Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 11422–11426, November 1994
Chemistry

Recursive deconvolution of combinatorial chemical libraries

ERIC ERB*†, KIM D. JANDA*‡, AND SYDNEY BRENNER§

*Departments of Molecular Biology and Chemistry, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, CA 92037; and §Department of Medicine, University of Cambridge, CB2 2QQ, United Kingdom

Contributed by Sydney Brenner, August 3, 1994

ABSTRACT **A recursive strategy that solves for the active members of a chemical library is presented. A pentapeptide library with an alphabet of Gly, Leu, Phe, and Tyr (1024 members) was constructed on a solid support by the method of split synthesis. One member of this library (NH_2-Tyr-Gly-Gly-Phe-Leu) is a native binder to a β-endorphin antibody. A variation of the split synthesis approach is used to build the combinatorial library. In four vials, a member of the library's alphabet is coupled to a solid support. After each coupling, a portion of the resin from each of the four reaction vials was set aside and catalogued. The solid support from each vial is then combined, mixed, and redivided. The steps of (*i*) coupling, (*ii*) saving and cataloging, and (*iii*) randomizing were repeated until a pentapeptide library was obtained. The four pentapeptide libraries where the N-terminal amino acid is defined were screened against the β-endorphin antibody and quantitated via an ELISA. The amino acid of the four pools that demonstrated the most binding was then coupled to the four tetrapeptide partial libraries that had been set aside and catalogued during the split synthesis. This recursive deconvolution was repeated until the best binders were deduced. Besides the anticipated native binder, two other members of the library displayed significant binding. This recursive method of deconvolution does not use a molecular tag, requires only one split synthesis, and can be applied to the deconvolution of nonlinear small-molecule combinatorial libraries and linear oligomeric combinatorial libraries, since it is based only on the procedure of the synthesis.**

There is increasing interest in synthesizing large numbers of molecules in parallel and in analyzing these pools for members with biological activity. So called "irrational" drug design, involving selection from combinatorial libraries, is becoming accepted as a useful method of finding pharmacologically active compounds.

The main difficulty with this approach is a way of finding the compounds with defined activity, especially when the libraries used are large. Peptides and oligonucleotides bound to an immobilized receptor can be eluted and directly sequenced (1–3). Although oligonucleotides can be amplified by PCR, peptides may require several runs to obtain sufficient material for analysis. Alternatively, peptides, synthesized on beads, have been identified by isolating beads that have bound a receptor and then sequencing the released peptide. Peptides have also been identified by synthesizing them in arrays or on small surfaces (4, 5).

These methods are restricted by the chemistries involved in the synthesis or the analysis. An extension that allows wider chemical diversity of the libraries is to "encode" the library in some way. Brenner and Lerner (6) proposed encoding each molecule of the library with an oligodeoxynucleotide, which could be used both for identification and for the enrichment of active members, and the chemistry for this has been implemented by Janda and coworkers (7). Others have encoded (8–11) beads, each carrying a single component of the chemical library, with peptides, oligonucleotides, and organohalide tags. The bead with the active component must be isolated and then the tag must be analyzed by mass spectrometry (8) or Edman degradation (9) for peptide analysis, by electron-capture gas chromatography for organohalide tags (10), or by PCR for deoxynucleotide tags (11).

Another method uses an iterated search process that consists of making the library in a number of pools, finding the active pool that defines the entity for that position in the molecule, and then repeating this until the active component has been identified. In the method proposed by Houghten *et al.* (12), hexapeptide libraries of 18 amino acids were synthesized as follows. By using split synthesis, four cycles of solid-phase peptide synthesis provided equimolar mixtures of 18^4 (104,976) tetrapeptides. The tetrapeptide-linked resin was then divided into 324 pools so that the synthesis of the next two positions could be defined to give the general formula $X_{[1]}X_{[2]}N_{[3]}N_{[4]}N_{[5]}N_{[6]}$, where X is a position to be defined, N is a randomized position, and the subscript indicates the position in the molecule. These were assayed and positive results for the first two residues ($A_{[1]}B_{[2]}$, say) were noted. Next, 18 new libraries were synthesized with the formula $A_{[1]}B_{[2]}X_{[3]}N_{[4]}N_{[5]}N_{[6]}$, one for each amino acid at position 3, and tested to define $X_{[3]}$. The process is repeated until all positions are defined. A similar iterative process called SURF (synthetic unrandomization of randomized fragments) was used by ISIS Pharmaceuticals (Carlsbad, CA) for an oligonucleotide library (13).

A virtue of the iterative method is that the multiplicity of components decreases with each step so that there is an enrichment process, and since molecules can be assayed in solution, it permits functional and binding assays. It would, of course, be possible to synthesize six libraries of general formula $X_{[1]}N_{[2]}N_{[3]}N_{[4]}N_{[5]}N_{[6]}$, $N_{[1]}X_{[2]}N_{[3]}N_{[4]}N_{[5]}N_{[6]}$, . . . , $N_{[1]}N_{[2]}N_{[3]}N_{[4]}N_{[5]}X_{[6]}$ with 18 members each, assay all 108 libraries, and define the residue in each position in the active molecule as suggested by Dooley and Houghten (14). However, this would not allow enrichment with progressive improvement of the signal-to-noise ratio.

Another iterative method that depends on mixed syntheses, that is, each $N_{[i]}$ step is achieved by adding a mixture of amino acids has also been proposed (15). Apart from the fact that it is difficult to know whether each amino acid is appended in the same yield in each stage, it is impossible to make libraries where the linking bond is itself a variable contributing to the chemical diversity. This must be carried out by the "split synthesis" method (16–18).

To implement a dictionary or bucket search method using split syntheses requires a different procedure in which the

Abbreviations: DMF, *N*,*N*-dimethylformamide; HBTU, 2-(1*H*-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate; Fmoc, 9-fluorenylmethoxycarbonyl; TG, TentaGel.

†Present address: CombiChem, Inc., 10975 Torreyana Road, Suite 230, San Diego, CA 92121.
‡To whom reprint requests should be addressed.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

libraries need to be made in a different way and that deconvolutes the structure of an active molecule by starting with the last component first and proceeding backwards. Here we present an implementation for this recursive method for analyzing combinatorial chemical libraries.

PRINCIPLE OF THE METHOD

Strategy for the Construction of the Combinatorial Library. The essence of our method is to build and hold a set of partially synthesized combinatorial libraries. The following example considers a library with a degree of three, made from an alphabet of four components, *A*, *B*, *C*, and *D* (Fig. 1). As in all split syntheses, we define four channels of synthesis, each channel involves only the addition of one member of the alphabet and each can be independently optimized. Step 1 consists of making four pools, in which *A*, *B*, *C*, and *D* may be added to a solid surface carrying a linker, for example. A portion of this library is set aside and labeled as partial library p{1}. Then the remaining material is combined and separated into four portions, each channel is loaded, and *A*, *B*, *C*, and *D* are added as before. Again, an aliquot of this library is set aside as partial library p{2}, which has four pools $N_{[1]}A_{[2]}$, $N_{[1]}B_{[2]}$, $N_{[1]}C_{[2]}$, and $N_{[1]}D_{[2]}$. The remainder is pooled and split, and the third step of addition is carried out to give the final library $N_{[1]}N_{[2]}A_{[3]}$, $N_{[1]}N_{[2]}B_{[3]}$, $N_{[1]}N_{[2]}C_{[3]}$, and $N_{[1]}N_{[2]}D_{[3]}$. More generally, we synthesize the partial combinatorial libraries, $X_{[1]}$, $N_{[1]}X_{[2]}$, $N_{[1]}N_{[2]}X_{[3]}$, . . . , $N_{[1]}N_{[2]}$, . . . , $N_{[n-1]}X_{[n]}$.

Recursive Deconvolution of the Library. Each of the four pools, $N_{[1]}N_{[2]}X_{[3]}$ (*X* is *A*, *B*, *C*, or *D*), which contain a total of 64 different molecules, are now tested by an appropriate assay and the active pool (if any) is determined (Fig. 2). Suppose $N_{[1]}N_{[2]}A_{[3]}$ from the final library is positive. Then, we proceed to partial library p{2} and add *A* to an aliquot of each of the four pools, to give four new libraries of general formula $N_{[1]}X_{[2]}A_{[3]}$. This library now contains a total of only 16 components, so 4-fold enrichment has been achieved. Again, after testing, suppose $N_{[1]}B_{[2]}A_{[3]}$ is active. We proceed to partial library p{1} and to each add *B*, followed by *A*, to give four pools with structure $X_{[1]}B_{[2]}A_{[3]}$, which can be tested to find $X_{[1]}$. Again, a 4-fold enrichment has been achieved. If there is more than one active component, this method will find both as the recursion can be pursued to completion down all tracks independently.

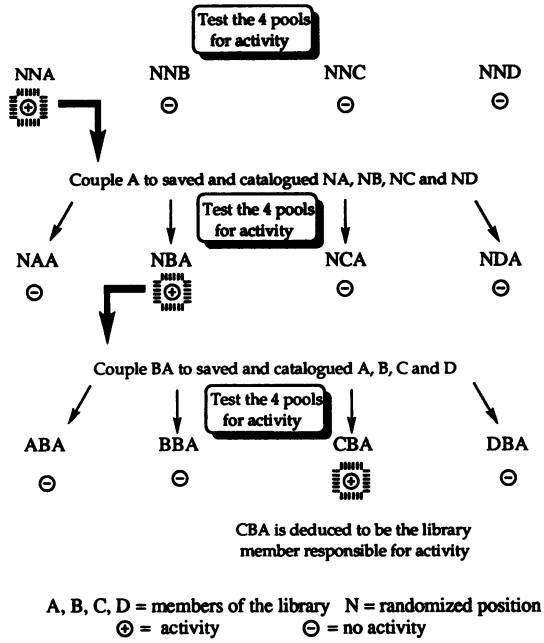
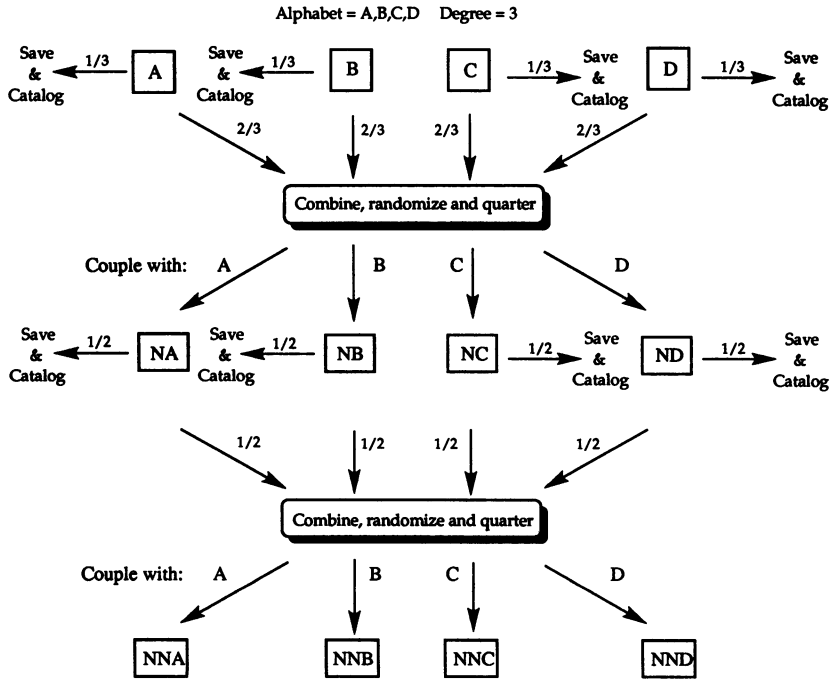


FIG. 2. Recursive deconvulatory pathway for solving the active member of a combinatorial library.

MATERIALS AND METHODS

Solvents. *N,N*-Dimethylformamide (DMF) (HPLC grade) was purchased from Baxter Scientific Products (McGaw Park, IL) and was used without further purification. CH_2Cl_2 was purchased from Fisher and was distilled over CaH_2 before use.

Instrumentation. UV absorbances were measured on a Hewlett–Packard 8452A diode array spectrophotometer. All



Each pool contains 16 molecules. There are a total of 64 unique molecules.

FIG. 1. Depiction of the split synthesis of a combinatorial library with an alphabet of four and a degree of three. After each step of the synthesis, a portion of each partial library is saved and catalogued before the randomization step. The fraction of each partial library to be saved and catalogued is the inverse of the degree of the library in the first step and the inverse of degree −1 for the subsequent steps.

centrifugations were carried out on an Eppendorf 5415 C centrifuge for 1 min at 10,000 rpm.

Formation of the Pentapeptide Library. The peptide libraries were synthesized manually by a solid-phase method using 90-μm TentaGel (TG) resin as the solid support. The amino acids used for the formation of the libraries were Gly-Fmoc, Leu-Fmoc, Phe-Fmoc, and Tyr(tBu)-Fmoc, where Fmoc is 9-fluorenylmethoxycarbonyl and tBu is *tert*-butyl. All amino acids were coupled with the aid of 2-(1*H*-benzotriazole-1-yl)-1,1,3,3-tetrauronium hexafluorophosphate (HBTU) and diisopropylethylamine (DIPEA). The TG, HBTU, DIPEA, and amino acids used were purchased from Nova Biochem. TG was added to four glass-fritted filter vials, shaken with DMF, and filtered. To each vial was added one of the Fmoc-protected amino acid components of the library, HBTU, DMF, and 20 μl of DIPEA. All four vials were shaken for 1 h, filtered, and washed with DMF. A 5% (vol/vol) solution of acetic anhydride in DMF was added to each vial and shaken for 20 min to cap any uncoupled free amino acid groups. The beads were then filtered, washed with DMF twice and CH_2Cl_2 twice, and dried for 2 h in a vacuum oven (55°C at 20 mmHg; 1 mmHg = 133 Pa). About 75 mg of the amino acid-coupled TG beads was removed from each vial and labeled in a manner describing the length of the peptide and the identity of the last amino acid coupled. At the same time, an Fmoc deprotection test was carried on 5 mg of the resin to measure the yield of the previous amino acid coupling step. The remaining resin was combined and shaken in DMF/CH_2Cl_2, 3:1 (vol/vol), for 30 min to mix the beads. The beads were divided into four equal portions by weight (each, therefore, containing an equivalent number of moles) and placed into four glass-fritted filter vials. Each was shaken twice for 10 min with 20% (vol/vol) piperidine in DMF to cleave the N-terminal Fmoc protecting group. The beads were filtered, washed with DMF three times, and coupled with one of the four amino acid components of the library to give the four dipeptide pools, Fmoc-Gly-Naa-TG, Fmoc-Leu-Naa-TG, Fmoc-Phe-Naa-TG, and Fmoc-Tyr(tBu)-Naa-TG. The steps of (*i*) capping the peptide with acetic anhydride; (*ii*) saving and labeling a 75-mg portion of the resin; (*iii*) combining, randomizing, and evenly dividing the beads; and (*iv*) coupling each division with an Fmoc-protected amino acid were repeated until the fifth coupling. At this point there are four pentapeptide libraries. The *t*-butyl protecting group on the Tyr residues was removed by mixing the resin with trifluoroacetic acid (Pierce) for 2 h, followed by filtering and washing with ethanol three times and DMF three times. The beads were then shaken with 20% piperidine in DMF for two 10-min periods to cleave the N-terminal Fmoc group. The deprotected beads were then washed with DMF three times and CH_2Cl_2 twice, and the residual solvent was removed in a vacuum oven overnight (55°C at 20 mmHg).

The efficiency of each coupling was measured by adding a TG resin sample of known mass from each library pool to 1000 μl of 20% piperidine in DMF, shaking the solution for 15 min, and measuring the UV absorbance of the resulting piperidine–benzofulvene complex at 302 nm. This absorbance was compared to that of a standard solution of Gly-Fmoc in 20% piperidine/DMF, which gives the loading capacity of the TG resin after each coupling. Each coupling of the Fmoc-protected amino acids to the TG with HBTU proceeded smoothly and quantitatively. The quantitativeness of the coupling is advantageous since we can assume that the loading capacity (and, therefore, the number of mol per g of resin) of the TG is approximately constant, no matter at what stage we are in the synthesis or deconvolution of the library.

ELISA Procedure. The library containing TG resin (≈3 mg) was added to a polypropylene tube previously coated with a 1:1 (vol/vol) blocking solution [3% (wt/vol) bovine serum albumin (BSA) in PBS and 0.05% Tween 20 (Sigma) in PBS]. The tube was incubated with 500 μl of a solution consisting of equal volumes of mouse anti-β-endorphin monoclonal antibody (Boehringer Mannheim) at 1 μg/ml in PBS, 3% BSA in PBS, and 0.05% Tween 20 in PBS at 37°C for 1 h. The tube was centrifuged and the supernatant was decanted. The tube was washed three times by adding 0.05% Tween 20 in PBS, centrifuging the tube, and decanting the supernatant. The tube was then incubated with 500 μl of goat anti-mouse antibody conjugated to glucose oxidase (from Cappel Laboratories diluted by a factor of 1000 with a 1:1 solution of 3% BSA in PBS and 0.05% Tween 20 in PBS). The contents of the tube were washed three times with 0.05% Tween 20 in PBS and two times with PBS: after each wash the tube was centrifugated, and the supernatant was decanted.

A 500-μl aliquot of a developing solution [25 ml of 0.1 M Na_3PO_4 (pH 6.0), 3 ml of 20% (wt/vol) glucose in H_2O, 200 μl of a 0.1% horseradish peroxidase in 0.1 M Na_3PO_4 (pH 6.0), and 200 μl of 2,2′-azinobis(3-ethylbenzothiazoline-6-sulfonic acid) dye (45 mg/ml) in 0.1 M Na_3PO_4 (pH 6.0)] was added to the tube to assay the H_2O_2 released by glucose oxidase. After 1-h incubation, 100 μl of the developed solution was added to 900 μl of 0.1 M Na_3PO_4 (pH 6.0). The UV absorbance at 416 nm was measured.

Data Analysis. Each ELISA assay also included simultaneously assayed positive and negative controls. The positive control was the independently synthesized pentapeptide, NH_2-Tyr-Gly-Gly-Phe-Leu-TG. The negative control was acetylated TG. The results of the assays were first normalized, taking into account the weight of the TG resin in each assay tube, and then the normalized value for the negative control was subtracted from that of each pool.

In the early steps of the deconvolution, the negative control often gave higher absorbance readings than some of the pools, which is reflected in negative values seen in Fig. 3 and is an indicator of the noise in the system. As the library became more defined, fewer negative values were obtained, showing that the signal-to-noise ratio progressively improves.

RESULTS AND DISCUSSION

As a simple test of our method of deconvolution, we screened libraries containing the well-studied pentapeptide sequence NH_2-Tyr-Gly-Gly-Phe-Leu, which displays nanomolar binding to commercially available anti-β-endorphin monoclonal antibody. A synthetic combinatorial peptide library with a degree of five and an alphabet consisting of Leu, Gly, Phe, and Tyr was synthesized on TG solid support. The total number of molecules in this library is 4^5 (1024 members).

Assays for binding to the β-endorphin antibody were most conveniently performed with the peptide libraries attached to the TG resin. In principle, a library could have been constructed with a cleavable linker in which the peptides could be detached from the beads and assayed in solution by a competitive ELISA. The deconvolution sequence can be followed pictorially in Fig. 4 and the corrected absorbances for each round are depicted in Fig. 3. In the four pools of the pentapeptide libraries, where only the N-terminal amino acid was defined, Tyr-Naa-Naa-Naa-Naa-TG clearly gave the strongest absorbance. From this point, the four tetrapeptide partial libraries, Fmoc-Gly-Naa-Naa-Naa-TG, Fmoc-Leu-Naa-Naa-Naa-TG, Fmoc-Phe-Naa-Naa-Naa-TG, and Fmoc-Tyr-Naa-Naa-Naa-TG, saved and catalogued during the formation of the pentapeptide combinatorial library, were all coupled with Tyr-Fmoc. Of the pools Tyr-Gly-Naa-Naa-Naa-TG, Tyr-Leu-Naa-Naa-Naa-TG, Tyr-Phe-Naa-Naa-Naa-TG, and Tyr-Tyr-Naa-Naa-Naa-TG, Tyr-Gly-Naa-Naa-Naa-TG clearly gives the strongest binding of the antibody. In the next deconvolution step, the highest values were found for Tyr-Gly-Gly-Naa-Naa-TG. Solving for the fourth amino acid did not give a unique result but Tyr-Gly-Gly-Phe-Naa-TG, the sequence

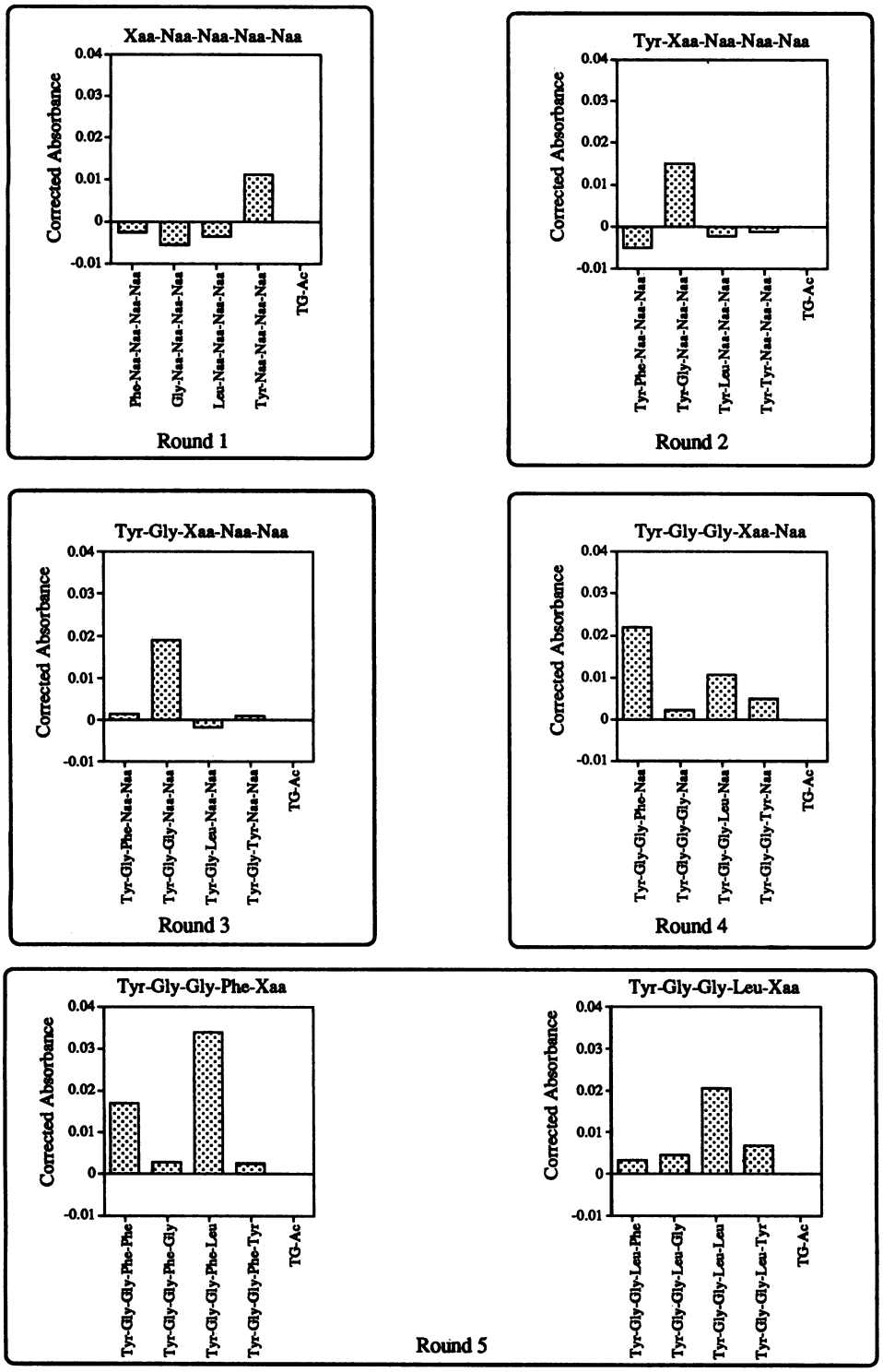


FIG. 3. Corrected absorbances of each pool in comparison to the acetylated TentaGel (TG-Ac) control pool. Xaa denotes the amino acid to be defined in that round. Naa is a randomized position.

corresponding to that of the native epitope, was the strongest binder. However, Tyr-Gly-Gly-Leu-Naa-TG also showed significant binding to the antibody. This was significant enough, in fact, to warrant the coupling of Tyr-Gly-Gly-Leu and Tyr-Gly-Gly-Phe to the monocoupled TG libraries saved from the first step of the combinatorial synthesis.

In the final analysis, Tyr-Gly-Gly-Phe-Leu-TG, the native epitope, was the most extensive binder. Other weaker binders were also deduced; Tyr-Gly-Gly-Phe-Phe-TG and Tyr-Gly-Gly-Leu-Leu-TG showed significant binding to the β-endorphin antibody. In fact, in experiments assaying support-bound pentapeptide libraries against a β-endorphin antibody, Lam *et al.* (17) using a more extensive amino acid alphabet, reported finding a peptide sequence with stronger binding than the native epitope.

Both the recursive and the iterative methods of deconvolution have the advantage of being easily applicable to all chemistry, but because the recursive method allows for the use of split synthesis it not only ensures that the yields of all components are the same but allows greater chemical versatility as well. Although many compounds are assayed together, especially in the first rounds of deconvolution, as

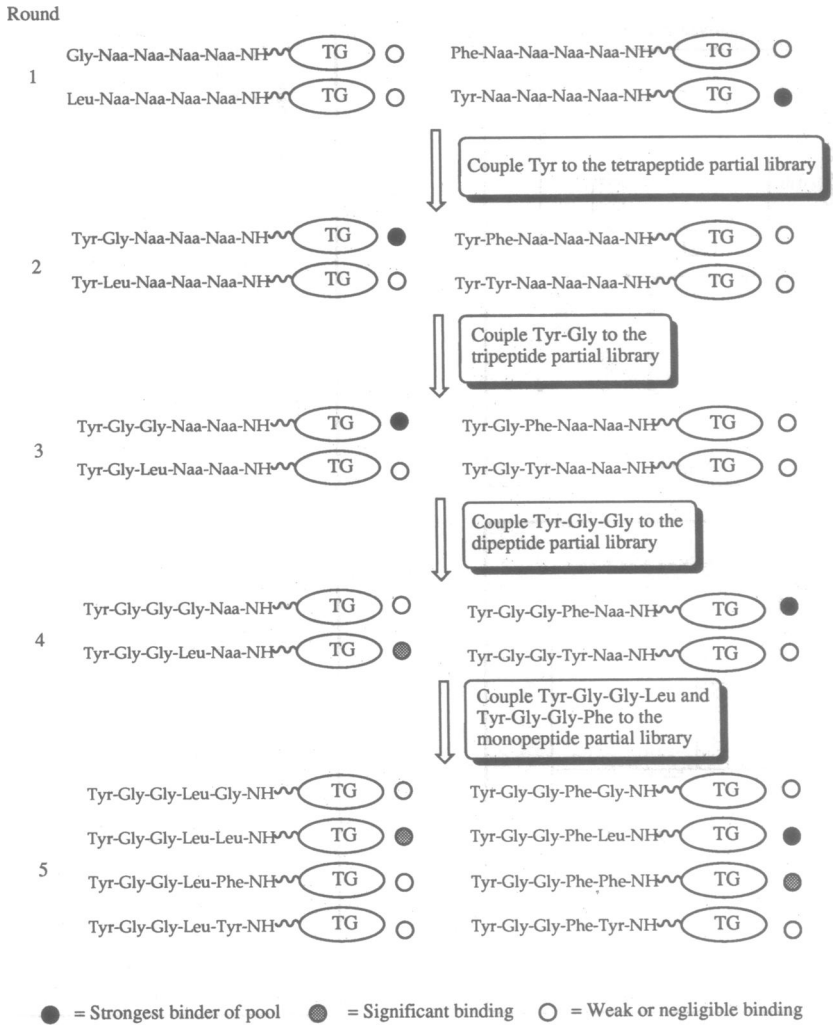


FIG. 4. Route of recursive deconvolution of the TG-bound combinatorial pentapeptide library. Each peptide pool of a round was tested against the anti-β-endorphin antibody. The defined positions of the pool with the strongest and significant binding were noted and coupled to the appropriate partial library for the next round.

the deconvolution proceeds, there is an enrichment of active components by the size of the alphabet in each step. The consequent improvement of the signal-to-noise ratio ensures that false positives are not followed up.

Making combinatorial libraries by split synthesis is a cumbersome process. By holding the partially synthesized libraries at each step, the process need only be carried out once. Thereafter, the entire ensemble may be used repeatedly to find active molecules for any assay. The deconvolution process recursively defines the synthesis of the active compound, so that in the last cycle, the active component is synthesized. If the mixture contains several compounds, each deconvolution pathway can be followed either in parallel or successively, to find all the active members of the library.

We thank Diane M. Schloeder for assistance in the antibody binding studies. This work was funded in part from the Alfred P. Sloan Foundation (K.D.J.) and the Markey Foundation (S.B.).

1. Stevanovic, S. & Jung, G. (1993) *Anal. Biochem.* **212,** 212–220.
2. Ellington, A. D. & Szostak, J. W. (1990) *Nature (London)* **346,** 818–822.
3. Sassanfar, M. & Szostak, J. W. (1993) *Nature (London)* **364,** 550–553.
4. Fodor, S. P. A., Read, J. L., Pirrung, M. C., Stryer, L., Lu, A. T. & Solas, D. (1991) *Science* **251,** 767–773.
5. Jacobs, J. W. & Fodor, S. P. A. (1994) *Trends Biotechnol.* **12,** 19–26.
6. Brenner, S. & Lerner, R. A. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 5381–5383.
7. Nielsen, J., Brenner, S. & Janda, K. D. (1993) *J. Am. Chem. Soc.* **115,** 9812–9813.
8. Youngquist, R. S., Fuentes, G. R., Lacey, M. P. & Keough, T. (1994) *Rapid Commun. Mass Spectrom.* **8,** 77–81.
9. Kerr, J. M., Banville, S. C. & Zuckermann, R. N. (1993) *J. Am. Chem. Soc.* **115,** 2529–2531.
10. Ohlmeyer, M. H. J., Swanson, R. N., Dillard, L. W., Reader, J. C., Asouline, G., Kobayashi, R., Wigler, M. & Still, W. C. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 10922–10926.
11. Needels, M. N., Jones, D. G., Tate, E. H., Heinkel, G. L., Kochersperger, L. M., Dower, W. J., Barrett, R. W. & Gallop, M. A. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 10700–10704.
12. Houghten, R. A., Pinilla, C., Blondelle, S. E., Appel, J. R., Dooley, C. T. & Cuervo, J. H. (1991) *Nature (London)* **354,** 84–86.
13. Ecker, D. J., Vickers, T. A., Hanecak, R., Driver, V. & Anderson, K. (1993) *Nucleic Acids Res.* **21,** 1853–1856.
14. Dooley, C. T. & Houghten, R. A. (1993) *Life Sci.* **52,** 1509–1517.
15. Eichler, J. & Houghten, R. A. (1993) *Biochemistry* **32,** 11035–11041.
16. Furka, Á., Sebestyén, F., Asgedom, M. & Dibó, G. (1991) *Int. J. Pept. Protein Res.* **37,** 487–493.
17. Lam, K. S., Salmon, S. E., Hersh, E. M., Hruby, V. J., Kazmierski, W. M. & Knapp, R. J. (1991) *Nature (London)* **354,** 82–84.
18. Furka, A., Sebestyen, F., Asgedom, M. & Dibo, G. (1988) in *Highlights of Modern Biochemistry: Proceedings of the 14th International Congress of Biochemistry* (VSP, Utrecht, The Netherlands), Vol. 5, p. 47.