μL. The apparatus for manual synthesis consisted of two standard manual synthesis reaction vessels, labeled **A** and **B,** and two small fritted funnels in a test tube rack in positions **1** and **1′**.

Following this procedure, half of the mixture of peptide resins in vessel **B** (25 μmol total) containing nine peptide analogues was removed and transferred to an Applied Biosystems 430A peptide synthesizer for addition of the remaining amino acids, cCrk 134−155. Following deprotection and cleavage from the resin, the lyophilized peptide array was analyzed by analytical reverse-phase HPLC and MALDI mass spectrometry.

Acknowledgment. We thank the Peptide Group at the R. W. Johnson Research Institute in La Jolla, CA, for use of the Vestec MALDI mass spectrometer and Ciphergen Biosystems (Palo Alto, CA) for use of their prototype MALDI mass spectrometer. M.C.F. was an Amgen Fellow. This work was supported by funds from NIH Grants PO1GM48870 and RO1GM48897.

Supporting Information Available: Full experimental details of the synthesis of the peptides and peptide arrays as well as the readout and affinity chromatography procedures (8 pages). See any current masthead page for ordering and Internet access instructions.

JA963122T