Combinatorial Chemistry

Gordon Lowe
The Dyson Perrins Laboratory, Oxford University, South Parks Road, Oxford OX1 3QY, U.K.

1 Introduction

Combinatorial chemistry is a novel and innovative way of rapidly generating a large number of related compounds. The solid phase method of synthesis, first introduced by Merrifield for peptide synthesis some 30 years ago, is generally used for preparing combinatorial chemical libraries. The solid phase method has many advantages over solution phase chemistry, most notably that excess reagents may be used to drive reactions rapidly to completion and then by-products and excess reagents removed by a simple washing procedure. The whole process is capable of automation and instrumentation is commercially available for this purpose. Although the solid phase method of synthesis initially received much criticism, primarily because intermediates are not characterized, as reagents and protecting groups for peptide synthesis improved and the purification of peptides by HPLC became routine, the method was accepted as the preferred method of peptide synthesis. The solid phase method has also been successfully developed for the synthesis of oligonucleotides which are now routinely used in molecular biology.

Combinatorial chemistry was first reported by Arpad Furka and his colleagues from the Eotvos Lorand University in Budapest at international scientific meetings in Prague and Budapest in 1988.[1] However, the concept was not published in the primary literature until 1991 when three papers from the groups of Furka,[2] Houghten,[3] and Lam[4] heralded the advent of combinatorial chemistry to the scientific community. Although the generation of mixtures of organic compounds is an anathema to the traditional synthetic organic chemist, the concept has been accepted with remarkable alacrity, especially in the pharmaceutical industry. One reason for this is that the development of automated high throughput screening methods over the last decade has meant that a large number of compounds can be screened rapidly for biological activity. Many companies now have the capability of screening millions of compounds each year using these robotic systems. Thus the bottle-neck to drug discovery for many companies has become the availability of compounds for screening. It is instructive to note that the total number of compounds described in the chemical literature up to 1991 was about 11 million. By combinatorial methods that number of compounds can be made by a single machine (a so-called 'librarian') in less than a week The power of the method is so remarkable that it seems certain to change the course of synthetic organic chemistry.

Professor Gordon Lowe, F.R.S. is Professor of Biological Chemistry at Oxford University and Fellow of Lincoln College, Oxford. He was educated at the Royal College of Science, Imperial College, London where on graduating he was awarded the Governors' Prize and the Edmund White Prize. After obtaining his Ph.D. and D.I.C. in 1957 he moved to Oxford to undertake postdoctoral research with Professor Sir Ewart Jones and was appointed University Demonstrator and Weir Junior Research Fellow of University College, Oxford in 1959. He was elected an Official Fellow of Lincoln College, Oxford in 1962. In 1988 he was appointed Aldrichian Praelector in Chemistry at Oxford University and in 1989 appointed to his present position. He was elected F.R.S.C. in 1981 and Fellow of the Royal Society in 1984 and awarded the D.Sc. (Oxon) in 1985. He was awarded the Charmian Medal for Enzyme Chemistry by the Royal Society of Chemistry in 1983 and the Award and Medal of the Society for Stereochemistry in 1992.

2 Strategies for Generating Combinatorial Libraries

2.1 Recombinant DNA Methods

Recombinant DNA techniques had been used to achieve the directed random mutagenesis of proteins before the introduction of chemically synthesized combinatorial libraries. The phage display method is probably the most useful expression of this technology. In 1985, Smith described a method for the display of peptides on filamentous phage particles and in 1988 this was extended and improved.[5] The phage particle bearing the peptide with the highest affinity for an antibody was selected by screening and the phage particle then grown in *E. coli* to produce more of the high affinity peptide for identification. Although the method is undoubtedly powerful it is limited to generating libraries with the 20 proteinogenic amino acids.

2.2 The Split Synthesis Method

The basic strategy for synthesizing a combinatorial chemical library is outlined in Figure 1.[2,4] A batch of resin (usually polystyrene) suitable for solid phase synthesis, say a gram, is made up of several million porous spherical beads which are more or less homogeneous in size and loading capacity. They are commercially available in functionalized form and with chemical linkers of various types for different applications. The beads are divided into a number of aliquots of equal size and to each aliquot is coupled a different monomer (A, B, C, *etc.*). In the case of a combinatorial peptide library the monomers would be protected amino acids. When coupling is complete (this should be tested if possible) on each aliquot of resin, excess reagent is removed by washing with an appropriate solvent and the aliquots recombined and thoroughly mixed. This process of dividing the resin into aliquots, quantitatively coupling the monomer units separately on to each aliquot, and washing the resin before recombining and thoroughly mixing, may be repeated several times depending on the size of the combinatorial library required and the capacity of the resin to accommodate all the members of the library. If the same monomers are used in each round of coupling, a complete combinatorial library of oligomers will be obtained. If three monomers are used (A, B and C), two rounds of coupling gives a combinatorial library of $3^2 = 9$, three rounds gives a library of $3^3 = 27$, four rounds gives a library of $3^4 = 81$ *etc.* If a full complement of proteinogenic amino acids is used as monomers then the number of peptides in the library will be 20^n where n is the number of coupling cycles (Figure 2). There must, of course, be more beads in the batch of resin than there are variants in the combinatorial library. The multiplication factor, f, multiplied by the number of variants in the combinatorial library indicates the number of beads

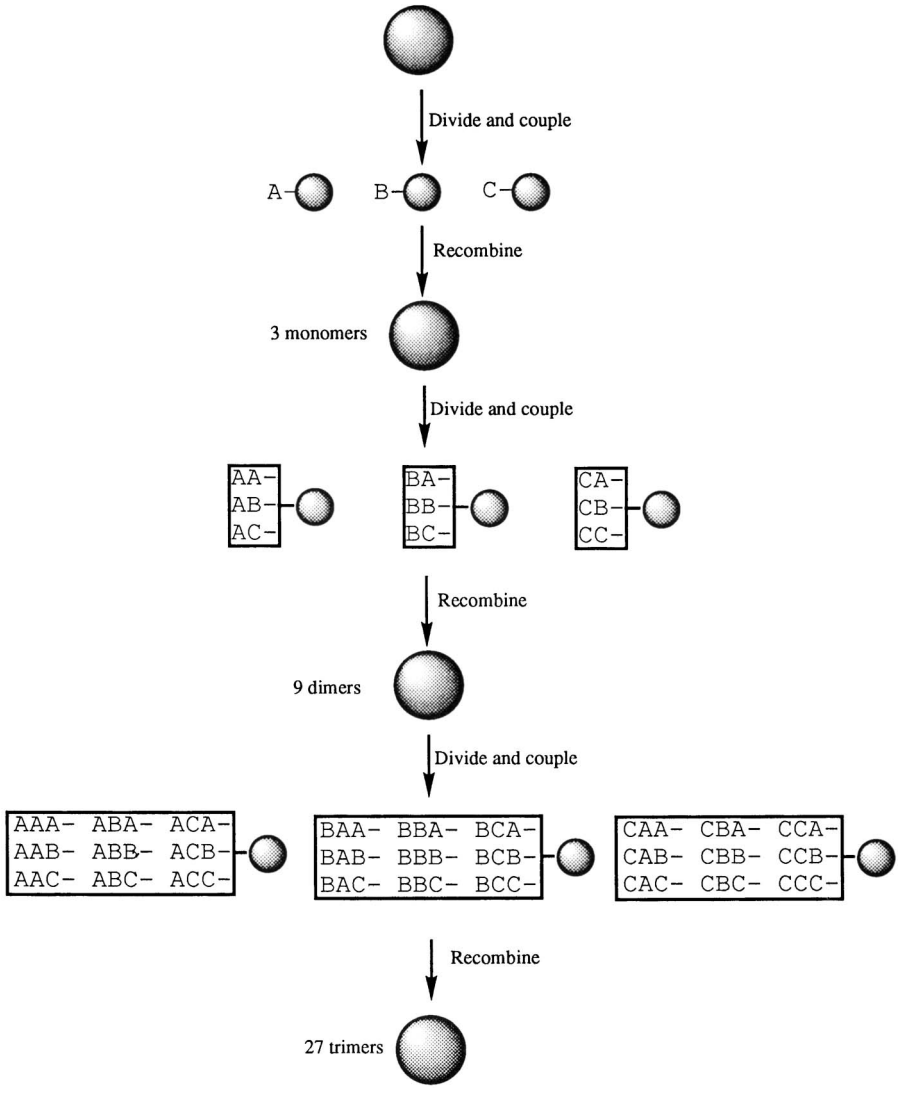


Figure 1 The split synthesis method of preparing a combinatorial library. In this illustration only three synthetic cycles with three monomers (A, B, and C) are used, giving a combinatorial library of 27 trimeric variants.

Dipeptides	20^2 =	400
Tripeptides	20^3 =	8,000
Tetrapeptides	20^4 =	160,000
Peptapeptides	20^5 =	3,200,000
Hexapeptides	20^6 =	64,000,000
Heptapeptides	20^7 =	1,280,000,000

Figure 2 The chemical diversity of combinatorial peptide libraries of increasing length using the 20 proteinogenic amino acids.

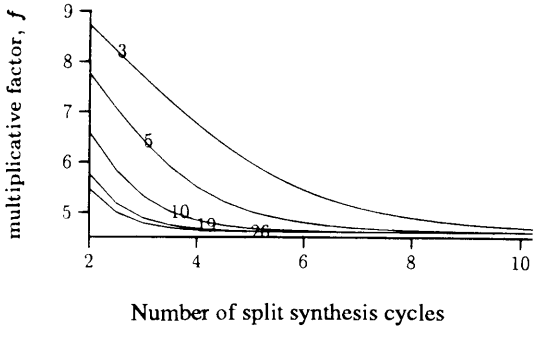


Figure 3 Statistical considerations in preparing a combinatorial library. The calculated multiplication factor, f, expressed as a function of the number of synthetic cycles for generating combinatorial libraries from 3, 5, 10, 19, and 26 monomers with a 99% probability that all variants will be represented.

required for a 99% probability that all variants of the library will be represented (Figure 3).[6]

It is important to appreciate that this synthetic strategy, generally called the 'split synthesis method', ensures that each bead in the batch of resin carries only a single member of the library, *i.e.* each bead behaves like a micro-reaction vessel. Of course, if sufficient beads have been used in the batch there will be several beads carrying the same structure. Indeed if sufficient beads have been used it will be possible to store a fraction of the beads for future use.

It will be obvious that there is a practical limit to the size of any library, which will depend on the amount of resin that can be conveniently handled. For example, a gram of resin containing several million beads would be capable of accommodating a tetrapeptide library of 160000 combinations with a high probability that all members would be represented, but the probability of all members of a pentapeptide library (3200000 combinations) being represented would be significantly reduced.

If the biological target used to screen the combinatorial library is soluble then there is a major advantage in screening the library while still attached to the beads. Since each bead carries a

single compound, by using a labelled biological target the beads carrying the compound with the highest affinity for the labelled target can be selected and the compound identified. Once identified, it may be synthesized on a larger scale and fully characterized.

2.3 Iterative Synthesis

2.3.1 Iterative Deconvolution or SURF (Synthetic Unrandomization of Randomized Fragments)

If the biological target is insoluble it is necessary to present the library of compounds in solution, which led Houghten to develop an iterative deconvolution approach. He chose to ask whether the naturally occuring met-enkephalin had the highest affinity for the μ-opioid receptor. Because the opioid receptor system consists of at least three sub-classes of receptors, a displacement assay of a radiolabelled analogue of met-enkephalin, [^{3}H]-DAGO, was used. Although the enkephalins are pentapeptides, Houghten chose to use a hexapeptide library with a free amino terminus and amidated C-terminus. 400 sub-libraries were initially prepared, each represented by the formula $O_1O_2XXXX\text{-}NH_2$, where O_1 and O_2 are one of the 20 proteinogenic amino acids defined in each sub-library and X represents any of the 19 proteinogenic amino acids, *i.e.* all except cysteine, thus giving 400 sub-libraries of 130 321 variants, *i.e.* a library of 52 128 400 hexapeptides. Each of the 400 peptide sub-libraries was used separately in solution to determine its ability to inhibit the binding of [^{3}H]-DAGO to the receptor. $YGXXXX\text{-}NH_2$ and $WWXXXX\text{-}NH_2$ were found to be the most effective sub-libraries. The $YGXXXX\text{-}NH_2$ sub-library was selected for further iterative deconvolution in which each X position was defined in turn with one of the 20 proteinogenic amino acids. Thus 20 sub-libraries of the general formula $YGOXXX\text{-}NH_2$ were prepared and screened, from which $YGGXXX\text{-}NH_2$ emerged as the preferred sequence. The next 20 sub-libraries of the sequence $YGGOXX\text{-}NH_2$ revealed $YGGFXX\text{-}NH_2$ to be the preferred sequence. The 20 sub-libraries $YGGFOX\text{-}NH_2$ revealed Met to be preferred in the fifth position giving the sequence YGGFM, which matches the natural sequence of met-enkephalin. On defining the final (sixth) position little improvement was made, the most effective peptide being $YGGFMA\text{-}NH_2$. It was encouraging that the natural sequence emerged as the most effective ligand for the μ-opioid receptor, although as yet iteration of the $WWXXXX\text{-}NH_2$ sequence has not been reported. It is noteworthy that the structure of the most effective ligand emerges from the iterative deconvolution approach without the need for sequencing.[3]

It is not obvious, however, that the iterative deconvolution strategy will always find the most active molecule. There may, for example, be many molecules that have nearly similar activity to the most active molecule in a combinatorial library which might make selection of the most active molecule difficult. A model study has been undertaken to investigate the importance of molecules with sub-optimal activity on the outcome of iterative deconvolution. Gratifyingly, it was found that for 9-mers or oligomers of greater length the molecule with the highest affinity should always be selected. For shorter oligomers, occasional sub-optimal selection may occur, but the free energy of binding of such sub-optimal molecules should be within 1 kcal/mol of that for the molecule with the highest affinity.[7]

2.3.2 Positional Scanning

Positional scanning was also developed in order to find the highest affinity peptide of a combinatorial library in solution, but unlike the iterative deconvolution approach, the number of molecules is the same in each successive set of combinatorial sub-libraries. A hexapeptide library was again used for determining the highest affinity sequence for the opioid receptor. Six positional combinatorial libraries were synthesized, each of which contained all possible combinations of 18 of the 20 proteinogenic amino acids (*i.e.* all except cysteine and tryptophan). Each positional library, $O_1XXXXX\text{-}NH_2$, $XO_2XXXX\text{-}NH_2$, $XXO_3XXX\text{-}NH_2$, $XXXO_4XX\text{-}NH_2$, $XXXXO_5X\text{-}NH_2$, and $XXXXXO_6\text{-}NH_2$ was kept as a set of sub-libraries (*i.e.* O is defined in each sub-library) and used to determine the most effective amino acid in each of the six positions of the hexapeptide. All possible sequences which were active at $IC_{50} \leq 125\mu M$ were synthesized and investigated separately. The most active was $YGGFMY\text{-}NH_2$, the first five amino acid residues being those found in met-enkephalin.[8] A similar strategy using 'indexed' combinatorial libraries has been used to generate a library of carbamates from alcohols and isocyanates from which an inhibitor of acetylcholine esterase was identified.[9]

2.4 Spatially Addressable or Array Synthesis

This form of combinatorial library is generated as a two-dimensional array, *e.g.* on a glass plate, each variant of the combinatorial library being identified by its position on the plate. By merging the techniques of solid phase chemistry and photolithography an array of peptides can be displayed which is spatially addressable.[10] The square plate is first functionalized with amino groups which are protected with the photolabile nitro-veratryloxycarbonyl (NVOC) group. By masking regions of the plate and removing the protecting group from exposed regions with light, amino groups become available for coupling at selected sites on the plate. The concept is shown in outline in Figure 4. Excellent spatial resolution can be achieved on sites of only 50 μm square, thus making it possible to provide 40 000 discrete synthesis sites on a 1 cm × 1 cm square plate. A combinatorial tetrapeptide library could thus be accommodated on a 2 cm × 2 cm square plate. The position of each member of the library depends on the strategy adopted for illuminating the plate. Many different masking strategies are possible.

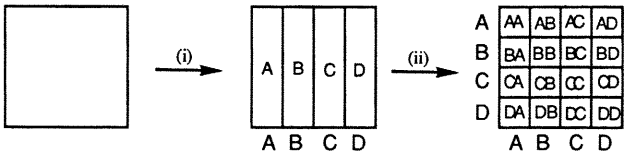
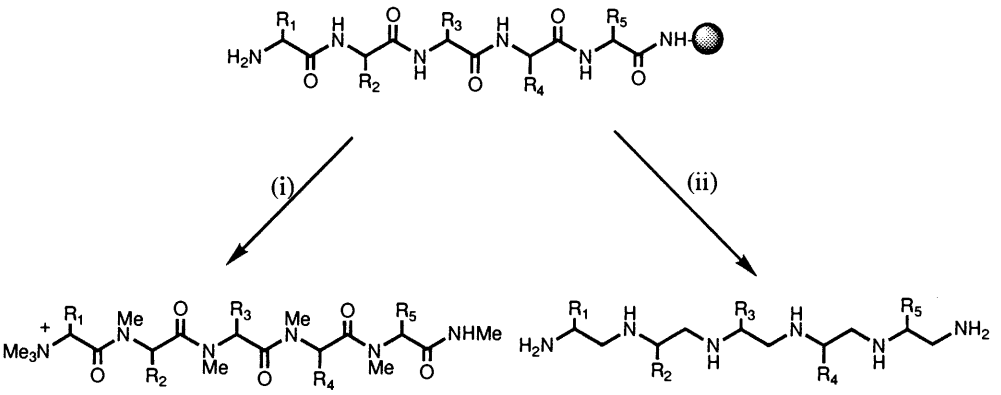


Figure 4 The orthogonal stripe strategy for array synthesis. (i) By masking the plate, vertical stripes may be exposed to illumination leading to deprotection and release of free amino groups. These are then coupled to photolabile protected amino acid A in the first stripe, B in the second stripe, *etc.* (ii) Orthogonal stripes are then generated by masking and illumination, and coupled to photolabile protected amino acid A in the first stripe, B in the second stripe, *etc.* If the same four monomers are used all 16 possible dimers will be generated and their position known from the synthetic strategy. Although not shown, each vertical stripe and each horizontal stripe could be divided again, say into four stripes, leading to a total of $4^4 = 256$ addressable tetramers.

There is considerable interest in the use of arrays of oligonucleotides for DNA sequencing and for diagnosing genetic mutations, and here spatial resolution has been improved, each synthesis site being 25 μm × 25 μm.[11] Hybridization of a labelled piece of DNA with an array of oligonucleotides gives rise to a pattern of complementary sequences from which the target DNA can be identified.[12] The recent demonstration that electron transfer through double-stranded DNA occurs extremely fast compared with single stranded DNA, the base pairs acting as a molecular wire, could enable an array of oligonucleotides to be used to identify the fragments from a restriction digest of DNA. By labelling the fragments with a ruthenium complex, on hybridization with an oligonucleotide array followed by illumination, electron transfer at sites bearing the complementary oligonucleotides could be detected on a silicon chip and the sequence of each displayed. A single nucleotide mismatch in a 15-mer leads to a marked reduction in the rate of electron



Reagents: (i)(a) NaH, (b) MeI, (c) HF
(ii) (a) reduction, (b) HF

Figure 5 Libraries from libraries. Permethylation and reduction of a combinatorial peptide library leading to new libraries.

transfer. Such signals could be eliminated by setting an appropriate threshold.[13]

2.5 Libraries from Libraries

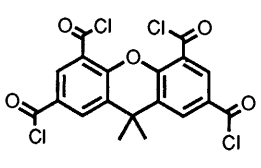
A combinatorial library may be transformed into an entirely new but structurally related library by one or more chemical reactions. Thus a combinatorial peptide library while still attached to a solid support may be permethylated or reduced (Figure 5).[14] If a suitable resin is used the new library may be released into solution.

2.6 Biased Combinatorial Libraries

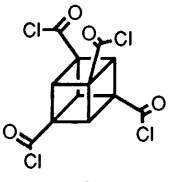
Src homology 3 (SH3) domain is a protein receptor commonly used in cellular signal transduction. Recognizing that several SH3 ligands contain the sequence PPXP Schreiber and his colleagues constructed a biased combinatoral peptide library, *i.e.* a library where some sites are predetermined. It took the form XXXPPXPXX (where P is proline and X is any of the proteinogenic amino acids except cysteine) on beads which were compatible with both the organic solvents required for the synthesis and aqueous solutions required for the biological assay. The PI3K SH3 and Src SH3 domains were labelled with fluorescein and used to screen the 2 million biased combinatorial peptide library whilst it was still attached to the resin beads. Two distinct classes of SH3 ligands emerged, and class I had the consensus sequence RXLPPZPXX (where Z represents arginine for PI3K SH3 and leucine for Src SH3). Since there was a conserved N-terminal arginine in the class I ligands and a conserved arginine in the class II ligands it was proposed that the peptides can bind to SH3 domains in either orientation. It was found that the palindromic PXXP motif was universally conserved in all SH3 domains. The highest affinity sequence for the PI3K SH3 domain was found to be RKLPPRPSK (K_D = 9.1 μM) and for the Src SH3 domain, RALPPLPRY (K_D = 7.8 μM). Because of the size of the peptide ligand recognized by the SH3 domains it was imperative to use a biased library in order to reduce the number of variants.[15]

2.7 Non-linear Combinatorial Libraries

The first combinatorial chemical libraries were made with amino acids as the monomeric building blocks, primarily because solid phase peptide synthesis has been greatly refined over the past 30 years. There is now, however, enormous interest in generating combinatorial libraries of non-linear organic molecules by solid phase methods. There is, of course, much less experience of performing organic synthesis generally on solid supports. Reagents which are themselves not freely soluble or which give rise to products which are not freely soluble are naturally unsuitable for the solid phase method of synthesis. But this is not such a limitation, in view of the vast array of reagents that are currently available. A more demanding requirement is that the reaction should be clean and occur rapidly and quantitatively, preferably at ambient temperature. This has been a long standing objective of synthetic organic chemists but is only rarely achieved. Nevertheless the advantage of performing reactions on a solid support is that an excess of reagent can be used without complicating the work-up procedure.



(1)



(2)

It is advantageous to start with a core structure not only for synthetic reasons but also because, if it is well chosen, a directed combinatorial library will be generated to explore structural and conformational space of the biological target. An example of this approach is provided by Rebek and his collaborators who used the 9,9-dimethylxanthene-2,4,5,7-tetracarboxylic acid chloride (1) and cubane-1,3,5,7-tetracarboxylic acid chloride (2) to generate disk-like and sphere-like combinatorial libraries using the four sites on each core structure.[16]

A piperazinedione combinatorial library has been synthesized as outlined in Figure 6. The core piperazinedione has three sites for incorporating molecular diversity and a library of a thousand members has been constructed.[17] Another example of this approach is the β-D-glucose based non-peptide mimetic library of a potent cyclic somatostatin agonist.[18]

2.8 Double Combinatorial Libraries

Pavia and his collaborators are developing double combinatorial libraries to generate even greater molecular diversity. First a combinatorial library is generated by adding a range of functional groups at various points on the first core molecule attached to a solid support. Then a second core molecule is attached to the first and again a combinatorial library of functional groups is attached to the second core molecule. The double combinatorial library is then ready to be released from the support or may be screened whilst still attached to the support (Figure 7). This strategy has been used to generate a

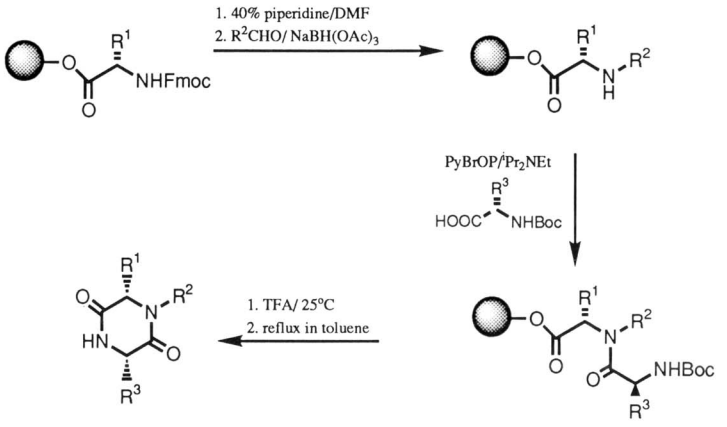


Figure 6 Synthesis of a piperazinedione combinatorial library.

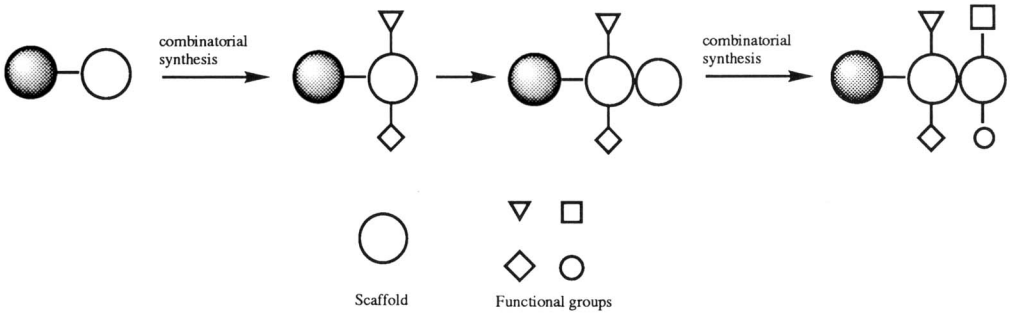


Figure 7 Synthesis of a double combinatorial library.

combinatorial library of biphenyls,[19] the two phenyl scaffolds being coupled by the Stille reaction.[20] Pavia and his colleagues have also developed equipment which is compatible with a wide range of reagents and temperatures and allows simultaneous multiple organic synthesis to be accomplished on solid supports.[21] Similar automated equipment is now commercially available.

2.9 Orthogonal Combinatorial Libraries

Suppose a combinatorial tripeptide library is required with 25 amino acid building blocks. In this approach, two orthogonal combinatorial libraries of the same size are prepared from the same building blocks. For the first library (A), the 25 amino acids are divided into 5 groups, A1 to A5, each group containing 5 amino acids. Each sub-library is produced by incorporating one group A*n* (n = 1,2,3,4, or 5) at each of the three positions in the tripeptide. Each sub-library of the A library contains, therefore, $5^3 = 125$ components, and there are 125 sub-libraries. For the second library (B), the same amino acid building blocks are used, giving the same 15625 tripeptides. However, they are generated orthogonally, using five different groups B1 to B5, each containing five amino acids (Figure 8). The important property of these libraries is that any sub-library from A and any sub-library from B have only one tripeptide in common. In a given assay, the most active sub-library from A and the most active sub-library from B will identify the most active compound of the whole library as the only component in common. Thus by assaying 250 sub-libraries the most active component of the 15625 tripeptides is identified. The orthogonal combinatorial library approach was used to identify a tripeptide which bound to the V2 vasopressin receptor at the nanomolar level.[22] The major advantage of this approach is that the highest affinity component of the library can be identified without analysis or iterative deconvolution.

2.10 Solid Supports

A number of solid supports for the synthesis of combinatorial libraries have been investigated. These include cellulose, cotton, glass as well as the large array of supports for solid phase peptide synthesis.[23] If the library is to be released into solution then the support does not need to be compatible with the aqueous medium required for the biological assay. However, if the library is to remain attached to the solid support for the biological assay then it is important that it is both compatible with the requirement of the synthetic methods used to prepare the library as well as the aqueous conditions needed for the biological assay. TentaGel beads, which have polyethylene glycol chains (average M_r 8000 daltons) grafted onto a low-cross-linked polystyrene backbone, are available in a variety of sizes from 10 μm to 750 μm and with loading from pmoles to nmoles per bead, are suitable for this purpose.

3 Screening Combinatorial libraries

3.1 Selection by Binding to a Biological Target

The screening of a combinatorial library for the highest affinity variant with a biological target is the commonest type of assay. Since it has been illustrated in Sections 2.3 and 2.6 it will not be discussed further.

3.2 Screening of Combinatorial Libraries on Solid Supports or in Solution

There has been considerable debate about the relative merits of screening combinatorial libraries in free solution or while still attached to the solid support on which they are synthesized. Clearly if the biological target is insoluble it is imperative to assay the library in solution. However, it has been argued that even for a soluble biological target, since a small structural change may change the activity of a bioactive molecule, leaving

(a)

	A1	A2	A3	A4	A5
B1	AA_1	AA_6	AA_{11}	AA_{16}	AA_{21}
B2	AA_2	AA_7	AA_{12}	AA_{17}	AA_{22}
B3	AA_3	AA_8	AA_{13}	AA_{18}	AA_{23}
B4	AA_4	AA_9	AA_{14}	AA_{19}	AA_{24}
B5	AA_5	AA_{10}	AA_{15}	AA_{20}	AA_{25}

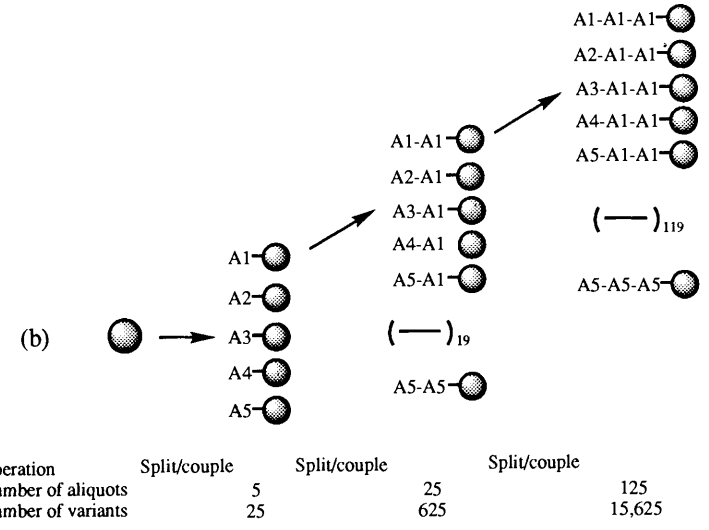


Figure 8 Synthesis of two orthogonal combinatorial libraries. The 25 amino acids are arranged in a 5 by 5 matrix, the columns A1 to A5 contain the groups of amino acids used to generate the A library and rows B1 to B5 contain the groups of amino acids used to generate the B library. For the synthesis of the A library, each group (A1 to A5) is coupled to an aliquot of beads separately and each aliquot of resin is then split into five aliquots for coupling to each group of amino acids (A1 to A5) into the second position. Each aliquot of resin is then split again into five aliquots for coupling to each group of amino acids (A1 to A5) in the third position. In this way 125 sub-libraries of the A library are generated. Likewise 125 sub-libraries of the B library are generated using the orthogonal groups B1 to B5.

the library attached to the massive structure of the solid support may affect the biological activity and accessibility and therefore it is preferable to assay the library in solution. An argument against assaying combinatorial libraries in solution is that the mixture of variants may have synergistic or antagonistic effects and so lead to erroneous conclusions being drawn. Moreover, if a combinatorial library is to be assayed in solution an iterative deconvolution protocol (Section 2.3.1) or an orthogonal combinatorial approach (Section 2.9) is required, which is labour intensive. On balance, therefore, it appears that where possible, *e.g.* with a soluble biological target, it is advantageous to synthesize the library on a support which is compatible with the biological assay and to screen the library while still attached to the support.

A number of experiments have been performed to investigate the ability of large biomolecules to penetrate the pores of TentaGel beads. A library of variants in the reactive site loop of the protein proteinase inhibitor CMTI-I (*Cucurbita maxima* trypsin inhibitor-I) were screened with fluorescein-labelled human leucocyte elastase (HLE). Using a confocal laser microscope, sections were taken through several of the most intensely fluorescent beads and shown to be homogeneously labelled from the surface to the core of the bead.[24] Some beads, however, were only partially labelled presumably because they bore variants with a lower affinity and competed less effectively for the limited amount of fluorescent-labelled enzyme.[25]

3.3 Determination of Enzyme Specificity

The mechanism of hydrolysis of a substrate (RCOX) by a serine proteinase (E) involves an acyl-enzyme intermediate (E-COR) formed through the hydroxyl group of the active site serine residue (Figure 9). Amines compete with water for the acyl-enzyme and divert substrate hydrolysis into peptide synthesis. If the amine is the N-terminus of a peptide which binds to the S′ subsites of the enzyme, it will compete for the acyl-enzyme more effectively than a peptide that does not bind, or does not bind so effectively. This strategy allows the binding preferences of the S and S′ subsites to be explored separately.

If a library of peptides is attached to TentaGel beads and the substrate ester is labelled with a fluorescent probe, it is possible by this strategy to identify and isolate those beads which carry peptides with high affinity for the S′ subsites of the enzyme. Using a fluorescently labelled substrate methyl ester and HLE, beads which became labelled were shown to be homogeneously fluorescent by confocal fluorescence microscopy.[25] If, however, the enzyme is not very specific for particular subsites then a large number of beads will become fluorescent and the analysis will be laborious. In this situation a sequential strategy can be adopted. In this approach a set of TentaGel beads carrying each of the 20 proteinogenic amino acids is incubated with the fluorescently labelled substrate and the enzyme, and the beads which become fluorescently labelled are identified. In fact this can be most

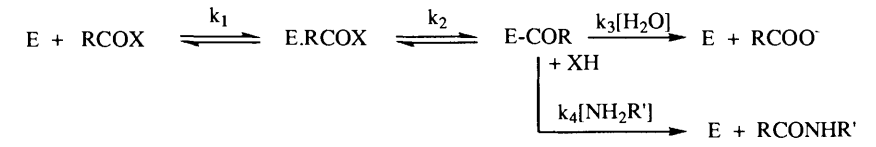


Figure 9 The mechanism of action of a serine proteinase. The acyl-enzyme intermediate can be diverted from hydrolysis into synthesis of an amide by an amine.

easily accomplished by attaching beads carrying each of the amino acids to separate sites on a microscope slide, thus generating an array of the 20 proteinogenic amino acids attached to the TentaGel beads. The position of the set of beads which becomes most intensely fluorescent can be identified by confocal laser microscopy. This technique is being used to explore the specificity of serine and cysteine proteinases.[26]

4 Analysis

4.1 Microsequencing of Peptides

The amino acid sequence on a resin bead, selected from a combinatorial peptide library by biological screening, may be obtained by Edman microsequencing. For this to be successful at least 2 pmole of peptide is required for each round of Edman sequencing. TentaGel beads of 90 μm diameter with a loading capacity of about 80 pmole per bead should be sufficient. Since a gram of resin contains about 2.86 million beads, a combinatorial tetrapeptide library is the largest that can be conveniently managed on these beads. If the analytical method becomes more sensitive with a new generation of sequencers, then it may be possible to use 35 μm beads which have a loading capacity of about 5.5 pmole/bead. There are about 45.5 million of these beads in a gram of resin which could accommodate a pentapeptide library. Moreover, fluorescent beads of this size could be isolated from the bulk of the beads by a standard cell sorter which has a nozzle size of 100 μm. The 90 μm beads are too large after being exposed to solvent as they swell to 150—160 μm.

4.2 Sequencing of Oligonucleotides

A combinatorial oligonucleotide library can be used on beads with a very low loading capacity since the oligonucleotide sequence on a selected bead can be amplified by the polymerase chain reaction (PCR) until sufficient is available for sequencing. This is a major advantage of oligonucleotides. Oligonucleotides, however, have not been widely explored for drug leads although an effective inhibitor of HIV envelope-mediated cell fusion was found in this way. Using a combinatorial phosphorothioate oligonucleotide library of 8-mers, a deconvolution strategy revealed that oligonucleotides with four contiguous guanosine nucleotides (G_4) were active inhibitors; $T_2G_4T_2$ was selected for further study. Such G_4 oligonucleotides form tetramers stabilized by Hoogstein-paired quartets of G in the presence of monovalent cations. These tetramers bind to the V3 loop of gp120 of HIV which contains a high percentage of positively charged amino acid residues. It was suggested that the G_4 tetramer provides a scaffold to present the phosphorothioate groups in a favourable orientation for binding to the cationic V3 loop, thus preventing HIV envelope cell fusion.[27] By screening combinatorial oligonucleotide libraries, inhibitors of thrombin and HIV reverse transcriptase have also been found.[28]

4.3 Encoded Combinatorial Libraries

4.3.1 Coding with Oligonucleotides

Because oligonucleotides may be amplified by PCR they provide an attractive and sensitive method of coding.[29] However, there are a number of matters which have to be addressed if this strategy is not to lead to errors. Only limited and carefully selected coding sequences should be used to avoid frame shifts. The solid support used must contain distinguishable sites for parallel synthesis and the protecting groups for the two synthetic procedures must be compatible. The coding region must be flanked by restriction sites. The encoding oligonucleotide should be synthesized on the solid support in parallel with the synthesis of the combinatorial chemical library. Thus, when a bead has been identified, the coding region for the combinatorial variant may be amplified by PCR and sequenced. Using this coding strategy a library of heptapeptides synthesized on 10 μm beads was screened for binding to an antipeptide monoclonal antibody using a fluorescence-activated cell sorter.[30]

4.3.2 Coding with Peptides

A peptide coding strategy was developed to enable a wider range of amino acid monomers to be used than the 20 proteinogenic amino acids for preparing combinatorial peptide libraries. Since Edman sequencing does not distinguish between D- and L-amino acids a coding strategy is necessary if D-amino acids are to be included. Orthogonal protecting groups (Fmoc and Ddz) were used to allow parallel synthesis of the combinatorial library and the encoding sequence. Ddz-Leu, Phe, Gly, and Ala were used as the encoding monomers so that side-chain protection was not required, and by using a triplet code, 64 amino acids could be encoded. By ensuring that the combinatorial library was blocked at the N-terminus, sequencing of the encoding peptide could be achieved.[31]

Recently a new range of TentaGel beads have become available which have orthogonal protective groups on the surface and interior synthesis sites. These beads could be used for this coding strategy although the size of the beads currently available (400—750 μm) would only allow relatively small combinatorial libraries to be synthesized. It seems likely, however, that orthogonally protected TentaGel beads of smaller size will become available.

4.3.3 Binary Coding with Molecular Tags

Coding of a combinatorial library generated by the split synthesis method may be achieved by coupling a molecular tag *via* a photolabile carbonate linker to about 0.5% of the synthesis sites at each synthesis cycle. The molecular tags consists of two components, a saturated aliphatic chain of defined but variable length and an aryl ether with defined but variable halide substitution (Figure 10). When active beads have been selected in the screening programme the molecular tags are released by irradiation at 350 nm and after silylation are identified by electron capture capillary gas chromatography. The high sensitivity of this analytical technique and the chemically inert nature of the molecular tags make this method of coding particularly attractive.[32] A variation of this encoding strategy has been reported in which the molecular tags are directly attached to the

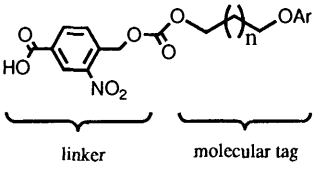


Figure 10 An encoded combinatorial library. Molecular tags are used to create a binary code during the synthesis of the combinatorial library. This can be subsequently used to identify the reaction history of individual beads.

resin matrix by a rhodium-catalysed carbene insertion reaction and released by oxidative cleavage.[33]

4.4 Mass spectrometry

Mass spectrometry is a sensitive and powerful method for structural analysis which has been used extensively for the determination of the sequence of peptides. Initially derivatization of peptides was essential but the recently introduced methods of electrospray ionization (ESI) and matrix assisted laser desorption ionization (MALDI) mass spectrometry make this unnecessary.

4.4.1 Ladder Sequencing

To obtain sequence information by this procedure a complete set of peptide fragments is generated by performing each cycle of Edman degradation (with phenyl isothiocyanate, PITC) in the presence of a few percent of terminating agent (*e.g.* phenyl isocyanate, PIC). When the initial PTC derivative is treated with trifluoroacetic acid to effect the cyclization and release the PTH amino acid, the phenyl carbamate (PC) peptide remains unaffected. Thus after several cycles of Edman degradation with a small amount of terminating agent used in each cycle, a mixture of PC peptides of decreasing length is obtained which can be submitted as a mixture for MALDI mass spectrometry (Figure 11). The mass spectrum contains the molecular ion of each terminated peptide. The mass difference between consecutive peaks corresponds to an amino acid, and the order of occurrence defines the amino acid sequence. If the peptide is attached to a bead of resin the nest of peptides must be detached before submission for mass spectrometry. Ladder sequencing lends itself to high sample throughput.[34]

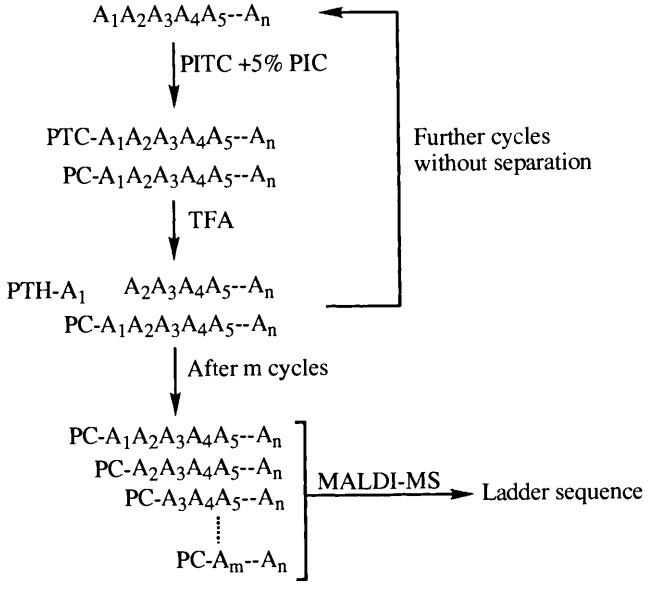


Figure 11 Peptide ladder sequencing. Reaction of the peptide with phenyl isothiocyanate (PITC) in the presence of about 5% phenyl isocyanate (PIC), followed by treatment with trifluoroacetic acid leaves a few percent of the stable phenyl carbamate (PC) derivative of the peptide and the truncated peptide. After Edman degradation is complete using this strategy, the mixture of PC-peptides is analysed by MALDI mass spectrometry.

4.4.2 Tandem Mass Spectrometry

Tandem mass spectrometry, also known as MS/MS, is the most effective method for directly determining the sequence of a peptide. The peptide (which does not need to be pure) is introduced *via* Fast Atom Bombardment (FAB), ESI, or MALDI into the first mass spectrometer where the MH^+ of the peptide of interest is selected for collisional induced dissociation (CID) prior to analysis in the second mass spectrometer. The m/z values of the fragments enable the sequence of the peptide to be read off from the mass spectrum. The latest refinement of the tandem method is the Fourier transform ion cyclotron mass spectrometer. This is a multiple tandem mass spectrometer (MS/MS^n) which can be used to analyse molecular or fragment ions produced by any ionization technique. This is an ultrasensitive method, capable of detecting femtomoles of sample, and being a multiple tandem method has the advantage of being able to distinguish, for example, between leucine and isoleucine (which are isobaric) by analysing the fragmentation of the side chain of these amino acids.[35]

5 Concluding Remarks

Although combinatorial chemistry is still in its infancy, there is intense interest in this new technology within the organic chemical community. There seems every reason to expect, therefore, that new solid supports will be developed with new linkers attached to them, so that a much wider range of chemistry can be used to generate combinatorial chemical libraries. Moreover, it seems likely that much more synthetic organic chemistry will be performed on solid supports in the future since this allows automated multiple synthesizers to be used which can perform several hundred reactions simultaneously. With these robotic systems synthetic organic chemistry could be transformed from a highly labour intensive discipline into a highly automated one, just as it has for peptide and oligonucleotide synthesis. This spin-off from combinatorial chemistry will be of immense benefit to organic chemistry in general and would further accelerate progress in the generation of new types of combinatorial chemical libraries for the exploration of structural and conformational space.

6 References

1 A. Furka, F. Sebestyen, M. Asgedom, and G. Dibo, *Abstr. 14th Int. Congr. Biochem., Prague, Czechoslovakia*, 1988, **5**, 47: *Abstr. 10th Int. Symp. Med. Chem.*, 1988, Budapest, Hungary, 288.
2 A. Furka, F. Sebestyen, M. Asgedom, and G. Dibo, *Int. J. Pept. Protein Res.*, 1991, **37**, 487.
3 R. A. Houghten, C. Pinilla, S. E. Blondelle, J. R. Appel, C. T. Dooley, and J. H. Cuervo, *Nature (London)*, 1991, **354**, 84.
4 K. Lam, S. Salmon, E. Hersh, V. Hruby, W. Kazmierski, and R. Knapp, *Nature (London)*, 1991, **354**, 82.
5 G. P. Smith, *Science*, 1985, **228**, 1315; S. F. Parmley and G. P. Smith, *Gene*, 1988, **73**, 305.
6 K. Burgess, A. I. Liaw, and N. Wang, *J. Med. Chem.*, 1994, **37**, 2985.
7 S. M. Freier, D. A. M. Konings, J. R. Wyatt, and D. J. Ecker, *J. Med. Chem.*, 1995, **38**, 344.
8 C. T. Dooley and R. A. Houghten, *Life Science*, 1993, **52**, 1509.
9 M. C. Pirrung and J. Chen, *J. Am. Chem. Soc.*, 1995, **117**, 1240.
10 S. P. A. Fodor, J. L. Read, M. C. Pirrung, L. Stryer, A. T. Lu, and D. Solas, *Science*, 1991, **251**, 767. C. P. Holmes and S. P. A. Fodor, 'Innovation and Perspectives in Solid Phase Synthesis', ed. Roger Epton, Mayflower Worldwide, 1994, 221.
11 A. C. Pease, D. Solas, E. J. Sullivan, M. T. Cronin, C. P. Holmes, and S. P. A. Fodor, *Proc. Natl. Acad. Sci., USA*, 1994, **91**, 5022.
12 E. M. Southern, U. Maskos, and J. K. Elder, *Genomics*, 1992, **13**, 1008: R. Drmanac, S. Drmanac, Z. Strezoska, T. Paunesku, I. Labat, M. Zeremski, J. Snoddy, W. K. Funkhouser, B. Koop, L. Hood, R. Crkvenjakov, *Science*, 1993, **260**, 1649.
13 T. J. Meade and J. F. Kayyem, *Angew. Chem., Int. Ed. Engl.*, 1995, **34**, 352.
14 J. M. Ostresh, G. M. Husar, S. E. Blondelle, B. Dorner, P. A. Weber, and R. A. Houghten, *Proc. Natl. Acad. Sci., USA*, 1994, **91**, 11138.
15 J. K. Chen and S. L. Schreiber, *Angew. Chem., Int. Ed. Engl.*, 1995, **34**, 953, and references there cited.
16 T. Carell, E. A. Winter, A. Bashir-Hashemi, and J. Rebek, *Angew. Chem., Int. Ed. Engl.*, 1994, **33**, 2059.
17 D. W. Gordon and J. Steele, *Bioorg. Med. Chem. Lett.*, 1995, **5**, 47.
18 R. Hirschmann, K. C. Nicolaou, S. Pietranico, J. Salvino, E. M. Leahy, P. A. Sprengeler, G. Furst, A. B. Smith, III, C. D. Strader, M. A. Cascieri, M. R. Candelore, C. Donaldson, W. Vale, and L. Maechler, *J. Am. Chem. Soc.*, 1992, **114**, 9217.

19 M. R. Pavia, Second European Conference on High Throughput Screening and on Exploiting Molecular Diversity, May 1995, Budapest.
20 F. W. Forman and I. Sucholeiki, *J. Org. Chem.*, 1995, **60**, 523.
21 S. H. Hobbs DeWitt, J. S. Kiely, C. J. Stankovic, M. C. Schroeder, D. M. Reynolds Cody, and M. R. Pavia, *Proc. Natl. Acad. Sci., USA*, 1993, **90**, 6909; D. R. Cody *et al.*, *US Patent*, 1994, No. 5,324, 483.
22 B. Déprez, X. Williard, L. Bourel, H. Coste, F. Hyafil, and A. Tartar, *J. Am. Chem. Soc.*, 1995, **117**, 5405.
23 Several contributions in 'Innovation and Perspectives in Solid Phase Synthesis', ed. R. Epton, Mayflower Worldwide, 1994.
24 G. Lowe and R. Quarrell, *METHODS: A Companion to Methods in Enzymology*, 1994, **6**, 411.
25 G. Lowe and R. Quarrell, submitted for publication.
26 O. W. Kingsbury, S. Leon, R. Quarrell, and G. Lowe, unpublished work.
27 J. R. Wyatt, T. A. Vickers, J. L. Robertson, R. W. Buckheit, Jr., T. Klimkait, E. N. Debaets, P. W. Davis, B. Rayner, J. L. Imbach, and D. J. Ecker, *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 1356.
28 L. C. Bock, L. C. Griffin, J. A. Latham, E. H. Vermaa, and J. J. Toole, *Nature (London)*, 1992, **355**, 546; C. Turek, S. Macdougal, and L. Gold. *Proc. Natl. Acad. Sci., USA*, 1992, **89**, 6988.
29 S. Brenner and R. A. Lerner, *Proc. Natl. Acad. Sci., USA*, 1992, **89**, 5381; J. Nielsen, S. Brenner, and K. D. Janda, *J. Am. Chem. Soc.*, 1993, **115**, 9812.
30 M. C. Needels, D. G. Jones, E. H. Tate, G. L. Heinkel, L. M. Kochersperger, W. J. Dower, R. W. Barrett, and M. A. Gallop, *Proc. Natl. Acad. Sci., USA*, 1993, **90**, 10700.
31 J. M. Kerr, S. C. Banville, and R. N. Zuckermann, *J. Am. Chem. Soc.*, 1993, **115**, 2529.
32 M. H. J. Ohlmeyer, R. N. Swanson, L. W. Dillard, J. C. Reader, G. Asouline, R. Kobayashi, M. Wigler, and W. C. Still, *Proc. Natl. Acad. Sci., USA*, 1993, **90**, 10922; A. Borchardt and W. C. Still, *J. Am. Chem. Soc.*, 1994, **116**, 373; S. S. Yoon and W. C. Still, *Tetrahedron*, 1995, **51**, 567.
33 H. P. Nestler, P. A. Bartlett, and W. C. Still, *J. Org. Chem.*, 1994, **59**, 4723; J. J. Baldwin, J. J. Burbaum, I. Henderson, and M. H. J. Ohlmeyer, *J. Am. Chem. Soc.*, 1995, **117**, 5588.
34 B. T. Chait, R. Wong, R. C. Beavis, and S. B. H. Kent, *Science*, 1993, **262**, 89.
35 F. W. McLafferty, *Acc. Chem. Res.*, 1994, **27**, 379.