



Office de la Propriété
Intellectuelle
du Canada

Un organisme
d'Industrie Canada

Canadian
Intellectual Property
Office

An agency of
Industry Canada

CA 2382545 C 2013/08/13

(11)(21) **2 382 545**

(12) **BREVET CANADIEN
CANADIAN PATENT**

(13) **C**

(86) Date de dépôt PCT/PCT Filing Date: 2000/08/25
(87) Date publication PCT/PCT Publication Date: 2001/03/08
(45) Date de délivrance/Issue Date: 2013/08/13
(85) Entrée phase nationale/National Entry: 2002/02/20
(86) N° demande PCT/PCT Application No.: US 2000/023414
(87) N° publication PCT/PCT Publication No.: 2001/016352
(30) Priorité/Priority: 1999/08/27 (US60/151,261)

(51) Cl.Int./Int.Cl. *C12P 21/06* (2006.01),
C07K 1/04 (2006.01), *C07K 1/107* (2006.01),
C07K 1/13 (2006.01)
(72) Inventeur/Inventor:
KUIMELIS, ROBERT G., US
(73) Propriétaire/Owner:
BRISTOL-MYERS SQUIBB COMPANY, US
(74) Agent: SMART & BIGGAR

(54) Titre : METHODES DE CODAGE ET DE TRI DE PROTEINES TRADUITES IN VITRO
(54) Title: METHODS FOR ENCODING AND SORTING IN VITRO TRANSLATED PROTEINS

(57) **Abrégé/Abstract:**

Described herein are methods and reagents for encoding and sorting in vitro translated proteins.



(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
8 March 2001 (08.03.2001)

PCT

(10) International Publication Number
WO 01/16352 A1

(51) International Patent Classification⁷: **C12P 21/06**

(21) International Application Number: PCT/US00/23414

(22) International Filing Date: 25 August 2000 (25.08.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/151,261 27 August 1999 (27.08.1999) US

(71) Applicant: PHYLOS, INC. [US/US]; 128 Spring Street,
Lexington, MA 02421 (US).

(72) Inventor: KUIMELIS, Robert, G.; 21 Malbert Road,
Brighton, MA 02135 (US).

(74) Agent: ELBING, Karen, L.; Clark & Elbing LLP, 176
Federal Street, Boston, MA 02110-02214 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

- With international search report.
- With amended claims.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 01/16352 A1

(54) Title: METHODS FOR ENCODING AND SORTING *IN VITRO* TRANSLATED PROTEINS

(57) Abstract: Described herein are methods and reagents for encoding and sorting *in vitro* translated proteins.

METHODS FOR ENCODING AND SORTING
IN VITRO TRANSLATED PROTEINS

Background of the Invention

5 In general, the present invention relates to methods of generating fixed arrays of proteins or coded sets of protein-conjugated microparticles.

 Certain macromolecules, such as proteins, are known to interact specifically with other molecules based on their three-dimensional shapes and electronic distributions. For example, proteins interact selectively with other
10 proteins, nucleic acids, and small-molecules. The identification of molecules that interact with proteins lays the groundwork for the development of compounds to treat diseases and their associated symptoms.

 The discovery of a single drug candidate can require the screening of thousands of compounds. It is therefore important to be able to screen large
15 numbers of compounds rapidly and efficiently. One method for screening a large number of compounds is to fix candidate binding partners, such as proteins, to a solid support.

Summary of the Invention

20 Various embodiments of this invention provide a method for encoding an *in vitro* translated protein with a unique tag, said method comprising the steps of: (a) providing an *in vitro* translated protein attached to a nucleic acid linker; and (b) binding an encoding molecule to said nucleic acid linker, thereby encoding said *in vitro* translated protein with a unique tag, wherein
25 the mRNA encoding said *in vitro* translated protein is degraded.

 Various embodiments of this invention provide a method for encoding an *in vitro* translated protein with a unique tag, said method comprising the steps of: (a) providing an *in vitro* translated protein; and (b) binding a nucleic

acid linker to said *in vitro* translated protein, wherein said nucleic acid linker contains an addressing element, thereby encoding said *in vitro* translated protein with a unique tag, wherein the mRNA encoding said *in vitro* translated protein is degraded.

5 Various embodiments of this invention provide a method for encoding an *in vitro* translated protein with a tag that facilitates recognition of the *in vitro* translated protein among a population of proteins, said method comprising the steps of: (a) providing the *in vitro* translated protein attached to a nucleic acid linker, wherein the nucleic acid linker does not comprise an RNA which encodes said protein; and (b) binding the tag to said *in vitro* translated protein or said nucleic acid linker, thereby encoding said *in vitro* translated protein with the tag; wherein the tag comprises a unique
10 portion having a sequence to give the tag an identity distinct from sequences in tags bound to other proteins (an addressing element); and, wherein mRNA encoding said *in vitro* translated protein is degraded.

Various embodiments of this invention provide a method for encoding an *in vitro* translated protein with a unique tag, said method comprising the steps of: (a) providing an *in vitro* translated
15 protein; and (b) binding a nucleic acid linker to said *in vitro* translated protein, wherein an addressing element branches off from said nucleic acid linker, thereby encoding said *in vitro* translated protein with a unique tag, wherein the mRNA encoding said *in vitro* translated protein is degraded.

Various embodiments of this invention provide a method for encoding an *in vitro* translated
20 protein with a tag that facilitates recognition of the *in vitro* translated protein among a population of proteins, said method comprising the steps of: (a) providing the *in vitro* translated protein; and (b) binding a nucleic acid linker, wherein the nucleic acid linker does not comprise an RNA which encodes said protein, to said *in vitro* translated protein, wherein said nucleic acid linker contains the tag, said tag having a unique portion having a sequence to give the tag an identity distinct from tags
25 bound to other proteins (an addressing element), thereby encoding said *in vitro* translated protein with said tag, wherein the mRNA encoding said *in vitro* translated protein is degraded.

Various embodiments of this invention provide a method for encoding an *in vitro* translated protein with a tag that facilitates recognition of the *in vitro* translated protein among a population of proteins, said method comprising the steps of: (a) providing the *in vitro* translated protein; and (b)
30 binding a nucleic acid linker, wherein the nucleic acid linker does not comprise an RNA which encodes said protein, to said *in vitro* translated protein, wherein the tag with a unique portion having a sequence to give the tag an identity distinct from tags bound to other proteins (an

addressing element) branches off from said nucleic acid linker, thereby encoding said *in vitro* translated protein with the tag, wherein the mRNA encoding said *in vitro* translated protein is degraded.

Various embodiments of this invention provide a method for detecting an interaction
 5 between a protein and a compound, said method comprising the steps of: (a) providing an encoded *in vitro* translated protein immobilized onto a solid support, wherein said *in vitro* translated protein is encoded with a unique tag, wherein the mRNA encoding said *in vitro* translated protein is degraded; (b) contacting said protein with a candidate compound under conditions which allow an interaction between said protein and said compound; and (c) analyzing said solid support for the
 10 presence of said compound as an indication of an interaction between said protein and said compound.

Various embodiments of this invention provide an *in vitro* translated protein attached to a nucleic acid linker molecule, wherein the nucleic acid linker does not comprise an RNA which encodes said protein, wherein the nucleic acid linker is encoded with a tag.

15 Various embodiments of this invention provide an *in vitro* translated protein attached to a branched nucleic acid linker molecule, wherein the nucleic acid linker does not comprise an RNA which encodes said protein, wherein the branched nucleic acid linker is encoded with a tag.

Various embodiments of this invention provide an *in vitro* translated protein attached to a nucleic acid linker, wherein the nucleic acid linker does not comprise an RNA which encodes said
 20 protein, and bound to a tag that facilitates recognition of the *in vitro* translated protein among a population of proteins in which mRNA encoding said *in vitro* translated protein is degraded, wherein the tag comprises an addressing element having a sequence that gives the tag a distinct identity.

The present invention features methods for tagging or “encoding” individual *in vitro*
 25 translated proteins, or groups of *in vitro* translated proteins, with unique and minimal encoding molecules, and related methods for subsequently sorting those encoded molecules onto solid supports or microparticles. The present invention also features methods for the identification of a desired binding partner (for example, a protein or other compound) using the encoded and sorted proteins of the invention. The

invention facilitates the isolation of proteins with desired properties from large pools of partially or completely random amino acid sequences. The invention also facilitates the use of automated approaches to protein or compound screening methods.

5 Accordingly, in a first aspect, the invention features a method for encoding and sorting an *in vitro* translated protein, involving the steps of providing an *in vitro* translated protein attached to a nucleic acid linker and attaching the protein, through the nucleic acid linker, to an encoding molecule, thereby encoding the protein.

10 In one embodiment, this method further involves immobilizing the encoded protein onto a solid support. In another embodiment, the candidate protein is derived from an RNA-protein fusion molecule. In yet another embodiment, the encoding molecule is made of nucleic acids, or nucleic acid analogs. Preferably, the encoding molecule comprises a unique addressing
15 element, a linker-specific alignment element, and a linkage element between the addressing element and the linker-specific alignment element. Furthermore, the linkage element of the encoding molecule may include polyethylene glycol units (preferably, hexaethylene oxide). In yet another embodiment, the candidate protein is attached to the encoding molecule
20 through hybridization of the linker-specific alignment element of the encoding molecule to the nucleic acid linker of the candidate protein, or to the protein itself.

 In a second aspect, the invention features a method for encoding an *in vitro* translated protein, involving the steps of providing an *in vitro*
25 translated protein and binding a nucleic acid linker to the protein, wherein the nucleic acid linker contains an addressing element, thereby encoding the protein.

In a preferred embodiment, this method further involves immobilizing the encoded protein onto a solid support. In another preferred embodiment, the candidate protein is derived from an RNA-protein fusion molecule.

5 In a third aspect, the invention features a method for encoding an *in vitro* translated protein, involving the steps of providing an *in vitro* translated protein and binding a nucleic acid linker to the protein, wherein an addressing element branches off from the nucleic acid linker, thereby encoding the protein.

10 In one embodiment, this method further involves immobilizing the encoded protein formed in the last step of the invention onto a solid support. In another embodiment, the candidate protein is derived from an RNA-protein fusion molecule. In yet another embodiment, the addressing element is bound to the nucleic acid linker by a linkage element. The linkage element of the
15 encoding molecule may include polyethylene glycol units. Preferably, the polyethylene glycol units are hexaethylene oxide.

In yet other embodiments of each of the above aspects of the invention, the solid support is a glass or silica-based chip, or a bead. A capture probe may be attached to the solid support, and may consist of nucleic acids or
20 nucleic acid analogs. The encoded candidate protein may be immobilized onto the solid support by hybridizing the encoded candidate protein to the nucleic acid capture probe, thus sorting the protein according to the information contained in the encoding molecule.

In further embodiments of all of the aspects of the invention, the
25 candidate protein is labeled with a reporter tag, which is preferably a fluorophore. An affinity tag may also be attached to the encoding molecule. One exemplary affinity tag is biotin.

In yet further embodiments, the encoding molecule and solid support are functionalized with a cross-linking moiety. Preferably, the cross-linking moiety is a psoralen, azido compound, or sulfur-containing molecule. In one embodiment, the 5' terminus of the encoding molecule is functionalized with an electrophile that cross-links regioselectively with a nucleophilic amino acid side chain of the protein.

In a fourth aspect, the invention features a method for detecting an interaction between a protein and a compound, involving the steps of providing an encoded *in vitro* translated protein immobilized onto a solid support; contacting the protein with a candidate compound under conditions which allow an interaction between the protein and the compound; and analyzing the solid support for the presence of the compound as an indication of an interaction between the protein and the compound. The compound may be a nucleic acid, a protein, a therapeutic, or an enzyme.

In a fifth aspect, the invention features an *in vitro* translated protein attached to a nucleic acid linker and bound to an encoding molecule.

In a sixth aspect, the invention features an *in vitro* translated protein attached to an encoded nucleic acid linker molecule.

In a seventh aspect, the invention features an *in vitro* translated protein attached to a branched encoded nucleic acid linker molecule.

In various preferred embodiments, the protein is attached to a solid support bearing a capture probe. In other embodiments, the encoded protein is attached to the capture probe through hybridization or a covalent bond.

As used herein, by a "protein" is meant any two or more naturally occurring or modified amino acids joined by one or more peptide bonds. "Protein," "peptide," and "polypeptide" are used interchangeably.

By an “encoding molecule” is meant a unique tag which may be attached to a protein or peptide and which facilitates recognition of the protein among a population of proteins. The encoding molecule may be composed of nucleic acids, nucleic acid analogs, or non-nucleosides, but it is not comprised
5 of the RNA that, when translated, yields the protein itself. By “encode” is meant to attach an encoding molecule.

By an “addressing element” is meant that portion of an encoding molecule which gives the encoding molecule its unique identity by differing sufficiently in sequence from other such elements in a given population.
10 Preferably the addressing element is between 4 and 40 nucleotide units in length. In addition, the addressing element may comprise nucleic acids or nucleic acid analogs.

By a “linker-specific alignment element” is meant that portion of an encoding molecule which hybridizes to the nucleic acid linker of an *in vitro*
15 translated protein, or to the protein itself. The addressing element may consist of nucleic acids or nucleic acid analogs.

By a “linkage element” is meant that portion of an encoding molecule that joins the addressing element and the linker-specific alignment element together. The linkage element may be composed of nucleic acids,
20 nucleic acid analogs, and non-nucleosides. Preferably the linkage element includes polyethylene glycol units, and more preferably the polyethylene glycol units are hexaethylene oxide.

By “sort” is meant to position in an organized manner or otherwise identify or separate. Encoded proteins may be sorted onto a solid support.

By a “solid support” is meant any solid surface including, without limitation, any chip (for example, silica-based, glass, or gold chip), glass slide, membrane, bead, solid particle (for example, agarose, sepharose, polystyrene or magnetic bead), column (or column material), test tube, or microtiter dish.

5 By a “microarray” is meant a fixed pattern of immobilized objects on a solid surface or membrane. Typically, the array is made up of encoded proteins bound to capture probes which themselves are immobilized on the solid surface or membrane. “Microarray” and “chip” are used interchangeably. Preferably the microarray has a density of between 10 and 1000 objects/cm².

10 By “capture probe” is meant a sequence of deoxyribonucleotides, ribonucleotides, or analogs thereof, which hybridize in a sequence dependent manner to the addressing element of a unique encoding molecule in a population. The capture probe may consist of nucleic acids or nucleic acid analogs.

15 By “nucleic acid linker” is meant a sequence of deoxyribonucleotides, ribonucleotides, or analogs thereof. The nucleic acid linker is not comprised of the RNA that, when translated, yields the protein to which it is attached.

20 By an “encoded DNA linker” is meant a sequence of deoxyribonucleotides which contains an addressing element. In a “branched encoded DNA linker,” the addressing element branches from an internal linker deoxyribonucleotide moiety. An encoded DNA linker may also comprise nucleic acid analogs.

25 By a “reporter tag” is meant a molecule whose presence can be monitored or detected. For example, the reporter tag can be a fluorophore.

By “therapeutic” is meant any molecule used to treat, ameliorate, improve, prevent, or stabilize a disease or symptom of a disease.

By an "RNA" is meant a sequence of two or more covalently bonded, naturally occurring or modified ribonucleotides. One example of a modified RNA included within this term is phosphorothioate RNA.

By "RNA-protein fusion" is meant an RNA molecule covalently
5 bound to a protein.

By "functionalize" is meant to chemically modify in a manner that results in the attachment of a functional group or moiety. For example, an encoding molecule may be functionalized with an electrophile that cross-links regioselectively with a nucleophilic amino acid side chain of a protein or
10 peptide. An encoding molecule or the capture probes of the solid support, in another example, can be functionalized with a cross-linking moiety such as psoralen, azido compounds, or sulfur-containing nucleosides.

The present invention provides a number of advantages. For example, the invention allows the employment of pre-made sets of universal
15 encoding molecules, such as nucleic acids or nucleic acid analogues. These encoding molecules can be used in conjunction with corresponding universal microarrays or sets of microparticles to create novel protein-display systems. A system of pre-made encoding molecules is flexible, modular, scalable, and cost-effective. Another advantage of the present invention is the option of
20 utilizing nucleic acid analogs which are not amenable to enzymatic incorporation or polymerization, but which are superior to conventional DNA or RNA in a number of respects. An additional advantage of the present invention is the ability to label proteins with fluorescent moieties, which can be used to monitor the protein in real time.

25 Yet another advantage of the present invention is the absence of RNA which encodes the protein in the final encoded and sorted product. This

is important for several reasons. In particular, DNA is simpler to work with due to its chemical stability and its resistance to nucleases. In addition, the length of a protein's RNA is directly related to the protein's size, with large proteins possessing long RNA messages. Regions of these long RNAs sometimes have a propensity to adopt stable secondary structures which are difficult to predict, and these secondary structures can interfere with hybridization steps and protein folding and function. Accordingly, the development of a method to encode and sort proteins in the absence of the RNA which encodes the protein represents an advance in this field.

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

Detailed Description

The drawings will first briefly be described.

Brief Description of the Drawings

FIGURE 1A is a schematic representation of one exemplary step involved in the production of encoded and sorted *in vitro* translated proteins. In this step, an RNA molecule which is attached to a peptide acceptor through a DNA linker is *in vitro* translated to form an RNA-protein fusion molecule. The RNA portion of the fusion molecule is subsequently degraded by an RNase. The remaining portion of the fusion molecule contains a protein attached to a DNA linker.

FIGURE 1B is a schematic representation of the protein portion of the fusion molecule remaining after degradation of the RNA as described in Figure 1A.

FIGURE 1C is a schematic representation of other exemplary steps involved in the production of encoded and sorted *in vitro* translated proteins. This step lists subsequent procedures involved in creating an encoded protein, including (i) hybridizing the DNA linker to a “unique encoding molecule,” (ii) 5 triggering a covalent cross-linking of the encoding molecule to the DNA linker, (iii) combining all encoded proteins into one solution, and (iv) isolating encoded proteins by affinity separation, followed by concentration of the protein products. The final product of these steps is a mixture of encoded *in vitro* translated protein.

10 FIGURE 1D is a schematic representation of further exemplary steps involved in the production of encoded and sorted *in vitro* translated proteins. These further steps include (i) hybridizing the molecule formed in Figure 1C, through the unique encoding molecule, to a universal chip or set of beads and (ii) triggering a covalent cross-link between the chip or bead and the unique 15 encoding element. This step also illustrates the specific binding of an encoded protein to the capture probe of the solid support, and the lack of binding between an encoded protein and a capture probe which does not correspond to the encoded protein.

FIGURE 2A is a schematic representation of an exemplary encoding 20 molecule in general terms, which is used to encode a protein as described in Figures 1A-1D. The molecule comprises an addressing element attached to a linker-specific alignment element containing a primary cross-linking moiety, through a linkage element.

FIGURE 2B is a schematic representation of an exemplary encoding 25 molecule used to encode a protein as described in Figures 1A-1D, in which the affinity tag is biotin, the secondary cross-linking moiety is a sulfur-containing nucleoside, the boxed nucleotide sequences represent the addressing elements,

and the primary cross-linking moiety is psoralen. The addressing element is joined to the linker-specific alignment element through a hexaethylene oxide linkage element.

FIGURE 3A is a schematic representation of an example of a solid support, a bead, which may be used in the present invention to create protein-display systems. The sequence of the capture probe is designed to bind the encoding molecule of example 2B.

FIGURE 3B is a schematic representation of an example of a solid support, a chip, which may be used in the present invention to create protein-display systems. The sequence of the capture probe at the indicated location of the chip is designed to bind the encoding molecule of example 2B.

FIGURE 4A is a schematic representation of one exemplary step involved in the production of encoded and sorted *in vitro* translated proteins using an encoded DNA linker. In this step an RNA molecule is ligated to an encoded DNA linker. The linker bears a code which will be used to bind the encoded protein to a solid support.

FIGURE 4B is a schematic representation of other exemplary steps involved in the production of encoded and sorted *in vitro* translated proteins using an encoded DNA linker. These steps depict the translation of the RNA molecule which is attached to a peptide acceptor through an encoded DNA linker to form an RNA-protein fusion molecule, and subsequent degradation of the RNA portion of the fusion molecule by an RNase. The remaining portion of the fusion molecule contains a protein attached to an encoded DNA linker.

FIGURE 4C is a schematic representation of a final exemplary step involved in the production of encoded and sorted *in vitro* translated proteins

using an encoded DNA linker. This step shows the hybridization of the molecule formed in Figure 4B, through the encoded DNA linker, to a universal chip or set of beads.

FIGURE 5A is a schematic representation of an exemplary encoding molecule used to encode a protein as described in Figures 4A-4C, in general terms, comprising a DNA linker containing addressing elements and a peptide acceptor, for example, puromycin.

FIGURE 5B is a schematic representation of an example of an encoding molecule used in the invention used to encode a protein as described in Figure 4A-4C, wherein the boxed nucleotide sequences represent the addressing elements.

FIGURE 6A is a phosphorimage of an exemplary microarray generated by the methods of the invention.

FIGURE 6B is a fluorescence scan of an exemplary microarray generated by the methods of the invention.

FIGURE 7A is a phosphorimage of an exemplary microarray generated by the methods of the invention.

FIGURE 7B is a fluorescence scan of an exemplary microarray generated by the methods of the invention.

FIGURE 8A is a schematic representation of one exemplary step involved in the production of encoded and sorted *in vitro* translated proteins using a branched encoded DNA linker. This step illustrates the ligation of an RNA molecule to a branched DNA linker.

FIGURE 8B is a schematic representation of other exemplary steps involved in the production of encoded and sorted *in vitro* translated proteins using a branched encoded DNA linker. These steps comprise the translation of

the RNA molecule which is attached to a peptide acceptor through a branched DNA linker to form an RNA-protein fusion molecule, and subsequent degradation of the RNA portion of the fusion molecule by an RNase. The remaining portion of the fusion molecule contains a protein attached to a
5 branched DNA linker. The product of these steps is the protein portion of the fusion molecule remaining after degradation of the RNA.

FIGURE 8C is a schematic representation of a final exemplary step involved in the production of encoded and sorted *in vitro* translated proteins using a branched encoded DNA linker. In this step the molecule formed in
10 FIGURE 8B is hybridized, through the addressing elements of the branched DNA linker, to a universal chip or set of beads.

FIGURE 9A is a schematic representation of an exemplary encoding molecule used to encode a protein as described in Figures 8A-8C, in general terms.

15 FIGURE 9B is a schematic representation of an example of an encoding molecule used in the invention, and as described in Figures 8A-8C, wherein X represents the branch point of the molecule, and the boxed nucleotide sequences represent the addressing elements. The addressing element is attached to the DNA linker through a hexaethylene oxide linkage
20 element.

Described herein are methods of encoding and sorting *in vitro* translated proteins. Techniques for carrying out each method of the invention are now described in detail, using particular examples. These examples are provided for the purpose of illustrating the invention, and should not be
25 construed as limiting.

Example 1

Encoding and Sorting *In Vitro* Translated Proteins Using an Encoding Molecule

In vitro translated proteins can be encoded and sorted, for example,
5 as shown in Figures 1A-1D. Individual RNA sequences (or a plurality of
sequences) are translated *in vitro*, and RNA-protein fusions are formed, for
example, according to the methods of Roberts and Szostak (Proc. Natl. Acad.
Sci. USA 94:12297-12302, 1997) and Szostak et al. (WO 98/31700,
WO 00/47775, and US 6,261,804). The RNA for the *in vitro*
10 translation reaction may be generated by any standard approach, including
normal cellular synthesis, recombinant techniques, and chemical synthesis, and
includes, without limitation, cellular RNA, mRNA libraries, and random
synthetic RNA libraries. A peptide acceptor (for example, puromycin) is
bonded to the RNA through a nucleic acid or nucleic acid analog linker.
15 Exemplary nucleic acid analogs may be, for example, a PNA (Nielsen et al.,
Science 254:1497-1500, 1991), a P-RNA (Krishnamurthy, Agnew. Chem.
35:1537, 1996), or a 3'N phosphoramidate (Gryaznov and Letsinger, Nucleic
Acids Res. 20:3403-3409, 1992). Such peptide acceptor molecules may be
generated by any standard technique, for example, the techniques described in
20 Roberts and Szostak (supra) and Szostak et al. (supra).

The RNA-protein fusion molecule preferably consists of an RNA
molecule, which includes a translation initiation sequence and a start codon
operably linked to a candidate protein coding sequence and a peptide acceptor
at the 3' end of the candidate protein coding sequence. A DNA or RNase
25 resistant nucleic acid analog sequence is included between the end of the
message and the peptide acceptor. If desired, groups or collections of RNA

sequences, for example, from a particular source or of a given type, may be translated together in a single reaction mixture according to the same general procedure.

5 If desired, the RNA protein fusion may be labeled with a reporter group, for example, a fluorescent reporter group. A fluorescent reporter group may be incorporated into the puromycin-containing DNA linker during the assembly of the DNA linker, for example, by modifying the general procedure as described by Roberts and Szostak (supra), replacing one or more of the nucleotides of the DNA linker with fluorescein dT (Glen Research, Sterling,
10 VA).

A suitable RNase devoid of DNase activity, for example, RNase I available from Ambion (Austin, TX), is then added to the fusion reaction to degrade the RNA portion of the RNA-protein fusion molecule.

Each remaining protein is then encoded as described below. For the
15 encoding steps, individual proteins may be encoded, each with a unique encoding molecule. In this strategy, each encoded protein may be designed to correspond to only one capture probe during the sorting process, and the exact “address” of each protein on the solid support is therefore known.

Alternatively, a plurality of proteins may be pooled and encoded with one or
20 more encoding molecules. In this strategy, the same encoding element may encode one or more different proteins, and, when the encoded proteins are sorted, more than one encoded protein may bind to a specific capture probe. Therefore each “address” on the solid support may contain a mixture of proteins, each possessing the same encoding molecule.

25 A unique encoding molecule, as shown, for example, in Figures 2A and 2B, in an approximately 1:1 molar ratio to the input RNA, is then added to each well. Each unique protein receives a different unique encoding molecule,

and, since the identity of the unique encoding molecule is known, the identity of the protein may be determined. As shown in Figure 2A, each unique encoding molecule includes three essential elements: a “linker-specific alignment element” which comprises nucleic acids or nucleic acid analogs and which binds either to the DNA linker positioned between the peptide acceptor and the *in vitro* translated protein or directly to the protein portion; an “addressing element” which comprises nucleic acids or nucleic acid analogs and which binds to a specific position on the solid support or to a specific microparticle; and a “linkage element” which connects the linker-specific alignment element and unique encoding element.

Simple unique encoding molecules can be assembled in the 3'→5' direction by conventional automated, solid-supported phosphoramidite oligonucleotide chemistry, for example, as described by Beaucage and Caruthers (Tetrahedron Letters 22:1859, 1981). For example, the synthesis of the exemplary encoding molecule shown in Figure 2B begins with a solid support (e.g., controlled-pore glass or polystyrene) functionalized with an A monomer at the nucleoside's 3'-hydroxyl group. Additional monomers are coupled stepwise until the desired encoding sequence has been built up. Design rules for specific encoding sequences are described in U.S. Pat. No. 5,863,722. After the encoding sequence has been assembled, four hexaethylene oxide monomer units (Glen Research) are added to provide the flexible linkage element. The linker-specific alignment element is added next. In the event that the fusions are prepared with the standard 30P DNA linker, as described by Roberts and Szostak (supra), a 15-mer poly-T alignment element is added. If other DNA linkers are utilized in the preparation of the fusions (i.e., not 30P), then the alignment element must be the reverse-complement of some region of the DNA linker sequence. Finally, if desired, a psoralen

phosphoramidite (Glen Research), or an equivalent phosphoramidite moiety, is added at the 5'-terminus to function as the primary crosslinking moiety.

Upon completion of synthesis, the encoding molecule is cleaved from the support and deprotected with ammonium hydroxide using methods
5 known to those skilled in the art. Final purification is accomplished by standard chromatographic or electrophoretic techniques.

The simple unique encoding molecules can be readily elaborated to give an expanded encoding molecule. For example, as described above, the unique encoding molecule, typically at the linker-specific alignment element,
10 may be functionalized with a primary cross-linking moiety, for example, psoralen, which is used to permanently crosslink the unique encoding species to the *in vitro* translated protein. The addressing element may also be functionalized, with a secondary cross-linking moiety, for example, 4-thio T, to form a covalent bond between the solid support and the unique encoding
15 molecule. Moreover, the addressing element may be further labeled with an affinity tag, for example, biotin, which may be used in the isolation and concentration of the encoded proteins. The affinity tag is incorporated as a 3'-support (Glen Research) and the rest of the encoding molecule is constructed as described above. In one particular embodiment, the linkage element of the
20 unique encoding molecule consists of polyethylene glycol units, for example, hexaethylene oxide.

The unique encoding molecules are hybridized to the proteins through an interaction between the linker-specific alignment element of the encoding molecule and the DNA linker of the protein fusion using standard
25 hybridization conditions known to those skilled in the art, as shown, for example, in Figure 1C. The unique encoding molecule is then covalently cross-linked to either the DNA linker of the protein or to the protein itself,

using ultraviolet light, typically at 350 nm. If the encoding molecule is to be covalently cross-linked to the DNA linker of the protein, a crosslinking agent such as psoralen may be used, for example according to the methods of Gasparro et al. (Nucleic Acids Research, 22:2845-2852, 1994). Alternatively, 5 if the encoding molecule is to be covalently crosslinked to the protein directly, then a cross-linking agent such as an azido compound may be used, for example, according to the methods of Bayley (Photogenerated Reagents in Biochemistry and Molecular Biology, Elsevier, New York, 1983).

The solutions containing proteins cross-linked to unique encoding 10 molecules are then combined and isolated by standard affinity separation, for example, by applying the biotinylated proteins to a streptavidin column using standard techniques. Other standard affinity separation techniques may be used according to the affinity tag which is attached to the encoding molecule.

The affinity-separated solution containing the encoded proteins is 15 then applied to a solid support, as shown, for example, in Figure 1D. Preferably, the solid support is a universal chip or a universal set of coded microparticles, for example as described by Fulton et al. (Clinical Chemistry 43:1749-1756, 1997) and, as shown in Figures 3A and 3B, containing capture probes designed to interact with the encoded proteins. The capture probes are 20 preferably nucleic acids or nucleic acid analogs which bind to the unique encoding element of the encoding molecule in a sequence specific manner, thereby linking the protein to the solid support and sorting the protein. Each capture probe on the solid support is designed to comprise a different nucleotide sequence, each of which binds a different encoding molecule- 25 protein complex. The capture probe may also contain a molecule, for example,

4-thio T, which upon oxidation can be used to crosslink the capture probe to the encoding molecule via the secondary crosslinking moiety (as described above).

The capture probes may be attached to the solid support by any method, for example, those methods described in Kuimelis et al. (WO 99/51773 and US 6,537,749).

In one exemplary method for attaching the capture probes to the solid support, the capture probes are adjusted to a concentration of 500 μ M in 100 mM sodium carbonate buffer (pH 9.0), and are applied to the derivatized surface of the solid support at defined positions. A three axis motion control apparatus coupled to a microvolume liquid delivery system may be used to accurately deposit the capture probes. The solid support containing the deposited capture probes is incubated at room temperature in a moisture-saturated environment for at least two hours. The attachment reaction is terminated by immersing the glass surface in an aqueous 1% ammonia solution for five minutes with gentle agitation. The glass surface is then subjected to three 5-minute washes, using fresh portions of distilled water for each wash. The array is then soaked in 1 M phosphate buffered saline (PBS) solution for 2 hours at room temperature, then rinsed again for 5 minutes in distilled water.

The "sorted" proteins can be covalently linked to the capture probes, for example, by triggering disulfide bond formation at the terminus of the duplex according to the methods of Cain et al. (Nucleic Acids Research, 23:2153-2160, 1995).

Example 2

Encoding and Sorting *In Vitro* Translated Proteins Using an Encoded DNA Linker

To further simplify the encoding process, *in vitro* translated proteins
5 may also be generated from RNA molecules joined directly to unique encoding
molecules in the form of encoded DNA linkers, as shown in Figures 4A-4C.
The encoded DNA linker may consist of nucleic acids or nucleic acid analogs,
and an “addressing element” which comprises nucleic acids or nucleic acid
analogues that bind to a specific position on the solid support or to a specific
10 microparticle.

Simple encoded DNA linker molecules can be assembled in the
3'→5' direction by conventional automated, solid-support phosphoramidite
oligonucleotide chemistry for example, as described by Beaucage and
Caruthers (supra). For example, the synthesis of the exemplary encoded DNA
15 linker molecule shown in Figure 5B begins with a solid support (e.g.,
controlled-pore glass or polystyrene) functionalized with a peptide acceptor,
for example, puromycin at the nucleoside's 2'-hydroxyl group (Glen
Research). Two C monomers are then added. Next, the desired encoding
sequence is built up by the stepwise addition of appropriate monomers. Design
20 rules for specific encoding sequences are disclosed in U.S. Pat. No. 5,863,722.
Twelve A monomers are coupled stepwise to finish the construction of the
DNA linker. The overall length of the encoded DNA linker molecule is
preferably the same as the 30P DNA linker employed by Roberts & Szostak
(supra). Upon completion of synthesis, the encoding molecule is cleaved from
25 the support and deprotected in the usual fashion with ammonium hydroxide.
Final purification is accomplished by standard chromatographic or
electrophoretic techniques.

If desired, the simple encoded DNA linker molecules can be readily elaborated with an affinity tag (e.g., biotin) and/or a fluorophore (e.g., fluorescein) to give an expanded encoded DNA linker molecule. The affinity tag and fluorophore are each incorporated as nucleobase-functionalized T monomers (Glen Research) in the poly A region of the encoding molecule, replacing one/two of the twelve A monomers as shown in Figure 5B. The remainder of the encoding molecule is constructed as described already for the simple encoded DNA linker.

The encoded DNA linker, which is phosphorylated at its 5' end, is ligated to the RNA using, for example T4 DNA ligase, as shown in Figure 4A. The ligated product is then *in vitro* translated according to the methods of Roberts and Szostak (supra) and Szostak et al. (supra), to form an RNA-protein fusion molecule, and the RNA portion of the molecule is then degraded as described in Example 1, and as shown in Figure 4B. This RNA degradation step results in an *in vitro* translated protein attached to an encoded DNA linker.

The encoded *in vitro* translated protein may be hybridized to a universal chip or set of beads as shown in Figure 4C. Hybridization occurs between the addressing element of the DNA linker and the capture probe of the solid support as described in Example 1.

Using the techniques essentially as described above, a polypeptide that binds TNF- α and a polypeptide that binds IL-13, were *in vitro* translated, encoded, sorted, and shown to bind TNF- α , and IL-13, as follows.

Unique sequences, based on four consecutive 4-nucleotide blocks, were selected to serve as capture points to both sort and anchor the encoded polypeptides. The following is a list of the capture (sorting) sequences that were employed (written 5'-->3'):

TAG_CP-4 = CAAGACACTCATAGCG-(HEO)₄-NH₂

TAG_CP-8 = CAAGACACACACCAAG-(HEO)₄-NH₂

TAG_CP-12 = CAAGACACTCATTTCAT-(HEO)₄-NH₂

TAG_CP-16 = ACACCAAGAGCGTCAT-(HEO)₄-NH₂

5 The oligonucleotides were prepared with an automated DNA
synthesizer (PE BioSystems Expedite 8909) using conventional
phosphoramidite chemistry, and reagents from Glen Research. Synthesis was
initiated with a solid support bearing an orthogonally protected amino
functionality, whereby the 3'-terminal amine is not unmasked until the final
10 deprotection step. The first four monomers to be added were hexaethylene
oxide units (HEO), followed by the standard A, G, C, and T monomers.
Oligonucleotides were cleaved from the solid support and deprotected with
ammonium hydroxide, concentrated to dryness, precipitated in ethanol, and
purified by reverse-phase HPLC using an acetonitrile gradient in
15 triethylammonium acetate buffer. Appropriate fractions from the HPLC were
collected, evaporated to dryness in a vacuum centrifuge, and then co-
evaporated with a portion of water.

 The purified, amine-labeled oligonucleotides were adjusted to a
concentration of 500 μ M in 50 mM sodium carbonate buffer (pH 9.0). These
20 sorting sequences were spotted onto the amine-reactive glass surface (3D-Link,
Surmodics) at defined positions in a 5x5x4 array pattern with a 3-axis robot
(MicroGrid, BioRobotics). A 4-pin tool was used to transfer the liquid from a
384-well microtiter plate, producing 200 micron features with a 600 micron
pitch. Each sub-grid of 24 features represents a single oligonucleotide (i.e., 24
25 replicate spots). The printed arrays were incubated at room temperature in a
moisture-saturated environment for 12-18 hours. The attachment reaction was

terminated by immersing the chips in 2% aqueous ammonium hydroxide for five minutes with gentle agitation, followed by rinsing with distilled water (3 X 5 minutes).

Unique encoded DNA linkers were then synthesized with a
5 5'-terminal dA11 tract, followed by unique 16-mers comprised of 4-nucleotide blocks, dC2 and finally a puromycin at the 3' terminus, similar to the encoding molecule of Figure 5B. The encoding sequences were as follows, and were designed to work in conjunction with the capture probe sequences described above (written 5'-->3'):

10 TAG_LN-8 = AAAAAAAAAAACTTGGTGTGTGTCTTGCC-puromycin
TAG_LN-16 = AAAAAAAAAAAATGACGCTCTTGGTGTCC-puromycin

The encoded DNA linkers were prepared with an automated DNA synthesizer essentially as described above, using conventional phosphoramidite chemistry. All reagents were from Glen Research. Synthesis was initiated
15 with a solid support bearing a protected puromycin moiety. Oligonucleotides were cleaved from the solid support and deprotected with ammonium hydroxide, concentrated to dryness, and then precipitated in ethanol. Purity and integrity were confirmed by anion-exchange HPLC.

Common PCR primers were used to amplify regions of DNA that
20 encode polypeptide sequences that bound TNF- α and IL-13, denoted as FnTNF and FnIL13, respectively. The 3'-primer contained an additional AGCGGATGC sequence at the end. Standard PCR amplifications were carried out using the primers and templates in the presence of PCR reagents (Ready-to-go beads, Amersham) for 25 cycles. The integrity of the PCR
25 products was confirmed on a 2% agarose gel. The FnTNF and FnIL13

sequences were then *in vitro* transcribed (Mega Short Script, Ambion) from the PCR products according to standard protocols and purified on a NAP-25 size-exclusion column (Amersham Pharmacia). The resulting RNA-containing fractions were precipitated and resuspended in H₂O.

5 The FnTNF and FnIL13 RNA constructs were enzymatically ligated to 5' phosphorylated unique encoding sequences. FnTNF RNA was ligated to TAG_LN-8, and FnIL13 was ligated to TAG_LN-16, yielding the following RNA-DNA linker-puromycin chimeras: FnTNF_LN-8 and FnIL13_LN-16, respectively. Ligation was performed on a 1 nanomole scale with equimolar
10 amounts of RNA and 5' phosphorylated encoding DNA-puromycin linker, utilizing 100 units T4 DNA ligase (Promega) in the presence of 1 nanomole of a common DNA splint, TTTTTTTTTTNAGCGGATGC. The incubations were carried out at 16°C for 12-18 hours, and the ligation products were separated by denaturing PAGE (6% TBE-Urea). The ligated products were
15 visualized by UV shadowing, excised, eluted from the gel by crushing and soaking, and subsequently precipitated and resuspended in H₂O.

In vitro translation of the purified ligation products was carried out according to the following procedure: 83 µL ligated RNA (120 pmol) in H₂O was added to 15 µL master mix (Ambion, without methionine), 2 µL ³⁵S-met
20 (15 µM) and 200 µL of rabbit reticulocyte lysate (Ambion) for a total volume of 300 µL. The reaction mixture was incubated for 30 minutes at 30°C, and then 100 µL 2 M KCl and 20 µL 1 M MgCl₂ was added. After a further 60 minutes incubation at room temperature, 47 µL 0.5 M EDTA was added.

The resulting encoded RNA-linker-protein fusions, FNTNF_LN-8
25 and FNIL13_LN-16, were subsequently isolated by oligo-dT chromatography. An equal volume of 2X oligo dT binding buffer (200 mM Tris; pH 8, 2 M

NaCl, 20 mM EDTA, and 0.1% Tween-20) was added to the reactions, and the RNA-linker-fusions were then bound to 100 mg oligo dT cellulose (Pharmacia), rinsed with wash buffer (100 mM Tris pH 8, 1 M NaCl, 0.05% Tween-20), and eluted with H₂O. Quantitation of the RNA-DNA linker fusions was done by scintillation counting, and the integrity of the fusions was confirmed by PAGE (4-20% Tris-glycine).

The FNTNF_LN-8 DNA-protein fusions were then sorted on a microarray. Fifty fmol of fusion was adjusted to 5X SSC containing 0.05% Tween-20 in a total volume of 350 μ L. The RNA was then digested by adding 2 uL RNase A (Ambion) at 37°C for 15 minutes, leaving just the 29-mer DNA containing the 16-nucleotide encoding sequence, in this case TAG_LN-8, fused to the protein. The entire volume was applied to the microarray under a 400 μ L gasket device, and the assembly was continuously rotated for 18 hours at room temperature. After sorting, the slide was washed sequentially with stirred 500 mL portions of 2.5X SSC, 1X SSC, and 0.5X SSC for 5 minutes each at room temperature. Traces of liquid were removed by centrifugation and the slide was allowed to air-dry.

The sorted polypeptide was visualized by direct detection of the ³⁵S methionine on the fusion protein by phosphorimage analysis with a Molecular Dynamics Storm system. Exposure time was 48 hours with direct contact between the microarray and the phosphor storage screen. Phosphorimage scanning was performed at the 50 μ m resolution setting and data was extracted with ImageQuant v.4.3 software.

The functionality of the sorted polypeptide was demonstrated by binding to labeled TNF- α protein. Recombinant human TNF- α (500 μ g, PeproTech) was taken up in 230 μ L 1X PBS and dialyzed against 700 mL stirred 1X PBS at 4°C for 18 hours in a Microdialyzer unit (3,500 MWCO,

Pierce). The dialyzed TNF- α was treated with EZ-Link NHS-LC-LC biotinylation reagent (20 μ g, Pierce) for 2 hours at 0°C and again dialyzed against 700 mL stirred 1X PBS at 4°C for 18 hours in a Microdialyzer unit (3,500 MWCO, Pierce). The resulting conjugate was analyzed by
5 MALDI-TOF mass spectrometry and was found to be almost completely functionalized with a single biotin moiety.

Each of the following processes was conducted at room temperature with continuous rotation or mixing. The protein microarray surface was passivated by treatment with 1X TBS containing 0.05% Tween-20, and 1%
10 BSA (200 μ L) for 60 minutes. Biotinylated TNF- α (100 nM in 1X TBS, 0.02% Tween-20, and 0.2% BSA) was contacted with the microarray for 120 minutes at room temperature. The microarray was washed with 1X TBS containing 0.05% Tween-20 (3 X 50 mL, 5 minutes for each wash).

Fluorescently labeled 2° agent (2.5 μ g/mL Cy3-labeled anti-biotin monoclonal
15 antibody (Sigma), made up in 1X TBS containing 0.05% Tween-20 and 0.2% BSA) was contacted with the microarray for 60 minutes. The microarray was washed with 1X TBS containing 0.05% Tween-20 (2 X 50 mL, 5 minutes each wash) followed by a 3 minute rinse with 1X TBS. Traces of liquid were removed by centrifugation, and the slide was allowed to air-dry at room
20 temperature.

Fluorescence laser scanning was performed with a GSI Lumonics ScanArray 5000 system using 10 μ m pixel resolution and preset excitation and emission wavelengths for Cy3 dye. Figures 6A and 6B are the phosphorimage and fluorescence scan, respectively, of a microarray containing sorted
25 FNTNF_LN-8. The phosphorimage demonstrates the location of the sorted

polypeptide based on the ^{35}S methionine signal. The fluorescence scan shows where the labeled TNF- α protein target has bound, demonstrating functionality of the sorted polypeptide.

5 The encoded IL-13 binder construct FNIL13_LN-16 was prepared as described above, and sorting was performed with the FNTIL13_LN-16 encoded DNA-protein fusions, as described above. The sorted polypeptide was visualized by direct detection of the ^{35}S methionine on the fusion protein by phosphorimage analysis, as described above.

10 The functionality of the sorted polypeptide was demonstrated by binding to labeled IL-13 protein. Recombinant human IL-13 (500 μg , PeproTech) was biotinylated as described above. The resulting conjugate was analyzed by MALDI-TOF mass spectrometry and was found to be almost completely functionalized with a single biotin moiety. Binding of the biotinylated IL-13 protein followed by detection with Cy3-labeled anti-biotin
15 monoclonal antibody was performed as described above.

Figures 7A and 7B are the phosphorimage and fluorescence scan, respectively, of a microarray containing sorted FNIL13_LN-16. The phosphorimage demonstrates the location of the sorted polypeptide based on the ^{35}S methionine signal. The fluorescence scan shows where the labeled IL-
20 13 protein target has bound, demonstrating functionality of the sorted polypeptide.

Example 3

Encoding and Sorting *In Vitro* Translated Proteins Using an Encoded Branched DNA Linker

Yet another method for encoding *in vitro* translated proteins is
5 shown in Figures 8A-8C. In this approach, an RNA encoding the desired
protein to be *in vitro* translated is joined to a unique encoding molecule in the
form of a branched encoded DNA linker. The DNA linker includes nucleic
acids or nucleic acid analogs as shown in Figures 9A-9B. An addressing
element, composed of nucleic acids or nucleic acid analogs, branches off from
10 the DNA linker, to which it is attached by a linkage element.

Simple branched encoded DNA linker molecules can be assembled
in the 3'→5' direction by conventional automated, solid-supported
phosphoramidite oligonucleotide chemistry, for example, as described by
Beaucage and Caruthers (supra). For example, the synthesis of the exemplary
15 branched encoded DNA linker molecule shown in Figure 9B begins with a
solid support (e.g., controlled-pore glass or polystyrene) functionalized with a
peptide acceptor, for example, puromycin at the nucleoside's 2'-hydroxyl
group (Glen Research). Two C monomers are then added followed by 18 A
monomers. A differentially protected asymmetric branching monomer
20 (Clontech, Palo Alto, CA) is then added and the DMT protecting group is
removed. Nine A monomers are subsequently added. The Lev protecting
group on the branching monomer is then removed and four hexaethylene oxide
monomer units (Glen Research) are added to provide the flexible linkage
element. Next, the desired encoding sequence is built up by the stepwise
25 addition of appropriate monomers. Design rules for specific encoding
sequences are disclosed in U.S. Pat. No. 5,863,722. Once completed, branched
encoded DNA linker molecules contain two 5'-termini. Upon completion of

synthesis, the encoding molecule is cleaved from the support and deprotected in the usual fashion with ammonium hydroxide. Final purification is accomplished by standard chromatographic or electrophoretic techniques.

The simple branched encoded DNA linker molecules described
5 above can be readily elaborated with affinity tags (e.g., biotin) and/or
fluorophores (e.g., fluorescein) to give expanded branched encoded DNA
linker molecules. Preferably, the affinity tag and fluorophore are each
incorporated as nucleobase-functionalized T monomers (Glen Research) in the
poly A region of the encoding molecule, replacing one/two of the 27 A
10 monomers. The remainder of the encoding molecule is constructed as
described above for the simple case. The RNA and branched DNA
linker are ligated using, for example, T4 DNA ligase. The ligated product is
then *in vitro* translated according to the methods of Roberts and Szostak
(supra) and Szostak et al. (supra), to form an RNA-protein fusion molecule,
15 and the RNA portion of the molecule is then degraded as described in Example
1, and as shown in Figure 8B. This RNA degradation step results in an *in vitro*
translated protein attached to a branched encoded DNA linker.

The encoded *in vitro* translated protein may be hybridized to a
universal chip or set of beads as shown in Figure 8C. Hybridization occurs
20 between the addressing element of the branched DNA linker and the capture
probe of the solid support as described in Example 1.

Other embodiments are within the claims.

SEQUENCE LISTING

<110> Phyllos, Inc.

<120> METHODS FOR CODING AND SORTING *IN VITRO* TRANSLATED PROTEINS

<130> 81331-90

<140> PCT/US00/23414

<141> 2000-08-25

<150> US 60/151,261

<151> 1999-08-27

<160> 11

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 31

<212> DNA

<213> Artificial Sequence

<220>

<223> Addressing elements

<400> 1

aacgacaaag gccaaactttt tttttttttt t

31

<210> 2

<211> 16

<212> DNA

<213> Artificial Sequence

<220>

<223> Capture probe

<400> 2

ttgctgtttc cggttg

16

<210> 3

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> Addressing elements

<400> 3

aaaaaaaaa aaaacgacaa aggccaaccc

30

<210> 4

<211> 30

<212> DNA

<213> Artificial Sequence

<220>
 <223> Encoding molecule
 <221> misc_feature
 <222> 10
 <223> n at position 10 can be a, t, c, or g.

<400> 4
 aaaaaaaaaan aaaaaaaaaa aaaaaaaacc

30

<210> 5
 <211> 16
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Capture sequence

<400> 5
 caagacactc atagcg

16

<210> 6
 <211> 16
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Capture sequence

<400> 6
 caagacacac accaag

16

<210> 7
 <211> 16
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Capture sequence

<400> 7
 caagacactc attcat

16

<210> 8
 <211> 16
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Capture Sequence

<400> 8
 acaccaagag cgtcac

16

<210> 9
 <211> 29

<212> DNA
<213> Artificial Sequence

<220>
<223> Encoding sequence
<400> 9
aaaaaaaaaa acttggtgtg tgtcttgcc

29

<210> 10
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Encoding sequence

<400> 10
aaaaaaaaaa aatgacgctc ttggtgtcc

29

<210> 11
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> DNA splint

<221> misc_feature
<222> 11
<223> n = A,T,C or G

<400> 11
tttttttttt nagcggatgc

20

Claims

1. A method for encoding an *in vitro* translated protein with a tag that facilitates recognition of the *in vitro* translated protein among a population of proteins, said method comprising the steps of:
 - (a) providing the *in vitro* translated protein attached to a nucleic acid linker, wherein the nucleic acid linker does not comprise an RNA which encodes said protein; and
 - (b) binding the tag to said *in vitro* translated protein or said nucleic acid linker, thereby encoding said *in vitro* translated protein with the tag; wherein the tag comprises a unique portion having a sequence to give the tag an identity distinct from sequences in tags bound to other proteins (an addressing element); and,

wherein mRNA encoding said *in vitro* translated protein is degraded.
2. The method of claim 1, wherein said *in vitro* translated protein is derived from an RNA-protein fusion molecule.
3. The method of claim 1 or 2, wherein said tag further comprises:
 - (a) a portion that hybridizes to the nucleic acid linker or to the *in vitro* translated protein, (a linker-specific alignment element); and
 - (b) a portion joining the addressing element and the linker-specific element (a linkage element) positioned between said addressing element and said linker-specific alignment element.
4. The method of claim 3, wherein said linkage element comprises polyethylene glycol units.
5. The method of claim 4, wherein said polyethylene glycol units are hexaethylene oxide.

6. The method of claim 3, 4 or 5, wherein said tag is bound through hybridization of said nucleic acid linker to said linker-specific alignment element.
7. The method of any one of claims 1 to 6, wherein said tag comprises nucleic acids or nucleic acid analogs.
8. A method for encoding an *in vitro* translated protein with a tag that facilitates recognition of the *in vitro* translated protein among a population of proteins, said method comprising the steps of:
 - (a) providing the *in vitro* translated protein; and
 - (b) binding a nucleic acid linker, wherein the nucleic acid linker does not comprise an RNA which encodes said protein, to said *in vitro* translated protein, wherein said nucleic acid linker contains the tag, said tag having a unique portion having a sequence to give the tag an identity distinct from tags bound to other proteins (an addressing element), thereby encoding said *in vitro* translated protein with said tag, wherein the mRNA encoding said *in vitro* translated protein is degraded.
9. The method of claim 8, wherein said *in vitro* translated protein is derived from an RNA-protein fusion molecule.
10. A method for encoding an *in vitro* translated protein with a tag that facilitates recognition of the *in vitro* translated protein among a population of proteins, said method comprising the steps of:
 - (a) providing the *in vitro* translated protein; and
 - (b) binding a nucleic acid linker, wherein the nucleic acid linker does not comprise an RNA which encodes said protein, to said *in vitro* translated protein, wherein the tag with a unique portion having a sequence to give the tag an identity distinct from tags bound to other proteins (an addressing element) branches off from said nucleic acid linker, thereby encoding said *in vitro* translated protein

with the tag, wherein the mRNA encoding said *in vitro* translated protein is degraded.

11. The method of claim 10, wherein said *in vitro* translated protein is derived from an RNA-protein fusion molecule.
12. The method of any one of claims 8 to 11, wherein said addressing element is bound to said nucleic acid linker by a portion of the tag that hybridizes to the nucleic acid linker (a linker-specific alignment element).
13. The method of claim 12, wherein the tag further comprises a linkage element that joins the addressing element to the linker-specific alignment element.
14. The method of claim 13, wherein said linkage element comprises polyethylene glycol units.
15. The method of claim 14, wherein said polyethylene glycol units are hexaethylene oxide.
16. The method of any one of claims 1 to 15, wherein said nucleic acid linker is functionalized with a cross-linking moiety and said method further comprises crosslinking said nucleic acid linker to said *in vitro* translated protein.
17. The method of any one of claims 1 to 16, wherein said tag is functionalized with a cross-linking moiety and said method further comprises cross-linking said tag to said *in vitro* translated protein.
18. The method of any one of claims 1 to 16, wherein a 5' terminus of said tag molecule is functionalized with an electrophile that cross-links regioselectively with a nucleophilic amino acid side chain of said *in vitro* translated protein, and said method further comprises cross-linking said tag to said *in vitro* translated protein.

19. The method of any one of claims 1 to 18, further comprising immobilizing said encoded protein formed in step (b) onto a solid support.
20. The method of claim 19, wherein said encoded protein is selected from a mixture of encoded proteins.
21. The method of claim 19 or 20, wherein said solid support is a glass or silica-based chip or a bead.
22. The method of claim 19, 20 or 21, wherein said solid support is functionalized with a cross-linking moiety and said method further comprises cross-linking said encoded protein to said solid support.
23. The method of claim 16, 17 or 22, wherein said cross-linking moiety is chosen from the group consisting of psoralen, azido compounds, and sulfur-containing molecules.
24. The method of claim 18, 19 or 20, wherein a capture probe which hybridizes in a sequence dependent manner to a unique portion of the tag having a sequence that gives the tag an identity distinct from tags bound to other proteins (an addressing element), is attached to said solid support.
25. The method of claim 24, wherein said capture probe comprises nucleic acids or nucleic acid analogs.
26. The method of claim 24 or 25, wherein said encoded protein is immobilized onto said solid support by hybridization to said capture probe.
27. The method of any one of claims 1 to 26, wherein said protein is labeled with a reporter tag.

28. The method of claim 27, wherein said reporter tag is a fluorophore.
29. The method of any one of claims 1 to 28, wherein an affinity tag is attached to said tag and the method further comprises isolating the encoded protein using the affinity tag.
30. The method of claim 29, wherein said affinity tag is biotin.
31. A method for detecting an interaction between a protein and a compound, said method comprising the steps of:
- (a) providing an *in vitro* translated protein immobilized onto a solid support, wherein said *in vitro* translated protein is encoded with a tag that facilitates recognition of the *in vitro* translated protein among a population of proteins, wherein the mRNA encoding said *in vitro* translated protein is degraded;
 - (b) contacting said protein with a candidate compound under conditions which allow an interaction between said protein and said compound; and
 - (c) analyzing said solid support for the presence of said compound as an indication of an interaction between said protein and said compound.
32. The method of claim 31, wherein said compound is selected from the group consisting of a nucleic acid, a protein, a therapeutic, and an enzyme.
33. An *in vitro* translated protein attached to a nucleic acid linker molecule, wherein the nucleic acid linker does not comprise an RNA which encodes said protein, wherein the nucleic acid linker is encoded with a tag.
34. An *in vitro* translated protein attached to a branched nucleic acid linker molecule, wherein the nucleic acid linker does not comprise an RNA which encodes said protein, wherein the branched nucleic acid linker is encoded with a tag.

35. An *in vitro* translated protein attached to a nucleic acid linker, wherein the nucleic acid linker does not comprise an RNA which encodes said protein, and bound to a tag that facilitates recognition of the *in vitro* translated protein among a population of proteins in which mRNA encoding said *in vitro* translated protein is degraded, wherein the tag comprises an addressing element having a sequence that gives the tag a distinct identity.
36. The protein of claim 33, 34 or 35, attached to a solid support bearing a capture probe, which hybridizes in a sequence dependent manner to a unique portion of the tag having a sequence to give the tag an identity distinct from tags bound to other proteins (an addressing element).
37. The protein of claim 36, wherein said protein is attached to said capture probe through hybridization.
38. The protein of claim 36, wherein said protein is attached to said capture probe through a covalent bond.

1/9

translation in microtiter tray --> fusion formation --> RNA degradation

Fig. 1A

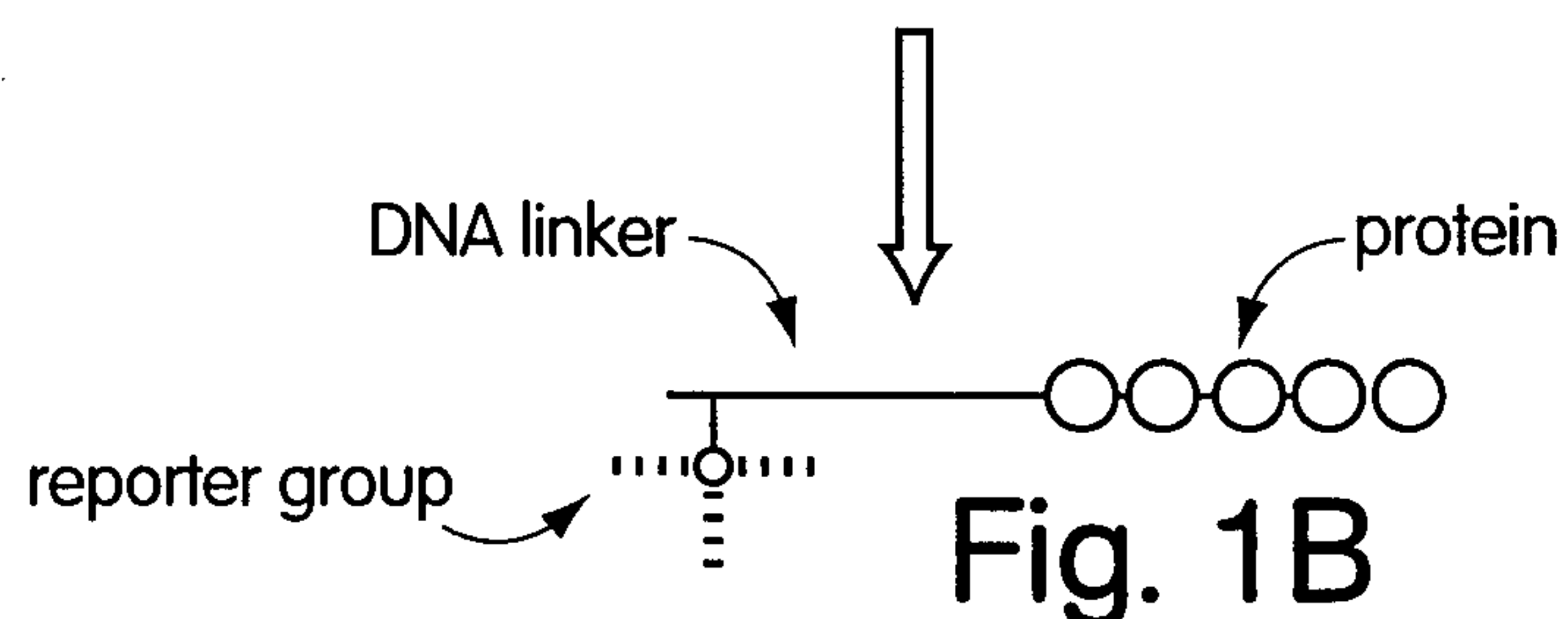


Fig. 1B

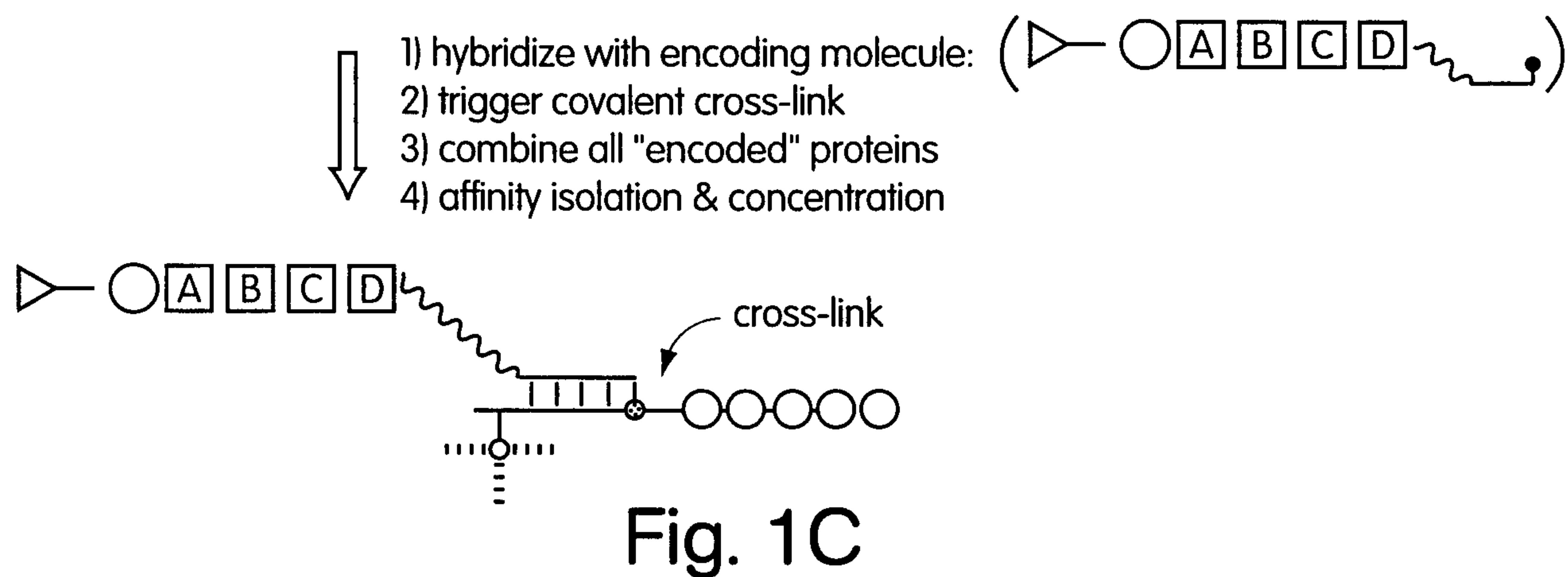


Fig. 1C

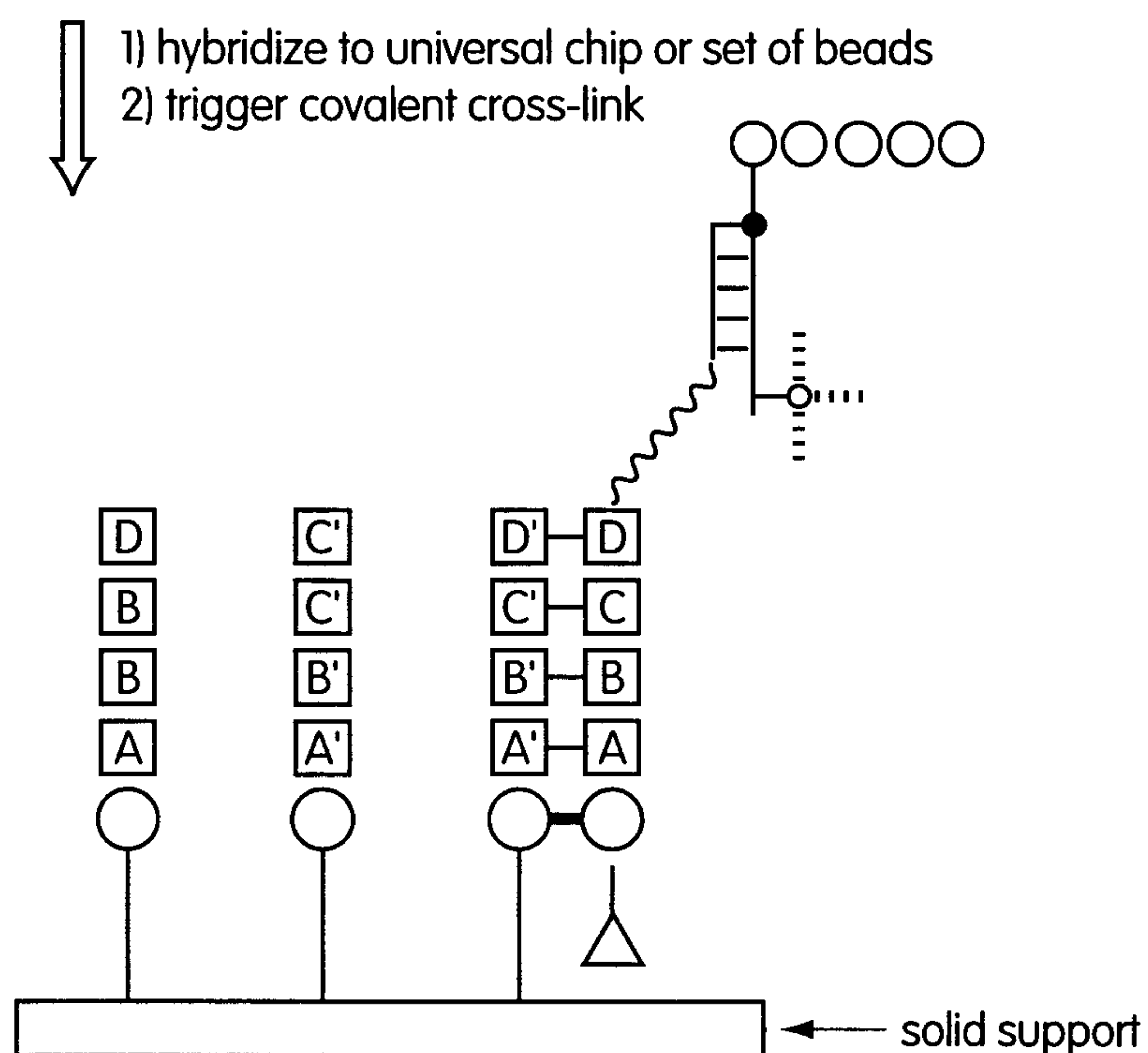


Fig. 1D

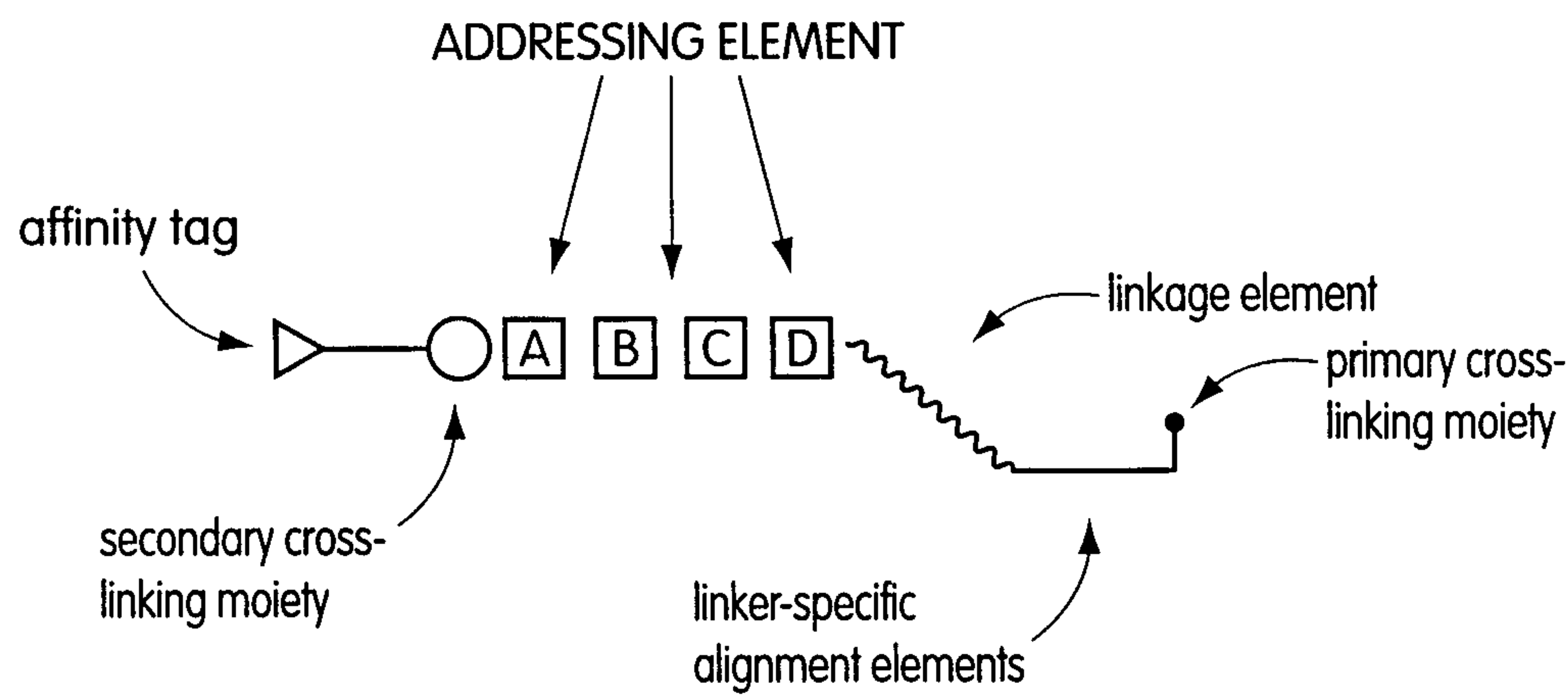


Fig. 2A

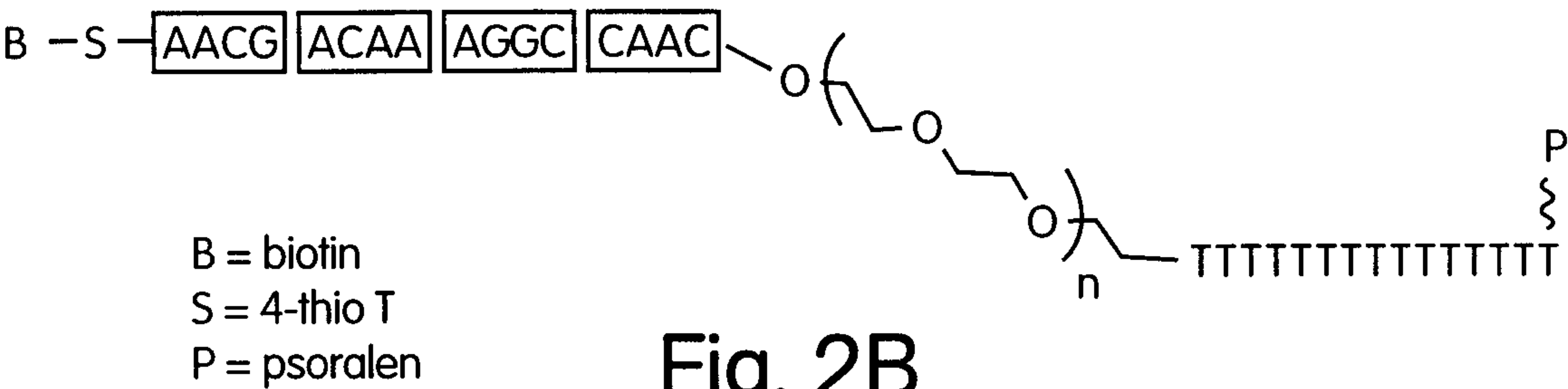
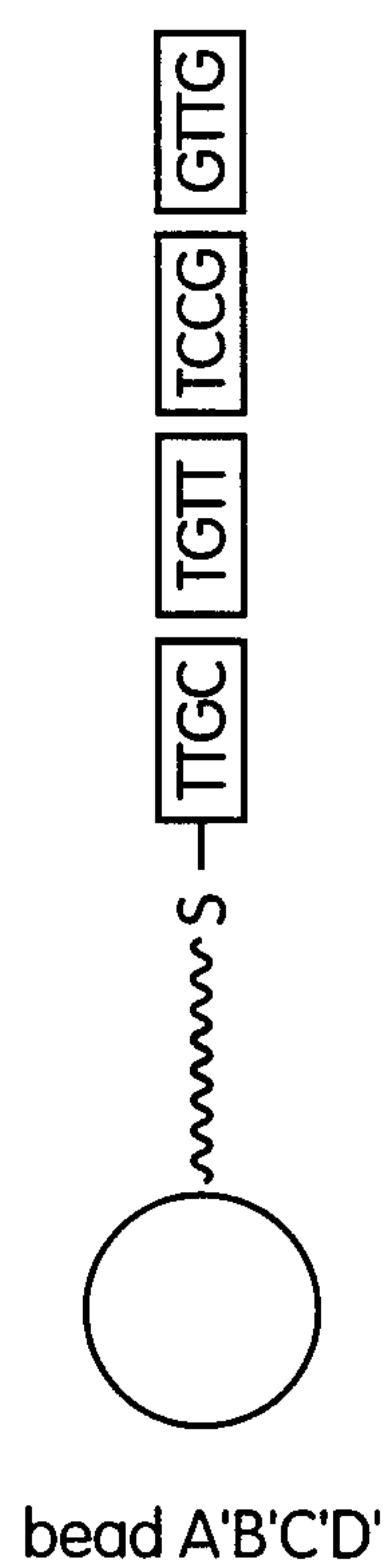


Fig. 2B

3/9



S=4-thio T

Fig. 3A

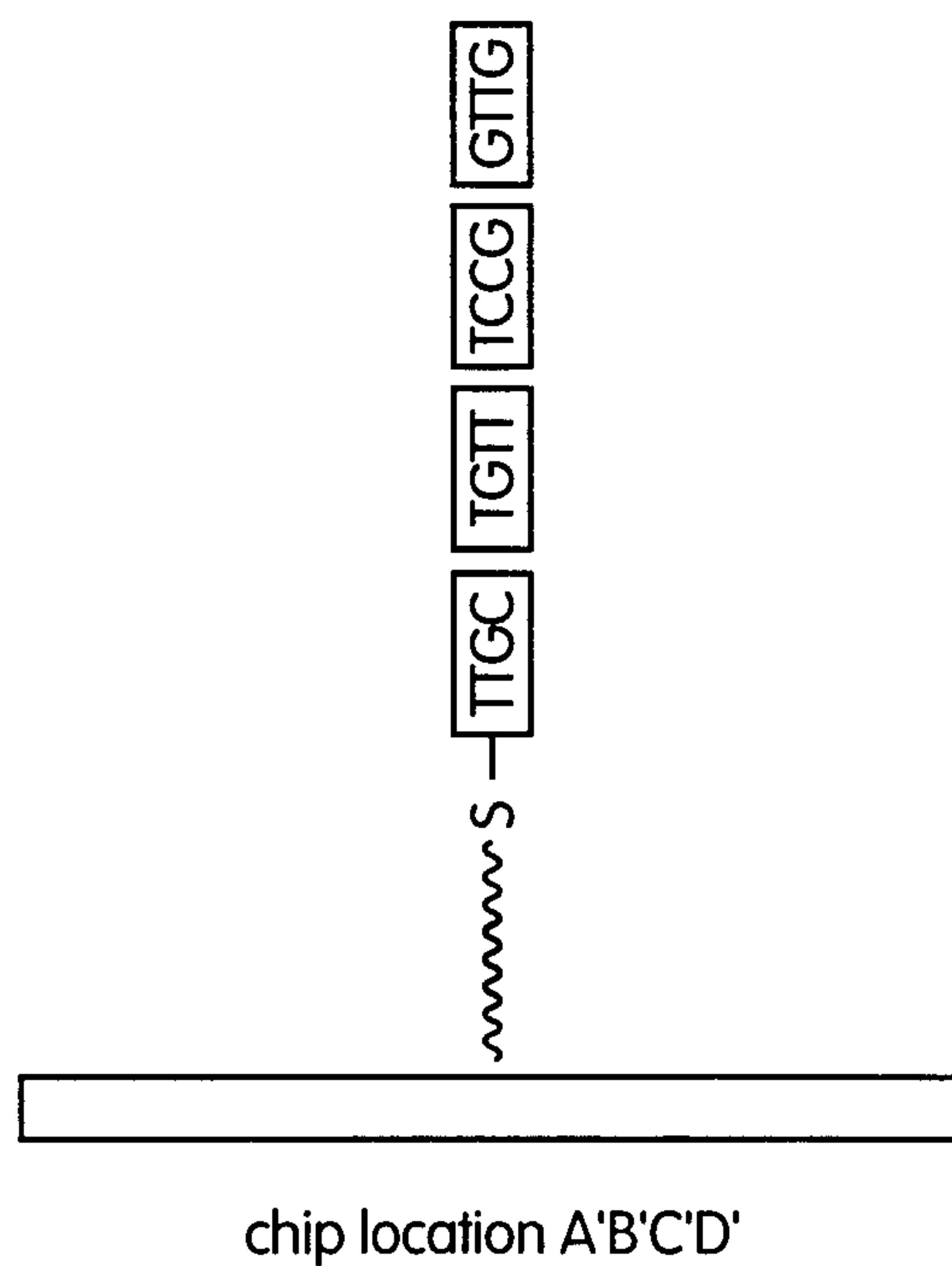


Fig. 3B

4/9

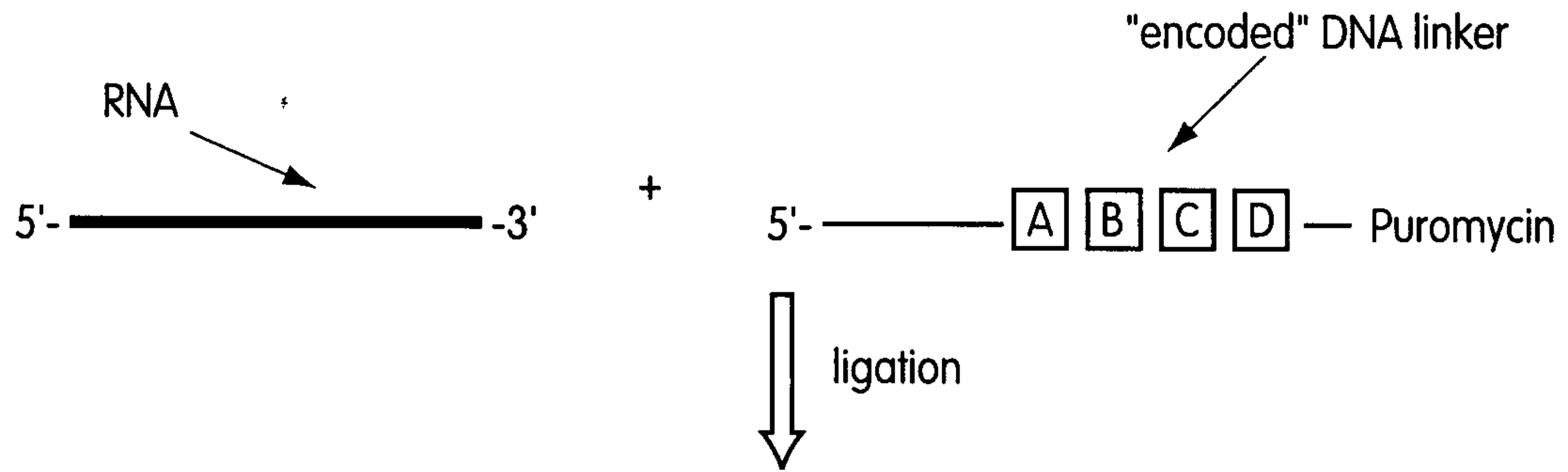


Fig. 4A

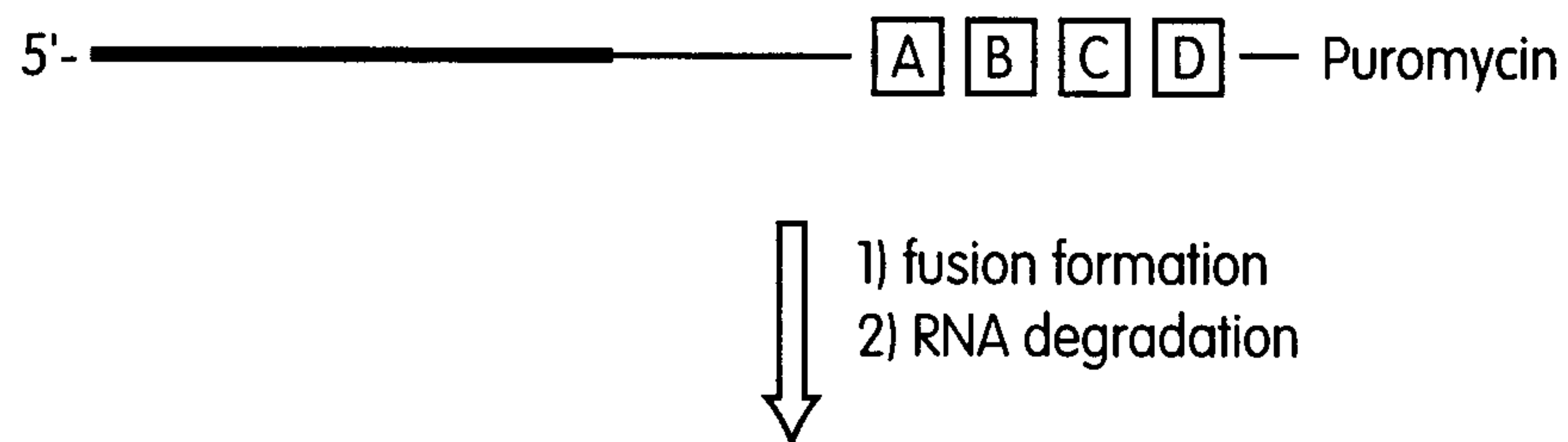


Fig. 4B

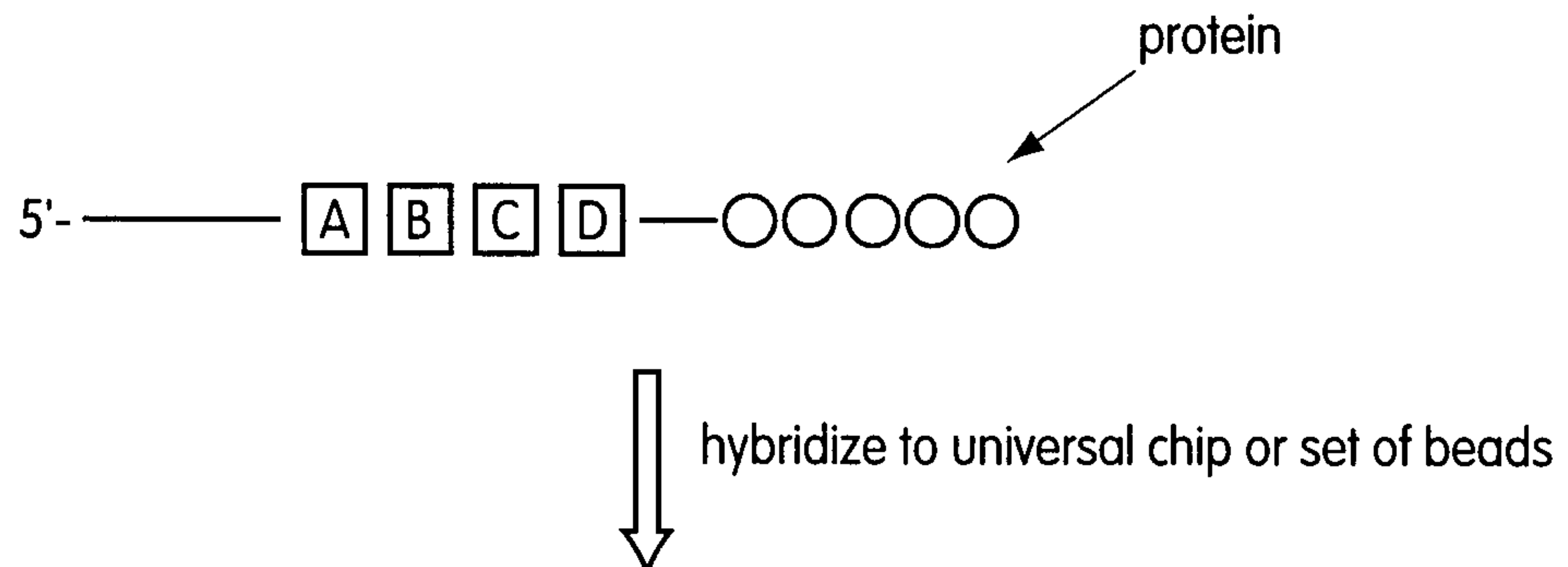
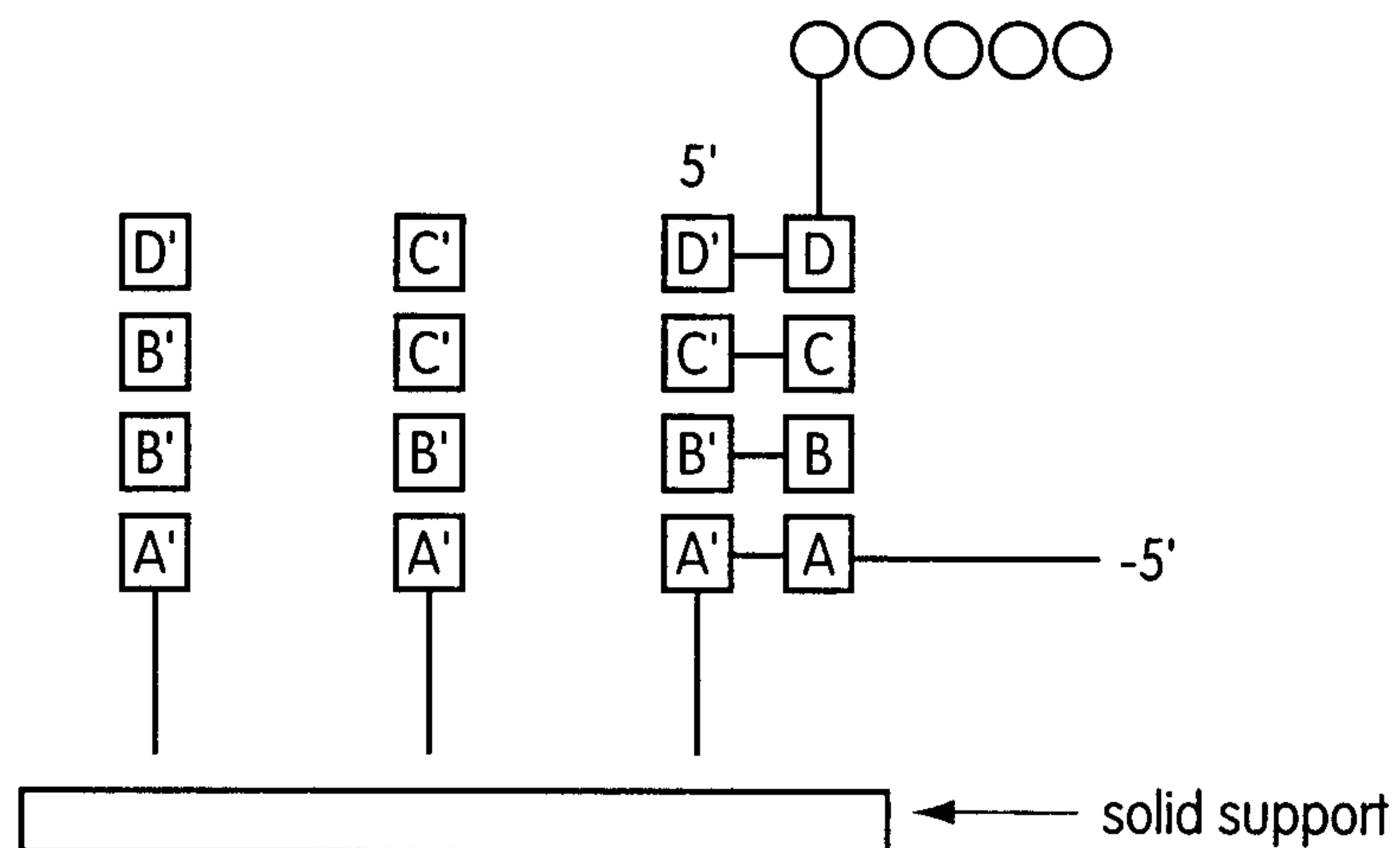


Fig. 4C



5/9

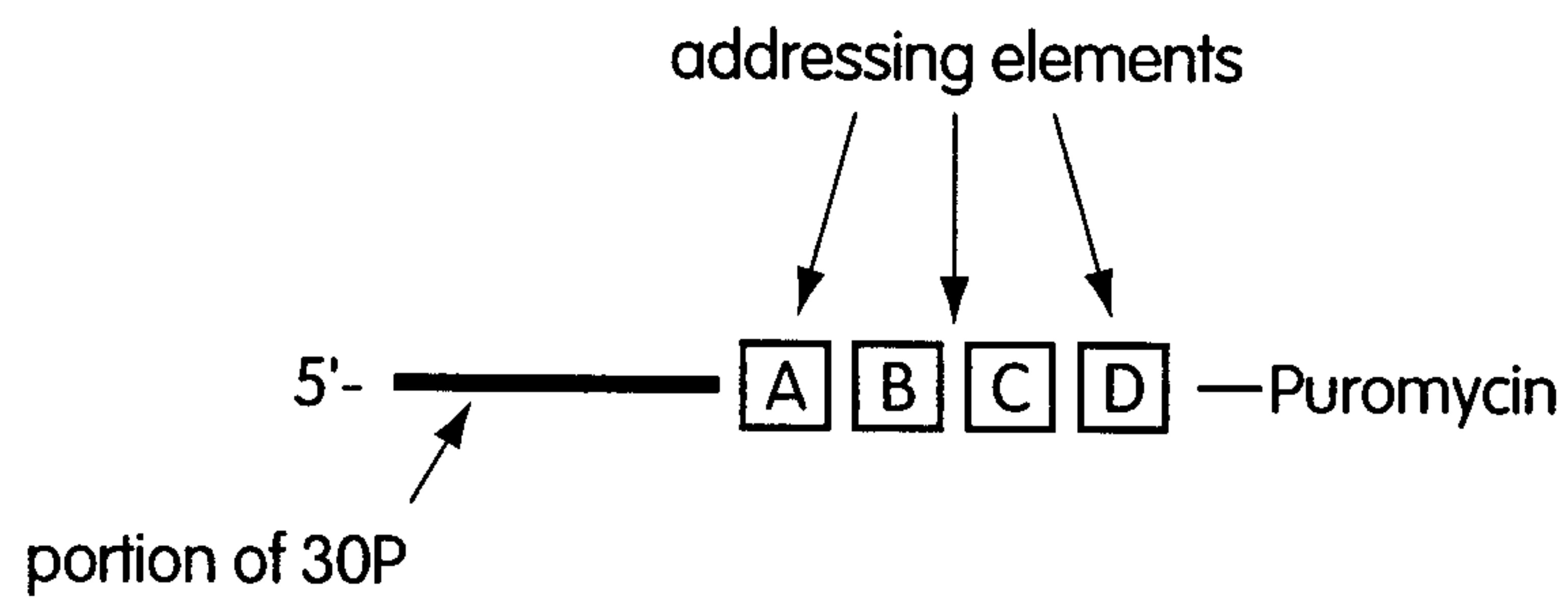


Fig. 5A

5'- AAA AAA AAA AAA AACG ACAA AGGC CAAC CC — Puromycin

Fig. 5B

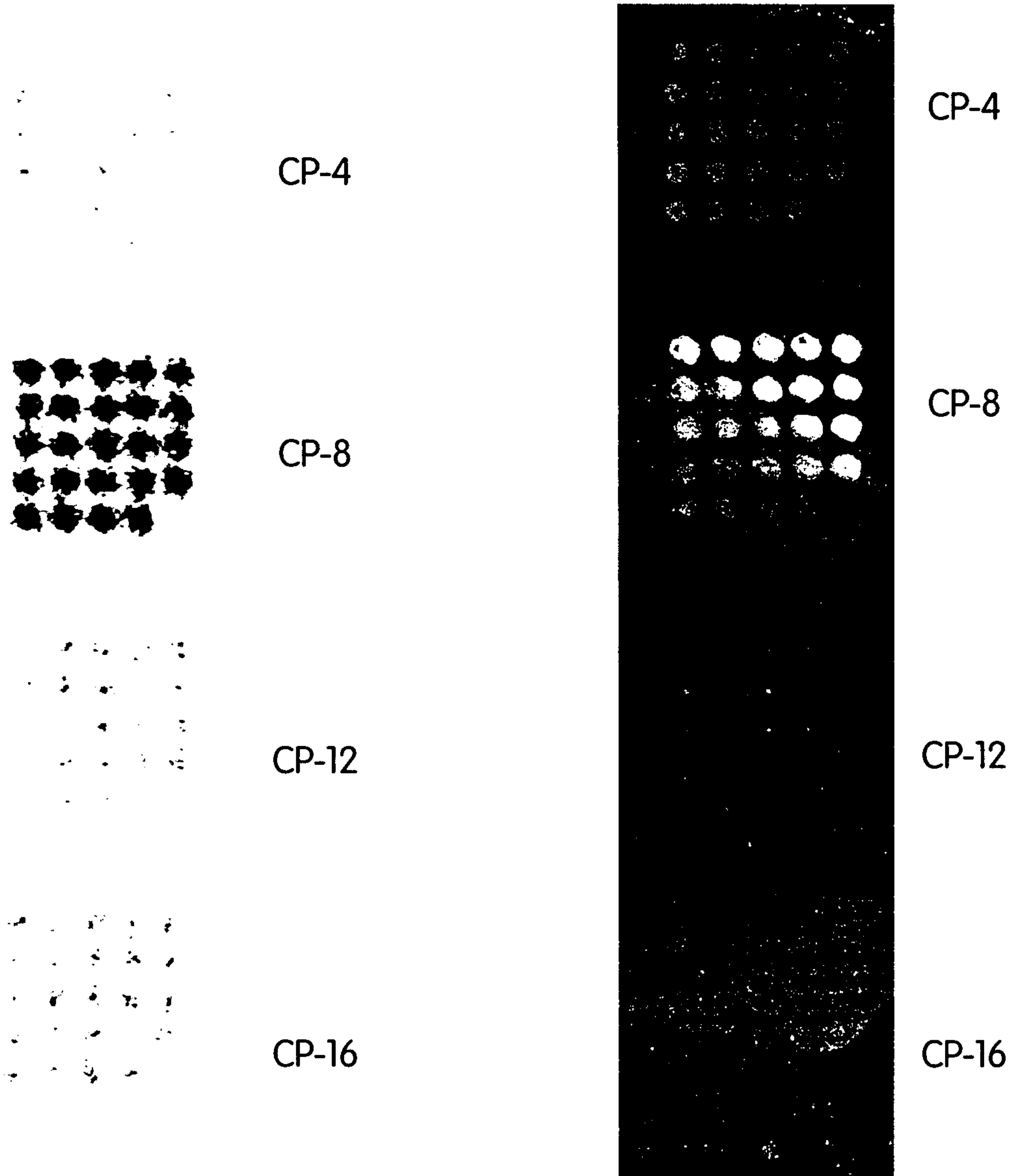


Fig. 6A

Fig. 6B

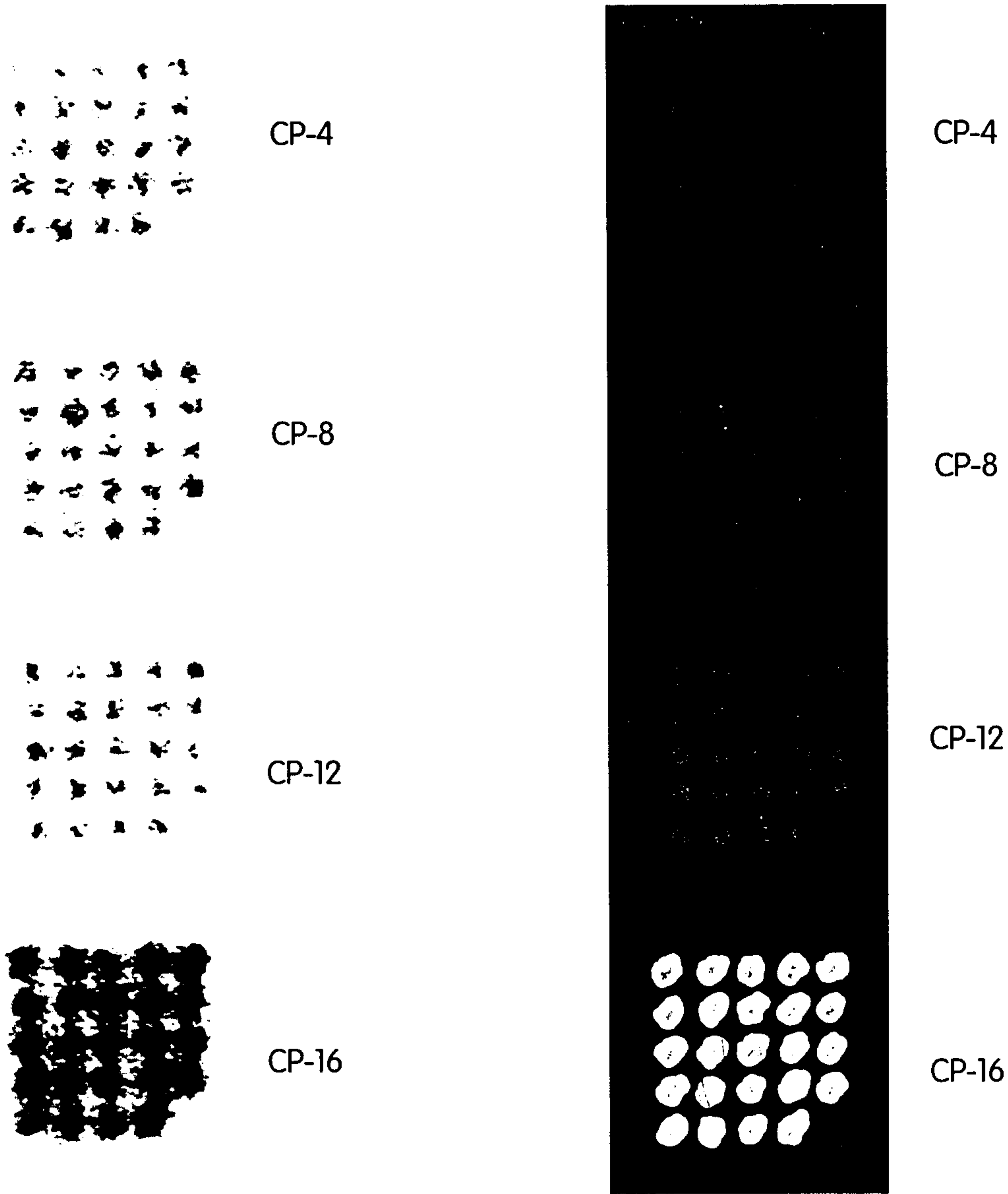


Fig. 7A

Fig. 7B

8/9

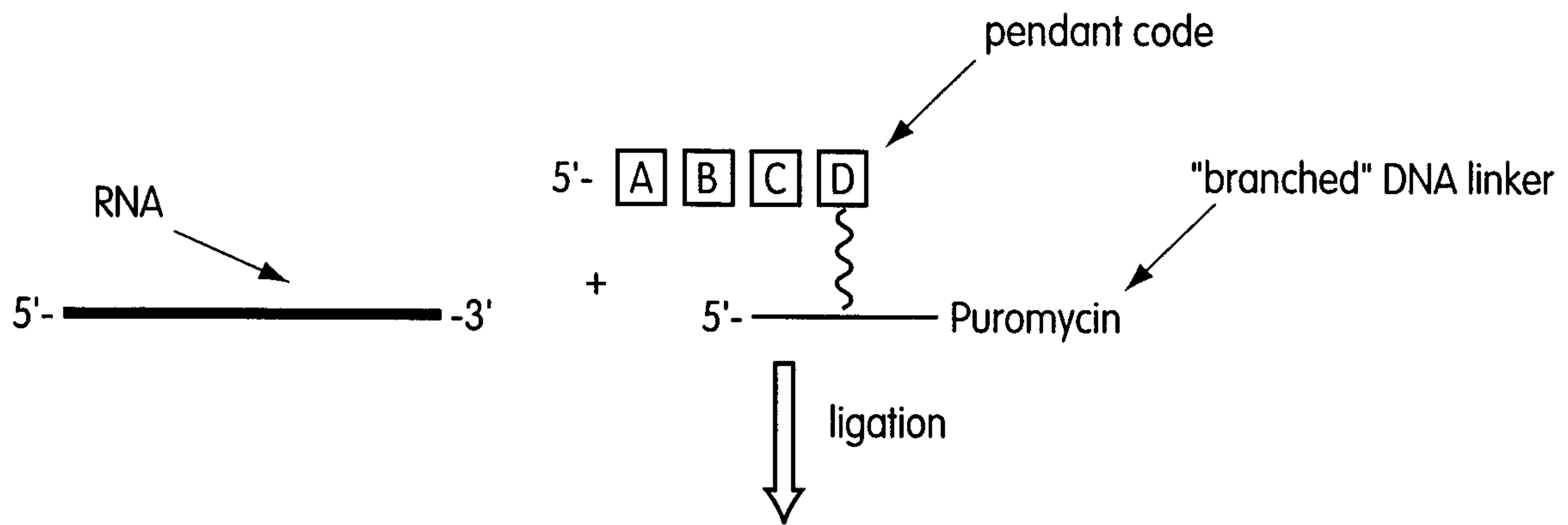


Fig. 8A

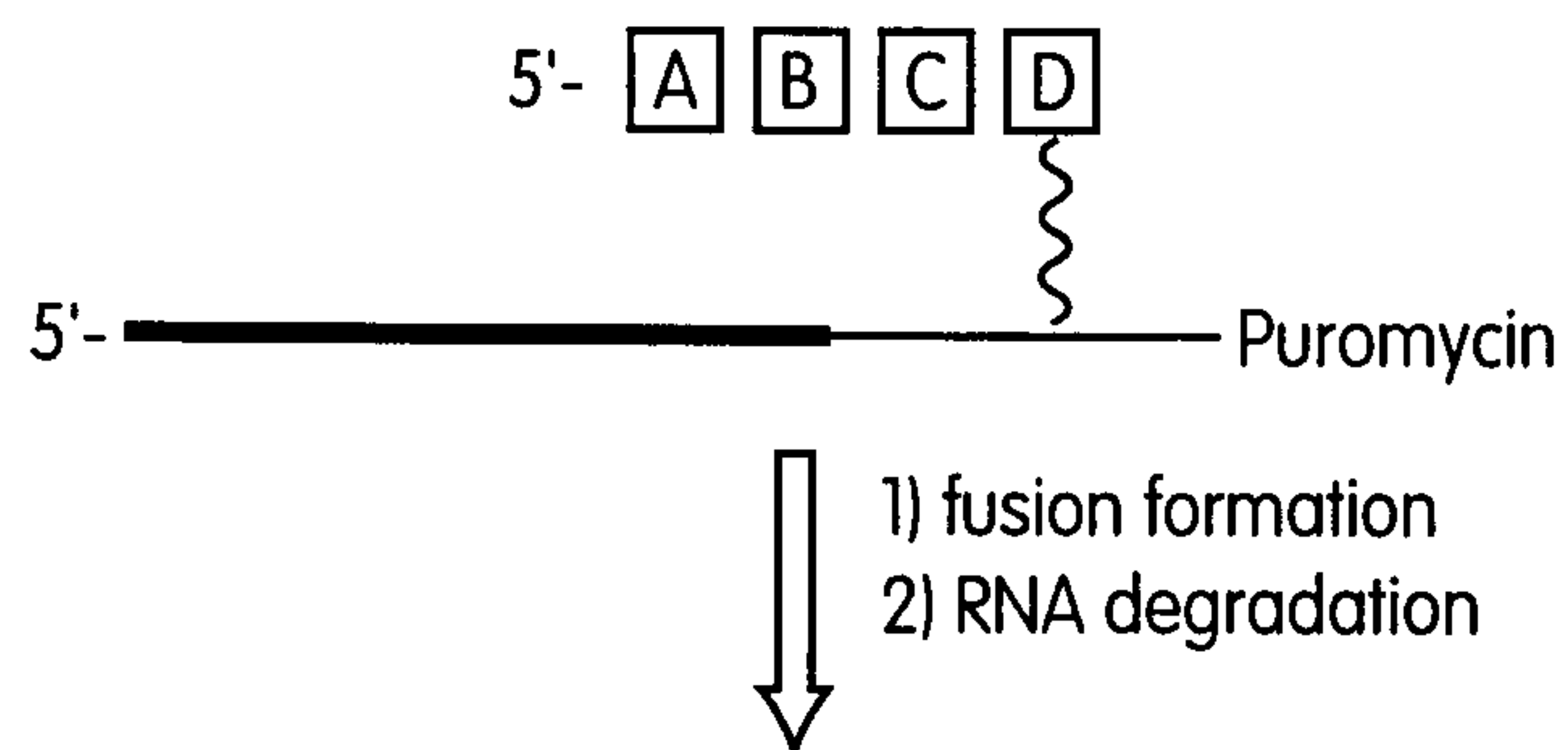


Fig. 8B

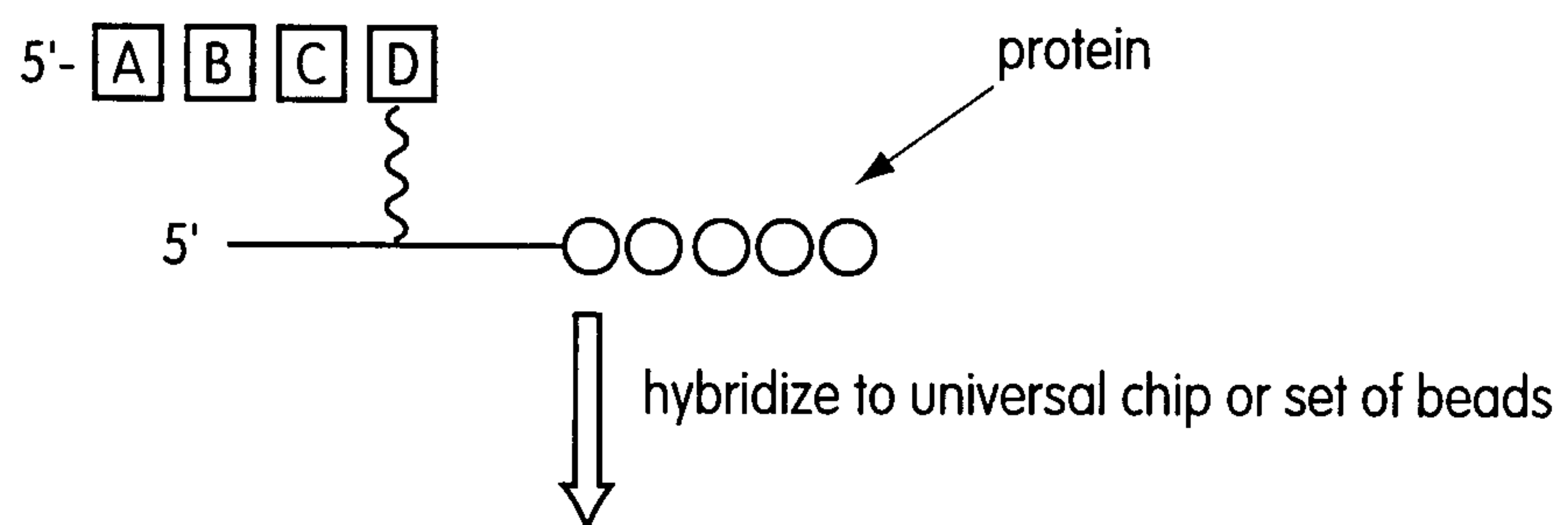
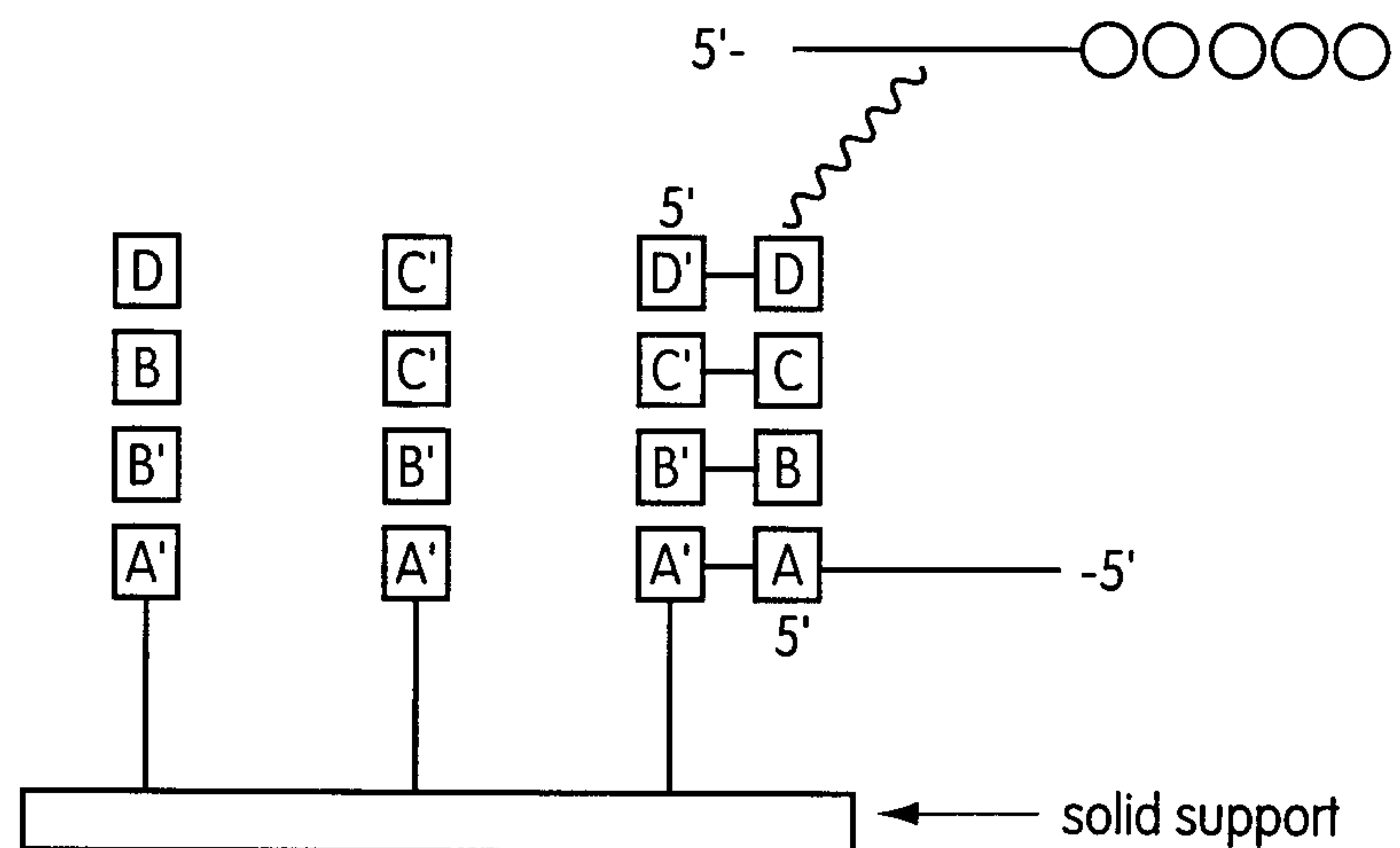


Fig. 8C



9/9

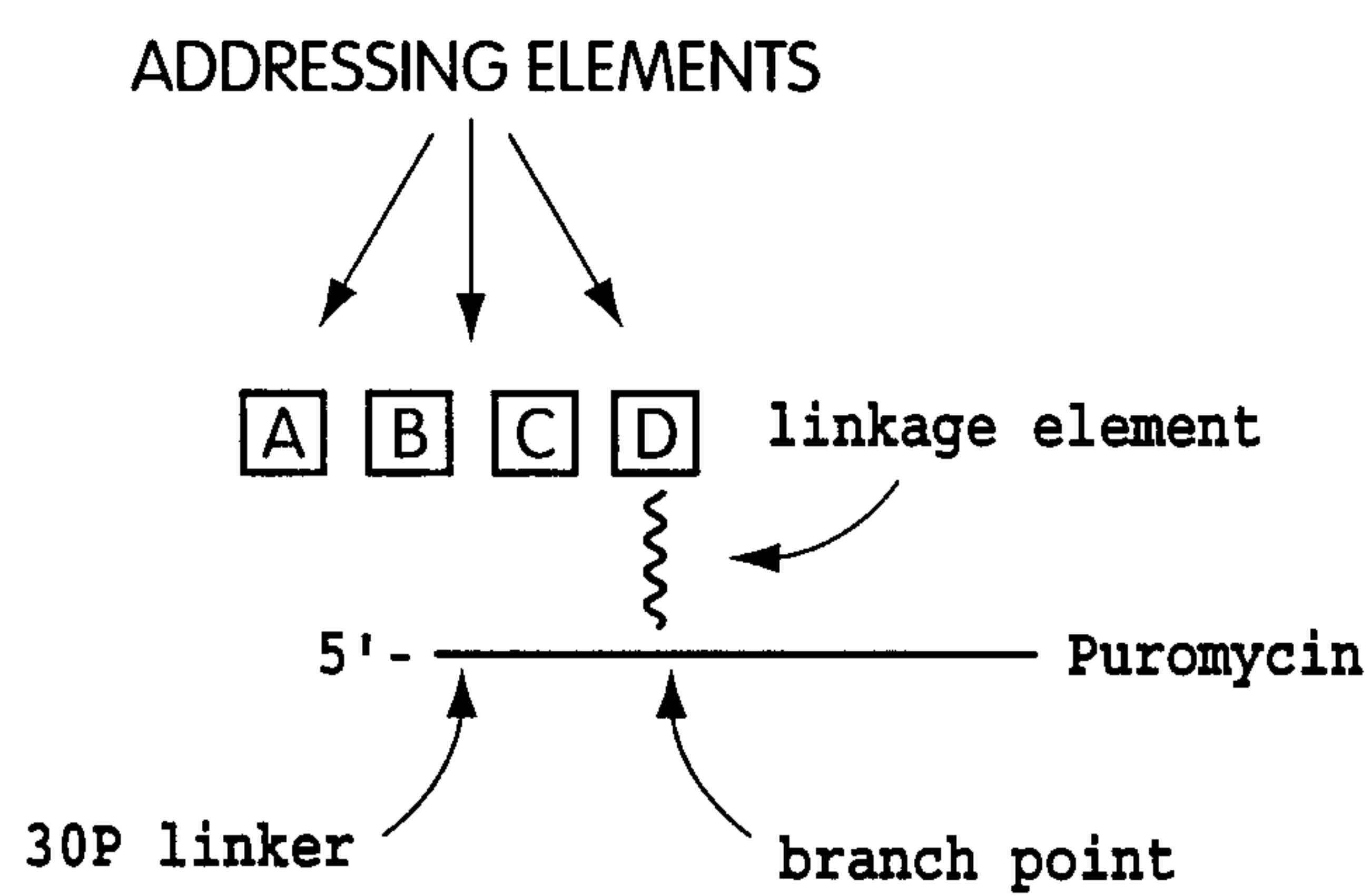


Fig. 9A

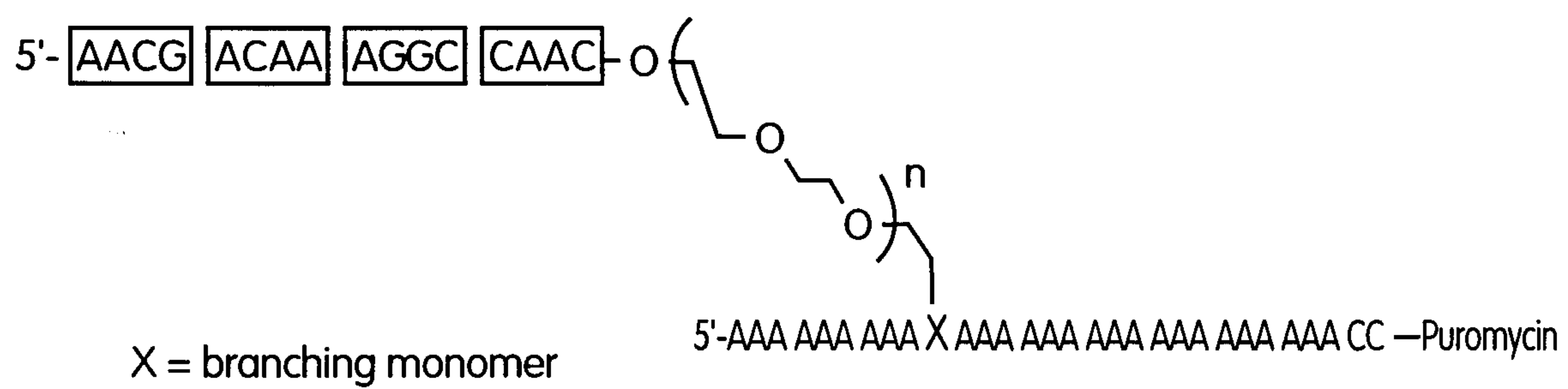


Fig. 9B