Spatially Addressable Combinatorial Libraries

Michael C. Pirrung

Department of Chemistry, P. M. Gross Chemical Laboratory, Duke University, Durham, North Carolina 27708-0346

Received July 31, 1996 (Revised Manuscript Received November 27, 1996)

Contents

I.	Introduction	473
II.	Peptides-on-Pins	474
	A. Method	474
	B. Applications	476
III.	Macroscopic DNA Arrays	478
	A. Method	478
	B. Applications	479
IV.	Light-Directed Synthesis	480
	A. Background	480
	B. Method	480
	C. Applications	483
	1. Peptides	483
	2. DNA	484
V.	Peptides on Paper	485
	A. Method	485
	B. Applications	486
VI.	Inorganic Combinatorial Libraries	486
	A. Method	486
	B. Applications	487
VII.	Conclusions	487
VIII.	References and Notes	487

Michael Pirrung (b. 1955) grew up in the rodeo capital of Texas, Mesquite. After graduating from the University of Texas—Austin, he pursued a Ph.D. at UC-Berkeley. He did research on stereoselective aldol condensations with Clayton Heathcock and an independent project on a photochemical total synthesis of isocomene. After a postdoc with Gilbert Stork at Columbia, he took a position at Stanford. His group became known for work on the mechanism of the biosynthesis of ethylene, the plant-ripening hormone. In 1988, he joined a "working group" of scientists meeting in Palo Alto to form a then-new company, Affymax. He established the chemistry program at Affymax, focusing on methods of creating chemical diversity. His work there on light-directed synthesis was recognized with the AAAS-Newcomb Cleveland Prize. In 1990, he moved to Duke University where he is now Professor of Chemistry and Director of the Program in Biological Chemistry. He has published ~100 peer-reviewed articles and holds several patents. A sabbatical at the Baylor College of Medicine in 1995 resulted in one current interest, chip-based comparison sequencing for mutation detection in highly mutated disease genes. Continuing interests include asymmetric catalysis, synthetic methodology, total synthesis, photochemistry, enzymology, combinatorial chemistry, light-directed synthesis, and ethylene biosynthesis.

I. Introduction

The explosion of interest in combinatorial chemistry has simultaneously created opportunities for neologisms and confusion. In an earlier commentary,[1] the distinction among the various binary, combinatorial, permutational, factorial, and binomial chemical libraries that can be created was discussed. A definition of combinatorial chemistry used in that publication will be restated here. Combinatorial chemistry is in its essence based on the principle of parallelism. Parallel processing enables the chemist to prepare many more substances than the number of chemical steps used to generate the library. This turns on its head the usual principles of chemical synthesis, particularly complex organic molecule synthesis, in which the steps required to reach a single target often number in the teens, or greater. Principles of parallel processing first arose in the preparation of peptides, where the advantages of using a solid-phase support for the preparation of molecular libraries are apparent. An example is Houghten's "tea bag" method, where physically isolated packets of polymer beads are treated batchwise for the common deprotection steps of peptide synthesis and individually for coupling of different amino acids.[2]

This review collects examples of libraries of compounds that have been prepared on monolithic supports so that the *position* of a molecule identifies its composition.[3] This arrangement offers the advantage that each molecule, when evaluated for some desired performance, is substantially "pure" in its local area, provided that the synthesis has proceeded with high efficiency in each stage. An absolute measure of its performance against some standard is therefore possible. Many areas of science still demand absolute, not relative, outcomes in their evaluation of molecular performance. Some other combinatorial chemistry methods, many described in this issue, test pools of compounds, in some cases pools that are so complex it is not possible to establish that each compound imputed to be present in the pool is in fact there, much less what its activity might be. The use of mixtures to discover new pharmacologically active agents was a major impetus for combinatorial chemistry, but such studies can generally only identify the most potent member(s) in a pool, and then only when the property of interest is based on a molecule-with-molecule interaction. For desired properties that are based on the bulk properties of molecules (e.g., liquid

S0009-2665(96)00013-1 CCC: $28.00 © 1997 American Chemical Society

crystals, magnetic materials, lubricants), a means to test each candidate in pure form is required. As a result, while combinatorial pool methods have profound applications, they are still limited because they cannot provide, at the time of testing, direct comparison between different substances. For this reason, the spatially addressed libraries have been applied in the broadest scientific areas, as described in later sections of this review, whereas the pool screening methods have primarily been applied in pharmaceutical discovery, where a specific molecular target relevant to biological activity is available. Combinatorial principles have now grown outside this initial, rather narrow, but deep, application.

Many library preparation methodologies extant today prepare multiple variants of a basic molecular framework as separated, single compounds, often using solid-phase synthesis for its advantages in robotics/automation. As a result, the location of a vial, microtiter well, or test tube and the record of the reagents that have been added to it can be used to identify each molecular structure. Early results in this field were provided by a Parke-Davis group.[4] Those methods are most appropriately called high-throughput or automated synthesis, rather than combinatorial chemistry. The preparation of all possible combinations of a set of reactants does not meet the definition of combinatorial chemistry applied here if the targets are prepared one at a time, regardless of whether robotics are used, how little human intervention is required, or how advantageous the application of solid-phase methods may be.

II. Peptides-on-Pins

A. Method

The common use of parallel processing in immunological methods provided both lessons and hardware for the development of the peptides-on-pins method at Commonwealth Serum Laboratories in Australia and the Central Veterinary Institute in the Netherlands. Geysen continued this work at Coselco Mimotopes, a subsidiary of the former.[5] The plastic 96-well microtiter plate is as commonly used in immunology as is the glass Erlenmeyer flask in chemistry. Its 8 × 12 format permits the simple execution of dilutions and replication of assays such as the ELISA (enzyme-linked immunosorbent assay), and equipment based on this format is available for both rapid solvent addition/removal (pipetting and "plate washing") and optical absorbance measurements. A simple example of the use of a microtiter plate for an ELISA (Figure 1) involves adding to the wells dilutions of a protein of interest, which becomes

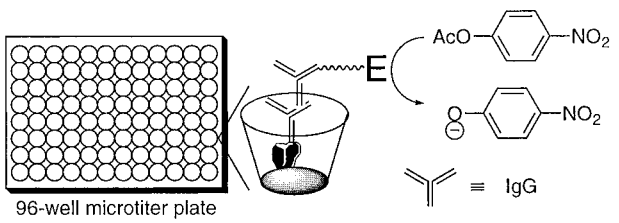


Figure 1. ELISA assay in a microtiter well with *p*-nitrophenylacetate as the chromogenic substrate. E is an esterase that is linked to the secondary antibody.

```
•••ASTWQIKLEFCNPEGLHKRR•••
   ASTWQI
    STWQIK
     TWQIKL
      WQIKLE
       QIKLEF
        IKLEFC
            •
            •
            •
```

Figure 2. Epitope scanning (PEPSCAN) synthesis of hexapeptides based on a protein sequence. One-letter amino acid codes: A, alanine; C, cysteine; D, aspartate; E, glutamate; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine.

weakly bound. An antibody specific for that protein is added, and through antigen−antibody recognition, also becomes bound in the well. A conjugate is obtained between an antibody that binds to a constant region of the first antibody (e.g., a goat anti-mouse antibody) and an enzyme that can produce an absorbing product (such as *p*-nitrophenolate) in the presence of a specific substrate. This enzyme-linked secondary antibody is added and is thereby also bound. Finally, the specific substrate is added and the amount of the absorbing product produced by the enzyme is related to the amount of the original protein present.

The problem that Geysen initially aimed to address by peptide synthesis was epitope mapping.[6,7] An epitope is simply the region of a protein antigen that is recognized by an antibody. Because of aspects of antigen processing that are far beyond the scope of this article, this is often a contiguous sequence within the protein, of 4−10 amino acids. Geysen sought a method to determine this region systematically and directly by simply preparing and testing all of the *n*-mers in the protein sequence. The method was designed so that the length of these *n*-mers would be approximately six and they would overlap by one amino acid. A hexapeptide "window" would therefore be scanned down the protein sequence, looking for a site of reaction with the antibody (Figure 2). The method was named epitope scanning. Since hundreds or thousands of amino acids can be present in the sequence of a given protein, this goal required methods to prepare about this number of peptides in isolated form for individual testing with an antibody. In today's world of combinatorial chemistry, the preparation of thousands of compounds may seem routine, but in the 1980s, there was no precedent for any synthesis of this magnitude.

The parallel processing offered by the microtiter plate was a natural adaptation for the peptide synthesis required for epitope scanning. While solid-phase peptide synthesis in the wells is an approach that might have been taken, Geysen chose to attach the growing peptide chain to solid polyethylene pins bearing functionality appropriate for peptide synthesis. These pins were pressed into holes in a plastic block in the same 8 × 12 arrangement as the microtiter plate (Figure 3). This accommodated not only the requirements of peptide synthesis, where the separation of reagents (in the wells) from synthesized peptides (on the pins) was straightforward, but of the

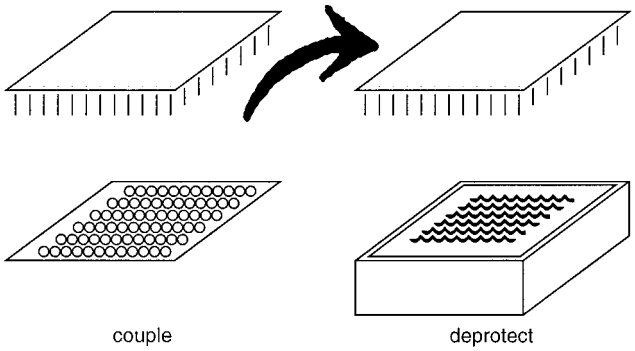


Figure 3. Peptides-on-pins synthesis with parallel coupling steps and batch deprotection steps.

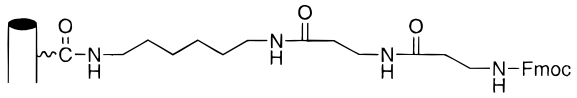


Figure 4. Structure of the linker on the pins.

ELISA, where an antibody could be absorbed to a peptide-bearing pin on the basis of its specific recognition, and then placed into a microtiter well containing a specific substrate to produce a color that would report that recognition.

A number of technical hurdles had to be overcome before these principles could be applied. A significant literature already existed on solid phases for peptide synthesis, and specific requirements for polymer physical state, permeability, and swelling were clear. Geysen chose polyethylene as the pin material, onto which acrylic acid polymers were radiation grafted. These provide a hydrophilic, chemically reactive surface for coupling, through a diamine linker and two β-alanine residues, to provide amino groups ready for N-to-C terminus peptide synthesis (Figure 4). The amino group loading on such pins is 10−100 nmol, resulting in $\sim 10^{11}$ copies of each peptide per pin. The peptide synthesis chemistry that is most successful for this format is [(fluorenylmethyl)oxy-]carbonyl (Fmoc), with PyBOP as the coupling agent.

The execution of a peptide synthesis on such pins involves pipetting into the microtiter wells the different individual protected amino acids that will be added for the next position of the peptides. A block of pins is then treated in batch mode to remove N-terminal protecting groups. The pins are dipped into the wells of reagents for the coupling step. Thus, 96 peptide coupling reactions are being conducted in parallel. This cycle is repeated until peptides of the desired length are prepared. The N-terminal residue is usually capped by acetylation, and the protecting groups on the reactive amino acid side chains are cleaved by a reagent consisting of trifluoroacetic acid and nucleophiles (anisole, ethanedithiol, etc.), as in conventional practice.

Geysen has developed several methods to remove synthesized peptides from pins so that their purity can be assessed and so that solution-phase assays, such as T-cell proliferation, can be used. One involves a Lys−Pro linkage onto the pin, with the peptide synthesis occurring at the ϵ-amino group of lysine (Figure 5).[8,9] After the chain synthesis is complete, removal of the side chain protecting groups also deprotects the Boc group of the lysine, and

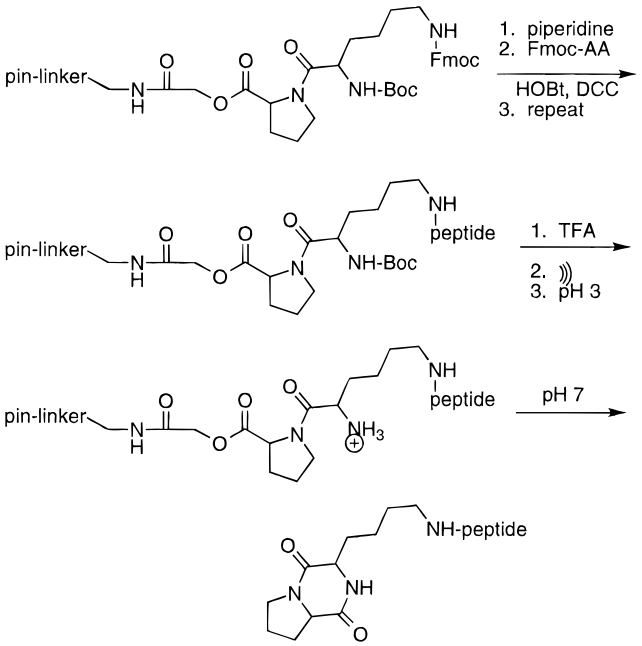


Figure 5. Chemistry of the cleavage of peptides from pins as LysPro-diketopiperazines at neutral pH for biological assay.

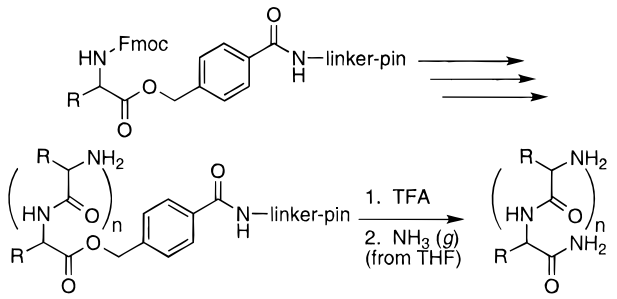


Figure 6. Chemistry of the cleavage of peptides from pins as C-terminal amides.

superfluous reagents are removed by sonication and an acidic wash. After the pin is placed into pH 7 buffer, the neutralized amine cyclizes onto the proline ester to form a diketopiperazine. All peptides bear this diketopiperazine amide at the C-terminus. This linker was used in the synthesis of a known difficult sequence with a Val−Gln coupling. Mass spectrometry, amino acid analysis, and HPLC analysis showed that no deletion peptides were present. Over 80% of the peptides prepared with the pin method and analyzed after cleavage proved to be homogeneous.

The Mimotopes group developed a second method to cleave pin-bound peptides not involving diketopiperazine formation.[10] The linker in this case was an ester, created by loading the first residue as an oxymethylbenzoic acid ester of an Fmoc amino acid (Figure 6). Glycolamide linkers have also been used. After synthesis, the side chain protecting groups were cleaved. Cleavage of the peptides from the pins was accomplished with ammonia vapor, producing the C-terminal amide. The peptide remains physically absorbed to the pin in this protocol, permitting the evaporation of excess ammonia and dissolution of the peptide amide ($\sim$30 nmol) in a buffer compatible with the intended biological screening. This method was also used to show that test sequences produced by the pin method are of good quality (HPLC purity of

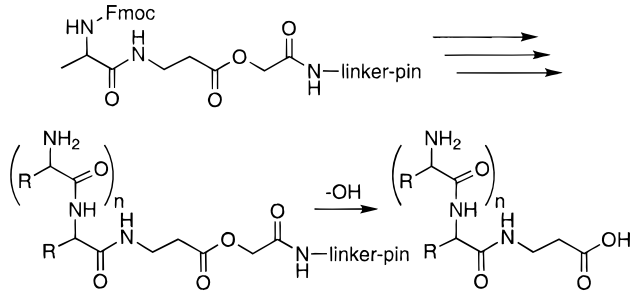


Figure 7. Chemistry of the cleavage of peptides from pins as C-terminal β-alanyl amides.

>85%, correct mass spectra). Continued work of Meloen from the laboratory in the Netherlands where Geysen began his studies has extended this cleavage method to (aminomethyl)polystyrene-grafted pins, permitting quantities up to ~2 mg to be prepared.[11] Screening of ammonia-cleaved peptides against the follicle-stimulating hormone receptor has revealed sequences with known antagonist activity.[12]

Additions to the repertoire of pin synthesis have included fully automated systems to produce peptides on pins.[13,14] Development of superior synthesis protocols has also been reported,[15] with a preferred embodiment being Fmoc amino acids activated by the BOP reagent in dimethylacetamide. Heptapeptide libraries can be prepared in only 3 days using this protocol. The cleavage of peptides with C-terminal β-alanine residues and glycolamide handles (Figure 7), either to the free acids or to the *N*-methyl amides, has also been reported.[16]

For a number of protein sequence/antibody pairs that were subjected to epitope scanning using pins, no high-binding contiguous peptides were discovered. It was presumed that this was because the antibody recognized a "discontinuous" or "assembled" epitope, one composed from amino acid residues that might be far apart in sequence but close to one another in the folded protein. The synthesis of candidate peptides could therefore not be based on sequence, since there would be an astronomical number of combinations of shorter peptides located in distant parts of the protein sequence. Hence, a solution called for an astronomical number of candidate peptides, a number far too large to be prepared one at a time even by the pin method. Geysen's approach to this problem was to prepare pools of peptides on each pin, where some residues were defined and some were mixtures of all genetically coded amino acids. (A seemingly equivalent approach, preparing and assaying short peptides to fix some residues and then extending the sequence to find the next residue outside this "kernel", is not attractive because of the relatively weak binding of antibodies to peptides smaller than the 5−8 residues usually recognized.) Pools based on sequences like XX⊗⊗XX, where X represents an equimolar mixture of all amino acids in the building block set and ⊗ represents defined amino acids, are well suited to pin technology. With 20 genetically coded amino acids, such a library could be held on 400 pins, a manageable number compared to a defined tripeptide library, which requires 8000 elements (84 blocks of pins!). Each of the peptide pools in a XX⊗⊗XX library would have 160 000 members (20 × 20 × 1 × 1 × 20 × 20). After the initial screen, the dipeptide resulting in the most potent signal could be used as the basis for a second-generation set of 400, to fix two more positions in the peptide of optimum activity. A third-generation set would fix the remaining residues. Since molecules discovered by such an approach might have no real relationship to the part of the protein that was immunogenic, but would have the correct size, shape, and charge to mimic the true epitope, they were called mimotopes.[17,18] Note that because this method has resorted to a hybrid of pooling and spatially defined synthesis, one of the main advantages of the latter, the ability to directly compare the activities of compounds in the library, has been lost.

At the height of the application of these pin peptide synthesis methods, over a quarter of a million peptides were being prepared yearly in the CSL labs. Today, the method has been commercialized by Chiron Mimotopes as PepSets made by Multipin technology. It can certainly be said that the principles that came from the peptides-on-pins efforts were extremely helpful to succeeding researchers, but it is also true that the validated applications of the method have been relatively meager. They are primarily in the area of immunology; pin synthesis has not held up to the demands of drug discovery as practiced today.

B. Applications

One of the greatest successes of the epitope scanning method with pins must be considered the discovery of an immunogenic peptide of foot-and-mouth disease virus.[5] An antibody that neutralizes this virus was scanned against the sequence of VP1, a coat protein known to be immunologically important. The sequence GDLQVL was discovered to be the most potent peptide, and replacement of each of the residues in this sequence in turn with all of the other amino acids defined the key residues of the epitope as LQ-L. The GDLQVL peptide was then used to create a peptide vaccine, which induced neutralizing antibodies to foot-and-mouth disease virus in guinea pigs.

Testing of the mimotope strategy, to find the peptide with highest binding affinity to a receptor without knowledge of the sequence that the receptor recognizes, was done with an antibody that binds to a known sequence DFLEKI. The sequence of screens required to (re)discover this linear epitope and the number of peptides synthesized shows the utility of the method (Figure 8).

With the diketopiperazine cleavable peptide technology, Geysen has investigated the binding of substance P and other tachykinins to the substance P receptor.[19] The peptides were cleaved with ammonia vapor and assayed in rat brain synaptosomal membranes, with results generally consistent with litera-

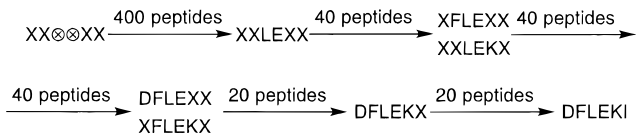


Figure 8. Execution of the mimotope strategy to elucidate an epitope without knowledge of the sequence of the immunogenic protein.

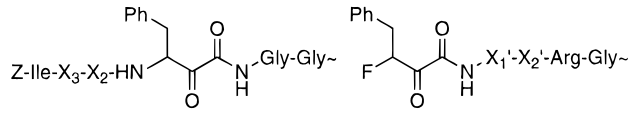


Figure 9. Two peptide libraries bearing serine protease inhibitory α-ketoamide functionalities.

ture structure−activity relationship data. The *in vivo* pharmacological screening of cleaved angiotensin II-related peptides synthesized by the multipin method has also been reported.[20] Synthesis of the native angiotensin sequence (DRVYIHPF) bearing the β-alanyl-amide-diketopiperazine C-terminus, with replacement at each position by the other 19 amino acids, produced a novel sequence DRVYKHPF with only slightly reduced agonist activity. Dose−response studies of more potent agonist peptides with acid C-termini could be conducted, with the discovery that Ser-2, Leu-3, and Thr-3 substitutions retain full agonist activity.

The mapping of determinants on T-cells has also been an excellent application of the cleavable peptide methodology.[21] A region of tetanus toxin containing a known epitope for a particular human T-cell clone was synthesized in an epitope scanning format as 7-mer, 8-mer, 9-mer, and 10-mer peptides. They were cleaved using the diketopiperazine method and examined for cell proliferation effects. The sequence YSYFPSVI was found, in agreement with earlier results for this clone.

Because of its relative simplicity, requiring little novel equipment and simple labware, the pin method has been applied in many laboratories. Many of these reports have entailed straightforward uses of Geysen's method of epitope mapping,[22] but others have gone farther afield.

Abeles has prepared on pins libraries of peptide analog enzyme inhibitors and screened them against chymotrypsin, the digestive enzyme, and recombinant human heart chymase, an enzyme that converts angiotensin I to angiotensin II.[23] Potency against the latter was a prime consideration, with selectivity over the former also important. Inhibitory residues used in these peptides were phenylalanine keto amides and 3-(fluorobenzyl)pyruvamides, which form tetrahedral intermediates like the tetrahedral intermediate in the acyl transfer reaction of serine proteases. The diversity in the library was created by variation of natural residues in the S and S′ substrate recognition sites (Figure 9). These libraries were evaluated by immersing the pins in a solution of target enzyme in microtiter wells and evaluating by chromogenic enzyme assay the amount of enzyme remaining in each well after the pins were removed. A comparison of the potencies of synthetic peptides in solution and the results of this depletion assay supported its validity. In the first library of 324 peptides (18 × 18), the Glu-Pro sequence was identified as superior, and the corresponding synthetic peptide Z-Ile-Glu-Pro-Phe-CO_2Me was shown to have a 1 nM K_i against chymase and a 10 nM K_i against chymotrypsin. In the second library of 400 peptides, Glu-Asp was identified as superior, and the synthetic peptide (F)-Phe-CO-Glu-Asp-Arg-OMe was shown to have a 1 μM K_i against chymase and a 100 μM K_i against chymotrypsin. These results were combined in a hybrid

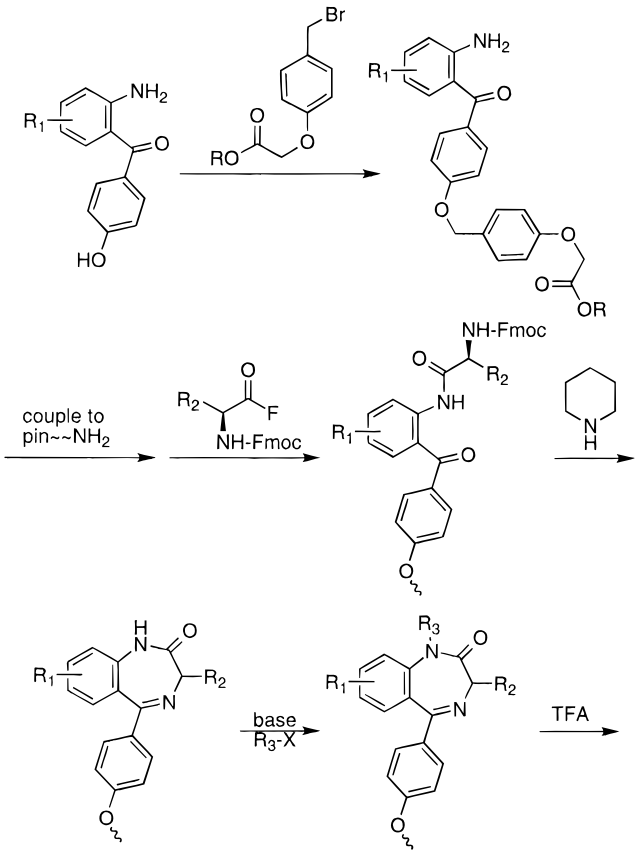


Figure 10. Chemistry of the solid-phase synthesis of benzodiazepines on polyethylene pins.

molecule of the sequence Z-Ile-Glu-Pro-Phe-CO-Glu-Asp-Arg-OMe, which is selective for chymase by a factor of 400!

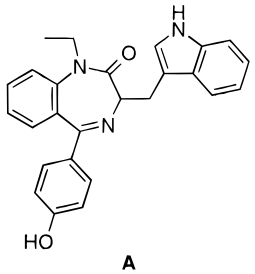
Inhibitors of angiotensin-converting enzyme based on the bradykinin sequence have also been developed with pin-based peptide synthesis.[24] The pins were modified with the acid-labile linker (hydroxymethyl)phenoxyacetic acid to permit cleavage of the peptide from the pin at the same time as protecting groups are removed. The optimum sequence is RAGFAPFR, or the bradykinin sequence with alanine substituted at the 2 and 6 positions and proline 3 deleted.

Localization of a receptor−recognition domain on the S3 cell adhesion subunit of pertussis toxin has been accomplished by peptide mapping against the glycoprotein fetuin.[25] The sequence RQITPGWSIY was shown to bind to fetuin.

Cell adhesion molecules that act through dimerization have also been studied using peptides on pins. A KYSFNY sequence involved in the dimerization of neural cell adhesion molecules (NCAMs) has been identified by epitope scanning with antibodies that inhibit cell adhesion.[26] Peptides containing this sequence inhibit NCAM-mediated binding interactions, while a growth substratum bearing them promotes cell attachment. Epitope scanning, using the sequence of the gp80 molecule and an antibody that both recognizes it and inhibits cell adhesion in *Dictyostelium*, has uncovered the YKLNV sequence that is thought to be involved in dimerization of gp80.[27]

Ellman has prepared 1,4-benzodiazepines on pins.[28] A library of 192 (two 96-well plates) structurally

diverse compounds was synthesized by a general route and tested for activity against the cholecystokinin A receptor (Figure 10). A phenolic 2-aminobenzophenone, one diversity input, was attached to an acid-cleavable linker, protected as an Fmoc derivative, and loaded onto amine-derivatized pins. The Fmoc was removed and Carpino's Fmoc amino acid fluoride coupling method was used to add a second diversity input. On removal of this Fmoc, cyclization to the benzodiazepine occurred. Finally, N-alkylation added the final diversity input. The ~100 nmol of product could be cleaved from the pin with TFA and analyzed by HPLC and FAB MS to verify that the synthetic route had produced the target. Yields were determined for a small sample of the compounds in the library by internal standard HPLC, and the racemization of two benzodiazepines was determined to be negligible by chiral HPLC. The screening of these compounds against the CCK-A receptor gave results consistent with previous structure−activity relationship data, and several compounds were selected for preparation on larger scale for direct comparison with known CCK-A ligands. Compound **A** provided an 80 nM IC_{50}, comparable to a lead compound from a pharmaceutical company that provided a structural basis for the library.



A

III. Macroscopic DNA Arrays

A. Method

Interest in analyzing and comparing nucleic acid sequences by hybridization (the formation of a double-stranded molecule by Watson−Crick base pairing between a known analyzing nucleic acid and an unknown nucleic acid) prompted Southern and Maskos to develop a method to prepare arrays of oligodeoxynucleotides on glass plates.[29,30] The approach initially involved an array of reaction "channels" that pressure seal onto the face of a flat glass synthesis surface that is derivatized through silanol chemistry to present a hydroxyl group. Each channel is formed from lines of 1.2 mm silicon rubber tubing, glued onto a glass plate at a spacing of 1−10 mm and acting as gaskets. Up to 64 channels per mask have been used. Conventional reagents for phosphoramidite or (preferably) H-phosphonate synthesis are passed down each channel by capillary action to synthesize DNA in stripes. The parallelism in this combinatorial synthesis is reflected both in the fact that the coupling chemistry of different nucleotide monomers happens simultaneously in adjacent stripes (literally, in *parallel* regions) and that each stripe will eventually constitute multiple square hybridization sites. This physical masking approach allows all sites on the surface to be extended in a single chemical step,

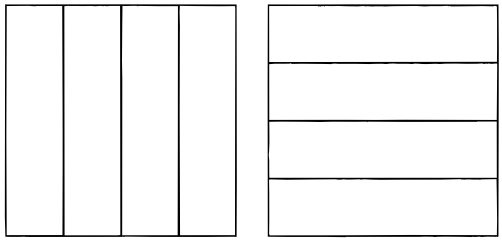


Figure 11. Two orientations of stripes in physical masks permit combinatorial synthesis of oligonucleotides.

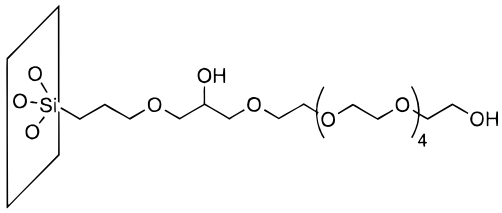


Figure 12. Linker chemistry for oligonucleotide synthesis on glass plates.

leading to the preparation of 4^l sequences in l steps, where l is the length of the oligodeoxynucleotides prepared. A synthesis cycle involves placing reagents in each channel for coupling, removal of the mask plate, treatment of the whole synthesis plate in batch mode with dichloroacetic acid to deprotect the 5′-protecting group, and repositioning of the same or a different mask plate for another coupling. This format permits two orientations of the DNA stripes (Figure 11). After the sequences are assembled, the oxidation of the H-phosphonates to phosphodiesters is accomplished in batch mode, as is the ammonia deprotection of base N-protecting groups. The novel linking chemistry (Figure 12) that was developed is stable to the conditions required to remove internucleotide phosphate and base N-protecting groups, so the DNA remains attached to the surface.[31] These arrays show good stability through up to 50 hybridization and wash cycles.

A second-generation array of support-bound DNA was also developed by Southern.[32] Rather than stripes, a region for chemical synthesis corresponding to a geometric object is defined by the physical mask. After a chemical coupling cycle, the mask is translated linearly so that it partially overlaps the area coupled in the preceding step and partially overlaps virgin territory. This coupling cycle, which requires ~10 min, can be repeated endlessly to prepare arrays that include all contiguous n-mers of a given nucleotide sequence. In that sense, it is somewhat like epitope scanning. The number of steps in the synthesis is simply the number of nucleotides in the sequence. The principle is demonstrated with a diamond-shaped mask (Figure 13), but it has so far proved impractical to actually fabricate arrays this way so that the resulting arrays have square cells. Circular masks lead to "scanning arrays" in which each oligonucleotide is found in a region of lenticular shape. Diamond-shaped arrays could be easily prepared using light-directed synthesis (*vide infra*), however.

Southern has summarized his work in this whole area.[33,34]

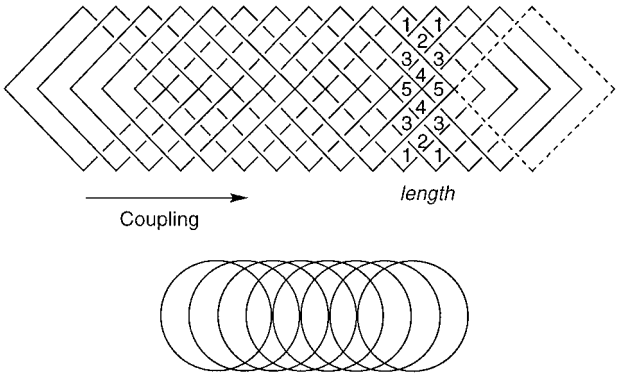


Figure 13. Scanning oligonucleotide arrays prepared (1) formally in an overlapping diamond pattern and (2) actually in an overlapping circular pattern.

B. Applications

DNA arrays have been used to study the factors influencing duplex formation with oligopurine sequences of length 8−12 on surfaces.[29] Room temperature hybridization was complete in 30−120 min. A number of additives have been investigated to improve the fidelity of hybridization of a pool of octapyrimidine probes to a grid of the 256 octapurines. The dependence of duplex formation on base composition can be reduced by a high concentration of tetramethylammonium chloride, and the duplex yield can be greatly increased.[35] The base pairing properties of deoxyinosine-substituted probes have also been investigated.[36] When placed at adjacent positions, the II dinucleotide sequence associates with bases on the complementary strand in the order CC > CA > AA > AC > GC > GA > CG > TA > TC > CT = AG > AT > GT > TT.

DNA arrays have been used for identifying alleles of polymorphic genes, such as the A, C, and S alleles of the β-globin sickle cell locus.[37] The target nucleic acid for this experiment was RNA, prepared by PCR from patient samples with primers bearing RNA polymerase promoters, followed by runoff transcription with radiolabeled UTP. By preparing probes in stripes in one dimension and applying the target in stripes in the other dimension, the grid position identifies the genotype. Hybridization gave data represented in Figure 14. DNA arrays have also been used for empirical analysis of the hybridization behavior of over 70 wild-type and mutant sites within the β-globin gene.[38]

The application of scanning arrays to the detection of mutations in the CFTR (cystic fibrosis transmembrane regulator) gene was problematic,[32] since the 528 bp RNA transcript used as a target contains some secondary structure that impedes hybrid formation. This study of the influence of dangling ends on duplex yields in decanucleotide arrays has supported previous suggestions that a 5′ overhang has a greater stabilizing effect on duplex formation than a 3′ overhang.[39]

It has been suggested that hybridization of an unknown DNA to a complete set of oligonucleotides of a given length could be used to determine its sequence, on the basis of the rules of Watson−Crick complementarity (Figure 15).[40] Octanucleotides are expected to be useful for determining sequence of up

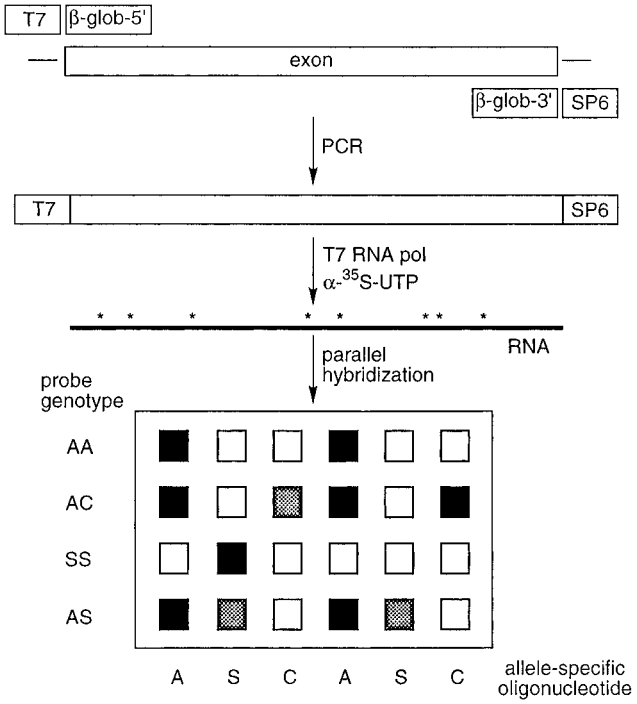


Figure 14. A method for the analysis of polymorphic genes based on oligonucleotide arrays. The region of interest in a sample of analyte DNA is subjected to polymerase chain reaction with primers bearing RNA polymerase promoters at the 5′ ends. The double-stranded product is then subjected to runoff transcription with labeled oligonucleotides to make a single-stranded RNA target. Different targets are placed in horizontal stripes on hybridization arrays of allele-specific oligonucleotides synthesized in duplicate in 2-mm-wide vertical stripes. In this case, the alleles are the wild-type β-globin sequence (A), the sickle-cell β-globin sequence (S), and another polymorphic sequence (C). The genotypes detected in this experiment include homozygotes (AA, SS) and heterozygote carriers of the C allele and the sickle-cell trait (AC, AS).

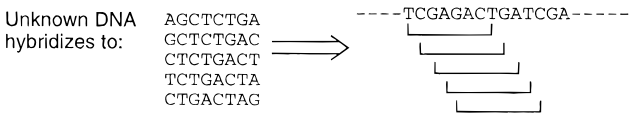


Figure 15. The theory of sequencing-by-hybridization (SBH), based on reading DNA by overlapping words instead of by individual letters.

to 200 bases, and decanucleotides may analyze up to a kilobase. To test this idea, Southern prepared an array displaying 256 octapurines in 3 mm × 3 mm sites using mask plates in a pattern as shown in Figure 16. The addition of a ^{32}P-labeled, mixed octapyrimidine probe (all T and C sequences, equimolar at each position) at 4 °C (necessary because of the relatively weak binding of such short duplexes) produced spots (detected by a storage phosphor imager), although of varying intensities, at each probe site. This background image was obtained as a control to allow compensation for variations in the synthesis yield, base composition, and base sequence that affect the yield of duplex (= signal) in each cell. This array was then probed with two 24-mers (eicosapyrimidines flanked by two AA sequences, for convenience in synthesis and labeling). *Ab initio*, one expects 13 hybridization spots from the 13 overlapping octapurines, but more are observed, and indeed a gradation exists between obvious positive and

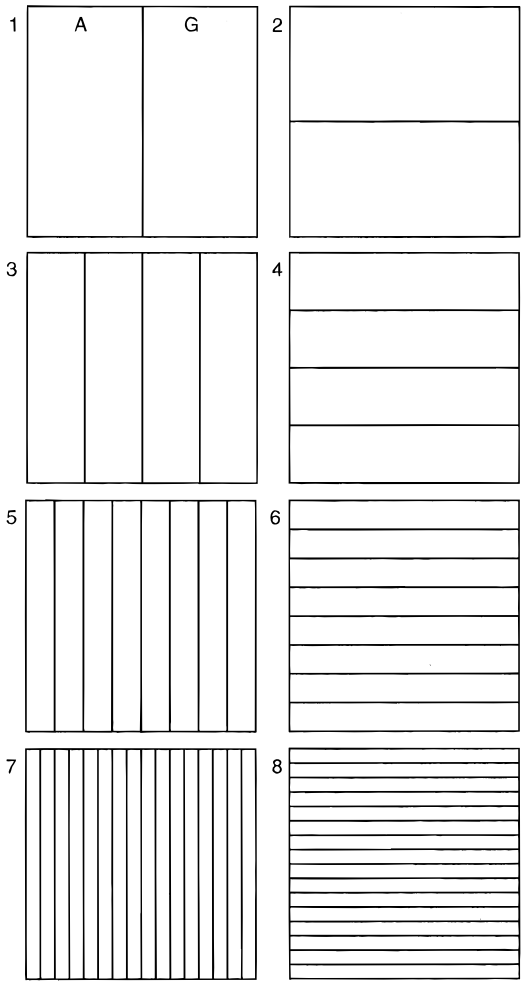


Figure 16. Striped physical masks that can be used to prepare an oligonucleotide array of all 256 octapurines.

negative spots. It was thus not possible to determine the sequence by inspection, and an informatics approach was required. A least-squares method comparing ideal and actual hybridization images was developed that provides a statistical estimate of the goodness of fit. The best fits in the octapyrimidine experiments, although correct, had *R* factors of 0.254 and 0.388, and the second-ranked solutions had *R* factors only slightly greater. Interestingly, all of the "mistake" sequences are single-base mismatches at the ends of the probes.

This experiment also provided a model for mutation detection, since the two 24-mers differed in only one base. Because a single base change affects the sequences of the eight octamers including that position, this site is highly overdetermined and redundantly analyzed. Overlaying and subtracting images from hybridization of the two different probes emphasizes the spots that differ. The least-squares analysis was again applied to and correctly identified the mutations with an *R* factor of 0.34; the second-ranked solution had an *R* factor of 0.39.

IV. Light-Directed Synthesis

A. Background

Early thinking concerning the founding of the Affymax company, which was to be aimed at the specific molecular interaction between drug and receptor, benefited greatly from the views of molecular pharmacologist and founding scientist Avram Goldstein. He emphasized the power of natural diversity in developing potent binding interactions for specific molecules, as exemplified by the mammalian immune system which, with its primary antibody repertoire of $\sim 10^8$, can develop a binding agent for essentially any antigen. The complementary idea, that a library containing all possible collections of charge, hydrophobic/hydrophilic groups, hydrogen bonding atoms, and shape in space would be able to fit any receptor, provided both a technological challenge to prepare such a collection and fundamental scientific interest to test the concept. This principle has almost become accepted wisdom, given the ubiquity of combinatorial concepts in, particularly, drug discovery today. It is likely, though, that it has never been truly tested. The failures of the available technologies to provide ligands for particular targets are not widely discussed, and none has really approached the ideal of providing a library of "all" molecules. Even at that time, however, the peptides-on-pins method as originated by Geysen had exposed the value of the ability to prepare many compounds in evaluating molecular interactions in biological systems. Questions existed, though, concerning the applicability of these libraries for drug discovery, particularly regarding their small size relative to the diversity of the immune system. In discussions with Affymax scientific advisors, we identified a need for what was only half-seriously called the "96 000-well plate" so that complete libraries of pentapeptides (3.2 million) could be prepared in pure form for screening, instead of in pools, as in the Geysen method. In an Affymax science meeting in mid-January, 1989, Leighton Read and I first described the "peptides-on-chips" concept based on the idea that, if arrays of peptides could be made the way computer chips are made, we could benefit from principles of miniaturization, automation, and parallelism that would make such large libraries reachable. The method came to be called VLSIPS, for very large-scale immobilized polymer synthesis, which was meant to make the connection between the concept and very large-scale integrated (VLSI) circuit technology of nearby Silicon Valley.

B. Method

This method for the preparation of arrays of molecules draws on semiconductor photolithographic techniques and permits the synthesis of thousands of different sequences of oligomers at small locations on a surface (Figure 17).[41] The pattern of illumination through a mask, using light to remove photolabile protecting groups from selected areas, determines which regions of the surface are activated for chemical coupling. Then, the first of a set of building blocks (each bearing a photolabile protecting group) is exposed to the entire surface, but it reacts only with regions that were addressed by light in the preceding step. The substrate is illuminated through another mask for deprotection and coupling with a second building block. The steps of masking/irradiation and coupling are repeated to build up various

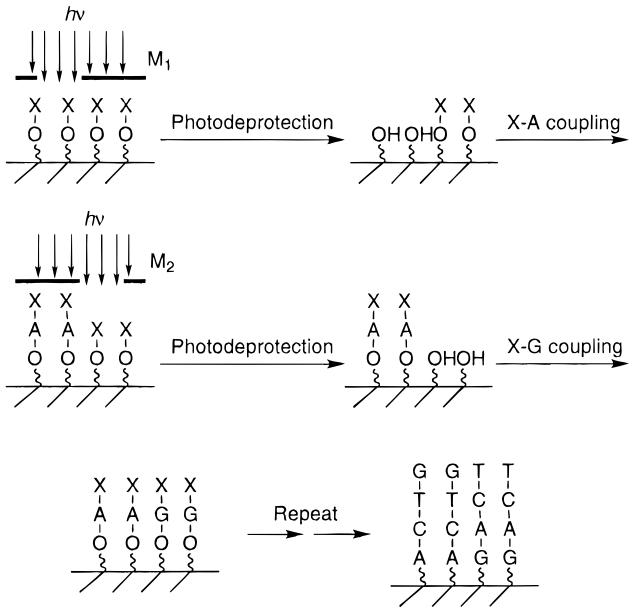


Figure 17. Principles of light-directed synthesis, or VLSIPS.

oligomers. The pattern of masks and the sequence of reagents define the ultimate sequences synthesized; their locations on the surface are known. Each site is accessible for interactions with other molecules and such an array can be used to conduct binding assays with vast parallelism. The technique accomplishes light-directed spatially addressable parallel chemical synthesis; for the applications envisioned, it is a method for VLSIPS. Light-directed synthesis is distinct from Southern's method because it selectively deprotects only certain areas and exposes the whole surface to coupling reagents, requiring one irradiation/coupling cycle for each monomer to complete the extension of the whole surface by one residue.

The patterns of the masks are a crucial part of light-directed synthesis and offer significant flexibility compared to physical masking, which can efficiently address only one contiguous region with the same reagent. Light is a "random access" reagent. Two principal masking strategies were developed. A striped masking strategy is similar to Southern's work except that each of the building blocks must be coupled individually. It produces libraries of n^l different molecules, where n is the number of building blocks and l is the length of the oligomers, in $l \times n$ chemical operations. A second strategy, termed binary, does not produce molecules of a fixed length; rather, it generates a nested, ordered set of peptides. Binary masking can produce up to 2^n compounds in n chemical steps. Half of the surface is coupled in each step, thereby doubling the number of compounds in the array. An example of a four-step binary synthesis (Figure 18) using the generic building blocks A, B, C, and D, in that order, demonstrates that the synthesis produces all single and multiple deletion and truncation peptides from the sequence ABCD.

The surface for light-directed peptide synthesis is a glass microscope slide derivatized with an aminopropylsilane linker capped with aminohexanoic acid derivatized at its N-terminus with the [(nitroveratryl)oxy]carbonyl (Nvoc) group (Figure 19). The amino acids are protected at their N-terminus with Nvoc and, for those that require it, at their side chain functionality with the same mild acid-sensitive blocking groups commonly used for Fmoc chemistry. A synthesis cycle involves first photochemical deprotection (365 nm) of an Nvoc amine and double coupling (1 h each) with an amino acid activated as its OBt (hydroxybenzotriazole) ester through the reagents HBTU/HOBt. All terminal Nvoc groups are removed through irradiation, and the N-terminus can be capped by acetylation. TFA removes side chain protecting groups. The equipment required for light-directed peptide synthesis includes a commercial automated peptide synthesizer, a flow cell that supplants its synthesis cartridge (to hold the substrate and deliver fluids from the synthesizer), a translation stage to move among the masks (all located on the same large-format glass plate, which is held against the synthesis substrate in a "contact" printing mode, Figure 20), and an arc lamp that can deliver collimated long-wavelength UV light to the mask. Each component is under computer control. The binding assays use a similar type of flow cell and a confocal fluorescence microscope.

The first examples of libraries prepared by light-directed synthesis were based on the N-terminal

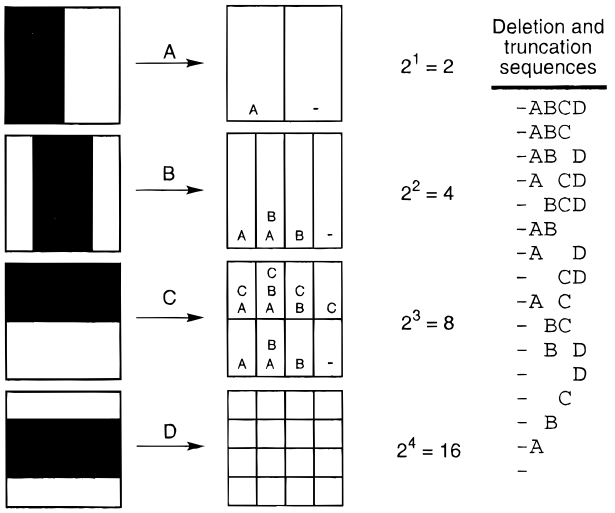


Figure 18. A binary masking strategy prepares the complete peptide on the basis of the chemical vector [ABCD] and all single and multiple deletion and truncation peptides of this sequence. The positions of the peptides prepared are as shown.

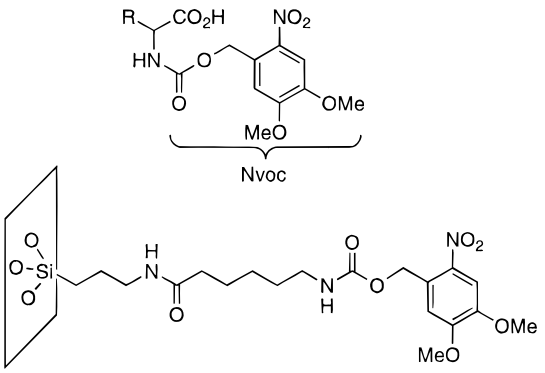


Figure 19. α-Amino acid building blocks protected with the [(nitroveratryl)oxy]carbonyl group and the surface linker for light-directed peptide synthesis.

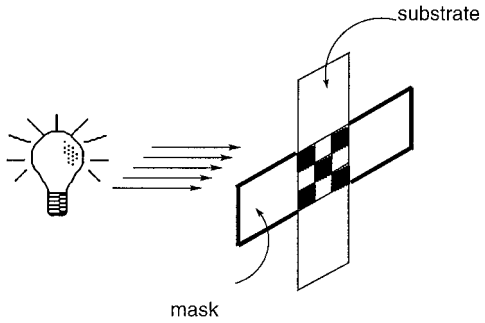


Figure 20. Contact printing in photolithography involves pressing the mask tightly onto the substrate and imposing an image of the mask onto it.

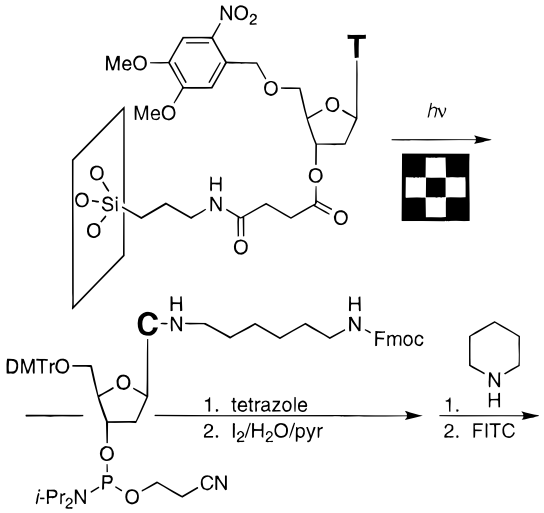


Figure 21. Preparation of a dinucleotide checkerboard by phosphoramidite chemistry using a surface functionalized with a nitroveratryl nucleoside.

pentapeptide of the leucine enkephalin sequence (YGGFL). A 10-step binary synthesis based on the sequence fYGAGTFLSF (where f is D-phenylalanine) produced an array of 1024 peptides in a 32 × 32 grid, with each peptide residing in a 400 μm × 400 μm area. When stained with the 3E7 antibody, which recognizes amino terminal enkephalin sequences, the array exhibited relatively few sequences with significant binding. They all begin YG(A/G)(F/L). Another binary synthesis identified YGAF peptides that give higher signals than the YGGFM immunogen used to raise the antibody.

A simple DNA grid was also prepared in this first report (Figure 21). The 5′-nitroveratryl ether of thymidine was linked onto an amino surface and irradiated in a 500 μm checkerboard pattern. A deoxycytidine phosphoramidite bearing a protected amino linker was then coupled using amidite chemistry, the Fmoc group was removed, and the free amine was stained with fluorescein.

As oligomeric molecules, libraries of peptides and DNA can be prepared by a permutational synthesis (using "stripe" masking strategies like in the Southern method) in which the number of different sequences is an exponential function of their length. The size of permutational libraries is therefore most sensitive to oligomer length. The number of steps in a permutational synthesis is the product of the length and the number of building blocks, so that when a smaller number of building blocks is used,

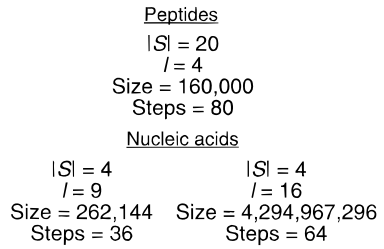


Figure 22. Comparison of the sizes of combinatorial libraries that can be prepared from peptides and amino acids. The library of hexadecanucleotides is larger than the whole human genome, in principle permitting such a library to uniquely address sites within it.

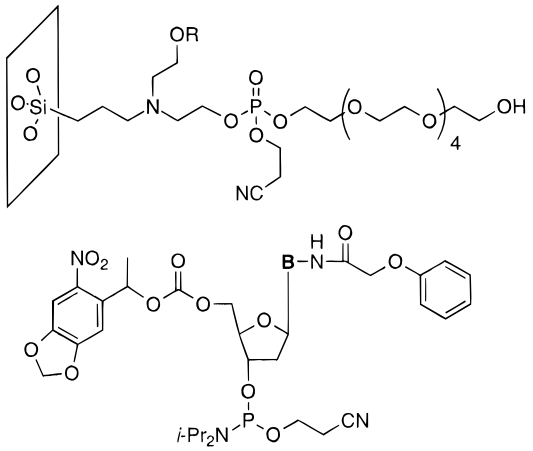


Figure 23. The surface linker for light-directed oligonucleotide synthesis and the N-protected nucleoside building blocks derivatized with the [(methylnitropiperonyl)-oxy])carbonyl group.

the number of chemical steps that can be used can increase. This is illustrated in a comparison between DNA and peptide libraries in Figure 22. A library of naturally occurring tetrapeptides requires 80 steps to prepare and contains 160 000 elements. A DNA library can be more than twice the length, yet be prepared in less than half the number of steps and generate a larger library. The ultimate expression of this power is that a library of hexadecanucleotides would encompass more sequences than present in the 3.5 billion nucleotide human genome, permitting sites within it to be individually addressed. Thus, light-directed synthesis (and any permutational synthesis method) is more powerful in making DNA than peptides.

The surface for light-directed oligonucleotide synthesis is a glass slide derivatized with poly(ethylene glycol) and protected with the [(methylnitropiperonyl)oxy]carbonyl (MeNPOC) group (Figure 23).[42] The reagents used for oligonucleotide synthesis are nucleoside 3′-[(cyanoethyl)phosphoramidites] protected at 5′ with MeNpoc and on the base nitrogen with "easy-off" phenoxyacetyl and isobutyryl protecting groups.

In initial proofing of the light-directed synthesis of oligonucleotides, the hydroxylated surface was coupled with a C phosphoramidite and patterned through a photolithographic mask. Conventional DMTr-phosphoramidite DNA synthesis was then conducted to generate the sequence 5′-GCCTACGC. The phosphate and amino protecting groups were removed with concentrated ammonia (4 h), and the

substrate was used for hybridization experiments. Incubation with fluoresceinated 10 nM 5′-GCG-TAGGC for 15 min at 15 °C produced 4-fold higher fluorescence in the irradiated areas compared to the background. The melting behavior of the probe was examined by increasing the temperature in 3 deg increments and imaging of the fluorescence. The melting temperature was 21−24 °C. Hybridization specificity was addressed by synthesis of 5′-GCCT-TCGC in stripes alternating with the above sequence. Formation of specific hybrids could be detected (>50:1 ratio of fluorescence) with each complementary target oligo, both by melting and annealing experiments.

The Affymax group has provided several summaries of their work in this whole area. [43−46]

C. Applications

1. Peptides

The initial result with the 1024 peptide array was extended through a 12-step binary synthesis based on the sequence YGPAFWGFMNLS. The so-produced 4096 peptide array was also used to map the epitope of the 3E7 antibody, with the result that the motif YGXFXX is dominant.[47] A 16-step binary synthesis based on the endorphin epitope, producing a 65 536 peptide chip in 50 μm^2 synthesis elements, has also been executed.

Light-directed synthesis has been applied by Holmes to epitope mapping of an antibody.[48−50] A new method developed in this work was the preparation of replicate arrays of peptides on a single surface so that direct comparisons could be made with different receptors or different chemical treatments (i.e., free amino terminus vs acetylation). Antibodies used were an anti-dynorphin B IgG (a bivalent receptor) raised against the opioid peptide YGGFLRRQFKVVT and a Fab fragment (a univalent receptor) derived from it. This work used a 10-step binary synthesis to prepare 1024 peptides based on the C-terminus of the antigen (FLRRQFKVVT). Since each synthesis site was 400 μm per side, the total array occupied a 1.28 cm square.

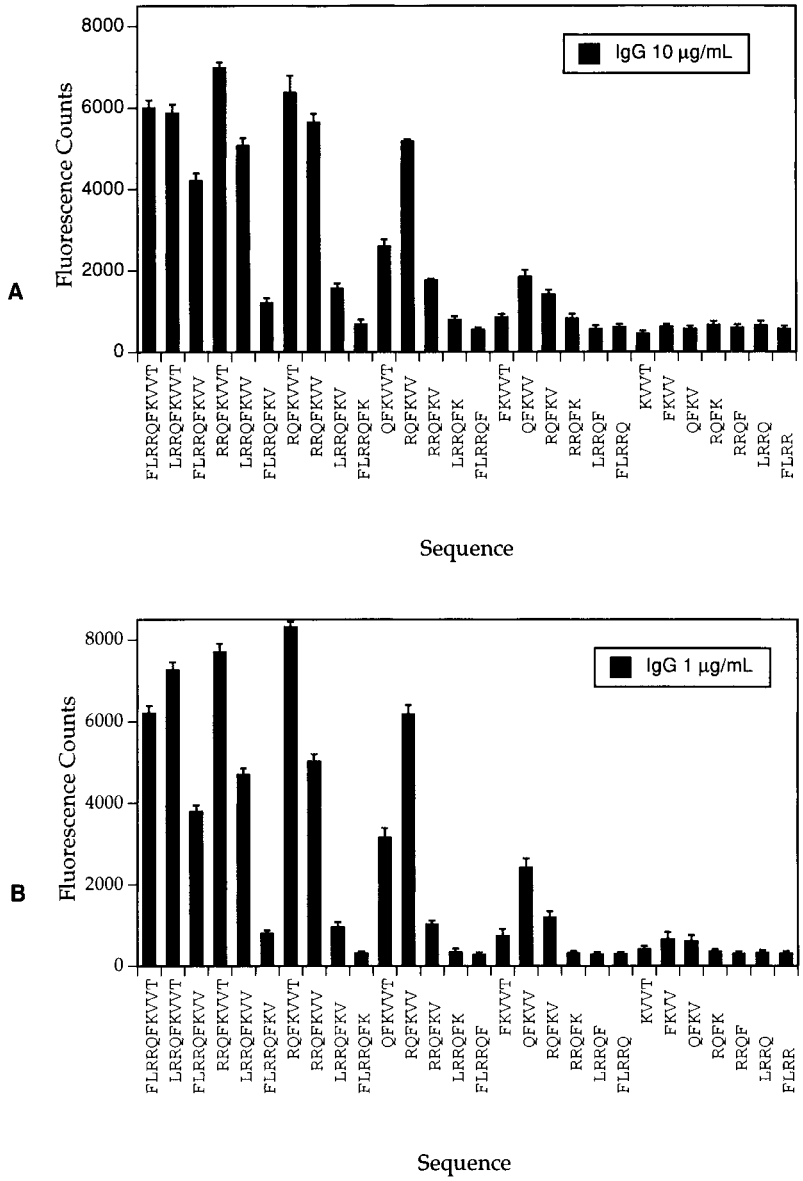


Figure 24. Truncation analysis of the peptides derived from the sequence FLRRQFKVVT. Surface fluorescence counts are given for each peptide sequence. (Reproduced with permission from ref 48. Copyright 1995 John Wiley & Sons.)

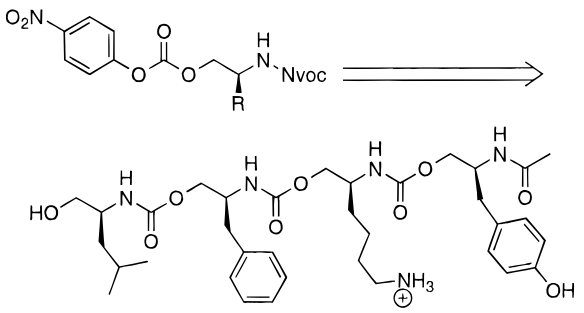


Figure 25. Building blocks used for the light-directed synthesis of oligocarbamates, and the YKFL oligocarbamate used to raise the antibody for probing for the oligocarbamate array.

These arrays comprise over 500 unique peptide sequences, including all truncation and deletion peptides based on the sequence used in the synthesis vector (defined in Figure 18 legend). The results of binding assays can therefore be organized to provide information about the epitope. The truncation data given in Figure 24 for both monovalent and bivalent receptors show excellent agreement, demonstrating that cooperative binding to peptide chains proximal on the surface does not lead to artifacts or influence the outcome of the screening. Deletion and frame shift data are also available from screening this array for binding. These extensive data permitted the characterization of the epitope recognized by the antibody as Ac-RQFKVV-NH_2. Noteworthy is that the preparation and screening of this array of peptides is fully parallelized and that all of the synthesis and data acquisition required to map this epitope required only three days.

Light-directed synthesis has also been applied to "epitope mapping" of nonnatural biopolymers.[51] An oligocarbamate derived from the amino acids YKFL was prepared and used to raise an antibody (Figure 25). A binary synthesis was used to prepare a library of 256 oligocarbamates, and the antibody was shown to recognize carbamates in the library containing the sequence FL.

2. DNA

The preparation of more complex DNA arrays was an important extension to the initial light-directed synthesis method. This has been accomplished at a subsidiary of Affymax, Affymetrix.[52] An array containing all 256 octanucleotides composed from dA and dT was prepared in 16 synthesis cycles requiring about 4 h. It was generated in 4-fold redundancy, resulting in 1024 elements in a 1.28 × 1.28 cm format, with 400 μm synthesis areas. Hybridization to a fluoresceinated A_8 probe stained sites besides T_8. Quantitation of the fluorescence signals allowed the influence of mismatch position on the formation of a probe−target duplex to be examined. Of the end (A−A) mismatches, a 3′-mismatch causes less perturbation than a 5′-mismatch. Internal mismatches cause even greater destabilization, consistent with expectations and earlier results concerning the importance of base stacking to duplex stability. A parallel experiment based on a dG/dT array hybridized to a fluoresceinated CCCAAACCCAA probe gave similar results.

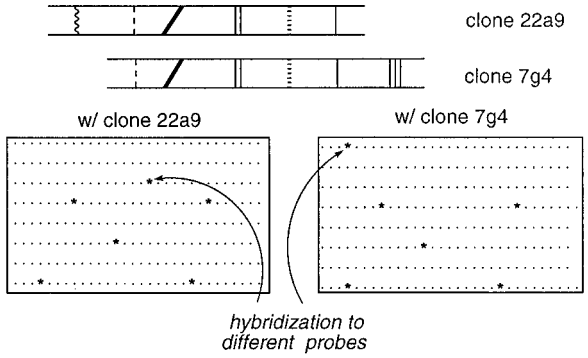


Figure 26. Principle of genetic mapping by hybridization. The two clones are mostly overlapping, but have regions of divergence. Most of the genetic markers along their sequences are held in common, but two are divergent. The probes on the array address a wide number of possible genetic markers (= unique DNA sequences), including those contained within these clones. When each clone is hybridized to the array, it will give signals at spots for genetic markers present in its sequence. Comparison of the pattern for each clone will identify the number of markers that they have in common and thereby define their degree of overlap.

Figure 27. Type IIS restriction enzymes used to cut four base regions out of clones for a fingerprinting process with oligonucleotide arrays.

Affymetrix has also reported efforts to use DNA chips in mock sequencing experiments.[42] An array of 256 octanucleotide probes was prepared on the basis of the sequence 5′-GC(A+G+C+T)^{4}GC. Hybridization of the fluorescently labeled target 5′-GCGGCGGC to this array at 15 °C yields the most signal at the perfect hybrid, but significant (up to 20% of the foregoing) signal is also seen at single-base mismatches. That is, hybridization is never a black/white proposition; there is always a gray area; hybridization will always be dependent on the exact sequences of the probes, the hybridization conditions, and the location of the mismatches (fraying at the 5′-end is common in mishybridization). Interestingly, the mismatches observed in this experiment are all internal.

Sapolsky has used oligonucleotide arrays to physically map, or place into order, clones from a genomic library.[53] The general idea of mapping clones by hybridization is that different specific sequences (markers) present in high molecular weight DNAs can be detected by hybridization, and that the pattern of hybridization of different clones will be similar if they overlap in the genome (Figure 26). A key to the Affymetrix mapping strategy is a special type of nuclease (type IIS restriction enzyme) that recognizes a specific sequence but cuts some distance (4−10 nts) away (Figure 27). The four bases in this adjacent region constitute the markers, which are enzymatically placed into a specific site in a 62 bp single-stranded, fluorescently labeled DNA by a series of enzymatic and PCR steps. This target is then used for hybridization to a DNA chip of 256 15-mers that are complementary to the target outside the marker region and constitute all possible combinations within

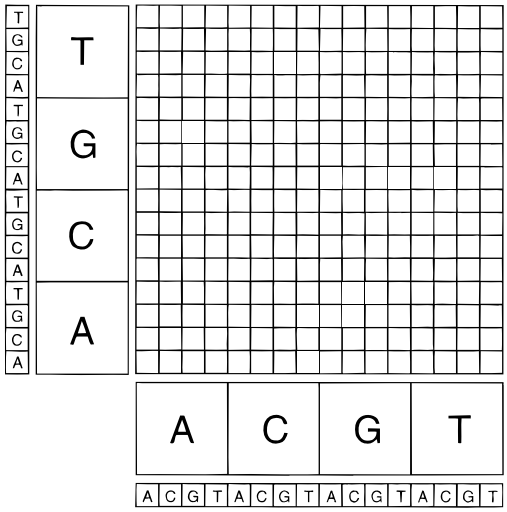


Figure 28. The oligonucleotide array pattern used to prepare all 256 tetranucleotides embedded within the sequence 5'-CTCTTCXXXXGCGTC.

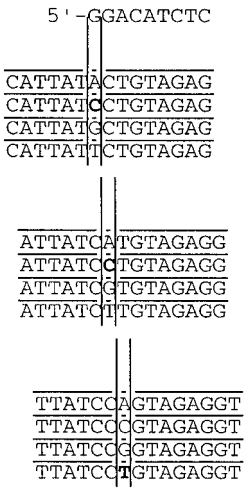


Figure 29. Mutation detection in the CFTR gene-based on arrays of pentadecanucleotides containing all four bases at the position being interrogated. The wild-type probe sequence is shown in bold.

it (Figure 28). A small library of cosmids containing yeast DNA was successfully mapped using this method, which offers the advantage that no gel electrophoresis is required.

DNA probe arrays that detect many mutations in many samples have been developed by Cronin for the interrogation of the cystic fibrosis transmembrane regulator (CFTR) gene.[54] Two approaches were taken. In one, a "tiling array" of 428 probes prepared in a 50-step light-directed synthesis scans 107 nts (including exon 11) of the CFTR gene. Probes (15-mers) are constructed in sets of four, each set interrogating successive nucleotide positions in the target (Figure 29). All single point mutations can be readily identified on the basis of hybridization to the non-wild-type probes. A "mutation array" is composed of two parallel probe sets, one complementary to the wild-type and one complementary to a previously known mutation. As in the tiling array, each position in a 5 nt mutation region is interrogated by a set of four 15-mer probes, each differing only at the center nucleotide. These arrays contain 1480 probes prepared in a 49-step synthesis and address 37 known mutation sites in CFTR. They readily identify heterozygous individuals on the basis of the presence of signals for both mutant and wild-type target DNA. These formats incorporate all one-base, internal mismatched comparison sequences in adjacent areas to compensate for signal variations across the array. Such variations exist because of differing thermal stability of hybrids in different sequence contexts. These arrays were hybridized with targets obtained by amplification from patient samples and successfully used for genotyping.

Affymetrix has also used "chip" technology to analyze genetic variations in HIV protease sequences across a population.[55] Such variations, *inter alia*, influence susceptibility to protease inhibitor AIDS drugs. These workers used a tiling array of 12 224 probes for "comparison" sequencing of over 100 viral isolates against a wild-type sequence, and they compared their results with conventional Sanger dideoxynucleotide chain termination sequencing. The agreement between the two was >98%. One of the problems identified in this study concerning the tiling array was difficulty in detecting multiple, closely spaced variations from the wild-type sequence. Since the sequence flanking the interrogation site in each probe is invariant, a mutation at one position in the DNA target not only causes it to bind preferentially to a variant probe at the interrogation site, it also diminishes its ability to hybridize to probes interrogating nearby sites.

A general description of the instrumentation system that Affymetrix is developing for research and clinical applications of DNA VLSIPS has been given.[56]

V. Peptides-on-Paper

A. Method

Frank has developed another method (SPOTs) for automated multiple peptide synthesis.[57] Simple cellulose paper (Whatman 540) is uniformly esterified with β-alanine, and another β-alanine residue is attached to the surface only in the 3 mm diameter spots that will be used for peptide synthesis. The spots are laid out at 4 mm spacing in a 17 × 25 array, permitting 425 peptides to be synthesized in a 6.7 × 9.8 cm area (smaller than an index card). This number of peptides is particularly suited to the goal of preparing libraries with two genetically coded amino acid positions identified (20 × 20 = 400). Fmoc N-protection is used, and coupling is promoted by diisopropylcarbodiimide and *N*-hydroxybenzotriazole. The N-termini of the peptides are generally prepared by N-acetylation, and the side-chain *tert*-butyl protecting groups are removed with a standard TFA/scavenger cocktail. Peptide quantities per spot are 5−10 nmol, potentially amenable to microsequencing. HPLC and mass spectroscopy have shown that excellent yields and purities are obtained. A robotic workstation can execute the synthesis, but synthesis can also be performed by relatively inexperienced personnel.

One strategy for discovery of peptide sequences that has been applied with the SPOTs method is

drawn closely from Geysen's mimotope approach. Libraries of octapeptides are prepared with an equimolar representation of all amino acids (X) at six positions and known amino acids (⊗) at two defined positions (i.e., XXX⊗⊗XXX). An initial array is screened, and the peptide resulting in the most potent signal is used as a kernel for a second-generation array, to fix two more positions in a peptide of optimum activity. Third- and fourth-generation arrays are used to fix the remaining residues. SPOTs kits for the preparation of peptides on paper are available commercially from Genosys.

B. Applications

Much of the early research with SPOTs libraries focused on epitope mapping. Frank showed their utility for studying alanine substitutions in the epitope of an anti-human ICP36 monoclonal antibody. The parent peptide is only twice as potent as a discovered sequence Ac-FAAAYVAASAK where alanines replace several amino acids from the native sequence.[57] Frank also prepared what has been called by Houghten a positional scanning library,[58] wherein sublibraries are prepared with one set position and the others randomized [i.e., ⊗XXX + X⊗XX + XX⊗X + XXX⊗]. Using this method, he rediscovered the Ac-HPQGG peptide known to bind to streptavidin.

Tegge and Dostmann have used SPOTs arrays for the discovery of optimal recognition sequences for the cAMP and cGMP-dependent serine/threonine protein kinases PKA and PKG.[59] These enzymes have structural homology, but perform different roles in distinct ways, so an understanding of the target sequences on which they operate would be useful in placing them into their biological context. Screening involved incubating peptide arrays with an enzyme and radioactive ATP. A storage phosphor imager identified the spots that had incorporated ^{32}P. After five generations of array screening, including extending the octapeptide by a residue at each end, optimum substrate sequences were identified as Ac-KRAERKASIY for PKA and Ac-TQAKRKKSLA for PKG. The serine at position 3 is the site of phosphorylation in each. The former sequence was consistent with earlier data for PKA, while the latter was a novel substrate for PKG. Kinetic data for these and related sequences in solution supported the assignments of the optimum substrates. Other groups have also used the SPOTs method for evaluating kinase substrates. Toomik has studied the specificity of protein kinase A, protein kinase C, and casein kinases I and II without use of the mimotope strategy and therefore with a far smaller selection of peptides.[60]

Another application of SPOTs arrays is discovery of metal-binding peptides. Schneider-Mergener has prepared libraries of the composition ⊗X⊗X⊗X to screen for binding to technetium-99m.[61] Such sequences might be useful to add to recombinant single-chain antibodies for use in tumor imaging. An impressive and beautiful graphical presentation of the binding data for the libraries, obtained by storage phosphor imaging, is best appreciated by viewing of the original journal article. Understandably, cysteine-containing peptides showed high binding activity, but other sequences were sought that might be especially selective for technetium. Two leads were identified, KXHXHX and KXMXHX, and further libraries were prepared on the basis of these sequences. The optimum sequences ultimately identified were KGHSHV and KAMYHG, and they were found to bind technetium sufficiently strongly that the complexes survive reverse-phase HPLC.

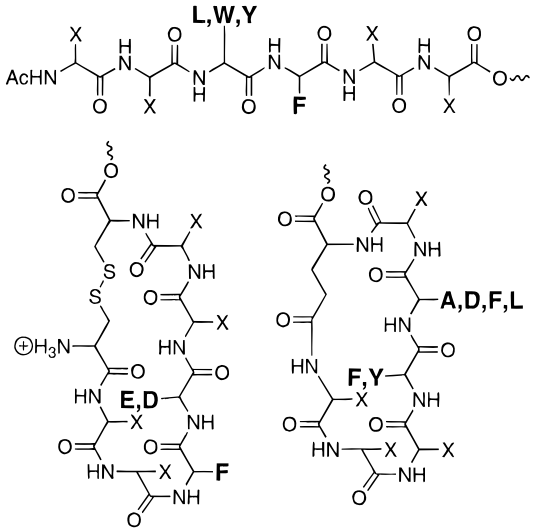


Figure 30. Linear and cyclic peptides that bind to transforming growth factor β.

Schneider-Mergener has screened XX⊗⊗XX SPOTs arrays for binding to nickel ion. Several cysteine-, histidine-, lysine-, and arginine-containing lead structures were identified, and some were subjected to iterative deconvolution. The identified sequences, such as HCRCHK, were prepared and shown to bind to a Ni^{2+} column in a pH-dependent manner.[62] SPOTs libraries including all L- and all D-hexapeptides of the sequence XX⊗⊗XX, and sequences that have been cyclized either through disulfide bond formation or γ-glutamine lactamation (Figure 30) were studied for binding to transforming growth factor β. Each gave different results, the deconvolution of which has not yet been reported.

Staining of a SPOTs library with silver ion, unsurprisingly, identified histidine and cysteine-containing peptides.[63] With other metals (iron, manganese, calcium, nickel), the pattern of affinities differed. The screening of a XX⊗⊗XX SPOTs library for binding to an unspecified double-stranded nucleic acid identified a family of lysine- and arginine-containing peptides, which excluded the acidic amino acids. Thus, it is reasonable that this binding process is primarily electrostatic in origin.

VI. Inorganic Combinatorial Libraries

A. Methods

Schultz has developed a shadow masking method for the assembly of spatially addressed arrays of metal oxide thin films with two different types of magnetic properties.[64] Overlaid on a single-crystal $LaAlO_3$ substrate is a primary mask that, in the most complex arrays made to date, forms 128 1 mm × 2 mm synthesis regions in a 16 × 8 format (Figure 31).

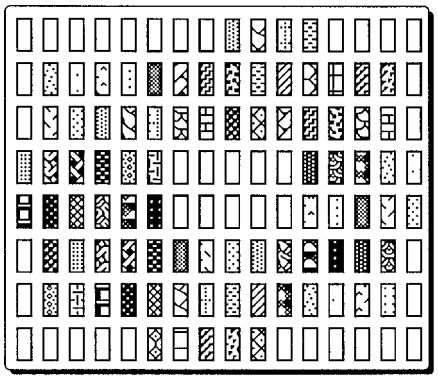


Figure 31. A 16 × 8 grid of magnetic materials prepared in thin film format by sputtering through shadow masks.

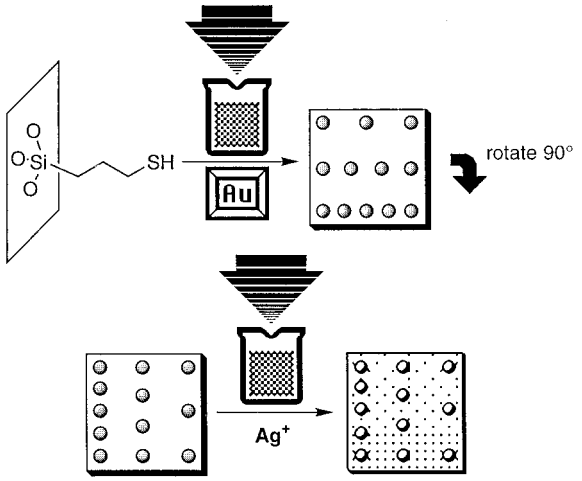


Figure 32. Protocol for the preparation of colloidal gold gradients in one dimension and of silver cladding in the second dimension for analysis of the enhancement factor in surface-enhanced Raman scattering.

Secondary masks are overlaid that conform to the binary synthesis protocols discussed above for peptide synthesis, and precursors are sputtered onto the surface through the secondary masks in a stepwise fashion, thereby generating an array containing all 2^n combinations that can be formed by deletions from the series of n masking−deposition steps. The stoichiometry of the phases can be modified by varying the sputtering time to vary the film thickness of each deposition. The coated substrate is subjected to thermal and oxidative treatment to form the final array of phases.

Natan has taken advantage of the relatively slow kinetics of the deposition of colloidal gold to form linear and step gradients of silver-clad gold particles on thiolated glass.[65] By dipping a substrate into a solution of a metal species at a controlled rate, the amount of metal deposited can be varied. Then, by turning the substrate at right angles and repeating the process with a different metal, a range of compositions in two dimensions can be generated (Figure 32). The composition at each site may not be immediately known, but the conditions required to generate it are, and the phase can be reproduced. Using a stepper motor with 3 μm resolution, it is possible to create on a 3 cm × 3 cm substrate $>10^8$ different compositions.

B. Applications

Schultz has prepared a 16-member binary library formed from Bi, Pb, Ca, and Sr precursors, in 1:1 stoichiometry, with Cu present in all locations.[66] The resistance of each site was measured by a probe as a function of temperature, and superconductivity (80−90 K T_c) was found in two sites: known high-temperature superconductors $BiCuCaSrO_x$ and $BiPbCuCaSrO_x$. The dependence of stoichiometry and deposition order of the properties of the former film was investigated in a 128-member library, and a $Bi_2Sr_2Ca_2Cu_3O_{10}$ phase with even higher (110 K) T_c was formed.

A second 128-member library of composition $Ln_xM_yCoO_z$ [Ln = Y or La; M = Pb, Ca, Sr, or Ba] aimed at discovering magnetoresistive materials was also prepared by Schultz.[67] The stoichiometries were varied, as was the thermal treatment of the deposited thin films. The resistivity of each sample as a function of temperature and applied magnetic field was determined by a probe, and three phases [$La_xM_yCoO_z$, where M = Ca, Sr, or Ba] that exhibit a large GMR (giant magnetoresistive) effect were found. Interestingly, the effect increases with the size of the alkaline earth ion, in contrast to the behavior of Mn-based perovskite oxides, such as $La_2CaMn_3O_9$. When one of these materials ($La_{0.58}Sr_{0.41}CoO_x$) was prepared in bulk, it exhibited twice the magnetoresistive ratio (60% to 30%) of the thin film. A significant aspect of this work is that previous giant and colossal magnetoresistive materials have been exclusively Mn-based.

Natan has applied his gradient deposition method to discover an optimum silver-clad gold surface for surface-enhanced Raman scattering (SERS). On a 2 cm × 2 cm substrate, the gold was deposited over 6 h, while the silver ion was deposited over 22 min. A SERS instrument with 2 mm spatial resolution was used to scan the surface to locate the most active region, which was further probed to locate a region over 100-fold more effective than background.

VII. Conclusion

The broad range of molecular functions that have already been addressed by combinatorial chemistry is impressive, and it is particularly notable that spatially addressable libraries have played such an important part in enabling this breadth of investigation. Future developments likely to impact on our ability to make chemistry spatially directable include extensions of ink-jet technology,[68] which should be able to deliver any fluid reagent to sites in the multimicron dimension. Indeed, this method already is under investigation in many industrial and academic laboratories.

VIII. References and Notes

(1) Pirrung, M. C. *Chemtracts: Org. Chem.* **1995**, *8*, 3−8.
(2) Houghten, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 5131−5135.
(3) A method for preparation of peptide libraries on polyethylene tiddly winks has been described as spatially addressable, but does not fit into that framework as used in this review: Luo,

K.; Zhou, P.; Lodish, H. F. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 11761−11765.
(4) DeWitt, S. H.; Kiely, J. S.; Stankovic, C. J.; Schroeder, M. C.; Cody, D. M. R.; Pavia, M. R. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6909−6913.
(5) Geysen, H. M.; Meloen, R. H.; Barteling, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 3998−4002. Geysen, H. M.; Barteling, S. J.; Meloen, R. H. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 178−182.
(6) Geysen, H. M. *Pept. Chem. 1992, Proc. Jpn. Symp., 2nd*; Yanaihara, N., Ed.; ESCOM: Leiden, Netherlands, 1993; pp 3−6.
(7) Geysen, H. M.; Rodda, S. J.; Mason, T. J. *Pept.: Chem. Biol., Proc. Am. Pept. Symp. 10th*; Marshall, G. R., Ed.; ESCOM Sci. Pub.: Leiden, Netherlands, 1988; pp 519−523.
(8) Bray, A. M.; Maeji, N. J.; Geysen, H. M. *Tetrahedron Lett.* **1990**, *31*, 5811−14.
(9) Bray, A. M.; Maeji, N. J.; Valerio, R. M.; Campbell, R. A.; Geysen, H. M. *J. Org. Chem.* **1991**, *56*, 6659−66.
(10) Bray, A. M.; Maeji, N. J.; Jhingran, A. G.; Valerio, R. M. *Tetrahedron Lett.* **1991**, *32*, 6163−6166.
(11) Schaaper, W. M. M.; Beekman, N. J. C. M.; van Lierop, M. J. C.; Hensen, E. J.; Meloen, R. H. *Pept. 1992, Proc. Eur. Pept. Symp., 22nd*; Schneider, C. H., Eberle, A. N., Eds.; ESCOM: Leiden, Netherlands, 1993; pp 312−313.
(12) Schaaper, W. M. M.; Beekman, N. J. C. M.; Hage-Van Noort, M.; Briel, P.; Kuperus, D.; Meleon, R. H. *Pept.: Chem. Biol., Proc. Am. Pept. Symp., 12th*; Smith, J. A., Rivier, J. E., Eds.; ESCOM: Leiden, Netherlands, 1992; pp 651−652.
(13) Wiesmueller, K. H.; Treffer, U.; Spohn, R.; Jung, G. *Pept. 1992, Proc. Eur. Pept. Symp., 22nd*; Schneider, C. H., Eberle, A. N., Eds.; ESCOM: Leiden, Netherlands, 1993; pp 308−309.
(14) Gausepohl, H.; Kraft, M.; Boulin, C.; Frank, R. W. *Innovation Perspect. Solid Phase Synth. Collect. Pap., Int. Symp., 1st* Epton, R., Ed.; SPCC (UK): Birmingham, UK, 1990; pp 487−490. Gausepohl, H.; Boulin, C.; Kraft, M.; Frank, R. W. *Pept. Res.* **1992**, *5*, 315−20.
(15) Arendt, A.; McDowell, J. H.; Hargrave, P. A. *Pept. Res.* **1993**, *6*, 346−352.
(16) Valerio, R. M.; Benstead, M.; Bray, A. M.; Campbell, R. A.; Maeji, N. J. *Anal. Biochem.* **1991**, *197*, 168−177.
(17) Geysen, H. M.; Mason, T. J. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 397−404.
(18) Geysen, H. M.; Mason, T. J.; Tribbick, G.; Edmundson, A. B. *Int. Congr. Ser.—Excerpta Med.* **1988**, *823*, 315−24.
(19) Wang, J. X.; Bray, A. M.; DiPasquale, A. J.; Maeji, N. J.; Geysen, H. M. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 447−450.
(20) Seldon, M. A.; Karabatsas, M.; Bray, A. M.; Maeji, N. J; Geysen, H. M. *ImmunoMethods* **1992**, *1*, 25−31.
(21) Maeji, N. J.; Bray, A. M.; Geysen, H. M. *J. Immunol. Methods* **1990**, *134*, 23−33.
(22) Carter, J. M. *Methods Mol. Biol. (Totowa, N.J.)* **1994**, *36*, 207−223.
(23) Bastos, M.; Maeji, N. J.; Abeles, R. H. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6738−6742.
(24) Chaturvedi, D.; Huelar, E.; Gunthorpe, M.; Gofman, M.; Krapf, D. S.; Apostol, E.; Lewis, W. S. *Pept. Res.* **1993**, *6*, 308−312.
(25) Tallett, A.; Seabrook, R. N.; Irons, L. I.; Robinson, A.; Van Heyningen, S.; Atkinson, T. *Eur. J. Biochem.* **1993**, *211*, 743−748.
(26) Rao, Yo.; Wu, X. F.; Gariepy, J.; Rutishauser, U.; Siu, C. H. *J. Cell Biol.* **1992**, *118*, 937−949.
(27) Wu, X. F.; Kamboj, R. K.; Gariepy, J.; Siu, C. H. *Biochem. Cell Biol.* **1992**, *70*, 246−249.
(28) Bunin, B. A.; Plunkett, M. J.; Ellman, J. A. *Proc. Natl. Acad. Sci., U.S.A.* **1994**, 91, 4708−4712.
(29) Maskos, U.; Southern, E. M. *Nucleic Acids Res.* **1992**, *20*, 1675−1678.
(30) Southern, E. M.; Maskos, U.; Elder, J. K. *Genomics* **1992**, *13*, 1008−1017.
(31) Maskos, U.; Southern, E. M. *Nucleic Acids Res.* **1992**, *20*, 1679−1684.
(32) Southern, E. M.; Case-Green, S. C.; Elder, J. K.; Johnson, M.; Mir, K. U.; Wang, L.; Williams, J. C. *Nucleic Acids Res.* **1994**, *22*, 1368−1373.
(33) Southern, E. M. *Trends Genet.* **1996**, *12*, 110−115.
(34) Case-Green, S. C.; Elder, J. K.; Mir, K. U.; Maskos, U.; Southern, E. M.; Williams, J. C. *Innovation Perspect. Solid Phase Synth. Collect. Pap., Int. Symp., 3rd*; Epton, R., Ed.; Mayflower Worldwide Ltd.: Birmingham, UK, 1994; pp 77−82.
(35) Maskos, U.; Southern, E. M. *Nucleic Acids Res.* **1993**, *21*, 4663−4669.
(36) Williams, J. C.; Southern, E. M. *Nucleic Acids Res.* **1994**, *22*, 131−136.
(37) Maskos, U.; Southern, E. M. *Nucleic Acids Res.* **1993**, *21*, 2269−2270.
(38) Maskos, U.; Southern, E. M. *Nucleic Acids Res.* **1993**, *21*, 2267−2268.
(39) Williams, J. C.; Case-Green, S. C.; Mir, K. U.; Southern, E. M. *Nucleic Acids Res.* **1994**, *22*, 1365−1367.
(40) Bains, W.; Smith, G. C. *J. Theor. Biol.* **1988**, *135*, 303−307. Drmanac, R.; Labat, I.; Brukner, I.; Crkvenjakov, R. *Genomics* **1989**, *4*, 114−128. Khrapko, K. R.; Lysov, Y. P.; Khorlyn, A. A.; Shick, V. V.; Florent'ev, V. L.; Mirzabekov, A. D. *FEBS Lett.* **1989**, *256*, 118−122.
(41) Fodor, S. P. A.; Read, J. L.; Pirrung, M. C.; Stryer, L.; Liu, A. T.; Solas, D. *Science* **1991**, *251*, 767.
(42) Pease, A. C.; Solas, D.; Sullivan, E. J.; Cronin, M. T.; Holmes, C. P.; Fodor, S. P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 5022−5026.
(43) Jacobs, J. W.; Fodor, S. P. A. *Trends Biotechnol.* **1994**, *12*, 19−26.
(44) Dower, W. J.; Fodor, S. P. A. *Annu. Rep. Med. Chem.* **1991**, *26*, 271−280.
(45) Holmes, C. P.; Fodor, S. P. A. *Innovation Perspect. Solid Phase Synth. Collect. Pap., Int. Symp., 3rd*; Epton, R., Ed.; Mayflower Worldwide Ltd.: Birmingham, UK, 1994, pp 221−226.
(46) Holmes, C. P.; Fodor, S. P. A. *Harnessing Biotechnol. 21st Century, Proc. Int. Biotechnol. Symp. Expo., 9th*; Ladisch, M. R., Bose, A., Eds.; American Chemical Society: Washington, DC, 1992; pp 241−243.
(47) Gruber, S. M.; Yu-Yang, P.; Fodor, S. P. A. *Pept.: Chem. Biol., Proc. Am. Pept. Symp., 12th*; Smith, J. A., Rivier, J. E., Eds.; ESCOM: Leiden, Netherlands, 1992; pp 489−491.
(48) Holmes, C. P.; Adams, C. L.; Fodor, S. P. A.; Yu-Yang, P. *Perspect. Med. Chem.*; Testa, B., Kyburz, E., Fuhrer, W., Giger, R., Eds.; Verlag Helvetica Chim. Acta: Basel, Switzerland, 1993; pp 489−500.
(49) Holmes, C. P.; Adams, C. L.; Kochersperger, L. M.; Mortensen, R. B.; Aldwin, L. A. *Biopolymers* **1995**, *37*, 199−211.
(50) Adams, C. L.; Kochersperger, L. M.; Mortensen, R. B.; Aldwin, L. A.; Holmes, C. P. *Tech. Prot. Chem. V.*; Crabb, J. W., Ed.; Verlag Academic Press: San Diego, 1994; pp 525−532.
(51) 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. *Science* **1993**, *261*, 1303−1305.
(52) Fodor, S. P. A.; Lipshutz, R. J.; Huang, X. *Proc. Robert A. Welch Found. Conf. Chem. Res.* **1994**, *37*, 3−9.
(53) Sapolsky, R. J.; Lipshutz, R. J. *Genomics* **1996**, *33*, 445−456
(54) Cronin, M. T.; Fucini, R. V.; Kim, S. M.; Masino, R. S.; Wespi, Ri. M.; Miyada, C. G. *Hum. Mutat.* **1996**, *7*, 244−255.
(55) Kozal, M. J.; Shah, N.; Shen, N.; Yang, R.; Fucini, R.; Merigan, T. C.; Richman, D. D.; Morris, D.; Hubbell, E.; Chee, M.; Gingeras, T. R. *Nature Medicine* **1996**, *2*, 753−759.
(56) Kreiner, T. *Am. Lab. (Shelton, CT)* **1996**, *28*, 39−43
(57) Frank, R. *Tetrahedron* **1992**, *48*, 9217−9232. Frank, R. *Innovation Perspect. Solid Phase Synth. Collect. Pap., Int. Symp., 3rd*; Epton, R., Ed.; Mayflower: Birmingham, UK, 1994; pp 509−512.
(58) Dooley, C. T.; Houghten, R. A. *Life Sci.* **1993**, *52*, 1509−1517.
(59) Tegge, W.; Frank, R.; Hofmann, F.; Dostmann, W. R. G. *Biochemistry* **1995**, *34*, 10569.
(60) Toomik, R.; Edlund, M.; Ek, P.; Öbrink, B.; Engström, L. *Pept. Res.* **1996**, *9*, 6−11.
(61) Malin, R.; Steinbrecher, R.; Jannsen, J.; Semmler, W.; Noll, B.; Johannsen, B.; Frömmel, C.; Höhne, W.; Schneider-Mergener, J. *J. Am. Chem. Soc.* **1995**, *117*, 11821−11822.
(62) Kramer, A.; Schuster, A.; Reineke, U.; Malin, R.; Volkmer-Engert, R.; Landgraf, C.; Schneider-Mergener, J. *METHODS (Com. Methods Enzymol.)* **1994**, *6*, 388−395.
(63) Kramer, A.; Volkmer-Engert, R.; Malin, R.; Reineke, U.; Schneider-Mergener, J. *Pept. Res.* **1993**, *6*, 314−319.
(64) Hsieh-Wilson, L. C.; Xiang, X.-D.; Schultz, P. G. *Acc. Chem. Res.* **1996**, *29*, 164−70.
(65) Baker, B. E.; Kline, N. J.; Treado, P. J.; Natan, M. J. *J. Am. Chem. Soc.* **1996**, *118*, 8721−8722.
(66) Xiang, X.-D.; Sun, X.; Briceno, G.; Lou, Y.; Wang, K.-A.; Chang, H.; Wallace-Freedman, W. G.; Chen, S.-W.; Schultz, P. G. *Science* **1995**, *268*, 1738−1740.
(67) Briceno, G.; Chang, H.; Sun, X.; Schultz, P. G.; Xiang, X.-D. *Science* **1995**, *270*, 273−275.
(68) Warnica, W. D.; Van Reenen, M.; Renksizbulut, M.; Strong, A. B. *Rev. Sci. Instrum.* **1993**, *64*, 2334−2339. Andresen, P.; Faubel, M.; Haeusler, D.; Kraft, G.; Luelf, H. W.; Skofronick, J. G. *Rev. Sci. Instrum.* **1985**, *56*, 2038−2042.

CR960013O