US005573905A

United States Patent [19]

Lerner et al.

[11] **Patent Number: 5,573,905**

[45] **Date of Patent: Nov. 12, 1996**

[54] **ENCODED COMBINATORIAL CHEMICAL LIBRARIES**

[75] Inventors: **Richard Lerner**, La Jolla; **Kim Janda**, San Diego, both of Calif.; **Sydney Brenner**, Cambridge, England

[73] Assignee: **The Scripps Research Institute**, La Jolla, Calif.

[21] Appl. No.: **860,445**

[22] Filed: **Mar. 30, 1992**

[51] **Int. Cl.**[6] **C12Q 1/68**; G01N 33/53

[52] **U.S. Cl.** **435/6**; 435/7.94

[58] **Field of Search** 435/6, 7.94; 935/77, 935/78; 530/350; 436/501

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,748,111	5/1988	Dattagupta et al.	435/7
4,965,188	10/1990	Mullis et al.	435/6
5,082,780	1/1992	Warren et al.	435/191
5,141,813	8/1992	Nelson	428/402

OTHER PUBLICATIONS

Devlin, et al., Science, 249: 404–406 (1990).

Nelson, et al., *Nucleic Acids Research,* 17: 7179–7195 (1989).

Nelson, et al., *Nucleic Acids Research,* 20: 6253–6259 (1992).

"Sales Literatuare" from Clontech Laboratories, Inc., p. 3 (1994).

Geysen, et al., "Use of Peptide Synthesis of Probe Viral Antigens for Epitopes to a Resolution of a Single Amino Acid," *Proc. Natl. Acad. Sci. U.S.A.,* 81: 3998–4002 (1984).

Maeji, et al., "Simultaneous Multiple Synthesis of Peptide–Carrier Conjugates," *J. Immunol. Met.,* 146: 83–90 (1992).

Fodor, et al., "Light–Directed, Spatially Addressable Parallel Chemical Synthesis," *Science,* 251: 767–775 (1991).

Lam, et al., "A New Type of Synthetic Peptide Library for Identifying Ligand–Binding Activity," *Nature,* 354: 82–84 (1991).

Houghton, et al., "Generation and Use of Synthetic Peptide Combinatorial Libraries for Basic Research and Drug Discovery," *Nature,* 354: 84–86 (1991).

Cwirla, et al., "Peptides on Phage: A Vast Library of Peptides for Idenifying Ligands," *Proc. Natl. Acad. Sci. U.S.A.,* 87: 6378–6382 (1990).

Scott, et al., "Searching for Peptide Ligands with an Epitope Library," *Science,* 249: 386–390 (1990).

Devlin, et al., "Random Peptide Libraries: A Source of Specific Protein Binding Molecules," *Science,* 249: 404–406 (1990).

Primary Examiner—W. Gary Jones
Assistant Examiner—Eggerton Campbell
Attorney, Agent, or Firm—Donald G. Lewis

[57] **ABSTRACT**

The present invention describes an encoded combinatorial chemical library comprised of a plurality of bifunctional molecules having both a chemical polymer and an identifier oligonucleotide sequence that defines the structure of the chemical polymer. Also described are the bifunctional molecules of the library, and methods of using the library to identify chemical structures within the library that bind to biologically active molecules in preselected binding interactions.

5 Claims, 2 Drawing Sheets

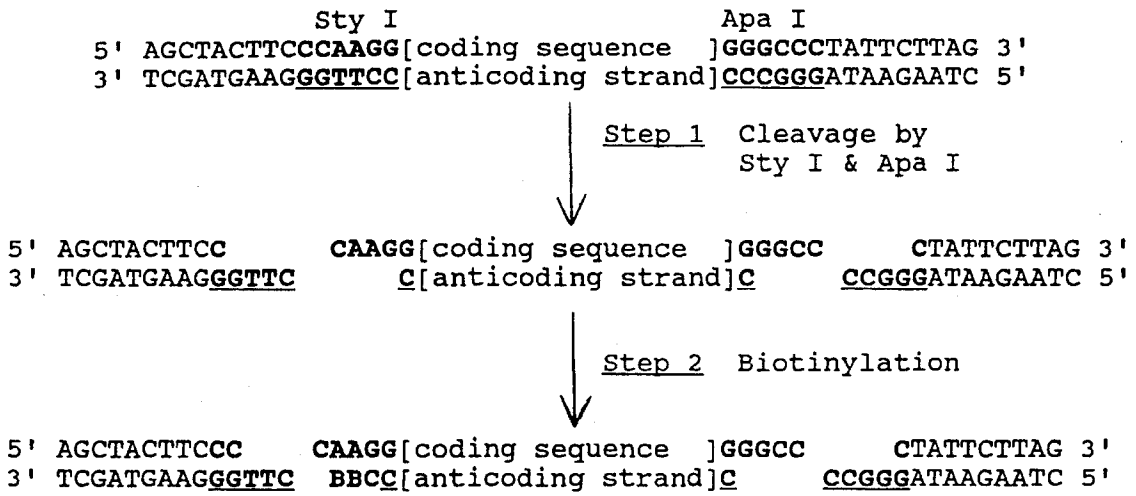
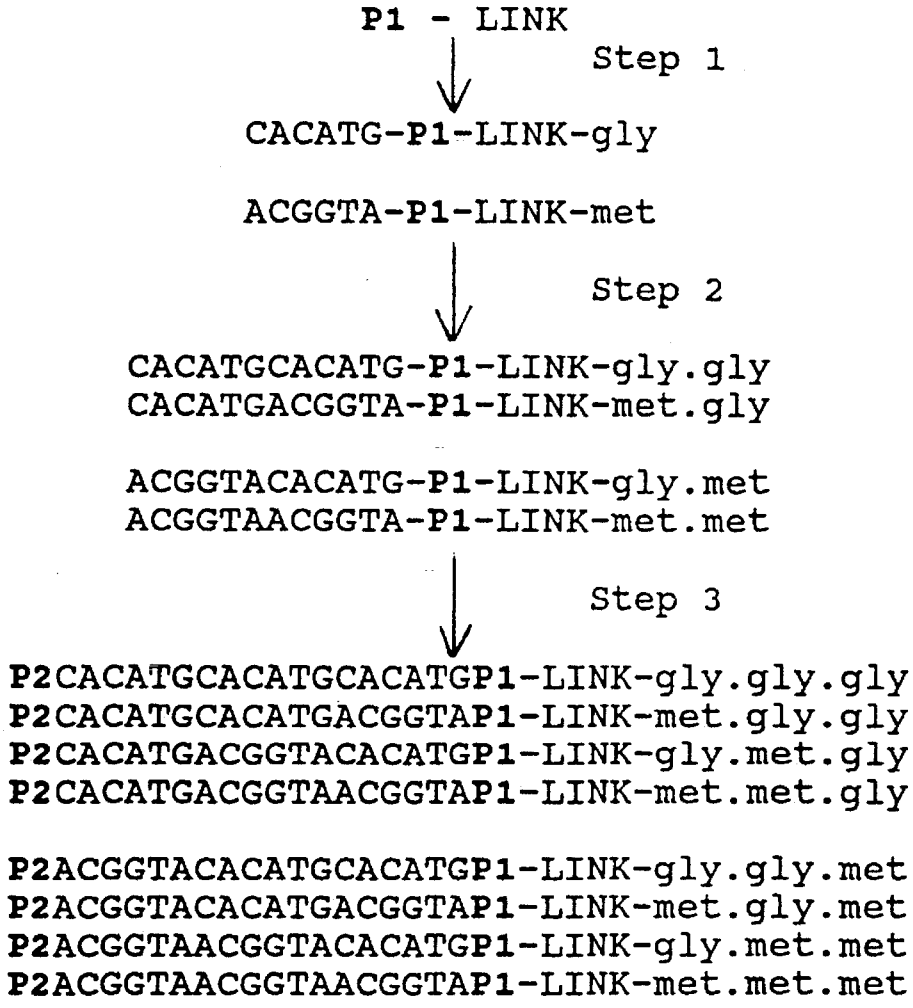


FIGURE 1



P1 = GGGCCCTATTCTTAG
P2 = AGCTACTTCCCAAGG

FIGURE 2

ENCODED COMBINATORIAL CHEMICAL LIBRARIES

TECHNICAL FIELD

The present invention relates to encoded chemical libraries that contain repertoires of chemical structures defining a diversity of biological structures, and methods for using the libraries.

BACKGROUND

There is an increasing need to find new molecules which can effectively modulate a wide range of biological processes, for applications in medicine and agriculture. A standard way for searching for novel bioactive chemicals is to screen collections of natural materials, such as fermentation broths or plant extracts, or libraries of synthesized molecules using assays which can range in complexity from simple binding reactions to elaborate physiological preparations. The screens often only provide leads which then require further improvement either by empirical methods or by chemical design. The process it time-consuming and costly but it is unlikely to be totally replaced by rational methods even when they are based on detailed knowledge of the chemical structure of the target molecules. Thus, what we might call "irrational drug design"—the process of selecting the right molecules from large ensembles or repertoires—requires continual improvement both in the generation of repertoires and in the methods of selection.

Recently there have been several developments in using peptides or nucleotides to provide libraries of compounds for lead discovery. The methods were originally developed to speed up the determination of epitopes recognized by monoclonal antibodies. For example, the standard serial process of stepwise search of synthetic peptides now encompasses a variety of highly sophisticated methods in which large arrays of peptides are synthesized in parallel and screened with acceptor molecules labelled with fluorescent or other reporter groups. The sequence of any effective peptide can be decoded from its address in the array. See for example Geysen et al., *Proc. Natl.Acad. Sci.USA*, 81:3998–4002 (1984); Maeji et al., *J.Immunol.Met.*, 146:83–90 (1992); and Fodor et al., *Science*, 251: 767–775 (1991).

In another approach, Lam et. al., *Nature*, 354:82–84 (1991) describes combinatorial libraries of peptides that are synthesized on resin beads such that each resin bead contains about 20 pmoles of the same peptide. The beads are screened with labelled acceptor molecules and those with bound acceptor are searched for by visual inspection, physically removed, and the peptide identified by direct sequence analysis. In principle, this method could be used with other chemical entities but it requires sensitive methods for sequence determination.

A different method of solving the problem of identification in a combinatorial peptide library is used by Houghten et al., *Nature*, 354:84–86 (1991). For hexapeptides of the 20 natural amino acids, 400 separate libraries are synthesized, each with the first two amino acids fixed and the remaining four positions occupied by all possible combinations. An assay, based on competition for binding or other activity, is then used to find the library with an active peptide. Then twenty new libraries are synthesized and assayed to determine the effective amino acid in the third position, and the process is reiterated in this fashion until the active hexapeptide is defined. This is analogous to the method used in searching a dictionary; the peptide is decoded by construction using a series of sieves or buckets and this makes the search logarithmic.

A very powerful biological method has recently been described in which the library of peptides is presented on the surface of a bacteriophage such that each phage has an individual peptide and contains the DNA sequence specifying it. The library is made by synthesizing a repertoire of random oligonucleotides to generate all combinations, followed by their insertion into a phage vector. Each of the sequences is cloned in one phage and the relevant peptide can be selected by finding those that bind to the particular target. The phages recovered in this way can be amplified and the selection repeated. The sequence of the peptide is decoded by sequencing the DNA. See for example Cwirla et al., *Proc. Natl.Acad. Sci.USA*, 87:6378–6382 (1990); Scott et al., *Science*, 249:386–390 (1990); and Devlin et al., *Science*, 249:404–406 (1990).

Another "genetic" method has been described where the libraries are the synthetic oligonucleotides themselves wherein active oligonucleotide molecules are selected by binding to an acceptor and are then amplified by the polymerase chain reaction (PCR). PCR allows serial enrichment and the structure of the active molecules is then decoded by DNA sequencing on clones generated from the PCR products. The repertoire is limited to nucleotides and the natural pyrimidine and purine bases or those modifications that preserve specific Watson-Crick pairing and can be copied by polymerases.

The main advantages of the genetic methods reside in the capacity for cloning and amplification of DNA sequences, which allows enrichment by serial selection and provides a facile method for decoding the structure of active molecules. However, the genetic repertoires are restricted to nucleotides and peptides composed of natural amino acids and a more extensive chemical repertoire is required to populate the entire universe of binding sites. In contrast, chemical methods can provide limitless repertoires but they lack the capacity for serial enrichment and there are difficulties in discovering the structures of selected active molecules.

BRIEF SUMMARY OF THE INVENTION

The present invention provides a way of combining the virtues of both of the chemical and genetic methods summarized above through the construction of encoded combinatorial chemical libraries, in which each chemical sequence is labelled by an appended "genetic" tag, itself constructed by chemical synthesis, to provide a "retrogenetic" way of specifying each chemical structure.

In outline, two alternating parallel combinatorial syntheses are performed so that the genetic tag is chemically linked to the chemical structure being synthesized; in each case, the addition of one of the particular chemical units to the structure is followed by the addition of an oligonucleotide sequence, which is defined to "code" for that chemical unit, ie., to function as an identifier for the structure of the chemical unit. The library is built up by the repetition of this process after pooling and division.

Active molecules are selected from the library so produced by binding to a preselected biological molecule of interest. Thereafter, the identity of the active molecule is determined by reading the genetic tag, i.e., the identifier oligonucleotide sequence. In one embodiment, amplified copies of their retrogenetic tags can be obtained by the polymerase chain reaction.

The strands of the amplified copies with the appropriate polarity can then be used to enrich for a subset of the library by hybridization with the matching tags and the process can then be repeated on this subset. Thus serial enrichment is achieved by a process of purification exploiting linkage to a nucleotide sequence which can be amplified. Finally, the structure of the chemical entities are decoded by cloning and sequencing the products of the PCR reaction.

The present invention therefore provides a novel method for identifying a chemical structure having a preselected binding activity through the use of a library of bifunctional molecules that provides a rich source of chemical diversity. The library is used to identify chemical structures (structural motifs) that interact with preselected biological molecules.

Thus, in one embodiment, the invention contemplates a bifunctional molecule according to the formula A-B-C, where A is a chemical moiety, B is a linker molecule operatively linked to A and C, and C is an identifier oligonucleotide comprising a sequence of nucleotides that identifies the structure of chemical moiety A.

In another embodiment, the invention contemplates a library comprising a plurality of species of bifunctional molecules, thereby forming a repertoire of chemical diversity.

Another embodiment contemplates a method for identifying a chemical structure that participates in a preselected binding interaction with a biologically active molecule, where the chemical structure is present in the library of bifunctional molecules according to this invention. The method comprises the steps of:

a) admixing in solution the library of bifunctional molecules with the biologically active molecule under binding conditions for a time period sufficient to form a binding reaction complex;

b) isolating the complex formed in step (a); and

c) determining the nucleotide sequence of the polymer identifier oligonucleotide in the isolated complex and thereby identifying the chemical structure that participated in the preselected binding interaction.

The invention also contemplates a method for preparing a library according to this invention comprising the steps of:

a) providing a linker molecule B having termini A' and C' according to the formula A'-B-C' that is adapted for reaction with a chemical precursor unit X' at termini A' and with a nucleotide precursor Z' at termini C';

b) conducting syntheses by adding chemical precursor unit X' to termini A' of said linker and adding precursor unit identifier oligonucleotide Z' to termini C' of said linker, to form a composition containing bifunctional molecules having the structure X_n—B—Z_n;

c) repeating step (b) on one or more aliquots of the composition to produce aliquots that contain a product containing a bifunctional molecule;

d) combining the aliquots produced in step (c) to form an admixture of bifunctional molecules, thereby forming said library.

BRIEF DESCRIPTION OF THE DRAWINGS

In the drawings, forming a portion of this disclosure:

FIG. **1** illustrates a scheme for the restriction endonuclease cleavage of a PCR amplification product derived from a bifunctional molecule of this invention (Step 1), and the subsequent addition of biotin to the cleaved PCR product (Step 2). The unique coding and non-coding nucleotide base sequences shown in FIG. **1** are listed in the Sequence Listing, SEQ ID NOs 15–22.

FIG. **2** illustrates the process of producing a library of bifunctional molecules according to the method described in Example 9. The nucleotide base sequences shown in FIG. **1** are listed in the Sequence Listing, SEQ ID NOs 15–22.

DETAILED DESCRIPTION OF THE INVENTION

A. Encoded Combinatorial Chemical Libraries

An encoded combinatorial chemical library is a composition comprising a plurality of species of bifunctional molecules that each define a different chemical structure and that each contain a unique identifier oligonucleotide whose nucleotide sequence defines the corresponding chemical structure.

1. Bifunctional Molecules

A bifunctional molecule is the basic unit in a library of this invention, and combines the elements of a polymer comprised of a series of chemical building blocks to form a chemical moiety in the library, and a code for identifying the structure of the chemical moiety.

Thus, a bifunctional molecule can be represented by the formula A-B-C, where A is a chemical moiety, B is a linker molecule operatively linked to A and C, and C is an identifier oligonucleotide comprising a sequence of nucleotides that identifies the structure of chemical moiety A.

a. Chemical Polymers

A chemical moiety in a bifunctional molecule of this invention is represented by A in the above formula A-B-C and is a polymer comprising a linear series of chemical units represented by the formula $(X_n)_a$, wherein X is a single chemical unit in polymer A and n is a position identifier for X in polymer A. n has the value of 1+i where i is an integer from 0 to 10, such that when n is 1, X is located most proximal to the linker (B).

Although the length of the polymer can vary, defined by a, practical library size limitations arise if there is a large alphabet size as discussed further herein. Typically, a is an integer from 4 to 50.

A chemical moiety (polymer A) can be any of a variety of polymeric structures, depending on the choice of classes of chemical diversity to be represented in a library of this invention. Polymer A can be any monomeric chemical unit that can be coupled and extended in polymeric form. For example, polymer A can be a polypeptide, oligosaccharide, glycolipid, lipid, proteoglycan, glycopeptide, sulfonamide, nucleoprotein, conjugated peptide (i.e., having prosthetic groups), polymer containing enzyme substrates, including transition state analogues, and the like biochemical polymers. Exemplary is the polypeptide-based library described herein.

Where the library is comprised of peptide polymers, the chemical unit X can be selected to form a region of a natural protein or can be a non-natural polypeptide, can be comprised of natural D-amino acids, or can be comprised of non-natural amino acids or mixtures of natural and non-natural amino acids. The non-natural combinations provide for the identification of useful and unique structural motifs involved in biological interactions.

Non-natural amino acids include modified amino acids and L-amino acids, stereoisomer of D-amino acids.

The amino acid residues described herein are preferred to be in the "L" isomeric form. NH_2 refers to the free amino group present at the amino terminus of a polypeptide. COOH refers to the free carboxy group present at the carboxy terminus of a polypeptide. In keeping with standard

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.