Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters 15 (2005) 5329–5334
Bioorganic & Medicinal Chemistry Letters

Geometric diversity through permutation of backbone configuration in cyclic peptide libraries

Zachary E. Perlman,[a] Jonathan E. Bock,[b] Jeffrey R. Peterson[c] and R. Scott Lokey[b,*]

[a]*Department of Systems Biology, Harvard Medical School, 240 Longwood Avenue, Boston, MA 02115l, USA*
[b]*Department of Chemistry and Biochemistry, University of California Santa Cruz, 1156 High St., Santa Cruz, CA 95064, USA*
[c]*Fox Chase Cancer Center, 333 Cottman Ave., Philadelphia, PA 19111-2497, USA*

Received 16 June 2005; revised 15 July 2005; accepted 25 July 2005

Abstract—Cyclic peptides offer the possibility of varying both scaffold geometry and R-group functionality. For example, parameters such as ring size and the placement of D-amino acid and proline residues can have a dramatic effect on the conformations of cyclic peptides, allowing access to structurally diverse species based on simple modifications in their linear sequences. We synthesized a cyclic peptide library in which ring size, α-carbon stereochemistry, and proline placement were varied. Analysis of the products showed that heptapeptides in general cyclized more readily than hexapeptides, and within these groups the scaffolds with a greater number of pralines cyclized with markedly lower yields than scaffolds with fewer pralines. Split-pool libraries based on a sample set of these scaffolds showed that, in general, scaffold geometry outweighed side chains variation in determining cyclization efficiency. These concepts were applied to the synthesis of cyclodimeric variants of an inhibitor of actin assembly in *Xenopus* egg extracts, yielding side chain variants with improved potency over the original scaffold.

Most combinatorial libraries consist of varied R-groups appended to a common core scaffold.[1] R-groups are typically chosen to maximize diversity, but their relative spatial orientations are usually fixed by the geometric constraints of the scaffold. Cyclic peptides offer the possibility of conveniently varying both scaffold geometry and R-group functionality. For example, parameters such as ring size and the placement of D-amino acid[2,3] and proline[4] residues can have a dramatic effect on the conformations of cyclic peptides, allowing access to structurally diverse species based on simple modifications in their linear sequences.

We have been interested in cyclic peptide scaffolds, inspired by natural products such as the phakellistatins[5] and yunnanins,[6] as sources of chemical diversity for input into phenotype-based assays. Previously we reported the synthesis of a 384-member cyclic peptide library in which ring size, proline position, and α-carbon stereochemistry were permuted. One of the compounds from this library inhibited lipid-induced actin assembly in *Xenopus* egg extract (IC_{50} = 2 μM), an activity that correlated with its ability to stabilize an auto-inhibited form of the protein N-WASP.[7] Interestingly, the active species in this assay was a 14-residue cyclodimeric side product resulting from inefficient cyclization of a heptapeptide precursor. In light of this study, we have attempted to assess the relative contributions of geometric and functional group variation in determining cyclization efficiency and biological activity, and the degree to which these two sources of variation can be separated. This report presents an exploration of some determinants of cyclization efficiency in scaffolds composed of

1. cyclo[D,L Lys - D,L Phe - D,L Phe - D,L Phe - D,L **Pro** - L Gln]
2. cyclo[D,L Lys - D,L Phe - D,L Phe - D,L **Pro** - D,L **Pro** - L Gln]
3. cyclo[D,L Lys - D,L Phe - D,L **Pro** - D,L Phe - D,L **Pro** - L Gln]
4. cyclo[D,L Lys - D,L **Pro** - D,L Phe - D,L Phe - D,L **Pro** - L Gln]
5. cyclo[D,L Lys - D,L Phe - D,L Phe - D,L Phe - D,L **Pro** - D,L **Pro** - L Gln]
6. cyclo[D,L Lys - D,L Phe - D,L Phe - D,L **Pro** - D,L Phe - D,L **Pro** - L Gln]
7. cyclo[D,L Lys - D,L Phe - D,L **Pro** - D,L Phe - D,L Phe - D,L **Pro** - L Gln]
8. cyclo[D,L Lys - D,L **Pro** - D,L Phe - D,L **Pro** - D,L Phe - D,L **Pro** - L Gln]

Figure 1. 384-member scaffold library of cyclic hexa- (rows 1–4) and heptapeptides (rows 5–8), categorized by ring size and number and position of proline residues. Within each sublibrary all diastereomers were included.

* Corresponding author. Tel.: +1 831 459 1307; fax: +1 831 459 2935; e-mail: lokey@chemistry.ucsc.edu

0960-894X/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2005.07.089

L- and D-amino acids with variable placement of proline residues, as well as an approach for exploiting cyclodimer formation toward the identification of a more potent inhibitor of actin assembly.

The scaffold library was designed as a set of eight hexa- and heptapeptide sublibraries categorized by ring size and proline position (Fig. 1). The side chain of phenylalanine was used as the generic R-group, and a lysine residue was incorporated into each sequence to facilitate analysis by HPLC and electrospray mass spectrometry (LC/MS). Within each sublibrary all possible diastereomers were included, yielding 384 unique sequences. The compounds were synthesized in parallel using standard Fmoc chemistry, with anchorage to the solid phase via a Rink linker through the glutamine side chain. The C-terminal Gln residue was protected as the allyl ester during chain elongation. Upon completion of the linear synthesis, the C- and N-termini were deprotected and the cyclization step was performed using PyBop/HOAT. After deprotection and cleavage from the solid phase, products were analyzed individually by LC/MS.

The relative yields of monomeric cyclic peptides (based on integration of the chromatogram traces) varied considerably among the different samples, ranging from 0 to >99%. Among the by-products were significant amounts of cyclo-oligomers, mostly cyclodimer,[8] in addition to linear oligomers. A random sample of 30 peptides was analyzed after the linear synthesis was complete but before cyclization. By LC/MS each sample was >90% pure with a mass corresponding to the expected linear peptide, indicating that differences in purity among the final products were mainly determined in the cyclization step.

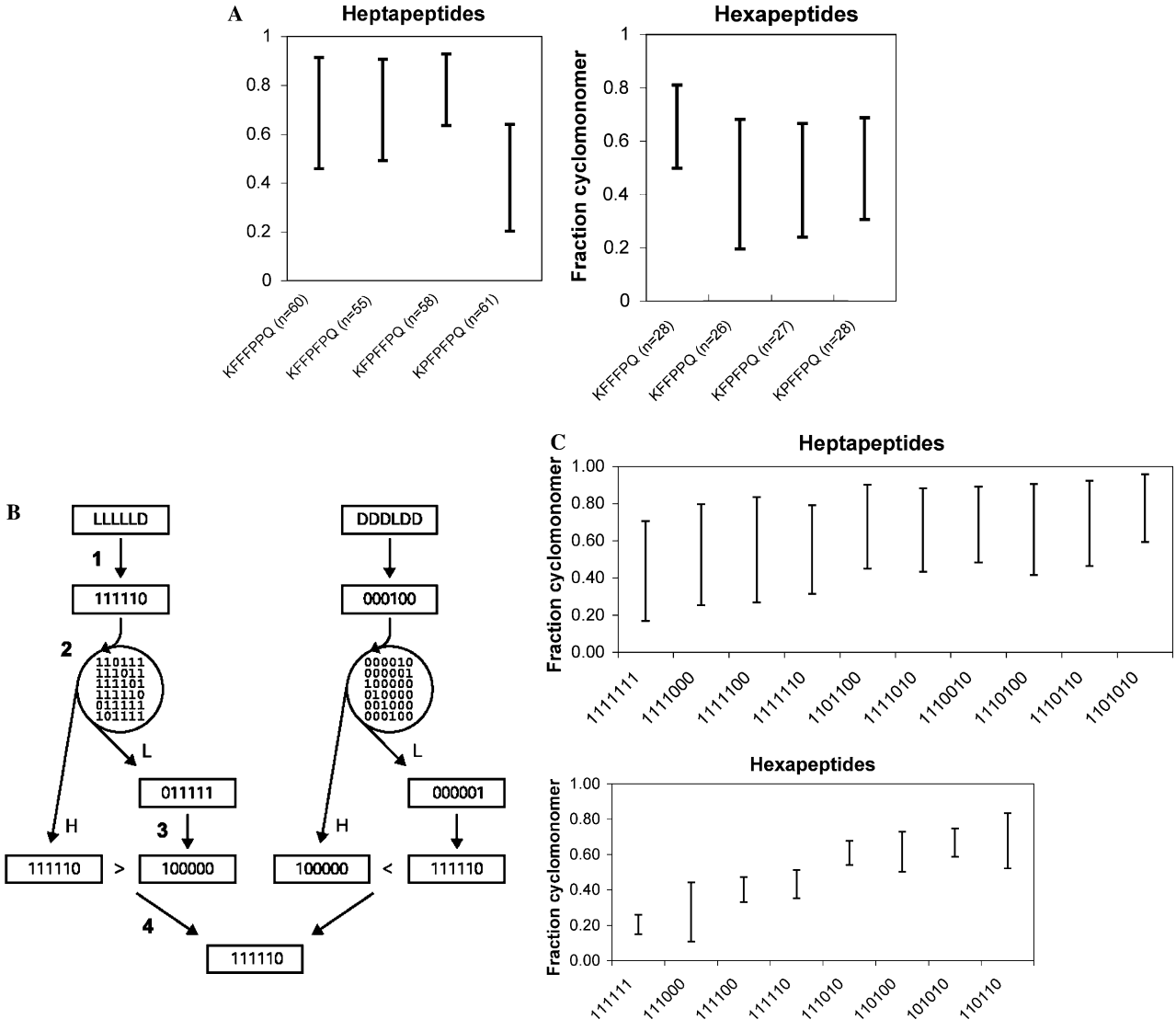

Figure 2. (A) Cyclization efficiency data for scaffold sublibraries. Plots indicate one standard deviation about the mean. (B) The stereoindex algorithm for grouping cyclic peptides having backbones identical up to rotation and reflection: (1) The linear backbone stereochemistry is encoded as a binary number, where $L = 1$ and $D = 0$. (2) The highest and lowest numbers were obtained from all circular permutations of the digits. Backbones which are similar up to rotation will yield the same pairs of numbers. (3) The binary complement of the lesser number was determined. Enantiomeric cyclic peptides will yield the same pairs. (4) The greater of the pair is defined as the stereoindex. (C) Cyclization efficiency as a function of stereoindex.

Marfey's test[9] was performed on 40 of the samples taken from both high- and low-yielding scaffolds, and there was little (av 8.8 ± 3.8%) racemization at the C-terminal Gln residue. For those samples that gave low yields of cyclomonomer, the extent of racemization was not significantly higher than for high-yielding scaffolds.[10] This suggests that even for those scaffolds with sluggish cyclization rates due to unfavorable backbone geometry, oligomerization on the resin is faster than C-terminal racemization.

We examined the dependence of cyclization efficiency on backbone geometry in more detail. Eighty-one out of the 384 sequences cyclized with greater than 80% efficiency to give the monomeric cyclic product, while a total of 183 sequences cyclized with greater than 60% efficiency. We set out to evaluate the dependence of cyclization efficiency on backbone geometry, beginning by testing the simple hypothesis that the variation of proline count and position might have a significant effect on cyclization efficiency within our hexapeptide and heptapeptide sublibraries. Heptapeptides in general cyclized more readily than hexapeptides, and within these groups the scaffolds with a greater number of prolines cyclized with markedly lower yields than scaffolds with fewer prolines (Fig. 2A). In both the hexapeptide and heptapeptide libraries, the scaffolds with two prolines can be further subdivided according to proline position; the differences in cyclization efficiency between these groups could not be proven statistically significant by one-way analysis of variance (ANOVA).

In order to test the hypothesis that the absolute pattern of stereocenters around the peptide backbone is predictive of cyclization efficiency, we developed an algorithm for partitioning the library into sets of cyclic peptides whose backbones are circular permutations or enantiomers of each other[11] (Fig. 2B). This parameter, which we term the 'stereoindex', proved to have a strong influence on cyclization efficiency among the hexapeptides (Fig. 2C; $P = <0.001$). Although still statistically significant ($P = 0.006$), the dependence is less pronounced for heptapeptides (Fig. 2C), presumably reflecting the generally greater ease of cyclization of the larger ring.

Each high-yielding sequence potentially represents a unique scaffold from which a combinatorial library could be constructed. To test the dependence of cyclization efficiency on side-chain variation, individual split-pool libraries were synthesized based on nine scaffolds with a range of cyclization efficiencies chosen from the original 384-member library (Fig. 3A). Eight side chains with varied functionality and steric bulk were chosen (Ala, Leu, Arg, Met, Trp, Thr, Phe, and Asn). The nine libraries were synthesized simultaneously in a 96-well polypropylene filter block (Robbins Scientific) on 450 μ-diameter polystyrene beads using the diketopiperazine linker of Lebl et al.[12] (Fig. 3B). The efficiency of cyclization was analyzed for 30 beads from each library by LC/MS using essentially the same methodology as for the scaffold library. The mass of each identified cyclomonomer was verified to correspond to an expected library product.

The resulting distributions suggest that scaffold geometry can outweigh the effect of side-chain variation in determining the efficiency of formation of monomeric cyclic peptides, although the effect is not as pronounced as one might expect (Fig. 4). The relative percentages of cyclic monomer[13] were clustered according to scaffold, with cyclization efficiencies for each scaffold varying from 10 to 30 percentage points about the mean. For scaffolds 6–9, cyclization efficiencies were >50% for all but two side-chain combinations, suggesting that synthetically robust libraries can be generated from these scaffolds. Among scaffolds 2–5, yields were generally lower, although the majority of side-chain combinations had yields in excess of 40%. The proportion of cyclomonomer product was markedly lower for many scaffolds than observed in the original all-phenylalanine library (Fig. 4, circles), suggesting that yield in the all-Phe library was not an ideal predictor of cyclization efficiency in its corresponding split-pool library. For scaffolds 2, 4, 5, and 6, the transition state of the cyclization step may be less tolerant of subtle differences in backbone conformation induced by certain combinations of side chains. On the other hand, in scaffolds 7, 8, and 9, the cyclization step may be more tolerant of side-chain differences, thus giving high cyclomonomer ratios for multiple residue combinations. The poor performance of scaffold 1 in the original library, in contrast, did predict the low yields observed in the corresponding split-pool library. This study suggests that while backbone features may not be the sole determinants of yield, a significant subset of the high-yielding sequences from the original scaffold library can be expected to cyclize efficiently upon side-chain variation.

The above studies identified a set of scaffolds that provided high yields of cylcomonomer; however, in some cases the cyclodimer (or cyclo-oligomer) may be the desired product. For example, we screened the 384-member scaffold library in an assay that models actin assembly in *Xenopus* egg extract[7] and identified the cyclodimeric byproduct corresponding to the sequence cyclo[L-Lys-D-Phe-D-Pro-D-Phe-L-Phe-D-Pro-L-Gln]$_2$ as the active inhibitor (IC_{50} = 2 μM). The corresponding cyclomonomer had no inhibitory activity up to 100 μM. We therefore set out to identify inhibitors with improved potency based on this cyclodimeric scaffold by generating a 512-member library in which the Phe residues were varied, keeping the stereochemistry, ring size, and proline residues constant. Since our previous results indicated that scaffold geometry generally outweighs side-chain identity in determining cyclomonomer/cyclo-oligomer ratios, we took advantage of the relatively inefficient cyclomonomer formation of this scaffold to create a split-pool library of cyclodimers (Fig. 3). After cyclization, the beads were treated with a standard TFA cleavage cocktail to remove protecting groups and set up the linker for auto-cleavage. The beads were arrayed into three 384-well plates. DMSO containing 0.1 M *N*-methylmorpholine and 0.1 M acetic acid was added to each well to induce compound cleavage, generating 1152 separate stock solutions. Using a plastic pin array, stock solutions were diluted approximately 1/100 into assay plates containing *Xenopus* egg extract and pyrene-labeled actin. We calculated a final

A

1: cyclo[L Lys - D Pro - **D** X^1 - D Pro - **L** X^2 - L Pro - L Gln]

2: cyclo[L Lys - **D** X^1 - D Pro - **D** X^2 - **L** X^3 - D Pro - L Gln]

3: cyclo[L Lys - **L** X^1 - **D** X^2 - L Pro - **D** X^3 - D Pro - L Gln]

4: cyclo[L Lys - **L** X^1 - **L** X^2 - L Pro - **D** X^3 - L Gln]

5: cyclo[L Lys - D Pro - **L** X^1 - **D** X^2 - L Pro - L Gln]

6: cyclo[D Lys - **L** X^1 - **D** X^2 - L Pro - **L** X^3 - L Pro - L Gln]

7: cyclo[D Lys - **L** X^1 - D Pro - **L** X^2 - **L** X^3 - L Pro - L Gln]

8: cyclo[L Lys - D Pro - **L** X^1 - **L** X^2 - L Pro - L Gln]

9: cyclo[D Lys - **D** X^1 - **L** X^2 - **L** X^3 - D Pro - L Gln]

B

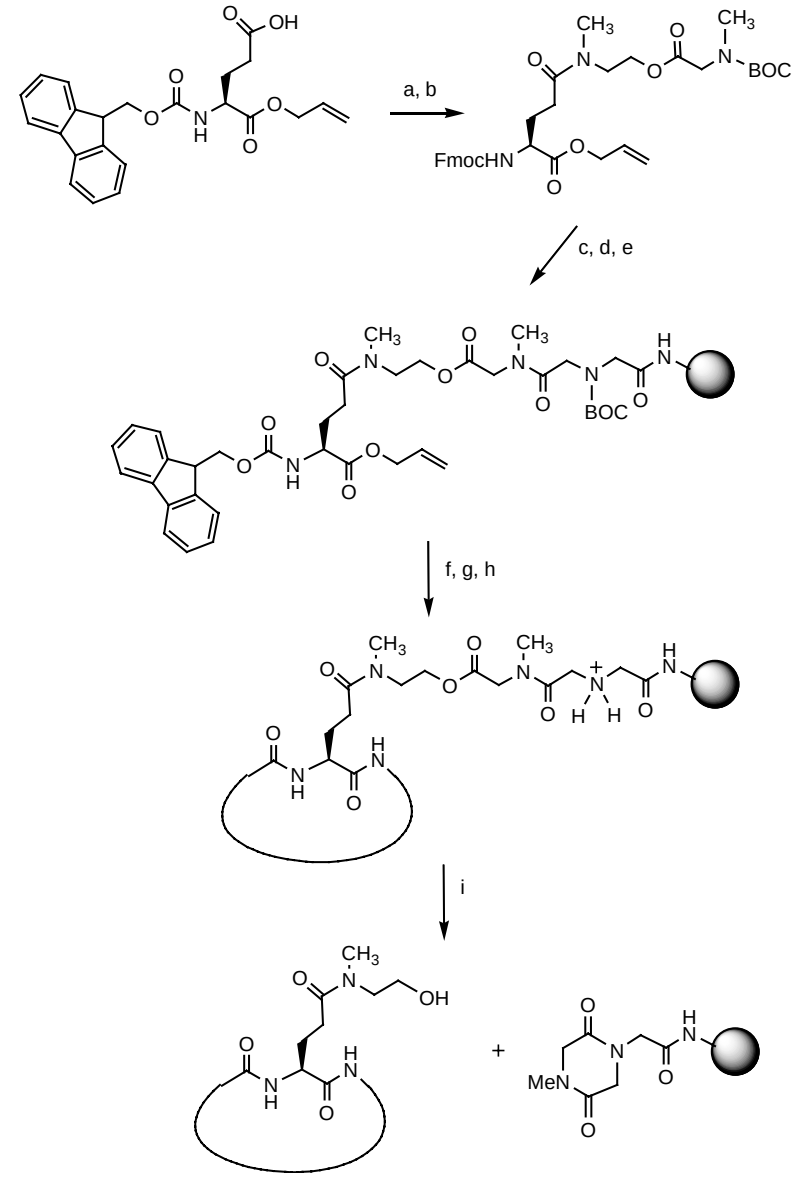

Figure 3. (A) Scaffolds selected for combinatorial expansion in which Phe residues were replaced with variable side chains X (X = Ala, Leu, Arg, Met, Trp, Thr, Phe, and Asn). (B) Synthetic scheme for split-pool synthesis of side-chain variants. Reagents: (a) DCC, HOBT, $HO(CH_2)_2N(H)Me$; (b) BocN(Me)Gly-OH, EDC, DMAP; (c) TFA; (d) *N*-Boc-iminodiacetic anhydride; (e) HBTU, DIPEA, amino-PS beads, 450 μ; (f) SPPS by Fmoc method; (g) Pd(Ph3P)4, *N*-methylmorpholine, HOAc followed by PyBOP/HOAT; (h) TFA, 5% tri-isopropylsilane, 2.5% phenol; (i) beads arrayed into 384-well plates and DMSO (15 μL) containing 0.1 M NMM/0.1 M HOAc was added to each well.

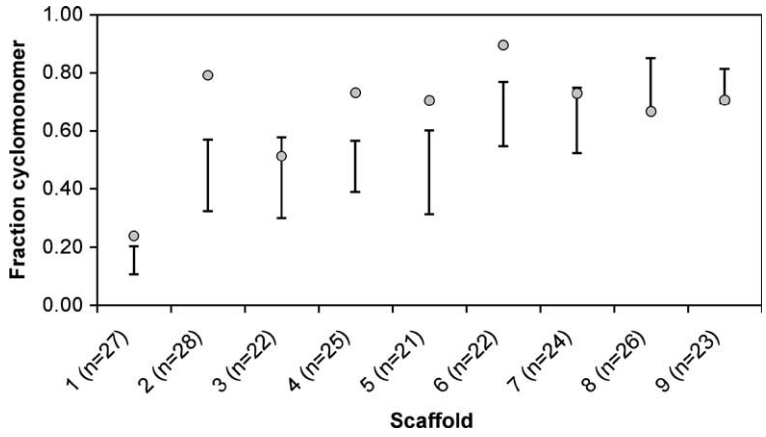

Figure 4. Cyclization efficiency of scaffolds after functional variation by split-pool synthesis. Plot indicates one standard deviation unit about the mean. ● indicates cyclization efficiency of corresponding scaffold in original 384-member scaffold library (X = Phe). Scaffold numbering is the same as in Figure 3A.

Table 1. IC_{50} values of sequence variants identified as being active in the *Xenopus* actin polymerization assay

Index of active well	Mass of crude product	Residue combination (X1–3)	Actin polymerization inhibition IC_{50} (μM)
D01	955.1	Trp, Phe, Leu	1–10 (±2)
N23	1037.6	Trp, Trp, Arg	0.5–20 (±2)
K02	1067.7	Trp, Trp, Trp	0.3 (±0.1)

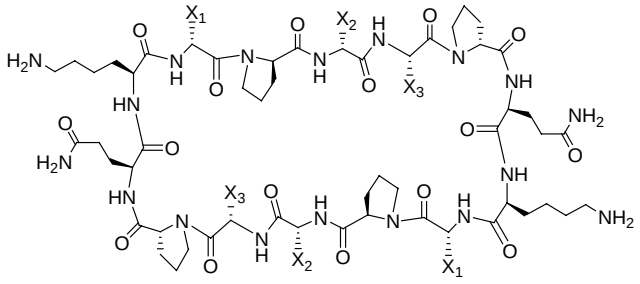

Figure 5. Structure of scaffold used in exploration of side chain influence on inhibitory potency in *Xenopus* actin polymerization assay.

compound concentration of approximately 1 μM, based on the resin loading and the estimated yield of cyclodimer from the original scaffold synthesis.

Three wells from the split-pool library were identified that significantly inhibited actin assembly in the *Xenopus* actin assay. LC/MS analysis of each of the crude products gave a unique mass, corresponding to the residue combinations shown in Table 1 (also see Fig. 5 for scaffold structure). The 16 possible sequence permutations corresponding to the three mass-decoded combinations (R1–3 = Trp3; R1–3 = Trp2Arg; X1–3 = Phe, Leu, Trp) were resynthesized via an alternate route: linear 14-mers were generated on low-loading resin, followed by cyclization to yield the cyclic 14-mer as the major product. IC_{50} values for each sequence permutation were determined in the *Xenopus* actin polymerization assay (Table 1). The most potent permutation was the compound in which the three Phe residues were replaced by Trp, with an IC_{50} of 0.3 μM, which represents a 7-fold improvement in potency over the original scaffold.

We have thus developed an iterative strategy in which a bioactive scaffold is identified from a small, geometrically diverse library, and then elaborated in a second step by varying side-chain functionality. There is certain to be a non-linear component to the relationship between side-chain functionality and scaffold geometry, a hypothesis supported by the fact that cyclization efficiencies among the split-pool libraries did not correlate perfectly with those of their corresponding all-Phe scaffolds. Our results indicate, however, that a stepwise approach to screening can lead to rapid improvement in potency when the screening of large libraries is prohibitive.

This study opens up a number of questions. First, what is the structural basis of the differences in cyclization efficiency that we observe? Can cyclization efficiency be predicted based on the rigidity and/or stability of the cyclic product, or are the effects of structure on the cyclization reaction only observable by conformational analysis of the linear precursors (i.e., to what extent is the cyclization transition state- or product-like)? We have demonstrated a statistical method for categorizing scaffolds and correlating various parameters with respect to cyclization efficiency, and have shown that conformational parameters and side-chain functionality can both influence cyclization efficiency. Future studies will be directed toward developing a structural hypothesis of cyclization efficiency, using our cyclization data in conjunction with modeling studies and solution state NMR. The discovery of a bioactive compound in an 'undesired' product, that is, the cyclodimer, suggests that this side-reaction can be exploited to further increase the diversity of geometrically defined scaffold libraries. Since cyclodimer can, in most cases, be chromatographically separated from cyclomonomer, screening mixtures derived from this side-reaction poses little problem with respect to deconvolution.

Acknowledgments

We are grateful to the Harvard Institute of Chemistry and Cell Biology, the NIH (CA78048), and Merck and Co. for financial support.

References and notes

1. For exceptions, see Burke, M. D.; Berger, E. M.; Schreiber, S. L. *Science* **2003**, *302*, 613; Huang, X.; Liu, Z. *J. Org. Chem.* **2002**, *67*, 6731; Couladouros, E. A.; Strongilos, A. T. *Angew Chem. Int. Ed. Engl.* **2002**, *41*, 3677, 3521; Ding, S.; Gray, N. S.; Wu, X.; Ding, Q.; Schultz, P. G. *J. Am. Chem. Soc.* **2002**, *124*, 1594; Kwon, O.; Park, S. B.; Schreiber, S. L. *J. Am. Chem. Soc.* **2002**, *124*, 13402.
2. Dutta, A. S.; Gormley, J. J.; McLachlan, P. F.; Woodburn, J. R. *Biochem. Biophys. Res. Commun.* **1989**, *159*, 1114; Saviano, M.; Isernia, C.; Rossi, F.; Di Blasio, B.; Iacovino, R.; Mazzeo, M.; Pedone, C.; Benedetti, E. *Biopolymers* **2000**, *53*, 189.
3. Gilon, C.; Huenges, M.; Matha, B.; Gellerman, G.; Hornik, V.; Afargan, M.; Amitay, O.; Ziv, O.; Feller, E.; Gamliel, A.; Shohat, D.; Wanger, M.; Arad, O.; Kessler, H. *J. Med. Chem.* **1998**, *41*, 919.
4. Bradley, E. K.; Ng, S. C.; Simon, R. J.; Spellmeyer, D. C. *Bioorg. Med. Chem.* **1994**, *2*, 279.
5. Pettit, G. R.; Cichacz, Z.; Barkoczy, J.; Dorsaz, A. C.; Herald, D. L.; Williams, M. D.; Doubek, D. L.; Schmidt, J. M.; Tackett, L. P.; Brune, D. C., et al. *J. Nat. Prod.* **1993**, *56*, 260.
6. Pettit, G. R.; Srirangam, J. K.; Herald, D. L.; Ericksen, K. L.; Doubek, D. L.; Schmidt, J. M.; Tackett, L. P.; Bakus, G. J. *J. Org. Chem.* **1992**, *57*, 7217.
7. Peterson, J. R.; Lokey, R. S.; Mitchison, T. J.; Kirschner, M. W. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10624.
8. The cyclodimer and higher-order oligomers were easily separated by HPLC from the cyclomonomer in most of the samples, and their identification was confirmed by MS/MS. Also, to confirm the assignment of cyclodimers two examples were synthesized independently by cyclization of linear 12- or 14-mer linear precursors. In both cases, their LCMS profiles were identical to those of the cyclodimers found in the original library.
9. Marfey, P. *Carlsberg Res. Commun.* **1984**, *49*, 591.
10. For scaffolds with high (av 91%) and low (av 23%) yields of cyclomonomer, the average extent of C-terminal racemization was $10 \pm 4.8\%$ and $7.4 \pm 1.2\%$, respectively.
11. Partitioning by stereoindex was performed using a script written in the Perl programming language.
12. Lebl, M.; Krchnak, V.; Kocis, P.; Lam, K. S. Patent 94-US7012 9500165, 19940621., **1995**.
13. In every case the mass of the major product correlated with the mass of one of the sequences expected for that scaffold. Amino acid analysis with Marfey's reagent was performed on 18 samples and confirmed the stereochemistry and identities of the amino acids. The results of Marfey's test correlated in each case with the mass obtained from LCMS analysis.