Annu. Rev. Pharmacol. Toxicol. 2000. 40:273–82

PARALLEL ARRAY AND MIXTURE-BASED SYNTHETIC COMBINATORIAL CHEMISTRY: Tools for the Next Millennium

Richard A. Houghten
Torrey Pines Institute for Molecular Studies, San Diego, California 92121;
e-mail: rhoughten@tpims.org

Key Words deconvolution, iterative, libraries, positional scanning, screening

■ **Abstract** Technological advances continue to be a central driving force in the acceleration of the drug discovery process. Combinatorial chemistry methods, developed over the past 15 years, represent a paradigm shift in drug discovery. Initially viewed as a curiosity by the pharmaceutical industry, combinatorial chemistry is now recognized as an essential tool that decreases the time of discovery and increases the throughput of chemical screening by as much as 1000-fold. The use of parallel array synthesis approaches and mixture-based combinatorial libraries for drug discovery is reviewed.

INTRODUCTION

The philosophical and practical concepts encompassed by the varied approaches now termed combinatorial chemistry have fundamentally altered drug discovery. The first synthetic combinatorial chemistry methods were presented by Geysen et al (individual compounds chemically synthesized on plastic ''pins'') (1) and Houghten (simultaneous multiple compound synthesis on polystyrene resin ''tea bags'') (2) in 1984 and 1985, respectively. Both of these methods are based on Merrifield's pioneering work in solid-phase peptide synthesis (3). Even with the advent of automated means to carry out solid-phase peptide synthesis, the number of peptides available remained the limiting factor in virtually all studies. This was also true for the solution phase synthesis of classical heterocycles and other small molecule organic compounds. It is of note that Leznoff & Wong (4, 5) and Crowley & Rapoport (6) first carried out the solid-phase synthesis of heterocycles in 1973 and 1976, respectively, but this powerful approach remained unappreciated until work by Bunin & Ellman in 1992 (7). In the author's laboratory, the value of the parallel preparation of large numbers of peptides was shown in a 1988 study, in which more than 500 analogs of the 23-residue peptide

magainin were prepared (8). This increased synthetic capability (at the time, approximately 10-fold greater for peptides) has led to the successful clinical trial of the magainin analog Cytolex™ (MSI-78) and other therapeutic peptides (9). Furthermore, these approaches have made possible a range of basic research studies that were previously impractical due to economic or timeframe factors (reviewed in 10, 11).

As is true for all combinatorial chemistry approaches, the pin and tea-bag methods for solid-phase synthesis were first used to prepare peptides (1, 2, 11–14). They are now widely used to prepare large individual compound libraries of virtually all types, including peptidomimetics (15–19), oligonucleotides (20, 21), oligosaccharides (22–24), and heterocycles (25, 26). The widespread success of solid-phase synthesis is due to the ability to drive reactions to completion (often $>$99.8%) on polymer supports, the ability to readily remove excess reagents or starting materials, and automation (2, 7, 25). Despite the great increase in synthesis capability made possible by the pin and tea-bag techniques, it was soon clear that the barrier in virtually all studies still remained the number of compounds that could be synthesized. The successful preparation of mixture-based combinatorial libraries, in combination with earlier parallel-array synthesis approaches, has now pushed aside the synthesis limitations that were inherent in the drug-discovery process. Our laboratory's efforts have remained focused on the use of parallel-array synthesis for the preparation of very large, mixture-based combinatorial libraries.

MIXTURES VERSUS INDIVIDUAL COMPOUND ARRAYS

It is well accepted that the screening of natural product extracts is a productive source of therapeutic compounds. Such extracts are typically composed of hundreds to thousands of different compounds in varying concentrations. Additionally, highly active individual compounds have been found that were present in extracts at one part per 100,000 or less. In spite of the fact that natural products and, indeed, the very nature of biological interactions are inherently mixture-based, and the fact that these interactions do not occur in an environment made up of single compounds and single acceptors, an intense debate continues on the relative merits of the preparation and screening of individual compound arrays versus mixture-based combinatorial libraries of the same compounds. The central issue is the tradeoff between the time and cost necessary to acquire complete information about every compound making up an individual compound array, versus a somewhat less complete information set derived from the same compounds in a mixture-based library. Our experience over the past 15 years has found the generation and use of mixture-based combinatorial libraries to be an

extremely effective and cost-efficient means to generate highly active, therapeutically relevant individual compounds.

SYNTHESIS OF MIXTURES

Divide-Couple-Recombine (Split-and-Mix) Synthesis

One important issue in library synthesis is to obtain as close to equimolar representation as possible of all individual compounds within a mixture-based library for ease of deconvolution. For peptides, therefore, the various amino acids have to be incorporated into each of the library positions in a ratio as close to equimolar as possible. When using resin beads as the solid support for library synthesis, this can be achieved using a process known as ''divide-couple-recombine'' (DCR) (27), ''split-and-mix synthesis'' (28), or ''portioning-mixing'' (29). This process involves the coupling of each protected amino acid to be used for the library to separate portions of resin, followed by combining and mixing all resin portions, before dividing the resin again for the next coupling step. By repeating this process for a total of five couplings, and using 20 amino acids as building blocks, a library of 3,200,000 (20^5) pentapeptides can be readily prepared. Due to the physical separation of the resins prior to incorporating the individual amino acids, the DCR process yields libraries containing an individual, unique compound on each resin bead.

Coupling of Mixtures of Incoming Reagents

An alternative means for the introduction of mixture positions is through the coupling of mixtures of incoming reagents, such as protected amino acids. Due to the differences in coupling rates of the various amino acids, coupling of an equimolar amino acid mixture to resin-bound amino groups, as typically used in solid-phase peptide synthesis, will lead to nonequimolar incorporation of amino acids into the mixture positions. This ultimately will result in a highly nonequimolar distribution of individual peptides within the library. To overcome this problem, the ratio of protected amino acids within the coupling mixture is adjusted according to their different coupling rates, i.e. the higher the coupling rate of a particular amino acid, the lower the concentration of that amino acid in the coupling mixture (30, 31). Such ratios are established by adjusting the relative concentration of each amino acid according to its incorporation ratio after coupling of an equimolar amino acid mixture, as determined by amino acid analysis or HPLC. Because these amino acid mixtures are coupled in a large (i.e. 5- to 10-fold) molar excess over resin-bound amino groups, the coupling can be considered a pseudo-first-order reaction. The ratio of coupling rates of the amino acids within the mixture is therefore independent of the amino acid it is being coupled to.

DECONVOLUTION—FINDING THE NEEDLE IN THE HAYSTACK

Iterative Deconvolution

Geysen's early use of very large peptide mixtures immobilized on pins, while clearly of conceptual importance, has not been found to be generally useful. To deconvolute pin-immobilized peptide mixtures, Geysen et al used an iterative approach (30), as did the author's laboratory for soluble libraries (27). This approach steadily decreases the number of compounds per mixture while steadily defining successive positions. Iterative deconvolution of soluble (i.e. not immobilized on a solid support) mixture-based libraries has been found to be successful in a wide range of studies (reviewed in 32). The primary limitation of iterative deconvolution is the cost and time associated with the need for repetitive synthesis and screening steps, typically equal to the number of variable positions.

Positional Scanning Deconvolution

A rapid means to gather information about all possible variable positions in a library was presented by the author's laboratory in 1992 (33). An illustration of the positional scanning deconvolution approach for a simple tripeptide combinatorial library is shown in Figure 1. Four different amino acids are incorporated at each of the three diversity positions, resulting in 64 (4^3) individual peptides. When the same diversity is arranged as a positional scanning synthetic combinatorial library (PS-SCL), only 12 peptide mixtures (4 separate mixtures for each of the 3 positions) need to be synthesized. Each of the three positional sublibraries, namely OXX, XOX, and XXO, contain exactly the same diversity of peptides, differing only in the location of the position defined. Each of the O positions is singularly defined with one of the four amino acids, whereas the remaining two positions are mixtures (X) of the same four amino acids.

In this example, assume that the sequence RAT is the sole tripeptide having activity. Since each positional sublibrary contains the exact same diversity of peptides, the RAT tripeptide (outlined below each sublibrary in Figure 1) is present in all three positional sublibraries. Thus, the only mixtures with activity will be RXX, XAX, and XXT because the only active sequence, RAT, is present only in those mixtures. These three amino acids in their respective positions yield the tripeptide RAT, which can then be synthesized and tested for its individual activity. It should be noted that the activity observed for each of the three mixtures (RXX, XAX, and XXT) is due to the presence of the tripeptide RAT within each of these mixtures and is not due to the individual amino acids (R, A, and T) that occupy the defined positions. As expected, and found experimentally in more complex libraries, more than one mixture is usually found to have activity at each position. Selection of the building blocks for the synthesis of individual compounds is based first on the overall activity of the mixture and then on differences

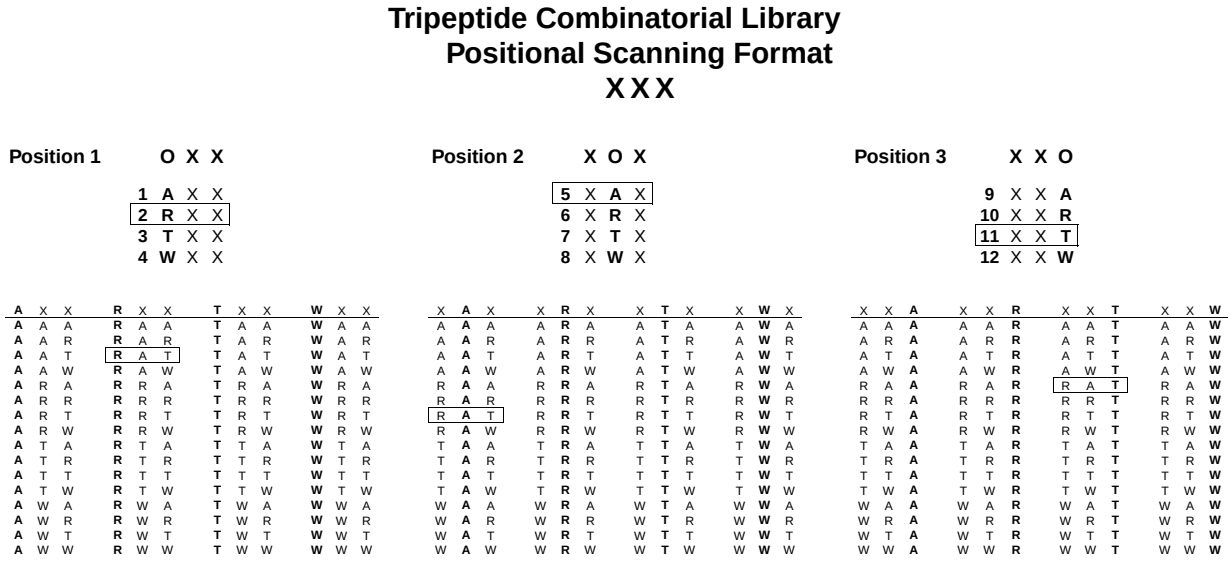

Figure 1 Tripeptide combinatorial library—positional scanning format.

in the chemical character of the building block (to reduce the number of individual compounds to be made). Freier and coworkers have presented an excellent discussion of the theoretical and experimental aspects of iterative and positional scanning deconvolution (34, 35).

The use of mixtures of compounds in a positional scanning combinatorial library format versus the use of individual compound arrays is clearly cost beneficial, thus allowing biotechnology companies, universities, and research institutes to carry out basic research and the initial stages of the drug-discovery process. Thus, previously unimagined numbers of compounds can be prepared and screened to yield highly active individual compounds.

The author's laboratory has successfully identified uniquely active and selective peptides from libraries made up of L-, D-, and unnatural amino acids totaling 6.25 million tetrapeptides (36), 52 million hexapeptides (27), and 6 trillion decapeptides (37). In practical terms, we believe that the cost and time savings make the use of mixtures in the positional scanning format a powerful alternative to the use of large individual compound arrays. A wide variety of classic heterocycles and other acyclic organic compounds can now be prepared using solid-phase parallel synthesis. The author's laboratory has found that existing libraries can be utilized to generate a diverse range of mixture-based heterocyclic libraries through the transformation of peptide and peptidomimetic libraries using the ''libraries from libraries'' approach (38). Such heterocyclic compounds have dramatically different physical and biological properties from the peptide libraries used as starting materials. Figures 2 and 3 illustrate a number of the transformations we have successfully carried out. The average library generated for the compound classes shown contains more than 50,000 compounds. We have recently generated a variation of the bicyclic guanidine library shown in Figure 2 (26), which is composed of more than 1.2 million bicyclic guanidines. It should be noted that each of the compounds illustrated in Figure 2 can be further transformed into other pharmacophores.

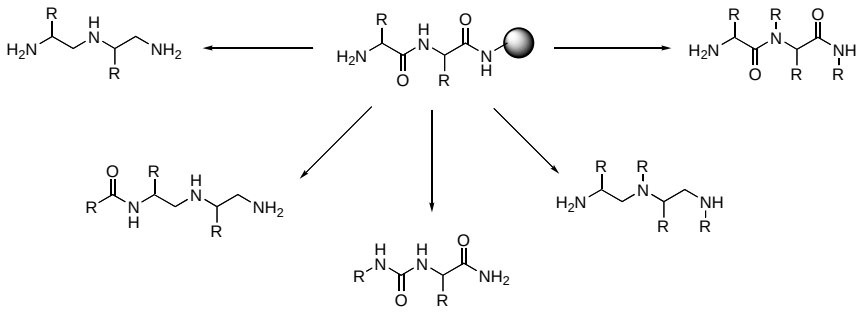

Figure 2 Solid-phase synthesis of acyclic compounds using dipeptides as starting material.

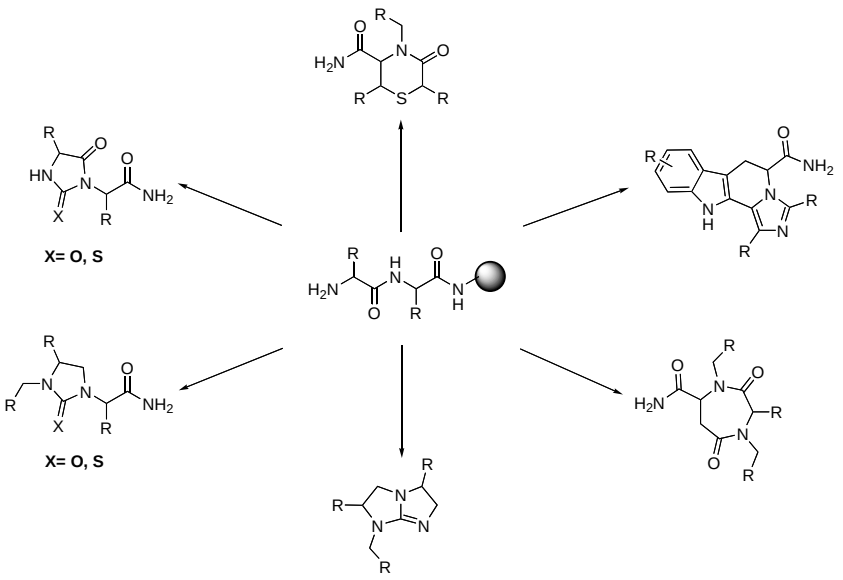

Figure 3 Solid-phase synthesis of heterocyclic compounds using dipeptides and acylated dipeptides as starting material.

CONCLUSIONS

Extensive studies on the use of mixture-based combinatorial libraries carried out by this laboratory (reviewed in 32) and others have enabled the rapid, cost-effective identification of highly active and specific individual compounds. These methods are extremely effective and broadly applicable, and have shown that there is nothing inherently unique about peptides or other oligomers that permits their successful use in mixture-based library formats as compared to heterocycles. Thus, the use of massive parallel synthesis in conjunction with classic high throughput screening methods versus the use of mixture-based combinatorial libraries must be tied to the balance between the need for complete data acquisition versus the pragmatic and rapid gathering of compound information for lead development. The use of extremely large mixture-based combinatorial libraries offers unique advantages that are simply not possible with other approaches.

Combinatorial chemistry, in all its manifestations over the past 15 years, has fundamentally changed synthetic chemistry in all areas of basic research and drug discovery. When coupled with methods such as computer-assisted design and molecular biology, combinatorial chemistry can be expected to enhance and continually increase the speed and thoroughness of drug discovery into the next millennium.

ACKNOWLEDGMENTS

The author's work is based on long-term collaborations with a number of scientific colleagues over the past 15 years. He would like to thank: John Ostresh for the development of a number of synthetic approaches, including ''libraries from libraries''; Clemencia Pinilla and Jon Appel for their work in developing the positional scanning deconvolution approach and the use of combinatorial libraries in immunological systems; Sylvie Blondelle for her work in identifying antibacterial and other highly active compounds from libraries; Colette Dooley for the development of radioreceptor screening assays and the identification of highly active opioid compounds; Jutta Eichler for the development of cyclic libraries; Barbara Dörner for the development of peralkylated peptidomimetic libraries; Adel Nefzi for heterocyclic chemistry; and Darcy Wilson for his T-cell immunology expertise.

LITERATURE CITED

1. Geysen HM, Meloen RH, Barteling SJ. 1984. Use of a peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid. *Proc. Natl. Acad. Sci. USA* 81:3998–4002
2. Houghten RA. 1985. General method for the rapid solid-phase synthesis of large numbers of peptides: specificity of antigen-antibody interaction at the level of individual amino acids. *Proc. Natl. Acad. Sci. USA* 82:5131–35
3. Merrifield RB. 1963. Peptide synthesis. I. The synthesis of a tetrapeptide. *J. Am. Chem. Soc.* 85:2149–54
4. Wong JY, Leznoff CC. 1973. The use of polymer supports in organic synthesis. II. The syntheses of monoethers of symmetrical diols. *Can. J. Chem.* 51:2452–56
5. Leznoff CC, Wong JY. 1973. The use of polymer supports in organic synthesis. III. Selective chemical reactions on one aldehyde group of symmetrical dialdehydes. *Can. J. Chem.* 51:3756–64
6. Crowley JI, Rapoport H. 1976. Solid-phase organic synthesis: novelty or fundamental concept? *Acc. Chem. Res.* 9:135–44
7. Bunin BA, Ellman JA. 1992. A general and expedient method for the solid-phase synthesis of 1,4-benzodiazepine derivatives. *J. Am. Chem. Soc.* 114:10997–98
8. Cuervo JH, Rodriguez B, Houghten RA. 1988. The magainins: sequence factors relevant to increased antimicrobial activity and decreased hemolytic activity. *Pept. Res.* 1:81–86
9. Suto MJ, Girten BE, Houghten RA, Loullis CC, Tuttle RR. 1995. Cytokine restraining agents. *U.S. Patent No. 5,420,109*
10. Pinilla C, Appel JR, Blondelle SE, Dooley CT, Eichler J, et al. 1994. Versatility of positional scanning of synthetic combinatorial libraries for the identification of individual compounds. *Drug Dev. Res.* 33:133–45
11. Eichler J, Appel JR, Blondelle SE, Dooley CT, Dörner B, et al. 1995. Peptide, peptidomimetic and organic synthetic combinatorial libraries. *Med. Res. Rev.* 15:481–96

12. Merrifield RB. 1986. Solid phase synthesis. *Science* 232:341–47
13. Stewart JM, Young JD. 1984. *Solid Phase Peptide Synthesis.* Rockford, IL: Pierce Chem. Co. 2nd ed.
14. Atherton E, Sheppard RC. 1989. *Solid Phase Peptide Synthesis—A Practical Approach.* Oxford, UK: IRL Press
15. Giannis A, Kolter T. 1993. Peptidomimetics for receptor ligands—discovery, development, and medical perspectives. *Angew. Chem. Int. Ed. Engl.* 32:1244–67
16. Liskamp RMJ. 1994. A new application of modified peptides and peptidomimetics: potential anticancer agents. *Angew. Chem. Int. Ed. Engl.* 33:305–7
17. Dörner B, Husar GM, Ostresh JM, Houghten RA. 1996. The synthesis of peptidomimetic combinatorial libraries through successive amide alkylation. *Bioorg. Med. Chem.* 4:709–15
18. Gersuk VH, Rose TM, Todaro GJ. 1995. Molecular cloning and chromosomal localization of a pseudogene related to the human acyl-CoA binding protein/diazepam binding inhibitor. *Genomics* 25:469–76
19. Ostresh JM, Blondelle SE, Dörner B, Houghten RA. 1996. Generation and use of nonsupport-bound peptide and peptidomimetic combinatorial libraries. *Methods Enzymol.* 267:220–34
20. Smythe ML, Huston SE, Marshall GR. 1993. Free energy profile of a 3_{10}- to α-helical transition of an oligopeptide in various solvents. *J. Am. Chem. Soc.* 115:11594–95
21. Fruchtel JS, Jung G. 1996. Organic chemistry on solid supports. *Angew. Chem. Int. Ed. Engl.* 35:17–42
22. Douglas SP, Whitfield DM, Krepinsky JJ. 1995. Polymer supported solution synthesis of oligosaccharides using a novel versatile linker for the synthesis of D-mannopentose, a structural unit of D-mannose of pathogenic yeasts. *J. Am. Chem. Soc.* 117:2116–17
23. Schuster M, Wang P, Paulson JC, Wong C. 1994. Solid-phase chemical-enzymatic synthesis of glycopeptides and oligosaccharides. *J. Am. Chem. Soc.* 116:1135–36
24. Frecht JM, Schuerch C. 1971. Solid phase synthesis of oligosaccharides. I. Preparation of the solid support. poly[p-(1-propen-3-ol-1-yl)styrene]. *J. Am. Chem. Soc.* 93:492–96
25. Nefzi A, Ostresh JM, Houghten RA. 1997. The current status of heterocyclic combinatorial libraries. *Chem. Rev.* 97:449–72
26. Ostresh JM, Schoner C, Hamashin VT, Nefzi A, Meyer JP, et al. 1998. The solid phase synthesis of tri-substituted bicyclic guanidines via cyclization of reduced N-acylated dipeptides. *J. Org. Chem.* 63:8622–23
27. Houghten RA, Pinilla C, Blondelle SE, Appel JR, Dooley CT, et al. 1991. Generation and use of synthetic peptide combinatorial libraries for basic research and drug discovery. *Nature* 354:84–86
28. Lam KS, Salmon SE, Hersh EM, Hruby VJ, Kazmierski WM. 1991. A new type of synthetic peptide library for identifying ligand binding activity. *Nature* 354:82–84
29. Furka A, Sebestyen F, Asgedom M, Dibo G. 1991. General method for rapid synthesis of multicomponent peptide mixtures. *Int. J. Pept. Protein Res.* 37:487–93
30. Geysen HM, Rodda SJ, Mason TJ. 1986. A priori delineation of a peptide which mimics a discontinuous antigenic determinant. *Mol. Immunol.* 23:709–15
31. Ostresh JM, Winkle JH, Hamashin VT, Houghten RA. 1994. Peptide libraries: determination of relative reaction rates of amino acids in competitive couplings. *Biopolymers* 34:1681–89
32. Houghten RA, Pinilla C, Appel JR, Blondelle SE, Dooley CT, et al. 1999. Mixture-based synthetic combinatorial libraries. *J. Med. Chem.* 42:3743–78
33. Pinilla C, Appel JR, Blanc P, Houghten

RA. 1992. Rapid identification of high affinity peptide ligands using positional scanning synthetic peptide combinatorial libraries. *Biotechniques* 13:901–5

34. Konings DAM, Wyatt JR, Ecker DJ, Freier SM. 1996. Deconvolution of combinatorial libraries for drug discovery: theoretical comparison of pooling strategies. *J. Med. Chem.* 39:2710–19
35. Wilson-Lingardo L, Davis PW, Ecker DJ, Hebert N, Acevedo O, et al. 1996. Deconvolution of combinatorial libraries for drug discovery: experimental comparison of pooling strategies. *J. Med. Chem.* 39:2720–26
36. Dooley CT, Ny P, Bidlack JM, Houghten RA. 1998. Selective ligands for the mu, delta and kappa opioid receptors identified from a single tetrapeptide positional scanning combinatorial library. *J. Biol. Chem.* 273:18848–56
37. Hemmer B, Pinilla C, Appel J, Pascal J, Houghten RH, et al. 1998. The use of soluble synthetic peptide combinatorial libraries to determine antigen recognition of T cells. *J. Pept. Res.* 52:338–45
38. Ostresh JM, Husar GM, Blondelle SE, Dörner B, Weber PA, et al. 1994. Libraries from libraries: chemical transformation of combinatorial libraries to extend the range and repertoire of chemical diversity. *Proc. Natl. Acad. Sci. USA* 91:11138–42

Annual Review of Pharmacology and Toxicology
Volume 40, 2000

CONTENTS