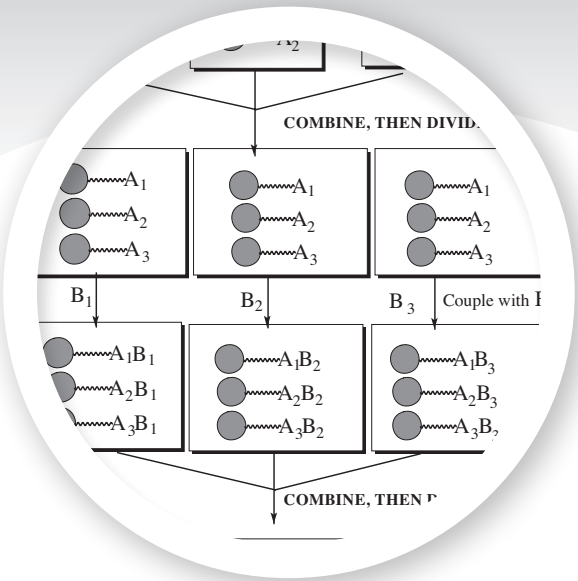
Using Stable Isotopes to Solve Problems in Combinatorial Chemistry

Dr. John Aberhart, PhD, Senior Staff Chemist
Cambridge Isotope Laboratories, Inc.
50 Frontage Road, Andover, MA 01810 USA



Virtually unheard of five years ago, combinatorial chemistry has grown to become today one of the hottest, fastest-moving areas of synthetic chemistry and biochemistry. Prompted by a continuing need to identify lead compounds for use as medicines, developments in combinatorial chemistry now permit the rapid generation of potentially vast numbers of compounds that can, in principle, be screened for interesting biological activity.[1-3]

Traditionally, searching for lead drug candidates was done by screening existing synthetic compound collections or natural materials (plants, microbial extracts, etc.). In attempts to optimize structure/activity relations or minimize side effects, promising lead compounds were (and continue to be) modified one by one, guided (where possible) by computer-aided design based on the fitting of drug candidate structures to the three-dimensional structures of their biological targets such as enzymes, receptors, etc. These 3D structures are, in many cases, determined by using X-ray crystallography or by using multidimensional NMR (a technique greatly aided by the application of stable-isotope labeling). Although the traditional approach is labor intensive, time consuming and very expensive, an impressive arsenal of pharmaceutical agents presently available to medical science attests to its success over the course of the past 50 years or so, and it is unlikely that these methods will be totally abandoned any time soon.

Nevertheless, as a result of needs for more efficiency in creating compounds for screening, virtually all pharmaceutical companies and many academic laboratories are now using combinatorial chemistry as a tool to supplement more traditional methods of drug discovery. In part, this increased need for compound creation results from advances in screening techniques that permit automatic, simultaneous screening of thousands of candidates.

Many reviews have been written on combinatorial chemistry,[4-10] and it is our intention in this essay to focus briefly on ways in which stable isotopes have been used as an adjunct to this technique. To our knowledge, this aspect of combinatorial chemistry has not previously been summarized.

In combinatorial chemistry, large arrays ("libraries") of structurally related compounds are synthesized by sequentially coupling collections of related components. Two basic methods are commonly used to generate combinatorial libraries: **parallel synthesis** and **split synthesis**.

In **parallel synthesis**, test compounds are synthesized simultaneously, step by step, in separate containers (often a microtitre plate containing small wells). Parallel synthesis can be done either in solution or on a solid support (which may simplify isolation of the synthesized library components at the end of the synthesis). Use of microtitre plates (with plate reading by microplate reading devices) facilitates the screening of synthesis libraries for promising lead compounds. The identity of synthesized compounds can generally be deduced from the location of their well (or other container), assuming that straightforward, well-established reactions are used and no unexpected chemistry occurs. (The relative advantages and disadvantages of solution vs solid-phase parallel synthesis have been discussed elsewhere; see reviews.) Automatic compound-synthesizing equipment has been developed for carrying out parallel syntheses of combinatorial libraries. Computers are commonly used to keep track of synthesis histories and the results of post-synthesis screening assays.

In contrast, **split synthesis**, often termed "one-bead-one-compound" combinatorial synthesis,[11] is performed by using solid-phase techniques somewhat analogous to those used in modern solid-phase peptide or oligonucleotide synthesis. As illustrated below, in preliminary steps, various types of monometric starting units, A (A_1,A_2,A_3...) are connected to beads via an appropriate

(continued)

linker that is a critical part of the synthetic apparatus. Each bead is connected to only one type of monomeric starting material. The linker must be compatible with subsequent chemistry to be used, compatible with screening, allow disconnection of the final test compounds from the bead and/or allow coding that permits tracking of the synthesis. As discussed later, linkers incorporating stable isotopes can be the key to deciphering the identity of library compounds.

As the synthesis proceeds, beads from preceding steps are *combined* and then *divided* for reaction with the next group of monomeric units, i.e. B_1,B_2,B_3...or C_1,C_2,C_3...etc.

At the conclusion of the above-depicted scheme, the result would be a combinatorial library consisting of 27 *unique*, polymer bound trimers, ABC.

By using the split-synthesis technique, the total number of coupling steps required to prepare a library is greatly reduced, as compared with parallel synthesis. At the end of two coupling steps, starting with three unique (but chemically related), bead-bound monomers A and three unique forms of monomers B and C (which would also normally be chemically related), the combinatorial library consists of 27 bead-bound trimers, each bead having bound to it only one unique trimer (of course, each bead would contain multiple copies of this trimer). (This does not count steps required to prepare the initial polymer-bound monomers, or mechanical combination/division steps.)

At the completion of the split combinatorial synthesis, beads would be separated and then screening performed, either with test compounds attached or after disconnection is performed. As a result of the *combined/divided* strategy of split combinatorial synthesis, the structures of test compounds *cannot be deduced* from their location in a synthesis apparatus or microtitre plate, and *some means of identification* must be used to assign structures (one would not need to determine the structures of all compounds in the library, only of promising lead compounds). For such purposes, various coding or ragging schemes have often been used. These will be discussed later.

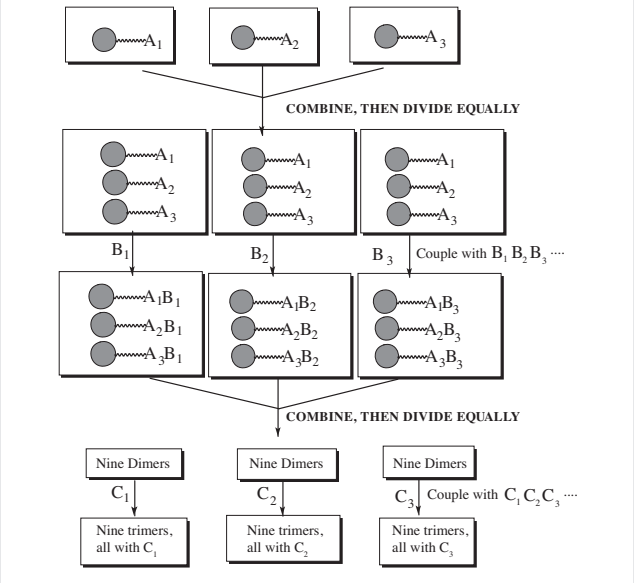


Figure 1.

Application of Stable Isotopes in Preparing Combinatorial Libraries

Typically, reactions carried out on solid-phase supports are driven to completion by using excess reagents. This works very well in the well-established methods for the syntheses of peptides and oligonucleotides. However, for other solid-phase synthetic chemistry, particularly when applied to small molecule synthesis, there is often a need to follow reactions (at least in the early development of a combinatorial synthesis strategy). Several approaches to this end, using stable-isotope labeling, have been recently published.

A group at Affymax Research Institute[12] has reported using ^{13}C-enriched compounds to monitor solid-phase preparations, using a technique they call "Fast ^{13}C NMR." In essence, this technique involves monitoring reactions taking place on resin samples, placed in NMR inserts, by using ^{13}C NMR. For example, formation of 4-thiazolidinone was followed on resin using resin-bound glycine (2-^{13}C) (Figure 2).

The signal enhancement from using ^{13}C-enriched substrates allowed them to obtain ^{13}C NMR spectra with excellent S/N with very few

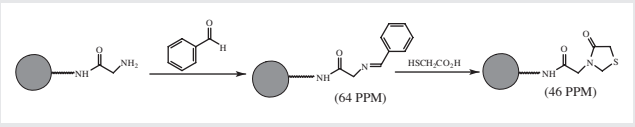


Figure 2.

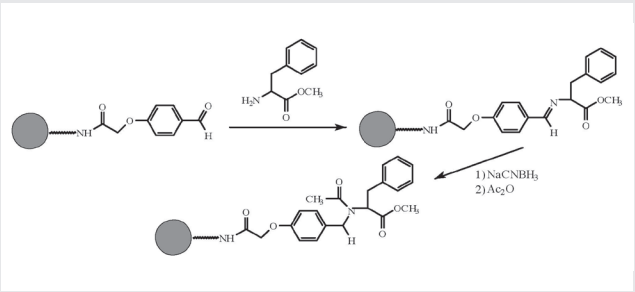


Figure 3.

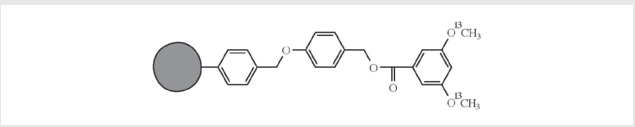


Figure 4.

Figure 5.

transients. Similarly, a stepwise reductive amination was followed by a triple-labeling scheme (Figure 3).

Sarkar and coworkers[13] at SmithKline Beecham have also shown ways in which the use of ^{13}C labeling of precursor molecules can be advantageous in monitoring the progress of solid-phase syntheses. By using magic-angle spinning, they were able to minimize linebroadening due to magnetic susceptibility that normally might be expected to be a problem with a heterogeneous sample (although the Affymax results did not show this problem). Combined with "Nano.NMR probe" techniques[14,15] and by using isotope-filtered NMR, Sarkar's group was able to selectively observe the ^{1}H resonances of the methoxy groups of a 3,5-dimethoxy-benzoate bound to a single bead of Wang resin (100 mm diameter).

A group at Sandoz[16] and Schreiber and coworkers[17] have also reported the use of magic-angle spinning techniques for analysis of resinbound samples, and clearly the incorporation of stable-isotope labeling would greatly enhance the sensitivity of these applications. Opella's group,[18] and others[19-22] have also recently reported NMR techniques for the study of resin-bound peptide structures that may have other applications in combinatorial chemistry. Extension of these techniques to more extensively labeled (e.g. uniformly ^{13}C labeled) molecules could have considerable utility in elucidating the structures of members of combinatorial libraries without detaching from the beads or in monitoring and controlling combinatorial synthesis reactions.

Russell and coworkers at Zeneca Pharmaceuticals[23-24] have used IR spectroscopy in combination with deuterium labeling to monitor solid-phase reactions. Although, in general, IR spectra of complex molecules are themselves quite complex, this group has taken advantage of the nearly unique spectral region (from 2300-2200 cm-^{1}) in which carbon-deuterium stretching frequencies appear. Unlabeled molecules, with few exceptions, would have no interfering absorbance in this region. By using benzoyl chloride (D_5), they were able to follow the benzoylation of aminomethyl polystyrene resin. Similarly, the progress of synthesis of a short peptide could be monitored by IR, using a chain extender L-Lysine, Nα-Fmoc,Nε-t-BOC labeled in the t-BOC sidechain protecting group with t-BOC (D_9), and following the increase in intensity of the C-D absorbance. In samples with multiple protected lysines, the number of lysines could be determined by measuring the intensity of the C-D absorbance at completion of the synthesis. Considering the wide range of deuterium-labeled compounds available from CIL, this technique clearly has potential for use in following solid-state reactions.

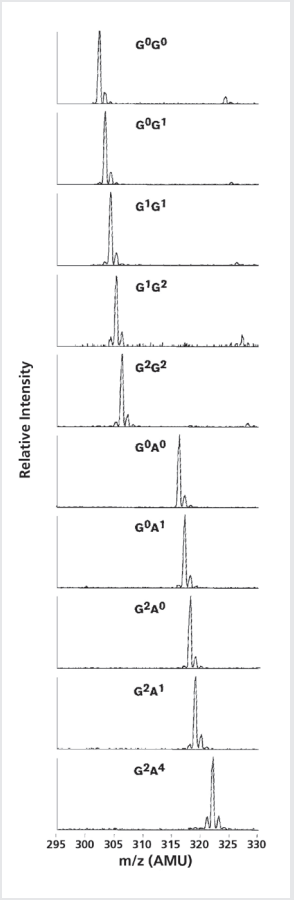


Figure 6.

While the above spectral techniques (with or without isotopic labeling) have some application in elucidating the structures of compounds attached to beads in combinatorial chemistry, in most cases, the structures and spectra will be too complicated or there will not be enough resin-bound material to assign structures.

Encoding of Combinatorial Libraries

To allow assignment of structures of attached compounds in some bead-one compound combinatorial chemistry, various tagging or coding strategies have been invented. The first of these, by Lerner and Brenner,[25] and developed independently by an Affymax group,[26] entailed synthesis of an oligonucleotide chain in parallel with the synthesis of library compounds (e.g. peptides). Thus, the bead would have two independently growing chains, one containing the peptide (or other library member) and the other,

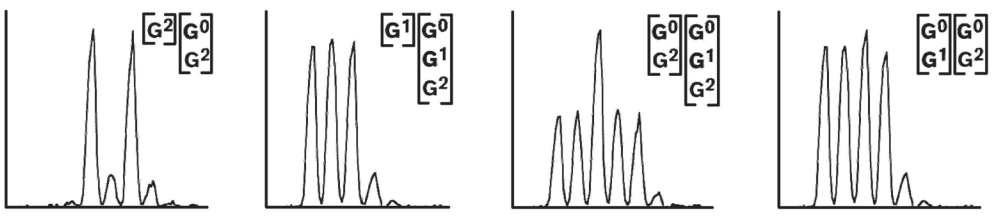


Figure 7.

(continued)

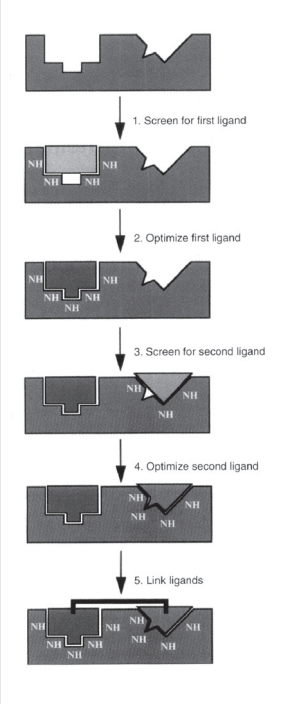


Figure 8.

synthesized in alternating steps, an oligonucleotide chain. By elucidating the oligonucleotide sequence using established ladder techniques, one could deduce the identity of units added to the other chain. Many other encoding/tagging schemes have been devised and reviewed elsewhere, and in this review, we focus on only one, recently reported by the Glaxo group,[27] which uses stable-isotope labeling.

In Geysen's work, a stable isotope-labeled coding element is incorporated into the linker which binds library components to the polymer bead (Figure 5). The code is read by mass spectrometric techniques, after cleavage of the linker from the bead. Alternatively, NMR can be used, in some cases, to read the code while still connected to the bead. The approach does not require adding a coding element at each stage of the synthesis, as long as the synthesis involves only two addition steps beyond connection of the starting unit to the linker.

For example, in their technique termed "single peak positional encoding," a code block, consisting of a dipeptide comprised of various stable isotope labeled and unlabeled forms of glycine and alanine, is used. In this scheme, only the total mass of the dipeptide, rather than labeling position (or even isotope used) would be important. In their published illustration of the technique, not all conceivable labeled forms of either amino acid were used, only readily available ^{13}C,^{15}N-labeled forms, e.g. alanine ($^{13}C_1$) and ($^{13}C_3$,^{15}N) but not ($^{13}C_3$) or ($^{13}C_2$). Also only gly-gly or gly-ala dipeptides were used as illustrations. Thus there is scope for expansion of the number of coding elements, e.g. by adding deuterium labeling and multielement labeling, or by adding other amino acids, e.g. leucine. Even so, ten coding elements were shown (see Figure 6).

To encode the library, each unique dipeptide is used to encode a unique monomer A_1,A_2.... After the usual type of split synthesis as outlined in Figure 1, the identity of the third monomer C_1,C_2.... is defined by the pool from which a particular bead is obtained. After the cleavage of link 2 and activity testing of the library, link 1 of promising leads is cleaved, followed by MS analysis of the dipeptide. This allows identification of monomer A. The total mass of the hit compound AxBxCx, also determined by MS, allows assignment of the identity of monomer B, provided that all forms of B used in the combinatorial synthesis have different masses. By using only the ten dipeptides shown in Figure 6, a fully encoded library of 1000 products can, in principle, be produced and decoded. In a variant of the above coding scheme, use is made of mixed isotopically distinct code monomers at each step of introduction of the coding element. For a code consisting solely of the dipeptide gly-gly, by using various combinations of only unlabeled or $^{13}C_1$ or $^{13}C_2$ glycine, 25 distinct mass spectral patterns (a type of "bar code") are possible (several examples are shown in Figure 6). By varying the ratios of labeled glycines used in the bar code (in 9% increments), as many as 3,750 distinct bar codes could be created. Other coding scenarios, also decodable by MS techniques, were presented.

The Glaxo group also described an isotope-encoding technique that could be decoded by NMR techniques. In the example presented, varying ratios of ^{13}C-labeled and unlabeled acetate were incorporated into their linker. The ratio code was analyzed by comparing the peak height of the enriched acetate with residual ^{13}C in the unlabeled methylene units of the polyethylene spacer of the resin. Obviously a wide range of other labeling options would be possible, as long as the tagging compound exhibited unique ^{13}C chemical shifts and sufficient taggant was used for obtaining a spectrum. This technique is related to the work of Look,[12] Sarkar,[13] and others described above, but with the labeling being in the linker rather than in the library compound itself. A group at SmithKline Beecham[27c] has also used NMR methods for decoding mass-encoded combinatorial libraries.

The mass spectrometric technique would seem to be more capable of rapid screening of large libraries. In any case, it is clear that isotopic encoding of combinatorial libraries has much, as yet underutilized, potential for use as an alternative or adjunct to other coding methods now in use. CIL offers a wide range of ^{13}C,^{15}N, D-labeled materials potentially useful for such purposes.

SAR by NMR

Recently introduced by Fesik and coworkers[28-35] at Abbott Laboratories as a new technique for discovering ligands that bind strongly to proteins or other biomolecules of medical interest, this method is conceptually related to combinatorial chemistry, but differs substantially in practice.

The technique is a "linked fragment" approach to identifying high activity ligands, wherein ligands are synthesized from components that have been optimized for binding to protein subsites. Optimization is performed by observing ^{15}N- or ^{1}H-amide chemical shift perturbations in the ^{15}N-heteronuclear single-quantum correlation (^{15}N-HSQC) spectra of the ^{15}N-labeled protein upon addition of test ligands. Proteins can, in principle, be readily labeled by using modern molecular biological methods, involving biosynthesis in the presence of ^{15}N-labeled amino acids or labeled growth media such as CIL's BioExpress.® Screening of large numbers of test compounds (individually, or in groups) can be performed, e.g. Abbott's group reported screening 1,000 compounds per day in groups of ten.

After identifying lead compounds, analogs are screened to optimize binding to a *particular* binding site. Next, a different

ligand is identified that binds to a *neighboring* site by observing chemical shift changes for a *different* set of amide resonances, and optimization is again performed. After determining the location and orientation of bound ligands in the ternary complex by NMR or X-ray methods, test compounds in which both ligands are conjoined are synthesized. Since the binding affinity, expressed as Kb, of a linked compound is, in principle, essentially a product of the binding constants of the unlinked fragments, corrected for linking contributions, it was possible to generate compounds having Kbs in the low nanomolar range by linking fragments having micromolar binding affinities. This was demonstrated for ligands binding to the FK506 binding protein,[28] and by the discovery of high-activity inhibitors of stromelysin.[29,30] The approach is outlined in Figure 8.

The technique differs from more conventional combinatorial chemistry in that only a few (linked) compounds need to be synthesized, because most screening and optimization is done with (presumably) readily available, existing compound collections prior to linkage. These methods would, of course, be limited to water-soluble ligands and to isotopically labeled biomolecules of relatively low molecular weight (MW<30kD) that can be studied by high-resolution NMR and that can be labeled by available labeling methods. However, the Abbott paper is obviously a ground-breaking development that will see increased use for discovery of useful medicinal agents.

References

1. Baum, R.M. **1994**. Combinatorial Approaches Provide Fresh Leads for Medicinal Chemisry. *Chemical and Engineering News,* 20-26.
2. Borman, S. **1996**. Combinatorial Chemists Focus on Small Molecules, Molecular Recognition, and Automation. *Chem Eng News,* 29-73.
3. Borman, S. **1997**. Combinatorial Chemistry. *Chem Eng News,* 43-63.
4. Moos, W.H.; Green, G.D.; Pavia, M.R. **1993**. Recent Advances in the Generation of Molecular Diversity. *Annu Rep Med Chem, 28,* 315-324.
5. Choong, I.C.; Ellman, J.A. **1996**. Solid-Phase Synthesis: Application to Combinatorial Libraries. *Annu Rep Med Chem, 31,* 309-318.
6. Abelson, J.N. **1996**. Combinatorial Chemistry. *Methods Enzymol, 267* (complete volume).
7. Terrett, N.K.; Gardner, M.; Gordon, D.W.; Kobylecki, R.J.; Steele, J. **1995**. Combinatorial Synthesis – The Design of Compound Libraries and their Application to Drug Discovery. *Tetrahedron, 51,* 8135-8173.
8. Chin, J.; Fell, J.B.; Jarosinski, M.; Shapiro, M.J.; Wareing, J.R. **1998**. *J Org Chem, 63,* 386-390.
9. Czarnik, A.W.; Ellman, J.A. ed. **1996**. Acc Chem Res, 29, 112-170.
10. Ellman, J.A.; Stoddard, B.; Wells, J. **1997**. Combinatorial Thinking in Chemistry and Biology. *Proc Natl Acad Sci USA*, 94, 2779-2782.
11. Lebl, M.; Krchnak, V.; Serperov, N.E.; Seligmann, B.; Felder, S.; Lam, K.S. **1995**. One-Bead-One-Structure Combinatorial Libraries. *Biopolymers, 37,* 177-198.
12. Look, G.C.; Holmes, C.P.; Chinn, J.P.; Gallop, M.A. **1994**. Methods for Combinatorial Organic Synthesis: The Use of Fast ^{13}C NMR Analysis for Gel Phase Reaction Monitoring. *J Org Chem, 59,* 7588-7590.
13. Sarkar, S.K.; Garagipati, R.S.; Adams, J.L.; Keifer, P.A. **1996**. An NMR Method to Identify Nondestructively Chemical Compounds Bound to a Single Solid-Phase Bead for Combinatorial Chemistry Applications. *J Am Chem Soc, 118,* 2306.
14. Fitch, W.L.; Detre, G.; Holmes, C.P.; Shoolery, J.N.; Keifer, P.A. **1994**. High-Resolution ^{1}H NMR in Solid-Phase Organic Synthesis. *J Org Chem, 59,* 7955-7956.
15. Garigipati, R.S.; Adams, B.; Adams, J.L.; Sarkar, S.K. **1996**. Use of Spin Echo Magic Angle ^{1}H NMR in Reaction Monitoring in Combinatorial Organic Synthesis. *J Org Chem, 61,* 2911-2914.
16. Anderson, R.C.; Jarema, M.A.; Shapiro, M.J.; Stokes, M.J.; Ziliox, M. **1995**. Analytical Techniques in Combinatorial Chemistry: MASCH Correlation in Solvent-Swollen Resins. *J Org Chem, 60,* 2650-51.
17. Chen, J.K.; Schreiber, S.L. **1995**. Combinatorial Synthesis and Multidimensional NMR Spectroscopy: An Approach to Understanding Protein-Ligand Interactions. *Angew Chem, int. ed., 34,* 953-968.
18. Jelinek, R.; Valente, A.P.; Valentine, K.G.; Opella, S.J. **1997**. Two-dimensional NMR Spectroscopy of Peptides on Beads. *J Magn Reson, 125,* 185-187.
19. Boutin, J.A.; Hennig, P.; Lambert, P.H.; Bertin, S.; Petir, L.; Mahieu, J.P.; Serkiz, B.; Volland, J.P.; Fauchere, J.L. **1996**. Combinatorial Peptide Libraries: Robotic Synthesis and Analysis by Nuclear Magnetic Resonance, Mass Spectrometry, Tandem Mass Spectrometry, and High-Performance Capillary Electrophoresis Techniques. *Anal Biochem, 234,* 126-141.
20. Keifer, P.A. **1996**. Influence of Resin Structure, Tether Length, and Solvent upon the High-Resolution ^{1}H NMR Spectra of Solid-Phase Synthesis Resins. *J Org Chem, 61,* 1558-1559.
21. Wilson, S.R.; Czarnik, A.W. **1997**. ed. Combinatorial Chemistry: Synthesis and Application. Wiley-Interscience.
22. Anderson, R.C.; Stokes, J.P.; Shapiro, M.J. **1995**. Structure Determination in Combinatorial Chemistry: Utilization of Magic Angle Spinning HMQC and TOCSY NMR Spectra in the Structure Determination of Wang-bound Lysine. *Tetrahedron Lett,* 5311-5314.
23. Russell, K.; Cole, D.C.; McLaren, F.M.; Pivonka, D.E. **1996**. Analytical Techniques for Combinatorial Chemistry: Quantitative Infrared Spectroscopic Measurements of Deuterium-Labeled Protecting Groups. *J Am Chem Soc, 118,* 7941-7945.
24. Russell, K.; Cole, D.C.; McLaren, D.F.; Pivonka, D.E. **1996**. Tools for Combinatorial Chemistry. Libr. Drug Discovery Conference, I.M. Chaiken and K.D. Janda, ed., *Am Chem Soc,* 246-254.
25. Brenner, S.; Lerner, R.A. **1992**. Encoded Combinatorial Chemistry. *Proc Natl Acad Sci, USA, 89,* 5381-5383.
26. Needles, M.C.; Jones, D.G.; Tate, E.H.; Heinkel, G.L.; Kochersperger, L.M.; Dower, W.J.; Barrett, R.W.; Gallop, M.A. **1993**. *ibid, 90,* 10700-10704.
27. Geysen, H.M.; Wagner, C.D.; Bodnar, W.M.; Mackworth, C.J.; Parke, G.J.; Schoenen, F.J.; Wagner, D.S.; Kinder, D.S. **1996**. Isotope or Mass Encoding of Combinatorial Libraries. *Chem Biol, 3,* 679-688.

27a. Wagner, D.S.; Markworth, C.J.; Wagner, C.D.; Schoenen, F.J.; Rewerts, C.E.; Kay, B.K.; Geysen, H.M. **1998**. Ratio Encoding Combinatorial Libraries with Stable Isotopes and their Utility in Pharmaceutical Research. *Comb Chem High Throughput Screening, 1,* 143-153.

27b. Geysen, H.M.; Kinder, D.S.; Wagner, C.D. Mass-Based Encoding and Qualitative Analysis of Combinatorial Libraries. World patent, WO 97/37953.

27c. Garigipati, R.S.; Sarkar, S.K. Mass Encoded Combinatorial Libraries. World patent, WO/9714814.

28. Shuker, S.B.; Hajduk, P.J.; Meadows, R.P.; Fesik, S.W. **1996**. *Science, 274,* 1531-1534.
29. Hajduk, P.J.; Sheppard, G.; Nettesheim, D.G.; Olejniczak, E.T.; Shuker, S.B.; Meadows, R.P.; Steinman, D.H.; Carrera, D.M.; Marcotte, P.A.; Severin, J.; Walter, K.; Smith, H.; Gubbins, E.; Simmer, R.; Holzman, T.F.; Morgan, D.W.; Davidsen, S.K.; Summers, J.B.; Fesik, S.W. **1997**. Discovery of Potent Nonpeptide Inhibitors of Stromelysin Using SAR by NMR. *J Am Chem Soc, 119,* 5818-5827.
30. Olejniczak, E.T.; Hajduk, P.J.; Marcotte, P.A.; Nettesheim, D.G.; Meadows, R.P.; Edalji, R.; Holtzman, T.F.; Fesik, S.W. **1997**. Stromelysin Inhihbitors Designed from Weakly Bound Fragments: Effects of Linking and Cooperativity. *J Amer Chem Soc, 119,* 5828-5832.
31. Hajduk, P.J.; Gerfin, T.; Boehlen, J.M.; Haberli, M.; Marek, D.; Fesik, S.W. **1999**. High-Throughput Nuclear Magnetic Resonance-Based Screening. *J Med Chem, 42,* 2315-2317.
32. Hajduk, P.J.; Dinges, J.; Miknis, G.F.; Merlock, M.; Middleton, T.; Kempt, D.J.; Egan, D.A.; Walter, K.A.; Robins, T.S.; Shuker, S.B.; Holzman, T.F.; Fesik, S.W. **1997**. NMR-Based Discovery of Lead Inhibitors That Block DNA Binding of the Human Papillomavirus E2 Protein. *J Med Chem, 40,* 3144--3150.
33. Fesik, S.W.; Hajduk, P.J. . Use of Nuclear Magnetic Resonance to Identify Ligands to Target Biomolecules. World patent, WO97/18471.
34. Fesik, S.W.; Hajduk, P.J. **1998**. Use of Nuclear Magnetic Resonance to Identify Ligands to Target Biomolecules. US patents, no. 5,698,401 1997; no. 5,804,390.
35. Fesik, S.W.; Hajduk, P.J.; Olejniczak, E.T. Use of Nuclear Magnetic Resonance to Design Ligands to Target Biomolecules. US patents no. 5,891,643 (1999) and 5,989,827 (1999).