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

(51) **International Patent Classification[7]: C12Q 1/68,** C12P 19/34, C12N 9/00, G01N 33/00, C07H 21/02, 21/04

(21) **International Application Number:** PCT/US02/08850

(22) **International Filing Date:** 22 March 2002 (22.03.2002)

(25) **Filing Language:** English

(26) **Publication Language:** English

(30) **Priority Data:**
09/833,498   11 April 2001 (11.04.2001)   US

(71) **Applicant** *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) **Inventors; and**
(75) **Inventors/Applicants** *(for US only)*: **FU, Glenn** [US/US]; 4215 Roscommon Way, Dublin, CA 94568 (US). **STUVE, Laura L.** [US/US]; 14630 Stetson Road, Los Gatos, CA 95033 (US). **LEE, Walter H.** [US/US]; 3762 Delgado Court, Campbell, CA 95008 (US). **NI, Irene** [US/US]; 1062 Bodmin Avenue, San Leandro, CA 94579 (US).

(74) **Agent: FRANCIS, Carol, L.**; Bozicevic, Field & Francis LLP, 200 Middlefield Road, Suite 200, Menlo Park, CA 94025 (US).

(81) **Designated States** *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, 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, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, 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, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

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.

(54) **Title:** IMPROVED CDNA LIBRARIES AND METHODS FOR THEIR PRODUCTION

(57) **Abstract:** The present invention provides new methods of synthesizing cDNAs, methods of verifying full-length cDNAs, methods of producing cDNA libraries enriched for full-length inserts, and the like.

IMPROVED CDNA LIBRARIES AND METHODS FOR THEIR PRODUCTION

FIELD OF THE INVENTION

The present invention relates to the construction of cDNA libraries, and in particular

5   to methods of constructing full-length cDNA libraries.

BACKGROUND OF THE INVENTION

Methods for constructing complementary DNA (cDNA) libraries from mRNA are
well known in the art. In a typical procedure, poly(A)+ mRNAs are isolated from cells,

10   preferably a cell type in which the mRNA encoding the desired polypeptide is produced in
large quantities. The mRNAs are then converted into cDNA *in vitro* using the enzyme
reverse transcriptase to synthesize complementary cDNA strands from the mRNA template.
General protocols are, for example, described in Chapter 5 of Ausubel et al., *Current
Protocols in Molecular Biology*, Volume 1 (1991). Two commonly used methods of

15   producing cDNA from mRNA are described in Okayama and Berg, *Mol. Cell Biol.* 2:161-
170 (1982) and Gubler and Hoffman, *Gene* 25:263-269 (1983).

In the conventional process of converting mRNA into double stranded cDNA *in
vitro*, a first cDNA strand is synthesized by the reverse transcriptase and separated from the
mRNA by treatment with alkali or using a nuclease such as the enzyme RNase H. *E. coli*

20   DNA polymerase then uses the first cDNA strand as a template for the synthesis of the
second cDNA strand, thereby producing a population of double stranded cDNA molecules
from the original poly(A)+ mRNA. After converting the 5' and 3' ends into blunt ends, the
cDNA can be ligated to linkers/adaptors and subsequently ligated into suitable vectors and
transformed or packaged into a cell line to form the library. The library can then be screened

25   for cells transformed with nucleic acid encoding the desired polypeptide.

While the conventional methods have been used to successfully create cDNA
libraries, and to identify a large number of polypeptides, they do have certain disadvantages.
For example, an intrinsic problem in the construction of high quality full-length cDNA
libraries is that, under *in vitro* conditions, the reverse transcriptase very often does not

30   extend the first strand cDNA up to the 5' end of the mRNA, with the result that some mRNA
sequences (often longer sequences) are not represented in the library. This is thought to
occur in part due to misincorporation of an incorrect base by the reverse transcriptase, which
destabilizes the cDNA/mRNA duplex. Enzymes or proteins present in the cell that normally

repair nicks or correct mistakes during DNA synthesis are not present when the cDNA is synthesized *in vitro*.

In addition, hairpin formation in the mRNA can lead to early termination in the conversion to cDNA. This is especially a problem in the cloning of polypeptides having a

5   signal sequence located at the 5' end of the gene, as these libraries are often screened by detecting polypeptide exported from the transformed cells. Thus, these methods require full-length cDNA, including the signal sequence.

A number of methods have been developed to attempt to address these problems. For example, a different method of synthesizing cDNA *in vitro* selects full length poly(A)+

10   mRNA by treatment with bacterial alkaline phosphatase and tobacco acid pyrophosphatase, and subsequently ligating the 5' end of the mRNA to a chimeric DNA-RNA linker containing a restriction site. See, Kato et al., *Gene 25*:243-250 (1994). The poly(A) 3' end of the mRNA is then hybridized to an oligo d(T) sequence of and the oligo d(T) used to prime cDNA synthesis. This procedure is also limited, however, by the efficiency of the

15   phosphatases and the ligation procedure. Moreover, the ligation procedure can work with mRNA in which the 5' end has degraded, since the method does not distinguish between full-length and partial mRNA.

In standard methods currently used for the preparation of cDNA libraries, the mRNA in the cell is isolated by virtue of the presence of a polyadenylated tail present at its 3' end,

20   which binds to a resin specific for this structure (oligo dT-chromatography). The purified mRNA is then copied into cDNA using a reverse transcriptase, which starts at the 3' end of the mRNA and proceeds towards the 5' end. Second strand synthesis is then performed. Linkers are added to the ends of the double stranded cDNA to allow for its packaging into virus or cloning into plasmids. At this stage, the cDNA is in a form that can be propagated.

25   One disadvantage observed with current cDNA library synthesis protocols is that current methods tend to produce libraries having a significant proportion of incomplete cDNAs, which results from inefficiencies in the reverse transcriptase employed to generate the library. To compensate for the incomplete cDNA constituents of the library, investigators must perform many rounds of isolation (screenings) and construct a "full-

30   length" cDNA from the accumulated pieces. Such processes are resource intensive and do not ensure that each initial mRNA is represented in the cDNA library.

In addition, there is significant under-representation of sequences close to the 5' end of mRNAs in cDNA libraries produced by conventional methods. This under-representation

results from the fact that the reverse transcriptase will usually "fall off" before reaching these sequences. In many instances, the information located at the 5' end is of great interest.

Thus, there remains a need in the art for improved cDNA libraries, and in particular for cDNA libraries that are enriched for full-length cDNAs.

5

SUMMARY OF THE INVENTION

The present invention provides methods for identifying cDNAs comprising sequences corresponding to the 5' end of a transcript, methods of producing libraries comprising cDNAs, methods of verifying the presence of a 5'-end in a cDNA, and

10   producing libraries particularly rich in 5' ends and full length cDNAs (40-70%) relative to libraries produced from conventional technologies (10-30%).

In one embodiment, the present invention provides an improved method for producing full-length cDNAs comprising 1) relaxing the mRNA secondary structure, *e.g.*, by adding an agent such as dimethyl sulfoxide (DMSO) to the first-strand synthesis reaction

15   mixture and 2) utilizing a thermostable enzyme that exhibits 3' to 5' exonuclease activity for template driven enzymatic deoxynucleotide synthesis during first strand synthesis.

In one embodiment, the invention provides a method of isolating a full-length cDNA by: 1) contacting a ribonucleic acid molecule with a primer under conditions sufficient for template driven enzymatic deoxyribonucleic acid synthesis to occur to form a hybrid

20   RNA:DNA molecule; 2) contacting the hybrid molecule with a detectably labeled oligo-dV primer under conditions sufficient for template driven enzymatic deoxyribonucleic acid synthesis to occur to produce hybrid primer:RNA sequences, provided the deoxyribonucleic acid of the first hybrid molecule does not extend to the 5' end of the ribonucleic acid; 3) isolating a hybrid molecule that does not contain label; and 4) converting the unlabelled

25   hybrid molecule to a first double-stranded deoxyribonucleic acid molecule. In a particular embodiment, the single-stranded ribonucleic acid molecule is an mRNA. Following cDNA production, the double-stranded cDNA molecule can be introduced into a vector.

The primer used for first strand cDNA synthesis may be any primer that allows for directed synthesis of a deoxyribonucleic acid from a ribonucleic acid (including a gene

30   specific primer and/or a random primer), but is preferably an oligo-dT primer. The detectably labeled oligo-dV primer can be labeled with anything known in the art, including but not limited to, biotin, digoxygenin, radioactivity, and the like.

The present invention also features a method of identifying a full-length first strand cDNA including the steps of contacting an RNA-cDNA hybrid with a detectably labeled

- 3 -

primer under conditions sufficient for template driven enzymatic deoxyribonucleic acid synthesis to occur, where the primer is composed of a plurality of deoxyadenosines, deoxycytidines, and/or deoxyguanosines. The primer mixture used in the reaction may contain a plurality of the same primers, or a mixture of primers having varying sequences.

5   The primer will result in primer:RNA hybrid sequences if the cDNA in the first hybrid molecule does not extend to the 5' end of the RNA, and the primers can be detected as labeled sequences in the RNA:cDNA hybrid. Labeled sequences in a hybrid is indicative of the hybrid having a non-full-length cDNA. This can be performed as a step in cDNA synthesis, or to verify the efficacy of a particular method and/or reagent (*e.g.*, an enzyme).

10   Another embodiment provides a method for producing a 5' enriched cDNA library from a sample of mRNA molecules by: 1) contacting said mRNA molecules with a first primer under conditions sufficient for template driven enzymatic deoxyribonucleic acid synthesis to occur, creating a population of hybrid molecules of mRNA molecules hybridized cDNA molecules; 2) isolating the population of hybrid molecules; 3) contacting

15   the isolated hybrid molecules with a detectably oligo-dV labeled primer under conditions sufficient for template driven enzymatic deoxyribonucleic acid synthesis to occur, so that the reaction results in additional production of hybrid sequences provided the first hybrid molecule does not extend to the 5' end of the ribonucleic acid; 4) isolating hybrid molecules that do not contain labeled primer sequences; and 5) converting the unlabeled hybrid

20   molecules to a first double-stranded deoxyribonucleic acid molecule. These molecules can then be separated and introduced into vectors.

A feature of the present invention is a method for increasing the production of full-length cDNAs.

Another feature of the present invention is to provide methods for identifying cDNAs

25   that comprise the 5' end of an RNA, and in particular the 5' end of an mRNA.

An advantage of the present invention is that it provides quick and effective methods for validating the presence of a 5' end on a cDNA.

Another advantage of the present invention is that it provides for improved full-length cDNA libraries.

30   These and other objects, advantages, and features of the invention will become apparent to those persons skilled in the art upon reading the details of the invention as more fully described below.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Before describing the present methods, constructs and reagents, it is to be understood that this invention is not limited to the particular methods, constructs and reagents described, and as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting, since the scope of the present invention will be limited only by the appended claims.

Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limits of that range is also specifically disclosed. Each smaller range between any stated value and intervening value in a stated range and any other stated or intervening value in that stated range is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included or excluded in the range, and each range where either, neither or both limits are included in the smaller ranges is also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the invention.

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are now described. All publications mentioned herein are incorporated herein by reference to disclose and describe the methods and/or materials in connection with which the publications are cited.

It must be noted that as used herein and in the appended claims, the singular forms "a," "and," and "the" include plural references unless the context clearly dictates otherwise. Thus, for example, reference to "an enzyme" includes a plurality of such enzymes and reference to "the cDNA" includes reference to one or more cDNAs and equivalents thereof known to those skilled in the art, and so forth.

The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention. Further, the dates of publication provided might be different from the actual publication dates, which may need to be independently confirmed.

DEFINITIONS

The term "nucleic acid" as used herein refers to any polynucleotide, and is intended to encompass ribonucleic acids ("RNA"), including mRNA, and deoxyribonucleic acids ("DNA"), including genomic and cDNA. The term is also intended to encompass RNA or

5   DNA having analogs or substitutions to the structure of the nucleic acid, provided the analogs or substitutions does not impede the ability to isolate and/or characterize the sequence of the desired region of the nucleic acid.

The term "primer" as used herein refers to a polymer of nucleotides capable of acting as a point of initiation of DNA synthesis when annealed to a nucleic acid template under

10   conditions in which synthesis of a primer extension product is initiated, *i.e.*, in the presence of four different nucleotide triphosphates and a polymerase in an appropriate buffer ("buffer" includes pH, ionic strength, cofactors, etc.) and at a suitable temperature. Generally, a primer will be between 12 and 100 nucleotides, more preferably between 15 and 80 nucleotides, and even more preferably between 18 and 50 nucleotides. The primer may be

15   composed of naturally occurring and/or modified nucleotides, and the modified nucleotides may have a base substitution (*e.g.*, an analog with improved binding), a modified internucleoside linkage, or a substitution of the ribose group.

A primer that hybridizes to a sequence, refers to a primer that is complementary to a strand of the nucleic acid and/or a strand of the adaptor. A primer that hybridizes to the

20   coding region of a nucleic acid, or the corresponding strand of the adaptor, will have an "antisense" sequence, *i.e.*, the primer will form Watson-Crick base pairing with the coding region. A primer that hybridizes to a sequence complementary to a sequence will have a "sense" sequence, *i.e.*, it will have the same sequence as the coding region of the nucleic acid or the corresponding strand of the adaptor. For an amplification reaction, generally one

25   primer hybridizes to the sense strand and a second primer hybridizes to a sequence complementary to the sense strand.

The term "oligo-dV primer" as used herein refers to a primer composed of non-deoxythymidine nucleotides, i.e. deoxyadenosine, deoxycytidine, and/or deoxyguanosine. The oligo-dV primers of the invention can consist of only one nucleotide (*e.g.*, oligo-dC) or

30   may be any combination of dA, dC or dG. The oligo-dV primers of the invention are preferably between 6 and 15 nucleotides in length, and more preferably from between 8 and 12 nucleotides in length.

The term "hybridization" as used herein, refers to the formation of a duplex structure by two single stranded nucleic acids due to complementary base pairing. Hybridization can

occur between complementary nucleic acid strands or between nucleic acid strands that contain minor regions of mismatch. The melting temperature, or "Tm" measures stability of a nucleic acid duplex. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the base pairs have dissociated. Those skilled in the art of nucleic acid

5   technology can determine duplex stability empirically considering a number of variables including, for example, the length of the nucleic acids, base composition and sequence, ionic strength, and incidence of mismatched base pairs.

The term "stringent hybridization conditions" as used herein refers to conditions under which only fully complementary nucleic acid strands will hybridize. Stringent

10   hybridization conditions are well known in the art (see, e.g., Sambrook et al., 1985, Molecular Cloning - A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York). Generally, stringent conditions are selected to be about 5°C lower than the T_m for the specific sequence at a defined ionic strength and pH. Typically, stringent conditions will be those in which the salt concentration is at least about 0.2 M at pH 7 and

15   the temperature is at least about 60°C. Relaxing the stringency of the hybridizing conditions will allow sequence mismatches to be tolerated; the degree of mismatch tolerated can be controlled by suitable adjustment of the hybridization conditions.

The term "substantially complementary" as used herein refers to two single-stranded nucleic acids that are complementary except for minor regions of mismatch. Stable duplexes

20   of substantially complementary sequences can be achieved under less stringent hybridization conditions. Those skilled in the art of nucleic acid technology can determine duplex stability empirically considering a number of variables including, for example, the length and base pair concentration of the nucleic acids, ionic strength, and incidence of mismatched base pairs.

25   The term "full-length" mRNA as used herein refers to an mRNA that encodes the entire translation region of an mRNA, including promoter or enhancer regions 5' of the translation start site. The term cam also used to encompass transcripts that comprise at least the start methionine of the coding region of an mRNA, i.e., a transcript that is comprised of the entire coding region of an mRNA.

30

EXAMPLES

Production of full-length cDNA

The present invention provides an improved method for producing full-length cDNA comprising 1) relaxing the mRNA secondary structure, e.g., by adding an agent such as

dimethyl sulfoxide (DMSO) to the first-strand synthesis reaction mixture and 2) utilizing a thermostable enzyme that exhibits 3' to 5' exonuclease activity for template driven enzymatic deoxynucleotide synthesis.

5   The protocol for producing a cDNA can use any variation known in the art, provided it is not contraindicative with the relaxation techniques used and/or the thermostable enzyme. In general, the first step in preparing a cDNA library is to purify the mRNA, which usually represents only about 1-3% of the total RNA of the cell, the remainder being ribosomal RNA, transfer RNA, and several other RNA species. Many mRNAs from eukaryotic organisms have a poly(A) "tail," a tract of 50-150 adenosine residues at their 3'

10   ends. A general practice for purifying mRNA from total cellular RNA involves specifically annealing, or binding, the poly(A) tail to oligo(dT), a single stranded DNA molecule of between about 12 and 30 consecutive dT residues (Jacobson, A. (1987) *Meth. Enzymol. 152*:254). Total cellular RNA can be incubated with a matrix to which oligo(dT) has been immobilized. Only RNA molecules containing poly(A) tails selectively anneal to the matrix.

15   Upon purification of poly(A)+ RNA, a double-stranded complementary DNA (cDNA) copy of this active RNA can be synthesized *in vitro* by two sequential enzymatic steps. An RNA-dependent DNA polymerase, known as a reverse transcriptase, is used to synthesize the first strand cDNA (complementary DNA), using the RNA as a template. Then, a DNA-dependent DNA polymerase copies the newly synthesized first cDNA strand

20   to form a complementary second cDNA strand.

Relaxation of the secondary structure of the mRNA can be achieved using any physical or chemical means, as will be apparent to one skilled in the art upon reading the present disclosure. In a particular embodiment, the relaxation of the mRNA secondary structure is achieved via chemical means, *e.g.*, the addition of a chemical such as DMSO or

25   dimethyl-formamide at a concentration of about 5-20%, glycerol at a concentration of about 5-40%, formamide in a concentration about 5-10%, and betaine in a concentration about 0.5-2.2 M.

These chemicals are added in sufficient quantities to prevent occurrence of mRNA self-annealing, up to and including 10% of the final reverse transcription reaction volume.

30   In one embodiment, the invention provides the use of a polymerase having proofreading ability to increase the efficacy of the polymerase reaction in first strand synthesis and thus the production of full-length cDNAs. Examples of such enzymes include, but are not limited to, Pfu DNA polymerase, Tna DNA polymerase, Tma DNA polymerase or Tli DNA polymerase. Examples of such enzymes are those isolated from the organisms

Pyrococcus furiosus (Pfu) (U.S. Pat. Nos. 5,948,663 and 5,866,395), *Thermotoga*

neapolitana (Tne); *Thermotoga maritima* (Tma) (U.S. Pat. Nos. 6,077,664; 6001645;

5,948,614; and 5,939,301); and *Pyrodictium* species (U.S. Pat. No. 5,491,086).

      These and other enzymes that have 3'-5' exonuclease (proofreading) activity are

5   useful to practice the invention. However, Pfu is preferred because it does not have strand

displacement activity.

      Once synthesized, double-stranded cDNA can be inserted into a prepared cloning

vector. To efficiently insert the cDNA into a cloning vector, the ends of the insert cDNA

and the vector DNA molecules must be prepared such that they are compatible. For

10   example, specialized linkers can be added to the cDNA ends, followed by digestion with the

relevant enzyme to create single stranded protrusions that will anneal to corresponding ends

in the vector. The insert and vector molecules are ligated together with T4 DNA ligase. The

ligated vectors carrying their cDNA molecule inserts are then introduced into *E. coli* and

screened.

15   Various approaches have been used to prepare the cDNA ends for vector insertion

(Kimmel, A. R. and Berger, S. L. (1987) *Meth. Enzymol. 152*:307). Most have used the

"linker" or "adapter" method described above. All methods using linkers require an

additional step to protect the cDNA from being cleaved at adventitious restriction sites

during digestion to create the cohesive ends (Wu, R., Wu, T. and Ray, A. (1987) *Meth.*

20   *Enzymol. 152*:343). The protection is accomplished either by treating the cDNA with on

site-specific methylases or by substituting a methylated dCTP analog for unmodified dCTP

in the synthesis reactions.

      The degradation of an oligonucleotide primer by a 3' exonuclease can be prevented

by the use of modified nucleotides at the 3' terminus. For example, the use of

25   dideoxynucleotides or deoxynucleotides having a phosphorothioate linkage at the 3' terminus

of an oligonucleotide would prevent degradation by 3' exonucleases.

Identification and/or validation of full-length cDNAs

      The methods of the present invention can also be used to validate and/or identify the

30   presence of a full-length first strand cDNA. Such a method can be used in any conventional

means for producing cDNA to determine whether or not the cDNA produced encompasses

the 5' end of the transcript.

      Following first strand synthesis of cDNA from the mRNA, the DNA:RNA hybrid

molecule can be used as a template for a further reverse transcription reaction, using a

mixture of detectably labeled oligo-dV primers to initiate the template driven enzymatic deoxyribonucleic acid synthesis. This mixture may contain a plurality of the same oligo-dV primer, or may contain oligo-dV primers of varying sequences.

5   The RNA:DNA hybrid is hybridized to a detectably labeled oligo-dV primer mixture under conditions sufficient for template driven enzymatic deoxyribonucleic acid synthesis to occur, wherein the primer results in additional production of hybrid sequences provided the deoxyribonucleic acid does not extend to the 5' end of the ribonucleic acid. The presence of labeled product following this second reverse transcription reaction is indicative of incomplete reverse transcription in the first reaction, i.e., a non-full-length cDNA in the
10   RNA:DNA hybrid.

Synthetic oligonucleotides may be prepared by the method of Efimov, V. A. et al. (*Nucl. Acids Res.* (1982) 6875-6894), and can be prepared using commercially available automated oligonucleotide synthesizers. Kinasing of single strands prior to annealing or for labeling is achieved using an excess, e.g., approximately 10 units of polynucleotide kinase to
15   1 nmole substrate in the presence of 50 mM Tris, pH 7.6, 10 mM $MgCl_2$, 5 mM dithiothreitol, 1-2 mM ATP, 1.7 pmoles γ32P-ATP (2.9 mCi/mmole), 0.1 mM spermidine, 0.1 mM EDTA.

Production of cDNA libraries

20   The methods of the present invention can also be used to produce cDNA libraries having an increased percentage of inserts comprising the 5' end of a transcript, (e.g., has an increased percentage of inserts encoding at least the start methionine of an mRNA.) each library comprising size fractionated cDNAs preferably enriched for full-length mRNA sequences. The methods used to produce the cDNAs can comprise 1) comprising relaxing
25   the mRNA secondary structure, e.g., by adding an agent such as dimethyl sulfoxide (DMSO) to the first-strand synthesis reaction mixture and utilizing a thermostable enzyme that exhibits 3' to 5' exonuclease activity for template driven enzymatic deoxynucleotide synthesis; 2) validating the presence of the 5' end of the first strand cDNA using the oligo-dV driven deoxynucleotide synthesis, as described above; or 3) a combination of the two
30   techniques. Where validation techniques are used, the full-length cDNAs can be identified by the absence of fluorescence in the product, and products lacking such fluorescence can be cloned into a vector, thus enriching the population of cDNAs comprising a 5'-end in the library. The incomplete (non-full length) cDNA is labeled with biotin and can be selectively removed using immobilized streptavidin.

Once the cDNA is produced, any number of methods can be used to produce a cDNA library, as will be apparent to one skilled in the art upon reading the present disclosure. The following is a general example of techniques that may be used, and as such is not meant to be limiting.

5   The cDNAs may be size selected prior to introduction into a vector, or may be directly introduced to a vector following production of the cDNAs. For example, the blunt-ended, double-stranded cDNA can be fractionated by size using gel filtration through a column (0.3 X 14 cm) Sepharose 4B (Pharmacia Fine Chemicals, Piscataway, NJ) or by ultracentrifugation in 5-20% glycerol gradient followed by fractionation of the gradient. The
10   cDNA is retained and recovered by precipitation with 70% ethanol. Short (10-30 nucleotide) polymeric tails of deoxycytosine are added to the 3' termini of the cDNA using a reaction containing 0.2M potassium cacodylate, 25 mM Tris, pH 6.9, 2 mM dithiothreitol, 0.5 mM $CoCl_2$, 200 mM cDTP, 400 µg/ml BSA, and 40 units calf thymus terminal deoxynucleotide transferase for 5 min at 22°C. The reaction is extracted with
15   phenol:chloroform, and small oligonucleotides are removed with three ethanol precipitations in the presence of ammonium acetate.

Following isolation, the cDNAs can be ligated to adaptor molecules that allow cloning, directional or not, into the vector of choice. In one particular embodiment, the restriction sites used in the adaptors, and thus in the introduction of the cDNA insert into a
20   vector, have relatively long recognition sites (*i.e.*, recognition sites of 8 or more residues). The use of adaptors and multiple cloning sites having longer recognition sites decreases the chance of the site occurring in a cDNA, and thus increases the chance that the cDNA will be inserted into the vector in its full-length form. Construction of suitable vectors containing the desired coding and control sequences employs standard ligation and restriction
25   techniques that are well understood in the art.

Site-specific DNA cleavage is performed by treating with the suitable restriction enzyme (or enzymes) under conditions that are generally understood in the art, the particulars of which are specified by the manufacturer of these commercially available restriction enzymes. See, *e.g.*, New England Biolabs, Product Catalog. In general, about 1
30   µg of plasmid or DNA sequence is cleaved by one unit of enzyme in about 20 µl of buffer solution. In the examples herein, typically, an excess of restriction enzyme is used to insure complete digestion of the DNA substrate. Incubation times of about one hour to two hours at about 37°C. are workable, although variations can be tolerated. After each incubation, protein is removed by extraction with phenol/chloroform, and may be followed by ether

extraction, and the nucleic acid recovered from aqueous fractions by precipitation with ethanol. If desired, size separation of the cleaved fragments may be performed by polyacrylamide gel or agarose gel electrophoresis using standard techniques. A general description of size separations is found in *Methods in Enzymology* (1980) 65:499-560.

5   Restriction cleaved fragments may be blunt ended by treating with the large fragment of *E. coli* DNA polymerase I (Klenow) in the presence of the four deoxynucleotide triphosphates (dNTPs) using incubation times of about 15 to 25 min at 20°C to 25°C in 50 mM Tris pH 7.6, 50 mM NaCl, 6 mM $MgCl_2$, 6 mM DTT and 5-10 µM dNTPs. The Klenow fragment fills in at 5' sticky ends but eliminates protruding 3' single strands, even

10   though the four dNTPs are present. If desired, selective repair can be performed by supplying only one of the, or selected, dNTPs within the limitations dictated by the nature of the sticky ends. After treatment with Klenow, the mixture is extracted with phenol/chloroform and ethanol precipitated. Treatment under appropriate conditions with S1 nuclease or Bal-31 results in hydrolysis of any single-stranded portion.

15   Ligations are performed in 15-50 µl volumes under the following standard conditions and temperatures: 20 mM Tris-Cl pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 33 µg/ml BSA, 10 mM-50 mM NaCl, and either 40 µM ATP, 0.01-0.02 (Weiss) units T4 DNA ligase at 0°C. (for "sticky end" ligation) or 1 mM ATP, 0.3-0.6 (Weiss) units T4 DNA ligase at 14°C. (for "blunt end" ligation). Intermolecular "sticky end" ligations are usually performed at 33-100

20   µg/ml total DNA concentrations (5-100 nM total end concentration). Intermolecular blunt end ligations (usually employing a 10-30 fold molar excess of linkers) are performed at 1 µM total ends concentration.

        In vector construction employing "vector fragments," the vector fragment is commonly treated with bacterial alkaline phosphatase (BAP) or calf intestinal alkaline

25   phosphatase (CIP) in order to remove the 5' phosphate and prevent religation of the vector. Digestions are conducted at pH 8 in approximately 150 mM Tris, in the presence of Na^+ and Mg^{-2} using about 1 unit of BAP or CIP per µg of vector at 60°C for about one hour. In order to recover the nucleic acid fragments, the preparation is extracted with phenol/chloroform and ethanol precipitated. Alternatively, religation can be prevented in vectors that have been

30   double digested by additional restriction enzyme digestion of the unwanted fragments.

        Once the cDNA library has been constructed as described above it can be screened by conventional methods designed to quickly determine whether a particular clone contains the desired nucleic acid sequence. In general, libraries are usually screened by hybridization to nick-translated DNA and synthetic oligonucleotides, or using immunoreactivity-based

assays, or by hybrid selection of mRNA and translation. Libraries are usually spread out on agarose plates, transferred to nitrocellulose filters and hybridized to labeled oligonucleotide probes complementary to a target sequence in the cDNA, or the target polypeptides are expressed and bound to labeled antibodies which recognize an epitope on a target

5   polypeptide. In both instances, the label can be radioactive (e.g., ^{32}P), fluorescent, biotinylated, and the like. When clones containing cDNA encoding the target polypeptide have been found, the cDNA can be readily isolated for larger scale expression, *e.g.*, by cutting the cDNA from the vector and amplifying it using PCR. This step can be facilitated where restriction sites are available in the cDNA sequence or engineered in the flanking

10   sequence of the vector.

EXAMPLES

       The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the present invention,

15   and are not intended to limit the scope of what the inventors regard as their invention nor are they intended to represent that the experiments below are all or the only experiments performed. Efforts have been made to ensure accuracy with respect to numbers used (*e.g.*, amounts, temperature, etc.) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is weight

20   average molecular weight, temperature is in degrees Centigrade, and pressure is at or near atmospheric.

EXAMPLE 1: Full-length cDNA Synthesis, Selection and Cloning
First Strand Synthesis

25   2µl DMSO was added to 1.5 µg mRNA and the volume to 11 µl with water. The mixture was heated to 70°C for 5 minutes, and then quick-chilled on ice. Contents were collected at the bottom of the tube by brief centrifugation. Primer mix was then added 0.75 µg in 1 µl water. Each primer mix was a combination of an antisense strand, 5'pCTCCAGTCGCGATCTAGAACTAGTC (SEQ ID NO:1) annealed equimolar to one of

30   five sense strands:
       5'pGACTAGTTCTAGATCGCGACTGGAGTTTTTTTTTTTTTTTTTTVV (SEQ ID NO:2);
       5'pGACTAGTTCTAGATCGCGACTGGAGTTTTTTTTTTTTTTTTTTTTVTV (SEQ ID NO:3);

5'pGACTAGTTCTAGATCGCGACTGGAGTTTTTTTTTTTTTTTTTTTTVTTV (SEQ ID NO:4);

5'pGACTAGTTCTAGATCGCGACTGGAGTTTTTTTTTTTTTTTTTTTTTVTTV (SEQ ID NO:5); and

5   5'pGACTAGTTCTAGATCGCGACTGGAGTTTTTTTTTTTTTTTTTTTTVVVVV (SEQ ID NO:6).

Each primer was gel-purified prior to dilution.

The following components were then added to the mRNA-primer mix: 4 μl 5X first strand buffer; 1 μl 0.1 M DTT;1 μl 10 mM methyl dNTP mix (10mM each of dG, dA, dT

10 and Methyl-dCTP); and 1 μl α-[^{32}P]dCTP (10 μCi/μl; Amersham). Contents were gently vortexed and collected by brief centrifugation. One μl of mut-MMLV-RT (200u/μl) was added gently and incubated at 42˚C for 1 hour. Following incubation, 0.5 μl pfu (Stratagene, 2.5u/μl) were added, and the mixture incubated at 55˚C for 10 minutes. 0.5 μl mut-MMLV-RT (RnaseH minus MMLV (Moloney murine leukemia virus) reverse transcriptase from

15 Stratascript by Stratagene) were added, and incubated at 42˚C for an additional 20 minutes. The temperature is then slowly brought to a final temperature of 55˚C over 20 minutes.

To assess the quantity, quality and average length of the synthesized first-strand cDNA, 0.5 μl sample of the reaction was added to 10.75 μl 20mM EDTA and 1.25 μl yeast tRNA (1 mg/ml) in a 1.5 ml microcentrifuge tube and mixed well.

20 For each sample a GF/C filter was labeled with an EtOH proof (VWR) marker, and 10μl of the sample spotted onto each of the filters and let dry. The CPM for each sample on the dried filters was calculated to determine the overall count. The filters were washed for five minutes in 300 ml of cold 10% TCA and 1% sodium pyrophosphate and 1X in 75 ml of cold 100% EtOH while shaking. The filters were dried again to determine the TCA

25 precipitable counts incorporated into the newly synthesized cDNA strand.

Non-full-length sequences were removed from the pools using primers complementary to the 3' end of the sense strand primers. Added to the mixture were:

4 μl 5X first strand buffer;

1 μl N1V5 (NVVVVV) 100 ng/μl;

30 3.5 μl Biotin-dATP 0.4mM;

3.5 μl Biotin dCTP 0.4mM;

3.5 μl dTTP 0.4mM;

3.5 μl dGTP 0.4mM; and

1 μl mut-MMLV RT.

The mixture was incubated at 37°C for 30 minutes, then 1 μl 0.5M EDTA was added. The reaction contents were spun through a chromospin 1000 column to remove unincorporated nucleotides and excess primer. Following spinning, an equal amount of 2X M280 streptavidin binding buffer (Dynal) was added. The sample was incubated with 200

5   μl of prewashed M280 beads with agitation at 37°C for 10 minutes and the supernatant removed.

To precipitate the first strand cDNA, one-half volume of 7.5 M NH4Oac and 1 μl of glycogen (20 μg) was added followed by 2.5 volumes of 100% ethanol. The mixture was vortexed thoroughly and centrifuged at 14,000 x g at room temperature for 15 minutes. The

10   supernatant was carefully removed and the pellet gently rinsed with 200 μl of cold 70% ethanol, followed by another centrifugation for 2 min. at 14,000 x g. The wash was repeated, and all of the ethanol removed. The pellet was allowed to dry for 5-10 minutes to evaporate residual ethanol.

15   **Second Strand Synthesis**

For second strand synthesis, the following were added to the first strand cDNA pellet:

4 μl 5X first strand buffer;

2 μl 0.1M DTT;

20   104 μl DEPC-treated water;

30 μl 5X second strand buffer;

4 μl 10 mM dNTP mix (unmethylated);

1 μl *E. coli* ligase (10U/μl);

4 μl *E. coli* DNA polymerase I (10U/μl); and

25   1 μl *E. coli* RNaseH (2U/μl).

This was gently mixed, and the reaction incubated at 16°C for 2 hr. 2 μl (5U/μl) of T4 DNA polymerase were added, and incubating continued at 16°C for 5 minutes. The reaction mixture was then placed on ice and 10 μl of 0.5 M EDTA was added.

An equal amount of 2X dynal M280 streptavidin binding buffer was also added, and

30   the reaction mixture was incubated with 100 μl of prewashed M280 beads with agitation at 37°C for 10 min, and the solution extracted to a new tube. An equal volume of phenol:chloroform:isoamyl alcohol (25:24:1; Amersham) was added, the mixture vortexed thoroughly, and centrifuged at room temperature for 5 minutes at 14,000 x g to separate the

two phases. The upper, aqueous layer was carefully removed and transferred to a new 1.5 ml tube.

To precipitate the cDNA, one-half volume of 7.5 M NH₄Oac and 1 µl of glycogen (20 µg) was added followed by 2.5 volumes of 100% ethanol. The mixture was vortexed

5   thoroughly and centrifuged at 14,000 x g at room temperature for 15 minutes. The supernatant was carefully removed and the pellet gently rinsed with 200 µl of cold 70% ethanol, followed by another centrifugation for 2 min. at 14,000 x g. The wash was repeated, and all of the ethanol removed. The pellet was allowed to dry for 5-10 minutes to evaporate residual ethanol. The cDNA pellet was resuspended in 70 µl of 1x TE buffer.

10

Iceu 1 Adapter Addition

An adaptor was added to the synthesized cDNAs to allow cloning of the cDNAs into an appropriate vector. The following reagents were added to the to 34 µl of cDNA:

10 µl 5X T4 ligase buffer;

15   1 µl of Iceu adapters (2 µg/µl) ("p"ggtagcga)(ptcgctaccttag);

5 µl of T4 DNA ligase; and

50 µl final volume.

The reaction mixture was mixed gently and incubated overnight at 16°C. The volume was then brought to 200 µl by adding 150 µl of DEPC water, and an equal volume of

20   phenol:chloroform: isoamyl alcohol (25:24:1) was added and the mixture vortexed thoroughly and centrifuged at room temperature for 5 minutes at 14,000 x g to separate the two phases. The upper, aqueous layer was carefully removed to a new 1.5 ml tube.

The cDNA with adaptor was precipitated by adding 1/2 volume 7.5 M NH₄OAc, 1 µl of glycogen (at 20µg/µl) followed by 2.5 volumes of 100 % EtOH. The precipitation mixture

25   was vortexed thoroughly and centrifuge at 14,000 x g at room temperature for 15 minutes. The supernatant was carefully removed and the pellet rinsed gently 2x with 200 µl of cold 70% EtOH and centrifuged for 2 minutes to remove the EtOH. The pellet was air-dried for 5-10 minutes to evaporate residual EtOH, and the pellet resuspended in 40.5 µl DEPC water.

30   **BpMI Digestion**

The 40.5 µl cDNA/adaptor mixture

5µl 10X buffer 3

0.5µl 100X BSA

4µl BpMI (New England Biolabs)

- 16 -

50 μl total volume

This was mixed gently and incubated at least 4-5 hrs at 37°C. The volume was then brought to 200 μl by adding 150 μl of DEPC water. An equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added, the mixture vortexed thoroughly, and centrifuged at room temperature for 5 minutes at 14,000 x g to separate the two phases. The upper, aqueous layer was carefully removed and transferred to a new 1.5 ml tube. One-half volume of 7.5 M NH₄Oac and 3.5 μl of yeast tRNA (1μg/μl) were added, followed by 2.5 volumes of 100% ethanol. The mixture was vortexed thoroughly and centrifuged at 14,000 x g at room temperature for 15 minutes. The supernatant was carefully removed and the pellet gently rinsed twice with 200 μl of 70% EtOH. Following rinsing, the pellet was centrifuged for 2 minutes at 14,000 x g, and the supernatant removed. The pellet was air dried for 5-10 minutes to evaporate residual ethanol and resuspended in 70 μl of STE.

The cDNA was then sized using a Bio-Rad Poly-prep Chromatography Column was filled with 2 ml STE-equilibrated Sepharose. The sample was added to the column and allowed to drip through. STE was added to elute the cDNA, and the run-off collected in fractions. The amount of cDNA in each fraction was determined, and fractions having a significant concentration of cDNA precipitated.

Add 20 μg glycogen, 1/2 volume 7.5 M NH₄OAc and 2.5 volumes 100% EtOH. Precipitate for greater than three hours to overnight. Vortex the mixture thoroughly and centrifuge at 14,000 x g at room temperature for 30 min. Wash with 200 μl cold 70% EtOH and air dry. Count the dry pellets.

EXAMPLE 2: Creating a cDNA library

The cDNA was then cloned into the vector pCMV_ICIS to create cDNA library using the methods of the invention.

The following components were added to a 1.5 ml micro-centrifuge tube:

2 μl 5X DNA ligase buffer;

1 μl (25 ng/μl) of pCMV_ICIS (IceuI/BsgI cut);

10 ng cDNA;

water to bring the volume up to 9 μl; and

1 μl of T4 DNA ligase (New England Biolabs, Beverly, MA).

Vortex gently to mix and quick-spin contents to bottom of the tube. The reaction was then incubated at room temperature for 2 hr.

While the present invention has been described with reference to the specific embodiments thereof, it should be understood by those skilled in the art that various changes may be made and equivalents may be substituted without departing from the true spirit and scope of the invention. In addition, many modifications may be made to adapt a particular

5   situation, material, composition of matter, process, process step or steps, to the objective, spirit and scope of the present invention. All such modifications are intended to be within the scope of the claims appended hereto.

CLAIMS

What is claimed is:

1. A method for producing a full-length cDNA, comprising the steps of:

contacting a ribonucleic acid to an agent that inhibits formation of secondary structure;

contacting the ribonucleic acid molecule with a primer and a polymerase to form a first-strand synthesis mixture, wherein said polymerase comprises 3'-5' exonuclease activity;

exposing the first-strand synthesis mixture to conditions sufficient for template driven enzymatic deoxyribonucleic acid synthesis to occur, whereby a hybrid molecule comprising said ribonucleic acid molecule hybridized to a first deoxyribonucleic acid molecule comprising the first primer is produced; and

converting said hybrid molecule to a double-stranded deoxyribonucleic acid molecule.

2. The method of claim 1, wherein the agent is present in the transcription mixture and is selected from the group consisting of dimethyl sulfoxide (DMSO) at a concentration of about 5-20%, glycerol at a concentration of about 5-40%, formamide at a concentration of about 5-10%, and betaine at a concentration of about 0.5-2.2 M.

3. The method of claim 1, wherein the reverse transcriptase is selected from the group consisting of: Pfu DNA polymerase, Tma DNA polymerase, Tne polymerase, and Tli DNA polymerase.

4. A method of isolating a full-length cDNA, said method comprising:

contacting a ribonucleic acid molecule with a first primer under conditions sufficient for template driven enzymatic deoxyribonucleic acid synthesis to occur, whereby a hybrid molecule comprising said ribonucleic acid molecule hybridized to a first deoxyribonucleic acid molecule comprising the first primer is produced;

contacting the hybrid molecule with a detectably labeled oligo-dV primer under conditions sufficient for template driven enzymatic deoxyribonucleic acid synthesis to occur, wherein said primer results in additional production of hybrid sequences provided first hybrid molecule does not extend to the 5' end of the ribonucleic acid;

isolating a hybrid molecule that does not contain the detectable label; and

converting said unlabelled hybrid molecule to a double-stranded deoxyribonucleic

acid molecule.

5   5. The method of claim 4, wherein said single-stranded ribonucleic acid molecule is an

mRNA.

6. The method of claim 4, wherein the first primer is an oligo-dT primer.

10   7. The method of claim 4, wherein the first primer is a random primer.

8. The method of claim 4, wherein the oligo-dV primer is labeled with biotin.

9. The method of claim 4, wherein the oligo-dV primer is radioactively labeled.

15

10. The method of claim 4, wherein the oligo-dV primer is comprised of 4-12 nucleotides.

11. The method of claim 4, wherein said method further comprises introducing said double-

stranded cDNA molecule into a vector.

20

12. A method of identifying a full-length first strand cDNA, the method comprising the

steps of:

contacting a ribonucleic acid molecule hybridized to a deoxyribonucleic acid

molecule with a detectably labeled oligo-dV primer under conditions sufficient for template

25   driven enzymatic deoxyribonucleic acid synthesis to occur, wherein said primer results in

additional production of hybrid sequences provided the deoxyribonucleic acid does not

extend to the 5' end of the ribonucleic acid; and

detecting label in the cDNA;

wherein label in the cDNA is indicative of a non-full-length cDNA.

30

13. The method of claim 12, further comprising:

contacting a ribonucleic acid molecule with a first primer under conditions sufficient

for template driven enzymatic deoxyribonucleic acid synthesis to occur to produce a hybrid

molecule comprising said ribonucleic acid molecule hybridized to a deoxyribonucleic acid molecule.

14. A method for producing a cDNA library from a sample of mRNA molecules, said
5   method comprising:

contacting said mRNA molecules with a first primer under conditions sufficient for template driven enzymatic deoxyribonucleic acid synthesis to occur, whereby a population of hybrid molecules comprising said mRNA molecules hybridized to a population of cDNA molecules comprising said first primer is produced;

10   isolating the population of hybrid molecules;

contacting said isolated hybrid molecules with a detectably labeled oligo-dV primer under conditions sufficient for template driven enzymatic deoxyribonucleic acid synthesis to occur, wherein said primer results in additional production of hybrid sequences provided the first hybrid molecule does not extend to the 5' end of the ribonucleic acid;

15   converting said population of hybrid molecules to a double-stranded deoxyribonucleic acid molecule;

separating said population of double-stranded cDNA molecules into labeled double-stranded cDNA molecules and unlabelled double-stranded cDNA molecules; and

introducing each unlabelled double-stranded cDNA molecules into a vector to
20   produce a library enriched in full-length cDNAs.

15. The method of claim 14, wherein the first primer is oligo-dT primer.

16. The method of claim 14, wherein the first primer is a random primer.
25

17. The method of claim 14, wherein said method further comprises introducing the population of double-stranded cDNA molecules into vectors.

18. The method of claim 14, wherein said method further comprises selectively cloning only
30   the 5' end of a full-length cDNA to produce a cDNA library comprising the 5' ends of expressed genes.

19. A method for producing a cDNA library from a sample of mRNA molecules, said method comprising:

contacting said mRNA molecules with an agent that inhibits formation of secondary structure;

contacting said mRNA molecules with a primer and a polymerase to form a first-strand synthesis mixture, wherein said polymerase comprises 3'-5' exonuclease activity;

5   exposing the first-strand synthesis mixture to conditions sufficient for template driven enzymatic deoxyribonucleic acid synthesis to occur whereby a population of hybrid molecules comprising said mRNA molecules hybridized to a population of cDNA molecules comprising said first primer is produced;

isolating the population of hybrid molecules;

10   contacting said isolated hybrid molecules with a detectably labeled oligo-dV primer under conditions sufficient for template driven enzymatic deoxyribonucleic acid synthesis to occur, wherein said primer results in additional production of hybrid sequences provided the first hybrid molecule does not extend to the 5' end of the ribonucleic acid;

converting said population of hybrid molecules to a double-stranded

15   deoxyribonucleic acid molecule;

separating said population of double-stranded cDNA molecules into labeled double-stranded cDNA molecules and unlabelled double-stranded cDNA molecules; and

introducing each unlabelled double-stranded cDNA molecules into a vector to produce a library enriched in full-length cDNAs.

20

SEQUENCE LISTING

<110> Fu, Glenn K.
      Stuve, Laura L.
      Lee, Walter H.
      Ni, Irene

<120> Improved cDNA Libraries and Methods for
  their Production

<130> 06514/091WO1

<150> 09/833,498
<151> 2001-04-11

<160> 6

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> antisense strand of primer

<400> 1
ctccagtcgc gatctagaac tagtc   25

<210> 2
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> sense strand of primer

<400> 2
gactagttct agatcgcgac tggagttttt tttttttttt ttttvv   46

<210> 3
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> sense strand of primer

<400> 3
gactagttct agatcgcgac tggagttttt tttttttttt ttttvtv   47

<210> 4
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<223> sense strand of primer

<400> 4
gactagttct agatcgcgac tggagttttt tttttttttt ttttvttv   48

```
<210> 5
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> sense strand of primer

<400> 5
gactagttct agatcgcgac tggagttttt tttttttttt ttttvtttv                    49

<210> 6
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> sense strand of primer

<400> 6
gactagttct agatcgcgac tggagttttt tttttttttt ttttvvvvv                    49
```

INTERNATIONAL SEARCH REPORT	International application No.
	PCT/US02/08850

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7)　:C12Q 1/68; C12P 19/34; C12N 9/00; G01N 33/00; C07H 21/02, 21/04

US CL　:435/6, 91.1, 91.2, 91.51, 183; 436/94; 536/23.1, 24.3, 24.33

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. :　435/6, 91.1, 91.2, 91.51, 183; 436/94; 536/23.1, 24.3, 24.33

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CARNINCI et al. Thermostabilization and thermoactivation of the thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. USA, January 1998, Vol. 95, Pages 520-524, especially page 521.	1-3
Y	US 5,374,553 A (GELFAND ET AL) 20 December 1994 (20/12/94), see entire document, especially column 28.	1-3

☐ Further documents are listed in the continuation of Box C.　☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance		
"E"	earlier document published on or after the international filing date	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed	"&"	document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
07 JUNE 2002	**20 AUG 2002**
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	Authorized officer FRANK LU
Facsimile No.　(703) 305-3230	Telephone No.　(703) 308-1235

Form PCT/ISA/210 (second sheet) (July 1998)★

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/08850

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest   ☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet(1)) (July 1998)★

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

STN and WEST
Search terms: cDNA synthesis, secondary structure, inhibitor, dimethyl sulfoxide,glycerol, formamide, betaine, primer without deoxythymidine, primer with no deoxythymidine, and nondeoxymidine nucleotide

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

Group I, claims 1-3, drawn to a method for producing a full length cDNA.
Group II , claims 4-11, drawn to a method of isolating a full-length cDNA.
Group III, claims 12 and 13, drawn to a method of identifying a full-length first strand cDNA.
Group IV, claims 14-18, drawn to a method for producing a cDNA library from a sample ofmRNA molecules.
Group V, claim 19, drawn to a method for producing a cDNA library from a sample ofmRNA molecules

The inventions listed as Groups I to V do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Groups I and II do not relate to a single inventive concept under PCT Rule 13.1 because they lack the same or corresponding special technical features. For example, the search required for Group I such as an agent that inhibits formation of secondary structure is not required for Group II while the search required for Group II such as isolating a hybrid molecule (mRNA-first stranded cDNA) that does not contain the detectable label is not required for Group I.

Groups I and III do not relate to a single inventive concept under PCT Rule 13.1 because they lack the same or corresponding special technical features. For example, the search required for Group I such as an agent that inhibits formation of secondary structure is not required for Group III while the search required for Group III such as detecting label in the cDNA is not required for Group I.

Groups I and IV do not relate to a single inventive concept under PCT Rule 13.1 because they lack the same or corresponding special technical features. For example, the search required for Group I such as polymerase with 3'-5'exonuclease activity is not required for Group IV while the search required for Group IV such as separating labeled double-stranded cDNA and unlabeled double-stranded cDNA molecules is not required for Group I.

Groups I and V do not relate to a single inventive concept under PCT Rule 13.1 because they lack the same or corresponding special technical features. For example,the search required for Group V such as separating labeled double-stranded cDNA and unlabeled double-stranded cDNA molecules is not required for Group I.

Groups II and III do not relate to a single inventive concept under PCT Rule 13.1 because they lack the same or corresponding special technical features. For example,the search required for Group II such as isolating a hybrid molecule (mRNA-first stranded cDNA) that does not contain the detectable label is not required for Group III while the search required for Group III such as non-full-length cDNA is not required for Group II.

Groups II and IV do not relate to a single inventive concept under PCT Rule 13.1 because they lack the same or corresponding special technical features. For example,the search required for Group II such as isolating a hybrid molecule (mRNA-first stranded cDNA) that does not contain the detectable label is not required for Group IV while the search required for Group IV such as separating labeled double-stranded cDNA and unlabeled double-stranded cDNA molecules is not required for Group II.

Groups II and V do not relate to a single inventive concept under PCT Rule 13.1 because they lack the same or corresponding special technical features. For example, the search required for Group V such as an agent that inhibits formation of secondary structure is not required for Group II.

Groups III and IV do not relate to a single inventive concept under PCT Rule 13.1 because they lack the same or corresponding special technical features. For example,the search required for Group III such as non-full-length cDNA is not required for Group IV while the search required for Group IV such as separating labeled double-stranded cDNA and unlabeled double-stranded cDNA molecules is not required for Group III.

Groups III and V do not relate to a single inventive concept under PCT Rule 13.1 because they

lack the same or corresponding special technical features. For example,the search required for Group
III such as non-full-length cDNA is not required for Group IV while the search required for Group
V such as separating labeled double-stranded cDNA and unlabeled double-stranded cDNA molecules
is not required for Group III.

Groups IV and V do not relate to a single inventive concept under PCT Rule 13.1 because they lack
the same or corresponding special technical features. For example,the search required for Group V
such as an agent that inhibits formation of secondary structure is not required for Group IV.