



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁵ : C12Q 1/68, C12P 19/34 C07H 21/04		A1	(11) International Publication Number: WO 94/02636 (43) International Publication Date: 3 February 1994 (03.02.94)
(21) International Application Number: PCT/US93/00999 (22) International Filing Date: 29 January 1993 (29.01.93) (30) Priority data: 07/922,522 28 July 1992 (28.07.92) US 07/974,406 12 November 1992 (12.11.92) US (60) Parent Applications or Grants (63) Related by Continuation US 07/922,522 (CIP) Filed on 28 July 1992 (28.07.92) US 07/974,406 (CIP) Filed on 12 November 1992 (12.11.92) (71) Applicant (for all designated States except US): HITACHI CHEMICAL COMPANY LTD. [JP/JP]; Shinjyku-Mit- sui Building, No. 1-1, Nishishinjuku 2-chome, Shinjuku- ku, Tokyo 163 (JP).		(72) Inventors; and (75) Inventors/Applicants (for US only): MITSUHASHI, Masa- to [JP/US]; 8 Brookmont, Irvine, CA 92714 (US). COOPER, Allan [US/US]; 3607 N.E. 172nd Avenue, Bellview, WA 98008 (US). (74) Agents: ALTMAN, Daniel, E. et al.; Knobbe, Martens, Olson and Bear, 620 Newport Center Drive, 16th Floor, Newport Beach, CA 92660 (US). (81) Designated States: CA, JP, KR, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report.</i>	
(54) Title: GENE DETECTION SYSTEM			
(57) Abstract A method for detecting the presence of a particular organism, infectious agent, or component of a cell or organism in a biological sample. A polynucleotide probe capable of hybridizing to an analyte polynucleotide belonging to the organism, infectious agent, or cellular component is bound to a solid support, such as a microtiter well, thereby forming a solid support-polynucleotide structure. A second, labeled polynucleotide probe is then hybridized to this support-polynucleotide structure. Alternative embodiments make use of the Polymerase Chain Reaction. Also included in the present invention are polynucleotide probes and primers.			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NE	Niger
BE	Belgium	GN	Guinea	NL	Netherlands
BF	Burkina Faso	GR	Greece	NO	Norway
BG	Bulgaria	HU	Hungary	NZ	New Zealand
BJ	Benin	IE	Ireland	PL	Poland
BR	Brazil	IT	Italy	PT	Portugal
BY	Belarus	JP	Japan	RO	Romania
CA	Canada	KP	Democratic People's Republic of Korea	RU	Russian Federation
CF	Central African Republic	KR	Republic of Korea	SD	Sudan
CG	Congo	KZ	Kazakhstan	SE	Sweden
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovak Republic
CM	Cameroon	LU	Luxembourg	SN	Senegal
CN	China	LV	Latvia	TD	Chad
CS	Czechoslovakia	MC	Monaco	TG	Togo
CZ	Czech Republic	MG	Madagascar	UA	Ukraine
DE	Germany	ML	Mali	US	United States of America
DK	Denmark	MN	Mongolia	UZ	Uzbekistan
ES	Spain			VN	Viet Nam
FI	Finland				

GENE DETECTION SYSTEM

Field of the Invention

5 The present invention relates to methods for detecting the presence of an organism or a member of a group of organisms in a biological sample by probing the sample for polynucleotides indicative of the presence of such organisms. The present invention also relates to methods for detecting the presence of other infectious agents or biological components in a biological sample which comprises polynucleotides.

Background of the Invention

10 At the present time, the identity of an organism or infectious agent suspected of infecting a subject is normally determined by culturing a sample of biological material from the subject. For example, if it is suspected that a subject is suffering from an infection of the lung caused by the fungus *Candida albicans*, a sputum sample can be cultured. After a period of time, the culture is visually
15 observed, and if a fungus grows in the culture in numbers sufficient to indicate a fungal infection, that fungus is identified by observing its morphological characteristics.

 This method of confirming a diagnosis, however, has serious drawbacks. For example, it requires that the biological sample be cultured for a long enough period of time to allow a detectable amount of the organism to grow. This method also requires that the cultured sample be inspected by
20 a technician trained in identifying different varieties of organisms. There is therefore a great need for an assay which can quickly and specifically identify an organism or infectious agent or a group of organisms or infectious agents. An assay which does not require a great deal of training to perform and interpret would also be advantageous.

 There is also a need for an improved method for identifying other biological components
25 present in a biological sample, where such components comprise polynucleotides or where the presence of such components is indicated by the presence of a polynucleotide. Present methods for detecting polynucleotides in a cell or tissue sample, such as the Northern blot method, require a relatively large amount of starting material. The Northern blot method is a widely accepted method of detecting specific genes (Sambrook J. et al., *Molecular Cloning: A Laboratory Manual*, 2nd ed., pp. 7.39-7.52) (hereinafter
30 "*Molecular Cloning*"), and has been adapted for detecting cellular components such as jun oncogenes (Sherman, et al., *Proc. Natl. Acad. Sci. USA*, **87**:5663-5666 (1990); Oursler MJ et al., *Proc. Natl. Acad. Sci. USA*, **88**:6613-6617 (1991)). In this method, mRNA is first purified from a tissue or cell culture, such as through electrophoresis on an agarose gel. Following electrophoresis, the mRNA is transferred onto membranes and hybridized with radioactive probes to identify positive band(s) through

-2-

autoradiography. However, this method is not sensitive enough to identify signals corresponding to specific genes or gene products if the cells from which the mRNA is extracted have only a small quantity of genetic material.

Alternatively, reverse PCR (discussed in *Molecular Cloning*) can also be used to detect a wide variety of genes from different tissues and cells. In this method, mRNA is first converted to cDNA by reverse transcriptase, and specific gene fragments are then amplified by PCR using a set of primers (sense and anti-sense primers). The amplified gene can then be seen through agarose gel electrophoresis.

Summary

The present invention provides an improved method of detecting the presence of an organism, an infectious agent, or a biological component of a cell or organism in a biological sample. In this method, a polynucleotide probe is hybridized to an analyte polynucleotide in the biological sample which belongs to the organism, infectious agent, or biological component or which is indicative of the presence of the organism, infectious agent or biological component. When only small amounts of such an analyte polynucleotide are available, an alternative method can be used which employs the Polymerase Chain Reaction (PCR). In addition, the present invention embodies polynucleotide probes and primers for use in the present methods, as well as kits which incorporate such probes and primers.

In one embodiment, the present invention includes a method of detecting the presence of a particular organism, infectious agent, or biological component of a cell or organism in a biological sample that contains polynucleotides. This method involves detecting an analyte polynucleotide in the sample that is indicative of the presence of the organism, infectious agent, or biological component and comprises the steps of:

(a) identifying a first polynucleotide probe, wherein the nucleotide sequence of the first polynucleotide probe is sufficiently complementary to a first nucleotide sequence contained in the analyte polynucleotide that the first polynucleotide probe can hybridize to the first nucleotide sequence of the analyte polynucleotide, the first nucleotide sequence of the analyte polynucleotide being specific to the particular organism, infectious agent, or biological component;

(b) immobilizing the first polynucleotide probe to a solid support;

(c) hybridizing the analyte polynucleotide in the sample with the first polynucleotide probe;

(d) identifying a second polynucleotide probe, wherein the nucleotide sequence of the second polynucleotide probe is sufficiently complementary to a second nucleotide sequence contained in the analyte polynucleotide that the second polynucleotide probe can hybridize to the second nucleotide sequence, the second nucleotide sequence being common to a plurality

-3-

of organisms, infectious agents, or biological components including the particular organism, infectious agent, or biological component;

(e) hybridizing the second polynucleotide probe with the analyte polynucleotide which is hybridized to the first polynucleotide probe; and

(f) determining the presence of the particular organism, infectious agent, or biological component in the sample by detecting the presence of the second polynucleotide probe on the solid support.

In another embodiment, the present method for detecting the presence of an organism, infectious agent, or biological component of a cell or organism in a biological sample comprises the steps of:

(a) immobilizing a first polynucleotide probe to a solid support, wherein the nucleotide sequence of the first polynucleotide probe is sufficiently complementary to a first nucleotide sequence contained in an analyte polynucleotide in the organism, infectious agent, or biological component that the first polynucleotide probe can hybridize to the first nucleotide sequence of the analyte polynucleotide of the organism, infectious agent, or biological component;

(b) contacting the polynucleotides present in the sample with the first polynucleotide probe;

(c) hybridizing the analyte polynucleotide in the sample to the first polynucleotide probe, if the analyte polynucleotide is present in the sample;

(d) contacting a second polynucleotide probe with the analyte polynucleotide hybridized to the first polynucleotide probe, if the analyte polynucleotide from the sample has hybridized to the first polynucleotide probe, wherein the nucleotide sequence of the second polynucleotide probe is sufficiently complementary to a second nucleotide sequence contained in the analyte polynucleotide of the organism, infectious agent, or biological component that the second polynucleotide probe can hybridize to the second nucleotide sequence;

(e) hybridizing the second polynucleotide probe to the analyte polynucleotide hybridized to the first polynucleotide probe, if the analyte polynucleotide has hybridized to the first polynucleotide probe; and

(f) determining the presence of the organism, infectious agent, or biological component in the sample by detecting the presence of the second polynucleotide probe hybridized to the analyte polynucleotide which has hybridized to the first polynucleotide probe.

In this method, the second polynucleotide probe can preferably have the same or lower T_m as the first polynucleotide probe. Preferably, the first polynucleotide probe also has a T_m within the range of from approximately 48°C to approximately 60°C. The first nucleotide sequence of the analyte polynucleotide can also in one embodiment be common to a plurality of organisms, infectious agents, or biological components of a cell or organism. Alternatively, the first nucleotide sequence of the

analyte polynucleotide can be specific to a particular organism, infectious agent, or biological component of a cell or organism.

5 The second nucleotide sequence of the analyte polynucleotide can have a sequence common to a plurality of organisms, infectious agents, or biological components of a cell or organism in this embodiment of the present method. In an alternative embodiment, the second nucleotide sequence of the analyte polynucleotide can have a sequence that is specific for a particular organism, infectious agent, or biological component of a cell or organism. Additionally, a label can advantageously be attached to the second polynucleotide probe. Any of a number of polynucleotide labels known to the art can be used, including a radionuclide, an enzyme, an enzyme substrate, a specific binding moiety, 10 an binding partner for a specific binding moiety, biotin, avidin, a nucleic acid stain, or a fluorescent material. If a nucleic acid stain is used as the label, the stain can consist of either ethidium bromide, yoyo-1, or toto-1. When the label is a light-emitting substance, the label can advantageously be detected by measuring the amount of light emitted therefrom. When measuring the amount of light emitted by the label, light can be recorded on film, after which the amount of exposure of the film is measured using a densitometer. In an even more preferable embodiment, the label comprises alkaline phosphatase and the label is detected by adding ATTOPHOS to the solution containing the labeled probe and then measuring the fluorescence emitted using a fluorimeter. 15

A solid support such as a microtiter plate having a plurality of wells can be used to perform the present method. Preferably, each of the wells has a specific polynucleotide probe immobilized thereon. 20 The first polynucleotide probe, which can be immobilized on the microtiter plate, advantageously comprises DNA, and more advantageously both the first and second polynucleotide probes comprise DNA.

A further preferable step of the method of the present invention involves washing the solid support after hybridizing the analyte polynucleotide in the sample to the first polynucleotide probe. In 25 this way, substantially all of the biological sample not annealed to the first polynucleotide probe is removed from the solid support. Yet another step of the method includes the step of washing the solid support after hybridizing the second polynucleotide probe with the analyte polynucleotide, which is itself hybridized to the first polynucleotide probe. After such washing, substantially all of the second polynucleotide probe not hybridized with the analyte polynucleotide is removed from the solid support. 30 The analyte polynucleotide in this method can be selected from the group consisting of mRNA, rRNA, and genomic DNA.

Another embodiment of the present invention is a method of detecting the presence of an organism, infectious agent, or biological component of a cell or organism in a biological sample containing polynucleotides, comprising the steps of:

35 (a) identifying a first polynucleotide probe and a second polynucleotide probe, wherein the nucleotide sequence of the first polynucleotide probe is sufficiently complementary to a first nucleotide sequence contained in an analyte polynucleotide of the organism, infectious

-5-

agent, or biological component that the first polynucleotide probe can hybridize to the first nucleotide sequence of the analyte polynucleotide of the organism, infectious agent, or biological component, and wherein the nucleotide sequence of the second polynucleotide probe is sufficiently complementary to a second nucleotide sequence contained in the analyte polynucleotide of the organism, infectious agent, or biological component that the second polynucleotide probe can hybridize to the second nucleotide sequence, the second nucleotide sequence being common to a plurality of organisms, infectious agents, or biological components;

(b) immobilizing the first polynucleotide probe to a solid support;

(c) contacting the polynucleotides present in the sample with the first polynucleotide probe;

(d) hybridizing an analyte polynucleotide in the sample to the first polynucleotide probe, if the analyte polynucleotide is present in the sample;

(e) contacting the second polynucleotide probe with the analyte polynucleotide hybridized to the first polynucleotide probe, if the analyte polynucleotide from the sample has hybridized to the first polynucleotide probe;

(f) hybridizing the second polynucleotide probe to the analyte polynucleotide hybridized to the first polynucleotide probe, if the analyte polynucleotide has hybridized to the first polynucleotide probe; and

(g) determining the presence of the organism, infectious agent, or biological component in the sample by detecting the presence of the second polynucleotide probe hybridized to the analyte polynucleotide which has hybridized to the first polynucleotide probe.

In this embodiment, the identifying step can advantageously comprise the use of a computer, preferably one which uses an H-site model to identify the first polynucleotide probe. Using the H-site model to identify the first polynucleotide probe involves the steps of:

specifying a minimum melting temperature for the first nucleotide probe and the nucleotide sequence specific to the organism;

specifying a nucleation threshold that places a minimum value on the number of base pairs at any nucleation site;

determining the melting temperatures (T_m) of the first nucleotide probe and the sequence specific to the organism at every possible hybridization point; and

selecting the nucleotide probe having the highest T_m value.

When using the H-site model, the melting temperature is preferably determined by the formula;

$T_m = 81.5 - 16.6(\log[Na]) - 0.63\%(\text{formamide}) + 0.41(\%/(G+C)) - 600/N$, wherein $\log[Na]$ is the log function of the sodium concentration, 0.063% (formamide) is the concentration of formamide, $\%(G+C)$ is the percentage of matched GC base pairs, and N is the probe length.

-6-

In this method the second polynucleotide probe can also be identified using a computer which makes use of the H-site model. Identifying the second polynucleotide probe with the H-site model is preferably accomplished by following the steps of:

- 5 specifying a minimum melting temperature for the second nucleotide probe and the nucleotide sequence specific to the organism;
- specifying a nucleation threshold that places a minimum value on the number of base pairs at any nucleation site;
- determining the melting temperatures (T_m) of the first nucleotide probe and the sequence specific to the organism at every possible hybridization point; and
- 10 selecting the nucleotide probe of the proper length having the lowest T_m value.

The melting temperature of the second polynucleotide probe can likewise be determined by the formula; $T_m = 81.5 - 16.6(\log[Na]) - 0.63\%(\text{formamide}) + 0.41(\%(G+C)) - 600/N$, wherein $\log[Na]$ is the log function of the sodium concentration, 0.063% (formamide) is the concentration of formamide, $\%(G+C)$ is the percentage of matched GC base pairs, and N is the probe length.

- 15 In this method, the second polynucleotide probe can advantageously have the same or lower T_m as the first polynucleotide probe. The first polynucleotide probe also preferably has a T_m within the range of from approximately 48°C to approximately 60°C. The first nucleotide sequence of the analyte polynucleotide can also in one embodiment be common to a plurality of organisms, infectious agents, or biological components of a cell or organism. Alternatively, the first nucleotide sequence of the
- 20 analyte polynucleotide can be specific to a particular organism, infectious agent, or biological component of a cell or organism.

- The second nucleotide sequence of the analyte polynucleotide can have a sequence common to a plurality of organisms, infectious agents, or biological components of a cell or organism in this embodiment of the present method. In an alternative embodiment, the second nucleotide sequence of
- 25 the analyte polynucleotide can have a sequence that is specific for a particular organism, infectious agent, or biological component of a cell or organism. Additionally, a label can advantageously be attached to the second polynucleotide probe. Any of a number of polynucleotide labels known to the art can be used, including a radionuclide, an enzyme, an enzyme substrate, a specific binding moiety, an binding partner for a specific binding moiety, biotin, avidin, a nucleic acid stain, or a fluorescent
 - 30 material. If a nucleic acid stain is used as the label, the stain can consist of either ethidium bromide, yoyo-1, or toto-1. When the label is a light-emitting substance, the label can advantageously be detected by measuring the amount of light emitted therefrom. When measuring the amount of light emitted by the label, light can be recorded on film, after which the amount of exposure of the film is measured using a densitometer. In an even more preferable embodiment, the label comprises alkaline
 - 35 phosphatase and the label is detected by adding ATTOPHOS to the solution containing the labeled probe and then measuring the fluorescence emitted using a fluorimeter.

-7-

A solid support such as a microtiter plate having a plurality of wells can be used to perform the present method. Preferably, each of the wells has a specific polynucleotide probe immobilized thereon. The first polynucleotide probe, which can be immobilized on the microtiter plate, advantageously comprises DNA, and more advantageously both the first and second polynucleotide probes comprise DNA.

A further preferable step of the method of the present invention involves washing the solid support after hybridizing the analyte polynucleotide in the sample to the first polynucleotide probe. In this way, substantially all of the biological sample not annealed to the first polynucleotide probe is removed from the solid support. Yet another step of the method includes the step of washing the solid support after hybridizing the second polynucleotide probe with the analyte polynucleotide, which is itself hybridized to the first polynucleotide probe. After such washing, substantially all of the second polynucleotide probe not hybridized with the analyte polynucleotide is removed from the solid support. The analyte polynucleotide in this method can be selected from the group consisting of mRNA, rRNA, and genomic DNA.

Another embodiment of the present invention comprises a solid support-polynucleotide structure for identifying the presence of an organism, infectious agent, or biological component of a cell or organism in a biological sample containing polynucleotides. This structure comprises a solid support having immobilized thereto a first polynucleotide probe, the first polynucleotide probe having a sequence complementary to a first nucleotide sequence specific to the organism, infectious agent, or biological component. The structure also includes an analyte polynucleotide from the cell, organism, or infectious agent which contains the first nucleotide sequence, the analyte polynucleotide being hybridized to the first polynucleotide probe at the first nucleotide sequence. The structure includes as well a second polynucleotide probe, preferably having a label, which is complementary to a second nucleotide sequence present on the analyte polynucleotide from the cell, organism, or infectious agent which is hybridized to the first polynucleotide probe, the second polynucleotide probe being hybridized to the analyte polynucleotide at the second nucleotide sequence.

The label of the above method is preferably selected from the group consisting of a radionuclide, an enzyme, an enzyme substrate, a specific binding moiety, an binding partner for a specific binding moiety, biotin, avidin, a nucleic acid stain, and a fluorescent material. Additionally, the solid support-polynucleotide structure can even more advantageously have a second polynucleotide probe that is common to polynucleotides contained in a plurality of organisms, infectious agents, or biological components of a cell or organism. In still another preferred embodiment of this method, the first and second polynucleotide probes are determined through the use of a computer system for designing oligonucleotide probes for use with a gene sequence data source. This the computer system can comprise an input means for retrieving the gene sequence data, a processor, and instructions directing the processor to determine the first oligonucleotide probe. Yet still more preferably, the solid support-

-8-

polynucleotide structure of the above method has a polynucleotide selected from the group consisting of mRNA, rRNA, and genomic DNA.

Another embodiment of the present invention is a kit for identifying the presence of an organism, infectious agent, or biological component of a cell or organism in a biological sample which contains:

a specific polynucleotide probe, the specific polynucleotide probe being complementary to or homologous to a first nucleotide sequence in an analyte polynucleotide specific to a particular organism, infectious agent, or biological component to be detected; and

a common polynucleotide probe complementary to or homologous to a second nucleotide sequence in the analyte polynucleotide of the organism, infectious agent, or biological component, the common polynucleotide probe being complementary to polynucleotides contained in a plurality of organisms, infectious agents, or biological components.

In addition, the above kit can advantageously have a solid support to which a polynucleotide can be immobilized. Even more preferably, the kit has a specific polynucleotide probe immobilized to the solid support. Still more advantageously, the above kit has a solid support with a plurality of specific polynucleotide probes immobilized thereto, each of the probes specific to a different organism, infectious agent, or biological component. Most preferably, the solid support of the kit has a plurality of wells, each of the specific polynucleotide probes being immobilized to a different well, and a buffer appropriate for the hybridization of the probes and polynucleotides, with the polynucleotides being selected from the group consisting of mRNA, rRNA, and genomic DNA.

Similarly, the second polynucleotide probe of the above kit can advantageously bear a label, with the label being selected from the group consisting of a radionuclide, an enzyme, an enzyme substrate, a specific binding moiety, an binding partner for a specific binding moiety, biotin, avidin, a nucleic acid stain, and a fluorescent material. Further, the above kit can have a specific polynucleotide probe comprising a first specific primer which is complementary or homologous to a sequence specific to a particular organism, infectious agent, or biological component, the kit additionally comprising a second specific polynucleotide primer which is complementary or homologous to a different sequence specific to the organism, infectious agent, or biological component.

When the above kit is designed to be used with PCR, it should have one or more of the following: dNTP's, a reverse transcriptase, a polymerase, and a buffer appropriate for the addition of dNTP's to a primer using a reverse transcriptase or polymerase. Additionally, the kit can include a DNA polymerase that has significant polymerase activity at temperatures above 50°C.

In yet another embodiment, the present method comprises a means of detecting the presence of one or more organisms, infectious agents, or biological components in a biological sample containing polynucleotides, wherein at least one of the polynucleotides is indicative of the presence of the one or

-9-

more organisms, infectious agents or biological components and is present in minute quantities. This method makes use of PCR and employs the steps of:

- (a) obtaining a biological sample containing polynucleotides;
- (b) contacting the sample with a first polynucleotide primer, the first primer having
5 a nucleotide sequence complementary to a nucleotide sequence common to a plurality of organisms, infectious agents, or biological components;
- (c) hybridizing the first primer to an analyte polynucleotide present in the sample that is complementary to the first primer, if such an analyte polynucleotide is present;
- (d) extending the first primer, thereby producing a double-stranded polynucleotide
10 including a complementary nucleotide strand comprising the first primer and having a nucleotide sequence complementary to the analyte polynucleotide;
- (e) contacting the sample with a second polynucleotide primer, the second primer being complementary to a sequence contained in the complementary nucleotide strand;
- (f) hybridizing the second primer to the complementary nucleotide strand;
- (g) extending the second primer to form a nucleotide strand homologous to the
15 analyte polynucleotide;
- (h) contacting the sample with a third and a fourth polynucleotide primer, the third and fourth primers having sequences complementary to the homologous nucleotide strand and the complementary nucleotide strand, respectively, wherein the third primer has a
20 nucleotide sequence complementary to a sequence common to a plurality of organisms, infectious agents, or biological components whose presence is to be determined and wherein the sequence of the third primer is different from that of the first primer;
- (i) hybridizing the third and fourth primers to the complementary nucleotide strand and the homologous nucleotide strand;
- (j) extending the third and fourth primers, thereby producing double-
25 stranded polynucleotides; and
- (k) determining the presence of the one or more organisms, infectious agents, or biological components in the sample by detecting the extension of at least one of the first, second, third, or fourth primers.

30 In this method, the second primer can have a nucleotide sequence that is common to a plurality of the organisms, infectious agents, or biological components whose presence is being determined. This method is preferably practiced such that the extending and hybridizing steps are repeated a plurality of times. In this method, the extension step can be accomplished with a reverse transcriptase when the primer is bound to RNA, while this step is accomplished with a DNA polymerase when the bound
35 polynucleotide is DNA. If a DNA polymerase is used, it preferably has significant polymerase activity at temperatures above 50°C. In the present method the nucleotide sequences of the first and second primers can be determined by a computer-assisted method, and preferably by a computer-assisted

-10-

method which determines the sequences of the first and second nucleotide probes using an H-Site model.

Yet another method of detecting the presence of one or more organisms, infectious agents, or biological components in a biological sample containing polynucleotides, wherein at least one of the polynucleotides is indicative of the presence of the one or more organisms, infectious agents or biological components and is present in minute quantities, comprises the steps of:

obtaining a biological sample containing polynucleotides;

contacting the sample with a first polynucleotide primer, the first primer having a nucleotide sequence complementary to a nucleotide sequence common to a plurality of organisms, infectious agents, or biological components;

hybridizing the first primer to an analyte polynucleotide present in the sample that is complementary to the first primer, if such an analyte polynucleotide is present;

extending the first primer, thereby producing a double-stranded polynucleotide including a complementary nucleotide strand comprising the first primer and having a nucleotide sequence complementary to the analyte polynucleotide;

contacting the sample with a second polynucleotide primer, the second primer being complementary to a sequence contained in the complementary nucleotide strand;

hybridizing the second primer to the complementary nucleotide strand;

extending the second primer to form a nucleotide strand homologous to the analyte polynucleotide;

contacting the sample with a third polynucleotide primer, the third primer having a sequence complementary to the homologous nucleotide strand, wherein the third primer has a nucleotide sequence complementary to a sequence that is specific to a particular organism, infectious agent, or biological component whose presence is to be determined;

hybridizing the third primer to the homologous nucleotide strand;

extending the third primer, thereby producing a double-stranded polynucleotide; and

determining the presence of the particular organism, infectious agent, or biological component in the sample by detecting the extension of at least one of the third primer.

This method is preferably practiced such that the extending and hybridizing steps are repeated a plurality of times. The extension step in particular can be accomplished with a reverse transcriptase when the primer is bound to RNA, while a DNA polymerase is used when the bound polynucleotide is DNA. Such a DNA polymerase preferably has significant polymerase activity at temperatures above 50°C.

In the present method the nucleotide sequences of the first and second primers can be determined by a computer-assisted method, and preferably by a computer-assisted method which determines the sequences of the first and second nucleotide probes using an H-Site model. In this

-11-

method, the second primer can have a nucleotide sequence that is common to a plurality of the organisms, infectious agents, or biological components whose presence is being determined. The third primer also preferably includes a label, so that the detecting step comprises the extension of this labeled primer. The label used can be selected from the group consisting of a radionuclide, an enzyme, an enzyme substrate, a specific binding moiety, a binding partner for a specific binding moiety, biotin, avidin, a nucleic acid stain, and a fluorescent material.

Yet another method of detecting the presence of a particular organism, infectious agent, or biological component in a biological sample containing polynucleotides, wherein at least one of the polynucleotides is indicative of the presence of the organism, infectious agent, or biological component and is present in minute quantities, comprises the steps of:

- (a) obtaining a biological sample containing polynucleotides;
- (b) contacting the sample with a first polynucleotide primer, the first primer having a nucleotide sequence complementary to a nucleotide sequence that is specific to the particular organism, infectious agent, or biological component, wherein the nucleotide sequence of the first primer is determined by means of a computer-assisted method;
- (c) hybridizing the first primer to a sample polynucleotide present in the sample that is complementary to the first primer, if such a sample polynucleotide is present;
- (d) extending the first primer, thereby producing a double-stranded polynucleotide including a complementary nucleotide strand comprising the first primer and having a nucleotide sequence complementary to the sample polynucleotide;
- (e) contacting the sample with a second polynucleotide primer, the second primer being complementary to a sequence contained in the complementary nucleotide strand;
- (f) hybridizing the second primer to the complementary nucleotide strand;
- (g) extending the second primer to form a nucleotide strand homologous to the sample polynucleotide; and
- (h) determining the presence of the particular organism, infectious agent, or biological component in the sample by detecting the extension of at least one of the first or second primers.

We have also discovered a number of useful probes and primers for use in the foregoing methods, including primers for detecting jun oncogenes, G proteins, β receptors, and Substance P receptors, such as those identified with a sequence identifier herein. Among the sequences we have discovered for use as probes and primers in the present method are SEQ ID NO:473, SEQ ID NO:600, SEQ ID NO:615, SEQ ID NO:622, SEQ ID NO:622, SEQ ID NO:730, SEQ ID NO:747, SEQ ID NO:748, SEQ ID NO:488, SEQ ID NO:513, SEQ ID NO:630, SEQ ID NO:639, SEQ ID NO: 728, SEQ ID NO:729, SEQ ID NO:733, SEQ ID NO:734, SEQ ID NO:739, SEQ ID NO:740, SEQ ID NO:741, SEQ ID NO:742, SEQ ID NO:743, SEQ ID NO:744, SEQ ID NO:759, SEQ ID NO:758, SEQ ID NO:751, SEQ ID NO:553, SEQ ID NO:670, SEQ ID NO:752, SEQ ID NO:753, SEQ ID NO:565, SEQ

-12-

ID NO:678, SEQ ID NO:686, SEQ ID NO:754, SEQ ID NO:577, SEQ ID NO:697, SEQ ID NO:704, SEQ ID NO:755, SEQ ID NO:756, SEQ ID NO:732, SEQ ID NO:642, SEQ ID NO:652, SEQ ID NO:757, SEQ ID NO:593, SEQ ID NO:710, SEQ ID NO:721, SEQ ID NO:528, SEQ ID NO:731, SEQ ID NO:749, and SEQ ID NO:750

5

Brief Description of the Figures

Figure 1 is a schematic representation of one embodiment of the present inventive method.

Figure 1a is a picture of a gel showing the effect of various RNase inhibitors on mRNA preparations containing SDS and EDTA.

5 Figure 1b is a graph showing the relationship between YOYO-1 fluorescence and amount of immobilized oligonucleotide.

Figure 1c is a graph showing the time course of YOYO-1 analysis.

Figure 1d is a graph showing the dose response of YOYO-1 concentration.

10 Figure 1e is a bar graph showing the reproducibility of YOYO-1 staining based on quantitation of immobilized oligonucleotides.

Figure 2 is a schematic representation of the common and specific sequences identified in various species of fungi.

Figure 3 is schematic representation of an embodiment of the method of the present invention in which PCR is used.

15 Figure 4 is a picture of a gel showing the outcome of an experiment in which jun-D, c-jun, and jun-D jun oncogene subtypes were amplified using SEQ ID NO:728 and SEQ ID NO:729.

Figure 5 is a picture of a gel showing the outcome of an experiment in which cloned rat G_{i-1} , G_{i-2} , G_{i-3} , G_s , and G_o G protein subtypes were amplified using SEQ ID NO:528 and SEQ ID NO:731.

20 Figure 6 is a schematic representation of an example of a microtiter plate used in the methods of the present invention.

Figure 7 is a picture of the gel referred to in Figure 7 which also provides Southern blots using each of the five G protein sequences as a probe.

Figure 7A-7C is a graphic representation of the results of an experiment which shows that SEQ ID NOS:470, 488 and 730 can be used to detect specific subtypes of jun oncogenes.

25 Figure 8 shows a gel of samples containing varying amounts of each of five different G protein oligonucleotides (as indicated by number of "+" symbols) amplified with the G_2 and G_4 PCR primers and also provides Southern blots using each of the five G protein sequences as a probe.

Figure 9 shows a gel of λ ZAP cDNA libraries from rat pituitary (P), kidney (K) and intestinal (I) amplified with each of five different G protein primers, as indicated.

30 Figure 10 shows a gel of 500 bp DNA from cDNAs of human IM9 and Jurkat cells amplified with G_2 and G_4 PCR primers.

Figure 11 is a simplified block diagram of a computer system illustrating the overall design of this invention;

35 Figures 12A-12C show display screen representations of the main oligoprobe design station dialog windows of this invention;

Figures 13A and 13B are flow charts of the overall invention illustrating the program and the invention's sequence and structure;

Figure 14 is a display screen representation of the Mitsuhashi probe selection diagram;

Figure 15 is a display screen representation of the probeinfo and matchinfo window;

5 Figure 16 is a display screen representation of the probesedit window;

Figures 16A and 16B are printouts of the probesedit output file;

Figure 17 is a flow chart of the overall k_diff program of the Mismatch Model of this invention, including its sequence and structure;

Figures 18A and 18B are flow charts of the k_diff module of this invention;

10 Figure 19 is a flow chart of the hashing module of this invention;

Figure 20 is a flow chart of the tran module of this invention;

Figure 21 is a flow chart of the let_dig module of this invention;

Figure 22 is a flow chart of the update module of this invention;

Figures 23A and 23B are flow charts of the assembly module of this invention;

15 Figures 24A and 24B are flow charts of the seqload module of this invention;

Figures 25A and 25B are flow charts of the read1 module of this invention;

Figure 26 is a flow chart of the dig_let module of this invention;

Figures 27A and 27B are flow charts of the q_colour module of this invention;

Figure 28 is a flow chart of the hit_ext module of this invention;

20 Figure 29 is a flow chart of the colour module of this invention;

Figure 30 is the first page of a printout of a sample file containing the output of the Mismatch Model program of this invention;

Figure 31 is a flow chart of the H-Site Model, stage I, covering the creation of a preprocessed preparation file of this invention;

25 Figure 32 is a flow chart of the H-Site Model, stage II, covering the preparation of the target sequence(s);

Figures 33A-33D are flow charts of the H-Site Model, stage III, covering the calculation of MPSD data;

30 Figure 34A is the first page of a printout of a sample file containing output of Mismatch Model program;

Figure 34B is the first page of a printout of a sample file containing output of H-Site Model program;

Figure 35 is a flow chart of the processing used to create the Mitsuhashi probe selection diagram (MPSD);

35 Figure 36 is a flow chart of processing used to create the matchinfo window;

Figure 37 is a printout of a sample target species file;

Figures 38A-38C are printouts of a sample preparation file.

Detailed Description of the Invention

The present method of detecting organisms, infectious agents, or other biological components in a biological sample which contains polynucleotides represents an advance over prior art diagnostic tests in its speed, ease of use, and accuracy. For example, the time necessary to make an accurate identification of an organism in a biological sample using prior art methods required enough time to culture such an organism, which could take from hours to days. Such methods also required that a technician trained to differentiate between different disease-causing organisms examine cultures of a biological sample. The present method, on the other hand, can be accomplished by someone without such specialized training. The present method can also be performed so as to identify either a specific organism or a group of organisms, such as a genus of fungus or bacteria, without the possibility of error due to the mis-identification of an organism by a technician examining a culture of a biological sample. Infectious agents such as viruses can also be detected. ***

The present method also includes an improved way of detecting other biological components which comprise polynucleotides or whose presence is indicated by a polynucleotide. A biological sample can thus be probed with oligonucleotides specific to a biological component, such as a jun oncogene or G protein mRNA. In the present application, "biological component" means a component of a cell or organism which comprises a polynucleotide, such as mRNA, rRNA, and genomic DNA. A component which does not contain a polynucleotide but whose presence is indicated by presence of a polynucleotide in the cell or organism is also included where the detection of a polynucleotide is an indication of the presence of the component. Other probes which can be specific to the biological component or which can be common to the biological component and other biological components can then be used to detect the binding of the specific probe to the biological component. Alternatively, a biological sample can be probed with oligonucleotides capable of binding to a number of polynucleotides with related gene sequences in order to increase the sensitivity of the assay. These and other aspects of the present invention will be described in greater detail below.

I. Common and Specific Sequences

We have discovered a number of specific sequences that are unique to the rRNA of a single fungal species or genus which can be used in the present method. These sequences are SEQ ID NO:81, SEQ ID NO:104, SEQ ID NO:131 through SEQ ID NO:133, SEQ ID NO:154 through SEQ ID NO:156, SEQ ID NO:176, SEQ ID NO:199, SEQ ID NO:267, SEQ ID NO:290, SEQ ID NO:312, SEQ ID NO:335, SEQ ID NO:364 through SEQ ID NO:376, and SEQ ID NO:391 through SEQ ID NO:392. Reference can be made to Tables V through XVI for a comparison of these specific sequences to corresponding sequences in other species or genera.

SEQ ID NO:227 and SEQ ID NO:250 are sequences that are specific to the rRNA of certain strains of *C. albicans*. However, reported sequences in other strains of the same species have slight changes in these sequences in their rRNA, as seen in SEQ ID NO:226 and SEQ ID NO:249,

-16-

respectively. Accordingly, these particular sequences are less preferred for use as specific sequences within the context of the present invention. However, these sequences can be useful for identifying the particular strain of *C. albicans* in a sample.

We have also discovered a number of common sequences that are common to the rRNA of several fungal species and genera. These sequences are SEQ ID NO:1 through SEQ ID NO:80. Reference can be made to Tables 1 through 4 to see that these sequences are common to all species shown.

In a further discovery, we have discovered sequences common to a number of rat and human G protein sequences, as well as sequences that are specific to particular G protein subtypes. Oligonucleotides containing such sequences can be useful as probes or primers for identifying the presence of such sequences in a sample in which such sequences are present in only small quantities. As will be discussed *infra* in more detail, oligonucleotides containing the inventive sequences can be used as primers in the Polymerase Chain Reaction (PCR) to detect G protein sequences in a biological sample.

Sequences have as well been identified which are common to the different jun oncogenes which have been identified in humans and in mice. Such sequences can also be used in PCR primers to identify the presence of jun gene sequences in a biological sample. In addition, sequences specific to particular subtypes of jun oncogenes have also been discovered. Other sequences useful in the present method are described below.

Advantageously, all of these probes, both common and specific, have been designed to have approximately, the same melting temperature (T_m) when annealed to a complementary sequence. Thus, various procedures requiring annealing of these sequences can all be performed under the same conditions. Those of skill in the art will recognize that longer or shorter sequences, with a correspondingly higher or lower T_m , respectively, can also be obtained upon reference to the full-length sequences available from GenBank. When the term " T_m " is used herein in connection with a single-stranded polynucleotide, this term refers to the melting temperature of that single-stranded polynucleotide when it is annealed to a complementary strand.

As is known to those of skill in the art, the T_m of a polynucleotide strand can be determined using the following formulas:

$$(a) \quad T_m = 69.3 + 0.41 (G + C)\% - 650/L$$

(where G is the number of guanine residues in the strand, C is the number of cytosine residues, and L is the total length, in bases, of the polynucleotide);

$$(b) \quad (T_m)_{u_2} - (T_m)_{u_1} = 18.5 \log_{10} u_2/u_1$$

(where u_1 and u_2 are the ionic strengths of two solutions); and

$$(c) \quad \text{The } T_m \text{ of duplex DNA decreases } 1^\circ\text{C with every increase of } 1\% \text{ in the number of mismatched base pairs.}$$

In a preferred embodiment of the present invention, a plurality of probes that have the same T_m are immobilized to one or more solid supports. When probes having the same T_m are used, such probes can be hybridized together under the same conditions because they require the same reaction temperatures. Preferably, the specific probes in this embodiment have a T_m between approximately 48°C and 60°C. Other probes that have the same T_m and are within this range can be determined by using the formulas above and by performing routine experimentation.

Unless otherwise specified, in the present application the term "specific sequence" denotes a sequence of nucleic acids which is present only in a specific organism or infectious agent, or which is present in a biological sample only as a result of the presence of a specific organism or infectious agent. A "specific sequence" can also be one present in a particular kind of biological component, such as the mRNA of a subtype of jun oncogene. Of course, the sequence complementary to a specific sequence can also be said to be specific. The term "complementary" is used herein to describe a polynucleotide sequence in which adenine is replaced by thymine or a nucleotide that reacts in an equivalent way to thymine such as uracil, and in which thymine (or uracil) is replaced by adenine or a nucleotide that reacts in a similar way to adenine. In such a complementary molecule the guanine residues would also be replaced by cytosines or equivalent nucleotide molecules, and the cytosine residues would be replaced by guanine or equivalent nucleotides.

In the present application, the term "homologous" is used to describe a polynucleotide having a sequence which contains the same nucleotides or equivalent nucleotides, in the same order, as another polynucleotide. For example, a second polynucleotide having the same sequence of nucleotides as a first polynucleotide but in which uracil residues have been substituted for the thymine residues is homologous to the first polynucleotide. Other equivalent nucleotide substitutions known to the art are also included.

Sequences useful in the methods of the present invention can be determined in any way known to the art. Preferably, such sequences are identified with a computer program, as will be detailed *infra*. Therefore, the sequences discussed herein are only examples of sequences which will work in the present invention and are not the only sequences which can be used.

II. Fungus Assay

One example of the method of the present invention involves the detection of a particular species of fungus in a biological sample. We have discovered that the presence of a particular species of fungus can be determined by probing the ribosomal RNA of a sample for sequences specific to the ribosomal RNA of a particular species of fungus. Probes which carry sequences complementary to sequences found in a group of fungi which include the specific fungus probed for can then be used to detect the specific species of fungus. Those of skill in the art will recognize that other organisms, infectious agents, and biological components present in a biological sample can also be detected using

the present method. Those of skill in the art will similarly recognize that other kinds of nucleic acids present in a biological sample can also be probed, including mRNA and genomic DNA.

In the present example, it has been found that each one of a plurality of fungal species carries ribosomal RNA sequences specific to only one species of fungus. The presence of a particular species of fungus in a biological sample can thus be detected by probing that sample for a sequence of ribosomal RNA found only in the ribosomal RNA of that species of fungus. The specific ribosomal RNA sequences found in a number of species of fungus appear to occur in regions that pick up mutations at a relatively high rate. Thus, many species of fungi are likely to have different nucleotide sequences in those regions. Although ribosomal RNA is not expressed, such regions would be analogous to unexpressed regions of genomic DNA, which pick up mutations at a relatively faster rate than expressed regions.

It has been further discovered that a number of species of fungi share sequences of ribosomal RNA common to all of those species. Thus, the presence of any of those fungal species in a biological sample can be determined by probing the sample for polynucleotides having such common sequences. If a fungus contains both specific and common sequences, probing for such common sequences can be used to detect the presence of a variety of that species of fungus which contains one or more mutations in its specific ribosomal RNA sequences. The existence of common sequences can also be exploited by annealing labeled probes to those sequences in order to facilitate the detection of polynucleotides which carry such common sequences.

In one group of pathogenic fungal species, two separate ribosomal RNA sequences have been identified in each of the species which are specific to the individual species carrying such sequences. This group comprises the following fungal species: *Pneumocystis carinii*, *Aspergillus fumigatus*, *Aspergillus fumigatus*, *Cryptococcus neoformans*, *Coccidioides immitis*, *Blastomyces dermatitidis*, and a number of species in the *Candida* group, including *Candida albicans* and *Candida tropicalis*. The Genbank accession numbers of the ribosomal RNA of the fungal species of this group are shown in Table A below:

Table 1

	<u>Fungal Species</u>	<u>Accession No.</u>
30	<i>Aspergillus fumigatus</i>	M55626
	<i>Aspergillus fumigatus</i>	M60300
	<i>Aspergillus fumigatus</i>	M60301
	<i>Blastomyces dermatitidis</i>	M55624
	<i>Candida albicans</i>	M60302
35	<i>Candida albicans</i>	X53497
	<i>Candida guilliermondii</i>	M60304

-19-

	<i>Candida glabrata</i>	X51831
	<i>Candida kefyr</i>	M60303
	<i>Candida krusei</i>	M55528
	<i>Candida krusei</i>	M60305
5	<i>Candida lusitanae</i>	M55526
	<i>Candida lusitanae</i>	M60306
	<i>Candida parapsilosis</i>	M60307
	<i>Candida tropicalis</i>	M55527
	<i>Candida tropicalis</i>	M60308
10	<i>Candida viswanathii</i>	M60309
	<i>Pneumocystis carinii</i>	X12708
	<i>Coccidioides immitis</i>	M55627
	<i>Cryptococcus neoformans</i>	M55625

15 In the embodiment of the present invention comprising a fungal assay, the term "specific sequence" is used to indicate a sequence of ribosomal RNA specific to one species of fungus. Such a sequence is specific to that fungus species and thus is not found in the ribosomal RNA of any other fungal species. The sequence is also different from other RNA sequences found in the cells being tested. A probe which has a sequence complementary to one of these specific sequences will therefore
20 anneal only to the ribosomal RNA of a particular species of fungus, or to a polynucleotide homologous to such ribosomal RNA.

In the group of pathogenic fungi containing specific sequences referred to above, four common sequences of ribosomal RNA have also been identified. "Common sequences" are those common to a group of organisms, such as those common to a particular genus or family of organisms. The term
25 "common" can also denote sequences shared by a group of infectious agents or biological components, such as sequences shared by different subtypes of *jun* oncogene. As shown in Figure 1, two such common sequences for the group of fungi listed above in Table 1 occur 5' of the specific sequences identified in such fungi, while two other common sequences are located 3' of these specific sequences.

30 Thus, in one embodiment of the present invention, primers complementary to the common sequences located 3' of the specific sequences in the ribosomal RNA of a fungal species of this group are used to create a polynucleotide, preferably a strand of cDNA, complementary to the portion of the ribosomal RNA of the species that contains the specific sequences. A probe homologous to one of the common sequences located 5' of the specific sequences can also be annealed to a strand complementary
35 to the ribosomal RNA of a species of fungus and then extended in order to create a polynucleotide strand homologous to the strand of fungal ribosomal RNA that contains at least one of the sequences

-20-

specific to a particular fungus. Further aspects of the method of detecting fungi of the present invention are detailed in the examples below.

III. Obtaining a Biological Sample

5 In order to obtain a biological sample containing an organism, an infectious agent, or a biological component of a cell or organism to be detected according to the method of the present invention, an organism or tissue suspected of harboring such an infectious agent, organism, or biological component can be identified. The identification can be made in any way known to the art. Preferably, the organism suspected of carrying an infectious agent or other organism is a human, and the identification is made by a physician who observes symptoms indicative of the presence of such an
10 organism or agent in such a human. For example, a patient diagnosed as having AIDS who comes down with pneumonia and who does not respond to anti-bacterial agents is identified by a physician as possibly harboring the fungus *Pneumocystis carinii*.

Alternatively, a biological sample taken from a host with no overt signs of having a medical condition or harboring an organism or agent can be tested. For example, a food sample or tissue from
15 an AIDS patient without signs of a fungal growth can be tested for the presence of a fungus. In this embodiment, any biological sample can be tested, even though overt signs of the presence of a fungus are lacking in that sample. Appropriate action may thereby be taken if a fungus is in fact found in such a biological sample.

The biological sample to be tested can be obtained by any means known to the art. For
20 example, if an AIDS patient is suspected of suffering from interstitial plasma pneumonia caused by the fungus *Pneumocystis carinii*, a sputum sample can be taken from the lungs of that patient. The sputum can be obtained by having the patient cough up phlegm from the lungs and deposit it into a cup. Alternatively, a sputum sample can be obtained by scraping the bronchial passage with a sterile swab, or by any other means known to the art. Any other biological sample which could possibly carry a
25 fungus is likewise obtained in an appropriate fashion.

IV. Preparing the Biological Sample

The biological sample is next prepared so that the RNA and/or DNA present in the biological sample can be probed. When a fungus is being probed for, ribosomal RNA of any fungi present in the sample can be probed in accordance with the methods of the present invention. In order to probe the
30 ribosomal RNA of any fungal cells present, these cells should first be lysed. Lysis of fungal cells or of other cells containing RNA or DNA of interest can be accomplished by any of a number of methods known to the art, including those set out in *Molecular Cloning*.

In one embodiment, the cells are lysed before they come into contact with the solid support. This embodiment might be used, for example, when the solid support is one which is not designed to
35 hold a sample of lysed cells, such as a nitrocellulose filter. In this embodiment, the cells are contacted with the solid support after they have been lysed.

-21-

A variety of techniques can be used for cell lysis. When the ribosomal RNA of fungi is being probed, for example, techniques that separate the ribosomal RNA from the ribosomal proteins are preferred. Example 1 is provided to show one technique believed to be useful in obtaining ribosomal RNA. However, techniques for obtaining ribosomal RNA are well known. Techniques for obtaining DNA and other kinds of RNA are also well known to those of skill in the art. Thus, the technique of Example 1 is not necessarily a preferred method of obtaining ribosomal RNA-containing samples. Example 1, like all of the examples provided herein, are provided merely to illustrate certain aspects of the present invention. As such, they are not intended to limit the invention in any way.

EXAMPLE 1

Lysing Cells in a Biological Sample

Cells present in a biological sample can be lysed by treatment with a solution of 10 mM ethylenediaminetetraacetic acid (EDTA) (pH 8.0), 0.2 M NaCl, 0.5% of sodium dodecyl sulfate (SDS), 500 Unit/ml of RNase inhibitor, 10mM of Vanadyl Ribonucleosyl Complex and 200 μ g/l of Proteinase K (hereafter called Lysis Buffer). After lysis of the cells, the NaCl concentration of the resulting cell lysate in solution is adjusted to 0.5M.

V. Solid Support

In a preferred embodiment, the solid support is capable of containing a biological sample and is resistant to the reagents used to lyse cells in the biological sample. The sample can thus be lysed in the solid support. However, the use of such a solid support is not necessary if the sample can be obtained without lysis or is lysed in a separate container. An example of a solid support that is resistant to a large number of treatments is a microtiter well or a plate made from a resistant plastic material.

The solid support can also be any of a variety of other solid supports known to the art, such as a membrane filter, a bead, or any other solid, insoluble support to which polynucleotides can be attached. The solid support is preferably made of a material which can immobilize a polynucleotide probe. Immobilization can be through covalent bonds or through any of a variety of interactions that are known to those having skill in the art. Plastic materials containing carboxyl or amino groups on their surfaces, such as polystyrene, are preferred for the solid support of the present invention because polynucleotide probes can be immobilized on their surfaces, because they are inexpensive and easy to make, and because they are resistant to the reagents used to lyse the cells of the biological samples used in the present invention. For example, the Sumilon microtiter plate MS-3796F made by Sumitomo Bakelite, which has a carboxyl group on its surface, can be used in such a preferred embodiment. A plastic plate having an amino group on its surface, such as the Sumilon microtiter plate MS-3696, can also be used.

VI. Contacting the Sample and First Polynucleotide Probe

After the cells in the biological sample have been lysed, the RNA and/or DNA contained in such cells is substantially released into solution or otherwise made available to being probed. If the

biological sample was not lysed in the solid support, the cell lysate is next brought into contact with the solid support. Immobilized to the solid support is a first polynucleotide probe which preferably contains a sequence complementary to a specific sequence in the RNA and/or DNA of a particular organism, infectious agent, or biological component of a cell or organism in the biological sample. In an alternate embodiment, the first polynucleotide probe can also contain a sequence common to a plurality of organisms, infectious agents, or biological components when any of a group of such organisms, infectious agents, or biological components is sought to be identified. When the RNA and/or DNA present in the cell lysate contacts the solid support, therefore, it also comes into contact with the first polynucleotide probe.

Preferably, the first polynucleotide probe is an oligodeoxyribonucleotide (DNA) rather than an oligoribonucleotide (RNA), since DNA is more stable than RNA. The number of nucleotides in the polynucleotide probe is not restricted. However, if an oligodeoxynucleotide is used as the specific polynucleotide probe, a preferred length for the oligodeoxynucleotide is from 15 to 100 nucleotides. Lengths longer than 100 nucleotides are usable within the scope of the present invention. However, lengths of 100 nucleotides or less are preferable because many automated polynucleotide synthesizers have a limit of 100 nucleotides. Longer sequences can be obtained by ligating two sequences of less than 100 nucleotides.

In one embodiment, the first polynucleotide probe is an oligonucleotide complementary to one of the following sequences: SEQ ID NO:81, SEQ ID NO:104, SEQ ID NO:131 through SEQ ID NO:133, SEQ ID NO:154 through SEQ ID NO:156, SEQ ID NO:176, SEQ ID NO:199, SEQ ID NO:267, SEQ ID NO:290, SEQ ID NO:312, SEQ ID NO:335, SEQ ID NO:364 through SEQ ID NO:376, or SEQ ID NO:391 through SEQ ID NO:392. These sequences are specific to the ribosomal RNA of various pathogenic species of fungi, as listed in the sequence listing, and Tables V through XII.

Example 2 is provided to show one particular probe that is useful for determining the presence of *Pneumocystis carinii* in a biological sample.

EXAMPLE 2

Preparing the First Polynucleotide Probe

A first polynucleotide probe that is specific to a sequence of ribosomal RNA in *Pneumocystis carinii* is prepared. The probe is produced with a DNA synthesizer such as a DNA synthesizer made by Applied Biosystems of Menlo Park, CA. The probe is complementary to a polynucleotide having the following sequence where A, T, G, and C stand for adenine, thymine, guanine and cytosine, respectively: 5'-GCGCAACTGATCCTTCCC-3' (SEQ ID NO:81).

VII. Immobilizing the First Polynucleotide Probe

Various methods of immobilizing polynucleotides to a solid support are known to the art, including covalent binding, ionic binding, and the physical absorbance method. In certain embodiments of the present invention, the polynucleotides, such as the first polynucleotide probe, are immobilized to

microtiter wells which exhibit functional groups such as carboxyl residues, amine residues, or hydroxyl residues on the surfaces thereof. Thus, in one procedure for the immobilization of the first polynucleotide probe to a solid support, the solid support exhibits a functional group and the 5'-terminal end of the polynucleotide is covalently linked to the functional group. Any of a variety of methods for the covalent binding of polynucleotides to these functional groups can be used. Examples of preferred, well-known methods include the maleimide method and the carbodiimide method.

The maleimide method involves a reaction between a substance containing a maleimide group and another material containing a sulfhydryl residue (SH). The 5' end of the specific polynucleotide probe is immobilized on a solid support in this method by reacting the 5' end of the polynucleotide with a maleimide compound. A suitable maleimide compound is sulfosuccinimidyl-4-(N-maleimidomethyl) cyclohexane-1-carboxylate (sulfo-SMCC).

The SH residue is provided on the support by a reaction between a support having an amino group and succinimidyl-S-acetylthioacetate (SATA), followed by deacetylation using hydroxylamine (NH_2OH). Sulfo-SMCC and SATA are readily available from a variety of commercial sources, including the Pierce Company. The resulting SH group on the support is reacted with the maleimide group on the 5' end of the first polynucleotide probe when these groups are brought into contact under the appropriate conditions, thereby immobilizing the first polynucleotide probe to a solid support.

One problem we have experienced in the use of the maleimide method is that the SH group on the support can react not only with an amino group at the 5' end of the first polynucleotide probe, but also with primary amino groups on the purine bases, adenine and guanine. In order to assure that the polynucleotides are immobilized at their 5' ends, so that the sequences complementary to the ribosomal RNA sequences specific to a particular species of fungus are available for hybridization, the amino groups on the purine bases can be protected by pairing the specific polynucleotide probe to a complementary polynucleotide prior to immobilization. After immobilization, the complementary polynucleotide can be removed through denaturation, such as through heating, leaving the single-stranded probe immobilized to the solid support.

Another method of immobilizing a polynucleotide to a solid support is the carbodiimide method. This method involves a reaction between an amino group and a material containing a carboxyl residue using a carbodiimide compound. An example of a carbodiimide compound is 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (hereafter called EDC). This reaction can be enhanced with N-hydroxysulfosuccinimide (hereafter called Sulfo-NHS). Both EDC and Sulfo-NHS are available from well known commercial sources, including the Pierce Company.

In the practice of a preferred carbodiimide method for attaching polynucleotides to a solid support, a support having a carboxyl residue attached is used. Before contacting EDC with the support, the EDC is activated by reacting it with Sulfo-NHS. This activated EDC is then reacted with the solid support containing surface-bound carboxyl residues. The support, after being so treated, can be reacted

-24-

with strands of the first polynucleotide probe, which have an amino group at their 5'-terminal ends, thereby immobilizing the specific polynucleotide probe to the support.

In order to assure that the first polynucleotide probe is immobilized at its 5' end, the primary amino groups on the probe (the adeny, guanyl and cytosyl groups) can be protected by hybridizing the nucleotide to a complementary polynucleotide prior to immobilization. After immobilization, the complementary polynucleotide can then be removed through denaturation, such as through heating, leaving the single-stranded probe immobilized to the solid support. In order to further prevent the non-specific binding of activated amino or carboxyl residues on solid supports, the solid supports to which the specific polynucleotide probes are immobilized can be treated with a primary amine compound, preferably glycine.

Example 3 is provided as an indication to those of skill in the art of but a single method of immobilizing a probe to a solid support. Those of skill in the art will recognize that any of a variety of methods of so immobilizing the probe can be used, including those described above.

EXAMPLE 3

Immobilizing the First Polynucleotide Probe onto a Solid Support with the Carbodiimide Method

Both EDC and sulfo-NHS (Pierce, IL) are dissolved in DEPC-treated water at concentrations of 20 mM and 10mM, respectively. EDC/Sulfo-NHS solution is then prepared by mixing equal volumes of both EDC and sulfo-NHS. The specific nucleotide probe is dissolved in DEPC-treated water at a concentration of $1\mu\text{g}/\mu\text{l}$ and then mixed with the EDC/Sulfo-NHS solution in the ration 1:25 (Vol:Vol).

50 μl of this probe solution is added to each well of a microtiter plate (MS-3796F, Sumitomo Bakelite, JAPAN), which is known to have carboxyl groups on the surface of the plate. After incubation at room temperature overnight, the reaction solution is removed by aspiration.

VIII. Hybridizing the First Polynucleotide Probe

In a preferred embodiment, the cell lysate or other sample containing polynucleotides is next hybridized to a first, specific polynucleotide probe immobilized to the solid support. The first polynucleotide probe can, however, also be a common probe, depending on the purpose of a particular assay performed according to the present invention. Hybridization can be accomplished by incubating the cell lysate and the first polynucleotide probe at a temperature dependent on a variety of factors, as is well known to those with ordinary skill in the art. These factors include the length of complementary nucleotide sequences, the ratio of guanine and cytosine bases to the entire base content in the complementary nucleotide sequences (the GC content), the NaCl concentration in the buffer solution, the number of bases which mismatch in the complementary nucleotide sequence, and the type of nucleotide. In a preferred form of this invention, the following equation can be used to calculate the preferred incubation temperature (T_{inc}):

$$T_{\text{inc}} = 16.6 \times \log (M) + 0.41 (\text{GC}) + 81.5 - 675/n - 15 (^\circ\text{C}).$$

-25-

In the equation shown above, M is the NaCl concentration in solution, GC represents the GC content (the percentage of guanine and cytosine residues in the sequence), and n represents the length of the nucleotide sequences (the number of hybridizing nucleotides). The incubation temperature can also be determined according to methods described in *Molecular Cloning*.

5 The time for incubation is preferably from 1 hour to overnight, and the sample is preferably gently rocked during incubation. Incubation is preferably performed in an appropriate buffer solution. The same buffer used to hybridize RNA and DNA in the Northern Blot or the Dot Blot methods, as described in the Maniatis treatise, can be used. The buffer is preferably prepared in a way so as not to contaminate it with RNase. If any RNase contamination is present, the activation of RNase should
10 be controlled so as to be as low as possible.

Example 4 illustrates one method of hybridizing ribosomal RNA in a sample to an immobilized probe.

EXAMPLE 4

Hybridizing the First Polynucleotide Probe to Ribosomal RNA

15 RNase is removed from a microtiter well to which a first polynucleotide probe having a sequence complementary to SEQ ID NO:81 has been bound by adding 250 μ l of Lysis Buffer containing 0.5 M NaCl and incubating the well at 45°C for one hour. The buffer is removed from individual wells by aspiration, and 50 μ l of Lysis Buffer containing the biological sample is added to each well. These solutions are incubated at 39°C for one hour ($T_m = 54^\circ$) to allow hybridization and then slowly cooled
20 over the course of 20-30 minutes.

In the methods of the present invention that involve probing samples containing RNA, or in any procedure in which it is desired to prevent the degradation of RNA, it is advantageous to inhibit the activity of any ribonucleases (RNases) which may be present. One RNase inhibitor which can be used is Vanadyl Ribonucleoside Complex (VRC) (Bethesda Research Laboratories, Gaithersburg, MD).
25 VRC has been reported to be useful during cell fractionation and in the preparation of RNA, and has been shown not to interfere with the phenol extraction or ethanol precipitation of RNA. In addition, VRC does not affect other cytoplasmic components of cells. Therefore, VRC is an ideal inhibitor of RNase in many experimental procedures.

However, prior art procedures for inhibiting RNases with VRC taught that VRC should not be
30 used in buffer systems containing EDTA or SDS, which are commonly used in the field of molecular biology. The reason for this prohibition was that it was believed that the Complex would dissociate in the presence of these buffers, leading to a loss of RNase inhibiting activity. In fact, the BRL insert accompanying the VRC product recommends that a five- to ten-fold molar excess of EDTA be added to an RNA solution containing VRC in order "destroy" the VRC prior to ethanol extraction of the RNA
35 from the solution. The apparent inability to use VRC together with common buffers that include EDTA and SDS thus presented a major impediment to the exploitation of VRC as an RNase inhibitor.

-26-

We have discovered, however, that VRC is in fact an effective RNase inhibitor even in the presence of SDS and/or EDTA. Thus, VRC can be used in assays which use buffers including SDS or EDTA, where heretofore it was believed that VRC would not be effective in such systems. VRC is an effective RNase inhibitor at the concentrations of EDTA and SDS that are normally used when manipulating RNA or when performing a variety of other molecular biology techniques. For example, we have found that VRC effectively inhibits RNase in a buffer solution comprising approximately 1 mM EDTA. We also found VRC to be effective in .5% SDS solutions, and is believed to be effective in solutions ranging up to 2% SDS, more preferably up to 1% SDS. VRC can, of course, also be used in other solutions including EDTA and SDS.

We have found that VRC is a particularly potent inhibitor of RNases when used in combination with Proteinase K. Proteinase K is also available from BRL. As shown in the gel in Figure 1a, mRNA prepared from U937 cells (human macrophage cell line) was protected from RNase degradation by a combination of VRC and Proteinase K. Lane 1 of the gel shows the mRNA from a cell preparation which included VRC, Proteinase K, and RNasin, while lane 2 represents the mRNA from a cell preparation that included VRC and Proteinase K. RNasin is available from Promega of Madison, WI. The distinct band 10 in lanes 1 and 2 matches the band seen in lane 7, which contains pure clonal cDNA from an RNase-free preparation of U937 cells, thus showing that the mRNA in the preparations of lanes 1 and 2 did not experience substantial mRNA degradation.

A comparison of lanes 1 and 2 with lanes 3-6 shows that VRC in combination with Proteinase K inhibits RNase activity in the above-mentioned mRNA preparation from U937 cells to a far greater extent than either Protein K alone (lane 4), Proteinase K in combination with RNasin (lane 3), or a commercial RNase-inhibiting preparation sold under the name "FastTrack" (available from In Vitrogen of San Diego, CA) (lane 5). None of lanes 3-5 exhibit the distinct band 10 representing undegraded mRNA. On the contrary, lane 4 (Proteinase K alone) and lane 5 (FastTrack) have the same smear of degraded mRNA as band 20 in lane 6 (no inhibitors). The gel shown in Figure 1a also shows the effectiveness of VRC in a buffer solution of 1 mM EDTA and .5% SDS, which is the buffer used in the U937 cell preparations tested, since lane 2 (VRC and Proteinase K) shows less mRNA degradation than lane 4 (Proteinase K alone) or lane 3 (Proteinase K and RNasin).

In order to eliminate RNase activity from water used in the methods of this invention, the water is preferably treated with diethylpyrocarbonate (DEPC). The preferred DEPC treatment involves addition of 0.1% DEPC to the water, followed by storage overnight at 37°C and sterilization in an autoclave. The DEPC is deactivated by such autoclaving so that it does not interfere with the enzymatic processes of the methods of the present invention. Alternatively, if the water is sterilized in some other manner, the DEPC in the water can be deactivated by other means known to the art.

IX. Washing the Solid Support

Following hybridization, the non-hybridized portions of the biological sample are preferably separated from the solid support, so that substantially all of the biological sample not annealed to the first polynucleotide probe is removed from the solid support. If the solid support is a microtiter well, for example, and the first polynucleotide probe is immobilized to the walls or bottom of the well, the non-hybridized cell lysate can be removed by pouring the lysate out of the well or by aspirating the cell lysate. The well itself can then be "washed" or rinsed with a washing solution such as the Lysis Buffer by applying the washing solution to the walls of the well and then removing the washing solution through aspiration. Any washing solution known to the art can be used, provided that the salt content and other parameters of the solution are controlled so that the washing solution does not remove any polynucleotide hybridized to the polynucleotide probe.

X. Contacting and Hybridizing the Second Polynucleotide Probe

When the solid support has been washed, a second polynucleotide probe is then contacted with and hybridized to the polynucleotide strand of RNA or DNA which is hybridized to the immobilized first polynucleotide probe, if such a polynucleotide was present in the cell lysate. The contacting and hybridization steps are performed as with the first polynucleotide probe, above, or by any other methods known to the art.

In a preferred embodiment, the second polynucleotide probe contains a polynucleotide sequence complementary to a sequence which is common to the RNA and/or DNA of a group of organisms, infectious agents, or biological components. For example, if a particular organism is being probed for and that organism is a fungus, the second probe can contain a sequence common to a plurality of species of fungus including the species being probed for. If an infectious agent such as a virus is being assayed for, the second probe can contain a sequence common to a number of related viruses. In this way, the same second polynucleotide probe can be hybridized to the RNA and/or DNA of any of a group of organisms, infectious agents, or biological components. In an even more preferred embodiment, the second, common probe has the same or lower T_m as the first, specific probe so that the hybridization of the second probe can be performed under the same conditions as the conditions used to hybridize the first probe.

This embodiment of the present invention is preferred when a plurality of specific probes are used to assay for the presence of a plurality of organisms, infectious agents, or biological components, because then the same second probe can be annealed to the RNA and/or DNA of the plurality of such organisms, agents, or biological components. The common polynucleotide probe of this embodiment can also advantageously be included in a kit in which a plurality of specific probes are used to detect the presence of a plurality of organisms or agents.

The second polynucleotide probe can alternatively comprise a second sequence specific to the polynucleotide of the particular organism, infectious agent, or biological component sought to be

-28-

identified in the sample. This second sequence is complementary to a different specific sequence of the RNA and/or DNA of the particular organism, infectious agent, or biological component than the sequence to which the first polynucleotide probe is complementary.

XI. Label

5 The second polynucleotide probe is preferably labeled in order to easily detect its presence and facilitate the detection of the RNA and/or DNA hybridized to the first probe. A variety of chemical substances are available which can label a polynucleotide probe when attached to that probe. For example, a variety of radionuclides can be used, such as the radioisotopes ^{32}P , ^{35}S , ^3H , and ^{125}I . Enzymes or enzyme substrates can also be attached to the second polynucleotide probe in order to label
10 it. Suitable enzymes include alkaline phosphatase, luciferase, and peroxidase.

Labels which provide a colorimetric indication or a radionuclide are preferred. Other labels which can be used include chemical compounds such as biotin, avidin, streptavidin, and digoxigenin. Colorimetric labels, such as fluorescein, are especially preferred because they avoid the health hazards and disposal problems associated with the use of radioactive materials. In a preferred embodiment,
15 biotin is attached to the nucleotide probe, followed by an avidin, such as streptavidin, which is itself conjugated to an enzyme such as alkaline phosphatase. The presence of the enzyme can be detected by various substrates, such as ATTOPHOS, which provide a fluorescent marker that can be detected through fluorimetry.

Nucleic acids can also be "labeled" by staining them with a nucleic acid stain. Thus, where a
20 relatively large amount of nucleic acids are present, ethidium bromide (EtBr) can be used to identify the presence of such nucleic acids. However, more sensitive stains are more preferable. Sensitive stains include various cyanine nucleic acid stains, such as POPO, BOBO, YOYO and TOTO, available from Molecular Probes (California). These stains are described, e.g., in *Science*, **257**:885 (1992). Particularly preferred stains for use in the context of the present invention include the shorter wavelength forms,
25 TOTO-1 and YOYO-1, still more preferably YOYO-1. As little as four picograms of stained DNA can be detected by visible fluorescence upon stimulation with a transilluminator or hand-held UV lamp. Thus, these stains provide a particularly easy and sensitive method of identifying the presence of nucleic acids.

We tested the ability of YOYO-1 to stain oligonucleotides immobilized to wells as discussed
30 herein. We first immobilized various known amounts of oligonucleotides to wells and washed to remove non-immobilized oligonucleotides. We then added 1:1000 dilution of YOYO-1 in water. We then used a fluorimeter directly without washing. The relation between pmoles of oligonucleotide is shown in Figure 1b with circles. We also incubated the TOTO-1 stained immobilized oligonucleotides for ten minutes, washed and added water, followed by use of the fluorimeter. The washed results are shown
35 in dark circles in Figure 1b, while results without washing are shown in open circles. It can be seen that

washing does not significantly effect the amount of staining and that the amount of staining is related to the amount of oligonucleotide.

We also tested the time course of YOYO-1 staining over the course of one hour. We included a control in which no oligonucleotide was immobilized to the plate. The oligonucleotide-immobilized plate is shown in open circles in Figure 1c and the control in dark circles in that figure. It can be seen from Figure 1c that after an initial spike, relatively constant staining is found.

We further tested the dose response of constant amounts of oligonucleotides with oligonucleotides immobilized in wells (open circles in Figure 1d) and control wells with no oligonucleotide (closed circles in Figure 1d). The difference between the immobilized and non-immobilized is shown in Figure 1d as triangles. It can be seen that a sharp increase in fluorescence occurs between 10^{-4} and 10^{-3} dilution.

We also tested the reproducibility of staining of both oligonucleotide-immobilized wells and non-immobilized wells. We repeated the experiment five times and graphically depicted the data for both oligonucleotide immobilized (+) and non-immobilized (-) wells in Figure 1e. It can be seen that the data was substantially similar for each experiment.

Thus, the data depicted in Figures 1b-1e shows that use of YOYO-1 as a staining agent provides a reliable indication of the amount of oligonucleotide present.

Other labels known to the art can also be used. For example, a specific binding moiety which binds to a binding partner for that moiety can be used. For example, a binding moiety such as an antibody can be labeled with a marker and then bound to a binding partner, such as an antigen, on the second polynucleotide probe. An agonist and its receptor can also be used, as is known to the art. A receptor such as the norepinephrine receptor attached to the second polynucleotide probe can be detected by the addition of a labeled agonist for that receptor, in this case norepinephrine.

Example 5 shows one method of using a labeled common polynucleotide probe to identify the presence of rRNA.

EXAMPLE 5

Preparing, Labeling, and Hybridizing the Second Nucleotide Probe

The second nucleotide probe is an oligodeoxynucleotide prepared as in Example 1 comprising a sequence complementary to the following sequence: 5'-GAGGGAGCCTGAGAAACG-3' (SEQ ID NO:1). This sequence is complementary to the ribosomal RNA of a number of fungal species, including the ribosomal RNA to which the specific polynucleotide probe of Example 1 is complementary. This second nucleotide probe is labeled with fluorescein at either the 3' or 5' end of the probe. The 3' end is labeled using terminal transferase and FITC-dUTP. alternatively, the 5' end is labeled by chemical reaction with FITC. Fifty μ l of Lysis Buffer including 0.5 M NaCl and 1 μ l of a solution containing the

5 After the second polynucleotide probe has been hybridized to any RNA and/or DNA of the organism, infectious agent, or biological component present in the biological sample, substantially all of the solution containing the second probe is removed from the solid support by aspiration or in any other appropriate way, and the solid support is again washed to remove any unhybridized second polynucleotide. The presence of the particular organism, infectious agent, or biological component sought to be detected is then finally determined by detecting the presence of the second polynucleotide probe immobilized to the solid support via the RNA and/or DNA strand of the organism, infectious agent, or biological component being probed for, which is itself hybridized to the first polynucleotide probe. If the second polynucleotide probe is detected, this indicates that the biological sample contained RNA and/or DNA being tested for, and hence that the biological sample harbors the organism, infectious agent, or biological component sought to be identified in the sample.

20 A positive control experiment is also preferably run when performing the methods of the present invention. A positive control is run by performing the present method with the same materials and using the same steps, with the exception that the first polynucleotide probe immobilized to the solid support contains a sequence common to the RNA and/or DNA of a group of organisms, infectious agents, or biological components of which the organism, infectious agent, or biological component being probed for is a member. For example, if a species of fungus is being probed for, the first polynucleotide probe immobilized to the solid support of the positive control can contain a sequence common to the genus of fungi of which the species being probed for is a member. Thus, if any of the species of this genus of fungus are present in the biological sample, the solid support comprising the positive control will indicate the presence of a fungus in the sample.

30

35

-31-

was removed from the wells, 10 mM glycine was added into each well and incubated at 37°C for 2 hours. Each of the wells was washed with 200 μ l of water six times and then stored at 4°C until use.

Approximately 50 μ l of a solution containing approximately 1 mg/ml of each of jun-B, c-jun, and jun-D mouse jun oncogene cDNA (biotinylated) in reaction buffer (10 mM Tris, pH 8.0, 1 mM EDTA, and 0.5 M NaCl) was placed in each of three wells in a microtiter plate. After incubating the wells at 37°C for 2 hours, the wells were washed with the reaction buffer 3 times. Alkaline phosphatase-conjugated streptavidin (Clontech, Palo Alto, CA) was diluted 1:1000 with the reaction buffer, and 50 μ l of the diluted solution was added into each well and incubated at room temperature for an additional 30 minutes. After the wells were washed with the reaction buffer 3 more times, 50 μ l of ATTOPHOS (JBL, San Luis Obispo, CA) was added into each well and incubated at room temperature for 10 minutes. The fluorescence of the individual wells was measured by CytoFluor 2300 (Millipore, Bedford, MA) at a filter setting of 485 nm for excitation and 590 nm for emission. As shown in Figure 7A-7C, each subtype-specific oligonucleotide hybridized only to the corresponding mouse jun oncogene subtype. Also, no appreciable cross-hybridization between subtypes occurred. Therefore, each of these probes is specific to only one subtype of jun oncogene.

XIII. Quantifying the Amount of an Organism, Infectious

Agent, or Biological Component Contained in the Sample

The amount of RNA and/or DNA in the biological sample from a particular organism or agent, or the amount of such RNA and/or DNA indicative of the presence of a particular biological component can be quantitated by measuring the amount of second polynucleotide probe bound to the solid support after being applied thereto in accordance with the method of the present invention. When testing for an organism or infectious agent, the amount of RNA and/or DNA in the sample can give a rough measurement of the number of organisms or infectious agents contained in the sample, and thus a rough estimate of the extent of an infection if the sample is from a diseased organism.

To quantify the amount of second polynucleotide probe attached to the solid support, a physical or chemical quantity or activity of the label on the second polynucleotide is measured. A number of techniques for measuring the label on a polynucleotide probe known to the art can be used, the technique used depending on the kind of label. Such techniques include measuring the optical density of the buffer solution, the emitted-light intensity of the buffer solution, or the amount of radiation given off by the immobilized second polynucleotide probe. The label itself can provide this indication or can require other compounds which bind thereto or which catalyze the label. Other mechanisms for detecting label include the use of compounds that can chemically react with the label, the detection of a colored label, the detection of light emission, the detection of radiation, or the catalytic ability of the label.

In one measurement technique, the label on the second polynucleotide is biotin. The presence of this label can be detected by reacting it with the enzymes peroxidase or alkaline phosphatase. These

-32-

enzymes can be specifically directed to biotin by conjugation with avidin or streptavidin. The presence of the enzymes is then detected by the addition of an appropriate substrate to provide a detectable color-developing or light-emitting reaction. Alkaline phosphatase-labeled streptavidin can be readily obtained from the commercial market. One substrate for alkaline phosphatase is adamantyl-1,2-dioxetane phosphate (AMPPD). Upon reaction with the alkaline phosphatase, AMPPD will emit light at a wavelength of 447 nm. This light can be detected in accordance with techniques known in the art. A more preferred light emitting substrate for alkaline phosphatase is ATTOPHOS.

In the reaction between alkaline phosphatase and AMPPD, an enhancer such as 5-N-tetradecanoyl-amino-fluorescein can be added. 5-N-tetradecanoyl-amino-fluorescein has the ability to convert light of 477 nm wavelength to light of 530 nm wavelength, which is more readily detectable.

Other labels include an antigen, such as digoxigenin or an antibody. An antigen can be detected by its ability to bind to an antibody directed thereto. Such antibodies, or an antibody directly used to label the second polynucleotide probe, can be detected by their ability to bind a protein. The antibody itself can be labeled directly with a radionuclide, such as ^{125}I , or can be labeled by binding thereto a protein labeled with the radionuclide. The radionuclide can then be detected in accordance with techniques well known in the art, such as using X-ray film or a radiation counter.

Well-known techniques for the detection of a label include detecting a label in a color-developing reaction with a spectrophotometer or fluorimeter. For example, fluorescein, which gives off a fluorescent pigment, can be used as a label, and the intensity of the pigment can be used to gauge the amount of label hybridized to target polynucleotides. The use of such color-developing reactions are preferred to using radioactive labels in the present methods because the problems associated with using radioactive nuclides are thereby avoided.

Other such techniques include the detection of a light emitting reaction using X-ray film or an instant camera. The emission reactions are recorded by X-ray film or instant camera film in the dark room. The X-ray film which is exposed by emission reactions is recorded as a blot, so that the shading of the blot can be measured by a densitometer. If one uses an instant camera such as a Polaroid, the picture is read by a scanner to decide the location of the blot on the computer, and the shading of the blot is determined using graphic analysis software.

An example of one embodiment of the present invention is illustrated in Figure 2. As shown in that Figure, following the collection and lysis of a sample of biological materials from a human patient, the ribosomal RNA of a species of fungus sought to be detected is hybridized. Following hybridization, a specific sequence 24 of the strand of ribosomal RNA 22 to be detected is annealed to a complementary sequence 28 on a first polynucleotide probe 26. The first probe 26 is immobilized on a solid support 20.

After removing substantially all of the unhybridized portions of the sample, a second polynucleotide probe 34 is hybridized to the strand of ribosomal RNA 22 at a different sequence 30, which is preferably one that is common to a plurality of fungal species. The second probe 34 contains

-33-

a sequence 32 that is complementary to the sequence 30. The second probe, in this diagram, also comprises a label 36 attached to the probe which facilitates the detection of the complex formed by the first probe 26, the second probe 34, the strand of ribosomal RNA 22, and the solid support 20.

Example 6 shows one method of measuring the amount of rRNA of a particular fungal species using a labeled common polynucleotide probe.

EXAMPLE 6

Measurement of Chemical Activities of the Labeled Second Nucleotide Probe

Following the hybridization described in Example 5, the hybridization solution (Lysis Buffer containing the labeled second polynucleotide probe) is removed by aspiration and the microtiter plate is washed once with 250 μ l of fresh Lysis Buffer. A blocking buffer consisting of 0.05% (w/v) Tween 20, 500 mM NaCl, and 100 mM Tris-HCl, pH 7.5 is added into each well and incubated at room temperature for five minutes to reduced nonspecific binding. These solutions are then removed by aspiration.

The fluorescein on the second polynucleotide probe which is bound to the ribosomal RNA of the species of fungus being tested for, if present, is then visually detected to determine whether a fungus of that species was present in the biological sample. The approximate quantity of fungus present in the sample is then measured by determining the amount of fluorescein bound to the microtiter plate with a spectrophotometer or in a fluorimeter.

XIV. Using PCR to Detect Small Quantities of DNA or RNA in a Sample

We have also discovered an alternative procedure which is especially useful in the detection of minute quantities of an organism, infectious agent, or biological component in a sample. This alternative procedure makes use of the polymerase chain reaction (PCR) procedure to create multiple copies of a polynucleotide strand which is complementary and/or homologous to a biological component or a strand of RNA and/or DNA which belongs to an organism or infectious agent in the sample.

In an example of this embodiment of the present invention, shown in Figure 3, a biological sample is first obtained as described previously. The sample is lysed so that the RNA and/or DNA of any organisms, infectious agents, or cells carrying biological components which are present in the sample can be probed. The lysed sample is then contacted with a first polynucleotide primer 42. This primer 42 is complementary to a sequence contained in an analyte polynucleotide of an organism, infectious agent, or biological component to be detected in the sample. The primer 42 is then contacted with such an analyte polynucleotide 40 with which it is complementary and hybridized to that polynucleotide. The hybridizing of the primer 42 can be accomplished in the same manner as was previously described in relation to the hybridizing or annealing of polynucleotide probes.

The sequence complementary to the primer 42 can comprise, for example, a sequence specific to the DNA or RNA of an organism, such as the ribosomal RNA of a species of fungus. Such a sequence can also comprise a sequence specific to another infectious agent or to a biological component,

-34-

such as an oncogene sequence. However, in another embodiment the primer 42 is complementary to a sequence located 3' of such a specific sequence, so that the primer 42 is positioned 5' of the specific sequence when it is annealed to a strand containing the specific sequence. In this way the sequence complementary to the specific sequence is incorporated when the primer 42 is extended.

5 In yet another embodiment, the primer 42 is complementary to a sequence common to a group of organisms, infectious agents, or biological components, such as a sequence common to a plurality of fungal species. In another example, the primer 42 is common to a number of related biological components, such as a plurality of subtypes of jun oncogenes. Thus, in this embodiment, the first polynucleotide primer 42 can be referred to as a common PCR primer.

10 After the common PCR primer 42 has been annealed to the analyte polynucleotide 40 present in the sample, this primer is extended using the four nucleotide triphosphates and a polymerase enzyme, thereby producing a double-stranded polynucleotide including a complementary nucleotide strand 41 comprising cDNA and having a sequence complementary to the analyte polynucleotide in the sample. If the nucleotide sequence being probed for is contained in RNA, the polymerase is preferably a reverse transcriptase and the nucleotide triphosphates are deoxynucleotide triphosphates (dNTP's) so that cDNA
15 complementary to the rRNA is produced. Further rounds of amplification can be accomplished by reannealing additional primer 42 to the analyte polynucleotide 40 and then extending that primer. As will be clear to those of skill in the art, the first polynucleotide primer 42 can also have a sequence specific to a particular organism, infectious agent, or biological component.

20 The first primer 42 is preferably extended using deoxynucleotides to form a strand of cDNA 41 that is complementary to the strand of RNA or DNA 40 in the sample. Using cDNA 41, amplification can also be accomplished by hybridizing to the strand of cDNA 41 a second polynucleotide primer 44 that is complementary to the newly synthesized complementary nucleotide strand 41. Thus, this second primer 44 is homologous to a portion of the analyte polynucleotide 40. This second primer
25 44 can preferably be common to the RNA or DNA of a plurality of organisms, infectious agents, or biological components and comprise a sequence different from that to which the first polynucleotide primer is complementary. Thus, in a preferred embodiment, the second primer 44 is a second common PCR primer, and the RNA and/or DNA of a plurality of organisms, infectious agents, or biological components can be amplified with it. Amplification is completed by extending the second primer 44 to
30 produce a strand of cDNA 43 that is homologous to at least a portion of the analyte polynucleotide 40 in a sample. This amplification step is also preferably repeated a plurality of times with further second primers 44.

These later rounds of amplification preferably use the four dNTP's in combination with a DNA polymerase so that during amplification, double-stranded DNA is produced which contains one strand
35 of cDNA 43 homologous to the analyte polynucleotide 40 present in the biological sample and one strand of cDNA 41 complementary to such homologous cDNA 43. Preferably, approximately 20 to 40 cycles of DNA synthesis are performed in order to produce an adequate amount of cDNA homologous

-35-

to the RNA and/or DNA present in the sample for later detection. A DNA polymerase is used in the synthesis of such cDNA. This polymerase preferably has significant polymerase activity at temperatures above 50°C, such as Taq DNA polymerase.

Following these rounds of amplification, another round of amplification is preferably performed.

5 In this second round of PCR, a third polynucleotide primer is used and if desired a fourth primer can also be used. One or both of these primers preferably contain a sequence that is homologous or complementary to a sequence specific to a particular organism, infectious agent, or biological component, in which case such primers can be called specific primers. However, when it is desired to identify or quantitate the presence of a group of organisms, infectious agents, or biological components,

10 one or both of the third and fourth primers can contain a sequence complementary or homologous to a sequence common to a plurality of organisms, infectious agents, or biological components. When PCR is performed using third and/or fourth primers containing sequences specific to an organism, infectious agent, or biological component, amplification will occur in significant amounts only if RNA and/or DNA of the particular organism, infectious agent, or biological component being tested for is present in the

15 biological sample and the initial amplification has produced cDNA corresponding to such RNA and/or DNA.

The amplification of sequences contained in the RNA and/or DNA of an organism, infectious agent, or biological component can be detected in several ways, as will be appreciated by those having ordinary skill in the art. For example, a common or specific primer can be labeled and the presence

20 of the label in an extended polynucleotide can be detected. As shown in Figure 3, for example, a label 45 can be attached to the primer 44, which is then extended to produce the homologous polynucleotide 43.

When it is desired to detect a specific organism, infectious agent, or biological component, the homologous polynucleotide strand 43 can be contacted with and hybridized to a sequence 48 on a

25 specific polynucleotide probe 46 which is complementary to a sequence 49 that is specific to a particular organism, infectious agent, or biological component and which is immobilized on a solid support 47, as in other embodiments of the present invention. Following this, the unhybridized portions of the sample are preferably washed from the solid support 47, and the labeled homologous polynucleotide 43 immobilized on the solid support 47 is detected. Alternatively, if the homologous polynucleotide 43 is

30 not labeled, following the washing of the solid support 47 a second polynucleotide probe (not shown in Figure 3) carrying a label can be hybridized to the homologous polynucleotide 43. This can also be followed by washing. After washing unhybridized second probe from the solid support, the labeled second probe can be detected using any of a variety of techniques. The homologous polynucleotide 43 will only be detected, of course, if it is hybridized to the immobilized polynucleotide probe 46.

35 As will be appreciated by those of skill in the art, the complementary cDNA strand 41 rather than the homologous polynucleotide 43 can also be annealed to a polynucleotide probe 46 immobilized on the solid support. In this method, the primer 42 is labeled rather than the primer 44. The strand

-36-

41 can then be detected directly, as described above, or can be detected by hybridizing a labeled polynucleotide to the strand 41.

Example 7 illustrates one method of identifying small quantities of the fungus *Pneumocystis carinii* in a sample.

EXAMPLE 7

Amplifying Ribosomal RNA Present in Minute Quantities with PCR

5 A sputum sample from a patient suspected of having pneumonia caused by the fungus *Pneumocystis carinii* is first lysed with Lysis Buffer and brought to a total volume of sample and lysis
10 buffer of 50 μ l. This mixture is then added to a well of a microtiter plate to which a common probe (SEQ ID NO: 1) has been immobilized, thereby contacting the probe with the mixture. The common probe is a polydeoxynucleotide complementary to a sequence common to the ribosomal RNA of a number of fungal species, including *P. carinii*. The sequence on the ribosomal RNA of *Pneumocystis carinii* to which this probe is complementary is located 3' of a specific sequence (SEQ ID NO: 81) of
15 such ribosomal RNA.

The common probe can then be hybridized to the fungal ribosomal RNA in the sample by incubating the mixture in the well at 39°C for one hour and then cooling the mixture over 20-30 minutes. Following the annealing of the common probe to the ribosomal RNA, the probe is extended with a reverse transcriptase to produce a cDNA strand having a sequence complementary to the
20 ribosomal RNA. The complementary strand is then melted off of the ribosomal RNA by heating the mixture to 94°C for 1 to 2 minutes. This process is repeated several times in order to amplify the number of complementary strands.

Following the denaturation of the complementary strand from the ribosomal RNA, a second primer homologous to a sequence of the ribosomal RNA that is specific to *Pneumocystis carinii* is added
25 to the mixture. The four dNTP's and a DNA polymerase capable of polymerase activity above 50°C is added to the mixture. The mixture is then heated to a temperature below the T_m of the second primer but high enough to assure specific binding of the primer, in this case approximately 50°C. After allowing enough time for the second primer to be extended, about 1-2 minutes, the mixture is heated to 94°C for 1 to 2 minutes to melt the newly synthesized strand homologous to the ribosomal RNA of
30 the sample from the complementary strand. This process is repeated from between 20-40 times, with the addition of primer as necessary.

Following this round of amplification, two labeled, specific primers are added to the mixture. One is complementary to a specific sequence on complementary strand, while the other is
35 complementary to a sequence on the homologous strand. After annealing such primers to the complementary and homologous strands, these primers are extended with a polymerase in the same

-37-

fashion as the second primer, after which such strands are melted off by raising the temperature of the mixture to 94°C for 1 to 2 minutes. This process is also repeated between 20-40 times.

After the final round of amplification, the mixture is heated to melt off any newly synthesized strands from the templates from which they were produced, and the mixture is cooled to 37°C and incubated at that temperature for an hour in order to allow the strands present in the mixture to anneal to a specific probe immobilized on the microtiter well. This probe is complementary to a specific sequence in the ribosomal RNA of *Pneumocystis carinii* and anneals to the ribosomal RNA of that fungus in addition to the synthesized sequences homologous to such ribosomal RNA.

Once the polynucleotide strands in the mixture have been allowed to hybridize to the specific probe, the non-hybridized portions of the mixture are removed by aspiration. The walls of the microtiter plate are then washed with Lysis Buffer to remove any non-specifically bound nucleotide strands from the well. The label attached to the specific primer is next detected in order to detect the presence of *Pneumocystis carinii* in the biological sample. If the label is detected, this indicates that the biological sample contained this fungus.

XV. Primers for Use in PCR

We have identified several probes and PCR primers for use in the present invention. In particular, sequences have been identified for detecting jun oncogenes and G protein sequences in humans and other animal species, as well as sequences for detecting substance P and Beta-receptor sequences. The design and use of such primers is described below.

A. *Primers for Detecting Jun Oncogenes*

It is well known that certain oncogenes, such as the jun oncogenes, are most rapidly expressed when cells are stimulated to proliferate. Therefore, the detection or quantification of the expression of jun oncogenes is a good marker for cellular mitogenic activity. Jun oncogenes were first reported by Maki, et al. (*Proc. Natl. Acad. Sci. USA*, **84**:2848-2852 (1987)), and are currently known to exist in at least three different forms or subtypes, jun-B, c-jun, and jun-D. It is still unclear how these three jun subtypes are involved in cell growth, however, so that one has to analyze all three genes during cell growth in order to detect mitogenic activity. In addition, although it has been discovered that mice also carry three different jun oncogene subtypes, the nucleotide sequences of these three subtypes differ from the subtypes present in human cells (Ryder K., Nathans D., *Proc. Natl. Acad. Sci. USA* **85**:8464-8467, 1988; Ryder K. et al., *Proc. Natl. Acad. Sci., USA* **85**:1487-1491, 1988; Hattori K. et al., *Proc. Natl. Acad. Sci. USA* **85**:9148-9152, 1988; Schuette J. et al., *Cell*, **59**:987-997, 1989).

In order to design sense and anti-sense PCR primers for detecting jun oncogenes in both humans and mice, sequences were designed with the following considerations in mind:

- (a) The nucleotide sequences should be common to the three subtypes of jun oncogenes (jun-B, c-jun and jun-D), with a maximum of 4 base mismatches.

-38-

- (b) The nucleotide sequences should be common to both humans and mice, with a maximum of 4 base mismatches.
- (c) At least 5 bases at the 3' end of the primers should be 100% identical to sequences in all three jun subtypes in both humans and mice.
- 5 (d) The length of the nucleotide sequences of both the sense primers and the anti-sense primers should be between about 17 and 50 bases.
- (e) The difference in T_m between the sense primers and anti-sense primers and their corresponding sequences should be within 2°C.
- 10 (f) There is no complementary structure more than 4 bases long in either the sense or the anti-sense primer.
- (g) There should be no complementary structure more than 4 bases long between the sense and the anti-sense primer.
- (h) Both the sense and the anti-sense primers should span part of the coding sequence of their complementary polynucleotides.
- 15 (i) The length of polynucleotide to be amplified should be greater than 200 bases.
- (j) The nucleotide sequence homology of amplified genes among the three subtypes of jun oncogenes should not be greater than 80%.

DNA fragments (both sense and anti-sense primers) which satisfied the above conditions were investigated and several candidate oligonucleotides were synthesized. One set of primers which fit these parameters particularly well is the sense primer 5'-CCCTGAAGGAGGAGCCGCAGAC-3' (SEQ ID NO:733) and the anti-sense primer 5'-CGTGGGTCAAGACTCTGCTTGAGCTG-3' (SEQ ID NO:734). The homology between the sense primer SEQ ID NO:733 and several jun gene subtypes is shown in Table 2 below.

Table 2

Sense primer 5'-CCCTGAAGGAGGAGCCGCAGAC-3'		
30	Mouse jun-B --T-T--A-----	(SEQ ID NO:735)
	c-jun -----A-----	(SEQ ID NO:736)
35	Human jun-B --T-C-----A-----	(SEQ ID NO:737)
	c-jun -----T-----	(SEQ ID NO:738)

:- indicates identical base to the sense primer.

40 As an alternative to the sense primer SEQ ID No:733 shown in Table 2, 5'-XCCCTGAAGGAGGAGCCGCAGAC-3' (SEQ ID No:739) can also be used as a primer for detecting jun oncogenes. In this sequence, X represents a primary amine residue, a nucleotide sequence

-39-

recognized by a restriction endonuclease, or an RNA promoter sequence. The attachment of a restriction site to the 5' end of the PCR primers is useful for the cloning of amplified genes, while the attachment of RNA promoter sequences at such 5' ends is useful in RNA transcription and RNA transcription-based amplification, as is known to those of skill in the art. Preferably, the RNA promoter is either a T7, SP6, or T3 RNA promoter sequence. The attachment of a primary amine at 5' end can also be useful for coupling reactions, whereby the primer can be attached to labeling compounds or to solid supports via the primary amine group. Such primary amine residues can be added onto the 5' end of a nucleotide during oligonucleotide synthesis. The antisense analog 5'-XCGTGGGTCAAGACTTCTGCTTGAGCTG-3' (SEQ ID No:740) can also be used to probe for the different jun gene subtypes, where X represents a primary amine residue, nucleotide sequence recognized by a restriction enzyme, or RNA promoter sequence.

The DNA fragments for the sense primer and anti-sense primer in this invention can be easily synthesized using a DNA synthesizer. These synthesized oligonucleotides can be purified by high pressure liquid chromatography or gel electrophoresis.

The test material to be analyzed is usually total RNA or purified mRNA from cells or tissues. If desired, cells or tissues can be tested in their natural state without any pretreatment. In the case of drug testing, a drug can first be administered to cells or to tissues in a test tube. A drug can alternatively be administered (through intravenous injection, subcutaneous injection, intramuscular injection, oral administration, or intra-abdominal injection) to a laboratory animal, after which cells or tissues are removed from the animal.

Total RNA or mRNA can be purified through standard protocols, such as those described in *Molecular Cloning*, or by using a commercially available kit such as FastTrack from Invitrogen (San Diego). In either case, in order to avoid introducing any RNase into a solution containing RNA, a researcher's hands should be protected with vinyl gloves, and the instruments used for experiments should not be touched with bare hands. Also, any glass containers to be used in the experiments should be heated prior to use at approximately 250°C for at least 4 hours. Furthermore, any water to be used in the procedures should be treated with 0.1% diethyl pyrocarbonate (DEPC) incubated overnight at 37°C and autoclaved.

The cDNA to be synthesized from mRNA is made using reverse transcriptase, as described in *Molecular Cloning*. Once such cDNA has been generated, it is mixed with sense primer, antisense primer, 4 types of deoxynucleotides (dATP, dCTP, dGTP and dTTP), Taq polymerase, inorganic salts, and other necessary materials, and a PCR reaction is undertaken in a thermal cycler (Perkin-Elmer Cetus).

In order to analyze the genes amplified in this way, it is appropriate to use electrophoresis. After the amplified gene undergoes electrophoresis in an agarose gel, the DNA is stained with ethidium bromide. The amplified DNA band will be then visible under fluorescent light. After taking

-40-

photographs of the DNA band, it is also possible to quantify the intensity of each DNA band by scanning such photographs and analyzing the scanned picture with a commercially available system such as Stratascan (Stratagene, La Jolla). Furthermore, after agarose gel electrophoresis, amplified genes can be transblotted onto membranes, and subtypes of specific jun genes can be detected by hybridizing the blotted membranes with labeled probes followed by exposing labeled signals, such as ³²P or chemiluminescence, to either Polaroid films or X-ray films (i.e., doing a Southern blot).

As with the other polynucleotides that can be detected in accordance with the methods of the present invention, a variety of jun sequences can serve as sense or antisense primers for PCR methods or as probes for the detection of DNA or RNA as described herein. A method for the identification of such sequences that are either common to a variety of jun oncogenes or specific to a particular species is provided hereinbelow. In the preferred embodiment of this method of identification, a computer program is used to identify the sequences. Through use of such a program, we have identified a large number of both common and specific primers and probes. Provided as Tables XVII through XXII and XXX through XXXII are various sense sequences, i.e., sequences homologous or approximately homologous to sequences found in an organism, infectious agent, or biological component, which have been identified through the use of such a computer program as being useful as jun gene probes and primers.

All of the sequences listed in these tables are useful within the context of the PCR methods of the present invention. The complementary antisense sequences are also useful as both probes and/or PCR primers in certain aspects of the invention. As will be known by those having ordinary skill in the art, for common probes that are similar but not identical to target sequences, stringency conditions can be varied (e.g. by changes in temperature and salinity) so that such probes will hybridize or fail to hybridize with a particular target sequence. Thus, also included within the present invention are sequences that are capable of hybridizing with the same sequences as either the sense sequences listed or their anti-sense counterparts. Additional probes for jun genes include the following:

Common jun gene probes

5'-CCATGTCGATGGGGGACAGCGG-3' SEQ ID NO:741
 5'-CTGTTTAAGCTGCGCCACCTG-3' SEQ ID NO:742
 5'-GTCTGCGGCTCCTCCTTCAGGG-3' SEQ ID NO:743
 5'-CGTGGGTCAAGACTTTCTGCTTGAGCTG-3' SEQ ID NO:744

Specific probes

B type: 5'-CACTTGGTGGCCGCCAG-3' SEQ ID NO:745
 C type: 5'-GAGCATGTTGGCCGTGG-3' SEQ ID NO:746
 Human D type: 5'-GATGCGCTCCTGCGTGT-3' SEQ ID NO:747
 Mouse D type: 5'-GCCTGTTCTGGCTTTTGAGGG-3' SEQ ID NO:748

EXAMPLE 8

-41-

Synthesis of DNA fragments (sense and anti-sense primer) and amplification of mouse clones of jun oncogenes.

SEQ ID No:728 (S943-2) and SEQ ID No:729 (AS1132-2) were synthesized with a 380 B type DNA synthesizer (Applied Biosystems Co.). After treatment with ammonium hydroxide at 55°C overnight, the synthesized oligonucleotides were dried in a Speed-Vac (Savant Co.), and the concentration of each oligonucleotide was adjusted to 1 microgram/ml with water. These oligonucleotides were then at -20°C until use.

One microliter (containing approximately 10 ng) of one of the three types of mouse jun clones (jun-B, c-jun, or jun-D, obtained from ATCC), was placed in each of three reaction tubes. To each of these reaction tubes was then added 1 microliter of sense primer SEQ ID No:728, 1 microliter of anti-sense primer SEQ ID No:729, 5 microliters of 10X buffer for PCR (Promega), 1 microliter of 25 mM magnesium chloride, 4 microliters of 10 mM dNTP mix, and 0.5 microliters of Taq polymerase (Promega) were mixed, and water was added up to a total volume of 50 microliters. After adding two drops of mineral oil to each tube, PCR was undertaken using the thermal cycler, model 480 (Perkin-Elmer Cetus). After the reaction mixture was heated at 95°C for 10 minutes, PCR was carried out with the following cycles 30 times: annealing at 55°C for 1.5 minutes, extension at 72°C for 4 minutes, and denaturing at 95°C for 1.5 minutes.

After PCR, 10 microliters of the sample was mixed with 1 microliter of 10x loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol FF, and 15% Ficoll, Type 400), and electrophoresis was carried out on a 1.5% agarose gel containing 5 microgram/ml ethidium bromide. After electrophoresis, the amplified DNA bands were visualized by an ultraviolet light.

As shown in Figure 4, the clones jun-B, c-jun, and jun-D were amplified using SEQ ID NO:728 and SEQ ID NO:729, and a single band for the amplified DNA was observed at the position of approximately 270 bp in each case. This indicates that the above-mentioned set of primers can recognize and amplify all three types of mouse jun oncogenes, jun-B, c-jun and jun-D.

EXAMPLE 9

The effect of pretreatment with EGF on expression of jun oncogenes in human mononuclear leukocytes.

The following protocol describes the effect of using EGF to stimulate the production of jun oncogene mRNA:

(1) Pretreatment of human leukocytes with EGF. 40 ml of phosphate buffered saline (PBS) is added to 20 ml of heparinized human blood and mixed. 10 ml each of this sample is overlaid onto 3 ml of IsoLymph, then centrifuged for 30 minutes at 400 x g. After washing the pellet three times with PBS, the pellet is resuspended in 3 ml of PBS. 1 ml each of this sample is then placed into three tubes (No. 1 to No. 3). 1 ml of PBS is placed in a fourth tube as control. The four tubes are incubated for 10 minutes at 37°C, and EGF (Epidermal Growth Factor) is then placed in tube No. 1 at a final concentration of 30 ng/ml. After 15 minutes, EGF is placed in the same manner in tube

-42-

No. 2, and incubated for another 5 minutes. After 5 min, RNA is extracted from all 4 tubes simultaneously. Thus, the time period of the pretreatment of human leukocytes by EGF is 20 minutes for tube No. 1, 5 minutes for tube No. 2, and 0 minutes for tube No. 3.

5 (2) Extraction of RNA from cells. The above four tubes are taken out and subjected to centrifugation by a microfuge for 10 seconds. Then, the supernatant is discarded and the below described lysis buffer is added to the pellet. The lysis buffer and pellet are thoroughly mixed, after which the mixture is incubated for 30 minutes at 45°C.

Contents of lysis buffer:

10 mM EDTA pH 8.0
10 0.5% SDS (with bacteria removed by non-bacterial filter)
0.2 M NaCl
DEPC treated water
RNA inhibitor 500 unit s/ml
Vanadyl Complex 10 mM
15 Proteinase K 200 microgram/ml

20 (3) Purification of mRNA. After 5 M NaCl is added to the above cell lysates to obtain a final concentration of 0.5 M, oligo (dT) cellulose (Stratagene Co.) is added, and is reacted for 30 minutes at room temperature. Then, after washing the cellulose in 10 ml binding buffer (20 mM Tris-HCl, pH 7.6, 1 mM EDTA, 0.5 M NaCl) 5 times, 0.35 ml of DEPC treated water is added, and mRNA is eluted from the solid phase. Then, 53 microliters of 2 M sodium acetate and 2.5 times the volume of ethanol are added, and after cooling in dry ice for 20 minutes, this mixture is centrifuged at 15,000 rpm for 20 minutes. After washing the pellet in 75% ethanol one time, the pellet is dried and then dissolved in 10 microliters of DEPC treated water. The resulting mixture contains mRNA from the cells being tested.

25 (4) Synthesis of cDNA. 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM magnesium chloride, 10 mM DTT, 0.5 mM dNTP (dATP, dCTP, dGTP, dTTP), 50 micrograms/ml oligo (dT) primer, and 10,000 units/ml reverse transcriptase are added to 10 microliters of the mRNA obtained above to a total volume of 20 microliters, and this undergoes a reaction for one hour at 37°C. After the reaction, 20 microliters of a phenol:chloroform:isoamyl alcohol mixture is added, and the mixture is cooled for 20 minutes in dry ice to precipitate the cDNA. After centrifugation for 20 minutes at 10,000 rpm, the pellet is washed one time in 75% ethanol. Then, after drying, the pellet is dissolved in 20 microliters of autoclaved water to form a cDNA solution, and stored at -20 degrees C.

30 (5) PCR. One microliter each of sense primer SEQ ID No:733 and anti-sense primer SEQ ID NO:740 (1 mg/ml) for jun gene amplification are added with 2 microliters of the cDNA solution above. After mixing this with 50 mM KCl, 10 mM Tris-HCl (pH 8.4), 2.0 mM magnesium chloride, 100 micrograms/ml gelatin, and 0.2 mM dNTP, 2.5 units of Taq polymerase are added (final volume is 50 ml). After the reaction mixture is heated at 95°C for 10 minutes, PCR is carried out with the following

-43-

cycles 30 times: annealing at 55°C for 1.5 minutes, extension at 72°C for 4 minutes, and denaturing at 95 degrees C for 1.5 minutes.

(6) Agarose gel electrophoresis. After completing PCR, 10 microliters of the reacted solution is taken and electrophoresed in the same method as described in example 1. The results showed that in the leukocytes which did not undergo treatment with EGF (i.e., the sample which underwent 0 treatment time in tube No. 3) is found to have a minimal band for the amplified DNA at the position of about 270 bp size. However, it has been found that the band of amplified DNA was increased after 5 min, then returns to basal levels within 20 min.

EXAMPLE 10

Effect of pretreatment with PHA on jun gene expression in human leukocytes.

In the place of EGF, PHA (at a final concentration of 10 micrograms/ml) is utilized and the pretreatment times are set at 0 minutes, 5 minutes, 15 minutes, and 30 minutes. The procedures followed in Example 2 are then followed. As a result, with the pretreatment with PHA done at 15 minutes, it was found that the band for the amplified DNA at the position of approximately 270 bp was maximized, but for the time periods afterwards, this band decreased in intensity and therefore in the quantity of DNA it contained.

B. Primers for Detecting G Protein Sequences

Cell surface receptors for hormones and neurotransmitters are known to be coupled to intracellular heterotrimeric GTP-binding proteins (G proteins) composed of α , β and γ subunits. Once receptors are activated by specific ligands, receptor-coupled G proteins transduce signals to intracellular secondary effector systems, such as adenylyl cyclase, phospholipase C, and ion channels.

G proteins are believed to be involved in causing various disease states. For example, a genetic deficiency of G_s proteins is the molecular basis of hereditary osteodystrophy. Pituitary tumors in acromegalic patients have been shown to contain mutant G_s proteins. G proteins are also involved in invasive and metastatic melanoma cells. Rat models of streptozotocin-induced experimental diabetes suggest that the levels of mRNA for various subclasses of $G\alpha$ proteins are significantly altered from normal control rats. Furthermore, cellular functions of pertussis toxin-sensitive G proteins were shown to be significantly impaired in atherosclerotic porcine coronary arteries, while G protein function in leukocytes of patients with mania was hyperfunctional. However, currently available immunological detection methods (Western blots) and mRNA detection methods (Northern blots) are not sensitive and require a lot of cellular material, making it difficult to study the role of G proteins in such diseases.

Although G proteins have been analyzed extensively from a biochemical and immunological point of view using various antibodies, antibody production without any cross-reactivity among various subclasses or with high species specificity has been quite difficult to obtain. Therefore, recent experiments have focused on Northern blot analyses to identify G protein-specific mRNA from various tissues or cells in different species. However, Northern blots require experienced handling and protection from RNase contamination, in addition to a large amount of starting cellular materials.

In contrast to these conventional methods, PCR technology is more convenient and practically useful, because it requires less material than a Northern blot analysis and has great sensitivity. However, it is difficult for PCR to quantify the amount of DNA or mRNA in starting materials.

We have identified two highly conserved oligonucleotide sequences among five different α subunits of G proteins, SEQ ID NO:528 and SEQ ID NO:731, which can be used as PCR primers. These sequences are able to amplify the sequences of all the subclasses of G proteins under the same PCR conditions, including G protein sequences obtained from a mixture of rat G α protein clones and cDNAs derived from various human tissues. Interestingly, the final PCR products obtained using the novel G-protein PCR primers of the present invention reflect the relative composition of each of the subclasses of G α proteins present in the starting materials. This is probably because the five different G α proteins cDNAs are amplified at a similar rate with a single set of PCR primers under the same PCR conditions. If known mixtures of each of the subclasses of G α protein clones are assayed together with unknown test samples, as shown in Fig. 6, the relative composition of G α proteins can be determined fairly precisely. Therefore, the present method is ideal for the characterization of G α proteins in various tissues and cells.

Performing PCR with the primers of the present invention is also useful in clinical and diagnostic assays in the detection of disease. Since G protein abnormalities have been associated with hereditary diseases, cancer, forms of diabetes, and other diseases, the present PCR primers for detecting and quantifying G proteins can be used to detect these diseases and assess their severity.

We have identified G_{i-1}, G_{i-2}, G_{i-3}, G_s and G₀ using the PCR primers of the present invention (SEQ ID NO:528 and SEQ ID NO:731). Although recent cloning has identified more subclasses of G proteins, all of these newly identified G protein cDNAs showed a high degree of homology to other known G proteins. Therefore, it is expected that the primers of the present invention will amplify these subclasses as well, and that the present PCR the present PCR technique can also be applied to these new G proteins. Moreover, this PCR method can be utilized to clone unique G protein genes as well.

We designed two 22-mer oligonucleotides, G₂ (SEQ ID NO:528) and G₄ (SEQ ID NO:731), as PCR primers for the detection of G protein sequences. As shown in Table 3, these oligonucleotides contain sequences which are highly conserved among five different G α protein cDNAs, having only 0-4 base mismatches per sequence. No mismatch was found in the 4 bases at the 3' end of the G₂-sense and G₄-antisense sequences. Furthermore, G₂ and G₄ have no self-complementary sequences more than 3 base pairs in a row (data not shown). In order to analyze whether G₂ and G₄ are common to all the G α proteins, but not to other unrelated sequences, a homology search (DNASIS) of G₂ and G₄ sequences was carried out against all mammalian sequences in GenBank. As a result, G₂ and G₄ were found to be common to all the types of G α proteins and rhodopsins of various species, but less homologous to other unrelated sequences (data not shown).

Table 3.

-45-

Two consensus oligonucleotides (G₂ and G₄) among five different cDNAs of G protein α subunits.

Consensus sequence (# of mismatch)			
5	G ₂	Length (bp)	G ₄
	AGCACCATTGTGAAGCAGATGA		TGTTTGATGTGGGAGGCCAGAG
Gi-1	AGCAC _a ATTGTGAAGCAGATGA (1)	476	TGTTTGAC _e GTGGGAGGCCAGAG (1)
Gi-2	AGCACCAT _e GT _e AAGCAGATGA (2)	479	TGTTTGATGTGGG _t GG _t CAG _e G (3)
10 Gi-3	AG _t AC _t ATTGTGAA _a CAGATGA (3)	476	TGTTTGATGT _a GG _t GGCCA _a AG (3)
G _s	AGCACCATTGTGAAGCAGATGA (0)	524	TGTT _e GATGTGGG _e GGCCAG _e G (3)
G ₀	AGCACCATTGTGAAGCAGATGA (0)	479	TGTTGAC _e GT _t GG _g GGCCAG _e G (4)

15 PCR was first carried out at different annealing temperatures ranging from 37°C to 65°C using the λ gt10 library of human HL-60 cells. As a result, PCR products were seen only at 45°C and 55°C with a size of approximately 500 bp (data not shown), which was similar to theoretical values (576 to 524 bp) (see Table 3 above). Therefore, all the PCR was then carried out at an annealing temperature of 45°C.

20 As shown in Fig. 5, cloned rat G α protein cDNAs (G_{i-1}, G_{i-2}, G_{i-3}, G_s, G₀) were successfully amplified using the same set of PCR primers (G₂ and G₄) with a size of approximately 500 bp in 1.2% agarose gels stained with ethidium bromide. According to the computer analysis (DNASIS), the nucleotide sequences of the amplified PCR products were less homologous among five G α proteins with the percentage of similarity ranging from 76.8% to 47.6%. This indicates that after PCR amplification, each of the components of G α proteins can be identified by Southern blot analysis, even though the sizes of the PCR products generated are very similar among the five G α proteins. Therefore, another PCR was carried out in which 35% of the dTTP was replaced with biotin-conjugated dUTP in order to prepare subclass-specific, biotin-labeled probes. Southern membranes were then probed with these biotin-PCR products. As shown in Fig. 5, these biotin-PCR probes were highly specific to each G α protein subclass with washing temperature at 65°C. At low stringent washing, these probes cross-hybridized with other subclasses of G α proteins (data not shown).

30 By using the G₂ and G₄ sequences, all the subclasses of G α protein cDNA were amplified with PCR when an equal amount of G_{i-1}, G_{i-2}, G_{i-3} and G₀ were present in test samples (Fig. 6, lane 4, 5, 10). However, if all the concentrations of G α protein cDNA are abundant, G₀ is less amplified (Fig. 6, lane 10), probably because the number of mismatches between G₀ and G₄ is higher than others between G₄ and the G₂ sequences. If 1 or 2 of the 5 G α protein cDNAs were present in smaller quantities than the others, the amounts of amplified cDNA were relatively correlated with the starting concentrations of cDNAs (Fig. 6, lane 1, 2, 3, 8, 9). Furthermore, if 1 of 5 of the G α proteins' cDNA is more abundant than that of the others, this G protein gene was amplified more than others (Fig. 6, lane 6, 7).

-46-

Using this PCR method, G α protein genes were amplified not only from cloned cDNAs, but also from various rat cDNAs (Fig. 7). In λ ZAP cDNA libraries from rat pituitary glands and cDNA from rat kidney KNRK cells, G₀ was more abundant than G_s, G_{i-2} and G_{i-3}, and G_{i-1} was undetectable (Fig. 7, lane 1, 2). λ ZAP cDNA library of rat intestine contained more G_{i-2}, G_{i-3}, and G_s and G₀ (Fig. 7, lane 3).

According to the sequence analyses, the PCR products of rat G_{i-1}, G_{i-2}, G_{i-3}, G_s and G₀ sequence amplification exhibited a high degree of homology to human G protein cDNAs, see Table 4 below). Furthermore, as shown in Fig. 8, PCR with a pair of G₂ and G₄ primers could amplify 500 bp DNA from cDNAs of human IM9 and Jurkat cells. Unlike rat cDNAs (Fig. 7), both IM9 and Jurkat cells contained all the subclasses of G α proteins (Fig. 8). However, G_{i-3} is relatively more abundant in IM9 cells, while G_s and G₀ were more in Jurkat cells than IM9 cells (Fig. 8).

Table 4.

Nucleotide sequence similarity of PCR products between rat and human G proteins.

	Length (bp)No. of		mismatch	% Similarity
	Rat	Human		
Gi-1	476	476	62	87.0%
Gi-2	479	479	44	90.8%
Gi-3	476	476	40	91.6%
Gs	524	482	52	89.2%
Go	479	479	39	91.9%

Example 10 describes a method of amplifying and detecting G proteins with another set of PCR primers of the present invention, G2-S and G4-AS.

EXAMPLE 10

Amplifying G Proteins With PCR Primers

Materials. The cDNAs of rat G protein α subunits (G_{i-1}, G_{i-2}, G_{i-3}, G_s and G₀) were provided by Dr. R.R. Reed (Johns Hopkins Univ., MD). λ ZAP libraries of rat pituitary and intestine were provided by Dr. D.G. Payan (Univ. Calif. San Francisco). Kirsten murine sarcoma virus transformed rat kidney cells (KNRK), human IM9 B-lymphocytes and human Jurkat T-lymphocytes were obtained from American Type Tissue Culture Collection, Rockville, MD). Cell culture media, Superscript (Gibco/BRL, Gaithersburg, MD), reagents for PCR (Promega, Madison, WI), ECL (Amersham, Arlington Height, IL), Genius, Lumi-Phos 530 (Boehringer-Mannheim, Indianapolis, IN), FastTrack (Invitrogen San Diego, CA), λ gt10 library of human HL-60 cells, biotin-dUTP, alkaline phosphatase-

-47-

conjugated streptavidine (Clontech, Palo Alto, CA), dNTP (Pharmacia, Piscataway, NJ) were obtained from the designated suppliers. Other chemicals were purchased from Sigma (St. Louis, MO).

Cell culture. KRNK cells were grown in Dulbecco's modified Eagles medium containing 10% fetal calf serum, 100 U/ml penicillin and 100 µg/ml streptomycin at 37°C in 5% CO₂/95% air. Cells were fed every other day and passaged at 70-90% confluency with 0.1% trypsin in CA²⁺-Mg²⁺-free saline containing 0.02% EDTA. IM9 and Jurkat cells was grown in RPMI 1640 containing 10% fetal calf serum, 100 U/ml penicillin and 100 µg/ml streptomycin at 37°C in 5% CO₂/95%. Cell viability was more than 90% as assessed by the exclusion of trypan blue.

Primer design. Rat clones of Gα proteins (G_{i-1} (RATBPGTPB), G_{i-2} (RATBPGTPA), G_{i-3} (RATBPGTP), G_s (RATBPGTPD), and G₀ (RATBPGTPC) were retrieved from GenBank release 65.0 (HIBIIO, Hitachi America, Brisbane, CA). The nucleotide sequence similarity among these clones were then analyzed by the multiple alignment program (DNASIS, Hitachi). We have initially identified 7 highly conserved areas among them. These conserved nucleotide sequences were then analyzed against all mammalian sequences in GenBank in order to identify other similar sequences. The designed oligonucleotides G2-S (SEQ ID NO:528) and G4-AS (SEQ ID NO:731) were synthesized by Genosys Biotechnologies (Woodlands, TX), and suspended in water at 100 pg/ml.

PCR. One µl of the template DNA was mixed with 1 mM each of dATP, dGTP, dCTP and dTTP, 1 µl of each PCR primers, 1 µl of 25 mM MgCl₂, 5 µl PCR buffer, and 0.5 µl of Taq polymerase (18). PCR was then carried out in a DNA thermal cycler (model 480, Perkin-Elmer Cetus, Norwalk, CT) with 30 cycles of annealing temperature at ranging from 37°C to 65°C for 1.5 min, 72°C extension for 4 min followed by 95°C denaturation for 1.5 min. In separate experiments, 35% of dTTP was replaced with biotin-dUTP in order to prepare biotin-labeled probes.

Southern blot. PCR products were separated by electrophoresis in 1.2% agarose, and stained with ethidium bromide (19). Gels were then depurinated in 0.25 N HCl for 30 minutes and denatured in 0.5 N NaOH containing 1.5 M NaCl for 30 minutes. The gels were then neutralized with 1.0 M Tris, pH 7.6 containing 1.5 M NaCl for 30 minutes. Gels were then placed onto nylon membranes (MagnaGraph, MSI, Westboro, MA) prewetted in 10X SSPE for 10 min, and DNA was transferred onto membranes by positive pressure at 75 mmHg for 60 minutes (Posiblot, Stratagene, La Jolla, CA). The DNA from the gel was then cross-linked to the membranes with ultraviolet light at 120 mJoules (Stratalinker, Stratagene), and the membranes were incubated with hybridization buffer (ECL) containing 5% blocking reagent (ECL) and 0.5 M NaCl at 40°C for more than 1 hour. Heat denatured biotin-labeled PCR probes were then added, and hybridization was continued overnight. The membranes were washed four times for 15 minutes each time with primary wash buffer (0.5x SSPE, 36 w/v% urea, 0.4w/v% SDS) at 45-65°C, then washed twice for 5 minutes with secondary wash buffer (2x SSPE) at room temperature, and were incubated with the blocking buffer (Genius) for at least 3 hours at room temperature. Alkaline phosphatase-conjugated streptavidine (1:5,000 dilution) was then added, and incubation was continued for an additional 30-60 minutes at room temperature. The membranes

-48-

were washed four times for 15 minutes with buffer A (100 mM Tris, pH 7.5, 150 mM NaCl) at room temperature, were washed for 2 minutes once with buffer C (100 mM Tris, pH 9.5, 100 mM NaCl, 50 mM MgCl₂), and soaked in Lumi-Phos 530 for approximately 1-2 minutes. The membranes were then wrapped with transparency films, and chemiluminescent signals were allowed to expose X-ray films (XAR-5, Kodak, Rochester, NY) for between 10 minutes and 1 hour.

mRNA preparation and cDNA synthesis. The cells were washed with phosphate buffered saline three times, homogenized in lysis buffer (FastTrack), and then incubated at 45°C for 1 hour to eliminate any RNase activity. NaCl concentrations were adjusted at 0.5 M, and an oligo (dT) cellulose tablet was added to lysis buffer. Incubation was then continued at room temperature for an additional 40 minutes. After oligo (dT) cellulose was washed with binding buffer (FastTrack) four times, bound mRNA was eluted with DEPC-treated water. Concentrations of mRNA were determined in a spectrophotometer (Hitachi, U-2000, Irvine, CA) at OD₂₆₀. The first strand cDNA was synthesized from a template mRNA in the presence of 50 mM Tris, pH 8.3, 75 mM KCl, 3 mM MgCl₂, 10 mM DTT, 0.5 mM each of dATP, dCTP, dGTP, and dTTP, poly (dT) as a primer, and reverse transcriptase (Superscript) at 37°C for 1 hour. Second strand cDNA was then synthesized in the same tube, containing 25 mM Tris, pH 7.5, 100 mM KCl, 5 mM MgCl₂, 10 mM (NH₄)₂SO₄, 0.15 mM β-NAD⁺, 250 μM each of dATP, dGTP, dCTP and dTTP, 1.2 mM DTT, 65 U/ml DNA ligase, 250 U/ml DNA polymerase, and 13 U/ml RNase H (Superscript) for 2 hours at 16°C. Synthesized cDNAs were then extracted once with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1), precipitated with ethanol, and resuspended in H₂O.

Graphic presentation. Data on Polaroid films and X-ray films was scanned by Stratascan (Stratagene) with optimization of signal-to-noise ratio, then edited with desk top publishing software (PageMaker, Aldus, Seattle, WA). As shown in Figure 5, the combination of SEQ ID NO:528 and SEQ ID NO:731 can amplify all of the subtypes of G protein α subunits.

It will be evident to one having ordinary skill in the art that a variety of sequences could serve as sense or antisense primers for PCR methods or as probes for the detection of DNA or RNA as described herein. A method for identification of such sequences that are either common to a variety of G proteins or specific to a particular species is provided hereinbelow. In the preferred embodiment of this method of identification, a computer program is used to identify the sequences. Through use of such a program, we have identified a large number of both common and specific primers and probes. Provided as Tables XXIII through XXIX and XXXIII through XXXVII are various sense sequences identified through the use of such a program that are useful as G protein probes and primers.

All of the sequences listed in these tables are useful within the context of the PCR methods of the present invention. The complementary antisense sequences are also useful in certain aspects of the invention. As will be known having ordinary skill in the art, for common probes that are similar, but not identical to target sequences, stringency conditions can be varied (e.g. by changes in temperature and salinity) so that such probes will hybridize or fail to hybridize with a particular target sequence.

-49-

Thus, also included within the present invention are sequences that are capable of hybridizing with the same sequences as either the sense sequences listed or their anti-sense counterparts.

Additional probes for G protein also include the following:

Common G protein probes

5 5'-CTCTGGCCTCCCACATCAAACA-3' SEQ ID NO:749
5'-TCATCTGCTTCACAATGGTGCT-3' SEQ ID NO:750

Specific probes (Human & Rat common)

Gi-1 5'-GTTTTCACTCTAGTTCTGAGAACATC-3' SEQ ID NO:751
Gi-2 5'-CAAAGTCGATCTGCAGGTTGC-3' SEQ ID NO:752
10 5'-ATGGTCAGCCCAGAGCCTCCGG-3' SEQ ID NO:753
Gi-3 5'-GTCTTCACTCTCGTCCGAAGA-3' SEQ ID NO:754
Gs 5'-GCCTTGGCATGCTCATAGAATT-3' SEQ ID NO:755
5'-TTCATCCTCCACAGAGCCTTG-3' SEQ ID NO:756
Go 5'-CGCATCATGGCAGAAAGCAG-3' SEQ ID NO:757

15

C. Primers for Detecting Other Biological Components

Another example of a primer or probe for detecting a biological component is a sequence specific for the mRNA of substance P. Substance P is a neurotransmitter expressed by nerves that are involved in pain receptor pathways. We have discovered that the sequence 5'-
20 TGGTACGCTTTCTCATAAGTCC-3' (SEQ ID NO:758) is very specific for Substance P.

Another biological component which can be probed for is the mRNA for the β receptor. The β receptor is a protein located in human nerve tissue. In particular, abnormalities in the β_2 receptor has been found to be closely correlated with asthma. Thus, measuring the mRNA for β_2 receptor can be used to determine the pathophysiology of asthma patients, and could also be used to assess the
25 effectiveness of anti-asthma agents. We have found that the sequence 5'-ATGCTGGCCGTGACGCACAGCA-3' (SEQ ID NO:759) is common to a number of human subtypes of β receptor, including β_1 (only one mismatch), β_2 (no mismatches), and β_3 (2 mismatches). Thus, SEQ ID NO:759 can be used to probe for all three of these subtypes of β receptor.

XVI. Identifying PCR Primers and Probes

30 PCR primers and probes for use in the methods of the present invention can be identified in any way known to the art. Preferably, however, such probes and primers are identified by a computer. We have developed a novel computer system for identifying the sequences to be used in such probes and primers. This system is an automated system which allows the user to calculate and design extremely accurate oligonucleotide probes and PCR primers.

35 The software of the present invention runs under Microsoft Windows® on IBM® compatible personal computers (PC's). This invention allows a researcher to design oligonucleotide probes based

-50-

on the GenBank database of DNA and mRNA sequences. The present invention further allows examination of probes for specificity or commonality with respect to a user-selected target gene sequences. Hybridization strength between a probe and a target subsequence of DNA or mRNA can be estimated through a hybridization strength model. Quantitatively, hybridization strength is given as the melting temperature (T_m).

Two models for estimating hybridization strength models are supported by this invention: 1) the Mismatch Model and 2) the H-Site Model. In either case, the user can select the following calculations for each probe, results of which are then made available for display and analysis: 1) Sequence, Melting Temperature (T_m) and Hairpin characteristics (a hairpin is a nucleotide sequence that is homologous to itself and can "fold back" with one portion of the probe hybridizing to another portion of the same probe); 2) Hybridization to other species within the preparation mixture; and (3) Location and T_m for the strongest hybridizations. The results of the invention's calculations are then displayed on a Mitsuhashi Probe Selection Diagram (MPSD) which is a graphic display of all potential hybridizations between the target mRNA and the probe sequences in the preparation.

The Main OligoProbe Design Station dialog window controls all user-definable settings in the program. The user is offered a number of options at this window. The File option allows the user to print, print in color, save selected probes, and exit the program. The Preparation option allows the user to open and create preparation (PRP) files. The Models option allows the user to choose between the two hybridization models currently supported by the OligoProbe DesignStation: 1) the H-Site Model and 2) the Mismatch Model.

If the user selects the H-Site Model option, the melting temperature for each probe and the nucleation threshold parameters can be set. The nucleation threshold is the number of base pairs constituting a nucleation site (a subsequence with an exact match). If the user selects the Mismatch Model option, the probe length and mismatches (N) can be set.

Mismatch Model

The Mismatch Model is used for designing DNA and mRNA probes utilizing sequence database information from sources such as GenBank. In this Model, hybridization strength is related only to the number of base pair mismatches between a probe and its target. Generally, the more mismatches a user allows when setting parameters, the more probes will be identified. The Mismatch Model does not take into account the GC content of candidate probes so there is no calculation of the probe's binding strength.

The basic technologies employed by the Mismatch model are hashing and continuous seed filtration. Hashing involves the application of an algorithm to the records in a set of data to obtain a symmetric grouping of the records. When using an indexed set of data such as a database, hashing is the process of transforming a record key to an index value for storing and retrieving a record. The Mismatch Model is essentially a quick process for determining exact and inexact matching between DNA and mRNA sequences to support the Mitsuhashi Probe Selection Diagram (MPSD).

The algorithm used by the Mismatch Model is based on the Waterman-Pevzner Algorithm (WPALG), which is a computer-based probe selection process. Essentially, this is a combination of new and improved pattern matching processes. See Hume and Sunday (1991, Ref. 4), Landau et al (1986-1990, Refs. 6, 7, 8), Grossi and Luccio (1989, Ref. 3), and Ukkonen (1982, Ref. 14).

There are three principal programs that make up the Mismatch Model in this implementation of the invention. The first is designated by the inventors as "k_diff." WPALG uses k_diff to find all locations of matches of length greater than or equal to one (1) (length is user-specified) with less than or equal to k number of mismatches (k is also user-specified) between the two sequences. If a candidate oligonucleotide probe fails to match that well, it is considered unique. k_diff uses hashing and continuous seed filtration, and looks for homologs by searching GenBank and other databases with similar file formats. The technique of continuous seed filtration allows for much more efficient searching than previously implemented techniques.

A seed is defined in this invention to be a subsequence having a length equal to the longest exact match in the worst case scenario. For example, suppose the user selects a probe length (l) of 18, with 2 or fewer mismatches (k). If a match exists with 2 mismatches, then there must be a perfectly matching subsequence of length equal to 6. Once the seed length has been determined, the Mismatch Model looks at all substrings of that seed length (in this example, the seed length would be 6), finds the perfectly matched base pair subsequence of length equals 6, and then looks to see if this subsequence extends to a sequence of length equal to the user selected probe length (i.e., 18 in this example). If so, a candidate probe has been found that meets the user's criteria.

Where the seed size is large (i.e., a long string of unique nucleotides), the program allocates a relatively large amount of memory for the hash table. This invention has an option that allows memory allocation for GenBank entries just once at the beginning of the program, instead of reallocating memory for each GenBank entry. This reduces input time for GenBank entries by as much as a factor of two (2), but this method requires the user to know the maximum GenBank entry size in advance.

A probe is found to hybridize if it has k or fewer mismatches with a target sequence from the database or file searched. The hit extension time for all appropriate parameters of the Mismatch Model has been found by experimentation to be less than thirty-five (35) seconds, except in one case where the minimum probe length (l) was set to 24 and the maximum number of mismatches (k) was set to four (4). This situation would rarely be used in real gene localization experiments because the hybridization conditions are too weak.

H-Site Model

In this embodiment of the invention, the second hybridization strength model is termed the H-Site Model. One aspect of the H-Site Model uses a generalization of an experimental formula to analyze nucleotide binding strength. The basic formula on which this aspect of the model is built is as follows:

$$T_m = 81.5 - 16.6(\log[Na]) - .63 \%(\text{formamide}) + .41 (\% (G + C)) - 600 / N$$

-52-

In this formula, $\log[\text{Na}]$ is the log of the sodium concentration, $\%(G + C)$ is the fraction of matched base pairs which are G-C complementary, and N is the probe length. This formula relates the fact that melting temperature is a function of both probe length and percent GC content. This basic formula has been modified in this invention to account for the presence of mismatches. Each percent of mismatch reduces the melting temperature by an average of 1.25° (2°C for an AT mismatch, and 4°C for a GC mismatch). This formula is, however, an approximation. The actual melting temperature might potentially differ from this approximation, especially for short probes or probes with a relatively large number of mismatches.

Hybridization strength in the H-Site Model is related to each of the following factors: 1) "binding region"; 2) type of mismatch (GC or AT substitution); 3) length of the probe; 4) GC content of the binding region; and 5) existence of a "nucleation site" (a subsequence with an exact match). The type of mismatch and GC content of the binding region from each sequence contributes to a candidate probe's binding strength. The binding strength from each probe is thereby determined enabling the user to select an optimal probe.

The fundamental assumption of the H-Site Model is that binding strength is mostly determined by a paired subsequence of the probe and target, called the binding region. If the subsequence binding region contains more GC pairs than AT pairs, the binding strength will be higher due to the greater number of hydrogen bonds between G and C bases (three bonds) in comparison to A and T bases (two bonds). Thus, GC rich probes have a higher melting temperature and subsequently form stronger hybridizations.

In the H-Site Model the program determines optimal probes, ideally without any mismatches to the target gene. With this model, however, a candidate probe can have more AT mismatches if the sequence is GC rich. The amount of allowable AT mismatches in a specific sequence is determined in the present invention program by looking primarily at subsequence regions of the probe and target that match without penalizing the probe for areas that mismatch. If the mismatches are located at either or both of the ends of the binding region, there is little effect on the overall stability of the base-pairing. Centrally located mismatches in the binding region are much more deleterious, as this will significantly lower the binding strength of the probe.

The formula cited above for the melting temperature applies within the binding region. The length of the probe is used to calculate percentages, but all other parameters of the formula are applied to the binding region only. The H-Site Model further assumes the existence of a nucleation site. The length of this nucleation site may be set by the user. Typically, a value of 8 to 10 base pairs is used. To complete the H-Site Model, the binding region is chosen so as to maximize the melting temperature T_m among all regions containing a nucleation site, assuming one exists (otherwise, $T_m = 0$).

The H-Site Model is more complex than the Mismatch Model discussed above in that hybridization strength is modeled as a sum of multiple subsequence contributions, with matches generally providing positive binding energy and mismatches generally providing negative binding energy.

-53-

The exact binding energies to be used depend only on the matched or mismatched pair. These coefficients may be specified by the user, although in the current version of this invention these coefficients are not explicitly user-selectable, but rather are selected to best fit the hybridization strength formulas developed by Itakura et al (1984, Ref. 5), Bolton and McCarthy (1962, Ref. 2), Benner et al (1973, Ref. 1), and Southern (1975, Ref. 13).

A unique aspect of the H-Site Model is that hybridization strength is determined by the optimal binding region between the candidate probe and binding locus. This binding region is called the hybridization site, or h-site, and is selected so as to maximize overall hybridization strength, so that mismatches outside the binding region do not detract from the estimated hybridization strength. Several other unique features of the H-Site Model include the fact that it is more oriented toward RNA and especially cDNA sequences than DNA sequences, and the fact that the user has control over preparation and environmental variables.

The emphasis on RNA and cDNA sequences allows the user to concentrate on coding regions of genes, rather than necessitating sorting through all of a genomic sequence for the desired probe. The enhanced user control over environmental and preparation variables allows the user to more accurately simulate laboratory conditions that closely correspond with any experiments he or she is conducting. Further, this implementation of the invention does some preliminary preprocessing of the GenBank database to sort out and select the cDNA sequences. This is done by locating a keyword (in this case CDS) in each GenBank record, thereby eliminating any sequences containing introns.

The Mitsuhashi Probe Selection Diagram (MPSD), Figure 14, is a key feature of this invention, as it is a unique way of visualizing the results of the probe designed by the Mismatch and H-Site Models. It is a graphic display of all of the hybridizations of candidate oligonucleotide probes and the target with all sequences in the preparation. Given a gene sequence database and a target mRNA sequence, the MPSD graphically displays all of the candidate probes and their hybridization strengths with all sequences from the database. In the present implementation, each melting temperature is displayed as a different color, from red (highest T_m) to blue (lowest T_m). The MPSD allows the user to see visually the number of false hybridizations at various temperatures for all candidate probes, and the sources of these false hybridizations (with a loci and sequence comparison). A locus may be a specific site or place, or, in the genetic sense, a locus is any of the homologous parts of a pair of chromosomes that may be occupied by allelic genes.

These probes may then be used to test for the presence of precursors of specific proteins in living tissues. The oligonucleotide probes designed with this invention may be used for medical diagnostic kits, DNA identification, and potentially continuous monitoring of metabolic processes in human beings. The present implementation of this computerized design tool runs under Microsoft® Windows™ v. 3.1 (made by Microsoft Corporation; Redmond, Washington) on IBM® compatible personal computers (PC's).

-54-

The H-Site Model of this invention is unique in that it offers a multitude of information on selected probes and original and distinctive means of visualizing, analyzing and selecting among candidate probes designed with the invention. Candidate probes are analyzed using the H-Site Model for their binding specificity relative to some known set of mRNA or DNA sequences, collected in a database such as the GenBank database. The first step involves selection of candidate probes at some or all the positions along a given target. Next, a melting temperature model is selected, and an accounting is made of how many false hybridizations each candidate probe will produce and what the melting temperature of each will be. Lastly, the results are presented to the researcher along with a unique set of tools for visualizing, analyzing and selecting among the candidate probes.

This invention is both much faster and much more accurate than the methods that are currently in use. It is unique because it is the only method that can find not only the most specific and unique sequence, but also the common sequences. Further, it allows the user to perform many types of analysis on the candidate probes, in addition to comparing those probes in various ways to the target sequences and to each other.

Therefore, it is the object of this invention to provide a practical and user-friendly system that allows a researcher to design both specific and common oligonucleotide probes, and to do this in less time and with much more accuracy than currently done.

This invention is employed in the form best seen in Figure 11. There, the combination of this invention consists of an IBM® compatible personal computer (PC), running software specific to this invention, and having access to a distributed database with the file formats found in the GenBank database and other related databases.

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 hereunder are incorporated herein by reference.

The preferred computer hardware capable of operating this invention involves of a system with at least the following specifications (Figure 11): 1) an IBM® compatible PC, generally designated 1A, 1B, and 1C, with an 80486 coprocessor, running at 33 Mhz or faster; 2) 8 or more MB of RAM, 1A; 3) a hard disk 1B with at least 200 MB of storage space, but preferably 1 GB; 4) a VGA color monitor (1C) with graphics capabilities of a size sufficient to display the invention's output in readable format, preferably with a resolution of 1024 x 768; and 5) a 580 MB CD ROM drive 5 (1B of Figure 11 generally refers to the internal storage systems included in this PC, clockwise from upper right, two floppy drives, and a hard disk). Because the software of this invention preferably has a Microsoft® Windows™ interface, the user will also need a mouse 2, or some other type of pointing device.

The preferred embodiment of this invention would also include a laser printer 3 and/or a color plotter 4. The invention may also require a modem (which can be internal or external) if the user does

-55-

not have access to the CD ROM versions of the GenBank database 8 (containing a variable number of gene sequences 6). If a modem is used, information and instructions are transmitted via telephone lines to and from the GenBank database 8. If a CD ROM drive 5 is used, the GenBank database (or specific portions of it) is stored on a number of CDs.

5 The computer system should preferably have at least the Microsoft® DOS version 5.0 operating system running Microsoft® Windows™ version 3.1. All of the programs in the preferred embodiment of the invention were written in the Borland® C++ (Borland International, Inc.; Scotts Valley, CA) computer language. It should be noted that subsequently developed computers, storage systems, and languages may be adapted to utilize this invention and vice versa.

10 This inventive computer program is designed to enable the user to access DNA, mRNA and cDNA sequences stored either in the GenBank or in databases with similar file formats. GenBank is a distributed flat file database made up of records, each record containing a variable number of fields in ASCII file format. The stored database itself is distributed, and there is no one database management system (DBMS) common to even a majority of its users. One general format, called the
15 line type format, is used both for the distributed database and for all of GenBank's internal record keeping. All data and system files and indexes for GenBank are kept in text files in this line type format.

 The primary GenBank database is currently distributed in a multitude of files or divisions, each of which represents the genome of a particular species (or at least as much of it as is currently known
20 and sequenced and publicly available). The GenBank provides a collection of nucleotide sequences as well as relevant bibliographic and biological annotation. Release 72.0 (6/92) of the GenBank CD distribution contains over 71,000 loci with a total of over ninety-two (92) million nucleotides. GenBank is distributed by IntelliGenetics, of Mountain View, CA, in cooperation with the National Center for Biotechnology Information, National Library of Medicine, in Bethesda, MD.

25

1. Overall Description of the OligoProbe DesignStation

a. **General Theory**

 The intent of this invention is to provide one or more fast processes for performing exact and
 inexact matching between DNA sequences to support the Mitsuhashi Probe Selection Diagram (MPSD)
30 discussed below, and other analysis with interactive graphical analysis tools. Hybridization strength between a candidate oligonucleotide probe and a target subsequence of DNA, mRNA or cDNA can be estimated through a hybridization strength model. Quantitatively, hybridization strength is given as the melting temperature (T_m). Currently, two hybridization strength models are supported by the invention:
 1) the Mismatch Model and 2) the H-Site Model.

-56-

b. Inputs**i. Main OligoProbe DesignStation Dialog Window**

5 The Main OligoProbe DesignStation dialog window, Figure 12, controls all user-definable settings. This window has a menu bar offering five options: 1) File 10; 2) Preparation 80; 3) Models 70; 4) Experiment 90; and 5) Help 50. The File 10 option allows the user to print, print in color, save selected probes, and exit the program. The Preparation 80 option allows the user to open and create preparation (PRP) files.

10 The Models 70 option allows the user to chose between the two hybridization models currently supported by the OligoProbe DesignStation: 1) the H-Site Model 71 and 2) the Mismatch Model 75. If the user selects the H-Site Model 71 option, the left hand menu of Figure 12C is displayed and the user sets the following model parameters: 1) the melting temperature T_m 72 for which probes are being designed (i.e., the melting temperature that corresponds to a particular experiment or condition the user desires to simulate); and 2) the nucleation threshold 73, which is the number of base pairs constituting a nucleation site. If the user selects the Mismatch Model 75 option, the right hand menu of Figure 12C
15 is displayed and the user sets the following model parameters: 1) probe length 76, which is the number of base pairs in probes to be considered; and 2) mismatch N 77, which is the maximum number of mismatches constituting a hybridization. Computation of the user's request takes longer with the H-Site Model if the threshold 73 setting is decreased, but longer with the Mismatch Model if the number of mismatches K 77 is increased.

20 In addition, for both Model options the user chooses the target species 11 DNA or mRNA for which probes are being designed and the preparation 12, a file of all sequences with which hybridizations are to be calculated. A sample of a target species file is shown in Figure 37 (humbjunix.cds), while a sample of a preparation file is shown in Figure 38 (junmix.seq). Each of these inputs is represented by a file name and extension in standard DOS format. In the target species and preparation fields, the file
25 format follows the GenBank format with each of the fields having a default file extension. Pressing the "OK" button 91 (Figure 12C) will initiate processing, while pressing the "Cancel" button 93 will stop the processing.

The Experiment 90 option and the Help 50 option are expansion options not yet available in the current implementation of the invention.

30 c. Processing

Figures 13A and 13B are flow charts of the overall OligoProbe DesignStation Program, illustrating its sequence and structure. Generally, the main or "control" program of the OligoProbe DesignStation performs overall maintenance and control functions. This program, as illustrated in Figures 13A and 13B, accomplishes the general housekeeping functions 51, such as defining global
35 variables. The user-friendly interface 53, carries out the user-input procedures 55, the file 57 or database 59 access procedures, calling of the model program 62 or 63 selected by the user, and the user-

selected report 65 or display 67, 69, 71 and 73 features. Each of these features is discussed in more detail in later sections, with the exception of the input procedures, which involves capturing the user's set-up and control inputs.

d. **Outputs**

5 i. The Mitsuhashi Probe Selection Diagram Window

The Mitsuhashi Probe Selection Diagram (MPSD), Figure 14, is a key feature of the invention as it is a unique way of visualizing the results of the program's calculations. It is a graphic display of all of the hybridizations of probes with the target oligonucleotides in the preparation. Specifically, given a nucleotide sequence database and a target mRNA, the MPSD graphically displays all of the candidate probes and their hybridization strengths with all sequences from the nucleotide database. The MPSD
10 allows the user to visualize the number of false hybridizations at various temperatures for all candidate probes, and the sources of these false hybridizations (with a loci and sequence comparison).

For each melting temperature selected, a graph showing the number of hybridizations for each probe is displayed. In the preferred embodiment, the graphs are color coded. In this implementation
15 of the invention, the color red 123 identifies the highest melting temperature and the color blue 124 identifies the lowest melting temperature. Each mismatch results in a reduction in the T_m value. The melting temperature is also a function of probe length and percent GC content. Within the window, the cursor 125 shape is changed from a vertical line bisecting the screen to a small rectangle when the user selects a particular probe. The current probe is defined to be that probe under the cursor position
20 (whether it be a line or a rectangle) in the MPSD window. More detailed information about the current probe is given in the ProbeInfo and MatchInfo windows, discussed below. Clicking the mouse button 2 once at the cursor 125 selects the current probe. Clicking the mouse button 2 a second time deselects the current probe. Moving the cursor across the screen causes the display to change and reflect the candidate probe under the current cursor position.

25 The x-axis 110 of the MPSD, Figure 14, shows the candidate probes' starting positions along the given mRNA sequence. The user may "slide" the display to the left or right in order to display other probe starting positions. The y-axis 115 of the MPSD displays the probe specificity, which is calculated by the program.

The menu options 116, 117, 118, 119, and 120 available to the user while in the MPSD, Figure
30 14, and are displayed along a menu bar at the top of the screen. The user can click the mouse 2 on the preferred option to briefly display the option choices, or can click and hold the mouse button on the option to allow an option to be selected. The user may also type a combination of keystrokes in order to display an option in accordance with well-known computer desk top interface operations. This combination usually involves holding down the ALT key while pressing the key representing the first
35 letter of the desired option (i.e. F, P, M, E or H).

-58-

The File option 116 allows the user to specify input files and databases. The Preparation option 117 allows the user to create a preparation file summarizing the sequence database. The Models option 118 allows the user to specify the hybridization model (i.e., H-Site or Mismatch) and its parameters. The Experiment option 119 and the Help option 120 are not available in the current implementation of this invention. These options are part of the original Main OligoProbe DesignStation dialog window, Figure 12.

Areas on the graphical display of the MPSD, Figure 14, where the hybridizations for the optimal probes are displayed are lowest and most similar, such as shown at 121, indicate that the particular sequence displayed is common to all sequences. Areas on the graphical display of the MPSD where the hybridizations for the optimal probes are displayed are highest and most dissimilar, such as shown at 122, indicate that the particular sequence displayed is extremely specific to that particular gene fragment. The high points on the MPSD show many loci in the database, to which the candidate probe will hybridize (i.e., many false hybridizations). The low points show few hybridizations, at least relative to the given database. Specifically, the sequence shown at 121 would reflect a probe common to all of the gene fragments tested, such that this probe could be used to detect each of these genes. The sequence shown at 122 would reflect a probe specific to the particular gene fragment, such that this probe could be used to detect this particular gene and no others.

ii. The ProbeInfo and MatchInfo Window

The combined ProbeInfo and MatchInfo Window, Figure 15, displays detailed information about the current candidate probe. The upper portion of the window is the ProbeInfo window, and the lower portion is the MatchInfo window. The ProbeInfo window portion displays the following types of information: the target locus (i.e., the mRNA, cDNA, or DNA from which the user is looking for probes) is displayed at 131, while the preparation used for hybridizations is displayed at 132. In the example shown in Figure 15, the target locus 131 is the file named HUMBJUNX.CDS, which is shown as being located on drive F in the subdirectory MILAN. The preparation 132 is shown as being the file designated JUNMIX.PRP, which is also shown as being located on drive F in the subdirectory MILAN. The JUNMIX.PRP preparation in this example is a mixture of human and mouse jun loci.

The current and optimal probe's starting position is shown at 135. The current candidate oligonucleotide probe is defined at 136, and is listed at 137 as having a length of 21 bases. The melting temperature for the probe 136 as hybridized with the targets is shown in column 140. The melting temperature for the optimal probe is given as 61.7 degrees C at 138. The ProbeInfo Window Figure 15 also displays hairpin characteristics of the probe at 139. In the example shown, the ProbeInfo Window shows that there are four (4) base pairs involved in the worst hairpin, and that the worst hairpin has a length of one (1) (see Figure 15, at 139).

The MatchInfo Window portion displays a list of hybridizations between the current probe and species within the preparation file, including hybridization loci and hybridization temperatures. The

hybridizations are listed in descending order by melting temperature. The display shows the locus with which the hybridization occurs, the position within the locus, and the hybridization sequence.

In the MatchInfo window portion, the candidate probe 136 is shown at 150 as hybridizing completely with a high binding strength. This is because the target DNA is itself represented in the database in this case, so the candidate probe is seen at 150 to hybridize with itself (a perfect hybridization). The locus of each hybridization from the preparation 132 are displayed in column 141, while the starting position of each hybridization is given in column 142. The calculated hybridizations are shown at 145.

iii. The ProbesEdit Window

The ProbesEdit Window, Figure 16, is a text editing window provided for convenient editing and annotation of OligoProbe DesignStation text file output. It is also used to accumulate probes selected from the MPSD, Figure 14, by mouse button 2 clicks. Standard text editing capabilities are available within the ProbesEdit Window. The user may accumulate selected probes in this window (see 155 for an example) and then save them to a file (which will bear the name of the preparation sequence with the file extension of "prb" 156, or may be another file name selected by the user). A sample of this file is shown in Figures 16A and 16B.

iv. Miscellaneous Output

The present embodiment of this invention also creates two output files, currently named "test.out" and "test1.out", depending upon which model the user has selected. The first file, "test.out", is created with both the Mismatch Model and the H-Site Model. This file is a textual representation of the Mitsuhashi Probe Selection Diagram (MPSD). It breaks the probe sequence down by position, length, delta Tm, screensN, and the actual probe sequence (i.e., nucleotides). An example of this file created by the Mismatch Model is shown in Figure 30, and example created by the H-Site Model is shown in Figure 34a. The second file, "test1.out", is created only by the H-Site Model. This file is a textual representation of the ProbeInfo and MatchInfo window that captures all hybridizations, along with their locus, starting position, melting temperature, and possible other hybridizations. A partial example of this file is shown in Figure 34b (10 pages out of a total of 190 pages created by the H-Site Model).

2. Description of the Mismatch Model Program

a. Overview

In this invention, one of the hybridization strength models is termed the Mismatch Model (see Figure 12 for selection of this model). The basic operation of this model involves the techniques of hashing and continuous seed filtration, as defined earlier but described in more detail below. The essence of the Mismatch Model is a fast process for doing exact and inexact matching between nucleotide sequences to support the Mitsuhashi Probe Selection Diagram (MPSD). There are a number of modules in the present implementation of the Mismatch Model contained in this invention, the most

-60-

significant of which are shown in the flow chart in Figure 17 and in more detail in Figures 18 through 28. The main k_diff module shown in the flow chart in Figure 18 is a structured program that provides overall control of the Mismatch Model, calling various submodules that perform different functions.

b. **Inputs**

5 The user-selected input variables for this model are minimum probe length 76 (which is generally from 18 to 30) and maximum number of mismatches 77 (which generally is from 1 to 5). These inputs are entered by the user in the Main OligoProbe DesignStation Dialog Window, Figure 12C.

c. **Processing**

10 i. k_diff Program

Some terms of art need to be defined before the processing performed by this module can be explained. A hash table basically is an array or table of data. A linked list is a classical data structure which is a chain of linked entries and involves pointers to other entry structures. Entries in a linked list do not have to be stored sequentially in memory, as is the case with elements contained in an array. 15 Usually there is a pointer to the list associated with the list, which is often initially set to point to the start of the list. A pointer to a list is useful for sequencing through the entries in the list. A null pointer (i.e., a pointer with a value of zero) is used to mark the end of the list.

As the flow charts in Figures 17 and 18 illustrate, the general process steps and implemented functions of this model can be outlined as follows:

20 Step 1: First, create a hash table and linked list from the query (Figure 17, hashing module 222).

Step 2: Next, while there are still GenBank entries available for searching (Figure 17, assembly module 230):

25 Step 2a: Read the current GenBank entry (record) sequence of user-specified length (Figure 17, seqload module 232), or read the current sequence (record) from the file selected by the user (Figure 17, read1 module 234).

Step 2b: For the current sequence for each position of the sequence from the first position (or nucleotide) to the last position (or nucleotide) (incrementing the position number once each iteration of the loop) (Figure 17, q_colour module 242),

30 Step 2c: set the variable dna_hash equal to the hash of the current position of the current sequence (Figure 17, q_colour module 242).

Step 2d: While not at the end of the linked list for dna_hash (Figure 17, q_colour module 242),

35 Step 2e: set the query_pos equal to the current position of dna_hash in the linked list (Figure 17, q_colour module 242) and

-61-

Step 2f: Extend the hit with the coordinates (query_pos, dna_pos) (Figure 17, hit_ext module 244),

Step 2g: If there exists a `k_mismatch` in the current extended hit (Figure 17, colour module 246), then

5 Step 2h: print the current hit (Figure 17, q_colour module 242), and repeat from Step 2.

As this illustrates, there are three (3) basic looping or iteration processes with functions being performed based on variables such as whether the GenBank section end has been reached (the first "WHILE" loop, Step 2), whether the end of the current DNA entry has been reached (the "FOR" loop, Step 2b), and whether the end of the dna_hash linked list has been reached (the second "WHILE" loop, Step 2d). A "hit" will only be printed if there are k mismatches in the current extended hit.

Figures 18 through 28 illustrate the functions of each of the modules of the present embodiment of this invention, all of which were generalized and summarized in the description above. Figure 18, which outlines the main "k_diff" module, shows that this module is primarily a program organization and direction module, in addition to performing routine "housekeeping" functions, such as defining the variables and hash tables 251, checking if the user-selected gene sequence file is open 252, extracting needed identification information from the GenBank 253, and ensuring valid user input 254. This module also performs a one-time allocation of memory for the gene sequences, and allocates memory for hit information, hashing, hybridization and frequency length profiles and output displays, 255 & 256. The "k_diff" module also initializes or "zeros out" the hashing table, the linked hashing list and the various other variables 257 in preparation for the hashing function. In addition, this module forms the hash tables 258 and extracts a sequence and finds the sequence length 259.

One of the most important functions performed by the "k_diff" module is to define the seed (or kernel or k_tuple) size. This is done by setting the variable k_tuple equal to $(\text{min_probe_length} - \text{max_mismatch_}\#) / (\text{max_mismatch_}\# + 1)$ Figure 18 at 265. Next, if the remainder of the aforementioned process is not equal to zero 266, then the value of the variable k_tuple is incremented by one 267. The resulting value is the size of the seed. The module then reads the query 268 and copies the LOCUS name 269 for identification purposes (a definition of the term locus is given earlier in the specification).

30 The "k_diff" module Figure 18 also calls the "assembly" module 260, writes the results to a file 261a, plots the results 261b (discussed below), calculates the hairpin characteristics 262 (i.e., the number of base pairs and the length of the worst hairpin) and the melting temperature (T_m) for each candidate probe 263, and saves the results to a file 264.

35 The screen graphs are plotted 261b by converting the result values to pixels, filling a pixel array and performing a binary search into the pixel array. Next, given the number of pixels per probe position and which function is of interest to the user (i.e., the three mismatch match numbers), the program

-62-

interpolates the values at the value of (pixelsPerPositionN-1) and computes the array of pixel values for drawing the graph. These values are then plotted on the MPSD.

5 The "hashing" module, Figure 19, performs hashing of the query. In other words, it creates the hash table and linked list of query positions with the same hash. The variable `hash_table[i]` equals the position of the first occurrence of hash `i` in the query. If `i` does not appear in the query, `hash_table[i]` is set to zero.

10 The "tran" module, Figure 20, is called by the "hashing" module 271, and performs the hashing of the sequence of `k_tuple` (kernel or seed) size. If the `k_tuple` exists (i.e., its length is greater than zero), the variable `uns` is set equal to `uns*ALF+p` 291. The variable `p` represents the digit returned by the "let_dig" module Figure 21 that represents the nucleotide being examined. `ALF` is a constant that is set by the program in this implementation to equal four. The query pointer is then incremented, while the size of `k_tuple` (the seed) is decremented 292. This process is repeated until the sequence of `k_tuple` has been entirely hashed. Then the "tran" module returns the variable `current_hash` 293 to the "hashing" module Figure 19.

15 The "let_dig" module, Figure 21, is called by the "tran" module 291, and transforms the nucleotides represented as the characters "A", "T", "U", "G" and "C" in the GenBank and the user's query into numeric digits for easier processing by the program. This module transforms "a" and "A" into "0" 301, "t", "T", "u" and "U" into "1" 302, "g" and "G" into "2" 303, and "c" and "C" into "3" 305. If the character to be transformed does not match any one of those listed above, the module returns "-1" 305. The "hashing" module, Figure 19, then calls the "update" module 272, Figure 22, which updates the hash with a sliding window (i.e., it forms a new hash after shifting the old hash by "1"). The remainder of `old_hash` divided by `power_1` is calculated 311 (a modulus operation), the remainder is multiplied by `ALF` 312 (i.e., four), and then the digit representing the nucleotide is added to the result 313. The "update" module then returns the result 314 to the "hashing" module Figure 19.

25 If the current hash has already occurred in the query, the program searches for the end of the linked list for the current hash 273 and marks the end of the linked list for the current hash 274. If the current hash has not already occurred in the query, the program puts the hash into the hash table 275. The resulting hash table and linked list are then returned to the "k_diff" module, Figure 18 at 258.

30 The "assembly" module, Figure 23, extracts sequences from the GenBank and performs hit locating and extending functions. This module is called by the "k_diff" module Figure 18 at 260 if the user has chosen to use the database to locate matches. The output from the "assembly" module (Figure 23) tells the user that the section of the database searched contains `E` number of entries 321 of `S` summary length 322 with `H` number of hits 323. Further, the program tells the user that the number of considered `l`-tuples equals `T` 324. The entry head line is also printed 326.

35 The "seqload" module, Figure 24, is called by the "k_diff" module Figure 18 at 259 once the query hash table and linked list have been formed by the "hashing" module Figure 19. The "seqload" module Figure 24 checks to see if the end of the GenBank file has been reached 327, and, if not,

-63-

searches until a record is found with LOCUS in the head-line 328. Next, the LOCUS name is extracted 329 for identification purposes, and the program searches for the ORIGIN field in the record 330.

The program then extracts the current sequence 331 from the GenBank and performs two passes on each sequence. The first is to determine the sequence length 332 and allocate memory for each sequence 333, and the second pass is to read the sequence into the allocated memory 334. Since the sequences being extracted can contain either DNA nucleotides or protein nucleotides, the "seqload" module can recognize the characters "A", "T", "U", "G", and "C". The bases "A", "T", "G" and "C" are used in DNA sequences, while the bases "A", "U", "G" and "C" are used in RNA and mRNA sequences. The extracted sequence is then positioned according to the type of nucleotides contained in the sequence 335, and the process is repeated. Once the end of the sequence has been reached, the "seqload" module returns the sequence length 336 to the "k_diff" module Figure 18.

If the user has chosen to use one or more files to locate matches, rather than the database, the "read1" module, Figure 25, rather than the "seqload" module Figure 24, is called by the "k_diff" module Figure 18. The "read1" module, Figure 25, reads the sequence from the user specified query file 341 and allocates memory 342. This module also determines the query length 343, extracts sequence identification information 344, determines the sequence length 345, transforms each nucleotide into a digit 346 by calling the "let_dig" module Figure 21, creates the query hash table 347 by calling the "dig_let" module Figure 26, and closes the file 348 once everything has been read in.

First, the "read1" module Figure 25 allocates space for the query 342. To do this, the "ckalloc" module, Figure 25 at 342, is called. This module allocates space and checks whether this allocation is successful (i.e., is there enough memory or has the program run out of memory). After allocating space, the "read1" module Figure 25 opens the user-specified file 349 (the "ckopen" module, Figure 25 at 349, is called to ensure that the query file can be successfully opened 349), determines the query length 343, locates a record with LOCUS in the head-line and extracts the LOCUS name 344 for identification purposes, locates the ORIGIN field in the record and then reads the query sequence from the file 341. Next, the sequence length is determined 345, memory is allocated for the sequence 342, and the sequence is read into the query file 350. If the string has previously been found, processing is returned to 344. If not, then each character in the query file is read into memory 350.

The characters are transformed into digits 346 using the "let_dig" module, Figure 21, until a valid digit has been found, and then the hash table containing the query is set up 347 using the module "dig_let", Figure 26, which transforms the digits into nucleotides represented by the characters "A" 371, "T" 371, "G" 373, "C" 374, and "X" 375 as a default. If the end of the file has not been reached, processing is returned to 344. If it has, the file is closed 348 and the query is then returned to the "read1" module Figure 25 at 347.

The "q_colour" module, Figure 17 (Figure 23 at 325), is called by the "assembly" module Figure 23 after the current sequence has been extracted from the GenBank. The "q_colour" module Figure 27 performs the heart of the Mismatch Model process in that it performs the comparison between the

-64-

query and the database or file sequences. If the module finds that there exists a long (i.e., greater than the min_hit_length) extended hit, it returns a "1" to the "assembly" module Figure 24. Otherwise, the "q_colour" module, Figure 27, returns a "0".

5 In the "q_colour" module, Figure 27, all DNA positions are analyzed in the following manner. First, the entire DNA sequence is analyzed 391 to see whether each position is equal to zero 392 (i.e., whether it is empty or the sequence is finished). If it is not equal to zero 393, the "q_colour" module Figure 30 calls the "tran" module, Figure 20 described above, which performs the hashing of k_tuples. The "tran" module Figure 20 calls other modules which transform the nucleotides represented by characters into digits for easier processing by the program and then updates the hash with a sliding window. If the position is equal to zero, the current_hash position is set to new_has after one shift of old_hash 390 by calling the "update" module Figure 22.

10 If the nucleotide at the current_hash position is equal to zero, processing is returned to 391. If not, the query position is set equal to (nucleotide at current hash position - 1). Next, the "q_colour" module Figure 27 looks for the current_hash in the hash table 394. If the current k_tuple does not match the query 395, then the next k_tuple is considered 395, and processing is returned to 391. If the current k_tuple does match the query, then the program checks the hit's (i.e., the match's) vicinity 396 by calling the "hit_ext" module, Figure 28 to determine if the hit is weak. The inventors have found that if the code for the module "hit_ext" is included within the module "q_colour", rather than being a separate module utilizing the parameter transfer machinery, 25% of CPU time can be saved.

20 The "hit_ext" module Figure 28 determines the current query position in the hit's vicinity 421, determines the current DNA position in the hit's vicinity 422, and creates the list of mismatch positions (i.e., the mismatch_location_ahead 423, the mismatch_location_behind 423 and the kernel match location). If the hit is weak 424, the "hit_ext" module Figure 28 returns "0" to the "q_colour" module Figure 27. If the hit has a chance to contain 425, the module returns "1" to the "q_colour" module Figure 27. A hit has a chance to contain, and is therefore not considered weak, if the mismatch_location_ahead - the mismatch_location_behind is greater than the min_hit_length. If not, it is a short hit and is too weak.

25 If the "hit_ext" module Figure 28 tells the "q_colour" module Figure 27 that the hit was not a weak one, then the "q_colour" module determines whether the current hit is long enough 398 by calling the "colour" module Figure 29. The "colour" module Figure 29 performs query_colour modification by the hit data, starting at pos_query and described by mismatch_location_ahead and mismatch_location_behind. After the variables to be used in this module are defined, variable isw_print (which is the switch indicating the hit length) is initialized to zero 430. The cur_length is then set equal to the length of the extending hit 431 (mismatch_location_behind[i] + mismatch_location_ahead[j]-1). 30 Next, if cur_length is greater than or equal to the min_hit_length 432 (i.e., the minimum considered probe size), the hit is considered long and isw_print is set equal to two 433. The value of isw_print is then returned 434 to the "q_colour" module Figure 27. 35

-65-

If the length of the extending hit is longer than the min_hit_length, the hit is considered long 399. Otherwise, the hit is considered short. If the hit is short, nothing more is done to the current hit and the module begins again. If, on the other hand, the hit is considered long 399, the "q_colour" module Figure 27 prints the current extended hit 400. The current extended hit can be printed in ASCII, printed in a binary file, or printed to a memory file. The "q_colour" module Figure 27 then repeats until the end of the linked list is reached.

d. Outputs

The output of the k_diff program may be either a binary file containing the number of extended hits and the k_mismatch hit locations (see Figure 30), or the output may be kept in memory without writing it to a file. See Section 1(d)(iv) for more detail.

3. Description of the H-Site Model Program

a. Overview

In this invention, the second hybridization strength model is termed the H-Site Model (see Figure 12 for user selection of this model). The formula used in the H-Site Model is an expression of the fact that melting temperature T_m is a function of both probe length and percent of GC content. This basic formula has been modified in this invention to account for the presence of mismatches. Each percent of mismatch reduces the melting temperature T_m by an average of 1.25 degrees (2 degrees C for an AT mismatch, and 4 degrees C for a GC mismatch).

In addition, this implementation of the invention does some preliminary preprocessing of the GenBank database to sort out and select the cDNA sequences. This is done by locating a keyword (in this case CDS) in each GenBank record.

There are a number of modules in the present embodiment of the H-Site Model contained in this invention. Each step of the processing involved in the H-Site Model is more fully explained below, and is accompanied by detailed flow charts.

b. Inputs

There are two basic user-selected inputs for the H-Site Model (see Figure 12C): 1) the melting temperature T_m 22 for which probes are being designed (i.e., the melting temperature that corresponds to a particular experiment or condition the user desires to simulate); and 2) the nucleation threshold 23, which is the number of base pairs constituting a nucleation site. The user is also required to select the 1) target species 11 gene sequence(s) (DNA, mRNA or cDNA) for which probes are being designed; 2) the preparation 12 of all sequences with which hybridizations are to be calculated; and 3) the probe output file 13. The preparation file is the most important, as discussed below.

c. Organization of the H-Site Model Program

The current implementation of the H-Site Model program of this invention is distributed between five files containing numerous modules. The main file is designated by the inventors as "ds.cpp" in its uncompiled version. This file provides overall control to the entire OligoProbe DesignStation

-66-

invention. It is divided into six sections. Section 0 defines and manipulates global variables. Section 1 controls general variable definition and initialization (including the arrays and memory blocks). It also reads and writes buffers for user input selections, and constructs multi buffers.

5 Section 2 sets up and initializes various "snippet" variables (see section below for a complete definition of the term snippet), converts base pair characters to a representation that is 96 base pairs long and to ASCII base pair strings, and performs other sequence file manipulation such as comparing snippets. This section also reads the sequence format file, reads base pairs, checks for and extracts sequence identification information (such as ORIGIN and LOCUS) and filters out sequences beginning with numbers.

10 Section 3 involves preparation file manipulation. This section performs the preprocessing on the PRP file discussed above. It also merges and sorts the snippet files, creates a PRP file and sorts it, and outputs the sorted snippets. Next, this section streams through the PRP file.

15 Section 4 contains the essential code for H-Site Model processing (see Figures 31 through 33 for details, discussed below). Streams are set up, and then RIBI comparisons are performed for hybridizations (see file "ribi.cpp" for definitions of RIBI search techniques). Next, probes are generated, binding strength is converted to melting temperature, and hybridizations are calculated and stored (including hybridization strength). Lastly, other H-Site calculations are performed.

20 Section 5 is concerned with formatting and presenting diagnostic and user file (test.out, test1.out, and test2.out files) output. This section also handles the graphing functions (the MPSD diagram in particular). In addition, this section calculates the hairpin characteristics for the H-Site Model candidate probes.

25 The second H-Site Model file, designated as "ds.h" defines data variables and structures. Section 1 of this file concerns generic data structures (including memory blocks and arrays, and file inputs and outputs). Section 2 defines the variables and structures used with sequences, probes and hybridizations. Section 3 defines variables and structures concerned with protocols (i.e., function prototypes, graphing, etc.).

The third H-Site Model file, designated as "funedoc.txt", contains very detailed documentation for this implementation of the H-Site Model program. Numerous variables and structures are also defined. The flow of the program is clearly shown in this file.

30 The fourth H-Site Model file, designated as "ribi.h" handles the sequence comparisons. The fifth and last H-Site Model file, designated as "ribi.cpp", performs internal B-Tree indexing. Definitions of Red-black Internal Binary Index (RIBI) searching are found in this file. Definitions are also included for the concepts keyed set, index, binary tree, internal binary index, paths, and red-black trees. Implementation notes are also included in this file.

d. **Processing**

Implementation of the H-Site Model in this invention is done in three stages. First, the invention creates the preparation (PRP) file, which contains all relevant information from the sequence database. This is the preprocessing stage discussed above. Next, the target is prepared by the program. Lastly, the invention calculates the MPSD data using the PRP file and target sequence to find probes.

i. Creation of the Preprocessed Preparation File

Figure 31. Step 1: The program first opens the sequence database for reading into memory 461, 462. Step 2: Next, as sequence base pairs are read in 462, "snippets" are saved to disk 463, along with loci information. A snippet is a fixed-length subsequence of a preparation sequence. The purpose of snippets is to allow the user to examine a small portion of a preparation sequence together with its surrounding base pairs. Snippets in the implementation of this invention are 96 base pairs long (except for snippets near the end or beginning of a sequence, which may have fewer base pairs). The "origin" of the snippet is in position 40. For snippets taken near the beginning of a sequence, some of the initial 40 bases are undefined. For snippets near the end of a sequence, some of the final 55 bases are undefined. Snippets are arranged in the preparation file (PRP) in sorted order (lexicographical order beginning at position 40). In this invention, the term "lexicographical order" means a preselected order, such as alphabetical, numeric or alphanumeric. In order to conserve space, snippets are only taken at every 4th position of the preparation sequence.

Step 3: The snippets are merge sorted 464 to be able to search quickly for sequences which pass the "screen", discussed below. Step 4: The merged file is prepended with identifiers for the sources of the snippets 465. This is done to identify the loci from which hybridizations arise.

ii. Target Preparation

Figure 32. Step 1: The target sequence file is opened 471 and read into memory 472. For each position in the target mRNA, the probe defined at that starting position is the shortest subsequence starting at that position whose hybridization strength is greater than the user specified melting temperature T_m . Typically, the probes are of length 18 to 50. Step 2: Four lists of "screens" are formed 473, 474, 475, each shifted by one base pair 475 to correspond to the fact that snippets are only taken at every four base pairs. A screen is a subsequence of the target mRNA of length equal to the screening threshold specified by the user. The screens are then indexed 476 and sorted in memory 477.

iii. Calculation of the MPSD Data

Figure 33. Step 3: This step is the heart of the process. Step 3a: The program streams through the following five items in sync, examining them in sequential order: the snippet file and the four lists of screens 481-484. Step 3b: Each snippet is compared to a screen 485. Step 3c: If the snippet does not match, whichever stream is behind is advanced 486 and Step 3b is repeated. If the snippet does match, Step 4 is performed.

-68-

Step 4: If a snippet and a matching screen were found in Step 3b 487, the hybridization strength of the binding between the sequence containing the snippet and all of the probes containing the screen is calculated (see Step 5). Double counting is avoided by doing this only for the first matched screen containing the probe. Each pair of bases is examined and assigned a numerical binding strength. An AT pair would be assigned a lower binding strength than a GC pair because AT pairs have a lower melting temperature T_m . The process is explained more fully below at Step 5b.

Step 5: The hybridization strengths between sequence and all the probes containing it are calculated using a dynamic programming process. The process is as follows: Step 5a: Begin at the position of the first probe containing the given screen but not containing any other screens which start at an earlier position and also match the sequence. This is done to avoid double counting. Two running totals are maintained: a) boundStrength, which represents the hybridization strength contribution which would result if the sequence and probe were to match exactly for all base pairs to the right of the current position, and b) unboundStrength, which represents the strength of the maximally binding region. Step 5b: At each new base pair, the variable boundStrength is incremented by 71 if the sequence and probe match and the matched base pair is GC 489, incremented by 30 if the matched base pair is AT 490 (i.e., this number is about 42.25% of the first number 71), and decremented by 74.5 if there is not a match 488 (i.e., this number is about 5% larger than the first number 71). Step 5c: If the current boundStrength exceeds the current unboundStrength 491 (which was originally initialized to zero), a new binding region has been found, and unboundStrength is set equal to boundStrength 492. Step 5d: If the current boundStrength is negative, boundStrength is reset to zero 493. Step 5e: If the current position is at the end of a probe, the results (the hybridization strengths) are tallied for that probe. Step 5f: If the current position is at the end of the last probe containing the screen, the process stops.

Step 6: A tally is kept of the number and melting temperature of the matches for each candidate probe, and the location of the best 20 candidates, using a priority queue (reverse order by hybridization strength number) 494. Step 7: A numerical "score" is kept for each preparation sequence by tallying the quantity $\exp$ (which can be expressed as $\sum e^{-T_m}$) for each match 495, where T_m is the melting temperature for the "perfect" match, the probe itself. In other words, the probe hybridizes "perfectly" to its target.

Step 8: Hairpins are calculated by first calculating the complementary probe. In other words, the order of the bases in the candidate probe are reversed (CTATAG to GATATC), and complementary base pairs are substituted (A for T, T for A, G for C, and C for G, changing GATATC to CTATAG in the above example). Next, the variable representing the maximum hairpin length for a candidate probe is initialized to zero, as is the variable representing a hairpin's distance. For each offset, the original candidate probe and the complementary probe just created are then aligned with each other and compared. The longest match is then found. If any two matches have the same length, the one with the longest hairpin distance (i.e., the number of base pairs separating the match) is then saved.

-69-

Step 9: The preparation sequences are then sorted 496 and displayed in rank order, from best to worst 497. Step 10: The resulting MPSD, which includes all candidate probes, is then displayed on the screen. Step 11: The best 20 matches are also printed or displayed in rank order, as the user requests 497.

5 e. Outputs

The outputs of the H-Site Model are fully described in Section 1(d)(iv), above, and illustrated in Figures 14 through 16. Samples of the two output files created by the H-Site Model are shown in Figures 34A and 34B.

4. Description of the Mitsuhashi Probe Selection Diagram Processing

10 Once the Mitsuhashi Probe Selection Diagram (MPSD) data has been calculated by the H-Site Model program (see stage three and Figure 33, discussed above), it is necessary to convert this data to pixel format and plot a graph. An overview of this process is shown in Figure 35. First, the program calculates the output (x,y) ranges 500. Next, these are converted to a logarithmic scale 501. The values are then interpolated 502, and a bitmap is created 503. Lastly, the bitmap is displayed on the screen 504 in MPSD format (discussed above in section 1(e)(i)). A sample MPSD is shown in Figure 14.

5. Description of the MatchInfo Window Processing

20 The ProbeInfo and MatchInfo windows are discussed in great detail in Section 1(e)(ii), and a sample of these windows is shown in Figure 15. An overview of the processing involved in creating the MatchInfo portion of the window is given in the flow chart in Figure 36. First, as the user moves the MPSD cursor 570 (seen as a vertical line bisecting the MPSD window), the program updates the position of the candidate probe shown under that cursor position 521. Next, based upon the candidate probe's position, the program updates the sequence 522 and hairpin information 523 for that probe. This updated information is then displayed in an updated match list 524, shown in the MatchInfo window.

25 **XVII. Detection Kits**

The present invention can preferably be embodied in a kit for the detection of an organism, infectious agent, or biological component contained in a biological sample. Such a kit can take a variety of forms, as will be apparent to those of skill in the art.

30 In one embodiment, the present invention comprises a kit for identifying the presence of a particular species of fungus in a biological sample. Such a kit includes at least one specific polynucleotide probe and a common probe, as described above. In a preferred embodiment, a plurality of specific probes are included, and such probes are preferably immobilized on one or more solid supports. In a more preferred embodiment, each of the plurality of specific probes is immobilized on a different solid support. For example, a microtiter plate having a plurality of wells can have a different polynucleotide probe immobilized to each well. If such probes contain sequences specific to the

35

-70-

ribosomal RNA of different species of fungi, the kit can be used to test a single biological sample for the presence of a plurality of fungi.

5 An example of such a kit using microtiter wells as the solid supports is shown schematically in Figure 6. In this figure, the presence of the ribosomal RNA of a particular species of fungus, in this case *Candida albicans*, is detected in a well 56, which is darkened to show a positive result. No probe has been immobilized to well 52 in order to provide a negative control. Well 54, on the other hand, has
10 immobilized to its walls a common probe, such as SEQ ID NO:1, which is complementary to a sequence present in the ribosomal RNA of the fungi being tested for. This provides a positive control, since well 54 should show a positive result whenever a positive result is detected in well 56 or any of the other wells in this embodiment of the kit.

Well 54 also provides a means of detecting the presence of fungi which do not contain the specific ribosomal RNA sequences being probed by the specific polynucleotide probes. For example, if a mutant strain of *Candida albicans* which does not contain the specific sequence complementary to the probe used in well 56 is present in a sample tested with the kit illustrated in Figure 6, well 56 would
15 not show a positive result. Well 54 would, however, indicate the presence of a fungal pathogen, as long as the mutant strain did not contain a mutation in the common sequence detected in well 54 which interfered with the hybridization of that sequence to the common probe immobilized to the walls of well 54.

In addition to specific probes, common probes, and solid supports, other elements can also be included in the present kit. For example, appropriate buffers for hybridizing DNA or RNA to the probes in the kit can be included. Labels, as described above, can also be incorporated which are attached to a common probe.
20

In an alternative embodiment, the kit can include PCR primers such as those previously described. Such a kit could comprise a common primer and a specific primer, 2 common primers, or
25 2 specific primers identified through the method of the present invention. Other components, such as a reverse transcriptase, a DNA polymerase like Taq polymerase and dNTP's can also be included in this embodiment of the present kit.

The forgoing embodiments of the kit of the present invention can be adapted to perform the methods of the present invention that involve PCR as well. In this embodiment, the kit additionally
30 includes a reverse transcriptase and a polymerase, preferably a DNA polymerase that has significant polymerase activity at temperatures above 50°C, such as Taq DNA polymerase.

XVIII. Conclusion

All references cited herein are hereby explicitly incorporated by this reference thereto. Although the invention has been described with reference to certain particular exemplary embodiments
35 of various aspects, these embodiments are intended only to illustrate and not to limit the present

-71-

invention. Accordingly, the scope of the present invention is to be determined upon reference to the appended claims.

-72-

Table I. Common probe (Com-392)* for Fungi

Species	GenBank name	Sequence I.D. No.	Com-392 GAGGGAGCCTGAGAAACG
I. Fungi			
Pneumocystis carinii	PMC16SRR1	1	-----
Cryptococcus neoformans	CPCDA	2	-----
Coccidioides immitis	COIDA	3	-----
Blastomyces dermatitidis	BLODA	4	-----
Aspergillus fumigatus	ASNDA	5	-----
fumigatus	ASNRR5SS	6	-----
fumigatus	ASNRRSSB	7	-----
Candida albicans	YSASRSUA	8	-----
albicans	YSAL16S	9	-----
lusitaniae	YSASRRNAA	10	-----
lusitaniae	YSASRSUE	11	-----
kefyr	YSASRSUB	12	-----
krusei	YSASRRNAC	13	-----
krusei	YSASRSUD	14	-----
tropicalis	YSASRRNAB	15	-----
tropicalis	YSASRSUG	16	-----
viswanathii	YSASRSUH	17	-----
parapsilosis	YSASRSUF	18	-----
guilliermondii	YSASRSUC	19	-----
glabrata	YS5CRRNAS	20	-----

* Com-392 is identical among 107 different rRNAs registered in GenBank

Table II. Common probe (Com-419)* for Fungi

Species	GenBank name	Sequence I.D. No.	Com-419 TCCAAGGAAGGCAGCAGG
I. Fungi			
Pneumocystis carinii	PMC16SRR1	21	-----
Cryptococcus neoformans	CPCDA	22	-----
Coccidioides immitis	COIDA	23	-----
Blastomyces dermatitidis	BLODA	24	-----
Aspergillus fumigatus	ASNDA	25	-----
fumigatus	ASNRR5SS	26	-----
fumigatus	ASNRRSSB	27	-----
Candida albicans	YSASRSUA	28	-----
albicans	YSAL16S	29	-----
lusitaniae	YSASRRNAA	30	-----
lusitaniae	YSASRSUE	31	-----
kefyr	YSASRSUB	32	-----
krusei	YSASRRNAC	33	-----
krusei	YSASRSUD	34	-----
tropicalis	YSASRRNAB	35	-----
tropicalis	YSASRSUG	36	-----
viswanathii	YSASRSUH	37	-----
parapsilosis	YSASRSUF	38	-----
guilliermondii	YSASRSUC	39	-----
glabrata	YS5CRRNAS	40	-----

* Com-419 is identical among 123 different rRNAs registered in GenBank

SUBSTITUTE SHEET

-73-

Table III. Common probe (Com-1205)* for Fungi

Species	GenBank name	Sequence I.D. No.	Com-1205 ACGGGGAACTCACCAGG
I. Fungi			
<i>Pneumocystis carinii</i>	PMC16SRR1	41	-----
<i>Cryptococcus neoformans</i>	CPCDA	42	-----
<i>Coccidioides immitis</i>	COIDA	43	-----
<i>Blastomyces dermatitidis</i>	BLODA	44	-----
<i>Aspergillus fumigatus</i>	ASNDA	45	-----
<i> fumigatus</i>	ASNRR5SS	46	-----
<i> fumigatus</i>	ASNRRSSB	47	-----
<i>Candida albicans</i>	YSASRSUA	48	-----
<i> albicans</i>	YSAL16S	49	-----
<i> lusitaniae</i>	YSASRRNAA	50	-----
<i> lusitaniae</i>	YSASRSUE	51	-----
<i> kefyr</i>	YSASRSUB	52	-----
<i> krusei</i>	YSASRRNAC	53	-----
<i> krusei</i>	YSASRSUD	54	-----
<i> tropicalis</i>	YSASRRNAB	55	-----
<i> tropicalis</i>	YSASRSUG	56	-----
<i> viswanathii</i>	YSASRSUH	57	-----
<i> parapsilosis</i>	YSASRSUF	58	-----
<i> guilliermondii</i>	YSASRSUC	59	-----
<i> glabrata</i>	YS5CRRNAS	60	-----

* Com-1205 is identical among 42 different rRNAs registered in GenBank

Table IV. Common probe (Com-1544)* for Fungi

Species	GenBank name	Sequence I.D. No.	Com-1544 TCGTGCTGGGGATAGAGC
I. Fungi			
<i>Pneumocystis carinii</i>	PMC16SRR1	61	-----
<i>Cryptococcus neoformans</i>	CPCDA	62	-----
<i>Coccidioides immitis</i>	COIDA	63	-----
<i>Blastomyces dermatitidis</i>	BLODA	64	-----
<i>Aspergillus fumigatus</i>	ASNDA	65	-----
<i> fumigatus</i>	ASNRR5SS	66	-----
<i> fumigatus</i>	ASNRRSSB	67	-----
<i>Candida albicans</i>	YSASRSUA	68	-----
<i> albicans</i>	YSAL16S	69	-----
<i> lusitaniae</i>	YSASRRNAA	70	-----
<i> lusitaniae</i>	YSASRSUE	71	-----
<i> kefyr</i>	YSASRSUB	72	-----
<i> krusei</i>	YSASRRNAC	73	-----
<i> krusei</i>	YSASRSUD	74	-----
<i> tropicalis</i>	YSASRRNAB	75	-----
<i> tropicalis</i>	YSASRSUG	76	-----
<i> viswanathii</i>	YSASRSUH	77	-----
<i> parapsilosis</i>	YSASRSUF	78	-----
<i> guilliermondii</i>	YSASRSUC	79	-----
<i> glabrata</i>	YS5CRRNAS	80	-----

* Com-1544 is identical among 40 different rRNAs registered in GenBank

SUBSTITUTE SHEET

-74-

Table V. Probes for *Pneumocystis carinii* (Cari-685)

Species	GenBank name	Sequence I.D. No.	Cari-685 GCGCAACTGATCCTTCCC
I. Fungi			
<i>Pneumocystis carinii</i>	PMC16SRR1	81	-----
<i>Cryptococcus neoformans</i>	CPCDA	82	-T--CGGC---G----AT
<i>Coccidioides immitis</i>	COIDA	83	A-CTGGT-----G--A
<i>Blastomyces dermatitidis</i>	BLODA	84	A-CTGGT-----G--A
<i>Aspergillus fumigatus</i>	ASNDA	85	A-CTGGT-----G--A
<i> fumigatus</i>	ASNRR5SS	86	----CGGC---GG---AT
<i> fumigatus</i>	ASNRRSSB	87	----CGGC---GG---AT
<i>Candida albicans</i>	YSASRSUA	88	-----G-C---AGCTTG
<i> albicans</i>	YSAL16S	89	-----G-C---AGCTTG
<i> lusitaniae</i>	YSASRRNAA	90	-----G-C---AGCTTG
<i> lusitaniae</i>	YSASRSUE	91	-----G-C---AGCTTG
<i> kefyr</i>	YSASRSUB	92	-----G-C---AGCTTG
<i> krusei</i>	YSASRRNAC	93	-----G-C---AGCTTG
<i> krusei</i>	YSASRSUD	94	-----G-C---AGCTTG
<i> tropicalis</i>	YSASRRNAB	95	-----G-C---AGCTTG
<i> tropicalis</i>	YSASRSUG	96	-----G-C---AGCTTG
<i> viswanathii</i>	YSASRSUH	97	-----G-C---AGCTTG
<i> parapsilosis</i>	YSASRSUF	98	-----G-C---AGCTTG
<i> guilliermondii</i>	YSASRSUC	99	-----G-C---AGCTTG
<i> glabrata</i>	YS5CRRNAS	100	-----G-C---AGCTTG
II. Highest homologous sequence in GenBank			
Human TAN-1	HUMTAN1	101	-----A-C-----
<i>Abccelsyn xylium</i>	ABCCELSYN	102	-A-G-----
Human mRNA neuron	HUMNSEMRNA	103	CTC-C-----

-75-

Table VI. Probes for *Pneumocystis Carinii* (Cari-1056)

Species	GenBank name	Sequence I.D. No.	Cari-1056 GGCGATGTTTTTCTTGACTCG
I. Fungi			
<i>Pneumocystis carinii</i>	PMC16SRR1	104	-----
<i>Cryptococcus neoformans</i>	CPCDA	105	-----C--CA---AAATA--T
<i>Coccidioides immitis</i>	COIDA	106	-----G--CA---AAATT--T
<i>Blastomyces dermatitidis</i>	BLODA	107	-----G--CA---AAATT--T
<i>Aspergillus fumigatus</i>	ASNDA	108	----G-----C-A-GA----C--
<i> fumigatus</i>	ASNRR5SS	109	----G-----C-A-GA----C--
<i> fumigatus</i>	ASNRRSSB	110	----G-----C-A-GA----C--
<i>Candida albicans</i>	YSASRSUA	111	-TT-T----C---A-----G-A
<i> albicans</i>	YSAL16S	112	-TT-T----C---A-----G-A
<i> lusitaniae</i>	YSASRRNAA	113	----GC---CA---AG---G--
<i> lusitaniae</i>	YSASRSUE	114	----GC---CA---AG---G--
<i> kefyr</i>	YSASRSUB	115	-----G--CA---AAATTTCT
<i> krusei</i>	YSASRRNAC	116	---CG-T---AG-C---GAGTG
<i> krusei</i>	YSASRSUD	117	TATTT---NG-----A-GACCA
<i> tropicalis</i>	YSASRRNAB	118	-TT-T----C---A-----G-A
<i> tropicalis</i>	YSASRSUG	119	-TT-T----C---A-----G-A
<i> viswanathii</i>	YSASRSUH	120	-TT-T----C---A-----G-A
<i> parapsilosis</i>	YSASRSUF	121	-TT-T----C---A-----G-A
<i> guilliemondii</i>	YSASRSUC	122	--T-T----C-K--T-----G-A
<i> glabrata</i>	YS5CRRNAS	123	--T-G-----AG---C-A
II. Highest homologous sequence in GenBank			
Tobacco chloroplast	TOBCPTGRG	124	C-A-CC-----
Aureobasidium pullulans	AURRR16S	125	-----A-CA-T-----
TOBCOCG	127	C-A-CC-----	

N. tabacum

-7(-

Table VII. Probes for Aspergillus (Asp-693)

Species	GenBank name	Sequence I.D. No.	Asp-693 CTTCTGGGGAACCTCATGG
I. Fungi			
<i>Pneumocystis carinii</i>	PMC16SRR1	127	AACAC-----A----CCA
<i>Cryptococcus neoformans</i>	CPCDA	128	AACAC-----A----CCA
<i>Coccidioides immitis</i>	COIDA	129	-----CT----
<i>Blastomyces dermatitidis</i>	BLODA	130	--C-----A-G--C-----
<i>Aspergillus fumigatus</i>	ASNDA	131	-----
<i> fumigatus</i>	ASNRR5SS	132	-----
<i> fumigatus</i>	ASNRRSSB	133	-----
<i>Candida albicans</i>	YSASRSUA	134	AAAGG---C-----TC
<i> albicans</i>	YSAL16S	135	AACAC-----A----CCA
<i> lusitaniae</i>	YSASRRNAA	136	AA---T-----A---CGTC
<i> lusitaniae</i>	YSASRSUE	137	AA---T-----A---CGTC
<i> kefyr</i>	YSASRSUB	138	-----CT-----GTACT
<i> krusei</i>	YSASRRNAC	139	AACAC-----A----CCA
<i> krusei</i>	YSASRSUD	140	-----CT-S---GG-C
<i> tropicalis</i>	YSASRRNAB	141	-----CT-G---TT---
<i> tropicalis</i>	YSASRSUG	142	-----CT-G---TT---
<i> viswanathii</i>	YSASRSUH	143	-----CT-G---TT---
<i> parapsilosis</i>	YSASRSUF	144	AACAC-----A----CCA
<i> guilliermondii</i>	YSASRSUC	145	-----CT----ATT-C
<i> glabrata</i>	YSSCRRNAS	146	-----CT----C--A-T
II. Highest homologous sequence in GenBank			
<i>Pennicillium notatum</i> sub	PNNDA	147	-----
HUMBAT3A 148	-----GC-A-		
Rat olfactory protein	RATOLFPRON	140	--C-----CA

Human HLA-B-AT3

-77-

Table VIII. Probes for *Aspergillus* (Asp-1046)

Species	Genbank name	Sequence I.D. No.	Asp-1046 GGCGGTGTTTCTATGATGACC
I. Fungi			
<i>Pneumocystis carinii</i>	PMC16SRR1	150	----A-----T-T-CT----T
<i>Cryptococcus neoformans</i>	CPCDA	151	TTGTTG-----G---CG--
<i>Coccidioides immitis</i>	COIDA	152	ACGT---G-----TT--TTG
<i>Blastomyces dermatitidis</i>	BLODA	153	-A---G---CT-----
<i>Aspergillus fumigatus</i>	ASNDA	154	-----
<i>fumigatus</i>	ASNRR5SS	155	-----
<i>fumigatus</i>	ASNRRSSB	156	-----
<i>Candida albicans</i>	YSASRSUA	157	CCTTCG-GC---T-----TT
<i>albicans</i>	YSAL16S	158	CCTTCG-GC---T-----TT
<i>lusitaniae</i>	YSASRRNAA	159	-----C---CA-T-AG---G
<i>lusitaniae</i>	YSASRSUE	160	-----C---CA-T-AG---G
<i>kefyr</i>	YSASRSUB	161	--T-----T-C-T-----
<i>krusei</i>	YSASRRNAC	162	-A---C-AC-----G-A-G-
<i>krusei</i>	YSASRSUD	163	-A---C-AC-----G-A-G-
<i>tropicalis</i>	YSASRRNAB	164	TCTTCG-AC---T-----TT
<i>tropicalis</i>	YSASRSUG	165	TCTTCG-AC---T-----TT
<i>viswanathii</i>	YSASRSUH	166	CCTTCG-GC---T-----TT
<i>parapsilosis</i>	YSASRSUF	167	----A--G---AT-C-AATTT
<i>guilliermondii</i>	YSASRSUC	168	TCTTTGAGC---T-----TT
<i>glabrata</i>	YS5CRRNAS	169	--T-----T-T-AG-----
II. Highest homologous sequence in GenBank			
<i>Nanochlorum eucaryotum</i>	NANRRN18S	170	-CG-----T-T-----
<i>Moraxella</i> sp. MspI	MBOMSPI	171	TCTA-----A-----T
<i>E. coli</i> cvaA,B operon	ECOCVAB	172	-T-----G-G-----TG

-78-

Table IX. Probes for Blastomyces (Blast-694)

Species	GenBank name	Sequence I.D. No.	Blast-694 TCCTGGGAAGCCCATG
I. Fungi			
<i>Pneumocystis carinii</i>	PMC16SRR1	173	GT---T--T----TTA-
<i>Cryptococcus neoformans</i>	CPCDA	174	-G---AA-----GAC
<i>Coccidioides immitis</i>	COIDA	175	-T-----G-A---T---
<i>Blastomyces dermatitidis</i>	BLODA	176	-----
<i>Aspergillus fumigatus</i>	ASNDA	177	-T-----G-A--T----
<i>Aspergillus fumigatus</i>	ASNRR5SS	178	-T-----G-A--T----
<i>Aspergillus fumigatus</i>	ASNRRSSB	179	-T-----G-A--T----
<i>Candida albicans</i>	YSASRSUA	180	-T-----T----ATT-A
<i>Candida albicans</i>	YSAL16S	181	GT---T--T----TTA-
<i>Candida lusitanae</i>	YSASRRNAA	182	GT---T--T----TTA-
<i>Candida lusitanae</i>	YSASRSUE	183	GT---T--T----TTA-
<i>Candida kefyr</i>	YSASRSUB	184	GT---T--T----TTA-
<i>Candida krusei</i>	YSASRRNAC	185	GT---T--T----TTA-
<i>Candida krusei</i>	YSASRSUD	186	GT---T--T----TTA-
<i>Candida tropicalis</i>	YSASRRNAB	187	GT---T--T----TTA-
<i>Candida tropicalis</i>	YSASRSUG	188	GT---T--T----TTA-
<i>Candida viswanathii</i>	YSASRSUH	189	GT---T--T----TTA-
<i>Candida parapsilosis</i>	YSASRSUF	190	GT---T--T----TTA-
<i>Candida guilliermondii</i>	YSASRSUC	191	GT---T--T----TTA-
<i>Candida glabrata</i>	YS5CRRNAS	192	G-----T----GGTCC
II. Highest homologous sequence in GenBank			
Avian influenza	FLAHA5	193	-----A-----
Mouse perlecan	MUSPERPA	194	-----C--G-----
Mouse basement membrane	MUSPGCBMA	195	-----C--G-----

-79-

Table X. Probes for Blastomyces (Blast-1046)

Species	GenBank name	Sequence ID No.	Blast-1046 GACGGGGTTCTTATGATGACC
I. Fungi			
<i>Pneumocystis carinii</i>	PMC16SRR1	196	CG----C-----GAGG---T
<i>Cryptococcus neoformans</i>	CPCDA	197	-GT-AAA-----GAT----G
<i>Coccidioides immitis</i>	COIDA	198	-----CAA---TGA--A---
<i>Blastomyces dermatitidis</i>	BLODA	199	-----
<i>Aspergillus fumigatus</i>	ASNDA	200	-G---T---TC-----
<i>fumigatus</i>	ASNRR5SS	201	-G---T---TC-----
<i>fumigatus</i>	ASNRRSSB	202	-G---T---TC-----
<i>Candida albicans</i>	YSASRSUA	203	-TT-TT-----T-AT---G
<i>albicans</i>	YSAL16S	204	-TT-TT-----T-AT---G
<i>lusitaniae</i>	YSASRRNAA	205	-G---C---A-T-AG---G
<i>lusitaniae</i>	YSASRSUE	206	-G---C---A-T-AG---G
<i>kefyr</i>	YSASRSUB	207	-GT--T---T--C-T-----
<i>krusei</i>	YSASRRNAC	208	-----TC-A-C---G-A-G-
<i>krusei</i>	YSASRSUD	209	-----TC-A-C---G-A-G-
<i>tropicalis</i>	YSASRRNAB	210	-TT-TT-----T-AT---G
<i>tropicalis</i>	YSASRSUG	211	-TT-TT-----T-AT---G
<i>viswanathii</i>	YSASRSUH	212	-TT-TT-----T-AT---G
<i>parapsilosis</i>	YSASRSUF	213	-TT-TT-----T-AT---G
<i>guilliermondii</i>	YSASRSUC	214	-GT-TT-----KT-TT---G
<i>glabrata</i>	YS5CRRNAS	215	-GT--T---T--T-AG-----
II. Highest homologous sequence in GenBank			
Rat ITPR2 Type2inositol	RATITPR2R	216	-----CC-CT-----
Canine mRNA	DOGSRPR	217	CTGCTAA-----
Mitochondrion Oenothera	OBEMTNAD12	218	C-GTCTT-----

-80-

Table XI. Probes for Candida (Cand-513)

Species	GenBank Sequence		Cand-513
	name	ID No.	
			GAGTACAATGTAAATACCTTAACGAG
I. Fungi			
<i>Pneumocystis carinii</i>	PMC16SRR1	357	-----T--G-----
<i>Cryptococcus neoformans</i>	CPCDA	358	-----T----C-----
<i>Coccidioides immitis</i>	COIDA	359	-----T----C-----
<i>Blastomyces dermatitidis</i>	BLODA	360	-----C----C-----
<i>Aspergillus fumigatus</i>	ASNDA	361	-----C----C-----
	ASNRR5SS	362	-----C----C-----
	ASNRRSSB	363	-----C----C-----
<i>Candida albicans</i>	YSASRSUA	364	-----
	YSAL16S	365	-----
	YSASRRNAA	366	-----
	YSASRSUE	367	-----
	YSASRSUB	368	-----
	YSASRRNAC	369	-----
	YSASRSUD	370	-----
	YSASRRNAB	371	-----
	YSASRSUG	372	-----
	YSASRSUH	373	-----
	YSASRSUF	374	-----
	YSASRSUC	375	-----
	YS5CRRNAS	376	-----
II. Highest homologous sequence in GenBank			
Yeast 18S rRNA	YSCRNA5	377	-----
Yeast (<i>S. cerevisiae</i>)	YSCRGEA	378	-----
<i>Kluyveromyces lactis</i>	YSK17SRRNA	379	-----
<i>Torulaspora delbrueckii</i>	TOUSRSR	380	-----
<i>T. glabrata</i> rRNA	YSLRSUA	381	-----
<i>H. polymorpha</i> rRNA	HASSRSUA	382	-----
<i>S. pombe</i> rRNA	YSPRRNASS	383	-----

-81-

Table XII. Probes for Candida (Cand-701)

Species	GenBank name	Sequence ID No.	Cand-701 GGTAGCCATTATGGCGAACC
I. Fungi			
<i>Pneumocystis carinii</i>	PMC16SRR1	384	TTA-----GC---T-T--GT
<i>Cryptococcus neoformans</i>	CPCDA	385	TTCG---C-C-----T---T-
<i>Coccidioides immitis</i>	COIDA	386	TTA-----GC---T-T--GT
<i>Blastomyces dermatitidis</i>	BLODA	387	TTA-----GC---T-T--GT
<i>Aspergillus fumigatus</i>	ASNDA	388	ATA---A-----ACG-TGAA
<i>fumigatus</i>	ASNRR5SS	389	ATA---A-----ACG-TGAA
<i>fumigatus</i>	ASNRRSSB	390	ATA---A-----ACG-TGAA
<i>Candida albicans</i>	YSASRSUA	391	-----
<i>albicans</i>	YSAL16S	392	-----
<i>lusitaniae</i>	YSASRRNAA	393	TTA-----GC---T-T--GT
<i>lusitaniae</i>	YSASRSUE	394	-C-----TNG-C-----CG-N
<i>kefyr</i>	YSASRSUB	395	-C---NC--GC---TTN--T
<i>krusei</i>	YSASRRNAC	396	C--TTT---A--CAA---G
<i>krusei</i>	YSASRSUD	397	TTA-----GC---T-T--GT
<i>tropicalis</i>	YSASRRNAB	398	-C-----T-----
<i>tropicalis</i>	YSASRSUG	399	-C-----T-----
<i>viswanathii</i>	YSASRSUH	400	-C-----T-----
<i>parapsilosis</i>	YSASRSUF	401	-C-----T-----
<i>guilliermondii</i>	YSASRSUC	402	---CCG-C---T-----GTA
<i>glabrata</i>	YS5CRRNAS	403	ATA---A-----ACA-TGAA
II. Highest homologous sequence in GenBank			
<i>S. enterica</i>	STYRFB	404	-----G-----CGCTT
<i>Clostridium pasteurianum</i>	CLONIFH5	405	-----T-----AA--GG
<i>C. pasteurianum nifH</i>	CLONIFH	406	-----T-----AA--GG
<i>C. pasteurianum nifH</i>	CLONIFH1	407	-----T-----AA--GG

-82-

Table XIII. Probes for Coccidioides (Cocc-659)

Species	GenBank name	Sequence ID No.	Cocc-659 CTTCGGCGCGTGCACTG
I. Fungi			
<i>Pneumocystis carinii</i>			PMC16SRR1 265 -----ATC---TG-
<i>Cryptococcus neoformans</i>	CPCDA	266	-C--A---A-----
<i>Coccidioides immitis</i>	COIDA	267	-----
<i>Blastomyces dermatitidis</i>	BLODA	268	-C--A-C-----
<i>Aspergillus fumigatus</i>	ASNDA	269	-C--AT---C-T----
<i> fumigatus</i>	ASNRR5SS	270	-C--AT---C-T----
<i> fumigatus</i>	ASNRRSSB	271	-C--AT---C-T----
<i>Candida albicans</i>	YSASRSUA	272	GCA----CGC-A-----
<i> albicans</i>	YSAL16S	273	GCA----CGC-A-----
<i> lusitaniae</i>	YSASRRNAA	274	-GCTTA----A-----
<i> lusitaniae</i>	YSASRSUE	275	TNGTCT---C--N--T
<i> kefyr</i>	YSASRSUB	276	GCA----CGC-A-----
<i> krusei</i>	YSASRRNAC	277	GCA----CGC-A-----
<i> krusei</i>	YSASRSUD	278	GCA----CGC-A-----
<i> tropicalis</i>	YSASRRNAB	279	GCA----CGC-A-----
<i> tropicalis</i>	YSASRSUG	280	GCA----CGC-A-----
<i> viswanathii</i>	YSASRSUH	281	GCA----CGC-A-----
<i> parapsilosis</i>	YSASRSUF	282	GCA----CGC-A-----
<i> guilliermondii</i>	YSASRSUC	283	---TTT---A-T----
<i> glabrata</i>	YS5CRRNAS	284	---G-----AA-CAG-
II. Highest homologous sequence in GenBank			
Rabbit progest. recep.	RABPRG1	285	-----A-----G-
<i>Streptomyces lividans</i> 66	STMTRNGM	286	-----C-----GA
Human mRNA cysteine	HUMCYSTCR	287	G - G T - - - - -

-83-

Table XIV. Probes for Coccidioides (Cocc-1050)

Species	GenBank Sequence		Cocc-1050
	.name	ID No.	GGCAACTTTGAATAACCCGTTTC
I. Fungi			
Pneumocystis carinii	PMC16SRR1	288	TTACTAC---G-----GTGGT
Cryptococcus neoformans	CPCDA	289	CTGCC-----T-C-CA---CC-C
Coccidioides immitis	COIDA	290	-----
Blastomyces dermatitidis	BLODA	291	--GTT---ATG--G-----
Aspergillus fumigatus	ASNDA	292	CTGCC-----T-C-CA---CC-
fumigatus	ASNRR5SS	293	TT----G-G----GC-TA--AG
fumigatus	ASNRR5SB	294	TT----G-G----GC-TA--AG
Candida albicans	YSASRSUA	295	TTACTAC---G-----GTGGT
albicans	YSAL16S	296	TTACTAC---G-----GTGGT
lusitaniae	YSASRRNAA	297	T-TT-----G---ATGAGAG
lusitaniae	YSASRSUE	298	T-TT-----G---ATGAGAG
kefyr	YSASRSUB	299	TTTT-----A--ATTAGAG
krusei	YSASRRNAC	300	-----CCCATGGGGCCGA
krusei	YSASRSUD	301	-----CCCATGGGGCCGA
tropicalis	YSASRRNAB	302	TTACTAC---G-----GTGGT
tropicalis	YSASRSUG	303	TTACTAC---G-----GTGGT
viswanathii	YSASRSUH	304	TTACTAC---G-----GTGGT
parapsilosis	YSASRSUF	305	TTACTAC---G-----GTGGT
guilliermondii	YSASRSUC	306	TTACTAC---G-----GTGGT
glabrata	YS5CRRNAS	307	TTTT-----A--ATTAGAG
II. Highest homologous sequence in GenBank			
Genome bacteriophage T7	PT7DOT7	308	--ACTTC-----
Bacteriophage T7, comple.	PT7CG	309	--ACTTC-----
Staphylococcus aureus	STATOXA	310	-----C-----TCGA

SUBSTITUTE SHEET

-84-

Table XV. Probes for *Cryptococcus* (Cryp-691)

Species	GenBank name	Sequence ID No.	Cryp-691 GTGGTCCTGTATGCTCTTTACT
I. Fungi			
<i>Pneumocystis carinii</i>	PMC16SRR1	311	-----GG--C---GC-G--CT-
<i>Cryptococcus neoformans</i>	CPCDA	312	-----
<i>Coccidioides immitis</i>	COIDA	313	C-----G-CTG-AC----C--
<i>Blastomyces dermatitidis</i>	BLODA	314	C-----G-CCG-AC----C--
<i>Aspergillus fumigatus</i>	ASNDA	315	C-----G-CTG-AC----C--
	ASNRRSSS	316	CA----TGTG----C---AGA-
	ASNRRSSB	317	CA----TGTG----C---AGA-
<i>Candida albicans</i>	YSASRSUA	318	-----GG--C---GC-G--CT-
	YSAL16S	319	-----GG--C---GC-G--CT-
	YSASRRNAA	320	-----GG--C---GC-G--CT-
	YSASRSUE	321	CA----TGTG----C---AGAC
	YSASRSUB	322	----NNG--C---GC-G--CT-
	YSASRRNAC	323	-----GG--C---GC-G--TT-
	YSASRSUD	324	CA----TGTG----C---AGAC
	YSASRRNAB	325	-----GG--C---GC-G--CT-
	YSASRSUG	326	----NG--C---GC-G--CT-
	YSASRSUH	327	-----GG--C---GC-G--CT-
	YSASRSUF	328	----NG--C---GC-G--CT-
	YSASRSUC	329	TAAAAA-CA-----GAG
	YS5CRRNAS	330	-----GG--C---GC-G--CT-
II. Highest homologous sequence in GenBank			
Rat α tropomyosin	RATTMA3	331	T-----T-----CGT-
Human ribonucl/angio	HUMRAJ	332	C-----C-ACA-----
Human ribonucl/angio inh	HUMRAI	333	C-----C-ACA-----

-85-

Table XVI. Probes for *Cryptococcus* (Cryp-1042)

Species	GenBank name	Sequence ID No.	Cryp-1042 CACGTCAATCTCTGACTGGG
I. Fungi			
<i>Pneumocystis carinii</i>	PMC16SRR1	334	TT-T-G-T---A--GG---T
<i>Cryptococcus neoformans</i>	CPCDA	335	-----
<i>Coccidioides immitis</i>	COIDA	336	GG-----G-A-TC-G---TC
<i>Blastomyces dermatitidis</i>	BLODA	337	GG-----G-A-TC-G---TC
<i>Aspergillus fumigatus</i>	ASNDA	338	-G--CGCTA-A-----A---
<i>Aspergillus fumigatus</i>	ASNRR5SS	339	-G--CGCTA-A-----A---
<i>Aspergillus fumigatus</i>	ASNRRSSB	340	-G--CGCTA-A-----A---
<i>Candida albicans</i>	YSASRSUA	341	GGG-G---C---ATT----A
<i>Candida albicans</i>	YSAL16S	342	--TTCA---T----C-CTAT
<i>Candida lusitanae</i>	YSASRRNAA	343	--TTCA---T----C-CTAT
<i>Candida lusitanae</i>	YSASRSUE	344	--TTCA---T----C-CTAT
<i>Candida kefyr</i>	YSASRSUB	345	A-AA-----G---TCGGACT
<i>Candida krusei</i>	YSASRRNAC	346	TG---A--G-C-C----TC-
<i>Candida krusei</i>	YSASRSUD	347	-T---G--A---C-T-GT-C
<i>Candida tropicalis</i>	YSASRRNAB	348	A-AA-----G---TCGGACT
<i>Candida tropicalis</i>	YSASRSUG	349	-TN-----A--TG-N-ATTT
<i>Candida viswanathii</i>	YSASRSUH	350	A-A---NNGGGNNCNN---
<i>Candida parapsilosis</i>	YSASRSUF	351	--TTCA---T----C-CTAT
<i>Candida guilliermondii</i>	YSASRSUC	352	--TTCA---T----C-CTAT
<i>Candida glabrata</i>	YS5CRRNAS	353	GTT---C-CT---T-GA---
II. Highest homologous sequence in GenBank			
Rat α -1-acid gly.pro.	RATAGPA1H	354	T-TTA-----T
Rat α -1-acid gly(sp-daw)	RATAGPA1G	355	T-TTA-----T
<i>S.pneumoniae</i> malX malM	STRMALMXP	356	T-ACC-----A-----

-86-

Table XVII. Jun-common sense primer (S943-2, SEQ ID NO:728).

5	Sequences (5'-3')		SEQ ID NO:	CCGCTGTCCCCATCGACATGG
	LocusPos			
	Human			
10	B:humjunca1189	408	---G-----A-----	
	C:humjuna1981	409	--C-----T-----	
	D:humjundr943	410	---T---G-----	
	Mouse			
15	B:musjunba1079	411	--TG-----A-----	
	C:musjunc1344	412	--C-----T-----	
	C:muscjun1646	413	--C-----T-----	
	C:musjun1084	414	--C-----T-----	
	D:musjund927	415	-----G-----	
20	D:musjunda782	416	-----G-----	
	D:musjundr793	417	-----G-----	
	Rat			
25	C:atjunap11082	418	--CT-----	
	C:ratrjg92984	419	--CT-----	
	Chicken			
	C:chkjun1470	420	--C-----T--T-----	
30	Quail			
	C:quljun1186	421	--C-----T--T-----	
	Drosophila			
35	C:drojun1038	422	A-CG-TAAT-----T-----	
	Highest matched sequences in EMBL			
	SDNAM2G	Yeast NAM2 gene	423	-----A-A-----GAAT
	PRK2TRFB	Plasmid PK2 trfB ope	424	GT-----GC-T-
	DMSYT	D.melanogaster synap	425	-----G-A---UC-----
40				

-87-

Table XVIII. Jun-common antisense primer (AS1132-2, SEQ ID NO:729).

5	Sequences (5'-3')		
	LocusPos	SEQ ID NO:	CCGCTGTCCCCATCGACATGG
10	Human		
	B:humjunca1378	426	-----G-----G-----
	C:humjuna2170	427	-----T-----
	D:humjundr1132	428	-----G-C---
15	Mouse		
	B:musjunba1268	429	--T-----G-C---
	C:musjunc1835	430	-----T-----C---
	C:muscjun1273	431	-----T-----C---
	C:musjun1533	432	-----T-----C---
	D:musjund982	433	-----GC-G-----
	D:musjunda1116	434	-----G-----
20	D:musjundr971	435	-----G-----
	Rat		
	C:atjunap11271	436	-----T-----
25	C:ratrjg93173	437	-----T-----
	Chicken		
30	C:chkjun1659	438	-----A--T-----C---
	Quail		
35	C:quljun1375	439	-----A--T-----C---
	Drosophila		
40	C:drojun1227	440	--A-----GC-C--C---
	Highest matched sequences in EMBL		
45	HSATFA Human mRNA for ATF-a	441	C-----G---A-C-----
	ECDCM E. coli dom gene	442	AA-----ACCA
	ECDCMA E. coli dom	443	AA-----ACCA
	OCIGKC1 Rabbit Ig germkine k	444	-----G-----G-CGA
	OCIG05 Rabbit Ig k2 L chain	445	-----G-----G-CGA
	OCK1B4 Rabbit Ig kappa L	446	-----G-----G-CGA
	OCIGKCG Rabbit Ig kappa2 J-C	447	-----G-----G-CGA
	OCIGKC02 Rabbit Ig kappa1 J-C	448	-----G-----G-CGA

-88-

Table XIX. Jun-B specific probe (B-1258, SEQ ID NO:730).

5	GenBank		B1258(5'-3')
			SEQ ID NO: CTGGCGGCCACCAAGTG
10	Human		
	B: HUMJUNCA	449	-----
	C: HUMJUNA	450	A-C--T---T-----
15	D: HUMJUNDR	451	A-----TCCAA
	Mouse		
	B: MUSJUNBA	452	-----
20	C: MUSJUNC	453	A-T--C---T-----
	MUSCJUN	454	A-T--C---T-----
	MUSJUN	455	A-T--C---T-----
25	D: MUSJUND	456	-----C-----CCCG-
	MUSJUNDA	457	-----T-----GCCA-
	MUSJUNDR	458	-----T-----GCCA-
30	Rat		
	C: RATRJG9	459	-----T-----GCCAA
	RATJUNAP	460	A-C--T---T-----
35	Chicken		
	C: CHKJUN	461	-C-----G-----CCC
	Quail		
35	C: QULJUN	462	GGC-----AG-----TG-
	Drosophila		
	C: DROJUN	463	G----T--AT-----
35	Homologous sequences in GenBank		
	J04695 Figure 2. Nucleo.	464	G-----A-----G
	M27884 Figure 2. Nucleo.	465	-----A-----GC
	HUMCNPG2 green cone photo.	466	-----T-AA
	HUMCNPR2 green cone photo.	467	-----T-AA
35	HUMPIGMF2 colour-blind pho.	468	-----T-AA

-89-

Table XX. c-jun specific probe (C2147).

	Subtype	GenBank name	C2147.DNA	
			SEQ ID NO:	CCACGGCCAACATGCTC
5	Human			
	B:	HUMJUNCA	469	-G---A---C---CAGC
	C:	HUMJUNA	470	-----
	D:	HUMJUNDR	471	-----G-G-C---G
10	Mouse			
	B:	MUSJUNBA	472	-G-----C---CAGC
	C:	MUSJUN	473	-----
		MUSCJUN	474	-----
		MUSJUNC	475	-----
15	D:	MUSJUND	476	---C---G-C---G
		MUSJUNDR	477	---C---G-C---G
		MUSJUNDA	478	---C---G-C---G
	Rat			
	C:	RATJUNAP	479	-----
20		RATRJG9	480	---C-----
	Chicken			
	C:	CHKJUN	481	-----
	Quail			
	C:	QULJUN	482	---T-----
25	Drosophila			
	C:	DROJUN	483	CCTACA-----ACC
	Homologous sequences in GenBank			
	MXBPALPA	L.enzymogenes	484	-----CG--
30	MXBPALP	L.enzymogenes	485	-----CG--

35

-90-

Table XXI. Human jun-D specific probe (HUMD965)

	Subtype	GenBank name	HUMD965	ACACGCAGGAGCGCATC
			SEQ ID NO:	
5		Human		
	B:	HUMJUNCA	486	GG-A--T-----
	C:	HUMJUNA	487	-GT-C-----G---
	D:	HUMJUNDR	488	-----
10		Mouse		
	B:	MUSJUNBA	489	-AGAC-----
	C:	MUSJUNC	490	-GT-T-----G---
		MUSCJUN	491	-GT-T-----G---
		MUSJUN	492	GT-T-----G---
15	D:	MUSJUND	493	-----A--A-----
		MUSJUNDA	494	-----A--A-----
		MUSJUNDR	495	-----A--A-----
		Rat		
	C:	RATJUNAP	496	-GT-T-----G---
20		RATRJG9	497	-GT-T-----G---
		Chicken		
	C:	CHKJUN	498	-GT-----A-A---
		Quail		
	C:	QULJUN	499	-GT-----A-A---
25		Drosophila		
	C:	DROJUN	500	G--A--T-----
		Homologous sequences in GenBank		
	CELPOLII	C.elegans RNA	501	-----A-AA
	ECOPRIAY	E.coli primo.	502	-ACA-----
30	SINOCK82	Ockelbo 82 g.	503	C-----TGC
	TRPPROC	Treponema pa.	504	G-----TGC
	ECOCYSJIHA	E.coli NADPH-	505	-----T-----G-G

-91-

Table XXII. Mouse jun-D specific probe (MUSD1063).

			MUSD1063.DNA	
	Subtype	GenBank name	SEQ ID NO:	CCCTCAAAGCCAGAACACCG
5		Human		
	B:	HUMJUNCA	506	-G----GGC-G----G-G-
	C:	HUMJUNA	507	AA--GC-C-----GC
	D:	HUMJUNDR	508	-----G-T-----G-
10		Mouse		
	B:	MUSJUNBA	509	TA---TCC---TCTG---C
	C:	MUSJUNC	510	AA--GC-T-----GC
		MUSJUN	511	AA--GC-T-----GC
15		MUSCJUN	512	AA--GC-T-----GC
	D:	MUSJUNDA	513	-----
		MUSJUNDR	514	-----
		MUSJUND	515	-----
20		Rat		
	C:	RATRJG9	516	G----GCC---GC-C---TT
		RATJUNAP	517	G----GCC---GC-C---TT
		Chicken		
	C:	CHKJUN	518	A---G-GG-A-A-----G---
25		Quail		
	C:	QULJUN	519	---C--G-T-----TG--G-A
		Drosophila		
	C:	DROJUN	520	A-----GGA---TGTGGCGC
30		Homologous sequences in GenBank		
	M27221	Figure 3.	521	TG-----T--T-----
	DROAMY	D.erecta	522	G-----G-A--T-----
	DROAMYQ	D.erecta	523	G-----G-A--T-----
	HUMIGCMUDE	Immunoglob.	524	A--C-----GTT
35				

-92-

Table XXIII. G protein common primer (G2-S)

			G2-S	
			SEQ ID NO:	AGCACCATTTGTGAAGCAGATGA
5	Human			
	Gi-1	HUMGNBPAI	525	--T--A-----
	Gi-2	HUMGIAA	526	-----C--C-----
	Gi-3	HUMGIAB	527	-----A-----
10	Gs	HUMGNPAS	528	-----
	Go	HUMGOAQ01	529	-----
	Rat			
	Gi-1	RATBPGTPB	530	-----A-----
15	Gi-2	RATBPGTPA	531	-----C--C-----
	Gi-3	RATBPGTP	532	--T--T-----
	Gs	RATBPGTPD	533	-----
	Go	RATBPGTPC	534	-----
20	Gx	RATGXA	535	-----C--C-----
	Highest matched sequences in GenBank			
	HUMADECYC	adenyl cyclase	536	-----
	RATACOA1	acyl-coA oxidase	537	GC-----G-----A-----
	HUMTGASE	transglutaminase	538	CT-----CC-AC
25				

-93-

Table XXIV. G protein common primer (G4-AS, SEQ ID NO:731).

		G4-AS	
		SEQ ID NO:	TGTTTGATGTGGGAGGCCAGAG
5	Human		
	Gi-1	HUMGNBPAL	539
	Gi-2	HUMGIAA	540
	Gi-3	HUMGIAB	541
10	Gs	HUMGNPAS	542
	Go	HUMGOAQ01	543
	Rat		
	Gi-1	RATBPGTPB	544
15	Gi-2	RATBPGTPA	545
	Gi-3	RATBPGTP	546
	Gs	RATBPGTPD	547
	Go	RATBPGTPC	548
	Gx	RATGXA	549
20	Highest matched sequences in GenBank		
	HUMLDLRRL	LDL-receptor	550
	MUSHEPGFA	hepatocyte growth	551
25	HUMKEREP	epidermal keratin	552

-94-

Table XXV. Gi-1 protein specific, human-rat common
primer (Gi-1)

5			SEQ ID NO:	Gi-1 GATGTTCTCAGAACTAGAGTGAAAAC
10	Human			
	Gi-1	HUMGNBPAI	553	-----
	Gi-2	HUMGIAA	554	TT-----CT-CCCCTGTCCCCT
	Gi-3	HUMGIAB	555	-----TC-G--G-----G--
	Gs	HUMGNPAS	556	AC---G-C---T---TCCTG--C--G
15	Rat			
	Gi-1	RATBPGTPB	558	-----
	Gi-2	RATBPGTPA	559	-----G--GC-G--CC-T----G--
	Gi-3	RATBPGTP	560	-----TC-G--G-----G--
	Gs	RATBPGTPD	561	A-GCACAATTA-TTA-----CG
20	Go	RATBPGTPC	562	--CA-C---C---C--G--C-----
	Gx	RATGXA	563	TCA----GAG--C---A-CC----CA

-95-

Table XXVI. Gi-2 protein specific, human-rat common primer (Gi-2)

		Gi-2	
		SEQ ID NO:	GCAACCTGCAGATCGACTTTG
5	Human		
	Gi-1	HUMGNBPAI	564 -G-GGT--A----A-----
	Gi-2	HUMGIAA	565 -----
	Gi-3	HUMGIAB	566 -ACGG--AA----T-----
	Gs	HUMGNPAS	567 AGT---A--T---T---G--
10	Go	HUMGOAQ01	568 CTTCT-----G-TG----C
	Rat		
	Gi-1	RATBPGTPB	569 -G-GAT--A-A-----
	Gi-2	RATBPGTPA	570 -----
	Gi-3	RATBPGTP	571 AA---G-----T-T-TT---
15	Gs	RATBPGTPD	572 AGT---A--T-----G--
	Go	RATBPGTPC	573 ATTCT-----A-TG----C
	Gx	RATGXA	574 -T---A-C---T-T-T---C-
20			

-9(-

Table XXVII. Gi-3 protein specific, human-rat common primer (Gi-3)

5	Gi-3			
	SEQ ID NO:			TCTTCGGACGAGAGTGAAGAC
10	Human			
	Gi-1	HUMGNBPAI	575	---CA-A--T-----A--
	Gi-2	HUMGIAB	576	G--A-----CC-C--A-----
	Gi-3	HUMGIAB	577	-----
	Gs	HUMGNPAS	578	GGGAAATCGA---T---G---
15	Go	HUMGOAQ01	579	---C---GA---CGTG-A--
	Rat			
	Gi-1	RATBPGTPB	580	---CA-A--T-----A--
	Gi-2	RATBPGTPA	581	G--G-----CC-T-----
	Gi-3	RATBPGTP	582	-----
20	Gs	RATBPGTPD	583	TTCCT----A---T---T-TG
	Go	RATBPGTPC	584	CGCAT---G--C-C---CCA
	Gx	RATGXA	585	AAC----G-A---CACCAT-

Table XXVIII. Gs Protein specific, human-rat common primer (Gs, Seq ID NO:732)

		Gs	
		Seq ID NO:	AATTCTATGACGATGCCAAGGC
Human	Gi-1	HUMGNBPAI	
	Gi-2	HUMGIAB	586 GGGCGG-----T---C-----CT
	Gi-3	HUMGIAB	587 ---ATG-----GCA-----GCTA
	Gs	HUMGNPAS	588 C-G-----ACTA---T-----CTC

Table XXIX. Go Protein specific, human-rat common primer (Go)

		SEQ ID NO:	Go CTGCTTTCTGCCATGATCGG
5	Human		
	Gi-1	HUMGNBPAI	
	Gi-2	HUMGTAA	589 GAC-----TAA---TGACA
10	Gi-3	HUMGTAB	590 ---TGA---C-A-G-C----
	Gs	HUMGNPAS	591 ---A---AGA-CTTCACA
	Go	HUMGOA01	592 TCAACGA-----G---CATC
			593 -----
15	Rat		
	Gi-1	RATBPGTPB	
	Gi-2	RATBPGTPA	594 ---A-----CCA-G-T---TT
	Gi-3	RATBPGTP	595 T---GA-CC---CCG----
	Gs	RATBPGTPD	596 ---G---T---TG-CAC
	Go	RATBPGTPC	597 ---G---AG-A-G-C-T
20	Gx	RATGXA	598 -----
			599 -A---C-T---C---CTG-
25			

Table XXX. Human-rodent common jun-B specific probes.

	Name in GenBank	B-504 (5'-3')	Seq. ID No:	B-739 (5'-3')	Seq. ID No:	Size (bp)
Mouse		CACGACTACAACTCCTGAAAC		GGACAGTACTTTTACCCCG		
jun-B	MUSJUNBA	-----	600	-----	607	251
c-JUN	MUSCJUN	A-AT--A-T-----AT-----A	601	ACC-----T---G-G-----AA	608	
jun-D	MUSJUND	T----G-C--C-T-----TTCC	602	ACG-----T-C-C-----GAA	609	
Human						
jun-B	HUMJUNCA	-----	603	-----	610	251
c-jun	HUMJUNA	AGTA--CC---GA-----	604	---A-C-GCG-----AC	611	
jun-D	HUMJUNDR	T----G-C--C-T-----TTTG	605	TC-----T-C-C-----AA	612	
GenBank						
	MUSMET1	-----T---CG-----GTA	606			
	RATMYHCD2			-----C-----AT---A	613	

*: Size of PCR products using B-504 and B-739 as primers.

**: The highest matched sequences in GenBank (release 68.0) next to the jun genes.

-100-

Table XXXI. Human-rodent common c-jun specific probes.

	Name in GenBank	C-2101 (5'-3')	Seq. ID No:	C-2219 (5'-3')	Seq. ID No:	Size (bp)
Mouse						
jun-B	MUSJUNBA	-----C--G-----G--AC-C--G--	614	----GT-CC-----GG-GTC	621	138
c-JUN	MUSCJUN	-----	615	-----	622	
jun-D	MUSJUND	-----G-----C--G---C-C---AG	616	T-----A-GCC--A-G-----A	623	
Human						
jun-B	HUMJUNCA	-----C--G-----G--GC-C--G--	617	----G-C-CC-AG-----GC	624	138
C-jun	HUMJUNA	-----	618	-----	625	
jun-D	HUMJUNDR	--A--G-----G---C-C--GAG	619	T-----G-GCC--A-G-----A	626	
GenBank						
	HUNGASTA	A-A-----AA-----A---A-----	620			
	MUSOCT22			C-----G---GGC	627	

*: Size of PCR products using C-2101 and C-2219 as primers.

**: The highest matched sequences in GenBank (release 68.0) next to the jun genes.

SUBSTITUTE SHEET

-101-

Table XXXII. Human-rodent common jun-D specific probes.

	Name in GenBank	D-916 (5'-3') GACGTCCGAGCTTCGG	Seq. ID No:	D-1153 (5'-3') AAAGTCCTCAGCCACGTCAAC	Seq. ID No:	Size (bp)
Mouse						
jun-B	MUSJUNBA	CTG-G-----ACT--	628	--G---A-G-C---T---G-	635	
c-JUN	MUSCJUN	T-A-----AG---AGAC	629	-----A-G-A-----T---	636	258
jun-D	MUSJUND	-----	630	-----	637	
Human						
jun-B	HUMJUNCA	C---C---CC-----C	631	--G---A-G-C-----G-	638	
C-jun	HUMJUNA	T-A---TT-----CG--	632	-----A-G-A-----T---	638	
jun-D	HUMJUNDR	-----	633	-----	640	258
GenBank						
	HUMTRKR	-----TG-----	634			
	HUMBSSL			---A-----A---CTG	641	

*: Size of PCR products using D-916 and D-1153 as primers.

**: The highest matched sequences in GenBank (release 68.0) next to the jun genes.

SUBSTITUTE SHEET

-102-

Table XXXIII. Human-rodent common Gs specific probes.

	Name in GenBank	Gs-246 (5'-3')	Seq. ID No:	Gs-824 (5'-3')	Seq. ID No:	Size (bp)
Rat						
Gs	RATBPGETD	-----	642	-----	652	600
Gi-1	RATBPGETB	CG--G---G-T-----CCGCA	643	-----G-CGGC-G-GG-CGGC	653	
Gi-2	RATBPGETA	CAA-GG-----CA-A-----	644	A-----C-C-C-G-T---T	654	
Gi-3	RATBPGETP	TCAGCA-----T--A-----T	645	-TCAGG---A-TA-----GC	655	
Go	RATBPGETPC	TTTGG-G-G-----TA-----T	646	CCGAGG--CGGCGA-G-----	656	
Mouse						
Gs-1	MUSGS	-----	647	-----	657	555
Gs-2	MUSGTPAMU	-----	648	-----	658	600
Human						
Gs-1	HUMGSA1R	-----	649	-----	659	558
Gs-2	HUMGSA2R	-----	650	-----	660	600
GenBank						
	HUMMBP2	A-----A-C-----A-----	651	A-----T---T-----A-	661	
	RATAFPM					

*: Size of PCR products using Gs-246 and Gs-824 as primers.

**: The highest matched sequences in GenBank (release 68.0) next to the Gs protein DNAs.

SUBSTITUTE SHEET

-103-

Table XXXIV. Human-rodent common Gi-1 specific probes.

	Name in GenBank	Gi1-735 (5'-3') GATGTTCTCAGAACTAGAGTGAAC	Seq. ID No:	Gi1-1131 (5'-3') TGTCAGTTTGAAGACCICATAAAG	Seq. ID No:	Size (bp)
Rat						
Gs	RATBP6TPD	A-GCACAATTA-TTA-----CG	662	AACAGAAA-A-----AA-A---G---T	669	
Gi-1	RATBP6TPB	-----	663	-----	670	422
Gi-2	RATBP6TPA	-----G--GC-G--CC-T-----G--	664	A-CA-----G-----G-----C-	671	
Gi-3	RATBP6TP	-----TC-G-G-----G--	665	--C-----T--G--CCG---	672	
Go	RATBP6TPC	--CA-C---C---C---G--C-----	666	ACA-----AG-AAA--CCGCTC	673	
Human						
Gs-1	HUMGNBPA	-----	667	-----	674	422
GenBank						
	HUMCPH192	C---C---G--C---G---A-----G	668			
	MUSGBPA			A-C---CC--T--C-----	675	

*: Size of PCR products using Gi1-735 and Gi1-1131 as primers.

**: The highest matched sequences in GenBank (release 68.0) next to the Gs protein DNAs.

SUBSTITUTE SHEET

Table XXXV. Human-rodent common Gi-2 specific probes.

	Name in GenBank	Gi2-742 (5'-3')	Seq. ID No:	Gi2-1102 (5'-3')	Seq. ID No:	Size (bp)
Rat		AAGATGTTTGATGTGGTGGTC		AAGGAGATCTACACGCACTTCA		
Gs	RATBPCTPD	G-C-A-A-----ATCAAG-	676	---C--G-----CG-GC-ACGC	684	
Gi-1	RATBPCTPB	--A-----C-----A--C-	677	-----A--T-----C-----	685	
Gi-2	RATBPCTPA	-----	678	-----	686	382
Gi-3	RATBPCTP	--A-----A-----C-	679	-----G-----T-----T-	687	
Go	RATBPCTPC	-G-C-----C--T--G--C-	680	CTA-----G---C-CTCA-C	688	
Mouse						
Gs-2	MUSGI	-----	681	-----	689	382
Human						
Gs-2	HUMGIR	-----	682	-----	690	382
GenBank						
	HUMCHPF	-----C-----TGG				
	MUSHSPB4A			-----A-----A-A-G---T	691	

*: Size of PCR products using Gi2-742 and Gi2-1102 as primers.

**: The highest matched sequences in GenBank (release 68.0) next to the Gs protein DNAs.

Rat Gi-2 was shown to be specifically amplified when PCR was performed with Gi-2 and SEQ ID NO:678 and SEQ ID NO:686 as PCR primers.

-105-

Table XXXVI. Human-rodent common Gi-3 specific probes.

	Name in GenBank	Gi3-407 (5'-3') TTGTTTAGCTGGCAGTCTGA	Seq. ID No:	Gi3-730 (5'-3') GAGGGAGTGACAGCAATTATCT	Seq. ID No:	Size (bp)
Rat						
Gs	RATBPGTDP	CC-GAAGT-GATC-----TC	692	A-T-AT-----T--C--C-----	699	
Gi-1	RATBPGTPB	A-----GA-----GGC-CTA-T	693	--A--C-----T--C--C-----	700	
Gi-2	RATBPGTPA	CCTACAC--A--AT-----C	694	-----T--C--G--C--C-----	701	
Gi-3	RATBPGTIP	-----	695	-----	702	345
Go	RATBPGTPC	CG-----T--AGTC-TTAC-		TC-----AT-----TCAACGAC	703	
Human						
Gi-3	HUMGTPBP	-----	697	-----	704	345
GenBank						
	HUMAPOB	G-----C--A-----T-----C	698			
	HUMINSR02			C-----G---AC---T---	705	

*: Size of PCR products using Gi3-407 and Gi3-730 as primers.

**: The highest matched sequences in GenBank (release 68.0) next to the Gs protein DNAs.

Rat Gi-3 was shown to be specifically amplified when PCR was performed with Gi-2 and SEQ ID NO:697 and SEQ ID NO:704 as PCR primers.

SUBSTITUTE SHEET

-106-

Table XXXVII. Human-rodent common Go specific probes.

	Name in GenBank	Go-1224 (5'-3')	Seq. ID No:	Go-1397 (5'-3')	Seq. ID No:	Size (bp)
Rat						
Gs	RATBPCTPD	-----AAAAC-A-CT	706	AG-----AC--G-TTC	717	
Gi-1	RATBPCTPB	G--G-CG-----GG----	707	-----T---T--TGCT-	718	
Gi-2	RATBPCTPA	CCAT-----TTCCT--A	708	-TCTGTT-----GG-GA	719	
Gi-3	RATBPCTP	G-CCA-AGA-----GGA	709	A-----TA-T-AATTT	720	
Go	RATBPCTPC	-----	710	-----	721	192
Mouse						
Go-A	MUSGOASA	-----	711	-----	722	192
Go-B	MUSGOASB	-----	712	-----	723	192
Hamster						
Go-1	HAMHITA01	-----	713	-----	724	192
Go-2	HAMHITA02	-----		-----	725	192
Human						
Go	HUMGOA1(2)	-----	715	-----	726	192
GenBank						
	MUSPC3AA	G-----T--TC-----	716			
	RATACACA			-----A-GT	727	

*: Size of PCR products using Go-1224 and Go-1397 as primers.

**: The highest matched sequences in GenBank (release 68.0) next to the Gs protein DNAs.

SUBSTITUTE SHEET

WHAT IS CLAIMED IS:

1. A method of detecting the presence of an organism, infectious agent, or biological component of a cell or organism in a biological sample containing polynucleotides, comprising the steps of:

5 (a) immobilizing a first polynucleotide probe to a solid support, wherein the nucleotide sequence of said first polynucleotide probe is sufficiently complementary to a first nucleotide sequence contained in an analyte polynucleotide in said organism, infectious agent, or biological component that said first polynucleotide probe can hybridize to said first nucleotide sequence of said analyte polynucleotide of said organism, infectious agent, or biological component;

10 (b) contacting the polynucleotides present in said sample with said first polynucleotide probe;

(c) hybridizing said analyte polynucleotide in said sample to said first polynucleotide probe, if said analyte polynucleotide is present in said sample;

15 (d) contacting a second polynucleotide probe with said analyte polynucleotide hybridized to said first polynucleotide probe, if said analyte polynucleotide from said sample has hybridized to said first polynucleotide probe, wherein the nucleotide sequence of said second polynucleotide probe is sufficiently complementary to a second nucleotide sequence contained in said analyte polynucleotide of said organism, infectious agent, or biological component that said second polynucleotide probe can hybridize to said second nucleotide sequence;

20 (e) hybridizing said second polynucleotide probe to said analyte polynucleotide hybridized to said first polynucleotide probe, if said analyte polynucleotide has hybridized to said first polynucleotide probe; and

25 (f) determining the presence of said organism, infectious agent, or biological component in said sample by detecting the presence of said second polynucleotide probe hybridized to said analyte polynucleotide which has hybridized to said first polynucleotide probe.

30 2. The method of Claim 1, wherein said second polynucleotide probe has the same or lower T_m as said first polynucleotide probe.

3. The method of Claim 1, wherein said first polynucleotide probe has a T_m within the range of from approximately 48°C to approximately 60°C.

35 4. The method of Claim 1, wherein said first nucleotide sequence of said analyte polynucleotide from said sample is common to a plurality of organisms, infectious agents, or biological components of a cell or organism.

-108-

5. The method of Claim 1, wherein said first nucleotide sequence of said analyte polynucleotide from said sample is specific to a particular organism, infectious agent, or biological component of a cell or organism.

5 6. The method of Claim 4, wherein said first polynucleotide probe has a sequence complementary to an rRNA sequence that is specific to a particular fungal species.

7. The method of Claim 1, wherein said second nucleotide sequence of said analyte polynucleotide from said sample is common to a plurality of organisms, infectious agents, or biological components of a cell or organism.

10 8. The method of Claim 7, wherein said second polynucleotide probe has a sequence complementary to an rRNA sequence that is common to to a plurality of fungal species.

9. The method of Claim 1, wherein said second nucleotide sequence of said analyte polynucleotide from said sample is specific to a particular organism, infectious agent, or biological component of a cell or organism.

15 10. The method of Claim 1, wherein a label is attached to said second polynucleotide probe.

11. The method of Claim 10, wherein said label is selected from the group consisting of a radionuclide, an enzyme, an enzyme substrate, a specific binding moiety, an binding partner for a specific binding moiety, biotin, avidin, a nucleic acid stain, and a flourescent material.

20 12. The method of Claim 11, wherein said label is a nucleic acid stain selected from the group consisting of ethidium bromide, yoyo-1, and toto-1.

13. The method of Claim 11, wherein said label comprises alkaline phosphatase, and wherein step (f) comprises adding ATTOPHOS and measuring the fluorescence emitted using a fluorimeter.

25 14. The method of Claim 10, wherein said label can be measured by light emitted therefrom, and wherein step (f) comprises measuring the amount of light emitted by said label.

15. The method of Claim 14, wherein the step of measuring the amount of light emitted by said label comprises:

recording the amount of light on film; and

measuring the exposure of the film using a densitometer.

30 16. The method of Claim 1, wherein said solid support comprises a microtiter plate having a plurality of wells, each of said wells having a specific polynucleotide probe immobilized thereon.

17. The method of Claim 1, wherein said first polynucleotide probe comprises DNA.

18. The method of Claim 1, wherein said first and second polynucleotide probes comprise DNA.

35 19. The method of Claim 1, additionally comprising the step of washing said solid support after hybridizing said analyte polynucleotide in said sample to said first polynucleotide probe so that

-109-

substantially all of said biological sample not annealed to said first polynucleotide probe is removed from said solid support.

20. The method of Claim 1, additionally comprising the step of washing said solid support after hybridizing said second polynucleotide probe with said analyte polynucleotide in said sample which is hybridized to said first polynucleotide probe so that substantially all of said second polynucleotide probe not hybridized with said analyte polynucleotide is removed from said solid support.

21. The method of Claim 1, wherein the polynucleotide hybridized to said first polynucleotide probe is selected from the group consisting of mRNA, rRNA, and genomic DNA.

22. The method of Claim 1, additionally comprising the step of identifying said first and second polynucleotide probes.

23. The method of Claim 22, wherein said identifying step is performed by means of a computer-assisted method.

24. The method of Claim 23, wherein said identifying step comprises the use of an H-site model.

25. The method of Claim 24, wherein the identification of said first polynucleotide probe using said H-site model comprises the steps of:

specifying a minimum melting temperature for the first nucleotide probe and the nucleotide sequence specific to said organism;

specifying a nucleation threshold that places a minimum value on the number of base pairs at any nucleation site;

determining the melting temperatures (T_m) of the first nucleotide probe and said sequence specific to said organism at every possible hybridization point; and

selecting the nucleotide probe having the highest T_m value.

26. The method of Claim 25, wherein said melting temperature is determined by the formula;

$$T_m = 81.5 - 16.6(\log[Na]) - 0.63\%(\text{formamide}) + 0.41(\%/(G+C)) - 600/N$$
, wherein $\log[Na]$ is the log function of the sodium concentration, 0.063% (formamide) is the concentration of formamide, $\%(G+C)$ is the percentage of matched GC base pairs, and N is the probe length.

27. The method of Claim 24, wherein the identification of said second polynucleotide probe using said H-site model comprises the steps of:

specifying a minimum melting temperature for the second nucleotide probe and the nucleotide sequence specific to said organism;

specifying a nucleation threshold that places a minimum value on the number of base pairs at any nucleation site;

determining the melting temperatures (T_m) of the first nucleotide probe and said sequence specific to said organism at every possible hybridization point; and

selecting the nucleotide probe of the proper length having the lowest T_m value.

-110-

28. The method of Claim 27, wherein said melting temperature is determined by the formula;

$T_m = 81.5 - 16.6(\log[Na]) - 0.63\%(\text{formamide}) + 0.41(\%(G+C)) - 600/N$, wherein $\log[Na]$ is the log function of the sodium concentration, 0.063% (formamide) is the concentration of formamide, $\%(G+C)$ is the percentage of matched GC base pairs, and N is the probe length.

29. The method of Claim 1, wherein said organism or infectious agent to be detected is a variety of fungus, said first polynucleotide probe being a polynucleotide strand comprising a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:81, SEQ ID NO:104, SEQ ID NO:131 through SEQ ID NO:133, SEQ ID NO:154 through SEQ ID NO:156, SEQ ID NO:176, SEQ ID NO:199, SEQ ID NO:267, SEQ ID NO:290, SEQ ID NO:312, SEQ ID NO:335, SEQ ID NO:364 through SEQ ID NO:376, SEQ ID NO:391 through SEQ ID NO:392, a sequence homologous to any of the foregoing sequences, and a sequence capable of hybridizing to any of the foregoing sequences.

30. The method of Claim 1, wherein said organism or infectious agent to be detected is a variety of fungus, said second polynucleotide probe being a polynucleotide strand comprising a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:1 through SEQ ID NO:80, a sequence homologous to any of SEQ ID NO:1 through SEQ ID NO:80, and a sequence capable of hybridizing to any of the foregoing sequences.

31. The method of Claim 1, wherein said biological component to be detected is a jun oncogene, said first polynucleotide probe being a polynucleotide strand comprising a sequence selected from the group consisting of SEQ ID NO:473, SEQ ID NO:600, SEQ ID NO:607, SEQ ID NO:615, SEQ ID NO:622, SEQ ID NO:637, SEQ ID NO:730, SEQ ID NO:747, SEQ ID NO:748, SEQ ID NO:488, SEQ ID NO:513, SEQ ID NO:630, AND SEQ ID NO:639, a sequence complementary to any of such sequences, and a sequence capable of hybridizing with any of these sequences.

32. The method of Claim 1, wherein said biological component to be detected is a jun oncogene, said second polynucleotide probe being a polynucleotide strand comprising a sequence complementary to a sequence selected from the group consisting of SEQ ID NO: 728, SEQ ID NO:729, SEQ ID NO:733, SEQ ID NO:734, SEQ ID NO:739, SEQ ID NO:740, SEQ ID NO:741, SEQ ID NO:742, SEQ ID NO:743, and SEQ ID NO:744, a sequence complementary to any of such sequence, and a sequence capable of hybridizing with any of these sequences.

33. The method of Claim 1, wherein said biological component to be detected is a Substance P receptor, said first polynucleotide probe being a polynucleotide strand comprising a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:758, a sequence complementary to SEQ ID NO:758, and a sequence capable of hybridizing with SEQ ID NO:758.

34. The method of Claim 1, wherein said biological component to be detected is a β -receptor, said first polynucleotide probe being a polynucleotide strand comprising a sequence

complementary to a sequence selected from the group consisting of SEQ ID NO:759, a sequence complementary to SEQ ID NO:759, and a sequence capable of hybridizing with SEQ ID NO:759.

35. The method of Claim 1, wherein said biological component to be detected is a G protein, said first polynucleotide probe being a polynucleotide strand comprising a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:751, SEQ ID NO:553, SEQ ID NO:670, SEQ ID NO:752, SEQ ID NO:753, SEQ ID NO:565, SEQ ID NO:678, SEQ ID NO:686, SEQ ID NO:754, SEQ ID NO:577, SEQ ID NO:697, SEQ ID NO:704, SEQ ID NO:755, SEQ ID NO:756, SEQ ID NO:732, SEQ ID NO:642, SEQ ID NO:652, SEQ ID NO:757, SEQ ID NO:593, SEQ ID NO:710, and SEQ ID NO:721, a sequence complementary to any of such sequences, and a sequence capable of hybridizing with any of these sequences.

36. The method of Claim 1, wherein said biological component to be detected is a G protein, said second polynucleotide probe being a polynucleotide strand comprising a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:528, SEQ ID NO:731, SEQ ID NO:749, and SEQ ID NO:750, a sequence complementary to any of such sequences, and a sequence capable of hybridizing with any of these sequences.

37. A solid support-polynucleotide structure for identifying the presence of an organism, infectious agent, or biological component of a cell or organism in a biological sample containing polynucleotides, comprising:

a solid support having immobilized thereto a first polynucleotide probe, said first polynucleotide probe having a sequence complementary to a first nucleotide sequence specific to said organism, infectious agent, or biological component;

an analyte polynucleotide from said cell, organism, or infectious agent containing said first nucleotide sequence, said analyte polynucleotide being hybridized to said first polynucleotide probe at said first nucleotide sequence;

a second polynucleotide probe complementary to a second nucleotide sequence present on said analyte polynucleotide from said cell, organism, or infectious agent which is hybridized to said first polynucleotide probe, said second polynucleotide probe being hybridized to said analyte polynucleotide at said second nucleotide sequence.

38. The solid support-polynucleotide structure of Claim 37, wherein said second polynucleotide probe includes a label.

39. The solid support-polynucleotide structure of Claim 38, wherein said label is selected from the group consisting of a radionuclide, an enzyme, an enzyme substrate, a specific binding moiety, an binding partner for a specific binding moiety, biotin, avidin, a nucleic acid stain, and a fluorescent material.

40. The solid support-polynucleotide structure of Claim 37, wherein said second polynucleotide probe is common to polynucleotides contained in a plurality of organisms, infectious agents, or biological components of a cell or organism.

-112-

41. The solid support-polynucleotide structure of Claim 37, wherein said first and second polynucleotide probes are determined through the use of a computer system for designing oligonucleotide probes for use with a gene sequence data source, said computer system comprising:

an input means for retrieving said gene sequence data;

5

a processor;

instructions directing said processor to determine said first oligonucleotide probe;

42. The solid support-polynucleotide structure of Claim 37, wherein said polynucleotide is selected from the group consisting of mRNA, rRNA, and genomic DNA.

10

43. The solid support-polynucleotide structure of Claim 37, wherein said first polynucleotide probe is a polynucleotide strand comprising a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:81, SEQ ID NO:104, SEQ ID NO:131 through SEQ ID NO:133, SEQ ID NO:154 through SEQ ID NO:156, SEQ ID NO:176, SEQ ID NO:199, SEQ ID NO:267, SEQ ID NO:290, SEQ ID NO:312, SEQ ID NO:335, SEQ ID NO:364 through SEQ ID NO:376, SEQ ID NO:391 through SEQ ID NO:392, a sequence homologous to any of the foregoing sequences, and a sequence capable of hybridizing to any of the foregoing sequences.

15

44. The solid support-polynucleotide structure of Claim 37, wherein said second polynucleotide probe is a polynucleotide strand comprising a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:1 through SEQ ID NO:80, a sequence homologous to any of SEQ ID NO:1 through SEQ ID NO:80, and a sequence capable of hybridizing to any of the foregoing sequences.

20

45. The solid support-polynucleotide structure of Claim 37, wherein said first polynucleotide probe is a polynucleotide strand comprising a sequence selected from the group consisting of SEQ ID NO:473, SEQ ID NO:600, SEQ ID NO:607, SEQ ID NO:615, SEQ ID NO:622, SEQ ID NO:637, SEQ ID NO:730, SEQ ID NO:747, SEQ ID NO:748, SEQ ID NO:488, SEQ ID NO:513, SEQ ID NO:630, AND SEQ ID NO:639, a sequence complementary to any of such sequences, and a sequence capable of hybridizing with any of these sequences.

25

46. The solid support-polynucleotide structure of Claim 37, wherein said second polynucleotide probe is a polynucleotide strand comprising a sequence complementary to a sequence selected from the group consisting of SEQ ID NO: 728, SEQ ID NO:729, SEQ ID NO:733, SEQ ID NO:734, SEQ ID NO:739, SEQ ID NO:740, SEQ ID NO:741, SEQ ID NO:742, SEQ ID NO:743, and SEQ ID NO:744, a sequence complementary to any of such sequence, and a sequence capable of hybridizing with any of these sequences.

30

47. The solid support-polynucleotide structure of Claim 37, wherein said first polynucleotide probe is a polynucleotide strand comprising a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:758, a sequence complementary to SEQ ID NO:758, and a sequence capable of hybridizing with SEQ ID NO:758.

35

48. The solid support-polynucleotide structure of Claim 37, wherein said first polynucleotide probe is a polynucleotide strand comprising a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:759, a sequence complementary to SEQ ID NO:759, and a sequence capable of hybridizing with SEQ ID NO:759.

5 49. The solid support-polynucleotide structure of Claim 37, wherein said first polynucleotide probe is a polynucleotide strand comprising a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:751, SEQ ID NO:553, SEQ ID NO:670, SEQ ID NO:752, SEQ ID NO:753, SEQ ID NO:565, SEQ ID NO:678, SEQ ID NO:686, SEQ ID NO:754, SEQ ID NO:577, SEQ ID NO:697, SEQ ID NO:704, SEQ ID NO:755, SEQ ID NO:756, SEQ ID NO:732, SEQ ID NO:642, 10 SEQ ID NO:652, SEQ ID NO:757, SEQ ID NO:593, SEQ ID NO:710, and SEQ ID NO:721, a sequence complementary to any of such sequences, and a sequence capable of hybridizing with any of these sequences.

50. The solid support-polynucleotide structure of Claim 37, wherein said second polynucleotide probe is a polynucleotide strand comprising a sequence complementary to a sequence 15 selected from the group consisting of SEQ ID NO:528, SEQ ID NO:731, SEQ ID NO:749, and SEQ ID NO:750, a sequence complementary to any of such sequences, and a sequence capable of hybridizing with any of these sequences.

51. A kit for identifying the presence of an organism, infectious agent, or biological component of a cell or organism in a biological sample, comprising the following components:

20 a specific polynucleotide probe, said specific polynucleotide probe being complementary to or homologous to a first nucleotide sequence in an analyte polynucleotide specific to a particular organism, infectious agent, or biological component to be detected; and

25 a common polynucleotide probe complementary to or homologous to a second nucleotide sequence in said analyte polynucleotide of said organism, infectious agent, or biological component, said common polynucleotide probe being complementary to polynucleotides contained in a plurality of organisms, infectious agents, or biological components.

52. The kit of Claim 51, additionally comprising a solid support to which a polynucleotide can be immobilized.

30 53. The kit of Claim 52, wherein said specific polynucleotide probe is immobilized to said solid support.

54. The kit of Claim 53, wherein said solid support has a plurality of specific polynucleotide probes immobilized thereto, each of said probes specific to a different organism, infectious agent, or biological component.

35 55. The kit of Claim 53, wherein said solid support comprises a plurality of wells, each of said specific polynucleotide probes being immobilized to a different well.

-114-

56. The kit of Claim 55, additionally comprising a buffer appropriate for the hybridization of said probes and polynucleotides, said polynucleotides being selected from the group consisting of mRNA, rRNA, and genomic DNA.

57. The kit of Claim 51, wherein said second polynucleotide probe bears a label.

58. The kit of Claim 57, wherein said label is selected from the group consisting of a radionuclide, an enzyme, an enzyme substrate, a specific binding moiety, an binding partner for a specific binding moiety, biotin, avidin, a nucleic acid stain, and a fluorescent material.

59. The kit of Claim 51, wherein said specific polynucleotide probe comprises a first specific primer which is complementary or homologous to a sequence specific to a particular organism, infectious agent, or biological component, said kit additionally comprising a second specific polynucleotide primer which is complementary or homologous to a different sequence specific to said organism, infectious agent, or biological component.

60. The kit of Claim 59, additionally comprising at least one of the following: dNTP's, a reverse transcriptase, a polymerase, and a buffer appropriate for addition of dNTP's to a primer using a reverse transcriptase or polymerase.

61. The kit of Claim 60, including a DNA polymerase that has significant polymerase activity at temperatures above 50°C.

62. The kit of Claim 61, additionally comprising at least one of the following: dNTP's, a reverse transcriptase, a polymerase, and a buffer appropriate for addition of dNTP's to a primer using a reverse transcriptase or polymerase.

63. The kit of Claim 51, wherein said specific polynucleotide probe comprises a sequence complementary to or homologous to a sequence selected from the group consisting of SEQ ID NO:81, SEQ ID NO:104, SEQ ID NO:131 through SEQ ID NO:133, SEQ ID NO:154 through SEQ ID NO:156, SEQ ID NO:176, SEQ ID NO:199, SEQ ID NO:267, SEQ ID NO:290, SEQ ID NO:312, SEQ ID NO:335, SEQ ID NO:364 through SEQ ID NO:376, and SEQ ID NO:391 through SEQ ID NO:392.

64. The kit of Claim 51, wherein said common polynucleotide probe comprises a sequence complementary to or homologous to a sequence selected from the group consisting of SEQ ID NO:1 through SEQ ID NO:80.

65. The kit of Claim 51, wherein said specific polynucleotide probe is a polynucleotide strand comprising a sequence selected from the group consisting of SEQ ID NO:473, SEQ ID NO:600, SEQ ID NO:607, SEQ ID NO:615, SEQ ID NO:622, SEQ ID NO:637, SEQ ID NO:730, SEQ ID NO:747, SEQ ID NO:748, SEQ ID NO:488, SEQ ID NO:513, SEQ ID NO:630, AND SEQ ID NO:639, a sequence complementary to any of such sequences, and a sequence capable of hybridizing with any of these sequences.

66. The kit of Claim 51, wherein said common polynucleotide probe is a polynucleotide strand comprising a sequence complementary to a sequence selected from the group consisting of SEQ ID NO: 728, SEQ ID NO:729, SEQ ID NO:733, SEQ ID NO:734, SEQ ID NO:739, SEQ ID NO:740,

-115-

SEQ ID NO:741, SEQ ID NO:742, SEQ ID NO:743, and SEQ ID NO:744, a sequence complementary to any of such sequence, and a sequence capable of hybridizing with any of these sequences.

67. The kit of Claim 51, wherein said specific polynucleotide probe is a polynucleotide strand comprising a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:758, a sequence complementary to SEQ ID NO:758, and a sequence capable of hybridizing with SEQ ID NO:758.

68. The kit of Claim 51, wherein said specific polynucleotide probe is a polynucleotide strand comprising a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:759, a sequence complementary to SEQ ID NO:759, and a sequence capable of hybridizing with SEQ ID NO:759.

69. The kit of Claim 51, wherein said specific polynucleotide probe is a polynucleotide strand comprising a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:751, SEQ ID NO:553, SEQ ID NO:670, SEQ ID NO:752, SEQ ID NO:753, SEQ ID NO:565, SEQ ID NO:678, SEQ ID NO:686, SEQ ID NO:754, SEQ ID NO:577, SEQ ID NO:697, SEQ ID NO:704, SEQ ID NO:755, SEQ ID NO:756, SEQ ID NO:732, SEQ ID NO:642, SEQ ID NO:652, SEQ ID NO:757, SEQ ID NO:593, SEQ ID NO:710, and SEQ ID NO:721, a sequence complementary to any of such sequences, and a sequence capable of hybridizing with any of these sequences.

70. The kit of Claim 51, wherein said common polynucleotide probe is a polynucleotide strand comprising a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:528, SEQ ID NO:731, SEQ ID NO:749, and SEQ ID NO:750, a sequence complementary to any of such sequences, and a sequence capable of hybridizing with any of these sequences.

71. An isolated oligonucleotide eight nucleotides or longer that is useful in the detection and quantification of jun oncogenes, said nucleotide containing a sequence specific to a particular jun oncogene, said sequence being selected from the group consisting of SEQ ID NO:473, SEQ ID NO:600, SEQ ID NO:607, SEQ ID NO:615, SEQ ID NO:622, SEQ ID NO:637, SEQ ID NO:730, SEQ ID NO:747, SEQ ID NO:748, SEQ ID NO:488, SEQ ID NO:513, SEQ ID NO:630, AND SEQ ID NO:639, a sequence complementary to any of such sequences, and a sequence capable of hybridizing with any of these sequences.

72. An isolated oligonucleotide eight nucleotides or longer that is useful in the detection and quantification of jun oncogenes, said nucleotide containing a sequence common to a plurality of jun oncogenes, said sequence being selected from the group consisting of SEQ ID NO: 728, SEQ ID NO:729, SEQ ID NO:733, SEQ ID NO:734, SEQ ID NO:739, SEQ ID NO:740, SEQ ID NO:741, SEQ ID NO:742, SEQ ID NO:743, and SEQ ID NO:744, a sequence complementary to any of such sequence, and a sequence capable of hybridizing with any of these sequences.

73. An isolated oligonucleotide eight nucleotides or longer that is useful in the detection and quantification of β receptors, said sequence being selected from the group consisting of SEQ ID

NO:759, a sequence complementary to SEQ ID NO:759, and a sequence capable of hybridizing with SEQ ID NO:759.

74. An isolated oligonucleotide eight nucleotides or longer that is useful in the detection and quantification of Substance P receptors, said sequence being selected from the group consisting of SEQ ID NO:758, a sequence complementary to SEQ ID NO:758, and a sequence capable of hybridizing with SEQ ID NO:758.

75. An isolated oligonucleotide eight nucleotides or longer that is useful in the detection and quantification of G proteins, said nucleotide containing a sequence specific to a particular G protein, said sequence being selected from the group consisting of SEQ ID NO:751, SEQ ID NO:553, SEQ ID NO:670, SEQ ID NO:752, SEQ ID NO:753, SEQ ID NO:565, SEQ ID NO:678, SEQ ID NO:686, SEQ ID NO:754, SEQ ID NO:577, SEQ ID NO:697, SEQ ID NO:704, SEQ ID NO:755, SEQ ID NO:756, SEQ ID NO:732, SEQ ID NO:642, SEQ ID NO:652, SEQ ID NO:757, SEQ ID NO:593, SEQ ID NO:710, and SEQ ID NO:721, a sequence complementary to any of such sequences, and a sequence capable of hybridizing with any of these sequences.

76. An isolated oligonucleotide eight nucleotides or longer that is useful in the detection and quantification of G proteins, said nucleotide containing a sequence common to a plurality of G proteins, said sequence being selected from the group consisting of SEQ ID NO:528, SEQ ID NO:731, SEQ ID NO:749, and SEQ ID NO:750, a sequence complementary to any of such sequences, and a sequence capable of hybridizing with any of these sequences.

77. An isolated segment of polynucleotide specific to the rRNA of a particular fungus, said sequence being complementary or homologous to a sequence selected from the group consisting of SEQ ID NO:81, SEQ ID NO:104, SEQ ID NO:131 through SEQ ID NO:133, SEQ ID NO:154 through SEQ ID NO:156, SEQ ID NO:176, SEQ ID NO:199, SEQ ID NO:267, SEQ ID NO:290, SEQ ID NO:312, SEQ ID NO:335, SEQ ID NO:364 through SEQ ID NO:376, SEQ ID NO:391 through SEQ ID NO:392 and a sequence homologous to any of the foregoing sequences.

78. An isolated segment of polynucleotide coding for or complementary to a sequence common to a plurality of fungal species, said sequence being complementary to or homologous to a sequence selected from the group consisting of SEQ ID NO:1 through SEQ ID NO:80 and a sequence homologous to any of the foregoing sequences.

79. A method of detecting the presence of one or more organisms, infectious agents, or biological components in a biological sample containing polynucleotides, wherein at least one of said polynucleotides is indicative of the presence of said one or more organisms, infectious agents or biological components and is present in minute quantities, comprising the steps of:

(a) obtaining a biological sample containing polynucleotides;

(b) contacting said sample with a first polynucleotide primer, said first primer having a nucleotide sequence complementary to a nucleotide sequence common to a plurality of organisms, infectious agents, or biological components;

-117-

(c) hybridizing said first primer to an analyte polynucleotide present in said sample that is complementary to said first primer, if such an analyte polynucleotide is present;

(d) extending said first primer, thereby producing a double-stranded polynucleotide including a complementary nucleotide strand comprising said first primer and having a nucleotide sequence complementary to said analyte polynucleotide;

(e) contacting said sample with a second polynucleotide primer, said second primer being complementary to a sequence contained in said complementary nucleotide strand;

(f) hybridizing said second primer to said complementary nucleotide strand;

(g) extending said second primer to form a nucleotide strand homologous to said analyte polynucleotide;

(h) contacting said sample with a third polynucleotide primer, said third primer having a sequence complementary to said homologous nucleotide strand, wherein said third primer has a nucleotide sequence complementary to a sequence that is specific to a particular organism, infectious agent, or biological component whose presence is to be determined;

(i) hybridizing said third primer to said homologous nucleotide strand;

(j) extending said third primer, thereby producing a double-stranded polynucleotide; and

(k) determining the presence of the particular organism, infectious agent, or biological component in said sample by detecting the extension of said third primer.

80. The method of Claim 79, wherein steps (c), (d), (f), (g), (i), and (j) are repeated a plurality of times.

81. The method of Claim 79, wherein step (d) comprises extension with a reverse transcriptase, and step (g) comprises extension with a DNA polymerase.

82. The method of Claim 81, wherein said DNA polymerase has significant polymerase activity at temperatures above 50°C.

83. The method of Claim 79, wherein the nucleotide sequence of said first primer is determined by a computer-assisted method.

84. The method of Claim 83, wherein said computer-assisted method determines the sequence of said first nucleotide probe using an H-Site model.

85. The method of Claim 79, wherein the nucleotide sequence of said second primer is determined by a computer-assisted method.

86. The method of Claim 85, wherein said computer-assisted method determines the sequence of said second nucleotide probe using an H-Site model.

87. The method of Claim 79, wherein said third primer includes a label, and wherein step (k) comprises detecting the extension of said labeled primer.

-118-

88. The method of Claim 87, wherein said label is selected from the group consisting of a radionuclide, an enzyme, an enzyme substrate, a specific binding moiety, a binding partner for a specific binding moiety, biotin, avidin, a nucleic acid stain, and a fluorescent material.

5 89. The method of Claim 79, wherein the nucleotide sequence of said third primer is determined by a computer-assisted method.

90. The method of Claim 89, wherein said computer-assisted method determines the sequence of said third primer using an H-Site model.

10 91. The method of Claim 79, wherein said second primer has a nucleotide sequence that is common to a plurality of the organisms, infectious agents, or biological components whose presence is being determined.

15 92. The method of Claim 79, additionally comprising the steps of:
contacting said sample with a fourth primer, said fourth primer having a sequence complementary to said complementary nucleotide strand;
hybridizing said fourth primer to said complementary nucleotide strand; and
extending said fourth primer.

1/61

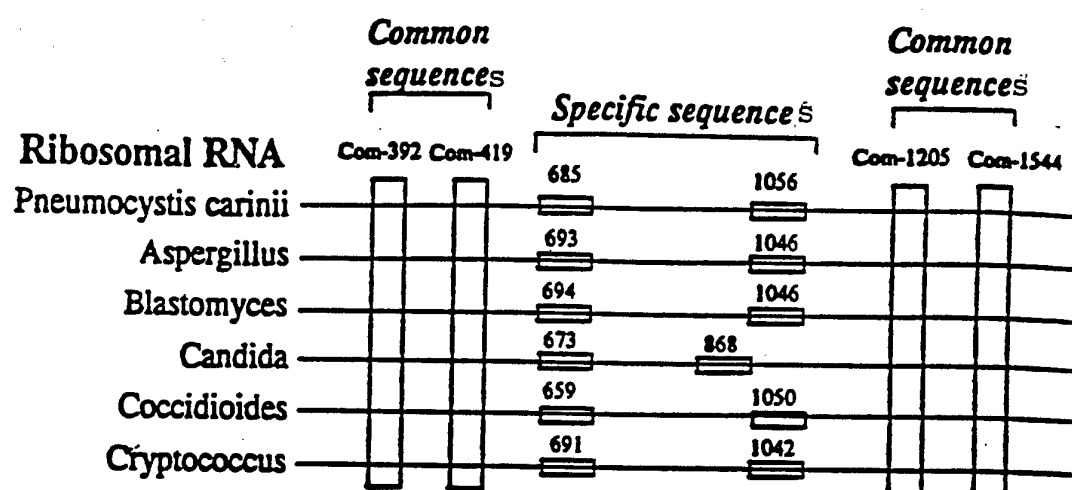


FIG. 1

2/61

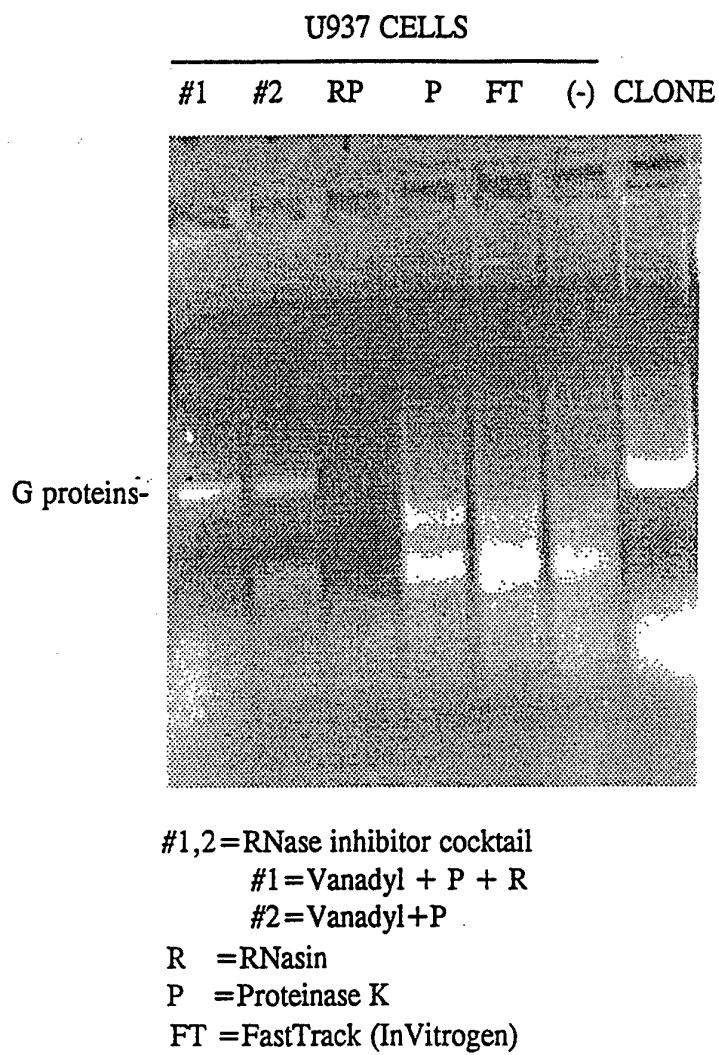


FIG. 1A

3/61

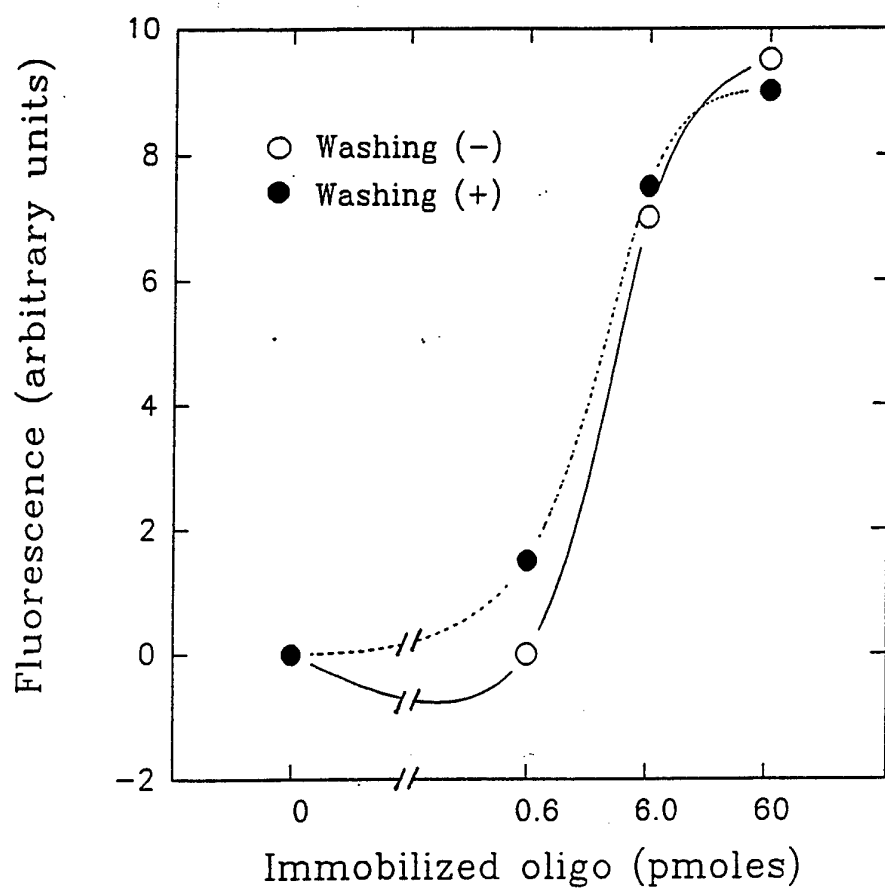


FIG. 1B

4/61

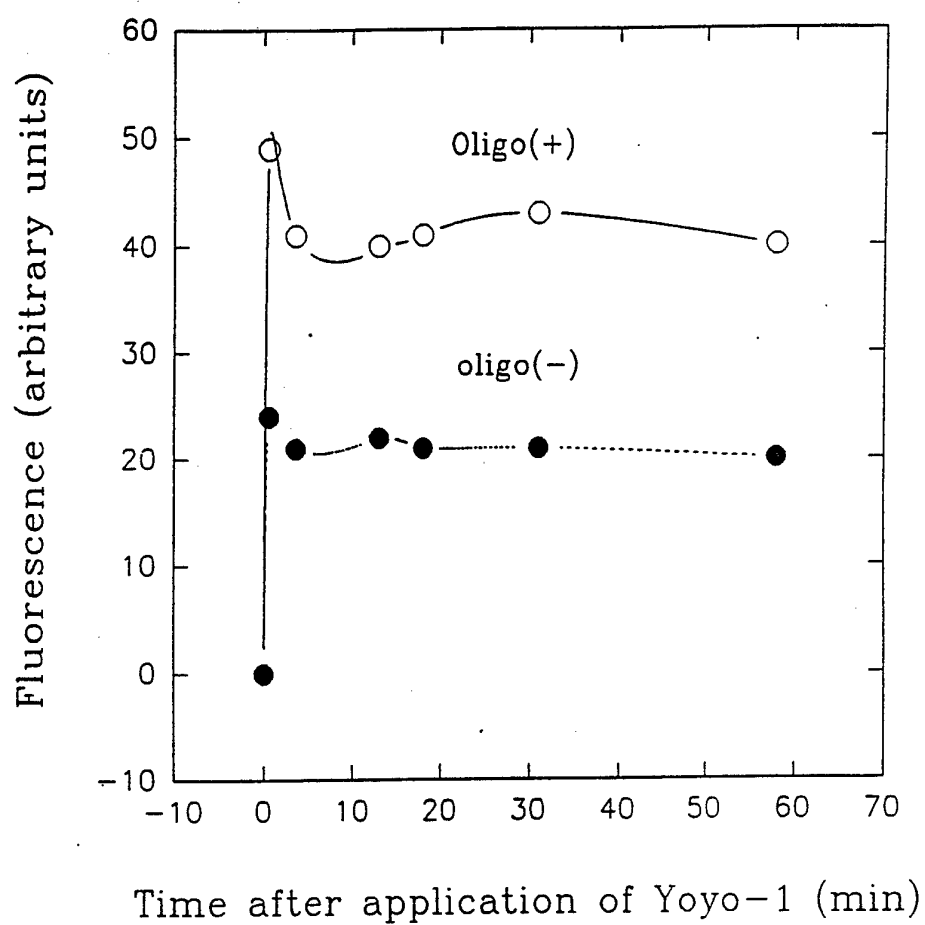


FIG. 1C

5/61

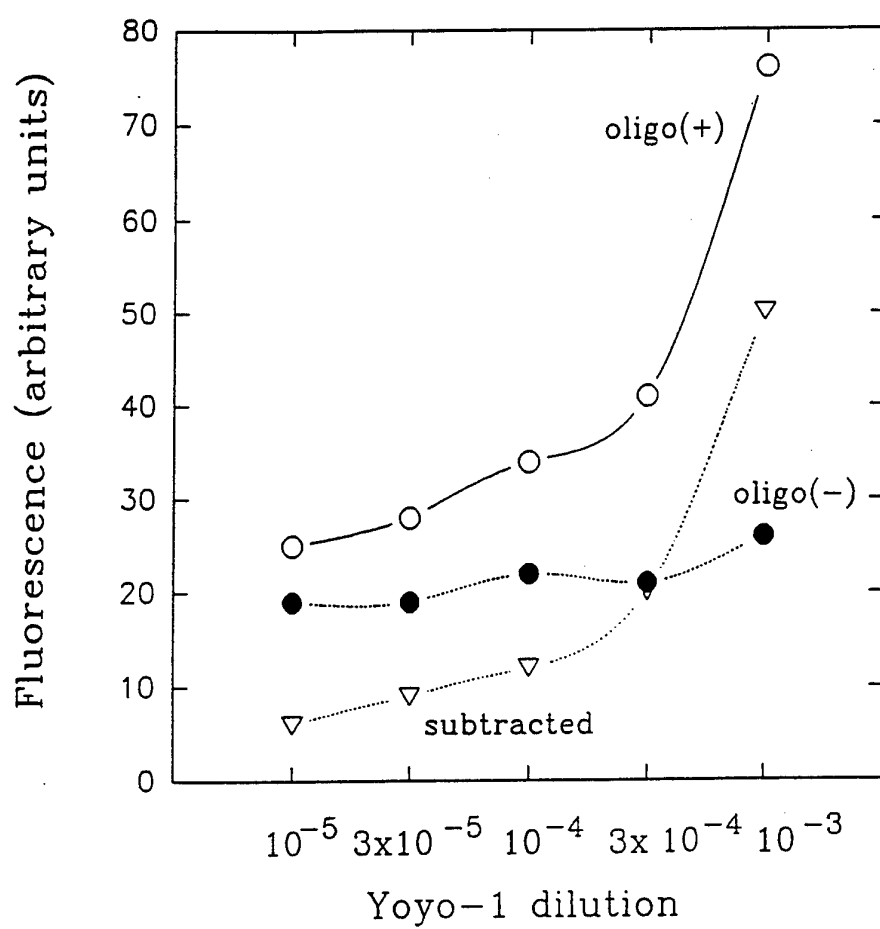


FIG. 1D

6/61

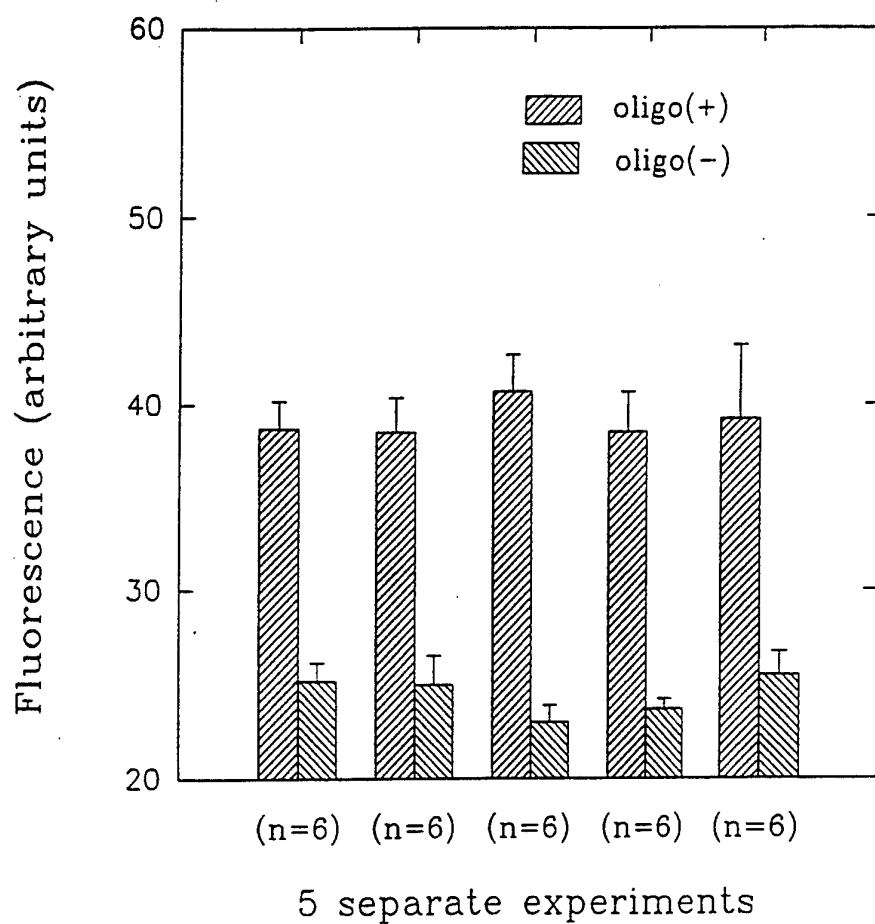
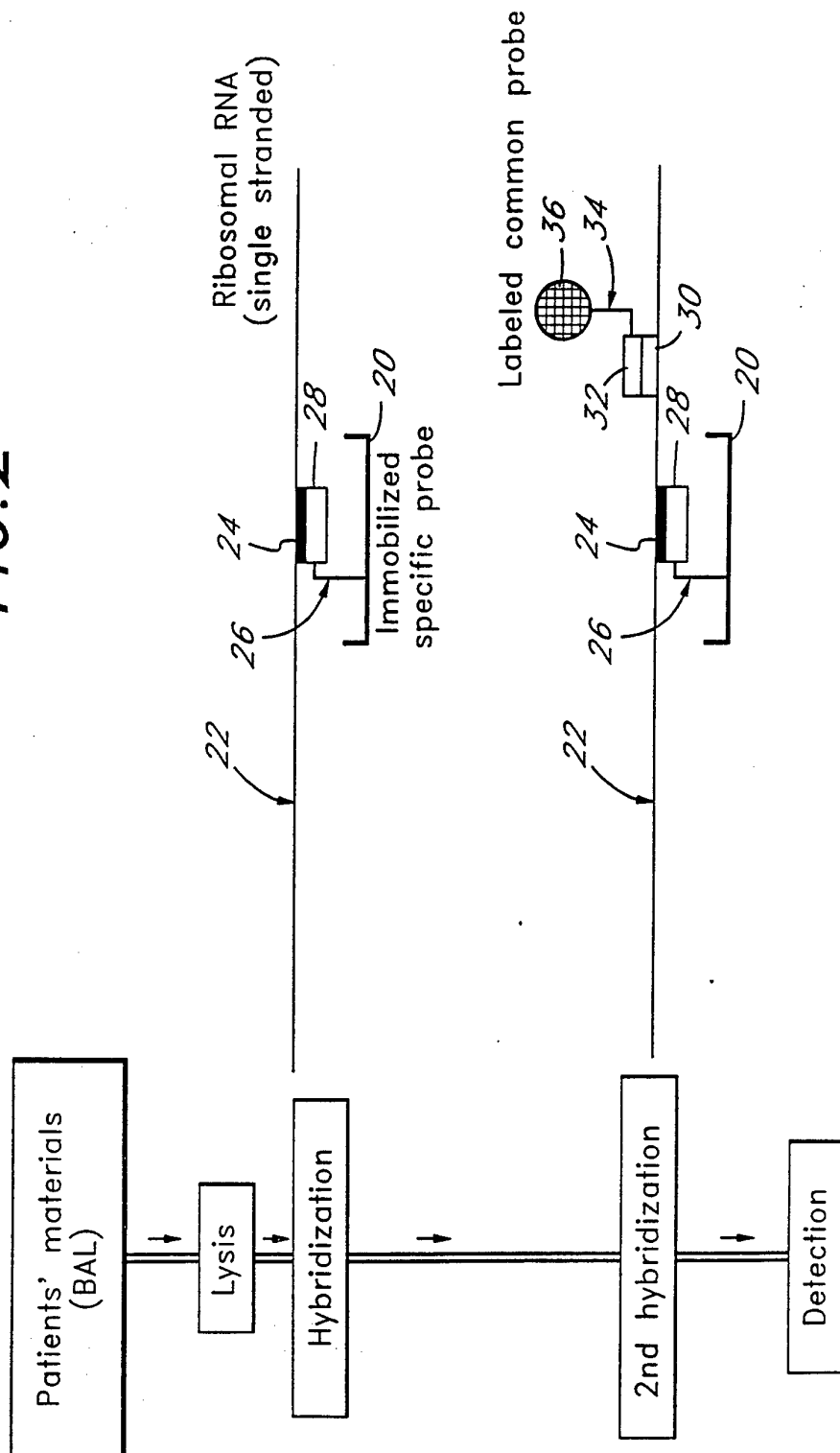


FIG. 1E

7/61

FIG. 2



8/61

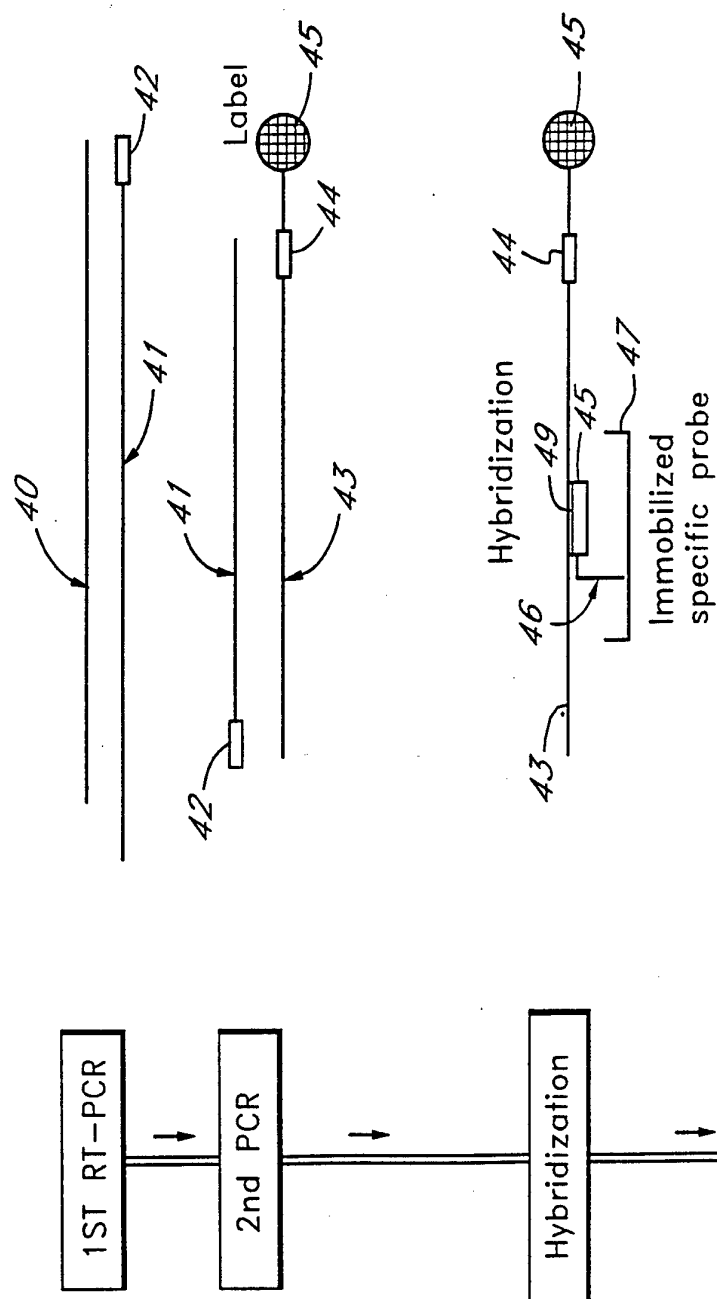


FIG. 3

9/61

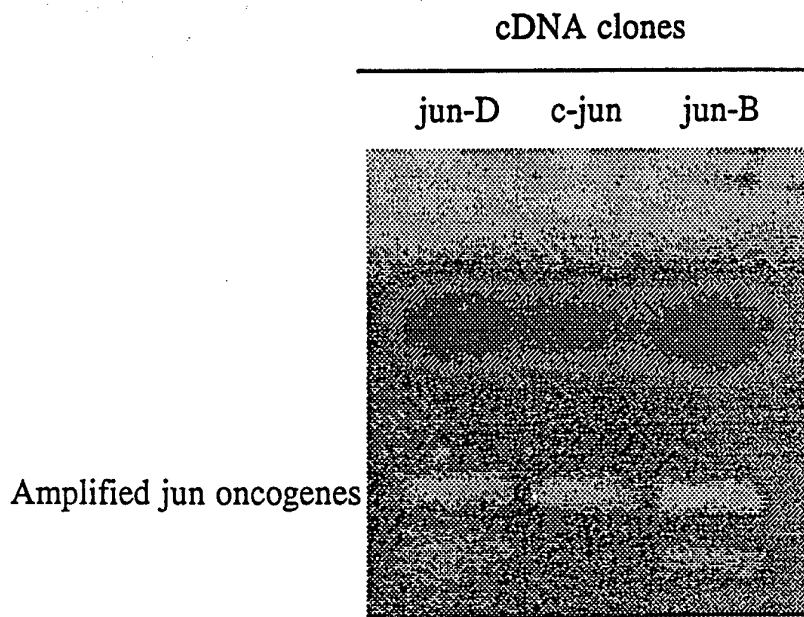


FIG. 4

10/61

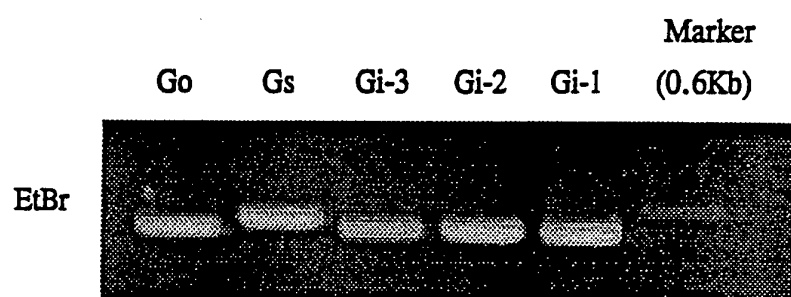


FIG. 5

11/61

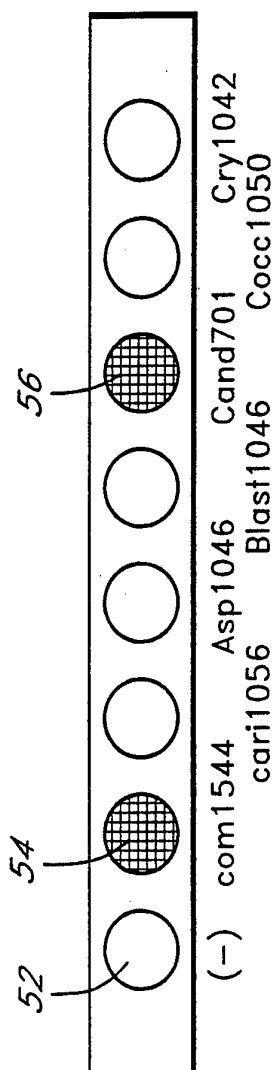


FIG. 6

12/61

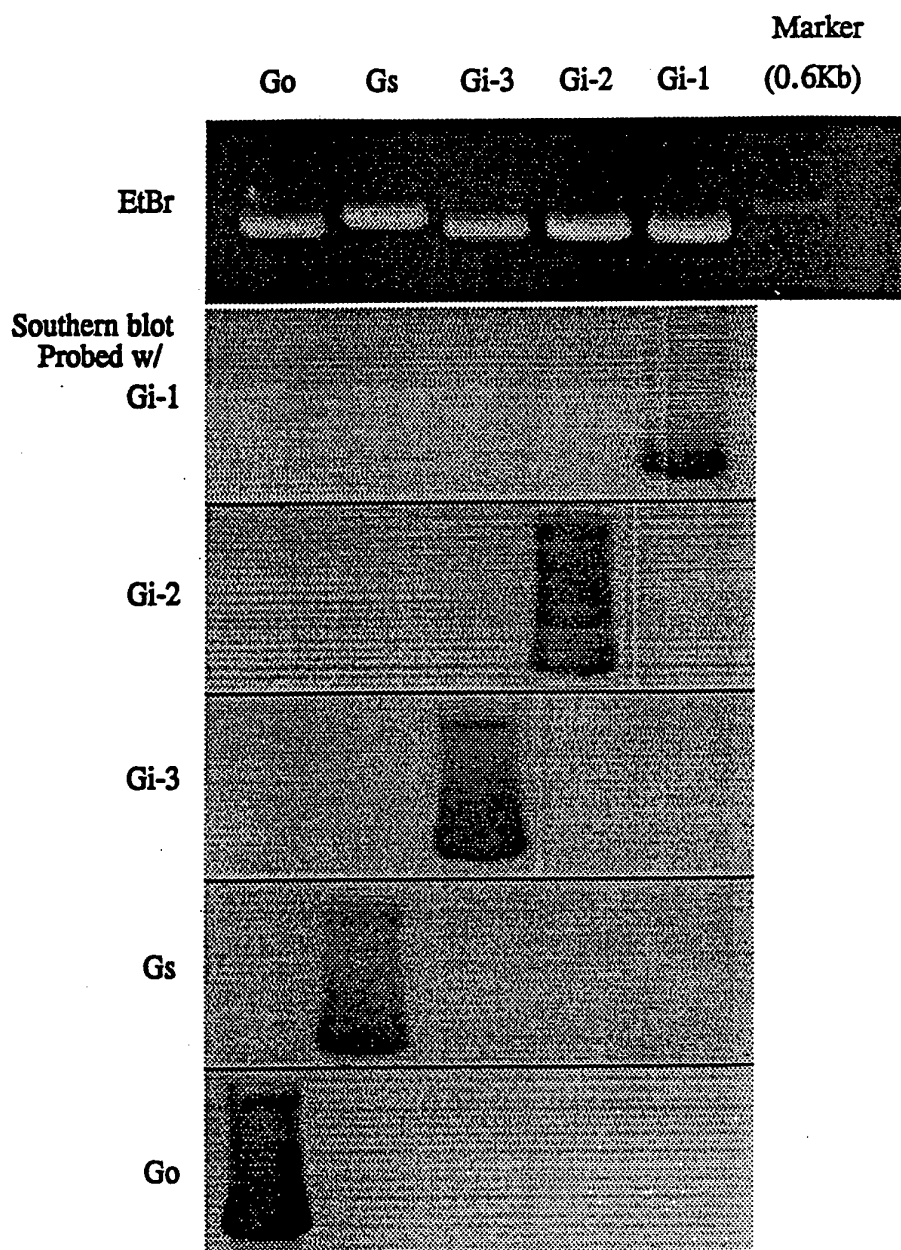


FIG. 7

13/61

FIG. 7A

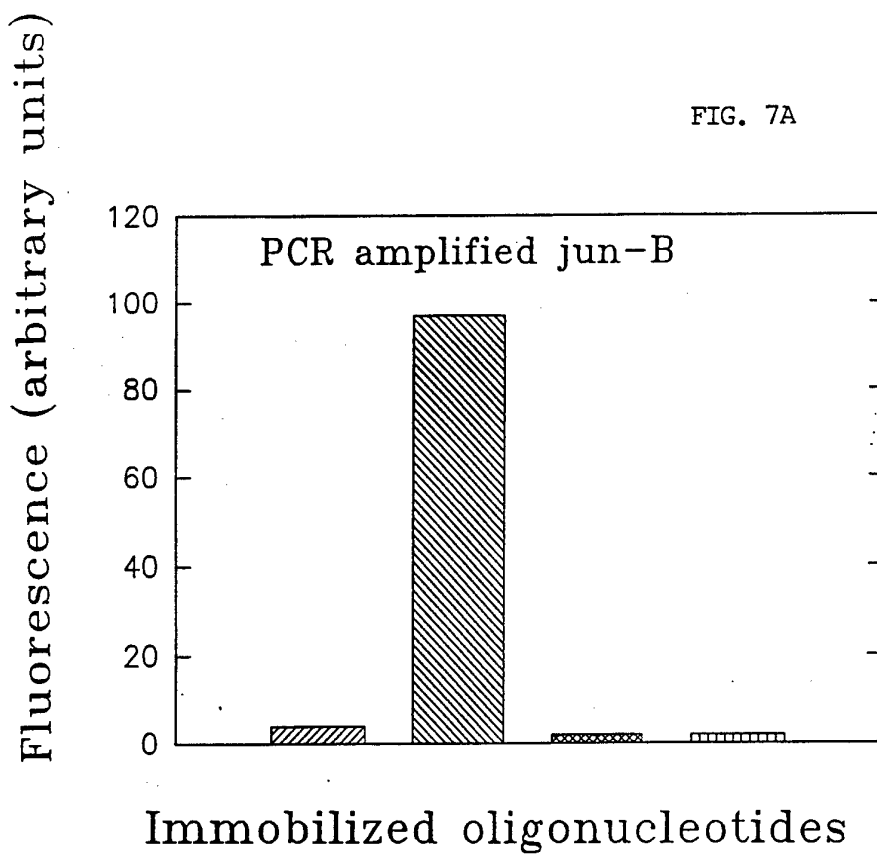
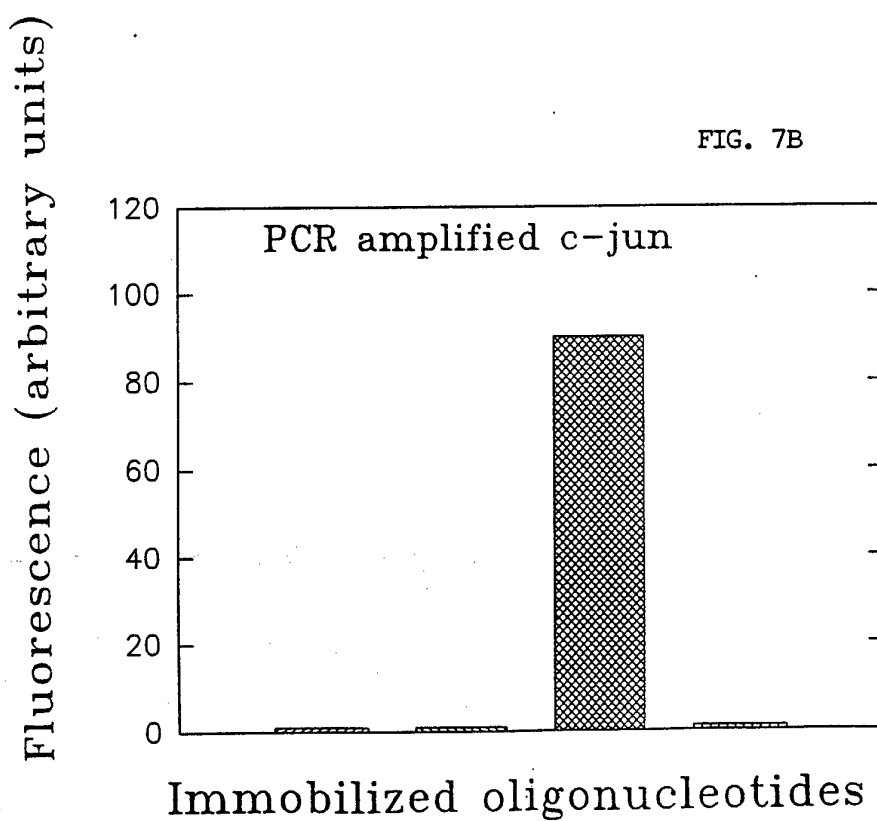
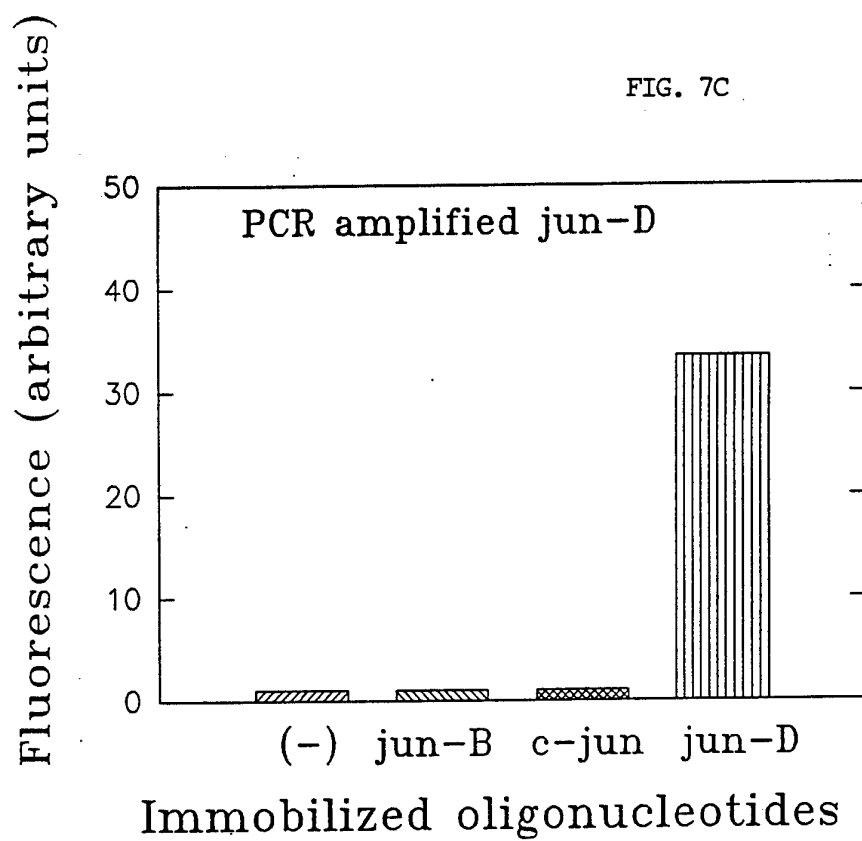


FIG. 7B



14/61



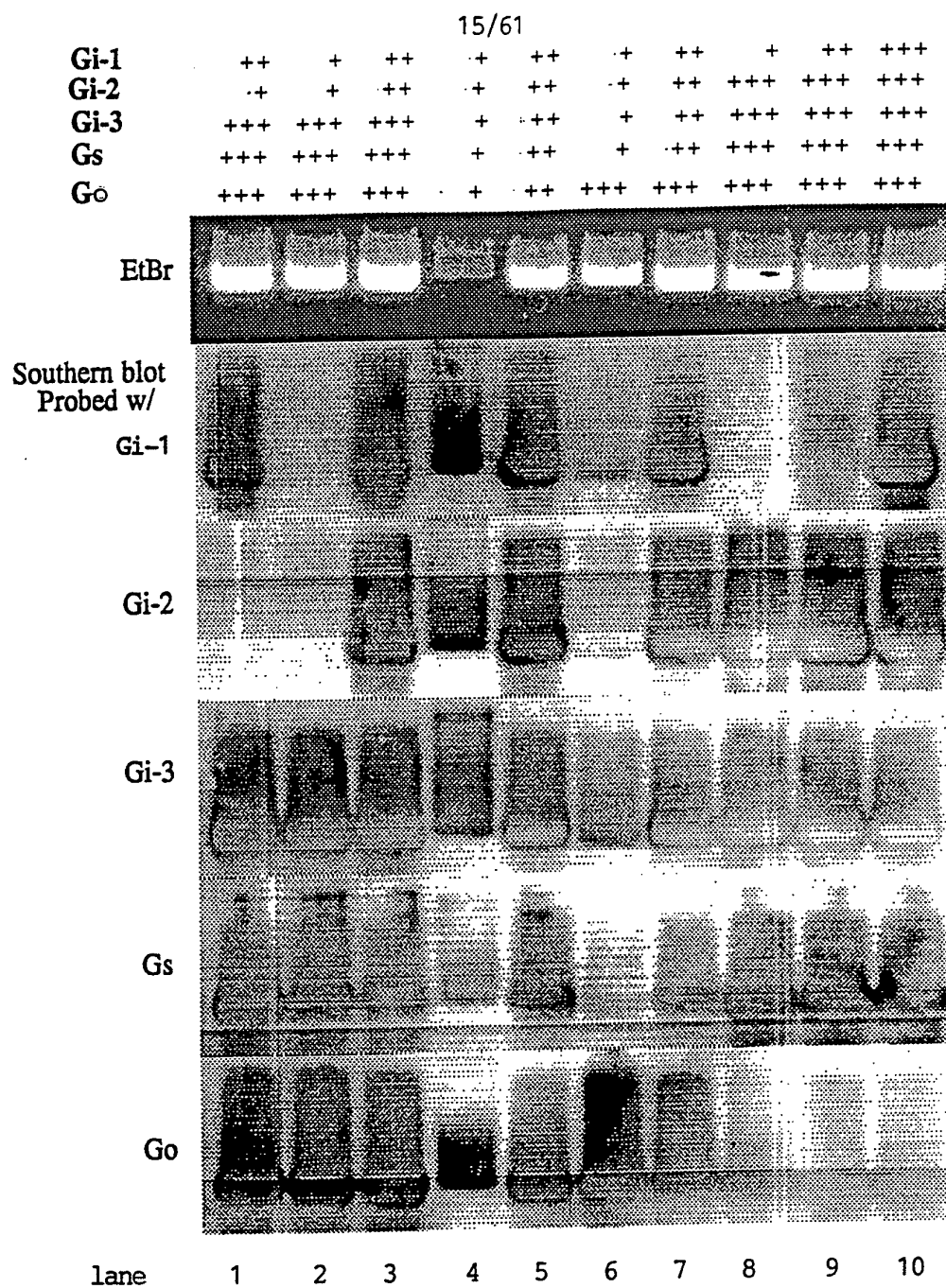


FIG. 8

16/61

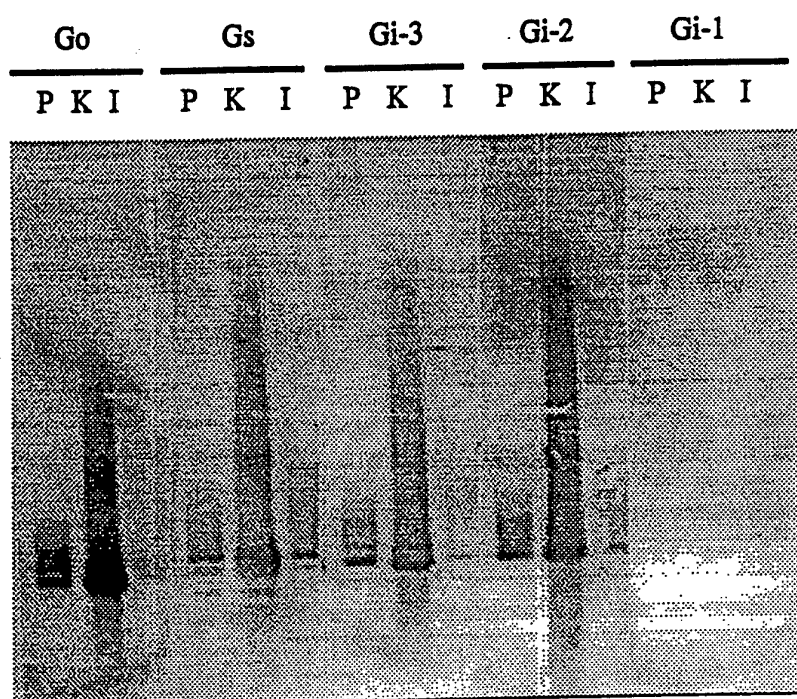


FIG. 9

17/61

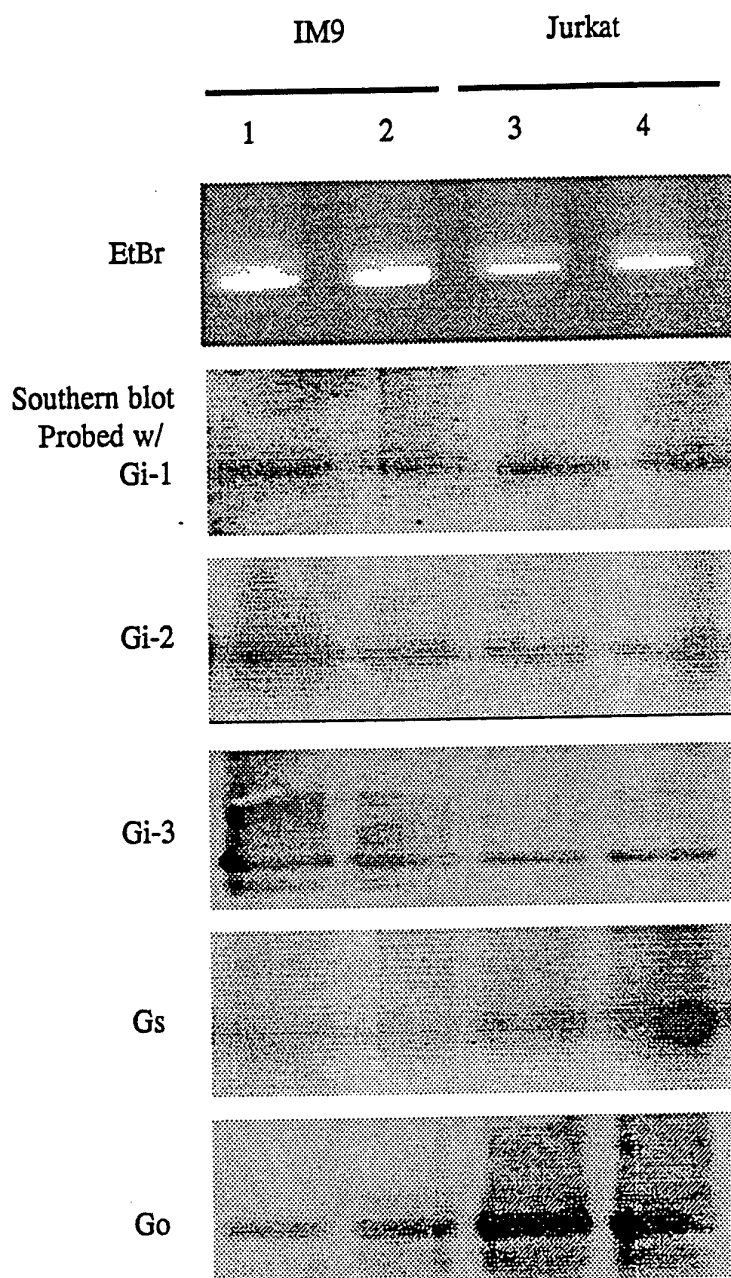
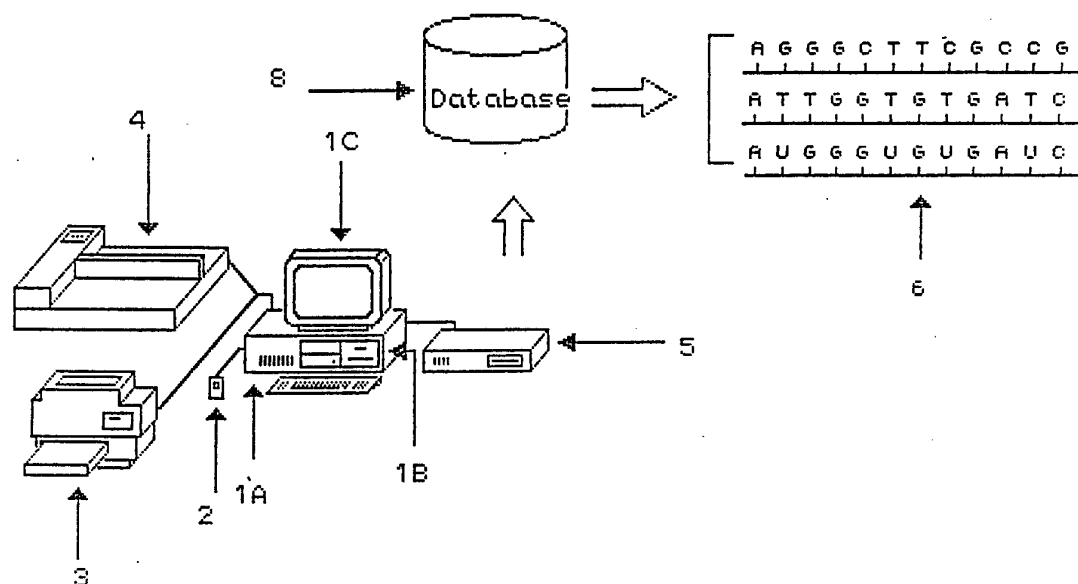


FIG. 10

18/61

FIG. 11



19/61

FIG. 12A

Oligo Probe Design Station				
File	Preparation	Models	Experiment	Help
↑	↑	↑	↑	↑
10	80	70	90	50

20/61

FIG. 12B

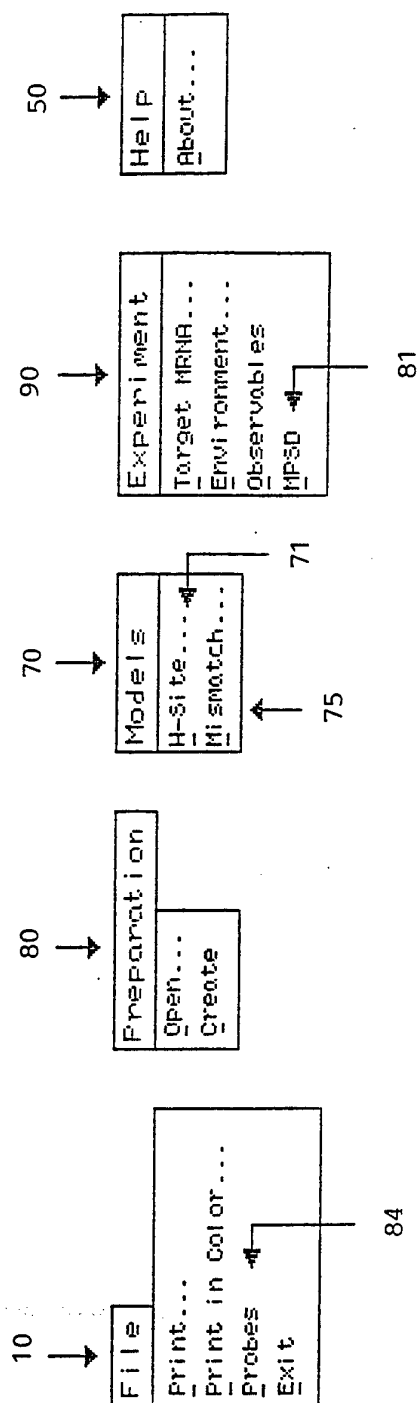
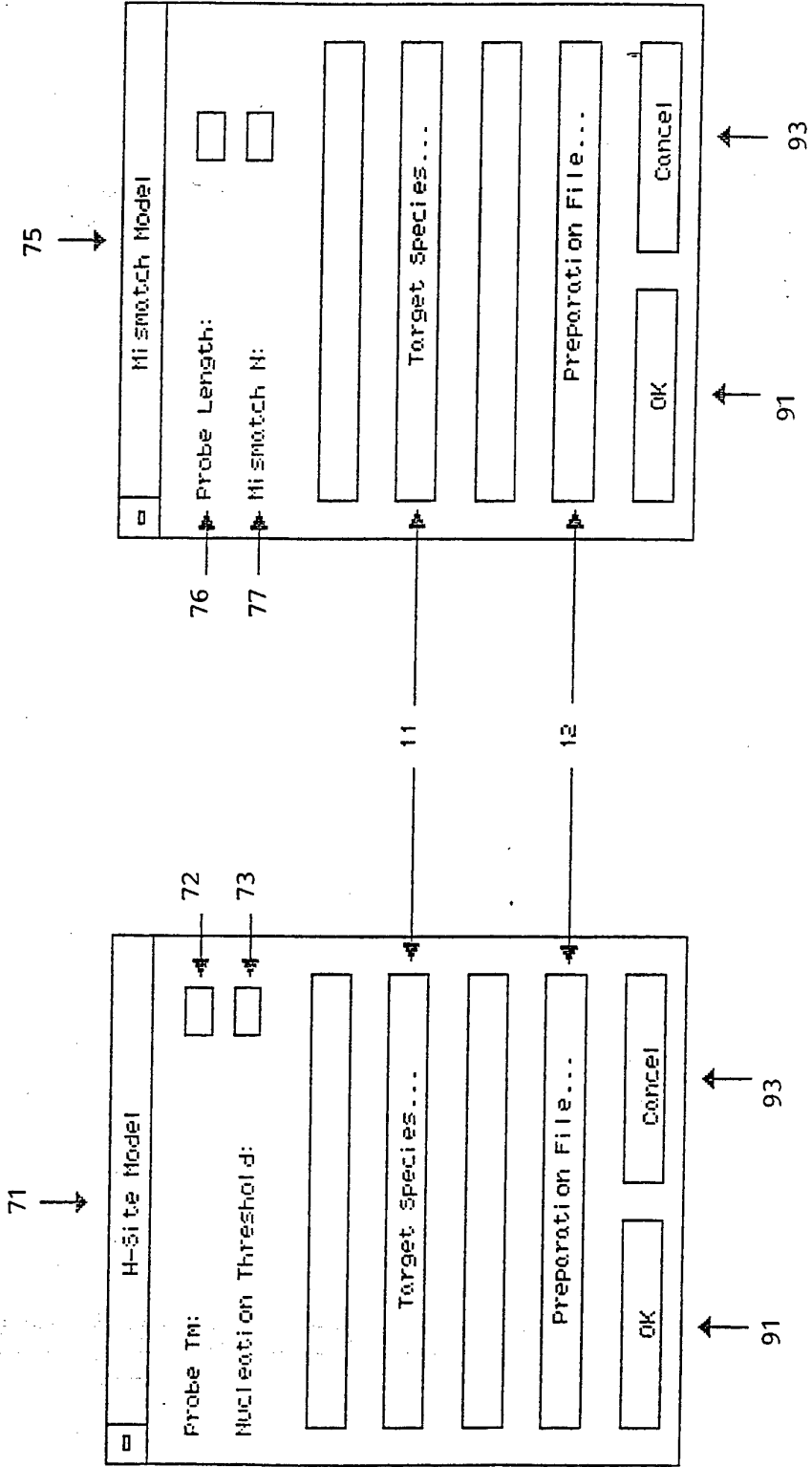
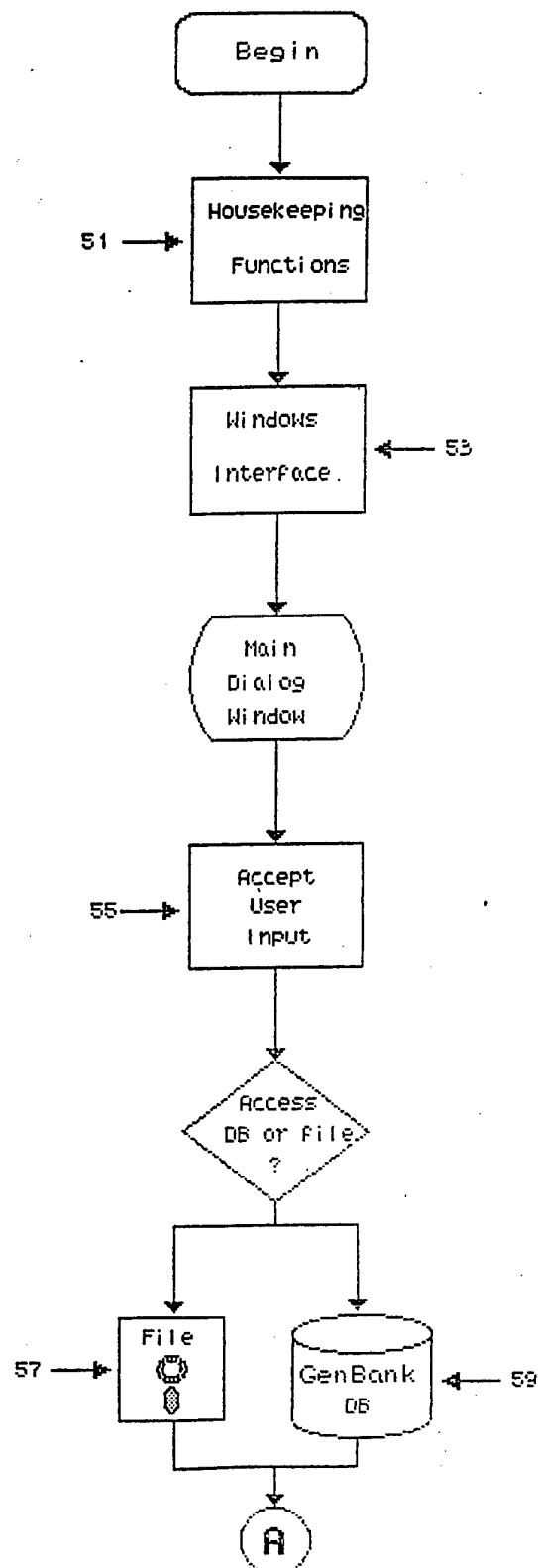


FIG. 12C



22/61

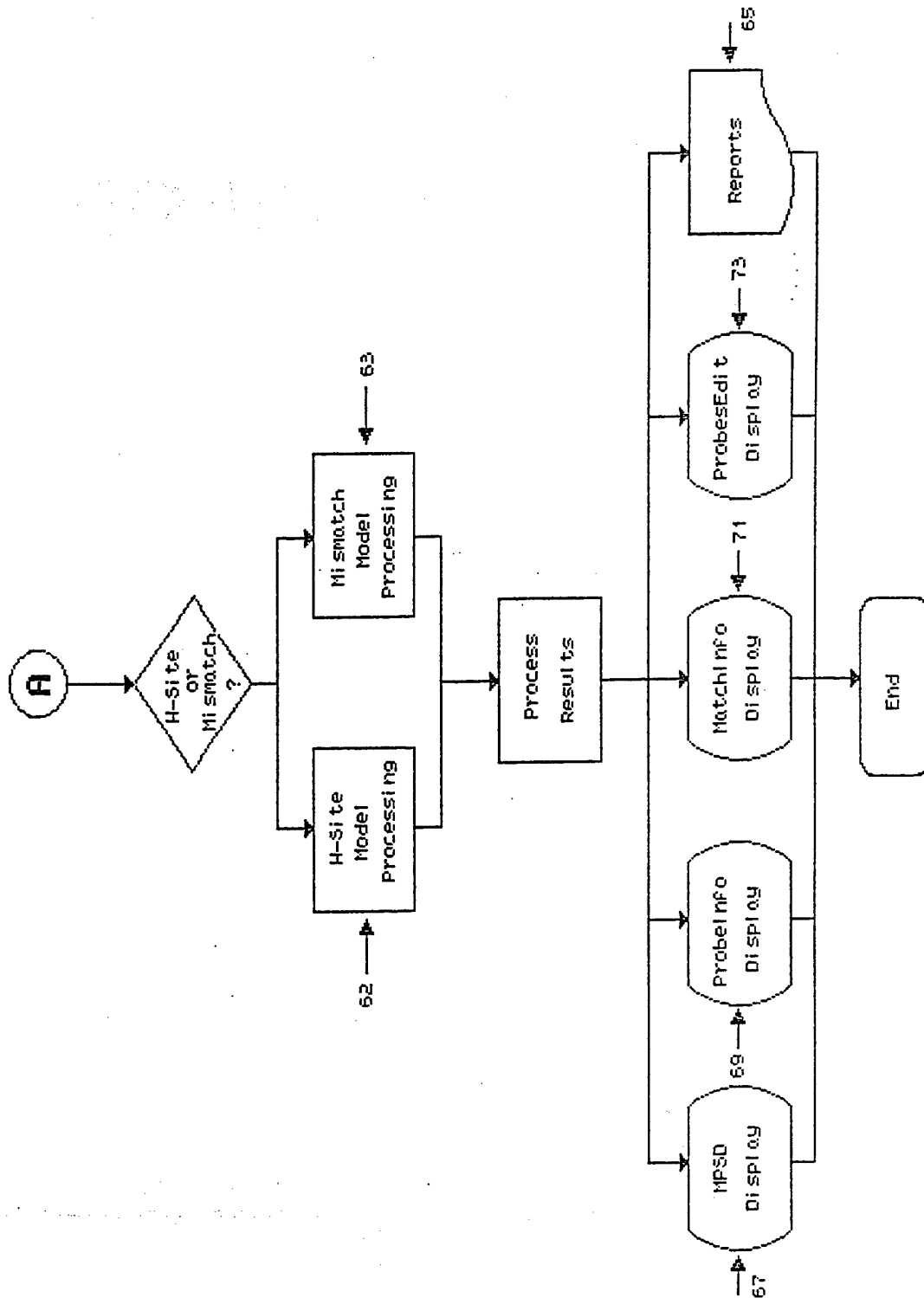
FIG. 13A



SUBSTITUTE SHEET

23/61

FIG. 13B



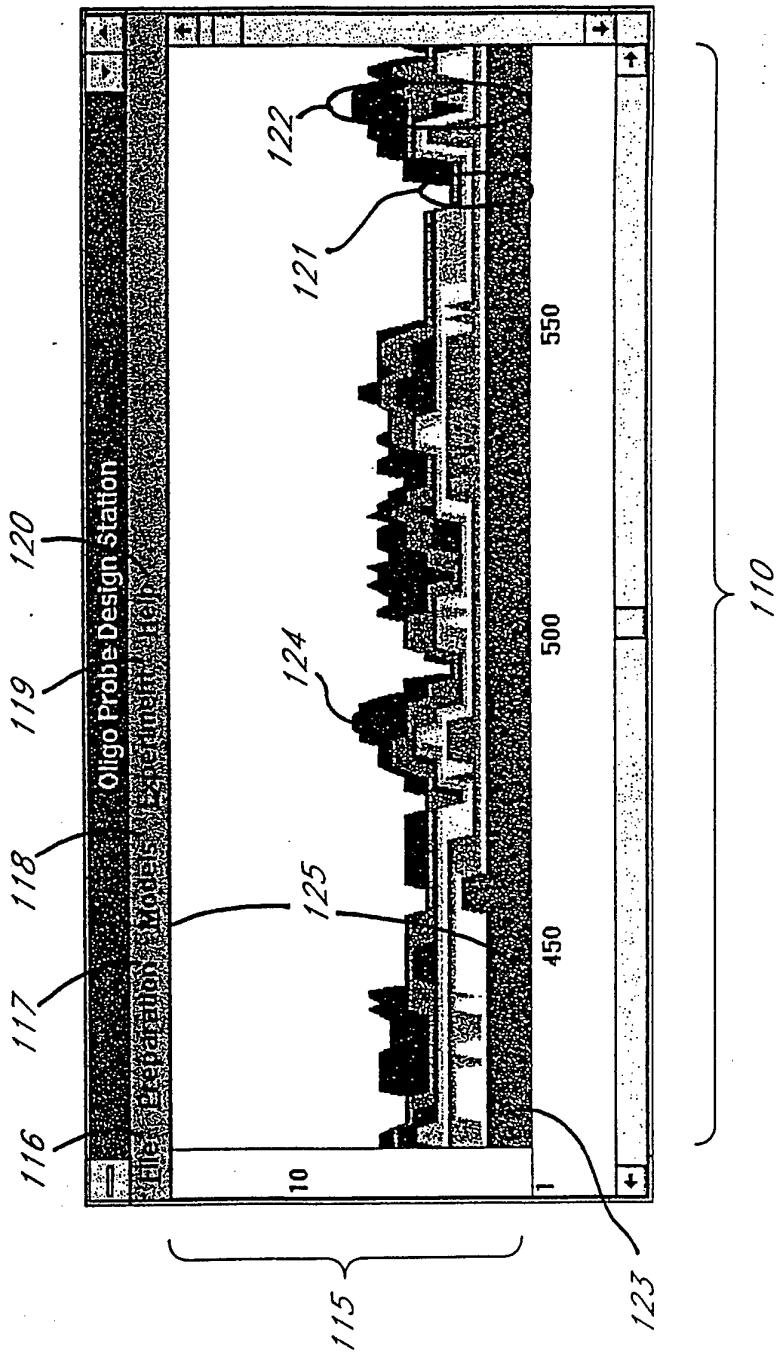


FIG. 14

26/61

FIG. 16

Probes Selected - JUNMI X.prb		156
File		
PROBE: C:\HITACHI\JUNMI X.PRP HYBRIDIZATION: C:\HITACHI\HUMBJU Length = 374 Hairpin = 3 5 Locus Pos Tm humbjunx 374 61.47 ----- musbjunx 365 61.47 ----- humdjunx 41 34.82 t-----g-g- humbjunx 162 31.12 a-----gtgg humdjunx 602 31.12 c-----c-gg		↑
PROBE: C:\HITACHI\JUNMI X.PRP HYBRIDIZATION: C:\HITACHI\HUMBJU Length = 467 Hairpin = 2 13 Locus Pos Tm humbjunx 467 61.7 ----- musbjunx 456 51.6 -----c- humdjunx 32 29.35 tgagcgg----- humdjunx 32 29.35 tgagcgg-----		↓
155 →		

27/61

FIG. 16A

PROBE: C:\HITACHI\JUNMIX.PRP
 HYBRIDIZATION: C:\HITACHI\HUMBJUNX.CDS
 Length = 374 Hairpin = 3 5
 Locus Pos Tm
 humbjunx 374 61.47 -----
 musbjunx 365 61.47 -----
 humdjunx 41 34.82 t-----g-g--agt
 humbjunx 182 31.12 a-----gtgg--gc
 humdjunx 602 31.12 c-----c-ggg-gc
 humdjunx 602 31.12 c-----c-ggg-gc

PROBE: C:\HITACHI\JUNMIX.PRP
 HYBRIDIZATION: C:\HITACHI\HUMBJUNX.CDS
 Length = 377 Hairpin = 2 14
 Locus Pos Tm
 humbjunx 377 61.55 -----
 musbjunx 368 61.55 -----
 humdjunx 383 28.12 tg-cg-c--g-----
 musdjunx 383 28.12 tg-ca-c--g-----
 musdjunx 383 28.12 tg-ca-c--g-----

PROBE: C:\HITACHI\JUNMIX.PRP
 HYBRIDIZATION: C:\HITACHI\HUMBJUNX.CDS
 Length = 389 Hairpin = 3 3
 Locus Pos Tm
 humbjunx 389 61.7 -----
 muscjunx 314 56.65 -c-----
 musbjunx 380 50.85 -----t--g
 humcjunx 314 49.35 -t-----g-----
 humdjunx 395 33.85 -----tt-gc--ag
 musdjunx 395 33.85 -----tt-gc--aa
 humcjunx 326 32.35 g-ttcgcc-----tg
 humdjunx 404 32.35 --ttcgcc-----t-
 muscjunx 326 32.35 gcttcgcc-----tg
 musdjunx 253 30.85 gacg-gct-ct-----
 humbjunx 953 30.65 g-----t--c-cagct-
 musdjunx 83 27.3 cc-gcgggt-gt-----g

PROBE: C:\HITACHI\JUNMIX.PRP
 HYBRIDIZATION: C:\HITACHI\HUMBJUNX.CDS
 Length = 397 Hairpin = 4 1
 Locus Pos Tm
 humbjunx 397 61.55 -----
 muscjunx 322 53.44 -----g---
 humcjunx 322 45.33 -----g-----g---
 musbjunx 388 41.38 -----t--g-----t
 humdjunx 214 36.83 cccctgc-----
 humdjunx 99 36.16 cg----gc-c-----
 musdjunx 261 34.55 -ct-----gatct
 humdjunx 400 33.27 c---ag-----g---
 musdjunx 400 33.27 c---ag-----a---
 humcjunx 334 32.28 -----tgcg--c-
 humdjunx 412 32.28 -----t-a-g-c-
 muscjunx 334 32.28 -----tgcg--c-
 humbjunx 658 30.17 cc-cc-----gt---
 humdjunx 241 28.95 -c--cacc-c-----
 humdjunx 342 28.95 c-cca-ca-----ag
 musbjunx 606 28.95 ---ct-a-ac-----
 musdjunx 229 28.95 -c-ctgcg-c-----
 musdjunx 91 26.67 -gt-----gcc-ccg

SUBSTITUTE SHEET

28/61

FIG. 16B

PROBE: C:\HITACHI\JUNMIX.PRP
 HYBRIDIZATION: C:\HITACHI\HUMBJUNX.CDS
 Length = 417 Hairpin = 2 15
 Locus Pos Tm
 humbjunx 417 60.08 -----
 musbjunx 408 55.52 -----c-----
 humdjunx 420 37.3 c-----g-----g---t-a-
 musbjunx 61 29.0 g---gg-----ca-cctgt-
 muscjunx 672 26.27 gc-gc-----a-g--aga--

PROBE: C:\HITACHI\JUNMIX.PRP
 HYBRIDIZATION: C:\HITACHI\HUMBJUNX.CDS
 Length = 461 Hairpin = 4 9
 Locus Pos Tm
 humbjunx 461 61.63 -----
 musbjunx 452 61.63 -----
 musbjunx 452 61.63 -----

PROBE: C:\HITACHI\JUNMIX.PRP
 HYBRIDIZATION: C:\HITACHI\HUMBJUNX.CDS
 Length = 467 Hairpin = 2 13
 Locus Pos Tm
 humbjunx 467 61.7 -----
 musbjunx 458 51.6 -----c-g-
 humdjunx 32 29.35 tgagcgg-----gcgg-
 humdjunx 32 29.35 tgagcgg-----gcgg-

PROBE: C:\HITACHI\JUNMIX.PRP
 HYBRIDIZATION: C:\HITACHI\HUMBJUNX.CDS
 Length = 477 Hairpin = 2 4
 Locus Pos Tm
 humbjunx 477 61.37 -----
 humdjunx 489 34.93 c-c---cg-----
 humdjunx 489 34.93 c-c---cg-----

PROBE: C:\HITACHI\JUNMIX.PRP
 HYBRIDIZATION: C:\HITACHI\HUMBJUNX.CDS
 Length = 487 Hairpin = 3 3
 Locus Pos Tm
 humbjunx 487 61.14 -----
 musdjunx 74 51.0 ct-----
 humdjunx 499 45.64 -----t---g
 humdjunx 527 30.72 cc-c-c-----
 musdjunx 97 30.72 ttc-c-----g
 musdjunx 580 30.72 -cc-----t-g
 musdjunx 637 30.72 cc-cc-----g
 musdjunx 637 30.72 cc-cc-----g

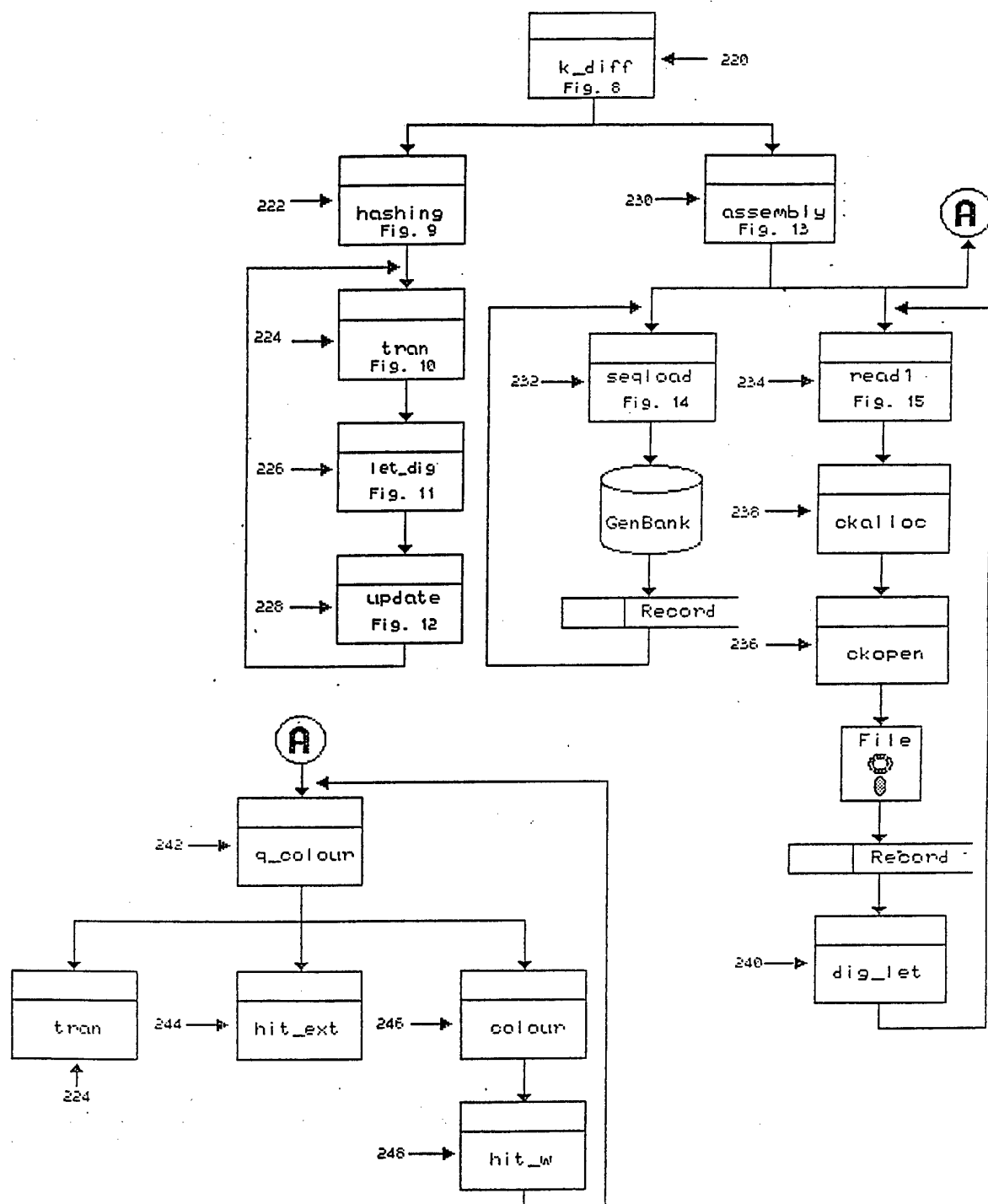
PROBE: C:\HITACHI\JUNMIX.PRP
 HYBRIDIZATION: C:\HITACHI\HUMBJUNX.CDS
 Length = 498 Hairpin = 3 2
 Locus Pos Tm
 humbjunx 498 61.26 -----
 humbjunx 498 61.26 -----

PROBE: C:\HITACHI\JUNMIX.PRP
 HYBRIDIZATION: C:\HITACHI\HUMBJUNX.CDS
 Length = 504 Hairpin = 3 2
 Locus Pos Tm
 humbjunx 504 61.47 -----
 musbjunx 495 40.35 c--a-----t-
 humdjunx 609 35.29 cg-----cgggg-
 humdjunx 609 35.29 cg-----cgggg-

SUBSTITUTE SHEET

29/61

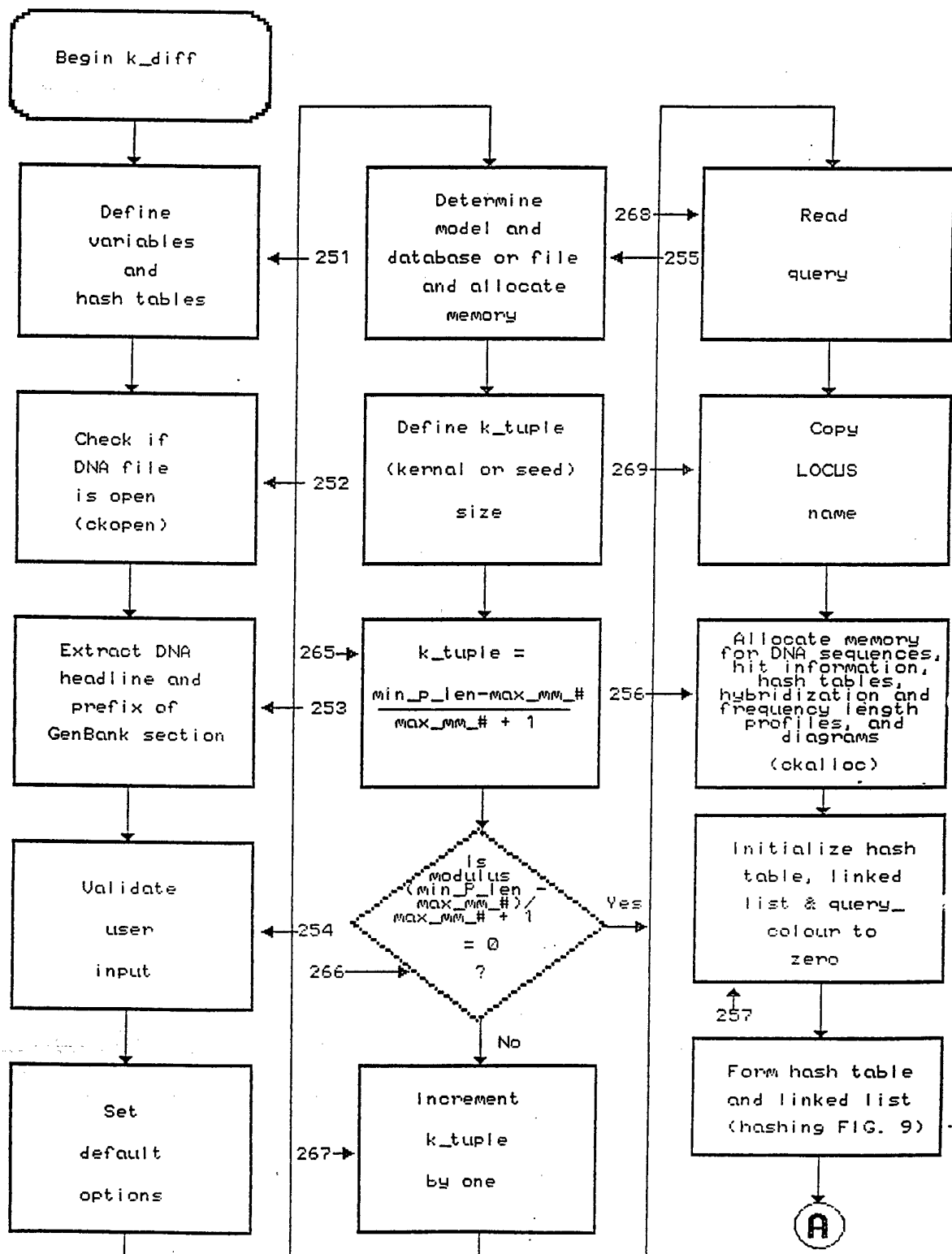
FIG. 17



SUBSTITUTE SHEET

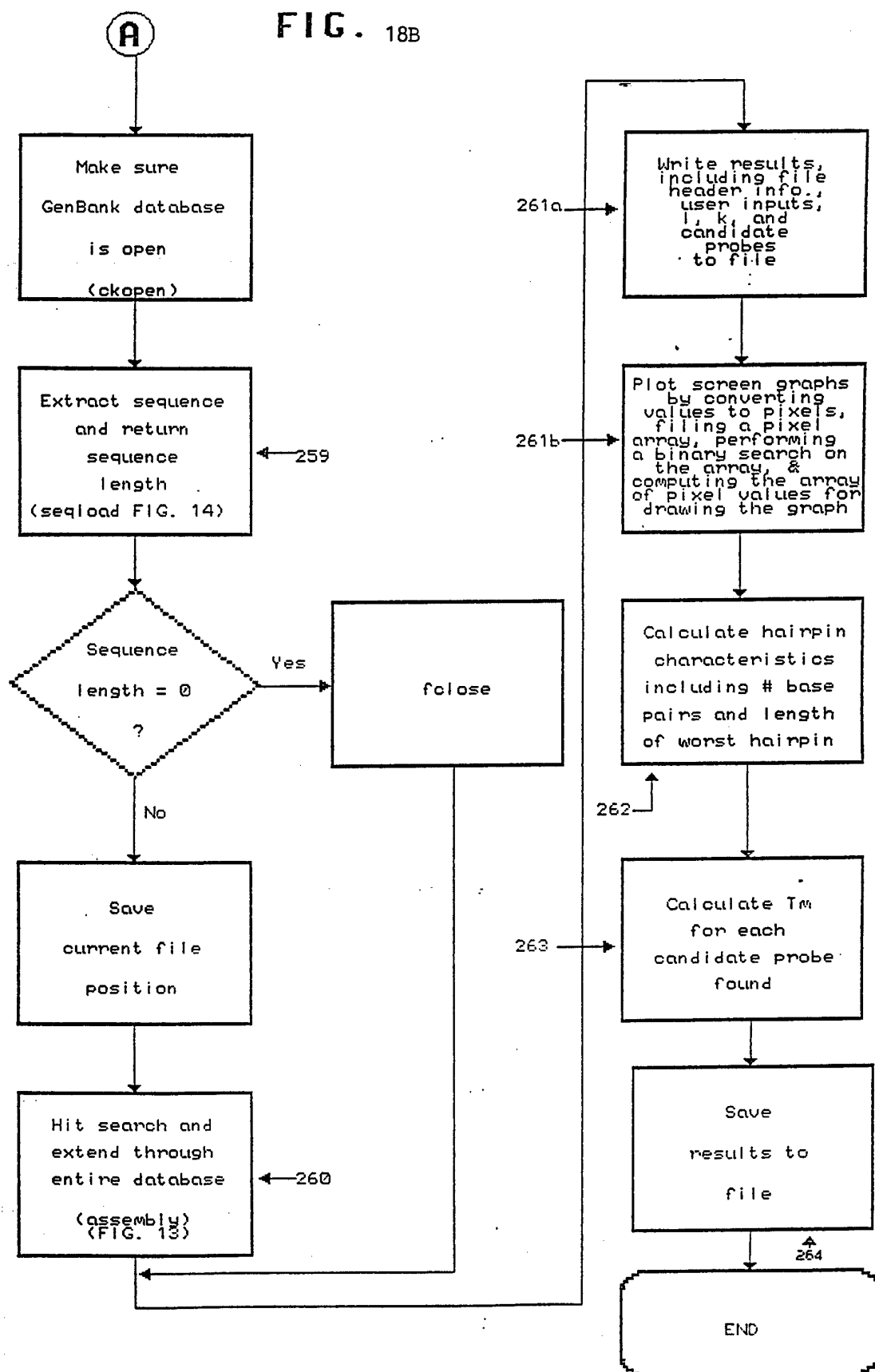
30/61

FIG. 18A



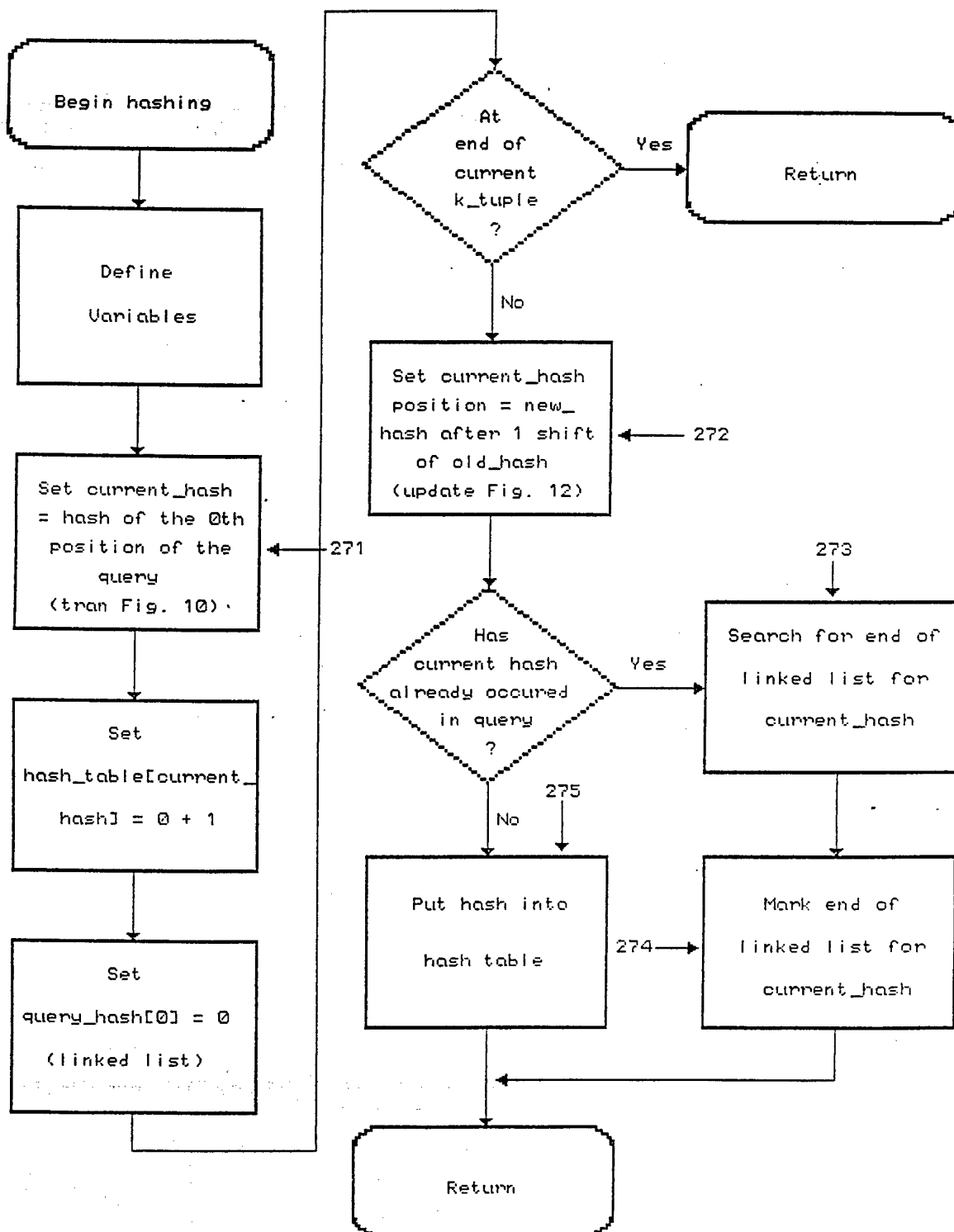
SUBSTITUTE SHEET

FIG. 18B



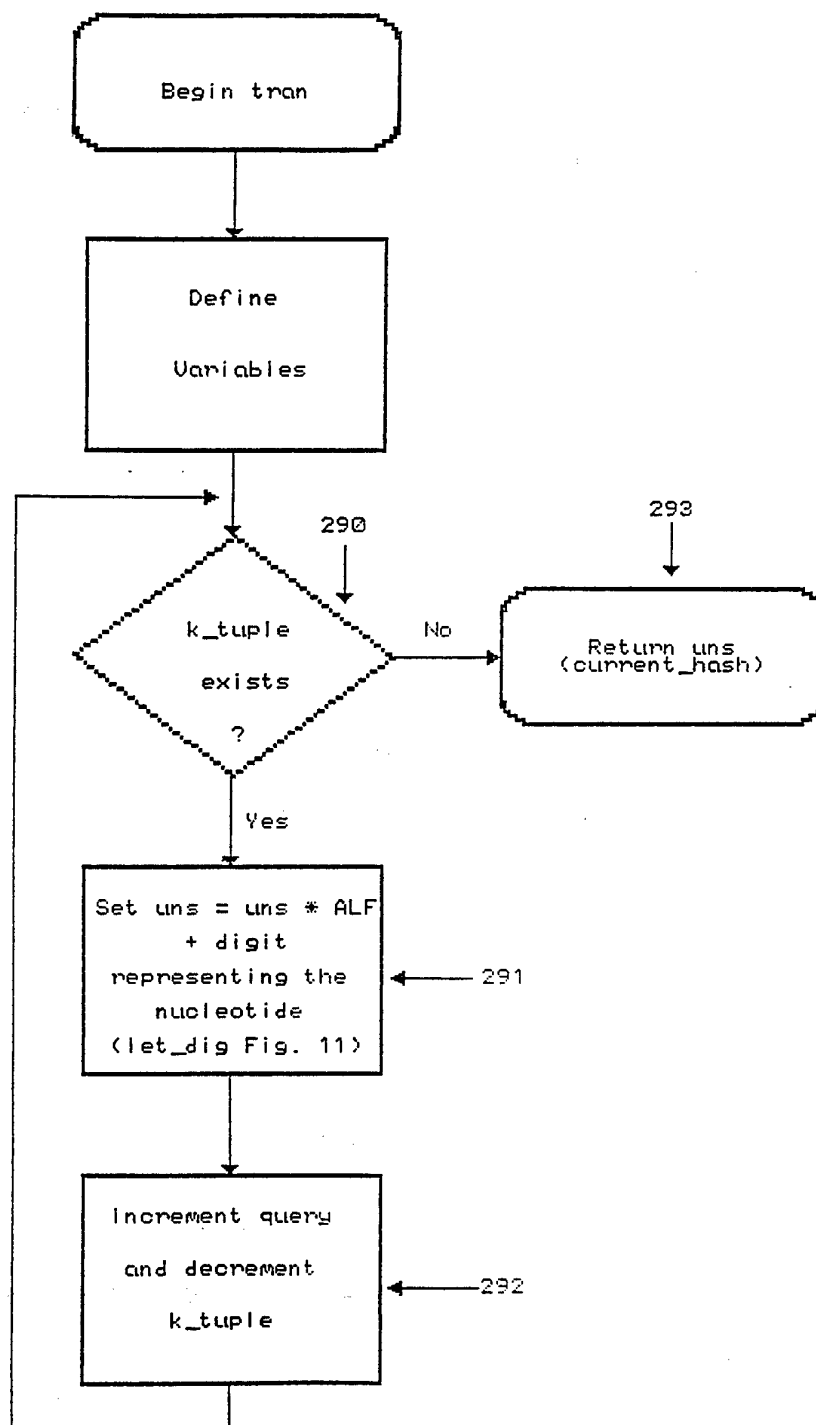
32/61

FIG. 19



33/61

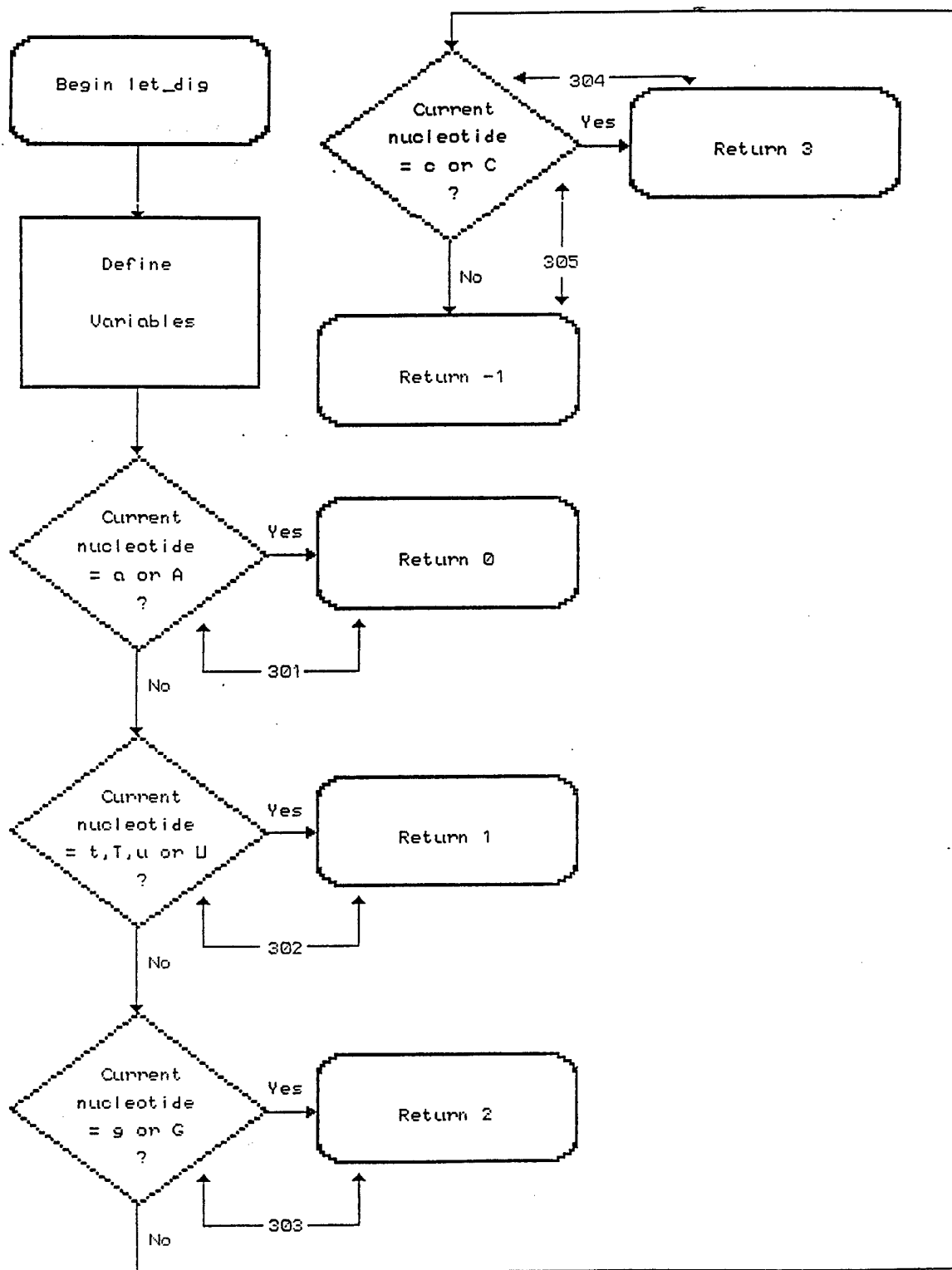
FIG. 20



SUBSTITUTE SHEET

34/61

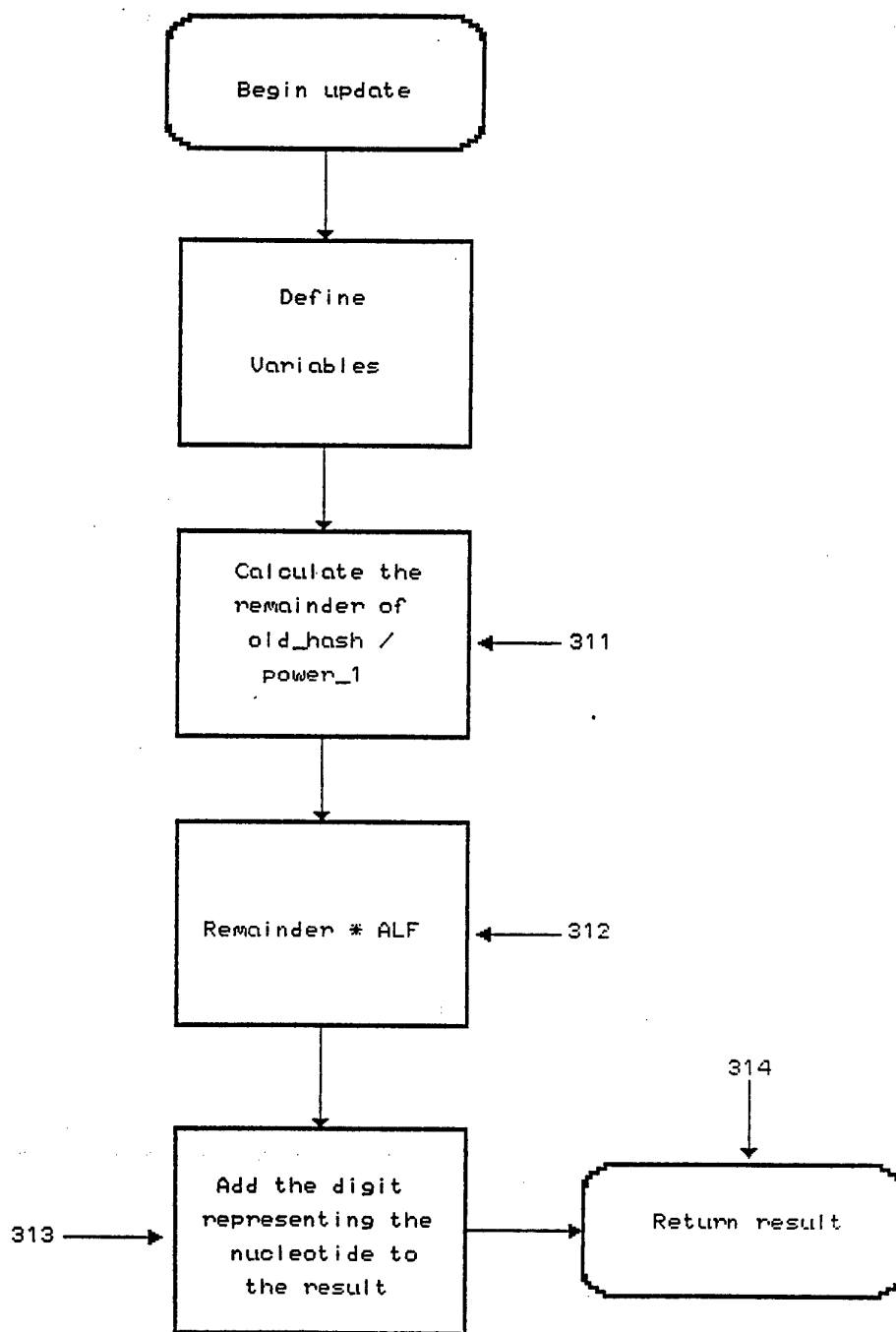
FIG. 21



SUBSTITUTE SHEET

35/61

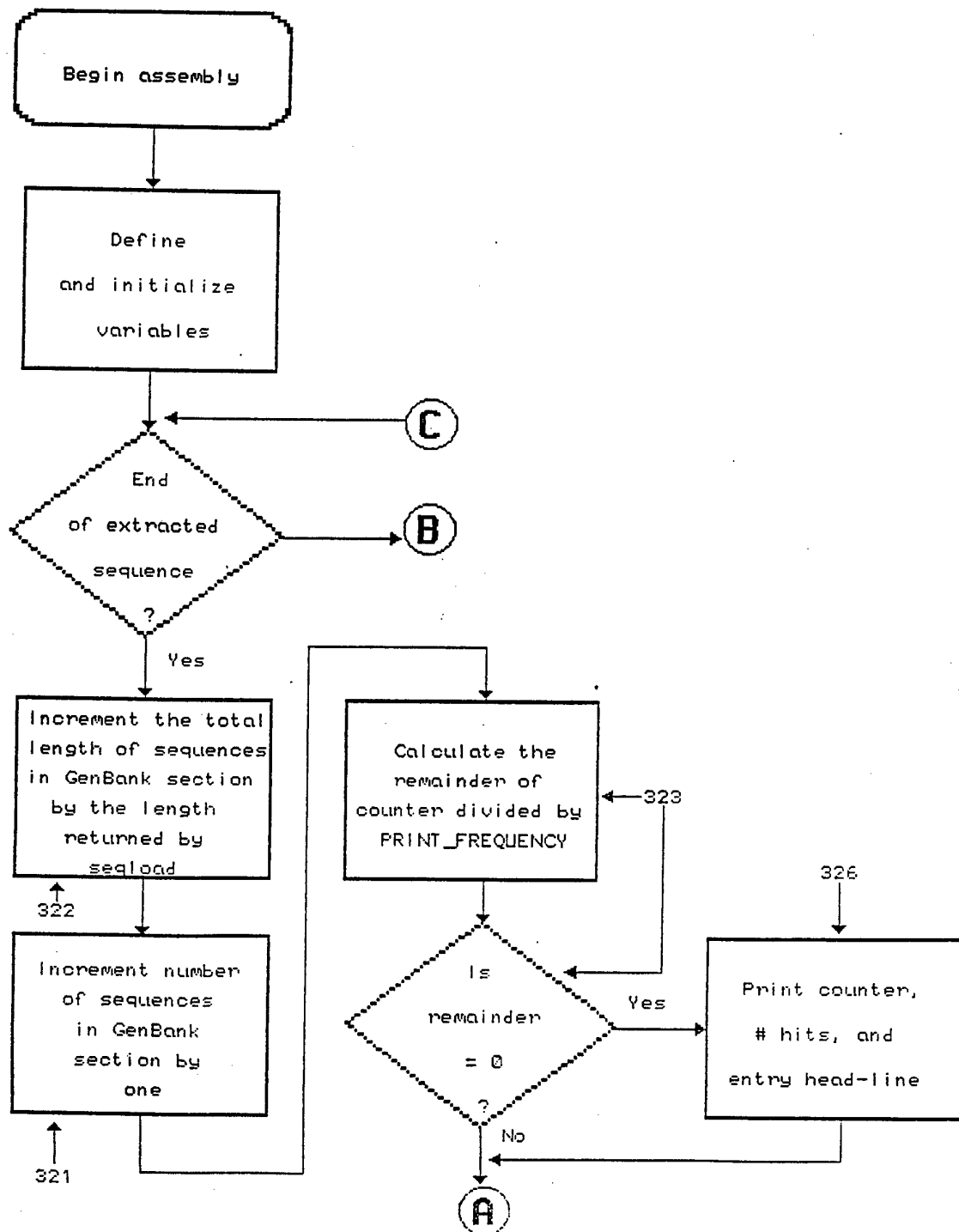
FIG. 22



SUBSTITUTE SHEET

36/61

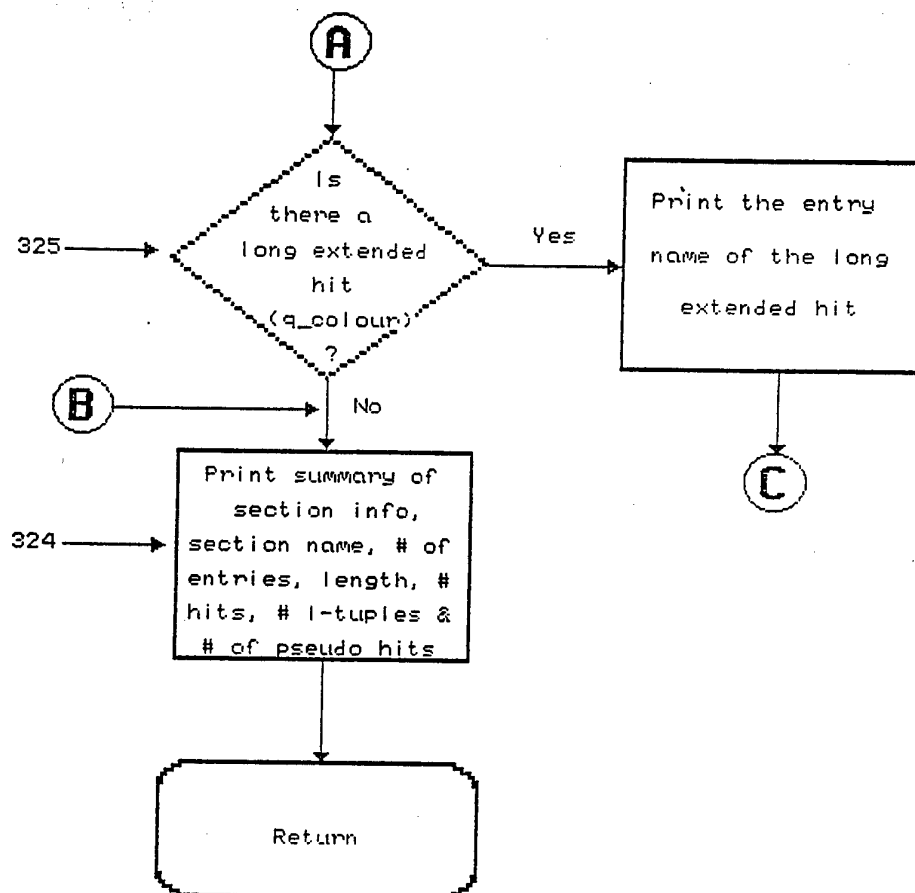
FIG. 23A



SUBSTITUTE SHEET

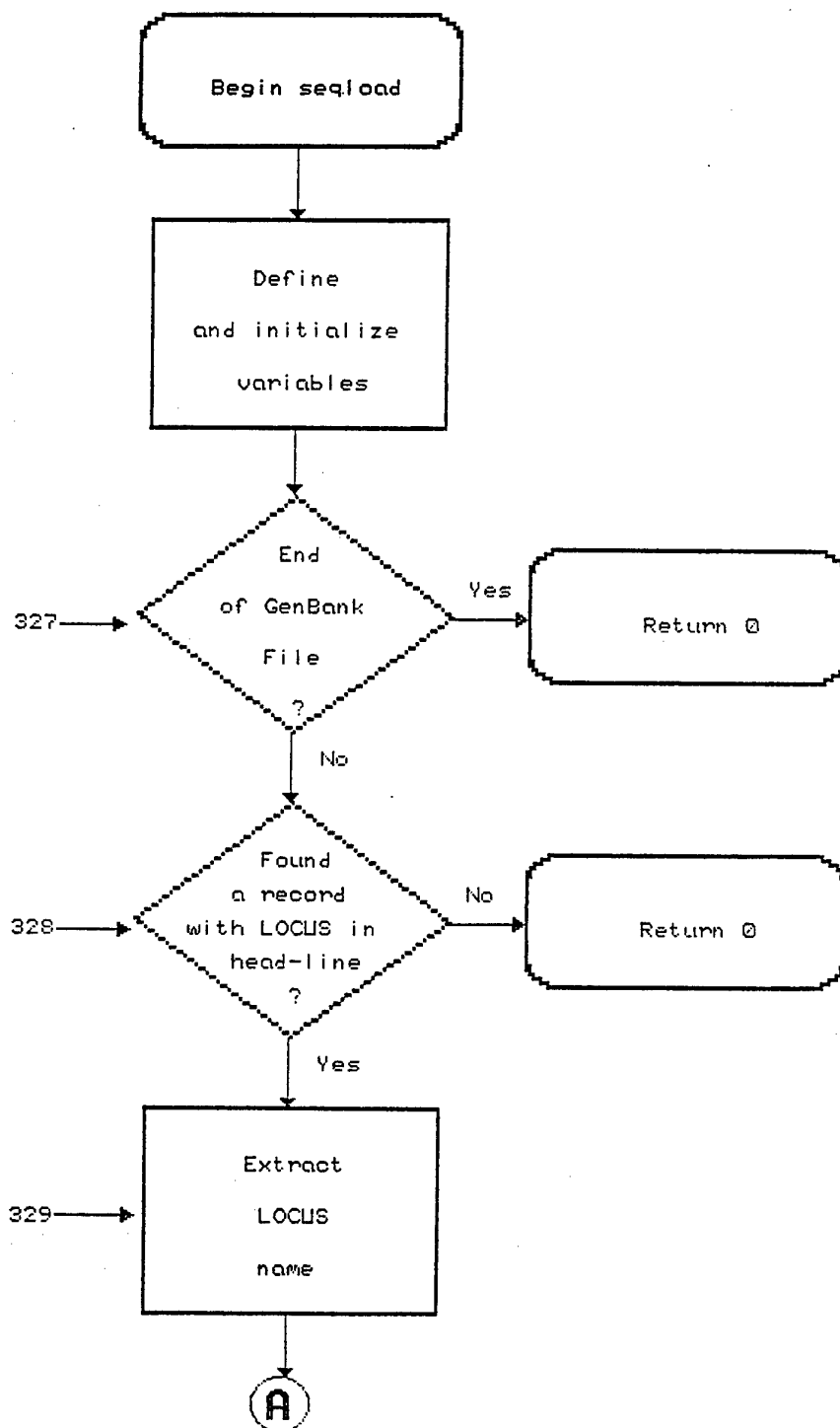
37/61

FIG. 23B



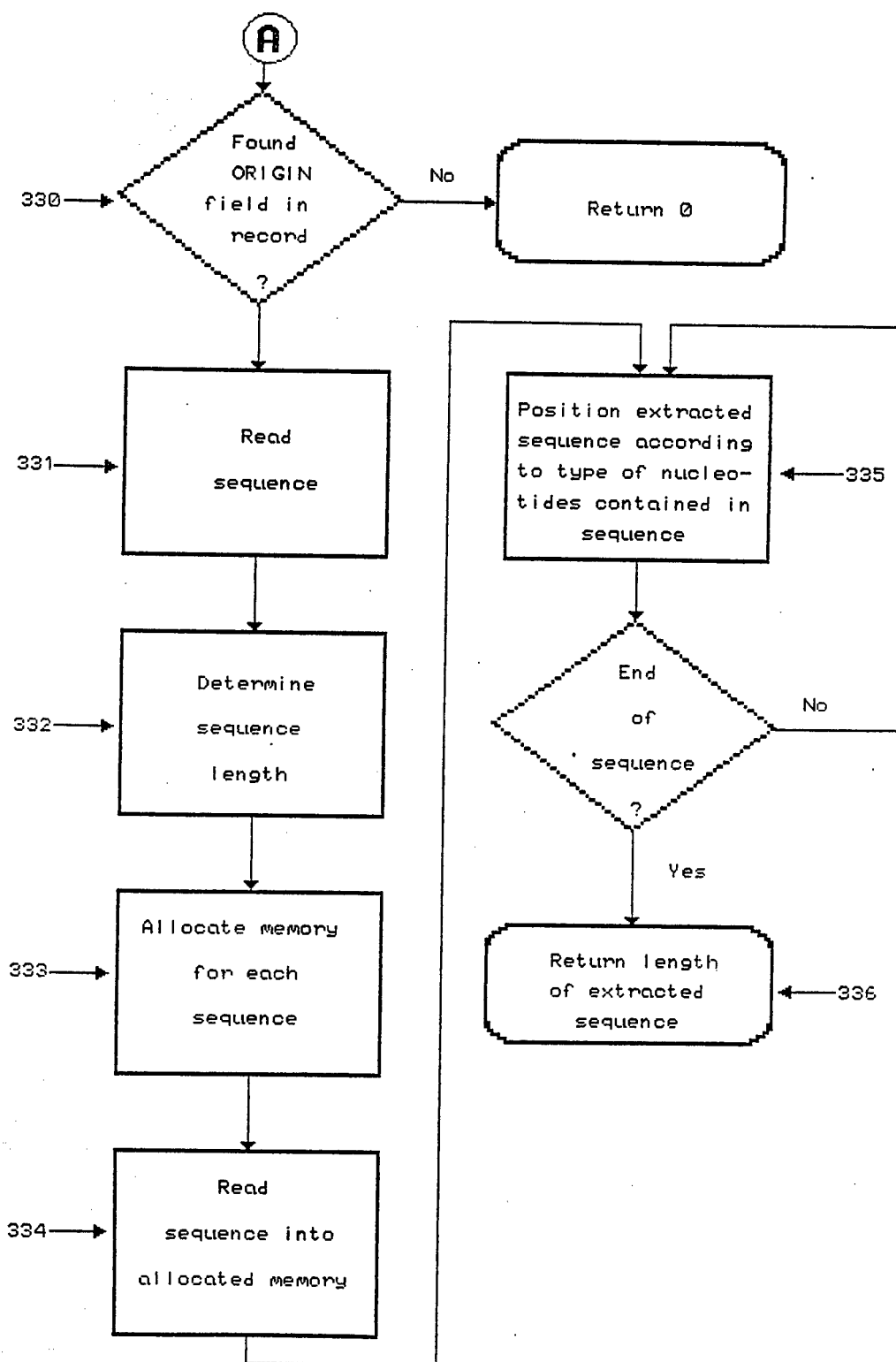
38/61

FIG. 24A



39/61

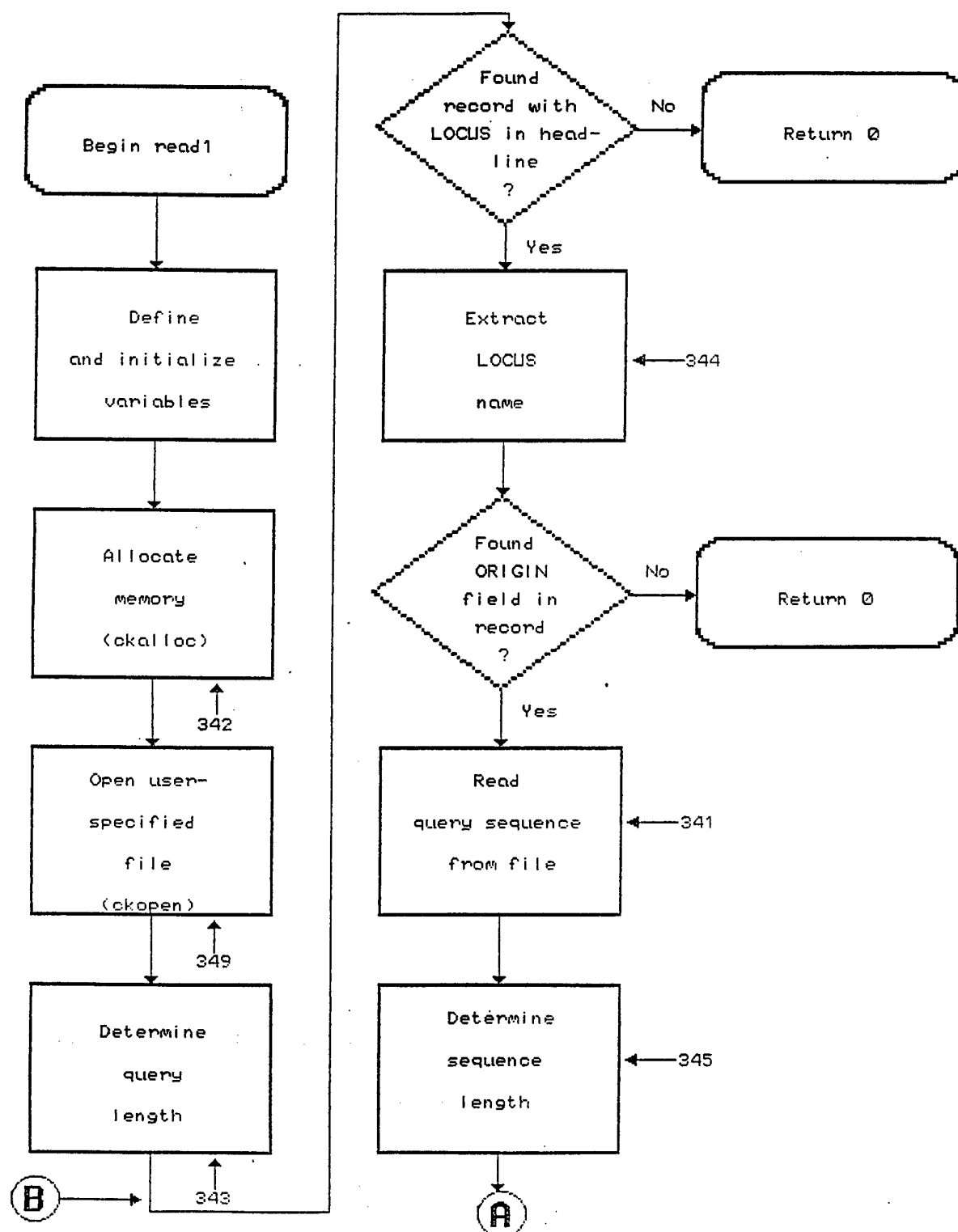
FIG. 24B



SUBSTITUTE SHEET

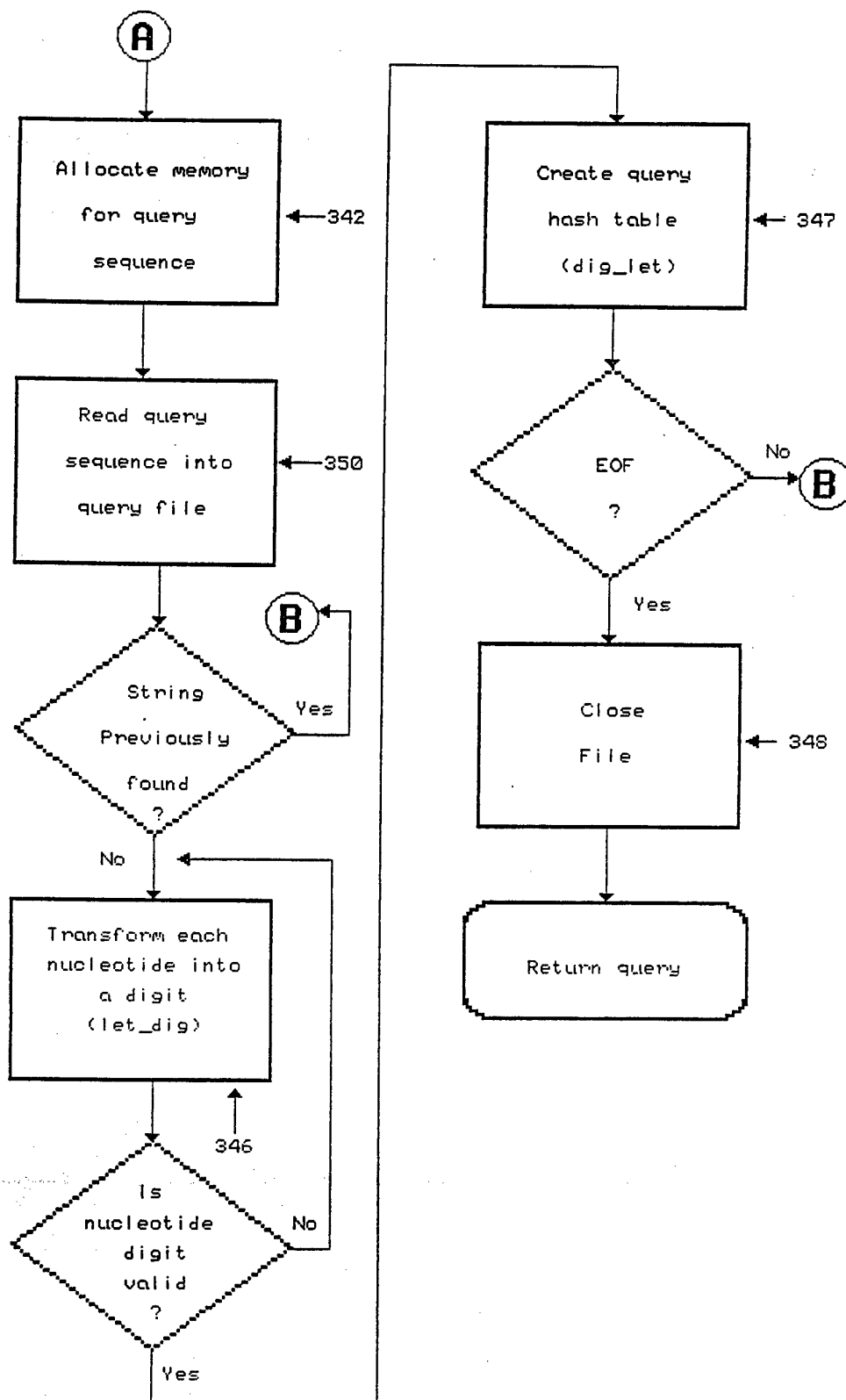
40/61

FIG. 25A



41/61

FIG. 25B



SUBSTITUTE SHEET

42/61

FIG. 26

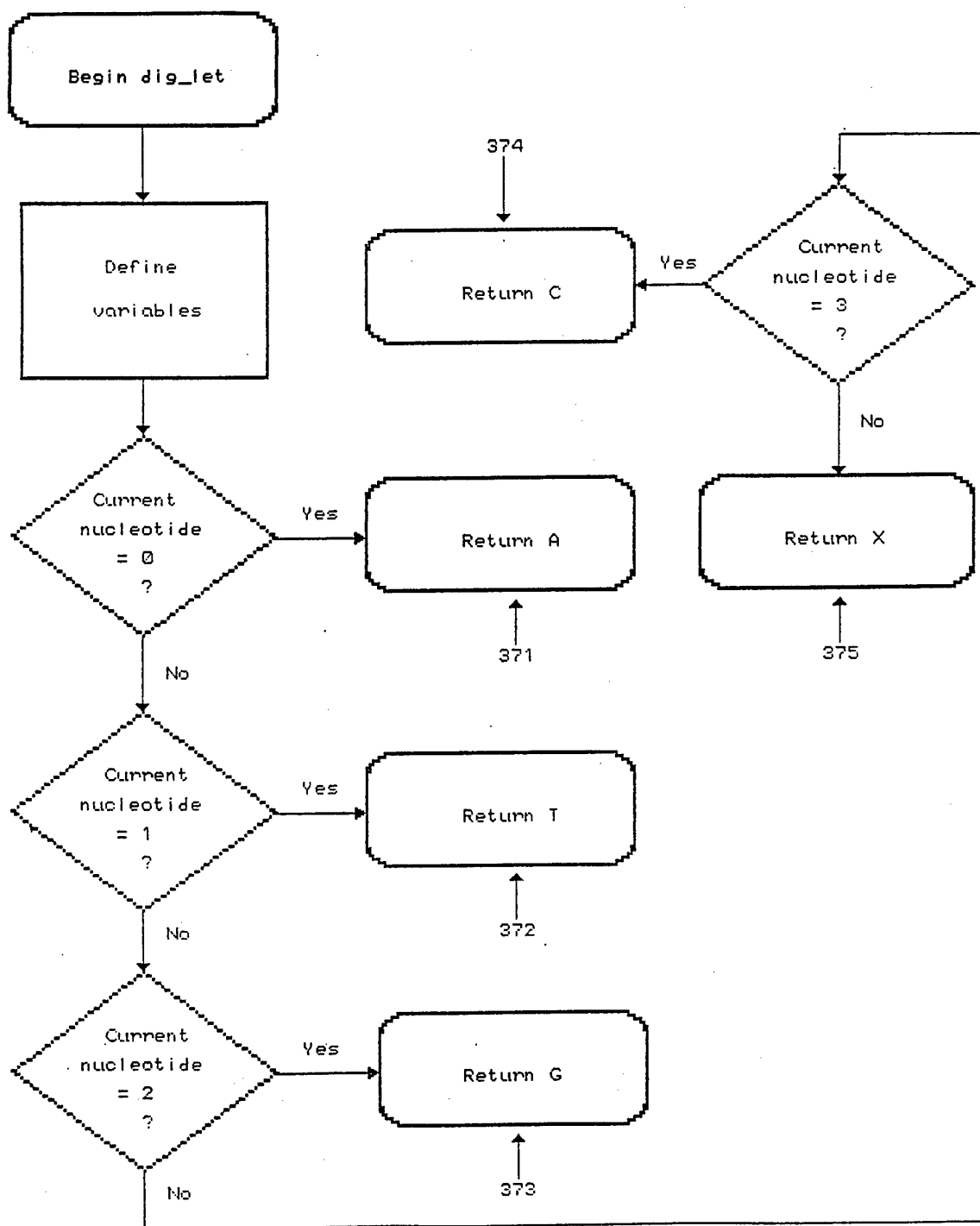
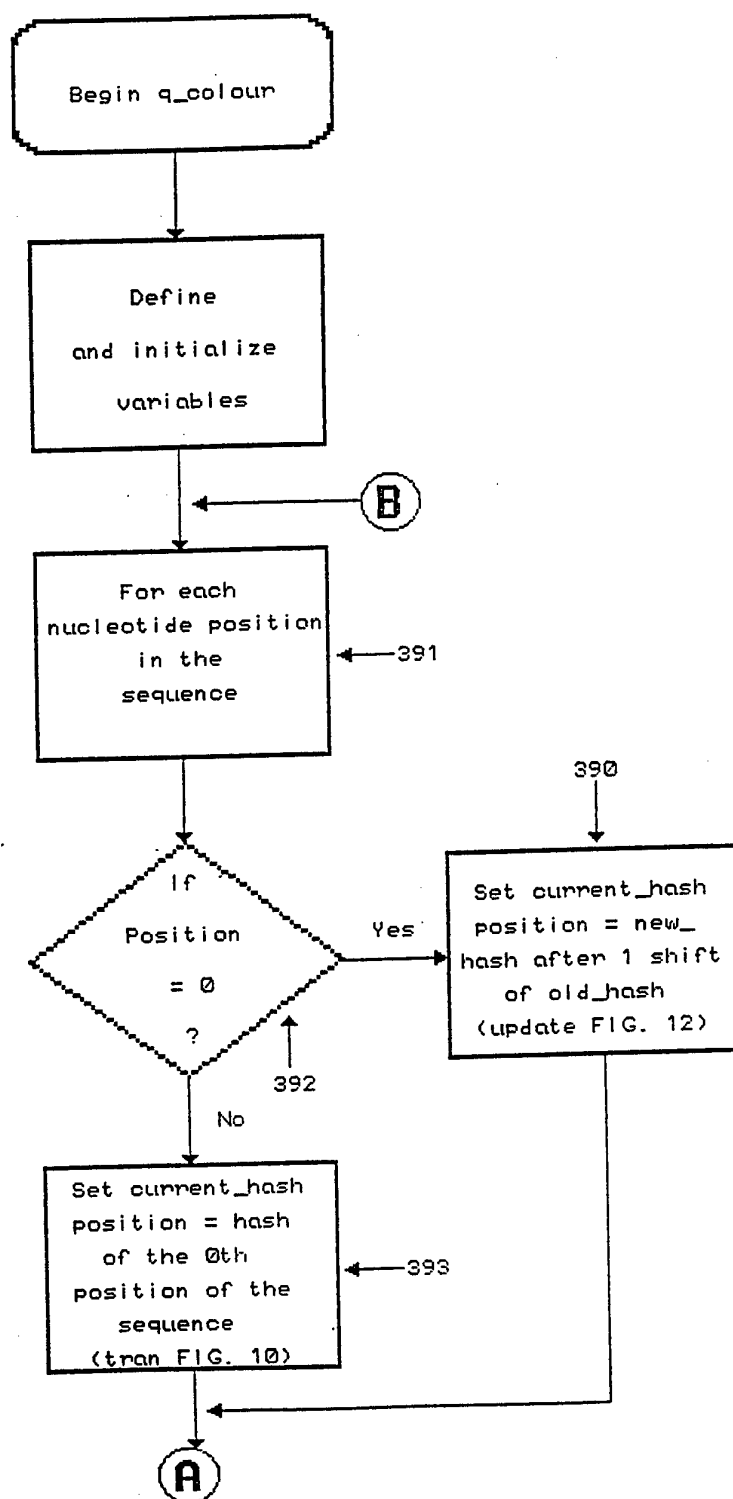
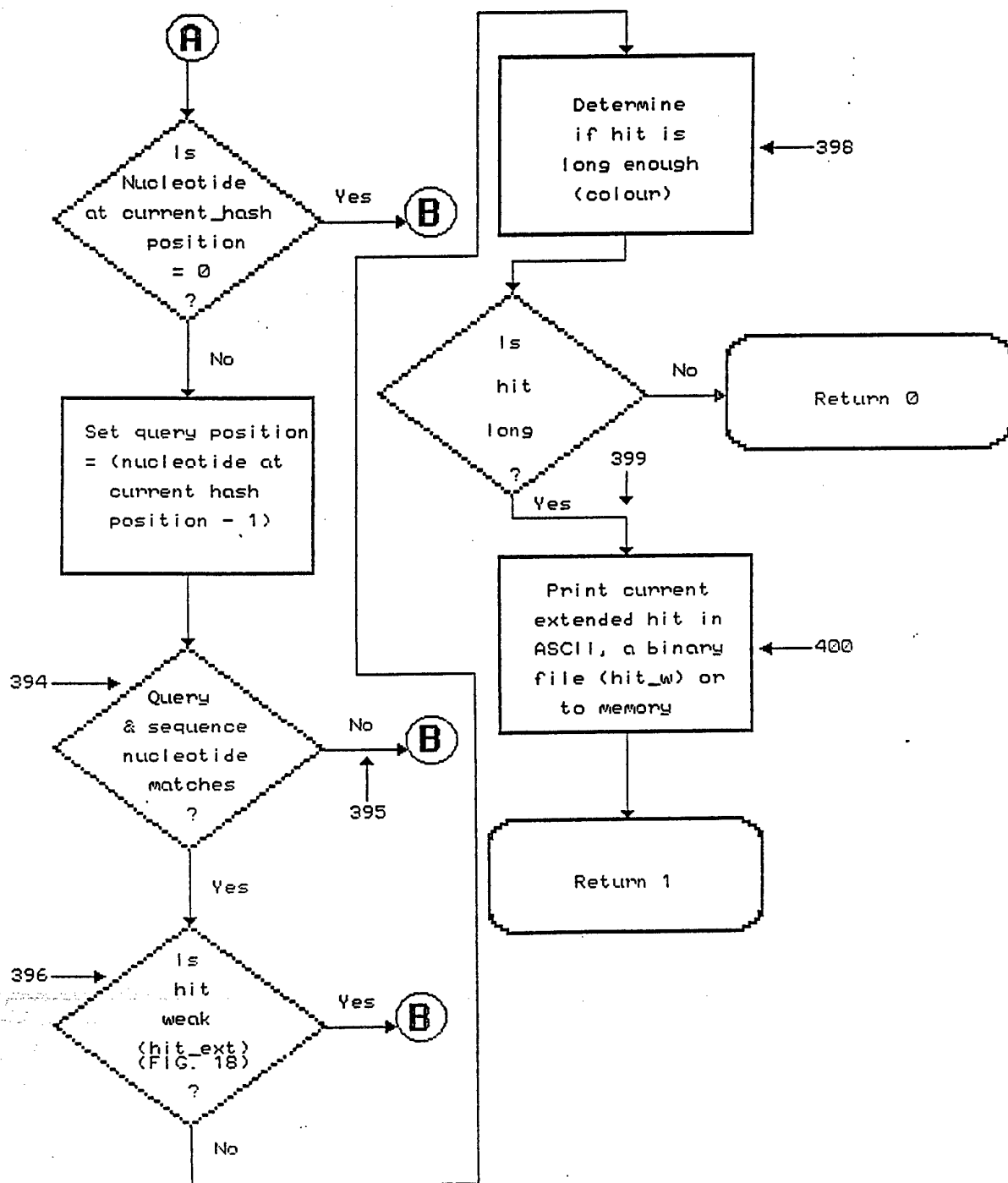


FIG. 27A



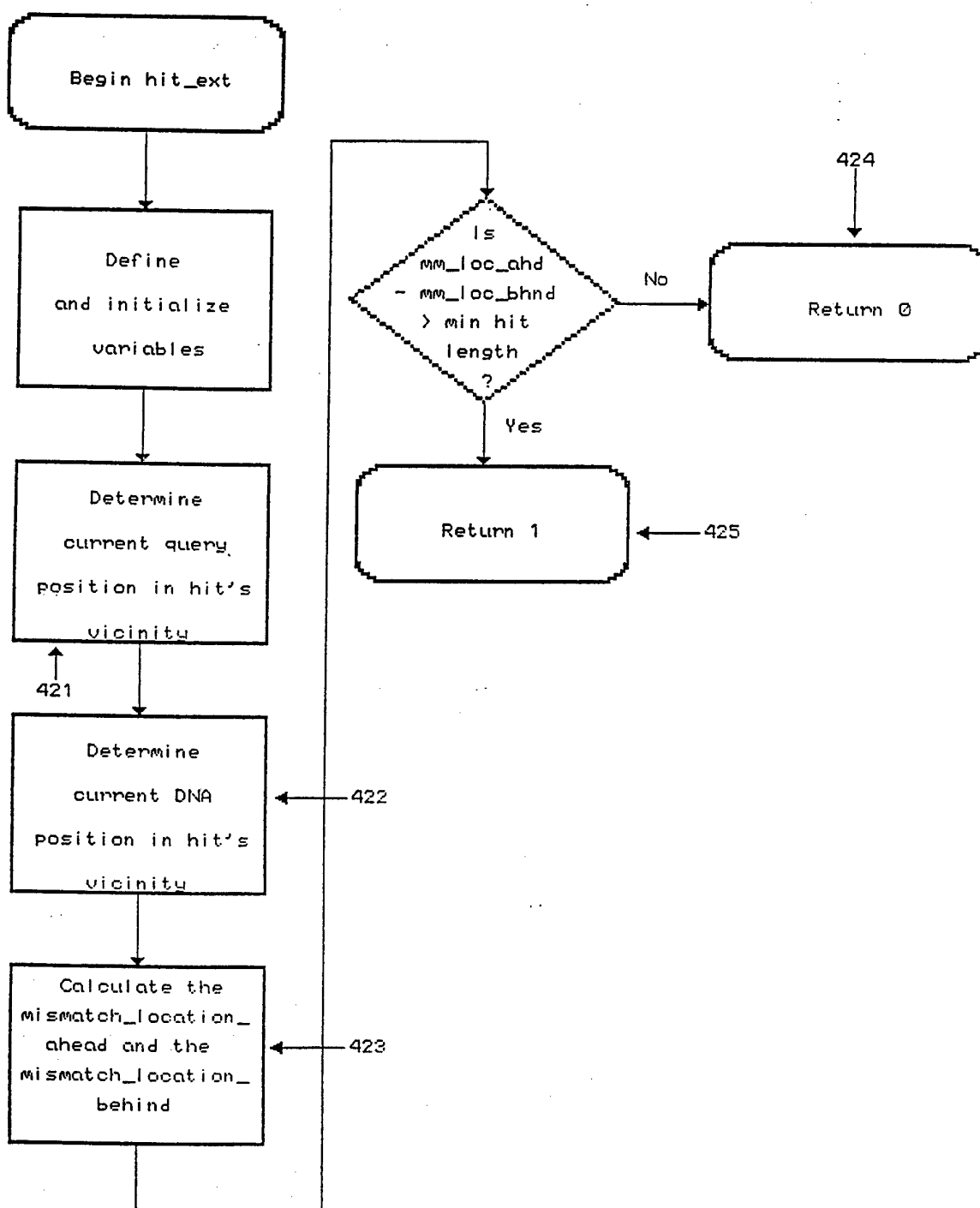
44/61

FIG. 27B



45/61

FIG. 28



46/61

FIG. 29

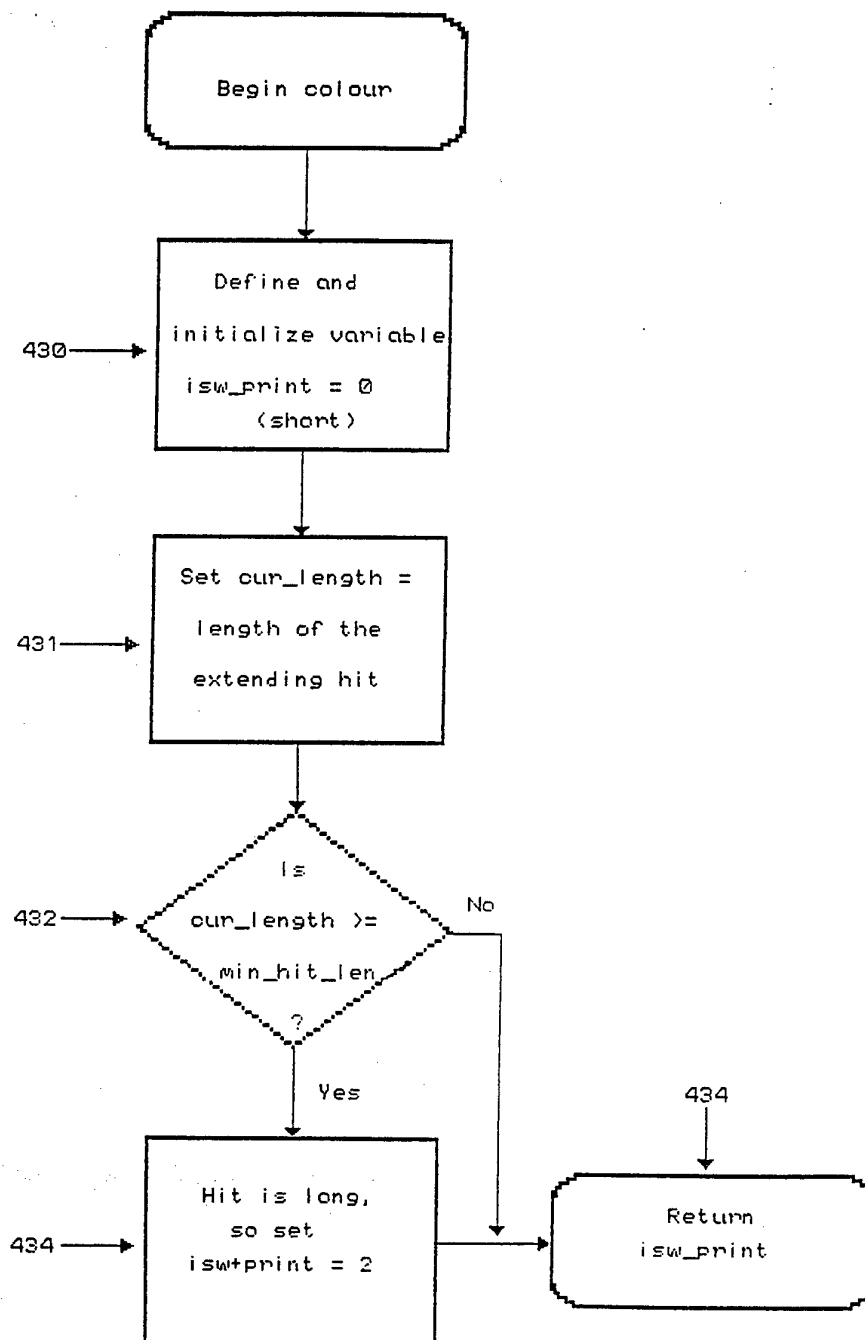


FIG 30

OligoProbe DesignStation

Probes: C:\HITACHI\HUMBJUNX.CDS

Datatbase: C:\HITACHI\JUNMIX.SEQ

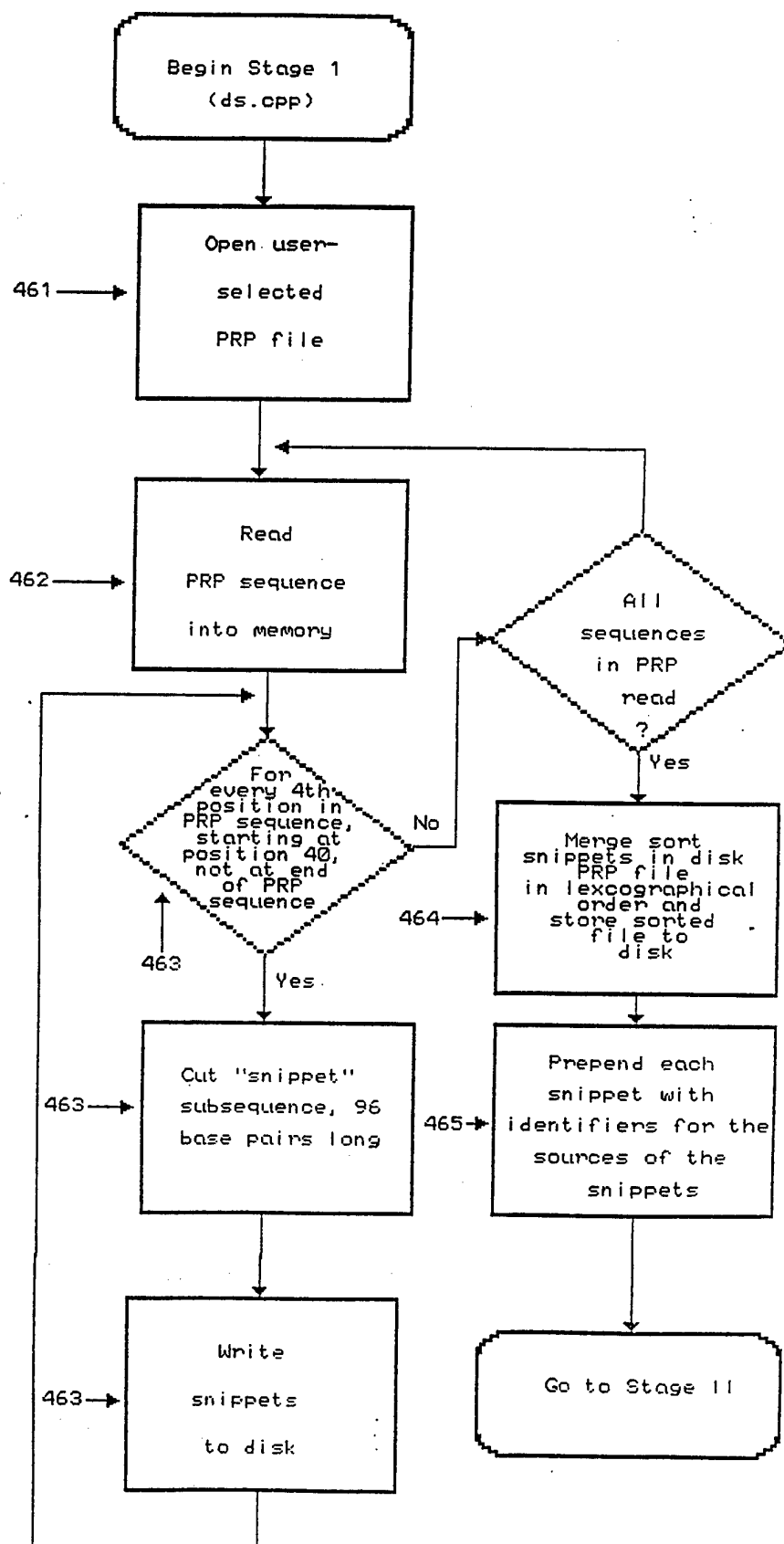
Mismatch Model, l = 21, k = 4

Position length	Mismatches								screensN Probe
	0	1	2	3	4	5	6	7	8
1 21	1	1	1	1	1	1	1	1	1
2 21	1	1	1	1	1	1	1	1	1
3 21	1	1	1	1	1	1	1	1	1
4 21	1	1	1	1	1	1	1	1	1
5 21	1	1	1	1	1	1	1	1	1
6 21	1	1	1	1	1	1	1	1	1
7 21	1	1	1	1	1	1	1	1	1
8 21	1	1	1	1	1	1	1	1	1
9 21	1	1	1	1	1	1	1	1	1
10 21	1	1	1	1	1	1	1	1	1
11 21	1	1	1	1	1	1	1	1	1
12 21	1	1	1	1	1	1	1	1	1
13 21	1	1	1	1	1	1	1	1	1
14 21	1	1	1	1	1	1	1	1	1
15 21	1	1	1	1	1	1	1	1	1
16 21	1	1	1	1	1	1	1	1	1
17 21	1	1	1	1	1	1	1	1	1
18 21	1	1	1	1	1	1	1	1	1
19 21	1	1	1	1	1	1	1	1	1
20 21	1	1	1	1	1	1	1	1	1
21 21	1	1	1	1	1	1	1	1	1
22 21	1	1	1	1	1	1	1	1	1
23 21	1	1	1	1	1	1	1	1	1
24 21	1	1	1	1	1	1	1	1	1
25 21	1	1	1	1	1	1	1	1	1
26 21	1	1	1	1	1	1	1	1	1
27 21	1	1	1	1	1	1	1	1	1
28 21	1	1	1	1	1	1	1	1	1
29 21	1	1	1	1	1	1	1	1	1
30 21	1	1	1	1	1	1	1	1	1
31 21	1	1	1	1	1	1	1	1	1
32 21	1	1	1	1	1	1	1	1	1
33 21	1	1	1	1	1	1	1	1	1
34 21	1	1	1	1	1	1	1	1	1
35 21	1	1	1	1	1	1	1	1	1
36 21	1	1	1	1	1	1	1	1	1
37 21	1	1	1	1	1	1	1	1	1
38 21	1	1	1	1	1	1	1	1	1
39 21	1	1	1	1	1	1	1	1	1
40 21	1	1	1	1	1	1	1	1	1
41 21	1	1	1	1	1	1	1	1	1
42 21	1	1	1	1	1	1	1	1	1
43 21	1	1	1	1	1	1	1	1	1
44 21	1	1	1	1	1	1	1	1	1
45 21	1	1	1	1	1	1	1	1	1
46 21	1	1	1	1	1	1	1	1	1
47 21	1	1	1	1	1	1	1	1	1
48 21	1	1	1	1	1	1	1	1	1
49 21	1	1	1	1	1	1	1	1	1
50 21	1	1	1	1	1	1	1	1	1
51 21	1	1	1	1	1	1	1	1	1
52 21	1	1	1	1	1	1	1	1	1
53 21	1	1	1	1	1	1	1	1	1

SUBSTITUTE SHEET

48/61

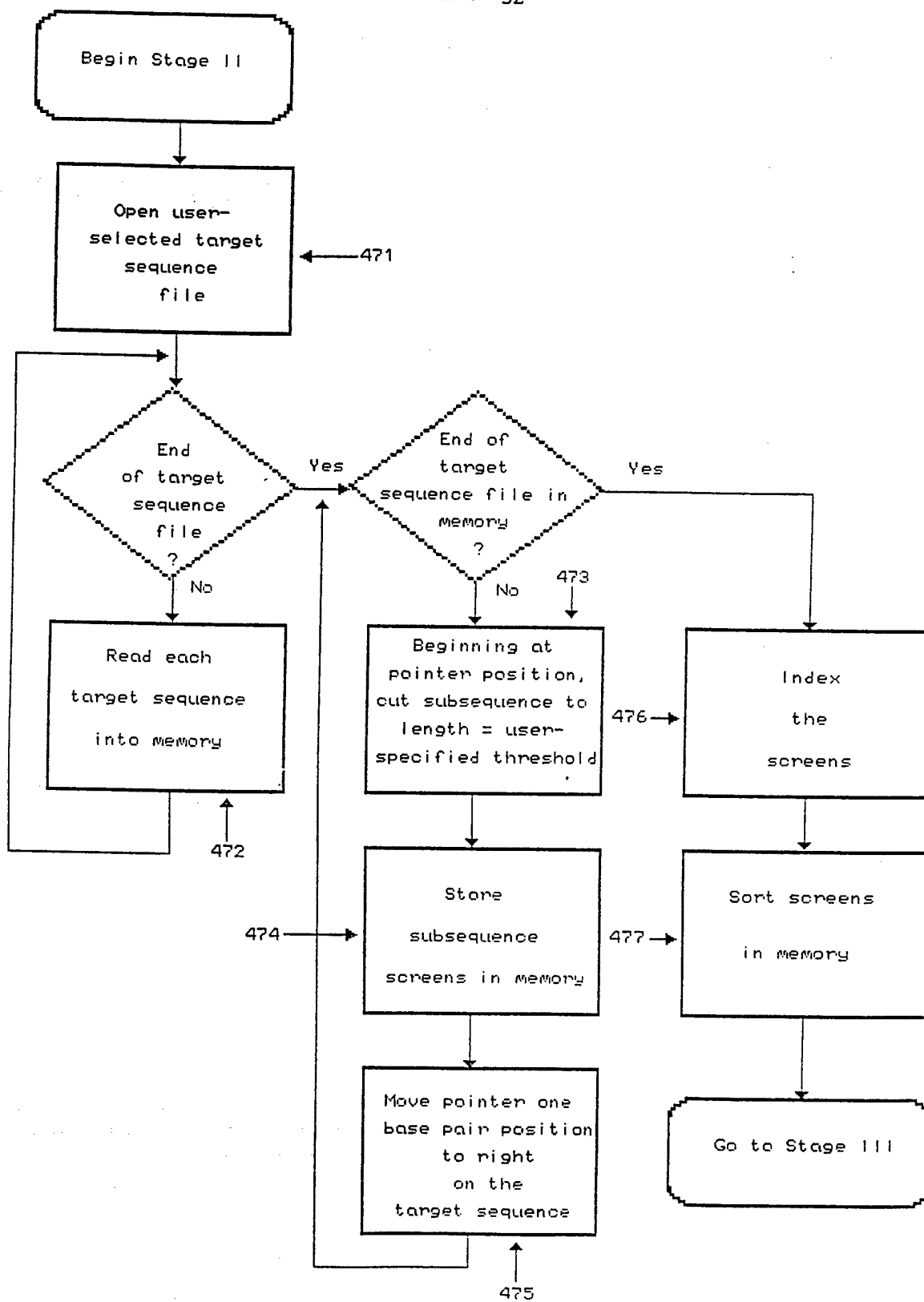
FIG. 31



SUBSTITUTE SHEET

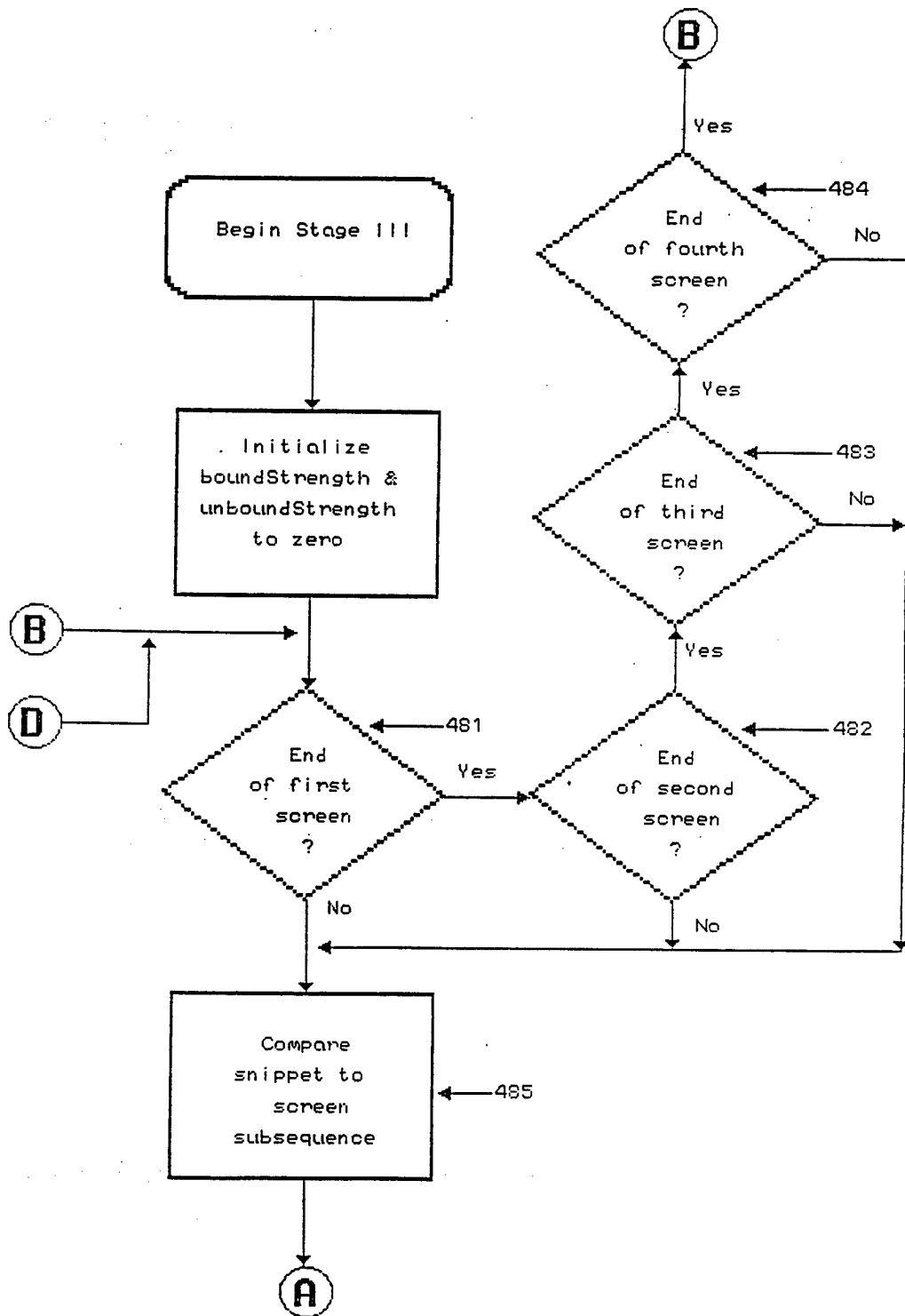
49/61

FIG. 32



50/61

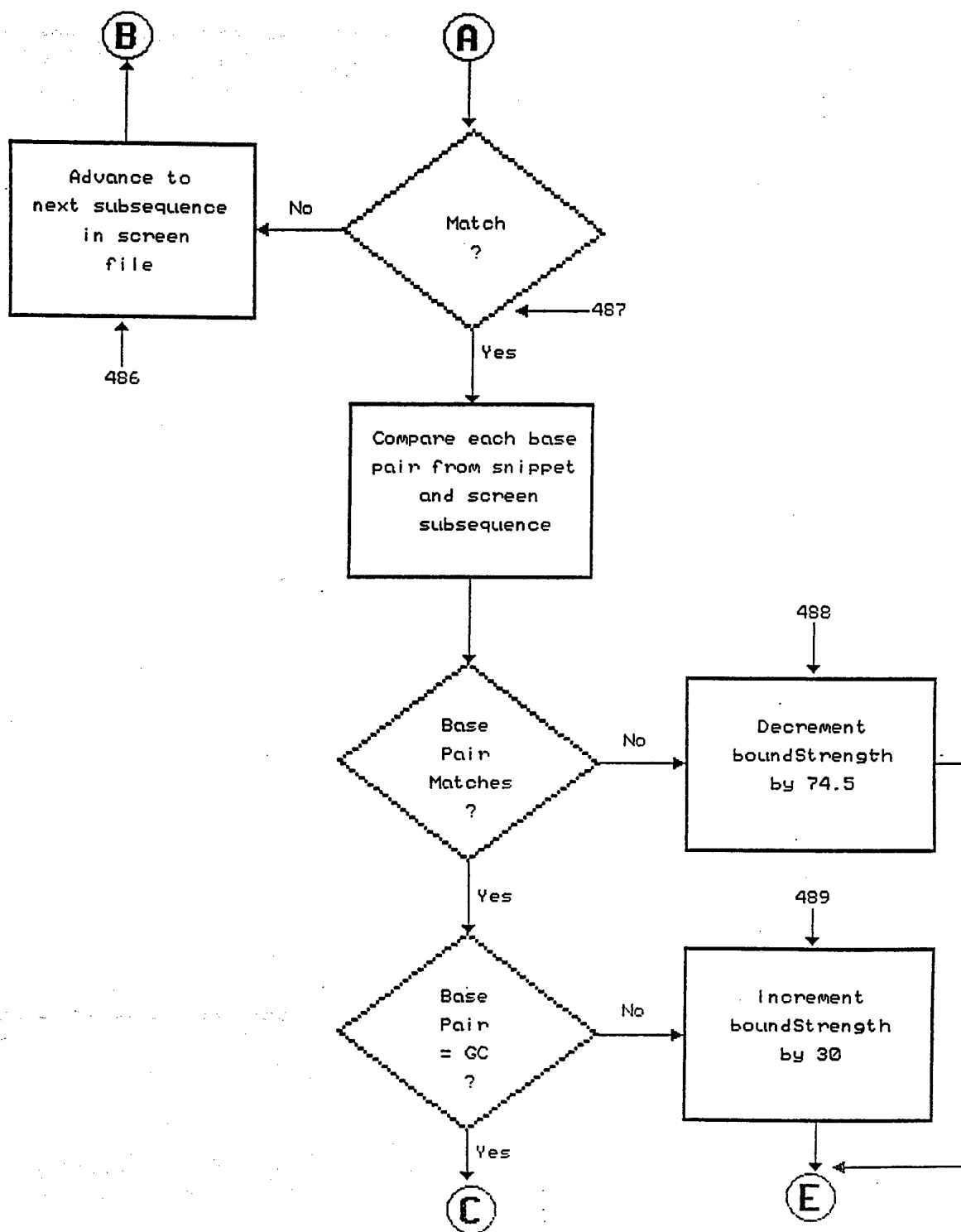
FIG. 33A



SUBSTITUTE SHEET

51/61

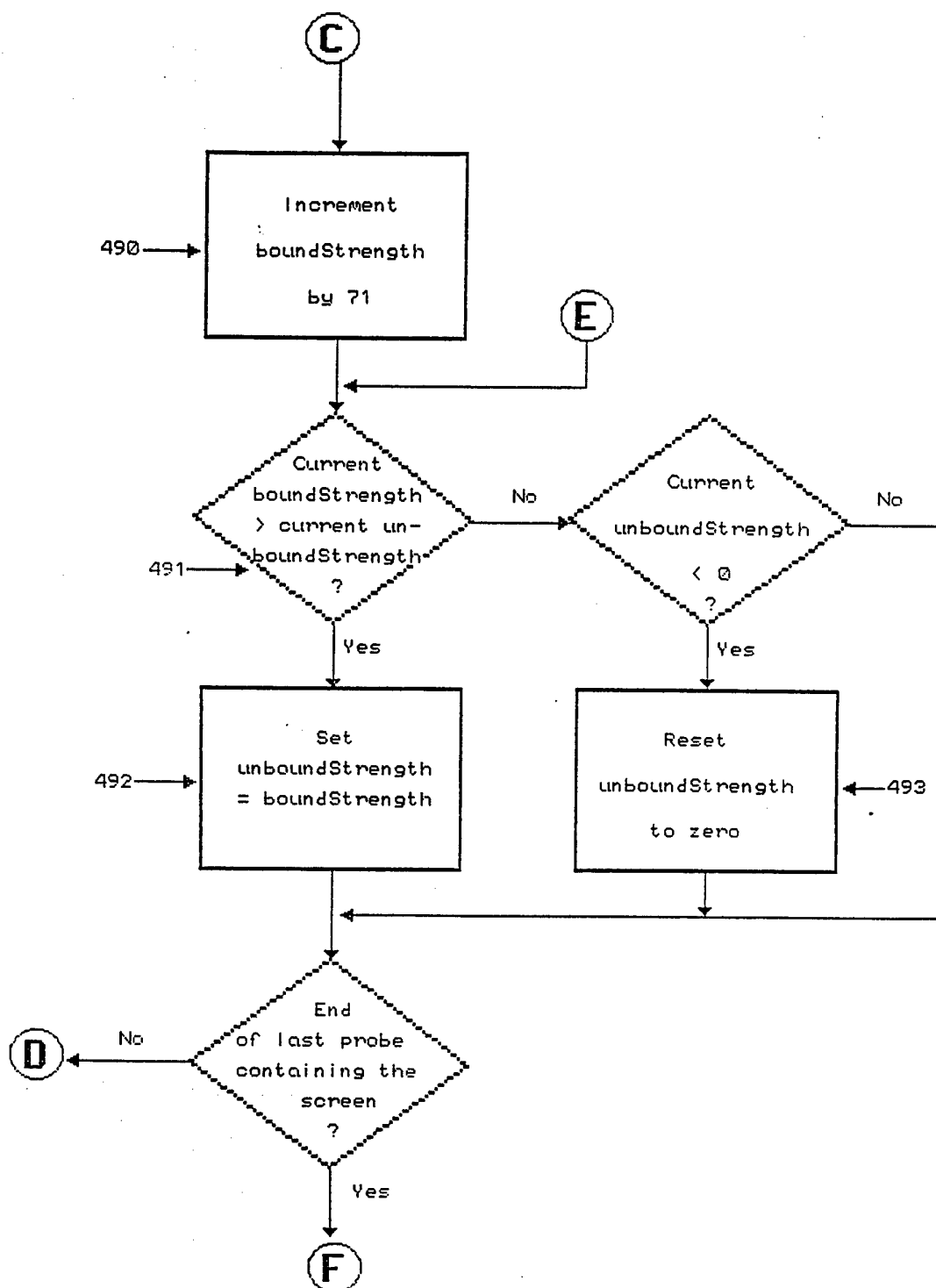
FIG. 33B



SUBSTITUTE SHEET

52/61

FIG. 33C



SUBSTITUTE SHEET

53/61

FIG. 33D

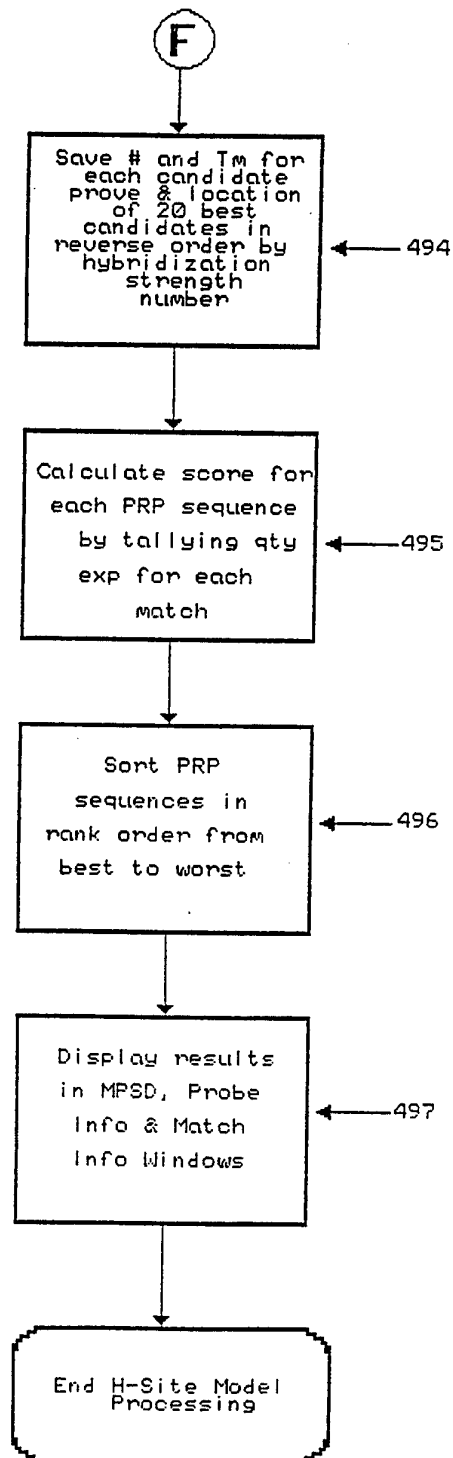


FIG. 34A

OligoProbe DesignStation

Probes: C:\HITACHI\HUMBJUNX.CDS
 Datatbase: C:\HITACHI\JUNMIX.SEQ

Mismatch Model, l = 21, k = 4

Position length	Mismatches								screensN Probe
	0	1	2	3	4	5	6	7	8
1 21	0	0	0	0	0	0	0	0	0
2 21	0	0	0	0	0	0	0	0	0
3 21	0	0	0	0	0	0	0	0	0
4 21	0	0	0	0	0	0	0	0	0
5 21	0	0	0	0	0	0	0	0	0
6 21	0	0	0	0	0	0	0	0	0
7 21	0	0	0	0	0	0	0	0	0
8 21	0	0	0	0	0	0	0	0	0
9 21	0	0	0	0	0	0	0	0	0
10 21	0	0	0	0	0	0	0	0	0
11 21	0	0	0	0	0	0	0	0	0
12 21	0	0	0	0	0	0	0	0	0
13 21	0	0	0	0	0	0	0	0	0
14 21	0	0	0	0	0	0	0	0	0
15 21	0	0	0	0	0	0	0	0	0
16 21	0	0	0	0	0	0	0	0	0
17 21	0	0	0	0	0	0	0	0	0
18 21	0	0	0	0	0	0	0	0	0
19 21	0	0	0	0	0	0	0	0	0
20 21	0	0	0	0	0	0	0	0	0
21 21	0	0	0	0	0	0	0	0	0
22 21	0	0	0	0	0	0	0	0	0
23 21	0	0	0	0	0	0	0	0	0
24 21	0	0	0	0	0	0	0	0	0
25 21	0	0	0	0	0	0	0	0	0
26 21	0	0	0	0	0	0	0	0	0
27 21	0	0	0	0	0	0	0	0	0
28 21	0	0	0	0	0	0	0	0	0
29 21	0	0	0	0	0	0	0	0	0
30 21	0	0	0	0	0	0	0	0	0
31 21	0	0	0	0	0	0	0	0	0
32 21	0	0	0	0	0	0	0	0	0
33 21	0	0	0	0	0	0	0	0	0
34 21	0	0	0	0	0	0	0	0	0
35 21	0	0	0	0	0	0	0	0	0
36 21	0	0	0	0	0	0	0	0	0
37 21	0	0	0	0	0	0	0	0	0
38 21	0	0	0	0	0	0	0	0	0
39 21	0	0	0	0	0	0	0	0	0
40 21	0	0	0	0	0	0	0	0	0
41 21	0	0	0	0	0	0	0	0	0
42 21	0	0	0	0	0	0	0	0	0
43 21	0	0	0	0	0	0	0	0	0
44 21	0	0	0	0	0	0	0	0	0
45 21	0	0	0	0	0	0	0	0	0
46 21	0	0	0	0	0	0	0	0	0
47 21	0	0	0	0	0	0	0	0	0
48 21	0	0	0	0	0	0	0	0	0
49 21	0	0	0	0	0	0	0	0	0
50 21	0	0	0	0	0	0	0	0	0
51 21	0	0	0	0	0	0	0	0	0
52 21	0	0	0	0	0	0	0	0	0
53 21	0	0	0	0	0	0	0	0	0

SUBSTITUTE SHEET

55/61

FIG. 34B

(Partial File -- 10 pages of 190 pages)

OligoProbe DesignStation

Probes: C:\HITACHI\HUMBJUNX.CDS

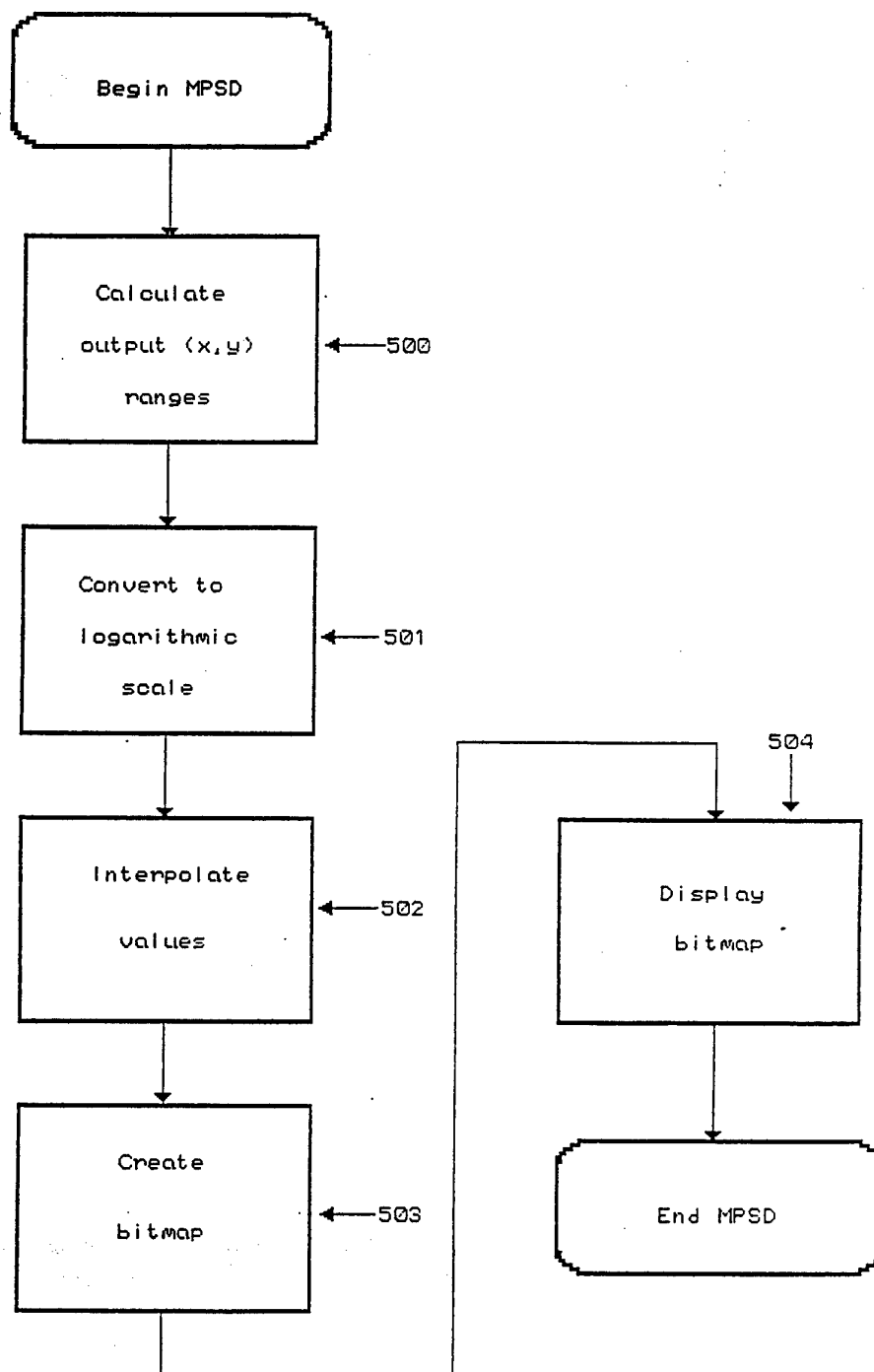
Preparation: C:\HITACHI\JUNMIX.PRP

Locus pos	Tm	Locus pos	Tm	Locus pos	Tm
atgtgcactaaaaatggaacagcccttctac					
1 30	1 1 1 2	2 2 2 2 3 4			
humbjunx 1	60.76				
musbjunx 1	50.03				
muscjunx 1	30.07				
musdjunx 721	27.84				
tgtgcactaaaaatggaacagcccttctac					
2 29	1 1 1 2	2 2 2 2 3 4			
humbjunx 65533	60.68				
musbjunx 65533	49.58				
muscjunx 1	29.97				
musdjunx 721	27.66				
gtgcactaaaaatggaacagcccttctac					
3 28	1 1 1 2	2 2 2 2 3 4			
humbjunx 65533	60.60				
musbjunx 65533	49.10				
muscjunx 1	29.86				
musdjunx 721	27.47				
tgccactaaaaatggaacagcccttctacc					
4 28	1 1 1 1	2 2 2 2 3 4			
humbjunx 65533	60.60				
musbjunx 65533	46.57				
muscjunx 1	29.86				
musdjunx 729	27.47				
gcactaaaaatggaacagcccttctacc					
5 27	1 1 1 1	2 2 2 2 3 4			
humbjunx 5	60.51				
musbjunx 5	45.96				
muscjunx 1	29.75				
musdjunx 729	27.26				

SUBSTITUTE SHEET

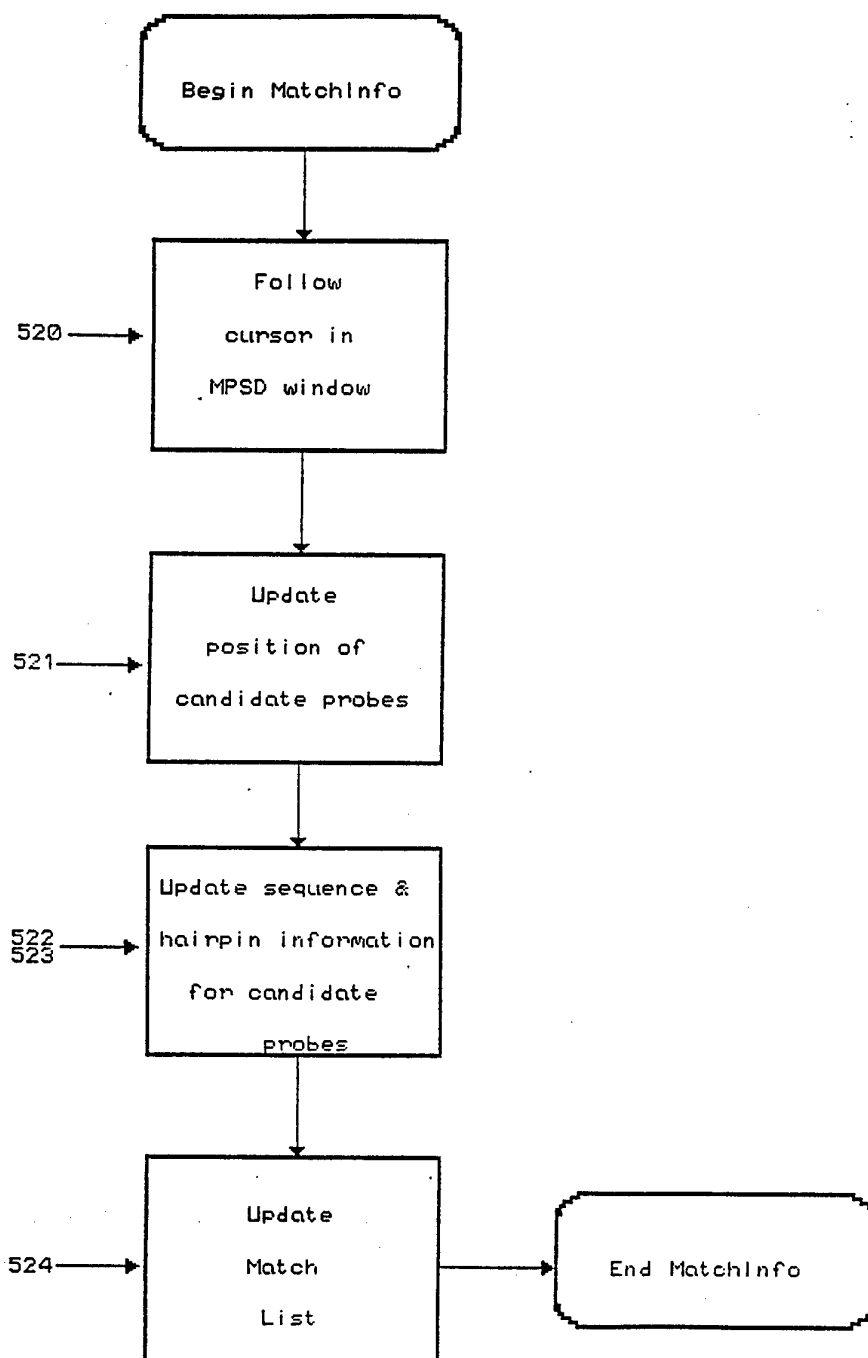
56/61

FIG. 35



57/61

FIG. 36



SUBSTITUTE SHEET

58/61

FIG. 37

LOCUS HUMBUNX 1044 bp DNA 19-DEC-1991
BASE COUNT 195 A 368 C 340 G 141 T
ORIGIN

```
1 ATGTGCACTA AAATGGAACA GCCCTTCTAC CACGACGACT CATACACAGC TACGGGATAC
61 GGCCGGGCCC CTGGTGGCCT CTCTCTACAC GACTACAAAC TCCTGAAACC GAGCCTGGCG
121 GTCAACCTGG CCGACCCCTA CCGGAGTCTC AAAGCGCCTG GGGCTCGCGG ACCCGGCCCA
181 GAGGGCGGCG GTGGCGGCAG CTACTTTTCT GGTCAGGGCT CGGACACCGG CGCGTCTCTC
241 AAGCTCGCCT CTTCGGAGCT GGAACGCCTG ATTGTCCCCA ACAGCAACGG CGTGATCACG
301 ACGACGCCTA CACCCCCGGG ACAGTACTTT TACCCCCGCG GGGGTGGCAG CGGTGGAGGT
361 GCAGGGGGCG CAGGGGGCGG CGTCACCGAG GAGCAGGAGG GCTTCGCCGA CGGCTTTGTC
421 AAAGCCCTGG ACGATCTGCA CAAGATGAAC CACGTGACAC CCCCCAACGT GTCCCTGGGC
481 GCTACCGGGG GGCCCCCGGC TGGGCCCGGG GCGGTCTACG CCGGCCCGGA GCCACCTCCC
541 GTTTACACCA ACCTCAGCAG CTACTCCCCA GCCTCTGCGT CCTCGGGAGG CGCCGGGGCT
601 GCCGTCGGGA CCGGGAGCTC GTACCCGACG ACCACCATCA GTACCTCCC ACACGCGCCG
661 CCCTTCGCCG GTGGCCACCC GGCGCAGCTG GGCTTGGGCC GCGGCQCCTC CACCTTCAAG
721 GAGGAACCGC AGACCGTGCC GGAGGCGCGC AGCCGGGACG CCACGCCGCC GGTGTCCCCC
781 ATCAACATGG AAGACCAAGA GCGCATCAA GTGGAGCGCA AGCGGCTGCG GAACCGGCTG
841 GCGGCCACCA AGTGCCGGAA GCGGAAGCTG GAGCGCATCG CGCGCCTGGA GGACAAGGTG
901 AAGACGCTCA AGGCCGAGAA CGCGGGGCTG TCGAGTACCG CCGGCCTCCT CCGGGAGCAG
961 GTGGCCCAGC TCAAACAGAA GGTGATGACC CACGTCAGCA ACGGCTGTCA GCTGCTGCTT
1021 GGGGTCAAGG GACACGCCTT CTGA
```

//

SUBSTITUTE SHEET

59/61

FIG. 38A

LOCUS	HUMBJUNX	1044 bp	DNA		19-DEC-1991	
BASE COUNT	195 A	368 C	340 G	141 T		
ORIGIN						
1	ATGTGCACTA	AAATGGAACA	GCCCTTCTAC	CACGACGACT	CATACACAGC	TACGGGATAC
61	GGCCGGGCCC	CTGGTGGCCT	CTCTCTACAC	GA CTACAAAC	TCCTGAAACC	GAGCCTGGCG
121	GTCAACCTGG	CCGACCCCTA	CCGGAGTCTC	AAAGCGCCTG	GGGCTCGCGG	ACCCGGCCCCA
181	GAGGGCGGCG	GTGGCGGCAG	CTACTTTTCT	GGTCAGGGCT	CGGACACCGG	CGCGTCTCTC
241	AAGCTCGCCT	CTTCGGAGCT	GGAACGCCTG	ATTGTCCCCA	ACAGCAACGG	CGTGATCAGC
301	ACGACGCCCTA	CACCCCGGGG	ACAGTACTTT	TACCCCGCGG	GGGGTGGCAG	CGGTGGAGGT
361	GCAGGGGGCG	CAGGGGGCGG	CGTCACCGAG	GAGCAGGAGG	GCTTCGCCGA	CGGCTTTGTC
421	AAAGCCCTGG	ACGATCTGCA	CAAGATGAAC	CACGTGACAC	CCCCCAACGT	GTCCCTGGGC
481	GCTACCGGGG	GGCCCCCGGC	TGGGGCCGGG	GGCGTCTACG	CCGGCCCCGA	GCCACCTCCC
541	GTTTACACCA	ACCTCAGCAG	CTACTCCCCA	GCCTCTGCGT	CCTCGGGAGG	CGCCGGGGCT
601	GCCGTCTGGGA	CCGGGAGCTC	GTACCCGACG	ACCACCATCA	GCTACCTCCC	ACACGCGCGG
661	CCCTTCGCCG	GTGGCCACCC	GGCGCAGCTG	GGCTTGGGCC	GCGGCGCCTC	CACCTTCAAG
721	GAGGAACCGC	AGACCGTGCC	GGAGGCGCGC	AGCCGGGACG	CCACGCGGCC	GGTGTCCCCC
781	ATCAACATGG	AAGACCAAGA	GCGCATCAAA	GTGGAGCGCA	AGCGGCTCGG	GAACCGGCTG
841	GCGGCCACCA	AGTGCCGGAA	GCGGAAGCTG	GAGCGCATCG	CGCGCCTGGA	GGACAAGGTG
901	AAGACGCTCA	AGGCCGAGAA	CGCGGGGCTG	TCGAGTACCG	CCGGCCTCCT	CCGGGAGCAG
961	GTGGCCACAGC	TCAAACAGAA	GGTCATGACC	CACGTACGCA	ACGGCTGTCA	GCTGCTGCTT
1021	GGGGTCAAGG	GACACGCCTT	CTGA			

LOCUS	HUMCJUNX	996 bp	DNA		19-DEC-1991	
BASE COUNT	226 A	342 C	299 G	129 T		
ORIGIN						
1	ATGACTGCAA	AGATGGAAAC	GACCTTCTAT	GACGATGCCC	TCAACGCCTC	GTTCTCTCCG
61	TCCGAGAGCG	GACCTTATGG	CTACAGTAAC	CCCAAGATCC	TGAAACAGAG	CATGACCCTG
121	AACCTGGCCG	ACCCAGTGGG	GAGCCTGAAG	CCGCACCTCC	GCGCCAAGAA	CTCGGACCTC
181	CTCAGCTCGC	CCGACGTGGG	GCTGCTCAAG	CTGGCGTTCG	CCGAGCTGGA	GCGCCTGATA
241	ATCCAGTCCA	GCAACGGGCA	CATCACCACC	ACGCCGACCC	CCACCCAGTT	CCTGTGCCCC
301	AAGAACGTGA	CAGATGAGCA	GGAGGGGTTC	GCCGAGGGCT	TCGTGCGCGC	CCTGGCCGAA
361	CTGCACAGCC	AGAACACGCT	GCCCAGCGTC	ACGTGCGCGG	CGCAGCCGGT	CAACGGGGCA
421	GGCATGGTGG	CTCCCGCGGT	AGCCTCGGTG	GCAGGGGGCA	GCGGCAGCGC	CGGCTTCAGC
481	GCCAGCCTGC	ACAGCGAGCC	GCCGGTCTAC	GCAAACCTCA	GCAACTTCAA	CCCAGGCGCG
541	CTGAGCAGCG	GCGGCGGGGC	GCCCTCTTAC	GGCGCGGCGG	GCCTGGCCTT	TCCCGCGCAA
601	CCCCAGCAGC	AGCAGCAGCC	GCCGCACCAC	CTGCCCCAGC	AGATGCCCCG	GCAGACCCCG
661	CGGCTGCAGG	CCCTGAAGGA	GGAGCCTCAG	ACAGTGCCCG	AGATGCCCCG	CGAGACACCG
721	CCCCTGTCCC	CCATCGACAT	GGAGTCCCGA	GAGCGGATCA	AGGCGGAGAG	GAAGCGCATG
781	AGGAACCGCA	TCGCTGCCTC	CAAGTGCCCA	AAAAGGAAGC	TGGAGAGAAT	CGCCCGGCTG
841	GAGGAAAAAG	TGAAAACCTT	GAAAGCTCAG	AACTCGGAGC	TGGCGTCCAC	GGCCAACATG
901	CTCAGGGAAC	AGGTGGCACA	GCTTAAACAG	AAAGTCATGA	ACCACGTTAA	CAGTGGGTGC
961	CAACTCATGC	TAACGCAGCA	GTTGCAAACA	TTTTGA		

LOCUS	HUMDJUNX	1044 bp	ss-mRNA	PRI	24-MAY-1991
DEFINITION	Human junD mRNA				
ACCESSION	X56681				
KEYWORDS	jun-D gene; oncogene.				
SOURCE	Homo sapiens RNA.				
ORGANISM	Homo sapiens				
	Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;				
	Theria; Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.				
REFERENCE	1 (bases 1 to 1891)				
AUTHORS	Shaul, Y.				
JOURNAL	Unpublished (1990)				
STANDARD	full automatic				
REFERENCE	2 (sites)				

SUBSTITUTE SHEET

60/61

FIG. 38B

AUTHORS Berger, I. and Shaul, Y.
TITLE Structure and function of human jun-D
JOURNAL Unpublished (1990)
STANDARD full staff review
COMMENT From EMBL 26 entry HSJUNDR; dated 18-MAR-1991.
FEATURES Location/Qualifiers
 mRNA 1..1891
 /gene="junD"
 /evidence=EXPERIMENTAL
 CDS 175..1218
 /product="junD protein"
 /gene="junD"
 /codon_start=1
 polya_site 1891..1891
BASE COUNT 162 A 405 C 360 G 117 T
ORIGIN
1 ATGGAAACAC CCTTCTACGG CGATGAGGCG CTGAGCGGCC TGGGCGGGCG GCCCAGTGGC
61 AGCGGCGGCA CGTTCGCGTC CCGGGGCCGC TTGTTCCCGG GGGCGCCCCG GACGGCCGCG
121 GCCGGCAGCA TGATGAAGAA GGACGCGCTG ACGCTGAGCC TGAGTGAGCA GGTGGCGGCA
181 GCGCTCAAGC CTGCGCCCGC GCCCGCCTCC TACCCCTG CCGCCGACGG CGCCCCCAGC
241 GCGGCACCCC CCGACGGCCT GCTCGCCTCT CCCGACCTGG GGCTGCTGAA GCTGGCCTCC
301 CCCGAGCTCG AGCGCCTCAT CATCCAGTCC AACGGGCTGG TCACCACCAC GCCGACGAGC
361 TCACAGTTCC TCTACCCCAA GGTGGCGGCC AGCGAGGAGC AGGAGTTCGC CGAGGGCTTC
421 GTCAAGGCC TGGAGGATT ACACAAGCAG AACCAGCTCG GCGCGGGCCG GGCCGCTGCC
481 GCCGCCGCCG CCGCCGCCGG GGGGCCCTCG GGCACGGCCA CGGGCTCCGC GCGCCCGGGC
541 GAGCTGGCCC CCGCGGCGGC CGCGCCCGAA GCGCCTGTCT ACGCGAACCT GAGCAGCTAC
601 GCGGGCGGCG CCGGGGGCGC GGGGGGCGCC GCGACGGTCG CCTTCGCTGC CGAACCTGTG
661 CCCTTCCCGC CGCCGCCACC CCCAGGCGCG TTGGGGCCGC CGCGCTGGC TCGCTCAAG
721 GACGAGCCAC AGACGGTGCC CGACGTGCCG AGCTTCGGCG AGAGCCCGCC GTTGTGCCCC
781 ATCGACATGG ACACGCAGGA GCGCATCAAG GCGGAGCGCA AGCGGCTGCG CAACCGCATC
841 GCCGCCTCCA AGTGCCCAA GCGCAAGCTG GAGCGCATCT CGCGCTGGA AGAGAAAGTG
901 AAGACCCTCA AGAGTCAGAA CACGGAGCTG GCGTCCACGG CGAGCCTGCT GCGCGAGCAG
961 GTGGCGCAGC TCAAGCAGAA AGTCCTCAGC CACGTCAACA GCGGCTGCCA GCTGCTGCC
1021 CAGCACCAGG TCCCGCGCTA CTGA

//

LOCUS MUSBJUNX 1035 bp DNA 19-DEC-1991
BASE COUNT 210 A 333 C 333 G 159 T
ORIGIN
1 ATGTGCACGA AAATGGAACA GCCTTTCAT CACGACGACT CTTACGCAGC GGCGGGATAC
61 GGTGCGGAGCC CTGGCAGCCT GTCTCTACAC GACTACAAAC TCCTGAAACC CACCTTGGCG
121 CTCAACCTGG CGGATCCCTA TCGGGGTCTC AAGGGTCCTG GGGCGCGGGG TCCAGGCCCC
181 GAGGGCAGTG GGGCAGGCAG CTACTTTTCG GGTCAAGGAT CAGACACAGG CGCATCTCTG
241 AAGCTAGCCT CCACGGAAC TGGAGCGCTG ATCGTCCCCA ACAGCAACGG CGTGATCAGC
301 ACGACGCCCC CGCCTCCGGG ACAGTACTTT TACCCCTG GGGGTGGCAG CGGTGGAGGT
361 ACAGGGGGCG GCGTCACCGA GGAGCAGGAG GGCTTTGCGG ACGGTTTTGT CAAAGCCCTG
421 GACGACCTGC ACAAGATGAA CCACGTGACG CCCCCAACG TGTCCCTGGG CGCCAGCGGG
481 GGTCCCCAGG CCGGCCCAGG GGGCGTCTAT GCTGCTCCG AGCCGCTCC CGTCTACACC
541 AACCTCAGCA GTTACTCTCC AGCCTCTGCA CCCTCTGGAG GCTCCGGGAC CGCCGTCGGG
601 ACTGGGAGCT CATACCCGAC GGCCACCATC AGTACCTCC CACATGCACC ACCCTTTGCG
661 GGCGGCCACC CGGCACAGCT GGGTTTGGT CGTGGCGCTT CCGCCTTTAA AGAGGAACCG
721 GACACCTTAC CGGAGGCAG CAGCCGCGAC GCCACGCCG CTGTGTCCCC CATCAACATG
781 GAAGACCAGG AGCGCATCAA AGTGGAGCGA AAGCGGCTGC GGAACAGGCT CGCGGCCACC
841 AAGTGCCGGA AGCGGAAGCT GGAGCGCATC GCGCGCTGG AGGACAAGGT GAAGACACTC
901 AAGGCTGAGA ACGCGGGGCT GTGAGTGTCT GCCGTCTCC TAAGGGAGCA AGTGGCGCAG
961 CTCAAGCAGA AGGTCATGAC CCAATGTCAGC AACGGCTGCC AGTTGCTGCT AGGGGTCAAG
1021 GGACACGCCT TCTGA

//

SUBSTITUTE SHEET

61/61

FIG. 38C

LOCUS MUSCJUNX 1005 bp DNA 19-DEC-1991
 BASE COUNT 223 A 334 C 300 G 148 T
 ORIGIN

```

1 ATGACTGCAA AGATGGAAAC GACCTTCTAC GACGATGCCC TCAACGCCTC GTTCCTCCAG
61 TCCGAGAGCG GTGCCTACGG CTACAGTAAC CCTAAGATCC TAAACAGAG CATGACCTTG
121 AACCTGGCCG ACCCGGTGGG CAGTCTGAAG CCGCACCTCC GCGCCAAGAA CTCGGACCTT
181 CTCACGTCGC CCGACGTCGG GCTGCTCAAG CTGGCGTCGC CGGAGCTGGA GCGCCTGATC
241 ATCCAGTCCA GCAATGGGCA CATCACCCT ACACCGACCC CCACCCAGTT CTTGTGCCCC
301 AAGAACGTGA CCGACGAGCA GGAGGGCTTC GCCGAGGGCT TCGTGC CGC CCGTGGCTGAA
361 CTGCATAGCC AGAACACGCT TCCCAGTGTC ACCTCCGCGG CACAGCCGGT CAGCGGGGCG
421 GGCATGGTGG CTCCCGCGGT GGCCTCAGTA GCAGGCGCTG GCGGCGGTGG TGGCTACAGC
481 GCCAGCCTGC ACAGTGAGCC TCCGCTCTAC GCCAACCTCA GCAACTTCAA CCCGGGTGCG
541 CTGAGCAGCG GCGGTGGGGC GCCCTCCTAT GGCGCGGCGG GGCTGGCCTT TCCCTCGCAG
601 CCGCAGCAGC AGCAGCAGCC GCCTCAGCCG CCGCACCCT TGCCCCAACA GATCCCGGTG
661 CAGCAGCCGC GGCTGCAAGC CCTGAAGGAA GAGCCGCAGA CCGTGCCGGA GATGCCGGGA
721 GAGACGCCGC CCCTGTCCCC TATCGACATG GAGTCTCAGG AGCGGATCAA GGCAGAGAGG
781 AAGCGCATGA GGAACCGCAT TGCCGCCTCC AAGTGCCGGA AAAGGAAGCT GGAGCGGATC
841 GCTCGGCTAG AGGAAAAAGT GAAAACCTTG AAAGCGCAAA ACTCCGAGCT GGCATCCACG
901 GCCAACATGC TCAGGGAACA GGTGGCACAG CTTAAGCAGA AAGTCATGAA CCACGTAAAC
961 AGTGGGTGCC AACTCATGCT AACGCAGCAG TTGCAAACGT TTTGA

```

//

LOCUS MUSDJUNX 1026 bp DNA 19-DEC-1991
 BASE COUNT 172 A 382 C 343 G 129 T
 ORIGIN

```

1 ATGGAACGC CTTTCTATGG CGAGGAGGCG CTGAGCGGCC TGGCTGCGGG TGCGTGGAGC
61 GTCGCTGGTG CTACTGGGGC CCGCGGCGGT GGTGGCTTCG CGCCCCGGG CCGCGCTTTC
121 CCCGGGGCGC CCGCGACGAG CAGCATGCTG AAGAAAGACG CGCTGACGCT CAGCCTGGCG
181 GAGCAGGGAG CCGCGGGATT GAAACCAGGG TCGGCCACTG CACCTTCTGC GCTGCGCCCC
241 GACGGCGCCC CCGACGGGCT GCTGGCTTCG CCGGATCTTG GGCTGCTCAA ACTCGEGTCG
301 CCGGAGCTGG AGAGGCTGAT CATCCAGTCC AACGGGCTGG TGACCACTAC CCCGACCACT
361 ACGCAGTTCC TCTACCCGAA GGTGGCAGCC AGCGAGGAGC AGGAGTTTCG CGAAGGCTTC
421 GTCAAGGCGC TGGAGGACCT GCACAAGCAA AGCCAGCTGG GTGCGGCCAC CGCGGCCACC
481 TCAGGGGCTC CCGCGCCTCC CGCGCCCGCG GACCTGGCCG CCACCCCGG GGCCACGGAG
541 ACCCGGTCT ACGCCAACCT GAGCAGTTTC CCGGGTGGCG CCGGGCCCCG TGGGGGCGCG
601 GCCACCGTGG CTTTCGCGCG GGAGCCAGTG CCCTTCCCGC CGCCCCGGG CGCGCTGGGG
661 CCGCCGCCAC CTCCGCATCC ACCGCGCCTG GCCGCGCTCA AGGACGAGCC GCAGACCGTG
721 CCGGACGTGC CGAGCTTCGG CGACAGCCCT CCGCTGTCGC CCATCGACAT GGACACGCAA
781 GAACGCATCA AGGCGGAGCG CAAGAGGCTG CGCAACCGCA TCGCCGCCTC CAAATGCCGC
841 AAGCGCAAGC TGGAGCGTAT CTCGCGCCTG GAGGAGAAAG TCAAGACCTT CAAAAGCCAG
901 AACACCGAGC TGGCGTCCAC CGCCAGCCTG CTGCGCGAGC AGGTGGCCGA GCTCAAACAG
961 AAAGTCCTCA GCCACGTCAA CAGCGGCTGC CAGCTGCTGC CCCAGACCA GGTCCCGGCG
1021 TACTGA

```

//

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US93/00999

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :C12Q 1/68; C12P 19/34; C07H 21/04

US CL :435/6, 91; 536/24.31, 24.32

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; 536/24.31, 24.32

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)

Medline, Biosis, APS

search terms: sandwich hybridization, rRNA, fungus, jun, substance P receptor, Beta receptor, G-protein

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X - Y	Clin. Chem., Vol. 31, Number 9, issued 1985, Polsky-Cynkin et al., "Use of DNA Immobilized on Plastic and Agarose Supports to Detect DNA by Sandwich Hybridization", pages 1438-1443, see entire document.	1, 5, 10-11, 17-18, 20-21 ----- 2-4, 6-9, 12-16, 19, 22-78
Y	Laboratory Investigation, Vol. 63, Number 4, issued 1990, Hayashi et al., "A Novel Diagnostic Method of Pneumocystis carinii", pages 576-580, see entire document.	4, 6, 8, 29-30, 77-78



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

A document defining the general state of the art which is not considered to be part of particular relevance

E earlier document published on or after the international filing date

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)

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

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

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

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

Z

document member of the same patent family

Date of the actual completion of the international search

11 May 1993

Date of mailing of the international search report

17 MAY 1993

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. NOT APPLICABLE

Authorized officer

KENNETH R. HORLICK

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US93/00999

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	J. Bacteriology, Vol. 170, Number 2, issued February 1988, Giovannoni et al., "Phylogenic Group-Specific Oligodeoxynucleotide Probes for Identification of Single Microbial Cells", pages 720-726, see entire document.	4,6,8,29-30,79-92
Y	Nucleic Acids Research, Vol. 18, Number 11, issued 1990, Kitabayashi et al., "Nucleotide Sequence of Rat c-jun Protooncogene", EMBL Accession No. X17215, see entire document.	31-32,45-46,65-66,71-72
Y	Science, Vol. 247, issued 23 February 1990, Hershey et al., "Molecular Characterization of a Functional cDNA Encoding the Rat Substance P Receptor", pages 958-961, see entire document.	33,47,67,74
Y	FEBS Letters, Vol. 211, Number 2, issued January 1987, Chung et al., "Cloning and Sequence Analysis of the Human Brain B-adrenergic Receptor", pages 200-206, see entire document.	34,48,68,73
Y	FEBS Letters, Vol. 219, Number 1, issued July 1987, Didsbury et al., "Molecular Cloning of a New Human G Protein", pages 257-263, see entire document.	35-36,49-50,69-70,75-76
Y	Science, Vol.239, issued 29 January 1988, Saiki et al., "Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase", pages 487-491, see entire document.	79-92

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US93/00999

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:

(Telephone Practice)

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.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US93/00999

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

- I. Claims 1-30, 37-44, 51-64, 77-78, drawn to methods of detecting organisms including fungi via hybridization, solid supports for use in said methods, kits for use in said methods, and polynucleotides for use in said methods, classified in class 435/6 for example.
- II. Claims 31-32, 45-46, 65-66, 71-72, drawn to methods of detecting a jun oncogene, and solid supports, kits, and polynucleotides for use in said methods, classified in class 435/6, for example.
- III. Claims 33, 47, 67, 73, drawn to methods of detecting a substance P receptor, and solid supports, kits, and polynucleotides for use in said methods, classified in class 435/6, for example.
- IV. Claims 34, 48, 68, 74, drawn to methods of detecting a B-receptor, and solid supports, kits, and polynucleotides for use in said methods, classified in class 435/6, for example.
- V. Claims 35-36, 49-50, 69-70, 75-76, drawn to methods of detecting a G-protein, and solid supports, kits, and polynucleotides for use in said methods, classified in class 435/6, for example.
- VI. Claims 79-92, drawn to methods of detecting organisms via nucleic acid amplification, classified in class 435/91.