The invention relates to nucleic acid fragments whose sequence is specific for the human villin gene.

By "specific sequence of the human villin gene" is meant any sequence contained in the coding sequence of this gene, any sequence contained in the messenger RNA resulting from the transcription of the said coding sequence, or also a sequence complementary of those just mentioned.

The invention also relates to the use of these nucleic acid fragments to detect the expression of the human villin gene. Among other things, this application of the nucleic acid fragments of the invention makes it possible to detect in vitro the presence of tumor cells derived from primary tumors of the gastric or intestinal region or metastases, starting from a biological sample (tissue or biological fluid) taken from the organs of a patient other than those of the digestive or intestinal sphere. Another use of these fragments consist in the detection of polymorphism of restriction fragments of the human genome.

Experiments performed hitherto have mentioned the importance of villin as a marker of metastases of tissues other than those of the digestive or intestinal region. In this connection the European patent application No. 0206849 filed on 30 April 1986 describes agents for the detection of villin in a biological sample taken from an organ other than those of the digestive or intestinal sphere. The patent application EP 0206849 takes advantage of the fact that villin is generally present, more particularly in the brush borders observed at the surface of the enterocytes, at the apical pole of the elongated cells flush with the internal wall of the intestine and in the cells of the kidney proximal tubule. On the other hand, villin is not detected in other organs such as the lungs, the brain etc...When metastases of digestive or intestinal tumors occur, cells producing villin can migrate into different organs. It is then possible to detect villin in these metastases and diagnose a primary tumor of digestive or intestinal origin.

These diagnostic agents described in the above-mentioned application EP 0206849 and designed to detect the protein translated from the messenger or the gene coding for it are constituted either by antibodies directed against villin or by all or part of a DNA corresponding to the complete mRNA of human villin, or also by a

cDNA fragment containing at least a part of the nucleotide sequence coding for the amino acids of the COOH extremity of human villin.

Up to now only this COOH-terminal part of the nucleic acid for villin has been sequenced.

Generally, villin belongs to the class of proteins having an actin binding activity. Villin exhibits a specificity which is intrinsic to it within the group of proteins capable of binding to actin. This specificity relates, on the one hand, to its double function in vitro in relation to the calcium concentration of the medium and, on the other, to its very specific tissue distribution. Nonetheless, in certain regions of its amino acid sequence, villin exhibits certain sequence homologies with another protein present in the higher eukaryotes, gelsolin.

In particular, comparison of the amino acid sequences of these proteins shows that, except for the COOH-terminal part, they exhibit a similar overall structural organization which comprises a duplicated domain of four motifs, three of which are homologous. Thus villin and gelsolin exhibit, in their common part, an overall amino acid homology of about 50%, this homology being greater in certain regions. It is essentially in the distinctive COOH-terminal part designated "head piece" (end region) that villin differs from gelsolin.

The invention results from the discovery after sequencing the regions coding for villin that, contrary to all expectation, the identity of organization of villin and gelsolin and the homology of their amino acid sequences did not have an equivalent at the level of the nucleic acids coding for villin and gelsolin, respectively.

The inventors have taken advantage of these observations to develop novel agents to detect specifically the expression of the gene coding for human villin. These agents make possible in particular the early detection of the presence of messenger RNA (mRNA) resulting from the transcription of the human villin gene.

The agents in conformity with the invention to detect the expression of the gene coding for human villin are constituted by nucleic acid fragments specific for the sequence coding for the human villin gene or for the transcription of this gene.

It follows from the observation made by the inventors that a nucleic acid fragment comprising at least 8 nucleotides and derived

from the DNA coding for human villin or the messenger RNA transcribed from this coding DNA, except for the sequence coding for the "head piece", may be specific for the human villin gene and, in particular, be distinguished from corresponding fragments of DNA coding for gelsolin.

The invention thus relates to any nucleic acid fragment comprising a sequence of from 8 to 40, and preferably 20 to 40, nucleotides, characterized in that its sequence is contained in the DNA sequence coding for the human villin gene shown in Figure I except for the sequence coding for the "head piece". It also relates to any DNA fragment exactly complementary to one of the fragments just mentioned and which consequently contains the same number of deoxyribonucleotides as well as any corresponding RNA fragment which contains the same number of ribonucleotides.

The invention relates more particularly to the sequences defined above in single stranded form but also in double stranded form in the case of the DNAs.

The previously defined fragments may be isolated by cleavage with optionally trimmed by a suitable enzyme, for example Bal31, starting from the gene coding for human villin, or from the cDNA prepared from the mRNA resulting from the transcription of this gene. These fragments may also be synthesized chemically or enzymatically from the sequence given in Figure I which presents the sequence coding for the human villin gene.

Chemical synthesis of one of the fragments can be performed by implementation of the "phosphodiester" method, the steps of which are summarized below:

The 5'-phosphate end of one of the nucleotides, after protection of the functional groups, reacts (condensation reaction) with the 3'-hydroxyl end of another nucleoside or nucleotide also protected by using dicyclohexylcarbodiimide or an arylsulphonyl chloride. The nucleotide chain is thus elongated by the implementation of similar condensation steps.

Protecting groups for nucleotide functions may consist, for example, of a monomethoxytrityl (MMTr) group, an acetyl (Ac) group or an anisoyl (An) group.

Chemical synthesis can also be carried out by the "phosphotriester" method which makes it possible to block each

internucleotide phosphodiester function as the synthesis of the sequence progresses.

In order to carry out this method a completely protected mononucleotide is used containing a masked 3'-phosphotriester group.

A third method for the chemical synthesis of a nucleic acid fragment is the "phosphite-triester" method. This method is preferably used with the aid of an insoluble support to which the 3' end of the oligonucleotide to be synthesized is bound. This immobilization of the oligonucleotide by coupling makes it possible to avoid the purification steps after each condensation cycle. In order to carry out this method, the mononucleotides suitably blocked beforehand are added sequentially and the reagents and undesired starting materials are removed by filtration.

The protecting groups of the nucleic acid synthesized are removed at the end of the reaction, the nucleic acid is then detached from the support and purified by electrophoresis or HPLC.

When the oligonucleotide synthesized is bound to a polymer-type support in a column, the filtration steps can be replaced by washing.

The methods of chemical synthesis described may profit from the teaching of chapter 2 "Isolation, identification and characterization of DNA fragments" of the monograph entitled "From genes to clones" by E.L. WINNACKER.

It is understood that any region of the coding sequence or the corresponding mRNA may serve as starting point for the production of a nucleic acid fragment of the invention identical with or complementary to these sequences provided that the specificity of the nucleotide sequence of the fragment with respect to the DNA or the mRNA coding for villin is respected.

The invention also relates to the distinctive application which consists of the use of the defined nucleic acid fragments for the constitution of nucleotide primers (designated also as primers) capable of inducing the initiation of the synthesis of all or part of a nucleic acid specific for the villin gene under defined conditions.

By primer is meant a nucleotide sequence (oligonucleotide) advantageously constituted by a nucleic acid fragment defined above, in particular any fragment advantageously comprising from

8 to 40 nucleotides and preferably from 20 to 40 nucleotides and possessing the following properties :

- it is such that it is capable of hybridizing specifically with a corresponding sequence complementary either to the DNA coding for the villin gene or the mRNA corresponding to this gene, or to a sequence which is complementary to them,
- it is capable of inducing the synthesis of a nucleic acid complementary to that to which it hybridizes (this latter is then called the matrix) in a buffered solution and under conditions of pH and temperature suitable for the synthesis in the presence of a synthesis- inducing agent, in particular a polymerase or reverse transcriptase, and appropriate nucleotides. The synthetic product obtained from the fragment serving as primer is called elongation product of the primer.

Buffered solutions and conditions of pH and temperature suitable for the production of the "primer elongation product" can be defined with reference to the teachings of the following patents : US.4683202 and US.4683195, the contents of which are incorporated into the present application.

The primer exists in a single-stranded or double-stranded form. The former is usually more favourable for the required property of hybridization with the nucleic acid in order to be able to initiate the synthesis of the elongation product starting from the primer which is in contact with the matrix.

The use of a double-stranded primer, denatured prior to the hybridization step, may cause competition between the rehybridization of the primer with its complementary strand, on the one hand, and the nucleic acid matrix capable of pairing with one of the strands of the primer, on the other.

The invention relates more particularly to the use of primers in conformity with the invention for the initiation of the synthesis in sufficient quantity to be detected of elongation products of the primer in contact with a nucleic acid coding for human villin, present in the sequence shown in figure I, or from an mRNA sequence resulting from the transcription of the human villin gene or from complementary sequences to those just mentioned, present in a biological sample, these sequences then acting as complementary matrices to the said primers.

Such an application permits for example the detection of a change (mutation, deletion ....) in the nucleic acid fragment constituting the elongation product of the primer.

In connection with the foregoing, the invention hence relates to a method for the in vitro detection of the presence of either a nucleic acid coding for human villin or an mRNA resulting from the transcription of the human villin gene in a biological sample suspected to contain this mRNA, comprising the following steps :

a/ the placing of the biological sample suspected to contain the mRNA resulting from the transcription of the human villin gene in contact with the first primer to which it has previously been made accessible under conditions permitting hybridization between these primers and the mRNA under investigation in the presence of nucleoside triphosphates and a polymerization-inducing agent, and the polymerization from these primers hybridized to the mRNA to produce a duplex formed between the elongation product of the primer hybridized either with the nucleic acid strand coding for human villin or with the corresponding mRNA when they are present in the biological sample,

b/ the denaturation of the duplex obtained in step a/ so as to "separate" the elongation product of the primer from the mRNA to be detected,

c/ the placing in contact of the elongation product obtained with a second primer (in the sense given above to this expression) having a nucleotide sequence (1) not complementary to that of the first primer and (2) complementary to a sequence of the previously formed elongation product,

d/ optionally repetition of steps a/ and b/ and c/ in the presence of the first and second primers used in excess and reagents necessary for the production of further elongation products used in their turn as matrix for additional syntheses until a sufficient quantity is obtained for the detection of the elongation products of the primers used.

e/ the detection of the presence of the elongation products characteristic of the presence of the DNA or mRNA of human villin.

Such a method may be used for example for the in vitro detection of the presence of tumor cells or tumors, or even tumor metastases, provided that they are characterized by the abnormal

presence of the mRNA of human villin. As an example mention should be made of the cells resulting from the tumorigenesis of renal or intestinal tissues.

The implementation of the above detection method is performed on a biological sample taken from a patient and constituted for example by a biopsy, tissue removed during an operation, a biological fluid.

According to a particular embodiment of the above method, the primer or the first and second primers as well as the nucleotides are added at the first step in sufficient quantity to carry out several cycles of synthesis of the elongation product(s), by utilization of the steps of the procedure defined above in order to obtain a desired level of amplification of the mRNA initially present in the biological sample.

According to an alternative embodiment of the method these constituents are incorporated throughout the performance of the steps of the method , as often as required. The nucleotides introduced in the initial step of the above detection method are constituted by deoxynucleotides or by ribonucleotides depending on whether the primers are constituted of DNA or RNA, respectively.

The agent inducing polymerization from the primer hybridized to the nucleic acid serving as matrix is generally constituted by a DNA polymerase when it is required to induce the synthesis of DNA or by a reverse transcriptase to synthesize DNA from a RNA matrix. As examples of polymerases which can be used mention should be made of the DNA polymerase of E. coli, the Klenow fragment of this polymerase, the Taq polymerase or the reverse transcriptase. This list is obviously not exhaustive.

On the contrary, any system, any enzyme which can be used as polymerization catalyst is included in the framework of the invention.

The duplex obtained after polymerization from a primer when the elongation product of the primer and the matrix of synthesis are of different kinds is a heteroduplex , i.e. when mRNA is initially present in the biological sample, then the elongation product of the DNA primer is the cDNA, complementary to the mRNA.

For more details concerning the metohd relating to the detection method described above or for the working out of variants of this method , the specialist will be able to make useful reference to the principles described in the patents US.4.683.202 and US 4.683.195.

If required, the implementation of the method previously described can result in the amplification of a specific part of the mRNA sequence of transcription of the villin gene. In this case recourse will be had to the use of two specific primers forming the 5' and 3' ends of the sequence to be amplified.

The size of the sequence to be amplified is selected in order to permit the detection of the presence of the mRNA under investigation. The polymerization from a primer may hence be deliberately limited by any appropriate means.

The detection of the elongation products of the primers can be performed by the standard detection method. According to one embodiment of the invention the method for the detection of the presence of mRNA is characterized in that the nucleotides incorporated in the course carrying out the method are labelled radioactively and in that the detection of the presence of the mRNA under investigation corresponds to the detection of the radioactivity of the elongation products synthesized after removal of the nucleotides not incorporated during the polymerization from the primers.

Another application of the fragments in conformity with the invention relates to their use for the detection of human genetic polymorphism.

By detection of human genetic polymorphism is meant both the genetic identification of an individual and the detection of certain diseases linked to genetic modifications relating to the villin gene or a genetically linked gene. The identification of an individual may include the analysis of the restriction fragments containing all or part of the villin gene.

In the framework of this application the previously defined nucleic acid fragments are used as probes.

These probes are used on the biological sample, for example a tissue or fluid prepared in order to be accessible to the said probes.

For use of these probes reference should be made to the European patent application 0186 271.

Other characteristics and advantages of the invention will become apparent in the examples which follow and in the Figures :
- Figure I represents the DNA coding sequence of the human villin gene,
- Figure II comprises the coding sequence of Figure I together with the correponding amino acid sequence of human villin,
- Figure III shows the comparison between the coding sequences for villin and gelsolin.
- Figure IV shown the comparative structures of villin and gelsolin and similar proteins present in lower eukaryotes.

SEQUENCING OF THE cDNA COMPLEMENTARY TO THE mRNA OF HUMAN VILLIN

Construction of a cDNA library

The total RNA was isolated from a subclone of a HT-29 cell line (HT29-18-C1; Huet et al., 1987) by the guanidine isothiocyanate method (Chirgwin et al., 1979) and enriched in polyadenylated RNA (poly A^+ RNA) by passage through an oligodT cellulose column (Aviv and Leder, 1972).

The polyadenylated RNA was fractionated according to size on a 5%-20% sucrose gradient and the fractions containing the mRNA of villin were identified by Northern blot analysis using a cRNA probe (complementary RNA) corresponding to the carboxy-terminal part of human villin (Pringault et al., 1986, reference 1).

The cDNA was synthesized from 1 µg of polyadenylated RNA according to the method described by Gubler and Hoffmann (1983). After methylation of the cDNA by a EcoRI methylase, EcoRI linkers were ligated to the blunt ends of the cDNA, digested with EcoRI and separated on a column of Ultrogel AcA34 (LKB) in a TE (20 mM Tris, 1 mM EDTA) buffer. 30 ng of cDNA were ligated with 1 µg of the DNA of a λgt10 vector digested by EcoRI and the particles were encapsidated in vitro ("in vitro packaging") to give a cDNA bank containing 10^6 independent recombinants.

In order to isolate the 5' part of villin cDNA, 3 successive extensions from primers were carried out using oligonucleotide

sequences complementary to the villin mRNA. The oligonucleotide probes are underlined in figure II. The oligonucleotides were hybridized in a molar ratio of 10 : 1 with 2 µg of mRNA polyA$^+$ fractionated according to size. The synthesis of cDNA was then carried out under the conditions already described. The λgt10 library was screened with cRNA probes obtained by in vitro transcription (Melton et al., 1984 reference 2) of the cDNA clones already characterized.

NORTHERN BLOT Analysis

The RNA poly A$^+$ (1.5 ug) isolated from the HT29-18-C1 cell line was fractionated by electrophoresis on 1% agarose gels in the presence of 1M formaldehyde (Lehrach et al., 1977 reference 3) and transferred to nitrocellulose. The blots were pre-hybridized at 55°C for 16 hours in 50% formamide, 4 x SSC, 0.05 M Na_2HPO_4 at pH 7.4 in 1x Denhardt's solution, 250 µg/ml of denatured salmon sperm DNA and 500 µg/ml tRNA. The hybridization with an RNA probe labelled with P^{32} (2×10^6 c.p.m./ml) was carried out for 24 hours at 55°C in the same solution without tRNA. The blots were washed twice in 1 x SSC, 0.1% SDS, once in 0.2 x SSC, 0.1% SDS at 65°C and finally once in 0.1 x SSC, 0.1% SDS for 30 min. at 70°C.

Sequence analysis

cDNA restriction fragments were subcloned in M13mp18-mp19 derivatives and sequenced by the dideoxy chain termination method described by SANGER et al. (1977 reference 4). The data on the sequences of the two strands are consistent.

Protein sequence analysis

Human villin was purified by B. West, L. West and M. Mooseker (Biology department, University of Yale) from brush border cells isolated from a human subject (Carboni et al., 1987 reference 5) by methods described by Coleman and Mooseker (1985 reference 6). Before sequencing 25 µg of villin were subjected to electrophoresis on a 10% SDS-polyacrylamide gel and electrotransferred to a glass fiber membrane (Whatman GF/C) coated with poly(4-vinyl-N-methylpyridine). The details concerning this method are given in BAUW et al. (1987 reference 7). The immobilized protein was detected by a dilute fluorescamine stain (1 mg/l acetone), detached from the membrane and transferred to the reaction chamber of a gas phase protein sequenator, Applied

Biosystems 470A operating according to the manufacturer's instructions.

An initial sequencing yield of about 10% of the protein deposited on the gel is obtained. It is known that the transfer of protein is rarely quantitative (at most between 60 and 90%) and that NH_2-terminal blocking resulting from an artefact occurs during gel electrophoresis (in most cases more than 50% of the protein is blocked, see Moos et al., 1988 reference 8). That is why the value obtained is not abnormally low and probably reflects the NH_2-terminal sequence of the majority of the villin molecules, rather than the result of a proteolytic cleavage close to the NH_2-terminal part of a fraction of the villin molecules which would be blocked at the NH_2-terminus.

LITERATURE REFERENCES

-(1) Pringault,E. et al 1986. A human villin cDNA clone to investigate the differentiation of intestinal and kidney cells in vivo and in culture. EMBO J.5:3119-3124

-(2) Melton,D.A. et al 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucl. Acids Res.12:7035-7056.

-(3) Lehrach,H. et al 1977. RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. Biochemistry,16: 4743-4751.

-(4) Sanger,F.et al. 1977. DNA sequencing with chain-terminating inhibitors. Proc.Nat.Acad.Sci.74:5463-5467.

-(5) Carboni,J.M.et alm.1987. Characterization of intestinal brush border cytoskeletal proteins of normal and neoplastic human epithelial cells : a comparison with the avian brush border. Am.J.Pathol. 129 : 589-600.

-(6) Coleman,T.R. et al.1985. Effects of actin filament cross-linking and filament length on actin-myosin interaction.J. Cell Biol.101 : 1850-1857.

-(7) Bauw,G.et al. 1987. Alterations in the phenotype of plan cells studied by NH_2-terminal amino acid-sequence analysis of proteins electroblotted from two-dimensional gel-separated total extracts. Proc.Natl.Acad.Sci.84 : 4806-4810.

-(8) Moos,M.Jr.et al.1988. Reproducible high yield sequencing of proteins electrphoretically separated and transferred to an inert support. J.Biol.Chem.263:6005-6008.

* C L M S *

CLAIMS

1/. Nucleic acid fragment characterized in that it comprises from 8 to 40 nucleotides and in that its sequence is contained either in the DNA sequence coding for the gene for human villin, with the exception of the sequence coding for the "head piece", or in an DNA fragment exactly complementary to one of the foregoing and consequently containing the same number of deoxynucleotides, or in any corresponding RNA fragment containing the same number of ribonucleotides.

2/. Nucleic acid fragment according to Claim 1, characterized in that its sequence comprises from 20 to 40 nucleotides.

3/. Use of the nucleic acid fragments according to any one of the Claims 1 or 2 as primers for the initiation of the synthesis of a detectable quantity of elongation products of the primer in contact with a nucleic acid coding for human villin present in the coding sequence presented in Figure I, or from a mRNA sequence resulting from the transcription of the gene for human villin or from sequences complementary to the foregoing, these sequences then acting as complementary matrices of the said primers.

4/. Method for in vitro detection of the presence of either a nucleic acid coding for human villin or an mRNA strand resulting from the transcription of the gene for human villin in a biological sample suspected to contain this mRNA, comprising the following steps :

a/ the placing of a biological sample suspected to contain the mRNA resulting from the transcription of the gene for human villin, after it has been made accessible, in contact with a first primer constituted by a nucleic acid fragment according to either of the Claims 1 or 2, in the presence of nucleoside triphosphates and polymerization-inducing agent and under conditions permitting the hybridization between this primer and the mRNA under investigation, and the polymerization from this primer hybridized to the mRNA, the nucleic acid complementary (cDNA) to this mRNA, in order to produce a duplex formed between the elongation product of this first primer, hybridized either with the nucleic acid strand coding for human villin or with the

corresponding mRNA when they are present in the biological sample,
b/ denaturation of the duplex obtained in step a/ so as to "separate" the elongation product of the primers from the mRNA to be detected,
c/ the placing of the elongation product obtained in contact with a second primer (in the sense given above to this expression) having a nucleotide sequence (1) not complementary to that of the first primer and (2) complementary to a sequence of the previously formed elongation product,
d/ optionally, the repetition of the steps a/, b/ and c/ in the presence of the first and second primers used in excess and the reagents necessary for the production of more elongation products used in turn as matrix for additional syntheses until a detectable quantity of the elongation products of the primers used is obtained,
e/ detection of the presence of the elongation products characteristic of the presence of the DNA or mRNA of human villin.
5/. Method : according to Claim 4, characterized in that the nucleoside triphosphates are radioactively labelled and that the detection of the presence of the mRNA under investigation corresponds to the detection of the radioactivity of the elongation products synthesized, after removal of the nucleoside triphosphates not incorporated during polymerization from the primers.
6/. Method according to Claim 4, characterized in that the nucleoside triphosphates are deoxyribonucleoside triphosphates and in that the polymerization-inducing agent is a DNA polymerase or a reverse transcriptase.
7/ Method according to either of the Claims 4 or 5 for the detection of a tumor or metastasis, characterized in that it is performed on a biopsy whose mRNAs resulting from the transcription of the villin gene possibly present have been made accessible to the primers used.
8/. Probe for the detection of polymorphism of the restriction fragments of the human genome characterized in that it comprises a nucleic acid fragment according to either of the Claims 1 or 2.
9/. Nucleic acid fragment, characterized in that it comprises at least 8 nucleotides and in that its sequence is contained either in the DNA sequence coding for the gene for human villin, with the exception of

* C L M S *

the sequence coding for the "head piece", or in a DNA fragment exactly complementary to one of those previously mentioned and which consequently contains the same number of nucleotides, or in any corresponding RNA fragment containing the same number of ribonucleotides, this fragment being, in addition, distinct from a fragment corresponding to the DNA coding for gelsoline.