OPIC
Office de la propriété intellectuelle du Canada

CIPO
Canadian Intellectual Property Office

(12) (19) (CA) **Demande-Application**

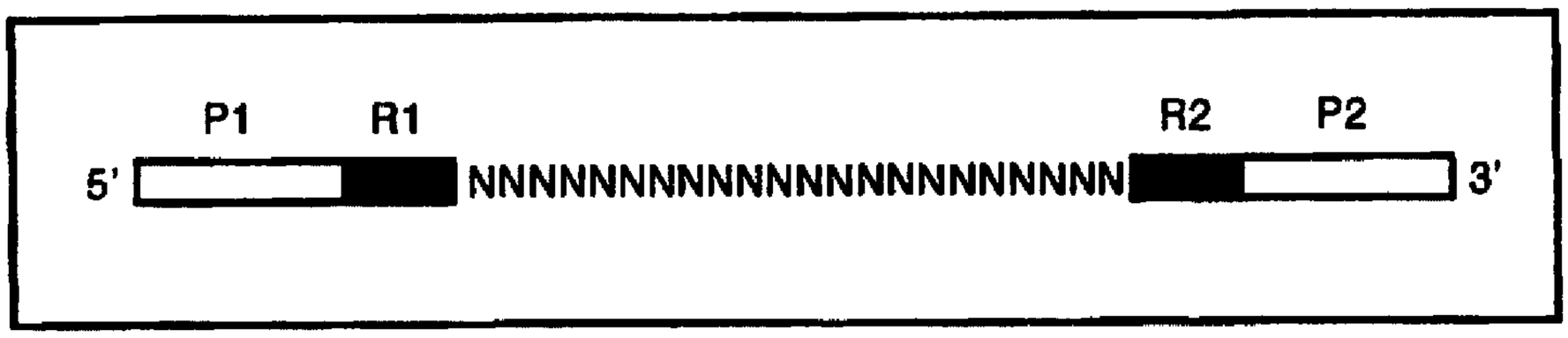
(21) (A1) **2,296,771**
(86) 1998/07/14
(87) 1999/01/21

(72) SCHWARTZ, ROBERT J., US
(72) EASTMAN, ERIC M., US
(72) LI, XUYANG, US
(72) NORDSTROM, JEFF, US
(71) VALENTIS, INC., US

(51) Int.Cl.[7] C12Q 1/68, C12N 15/10
(30) 1997/07/14 (60/052,403) US

(54) **PROCEDE D'IDENTIFICATION DE REGIONS DE SYNTHESE REGULATRICES DE TRANSCRIPTION SPECIFIQUES DE CELLULES OU DE TISSUS**

(54) **METHOD FOR THE IDENTIFICATION OF SYNTHETIC CELL- OR TISSUE-SPECIFIC TRANSCRIPTIONAL REGULATORY REGIONS**



(57) L'invention concerne la production et l'évaluation de régions de synthèse régulatrices destinées a réguler l'expression de gènes. L'invention a trait à un procédé d'identification de sites de liaison de facteur de transcription et à un procédé d'évaluation des fonctions régulatrices de régions de synthèse régulatrices.

(57) The invention concerns making and evaluating synthetic regulatory regions for controlling gene expression. The invention features a method for identifying transcription factor binding sites and a method for evaluating the regulatory functions of synthetic regulatory regions.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68, C12N 15/10 — A1

(11) International Publication Number: **WO 99/02737**

(43) International Publication Date: 21 January 1999 (21.01.99)

(21) International Application Number: PCT/US98/14901

(22) International Filing Date: 14 July 1998 (14.07.98)

(30) Priority Data:
60/052,403 14 July 1997 (14.07.97) US

(71) Applicant *(for all designated States except US)*: GENEMEDICINE, INC. [US/US]; 8301 New Trails Drive, The Woodlands, TX 77381–4248 (US).

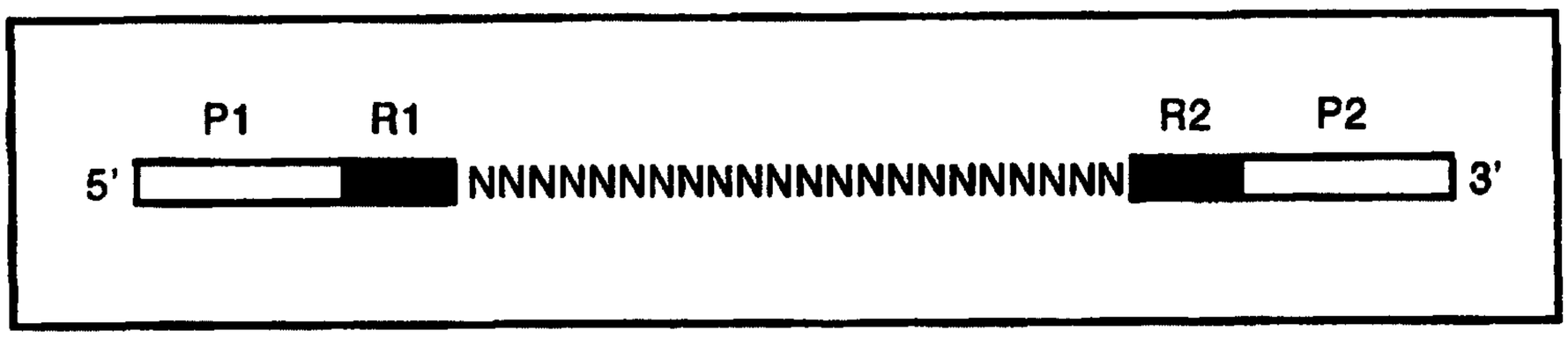
(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: SCHWARTZ, Robert, J. [US/US]; 4019 Marlowe, Houston, TX 77005 (US). EASTMAN, Eric, M. [US/US]; 7465 Mink Hollow Road, Highland, MD 20777 (US). LI, Xuyang [CN/US]; 2111 Holly Hall #310, Houston, TX 77054 (US). NORDSTROM, Jeff [US/US]; 1016 Puryear Drive, College Station, TX 77840 (US).

(74) Agents: WARBURG, Richard, J. et al.; Lyon & Lyon LLP, Suite 4700, 633 West Fifth Street, Los Angeles, CA 90071–2066 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHOD FOR THE IDENTIFICATION OF SYNTHETIC CELL– OR TISSUE–SPECIFIC TRANSCRIPTIONAL REGULATORY REGIONS



(57) Abstract

The invention concerns making and evaluating synthetic regulatory regions for controlling gene expression. The invention features a method for identifying transcription factor binding sites and a method for evaluating the regulatory functions of synthetic regulatory regions.

DESCRIPTION

METHOD FOR THE IDENTIFICATION OF SYNTHETIC CELL- OR TISSUE-SPECIFIC TRANSCRIPTIONAL REGULATORY REGIONS

Background of the Invention

This invention relates to natural and synthetic cell- or tissue-specific transcriptional regulatory regions that regulate gene transcription in particular cells or tissues. In addition, this invention also relates to the methods for the selection, identification and evaluation of the synthetic cell- or tissue- specific transcriptional regulatory regions.

None of the information described herein is admitted to be prior art to the present invention, but is provided solely to assist the understanding of the reader.

Cell- or tissue-specific gene expression plays a central role in the proliferation and differentiation of cells. As the first step of gene expression, transcription is an important step for regulation. The study of transcriptional regulatory regions is one of the major fields in modern biology. The transcriptional regulatory regions are also very important for applications in biotechnology, such as in gene therapy and the production of recombinant proteins.

Transcriptional regulatory regions generally have two portions: transcription initiation sites and enhancers which are capable of regulating the transcription level from a distance to the initiation sites. The binding of transcription factors to the regulatory regions is necessary for the regulatory regions to regulate transcription. The regulatory regions fall into several categories: general regulatory regions which regulate transcription in all cells of an organism, inducible regulatory regions which only regulate

transcription in response to certain signals, and cell- or tissue-specific regulatory regions which only regulate transcription in certain cells.

Several methods have been used to identify the regulatory regions. One of these methods is the analysis of regions that are important for the proper expression of cloned genes. The first step is usually to identify rough boundaries of the regulatory regions using deletion and mutation analysis of the cloned genes. These regions include the 5' upstream regions, 3' downstream regions, and sometimes introns or coding sequences within the gene itself. Most studies are performed using chimeric constructs containing a reporter gene such as β-galactosidase (β-gal), chloramphenicol acetyltransferase (CAT), luciferase or growth hormone (GH). The regions that actually bind protein factors can be more accurately defined using DNA footprinting techniques followed by mutation analysis. The sequences that bind protein transcription factors are often referred to as transcription factor binding sites.

Consensus sequences for a number of common binding sites have been determined. One example is the binding site recognized by the family of basic-helix-loop-helix (bHLH) transcription factors. The consensus sequence of binding sites for bHLH proteins is 5'-CANNTG-3', where "N" can be any nucleotide. This binding site is called the "E box" and is found in the regulatory regions of a number of genes that are expressed in diverse cell types, including lymphocytes, muscle cells and fibroblasts. Some bHLH proteins are common to most or all cells while others are cell-specific. In addition, bHLH proteins form heterodimers and the interaction of some of these dimers with DNA is cell-specific. The binding of different bHLH proteins to specific regulatory regions appears to be

affected by the variable dinucleotide sequence within the core consensus sequence and the sequence adjacent to the core sequence (Sun, et al., *Cell* 64:459-470 (1991)).

Binding sites associated with newly cloned and sequenced genes can also be identified by searching the sequence for homology with the sequences of known binding sites that have been characterized from other, sometimes related, genes.

In addition, several methods were developed to identify the binding sites of transcription factors without cloning of the target genes. Selected and amplified binding site (SAAB) method was used to identify the binding sites for known transcription factors (Blackwell, et al., *Science* 250:1104-1110 (1990)). By using this method, synthesized templates with random sequences are incubated with purified transcription factors. Those bound to transcription factors are isolated with electrophoretic mobility shift assay (EMSA). The templates are then amplified by the polymerase chain reaction (PCR). After reiteratively being rebound and reamplified, the binding site of the transcription factor is sequenced and identified. The binding site of transcription factor myc was identified with this method (Blackwell, et al., *Science* 250:1149-1151 (1990)).

It is often difficult, however, to identify and purify transcription factors for use in such assays. Indeed, the binding sites are often identified first and then are used to facilitate the identification and purification of transcription factors binding to the sites. Moreover, in many studies, it is crucial to understand the characteristics of certain regulatory regions, whereas it is not necessary to know the transcription factors binding to the regulatory regions. A method similar to SAAB, multiplex selection technique (MuST) was therefore developed (Nullur, et al., *PNAS* 93:1184-1189

(1996)). In the multiplex selection technique, purified transcription factors are replaced with crude nuclear extract, so that binding sites can be identified without the identification of transcription factors. The identified binding sites can then be used to identify the corresponding transcription factors.

The regulatory regions often consist of multiple different binding sites for transcription factors. The characteristics of a regulatory region are determined by the composition and arrangement of the binding sites. In addition to naturally-occurring regulatory regions, synthetic regulatory regions can be constructed through the combination and modification of binding sites.

Available naturally-occurring regulatory regions are not always capable of regulating transcription in a desired manner. In these cases, as well as others, synthetic regulatory regions may be utilized to provide the desired functional characteristics. As an example, synthetic herpes simplex virus (HSV) regulatory regions were constructed by linking the 5' nontranscribed domain of an HSV α gene to a fragment containing the transcription initiation site and the 5' transcribed noncoding region from an HSV λ gene (Roizman, PCT 94/14971). The resulting synthetic regulatory regions direct constitutive transcription of the heterologous gene throughout the reproductive cycle of the virus at a high cumulative level. Synthetic regulatory regions were also constructed to achieve high inducible transcription levels and low basal transcription levels (Filmus, et al., PCT 93/20218).

In both of the above cases, the binding sites are well-understood transcription factor response elements. Many binding sites, however, are not well-understood, especially those identified without the cloning of the corresponding

transcription factors. These binding sites are therefore only potential transcription factor response elements until they are confirmed to be functional for transcription regulation using functional assays. These assays are usually a laborious and costly task. It is even more complicated for synthetic regulatory regions produced by the combination, modification and rearrangement of various binding sites.

Summary of the Invention

Applicant has designed useful methods to create, identify and evaluate cell- or tissue-specific synthetic regulatory regions. Specifically, the methods include the selection of transcription factor binding sites, the creation of synthetic regulatory regions using the binding sites and/or portions of known regulatory regions, and the evaluation of the synthetic regulatory regions. The synthetic regulatory regions acquired with this method can be used in gene delivery or gene therapy to achieve desired gene expression in targeted cells. The acquired synthetic regulatory regions can also be used to achieve the production of recombinant proteins at high levels.

The present invention utilizes the recognition that the cells themselves contain all the information required to identify the binding sites that are most important or are recognized by the key transcription factors in the cells. The methods described for the selection of binding sites do not require any previous knowledge of the genes that are expressed or the transcription factors that are present in the cells. Thus, these methods bypass the extensive work needed for the purification, identification, and analysis of transcription factors. In addition, these methods eliminate the need to know the tissue specific transcription factor binding sites. Furthermore, many more potential binding sites can be

identified using these methods than using the methods with purified transcription factors. Similarly, the methods for the creation and evaluation of synthetic regulatory regions do not require complete understanding of the binding sites. The binding sites can be linked together in various combinations and with various arrangements, and can then be evaluated to select particular synthetic regulatory regions which are functional in a certain cell line. Therefore, these methods make it possible to create and identify useful synthetic regulatory regions on a large-scale.

As indicated above, the methods discussed herein are useful for identifying regulatory region sequences for gene delivery or gene therapy. One of the major obstacles for gene delivery or gene therapy is the difficulty in expressing genes at preferred levels in certain cells or tissues. The difficulties are partly due to the lack of proper regulatory regions to direct the desired gene transcription. The functional synthetic regulatory regions identified from these methods will provide many candidates for the regulatory regions needed in gene delivery or gene therapy. Moreover, these synthetic regulatory regions will also be candidates for the regulatory regions needed in large-scale production of recombinant proteins, which also requires gene transcription at high level in certain cell lines.

A first aspect of the present invention features a method of identifying binding sites for transcription factors. The method involves identifying the oligonucleotides in protein-oligonucleotide complexes formed between a cellular or nuclear extract from a group of cells and any of a plurality of double-stranded oligonucleotide fragments. Preferably the complexes are separated from free oligonucleotides using size exclusion chromatography. The presence of an oligonucleotide in a complex is indicative that the oligonucleotide includes a

binding site.

In preferred embodiments, the double-stranded oligonucleotides are made through the synthesis of single-stranded oligonucleotide and conversion of the single-stranded oligonucleotide to double-stranded oligonucleotide. Also in preferred embodiments, the oligonucleotide fragment has a central random sequence and both restriction sites and primer sequences on both ends. In preferred embodiments, the identifying step includes amplifying, cloning and sequencing the oligonucleotide fragments from the protein-oligonucleotide complexes to identify the binding sites. The amplifying step is preferably performed by polymerase chain reaction.

The oligonucleotide fragments can be of various sizes, but preferably include test sequences between about 5 and 500 bp in length, more preferably between about 5 and 100 bp, still more preferably between 20 and 50 bp.

The term "transcribe" or "transcription" as used herein refers to the synthesis of RNA by RNA polymerase, following a DNA template. Transcription is the first step of gene expression and the most important step for the regulation of gene expression. That is, the regulation of gene expression is achieved mainly through the regulation of transcription.

The term "gene expression" refers to the process in which genetic information flows from DNA to functional molecules, such as proteins or RNA molecules. The regulation of transcription, as a part of gene expression is achieved with the interaction between the regulatory region of a gene and various transcription factors.

As used herein, the term "transcriptional regulatory regions" or regulatory regions" refers to the regions of a gene controlling the transcription of the gene. A regulatory region often includes several portions. Some of these portions are in the initiation site for transcription, whereas others are

located a distance to the initiation site. The term thus includes regions commonly referred to as enhancers.

The term "synthetic regulatory regions" as used herein refers to regulatory regions which are artificially made (i.e., made by humans using molecular biology techniques) such as by the creation with one or more modifications, combinations, or rearrangements of various transcription factor binding sites.

The term "transcription factors" as used herein refers to proteins which bind to the elements of regulatory regions and regulate the transcription of the corresponding genes. According to their functions, transcription factors fall into several categories. These include general transcription factors which are needed by most genes in most cells, cell- or tissue-specific transcription factors which only regulate gene transcription in certain cells, and inducible transcription factors which regulate gene transcription in response to certain signals.

The term "transcription factor binding site" or binding site" refers to any nucleic acid sequence which can bind transcription factors under transcription conditions or conditions approximating intracellular physical conditions.

As used herein, the term "transcription factor response elements" or "response elements" refers to the functional regulatory region components which can bind transcription factors and thereby regulate transcription of the corresponding genes. Thus, binding sites are potential response elements, their regulatory function can readily be tested and characterized.

As used herein, the term "restriction sites" refers to deoxyribonucleic acid sequences at which specific restriction endonucleases can cleave in a sequence-specific manner.

The term "cells" or "cell" as used herein refers to a membrane-enveloped protoplasmic body capable of independent

reproduction. Cells can be maintained, or propagated, *in vivo*, *in vitro* or in tissue culture and are capable of being transformed by plasmids as discussed herein.

As used herein "tissue" refers to a population consisting of cells of the same kind performing the same function.

The term "nuclear or cellular extract" refers to a preparation containing all or some of the cellular contents from inside the nuclear membrane or the plasma membrane of cells respectively, particularly including protein components. Such an extract is distinguished from a purified transcription factor.

As used in this context, the term "mixing" refers to putting together oligonucleotides and nuclear or cellular extract, such that the oligonucleotides and components of the extract can contact each other. Preferably a nuclear extract is used.

The term "oligonucleotide" as used herein refers to a nucleic acid molecule consisting of same or different individual nucleotides which are covalently linked together. Oligonucleotides can be single-stranded or double-stranded, consisting of two anti-parallel single-stranded oligonucleotides with complementary sequences. For use in the identification of binding sites, each oligonucleotide strand is preferably between about 5 and 500 nucleotides in length, more preferably between 5 and 100, still more preferably between about 7 and 50, and most preferably between about 20 and 50 nucleotides in length. The term "free oligonucleotide" refers to the oligonucleotides which are not bound to proteins or any other compounds. The term "protein-oligonucleotide complexes" as used herein refers to the complexes comprising oligonucleotides and the proteins bound with the oligonucleotides.

As used in the context of the oligonucleotide fragments, the term "conversion" is used to refer to the synthesis of a single-stranded DNA molecule complementary to another DNA molecule to form a double-stranded DNA molecule.

The term "primer" as used herein refers to a single-stranded oligonucleotide, the 3' end of which can be used as the initiation site for the DNA synthesis with a DNA polymerase. As used herein, the term "primer sequence" refers to the sequence of the primer or the complementary sequence.

As used herein, the terms "5'" and "3'" refer to the two different ends of a single-stranded DNA molecule respectively in accord with common usage. When used in relation to a coding sequence, the terms refer to being in the 5' direction from the coding sequence or in the 3' direction from the coding sequence. For a sequence on a circular nucleic acid molecule, e.g., on a circular plasmid, the terms refer to the direction from a reference sequence but not fully around the chain, and preferably includes a functional relationship. Thus, for example, a regulatory region is 5' to a coding sequence if it is in a position in which it would be expected to functionally affect transcription if in a 5' position on a linear molecule. Usually, a 5' position is closer to the 5' end of a coding sequence than to the 3' end.

As used herein, the term "size exclusion chromatography" refers to a technique for the separation of biomolecules. This approach separates molecules into two groups, one which is smaller than the exclusion size of the chromatographic media and another which is larger than the exclusion size. The protein-oligonucleotide complexes are much larger than free oligonucleotides, so they can be readily separated, utilizing an exclusion size greater than the size of the free oligonucleotides and smaller than the size of protein-oligonucleotide complex. In this context, size refers to the

effective radius of the molecule or complex. As indicated above, nuclear or cellular extract, which includes many different transcription factors, is used instead of purified transcription factors in the present invention. The protein-oligonucleotide complexes resulting from the mixing of oligonucleotide fragments and nuclear or cellular extract therefore have many different sizes. As a result, size exclusion chromatography provides a more useful separation than electrophoretic mobility shift assay (EMSA) because size exclusion chromatography produces a simple separation of bound and unbound oligonucleotides while EMSA produces a series of bands distributed over a gel. Due to the nature of the gels typically utilized, EMSA generally also requires an extraction step to recover the bound oligonucleotide from the gel for further manipulation.

The term "amplifying" as used herein refers to increasing the numbers of DNA molecules. The approaches for amplifying include, but are not limited to, polymerase chain reaction.

As used herein, the term "sequencing" refers to the process of identifying the nucleotide sequence of DNA molecules. The term "nucleotide sequence" refers to the linear order of nucleotides in a DNA molecule or other nucleic acid molecules. Methods for sequencing of nucleic acid molecules are well-known to those skilled in the art.

A second aspect of the present invention features a method for evaluating a cell- or tissue-specific synthetic regulatory region or regions. This method involves determining whether a cell is selected under selective conditions. The method uses cells which contain different putative transcriptional regulatory regions located in transcriptional regulatory positions to a selective gene. A cell can only be selected if the selective gene is expressed at sufficiently high levels, and the selective gene will be expressed at the

sufficiently high level if the putative transcriptional regulatory region is active in the particular cell. The capability of a cell to be selected in response to the selection condition indicates that the nucleic acid test sequence contains a transcriptional regulatory region active in the cell. The selection condition can be adjusted so that only strong regulatory regions will be effective to be selected in the selection condition. In general, the method involves culturing the cell or cells having the putative transcriptional regulatory sequence

The term "sufficiently high level" refers to a functional level of expression which depends on the type of selection used and the stringency applied to the selection. Thus, for positive selection, the level is sufficient to allow discrimination of a cell expressing the selective gene at a "sufficiently high level" from an otherwise isogenic cell not expressing the gene at a sufficiently high level. For negative selection, a "sufficiently high level" is a level which allows the cell to grow in the presence of the selection condition.

In a preferred embodiment, the selection condition is a positive selection condition. The capability of at least one cell to be selected in the presence of the selective condition is indicative that the nucleic acid test sequence contains a transcriptional region active in the cell. The selection condition can be adjusted so that only strong regulatory regions will be effective to be selected in the selection condition.

In another preferred embodiment, the selection condition is a negative selection condition, i.e., stress condition; and the selective gene is a protective gene. The growth of the cells is inhibited under the stress condition in the absence of high level expression of the protective gene. Growth of at least one cell in the presence of the stress condition is

indicative that the nucleic acid test sequence contains a transcriptional region active in the cell. The stress condition can be adjusted so that only strong regulatory regions will be effective to overcome the stress condition.

The term "regulates" or "regulation" as used herein refers to the effect of nucleic acid sequences or other molecules involved in control of a response or action. In particular, this includes the effects of sequences involved in regulating, controlling or affecting the expression level or rate of structural genes. Generally this includes the binding of transcription factors to sequences, affecting transcription rates or other steps in gene expression.

As used in this context, the term "transcriptional regulatory position" refers to the position where functional regulatory regions can influence the transcription of the selective gene. Transcriptional regulatory positions include, but are not limited to, 5' to the coding sequence of the selective gene, 3' to the coding sequence of the selective gene, and within the intron or signal sequence of the selective gene. For identification and/or evaluation of synthetic regulatory regions, the region 5' to the coding sequence of the selective gene is of particular interest, however, other positions are also of interest and can be utilized in this invention.

The term "cell- or tissue-specific transcriptional regulatory region" as used herein refers to a nucleic acid sequence which is involved in controlling transcription through one or more coding sequences in a cell- or tissue-specific manner. As used herein, the term "cell- or tissue-specific transcription" refers to the gene transcription which occurs at a higher level in cells of a group or in certain tissue as compared to other cells or tissue of the corresponding organism generally.

As used herein the term "transfected" or "transfection" refers to the incorporation of foreign DNA into cultured cells by exposing them to such DNA. This would include the introduction of DNA by various delivery methods, *e.g.*, via vectors or plasmids using naked DNA, DNA-cationic lipid complexes, DNA in liposomes. The methods may include techniques to enhance penetration of the cellular membrane, such as electroporation or use of lytic peptides.

The term "cells of a group" as used herein refers to cells which are differentiated into the same or similar stage, and thereby have the same or similar characteristics, e.g., the same or similar characteristics with respect to control of transcription.

As used herein, the term "vector" refers to a DNA construct which can be transfected into cells. Vectors can be of a variety of different types, including plasmids, viral vectors, and others. Various genes can be inserted into a vector so that the gene can be delivered into cells. The term "insert" as in this context refers to incorporating a nucleic acid sequence into the vector nucleic acid sequence. Vector can include both linear and circular DNA constructs.

The term "selection condition", refers to conditions, under which cells expressing a selective gene show distinguishing features, and thereby can be easily separated from cells not expressing a selective gene. Selection condition can be positive selection condition, or negative selection condition, i.e., stress condition.

The term "positive selection conditions" refers to conditions which distinguish cells expressing the selective gene so that these cells can be easily isolated. The positive selection can be, but not limited to, Fluorescence Activated Cell Sorting (FACS) and magnetic bead sorting.

The term "selective gene" refers to a gene whose expression confers on its host cells a special feature which allows the host cell to be distinguished from other cells with which the host cell is associated. The selective gene can be, but is not limited to, a gene coding a particular antigen or antibody, or a protective gene.

The term "stress conditions" refers to conditions which either kill the cells or inhibit the division and proliferation of the cells. Such stress conditions include, but are not limited to, 1) elevated temperatures; 2) radiation; and 3) contact with particular biochemical agents.

The term "protective gene" means a gene encoding a protein which is capable of protecting cells from a stress condition. Such protective genes include, but are not limited to, genes for 1) adenosine deaminase; 2) dihydrofolate reductase; and 3) heat shock proteins.

The term "biochemical agents" as used herein refers to compounds which kill certain cells or inhibit the division and proliferation of certain cells. These biochemical agents include, but are not limited to, 1) xylofuranosyl-adenine; 2) methotrexate; 3) xylofuranosyl-adenine and deoxycorformacin; 4) alanosine, adenosine, and uridine.

As used in connection with binding sites and regulatory regions, the term "combination" refers to linking together two or more of the same or different kinds of oligonucleotides.

The term "modification" refers to a change in the sequence of a DNA molecule, which includes, but is not limited to, the substitution of one or a few nucleotides, or the addition or deletion of one or a few nucleotides as compared to a reference sequence. The term "rearrangement" refers to one or more changes in the order of subsequences of a regulatory region, and can include the insertion of a new subsequence or replacement of a subsequence with a new subsequence. This

includes combinations of re-ordering, substitution, and insertion of subsequences.

A third aspect of the present invention features a method, which combines both of the above aspects, for evaluating a cell- or tissue-specific transcriptional regulatory region. The method involves identifying the oligonucleotides in protein-oligonucleotide complexes formed between a cellular or nuclear extract from a group of cells and any of a plurality of double-stranded oligonucleotide fragments. The presence of an oligonucleotide in a complex is indicative that the oligonucleotide includes a binding site. One or more cells are then cultured under a selection condition. Among the cells, at least one cell, and preferably a plurality of cells, contains a nucleic acid test sequence inserted in a transcriptional regulatory position to a selective gene. The test sequence consists of at least one of the binding sites identified using the cellular or nuclear extract. The capability of at least one cell to be selected in the presence of the selection condition is indicative that the nucleic acid test sequence contains a transcriptional region active in the cell. The selection condition can be adjusted so that only strong regulatory regions will be effective to be selected in the selection condition.

In addition, in another aspect, the invention provides synthetic regulatory regions which include all or portions of the synthetic regulatory regions described in Example 5 and in the Drawings. Preferably the synthetic regulatory region is in a transcriptional regulatory position with respect to a coding sequence of interest. A portion of one of the described regions preferably includes at least 20 contiguous nucleotides, more preferably at least 40 contiguous nucleotides, and still more preferably at least 80 contiguous nucleotides of one of the described synthetic regulatory regions. Preferably the

portion is placed at about the same position relative to a coding sequence as it occupied in the plasmids used for analysis as described herein. Thus, the portion is preferably within 100 nucleotides, more preferably within 60 nucleotides, and still more preferably within 30 nucleotides of the position it occupied in a corresponding described synthetic regulatory region.

Other features and advantages of the invention will be apparent from the following detailed description of the invention in conjunction with the accompanying drawings and from the claims.

Brief Description of the Drawings

Figure 1 shows five important features for the synthetic single-stranded oligonucleotides (oligos) used in the described selection method.

Figure 2 outlines the overall scheme for an embodiment of transcription factor selection of regulatory regions.

Figure 3 is a comparison of relative regulatory region activity of a number of different regulatory regions during differentiation in primary myoblast cells.

Figure 4 shows the differential SRF activity on c-Fos SRE vs muscle SRE.

Figure 5 shows the arrangement of sub-elements of some exemplary synthetic regulatory regions.

Figure 6 is a bar graph showing the expression levels in myotubes of the luciferase reporter gene driven by various synthetic regulatory regions in comparison to the expression driven by the skeletal α-actin promoter, and the expression level of each of the synthetic regulatory regions in the presence of KCl depolarization.

Figure 7 shows the activities of exemplary regulatory

regions under the nerve-injury induced down-regulation of skeletal actin. Tibiales muscle of ICR mice were injected with 100 μg of clone skeletal α-actin promoter 448 (control), synthetic regulatory region C1-28, and C5-12 luciferase vectors. Two weeks post sciatic nerve crush, the muscle was harvested and assayed for luciferase reporter gene activity.

Figure 8 shows the sequence of a portion of the plasmid containing the synthetic regulatory region of clone C1-28, including the sequence of the synthetic regulatory region insert.

Figure 9A & B show two independently determined sequences of portions of the plasmid containing the synthetic regulatory region of clone C2-27, including the sequence of the synthetic regulatory region insert.

Figure 10A & B show two independently determined sequences of portions of the plasmid containing the synthetic regulatory region of clone C5-12, including the sequence of the synthetic regulatory region insert.

Figure 11A & B show two independently determined sequences of portions of the plasmid containing the synthetic regulatory region of clone C6-16, including the sequence of the synthetic regulatory region insert.

Figure 12 shows the sequence of a portion of the plasmid containing the synthetic regulatory region of clone C6'-7, including the sequence of the synthetic regulatory region insert.

Figure 13A & B show two independently determined sequences of portions of the plasmid containing the synthetic regulatory region of clone C5-1, including the sequence of the synthetic regulatory region insert.

Figure 14A & B show two independently determined sequences of portions of the plasmid containing the synthetic

regulatory region of clone C5-5, including the sequence of the synthetic regulatory region insert.

Figure 15A & B show two independently determined sequences of portions of the plasmid containing the synthetic regulatory region of clone C6-5, including the sequence of the synthetic regulatory region insert.

Figure 16A & B show two independently determined sequences of portions of the plasmid containing the synthetic regulatory region of clone C1-1, including the sequence of the synthetic regulatory region insert.

Figure 17A & B show two independently determined sequences of portions of the plasmid containing the synthetic regulatory region of clone C1-14, including the sequence of the synthetic regulatory region insert.

Figure 18 shows the sequence of a portion of the plasmid containing the synthetic regulatory region of clone C1-20, including the sequence of the synthetic regulatory region insert.

Figure 19 shows the sequence of a portion of the plasmid containing the synthetic regulatory region of clone C1-21, including the sequence of the synthetic regulatory region insert.

Figure 20A & B show two independently determined sequences of portions of the plasmid containing the synthetic regulatory region of clone C1-26, including the sequence of the synthetic regulatory region insert.

Figure 21A & B show two independently determined sequences of portions of the plasmid containing the synthetic regulatory region of clone C2-26, including the sequence of the synthetic regulatory region insert.

Figure 22A & B show two independently determined sequences of portions of the plasmid containing the synthetic regulatory region of clone C5-13, including the sequence of the

synthetic regulatory region insert.

Figure 23 shows the sequence of a portion of the plasmid containing the synthetic regulatory region of clone C5'-3, including the sequence of the synthetic regulatory region insert.

Figure 24 shows the sequence of a portion of the plasmid containing the synthetic regulatory region of clone C5'-5, including the sequence of the synthetic regulatory region insert.

Figure 25 shows the sequence of a portion of the plasmid containing the synthetic regulatory region of clone C5'-9, including the sequence of the synthetic regulatory region insert.

Figure 26 shows the sequence of a portion of the plasmid containing the synthetic regulatory region of clone C5'-12, including the sequence of the synthetic regulatory region insert.

Figure 27A & B show two independently determined sequences of portions of the plasmid containing the synthetic regulatory region of clone C6-12, including the sequence of the synthetic regulatory region insert.

Figure 28 shows the sequence of a portion of the plasmid containing the synthetic regulatory region of clone C6'-8, including the sequence of the synthetic regulatory region insert.

Figure 29 shows the sequence of a portion of the plasmid containing the synthetic regulatory region of clone C6'-10, including the sequence of the synthetic regulatory region insert.

Figure 30 shows the sequence of a portion of the plasmid containing the synthetic regulatory region of clone C6'-11, including the sequence of the synthetic regulatory region insert.

Figure 31 shows the sequence of a portion of the plasmid containing the synthetic regulatory region of clone C6'-22, including the sequence of the synthetic regulatory region insert.

Detailed Description of the Invention

The present invention provides methods for identifying and selecting transcription factor binding sites and methods for creating and evaluating synthetic regulatory regions or identified transcriptional regulatory regions. The following description is offered by way of illustration and is not intended to limit the invention in any manner.

The description includes specific examples of preferred embodiments of the present invention. These examples demonstrate how oligonucleotide fragments and nuclear or cellular extracts are used to identify transcription factor binding sites. These examples also demonstrate how synthetic regulatory regions can be created through the modification, combination, and rearrangement of these binding sites or portions thereof and/or of known regulatory regions or binding sites. Furthermore, these examples demonstrate how the synthetic regulatory regions can be evaluated. Such evaluation can identify functional synthetic regulatory regions which direct transcription of a gene at a high level in a particular cell line. These examples include *in vivo* and *in vitro* techniques.

Identification of Transcription Factor Binding Sites

The present invention provides a method for identifying nucleic acid sequences which bind cellular proteins, and which are therefore putative transcriptional regulatory sequences. The method can use any of a variety of mixtures of DNA binding

proteins, in particular including crude transcription factor preparations from nuclear extracts or whole cell extracts of specific cells or tissues. Certain proteins in such mixtures or extracts will bind to and select specific oligonucleotide sequences from a mixture of oligonucleotide sequences. The oligonucleotide sequences can be random sequences, or fragments of DNA from a genomic or cDNA source, or portions, modifications or rearrangements of known binding sites or other selections of nucleic acid sequences.

The protein-bound or selected oligonucleotides are then identified, such as by amplification, cloning and sequencing. The sequences of selected oligonucleotides will reveal consensus sequences which are recognized by the more abundant transcription factors in these cells. Some of the selected sequences will be recognized by common, non-cell-specific transcription factors but a number of selected sequences will be recognized by cell-specific transcription factors.

As a first step of an exemplary selection method for the identification of synthetic regulatory regions, synthetic single-stranded oligonucleotides (oligos) are constructed or obtained which preferably have five important features. These oligos preferably contain the following:

1. a specific sequence of 10-30 nucleotides at the 5' end to act as a primer annealing site for DNA amplification after the selection process has been performed. This sequence will be identical in all oligos and is labeled "P1" in Figure 1.

2. a specific restriction enzyme cleavage site located immediately 3' to or within the 3' end of the 5' primer sequence. This site will be used for the cloning of the selected oligos. This site will be identical in all oligos and is labeled "RI" in Figure 1.

3. a region within the central part of the oligo that contains a number of random nucleotides (preferably ≥10

nucleotides). The sequence in this region will be responsible for the selection of oligos during the selection process.

4. a specific restriction enzyme cleavage site located immediately 3' to the region of random nucleotides. This site will be used with the other restriction site for the cloning of the selected oligos. This site will be identical in all oligos and may be different from the restriction enzyme cleavage site (R1) at the 5' side of the region of random nucleotides and is labeled "R2" in Figure 1.

5. a specific sequence of 10-30 nucleotides at the 3' end of the oligos to act as a primer annealing site for both the synthesis of a second strand complementary to the original oligos prior to selection and DNA amplification after the selection process has been performed. This sequence will be identical in all oligos but different from the sequence at the 5' end of the oligos (P1) and is labeled "P2" in Figure 1.

As outlined in Figure 2, the overall scheme for the selection of binding sites in this embodiment is as follows.

The single-stranded oligo is first converted to a double-stranded oligo by extending primer P2 using a DNA polymerase such as the Klenow fragment of *E. coli* DNA polymerase I, T4 DNA polymerase, or T7 DNA polymerase. The double-stranded oligos are gel-purified and incubated with the crude transcription factor preparation, which preferably would be prepared from isolated nuclei but could also be prepared from whole cell extracts (Dent, et al., In *Transcription Factors: A Practical Approach*, D. S. Latchman (ed.) IRL Press, Oxford, 1-26, (1993)). Transcription factors in the protein extracts will bind to oligos which contain the appropriate recognition sequence or binding site. In preferred embodiments, protein-DNA complexes are separated from unbound oligos by size exclusion chromatography (SEC). SEC is preferable for this

step because the protein-DNA complexes will be heterogeneous in size due to differences in the molecular weights of the bound transcription factors and the possibility that multimeric protein complexes may bind to some binding sites. Thus, electrophoresis would result in a distribution of bands across the gel which would require separate extraction. In contrast, SEC media and conditions can be selected to provide a sharp separation of free and protein-bound oligos.

The selected oligos are then purified, amplified using primers P1 and P2, and digested with restriction enzymes R1 and R2 to excise the central protein-binding regions from the flanking primer sequences. Those skilled in the art can readily determine appropriate primers and restriction enzymes.

The digested oligos are then ligated to form concatamers, and fragments in the 200 - 400 bp range are purified and cloned into an appropriate cloning/sequencing vector. Cloning 200-400 bp concatamers, which contain 20 or more different selected sequences, allows the acquisition of much more sequence information per sequencing reaction than would be obtained if single selected oligos were cloned and sequenced. The method, however, can also utilize single oligos or other size concatamers.

The sequences of individual selected oligos are aligned to identify consensus sequences for the most abundant transcription factors. These sequences are tested for cell specificity, either individually or in combination, by cloning them upstream of a basal heterologous regulatory region driving a reporter gene. The selected oligos can also be used in combination with known transcription factor response elements to make synthetic regulatory regions.

Since this method does not require knowledge of the genes that are expressed or the transcription factors that are present in the cells of interest, this method can be used to

identify transcriptional regulatory sequences which are utilized in cell types or under conditions in which gene regulation is poorly understood. The process can be used to identify and characterize regulatory regions that are highly active in a specific cell type or tissue, as well as cell-specific regulatory regions. This can be extended to include different developmental stages, induction states, or transformation states of cells.

Evaluation Method for Synthetic Regulatory Regions

Because of the limitations in previous methods, as discussed above, new methods are needed to evaluate the functions of synthetic regulatory regions. This invention provides an approach utilizing the expression of proteins capable of protecting cells from stress conditions, such as drugs, to select functional synthetic regulatory regions. In addition to evaluating synthetic regulatory regions, this method can be used to evaluate any of a variety of other transcriptional regulatory sequences.

A number of different proteins are capable of protecting eukaryotic cells from the toxic effects of specific biochemical agents (drugs). The genes coding for some of these proteins (protective genes) have been used to select for the amplification of other non-selectable genes that are linked to the protective gene. This amplification occurs after integration of the two linked genes into the same site of the genome of transfected cells. These selection systems have been used to amplify exogenous genes to increase the production of recombinant proteins (Kaufman, *Meth. Enzymol.* 185:537-566 (1990); Kellems, *Current Opinion in Biotechnology* 2:723-729 (1991); Kellems, *Methods in Molecular Genetics* 5:143-155 (1994)).

The gene most frequently used in gene amplification schemes is the gene coding for dihydrofolate reductase (DHFR), which provides protection against the toxic effects of the drug methotrexate. After transfection of methotrexate sensitive cells with an expression plasmid containing both the DHFR gene and the gene of interest, these genes can be induced to coamplify by treating the cells with increasing concentrations of methotrexate (Kaufman, *Meth. Enzymol.* 185:537-566 (1990)).

The gene for adenosine deaminase (ADA) can also be used to select for the amplification of linked genes (Kellems et al., in *Genetics and Molecular Biology of Industrial Microorganisms*, Hershberger et al., (ed.) American Society for Microbiology, Washington, 215-225 (1989); Kellems, *Current Opinion in Biotechnology* 2:723-729 (1991); Kellems, in *Gene Amplification in Mammalian Cells*, Marcel Dekker, Inc., New York, 207-221 (1992); Kellems, *Methods in Molecular Genetics* 5:143-155 (1994)). ADA is an enzyme involved in purine metabolism in mammalian cells and can provide protection against the toxic effects of the drug such as xylofuranosyl-adenine (xyl-A).

Applicant has found that the ADA gene can be used in a method for evaluating the transcriptional activity of transcriptional regulatory regions. In this method, a high level of ADA gene expression is required to allow growth of a cell. Such high level expression will only be provided if a test sequence inserted in a transcriptional regulatory position, e.g., upstream to the ADA gene, is effective in allowing sufficient transcription of the ADA gene.

In this system, synthetic regulatory regions/enhancers will be assembled from mixtures of synthetic oligonucleotides, fragments of cloned natural regulatory regions, and/or protein binding sites using a random combinatorial approach. The

synthetic regulatory regions will be inserted upstream of a basal TATA box and functional ADA minigene (cDNA) contained in a plasmid. This will produce libraries of synthetic or recombined regulatory regions which can contain millions of different combinations. These plasmid libraries will then be transfected into cells of different origins and the transfected cells will be selected for increased ADA activity in transient assays. Cells that express no or low levels of ADA will be killed and lost from the culture due to insufficient ADA activity. Cells that express high levels of ADA, due to the strength of the synthetic regulatory region, will survive. This procedure thus selects for synthetic regulatory regions that drive the expression of ADA in that specific cell type. This approach can be used to develop strong regulatory region that will function in cells or tissues for which there is poor understanding of patterns of gene expression or the regulatory regions of specific genes have not been characterized.

This approach is not limited to the use of ADA-based selection protocols but can also utilize selection strategies developed based on the expression of other genes, including but not limited to dihydrofolate reductase (DHFR), metallothienin, CAD, thymidylate synthetase, ornithine decarboxylase, etc. (see Kellems, *Current Opinion in Biotechnology* 2:723-729 (1991) for a more extensive list).

Examples of how this type of selection system could be used are outlined below:

Creation of Synthetic Regulatory Regions from Transcription Factor Binding Sites

As discussed previously, the synthetic regulatory regions are created to have altered composition, order, and/or spacing of individual binding sites for transcription factors.

Creation of the synthetic regulatory regions usually uses a combination of specific restriction sites. If convenient sites are not available, alternatives can be used, such as chemical resynthesis or engineering of different restriction sites onto the ends of the binding sites. A variety of methods can be used to assemble the different components, such as the method of nucleic acid ordered assembly with directionality (NOMAD) (Rebatchouk, et al., *PNAS* 93:10891-10896 (1996)).

NOMAD is a general cloning strategy (WWW resource locator http://Lmb1.bios.uic. edu/NOMAD/NOMAD.html). NOMAD can manipulate the binding sites in the form of "module" having a standardized cohesive structure. Specially designed "assembly vectors" allow for sequential and directional insertion of any number of binding sites in an arbitrary predetermined order, using the ability of type IIS restriction enzymes to cut DNA outside of their recognition sequences (Rebatchouk, et al., *PNAS* 93:10891-10896 (1996)). NOMAD ensures the convenient construction of the synthetic regulatory regions with altered composition, order, or spacing of individual binding sites for transcription factors. The acquired synthetic regulatory regions can then be evaluated, such as with the ADA selection method.

Biochemical Agents Used in ADA Selection

A number of protocols have been developed that use ADA selection to amplify genes (Kellems et al., in *Genetics and Molecular Biology of Industrial Microorganisms*, Hershberger et al., (ed.) American Society for Microbiology, Washington, 215-225 (1989); Kellems, *Current Opinion in Biotechnology* 2:723-729 (1991); Kellems, in *Gene Amplification in Mammalian Cells*, Marcel Dekker, Inc., New York, 207-221 (1992); Kellems, *Methods in Molecular Genetics* 5:143-155 (1994)).

In this invention, a method has been developed which uses ADA to identify and evaluate regulatory regions, such as synthetic regulatory regions, or other regulatory sequences. This method can be performed in a number of different ways, including the following.

The simplest method uses increasing concentrations of xylofuranosyl-adenine (xyl-A) alone. In cells expressing low levels of ADA, xyl-A is converted to xyl-AMP by adenosine kinase. Xyl-AMP is subsequently converted to xyl-ATP which can then be incorporated into RNA by RNA polymerase where it acts to block further extension of the RNA chain. This chain termination is due to the fact that, unlike the normal sugar contained in ribonucleosides, xylose lacks a 3' hydroxyl group which is required for RNA chain extension. ADA is capable of detoxifying xyl-A by converting it to hypoxanthine and xylose-P_i, both of which are non-toxic. Since the chain terminating effect of xyl-A is independent of DNA synthesis, xyl-A will readily kill non-dividing as well as dividing cells (Kellems et al., in *Genetics and Molecular Biology of Industrial Microorganisms*, Hershberger et al., (ed.) American Society for Microbiology, Washington, 215-225 (1989)). The concentration of xyl-A required to kill a specific type of cell depends on the level of endogenous ADA expressed by those cells. Most cells normally produce relatively low levels of ADA and are, therefore, killed quickly by low (micromolar) concentrations of xyl-A. Endogenous ADA can be selectively inhibited by incubation with deoxycoformacin. This protocol has the limitation that ADA expression increases with increasing concentrations of xyl-A up to only about 10 μM. Cells can be selected that are resistant to higher concentrations of xyl-A but they do not express higher levels of ADA. It was found that cells selected for resistance to more than about 10 μM

xyl-A were deficient in the activity of adenosine kinase, which is responsible for converting xyl-A to xyl-AMP, the first step in producing xyl-ATP which is a substrate for RNA polymerase.

An alternative method of ADA selection, termed 11AAU selection (Yeung et al., *J. Biol. Chem.* 258:8338-8345 (1983); Yeung et al., *J. Biol. Chem.* 258:8330-8337 (1983)), was subsequently developed that used a combination of 1) alanosine, which inhibits the *de novo* synthesis of AMP; 2) adenosine, which then becomes a required substrate for adenosine kinase via the salvage biosynthetic pathway; and 3) uridine, which overcomes the inhibitory effect of high concentrations of adenosine on UNT synthesis. This selection protocol requires adenosine kinase to produce AMP and thus greatly reduces the chance that this enzyme will be affected during the selection process. In this protocol adenosine is used at a concentration that is cytotoxic to normal cells. Thus, this protocol selects for increased expression of ADA which is required to detoxify the excess adenosine. ADA activity can be further increased by exposing cells to both 11AAU selection and increasing concentrations of deoxycoformacin (Yeung et al., *J. Biol. Chem.* 258:8330-8337 (1983)). However, some cells do not tolerate the 11AAU/deoxycoformacin selection system well.

Yet another selection system uses xyl-A as the cytotoxic agent in combination with deoxycorformacin to inhibit endogenous ADA activity (Kaufman et al., *PNAS* 83:3136-3140 (1986); Kellems et al., in *Genetics and Molecular Biology of Industrial Microorganisms*, Hershberger et al., (ed.) American Society for Microbiology, Washington, 215-225 (1989)). This is a very effective method to select for increased ADA levels but does not provide any selection for the maintenance of adenosine kinase activity. Therefore, this method should not be used for long periods of time as this increases the

probability that adenosine kinase mutants will arise.

These selection methods can be used in the selection and evaluation of synthetic regulatory regions, as discussed previously. An exogenous ADA gene under the control of one of the synthetic regulatory regions to be evaluated is transfected into cells that are then placed under selective pressure. The surviving cells should carry the functional synthetic regulatory regions which direct the strong transcription of ADA gene, protecting the cells from the toxic effect of the biochemical agents.

As indicated, a variety of different selection methods can be used to identify effective synthetic regulatory regions. Generally a selection method based on expression of a protective gene can be used, where the selection method is able to distinguish between low or moderate expression levels and high expression levels. This allows a semi-quantitative comparison of the relative effects of different synthetic and natural promoters or other regulatory regions.

Positive selection systems can also be used, such as magnetic sorting and FACS. An example of such systems is the MAC Selecting System (Miltenvi Biotec, Auburn, CA). In this system, a gene encoding CD4 antigen is the selective gene and CD4 antibody complexed to magnetic beads is used to separate cells expressing CD4 antigen from non-expressing cells. Alternatively, florescence labeled CD4 antibody can be used to detect CD4 expressing cells, and expressing cells can then be separated by FACS.

Synthetic Regulatory Regions for Muscle Cells

The development of synthetic regulatory regions with high level activity in a particular cell type or state can be illustrated by the identification of regions producing high level expression in muscle cells. Individual synthetic

oligonucleotides can be synthesized containing known consensus sequences capable of binding cell-specific transcription factors (transcription factor binding sites), ligated together in random combinations and cloned upstream of the ADA gene as described above. For example, consensus sequences for muscle-specific binding sites, including serum binding sites (SREs), MEF-1 sites, MEF-2 sites, and/or TEF-1 sites, can be used. This library of synthetic regulatory regions can then be transfected into muscle cells (e.g., C_2C_{12}, SOL8, or primary myoblast cells). The ADA selection system allows the selection against clones containing weak muscle regulatory regions and for clones containing strong muscle regulatory regions.

Also, cloned or PCR-amplified cell-specific regulatory elements can be digested with one or more frequent cutting restriction enzymes to produce mixtures of small DNA fragments containing sequences capable of binding cell-specific transcription factors. These fragments would be ligated together in random combinations and cloned upstream of the ADA gene as described above. For example, regulatory regions for the skeletal α-actin, cardiac α-actin, myosin heavy chain, and myosin light chain genes, which contain the muscle-specific binding sites, can be used.

This library of synthetic regulatory regions would then be transfected into muscle cells (e.g., C_2C_{12}, SOL8, or primary myoblast cells). The ADA selection system would allow the selection against clones containing weak muscle regulatory regions and for clones containing strong muscle regulatory regions.

Identification of 3', 5', and Intron Regions that Enhance Gene Expression

Alone, or in combination with the promoter selection

methodology described herein, one may use the combinatorial approach combined with a selection methodology to identify gene control regions, including novel regions, such as 3' untranslated regions (3'UTR), 5' untranslated regions (5'UTR), and intron elements that have the effect of enhancing gene expression when inserted into a plasmid construct in the proper orientation to the gene. One skilled in the art will immediately recognize the proper position of the element to be inserted from the terms 3'UTR, 5'UTR, and intron. 3'UTR, 5'UTR, or intron regions from known gene are randomly combined, for example, by the method described herein in connection with promoter/enhancer sequences, and inserted into the appropriate position relative to the coding sequence of the gene of interest. As indicated above, other sequences can also be used, including but not limited to random sequences and combinatorial rearrangements of known sequences. A selection procedure, such as that described above, is then employed to identify control regions which have the effect of enhancing the expression of the gene with which they are associated.

Selection of Transcriptional Regulatory Regions from Various Tissue Types

While the methods described herein are exemplified by selection of muscle-specific promoter sequences, the use of these methods is by no means restricted to muscle cells. For example, cells of lung, kidney, brain, heart, eye, inner ear, epithelial, endothelial, mesothelial, smooth muscle, neuronal, lymphocyte, macrophage, glial, microglial, intestinal, colon, bone, hematopoietic, skin, liver, cancerous, precancerous, metastatic, fetal, or vascular origin may be used to identify expression enhancing regulatory regions. In addition, regulatory elements derived from one cell type may be selected for in a different cell type for expression enhancing capacity.

Such a procedure would also fall within the scope of this invention.

Identification of Reduced-Size Active Portion of Synthetic Regulatory Region

Using methods described above, one can identify synthetic regulatory regions which provide appropriate expression levels in a selected type or group of cells. Depending on the oligonucleotide length utilized in the identification, it can be useful to reduce the size of the synthetic regulatory region by identifying and utilizing a portion or portions of the larger region which provide the enhanced transcriptional regulatory effects. Such identification can be performed by routine methods, such as by replacement of portions of an effective regulatory region with equal length inactive sequences and determining the activity of the resulting modified region. If the expression enhancing activity is significantly reduced, this indicates that the modified region includes at least part of a sequence which provides the expression enhancing activity. On the other hand, if the modification does not significantly affect the resulting expression, this indicates that the modified portion does not contribute to the activity of the synthetic regulatory region.

Thus, the portion or portions which significantly contribute to the transcriptional regulatory activity can be used as new smaller synthetic regulatory regions separately from other parts of the original synthetic regulatory region. Generally the position of the active portion or portions with respect to the coding sequence should be maintained at approximately the position it occupied in the original synthetic region. However, it will not usually be necessary to maintain exactly the same position, but will preferably be within 100, 60, 30, or fewer bases of the original position.

While the active portions can be of various sizes, preferably the portion providing a small synthetic transcriptional regulatory region includes at least 20 contiguous nucleotides, and more preferably includes at least 40, 60, 80, or 100 contiguous nucleotides of the original synthetic region.

The present invention is further illustrated by the following examples, which are not intended to limit the present invention in any way.

Example 1. Generating the libraries of synthetic muscle specific regulatory regions by random combination of regulatory elements.

Available naturally-occurring muscle specific regulatory regions cannot regulate transcription in all desired manners in muscle cells. Synthetic muscle specific regulatory regions are therefore needed to provide new candidates for controlling the transcription. The synthetic muscle specific regulatory regions can be constructed by random combination of transcription factor binding sites which are known to be important in the regulation of general transcription or muscle cell-specific transcription. This example illustrates how synthetic muscle-specific regulatory regions can be constructed using a selection of known binding sites.

The sequences which are shown in the following include MRE (muscle response element), E-box which is the binding site recognized by the family of basic-helix-loop-helix (bHLH) transcription factors, and the binding sites for transcription factors MEF-2, TEF-1 and Sp1.

MEF-2 CTCTAAAAATAACCCT

MRE	GCCCAACACCCAAATATGGCTT
E-box	CTCACCTGCTG
TEF-1	GCCGCATTCCTGGG
Sp1	CCCCGCCC

The first step in constructing the synthetic regulatory regions is to synthesize double-stranded oligonucleotides containing one of the above binding sites. This synthesis is performed for each of the binding sites to be included. The oligonucleotides should be sticky ended, i.e., have ends which are single-stranded with sequences complementary to each other. The oligonucleotides preferably fit in one or two helical turns so that elements reside on the same face after being linked together. This can be achieved by constructing a sequence so that the contact points contained in the elements are approximately 10 base pairs apart from each other (or approximately 20 base pairs apart). Those skilled in the art will know appropriate techniques to provide appropriate spacing and sticky ends.

These oligonucleotides are mixed together using a particular ratio of different oligonucleotides. This ratio can be varied to favor the presence of a particular element. For example, MEF-2, E-box, MRE, TEF-1, and Sp1 can be mixed at a ratio of 4:2:2:2:1, in order to increase the probability of MEF-2 presence in the synthetic regulatory regions. Similarly, the ratios can be biased in favor of other binding sites. The mixed oligonucleotides can automatically be linked together non-covalently through annealing of the sticky ends. The oligonucleotides are then ligated using a DNA ligase. The oligonucleotides are therefore covalently linked together to form new and longer oligonucleotides.

The ligated oligonucleotides are cut through partial digestion with a nuclease. The digested oligonucleotides are

separated by gel electrophoresis and the oligonucleotides with a particular size, e.g., 200 bp, are recovered from the gel. The recovered oligonucleotides are then capped with a sticky ended adaptor using a DNA ligase.

The capped oligonucleotides are then cloned into appropriate vectors for expression analysis. For example, for identification of effective myogenic promoter/enhancer sequences, the capped oligonucleotides can be inserted at a site adjacent to the Sk-actin TATA -box in a myogenic vector system (MVS) β-gal construct or at -200 in MVS β-gal construct.

Example 2. Comparison of relative regulatory region activity during differentiation at primary myoblast cells

The synthetic regulatory regions should be evaluated to confirm they are functional in the regulation of transcription. Large-scale evaluation can be done with the stress condition selection (e.g., ADA), as discussed above; medium-scale evaluation can be done either with the stress condition selection, or with the following approach or with other analyses of expression level. This example also illustrates the selection of synthetic regulatory regions which regulate transcription rates in particular cells, in this example, muscle cells.

In this approach, the synthetic regulatory regions are inserted into a vector to regulate the transcription of a reporter gene, instead of a selective gene. The reporter genes include, but are not limited to, the genes encoding β-gal and luciferase. Minilysate prepared DNA, such as the constructs of example 1, is transferred into myogenic cultures in 96 well microtiter dishes. β-gal activity is assayed by routine methods, e.g., mini ONPG assay, and compared to β-gal

expression driven by the cytomegalovirus immediate early promoter (CMV-β-gal). High β-gal activities represent the strong synthetic regulatory regions. Of course, other non-cell-specific regulatory regions could also be used for a reference expression level.

The above approach can also be used for the further evaluation of synthetic regulatory regions acquired using the stress condition approach, as the β-gal activity assay can provide quantitative information about the regulatory regions being evaluated.

Fig. 3 shows the comparison of relative regulatory region activity during differentiation of primary myoblast cells. This experiment was done using reporter gene product assay. The regulatory region containing 2 x MEF-2 has about a five-fold higher activity than other regulatory regions tested. This result indicates that the regulatory region containing 2 x MEF is capable of stimulating gene transcription at a high level in myoblast cells.

Example 3. Differential SRF activity on c-Fos SRE vs muscle SRE

The above approach (Example 2) using a reporter gene product assay was used to determine the differential SRF activity on c-Fos SRE and muscle SRE, the sequences of which are shown in the following. These sequences have sequence similarity in the SRF binding sites, which are underlined.

C-FOS SRE: ACAGGATGTCCATATTAGGACATCTGCG
MUSCLE SRE: GCCCGACACCCAAATATGGCGACGGCCG

The c-Fos SRE and muscle SRE were inserted into a vector to regulate a reporter gene encoding a luciferase. The vector constructs were transferred into C_2C_{12} myoblasts. The

luciferase gene is transcribed in the presence of various SRF□s. The luciferase activity was then assayed. All the transcription factors tested except GCN1 showed similar activities on c-Fos SRE and muscle SRE. On c-Fos SRE, GCN1 has about 3-fold higher activity than SRFwt does. On muscle SRE, in contrast, GCN1 has about 2-fold lower activity than SRFwt does (Fig. 4). These results indicate that minor variations in transcription binding sites can result in a major difference in regulatory region activity in the presence of a particular transcription factor.

Example 4. Selection of Tissue- or Cell- Specific Elements *in vivo*

In addition to *in vitro* selection approaches, synthetic tissue- or cell- specific transcriptional regulatory regions can be selected and evaluated *in vivo*. One of the most important uses of the synthetic elements is to regulate tissue- or cell- specific gene expression in an organism. The synthetic elements identified *in vitro* may be further studied *in vivo* to better evaluate or understand their functions. Useful *in vivo* approaches include, but are not limited to, transgenic animals and muscle injection.

A. Insertion of Vectors into Transgenic Mice

Vectors are constructed containing a reporter gene, e.g. β-gal, under the control of the synthetic elements identified as having *in vitro* activity in a particular type or types of cells, e.g., in muscle cells. Transgenic mice carrying the vectors can be generated by standard oocyte injection (Brinster, et al, *Proc. Natl. Acad. Sci. USA* 82:4438-4442 (1958)) and bred to demonstrate stable transmission of transgenes to subsequent generations. Transgenics can be identified by polymerase chain reaction or Southern genomic DNA blotting

analysis, such as from tail cut DNA.

Transgenics can be tested for tissue specific expression, e.g., muscle specific expression, of the transferred vector by RNA blotting of total RNA isolated from several tissues, or by β-gal assay. For example, samples can be taken and analyzed from skeletal muscle, gonad, lymph nodes, liver, spleen, kidney, lungs, heart, brain, bone marrow, blood, and other tissues. The analysis and comparison of expression levels, such as by the determination of β-gal activity in the different tissues, will reveal the regulatory pattern of the synthetic regulatory regions in the organism. Expression in one tissue at a significantly higher level than in other tissues indicates that the regulatory regions on the plasmid a specific for that tissue.

Such *in vivo* analysis of tissue specific expression is applicable to the evaluation of regulatory regions in any position with respect to the coding sequences, such as in the 5' UTR, the 3' UTR, and in introns.

B. Somatic Gene Transfer to Skeletal Muscle *In Vivo*

To demonstrate the effects of the synthetic elements as used in *in vivo* gene therapy and/or to identify elements having muscle specific activity, vectors can be injected into adult muscle (e.g., avian or mammalian) for the expression of a reporter gene such as the gene encoding β-gal or luciferase.

Vectors carrying β-gal under the control of the synthetic elements, or under the control of known regulatory regions (used as controls), are pelleted by centrifugation, dried under vacuum, resuspended in an appropriate formulation, and injected into the quadriceps muscle (20 μg/pellet - 3 pellets/muscle) of 2 sets of 6 mice (injection into other muscles can also be used). The animal is sacrificed 48 hours following introduction of the DNA and the entire muscle (the muscle

injected) from each animal that received an inoculation is removed and assayed for β-gal activity in the tissue. If sufficient experimental animals are available, it is preferable to assay for expression at a number of different time points, such as 24 hrs, 48 hrs, 7 days, 14 days, and 28 days following DNA introduction. In this way additional information is provided on the time course of expression of the reporter gene.

As described above, expression of the reporter gene is determined by assay for activity of the product of that gene, e.g., β-gal activity, however, other methods can also be used, including reverse transcriptase PCR analysis.

Muscle specific expression is demonstrated by showing that expression occurs only or at a significantly higher level in muscle than in other tissues. Therefore, the evaluation preferably also includes assaying for expression of the reporter gene in tissues other than skeletal muscle. It is expected that some amount of the injected vector will migrate to other tissues. Thus, at each of the time points for which muscle samples are taken, samples can also be taken from a set of other tissues, such as gonad, lymph nodes, liver, spleen, kidney, lungs, heart, brain, bone marrow, and blood. Each of the samples is assayed for reporter gene expression.

The pattern of reporter gene expression can also be correlated with the presence of the vector. The presence of the vector in a tissue can be determined by amplification and hybridization of a vector-specific sequence.

Example 5. The Development of Synthetic Regulatory Regions

The above examples describe approaches to constructing, screening, and evaluating synthetic regulatory regions. The combination of these approaches can identify regulatory regions with advantageous properties for particular applications. The

following example demonstrates that synthetic regulatory regions constructed using binding sequences in a combinatorial approach can be identified which provide advantageous expression characteristics in a particular tissue and state of that tissue.

To aid in understanding the results of this example, a short background discussion may be of assistance. IGF-1 plays a role as a neurotrophic agent in repairing crushed motor neurons. Localized expression of IGF-I hastens the repair of crushed motor neurons. Although it is one of the strongest muscle specific promoters, skeletal α-actin promoter is not an ideal regulatory region for this expression as intact innervation of muscle is required to maintain skeletal α-actin promoter activity at a high level. In transgenic mice having α-actin/hIGF-1 transgene and showing high level expression of hIGF-1, following sciatic nerve crush the expression level of hIGF-1 was down regulated. hIGF-1 expression was at a minimum about 2 weeks post crush (matching the time of greatest muscle atrophy), and only began to return to normal levels at about 3 weeks post crush.

Thus, nerve crush effectively represses skeletal α-actin promoter, which only recovers with reinnervation. This is in accord with observations that injected α-actin/IGF-1 plasmids take at least three weeks to show effectiveness. Earlier expression of IGF-1 would therefore be desirable in order to maintain high level expression of neurotrophic genes during the early stages of nerve and muscle regeneration.

It is, therefore, beneficial to develop synthetic myogenic regulatory regions to drive IGF-I expression which are insensitive to the innervation state of muscle. Thus, having a myogenic regulatory region that is turned on all the time in muscle should even further speed the nerve repair process. In

order to develop such a regulatory region, we took the following steps.

A. Construction of libraries of synthetic regulatory regions

We first constructed a series of synthetic regulatory regions based on the sequences of transcriptional control elements involved in the activation and regulation of genes in mammalian cells.

The portion of the skeletal α-actin promoter upstream of the ATAAAA box was removed from plasmid p612aACATMLC (which contains a pBluescript polylinker upstream of a skeletal α-actin promoter) by digestion with EagI, which cuts in the pBluescript polylinker upstream of the promoter and 47 bp upstream of the ATAAAA box. The luciferase gene was linked downstream of the resulting minimal α-actin promoter. The synthetic regulatory regions were randomly cloned into this minimal α-actin/luciferase test plasmid.

The control elements that were tested include:

SRE	5'-GACACCCAAATATGGCGACCG-3' 3'-CTGTGGGTTTATACCGCTGGC-5'
MEF-1	5'-CCAACACCTGCTGCCTGCC-3' 3'-GGTTGTGGACGACGGACGG-5'
MEF-2	5'-CGCTCTAAAAATAACTCCC-3' 3'-GCGAGATTTTTATTGAGGG-5'
TEF-1	5'-CACCATTCCTCAC-3' 3'-GTGGTAAGGAGTG-5'
SP1	5'-CCGTCCGCCCTCGG-3' 3'-GGCAGGCGGGAGCC-5'

The SRE sequence corresponds to the proximal skeletal α-actin SRE sequence. The SRE core sequence is overlined. The MEF-1 sequence (complemented and overlined) and the adjacent

GCTGC motif (asterisked) are conserved in the muscle creating kinase gene and rat myosin light chain gene (Lasser et al., 1989). The SP1 sequence (overlined) has an Eag1 half restriction site at each end. Sp1 sites were included as spacers between the other control elements.

Oligonucleotide pairs (dsDNA) were annealed and then ligated together in various combinations to form larger fragments of randomly oriented control elements. Since each of the Sp1 elements contains EagI half-sites at each end, an intact EagI restriction site will be generated wherever two Sp1 elements are ligated together. DNA fragments contain from 8 to 14 control elements in random combinations with EagI cohesive ends, and thus represent synthetic regulatory regions.

Fragments formed from each of the combinations of elements resulted in a separate pool of fragments. Each of the combinations contains a heterogenous set of fragments resulting from the particular starting combination of oligonucleotides, as the oligonucleotides can anneal together in various orders and numbers.

DNA fragments from each pool of synthetic regulatory regions was ligated into the EagI site of the minimal α-actin/luciferase plasmid. Approximately twenty clones were picked for each combination, which were then grown, purified with Qiagen kits and used to transfect primary myoblasts.

The clones were named Cm-n, where m is the number of a particular combination and n is the number of a particular clone picked from that combination. For example, C5-1 represents clone number 1 of combination number 5. Figure 5 shows the arrangement of sub-elements of some exemplary synthetic regulatory regions. The sequences of portions of the plasmids containing exemplary synthetic regulatory regions, including the sequence of the synthetic regulatory region, are

shown in Figs. 8-31. The sequences are believed to be correct, however a small percentage of sequence errors may be present.

One skilled in the art could readily obtain the correct synthetic regulatory region by identifying the particular elements and their positions in the region from the sequence provided, and constructing the synthetic regulatory regions from those elements in the same positions and orientations.

A p448 Sk α-actin promoter/luciferase vector was used as a control. This promoter is a standard representative of strong muscle specific promoters, being one of the strongest such promoters currently available. Expression from this vector was used as a standard for comparison of the expression levels regulated by the test synthetic regulatory regions.

B. Screening of Library of Synthetic Regulatory Regions *in vitro*

Plasmids of the synthetic regulatory region library described in A. were transfected into muscle cells with lipofectamine transfections in two series. The transfected cells from these transfection series were grown and collected for luciferase activity assay.

We observed from the first series of lipofectamine transfections done in duplicate in primary myoblast cultures, that none of the eight constructions grown for each of the multimerized SREs, E-boxes, MEF-2, and TEF-1 regulatory regions (32 separate plasmids) had activity greater or equal to the activity of the skeletal α-actin promoter/enhancer driven luciferase plasmid (p448).

In the second series, six different combinations of synthetic regulatory regions were then tested in mature myotubes. Luciferase activities up to 5-fold greater than that driven by the skeletal α-actin promoter/enhancer were detected by transfections in a subset of clones, namely C1-28 (Fig. 8),

C2-27 (Fig. 9), C5-12 (Fig. 10), C6-16 (Fig. 11) and C6'-7 (Fig. 12). In muscle cells, therefore, these synthetic regulatory regions stimulate higher transcription levels than skeletal α-actin promoter.

Moreover, we used a simple assay to check the effect of myoblast depolarization as a way to evaluate the potential for innervation effects on muscle gene expression. We found that the skeletal α-actin promoter is up-regulated 3-4 fold by applying KCl for 20 minutes to the media of myotube cultures. Clones C1-28 (Fig. 8), C5-1 (Fig. 13), C5-5 (Fig. 14), C6-5 (Fig. 15), C5-12 (Fig. 10), C6-16 (Fig. 11), and C6'-7 (Fig. 12) provided high levels of rather stable expression in depolarized myotubes. Thus, these synthetic regulatory regions may be much less affected by innervation effects than skeletal α-actin promoter and are ready for further evaluation. Results of the reporter expression levels and of the expression levels in the KCl depolarization test are shown in Fig. 6.

Method

A. First Transfection

1 μg synthetic regulatory region/luciferase plasmid was transfected into 24 hr primary chick myoblast in 60 mm plates (500,000 cells/plate). 200 ng CMV β-gal plasmid was cotransfected in each transfection.

40 hours after transfection, KCl was added directly to the medium to a concentration of 50 μM and cells were treated at 37°C for 2 hours. The medium containing KCl was aspirated, the cells rinsed once with HBSS, and fresh medium was added. The control plates without KCl treatment were left untouched in the original medium.

20 hours after KCL treatment, cells were collected and luciferase activity was assayed.

B. Second Transfection

100 ng synthetic regulatory region-luciferase plasmid, along with 200 ng CMV β-gal plasmid was transfected to 24-hour primary chick myoblast in 60 mm plates (500,000 cells/plate). 700 ng YEAST MARKER carrier DNA was added to each transfection to make the total amount of DNA transfected 1 μg.

36 hours after transfection, cells were rinsed once with HBSS, MEM (no serum) containing 50 μM KCl (for control) was added, and the cells were incubated at 37°C for 40 minutes. Then the above medium was aspirated, the cells rinsed once with HBSS, and full medium was added.

24 hours after KCl treatment, cells were collected, and luciferase activity was assayed.

C. Evaluation of Synthetic Regulatory Regions in Nerve Crush Model

To demonstrate the evaluation and identification of synthetic regulatory regions effective in a specific *in vivo* environment, we tested some of the constructs from above which were shown to provide high level myogenic expression and for which the *in vitro* test suggested less sensitivity to innervation effects than the Sk α-actin promoter/enhancer. Results for two of the constructs in a nerve crush model are described. Experiments were designed to test synthetic regulatory regions that are resistant to nerve-injury induced down-regulation of expression driven by skeletal actin promoter.

Tibiales muscles of ICR mice were injected with 100 μg of clone skeletal α-actin promoter 448 (control), synthetic regulatory region luciferase vectors C1-28 (Fig. 6), and C5-12 (Fig. 8), which had been shown to be less affected by myoblast depolarization effect than the control (see Section B.). Two weeks post sciatic nerve crush, the injected muscle was harvested and assayed for luciferase activity. The expression

levels from C1-28 and C5-12 were approximately 7-fold and 15-fold greater respectively than from the skeletal α-actin promoter (Fig. 7).

These results demonstrate that the two new regulatory regions were more resistant to injury induced regulation. A benefit of these regulatory regions will be to sustain high expression levels of neurotrophic genes during the initial stages of nerve and muscle regeneration, when skeletal α-actin promoter is down-regulated. The higher expression levels provided by synthetic regulatory regions such as these may allow the use of significantly lower amounts of DNA, e.g., 1/10 the amount of DNA, to achieve the same biological effects as that provided by expression driven by promoters such as the skeletal α-actin promoter.

One skilled in the art would readily appreciate that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those inherent therein. The molecular complexes and the methods, procedures, treatments, molecules, specific compounds described herein are presently representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the invention. It will be readily apparent to one skilled in the art that varying substitutions and modifications may be made to the invention disclosed herein without departing from the scope and spirit of the invention.

All patents and publications mentioned in the specification are indicative of the levels of those skilled in the art to which the invention pertains. All patents and publications are herein incorporated by reference to the same extent as if each individual publication was specifically and individually indicated to be incorporated

by reference.

The invention illustratively described herein suitably may be practiced in the absence of any element or elements, limitation or limitations which is not specifically disclosed herein. Thus, for example, in each instance herein any of the terms "comprising," "consisting essentially of" and "consisting of" may be replaced with either of the other two terms. The terms and expressions which have been employed are used as terms of description and not of limitation, and there is no intention that in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification and variation of the concepts herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention as defined by the appended claims.

In addition, where features or aspects of the invention are described in terms of Markush groups, those skilled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group. For example, if X is described as selected from the group consisting of bromine, chlorine, and iodine, claims for X being bromine and claims for X being bromine and chlorine are fully described.

In view of the degeneracy of the genetic code, other combinations of nucleic acids also encode the claimed peptides and proteins of the invention. For example, all four nucleic acid sequences GCT, GCC, GCA, and GCG encode

the amino acide alanine. Therefore, if for an amino acid there exists an average of three codons, a polypeptide of 100 amino acids in length will, on average, be encoded by 3^{100}, or 5×10^{47}, nucleic acid sequences. Thus, a nucleic acid sequence can be modified to form a second nucleic acid sequence, encoding the same polypeptide as endoded by the first nucleic acid sequences, using routine procedures and without undue experimentation. Thus, all possible nucleic acids that encode the claimed peptides and proteins are also fully described herein, as if all were written out in full taking into account the codon usage, especially that preferred in humans. Furthermore, changes in the amino acid sequences of polypeptides, or in the corresponding nucleic acid sequence encoding such polypeptide, may be designed or selected to take place in an area of the sequence where the significant activity of the polypeptide remains unchanged. For example, an amino acid change may take place within a b-turn, away from the active site of the polypeptide. Also changes such as deletions (*e.g.*, removal of a segment of the polypeptide, or in the corresponding nucleic acid sequence encoding such polypeptide, which does not affect the active site) and additions (*e.g.*, addition of more amino acids to the polypeptide sequence without affecting the function of the active site, such as the formation of GST-fusion proteins, or additions in the corresponding nucleic acid sequence encoding such polypeptide without affecting the function of the active site) are also within the scope of the present invention. Such changes to the polypeptides can be performed by those with ordinary skill in the art using routine procedures and without undue experimentation. Thus, all possible nucleic and/or amino acid sequences that can readily be determined not to affect a significant activity

of the peptide or protein of the invention are also fully described herein.

Other embodiments are within the following claims.

Claims

We claim:

1. A method of identifying binding sites for transcription factors, comprising the step of:

identifying the oligonucleotides in oligonucleotide-protein complexes formed between one or more proteins of a cellular or nuclear extract and any of a plurality of double-stranded oligonucleotide fragments in a mixture of said fragments and said extract wherein said complexes are separated from free oligonucleotides in said mixture using size exclusion chromatography; and

wherein the presence of a said oligonucleotide in a said complex is indicative that said oligonucleotide comprises a said binding site.

2. The method of claim 1, wherein a said double-stranded oligonucleotide fragment is made by synthesizing a single-stranded oligonucleotide and converting said single-stranded oligonucleotide to a double-stranded oligonucleotide.

3. The method of claim 1, wherein said oligonucleotide fragment comprises a central random sequence and both restriction sites and primer sequences in the 5' and 3' ends.

4. The method of claim 1, wherein said identifying comprises amplifying and sequencing said oligonucleotides from said protein-oligonucleotide complexes.

5. The method of claim 4, wherein said amplifying is performed by polymerase chain reaction.

6. A method for evaluating a putative cell- or tissue-

specific transcriptional regulatory region, comprising the step of:

determining whether a cell comprising said putative transcriptional regulatory region in a transcriptional regulatory position to a selective gene is selected under selective conditions, wherein said selection of said cell will only occur if said selective gene is expressed at a sufficiently high level in said cell and wherein said selective gene will only be expressed at said sufficiently high level if said putative transcriptional regulatory region is active in said cell.

7. The method of claim 6, wherein said selection condition is a positive selection condition, and wherein said method comprises the steps of:

culturing one or more cells under said positive selection condition, wherein said at least one said cell each contain a different nucleic acid test sequence inserted in a transcriptional regulatory position to a selective gene; and wherein a cell of said one or more cells cannot be selected under said positive selection condition in the absence of high level expression of said selective gene; and

determining whether any of said cells can be selected, wherein selection of a cell is indicative that the said nucleic acid test sequence in said cell contains a transcriptional region active in said cell.

8. The method of claim 7, wherein said selective gene is a gene encoding an antigen.

9. The method of claim 7, wherein said positive selection condition comprises the steps of:

contacting said cells with florescence-labeled antibody specific for cells expressing said selective gene; and

selecting cells expressing said selective gene with florescence activated cell sorting.

10. The method of claim 7, wherein said positive selection condition comprises the steps of:

contacting said cells with iron-labeled antibody specific for cells expressing said selective gene; and

selecting cells expressing said selective gene with magnetic bead sorting.

11. The method of claim 6, wherein said selection condition is a stress condition, said method comprising the steps of:

culturing one or more cells under said stress condition, wherein each of said at least one cell contains a different nucleic acid test sequence inserted in a transcriptional regulatory position to a selective gene, wherein said selective gene is a protective gene and wherein the growth of said one or more cells is inhibited under said stress condition in the absence of high level expression of said protective gene; and

wherein growth of a cell of said at least one cell in the presence of said stress condition is indicative that said nucleic acid test sequence comprises a transcriptional regulatory region active in said cell.

12. The method of claim 11, wherein said stress condition is the presence of at least one biochemical agent.

13. The method of claim 11, wherein said protective gene is an adenosine deaminase gene.

14. The method of claim 12, wherein said at least one

biochemical agent is xylofuranosyl-adenine.

15. The method of claim 11, wherein said protective gene is a dihydrofolate reductase gene.

16. The method of claim 12, wherein said at least one biochemical agent is methotrexate.

17. The method of claim 12, wherein said at least one biochemical agent consists of xylofuranosyl-adenine and deoxycorformacin.

18. The method of claim 12, wherein said at least one biochemical agent consists of alanosine, adenosine, and uridine.

19. The method of claims 8 or 11, wherein said test sequence comprises a combination or modification of known transcription factor response elements.

20. The method of claims 8 or 11, wherein said test sequence comprises one or more binding sites of unknown function.

21. The method of claims 8 or 11, wherein said test sequence comprises a combination of at least one known transcription factor response element and at least one binding site of unknown function.

22. The method of claims 8 or 11, wherein said one or cells are muscle cells.

23. A method of identifying cell- or tissue-specific

synthetic regulatory regions, comprising the steps of:

identifying the oligonucleotides in protein-oligonucleotide complexes formed between a cellular or nuclear extract from a group of cells and a plurality of double-stranded oligonucleotide fragments, wherein said complexes are separated from free said oligonucleotide fragments using size exclusion chromatography; and

wherein the presence of a said oligonucleotide in a said complex is indicative that said oligonucleotide comprises a said binding site.

evaluating a putative cell- or tissue-specific transcriptional regulatory region comprising said binding site under a selection condition by determining whether a cell comprising said putative transcriptional regulatory region in a transcriptional regulatory position to a selective gene is selected under selective conditions, wherein said selection of said cell will only occur if said selective gene is expressed at a sufficiently high level in said cell and wherein said selective gene will only be expressed at said sufficiently high level if said putative transcriptional regulatory region is active in said cell.

24. The method of claim 23, wherein a said double-stranded oligonucleotide fragment is made by synthesizing a single-stranded oligonucleotide and converting said single-stranded oligonucleotide to a double-stranded oligonucleotide.

25. The method of claim 23, wherein said oligonucleotide fragment comprises a central random sequence and both restriction sites and primer sequences in the 5' and 3' ends.

26. The method of claim 23, wherein said complexes are separated from free oligonucleotides using size exclusion

chromatography.

27. The method of claim 23, wherein said identifying comprises amplifying and sequencing said oligonucleotides from said protein-oligonucleotide complexes.

28. The method of claim 23, wherein said amplifying is performed by polymerase chain reaction.

29. A method of claim 23, wherein said selection condition is a positive selection condition, and wherein said evaluating comprises the steps of:

culturing one or more cells under said positive selection condition, wherein at least one said cell contains a nucleic acid test sequence inserted in a transcriptional regulatory position to a selective gene; and wherein said one or more cells cannot be selected under said positive selection condition in the absence of high level expression of said selective gene; and

wherein at least one cell capable to be selected in the presence of the positive selection condition is indicative that the nucleic acid test sequence contains a transcriptional region active in said cell.

30. A method of claim 29, wherein said selective gene is a gene encoding an antigen.

31. A method of claim 29, wherein said positive selection condition comprises the steps of:

contacting said cells with florescence-labeled antibody specific for cells expressing said selective gene; and

selecting out cells expressing said selective gene with florescence activated cell sorting.

32. A method of claim 29, wherein said positive selection condition comprises the steps of:

contacting said cells with magnetic bead-labeled antibody specific for cells expressing said selective gene; and

selecting out cells expressing said selective gene with magnetic sorting.

33. A method of claim 23, wherein said selection condition is a stress condition, and wherein said evaluating comprises the steps of:

culturing one or more cells under said stress condition, wherein at least one said cell contains a nucleic acid test sequence inserted in a transcriptional regulatory position to a protective gene; the growth of said one or more cells being inhibited under said stress condition in the absence of high level expression of said protective gene; and

wherein growth of a cell of said at least one cell in the presence of said stress condition is indicative that said nucleic acid test sequence comprises a transcriptional regulatory region active in said cell.

34. The method of claim 33, wherein said stress condition is the presence of at least one biochemical agent.

35. The method of claim 33, wherein said protective gene is an adenosine deaminase gene.

36. The method of claim 34, wherein said at least one biochemical agent is xylofuranosyl-adenine.

37. The method of claim 33, wherein said protective gene is a dihydrofolate reductase gene.

38. The method of claim 34, wherein said at least one biochemical agent is methotrexate.

39. The method of claim 34, wherein said at least one biochemical agent consists of xylofuranosyl-adenine and deoxycorformacin.

40. The method of claim 34, wherein said at least one biochemical agent consists of alanosine, adenosine, and uridine.

41. The method of claim 29 or 34, wherein said test sequence comprises a combination or modification of at least one known transcription factor response element and at least one said binding site.

42. The method of claims 29 or 34, wherein said one or more cells are muscle cells.

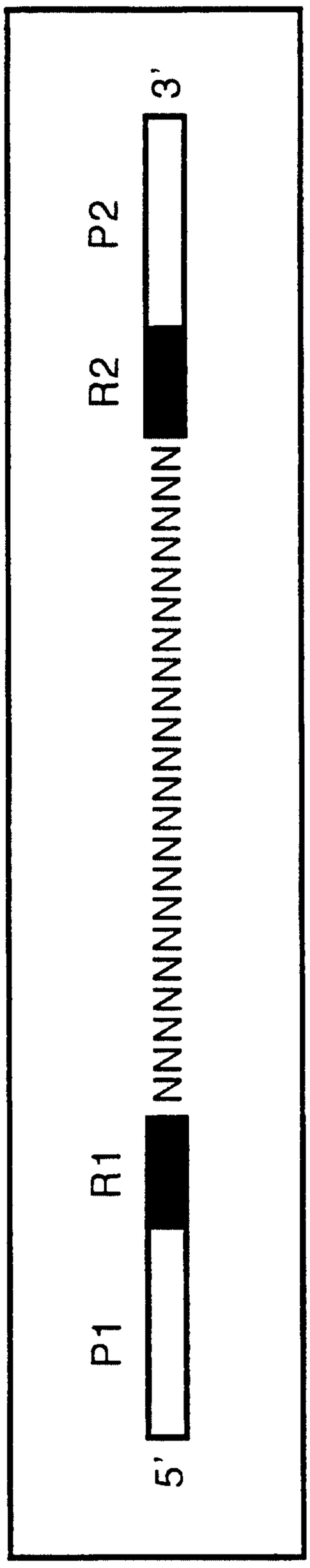


Fig. 1

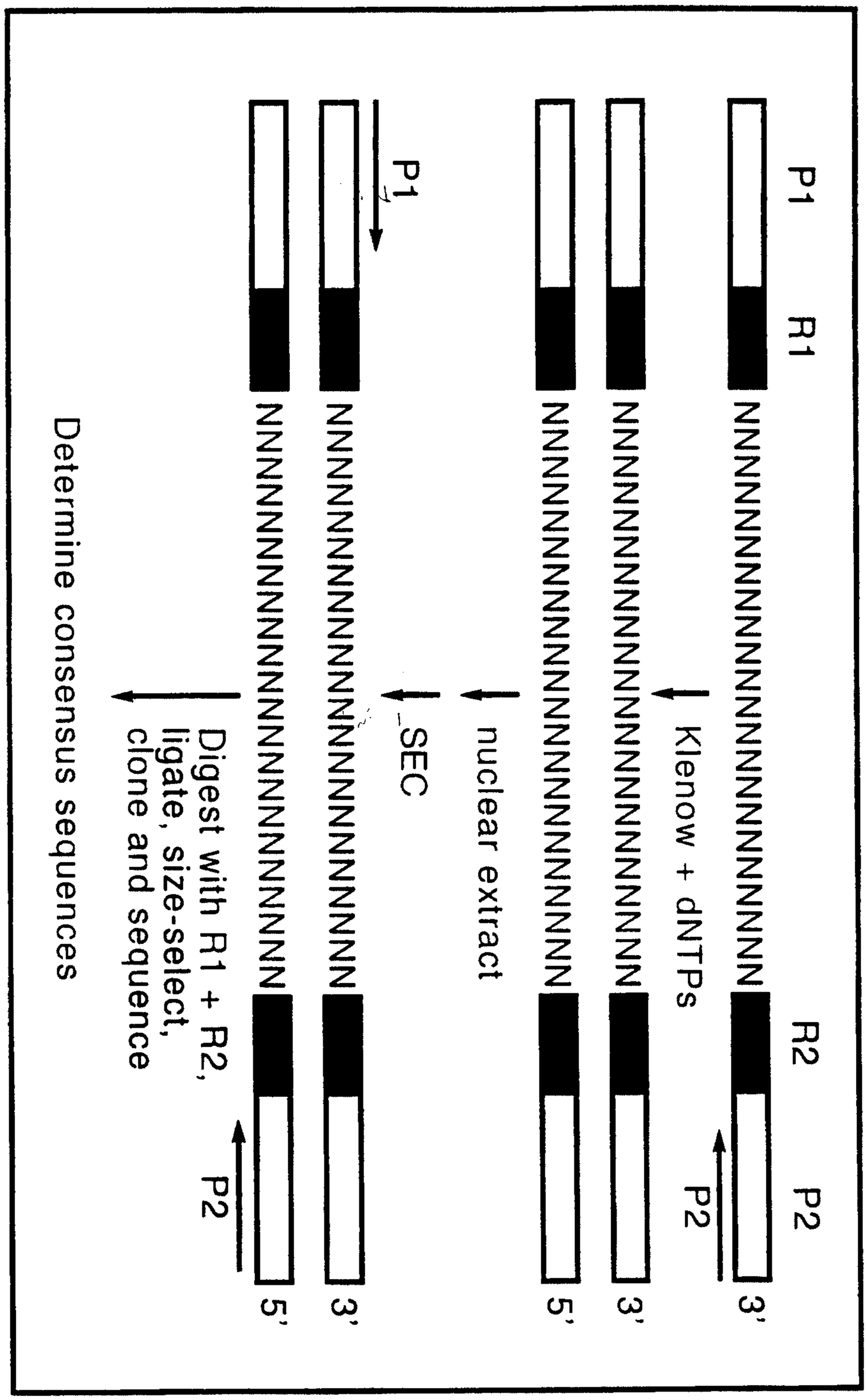

P1
R1
NNNNNNNNNNNNNNN
NNNNNNNNNNNNNNN
NNNNNNNNNNNNNNN
P1
NNNNNNNNNNNNNNN
NNNNNNNNNNNNNNN
Determine consensus sequences
Digest with R1 + R2,
ligate, size-select,
clone and sequence
nuclear extract
SEC
Klenow + dNTPs
R2
P2
P2
P2
5'
3'
5'
3'
3'

Fig. 2

COMPARISON OF RELATIVE PROMOTER ACTIVITY DURING DIFFERENTIATION AT PRIMARY MYOBLAST CELLS

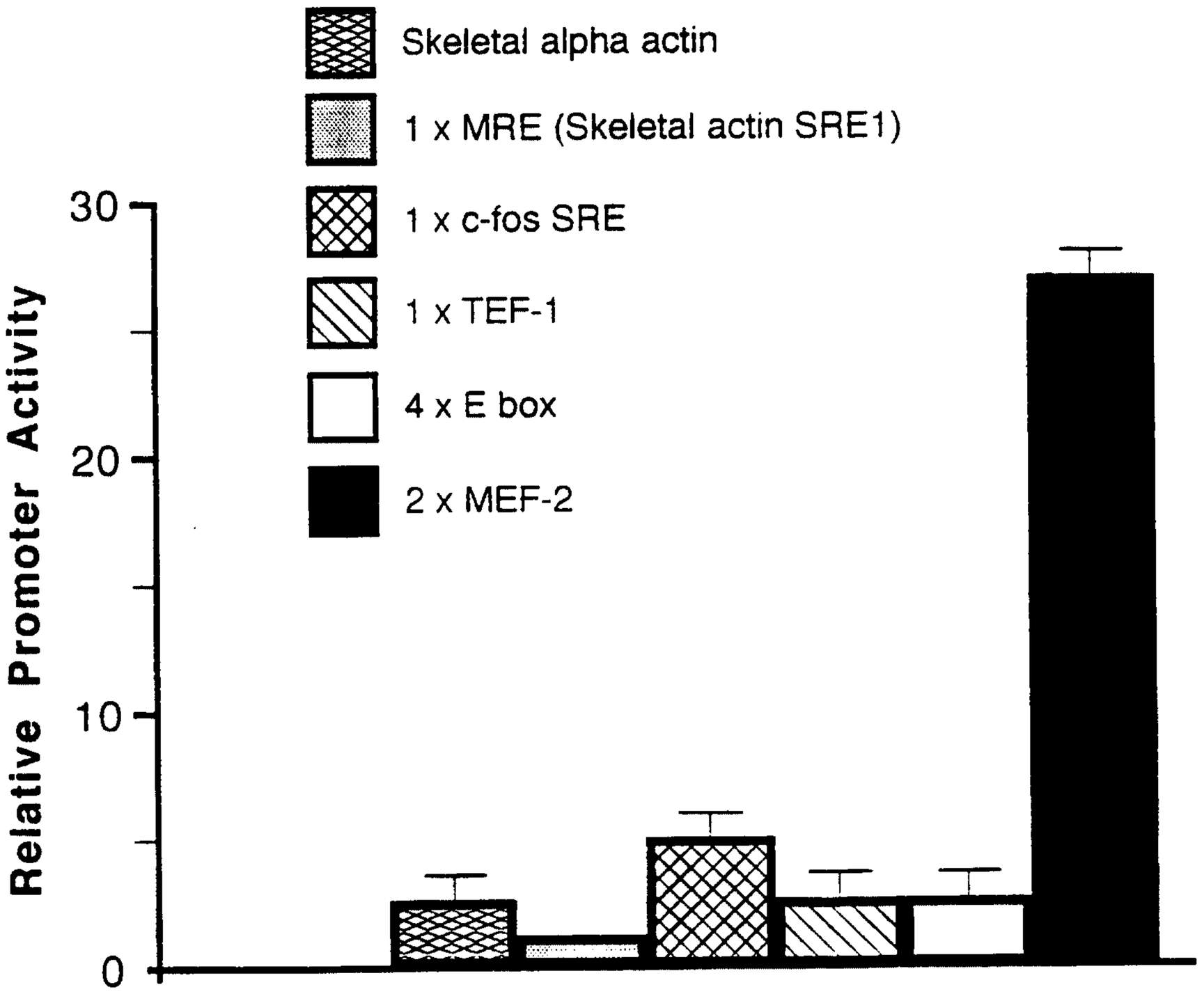
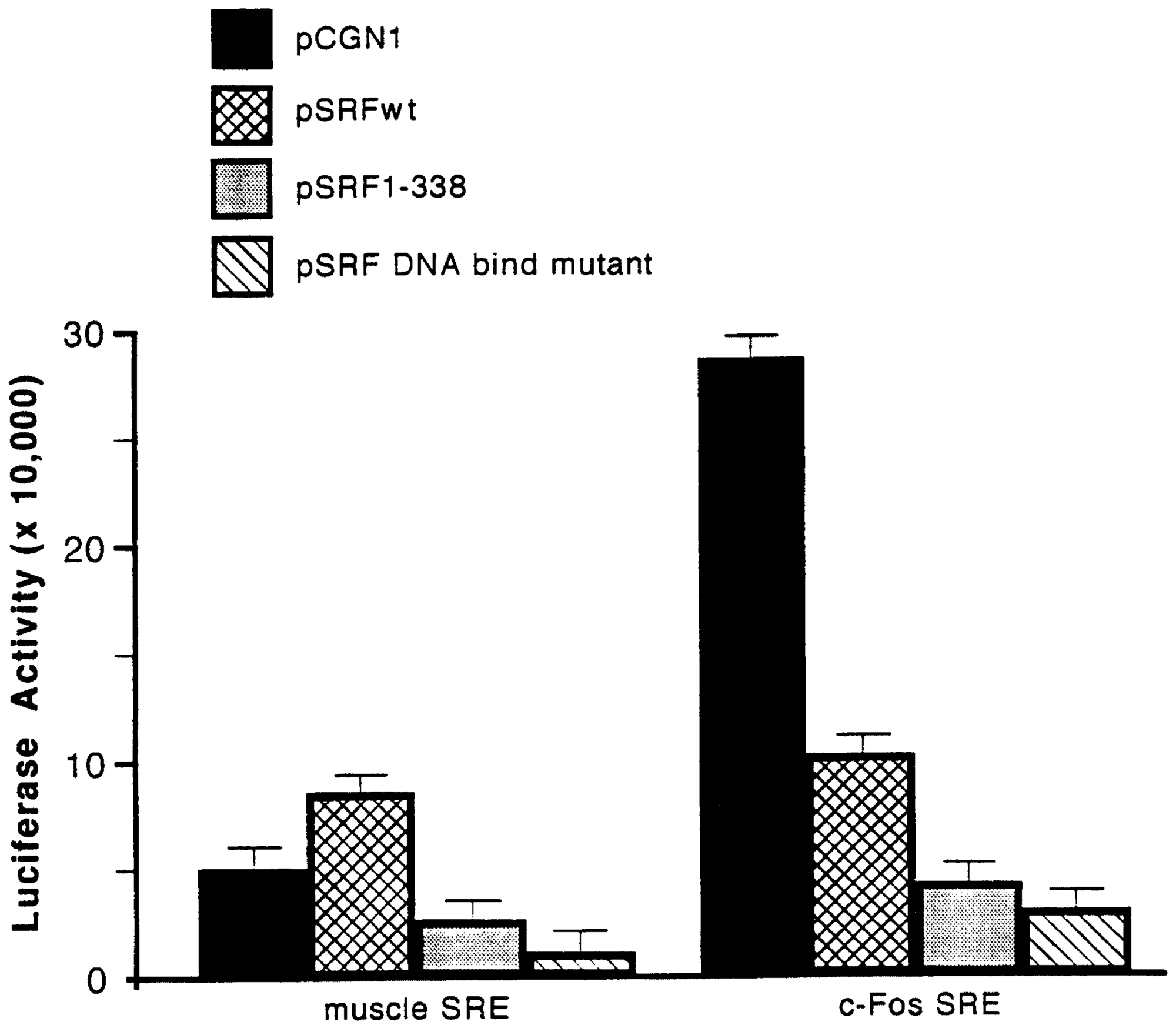


Fig. 3

DIFFERENTIAL SRF ACTIVITY ON c-Fos SRE VS MUSCLE SRE



C-FOS SRE: 5' ACAGGATGTCCATATTAGGACATCTGCG 3'

MUSCLE SRE: 5' GCCCGACACCCAAATATGGCGACGGCCG 3'

Fig. 4

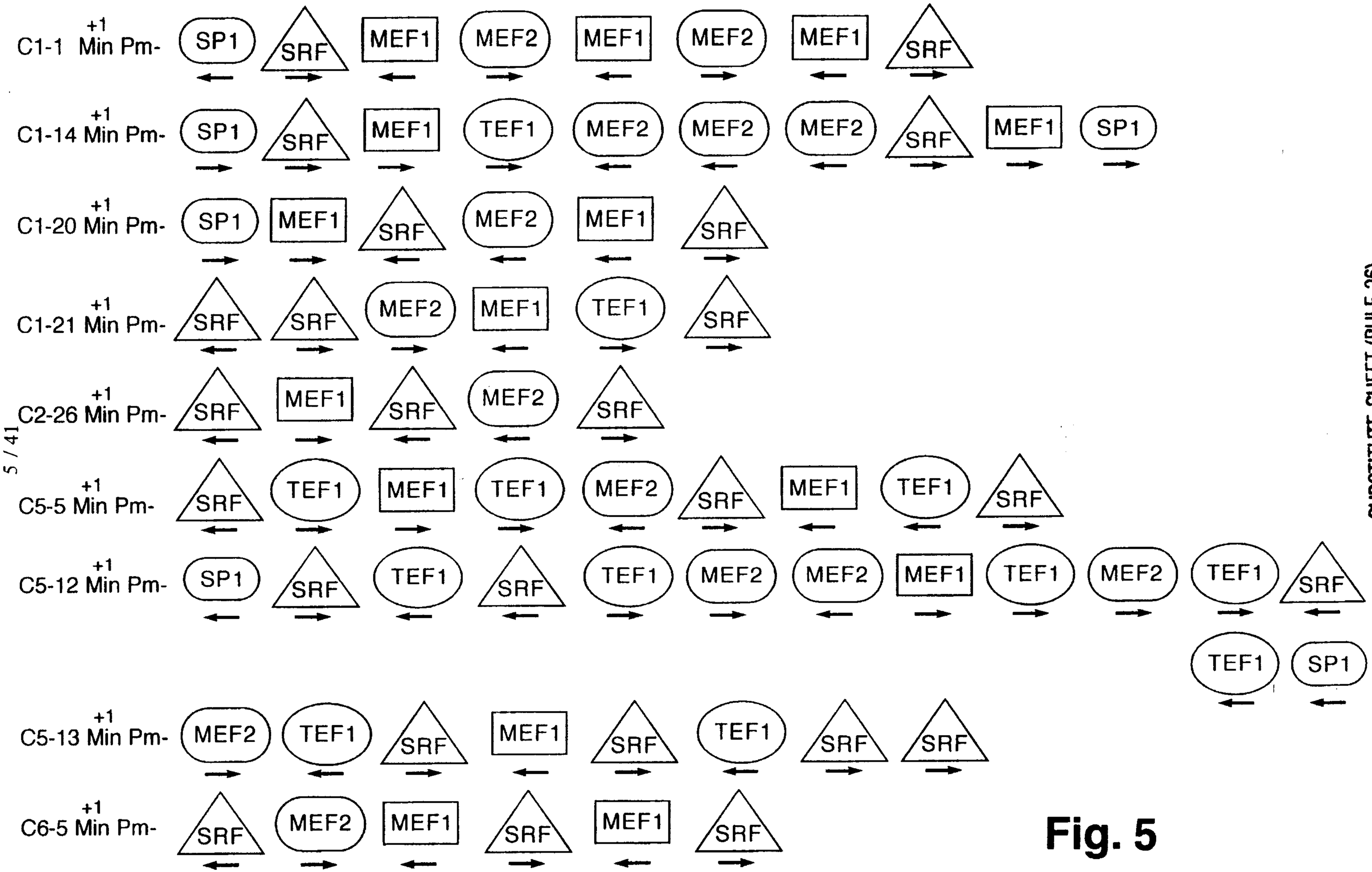


Fig. 5

SUBSTITUTE SHEET (RULE 26)

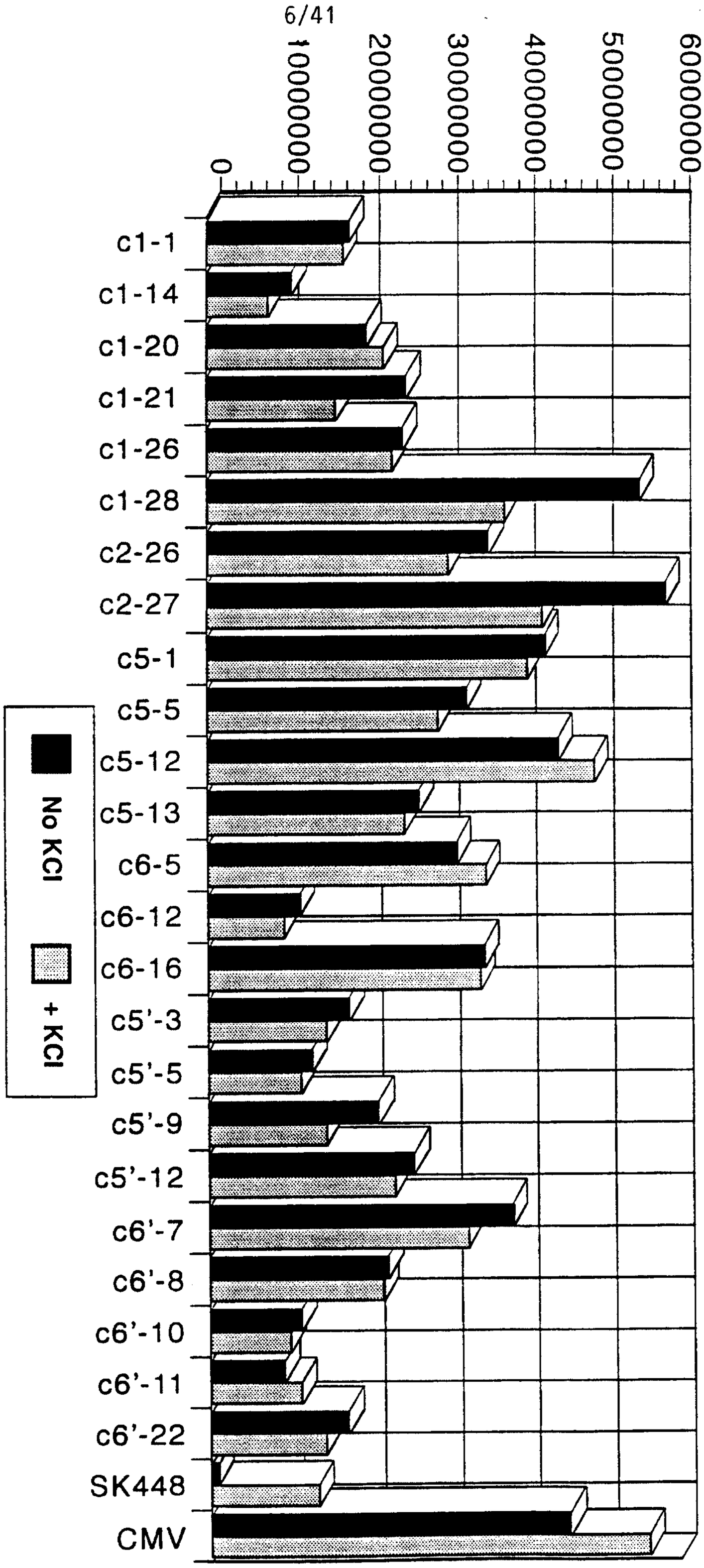

0
10000000
20000000
30000000
40000000
50000000
60000000
c1-1
c1-14
c1-20
c1-21
c1-26
c1-28
c2-26
c2-27
c5-1
c5-5
c5-12
c5-13
c6-5
c6-12
c6-16
c5'-3
c5'-5
c5'-9
c5'-12
c6'-7
c6'-8
c6'-10
c6'-11
c6'-22
SK448
CMV
No KCl
+ KCl

Fig. 6

	1	2	3	4	mean	std
448 crush	124.15	315.14	216.00	314.55	242.46	91.61
C1 28 crush	1523.00	2784.80	947.00	1756.00	1747.70	758.55
C5 12 crush	2770.00		4549.90	3366.00	3561.97	905.99

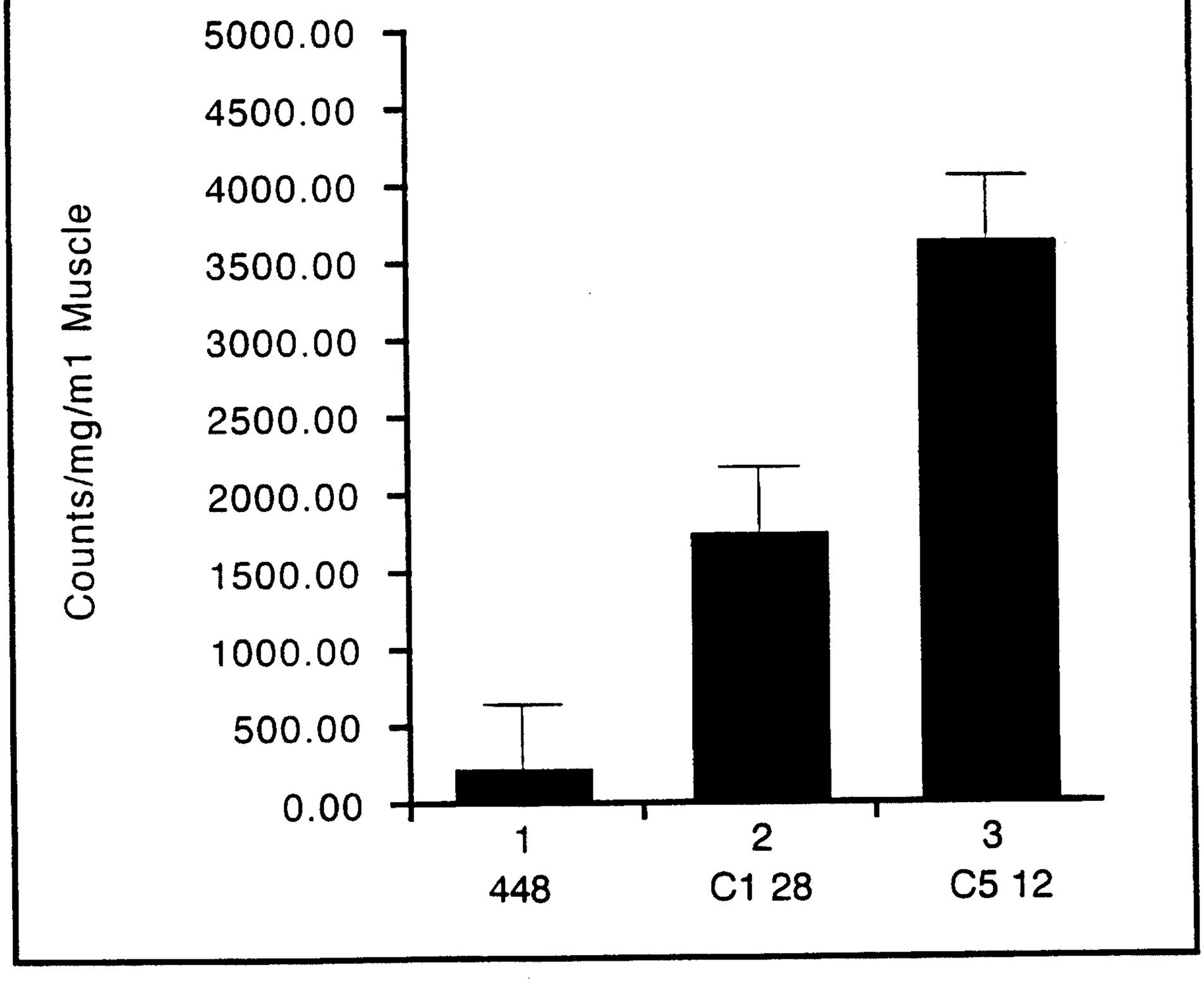


Fig. 7

CA 02296771 2000-01-13

WO 99/02737

SUBSTITUTE SHEET (RULE 26)

10	20	30	40	50	60
ATTTTACAAC	AGTACCGGAA	TGCCAAGCTT	GATATCGAAT	TCCTGCAGCC	CGGGGGATCC
70	80	90	100	110	120
ACTAGTTCTA	GAGCTTGGCG	CCTCCCGCTC	CTCCGGGTAG	CTCGTGGGCC	GCCGCCGGCC
130	140	150	160	170	180
CCGGAGCCTT	TTATCGAGGC	GGGCGGGAGC	ACCGCCCGGC	CCCCAGGAAT	GCGGCCCCGG
190	200	210	220	230	240
CCGTCCGCCC	TCGGGAGTTA	TTTTAGANCG	GTGAGGAATG	GTGCCAACAC	CTGCTGCCTG
250	260	270	280	290	300
CCCCGTCGCC	ATATTTGGGT	GTCGTGAGGA	ATGGTGCCGT	CGCCATATTT	CCGTCGCCAT
310	320	330	340	350	360
ATTTGGGTGT	CCACCATTCC	TCACCGCTCT	AAAAATAACT	CCCGGGAGTT	ATTTTTAGAG
370	380	390	400	410	420
CGCCGTCGCC	ATATTTGGGT	GTCGTGAGGA	ATGGTGCACC	ATTCCTCACC	GCTCTAAAAA
430	440	450	460	470	480
TAACTCCCCC	AACACCTGCT	GCCTGCCCGC	TCTAAAATAA	CTCCCGACAC	CCAAATATGG
490	500	510	520	530	540
CGACGGCCGC	CACCGCGGTG	GANCTCGGTA	CCTCCCGGGT	TATGTTAACT	CANTTACAGT
550	560	570	580	590	600
ACCATAANAT	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .

Fig. 8

10	20	30	40	50	60
ATTTTACAAC	TTCGNGAGAN	TGCCAAGCTT	GATATCGAAT	TCCTGCAGCC	CGGGGGATCC
70	80	90	100	110	120
ACTAGTTCTA	GAGCTTGGCG	CCTCCCGCTC	CTCCGGGTAG	CTCGTGGGCC	GCCGCCGGCC
130	140	150	160	170	180
CCGGAGCCTT	TTATCGAGGC	GGGCGGGAGC	ACCGCCCGGC	CCCCAGGAAT	GCGGCCCCGG
190	200	210	220	230	240
CCGTCGCCAT	ATTTGGGTGT	CCCAACACTG	CTGCCTGCCG	ACACCCAAAT	ATGGCGACGG
250	260	270	280	290	300
GTGAGGAATG	GTGCCAACAC	CTGCTGCCTG	CCGACACCCA	AATATGGCGA	CGGCCGCCAC
310	320	330	340	350	360
CGCGGTGGAG	CTCGGTACCT	CCCGGGTTAT	GTTAGCTCAG	TTACAGTACC	ATAANATACA
370	380	390	400	410	420
TTGATGAGTT	TGGACAAACC	ACAACTANAA	TGCAGTGAAA	AAAATGCTTT	ATTTGTGAAA
430	440	450	460	470	480
TTTGTGATGC	TATTGCTTTA	TTTGTAACCA	TTATAAGCTG	CAATAAACAA	GTTAACAACA
490	500	510	520	530	540
ACAATTGCAT	TCATTTTATG	TTTCAAGTTC	AGGGGGANGT	GTGGGAAGTT	TTTTAAAGCA
550	560	570	580	590	600
AGTAAAACCT	CCACGTACCT	TAATATTACT	TACTTATCAT	GGTACTTGGG	CTGGCGTAAT
610	620	630	640	650	660
. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .

Fig. 9A

BCM17.CP1
BCM17.CP1

```
  1   AATGCCAAGC  TTGATATCGA  ATTCCTGCAG  CCCGGGGGAT  CCACTAGTTC  TAGAGCTTGG
 61   CGCCTCCCGC  TCCTCCGGGT  AGCTCGTGGG  CCGCCGCCGG  CCCCGGAGCC  TTTTATCGAG
121   GCGGGCGGGA  GCACCGCCCG  GCCCCCAGGA  ATGCGGCCCC  GG CCGTCGCC  ATATTTGGGT
                                                          SRE
181   GTC CCAACAC  TGCTGCCTGC  C GACACCCAA  ATATGGCGAC  GG GTGAGGAA  TGGTG CCAAC
                   MEF-1                       SRE                TEF-1
241   ACCTGCTGCC  TGCC GACACC  CAAATATGGC  GACGG CCGCC  ACCGCGGTGG  AGCTCGGTAC
          MEF-1                   SRE
301   CTCCCGGGTT  ATGTTAGCTC  AGTTACAGTA  CCATAAGATA  CATTGATGAG  TTTGGACAAA
361   CCACAACTAG  AATGCAGTGA  AAAAAATGCT  TTATTTGTGA  AATTTGTGAT  GCTATTGCTT
421   TATTTTGTAA  CCATTATAAC  TGCAATAAAC  AATTTAACAA  CAACAATTGC  ATTCCATTTT
481   ATTTTTCAAG  TTCAAGGGGA
```



Fig. 9B

DNASIS
4573-12g.seq

10	20	30	40	50	60
ATTTTACAAC	AGTACGGAAT	GCCAAGCTTG	ATATCGAATT	CCTGCAGCCC	GGGGGATCCA
70	80	90	100	110	120
CTAGTTCTAG	AGCTTGGCGC	CTCCCGCTCC	TCCGGGTAGC	TCGTGGGCCG	CCGCCGGCCC
130	140	150	160	170	180
CGGAGCCTTT	TATCGAGGCG	GGCGGGAGCA	CCGCCCGGCC	CCCAGGAATG	CGGCCCCGGC
190	200	210	220	230	240
CGAGGGCGGA	CACCCAAATA	TGGCGACGGG	TGAGGAACCG	TCGCCATATT	TGGGTGTCCA
250	260	270	280	290	300
CCATTCCTCC	GCTCTAAAAA	TAACTCCCGG	GAGTTATTTT	TAAAGCGCCA	ACACCTGCTG
310	320	330	340	350	360
CCTGCCCACC	TTCCTCACCG	CTCTAAAAAT	AACTCCCCAC	CATTCCTCAC	CCGTCGCCAT
370	380	390	400	410	420
ATTTGGGTGT	CGTGAGGATG	GTGCCGAAGG	CGGACGGCCG	CCACCGCGGT	GGANCTCGGT
430	440	450	460	470	480
ACCTCCCGGG	TTATGTTANC	TCANTTACAN	TACCATAANA	TACATTGATG	AATTTGGACA
490	500	510	520	530	540
AACCACAACT	ANAATGCATG	AAAAAAATGC	TTTATTTGTN	AAATTTGTNA	TGCTATTGCT
550	560	570	580	590	600
TTATTTGTTA	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .

Fig. 10A

BCM12.CP1

BCM12.CP1

```
1    GATATCGANN TCGNGCAGCC CGGGGGATCC ACTNNTTCTA GAGCTTGGCG CCTCCCGCTC
61   CTCCGGGTAG CTCGTGGGCC GCCGCCGGCC CCGGAGCCTT TTATCNAGGC GGGCGGGAGC
121  ACCGCCCGGC CCCCACGAAT GCNGCCCCGG CCGAGGGCGG ACACCCAAAT ATGGCGACGG
181  GTGAGGAACC GTCGCCATAT TTGGGTGTCC ACCATTCCTC CGCTCTAAAA ATAACTCCCG
241  GGAGTTATTT TTAGAGCGCC AACACCTGCT GCCTGCCCAC CTTCCTCACC GCTCTAAAAA
301  TAACTCCCCA CCATTCCTCA CCCGTCGCCA TATTTGGGTG TCGTGAGGAT GGTGCCGAGG
361  GCGGACGGCC GCCACCGCGG TGGAGCTCGG TACCTCCCGG GTTATGTTAG CTCAGTTACA
421  GTACCATAAG ATACATTGAT GAGTTT
```

SPI SRE TEF-1 SRE TEF-1 MEF-2 MEF-2 MEF-1 TEF-1 TEF-1 MEF-2 TEF-1 SRE TEF-1 SPI

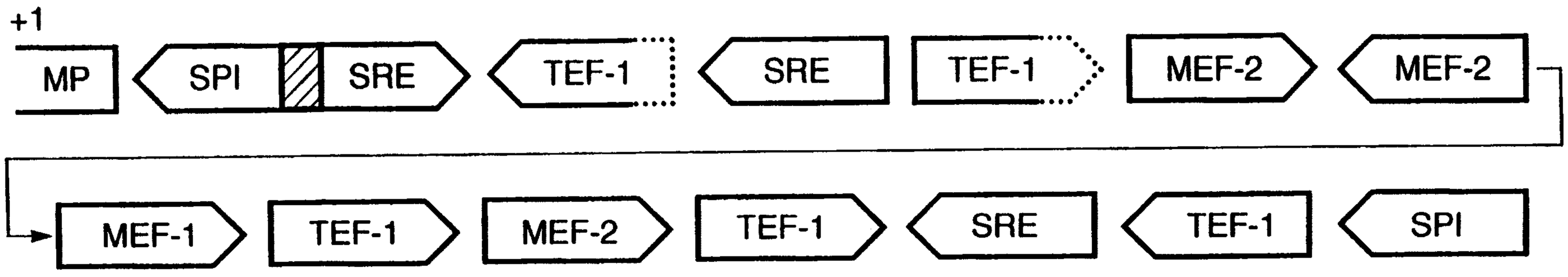


Fig. 10B

CA 02296771 2000-01-13

WO 99/02737

DNASIS
4573-16g.seq

10	20	30	40	50	60
ATTTTACCAA	CAGTACCGGA	ATGCCAAGCT	TGATATCGAA	TTCCTGCAGC	CCGGGGGATC
70	80	90	100	110	120
CACTAGTTCT	AGAGCTTGGC	GCCTCCCGCT	CCTCCGGGTA	GCTCGTGGGC	CGCCGCCGGC
130	140	150	160	170	180
CCCGGAGCCT	TTTATCGAGG	CGGGCGGGAG	CACCGCCCGG	CCCCCAGGAA	TGCGGCCCCG
190	200	210	220	230	240
GCCGAGGGCG	GACACCAAAT	ATGGCGACGG	GGCAGGCAGC	AGGTGTTGGG	GCAGGCAGCA
250	260	270	280	290	300
GGTGTTGGCC	AACACCTGCT	GCCTGCCGAC	ACCCAAATAT	GGCGACGGGG	CAGGCAGCAG
310	320	330	340	350	360
GTGTTGGGGG	AGTTATTTTT	AGAGCGGACA	CCCAAATATG	GCGACGGCCG	CCACCGCGGT
370	380	390	400	410	420
GGAGCTCGGT	ACCTCCCGGG	TTATGTTAGC	TCAGTTACAG	TACCATAAGA	TACATTGATG
430	440	450	460	470	480
AGTTTGGACA	AACCACAACT	ANAATGCAGT	TGAAAAAAAT	GCTTTATTTG	TGAAATTTGT
490	500	510	520	530	540
GATGCTATTG	CTTTATTTGT	AACCATTATA	AGCTGCAATA	AACAATTTAA	CAACAACAAT
550	560	570	580	590	600
TGCATTCCAT	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .

Fig. 11A

BCM16.CP1
BCM16.CP1

1	GCTTGATATC	GAATTCCTGC	AGCCCGGGGG	CATCCACTAT	CTACTAGNGC	TTGACNCCTC	
61	CCGCTCCTCC	GGGTAGCTCG	TGGGCCGCCG	CCGGCCCCGG	ACCCTATNAT	CGAAGCGGGC	
121	NGGANCACNG	CCCGGCCCCC	ACCCAATGCA	GTCCCGGCCC	GAGGGCNCGA	CACCAAATAT	
181	GTGTCACAGG	GCNGGCACCA	GGTGTTGGGG	CAAGCNGCAG	GTGTTTGCCA	ACTCCTGCTG	
241	CCTGCCGACA	CCCANATATG	GCCACNGGGC	ACGNAGCACG	TGTTNGGGGA	GTNATTTTTA	
301	NACCCNACAC	NCANATATGG	NCACNGCCGC	CACCGCGGTN	GANCTCGGTA	ACTCCCGGGT	
361	TATGTTANCT	CAATTACAGT	ACCATAATAT	NCTTTGATNA	ATTTGGACAA	ACCACAACTA	
421	TAATGCAGTG	AAAAAAATGC	TTTATTTGTG	AAATTTGTNA	TGCTATTGCT	TTTATNTNTT	
481	AANCATTANA	AGCTCCAATA	A				

Fig. 11B

DNASIS
4585-1g.seq

10	20	30	40	50	60
ATTTTACAAC	AGTACGGAAT	GCCAAGCTTG	ATATCGAATT	CCTGCAGCCC	GGGGGATCCA
70	80	90	100	110	120
CTAGTTCTAG	AGCTTGGCGC	CTCCCGCTCC	TCCGGGTAGC	TCGTGGGCCG	CCGCCGGCCC
130	140	150	160	170	180
CGGAGCCTTT	TATCGAGGCG	GGCGGGAGCA	CCGCCCGGCC	CCCAGGAATG	CGGCCCCGGC
190	200	210	220	230	240
CGTCCGCCCT	CGGGACACCC	AAATATGGCG	ACGGCGCTCT	AAAAATAACT	CCCCCAACAC
250	260	270	280	290	300
CTGCTGCCTG	CCGACACCCA	AATATGGCAA	CGGGGCNAGG	CAGCAGGTGT	TTGGCGCTCT
310	320	330	340	350	360
AAAAATAACT	CCCCCCGAGG	GCGGACGGCC	CGCCACCGCG	GTNGGAGCTC	GGTACCTCCC
370	380	390	400	410	420
GGGTTATGTT	TAGCTCCAGT	TACAGTACCA	TAAGATACAT	TGAATGATTT	NGGACAAACC
430	440	450	460	470	480
ACAACTAAAA	ATGCAATTGA	AAAAAAATGC	TTTATTTGTT	GAAATTTGTT	GAATGCTATT
490	500	510	520	530	540
GCTTTATTTT	GTTAACCATT	. . . . .	. . . . .	. . . . .	. . . . .

Fig. 12

10	20	30	40	50	60
ATTTTACAAC	AGTACCGGAA	TGCCAAGCTT	GATATCGAAT	TCCTGCAGCC	CGGGGGATCC
70	80	90	100	110	120
ACTAGTTCTA	GAGCTTGGCG	CCTCCCGCTC	CTCCGGGTAG	CTCGTGGGCC	GCCGCCGGCC
130	140	150	160	170	180
CCGGAGCCTT	TTATCGAGGC	GGGCGGGAGC	ACCGCCCGGC	CCCCAGGAAT	GCGGCCCCGG
190	200	210	220	230	240
CCGAGGGCCG	ACGGCCGA..	. . . . .	. . . . .	. . . . .	. . . . .

Fig. 13A

BCM10.CP1
BCM10.CP1

1	AAGCTTGATA	TCGAATTCCT	GCAGCCCGGG	GGATCCACTA	GTTCTAGAGC TTGGCGCCTC
61	CCGCTCCTCC	GGGTAGCTCG	TGGGCCGCCG	CCGGCCCCGG	AGCCTTTTAT CGAGGCGGGC
121	GGGAGCACCG	CCCGGCCCCC	AGGAATGCGG	CCCCGGCCGA	TGGCGGACGG CCGAT

Fig. 13B

WO 99/02737
CA 02296771 2000-01-13
16 / 41

SUBSTITUTE SHEET (RULE 26)

```
10 ATTTTACAAC 20 AGTACGGAAT 30 GCCAAGCTTG 40 ATATCGAATT 50 CCTGCAGCCC 60 GGGGGAATCC
70 ACTAGTTCTA 80 GAGCTTGGCG 90 CCTCCCGCTC 100 CTCCGGGTAG 110 CTCGTGGGCC 120 GCCGCCGGCC
130 CCGGAGCCTT 140 TTATCGAGGC 150 GGGCGGGAGC 160 ACCGCCCGGC 170 CCCCAGGAAT 180 GCGGCCCCGG
190 CCGTCGCCAT 200 ATTTGGGTGT 210 CCACCATTCC 220 TCACCGCTCT 230 AAAAATAACT 240 CCCGTGAGGA
250 ATGGTGCACC 260 ATTCCTCACC 270 CGTCGCCATA 280 TTTGGGTGTC 290 CCGAGGGCGG 300 ACGGCCGCCA
310 CCGCGGTGGA 320 GCTCGGTACC 330 TCCCGGGTTA 340 TGTTAGCTCA 350 GTTACAGTAC 360 CATAAGATAC
370 ATTGATGAGT 380 TTGGACAAAC 390 CACAACTAGA 400 ATGCAGTGAA 410 AAAAATGCTT 420 TATTTGTGAA
430 ATTTGTGATG 440 CTATTGCTTT 450 ATTTGTAACC 460 ATTATAAGCT 470 GCAATAAACA 480 AGTTAACAAC
490 AACAATTGCA 500 TTCATTTTAT 510 GTTTCANGTT 520 CAAGGGGAAG 530 TNTTGGAAGT 540 TTTTTTAAAN
550 CAATTAAAAC 560 ..... 570 ..... 580 ..... 590 ..... 600 .....
```

Fig. 14A

BCM11.CP1
BCM11.CP1

```
1    AAGCTTGATA  TCGAATTCCT  GCAGCCCGGG GGATCCACTA  GTTCTAGAGC  TTGGCGCCTC
61   CCGCTCCTCC  GGGTAGCTCG  TGGGCCGCCG CCGGCCCCGG  AGCCTTTTAT  CGAGGCGGGC
121  GGGAGCACCG  CCCGGCCCCC  AGGAATGCGG  CCCCGGCCGT  CGCCATATTT  GGGTGTCCAC
                                                                SRE
181  CATTCCTCAC  CCAACACCTG  CTGCCTGCCC  ACCATTCCTC  ACGGGAGTTA  TTTTTAGAGC
     TEF-1                   MEF-1                  TEF-1                  MEF-2
241  GGACACCCAA  ATATGGCGAC  GGGGCAAGCA  NCANGTGTTG  GGTNAGGAAT  GGTGGACACC
                                SRE                  MEF-1                  TEF-1
301  CAAATATGGC  GACGGCCGGG  GCCGCATTCC  TGGGGGCCGG  GCGGTGCTCC  CGCCCGCCTC
                SRE
361  GATAAAAGCT  CCGGGGCCGG  CGGCGGCCAC  GAACTACCCG  GANGAACGGG  AAGCGCCAAN
421  CTCTANAACT  AATGGATCCC  CCGGGCTGCA  AGAATTCGAT  ATCAAGCTTG  GCATTCCGGG
481  TACTGTTGGT  AA
```

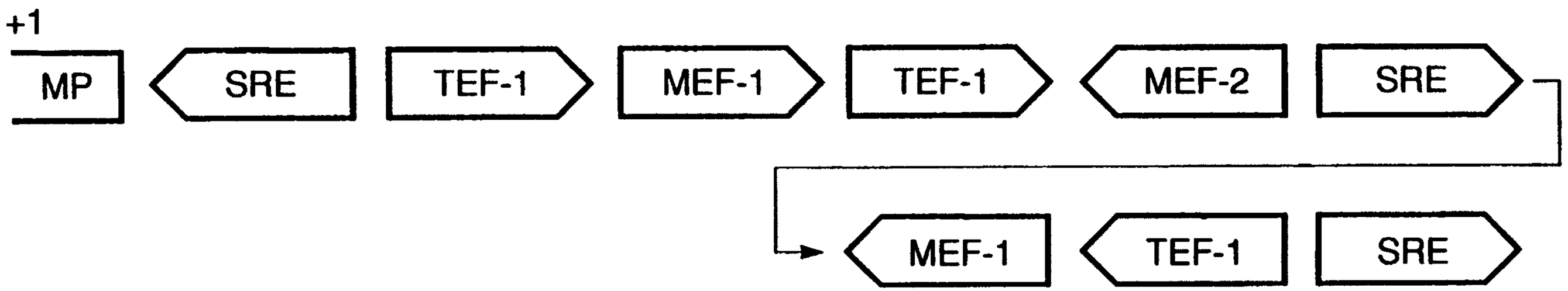


Fig. 14B

DNASIS
4573-14g.seq

```
                10                  20                  30                  40                  50                  60
ATTTTACAAC  AGTACCGGAA  TGCCAAGCTT  GATATCGAAT  TCCTGCAGCC  CGGGGGATCC

                70                  80                  90                 100                 110                 120
ACTAGTTCTA  GAGCTTGGCG  CCTCCCGCTC  CTCCGGGTAG  CTCGTGGGCC  GCCGCCGGCC

               130                 140                 150                 160                 170                 180
CCGGAGCCTT  TTATCGAGGC  GGGCGGGAGC  ACCGCCCGGC  CCCCAGGAAT  GCGGCCCCGG

               190                 200                 210                 220                 230                 240
CCGTCGCCAT  ATTTGGGTGT  ..........  ..........  ..........  ..........
```

Fig. 15A

RECTIFIED SHEET (RULE 91)
ISA/EP

BCM14.CP1

BCM14.CP1

```
1     GATATCGAAT  TCNTGCAGCC  CGGGGGATCC  ACTAGTTCTA  GAGCTTGGCG  CCTCCCGCTC
61    CTCCGGGTAG  CTCGTGGGCC  GCCGCCGGCC  CCGGAGCCTT  TTATCGAGGC  GGGCGGGAGC
121   ACCGCCCGGC  CCCCAGGAAT  GCGGCCCCGG  CCGTCGCCAT  ATTTGGGTGT  CCGCTCTAAA
                                          SRE
181   AATAACTCCC  GGCAGGCAGC  AGGTGTTGGC  CAACACCTGC  TGCCTGCCGA  CACCAAATAT
      MEF-2       MEF-1                  MEF-1                  SRE
241   GGCGACGGGG  CAGGCAGCAG  GTGTTGGGAC  ACCCAAATAT  GGCGACGGCC  GCCACCGCGG
                  MEF-1                  SRE
301   TGGAGCTCGG  TACCTCCCGG  GTTATGTTAG  CTCAGTTACA  GTACCATAAG  ATACATTGAT
361   GAGTTTGGAC  AAACCACAAC  TAGAATGCAG  TGAAAAAAAT  GCTTTATTTG  TGAAATTTGT
421   GATGCTATTG  CTTTATTTGT  AACCATTATA  AGCTGCAATA  AACAAGTTAA  CAACAACAAT
481   TGCATTCATT  TTATTTTCA
```

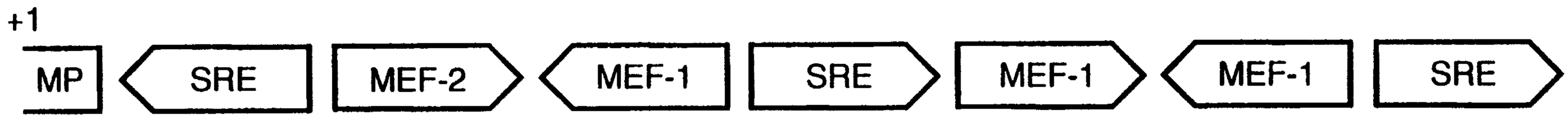


Fig. 15B

SUBSTITUTE SHEET (RULE 26)

10	20	30	40	50	60
ACAACAGTAC	CGGAATGCCA	AGCTTGATAT	CGAATTCCTG	CAGCCCGGGG	GATCCACTAG
70	80	90	100	110	120
TTCTAGAGCT	TGGCGCCTCC	CGCTCCTCCG	GGTAGCTCGT	GGGCCGCCGC	CGGCCCCGGA
130	140	150	160	170	180
GCCTTTTATC	GAGGCGGGCG	GGAGCACCGC	CCGGCCCCCA	GGAATGCGGC	CCCGGCCGAG
190	200	210	220	230	240
GGCGGACACC	AATATGGCGA	CGGGGCAGGC	AGCAGGTGTT	GGCGCTCTAA	AAATAACTCC
250	260	270	280	290	300
CGGCAGGCAG	CAGGTGTTGG	CGCTCTAAAA	ATAACTCCCG	GCAGGCAGCA	GGTGTTGGGA
310	320	330	340	350	360
CACCCAAATA	TGGCGACGGC	CGCCACCGCG	GTGGAGCTCG	GTACCTCCCG	GGTTATGTTA
370	380	390	400	410	420
GCTCAGTTAC	AGTACCATAA	GATACATTGA	TGAGTTTGGA	CAAACCACAA	CTAGAATGCA
430	440	450	460	470	480
GTGAAAAAAA	TGCTTTATTT	GTGAAATTTG	TGATGCTATT	GCTTTATTTG	TAACCATTAT
490	500	510	520	530	540
AAGCTGCAAT	AAACAAGTTA	ACAACAACAA	TTGCATTCAT	TTTATGTTTC	ANGTTCANGG
550	560	570	580	590	600
GGAAGTGTGG	GAAGTTTTTT	AAAGCAAGTA	AAACTCCACG	TACCTTAATA	TTACTTACTT
610	620	630	640	650	660
. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .

Fig. 16A

BCM3.CP1
BCM3.CP1

```
   1  GATATCGAAT  TCCTGCAGCC  CGGGGGATCC  ACTAGTTCTA  GAGCTTGGCG  CCTCCCGCTC
  61  CTCCGGGTAG  CTCGTGGGCC  GCCGCCGGCC  CCGGAGCCTT  TTATCGAGGC  GGGCGGGAGC
 121  ACCGCCCGGC  CCCCAGGAAT  GCGGCCCCGG  CCGAGGGCGG  ACACCAATAT  GGCGACGGGG
                                          SPI         SRE
 181  CAGGCAGCAG  GTGTTGGCGC  TCTAAAAATA  ACTCCCGGCA  GGCAGCAGGT  GTTGGCGCTC
      MEF-2                              MEF-2
 241  TAAAAATAAC  TCCCGGCAGG  CAGCAGGTGT  TGGGACACCC  AAATATGGCG  ACGGCCGCCA
           MEF-2       MEF-1                  SRE
 301  CCGCGGTGGA  GCTCGGTACC  TCCCGGGTTA  TGTTAGCTCA  GTTACAGTAC  CATAAGATAC
 361  ATTGATGAGT  TTGGACAAAC  CACAACTAAG  AATGCAGTGA  AAAAAATGCT  TTATTTGTTG
 421  AAATTTGTTG  ATGCTATTGC  TTTATTTGTT  AACCCATTAT  AAGCTTGCCA  ATAAACAA
```

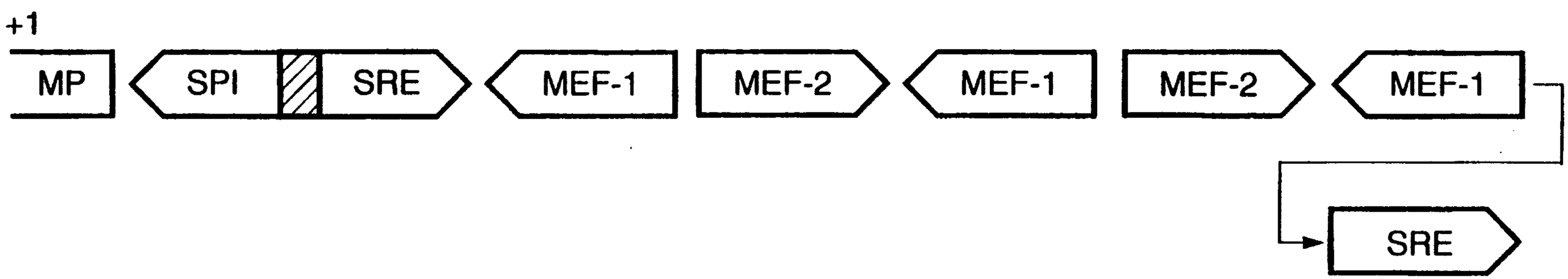


Fig. 16B

CA 0229677 2000-01-13
WO 99/02737
22 / 41
SUBSTITUTE SHEET (RULE 26)

10	20	30	40	50	60
ATTTTACAAC	AGTACTGGAA	TGCCAAGCTT	GATATCGAAT	TCCTGCAGCC	CGGGGGNTCC
70	80	90	100	110	120
ACTAGTTCTA	GAGCTTGGCG	CCTCCCGCTC	CTCCGGGTAG	CTCGTGGGCC	GCCGCCGGCC
130	140	150	160	170	180
CCGGAGCCTT	TTATCGAGGC	GGGCGGGAGC	ACCGCCCGGC	CCCCAGGAAA	TGCGGCCCCG
190	200	210	220	230	240
GCCGTCCGCC	CTCGGCCGTC	GCCATATTTG	GGTGTCCCAA	CACCTGCTGC	CTGCCCACCA
250	260	270	280	290	300
TCCTCACGGG	AGTTATTTTT	ANAGCGGGGA	GTTATTTTAN	ANCGGGGANT	TATTTTANA .

Fig. 17A

BCM5.CP1

BCM5.CP1

1	AAGCTTGATA	TCGAATTCCT	GCAGCCCGGG	GGATCCACTA	GTTCTAGAGC	TTGGCGCCTC
61	CCGCTCCTCC	GGGTAGCTCG	TGGGCCGCCG	CCGGCCCCGG	AGCCTTTTAT	CGAGGCGGGC
121	GGGAGCACCG	CCCGGCCCCC	AGGAATGCGG	NCCCGGCCGT	CCGCCCTGCT	GCCTGCGCCG
181	TCGCCATATT	TGGGTGTGGG	GAGTTATTTT	TAGAGCGGGC	AGGCANCAGG	TGTTGGGACA
241	CCCAAATATG	GCGACGGCCG	CCACCGCGGT	GGAGCTCGGT	ACCTCCCGGG	TTATGTTAGC
301	TCAGTTACAG	TACCATAAGA	TACATTGATG	AGTTTGGACA	AACCACAACT	AGAATGCAGT
361	GAAAAAAATG	CTTTATTTGT	GAAATTTGTG	ATGCTATTGC	TTTATTTGTA	ACCATTATAA
421	CTGCAATAAA	CAATTTAACA	ACAACAATTG	CATTCATTTT	ATGTTTCAGG	TTCAGGGGAA
481	GTTTTGGAAG	TTTTTTAAAC	CAATTAAACC	CCAC		

SPI MEF-1

SRE MEF-2 MEF-1

SRE

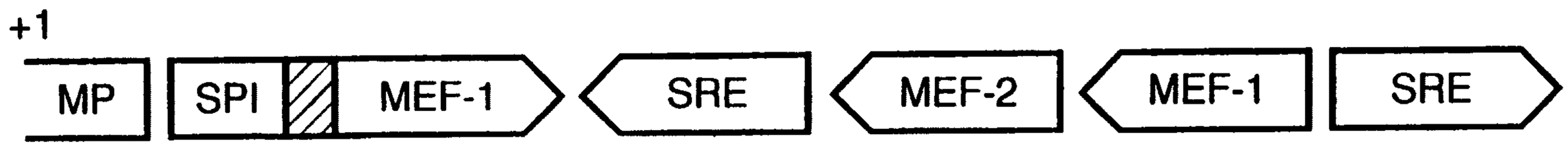


Fig. 18

BCM4.CP1
BCM4.CP1

```
1    GATATCGAAT  TCCTGCAGCC  CGGGGGATCC  ACTAGTTCTA  GAGCTTGGCG  CCTCCCGCTC
61   CTCCGGGTAG  CTCGTGGGCC  GCCGCCGGCC  CCGGAGCCTT  TTATCGAGGC  GGGCGGGAGC
121  ACCGCCCGGC  CCCCAGGAAT  GCGGCCCCGG  CCGTCCGCCC  TCGGCCGTCG  CCATATTTGG
                                                SPI         SRE
181  GTGTCCCAAC  ACCTGCTGCC  TGCCCACCAT  CCTCACGGGA  GTTATTTTTA  GAGCGGGGAG
                     MEF-1                TEF-1          MEF-2
241  TTATTTTAGA  GCGGGGAGTT  ATTTTAGAGC  GGACACCCAA  ATATGGCGAC  GGCCAACACC
          MEF-2                  MEF-2                  SRE
301  TGTGCCTGCC  CCGAGGGCGG  ACGGCCGCCA  CCGCGGTGGA  NCTCGGTACC  TCCCGGGTTA
     MEF-1            SPI
361  TGTTANCTCA  GTTACAGTAC  CATAAGATAC  ATTGATGAAT  TTGGACAAAC  CACAACTAGA
421  ATGCAGTGAA  AAAAATGCTT  TATTTGTTAA  ATTTGTGATG  CTATTGCTTT  ATTTGTTAAC
481  CATTATAAGC  TGCAATAAAC  AA
```

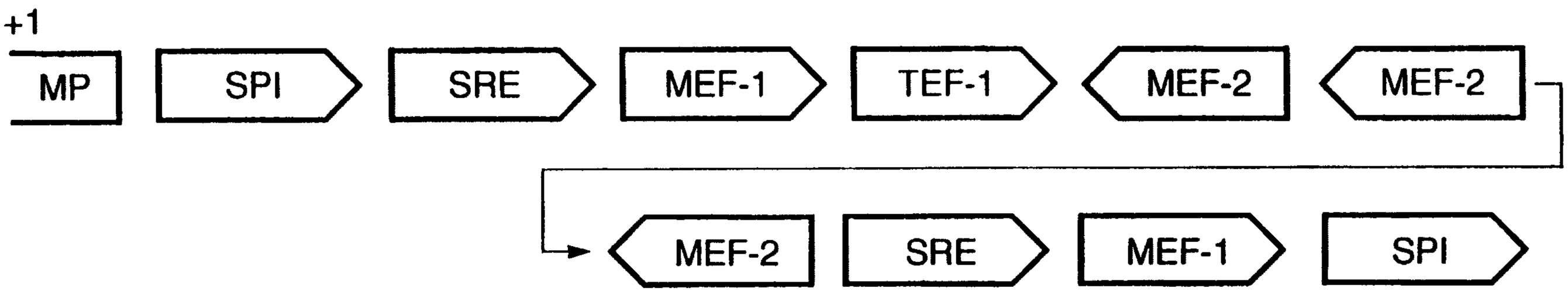


Fig. 17B

BCM6.CP1

BCM6.CP1

```
1     AAGCTTGATA  TCGAATTCCT  GCAGCCCGGG  GGATCCACTA  GTTCTAGAGC  TTGGCGCCTC
61    CCGCTCCTCC  GGGTAGCTCG  TGGGCCGCCG  CCGGCCCCGG  AGCCTTTTAT  CGAGGCGGGC
121   GGGAGCACCG  CCCGGCCCCC  AGGAAATGCG  GCCCCGGCCG  TCGCCATATT  TGGGTGTCGA
                                                     SRE
181   CACGCAAATA  TGGCGACGGC  GCTCTAAGAA  TNNCTCCCGG  CAGGCAGCAN  GTGTTGGCAC
                  SRE                   MEF-2                  MEF-1
241   CATTCCTCAC  GACACCCAAA  TATGGCGACG  GCCGCCACCG  CGGTGGANCT  CGGTACCTCC
      TEF-1                 SRE
301   CGGGTTATGT  TANCTCANTT  ACAGTACCAT  AANATACATT  GATGAGTTTG  GACAAACCAC
361   AACTANAATG  CANTGAAAAA  AATGCTTTAT  TTGTNAAATT  TGTTGATGCT  ATTGCTTTAT
421   TTGTAACCAT  TATAACTGCA  ATAAACAATT  TAACAACAAC  AATTGCATTC  ATTTTATGTT
```



Fig. 19

CA 02296771 2000-01-13

WO 99/02737

SUBSTITUTE SHEET (RULE 26)

```
                10          20          30          40          50          60
AGTACCGGAA  TGCCAAGCTT  GATATCGAAT  TCCTGCAGCC  CGGGGGATCC  ACTAGTTCTA

                70          80          90         100         110         120
GAGCTTGGCG  CCTCCCGCTC  CTCCGGGTAG  CTCGTGGGCC  GCCGCCGGCC  CCGGAGCCTT

               130         140         150         160         170         180
TTATCGAGGC  GGGCGGGAGC  ACCGCCCGGC  CCCCAGGAAT  GCGGCCCCGG  CCGTCGCCAT

               190         200         210         220         230         240
ATTTGGGTGT  CCACCATTCC  TCACCGCTCT  AAAAATAACT  CCCCGCTCTA  AAAATAACTC

               250         260         270         280         290         300
CCGGCAGGCA  GCAGGTGTTG  G.........  ..........  ..........  ..........
```

Fig. 20A

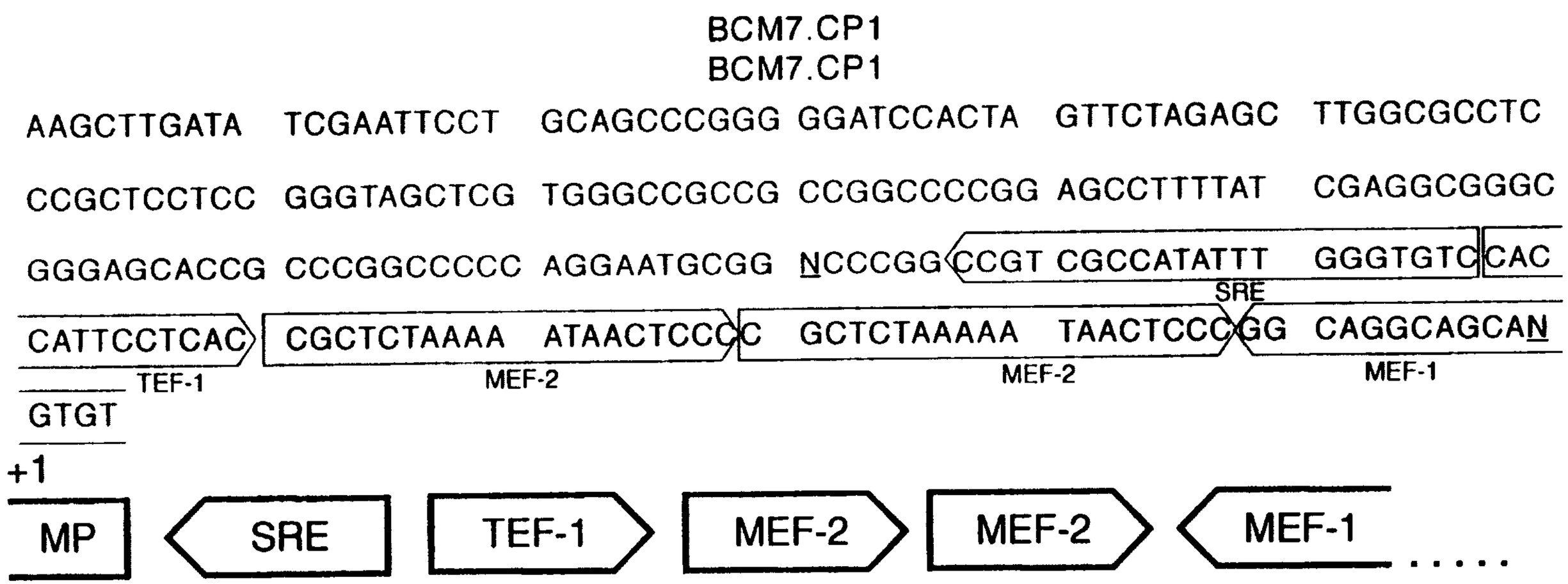


Fig. 20B

10	20	30	40	50	60
ACGAGAATGC	NAAGCTTGAT	ATCGAATTCC	NGCAGCCCGG	GGGATNCACT	AGTTCTACAN
70	80	90	100	110	120
CTTGGCGCCT	CCCGCTCCTC	CGGGTACCTC	GTGGGCCGCC	GCCGGCCCCG	GAGCCTTTTA
130	140	150	160	170	180
TCGAGGCGGG	CGGGAGCACC	GCCNGGCCCC	CANGAATGCG	GCCCCGGCCG	TCGCCATATT
190	200	210	220	230	240
TGGGTGTCCC	AACACCTGCT	GCCTGCCCCG	TCGCCATATT	TGGGTGTCGG	GAGTTATTTT
250	260	270	280	290	300
TAGANCNGAC	ACCCAAATAT	GGCGACGGCC	GCCACCGCGG	TGGAGCTCGG	TACCTCCCGG
310	320	330	340	350	360
GTTATGTTAN	CTCAGTTACA	GTACNATAAN	ATACATTGAT	GACTTTGGAC	AAACCNCAAC
370	380	390	400	410	420
TAAAATGCAG	TGAAAAAAAT	GCTTTATNTG	TGAAATTTGT	GATNCTATTG	CTTTATTTGT
430	440	450	460	470	480
AACCATTATA	AGCTGCAATA	AACAANTTAA	CAACNACAAT	GGCATNCATT	TTATGTATCA
490	500	510	520	530	540
CGTTCACGGG	GAGGTGTGGG	. . . . .	. . . . .	. . . . .	. . . . .

Fig. 21A

BCM9.CP1
BCM9.CP1

```
  1   AAGCTTGATA  TCGAATTCCT  GCAGCCCGGG GGATCCACTA  GTTCTAGAGC  TTGGCGCCTC
 61   CCGCTCCTCC  GGGTAGCTCG  TGGGCCGCCG CCGGCCCCGG  AGCCTTTTAT  CGAGGCGGGC
121   GGGAGCACCG  CCCGGCCCCC  AGGAATGCGG  CCCCGGCCGT  CGCCATATTT  GGGTGTCCCA
                                                                  SRE
181   ACACCTGCTG  CCTGCCGCGT  CGCCATATTT  GGGTGTCGGG  AGTTATTTTT  AGAGCGGACA
                MEF-1                      SRE                    MEF-2
241   CCCAAATATG  GCGACGGCCG  CCACCGCGGT  GGAGCTCGGT  ACCTCCCGGG  TTATGTTAGC
                SRE
301   TCAGTTACAG  TACCATAAGA  TACATTGATG  AGTTTGGACA  AACCACAACT  AGAATGCAGT
361   GAAAAAAATG  CTTTATTTGT  GAAATTTGTG  ATGCTATTGC  TTTATTTGTA  ACCATTATAA
421   CTGCAATAAA  CAATTTAACA  ACAACAATTG  CATTCATT
```

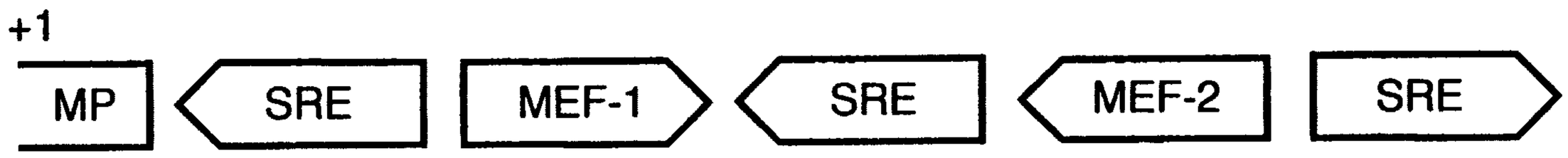


Fig. 21B

DNASIS
4573-13g.seq

10	20	30	40	50	60
ATTTTACAAC	AGTACCGGAA	TGCCAAGCTT	GATATCGAAT	TCCTGCAGCC	CGGGGGATCC
70	80	90	100	110	120
ACTAGTTCTA	GAGCTTGGCG	CCTCCCGCTC	CTCCGGGTAG	CTCGTGGGCC	GCCGCCGGCC
130	140	150	160	170	180
CCGGAGCCTT	TTATCGAGGC	GGGCGGGAGC	ACCGCCCGGC	CCCCAGGAAT	GCGGCCCCGG
190	200	210	220	230	240
ACGCCATTTC	TCTCCTCTAA	AATAACTCCC	GTGAGGAATG	GTGGACACCC	AAATATGGCG
250	260	270	280	290	300
ACGGGGCAGG	CAGCAGGTGT	TGGGACACCC	AAATATGGCG	ACGGGTGAGG	AATGGTGGAC
310	320	330	340	350	360
ACCCAAATAT	GGCGACGGGA	CACCCAAATA	TTTGG.....	....	....

Fig. 22A

WO 99/02737

CA 02296771 2000-01-13

SUBSTITUTE SHEET (RULE 26)

BCM13.CP1
BCM13.CP1

```
1    AAGCTTGATA  TCGACTTCCT  GCAGCCCGGG GGATCCACTA  GTTCTAGAGC  TTGGCGCCTC
61   CCGCTCCTCC  GGGTAGCTCG  TGGGCCGCCG CCGGCCCCGG  AGCCTTTTAT  CGAGGCGGGC
121  GGGAGCACCG  CCCGGCCCCC  AGGAATGCGG  CCCCGGACGC  CATTTCTCTC  CTCTAAAATA
                                                                MEF-2
181  ACTCCC GTGA GGAATGGTGG  ACACCCAAAT  ATGGCGACGG  GGCAGGCAGC  AGGTGTTGG G
                TEF-1                   SRE                MEF-1
241  ACACCCAAAT  ATGGCGACGG  GTGAGGAATG  GTGGACACCC  AAATATGGCG  ACGGGACACC
                SRE                     TEF-1                  SRE        SRE
301  CA....
```

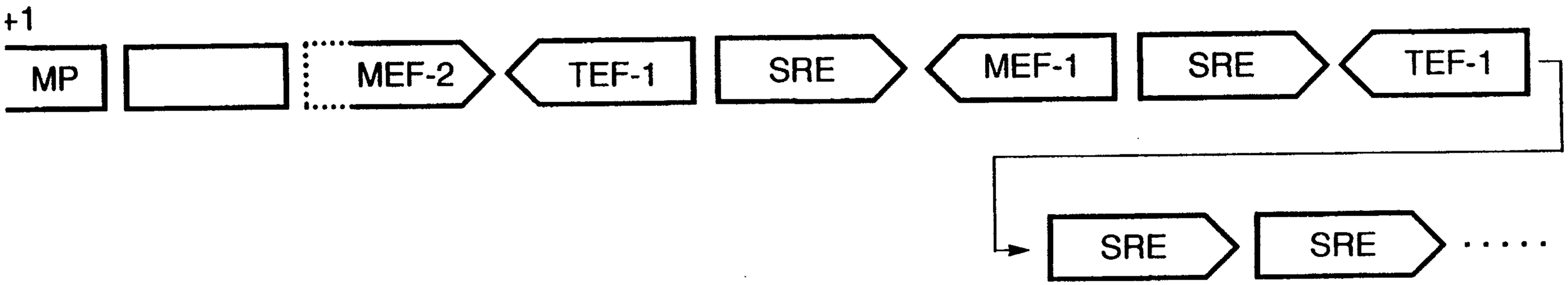


Fig. 22B

CA 022967 2000-01-13

WO 99/02737

SUBSTITUTE SHEET (RULE 26)

DNASIS
4573-17g.seq

10	20	30	40	50	60
ATTTTACAAC	AGTACCGGAA	TGCCAAGCTT	GATATCGAAT	TCCTGCAGCC	CGGGGGATCC
70	80	90	100	110	120
ACTAGTTCTA	GAGCTTGGCG	CCTCCCGCTC	CTCCGGGTAG	CTCGTGGGCC	GCCGCCGGCC
130	140	150	160	170	180
CCGGAGCCTT	TTATCGAGGC	GGGCGGGAGC	ACCGCCCGGC	CCCCAGGAAT	GCGGCCCCGG
190	200	210	220	230	240
CCGTCGCCAT	ATTGGGTGTC	CCAACACCTG	CTGCCTCCCG	CTCTAAAAAT	AACTCCCGAC
250	260	270	280	290	300
ACCCAAATAT	GGCGACGGCC	GCCACCGCGG	TGGAGCTCGG	TACCTCCCGG	GTTATGTTAG
310	320	330	340	350	360
CTCAGTTACA	GTACCATAAG	ATACATTGAT	GAGTTTGGAC	AAACCACAAC	TAGAATGCAG
370	380	390	400	410	420
TGAAAAAAAT	GCTTTATTTG	TGAAATTTGT	GATGCTATTG	CTTTATTTGT	AACCATTATA
430	440	450	460	470	480
AGCTGCAATA	AACAAGTTAA	CAACAACAAT	TGCATTCATT	TTATGTTTCA	NGTTCANGGG
490	500	510	520	530	540
GAAGTGTNGG	AAGTTTTTTA	AAACAATTNA	AACTCCACGT	TACTTTAATA	TTACTTACTT
550	560	570	580	590	600
ATCATGGTA .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .

Fig. 23

DNASIS
4573-18g.seq

```
                10                  20                  30                  40                  50                  60
ATTTTACAAC  AGTACCGGAA  TGCCAAGCTT  GATATCGAAT  TCCTGCAGCC  CGGGGGATCC

                70                  80                  90                 100                 110                 120
ACTAGTTCTA  GAGCTTGGCG  CCTCCCGCTC  CTCCGGGTAG  CTCGTGGGCC  GCCGCCGGCC

               130                 140                 150                 160                 170                 180
CCGGAGCCTT  TTATCGAGGC  GGGCGGGAGC  ACCGCCCGGC  CCCCAGGAAT  GCGGCCCCGG

               190                 200                 210                 220                 230                 240
CCGAGGGCGG  ACGGCTCCGC  CATATTTGGG  ..........  ..........  ..........
```

Fig. 24

CA 02296771 2000-01-13

WO 99/02737

DNASIS
4573-19g.seq

10	20	30	40	50	60
ATTTTACAAC	AGTACCGGAA	TGCCAAGCTT	GATATCGAAT	TCCTGCAGCC	CGGGGGAATC
70	80	90	100	110	120
CACTAGTTCT	AGAGCTTGGC	GCCTCCCGCT	CCTCCGGGTA	GCTCGTGGGC	CGCCGCCGGC
130	140	150	160	170	180
CCCGGAGCCT	TTTATCGAGG	CGGGCGGGAG	CACCGCCCGG	CCCCCAGGAA	TGCGGCCCCG
190	200	210	220	230	240
GATGGTGGGC	AGGCAGCAGG	TGTTGGCGCT	CTAAAAATAA	CTCCCCACCA	TTCCTCACGA
250	260	270	280	290	300
CACCCAAATA	TGGCGACGGN	ACCATTCCTC	ACCCGTCCGC	CCTCGGCCGC	CACCGCGGTG
310	320	330	340	350	360
GANCTCGGTA	CCTCCCGGGT	TATGTTANCT	CAGTTACAGT	ACCATAAGAT	ACATTGATGA
370	380	390	400	410	420
NTTTGGACAA	ACCACAACTA	NAATGCAGTG	AAAAAAATGC	TTTATTTGTG	AAATTTGTGA
430	440	450	460	470	480
TGCTATTGCT	TTATTTGTNA	CCATTATAAG	CTGCAATAAA	CAANTTAACA	ACAACAATTG
490	500	510	520	530	540
CATTCATTTT	ATGTTTCANG	. . . . .	. . . . .	. . . . .	. . . . .

Fig. 25

DNASIS
4573-20g.seq

10	20	30	40	50	60
ATTTTACAAC	AGTACCGGAA	TGCCAAGCTT	GATATCGAAT	TCCTGCAGCC	CGGGGGATCC
70	80	90	100	110	120
ACTAGTTCTA	GAGCTTGGCG	CCTCCCGCTC	CTCCGGGTAG	CTCGTGGGCC	GCCGCCGGCC
130	140	150	160	170	180
CCGGAGCCTT	TTATCGAGGC	GGGCGGGAGC	ACCGCCCGGC	CCCCAGGAAT	GCGGCCCCGG
190	200	210	220	230	240
CCGTCGCCAT	ATTTGGGTGT	CCACCATTCC	TCACCCAACA	CCTGCTGCCT	GCCCCAACAC
250	260	270	280	290	300
CTGCTGCCTG	CCGGGAGTTA	TTTTTAGAGC	GCCAACACCT	GCTGCCTGCC	CCGAGGGCGG
310	320	330	340	350	360
ACGGCCGCCA	CCGCGGTGGA	GCTCGGTACC	TCCCGGGTTA	TGTTAGCTCA	GTTACAGTAC
370	380	390	400	410	420
CATAAGATAC	ATTGATGAGT	TTGGACAAAC	CACAACTAGA	ATGCAGTGAA	AAAAATGCTT
430	440	450	460	470	480
TATTTGTTGA	AATTTGTGAT	GCTATTGCTT	TATTTGTAAC	CATTATAAGC	TGCAATAAAC
490	500	510	520	530	540
AANTTAACAA	CAACAATTGC	ATTCATTTTA	TTTTCANGTT	CANGGGAAGT	GTNGGAAGTT
550	560	570	580	590	600
TTTTAAAACC	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .

Fig. 26

CA 02296771 2000-01-13

WO 99/02737

DNASIS
4573-15g.seq

10	20	30	40	50	60
ATTTTACAAC	AGTACCGGAA	TGCCAAGCTT	GATATCGAAT	TCCTGCAGCC	CGGGGGTATC
70	80	90	100	110	120
CACTAGTTCT	AGAGCTTGGC	GCCTCCCGCT	CCTCCGGGTA	GCTCGTGGGC	CGCCGCCGGC
130	140	150	160	170	180
CCCGGAGCCT	TTTATCGAGG	CGGGCGGGAG	CACCGCCCGG	CCCCCAGGAA	TGCGGCCCCG
190	200	210	220	230	240
GCCGTCCGCC	CTCGGCCGAG	GGGGACGGCG	CTCTAAAAAT	AACTCCCCCA	ACACCTGCTG
250	260	270	280	290	300
CCTGCCGGCA	GGCAGCAGGT	GTTGGGACAC	CCAAATATGG	CGACGGCCGC	CACCGCGGTG
310	320	330	340	350	360
GAGCTCGGTA	CCTCCCGGGT	TATGTTAGCT	CAGTTACAGT	ACCATAAGAT	ACATTGATGA
370	380	390	400	410	420
GTTTGGACAA	ACCACAACTA	GAATGCAGTG	AAAAAAATGC	TTTATTTGTT	GAAATTTGTG
430	440	450	460	470	480
ATGCTATTGC	TTTATTTGTA	ACCATTATAA	GCTGCAATAA	ACAAGTTAAC	AACAACAATT
490	500	510	520	530	540
GCATTCATTT	TATGTTTCAA	GTTCAAGGGG	AAGTTTTNGG	AAGTTTTTTA	AAACAAATTA
550	560	570	580	590	600
AAACTCCACT	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .

Fig. 27A

BCM15.CP1
BCM15.CP1

```
   1  AAGCTTGATA TCGACCTCCT GCANCCCGGG GGATCCACTA GTTCTAGAGC TTGGCGCCTC
  61  CCGCTCCTCC GGGTAGCTCG TGGGCCGCCG CCGGCCCCGG AGCCTTTTAT CGAGGCGGGC
 121  GGGAGCACCG CCCGGCCCCC AGGAATGCGG CCCCGGCCGT CCGCCCTCGG CCGAGGGGGA
                                                  SRE
 181  ACGGGCTCNA AAAATNACTC CCCCNACACC TGCTGCCTGC CGGCAAGNAA CAAGTTTTGG
        MEF-1
 241  GAAACCCNAA TATNGCNAAC GGCGCCACCN CNGTGGAACT CCGTNCCTCC CNGGTTATGT
 301  TAACTCNATT ACCGTNCCNT NANAANCNTT NANNAATTTG GAACAACCNC NACTAAAATN
 361  CNATNAAAAA AATNCNTTAT TTGTTAAATT TGTTAAGCNA
```

Fig. 27B

DNASIS
4585-2g.seq

10	20	30	40	50	60
ATTTTACAAC	AGTACCGGAA	TGCCAAGCTT	GATATCGAAT	TCCTGCAGCC	CGGGGGATCC
70	80	90	100	110	120
ACTAGTTCTA	GAGCTTGGCG	CCTCCCGCTC	CTCCGGGTAG	CTCGTGGGCC	GCCGCCGGCC
130	140	150	160	170	180
CCGGAGCCTT	TTATCGAGGC	GGGCGGGAGC	ACCGCCCGGC	CCCCAGGAAT	GCGGCCCCGG
190	200	210	220	230	240
CCGTCGCCAT	ATTTGGTGTC	GGGAGTTATT	TTTAGAGCGG	ACACCCAAAT	ATGGCGACGG
250	260	270	280	290	300
GGCAGGCAGC	AGGTGTTGGG	ACACCCAAAT	ATGGCGACGG	CCGCCACCGC	GGTGGAGCTC
310	320	330	340	350	360
GGTACCTCCC	GGGTTATGTT	AGCTCAGTTA	CAGTACCATA	AGATACATTG	ATGAGTTTGG
370	380	390	400	410	420
ACAAACCACA	ACTAGAAATG	CAGTTGAAAA	AAATGCTTTA	TTTGTTGAAA	TTTGTTGATG
430	440	450	460	470	480
CTATTGCTTT	ATTTGTTAAC	CCATTATAAG	CCTGCAATAA	ACAATTTAAC	AACAACAATT
490	500	510	520	530	540
GCATTCCATT	TTATNTTTCC	. . . . .	. . . . .	. . . . .	. . . . .

Fig. 28

CA 02296771 2000-01-13

WO 99/02737

DNASIS
4573-21g.seq

```
                10                  20                  30                  40                  50                  60
ATTTTACAAC  AGTACCGGAA  TGCCAAGCTT  GATATCGAAT  TCCTGCAGCC  CGGGGGATCC

                70                  80                  90                 100                 110                 120
ACTAGTTCTA  GAGCTTGGCG  CCTCCCGCTC  CTCCGGGTAG  CTCGTGGGCC  GCCGCCGGCC

               130                 140                 150                 160                 170                 180
CCGGAGCCTT  TTATCGAGGC  GGGCGGGAGC  ACCGCCCGGC  CCCCAGGAAT  GCGGCCCCGG

               190                 200                 210                 220                 230                 240
CCGTCGCCAT  ATTTGGGTGT  CGGGAGTTAT  TTTTAGAGGT  GAGGAATGGT  GCCGTCCGC .
```

Fig. 29

DNASIS
4573-22g.seq

10	20	30	40	50	60
ATTTTACAAC	AGTACCGGAA	TGCCAAGCTT	GATATCGAAT	TCCTGCAGCC	CGGGGGATCC
70	80	90	100	110	120
ACTAGTTCTA	GAGCTTGGCG	CCTCCCGCTC	CTCCGGGTAG	CTCGTGGGCC	GCCGCCGGCC
130	140	150	160	170	180
CCGGAGCCTT	TTATCGAGGC	GGGCGGGAGC	ACCGCCCGGC	CCCCAGGAAT	GCGGCCCCGG
190	200	210	220	230	240
CCGTCGCCAT	ATTTGGGTGT	CCCGTCGCCA	TATTTGGGTG	TCGGGAGTTA	TTTTTAGAGC
250	260	270	280	290	300
GGACACCCAA	ATATGGCGAC	GGCCGCCACC	GCGGTGGAGC	TCGGTACCTC	CCGGGTTATG
310	320	330	340	350	360
TTAGCTCAGT	TACAGTACCA	TAAGATACAT	TGATGAGTTT	GGACAAACCA	CAACTANAAT
370	380	390	400	410	420
GCAGTGAAAA	AAATGCTTTA	TTTGTGAAAT	TTGTGATGCT	ATTGCTTTAT	TTGTAACCAT
430	440	450	460	470	480
TATAAGCTGC	AATAAACAAG	TTAACAACAA	CAATTGCATT	CATTTTATGT	TTCANGTTCA
490	500	510	520	530	540
AGGGGAAGTG	TTNGAAGTTT	. . . . .	. . . . .	. . . . .	. . . . .

Fig. 30

CA 02296771 2000-01-13

WO 99/02737

SUBSTITUTE SHEET (RULE 26)

10	20	30	40	50	60
ATTTTACAAC	AGTACGGAAT	GCCAAGCTTG	ATATCGAATT	CCTGCAGCCC	GGGGGATCCA
70	80	90	100	110	120
CTAGTTCTAG	AGCTTGGCGC	CTCCCGCTCC	TCCGGGTAGC	TCGTGGGCCG	CCGCCGGCCC
130	140	150	160	170	180
CGGAGCCTTT	TATCGAGGCG	GGCGGGAGCA	CCGCCCGGCC	CCCAGGAATG	CGGCCCCGGC
190	200	210	220	230	240
CGTCGCCATA	TTTGGTGTCG	ACACCCAAAT	ATGGCGACGG	GGCAGGCAGC	AGGTGTTGGG
250	260	270	280	290	300
ACACCCAAAT	ATGGCGACGG	GTGAGGAATG	GTGGGGAGTT	ATTTTTAGAG	CGGACACCCA
310	320	330	340	350	360
AATATGGCGA	CGGCCGCCAC	CGCGGTGGAG	CTCGGTACCT	CCCGGGTTAT	GTTAGCTCAG
370	380	390	400	410	420
TTACAGTACC	ATAAGATACA	TTGATGAGTT	TGGACAAACC	ACAACTAGAA	TGCAGTGAAA
430	440	450	460	470	480
AAAATGCTTT	ATTTGTTGAA	ATTTGTGATG	CTATTGCTTT	ATTTGTAACC	ATTATAAGCT
490	500	510	520	530	540
GCAATAAACA	ATTTAACAAC	AACAATTGCA	TTCATTTTAT	GTTTCANGTT	CCAGGGGAAG
550	560	570	580	590	600
TTTTTGGAAG	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .

Fig. 31