Published in final edited form as:

Chem Commun (Camb). 2016 May 4; 52(36): 6038–6059. doi:10.1039/c6cc00617e.

Towards Vast Libraries of Scaffold-Diverse, Conformationally Constrained Oligomers

Thomas Kodadek* and Patrick McEnaney

Departments of Chemistry and Cancer Biology, The Scripps Research Institute, 130 Scripps Way, Jupiter, FL 33458 USA

Abstract

There is great interest in the development of probe molecules and drug leads that would bind tightly and selectively to protein surfaces that are difficult to target with traditional molecules, such as those involved in protein-protein interactions. The currently available evidence suggests that this will require molecules that are larger and have quite different chemical properties than typical Lipinski-compliant molecules that target enzyme active sites. We describe here efforts to develop vast libraries of conformationally constrained oligomers as a potentially rich source of these molecules.

Introduction

Foldamers, as defined originally by Gellman¹, are synthetic oligomers with backbones that exhibit “well defined and predictable folding propensities”. Some foldamers closely mimic secondary structures found in proteins, while some adopt novel folds not explored by nature. This article will focus on a particular slice of foldamer science, specifically the goal of developing this class of compounds as a rich source of chemical probes (and perhaps drug leads) that target biological macromolecules. This has only crystallized as a key goal in our minds over the last few years. In other words, we tried to get away with using “floppy” molecules as probes and, despite some progress, eventually hit a wall. Clearly, in order to identify high affinity, high selectivity chemical probes for most proteins, it will be necessary to focus on oligomers that incorporate significant conformational constraints. Such probes could come in two flavors. One is comprised of large molecules made from monomers with only modest intrinsic conformational constraints. These would take inspiration from biopolymers like proteins or RNA, which form secondary and tertiary structures that are stable under physiological conditions only if they are quite long, at least by organic chemistry standards. This is due to the fact that interactions between non-neighbors in the primary chain are required to stabilize the fold. Such molecules may be quite useful as probes for extracellular targets, but will often be too large to pass through the cell membrane. Therefore, to tackle the more general problem of probe development for intracellular targets, the challenge is to design foldamers of modest molecular mass, roughly in the 600–1300 Dalton range, from building blocks that do not compromise cell

*To whom correspondence should be addressed: Kodadek@scripps.edu.

permeability. This will require building blocks that are intrinsically more rigid, since the oligomers will be too short to rely on long-range intramolecular interactions.

We will discuss progress towards this general goal here, with a focus on cell-permeable compounds. We first present a historical perspective on how we got into this field and how it evolved in the first decade of the 21st century. We will then focus on work from several laboratories, including our own, that have more recently advanced the goals of developing high quality foldamer probes through both designer approaches as well as unbiased screening campaigns. Finally, we will look to the future and consider what an ideal system for foldamer design/discovery would look like and what will be necessary to achieve it. Because of the wealth of literature in this area, we make no attempt to be comprehensive, considering only a few examples that we believe to be highly illustrative of the points we wish to make, and therefore apologize in advance to authors whose meritorious work we do not have space to discuss.

Protein-binding oligomers circa early 2000s

About a dozen years ago, our research group began to work on developing methods to discover probe molecules for chemical biology far more rapidly and efficiently than was possible with the methods in place at the time. The impetus for doing so was that progress on our own biological projects, which at the time were mostly focused on the role of the proteasome in eukaryotic transcription^{2–6}, were being held up significantly because of the almost complete lack of good probe molecules for most proteasomal proteins, with the major exception of active site inhibitors, which are used clinically⁷. At that time, what we now consider to be traditional, robotically-driven high-throughput screening (HTS) facilities were rare in an academic setting and, even if we had access to one, obtaining all of the molecules that we would have ideally like to have had via this technology would have been far beyond our modest budget. Moreover, since many of the proteins that we wished to target were not enzymes, it was questionable that traditional small molecules would fit the bill. Rather than just shrugging our shoulders, we considered how one might set up a reliable and general system for the discovery of high affinity, high selectivity probe molecules for almost any protein of interest in 3–4 weeks at a cost of \$1,000 or less.

Obviously, this is a tall order, but in the mind of one of us (T.K.), it did not seem completely ridiculous based on the precedent provided by mining ligands from peptide libraries, usually by phage display⁸. In a phage display experiment (Fig. 1), up to 10⁹ phage particles, each of which displays multiple copies of a single unique peptide sequence fused to the end of a coat protein, are incubated with a target protein that carries some sort of tag, such as a metal-binding site (His₆) or an epitope recognized by a commercially available antibody (FLAG, HA, etc.), as shown schematically in Fig. 1. The protein is pulled out of solution by virtue of the tag. Phage particles that display a peptide ligand for the tagged protein come along for the ride and can be recovered by using the precipitate to infect bacteria, which amplifies the protein-binding phage. A few rounds of this type of experiment, sometimes with negative selection steps inserted into the workflow⁹, can provide a population of phages that is highly enriched for those that display good peptide ligands for the protein of interest. Many phage clones are then Sanger sequenced (this was before deep sequencing existed) to reveal the

identities of the putative binding peptides, which then must be synthesized and validated in biochemical binding experiments. This technology has been used to identify peptide ligands for many proteins that were structurally uncharacterized. Like everything else in science, phage display is not completely general. It fails sometimes and there are many tricks of the trade that must be learned in order to use the technology to good effect. But in general it is fair to say that this (and now more modern and powerful methods of peptide library synthesis and screening^{10, 11}) comes close to satisfying the high bar of a cheap, rapid and general system for the discovery of protein ligands.

Of course, there are limitations. Big ones. Simple linear peptides of the type isolated from a phage display experiment make poor probe molecules. They lack serum stability and generally display poor cell permeability. Highly germane to the present discussion, short peptides rarely form stable secondary structures. This “floppiness” surely limits their affinity for protein binding partners. Indeed, it is rare to identify linear peptide ligands that bind to proteins with K_D s below the high nM to low μ M range, whereas dissociation constants 100-fold lower than this are desirable. Finally, the chemical diversity of peptide libraries is rather limited. To a first approximation, one is limited to the 20 natural amino acids as building blocks for the oligomer because of the substrate tolerance of the ribosome. Therefore, the problem becomes how one can combine the best features of phage display and other biological screening systems (big numbers, simple binding assays, modest cost) with the strengths of small molecules (high cell permeability, more drug-like properties, greater structural diversity).

Many laboratories have attempted to do so by developing ways to incorporate unnatural units into ribosomally synthesized libraries. For example, tRNAs conjugated to unnatural amino acids were developed that are accepted by the ribosome, allowing insertion of side chains with functional groups different than those of the 20 canonical amino acids at one or a few positions^{12, 13}. This technology has been particularly useful in developing proteins that can be tagged at a specific position. To address the important issue of floppiness, several clever strategies have been developed to post-translationally modify peptides in the library so as to limit their flexibility, most of which involve macrocyclization^{14, 15}. A representative example is the “bicycle” strategy of Heinis and Winter¹⁶, in which three cysteines are placed at invariant positions in an otherwise random peptide library displayed on phage. The phage population is treated with the tribromide shown in Fig. 2, which undergoes three thioalkylation reactions with the peptide cysteines, thus producing bicyclic peptides on the surface of the phage. These much stiffer molecules are capable of binding to some proteins with low nM K_D s¹⁷. To foreshadow a point that will be discussed in more detail later, Heinis and co-workers have demonstrated that, as one might imagine, altering the structure of the organic molecule used to zip up the bicycle forces the peptide into a very different conformation¹⁸, so some level of scaffold diversity is achievable in this system. Moreover, peptide cyclization increases the serum stability of peptides greatly relative to the linear molecule¹⁹. This is a combination of the lack of an end for exopeptidases to recognize, as well as the reduced propensity of a cyclic molecule to be able to adopt the extended conformation that most endopeptidases require for cleavage.

Alternatively, some groups turned to synthetic chemistry. Almost all such efforts use the split and pool strategy (Fig. 3) for library construction pioneered by Lam²⁰ and Houghten²¹. The Lam process produces one bead one compound (OBOC) libraries in which each bead displays many different copies of a single peptide, assuming perfect chemistry. To screen these libraries, a labeled protein is incubated with the beads in the presence of an excess of unlabeled competitor proteins (to block non-specific interactions). After washing, the label is somehow visualized and beads that retain high levels of the target protein are picked. The identity of the bead-displayed peptide can then be deduced by Edman degradation or, as is now usually the case, by releasing the compound from the bead and sequencing it by tandem mass spectrometry²².

The major advantage of OBOC peptide libraries over biological techniques is that one can easily incorporate unnatural amino acids, including D amino acids, which the ribosome would not accept. Macrocyclization and bicycle formation can be employed in these systems to increase structural rigidity²³. On the other hand, OBOC libraries are much smaller than ribosomally synthesized peptide libraries. A gram of 90 μm TentaGel beads, a support commonly used for OBOC library synthesis and screening²⁴, contains about 2.3 million particles. One gram is a comfortable amount of beads to employ in a library synthesis, so synthetic OBOC libraries come nowhere near the gaudy numbers ($10^9 - 10^{14}$) achievable using ribosomally synthesized peptide libraries. Having said that, a million compounds is still a respectable number, competitive with the size of even the best traditional HTS collections. The real limitation is that peptides are still peptides with respect to cell permeability. The root problem is that the highly polar N-H bond of the amide bond is quite well hydrated. Desolvation of this group must occur in order for the molecule to pass through the membrane and this is unfavorable.

Libraries of non-peptidic compounds

Given this backdrop, it seemed to us that one way to go to solve the “protein ligand problem” was to extend OBOC technology to non-peptidic molecules that would retain many of the favorable protein-binding characteristics of peptides, including the ability to engage relatively shallow protein surfaces, but exhibit far greater cell permeability. What kind of non-peptidic molecules can be used in this system? First, since it is impossible to keep track of what molecule is on what bead in a split and pool synthesis, the molecule displayed by any bead deemed to be interesting in a screening experiment must be characterized de novo. Unfortunately, there is not enough compound on a 90 μm TentaGel bead to allow NMR to be employed for this purpose. Instead, much more sensitive techniques such as Edman degradation or mass spectrometry must be used. Both of these techniques have technical drawbacks that limit their utility, Edman degradation only works on α -amino acids and while tandem mass spectrometry is somewhat more flexible, it only works well with oligomers that fragment in a logical and pre-determined fashion. Therefore, only oligomers that are readily characterized by tandem mass spectrometry can be used unless encoding strategies are employed (vide infra). Second, split and pool synthesis demands the use of extremely efficient reactions. Over the course of a multi-step library synthesis, incomplete reactions or those that produce significant side products will result in a complex mixture of different molecules on the bead, making post-screening characterization

difficult, if not impossible. Not many organic reactions demonstrate such high efficiency over a broad spectrum of structurally distinct reactants. This has largely limited OBOC libraries to peptide-like compounds where the monomeric units are joined by amide bond formation.

Given these severe restrictions, many chemists turned away from split and pool synthesis to focus instead on the creation of libraries by split synthesis, where one can keep track of things, in order to be able to create libraries of far more sophisticated, structurally diverse molecules. A good example is the build-couple-pair strategy reported by Schreiber and co-workers²⁵ as part of their diversity-oriented synthesis (DOS) program²⁶. These libraries are more attractive to chemists given the beautiful, stereochemically diverse structures that can be accessed²⁷. But the libraries are far smaller than those accessible by split and pool techniques, with a top end in the neighborhood of 20,000 compounds and even that requiring an enormous effort and a cost well above that of an OBOC library.

There was a third possibility that, to some degree, promised to combine the strengths of split and pool synthesis with the much greater structural diversity accessible if one didn't have to worry about de novo compound characterization by mass spectrometry. This was the idea of encoded split and pool synthesis libraries. In an encoded library, the library molecule is linked physically to a second molecule whose structure informs that of the library molecule and can be deduced easily by highly sensitive methods, an approach first reported by Clark Still and colleagues²⁸. Amongst the various strategies suggested for making encoded libraries, the most attractive, in theory, was an idea promulgated by Brenner and Lerner all the way back in 1992²⁹. They suggested that covalently linked hybrids of a growing synthetic molecule and a DNA chain could be made. In each split, every synthetic operation would be followed by a corresponding ligation of an oligonucleotide whose sequence would encode the structure of the synthetic unit added at that step. Only then would the molecules be pooled. As will be discussed below, this paper was prescient, but in 1992 the technology did not really exist to pursue this strategy effectively.

An alternative idea was proposed by Lam in 2002 that was specifically tailored to TentaGel-based OBOC libraries (Fig. 4)³⁰. These beads are comprised of an amino-polystyrene (PS) core onto which is grafted a thick layer of amine-terminated polyethylene glycol (PEG). TentaGel beads are uniquely useful for OBOC synthesis and screening since they swell well in organic and aqueous solvents, are highly stable mechanically and strongly resist non-specific protein binding²⁴, thanks to the PEG coating. Since the outer PEG layer is quite hydrophilic, while the inner PS core is hydrophobic, by exposing the beads briefly to an aqueous solution containing succinimidyl Fmoc (Fmoc-OSu), only the amines in the surface layer are protected. The beads are then switched to an organic solvent and allowed to swell completely, allowing chemistry to be done on the bead interior. If this is peptide addition, for example, the amino acid is protected with a group orthogonal to Fmoc, such as Boc, and the surface amine groups are then exposed. At this point split and pool library synthesis commences. In each split, the surface amine is coupled to a building block that may not be amenable to mass spectrometry-based characterization. To encode that unit, the interior amine is exposed and a Boc-protected amino acid is coupled only to the interior of the bead. This process of differential deprotection is continued until the desired library is present on

the surface, encoded by a particular peptide sequence on the interior that is easily decipherable by tandem MS/MS. Since proteins can only penetrate a short distance into the PEG layer of TentaGel beads³¹ the encoding molecule is never exposed to the target. The major downside of this technique is that it is quite tedious with multiple steps and solvent exchanges at each split. For example, we constructed an encoded library of interesting macrocycles recently using this technology³² and the library construction required over a month.

OBOC peptoid libraries

After considering the information available at the time (circa 2002), our research group made the decision to explore OBOC libraries of peptoids (Fig. 5), a class of molecules created by Zuckermann and colleagues³³. Peptoids have many potential advantages for our purposes. They can be made using the remarkably facile “sub-monomer” method³⁴ in which each unit of the peptide is created in two steps: acylation of an amine with the activated ester of chloro- or bromoacetic acid, followed by displacement of the halide with a primary amine (Fig. 3). Both reactions are extraordinarily efficient on beads and there exist hundreds of commercially available primary amines, providing a great deal of diversity for library construction³⁵. Peptoids can be sequenced by Edman degradation²⁴ or mass spectrometry²². They are immune to peptidases and proteases³⁶. Moreover, it seemed likely that peptoids would be far more cell permeable than peptides due to the replacement of the highly hydrated N-H groups in the main chain with N-alkyl moieties, and this was later demonstrated to be true³⁷. With respect to cell permeability, peptoids are much more like typical organic molecules than they are like peptides. Unless the molecule carries multiple negative charges, almost all hexameric peptoids in a library that we created are highly cell permeable and even most octamers are as well³⁷.

Of course, a significant limitation of peptoids is that they lack generally lack conformational constraints³⁸. Indeed, peptoids are even less conformationally restricted than peptides because both the cis and trans amide bond conformations are populated³⁸, whereas native peptide bonds are mostly transoid. Therefore, one would imagine that there will be a high entropic price for peptoid binding to a target protein, since the protein-binding side chains will not be pre-organized into the geometry capable of binding to the protein. As will be discussed below, peptoids can be coaxed into forming folded structures, but this requires special circumstances^{39–45} that are not conducive to the creation of diverse combinatorial libraries.

Nonetheless, given the aforementioned numerous advantages of this compound class, we decided to see how far we could get with these compounds. At a minimum, it seemed reasonable to think that we could at least use peptoids as a vehicle to iron out the many technical kinks that had limited the utility of OBOC libraries. Indeed, while a detailed review of this issue is beyond the scope of this article, it is important to point out that when we began this effort, bead-based libraries had a bad reputation that persists to some extent even today⁴⁶. Suffice to say that a great deal of troubleshooting and platform optimization was required to get to the point of being able to identify protein-binding ligands fairly routinely^{24, 47–50}.

These efforts^{51–63}, as well as peptoid screening campaigns (Fig. 5) conducted by other groups^{64–72} have yielded a number of protein ligands and some have proven to be useful bioactive probes^{57, 58, 73}. However, with rare exceptions⁶⁴, peptoids have failed to provide protein ligands with low nM affinity. Of course, it is important to point out that obtaining low nM primary screening hits from any unbiased library of a few hundred thousand compounds would be uncommon. But more troubling is that: 1) even modest affinity binding in the low μ M region usually requires relatively long peptoids of 6–9 residues, and 2) unpublished efforts to optimize primary screening hits into compounds with low nM affinities without making them much larger⁵⁸ have not been fruitful so far. We interpret these observations to reflect the idea that a pre-requisite for high affinity binding and facile hit optimization is a relatively stiff scaffold. This point of view is common in the drug development community, where primary screens are usually focused on identification of a stiff scaffold that is suitable to bind the protein target. This scaffold is then elaborated in a subsequent medicinal chemistry effort to improve both the potency and pharmacokinetic properties. If this is true, in what direction should we and others interested in this problem turn to most efficiently discover much more conformationally constrained molecules that will serve as more potent and selective probes? It seemed clear that the answer was to move from peptoids to molecules with far greater conformational constraints, i.e. foldamers. Therefore, we began to assess what type of foldamers were available that might suit our purposes.

Biomimetic foldamers: Rational design of α -helix surrogates

While large combinatorial libraries of foldamers are uncommon (vide infra), many laboratories have focused on the rational design of foldamers that seek to mimic one of the binding partners in a native protein-protein interaction. Essentially, the goal is to graft the side chains of the binding interface of one of the proteins onto an unnatural, cell permeable scaffold.

This approach has seen success in the development of synthetic α -helix surrogates. Many protein-protein (not to mention protein-nucleic acid) complexes contain a critical α -helix at their interface. α -helices in proteins can be thought of as a cylindrical structure from which side chains protrude with a periodicity of 3.5 residues per turn. The side chains at residues i , $i+4$ and $i+7$, point in more or less the same direction and thus are well-positioned to interact with a target protein (Fig. 6). In native proteins, the residues on the opposite face of the helical cylinder often pack against other elements of the protein, thus affording stability to the helical fold. Synthetic scaffolds that point three side chains in the same direction with spacing between them approximately equal to that seen between the i , $i+4$ and $i+7$ residues are thus anticipated to be good α -helix mimics.

One of the earliest examples of this approach was the work of Hamilton and colleagues. They sought to create fully synthetic molecules capable of structurally mimicking the natural α -helix structure found in Bcl-x_L binding proteins. To achieve this a terphenyl scaffold was designed (Fig. 7). This scaffold exists in a staggered conformation, allowing the display of side chains with a spacing and geometry that mimics that of the i , $i+4$ and $i+7$ residues in a native helix quite well. A small collection of terphenyl molecules were

synthesized that included the critical side chains from the Bak and Bad proteins. This rational design approach led to the discovery of a helix mimetic capable of binding Bcl-x_L with an affinity of 114 nM⁷⁴.

Unfortunately, the approach behind this pioneering work is not likely to be employed in the future. The terphenyl scaffold is simply too hydrophobic and the synthesis required to construct these species is long, requiring 8–11 linear steps involving sequential Suzuki couplings and column chromatography at each step. It is certainly not suitable for the construction of even modest libraries.

Recognizing this, numerous groups expanded on this idea utilizing a benzamide scaffold. Boger and colleagues used this design element as the basis for the synthesis of a small library for the discovery of protein-protein interaction mimics^{75–77} (Fig. 8). While these modifications will result in a different geometry of side chain presentation as compared to the purely arene scaffolds, modeling suggested that they might still be effective. In the course of investigating these compounds, Boger and colleagues further modified the design, largely due to the desire to further increase water solubility, to a single central ring flanked by amino acids. A mixture library of a few thousand compounds of this type was made by solution phase synthesis (Fig. 8) and a compound that antagonized the p53 helix-Mdm2 interaction with a low μM IC₅₀ was isolated from it⁷⁸.

Lim and colleagues reported an interesting design that employed chemistry a good bit more practical than that described above (Fig. 9). By taking advantage of highly efficient solid-phase chemistry reactions the Lim group generated an alpha-helix mimetic based on a phenyl-piperazine-triazine scaffold.⁷⁹ The synthesis employed only a single column purification step at the end of the eight-step solid-phase synthesis, but even the crude product was routinely 90% pure.

This scaffold synthesis strategy was used to generate a small compound collection (36 members) with the aim to disrupt the protein-protein interaction between BH-3 and Mcl-1, an interaction dependent on a short alpha-helical peptide in BH-3. The compounds were analyzed using a competitive FP assay with fluorescently labeled BH-3 peptide and the best compound was found to bind Mcl-1 with a K_i of 7.3 μM . Importantly this compound was highly selective for the BH3 binding pocket on Mcl-1 as compared to the very closely related Bcl-x_L. The synthetic ease and the properties of the compounds generated make this strategy particularly applicable to the generation of significantly larger combinatorial libraries for screening. In addition to being easier to make, the molecules exhibit much better aqueous solubility than the highly hydrophobic terphenyl scaffold (predicted cLogPs of 0.96 and 6.02, respectively). Thus, it appears that this scaffold design will see use in the future.

A shared feature of both the Hamilton and Lim designs is that they rely on intrinsic conformational preferences through non-binding steric interactions of their synthetic building blocks to position the side chains in the appropriate geometry. This is in stark contrast to native α -helices, in which hydrogen bonding between the main chain amides at positions *i* and *i*+4 are critical for holding the structure together. This feature of these

foldamer designs was interesting to us and, as will be described below, we have employed different types of non-bonded steric constraints into novel foldamer designs.

One of the most successful and sophisticated examples of helix mimicry can be found in the recent work of Arora and colleagues. They described the rational, computation-based design and synthesis of compounds based on an oxopiperazine helix mimetic (OHM) scaffold (Fig. 10) and their biological activities^{80, 81}. The OHM scaffold takes advantage of an efficient solid-phase synthetic scheme utilizing Fmoc-protected amino acids to generate conformationally constrained oligomers with backbone chirality. Utilizing computational analysis of the hotspot residues at the interface of two important protein-protein interactions, p53/Mdm2 and HIF1 α /p300, a small compound collection was designed and synthesized. From this collection selective hits for both Mdm2 and p300 were discovered, with dissociation constants of 300 nM and 530 nM respectively. Importantly these designed compounds were highly selective for their protein of interest.

They also explored the ability of the p300 helix mimic to down regulate the hypoxia-inducible gene expression caused by the Hif1 α /p300 interaction. *In vitro* studies of gene expression utilizing cancer cell models demonstrated that the helix mimetic targeting the p300-HIF-1 α interaction down-regulated the expression of hypoxia-inducible genes responsible for promoting invasion, glycolysis and angiogenesis. The results were recapitulated in an *in vivo* mouse model, with compound treatment reducing the median tumor volume by roughly 50% as compared to controls, with no observable signs of toxicity.

The oxopiperazine scaffold is attractive because it is readily synthesized on the solid phase and has good solubility. It seems likely that more designer inhibitors of protein-protein inhibitors will likely arise from this elegant design.

There is much to learn from these efforts, as well as similar excursions into the design of mimics for other types of protein secondary structures. While the investigators who developed these scaffolds originally intended to simply graft onto them native side chains found in protein-protein interaction hotspots, there is no reason that they could not be used to display diverse side chains, both natural and unnatural, in a library format. The Lim design is certainly efficient enough to be used in library work⁸² and that may also be the case for the oxopiperazine scaffold⁸⁰. Indeed, we independently reported a highly efficient on-resin synthesis of oxopiperazines⁸³ and demonstrated that it is useful in the creation of libraries, though we did not attempt to string several such units together to create helices. Even in cases where the stepwise synthesis of these types of scaffolds is not good enough for split and pool synthesis, an alternative strategy would be to pre-synthesize and purify a single scaffold with three chemically differentiable groups on each of the three side chains. This molecule could be affixed to beads and then diverse side chains could be added to each site using the split and pool strategy. Several years ago, we employed exactly such a strategy to create a library of diverse macrocycles⁸⁴, thus avoiding the classical problem that the efficiency of macrocyclization is often highly dependent on the nature of the residues between the reactive ends of the chain.

With the above examples it is important to note that this designed helix approach is limited to situations in which the target protein has been crystallized with an α -helical binding partner. While there are surely many important instances where this is the case and we believe this work to be of great value, we are trying to address the more general problem of having a platform to tackle almost any protein target, even those for which no structural information is available.

Bioinspired foldamers: Accessing secondary structures different than those found in proteins

Another approach begins with the reasonable hypothesis that natural oligomers are not the only structures that can fold into discreet conformations. Thus, if oligomers are built from unnatural, conformationally constrained building blocks, interesting secondary structures might emerge that would be interesting and perhaps useful in probe discovery even if they are not designed to precisely mimic a biological fold.

Given our history working with peptoid libraries, we have watched with interest the development of an body of work from Kirshenbaum, Zuckermann, Baron, Blackwell and others demonstrating that peptoids can be coaxed into secondary structures (reviewed in ref.³⁸). The first example of this sort was the finding that peptoids with runs of α -chiral side chains will fold into a helical conformation (Fig. 11)^{85, 86}. More recently, methods to control the ratio of amide bond isomers have been uncovered, allowing one to “dial in” mostly the cis or trans isomer^{40, 41, 44, 87}. However, the structural requirements to force peptoids adopt these folded structures are rather stringent and either dictate what the structure of almost every side chain would be or require couplings that are not efficient enough to support the construction of high quality combinatorial libraries. Indeed, to the best of our knowledge, there are no reports of the synthesis of a library of structured peptoids, nor have ligands to proteins been identified using these design strategies.

A different class of unnatural foldamers that has seen application to protein binding is β -peptides, originally developed by the Gellman¹ and Seebach^{88, 89} groups. These compounds have two carbons atoms in between the carbonyl and amino groups of the amino acids, rather than one as is the case in natural amino acids. β -peptides fold into distinct structures, which are dictated by the sequence of the molecule. A detailed overview of this interesting field is beyond the scope of this article, but β -peptide-containing oligomers are clearly of interest, because Gellman and colleagues have developed synthetic routes to them that are efficient enough for use in library preparation by parallel synthesis⁹⁰ and Schepartz and co-workers have reported the creation and screening of an OBOC β -peptide library⁹¹. Thus β -amino acids will almost certainly be valuable building blocks in the future for the creation of conformationally constrained libraries. This is especially true since advances in organocatalytic methods, including highly enantioselective Mannich reactions^{92, 93}, have made the synthesis of some members of this class of molecules far more straightforward.

Another advance in this field of great interest to us is that β -amino acid units have been used in conjunction with α -amino acid monomers to build hybrid structures, one of the few examples of an oligomer with more than one type of backbone element. The Gellman group

synthesized a 1000 compound library via split and pool synthesis to rapidly explore the structure activity relationship of certain key α and β units in a previously discovered hit with a 60 nM affinity for Bcl-x_L.⁹⁴ This library design maintained the spacing required for the display of key side chains that interact with the protein. Single library beads were arrayed into numerous 384 well plates, the compound released, and the crude cleavage products tested for their ability to inhibit the binding of BH3 to Bcl-x_L. From this screen several hits with mid-nM potency were isolated and information regarding structure activity relationships (SAR) was garnered.

Libraries of conformationally-restricted oligomers: PTAs and COPAs

The discussion above makes clear that in certain specific cases (i.e., where structures are available to guide the design of α -helix or other strict protein mimics), good inhibitors of protein-protein interactions can be designed. Moreover, as the β -peptide work shows, synthetic methods are emerging that will make libraries of certain bio-inspired, but unnatural, foldamer classes available for screening. As we considered our own plans in early 2010, shortly after moving to the Scripps Research Institute, we decided that rather than pursue an approach like those described above, for example exploring new foldamers comprised of a single, novel backbone element, we would try to make libraries of conformationally constrained molecules containing multiple scaffolds. As mentioned above, Gellman's work with peptides comprised of both α - and β -amino acid backbone elements represents one of the very few reports of libraries constructed with heterogeneous backbone oligomers, and even in this case, they were used in an ordered fashion (Fig. 12). We thought that using different backbone building blocks as a full diversity element in a library synthesis would be an interesting and, hopefully, useful extension of foldamer technology. As illustrated schematically for a hypothetical, peptoid-like library in Fig. 13, if the different backbone units employed in the split and pool synthesis are relatively stiff, the library will essentially contain many different scaffolds, each of which presents the side chains in very different orientations (Fig. 13).

This decision to move towards scaffold-diverse libraries reflected our desire to be able to use them as a source of ligands for almost any protein without the foreknowledge that there is a binding surface for α -helical ligands or some other defined secondary structure. While single scaffold libraries are interesting and elegant, if the scaffold on which it is built is simply not suitable for docking with a protein target, then nothing useful will come from a screen.

This lesson was learned in the early days of HTS, when one particular class of molecule, such as a benzodiazepine, dominated compound collections. These included an impressive number of compounds, but most were variations on a single theme in which diverse moieties were displayed from only a single scaffold, and would often fail to yield useful hits. Nowadays, the goal in assembling HTS collections is to include as much scaffold diversity as possible in the primary screening collection, with the goal of identifying a favorable scaffold for binding to the protein of interest. In a second, medicinal chemistry-driven screening campaign, the goal is to then diversify the groups displayed on that scaffold to improve the fit with the target protein. Translating this kind of thinking to the foldamer/oligomer world would suggest that the ideal primary screening library would consist of a

large number of different foldamer scaffolds, even it means reducing the diversity of side chains. In analogy with traditional HTS, the hope is that the primary screen would provide modest affinity hits that would reveal a favorable scaffold to target the protein of interest and that a “derivative library” could then be built using only that single scaffold, or perhaps also include modest variations of it, but with much greater side chain diversity⁹⁵.

Our first step toward this goal involved elaborating the synthesis of peptoids by employing new backbones in place of bromoacetic acid sub-monomers that imposed conformational constraints on the main chain, as suggested in Fig. 13. Like the Hamilton and Lim helix mimic designs, we chose to rely on local steric repulsion rather than longer range hydrogen bonding to achieve conformational restriction. We focused initially on using the well-established concepts of allylic 1,3 ($A_{1,3}$) strain to provide the desired constraint in two related classes of compounds called peptide tertiary amides (PTAs, a.k.a. N-alkylated peptides)⁹⁶ and chiral oligomers of pentenoic acids (COPAs)⁹⁷ respectively (Fig. 14A). In these compounds a chiral center is sandwiched between sp^2 -hybridized atoms. The conformation in which the hydrogen on the chiral center is in the same plane as the substituent on the double bond is highly favored due to $A_{1,3}$ strain, grossly limiting the conformational space of the chain (Fig. 14D). Therefore, by altering the absolute configuration of the stereocenters in PTAs and COPAs, one achieves different backbone “folds”, as illustrated in Fig. 14E for COPAs. In analogy with the cartoon at the bottom of Fig. 13, if one made a library comprised of a peptoid unit followed by three COPA units, using both enantiomers at each position, then a modestly scaffold-diverse library displaying the side chains in eight (2^3) distinct geometries would result.

As stressed above, for the purposes of OBOC library construction, an issue of paramount importance is the ease and efficiency with which these compounds can be made. COPAs were the brainchild of our former colleague Glenn Micalizio, who devised an elegant and efficient four step-synthesis of the required optically active sub-monomers⁹⁷ (Fig. 14B). Conditions for acylating secondary amines efficiently with these sub-monomers were developed, as were conditions for displacing the allylic chloride with diverse amines, allowing the facile construction of large libraries of COPAs^{97, 98} (Fig. 14C). Since screening hits are identified by mass spectrometry, it was necessary to mass encode the orientation of the stereocenter. This was accomplished either by using a deuterated substrate for the synthesis of one of the enantiomers (Fig. 14A), or, for the COPA sub-monomer, placing an ethyl substituent on the alkene in one enantiomer and a methyl group on the other. COPAs do not fragment as readily as common peptides or peptoids in MALDI tandem mass spectrometry, but fragmentation by electron transfer dissociation (ETD) allowed reliable MS-based assignment of COPA structures from the material available on a single bead⁹⁹.

PTAs can also be made using sub-monomer synthesis. Many amino acids can be transformed into 2-bromo carboxylic acids with retention of stereochemistry¹⁰⁰. These can then be used as sub-monomers in peptoid-like solid phase synthesis. However, as was well known in the literature from studies of N-methyl peptides, the synthesis of PTA oligomers is limited by fact that the acylation reactions are difficult due to steric congestion. Indeed, the use of highly active coupling reagents, such as BTC¹⁰¹, is essential to create high quality libraries⁹⁶. Even then, it is quite difficult to create PTAs longer than 3–4 units without some

break using a non-PTA unit to relieve steric congestion. Perhaps more unfortunately it was found not to be synthetically feasible to generate libraries in which any group larger than methyl is present at the chiral center. Thus, we have limited our efforts in this class to libraries of diversely substituted N-alkylated alanine oligomers (Fig. 14A). It is possible to include N-alkylated derivatives of many different amino acids into a library¹⁰², but in this case, the N-alkylated amino acid (where the side chain is larger than methyl) must be followed by a peptoid unit since chloroacetyl chloride is the only acylating agent that couples in high enough yield to the sterically congested amine to support the synthesis of high-quality libraries³².

Oligo-N-alkylated alanines are highly constrained conformationally. Indeed, we have been able to resolve conformers by HPLC at room temperature in some cases (Y. Gao and T.K., unpublished results). Even very short PTAs of four residues exhibit CD spectra clearly indicative of a folded conformation⁹⁶. COPAs are somewhat less stiff. The less crowded amide bond exists as a mixture of cis and trans isomers, as is the case in peptoids. Nonetheless, calculations indicate that for COPA oligomers, only a small number of major main chain conformers are predicted to be populated at room temperature (Fig. 14E). Therefore, by using the absolute stereochemistry of PTA and COPA sub-monomers as a diversity element, a modest number of diverse scaffold architectures can be included in a single library. The potential number of scaffolds in this case is n^2 , where n is the number of monomers in the oligomer library. To the best of our knowledge, the COPA- and PTA-based libraries that we have created represent one of the few examples of combinatorial oligomer libraries with more than one conformationally restricted scaffold.

High affinity and selectivity protein ligands from PTA- and COPA-based libraries

This is all well and good, but the real question is whether these more conformationally constrained, modestly scaffold-diverse libraries provide superior protein ligands. While only a handful of screens have been completed, the currently available evidence indicates the answer is yes. Three examples will be reviewed here.

The first screen involving COPAs was carried out against p53, a transcription factor of intense interest in cancer biology¹⁰³. Despite a great deal of effort, there are very few non-covalent small molecules that recognize p53 with good affinity and selectivity. It is considered an undruggable target. Recombinant p53 DNA-binding domain (DBD) was incubated with an OBOC library of 160,000 COPA tetramers and beads that retained the protein after thorough washing were picked. The same experiment was conducted with a library of peptoid tetramers. Importantly the same amines were employed to create both the COPA and peptoid libraries. Therefore, the only difference between the two is the conformationally restricted main chain in the COPA library. The best hit from the COPA library was found to bind to the p53 DBD with a K_D of 10 μM ⁹⁷. Binding appeared quite selective, since no interaction could be detected with several other proteins. While 10 μM may not sound terribly impressive, we considered this an excellent result since p53 is such a difficult target. More importantly, the peptoid library provided no p53 ligands,

demonstrating the superiority of the COPA architecture. To determine if the scaffold geometry was critical to binding, several diastereomers of the p53 ligand were synthesized and tested for binding. All showed greatly reduced affinity⁹⁷.

Another example was done with the library shown in Fig. 15, which was screened against immunoglobulins cloned from the B cells of chronic lymphocytic leukemia¹⁰⁴ (CLL) patients. The detailed biological underpinnings of this effort are not relevant to the theme of this article. But briefly, CLL is a disease in which a single clonal B cell proliferates relentlessly, eventually crowding out other B cells, forming tumors while at the same time and eventually compromising the normal function of the immune system. There exist good drugs for CLL, but these function by killing or blocking the activation of all B cells, not simply those that are pathogenic. We are interested in exploring a therapeutic strategy in which cytotoxic compounds or immune effector functions are delivered selectively to these pathogenic cells such they could be wiped out without affecting normal immune system function. To do so requires a ligand that would bind with high affinity and extremely high selectivity to a cell surface receptor highly restricted to these pathogenic cells. One target to consider in this vein is the antigen-binding site of the CLL BCR itself, which is found uniquely on the CLL cells and is not present on healthy B cells. Towards this end, the antibodies that we screened against were soluble IgG versions of CLL BCRs.

A OBOC library containing more than a million COPA molecules was first denuded of beads that displayed ligands to non-CLL IgGs. The remainder of the beads were then incubated with three CLL patient-derived monoclonal IgGs, each of which has a distinct antigen-binding site. Ligands for two of the three IgGs were identified, re-synthesized and characterized biophysically. The best of these ligands bound to the antibody target with K_D s of 400 nM and 600 nM, respectively, in solution, and 5- to 10-fold higher affinity in ELISA assays using immobilized compound, the difference presumably reflecting avidity effects. When these ligands were mounted onto a dextran oligomer that was also modified with biotin, highly selective binding to the CLL cells displaying the target antigen-specific BCR was observed. The compounds ignored B cells with different antigen-specific BCRs.

Finally, another antibody screen was conducted with the PTA-based OBOC library shown in Fig. 16, which contained 150,000 compounds¹⁰⁵. Note that three possible acid sub-monomers were used at each position, bromoacetic acid or either enantiomer of 2-bromopropionic acid. Thus, one could consider this library to have 3^4 , or 81 different scaffolds, if we count the conformationally unrestricted peptoid units. If one only considers the chiral sub-monomers, then the library contains $2^4 = 16$ scaffolds. In this case, the screening strategy involved first denuding the library of beads that retained IgG antibodies from the serum of healthy mice. The remainder of the library was then screened against serum collected from mice with Type 1 diabetes, the goal being to identify antibodies that would prove to be good serum biomarkers for this disease and, simultaneously, generate synthetic "epitope surrogates" that recognize them with sufficient selectivity and affinity to pull them out of serum, allowing their levels to be measured^{106–108}. In this case, a single ligand was isolated from the library (Fig. 16). This PTA was used an affinity reagent to highly enrich the antibodies to which it binds from the serum of diabetic mice, which then allowed us do relatively accurate biophysical measurements of its affinity for these

antibodies. Remarkably, in an ELISA format using immobilized PTA, the “ K_D ”¹⁰⁹ was found to be 2 nM, with excellent selectivity, as no binding was detected to antibodies from non-diabetic mice. This impressive binding probably reflects some level of avidity enhancement for the binding of the bivalent antibody to the immobilized small molecule, but in our experience, this is never worth more than 10–20-fold, thus likely placing the solution dissociation constant somewhere in the low-mid nM range, still an impressive result.

A thorough study of the importance of the main chain scaffold was conducted. We found that the des-methyl compound (i.e., the peptoid analogue of the screening hit shown in Fig. 15C) failed to bind the Type 1 diabetes-linked antibodies. Remarkably, so did every diastereomer that resulted from altering the absolute chirality of any of the three stereocenters in the molecule. This provides a striking example of the importance of conformational constraints in this small molecule-protein interaction. Given that two of the stereocenters in the screening hit are of the S configuration, while the other is in the R configuration, and all are absolutely critical for binding, the importance of scaffold diversity is clear. If we had used only one of the two chiral sub-monomers, giving a library with a single scaffold, the hit would not have been present in the library.

It is important to emphasize that these screens employed completely unbiased libraries. In other words, we had no way to know what building blocks to put into the library that would increase the chances of identifying a high affinity hit. Nor was any structural information available for the protein targets that would help in this regard. Thus, the identification of primary screening hits with high selectivity for the target and affinities in the nM range is quite encouraging and suggests that we are moving in the right direction with respect to our goal of developing a rapid and general system for the identification of high quality probe molecules.

Towards highly scaffold-diverse libraries

So where do we go from here? First, if having a few different scaffolds in the library is good, then increasing this number would be better. PTA- and COPA-based libraries are restricted to n^2 scaffolds (where n = units in the oligomer) since the absolute stereochemistry at the chiral center causes a distinctively different solution phase geometry. Therefore, one way forward is to develop more conformationally restricted, γ -halogenated acid sub-monomers with different geometries. Towards this goal, we have reported several heterocycle-containing sub-monomers¹¹⁰ (1–5 in Fig. 17). In unpublished work we have employed the bromobenzyl carboxylic acid (8 in Fig. 17) first reported by Lokey's group¹¹¹ in libraries, as well as a sulfonyl analogue of it (9 in Fig. 17), as well as o- and p-bromobenzyl isocyanates that introduce a urea molecule into the main chain (6–7 in Fig. 17). Unlike the COPA and PTA units, these sub-monomers are achiral and flat. They introduce conformational constraints by virtue of their ring structures.

We have also developed multi-step syntheses of interesting conformationally constrained ring systems on resin. For example, 2-oxopiperidine (OP) and diketopiperazine (DKP) units can be incorporated into oligomers very efficiently using the chemistry shown in Fig. 18. These units have the advantage of stiffening the chain by virtue of their cyclic structure and

also contain chiral centers, providing a site of variability in the geometry by which side chains are displayed from them.

Through ongoing efforts in our laboratory, we anticipate that this list of conformationally constrained sub-monomers will grow steadily. We call oligomers made from these building blocks and amines PICCOs (Peptoid-Inspired Conformationally Constrained Oligomers). However, even with the modest number of bromoacetic acid replacements that we have in hand now, a few interesting problems have arisen in thinking about making and exploiting PICCO libraries.

The first has to do with the numbers game inherent in split and pool library synthesis. As mentioned above, a gram of 90 μm TentaGel beads contains about 2.3 million particles. While scale-up could undoubtedly be pursued, our current protocols limit the number of beads in a single library synthesis to between 1–2 grams. Furthermore, one of the important lessons that we learned in optimizing bead-based screening protocols is that the best way to distinguish false positives from bona fide hits is to employ redundant libraries and only pay attention to compounds isolated on more than one bead in a screen. These are almost always good ligands for the target protein whereas the “singleton” hits are more often than not screening artifacts⁴⁸. Ideally, we like to use at least a five-fold redundant library. This constraint means that from one gram of 90 μm TentaGel the largest library that we can make will have 460,000 compounds. This is certainly a respectable number, but pales in significance compared to that which one would like to have to explore highly scaffold diverse libraries. For example, if we consider a library of peptoid-like tetramers made from 10 acid or acid-like sub-monomers and ten amine sub-monomers at each position, the theoretical diversity of this library would be 10^8 molecules, a number more than 200 times greater than the number of beads available. Even if a more modest tetrameric PICCO library design that included six acid sub-monomers would have 1296 distinct scaffold. This means that only four different amines could be employed to keep the theoretical diversity below 460,000 compounds.

Given that compromises must be made, what is the best way to go? High scaffold diversity, but low side chain diversity? Low scaffold diversity, but high side chain diversity? Or something in the middle? Our bias, based on analogy with the traditional mode of drug development that focuses on finding the optimal scaffold first, is the former. But it must be admitted that this is pure speculation, though we are working on ways to test this¹¹³ (*vide infra*).

An issue of more immediate significance is that PICCOs containing diverse backbone elements are much harder to characterize by tandem MALDI TOF-TOF mass spectrometry. Depending on the nature of the building block, the fragmentation efficiencies of the different types of amide bonds can vary dramatically. Thus, we have generally found that it is impossible to find enough intense fragment ions to unequivocally assign the structure of most backbone-diverse PICCOs. We are currently experimenting with other forms of fragmentation, including ETD in hopes of solving this problem, but whether or not this will be feasible is an open question. Thus, no matter what the answer to the interesting

theoretical question posed in the previous paragraph, we cannot currently use highly scaffold-diverse libraries.

DNA-encoded OBOC libraries

This brings us back to encoding. One option is to use Lam's two-phase strategy (see Fig. 4), using peptoids as the internal encoding elements since they fragment more efficiently than peptides^{32, 114}. But if one is building PICCO on the bead surface and peptoids in the interior (i.e., using the same basic chemistry in both domains) this becomes extremely tedious with several operations required at each split, including differential protection and deprotection of the exterior and internal amine groups³². Therefore, we have chosen to explore DNA encoding in collaboration with our colleague Brian Paegel.

Brenner and Lerner's original idea²⁹ has been realized experimentally over the last several years and DNA-encoded libraries are now used by a number of groups in both academia and industry^{115, 116}. A "headpiece" (Fig. 19), which has on one end a partially duplex DNA oligomer with a single-stranded overhang and an amine on the other end serves as the starting material for library synthesis¹¹⁷. After ligation of a template sequence to all of the molecules, the modified headpiece is carried through a split and pool sequence. In each split a chemical step is carried out, followed by the ligation of a partially duplex oligonucleotide whose sequence encodes the structure of the unit added in the chemical step. The molecules are pooled and the sequence is repeated. At the conclusion of the synthesis, another PCR primer template sequence is added. All of this is done in solution.

DNA-encoded libraries are screened in much the same way that a phage display library is. A protein of interest is mixed with the library and, after a suitable incubation, pulled down. The DNA-small molecule conjugates that come along are amplified by PCR using primers complementary to the two ends. All of these amplicons are then deep sequenced. The entire library is also sequenced. Encoding molecules that are highly enriched in the protein-associated fraction are taken to codes for ligands that bind the protein of interest. Note that because of the enormous capacity of deep sequencing (tens of millions of reads), one can combine hits from many different screens into one deep sequencing run. What hits are from what screen can be kept straight by using PCR primers for each hit pool that have different non-homologous 5' ends. This essentially provides a "barcode" that uniquely identifies the screen.

The Paegel laboratory adapted this technology to solid-phase synthesis¹¹⁸. The headpiece shown in Fig. 19A is added to a small percentage of the sites on the surface of TentaGel beads using Click chemistry. Split and pool solid-phase synthesis is then done with a chemical step and a ligation of an encoding DNA carried out at each split, following the precedent developed by solution-phase DNA-encoded library synthesis. The "encoding language" (i.e., the sequences of all of the encoding DNAs was designed carefully to eliminate problems due to hairpin formation or other troublesome secondary structures¹¹⁸. The single-stranded overhangs were designed to discourage inappropriate ligations (for example the coding units for position 3 in a library cannot be ligated to encoding units for

position 1). This process is continued until the library synthesis is complete and a reverse PCR template is added (Fig. 19B).

An important feature of the protocols now in use is that quality control (QC) is a major focus. The library is constructed on 10 μm TentaGel bead, which is highly advantageous for reasons discussed below, but makes it impossible to conduct lot sampling-based quality control checks on the synthetic quality of the library because there is so little material on each bead. Therefore, a few thousand 160 μm TentaGel beads with an acid-cleavable linker are included in the split and pool synthesis. Control experiments have demonstrated that the same chemistry proceeds with the same efficiency on the large and small beads¹¹⁸. The 160 μm beads can be separated easily from the 10 μm beads at any time and the purity of the compound on any single bead is easily checked by LC-MS analysis. The encoding DNA is Sanger sequenced to ensure that the DNA code predicts the mass of the major ion (Fig. 19C)¹¹⁸. It is difficult to overestimate the importance of library QC. If one enters into a screening campaign blindly and employs a library of poor quality, enormous amounts of time can be wasted. This is perhaps one advantage of using solid-phase synthesis for making DNA-encoded libraries, since there is no simple way to QC solution phase libraries. Another may be that DNA does not dissolve in most organic solvents, restricting what conditions can be used for solution phase synthesis. This is irrelevant on the solid phase.

DNA encoding is quite attractive for the construction of PICCO libraries for at least two reasons. The first, obviously, is that it eliminates the need for MS-based compound characterization of hits, making it completely feasible to use mixed backbones. The second is that there are far more 10 μm beads (≈ 2 billion) in a gram than there are 90 μm beads (≈ 2.3 million). So much larger libraries can be made.

A third advantage of this platform relative to screening with larger beads is that 10 μm TentaGel beads pass easily through a flow cytometer. This means that hits can be separated from non-hits by fluorescence activated "cell" sorting (FACS) (McEnaney, et al., in preparation), which is far easier and more efficient than manual picking or even magnetic separation. It also makes multi-color screening straightforward, which might be interesting if one is concerned about avoiding hits that have an affinity for particular off-targets. These off-targets could be included in the screen, but labeled with a different fluorochrome than the target protein. FACS instruments can be gated to only collect beads that exhibit a user-determined ratio of target/off-target binding. We have demonstrated the utility of this protocol in serum screening for diagnostic antibodies (Mendes, K., et al., in preparation).

The major downside of DNA-encoded libraries is that DNA is not Teflon. It can easily be damaged in such a way that it fails to serve as a template for PCR-based amplification. For example, strongly acidic or basic conditions are out of the question. This places constraints on the type of protecting groups that can be employed, amongst other things. Unfortunately for the synthesis of PICCO libraries, strong acylating agents also degrade the DNA to some extent. The more difficult the acylation, the more DNA damage is incurred. The Paegel laboratory has also developed a useful quantitative PCR-based assay to measure DNA damage during synthesis, allowing one to predict the percentage of the original templates that will survive library synthesis under a given set of conditions (Malone, M., et al.,

submitted). This is an important tool for planning a library synthesis with the expectation that there will be enough encoding DNA left on the beads at the end to allow compound identification. In general terms, we have found that the usual amount of DNA that we append to the beads¹¹⁸ will tolerate 3–4 acylations of secondary amines, or 5–6 acylations of primary amines under our standard conditions before the amplifiable level of DNA drops too low to be of use as an encoding tag. Some organometallic reactions, such as Suzuki couplings¹¹⁹, also inactivate about 50% of the DNA templates on the bead. Therefore, in the future it will be important to develop new library designs of different types of molecules that can be made with gentler reactions. For example, we have found that reductive amination¹⁰² is a “freebie” with little or no damage to the DNA. We are working on adapting to the solid phase various reactions that are efficient in solution under conditions that are not DNA damaging. For example a variety of organocatalytic reactions such as Mannich⁹³ and aldol¹²⁰ condensations may be useful in this regard. Since our laboratory is in the initial phases of this work and primary papers have not yet been submitted, a detailed discussion of these efforts would be premature. Nonetheless our preliminary efforts have lent confidence to the idea that DNA encoding will be a useful strategy for the synthesis of at highly scaffold-diverse synthetic libraries.

Interrogating library fitness

Finally, it is useful to consider whether we, or anyone else in the business of making combinatorial libraries or assembling screening collections really know what they are talking about when speculating about what makes one library better than another. For example, as discussed above, is it really better, on average, to have libraries with diverse scaffolds as opposed to libraries with only one, all other things being equivalent? Or, as one often hears at meetings, is it really important to “escape flatland” and develop libraries of molecules with more diverse three-dimensional shapes, more chiral centers, etc. in hopes of achieving better target selectivity¹²¹? These seem like appealing ideas, but what is the evidence to back them? Of course, the answer to this question almost certainly will be different for different target proteins. If a target has a pocket that is perfect for α -helix binding, then a library with this single scaffold and highly diverse side chains will likely provide more and better hits than a scaffold-diverse library in which an α -helix-like scaffold is only lightly represented or not present at all. Nonetheless, at least within a class of target proteins, such as kinases, GPCRs, PDZ domains, etc., it seems reasonable to believe that the question of what constitutes a superior screening collection is an important and interesting one.

Only anecdotal information of this type is available. Careful, quantitative comparisons of the “fitness” of two libraries that differ some well-defined way are almost non-existent. Some of our experiments have addressed this issue in a qualitative way. For example the p53 screen⁹⁷ was done with two libraries that were identical except that one had a COPA backbone and one a peptoid backbone. Only the COPA library provided bona fide ligands⁹⁷. In the antibody screen summarized in Fig. 16, only PTAs, not peptoids, were isolated as ligands although both were present in the same library¹⁰⁵. When the absolute stereochemistry of any of the chiral centers in the hit was changed, binding was abolished. These are fair comparisons between peptoids and chemically similar, but much more conformationally

constrained species. Thus, we feel justified in concluding that some degree of stiffness increases the fitness of an oligomer library.

But we were interested in developing a more comprehensive and quantitative assay for library fitness comparisons and recently published the first application of this system. We chose to ask if libraries of macrocycles, which are all the rage currently^{122, 123}, really perform better as a source of protein ligands than comparable libraries of linear compounds. This was investigated in the context of peptoids. Since peptoids are so floppy, one might imagine that this might be a best case for macrocyclization being beneficial. Indeed, studies of individual peptoid macrocycles have shown that they can be quite constrained conformationally¹²⁴.

Six libraries were synthesized, each containing four variable elements comprised of 15 different R groups (highlighted in red in Fig. 20A), resulting in a theoretical diversity of 15⁴, or 50,625 peptoids. The same amines were employed for all six libraries. Three of the libraries were subjected to macrocyclization while the other three were not. The three macrocyclic and three linear libraries, called C1–C3 and L1–L3, respectively, differing from one another in the number of atoms in the chain. This was done by inserting 0, 1 or 2 invariant peptoids (n=0 is C1 or L1, n=1 C2 or L2, etc.) into the oligomer following the variable elements (blue in Fig. 20A).

Ten copies of each library were screened separately against streptavidin (SA) as a simple model protein target. This was done by incubating the beads with SA conjugated to magnetic FeO nanoparticles (Dynabeads) and collecting the peptoid-displaying TentaGel beads that had bound significant FeO-SA using a strong magnet. These “hit beads” were then analyzed using the powerful platform developed by Auer and co-workers⁴⁹. The beads were stripped of protein using a denaturing buffer wash and placed into the wells of a 96 well microtiter filter plate (one bead per well). Each compound was labeled with azidofluorescein whilst still on the beads by Cu-catalyzed cycloaddition¹²⁵ to an alkyne side chain in the invariant linker (Fig. 20B), after which the compounds were released into solution and filtered away from the beads. A small amount of compound was used for structure determination by tandem mass spectrometry. The remainder was used in a plate-based titration experiment in which increasing amounts of SA or a control protein were added to the well and binding was tracked by an increase in fluorescence polarization (FP). This protocol⁴⁹ allows the K_D s of hundreds of small molecule-protein complexes to be determined without the need for hit re-synthesis. In our studies, we re-synthesized and purified many of the hits and found that the K_D for their complex with SA matched that measured in the high-throughput format quite well.^{95, 113}

The results of this analysis were clear cut¹¹³. The number and quality of SA ligands isolated from all of the libraries were similar, with the clear exception of C1, the macrocyclic library with the smallest ring size (i.e., no blue residues: Fig. 20A). This library yielded the highest affinity SA ligands and in greater numbers than the other five. Moreover, the linear analogues of the best cyclic ligands bind SA at least 100-fold more poorly. In contrast the difference between the binding constants of the cyclic and linear versions of the best hit from the largest ring library, C3, was only 3-fold. Finally, the best hits from library C1 had a

fundamentally different structure than those isolated from libraries C2, C3 and L1–L3. All of these data point toward macrocyclization affording little advantage with respect to, or differentiation from, the linear libraries for libraries C2 and C3, whereas, ring closure clearly differentiated the smallest macrocyclic library, C1, from the other libraries. Thus, only when the ring becomes small enough to impose significant conformational restriction on the molecule does a macrocyclic library afford an advantage over the equivalent linear library. To the best of my knowledge, this is the first “apples vs. apples” test of the utility of macrocyclization at the level of complete combinatorial library analysis. We believe that this approach, which provides a comprehensive analysis of the entire hit pool in a library screen, will be quite useful in comparing the utility of different kinds of foldamer libraries as a source of high affinity protein ligands. As an aside, it is also worth mentioning that the same protocol provides a powerful tool to analyze focused libraries designed around primary screening hits with the intention of identifying improved compounds⁹⁵. A tremendous amount of structure activity relationship information is obtained.

Summary and Conclusions

Our research group has come into the foldamer field by a quite different route than most. Most foldamer science is focused on the design of biomimetic folds and structural studies aimed at increasing our understanding of the factors that influence the degree of folding and the fine structure of the molecules. We, on the other hand, have turned to foldamers for utterly practical reasons having to do with our desire to develop a relatively inexpensive technology platform that will allow the routine and rapid identification of high quality protein ligands to be used as probe molecules and drug leads. Since protein-protein interaction surfaces are of great interest to us, collections/libraries of molecules with a large “wingspan” are preferred over traditional small molecules. As detailed above, a good deal of work from our laboratory comparing the properties of floppy peptoids with stiffer PICCOs as protein ligands has clearly confirmed the desirability of including significant conformational constraints into the molecules that are screened. Thus, foldamers are prime candidates for these libraries. In particular, we speculate that libraries containing many different types of foldamers (scaffold-diverse libraries) will be particularly useful as primary screening collections, though this hypothesis remains to be tested. This effort brings challenges. Since most investigators are focused on design and analysis of one or a few foldamer molecules, moderately efficient synthetic processes are acceptable. This is certainly not the case for split and pool library synthesis, which demands highly efficient reactions. So a good deal of work lies ahead of us with respect to synthetic optimization to make full use of many different foldamers in this application. Likewise, the challenge of characterizing molecules with chemically diverse backbone units in the same chain is not to be underestimated. DNA encoding is one solution, but limits the range of chemical reactions that can be employed. Finally, there is great opportunity and challenges in the development of new classes of foldamers inspired by biomolecules different than proteins, for example polyketides. While we can predict with confidence that nothing will be easy, we believe strongly that medium-sized foldamer molecules that incorporate the best features of small molecules and biologics have a bright future as probes and drug leads.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The work from our laboratory described in this proposal was supported by grants from the National Institutes of Health (N01-HV-00242, 1 DP3 DK094309-01) and a grant from DARPA's FoldRx program.

References

1. Appella DH, Christianson LA, Klein DA, Powell DR, Huang X, Barchi JJ Jr, Gellman SH. *Nature*. 1997; 387:381–384. [PubMed: 9163422]
2. Kodadek T. *J Biol Chem*. 2010; 285:2221–2226. [PubMed: 19955182]
3. Ferdous A, Gonzalez F, Sun L, Kodadek T, Johnston SA. *Mol. Cell*. 2001; 7:981–991. [PubMed: 11389845]
4. Gonzalez F, Delahodde A, Kodadek T, Johnston SA. *Science*. 2002; 296:548–550. [PubMed: 11964484]
5. Ferdous A, Sikder D, Gillette T, Nalley K, Kodadek T, Johnston SA. *Genes Dev*. 2007; 21:112–123. [PubMed: 17167105]
6. Kodadek T, Sikder D, Nalley K. *Cell*. 2006; 127:261–264. [PubMed: 17055428]
7. Goldberg AL. *J Cell Biol*. 2012; 199:583–588. [PubMed: 23148232]
8. Smith GP, Petrenko VA. *Chem. Rev*. 1997; 97:391–410. [PubMed: 11848876]
9. Han Y, Kodadek T. *J Biol Chem*. 2000; 275:14979–14984. [PubMed: 10809742]
10. Roberts RW, Szostak JW. *Proc. Natl. Acad. Sci. USA*. 1997; 94:12297–12302. [PubMed: 9356443]
11. Ohuchi M, Murakami H, Suga H. *Curr Opin Chem Biol*. 2007; 11:537–542. [PubMed: 17884697]
12. Wu X, Schultz PG. *J. Amer. Chem. Soc*. 2009; 131:12497–12515. [PubMed: 19689159]
13. Yoo TH, Tirrell DA. *Angew Chem Int Ed Engl*. 2007; 46:5340–5343. [PubMed: 17568466]
14. Goto Y, Ohta A, Sako Y, Yamagishi Y, Murakami H, Suga H. *ACS Chem Biol*. 2008; 3:120–129. [PubMed: 18215017]
15. Walensky LD, Kung AL, Escher I, Malia TJ, Barbuto S, Wright RD, Wagner G, Verdine GL, Korsmeyer SJ. *Science*. 2004; 305:1466–1470. [PubMed: 15353804]
16. Heinis C, Rutherford T, Freund S, Winter G. *Nature Chem Biol*. 2009; 5:502–507. [PubMed: 19483697]
17. Chen S, Heinis C. *Methods Mol Biol*. 2015; 1248:119–137. [PubMed: 25616330]
18. Chen S, Morales-Sanfrutos J, Angelini A, Cutting B, Heinis C. *Chembiochem*. 2012; 13:1032–1038. [PubMed: 22492661]
19. Pakkala M, Hekim C, Soininen P, Leinonen J, Koistinen H, Weisell J, Stenman UH, Vepsäläinen J, Narvanen A. *J Pept Sci*. 2007; 13:348–353. [PubMed: 17436344]
20. Lam KS, Salmon SE, Hersh EM, Hruby VJ, Kazmierski WM, Knapp RJ. *Nature*. 1991; 354:82–84. [PubMed: 1944576]
21. Houghten RA, Pinilla C, Blondelle SE, Appel JR, Dooley CT, Cuervo JH. *Nature*. 1991; 354:84–86. [PubMed: 1719428]
22. Paulick MG, Hart KM, Brinner KM, Tjandra M, Charych DH, Zuckermann RN. *J. Comb. Chem*. 2006; 8:417–426. [PubMed: 16677012]
23. Lian W, Upadhyaya P, Rhodes CA, Liu Y, Pei D. *J Am Chem Soc*. 2013; 135:11990–11995. [PubMed: 23865589]
24. Alluri PG, Reddy MM, Bachhawat-Sikder K, Olivos HJ, Kodadek T. *J Am Chem Soc*. 2003; 125:13995–14004. [PubMed: 14611236]
25. Uchida T, Rodriguez M, Schreiber SL. *Org Lett*. 2009; 11:1559–1562. [PubMed: 19275223]
26. Schreiber SL. *Science*. 2000; 287:1964–1969. [PubMed: 10720315]

27. Morton D, Leach S, Cordier C, Warriner S, Nelson A. *Angew Chem Int Ed Engl.* 2009; 48:104–109. [PubMed: 19016294]
28. Ohlmeyer MH, Swanson RN, Dillard LW, Reader JC, Asouline G, Kobayashi R, Wigler M, Still WC. *Proc. Natl. Acad. Sci. USA.* 1993; 90:10922–10926. [PubMed: 7504286]
29. Brenner S, Lerner RA. *Proc Natl Acad Sci U S A.* 1992; 89:5381–5383. [PubMed: 1608946]
30. Liu R, Marik J, Lam KS. *J. Amer. Chem. Soc.* 2002; 124:7678–7680. [PubMed: 12083920]
31. Hintersteiner M, Buehler C, Auer M. *ChemPhysChem.* 2012; 13:3472–3480. [PubMed: 22829563]
32. Morimoto J, Kodadek T. *Mol Biosyst.* 2015; 11:2770–2779. [PubMed: 26067000]
33. Simon RJ, Kania RS, Zuckermann RN, Huebner VD, Jewell DA, Banville S, Ng S, Wang L, Rosenberg S, Marlowe CK, et al. *Proc Natl Acad Sci U S A.* 1992; 89:9367–9371. [PubMed: 1409642]
34. Zuckermann RN, Kerr JM, Kent SBH, Moos WH. *J. Amer. Chem. Soc.* 1992; 114:10646–10647.
35. Figliozzi GM, Goldsmith R, Ng SC, Banville SC, Zuckermann RN. *Methods Enzymol.* 1996; 267:437–447. [PubMed: 8743331]
36. Miller SM, Simon RJ, Ng S, Zuckermann RN, Kerr JM, Moos WH. *Bioorg. Med. Chem. Lett.* 1994; 4:2657–2662.
37. Yu P, Liu B, Kodadek T. *Nat Biotechnol.* 2005; 23:746–751. [PubMed: 15908941]
38. Yoo B, Kirshenbaum K. *Curr Opin Chem Biol.* 2008; 12:714–721. [PubMed: 18786652]
39. Gorske B, Jewell SA, Guerard EJ, Blackwell HE. *Org. Lett.* 2005; 7:1521–1524. [PubMed: 15816742]
40. Gorske BC, Bastian BL, Geske GD, Blackwell HE. *J Am Chem Soc.* 2007; 129:8928–8929. [PubMed: 17608423]
41. Gorske BC, Blackwell HE. *J. Amer. Chem. Soc.* 2006; 128:14378–14387. [PubMed: 17076512]
42. Huang K, Wu CW, Sanborn TJ, Patch JA, Kirshenbaum K, Zuckermann RN, Barron AE, Radhakrishnan I. *J Am Chem Soc.* 2006; 128:1733–1738. [PubMed: 16448149]
43. Kirshenbaum K, Barron AE, Goldsmith RA, Armand P, Bradley EK, Truong KT, Dill KA, Cohen FE, Zuckermann RN. *Proc. Natl. Acad. Sci. USA.* 1998; 95:4303–4308. [PubMed: 9539732]
44. Shah NH, Butterfoss GL, Nguyen K, Yoo B, Bonneau R, Rabenstein DL, Kirshenbaum K. *J Am Chem Soc.* 2008; 130:16622–16632. [PubMed: 19049458]
45. Wu CW, Kirshenbaum K, Sanborn TJ, Patch JA, Huang K, Dill KA, Zuckermann RN, Barron AE. *J Am Chem Soc.* 2003; 125:13525–13530. [PubMed: 14583049]
46. Kodadek T. *Chem Commun (Camb).* 2011; 47:9757–9763. [PubMed: 21701754]
47. Olivos HJ, Baccawat-Sikder K, Kodadek T. *ChemBiochem.* 2003; 4:1242–1245. [PubMed: 14613120]
48. Doran TM, Gao Y, Mendes K, Dean S, Simanski S, Kodadek T. *ACS Comb. Sci.* 2014; 16:259–270. [PubMed: 24749624]
49. Hintersteiner M, Kimmerlin T, Kalthoff F, Stoeckli M, Garavel G, Seifert JM, Meisner NC, Uhl V, Buehler C, Weidemann T, Auer M. *Chem Biol.* 2009; 16:724–735. [PubMed: 19635409]
50. Mendes K, Ndungu JM, Clark L, Kodadek T. *ACS Comb Sci.* 2015
51. Alluri P, Liu B, Yu P, Xiao X, Kodadek T. *Mol Biosyst.* 2006; 2:568–579. [PubMed: 17216038]
52. Lee J, Reddy MM, Kodadek T. *Chem. Sci.* 2010; 1:48–54.
53. Lim H-S, Archer CT, Kodadek T. *J. Amer. Chem. Soc.* 2007; 129:7750–7751. [PubMed: 17536803]
54. Raveendra B, Hao W, Baccala R, Reddy MM, Schilke J, Bennett JL, Theofilopolous AN, Kodadek T. *Chem. & Biol.* 2013; 20:350–359.
55. Reddy MM, Bachhawat-Sikder K, Kodadek T. *Chem Biol.* 2004; 11:1127–1137. [PubMed: 15324814]
56. Reddy MM, Wilson R, Wilson J, Connell S, Gocke A, Hynan L, German D, Kodadek T. *Cell.* 2011; 144:132–142. [PubMed: 21215375]
57. Trader DJ, Simanski S, Kodadek T. *J. Amer. Chem. Soc.* 2015; 137:6312–6319. [PubMed: 25914958]

58. Udugamasooriya DG, Dineen SP, Brekken RA, Kodadek T. J. Amer. Chem. Soc. 2008; 130:5744–5752. [PubMed: 18386897]
59. Xiao X, Yu P, Lim H-S, Sikder D, Kodadek T. J. Comb. Chem. 2007; 9:592–600. [PubMed: 17530904]
60. Liu X, Dix M, Speers AE, Bachovchin DA, Zuhl AM, Cravatt BF, Kodadek TJ. Chembiochem. 2012; 13:2082–2093. [PubMed: 22907802]
61. Simpson LS, Burdine L, Dutta AK, Feranchak AP, Kodadek T. J Am Chem Soc. 2009; 131:5760–5762. [PubMed: 19351156]
62. Simpson LS, Kodadek T. Tetrahedron Lett. 2012; 53:2341–2344. [PubMed: 22736872]
63. Gocke AR, Udugamasooriya DG, Archer CT, Lee J, Kodadek T. Chem. & Biol. 2009; 16:1133–1139. [PubMed: 19942136]
64. Zuckermann RN, Martin EJ, Spellmeyer DC, Stauber GB, Shoemaker KR, Kerr JM, Figliozzi GM, Goff DA, Siani MA, Simon RJ, et al. J. Med. Chem. 1994; 37:2678–2685. [PubMed: 8064796]
65. Heizmann G, Hildebrand P, Tanner H, Ketterer S, Pansky A, Froidevaux S, Beglinger C, Eberle AN. J. Recept. Signal Transduct. Res. 1999; 19:449–466. [PubMed: 10071777]
66. Garcia-Martinez C, Humet M, Planells-Cases R, Gomis A, Caprini M, Viana F, De La Pena E, Sanchez-Baeza F, Carbonell T, De Felipe C, Perez-Paya E, Belmonte C, Messeguer A, Ferrer-Montiel A. Proc Natl Acad Sci U S A. 2002; 99:2374–2379. [PubMed: 11854530]
67. Humet M, Carbonell T, Masip I, Sanchez-Baeza F, Mora P, Canton E, Gobernado M, Abad C, Perez-Paya E, Messeguer A. J Comb Chem. 2003; 5:597–605. [PubMed: 12959560]
68. Luo Y, Vali S, Sun S, Chen X, Liang X, Drozhzhina T, Popugaeva E, Bezprozvanny I. ACS Chem Neurosci. 2013
69. Marquez N, Sancho R, Macho A, Moure A, Masip I, Messeguer A, Munoz E. Biochem Pharmacol. 2006; 71:596–604. [PubMed: 16405868]
70. Mas-Moruno C, Cruz LJ, Mora P, Francesch A, Messeguer A, Perez-Paya E, Albericio F. J Med Chem. 2007; 50:2443–2449. [PubMed: 17432841]
71. Masip I, Cortes N, Abad MJ, Guardiola M, Perez-Paya E, Ferragut J, Ferrer-Montiel A, Messeguer A. Bioorg Med Chem. 2005; 13:1923–1929. [PubMed: 15727848]
72. Masip I, Perez-Paya E, Messeguer A. Comb Chem High Throughput Screen. 2005; 8:235–239. [PubMed: 15892625]
73. Lim HS, Archer CT, Kodadek T. J Am Chem Soc. 2007; 129:7750–7751. [PubMed: 17536803]
74. Kutzki O, Park HS, Ernst JT, Orner BP, Yin H, Hamilton AD. J. Amer. Chem. Soc. 2002; 124:11838–11839. [PubMed: 12358513]
75. Ahn J, Han S-Y. Tet. Lett. 2007; 48:3543–3547.
76. Plante J, Campbell F, Malkova B, Kilner C, Warriner SL, Wilson AJ. Org. Biomol. Chem. 2008; 6:138–146. [PubMed: 18075658]
77. Sarogi I, Incarvito CD, Hamilton AD. Angew. Chem. Int. Ed. 2008; 47:9691–9694.
78. Shaginian A, Whitby LR, Hong S, Hwang I, Farooqi B, Searcey M, Chen J, Vogt PK, Boger DL. J Am Chem Soc. 2009; 131:5564–5572. [PubMed: 19334711]
79. Moon H, Lee WS, Oh M, Lee H, Lee JH, Im W, Lim HS. ACS Comb Sci. 2014
80. Lao BB, Drew K, Guarracino DA, Brewer TF, Heindel DW, Bonneau R, Arora PS. J Am Chem Soc. 2014; 136:7877–7888. [PubMed: 24972345]
81. Lao BB, Grishagin I, Mesallati H, Brewer TF, Olenyuk BZ, Arora PS. Proc Natl Acad Sci U S A. 2014; 111:7531–7536. [PubMed: 24821806]
82. Moon H, Lim HS. Curr Opin Chem Biol. 2015; 24:38–47. [PubMed: 25461722]
83. Suwal S, Kodadek T. Org. Biomol. Chem. 2013; 11:2088–2092. [PubMed: 23440085]
84. Li S, Marthandan N, Bowerman D, Garner HR, Kodadek T. Chem. Comm. 2005:581–583. [PubMed: 15672142]
85. Armand P, Kirshenbaum K, Falicov A, Dunbrack RL Jr, Dill KA, Zuckermann RN, Cohen FE. Fold Des. 1997; 2:369–375. [PubMed: 9427011]

86. Armand P, Kirshenbaum K, Goldsmith RA, Farr-Jones S, Barron AE, Truong KT, Dill K, Mierke DF, Cohen FE, Zuckermann RN, Bradley EK. *Proc. Natl. Acad. Sci. USA*. 1998; 95:4309–4314. [PubMed: 9539733]
87. Gorske BC, Stringer JR, Bastian BL, Fowler SA, Blackwell HE. *J Am Chem Soc*. 2009; 131:16555–16567. [PubMed: 19860427]
88. Gademann K, Kimmerlin T, Hoyer D, Seebach D. *J Med Chem*. 2001; 44:2460–2468. [PubMed: 11448228]
89. Seebach D, Gardiner J. *Acc Chem Res*. 2008; 41:1366–1375. [PubMed: 18578513]
90. Murray JK, Farooqi B, Sadowsky JD, Scalf M, Freund WA, Smith LM, Chen J, Gellman SH. *J Am Chem Soc*. 2005; 127:13271–13280. [PubMed: 16173757]
91. Kritzer JA, Luedtke NW, Harker EA, Schepartz A. *J Am Chem Soc*. 2005; 127:14584–14585. [PubMed: 16231906]
92. Yang JW, Stadler M, List B. *Nat Protoc*. 2007; 2:1937–1942. [PubMed: 17703205]
93. Notz W, Tanaka F, Watanabe S, Chowdari NS, Turner JM, Thayumanavan R, Barbas CF 3rd. *J Org Chem*. 2003; 68:9624–9634. [PubMed: 14656087]
94. Murray JK, Sadowsky JD, Scalf M, Smith LM, Tomita Y, Gellman SH. *Journal of combinatorial chemistry*. 2008; 10:204–215. [PubMed: 18275161]
95. Gao Y, Amar S, Pahwa S, Fields G, Kodadek T. *ACS Comb Sci*. 2015; 17:49–59. [PubMed: 25434974]
96. Gao Y, Kodadek T. *Chem. & Biol*. 2013; 20:360–369. [PubMed: 23521794]
97. Aquino C, Sarkar M, Halmers MJC, Mendes K, Kodadek T, Micalizio G. *Nature Chem*. 2011; 4:99–104. [PubMed: 22270625]
98. Sarkar M, Liu Y, Morimoto J, Peng H, Aquino C, Rader C, Chiorazzi N, Kodadek T. *Chem. & Biol*. 2014; 111:1670–1679. [PubMed: 25467125]
99. Sarkar M, Pascal BD, Steckler C, Micalizio GC, Kodadek T, Chalmers MJ. *J. Amer. Soc. Mass Spec*. 2013; 24:1026–1036.
100. Izumiya N, Nagamatsu A. *Bull. Chem. Soc. Japan*. 1952; 25:265–267.
101. Thern B, Rudolph J, Jung G. *Angew Chem Int Ed Engl*. 2002; 41:2307–2309. [PubMed: 12203574]
102. Pels K, Kodadek T. *ACS Comb. Sci*. 2015; 17:152–155. [PubMed: 25695359]
103. Brown CJ, Lain S, Verma CS, Fersht AR, Lane DP. *Nat Rev Cancer*. 2009; 9:862–873. [PubMed: 19935675]
104. Chiorazzi N, Rai KR, Ferrarini M. *N Engl J Med*. 2005; 352:804–815. [PubMed: 15728813]
105. Doran TM, Gao Y, Simanski S, McEnaney P, Kodadek T. *Bioorg. Med. Chem. Lett*. 2015
106. Doran TM, Morimoto J, Simanski S, McEnaney PJ, Kodadek T. *Mol Biosyst*. 2015; 11:3156–3163. [PubMed: 26390856]
107. Doran TM, Simanski S, Kodadek T. *ACS Chem. Biol*. 2015; 10:401–412. [PubMed: 25474415]
108. Kodadek T. *Chem. & Biol*. 2014; 21:1066–1074. [PubMed: 25237855]
- 109.
- The term "K_D" is used loosely here as the antibody population is almost certainly oligoclonal and thus represents a weighted aggregate of many true small molecule-antibody K_Ds.
110. Aditya A, Kodadek T. *ACS Comb Sci*. 2012; 14:164–169. [PubMed: 22320121]
111. Combs DJ, Lokey RS. *Tetrahedron Lett*. 2007; 48:2679–2682. [PubMed: 18398447]
112. Suwal S, Kodadek T. *Org Biomol Chem*. 2014; 12:5831–5834. [PubMed: 24976298]
113. Gao Y, Kodadek T. *ACS Comb Sci*. 2015; 17:190–195. [PubMed: 25623285]
114. Lee JH, Zhang Q, Jo S, Chai SC, Oh M, Im W, Lu H, Lim HS. *J Am Chem Soc*. 2011; 133:676–679. [PubMed: 21171592]
115. Clark MA. *Curr Opin Chem Biol*. 2010; 14:396–403. [PubMed: 20346729]
116. Scheuermann J, Neri D. *Chembiochem*. 2010; 11:931–937. [PubMed: 20391457]
117. Clark MA, Acharya RA, Arico-Muendel CC, Belyanskaya SL, Benjamin DR, Carlson NR, Centrella PA, Chiu CH, Creaser SP, Cuzzo JW, Davie CP, Ding Y, Franklin GJ, Franzen KD,

Gefter ML, Hale SP, Hansen NJ, Israel DI, Jiang J, Kavarana MJ, Kelley MS, Kollmann CS, Li F, Lind K, Mataruse S, Medeiros PF, Messer JA, Myers P, O'Keefe H, Oliff MC, Rise CE, Satz AL, Skinner SR, Svendsen JL, Tang L, van Vloten K, Wagner RW, Yao G, Zhao B, Morgan BA. *Nature Chem Biol.* 2009; 5:647–654. [PubMed: 19648931]

118. MacConnell AB, McEnaney PJ, Cavett VJ, Paegel BM. *ACS Comb Sci.* 2015; 17:518–534. [PubMed: 26290177]
119. Ding Y, Clark MA. *ACS Comb Sci.* 2015; 17:1–4. [PubMed: 25459065]
120. Ma X, Da C-S, Yi L, Jia Y-N, Guo Q-P, Che L-P, Wu F-C, Eang J-R, Li W-P. *Tetrahedron Asymm.* 2009; 20:1419–1424.
121. Shang S, Tan DS. *Curr Opin Chem Biol.* 2005; 9:248–258. [PubMed: 15939326]
122. Heinis C. *Nature Chem Biol.* 2014; 10:696–698. [PubMed: 25038789]
123. *Nature Chem Biol.* 2014; 10:693. Editorial Staff. [PubMed: 25271344]
124. Shin SB, Yoo B, Todaro LJ, Kirshenbaum K. *J Am Chem Soc.* 2007; 129:3218–3225. [PubMed: 17323948]
125. Kolb HC, Sharpless KB. *Drug Discovery Today.* 2003; 8:1128–1137. [PubMed: 14678739]

Fig. 1. Schematic representation of phage display-based screening. The blue ovals represent phages. The shapes fused to them are different displayed peptides. The protein (red) is precipitated by virtue of the tag and phages that are co-precipitated with the protein are amplified and sequenced.

Fig. 2. Post-translational modification of peptide-displaying phage to create bicyclic peptide libraries. The black line represents the peptide backbone with three conserved cysteines. The blue oval represents a phage.

Fig. 3. Schematic depiction of the split and pool synthesis of peptoid libraries using the "sub-monomer" synthesis (top box). A simple example is shown in which all nine possible dimers that can be made using three amines (colored rectangles) are synthesized in two steps.

Fig. 4.

A scheme for encoding compounds displayed on the surface of TentaGel beads with peptides synthesized on the protein inaccessible interior of the beads. Reproduced with permission from ref.³⁰

Fig. 5. Peptoid library screening. Top: general structure of peptides and peptoids. Bottom: schematic illustration of screening. A peptoid library displayed on TentaGel beads (only one bead is shown, but 50,000–1,000,000 are used routinely) is mixed with a fluorescently labeled protein or cell, along with a large excess of unlabeled competitor proteins (not shown). In this format, hits are identified under a low power fluorescence microscope (micrographs of protein- and cell-binding hits are shown). They are picked manually with a Pipetteman. The peptoid is then cleaved from the bead and sequenced by tandem mass spectrometry. For an early description of this procedure, see ref.²⁴.

Fig. 6.
Graphic depiction of the orientation of the i , $i+4$ and $i+7$ side chains (shown in pink) in a protein α -helix.

Fig. 7. Terphenyl scaffolds as α -helix mimetics by Hamilton and colleagues. A molecule designed to be a mimic for helices that bind to Bcl-x_L (box). The synthesis of this species is shown.

Fig. 8. α -helix mimetics developed by Boger and colleagues. Top left: the original design that elaborated the Hamilton terphenyls by placing an amide bond between the aromatic rings. Box: A second generation design that flanked a central phenyl ring with simple amino acids. Bottom: synthesis of the boxed molecule.

Fig. 9.
Synthesis of a BH3 helix mimic (box) by Lim and colleagues⁷⁹.

Fig. 10.

Arora's oxopiperazine-based α -helix mimetic scaffold and its synthesis. Box: inhibitors of the p53-Mdm2 and p300-HIF-1 α interactions. Right: General synthesis of the scaffold.

Fig. 11.
Synthesis and structure of helical peptoids.

Fig. 12.

Gellman's synthesis of β -amino acids (top) and the structure of a small library of peptides made from both α - and β -amino acids. The mixed α/β -peptide shown was a previously discovered Bcl-x_L hit that was diversified only at the positions highlighted in red with the units shown boxed directly below each altered region to generate a SAR library.

Fig. 13.

Illustration of a hypothetical scaffold-diverse library of peptoid-like compounds. In peptoid synthesis, only a single “acid sub-monomer” (chloro- or bromo-acetic acid) is employed, along with diverse amines. If the ten structurally diverse acid sub-monomers, such as those shown in the box, were used at each of three positions following a standard peptoid residue, then the four side chains (represented by red balls) would be presented to a protein target with $10^3 = 1,000$ different geometries. In oligomer-speak, we consider these to represent 1,000 different scaffolds. Five scaffolds that would be present in the library are shown to

illustrate (albeit in two dimensions) the spatially distinct arrangement of the side chains depending on the nature of the acid sub-monomers used in the synthesis.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 14.

Oligomers with modest scaffold diversity: PTAs and COPAs. A. Basic structure of the PTA and COPA oligomers. B. Synthesis of a COPA sub-monomer. Only one enantiomer is shown. C. Synthesis of COPA oligomers using a peptoid-like sub-monomer approach. D. Cartoon of two possible conformations of a COPA monomer showing the effect of A1,3 strain in favoring placement of the hydrogen on the chiral center in the same plane as the amide and alkenyl substituents. E. Distribution of conformers found within 1.1 kcal/mol of the lowest energy conformation identified for four of the 16 possible COPA tetramers. The

green and blue balls represent the heteroatoms and alkenyl carbons of the COPA oligomer, respectively.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 16.

A PTA that binds antibodies linked to murine Type 1 diabetes with nM affinity. A. General structure of the library screened. The common linker is shown in gray. B. Acid submonomers used at each of the positions labeled with the red ball. C. Structure of the high affinity ligand.

Fig. 17.

Aromatic sub-monomers for the construction of conformationally constrained oligomers.

Fig. 18. Solid-phase, multi-step synthesis of oxopiperazine⁸³ (A) and diketopiperazine¹¹² (B) units.

Fig. 19.

DNA-encoded OBOC libraries. A. Structure of the headpiece used to prime the synthesis. B. Schematic of the structure of a complete molecule and its encoding DNA. C. Workflow for the synthesis of the library, which is carried out in 96 well plates. A few thousand 160 μm TentaGel beads are included along with hundreds of millions of 10 μm beads for quality control purposes (see text for details).

Fig. 20.

A strategy to compare the utility of linear and macrocyclic peptoid libraries as a source of protein ligands. A. Structure of the libraries. Where R=variable side chains, n=0–2 units B. The protocol by which the K_D s of all of the hits from an OBOC library screen can be determined without the need for re-synthesis⁴⁹.