United States Patent [19]

Paddock

[11] **4,362,867**

[45] **Dec. 7, 1982**

[54] **RECOMBINANT CDNA CONSTRUCTION METHOD AND HYBRID NUCLEOTIDES USEFUL IN CLONING**

[75] Inventor: **Gary V. Paddock,** Mount Pleasant, S.C.

[73] Assignee: **Research Corporation,** New York, N.Y.

[21] Appl. No.: **214,648**

[22] Filed: **Dec. 10, 1980**

[51] **Int. Cl.**[3] **C07H 15/12;** C12P 21/00; C12P 21/02; C12P 19/34; C12P 15/00; C12N 1/00

[52] **U.S. Cl.** **536/27;** 536/28; 435/68; 435/70; 435/91; 435/172; 435/317

[58] **Field of Search** 435/91; 536/27, 28

[56] **References Cited**

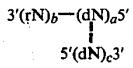
PUBLICATIONS

Salser, in *Genetic Engineering,* Chakrabarty (ed.), CRC Press, West Palm Beach, Fla., pp. 53–81 (1978).
Kavsan et al., Chem. Abstr. 92, 224 123526h (1980).
Romano et al., Chem. Abstr. 91:170822r (1979).
Hartman et al., Chem. Abstr. 87:147246q (1977).
Waqar et al., Chem. Abstr. 84:100898b (1976).
Ogawa et al., Chem. Abstr. 87, 17513n (1977).

Primary Examiner—Alvin E. Tanenholtz
Assistant Examiner—James Martinell
Attorney, Agent, or Firm—Cooper, Dunham, Clark, Griffin & Moran

[57] **ABSTRACT**

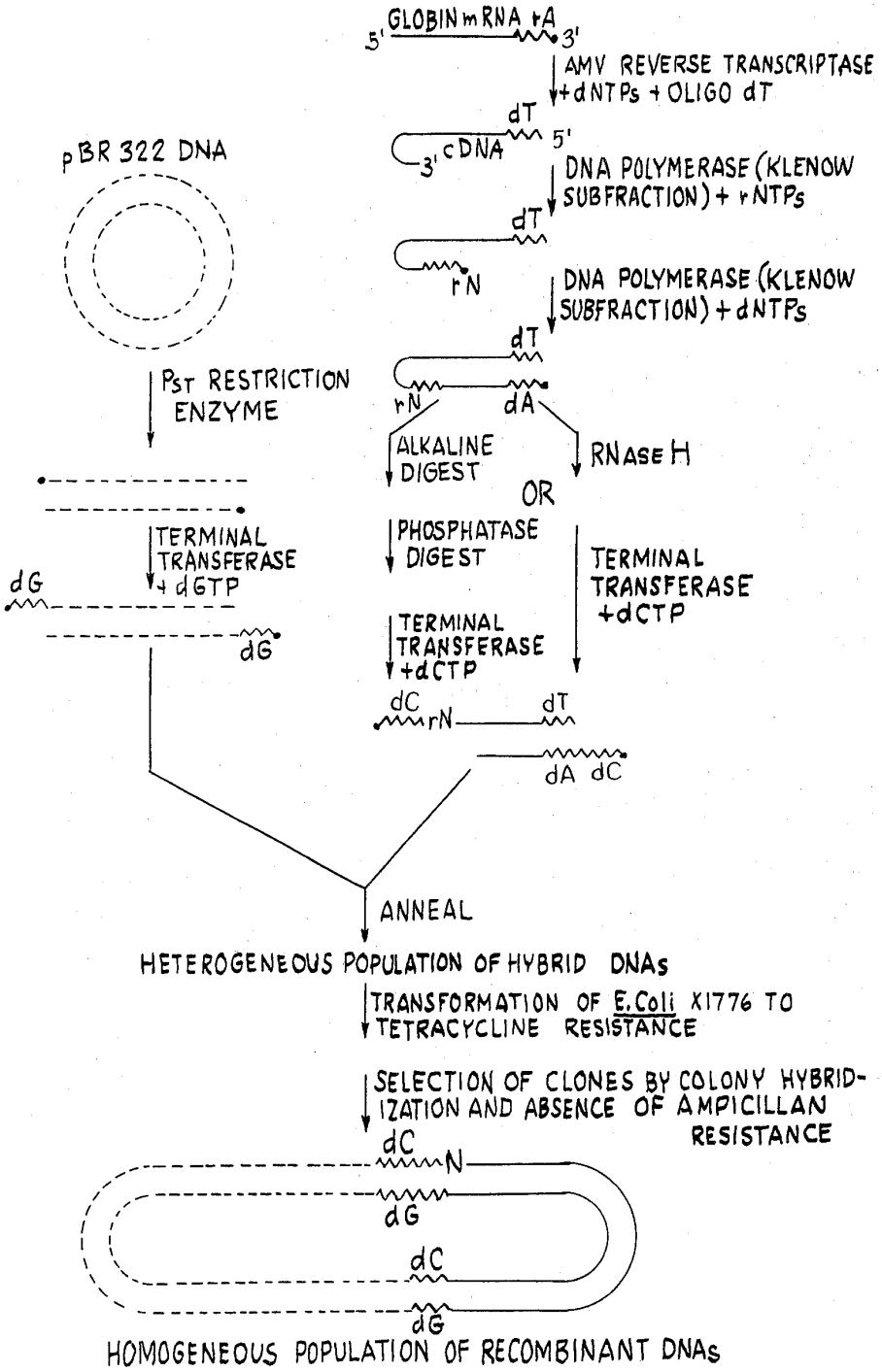
Compounds useful as complementary DNA (cDNA) include deoxyribonucleotides and at least one ribonucleotide. They may be depicted by the general formula:



wherein $(dN)_a$ and $(dN)_c$ represent series of deoxyribonucleotides and $(rN)_b$ represents a series of ribonucleotides; wherein a, b, and c are the number of nucleotides in the series, with the proviso that b is $\geqq 1$, a is $\geqq 35$, and c is $\geqq 10$; wherein the series of deoxyribonucleotides $(dN)_a$ includes a series of deoxyribonucleotides which is substantially complementary to the series of deoxyribonucleotides $(dN)_c$ and the dashed line represents noncovalent bonding between the complementary deoxyribonucleotide series; and wherein the solid line represents a covalent phosphodiester bond.

These compounds may be prepared from messenger RNA (mRNA) containing the genetic information necessary for cellular production of desired products such as polypeptides. After appropriate modification, they may be combined with DNA from a suitable cloning vehicle such as a plasmid and the resulting combined DNA used to transform bacterial cells. The transformed bacterial cells may then be grown and harvested; and the desired product or products recovered.

4 Claims, 1 Drawing Figure



RECOMBINANT CDNA CONSTRUCTION METHOD AND HYBRID NUCLEOTIDES USEFUL IN CLONING

The invention described herein was made in the course of work partly supported by Grant Number GM-24783 from the National Institute of Health, Department of Health and Human Services.

FIELD OF THE INVENTION

This invention generally concerns the preparation of recombinant complementary DNAs (cDNAs) and cDNA analogs coding for cellular production of desirable products such as polypeptides. It also concerns novel compounds which include both deoxyribonucleotides and ribonucleotides. Finally, it concerns the use of such compounds in bacterial cloning.

BACKGROUND OF THE INVENTION

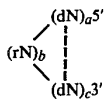
One of the major areas of research in molecular biology today concerns gene organization and expression in eukaryotic cells. Much effort has been spent on studies of RNA transcription and its subsequent processing to mRNA. It is currently thought that the genome sequences surrounding the cap site contain the signal for initiation of mRNA transcription or, alternatively, that very rapid processing cleaves away the first few nucleotides followed by capping at the 5′ end. [Konkel, D. A., Tilghman, S. M., and Leder, P., (1978), Cell, 15: 1125–1132; Konkel, D. A., Maizel, J. V., Jr., and Leder, P., (1979), Cell, 18: 865–873; Gannon, F., O'Hare, K., Perrin. F., LePennec, J. P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B., and Chambon, P., (1979), Nature, 278: 428–434; Nishioka, Y. and Leder, P., (1979), Cell, 18: 875–882; and Kinniburgh, A. J. and Ross, J., (1979), Cell, 17: 915–921.] In either case, the nucleotide sequences contained in the 5′ untranslated regions of mRNA, especially those near the cap site, are of prime importance to proper gene regulation, as illustrated by the extensive conservation of sequences found in this region for alpha and beta globin and other mRNA species. [Konkel, D. A., Tilghman, S. M., and Leder, P., (1978), Cell, 15: 1125–1132; Konkel, D. A., Maizel, J. V., Jr., and Leder, P., (1979), Cell, 18: 865–873; and Lockard, R. E. and RajBhandary, U. L., (1976), Cell, 9: 747–760.] In addition to transcription and processing of mRNA, these sequences undoubtedly play an important role in the translation of protein from mRNA. Indeed, the importance of the nucleotides contained in the 5′ untranslated regions of mRNA is emphasized by the variety of methods designed to sequence them. [Lockard, R. E. and RajBhandary, U. L., (1976), Cell, 9: 747–760; Baralle, F. E., (1977), Cell 10: 549 –558; Baralle, F. E., (1977), Nature, 267: 279–281; Baralle, F. E., (1977), Cell, 12: 1085–1095; Legon, S., (1976). J. Mol. Biol., 106: 37–53; Chang, J. C., Temple, G. F., Poon, R., Neumann, K. H. and Kan Y. W., (1977)., Proc. Natl. Acad. Sci. U.S.A., 74: 5145–5149; and Chang, J. C., Poon, R., Neumann, K. H. and Kan, Y. W., (1978), Nucl. Acids. Res., 5: 3515–3522.] Yet none of these methods permits sequencing of the 5′ end of an impure mRNA obtained in low yield as is the case for most mRNAs. Furthermore, none of the cloning techniques developed thus far, (Higuchi, R., Paddock, G. V., Wall, R., and Salser, W., (1976), Proc. Natl. Acad. Sci. U.S.A., 73: 3146–3150; Maniatis, T., Kee, S. G., Efstratiadis, A., and Kafatos, F. C., (1976), Cell, 8: 163–182; Rougeon, F., Kourilski, P., and Mach, B., (1975), Nucl. Acids Res. 2: 2365–2378; Efstratiadis, A., Kafatos, F. C., and Maniatis, T., (1977), Cell, 10: 571–585; Rabbits, T. H., (1976), Nature, 260: 221–225; Rougeon, F. and Mach, B., (1976), Proc. Natl. Acad. Sci. U.S.A., 73: 3418–3422; and Wood, K. O. and Lee, J. C., (1976), Nucl. Acids Res., 3: 1961–1971.] have been successful in preserving these important terminal sequences. In fact, the most popular of these techniques is destined to destroy these sequences in part, because it relies upon use of Sl nuclease. [Higuchi, R., Paddock, G. V., Wall, R., and Salser, W., (1976), Proc. Natl. Acad. Sci. U.S.A., 73: 3146–3150.]

In order to preserve these important 5′-end signals, efforts have been undertaken to develop methodology which avoids the need for Sl nuclease. [Frankis, R., Gaubatz, J., Lin, F. K., and Paddock, G. V., The Twelfth Miami Winter Symposium (ed. Whelan, W. J., and Schultz, J., Academic Press, New York), vol. 17, in press (1980); and Gaubatz, J. and Paddock, G. V., (1980), Fed. Proc., 39: 1782.] These efforts have resulted in the discovery of the floppy loop method described herein. This method employs a ribosubstitution step so that cleavage of the hairpin loop can be carried out by alkali or ribonuclease. It avoids destruction of nucleotide sequence information which is lost if the hairpin is opened in the conventional manner with Sl nuclease. Thus, by elimination of the Sl nuclease step, whole genes can be synthesized without loss of genetic information. Moreover, the Sl nuclease technique is known to introduce errors in the sequence [Richards, R. I., Shine, J., Ulbrich, A., Wells, J. R. E., and Goodman, H. M., (1979), Nucl. Acids Res. 7: 1137–1146] through a mechanism which the present invention avoids. Finally, although it has been demonstrated that hormones (insulin) and interferon can be cloned via recombinant cDNA, it may not be possible to clone some genes in their entireties with the Sl nuclease technique because the hairpin loop may be extremely large and may even include part of the structural gene (i.e., part of the mRNA coding for protein.)

SUMMARY OF THE INVENTION

This invention provides various compounds which include both deoxyribonucleotides and at least one ribonucleotide. Certain of these compounds are useful in the preparation of cDNA and cDNA analogs. Others are useful in bacterial cloning. This invention also provides processes for preparing such molecules and using them in the production of desirable products such as polypeptides.

Specifically, compounds useful in the preparation of cDNAs and cDNA analogs may be prepared. These compounds may be depicted by the formula:



I

wherein $(dN)_a$ and $(dN)_c$ represent series of deoxyribonucleotides and $(rN)_b$ represents a series of ribonucleotides; wherein a, b, and c are numbers of nucleotides in the series provided that b is $\geqq 1$, a is $\geqq 35$, and c is $\geqq 10$; wherein the series of deoxyribonucleotides and $(dN)_a$ includes a series of deoxyribonucleotides which is substantially complementary to the series of deox-

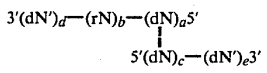
yribonucleotides $(dN)_c$ and the dashed line represents non-covalent bonding between the complementary deoxyribonucleotide series; and wherein the solid lines represent covalent phosphodiester bonds. Such compounds may be prepared as follows. A first molecule having the formula $3'(dN)_a5'$ is prepared. At least one ribonucleotide is added to the 3'-end of this molecule to produce a molecule having the formula $3'(rN)_b$—$(dN)_a5'$, and additional deoxyribonucleotides are then added to the 3'-end of the latter to produce the compounds.

If these compounds are treated with a reagent capable of breaking or disrupting either 5'(rN)—(dN)3' or (rN)—(rN) bonds, or both, compounds useful as cDNA analogs or precursors may be prepared having the formula:

II

wherein $(dN)_a$ and $(dN)_c$ represent series of deoxyribonucleotides and $(rN)_b$ represents a series of ribonucleotides; wherein a, b, and c are the numbers of nucleotides in the series; provided that b is $\geqq 1$, a is $\geqq 35$, and c is $\geqq 10$; wherein the series of deoxyribonucleotides $(dN)_a$ includes a series of deoxyribonucleotides which is substantially complementary to the series of deoxyribonucleotides $(dN)_c$ and the dashed line represents non-covalent bonding between the complementary deoxyribonucleotide series; and wherein the solid line represents a covalent phosphodiester bond.

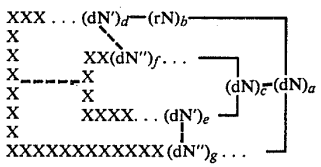
Further compounds having the formula:



III

may be prepared, wherein $(dN)_a$, $(dN)_c$, and $(rN)_b$ are as above and $(dN')_d$ and $(dN')_e$ represent series of identical deoxyribonucleotides; wherein a, b, and c are as above, and d and e are integral numbers of nucleotides in the series and are $\geqq 10$; and wherein the solid lines represent covalent phosphodiester bonds. These compounds are prepared by adding deoxyribonucleotides dN' to the 3'-ends of compounds II.

Still other compounds may be prepared having the formula:



IV

wherein $(dN)_a$ and $(dN)_c$ represent series of deoxyribonucleotides, $(dN')_d$, $(dN')_e$, $(dN'')_f$ and $(dN'')_g$ represent series of identical deoxyribonucleotides, the series $(dN')_d$ and $(dN')_e$ being complementary to the series $(dN'')_f$ and $(dN'')_g$ respectively, and $(rN)_b$ represents a series of ribonucleotides; wherein, a, b, c, d, e, f, and g are integral numbers of nucleotides in the series; provided that b is $\geqq 1$, a is $\geqq 35$, and c, d, e, f, and g are $\geqq 10$; wherein the series of deoxyribonucleotides $(dN)_a$ includes a series of deoxyribonucleotides which is substantially complementary to the series of deoxyribonucleotides $(dN)_c$; wherein the xxx- lines represent double-stranded DNA derived from a cloning vehicle such as a plasmid, bacteriophage, or virus; wherein the dotted lines may be either no bond or covalent bonds; wherein the dashed lines represent non-covalent bonding between complementary deoxyribonucleotide series; and wherein the solid lines represent covalent phosphodiester bonds.

These molecules may be used to transform bacterial or eucaryotic cells, e.g., *Escherichia coli* cells, which may be grown in culture to produce desired products, including polypeptides.

BRIEF DESCRIPTION OF THE DRAWING

FIG. **1** is a schematic illustration of the ribosubstitution floppy loop recombinant cDNA technique showing alternative means for cleavage of the ribosubstituted hairpin double-stranded cDNA.

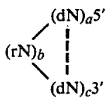
DETAILED DESCRIPTION OF THE INVENTION

Much publicity has been given to recent efforts to employ genetic engineering, particularly recombinant DNA technology, in the production of useful products such as insulin, interferon, growth hormone and the like. These attempts have often involved the insertion of a DNA molecule containing the genetic information necessary for cellular production of the desired product into host cells, especially bacterial cells. Bacterial cells into which such DNA molecules have been inserted can be grown in culture, resulting in the production of increased quantities of recoverable products.

One limitation upon such efforts is the availability of appropriate DNA molecules. Approaches which have been pursued to obtain these DNA molecules include synthesis of the molecules by conventional chemical methods and by reverse transcription of mRNA molecules which have been recovered from cells which contain DNA sequences coding for the desired product.

This invention is directed to improved methods of preparing these DNA molecules, called complementary DNA (cDNA). It is also directed to the preparation of novel hybrid molecules which include both deoxyribonucleotides and at least one ribonucleotide. Certain of these molecules are useful in the preparation of cDNA and cDNA analogs. Others are useful in bacterial cloning.

One such compound may be represented by the general formula:



I

wherein $(dN)_a$ and $(dN)_c$ represent series of deoxyribonucleotides and $(rN)_b$ represents a series of ribonucleotides; wherein a, b, and c are the numbers of nucleotides in the series; wherein b is $\geqq 1$, a is $\geqq 35$, and c is $\geqq 10$; wherein the series of deoxyribonucleotides $(dN)_a$ includes a series of deoxyribonucleotides which is substantially complementary to the series of deoxyribonucleotides $(dN)_c$ and the dashed line represents non-covalent bonding, particularly hydrogen bonding, between complementary deoxyribonucleotide series; and wherein the solid lines represent covalent phosphodiester bonds.

In such molecules the deoxyribonucleotide series $(dN)_a$ is an ordered polymer of deoxyribonucleotides which include the purines, adenine and guanine, and the pyrimidines, thymine and cytosine. The particular order of deoxyribonucleotides contains information necessary for cellular production of a desired product in accordance with the established genetic code, whereby groups of three nucleotides correspond to single amino acids which are assembled by cells into polypeptides. The precise order of the nucleotides may vary widely, depending upon the product for which the series codes. However, in the aforementioned compound I, the number of nucleotides, a, must generally be $\geqq 35$ since fewer nucleotides do not permit sufficient formation of non-covalent bonding between complementary nucleotides in the series $(dN)_a$ and $(dN)_c$. More often, the number of nucleotides will be even greater, varying from as few as about 60 in the case of DNA sequences which code for oligopeptides, to about 10^3 nucleotides for an average protein containing about 300 amino acids, and to $\geqq 10^5$ for particularly large polypeptides.

The series $(dN)_c$ is also an ordered polymer, of at least 10 deoxyribonucleotides. The order of nucleotides is such that the nucleotide series $(dN)_c$ is complementary to an equal number of nucleotides in the series $(dN)_a$. In general, the length of the deoxyribonucleotide polymer $(dN)_c$ will be about 25 nucleotides shorter than the length of the series $(dN)_a$. The approximately 25 nucleotide difference may be attributed to nucleotides in the series $(dN)_a$ which are not base-paired with nucleotides in the series $(dN)_c$, but are involved in formation of a folded segment of nucleotides known to those skilled in the art as a hairpin structure. The minimum length of about ten nucleotides is necessary to permit sufficient non-covalent, hydrogen bonding between complementary nucleotides in the $(dN)_c$ and $(dN)_a$ polymer strands to form a double-stranded helical structure.

Situated between and covalently bonded to the series $(dN)_a$ and $(dN)_c$ is a series of ribonucleotides $(rN)_b$, wherein b is the integral number of nucleotides in the series and is $\geqq 1$. It will be appreciated that, when $b=0$, the resulting molecule is a homopolymer of deoxyribonucleotides. Although b may vary considerably, it will generally be less than about 50 and oftentimes less than about 20. Moreover, in practicing the invention, it may be preferable to first form a molecule having the formula $3'(rN)_b—(dN)_a5'$, wherein b is $\geqq 1$, then remove ribonucleotides until $b=1$, and finally add deoxyribonucleotides $(dN)_c$ to form molecules I, wherein $b=1$. In all molecules I, the ribonucleotide series $(rN)_b$ is covalently joined to the series $(dN)_a$ and $(dN)_c$ by means of $5'\rightarrow 3'$ and $3'\rightarrow 5'$ phosphodiester bonds, respectively.

Although the experiments described hereinafter involve globin polypeptides, it will be readily understood by those skilled in the art that the practices of this invention are widely applicable to polypeptides generally and to other desirable products. Thus, the deoxyribonucleotide series $(dN)_a$ may contain information in the form of its ordered nucleotide sequence for cellular production of any desired product, e.g., proinsulin, the polypeptide A chain of insulin, the polypeptide B chain of insulin, a growth hormone, an enzyme, a clotting factor, an antibody, or the polypeptide portion of one of the interferon glycoproteins. These examples are set forth to illustrate some of the better known commercial products which may be prepared in accordance with the present invention, but are not intended to be limiting, since countless other products may also be prepared.

In general, the deoxyribonucleotide series $(dN)_a$ will be obtained by reverse transcription of a messenger RNA (mRNA) molecule which itself will have been obtained by standard methods from a natural source, such as a eukaryotic cell known to contain a gene or genes coding for or otherwise associated with production of the desired product. However, it is also contemplated that the series $(dN)_a$ might be directly synthesized to create a polynucleotide having a sequence coding for a desired product.

Compounds I may be prepared as follows. Initially, a first molecule having the formula $3'(dN)_a5'$ is prepared either by conventional chemical synthesis, or preferably by reverse transcription of a mRNA molecule corresponding thereto. One method of accomplishing this result is to treat the mRNA molecule with a suitable enzyme, such as AMV reverse transcriptase, and a mixture of the deoxyribonucleotides, dATP, dCTP, dGTP, and dTTP, under appropriate conditions which permit formation of the $(dN)_a$ molecule. Suitable conditions are well known in the art and may include: a temperature of about 25°–45° C., e.g., 37° C.; a buffer having a pH of about 7.0 to 9.0, e.g., 8.3; a catalytic amount of enzyme; and a molar excess of the deoxyribonucleotide triphosphates.

In one embodiment of the invention, the mRNA molecule will include a series of repeating adenylate ribonucleotides at its 3'-end, the number of such being n, and formation of $(dN)_a$ involves addition of oligo dT, e.g., oligo $(dT)_{12\text{-}18}$, to the reaction mixture in sufficient quantity to permit formation of a $3'(dN)_a5'$ molecule having n repeating thymidylate deoxyribonucleotides at its 5'-end. In such molecules n is necessarily less than a.

Next, at least one ribonucleotide is added to the 3'-end of the $3'(dN)_a5'$ molecule to form a molecule having the formula $3'(rN)_b—(dN)_a5'$. Although ribonucleotide addition could be accomplished by conventional chemical methods, it is preferably effected by a ribosubstitution addition reaction utilizing a DNA polymerase, e.g., DNA polymerase I, and a mixture of the ribonucleotide triphosphates rATP, rGTP, rCTP, and rUTP under appropriate conditions to permit formation of the $3'(rN)_b—(dN)_a5'$ molecule. Once again, suitable conditions are known. They may include temperatures of about 25°–45° C., e.g., 37° C.; buffers having pH's of about 7.0, e.g., 7.4; molar excess of ribonucleotide triphosphates; and catalytic amounts of enzyme.

Finally, additional deoxyribonucleotides are added to the 3'-end of the $3'(rN)_b—(dN)_a5'$ molecule to form compound I. Here again, conventional chemical methods may be employed, or preferably enzymatic addition of the nucleotides may be utilized. In this regard, treatment of the $3'(rN)_b—(dN)_a5'$ molecule with a DNA polymerase, e.g., DNA polymerase I, and a mixture of dATP, dCTP, dGTP, and dTTP under suitable reaction conditions may be used. Suitable conditions may include temperatures from about 25°–45° C., e.g., 37° C.; buffers having pH's of about 7.0, e.g., 7.3; molar excess of nucleotide triphosphates; and catalytic amounts of polymerase.

The resulting compound I may then be recovered and purified by conventional techniques. Thereafter, it may be converted to a double-stranded molecule useful in bacterial cloning as described more fully hereinafter. Also, it may have uses in other areas, e.g., in pharmaceutical preparations or diagnostic tests.

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.