Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 10779–10785, November 1994

Review

Tagged versus untagged libraries: Methods for the generation and screening of combinatorial chemical libraries

Kim D. Janda

Departments of Molecular Biology and Chemistry, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, CA 92037

ABSTRACT **Over the past two decades the pharmaceutical industry has been driven by the biological sciences. The discovery and description of the biological mechanisms that underlie disease states accompanied by an unraveling of these mechanisms has provided drug, and more recently biotechnological, companies with a barrage of new therapeutic targets. Paradoxically, as a result of such biological and biochemical advances, new sources of drug leads are in short supply. Considerable effort in trying to create potential drug candidates has led to the parturition of combinatorial chemical libraries. In this review I will examine some of the main technologies for generating and deducing active components from combinatorial libraries that have been segregated into two schools of thought: (*i*) the creation and decoding of combinatorial libraries by so-called tagged methodologies, and (*ii*) the production and deconvolution of chemical libraries by untagged protocols.**

The random screening of natural products from microbial fermentations, plant extracts, and marine sources for possible activity as therapeutic agents has been a rich source of new drug discoveries for years in the pharmaceutical sciences (1). However, with the advent of molecular biology and progress in crystallography and computational chemistry, "rational drug design" has found many advocates. The knowledge of the three-dimensional structure of receptors or enzymes and the manipulation of this information has lead to the development of a number of drug candidates including angiotensin converting enzyme (2, 3) and renin inhibitors (4–6). But even such "rational" success stories, initiated from first principles and based solely on structural information about the macromolecule alone, have not proved as fast or reliable as once claimed. In recent years there has been a renaissance in drug screening with an ingress of new technologies based on combinatorial chemical libraries (for review, see ref. 7). These methods expose many compounds to a target and allow the compounds that bind the target with the highest affinity to be filtered out from a pool of statistical sequences. In its purest form, a combinatorial chemical library can be defined as any ensemble of molecules, whereas the production and ultimately the methods of screening these combinatorial chemical libraries determine the "hit rate"—i.e., the success or failure of these collections of molecules (8).

Untagged Approaches to Generating Combinatorial Chemical Libraries

The two general procedural methodologies in formulating untagged combinatorial libraries are "mixture synthesis" and "portioning-mixing." Mixture synthesis is exactly what it sounds like, in that mixtures of chemical units are coupled to an activated support to produce chemical diversity (9–12). The advantage of such an approach is that combinatorial libraries of vast complexity, in theory, should be accessible. The problem of achieving such diversity is that product dispersal is strongly influenced by the relative kinetic rates of each competing chemical unit being coupled. Deviating from reactions where the kinetic constants for the addition of each individual component are unknown could be disastrous to the product distribution. Portioning-mixing (13–15), or what has come to be known as combinatorial library split synthesis, is a two-step operation based on a divide–couple and recombine procedure. The essentials of this strategy are that a polymeric support is divided into equal portions for coupling to modular units [such units have typically been amino acids; however, in principle, any chemical moiety (say X) could be appended to the support]. The matrices are combined in a single vessel for washing and/or deprotection and then divided again for the next coupling. Repeating this protocol for a total of n cycles can produce a stochastic collection of up to X^n different molecules. More important, this strategy allows for an equal distribution of the coupled chemical units and for uniform couplings to occur.

Various approaches for generating and screening untagged combinatorial chemical libraries have been developed. Each has advantages and disadvantages in its efforts to create chemical diversity. Several of the most prominent will be considered below.

Mimotope Strategy

Geysen and coworkers (11, 16, 17) have presented an amide linkage strategy wherein peptides can be synthesized in a reusable format. The peptides are synthesized on polymer matrices, "pins," positioned within a microtiter 96-well plate. In this approach, natural L-amino acids, as well as unnatural D-amino acids (we will term either enantiomer X), can be incorporated into the library; thus, X^n peptides are conceivable. However, to manage such large numbers with this pin format dipeptide units are initially held constant, whereas the rest of the sequence is formed from randomly incorporated amino acids. To synthesize a hexapeptide library in this regime requires $NNX_{[3]}X_{[4]}NN$, where now $X_{[3]}$ and $X_{[4]}$* are defined amino acids at positions 3 and 4, respectively, and N represents positions where residues are randomly incorporated by using reaction mixtures containing natural or unnatural amino acids. If a primary set is made using an alphabet of 20 amino acids, then the size of this set will be 400 (20 × 20 preparations), each of which is now a mixture of peptides theoretically consisting of 160,000 different peptides. The derivatization level will be at ≈100 nmol per pin or an average of $\approx 4 \times 10^{11}$ copies of each of the individual peptides.

Examining this primary set of 400 peptide mixtures for binding with an antibody or receptor allows identification of the optimum dipeptide sequence for $X_{[3]}X_{[4]}$. A second screen (40 peptide mixtures) $NNX_{[3]}X_{[4]}D_{[5]}N$ and $ND_{[2]}X_{[3]}X_{[4]}NN$, wherein the amino acids $X_{[3]}X_{[4]}$ are fixed and D equates to positions where single amino acids are to be incorporated, allows resolution or absolute identification of additional residues. This cycle of synthesis and screening is reiterated until the entire hexapeptide sequence is optimized for binding to the target of interest.

Originally this mimotope strategy was used to determine discontinuous epitopes with a protein antigen (12). Recently, its scope has been expanded to include the screening of other relevant receptors

*Subscripts in brackets throughout text represent residue positions rather than multiples.

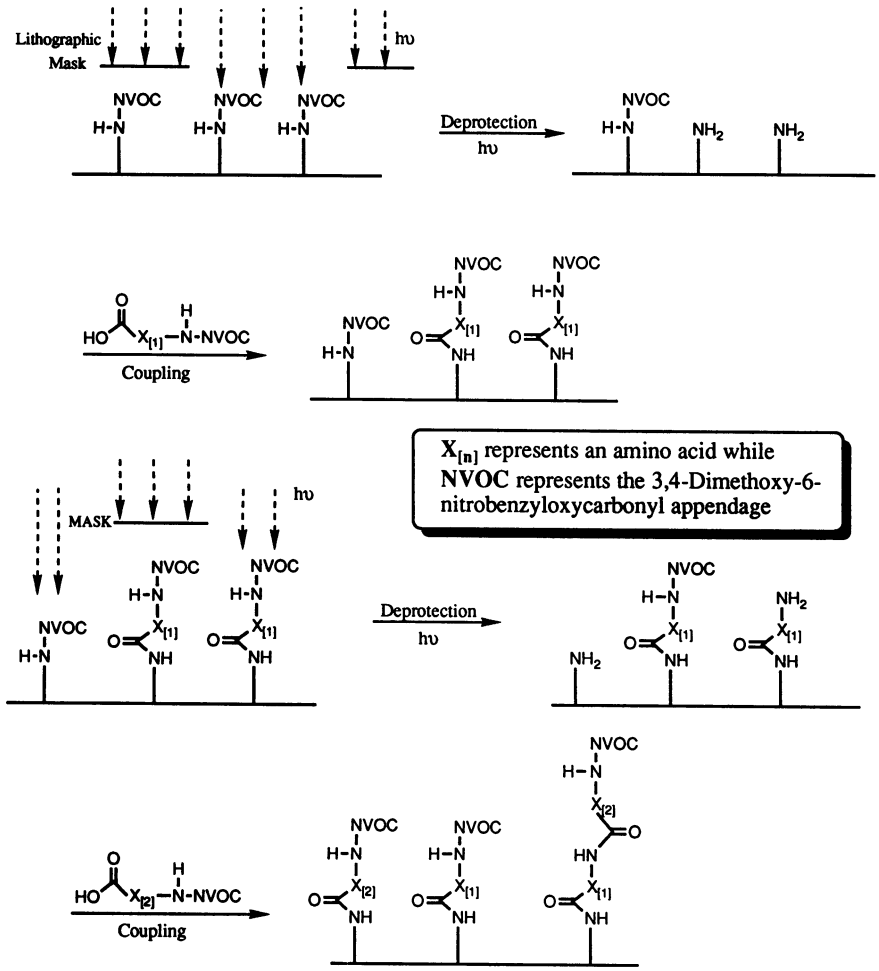

FIG. 1. Scheme of how peptide combinatorial libraries can be constructed using the light-directed, spatially addressable parallel chemical synthetic approach developed by Foder *et al.* (19).

(18). The positive aspects of the mimotope scheme are that modular synthetic strategies are ultimately used in secondary, tertiary, etc. . . . screens. The library diversity can also be increased by applying α-disubstituted and β-amino acids to the process.

Light-Directed, Spatially Addressable Parallel Synthesis of Combinatorial Libraries

Fodor and his colleagues at Affymax (19) have demonstrated a technology that intertwines solid-phase synthesis with photolithography in the preparation of peptide and oligonucleotide combinatorial libraries (Fig. 1). Their general procedure was to derivatize an aliphatic amino-terminated matrix with the light-sensitive amino-protecting group 3,4-dimethoxy-6-nitrobenzyloxycarbonyl (NVOC). Photolysis of the surface removes the protecting group and thereby activates the area for further synthesis. The synthetic scaling process that ensues depends on the photolithographic masking pattern used. Thus, after photolysis (deprotection) the entire surface is exposed to the next NVOC-protected amino acid or oligonucleotide, wherein coupling of either unit occurs only in the regions exposed to light. The procedure is repeated until all the building blocks are coupled to the support. The pattern of masks and the sequence of reactants define the products and their locations. In other words, the identity of the sequence and its location are known.

Screening for activity within these synthesized libraries is done via a fluorescent reporter-group assay, wherein a fluorescently tagged receptor or enzyme is allowed to interact with the library. The fluorescent intensity at each site will depend on the affinity of the receptor for the compound, the concentration of the receptor, and the number and density of the sites resulting in the binding event.

The strengths of this approach are that high-density arrays of chemical compounds can be created in a very small area. A 10-step binary synthesis with amino acids results in the formation of 1024 peptides in 1.6 cm^2. Furthermore, the chemistry is not limited to just peptide and oligonucleotide synthetic strategies. As evidence, the Schultz group (20) has extended the Fodor technology to an oligocarbamate library (Fig. 2).

One-Bead, One-Peptide Solid-Support Technology

A one-bead, one-peptide combinatorial library using split synthesis was pioneered in a collaborative effort by workers at the University of Arizona and the Selectide Corporation (21). Their approach is defined by a noncleavable linker moiety that attaches the peptide unit to the bead. Library screening is accomplished by using receptors conjugated to fluorescein or an enzyme; either entity distinguishes an association event between relevant library members and an acceptor molecule from the rest of the population. Individual library member beads that are stained by this process are removed and analyzed by Edman microsequencing to deduce the sequence of the corresponding peptide ligand.

Using this methodology, the Tucson group has synthesized a 19-amino acid (less cysteine) pentapeptide library (2,476,099 members). The library was screened against a monoclonal antibody known to bind the β-endorphin epitope (YGGFL) with a K_d of 17.5 nM. The

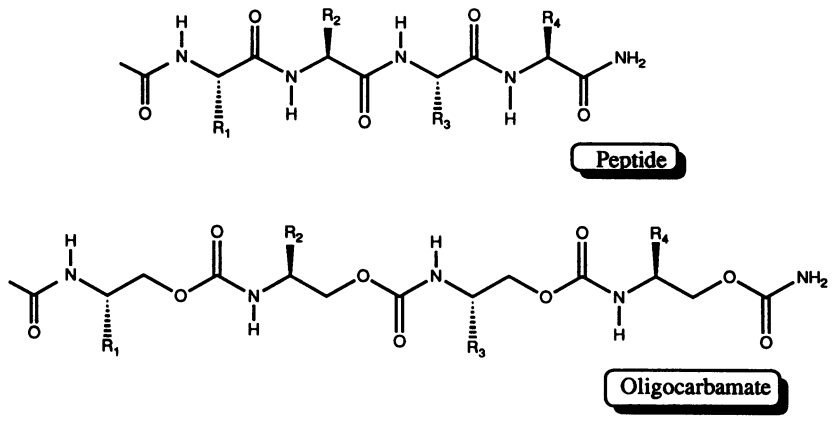

FIG. 2. A peptide and the corresponding oligocarbamate. R represents any standard amino acid side chain.

antibody 3E7 conjugated to alkaline phosphatase retrieved six new pentapeptide-binding sequences with K_d values from 15 to 8780 nM.

The power of this one-bead, one-peptide split synthesis concept lies in the fact that any solid-phase chemistry can theoretically be applied and investigated in a receptor-based and/or enzyme assay. However, this advantage also leads to drawbacks, one of which is that libraries must be screened attached to the support. Subsequent work has been directed at using multicleavable linker appendages to allow for screening of soluble peptide mixtures (22).

Dual-Defined, Positional Scanning, and Robotics Library Technology

Houghten *et al.* (23–25), using what they termed a "dual-defined iterative" methodology, have assembled soluble combinatorial peptide libraries via split synthesis. In the seminal paper (23) a dual-defined hexapeptide library containing 18 amino acids was constructed as follows: Four cycles using solid-phase peptide synthesis and "portioning-mixing" provided 104,976 protected tetrapeptide resin sequences (18^4). This partial library of NNNN-resins (where N is a randomized amino acid position) was divided into 324 aliquots, so that the synthesis of the next two positions $X_{[1]}X_{[2]}$ NNNN-resin could be defined (X is an amino acid position that is defined; the bracketed subscript indicates residue position); upon deprotection and cleavage from these respective resins, a now complete soluble hexapeptide library of >34 million members was obtained. These 324 pools are assayed, and positive results for the first two residues (say $A_{[1]}B_{[2]}$) were noted. Next, 18 new libraries were synthesized with the formula $A_{[1]}B_{[2]}X_{[3]}$NNN, one for each amino acid at position 3, and tested to define $X_{[3]}$. The process is repeated until all positions are defined. Essentially, this methodology is an iterated search process that consists of making the library in a number of segregated pools, finding the active pool that defines the entity for the position on the molecule, and then repeating the process until the active component has been identified. A similar iterative process called SURF (synthetic unrandomization of randomized fragments) was used by ISIS Pharmaceuticals (Carlsbad, CA) for an oligonucleotide library (26).

A virtue of this dual-defined iterative technology is that the multiplicity of components decreases with each step, so that an enrichment process occurs, and because molecules can be assayed in solution, it permits functional, as well as binding, assays. An application of this iterative strategy includes the discovery of antimicrobial peptides with activities against Gram-negative and Gram-positive bacteria (23, 24).

The positional scanning format is again based on soluble combinatorial libraries; however, its proof of concept has been shown to be viable using mixture synthesis (27, 28); now for the same hexapeptide library of 18 amino acids, six different libraries (i.e., 108 positional scanning sublibraries) of the general formula $X_{[1]}N_{[2]}N_{[3]}N_{[4]}N_{[5]}N_{[6]}$, $N_{[1]}X_{[2]}N_{[3]}N_{[4]}N_{[5]}N_{[6]}$, $N_{[1]}N_{[2]}X_{[3]}N_{[4]}N_{[5]}N_{[6]}$, $N_{[1]}N_{[2]}N_{[3]}X_{[4]}N_{[5]}N_{[6]}$, $N_{[1]}N_{[2]}N_{[3]}N_{[4]}X_{[5]}N_{[6]}$, and $N_{[1]}N_{[2]}N_{[3]}N_{[4]}N_{[5]}X_{[6]}$ must be synthesized and assayed. This technology defines the preferred residue at each position of the sequence. The technology also alleviates the unwieldy iterative synthesis and selection steps required in the dual-defined methodology. However, this strategy, unlike the dual-defined methodology, is not endowed with an enrichment process or a progressive improvement of the signal-to-noise ratio.

The Chiron group has developed a fully automated peptide synthesizer that allows combinatorial peptide libraries to be created in the split-synthesis format (29, 30). The instrument consists of an array of reaction vessels, solenoid valves, and a Zymark robot that is computer controlled. The use of this instrument was shown by synthesizing a 361-member decapeptide library. Through competitive ELISA and an affinity-selection methodology decapeptide library members were identified that bound an anti-gp120 monoclonal antibody.

A Recursive Deconvolution Strategy

The final "nontagged" combinatorial library methodology to be described is that of my group (31). The essence of our method is to build and hold a set of partially synthesized combinatorial libraries. A formal example of our method is shown in Fig. 3 for a library of degree 3, made from an alphabet of three components, A, B, and C. In our process we define three channels of synthesis, and each involves only the addition of a single component. Initiating the process requires the making of three pools, in which A, B, and C are added to a matrix adapted with a linker for coupling of the components. A portion of this library is set aside and labeled as partial library p(1). This portion or fraction to be saved and catalogued is the inverse of the degree of the library in the first step. Hence, in the example presented, a degree of three, one-third of each pool is saved and labeled; for subsequent steps the amount saved and catalogued is the inverse of degree minus one. The remaining material is combined and separated into three portions, each channel is loaded, and A, B, and C are attached as before. Again, an aliquot of this library is set aside as partial library p(2), which now is three pools made up of $N_{[1]}A_{[2]}$, $N_{[1]}B_{[2]}$, and $N_{[1]}C_{[2]}$. The remainder is again pooled and split, and the

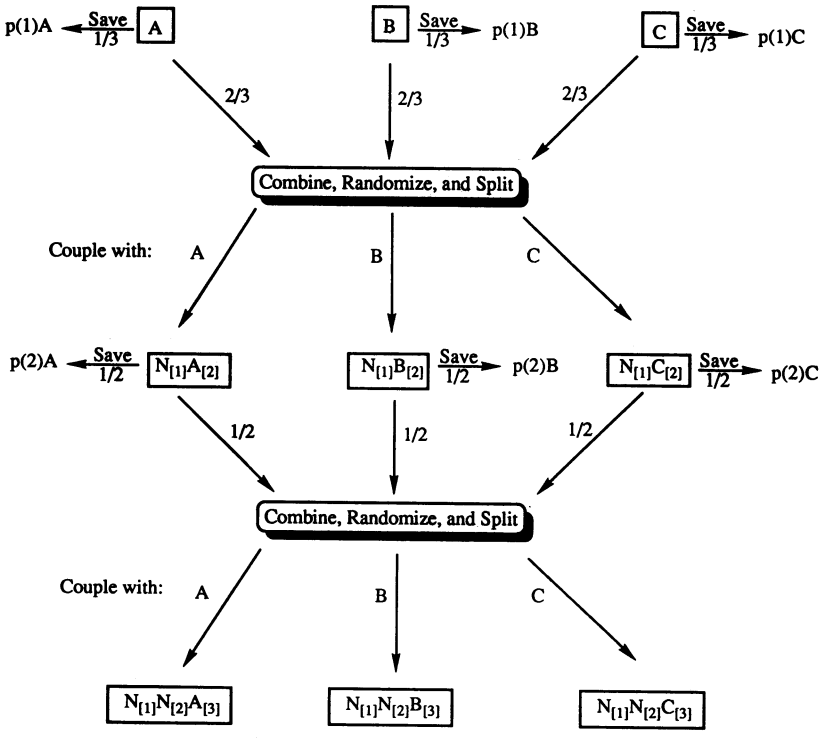

FIG. 3. General scheme of how a chemical combinatorial library using the concept of recursive deconvolution is synthesized and cataloged. Each final pool contains 9 molecules; there is a total of 27 unique molecules in this example. Subscripts denote residue position.

third step of addition is carried out to give the final library $N_{[1]}N_{[2]}A_{[3]}$, $N_{[1]}N_{[2]}B_{[3]}$, and $N_{[1]}N_{[2]}C_{[3]}$.

Screening against a receptor, ligand, or even an enzyme and then determining the most active library member(s) is done simply by examining the final three pools first and proceeding backward to the partial libraries saved. Thus, for the example presented in Fig. 3, we have three pools, $N_{[1]}N_{[2]}X_{[3]}$ (X is A, B, or C), nine compounds in each pool (i.e., a total of 27 different compounds), which are tested by an appropriate assay, and the active pool is determined. Suppose $N_{[1]}N_{[2]}B_{[3]}$ from the final library shown in Fig. 3 is positive. We then go back to library p(2) and add B to an aliquot of each of the three pools, P(2)A, P(2)B, P(2)C, to give three new libraries of the general formula $N_{[1]}X_{[2]}B_{[3]}$. These three new libraries now contain only nine components, so a 3-fold enrichment has been achieved. Again, after testing, suppose pool $N_{[1]}A_{[2]}B_{[3]}$ is active. We proceed to partial library P(1) and add A to each followed by B to give three new pools with the structure $X_{[1]}A_{[2]}B_{[3]}$, which can be tested to find $X_{[1]}$. Again, a 3-fold enrichment has been achieved; the structure is synthesized, and the sequence is deduced.

We have examined this recursive deconvolution in a peptide combinatorial library that was tailored to contain pentapeptide sequences that display binding to the commercially available anti-β-endorphin monoclonal antibody 3E7. In the final analysis, the native epitope YGGFL was found to be the most extensive binder; however, other weaker binders were also deduced through this strategy (31).

There are a number of advantages in using this recursive deconvolution strategy. (*i*) Split synthesis, a rather cumbersome process, need only be done once for each combinatorial library made. In stark contrast is the dual-defined method, which requires numerous split synthetic operations. (*ii*) The deconvolution process recursively defines the synthesis of the active component, so that in the last cycle, the active compound is synthesized. In addition, this methodology allows the deduction of alternative active members, as each deconvolution pathway can be followed either in parallel or successively. (*iii*) Any chemistry is applicable with this technology, which, as we will see, can be problematic in generation of encoded combinatorial libraries.

Tagged Methodologies to Generating Combinatorial Chemical Libraries

Evident from the discussion presented is that there are three critical aspects in any combinatorial library: (*i*) The chemical units that go into the library, (*ii*) the technique for generating the library, and (*iii*) identification of library members that interact with the biological target of interest. Although these three points have been addressed in several of the "untagged" combinatorial library protocols described, limitations do exist. In this next section alternative strategies will be provided, which are termed "tagged" combinatorial libraries.

Phage Technology

Arguably, one of the most powerful tag technologies is that of strictly biological origin. The general concept is one in which a library of peptides is presented on the surface of a bacteriophage such that each phage displays a unique peptide and contains within its genome the corresponding DNA sequence (32–34). In detail, foreign DNA can be inserted into the minor coat protein locus (gene III) of filamentous phage to create fusion phage that express these corresponding peptides at the N terminus of the absorption protein (pIII), which is displayed on the phage surface (Fig. 4). The diversity of displayed peptides is generated by cloning randomly synthesized oligonucleotides that are inserted into a specific region of gene III. These phage libraries encoding peptide units can possess as many as 10^9 unique sequences, which can be screened for binding to any type of immobilized receptor in a selection process known as "panning." This method uses an affinity capture technique to select peptide display phage that bind to the receptor of interest. Selected phages are amplified by infecting *E. coli*, whereas each cycle of panning and amplification enriches certain peptide display phage sequences that bind most tightly to the receptor molecule. After the panning process, DNA from the isolated phage is sequenced, and the peptide responsible for binding is elucidated.

The application of phage technology to the binding of receptors (antibodies) has been demonstrated by several groups (35, 36), including Smith's group (34), who have shown that an epitope hexapeptide library ($\approx 10^7$ members) displayed within the pIII protein could provide individual members that bound to two different antibodies that were specifically made to the DFEKI peptide unit found on the surface of myohemerythrin. In a library designed by Devlin *et al.* (37) phage display epitope libraries with 15, rather than 6, amino acids were made. This approach not only increases the effective size of the library, it also allows the possibility of the display of discontinuous epitopes.

The strengths of phage technology are that large combinatorial libraries of peptide fragments can be generated very quickly and efficiently. Of greater importance is that these peptide libraries are physically linked to their own encoding tag (DNA), which allows for user-friendly amplification, enrichment, and decoding of pertinent binding sequences. At first glance, this methodology seems very appealing; however, a glaring weakness inherent in this technology and other "genetic" library methodologies is that the overall diversity of the chemical units that can be applied within these systems is finite. Exploitation of these libraries for drug discovery may thus be limited.

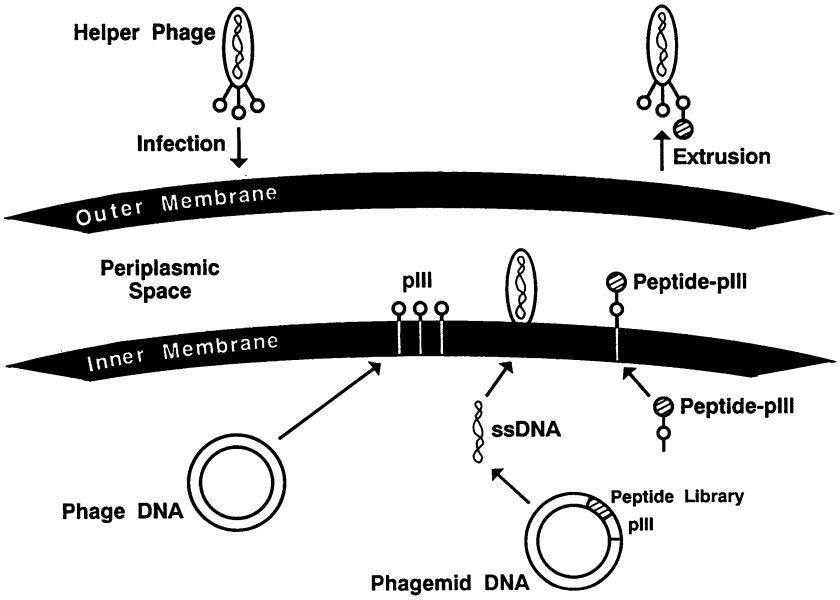

FIG. 4. Scheme illustrating the proposed pathway for peptide display on filamentous bacteriophage. Helper phage infects *Escherichia coli* cells harboring phagemid DNA that contains the genes for the peptide libraries. Helper phage DNA is used to express native pIII, whereas phagemid DNA is used to express peptide–pIII. Single-stranded (ss) phagemid DNA is packaged into phage particles through the aid of helper phage-encoded proteins.

Peptides-on-Plasmids

An alternative to phage technology, which also relies on a biological tack is a genetic combinatorial library approach termed "peptides-on-plasmids." What Schatz and his coworkers at Affymax (38) have done is establish another efficient methodology for a physical connection between a peptide and nucleic acid that encodes for it. In their procedure, a library of peptides is constructed so that the genetic material encoding them is linked through the DNA-binding activity of the lac repressor protein. The random peptides are fused to the C terminus of the repressor by cloning degenerate oligonucleotides at the 3′ of the repressor gene (lac I) present on a plasmid. This plasmid also has lac repressor-binding sites, so the fusions bind the same plasmid that encodes them. Assay for peptide-receptor binding in the Schatz approach is quite similar to the panning process in phage methodology; however, now cell lysis is used to liberate the peptide–lac-plasmid complexes that are screened repeatedly for affinity enrichment.

A proof of principle of this work came from the construction of a random dodecamer library. This library was used to probe for potential peptide members that could bind to IgG D32.39, a specific hybridoma to the dynorphin B epitope. As expected, a consensus sequence was discovered that corresponded to similar binding studies of phage libraries (35). More recently, this peptides-on-plasmid approach has been extended to a more challenging system—the discovery of new substrates for *E. coli* biotin haloenzyme synthetase. The results were that smaller peptide units (13-residue sequences) could be used as substrates (39).

Two features distinguish the peptides-on-plasmid approach from phage technology. (*i*) Random peptides are displayed with a free C terminus. (*ii*) The repressor fusions are cytoplasmic, whereas phage fusions are periplasmic. However, as with phage technology this methodology is restricted in its repertoire of chemical building blocks to oligopeptide units.

Peptide Coded Libraries

In an attempt to address the diversity problem inherent in genetic libraries while still retaining the advantages of a tagging unit, the Chiron Corporation group devised a chemical approach with peptides as the encoding unit (40). The chemical combinatorial strategy uses resin-splitting peptide synthesis to alternately synthesize a "binding" strand and a "coding" strand. An orthogonal protecting-group scheme is also used to allow for the parallel synthesis of both chemical units on the resin shown in Fig. 5.

The isolation of receptor-binding ligands from a tagged library of this type can be done by affinity-selection or bead-staining techniques. The identification of the binding sequence can be determined by Edman sequencing. However, when using such sequencing technology, the binding strand, if of peptide composition, must be made nonsequenceable. In an analogous fashion, the Selectide Corporation group (41) has also described a peptidic coding approach. In their report, protecting-group schemes are used that allow an assay for library receptor binding, either on the matrix or in a soluble format.

The most desirable feature of this type of peptide encoding strategy is that it grants a potential for alternative chemical units other than the natural amino acids (or nucleotides) to be incorporated into the binding strand. However, this methodology does not allow for enrichment by serial selection, the cornerstone upon which genetic library selection methods are founded.

Electrophoric Polyhalobenzene Coded Libraries

Another solution to the chemical coding of combinatorial libraries has been disclosed recently by Still and coworkers (42). In their approach, a combinatorial peptide library attached to beads was assembled using the split synthesis method, while simultaneously being indexed by what they refer to as electrophoric tagging. In their scheme, a series of aromatic electrophores varying in hydrocarbon chain length (Fig. 6) are used as tagging units. Tagging is done by an alternating synthetic process that coincides with the addition of each library binding unit; however, the tagging process here does not require sequential connection. To further simplify the entire scheme, the tags are used in a binary code to record the addition of each building block and, thus, the reaction history of each bead.

Screening of the library is done by a reporter-group assay (*vide infra*), and identified beads with affinity to the receptor are individually picked out by a micropipetter. Use of the *o*-nitrobenzylcarbonate moiety on the linker portion of each coding unit allows release of tag alcohols through photolysis, which can be analyzed by electron capture gas chromatography and identified by the binary synthesis code. Borchardt and Still (43) have advanced this technology to the synthesis of a combinatorial N-acylated tripeptide library using both D- and L-amino acids. This library was used to probe the binding requirements of a synthetic receptor. The outcome of this study was that it allowed new types of host–guest interactions to be observed that might not have been discovered by conventional studies.

A clear advantage of the Still coding technology is that numerous types of chemical processes are amenable to it. Thus, the potential to create highly diversified chemical libraries is well within its purview. The tagging methodology is elegant, as no cosynthesis is required. Again, however, this methodology, like the peptide-tagging technology, does not allow for amplification or enrichment to operate.

Encoded Combinatorial Libraries

If the breadth and versatility of chemical synthesis could be linked to the power of genetics, a potentially more diverse approach to tagged combinatorial libraries could be achieved. A conceptual method to bridge the gap between these two different realms and apply it to combinatorial libraries was advanced by Brenner and Lerner (44) and termed encoded combinatorial chemistry. In their theory, to carry out encoded combinatorial chemistry (Fig. 7) two alternating parallel combinatorial syntheses must be done so that a genetic tag is chemically linked to the chemical structure being synthesized. Thus, addition of a chemical unit to a matrix is followed by addition of an oligonucleotide sequence that codes the chemical unit appended. As with all chem-

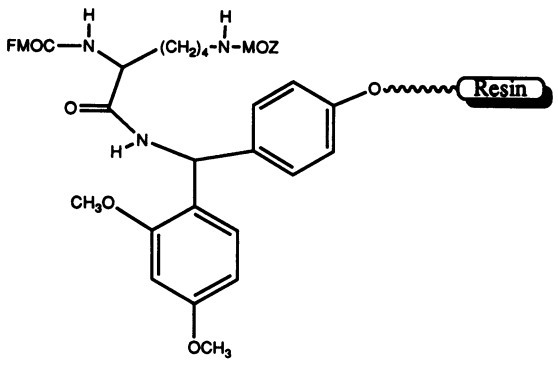

FIG. 5. Resin and bifunctional linker unit containing orthogonal protecting groups to allow for alternating synthesis of the binding and coding strands in peptide coded libraries. FMOC, 9-fluorenylmethyloxycarbonyl functionality; MOZ, 4-methoxybenzyloxycarbonyl.

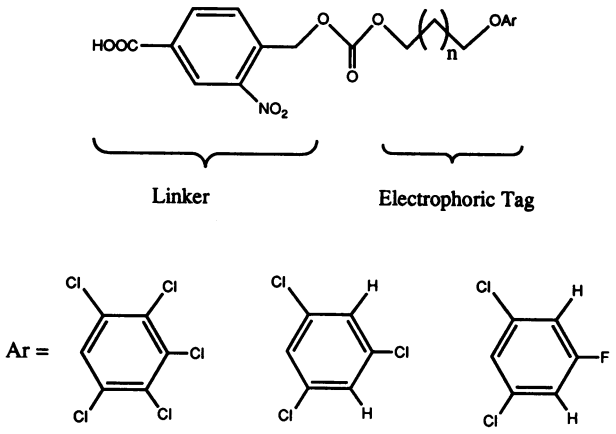

FIG. 6. Tagging units used to create binary synthesis code for electrophoric polyhalobenzene coded libraries.

ical tagging technologies, split synthesis is applied in making the library. However, where this encoded approach differs from other coded synthetic methodologies is in the selection process between library members and receptors. Now active library molecules are affinity-selected to a receptor, and amplified copies of their retrogenetic tags are obtained via PCR. DNA strands that are amplified can be used to enrich for a subset of the library by hybridization with the matching tags, and this type of panning process can be repeated with this subset. Thus, serial enrichment is achieved by a process of purification. Ultimately, the DNA of binding members can be decoded to provide the chemical history and, hence, the structure of the binding unit.

The general principles of this technology are straightforward. The chemical manipulations to create such libraries can be complex. If one just considers synthesizing a chemical library of peptides, new linker technologies, protecting-group schemes, and, moreover, synthetic protocols must be contemplated and devised. At the time when this encoded combinatorial chemical library approach was disclosed, the alternating parallel synthesis of peptides and oligonucleotides had yet to be described. My group recently demonstrated the synthetic methods needed for implementation of this type of mixed chemical synthesis (45, 46). The gist of our tack was to make solid-phase peptide and oligonucleotide synthesis compatible. We accomplished this through the use of 9-fluorenylmethyloxycarbonyl (Fmoc) chemistry for the peptide portion of the molecule, whereas a methyl phosphate scheme was engaged for the oligonucleotide-encoding unit. Following our lead, similar synthetic methodologies applied to beads have also been published by the group at Affymax (47). In their program, an oligonucleotide-encoded D, L peptide library was made on a 10-μm support. Screening the library for binding to an epitope on dynorphin B provided a number of peptide members after amplification and decoding.

The fervor that encoded combinatorial libraries brings to the chemical library field is evident, as it exploits the best of both worlds—chemical diversity and the power of genetics. The main constraints in using this technology come from its strength—namely, the tagging unit. Oligonucleotides can be chemically labile and incompatible with certain synthetic procedures. However, even with these limitations, the elegance of this approach is hard to surpass.

Concluding Remarks and Future Directions

Within a very short time span, a large body of work has accumulated within the field of combinatorial chemical libraries. As with any new scientific endeavor, different paths of research are explored in trying to advance the subject. I have presented what has come to be a philosophical divide within the scientific community on combinatorial libraries: the

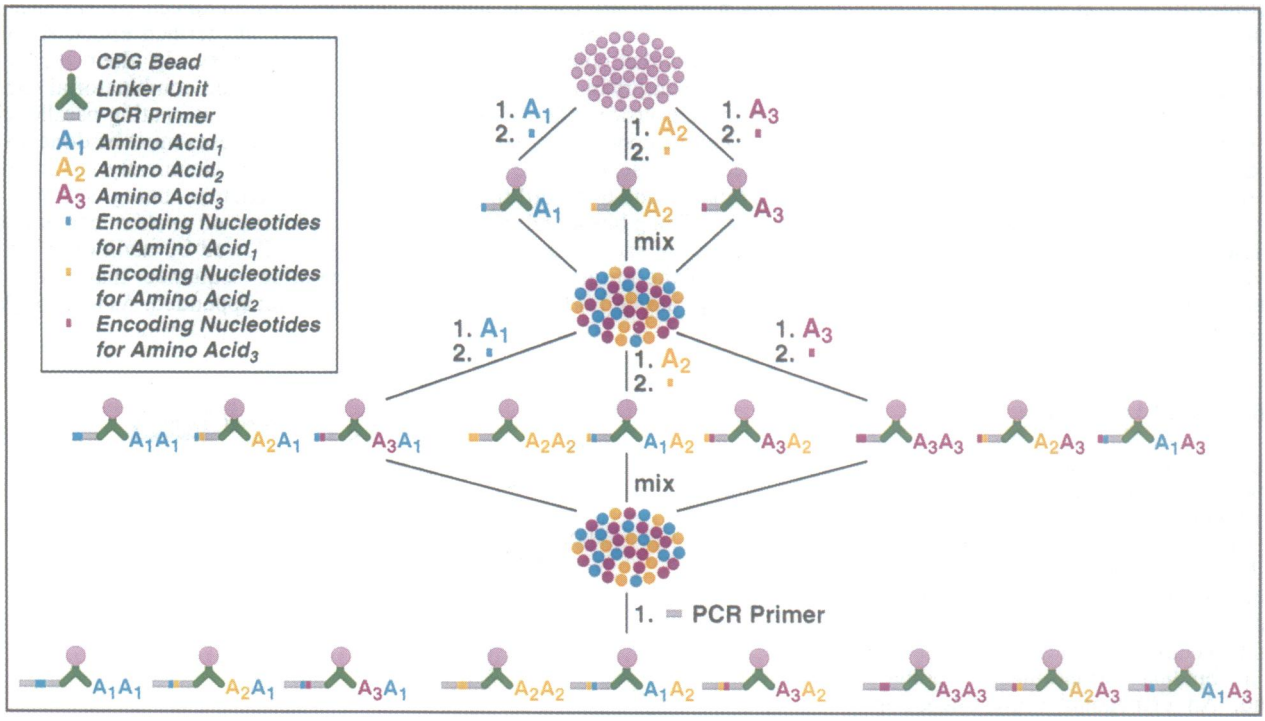

FIG. 7. Scheme that describes the general format used in obtaining encoded combinatorial chemical libraries. CPG, controlled pore glass. Subscripts denote residue position. The figure is reproduced with permission from ref. 45 [copyright (1993) American Chemical Society].

notion of tagging the combinatorial entities or leaving them as untagged components. Clearly, strong arguments can be made in defense of either technology, but ultimately the deciding factor as to what methodology will be engaged depends on user need. Addressable libraries or ones procured by split synthesis for an iterative or recursive deconvolution approach have the potential to create the most diverse combinatorial libraries. In contrast, coded libraries, specifically those of genetic origin, allow for the screening of receptors that are rare and/or of limited concentration.

Currently, most available combinatorial libraries (tagged or untagged) are of either peptide or nucleotide origin. It appears that the next wave of combinatorial research will be directed at the design of libraries that are devoid of the repetitive backbone linkage found within peptides or nucleotides. It is here that structural/stereoelectronic variation and unconstrained motifs will be allowed to expand to unparalleled combinatorial chemical diversity. Already important advances in this area by Bunin and Ellman (48) and Hobbs Dewitt *et al.* (49) have provided us with a glimpse of possible developments.

Finally, it should be noted that combinatorial chemical libraries can provide us with another paradigm for drug discovery and development. Although the field is still young, the methods/technologies for generating and screening these libraries is already quite diverse and is becoming increasingly more sophisticated. The quest for the future is to find whether combinatorial chemical libraries can provide us with more than just "lead" sources for drug discovery. As long as the essence of drug discovery is to find quick and cost-effective new drug candidates, combinatorial chemical libraries provide a unique and ever changing source of chemical diversity.

I thank Peter Wirsching, Eric Erb, Larry Lo, and Richard Lerner for helpful comments on the manuscript. This work was supported in part through a fellowship from the Alfred P. Sloan Foundation and by The Scripps Research Institute.

1. Granellin, C. R. (1992) in *Medicinal Chemistry for the 21st Century*, eds. Wermuth, C. G., Koga, N., Konig, H. & Metcalf, B. W. (Blackwell, London), pp. 3–12.
2. Ondetti, M. D., Rubin, B. & Cushman, D. W. (1977) *Science* **196**, 441–444.
3. Patchett, A. A., Harris, E., Tristam, E. W., Wyvratt, M. J., Wu, D., *et al.* (1980) *Nature (London)* **288**, 280–283.
4. Marciniszyn, J., Hartsuck, J. A. & Tang, J. (1976) *J. Biol. Chem.* **251**, 7088–7094.
5. Marshall, G. R. (1976) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **35**, 2494–2501.
6. Rich, D. H., Sun, E. T. & Singh, J. (1977) *Biochem. Biophys. Res. Commun.* **74**, 762–767.
7. Gallop, M. A., Barrett, R. W., Dower, W. J., Fodor, S. P. A. & Gordon, E. M. (1994) *J. Med. Chem.* **37**, 1233–1251.
8. Hodgson, J. (1993) *Bio/Technology* **11**, 683–688.
9. Rutter, W. J. & Santi, D. V. (1991) U.S. Pat. 5,010,175.
10. Blake, J. & Litzi-Davis, L. (1992) *Bioconjugate Chem.* **3**, 510–513.
11. Geysen, H. M., Rodda, S. J., Mason, T. J., Tribbick, G. & Schoofs, P. G. (1987) *J. Immunol. Methods* **102**, 259–274.
12. Geysen, H. M., Rodda, S. J. & Mason, T. J. (1986) *Mol. Immunol.* **23**, 709–715.
13. Furka, A., Sebestyen, F., Asgedom, M. & Dibo, G. (1988) in *Highlights of Modern Biochemistry: Proceedings of the 14th International Congress of Biochemistry* (VSP, Utrecht, The Netherlands), Vol. 5, p. 47.
14. Furka, A., Sebestyen, F., Asgedum, M. & Dibo, G. (1991) *Int. J. Pept. Protein Res.* **37**, 487–493.
15. Sebestyen, F., Dibo, G., Kovacs, A. & Furka, A. (1993) *Bioorg. Med. Chem. Lett.* **3**, 413–418.
16. Brag, A. M., Maeji, N. J. & Geysen, H. M. (1990) *Tetrahedron Lett.* **31**, 5811–5814.
17. Geysen, H. M. & Mason, T. J. (1993) *Bioorg. Med. Chem. Lett.* **3**, 397–404.
18. Wang, J.-X., Brag, A. M., DiPasquale, A. J., Maeji, N. J. & Geysen, H. M. (1993) *Bioorg. Med. Chem. Lett.* **3**, 447–450.
19. Fodor, S. P. A., Read, J. L., Pirrung, M. C., Stryer, L., Lu, A. T. & Solas, D. (1991) *Science* **251**, 767–773.
20. Cho, C. Y., Moran, E. J., Cherry, S. R., Stephans, J. C., Fodor, S. P. A., Adams, C. L., Sundaram, A., Jacobs, J. W. & Schultz, P. G. (1993) *Science* **261**, 1303–1305.
21. Lam, K. S., Salmon, S. E., Hersh, E. M., Hruby, V. J., Kazmierski, W. M. & Knapp, R. J. (1991) *Nature (London)* **354**, 82–84.
22. Lebl, M., Patek, M., Kocis, P., Krchnak, V., Hruby, V. J., Salman, S. E. & Lam, K. S. (1993) *Int. J. Pept. Protein Res.* **41**, 201–203.
23. Houghten, R. A., Pinilla, C., Blondelle, S. E., Appel, J. R., Dooley, C. T. & Cuervo, J. H. (1991) *Nature (London)* **354**, 84–86.
24. Houghten, R. A., Appel, J. R., Blondelle, S. E., Cuervo, J. H., Dooley, C. T. & Pinilla, C. (1992) *Biotechniques* **13**, 412–421.
25. Houghten, R. A. & Dooley, C. T. (1993) *Bioorg. Med. Chem. Lett.* **3**, 405–412.
26. Ecker, D. J., Vickers, T. A., Hanecak, R., Driver, V. & Anderson, K. (1993) *Nucleic Acids Res.* **21**, 1853–1856.
27. Pinilla, C., Appel, J. R., Blanc, P. & Houghten, R. A. (1992) *Biotechniques* **13**, 901–905.
28. Dooley, C. T. & Houghten, R. A. (1993) *Life Sci.* **52**, 1509–1517.
29. Zuckermann, R. N., Kerr, J. M., Siani, M. A., Banville, S. C. & Santi, D. V. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 4505–4509.
30. Kerr, J. M., Banville, S. C. & Zuckermann, R. N. (1993) *Bioorg. Med. Chem. Lett.* **3**, 463–468.
31. Erb, E., Janda, K. D. & Brenner, S. J. (1994) *Proc. Natl. Acad. Sci. USA* **91**, in press.
32. Smith, G. P. (1985) *Science* **228**, 1315–1317.
33. Parmley, S. F. & Smith, G. P. (1988) *Gene* **73**, 305–318.
34. Scott, J. K. & Smith, G. P. (1990) *Science* **249**, 386–390.
35. Cwirla, S., Peters, E. A., Barrett, R. W. & Dower, W. J. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6378–6382.
36. Scott, J. K. (1993) in *Drug Design*, eds. Weiner, D. B. & Williams, W. V. (CRC, Boca Raton, FL), pp. 1–61.
37. Devlin, J. J., Panganiban, L. C. & Devlin, P. E. (1990) *Science* **249**, 404–406.
38. Cull, M. G., Miller, J. F. & Schatz, P. J. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 1865–1869.
39. Schatz, P. J. (1993) *Bio/Technology* **11**, 1138–1143.
40. Kerr, J. M., Banville, S. C. & Zuckermann, R. N. (1993) *J. Am. Chem. Soc.* **115**, 2529–2531.
41. Nikolaiev, V., Stierandova, A., Krchnak, V., Seligmann, B., Lam, K. E., Salmon, S. E. & Lebl, M. (1993) *Pept. Res.* **6**, 161–170.
42. Ohlemeyer, M. H. J., Swanson, R. N., Dillard, L. W., Reader, J. C., Asouline, G., Kobayashi, R., Wigler, M. & Still, W. C. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 10922–10926.
43. Borchardt, A. & Still, W. C. (1994) *J. Am. Chem. Soc.* **116**, 373–374.
44. Brenner, S. & Lerner, R. A. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 5181–5183.
45. Nielsen, J., Brenner, S. & Janda, K. D. (1993) *J. Am. Chem. Soc.* **115**, 9812–9813.
46. Nielsen, J. & Janda, K. D. (1994) *Methods*, in press.
47. Needels, M. N., Jones, D. G., Tate, E. H., Heinkel, G. L., Kochersperger, L. M., Dower, W. J., Barrett, R. W. & Gallop, M. A. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 10700–10704.
48. Bunin, B. A. & Ellman, J. A. (1992) *J. Am. Chem. Soc.* **114**, 10997–10998.
49. Hobbs DeWitt, S., Kiely, J. S., Stankovic, C. J., Schroeder, M. C., Reynolds Cody, D. M. & Pavia, M. R. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 6909–6913.