JOURNAL OF

MEDICINAL CHEMISTRY

© Copyright 1994 by the American Chemical Society

Volume 37, Number 9 April 29, 1994

Perspective

Applications of Combinatorial Technologies to Drug Discovery. 1. Background and Peptide Combinatorial Libraries

Mark A. Gallop, Ronald W. Barrett, William J. Dower, Stephen P. A. Fodor, and Eric M. Gordon*

Affymax Research Institute, 4001 Miranda Avenue, Palo Alto, California 94304

Received December 21, 1993

Biographies

Ronald W. Barrett received his Ph.D. in Pharmacology from Rutgers University in 1983 and did postdoctoral work at the Addiction Research Foundation in Palo Alto in the field of opiate receptors. From 1986 to 1989, he worked in the Neuroscience Research Area at Abbott Laboratories in Abbott Park, IL. He joined Affymax in 1989 where he established the receptor pharmacology group. Dr. Barrett is presently Vice President and Director of Receptor Pharmacology.

William J. Dower received his Ph.D. in Biology from the University of California, San Diego, and did postdoctoral work at Stanford University where he studied steroid control of specific mRNA metabolism and the amplification of genes mediated by chemotherapeutic agents. He joined BioRad Laboratories in 1984 where as a member of the newly formed molecular biology unit, he introduced the electroporation technique for the high-efficiency transformation of bacteria. He joined Affymax in 1989 to establish a molecular biology research group. Dr. Dower is Director of Molecular Biology at Affymax.

Stephen P. A. Fodor received his Ph.D. in Chemistry at Princeton University and was an NIH postdoctoral fellow in Chemistry at the University of California, Berkeley. He joined Affymax in 1989 as a Staff Scientist in Optical Technologies and became Director of Physical Sciences. Dr. Fodor's group led the development of new technologies, merging photolithography with combinatorial solid-phase chemistry. In 1993 he joined Affymetrix as Scientific Director where he is currently using oligonucleotide arrays to study a variety of DNA molecular recognition processes.

Mark A. Gallop received his Ph.D. from the University of Cambridge, England, for research in the area of organo-transition metal chemistry and did postdoctoral work in the laboratories of Peter G. Schultz at the University of California, Berkeley. He joined Affymax in 1990 and is currently a Senior Scientist and leader of the Combinatorial Chemistry Group.

Eric M. Gordon received his Ph.D. in 1973 from the University of Wisconsin—Madison and did postdoctoral work at Yale University. His research interests include the rational design of enzyme inhibitors, the chemistry of amino acids, peptides, and natural products, and combinatorial chemistry. Dr. Gordon has authored 150 papers and U.S. patents in these areas. From 1974 to 1992 he worked at the Squibb Institute for Medical Research (subsequently the Bristol Myers Squibb Pharmaceutical Research Institute) in Princeton, NJ, most recently as a Director of Medicinal Chemistry. In 1992, he joined the Affymax Research Institute where he is currently Vice President of Research and Director of Chemistry.

A. Introduction

Recent trends in the search for novel pharmacological agents have focused on the preparation of "chemical libraries" as potential sources of new leads for drug discovery. Chemical libraries are intentionally created collections of differing molecules which can be prepared either synthetically or biosynthetically and screened for biological activity in a variety of different formats (e.g., libraries of soluble molecules; libraries of compounds tethered to resin beads, silica chips, or other solid supports; recombinant peptide libraries on bacteriophage and other biological display vectors; etc). Combinatorial chemistry is a type of synthetic strategy which leads to large chemical libraries. For the purposes of this discussion, combinatorial chemistry may be defined as the systematic and repetitive, covalent connection of a set of different "building blocks" of varying structures to each other to yield a large array of diverse molecular entities.

Traditionally, new medicinal chemical lead structures have originated from the isolation of natural products from microbiological fermentations, plant extracts, and animal sources; from screening of pharmaceutical company compound databases; and more recently through the application of both mechanism-based and structure-based approaches to rational drug design.[1,2] The advent of high throughput, automated techniques has made possible the robotized screening of in excess of hundreds of thousands

0022-2623/94/1837-1233$04.50/0 © 1994 American Chemical Society

of individual compounds per year, per drug target. The availability of this capability, combined with a major, worldwide shift in emphasis by the drug industry toward more cost-effective pharmaceutical products, has exacerbated the need for a continuous flow of huge numbers of novel molecules. Several studies have shown that the average cost of creating a new molecular enitity in a major pharmaceutical company is around \$7500/compound.[3] A significant portion of this figure relates to an intrinsic feature of contemporary medicinal chemistry which has, by and large, involved the "hand-crafted" serial synthesis and testing of individual molecular entities. A major theme of the industrial revolution was replacement of highly expensive, individualized manufacturing with the concept of mass production to greatly decrease manpower and cost. Though this period began in America more than a century ago, to date the idea has not been significantly applied to the preparation of pharmaceutical leads.

To address inefficiencies inherent in the contemporary new lead discovery process, researchers have recently turned to the concept of using combinatorial chemical technologies and devising novel assay formats for rapidly evaluating these compound collections.[4-7] Thus by employing a building-block collection (*vide infra*), and systematically assembling these blocks in many combinations using chemical, biological, or biosynthetic procedures, it is possible to create "chemical libraries" as vast populations of molecules. This approach has clear parallels in Nature, which eons ago applied the building block approach for creation of oligonucleotides, carbohydrates, and peptides/proteins, through the combination of nucleosides, sugars, and amino acids respectively.

An essential starting point for the generation of molecular diversity is an assortment of small, reactive molecules which may be considered chemical building blocks. The universe of structural diversity accessible through assembly of even a small set of building-block elements is potentially large, and unleashing the power inherent in the building-block approach is crucial to the success of the combinatorial method. The building-block argument is easily illustrated, and its implications are profound. Theoretically, the number of possible different individual compounds, N, prepared by an ideal combinatorial synthesis is determined by two factors; the number of blocks available for each step "b", and the number of synthetic steps in the reaction scheme, x. If an equal number of building blocks are used in each reaction step, then $N = b^x$. If the number of building blocks for each step varies (e.g., b, c, d in a three-step synthesis), then $N = bcd$. Exploitation of a basis set of (for example) 100 interchangeable building blocks permits the theoretical synthesis of 100 million tetrameric or 10 billion pentameric chemical entities (see Figure 1).

In this perspective we will review the technologies and strategies that have recently emerged for using combinatorial chemical methods in ligand discovery. Historically, these methods made their first appearance in the development of peptide libraries. The ready availability of a large and structurally diverse range of amino acid building blocks, a highly refined, generic coupling chemistry, and the fact that small peptides are biologically and pharmaceutically key molecules, focused early efforts on peptide chemistry as a useful vehicle for exploring the power and conceptual issues attached to combinatorial ligand discovery. In addition, biological systems have been

	Units	Library Entities
Basis Set of 20 (e.g. natural amino acids)	20^3	8,000
	20^4	160,000
	20^5	3.2 Million
Basis Set of 100	100^3	1 Million
	100^4	100 Million
	100^5	10 Billion
Basis Set of 1000	1000^3	1 Billion
	1000^4	1 Trillion
	1000^5	1 Quadrillion

Figure 1. Creating chemical diversity from a basis set of building blocks.

used to generate vast collections of peptides through the expression of randomly synthesized DNA. The ability to selectively enrich individual peptide sequences from these recombinant libraries has demonstrated the power and versatility of peptide diversity for *de novo* ligand identification. Generalization of the combinatorial strategy has led to construction of collections of other natural polymers (e.g., oligonucleotides) and unnatural (synthetic) polymeric libraries. Currently there is an acute interest in the generation of small-molecule libraries, through "combinatorial organic synthesis (COS)": the rapid synthesis of enormous numbers of diverse, low molecular weight, nonpolymeric organic molecules. In surveying progress in the field to date, we will also attempt to define criteria and constraints which must be imposed on the creation of libraries useful for drug discovery.

An essential element of the combinatorial discovery process is that *one must be able to extract the information made available by library screening*. Put another way, creating large quantities of molecular diversity for ligand discovery is insufficient, unless there is a format at hand to capture the information, which in this case is the chemical structures of active compounds. In the course of this overview, particular attention will be directed to the assay methodologies employed in screening combinatorial libraries and to the interrelationship between these methods, the diversity display formats, and the mechanisms for determining the structure of selected molecules that emerge from library evaluation. Two divergent embodiments of the combinatorial discovery exercise will be highlighted: (i) *random screening*, where the task is to identify a lead compound in the absence of any structural information about active molecules, and (ii) *directed screening* or *chemical analoging*, where the objective is to evaluate closely related structural analogs of a lead molecule, establish SAR, and optimize biological potency. These activities raise different issues in terms of both the synthetic combinatorial strategies and assay parameters that must be applied.

The above cursory analysis, at face value, makes clear that one or a few combinatorial experiments is sufficient to create more different molecules than exist in the combined compound files of the worldwide pharmaceutical industry! Recent research suggests that medicinal chem-

istry and drug discovery may be entering a new era in which vast numbers of small molecules may be readily accessible by combinatorial approaches. The power to synthetically create and evaluate huge numbers of known and future pharmacophores is unprecedented and suggests that combinatorial technologies may rapidly intersect and ultimately shortcut the traditional path of medicinal chemistry. The generation and screening of such immense numbers of compounds will demand innovative tools for data handling and analysis. Questions such as what should be the composition of a building-block collection suitable for drug discovery, what kinds of new synthetic and screening strategies are suitable, what formats for diversity creation are relevant, how to integrate both diversity generation and screening, and ultimately how to develop a new paradigm to dramatically shorten the drug-discovery process will be considered and discussed.

B. Biological Approaches to Generating Molecular Diversity

The harnessing of biological systems for the generation of peptide diversity mimics the evolutionary creation of protein diversity; however, artificial evolution is greatly accelerated by the introduction of diversity into the system at a much higher rate than occurs naturally. The source of diversity is the combinatorial chemical synthesis of mixtures of oligonucleotides. Oligonucleotide synthesis is a well-characterized chemistry that allows tight control of the composition of the mixtures created. The degenerate sequences produced are then placed into an appropriate genetic context for expression as peptides.

The particular context of expression of the randomly-encoded peptides is a key feature in the practical application to screening. Because a very large number of different peptides can be produced by biological expression, a highly efficient means of selecting the most interesting of these peptides is required. Depending on the type of activity that is sought, a variety of selection schemes are possible. If, for example, a particular catalytic activity is desired, a genetic screen might be devised that would endow organisms that express that activity with a growth advantage. The application emphasized in this review is the identification of new ligands, an approach based on the affinity purification of active peptides by adsorption on a target receptor. With each of the biological systems that will be described, identification of individual active peptides is enabled by physical linkage of the peptides to their encoding DNA. Capturing the peptides also recovers the DNA, and the replicative nature of the biological system is employed to amplify and sequence the DNA, thereby identifying the active peptides.

In practice, the process begins with the synthesis of single-stranded DNA. Synthetic oligonucleotides have been used to create genetic mutants since the 1970s. Initially, precise changes were made in regulatory and protein-encoding regions in order to study the phenotypes of the resulting mutants.[8,9] Later, pools of mutants were produced by introducing a variety of base changes into sites in the genome. It was soon realized that large collections, or libraries, of mutant clones could be acquired by inserting fully random cassettes of oligonucleotides into the targeted loci.[10-13] Because of the combinatorial nature of random cassette mutagenesis, the number of mutant strains in these libraries becomes much too large to analyze individually, and population selection schemes of the sort mentioned above are required.

Conventional expression cloning techniques rely on the detection of the binding activity of an expressed protein that is associated geographically or physically with its encoding DNA. The identification of individual clones that bind to a particular probe can be detected among thousands to millions of inert clones by the use of phage- or colony-lift techniques. Here, the clones are grown at rather high density on agar surfaces and "lifted" onto polymeric membranes to adsorb the macromolecules from the clones in a pattern that mirrors the arrangement of the colonies on the growth plates. Detection is achieved with a binding protein, such as an antibody, that is appropriately labeled. Once binding is detected on the membrane, the corresponding colony or plaque on the growth plate can be located and isolated for further analysis.[14,15] A method that has been employed in the cloning of receptors invokes the surface expression of clone libraries on eukaryotic cells, followed by selection of the desired clones through affinity adsorption of cells displaying the receptor on a matrix of immobilized ligand. This selection method has been dubbed "panning".[16,17] Upon isolation of the cells, the cloned fragments of DNA they contain are rescued and sequenced. These methods are commonly used, powerful means of isolating clones from cDNA and genomic libraries of normal size (up to $\sim 10^7$ clones), but they become impractical with the very large random libraries of 10^8–10^9 clones typically employed in ligand discovery projects. Recently, the advent of extraordinarily efficient cloning tools has permitted the construction of clone libraries of up to 10^{11} members.[18]

Because the number of different peptides one can create by this combinatorial approach is astronomical, and because the expectation is that peptides with the appropriate structural characteristics to serve as ligands for a given receptor will be rare, the need for methods capable of conveniently screening billions of clones is apparent. Several strategies for selecting very rare peptide ligands have been described. These are categorized as peptide display techniques and are distinguished by the peptide presentation format and by the nature of the linkage between the peptide-encoding DNA and the peptide itself. This association effectively tags the peptide for identification during screening, since in practice the peptide sequence is revealed by sequencing of the DNA.

Several approaches for displaying peptides or proteins on the surface of microorganisms have been developed. Each of these methods employs the fusion of the "guest" peptides to a cell surface protein of the host. A number of reports describe the use of the maltose receptor of *E. coli* (the outer membrane protein, LamB) as a peptide fusion partner.[19] Oligonucleotides have been inserted into plasmids encoding the LamB gene to produce peptides fused into one of the extracellular loops of the protein. These peptides are available for binding to antibodies and can elicit an immune response when the cells are administered to animals. Other cell surface proteins, e.g., OmpA,[20] PhoE,[21] and PAL,[22] as well as large bacterial surface structures have served as vehicles for peptide display.[†] Peptides have been fused to pilin, a protein which polymerizes to form the pilus—a conduit for interbacterial exchange of genetic information.[23] Because of its role in

[†] OmpA, outer membrane protein A; PhoE, phosphate-limitation inducible outer membrane protein; PAL, peptidoglycan-associated lipoprotein.

interacting with other cells, the pilus provides a useful support for the presentation of peptides to the extracellular environment. Another large surface structure pressed into service for peptide display is the bacterial motive organ, the flagellum. Fusion of peptides to the subunit protein flagellin offers a dense array of many peptide copies on the host cells.[24] Surface proteins of other bacterial species have also served as peptide fusion partners. Examples include the *Staphylococcus* protein A and the outer membrane protease IgA of *Neisseria*.[25,26]

Peptides and Proteins Displayed on Phage Particles. Systems for presenting peptides on bacterial cells have potential as random library vectors for affinity selection of ligands, but to date, only the LamB system has been used in this way.[27] An alternative system, which presents peptides on the much less complex surface of bacteriophage particles, has found wide acceptance. These vectors are derived from a family of filamentous phages of *Escherichia coli* that include phage m13, f1, and fd. In 1985, Smith published a seminal paper describing the display of peptides on filamentous phage particles.[28] The minor coat protein pIII was chosen as the fusion partner, and a fragment of the restriction enzyme EcoRI was placed in the middle of pIII. The guest peptide could be detected on the surface of mature phage with anti-EcoRI antibodies. The normal function of pIII is to mediate adsorption to the host cell (via the sex pilus) as a prelude to the entry of the phage into the bacterial cell. pIII is the largest protein of the virion (~40 kD) and is well suited by its natural function to the display of peptides for binding to exogenous proteins. The C-terminus of the pIII protein is embedded in the phage coat while the N-terminus projects into the environment. In 1988, Parmley and Smith extended this work,[29] relocating the fused peptides to the N-terminal domain of pIII, an insertion site less disruptive to the normal function of the protein. This report presented a detailed method for selecting phage bearing peptide ligands and demonstrated the isolation of a few phage carrying fragments of β-galactosidase from an enormous excess of inert phage by affinity purification on an anti-β-galactosidase antibody. Although the authors proposed the method primarily as a tool for cDNA expression library cloning, they also suggested that a library of short, randomly created peptides could serve as an "epitope library" for mapping the binding specificities of antibodies.

To realize the potential of such random libraries required the construction of libraries of extraordinary size. In 1990 three groups, each utilizing a variant of filamentous phage as a pIII–peptide expression vector, reported the creation of libraries of $\sim 10^7$ to $>10^8$ recombinants, and the successful selection from these libraries of specific ligands for monoclonal antibodies,[30,31] and for the biotin-binding protein, streptavidin.[32] The much more abundant major coat protein, pVIII, of the filamentous phage can also serve as a peptide expression vehicle, as demonstrated initially by Il'chev.[33] Greenwood *et al.*[34] cloned oligonucleotides into the single pVIII gene of the fd genome, leading to fusion of the expressed peptide to all of the several thousand copies of pVIII on each phage particle. This format was effective in expressing peptides up to six residues in length. However, peptides longer than six residues impeded the assembly of the virions, greatly reducing the yield of phage produced. This limitation was addressed by utilizing a phagemid that contained a pVIII gene into which the oligonucleotide was inserted. Infection of cells carrying these phagemids with a "helper" phage provided all of the phage functions including the wild-type pVIII. Adjustment of the level of expression of the peptide-fused pVIII permitted expression of the recombinant peptide on a minority of the coat molecules and allowed efficient assembly of phage carrying peptides of a range of sizes. The phagemid gene specifying the pVIII–peptide fusion protein was placed under control of the inducible tac promotor, and the authors estimated the fusion protein to be incorporated into mature virions as 10% of the total pVIII for one 12-mer peptide and as 30% for another unrelated 12-mer.[34] Expression levels of decapeptide–pVIII fusions driven by an arabinose promotor are 8–25% of total pVIII, as measured by peptide sequencing analysis.[35] Felici *et al.*[36] employed a pVIII phagemid vector to prepare a library of $\sim 10^8$ clones displaying random peptides nine residues in length. This library was screened on an antibody raised against a fragment of IL-1β, and the recovered peptides contained a three-residue subset of the immunogenic peptide.

Phage display methods have found wide use in the engineering of proteins and protein domains. Although the focus in this review is the use of display techniques for small peptide discovery, a brief discussion of their role in protein engineering is relevant. The availability of recombinant display technology is revolutionizing the field of antibody engineering. Since the first reports of antibody domains displayed on the major[37] or minor[38–40] coat proteins of phage, the field has progressed very rapidly. cDNA libraries of single-chain Fvs and assembled Fabs representing the repertoire of heavy and light chains of an animal are cloned into a phage vector to create many possible combinations of the associated chains, potentially creating new antigen specificities. The libraries may be derived from animals immunized with the antigen of interest or may even be made from a "naive" source not previously exposed to the target antigen.[40,41] The antibody library is then selected against an immobilized antigen, often yielding a large number of reactive clones. The antibodies recovered in the early rounds of selection are usually of modest affinities, with K_d values in the range of high nanomolar to mid micromolar. The affinities can be improved using several strategies, collectively referred to as "*in vitro* affinity maturation"—a process that mimics the selection of high-affinity antibodies that occurs *in vivo* during the course of an immune response. *In vitro* affinity improvement is accomplished by simply continuing selection of the pool of antibodies on the immobilized antigen, or by introducing sequence variation into the enriched antibody pool and reapplying selection. The variation is produced by shuffling heavy and light chains to create new chain pairings and by mutagenizing the complementarity determining regions (CDRs) to create new variations of the combining site with improved complementarity to the antigen. Among the benefits of this burgeoning area are the rapid development of antibodies with customized specificities and high affinities, and most importantly, the development of completely human antibodies, some with specificity for even human antigens, and with the potential for application as therapeutic agents. Excellent reviews of this fast-moving field are available.[42,43]

The engineering of other classes of proteins has also benefited from the application of recombinant display technology. Human growth hormone (hGH) was the first

example of a foreign protein displayed on filamentous phage in a functionally active form with respect to binding its receptor.[44] Lowman *et al.*[45] extended this work, accomplishing the rapid artificial evolution of the wild-type hGH to a much more potent form via phagemid display selection. Markland and co-workers[46] expressed functional bovine pancreatic trypsin inhibitor (BPTI) on the major coat protein of phage particles. This was achieved by introducing a second pVIII gene, controlled by an inducible promotor, into the genome of phage m13. The level of expression of the BPTI-pVIII fusion was estimated to be 30–60 molecules per phage. This group also reported constructing a small library of BPTI mutants and screening it against human neutrophil elastase (HNE) to find the most potent inhibitor of HNE yet described.[47] Pannekoek *et al.*[48] expressed another protease inhibitor, human plasminogen-activator inhibitor 1 (PAI-1) on pIII, and showed retention of the ability to bind t-PA. A library of lightly mutagenized variants of PAI-1 was constructed for the purpose of probing structure–activity relationships of the inhibitor molecules. The enzymes, alkaline phosphatase and trypsin, have been expressed in catalytically active forms on phage surfaces,[49,50] opening the way for applied evolution of altered catalytic properties. The means to produce huge collections of mutants and to rapidly select for those with enhanced binding and other features is destined to have an enormous impact on the field of protein engineering.

As profound as developments in the area of protein engineering may be, the greatest significance of recombinant display methods for medicinal chemistry is the promise of small peptides as lead compounds for drug development. Phage-borne peptide libraries have been screened against a variety of receptors. Since the initial reports identifying families of peptide ligands for antibodies appeared, a number of antibodies have been mapped for their epitope specificities.[51] Those with continuous sequence peptide epitopes are usually quite easily mapped.[30,31,36,52–54] There are also examples of the identification of peptide ligands for antibodies that recognize conformation-dependent peptide epitopes[55–57] and non-peptide epitopes.[58] Peptide ligands for other proteins that normally bind nonpeptidic compounds have also been found with phage libraries. Examples are peptides that compete with the binding of biotin to streptavidin[32,59] and peptides that displace α-methylmannoside from the lectin, concanavalin A.[60,61] Smith *et al.*[62] recovered from a phage library peptides that are competitive inhibitors of the interaction of S-peptide with S-protein (a fragment of RNase). These peptides have little or no sequence similarity to the natural ligand S-peptide. A family of peptides binding to gpIIb/IIIa was obtained by screening the integrin with a library of random hexapeptides flanked by cysteine residues, capable of forming a disulfide-bridged loop containing the variable hexamer region. The peptide ligands selected contained the tripeptide sequence, RGD, and were much more active in binding the integrin when prepared in the cyclic, oxidized form.[63] Another integrin, $\alpha_5\beta_1$, was screened with a phage hexapeptide library, and most of the sequences recovered contained the RGD motif.[64] One of these, having the potential of forming a disulfide-constrained cyclic structure, was the most potent inhibitor of $\alpha_5\beta_1$ binding to fibronectin. By screening a cell surface antigen receptor derived from the B-cells of a lymphoma patient, Renschler *et al.*[65] identified several families of peptides that compete with an anti-idiotype IgG prepared against the receptor. One of the peptides, when dimerized or tetramerized, activated the receptor leading to the specific effects of tyrosine phosphorylation, apoptosis, and death in culture of cells derived from the tumor. Hammer *et al.*[66] have employed phage peptide libraries to define the consensus binding motifs of three alleles of the MHC class II peptide receptor. Blond-Elguindi *et al.*[67] analyzed the peptide-binding preferences of another promiscuous receptor, the chaperonin BiP. The motif identified allowed the formulation of rules that predicted additional peptide ligand sequences in proteins that interact with BiP.

Phage–peptide libraries have been actively used against receptors for several years now. Because peptides discovered as ligands to pharmaceutically important receptors serve only as very early leads however, very little information has been made publically available concerning the successful identification of novel, low molecular weight peptides for important receptor targets.

Besides their use as sources of ligands, peptide libraries can be adapted, with a little ingenuity, to the selection of peptides with other properties. Matthews and Wells[68] recently described a method for identifying protease substrates using phage libraries. A library of peptides fused to pIII was constructed with the additional feature of a "tether" added at the N-terminus of the peptide-pIII fusion. The tether was chosen to be a high-affinity ligand for some binding protein (these workers used hGH as their tether). The library phage, all of which possessed the tether, were bound to the hGH receptor previously coated onto microtitre wells. The immobilized phage were then treated with the protease subtilisin, and phage-bearing peptides cleaved by the protease were released from their tether and freed from the support. The released phage were recovered for amplification and DNA sequencing. A similar method, utilizing a small peptide epitope as the tether and allowing the protease digestion to occur in solution, has been independently developed and utilized to identify new substrates for a matrix metalloprotease.[69] Substrates for peptide-modifying enzymes other than proteases can be identified with peptide libraries if an appropriate antibody or other binding protein specific for either the substrate or the product of the reaction is available (*vide infra*).

Peptides-on-Plasmids. In each of the filamentous phage systems and the LamB system described above, the physical link between the peptide and its encoding DNA occurs by the containment of the DNA within a particle (cell or phage) that carries the peptide on its surface. Capturing the peptide captures the particle and the DNA within. Cull *et al.*[70] devised an alternative scheme that utilizes the DNA-binding protein LacI to form a link between peptide and DNA. They constructed a plasmid containing the LacI gene with an oligonucleotide cloning site at its 3′-end. Under the controlled induction by arabinose, a LacI–peptide fusion protein is produced. This fusion retains the natural ability of LacI to bind to a short DNA sequence known as the Lac operator (LacO). By installing two copies of LacO on the expression plasmid, the LacI–peptide fusion binds tightly to the plasmid that encoded it. Because the plasmids in each cell contain only a single oligonucleotide sequence and each cell expresses only a single peptide sequence, the peptides become specifically and stably associated with the DNA sequence

that directed its synthesis. The cells of the library are gently lysed and the peptide–DNA complexes are exposed to a matrix of immobilized receptor to recover the complexes containing active peptides. The associated plasmid DNA is then reintroduced into cells for amplification and DNA sequencing to determine the identity of the peptide ligands. As a demonstration of the practical utility of the method, a large random library of dodecapeptides was made and selected on a monoclonal antibody raised against the opioid peptide dynorphin B. A cohort of peptides was recovered, all related by a consensus sequence corresponding to a six-residue portion of dynorphin B.[70]

This scheme, named peptides-on-plasmids by the designers, differs in two important ways from the phage display methods. First, the peptides are attached to the C-terminus of the fusion protein, resulting in the display of the library members as peptides having free carboxy termini. Both of the filamentous phage coat proteins, pIII and pVIII, are anchored to the phage through their C-termini, and the guest peptides are placed into the outward-extending N-terminal domains. In some designs, the phage-displayed peptides are presented right at the amino terminus of the fusion protein.[30] A second difference is the set of biological biases affecting the population of peptides actually present in the libraries. The LacI fusion molecules are confined to the cytoplasm of the host cells. The phage coat fusions are exposed briefly to the cytoplasm during translation but are rapidly secreted through the inner membrane into the periplasmic compartment, remaining anchored in the membrane by their C-terminal hydrophobic domains, with the N-termini, containing the peptides, protruding into the periplasm while awaiting assembly into phage particles. The peptides in the LacI and phage libraries may differ significantly as a result of their exposure to different proteolytic activities. The phage coat proteins require transport across the inner membrane and signal peptidase processing as a prelude to incorporation into phage. Certain peptides exert a deleterious effect on these processes and are underrepresented in the libraries.[71] These particular biases are not a factor in the LacI display system.

Schatz[72] has used the peptides-on-plasmid system to discover a new enzyme substrate: a completely novel and much smaller recognition sequence for the protein-biotinylating enzyme, BirA. Functional homologs of BirA are found in species ranging from bacteria to tomatoes. In *E. coli*, the enzyme adds biotin to the epsilon amino group of a specific lysine in the 75 residue enzyme recognition domain of a single protein made by the bacterium. The recognition domain is highly conserved among prokaryotic and eukaryotic species that possess the BirA-like activity, and attempts to reduce the size of the domain by deletion analysis have not been fruitful. Schatz designed a library with a fixed lysine flanked by 10 random residues on each side and fused to the C-terminus of LacI. The LacI fusions are expressed in the bacterial cytoplasm where the peptides are exposed to the BirA enzyme. The LacI–plasmid complexes were isolated in the usual way and the library screened for binding to immobilized streptavidin to retain peptides that had become biotinylated. An important counter selection was employed to bias against the recovery of non-biotinylated peptides that contained streptavidin-binding motifs of the types reported by Devlin[32] and Kay.[59] The counter selection was based on the assumption that the nonbiotinylated motifs would bind streptavidin with much lower affinity than the peptides containing a biotinylated lysine. After binding the plasmid library to immobilized streptavidin, biotin was added to displace the low-affinity streptavidin-binding peptides. Sequences of the clones that survived this process revealed a 13-residue consensus around the recipient lysine that served as a substrate for the BirA enzyme.

Peptide-modifying enzymes can be used to create additional diversity in the libraries. To illustrate, libraries of C-terminally amidated peptides have been constructed. The dual activity enzyme, peptidyl amino monooxygenase (PAM), cleaves the C-terminal glycine of peptides to leave a carboxamide on the penultimate residue. Because the enzyme is quite permissive in its specificity for the residues adjacent to the glycine, most peptides should serve as substrates for this transformation. A library was designed to contain 12 random residues followed by a single fixed glycine, all fused to the C-terminus of LacI. The LacI–peptide–plasmid complexes were isolated and treated with PAM *in vitro*, providing display of a collection of random peptides terminating with an amide rather than a free carboxylic acid residue.[73] This type of library may be a fruitful source of new ligands for the family of G-protein receptors that have C-terminally amidated peptides as their natural ligands.

Peptides Displayed on Polysomes: An *in Vitro* Approach to Random Library Construction. The number of small peptides available in recombinant random libraries is enormous. Libraries of 10^7–10^9 independent clones are routinely prepared. Libraries as large as 10^{11} recombinants have been created,[74] but this size approaches the practical limit for clone libraries. This limitation in library size occurs at the step of transforming the DNA containing randomized segments into the host bacterial cells. To circumvent this limitation, an *in vitro* system based on the display of nascent peptides in polysome complexes has recently been developed. This new addition to the repertoire of L-amino acid display libraries has the potential of producing libraries 3–6 orders of magnitude larger than the currently available phage/phagemid or plasmid libraries. The salient feature of this approach is that the construction of the libraries, expression of the peptides, and screening, is done in an entirely cell-free format.

As a demonstration of this method, Mattheakis *et al.*[75] have constructed a molecular DNA library encoding 10^{12} decapeptides and have expressed the library in an *E. coli* S30 *in vitro* coupled transcription/translation system. Conditions were chosen to stall the ribosomes on the mRNA, causing the accumulation of a substantial proportion of the RNA in polysomes and yielding complexes containing nascent peptides still linked to their encoding RNA. The polysomes are sufficiently robust to be affinity purified on immobilized receptors in much the same way as the more conventional recombinant peptide display libraries are screened. RNA from the bound complexes is recovered, converted to cDNA, and amplified by PCR to produce a template for the next round of synthesis and screening. The polysome display method can be coupled to the phage display system, as shown in this report. Following several rounds of screening, cDNA from the enriched pool of polysomes was cloned into a phagemid vector. This vector serves as both a peptide expression vector, displaying peptides fused to the coat proteins, and

as a DNA sequencing vector for peptide identification. By expressing the polysome-derived peptides on phage, one can either continue the affinity selection procedure in this format or assay the peptides on individual clones for binding activity in a phage ELISA (*vide infra*), or for binding specificity in a competition phage ELISA.[76] To identify the sequences of the active peptides one simply sequences the DNA produced by the phagemid host.

The authors screened the polysome library for ligands to the anti-dynorphin B antibody, 32.39. This antibody recognizes as its epitope a six-residue portion of dynorphin B, RQFKVV, and has been shown to bind a large number of peptides closely related to this sequence.[70] Three to five rounds of screening yielded a family of peptides clearly related to the known ligands and having affinities, as free peptides, of 7–140 nM. The selection of peptides in this affinity range indicates the polysomes may form a monovalent association with the immobilized receptors (see discussion below on the effect of valency in library selection schemes). The molecular library described, 10^{12} DNA molecules, was expressed in an *in vitro* transcription/translation reaction of 50 μL. Reactions 100-fold larger are easily attained, and modifications to the S30 system could also provide additional synthetic capacity. It may therefore be possible to create and screen polysome libraries of 10^{14}–10^{15} members. The peptide population of the polysome libraries should be less influenced by the biological biases discussed above, and additional building block diversity might potentially be employed by incorporating tRNAs charged with unnatural amino acids into the reaction, as pioneered by Schultz *et al.*[77]

Designing Random Libraries for Biological Display Systems. Frequently, a drug-discovery program may be initiated with no information about the structural requirements of a small ligand for a target receptor. This is especially true for programs in which the goal is to disrupt protein/protein interactions. The large families of cytokine and growth factor receptors are examples of targets of this type. Historically, there have been few examples of low molecular weight compounds that block such interactions. For receptors where small molecule leads are available, it may still be desirable to identify new leads that bind to the target in a manner that is either similar or dissimilar to that of a known ligand. A likely strategy for finding such ligands may be searching the huge collections of compounds available in randomly created peptide display libraries. In this section the salient points in the design of these libraries are discussed.

In the construction of synthetic libraries, the choice of appropriate building blocks from among the huge number of available monomers can potentially be problematic. This is usually not an issue in the construction of random biological peptide libraries, where one simply uses all of the 20 genetically encoded amino acids. More pressing design issues relate to the length of the peptides to be displayed and the fixing of amino acids in certain positions—often for the purposes of introducing structural constraint into the peptides. Choosing a peptide length so short that the library size greatly exceeds the number of peptides theoretically accessible is obviously a defective strategy. If, for example, a library of L-tetrapeptides were to be made (1.6×10^5 possible tetrapeptides, $\sim 10^6$ possible tetracodons), there would be little point in constructing the library to contain 10^8 members. Often the goal in library design is to create the greatest number of *different* compounds. This requires making a library of the largest practical size and choosing a peptide length whose combinatorial possibilities greatly exceed the library size. This results in a low likelihood that the same peptide will be made by more than one clone of the library. To illustrate, there are 2.6×10^{10} L-octapeptides; even the more abundant peptides comprised entirely of 3-codon amino acids are unlikely to appear more than once in a library of 10^8 members; thus, this library may express nearly 10^8 different peptides. Of course, peptides containing many of the 12 amino acids encoded by a single codon will be poorly represented in this library. To include most of the octapeptides, a library exceeding 10^{12} members is required.

The nature of the target receptor should also be considered when designing the library. Although some binding sites for small ligands are known to bind peptides of two or three residues, other important targets, especially those involved in protein/protein interactions, may prefer larger molecules as ligands. Libraries containing peptides of 8, 10, 12 or more random residues may be the most productive for early screening of receptors of this type.

Biological display libraries may be designed to contain randomized residues placed within the context of fixed "scaffolds" of constraining residues. Obvious examples include libraries of disulfide-bridged loops, made by flanking strings of random residues with pairs of fixed cysteines. Restricting the global conformations of peptides may provide ligands of higher affinity that are more amenable to medicinal chemical development, but this may come at the cost of *effective* library size. Each member of a peptide library exists as a family of conformational states, and as one deliberately constrains the peptides, the universe of molecular shapes to be sampled by the library declines. To exploit the conformational diversity in unconstrained libraries requires screening techniques that, in the early rounds, can recover rather low affinity ligands. Some of the multivalent, high-avidity systems described above are capable of selecting ligands having K_d values as weak as 100 μM. Early, low-affinity leads can then be improved by using analog libraries and affinity-sensitive screening strategies as described in later sections of this review.

Designing Libraries of Variants Based on a Lead Peptide. Often, one has some knowledge about the primary structures of small peptide ligands of the target receptor. This information may come from the structure of the receptor, from the structure of the natural ligand, or from early leads identified in random library screening. This information can be used to construct libraries of analogs ("mutants") based on the starting sequence ("wild type"). Screening the analog libraries aids in determining the minimum size of the active sequence and in identifying those residues critical for binding and intolerant of substitution. The mutant libraries may also contain peptides of higher affinities than the starting sequence, and these are recovered with specialized screening techniques (*vide infra*).

Several methods are available to construct analog libraries. Beginning with a single peptide sequence, or with a consensus sequence representing a group of related peptides, the encoding oligonucleotide is resynthesized, and a controlled level of incorrect bases is introduced at each coupling step.[78–80] Both low and high levels of mutagenesis are feasible. A typical protocol calls for a

mixture of 70% of the correct base and 10% of each of the other three bases (70/10/10/10 strategy) and produces amino acid changes in about 50% of the positions. The resulting mixture of oligonucleotides is cloned into the display vector to create a library of 10^8–10^9 variants for rescreening. The method is very convenient but somewhat biased in the distribution of changes produced from the starting sequence. For example, a Val to Ala change requires that only one base of the codon change (GTN to GCN), an event with 0.1 probability. By contrast, a Asp to Trp change requires a different base in all three codon positions (GAT/C to TGG), an event with only 0.001 likelihood. Nevertheless, because very large libraries can be made and very rare events detected, this approach is highly productive in defining the critical residues involved in binding.

A less biased, but technically more difficult alternative is the codon-by-codon mutagenesis strategy. The pool of synthesis resin is apportioned between two (or more) reactions, one dedicated to making the wild-type codons, and the other(s) to making random codons (or other specific amino acid codons). The ratio of wild-type to mutant codon is determined by the proportion of the resin placed in each reaction. At each codon boundary, the resin is pooled, mixed, and reapportioned to the reaction chambers.

Both of these methods cause mutations to be distributed throughout the starting peptide at a controllable frequency. A different approach fixes certain residues and randomizes the others. One might make a mutant library based on an octapeptide lead by fixing the first four residues as those in the starting sequence and randomizing positions 5–8. A complementary library fixing positions 5–8 and randomizing the first four residues would also be made. From screening these libraries, much can be learned about the preferences of the receptor for each position in the ligand sequence. Of course one could make all 70 libraries of 8 residues taken 4 ways ($n!/[x!(n-x)!]$), but enough information is usually obtained with a very few such libraries. A variation on this theme may be employed when, in early screening, an extraordinary conservation of particular amino acids at certain positions is observed. A sublibrary is made to fix (or lightly mutate) these residues and randomize the remaining positions.

Finally, short, random regions can be added to the starting sequence. Although, in general, reducing the size of the initial lead is desirable, peptides of higher affinity may be obtained from libraries constructed to contain short random regions flanking the starting sequence. Selection of the higher affinity members of mutagenesis libraries requires the application of screening conditions sensitive to intrinsic affinity differences among the library members, as discussed in Part 2 of this Perspective (scheduled to appear in *J. Med. Chem.* **1994**, *37* (10)). The construction and screening of mutagenesis libraries serves to compile structure–activity relationships around analogs of a starting compound. As such, the process is smoothly coupled to the parallel synthetic library techniques described in the next section, extending the array of building block monomers utilized in the analog constructions.

C. Synthetic Chemical Approaches to Generating Peptide Diversity

During the last two decades peptide chemistry has been steadily attaining ever greater prominence in pharmaceutical research, fueled, in part, by the discovery of many biologically active peptide hormones and an increased understanding of the roles that these molecules play in regulating a multitude of human physiological responses. Consequently many low molecular weight bioactive peptides have become the focus of medicinal chemistry research efforts to develop hormone receptor agonists or antagonists.[1] In addition, peptidic structures commonly are found in molecules designed to inhibit enzymes that catalyze proteolysis, phosphorylation, and other post-translation protein modifications that may play central roles in the pathologies of various disease states.[81]

The desire to develop and explore structure–activity relationships around peptide lead compounds has placed tremendous demands on the productivity of peptide chemistry. Over the last 10 years a variety of methodologies have been developed that permit simultaneous synthesis of multiple peptides, and this area has recently been thoroughly reviewed.[82] Because many of the technologies underlying the construction of diverse synthetic peptide libraries are rooted in multiple peptide synthesis (MPS) strategies, a brief overview of the principal MPS methods and their applications is provided below.

Multiple Peptide Synthesis. The various approaches to simultaneous preparation and analysis of large numbers of synthetic peptides each rely on the fundamental concept of synthesis on a solid support introduced by Merrifield in 1963.[83] Generally, these techniques are not dependent on the protecting group or activation chemistry employed, although most workers today avoid Merrifield's original ^{t}Boc/Bzl strategy in favor of the more mild Fmoc/^{t}Bu chemistry and efficient hydroxybenzotriazole-based coupling agents. Many types of solid matrices have been successfully used in MPS, and yields of individual peptides synthesized vary widely with the technique adopted (e.g., nanomoles to millimoles). Interestingly, the early architects of these methodologies were molecular immunologists, primarily concerned with determining the immunodominant B-cell epitopes of viral proteins and in deducing the fine specificity of peptide recognition by monoclonal antibodies.

(A) Multipin Synthesis. Geysen and co-workers[84] introduced a method for generating 96 peptide sequences by a parallel synthesis on polyacrylic acid-grafted polyethylene pins arrayed in the microtitre plate format. In the original experiments ~50 nmol of (ideally) a single peptide sequence was covalently linked to the spherical head of each pin, and interactions of each peptide with monoclonal antibodies or polyclonal sera could be determined in a direct binding assay by ELISA. The Geysen group claims to be able to synthesize and screen thousands of peptides per week using the multipin method, and the tethered peptides may be reused in many assays. In subsequent work the level of peptide loading on individual pins was increased to as much as 2 μmol/pin by grafting greater amounts of functionalized acrylate derivatives to detachable pin heads.[85] Appropriate linker moieties have also been appended to the pins so that the peptides may be cleaved from the supports after synthesis for assessment of purity and evaluation in competition binding or functional bioassays.[86–88]

The most significant applications of the multipin technology have been in the area of anti-protein antibody epitope analysis[89,90]. Continuous epitopes are readily mapped by synthesizing on pins a complete set of

overlapping peptides derived from the antigen's primary structure and measuring the reactivity of the antibody or antiserum by ELISA (Pepscan technique[89]). For example, a protein antigen of 200 amino acids may be scanned by evaluating antibody interactions with 193 overlapping octapeptide fragments. Upon its identification, the fine specificity of an antibody epitope can be further deduced by "replacement set analysis" in which the effect on antibody recognition of substituting each of the amino acids within the epitope by other amino acids is observed. Those residues which poorly tolerate substitution are implicated as being directly involved in antigen–antibody interaction. Antibody binding sites on a wide range of viral proteins have been elucidated by the Pepscan method.[82,84,91–93] More recent applications of the multipin method of MPS have taken advantage of the cleavable linker strategy to prepare soluble peptides for T-cell proliferation studies.[94–96] An investigation of the binding activity of 512 stereoisomers of the tachykinin peptide, substance P, generated by synthesizing all possible D-amino acid replacements at 9 of the 11 positions in this neuropeptide was recently reported.[97,98] Radioligand binding experiments indicated that while substitutions toward the C-terminus of the peptide generally abrogated NK1 receptor binding, substitutions in the N-terminal region were often well tolerated. In structure–activity relationship studies of a hexapeptide endothelin receptor antagonist conducted by Spellmeyer *et al.*, multipin synthesis was used in the systematic substitution of 50 different amino acids at each position in the lead peptide.[99,100] Competition binding assays run with crude peptides cleaved from the pins provided an initial screen from which a number of more potent analogs were identified.

(B) Tea-Bag Method. In the so-called "tea-bag" MPS method first developed by Houghten, peptide synthesis occurs on resin that is sealed inside porous polypropylene bags.[101] Amino acids are coupled to the resins by placing the bags in solutions of the appropriate individual activated monomers, while all common steps such as resin washing and α-amino group deprotection are performed simultaneously in one reaction vessel. At the end of the synthesis, each bag contains a single peptide sequence, and the peptides may be liberated from the resins using a multiple cleavage apparatus.[102] This technique offers advantages of considerable synthetic flexibility and has been partially automated to permit the simultaneous synthesis of up to 150 different peptides.[103] Moreover, soluble peptides of greater than 15 amino acids in length can be produced in sufficient quantities (*ca.* 500 μmol) for purification and complete characterization if desired.

Multiple peptide synthesis using the tea-bag approach has been applied to a range of molecular recognition problems including antibody epitope analysis,[101] peptide hormone structure–function studies,[104,105] and protein conformational mapping.[106] For example, using a combination of alanine and D-amino acid scanning, the C-terminal tetrapeptide fragment of the 36-amino acid peptide hormone, neuropeptide Y, has been shown to be essential for receptor recognition. Neuropeptide Y analogs that link regions of the amino and carboxy termini of the parent peptide via a spacer residue were shown to behave as receptor agonists, mimicking a discontinuous pharmacophore of the hormone.[104,105] Families of synthetic peptides modeled around the Arg-Gly-Asp and γ chain fragments of fibrinogen have been prepared using the tea-bag method to define antagonists of the platelet integrin receptor gpIIb/IIIa.[107]

(C) Multiple Peptide Synthesis through Coupling of Amino Acid Mixtures. Simultaneous coupling of mixtures of activated amino acids to a single resin support has been used as a multiple peptide synthesis strategy on several occasions.[90,108–111] For example, four to seven analogs of the magainin 2 and angiotensinogen peptides were successfully synthesized and resolved in one HPLC purification after coupling a mixture of amino acids at a single position in each sequence.[108] This approach has also been used by two groups to prepare degenerate peptide mixtures for defining the substrate specificity of endoproteolytic enzymes.[110,111] In these experiments a series of amino acids was substituted at a single P′ position within the substrate sequence. After proteolysis, Edman degradation was used to quantitate the yield of each amino acid component in the hydrolysis product and hence to evaluate the relative k_{cat}/K_m values for each substrate in the mixture.

The operational simplicity of synthesizing many peptides by coupling monomer mixtures is offset by the difficulty in controlling the composition of the products. The product distribution reflects the individual rate constants for the competing coupling reactions, with activated derivatives of sterically hindered residues such as valine or isoleucine adding at a significantly slower rate than glycine or alanine for example. The nature of the resin-bound component of the acylation reaction also influences the addition rate, and the relative rate constants for the formation of 400 dipeptides from the 20 genetically coded amino acids have been determined by Rutter and Santi.[109] These reaction rates can be used to guide the selection of appropriate relative concentrations of amino acids in the mixture to favor more closely equimolar coupling yields.

(D) Multiple Peptide Synthesis on Nontraditional Solid Supports. The search for innovative methods of multiple peptide synthesis has led to the investigation of alternative polymeric supports to the polystyrene–divinylbenzene matrix originally popularized by Merrifield. Cellulose, either in the form of paper disks[112–115] or cotton fragments[116,117] has been successfully functionalized for peptide synthesis. Typical loadings attained with cellulose paper range from 1 to 3 μmol/cm^2, and HPLC analysis of material cleaved from these supports indicates a reasonable quality for the synthesized peptides. Alternatively, peptides may be synthesized on cellulose sheets via noncleavable linkers and then used in ELISA-based receptor binding studies.[118] The porous, polar nature of this support may help suppress unwanted nonspecific protein binding effects. By controlling the volume of activated amino acids and other reagents spotted on the paper, the number of peptides synthesized at discreet locations on the support can be readily varied. In one convenient configuration spots are made in an 8×12 microtiter plate format. Frank has used this technique to map the dominant epitopes of an antiserum raised against a human cytomegalovirus protein, following the overlapping peptide screening (Pepscan) strategy of Geysen.[118] Other membrane-like supports that have been investigated for multiple solid-phase synthesis include polystyrene-grafted polyethylene films.[119]

Automated Multiple Peptide Synthesis Methods. There has been considerable effort devoted to constructing

automated multiple peptide synthesizers, and several designs have now been successfully commercialized. The groups of Schnorrenberg[120] and Gausepohl[121] have employed pipetting robots to deliver solvents, activated amino acids and other reagents to peptide resins contained within an array of 96 and 48 reaction vessels, respectively. Individual peptides may be synthesized on a scale of 10–50 μmol using these instruments. Zuckermann *et al.* have designed a related robotic delivery system that can simultaneously prepare 36 peptides (at up to 125 μmol per peptide) when operated in multiple synthesis mode.[122] This workstation has the additional capability of mixing and redistributing resin from multiple reaction vessels which significantly facilitates the assembly of combinatorial peptide libraries (*vide infra*). Although the operation of these instruments differs in details such as resin agitation and washing, similar synthesis chemistries can be utilized with each, and the purities and yields of the resulting peptides obtained are typically respectable.

Synthesizers based on the continuous flow principle[123] have also been successfully used to prepare milligram quantities of multiple peptides.[113,124,125] Novel, porous polymer supports including cellulose and poly(dimethylacrylamide) have typically been used in these column reactor formats. Other innovations in multiple synthesis relate to strategies for automating the deprotection and cleavage of many peptides in parallel.[126,127] An integrated multichannel synthesizer and cleavage device has recently been described by Neimark and Briand.[127]

Combinatorial Synthetic Peptide Libraries. The generation of molecular diversity using strategies that covalently connect together members of a set of chemical building blocks in all possible combinations represents a revolution in multiple synthesis. Parallel development of a variety of new high-throughput screening methodologies has made evaluation of these combinatorial libraries in biological assays both practical and efficient. The various technologies that have emerged for generating and screening peptide libraries may be fundamentally distinguished by the format in which the diversity is presented (tethered vs soluble libraries; physically segregated ligands vs mixed pools). This in turn dictates the type of biological assay methods that may be utilized, influences the strategy followed in ligand structure elucidation, and ultimately determines the size of libraries that may be practically screened. The common hallmark of these combinatorial techniques is that within each chemical coupling step, multiple compounds are generated simultaneously such that each synthesis cycle results in an exponential increase in library size.

Two distinct mechanisms for elaborating molecular diversity may be defined. In the first, mixtures of activated monomers are coupled to one (or more) solid supports at each cycle of the synthesis.[90,109,128,129] As previously noted, the product distribution here is influenced by the relative kinetics of the competing reactions, making it difficult, in general, to ensure an equimolar representation of each component in the library. Ligand equimolarity is an important criterion if one wishes to make even semiquantitative determinations of the affinities of receptor–ligand interactions by equilibrium measurements with compound mixtures. The second mechanism circumvents the problem of competing coupling reactions by physically segregating the support into multiple aliquots or spatially discreet fractions to which addition of a single monomer

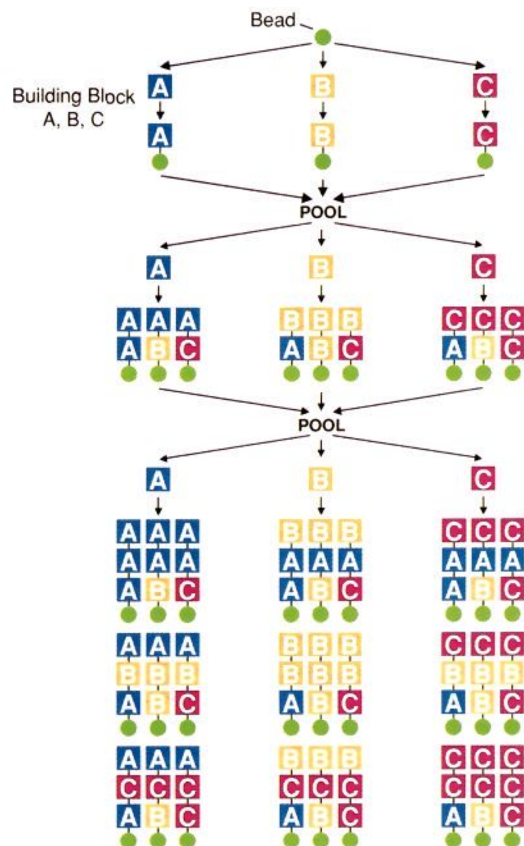


Figure 2. Preparation of a combinatorial library by the split synthesis method.

can occur. The earliest description of this approach was the "portioning–mixing" method of peptide synthesis of Furka and co-workers,[130–132] and termed the "split synthesis"[133] or "divide, couple, and recombine" process[134] by other workers. This simple procedure involves dividing the resin support into n equal fractions, coupling a single monomer to each aliquot in a separate reaction, and then thoroughly mixing all the resin particles together. Repeating this protocol for a total of x cycles can produce a stochastic collection of up to n^x different peptides, as governed by the Poisson distribution. A schematic illustration of a split synthesis using three building blocks A, B, and C to generate all 27 possible trimer combinations is shown in Figure 2.

Combinatorial Libraries Using Multipin Synthesis. The problem of finding peptide ligands that bind to monoclonal antibodies which recognize discontinuous epitopes within a protein antigen was first approached by Geysen using a combinatorial synthesis procedure referred to as the "mimotope strategy".[89,90] Peptide mixtures are synthesized on pins in the format *-*-*-D_1-D_2-*-*-* where the positions D_1 and D_2 are occupied by single defined amino acids and the * positions represent randomly incorporated residues resulting from coupling mixtures of activated amino acids. Using the 20 common L-amino acid monomers, a total of 400 such mixtures may be synthesized, each ideally containing $20^6 = 6.4 \times 10^7$ different sequences. The peptide mixture on each pin is screened for antibody binding by ELISA to identify an optimum dipeptide sequence X-Y. This sequence then provides the basis for a further round of synthesis in the format *-*-D-X-Y-*-*-*, where the number of degenerate positions is reduced by one, allowing resolution of a third

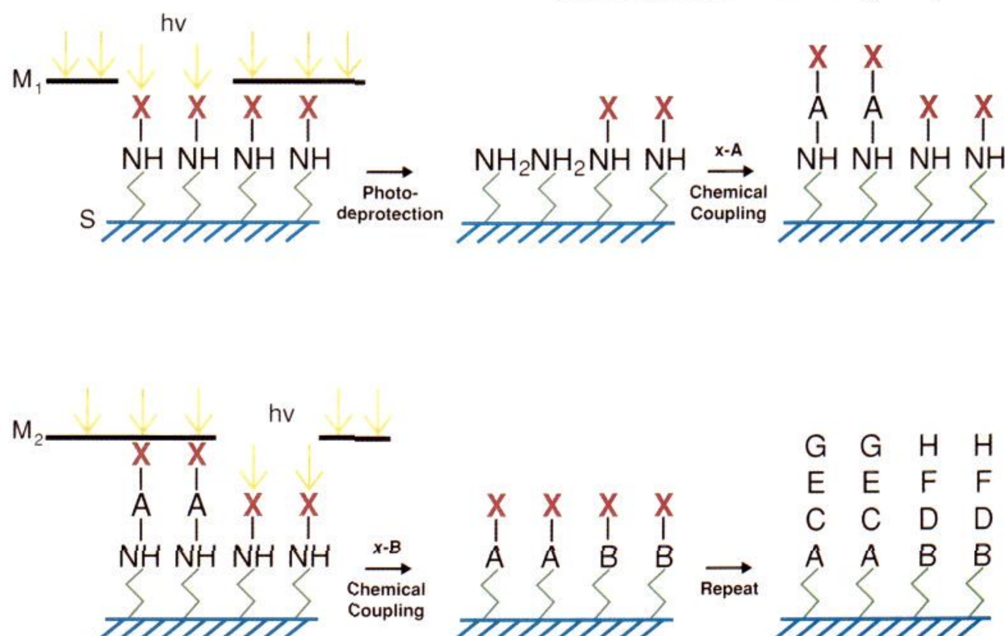


Figure 3. Concept of light-directed spatially addressable parallel chemical synthesis. A substrate S bears amino groups that are blocked with a photolabile protecting group X. Illumination of specific regions through a lithographic mask M_1 leads to photodeprotection. Amino groups in the exposed sector of the substrate are now accessible for coupling (for example by the Merrifield solid-phase peptide synthesis method). The first chemical building block A containing a photolabile protecting group X is then attached. A different mask M_2 is used to photoactivate a different region of the substrate. A second labeled group X–B is added and condensed to the newly exposed amino groups. Additional cycles of photodeprotection and coupling are carried out to obtain the desired set of products.

residue in the sequence. This cycle of synthesis and screening is reiterated until the entire sequence is optimized, and the resulting peptide is termed a "mimotope" (for epitope mimetic) for the antibody.

Geysen has recently reviewed use of the mimotope strategy for the identification and optimization of ligands for antibodies and other receptors.[135] A preferred approach is to screen hexapeptides that incorporate both L- and D-amino acids ($39^6 = 3.52 \times 10^9$ hexamers from the common monomers), though non-α or other nonnatural amino acids can certainly be used in these syntheses.

Combinatorial Libraries by Light-Directed, Spatially Addressable Parallel Chemical Synthesis. A scheme of combinatorial synthesis in which the identity of a compound is given by its location on a synthesis substrate is termed a spatially addressable synthesis. Here, the combinatorial process is carried out by controlling the addition of a chemical reagent to specific locations on a solid support.[4,136–138] The technique combines two well-developed technologies: solid-phase peptide synthesis chemistry and photolithography. The high coupling yields of Merrifield chemistry allow efficient peptide synthesis, and the spatial resolution of photolithography affords miniaturization. The merging of these two technologies is done through the use of photolabile amino protecting groups in the Merrifield synthetic procedure.

The key points of this technology are illustrated in Figure 3. A synthesis substrate is prepared for amino acid coupling through the covalent attachment of photolabile nitroveratryloxycarbonyl (NVOC) protected amino linkers. Light is used to selectively activate a specified region of the synthesis support for coupling. Removal of the photolabile protecting groups by light (deprotection) results in activation of selected areas. After activation, the first of a set of amino acids, each bearing a photolabile *protecting group on the amino terminus, is exposed to the* entire surface. Amino acid coupling only occurs in regions that were addressed by light in the preceding step. The solution of amino acid is removed, and the substrate is again illuminated through a second mask, activating a different region for reaction with a second protected building block. The pattern of masks and the sequence of reactants define the products and their locations. Since this process utilizes photolithographic techniques, the number of compounds that can be synthesized is limited only by the number of synthesis sites that can be addressed with appropriate resolution. The position of each compound is precisely known; hence, its interactions with other molecules can be directly assessed. The assay molecule can be labeled with a fluorescent reporter group to facilitate the identification of specific interactions with individual members of the matrix. This technology requires only micropreparation and consumption of chemical reagents as well as micropreparation and consumption of biological targets.

The light-directed synthesis of an array of two peptides YGGFL and PGGFL was performed. Following synthesis, the pentapeptide array was probed with a mouse monoclonal directed against β-endorphin (mAb 3E7). This antibody requires an amino-terminal tyrosine residue for high affinity recognition. A second incubation with fluorescein-labeled goat-anti-mouse IgG was used to label the regions containing bound 3E7. The array was then scanned in an epifluorescence microscope. The result of the scan is shown in Figure 4. A high-contrast fluorescence checkerboard image shows that YGGFL and PGGFL were synthesized in alternating 50-μm squares, that the YGGFL synthesized on the surface is accessible for binding to the antibody 3E7, and that the antibody does not recognize PGGFL. The contrast between synthesis sites is very high at 50-μm resolution. This spatial resolution provides 40 000 discrete synthesis sites in a 1-cm × 1-cm square.

In a light-directed chemical synthesis, the products depend on the pattern of illumination and on the order

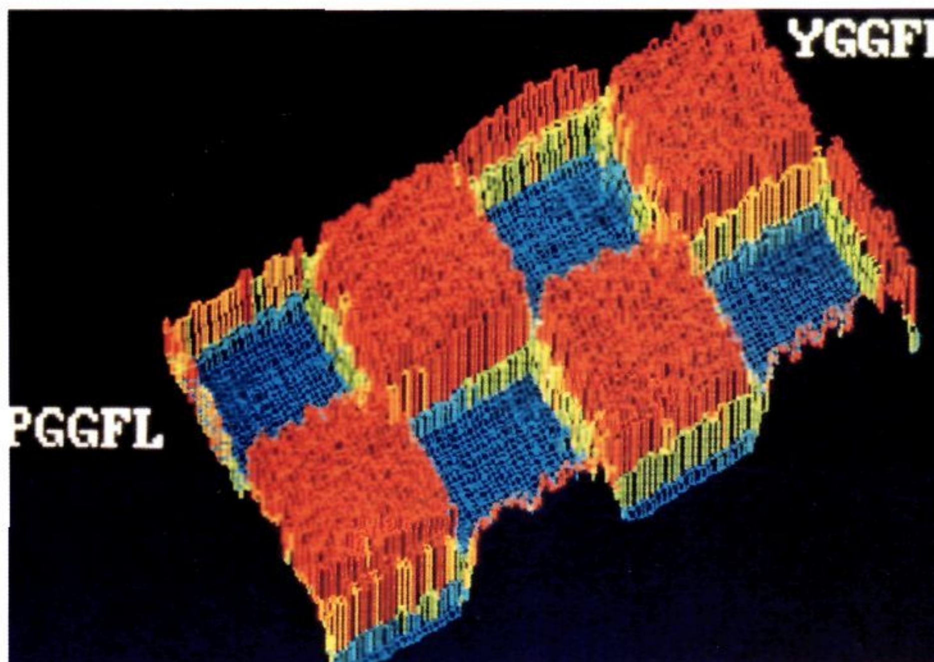


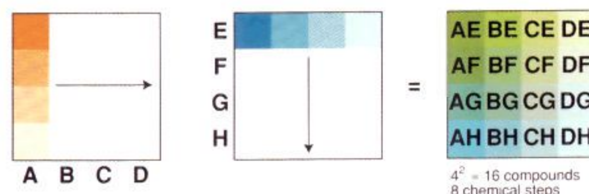
Figure 4. Detection of binding events with laser confocal fluorescence microscopy. After synthesis, the array of compounds is mounted on a thermostatically regulated flow cell and then interrogated with fluorescently-labeled target molecules. Binding events are detected by their fluorescence emissions, which are scored with a photomultiplier tube as described in the text.

of addition of reactants. By varying the lithographic patterns, many different sets of compounds can be synthesized in the same number of steps; this leads to the generation of many different masking strategies. As an example, an eight-step synthesis is shown in Figure 5. The reactants are the ordered set {A, B, C, D, E, F, G, H}. In the first cycle, illumination through photolithographic mask M_1 activates the left one-quarter of the synthesis area. Building block A is added and reacts only at the previously illuminated region. Illumination through M_2 activates the second one-quarter of the substrate, followed by addition of building block B. This process is continued through the first four cycles and results in the pattern of compounds shown in the first half of Figure 5a. M_5 then illuminates *across* the patterns generated in the first four cycles. This one illumination through M_5 now generates four distinct dimers. The process is continued to generate the set of 16 compounds shown in the Figure 5. The final product set is {AE, AF, AG, AH, BE, BF, BG, BH, CE, CF, CG, CH, DE, DF, DG, DH}.

This strategy can be applied to a large set of building blocks. For example, the complete set of 400 dipeptides can be formed from the set of 20 L- or D-amino acids. Analogous to Figure 5a, the synthesis may be performed via two rounds of 20 coupling cycles per round. In cycle 1 of the first round, M_1 activates 1/20th of the substrate for coupling with the first of 20 amino acids. An additional 19 translations and illuminations followed by 20 coupling cycles are required to complete round 1. The substrate now consists of 20 rectangular stripes each bearing a distinct member of the 20 amino acid "building blocks". The masks of round 2 are perpendicular to round 1 masks, and therefore, a single illumination–coupling cycle in round 2 yields 20 dipeptides. The other 19 cycles in round 2 complete the synthesis of 400 dipeptides.

A general formalism describes the combinatorial synthesis for any spatially addressable chemical synthesis. The process is coveniently expressed in matrix notation[136]

A. Orthogonal Stripes



B. Binary Synthesis

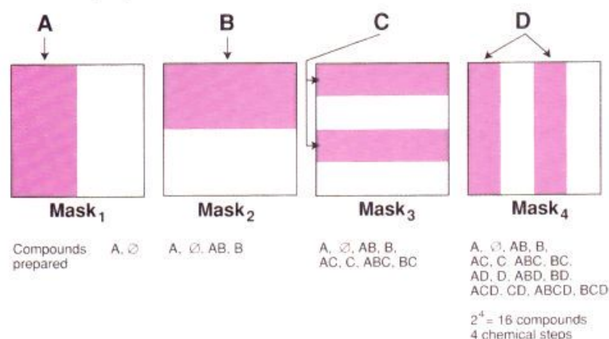


Figure 5. Synthesis strategies. (a) Orthogonal stripe method. Using a strategy similar to the split-resin method, a layer of monomers is formed by photolyzing stripes for each building block. Dimers are formed by photolyzing stripes orthogonal to the first set, preparing n^2 compounds in $2n$ chemical steps. (b) Binary synthesis. Half of the synthesis surface is photolyzed during each coupling step, with subsequent photochemical steps overlapping one-half of the previous synthesis space. With this strategy, 2^n compounds are made in n chemical steps.

and is based on the notion that at any given synthesis site the decision as to whether or not to add a monomer of the building block set is a binary process. If a particular position is to receive a new monomer group, it is addressed

by a 1 (light on condition), if not, it receives a 0. For example, to form ACD from the ordered set {A, B, C, D}, the light switch consists of a vector (switch vector) [1,0,1,1]. Each product element is, therefore, given by the dot product of two vectors, a chemical reactant vector and a binary vector ([A,B,C,D]·[1,0,1,1] = ACD in the previous example). This formalism is very powerful in defining a complex chemical synthesis. For example, consider a four-step synthesis with the ordered set of reactants {A, B, C, D}. Any switch vector consists of four bits for the four chemical reactants. There are 16 four-bit vectors [0,0,0,0] to [1,1,1,1]; hence, a maximum of 16 different products can be made from the ordered set of four reagents (one of the four compounds is the null or zero compound). For the example shown in Figure 5a, there was an ordered set of eight chemical reactants, and 16 compounds were formed. A more efficient synthesis strategy would create all 2^8 (256) possible compounds. Thus, a "binary masking strategy" where the maximum number of compounds are generated in the fewest number of steps when using pure monomers in each step (Figure 5b) was developed. For n number of steps, 2^n number of compounds are generated. Since one of these compounds is the linker where no polymer has been assembled, $2^n - 1$ polymers are made in this strategy. A 16-step binary synthesis, for example, would result in 2^{16} (65 536) regions for screening, of which 65 535 would represent peptides or oligonucleotides. This strategy also creates a distribution of chain lengths, ranging from 1 to n, with the maximum number of polymers being $n/2$ in length. The binary strategy generates all possible permutations of the building blocks while preserving their order of introduction. Hence all deletion sequences possible in the starting sequence are obtained, as well as all possible truncation sequences. This type of strategy has proven useful in epitope mapping experiments with an anti-dynorphin B antibody, D32.39. A binary strategy was employed to generate all permutations about the C-terminus region of dynorphin B, from which RQFKVVT was identified as the minimal binding subunit.[139-141]

There are circumstances where biopolymers in the array should be the same length, and other masking strategies have been developed to accomodate these applications. It may also be desirable to generate sets of biopolymers where all members of a certain class are represented in one array and to rank order the *relative* affinity of each member of the set. A common strategy in ligand optimization is the substitution of a set of novel monomers (e.g., 20 L-amino acids or D-amino acids) for each monomer in a lead compound. This can be accomplished with relatively few steps in a combinatorial synthesis. The number of steps required to substitute x monomers into a polymer of length n is given by $2(n - 1) + x$. For example, substitution of 20 novel amino acids into each position of an octapeptide represents 160 peptides (8 × 20), which would require up to 1280 individual steps (160 × 8) if carried out conventionally; combinatorial strategies can create an array of these 160 peptides in 34 reactions.

The maximum number of compounds that can be made in the array is governed largely by the lithographic performance of the system. Practical limitations on spatial resolution, or the desire for a specific class of compounds, may require one to select a particular subset of the possible compounds for a given synthesis. Among other things, the masking resolution is dependent on the wavelength of the light employed and the distance between the mask and the substrate. It has been possible to generate arrays of up to 65 536 compounds in an area slightly greater than 1 cm^2. Further improvements in resolution will allow even higher density arrays to be generated.

Combinatorial Libraries of Soluble Peptides. The "split synthesis" algorithim is readily adapted to generating equimolar mixtures of soluble peptides that may be screened in a variety of competition binding or functional bioassays. This approach has been most extensively investigated by Houghten *et al.*, and bioactive peptides have been identified from libraries containing more than 50 million different sequences.[134,142-146] These hexamer libraries consist of sets of sublibraries that are conveniently prepared by the tea-bag technique in one of two formats: (i) the dual-fixed position series D_1-D_2-*-*-*-* in which a mixture of amino acids is randomly incorporated at four positions of the peptide while two residues (D_i) are explicitly defined, and (ii) the positional scanning series D_1-*-*-*-*-*, *-D_2-*-*-*-*, *-*-D_3-*-*-*, *-*-*-D_4-*-*, *-*-*-*-D_5-*, and *-*-*-*-*-D_6 where single positions are defined throughout the sequence. For libraries constructed from 20 amino acids, 400 distinct dual position sublibraries and 120 positional scanning sublibraries must be synthesized. Active peptides are identified from dual-position libraries by an iterative process of screening and sublibrary resynthesis in a manner that is completely analogous to the mimotope resolution strategy. Thus at each stage of the screening, an additional residue is defined explicitly by evaluating 20 sublibraries in the assay. By contrast, the positional scanning strategy is used to define the most preferred residue or residues at each position of the sequence. All possible combinations of amino acids with activity exceeding some threshold level are then prepared and tested to identify explicit sequences with potent activities.[143,144]

Using competition ELISA experiments to map the linear epitopes recognized by two different monoclonal antibodies, both screening strategies converged to equivalent peptide sequences.[134,142,143] Similarly, either dual fixed residue or positional scanning screening of these hexapeptide libraries rediscovered the methionine and leucine enkephalin peptide families as competitive inhibitors of the binding of the μ-opioid-specific ligand [^{3}H]DAMGO to crude rat brain homogenates.[144,145] As the presence of a positively charged N-terminal amino group is generally regarded as a critical determinant for high-affinity peptide binding at the opioid receptors, the recent finding that potent ligands can identified from an N-terminally acetylated hexapeptide library is of particular interest.[146] The peptides Ac-RFMWMT-NH_2, Ac-RFMWMK-NH_2, and Ac-RFMWMR-NH_2 have low nanomolar affinities for the μ-, δ-, and κ_3- receptors and behave as potent μ-antagonists in the guinea pig ileum assay. Removal of the N-terminal acetyl group from these peptides results in a 500–1000-fold decrease in affinity for μ-receptor, and the compounds have been termed "acetalins" to reflect the critical contribution of the N-terminus to the opioid binding characteristics.[146] Another noteworthy application of these soluble hexamer libraries includes the discovery of novel antimicrobial peptides with activities against both Gram-positive and Gram-negative bacteria by iterative microdilution assays.[134,142]

Inhibitors of the human immunodeficiency virus (HIV) protease have been identified by screening sets of equimolar peptide mixtures, together containing more than

240 000 soluble tetrapeptides.[147] The library incorporated the aspartyl protease transition-state mimetic statine at one position of the tetramer sequence, along with 22 other amino acid monomers. The preferred residue at each position in the sequence was identified through the standard iterative cycle of assay and library resynthesis and yielded the peptide Ac-Phe-Ile-Sta-D-Leu-NH_2 (IC_{50} value of 1.4 μM) as the most potent protease inhibitor. This compound provided the basis for further substitutional analysis and was elaborated to a pentapeptide analog with an IC_{50} value of 5 nM. Antibody epitopes have been mapped by another iterative assay strategy using competition ELISA in a way that attempts to minimize the total number of sublibraries that need be synthesized and screened.[129]

Zuckermann and co-workers have employed an automated robotic synthesizer to prepare equimolar peptide mixtures according to the split synthesis method.[148] A library of 361 different compounds formed by randomly incorporating 19 amino acids at each of two positions within a decapeptide known to bind an anti-gp120 monoclonal antibody was screened with this antibody in a two-step procedure. After establishing an amino acid preference at one of the degenerate positions by competition ELISA measurements, the optimal substitution at the other residue was identified by affinity purifying a 19-component peptide mixture on the antibody using gel-filtration chromatography. The affinity-selected peptides were identified by mass spectrometry and amino acid analysis, and the relative amounts of each component recovered in the purification correlated with the peptides' receptor-binding affinities. This strategy has also been applied to the analysis of a 512-component hexamer library incorporating a limited set of nonnatural amino acid substitutions.[149]

Peptide Libraries Tethered to Solid Supports. When the split synthesis method of library construction is performed with beaded resin equipped with a noncleavable linker moiety, the resulting libraries consist of collections of particles, each bearing (ideally) a single peptide sequence. These libraries can be screened for interaction with soluble receptors bearing some reporter group, and individual beads that bind the target protein subjected to Edman microsequence analysis to determine the sequence of the corresponding peptide ligands.[133,150,151] Lam *et al.* have used this method to prepare a library of $\sim 2.5 \times 10^6$ pentapeptides from 19 L-amino acids. In one assay the library was screened for binding to an anti-β-endorphin mAb 3E7 that was conjugated to alkaline phosphatase, and beads that were stained by ELISA were selected with a micromanipulator and sequenced.[133,150] The majority of the deduced sequences were clearly enkephalin-related, and as soluble peptides, several had affinities for the mAb within the range of 15 nM to 10 μM. That lower affinity ($K_i > 10$ μM) enkephalin analogs could also be identified in this screen was attributed to the presence of a high density of ligands on the bead surface simulating a high concentration of the ligand in solution.[150] Although this explanation is probably partially accurate, it is likely that the bivalency of the antibody used in this assay permits the formation of high avidity, multivalent interactions between the receptor and tethered ligands with low intrinsic affinity for the mAb. This method of library synthesis and ligand identification has been subsequently adapted for screening soluble peptide mixtures. Incorporating multicleavable linker groups into the beads permits the release of equimolar quantities of ligand at independent steps, such that after the screening sufficient peptide is retained on each bead for Edman sequence analysis.[152]

The nature of the bead material used in the construction and screening of support-tethered libraries has an important impact on the utility of this display format in ligand discovery. The beads must be resistant to all the organic solvents and reagents used in ligand synthesis and yet be well-behaved in the aqueous media where protein receptor binding studies are conducted. In particular, the polymer support should have low nonspecific protein binding characteristics to maximize the opportunity of detecting specific receptor interactions with the immobilized ligands. Some workers have preferred to use receptors that are directly radiolabeled in their binding studies with bead libraries to minimize the background signals associated with enzyme-linked immunoassays.[151] The choice of the linker or spacer moiety on the bead from which the compounds are synthesized can also influence the accessibility of the ligands to macromolecular receptors in solution. Poly(dimethylacrylamide) resins derivatized with ethylenediamine, aminocaproic acid and β-alanine spacers have been successfully used in several studies.[133,150] Recently, a polyethylene-grafted cross-linked polystyrene resin termed TentaGel has been made commercially available by Rapp Polymere (Tubingen, Germany) for use as a carrier in tethered peptide library construction. The hydrophilic poly(oxyethylene) spacer moiety is equally well solvated in organic and aqueous solvents, and the relatively monodisperse bead material is available in a variety of diameters and with several different reactive functional groups (e.g., NH_2, OH, SH, Br, CO_2H). The 90-μm-diameter NH_2 beads for peptide synthesis give rise to *ca.* 2.9×10^6 beads/g of resin with a loading of 80–100 pmol per bead. Since the detection limit in conventional phenylthiohydantoin (PTH)-based Edman peptide microsequence analysis is *ca.* 1–5 pmol, each bead provides adequate material for direct ligand analysis after selection from a binding assay.

Encoded Combinatorial Synthesis. An important goal in screening diverse libraries of molecules produced through combinatorial synthesis is, of course, the identification of novel structures that interact with receptor targets of biological interest. For any given target protein, the probability of successfully identifying potent ligands through a process of randomly screening molecular repertoires in affinity-sensitive assays will undoubtedly increase as the size and structural diversity of the library is also increased. The generation and screening of very large combinatorial libraries assembled from an expanded set of molecular building blocks provides substantial challenges to the diversity technologies previously outlined in this section.

For example, to screen an immobilized synthetic peptide library containing 10^9 members constructed from the bead support described above would require treating kilograms of resin with liters of soluble receptor, clearly an impractical undertaking. A decrease in the bead diameter might permit a library of this diversity to occupy a manageable volume, but the quantity of peptide associated with any selected particle would quickly fall below the threshold required for microsequence analysis. The ability to dramatically expand the amino acid monomer set to

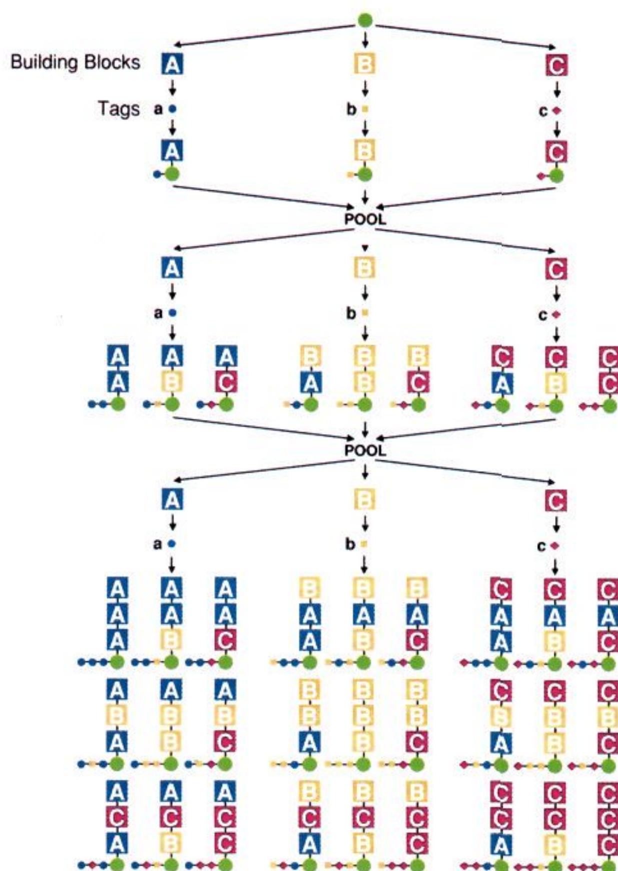


Figure 6. Schematic assembly of an encoded synthetic library.

include D-amino acids and nonnatural amino acids is also limited by the necessity of resolving chromatographically the PTH-derivatives of every building block used in the library. Moreover, direct ligand microsequencing remains an analysis option for peptides (Edman degradation) and oligonucleotides (Sanger dideoxy sequencing) only.

The iterative strategy of assaying compound mixtures and then synthesizing and retesting sublibraries avoids the problems associated with direct ligand structural analysis. However, this process can be very laborious, particularly with large libraries, since unique structural deconvolutions that lead to potent ligands are likely to be very rare. Furthermore, the complexity of nontethered libraries is limited in practice by the solubility of the mixtures, since the concentrations of individual components in the pools eventually become too small to exert any independent measurable effect on receptor binding or function.

In attempting to address some of these limitations, it was recognized that the products of a combinatorial synthesis on resin beads could be explicitly specified if it were possible to attach an identifier tag to the beads coincident with each monomer coupling step in the synthesis.[153] Each tag would then convey which monomer was coupled in a particular step of the synthesis, and the overall structure of a ligand on any bead could be deduced by reading the set of tags on that bead. An encoded version of the combinatorial synthesis previously outlined in Figure 2 is illustrated schematically in Figure 6.

Ideally, such tags should have a high information content, be amenable to very high sensitivity detection and decoding, and must be stable to reagents used in the ligand synthesis. A method using single-stranded oligonucleotides to encode combinatorial peptide syntheses on 10-μm-diameter polystyrene beads has recently been developed.[154] Peptides and nucleotides are assembled in parallel, alternating syntheses so that each bead bears many copies of both a single peptide sequence and a unique oligonucleotide identifier tag. The average stoichiometry of peptide to oligonucleotide per particle is readily controlled by varying the ratio of differentially functionalized linker residues coupled to the beads and can be heavily weighted in favor of the peptide molecules. Each amino acid monomer used in the synthesis is assigned a distinct contiguous nucleotide sequence or "codon" and hence the structure of the peptide assembled on any bead is reflected in the oligonucleotide sequence of the corresponding tag. The oligonucleotides share common 5'- and 3'-PCR priming sites and thus the beads can serve as templates for the PCR.

An oligonucleotide-encoded synthetic library of approximately 8.2×10^5 heptapeptides was screened for binding to an anti-peptide monoclonal antibody using a fluorescence activated cell sorting (FACS) instrument to select individual beads that strongly bind the antibody.[154] After PCR amplification of the oligonucleotide tags on sorted beads, the DNA was sequenced to determine the identity of the peptide ligands. This experiment clearly defined a consensus peptide sequence recognized by the mAb, and the affinities of peptides identified in this assay for the antibody ranged from 0.3 nM to 1.4 μM.

The concept of an oligonucleotide-encoded chemical synthesis has also been proposed independently by Brenner and Lerner.[155,156] They envisaged screening a library of soluble ligand–oligonucleotide hybrid molecules for receptor binding and proposed that the nucleotide tag could serve both as a coding device and as a handle for selectively enriching active ligands from the pool. More recently this group has disclosed the design of an orthogonally-protected polyfunctional linker group that can be used for the parallel, alternating solid-phase synthesis of peptides and oligonucleotides. They report that soluble peptide–oligonucleotide chimeras generated on this matrix can bind as anticipated to an anti-peptide antibody, as well as functioning as templates for amplification by the PCR.[157]

Zuckermann and co-workers have shown that by starting with an orthogonally differentiated diamine linker, parallel combinatorial synthesis can be used to generate a library of soluble chimeric peptides comprising a "binding" strand and a "coding" strand.[158] The coupling of either natural or nonnatural amino acid monomers to the binding strand was recorded by building an amino acid code comprised of four L-amino acids on the "coding" strand. Compounds were selected from equimolar peptide mixtures by affinity purification on a receptor and were resolved by HPLC. The sequence of the coding strand of individual purified molecules was then determined by Edman degradation to reveal the structure of the binding strand. An analogous peptidic coding scheme has also been recently described in detail by Nikolaiev *et al.*[159] With the appropriate choice of linker residues, either the binding or coding sequences or both can be cleaved from the resin. Thus it is possible to screen for receptor binding to tethered and to soluble ligands using this technique and to separate the contributions of the ligand and coding sequences to receptor interaction.

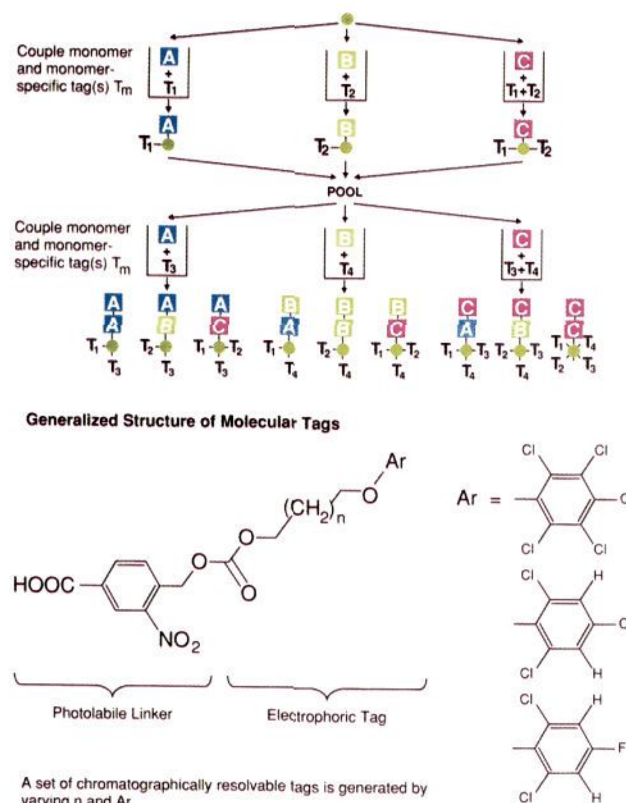


Figure 7. Binary encoded synthesis using gas chromatographically detectable chemically inert tags.

Still and colleagues[160] have disclosed an alternate approach to encoding combinatorial syntheses on resin beads wherein electron capture capillary gas chromatography (ECGC) is used in the analysis of the coding moieties. This scheme employs a series of chromatographically resolvable halocarbon derivatives as molecular tags which, when appended to reactive groups on the bead surface, can constitute a binary code that reflects the chemical history of any member of a library. In contrast to the oligonucleotide or peptide-coding approaches where the order of assembly of the chemical building blocks for any library member is preserved in the sequence of a single cognate tagging molecule, the binary strategy uses a uniquely defined mixture of tags to represent each building block at each particular step of the synthesis. Thus a set of N distinguishable tags can be used to encode the combinatorial synthesis of a library containing as many as 2^N different members.

The Still group has synthesized peptide libraries on 50–80-μm-diameter beads and determined the structures of ligands for a monoclonal antibody[160] and synthetic receptor[161] through the use of this coding strategy. The library construction follows the "split synthesis" protocol except that prior to each amino acid coupling, a substoichiometric addition of a tag mixture is made to the beads so that ~0.5% of the growing ligand chain is terminated for each tag added. The tags are coupled to the beads via a photolabile carbonate linker that can be cleaved upon irradiation at 365 nm (see Figure 7). After assembly, the library is screened for interaction with labeled receptor and stained beads are selected for analysis. Photolysis of individual beads in ~1 μL of DMF releases the haloarene tags for rapid analysis by ECGC. This technology seems well suited for analyzing combinatorial libraries comprised of up to ~10^6 members, although the sensitivity of electron capture detection may ultimately enable accurate sequence information to be obtained from beads of even smaller size than those used in this work. The chemically inert nature of the halocarbon tags is a particularly attractive feature of this coding approach, and should facilitate the screening of nonsequenceable organic molecules prepared by multistep combinatorial synthesis.

By contrast, encoding a combinatorial synthesis with oligonucleotides offers the advantage that tremendous levels of tag amplification can be achieved through the polymerase chain reaction. Thus it is possible to work with tiny quantities of DNA template and hence to use solid supports of microscopic dimensions in the syntheses. For example, the 10-μm-diameter beads described above give rise to 5×10^9 beads/g with a maximum ligand loading of ~20 fmol/bead. By using a combination of magnetic-activated and fluorescence-activated bead sorting techniques together with DNA sequence analysis, it should be practical to screen libraries containing as many as 10^8–10^9 synthetic compounds in soluble receptor binding assays. It seems likely that constraints on the sensitivity and/or throughput of other analytical procedures will ultimately restrict the scope of the nonamplifiable coding approaches to analyzing libraries of more limited diversity.

Acknowledgment. We thank Drs. A. Zaffaroni and G. Ringold for their interest, insights, and support in undertaking the application of combinatorial technologies to drug discovery and Mrs. Terri Nebozuk for help in preparing this manuscript.

References

(1) Hirschmann, R. Medicinal Chemistry in the Golden Age of Biology: Lessons from Steroid and Peptide Research. *Angew. Chem.* **1991**, *30*, 1278–1301.
(2) Patchett, A. A. Excursions in Drug Discovery. *J. Med. Chem.* **1993**, *36*, 2051–2058.
(3) Chabala, J. C. Pharmaceutical Manufacturers Association, Drug Discovery Management Subsection, September 19–21, 1993, Philadelphia, PA.
(4) Dower, W. J.; Fodor, S. P. A. The Search for Molecular Diversity (II) Recombinant and Synthetic Randomized Peptide Libraries. *Annu. Rep. Med. Chem.* **1991**, *26*, 271–280.
(5) Moos, W. H.; Green, G. D.; Pavia, M. R. Recent Advances in the Generation of Molecular Diversity. *Annu. Rep. Med. Chem.* **1993**, *28*, 315–324.
(6) Zuckermann, R. N. The Chemical Synthesis of Peptidomimetic Libraries. *Curr. Opin. Struct. Biol.* **1993**, *3*, 580–584.
(7) Pavia, M. R., Sawyer, T. K., Moos, W. H., Eds. The Generation of Molecular Diversity. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 381–475.
(8) Zoller, M. J.; Smith, M. Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any fragment of DNA. *Nucleic Acids Res.* **1982**, *10*, 6487–6500.
(9) Dalbadie-McFarland, G.; Cohen, L. W.; Riggs, A. D.; Morin, C.; Itakura, K.; Richards, J. H. Oligonucleotide-directed mutagenesis as a general and powerful method for studies of protein function. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 6409–6413.
(10) Matteucci, M. D.; Heyneker, H. L. Targeted random mutagenesis: the use of ambiguously synthesized oligonucleotides to mutagenize sequences immediately 5′ of an ATG initiation codon. *Nucleic Acids Res.* **1983**, *11*, 3113–3121.
(11) Wells, J. A.; Vasser, M.; Powers, D. B. Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites. *Gene* **1985**, *34*, 315–323.
(12) Horwitz, M. S. Z.; Loeb, L. A. Promoters Selected from Random DNA Sequences. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 7405–7409.
(13) Reidhaar-Olson, J. F.; Sauer, R. T. Combinatorial Cassette Mutagenesis as a Probe of the Informational Content of Protein Sequences. *Science* **1988**, *241*, 53–57.
(14) Young, R. A.; Davis, R. W. Yeast RNA polymerase II genes: isolation with antibody probes. *Science* **1983**, *222*, 778–782.
(15) Young, R. A.; Davis, R. W. Efficient isolation of genes by using antibody probes. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 1194–1198.

(16) Aruffo, A.; Seed, B. Molecular cloning of a CD28 cDNA by a high-efficiency COS cell expression system. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 8573–8577.
(17) Seed, B.; Aruffo, A. Molecular cloning of the CD2 antigen, the T-cell erythrocyte receptor, by a rapid immunoselection procedure. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 3365–3369.
(18) Wagstrom, C.; Cwirla, S.; Dower, W. J. Unpublished results.
(19) Charbit, A.; Boulain, J. C.; Ryter, A.; Hofnung, M. Probing the topology of a bacterial membrane protein by genetic insertion of a foreign epitope; expression at the cell surface. *EMBO J.* **1986**, *5*, 3029–3037.
(20) Schorr, J.; Kanpp, B.; Hundt, E.; Kupper, H.; Amann, E. Surface expression of malarial antigens in *E. coli* and *S. typhimurium*: induction of seum antibody response upon oral vaccination of mice. *Vaccines 91*; Chanock, R. M., Ginsberg, H. S., Brown, F., Lerner, R. A., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1991; pp 387–392.
(21) Agterberg, M.; Adriaanse, H.; van Bruggen, A.; Karperien, M.; Tommassen, J. Outer-membrane PhoE protein of *E. coli* K-12 as an exposure vector: possibilities and limitations. *Gene* **1990**, *88*, 37–45.
(22) Fuchs, P.; Breitling, F.; Dubel, S.; Seehaus, T.; Little, M. Targeting recombinant antibodies to the surface of *Escherichia coli*: fusion to a peptidoglycan associated lipoprotein. *Bio/Tech.* **1991**, *9*, 1369–1372.
(23) Thiry, G.; Clippe, A.; Scarcez, T.; Petre, J. Cloning of DNA sequences encoding foreign peptides and their expression in the K88 pili. *Appl. Environ. Microbiol.* **1989**, *55*, 984–993.
(24) Kuwajima, G.; Asaka, J.-I.; Fujiwara, T.; Fujiwara, T.; Nakano, K.; Kondoh, E. Presentation of an antigenic determinant from hen egg-white lysozyme on the flagellar filament of *Escherichia coli*. *Bio/Tech.* **1988**, *6*, 1080–1083.
(25) Hansson, M.; Stahl, S.; Nguyen, T. N.; Bachi, T.; Robert, A.; Binz, H.; Sjolander, A.; Uhlen, M. Expression of recombinant proteins on the surface of the coagulase-negative bacterium *Staphylococcus xylosus*. *J. Bacteriol.* **1992**, *174*, 4239–4245.
(26) Klauser, T.; Pohler, J.; Meyer, T. F. Extracellular transport of cholera toxin B subunit using *Neisseria* IgA protease β-domain: conformation-dependent outer membrane translocation. *EMBO J.* **1990**, *9*, 1991–1999.
(27) Brown, S. Engineered iron oxide-adhesion mutants of the *Escherichia coli* phage λ receptor. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 8651–8655.
(28) Smith, G. P. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **1985**, *228*, 1315–1317.
(29) Parmley, S. F.; Smith, G. P. Antibody-selectable filamentous fd phage vectors: affinity purification of target genes. *Gene* **1988**, *73*, 305–318.
(30) Cwirla, S.; Peters, E. A.; Barrett, R. W.; Dower, W. J. Peptides on phage: a vast library of peptides for identifying ligands. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 6378–6382.
(31) Scott, J. K.; Smith, G. P. Searching for peptide ligands with an epitope library. *Science* **1990**, *249*, 386–390.
(32) Devlin, J. J.; Panganiban, L. C.; Devlin, P. E. Random peptide libraries: a source of specific protein binding molecules. *Science* **1990**, *249*, 404–406.
(33) Il'ichev, A. A.; Minenkova, O. O.; Tat'kov, S. I.; Karpyshev, N. N.; Eroshkin, A. M.; Petrenko, V. A.; Sandakhchiev, L. S. Production of a viable variant of the M13 phage with a foreign peptide inserted into the basic coat protein. *Doklady Akademii Nauk. S.S.S.R.* **1989**, *307*, 481–483.
(34) Greenwood, J.; Willis, A. E.; Perham, R. N. Multiple display of foreign peptides on a filamentous bacteriophage. *J. Mol. Biol.* **1991**, *220*, 821–827.
(35) Peters, E. A.; Cwirla, S.; Dower, W. J. Unpublished results.
(36) Felici, F.; Castagnoli, L.; Musacchio, A.; Jappelli, R.; Cesareni, G. Selection of Antibody Ligands from a Large Library of Oligopeptides Expressed on a Multivalent Exposition Vector. *J. Mol. Biol.* **1991**, *222*, 301–310.
(37) Kang, A. S.; Barbas, C. F.; Janda, K. K.; Benkovic, S. J.; Lerner, R. A. Linkage of recognition and replication functions by assembling combinatorial antibody Fab libraries along phage surfaces. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 4363–4366.
(38) McCafferty, J.; Griffiths, A. D.; Winter, G.; Chiswell, D. J. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* **1990**, *348*, 552–554.
(39) Hoogenboom, H. R.; Griffiths, A. D.; Johnson, K. S.; Chiswell, D. J.; Hudson, P.; Winter, G. Multi-subunit proteins on the surface of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains. *Nucleic Acids Res.* **1991**, *19*, 4133–4137.
(40) Marks, J. D.; Hoogenboom, H. R.; Bonnert, T. P.; McCafferty, J.; Griffiths, A. D.; Winter, G. By-Passing Immunization: Human Antibodies from V-Gene Libraries Displayed on Phage. *J. Mol. Biol.* **1991**, *222*, 581.
(41) Barbas, C. F., III; Bain, J. D.; Hoekstra, D. M.; Lerner, R. A. Semisynthetic Combinatorial Antibody Libraries: A Chemical Solution to the Diversity Problem. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 4457–4461.
(42) Hoogenboom, H. R.; Marks, J. D.; Griffiths, A. D.; Winter, G. Building antibodies from their genes. *Immunol. Rev.* **1992**, *130*, 41–68.
(43) Soderlind, E.; Simonsson, A. C.; Borrebaeck, C. A. K. Phage display technology in antibody engineering: Design of phagemid vectors and *in vitro* maturation systems. *Immunol. Rev.* **1992**, *130*, 109–124.
(44) Bass, S.; Green, R.; Wells, J. A. Hormone Phage: An enrichment method for variant proteins with altered binding properties. *Proteins: Struct. Funct. Genet.* **1990**, *8*, 309–314.
(45) Lowman, H. B.; Bass, S. H.; Simpson, N.; Wells, J. A. Selecting high-affinity binding proteins by monovalent phage display. *Biochemistry* **1991**, *30*, 10832–10838.
(46) Markland, W.; Roberts, B. L.; Saxena, M. J.; Guterman, S. K.; Ladner, R. C. Design, construction and function of a multicopy display vector using fusions to the major coat protein of bacteriophage M13. *Gene* **1991**, *109*, 13–19.
(47) Roberts, B. L.; Markland, W.; Ley, A. C.; Kent, R. B.; White, D. W.; Guterman, S. K.; Ladner, R. C. Directed evolution of a protein: selection of potent neutrophil elastase inhibitors displayed on M13 fusion phage. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 2429–2433.
(48) Pannekoek, H.; van Meijer, M.; Schleef, R. R.; Loskutoff, D. J.; Barbas, C. F., III. Functional display of human plasminogen-activator inhibitor 1 (PAI-1) on phages: novel perspectives for structure-function analysis by error-prone DNA synthesis. *Gene* **1993**, *128*, 135–140.
(49) McCafferty, J.; Jackson, R. H.; Chiswell, D. J. Phage-enzymes: expression and affinity chromatography of functional alkaline phosphatase on the surface of bacteriophage. *Prot. Eng.* **1991**, *4*, 955–961.
(50) Corey, D. R.; Shiau, A. K.; Yang, Q.; Janowski, B. A.; Craik, C. S. Trypsin display on the surface of bacteriophage. *Gene* **1993**, *128*, 129–134.
(51) Scott, J. K. Identifying lead peptides from epitope libraries. *Drug Design*; Weiner, D. B., Williams, W. V., Eds.; CRC Press: Boca Raton, FL, 1993; pp 1–61.
(52) Dower, W. J. Phage power. *Curr. Biol.* **1992**, *2*, 251–253.
(53) Stephen, C. W.; Lane, D. P. Mutant Conformation of p53: Precise epitope mapping using a filamentous phage epitope library. *J. Mol. Biol.* **1992**, *225*, 577–583.
(54) Yayon, A.; Aviezer, D.; Safran, M.; Gross, J. L.; Heldman, Y.; Cabilly, S.; Givol, D.; Katchalski-Katzir, E. Isolation of peptides that inhibit binding of basic fibroblast growth factor to its receptor from a random phage-epitope library. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 10643–10647.
(55) Felici, F.; Luzzago, A.; Folgori, A.; Cortese, R. Mimicking of discontinuous epitopes by phage-displayed peptides, II. Selection of clones recognized by a protective monoclonal antibody against the *Bordetella pertussis* toxin from phage peptide libraries. *Gene* **1993**, *128*, 21–27.
(56) Luzzago, A.; Felici, F.; Tramontano, A.; Pessi, A.; Cortese, R. Mimicking of discontinuous epitopes by phage-displayed peptides, I. Epitope mapping of human H ferritin using a phage library of constrained peptides. *Gene* **1993**, *128*, 51–57.
(57) Balass, M.; Heldman, Y.; Cabilly, S.; Givol, D.; Katchalski-Katzir, E.; Fuchs, S. Identification of a hexapeptide that mimics a conformation-dependent binding site of acetylcholine receptor by use of a phage-epitope library. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 10638–10642.
(58) Hoess, R.; Brinkmann, U.; Handel, T.; Pastan, I. Identification of a peptide which binds to the carbohydrate-specific monoclonal antibody B3. *Gene* **1993**, *128*, 43–49.
(59) Kay, B. K.; Adey, N. B.; He, Y.-S.; Manfredi, J. P.; Mataragnon, A. H.; Fowlkes, D. M. An M13 phage library displaying random 38-amino-acid peptides as a source of novel sequences with affinity to selected targets. *Gene* **1993**, *128*, 59–65.
(60) Oldenburg, K. R.; Loganathan, D.; Goldstein, I. J.; Schultz, P. G.; Gallop, M. A. Peptide ligands for a sugar-binding protein isolated from a random peptide library. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5393–5397.
(61) Scott, J. K.; Loganathan, D.; Easley, R. B.; Gong, X.; Goldstein, I. J. A family of concanavalin A-binding peptides from a hexapeptide epitope library. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5398–5402.
(62) Smith, G. P.; Schultz, D. A.; Ladbury, J. E. A Ribonuclease S-Peptide Antagonist Discovered with a Bacteriophage Display Library. *Gene* **1993**, *128*, 37–42.
(63) O'Neil, K. T.; Hoess, R. H.; Jackson, S. A.; Ramachandran, N. S.; Mousa, S. A.; DeGrado, W. F. Identification of novel peptide antagonists for GPIIb/IIIa from a conformationally constrained phage peptide library. *Proteins: Struct. Funct. Genet.* **1992**, *14*, 509–515.
(64) Koivunen, E.; Gay, D. A.; Ruoslahti, E. Selection of peptides binding to the $\alpha_5\beta_1$ integrin from a phage display library. *J. Biol. Chem.* **1993**, *268*, 20205–20210.
(65) Renschler, M. F.; Bhatt, R. R.; Dower, W. J.; Levy, R. Synthetic peptide ligands of the immunoglobulin receptor induce programmed cell death in a human B-cell lymphoma. *Proc. Natl. Acad. Sci. U.S.A.* In press.

(66) Hammer, J.; Valsasnini, P.; Tolba, K.; Bolin, D.; Higelin, J.; Takacs, B.; Sinigaglia, F. Promiscuous and Allele-specific anchors in HLA-DR-binding peptides. *Cell* **1993**, *74*, 197–203.
(67) Blond-Elguindi, S.; Cwirla, S. E.; Dower, W. J.; Lipshutz, R. J.; Sprang, S. R.; Sambrook, J. F.; Gething, M.-J. Affinity panning of a library of peptides displayed on bacteriophages reveals the binding specificity of BIP. *Cell* **1993**, *75*, 717–728.
(68) Matthews, D. J.; Wells, J. A. Substrate phage: selection of protease substrates by monovalent phage display. *Science* **1993**, *260*, 1113–1117.
(69) Smith, M. M.; Shi, L.; Navre, M. Manuscript in preparation.
(70) Cull, M. G.; Miller, J. F.; Schatz, P. J. Screening for receptor ligands using large libraries of peptides linked to the C terminus of the *lac* repressor. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 1865–1869.
(71) Peters, E. A.; Schatz, P. J.; Dower, W. J. Submitted for publication.
(72) Schatz, P. J. Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme: A 13 residue consensus peptide specifies biotinylation in *Escherichia coli*. *Bio/Tech.* **1993**, *11*, 1138–1143.
(73) Schatz, P. J.; Barrett, R. W. Unpublished results.
(74) Wagstrom, C.; Cwirla, S.; Dower, W. J. Unpublished results.
(75) Mattheakis, L. C.; Bhatt, R.; Dower, W. J. An *in vitro* polysome display system for identifying ligands from very large peptide libraries. Submitted.
(76) Barrett, R. W.; Cwirla, S. E.; Ackerman, M. S.; Olson, A. M.; Peters, E. A.; Dower, W. J. Selective enrichment and characterization of high affinity ligands from collections of random peptides on filamentous phage. *Anal. Biochem.* **1992**, *204*, 357–364.
(77) Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. A General Method for Site-Specific Incorporation of Unnatural Amino Acids into Proteins. *Science* **1989**, *244*, 182–188.
(78) McNeil, J. B.; Smith, M. Saccharomyces cerevisiae CYC1 mRNA 5′-end positioning: analysis by in vitro mutagenesis, using synthetic duplexes random mismatch base pairs. *Mol. Cell Biol.* **1985**, *5*, 3545–3551.
(79) Ner, S. S.; Goodin, D. B.; Smith, M. A simple and efficient procedure for generating random point mutations and for codon replacements using mixed oligodeoxynucleotides. *DNA* **1988**, *7*, 127–134.
(80) Arkin, A.; Youvan, D. C. Optimizing nucleotide mixtures to encode specific subsets of amino acids for semi-random mutagenesis. *Bio/Tech.* **1992**, *10*, 297–300.
(81) *Comprehensive Medicinal Chemistry*; Hansch, C., Sammes, P. G., Taylor, J. B., Eds.; Pergamon Press: Oxford, 1990; Vol. 2.
(82) Jung, G.; Beck-Sickinger, A. G. Multiple Peptide Synthesis Methods and Their Applications. *Angew. Chem., Int. Ed. Engl.* **1992**, *31*, 367–383.
(83) Merrifield, R. B. Solid Phase Peptide Synthesis I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **1963**, *85*, 2149–2154.
(84) Geysen, H. M.; Meloen, R. H.; Barteling, S. J. Use of Peptide Synthesis to Probe Viral Antigens for Epitopes to a Resolution of a Single Amino Acid. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 3998–4002.
(85) Valerio, R. M.; Bray, A. M.; Campbell, R. A.; Dipasquale, A.; Margellis, C.; Rodda, S. J.; Geysen, H. M.; Maeji, N. J. Multipin Peptide Synthesis at the Micromole Scale Using 2-Hydroxyethyl Methacrylate Grafted Polyethylene Supports. *Int. J. Pept. Protein Res.* **1993**, *42*, 1–9.
(86) Bray, A. M.; Maeji, N. J.; Geysen, H. M. The Simultaneous Multiple Production of Solution Phase Peptides: Assessment of the Geysen Method of Simultaneous Peptide Synthesis. *Tetrahedron Lett.* **1990**, *31*, 5811–5814.
(87) Valerio, R. M.; Benstead, M.; Bray, A. M.; Campbell, R. A.; Maeji, N. J. Synthesis of Peptide Analogs Using the Multipin Peptide Synthesis Method. *Anal. Biochem.* **1991**, *197*, 168–177.
(88) Bray, A. M.; Maeji, N. J.; Jhingran, A. G.; Valerio, R. M. Gas Phase Cleavage of Peptides from a Solid Support with Ammonia Vapor. Application in Simultaneous Multiple Peptide Synthesis. *Tetrahedron Lett.* **1991**, *32*, 6163–6166.
(89) Geysen, H. M.; Rodda, S. J.; Mason, T. J.; Tribbick, G.; Schoofs, P. G. Strategies for Epitope Analysis Using Peptide Synthesis. *J. Immunol. Methods* **1987**, *102*, 259–274.
(90) Geysen, H. M.; Rodda, S. J.; Mason, T. J. A Priori Delineation of a Peptide Which Mimics a Discontinuous Antigenic Determinant. *Mol. Immunol.* **1986**, *23*, 709–715.
(91) Geysen, H. M.; Barteling, S. J.; Meloen, R. H. Small Peptides Induce Antibodies with a Sequence and Structural Requirement for Binding Antigen Comparable to Antibodies Raised Against the Native Protein. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 178–182.
(92) Middeldorp, J. M.; Meloen, R. H. Epitope-Mapping on the Epstein-Barr Virus Major Capsid Protein Using Systematic Synthesis of Overlapping Peptides. *J. Virol. Methods* **1988**, *21*, 147–159.
(93) Gombert, F.; Blecha, W.; Tahtinen, M.; Ranki, A.; Pfeiffer, S.; Troger, W.; Braun, R.; Muller-Lantzsch, N.; Jung, G.; Rubsamen-Waigmann, H.; Krohn, K. Antigenic Epitopes of NEF Proteins from Different HIV-1 Strains as Recognized by Sera from Patients with Manifest and Latent HIV Infection. *Virology* **1990**, *176*, 458–466.
(94) Maeji, N. J.; Bray, A. M.; Geysen, H. M. Multi-Pan Peptide Synthesis Strategy for T Cell Determinant Analysis. *J. Immunol. Methods* **1990**, *134*, 23–33.
(95) Gammon, G.; Geysen, H. M.; Apple, R. J.; Pickett, E.; Palmer, M.; Ametani, A.; Sercarz, E. E. T Cell Determinant Structure: Cores and Determinant Envelopes in Three Mouse Major Histocompatibility Complex Haplotypes. *J. Exp. Med.* **1991**, *173*, 609–617.
(96) Mutch, D. A.; Rodda, S. J.; Benstead, M.; Valerio, R. M.; Geysen, H. M. Effects of End Groups on the Stimulatory Capacity of Minimal Length T Cell Determinant Peptides. *Pept. Res.* **1991**, *4*, 132–137.
(97) Wang, J.-x.; Bray, A. M.; DiPasquale, A. J.; Maeji, N. J.; Geysen, H. M. Application of the Multipin Peptide Synthesis Technique for Peptide Receptor Binding Studies: Substance P as a Model System. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 447–450.
(98) Wang, J.-x.; DiPasquale, A. J.; Bray, A. M.; Maeji, N. J.; Geysen, H. M. Study of Stereo-Requirements of Substance P Binding to NK1 Receptors Using Analogs with Systematic D-Amino Acid Replacements. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 451–456.
(99) Spellmeyer, D. C.; Brown, S.; Stauber, G. B.; Geysen, H. M.; Valerio, R. Endothelin Receptor Ligands. Replacement Net Approach to SAR Determination of Potent Hexapeptides. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 519–524.
(100) Spellmeyer, D. C.; Brown, S.; Stauber, G. B.; Geysen, H. M.; Valerio, R. Endothelin Receptor Ligands. Multiple D-Amino Acid Replacement Net Approach. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 1253–1256.
(101) Houghten, R. A. General Method for the Rapid Solid-Phase Synthesis of Large Numbers of Peptides: Specificity of Antigen-Antibody Interaction at the Level of Individual Amino Acids. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 5131–5135.
(102) Houghten, R. A.; Bray, M. K.; Degraw, S. T.; Kirby, C. J. Simplified Procedure For Carrying Out Simultaneous Multiple Hydrogen Fluoride Cleavages of Protected Peptide Resins. *Int. J. Pept. Protein Res.* **1986**, *27*, 673–678.
(103) Beck-Sickinger, A. G.; Durr, H.; Jung, G. Semi-Automated T-Bag Peptide Synthesis Using 9-Fluorenylmethoxycarbonyl Strategy and Benzotriazol-1-yl-tetramethyluronium Tetrafluoroborate Activation. *Pept. Res.* **1991**, *4*, 88–94.
(104) Beck-Sickinger, A. G.; Gaida, W.; Schnorrenberg, G.; Lang, R.; Jung, G. Neuropeptide Y: Identification of the Binding Site. *Int. J. Pept. Protein Res.* **1990**, *36*, 522–530.
(105) Beck-Sickinger, A. G.; Jung, G.; Gaida, W.; Koppen, H.; Schnorrenberg, G.; Lang, R. Structure/Activity Relationships of C-Terminal Neuropeptide Y Peptide Segments and Analogs Composed of Sequence 1–4 Linked to 25–36. *Eur. J. Biochem.* **1990**, *194*, 449–456.
(106) Zimmermann, N.; Beck-Sickinger, A. G.; Folkers, G.; Krickl, S.; Muller, I.; Jung, G. Conformational and Epitope Mapping of Herpes-Simplex-Virus Type-1 Thymidine Kinase Using Synthetic Peptide Segments. *Eur. J. Biochem.* **1991**, *200*, 519–528.
(107) Ruggeri, Z. M.; Houghten, R. A.; Russell, S. R.; Zimmerman, T. S. Inhibition of Platelet Function with Synthetic Peptides Designed to be High-Affinity Antagonists of Fibrinogen Binding to Platelets. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 5708–5712.
(108) Tjoeng, F. S.; Towery, D. S.; Bulock, J. W.; Whipple, D. E.; Fok, K. F.; Williams, M. H.; Zupec, M. E.; Adams, S. P. Multiple Peptide Synthesis Using a Single Solid Support (MPS3). *Int. J. Pept. Protein Res.* **1990**, *35*, 141–146.
(109) Rutter, W. J.; Santi, D. V. General Method for Producing and Selecting Peptides With Specific Properties. U.S. Pat. 5,010,175, 1991.
(110) Birkett, A. J.; Soler, D. F.; Wolz, R. L.; Bond, J. S.; Wiseman, J.; Berman, J.; Harris, R. B. Determination of Enzyme Specificity in a Complex Mixture of Peptide Substrates by N-Terminal Sequence Analysis. *Anal. Biochem.* **1991**, *196*, 137–143.
(111) Petithory, J. R.; Masiarz, F. R.; Kirsch, J. F.; Santi, D. V.; Malcolm, B. A. A Rapid Method for Determination of Endoproteinase Substrate Specificity: Specificity of the 3C Proteinase from Hepatitis A Virus. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 11510–11514.
(112) Blankemeyer-Menge, B.; Frank, R. Simultaneous Multiple Synthesis of Protected Peptide Fragments on 'Allyl'-Functionalized Cellulose Disk Supports. *Tetrahedron Lett.* **1988**, *29*, 5871–5874.
(113) Frank, R.; Doring, R. Simultaneous Multiple Peptide Synthesis Under Continuous Flow Conditions on Cellulose Paper Disks as Segmental Solid Supports. *Tetrahedron* **1988**, *44*, 6031–6040.
(114) Eichler, J.; Beyermann, M.; Bienert, M. Application of Cellulose Paper as Support Material in Simultaneous Solid-Phase Peptide Synthesis. *Collect. Czech. Chem. Commun.* **1989**, *54*, 1746–1752.
(115) Frank, R. Strategies and Techniques in Simultaneous Multiple Peptide Synthesis Based on the Segmentation of Membrane Type Supports. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 425–430.
(116) Eichler, J.; Bienert, M.; Stierandova, A.; Lebl, M. Evaluation of Cotton as a Carrier for Solid-Phase Peptide Synthesis. *Pept. Res.* **1991**, *4*, 296–307.
(117) Schmidt, M.; Eichler, J.; Odarjuk, J.; Krause, E.; Beyermann, M.; Bienert, M. Multiple Peptide Synthesis Using Cellulose-Based Carriers: Synthesis of Substance P Diastereomers and their Histamine-Releasing Activity. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 441–446.

(118) Frank, R. Spot Synthesis: An Easy Technique for the Postionally Addressable, Parallel Chemical Synthesis on a Membrane Support. *Tetrahedron* **1992**, *48*, 9217–9232.

(119) Berg, R. H.; Almdal, K.; Batsberg Pedersen, W.; Holm, A.; Tam, J. P.; Merrifield, R. B. Long-Chain Polystyrene-Grafted Polyethylene Film Matrix: A New Support for Solid-Phase Peptide Synthesis. *J. Am. Chem. Soc.* **1989**, *111*, 8024–8026.

(120) Schnorrenberg, G.; Gerhardt, H. Fully Automatic Simultaneous Multiple Peptide Synthesis in Micromolar Scale-Rapid Synthesis of Series of Peptides for Screening in Biological Assays. *Tetrahedron* **1989**, *45*, 7759–7764.

(121) Gausepohl, H.; Boulin, C.; Kraft, M.; Frank, R. W. Automated Multiple Peptide Synthesis. *Pept. Res.* **1992**, *5*, 315–320.

(122) Zuckermann, R. N.; Kerr, J. M.; Siani, M. A.; Banville, S. C. Design, Construction and Application of a Fully Automated Equimolar Peptide Mixture Synthesizer. *Int. J. Pept. Protein Res.* **1992**, *40*, 497–506.

(123) Dryland, A.; Sheppard, R. C. Peptide Synthesis. Part 8. A System for Solid-Phase Synthesis Under Low Pressure Continuous Flow Conditions. *J. Chem. Soc., Perkin Trans. 1* **1986**, 125–137.

(124) Krchnak, V.; Vagner, J.; Mach, O. Multiple Continuous-Flow Solid-Phase Peptide Synthesis. Synthesis of an HIV Antigenic Peptide and its Omission Analogs. *Int. J. Pept. Protein Res.* **1989**, *33*, 209–213.

(125) Meldal, M.; Bisgaard Holm, C.; Bojesen, G.; Havsteen Jakobsen, M.; Holm, A. Multiple Column Peptide Synthesis, Part 2. *Int. J. Pept. Protein Res.* **1993**, *41*, 250–260.

(126) Zuckermann, R. N.; Banville, S. C. Automated Peptide-Resin Deprotection Cleavage by a Robotic Workstation. *Pept. Res.* **1992**, *5*, 169–174.

(127) Neimark, J.; Briand, J.-P. Development of a Fully Automated Multichannel Peptide Synthesizer with Integrated TFA Cleavage Capability. *Pept. Res.* **1993**, *6*, 219–228.

(128) Gras-Masse, H.; Ameisen, J.-C.; Boutillon, C.; Rouaix, F.; Bossus, M.; Deprez, B.; Capron, A.; Tartar, A. Synthetic Vaccines: The Mixotope Strategy. *Peptides: Chemistry, Structure and Biology, Proceedings of the 12th American Peptide Symposium*; Smith, J. A., Rivier, J. E., Eds.; Escom: Leiden, 1992; pp 842–844.

(129) Blake, J.; Litzi-Davis, L. Evaluation of Peptide Libraries: An Iterative Strategy To Analyze the Reactivity of Peptide Mixture with Antibodies. *Bioconj. Chem.* **1992**, *3*, 510–513.

(130) Furka, A.; Sebestyen, F.; Asgedom, M.; Dibo, G. 1988, Abstr. 14th Int. Congr. Biochem., Prague, Czechoslovakia, Vol. 5, p 47. Abstr. 10th Int. Symp. Med. Chem., Budapest, Hungary, p 288.

(131) Furka, A.; Sebestyen, F.; Asgedom, M.; Dibo, G. General Method for Rapid Synthesis of Multicomponent Peptide Mixtures. *Int. J. Pept. Protein Res.* **1991**, *37*, 487–493.

(132) Sebestyen, F.; Dibo, G.; Kovacs, A.; Furka, A. Chemical Synthesis of Peptide Libraries. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 413–418.

(133) Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. A New Type of Synthetic Peptide Library for Identifying Ligand-Binding Activity. *Nature* **1991**, *354*, 82–84.

(134) Houghten, R. A.; Pinilla, C.; Blondelle, S. E.; Appel, J. R.; Dooley, C. T.; Cuervo, J. H. Generation and Use of Synthetic Peptide Combinatorial Libraries for Basic Research and Drug Discovery. *Nature* **1991**, *354*, 84–86.

(135) Geysen, H. M.; Mason, T. J. Screening Chemically Synthesized Peptide Libraries for Biologically Relevant Molecules. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 397–404.

(136) Fodor, S. P. A.; Read, J. L.; Pirrung, M. C.; Stryer, L.; Lu, A. T.; Solas, D. Light-Directed, Spatially Addressable Parallel Chemical Synthesis. *Science* **1991**, *251*, 767.

(137) Pirrung, M. C.; Read, J. L.; Fodor, S. P. A.; Stryer, L. U.S. Pat. 5,143,854 1992.

(138) Jacobs, J. W.; Fodor, S. P. A. Combinatorial Chemistry: Applications of Light Directed Chemical Synthesis. *Trends Biotechnol.* **1994**, *12*, 19–26.

(139) Holmes, C. P.; Adams, C. L.; Fodor, S. P. A.; Yu-Yang, P. New Techniques in Random Screening: The use of Light-Directed Spatially Addressable Chemical Synthesis. In *Perspectives in Medicinal Chemistry*; Testa, B., Kyburz, E., Fuher, W., Giger, R., Eds.; VHCA: Basel, 1992; pp 489–500.

(140) Adams, C. L.; Kochersperger, L. M.; Mortensen, R. B.; Aldwin, L. A.; Holmes, C. P. New Methodologies in the Preparation of Peptide Libraries. In *Techniques in Protein Chemistry V*; Crabb, J. W., Ed.; Academic Press: San Diego, 1994; pp 525–532.

(141) Holmes, C. P.; Fodor, S. P. A. Photolithography and Combinatorial Chemistry in Screening. In *Innovation and Perspectives in Solid Phase Synthesis and Complementary Technologies*; Epton, R., Ed.; I Intercept, UK, 1994, in press.

(142) Houghten, R. A.; Appel, J. R.; Blondelle, S. E.; Cuervo, J. H.; Dooley, C. T.; Pinilla, C. The Use of Synthetic Peptide Combinatorial Libraries for the Identification of Bioactive Peptides. *BioTechniques* **1992**, *13*, 412–421.

(143) Pinilla, C.; Appel, J. R.; Blanc, P.; Houghten, R. A. Rapid Identification of High Affinity Peptide Ligands Using Positional Scanning Synthetic Peptide Combinatorial Libraries. *BioTechniques* **1992**, *13*, 901–905.

(144) Dooley, C. T.; Houghten, R. A. The Use of Positional Scanning Synthetic Peptide Combinatorial Libraries for the Rapid Determination of Opioid Receptor Ligands. *Life Sci.* **1993**, *52*, 1509–1517.

(145) Houghten, R. A.; Dooley, C. T. The Use of Synthetic Peptide Combinatorial Libraries for the Determination of Peptide Ligands in Radio-Receptor Assays: Opioid Peptides. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 405–412.

(146) Dooley, C. T.; Chung, N. N.; Schiller, P. W.; Houghten, R. A. Acetalins: Opioid Receptor Antagonists Determined Through the Use of Synthetic Peptide Combinatorial Libraries. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 10811–10815.

(147) Owens, R. A.; Gesellchen, P. D.; Houchins, B. J.; DiMarchi, R. D. The Rapid Identification of HIV Protease Inhibitors Through the Synthesis and Screening of Defined Peptide Mixtures. *Biochem. Biophys. Res. Commun.* **1991**, *181*, 402–408.

(148) Zuckermann, R. N.; Kerr, J. M.; Siani, M. A.; Banville, S. C.; Santi, D. V. Identification of Highest-Affinity Ligands by Affinity Selection from Equimolar Peptide Mixtures Generated by Robotic Synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 4505–4509.

(149) Kerr, J. M.; Banville, S. C.; Zuckermann, R. N. Identification of Antibody Mimotopes Containing Non-Natural Amino Acids by Recombinant and Synthetic Peptide Library Affinity Selection Methods. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 463–468.

(150) Lam, K. S.; Hruby, V. J.; Lebl, M.; Knapp, R. J.; Kazmierski, W. M.; Hersh, E. M.; Salmon, S. E. The Chemical Synthesis of Large Random Peptide Libraries and Their Use for the Discovery of Ligands for Macromolecular Acceptors. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 419–424.

(151) Kassarjian, A.; Schellenberger, V.; Turck, C. W. Screening of Synthetic Peptide Libraries with Radiolabeled Acceptor Molecules. *Pept. Res.* **1993**, *6*, 129–133.

(152) Lebl, M.; Patek, M.; Kocis, P.; Krchnak, V.; Hruby, V. J.; Salmon, S. E.; Lam, K. S. Multiple Release of Equimolar Amounts of Peptides from a Polymeric Carrier Using Orthogonal Linkage-Cleavage Chemistry. *Int. J. Pept. Protein Res.* **1993**, *41*, 201–203.

(153) Dower, W. J.; Barrett, R. W.; Gallop, M. A. Method of Synthesizing Diverse Collection of Oligomers. PCT Application WO 93/06121.

(154) Needels, M. N.; Jones, D. G.; Tate, E. H.; Heinkel, G. L.; Kochersperger, L. M.; Dower, W. J.; Barrett, R. W.; Gallop, M. A. Generation and Screening of an Oligonucleotide-Encoded Synthetic Peptide Library. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 10700–10704.

(155) Lerner, R.; Janda, K.; Brenner, S.; Nielsen, J. Encoded Combinatorial Chemical Libraries. PCT Application WO 93/20242.

(156) Brenner, S.; Lerner, R. A. Encoded Combinatorial Chemistry. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5181–5183.

(157) Nielsen, J.; Brenner, S.; Janda, K. D. Synthetic Methods for the Implementation of Encoded Combinatorial Chemistry. *J. Am. Chem. Soc.* **1993**, *115*, 9812–9813.

(158) Kerr, J. M.; Banville, S. C.; Zuckermann, R. N. Encoded Combinatorial Peptide Libraries Containing Non-Natural Amino Acids. *J. Am. Chem. Soc.* **1993**, *115*, 2529–2531.

(159) Nikolaiev, V.; Stierandova, A.; Krchnak, V.; Seligmann, B.; Lam, K. S.; Salmon, S. E.; Lebl, M. Peptide-Encoding for Structure Determination of Nonsequenceable Polymers Within Libraries Synthesized and Tested on Solid-Phase Supports. *Pept. Res.* **1993**, *6*, 161–170.

(160) Ohlmeyer, M. H. J.; Swanson, R. N.; Dillard, L. W.; Reader, J. C.; Asouline, G.; Kobayashi, R.; Wigler, M.; Still, W. C. Complex Synthetic Chemical Libraries Indexed with Molecular Tags. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 10922–10926.

(161) Borchardt, A.; Still, W. C. Synthetic Receptor Binding Elucidated with an Encoded Combinatorial Library. *J. Am. Chem. Soc.* **1994**, *116*, 373–374.