



(12) **DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION**

(13) **A1**

(22) Date de dépôt/Filing Date: 2005/05/18
(41) Mise à la disp. pub./Open to Public Insp.: 2005/12/01
(62) Demande originale/Original Application: 2 563 666
(30) Priorité/Priority: 2004/05/18 (DE10 2004 024 617.3)

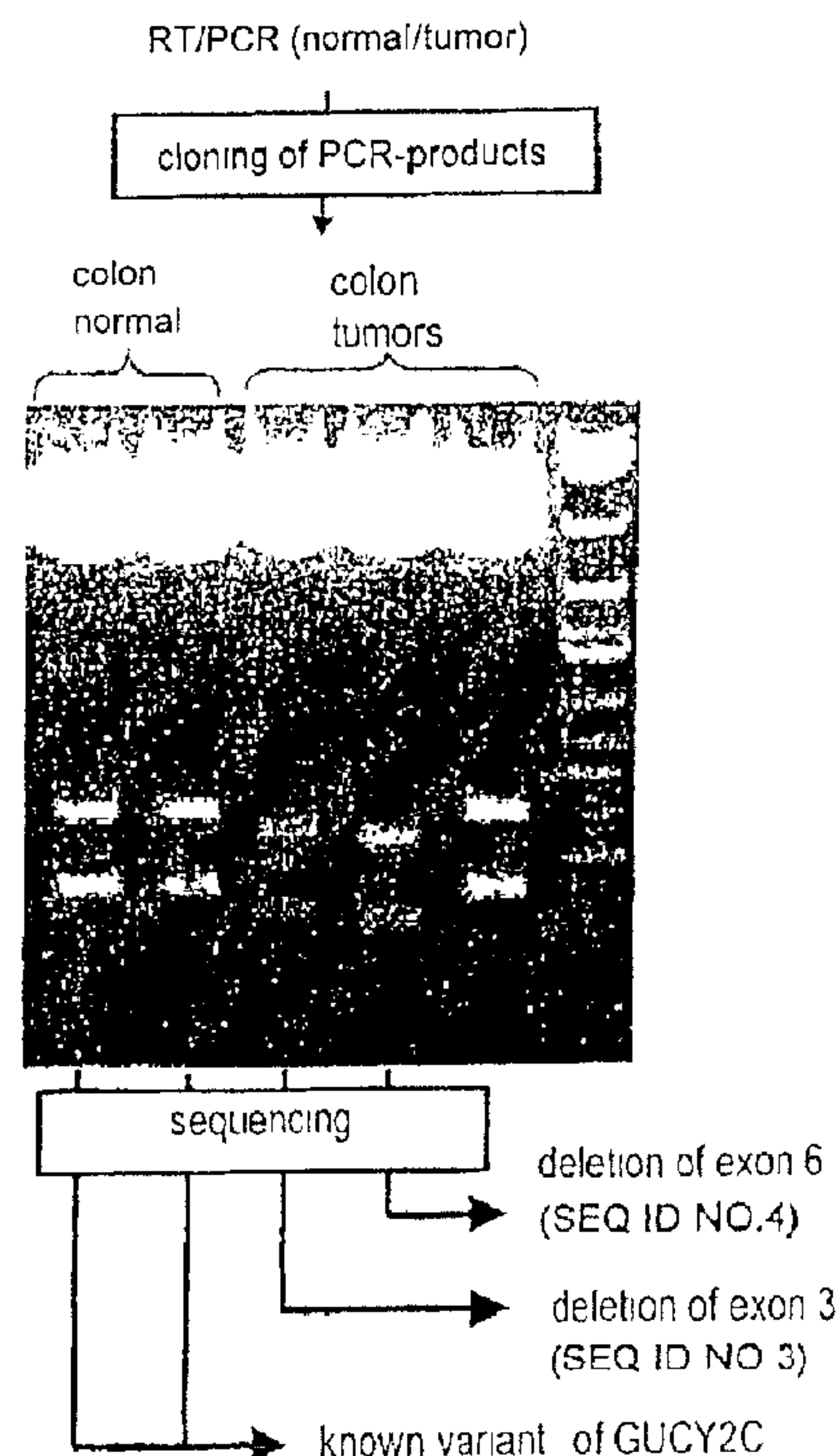
(51) Cl.Int./Int.Cl. *C12N 15/12* (2006.01),
A61K 31/7088 (2006.01), *A61K 39/00* (2006.01),
A61K 39/395 (2006.01), *A61K 47/66* (2017.01),
A61K 47/68 (2017.01), *A61K 49/00* (2006.01),
A61P 35/00 (2006.01), *C07K 14/705* (2006.01),
C07K 16/28 (2006.01), *C07K 16/30* (2006.01),
C12N 15/113 (2010.01), ...

(71) Demandeur/Applicant:
GANYMED PHARMACEUTICALS AG, DE

(72) Inventeurs/Inventors:
TURECI, OZLEM, DE;
SAHIN, UGUR, DE;
KOSLOWSKI, MICHAEL, DE;

(54) Titre : VARIANTES DE LA PROTEINE CLAUDIN-18 D'EXPRESSION DIFFERENTIELLE DANS LE TUMEURS ET
LEUR UTILISATION

(54) Title: DIFFERENTIALLY EXPRESSED CLAUDIN-18 VARIANTS IN TUMORS AND USE THEREOF



(57) Abrégé/Abstract:

The invention relates to the therapy and diagnosis of diseases wherein gene products expressed in a tumor-associated manner are aberrantly expressed. Specifically disclosed are antibodies which bind specifically to certain regions of the tumor-associated

(51) Cl.Int./Int.Cl. (suite/continued) *C12N 15/63* (2006.01), *C12Q 1/6809* (2018.01), *C12Q 1/6886* (2018.01),
G01N 33/48 (2006.01), *G01N 33/574* (2006.01)

(72) Inventeurs(suite)/Inventors(continued): FRITZ, STEFAN, DE; GEPPERT, HARALD-GERHARD, DE...

(74) Agent: SMART & BIGGAR

(57) Abrégé(suite)/Abstract(continued):

antigen Claudin-18, such as the regions defined by SEQ ID NOs: 142-148. The antibody binds selectively to an epitope of a glycosylation variant of Claudin-18, which epitope is non-glycosylated in Claudin-18 when expressed in tumor tissue and glycosylated in Claudin-18 when expressed in normal tissue, said glycosylation in normal tissue preventing binding of the antibody. Such antibodies may be used in the diagnosis of cancer characterized by the expression of the Claudin-18 variant.

ABSTRACT

The invention relates to the therapy and diagnosis of diseases wherein gene products expressed in a tumor-associated manner are aberrantly expressed. Specifically
5 disclosed are antibodies which bind specifically to certain regions of the tumor-associated antigen Claudin-18, such as the regions defined by SEQ ID NOs: 142-148. The antibody binds selectively to an epitope of a glycosylation variant of Claudin-18, which epitope is
10 non-glycosylated in Claudin-18 when expressed in tumor tissue and glycosylated in Claudin-18 when expressed in normal tissue, said glycosylation in normal tissue preventing binding of the antibody. Such antibodies may be used in the diagnosis of cancer characterized by the
15 expression of the Claudin-18 variant.

SMART & BIGGAR

Intellectual Property & Technology Law

PATENT RULES

SECTION 104(4) NOTICE

It is the applicant's wish that, until either a patent has issued on the basis of the application or the application is refused, or is abandoned and no longer subject to reinstatement, or is withdrawn, the Commissioner only authorize the furnishing of a sample of any deposited biological material referred to in the specification to an independent expert nominated by the Commissioner in accordance with section 109 of the Patent Rules.

**DIFFERENTIALLY EXPRESSED CLAUDIN-18 VARIANTS IN TUMORS
AND USE THEREOF**

5 This application is a divisional of Canadian Patent Application No. 2,563,666.

Despite interdisciplinary approaches and exhaustive use of classical therapeutic procedures, cancers are still
10 among the leading causes of death. More recent therapeutic concepts aim at incorporating the patient's immune system into the overall therapeutic concept by using recombinant tumor vaccines and other specific measures such as antibody therapy. A prerequisite for
15 the success of such a strategy is the recognition of tumor-specific or tumor-associated antigens or epitopes by the patient's immune system whose effector functions are to be interventionally enhanced. Tumor cells biologically differ substantially from their
20 nonmalignant cells of origin. These differences are due to genetic alterations acquired during tumor development and result, inter alia, also in the formation of qualitatively or quantitatively altered molecular structures in the cancer cells. Tumor-
25 associated structures of this kind which are recognized by the specific immune system of the tumor-harboring host are referred to as tumor-associated antigens. The specific recognition of tumor-associated antigens involves cellular and humoral mechanisms which are two
30 functionally interconnected units: CD4⁺ and CD8⁺ T lymphocytes recognize the processed antigens presented on the molecules of the MHC (major histocompatibility complex) classes II and I, respectively, while B lymphocytes produce circulating antibody molecules
35 which bind directly to unprocessed antigens. The potential clinical-therapeutical importance of tumor-associated antigens results from the fact that the recognition of antigens on neoplastic cells by the

wrong target structures.

It was the object of the present invention to provide target structures for a diagnosis and therapy of
5 cancers.

According to the invention, this object is achieved by the subject matter of the claims.

10 According to the invention, a strategy for identifying and providing antigens expressed in association with a tumor and the nucleic acids coding therefor was pursued. This strategy is based on the fact that particular genes which are expressed in an organ
15 specific manner, e.g. exclusively in colon, lung or kidney tissue, are reactivated also in tumor cells of the respective organs and moreover in tumor cells of other tissues in an ectopic and forbidden manner. First, data mining produces a list as complete as
20 possible of all known organ-specific genes which are then evaluated for their aberrant activation in different tumors by expression analyses by means of specific RT-PCR. Data mining is a known method of identifying tumor-associated genes. In the conventional
25 strategies, however, transcriptoms of normal tissue libraries are usually subtracted electronically from tumor tissue libraries, with the assumption that the remaining genes are tumor-specific (Schmitt et al., *Nucleic Acids Res.* 27:4251-60, 1999; Vasmatazsis et al.,
30 *Proc. Natl. Acad. Sci. USA.* 95:300-4, 1998; Scheurle et al., *Cancer Res.* 60:4037-43, 2000).

The concept of the invention, which has proved much more successful, however, is based on utilizing data
35 mining for electronically extracting all organ-specific genes and then evaluating said genes for expression in tumors.

The invention thus relates in one aspect to a strategy for identifying tissue-specific genes differentially expressed in tumors. Said strategy combines data mining of public sequence libraries ("*in silico*") with
5 subsequent evaluating laboratory-experimental ("wet bench") studies.

According to the invention, a combined strategy based on two different bioinformatic scripts enabled new
10 tumor genes to be identified. These have previously been classified as being purely organ-specific. The finding that these genes are aberrantly activated in tumor cells allows them to be assigned a substantially new quality with functional implications. According to
15 the invention, these tumor-associated genes and the genetic products encoded thereby were identified and provided independently of an immunogenic action.

The tumor-associated antigens identified according to
20 the invention have an amino acid sequence encoded by a nucleic acid which is selected from the group consisting of (a) a nucleic acid which comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 1-8, 41-44, 51-59, 84, 117,
25 119 and 138, a part or derivative thereof, (b) a nucleic acid which hybridizes with the nucleic acid of (a) under stringent conditions, (c) a nucleic acid which is degenerate with respect to the nucleic acid of (a) or (b), and (d) a nucleic acid which is
30 complementary to the nucleic acid of (a), (b) or (c). In a preferred embodiment, a tumor-associated antigen identified according to the invention has an amino acid sequence encoded by a nucleic acid which is selected from the group consisting of SEQ ID NOs: 1-8, 41-44,
35 51-59, 84, 117, 119 and 138. In a further preferred embodiment, a tumor-associated antigen identified according to the invention comprises an amino acid sequence selected from the group consisting of SEQ ID

NOs: 9-19, 45-48, 60-66, 85, 90-97, 100-102, 105, 106, 111-116, 118, 120, 123, 124, 135-137, 139, and 142-150, a part or derivative thereof.

5 The present invention generally relates to the use of tumor-associated antigens identified according to the invention or of parts or derivatives thereof, of nucleic acids coding therefor or of nucleic acids directed against said coding nucleic acids or of
10 antibodies directed against the tumor-associated antigens identified according to the invention or parts or derivatives thereof for therapy and diagnosis. This utilization may relate to individual but also to combinations of two or more of these antigens,
15 functional fragments, nucleic acids, antibodies, etc., in one embodiment also in combination with other tumor-associated genes and antigens for diagnosis, therapy and progress control.

20 Preferred diseases for a therapy and/or diagnosis are those in which one or more of the tumor-associated antigens identified according to the invention are selectively expressed or abnormally expressed.

25 The invention also relates to nucleic acids and genetic products which are expressed in association with a tumor cell.

Furthermore, the invention relates to genetic products,
30 i.e. nucleic acids and proteins or peptides, which are produced by altered splicing (splice variants) of known genes or altered translation using alternative open reading frames. In this aspect the invention relates to nucleic acids which comprise a nucleic acid sequence
35 selected from the group consisting of sequences according to SEQ ID NOs: 3-5 of the sequence listing. Moreover, in this aspect, the invention relates to proteins or peptides which comprise an amino acid

sequence selected from the group consisting of the sequences according to SEQ ID NOs: 10 and 12-14 of the sequence listing. The splice variants of the invention can be used according to the invention as targets for
5 diagnosis and therapy of tumor diseases.

In particular, the invention relates to the amino acid sequence according to SEQ ID NO: 10 of the sequence listing which is encoded by an alternative open reading
10 frame identified according to the invention and differs from the previously described protein sequence (SEQ ID NO: 9) in additional 85 amino acids at the N terminus of the protein.

- 15 Very different mechanisms may cause splice variants to be produced, for example
- utilization of variable transcription initiation sites
 - utilization of additional exons
 - 20 - complete or incomplete splicing out of single or two or more exons,
 - splice regulator sequences altered via mutation (deletion or generation of new donor/acceptor sequences),
 - 25 - incomplete elimination of intron sequences.

Altered splicing of a gene results in an altered transcript sequence (splice variant). Translation of a splice variant in the region of its altered sequence
30 results in an altered protein which may be distinctly different in the structure and function from the original protein. Tumor-associated splice variants may produce tumor-associated transcripts and tumor-associated proteins/antigens. These may be utilized as
35 molecular markers both for detecting tumor cells and for therapeutic targeting of tumors. Detection of tumor cells, for example in blood, serum, bone marrow, sputum, bronchial lavage, bodily secretions and tissue

biopsies, may be carried out according to the invention, for example, after extraction of nucleic acids by PCR amplification with splice variant-specific oligonucleotides. In particular, pairs of primers are
5 suitable as oligonucleotides at least one of which binds to the region of the splice variant which is tumor-associated under stringent conditions. According to the invention, oligonucleotides described for this purpose in the examples are suitable, in particular
10 oligonucleotides which have or comprise a sequence selected from SEQ ID NOs: 34-36, 39, 40, and 107-110 of the sequence listing. According to the invention, all sequence-dependent detection systems are suitable for detection. These are, apart from PCR, for example gene
15 chip/microarray systems, Northern blot, RNase protection assays (RDA) and others. All detection systems have in common that detection is based on a specific hybridization with at least one splice variant-specific nucleic acid sequence. However, tumor
20 cells may also be detected according to the invention by antibodies which recognize a specific epitope encoded by the splice variant. Said antibodies may be prepared by using for immunization peptides which are specific for said splice variant. In this aspect, the
25 invention relates, in particular, to peptides which have or comprise a sequence selected from SEQ ID NOs: 17-19, 111-115, 120, and 137 of the sequence listing and specific antibodies which are directed thereto.

30 Tumor cells can also be detected by using antibodies which recognize glycosylation variants which are modified in a tumor specific manner. Suitable for the generation of such antibodies are peptide regions which differ between tumor cells and healthy cells with
35 respect to glycosylation. In this aspect, the invention relates, in particular, to peptides which have or comprise a sequence selected from SEQ ID NOs: 17-19, 111-115, 120, 137 and 142-145 of the sequence listing

and specific antibodies which are directed thereto. Asparagin is transformed into aspartic acid by endogenous deglycosylation of N coupled sugar residues. According to the invention, the proteins described
5 herein can be modified with respect to their sequences in a tumor specific manner and, thus, have different biochemical and antibody binding properties. In this aspect, the invention relates, in particular, to peptides which have or comprise a sequence selected
10 from SEQ ID NOs: 146-150 of the sequence listing and specific antibodies which are directed thereto.

Suitable for immunization are particularly the amino acids whose epitopes are distinctly different from the
15 variant(s) of the genetic product, which is (are) preferably produced in healthy cells. Detection of the tumor cells with antibodies may be carried out here on a sample isolated from the patient or as imaging with intravenously administered antibodies. In addition to
20 diagnostic usability, splice variants having new or altered epitopes are attractive targets for immunotherapy. The epitopes of the invention may be utilized for targeting therapeutically active monoclonal antibodies or T lymphocytes. In passive
25 immunotherapy, antibodies or T lymphocytes which recognize splice variant-specific epitopes are adoptively transferred here. As in the case of other antigens, antibodies may be generated also by using standard technologies (immunization of animals, panning
30 strategies for isolation of recombinant antibodies) with utilization of polypeptides which include these epitopes. Alternatively, it is possible to utilize for immunization nucleic acids coding for oligo- or polypeptides which contain said epitopes. Various
35 techniques for in vitro or in vivo generation of epitope-specific T lymphocytes are known and have been described in detail (for example Kessler JH, et al. 2001, Sahin et al., 1997) and are likewise based on

utilizing oligo- or polypeptides which contain the splice variant-specific epitopes or nucleic acids coding for said oligo- or polypeptides. Oligo- or polypeptides which contain the splice variant-specific epitopes or nucleic acids coding for said polypeptides may also be used as pharmaceutically active substances in active immunotherapy (vaccination, vaccine therapy).

The present invention also describes proteins which differ in nature and degree of their secondary modifications in normal and tumor tissue (for example Durand & Seta, 2000; Clin. Chem. 46: 795-805; Hakomori, 1996; Cancer Res. 56: 5309-18).

The analysis of protein modifications can be done in Western blots. In particular, glycosylations which as a rule have a size of several kDa result in a higher overall mass of the target protein which can be separated in an SDS-PAGE. For the detection of specific O- and N-glycosidic bonds protein lysates are incubated with O- or N-glycosylases (according to the instructions of the respective manufactures, for example, PNGase, endoglycosidase F, endoglycosidase H, Roche Diagnostics) prior to denaturation using SDS. Thereafter, a Western blot is performed. If the size of target protein is reduced a specific glycosylation can be detected in this manner following incubation with a glycosidase and thus, also the tumor specificity of a modification can be analyzed. Protein regions which are differentially glycosylated in tumor cells and healthy cells are of particular interest. Such differences in glycosylation, however, have hitherto only been described for a few cell surface proteins (for example, Mucl).

According to the invention, it was possible to detect a differential glycosylation for Claudin-18 in tumors. Gastrointestinal carcinomas, pancreas carcinomas,

esophagus tumors, prostate tumors as well as lung tumors have a form of Claudin-18 which is glycosylated at a lower level. Glycosylation in healthy tissues masks protein epitopes of Claudin-18 which are not covered on tumor cells due to lacking glycosylation. Correspondingly it is possible according to the invention to select ligands and antibodies which bind to these domains. Such ligands and antibodies according to the invention do not bind to Claudin-18 on healthy cells since here the epitops are covered due to glycosylation.

As has been described above for protein epitopes which are derived from tumor-associated splice variants it is thus possible to use the differential glycosylation to distinguish normal cells and tumor cells with diagnostic as well as therapeutic intention.

In one aspect, the invention relates to a pharmaceutical composition comprising an agent which recognizes the tumor-associated antigen identified according to the invention and which is preferably selective for cells which have expression or abnormal expression of a tumor-associated antigen identified according to the invention. In particular embodiments, said agent may cause induction of cell death, reduction in cell growth, damage to the cell membrane or secretion of cytokines and preferably have a tumor-inhibiting activity. In one embodiment, the agent is an antisense nucleic acid which hybridizes selectively with the nucleic acid coding for the tumor-associated antigen. In a further embodiment, the agent is an antibody which binds selectively to the tumor-associated antigen, in particular a complement-activated or toxin conjugated antibody which binds selectively to the tumor-associated antigen. In a further embodiment, the agent comprises two or more agents which each selectively recognize different

tumor-associated antigens, at least one of which is a tumor-associated antigen identified according to the invention. Recognition needs not be accompanied directly with inhibition of activity or expression of the antigen. In this aspect of the invention, the antigen selectively limited to tumors preferably serves as a label for recruiting effector mechanisms to this specific location. In a preferred embodiment, the agent is a cytotoxic T lymphocyte which recognizes the antigen on an HLA molecule and lyses the cells labeled in this way. In a further embodiment, the agent is an antibody which binds selectively to the tumor-associated antigen and thus recruits natural or artificial effector mechanisms to said cell. In a further embodiment, the agent is a T helper lymphocyte which enhances effector functions of other cells specifically recognizing said antigen.

In one aspect, the invention relates to a pharmaceutical composition comprising an agent which inhibits expression or activity of a tumor-associated antigen identified according to the invention. In a preferred embodiment, the agent is an antisense nucleic acid which hybridizes selectively with the nucleic acid coding for the tumor-associated antigen. In a further embodiment, the agent is an antibody which binds selectively to the tumor-associated antigen. In a further embodiment, the agent comprises two or more agents which each selectively inhibit expression or activity of different tumor-associated antigens, at least one of which is a tumor-associated antigen identified according to the invention.

The invention furthermore relates to a pharmaceutical composition which comprises an agent which, when administered, selectively increases the amount of complexes between an HLA molecule and a peptide epitope from the tumor-associated antigen identified according

to the invention. In one embodiment, the agent comprises one or more components selected from the group consisting of (i) the tumor-associated antigen or a part thereof, (ii) a nucleic acid which codes for
5 said tumor-associated antigen or a part thereof, (iii) a host cell which expresses said tumor-associated antigen or a part thereof, and (iv) isolated complexes between peptide epitopes from said tumor-associated antigen and an MHC molecule. In one embodiment, the
10 agent comprises two or more agents which each selectively increase the amount of complexes between MHC molecules and peptide epitopes of different tumor-associated antigens, at least one of which is a tumor-associated antigen identified according to the
15 invention.

The invention furthermore relates to a pharmaceutical composition which comprises one or more components selected from the group consisting of (i) a tumor-associated antigen identified according to the
20 invention or a part thereof, (ii) a nucleic acid which codes for a tumor-associated antigen identified according to the invention or for a part thereof, (iii) an antibody which binds to a tumor-associated antigen
25 identified according to the invention or to a part thereof, (iv) an antisense nucleic acid which hybridizes specifically with a nucleic acid coding for a tumor-associated antigen identified according to the invention, (v) a host cell which expresses a tumor-associated antigen identified according to the
30 invention or a part thereof, and (vi) isolated complexes between a tumor-associated antigen identified according to the invention or a part thereof and an HLA molecule.

35

A nucleic acid coding for a tumor-associated antigen identified according to the invention or for a part thereof may be present in the pharmaceutical

composition in an expression vector and functionally linked to a promoter.

5 A host cell present in a pharmaceutical composition of the invention may secrete the tumor-associated antigen or the part thereof, express it on the surface or may additionally express an HLA molecule which binds to said tumor-associated antigen or said part thereof. In one embodiment, the host cell expresses the HLA
10 molecule endogenously. In a further embodiment, the host cell expresses the HLA molecule and/or the tumor-associated antigen or the part thereof in a recombinant manner. The host cell is preferably nonproliferative. In a preferred embodiment, the host cell is an antigen-
15 presenting cell, in particular a dendritic cell, a monocyte or a macrophage.

An antibody present in a pharmaceutical composition of the invention may be a monoclonal antibody. In further
20 embodiments, the antibody is a chimeric or humanized antibody, a fragment of a natural antibody or a synthetic antibody, all of which may be produced by combinatory techniques. The antibody may be coupled to a therapeutically or diagnostically useful agent.

25 An antisense nucleic acid present in a pharmaceutical composition of the invention may comprise a sequence of 6-50, in particular 10-30, 15-30 and 20-30, contiguous nucleotides of the nucleic acid coding for the tumor-
30 associated antigen identified according to the invention.

In further embodiments, a tumor-associated antigen, provided by a pharmaceutical composition of the
35 invention either directly or via expression of a nucleic acid, or a part thereof binds to MHC molecules on the surface of cells, said binding preferably causing a cytolytic response and/or inducing cytokine

release.

A pharmaceutical composition of the invention may comprise a pharmaceutically compatible carrier and/or
5 an adjuvant. The adjuvant may be selected from saponin, GM-CSF, CpG nucleotides, RNA, a cytokine or a chemokine. A pharmaceutical composition of the invention is preferably used for the treatment of a
10 disease characterized by selective expression or abnormal expression of a tumor-associated antigen. In a preferred embodiment, the disease is cancer.

The invention furthermore relates to methods of treating, diagnosing and/or monitoring a disease
15 characterized by expression or abnormal expression of one of more tumor-associated antigens. In one embodiment, the treatment comprises administering a pharmaceutical composition of the invention.

20 Preferably, the disease is cancer wherein the term "cancer" comprises but is not limited to leukemias, seminomas, melanomas, teratomas, gliomas, renal, adrenal, thyroid, intestinal, liver, colon, stomach, gastrointestinal, lymph node, esophageal, colorectal,
25 pancreatic, ear, nose and throat (ENT), breast, prostate, uterus, ovarian and lung cancer and the metastases thereof.

In one aspect, the invention relates to a method of
30 diagnosing a disease characterized by expression or abnormal expression of a tumor-associated antigen identified according to the invention. The method comprises detection of (i) a nucleic acid which codes for the tumor-associated antigen or of a part thereof
35 and/or (ii) detection of the tumor-associated antigen or of a part thereof, and/or (iii) detection of an antibody to the tumor-associated antigen or to a part thereof and/or (iv) detection of cytotoxic or T helper

lymphocytes which are specific for the tumor-associated antigen or for a part thereof in a biological sample isolated from a patient. In particular embodiments, detection comprises (i) contacting the biological sample with an agent which binds specifically to the nucleic acid coding for the tumor-associated antigen or to the part thereof, to said tumor-associated antigen or said part thereof, to the antibody or to cytotoxic or T helper lymphocytes specific for the tumor-associated antigen or parts thereof, and (ii) detecting the formation of a complex between the agent and the nucleic acid or the part thereof, the tumor-associated antigen or the part thereof, the antibody or the cytotoxic or T helper lymphocytes. In one embodiment, the disease is characterized by expression or abnormal expression of two or more different tumor-associated antigens and detection comprises detection of two or more nucleic acids coding for said two or more different tumor-associated antigens or of parts thereof, detection of two or more different tumor-associated antigens or of parts thereof, detection of two or more antibodies binding to said two or more different tumor-associated antigens or to parts thereof or detection of two or more cytotoxic or T helper lymphocytes specific for said two or more different tumor-associated antigens. In a further embodiment, the biological sample isolated from the patient is compared to a comparable normal biological sample.

In a further aspect, the invention relates to a method for determining regression, course or onset of a disease characterized by expression or abnormal expression of a tumor-associated antigen identified according to the invention, which method comprises monitoring a sample from a patient who has said disease or is suspected of falling ill with said disease, with respect to one or more parameters selected from the group consisting of (i) the amount of nucleic acid

which codes for the tumor-associated antigen or of a part thereof, (ii) the amount of the tumor-associated antigen or a part thereof, (iii) the amount of antibodies which bind to the tumor-associated antigen or to a part thereof, and (iv) the amount of cytolytic T cells or T helper cells which are specific for a complex between the tumor-associated antigen or a part thereof and an MHC molecule. The method preferably comprises determining the parameter(s) in a first sample at a first point in time and in a further sample at a second point in time and in which the course of the disease is determined by comparing the two samples. In particular embodiments, the disease is characterized by expression or abnormal expression of two or more different tumor-associated antigens and monitoring comprises monitoring (i) the amount of two or more nucleic acids which code for said two or more different tumor-associated antigens or of parts thereof, and/or (ii) the amount of said two or more different tumor-associated antigens or of parts thereof, and/or (iii) the amount of two or more antibodies which bind to said two or more different tumor-associated antigens or to parts thereof, and/or (iv) the amount of two or more cytolytic T cells or of T helper cells which are specific for complexes between said two or more different tumor-associated antigens or of parts thereof and MHC molecules.

According to the invention, detection of a nucleic acid or of a part thereof or monitoring the amount of a nucleic acid or of a part thereof may be carried out using a polynucleotide probe which hybridizes specifically to said nucleic acid or said part thereof or may be carried out by selective amplification of said nucleic acid or said part thereof. In one embodiment, the polynucleotide probe comprises a sequence of 6-50, in particular 10-30, 15-30 and 20-30, contiguous nucleotides of said nucleic acid.

In particular embodiments, the tumor-associated antigen to be detected or the part thereof is present intracellularly or on the cell surface. According to
5 the invention, detection of a tumor-associated antigen or of a part thereof or monitoring the amount of a tumor-associated antigen or of a part thereof may be carried out using an antibody binding specifically to said tumor-associated antigen or said part thereof.

10

In further embodiments, the tumor-associated antigen to be detected or the part thereof is present in a complex with an MHC molecule, in particular an HLA molecule.

15 According to the invention, detection of an antibody or monitoring the amount of antibodies may be carried out using a protein or peptide binding specifically to said antibody.

20 According to the invention, detection of cytolytic T cells or of T helper cells or monitoring the amount of cytolytic T cells or of T helper cells which are specific for complexes between an antigen or a part thereof and MHC molecules may be carried out using a
25 cell presenting the complex between said antigen or said part thereof and an MHC molecule.

The polynucleotide probe, the antibody, the protein or peptide or the cell, which is used for detection or
30 monitoring, is preferably labeled in a detectable manner. In particular embodiments, the detectable marker is a radioactive marker or an enzymic marker. T lymphocytes may additionally be detected by detecting their proliferation, their cytokine production, and
35 their cytotoxic activity triggered by specific stimulation with the complex of MHC and tumor-associated antigen or parts thereof. T lymphocytes may also be detected via a recombinant MHC molecule or else

a complex of two or more MHC molecules which are loaded with the particular immunogenic fragment of one or more of the tumor-associated antigens and which can identify the specific T lymphocytes by contacting the specific T
5 cell receptor.

In a further aspect, the invention relates to a method of treating, diagnosing or monitoring a disease characterized by expression or abnormal expression of a
10 tumor-associated antigen identified according to the invention, which method comprises administering an antibody which binds to said tumor-associated antigen or to a part thereof and which is coupled to a therapeutic or diagnostic agent. The antibody may be a
15 monoclonal antibody. In further embodiments, the antibody is a chimeric or humanized antibody or a fragment of a natural antibody.

In certain embodiments, the methods of diagnosing or
20 monitoring a disease characterized by expression or abnormal expression of a tumor associated antigen identified according to the invention are performed with aid of or by means of detecting disseminating tumor cells or tumor metastases. Disseminating tumor
25 cells can be detected, for example, in blood, serum, bone marrow, sputum, bronchial aspirate and/or bronchial lavage.

The invention also relates to a method of treating a
30 patient having a disease characterized by expression or abnormal expression of a tumor-associated antigen identified according to the invention, which method comprises (i) removing a sample containing immunoreactive cells from said patient, (ii) contacting
35 said sample with a host cell expressing said tumor-associated antigen or a part thereof, under conditions which favor production of cytolytic T cells against said tumor-associated antigen or a part thereof, and

(iii) introducing the cytolytic T cells into the patient in an amount suitable for lysing cells expressing the tumor-associated antigen or a part thereof. The invention likewise relates to cloning the
5 T cell receptor of cytolytic T cells against the tumor-associated antigen. Said receptor may be transferred to other T cells which thus receive the desired specificity and, as under (iii), may be introduced into the patient.

10

In one embodiment, the host cell endogenously expresses an HLA molecule. In a further embodiment, the host cell recombinantly expresses an HLA molecule and/or the tumor-associated antigen or the part thereof. The host
15 cell is preferably nonproliferative. In a preferred embodiment, the host cell is an antigen-presenting cell, in particular a dendritic cell, a monocyte or a macrophage.

20 In a further aspect, the invention relates to a method of treating a patient having a disease characterized by expression or abnormal expression of a tumor-associated antigen, which method comprises (i) identifying a nucleic acid which codes for a tumor-associated antigen
25 identified according to the invention and which is expressed by cells associated with said disease, (ii) transfecting a host cell with said nucleic acid or a part thereof, (iii) culturing the transfected host cell for expression of said nucleic acid (this is not
30 obligatory when a high rate of transfection is obtained), and (iv) introducing the host cells or an extract thereof into the patient in an amount suitable for increasing the immune response to the patient's cells associated with the disease. The method may
35 further comprise identifying an MHC molecule presenting the tumor-associated antigen or a part thereof, with the host cell expressing the identified MHC molecule and presenting said tumor-associated antigen or a part

thereof. The immune response may comprise a B cell response or a T cell response. Furthermore, a T cell response may comprise production of cytolytic T cells and/or T helper cells which are specific for the host
5 cells presenting the tumor-associated antigen or a part thereof or specific for cells of the patient which express said tumor-associated antigen or a part thereof.

10 The invention also relates to a method of treating a disease characterized by expression or abnormal expression of a tumor-associated antigen identified according to the invention, which method comprises (i) identifying cells from the patient which express
15 abnormal amounts of the tumor-associated antigen, (ii) isolating a sample of said cells, (iii) culturing said cells, and (iv) introducing said cells into the patient in an amount suitable for triggering an immune response to the cells.

20

Preferably, the host cells used according to the invention are nonproliferative or are rendered nonproliferative. A disease characterized by expression or abnormal expression of a tumor-associated antigen is
25 in particular cancer.

The present invention furthermore relates to a nucleic acid selected from the group consisting of (a) a nucleic acid which comprises a nucleic acid sequence
30 selected from the group consisting of SEQ ID NOs: 3-5, a part or derivative thereof, (b) a nucleic acid which hybridizes with the nucleic acid of (a) under stringent conditions, (c) a nucleic acid which is degenerate with respect to the nucleic acid of (a) or (b), and (d) a
35 nucleic acid which is complementary to the nucleic acid of (a), (b) or (c). The invention furthermore relates to a nucleic acid, which codes for a protein or polypeptide comprising an amino acid sequence selected

from the group consisting of SEQ ID NOs: 10, 12-14, and 146-150, a part or derivative thereof.

5 In a further aspect, the invention relates to promoter sequences of nucleic acids of the invention. These sequences may be functionally linked to another gene, preferably in an expression vector, and thus ensure selective expression of said gene in appropriate cells.

10 In a further aspect, the invention relates to a recombinant nucleic acid molecule, in particular DNA or RNA molecule, which comprises a nucleic acid of the invention.

15 The invention also relates to host cells which contain a nucleic acid of the invention or a recombinant nucleic acid molecule comprising a nucleic acid of the invention.

20 The host cell may also comprise a nucleic acid coding for a HLA molecule. In one embodiment, the host cell endogenously expresses the HLA molecule. In a further embodiment, the host cell recombinantly expresses the HLA molecule and/or the nucleic acid of the invention
25 or a part thereof. Preferably, the host cell is nonproliferative. In a preferred embodiment, the host cell is an antigen-presenting cell, in particular a dendritic cell, a monocyte or a macrophage.

30 In a further embodiment, the invention relates to oligonucleotides which hybridize with a nucleic acid identified according to the invention and which may be used as genetic probes or as "antisense" molecules. Nucleic acid molecules in the form of oligonucleotide
35 primers or competent probes, which hybridize with a nucleic acid identified according to the invention or parts thereof, may be used for finding nucleic acids which are homologous to said nucleic acid identified

according to the invention. PCR amplification, Southern and Northern hybridization may be employed for finding homologous nucleic acids. Hybridization may be carried out under low stringency, more preferably under medium
5 stringency and most preferably under high stringency conditions. The term "stringent conditions" according to the invention refers to conditions which allow specific hybridization between polynucleotides.

10 In a further aspect, the invention relates to a protein, polypeptide or peptide which is encoded by a nucleic acid selected from the group consisting of (a) a nucleic acid which comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 3-5,
15 a part or derivative thereof, (b) a nucleic acid which hybridizes with the nucleic acid of (a) under stringent conditions, (c) a nucleic acid which is degenerate with respect to the nucleic acid of (a) or (b), and (d) a nucleic acid which is complementary to the nucleic acid
20 of (a), (b) or (c). In a preferred embodiment, the invention relates to a protein or polypeptide or peptide which comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 10, 12-14, and 146-150, a part or derivative thereof.

25 In a further aspect, the invention relates to an immunogenic fragment of a tumor-associated antigen identified according to the invention. Said fragment preferably binds to a human HLA receptor or to a human
30 antibody. A fragment of the invention preferably comprises a sequence of at least 6, in particular at least 8, at least 10, at least 12, at least 15, at least 20, at least 30 or at least 50, amino acids.

35 In this aspect the invention relates, in particular, to a peptide which has or comprises a sequence selected from the group consisting of SEQ ID NOs: 17-19, 90-97, 100-102, 105, 106, 111-116, 120, 123, 124, 135-137,

139, and 142-150, a part or derivative thereof.

In a further aspect, the invention relates to an agent which binds to a tumor-associated antigen identified according to the invention or to a part thereof. In a preferred embodiment, the agent is an antibody. In further embodiments, the antibody is a chimeric, a humanized antibody or an antibody produced by combinatory techniques or is a fragment of an antibody. Furthermore, the invention relates to an antibody which binds selectively to a complex of (i) a tumor-associated antigen identified according to the invention or a part thereof and (ii) an MHC molecule to which said tumor-associated antigen identified according to the invention or said part thereof binds, with said antibody not binding to (i) or (ii) alone. An antibody of the invention may be a monoclonal antibody. In further embodiments, the antibody is a chimeric or humanized antibody or a fragment of a natural antibody.

In particular, the invention relates to such an agent, in particular an antibody, which specifically binds to a peptide which has or comprises a sequence selected from the group consisting of SEQ ID NOs: 17-19, 90-97, 100-102, 105, 106, 111-116, 120, 123, 124, 135-137, 139, and 142-150, a part or derivative thereof.

With respect to claudin-18, the invention also relates to agents, in particular antibodies, which specifically bind to one variant of claudin-18. In one embodiment, the agent, in particular the antibody, specifically binds to the variant claudin-18A1 (SEQ ID NO: 118). In another embodiment, the agent, in particular the antibody, binds to the variant claudin-18A2 (SEQ ID NO: 16). Such specific antibodies may, for example, be obtained by immunizing using the peptides described in Example 4.

Furthermore, the invention with respect to claudin-18 relates to agents, in particular antibodies, specifically binding to a form of claudin-18A2 having a particular glycosylation pattern. In one embodiment, 5 the agent, in particular the antibody, specifically binds to a form of claudin-18A2 which is not glycosylated at one or more potential glycosylation sites. In another embodiment, the agent, in particular the antibody, specifically binds to a form of claudin-10 18A2 which is glycosylated at one or more potential glycosylation sites. Preferably, such potential glycosylation site relates to one or more positions selected from the group consisting of the amino acid positions 37, 38, 45, 116, 141, 146 and 205 of claudin-15 18A2. Furthermore, such a potential glycosylation preferably relates to a N glycosylation.

An agent which is specific for a variant or form of claudin-18, in particular an antibody which is specific 20 for a variant or form of claudin-18, in this respect means that the agent or the antibody binds stronger to the variant or form for which it is specific than to another variant or form. An agent, in particular an antibody, binds stronger to a first variant or form or 25 a first epitope compared to the binding to a second variant or form or a second epitope, if it binds to the first variant or form or to the first epitope with a dissociation constant (K_D) which is lower than the dissociation constant for the second variant or form or 30 the second epitope. Preferably, the dissociation constant (K_D) for the variant or form or the epitope to which the agent, in particular the antibody, binds specifically is more than 10-fold, preferably more than 20-fold, more preferably more than 50-fold, even more 35 preferably more than 100-fold and, in particular, more than 200-fold, 500-fold or 1000-fold lower than the dissociation constant (K_D) for the variant or form for the epitope to which the agent, in particular the

antibody, does not bind specifically. Preferably, an agent, in particular an antibody, does not bind or does not essentially bind to the variant or form or the epitope for which the agent, in particular the
5 antibody, is not specific.

The agents described above, in particular the antibodies and derivatives thereof as described herein, which specifically bind to a variant or a form of
10 claudin-18 may also be used in the compositions and methods of the invention.

The invention furthermore relates to a conjugate between an agent of the invention which binds to a
15 tumor-associated antigen identified according to the invention or to a part thereof or an antibody of the invention and a therapeutic or diagnostic agent. In one embodiment, the therapeutic or diagnostic agent is a toxin.

20

In a further aspect, the invention relates to a kit for detecting expression or abnormal expression of a tumor-associated antigen identified according to the invention, which kit comprises agents for detection (i)
25 of the nucleic acid which codes for the tumor-associated antigen or of a part thereof, (ii) of the tumor-associated antigen or of a part thereof, (iii) of antibodies which bind to the tumor-associated antigen or to a part thereof, and/or (iv) of T cells which are
30 specific for a complex between the tumor-associated antigen or a part thereof and an MHC molecule. In one embodiment, the agents for detection of the nucleic acid or the part thereof are nucleic acid molecules for selective amplification of said nucleic acid, which
35 comprise, in particular a sequence of 6-50, in particular 10-30, 15-30 and 20-30, contiguous nucleotides of said nucleic acid.

Detailed description of the invention

According to the invention, genes are described which are expressed in tumor cells selectively or aberrantly and which are tumor-associated antigens.

According to the invention, these genes and/or their genetic products and/or their derivatives and/or parts are preferred target structures for therapeutic approaches. Conceptionally, said therapeutic approaches may aim at inhibiting the activity of the selectively expressed tumor-associated genetic product. This is useful, if said aberrant respective selective expression is functionally important in tumor pathogenecity and if its ligation is accompanied by selective damage of the corresponding cells. Other therapeutic concepts contemplate tumor-associated antigens as labels which recruit effector mechanisms having cell-damaging potential selectively to tumor cells. Here, the function of the target molecule itself and its role in tumor development are totally irrelevant.

"Derivative" of a nucleic acid means according to the invention that single or multiple nucleotide substitutions, deletions and/or additions are present in said nucleic acid. Furthermore, the term "derivative" also comprises chemical derivatization of a nucleic acid on a nucleotide base, on the sugar or on the phosphate. The term "derivative" also comprises nucleic acids which contain nucleotides and nucleotide analogs not occurring naturally.

According to the invention, a nucleic acid is preferably deoxyribonucleic acid (DNA) or ribonucleic acid (RNA). Nucleic acids comprise according to the invention genomic DNA, cDNA, mRNA, recombinantly produced and chemically synthesized molecules.

According to the invention, a nucleic acid may be present as a single-stranded or double-stranded and linear or covalently circularly closed molecule.

5 The nucleic acids described according to the invention have preferably been isolated. The term "isolated nucleic acid" means according to the invention that the nucleic acid was (i) amplified *in vitro*, for example by polymerase chain reaction (PCR), (ii) recombinantly
10 produced by cloning, (iii) purified, for example by cleavage and gel-electrophoretic fractionation, or (iv) synthesized, for example by chemical synthesis. An isolated nucleic acid is a nucleic acid which is available for manipulation by recombinant DNA
15 techniques.

A nucleic acid is "complementary" to another nucleic acid if the two sequences are capable of hybridizing and forming a stable duplex with one another, with
20 hybridization preferably being carried out under conditions which allow specific hybridization between polynucleotides (stringent conditions). Stringent conditions are described, for example, in Molecular Cloning: A Laboratory Manual, J. Sambrook et al.,
25 Editors, 2nd Edition, Cold Spring Harbor Laboratory press, Cold Spring Harbor, New York, 1989 or Current Protocols in Molecular Biology, F.M. Ausubel et al., Editors, John Wiley & Sons, Inc., New York and refer, for example, to hybridization at 65°C in hybridization
30 buffer (3.5 x SSC, 0.02% Ficoll, 0.02% polyvinylpyrrolidone, 0.02% bovine serum albumin, 2.5 mM NaH₂PO₄ (pH 7), 0.5% SDS, 2 mM EDTA). SSC is 0.15 M sodium chloride/0.15 M sodium citrate, pH 7. After hybridization, the membrane to which the DNA has
35 been transferred is washed, for example, in 2 x SSC at room temperature and then in 0.1-0.5 x SSC/0.1 x SDS at temperatures of up to 68°C.

According to the invention, complementary nucleic acids have at least 40%, in particular at least 50%, at least 60%, at least 70%, at least 80%, at least 90% and preferably at least 95%, at least 98% or at least 99%,
5 identical nucleotides.

Nucleic acids coding for tumor-associated antigens may, according to the invention, be present alone or in combination with other nucleic acids, in particular
10 heterologous nucleic acids. In preferred embodiments, a nucleic acid is functionally linked to expression control sequences or regulatory sequences which may be homologous or heterologous with respect to said nucleic acid. A coding sequence and a regulatory sequence are
15 "functionally" linked to one another, if they are covalently linked to one another in such a way that expression or transcription of said coding sequence is under the control or under the influence of said regulatory sequence. If the coding sequence is to be
20 translated into a functional protein, then, with a regulatory sequence functionally linked to said coding sequence, induction of said regulatory sequence results in transcription of said coding sequence, without causing a frame shift in the coding sequence or said
25 coding sequence not being capable of being translated into the desired protein or peptide.

The term "expression control sequence" or "regulatory sequence" comprises according to the invention
30 promoters, enhancers and other control elements which regulate expression of a gene. In particular embodiments of the invention, the expression control sequences can be regulated. The exact structure of regulatory sequences may vary as a function of the
35 species or cell type, but generally comprises 5'untranscribed and 5'untranslated sequences which are involved in initiation of transcription and translation, respectively, such as TATA box, capping

sequence, CAAT sequence, and the like. More specifically, 5'untranscribed regulatory sequences comprise a promoter region which includes a promoter sequence for transcriptional control of the
5 functionally linked gene. Regulatory sequences may also comprise enhancer sequences or upstream activator sequences.

Thus, on the one hand, the tumor-associated antigens
10 illustrated herein may be combined with any expression control sequences and promoters. On the other hand, however, the promoters of the tumor-associated genetic products illustrated herein may, according to the invention, be combined with any other genes. This
15 allows the selective activity of these promoters to be utilized.

According to the invention, a nucleic acid may furthermore be present in combination with another
20 nucleic acid which codes for a polypeptide controlling secretion of the protein or polypeptide encoded by said nucleic acid from a host cell. According to the invention, a nucleic acid may also be present in combination with another nucleic acid which codes for a
25 polypeptide causing the encoded protein or polypeptide to be anchored on the cell membrane of the host cell or compartmentalized into particular organelles of said cell. Similarly, a combination with a nucleic acid is possible which represents a reporter gene or any "tag".

30

In a preferred embodiment, a recombinant DNA molecule is according to the invention a vector, where appropriate with a promoter, which controls expression of a nucleic acid, for example a nucleic acid coding
35 for a tumor-associated antigen of the invention. The term "vector" is used here in its most general meaning and comprises any intermediary vehicle for a nucleic acid which enables said nucleic acid, for example, to

be introduced into prokaryotic and/or eukaryotic cells and, where appropriate, to be integrated into a genome. Vectors of this kind are preferably replicated and/or expressed in the cells. An intermediary vehicle may be adapted, for example, to the use in electroporation, in bombardment with microprojectiles, in liposomal administration, in the transfer with the aid of agrobacteria or in insertion via DNA or RNA viruses. Vectors comprise plasmids, phagemids or viral genomes.

10

The nucleic acids coding for a tumor-associated antigen identified according to the invention may be used for transfection of host cells. Nucleic acids here mean both recombinant DNA and RNA. Recombinant RNA may be prepared by in-vitro transcription of a DNA template. Furthermore, it may be modified by stabilizing sequences, capping and polyadenylation prior to application.

20 According to the invention, the term "host cell" relates to any cell which can be transformed or transfected with an exogenous nucleic acid. The term "host cells" comprises according to the invention prokaryotic (e.g. *E. coli*) or eukaryotic cells (e.g. dendritic cells, B cells, CHO cells, COS cells, K562 cells, yeast cells and insect cells). Particular preference is given to mammalian cells such as cells from humans, mice, hamsters, pigs, goats, primates. The cells may be derived from a multiplicity of tissue types and comprise primary cells and cell lines. Specific examples comprise keratinocytes, peripheral blood leukocytes, stem cells of the bone marrow and embryonic stem cells. In further embodiments, the host cell is an antigen-presenting cell, in particular a dendritic cell, monocyte or a macrophage. A nucleic acid may be present in the host cell in the form of a single copy or of two or more copies and, in one embodiment, is expressed in the host cell.

According to the invention, the term "expression" is used in its most general meaning and comprises the production of RNA or of RNA and protein. It also
5 comprises partial expression of nucleic acids. Furthermore, expression may be carried out transiently or stably. Preferred expression systems in mammalian cells comprise pcDNA3.1 and pRc/CMV (Invitrogen, Carlsbad, CA), which contain a selectable marker such
10 as a gene imparting resistance to G418 (and thus enabling stably transfected cell lines to be selected) and the enhancer-promoter sequences of cytomegalovirus (CMV).

15 In those cases of the invention in which an HLA molecule presents a tumor-associated antigen or a part thereof, an expression vector may also comprise a nucleic acid sequence coding for said HLA molecule. The nucleic acid sequence coding for the HLA molecule may
20 be present on the same expression vector as the nucleic acid coding for the tumor-associated antigen or the part thereof, or both nucleic acids may be present on different expression vectors. In the latter case, the two expression vectors may be cotransfected into a
25 cell. If a host cell expresses neither the tumor-associated antigen or the part thereof nor the HLA molecule, both nucleic acids coding therefor are transfected into the cell either on the same expression vector or on different expression vectors. If the cell
30 already expresses the HLA molecule, only the nucleic acid sequence coding for the tumor-associated antigen or the part thereof can be transfected into the cell.

The invention also comprises kits for amplification of
35 a nucleic acid coding for a tumor-associated antigen. Such kits comprise, for example, a pair of amplification primers which hybridize to the nucleic acid coding for the tumor-associated antigen. The

primers preferably comprise a sequence of 6-50, in particular 10-30, 15-30 and 20-30 contiguous nucleotides of the nucleic acid and are nonoverlapping, in order to avoid the formation of primer dimers. One
5 of the primers will hybridize to one strand of the nucleic acid coding for the tumor-associated antigen, and the other primer will hybridize to the complementary strand in an arrangement which allows amplification of the nucleic acid coding for the tumor-
10 associated antigen.

"Antisense" molecules or "antisense" nucleic acids may be used for regulating, in particular reducing, expression of a nucleic acid. The term "antisense
15 molecule" or "antisense nucleic acid" refers according to the invention to an oligonucleotide which is an oligoribonucleotide, oligodeoxyribonucleotide, modified oligoribonucleotide or modified oligo-
deoxyribonucleotide and which hybridizes under
20 physiological conditions to DNA comprising a particular gene or to mRNA of said gene, thereby inhibiting transcription of said gene and/or translation of said mRNA. According to the invention, an "antisense
molecule" also comprises a construct which contains a
25 nucleic acid or a part thereof in reverse orientation with respect to its natural promoter. An antisense transcript of a nucleic acid or of a part thereof may form a duplex with the naturally occurring mRNA specifying the enzyme and thus prevent accumulation of
30 or translation of the mRNA into the active enzyme. Another possibility is the use of ribozymes for inactivating a nucleic acid. Antisense oligonucleotides preferred according to the invention have a sequence of 6-50, in particular 10-30, 15-30 and 20-30, contiguous
35 nucleotides of the target nucleic acid and preferably are fully complementary to the target nucleic acid or to a part thereof.

In preferred embodiments, the antisense oligonucleotide hybridizes with an N-terminal or 5' upstream site such as a translation initiation site, transcription initiation site or promoter site. In further
5 embodiments, the antisense oligonucleotide hybridizes with a 3'untranslated region or mRNA splicing site.

In one embodiment, an oligonucleotide of the invention consists of ribonucleotides, deoxyribonucleotides or a
10 combination thereof, with the 5' end of one nucleotide and the 3' end of another nucleotide being linked to one another by a phosphodiester bond. These oligonucleotides may be synthesized in the conventional manner or produced recombinantly.

15 In preferred embodiments, an oligonucleotide of the invention is a "modified" oligonucleotide. Here, the oligonucleotide may be modified in very different ways, without impairing its ability to bind its target, in
20 order to increase, for example, its stability or therapeutic efficacy. According to the invention, the term "modified oligonucleotide" means an oligonucleotide in which (i) at least two of its nucleotides are linked to one another by a synthetic
25 internucleoside bond (i.e. an internucleoside bond which is not a phosphodiester bond) and/or (ii) a chemical group which is usually not found in nucleic acids is covalently linked to the oligonucleotide. Preferred synthetic internucleoside bonds are
30 phosphorothioates, alkyl phosphonates, phosphorodithioates, phosphate esters, alkyl phosphonothioates, phosphoramidates, carbamates, carbonates, phosphate triesters, acetamidates, carboxymethyl esters and peptides.

35 The term "modified oligonucleotide" also comprises oligonucleotides having a covalently modified base and/or sugar. "Modified oligonucleotides" comprise, for

example, oligonucleotides with sugar residues which are covalently bound to low molecular weight organic groups other than a hydroxyl group at the 3' position and a phosphate group at the 5' position. Modified
5 oligonucleotides may comprise, for example, a 2'-O-alkylated ribose residue or another sugar instead of ribose, such as arabinose.

Preferably, the proteins and polypeptides described
10 according to the invention have been isolated. The terms "isolated protein" or "isolated polypeptide" mean that the protein or polypeptide has been separated from its natural environment. An isolated protein or polypeptide may be in an essentially purified state.
15 The term "essentially purified" means that the protein or polypeptide is essentially free of other substances with which it is associated in nature or *in vivo*.

Such proteins and polypeptides may be used, for
20 example, in producing antibodies and in an immunological or diagnostic assay or as therapeutics. Proteins and polypeptides described according to the invention may be isolated from biological samples such as tissue or cell homogenates and may also be expressed
25 recombinantly in a multiplicity of pro- or eukaryotic expression systems.

For the purposes of the present invention, "derivatives" of a protein or polypeptide or of an
30 amino acid sequence comprise amino acid insertion variants, amino acid deletion variants and/or amino acid substitution variants.

Amino acid insertion variants comprise amino- and/or
35 carboxy-terminal fusions and also insertions of single or two or more amino acids in a particular amino acid sequence. In the case of amino acid sequence variants having an insertion, one or more amino acid residues

are inserted into a particular site in an amino acid sequence, although random insertion with appropriate screening of the resulting product is also possible. Amino acid deletion variants are characterized by the
5 removal of one or more amino acids from the sequence. Amino acid substitution variants are characterized by at least one residue in the sequence being removed and another residue being inserted in its place. Preference is given to the modifications being in positions in the
10 amino acid sequence which are not conserved between homologous proteins or polypeptides. Preference is given to replacing amino acids with other ones having similar properties such as hydrophobicity, hydrophilicity, electronegativity, volume of the side
15 chain and the like (conservative substitution). Conservative substitutions, for example, relate to the exchange of one amino acid with another amino acid listed below in the same group as the amino acid to be substituted:

20

1. small aliphatic, nonpolar or slightly polar residues: Ala, Ser, Thr (Pro, Gly)
2. negatively charged residues and their amides: Asn, Asp, Glu, Gln
- 25 3. positively charged residues: His, Arg, Lys
4. large aliphatic, nonpolar residues: Met, Leu, Ile, Val (Cys)
5. large aromatic residues: Phe, Tyr, Trp.

30 Owing to their particular part in protein architecture, three residues are shown in brackets. Gly is the only residue without a side chain and thus imparts flexibility to the chain. Pro has an unusual geometry which greatly restricts the chain. Cys can form a
35 disulfide bridge.

The amino acid variants described above may be readily prepared with the aid of known peptide synthesis

techniques such as, for example, by solid phase synthesis (Merrifield, 1964) and similar methods or by recombinant DNA manipulation. Techniques for introducing substitution mutations at predetermined
5 sites into DNA which has a known or partially known sequence are well known and comprise M13 mutagenesis, for example. The manipulation of DNA sequences for preparing proteins having substitutions, insertions or deletions, is described in detail in Sambrook et al.
10 (1989), for example.

According to the invention, "derivatives" of proteins, polypeptides or peptides also comprise single or multiple substitutions, deletions and/or additions of
15 any molecules associated with the enzyme, such as carbohydrates, lipids and/or proteins, polypeptides or peptides. The term "derivative" also extends to all functional chemical equivalents of said proteins, polypeptides or peptides.

20 According to the invention, a part or fragment of a tumor-associated antigen has a functional property of the polypeptide from which it has been derived. Such functional properties comprise the interaction with
25 antibodies, the interaction with other polypeptides or proteins, the selective binding of nucleic acids and an enzymatic activity. A particular property is the ability to form a complex with HLA and, where appropriate, generate an immune response. This immune
30 response may be based on stimulating cytotoxic or T helper cells. A part or fragment of a tumor-associated antigen of the invention preferably comprises a sequence of at least 6, in particular at least 8, at least 10, at least 12, at least 15, at
35 least 20, at least 30 or at least 50, consecutive amino acids of the tumor-associated antigen.

A part or a fragment of a nucleic acid coding for a

tumor-associated antigen relates according to the invention to the part of the nucleic acid, which codes at least for the tumor-associated antigen and/or for a part or a fragment of said tumor-associated antigen, as
5 defined above.

The isolation and identification of genes coding for tumor-associated antigens also make possible the diagnosis of a disease characterized by expression of
10 one or more tumor-associated antigens. These methods comprise determining one or more nucleic acids which code for a tumor-associated antigen and/or determining the encoded tumor-associated antigens and/or peptides derived therefrom. The nucleic acids may be determined
15 in the conventional manner, including by polymerase chain reaction or hybridization with a labeled probe. Tumor-associated antigens or peptides derived therefrom may be determined by screening patient antisera with respect to recognizing the antigen and/or the peptides.
20 They may also be determined by screening T cells of the patient for specificities for the corresponding tumor-associated antigen.

The present invention also enables proteins binding to
25 tumor-associated antigens described herein to be isolated, including antibodies and cellular binding partners of said tumor-associated antigens.

According to the invention, particular embodiments
30 ought to involve providing "dominant negative" polypeptides derived from tumor-associated antigens. A dominant negative polypeptide is an inactive protein variant which, by way of interacting with the cellular machinery, displaces an active protein from its
35 interaction with the cellular machinery or which competes with the active protein, thereby reducing the effect of said active protein. For example, a dominant negative receptor which binds to a ligand but does not

generate any signal as response to binding to the ligand can reduce the biological effect of said ligand. Similarly, a dominant negative catalytically inactive kinase which usually interacts with target proteins but
5 does not phosphorylate said target proteins may reduce phosphorylation of said target proteins as response to a cellular signal. Similarly, a dominant negative transcription factor which binds to a promoter site in the control region of a gene but does not increase
10 transcription of said gene may reduce the effect of a normal transcription factor by occupying promoter binding sites, without increasing transcription.

The result of expression of a dominant negative polypeptide in a cell is a reduction in the function of
15 active proteins. The skilled worker may prepare dominant negative variants of a protein, for example, by conventional mutagenesis methods and by evaluating the dominant negative effect of the variant
20 polypeptide.

The invention also comprises substances such as polypeptides which bind to tumor-associated antigens. Such binding substances may be used, for example, in
25 screening assays for detecting tumor-associated antigens and complexes of tumor-associated antigens with their binding partners and in the purification of said tumor-associated antigens and of complexes thereof with their binding partners. Such substances may also
30 be used for inhibiting the activity of tumor-associated antigens, for example by binding to such antigens.

The invention therefore comprises binding substances such as, for example, antibodies or antibody fragments,
35 which are capable of selectively binding to tumor-associated antigens. Antibodies comprise polyclonal and monoclonal antibodies which are produced in the conventional manner.

Such antibodies can recognize proteins in the native and/or denaturated state (Anderson et al., J. Immunol. 143: 1899-1904, 1989; Gardsvoll, J. Immunol. Methods 234: 107-116, 2000; Kayyem et al., Eur. J. Biochem. 208: 1-8, 1992; Spiller et al., J. Immunol. Methods 224: 51-60, 1999).

Antisera which contain specific antibodies specifically binding to the target protein can be prepared by various standard processes; see, for example, "Monoclonal Antibodies: A Practical Approach" by Philip Shepherd, Christopher Dean ISBN 0-19-963722-9; "Antibodies: A Laboratory Manual" by Ed Harlow, David Lane, ISBN: 0879693142 and "Using Antibodies: A Laboratory Manual: Portable Protocol NO" by Edward Harlow, David Lane, Ed Harlow ISBN 0879695447. Thereby it is also possible to generate affine and specific antibodies which recognize complex membrane proteins in their native form (Azorsa et al., J. Immunol. Methods 229: 35-48, 1999; Anderson et al., J. Immunol. 143: 1899-1904, 1989; Gardsvoll, J. Immunol. Methods 234: 107-116, 2000). This is in particular relevant for the preparation of antibodies which are to be used therapeutically, but also for many diagnostic applications. In this respect, it is possible to immunize with the whole protein, with extracellular partial sequences as well as with cells which express the target molecule in physiologically folded form.

30

Monoclonal antibodies are traditionally prepared using the hybridoma technology. (for technical details see: "Monoclonal Antibodies: A Practical Approach" by Philip Shepherd, Christopher Dean ISBN 0-19-963722-9; "Antibodies: A Laboratory Manual" by Ed Harlow, David Lane ISBN: 0879693142; "Using Antibodies: A Laboratory Manual: Portable Protocol NO" by Edward Harlow, David Lane, Ed Harlow ISBN: 0879695447).

It is known that only a small part of an antibody molecule, the paratope, is involved in binding of the antibody to its epitope (cf. Clark, W.R. (1986), *The Experimental Foundations of Modern Immunology*, Wiley & Sons, Inc., New York; Roitt, I. (1991), *Essential Immunology*, 7th Edition, Blackwell Scientific Publications, Oxford). The pFc' and Fc regions are, for example, effectors of the complement cascade but are not involved in antigen binding. An antibody from which the pFc' region has been enzymatically removed or which has been produced without the pFc' region, referred to as F(ab')₂ fragment, carries both antigen binding sites of a complete antibody. Similarly, an antibody from which the Fc region has been enzymatically removed or which has been produced without said Fc region, referred to as Fab fragment, carries one antigen binding site of an intact antibody molecule. Furthermore, Fab fragments consist of a covalently bound light chain of an antibody and part of the heavy chain of said antibody, referred to as Fd. The Fd fragments are the main determinants of antibody specificity (a single Fd fragment can be associated with up to ten different light chains, without altering the specificity of the antibody) and Fd fragments, when isolated, retain the ability to bind to an epitope.

Located within the antigen-binding part of an antibody are complementary-determining regions (CDRs) which interact directly with the antigen epitope and framework regions (FRs) which maintain the tertiary structure of the paratope. Both the Fd fragment of the heavy chain and the light chain of IgG immunoglobulins contain four framework regions (FR1 to FR4) which are separated in each case by three complementary-determining regions (CDR1 to CDR3). The CDRs and, in particular, the CDR3 regions and, still more particularly, the CDR3 region of the heavy chain are

responsible to a large extent for antibody specificity.

Non-CDR regions of a mammalian antibody are known to be able to be replaced by similar regions of antibodies with the same or a different specificity, with the specificity for the epitope of the original antibody being retained. This made possible the development of "humanized" antibodies in which nonhuman CDRs are covalently linked to human FR and/or Fc/pFc' regions to produce a functional antibody.

This is utilized in the so called "SLAM" technology, wherein B cells from whole blood are isolated and the cells are monocloned. Then, the supernatant of the single B cells is analyzed with respect to its antibody specificity. In contrast to the hybridoma technology the variable region of the antibody gene is amplified using single cell PCR and cloned into a suitable vector. In this way, the provision of monoclonal antibodies is accelerated (de Wildt et al., J. Immunol. Methods 207: 61-67, 1997).

As another example, WO 92/04381 describes the production and use of humanized murine RSV antibodies in which at least part of the murine FR regions have been replaced with FR regions of a human origin. Antibodies of this kind, including fragments of intact antibodies with antigen-binding capability, are often referred to as "chimeric" antibodies.

The invention also provides $F(ab')_2$, Fab, Fv, and Fd fragments of antibodies, chimeric antibodies, in which the Fc and/or FR and/or CDR1 and/or CDR2 and/or light chain-CDR3 regions have been replaced with homologous human or nonhuman sequences, chimeric $F(ab')_2$ -fragment antibodies in which the FR and/or CDR1 and/or CDR2 and/or light chain-CDR3 regions have been replaced with homologous human or nonhuman sequences, chimeric Fab-

fragment antibodies in which the FR and/or CDR1 and/or CDR2 and/or light chain-CDR3 regions have been replaced with homologous human or nonhuman sequences, and chimeric Fd-fragment antibodies in which the FR and/or CDR1 and/or CDR2 regions have been replaced with homologous human or nonhuman sequences. The invention also comprises "single-chain" antibodies.

The invention also comprises polypeptides which bind specifically to tumor-associated antigens. Polypeptide binding substances of this kind may be provided, for example, by degenerate peptide libraries which may be prepared simply in solution in an immobilized form or as phage-display libraries. It is likewise possible to prepare combinatorial libraries of peptides with one or more amino acids. Libraries of peptoids and nonpeptidic synthetic residues may also be prepared.

Phage display may be particularly effective in identifying binding peptides of the invention. In this connection, for example, a phage library is prepared (using, for example, the M13, fd or lambda phages) which presents inserts of from 4 to about 80 amino acid residues in length. Phages are then selected which carry inserts which bind to the tumor-associated antigen. This process may be repeated via two or more cycles of a reselection of phages binding to the tumor-associated antigen. Repeated rounds result in a concentration of phages carrying particular sequences. An analysis of DNA sequences may be carried out in order to identify the sequences of the expressed polypeptides. The smallest linear portion of the sequence binding to the tumor-associated antigen may be determined. The "two-hybrid system" of yeast may also be used for identifying polypeptides which bind to a tumor-associated antigen. Tumor-associated antigens described according to the invention or fragments thereof may be used for screening peptide libraries,

including phage-display libraries, in order to identify and select peptide binding partners of the tumor-associated antigens. Such molecules may be used, for example, for screening assays, purification protocols, 5 for interference with the function of the tumor-associated antigen and for other purposes known to the skilled worker.

The antibodies described above and other binding 10 molecules may be used, for example, for identifying tissue which expresses a tumor-associated antigen. Antibodies may also be coupled to specific diagnostic substances for displaying cells and tissues expressing tumor-associated antigens. They may also be coupled to 15 therapeutically useful substances. Diagnostic substances comprise, in a nonlimiting manner, barium sulfate, iocetamic acid, iopanoic acid, calcium ipodate, sodium diatrizoate, meglumine diatrizoate, metrizamide, sodium tyropanoate and radio diagnostic, 20 including positron emitters such as fluorine-18 and carbon-11, gamma emitters such as iodine-123, technetium-99m, iodine-131 and indium-111, nuclides for nuclear magnetic resonance, such as fluorine and gadolinium. According to the invention, the term 25 "therapeutically useful substance" means any therapeutic molecule which, as desired, is selectively guided to a cell which expresses one or more tumor-associated antigens, including anticancer agents, radioactive iodine-labeled compounds, toxins, 30 cytostatic or cytolytic drugs, etc. Anticancer agents comprise, for example, aminoglutethimide, azathioprine, bleomycin sulfate, busulfan, carmustine, chlorambucil, cisplatin, cyclophosphamide, cyclosporine, cytarabidine, dacarbazine, dactinomycin, daunorubin, 35 doxorubicin, taxol, etoposide, fluorouracil, interferon- α , lomustine, mercaptopurine, methotrexate, mitotane, procarbazine HCl, thioguanine, vinblastine sulfate and vincristine sulfate. Other anticancer

agents are described, for example, in Goodman and Gilman, "The Pharmacological Basis of Therapeutics", 8th Edition, 1990, McGraw-Hill, Inc., in particular Chapter 52 (Antineoplastic Agents (Paul Calabresi and
5 Bruce A. Chabner). Toxins may be proteins such as pokeweed antiviral protein, cholera toxin, pertussis toxin, ricin, gelonin, abrin, diphtheria exotoxin or *Pseudomonas* exotoxin. Toxin residues may also be high energy-emitting radionuclides such as cobalt-60.

10

The term "patient" means according to the invention a human being, a nonhuman primate or another animal, in particular a mammal such as a cow, horse, pig, sheep, goat, dog, cat or a rodent such as a mouse and rat. In
15 a particularly preferred embodiment, the patient is a human being.

According to the invention, the term "disease" refers to any pathological state in which tumor-associated
20 antigens are expressed or abnormally expressed. "Abnormal expression" means according to the invention that expression is altered, preferably increased, compared to the state in a healthy individual. An increase in expression refers to an increase by at
25 least 10%, in particular at least 20%, at least 50% or at least 100%. In one embodiment, the tumor-associated antigen is expressed only in tissue of a diseased individual, while expression in a healthy individual is repressed. One example of such a disease is cancer,
30 wherein the term "cancer" according to the invention comprises leukemias, seminomas, melanomas, teratomas, gliomas, kidney cancer, adrenal cancer, thyroid cancer, intestinal cancer, liver cancer, colon cancer, stomach cancer, gastrointestinal cancer, lymph node cancer,
35 esophagus cancer, colorectal cancer, pancreas cancer, ear, nose and throat (ENT) cancer, breast cancer, prostate cancer, cancer of the uterus, ovarian cancer and lung cancer and the metastases thereof.

According to the invention, a biological sample may be a tissue sample and/or a cellular sample and may be obtained in the conventional manner such as by tissue
5 biopsy, including punch biopsy, and by taking blood, bronchial aspirate, sputum, urine, feces or other body fluids, for use in the various methods described herein.

10 According to the invention, the term "immunoreactive cell" means a cell which can mature into an immune cell (such as B cell, T helper cell, or cytolytic T cell) with suitable stimulation. Immunoreactive cells comprise CD34⁺ hematopoietic stem cells, immature and
15 mature T cells and immature and mature B cells. If production of cytolytic or T helper cells recognizing a tumor-associated antigen is desired, the immunoreactive cell is contacted with a cell expressing a tumor-associated antigen under conditions which favor
20 production, differentiation and/or selection of cytolytic T cells and of T helper cells. The differentiation of T cell precursors into a cytolytic T cell, when exposed to an antigen, is similar to clonal selection of the immune system.

25 Some therapeutic methods are based on a reaction of the immune system of a patient, which results in a lysis of antigen-presenting cells such as cancer cells which present one or more tumor-associated antigens. In this
30 connection, for example autologous cytotoxic T lymphocytes specific for a complex of a tumor-associated antigen and an MHC molecule are administered to a patient having a cellular abnormality. The production of such cytotoxic T lymphocytes *in vitro* is
35 known. An example of a method of differentiating T cells can be found in WO-A-9633265. Generally, a sample containing cells such as blood cells is taken from the patient and the cells are contacted with a cell which

presents the complex and which can cause propagation of cytotoxic T lymphocytes (e.g. dendritic cells). The target cell may be a transfected cell such as a COS cell. These transfected cells present the desired
5 complex on their surface and, when contacted with cytotoxic T lymphocytes, stimulate propagation of the latter. The clonally expanded autologous cytotoxic T lymphocytes are then administered to the patient.

10 In another method of selecting antigen-specific cytotoxic T lymphocytes, fluorogenic tetramers of MHC class I molecule/peptide complexes are used for detecting specific clones of cytotoxic T lymphocytes (Altman et al., *Science* 274:94-96, 1996; Dunbar et al.,
15 *Curr. Biol.* 8:413-416, 1998). Soluble MHC class I molecules are folded *in vitro* in the presence of β_2 microglobulin and a peptide antigen binding to said class I molecule. The MHC/peptide complexes are purified and then labeled with biotin. Tetramers are
20 formed by mixing the biotinylated peptide-MHC complexes with labeled avidin (e.g. phycoerythrin) in a molar ratio of 4:1. Tetramers are then contacted with cytotoxic T lymphocytes such as peripheral blood or lymph nodes. The tetramers bind to cytotoxic T
25 lymphocytes which recognize the peptide antigen/MHC class I complex. Cells which are bound to the tetramers may be sorted by fluorescence-controlled cell sorting to isolate reactive cytotoxic T lymphocytes. The isolated cytotoxic T lymphocytes may then be propagated
30 *in vitro*.

In a therapeutic method referred to as adoptive transfer (Greenberg, *J. Immunol.* 136(5):1917, 1986; Riddell et al., *Science* 257:238, 1992; Lynch et al.,
35 *Eur. J. Immunol.* 21:1403-1410, 1991; Kast et al., *Cell* 59:603-614, 1989), cells presenting the desired complex (e.g. dendritic cells) are combined with cytotoxic T lymphocytes of the patient to be treated, resulting in

a propagation of specific cytotoxic T lymphocytes. The propagated cytotoxic T lymphocytes are then administered to a patient having a cellular anomaly characterized by particular abnormal cells presenting the specific complex. The cytotoxic T lymphocytes then lyse the abnormal cells, thereby achieving a desired therapeutic effect.

Often, of the T cell repertoire of a patient, only T cells with low affinity for a specific complex of this kind can be propagated, since those with high affinity have been extinguished due to development of tolerance. An alternative here may be a transfer of the T cell receptor itself. For this too, cells presenting the desired complex (e.g. dendritic cells) are combined with cytotoxic T lymphocytes of healthy individuals or another species (e.g. mouse). This results in propagation of specific cytotoxic T lymphocytes with high affinity if the T lymphocytes are derived from a donor organism which had no previous contact with the specific complex. The high affinity T cell receptor of these propagated specific T lymphocytes is cloned. If the high affinity T cell receptors have been cloned from another species they can be humanized to a different extent. Such T cell receptors are then transduced via gene transfer, for example using retroviral vectors, into T cells of patients, as desired. Adoptive transfer is then carried out using these genetically altered T lymphocytes (Stanislowski et al., Nat Immunol. 2:962-70, 2001; Kessels et al., Nat Immunol. 2:957-61, 2001).

The therapeutic aspects above start out from the fact that at least some of the abnormal cells of the patient present a complex of a tumor-associated antigen and an HLA molecule. Such cells may be identified in a manner known per se. As soon as cells presenting the complex have been identified, they may be combined with a

sample from the patient, which contains cytotoxic T lymphocytes. If the cytotoxic T lymphocytes lyse the cells presenting the complex, it can be assumed that a tumor-associated antigen is presented.

5

Adoptive transfer is not the only form of therapy which can be applied according to the invention. Cytotoxic T lymphocytes may also be generated *in vivo* in a manner known per se. One method uses nonproliferative cells
10 expressing the complex. The cells used here will be those which usually express the complex, such as irradiated tumor cells or cells transfected with one or both genes necessary for presentation of the complex (i.e. the antigenic peptide and the presenting HLA
15 molecule). Various cell types may be used. Furthermore, it is possible to use vectors which carry one or both of the genes of interest. Particular preference is given to viral or bacterial vectors. For example, nucleic acids coding for a tumor-associated antigen or
20 for a part thereof may be functionally linked to promoter and enhancer sequences which control expression of said tumor-associated antigen or a fragment thereof in particular tissues or cell types. The nucleic acid may be incorporated into an expression
25 vector. Expression vectors may be nonmodified extrachromosomal nucleic acids, plasmids or viral genomes into which exogenous nucleic acids may be inserted. Nucleic acids coding for a tumor-associated antigen may also be inserted into a retroviral genome,
30 thereby enabling the nucleic acid to be integrated into the genome of the target tissue or target cell. In these systems, a microorganism such as vaccinia virus, pox virus, Herpes simplex virus, retrovirus or adenovirus carries the gene of interest and de facto
35 "infects" host cells. Another preferred form is the introduction of the tumor-associated antigen in the form of recombinant RNA which may be introduced into cells by liposomal transfer or by electroporation, for

example. The resulting cells present the complex of interest and are recognized by autologous cytotoxic T lymphocytes which then propagate.

5 A similar effect can be achieved by combining the tumor-associated antigen or a fragment thereof with an adjuvant in order to make incorporation into antigen-presenting cells *in vivo* possible. The tumor-associated antigen or a fragment thereof may be represented as
10 protein, as DNA (e.g. within a vector) or as RNA. The tumor-associated antigen is processed to produce a peptide partner for the HLA molecule, while a fragment thereof may be presented without the need for further processing. The latter is the case in particular, if
15 these can bind to HLA molecules. Preference is given to administration forms in which the complete antigen is processed *in vivo* by a dendritic cell, since this may also produce T helper cell responses which are needed for an effective immune response (Ossendorp et al.,
20 *Immunol Lett.* 74:75-9, 2000; Ossendorp et al., *J. Exp. Med.* 187:693-702, 1998). In general, it is possible to administer an effective amount of the tumor-associated antigen to a patient by intradermal injection, for example. However, injection may also be carried out
25 intranodally into a lymph node (Maloy et al., *Proc Natl Acad Sci USA* 98:3299-303, 2001). It may also be carried out in combination with reagents which facilitate uptake into dendritic cells. Preferred tumor-associated antigens comprise those which react with allogenic
30 cancer antisera or with T cells of many cancer patients. Of particular interest, however, are those against which no spontaneous immune responses pre-exist. Evidently, it is possible to induce against these immune responses which can lyse tumors (Keogh et al.,
35 *J. Immunol.* 167:787-96, 2001; Appella et al., *Biomed Pept Proteins Nucleic Acids* 1:177-84, 1995; Wentworth et al., *Mol Immunol.* 32:603-12, 1995).

The pharmaceutical compositions described according to the invention may also be used as vaccines for immunization. According to the invention, the terms "immunization" or "vaccination" mean an increase in or
5 activation of an immune response to an antigen. It is possible to use animal models for testing an immunizing effect on cancer by using a tumor-associated antigen or a nucleic acid coding therefor. For example, human cancer cells may be introduced into a mouse to generate
10 a tumor, and one or more nucleic acids coding for tumor-associated antigens may be administered. The effect on the cancer cells (for example reduction in tumor size) may be measured as a measure for the effectiveness of an immunization by the nucleic acid.

15 As part of the composition for an immunization, one or more tumor-associated antigens or stimulating fragments thereof are administered together with one or more adjuvants for inducing an immune response or for
20 increasing an immune response. An adjuvant is a substance which is incorporated into the antigen or administered together with the latter and which enhances the immune response. Adjuvants may enhance the immune response by providing an antigen reservoir
25 (extracellularly or in macrophages), activating macrophages and/or stimulating particular lymphocytes. Adjuvants are known and comprise in a nonlimiting way monophosphoryl lipid A (MPL, SmithKline Beecham), saponins such as QS21 (SmithKline Beecham), DQS21
30 (SmithKline Beecham; WO 96/33739), QS7, QS17, QS18 and QS-L1 (So et al., Mol. Cells 7:178-186, 1997), incomplete Freund's adjuvant, complete Freund's adjuvant, vitamin E, montanide, alum, CpG oligonucleotides (cf. Kreig et al., Nature 374:546-9,
35 1995) and various water-in-oil emulsions prepared from biologically degradable oils such as squalene and/or tocopherol. Preferably, the peptides are administered in a mixture with DQS21/MPL. The ratio of DQS21 to MPL

is typically about 1:10 to 10:1, preferably about 1:5 to 5:1 and in particular about 1:1. For administration to humans, a vaccine formulation typically contains DQS21 and MPL in a range from about 1 µg to about 5 100 µg.

Other substances which stimulate an immune response of the patient may also be administered. It is possible, for example, to use cytokines in a vaccination, owing 10 to their regulatory properties on lymphocytes. Such cytokines comprise, for example, interleukin-12 (IL-12) which was shown to increase the protective actions of vaccines (cf. *Science* 268:1432-1434, 1995), GM-CSF and IL-18.

15 There are a number of compounds which enhance an immune response and which therefore may be used in a vaccination. Said compounds comprise costimulating molecules provided in the form of proteins or nucleic 20 acids. Examples of such costimulating molecules are B7-1 and B7-2 (CD80 and CD86, respectively) which are expressed on dendritic cells (DC) and interact with the CD28 molecule expressed on the T cells. This interaction provides a costimulation (signal 2) for an 25 antigen/MHC/TCR-stimulated (signal 1) T cell, thereby enhancing propagation of said T cell and the effector function. B7 also interacts with CTLA4 (CD152) on T cells, and studies involving CTLA4 and B7 ligands demonstrate that B7-CTLA4 interaction can enhance 30 antitumor immunity and CTL propagation (Zheng, P. et al., *Proc. Natl. Acad. Sci. USA* 95(11):6284-6289 (1998)).

B7 is typically not expressed on tumor cells so that 35 these are no effective antigen-presenting cells (APCs) for T cells. Induction of B7 expression would enable tumor cells to stimulate more effectively propagation of cytotoxic T lymphocytes and an effector function.

Costimulation by a combination of B7/IL-6/IL-12 revealed induction of IFN-gamma and Th1-cytokine profile in a T cell population, resulting in further enhanced T cell activity (Gajewski et al., *J. Immunol.* 5 154:5637-5648 (1995)).

A complete activation of cytotoxic T lymphocytes and a complete effector function require an involvement of T helper cells via interaction between the CD40 ligand 10 on said T helper cells and the CD40 molecule expressed by dendritic cells (Ridge et al., *Nature* 393:474 (1998), Bennett et al., *Nature* 393:478 (1998), Schönberger et al., *Nature* 393:480 (1998)). The mechanism of this costimulating signal probably relates 15 to the increase in B7 production and associated IL-6/IL-12 production by said dendritic cells (antigen-presenting cells). CD40-CD40L interaction thus complements the interaction of signal 1 (antigen/MHC-TCR) and signal 2 (B7-CD28).

20

The use of anti-CD40 antibodies for stimulating dendritic cells would be expected to directly enhance a response to tumor antigens which are usually outside the range of an inflammatory response or which are 25 presented by nonprofessional antigen-presenting cells (tumor cells). In these situations, T helper and B7-costimulating signals are not provided. This mechanism could be used in connection with therapies based on antigen-pulsed dendritic cells.

30

The invention also provides for administration of nucleic acids, polypeptides or peptides. Polypeptides and peptides may be administered in a manner known per se. In one embodiment, nucleic acids are administered 35 by *ex vivo* methods, i.e. by removing cells from a patient, genetic modification of said cells in order to incorporate a tumor-associated antigen and reintroduction of the altered cells into the patient.

This generally comprises introducing a functional copy of a gene into the cells of a patient *in vitro* and reintroducing the genetically altered cells into the patient. The functional copy of the gene is under the functional control of regulatory elements which allow the gene to be expressed in the genetically altered cells. Transfection and transduction methods are known to the skilled worker. The invention also provides for administering nucleic acids *in vivo* by using vectors such as viruses and target-controlled liposomes.

In a preferred embodiment, a viral vector for administering a nucleic acid coding for a tumor-associated antigen is selected from the group consisting of adenoviruses, adeno-associated viruses, pox viruses, including vaccinia virus and attenuated pox viruses, Semliki Forest virus, retroviruses, Sindbis virus and Ty virus-like particles. Particular preference is given to adenoviruses and retroviruses. The retroviruses are typically replication-deficient (i.e. they are incapable of generating infectious particles).

Various methods may be used in order to introduce according to the invention nucleic acids into cells *in vitro* or *in vivo*. Methods of this kind comprise transfection of nucleic acid CaPO_4 precipitates, transfection of nucleic acids associated with DEAE, transfection or infection with the above viruses carrying the nucleic acids of interest, liposome-mediated transfection, and the like. In particular embodiments, preference is given to directing the nucleic acid to particular cells. In such embodiments, a carrier used for administering a nucleic acid to a cell (e.g. a retrovirus or a liposome) may have a bound target control molecule. For example, a molecule such as an antibody specific for a surface membrane protein on the target cell or a ligand for a receptor on the

target cell may be incorporated into or attached to the nucleic acid carrier. Preferred antibodies comprise antibodies which bind selectively a tumor-associated antigen. If administration of a nucleic acid via
5 liposomes is desired, proteins binding to a surface membrane protein associated with endocytosis may be incorporated into the liposome formulation in order to make target control and/or uptake possible. Such proteins comprise capsid proteins or fragments thereof
10 which are specific for a particular cell type, antibodies to proteins which are internalized, proteins addressing an intracellular site, and the like.

The therapeutic compositions of the invention may be
15 administered in pharmaceutically compatible preparations. Such preparations may usually contain pharmaceutically compatible concentrations of salts, buffer substances, preservatives, carriers, supplementing immunity-enhancing substances such as
20 adjuvants, CpG and cytokines and, where appropriate, other therapeutically active compounds.

The therapeutically active compounds of the invention may be administered via any conventional route,
25 including by injection or infusion. The administration may be carried out, for example, orally, intravenously, intraperitoneally, intramuscularly, subcutaneously or transdermally. Preferably, antibodies are therapeutically administered by way of a lung aerosol.
30 Antisense nucleic acids are preferably administered by slow intravenous administration.

The compositions of the invention are administered in effective amounts. An "effective amount" refers to the
35 amount which achieves a desired reaction or a desired effect alone or together with further doses. In the case of treatment of a particular disease or of a particular condition characterized by expression of one

or more tumor-associated antigens, the desired reaction relates to inhibition of the course of the disease. This comprises slowing down the progress of the disease and, in particular, interrupting the progress of the
5 disease. The desired reaction in a treatment of a disease or of a condition may also be delay of the onset or a prevention of the onset of said disease or said condition.

10 An effective amount of a composition of the invention will depend on the condition to be treated, the severeness of the disease, the individual parameters of the patient, including age, physiological condition, size and weight, the duration of treatment, the type of
15 an accompanying therapy (if present), the specific route of administration and similar factors.

The pharmaceutical compositions of the invention are preferably sterile and contain an effective amount of
20 the therapeutically active substance to generate the desired reaction or the desired effect.

The doses administered of the compositions of the invention may depend on various parameters such as the
25 type of administration, the condition of the patient, the desired period of administration, etc. In the case that a reaction in a patient is insufficient with an initial dose, higher doses (or effectively higher doses achieved by a different, more localized route of
30 administration) may be used.

Generally, doses of the tumor-associated antigen of from 1 ng to 1 mg, preferably from 10 ng to 100 µg, are formulated and administered for a treatment or for
35 generating or increasing an immune response. If the administration of nucleic acids (DNA and RNA) coding for tumor-associated antigens is desired, doses of from 1 ng to 0.1 mg are formulated and administered.

The pharmaceutical compositions of the invention are generally administered in pharmaceutically compatible amounts and in pharmaceutically compatible compositions. The term "pharmaceutically compatible" refers to a nontoxic material which does not interact with the action of the active component of the pharmaceutical composition. Preparations of this kind may usually contain salts, buffer substances, preservatives, carriers and, where appropriate, other therapeutically active compounds. When used in medicine, the salts should be pharmaceutically compatible. However, salts which are not pharmaceutically compatible may be used for preparing pharmaceutically compatible salts and are included in the invention. Pharmacologically and pharmaceutically compatible salts of this kind comprise in a nonlimiting way those prepared from the following acids: hydrochloric, hydrobromic, sulfuric, nitric, phosphoric, maleic, acetic, salicylic, citric, formic, malonic, succinic acids, and the like. Pharmaceutically compatible salts may also be prepared as alkali metal salts or alkaline earth metal salts, such as sodium salts, potassium salts or calcium salts.

A pharmaceutical composition of the invention may comprise a pharmaceutically compatible carrier. According to the invention, the term "pharmaceutically compatible carrier" refers to one or more compatible solid or liquid fillers, diluents or encapsulating substances, which are suitable for administration to humans. The term "carrier" refers to an organic or inorganic component, of a natural or synthetic nature, in which the active component is combined in order to facilitate application. The components of the pharmaceutical composition of the invention are usually such that no interaction occurs which substantially impairs the desired pharmaceutical efficacy.

The pharmaceutical compositions of the invention may contain suitable buffer substances such as acetic acid in a salt, citric acid in a salt, boric acid in a salt
5 and phosphoric acid in a salt.

The pharmaceutical compositions may, where appropriate, also contain suitable preservatives such as benzalkonium chloride, chlorobutanol, parabens and
10 thimerosal.

The pharmaceutical compositions are usually provided in a uniform dosage form and may be prepared in a manner known per se. Pharmaceutical compositions of the
15 invention may be in the form of capsules, tablets, lozenges, suspensions, syrups, elixir or in the form of an emulsion, for example.

Compositions suitable for parenteral administration
20 usually comprise a sterile aqueous or nonaqueous preparation of the active compound, which is preferably isotonic to the blood of the recipient. Examples of compatible carriers and solvents are Ringer solution and isotonic sodium chloride solution. In addition,
25 usually sterile, fixed oils are used as solution or suspension medium.

The present invention is described in detail by the figures and examples below, which are used only for
30 illustration purposes and are not meant to be limiting. Owing to the description and the examples, further embodiments which are likewise included in the invention are accessible to the skilled worker.

35 **Figures:**

Fig. 1. GPR35 mRNA expression in colon tumor biopsies

RT-PCR investigations with DNA-free RNA show GPR35

expression in most of the colon tumor biopsies. By contrast, there is no detectable expression in normal tissues. (1-Breast, 2-lung, 3-lymph nodes, 4-thymus, 5-colon, 6-15 colon tumors, 16-neg. control).

5

Fig. 2. Quantitative PCR analysis of GUCY2C mRNA expression in normal and tumor tissues

Real-time PCR investigation with GUCY2C-specific primers (SEQ ID NO: 22-23) shows selective mRNA
10 expression in normal ileum, colon, and in all colon tumor biopsies. Distinct quantities of GUCY2C transcripts were also detected in a colon tumor metastasis in the liver.

15 **Fig. 3. Identification of tumor-specific GUCY2C splice variants**

PCR products from normal colon tissues and colon tumors were cloned, and clones from both groups were checked by restriction analysis (EcoR I) and sequenced.

20

Fig. 4. Selective SCGB3A expression in normal lung and lung tumors

RT-PCR analysis with gene-specific SCGB3A2 primers (SEQ ID NO: 37, 38) shows cDNA amplification
25 exclusively in normal lung (lane 8, 14-15) and in lung tumor biopsies (lane 16-24). (1-15 normal tissues, 1: Liver, 2: PBMC, 3: lymph node, 4: stomach, 5: testis, 6: breast, 7: kidney, 8: lung, 9: thymus, 10: ovary, 11: adrenal, 12: spleen, 14-15: lung, 16-24: lung
30 tumors, 25: negative control).

Fig. 5. Claudin-18A2.1 expression in stomach and esophagus, as well as stomach and pancreas tumors

RT-PCR analysis with claudin-18A2.1-specific primers
35 (SEQ ID NO: 39, 40) showed according to the invention pronounced claudin-18A2.1 expression in 8/10 stomach tumor biopsies and in 3/6 pancreatic tumor biopsies. Distinct expression was also detected in stomach and

esophageal normal tissues. In contrast thereto, no expression was detected in the ovary and in ovarian tumors.

5 **Fig. 6. SLC13A1 expression in the kidney and renal tumors**

RT-PCR analysis with SLC13A1-specific primers (SEQ ID NO: 49, 50) showed expression in 7/8 renal tumor samples. Otherwise, transcripts within normal
10 tissues were detected exclusively in the kidney. (1-2: kidney normal tissue, 3-10: renal tumors, 11: breast, 12: lung, 13: liver, 14: colon, 15: lymph node, 16: spleen, 17: esophagus, 18: thymus, 19: thyroid, 20: PBMC, 21: ovary, 22: testis normal tissues).

15

Fig. 7. CLCA1 expression in colon normal tissue, as well as colon and stomach tumors

RT-PCR investigations with CLCA1-specific primers (SEQ ID NO: 67, 68) confirmed selective expression in
20 the colon and showed high expression in 3/7 investigated colon carcinoma and 1/3 investigated stomach tumor samples. The other normal tissues showed no or only very weak expression.

25 **Fig. 8. FLJ21477 expression in the colon tumor tissue and colon tumors**

RT-PCR investigations with FLJ21477-specific primers (SEQ ID NO: 69, 70) showed selective expression in the colon and additionally various levels of expression in
30 7/12 investigated colon tumor samples. The other normal tissues showed no expression.

Fig. 9. FLJ20694 expression in the colon normal tissue and colon tumors

35 RT-PCR investigations with FLJ20694-specific primers (SEQ ID NO: 71, 72) showed selective expression in the colon and additionally various levels of expression in 5/9 investigated colon tumor samples. The other normal

tissues showed no expression.

Fig. 10. von Ebner expression in stomach and lung normal tissues, as well as lung tumors

5 RT-PCR investigations with von Ebner-specific primers (SEQ ID NO: 73, 74) showed selective expression in the stomach, in the lung and in 5/10 investigated lung tumor samples. The other normal tissues showed no expression.

10

Fig. 11. Plunc expression in thymus and lung normal tissues, as well as lung tumors

RT-PCR investigations with Plunc-specific primers (SEQ ID NO: 75, 76) showed selective expression in the
15 thymus, in the lung and in 6/10 investigated lung tumor samples. The other normal tissues showed no expression.

Fig. 12. SLC26A9 expression in lung, lung tumors and thyroid

20 RT-PCR investigations with SLC26A9-specific primers (SEQ ID NO: 77, 78) showed selective expression in the lung and in all (13/13) investigated lung tumor samples. The other normal tissues showed no expression with the exception of the thyroid.

25

Fig. 13. THC1005163 expression in stomach, ovary, lung and lung tumors

RT-PCR investigations with a THC1005163-specific primer (SEQ ID NO: 79) and a nonspecific oligo dT tag primer
30 showed expression in stomach, ovary, lung and in 5/9 lung tumor biopsies. The other normal tissues showed no expression.

Fig. 14. LOC134288 expression in kidney and renal tumors

35 RT-PCR investigations with LOC134288-specific primers (SEQ ID NO: 80, 81) showed selective expression in the kidney and in 5/8 investigated renal tumor biopsies.

Fig. 15. THC943866 expression in kidney and renal tumors

RT-PCR investigations with THC943866-specific primers
5 (SEQ ID NO: 82, 83) showed selective expression in the
kidney and in 4/8 investigated renal tumor biopsies.

Fig. 16. FLJ21458 expression in colon and colon tumors

RT-PCR investigations with FLJ21458-specific primers
10 (SEQ ID NO: 86, 87) showed selective expression in the
colon and in 7/10 investigated colon tumor biopsies.
(1-2: colon, 3: liver, 4: PBMC, 5: spleen, 6: prostate,
7: kidney, 8: ovary, 9: skin, 10: ileum, 11: lung, 12:
testis normal tissues, 13-22: colon tumors, 23: neg.
15 control).

Fig. 17. Cellular localization of GPR35

Immunofluorescence for detecting the cellular
localization of GPR35 after transfection of a plasmid
20 that expresses a GPR35-GFP fusion protein. The arrows
identify the membrane-associated fluorescence of the
fluorescent GFP.

Fig. 18. Quantitative expression of GPR35

- 25 A. Quantitative RT-PCR with GPR35-specific primers
(SEQ ID NO: 88, 89) show selective expression in
various regions of the intestine, in colon tumor
samples and in metastases from colon tumors. The
following normal tissues were analyzed: liver,
30 lung, lymph nodes, stomach, spleen, adrenal,
kidney, esophagus, ovary, testis, thymus, skin,
breast, pancreas, lymphocytes, activated
lymphocytes, prostate, thyroid, ovary,
endometrium, cerebellum, brain.
- 35 B. Prevalence of GPR35 in colon tumors and metastases
thereof. GPR35 is expressed both in the tumor and
in metastases in more than 90% of the cases.

Fig. 19. Quantitative expression of GUCY2C

Quantitative RT-PCR with GUCY2C-specific primers (SEQ ID NO: 98, 99) show high and selective expression in normal colonic and gastric tissue (A) and GUCY2C-specific expression in colonic and gastric tumor samples (B). GUCY2C is detectable in 11/12 colon tumors and in 7/10 stomach tumors.

Fig. 20. Quantitative expression of SCGB3A2

Quantitative RT-PCR with SCGB3A2-specific primers (SEQ ID NO: 103, 104) show selective expression in lung samples and lung tumor samples. 19/20 lung tumor samples are SCGB3A2-positive, and SCGB3A2 is over-expressed by a factor of at least 10 in more than 50% of the samples. The following normal tissues were analyzed: liver, lung, lymph nodes, stomach, spleen, adrenal, kidney, esophagus, ovary, testis, thymus, skin, breast, pancreas, lymphocytes, activated lymphocytes, prostate, thyroid, ovary, endometrium, cerebellum, brain.

Fig. 21. Immunofluorescence with SCGB3A2-specific antibodies

COS7 cells were transfected with a plasmid which codes for an SCGB3A2-GFP fusion protein. A. Detection of the transfected fusion protein with an SCGB3A2-specific rabbit antiserum (immunization with SEQ ID NO: 105). B. Detection of the transfected fusion protein by GFP fluorescence. C. Superimposition of the two fluorescences from A and B. The yellow color is produced at the points where the two fluorescences are superimposed and thus demonstrates the specificity of the SCGB3A2 antiserum.

Fig. 22. Diagrammatic depiction of claudin-18 conformations

According to the invention, the claudin-18A2 polypeptide can exist on the cell in two conformations.

5 In conformation 1, the protein is present as membrane molecule having four transmembrane domains (TM) and two separate, extracellularly localized domains. In conformation 2, the two hydrophobic regions in the middle (h-phob) do not exert a transmembrane domain
10 function. Thus, in this conformation, compared to conformation 1, additional peptide regions are located extracellularly. In addition, an additional N glycosylation site results in this conformation at position 116 (thicker arrow). All predicted
15 glycosylation domains are shown in the lower part of the figure. Ex1: extracellular domain 1, Ex2: extracellular domain 2, TM: transmembrane domain, H-phob: extracellular hydrophobic region.

20 **Fig. 23. Quantitative expression of claudin-18, variant A1**

Claudin-18A1 is detectable in no normal tissue except lung and stomach tissue. Claudin-18A1 is highly expressed in a large number of tumor tissues.
25 Particularly strong expression is found in gastric tumors, lung tumors, pancreatic tumors and esophageal tumors.

30 **Fig. 24. Quantitative expression of claudin-18, variant A2**

Claudin-18A2 is detectable in no normal tissue except stomach tissue. Claudin-18A2 is highly expressed in a large number of tumor tissues, in particular gastric tumors, lung tumors, pancreatic tumors and esophageal
35 tumors.

Fig. 25. Use of claudin-18A2-specific antibodies (extracellular domain)

A: Staining of claudin-18A2-positive gastric tumor cells (SNU-16, fixed with methanol) with an antibody which was produced by immunization with a peptide (SEQ ID NO: 17). Membrane staining appears particularly
5 strongly in the cell/cell interaction regions. The protein aggregates in focal membrane regions.

B, C, D: Demonstration of the specificity of the antibody by colocalization analysis in claudin-18A2-GFP-transfected 293T cells. B: GFP fluorescence; C:
10 anti-claudin-18A2; D: superimposition.

Fig. 26. Use of claudin-18A2-specific antibodies (extracellular domain)

Membrane staining of claudin-18A2-positive gastric
15 tumor cells (SNU-16) with an antibody which was produced by immunization with a peptide (SEQ ID NO: 113, N-terminally located extracellular domain). A monoclonal antibody which is directed against E-cadherin was used for counterstaining. A:
20 claudin-18A2 antibody; B: anti-E-cadherin counterstaining; C: superimposition.

Fig. 27. Use of antibodies against the C-terminal extracellular domain of claudin-18

25 Left figures: Membrane staining of claudin-18A2-positive gastric tumor cells (SNU-16) with an antibody which was produced by immunization with a peptide (SEQ ID NO: 116, C-terminally located extracellular domain). A monoclonal antibody which is directed
30 against E-cadherin was used for counterstaining (right figures).

Fig. 28. Use of claudin-18A1-specific antibodies

Top: Weak to absent staining of gastric tumor cells
35 (SNU-16; claudin18A2 positive) with an antibody which was produced by immunization with a claudin-18A1-specific peptide (SEQ ID NO: 115). A: anti-E-cadherin; B: anti-claudin-18A1; C: superimposition.

Below: Demonstration of the specificity of the antibody by colocalization analysis in claudin-18A1-GFP-transfected 293T cells. A: GFP fluorescence; B: anti-claudin-18A1; C: superimposition.

5

Fig. 29. Detection of claudin-18A2 in a Western blot.

Western blotting with lysates from various healthy tissues with a claudin-18A2-specific antibody directed against the epitope with SEQ ID NO: 17. 1: Stomach; 2: testis; 3: skin; 4: breast; 5: liver; 6: colon; 7: lung; 8: kidney; 9: lymph node normal tissues.

Fig. 30. Claudin-18A2 Western blotting with samples from stomach and stomach tumors, as well as different tumor cell lines

Lysates from stomach and stomach tumors (A, B) and tumor cell lines (C, D) were blotted and tested using a claudin-18A2-specific antibody against the epitope having SEQ ID NO: 17. Stomach tumors show a less glycosylated form of claudin-18A2. PNGase F treatment of stomach lysates leads to the formation of the low-glycosylated form.

A: 1: stomach normal tissue #A; 2: stomach tumor #A; 3: stomach normal tissue #B; 4: stomach tumor #B

B: 1: stomach normal tissue #A; 2: stomach normal tissue #B; 3: stomach normal tissue #B + PNGase F; 4: stomach tumor #C; 5: stomach tumor #D; 6: stomach tumor #D + PNGase F

C: 1: stomach normal tissue; 2: MDA-MB-231; 3: SK-MEL-37; 4: AGS; 5: SNU-1; 6: SNU-16; 7: EFO27; 8: TOV-112D; 9: OVCAR. Note that the tumor cell lines express the deglycosylated variant of claudin-18A2.

D: Summary table of the Western blot data for a selection of cell lines which have been tested using the claudin-18A2 specific antibody.

Fig. 31. Expression of claudin-18 in lung tumors

Low-glycosylated claudin-18A2 variants were detected in

lung tumors in accordance with fig. 30. 1: Stomach normal tissue; 2: stomach tumor; 3-9: lung tumors.

5 **Fig. 32. Immunohistochemical analysis of claudin-18 using claudin-18A2-specific antibodies in normal tissues**

In gastric mucosa only differentiated epithelial cells at the orifice as well as at the bottom of the glands
10 are stained. Claudin-18A2 is not detectable in stem cells of the stomach. All other investigated normal tissues also do not express this gene such as, for example, shown for kidney, lung and colon.

15 **Fig. 33. Results of the immune histology using claudin-18A2 specific polyclonal antiserum.**

A: Examples for specific staining of lung tumor tissues. Note that the normal lung tissue expressing the variant claudin-18A1 is not recognized by the
20 claudin-18A2 specific antiserum.

B: Examples for specific tumor staining of esophageal tumors. Note that healthy cells in the vicinity are not stained.

C: Examples for specific tumor staining of stomach
25 tumor epithelia. Also here healthy cells in the vicinity are not stained.

D: Exemplary summary table of immunohistochemical staining data using claudin-18A2 specific antibodies. AdenoCa: adenocarcinoma; SCC: squamous epithelium
30 carcinoma; RCC: renal cell carcinoma.

Fig. 34. Quantitative expression of SLC13A1

Quantitative RT-PCR with SLC13A1-specific primers (SEQ ID NO: 121, 122) show high and selective
35 expression in normal kidney tissue (A) and SLC13A1-specific expression in renal tumors (B). SLC13A1 transcription is detectable in 5/8 renal tumors.

Fig. 35. Cellular localization of SLC13A1

Immunofluorescence to demonstrate the cellular localization of SLC13A1 after transfection of a plasmid which provides an SLC13A1-GFP fusion protein. The
5 membrane-associated fluorescence of the SLC13A1 fusion protein is to be seen clearly (as ring around the transfected cell).

Fig. 36. Quantitative expression of CLCA1

10 Quantitative RT-PCR with CLCA1-specific primers (SEQ ID NO: 125, 126) show high and selective expression in normal colonic tissue and stomach tissue (A) and CLCA1-specific expression in colonic and gastric tumor samples (B). CLCA1 is detectable in 6/12
15 colon tumors and in 7/10 stomach tumors.

Fig. 37. Quantitative expression of FLJ21477

Quantitative RT-PCR with FLJ21477-specific primers (SEQ ID NO: 127, 128) show high and selective expres-
20 sion in normal colonic and gastric tissue and weak expression in thymus, esophagus and brain (A) and the FLJ21477-specific expression in colonic tumor samples (B). FLJ21477 is detectable in 11/12 colon tumors.

Fig. 38. Quantitative expression of FLJ20694

25 Quantitative RT-PCR with FLJ20694-specific primers (SEQ ID NO: 129, 130) show high and selective expres- sion in normal colonic and gastric tissue (A) and FLJ20694-specific overexpression in colonic and gastric
30 tumor samples (B). FLJ20694 is detectable in 11/12 colon tumors and in 7/10 stomach tumors.

Fig. 39. Quantitative expression of FLJ21458

Quantitative RT-PCR with FLJ21458-specific primers
35 (SEQ ID NO: 133, 134) show selective expression in testis, gastric tissue and different intestinal areas. In addition, FLJ21458-specific transcripts were detectable in 20/20 colonic tumors and in 7/11 colonic

metastases. The following normal tissues were analyzed:
liver, lung, lymph nodes, spleen, adrenal, kidney,
esophagus, ovary, testis, thymus, skin, breast,
pancreas, lymphocytes, activated lymphocytes, prostate,
5 thyroid, ovary, endometrium, cerebellum, brain.

Fig. 40. Immunofluorescence with FLJ21458-specific antibodies

Top: 293 cells were transfected with a plasmid which
10 codes for an FLJ21458-GFP fusion protein. A: detection
of the transfected fusion protein with an FLJ21458-
specific rabbit antiserum (immunization with
SEQ ID NO: 136). B: detection of the transfected fusion
protein by GFP fluorescence. C: superimposition of the
15 two fluorescences from A and B. The yellow color is
produced at the points where the two fluorescences are
superimposed and thus demonstrates the specificity of
the FLJ21458 antiserum.

Below: Analysis of Snul6 cells which endogenously
20 synthesize FLJ21458. A: protein detection using an
FLJ21458-specific rabbit antiserum (immunization with
SEQ ID NO: 136). B: detection of the membrane protein
E-cadherin. C: superimposition of the two fluorescences
from A and B. The yellow color is produced at the
25 points where the two fluorescences are superimposed,
and demonstrates the membrane localization of FLJ21458.

Fig. 41. Sequences

The sequences to which reference is made herein are
30 shown.

Fig. 42. Determination of extracellular regions of claudin-18A2

Three constructs were prepared which each had a marker
35 sequence (myc or HA tag) in one of the domains EX1 (= extracellular domain 1), EX2 (= extracellular domain 2) or D3 (= domain 3) (top). These were transfected into cell lines and then tested whether an antibody directed

against these marker sequences binds to non-permeabilized cells. This requires the respective region of the protein to be topologically extracellular. The flow-through cytometry demonstrated
5 that all three regions of the molecule are accessible for the antibody (below).

Fig. 43. Claudin-18A2 membrane topology

According to our data, claudin-18A2 can exist in
10 conformation 2 wherein the inner two hydrophobic domains do not pass through the cell membrane in an integral manner. In this way, larger regions of this molecule are extracellular. Located herein are also glycosylation domains which, according to our data, are
15 glycosylated in stomach normal tissue, but not in tumors. Thus, epitopes emerge which are specific for tumor tissue.

**Fig. 44. FACS analysis for determining the
20 extracellular localization of claudin-18.**

The figure shows flow-through cytometric analyses with non-permeabilized cells transfected with full-length claudin-18A1, claudin-18A2 and Mock transfected as well as transfected with portions of claudin-18A2. It is
25 shown that the antibodies mAB1 and mAB2 recognize specifically claudin-18A2 (left column) and the extracellular domain 2 (Ex2, third column) on the cell surface, while claudin-18A1 (second column) and the negative control (last column) are negative. The
30 antibody mAB1 in contrast to mAB2 also binds specifically to the extracellular domain 1 (Ex1, fourth column).

Examples:**Material and methods**

5

The terms "*in silico*", "electronic" and "virtual cloning" refer solely to the utilization of methods based on databases, which may also be used to simulate laboratory experimental processes.

10 Unless expressly defined otherwise, all other terms and expressions are used so as to be understood by the skilled worker. The techniques and methods mentioned are carried out in a manner known per se and are described, for example, in Sambrook et al., Molecular
15 Cloning: A Laboratory Manual, 2nd Edition (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. All methods including the use of kits and reagents are carried out according to the manufacturers' information.

20

Datamining-based strategy for determining new tumor-associated genes

Two *in silico* strategies, namely GenBank keyword search and the cDNAxProfiler, were combined. Utilizing the
25 NCBI ENTREZ Search and Retrieval System (<http://www.ncbi.nlm.nih.gov/Entrez>), a GenBank search was carried out for candidate genes annotated as being specifically expressed in specific tissues (Wheeler et al., *Nucleic Acids Research* 28:10-14, 2000).

30 Carrying out queries with keywords such as "colon-specific gene", "stomach-specific gene" or "kidney-specific gene", candidate genes (GOI, genes of interest) were extracted from the databases. The search was restricted to part of the total information of
35 these databases by using the limits "homo sapiens", for the organism, and "mRNA", for the type of molecule.

The list of the GOI found was curated by determining different names for the same sequence and eliminating such redundancies.

All candidate genes obtained by the keyword search were
5 in turn studied with respect to their tissue
distribution by the "electronic Northern" (eNorthern)
method. The eNorthern is based on aligning the sequence
of a GOI with an EST (expressed sequence tag) database
(Adams et al., *Science* 252:1651, 1991)
10 (<http://www.ncbi.nlm.nih.gov/BLAST>). The tissue origin
of each EST which is found to be homologous to the
inserted GOI can be determined and in this way the sum
of all ESTs produces a preliminary assessment of the
tissue distribution of the GOI. Further studies were
15 carried out only with those GOI which had no homologies
to EST from non organ-specific normal tissues. This
evaluation also took into account that the public
domain contains wrongly annotated cDNA libraries
(Scheurle et al., *Cancer Res.* 60:4037-4043, 2000)
20 ([www.fau.edu/cmabb/publications/](http://www.fau.edu/cmabb/publications/cancergenes6.htm)
[cancergenes6.htm](http://www.fau.edu/cmabb/publications/cancergenes6.htm)).

The second datamining method utilized was the **cDNA
xProfiler** of the NCBI Cancer Genome Anatomy Project
(<http://cgap.nci.nih.gov/Tissues/xProfiler>) (Hillier et
25 al., *Genome Research* 6:807-828, 1996; Pennisi, *Science*
276:1023-1024, 1997). This allows pools of
transcriptomes deposited in databases to be related to
one another by logical operators. We have defined a
pool A to which all expression libraries prepared for
30 example from colon were assigned, excluding mixed
libraries. All cDNA libraries prepared from normal
tissues other than colon were assigned to pool B.
Generally, all cDNA libraries were utilized
independently of underlying preparation methods, but
35 only those with a size > 1000 were admitted. Pool B was
digitally subtracted from pool A by means of the BUT
NOT operator. The set of GOI found in this manner was

also subjected to eNorthern studies and validated by a literature research.

This combined datamining includes all of the about 13 000 full-length genes in the public domain and
5 predicts out of these genes having potential organ-specific expression.

All other genes were first evaluated in normal tissues by means of specific RT-PCR. All GOI which had proved
10 to be expressed in non-organ specific normal tissues had to be regarded as false-positives and were excluded from further studies. The remaining ones were studied in a large panel of a wide variety of tumor tissues. The antigens depicted below proved here to be activated
15 in tumor cells.

RNA extraction, preparation of poly-d(T) primed cDNA and conventional RT-PCR analysis

Total RNA was extracted from native tissue material by
20 using guanidium isothiocyanate as chaotropic agent (Chomczynski & Sacchi, *Anal. Biochem.* 162:156-9, 1987). After extraction with acidic phenol and precipitation with isopropanol, said RNA was dissolved in DEPC-treated water.

25 First strand cDNA synthesis from 2-4 µg of total RNA was carried out in a 20 µl reaction mixture by means of Superscript II (Invitrogen), according to the manufacturer's information. The primer used was a dT(18) oligonucleotide. Integrity and quality of the
30 cDNA were checked by amplification of p53 in a 30 cycle PCR (sense CGTGAGCGCTTCGAGATGTTCCG, antisense CCTAACCAGCTGCCCAACTGTAG, hybridization temperature 67°C).

An archive of first strand cDNA was prepared from a
35 number of normal tissues and tumor entities. For expression studies, 0.5 µl of these cDNAs was amplified in a 30 µl reaction mixture, using GOI-specific primers (see below) and 1 U of HotStarTaq DNA polymerase

(Qiagen). Each reaction mixture contained 0.3 mM dNTPs, 0.3 μ M of each primer and 3 μ l of 10 $\times$ reaction buffer. The primers were selected so as to be located in two different exons, and elimination of the interference by
 5 contaminating genomic DNA as the reason for false-positive results was confirmed by testing nonreverse-transcribed DNA as template. After 15 minutes at 95°C to activate the HotStarTaq DNA polymerase, 35 cycles of PCR were carried out (1 min at 94°C, 1 min at the
 10 particular hybridization temperature, 2 min at 72°C and final elongation at 72°C for 6 min).
 20 μ l of this reaction were fractionated and analyzed on an ethidium bromide-stained agarose gel.

15 The following primers were used for expression analysis of the corresponding antigens at the hybridization temperature indicated.

GPR35 (65°C)

20 Sense: 5'-AGGTACATGAGCATCAGCCTG-3'

Antisense: 5'-GCAGCAGTTGGCATCTGAGAG-3'

GUCY2C (62°C)

Sense: 5'-GCAATAGACATTGCCAAGATG-3'

Antisense: 5'-AACGCTGTTGATTCTCCACAG-3'

25 SCGB3A2 (66°C)

Sense: 5'-CAGCCTTTGTAGTTACTCTGC-3'

Antisense: 5'-TGTCACACCAAGTGTGATAGC-3'

Claudin18A2 (68°C)

Sense1: 5'-GGTTCGTGGTTTCACTGATTGGGATTGC-3'

30 Antisense1: 5'-CGGCTTTGTAGTTGGTTTCTTCTGGTG-3'

Sense2: 5'-TGTTTTCAACTACCAGGGGC-3'

Antisense2: 5'-TGTTGGCTTTGGCAGAGTCC-3'

Claudin18A1 (64°C)

Sense: 5'-GAGGCAGAGTTCAGGCTTCACCGA-3'

35 Antisense: 5'-TGTTGGCTTTGGCAGAGTCC-3'

SLC13A1 (64°C)

Sense: 5'-CAGATGGTTGTGAGGAGTCTG-3'

Antisense: 5'-CCAGCTTTAACCATGTCAATG-3'

CLCA1 (62°C)

Sense: 5'-ACACGAATGGTAGATACAGTG-3'

5 Antisense: 5'-ATACTTGTGAGCTGTTCCATG-3'

FLJ21477 (68°C)

Sense: 5'-ACTGTTACCTTGCATGGACTG-3'

Antisense: 5'-CAATGAGAACACATGGACATG-3'

FLJ20694 (64°C)

10 Sense: 5'-CCATGAAAGCTCCATGTCTA-3'

Antisense: 5'-AGAGATGGCACATATTCTGTC

Ebner (70°C)

Sense: 5'-ATCGGCTGAAGTCAAGCATCG-3'

Antisense: 5'-TGGTCAGTGAGGACTCAGCTG-3'

15 Plunc (55°C)

Sense: 5'-TTTCTCTGCTTGATGCACTTG-3'

Antisense: 5'-GTGAGCACTGGGAAGCAGCTC-3'

SLC26A9 (67°C)

Sense: 5'-GGCAAATGCTAGAGACGTGA-3'

20 Antisense: 5'-AGGTGTCCTTCAGCTGCCAAG-3'

THC1005163 (60°C)

Sense: 5'-GTTAAGTGCTCTCTGGATTTG-3'

LOC134288 (64°C)

Sense: 5'-ATCCTGATTGCTGTGTGCAAG-3'

25 Antisense: 5'-CTCTTCTAGCTGGTCAACATC-3'

THC943866 (59°C)

Sense: 5'-CCAGCAACAACCTTACGTGGTC-3'

Antisense: 5'-CCTTTATTCAACCAATCACTC-3'

FLJ21458 (62°C)

30 Sense: 5'-ATTCATGGTTCCAGCAGGGAC-3'

Antisense: 5'-GGGAGACAAAGTCACGTACTC-3'

**Preparation of random hexamer-primed cDNA and
quantitative real-time PCR**

The expression of several genes was quantified by real-time PCR. The PCR products were detected using SYBR Green as intercalating reporter dye. The reporter fluorescence of SYBR Green is suppressed in solution and the dye is active only after binding to double-stranded DNA fragments. The increase in the SYBR Green fluorescence as a result of the specific amplification using GOI-specific primers after each PCR cycle is utilized for quantification. Expression of the target gene is quantified absolutely or relative to the expression of a control gene with constant expression in the tissues to be investigated. Expression was measured after standardization of the samples against 18s RNA as so-called housekeeping gene using the $\Delta\Delta\text{-C}_t$ method (PE Biosystems, USA). The reactions were carried out in duplicates and determined in triplicates. The QuantiTect SYBR Green PCR kit (Qiagen, Hilden) was used in accordance with the manufacturer's instructions. The cDNA was synthesized using the high capacity cDNA Archive Kit (PE Biosystems, USA) with use of hexamer primers in accordance with the manufacturer's instructions. Each 5 μl portions of the diluted cDNA were employed in a total volume of 25 μl for the PCR: sense primer 300 nM, antisense primer 300 nM; initial denaturation 95°C for 15 min; 95°C for 30 sec; annealing for 30 sec; 72°C for 30 sec; 40 cycles. The sequences of the primers used are indicated in the respective examples.

30 **Cloning and sequence analysis**

Cloning of full-lengths and gene fragments took place by conventional methods. To ascertain the sequence, corresponding antigens were amplified using the proofreading polymerase pfu (Stratagene). After completion of the PCR, adenosine was ligated by means of HotStarTaq DNA polymerase to the ends of the amplicon in order to clone the fragments in accordance with the manufacturer's instructions into the TOPO-TA

vector. The sequencing was carried out by a commercial service. The sequences were analysed using conventional prediction programs and algorithms.

5 **Western blotting**

Cells from cell culture (endogenous expression of the target gene or synthesis of the target protein after transfection of an expression vector which encodes the target protein) or tissue samples which might contain the target protein are lysed in a 1% SDS solution. The SDS denatures the proteins present in the lysate. The lysates of an experimental mixture are fractionated according to size by electrophoresis on 8-15% denaturing polyacrylamide gels (containing 1% SDS) depending on the expected protein size (SDS polyacrylamide gel electrophoresis, SDS-PAGE). The proteins are then transferred by the semi-dry electroblotting method (Biorad) to nitrocellulose membrane (Schleicher & Schüll) on which the desired protein can be detected. For this purpose, the membrane is initially blocked (e.g. with milk powder) and then incubated with the specific antibody in a dilution of 1:20-1:200 (depending on the specificity of the antibody) for 60 minutes. After a washing step, the membrane is incubated with a second antibody coupled to a marker (e.g. enzymes such as peroxidase or alkaline phosphatase) which recognizes the first antibody. After a further washing step, subsequently the target protein is visualized in a color or chemiluminescence reaction on the membrane by means of an enzyme reaction (e.g. ECL, Amersham Bioscience). The result is documented by photographing with a suitable camera.

Analysis of protein modifications usually takes place by Western blotting. Glycosilations, which usually have a size of several kDa, lead to a larger total mass of the target protein, which can be fractionated in the SDS-PAGE. To detect specific O- and N-glycosidic

linkages, protein lysates from tissues or cells are incubated before denaturation by SDS with O- or N-glycosidases (in accordance with their respective manufacturer's instructions, e.g. PNGase, endoglycosidase F, endoglycosidase H, Roche Diagnostics). This is followed by Western blotting as described above. Thus, if there is a reduction in the size of a target protein after incubation with a glycosidase it is possible to detect a specific glycosilation and, in this way, also analyse the tumor specificity of a modification. The exact position of the glycosilated amino acid can be predicted with algorithms and prediction programs.

15 **Immunofluorescence**

Cells of established cell lines which either synthesize the target protein endogenously (detection of the RNA in RT-PCR or of the protein by Western blotting) or else have been transfected with plasmid DNA before the IF are used. A wide variety of methods (e.g. electroporation, liposome-based transfection, calcium phosphate precipitation) are well established for transfecting cell lines with DNA (e.g. Lemoine et al. Methods Mol. Biol. 1997; 75: 441-7). The transfected plasmid may in the immunofluorescence encode the unmodified protein or else couple various amino acid markers to the target protein. The most important markers are, for example, the fluorescing "green fluorescent protein" (GFP) in its various differentially fluorescing forms and short peptide sequences of 6-12 amino acids for which high-affinity and specific antibodies are available. Cells which synthesize the target protein are fixed with paraformaldehyde, saponin or methanol. The cells can then if required be permeabilized by incubation with detergents (e.g. 0.2% Triton X-100). After the fixation/permeabilization, the cells are incubated with a primary antibody which is directed against the target

protein or against one of the coupled markers. After a washing step, the mixture is incubated with a second antibody coupled to a fluorescent marker (e.g. fluorescein, Texas Red, Dako) which binds to the first antibody. The cells labeled in this way are then covered with a layer of glycerol and analysed with the aid of a fluorescence microscope according to the manufacturer's instructions. Specific fluorescence emissions are achieved in this case by specific excitation depending on the substances employed. The analysis normally allows reliable localization of the target protein, the antibody quality and the target protein being confirmed in double stainings to stain in addition to the target protein also the coupled amino acid markers or other marker proteins whose localization has been described in the literature. GFP and its derivatives represents a special case that can be directly excited and itself fluoresces, so that no antibodies are necessary for the detection.

20

Immunohistochemistry

IHC serves specifically for (1) being able to estimate the amount of target protein in tumor and normal tissues, (2) analysing how many cells in the tumor and healthy tissue synthesize the target gene, and/or (3) defining the cell type in a tissue (tumor, healthy cells) in which the target protein is detectable.

Different protocols must be used depending on the individual antibody (e.g. "Diagnostic Immunohistochemistry by David J., MD Dabbs ISBN: 0443065667" or in "Microscopy, Immunohistochemistry, and Antigen Retrieval Methods: For Light and Electron Microscopy ISBN: 0306467704").

Immunohistochemistry (IHC) on specific tissue samples serves to detect protein in the corresponding tissue. The aim of this method is to identify the localization of a protein in a functionally intact tissue aggregate.

IHC serves specifically for (1) being able to estimate the amount of target protein in tumor and normal tissues, (2) analysing how many cells in tumor and healthy tissue synthesize the target gene, and (3) defining the cell type in a tissue (tumor, healthy cells) in which the target protein is detectable. Alternatively, the amounts of protein of a target gene can be quantified by tissue immunofluorescence using a digital camera and suitable software (e.g. Tillvision, Till-photonics, Germany). The technology has frequently been published, and details of staining and microscopy can therefore be found for example in "Diagnostic Immunohistochemistry" by David J., MD Dabbs ISBN: 0443065667 or "Microscopy, Immunohistochemistry, and Antigen Retrieval Methods: For Light and Electron Microscopy" ISBN: 0306467704. It should be noted that, because of the properties of antibodies, different protocols have to be used (an example is described below) in order to obtain a valid result.

20

Ordinarily, histologically defined tumor tissues and, as reference, comparable healthy tissues are employed in the IHC. It is moreover possible to use as positive and negative controls cell lines in which the presence of the target gene is known through RT-PCR analyses. A background control must always be included.

Fixed tissue (e.g. fixation with aldehyde-containing substances, formaldehyde, paraformaldehyde or in alcoholic solutions) or shock-frozen tissue pieces with a thickness of 1-10 µm are applied to a glass support. Paraffin-embedded samples are deparaffinated for example with xylene. The samples are washed with TBS-T and blocked in serum. This is followed by incubation with the first antibody (dilution: 1:2 to 1:2000) for 1-18 hours, with affinity-purified antibodies normally being used. A washing step is followed by incubation with a second antibody which is coupled to an alkaline

phosphatase (alternative: for example peroxidase), and is directed against the first antibody, for about 30-60 minutes. This is followed by color reaction using color substrates which are converted by the bound enzymes (cf. for example, Shi et al., *J. Histochem. Cytochem.* 39: 741-748, 1991; Shin et al., *Lab. Invest.* 64: 693-702, 1991). To demonstrate the antibody specificity, the reaction can be blocked by previous addition of the immunogen.

10

Immunization

(See also Monoclonal Antibodies: A Practical Approach by Philip Shepherd, Christopher Dean isbn 0-19-963722-9; Antibodies: A Laboratory Manual by Ed Harlow, David Lane ISBN: 0879693142; Using Antibodies: A Laboratory Manual: Portable Protocol NO. by Edward Harlow, David Lane, Ed Harlow ISBN: 0879695447).

The process for preparing antibodies is described briefly below, and details can be found in the cited publications. Firstly, animals (e.g. rabbits) are immunized by a first injection of the desired target protein. The animal's immune response to the immunogen can be enhanced by a second or third immunization within a defined period (about 2-4 weeks after the preceding immunization). Again after various defined periods (first bleeding after 4 weeks, then about every 2 weeks with a total of up to 5 samplings), blood is taken from the animals, and an immune serum is obtained therefrom.

The animals are usually immunized by one of four well-established methods, with other methods also being available. It is moreover possible to immunize with peptides which are specific for the target protein, with the complete protein or with extracellular partial sequences of a protein which can be identified experimentally or via prediction programs.

(1) In the first case, peptides (length: 8-12 amino acids) conjugated to KLH (keyhole

limpet hemocyanin) are synthesized by a standardized in vitro method, and these peptides are used for the immunization. Usually, 3 immunizations are carried out with a concentration of 5-1000 µg/immunization. The immunization can also be carried out as service from service providers.

5

10

15

20

25

30

35

(2) Alternatively, the immunization can be carried out with recombinant proteins. For this purpose, the cloned DNA of the target gene is cloned into an expression vector, and the target protein is synthesized in analogy to the conditions of the particular manufacturer (e.g. Roche Diagnostics, Invitrogen, Clontech, Qiagen) for example cell-free in vitro, in bacteria (e.g. E. coli), in yeast (e.g. S. pombe), in insect cells or in mammalian cells. After synthesis in one of the systems, the target protein is purified, the purification in this case usually taking place by standardized chromatographic methods. It is also possible in this connection to use for the immunization proteins which have a molecular anchor as aid for purification (e.g. His tag, Qiagen; FLAG tag, Roche Diagnostics; Gst fusion proteins). A large number of protocols is to be found for example in the "Current Protocols in Molecular Biology", John Wiley & Sons Ltd., Wiley Interscience.

(3) If a cell line which synthesizes the desired protein endogenously is available, this cell line can also be used to produce the specific antiserum. In this case, the immunization takes place in 1-3 injections in each case with about $1-5 \times 10^7$ cells.

(4) The immunization can also take place by injection of DNA (DNA immunization). For this

purpose, the target gene is initially cloned into an expression vector so that the target sequence is under the control of a strong eukaryotic promoter (e.g. CMV promoter).
5 Subsequently, 5-100 µg of DNA are transferred as immunogen using a "gene gun" into capillary regions with a strong blood flow in an organism (e.g. mouse, rabbit). The transferred DNA is taken up by the animal's
10 cells, the target gene is expressed, and the animal finally develops an immune response to the target gene (Jung et al., Mol Cells 12:41-49, 2001; Kasinrerk et al., Hybrid Hybridomics 21:287-293, 2002).

15

Quality control of the polyclonal serum or antibody

Assays based on cell culture with subsequent Western blotting are most suitable for demonstrating specificity (various variations are described for
20 example in "Current Protocols in Protein Chemistry", John Wiley & Sons Ltd., Wiley InterScience). For the demonstration, cells are transfected with a cDNA, which is under the control of a strong eukaryotic promoter (e.g. cytomegalovirus promoter), for the target
25 protein. A wide variety of methods (e.g. electroporation, liposome-based transfection, calcium phosphate precipitation) are well established for transfecting cell lines with DNA (e.g. Lemoine et al., *Methods Mol. Biol.* 75:441-7, 1997). It is also possible
30 alternatively to use cell lines which express the target gene endogenously (demonstration by target gene-specific RT-PCR). As control, in the ideal case homologous genes are also transfected in the experiment, in order to be able to demonstrate in the
35 following Western blot the specificity of the analysed antibody.

In the subsequent Western blot, cells from cell culture

or tissue samples which might contain the target protein are lysed in a 1% SDS solution, and the proteins are denatured thereby. The lysates are fractionated according to size by electrophoresis on
5 8-15% denaturing polyacrylamide gels (contain 1% SDS) (SDS polyacrylamide gel electrophoresis, SDS-PAGE). The proteins are then transferred by one of a plurality of blotting methods (e.g. semi-dry electroblot; Biorad) to a specific membrane (e.g. nitrocellulose, Schleicher &
10 Schüll). The desired protein can be visualized on this membrane. For this purpose, the membrane is first incubated with the antibody which recognizes the target protein (dilution about 1:20-1:200, depending on the specificity of the antibody) for 60 minutes. After a
15 washing step, the membrane is incubated with a second antibody which is coupled to a marker (e.g. enzymes such as peroxidase or alkaline phosphatase) and which recognizes the first antibody. It is then possible in a color or chemiluminescent reaction to visualize the
20 target protein on the membrane (e.g. ECL, Amersham Bioscience). An antibody with a high specificity for the target protein should in the ideal case recognize only the desired protein itself.

25 Various methods are used to confirm the membrane localization of the target protein identified in the in silico approach. An important and well-established method using the antibodies described above is immuno-
fluorescence (IF). Cells of established cell lines
30 which either synthesize the target protein (detection of the RNA in an RT-PCR or of the protein in a Western blot) or else have been transfected with plasmid DNA are used for this. A wide variety of methods (e.g. electroporation, liposome-based transfection, calcium
35 phosphate precipitation) are well established for transfection of cell lines with DNA (e.g. Lemoine et al., *Methods Mol. Biol.* 75:441-7, 1997). The plasmid transfected into the cells can in the

immunofluorescence encode the unmodified protein or else couple various amino acid markers to the target protein. The principal markers are, for example, the fluorescent "green fluorescent protein" (GFP) in its various differentially fluorescent forms, short peptide sequences of 6-12 amino acids for which high-affinity and specific antibodies are available, or the short amino acid sequence Cys-Cys-X-X-Cys-Cys which can bind via its cysteine specific fluorescent substances (Invitrogen). Cells which synthesize the target protein are fixed for example with paraformaldehyde or methanol. The cells can then, if required, be permeabilized by incubation with detergents (e.g. 0.2% Triton X-100). The cells are then incubated with a primary antibody which is directed against the target protein or against one of the coupled markers. After a washing step, the mixture is incubated with a second antibody which is coupled to a fluorescent marker (e.g. fluorescein, Texas Red, Dako) and which binds to the first antibody. The cells labeled in this way are then covered with a layer of glycerol and analysed with the aid of a fluorescence microscope according to the manufacturer's instructions. Specific fluorescence emissions are achieved in this case by specific excitation depending on the substances employed. The analysis usually permits reliable localization of the target protein, the antibody quality and the target protein being confirmed in double stainings to stain in addition to the target protein also the coupled amino acid markers or other marker proteins whose localization has already been described in the literature. GFP and its derivatives represents a special case, being excitable directly and themselves fluorescing. The membrane permeability, which can be controlled through the use of detergents, permits demonstration in the immunofluorescence of whether an immunogenic epitope is located inside or outside the cell. The prediction of the selected proteins can thus

be supported experimentally. An alternative possibility is to detect extracellular domains by means of flow cytometry. For this purpose, cells are fixed under non-permeabilizing conditions (e.g. with PBS/Na azide/2% FCS/5 mM EDTA) and analysed in a flow cytometer in accordance with the manufacturer's instructions. Only extracellular epitopes can be recognized by the antibody to be analysed in this method. A difference from immunofluorescence is that it is possible to distinguish between dead and living cells by use of, for example, propidium iodide or Trypan blue, and thus avoid false-positive results.

Affinity purification

Purification of the polyclonal sera took place in the case of the peptide antibodies entirely, or in the case of the antibodies against recombinant proteins in part, as service by the contracted companies. For this purpose, in both cases, the appropriate peptide or recombinant protein was covalently bonded to a matrix, and the latter was, after the coupling, equilibrated with a native buffer (PBS: phosphate buffered saline) and then incubated with the crude serum. After a further PBS washing step, the antibody was eluted with 100 mM glycine, pH 2.7, and the eluate was immediately neutralized in 2M TRIS, pH 8. The antibodies purified in this way could then be employed for specific detection of the target proteins both by Western blotting and by immunofluorescence.

30

Preparation of GFP transfectants

For the immunofluorescence microscopy of heterologously expressed tumor-associated antigens, the complete ORF of the antigens was cloned in pGFP-C1 and pGFP-N3 vectors (Clontech). CHO and NIH3T3 cells cultivated on slides were transfected with the appropriate plasmid constructs using Eugene transfection reagent (Roche) in accordance with the manufacturer's instructions and,

after 12-24 h, analysed by immunofluorescence microscopy.

Flow-through cytometry

- 5 Flow-through cytometric measurements were performed in a manner known *per se* (e.g. Robinson (editor) Handbook of flow cytometry methods. Wiley-Liss, New York, 1993).

Example 1: Identification of GPR35 as diagnostic and therapeutic cancer target

- 10 GPR35 (SEQ ID NO:1) and its translation product (SEQ ID NO:9) have been described as putative G protein-coupled receptor. The sequence is published in Genbank under accession No. AF089087. This transcript
15 codes for a protein of 309 amino acids with a molecular weight of 34 kDa. It was predicted that GPR35 belongs to the superfamily of G protein-coupled receptors with 7 transmembrane domains (O'Dowd et al., *Genomics* 47:310-13, 1998). In order to confirm the predicted
20 localization of GPR35 in the cell, the protein was fused to GFP as reporter molecule and, after transfection of the appropriate plasmid, expressed heterologously in 293 cells. The localization was then analysed in a fluorescence microscope. It was confirmed
25 according to the invention that GPR35 is an integral transmembrane molecule (fig. 17). Investigation to date on human GPR35 (see, inter alia, Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, Hinokio Y, Lindner TH, Mashima H, Schwarz PE, del Bosque-Plata L, Horikawa
30 Y, Oda Y, Yoshiuchi I, Colilla S, Polonsky KS, Wei S, Concannon P, Iwasaki N, Schulze J, Baier LJ, Bogardus C, Groop L, Boerwinkle E, Hanis CL, Bell GI *Nat Genet.* 2000 Oct; 26(2):163-75) suggested that GPR35 is activated in many healthy tissues. The reading frame of
35 the gene comprises a single exon. According to the invention, a gene-specific primer pair (SEQ ID NO:20, 21) for GPR35 was used in RT-PCR analyses to amplify cDNA in the colon and in colon tumors (13/26). By

contrast, no significant expression is detectable in other normal tissues. Because of the particular fact that GPR35 consists of a single exon, genomic DNA impurities cannot be detected with intron-spanning
5 primers. In order to preclude genomic contamination of the RNA samples, therefore, all RNAs were treated with DNase. GPR35 transcripts were detected according to the invention only in the colon, in the rectum, in the testis and in colon tumors using DNA-free RNA.

10

Tab. 1 GPR35 expression in normal tissues

Normal tissue	Expression
Brain	-
Cerebellum	-
Myocardium	-
Skeletal muscle	-
Rectum	++
Stomach	-
Colon	++
Pancreas	-
Kidney	-
Testis	-
Thymus	-
Mammary glands	-
Ovary	-
Uterus	n.d.
Skin	-
Lung	-
Thyroid	-
Lymph nodes	-
Spleen	-
PBMC	-
Adrenal	-
Esophagus	-
Small intestine	+
Prostate	-

(nd = not determined)

The selective and high expression of GPR35 transcripts in normal colonic tissue and in colon tumor biopsies (fig. 1) was not previously known and can be utilized according to the invention for molecular diagnostic methods such as RT-PCR for detecting disseminating tumor cells in the serum and bone marrow and for detecting metastases in other tissues. Quantitative RT-PCR with specific primers (SEQ ID NO:88 and 89) also confirms that GPR35 is a highly selective colon-specific differentiation antigen which is also contained in colon tumors and in colon tumor metastases. In some colon tumors, it is in fact overexpressed by one log compared with normal colon (fig. 18). Antibodies were produced by immunizing rabbits for detecting GPR35 protein. The following peptides were used to propagate these antibodies:

SEQ ID NO:90 GSSDLTWPPAIKLGC (AA 9-23)
SEQ ID NO:91: DRYVAVRHPLRARGLR (AA 112-127)
SEQ ID NO:92: VAPRAKAHKSQDSLC (C terminus)
SEQ ID NO:93 CFRSTRHNFNSMR (extracell. domain 2)

Stainings with these antibodies for example in a Western blot confirm the expression in tumors. All 4 extracellular domains of GPR35 (position of the predicted extracellular domains in the sequence of SEQ ID NO:9 AA 1-22 (SEQ ID NO:94); AA 81-94 (SEQ ID NO:95); AA 156-176 (SEQ ID NO:96); AA 280-309 (SEQ ID NO:97)) can be used according to the invention as target structures of monoclonal antibodies. These antibodies bind specifically to the cell surface of tumor cells and can be used both for diagnostic and for therapeutic methods. Overexpression of GPR35 in tumors provides additional support for such a use. In addition, the sequences coding for proteins can be used according to the invention as vaccine (RNA, DNA, peptide, protein) for inducing tumor-specific immune responses (T-cell and B-cell-mediated immune responses). In addition, it has surprisingly been found that a further start codon exists 5' in front of the

generally known start codon and expresses an N-terminally extended protein.

It has thus been found according to the invention that
5 GPR35, a protein which was previously described as
expressed ubiquitously, is tumor-associated
overexpressed, selectively in gastrointestinal tumors,
especially in tumors of the colon. GPR35 is therefore
suitable in particular as molecular target structure
10 for the diagnosis and treatment of these tumors.
Investigation to date of human GPR35, cf., for example,
Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara
M, Hinokio Y, Lindner TH, Mashima H, Schwarz PE, del
Bosque-Plata L, Horikawa Y, Oda Y, Yoshiuchi I, Colilla
15 S, Polonsky KS, Wei S, Concannon P, Iwasaki N, Schulze
J, Baier LJ, Bogardus C, Groop L, Boerwinkle E, Hanis
CL, Bell GI Nat Genet. 2000 Oct;26(2):163-75 suggested
that GPR35 is expressed in many healthy tissues. By
contrast, the investigations according to the invention
20 showed that GPR35 is surprisingly not significantly
detectable in most normal tissues and, in contrast
thereto, is highly activated in primary and metastatic
colon tumors. In addition, besides the described GPR35
sequence, according to the invention a novel
25 translation variant which makes use of an alternative
start codon has been found (SEQ ID NO:10).

GPR35 is a member of the group of G-coupled receptors
(GPCR), a very large protein family whose structure and
30 function has been very well investigated. GPCR are
outstandingly suitable as target structures for the
development of pharmaceutically active substances,
because the methods necessary therefor (e.g. receptor
expression, purification, ligand screening,
35 mutagenizing, functional inhibition, selection of
agonistic and antagonistic ligands, radiolabeling of
ligands) is very well developed and described in
detail, cf., for example, "G Protein-Coupled Receptors"

by Tatsuya Haga, Gabriel Bernstein and Gabriel Bernstein
ISBN: 0849333849 and in "Identification and Expression
of G-Protein Coupled Receptors Receptor Biochemistry
and Methodology" by Kevin R. Lynch ASIN: 0471183105.

5 Realization according to the invention that GPR35 is
undetectable in most healthy tissues but undergoes
tumor-associated expression on the cell surface,
enables it to be used as tumor-associated target
structure for example for pharmaceutically active
10 ligands, especially in conjugation for example with
radioactive molecules as pharmaceutical substances. It
is possible in a particular embodiment to use
radiolabeled ligands which bind to GPR35 for detecting
tumor cells or for treating colon tumors in vivo.

15

**Example 2: Identification of GUCY2C in hepatic and
ovarian tumors and novel GUCY2C splice variants as
diagnostic and therapeutic cancer targets**

20 Guanylate cyclase 2C (GUCY2C; SEQ ID NO:2; translation
product: SEQ ID NO:11) - a type I transmembrane protein
- belongs to the family of natriuretic peptide
receptors. The sequence is published in Genbank under
the accession number NM_004963. Binding of the peptides
25 guanylin and uroguanylin or else heat-stable
enterotoxins (STa) increases the intracellular cGMP
concentration, thus inducing signal transduction
processes inside the cell.

Recent investigations indicate that expression of
30 GUCY2C also extends to extraintestinal regions such as,
for example, primary and metastatic adenotumors of the
stomach and of the esophagus (Park et al., *Cancer
Epidemiol Biomarkers Prev.* 11: 739-44, 2002). A splice
variant of GUCY2C which is found both in normal and
35 transformed tissue of the intestine comprises a 142 bp
deletion in exon 1, thus preventing translation of a
GUCY2C-like product (Pearlman et al., *Dig. Dis. Sci.*
45:298-05, 2000). The only splice variant described to

date leads to no translation product.

The aim according to the invention was to identify tumor-associated splice variants for GUCY2C which can be utilized both for diagnosis and for therapy.

RT-PCR investigations with a GUCY2C-specific primer pair (SEQ ID NO:22, 23, 98, 99) show pronounced expression of GUCY2C transcripts in normal colon and stomach, and weak expression in liver, testis, ovary, thymus, spleen, brain and lung (tab. 2, fig. 19). Expression in colon and stomach was at least 50 times higher than in all other normal tissues. Marked GUCY2C transcript levels were detected in colon tumors and stomach tumors (tab. 2). These results were specified by a quantitative PCR analysis and showed pronounced GUCY2C expression in normal colon, ileum, and in almost all colon tumor samples investigated (fig. 2, 19B). A massive overexpression was detectable in some colon tumor samples. In addition, expression is found in 7/10 stomach tumors. We also surprisingly found that the gene is activated in many other previously undescribed tumors, inter alia ovarian, breast, liver and prostate tumors (fig. 19B, tab. 2).

Table 2: GUC2C expression in normal and tumor tissues

Normal tissues	Expression	Tumor type	Expression
Brain	+	Colon	+++
Cerebellum		Pancreas	-
Myocardium		Esophagus	-
Skeletal muscle	-	Stomach	+++
Myocardium		Lung	-
Stomach	+++	Mamma	-+
Colon	+++	Ovary	+
Pancreas	-	Endometrium	
Kidney	-	ENT	
Liver	+	Kidney	

Normal tissues	Expression
Testis	++
Thymus	+
Breast	-
Ovary	+
Uterus	+
Skin	
Lung	+
Thyroid	
Lymph nodes	-
Spleen	+
PBMC	-
Prostate	-

Tumor type	Expression
Prostate	+
Liver	+

The following primer pairs were used to detect splice variants in colonic tissue and colon tumor tissue:

GUCY2C-118s/GUCY2C-498as (SEQ ID NO:24, 29); GUCY2C-
 5 621s/GUCY2C-1140as (SEQ ID NO:25, 30);
 GUCY2C-1450s/GUCY2C-1790as (SEQ ID NO:26, 31);
 GUCY2C-1993s/GUCY2C-2366as (SEQ ID NO:27, 32);
 GUCY2C-2717s/GUCY2C-3200as (SEQ ID NO:28, 33);
 GUCY2C-118s/GUCY2C-1140as (SEQ ID NO:24, 30);
 10 GUCY2C-621s/GUCY2C-1790as (SEQ ID NO:25, 31);
 GUCY2C-1450s/GUCY2C-2366as (SEQ ID NO:26, 32);
 GUCY2C-1993s/GUCY2C-3200as (SEQ ID NO:27, 33).

On investigation of splice variants in colon tumor
 15 tissue, three previously unknown forms were identified
 according to the invention.

a) A deletion of exon 3 (SEQ ID NO:3) which leads
 to a variant of GUCY2C which is only 111 amino
 acids long and in which the asparagine at
 20 position 111 is replaced by a proline.

b) A deletion of exon 6 (SEQ ID NO:4) which
 results in an expression product 258 amino
 acids long. This would generate a C-terminal
 neoepitope comprising 13 amino acids.

c) A variant in which the nucleotides at positions 1606-1614, and the corresponding amino acids L(536), L(537) and Q(538), are deleted (SEQ ID NO:5).

5 The splice variants according to the invention with deletions respectively in exon 3 and exon 6 (SEQ ID NO:3, 4) are distinguished in particular by the translation products (SEQ ID NO:12, 13) having no transmembrane domain. The result in the case of exon 6
10 deletion is a C-terminal neoepitope of 13 amino acids which shows no homology whatsoever with previously known proteins. This neoepitope is thus predestined to be a target structure for immunotherapy. The splice variant of the invention with base deletions at
15 positions 1606-1614 (SEQ ID NO:5) and its translation product (SEQ ID NO:14) likewise comprises a neoepitope. Antibodies for detecting GUCY2C protein were produced by immunizing rabbits. The following peptides were used to propagate these antibodies:

20 SEQ ID NO:100: HNGSYEISVLMMGNS (AA 31-45)

SEQ ID NO:101: NLPTPPTVENQQRLA (AA 1009-1023)

Such antibodies can in principle be used for diagnostic and therapeutic purposes.

25 In particular, the extracellular domain of GUCY2C (position of the predicted extracellular domain from the sequence of SEQ ID NO:11: AA 454-1073 (SEQ ID NO:102)) can be used according to the invention as target structure of monoclonal antibodies. However,
30 the structural prediction is somewhat ambiguous and not yet verified experimentally, so that an alternative membrane orientation is also conceivable. In this case, amino acids 1-431 would be outside the cell and be suitable as target for monoclonal antibodies. These
35 antibodies bind specifically to the cell surface of tumor cells and can be used both for diagnostic and for therapeutic methods. Overexpression of GUCY2C, especially in the colon tumors, provides additional

support for such a use. Sequences coding for proteins can moreover be used according to the invention as vaccine (RNA, DNA, peptides, protein) for inducing tumor-specific immune responses (T-cell- and B-cell-mediated immune responses).

It is moreover possible in accordance with the cellular function of the GUCY2C molecule to develop according to the invention substances, especially small molecules, which modulate the function of the enzyme on tumor cells. The product of the enzymic reaction, cGMP, is a known cellular signal molecule with a wide variety of functions (Tremblay et al. Mol Cell Biochem 230, 31, 2002).

Example 3: Identification of SCGB3A2 as diagnostic and therapeutic cancer target

SCGB3A2 (SEQ ID NO:6) (translation product: SEQ ID NO:15) belongs to the secretoglobin gene family. The sequence is published in GenBank under accession number NM_054023. SCGB3A2 (UGRP1) is a homodimeric secretory protein with a size of 17 kDa, which is expressed exclusively in the lung and in the spiracles (Niimi et al., Am J Hum Genet 70:718-25, 2002). RT PCR investigations with a primer pair (SEQ ID NO:37, 38) confirmed selective expression in normal lung tissue. Lung- and trachea-specific genes, e.g. for surfactant proteins, are highly downregulated in malignant tumors during dedifferentiation and are normally undetectable in lung tumors. It was surprisingly found that SCGB3A2 is active in primary and metastatic lung tumors. The investigations according to the invention showed that SCGB3A2 is strongly and frequently expressed in lung tumors (fig. 4). All the other 23 normal tissues tested, apart from lung and trachea, show no expression (cf. fig. 20).

This was additionally confirmed in a specific quantitative RT-PCR (SEQ ID NO:103, 104) (fig. 20) which additionally shows overexpression by at least one

log in more than 50% of lung tumors.

The selective and high expression of SCGB3A2 in normal lung tissue and in lung tumor biopsies can be used according to the invention for molecular diagnostic methods such as RT-PCR for detecting disseminating tumor cells in blood and bone marrow, sputum, bronchial aspirate or lavage and for detecting metastases in other tissues, e.g. in local lymph nodes. In the healthy lung, SCGB3A2 is secreted by specialized cells exclusively into the bronchi. Accordingly, it is not to be expected that SCGB3A2 protein will be detectable in body fluids outside the respiratory tract in healthy individuals. By contrast, in particular metastatic tumor cells secrete their protein products directly into the bloodstream. One aspect of the invention therefore relates to detection of SCGB3A2 products in serum or plasma of patients via a specific antibody assay as diagnostic finding for lung tumors.

Antibodies for detecting SCGB3A2 protein were produced by immunizing rabbits. The following peptides were used to propagate these antibodies:

SEQ ID NO:105: LINKVPLPVDKLAPL

SEQ ID NO:106: SEAVKKLLEALSHLV

An SCGB3A2-specific reaction was detectable in immunofluorescence (fig. 21). As expected for a secreted protein, the distribution of SCGB3A2 in the cell was assignable to the endoplasmic reticulum and secretion granules (fig. 21A). To check the specificity, the cells were transfected in parallel with a plasmid that synthesizes an SCGB3A2-GFP fusion protein. Protein detection took place in this case via the autofluorescent GFP (green fluorescent protein) (fig. 21B). Superimposition of the two fluorescence diagrams shows unambiguously that the immune serum specifically recognizes SCGB3A2 protein (fig. 21C).

Such antibodies can be used according to the invention for example in the form of immunoassays for diagnostic

and therapeutic purposes.

Example 4: Identification of claudin-18A1 and claudin-18A2 splice variants as diagnostic and therapeutic cancer targets

The claudin-18 gene codes for a surface membrane molecule having 4 hydrophobic regions. According to prediction programs (TMHMM, TMPred) and in accordance with the topology described for many other members of this family, claudin-18 has four transmembrane domains and two extracellular domains EX1 and EX2, whose extracellular localisation (conformation 1) is shown in Figure 22. The domain D3 which is located between the two extracellular epitopes for claudin-18 and other members of this family is described in the literature as being located intracellularly and this is also predicted by commonly used prediction programs. The N and C termini are intracellular. Niimi and colleagues (*Mol. Cell. Biol.* 21:7380-90, 2001) described two splice variants of the murine and human claudin-18 which have been described as expressed selectively in lung tissue (claudin-18A1) and in stomach tissue (claudin-18A2), respectively. These variants differ in the N terminus.

It was investigated according to the invention how far the splice variants claudin-18A2 (SEQ ID NO:7) and claudin-18A1 (SEQ ID NO:117), and their respective translation products (SEQ ID NO:16 and 118), can be used as markers or therapeutic target structures for tumors. A quantitative PCR able to distinguish between the two variants was established by selecting A1-specific (SEQ ID NO:109 & 110) and A2-specific (SEQ ID NO:107 & 108) primer pairs. The A2 splice variant was additionally tested with a second primer pair in a conventional PCR (SEQ ID NO:39 & 40). The A1 variant is described to be active only in healthy lung tissue. However, it was surprisingly found according to the invention that the A1 variant is also active in the

gastric mucosa (fig. 23). Stomach and lung are the only normal tissues showing significant activation. All other normal tissues are negative for claudin-A1. On investigating tumors, it was surprisingly found that

5 claudin-A1 is highly activated in a large number of tumor tissues. Particularly strong expression is to be found in stomach tumors, lung tumors, pancreatic tumors, esophageal tumors (fig. 23), ENT tumors and prostate tumors. The claudin-A1 expression levels in

10 ENT, prostate, pancreatic and esophageal tumors are 100-10 000 higher than the levels in the corresponding normal tissues. The oligonucleotides used to investigate the claudin-A2 splice variant specifically enable this transcript to be amplified (SEQ ID NO:39 &

15 40 and 107 & 108). Investigation revealed that the A2 splice variant is expressed in none of the more than 20 normal tissues investigated apart from gastric mucosa and to a small extent also testis tissue (fig. 24). We have found that the A2 variant is also, like the A1

20 variant, activated in many tumors (fig. 24). These include stomach tumors, pancreatic tumors, esophageal tumors and liver tumors. Although no activation of claudin-18A2 is detectable in healthy lung, it was surprisingly found that some lung tumors express the A2

25 splice variant.

Table 3A. Expression of claudin-18A2 in normal and tumor tissues

Normal tissue	Expression
Brain	-
Cerebellum	-
Myocardium	-
Skeletal muscle	-
Endometrium	-
Stomach	+++
Colon	-
Pancreas	-
Kidney	-
Liver	-
Testis	+
Thymus	-
Breast	-
Ovary	-
Uterus	-
Skin	-
Lung	-
Thyroid	-
Lymph nodes	-
Spleen	-
PBMC	-
Esophagus	-

5

Tumor type	Expression
Colon	-
Pancreas	++
Esophagus	++
Stomach	+++
Lung	++
Breast	-
Ovary	-
Endometrium	n.i.
ENT	++
Kidney	-
Prostate	-

Table 3B. Expression of claudin-18A1 in normal and tumor tissues

5

Normal tissue	Expression
Brain	-
Cerebellum	-
Myocardium	-
Skeletal muscle	-
Endometrium	-
Stomach	+++
Colon	-
Pancreas	-
Kidney	-
Liver	-
Testis	+
Thymus	-
Breast	-
Ovary	-
Uterus	-
Skin	-
Lung	+++
Thyroid	-
Lymph nodes	-
Spleen	-
PBMC	-
Esophagus	-

Tumor type	Expression
Colon	-
Pancreas	++
Esophagus	++
Stomach	+++
Lung	++
Breast	+
Ovary	n.i.
Endometrium	n.i.
ENT	++
Kidney	-
Prostate	++

Conventional PCR as independent control investigation also confirmed the results of the quantitative PCR. The oligonucleotides (SEQ ID NO:39, 40) used for this permit specific amplification of the A2 splice variant.

5 It was shown according to the invention that most gastric tumors and half of the tested pancreatic tumors showed strong expression of this splice variant (fig. 5). By contrast, expression is not detectable in other tissues by conventional PCR. In particular, there
10 is no expression in important normal tissues such as lung, liver, blood, lymph nodes, breast and kidney (tab. 3).

The splice variants thus represent according to the
15 invention highly specific molecular markers for tumors of the upper gastrointestinal tract as well as lung tumors, ENT tumors, prostate tumors and metastases thereof. These molecular markers can be used according to the invention for detecting tumor cells. Detection
20 of the tumors is possible according to the invention with the oligonucleotides described (SEQ ID NO:39, 40, 107-110). Particularly suitable oligonucleotides are primer pairs of which at least one binds under stringent conditions to a segment of the transcript
25 which is 180 base pairs long and is specific for one (SEQ ID NO:8) or the other splice variant (SEQ ID NO:119).

These genetic products are attractive therapeutic
30 target structures since due to the fact that they are missing in most toxicity relevant organs no side effects on these organs are to be expected, while due to the strong activation in cells of the cancer types mentioned strong binding to these cells and mediation
35 of corresponding cell damaging effects can be expected.

In order to confirm these data at the protein level, claudin-specific antibodies and immune sera were

generated by immunizing animals. The N-terminal extracellular domain EX1 differs in sequence in the two splice variants A1 and A2 (SEQ ID NO:111 for A1 and SEQ ID NO:112 for A2). The C-terminal extracellular domain EX2 is identical for both variants (SEQ ID NO:137). To date, no antibodies which bind to the extracellular domains of claudin-18 have yet been described. Also no antibodies which are able to discriminate specifically between A1 and A2 variants have yet been described. According to the invention, peptide epitopes and protein fragments which are located extracellularly and are specific for variant A1 or A2 or occur in both variants were selected for the immunization in order to produce antibodies. The following peptides, inter alia, were selected for the immunization in order to produce antibodies:

- SEQ ID NO:17: DQWSTQDLYN (N-terminal extracellular domain, A2-specific, binding independent of glycosylation)
- SEQ ID NO:18: NNPVTAVFNYQ (N-terminal extracellular domain, A2-specific, binding mainly to unglycosylated form, N37)
- SEQ ID NO:113: STQDLYNNPVTAVF (N-terminal extracellular domain, A2-specific, binding only to non-glycosylated form, N37)
- SEQ ID NO:114: DMWSTQDLYDNP (N-terminal extracellular domain, A1-specific)
- SEQ ID NO:115: CRPYFTILGLPA (N-terminal extracellular domain, mainly specific for A1)
- SEQ ID NO:116: TNFWMSTANMYTG (C-terminal extracellular domain, recognizes both A1 and A2).

Inter alia, antibodies could be produced which selectively recognize the N terminal domain of the splice variant claudin-18A1 but not the A2 variant (fig. 28). Using epitopes for immunizations located in the C terminal extracellular domain which is identical in both splice variants, antibodies could be produced

which recognize both variants (fig. 27).

The data for a A2-specific antibody produced by immunization with SEQ ID NO:17 are shown by way of example. The specific antibody can be utilized under various fixation conditions for immunofluorescence investigations. With comparative stainings of RT-PCR-positive and negative cell lines, in an amount which is readily detectable, the corresponding protein can be specifically detected inter alia in the gastric tumor, esophageal tumor and pancreatic tumor cell lines typed as positive (fig. 25). The endogenous protein is membrane-located and forms relatively large focal aggregates on the membrane (fig. 25). This antibody was used for immunohistochemical stainings of human tissues. The selective tissue distribution of this protein was confirmed. A large series of different normal tissues was investigated in most of which claudin-18A2 protein was not detectable as shown by way of example for liver, lung, kidney and colon. Activation of this protein was only found in normal stomach tissue (fig. 32). Surprisingly, the A2 variant of claudin-18 was detectable in the differentiated cells of stomach mucosa but not in stem cells. Differentiated stomach mucosa cells are subject to permanent regeneration. Physiologically, the total stomach epithelium is continuously replaced from the stem cells of the stomach. This supports the usefulness of the A2 variant as therapeutic target structure since it was shown according to the invention that stem cells of the stomach as the indispensable cell population of stomach mucosa do not harbour the A2 variant as all other healthy organs and, thus, are not attacked by a substance which is specifically directed against the A2 variant. Using this antibody, the A2 variant of claudin-18 was detected in a series of human tumors (fig. 33), in particular in tumors of stomach, esophagus and lung, which attracted already attention

in RT-PCR investigations. According to the invention, these tumors are therapeutically accessible. The antibody described above was additionally employed for protein detection in Western blotting. As expected, protein is detected only in stomach and in no other normal tissue, not even lung where only the A1 variant is activated (fig. 29). The comparative staining of stomach tumors and adjacent normal stomach tissue from patients surprisingly revealed that claudin-18 A2 has a smaller mass weight in all stomach tumors in which this protein is detected (fig. 30, left). It was found according to the invention in a series of experiments that a band also appears at this position when lysate of normal stomach tissue is treated with the deglycosylating agent PNGase F (fig. 30, right). Whereas exclusively the glycosylated form of the A2 variant is detectable in all normal stomach tissues, A2 is detectable as such in more than 60% of the investigated gastric tumors, in particular exclusively in the deglycosylated form. Although the A2 variant of claudin-18 is not detected in normal lung even at the protein level, it is to be found in bronchial tumors, as also previously in the quantitative RT-PCR. Once again, only the deglycosylated variant is present (fig. 31). Claudin-18 is a highly selective differentiation antigen of stomach (variant A2) or lung and stomach (variant A1). Our data indicate that it is obviously subject to tumor-associated alterations of the glycosylation machinery and that in tumors a specific form of the variant A2 is produced which is deglycosylated. The results of the PNGaseF-treatment show that claudin-18A2 differs in its N glycosylation in tumor and normal tissue.

The glycosylation of an epitope can prevent binding of an antibody specific for this epitope and can in the present case contribute to the inability of such an antibody to bind to claudin-18A2 in normal tissues but

to the exclusive binding to the non-glycosylated form in cancer cells. To produce antibodies according to the invention which selectively bind to non-glycosylated epitopes, this was considered in selecting the immunogens. According to the invention, different regions of claudin-18A2 were identified which can be present in tumor and normal tissue in a differentially glycosylated form. Among others, the regions comprising the amino acids 37, 38, 45, 116, 141, 146, 205 of claudin-18A2 were identified as potential glycosylation sites for claudin-18A2 (fig. 22, below). According to the invention, tumor cells and normal tissues differ in glycosylation at one or more of these positions. Most of these regions do not represent a classical glycosylation site but contain asparagin, serine and threonine which infrequently can also be glycosylated (prediction of fig. 22, below). Both variants of claudin-18 have a unique classical glycosylation motive in the D3 domain which according to the literature and commonly used prediction algorithms is supposed to be intracellularly located.

However, for PMP 22 which is a tetraspanine which is structurally similar to claudin-18, it was shown that the hydrophobic membrane domains 2 and 3 do not span entirely through the cell membrane but intercalate only partially in the plasma membrane (Taylor et al., J. Neurosc. Res. 62:15-27, 2000). For this reason, the entire region between the two outer transmembrane domains of PMP22 is located extracellularly. The possibility for such a topology was hypothesized and verified for claudin-18A2. To this end, three constructs were prepared which each carried a marker sequence (His or HA tag) in one of the EX1, EX2 or D3 domains (fig. 42, top). These were transfected into cell lines and it was tested whether an antibody directed against these marker sequences binds to non-permeabilized cells which requires that the

corresponding region of the protein is located topologically in an extracellular manner. Since all three regions of the molecule were determined to be extracellular by flow-through cytometry (fig. 42, 5 below), it was confirmed that claudin-18A2 can be present in a conformation having two transmembrane domains and one large extracellularly located domain (fig. 22, conformation 2). This conformation is biochemically and therapeutically relevant since it 10 contains additional binding sites for therapeutic antibodies (SEQ ID NO: 142, 143).

According to the invention, antibodies are preferably produced which discriminate between glycosylated and 15 non-glycosylated variants of claudin-18A2. These have a particularly high specificity for tumor cells. In preparing antibodies which are specific for the glycosylation also these different conformations besides the glycosylation domains were considered.

20 Preferably, protein fragments from the D3 region of claudin-18A2 are suitable for immunizing animals in a non-limiting manner. This is shown for two antibodies mAB1 and mAB2 by way of example (fig. 44). The binding 25 properties of these antibodies to cell lines which express the A1 or A2 variant of claudin-18 were investigated. It was shown that claudin-18A2 is accessible for antibodies on the cell surface. According to the invention, such antibodies are 30 specific for the A2 variant and do not bind to the A1 variant (fig. 44). Short foreign sequences (myc tag) were each introduced into the region of the extracellular domains Ex1 and Ex2. For example, it is shown for mAB1 that the binding properties of the 35 antibody are not affected thereby and that the actual epitope is located in the D3 domain.

The antibodies generated can be utilized diagnostically

as well as therapeutically. Immune sera such as the one described herein (directed against peptide SEQ ID NO: 17) can be utilized diagnostically, for example, for Western blotting. According to the invention, antibodies which do not bind to the glycosylated epitope can be produced by immunizing with peptides which contain at least one of these regions (for example, peptide SEQ ID NO: 113 (fig. 26), peptide SEQ ID NO: 142-145). According to the invention, such antibodies specifically bind to the deglycosylated epitopes on tumor cells. The glycosylation which is missing compared to normal tissues at one of the positions mentioned might also be due to a secondary endogenous deglycosylation in tumor cells. Such a deglycosylation is associated with a Asn (N) → Asp (D) transformation of the respective amino acid. For the production of antibodies against tumor-associated variants which are modified in such a manner, peptides derived from claudin-18A2 can thus be used according to the invention in which the amino acid Asn (N) at at least one of the positions 37, 38, 45, 116, 141, 146, 205 of the claudin-18A2 peptide is substituted by Asp (D) (e.g. SEQ ID NO: 146-150). It is possible in particular to employ such antibodies therapeutically because they are highly selective for tumor cells. The produced antibodies can be used directly also for producing chimeric or humanized recombinant antibodies. This can also take place directly with antibodies obtained from rabbits (concerning this, see J Biol Chem. 2000 May 5;275(18):13668-76 by Rader C, Ritter G, Nathan S, Elia M, Gout I, Jungbluth AA, Cohen LS, Welt S, Old LJ, Barbas CF 3rd. "The rabbit antibody repertoire as a novel source for the generation of therapeutic human antibodies"). For this purpose, lymphocytes from the immunized animals were preserved. The amino acids 1-47 (SEQ ID NO:19 and 120) also represent particularly good epitopes for immunotherapeutic methods such as vaccines and the

adoptive transfer of antigen-specific T lymphocytes.

Example 5: Identification of SLC13A1 as diagnostic and therapeutic cancer target

5

SLC13A1 belongs to the family of sodium sulfate cotransporters. The human gene is, in contrast to the mouse homolog of this gene, selectively expressed in the kidney (Lee et al., *Genomics* 70:354-63, 2000).

10 SLC13A1 codes for a protein of 595 amino acids and comprises 13 putative transmembrane domains. Alternative splicing results in 4 different transcripts (SEQ ID NO:41-44) and its corresponding translation products (SEQ ID NO:45-48). It was investigated whether
15 SLC13A1 can be used as marker for kidney tumors. Oligonucleotides (SEQ ID NO:49, 50) which enable specific amplification of SLC13A1 were used for this purpose.

Table 4. Expression of SLC13A1 in normal and tumor tissues

5

Normal tissue	Expression
Brain	-
Cerebellum	nd
Myocardium	nd
Skeletal muscle	nd
Myocardium	-
Stomach	-
Colon	-
Pancreas	nd
Kidney	+++
Liver	-
Testis	+
Thymus	-
Breast	-
Ovary	-
Uterus	nd
Skin	nd
Lung	-
Thyroid	-
Lymph nodes	-
Spleen	-
PBMC	-
Sigmoid	-
Esophagus	-

Tumor type	Expression
Colon	nd
Pancreas	nd
Esophagus	nd
Stomach	nd
Lung	nd
Breast	nd
Ovary	nd
Endometrium	nd
ENT	nd
Kidney	+++
Prostate	nd

RT-PCR investigations with an SLC13A1-specific primer pair (SEQ ID NO:49, 50) confirmed virtually selective expression in the kidney, and showed according to the invention a high expression in virtually all (7/8) investigated renal tumor biopsies (tab. 4, fig. 6). Quantitative RT-PCR with specific primers (SEQ ID NO:121, 122) also confirmed these data (fig. 34). Weak signals were detectable in the following normal tissues: colon, stomach, testis, breast, liver and brain. Expression in renal tumors was, however, at least 100 times higher than in all other normal tissues.

In order to analyse the subcellular localization of SLC13A1 in the cell, the protein was fused to eGFP as reporter molecule and, after transfection of the appropriate plasmid, expressed heterologously in 293 cells. The localization was then analysed under the fluorescence microscope. Our data impressively confirmed that SLC13A1 is an integral transmembrane molecule (fig. 35).

Antibodies for detecting the SLC13A1 protein were produced by immunizing rabbits. The peptides of SEQ ID NO:123 and 124 were used for propagating these antibodies. Such antibodies can in principle be used for diagnostic and therapeutic purposes.

The SLC13A1 protein has 13 transmembrane domains and 7 extracellular regions. These extracellular domains of SLC13A1 in particular can be used according to the invention as target structures for monoclonal antibodies. SLC13A1 is involved as channel protein in the transport of ions. The extracellular domains of SLC13A1 in the healthy kidney are directed polarically in the direction of the urinary tract (luminally). However, high molecular weight monoclonal antibodies employed therapeutically are not excreted into the urinary tract, so that no binding to SLC13A1 takes place in the healthy kidney. By contrast, the polarity of SLC13A1 is abolished in tumor cells, and the protein

is available for antibody targeting directly via the bloodstream. The pronounced expression and high incidence of SLC13A1 in renal tumors make this protein according to the invention a highly interesting
5 diagnostic and therapeutic marker. This includes according to the invention the detection of disseminated tumor cells in serum, bone marrow, urine, and detection of metastases in other organs by means of RT-PCR. It is additionally possible to use the
10 extracellular domains of SLC13A1 according to the invention as target structure for immunodiagnosis and therapy by means of monoclonal antibodies. SLC13A1 can moreover be employed according to the invention as vaccine (RNA, DNA, protein, peptides) for inducing
15 tumor-specific immune responses (T and B cell-mediated immune responses). This includes according to the invention also the development of so-called small compounds which modulate the biological activity of SLC13A1 and can be employed for the therapy of renal
20 tumors.

Example 6: Identification of CLCA1 as diagnostic and therapeutic cancer target

25 CLCA1 (SEQ ID NO:51; translation product: SEQ ID NO:60) belongs to the family of Ca^{++} -activated Cl^- channels. The sequence is published in Genbank under the accession No. NM_001285. CLCA1 is exclusively expressed in the intestinal crypt epithelium and in the goblet
30 cells (Gruber et al., *Genomics* 54:200-14, 1998). It was investigated whether CLCA1 can be used as marker for colonic and gastric tumors. Oligonucleotides (SEQ ID NO:67, 68) which enable specific amplification of CLCA1 were used for this purpose. RT-PCR
35 investigations with this primer set confirmed selective expression in the colon, and showed according to the invention high expression in 3/7 investigated colonic and 1/3 investigated gastric tumor samples (fig. 7).

The other normal tissues showed no or only very weak expression. This was additionally confirmed with a specific quantitative RT-PCR (SEQ ID NO:125, 126), in which case no expression could be detected in the
5 normal tissues analyzed (fig. 36). Of the tumor samples investigated in this experiment, 6/12 colonic tumor samples and 5/10 gastric tumor samples were positive for CLCA1. Overall, expression of the gene in tumors appears to be dysregulated. Besides samples with very
10 strong expression, CLCA1 was markedly downregulated in other samples.

The protein is predicted to have 4 transmembrane domains with a total of 2 extracellular regions. These extracellular domains of CLCA1 in particular can be
15 used according to the invention as target structures for monoclonal antibodies.

The pronounced expression and high incidence of CLCA1 in gastric and colonic tumors make this protein according to the invention an interesting diagnostic
20 and therapeutic marker. This includes according to the invention the detection of disseminated tumor cells in serum, bone marrow, urine, and detection of metastases in other organs by means of RT-PCR. It is additionally possible to use the extracellular domains of CLCA1
25 according to the invention as target structure for immunodiagnosis and therapy by means of monoclonal antibodies. CLCA1 can moreover be employed according to the invention as vaccine (RNA, DNA, protein, peptides) for inducing tumor-specific immune responses (T and B
30 cell-mediated immune responses). This includes according to the invention also the development of so-called small compounds which modulate the biological activity as transport proteins of CLCA1 and can be employed for the therapy of gastrointestinal tumors.

35

Example 7: Identification of FLJ21477 as diagnostic and therapeutic cancer target

FLJ21477 (SEQ ID NO:52) and its predicted translation
5 product (SEQ ID NO:61) was published as hypothetical
protein in Genbank under the accession No. NM_025153.
It is an integral membrane protein having ATPase
activity and 4 transmembrane domains, which is
accordingly suitable for therapy with specific
10 antibodies. RT-PCR investigations with FLJ21477-
specific primers (SEQ ID NO:69, 70) showed selective
expression in the colon, and additionally various
levels of expression in 7/12 investigated colonic tumor
samples (fig. 8). The other normal tissues showed no
15 expression. This was confirmed additionally by a
specific quantitative RT-PCR (SEQ ID NO:127, 128).
FLJ21477-specific expression was detectable both in
colon (fig. 37A) and in 11/12 of colonic tumors.
Besides the expression in colon tissue, expression was
20 additionally detectable in stomach tissue. In addition,
under the conditions of the quantitative RT-PCR, the
expression detectable in brain, thymus and esophagus
was distinctly weaker compared with colon and stomach
(fig. 37A). It was moreover additionally possible to
25 detect FLJ21477-specific expression in the following
tumor samples: stomach, pancreas, esophagus and liver.
The protein is predicted to have 4 transmembrane
domains with a total of 2 extracellular regions. These
extracellular domains of FLJ21477 in particular can be
30 used according to the invention as target structures
for monoclonal antibodies.

The expression and the high incidence of FLJ21477 for
gastric and colonic tumors make this protein according
to the invention a valuable diagnostic and therapeutic
35 marker. This includes according to the invention the
detection of disseminated tumor cells in serum, bone
marrow, urine, and the detection of metastases in other
organs by means of RT-PCR. In addition, the

extracellular domains of FLJ21477 can be used according to the invention as target structure for immunodiagnosis and therapy by means of monoclonal antibodies. In addition, FLJ21477 can be employed
5 according to the invention as vaccine (RNA, DNA, protein, peptides) for inducing tumor-specific immune responses (T and B cell-mediated immune responses).

**Example 8: Identification of FLJ20694 as diagnostic and
10 therapeutic cancer target**

FLJ20694 (SEQ ID NO:53) and its translation product (SEQ ID NO:62) were published as hypothetical protein in Genbank under accession No. NM_017928. This protein
15 is an integral transmembrane molecule (transmembrane domain AA 33-54), very probably with thioredoxin function. RT-PCR investigations with FLJ20694-specific primers (SEQ ID NO:71, 72) showed selective expression in the colon, and additionally various levels of
20 expression in 5/9 investigated colonic tumor samples (fig. 9). The other normal tissues showed no expression. This was additionally confirmed by a specific quantitative RT-PCR (SEQ ID NO:129, 130) (fig. 38). FLJ20694 expression was undetectable in any
25 other normal tissue apart from colon and stomach (not analysed in the first experiment).

The protein is predicted to have one transmembrane domain with an extracellular region. These extracellular domains of FLJ20694 in particular can be
30 used according to the invention as target structures for monoclonal antibodies.

In addition, FLJ20694 can be employed according to the invention as vaccine (RNA, DNA, protein, peptides) for inducing tumor-specific immune responses (T and B cell-mediated immune responses). This includes according to
35 the invention also the development of so-called small compounds which modulate the biological activity of FLJ20694 and can be employed for the therapy of

gastrointestinal tumors.

Example 9: Identification of von Ebner's protein (c20orf114) as diagnostic and therapeutic cancer target

5

von Ebner's protein (SEQ ID NO:54) and its translation product (SEQ ID NO:63) were published as Plunc-related protein of the upper airways and of the nasopharyngeal epithelium in Genbank under the accession No. AF364078.

10 It was investigated according to the invention whether mRNA encoding von Ebner's protein can be used as marker of lung tumors. Oligonucleotides (SEQ ID NO:73, 74) which enable specific amplification of cDNA encoding Ebner's protein were used for this purpose. RT-PCR
15 investigations with this primer set showed selective expression in the lung and in 5/10 investigated lung tumor samples (fig. 10). In the group of normal tissues there was also expression in the stomach. The other normal tissues showed no expression.

20

Example 10: Identification of Plunc as diagnostic and therapeutic cancer target

Plunc (SEQ ID NO:55) and its translation product
25 (SEQ ID NO:64) were published in Genbank under the accession No. NM_016583. Human Plunc mRNA codes for a protein of 256 amino acids and shows 72% homology with the murine Plunc protein (Bingle and Bingle, *Biochem Biophys Acta* 1493:363-7, 2000). Expression of Plunc is
30 confined to the trachea, the upper airways, nasopharyngeal epithelium and salivary gland.

It was investigated according to the invention whether Plunc can be used as marker of lung tumors. Oligonucleotides (SEQ ID NO:75, 76) which enable
35 specific amplification of Plunc were used for this purpose.

RT-PCR investigations with this primer set showed selective expression in the thymus, in the lung and in

6/10 investigated lung tumor samples (fig. 11). Other normal tissues showed no expression.

Example 11: Identification of SLC26A9 as diagnostic and therapeutic cancer target

SLC26A9 (SEQ ID NO:56) and its translation product (SEQ ID NO:65) were published in Genbank under the accession No. NM_134325. SLC26A9 belongs to the family of anion exchangers. Expression of SLC26A9 is confined to the bronchiolar and alveolar epithelium of the lung (Lohi et al., J Biol Chem 277:14246-54, 2002).

It was investigated whether SLC26A9 can be used as marker of lung tumors. Oligonucleotides (SEQ ID NO:77, 78) which enable specific amplification of SLC26A9 were used for this purpose. RT-PCR investigations with SLC26A9-specific primers (SEQ ID NO:77, 78) showed selective expression in the lung and in all (13/13) investigated lung tumor samples (fig. 12). The other normal tissues showed no expression, with the exception of the thyroid. It was possible in quantitative RT-PCR experiments with the primers of SEQ ID NO:131 and 132 firstly to confirm these results, and to obtain additional information. It was possible in pooled samples of 4-5 tumor tissues to detect high expression levels for SLC26A9-specific RNA in lung, colon, pancreas and stomach tumors. SLC26A9 is member of a family of transmembrane anion transporters. In the healthy lung, the protein is lumenally directed in the direction of the airways and thus not directly available to IgG antibodies from the blood. By contrast, the polarity of the protein is abolished in tumors. It is therefore possible according to the invention to address SLC26A9 as therapeutic target using monoclonal antibodies in the defined tumors, inter alia lung, gastric, and pancreatic tumors. The pronounced, high expression and high incidence of SLC26A9 for lung, stomach, pancreatic and esophageal

tumors make this protein according to the invention an excellent diagnostic and therapeutic marker. This includes according to the invention the detection of disseminated tumor cells in serum, bone marrow and urine, and detection of metastases in other organs by means of RT-PCR. In addition, the extracellular domains of SLC26A9 can be used according to the invention as target structure for immunodiagnosis and therapy by means of monoclonal antibodies. It is additionally possible to employ SLC26A9 according to the invention as vaccine (RNA, DNA, protein, peptides) for inducing tumor-specific immune responses (T and B cell-mediated immune responses). This includes according to the invention also the development of so-called small compounds which modulate the biological activity of SLC26A9 and can be employed for the therapy of lung tumors and gastrointestinal tumors.

Example 12: Identification of THC1005163 as diagnostic and therapeutic cancer target

THC1005163 (SEQ ID NO:57) is a gene fragment from the TIGR gene index. The gene is defined only in the 3' region, while an ORF is lacking. RT-PCR investigations took place with a THC1005163-specific primer (SEQ ID NO:79) and an oligo dT₁₈ primer which had a specific tag of 21 specific bases at the 5' end. This tag was examined using database search programs for homology with known sequences. This specific primer was initially employed in the cDNA synthesis in order to preclude genomic DNA contaminations. RT-PCR investigations with this primer set showed expression in the stomach, ovary, lung and in 5/9 lung tumor biopsies (fig. 13). Other normal tissues showed no expression.

Example 13: Identification of LOC134288 as diagnostic and therapeutic cancer target

LOC134288 (SEQ ID NO:58) and its predicted translation product (SEQ ID NO:66) were published in Genbank under accession No. XM_059703.

5 It was investigated according to the invention whether LOC134288 can be used as marker of renal tumors. Oligonucleotides (SEQ ID NO:80, 81) which enable specific amplification of LOC134288 were used for this purpose. RT-PCR investigations showed selective
10 expression in the kidney and in 5/8 investigated renal tumor biopsies (fig. 14).

Example 14: Identification of THC943866 as diagnostic and therapeutic cancer target

15

THC 943866 (SEQ ID NO:59) is a gene fragment from the TIGR gene index. It was investigated whether THC943866 can be used as marker of renal tumors. Oligonucleotides (SEQ ID NO:82, 83) which enable specific amplification
20 of THC943866 were used for this purpose.

RT-PCR investigations with THC943866-specific primers (SEQ ID NO:82, 83) showed selective expression in the kidney and in 4/8 investigated renal tumor biopsies (fig. 15).

25

Example 15: Identification of FLJ21458 and B7h.4 as diagnostic and therapeutic cancer targets

FLJ21458 (SEQ ID NO:84) and B7h.4 (SEQ ID NO: 138) and
30 their predicted translation products (SEQ ID NO:85, 139) represent splice variants of one gene and were published in Genbank under the accession No. NM_034850 and AY358523, respectively. Sequence analyses revealed that the proteins represent members of the
35 butyrophillin family. Structural analyses revealed that they represent type 1 transmembrane proteins with an extracellular immunoglobulin domain. Oligonucleotides (SEQ ID NO:86, 87 or SEQ ID NO: 140, 141) which enable

specific amplification of FLJ21458 or B7h.4 were used for investigating expression. RT-PCR investigations with FLJ21458-specific primers (SEQ ID NO:86, 87) showed selective expression in colon and in 7/10 investigated colonic tumor biopsies (fig. 16, tab. 5). Quantitative RT-PCR with specific primers (SEQ ID NO:133, 134) confirmed this selective expression profile (fig. 39). It was additionally possible in the experiment to detect FLJ21458 gastrointestinal-specifically in the colon, and in stomach, in the rectum and cecum and in testis. 7/11 colon metastasis samples were also positive in the quantitative PCR. FLJ21458-specific expression was extended to other tumors, and a protein-specific expression was detectable in stomach, pancreas and liver tumors (tab. 5). RT-PCR investigations with B7h.4 specific primers (SEQ ID NO: 140, 141) showed strong selective expression in lung tumors but not in normal lung tissue. Thus, both splice variants of this butyrophillin show tumor-associated expression and can be utilized as diagnostic and therapeutic tumor targets. Antibodies for detecting FLJ21458 and B7h.4 protein were produced by immunizing rabbits. Peptides which are contained in both proteins (FLJ21458 and B7h.4) were used as epitopes to propagate these antibodies:

SEQ ID NO:135: QWQVFGPDKPVQAL

SEQ ID NO:136: AKWKGPQGQDLSTDS

An FLJ21458- or B7h.4-specific reaction was detectable in immunofluorescence (fig. 40). To check the specificity of the antibodies, 293 cells were transfected with a plasmid that codes for an FLJ21458-GFP fusion protein. Specificity was demonstrated on the one hand by colocalization investigations using the specific antibody, and on the other hand via the auto-fluorescent GFP. Superimposition of the two fluorescent diagrams showed unambiguously that the immune serum recognises FLJ21458 protein (fig. 40, top). Due to the

identical epitopes in B7h.4, these antibodies can also be utilized for binding to and detection of the B7h.4 protein in tumors. Owing to the overexpression of the protein, the resultant cell staining was diffuse and
5 did not allow unambiguous protein localization. For this reason, a further immunofluorescence experiment was carried out with the stomach tumor-specific cell line Snul6 which expresses FLJ21458 endogenously (fig. 40, below). The cells were stained with the
10 FLJ21458-specific antiserum and with another antibody which recognizes the membrane protein E-cadherin. The FLJ21458-specific antibody stains the cell membranes at least weakly and is thus evidence that FLJ21458 is localized in the cell membrane.

15 Bioinformatic investigations showed that the protein encoded by FLJ21458 represents a cell surface molecule and has an immunoglobulin supermolecule domain. Selective expression of this surface molecule makes it
20 a good target for developing diagnostic methods for the detection of tumor cells and therapeutic methods for the elimination of tumor cells.

The pronounced expression and high incidence of FLJ21458 for gastric and colonic tumors make this
25 protein according to the invention a highly interesting diagnostic and therapeutic marker. This includes according to the invention the detection of disseminated tumor cells in serum, bone marrow and urine, and the detection of metastases in other organs
30 by means of RT-PCR. It is additionally possible to employ the extracellular domains of FLJ21458 according to the invention as target structure for immuno-diagnosis and therapy by means of monoclonal antibodies. It is additionally possible to employ
35 FLJ21458 according to the invention as vaccine (RNA, DNA, protein, peptides) for inducing tumor-specific immune responses (T and B cell-mediated immune responses). This includes according to the invention

also the development of so-called small compounds which modulate the biological activity of FLJ21458 and can be employed for the therapy of gastrointestinal tumors.

5 **Table 5 FLJ21458 or B7h.4* expression in normal and tumor tissues**

Normal tissue	Expression
Brain	-
Cerebellum	-
Myocardium	nd
Skeletal muscle	-
Myocardium	-
Stomach	++
Colon	+++
Pancreas	-
Kidney	-
Liver	-
Testis	++
Thymus	nd
Breast	nd
Ovary	-
Uterus	-
Skin	-
Lung	-
Thyroid	nd
Lymph nodes	-
Spleen	-
PBMC	-
Adrenal	nd
Esophagus	-
Small intestine	-
Prostate	-

Tumor type	Expression
Colon	7/10
Pancreas	5/6
Esophagus	nd
Stomach	8/10
Lung	6/8*
Breast	nd
Ovary	nd
Endometrium	nd
ENT	nd
Kidney	nd
Prostate	nd
Colonic metastases	7/11
Liver	5/8

SEQUENCE LISTING IN ELECTRONIC FORM

In accordance with Section 111(1) of the Patent Rules, this description contains a sequence listing in electronic form in ASCII text format (file: 84060076 Seq 01-NOV-17 v1.txt).

A copy of the sequence listing in electronic form is available from the Canadian Intellectual Property Office.

The sequences in the sequence listing in electronic form are reproduced in the following table.

SEQUENCE TABLE

<110> Ganymed Pharmaceuticals AG

<120> DIFFERENTIALLY EXPRESSED CLAUDIN-18 VARIANTS IN TUMORS AND USE THEREOF

<130> 84060076

<140> Division of CA 2,563,666

<141> 2005-05-18

<150> DE 10 2004 024 617.3

<151> 2004-05-18

<160> 150

<170> PatentIn version 3.1

<210> 1

<211> 1875

<212> DNA

<213> Homo sapiens

<400> 1

caggccagag	tcccagctgt	cctggactct	gctgtgggga	agggctgatg	caggtgtgga	60
gtcaaatgtg	ggtgcctect	gcagccgggt	gccaggaggg	gtggaggggc	caccctgggc	120
tttgtccggg	agcctggtct	tcccgtcctt	gggctgacag	gtgctgctgc	ctctgagccc	180
tccctgctaa	gagctgtgtg	ctgggtaagg	ctgggtggccc	tttgggctcc	ctgtccagga	240
tttgtgctct	ggagggtagg	gcttgctggg	ctggggactg	gaggggaacg	tggagctcct	300
tctgcctect	ttcctgcccc	atgacagcag	gcagatccca	ggagagaaga	gctcaggaga	360
tgggaagagg	atctgtccag	gggttagacc	tcaaggggtg	cttggaattc	tttacggcac	420
ccatgctttc	tttgaggagt	tttgtgtttg	tgggtgtggg	gtcggggctc	acctcctccc	480
acatccctgc	ccagaggtgg	gcagagtggg	ggcagtgcct	tgctccccct	gctcgtctct	540
tgctgacctc	cggctccctg	tgctgcccc	ggaccatgaa	tggcacctac	aacacctgtg	600
gctccagcga	cctcacctgg	ccccagcga	tcaagctggg	cttctacgcc	tacttgggcg	660
tcctgctggt	gctaggcctg	ctgctcaaca	gcctggcgct	ctgggtgttc	tgctgccgca	720

tgcagcagtg	gacggagacc	cgcattctaca	tgaccaacct	ggcgggtggcc	gacctctgcc	780
tgetgtgcac	cttgcccttc	gtgetgcact	ccctgcgaga	cacctcagac	acgccgctgt	840
gccagctctc	ccagggcacc	tacctgacca	acaggtacat	gagcatcagc	ctggtcacgg	900
ccatcgccgt	ggaccgctat	gtggccgtgc	ggcaccgct	gcgtgccgc	gggctgcgt	960
ccccaggca	ggetgcggcc	gtgtgcgcgg	tcctctgggt	gctggtcacc	ggctccctgg	1020
tggctcgctg	gctcctgggg	attcaggagg	gcggcttctg	cttcaggagc	acccggcaca	1080
atttcaactc	catggcgctc	ccgtgctgg	gattctacct	gcccctggcc	gtggtggtct	1140
tctgctccct	gaaggtggtg	actgccttg	cccagaggcc	acccaccgac	gtggggcagg	1200
cagaggccac	ccgcaaggct	gcccgcctgg	tctgggccaa	cctcctgggt	ttcgtggtct	1260
gcttcctgcc	cctgcacgtg	gggtgcacag	tgcgcctcgc	agtgggctgg	aacgcctgtg	1320
ccctcctgga	gacgatccgt	cgcgccctgt	acataaccag	caagctctca	gatgccaaact	1380
gctgcctgga	cgcctctctg	tactactaca	tggccaagga	gttcacaggag	gcgtctgcac	1440
tggccgtggc	tcccagtgct	aaggcccaca	aaagccagga	ctctctgtgc	gtgaccctcg	1500
cctaagaggc	gtgctgtggg	cgtgtgtggc	caggtctcgg	gggtccggg	aggtgctgcc	1560
tgccagggga	agctggaacc	agtagcaagg	agcccgggat	cagccctgaa	ctcactgtgt	1620
attctcttgg	agccttgggt	gggcagggac	ggcccaggta	cctgctctct	tgggaagaga	1680
gagggacagg	gacaagggca	agaggactga	ggccagagca	aggccaatgt	cagagacccc	1740
cgggatgggg	cctcacactt	gccaccccca	gaaccagctc	acctggccag	agtgggttcc	1800
tgetggccag	ggtgcagcct	tgatgacacc	tgcgcgtgcc	cctcggggct	ggaataaaac	1860
tccccaccca	gagtc					1875

<210> 2
 <211> 3222
 <212> DNA
 <213> Homo sapiens

atgaagacgt	tgetgttggg	cttggctttg	tggctactgc	tcttccagcc	cgggtggctg	60
tccttttagtt	cccagggtgag	tcagaactgc	cacaatggca	gctatgaaat	cagcgtccctg	120
atgatgggca	actcagcctt	tgcagagccc	ctgaaaaact	tgggaagatgc	ggtgaatgag	180
gggctggaaa	tagtgagagg	acgtctgcaa	aatgctggcc	taaatgtgac	tgtgaacgct	240
actttcatgt	attcggatgg	tctgattcat	aactcaggcg	actgccggag	tagcacctgt	300
gaaggcctcg	acctactcag	gaaaatttca	aatgcacaa	ggatgggctg	tgtcctcata	360
gggccctcat	gtacatactc	caccttccag	atgtaccttg	acacagaatt	gagctacccc	420
atgatctcag	ctggaagtgt	tggattgtca	tgtgactata	aagaaacctt	aaccaggctg	480
atgtctccag	ctagaaagtt	gatgtacttc	ttggttaact	tttggaaaac	caacgatctg	540
cccttcaaaa	cttattcctg	gagcacttcg	tatgtttaca	agaatgggtac	agaaactgag	600
gactgtttct	ggtaccttaa	tgtctctggg	gctagcgttt	cctatttctc	ccacgaactc	660
ggctttaagg	tgggtgttaag	acaagataag	gagtttcagg	atatcttaat	ggaccacaa	720
aggaaaagca	atgtgattat	tatgtgtggt	ggtccagagt	tcctctacaa	gctgaagggt	780
gaccgagcag	tggctgaaga	cattgtcatt	attctagtgg	atcttttcaa	tgaccagtac	840
ttggaggaca	atgtcacagc	ccctgactat	atgaaaaatg	tccttgttct	gacgctgtct	900
cctgggaatt	cccttctaaa	tagctctttc	tccagggaatc	tatcaccaac	aaaacgagac	960
tttgctcttg	cctatttgaa	tggaaacctg	ctctttggac	atatgctgaa	gatatttctt	1020
gaaaatggag	aaaatattac	cacccccaaa	tttgctcatg	ctttcaggaa	tctcactttt	1080
gaagggtatg	acgggtccagt	gaccttggat	gactgggggg	atggtgacag	taccatgggtg	1140
cttctgtata	cctctgtgga	caccaagaaa	tacaagggtc	ttttgacctc	tgatacccac	1200
gtaaataaga	cctatcctgt	ggatatgagc	cccacattca	cttggaagaa	ctctaaactt	1260
cctaatagata	ttacaggccg	gggccctcag	atcctgatga	ttgcagtctt	cacctcact	1320
ggagctgtgg	tgetgtcct	gctcgtcgt	ctcctgatgc	tcagaaaata	tagaaaagat	1380
tatgaacttc	gtcagaaaaa	atgggtccac	attcctcctg	aaaatatctt	tcctctggag	1440
accaatgaga	ccaatcatgt	tagcctcaag	atcgatgatg	acaaaagacg	agatacaatc	1500
cagagactac	gacagtgcaa	atacgacaaa	aagcgagtga	ttctcaaaga	tctcaagcac	1560
aatgatggta	atttcactga	aaaacagaag	atagaattga	acaagtgtgt	tcagattgac	1620
tattacaacc	tgaccaagtt	ctacggcaca	gtgaaacttg	ataccatgat	cttcgggggtg	1680

atagaataact	gtgagagagg	atccctccgg	gaagttttta	atgacacaat	ttcctaccct	1740
gatggcacat	tcattggattg	ggagttttaag	atctctgtct	tgtatgacat	tgctaaggga	1800
atgtcatatc	tgcactccag	taagacagaa	gtccatggtc	gtctgaaatc	taccaactgc	1860
gtagtggaca	gtagaatggg	ggggaagatc	actgattttg	gctgcaattc	catttttacct	1920
ccaaaaaagg	acctgtggac	agctccagag	cacctccgcc	aagccaacat	ctctcagaaa	1980
ggagatgtgt	acagctatgg	gatcatcgca	caggagatca	ttctgcggaa	agaaaccttc	2040
tacactttga	gctgtcggga	ccggaatgag	aagattttca	gagtggaaaa	ttccaatgga	2100
atgaaacctt	tccgcccaga	tttattcttg	gaaacagcag	aggaaaaaga	gctagaagtg	2160
tacctacttg	taaaaaactg	ttgggaggaa	gatccagaaa	agagaccaga	tttcaaaaaa	2220
attgagacta	cacttgccaa	gatatttgga	ctttttcatg	accaaaaaaa	tgaaagctat	2280
atggatacct	tgatccgacg	tctacagcta	tattctcgaa	acctggaaca	tctggtagag	2340
gaaaggacac	agctgtacaa	ggcagagagg	gacagggctg	acagacttaa	ctttatgttg	2400
cttccaaggc	tagtggtaaa	gtctctgaag	gagaaaggct	ttgtggagcc	ggaactatat	2460
gaggaagtta	caatctactt	cagtgcattt	gtaggtttca	ctactatctg	caaatacagc	2520
acccccatgg	aagtgggtga	catgcttaat	gacatctata	agagttttga	ccacattgtt	2580
gatcatcatg	atgtctacaa	gggtggaaacc	atcgggtgatg	cgtacatggg	ggctagtggg	2640
ttgcctaaga	gaaatggcaa	tccgcatgca	atagacattg	ccaagatggc	cttggaaatc	2700
ctcagcttca	tggggacctt	tgagctggag	catcttcctg	gcctcccaat	atggattcgc	2760
attggagttc	actctggtec	ctgtgctgct	ggagttgtgg	gaatcaagat	gcctcgttat	2820
tgtctatttg	gagatacggg	caacacagcc	tctaggatgg	aatccactgg	cctccctttg	2880
agaattcacg	tgagtggctc	caccatagcc	atcctgaaga	gaactgagtg	ccagttcctt	2940
tatgaagtga	gaggagaaac	atacttaaag	ggaagaggaa	atgagactac	ctactggctg	3000
actgggatga	aggaccagaa	attcaacctg	ccaacccttc	ctactgtgga	gaatcaacag	3060
cgtttgcaag	cagaattttc	agacatgatt	gccaactctt	tacagaaaag	acaggcagca	3120
gggataagaa	gccaaaaacc	cagacgggta	gccagctata	aaaaaggcac	tctggaatac	3180
ttgcagctga	ataccacaga	caaggagagc	acctattttt	aa		3222

<210> 3
 <211> 336
 <212> DNA
 <213> Homo sapiens

<400> 3						
atgaagacgt	tgetgttgga	cttggtcttg	tgggtcactgc	tcttccagcc	cgggtggctg	60
tccttttagtt	cccaggtgag	tcagaactgc	cacaatggca	gctatgaaat	cagcgtcctg	120
atgatgggca	actcagcctt	tgcagagccc	ctgaaaaact	tggaagatgc	gggtgaatgag	180
gggctggaaa	tagtgagagg	acgtctgcaa	aatgctggcc	taaatgtgac	tgtgaacgct	240
actttcatgt	attcggatgg	tctgattcat	aactcaggcg	actgccggag	tagcacctgt	300
gaaggcctcg	acctactcag	gaaaatttca	ccttga			336

<210> 4
 <211> 777
 <212> DNA
 <213> Homo sapiens

<400> 4						
atgaagacgt	tgetgttgga	cttggtcttg	tgggtcactgc	tcttccagcc	cgggtggctg	60
tccttttagtt	cccaggtgag	tcagaactgc	cacaatggca	gctatgaaat	cagcgtcctg	120
atgatgggca	actcagcctt	tgcagagccc	ctgaaaaact	tggaagatgc	gggtgaatgag	180
gggctggaaa	tagtgagagg	acgtctgcaa	aatgctggcc	taaatgtgac	tgtgaacgct	240
actttcatgt	attcggatgg	tctgattcat	aactcaggcg	actgccggag	tagcacctgt	300
gaaggcctcg	acctactcag	gaaaatttca	aatgcacaac	ggatgggctg	tgtcctcata	360
gggccctcat	gtacatactc	caccttccag	atgtaccttg	acacagaatt	gagctacccc	420
atgatctcag	ctggaagttt	tggattgtca	tgtgactata	aagaaacctt	aaccaggctg	480

atgtctccag	ctagaaagtt	gatgtacttc	ttgggttaact	tttggaaaac	caacgatctg	540
cccttcaaaa	cttattcctg	gagcacttcg	tatgttttaca	agaatgggtac	agaaactgag	600
gactgtttct	ggtaccttaa	tgtctctggag	gctagecgttt	cctattttctc	ccacgaactc	660
ggctttaagg	tgggtgttaag	acaagataag	gagtttcagg	atatcttaat	ggaccacaac	720
aggaaaagca	atgtgaccag	tacttggagg	acaatgtcac	agcccttgac	tatatga	777

<210> 5
 <211> 3213
 <212> DNA
 <213> Homo sapiens

<400> 5						
atgaagacgt	tgtctgttga	cttggcctttg	tgggtcaactgc	tcttccagcc	cgggtggctg	60
tccttttagtt	cccaggtgag	tcagaactgc	cacaatggca	gctatgaaat	cagcgtcctg	120
atgatgggca	actcagcctt	tgcagagccc	ctgaaaaact	tgggaagatgc	ggtgaatgag	180
gggctggaaa	tagtgagagg	acgtctgcaa	aatgctggcc	taaatgtgac	tgtgaacgct	240
actttcatgt	attcggatgg	tctgattcat	aactcaggcg	actgccggag	tagcacctgt	300
gaaggcctcg	acctaactcag	gaaaatttca	aatgcacaac	ggatgggctg	tgtcctcata	360
gggccctcat	gtacatactc	caccttccag	atgtaccttg	acacagaatt	gagctacccc	420
atgatctcag	ctggaagttt	tggattgtca	tgtgactata	aagaaacctt	aaccaggctg	480
atgtctccag	ctagaaagtt	gatgtacttc	ttgggttaact	tttggaaaac	caacgatctg	540
cccttcaaaa	cttattcctg	gagcacttcg	tatgttttaca	agaatgggtac	agaaactgag	600
gactgtttct	ggtaccttaa	tgtctctggag	gctagecgttt	cctattttctc	ccacgaactc	660
ggctttaagg	tgggtgttaag	acaagataag	gagtttcagg	atatcttaat	ggaccacaac	720
aggaaaagca	atgtgattat	tatgtgtggt	gggtccagagt	tcctctacaa	gctgaagggt	780
gaccgagcag	tggctgaaga	cattgtcatt	attctagtgg	atcttttcaa	tgaccagtac	840
ttggaggaca	atgtcacagc	ccctgactat	atgaaaaatg	tccttgttct	gacgctgtct	900
cctgggaatt	cccttctaaa	tagctctttc	tcagggaatc	tatcaccaac	aaaacgagac	960
tttgctcttg	cctatttgaa	tggaatcctg	ctctttggac	atatgctgaa	gatatttctt	1020
gaaaatggag	aaaatattac	cacccccaaa	tttgctcatg	ctttcaggaa	tctcactttt	1080
gaagggatatg	acgggtccagt	gaccttggat	gactgggggg	atggtgacag	taccatggtg	1140
cttctgtata	cctctgtgga	caccaagaaa	tacaaggttc	ttttgacctt	tgatacccac	1200
gtaaataaga	cctatcctgt	ggatatgagc	cccacattca	cttggagaag	ctctaaactt	1260
cctaatagata	ttacaggccg	gggccctcag	atcctgatga	ttgcagtctt	caccctcact	1320
ggagctgtgg	tgtctctcct	gctcgtcgtc	ctcctgatgc	tcagaaaata	tagaaaagat	1380
tatgaacttc	gtcagaaaaa	atgggtccac	attcctcctg	aaaatatctt	tcctctggag	1440
accaatgaga	ccaatcatgt	tagcctcaag	atcgatgatg	acaaaagacg	agataacaac	1500
cagagactac	gacagtgcaa	atacgacaaa	aagcgagtga	ttctcaaaga	tctcaagcac	1560
aatgatggta	attttcactga	aaaacagaag	atagaattga	acaagattga	ctattacaac	1620
ctgaccaagt	tctacggcac	agtgaactt	gataccatga	tcttcggggg	gatagaatac	1680
tgtgagagag	gatccctccg	ggaagtttta	aatgacacaa	tttcctaccc	tgatggcaca	1740
ttcatggatt	gggagtttaa	gatctctgtc	ttgtatgaca	ttgctaaggg	aatgtcatat	1800
ctgcactcca	gtaagacaga	agtcctatgg	cgtctgaaat	ctaccaactg	cgtagtggac	1860
agtagaatgg	tgggtgaagat	cactgatttt	ggctgcaatt	ccatttttacc	tccaaaaaag	1920
gacctgtgga	cagctccaga	gcacctccgc	caagccaaca	tctctcagaa	aggagatgtg	1980
tacagctatg	ggatcatcgc	acaggagatc	attctgcgga	aagaaacctt	ctacactttg	2040
agctgtcggg	accggaatga	gaagattttc	agagtggaaa	attccaatgg	aatgaaaccc	2100
ttccgcccag	atttatttct	ggaaacagca	gaggaaaaag	agctagaagt	gtacctactt	2160
gtaaaaaact	gttgggagga	agatccagaa	aagagaccag	atttcaaaaa	aattgagact	2220
acacttgcca	agatattttg	actttttcat	gaccaaataa	atgaaagcta	tatggatacc	2280
ttgatccgac	gtctacagct	atattctcga	aacctggaac	atctggtaga	ggaaaggaca	2340
cagctgtaca	aggcagagag	ggacagggct	gacagactta	actttatgtt	gcttccaagg	2400
ctagtggtaa	agtctctgaa	ggagaaaggc	tttgtggagc	cggaaactata	tgaggaagtt	2460
acaatctact	tcagtgcacat	tgtaggtttc	actactatct	gcaaatacac	cacccccatg	2520
gaagtgggtg	acatgcttaa	tgacatctat	aagagttttg	accacattgt	tgatcatcat	2580

gatgtctaca	aggtggaaac	catcggtgat	gcgtacatgg	tggctagtgg	tttgcctaag	2640
agaaatggca	atcggcatgc	aatagacatt	gccaagatgg	ccttggaaat	cctcagcttc	2700
atggggacct	ttgagctgga	gcattcttct	ggcctcccaa	tatggattcg	cattggagtt	2760
cactctggtc	cctgtgctgc	tggagttgtg	ggaatcaaga	tgcctcgtaa	ttgtctattt	2820
ggagatacgg	tcaacacagc	ctctaggatg	gaatccactg	gcctcccttt	gagaattcac	2880
gtgagtggct	ccaccatagc	catcctgaag	agaactgagt	gccagttcct	ttatgaagtg	2940
agaggagaaa	catacttaaa	gggaagagga	aatgagacta	cctactggct	gactgggatg	3000
aaggaccaga	aattcaacct	gccaaccctt	cctactgtgg	agaatcaaca	gcgtttgcaa	3060
gcagaatttt	cagacatgat	tgccaactct	ttacagaaaa	gacaggcagc	agggataaga	3120
agccaaaaac	ccagacgggt	agccagctat	aaaaaaggca	ctctggaata	cttgcagctg	3180
aataccacag	acaaggagag	cacctatttt	taa			3213

<210> 6
 <211> 550
 <212> DNA
 <213> Homo sapiens

<400> 6						
ggggacactt	tgtatggcaa	gtggaaccac	tggcttggtg	gatttttgcta	gattttttctg	60
attttttaaac	tcctgaaaaa	tatcccagat	aactgtcatg	aagctggtaa	ctatcttcct	120
gctggtgacc	atcagccttt	gtagtacttc	tgtactgcc	ttcctcatca	acaaagtgcc	180
ccttcctggt	gacaagttgg	cacctttacc	tctggacaac	attcttcctt	ttatggatcc	240
attaaagctt	cttctgaaaa	ctctgggcat	ttctgttgag	caccttgtgg	aggggctaag	300
gaagtgtgta	aatgagctgg	gaccagaggc	ttctgaagct	gtgaagaaac	tgctggagggc	360
gctatcacac	ttggtgtgac	atcaagataa	agagcggagg	tggatgggga	tggaagatga	420
tgtctctatc	ctccctgcct	gaaacctgtt	ctaccaatta	tagatcaaat	gccctaaaat	480
gtagtgaccc	gtgaaaagga	caaataaagc	aatgaatact	aaaaaaaaaa	aaaaaaaaaa	540
aaaaaaaaaa						550

<210> 7
 <211> 786
 <212> DNA
 <213> Homo sapiens

<400> 7						
atggccgtga	ctgcctgtca	gggcttgggg	ttcgtgggtt	cactgattgg	gattgcgggc	60
atcattgctg	ccacctgcat	ggaccagtgg	agcacccaag	acttgtacaa	caacccegtg	120
acagctgttt	tcaactacca	ggggctgtgg	cgctcctgtg	tcogagagag	ctctggcttc	180
accgagtgcc	ggggctactt	caccttgcct	gggctgccag	ccatgctgca	ggcagtgcga	240
gccctgatga	tcgtaggcat	cgctcctggg	gccattggcc	tcctgggtatc	catctttgcc	300
ctgaaatgca	tcgcatttgg	cagcatggag	gactctgcca	aagccaacat	gacactgacc	360
tcggggatca	tgttcattgt	ctcaggtctt	tgtgcaattg	ctggagtgtc	tgtgtttgcc	420
aacatgctgg	tgactaactt	ctggatgtcc	acagctaaca	tgtacaccgg	catgggtggg	480
atggtgcaga	ctgttcagac	caggtacaca	tttggtgccg	ctctgttcgt	gggctgggtc	540
gctggaggcc	tcacactaat	tgggggtgtg	atgatgtgca	tcgcctgccg	gggcctggca	600
ccagaagaaa	ccaactacaa	agccgtttct	tatcatgcct	caggccacag	tgttgcttac	660
aagcctggag	gcttcaaggc	cagcactggc	tttgggtcca	acaccaaaaa	caagaagata	720
tacgatggag	gtgcccgcac	agaggacgag	gtacaatctt	atccttccaa	gcacgactat	780
gtgtaa						786

<210> 8
 <211> 180

<212> DNA

<213> Homo sapiens

<400> 8

```

tgcgccacca tggccgtgac tgcctgtcag ggcttgggggt tcgtgggtttc actgattggg      60
attgcgggca tcattgctgc cacctgcatg gaccagtgga gcacccaaga cttgtacaac      120
aaccgccgtaa cagctgtttt caactaccag gggctgtggc gctcctgtgt ccgagagagc      180

```

<210> 9

<211> 309

<212> PRT

<213> Homo sapiens

<400> 9

```

Met Asn Gly Thr Tyr Asn Thr Cys Gly Ser Ser Asp Leu Thr Trp Pro
1          5          10          15
Pro Ala Ile Lys Leu Gly Phe Tyr Ala Tyr Leu Gly Val Leu Leu Val
          20          25          30
Leu Gly Leu Leu Leu Asn Ser Leu Ala Leu Trp Val Phe Cys Cys Arg
          35          40          45
Met Gln Gln Trp Thr Glu Thr Arg Ile Tyr Met Thr Asn Leu Ala Val
          50          55          60
Ala Asp Leu Cys Leu Leu Cys Thr Leu Pro Phe Val Leu His Ser Leu
65          70          75          80
Arg Asp Thr Ser Asp Thr Pro Leu Cys Gln Leu Ser Gln Gly Ile Tyr
          85          90          95
Leu Thr Asn Arg Tyr Met Ser Ile Ser Leu Val Thr Ala Ile Ala Val
          100          105          110
Asp Arg Tyr Val Ala Val Arg His Pro Leu Arg Ala Arg Gly Leu Arg
          115          120          125
Ser Pro Arg Gln Ala Ala Ala Val Cys Ala Val Leu Trp Val Leu Val
          130          135          140
Ile Gly Ser Leu Val Ala Arg Trp Leu Leu Gly Ile Gln Glu Gly Gly
145          150          155          160
Phe Cys Phe Arg Ser Thr Arg His Asn Phe Asn Ser Met Arg Phe Pro
          165          170          175
Leu Leu Gly Phe Tyr Leu Pro Leu Ala Val Val Val Phe Cys Ser Leu
          180          185          190
Lys Val Val Thr Ala Leu Ala Gln Arg Pro Pro Thr Asp Val Gly Gln
          195          200          205
Ala Glu Ala Thr Arg Lys Ala Ala Arg Met Val Trp Ala Asn Leu Leu
          210          215          220
Val Phe Val Val Cys Phe Leu Pro Leu His Val Gly Leu Thr Val Arg
225          230          235          240
Leu Ala Val Gly Trp Asn Ala Cys Ala Leu Leu Glu Thr Ile Arg Arg
          245          250          255
Ala Leu Tyr Ile Thr Ser Lys Leu Ser Asp Ala Asn Cys Cys Leu Asp
          260          265          270
Ala Ile Cys Tyr Tyr Tyr Met Ala Lys Glu Phe Gln Glu Ala Ser Ala
          275          280          285
Leu Ala Val Ala Pro Arg Ala Lys Ala His Lys Ser Gln Asp Ser Leu
          290          295          300
Cys Val Thr Leu Ala
305

```

<210> 10
 <211> 394
 <212> PRT
 <213> Homo sapiens

<400> 10
 Met Thr Ala Gly Arg Ser Gln Glu Arg Arg Ala Gln Glu Met Gly Arg
 1 5 10 15
 Gly Ser Val Gln Gly Leu Asp Leu Lys Gly Asp Leu Glu Phe Phe Thr
 20 25 30
 Ala Pro Met Leu Ser Leu Arg Ser Phe Val Phe Val Gly Val Gly Ser
 35 40 45
 Gly Leu Thr Ser Ser His Ile Pro Ala Gln Arg Trp Ala Glu Trp Gly
 50 55 60
 Gln Cys Leu Ala Pro Pro Ala Arg Ser Leu Leu Thr Ser Gly Ser Leu
 65 70 75 80
 Cys Cys Pro Arg Thr Met Asn Gly Thr Tyr Asn Thr Cys Gly Ser Ser
 85 90 95
 Asp Leu Thr Trp Pro Pro Ala Ile Lys Leu Gly Phe Tyr Ala Tyr Leu
 100 105 110
 Gly Val Leu Leu Val Leu Gly Leu Leu Leu Asn Ser Leu Ala Leu Trp
 115 120 125
 Val Phe Cys Cys Arg Met Gln Gln Trp Thr Glu Thr Arg Ile Tyr Met
 130 135 140
 Thr Asn Leu Ala Val Ala Asp Leu Cys Leu Leu Cys Thr Leu Pro Phe
 145 150 155 160
 Val Leu His Ser Leu Arg Asp Thr Ser Asp Thr Pro Leu Cys Gln Leu
 165 170 175
 Ser Gln Gly Ile Tyr Leu Thr Asn Arg Tyr Met Ser Ile Ser Leu Val
 180 185 190
 Thr Ala Ile Ala Val Asp Arg Tyr Val Ala Val Arg His Pro Leu Arg
 195 200 205
 Ala Arg Gly Leu Arg Ser Pro Arg Gln Ala Ala Ala Val Cys Ala Val
 210 215 220
 Leu Trp Val Leu Val Ile Gly Ser Leu Val Ala Arg Trp Leu Leu Gly
 225 230 235 240
 Ile Gln Glu Gly Gly Phe Cys Phe Arg Ser Thr Arg His Asn Phe Asn
 245 250 255
 Ser Met Ala Phe Pro Leu Leu Gly Phe Tyr Leu Pro Leu Ala Val Val
 260 265 270
 Val Phe Cys Ser Leu Lys Val Val Thr Ala Leu Ala Gln Arg Pro Pro
 275 280 285
 Thr Asp Val Gly Gln Ala Glu Ala Thr Arg Lys Ala Ala Arg Met Val
 290 295 300
 Trp Ala Asn Leu Leu Val Phe Val Val Cys Phe Leu Pro Leu His Val
 305 310 315 320
 Gly Leu Thr Val Arg Leu Ala Val Gly Trp Asn Ala Cys Ala Leu Leu
 325 330 335
 Glu Thr Ile Arg Arg Ala Leu Tyr Ile Thr Ser Lys Leu Ser Asp Ala
 340 345 350
 Asn Cys Cys Leu Asp Ala Ile Cys Tyr Tyr Tyr Met Ala Lys Glu Phe
 355 360 365
 Gln Glu Ala Ser Ala Leu Ala Val Ala Pro Ser Ala Lys Ala His Lys
 370 375 380
 Ser Gln Asp Ser Leu Cys Val Thr Leu Ala
 385 390

<210> 11
 <211> 1073
 <212> PRT
 <213> Homo sapiens

<400> 11

Met	Lys	Thr	Leu	Leu	Leu	Asp	Leu	Ala	Leu	Trp	Ser	Leu	Leu	Phe	Gln
1			5						10					15	
Pro	Gly	Trp	Leu	Ser	Phe	Ser	Ser	Gln	Val	Ser	Gln	Asn	Cys	His	Asn
			20					25					30		
Gly	Ser	Tyr	Glu	Ile	Ser	Val	Leu	Met	Met	Gly	Asn	Ser	Ala	Phe	Ala
		35					40					45			
Glu	Pro	Leu	Lys	Asn	Leu	Glu	Asp	Ala	Val	Asn	Glu	Gly	Leu	Glu	Ile
	50					55					60				
Val	Arg	Gly	Arg	Leu	Gln	Asn	Ala	Gly	Leu	Asn	Val	Thr	Val	Asn	Ala
65					70					75					80
Thr	Phe	Met	Tyr	Ser	Asp	Gly	Leu	Ile	His	Asn	Ser	Gly	Asp	Cys	Arg
				85					90					95	
Ser	Ser	Thr	Cys	Glu	Gly	Leu	Asp	Leu	Leu	Arg	Lys	Ile	Ser	Asn	Ala
			100					105						110	
Gln	Arg	Met	Gly	Cys	Val	Leu	Ile	Gly	Pro	Ser	Cys	Thr	Tyr	Ser	Thr
		115					120						125		
Phe	Gln	Met	Tyr	Leu	Asp	Thr	Glu	Leu	Ser	Tyr	Pro	Met	Ile	Ser	Ala
	130					135						140			
Gly	Ser	Phe	Gly	Leu	Ser	Cys	Asp	Tyr	Lys	Glu	Thr	Leu	Thr	Arg	Leu
145					150					155					160
Met	Ser	Pro	Ala	Arg	Lys	Leu	Met	Tyr	Phe	Leu	Val	Asn	Phe	Trp	Lys
				165					170					175	
Thr	Asn	Asp	Leu	Pro	Phe	Lys	Thr	Tyr	Ser	Trp	Ser	Thr	Ser	Tyr	Val
			180					185						190	
Tyr	Lys	Asn	Gly	Thr	Glu	Thr	Glu	Asp	Cys	Phe	Trp	Tyr	Leu	Asn	Ala
		195					200						205		
Leu	Glu	Ala	Ser	Val	Ser	Tyr	Phe	Ser	His	Glu	Leu	Gly	Phe	Lys	Val
	210					215					220				
Val	Leu	Arg	Gln	Asp	Lys	Glu	Phe	Gln	Asp	Ile	Leu	Met	Asp	His	Asn
225					230					235					240
Arg	Lys	Ser	Asn	Val	Ile	Ile	Met	Cys	Gly	Gly	Pro	Glu	Phe	Leu	Tyr
			245						250					255	
Lys	Leu	Lys	Gly	Asp	Arg	Ala	Val	Ala	Glu	Asp	Ile	Val	Ile	Ile	Leu
			260					265						270	
Val	Asp	Leu	Phe	Asn	Asp	Gln	Tyr	Leu	Glu	Asp	Asn	Val	Thr	Ala	Pro
		275					280						285		
Asp	Tyr	Met	Lys	Asn	Val	Leu	Val	Leu	Thr	Leu	Ser	Pro	Gly	Asn	Ser
	290					295						300			
Leu	Leu	Asn	Ser	Ser	Phe	Ser	Arg	Asn	Leu	Ser	Pro	Thr	Lys	Arg	Asp
305					310					315					320
Phe	Ala	Leu	Ala	Tyr	Leu	Asn	Gly	Ile	Leu	Leu	Phe	Gly	His	Met	Leu
			325						330					335	
Lys	Ile	Phe	Leu	Glu	Asn	Gly	Glu	Asn	Ile	Thr	Thr	Pro	Lys	Phe	Ala
			340					345						350	
His	Ala	Phe	Arg	Asn	Leu	Thr	Phe	Glu	Gly	Tyr	Asp	Gly	Pro	Val	Thr
		355					360					365			
Leu	Asp	Asp	Trp	Gly	Asp	Val	Asp	Ser	Thr	Met	Val	Leu	Leu	Tyr	Thr
	370					375						380			
Ser	Val	Asp	Thr	Lys	Lys	Tyr	Lys	Val	Leu	Leu	Thr	Tyr	Asp	Thr	His
385					390					395					400

Val	Asn	Lys	Thr	Tyr	Pro	Val	Asp	Met	Ser	Pro	Thr	Phe	Thr	Trp	Lys	405	410	415
Asn	Ser	Lys	Leu	Pro	Asn	Asp	Ile	Thr	Gly	Arg	Gly	Pro	Gln	Ile	Leu	420	425	430
Met	Ile	Ala	Val	Phe	Thr	Leu	Thr	Gly	Ala	Val	Val	Leu	Leu	Leu	Leu	435	440	445
Val	Ala	Leu	Leu	Met	Leu	Arg	Lys	Tyr	Arg	Lys	Asp	Tyr	Glu	Leu	Arg	450	455	460
Gln	Lys	Lys	Trp	Ser	His	Ile	Pro	Pro	Glu	Asn	Ile	Phe	Pro	Leu	Glu	465	470	475
Thr	Asn	Glu	Thr	Asn	His	Val	Ser	Leu	Lys	Ile	Asp	Asp	Asp	Lys	Arg	485	490	495
Arg	Asp	Thr	Ile	Gln	Arg	Leu	Arg	Gln	Cys	Lys	Tyr	Asp	Lys	Lys	Arg	500	505	510
Val	Ile	Leu	Lys	Asp	Leu	Lys	His	Asn	Asp	Gly	Asn	Phe	Thr	Glu	Lys	515	520	525
Gln	Lys	Ile	Glu	Leu	Asn	Lys	Leu	Leu	Gln	Ile	Asp	Tyr	Tyr	Asn	Leu	530	535	540
Thr	Lys	Phe	Tyr	Gly	Thr	Val	Lys	Leu	Asp	Thr	Met	Ile	Phe	Gly	Val	545	550	555
Ile	Glu	Tyr	Cys	Glu	Arg	Gly	Ser	Leu	Arg	Glu	Val	Leu	Asn	Asp	Thr	565	570	575
Ile	Ser	Tyr	Pro	Asp	Gly	Thr	Phe	Met	Asp	Trp	Glu	Phe	Lys	Ile	Ser	580	585	590
Val	Leu	Tyr	Asp	Ile	Ala	Lys	Gly	Met	Ser	Tyr	Leu	His	Ser	Ser	Lys	595	600	605
Thr	Glu	Val	His	Gly	Arg	Leu	Lys	Ser	Thr	Asn	Cys	Val	Val	Asp	Ser	610	615	620
Arg	Met	Val	Val	Lys	Ile	Thr	Asp	Phe	Gly	Cys	Asn	Ser	Ile	Leu	Pro	625	630	635
Pro	Lys	Lys	Asp	Leu	Trp	Thr	Ala	Pro	Glu	His	Leu	Arg	Gln	Ala	Asn	645	650	655
Ile	Ser	Gln	Lys	Gly	Asp	Val	Tyr	Ser	Tyr	Gly	Ile	Ile	Ala	Gln	Glu	660	665	670
Ile	Ile	Leu	Arg	Lys	Glu	Thr	Phe	Tyr	Thr	Leu	Ser	Cys	Arg	Asp	Arg	675	680	685
Asn	Glu	Lys	Ile	Phe	Arg	Val	Glu	Asn	Ser	Asn	Gly	Met	Lys	Pro	Phe	690	695	700
Arg	Pro	Asp	Leu	Phe	Leu	Glu	Thr	Ala	Glu	Glu	Lys	Glu	Leu	Glu	Val	705	710	715
Tyr	Leu	Leu	Val	Lys	Asn	Cys	Trp	Glu	Glu	Asp	Pro	Glu	Lys	Arg	Pro	725	730	735
Asp	Phe	Lys	Lys	Ile	Glu	Thr	Thr	Leu	Ala	Lys	Ile	Phe	Gly	Leu	Phe	740	745	750
His	Asp	Gln	Lys	Asn	Glu	Ser	Tyr	Met	Asp	Thr	Leu	Ile	Arg	Arg	Leu	755	760	765
Gln	Leu	Tyr	Ser	Arg	Asn	Leu	Glu	His	Leu	Val	Glu	Glu	Arg	Thr	Gln	770	775	780
Leu	Tyr	Lys	Ala	Glu	Arg	Asp	Arg	Ala	Asp	Arg	Leu	Asn	Phe	Met	Leu	785	790	795
Leu	Pro	Arg	Leu	Val	Val	Lys	Ser	Leu	Lys	Glu	Lys	Gly	Phe	Val	Glu	805	810	815
Pro	Glu	Leu	Tyr	Glu	Glu	Val	Thr	Ile	Tyr	Phe	Ser	Asp	Ile	Val	Gly	820	825	830
Phe	Thr	Thr	Ile	Cys	Lys	Tyr	Ser	Thr	Pro	Met	Glu	Val	Val	Asp	Met	835	840	845

Leu Asn Asp Ile Tyr Lys Ser Phe Asp His Ile Val Asp His His Asp
 850 855 860
 Val Tyr Lys Val Glu Thr Ile Gly Asp Ala Tyr Met Val Ala Ser Gly
 865 870 875 880
 Leu Pro Lys Arg Asn Gly Asn Arg His Ala Ile Asp Ile Ala Lys Met
 885 890 895
 Ala Leu Glu Ile Leu Ser Phe Met Gly Thr Phe Glu Leu Glu His Leu
 900 905 910
 Pro Gly Leu Pro Ile Trp Ile Arg Ile Gly Val His Ser Gly Pro Cys
 915 920 925
 Ala Ala Gly Val Val Gly Ile Lys Met Pro Arg Tyr Cys Leu Phe Gly
 930 935 940
 Asp Thr Val Asn Thr Ala Ser Arg Met Glu Ser Thr Gly Leu Pro Leu
 945 950 955 960
 Arg Ile His Val Ser Gly Ser Thr Ile Ala Ile Leu Lys Arg Thr Glu
 965 970 975
 Cys Gln Phe Leu Tyr Glu Val Arg Gly Glu Thr Tyr Leu Lys Gly Arg
 980 985 990
 Gly Asn Glu Thr Thr Tyr Trp Leu Thr Gly Met Lys Asp Gln Lys Phe
 995 1000 1005
 Asn Leu Pro Thr Pro Pro Thr Val Glu Asn Gln Gln Arg Leu Gln
 1010 1015 1020
 Ala Glu Phe Ser Asp Met Ile Ala Asn Ser Leu Gln Lys Arg Gln
 1025 1030 1035
 Ala Ala Gly Ile Arg Ser Gln Lys Pro Arg Arg Val Ala Ser Tyr
 1040 1045 1050
 Lys Lys Gly Thr Leu Glu Tyr Leu Gln Leu Asn Thr Thr Asp Lys
 1055 1060 1065
 Glu Ser Thr Tyr Phe
 1070

<210> 12
 <211> 111
 <212> PRT
 <213> Homo sapiens

<400> 12
 Met Lys Thr Leu Leu Leu Asp Leu Ala Leu Trp Ser Leu Leu Phe Gln
 1 5 10 15
 Pro Gly Trp Leu Ser Phe Ser Ser Gln Val Ser Gln Asn Cys His Asn
 20 25 30
 Gly Ser Tyr Glu Ile Ser Val Leu Met Met Gly Asn Ser Ala Phe Ala
 35 40 45
 Glu Pro Leu Lys Asn Leu Glu Asp Ala Val Asn Glu Gly Leu Glu Ile
 50 55 60
 Val Arg Gly Arg Leu Gln Asn Ala Gly Leu Asn Val Thr Val Asn Ala
 65 70 75 80
 Thr Phe Met Tyr Ser Asp Gly Leu Ile His Asn Ser Gly Asp Cys Arg
 85 90 95
 Ser Ser Thr Cys Glu Gly Leu Asp Leu Leu Arg Lys Ile Ser Pro
 100 105 110

<210> 13
 <211> 258

<212> PRT
 <213> Homo sapiens

<400> 13
 Met Lys Thr Leu Leu Leu Asp Leu Ala Leu Trp Ser Leu Leu Phe Gln
 1 5 10 15
 Pro Gly Trp Leu Ser Phe Ser Ser Gln Val Ser Gln Asn Cys His Asn
 20 25 30
 Gly Ser Tyr Glu Ile Ser Val Leu Met Met Gly Asn Ser Ala Phe Ala
 35 40 45
 Glu Pro Leu Lys Asn Leu Glu Asp Ala Val Asn Glu Gly Leu Glu Ile
 50 55 60
 Val Arg Gly Arg Leu Gln Asn Ala Gly Leu Asn Val Thr Val Asn Ala
 65 70 75 80
 Thr Phe Met Tyr Ser Asp Gly Leu Ile His Asn Ser Gly Asp Cys Arg
 85 90 95
 Ser Ser Thr Cys Glu Gly Leu Asp Leu Leu Arg Lys Ile Ser Asn Ala
 100 105 110
 Gln Arg Met Gly Cys Val Leu Ile Gly Pro Ser Cys Thr Tyr Ser Thr
 115 120 125
 Phe Gln Met Tyr Leu Asp Thr Glu Leu Ser Tyr Pro Met Ile Ser Ala
 130 135 140
 Gly Ser Phe Gly Leu Ser Cys Asp Tyr Lys Glu Thr Leu Thr Arg Leu
 145 150 155 160
 Met Ser Pro Ala Arg Lys Leu Met Tyr Phe Leu Val Asn Phe Trp Lys
 165 170 175
 Thr Asn Asp Leu Pro Phe Lys Thr Tyr Ser Trp Ser Thr Ser Tyr Val
 180 185 190
 Tyr Lys Asn Gly Thr Glu Thr Glu Asp Cys Phe Trp Tyr Leu Asn Ala
 195 200 205
 Leu Glu Ala Ser Val Ser Tyr Phe Ser His Glu Leu Gly Phe Lys Val
 210 215 220
 Val Leu Arg Gln Asp Lys Glu Phe Gln Asp Ile Leu Met Asp His Asn
 225 230 235 240
 Arg Lys Ser Asn Val Thr Ser Thr Trp Arg Thr Met Ser Gln Pro Leu
 245 250 255
 Thr Ile

<210> 14
 <211> 1070
 <212> PRT
 <213> Homo sapiens

<400> 14
 Met Lys Thr Leu Leu Leu Asp Leu Ala Leu Trp Ser Leu Leu Phe Gln
 1 5 10 15
 Pro Gly Trp Leu Ser Phe Ser Ser Gln Val Ser Gln Asn Cys His Asn
 20 25 30
 Gly Ser Tyr Glu Ile Ser Val Leu Met Met Gly Asn Ser Ala Phe Ala
 35 40 45
 Glu Pro Leu Lys Asn Leu Glu Asp Ala Val Asn Glu Gly Leu Glu Ile
 50 55 60
 Val Arg Gly Arg Leu Gln Asn Ala Gly Leu Asn Val Thr Val Asn Ala
 65 70 75 80

Thr	Phe	Met	Tyr	Ser	Asp	Gly	Leu	Ile	His	Asn	Ser	Gly	Asp	Cys	Arg	
				85					90					95		
Ser	Ser	Thr	Cys	Glu	Gly	Leu	Asp	Leu	Leu	Arg	Lys	Ile	Ser	Asn	Ala	
			100					105						110		
Gln	Arg	Met	Gly	Cys	Val	Leu	Ile	Gly	Pro	Ser	Cys	Thr	Tyr	Ser	Thr	
		115						120						125		
Phe	Gln	Met	Tyr	Leu	Asp	Thr	Glu	Leu	Ser	Tyr	Pro	Met	Ile	Ser	Ala	
		130					135							140		
Gly	Ser	Phe	Gly	Leu	Ser	Cys	Asp	Tyr	Lys	Glu	Thr	Leu	Thr	Arg	Leu	
145						150				155					160	
Met	Ser	Pro	Ala	Arg	Lys	Leu	Met	Tyr	Phe	Leu	Val	Asn	Phe	Trp	Lys	
				165					170					175		
Thr	Asn	Asp	Leu	Pro	Phe	Lys	Thr	Tyr	Ser	Trp	Ser	Thr	Ser	Tyr	Val	
			180					185						190		
Tyr	Lys	Asn	Gly	Thr	Glu	Thr	Glu	Asp	Cys	Phe	Trp	Tyr	Leu	Asn	Ala	
		195					200							205		
Leu	Glu	Ala	Ser	Val	Ser	Tyr	Phe	Ser	His	Glu	Leu	Gly	Phe	Lys	Val	
		210					215							220		
Val	Leu	Arg	Gln	Asp	Lys	Glu	Phe	Gln	Asp	Ile	Leu	Met	Asp	His	Asn	
225					230					235					240	
Arg	Lys	Ser	Asn	Val	Ile	Ile	Met	Cys	Gly	Gly	Pro	Glu	Phe	Leu	Tyr	
				245					250					255		
Lys	Leu	Lys	Gly	Asp	Arg	Ala	Val	Ala	Glu	Asp	Ile	Val	Ile	Ile	Leu	
			260					265						270		
Val	Asp	Leu	Phe	Asn	Asp	Gln	Tyr	Leu	Glu	Asp	Asn	Val	Thr	Ala	Pro	
		275					280							285		
Asp	Tyr	Met	Lys	Asn	Val	Leu	Val	Leu	Thr	Leu	Ser	Pro	Gly	Asn	Ser	
		290				295								300		
Leu	Leu	Asn	Ser	Ser	Phe	Ser	Arg	Asn	Leu	Ser	Pro	Thr	Lys	Arg	Asp	
305					310					315					320	
Phe	Ala	Leu	Ala	Tyr	Leu	Asn	Gly	Ile	Leu	Leu	Phe	Gly	His	Met	Leu	
				325					330					335		
Lys	Ile	Phe	Leu	Glu	Asn	Gly	Glu	Asn	Ile	Thr	Thr	Pro	Lys	Phe	Ala	
			340					345						350		
His	Ala	Phe	Arg	Asn	Leu	Thr	Phe	Glu	Gly	Tyr	Asp	Gly	Pro	Val	Thr	
		355					360							365		
Leu	Asp	Asp	Trp	Gly	Asp	Val	Asp	Ser	Thr	Met	Val	Leu	Leu	Tyr	Thr	
		370				375						380				
Ser	Val	Asp	Thr	Lys	Lys	Tyr	Lys	Val	Leu	Leu	Thr	Tyr	Asp	Thr	His	
385					390					395					400	
Val	Asn	Lys	Thr	Tyr	Pro	Val	Asp	Met	Ser	Pro	Thr	Phe	Thr	Trp	Lys	
				405					410					415		
Asn	Ser	Lys	Leu	Pro	Asn	Asp	Ile	Thr	Gly	Arg	Gly	Pro	Gln	Ile	Leu	
			420					425						430		
Met	Ile	Ala	Val	Phe	Thr	Leu	Thr	Gly	Ala	Val	Val	Leu	Leu	Leu	Leu	
		435					440							445		
Val	Ala	Leu	Leu	Met	Leu	Arg	Lys	Tyr	Arg	Lys	Asp	Tyr	Glu	Leu	Arg	
		450				455						460				
Gln	Lys	Lys	Trp	Ser	His	Ile	Pro	Pro	Glu	Asn	Ile	Phe	Pro	Leu	Glu	
465					470					475					480	
Thr	Asn	Glu	Thr	Asn	His	Val	Ser	Leu	Lys	Ile	Asp	Asp	Asp	Lys	Arg	
				485					490					495		
Arg	Asp	Thr	Ile	Gln	Arg	Leu	Arg	Gln	Cys	Lys	Tyr	Asp	Lys	Lys	Arg	
		500						505						510		
Val	Ile	Leu	Lys	Asp	Leu	Lys	His	Asn	Asp	Gly	Asn	Phe	Thr	Glu	Lys	
		515					520					525				

Gln	Lys	Ile	Glu	Leu	Asn	Lys	Ile	Asp	Tyr	Tyr	Asn	Leu	Thr	Lys	Phe
530						535					540				
Tyr	Gly	Thr	Val	Lys	Leu	Asp	Thr	Met	Ile	Phe	Gly	Val	Ile	Glu	Tyr
545					550					555					560
Cys	Glu	Arg	Gly	Ser	Leu	Arg	Glu	Val	Leu	Asn	Asp	Thr	Ile	Ser	Tyr
				565					570					575	
Pro	Asp	Gly	Thr	Phe	Met	Asp	Trp	Glu	Phe	Lys	Ile	Ser	Val	Leu	Tyr
			580					585					590		
Asp	Ile	Ala	Lys	Gly	Met	Ser	Tyr	Leu	His	Ser	Ser	Lys	Thr	Glu	Val
		595					600					605			
His	Gly	Arg	Leu	Lys	Ser	Thr	Asn	Cys	Val	Val	Asp	Ser	Arg	Met	Val
610						615					620				
Val	Lys	Ile	Thr	Asp	Phe	Gly	Cys	Asn	Ser	Ile	Leu	Pro	Pro	Lys	Lys
625					630					635					640
Asp	Leu	Trp	Thr	Ala	Pro	Glu	His	Leu	Arg	Gln	Ala	Asn	Ile	Ser	Gln
				645					650					655	
Lys	Gly	Asp	Val	Tyr	Ser	Tyr	Gly	Ile	Ile	Ala	Gln	Glu	Ile	Ile	Leu
			660					665						670	
Arg	Lys	Glu	Thr	Phe	Tyr	Thr	Leu	Ser	Cys	Arg	Asp	Arg	Asn	Glu	Lys
			675				680					685			
Ile	Phe	Arg	Val	Glu	Asn	Ser	Asn	Gly	Met	Lys	Pro	Phe	Arg	Pro	Asp
690						695					700				
Leu	Phe	Leu	Glu	Thr	Ala	Glu	Glu	Lys	Glu	Leu	Glu	Val	Tyr	Leu	Leu
705					710					715					720
Val	Lys	Asn	Cys	Trp	Glu	Glu	Asp	Pro	Glu	Lys	Arg	Pro	Asp	Phe	Lys
				725					730					735	
Lys	Ile	Glu	Thr	Thr	Leu	Ala	Lys	Ile	Phe	Gly	Leu	Phe	His	Asp	Gln
			740					745					750		
Lys	Asn	Glu	Ser	Tyr	Met	Asp	Thr	Leu	Ile	Arg	Arg	Leu	Gln	Leu	Tyr
		755				760						765			
Ser	Arg	Asn	Leu	Glu	His	Leu	Val	Glu	Glu	Arg	Thr	Gln	Leu	Tyr	Lys
770						775						780			
Ala	Glu	Arg	Asp	Arg	Ala	Asp	Arg	Leu	Asn	Phe	Met	Leu	Leu	Pro	Arg
785					790					795					800
Leu	Val	Val	Lys	Ser	Leu	Lys	Glu	Lys	Gly	Phe	Val	Glu	Pro	Glu	Leu
				805					810					815	
Tyr	Glu	Glu	Val	Thr	Ile	Tyr	Phe	Ser	Asp	Ile	Val	Gly	Phe	Thr	Thr
			820					825					830		
Ile	Cys	Lys	Tyr	Ser	Thr	Pro	Met	Glu	Val	Val	Asp	Met	Leu	Asn	Asp
		835					840					845			
Ile	Tyr	Lys	Ser	Phe	Asp	His	Ile	Val	Asp	His	His	Asp	Val	Tyr	Lys
850						855					860				
Val	Glu	Thr	Ile	Gly	Asp	Ala	Tyr	Met	Val	Ala	Ser	Gly	Leu	Pro	Lys
865					870					875					880
Arg	Asn	Gly	Asn	Arg	His	Ala	Ile	Asp	Ile	Ala	Lys	Met	Ala	Leu	Glu
				885					890					895	
Ile	Leu	Ser	Phe	Met	Gly	Thr	Phe	Glu	Leu	Glu	His	Leu	Pro	Gly	Leu
			900					905					910		
Pro	Ile	Trp	Ile	Arg	Ile	Gly	Val	His	Ser	Gly	Pro	Cys	Ala	Ala	Gly
		915					920					925			
Val	Val	Gly	Ile	Lys	Met	Pro	Arg	Tyr	Cys	Leu	Phe	Gly	Asp	Thr	Val
		930				935					940				
Asn	Thr	Ala	Ser	Arg	Met	Glu	Ser	Thr	Gly	Leu	Pro	Leu	Arg	Ile	His
945					950					955					960
Val	Ser	Gly	Ser	Thr	Ile	Ala	Ile	Leu	Lys	Arg	Thr	Glu	Cys	Gln	Phe
				965					970					975	

Leu Tyr Glu Val Arg Gly Glu Thr Tyr Leu Lys Gly Arg Gly Asn Glu
 980 985 990
 Thr Thr Tyr Trp Leu Thr Gly Met Lys Asp Gln Lys Phe Asn Leu Pro
 995 1000 1005
 Thr Pro Pro Thr Val Glu Asn Gln Gln Arg Leu Gln Ala Glu Phe
 1010 1015 1020
 Ser Asp Met Ile Ala Asn Ser Leu Gln Lys Arg Gln Ala Ala Gly
 1025 1030 1035
 Ile Arg Ser Gln Lys Pro Arg Arg Val Ala Ser Tyr Lys Lys Gly
 1040 1045 1050
 Thr Leu Glu Tyr Leu Gln Leu Asn Thr Thr Asp Lys Glu Ser Thr
 1055 1060 1065
 Tyr Phe
 1070

<210> 15
 <211> 93
 <212> PRT
 <213> Homo sapiens

<400> 15
 Met Lys Leu Val Thr Ile Phe Leu Leu Val Thr Ile Ser Leu Cys Ser
 1 5 10 15
 Tyr Ser Ala Thr Ala Lys Leu Ile Asn Lys Cys Pro Leu Pro Val Asp
 20 25 30
 Lys Leu Ala Pro Leu Pro Leu Asp Asn Ile Leu Pro Phe Met Asp Pro
 35 40 45
 Leu Lys Leu Leu Leu Lys Thr Leu Gly Ile Ser Val Glu His Leu Val
 50 55 60
 Glu Gly Leu Arg Lys Cys Val Asn Glu Leu Gly Pro Glu Ala Ser Glu
 65 70 75 80
 Ala Val Lys Lys Leu Leu Glu Ala Leu Ser His Leu Val
 85 90

<210> 16
 <211> 261
 <212> PRT
 <213> Homo sapiens

<400> 16
 Met Ala Val Thr Ala Cys Gln Gly Leu Gly Phe Val Val Ser Leu Ile
 1 5 10 15
 Gly Ile Ala Gly Ile Ile Ala Ala Thr Cys Met Asp Gln Trp Ser Thr
 20 25 30
 Gln Asp Leu Tyr Asn Asn Pro Val Thr Ala Val Phe Asn Tyr Gln Gly
 35 40 45
 Leu Trp Arg Ser Cys Val Arg Glu Ser Ser Gly Phe Thr Glu Cys Arg
 50 55 60
 Gly Tyr Phe Thr Leu Leu Gly Leu Pro Ala Met Leu Gln Ala Val Arg
 65 70 75 80
 Ala Leu Met Ile Val Gly Ile Val Leu Gly Ala Ile Gly Leu Leu Val
 85 90 95
 Ser Ile Phe Ala Leu Lys Cys Ile Arg Ile Gly Ser Met Glu Asp Ser
 100 105 110

Ala Lys Ala Asn Met Thr Leu Thr Ser Gly Ile Met Phe Ile Val Ser
 115 120 125
 Gly Leu Cys Ala Ile Ala Gly Val Ser Val Phe Ala Asn Met Leu Val
 130 135 140
 Thr Asn Phe Trp Met Ser Thr Ala Asn Met Tyr Thr Gly Met Gly Gly
 145 150 155 160
 Met Val Gln Thr Val Gln Thr Arg Tyr Thr Phe Gly Ala Ala Leu Phe
 165 170 175
 Val Gly Trp Val Ala Gly Gly Leu Thr Leu Ile Gly Gly Val Met Met
 180 185 190
 Cys Ile Ala Cys Arg Gly Leu Ala Pro Glu Glu Thr Asn Tyr Lys Ala
 195 200 205
 Val Ser Tyr His Ala Ser Gly His Ser Val Ala Tyr Lys Pro Gly Gly
 210 215 220
 Phe Lys Ala Ser Thr Gly Phe Gly Ser Asn Thr Lys Asn Lys Lys Ile
 225 230 235 240
 Tyr Asp Gly Gly Ala Arg Thr Glu Asp Glu Val Gln Ser Tyr Pro Ser
 245 250 255
 Lys His Asp Tyr Val
 260

<210> 17
 <211> 10
 <212> PRT
 <213> Homo sapiens

<400> 17
 Asp Gln Trp Ser Thr Gln Asp Leu Tyr Asn
 1 5 10

<210> 18
 <211> 11
 <212> PRT
 <213> Homo sapiens

<400> 18
 Asn Asn Pro Val Thr Ala Val Phe Asn Tyr Gln
 1 5 10

<210> 19
 <211> 47
 <212> PRT
 <213> Homo sapiens

<400> 19
 Met Ala Val Thr Ala Cys Gln Gly Leu Gly Phe Val Val Ser Leu Ile
 1 5 10 15
 Gly Ile Ala Gly Ile Ile Ala Ala Thr Cys Met Asp Gln Trp Ser Thr
 20 25 30
 Gln Asp Leu Tyr Asn Asn Pro Val Thr Ala Val Phe Asn Tyr Gln
 35 40 45

<210> 20
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 20
 aggtacatga gcatcagcct g 21

<210> 21
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 21
 gcagcagttg gcatctgaga g 21

<210> 22
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 22
 gcaatagaca ttgccaagat g 21

<210> 23
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 23
 aacgctgttg attctccaca g 21

<210> 24
 <211> 33
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 24
 ggatcctcct ttagttccca ggtgagtcag aac 33

<210> 25
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 25
 tgctctggag gctagcgttt c 21

<210> 26
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 26
 accaatcatg ttagcctcaa g 21

<210> 27
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 27
 agctatggga tcatcgaca g 21

<210> 28
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 28
 cctttgagct ggagcatctt c 21

<210> 29
 <211> 21

<212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 29
 ctttctagct ggagacatca g 21

<210> 30
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 30
 caccatggta ctgtcaacat c 21

<210> 31
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 31
 atgtcataca agacagagat c 21

<210> 32
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 32
 tctgccttgt acagctgtgt c 21

<210> 33
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 33
 tctgtggtat tcagctgcaa g 21

<210> 34
 <211> 22
 <212> DNA
 <213> Artificial Sequence

<220>

<223> Description of the artificial sequence: Oligonucleotide

<400> 34
 tactcaggaa aatttcacct tg

22

<210> 35
 <211> 27
 <212> DNA
 <213> Artificial Sequence

<220>

<223> Description of the artificial sequence: Oligonucleotide

<400> 35
 gaccacaaca ggaaaagcaa tgtgacc

27

<210> 36
 <211> 22
 <212> DNA
 <213> Artificial Sequence

<220>

<223> Description of the artificial sequence: Oligonucleotide

<400> 36
 gatagaattg aacaagattg ac

22

<210> 37
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>

<223> Description of the artificial sequence: Oligonucleotide

<400> 37
 cagcctttgt agttactctg c

21

<210> 38
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>

<223> Description of the artificial sequence: Oligonucleotide

<400> 38
tgtcacacca agtgtgatag c 21

<210> 39
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of the artificial sequence: Oligonucleotide

<400> 39
ggttcgtggt ttcactgatt gggattgc 28

<210> 40
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of the artificial sequence: Oligonucleotide

<400> 40
cggctttgta gttggtttct tctggtg 27

<210> 41
<211> 3814
<212> DNA
<213> Homo sapiens

<400> 41
ctattgaagc cacctgctca ggacaatgaa attcttcagt tacattctgg tttatcgccg 60
atttctcttc gtggttttca ctgtgttggg tttactacct ctgcccatcg tctccacac 120
caaggaagca gaatgtgcct acacactctt tgtggtcgcc acattttggc tcacagaagc 180
attgcctctg tcggtaacag ctttgctacc tagtttaatg ttacccatgt ttgggatcat 240
gccttctaag aaggtggcat ctgcttattt caaggatttt cacttactgc taattggagt 300
tatctgttta gcaacatcca tagaaaaatg gaatttgca aagagaattg ctctgaaaat 360
ggtgatgatg gttggtgtaa atcctgcatg gctgacgctg gggttcatga gcagcactgc 420
ctttttgtct atgtggetca gcaacacctc gacggctgcc atggtgatgc ccattgcgga 480
ggctgtagtg cagcagatca tcaatgcaga agcagagggtc gagggccactc agatgactta 540
cttcaacgga tcaaccaacc acggactaga aattgatgaa agtggttaatg gacatgaaat 600
aatgagagg aaagagaaaa caaaaccagt tccaggatac aataatgata cagggaaaat 660
ttcaagcaag gtggagttgg aaaagaactc aggcattgaga accaaatata gaacaaagaa 720
gggccacgtg acacgtaaac ttacgtgttt gtgcattgcc tactcttcta ccattggtgg 780
actgacaaca atcactggta cctccaccaa cttgatcttt gcagagtatt tcaatacacg 840
ctatcttgac tgtcgttgcc tcaactttgg atcatggttt acgttttcct tcccagctgc 900
ccttatcatt ctactcttat cctggatctg gcttcagtgg cttttcctag gattcaattt 960
taaggagatg ttcaaatgtg gcaaaaccaa aacagtccaa caaaaagctt gtgctgaggt 1020
gattaagcaa gaataccaaa agcttgggcc aataaggtat caagaaattg tgaccttggg 1080
cctcttcatt ataatggctc tgctatgggt tagtcgagac cccggatttg ttcctgggtg 1140
gtctgcactt ttttcagagt accctgggtt tgctacagat tcaactgttg ctttacttat 1200
agggtgcta ttctttctta tcccagctaa gacactgact aaaactacac ctacaggaga 1260
aattgttgct tttgattact ctccactgat tacttggaag gaattccagt cattcatgcc 1320

ctgggatata	gccattcttg	ttggtggagg	gtttgccctg	gcagatgggt	gtgaggagtc	1380
tggattatct	aagtggatag	gaaataaatt	atctcctctg	ggttcattac	cagcatggct	1440
aataattctg	atatcttctt	tgatgggtgac	atctttaact	gaggtagcca	gcaatccagc	1500
taccattaca	ctctttctcc	caatattatc	tccattggcc	gaagccattc	atgtgaaccc	1560
tctttatatt	ctgatacctt	ctactctgtg	tacttcattt	gcattcctcc	taccagtagc	1620
aaatccaccc	aatgctattg	tcttttcata	tgggtcatctg	aaagtcattg	acatgggttaa	1680
agctggactt	ggtgtcaaca	ttggttggtgt	tgctgtgggt	atgcttggca	tatgtacttg	1740
gattgtaccc	atgtttgacc	tctacactta	cccttcgttg	gctcctgcta	tgagtaatga	1800
gaccatgcca	taataagcac	aaaatttctg	actatcttgc	ggtaatttct	ggaagacatt	1860
aatgattgac	tgtaaaatgt	ggctctaaat	aactaatgac	acacatttaa	atcagttatg	1920
gtgtagctgc	tgcaattccc	gtgaataccc	gaaacctgct	ggataaactc	agagtcacata	1980
tttgttattg	cagtgcact	aaagagcatc	tatgtgcctt	catcaagaag	cccatgtttt	2040
gagattttgc	tcatagaacca	tctgcaactt	gcttcacat	aagaataatt	tataacttga	2100
ccttcaaaga	gattagagca	tttgtttcat	cttacagttg	gagttcaatg	taacatttta	2160
aatgcaattt	attatttcag	aaatttccca	tgaaactaaa	aatagaaaat	aagatataca	2220
agttaattcg	gtacttggat	aaatcatttc	tgcattgttg	ttccagagaa	tttgctgaga	2280
aatcaaagcc	atgggtcatct	ggtgatgaag	agaaaagggt	aatctaaatg	atatgtgcat	2340
ttcctcattt	aaaaaatcca	attggattat	tcttaataata	tacatgtaat	atgaaaattg	2400
agattgaagc	actaattcca	aaattatggc	tgaatatact	aaataacaga	aaagttacag	2460
ataagaattt	atttctactg	aactctatag	ttagtgtaat	ataattcata	tttttatgat	2520
attggcacac	tgagaaattc	attttgtaga	gctatggata	aggcttgcta	tgatttgcac	2580
tattagtaca	gtatagttag	aaaggaaagc	tgaacactat	aaaactatta	acatattttc	2640
gtatatgagt	aacaactttg	cttaagtgtt	tatcttagtt	cagaaataca	taatgtcata	2700
tgttaaaaat	aaagagatgt	agaaatctaa	atgaattatc	actgtgtata	cagacagaaa	2760
aatcacataa	ctctgggtgtg	ttaacattgc	aatgaaaaaa	tgaaaaaaag	aaggaaaaaa	2820
gaataagaat	gaaaactgct	gacgtattac	aaaacagaaa	aataaatgat	ttaaaatcaa	2880
atcaaaaaga	aaaaaactaa	acattttaaac	aaaaatggga	taagaatagt	cttctagaag	2940
tgaggatgcg	taaaagaatg	agtttccaat	taccctgatg	tgacaattac	acattgtaga	3000
caggtagcaa	aatatcacat	acacccccaa	aatatgtaca	aatattatat	atcaataaat	3060
aaatttttaa	agagtaagtg	ctattggcat	tccaaaattc	agctaaagga	aaaatgatca	3120
aaaacaaagt	aaggtgcaca	gttagcaaaa	gatgcagatg	ttatatcaca	gcaattctca	3180
tgctaaaaat	acaacaaaag	acaaagcaaa	aaataaacct	ttgctttttt	tttttttttt	3240
tttttttttt	gagacggagt	ctcgtctctgt	cgcccaggct	ggagtgcagt	ggcgggatct	3300
cggctcactg	caagctccgc	ctcccagggt	cacgccattc	tcctgcctca	gccaaacctt	3360
tgctattttt	aatcttcgtt	ggcactttcc	agctgttact	gaccttgcta	ttttttgttc	3420
aaataagatt	atttacaaac	ttattcttga	aactaaatat	agtaaagagg	gttttttaaaa	3480
taatatttaa	catacgaatt	attaattggc	catgttcatt	atttatctat	gtttattaat	3540
gggccaatgc	aaaaaatcat	tttttcaaag	aaaaatttgt	ccatgtaaag	cttaaattat	3600
aatattgctg	ctttgtataa	ctcttctatg	tttattctat	tcatttggtc	ctttccctac	3660
catattttac	acatgtattt	ataatctgta	gtatttatta	catttctgct	tttttctagt	3720
cattcaattt	atcactgctg	aattgcatca	gatcatggat	gcatttttat	tatgaaaaaa	3780
taaaatgact	tttcaaatta	aaaaaaaaaa	aaaa			3814

<210> 42
 <211> 734
 <212> DNA
 <213> Homo sapiens

<400> 42						
caggacaatg	aaattcttca	gttacattct	ggtttatcgc	cgattttctct	tcgtgggtttt	60
cactgtgttg	gttttactac	ctctgcccat	cgctcctccac	accaaggaag	cagaatgtgc	120
ctacacactc	tttgtggteg	ccacattttg	gctcacagaa	gcattgcctc	tgtcggtaac	180
agctttgcta	cctagttaa	tgttacccat	gtttgggac	atgccttcta	agaagggtggc	240
atctgcttat	ttcaaggatt	ttcacttact	gctaattgga	gttatctgtt	tagcaacatc	300
catagaaaaa	tggaatttgc	acaagagaat	tgctctgaaa	atggtgatga	tggttggtgt	360

```

aaatcctgca tggctgacgc tgggggttcat gagcagcact gccttttttgt ctatgtggct 420
cagcaacacc tcgacggctg ccatggtgat gccattgcg gaggtgtag tgcagcagat 480
catcaatgca gaagcagagg tcgagggcac tcagatgact tacttcaacg gatcaaccaa 540
ccacggacta gaaattgatg aaagtgttaa tggacatgaa ataaatgaga ggaaagagaa 600
aacaaaacca gttccaggat acaataatga tacagggaaa atttcaagca aggtggagtt 660
ggaaaagact gtttaactac tgaaatgaag ctattctcct gactaaacat aactgaaaaa 720
ccattcatta aatg 734

```

```

<210> 43
<211> 539
<212> DNA
<213> Homo sapiens

```

```

<400> 43
gccactcaga tgacttactt caacggatca accaaccacg gactagaaat tgatgaaagt 60
gttaatggac atgaaataaa tgagaggaaa gagaaaacaa aaccagttcc aggatacaat 120
aatgatacag ggaaaatttc aagcaagggtg gagttggaaa agcactggaa acttgcagtt 180
caagatggct ccccatctcc ctctgtccat tctgtatcgc agctagctgc tcaaggaaa 240
gagaaagtgg aaggcatatg tacttagaaa ttattctatt actttcctgg atttaagagt 300
attcagattt tctatttcaa catcaaacaa ttgcattttt aaaaagaaat ttatgtgttc 360
catgtcaaat ttagtagtgt gtggttgttt ataatatattt cttatatcta ctttaatttct 420
atagtattta tagttatatg tctttatttc taacattttt cttgtgcttt taaagattat 480
ttaaagatta tttttaaata atctttattt catttaaata aaatatttta tttaagtct 539

```

```

<210> 44
<211> 556
<212> DNA
<213> Homo sapiens

```

```

<400> 44
cacggactag aaattgatga aagtgttaat ggacatgaaa taaatgagag gaaagagaaa 60
acaaaaccag ttccaggata caataatgat acagggaaaa tttcaagcaa ggtggagttg 120
gaaaagaact caggcatgag aaccaaatat cgaacaaaga agggccacgt gacacgtaaa 180
cttacgtgtt tgtgcattgc ctactcttct accattgggtg gactgacaac aatcactggt 240
acctccacca acttgatctt tgcagagtat ttcaatacat tccatccaca cagaagagga 300
gatcgtacaa ggcatgtaca ccaggaggca gaaatttgag gcatactctt gaactctgtc 360
taccacatcc tgaacatcac acagtttcca ctcttgttgc cttcaatcct gagaatgcat 420
ccaggagcca ttctgtttta tgtcaattac taattagatc atgtcacgtt actaacttac 480
tacgttccaa ttagtcctta ttgcatttgt aataaaatcc gcatactttc ggactggcta 540
caaggttata catgat 556

```

```

<210> 45
<211> 595
<212> PRT
<213> Homo sapiens

```

```

<400> 45
Met Lys Phe Phe Ser Tyr Ile Leu Val Tyr Arg Arg Phe Leu Phe Val
1          5          10          15
Val Phe Thr Val Leu Val Leu Leu Pro Leu Pro Ile Val Leu His Thr
20          25          30
Lys Glu Ala Glu Cys Ala Tyr Thr Leu Phe Val Val Ala Thr Phe Trp
35          40          45

```

Leu	Thr	Glu	Ala	Leu	Pro	Leu	Ser	Val	Thr	Ala	Leu	Leu	Pro	Ser	Leu
50						55					60				
Met	Leu	Pro	Met	Phe	Gly	Ile	Met	Pro	Ser	Lys	Lys	Val	Ala	Ser	Ala
65					70					75					80
Tyr	Phe	Lys	Asp	Phe	His	Leu	Leu	Leu	Ile	Gly	Val	Ile	Cys	Leu	Ala
				85					90					95	
Thr	Ser	Ile	Glu	Lys	Trp	Asn	Leu	His	Lys	Arg	Ile	Ala	Leu	Lys	Met
			100					105					110		
Val	Met	Met	Val	Gly	Val	Asn	Pro	Ala	Trp	Leu	Thr	Leu	Gly	Phe	Met
		115					120					125			
Ser	Ser	Thr	Ala	Phe	Leu	Ser	Met	Trp	Leu	Ser	Asn	Thr	Ser	Thr	Ala
130						135					140				
Ala	Met	Val	Met	Pro	Ile	Ala	Glu	Ala	Val	Val	Gln	Gln	Ile	Ile	Asn
145					150					155					160
Ala	Glu	Ala	Glu	Val	Glu	Ala	Thr	Gln	Met	Thr	Tyr	Phe	Asn	Gly	Ser
				165					170					175	
Thr	Asn	His	Gly	Leu	Glu	Ile	Asp	Glu	Ser	Val	Asn	Gly	His	Glu	Ile
			180					185					190		
Asn	Glu	Arg	Lys	Glu	Lys	Thr	Lys	Pro	Val	Pro	Gly	Tyr	Asn	Asn	Asp
		195					200					205			
Thr	Gly	Lys	Ile	Ser	Ser	Lys	Val	Glu	Leu	Glu	Lys	Asn	Ser	Gly	Met
210						215					220				
Arg	Thr	Lys	Tyr	Arg	Thr	Lys	Lys	Gly	His	Val	Thr	Arg	Lys	Leu	Thr
225					230					235					240
Cys	Leu	Cys	Ile	Ala	Tyr	Ser	Ser	Thr	Ile	Gly	Gly	Leu	Thr	Thr	Ile
				245					250					255	
Thr	Gly	Thr	Ser	Thr	Asn	Leu	Ile	Phe	Ala	Glu	Tyr	Phe	Asn	Thr	Arg
			260					265					270		
Tyr	Pro	Asp	Cys	Arg	Cys	Leu	Asn	Phe	Gly	Ser	Trp	Phe	Thr	Phe	Ser
		275					280					285			
Phe	Pro	Ala	Ala	Leu	Ile	Ile	Leu	Leu	Leu	Ser	Trp	Ile	Trp	Leu	Gln
290						295					300				
Trp	Leu	Phe	Leu	Gly	Phe	Asn	Phe	Lys	Glu	Met	Phe	Lys	Cys	Gly	Lys
305					310					315					320
Thr	Lys	Thr	Val	Gln	Gln	Lys	Ala	Cys	Ala	Glu	Val	Ile	Lys	Gln	Glu
				325					330					335	
Tyr	Gln	Lys	Leu	Gly	Pro	Ile	Arg	Tyr	Gln	Glu	Ile	Val	Thr	Leu	Val
			340					345					350		
Leu	Phe	Ile	Ile	Met	Ala	Leu	Leu	Trp	Phe	Ser	Arg	Asp	Pro	Gly	Phe
		355					360					365			
Val	Pro	Gly	Trp	Ser	Ala	Leu	Phe	Ser	Glu	Tyr	Pro	Gly	Phe	Ala	Thr
		370				375					380				
Asp	Ser	Thr	Val	Ala	Leu	Leu	Ile	Gly	Leu	Leu	Phe	Phe	Leu	Ile	Pro
385					390					395					400
Ala	Lys	Thr	Leu	Thr	Lys	Thr	Thr	Pro	Thr	Gly	Glu	Ile	Val	Ala	Phe
				405					410					415	
Asp	Tyr	Ser	Pro	Leu	Ile	Thr	Trp	Lys	Glu	Phe	Gln	Ser	Phe	Met	Pro
			420					425					430		
Trp	Asp	Ile	Ala	Ile	Leu	Val	Gly	Gly	Gly	Phe	Ala	Leu	Ala	Asp	Gly
		435					440					445			
Cys	Glu	Glu	Ser	Gly	Leu	Ser	Lys	Trp	Ile	Gly	Asn	Lys	Leu	Ser	Pro
450						455					460				
Leu	Gly	Ser	Leu	Pro	Ala	Trp	Leu	Ile	Ile	Leu	Ile	Ser	Ser	Leu	Met
465					470					475					480
Val	Thr	Ser	Leu	Thr	Glu	Val	Ala	Ser	Asn	Pro	Ala	Thr	Ile	Thr	Leu
				485					490					495	

Phe Leu Pro Ile Leu Ser Pro Leu Ala Glu Ala Ile His Val Asn Pro
 500 505 510
 Leu Tyr Ile Leu Ile Pro Ser Thr Leu Cys Thr Ser Phe Ala Phe Leu
 515 520 525
 Leu Pro Val Ala Asn Pro Pro Asn Ala Ile Val Phe Ser Tyr Gly His
 530 535 540
 Leu Lys Val Ile Asp Met Val Lys Ala Gly Leu Gly Val Asn Ile Val
 545 550 555 560
 Gly Val Ala Val Val Met Leu Gly Ile Cys Thr Trp Ile Val Pro Met
 565 570 575
 Phe Asp Leu Tyr Thr Tyr Pro Ser Trp Ala Pro Ala Met Ser Asn Glu
 580 585 590
 Thr Met Pro
 595

<210> 46
 <211> 224
 <212> PRT
 <213> Homo sapiens

<400> 46
 Arg Thr Met Lys Phe Phe Ser Tyr Ile Leu Val Tyr Arg Arg Phe Leu
 1 5 10 15
 Phe Val Val Phe Thr Val Leu Val Leu Leu Pro Leu Pro Ile Val Leu
 20 25 30
 His Thr Lys Glu Ala Glu Cys Ala Tyr Thr Leu Phe Val Val Ala Thr
 35 40 45
 Phe Trp Leu Thr Glu Ala Leu Pro Leu Ser Val Thr Ala Leu Leu Pro
 50 55 60
 Ser Leu Met Leu Pro Met Phe Gly Ile Met Pro Ser Lys Lys Val Ala
 65 70 75 80
 Ser Ala Tyr Phe Lys Asp Phe His Leu Leu Leu Ile Gly Val Ile Cys
 85 90 95
 Leu Ala Thr Ser Ile Glu Lys Trp Asn Leu His Lys Arg Ile Ala Leu
 100 105 110
 Lys Met Val Met Met Val Gly Val Asn Pro Ala Trp Leu Thr Leu Gly
 115 120 125
 Phe Met Ser Ser Thr Ala Phe Leu Ser Met Trp Leu Ser Asn Thr Ser
 130 135 140
 Thr Ala Ala Met Val Met Pro Ile Ala Glu Ala Val Val Gln Gln Ile
 145 150 155 160
 Ile Asn Ala Glu Ala Glu Val Glu Ala Thr Gln Met Thr Tyr Phe Asn
 165 170 175
 Gly Ser Thr Asn His Gly Leu Glu Ile Asp Glu Ser Val Asn Gly His
 180 185 190
 Glu Ile Asn Glu Arg Lys Glu Lys Thr Lys Pro Val Pro Gly Tyr Asn
 195 200 205
 Asn Asp Thr Gly Lys Ile Ser Ser Lys Val Glu Leu Glu Lys Thr Val
 210 215 220

<210> 47
 <211> 88

<212> PRT
 <213> Homo sapiens

<400> 47
 Ala Thr Gln Met Thr Tyr Phe Asn Gly Ser Thr Asn His Gly Leu Glu
 1 5 10 15
 Ile Asp Glu Ser Val Asn Gly His Glu Ile Asn Glu Arg Lys Glu Lys
 20 25 30
 Thr Lys Pro Val Pro Gly Tyr Asn Asn Asp Thr Gly Lys Ile Ser Ser
 35 40 45
 Lys Val Glu Leu Glu Lys His Trp Lys Leu Ala Val Gln Asp Gly Ser
 50 55 60
 Pro Ser Pro Ser Val His Ser Val Ser Gln Leu Ala Ala Gln Gly Lys
 65 70 75 80
 Glu Lys Val Glu Gly Ile Cys Thr
 85

<210> 48
 <211> 112
 <212> PRT
 <213> Homo sapiens

<400> 48
 His Gly Leu Glu Ile Asp Glu Ser Val Asn Gly His Glu Ile Asn Glu
 1 5 10 15
 Arg Lys Glu Lys Thr Lys Pro Val Pro Gly Tyr Asn Asn Asp Thr Gly
 20 25 30
 Lys Ile Ser Ser Lys Val Glu Leu Glu Lys Asn Ser Gly Met Arg Thr
 35 40 45
 Lys Tyr Arg Thr Lys Lys Gly His Val Thr Arg Lys Leu Thr Cys Leu
 50 55 60
 Cys Ile Ala Tyr Ser Ser Thr Ile Gly Gly Leu Thr Thr Ile Thr Gly
 65 70 75 80
 Thr Ser Thr Asn Leu Ile Phe Ala Glu Tyr Phe Asn Thr Phe His Pro
 85 90 95
 His Arg Arg Gly Asp Arg Thr Arg His Val His Gln Glu Ala Glu Ile
 100 105 110

<210> 49
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 49
 ccagctttaa ccatgtcaat g

21

<210> 50
 <211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of the artificial sequence: Oligonucleotide

<400> 50

cagatggttg tgaggagtct g

21

<210> 51

<211> 3311

<212> DNA

<213> Homo sapiens

<400> 51

tgctaattgct	tttggtacaa	atggatgtgg	aatataattg	aatatTTTTct	tgtttaaggg	60
gagcatgaag	aggtgttgag	gttatgtcaa	gcatctggca	cagctgaagg	cagatggaaa	120
tatttacaag	tacgcaattt	gagactaaga	tattgttata	attctcctat	tgaagacaag	180
agcaatagta	aaacacatca	ggtcaggggg	ttaaagacct	gtgataaacc	acttccgata	240
agttggaaac	gtgtgtctat	attttcatat	ctgtatatat	ataatggtaa	agaaagacac	300
cttcgtaacc	cgcattttcc	aaagagagga	atcacaggga	gatgtacagc	aatggggcca	360
tttaagagtt	ctgtgttcat	cttgattctt	caccttctag	aaggggccct	gagtaattca	420
ctcattcagc	tgaacaacaa	tggctatgaa	ggcattgtcg	ttgcaatcga	ccccaatgtg	480
ccagaagatg	aaacactcat	tcaacaaata	aaggacatgg	tgaccagggc	atctctgtat	540
ctgtttgaag	ctacaggaaa	gcgattttat	ttcaaaaatg	ttgccatttt	gattcctgaa	600
acatggaaga	caaaggctga	ctatgtgaga	ccaaaacttg	agacctacaa	aatgctgat	660
gttctgggtg	ctgagtctac	tcctccaggt	aatgatgaac	cctacactga	gcagatgggc	720
aactgtggag	agaagggtga	aaggatccac	ctcactcctg	atttcattgc	aggaaaaaag	780
ttagctgaat	atggaccaca	aggtaaggca	tttgtccatg	agtgggctca	tctacgatgg	840
ggagtatttg	acgagtacaa	taatgatgag	aaattctact	tatccaatgg	aagaatacaa	900
gcagtaagat	gttcagcagg	tattactggt	acaaatgtag	taaagaagtg	tcagggagggc	960
agctgttaca	ccaaaagatg	cacattcaat	aaagttacag	gactctatga	aaaaggatgt	1020
gagtttggtc	tccaatcccg	ccagacggag	aaggcttcta	taatgtttgc	acaacatggt	1080
gattctatag	ttgaattctg	tacagaacaa	aaccacaaca	aagaagctcc	aaacaagcaa	1140
aatcaaaaat	gcaatctccg	aagcacatgg	gaagtgatec	gtgattctga	ggactttaag	1200
aaaaccactc	ctatgacaac	acagccacca	aatcccacct	tctcattgct	gcagattgga	1260
caaagaattg	tgtgttttagt	ccttgacaaa	tctggaagca	tggcgactgg	taaccgcctc	1320
aatcgactga	atcaagcagg	ccagcttttc	ctgctgcaga	cagttgagct	ggggtcctgg	1380
gttgggatgg	tgacatttga	cagtgtctgc	catgtacaaa	gtgaactcat	acagataaac	1440
agtggcagtg	acagggacac	actcgccaaa	agattacctg	cagcagcttc	aggaggggacg	1500
tccatctgca	gcgggcttcg	atcggcattt	actgtgatta	ggaagaaata	tccaactgat	1560
ggatctgaaa	ttgtgctgct	gacggatggg	gaagacaaca	ctataagtgg	gtgctttaac	1620
gaggtcaaac	aaagtgggtg	catcatccac	acagtcgctt	tggggccctc	tgcagctcaa	1680
gaactagagg	agctgtccaa	aatgacagga	ggtttacaga	catatgcttc	agatcaagtt	1740
cagaacaatg	gcctcattga	tgtttttggg	gccctttcat	caggaaatgg	agctgtctct	1800
cagcgtcca	tccagcttga	gagtaaggga	ttaaccctcc	agaacagcca	gtggatgaat	1860
ggcacagtga	tcgtggacag	caccgtggga	aaggacactt	tgtttcttat	cacctggaca	1920
acgcagcctc	cccaaatect	tctctgggat	cccagtggac	agaagcaagg	tggctttgta	1980
gtggacaaaa	acaccaaaaat	ggcctaccte	caaateccag	gcattgctaa	ggttggcact	2040
tggaaataca	gtctgcaagc	aagctcacia	accttgacce	tgactgtcac	gtcccgtgcg	2100
tccaatgcta	ccctgcctcc	aattacagtg	acttccaaaa	cgaacaagga	caccagcaaa	2160
ttccccagcc	ctctggtagt	ttatgcaaat	attcgccaag	gagcctcccc	aattctcagg	2220
gccagtgtca	cagccctgat	tgaatcagtg	aatggaaaaa	cagttacctt	ggaactactg	2280
gataatggag	caggtgctga	tgtactaag	gatgacggtg	tctactcaag	gtatttcaca	2340
acttatgaca	cgaatggtag	atacagtgtg	aaagtgcggg	ctctgggagg	agttaacgca	2400

gccagacgga	gagtataacc	ccagcagagt	ggagcactgt	acatacctgg	ctggattgag	2460
aatgatgaaa	tacaatggaa	tecaccaaga	cctgaaatta	ataaggatga	tgttcaacac	2520
aagcaagtgt	gtttcagcag	aacatcctcg	ggaggctcat	ttgtggcttc	tgatgtccca	2580
aatgctccca	tacctgatct	cttcccacct	ggccaaatca	ccgacctgaa	ggcggaaatt	2640
cacgggggca	gtctcattaa	tctgacttgg	acagctcctg	gggatgatta	tgacctatga	2700
acagctcaca	agtatatcat	tcgaataagt	acaagtattc	ttgatctcag	agacaagttc	2760
aatgaatctc	ttcaagtga	tactactgct	ctcatcccaa	aggaagccaa	ctctgaggaa	2820
gtctttttgt	ttaaaccaga	aaacattact	tttgaaaatg	gcacagatct	tttcattgct	2880
attcaggctg	ttgataaggt	cgatctgaaa	tcagaaatat	ccaacattgc	acgagtatct	2940
ttgttttattc	ctccacagac	tcggccagag	acacctagtc	ctgatgaaac	gtctgctcct	3000
tgctctaata	ttcatatcaa	cagcaccatt	cctggcattc	acatttttaa	aattatgtgg	3060
aagtggatag	gagaactgca	gctgtcaata	gcctagggct	gaatttttgt	cagataaata	3120
aaataaatca	ttcatccttt	ttttgattat	aaaattttct	aaaatgtatt	ttagacttcc	3180
tgtagggggc	gatatactaa	atgtatatag	tacatttata	ctaaatgtat	tcctgtaggg	3240
ggcgatatatac	taaatgtatt	ttagacttcc	tgtagggggc	gataaaataa	aatgctaaac	3300
aactgggtaa	a					3311

<210> 52
 <211> 3067
 <212> DNA
 <213> Homo sapiens

<400> 52						
aattaaatta	tgagaattaa	aaagacaaca	ttgagcagag	atgaaaaagg	aagggaggaa	60
aaggtggaaa	agaaaagaag	acaagaagcg	agtagtggtc	tctaacttgc	tctttgaagg	120
atggtctcac	aaagagaacc	ccaacagaca	tcatcgtggg	aatcaaatac	agaccagcaa	180
gtacaccgtg	ttgtccttcg	tccccaaaaa	catttttgag	cagctacacc	ggtttgccaa	240
tctctatttt	gtgggcattg	cggttctgaa	ttttatccct	gtggteaagt	ctttccagcc	300
tgaggtgagc	atgataccaa	tctgtgttat	cctggcagtc	actgccatca	aggacgcttg	360
ggaagacctc	cggaggtaca	aatcggataa	agtcacatca	aaccgagagt	gcctcatcta	420
cagcagaaaa	gagcagacct	atgtgcagaa	gtgctggaag	gatgtgcgtg	tgggagactt	480
catccaaatg	aaatgcaatg	agattgtccc	agcagacata	ctcctccttt	tttcctctga	540
ccccaatggg	atatgccatc	tggaaaactgc	cagcttggtat	ggagagacaa	acctcaagca	600
aagacgtgtc	gtgaagggct	tctcacagca	ggaggtacag	ttcgaaccag	agcttttcca	660
caataccatc	gtgtgtgaga	aacccaacaa	ccacctcaac	aaattttaagg	gttatatgga	720
gcattcctgac	cagaccagga	ctggcttttg	ctgtgagagt	cttctgcttc	gaggctgcac	780
catcagaaac	accgagatgg	ctggttggcat	tgtcatctat	gcaggccatg	agacgaaagc	840
catgctgaac	aacagtggcc	cccggataca	acgcagcaag	attgagcggc	gcattgaatat	900
agacatcttc	ttctgcattg	ggatcctcat	cctcatgtgc	cttattggag	ctgtaggtca	960
cagcatctgg	aatgggacct	ttgaagaaca	ccctcccttc	gatgtgccag	atgccaatgg	1020
cagcttcctt	cccagtggcc	ttgggggctt	ctacatgttc	ctcacaatga	tcattcctgct	1080
ccaggtgctg	atccccatct	ctttgtatgt	ctccattgag	ctgggtgaagc	tcgggcaagt	1140
gttcttcttg	agcaatgacc	ttgacctgta	tgatgaagag	accgatttat	ccattcaatg	1200
tcgagccctc	aacatcgcag	aggacttggg	ccagatccag	tacatcttct	ccgataagac	1260
ggggaccctg	acagagaaca	agatgggtgt	ccgacgttgc	accatcatgg	gcagcgagta	1320
ttctcaccaa	gaaaatggta	tagaagctcc	caagggctcc	atccctcttt	ctaaaaggaa	1380
ataccctgct	ctcctaagaa	acgaggagat	aaaagacatt	ctcctggctc	tcttagaggc	1440
tgtgtggcat	ttccacaagt	tgcttctctg	atccctgtgg	tcttctctgt	cacagatcag	1500
ggctgttcca	attacttgta	aactttcatt	tgtttacaaa	ggttagaagt	tatcccatat	1560
gtggttcccc	ttcagctgat	ctttgtctgg	tgccagacaa	agcactttat	gagacgagtt	1620
ttttatctgt	cagcaatgga	ttggagacat	ttcccaattg	tgtgccagtc	acacaaccaa	1680
ggcttaggaa	tttctcaggc	caccttacct	gacatgtcag	ggcaggtctg	tgtctaggtg	1740
catggtcaga	tttaatacat	ccagaagatg	tcttctattc	taacagatct	cttagcttgt	1800
cactgaggca	aagttttgat	ttaggagata	gggctataaa	atgcctggac	tgttaccttg	1860
catggactga	atatgactca	taaaactgat	ctgattcctt	cagccatcat	ctgcccact	1920

tggttcccct	ccccaccccc	ccacaacaca	cacacacact	ttctaagaaa	agaaaagaaa	1980
ttcttttttt	tcaatacttt	aagttctggg	atacatgtgc	agaatgtgca	ggtttggttac	2040
ataggtatac	atgtgtcatg	gtggtttgca	gcacccacca	acccatcatc	taccttaggt	2100
atttctccta	atgctatccc	tcccctagcc	cccaaccccc	cgatgggctc	cagtgtgtga	2160
tgttcccctc	catgtccatg	tgttctcatt	gttcaattcc	cacttatgag	tgagaacatg	2220
cagtatttgg	ttttctgttc	ttgtgttagt	ttgctgatgg	tttcctgttc	atccgtgtcc	2280
ctgcaaagga	catgaactca	tcctttttta	tggctgcata	atattccatg	gtgtatatgt	2340
gccacatttt	ctttatccag	tctatcgctg	atgggcactg	gggttggttc	caagtctttg	2400
ctattgtgaa	cagtgtctgca	ataaacttac	atgtgcatgt	gtcttttagta	gaatgattta	2460
taatcctttg	ggatatatac	cagtaatggg	attgctggtc	aaatgggtatt	tctggttcta	2520
gaccttgag	gaatctttgt	cttcacacaat	ggttgaacta	atttgtactc	ccaccaacag	2580
tgtaaaagta	ttcctgtttc	tctacatcct	cttcagcatc	tgttgtgtcc	tgacatttta	2640
atgatcacta	ttctcactgg	cgtgagatgt	tatctcattg	tggttttgat	ttgcatttct	2700
ctaatgacca	gtaatgatga	gctttttttc	atatgtttgt	tggctgcata	aatgtcttct	2760
tttgagaagt	gtctgttcat	atccttcacc	cattttttga	agaaaacaaa	ctcttaagag	2820
agcagtattc	attcttttga	gtgtgagggg	tggagaaaag	gaaagatgga	gagagtatta	2880
taagcagctg	tatccccctt	gccatggtga	tagcagacca	ttcacatggg	agcttctggt	2940
ctctttgtaa	taataataag	agccacatta	ccagtactta	gagtatgcta	gttattttta	3000
cacattgtat	cattaaatct	tcaaaacatc	cctatgagtt	agaaacctaa	aaaaaaaaaa	3060
aaaaaaa						3067

<210> 53
 <211> 2778
 <212> DNA
 <213> Homo sapiens

<400> 53						
ctcatttttga	tgtctagaat	caggggatcc	aggatcatca	ccaaggteat	tttcccaggt	60
atggaggggt	ctttctgctt	ctttcttgct	atgcacagct	gctgaggaag	gggctgggag	120
taaagacagt	gaaatgggga	ggaggagtc	attcaaaccg	agaaacaaag	tgtttggttt	180
ttcttaccce	tgggtgtagaa	gctaccaacc	ttttccaaga	aagagggcct	ggcccccttc	240
tcgggtctgg	ctgggtgcct	gctgtgcctc	tctggcctcc	cctccgaagg	gcaccattcc	300
ctcgggtgag	tactaccggc	ctgcaccgct	ttccagtggg	gacagcctga	gaagagagtc	360
tggggcctta	cttcagtacc	ttccttcact	ggcctcaccc	tgtgcaaate	atgccacacg	420
ctgcagcctc	cttttcccta	tctataaaat	aaaaatgacc	ctgctctatc	tactgggct	480
ggcaagaaca	cactgttggt	gccttgacga	cagatgtgct	gaggctgtag	aaagtgcctt	540
ttatttggtt	gggagcttgt	gcataaatgc	gagaggggct	gcacatctga	cggactagag	600
gtgactcatg	gctgaaccgg	aacaggacat	cggggagaag	ccagcagcca	tgctgaactc	660
tccacagggc	cctgtgaaaa	gctcttcacc	tcctctgccc	tctggatcta	gtgaagccta	720
ttcatccttc	agatgtcagc	tcaaataate	aaccttcatt	gaggcctccc	ttgaccctta	780
acatgctttc	aaagtactgt	gtatttcaca	ttcatcatgc	cccagacaact	gtgatttccc	840
atttattaat	atctgtctct	tctgctggcc	tgcaaaactc	aggagcacag	agacatcttt	900
gggatttttg	aacatgatatt	ccccagggct	tagcccagtg	cctgggtgcaa	agcaggcttt	960
caacatgttc	agtggatatt	gtaagaaaga	aagaaatata	caaaaggcct	ggcatatgca	1020
aagcactcta	aatattcact	cctttccctt	ccctctgggt	gagaaaattt	ctccttataa	1080
agacaccctc	ctaactgtat	ctctgctaga	gaactgaaga	cataaagcac	tctgtgccaa	1140
aaatatttta	gtaaaaactt	gagctaagca	cagagattat	aaatatttct	tccccagatt	1200
acgcaccatt	taaaaatact	gtctcagctc	cttttcatga	tttgggtggg	gattaaagaa	1260
aattactctt	caagactgaa	agtcattact	gcccttttcc	tgacttgcc	tttcccttga	1320
gaaggggagg	ataagctgca	gggcaggaag	tggaaagtggg	gcatecttgt	cctttgtctg	1380
gcagacagcc	aactggtcag	gtactgctcc	ttctcaactc	tttctgatt	cccagggtgaa	1440
tataaacaag	aaggcacaaa	tccacacttg	ccaacaacgg	acccaagtga	taacaagaaa	1500
cccagtgaca	cctgtctagg	tgaagactca	gccctatgt	gaccagggtg	caaagccaaa	1560
ctgaccatct	gctttccatt	tggactttta	gttcatactg	tatcttctca	ggacagttaa	1620
gttggaatac	aatgccactg	tcctgaaaga	tggtagaatt	atcctatttc	tggaggagtg	1680

```

ggggtggtgg gtaggaatct caagagcgat ttgctcctct gcacaatagc ttctttaagg 1740
acaccagggc ccccagggct atacatttcc ctgaagcttt ccagataagc aacaaggat 1800
gagcacctgc tatgtattgc ccaaggggtga tgtgttttaa tatccattgc atattttaaa 1860
tccttggctg gcttaaagct gcaagctttc tgtcttcagt ggatataatg ggggcataca 1920
tcccagagct tgcccaacac tccaagaaaa gaacctcag ctaatgcaaa gtgtgtatgt 1980
gcccatgaaa gctccatgtc tacttaacat tcagttttta ggattattta tgctgtaata 2040
atagatatga aaatctctga caggatattt gtttccttta caaactgtat ttgaatttat 2100
gggtgattta gagcttgtgt ttaaagtcag aattcagaac cccaaagaaa atgacttcat 2160
tgaaattgaa ctgaagagac aagaactgag ttaccaaaac ctactaaacg tgagttgctg 2220
tgaactgggg attaaaccag aacgagtgga gaagatcaga aagctacca acacactgct 2280
cagaaaggac aaagacattc gaagactgcg ggactttcag gaagtggaa tcattttaat 2340
gaaaaatgga agctccagat tgacagaata tgtgccatct ctgacagaaa ggccctgcta 2400
tgatagcaaa gctgcaaaaa tgacttatta aatactccca ggaatggcgc cgcattggtg 2460
ctcacccctt gtaatcccag cactttggga agccaagggt ggccgatcac ctgaggtcag 2520
gagttctaga ccagcctggc caacatatag tgaaaccag tctctactaa aaaaaataca 2580
aaaattagct aggtgtggtg gcgcacacct gtagtagtcc cagctacatg ggaagctgag 2640
gcaggagaat cacctgaacc caggaggcag aggttgagc gagctgagat tgcgccactg 2700
cactccagcc tggcgacaga gcaagactct gtctctcaaa ataaataaat aaataaataa 2760
ataaataaat aaataatc
2778

```

```

<210> 54
<211> 1646
<212> DNA
<213> Homo sapiens

```

```

<400> 54
gcccgggaga ggagaggagc gggccgagga ctccagcgtg cccaggctctg gcacccctgca 60
cttgctgccc tctgacacct gggaagatgg ccggcccgctg gaccttcacc cttctctgtg 120
gtttgctggc agccaccttg atccaagcca cctcagtcct cactgcagtt ctcatcctcg 180
gcccaaaagt catcaaagaa aagctgacac aggagctgaa ggaccacaac gccaccagca 240
tcctgcagca gctgcccgtg ctgagtgcca tgcgggaaaa gccagccgga ggcateccctg 300
tgctgggcag cctgggtgaac accgtcctga agcacatcat ctggctgaag gtcacacag 360
ctaacatcct ccagctgcag gtgaagccct cggccaatga ccaggagctg ctagtcaaga 420
tccccctgga catgggtggt ggattcaaca cggccctggt caagaccatc gtggagttcc 480
acatgacgac tgaggcccaa gccaccatcc gcattggacac cagtgcaggt ggccccacce 540
gcctggctct cagtgaactg gccaccagcc atgggagcct gcgcattcaa ctgctgcata 600
agctctcctt cctgggtgaac gccttagcta agcaggctcat gaacctccta gtgccatccc 660
tgcccaatct agtgaaaaac cagctgtgtc ccgtgatcga ggcttccttc aatggcatgt 720
atgcagacct cctgcagctg gtgaagggtg ccatttcctt cagcattgac cgtctggagt 780
ttgaccttct gtatcctgcc atcaaggggtg acaccattca gctctacctg ggggccaagt 840
tgttggactc acagggaaag gtgaccaagt ggttcaataa ctctgcagct tcctgacaa 900
tgcccacctt ggacaacatc ccgttcagcc tcatcgtgag tcaggacgtg gtgaaagctg 960
cagtggctgc tgtgctctct ccagaagaat tcatggctct gttggactct gtgcttctctg 1020
agagtgccca tcggctgaag tcaagcatcg ggctgatcaa tgaaaaggct gcagataagc 1080
tgggatctac ccagatcgtg aagatcctaa ctccaggacac tcccagagttt tttatagacc 1140
aaggccatgc caaggtggcc caactgatcg tcttggaagt gtttcctctc agtgaagccc 1200
tccgcccttt gttcacctct ggcacgaaag ccagctcgga agctcagttt tacaccaaaag 1260
gtgaccaact tatactcaac ttgaataaca tcagctctga tcggatccag ctgatgaact 1320
ctgggattgg ctggttccaa cctgatgttc tgaaaaacat catcactgag atcatccact 1380
ccatcctgct gccgaaccag aatggcaaata taagatctgg ggtcccagtg tcattggtga 1440
aggccttggg attcgaggca gctgagtcct cactgaccaa ggatgccctt gtgcttactc 1500
cagcctcctt gtggaaaccc agctctcctg tctcccagtg aagacttgga tggcagccat 1560
cagggaaaggc tgggtcccag ctgggagtat ggggtgtgagc tctatagacc atccctctct 1620
gcaatcaata aacacttgcc tgtgat
1646

```

<210> 55
 <211> 1049
 <212> DNA
 <213> Homo sapiens

<400> 55
 ggagtggggg agagagagga gaccaggaca gctgctgaga cctctaagaa gtccagatac 60
 taagagcaaa gatgtttcaa actggggggc tcattgtctt ctacgggctg ttagcccaga 120
 ccatggccca gtttgagggc ctgcccgtgc ccctggacca gaccctgcc ttgaatgtga 180
 atccagccct gcccttgagt cccacaggtc ttgcaggaag cttgacaaat gccctcagca 240
 atggcctgct gtctgggggc ctgttgggca ttctggaaaa ccttccgctc ctggacatcc 300
 tgaagcctgg aggaggtact tctgggtggc tccttggggg actgcttgga aaagtgacgt 360
 cagtgattcc tggcctgaac aacatcattg acataaaggc cactgacccc cagctgctgg 420
 aacttggcct tgtgcagagc cctgatggcc accgtctcta tgtcaccatc cctctcggca 480
 taaagctcca agtgaatacg cccctggctg gtgcaagtct gttgaggctg gctgtgaagc 540
 tggacatcac tgcagaaatc ttagctgtga gagataagca ggagaggatc cacctgggcc 600
 ttggtgactg caccattcc cctggaagcc tgcaaatttc tctgcttgat ggacttggcc 660
 cctccccat tcaaggctct ctggacagcc tcacagggat cttgaataaa gtccctgcctg 720
 agttggttca gggcaacgtg tgcctctctg tcaatgaggt tctcagaggc ttggacatca 780
 ccctggtgca tgacattggt aacatgctga tccacggact acagtttgct atcaaggctc 840
 aagccttcca ggaaggggct ggctctctgt gagctgcttc ccagtgtca cagatggctg 900
 gcccatgtgc tggaagatga cacagttgcc ttctctccga ggaacctgcc cctctcctt 960
 tcccaccagg cgtgtgtaac atcccatgtg cctcacctaa taaaatggct cttcttctgc 1020
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1049

<210> 56
 <211> 4815
 <212> DNA
 <213> Homo sapiens

<400> 56
 gagcagagcc ctttcacaca cctcaggaa acctttcggc tgcccgcctc ccagacacac 60
 ctgcagccct gccagccgg ctttgctcac ccactgcttg taaatgcccc agatatgagc 120
 cagcccaggc cccgctacgt ggtagacaga gccgcatact cccttaccct cttcgacgat 180
 gagtttgaga agaaggaccg gacataccca gtgggagaga aacttcgcaa tgccttcaga 240
 tgttcctcag ccaagatcaa agctgtggtg tttgggctgc tgccctgtgt ctccctggctc 300
 cccaagtaca agattaaaga ctacatcatt cctgacctgc tcggtggact cagcggggga 360
 tccatccagg tcccacaagg catggcattt gctctgctgg ccaaccttc tgcagtcaat 420
 ggctctact cctccttctt cccctcctg acctacttct tcctgggggg tgttcaccag 480
 atggtgccag gtacctttgc cgttatcagc atcctgggtg gtaacatctg tctgcagctg 540
 gcccagagt cgaaattcca ggtcttcaac aatgccacca atgagagcta tgtggacaca 600
 gcagccatgg aggtgagag gctgcacgtg tcagctacgc tagcctgcct caccgccatc 660
 atccagatgg gtctgggctt catgcagttt ggctttgtgg ccactctacct ctccgagtec 720
 ttcattccgg gcttcatgac ggccgccggc ctgcagatcc tgatttcggt gctcaagtac 780
 atcttcggac tgaccatccc ctctacaca ggcccagggt ccactcgtct taccttcatt 840
 gacatttgca aaaacctccc ccacaccaac atcgctcgc tcactctcgc tctcatcagc 900
 ggtgccttcc tgggtgctgt gaaggagctc aatgctcgt acatgcacaa gattegcttc 960
 cccatcccta cagagatgat tgtggtggtg gtggcaacag ctatctccgg gggctgtaag 1020
 atgccccaaa agtatcacat gcagatcgtg ggagaaatcc aacgcggggt cccaccccg 1080
 gtgtcgctg tggcttcaca gtggaaggac atgataggca cagccttctc cctagccatc 1140
 gtgagctacg tcatcaacct ggctatgggc cggacctgg ccaacaagca cggctacgac 1200
 gtggattcga accaggagat gatcgctctc ggctgcagca acttcttttg ctcttcttt 1260
 aaaattcatg tcatttgctg tgcgctttct gtcactctgg ctgtggatgg agctggagga 1320
 aaatcccagg tggccagcct gtgtgtgtct ctggtggtga tgatcccat gctggctctg 1380
 gggatctatc tgtatcctct ccctaagtct gtgctaggag ccctgatcgc tgtcaatctc 1440

aagaactccc	tcaagcaact	caccgacccc	tactacctgt	ggaggaagag	caagctggac	1500
tgttgcatct	gggtagtgag	cttcctctcc	tccttcttcc	tcagcctgcc	ctatggtgtg	1560
gcagtgggtg	tcgccttctc	cgtcctggtc	gtggtcttcc	agactcagtt	tcgaaatggc	1620
tatgcaactg	cccaggtcat	ggacactgac	atztatgtga	atcccaagac	ctataatagg	1680
gcccaggata	tcaggggat	taaaatcatc	acgtactgct	cccctctcta	ctttgccaac	1740
tcagagatct	tcaggcaaaa	ggtcatcgcc	aagacaggca	tggacccccca	gaaagtatta	1800
ctagccaagc	aaaaatacct	caagaagcag	gagaagcgga	gaatgaggcc	cacacaacag	1860
aggaggtctc	tattcatgaa	aaccaagact	gtctccctgc	aggagctgca	gcaggacttt	1920
gagaatgcgc	ccccaccga	ccccacaac	aaccagaccc	cggctaacgg	caccagcgtg	1980
tcctatatca	ccttcagccc	tgacagctcc	tcacctgccc	agagtgagcc	accagcctcc	2040
gctgaggccc	cggcgagcc	cagtgcacatg	ctggccagcg	tcccaccctt	cgtcaccttc	2100
cacaccctca	tcctggacat	gagtggagtc	agcttcgtgg	acttgatggg	catcaaggcc	2160
ctggccaagc	tgagctccac	ctatgggaag	atcggcgtga	aggtcttctt	ggtgaacatc	2220
catgcccagg	tgtacaatga	cattagccat	ggaggcgtct	ttgaggatgg	gagtctagaa	2280
tgcaagcacg	tctttcccag	catacatgac	gcagtcctct	ttgcccaggc	aaatgctaga	2340
gacgtgaccc	caggacacaa	cttccaaggg	gctccagggg	atgctgagct	ctccttgtac	2400
gactcagagg	aggacattcg	cagctactgg	gacttagagc	aggagatggt	cgggagcatg	2460
tttcacgcag	agaccctgac	cgccctgtga	gggctcagcc	agtcctcatg	ctgcctacag	2520
agtgcctggc	acttgggact	tccataaagg	atgagcctgg	ggtcacaggg	ggtgtcgggc	2580
ggaggaaagt	gcatecccca	gagcttgggt	tcctctctcc	tctccccctc	tctcctccct	2640
tccttccctc	cccgcatctc	cagagagagc	ctctcagcag	caggggggtg	ctacccttac	2700
gggagtgaga	gtctggtgag	cccactcttc	acccgtcagg	ccctggccgc	aatggacaag	2760
cctcctgctc	actccacccc	acccacatct	gccctgtcct	tggcagctga	aggacacctt	2820
gacttccagc	ttttacgagt	gagccaaaaa	cagaaggaca	agtacaactg	tgctggcctg	2880
ctgtacaagc	ttcaaaaagt	gtcccagagc	ccgcacggct	cgggtgtcaga	tggtgtcagg	2940
ctgtcacgga	catagggata	aacttggtta	ggactctggc	ttgccttccc	cagctgcctc	3000
aactctgtct	ctggcagctc	tgcacccagg	gaccatgtgc	tctccacacc	caggagtcta	3060
ggccttggtg	actatgcgcc	ccccctccat	catccccaag	gctgcccaca	ccaccactgc	3120
tgtcagcaag	cacatcagac	tctagcctgg	acagtggcca	ggaccgtcga	gaccaccaga	3180
gctacctccc	cggggacagc	ccactaaggt	tctgcctcag	cctcctgaaa	catcactgcc	3240
ctcagaggct	gctcccttcc	cctggagggt	ggctagaaac	cccaaagagg	gggatgggta	3300
gctggcagaa	tcctctggca	tcctagtaat	agataaccagt	tattctgcac	aaaacttttg	3360
ggaattcctc	tttgacccca	gagactcaga	ggggaagagg	gtgctagtac	caacacaggg	3420
aaaacggatg	ggacctgggc	ccagacagtc	ccccttgacc	ccagggccca	tcagggaaat	3480
gcctcccttt	ggtaaatctg	ccttatecct	ctttacctgg	caaagagcca	atcatgttaa	3540
ctcttcctta	tcagcctgtg	gccagagac	acaatggggg	ccttctgtag	gcaaagggtg	3600
aagtcctcca	gggatccgct	acatcccccta	actgcacgca	gatgtggaaa	ggggctgac	3660
cagattgggt	cttcctgcac	aggaagactc	tttaacaccc	ttaggacctc	aggccatctt	3720
ctcctatgaa	gatgaaaata	gggggttaagt	tttccatatg	tacaaggagg	tattgagagg	3780
aaccctactg	ttgacttgaa	aataaatagg	ttccatgtgt	aagtgttttg	taaaatttca	3840
gtggaaatgc	acagaaaatc	ttctggcctc	tcactactgc	ttttctcaag	cttcttcagc	3900
ttaacaaccc	cttccctaac	aggttgggct	ggcccagcct	aggaaaacat	ccccatttct	3960
aacttcagcc	agacctgcgt	tgtgtgtctg	tgtgttgagt	gagctgggtca	gctaacaagt	4020
cttcttagag	ttaaaggagg	gggtgctggc	caagagccaa	cacattcttg	gcccaggagc	4080
attgcttttc	tgtgaattca	ttatgccatc	tggctgccaa	tggaaactcaa	aacttggaag	4140
gcgaaggaca	atgttatctg	ggattcaccc	tgccagcac	ccgaagtgcc	aaattccagg	4200
aggacaagag	ccttagccaa	tgacaactca	ctctcccccta	ctccacctcc	ttccaagtcc	4260
agctcaggcc	caggaggtgg	gagaaggtea	cagagcctca	ggaatttcca	agtcagagtc	4320
ccctttgaac	caagtatcta	gatcccctga	ggacttgatg	aagtgatcct	taacccccaa	4380
gtaatcatta	acccccagac	cagcctcaga	actgaaggag	attgttgacc	cagtgcactg	4440
gagttgaggc	tcagggagag	atctgccaca	tgtctgaggg	ttgcagagcc	cgctgtggag	4500
gtaagattgg	aaacacatga	ggcagaggga	agacattgaa	gaaaacatct	ctgctggaat	4560
atttggaaaa	gaacactctt	ctggacctgg	ttgaagcagg	aaagatggag	gcaaagtagt	4620
gaaataatcc	agaatttcaa	tgcttttgaa	tgttcttagt	gatactgacc	tgtgataata	4680
taattcccag	ggaggactgg	gaaccttata	tcttgagata	tttgcataat	ttatttaatt	4740

taagcctcat tctccttttg ttcatttttg taataaactg gatttgaatt gtgaacaaaa 4800
 aaaaaaaaaa aaaaa 4815

<210> 57
 <211> 2572
 <212> DNA
 <213> Homo sapiens

<400> 57
 aatgctctaa gacctctcag cacgggcgga agaaactccc ggagagctca cccaaaaaac 60
 aaggagatcc catctagatt tcttcttgct tttgactcac agctggaagt tagaaaagcc 120
 tcgatttcat ctttgagag gccaaatggg cttagcctca gtctctgtct ctaaataattc 180
 caccataaaa cagctgaggt atttatgaat tagaggctat agctcacatt ttcaatcctc 240
 tatttctttt tttaaataa acttttctact ctgatgagag aatgtgggtt taatctctct 300
 ctcacatttt gatgatttag acagactccc cctcttctc ctagtcaata aaccattga 360
 tgatctattt ccagcttat ccccaagaaa acttttgaaa ggaaagagta gacccaaaga 420
 tgttattttc tgctgtttga attttgtctc cccaccccca acttggtctag taataaacac 480
 ttactgaaga agaagcaata agagaaagat atttgtaate tctccagccc atgatctcgg 540
 ttttcttaca ctgtgatctt aaaagttacc aaaccaaaagt cattttcagt ttgaggcaac 600
 caaacctttc tactgctgtt gacatcttct tattacagca acaccattct aggagtttcc 660
 tgagctctcc actggagtc tctttctgtc gcgggtcaga aattgtccct agatgaatga 720
 gaaaattatt ttttttaatt taagtcctaa atatagttaa aataaataat gtttttagtaa 780
 aatgatacac tatctctgtg aaatagcctc acccctacat gtggatagaa ggaaatgaaa 840
 aaataattgc tttgacattg tctatatggg acttttgtaa gtcattgctta agtacaaatt 900
 ccatgaaaag ctactgatc ctaattcttt ccctttgagg tctctatggc tctgattgta 960
 catgatagta agtgtaagcc atgtaaaaag taaataatgt ctgggcacag tggctcacgc 1020
 ctgtaatcct agcactttgg gaggtgagg aggaaggatc acttgagccc agaagttcga 1080
 gactagcctg ggcaacatgg agaagccctg tctctacaaa atacagagag aaaaaatcag 1140
 ccagtcatgg tggcatacac ctgtagtccc agcattccgg gaggtgagg tgggaggatc 1200
 acttgagccc agggaggttg gggtgcagt gagccatgat cacaccactg cactccagcc 1260
 aggtgacata gcgagatcct gtctaaaaaa ataaaaaata aataatggaa cacagcaagt 1320
 cctaggaagt aggttaaaac taattcttta aaaaaaaaaa aaagttgagc ctgaattaaa 1380
 tgtaatgttt ccaagtgaca ggtatccaca tttgcatggg tacaagccac tgccagttgg 1440
 cagtagcact ttcctggcac tgtggtcggg tttgttttgt tttgctttgt ttagagacgg 1500
 ggtctcactt tccaggctgg cctcaaactc ctgcactcaa gcaattcttc taccctggcc 1560
 tcccaagtag ctggaattac aggtgtgcgc catcacaact agctgggtgg cagtttttgtt 1620
 actctgagag ctgttcactt ctctgaattc acctagagtg gttggaccat cagatgtttg 1680
 ggcaaaaactg aaagctcttt gcaaccacac accttccctg agcttacatc actgcccttt 1740
 tgagcagaaa gtctaaattc ctccaagac agtagaattc catcccagta ccaaagccag 1800
 ataggccccc taggaaactg aggttaagagc agtctctaaa aactaccac agcagcattg 1860
 gtgcagggga acttgcccat taggttatta tttgagagga aagtcctcac atcaatagta 1920
 catatgaaag tgacctcaa ggggattggg gaatactcat aaggatcttc aggctgaaca 1980
 gactatgtct ggggaaagaa cggattatgc ccattaaat aacaagttgt gttcaagagt 2040
 cagagcagtg agctcagagg cccttctcac tgagacagca acatttaaac caaaccagag 2100
 gaagtatttg tggaactcac tgcctcagtt tgggtaaagg atgagcagac aagtcaacta 2160
 aagaaaaaag aaaagcaagg aggaggggtg agcaatctag agcatggagt ttgttaagtg 2220
 ctctctggat ttgagttgaa gagcatccat ttgagttgaa ggccacaggg cacaatgagc 2280
 tctcccttct accaccagaa agtccctggg caggtctcag gtagtgcggt gtggctcagc 2340
 tgggttttta attagcgcac tctctatcca acatttaatt gtttgaaagc ctccatatag 2400
 ttagattgtg ctttgtaatt ttgttggtgt tgctctatct tattgtatat gcattgagta 2460
 ttaacctgaa tgttttgtta cttaaataat aaaaacactg ttatctaca aaaaaaccct 2520
 caaaggctga aaataaagaa ggaagatgga gacaccctct gggggctctc tc 2572

<210> 58
 <211> 1324
 <212> DNA
 <213> Homo sapiens

<400> 58
 ctttgcagtg gatgcccttg gcaggggtgag cccacaagga gcaatggagc agggcagcgg 60
 ccgcttggag gacttccctg tcaatgtgtt ctccgtcact cettacacac ccagcaccgc 120
 tgacatccag gtgtccgatg atgacaaggc gggggccacc ttgctcttct caggcatctt 180
 tctgggactg gtggggatca cattcactgt catgggctgg atcaaatacc aaggtgtctc 240
 ccactttgaa tggacccagc tccttgggcc cgtcctgctg tcagttgggg tgacattcat 300
 cctgattgct gtgtgcaagt tcaaaatgct ctectgccag ttgtgcaaag aaagtgagga 360
 aaggggtccg gactcggaac agacaccagg aggaccatca tttgttttca ctggcatcaa 420
 ccaacccatc accttccatg gggccactgt ggtgcagtac atccctctc cttatgggtc 480
 tccagagcct atggggataa ataccagcta cctgcagtct gtggtgagcc cctgcggcct 540
 cataacctct ggaggggagc cagccgccat gtcaagtcct cctcaatact acaccatcta 600
 ccctcaagat aactctgcat ttgtggttga tgagggctgc ctttctttca cggacggtgg 660
 aatcacagg cccaatcctg atgttgacca gctagaagag acacagctgg aagaggaggc 720
 ctgtgcctgc ttctctctc ccccttatga agaaatatac tctctccctc gctagaggct 780
 attctgatat aataacacaa tgctcagctc agggagcaag tgtttccgctc attgttacct 840
 gacaaccgtg gtgttctatg ttgtaacctt cagaagttac agcagcgccc aggcagcctg 900
 acagagatca ttcaaggggg gaaaggggaa gtgggagggt caatttctca gattggtaaa 960
 aattaggctg ggctggggaa attctcctcc ggaacagttt caaattccct cgggtaagaa 1020
 atctcctgta taagggtcag gagcaggaat ttcacttttt catccaccac cctccccctt 1080
 ctctgtagga aggcattggt ggctcaattt taaccccagc agccaatgga aaaatcacga 1140
 cttctgagac tttgggagtt tccacagagg tgagagtcgg gtgggaagga agcagggaag 1200
 agaaagcagg cccagctgga gatttcctgg tggctgtcct tggcccaaaa gcagactcac 1260
 taatcccaaa caactcagct gccatctggc ctctctgagg actctgggta ccttaaagac 1320
 tata 1324

<210> 59
 <211> 683
 <212> DNA
 <213> Homo sapiens

<400> 59
 caggaaagtt cgtgctgcta ggcagaggaa ctgcagcttg ttggcagggt aagggagcct 60
 gtttagctgt gtccagcaac aacttacgtg gtccctgcttg tgttccagggt gaagcgtctg 120
 gccgccgagc agaggaatca agacctgctc attctttcct cgggggatcc atccagcaat 180
 gacatcatct catgctgcca caaggacccc aagtctgggc tgctggggac cagccacgct 240
 ccccactgct cattccttca tcctagagac attctgactc tcctccgact gcgctgtgca 300
 caggcgtgac aagctctttt acatctcagt ctgcacaact tcaggcaatt agcagattga 360
 tatgcatcca acaaataattg attgaatata tgctaaatac ccagtaatgt ttcattgagt 420
 attgggtgaa taaaggaatg ctggttcctt ctggccatat taactcctgc acaatactaa 480
 gaaaaataaa ttgcactagc tgtggaataa tgtgaatccc aatgtcatct attgaaatat 540
 tacctgacta ttaagaggta tttatttttg tatcttttct agcaaagtaa ataaaattct 600
 taatacagca tatcccccta ttcacggggg gtatgttcca agacccccgg tggtatgcctg 660
 aaactatgga taataccaga tcc 683

<210> 60
 <211> 914
 <212> PRT
 <213> Homo sapiens

<400> 60

Met	Gly	Pro	Phe	Lys	Ser	Ser	Val	Phe	Ile	Leu	Ile	Leu	His	Leu	Leu
1				5					10					15	
Glu	Gly	Ala	Leu	Ser	Asn	Ser	Leu	Ile	Gln	Leu	Asn	Asn	Asn	Gly	Tyr
			20					25					30		
Glu	Gly	Ile	Val	Val	Ala	Ile	Asp	Pro	Asn	Val	Pro	Glu	Asp	Glu	Thr
		35					40					45			
Leu	Ile	Gln	Gln	Ile	Lys	Asp	Met	Val	Thr	Gln	Ala	Ser	Leu	Tyr	Leu
	50					55					60				
Phe	Glu	Ala	Thr	Gly	Lys	Arg	Phe	Tyr	Phe	Lys	Asn	Val	Ala	Ile	Leu
65					70					75					80
Ile	Pro	Glu	Thr	Trp	Lys	Thr	Lys	Ala	Asp	Tyr	Val	Arg	Pro	Lys	Leu
				85					90					95	
Glu	Thr	Tyr	Lys	Asn	Ala	Asp	Val	Leu	Val	Ala	Glu	Ser	Thr	Pro	Pro
			100					105					110		
Gly	Asn	Asp	Glu	Pro	Tyr	Thr	Glu	Gln	Met	Gly	Asn	Cys	Gly	Glu	Lys
		115					120					125			
Gly	Glu	Arg	Ile	His	Leu	Thr	Pro	Asp	Phe	Ile	Ala	Gly	Lys	Lys	Leu
	130					135					140				
Ala	Glu	Tyr	Gly	Pro	Gln	Gly	Lys	Ala	Phe	Val	His	Glu	Trp	Ala	His
145					150					155					160
Leu	Arg	Trp	Gly	Val	Phe	Asp	Glu	Tyr	Asn	Asn	Asp	Glu	Lys	Phe	Tyr
				165					170					175	
Leu	Ser	Asn	Gly	Arg	Ile	Gln	Ala	Val	Arg	Cys	Ser	Ala	Gly	Ile	Thr
			180					185					190		
Gly	Thr	Asn	Val	Val	Lys	Lys	Cys	Gln	Gly	Gly	Ser	Cys	Tyr	Thr	Lys
		195					200					205			
Arg	Cys	Thr	Phe	Asn	Lys	Val	Thr	Gly	Leu	Tyr	Glu	Lys	Gly	Cys	Glu
	210					215					220				
Phe	Val	Leu	Gln	Ser	Arg	Gln	Thr	Glu	Lys	Ala	Ser	Ile	Met	Phe	Ala
225					230					235					240
Gln	His	Val	Asp	Ser	Ile	Val	Glu	Phe	Cys	Thr	Glu	Gln	Asn	His	Asn
				245					250					255	
Lys	Glu	Ala	Pro	Asn	Lys	Gln	Asn	Gln	Lys	Cys	Asn	Leu	Arg	Ser	Thr
			260					265					270		
Trp	Glu	Val	Ile	Arg	Asp	Ser	Glu	Asp	Phe	Lys	Lys	Thr	Thr	Pro	Met
		275					280					285			
Thr	Thr	Gln	Pro	Pro	Asn	Pro	Thr	Phe	Ser	Leu	Leu	Gln	Ile	Gly	Gln
	290					295						300			
Arg	Ile	Val	Cys	Leu	Val	Leu	Asp	Lys	Ser	Gly	Ser	Met	Ala	Thr	Gly
305					310					315					320
Asn	Arg	Leu	Asn	Arg	Leu	Asn	Gln	Ala	Gly	Gln	Leu	Phe	Leu	Leu	Gln
			325						330					335	
Thr	Val	Glu	Leu	Gly	Ser	Trp	Val	Gly	Met	Val	Thr	Phe	Asp	Ser	Ala
			340					345					350		
Ala	His	Val	Gln	Ser	Glu	Leu	Ile	Gln	Ile	Asn	Ser	Gly	Ser	Asp	Arg
		355					360					365			
Asp	Thr	Leu	Ala	Lys	Arg	Leu	Pro	Ala	Ala	Ala	Ser	Gly	Gly	Thr	Ser
	370					375					380				
Ile	Cys	Ser	Gly	Leu	Arg	Ser	Ala	Phe	Thr	Val	Ile	Arg	Lys	Lys	Tyr
385					390					395					400
Pro	Thr	Asp	Gly	Ser	Glu	Ile	Val	Leu	Leu	Thr	Asp	Gly	Glu	Asp	Asn
				405					410					415	
Thr	Ile	Ser	Gly	Cys	Phe	Asn	Glu	Val	Lys	Gln	Ser	Gly	Ala	Ile	Ile
			420					425					430		

His	Thr	Val	Ala	Leu	Gly	Pro	Ser	Ala	Ala	Gln	Glu	Leu	Glu	Glu	Leu	
		435					440					445				
Ser	Lys	Met	Thr	Gly	Gly	Leu	Gln	Thr	Tyr	Ala	Ser	Asp	Gln	Val	Gln	
	450					455					460					
Asn	Asn	Gly	Leu	Ile	Asp	Ala	Phe	Gly	Ala	Leu	Ser	Ser	Gly	Asn	Gly	
465					470					475					480	
Ala	Val	Ser	Gln	Arg	Ser	Ile	Gln	Leu	Glu	Ser	Lys	Gly	Leu	Thr	Leu	
				485					490						495	
Gln	Asn	Ser	Gln	Trp	Met	Asn	Gly	Thr	Val	Ile	Val	Asp	Ser	Thr	Val	
			500					505					510			
Gly	Lys	Asp	Thr	Leu	Phe	Leu	Ile	Thr	Trp	Thr	Thr	Gln	Pro	Pro	Gln	
	515					520						525				
Ile	Leu	Leu	Trp	Asp	Pro	Ser	Gly	Gln	Lys	Gln	Gly	Gly	Phe	Val	Val	
530						535					540					
Asp	Lys	Asn	Thr	Lys	Met	Ala	Tyr	Leu	Gln	Ile	Pro	Gly	Ile	Ala	Lys	
545					550					555					560	
Val	Gly	Thr	Trp	Lys	Tyr	Ser	Leu	Gln	Ala	Ser	Ser	Gln	Thr	Leu	Thr	
				565					570						575	
Leu	Thr	Val	Thr	Ser	Arg	Ala	Ser	Asn	Ala	Thr	Leu	Pro	Pro	Ile	Thr	
			580					585						590		
Val	Thr	Ser	Lys	Thr	Asn	Lys	Asp	Thr	Ser	Lys	Phe	Pro	Ser	Pro	Leu	
		595					600					605				
Val	Val	Tyr	Ala	Asn	Ile	Arg	Gln	Gly	Ala	Ser	Pro	Ile	Leu	Arg	Ala	
610						615					620					
Ser	Val	Thr	Ala	Leu	Ile	Glu	Ser	Val	Asn	Gly	Lys	Thr	Val	Thr	Leu	
625					630					635					640	
Glu	Leu	Leu	Asp	Asn	Gly	Ala	Gly	Ala	Asp	Ala	Thr	Lys	Asp	Asp	Gly	
				645					650					655		
Val	Tyr	Ser	Arg	Tyr	Phe	Thr	Thr	Tyr	Asp	Thr	Asn	Gly	Arg	Tyr	Ser	
			660					665					670			
Val	Lys	Val	Arg	Ala	Leu	Gly	Gly	Val	Asn	Ala	Ala	Arg	Arg	Arg	Val	
		675				680						685				
Ile	Pro	Gln	Gln	Ser	Gly	Ala	Leu	Tyr	Ile	Pro	Gly	Trp	Ile	Glu	Asn	
690						695					700					
Asp	Glu	Ile	Gln	Trp	Asn	Pro	Pro	Arg	Pro	Glu	Ile	Asn	Lys	Asp	Asp	
705					710					715					720	
Val	Gln	His	Lys	Gln	Val	Cys	Phe	Ser	Arg	Thr	Ser	Ser	Gly	Gly	Ser	
				725					730					735		
Phe	Val	Ala	Ser	Asp	Val	Pro	Asn	Ala	Pro	Ile	Pro	Asp	Leu	Phe	Pro	
			740					745					750			
Pro	Gly	Gln	Ile	Thr	Asp	Leu	Lys	Ala	Glu	Ile	His	Gly	Gly	Ser	Leu	
		755					760					765				
Ile	Asn	Leu	Thr	Trp	Thr	Ala	Pro	Gly	Asp	Asp	Tyr	Asp	His	Gly	Thr	
770						775					780					
Ala	His	Lys	Tyr	Ile	Ile	Arg	Ile	Ser	Thr	Ser	Ile	Leu	Asp	Leu	Arg	
785					790					795					800	
Asp	Lys	Phe	Asn	Glu	Ser	Leu	Gln	Val	Asn	Thr	Thr	Ala	Leu	Ile	Pro	
				805					810					815		
Lys	Glu	Ala	Asn	Ser	Glu	Glu	Val	Phe	Leu	Phe	Lys	Pro	Glu	Asn	Ile	
			820					825					830			
Thr	Phe	Glu	Asn	Gly	Thr	Asp	Leu	Phe	Ile	Ala	Ile	Gln	Ala	Val	Asp	
		835				840						845				
Lys	Val	Asp	Leu	Lys	Ser	Glu	Ile	Ser	Asn	Ile	Ala	Arg	Val	Ser	Leu	
	850					855					860					
Phe	Ile	Pro	Pro	Gln	Thr	Pro	Pro	Glu	Thr	Pro	Ser	Pro	Asp	Glu	Thr	
865					870					875					880	

Ser Ala Pro Cys Pro Asn Ile His Ile Asn Ser Thr Ile Pro Gly Ile
 885 890 895
 His Ile Leu Lys Ile Met Trp Lys Trp Ile Gly Glu Leu Gln Leu Ser
 900 905 910
 Ile Ala

<210> 61
 <211> 501
 <212> PRT
 <213> Homo sapiens

<400> 61
 Met Lys Lys Glu Gly Arg Lys Arg Trp Lys Arg Lys Glu Asp Lys Lys
 1 5 10 15
 Arg Val Val Val Ser Asn Leu Leu Phe Glu Gly Trp Ser His Lys Glu
 20 25 30
 Asn Pro Asn Arg His His Arg Gly Asn Gln Ile Lys Thr Ser Lys Tyr
 35 40 45
 Thr Val Leu Ser Phe Val Pro Lys Asn Ile Phe Glu Gln Leu His Arg
 50 55 60
 Phe Ala Asn Leu Tyr Phe Val Gly Ile Ala Val Leu Asn Phe Ile Pro
 65 70 75 80
 Val Val Asn Ala Phe Gln Pro Glu Val Ser Met Ile Pro Ile Cys Val
 85 90 95
 Ile Leu Ala Val Thr Ala Ile Lys Asp Ala Trp Glu Asp Leu Arg Arg
 100 105 110
 Tyr Lys Ser Asp Lys Val Ile Asn Asn Arg Glu Cys Leu Ile Tyr Ser
 115 120 125
 Arg Lys Glu Gln Thr Tyr Val Gln Lys Cys Trp Lys Asp Val Arg Val
 130 135 140
 Gly Asp Phe Ile Gln Met Lys Cys Asn Glu Ile Val Pro Ala Asp Ile
 145 150 155 160
 Leu Leu Leu Phe Ser Ser Asp Pro Asn Gly Ile Cys His Leu Glu Thr
 165 170 175
 Ala Ser Leu Asp Gly Glu Thr Asn Leu Lys Gln Arg Arg Val Val Lys
 180 185 190
 Gly Phe Ser Gln Gln Glu Val Gln Phe Glu Pro Glu Leu Phe His Asn
 195 200 205
 Thr Ile Val Cys Glu Lys Pro Asn Asn His Leu Asn Lys Phe Lys Gly
 210 215 220
 Tyr Met Glu His Pro Asp Gln Thr Arg Thr Gly Phe Gly Cys Glu Ser
 225 230 235 240
 Leu Leu Leu Arg Gly Cys Thr Ile Arg Asn Thr Glu Met Ala Val Gly
 245 250 255
 Ile Val Ile Tyr Ala Gly His Glu Thr Lys Ala Met Leu Asn Asn Ser
 260 265 270
 Gly Pro Arg Tyr Lys Arg Ser Lys Ile Glu Arg Arg Met Asn Ile Asp
 275 280 285
 Ile Phe Phe Cys Ile Gly Ile Leu Ile Leu Met Cys Leu Ile Gly Ala
 290 295 300
 Val Gly His Ser Ile Trp Asn Gly Thr Phe Glu Glu His Pro Pro Phe
 305 310 315 320
 Asp Val Pro Asp Ala Asn Gly Ser Phe Leu Pro Ser Ala Leu Gly Gly
 325 330 335

Phe Tyr Met Phe Leu Thr Met Ile Ile Leu Leu Gln Val Leu Ile Pro
 340 345 350
 Ile Ser Leu Tyr Val Ser Ile Glu Leu Val Lys Leu Gly Gln Val Phe
 355 360 365
 Phe Leu Ser Asn Asp Leu Asp Leu Tyr Asp Glu Glu Thr Asp Leu Ser
 370 375 380
 Ile Gln Cys Arg Ala Leu Asn Ile Ala Glu Asp Leu Gly Gln Ile Gln
 385 390 395 400
 Tyr Ile Phe Ser Asp Lys Thr Gly Thr Leu Thr Glu Asn Lys Met Val
 405 410 415
 Phe Arg Arg Cys Thr Ile Met Gly Ser Glu Tyr Ser His Gln Glu Asn
 420 425 430
 Gly Ile Glu Ala Pro Lys Gly Ser Ile Pro Leu Ser Lys Arg Lys Tyr
 435 440 445
 Pro Ala Leu Leu Arg Asn Glu Glu Ile Lys Asp Ile Leu Leu Ala Leu
 450 455 460
 Leu Glu Ala Val Trp His Phe His Lys Leu Leu Pro Val Ser Leu Trp
 465 470 475 480
 Ser Ser Leu Ser Gln Ile Arg Ala Val Pro Ile Thr Cys Lys Leu Ser
 485 490 495
 Phe Val Tyr Lys Gly
 500

<210> 62
 <211> 154
 <212> PRT
 <213> Homo sapiens

<400> 62
 Met Gly Arg Arg Ser Pro Phe Lys Pro Arg Asn Lys Val Phe Gly Phe
 1 5 10 15
 Ser Tyr Pro Trp Cys Arg Ser Tyr Gln Pro Phe Pro Arg Lys Arg Ala
 20 25 30
 Trp Pro Pro Ser Arg Val Trp Leu Gly Ala Cys Cys Ala Ser Leu Ala
 35 40 45
 Ser Pro Pro Lys Gly Thr Ile Pro Ser Gly Glu Tyr Tyr Arg Pro Ala
 50 55 60
 Pro Ser Ser Ser Gly Asp Ser Leu Arg Arg Glu Ser Gly Ala Leu Leu
 65 70 75 80
 Gln Tyr Leu Pro Ser Leu Ala Ser Pro Cys Ala Asn His Ala Thr Arg
 85 90 95
 Cys Ser Leu Leu Phe Pro Ile Tyr Lys Ile Lys Met Thr Leu Leu Tyr
 100 105 110
 Leu Thr Gly Leu Ala Arg Thr His Cys Cys Cys Leu Ala Asp Arg Cys
 115 120 125
 Ala Glu Ala Val Glu Ser Ala Phe Tyr Leu Val Gly Ser Leu Cys Ile
 130 135 140
 Asn Ala Arg Gly Ala Ala His Leu Thr Asp
 145 150

<210> 63
 <211> 484

<212> PRT

<213> Homo sapiens

<400> 63

Met	Ala	Gly	Pro	Trp	Thr	Phe	Thr	Leu	Leu	Cys	Gly	Leu	Leu	Ala	Ala	
1				5				10						15		
Thr	Leu	Ile	Gln	Ala	Thr	Leu	Ser	Pro	Thr	Ala	Val	Leu	Ile	Leu	Gly	
			20					25					30			
Pro	Lys	Val	Ile	Lys	Glu	Lys	Leu	Thr	Gln	Glu	Leu	Lys	Asp	His	Asn	
		35					40					45				
Ala	Thr	Ser	Ile	Leu	Gln	Gln	Leu	Pro	Leu	Leu	Ser	Ala	Met	Arg	Glu	
	50					55					60					
Lys	Pro	Ala	Gly	Gly	Ile	Pro	Val	Leu	Gly	Ser	Leu	Val	Asn	Thr	Val	
65					70					75					80	
Leu	Lys	His	Ile	Ile	Trp	Leu	Lys	Val	Ile	Thr	Ala	Asn	Ile	Leu	Gln	
			85					90					95			
Leu	Gln	Val	Lys	Pro	Ser	Ala	Asn	Asp	Gln	Glu	Leu	Leu	Val	Lys	Ile	
			100					105					110			
Pro	Leu	Asp	Met	Val	Ala	Gly	Phe	Asn	Thr	Pro	Leu	Val	Lys	Thr	Ile	
		115					120					125				
Val	Glu	Phe	His	Met	Thr	Thr	Glu	Ala	Gln	Ala	Thr	Ile	Arg	Met	Asp	
	130					135					140					
Thr	Ser	Ala	Ser	Gly	Pro	Thr	Arg	Leu	Val	Leu	Ser	Asp	Cys	Ala	Thr	
145					150					155					160	
Ser	His	Gly	Ser	Leu	Arg	Ile	Gln	Leu	Leu	His	Lys	Leu	Ser	Phe	Leu	
			165					170						175		
Val	Asn	Ala	Leu	Ala	Lys	Gln	Val	Met	Asn	Leu	Leu	Val	Pro	Ser	Leu	
			180					185					190			
Pro	Asn	Leu	Val	Lys	Asn	Gln	Leu	Cys	Pro	Val	Ile	Glu	Ala	Ser	Phe	
	195					200						205				
Asn	Gly	Met	Tyr	Ala	Asp	Leu	Leu	Gln	Leu	Val	Lys	Val	Pro	Ile	Ser	
	210					215					220					
Leu	Ser	Ile	Asp	Arg	Leu	Glu	Phe	Asp	Leu	Leu	Tyr	Pro	Ala	Ile	Lys	
225					230					235					240	
Gly	Asp	Thr	Ile	Gln	Leu	Tyr	Leu	Gly	Ala	Lys	Leu	Leu	Asp	Ser	Gln	
			245					250						255		
Gly	Lys	Val	Thr	Lys	Trp	Phe	Asn	Asn	Ser	Ala	Ala	Ser	Leu	Thr	Met	
		260						265					270			
Pro	Thr	Leu	Asp	Asn	Ile	Pro	Phe	Ser	Leu	Ile	Val	Ser	Gln	Asp	Val	
		275					280					285				
Val	Lys	Ala	Ala	Val	Ala	Ala	Val	Leu	Ser	Pro	Glu	Glu	Phe	Met	Val	
	290					295					300					
Leu	Leu	Asp	Ser	Val	Leu	Pro	Glu	Ser	Ala	His	Arg	Leu	Lys	Ser	Ser	
305					310					315					320	
Ile	Gly	Leu	Ile	Asn	Glu	Lys	Ala	Ala	Asp	Lys	Leu	Gly	Ser	Thr	Gln	
			325						330					335		
Ile	Val	Lys	Ile	Leu	Thr	Gln	Asp	Thr	Pro	Glu	Phe	Phe	Ile	Asp	Gln	
		340						345					350			
Gly	His	Ala	Lys	Val	Ala	Gln	Leu	Ile	Val	Leu	Glu	Val	Phe	Pro	Ser	
	355					360					365					
Ser	Glu	Ala	Leu	Arg	Pro	Leu	Phe	Thr	Leu	Gly	Ile	Glu	Ala	Ser	Ser	
	370					375					380					
Glu	Ala	Gln	Phe	Tyr	Thr	Lys	Gly	Asp	Gln	Leu	Ile	Leu	Asn	Leu	Asn	
385					390					395					400	
Asn	Ile	Ser	Ser	Asp	Arg	Ile	Gln	Leu	Met	Asn	Ser	Gly	Ile	Gly	Trp	
			405					410						415		

Phe Gln Pro Asp Val Leu Lys Asn Ile Ile Thr Glu Ile Ile His Ser
 420 425 430
 Ile Leu Leu Pro Asn Gln Asn Gly Lys Leu Arg Ser Gly Val Pro Val
 435 440 445
 Ser Leu Val Lys Ala Leu Gly Phe Glu Ala Ala Glu Ser Ser Leu Thr
 450 455 460
 Lys Asp Ala Leu Val Leu Thr Pro Ala Ser Leu Trp Lys Pro Ser Ser
 465 470 475 480
 Pro Val Ser Gln

<210> 64
 <211> 256
 <212> PRT
 <213> Homo sapiens

<400> 64
 Met Phe Gln Thr Gly Gly Leu Ile Val Phe Tyr Gly Leu Leu Ala Gln
 1 5 10 15
 Thr Met Ala Gln Phe Gly Gly Leu Pro Val Pro Leu Asp Gln Thr Leu
 20 25 30
 Pro Leu Asn Val Asn Pro Ala Leu Pro Leu Ser Pro Thr Gly Leu Ala
 35 40 45
 Gly Ser Leu Thr Asn Ala Leu Ser Asn Gly Leu Leu Ser Gly Gly Leu
 50 55 60
 Leu Gly Ile Leu Glu Asn Leu Pro Leu Leu Asp Ile Leu Lys Pro Gly
 65 70 75 80
 Gly Gly Thr Ser Gly Gly Leu Leu Gly Gly Leu Leu Gly Lys Val Thr
 85 90 95
 Ser Val Ile Pro Gly Leu Asn Asn Ile Ile Asp Ile Lys Val Thr Asp
 100 105 110
 Pro Gln Leu Leu Glu Leu Gly Leu Val Gln Ser Pro Asp Gly His Arg
 115 120 125
 Leu Tyr Val Thr Ile Pro Leu Gly Ile Lys Leu Gln Val Asn Thr Pro
 130 135 140
 Leu Val Gly Ala Ser Leu Leu Arg Leu Ala Val Lys Leu Asp Ile Thr
 145 150 155 160
 Ala Glu Ile Leu Ala Val Arg Asp Lys Gln Glu Arg Ile His Leu Val
 165 170 175
 Leu Gly Asp Cys Thr His Ser Pro Gly Ser Leu Gln Ile Ser Leu Leu
 180 185 190
 Asp Gly Leu Gly Pro Leu Pro Ile Gln Gly Leu Leu Asp Ser Leu Thr
 195 200 205
 Gly Ile Leu Asn Lys Val Leu Pro Glu Leu Val Gln Gly Asn Val Cys
 210 215 220
 Pro Leu Val Asn Glu Val Leu Arg Gly Leu Asp Ile Thr Leu Val His
 225 230 235 240
 Asp Ile Val Asn Met Leu Ile His Gly Leu Gln Phe Val Ile Lys Val
 245 250 255

<210> 65
 <211> 791
 <212> PRT
 <213> Homo sapiens

```

<400> 65
Met Ser Gln Pro Arg Pro Arg Tyr Val Val Asp Arg Ala Ala Tyr Ser
1      5      10
Leu Thr Leu Phe Asp Asp Glu Phe Glu Lys Lys Asp Arg Thr Tyr Pro
20      25      30
Val Gly Glu Lys Leu Arg Asn Ala Phe Arg Cys Ser Ser Ala Lys Ile
35      40      45
Lys Ala Val Val Phe Gly Leu Leu Pro Val Leu Ser Trp Leu Pro Lys
50      55      60
Tyr Lys Ile Lys Asp Tyr Ile Ile Pro Asp Leu Leu Gly Gly Leu Ser
65      70      75      80
Gly Gly Ser Ile Gln Val Pro Gln Gly Met Ala Phe Ala Leu Leu Ala
85      90      95
Asn Leu Pro Ala Val Asn Gly Leu Tyr Ser Ser Phe Phe Pro Leu Leu
100     105     110
Thr Tyr Phe Phe Leu Gly Gly Val His Gln Met Val Pro Gly Thr Phe
115     120     125
Ala Val Ile Ser Ile Leu Val Gly Asn Ile Cys Leu Gln Leu Ala Pro
130     135     140
Glu Ser Lys Phe Gln Val Phe Asn Asn Ala Thr Asn Glu Ser Tyr Val
145     150     155     160
Asp Thr Ala Ala Met Glu Ala Glu Arg Leu His Val Ser Ala Thr Leu
165     170     175
Ala Cys Leu Thr Ala Ile Ile Gln Met Gly Leu Gly Phe Met Gln Phe
180     185     190
Gly Phe Val Ala Ile Tyr Leu Ser Glu Ser Phe Ile Arg Gly Phe Met
195     200     205
Thr Ala Ala Gly Leu Gln Ile Leu Ile Ser Val Leu Lys Tyr Ile Phe
210     215     220
Gly Leu Thr Ile Pro Ser Tyr Thr Gly Pro Gly Ser Ile Val Phe Thr
225     230     235     240
Phe Ile Asp Ile Cys Lys Asn Leu Pro His Thr Asn Ile Ala Ser Leu
245     250     255
Ile Phe Ala Leu Ile Ser Gly Ala Phe Leu Val Leu Val Lys Glu Leu
260     265     270
Asn Ala Arg Tyr Met His Lys Ile Arg Phe Pro Ile Pro Thr Glu Met
275     280     285
Ile Val Val Val Val Ala Thr Ala Ile Ser Gly Gly Cys Lys Met Pro
290     295     300
Lys Lys Tyr His Met Gln Ile Val Gly Glu Ile Gln Arg Gly Phe Pro
305     310     315     320
Thr Pro Val Ser Pro Val Val Ser Gln Trp Lys Asp Met Ile Gly Thr
325     330     335
Ala Phe Ser Leu Ala Ile Val Ser Tyr Val Ile Asn Leu Ala Met Gly
340     345     350
Arg Thr Leu Ala Asn Lys His Gly Tyr Asp Val Asp Ser Asn Gln Glu
355     360     365
Met Ile Ala Leu Gly Cys Ser Asn Phe Phe Gly Ser Phe Phe Lys Ile
370     375     380
His Val Ile Cys Cys Ala Leu Ser Val Thr Leu Ala Val Asp Gly Ala
385     390     395     400
Gly Gly Lys Ser Gln Val Ala Ser Leu Cys Val Ser Leu Val Val Met
405     410     415
Ile Thr Met Leu Val Leu Gly Ile Tyr Leu Tyr Pro Leu Pro Lys Ser
420     425     430

```

Val	Leu	Gly	Ala	Leu	Ile	Ala	Val	Asn	Leu	Lys	Asn	Ser	Leu	Lys	Gln
		435						440				445			
Leu	Thr	Asp	Pro	Tyr	Tyr	Leu	Trp	Arg	Lys	Ser	Lys	Leu	Asp	Cys	Cys
		450						455				460			
Ile	Trp	Val	Val	Ser	Phe	Leu	Ser	Ser	Phe	Phe	Leu	Ser	Leu	Pro	Tyr
					470					475					480
Gly	Val	Ala	Val	Gly	Val	Ala	Phe	Ser	Val	Leu	Val	Val	Val	Phe	Gln
				485					490					495	
Thr	Gln	Phe	Arg	Asn	Gly	Tyr	Ala	Leu	Ala	Gln	Val	Met	Asp	Thr	Asp
			500					505					510		
Ile	Tyr	Val	Asn	Pro	Lys	Thr	Tyr	Asn	Arg	Ala	Gln	Asp	Ile	Gln	Gly
		515					520					525			
Ile	Lys	Ile	Ile	Thr	Tyr	Cys	Ser	Pro	Leu	Tyr	Phe	Ala	Asn	Ser	Glu
	530					535					540				
Ile	Phe	Arg	Gln	Lys	Val	Ile	Ala	Lys	Thr	Gly	Met	Asp	Pro	Gln	Lys
	545				550					555					560
Val	Leu	Leu	Ala	Lys	Gln	Lys	Tyr	Leu	Lys	Lys	Gln	Glu	Lys	Arg	Arg
				565					570					575	
Met	Arg	Pro	Thr	Gln	Gln	Arg	Arg	Ser	Leu	Phe	Met	Lys	Thr	Lys	Thr
			580					585					590		
Val	Ser	Leu	Gln	Glu	Leu	Gln	Gln	Asp	Phe	Glu	Asn	Ala	Pro	Pro	Thr
		595					600					605			
Asp	Pro	Asn	Asn	Asn	Gln	Thr	Pro	Ala	Asn	Gly	Thr	Ser	Val	Ser	Tyr
	610					615					620				
Ile	Thr	Phe	Ser	Pro	Asp	Ser	Ser	Ser	Pro	Ala	Gln	Ser	Glu	Pro	Pro
	625				630					635					640
Ala	Ser	Ala	Glu	Ala	Pro	Gly	Glu	Pro	Ser	Asp	Met	Leu	Ala	Ser	Val
				645					650					655	
Pro	Pro	Phe	Val	Thr	Phe	His	Thr	Leu	Ile	Leu	Asp	Met	Ser	Gly	Val
			660					665					670		
Ser	Phe	Val	Asp	Leu	Met	Gly	Ile	Lys	Ala	Leu	Ala	Lys	Leu	Ser	Ser
		675				680						685			
Thr	Tyr	Gly	Lys	Ile	Gly	Val	Lys	Val	Phe	Leu	Val	Asn	Ile	His	Ala
	690					695						700			
Gln	Val	Tyr	Asn	Asp	Ile	Ser	His	Gly	Gly	Val	Phe	Glu	Asp	Gly	Ser
	705				710						715				720
Leu	Glu	Cys	Lys	His	Val	Phe	Pro	Ser	Ile	His	Asp	Ala	Val	Leu	Phe
				725					730					735	
Ala	Gln	Ala	Asn	Ala	Arg	Asp	Val	Thr	Pro	Gly	His	Asn	Phe	Gln	Gly
			740					745					750		
Ala	Pro	Gly	Asp	Ala	Glu	Leu	Ser	Leu	Tyr	Asp	Ser	Glu	Glu	Asp	Ile
		755					760					765			
Arg	Ser	Tyr	Trp	Asp	Leu	Glu	Gln	Glu	Met	Phe	Gly	Ser	Met	Phe	His
	770					775					780				
Ala	Glu	Thr	Leu	Thr	Ala	Leu									
	785				790										

<210> 66
 <211> 243
 <212> PRT
 <213> Homo sapiens

<400> 66
 Met Glu Gln Gly Ser Gly Arg Leu Glu Asp Phe Pro Val Asn Val Phe
 1 5 10 15

Ser	Val	Thr	Pro	Tyr	Thr	Pro	Ser	Thr	Ala	Asp	Ile	Gln	Val	Ser	Asp
			20					25					30		
Asp	Asp	Lys	Ala	Gly	Ala	Thr	Leu	Leu	Phe	Ser	Gly	Ile	Phe	Leu	Gly
		35					40					45			
Leu	Val	Gly	Ile	Thr	Phe	Thr	Val	Met	Gly	Trp	Ile	Lys	Tyr	Gln	Gly
	50					55					60				
Val	Ser	His	Phe	Glu	Trp	Thr	Gln	Leu	Leu	Gly	Pro	Val	Leu	Leu	Ser
65					70					75					80
Val	Gly	Val	Thr	Phe	Ile	Leu	Ile	Ala	Val	Cys	Lys	Phe	Lys	Met	Leu
				85					90					95	
Ser	Cys	Gln	Leu	Cys	Lys	Glu	Ser	Glu	Glu	Arg	Val	Pro	Asp	Ser	Glu
			100					105					110		
Gln	Thr	Pro	Gly	Gly	Pro	Ser	Phe	Val	Phe	Thr	Gly	Ile	Asn	Gln	Pro
		115					120					125			
Ile	Thr	Phe	His	Gly	Ala	Thr	Val	Val	Gln	Tyr	Ile	Pro	Pro	Pro	Tyr
	130					135					140				
Gly	Ser	Pro	Glu	Pro	Met	Gly	Ile	Asn	Thr	Ser	Tyr	Leu	Gln	Ser	Val
145					150					155					160
Val	Ser	Pro	Cys	Gly	Leu	Ile	Thr	Ser	Gly	Gly	Ala	Ala	Ala	Ala	Met
				165					170					175	
Ser	Ser	Pro	Pro	Gln	Tyr	Tyr	Thr	Ile	Tyr	Pro	Gln	Asp	Asn	Ser	Ala
			180					185					190		
Phe	Val	Val	Asp	Glu	Gly	Cys	Leu	Ser	Phe	Thr	Asp	Gly	Gly	Asn	His
		195					200					205			
Arg	Pro	Asn	Pro	Asp	Val	Asp	Gln	Leu	Glu	Glu	Thr	Gln	Leu	Glu	Glu
	210					215					220				
Glu	Ala	Cys	Ala	Cys	Phe	Ser	Pro	Pro	Pro	Tyr	Glu	Glu	Ile	Tyr	Ser
225					230					235					240
Leu	Pro	Arg													

<210> 67
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 67
 acacgaatgg tagatacagt g

21

<210> 68
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 68
 atacttgtga gctgttccat g

21

<210> 69
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>

<223> Description of the artificial sequence: Oligonucleotide

<400> 69

actggttacct tgcattggact g

21

<210> 70
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>

<223> Description of the artificial sequence: Oligonucleotide

<400> 70

caatgagaac acatggacat g

21

<210> 71
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>

<223> Description of the artificial sequence: Oligonucleotide

<400> 71

ccatgaaagc tccatgtcta c

21

<210> 72
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>

<223> Description of the artificial sequence: Oligonucleotide

<400> 72

agagatggca catattctgt c

21

<210> 73
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>

<223> Description of the artificial sequence: Oligonucleotide

<400> 73
atcggctgaa gtcaagcatc g

21

<210> 74
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of the artificial sequence: Oligonucleotide

<400> 74
tggtcagtga ggactcagct g

21

<210> 75
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of the artificial sequence: Oligonucleotide

<400> 75
tttctctgct tgatgcactt g

21

<210> 76
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of the artificial sequence: Oligonucleotide

<400> 76
.gtgagcactg ggaagcagct c

21

<210> 77
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of the artificial sequence: Oligonucleotide

<400> 77
ggcaaattgct agagacgtga c

21

<210> 78
<211> 21

<212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 78
 aggtgtcctt cagctgccaa g 21

<210> 79
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 79
 gttaagtgct ctctggattt g 21

<210> 80
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 80
 atcctgattg ctgtgtgcaa g 21

<210> 81
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 81
 ctcttctagc tggtaacat c 21

<210> 82
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 82
 ccagcaacaa cttacgtggt c 21

<210> 83
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>

<223> Description of the artificial sequence: Oligonucleotide

<400> 83

cctttattca cccaatcact c

21

<210> 84

<211> 2165

<212> DNA

<213> Homo sapiens

<400> 84

agaacagcgc	agtttgcct	ccgctcacgc	agagcctctc	cgtggcctcc	gcaccttgag	60
cattaggcca	gttctcctct	tctctctaat	ccatccgtca	cctctcctgt	catccgtttc	120
catgcogtga	ggtccattca	cagaacacat	ccatggctct	catgctcagt	ttggttctga	180
gtctcctcaa	gctgggatca	gggcagtggc	aggtgttttg	gccagacaag	cctgtccagg	240
ccttggtggg	ggaggacgca	gcattctcct	gtttcctgtc	tcctaagacc	aatgcagagg	300
ccatggaagt	gcggttcttc	aggggccagt	tctctagcgt	ggtccacctc	tacagggacg	360
ggaaggacca	gccatttatg	cagatgccac	agtatcaagg	caggacaaaa	ctggtgaagg	420
attctattgc	ggagggg'gcg	atctctctga	ggctggaaaa	cattactgtg	ttggatgctg	480
gcctctatgg	gtgcaggatt	agttcccagt	cttactacca	gaaggccatc	tgggagctac	540
aggtgtcagc	actgggctca	gttcctctca	tttccatcac	gggatatgtt	gatagagaca	600
tcagctact	ctgtcagtec	tcgggctggg	tcccccgccc	cacagcgaag	tggaaaggtc	660
cacaaggaca	ggatttgtec	acagactcca	ggacaaacag	agacatgcat	ggcctgtttg	720
atgtggagat	ctctctgacc	gtccaagaga	acgccgggag	catatcctgt	tccatgcggc	780
atgtcctct	gagccgagag	gtggaatcca	gggtacagat	aggagatacc	tttttcgagc	840
ctatatcgtg	gcacctgggt	accaaagtac	tgggaatact	ctgctgtggc	ctattttttg	900
gcattgttgg	actgaagatt	ttcttctcca	aattccagtg	taagcgagag	agagaagcat	960
gggccggtgc	cttattcatg	gttccagcag	ggacaggatc	agagatgctc	ccacatccag	1020
ctgcttctct	tcttctagtc	ctagcctcca	ggggcccagg	cccaaaaaag	gaaaatccag	1080
gcggaactgg	actggagaag	aaagcacgga	caggcagaat	tgagagacgc	ccggaaacac	1140
gcagtggagg	tgactctgga	tcagagacgc	gctcacccga	agctctgcgt	ttctgatctg	1200
aaaactgtaa	cccatagaaa	agctccccag	gaggtgcctc	actctgagaa	gagatttaca	1260
aggaagagtg	tgggtggcttc	tcagagtttc	caagcaggga	aacattactg	ggaggtggac	1320
ggaggacaca	ataaaaagggtg	gcgcgtggga	gtgtgccggg	atgatgtgga	caggaggaag	1380
gagtacgtga	ctttgtctcc	cgatcatggg	tactgggtcc	tcagactgaa	tggagaacat	1440
ttgtatttca	cattaaatcc	ccgtttttatc	agcgtcttcc	ccaggacccc	acctacaaaa	1500
ataggggtct	tcctggacta	tgagtgtggg	accatctcct	tcttcaacat	aaatgaccag	1560
tcctttattt	ataccctgac	atgtcggttt	gaaggcttat	tgaggcccta	cattgagtat	1620
ccgtcctata	atgagcaaaa	tggaaactccc	atagtcctct	gcccagtcac	ccaggaatca	1680
gagaaagagg	cctcttggca	aagggcctct	gcaatcccag	agacaagcaa	cagtgagtcc	1740
tcctcacagg	caaccacgcc	cttcctcccc	aggggtgaaa	tgtaggatga	atcacatccc	1800
acattcttct	ttagggatat	taaggctctct	ctcccagatc	caaagtcccg	cagcagccgg	1860
ccaaggtggc	ttccagatga	agggggactg	gcctgtccac	atgggagtca	ggtgtcatgg	1920
ctgccctgag	ctgggaggga	agaaggctga	cattacattt	agtttgcctc	cactccatct	1980
ggctaagtga	tcttgaaata	ccacctctca	ggtgaagaac	cgtcaggaat	tcccatctca	2040
caggctgtgg	tgtagattaa	gtagacaagg	aatgtgaata	atgcttagat	cttattgatg	2100
acagagtgtg	tcctaattgg	ttgttcatta	tattacactt	tcagtaaaaa	aaaaaaaaaa	2160
aaaaa						2165

<210> 85
 <211> 347
 <212> PRT
 <213> Homo sapiens

<400> 85
 Met Ala Leu Met Leu Ser Leu Val Leu Ser Leu Leu Lys Leu Gly Ser
 1 5 10 15
 Gly Gln Trp Gln Val Phe Gly Pro Asp Lys Pro Val Gln Ala Leu Val
 20 25 30
 Gly Glu Asp Ala Ala Phe Ser Cys Phe Leu Ser Pro Lys Thr Asn Ala
 35 40 45
 Glu Ala Met Glu Val Arg Phe Phe Arg Gly Gln Phe Ser Ser Val Val
 50 55 60
 His Leu Tyr Arg Asp Gly Lys Asp Gln Pro Phe Met Gln Met Pro Gln
 65 70 75 80
 Tyr Gln Gly Arg Thr Lys Leu Val Lys Asp Ser Ile Ala Glu Gly Arg
 85 90 95
 Ile Ser Leu Arg Leu Glu Asn Ile Thr Val Leu Asp Ala Gly Leu Tyr
 100 105 110
 Gly Cys Arg Ile Ser Ser Gln Ser Tyr Tyr Gln Lys Ala Ile Trp Glu
 115 120 125
 Leu Gln Val Ser Ala Leu Gly Ser Val Pro Leu Ile Ser Ile Thr Gly
 130 135 140
 Tyr Val Asp Arg Asp Ile Gln Leu Leu Cys Gln Ser Ser Gly Trp Phe
 145 150 155 160
 Pro Arg Pro Thr Ala Lys Trp Lys Gly Pro Gln Gly Gln Asp Leu Ser
 165 170 175
 Thr Asp Ser Arg Thr Asn Arg Asp Met His Gly Leu Phe Asp Val Glu
 180 185 190
 Ile Ser Leu Thr Val Gln Glu Asn Ala Gly Ser Ile Ser Cys Ser Met
 195 200 205
 Arg His Ala His Leu Ser Arg Glu Val Glu Ser Arg Val Gln Ile Gly
 210 215 220
 Asp Thr Phe Phe Glu Pro Ile Ser Trp His Leu Ala Thr Lys Val Leu
 225 230 235 240
 Gly Ile Leu Cys Cys Gly Leu Phe Phe Gly Ile Val Gly Leu Lys Ile
 245 250 255
 Phe Phe Ser Lys Phe Gln Cys Lys Arg Glu Arg Glu Ala Trp Ala Gly
 260 265 270
 Ala Leu Phe Met Val Pro Ala Gly Thr Gly Ser Glu Met Leu Pro His
 275 280 285
 Pro Ala Ala Ser Leu Leu Leu Val Leu Ala Ser Arg Gly Pro Gly Pro
 290 295 300
 Lys Lys Glu Asn Pro Gly Gly Thr Gly Leu Glu Lys Lys Ala Arg Thr
 305 310 315 320
 Gly Arg Ile Glu Arg Arg Pro Glu Thr Arg Ser Gly Gly Asp Ser Gly
 325 330 335
 Ser Arg Asp Gly Ser Pro Glu Ala Leu Arg Phe
 340 345

<210> 86
 <211> 21

<212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 86
 attcatgggtt ccagcagggga c 21

<210> 87
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 87
 gggagacaaa gtcacgtact c 21

<210> 88
 <211> 22
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 88
 tcctgggtgtt cgtggtctgc tt 22

<210> 89
 <211> 22
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 89
 gagagtcctg gcttttgtgg gc 22

<210> 90
 <211> 15
 <212> PRT
 <213> Homo sapiens

<400> 90
 Gly Ser Ser Asp Leu Thr Trp Pro Pro Ala Ile Lys Leu Gly Cys
 1 5 10 15

<210> 91
 <211> 16
 <212> PRT
 <213> Homo sapiens

<400> 91
 Asp Arg Tyr Val Ala Val Arg His Pro Leu Arg Ala Arg Gly Leu Arg
 1 5 10 15

<210> 92
 <211> 15
 <212> PRT
 <213> Homo sapiens

<400> 92
 Val Ala Pro Arg Ala Lys Ala His Lys Ser Gln Asp Ser Leu Cys
 1 5 10 15

<210> 93
 <211> 13
 <212> PRT
 <213> Homo sapiens

<400> 93
 Cys Phe Arg Ser Thr Arg His Asn Phe Asn Ser Met Arg
 1 5 10

<210> 94
 <211> 22
 <212> PRT
 <213> Homo sapiens

<400> 94
 Met Asn Gly Thr Tyr Asn Thr Cys Gly Ser Ser Asp Leu Thr Trp Pro
 1 5 10 15
 Pro Ala Ile Lys Leu Gly
 20

<210> 95
 <211> 14
 <212> PRT
 <213> Homo sapiens

<400> 95
 Arg Asp Thr Ser Asp Thr Pro Leu Cys Gln Leu Ser Gln Gly
 1 5 10

<210> 96
 <211> 22
 <212> PRT
 <213> Homo sapiens

<400> 96

Gly Ile Gln Glu Gly Gly Phe Cys Phe Arg Ser Thr Arg His Asn Phe
 1 5 10 15
 Asn Ser Met Arg Phe Pro
 20

<210> 97

<211> 30

<212> PRT

<213> Homo sapiens

<400> 97

Ala Lys Glu Phe Gln Glu Ala Ser Ala Leu Ala Val Ala Pro Arg Ala
 1 5 10 15
 Lys Ala His Lys Ser Gln Asp Ser Leu Cys Val Thr Leu Ala
 20 25 30

<210> 98

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of the artificial sequence: Oligonucleotide

<400> 98

tcctgctcgt cgctctcctg at

22

<210> 99

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of the artificial sequence: Oligonucleotide

<400> 99

tcgcttttttg tcgtatttgc

20

<210> 100

<211> 15

<212> PRT

<213> Homo sapiens

<400> 100

His Asn Gly Ser Tyr Glu Ile Ser Val Leu Met Met Gly Asn Ser
 1 5 10 15

<210> 101

<211> 15

<212> PRT

<213> Homo sapiens

<400> 101

Asn	Leu	Pro	Thr	Pro	Pro	Thr	Val	Glu	Asn	Gln	Gln	Arg	Leu	Ala
1				5					10					15

<210> 102

<211> 619

<212> PRT

<213> Homo sapiens

<400> 102

Arg	Lys	Tyr	Arg	Lys	Asp	Tyr	Glu	Leu	Arg	Gln	Lys	Lys	Trp	Ser	His
1				5					10					15	
Ile	Pro	Pro	Glu	Asn	Ile	Phe	Pro	Leu	Glu	Thr	Asn	Glu	Thr	Asn	His
			20					25					30		
Val	Ser	Leu	Lys	Ile	Asp	Asp	Asp	Lys	Arg	Arg	Asp	Thr	Ile	Gln	Arg
		35					40					45			
Leu	Arg	Gln	Cys	Lys	Tyr	Asp	Lys	Lys	Arg	Val	Ile	Leu	Lys	Asp	Leu
	50					55					60				
Lys	His	Asn	Asp	Gly	Asn	Phe	Thr	Glu	Lys	Gln	Lys	Ile	Glu	Leu	Asn
65					70					75					80
Lys	Leu	Leu	Gln	Ile	Asp	Tyr	Tyr	Asn	Leu	Thr	Lys	Phe	Tyr	Gly	Thr
				85					90					95	
Val	Lys	Leu	Asp	Thr	Met	Ile	Phe	Gly	Val	Ile	Glu	Tyr	Cys	Glu	Arg
			100					105					110		
Gly	Ser	Leu	Arg	Glu	Val	Leu	Asn	Asp	Thr	Ile	Ser	Tyr	Pro	Asp	Gly
		115					120					125			
Thr	Phe	Met	Asp	Trp	Glu	Phe	Lys	Ile	Ser	Val	Leu	Tyr	Asp	Ile	Ala
	130					135					140				
Lys	Gly	Met	Ser	Tyr	Leu	His	Ser	Ser	Lys	Thr	Glu	Val	His	Gly	Arg
145					150					155					160
Leu	Lys	Ser	Thr	Asn	Cys	Val	Val	Asp	Ser	Arg	Met	Val	Val	Lys	Ile
				165					170					175	
Thr	Asp	Phe	Gly	Cys	Asn	Ser	Ile	Leu	Pro	Pro	Lys	Lys	Asp	Leu	Trp
			180					185					190		
Thr	Ala	Pro	Glu	His	Leu	Arg	Gln	Ala	Asn	Ile	Ser	Gln	Lys	Gly	Asp
		195					200						205		
Val	Tyr	Ser	Tyr	Gly	Ile	Ile	Ala	Gln	Glu	Ile	Ile	Leu	Arg	Lys	Glu
	210					215						220			
Thr	Phe	Tyr	Thr	Leu	Ser	Cys	Arg	Asp	Arg	Asn	Glu	Lys	Ile	Phe	Arg
225					230					235					240
Val	Glu	Asn	Ser	Asn	Gly	Met	Lys	Pro	Phe	Arg	Pro	Asp	Leu	Phe	Leu
				245					250					255	
Glu	Thr	Ala	Glu	Glu	Lys	Glu	Leu	Glu	Val	Tyr	Leu	Leu	Val	Lys	Asn
			260					265					270		
Cys	Trp	Glu	Glu	Asp	Pro	Glu	Lys	Arg	Pro	Asp	Phe	Lys	Lys	Ile	Glu
		275					280					285			
Thr	Thr	Leu	Ala	Lys	Ile	Phe	Gly	Leu	Phe	His	Asp	Gln	Lys	Asn	Glu
	290					295					300				
Ser	Tyr	Met	Asp	Thr	Leu	Ile	Arg	Arg	Leu	Gln	Leu	Tyr	Ser	Arg	Asn
305					310					315					320
Leu	Glu	His	Leu	Val	Glu	Glu	Arg	Thr	Gln	Leu	Tyr	Lys	Ala	Glu	Arg
				325					330					335	
Asp	Arg	Ala	Asp	Arg	Leu	Asn	Phe	Met	Leu	Leu	Pro	Arg	Leu	Val	Val
			340					345					350		

Lys Ser Leu Lys Glu Lys Gly Phe Val Glu Pro Glu Leu Tyr Glu Glu
 355 360 365
 Val Thr Ile Tyr Phe Ser Asp Ile Val Gly Phe Thr Thr Ile Cys Lys
 370 375 380
 Tyr Ser Thr Pro Met Glu Val Val Asp Met Leu Asn Asp Ile Tyr Lys
 385 390 395 400
 Ser Phe Asp His Ile Val Asp His His Asp Val Tyr Lys Val Glu Thr
 405 410 415
 Ile Gly Asp Ala Tyr Met Val Ala Ser Gly Leu Pro Lys Arg Asn Gly
 420 425 430
 Asn Arg His Ala Ile Asp Ile Ala Lys Met Ala Leu Glu Ile Leu Ser
 435 440 445
 Phe Met Gly Thr Phe Glu Leu Glu His Leu Pro Gly Leu Pro Ile Trp
 450 455 460
 Ile Arg Ile Gly Val His Ser Gly Pro Cys Ala Ala Gly Val Val Gly
 465 470 475 480
 Ile Lys Met Pro Arg Tyr Cys Leu Phe Gly Asp Thr Val Asn Thr Ala
 485 490 495
 Ser Arg Met Glu Ser Thr Gly Leu Pro Leu Arg Ile His Val Ser Gly
 500 505 510
 Ser Thr Ile Ala Ile Leu Lys Arg Thr Glu Cys Gln Phe Leu Tyr Glu
 515 520 525
 Val Arg Gly Glu Thr Tyr Leu Lys Gly Arg Gly Asn Glu Thr Thr Tyr
 530 535 540
 Trp Leu Thr Gly Met Lys Asp Gln Lys Phe Asn Leu Pro Thr Pro Pro
 545 550 555 560
 Thr Val Glu Asn Gln Gln Arg Leu Gln Ala Glu Phe Ser Asp Met Ile
 565 570 575
 Ala Asn Ser Leu Gln Lys Arg Gln Ala Ala Gly Ile Arg Ser Gln Lys
 580 585 590
 Pro Arg Arg Val Ala Ser Tyr Lys Lys Gly Thr Leu Glu Tyr Leu Gln
 595 600 605
 Leu Asn Thr Thr Asp Lys Glu Ser Thr Tyr Phe
 610 615

<210> 103
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 103
 gctggtaact atcttcctgc

20

<210> 104
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 104
gaagaatggt gtccagaggt

20

<210> 105
<211> 15
<212> PRT
<213> Homo sapiens

<400> 105
Leu Ile Asn Lys Val Pro Leu Pro Val Asp Lys Leu Ala Pro Leu
1 5 10 15

<210> 106
<211> 15
<212> PRT
<213> Homo sapiens

<400> 106
Ser Glu Ala Val Lys Lys Leu Leu Glu Ala Leu Ser His Leu Val
1 5 10 15

<210> 107
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of the artificial sequence: Oligonucleotide

<400> 107
tgttttcaac taccaggggc

20

<210> 108
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of the artificial sequence: Oligonucleotide

<400> 108
tgttggcttt ggcagagtcc

20

<210> 109
<211> 24
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of the artificial sequence: Oligonucleotide

<400> 109

gaggcagagt tcaggcttca ccga

24

<210> 110

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of the artificial sequence: Oligonucleotide

<400> 110

tggttggttt ggcagagtcc

20

<210> 111

<211> 56

<212> PRT

<213> Homo sapiens

<400> 111

Thr	Gly	Met	Asp	Met	Trp	Ser	Thr	Gln	Asp	Leu	Tyr	Asp	Asn	Pro	Val
1				5				10					15		
Thr	Ser	Val	Phe	Gln	Tyr	Glu	Gly	Leu	Trp	Arg	Ser	Cys	Val	Arg	Gln
			20					25					30		
Ser	Ser	Gly	Phe	Thr	Glu	Cys	Arg	Pro	Tyr	Phe	Thr	Ile	Leu	Gly	Leu
		35					40					45			
Pro	Ala	Met	Leu	Gln	Ala	Val	Arg								
	50						55								

<210> 112

<211> 53

<212> PRT

<213> Homo sapiens

<400> 112

Asp	Gln	Trp	Ser	Thr	Gln	Asp	Leu	Tyr	Asn	Asn	Pro	Val	Thr	Ala	Val
1				5				10					15		
Phe	Asn	Tyr	Gln	Gly	Leu	Trp	Arg	Ser	Cys	Val	Arg	Glu	Ser	Ser	Gly
			20					25					30		
Phe	Thr	Glu	Cys	Arg	Gly	Tyr	Phe	Thr	Leu	Leu	Gly	Leu	Pro	Ala	Met
		35					40					45			
Leu	Gln	Ala	Val	Arg											
	50														

<210> 113

<211> 14

<212> PRT

<213> Homo sapiens

<400> 113
 Ser Thr Gln Asp Leu Tyr Asn Asn Pro Val Thr Ala Val Phe
 1 5 10

<210> 114
 <211> 12
 <212> PRT
 <213> Homo sapiens

<400> 114
 Asp Met Trp Ser Thr Gln Asp Leu Tyr Asp Asn Pro
 1 5 10

<210> 115
 <211> 12
 <212> PRT
 <213> Homo sapiens

<400> 115
 Cys Arg Pro Tyr Phe Thr Ile Leu Gly Leu Pro Ala
 1 5 10

<210> 116
 <211> 13
 <212> PRT
 <213> Homo sapiens

<400> 116
 Thr Asn Phe Trp Met Ser Thr Ala Asn Met Tyr Thr Gly
 1 5 10

<210> 117
 <211> 816
 <212> DNA
 <213> Homo sapiens

<400> 117
 gccaggatca tgtccaccac cacatgccaa gtgggtggcgt tcctcctgtc catcctgggg 60
 ctggccggct gcatcgcggc caccgggatg gacatgtgga gcaccagga cctgtacgac 120
 aaccccgtca cctccgtggt ccagtacgaa gggctctgga ggagctgcgt gaggcagagt 180
 tcaggcttca ccgaatgcag gccctatttc accatcctgg gacttccagc catgctgcag 240
 gcagtgcgag ccctgatgat cgtaggcatc gtccctgggtg ccattggcct cctgggtatcc 300
 atctttgccc tgaaatgcat ccgcattggc agcatggagg actctgccaa agccaacatg 360
 aactgacct ccgggatcat gttcattgtc tcagggtcttt gtgcaattgc tggagtgtct 420
 gtgtttgcca acatgctggt gactaacttc tggatgtcca cagctaacat gtacaccggc 480
 atgggtggga tgggtgcagac tgttcagacc aggtacacat ttgggtgcggc tctgttcgtg 540
 ggctgggtcg ctggaggcct cacactaatt gggggtgtga tgatgtgcat cgcctgccgg 600
 ggctggcac cagaagaaac caactacaaa gccgtttctt atcatgcctc aggccacagt 660
 gttgcctaca agcctggagg cttcaaggcc agcactggct ttgggtccaa caccaaaaac 720
 aagaagatat acgatggagg tgcccgaca gaggacgagg tacaatctta tccttccaag 780
 cacgactatg tgtaatgctc taagacctct cagcac 816

<210> 118
<211> 261
<212> PRT
<213> Homo sapiens

<400> 118
Met Ser Thr Thr Thr Cys Gln Val Val Ala Phe Leu Leu Ser Ile Leu
1 5 10 15
Gly Leu Ala Gly Cys Ile Ala Ala Thr Gly Met Asp Met Trp Ser Thr
20 25 30
Gln Asp Leu Tyr Asp Asn Pro Val Thr Ser Val Phe Gln Tyr Glu Gly
35 40 45
Leu Trp Arg Ser Cys Val Arg Gln Ser Ser Gly Phe Thr Glu Cys Arg
50 55 60
Pro Tyr Phe Thr Ile Leu Gly Leu Pro Ala Met Leu Gln Ala Val Arg
65 70 75 80
Ala Leu Met Ile Val Gly Ile Val Leu Gly Ala Ile Gly Leu Leu Val
85 90 95
Ser Ile Phe Ala Leu Lys Cys Ile Arg Ile Gly Ser Met Glu Asp Ser
100 105 110
Ala Lys Ala Asn Met Thr Leu Thr Ser Gly Ile Met Phe Ile Val Ser
115 120 125
Gly Leu Cys Ala Ile Ala Gly Val Ser Val Phe Ala Asn Met Leu Val
130 135 140
Thr Asn Phe Trp Met Ser Thr Ala Asn Met Tyr Thr Gly Met Gly Gly
145 150 155 160
Met Val Gln Thr Val Gln Thr Arg Tyr Thr Phe Gly Ala Ala Leu Phe
165 170 175
Val Gly Trp Val Ala Gly Gly Leu Thr Leu Ile Gly Gly Val Met Met
180 185 190
Cys Ile Ala Cys Arg Gly Leu Ala Pro Glu Glu Thr Asn Tyr Lys Ala
195 200 205
Val Ser Tyr His Ala Ser Gly His Ser Val Ala Tyr Lys Pro Gly Gly
210 215 220
Phe Lys Ala Ser Thr Gly Phe Gly Ser Asn Thr Lys Asn Lys Lys Ile
225 230 235 240
Tyr Asp Gly Gly Ala Arg Thr Glu Asp Glu Val Gln Ser Tyr Pro Ser
245 250 255
Lys His Asp Tyr Val
260

<210> 119
<211> 227
<212> DNA
<213> Homo sapiens

<400> 119
gccaggatca tgtccaccac cacatgccaa gtggtggcgt tcctcctgtc catcctgggg 60
ctggccggct gcatcgcggc caccgggatg gacatgtgga gcaccagga cctgtacgac 120
aaccctgtca cctccgtgtt ccagtacgaa gggctctgga ggagctgcgt gaggcagagt 180
tcaggcttca ccgaatgcag gccctatttc accatcctgg gacttcc 227

<210> 120
<211> 69

<212> PRT
 <213> Homo sapiens

<400> 120
 Met Ser Thr Thr Thr Cys Gln Val Val Ala Phe Leu Leu Ser Ile Leu
 1 5 10 15
 Gly Leu Ala Gly Cys Ile Ala Ala Thr Gly Met Asp Met Trp Ser Thr
 20 25 30
 Gln Asp Leu Tyr Asp Asn Pro Val Thr Ser Val Phe Gln Tyr Glu Gly
 35 40 45
 Leu Trp Arg Ser Cys Val Arg Gln Ser Ser Gly Phe Thr Glu Cys Arg
 50 55 60
 Pro Tyr Phe Thr Ile
 65

<210> 121
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 121
 aatgagagga aagagaaaac 20

<210> 122
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 122
 atggtagaag agtaggcaat 20

<210> 123
 <211> 15
 <212> PRT
 <213> Homo sapiens

<400> 123
 Glu Lys Trp Asn Leu His Lys Arg Ile Ala Leu Lys Met Val Cys
 1 5 10 15

<210> 124
 <211> 11
 <212> PRT
 <213> Homo sapiens

<400> 124
 Cys Leu Gly Phe Asn Phe Lys Glu Met Phe Lys
 1 5 10

<210> 125
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 125
 taatgatgaa ccctacactg agc 23

<210> 126
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 126
 atggacaaat gccctacctt 20

<210> 127
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 127
 agtgctggaa ggatgtgcgt gt 22

<210> 128
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 128
 ttgaggtggt tggtgggttt 20

<210> 129
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>

<223> Description of the artificial sequence: Oligonucleotide

<400> 129

agatgtgctg aggctgtaga

20

<210> 130

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of the artificial sequence: Oligonucleotide

<400> 130

atgaaggttg attatttgag

20

<210> 131

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of the artificial sequence: Oligonucleotide

<400> 131

agccgcatag tcccttaccc tct

23

<210> 132

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of the artificial sequence: Oligonucleotide

<400> 132

gcagcagccc aaacaccaca

20

<210> 133

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of the artificial sequence: Oligonucleotide

<400> 133

ctgagccgag aggtggaatc

20

<210> 134
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of the artificial sequence: Oligonucleotide

<400> 134
 ctctctcgct tacactggaa 20

<210> 135
 <211> 14
 <212> PRT
 <213> Homo sapiens

<400> 135
 Gln Trp Gln Val Phe Gly Pro Asp Lys Pro Val Gln Ala Leu
 1 5 10

<210> 136
 <211> 15
 <212> PRT
 <213> Homo sapiens

<400> 136
 Ala Lys Trp Lys Gly Pro Gln Gly Gln Asp Leu Ser Thr Asp Ser
 1 5 10 15

<210> 137
 <211> 32
 <212> PRT
 <213> Homo sapiens

<400> 137
 Asn Met Leu Val Thr Asn Phe Trp Met Ser Thr Ala Asn Met Tyr Thr
 1 5 10 15
 Gly Met Gly Gly Met Val Gln Thr Val Gln Thr Arg Tyr Thr Phe Gly
 20 25 30

<210> 138
 <211> 2052
 <212> DNA
 <213> Homo sapiens

<400> 138
 gacagctgtg tctcgatgga gtagactctc agaacagcgc agtttgccct cgcctcacgc 60
 agagcctctc cgtggcttcc gcaccttgag cattaggcca gttctcctct tctctctaat 120
 ccacccgtca cctctcctgt catccgtttc catgccgtga ggtccattca cagaacacat 180
 ccatggctct catgctcagt ttggttctga gtctcctcaa gctgggatca gggcagtggc 240
 aggtggttgg gccagacaag cctgtccagg ccttggtggg ggaggacgca gcattctcct 300
 gtttctgtc tcctaagacc aatgcagagg ccatggaagt gcggttcttc aggggccagt 360

```

tctctagcgt ggtccacctc tacagggacg ggaaggacca gccatztatg cagatgccac 420
agtatcaagg caggacaaaa ctggtgaagg attctattgc ggaggggcgc atctctctga 480
ggctggaaaa cattactgtg ttggatgctg gcctctatgg gtgcaggatt agttcccagt 540
cttactacca gaaggccatc tgggagctac aggtgtcagc actgggctca gttcctctca 600
tttccatcac gggatatgtt gatagagaca tccagctact ctgtcagtcc tcgggctggt 660
tcccccggcc cacagcgaag tggaaaggtc cacaaggaca ggatttgtcc acagactcca 720
ggacaaacag agacatgcat ggctgtttg atgtggagat ctctctgacc gtccaagaga 780
acgccgggag catatcctgt tccatgcggc atgctcatct gagccgagag gtggaatcca 840
gggtacagat aggagatacc tttttcgagc ctatatcgtg gcacctggct accaaagtac 900
tgggaatact ctgctgtggc ctattttttg gcattgttgg actgaagatt ttcttctcca 960
aattccagtg gaaaatccag gcggaactgg actggagaag aaagcacgga caggcagaat 1020
tgagagacgc ccggaacac gcagtggagg tgactctgga tccagagacg gctcaccoga 1080
agctctgcgt ttctgatctg aaaactgtaa cccatagaaa agctccccag gaggtgcctc 1140
actctgagaa gagatttaca aggaagagtg tgggtggcttc tcagagtttc caagcaggga 1200
aacattactg ggaggtggac ggaggacaca ataaaagggt gcgcgtggga gtgtgccggg 1260
atgatgtgga caggaggaag gagtacgtga ctttgtctcc cgatcatggg tactgggtcc 1320
tcagactgaa tggagaacat ttgtatttca cattaaatcc ccgttttata agcgtcttcc 1380
ccaggacccc acctacaaaa ataggggtct tcctggacta tgagtgtggg accatctcct 1440
tcttcaacat aatgaccag tcccttattt ataccctgac atgtcggttt gaaggcttat 1500
tgaggcccta cattgagtat ccgtcctata atgagcaaaa tggaaactccc atagtcatct 1560
gcccagtcac ccaggaatca gagaaagagg cctcttggca aagggcctct gcaatcccag 1620
agacaagcaa cagtgagtcc tcctcacagg caaccacgcc ctctctcccc aggggtgaaa 1680
tgtaggatga atcacatccc acattcttct ttagggatat taaggctctct ctcccagatc 1740
caaagtcccg cagcagccgg ccaagggtggc ttccagatga agggggactg gcctgtccac 1800
atgggagtca ggtgtcatgg ctgccctgag ctgggaggga agaaggctga cattacattt 1860
agtttgctct cactccatct ggctaagtga tcttgaaata ccacctctca ggtgaagaac 1920
cgtcaggaat tcccatctca caggctgtgg tgtagattaa gtagacaagg aatgtgaata 1980
atgcttagat cttattgatg acagagtgtg tcctaattgt ttgttcatta tattacactt 2040
tcagtaaaaa aa 2052

```

```

<210> 139
<211> 500
<212> PRT
<213> Homo sapiens

```

```

<400> 139
Met Ala Leu Met Leu Ser Leu Val Leu Ser Leu Leu Lys Leu Gly Ser
1          5          10          15
Gly Gln Trp Gln Val Phe Gly Pro Asp Lys Pro Val Gln Ala Leu Val
20          25          30
Gly Glu Asp Ala Ala Phe Ser Cys Phe Leu Ser Pro Lys Thr Asn Ala
35          40          45
Glu Ala Met Glu Val Arg Phe Phe Arg Gly Gln Phe Ser Ser Val Val
50          55          60
His Leu Tyr Arg Asp Gly Lys Asp Gln Pro Phe Met Gln Met Pro Gln
65          70          75          80
Tyr Gln Gly Arg Thr Lys Leu Val Lys Asp Ser Ile Ala Glu Gly Arg
85          90          95
Ile Ser Leu Arg Leu Glu Asn Ile Thr Val Leu Asp Ala Gly Leu Tyr
100         105         110
Gly Cys Arg Ile Ser Ser Gln Ser Tyr Tyr Gln Lys Ala Ile Trp Glu
115         120         125
Leu Gln Val Ser Ala Leu Gly Ser Val Pro Leu Ile Ser Ile Thr Gly
130         135         140

```

```

Tyr Val Asp Arg Asp Ile Gln Leu Leu Cys Gln Ser Ser Gly Trp Phe
145                               150               155               160
Pro Arg Pro Thr Ala Lys Trp Lys Gly Pro Gln Gly Gln Asp Leu Ser
                               165               170               175
Thr Asp Ser Arg Thr Asn Arg Asp Met His Gly Leu Phe Asp Val Glu
                               180               185               190
Ile Ser Leu Thr Val Gln Glu Asn Ala Gly Ser Ile Ser Cys Ser Met
                               195               200               205
Arg His Ala His Leu Ser Arg Glu Val Glu Ser Arg Val Gln Ile Gly
                               210               215               220
Asp Thr Phe Phe Glu Pro Ile Ser Trp His Leu Ala Thr Lys Val Leu
225                               230               235               240
Gly Ile Leu Cys Cys Gly Leu Phe Phe Gly Ile Val Gly Leu Lys Ile
                               245               250               255
Phe Phe Ser Lys Phe Gln Trp Lys Ile Gln Ala Glu Leu Asp Trp Arg
                               260               265               270
Arg Lys His Gly Gln Ala Glu Leu Arg Asp Ala Arg Lys His Ala Val
                               275               280               285
Glu Val Thr Leu Asp Pro Glu Thr Ala His Pro Lys Leu Cys Val Ser
                               290               295               300
Asp Leu Lys Thr Val Thr His Arg Lys Ala Pro Gln Glu Val Pro His
305                               310               315               320
Ser Glu Lys Arg Phe Thr Arg Lys Ser Val Val Ala Ser Gln Ser Phe
                               325               330               335
Gln Ala Gly Lys His Tyr Trp Glu Val Asp Gly Gly His Asn Lys Arg
                               340               345               350
Trp Arg Val Gly Val Cys Arg Asp Asp Val Asp Arg Arg Lys Glu Tyr
                               355               360               365
Val Thr Leu Ser Pro Asp His Gly Tyr Trp Val Leu Arg Leu Asn Gly
                               370               375               380
Glu His Leu Tyr Phe Thr Leu Asn Pro Arg Phe Ile Ser Val Phe Pro
385                               390               395               400
Arg Thr Pro Pro Thr Lys Ile Gly Val Phe Leu Asp Tyr Glu Cys Gly
                               405               410               415
Thr Ile Ser Phe Phe Asn Ile Asn Asp Gln Ser Leu Ile Tyr Thr Leu
                               420               425               430
Thr Cys Arg Phe Glu Gly Leu Leu Arg Pro Tyr Ile Glu Tyr Pro Ser
                               435               440               445
Tyr Asn Glu Gln Asn Gly Thr Pro Ile Val Ile Cys Pro Val Thr Gln
450                               455               460
Glu Ser Glu Lys Glu Ala Ser Trp Gln Arg Ala Ser Ala Ile Pro Glu
465                               470               475               480
Thr Ser Asn Ser Glu Ser Ser Ser Gln Ala Thr Thr Pro Phe Leu Pro
                               485               490               495
Arg Gly Glu Met
                               500

```

<210> 140

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of the artificial sequence: Oligonucleotide

<400> 140
tccaaattcc agtggaaaat c

21

<210> 141
<211> 21
<212> DNA
<213> Artificial Sequence
<220>
<223> Description of the artificial sequence: Oligonucleotide

<400> 141
ccacactcat agtccaggaa g

21

<210> 142
<211> 116
<212> PRT
<213> Homo sapiens

<400> 142
Ala Leu Met Ile Val Gly Ile Val Leu Gly Ala Ile Gly Leu Leu Val
1 5 10 15
Ser Ile Phe Ala Leu Lys Cys Ile Arg Ile Gly Ser Met Glu Asp Ser
20 25 30
Ala Lys Ala Asn Met Thr Leu Thr Ser Gly Ile Met Phe Ile Val Ser
35 40 45
Gly Leu Cys Ala Ile Ala Gly Val Ser Val Phe Ala Asn Met Leu Val
50 55 60
Thr Asn Phe Trp Met Ser Thr Ala Asn Met Tyr Thr Gly Met Gly Gly
65 70 75 80
Met Val Gln Thr Val Gln Thr Arg Tyr Thr Phe Gly Ala Ala Leu Phe
85 90 95
Val Gly Trp Val Ala Gly Gly Leu Thr Leu Ile Gly Gly Val Met Met
100 105 110
Cys Ile Ala Cys
115

<210> 143
<211> 24
<212> PRT
<213> Homo sapiens

<400> 143
Arg Ile Gly Ser Met Glu Asp Ser Ala Lys Ala Asn Met Thr Leu Thr
1 5 10 15
Ser Gly Ile Met Phe Ile Val Ser
20

<210> 144
<211> 8
<212> PRT
<213> Homo sapiens

<400> 144

Ala Lys Ala Asn Met Thr Leu Thr
1 5

<210> 145

<211> 14

<212> PRT

<213> Homo sapiens

<400> 145

Met Glu Asp Ser Ala Lys Ala Asn Met Thr Leu Thr Ser Gly
1 5 10

<210> 146

<211> 14

<212> PRT

<213> Homo sapiens

<400> 146

Met Glu Asp Ser Ala Lys Ala Asp Met Thr Leu Thr Ser Gly
1 5 10

<210> 147

<211> 9

<212> PRT

<213> Homo sapiens

<400> 147

Ser Ala Lys Ala Asp Met Thr Leu Thr
1 5

<210> 148

<211> 9

<212> PRT

<213> Homo sapiens

<400> 148

Ala Lys Ala Asp Met Thr Leu Thr Leu
1 5

<210> 149

<211> 53

<212> PRT

<213> Homo sapiens

<400> 149

Asp Gln Trp Ser Thr Gln Asp Leu Tyr Asp Asn Pro Val Thr Ala Val
1 5 10 15
Phe Asn Tyr Gln Gly Leu Trp Arg Ser Cys Val Arg Glu Ser Ser Gly
20 25 30

Phe Thr Glu Cys Arg Gly Tyr Phe Thr Leu Leu Gly Leu Pro Ala Met
35 40 45
Leu Gln Ala Val Arg
50

<210> 150
<211> 14
<212> PRT
<213> Homo sapiens

<400> 150
Ser Thr Gln Asp Leu Tyr Asp Asn Pro Val Thr Ala Val Phe
1 5 10

Claims

1. A pharmaceutical composition, comprising an agent
which inhibits expression or activity of a tumor-
associated antigen, said tumor-associated antigen
having a sequence encoded by a nucleic acid which
is selected from the group consisting of:
(a) a nucleic acid which comprises a nucleic acid
sequence selected from the group consisting of SEQ
ID NOs: 1-8, 41-44, 51-59, 84, 117, 119, and 138,
a part or derivative thereof,
(b) a nucleic acid which hybridizes with the
nucleic acid of (a) under stringent conditions,
(c) a nucleic acid which is degenerate with
respect to the nucleic acid of (a) or (b), and
(d) a nucleic acid which is complementary to the
nucleic acid of (a), (b) or (c).
2. A pharmaceutical composition, comprising an agent
with tumor-inhibiting activity, which is selective
for cells expressing or abnormally expressing a
tumor-associated antigen, said tumor-associated
antigen having a sequence encoded by a nucleic
acid which is selected from the group consisting
of:
(a) a nucleic acid which comprises a nucleic acid
sequence selected from the group consisting of SEQ
ID NOs: 1-8, 41-44, 51-59, 84, 117, 119, and 138,
a part or derivative thereof,
(b) a nucleic acid which hybridizes with the
nucleic acid of (a) under stringent conditions,
(c) a nucleic acid which is degenerate with
respect to the nucleic acid of (a) or (b), and
(d) a nucleic acid which is complementary to the
nucleic acid of (a), (b) or (c).
3. The pharmaceutical composition as claimed in claim
2, in which the agent causes induction of cell

death, reduction in cell growth, damage to the cell membrane or secretion of cytokines.

4. The pharmaceutical composition as claimed in claim
5 1 or 2, in which the agent is an antisense nucleic acid which hybridizes selectively with the nucleic acid coding for the tumor-associated antigen.
5. The pharmaceutical composition as claimed in claim
10 1 or 2, in which the agent is an antibody which binds selectively to the tumor-associated antigen.
6. The pharmaceutical composition as claimed in claim
15 2, in which the agent is a complement-activating antibody which binds selectively to the tumor-associated antigen.
7. A pharmaceutical composition, comprising an agent
20 which, when administered, selectively increases the amount of complexes between an HLA molecule and a tumor-associated antigen or a part thereof, said tumor-associated antigen having a sequence encoded by a nucleic acid which is selected from the group consisting of:
25 (a) a nucleic acid which comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 1-8, 41-44, 51-59, 84, 117, 119, and 138, a part or derivative thereof,
30 (b) a nucleic acid which hybridizes with the nucleic acid of (a) under stringent conditions,
(c) a nucleic acid which is degenerate with respect to the nucleic acid of (a) or (b), and
(d) a nucleic acid which is complementary to the nucleic acid of (a), (b) or (c).
35
8. The pharmaceutical composition as claimed in claim 7, in which the agent comprises one or more components selected from the group consisting of:

- (i) the tumor-associated antigen or a part thereof,
- (ii) a nucleic acid which codes for the tumor-associated antigen or a part thereof,
- 5 (iii) a host cell which expresses the tumor-associated antigen or a part thereof, and
- (iv) isolated complexes between the tumor-associated antigen or a part thereof and an HLA molecule.

10

9. The pharmaceutical composition as claimed in claim 1, 2 or 7, in which the agent comprises two or more agents which in each case selectively inhibit expression or activity of different tumor-associated antigens, which are in each case selective for cells expressing different tumor-associated antigens or which increase the amount of complexes between HLA molecules and different tumor-associated antigens or parts thereof, with at least one of said tumor-associated antigens having a sequence encoded by a nucleic acid which is selected from the group consisting of:

15

20

25

30

- (a) a nucleic acid which comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 1-8, 41-44, 51-59, 84, 117, 119, and 138, a part or derivative thereof,
- (b) a nucleic acid which hybridizes with the nucleic acid of (a) under stringent conditions,
- (c) a nucleic acid which is degenerate with respect to the nucleic acid of (a) or (b), and
- (d) a nucleic acid which is complementary to the nucleic acid of (a), (b) or (c).

10. A pharmaceutical composition, comprising one or more components selected from the group consisting of:

35

- (i) a tumor-associated antigen or a part thereof,
- (ii) a nucleic acid which codes for a tumor-

- associated antigen or a part thereof,
(iii) an antibody which binds to a tumor-associated antigen or a part thereof,
(iv) an antisense nucleic acid which hybridizes specifically with a nucleic acid coding for a tumor-associated antigen,
(v) a host cell which expresses a tumor-associated antigen or a part thereof, and
(vi) isolated complexes between a tumor-associated antigen or a part thereof and an HLA molecule, said tumor-associated antigen having a sequence encoded by a nucleic acid which is selected from the group consisting of:
(a) a nucleic acid which comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 1-8, 41-44, 51-59, 84, 117, 119, and 138, a part or derivative thereof,
(b) a nucleic acid which hybridizes with the nucleic acid of (a) under stringent conditions,
(c) a nucleic acid which is degenerate with respect to the nucleic acid of (a) or (b), and
(d) a nucleic acid which is complementary to the nucleic acid of (a), (b) or (c).
11. The pharmaceutical composition as claimed in claim 8 or 10, in which the nucleic acid of (ii) is present in an expression vector.
12. The pharmaceutical composition as claimed in claim 8 or 10, in which the nucleic acid of (ii) is functionally linked to a promoter.
13. The pharmaceutical composition as claimed in claim 8 or 10, in which the host cell secretes the tumor-associated antigen or the part thereof.
14. The pharmaceutical composition as claimed in claim 8 or 10, in which the host cell additionally

expresses an HLA molecule which binds to the tumor-associated antigen or the part thereof.

15. The pharmaceutical composition as claimed in claim
5 14, in which the host cell expresses the HLA molecule and/or the tumor-associated antigen or the part thereof in a recombinant manner.
16. The pharmaceutical composition as claimed in claim
10 14, in which the host cell expresses the HLA molecule endogenously.
17. The pharmaceutical composition as claimed in claim
15 8, 10, 14 or 16, in which the host cell is an antigen-presenting cell.
18. The pharmaceutical composition as claimed in claim
20 17, in which the antigen-presenting cell is a dendritic cell or a macrophage.
19. The pharmaceutical composition as claimed in any
of claims 8, 10 and 13-18, in which the host cell is nonproliferative.
20. The pharmaceutical composition as claimed in claim
25 5 or 10, in which the antibody is a monoclonal antibody.
21. The pharmaceutical composition as claimed in claim
30 5 or 10, in which the antibody is a chimeric or humanized antibody.
22. The pharmaceutical composition as claimed in claim
35 5 or 10, in which the antibody is a fragment of a natural antibody.
23. The pharmaceutical composition as claimed in claim
5 or 10, in which the antibody is coupled to a

therapeutic or diagnostic agent.

24. The pharmaceutical composition as claimed in claim
4 or 10, in which the antisense nucleic acid
5 comprises a sequence of 6-50 contiguous
nucleotides of the nucleic acid coding for the
tumor-associated antigen.
25. The pharmaceutical composition as claimed in any
10 of claims 8 and 10-13, in which the tumor-
associated antigen or the part thereof, provided
by said pharmaceutical composition, binds to MHC
molecules on the surface of cells which express an
abnormal amount of said tumor-associated antigen
15 or of a part thereof.
26. The pharmaceutical composition as claimed in claim
25, in which the binding causes a cytolytic
reaction and/or induces cytokine release.
20
27. The pharmaceutical composition as claimed in any
of claims 1-26, further comprising a
pharmaceutically acceptable carrier and/or an
adjuvant.
25
28. The pharmaceutical composition as claimed in claim
27, in which the adjuvant is saponin, GM-CSF, CpG,
cytokine or a chemokine.
- 30 29. The pharmaceutical composition as claimed in any
of claims 1-28, which may be used for the
treatment of a disease characterized by expression
or abnormal expression of a tumor-associated
antigen.
35
30. The pharmaceutical composition as claimed in claim
29, in which the disease is cancer.

31. The pharmaceutical composition as claimed in claim 29, in which the disease is a lung tumor, a breast tumor, a prostate tumor, a melanoma, a colon tumor, a gastric tumor, a pancreatic tumor, an ENT tumor, a renal cell carcinoma or a cervical carcinoma, a colon carcinoma or a mammary carcinoma.
32. The pharmaceutical composition as claimed in any of claims 1-31, in which the tumor-associated antigen comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 9-19, 45-48, 60-66, 85, 90-97, 100-102, 105, 106, 111-116, 118, 120, 123, 124, 135-137, 139, and 142-150, a part or derivative thereof.
33. A method of diagnosing a disease characterized by expression or abnormal expression of a tumor-associated antigen, which method comprises
- (i) detection of a nucleic acid which codes for the tumor-associated antigen or of a part thereof, and/or
 - (ii) detection of the tumor-associated antigen or of a part thereof, and/or
 - (iii) detection of an antibody to the tumor-associated antigen or of a part thereof and/or
 - (iv) detection of cytotoxic or T helper lymphocytes which are specific to the tumor-associated antigen or to a part thereof in a biological sample isolated from a patient, with said tumor-associated antigen having a sequence encoded by a nucleic acid which is selected from the group consisting of:
 - (a) a nucleic acid which comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 1-8, 41-44, 51-59, 84, 117, 119, and 138, a part or derivative thereof,
 - (b) a nucleic acid which hybridizes with the

nucleic acid of (a) under stringent conditions,
(c) a nucleic acid which is degenerate with
respect to the nucleic acid of (a) or (b), and
(d) a nucleic acid which is complementary to the
5 nucleic acid of (a), (b) or (c).

34. The method as claimed in claim 33, in which the
detection comprises
(i) contacting the biological sample with an
10 agent which binds specifically to the nucleic acid
coding for the tumor-associated antigen or to the
part thereof, to the tumor-associated antigen or
the part thereof, to the antibody or to the
cytotoxic or T helper lymphocytes, and
15 (ii) detecting the formation of a complex between
the agent and the nucleic acid or the part
thereof, the tumor-associated antigen or the part
thereof, the antibody or the cytotoxic or T helper
lymphocytes.

20

35. The method as claimed in claim 33 or 34, in which
the detection is compared to detection in a
comparable normal biological sample.

25 36. The method as claimed in any of claims 33-35, in
which the disease is characterized by expression
or abnormal expression of two or more different
tumor-associated antigens and in which detection
comprises detection of two or more nucleic acids
30 coding for said two or more different tumor-
associated antigens or of parts thereof, detection
of said two or more different tumor-associated
antigens or of parts thereof, detection of two or
more antibodies binding to said two or more
35 different tumor-associated antigens or to parts
thereof or detection of two or more cytotoxic or T
helper lymphocytes specific for said two or more
different tumor-associated antigens.

37. The method as claimed in any of claims 33-36, in which the nucleic acid or the part thereof is detected using a polynucleotide probe which hybridizes specifically to said nucleic acid or to said part thereof.
38. The method as claimed in claim 37, in which the polynucleotide probe comprises a sequence of 6-50 contiguous nucleotides of the nucleic acid coding for the tumor-associated antigen.
39. The method as claimed in any of claims 33-36, in which the nucleic acid or the part thereof is detected by selectively amplifying said nucleic acid or said part thereof.
40. The method as claimed in any of claims 33-36, in which the tumor-associated antigen to be detected or the part thereof are in a complex with an MHC molecule.
41. The method as claimed in claim 40, in which the MHC molecule is an HLA molecule.
42. The method as claimed in any of claims 33-36 and 40-41, in which the tumor-associated antigen or the part thereof is detected using an antibody binding specifically to said tumor-associated antigen or to said part thereof.
43. The method as claimed in any of claims 33-36, in which the antibody is detected using a protein or peptide binding specifically to said antibody.
44. A method for determining regression, course or onset of a disease characterized by expression or abnormal expression of a tumor-associated antigen,

which method comprises monitoring a sample from a patient who has said disease or is suspected of falling ill with said disease, with respect to one or more parameters selected from the group consisting of:

5

(i) the amount of nucleic acid which codes for the tumor-associated antigen or of a part thereof,
(ii) the amount of the tumor-associated antigen or of a part thereof,

10

(iii) the amount of antibodies which bind to the tumor-associated antigen or to a part thereof, and

(iv) the amount of cytolytic or cytokine-releasing T cells which are specific for a complex between the tumor-associated antigen or a part thereof and an MHC molecule, said tumor-associated antigen having a sequence encoded by a nucleic acid which is selected from the group consisting of:

15

(a) a nucleic acid which comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 1-8, 41-44, 51-59, 84, 117, 119, and 138, a part or derivative thereof,

20

(b) a nucleic acid which hybridizes with the nucleic acid of (a) under stringent conditions,

(c) a nucleic acid which is degenerate with respect to the nucleic acid of (a) or (b), and

25

(d) a nucleic acid which is complementary to the nucleic acid of (a), (b) or (c).

45. The method as claimed in claim 44, which comprises determining the parameter(s) in a first sample at a first point in time and in a further sample at a second point in time and in which the course of the disease is determined by comparing the two samples.

35

46. The method as claimed in claim 44 or 45, in which the disease is characterized by expression or abnormal expression of two or more different

tumor-associated antigens and in which monitoring comprises monitoring

- 5 (i) the amount of two or more nucleic acids which code for said two or more different tumor-associated antigens or of parts thereof,
- (ii) the amount of said two or more different tumor-associated antigens or of parts thereof,
- 10 (iii) the amount of two or more antibodies which bind to said two or more different tumor-associated antigens or to parts thereof, and/or
- (iv) the amount of two or more cytolytic or cytokine-releasing T cells which are specific for complexes between said two or more different tumor-associated antigens or of parts thereof and
- 15 MHC molecules.

47. The method as claimed in any of claims 44-46, in which the amount of the nucleic acid or of the part thereof is monitored using a polynucleotide

20 probe which hybridizes specifically to said nucleic acid or said part thereof.

48. The method as claimed in claim 47, in which the polynucleotide probe comprises a sequence of 6-50

25 contiguous nucleotides of the nucleic acid coding for the tumor-associated antigen.

49. The method as claimed in any of claims 44-46, in which the amount of the nucleic acid or of the part thereof is monitored by selectively

30 amplifying said nucleic acid or said part thereof.

50. The method as claimed in any of claims 44-46, in which the amount of the tumor-associated antigen or of the part thereof is monitored using an

35 antibody binding specifically to said tumor-associated antigen or said part thereof.

51. The method as claimed in any of claims 44-46, in which the amount of antibodies is monitored using a protein or peptide binding specifically to the antibody.
- 5
52. The method as claimed in any of claims 44-46, in which the amount of cytolytic or cytokine-releasing T cells is monitored using a cell presenting the complex between the tumor-associated antigen or the part thereof and an MHC molecule.
- 10
53. The method as claimed in any of claims 37-38, 42-43, 47-48 and 50-52, in which the polynucleotide probe, the antibody, the protein or peptide or the cell is labeled in a detectable manner.
- 15
54. The method as claimed in claim 53, in which the detectable marker is a radioactive marker or an enzymic marker.
- 20
55. The method as claimed in any of claims 33-54, in which the sample comprises body fluid and/or body tissue.
- 25
56. A method of treating a disease characterized by expression or abnormal expression of a tumor-associated antigen, which method comprises administration of a pharmaceutical composition as claimed in any of claims 1-32, said tumor-associated antigen having a sequence encoded by a nucleic acid which is selected from the group consisting of:
- 30
- 35 (a) a nucleic acid which comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 1-8, 41-44, 51-59, 84, 117, 119, and 138, a part or derivative thereof,

- (b) a nucleic acid which hybridizes with the nucleic acid of (a) under stringent conditions,
(c) a nucleic acid which is degenerate with respect to the nucleic acid of (a) or (b), and
5 (d) a nucleic acid which is complementary to the nucleic acid of (a), (b) or (c).
57. A method of treating, diagnosing or monitoring a disease characterized by expression or abnormal
10 expression of a tumor-associated antigen, which method comprises administering an antibody binding to said tumor-associated antigen or to a part thereof and coupled to a therapeutic or diagnostic agent, said tumor-associated antigen having a
15 sequence encoded by a nucleic acid which is selected from the group consisting of:
(a) a nucleic acid which comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 1-8, 41-44, 51-59, 84, 117, 119, and 138,
20 a part or derivative thereof,
(b) a nucleic acid which hybridizes with the nucleic acid of (a) under stringent conditions,
(c) a nucleic acid which is degenerate with respect to the nucleic acid of (a) or (b), and
25 (d) a nucleic acid which is complementary to the nucleic acid of (a), (b) or (c).
58. The method as claimed in claim 42, 50 or 57, in which the antibody is a monoclonal antibody.
30
59. The method as claimed in claim 42, 50 or 57, in which the antibody is a chimeric or humanized antibody.
- 35 60. The method as claimed in claim 42, 50 or 57, in which the antibody is a fragment of a natural antibody.

61. A method of treating a patient having a disease characterized by expression or abnormal expression of a tumor-associated antigen, which method comprises:
- 5 (i) removing a sample containing immunoreactive cells from said patient,
 - (ii) contacting said sample with a host cell expressing said tumor-associated antigen or a part thereof, under conditions which favor production
10 of cytolytic or cytokine-releasing T cells against said tumor-associated antigen or a part thereof, and
 - (iii) introducing the cytolytic or cytokine-releasing T cells into the patient in an amount
15 suitable for lysing cells expressing the tumor-associated antigen or a part thereof, said tumor-associated antigen having a sequence encoded by a nucleic acid which is selected from the group consisting of:
 - 20 (a) a nucleic acid which comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 1-8, 41-44, 51-59, 84, 117, 119, and 138, a part or derivative thereof,
 - (b) a nucleic acid which hybridizes with the
25 nucleic acid of (a) under stringent conditions,
 - (c) a nucleic acid which is degenerate with respect to the nucleic acid of (a) or (b), and
 - (d) a nucleic acid which is complementary to the nucleic acid of (a), (b) or (c).
 - 30
62. The method as claimed in claim 61, in which the host cell recombinantly expresses an HLA molecule binding to the tumor-associated antigen or to a part thereof.
- 35
63. The method as claimed in claim 62, in which the host cell endogenously expresses an HLA molecule binding to the tumor-associated antigen or to a

part thereof.

64. A method of treating a patient having a disease characterized by expression or abnormal expression of a tumor-associated antigen, which method comprises:
- (i) identifying a nucleic acid which is expressed by cells associated with said disease, said nucleic acid being selected from the group consisting of:
 - (a) a nucleic acid which comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 1-8, 41-44, 51-59, 84, 117, 119, and 138, a part or derivative thereof,
 - (b) a nucleic acid which hybridizes with the nucleic acid of (a) under stringent conditions,
 - (c) a nucleic acid which is degenerate with respect to the nucleic acid of (a) or (b), and
 - (d) a nucleic acid which is complementary to the nucleic acid of (a), (b) or (c),
 - (ii) transfecting a host cell with said nucleic acid or a part thereof,
 - (iii) culturing the transfected host cell for expression of said nucleic acid, and
 - (iv) introducing the host cells or an extract thereof into the patient in an amount suitable for increasing the immune response to the patient's cells associated with the disease.
65. The method as claimed in claim 64, which further comprises identifying an MHC molecule presenting the tumor-associated antigen or a part thereof, with the host cell expressing the identified MHC molecule and presenting the tumor-associated antigen or a part thereof.
66. The method as claimed in claim 64 or 65, in which the immune response comprises a B cell response or

a T cell response.

67. The method as claimed in claim 66, in which the immune response is a T cell response comprising production of cytolytic or cytokine-releasing T cells which are specific for the host cells presenting the tumor-associated antigen or a part thereof or specific for cells of the patient which express the tumor-associated antigen or a part thereof.
68. The method as claimed in any of claims 61-67, in which the host cells are nonproliferative.
69. A method of treating a disease characterized by expression or abnormal expression of a tumor-associated antigen, which method comprises:
- (i) identifying cells from the patient which express abnormal amounts of the tumor-associated antigen,
 - (ii) isolating a sample of said cells,
 - (iii) culturing said cells, and
 - (iv) introducing said cells into the patient in an amount suitable for triggering an immune response to the cells, said tumor-associated antigen having a sequence encoded by a nucleic acid which is selected from the group consisting of:
 - (a) a nucleic acid which comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 1-8, 41-44, 51-59, 84, 117, 119, and 138, a part or derivative thereof,
 - (b) a nucleic acid which hybridizes with the nucleic acid of (a) under stringent conditions,
 - (c) a nucleic acid which is degenerate with respect to the nucleic acid of (a) or (b), and
 - (d) a nucleic acid which is complementary to the nucleic acid of (a), (b) or (c).

70. The method as claimed in any of claims 33-69, in which the disease is cancer.
- 5 71. A method of inhibiting the development of cancer in a patient, which method comprises administering an effective amount of a pharmaceutical composition as claimed in any of claims 1-32.
- 10 72. The method as claimed in any of claims 33-71, in which the tumor-associated antigen comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 9-19, 45-48, 60-66, 85, 90-97, 100-102, 105, 106, 111-116, 118, 120, 123,
15 124, 135-137, 139, and 142-150, a part or derivative thereof.
73. A nucleic acid, selected from the group consisting of:
- 20 (a) a nucleic acid which comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 3-5, a part or derivative thereof,
(b) a nucleic acid which hybridizes with the nucleic acid of (a) under stringent conditions,
25 (c) a nucleic acid which is degenerate with respect to the nucleic acid of (a) or (b), and
(d) a nucleic acid which is complementary to the nucleic acid of (a), (b) or (c).
- 30 74. A nucleic acid, which codes for a protein or polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 10, 12-14, and 146-150, a part or derivative thereof.
- 35 75. A recombinant DNA or RNA molecule, which comprises a nucleic acid as claimed in claim 73 or 74.

76. The recombinant DNA molecule as claimed in claim 75, which is a vector.
77. The recombinant DNA molecule as claimed in claim 5 76, in which the vector is a viral vector or a bacteriophage.
78. The recombinant DNA molecule as claimed in any of claims 75-77, which further comprises expression control sequences controlling expression of the nucleic acid. 10
79. The recombinant DNA molecule as claimed in claim 78, in which the expression control sequences are homologous or heterologous to the nucleic acid. 15
80. A host cell, which comprises a nucleic acid as claimed in claim 73 or 74 or a recombinant DNA molecule as claimed in any of claims 75-79. 20
81. The host cell as claimed in claim 80, which further comprises a nucleic acid coding for an HLA molecule.
- 25 82. A protein or polypeptide, which is encoded by a nucleic acid as claimed in claim 73.
83. A protein or polypeptide, which comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 10, 12-14, 17-19, 90-97, 100-102, 30 105, 106, 111-116, 120, 123, 124, 135-137, and 142-150, a part or derivative thereof.
84. An immunogenic fragment of the protein or polypeptide as claimed in claim 82 or 83. 35
85. A fragment of the protein or polypeptide as claimed in claim 82 or 83, which binds to human

HLA receptor or human antibody.

86. An agent, which binds specifically to a protein or polypeptide or to a part thereof, said protein or polypeptide being encoded by a nucleic acid selected from the group consisting of:
- (a) a nucleic acid which comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 1-8, 41-44, 51-59, 84, 117, 119, and 138, a part or derivative thereof,
 - (b) a nucleic acid which hybridizes with the nucleic acid of (a) under stringent conditions,
 - (c) a nucleic acid which is degenerate with respect to the nucleic acid of (a) or (b), and
 - (d) a nucleic acid which is complementary to the nucleic acid of (a), (b) or (c).
87. The agent as claimed in claim 86, in which the protein or polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 9-19, 45-48, 60-66, 85, 90-97, 100-102, 105, 106, 111-116, 118, 120, 123, 124, 135-137, 139, and 142-150, a part or derivative thereof.
88. The agent as claimed in claim 86 or 87, which is an antibody.
89. The agent as claimed in claim 88, in which the antibody is a monoclonal, chimeric or humanized antibody or a fragment of an antibody.
90. An antibody, which binds selectively to a complex of:
- (i) a protein or polypeptide or a part thereof and
 - (ii) an MHC molecule to which said protein or polypeptide or said part thereof binds, with said antibody not binding to (i) or (ii) alone and said

protein or polypeptide being encoded by a nucleic acid selected from the group consisting of:

- 5 (a) a nucleic acid which comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 1-8, 41-44, 51-59, 84, 117, 119, and 138, a part or derivative thereof,
- (b) a nucleic acid which hybridizes with the nucleic acid of (a) under stringent conditions,
- 10 (c) a nucleic acid which is degenerate with respect to the nucleic acid of (a) or (b), and
- (d) a nucleic acid which is complementary to the nucleic acid of (a), (b) or (c).

- 15 91. The antibody as claimed in claim 90, in which the protein or polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 9-19, 45-48, 60-66, 85, 90-97, 100-102, 105, 106, 111-116, 118, 120, 123, 124, 135-137, 139, and 142-150, a part or derivative thereof.

20

92. The antibody as claimed in claim 90 or 91, which is a monoclonal, chimeric or humanized antibody or a fragment of an antibody.

- 25 93. A conjugate between an agent as claimed in any of claims 86-89 or an antibody as claimed in any of claims 90-92 and a therapeutic or diagnostic agent.

- 30 94. The conjugate as claimed in claim 93, in which the therapeutic or diagnostic agent is a toxin.

95. A kit for detecting expression or abnormal expression of a tumor-associated antigen, which
- 35 kit comprises agents for detection
- (i) of the nucleic acid which codes for the tumor-associated antigen or of a part thereof,
- (ii) of the tumor-associated antigen or of a part

thereof,

(iii) of antibodies which bind to the tumor-associated antigen or to a part thereof, and/or

5 (iv) of T cells which are specific for a complex between the tumor-associated antigen or a part thereof and an MHC molecule, said tumor-associated antigen having a sequence encoded by a nucleic acid which is selected from the group consisting of:

10 (a) a nucleic acid which comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 1-8, 41-44, 51-59, 84, 117, 119, and 138, a part or derivative thereof,

15 (b) a nucleic acid which hybridizes with the nucleic acid of (a) under stringent conditions,

(c) a nucleic acid which is degenerate with respect to the nucleic acid of (a) or (b), and

(d) a nucleic acid which is complementary to the nucleic acid of (a), (b) or (c).

20

96. The kit as claimed in claim 95, in which the agents for detection of the nucleic acid which codes for the tumor-associated antigen or of a part thereof are nucleic acid molecules for
25 selective amplification of said nucleic acid.

97. The kit as claimed in claim 96, in which the nucleic acid molecules for selective amplification of the nucleic acid comprise a sequence of 6-50
30 contiguous nucleotides of the nucleic acid which codes for the tumor-associated antigen.

98. A recombinant DNA molecule, comprising a promoter region which is derived from a nucleic acid
35 sequence selected from the group consisting of SEQ ID NOs: 1-8, 41-44, 51-59, 84, 117, 119, and 138.

99. Pharmaceutical composition as claimed in any of

claims 5, 10, and 20-23 or method as claimed in any of claims 42, 50, 53, and 57-60, in which the antibody is obtainable by immunizing using a protein, peptide or polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 9-19, 45-48, 60-66, 85, 90-97, 100-102, 105, 106, 111-116, 118, 120, 123, 124, 135-137, 139, and 142-150, a part or a derivative thereof.

10

100. Pharmaceutical composition as claimed in any of claims 5, 10, and 20-23 or method as claimed in any of claims 42, 50, 53, and 57-60, in which the selective binding of the antibody to the tumor-associated antigen involves the recognition of a glycosylation of the tumor-associated antigen.

15

101. Pharmaceutical composition as claimed in any of claims 5, 10, and 20-23 or method as claimed in any of claims 42, 50, 53, and 57-60, in which the antibody selectively binds to a specific glycosylation form of the tumor-associated antigen, preferably the non-glycosylated form.

20

Fig. 1

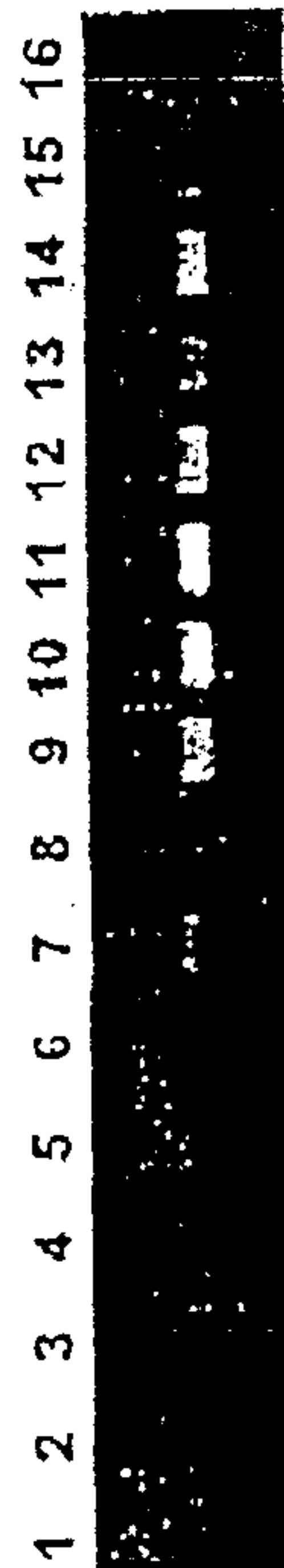


Fig. 2

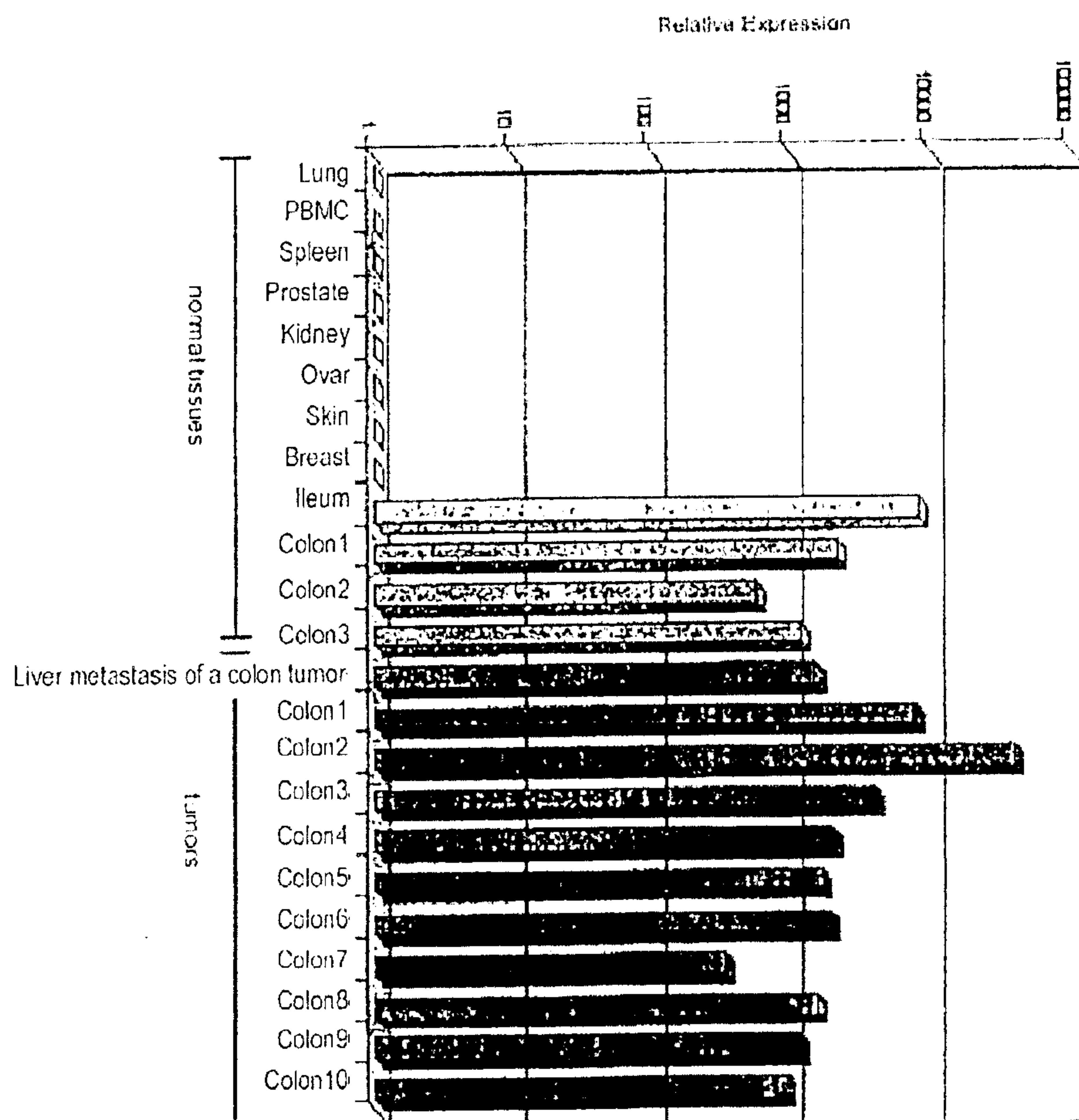


Fig. 3

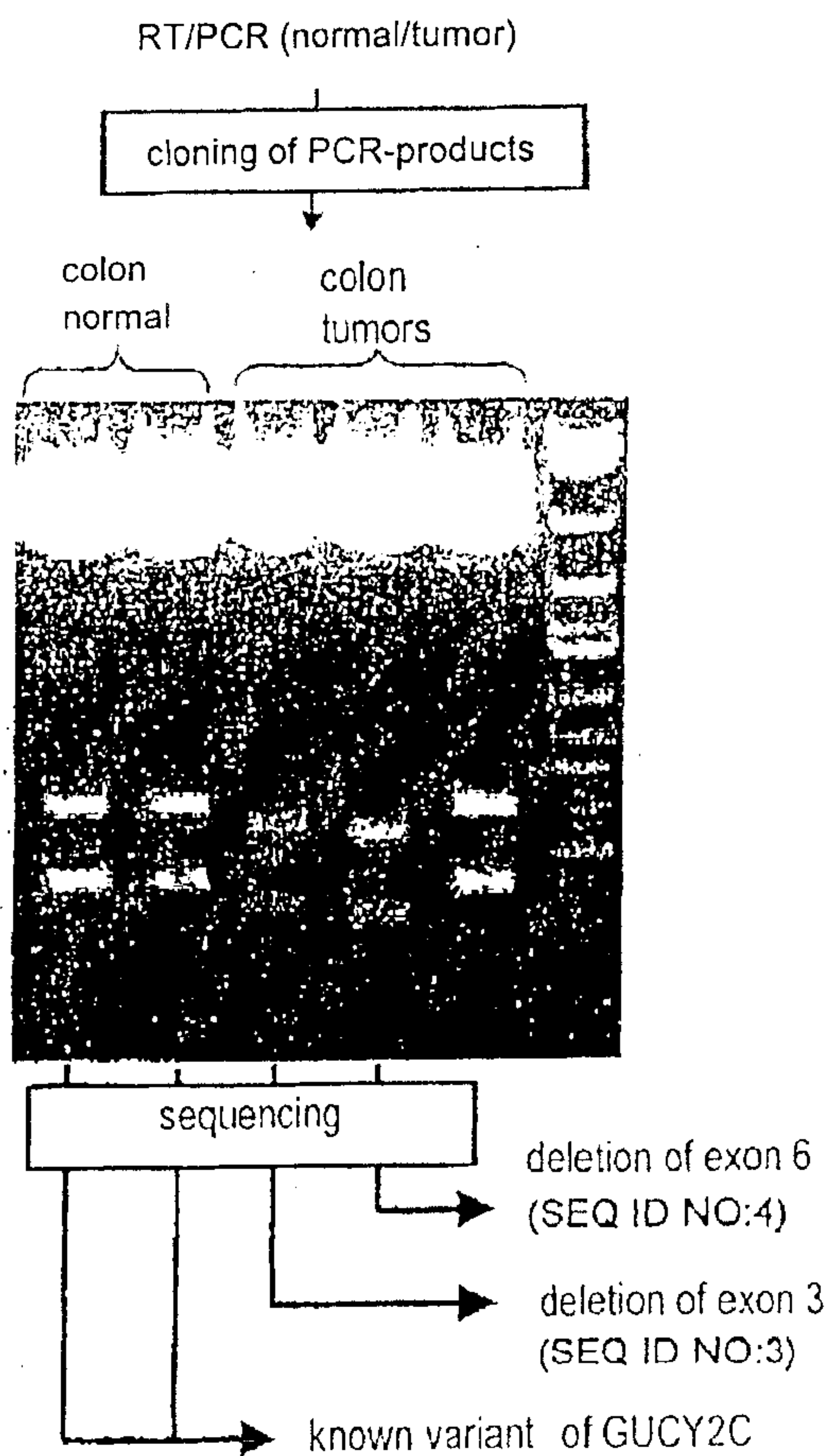
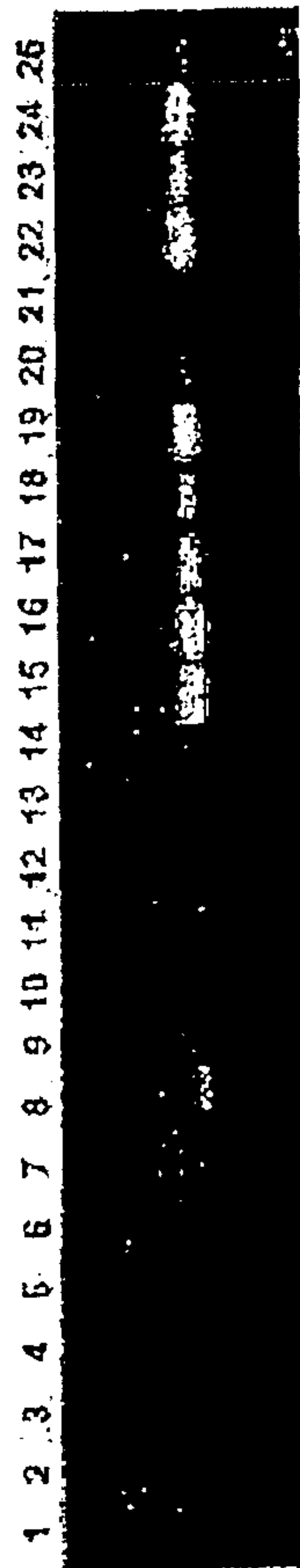


Fig. 4



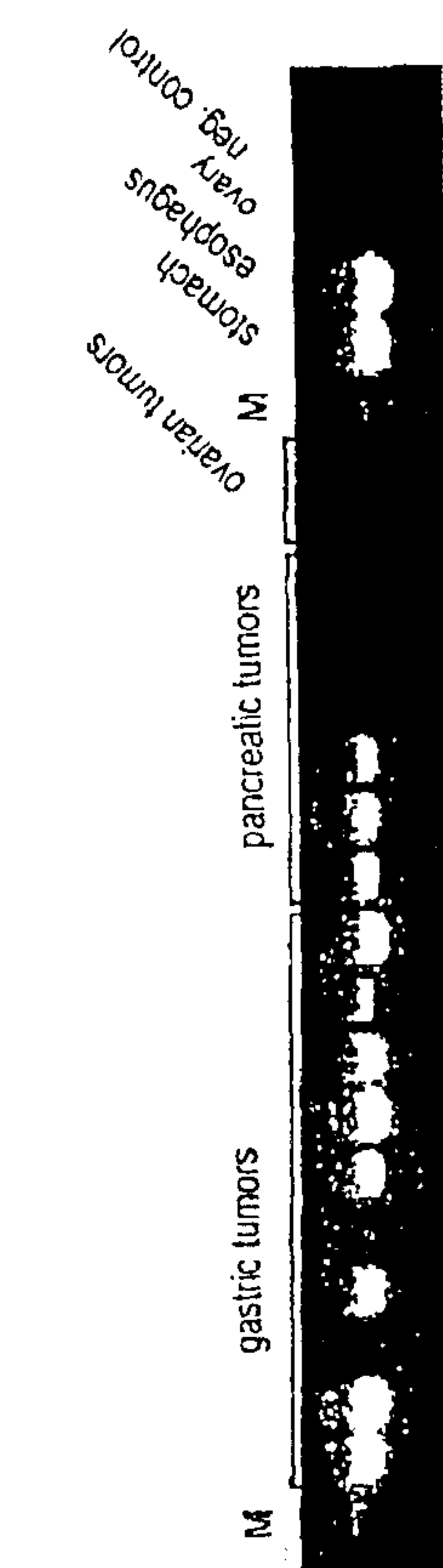


Fig. 5

Fig. 6

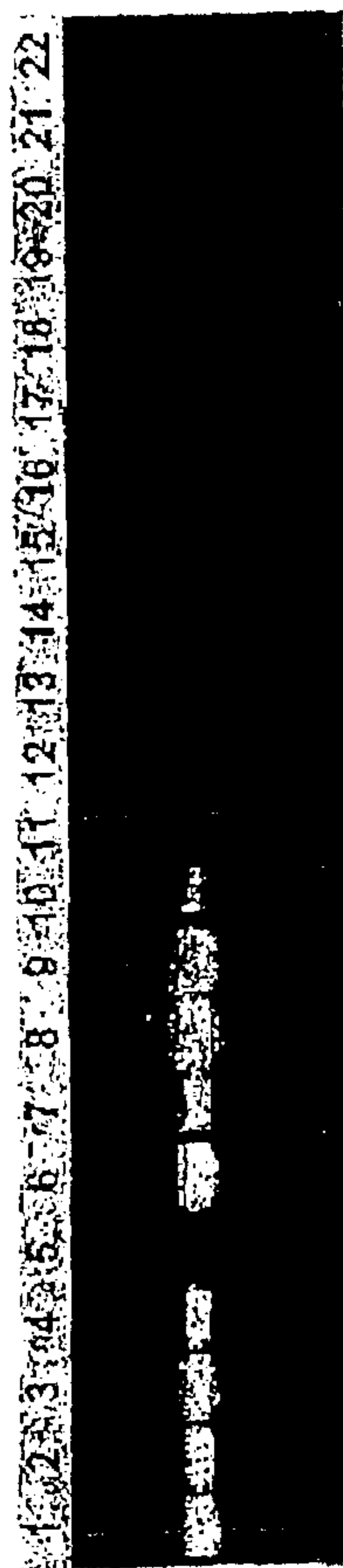


Fig. 7

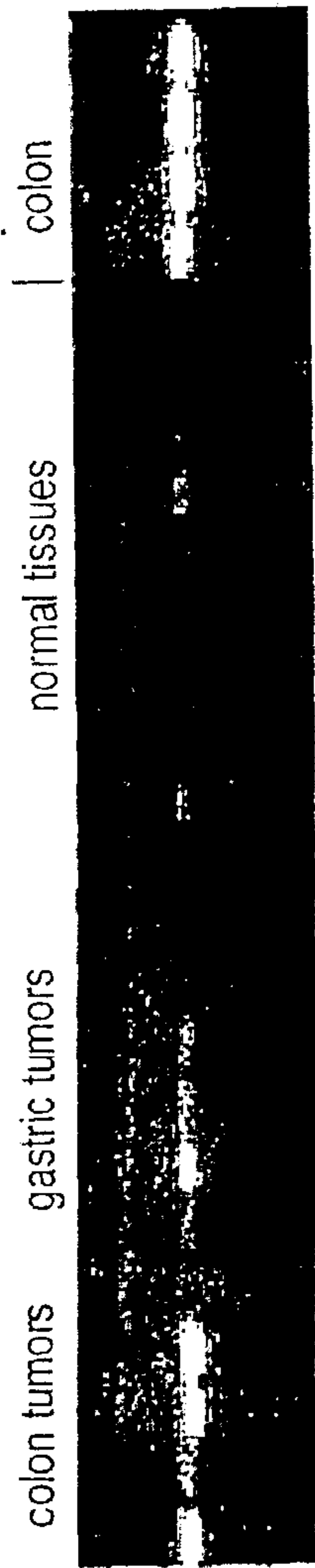


Fig. 8



Fig. 9

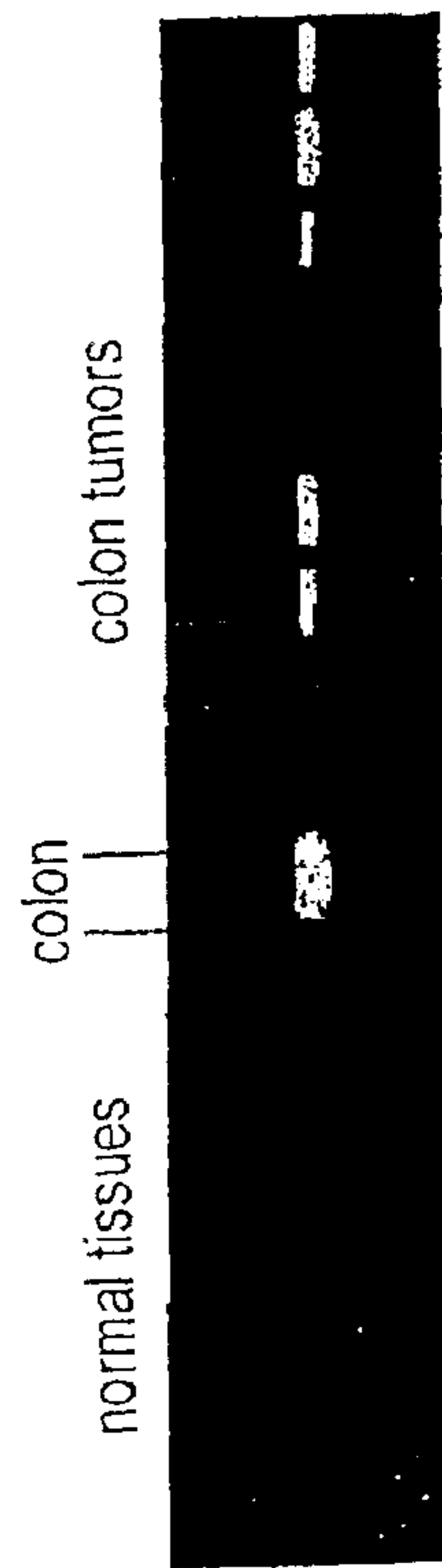


Fig. 10

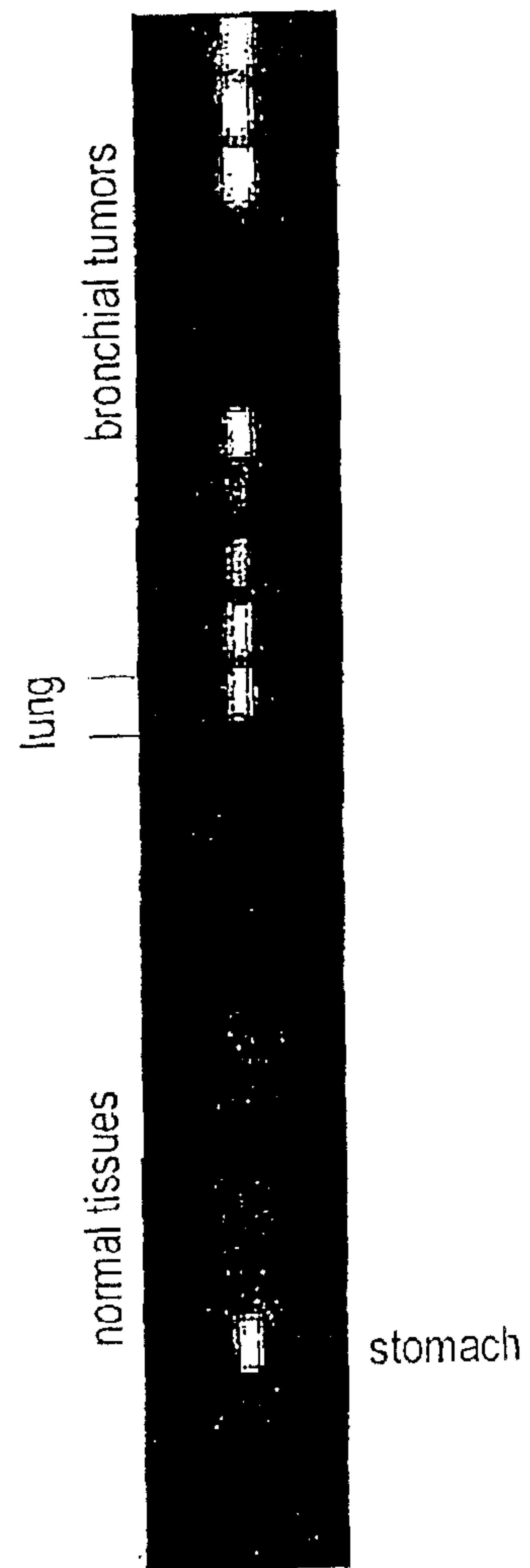


Fig. 11

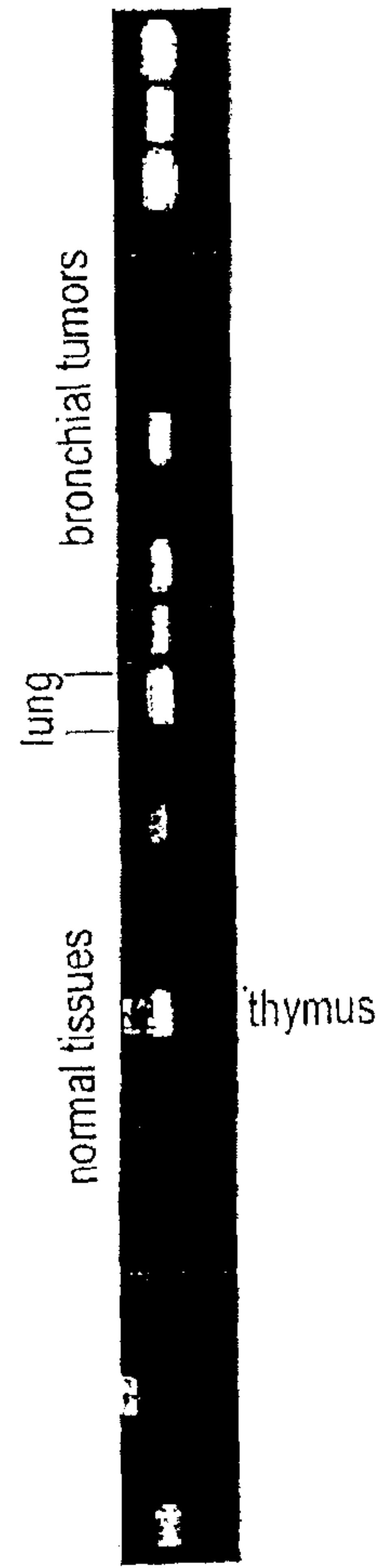


Fig. 12

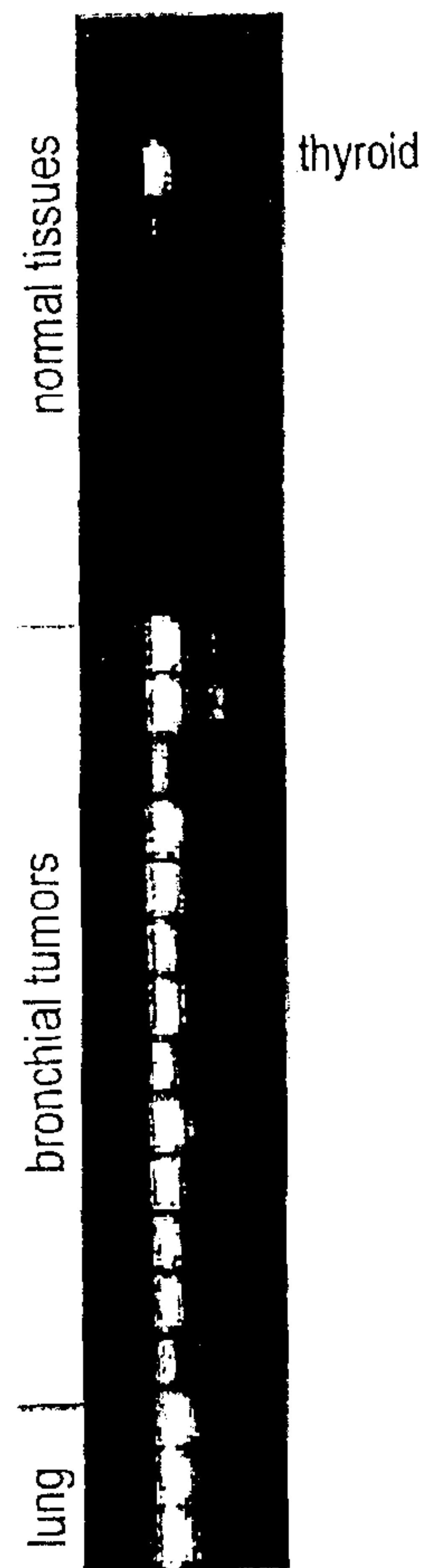


Fig. 13

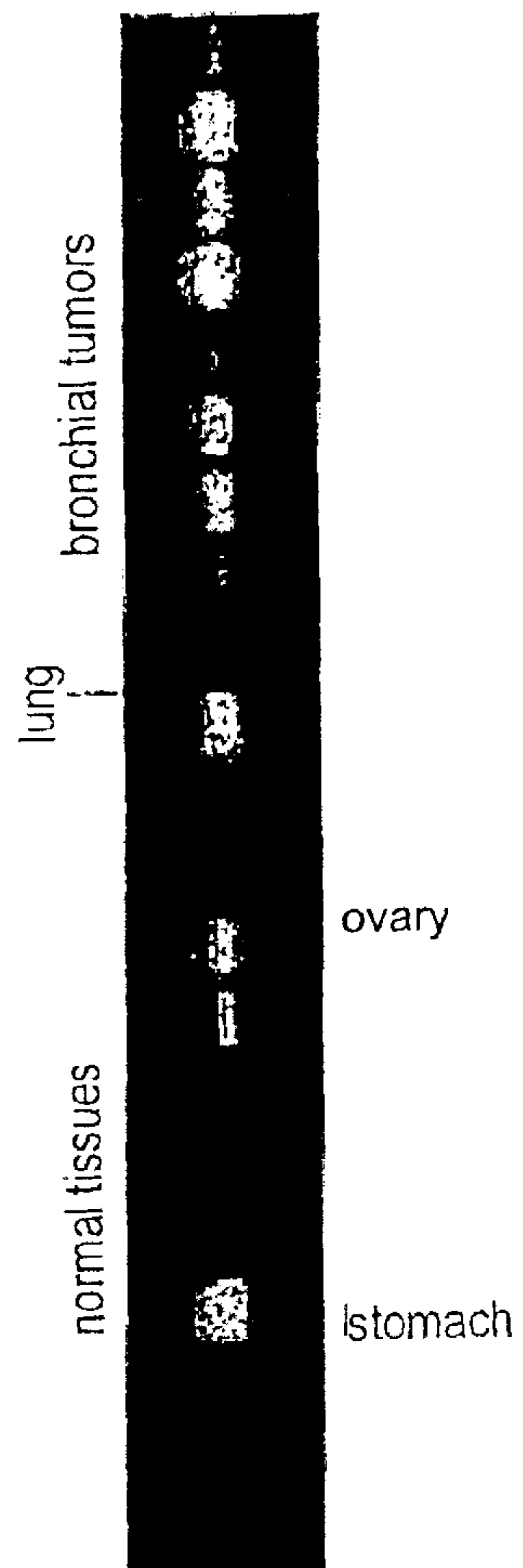


Fig. 14



Fig. 15

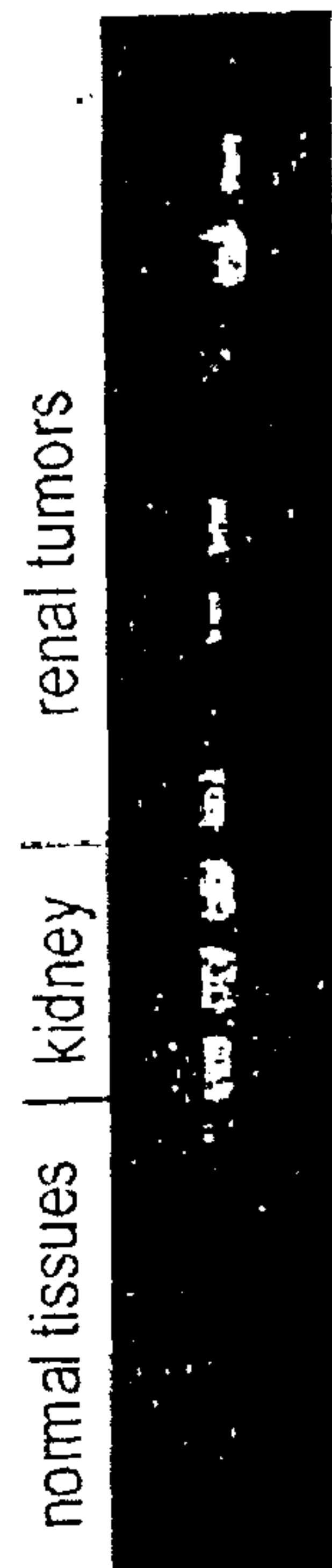
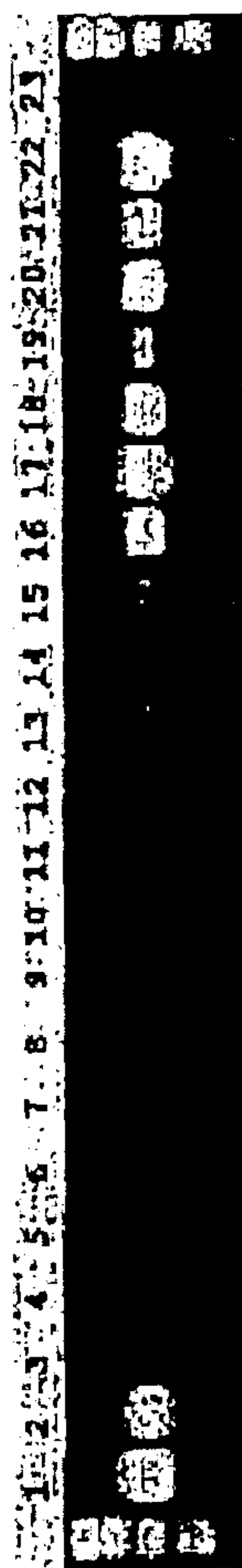


Fig. 16



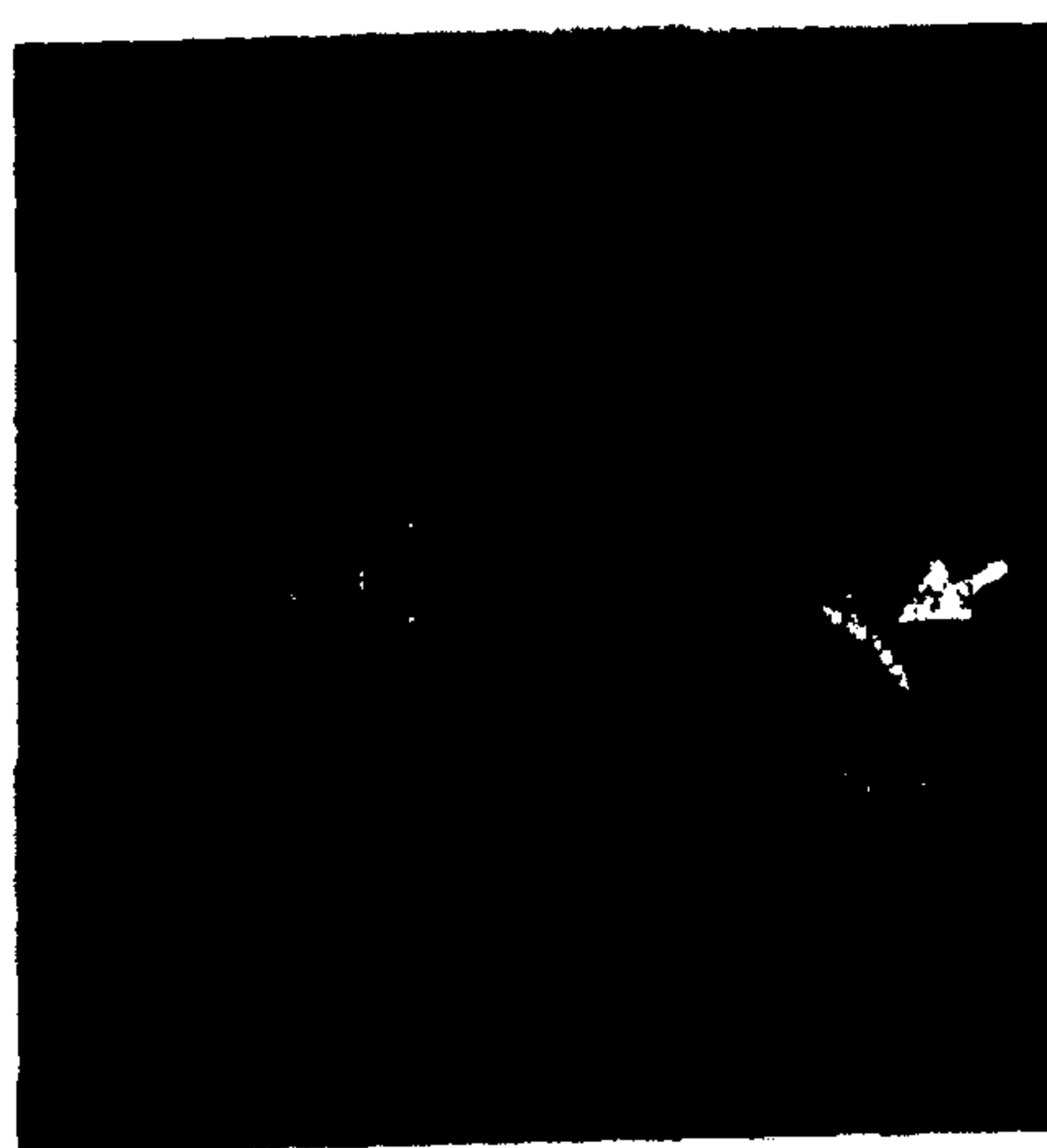
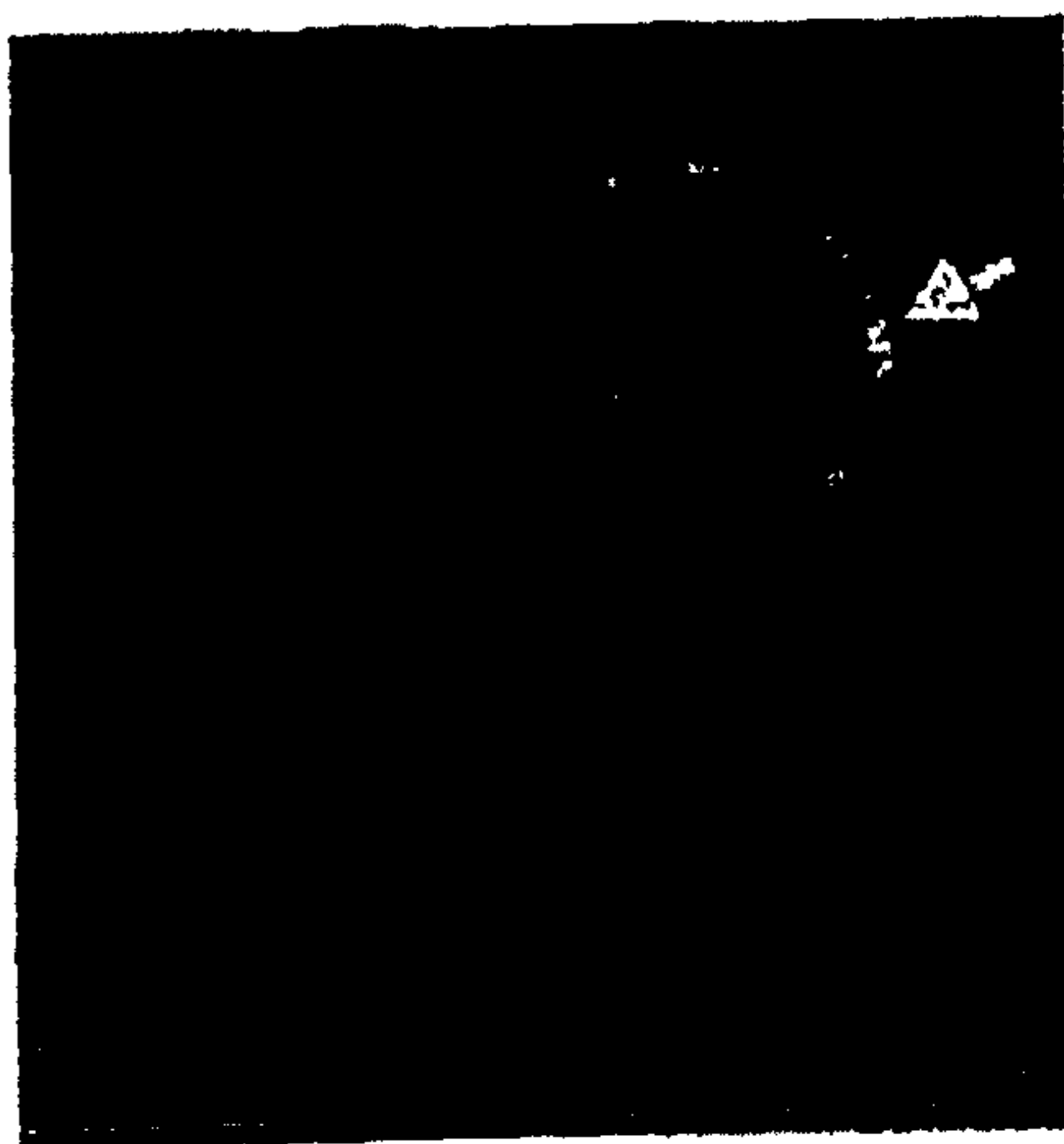
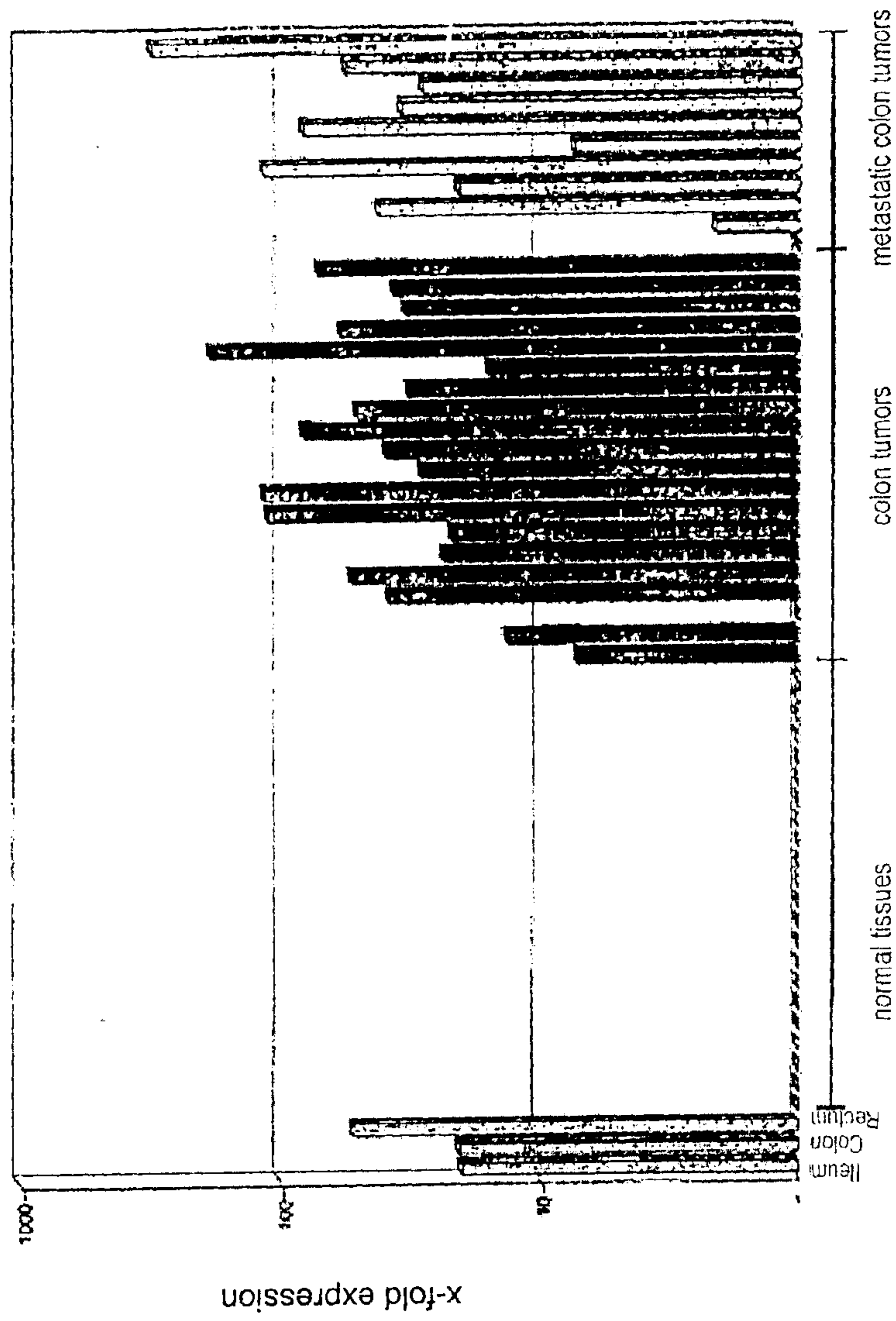


Fig. 17



18/62

cancer type	expression	%
colorectal, primary	19/20	95
colorectal, metastatic	14/15	93

Fig. 18B

Fig. 19A

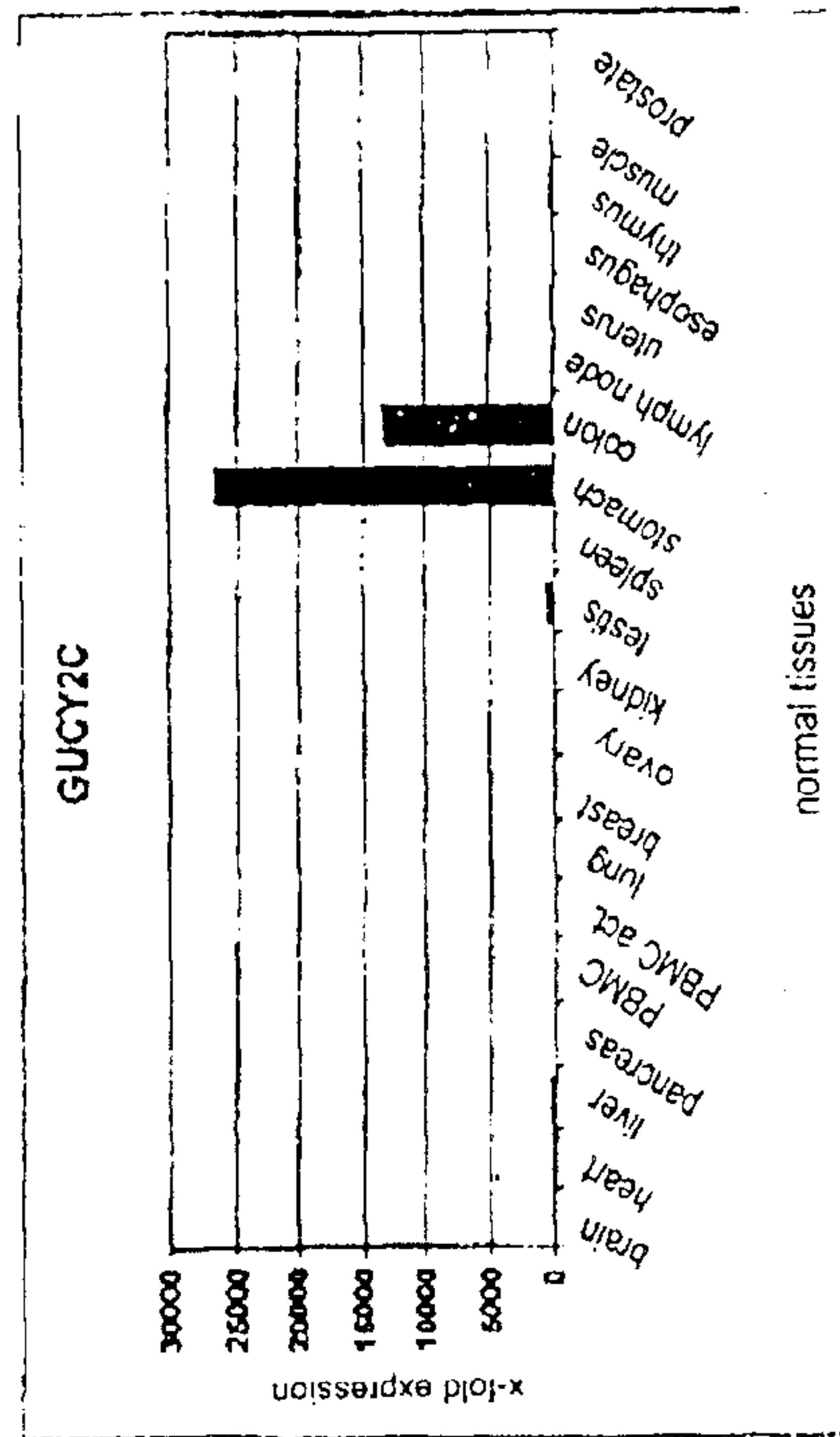
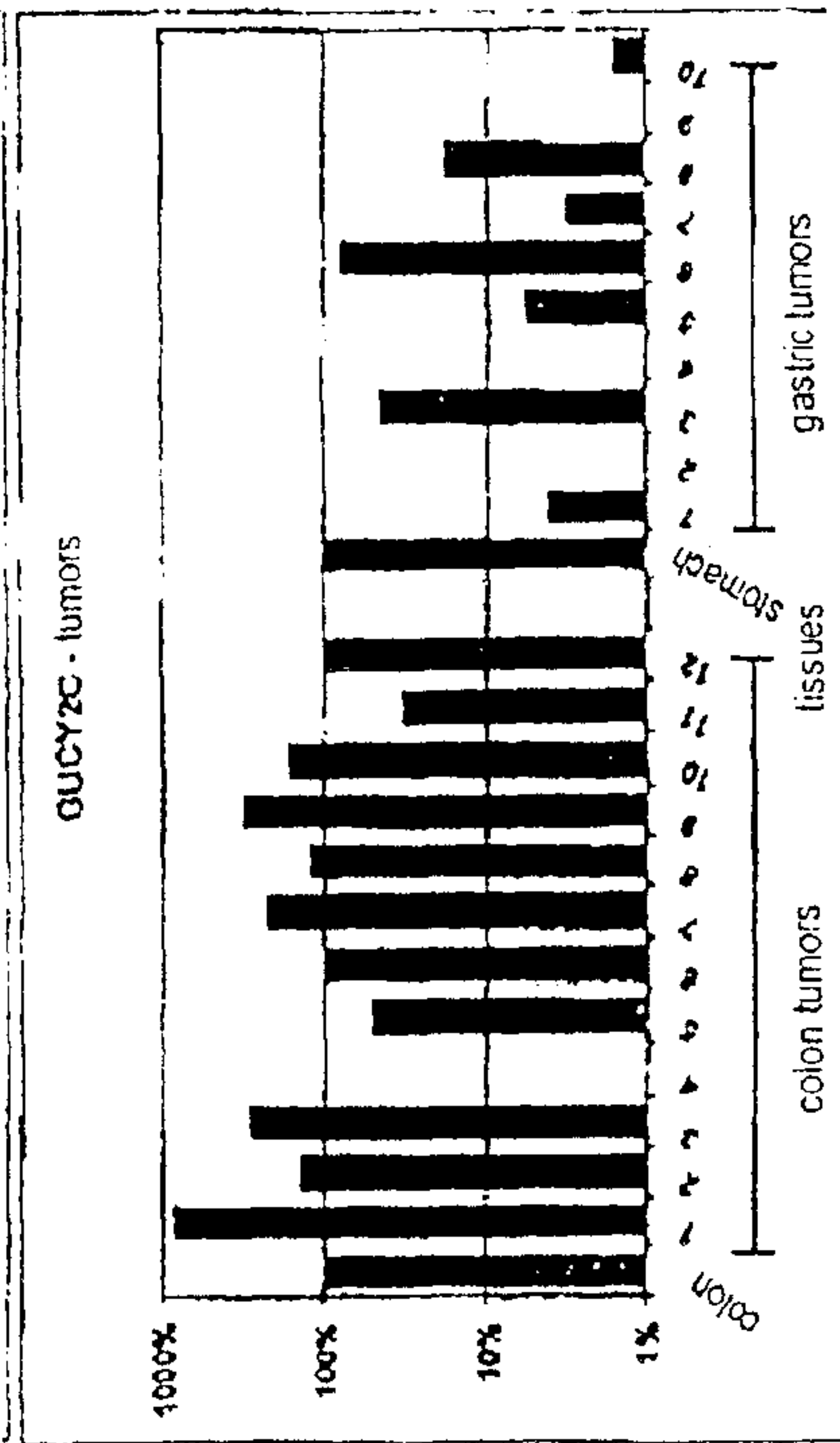


Fig. 19B



2062

Fig. 20

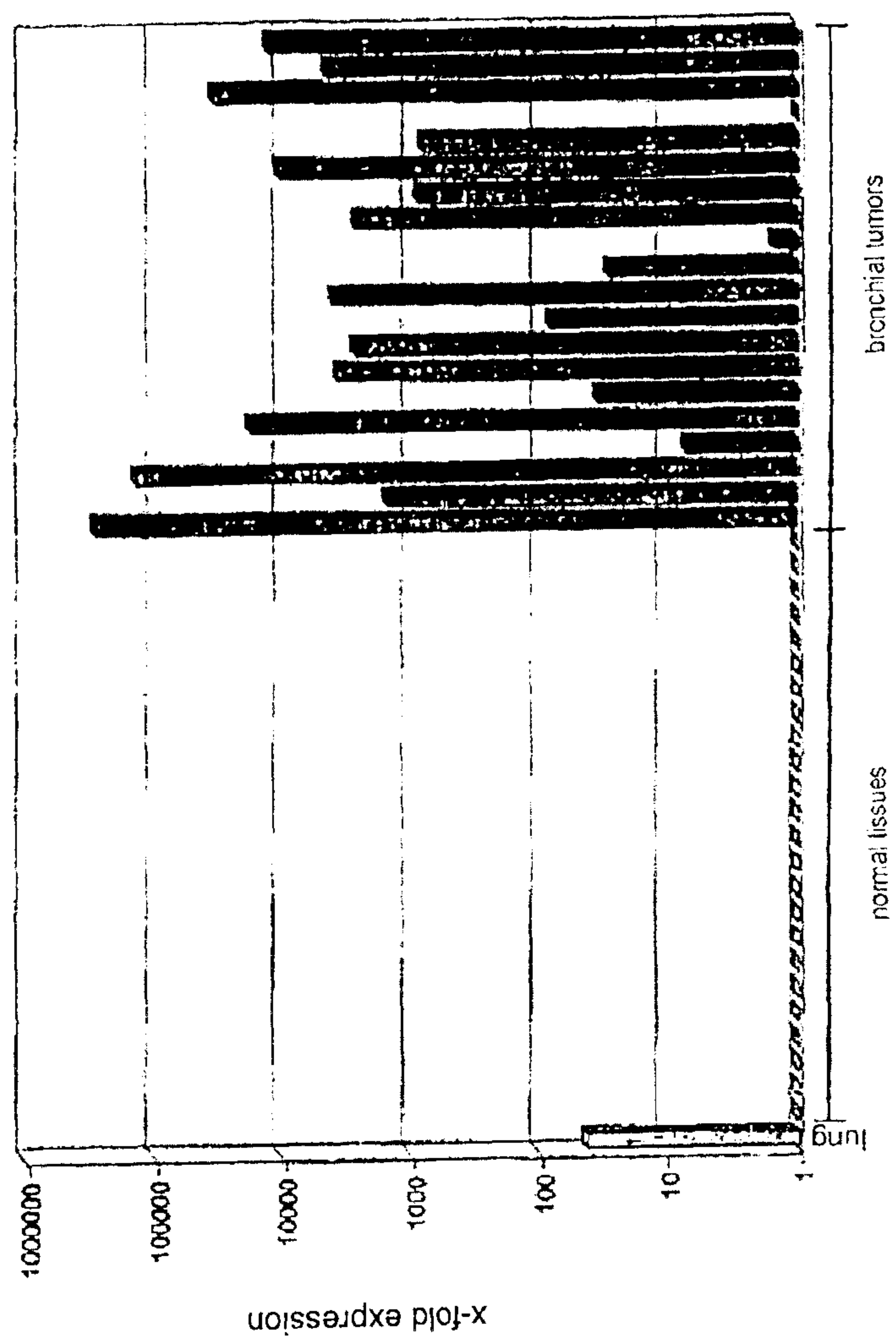


Fig. 21A

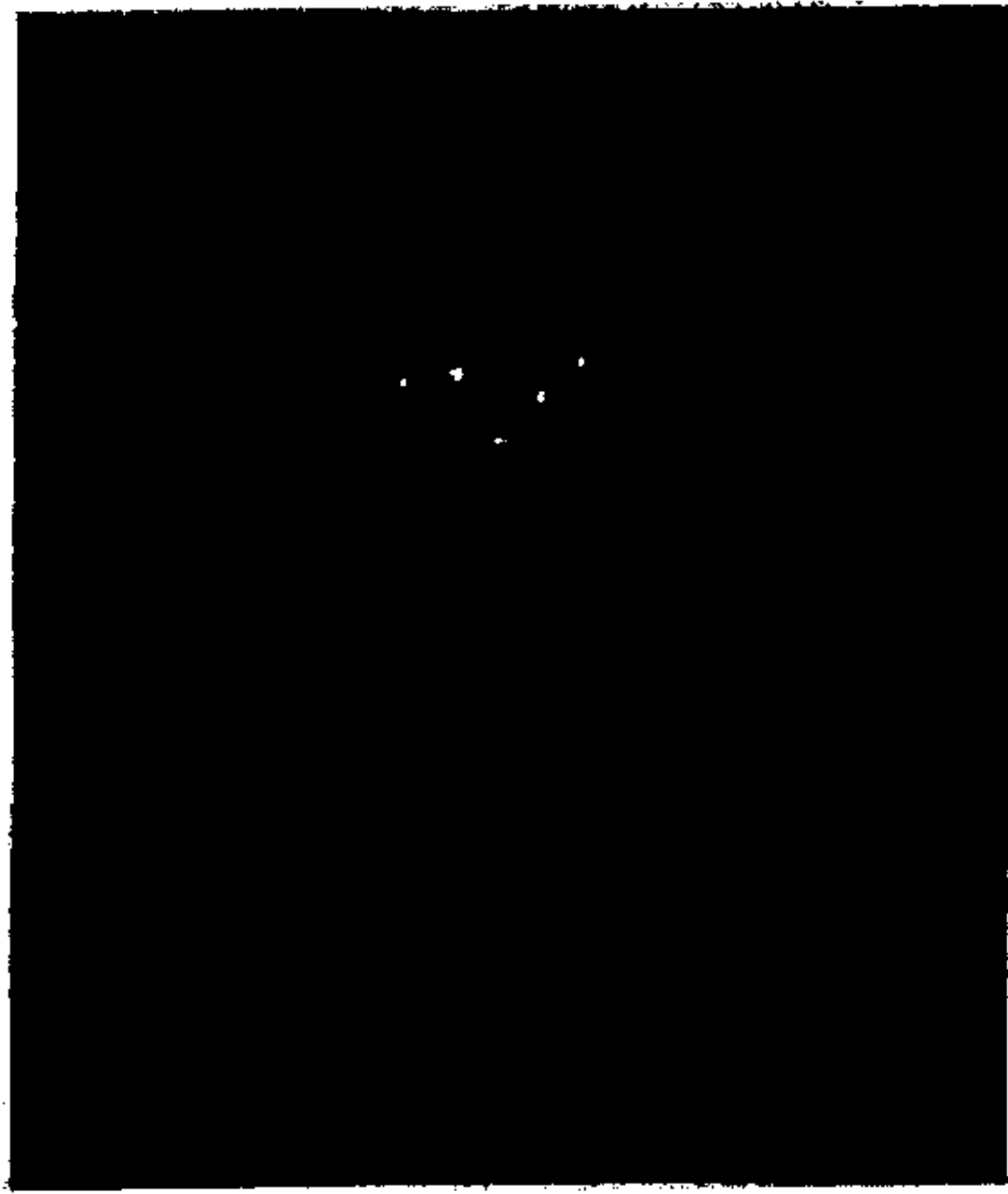
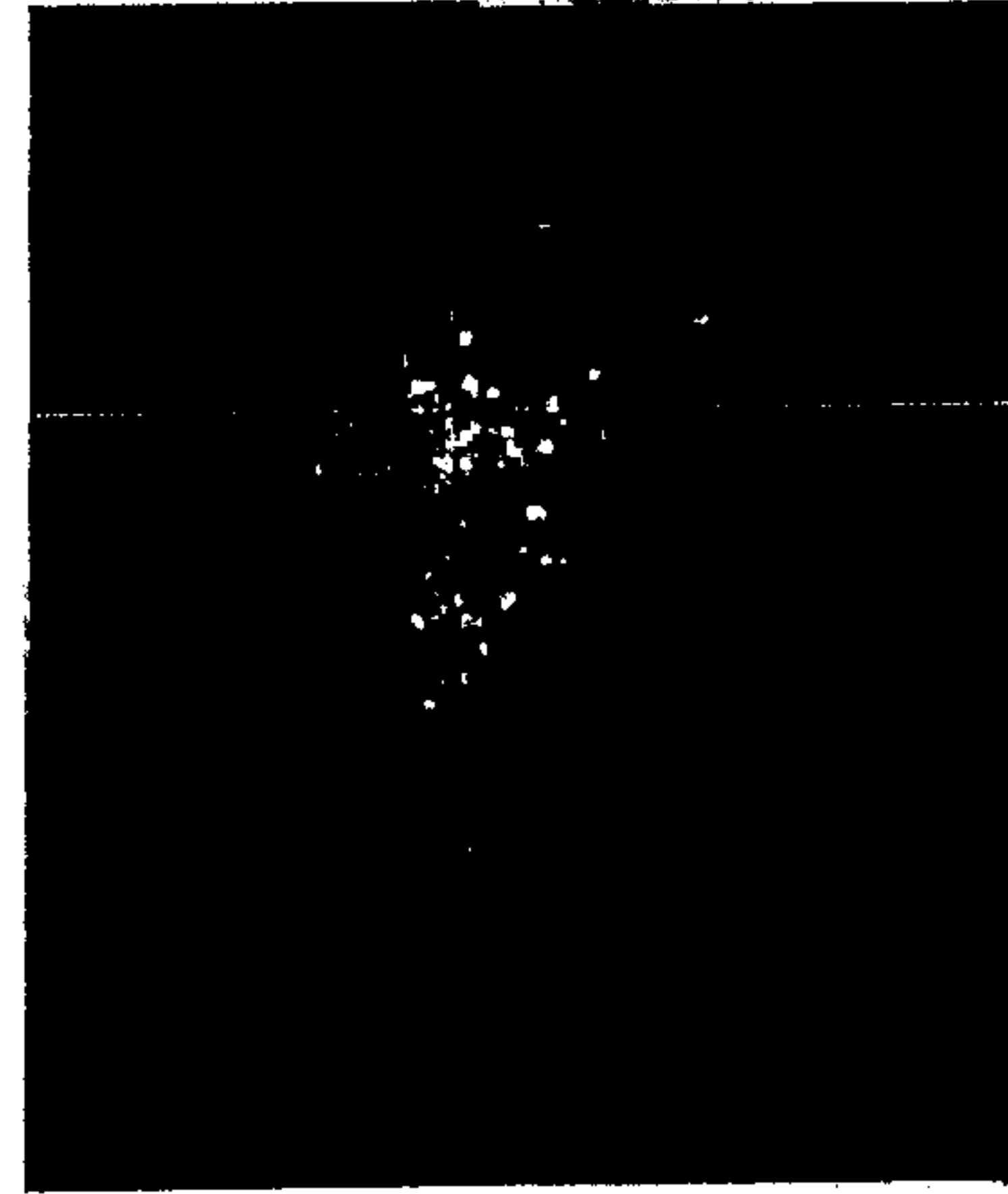
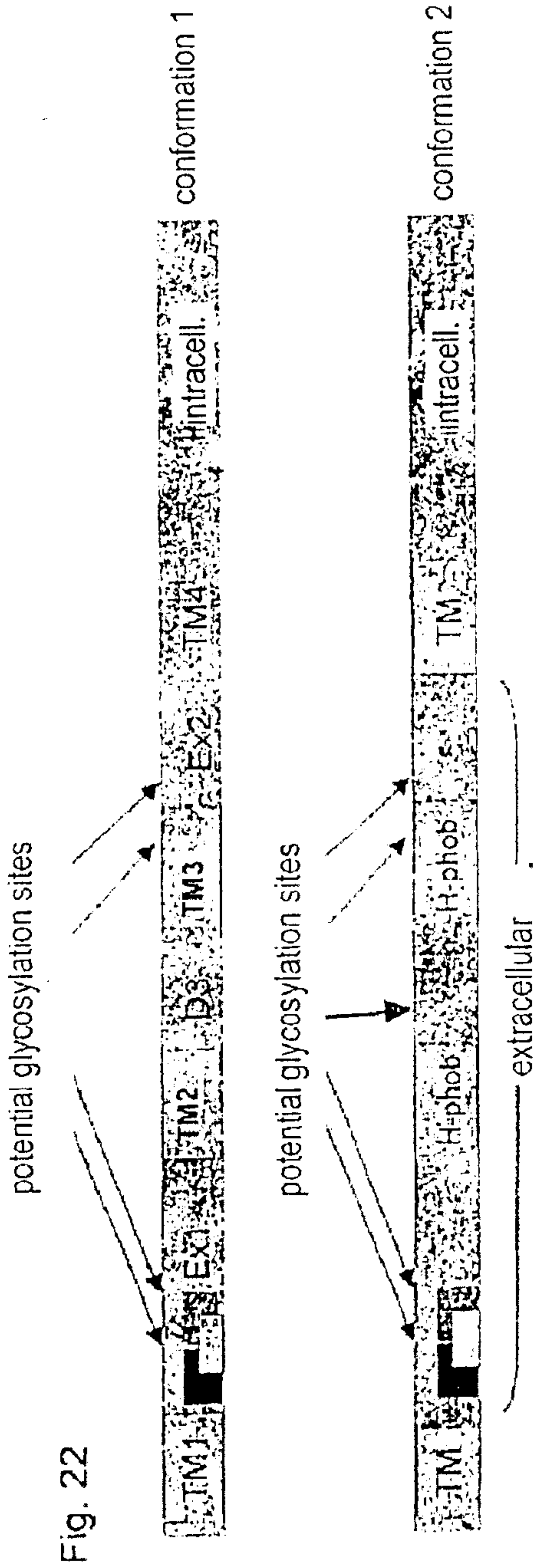


Fig. 21B



Fig. 21C





predicted glycosylation sites (amino acid positions)
Claudin18A2

SeqName	Position	Ranking of the glycosylation sites	Result
Sequence	37	0.7219 (9/9)	++
Sequence	38	0.6502 (8/9)	+
Sequence	45	0.6026 (8/9)	+
Sequence	116	0.5713 (7/9)	+
Sequence	141	0.6348 (7/9)	•
Sequence	146	0.5187 (6/9)	+
Sequence	153	0.4696 (5/9)	•
Sequence	205	0.6011 (8/9)	+
Sequence	234	0.3960 (6/9)	•
Sequence	237	0.4602 (6/9)	•

Fig. 23

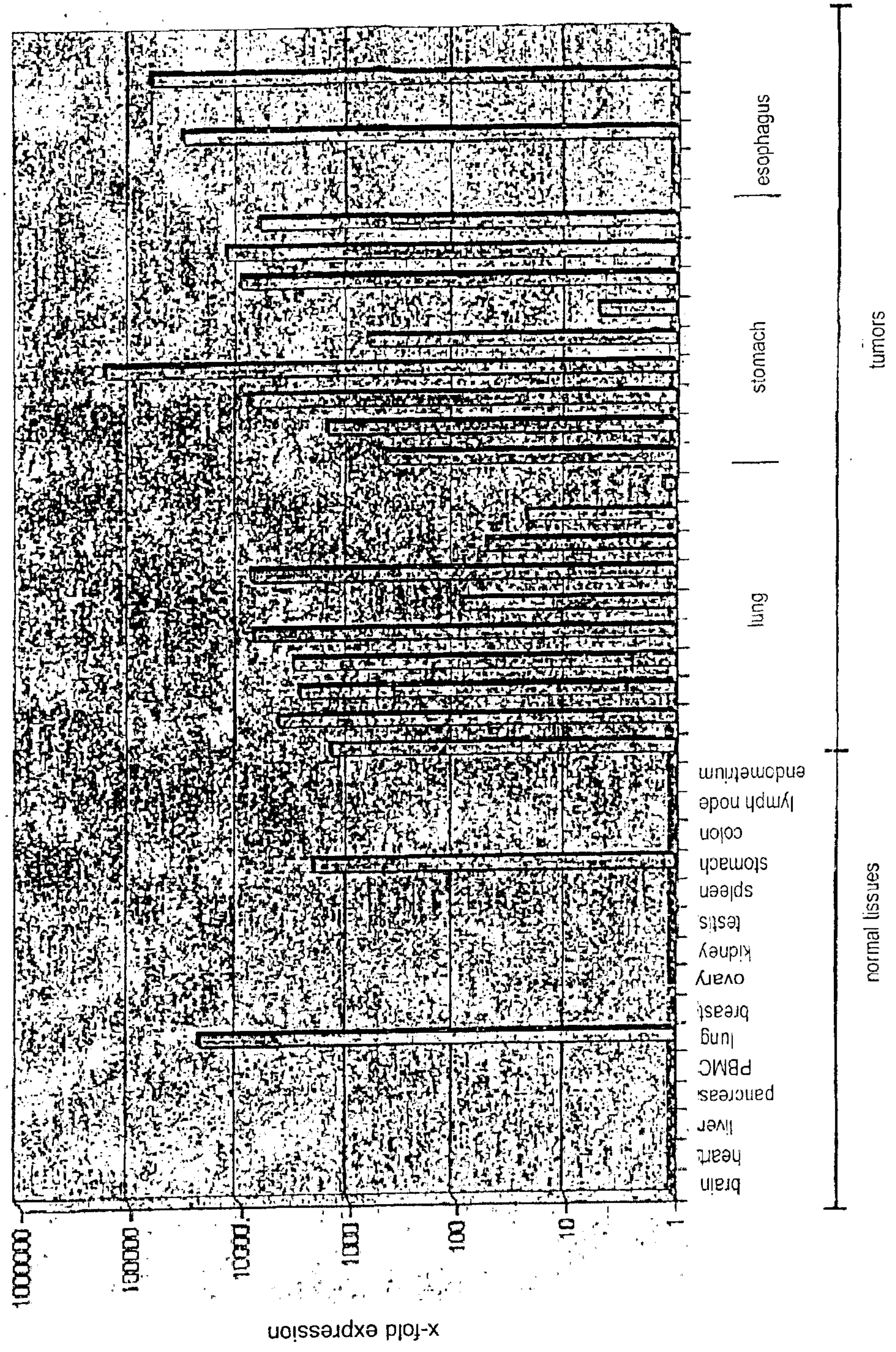


Fig. 24

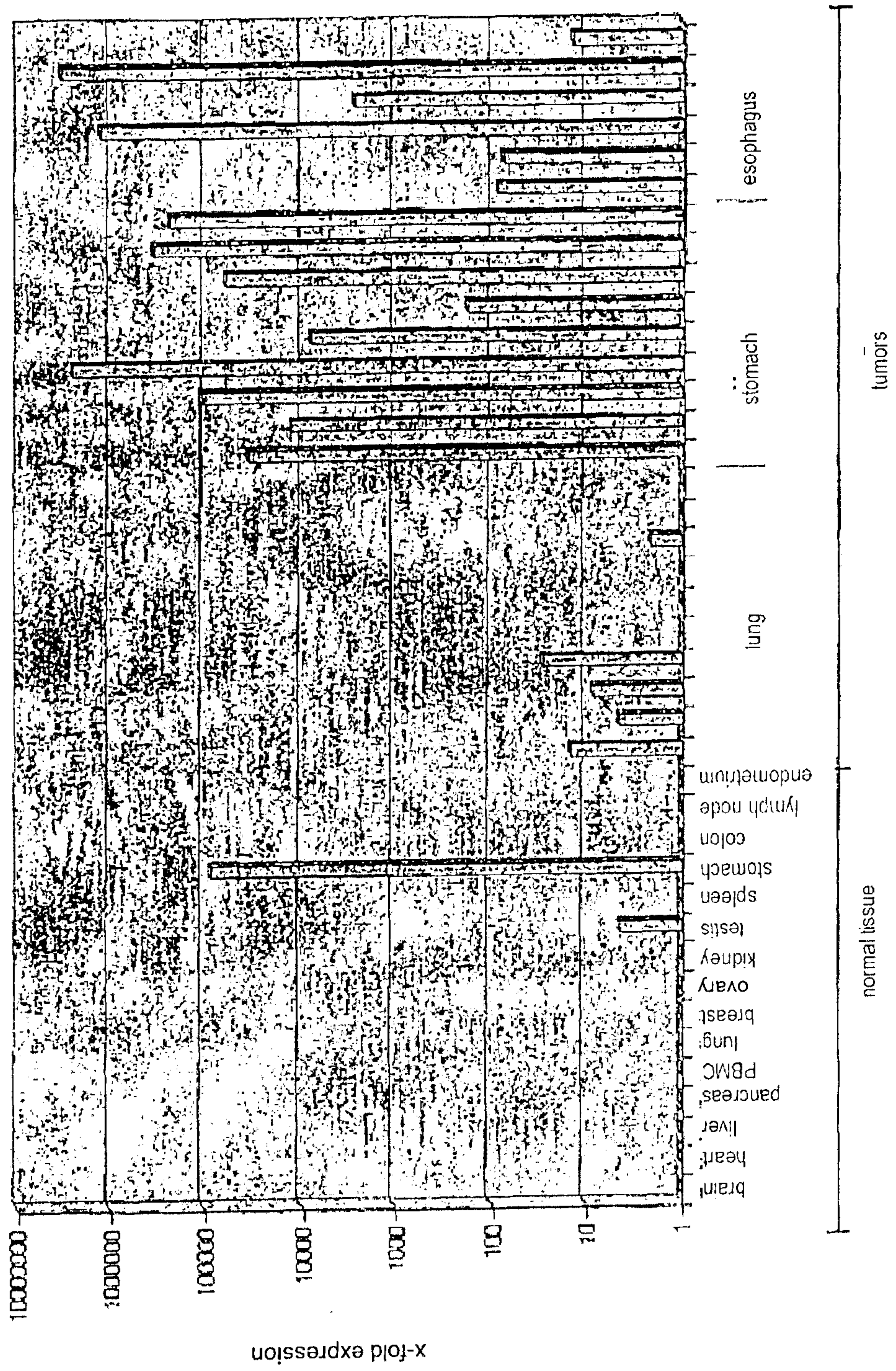


Fig. 25A

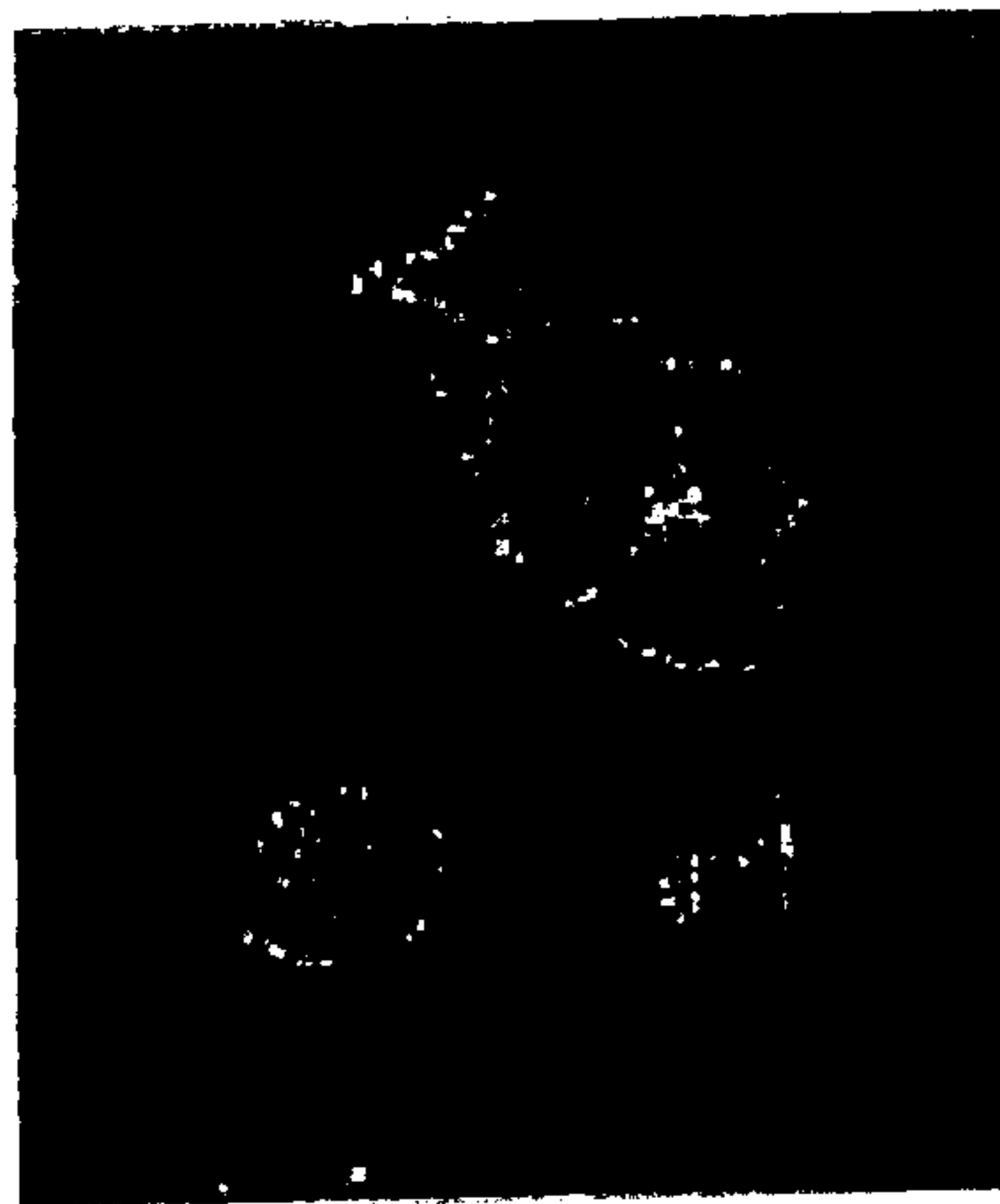


Fig. 25B

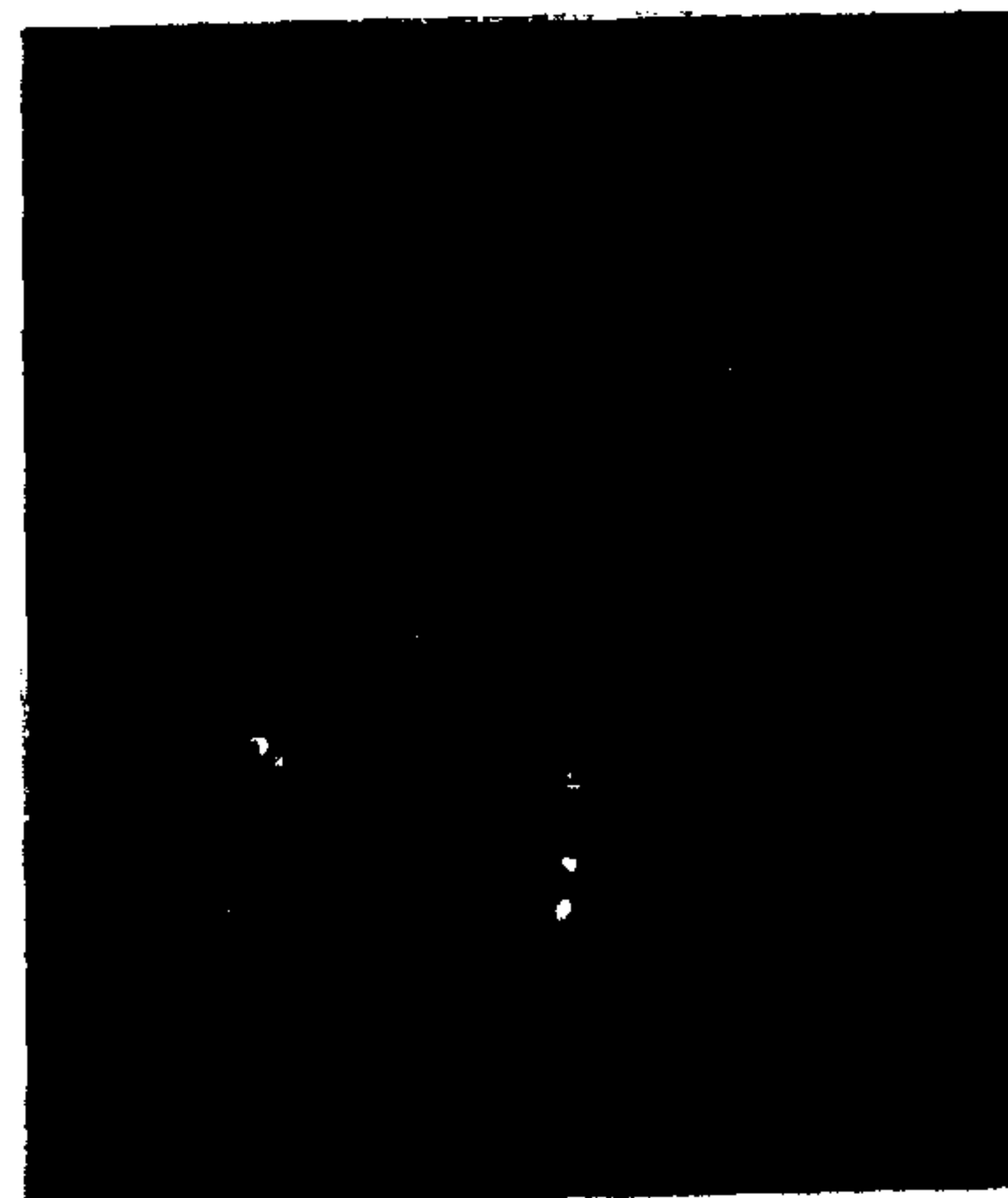


Fig. 25C



Fig. 25D

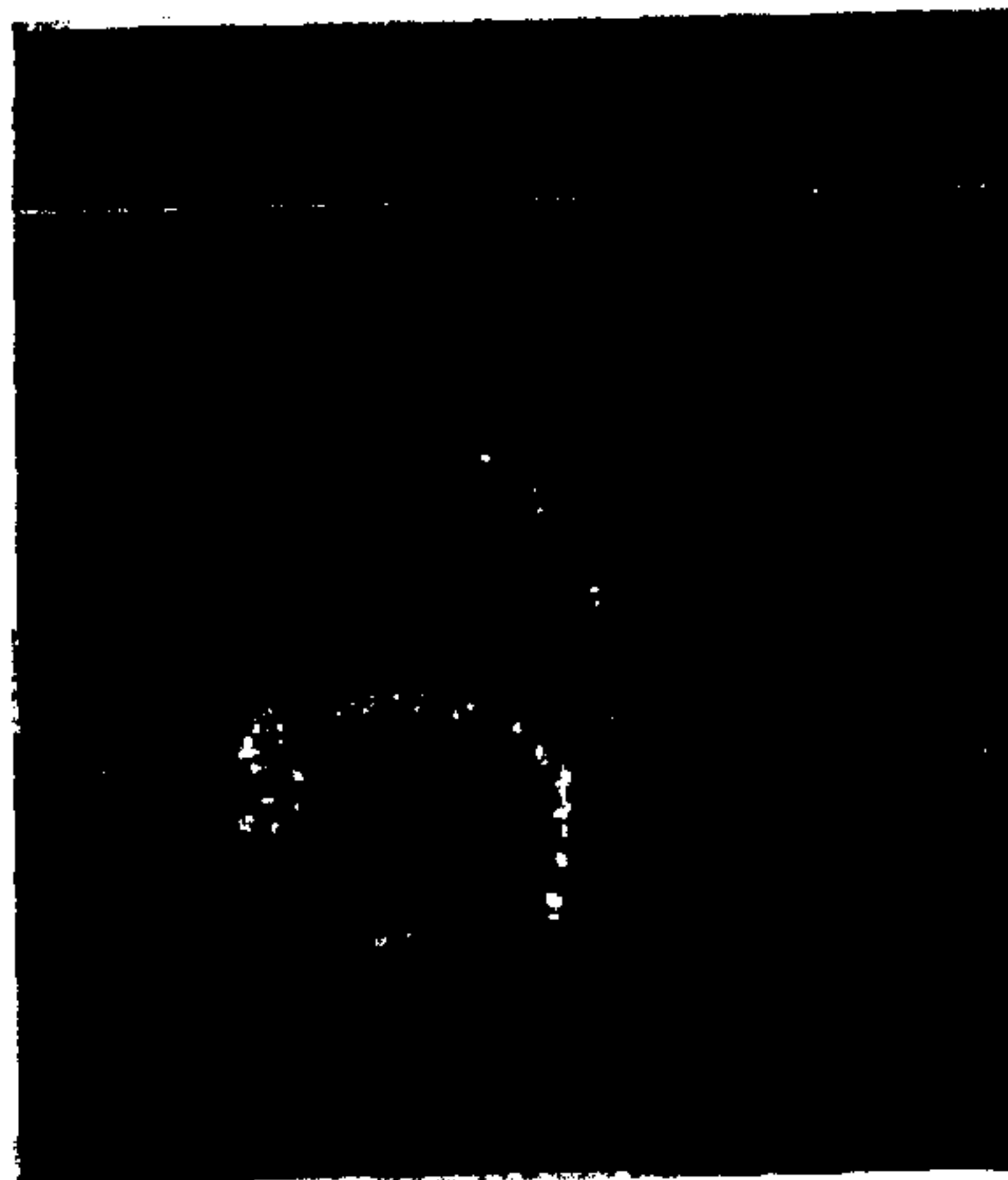


Fig. 26A

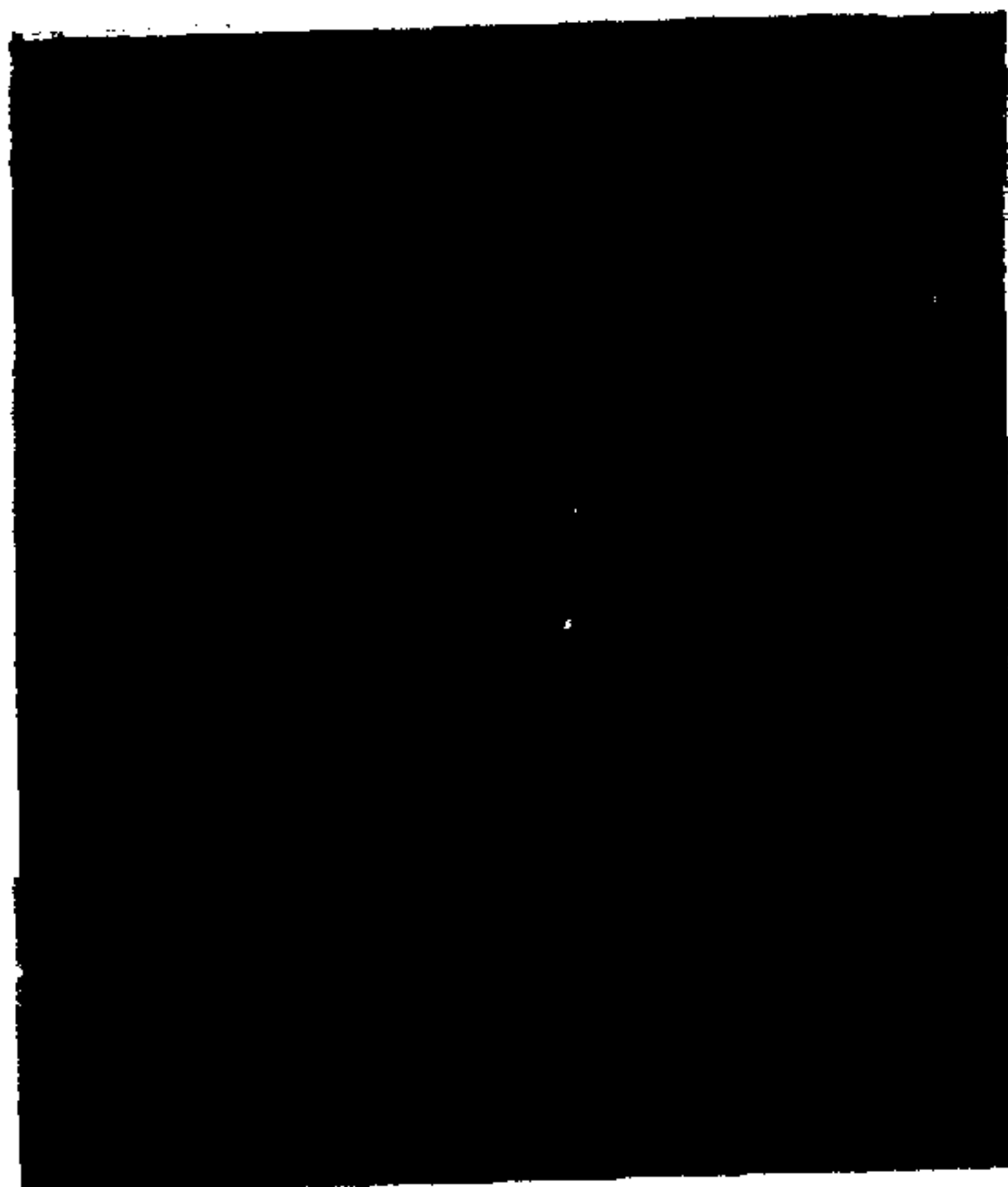


Fig. 26B



Fig. 26C



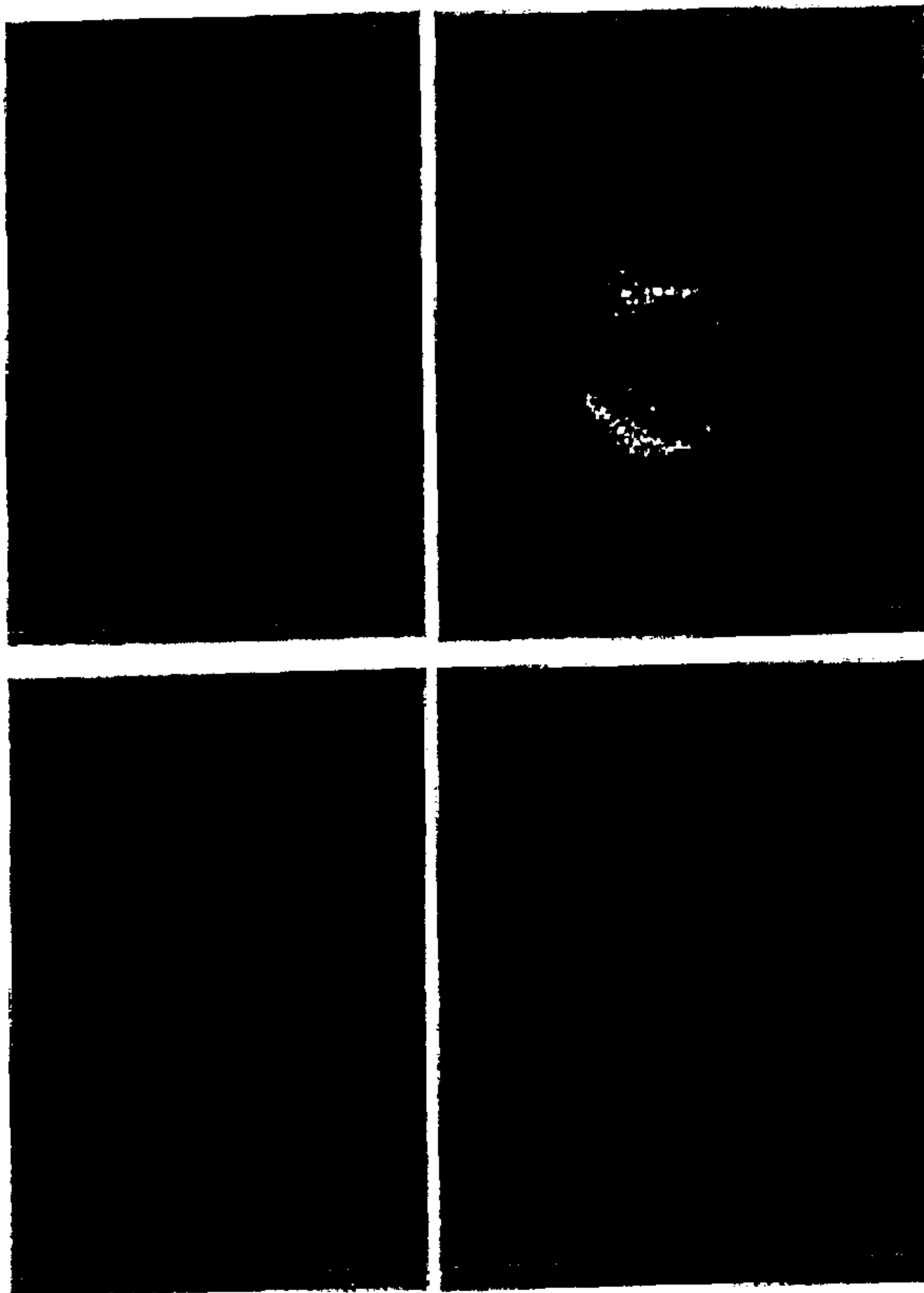


Fig. 27

Fig. 28

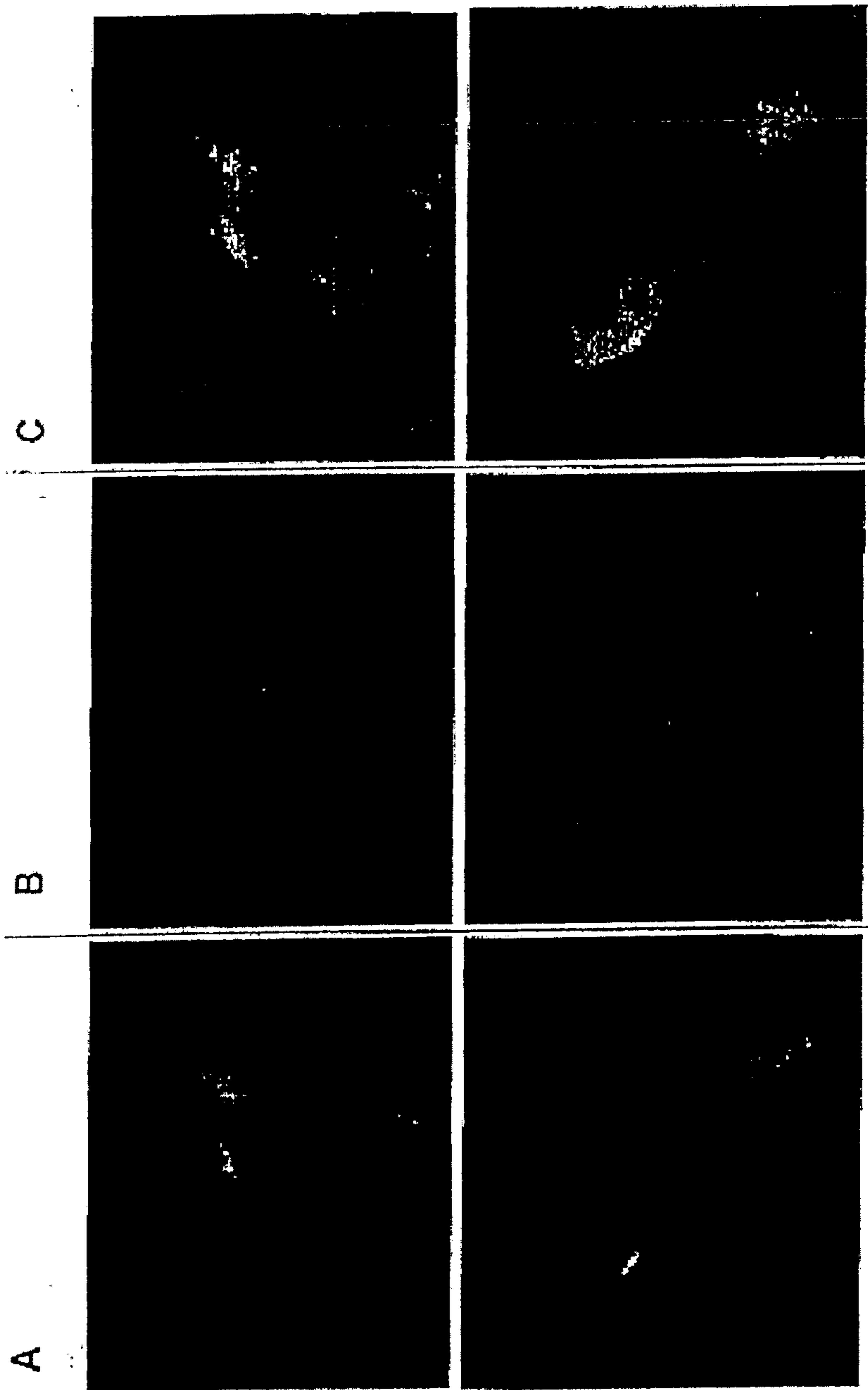
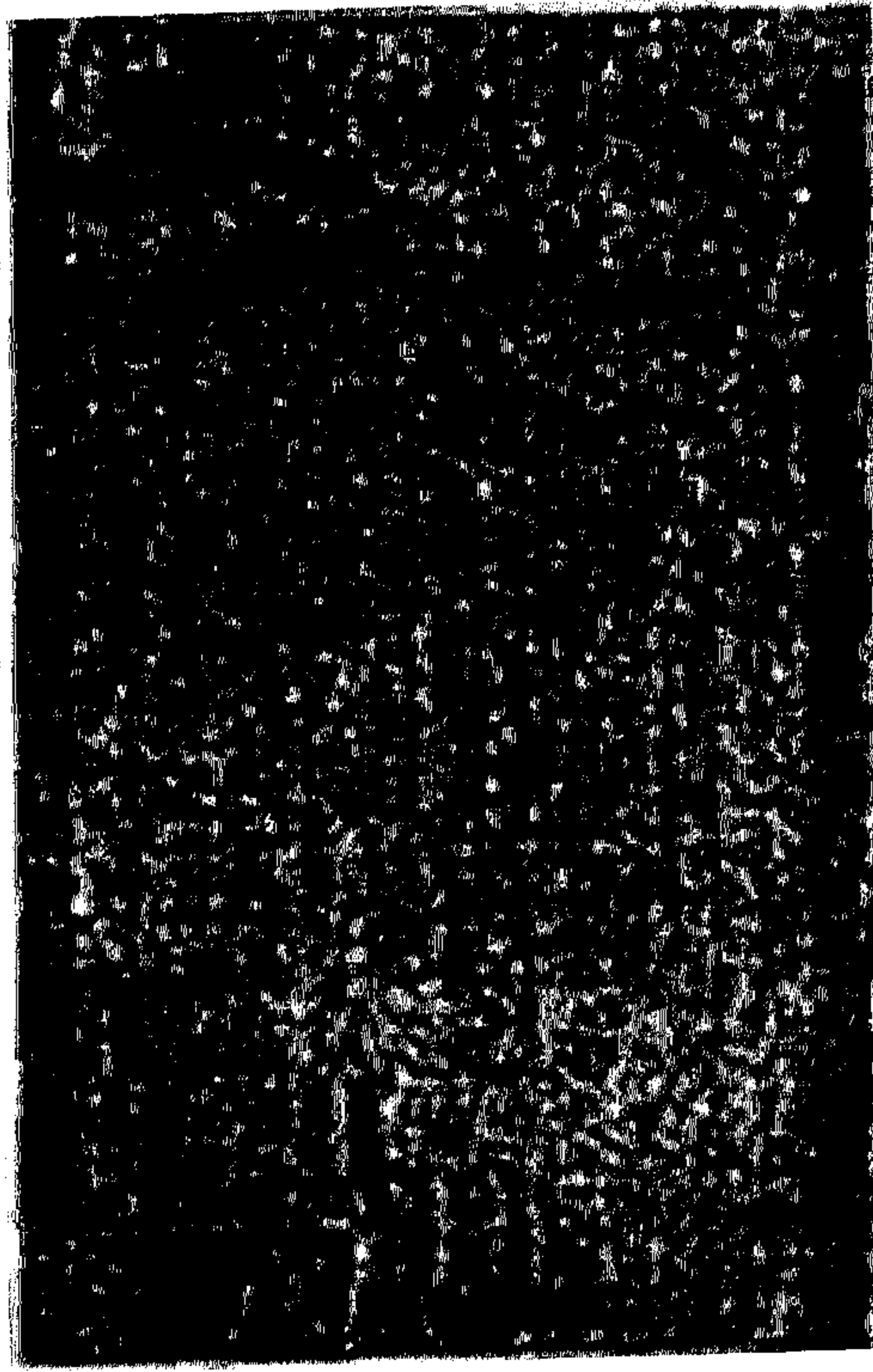


Fig. 29

M 1 2 3 4 5 6 7 8 9



19/61

Fig. 30A

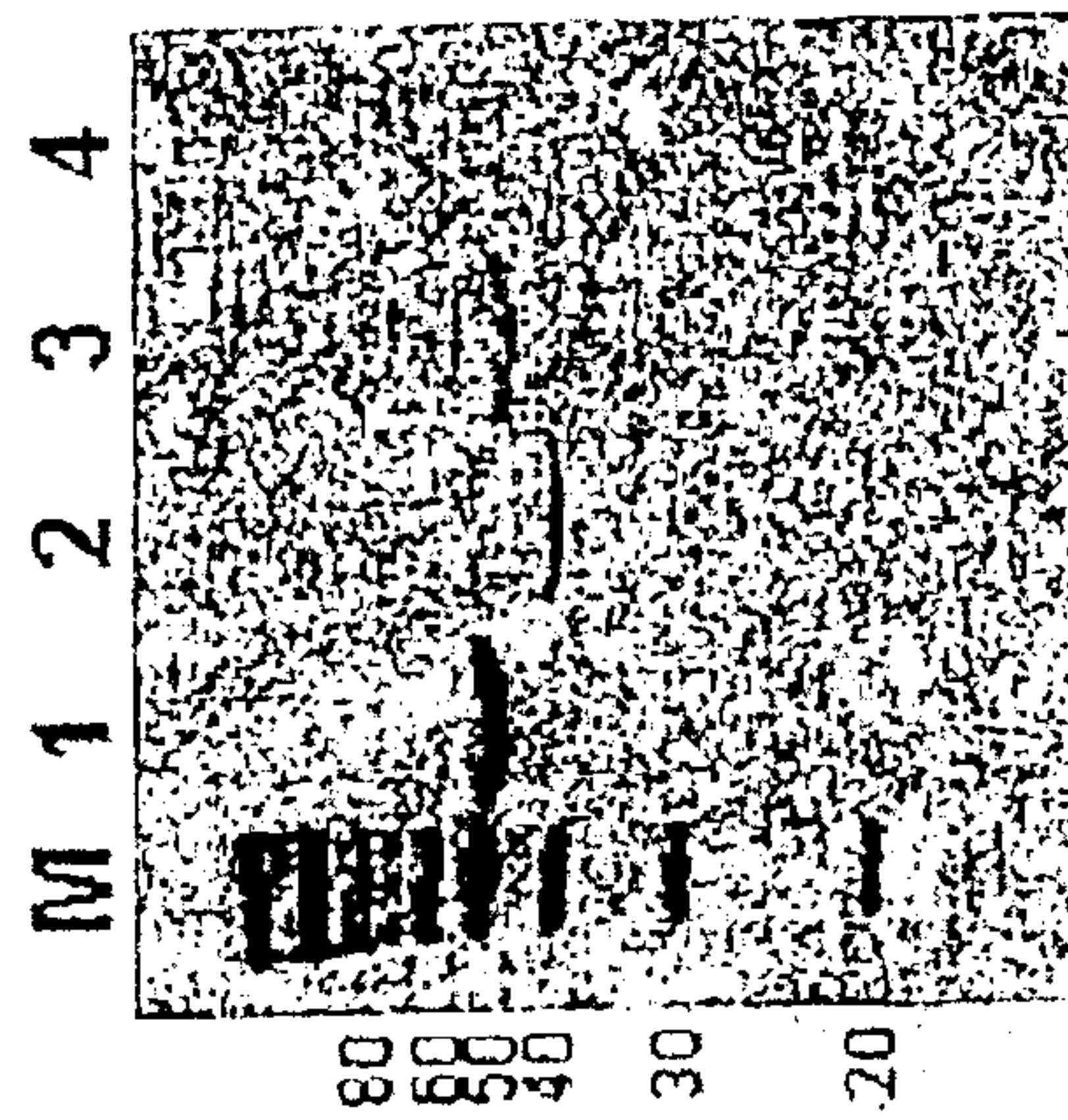


Fig. 30B

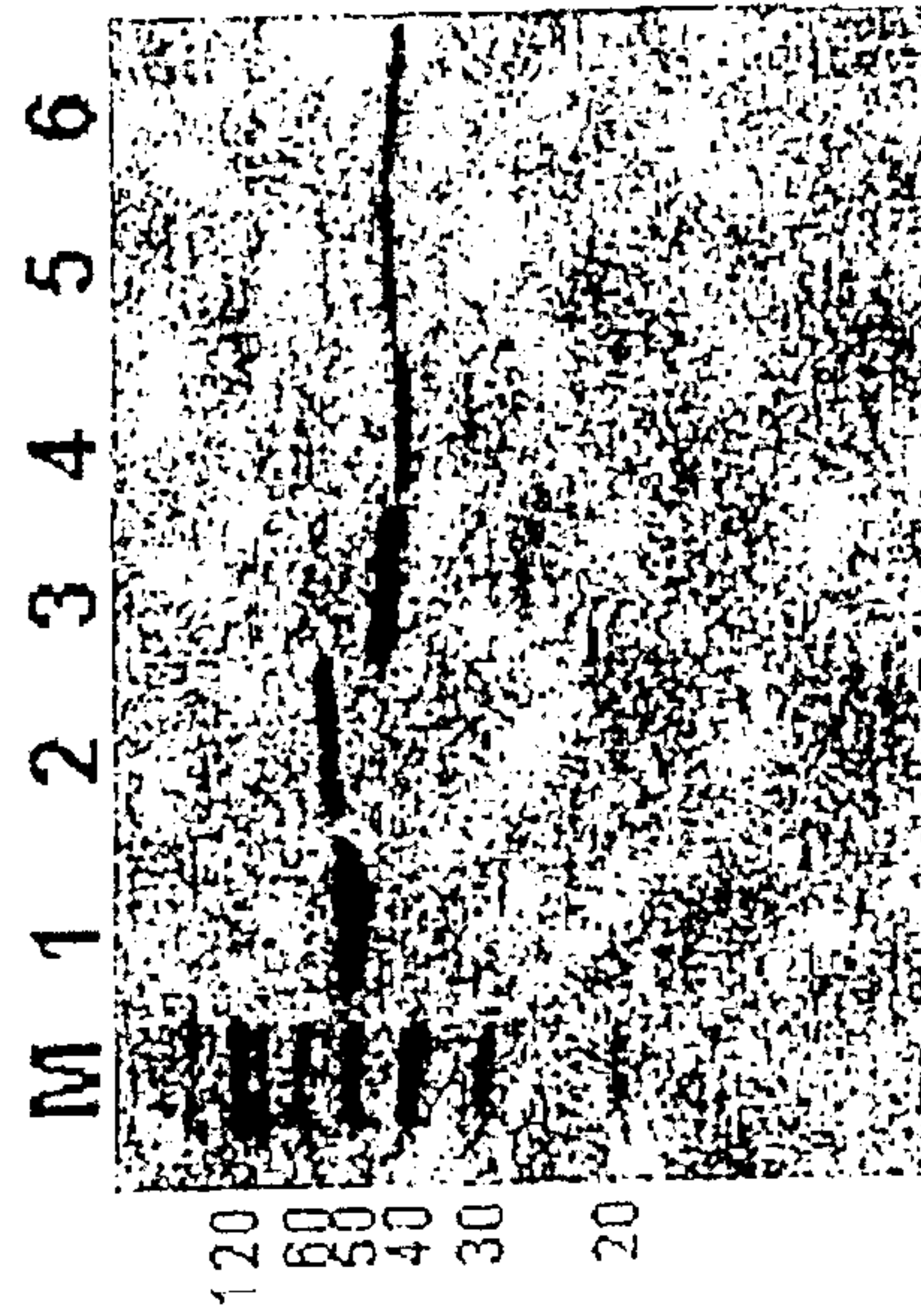


Fig. 30C

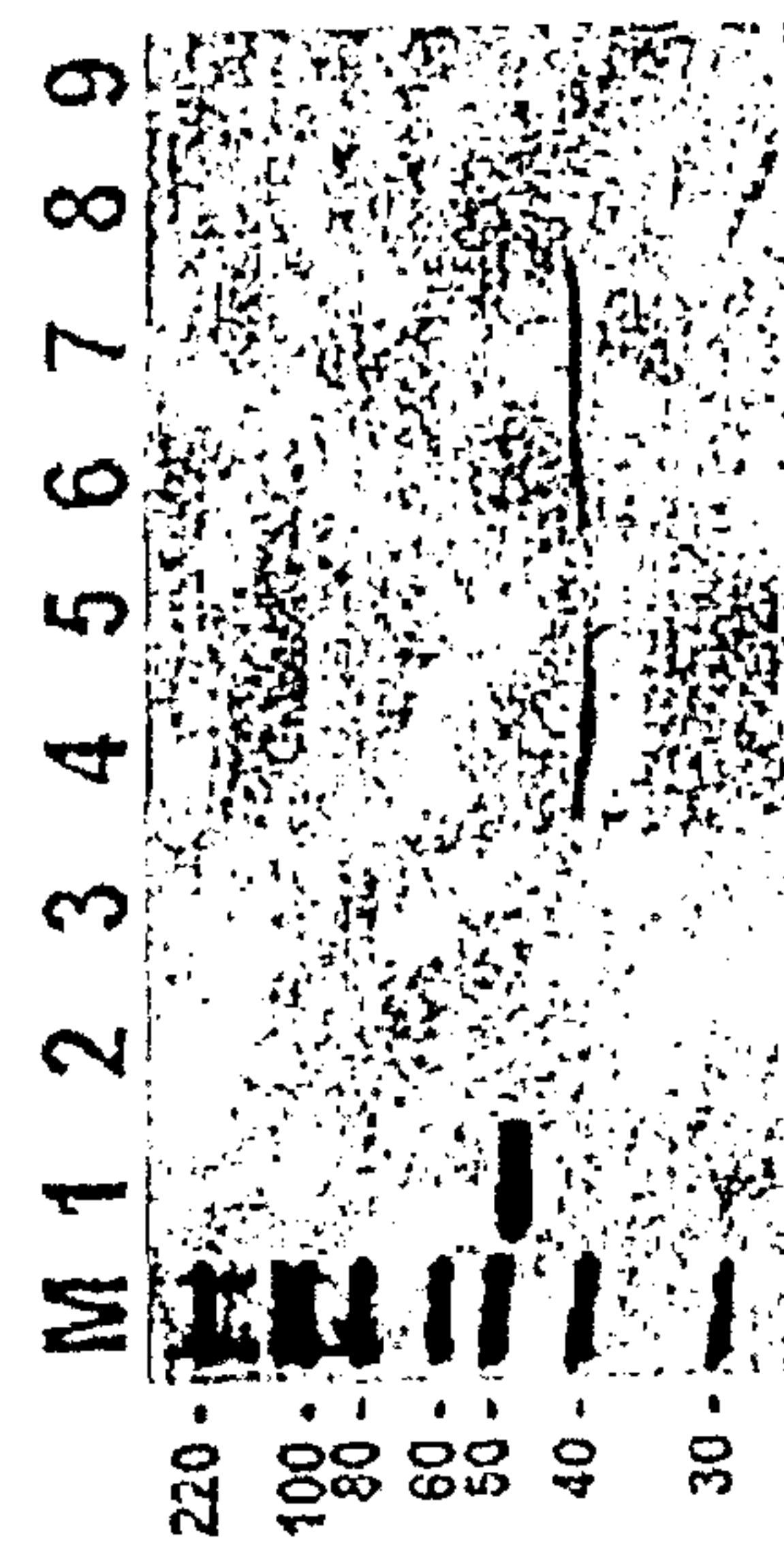
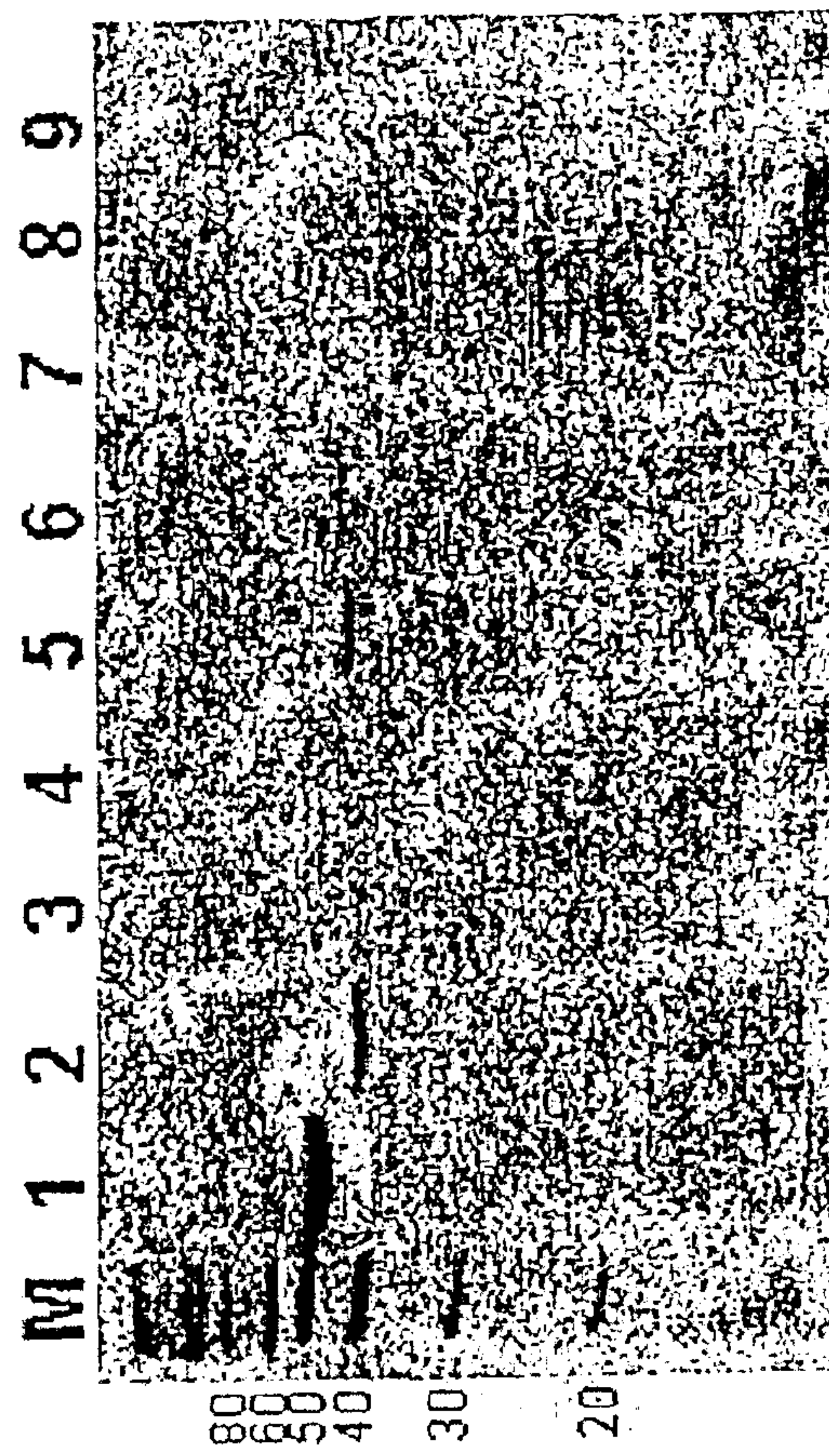


Fig. 30D

tissue specificity of the cell lines	positive/tested cell lines
stomach	4 / 5
esophagus	1 / 3
lung	2 / 5
pancreas	4 / 5
breast	1 / 4
colon	1 / 3
kidney	0 / 1
skin	0 / 1
ovary	1 / 2
liver	0 / 1

Fig. 31



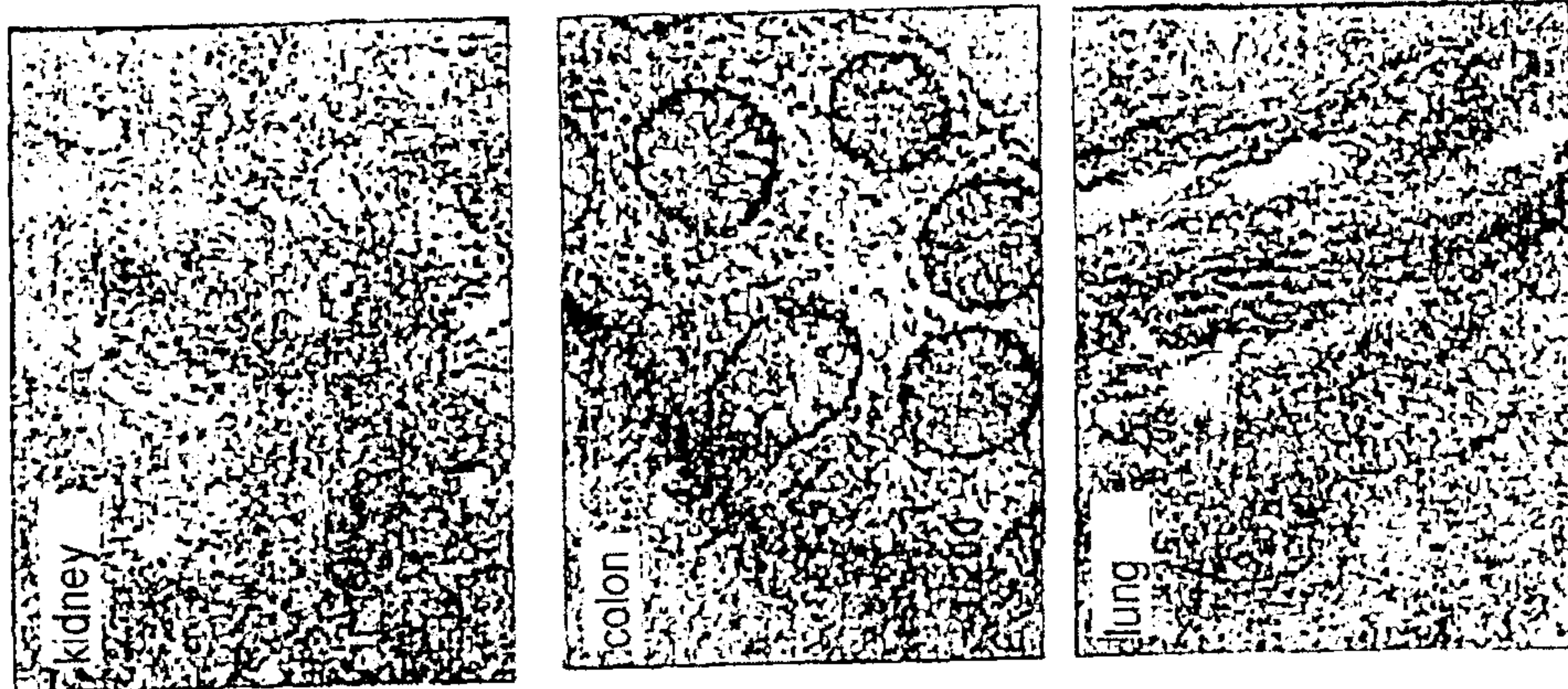
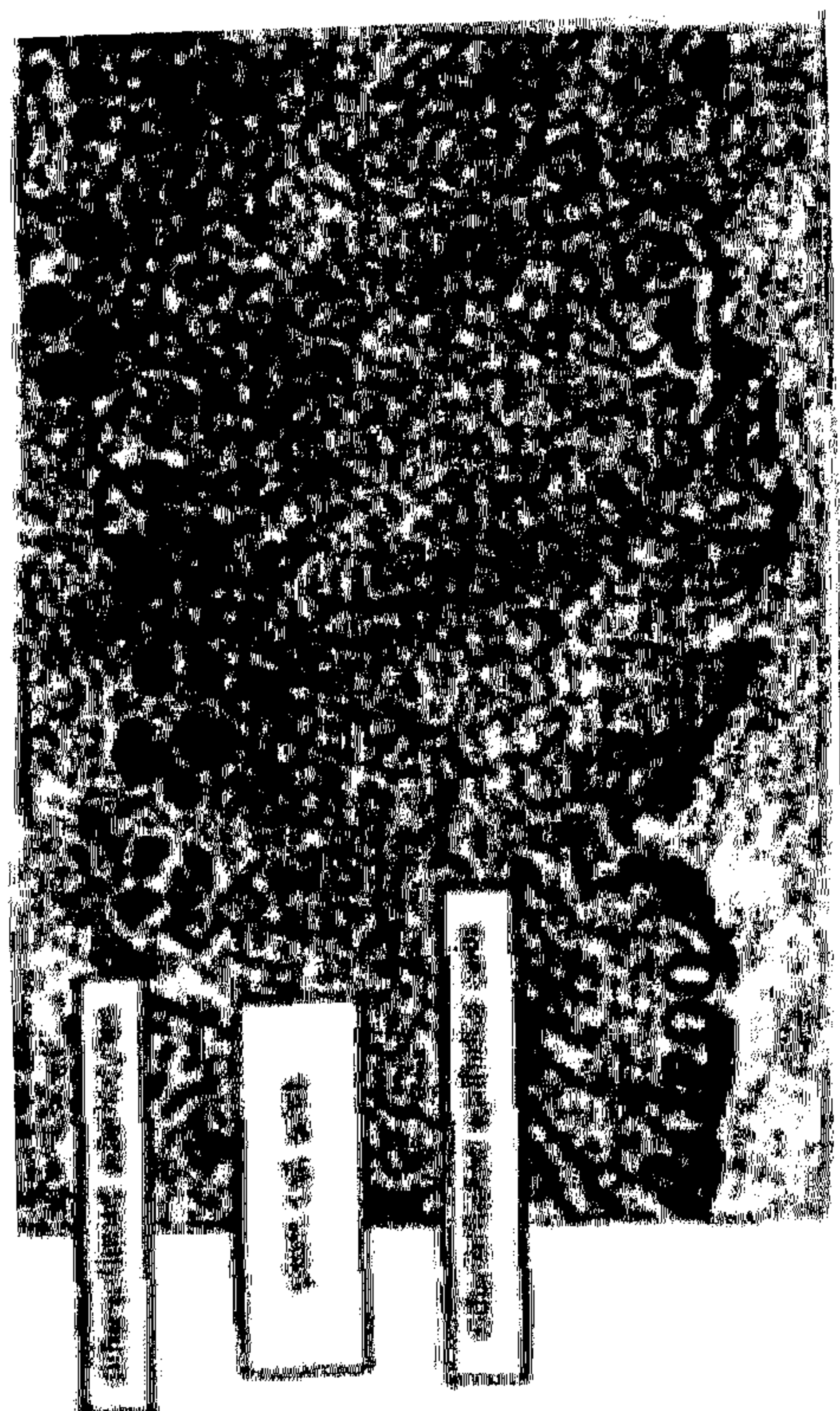


Fig. 32



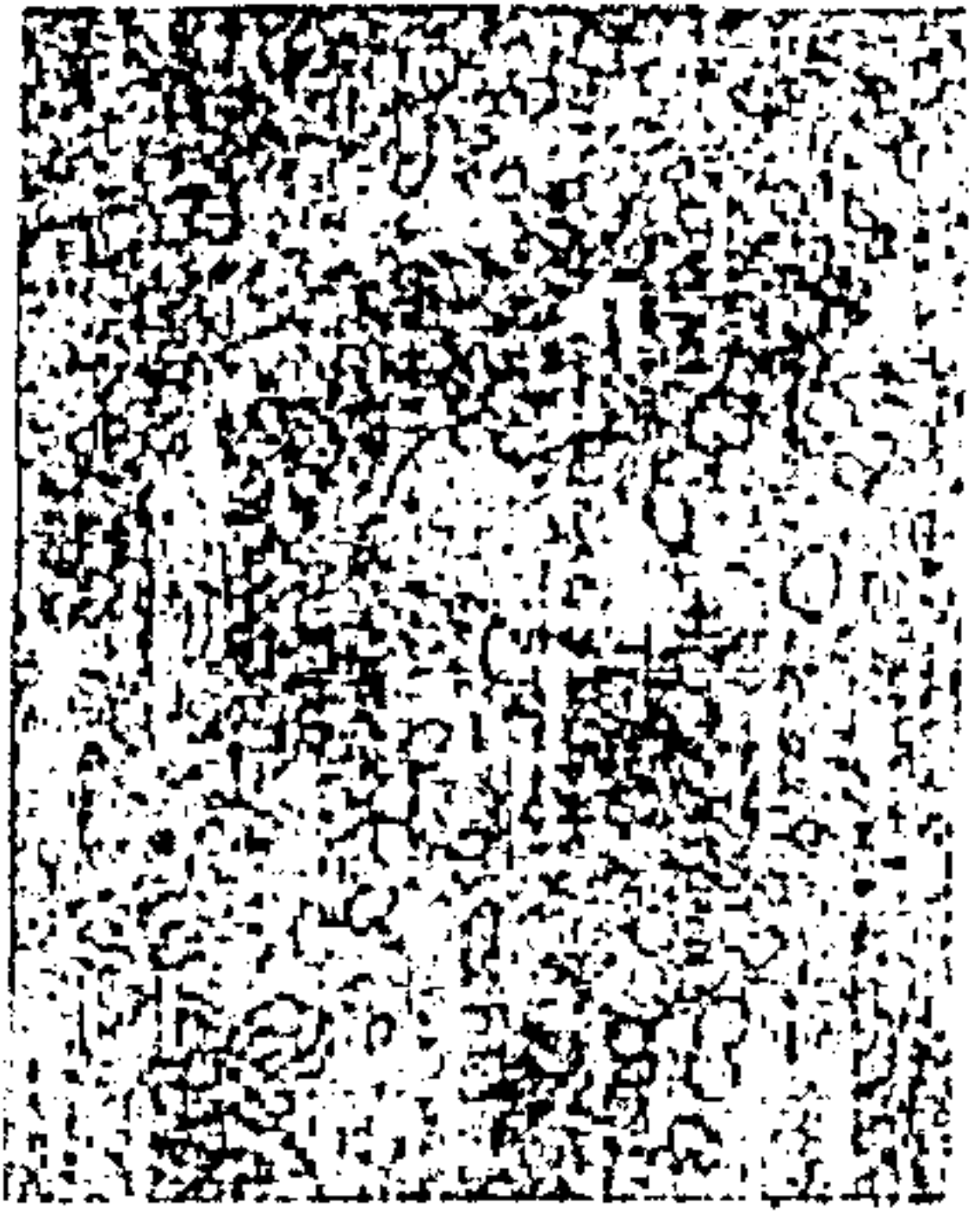
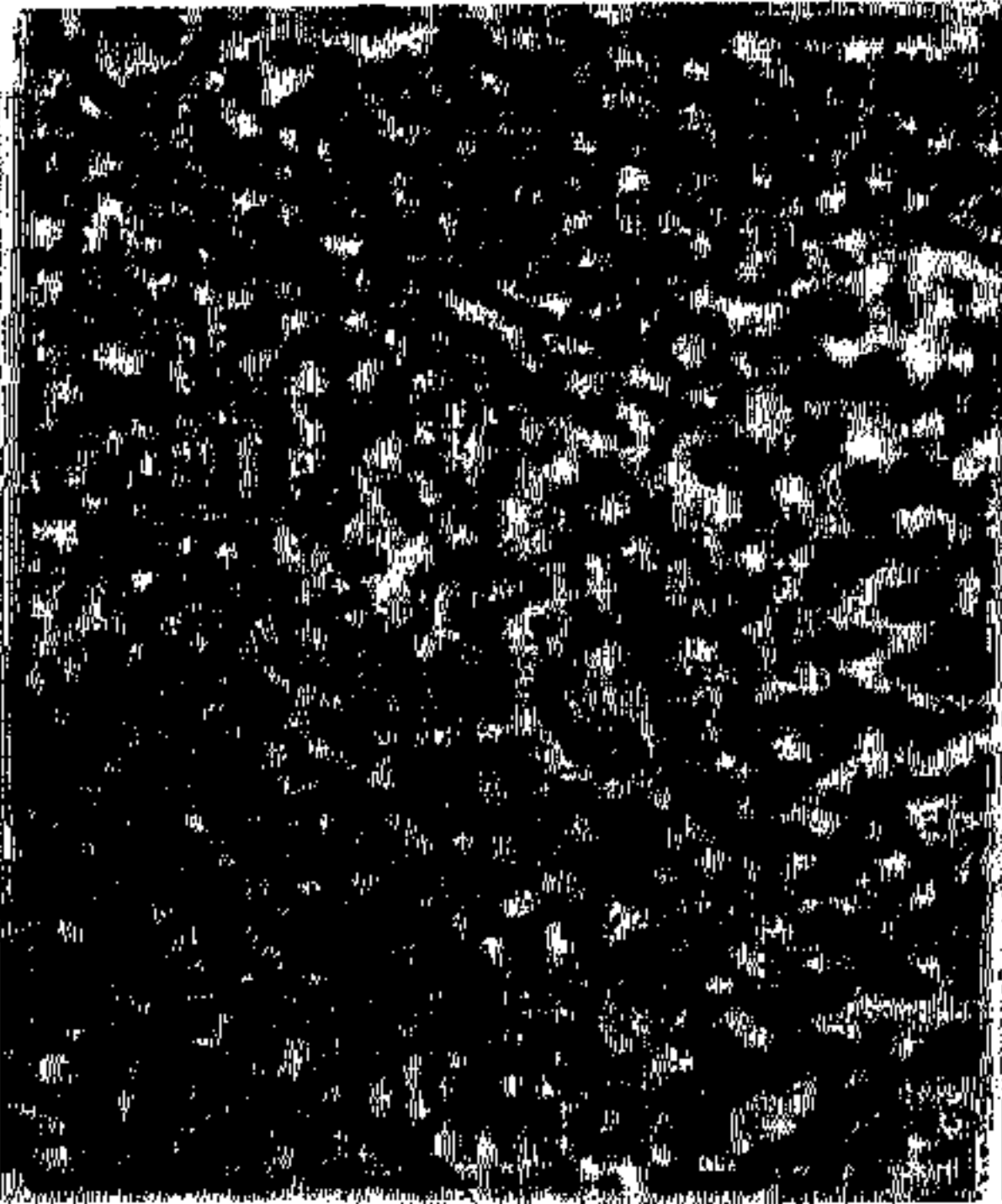
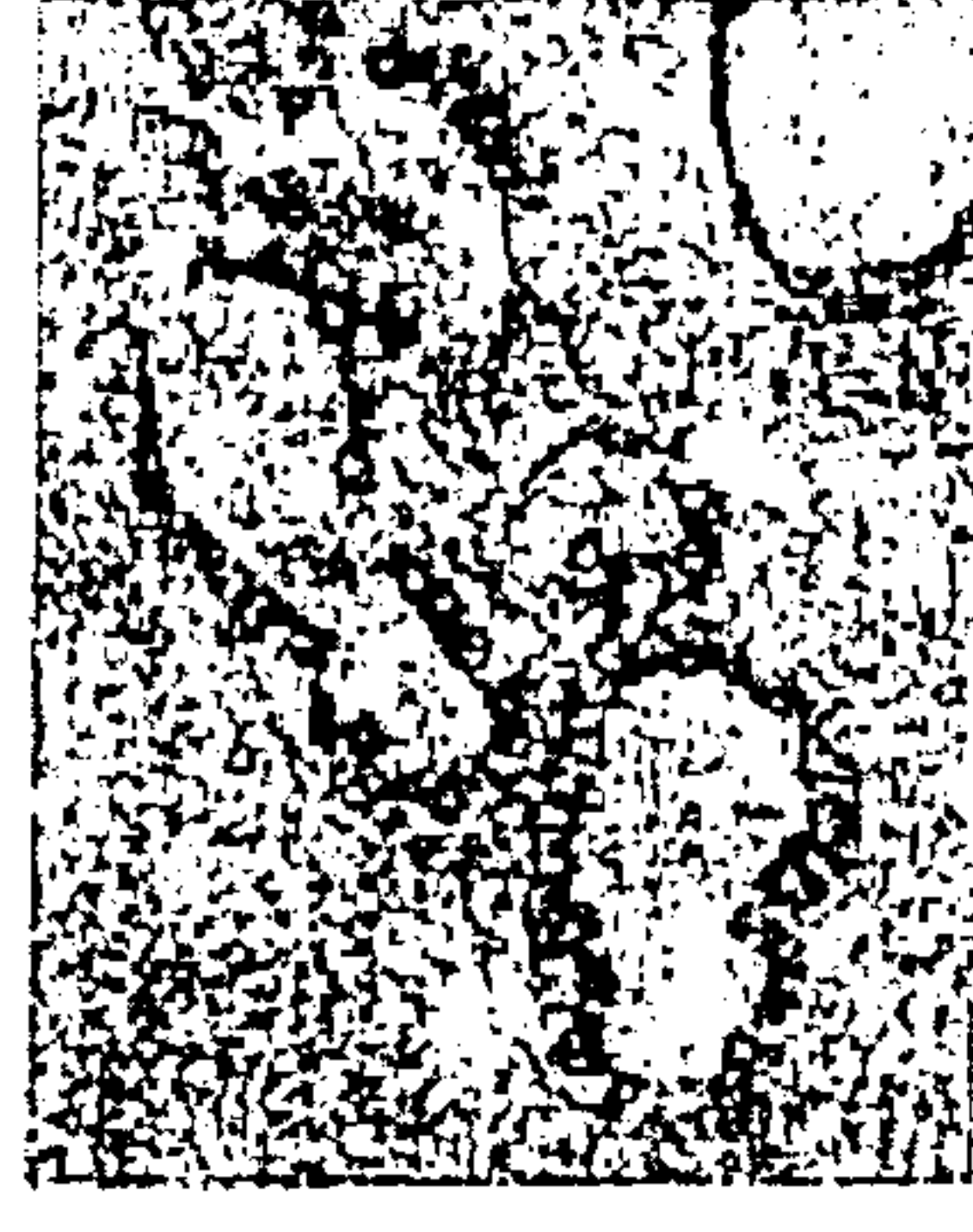


Fig. 33A

Fig. 33B

Fig. 33C

Fig. 33D

tissue	tumor	
	tumor type	positive/ total
stomach	AdenoCa	9/10
lung	AdenoCa	11/12
lung	SCC	0/2
esophagus	AdenoCa	5/5
esophagus	SCC	0/3
prostate		1/3
breast		1/3
kidney	RCC	0/3
colon		0/3
brain	glioblastoma	0/3

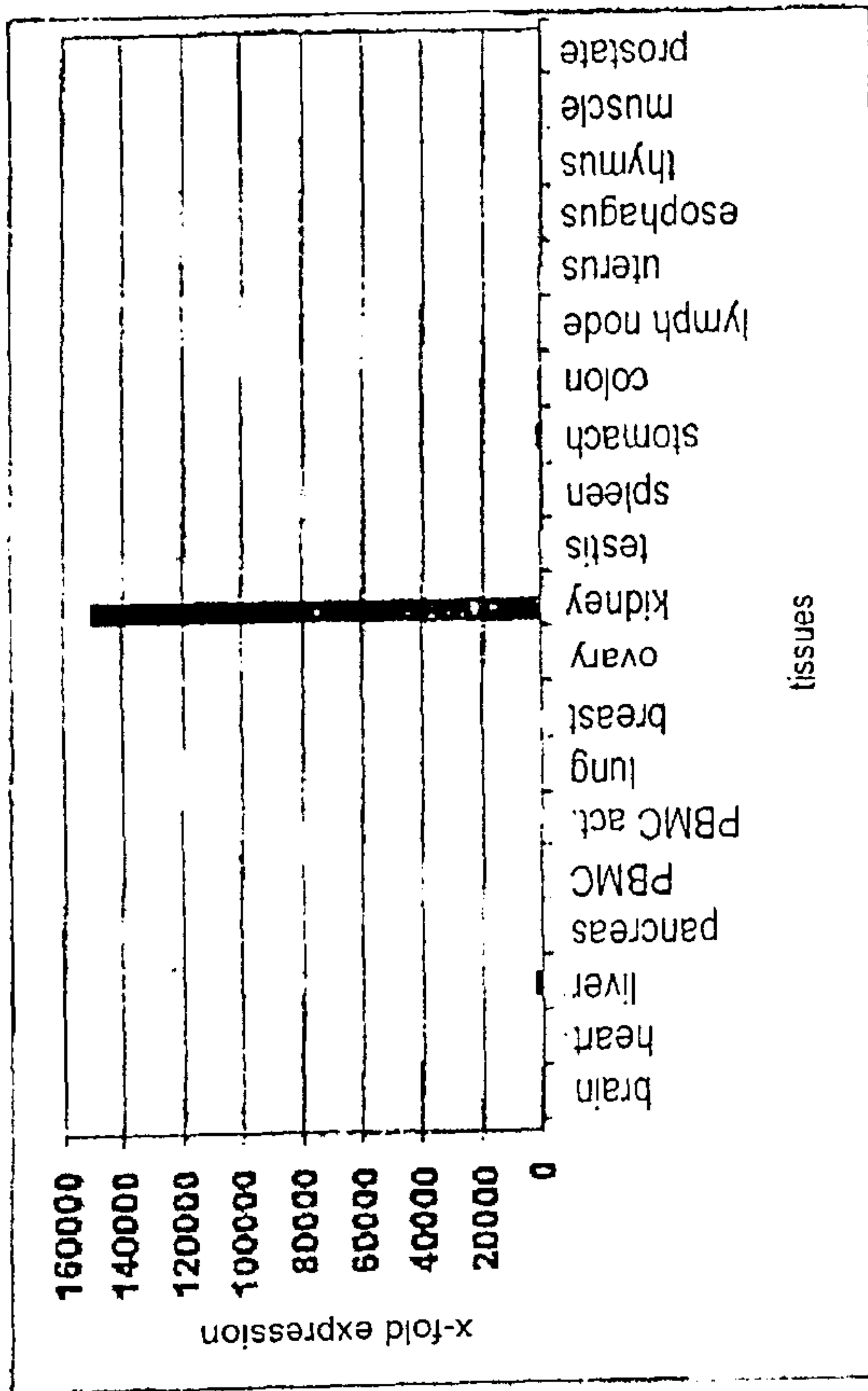


Fig. 34A

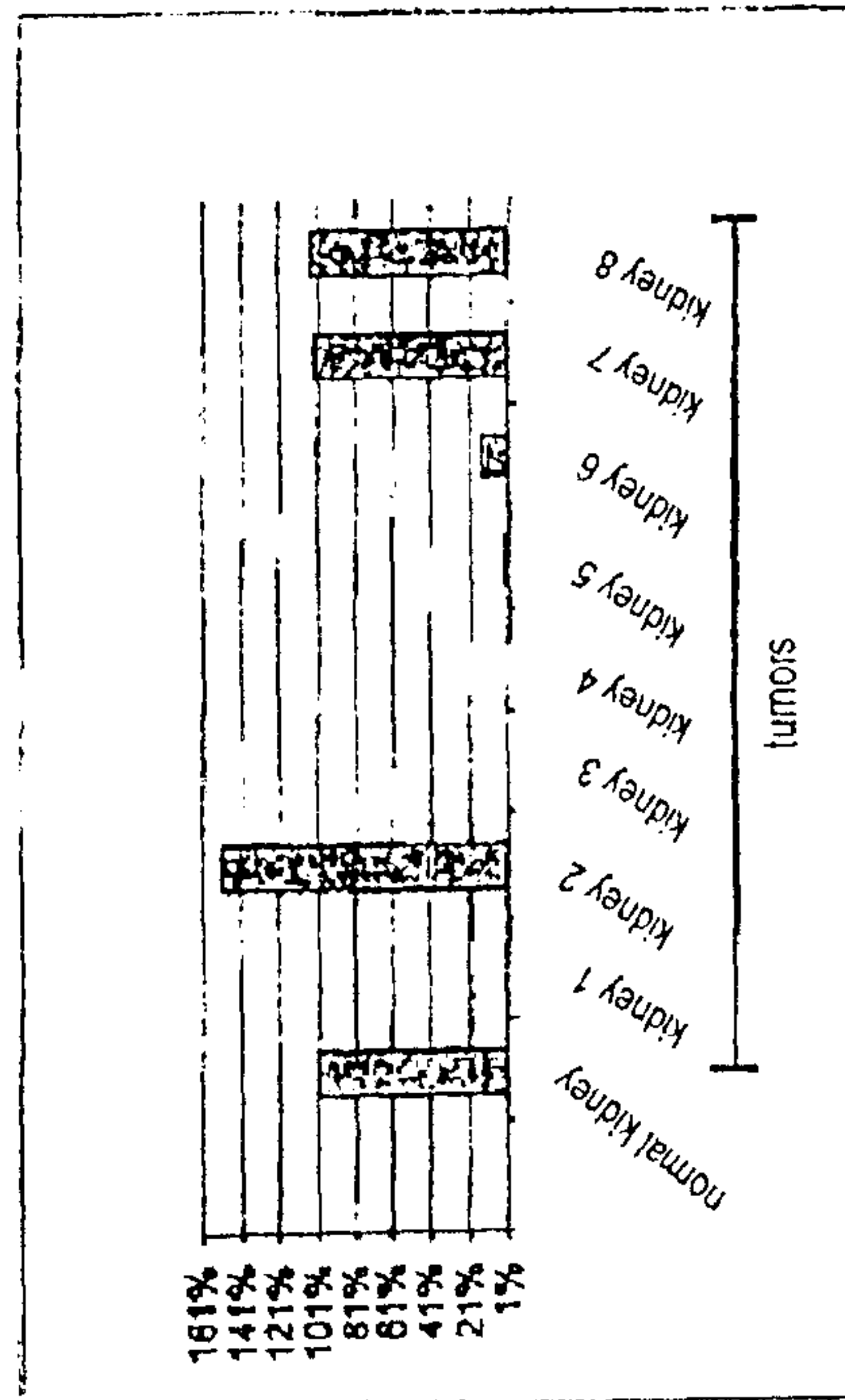
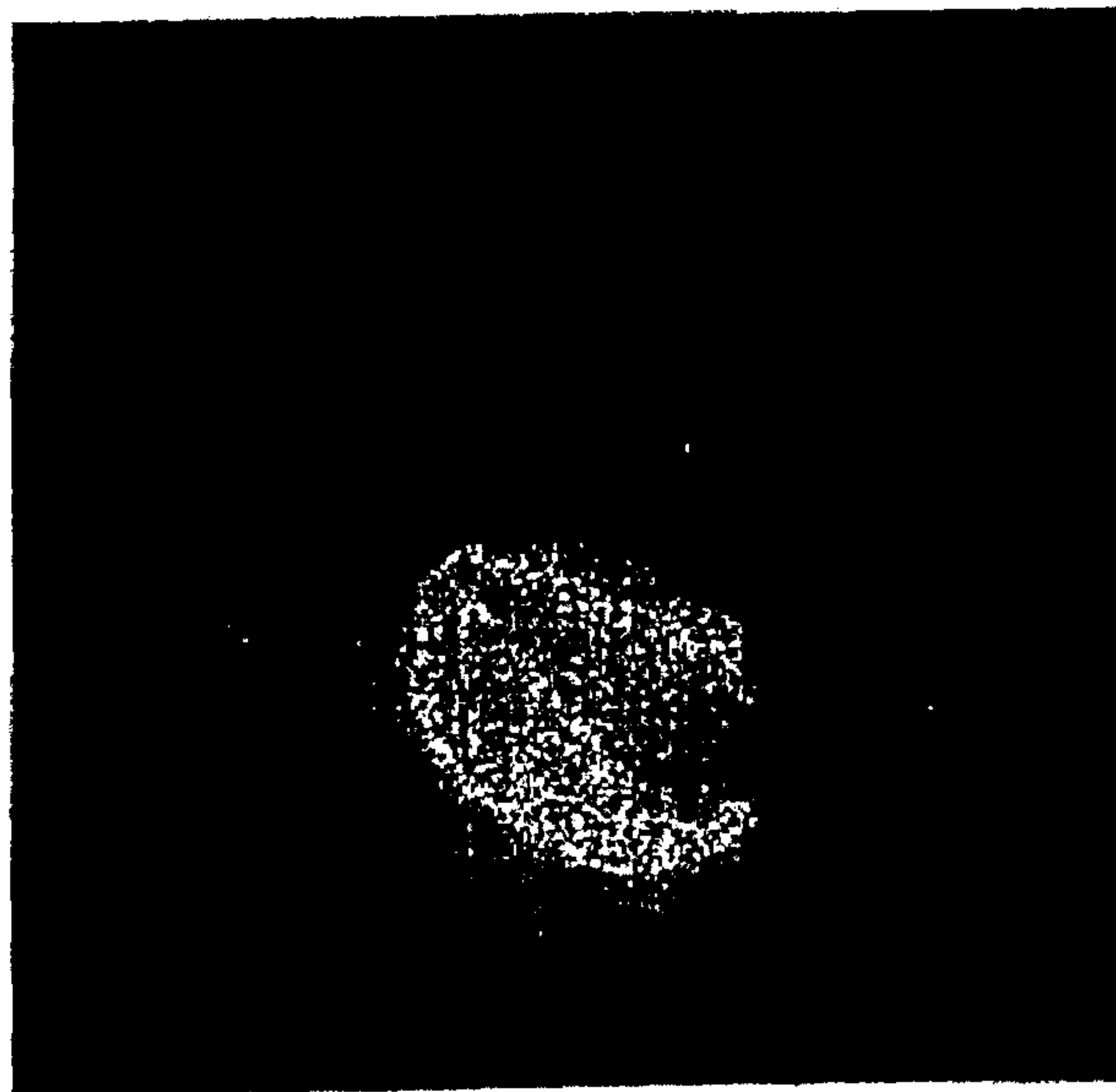


Fig. 34B

Fig. 35



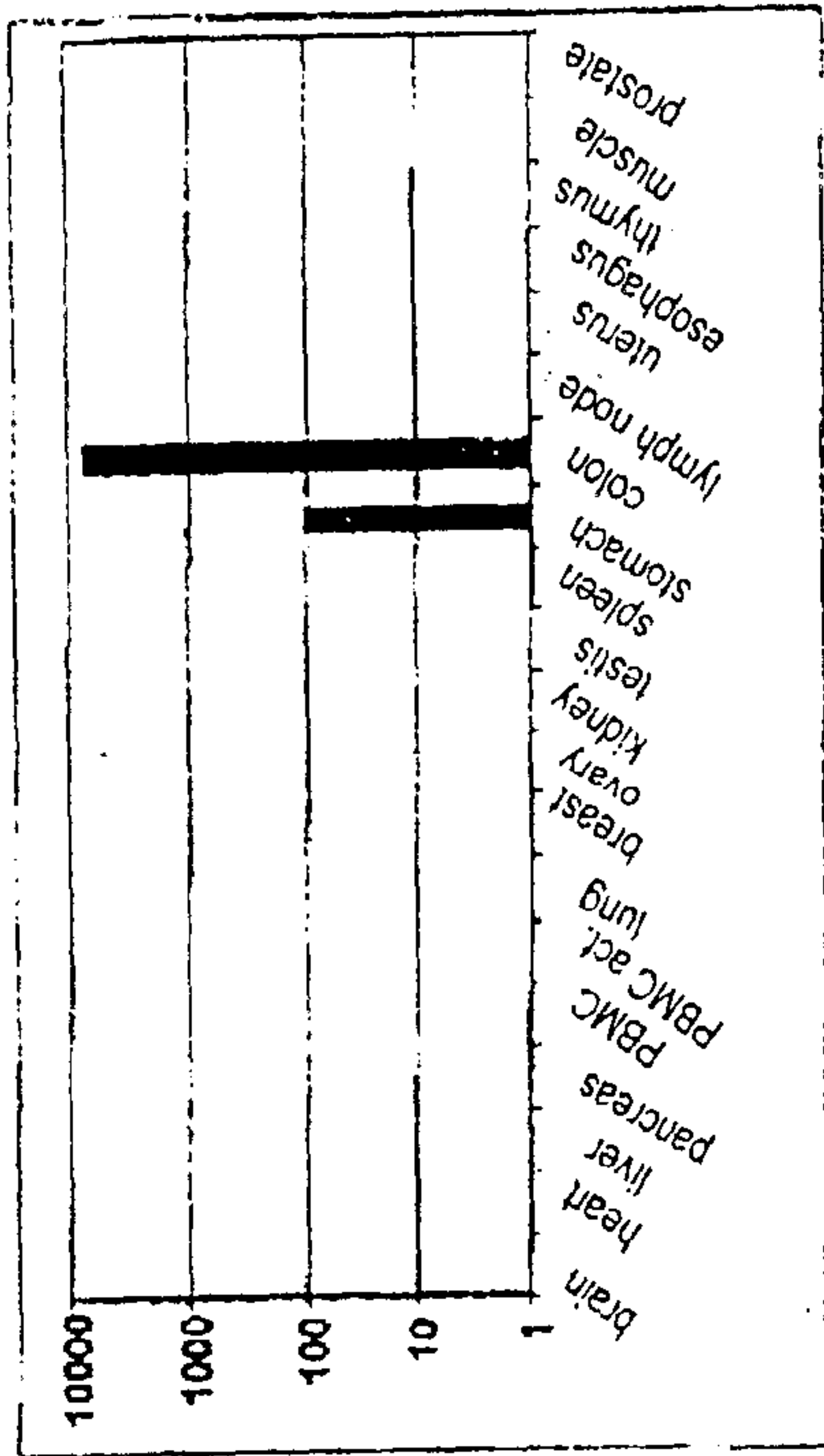


Fig. 36A

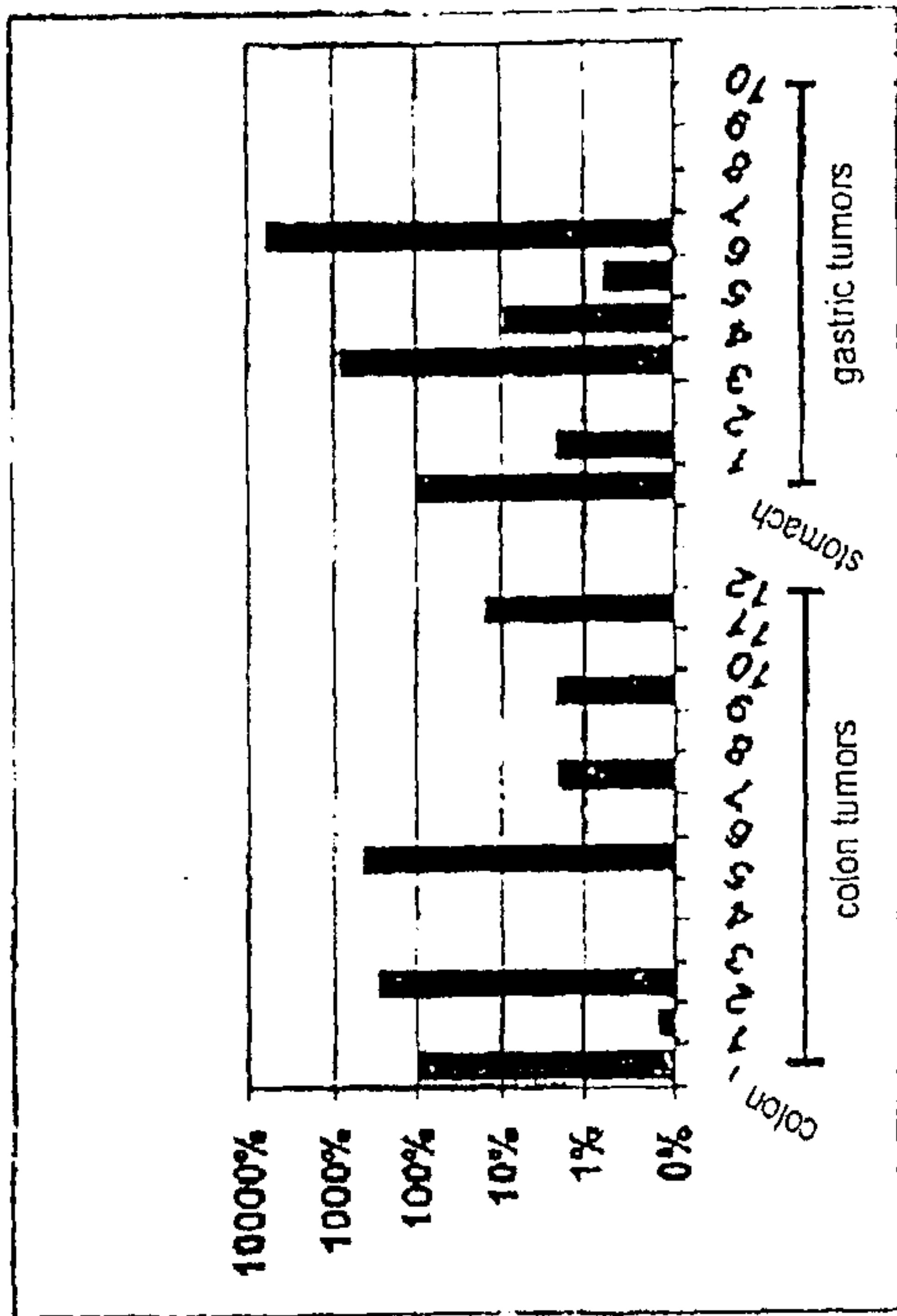


Fig. 36B

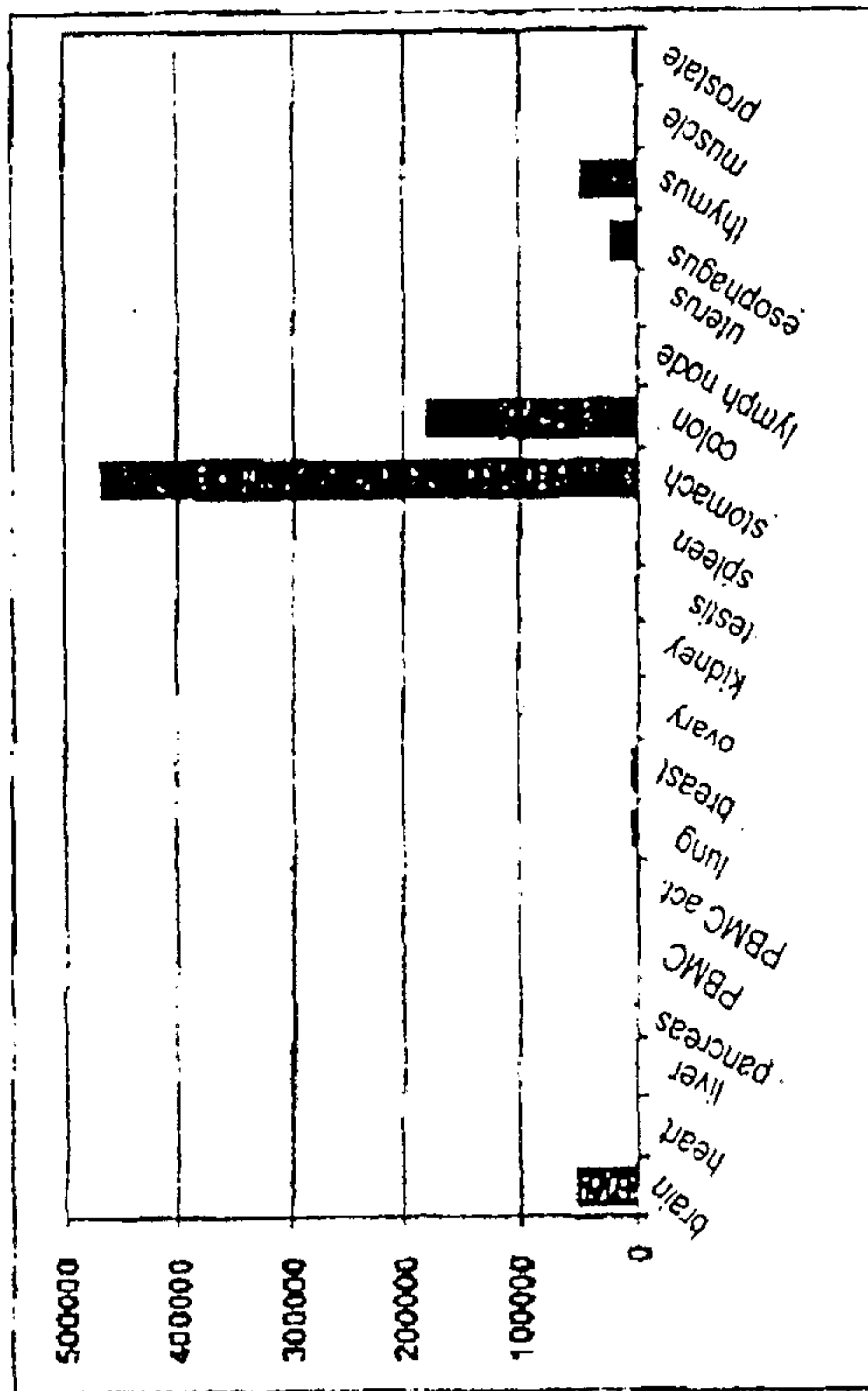


Fig. 37A

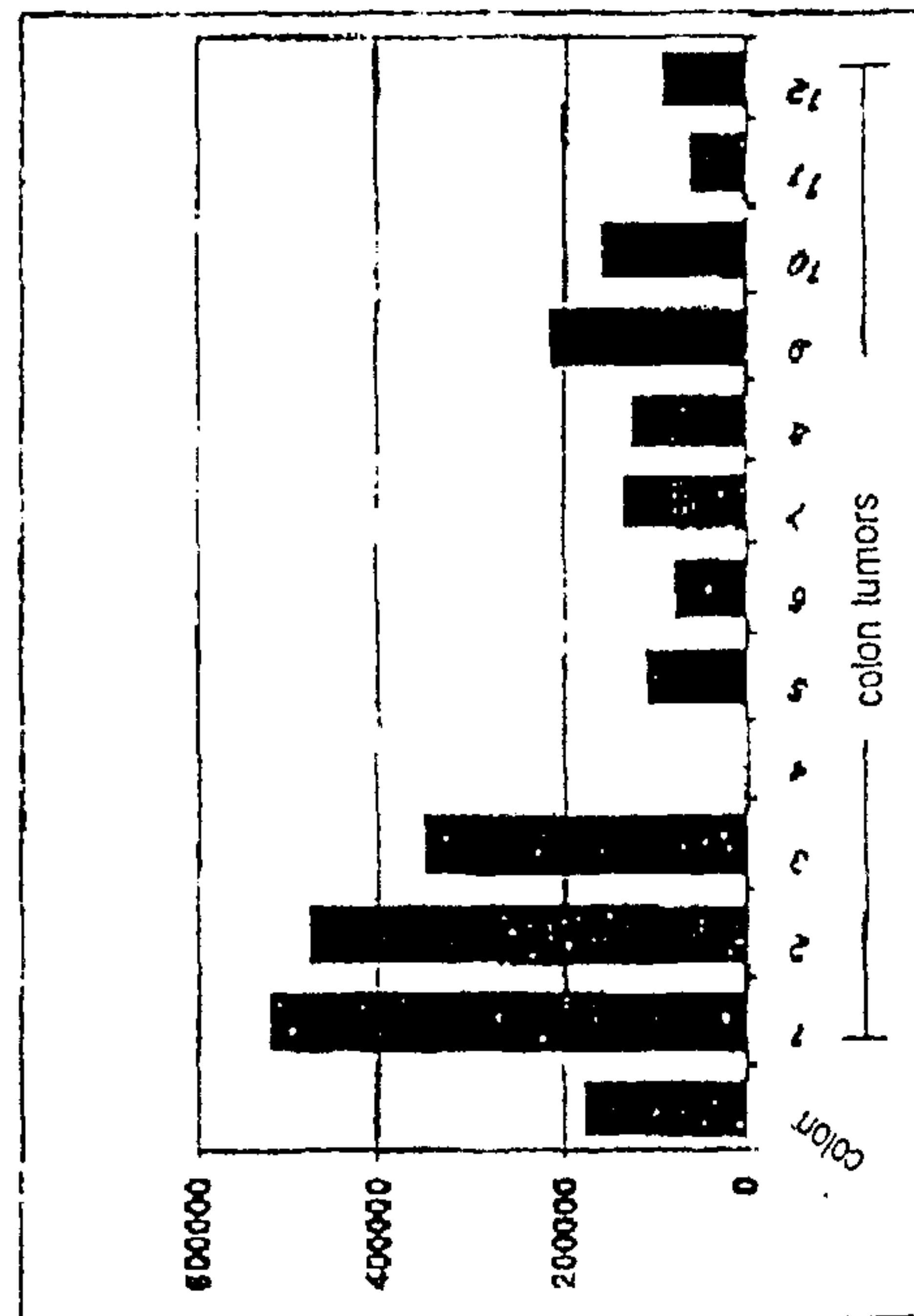


Fig. 37B

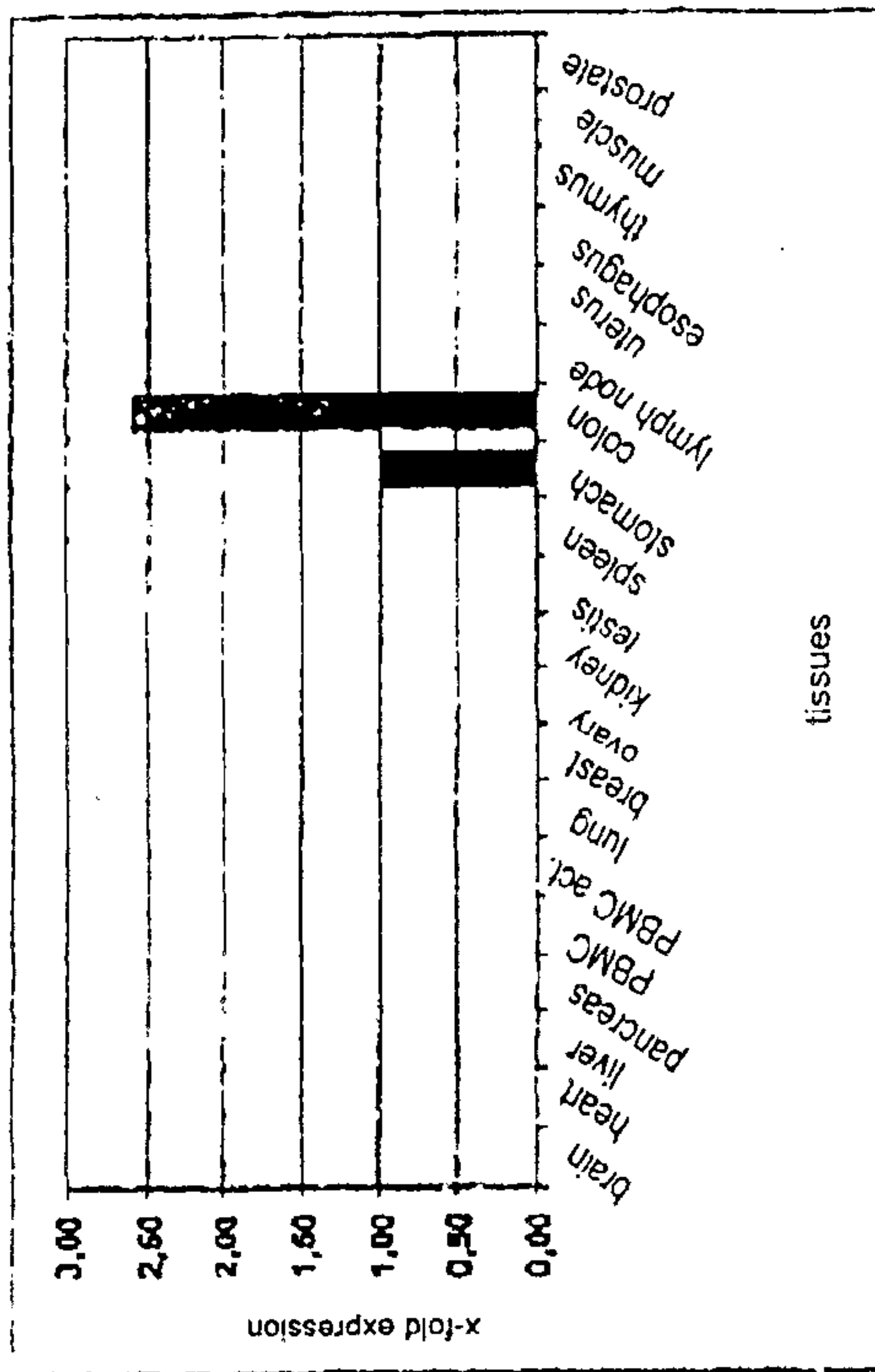


Fig. 38A

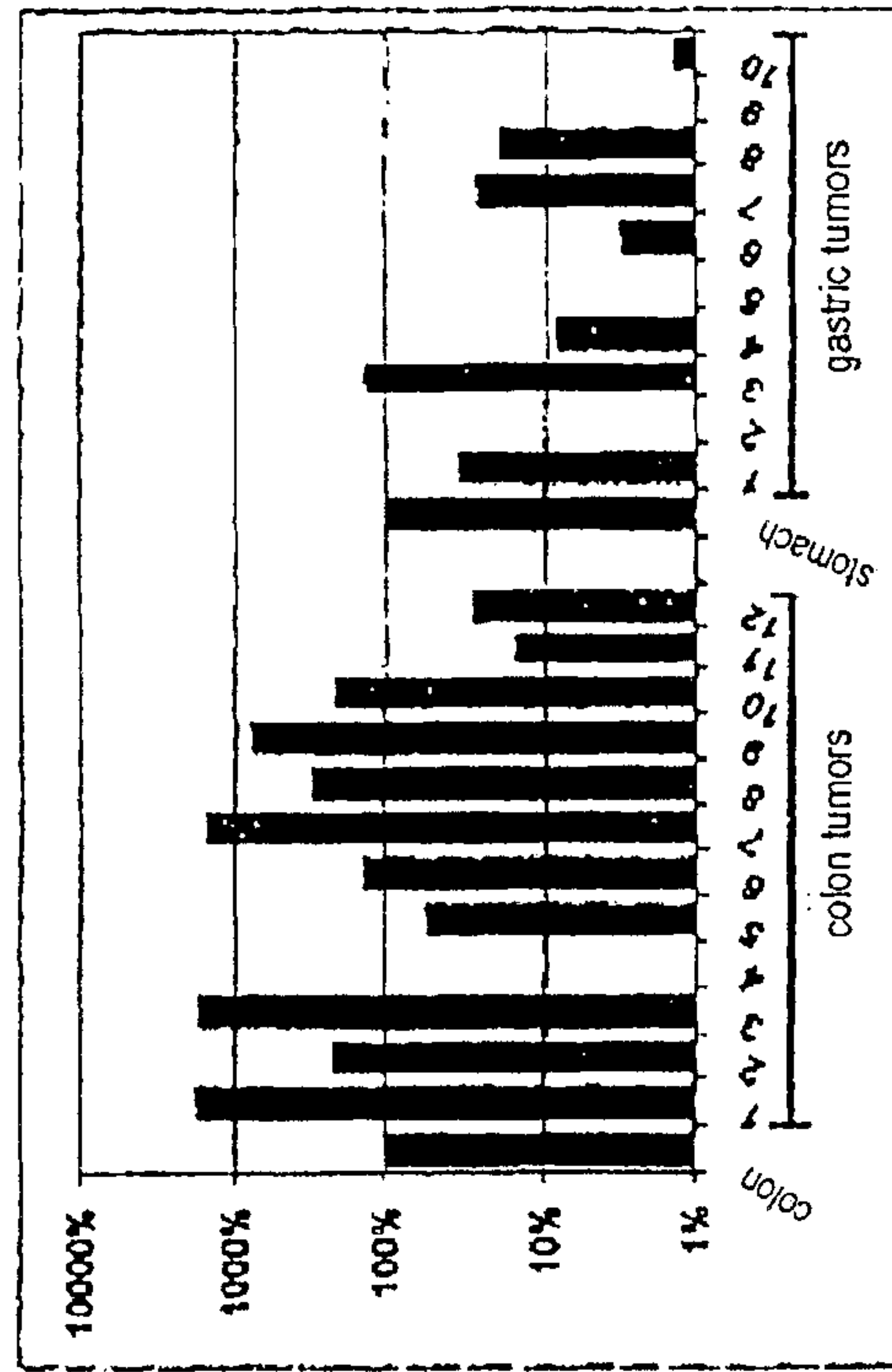


Fig. 38B

Fig. 39

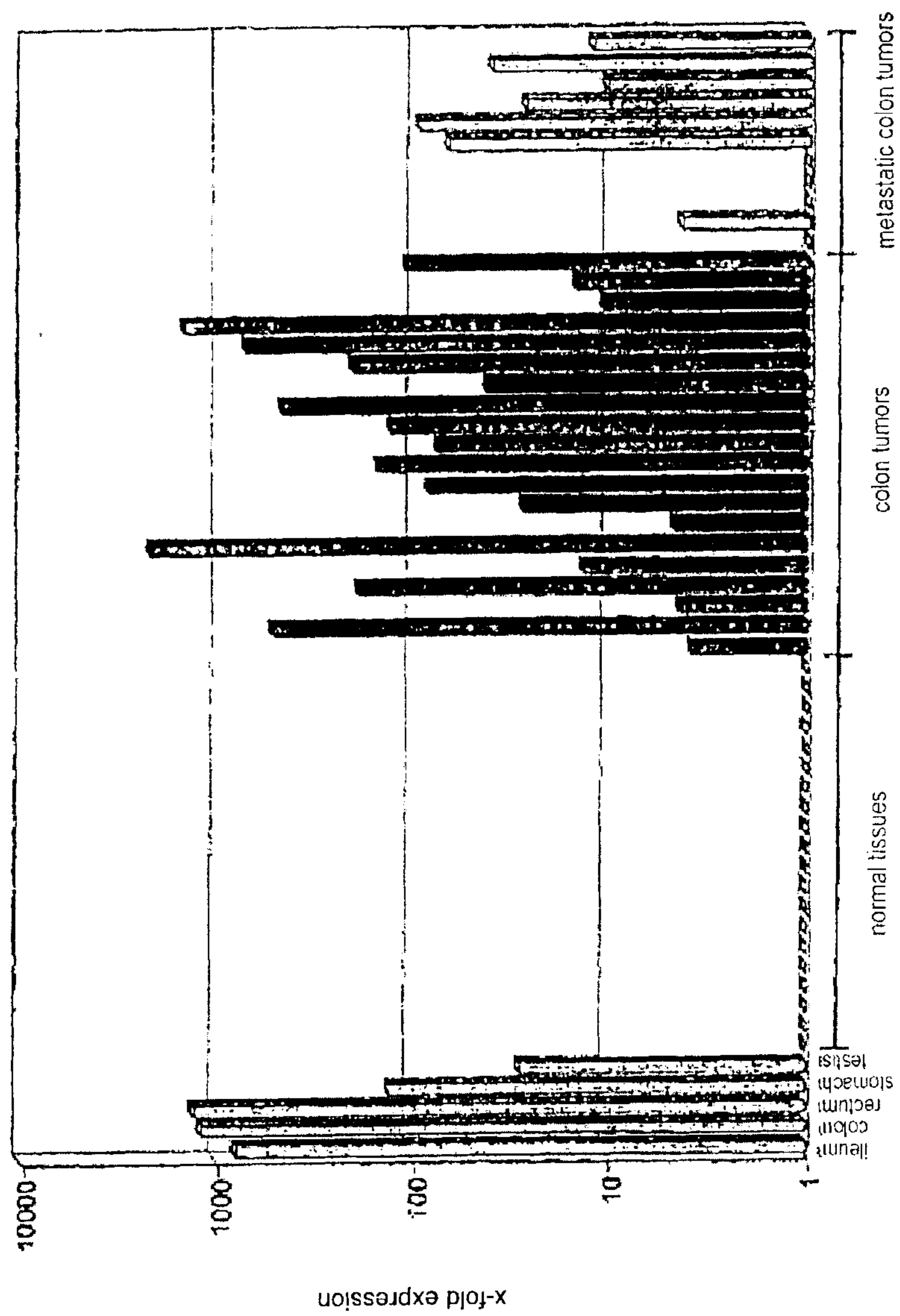
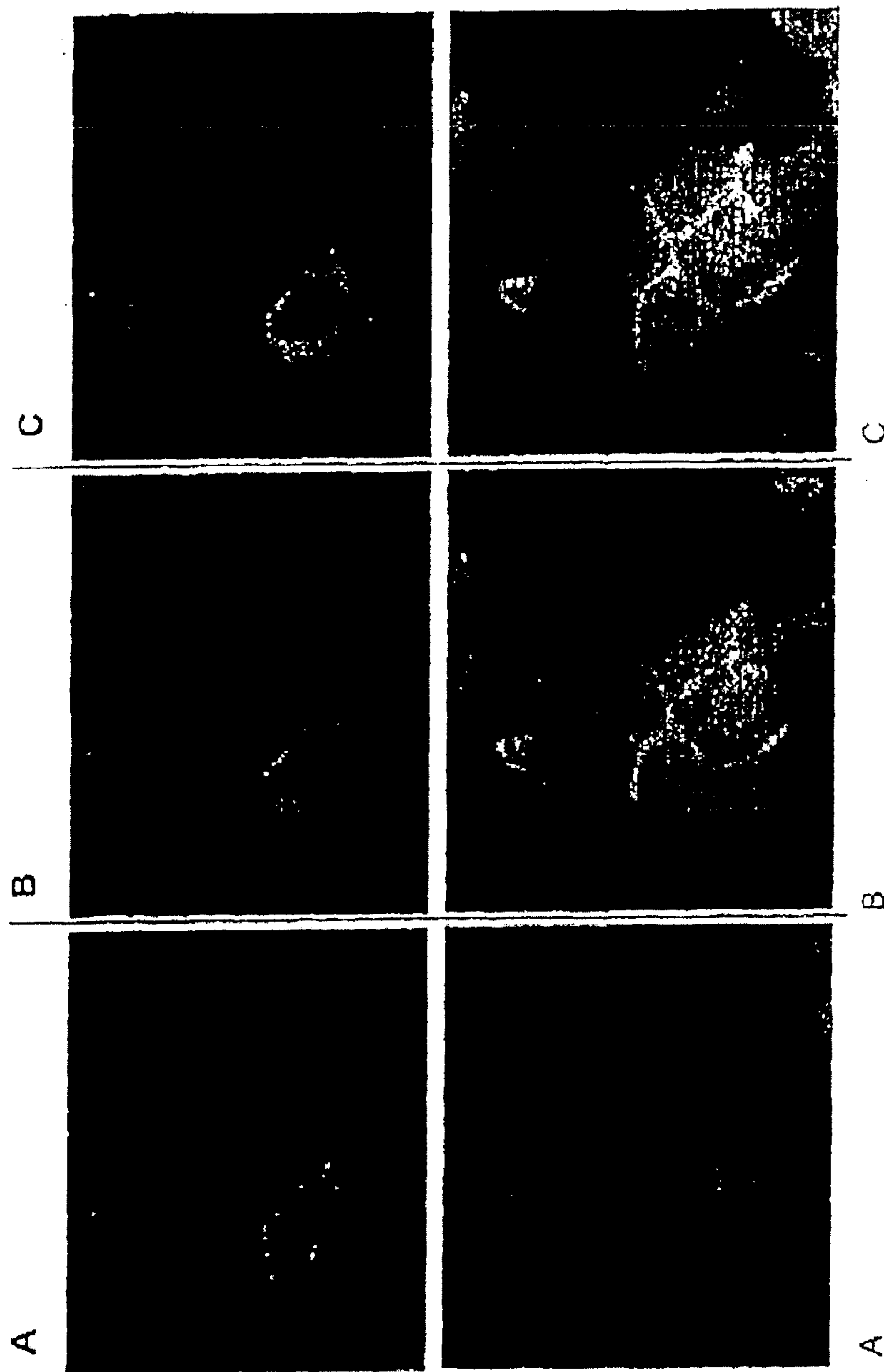


Fig. 40



ID	sequence
#1	caggccagagtcaccagctgtcctggactctgctgtggggaagggctgatgcaggtgtgga gtcaaatgtgggtgcctcctgcagccgggtgccaggaggggtggaggggccaccctgggc tttgtccgggagcctgggtcttcccgtccttgggctgacaggtgctgctgcctctgagccc tccctgctaagagctgtgtgctgggttaaggctgggtggcccttgggctccctgtccagga tttgtgctctggagggtagggcttggctgggctgggactggaggggaacgtggagctcct tctgctcctttcctgccccatgacagcagggcagatcccaggagagaagagctcaggaga tgggaagaggatctgtccaggggttagacctcaagggtgacttggagttctttacggcac ccatgctttctttgaggagttttgtgtttgtgggtgtggggtcggggctcacctcctccc acatccctgcccagaggtgggcagagtgggggcagtgcccttgcctccctgctcgtctc tgcctgacctccggctcctgtgctgcccaggacatgaatggcacctacaacacctgtg gctccagcgacctcacctggccccagcgatcaagctgggcttctacgctacttggggc tctgtggtgctagggctgctgctcaacagcctggcgctctgggtgttctgctgcccga tgcagcagtgagcgagaccccgcatctacatgaccaacctggcggtggccgacctctgcc tgctgtgcaccttgccttctgtgctgacctccctgcgagacacctcagacacggcgtgt gccagctctcccagggcattctacctgaaccaacaggtacatgagcatcagcctggtcacgg ccatcgccgtggacggctatgtggcgtgcccagcccgctgcgtgcccggggctgcccgt ccccaggcaggtgcccggcgtgtgcccggctcctctgggtgctggtcagcgtccctgg tggctcgtggctcctggggattcaggagggcggtctctgcttcaggagcaccggcaca atttcaactccatggcgttcccgtgctgggattctacctgcccctggcgtggtggtct tctgctccctgaagggtggtgactgcccctggcccagagggcaccaccgagctggggcagg cagagggcaccggcaaggctgcccgcattggtctgggccaacctcctggtgttctggtct gcttctgcccctgcagctggggctgacagtgccctcgcagtggggtggaaacgctgtg ccctcctggagacgatccgtcgcgcctgtacataaccagcaagctctcagatgccaact gctgctggagcgccatctgctactactacatggccaaggagttccaggaggcgtctgcac tggcctggtgctcccagtgctaaaggcccacaaaagccaggactctctgtgcgtgacctcg cctaagaggcgtgctgtgggcgtgtgggcccaggctctgggggctccgggaggtgctgce tgccaggggaagctggaaccagtagcaaggagcccgggatcagccctgaactcactgtgt attctcttggagccttgggtgggcagggacggcccagggtacctgctctcttgggaagaga gagggacagggacaagggcaagaggactgagggcagagcaaggccaatgtcagagacccc cgggatggggcctcacacttgcacccccagaaaccagctcacctggccagagtggttcc tgctggccaggggtgcagccttgatgacacctgcccgtgcccctcggggctggaataaaac tccccaccagagtc
#2	ATGAAGACGTTGCTGTTGGACTTGGCTTTGTGGTCACTGCTCTTCCAGCCCGGGTGGCTGTCTTTAGTT CCCAGGTGAGTCAGAACTGCCACAATGGCAGCTATGAAATCAGCGTCCTGATGATGGGCAACTCAGCCTT TGCAGAGCCCCGAAAACTTGGGAAGATGCGGTGAATGAGGGGCTGGAAATAGTGAGAGGACGTCTGCAA AATGCTGGCCTAATGTGACTGTGAACGCTACTTTTCATGTATTCGGATGGTCTGATTTCATAACTCAGGCG ACTGCCGGAGTAGCACCTGTGAAGGCCTCGACCTACTCAGGAAAATTTCAAATGCACAACGGATGGGCTG TGCTCTCATAGGGCCCTCATGTACATACTCCACCTTCCAGATGTACCTTGACACAGAATTGAGCTACCCC ATGATCTCAGCTGGAAGTTTGGATTGTCTGTGACTATAAAGAAACCTTAACCAGGCTGATGTCTCCAG CTAGAAAGTTGATGTACTTCTTGGTTAACTTTTGGAAAACCAACGATCTGCCCTTCAAACTTATTCCTG GAGCACTTCGTATGTTTACAAGAATGGTACAGAACTGAGGACTGTTTCTGGTACCTTAATGCTGTGGAG GCTAGCGTTTCTTATTTCTCCACGAACCTCGGCTTTAAGGTGGTGTAAAGACAAGATAAGGAGTTTCAGG ATATCTTAATGGACCACAACAGGAAAAGCAATGTGATTATTATGTGTGGTGGTCCAGAGTTCTCTACAA GCTGAAGGGTGACCGAGCAGTGGCTGAAGACATTGTCAATTATCTAGTGGATCTTTTCAATGACCAGTAC TTGGAGGACAATGTACAGCCCTGACTATATGAAAAATGTCCTTGTCTGACGCTGTCTCTGGGAATT CCCTTCTAAATAGCTCTTTCTCCAGGAATCTATCACCAACAAAACGAGACTTTGCTCTTGCTTATTTGAA TGGAATCCTGCTCTTTGGACATATGCTGAAGATATTTCTTGAAAATGGAGAAAATATTACCACCCCAAA TTTGCTCATGCTTTTCAAGGAATCTCACTTTTGAAGGGTATGACGGTCCAGTGACCTTGGATGACTGGGGGG ATGTTGACAGTACCATGGTGTCTGTATACCTCTGTGGACACCAAGAAATACAAGGTCTTTTGGACCTA TGATACCCACGTAAATAAGACCTATCCTGTGGATATGAGCCCCACATTCACTTGGAGAAGCTCTAACTT CCTAATGATATTACAGGCCGGGGCCCTCAGATCCTGATGATTGCAGTCTTCACCTCACTGGAGCTGTGG TGCTGCTCCTGCTCGTCCCTCTCTGATGCTCAGAAAATATAGAAAAGATTATGAACCTCGTCAGAAAAA ATGGTCCCACATTCTCTCTGAAAATATCTTTCTCTGGAGACCAATGAGACCAATCATGTTAGCCTCAAG ATCGATGATGACAAAAGACGAGATACAATCCAGAGACTACGACAGTGCAAAATACGACAAAAAGCGAGTGA TTCTCAAAGATCTCAAGCACAATGATGTAATTTCACTGAAAAACAGAAGATAGAATTGAACAAGTTGCT TCAGATTGACTATTACAACCTGACCAAGTTCTACGGCACAGTGAACTTGATACCATGATCTTCGGGGTG ATAGAATACTGTGAGAGAGGATCCCTCCGGGAAGTTTAAATGACACAATTTCTTACCTGATGGCACAT TCATGGATTGGGAGTTTAAAGATCTCTCTGTATGACATTGCTAAGGGGAATGTCAATCTGCACTCCAG TAAGACAGAAGTCCATGGTCTGAAATCTACCAACTGCGTAGTGACAGTAGAATGGTGGTGAAGATC

	ACTGATTTTGGCTGCAATTCCATTTTACCTCCAAAAAGGACCTGTGGACAGCTCCAGAGCACCTCCGCC AAGCCAACATCTCTCAGAAAGGAGATGTGTACAGCTATGGGATCATCGCACAGGAGATCATCTGCGGAA AGAAACCTTCTACACTTTGAGCTGTGCGGACCGGAATGAGAAGATTTTCAGAGTGGAAAATCCAATGGA ATGAAACCCTTCCGCCCAGATTTATTCTTGGAACAGCAGAGGAAAAAGAGCTAGAAGTGTACCTACTTG TAAAAAAGTGTGGGAGGAAGATCCAGAAAAGAGACCAGATTTCAAAAAAATTGAGACTACACTTGCCAA GATATTTGGACTTTTTCATGACCAAAAAAATGAAAGCTATATGGATACCTTGATCCGACGTCTACAGCTA TATCTCGAAACCTGGAACATCTGGTAGAGGAAAGGACACAGCTGTACAAGGCAGAGAGGGACAGGGCTG ACAGACTTAACCTTTATGTTGCTTCCAAGGCTAGTGGTAAAGTCTCTGAAGGAGAAAGGCTTTGTGGAGCC GGAATATATGAGGAAGTTACAATCTACTTCAGTGACATTGTAGGTTTCACTACTATCTGCAAATACAGC ACCCCATGGAAGTGGTGGACATGCTTAATGACATCTATAAGAGTTTTGACCACATTGTTGATCATCATG ATGCTTACAAGGTGGAACCATCGGTGATGCGTACATGGTGGCTAGTGGTTTGCCTAAGAGAAATGGCAA TCGGCATGCAATAGACATTGCCAAGATGGCCTTGGAATCCTCAGCTTCATGGGGACCTTTGAGCTGGAG CATCTTCTGGCCTCCCAATATGGATTGCGATTGGAGTTCACTCTGGTCCCTGTGCTGCTGGAGTTGTGG GAATCAAGATGCCTCGTTATTGTCTATTGGAGATACGGTCAACACAGCCTCTAGGATGGAATCCACTGG CCTCCCTTTGAGAATTCACGTGAGTGGCTCCACCATAGCCATCCTGAAGAGAACTGAGTGCCAGTTCCCTT TATGAAGTGAGAGGAGAAACATACTTAAGGGAAGAGGAAATGAGACTACCTACTGGCTGACTGGGATGA AGGACCAGAAATTCACCTGCCAACCCTCCTACTGTGGAGAATCAACAGCGTTTGCAAGCAGAAATTTTC AGACATGATTGCCAACTCTTTACAGAAAAGACAGGCAGCAGGGATAAGAAGCCAAAAACCAGACGGGTA GCCAGCTATAAAAAAGGCACTCTGGAATACTTGACAGCTGAATACCACAGACAAGGAGAGCACCTATTTT AA
#3	ATGAAGACGTTGCTGTTGGACTTGGCTTTGTGGTCACTGCTCTTCCAGCCCGGGTGGCTGTCTTTAGTT CCCAGGTGAGTCAGAACTGCCACAATGGCAGCTATGAAATCAGCGTCCTGATGATGGGCAACTCAGCCTT TGCAGAGCCCTGAAAAACTTGGAAGATGCGGTGAATGAGGGGCTGGAAATAGTGAGAGGACGTCTGCAA AATGCTGGCCTAAATGTGACTGTGAACGCTACTTTCATGTATTCCGGATGGTCTGATTTCATAACTCAGGCG ACTGCCGGAGTAGCACCTGTGAAGGCTCGACCTACTCAGGAAAATTTCAAATGCACAACGGATGGGCTG TGCTCTCATAGGGCCCTCATGTACATACTCCACCTTCCAGATGTACCTTGACACAGAATTGAGCTACCCC ATGATCTCAGCTGGAAGTTTGGATTGTGACTATAAAGAAACCTTAACCAGGCTGATGTCTCCAG CTAGAAAGTTGATGTACTTCTTGGTTAACTTTTGGAAAACCAACGATCTGCCCTTCAAACCTTATTCCTG GAGCACTTCGTATGTTTACAAGAATGGTACAGAACTGAGGACTGTTTCTGGTACCTTAATGCTCTGGAG GCTAGCGTTTCTTATTTCTCCACGAACCTCGGCTTTAAGGTGGTGTAAAGACAAGATAAGGAGTTTCAGG ATATCTTAATGGACCACAACAGGAAAAGCAATGTGACCAGTACTTGGAGGACAATGTACAGCCCTGAC TATATGA
#4	ATGAAGACGTTGCTGTTGGACTTGGCTTTGTGGTCACTGCTCTTCCAGCCCGGGTGGCTGTCTTTAGTT CCCAGGTGAGTCAGAACTGCCACAATGGCAGCTATGAAATCAGCGTCCTGATGATGGGCAACTCAGCCTT TGCAGAGCCCTGAAAAACTTGGAAGATGCGGTGAATGAGGGGCTGGAAATAGTGAGAGGACGTCTGCAA AATGCTGGCCTAAATGTGACTGTGAACGCTACTTTCATGTATTCCGGATGGTCTGATTTCATAACTCAGGCG ACTGCCGGAGTAGCACCTGTGAAGGCTCGACCTACTCAGGAAAATTTCAAATGCACAACGGATGGGCTG TGCTCTCATAGGGCCCTCATGTACATACTCCACCTTCCAGATGTACCTTGACACAGAATTGAGCTACCCC ATGATCTCAGCTGGAAGTTTGGATTGTGACTATAAAGAAACCTTAACCAGGCTGATGTCTCCAG CTAGAAAGTTGATGTACTTCTTGGTTAACTTTTGGAAAACCAACGATCTGCCCTTCAAACCTTATTCCTG GAGCACTTCGTATGTTTACAAGAATGGTACAGAACTGAGGACTGTTTCTGGTACCTTAATGCTCTGGAG GCTAGCGTTTCTTATTTCTCCACGAACCTCGGCTTTAAGGTGGTGTAAAGACAAGATAAGGAGTTTCAGG ATATCTTAATGGACCACAACAGGAAAAGCAATGTGACCAGTACTTGGAGGACAATGTACAGCCCTGAC TATATGA
#5	ATGAAGACGTTGCTGTTGGACTTGGCTTTGTGGTCACTGCTCTTCCAGCCCGGGTGGCTGTCTTTAGTT CCCAGGTGAGTCAGAACTGCCACAATGGCAGCTATGAAATCAGCGTCCTGATGATGGGCAACTCAGCCTT TGCAGAGCCCTGAAAAACTTGGAAGATGCGGTGAATGAGGGGCTGGAAATAGTGAGAGGACGTCTGCAA AATGCTGGCCTAAATGTGACTGTGAACGCTACTTTCATGTATTCCGGATGGTCTGATTTCATAACTCAGGCG ACTGCCGGAGTAGCACCTGTGAAGGCTCGACCTACTCAGGAAAATTTCAAATGCACAACGGATGGGCTG TGCTCTCATAGGGCCCTCATGTACATACTCCACCTTCCAGATGTACCTTGACACAGAATTGAGCTACCCC ATGATCTCAGCTGGAAGTTTGGATTGTGACTATAAAGAAACCTTAACCAGGCTGATGTCTCCAG CTAGAAAGTTGATGTACTTCTTGGTTAACTTTTGGAAAACCAACGATCTGCCCTTCAAACCTTATTCCTG GAGCACTTCGTATGTTTACAAGAATGGTACAGAACTGAGGACTGTTTCTGGTACCTTAATGCTCTGGAG GCTAGCGTTTCTTATTTCTCCACGAACCTCGGCTTTAAGGTGGTGTAAAGACAAGATAAGGAGTTTCAGG ATATCTTAATGGACCACAACAGGAAAAGCAATGTGATTATTATGTGTGGTGGTCCAGAGTTCTCTACAA GCTGAAGGGTGACCGAGCAGTGGCTGAAGACATGTGATTATTCTAGTGGATCTTTCAATGACAGTAC TTGGAGGACAATGTACAGCCCTGACTATATGAAAATGTCTTGTCTGACGCTGTCTCCTGGGAATT CCCTTCTAAATAGCTCTTTCTCCAGGAATCTATCACCACAAAACGAGACTTTGCTCTTGCTTATTGAA TGGAACTCTGCTCTTTGGACATATGCTGAAGATATTTCTTGAAGATGGAGAAATATTACCACCCCAAA TTTGCTCATGCTTTTCAAGGAATCTCACTTTTGAAGGGTATGACGGTCCAGTGACCTTGGATGACTGGGGG ATGTTGACAGTACCATGGTGTCTGTATACCTCTGTGGACACCAAGAAATACAAGGTTCTTTTGACCTA TGATACCCACGTAAATAAGACCTATCCTGTGGATATGAGCCCACTTCACTTGGGAAGAACTCTAACTT CCTAATGATATTACAGGCCCGGGGCCCTCAGATCCTGATGATTGCAGTCTTACCCTCACTGGAGCTGTGG TGCTGCTCCTGCTCGTCTCTCTGATGCTCAGAAAATATAGAAAAGATTATGAACCTTCGTGAGAAA ATGGTCCACATTCCTCCTGAAAATATCTTCTCTGGAGACCAATGAGACCAATCATGTTAGCCTCAAG ATCGATGATGACAAAAGACGAGATACAATCCAGAGACTACGACAGTGCAATACGACAAAAGCGACTGA

	TTCTCAAAGATCTCAAGCACAATGATGGTAATTTCACTGAAAAACAGAAGATAGAATTGAACAAGATTGA CTATTACAACCTGACCAAGTTCTACGGCACAGTGAACTTGATACCATGATCTTCGGGGTGATAGAATAC TGTGAGAGAGGATCCCTCCGGGAAGTTTTAAATGACACAATTTCTACCTGATGGCACATTCATGGATT GGGAGTTTAAAGATCTCTGTCTTGTATGACATTGCTAAGGGAATGTCATATCTGCACCTCCAGTAAGACAGA AGTCCATGGTCGTCTGAAATCTACCAACTGCGTAGTGGACAGTAGAATGGTGGTGAAGATCACTGATTTT GGCTGCAATTCCATTTTACCTCCAAAAAGGACCTGTGGACAGCTCCAGAGCACCTCCGCCAAGCCAACA TCTCTCAGAAAGGAGATGTGTACAGCTATGGGATCATCGACAGGAGATCATTCCTGCGGAAAGAAACCTT CTACACTTTGAGCTGTGGGACCGGAATGAGAAGATTTTCAGAGTGGAAAATTCGAATGGAATGAAACCC TTCCGCCAGATTTATTCTTGGAACAGCAGAGGAAAAGAGCTAGAAGTGTACCTACTTGTAAAAAACT GTTGGGAGGAAGATCCAGAAAAGAGACCAGATTTCAAAAAAATTGAGACTACACTTGCCAAGATATTTGG ACTTTTTCATGACCAAAAAATGAAAGCTATATGGATACCTTGATCCGACGTCTACAGCTATATTCTCGA AACCTGGAACATCTGGTAGAGGAAAGGACACAGCTGTACAAGGCAGAGAGGGACAGGGCTGACAGACTTA ACTTTATGTTGCTTCCAAGGCTAGTGGTAAAGTCTCTGAAGGAGAAAGGCTTGTGGAGCCGGAACATA TGAGGAAGTTACAATCTACTTCAGTGACATTGTAGGTTTCACTACTATCTGCAAATACAGCACCCCATG GAAGTGGTGGACATGCTTAATGACATCTATAAGAGTTTTGACCACATTGTTGATCATCATGATGTCTACA AGGTGGAACCATCGGTGATGCGTACATGGTGGCTAGTGGTTTGCCTAAGAGAAATGGCAATCGGCATGC AATAGACATTGCCAAGATGGCCTTGGAATCCTCAGCTTCATGGGGACCTTTGAGCTGGAGCATCTTCCT GGCCTCCCAATATGGATTTCGATTGGAGTTCACCTCTGGTCCCTGTGCTGCTGGAGTTGTGGGAATCAAGA TGCCTCGTTATTGTCTATTGGAGATACGGTCAACACAGCTCTAGGATGGAATCCACTGGCCTCCCTTT GAGAATTCACGTGAGTGGCTCCACCATAGCCATCCTGAAGAGAACTGAGTGCCAGTTCCCTTATGAAGTG AGAGGAGAAACATACTTAAAGGGAAGAGGAAATGAGACTACCTACTGGCTGACTGGGATGAAGGACCAGA AATTCAACCTGCCAACCCCTCCTACTGTGGAGAATCAACAGCGTTTGCAAGCAGAAATTTTCAGACATGAT TGCCAACTCTTTACAGAAAAGACAGGCAGCAGGGATAAGAAGCCAAAAACCCAGACGGGTAGCCAGCTAT AAAAAGGCACCTCTGGAATACTTGCAGCTGAATACCACAGACAAGGAGAGCACCTATTTTTTAA
#6	ggggacactttgtatggcaagtggaaacctggcttggtg gattttgctagatttttctgatttttaaaactcctgaaaaatatccagataactgtcatgaagctggtaacta tcttct gctggtgacctcagcctttgtagttactctgctactgcc ttctcatcaacaaagtggcccttctgttgacaagttggcacctttacctctggacaacattcttcccttta tggatcc attaaagcttcttctgaaaactctgggcatttctgttgag caccttgtggaggggctaaggaagtgtgtaaatgagctgggaccagaggcttctgaagctgtgaagaaactgc tggaggc gctatcacacttgggtgtgacatcaagataaagagcggagg tggatggggatggaagatgatgtcctatcctccctgcctgaaacctgttctaccaattatagatcaaagtc ctaaatgtagtgacctgtgaaaaggacaaataaagcaatgaatactaaaaa aaaaa
#7	ATGGCCGTGACTGCCTGTCAGGGCTTGGGGTTCTGTTTCACTGATTGGGATTGCGGGCATCATTGCTG CCACCTGCATGGACAGTGGAGCACCCAAGACTTGTACAACAACCCGTAACAGCTGTTTCAACTACCA GGGGCTGTGGCGCTCCTGTGTCCGAGAGAGCTCTGGCTTCACCGAGTCCCGGGGCTACTTCACCTGCTG GGGCTGCCAGCCATGCTGCAGGCAGTCCGAGCCCTGATGATCGTAGGCATCGTCCTGGGTGCCATTGGCC TCCTGGTATCCATCTTGGCCTGAAATGCATCCGCATTGGCAGCATGGAGGACTCTGCCAAGCCAAACAT GACACTGACCTCCGGGATCATGTTTATTGTCTCAGGTCTTTGTGCAATTGCTGGAGTGTCTGTGTTGCC AACATGCTGGTGACTAACTTCTGGATGTCCACAGCTAACATGTACACCGGCATGGGTGGGATGGTGCAGA CTGTTGAGACCAGGTACACATTTGGTGCGGCTCTGTTCTGGGCTGGGTGCTGGAGGCCTCACACTAAT TGGGGGTGTGATGATGTGCATCGCCTGCCGGGCTGGCACCAGAAGAAACCAACTACAAAGCCGTTTCT TATCATGCCTCAGGCCACAGTGTTCCTACAAGCCTGGAGGCTTCAAGGCCAGCACTGGCTTTGGGTCCA ACACCAAAAACAAGAAGATATACGATGGAGGTGCCCCGACAGAGGACGAGGTACAATCTTATCCTTCCAA GCACGACTATGTGTAA
#8	tgcgccaccatggcctgactgcctgtcagggttggggttcgtggtttcactgattggg attgcgggcatcattgtctgcacotgcatggaccagtggagcaccacagacttgtacaac aaccctgtaacagctgttttcaactaccaggggtgtggtgctcctgtgtccgagagagc
#9	MNGTYNTCGSSDLTWPPAIKLGFYAYLGVLLVLGLLLNSLALWVFCCRMQQWTETRIYMT NLAVADLCLLCTLPFVLHSLRDTSDTPLCQLSQGIYLTNRYMSISLVTALAVDRYVAVRH PLRARGLRSPQAAAVCAVLWVLVIGSLVARWLLGIQEGGFCFRSTRHNENSMRFPLLGF

	YLPLAVVVFCSLKVVTTALAQRPPPTDVGQAEATRKAARMVWANLLVFVVCFLPLHVGLTVR LAVGWNACALLETIRRALYITSKLSDANCCLDACYYYMAKEFQEASALAVAPRAKAHKS QDSL CVTLA
#10	MTAGRSQERRAQEMGRGSVQGLDLKGDLEFFTAPMLSLRSFVFVGVGSGLTSSHIPAQRKAEWGQCLAPPARS LLTEGSLCCPRTMNGTYNTCGSSDLTWPPAIKLGFIAYLGVLLVLGL LINSALWVFCRMQQTETRIYMTNLAVADLCLLCTLPFVLHSLRDTSDTPLCQLSQGI YLTNRYMSISIVTAIAVDYVAVRHPLRAGLRSPROAAVCAVLWVLVIGSLVARWLLG IQEGGFCFRSTRHNFNMAFFLLGFYLPPLAVVVFCSLKVVTTALAQRPPPTDVGQAEATRKA ARMVWANLLVFVVCFLPLHVGLTVRLAVGWNACALLETIRRALYITSKLSDANCCLDAC YYMAKEFQEASALAVAPSAKAHKSQDSL CVTLA
#11	MKTLLDLALWSLLFQPGWLSFSSQVSQNCNHSYEISVLMMGNSAFAEPLKNLEDAVNEGLEIVRGRLO NAGLNVTVNATFMYSDGLIHNSGDCRSSTCEGLDLLRKISNAQRMGCVLIGPSTYSTFQMYLDTLSYP MISAGSFGLSCDYKETLTRIMSPARKLMYFLVNEFKTNDLFFKTYSWSTSYVYKNGTETEDCFWYLNAL ASVSYFSHELGFKVLRQDKEFQDILMDHNRKSNVIIMCGGPEFLYKLGDRVAEDIVIIIVDLFNDQY LEDNVTAPDYMKNVLVLTLSPGNSLLNSSFSRNLSPTKRDFALAYLNGILLFGHMLKIFLNGENITTPK FAHAFRNLTFEGYDGPVTLDDWGDVDSTMVLLYTSVDTKKYKVLTYDTHVNKTYFVDMSPFTWKNKSL PNDITGRGPQILMIAVFTLTGAVVLLLVALLMLRKYRKDYELRQKKWSHIPPENIFPLETNETNHSVSLK IDDDKRRDTIQRLRQCKYDKKRVILKDLKHNDGNFTEKQKIELNKLLQIDYNNLTKFYGTVKLDTMIFGV LEYCERGSLEVLNDTISYPDGTDFMDWEFKISVLYDIAKGMSYLHSSKTEVHGRLKSTNCVVDSDRMVVKI TDFGCNSILPPKKDLWTAPHEHLRQANISQKGDVYSYGIIAQEIIIRKETFYTLSCRDRNEKIFRVENSNG MKPFRPDLFLETAEEKELEVYLLVKNCWEEDEKRPDFKKIETTLAKIFGLFHDQKNESYMDTLIRRLQL YSRNLHLVEERTQLYKAERDRADRLNFMLLPRLVVKSLKEKGFVEPELYEEVTIYFSDIVGFTTICKYS TPMEVDMNDIYKSFHDHVDHVDYKVTIGDAYMVASGLPKRNGNRHAIDIAKMALEILSFMGTFELE HLPGLPIWIRIGVHSGPCAAGVVGIMPRYCLFGDTVNTASRMESTGLPLRIHVSGSTIAILKRTECQFL YEVGETYLGKRGNETTYWLTGMKDQKFNLPPTVENQORLQAEFSDMIANSLOKROAAGIRSQKPRRV ASYKGTLEYLQLNTTDKESTYF*
#12	MKTLLDLALWSLLFQPGWLSFSSQVSQNCNHSYEISVLMMGNSAFAEPLKNLEDAVNEGLEIVRGRLO NAGLNVTVNATFMYSDGLIHNSGDCRSSTCEGLDLLRKISP*
#13	MKTLLDLALWSLLFQPGWLSFSSQVSQNCNHSYEISVLMMGNSAFAEPLKNLEDAVNEGLEIVRGRLO NAGLNVTVNATFMYSDGLIHNSGDCRSSTCEGLDLLRKISNAQRMGCVLIGPSTYSTFQMYLDTLSYP MISAGSFGLSCDYKETLTRIMSPARKLMYFLVNEFKTNDLFFKTYSWSTSYVYKNGTETEDCFWYLNAL ASVSYFSHELGFKVLRQDKEFQDILMDHNRKSNVTSTWRTMSQPLTI*
#14	MKTLLDLALWSLLFQPGWLSFSSQVSQNCNHSYEISVLMMGNSAFAEPLKNLEDAVNEGLEIVRGRLO NAGLNVTVNATFMYSDGLIHNSGDCRSSTCEGLDLLRKISNAQRMGCVLIGPSTYSTFQMYLDTLSYP MISAGSFGLSCDYKETLTRIMSPARKLMYFLVNEFKTNDLFFKTYSWSTSYVYKNGTETEDCFWYLNAL ASVSYFSHELGFKVLRQDKEFQDILMDHNRKSNVIIMCGGPEFLYKLGDRVAEDIVIIIVDLFNDQY LEDNVTAPDYMKNVLVLTLSPGNSLLNSSFSRNLSPTKRDFALAYLNGILLFGHMLKIFLNGENITTPK FAHAFRNLTFEGYDGPVTLDDWGDVDSTMVLLYTSVDTKKYKVLTYDTHVNKTYFVDMSPFTWKNKSL PNDITGRGPQILMIAVFTLTGAVVLLLVALLMLRKYRKDYELRQKKWSHIPPENIFPLETNETNHSVSLK IDDDKRRDTIQRLRQCKYDKKRVILKDLKHNDGNFTEKQKIELNKIDYNNLTKFYGTVKLDTMIFGVIEY CERGSLEVLNDTISYPDGTDFMDWEFKISVLYDIAKGMSYLHSSKTEVHGRLKSTNCVVDSDRMVVKITDF GCNSILPPKKDLWTAPHEHLRQANISQKGDVYSYGIIAQEIIIRKETFYTLSCRDRNEKIFRVENSNGMKP FRPDLFLETAEEKELEVYLLVKNCWEEDEKRPDFKKIETTLAKIFGLFHDQKNESYMDTLIRRLQLYSR NLEHLVEERTQLYKAERDRADRLNFMLLPRLVVKSLKEKGFVEPELYEEVTIYFSDIVGFTTICKYSTPM EVVDMNDIYKSFHDHVDHVDYKVTIGDAYMVASGLPKRNGNRHAIDIAKMALEILSFMGTFELEHLP GLPIWIRIGVHSGPCAAGVVGIMPRYCLFGDTVNTASRMESTGLPLRIHVSGSTIAILKRTECQFLYEV RGETYLGKRGNETTYWLTGMKDQKFNLPPTVENQORLQAEFSDMIANSLOKROAAGIRSQKPRRVASY KGTLEYLQLNTTDKESTYF*
#15	MKLVTIFLLVTISLCSYSATAKLINCKPLPVDKLAFLPLDNILPFMDPLK LLLKTLGISVEHLVEGLRKCVNELGPEASEAVKKLEALSHLV
#16	MAVTACQGLGFVVSLLIGIAGIIAATCMDQWSTQDLYNNPVTAVFNYQGLWRSCVRESSGFTECRGYFTLL

	GLPAMLQAVRALMIVGIVLGAIGLLVSI FALKCIRIGSMEDSAKANMTLTSGIMFIVSGLCAIAGVSVFA NMLVTNFWMSTANMYTGMGGMVQTVQTRYTFGAALFVGWVAGGLTLIGGVMMCIACRGLAPEETNYKAVS YHASGHSVAYKPGGFKASTGFGSNTKNKKIYDGGARTEDEVQSYPSKHDYV*
#17	DQWSTQDLYN
#18	NNPVTAVFNYQ
#19	MAVTACQGLGFVVSLLIGIAGIIAATCMDQWSTQDLYNPNPVTAVFNYQ
#20	AGGTACATGAGCATCAGCCTG
#21	GCAGCAGTTGGCATCTGAGAG
#22	GCAATAGACATTGCCAAGATG
#23	AACGCTGTTGATTCTCCACAG
#24	GGATCCTCCTTTAGTTCCCAGGTGAGTCAGAAC
#25	TGCTCTGGAGGCTAGCGTTTC
#26	ACCAATCATGTTAGCCTCAAG
#27	AGCTATGGGATCATCGCACAG
#28	CCTTTGAGCTGGAGCATCTTC
#29	CTTTCTAGCTGGAGACATCAG
#30	CACCATGGTACTGTCAACATC
#31	ATGTCATACAAGACAGAGATC
#32	TCTGCCTTGTACAGCTGTGTC
#33	TCTGTGGTATTGAGCTGCAAG
#34	TACTCAGGAAAATTCACCTTG
#35	GACCACAACAGGAAAAGCAATGTGACC

#36	GATAGAATTGAACAAGATTGAC
#37	CAGCCTTTGTAGTTACTCTGC
#38	TGTCACACCAAGTGTGATAGC
#39	GGTTCGTGGTTTCACTGATTGGGATTGC
#40	CGGCTTTGTAGTTGGTTTCTTCTGGTG
#41	ctattgaagccacctgctcaggacaatgaaattcttcagttacattctggtttatcgccg atctctcttctggtgttttactgtgttgggttttactacctctgcccacgtccctccacac caaggaagcagaatgtgcctacacactctttgtggtcgccacattttgggtcacagaagc attgcctctgtcggtaacagctttgtacctagttaatgttaccatgtttgggatcat gccttctaagaaggtggcatctgcttattttcaaggattttcacttactgctaattggagt tatctgttttagcaacatccatagaaaaatggaatttgacaaagagaattgctctgaaaat ggatgatggttgggtgtaaatcctgcatggctgacgctggggttcagagcagcactgc ctttttgtctatgtggctcagcaacacctcgacggctgocatggatgcccattgcccga ggctgtagtgcagcagatcatcaatgcagaagcagaggtcgaggccactcagatgactta cttcaacggatcaaccaaccacggactagaaattgatgaaagtgttaatggacatgaaat aatgagaggaagagaaaaacaaaccagttccaggatataaataatgatacagggaat ttcaagcaaggtggagttggaaaagaactcaggcatgagaaccaaataatcgaacaaagaa gggccacgtgacacgttaacttacgtgtttgtgcatgctactcttctaccattggtgg actgacaacaatcactggtacctccaccaacttgatctttgcagagtatttcaatacacg ctatcctgactgtcgttgcctcaactttggatcatggtttacgttttcttcccagctgc ccttatcattctactcttctcctggatctggtctcagtggttttcttctaggattcaattt taaggagatgttcaaagtgtggcaaaaccaaaccagttccaaacaaagcttgtgctgaggt gattaagcaagaataccaaaagcttgggccaataaggtatcaagaattgtgacctggt cctcttcattataatggtctgtctatggttttagtcgagaccccggatttgttctggttg gtctgcacttttttcagagtacctggttttgcacagattcaactgttgccttacttat agggctgctattcttcttctcagctaaagacactgactaaaactacacctacaggaga aattgttgccttttgattactctccactgattacttggaaagaattccagtcattcatgcc ctgggatatagccattcttgttgggtggagggtttgcctggcagatggttgtgaggagtc tggattatctaatggataggaataaattatctctctggttccattaccagcatggct aataattctgatattcttcttctggtgacatctttaactgaggttagccagcaatccagc taccattacactcttctctccaatattatctccattggccgaagccattcatgtgaacc tctttatattctgataccttctactctgtgtacttcaatttgcattctctctaccagtagc aaatccaccaatgctattgtcttttcatatggtcatctgaaagtcattgacatggttaa agctggacttgggtgtcaacattgttgggtgttgcgtgtggttatgcttggcatatgtaactg gattgtacctatgtttgacctctacacttaccttctggtggctcctgctatgagtaatga gacctgccataataagcacaataattctgactatcttgcggtaatttctggaagacatt aatgattgactgtaaaatgtggctctaaataaactaatgacacacatttaaatcagttatg gtgtagctgtgcaattcccggtgaatacccgaaacctgctggtataactcagagtcacata tttgttattgcagtgcactaaagagcatctatgtgccttcatcaagaagcccatgtttt gagattttgctcatgaaccatctgcaacttgccttcatcataagaataatttataacttga ccttcaaagagatttagagcatttgtttcatcttacagttggagttcaatgtaacatttta aatgcaatttattatttccagaaatttcccatgaaactaaaaatagaaaaataagatataca agtttaattcgttacttggataaatcatttctgcattgttgttccagagaatttgcctgaga aatcaaagccatggtcatctggtgatgaagagaaaaggttaattctaaatgatatgtgcat ttcctcatttataaaaaatccaattggattattcttaatatatacatgtaatatgaaaattg agattgaagcactaattccaaaattatggctgaatataactaaataacagaaaagttacag ataagaatttatttctactgaactctatagtttagtgtaataatcatatttttatgat attggcacactgagaaattcattttgtagagctatggataaggcttgcctatgatttgcac tattagtacagtatagttagaagggaagctgaacactataaaactattaacataattttc gtatatgagtaacaactttgcttaagtgtttatcttagttcagaaatacataatgtcata tggttaaaaatasagagatgtagaaatctaaatgaattatcactgtgtatacagacagaaa aatcacataactctggtgtgttaacattgcaatgaaaaaatgaaaaaagaaggaaaaaa gaataagaatgaaaactgctgacgtattacaaaacagaaaaataaatgattttaaatacaa

	atcaaaaagaaaaaaactaaacattttaacaaaaatgggataagaatagtccttctagaag tgaggatgcgtaaaagaatgagtttccaattaccctgatgtgacaattacacattgtaga caggtagcaaaatatacacatacacccccaaaatgtacaaaatattatataatcaataaat aaatttttaagagtaagtgtattggcattccaaaattcagctaaaaggaaaaatgatca aaaacaaagtaaggtgcacagtttagcaaaagatgcagatgttatatacagcaattotca tgctaaaaatacaacaaaagacaaaagcaaaaaataaacctttgctttttttttttttt tttttttttgagacggagtcctcgctctgtcgcccaggctggagtgagtgaggatct cggtcactgcaagctccgcctccaggttcacgccattctcctgcctcagccaaacctt tgctatttttaattcttgcgttggaactttccagctgttactgaccttgctattttttgtt aaataagattattttacaaacttattcttgaactaaatatagtaaaggaggttttttaaaa taatatttaacatacgaattattaattggccatgttcattatttatctatgtttattaat gggccaatgcaaaaaatcattttttcaaagaaaaatttgccatgtaaagcttaaatat aatattgctgctttgtataactcttctatgtttattctattcatttgcttcttccctac catattttacacatgtatttataatctgtagtattttattacatttctgctttttctagt cattcaatttatcactgctgaattgcatcagatcatggatgcattttttattatgaaaaa taaaatgacttttcaaattaaaaaataaaaaa
#42	caggacaatgaaattcttcagttacattctgggtttatcgccgatttctcttcgtgggttttactgtgttggt ttactacctctgccatcgctctccacaccaagggaagcagaatgtgctacacactctttgtgggtggccacat tttggctcacagaagcattgcctctgtcggttaacagctttgtacctagttaaattgttacctatgtttgggat catgcttcttaagaaggtggcatctgcttatttcaaggattttcacttactgctaattggagttatctgttta gcaacatccatagaaaaatggaatttgcaagaagaattgctctgaaaatgggtgatgatggttggtgtaaatc ctgcatggctgacgctggggttcattgagcagcactgcctttttgtctatgtggctcagcaacacctogacggc tgccatgggtgatgccattgaggaggtgtagtgcagcagatcatcaatgcagaagcagaggtcgaggccact cagatgacttacttcaacggatcaaccaaccaggactagaaattgatgaaagtgttaattggacatgaaataa atgagaggaaagagaaaaacaaaccaggtccaggatataataatgatacagggaatttcaagcaaggtgga gttggaagagactgtttaactactgaaatgaagctattctcctgactaaacataactgaaaaccattcatta aatg
#43	gccactcagatgacttacttcaacggatcaaccaaccacggactagaaattgatgaaagtgttaattggacatg aaataaatgagaggaaagagaaaaacaaaccagttccaggatataataatgatacagggaatttcaagcaa gggtggagtggaaaagcactggaaacttgcaagtggtcccccctctccctctgtccattctgtatcg cagctagctgctcaaggaaaggagaaagtgggaaggcatatgtacttagaaattattctattactttcctggat ttaagagtattcagattttctatttcaacatcaacaattgcatttttaaaaagaaatttatgtgttccatgt caaatttagtagtggtgtgtgtttataatattttcttatatctacttaatttctatagttattatagttata tgtctttatttctaaccattttcttgtgcttttaagattattttaagattatttttaataatctttatttc atttaataaaaatattttattttaagctc
#44	cacggactagaaattgatgaaagtgttaattggacatgaaataaatgagaggaaagagaaaaacaaaccagttc caggatacaataatgatacagggaatttcaagcaaggtggagttggaaaagaactcaggcatgagaaccaa atatcgaaacaaagaaggccacgtgacacgtaaacttacgtgtttgtgcattgcctactcttctaccattggt ggactgacaacaatcactgggtacctccaccaacttgatctttgcagagtatttcaatacattccatccacaca gaagaggagatcgtaacaggcatgtacaccaggaggcagaaatttgaggcatatcttggaaactctgtctacca catcctgaacatcacacagtttccactcttgttgccctcaatcctgagaatgcatecaggagccattctgttt tatgtcaattactaattagatcatgtcacgttactaacttactacgttccaattagtccttattgcatttgta ataaaatccgcatactttcggactggctacaaggttatcatgat
#45	MKFFSYILVYRRFLFVFTVLVLLPLPIVLHTKEAECAYTLFVV ATFWLTEALPLSVTALLPSLMLPMFGIMPSKKVASAYFKDFHLLIGVICLATSIEKW NLHKRIALKMVMVGVNPAWLTIGFMSSTAFLSMWLSNTSTAAMVMPPIAEAVVQIIN AEAEVEATQMTYFNGSTNHGLEIDESVNGHEINERKEKTKPVPGYNNDTGKISSKVEL EKNSGMRTKYRTKKGHVTRKLTCLCIAYSSTIGGLTTITGTSTNLIFAEYFNTRYPC RCLNFGSWFTFSFPAALIILLLSWIWLQWLFLGFNFKEMFKCGKTKTVQQKACAEVIK QEYQKLGPIRYQEIVTLVLFIIIMALLWFSRDPGFVPGWSALFSEYPGFATDSTVALLI GILFFLIPAKTLTKTTPTEIIVAFDYSPLITWKEFQSFMPWDIAILVGGGFALADGCE ESGLSKWIGNKLSPLGSLPAWLIILISSLMVTSLTEVASNPATITLFLPILSPLAEAI HVNPLYILIPSTLCTSFALLPVANPPNAIVFSYGLKVIDMVKAGLGVNIVGVAVVM LGICTWIVPMFDLYTYPSPWAPAMSNETMP"

#46	RTMKFFSYILVYRRFLFVVFTVLVLLPLPIVLHTKEACAYTLEVVATFWLTEALPLSVTALLPSLMDPMFGI MPSKKVASAYFKDFHLLIGVICLATSIEKWNLHKRIALKMVMVGVNPAWLTLGEMSSTAFLSMWLSNTSTA AMVMPIAEAVVQQIINAEAEVEATQMTYFNGSTNHGLEIDESVNGHEINERKEKTKPVPGYNNDTGKISSKVE LEKTV*
#47	ATQMTYFNGSTNHGLEIDESVNGHEINERKEKTKPVPGYNNDTGKISSKVELEKHWKLAVQDGSPSPSVHSVS QLAAQGKEKVEGICT*
#48	HGLEIDESVNGHEINERKEKTKPVPGYNNDTGKISSKVELEKNSGMRTKYRTKKGHVTRKLTCLCIAYSSTIG GLTTITGTSTNLIFAEIFNTFHPHRRGDRTRHVVHQBIEI*
#49	CCAGCTTTAACCATGTCAATG
#50	CAGATGGTTGTGAGGAGTCTG
#51	TGCTAATGCTTTTGGTACAAATGGATGTGGAATATAATTGAATATTTTCTTGTTTAAGGGGAGCATGAAGAGG TGTGAGGTTATGTCAAGCATCTGGCACAGCTGAAGGCAGATGGAATATTTACAAGTACGCAATTTGAGACT AAGATATTGTTATCATTCTCCTATTGAAGACAAGAGCAATAGTAAAACACATCAGGTCAGGGGGTTAAAGACC TGTGATAAACCCTTCCGATAAGTTGGAACGTGTGTCTATATTTTCATATCTGTATATATATAATGGTAAAG AAAGACACCTTCGTAACCCGCATTTTCCAAAGAGAGGAATCACAGGGAGATGTACAGCAATGGGGCCATTTAA GAGTTCTGTGTTTCATCTTGATTCTTCACCTCTAGAAGGGGCCCTGAGTAATTCATCTCATTGAGTGAACAAC AATGGCTATGAAGGCATTGTCTGTGCAATCGACCCCAATGTGCCAGAAGATGAAACACTCATTCAACAAATAA AGGACATGGTGACCCAGGCATCTCTGTATCTGTTTGAAGCTACAGGAAAGCGATTTTATTTCAAAAATGTTGC CATTTTGATTCTCTGAAACATGGAAGACAAAGGCTGACTATGTGAGACCAAAACTTGAGACCTACAAAATGCT GATGTTCTGGTTGCTGAGTCTACTCTCCAGGTAATGATGAACCTTACCTGAGCAGATGGGCAACTGTGGAG AGAAGGGTGAAAGGATCCACCTCACTCTGATTTTATTGAGGAAAAAGTTAGCTGAATATGGACCACAAGG TAAGGCATTTGTCCATGAGTGGGCTCATCTACGATGGGGAGTATTTGACGAGTACAATAATGATGAGAAATTC TACTTATCCAATGGAAGAATACAAGCAGTAAGATGTTTACGAGGTATTACTGGTACAAATGTAGTAAAGAAGT GTCAGGGAGGCAGCTGTTACACCAAAAGATGCACATTCAATAAAGTTACAGGACTCTATGAAAAAGGATGTGA GTTTGTCTCCAATCCCGCCAGACGGAGAAGGCTTCTATAATGTTTGCACACATGTTGATTCTATAGTTGAA TTCTGTACAGAACAAACCACAAACAAAGAGCTCCAAACAAGCAAATCAAAAATGCAATCTCCGAAGCACAT GGGAAGTGATCCGTGATTCTGAGGACTTTAAGAAAACCACTCTATGACACACAGCCACCAATCCCACCTT CTCATTGCTGCAGATTGGACAAAGAATTGTGTGTTAGTCTTGACAAATCTGGAAGCATGGCGACTGGTAAC CGCCTCAATCGACTGAATCAAGCAGGCCAGCTTTCTCTGCTGCAGACAGTTGAGCTGGGGTCTGGGTTGGGA TGCTGACATTTGACAGTCTGCCCATGTACAAAGTGAAGTCTATACAGATAAACAGTGGCAGTGACAGGGACAC ACTCGCCAAAAGATTACCTGCAGCAGCTTCAGGAGGGACGTCCATCTGCAGCGGGCTTCGATCGGCATTTACT GTGATTAGGAAGAAATATCCAAGTATGGATCTGAAATTTGTGCTGCTGACGGATGGGGAAGACAACTATAA GTGGGTGCTTTAACGAGGTCAAACAAAGTGGTGCCATCATCCACACAGTCGCTTTGGGGCCCTCTGCAGCTCA AGAACTAGAGGAGCTGTCCAAATGACAGGAGGTTTACAGACATATGCTTCAGATCAAGTTCAGAACAAATGGC CTCATTGATGCTTTTGGGGCCCTTTCTATCAGGAAATGGAGCTGTCTCTCAGCGCTCCATCCAGCTTGAGAGTA AGGGATTAACCTCCAGAACAGCCAGTGGATGAATGGCACAGTATCGTGGACAGCACCGTGGGAAAGGACAC TTTGTCTTATCAGCTGGACAAACGAGCCTCCCCAAATCCTTCTCTGGGATCCCAGTGGACAGAAGCAAGGT GGCTTTGTAGTGGACAAAACACCAAAATGGCCTACCTCCAAATCCCAGGCATTGCTAAGGTTGGCACTTGGA AATACAGTCTGCAAGCAAGCTCACAAACCTTGACCTGACTGTACAGTCCCCTGCGTCCATGCTACCTGCC TCCAATTACAGTGAATTTCCAAACGAACAAGGACACCAAGCAAAATCCCCAGCCCTCTGGTAGTTTATGCAAT ATTGCGCAAGGAGCCTCCCCAATTTCTCAGGGCCAGTGTACAGCCCTGATTGAATCAGTGAATGGAAAACAG TTACCTTGGAATACTGGATAATGGAGCAGGTGCTGATGCTACTAAGGATGACGGTGTCTACTCAAGGTATTT CACAACTTATGACACGAATGGTAGATACAGTGTAAAAGTGGGGCTCTGGGAGGAGTTAACGCAGCCAGACGG AGAGTGATACCCAGCAGAGTGGAGCACTGTACATACCTGGCTGGATTGAGAATGATGAAATACAATGGAATC CACCAGACCTGAAATTAATAAGGATGATGTTCAACACAAGCAAGTGTGTTTACAGAGAACATCCTCGGGAGG CTCATTGTGGCTTCTGATGTCCCAATGCTCCCATACCTGATCTCTCCACCTGGCCAAATCACCGACCTG AAGCGGAAATTCACGGGGGAGTCTCATTAACTGACTTGGACAGCTCCTGGGGATGATTATGACCATGGAA CAGCTCACAAGTATATCATTGGAATAAGTACAAGTATCTTGATCTCAGAGACAAGTCAATGAATCTCTTCA AGTGAATACTACTGCTCTCATCCCAAGGAAGCCAACTCTGAGGAAGTCTTTTGTTTAAACCAGAAAACATT ACTTTTGAATATGGCACAGATCTTTTCAATTGCTATTGAGGCTGTTGATAAGGTCGATCTGAAATCAGAAATAT CCAACATTGCACGAGTATCTTTGTTTATTTCTCCACAGACTCCGCCAGAGACACCTAGTCTGATGAAACGTC TGCTCCTTGTCTAATATTCATATCAACAGCACCATTCTGGCATTACATTTTAAAAATTATGTGGAAGTGG ATAGGAGAACTGCAGCTGTCAATAGCCTAGGGCTGAATTTTGTGAGATAAATAAATAAATCATTATCCTT TTTTGATTATAAAATTTTCTAAAATGATTTTAGACTTCTGTAGGGGGGATATACTAAATGTATATAGTA

	CATTATACTAAATGTATTCTGTAGGGGGCGATATACTAAATGTATTTTAGACTTCCTGTAGGGGGCGATAA AATAAAATGCTAAACAACCTGGGTAAA
#52	AATTAAATTATGAGAATTAAAAAGACAACATTGAGCAGAGATGAAAAAGGAAGGGAGGAAAAGGTGGAAAAGA AAAGAAGACAAGAAGCGAGTAGTGGTCTCTAACTTGCTCTTTGAAGGATGGTCTCACAAAGAGAACCCCAACA GACATCATCGTGGGAATCAAATCAAGACCAGCAAGTACACCGTGTGTCTTCCGTCCCAAAAACATTTTTGA GCAGCTACACCGGTTTGCCAATCTCTATTTTGTGGGCATTGCGGTTCTGAATTTTATCCCTGTGGTCAATGCT TTCCAGCCTGAGGTGAGCATGATACCAATCTGTGTTATCCTGGCAGTCACTGCCATCAAGGACGCTTGGGAAG ACCTCCGGAGGTACAAATCGGATAAAGTCATCAATAACCGAGAGTGCCTCATCTACAGCAGAAAAGAGCAGAC CTATGTGCAGAAGTGTGGAAGGATGTGCGTGTGGGAGACTTCATCCAAATGAAATGCAATGAGATTGTCCCA GCAGACATACTCCTCCTTTTTTTCCTCTGACCCCAATGGGATATGCCATCTGGAAGTGGCAGCTTGGATGGAG AGACAAACCTCAAGCAAAGACGTGTGCTGAAGGGCTTCTCACAGCAGGAGGTACAGTTGGAACAGAGCTTTT CCACAATACCATCGTGTGTGAGAAACCAACAACCACCTCAACAAATTTAAGGGTTATATGGAGCATCCTGAC CAGACCAGGACTGGCTTTGGCTGTGAGAGTCTTCTGCTTCGAGGCTGCACCATCAGAAACACCGAGATGGCTG TTGGCATTTGTCTATGTCAGGCCATGAGACGAAAGCCATGCTGAACAACAGTGGCCCCGGTACAAACGCAG CAAGATTGAGCGGCGCATGAATATAGACATCTTCTTCTGCATTGGGATCCTCATCCTCATGTGCCCTTATTGGA GCTGTAGGTACAGCATCTGGAATGGGACCTTTGAAGAACACCCCTCCCTTCGATGTGCCAGATGCCAATGGCA GCTTCCCTCCAGTGGCTTGGGGGCTTCTACATGTTCTCACAATGATCATCCTGCTCCAGGTGCTGATCCC CATCTCTTTGTATGTCTCCATTGAGCTGGTGAAGCTCGGGCAAGTGTCTTCTTGAGCAATGACCTTGACCTG TATGATGAAGAGACCGATTTATCCATTCAATGTGAGCCCTCAACATCGCAGAGGACTTGGGCCAGATCCAGT ACATCTTCTCCGATAAGACGGGGACCTTGACAGAGAACAAGATGGTGTTCGACGTTGCACCATCATGGGCAG CGAGTATTCTACCAAGAAAATGGTATAGAAGCTCCCAAGGGCTCCATCCCTCTTTCTAAAAGGAAATACCCCT GCTCTCCTAAGAAACGAGGAGATAAAAGACATTTCTCTGGCTCTCTTAGAGGCTGTGTGGCATTTCACAAAGT TGCTTCCCTGTATCCCTGTGGTCTTCTTGTACAGATCAGGGCTGTTCCAATTACTTGTAACTTTTCAATTTGT TTACAAAGGTTAGAAGTTATCCCATATGTGGTTCCTTTCAGCTGATCTTTGTCTGGTGGCAGACAAAGCACT TTATGAGACGAGTTTTTTATCTGTGAGCAATGGATTGGAGACATTTCCCAATTGTGTGCCAGTCACACAACCA AGGCTTAGGAATTTCTCAGGCCACCTTACCTGACATGTGAGGGCAGGTCTGTGTCTAGGTGCATGGTCAGATT TAATACATCCAGAAGATGTCTTCTATTCTAACAGATCTCTTAGCTTGTCACTGAGGCAAAGTTTGTATTAGG AGATAGGGCTATAAAATGCCTGGACTGTTACCTTGCATGGACTGAATATGACTCATAAACTGATCTGATTCC TTCAGCCATCATCTGCCCAACTTGGTTCCCTTCCCCACCCCCCACAACACACACACACTTTCTAAGAAAA GAAAAGAAATTTCTTTTTTTCAATACTTTAAGTTCTGGGATACATGTGCAGAATGTGCAGGTTTGTACATAG GTATACATGTGTCTATGGTGGTTTGCAGACCCACCAACCCATCATCTACCTTAGGTATTTCTCCTAATGCTAT CCCTCCCTAGCCCCCAACCCCCGATGGGCTCCAGTGTGTGATGTTCCCTCCATGTCCATGTGTCTCAAT GTTCAATTTCCACTTATGAGTGAGAACATGCAGTATTTGGTTTTCTGTTCTTGTGTAGTTTGCTGATGGTTT CCTGTTCTATCCGTGTCCCTGCAAAGGACATGAACATCATCTTTTATGGCTGCATAATATCCATGGTGTAT ATGTGCCACATTTCTTTATCCAGTCTATCGCTGATGGGCACTGGGGTTGGTTCCAAGTCTTGTCTATTGTGA ACAGTGTCTGCAATAAACTTACATGTGCATGTGTCTTAGTAGAATGATTTATAATCCTTTGGGTATATACCCA GTAATGGGATTGCTGGTCAAATGGTATTTCTGGTTCTAGATCCTTGAGGAATCTTGTCTTCCACAATGGTTG AACIAATTTGTACTCCACCAACAGTGTAAAGTATTCCTGTTTCTCTACATCCTCTTCAGCATCTGTTGTGT CCTGACATTTTAAATGATCACTATTCTCACTGGCGTGAGATGTTATCTCATTGTGGTTTTGATTGCAATTTCTC TAATGACCAGTAATGATGAGCTTTTTTTCATATGTTTGTGGCTGCATAAATGTCTTCTTTTGAGAAGTGTCT GTTCAATCCTTCACCCATTTTTTGAAGAAAACAACTCTTAAGAGAGCAGTATTCATTCTTTTGAGTGTGAG GGATGGAGAAAGAGAAAGATGGAGAGAGTATTATAAGCAGCTGTATCCCTTTGCCATGGTGATAGCAGACCA TTCACATGGGAGCTTCTGGTCTCTTTGTAATAATAAAGAGCCACATTACCAGTACTTAGAGTATGCTAGTT ATTTTAACACATTGTATCATTAATCTTCAAAACATCCCTATGAGTTAGAAACCTAAAAAATAAAAAA A
#53	CTCATTTTGATGTCTAGAATCAGGGGATCCAGGATCATACCAAGGTCAATTTCCCAGGTATGGAGGGGTCTT TCTGCTTCTTTCTTGTGATGCACAGCTGCTGAGGAAGGGGCTGGGAGTAAAGACAGTGAATGGGGAGGAGGA GTCCATTCAAACCGAGAAACAAAGTGTGTTGGTTTTTCTTACCCCTGGTGTAGAAGCTACCAACCTTTTCCAAG AAAGAGGGCCTGGCCCCCTTCTCGGGTCTGGCTGGGTGCTGCTGTGCTCTCTGGCCTCCCTCCGAAGGGC ACCATTCCTCGGGTGAGTACTACCGGCTGCACCGTCTTCCAGTGGGGACAGCCTGAGAAGAGAGTCTGGGG CCTTACTTCAGTACCTTCCCTTCACTGGCCTCACCTGTGCAATCATGCCACACGCTGCAGCCTCCTTTTCCC TATCTATAAAATAAAATGACCCTGCTCTATCTCACTGGGCTGGCAAGAACAACACTGTTGTTGCCCTTGCAGAC AGATGTGCTGAGGCTGTAGAAAGTGTCTTTTATTTGGTGGGAGCTTGTGCATAAATGCCAGAGGGGCTGCAC ATCTGACGAGTAGAGGTGACTCATGGCTGAACCGGAACAGGACATCGGGGAGAAGCCAGCAGCCATGCTGAA CTCTCCACAGGGCCCTGTGAAAAGCTCTTACCTCCTCTGCCCTCTGGATCTAGTGAAGCCTATTCATCCTTC AGATGTGAGCTCAAATAATCAACCTTCATGGAGGCTCCCTTGACCCCTAACATGCTTTCAAAGTACTGTGTA TTTCACATTCATCATGCCCGACAACGTGATTTCCCATTTATTAATATCTGTCTCTTCTGCTGGCCTGCAAA CTCCAGGAGCACAGAGACATCTTTGGGATTTTTGAACATGATTTCCCAGGGCTTAGCCAGTGCCTGGTGCA AAGCAGGCTTTCAACATGTTCAAGTGGATATTGTAAGAAAGAAAGAAATACACAAAGGCTTGGCATATGCAAA GCACTCTAAATATTCACTCCTTTCCCTTCCCTCTGGGTGAGAAAATTTCTCTTATAAAGACACCCTCCTAAC TGTATCTCTGTAGAGAACTGAAGACATAAAGCACTCTGTGCCAAAAATTTAAGTAAAACTTGAGCTAAG CACAGAGATTATAAATATTCTTCCCCAGATTACGCACCATTTAAAAATACTGTCTCAGCTCCTTTTCATGAT

	TTGGGTGGTGATTAAAGAAAATTACTCTTCAAGACTGAAAGTCATTACTGCCCTTTTCTGACTTGCCTTTTC CCTTGAGAAGGGGAGGATAAGCTGCAGGGCAGGAAGTGGAAGTGGGGCATCCTTGTCTTTGTCTGGCAGACA GCCAAGTGGTCAGGTACTGCTCCTTCTCAACTCTTTCTGATTCCCAGGTGAATATAAACAAGAAGGCACAAA TCCCACTTGCCAACAACGGACCCCAAGTGATAACAAGAAACCCAGTGACACCTGTCTAGGTGAAGACTCAGCC CCTATGTGACCAGGTGCAAAGCCAACTGACCATCTGCTTTCCATTTGGACTTTTAGTTTACTACTGTATCTT CTCAGGACAGTTAAGTTGGAATACAATGCCACTGTCCTGAAAGATGGTAGAATTATCCTATTTCTGGAGGAGT GGGGGTGGTGGGTAGGAATCTCAAGAGCGATTGCTCCTCTGCACAATAGCTTCTTTAAGGACACCAGGGCCC CCAGGGCTATACATTTCCCTGAAGCTTTCCAGATAAGCAACAAGGTATGAGCACCTGCTATGTATTGCCCAAG GGTGATGTGTTTAAATATCCATTGCATATTTTAAATCCTTGGCTGGCTTAAAGCTGCAAGCTTTCTGTCTTCA GTGGATATAATGGGGGCATACATCCCAGAGCTTGCCCAACACTCCAAGAAAAGAACCCTCAGCTAATGCAAG TGTGTATGTGCCCATGAAAGCTCCATGTCTACTTAACATTCACTTTTAGGATTATTTATGCTGTATAAATAG ATATGAAAATCTCTGACAGGTATTTTGTTCCTTTACAACTGTATTTGAATTTATGGGTGATTTAGAGCTTG TGTTTAAAGTCAGAATTCAGAACCCCAAAGAAATGACTTCATTGAATTTGAAGTGAAGAGACAAGAAGTGA TTACCAAAACCTACTAAACGTGAGTTGCTGTGAAGTGGGGATTAAACCAGAACGAGTGGAGAAGATCAGAAAG CTACCAAAACACTGCTCAGAAAGGACAAAGACATTCGAAGACTGCGGGACTTTCAGGAAGTGAAGTCAATTT TAATGAAAATGGAAGCTCCAGATTGACAGAATATGTGCCATCTCTGACAGAAAGGCCCTGCTATGATAGCAA AGCTGCAAAAATGACTTATTAAATACTCCCAGGAATGGCCGCGCATGGTGGCTCACCCCTGTAATCCCAGCA CTTTGGGAAGCCAAGGTGGGCGGATCACCTGAGGTGAGGAGTCTAGACCAGCTGGCCAACATATAGTGAAA CCCAGTCTCTACTAAAAAAATAACAAAATTAGCTAGGTGTGGTGGCGCACACCTGTAGTAGTCCCAGCTACA TGGGAAGCTGAGGCAGGAGAATCACCTGAACCCAGGAGGCAGAGGTTGCAGTGAGCTGAGATTGCGCCACTGC ACTCCAGCCTGGCGACAGAGCAAGACTCTGTCTCTCAAAATAAATAAATAAATAAATAAATAAATAAATAAAT AATC
#54	GCCCGGAGAGAGAGAGAGAGCGGGCCGAGGACTCCAGCGTGCCAGGTCTGGCATCCTGCACTTGCTGCCCTCT GACACCTGGGAAGATGGCCGGCCCGTGGACCTTCACCTTCTCTGTGGTTTGCTGGCAGCCACCTTGATCCAA GCCACCTCAGTCCCAGTCTCATCTCTCGGCCAAAAGTCATCAAAGAAAAGCTGACACAGGAGCTGA AGGACCACAACGCCACCAGCATCCTGCAGCAGCTGCCGCTGCTCAGTGCCATGCGGGAAAAGCCAGCCGGAGG CATCCCTGTGCTGGGCGCCTGGTGAACACCGTCTGAAGCACATCATCTGCTGAAGGTGATCACAGCTAAC ATCCTCCAGCTGCAGGTGAAGCCCTCGGCCAATGACCAGGAGCTGCTAGTCAAGATCCCCCTGGACATGGTGG CTGGATTCAACACGCCCTGGTCAAGACCATCGTGGAGTTCCACATGACGACTGAGGCCCAAGCCACCATCCG CATGGACACCAGTGCAAGTGGCCCCACCGCCCTGGTCTCAGTGACTGTGCCACCAGCCATGGGAGCCTGCGC ATCCAAGTCTGCATAAGCTCTCCTTCTGGTGAACGCCCTAGCTAAGCAGGTGATGAACCTCCTAGTGCCAT CCCTGCCCAATCTAGTGAAAACAGCTGTGTCCCGTGATCGAGGCTTCTTCAATGGCATGTATGCAGACCT CCTGCAGCTGGTGAAGGTGCCCATTTCCCTCAGCATTGACCGTCTGGAGTTGACCTTCTGTATCCTGCCATC AAGGGTGACACCATTGAGCTCTACCTGGGGGCCAAGTTGTGGACTCACAGGGAAAGGTGACCAAGTGGTTCA ATAACTCTGCAGCTTCCCTGACAATGCCACCCCTGGACAACATCCCGTTGAGCCTCATCGTGAGTCAGGACGT GGTGAAAGCTGCAGTGGCTGCTGTGCTCTCTCCAGAAGAATTCATGGTCTGTGGACTCTGTGCTTCTGAG AGTGCCCATCGGCTGAAGTCAAGCATCGGGCTGATCAATGAAAAGGCTGCAGATAAGCTGGGATCTACCCAGA TCGTGAAGATCCTAACTCAGGACACTCCCGAGTTTTTTATAGACCAAGGCCATGCCAAGGTGGCCCAACTGAT CGTGCTGGAAGTGTTCCTCCAGTGAAGCCCTCGGCCCTTTGTTACCCCTGGGCATCGAAGCCAGCTCGGAA GCTCAGTTTTACACCAAGGTGACCAACTTATACTCAACTTGAATAACATCAGCTCTGATCGGATCCAGCTGA TGAAGTCTGGGATTGGCTGGTTCCAACCTGATGTTCTGAAAACATCATCACTGAGATCATCCACTCCATCCT GCTGCCGAACCAGAATGGCAAATTAAGATCTGGGGTCCAGTGTCTATTGGTGAAGGCCTTGGGATTCCAGGCA GCTGAGTCTCTCACTGACCAAGGATGCCCTTGTGCTTACTCCAGCCTCCTTGTGGAAACCCAGCTCTCCTGTCT CCCAGTGAAGACTTGGATGGCAGCCATCAGGGAAGGCTGGGTCCAGCTGGGAGTATGGGTGTGAGCTCTATA GACCATCCCTCTCTGCAATCAATAAACACTTGCCTGTGAT
#55	GGAGTGGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG GTTTCAAACTGGGGGCTCATTGTCTTCTACGGGCTGTAGCCAGACCATGGCCAGTTTGGAGGCTGCC GTGCCCTGGACAGACCTGCCCTTGAATGTGAATCCAGCCCTGCCCTTGAGTCCCACAGGTCTTGAGGAA GCTTGACAAATGCCCTCAGCAATGGCCTGCTGTCTGGGGGCTGTGGGCAATCTGGAAAACCTTCCGCTCCT GGACATCCTGAAGCCTGGAGGAGGTACTTCTGGTGGCTCCTTGGGGGACTGCTTGGAAAAGTGACGTGAGT ATTCTTGGCCTGAACAACATCATTGACATAAAGGTCACTGACCCCGAGCTGCTGGAAGTTGGCCTTGTGCAG GCCCTGATGGCCACCGTCTCTATGTACCATCCCTCTCGGCATAAAGCTCCAAGTGAATACGCCCTGGTGGG TGCAAGTCTGTGAGGCTGGCTGTGAAGCTGGACATCACTGCAGAAATCTTAGCTGTGAGAGATAAGCAGGAG AGGATCCACCTGGTCTTGGTGAAGTGCACCATTCCTTGGAGCCTGCAATTTCTCTGCTTGTGAGTGGTCA GCCCCCTCCCATTCAAGGTCTTCTGGACAGCCTCACAGGGATCTTGAATAAAGTCCCTGCTGAGTTGGTTCA GGGCAAGGTGTGCCCTCTGGTCAATGAGGTTCTCAGAGGCTTGGACATCACCTGGTGCATGACATTGTTAAC ATGCTGATCCACGGACTACAGTTTGTCAATCAAGGTCTAAGCCTTCCAGGAAGGGGCTGGCCTCTGCTGAGCTG CTTCCAGTCTCACAGATGGCTGGCCCATGTGCTGGAAGATGACACAGTTGCCCTTCTCTCGAGGAACCTGC CCCCCTCTCTTTCCCACAGGCGTGTGTAACATCCCATGTGCCTCACCTAATAAATGGCTCTTCTCTGCAA AAAAAAAAAAAAAAAAAAAAAAAAAAAA

GAGCAGAGCCCTTTACACACCTCAGGAACACCTTTGGGCTGCCCGCTCCCCAGACACACCTGCAGCCCTGCC
CAGCCGGCTTTGCTCACCCTACTGCTTGTAAATGCCCCAGATATGAGCCAGCCAGGCCCGCTACGTGGTAGA
CAGAGCCGCATACCTCCTTACCCTCTTCGACGATGAGTTTGAGAAGAAGGACCGGACATACCCAGTGGGAGAG
AACTTCGCAATGCCCTCAGATGTTCTCAGCCAAGATCAAAGCTGTGGTGTGTTGGGCTGCTGCCTGTGCTCT
CCTGGCTCCCCAAGTACAAGATTAAAGACTACATCATTCCTGACCTGCTCGSTGGACTCAGCGGGGGATCCAT
CCAGGTCCCACAAGGCATGGCATTGCTCTGCTGGCCAACTTCTCAGTCAATGGCCTCTACTCCTCCTTC
TTCCCCCTCCTGACCTACTTCTTCTGGGGGGTGTTCACCAGATGGTGGCAGGTACCTTTGCCGTTATCAGCA
TCCTGGTGGGTAACATCTGTCTGCAGCTGGCCCCAGAGTCGAAATTCAGGTCTTCAACAATGCCACCAATGA
GAGCTATGTGGACACAGCAGCCATGGAGGCTGAGAGGCTGCACGTGTGAGCTACGCTAGCCTGCCTCACC GCC
ATCATCCAGATGGGTCTGGGCTTCATGCAGTTTGGCTTTGTGGCCATCTACCTCTCGAGTCTTCATCCGGG
GCTTCATGACGGCCGCCGGCCTGCAGATCCTGATTTGGGTGCTCAAGTACATCTTCGGAAGTACCATCCCCTC
CTACACAGGCCCCAGGGTCCATCGTCTTTACCTTCATTGACATTTGCAAAAACCTCCCCACACCAACATCGCC
TCGCTCATCTTCGCTCTCATCAGCGGTGCCTTCCTGGTGTGGTGAAGGAGCTCAATGCTCGCTACATGCACA
AGATTCGCTTCCCCATCCCTACAGAGATGATTGTGGTGGTGGTGGCAACAGCTATCTCCGGGGGCTGTAAAT
GCCCAAAAGTATCACATGCAGATCGTGGGAGAAATCCAACGCGGGTTCCCCACCCCGGTGTGCGCTGTGGTC
TCACAGTGAAGGACATGATAGGCACAGCCTTCTCCCTAGCCATCGTGAGCTACGTATCAACCTGGCTATGG
GCCGGACCTCGCCAACAAGCAGGCTACGACGTGGATTGCAACAGGAGATGATCGCTCTCGGCTGCAGCAA
CTTCTTTGGCTCCTTCTTTAAATTCATGTCTTTGCTGTGCGCTTTCTGTCACTCTGGCTGTGGATGGAGAT
GGAGGAAAATCCCAGGTGGCCAGCCTGTGTGTGTCTCTGGTGGTGTGATGATCACCATGCTGGTCTCGGGGATCT
ATCTGTATCCTCTCCCTAAGTCTGTGCTAGGAGCCCTGATCGCTGTCAATCTCAAGAACTCCCTCAAGCAACT
CACCGACCCCTACTACCTGTGGAGGAAGAGCAAGCTGGACTGTTGCATCTGGGTAGTGAGCTTCCTCTCCTCC
TTCTTCTCAGCCTGCCCTATGGTGTGGCACTGGGTGTGCGCTTCTCGCTCCTGGTGTGGTCTTCCAGACTC
AGTTTCGAAATGGCTATGCACTGGCCAGGTCTGGACACTGACATTTATGTGAATCCCAAGACCTATAATAG
GGCCAGGATATCCAGGGGATTAATCATCAGTACTGCTCCCTCTCTACTTTGCCAACTCAGAGATCTTC
AGGCAAAAGGTCTCGCCAAGACAGGCATGGACCCCAAGAAATATTACTAGCCAAGCAAAATACCTCAAGA
AGCAGGAGAAGCGGAGAATGAGGCCACACAACAGAGGAGTCTCTATTATGAATAACCAAGACTGTCTCCCT
GCAGGAGCTGCAGCAGGACTTTGAGAATGCGCCCCCAGCGACCCCAACAACAACAGACCCCGGCTAACGGC
ACCAGCGTGTCTATATACCTTCAGCCCTGACAGCTCTCACCTGCCAGAGTGAGCCACCAGCCTCCGCTG
AGGCCCCCGCGAGCCAGTGCATGCTGGCCAGCGTCCCACCTTCGTACCTTCCACACCTCATCTGGA
CATGAGTGGAGTCACTTCGTGGACTTGATGGGCATCAAGGCCCTGGCCAAGCTGAGCTCCACCTATGGGAAG
ATCGGCGTGAAGGTCTTCTTGGTGAACATCCATGCCAGGTGTACAATGACATTAGCCATGGAGGCGTCTTTG
AGGATGGGAGTCTAGAATGCAAGCAGTCTTTCCAGCATAACATGACGAGTCTCTTTGCCAGGCAATGC
TAGAGACGTGACCCAGGACACAACCTTCCAAGGGGCTCCAGGGGATGCTGAGCTCTCCTTGTACGACTCAGAG
GAGGACATTCGCACTACTGGGACTTAGAGCAGGAGATGTTGGGGAGCATGTTTACGCAGAGACCTGACCG
CCCTGTGAGGGCTCAGCCAGTCTCATGCTGCCCTACAGAGTGCCTGGCACTTGGGACTTCCATAAGGATGAG
CCTGGGGTCACAGGGGGTGTGGGGCGGAGGAAAGTGCATCCCCAGAGCTTGGGTTCCTCTCTCCTCTCCCC
TCTCTCCTCCTTCTTCTCCTCCCCGCATCTCCAGAGAGAGCCTCTCAGCAGCAGGGGGGTGCTACCCTTACG
GGAGTGAGAGTCTGGTGAGCCCACTCTTACCCGTGAGCCCTGGCCGCAATGGACAAGCCTCTGCTCACTC
CACCCACCCACATCTGCCCTGTCTTGGCAGCTGAAGGACACCTTGACTTCCAGCTTTTACGAGTGAGCCAA
AAACAGAAGGACAAGTACAACCTGTGCTGGCCTGCTGTACAAGCTTCAAAAAGTGTCCAGAGCCCGCAGCGCT
CGGTGTGAGATGGTGTGAGGCTGTACGGACATAGGGATAAAGTTGGTTAGGACTCTGGCTTGCTTCCCCAG
CTGCCCTCACTCTGTCTCTGGCAGCTCTGCACCCAGGGACCATGTGCTCTCCACACCCAGGAGTCTAGGCCTT
GGTAACTATGCGCCCCCTCCATCATCCCCAAGGTGCCCAAACCACTGCTGTGAGCAAGCACATCAGA
CTCTAGCCTGGACAGTGGCCAGGACCGTCGAGACCACAGAGCTACCTCCCCGGGGACAGCCCACTAAGGTTT
TGCTCAGCCTCCTGAAACATCACTGCCCTCAGAGGCTGCTCCCTTCCCTGGAGGCTGGCTAGAAACCCAA
AGAGGGGATGGGTAGCTGGCAGAATCATCTGGCATCCTAGTAATAGATACCAATATTCTGCACAAAACCTT
TGGAATTCCTCTTTGCCACCCAGAGACTCAGAGGGGAAGAGGGTGTAGTACCAACACAGGGGAAACGGATGG
GACCTGGGGCCAGACAGTCCCCCTTGACCCCAAGGGCCATCAGGGGAATGCCTCCCTTGGTAAATCTGCCTT
ATCCTTCTTTACCTGGCAAGAGGCCAATCATGTTAACTCTTCTTATCAGCCTGTGGCCAGAGACACAATGG
GGTCTTCTGTAGGCAAGGTGGAAGTCTCCAGGGATCCGCTACATCCCTAACTGCATGCAGATGTGGAAA
GGGGCTGATCCAGATTGGGTCTTCTGSCACAGGAAGACTCTTTAACACCTTAGGACCTCAGGCCATCTTCTC
CTATGAAGATGAAAATAGGGGTAAAGTTTCCATATGTACAAGGAGGTATTGAGAGGAACCTACTGTTGACT
TGAAAATAAATAGGTTCCATGTGTAAGTGTTTTGTAAAATTTCACTGGAAATGCACAGAAAATCTTCTGGCT
CTCATCACTGCTTTTCTCAAGCTTCTTCAAGCTTAAACAACCCCTTCCCTAACAGGTGGGGTGGCCAGCCTAG
GAAACATCCCCATTTCTAACTTCAGCCAGACCTGCGTTGTGTGTCTGTGTGTGAGTGAGCTGGTCAGCTAA
CAAGTCTTCTTAGAGTTAAAGGAGGGGGTGTGGCCAAGAGCCAACACATCTTGGCCCAAGGAGCATGCTTT
TCTGTGAATTCATTATGCCATCTGGCTGCCAATGGAAGTCAAACTTGAAGGCGAAGGACAATGTTATCTGG
GATTCACCGTGGCCAGCACCAGAGTGCCAAATTCAGGAGGACAAGAGCCTTAGCCAATGACAACTCACTCT
CCCCACTCCACCTCCTTCAAGTCCAGCTCAGGCCAGGAGGTGGGAGAAGGTACAGAGCCTCAGGAATTT
CCAAGTCAGAGTCCCCCTTGAACCAAGTATCTAGATCCCTGAGGACTTGATGAAGTATCTTAACCCCCAA
GTAATCATTAACCCCAAGACAGCCTCAGAACTGAAGGAGATTGTTGACCCAGTGACCTGGAGTTGAGGCTCA

	GGCCCCAAGCAGACTCACTAATCCCAAACAACTCAGCTGCCATCTGGCCTCTCTGAGGACTCTGGGTACCTTAAAGACTATA
#59	CAGGAAAGTTCTGTGCTAGGCAGAGGAACTGCAGCTTGTTEGCAGGTGAAGGGAGCCTGTTTAGCTGTGTC CAGCAACAACCTTACGTGGTCCCTGCTTGTGTTCCAGGTGAAGCCTCTGGCCGCCGAGCAGAGGAATCAAGACCT GCTCATTCTTTCTCGGGGGATCCATCCAGCAATGACATCATCTCATGCTGCCACAAGGACCCCAAGTCTGGG CTGCTEGGGACCAGCCACGCTCCCCACTGCTCATTCTTCATCCTAGAGACATTCTGACTCTCCTCCGACTGC GCTGTGCACAGCGGTGACAAGCTCTTTTACATCTCAGTCTGCACAACCTCAGGCACCTAGCAGATTGATATGC ATCCAACAAATATTGATTGAATATCTGCTAAATACCCAGTAATGTTTCATGAGTGATTGGGTGAATAAAGGAA TGCTSGTTCCTTCTGGCCATATTAACCTCTGCACAATACTAAGAAAAATAAATTGCAGTACTGTGGAATAAT GTGAATCCCAATGTCATCTATTGAAATATTACCTGACTATTAAGAGGTATTTATTTTGTATCTTTTCTAGCA AAGTAAATAAAATCTTAATACAGCATATCCCCTTATTCACGGGGGGTATGTTCCAAGACCCCCGGTGGATGC CTGAAACTATGGATAATACCAGATCC
#60	MGPFXSSVFILILHLLEGALSNSLIQLNNNGYEGIVVAIDPNVPEDETLIQQIKDMVTQASLYLFEATGKRIFY FKNVAILIPETWTKADYVRPKLETYKNADVLVAESTPPGNDEPYTEQMGNCGEKGERIHLTPDFIAGKKLAE YGPQKAFVHEWAHLRWGVFDEYNNDKFXLSNGRIQAVRCSAGITGTNVVKKCQGGSCYTKRCTFNKVTGLY EKGCEFLVLSRQTEKASIMEAQHVDSIVEFCTEQNHKEAPNKQKCNLRSTWEVIRDSDFKKTTPMTTQP PNPTFSLQIGQRIVCLVLDKSGSMATGNRLNRLNQGQLFLLQTVELGSWVGMVTFDSAHHVQSELIQINS SDROTLAKRLPAAASGGTSICSGLSAFTVIRKKYPTDGSEIVLLTDGEDNTISGCFNEVKQSGAIHTVALG PSAAQELEELSKMTGGLQTYASDQVQNNGLIDAFGALSSGNGAVSQRSIQLESKGLTLQNSQWMNGTVIVDST VGKDTLFLITWTTQPPQILLWDPGQKQGGFVVDKNTKMAYLQIPGIKVGWTKYSLQASSQTLTLTVTSRAS NATLPPTVTSTKNKDTSKFSPPLVYANIRQGASPILRASVTALIESVNGKTVTLELLDNGAGADATKDDGV YSRYFTTYDTNGRYSVKVRALGGVN AARRRVIPOQSGALYIPGWIENDEIQWNPPEINKDDVQHKQVCFSTRSSGGSFVASDVNPAPIPDLFPFGQ ITDLKAEIHGGSILNLTWTAPGDDYDHGTAKHYIIRISTSIDLRDKFNESLQVNTTALIPKEANSEEVFLFK PENITFENGTDLFIAIQAVDKVDLKEISNIARVSLFIPPQTPPETPSPDETSAPCPNIHINSTIPGIHILKI MWKWIGELQLSIA
#61	MKKEGRKRWRKEDDKRVRVVSNNLLFEGWSHKENPNRHRGNQIKTSKYTVLSFVPKNIFEQLHRFANLYFVGI AVLNFIPIVNAEQPEVSMIPICVILAVTAIKDAWEDLRRYKSDKVINNRECLIYSRKEQTYVQKWKDVRVED FIQMKCNEIVPADILLFSSDPNGTCHLETASLDGETNLKQRRVVKGFSSQVEVQFEPELFHNTIVCEKPNHL NKFKGMEHPDQTRTGFGCESLLLRGCTIRNTEMAVGIVYAGHETKAMLNNSGPRYKRSKIERRMNIIDIFFC IGILILMCLIGAVGHSIWNNGTFEEHPFDVDPANGSFLPSALGGFYMTMTILLQVLIPISLYVSIELVKLG QVFFLSNDLDLYDEETDLISIQCRALNIAEDLQIQYIFSDKTGTLTENKMFRRCTIMGSEYSHQENGIEAPK GSIPLSKRKYPALRNEFIKDILLALLEAVWHFHKLLPVSLWSSLSQIRAVPITCKLSFVYKG
#62	MGRRSFPKPRNKVFGFSYPWCRSYQPFPRKRAWPPSRVWLGAACASLASPPKGTIPSGEYYPAPSSSGDSL RESGALLQYLPPLASPCANHATRCSLFFIYKIKMTLLYLTGLARTHCCLADRCAEAVESAFYLVGSLCINA RGAHLTD
#63	MAGPWTFTLLCGLLAATLIQATLSPTAVLILGPKVIEKLTQELKDNATSILOQLPLLSAMREKPAAGGIPVL GSLVNTVLKHIIWLKVITANILQLQVKPSANDQELLVKIPLDMVAGENTPLVKTIVEFHMTEAQATIRMDTS ASGPTRLVLSDCATSHGSLRIQLLHKLSFLVNALAKQVMNLLVPSLPNLVKNQLCPVIEASFNGMYADLLQLV KVPISLSIDRLDFDLYPAIKGDTIQLYLGAKLLDSQGVTKWFNNSAASLTMTLDNIPFSLIVSQDVVKAA VAVLSPEEFMVLLDSVLPESAHLKSSIGLINEKAADKLGSTQIVKILTQDTPEFFIDQGHAKVAQLIVLEV FSSSEALRPLFTLGIEASSEAQFYTKGQQLILNLNNISSDRIQLMNSGIGWFQPDVLKNIITEIHSILLPNQ NGKLRSGVPVSLVKALGFEEAESSLTKDALVLTASLWKPSSPVSQ
#64	MFQTGGLIVFYGLLAQTMAQFGGLPVPLDQTLPLNVNPAIPLSPTGLAGSLTNALSNGLLSGGLLGILENPL LDILKPGGGTSGGLLGGLLGRVTSVIPGLNNIIDIKVTDQQLLELGLVQSPDGRRLYVTIELGKILQVNTPLV GASLLRLAVKLDITAEILAVRDKQERIHVLGDCTHSPGSLQISLLDGLGPIPIQGLDLSLTGILNKVLPPLV QGNVCEPLVNEVLRGLDITLVHDIVNMLIHGLQFVIKV
#65	MSQPRERYVDRAAYSLTDFDEFEKKDRTPVGEKLRNAFRCSAKIKAVVEGLLEVLVSWLPKYKIKDYIIP DLGGLSGGSIQVPOGMAFALLANLPAVNGLYSSFFPLTYFFLGGVHQMVPGTFAVISILVGNICLQAPES KFQVFNATNESYVDTAAMEAERLHVSATLACLTAIQMGGLGFMQFGEVAIYLSSESFIRGEMTAAGLQILISV

	LKXIFGLTIPSYTGPGSIVFTFIDICKNLPHNTIASLIFALISGAFLVLVKELNARYMHKIRFPIPTFMIVVV VATAISGGCKMPKHYMQIVGEIQRGFTTPVSPVVSQWKDMIGTAFSLAIVSYVINLAMGRTLANKHGYDVDS NQEMIALGCSNFTGSEFKIHVICCALSVTLAVDGAGGKSQVASLCVSLVVMITMLVLGIYLYPLPKSVLGALI AVNLKNSLKQLTDPYYLWRKSKLDCCIWVVSFLSSFFLSLPYGVAVGVAFSVLVVVFQTFRNGYALAQVMDT DIYVNPNTYNRAQDIQGIKIITYCSPLYFANSEIFRQKVIKTGMDPQKVLAKQKYLKKQEKRRMRPTQQR SLEMKTKTVSLQELQDDFENAPPTDPNNNQTTPANGTSVSYITFSPDSSSPAQSEPPASAEAPGEPDMLASVP PFVTFHTLILMSGVSFVDLMGIKA LAKLSSTYGKIGVKVFLVNIHAQVYNDISHGGVFEDGSLECKHVFPSIHDAVLEAQANARDVTPGHNFGAPG DAELSLYDSEEDIRSYWDLQEMFGSMFHAETLTAL
#66	MEQSGRLEDFPVNVFSVTPYTPSTADIQVSDDDKAGATLLESGIFLGLVGITFTVMGWIKYQGVSHFEWTQL LGPVLLSVGVTFILIAVCKFKMLSCQLCKESEERVDPDSEQTPGGPSFVFTGINQPIFTHGATVVQYIPPPYGS PEPMGINTSYLQSVVSPCGLITSGGAAAAMSSPPQYYTIYPQDNSAFVVDEGCLSFTDGGNHRPNPDVDQLEE TQLEEEACACFSPPPYEEIYSLPR
#67	ACACGAATGGTAGATACAGTG
#68	ATACTTGTGAGCTGTTCCATG
#69	ACTGTTACCTTGCATGGACTG
#70	CAATGAGAACACATGGACATG
#71	CCATGAAAGCTCCATGTCTAC
#72	AGAGATGGCACATATTCTGTC
#73	ATCGGCTGAAGTCAAGCATCG
#74	TGGTCAGTGAGGACTCAGCTG
#75	TTTCTCTGCTTGATGCACTTG
#76	GTGAGCACTGGGAAGCAGCTC
#77	GGCAAATGCTAGAGACGTGAC
#78	AGGTGTCCTTCAGCTGCCAAG
#79	GTAAAGTGCTCTCTGGATTG
#80	ATCCTGATTGCTGTGTGCAAG
#81	CTCTTCTAGCTGGTCAACATC
#82	CCAGCAACAACCTTACGTGGTC
#83	CCTTTATTACCCCAATCACTC
#84	agaacagcgcagtttgcctccgctcacgcagagcctctccgtggcctccgcaccttgag cattaggccagttctctctctctctctaatccatccgtcacctctcctgtcatccgttcc catgocgtgaggtccattcacagaacacatccatggctctcatgctcagtttggttctga gtctcctcaagctgggatcagggcagtggtggttgggcccagacaagcctgtccagg ccttggtggggaggacgcagcattctcctgttctcctgtctcctaagaccaatgcagagg ccatggaagtgcggttcttcagggccagttctctagcgtggtccacctctacagggacg

	ggaaggaccagccatttatgcagatgccacagtatcaaggcaggacaaaactggtgaagg attctattgoggagggggcgcacatctctctgaggctggaaaacattactgtgttgatgctg gcctctatgggtgcaggattagttccagctcttactaccagaaggccatctgggagctac aggtgtcagcactgggctcagttcctctcatttccatcaagggatatgttgatagagaca tccagctactctgtcagtcctcgggctgggtcccccgccacagcgaagtggaaaggctc cacaaggacaggatttgtccacagactccaggacaaacagagacatgcatggcctgtttg atgtggagatctctctgaccgtccaagagaacgcccgggagcatactctgttccatgcggc atgctcatctgagccgagaggtggaatccagggtacagataggagatacctttttcgagc ctatatcgtggcacctggctaccaaagtaactgggaatactctgctgtggcctatTTTTTg gcattgttggactgaagattttcttctccaaattccagtgtgagcagagagagagaagcat gggcgggtgccttattcatggttccagcagggaacaggatcagagatgctcccatccag ctgcttctctctcttagtcctagcctccaggggccaggcccaaaaaaggaaaatccag gcggaactggactggagaagaaagcacggacaggcagaattgagagacggccggaaacac gcagtggaggtgactctggatccagagacgggtcaccggaagctctgcgtttctgatctg aaaactgtaacccatagaaaagctcccaggaggtgcctcactctgagaagagatttaca aggaagagtgtggtggcttctcagagtttccaagcagggaacattactgggaggtggac ggaggacacaataaaaaggtggcgctgggagtggtccgggatgatgtggacaggaggaag gagtacgtgactttgtctcccgatcatgggtactgggtcctcagactgaatggagaacat ttgtatttcacattaaatccccgttttatcagcgtcttcccaggacccccctacaaaa ataggggtcttccctggactatgagtgtgggacctctcttcttcaacataaatgaccag tcccttatttataacctgacatgtcggtttgaaggcttattgaggccctacattgagtat ccgtcctataatgagcaaatggaactcccatagtcctctgccagtcacccagggaatca gagaaaggccctcttggcaaggccctctgcaatcccagagacaagcaacagtgaagtc tctcacaggcaaccacgccccttctcccccagggtgaaatgtaggatgaatcacatccc acattcttctttagggatattaaggtctctctcccagatccaaagtcgccagcagccgg ccaaggtggcttccagatgaagggggactggcctgtccacatgggagtcaggtgtcatgg ctgccttgagctgggaggaagaaggctgacattacatttagtttgctctcactccatct ggctaagtgatcttgaataccacctctcaggtgaagaaccgtcaggaattcccatctca caggctgtggtgtagattaagtagacaaggaatgtgaataatgcttagatcttattgatg acagagtgtatcctaattggtttgttcattatattacactttcagtaaaaaaaaaaaaaa aaaa.
#85	malmlslvlsllklsggqwqvfgpdkpvqalvgedaafscflspktnaeamevrffrgqf ssvvhlyrdgkdqpfmqmpqyqgrtklvkdsiaegrislrlnitvldaglygcrissqs yyqkaiwelqvsaalgsvglisitgyvdrdiqlcqsagwfprrptakwkpgggqdlstdsr tnrdmhglfdveislvtqenagsiscsmrhahlsrevesrvqigdtffepiswhlatkvl gilccglffgivglkiffskfqcrrereawagalfmvpagtgsmlphpaaslllvlasr gpgpkkenpggtglekkartgrlerpetrsggdagsrdgspealrf
#86	ATTCATGGTTCCAGCAGGGAC
#87	GGGAGACAAAGTCACGTACTC

#88	TCCTGGTGTTCGTGGTCTGCTT
#89	GAGAGTCCTGGCTTTTGTGGGC
#90	GSSDLTWPPAIKLG
#91	DRYVAVRHPLRARGLR
#92	VAPRAKAHKSQDSL
#93	CFRSTRHNENSMR
#94	MNGTYNTCGSSDLTWPPAIKLG
#95	RDTSDFLCQLSQG
#96	GIQEGGFCFRSTRHNENSMRFP
#97	AKEFQEASALAVAPRAKAHKSQDSLCLVTLA
#98	TCCTGCTCGTCGCTCTCCTGAT
#99	TCGCTTTTGTGCTATTGTC
#100	HNGSYELSVLMMGNS
#101	NLPTPPTVENQORLA

#102	RKYRKDYELRQKKWSHIPPENIFFLETNETNHVSLKIDDDKRRDTIQRLRQCKYDKKRVILKDLKHNDGN FTEKQKIELNKLQIDYNNLTKFYGTVKLDTMIFGVIEYCERGLREVINDTISYPDGTMDWEEKISVL YDIAGMSYLHSSKTEVHGRLKSTNCVVDSDRMVVKITDFGCNSILPPKKDLWTAPHLRQANISQKGDVY SYGIIAQEIILRKETFYTLSCDRNEKIFRVENSNGMKPFRPDLFLETAEKELEVYLLVKNCWEEDEPK RPFDFKKIETTLAKIFGLFHDQKNESYMDTLIRRLQLYSRNLEHLVEERTQLYKAERDRADRLNFMLLPRL VVKSLKEKGFVEPELYEEVTIYFSDIVGFTTICKYSTPMEVVDMLNDIYKSFHDHVDHHDVYKVETIGDA YMVASGLPKRNGNRHAIDIAKMALEILSFMGTFELEHLPGLEFIWIRIGVHSGPCAAGVVGKMPRYCLFG DTVNTASRMESTGLPLRIHVSGSTIAILKRTECQFLYEVGETYKGRGNETTYWLTGMKDQKFNLPFP TVENQORLOAEFSDMIANSLOKQQAAGIRSQKPRRVASYKKGTTLEYLQLNTTDKESTYF
#103	GCTGGTAACATCTTCCTGC
#104	GAAGAATGTTGTCCAGAGGT
#105	LINKVPLPVDKLAFL
#106	SEAVKKLLEALSHLV
#107	TGTTTTCAACTACCAGGGGC
#108	TGTTGGCTTTGGCAGAGTCC
#109	GAGGCAGAGTTCAGGCTTCACCGA
#110	TSTTGGCTTTGGCAGAGTCC
#111	TGMDMWSTQDLYDNPVTSVFQYEGLRSCVRQSSGFTECRPYFTILGLPAMLQAVR
#112	DQWSTQDLYNNPVTAVFNYQGLWRSCVRESSGFTECRGYFTLL GLPAMLQAVR
#113	STQDLYNNPVTAVE
#114	DMWSTQDLYDNP
#115	CRPYFTILGLPA
#116	TNFWMSTANMYTG
#117	gccaggatca tgtccaccac cacatgccaa gtgggtggcgt tcctcctgtc catcctgggg ctggccggct gcatcgcgcc caccgggatg gacatgtgga gcaccagga cctgtacgac aaccgccgta cctccgtgtt ccagtacgaa gggctctgga ggagctgctg gaggcagagt tcaggcttca ccgaatgcag gccctatttc accatcctgg gacttcacag catgctgcag gcagtgcgag cctgatgat cgtaggcatc gtctgggtg ccattggcct cctgggatcc atctttgccc tgaaatgcat ccgcattggc agcatggagg actctgccaa agccaacatg acactgacct ccgggatcat gttcattgtc tcaggctctt gtgcaattgc tggagtgtct gtgtttgcca acatgctggt gactaacctc tggatgtcca cagctaaccat gtacaccggc atgggtggga tgggtgcagac tgttcagacc aggtacacat ttgggtgggc tctgttcgtg ggctgggtcg ctggaggcct cacactaatt ggggtgtgta tgatgtgcat cgcctgccgg ggcctggcac cagaagaaac caactacaaa gccgtttctt atcatgcctc aggccacagt gttgectaca agcctggagg cttcaaggcc agcactggtt ttgggtccaa caccaaaaac aagaagatat acgatggagg tgcccgaca gaggacgagg tacaatotta tccttccaag cacgactatg tgtaatgtc taagacctct cagcac
#118	MSTTTCQVVAFLLSILGLAGCIAATGMDMWSTQDLYDNPVTSVF QYEGLRSCVRQSSGFTECRPYFTILGLPAMLQAVRALMIVGIVLGAIGLLVSIFALK CIRIGSMEDSAKANMTLTSGIMFIVSGLCAIAGVSVEANMLVTNFWMSTANMYTGMGG MVQTVQTRYTFGAALFVGWVAGGLTLGGVMMCIACRGLAPEETNYKAVSYHASGHSV AYKPGGFKASTGFGSNTKNKLYDGGARTEDEVQSYPSKHDYV
#119	gccaggatca tgtccaccac cacatgccaa gtgggtggcgt tcctcctgtc catcctgggg ctggccggct gcatcgcgcc caccgggatg gacatgtgga gcaccagga cctgtacgac aaccgccgta cctccgtgtt ccagtacgaa gggctctgga ggagctgctg gaggcagagt tcaggcttca ccgaatgcag gccctatttc accatcctgg gacttc
#120	MSTTTCQVVAFLLSILGLAGCIAATGMDMWSTQDLYDNPVTSVFQYEGLRSCVRQSSGFTECRPYFTI
#121	AATGAGAGGAAAGAGAAAAC
#122	ATGGTAGAAGAGTAGGCAAT
#123	EKWNLHKRIALKMVC
#124	CLGFNFKEMFK
#125	TAATGATGAACCCTACACTGAGC

#126	ATGGACAAATGCCCTACCTT
#127	AGTGCTGGAAGGATGTGCGTGT
#128	TTGAGGTGGTTGTTGGGTTT
#129	AGATGTGCTGAGGCTGTAGA
#130	ATGAAGGTTGATTATTGAG
#131	AGCCGCATACTCCCTTACCCTCT
#132	GCAGCAGCCCAACACCACA
#133	CTGAGCCGAGAGGTGGAATC
#134	CTCTCTCGCTTACACTGGAA
#135	QWQVFGPDKPQAL
#136	AKWKGPQGQDLSTDS
#137	NMLVTNFWMSTANMYTGMGGMVQTVQTRYTFG
#138	1 gacagctgtg tctcgatgga gtagactctc agaacagcgc agtttgccct ccgctcacgc 61 agagcctctc cgtggcttcc gcaccttgag cattaggcca gttctcctct tctctetaat 121 ccattcgtca cctctcctgt catccgtttc catgccgtga ggtccattca cagaacacat 181 ccattggtct catgctcagt ttggttctga gtctcctcaa gctgggatca gggcagtggc 241 aggtgttttg gccagacaag cctgtccagg ccttggtggg ggaggacgca gcattctcct 301 gtttctctgt tcttaagacc aatgcagagg ccattggaag gcggttcttc aggggccagt 361 tctctagcgt ggtccacctc tacagggacg ggaaggacca gccatttatg cagatgccac 421 agtatcaagg caggacaaaa ctggtgaagg attctattgc ggaggggcgc atctctctga 481 ggctggaaaa cattactgtg ttggatgctg gcctctatgg gtgcaggatt agttcccagt 541 cttactacca gaaggccatc tgggagctac aggtgtcagc actgggctca gttcctctca 601 tttccatcac gggatatgtt gatagagaca tccagctact ctgtcagtcg tcgggctggt 661 tcccccgcc cccagcgaag tggaaaggto cacaaggaca ggatttgtcc acagactcca 721 ggacaaacag agacatgcat ggcctgtttg atgtggagat ctctctgacc gtccaagaga 781 acgcccggag catatcctgt tccatgcggc atgtctatct gagccgagag gtggaatcca 841 gggtaacagat aggagatacc tttttcgagc ctatatctgt gcacctggct accaaagtac 901 tgggaatact ctgctgtggc ctattttttg gcattgttgg actgaagatt ttcttctcca 961 aattccagtg gaaaatccag gcggaactgg actggagaag aaagcacgga caggcagaat 1021 tgagagacgc ccggaacac gcagtggagg tgactctgga tccagagaag gctcacccga 1081 agctctgcgt ttctgatctg aaaactgtaa cccatagaaa agctccccag gagggtgctc 1141 actctgagaa gagatttaca aggaagagtg tgggtggctc tcagagtttc caagcaggga 1201 aacattactg ggaggtggac ggaggacaca ataaagggtg gcgctgagg gtgtgcccgg 1261 atgatgtgga caggaggaag gactacgtga ctttctctcc cgatcatggg tactgggtcc 1321 tcagactgaa tggagaacat ttgtatttca cattaaatcc ccgttttctc agcgtcttcc 1381 ccaggacccc acctacaaaa ataggggtct tccctggacta tgagtgtggg accatctcct 1441 tcttcaacat aatgaccag tcccttattt ataccctgac atgtcgggtt gaaggcttat 1501 tgaggcccta cattagatg cctctctata atgagcaaaa tggaaactcc atagtcatct 1561 gcccagtcac ccaggaatca gagaaagagg cctcttgcca aagggcctct gcaatcccag 1621 agacaagcaa cagtgaagtc tctcagagg caaccacgcc ctctctccc aggggtgaaa 1681 tgtaggatga atcacatccc acattcttct ttagggatat taaggctctc ctcccagatc 1741 caaagtcccg cagcagccgg ccaaggtggc ttccagatga agggggactg gctgtctcac 1801 atgggagtcg ggtgtcatgg ctgcctgag ctgggaggga agaaggctga cattacattt 1861 agtttgctct cactccatct ggctaagtga tcttgaaata ccacctctca ggtgaagaac 1921 cgtcaggaat tccatctca caggctgtgg tgtagaltaa gtagacaagg aatgtgaata 1981 atgcttagat cttattgatg acagagtga tccaatggt ttgttcatta tattacatt 2041 tcagtaaaaa aa
#139	MALMLSLVLSLLKLGSGQWQVFGPDKPQALVGEDAAFSCLSPKTNAEAMEVRFFRGQFSSVVHLYRDG KDQPFMQMPQYQGRITLVKDSIAEGRISLRLENITVLDAGLYGCRISQSYQKAIWELQVSAIGSVPLI SITGYVDRDIQLLCQSSGWFPRETAWKGPQGQDLSTDSRTNRMHGLFDVEISLTVQENAGSISCSMRH AHLREVESRVQIGDTFFEPISWHLATKVLGILCCGLFFGIVGLKIFFSKFQWKIQAELEDWRRKHGQAE RDARKHAVEVTLPETAHPKLCVSDLKTVTHRKAPQEVPHSEKRFTRKSVVASQSFOAGKHYWEVDGGHN KRWRVGVCRDDVDRKEYVTLSPDHGYWVLRNGEHLFTLNPRFISVFPRTPTKIGVFLDYECGTISF FNINDQSLIYTLTCRFEGLLRPYIEYPSYNEQNGTPIVICPVTQSEKEASWQRASAIPEITSNSESQA TTPFLPRGEM
#140	Tccaaattccagtggaaaatc
#141	ccacactcatagtccaggaag
#142	almivgvlgaigllvsifalkcirigsmedsakanmtltsgimfivsglcalagvsvfanmlvtnfwms tanmytgmggmvqtvtqtrytfgaalfvgwvaggltllogvmmciac
#143	rigsmedsakanmtltsgimflvs
#144	Akanmtlt
#145	Medsakanmtltsg

#146	Medsakadmtltsg
#147	sakadmtlt
#148	akadmtltl
#149	DQWSTQDLYDNPVTAVFNYQGLWRSCVRESSGFTECRGYFTLLGLPAMLOAVR
#150	STQDLYDNPVTAVF

Fig. 42

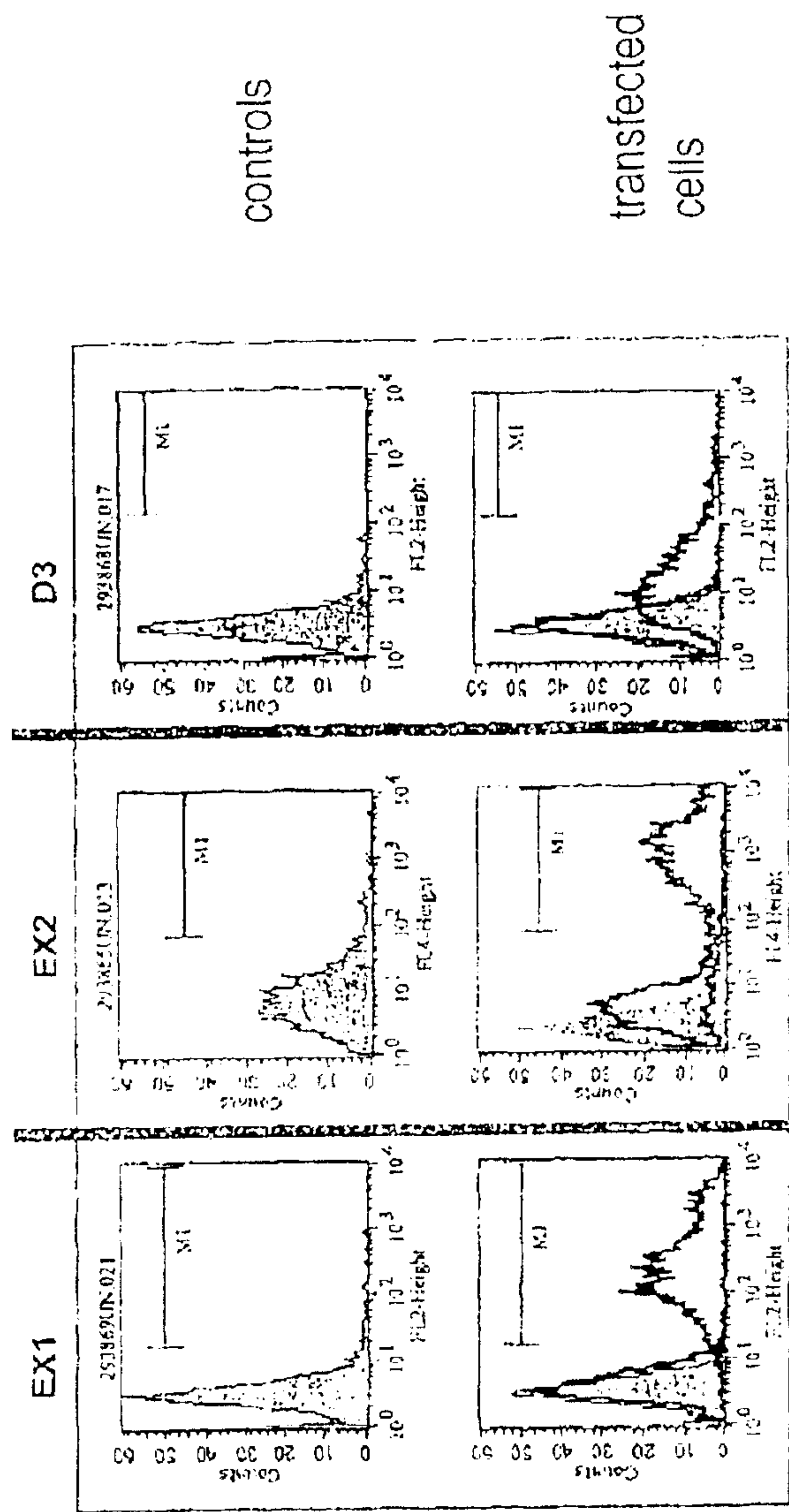
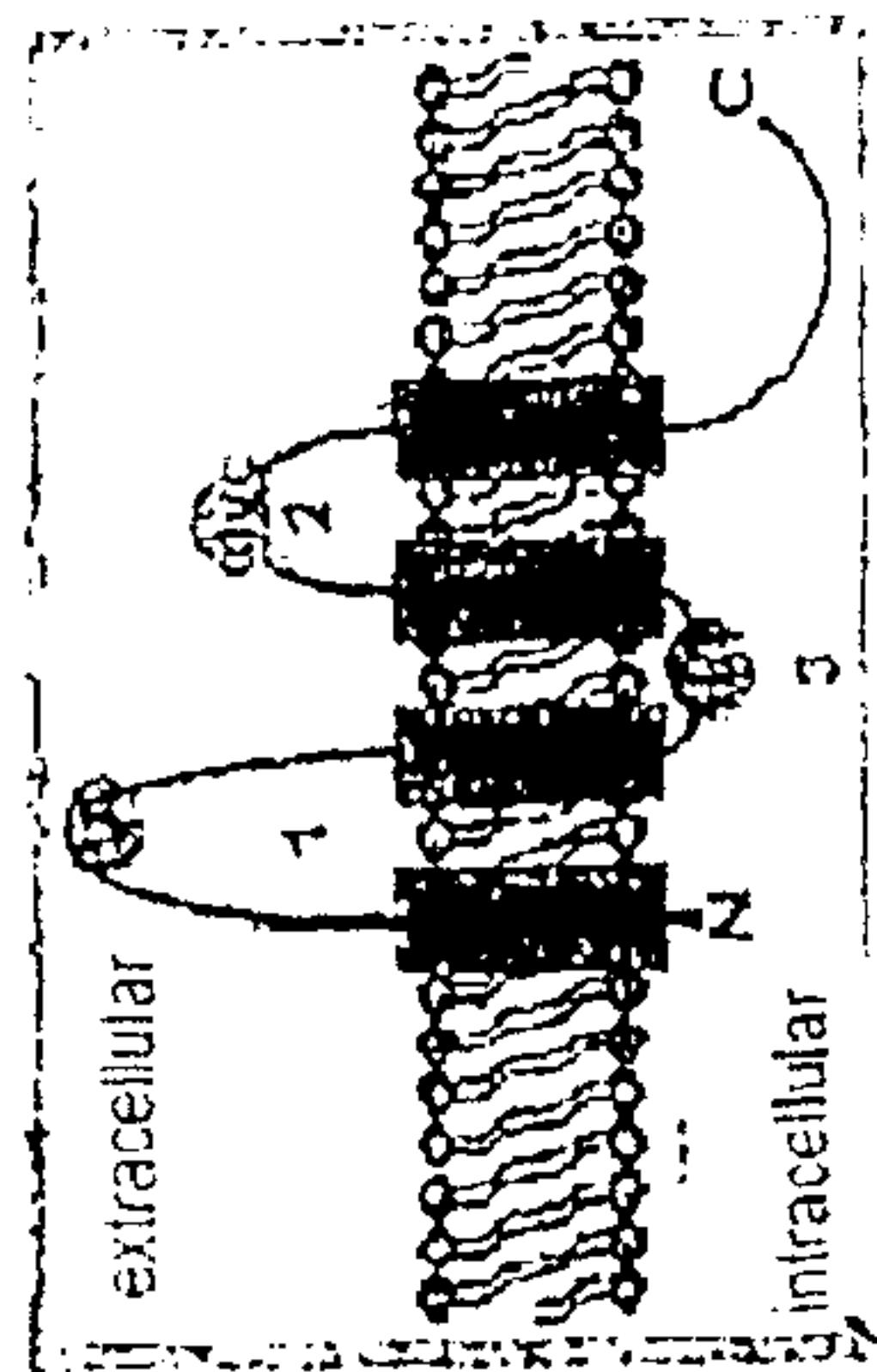


Fig. 43

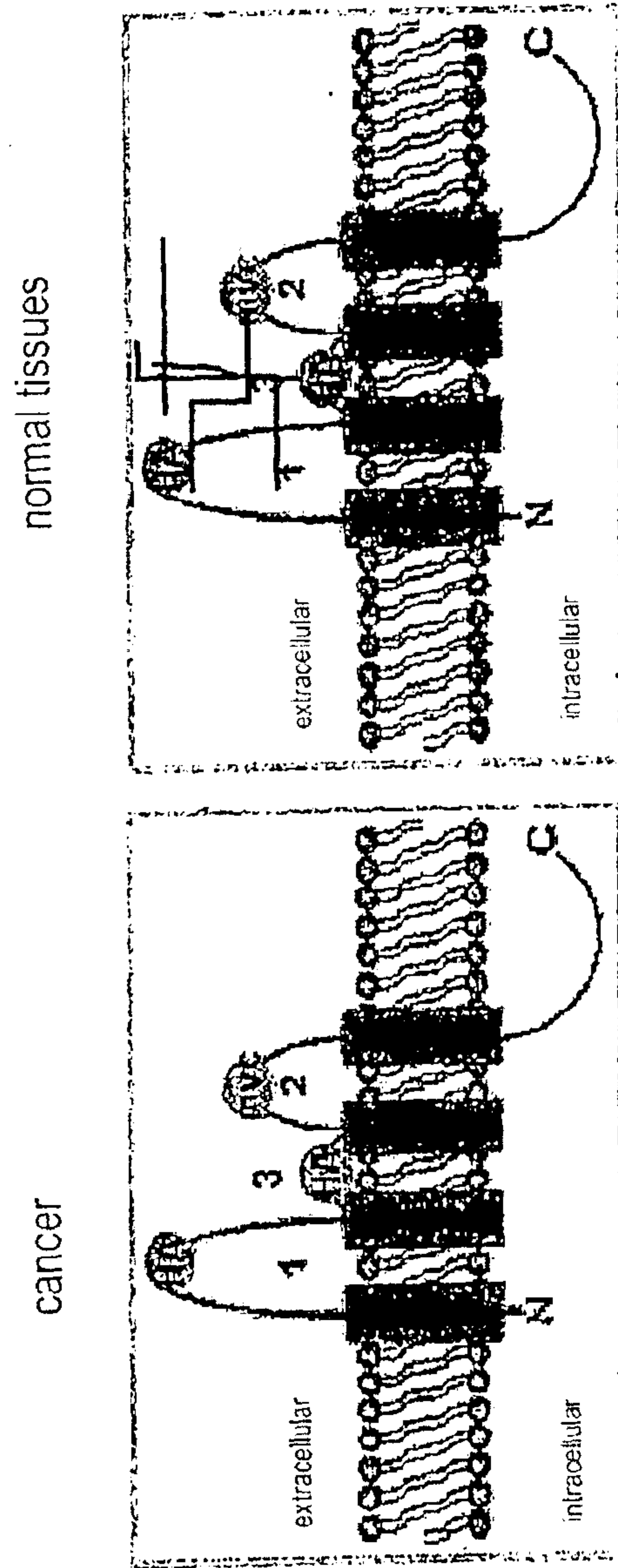
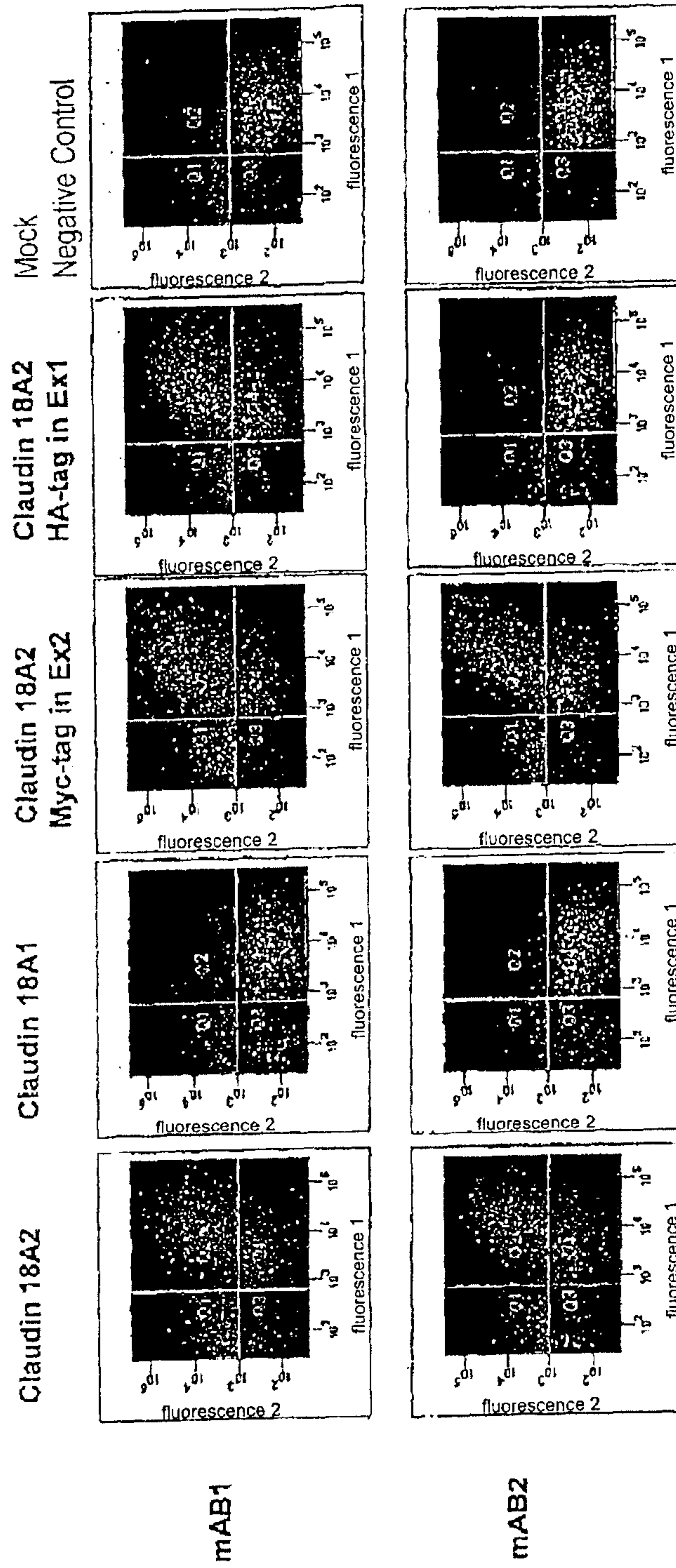


Fig. 44



RT/PCR (normal/tumor)

cloning of PCR-products

colon
normal

colon
tumors



sequencing

deletion of exon 6
(SEQ ID NO.4)

deletion of exon 3
(SEQ ID NO.3)

known variant of GUCY2C