Volume 17 Number 10 1989 Nucleic Acids Research

Characteristics of a multicopy gene family predominantly consisting of processed pseudogenes

Ans M.W.Van den Ouweland[1], Hans L.P.Van Duijnhoven[1], Klaus A.Deichmann[3], Jan J.M.Van Groningen[1], Lou de Leij[2] and Wim J.M.Van de Ven[1,3]

[1]Molecular Oncology Section, Department of Biochemistry, University of Nijmegen, St Adelbertusplein 1, 6525 EK Nijmegen, [2]Department of Clinical Immunology, University of Groningen, Groningen, The Netherlands and [3]Center for Human Genetics, University of Leuven, Belgium

Received March 9, 1989; Revised and Accepted April 21, 1989 EMBL accession nos*

ABSTRACT

The monoclonal antibody MOC-32 detected a 40 kDa protein in Western blot analysis. Immunological screening of an expression library of human SCLC cells with MOC-32 led to the isolation of overlapping cDNA clones. One of these, cHD4, was 1.0 kbp long and of about the same size as its corresponding mRNA. Preceded by an in phase stop codon, an open reading frame of 885 bp was present in cHD4 and a translational product of only 33 kDa could be calculated. Biochemical and immunological analysis established the relationship between the 40 kDa antigen and the isolated coding sequences and resolved the apparent discrepancy between the calculated molecular weight and the observed electrophoretic mobility. Nucleotide sequence comparison of cHD4 to the EMBL database revealed that cHD4 was nearly identical to a sequence claimed to encode a laminin binding protein. Southern blot and nucleotide sequence analysis indicated the presence of multiple copies of the gene in the human genome. At least five of these appeared to represent processed pseudogenes.

INTRODUCTION

Processes involved in tumorigenesis can be defined in terms of alterations in gene expression patterns. One approach to study such alterations uses differential hybridization analysis to identify genes whose expression is affected (1,2). Another approach is based upon the development of monoclonal antibodies with specificity for proteins whose synthesis is induced or inhibited in the process of tumorigenesis (3,4). One should be very careful however in interpreting results obtained by this approach. It is well known that epitopes which are recognized by antibodies often exhibit a limited complexity. Therefore, they may be present in a number of unrelated proteins and this emphasizes the need for complementary data, for instance at the mRNA level, when the monoclonal antibody approach is used in defining tumor cells.

MOC-32 is a monoclonal antibody that was obtained by immunizing mice with the human small cell lung carcinoma (SCLC) cell line GLC-1-M13 (5). By immunoperoxidase analysis of specimens of a variety of organs it was shown

that MOC-32 reacted with endocrine cells of the pancreas, the medulla of the thyroid, stromal cells of the ovary, brain cells, peripheral nerves and ganglion cells (6). The reactivity was found mainly at the surface of the cells (L. de Leij, unpublished observations). The protein detected by MOC-32 is probably a (neuro)endocrine differentiation antigen, which is also present in fetal lung tissue (6). Examination of lung tissue in various stages of development revealed expression of MOC-32-specific antigenic determinants in epithelial cells of very young fetal lungs (first trimester in gestation only). These epithelial cells still lacked the typical morphology of pulmonary neuroendocrine cells. Interestingly, expression of the MOC-32 antigen was also observed in small cell lung carcinomas (6). As outlined above, these observations required further substantiation at the nucleic acid level to exclude the possibility that the observed differential expression pattern was due in part to the presence of a common epitope on different proteins. Therefore, experiments were performed to identify the protein(s) containing the epitope recognized by MOC-32 and to characterize the gene(s) encoding them. In this study, we describe the identification of a 40 kDa protein as target for MOC-32 and characterization of the genetic sequences that encode this protein. Based upon nucleotide sequence data, features of the protein are predicted. Furthermore, results are presented indicating that the gene encoding the 40 kDa protein is a member of a multigene family which includes at least five processed pseudogenes.

MATERIALS AND METHODS

Cell lines: Cell lines were grown in RPMI (GIBCO) supplemented with 10% fetal calf serum. Cell lines used in this study included GLC-1, GLC-1-M13 (5), SCLC-16HV, SCLC-16HC (7) and human breast carcinoma cell line MCF 7 (This cell line was obtained from EG & G Mason, Research Institute, Worcester, MA, USA).

Immunoblotting analysis: Cell lysates were prepared in lysis buffer containing 350 mM NaCl, 10 mM Tris-HCl pH 7.6 and 1.5 mM $MgCl_2$. Protein samples were analyzed by SDS-polyacrylamide gel electrophoresis and, thereafter, proteins were transferred to nitrocellulose electrophoretically (8). Incubation of nitrocellulose filters with antisera was performed essentially as described by Mulders et al. (9).

RNA isolation and hybridization: Total cellular RNA was isolated using the lithium-urea procedure described by Auffray and Rougeon (10). Ten µg of oligo(dT)-cellulose purified mRNA was glyoxalated, fractionated on 1% agarose

gels (11) and transferred to Hybond-N (Amersham). Isolation of DNA probes, their nick-translation and hybridization of blots was carried out as described before (12).

Construction and screening of cDNA libraries: An oligo(dT)-primed cDNA library of GLC-1-M13 was constructed in lambda gt11 as described by Roebroek et al. (13). About $7x10^5$ plaques obtained upon infection of E.coli Y1090 (14) were screened using monoclonal antibody MOC-32, diluted in RIA buffer (0.3% BSA, 150mM NaCl, 10 mM Tris-HCL pH 7.5, 1% w/v Triton X-100, 0.5% Na-deoxycholate and 0.1% SDS), as described by Young and Davis (15). To detect bound antibody, ^{125}I-labelled sheep anti-mouse Ig[F(ab)$_2$ fragment] was used. With the cDNA insert of cHD2, a positive clone obtained from the screening with the MOC-32 antiserum, another $1x10^6$ plaques were screened using the method of Hanahan and Meselson (16).

Construction and screening of a genomic library: For the isolation of human genomic DNA sequences homologous to the cDNA clones obtained from the lambda gt11 expression library described above, about 300.000 colonies of a cosmid library were screened using cHD4 as a molecular probe under hybridization conditions as described before (17).

DNA sequence analysis: DNA fragments were ligated into the polylinker region of M13mp18-19 (18). Sequencing of the DNA fragments was according to the dideoxy method of Sanger et al. (19). Gel readings were recorded, edited and compared using the Staden programs (20).

In vitro transcription and translation: For in vitro transcription, cDNA inserts were subcloned in pSP65 and digested with restriction endonuclease BamHI. The linearized cDNA subclones were transcribed in vitro using SP6 polymerase (Promega) and the RNA's were analyzed on non-denaturated 1% agarose gels. In vitro translations in a nuclease-treated reticulocyte lysate (Amersham) was performed for 90 minutes at 30°C. Reaction mixtures were analyzed by SDS-PAGE. Gels containing ^{35}S-labelled proteins were impregnated with 2,5-diphenyloxazole (PPO) before being dried and autoradiography was performed using Kodak XAR-5 X-ray film.

RESULTS

Identification of a 40 kDa protein as target for MOC-32.

As an approach to characterize the protein(s) recognized by monoclonal antibody MOC-32, analysis of the corresponding gene(s) was persued. Therefore, experimental conditions were selected to isolate the coding sequences for the protein(s) that contained the MOC-32-specific epitope from

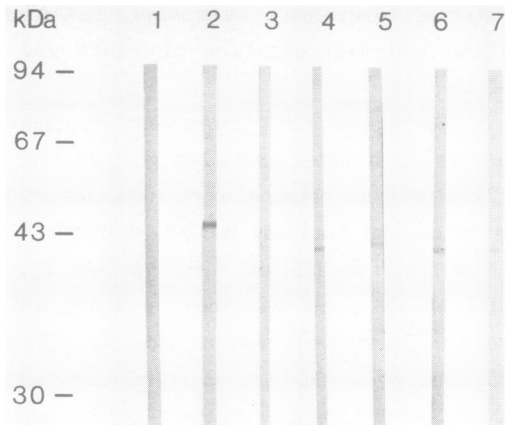


Fig. 1.
Western blot analysis of small cell lung carcinoma cell lines. Cell lysates of GLC-1-M13 (lanes 1,2 and 4), GLC-1 (lane 3), SCLC-16HV (lane 5), SCLC-16HC (lane 6) and MCF 7 (lane 7) were analyzed by SDS-polyacrylamide gel electrophoresis and proteins were transferred to nitrocellulose for immunological screening. As a negative control, medium of the SP2/0 cell line was used (lane 1) and as a positive control, a monoclonal antibody against cytokeratine (RCK 106; lane 2). Proteins in lanes 3-6 were analyzed with MOC-32.

a cDNA expression library using the monoclonal antibody in an immunological screening procedure.

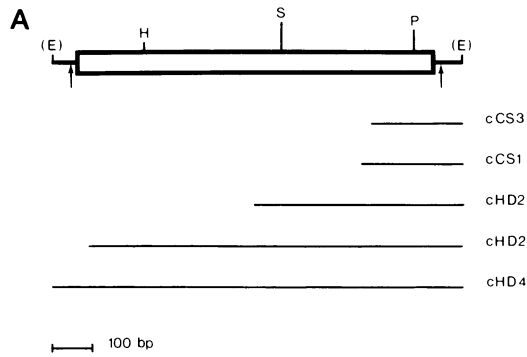
To first establish the identity of protein(s) containing the epitope recognized by MOC-32, Western blot analysis of a number of small cell lung carcinoma (SCLC) cell lines was performed. Such analysis of the cell lines GLC-1, GLC-1-M13, SCLC-16HV and SCLC-16HC revealed only the detection of a 40 kDa protein (Fig. 1, lanes 3-6). The 40 kDa protein was also detected in the breast carcinoma cell line MCF7 (Fig. 1 lane 7). The highest levels of the protein were present in GLC-1-M13 and, therefore, this cell line was selected for further studies. Detection of the 40 kDa protein was also possible under experimental conditions required for immunological screening of a cDNA expression library (for conditions, see Materials and Methods; results not shown). Apparently, the epitope recognized by MOC-32 is not altered significantly under these conditions. The 40 kDa protein is probably not a glycoprotein, because in GLC-1-M13 cells grown in the presence of tunicamycin (2 µg/ml), still a protein with a molecular weight of 40 kDa could be detected (data not shown). This is an important observation because detection of the unglycosylated protein moiety synthesized in bacteria using an

expression vector could be hampered if the 40 kDa eukaryotic protein would have been a glycoprotein.

Isolation of cDNA clones containing the coding sequences for the 40 kDa protein.

To isolate cDNA clones containing the coding sequences for the 40 kDa protein, a cDNA library was constructed in lambda gt11 using poly(A)-selected mRNA from GLC-1-M13. About $7x10^5$ plaques of the amplified cDNA library were screened with MOC-32 in RIA buffer. Screening resulted in the isolation of three positive cDNA clones (cCS1, cCS3 and cHD2) (Fig. 2A). The cDNA clones varied in lenght between 200 and 500 bp. Nucleotide sequence analysis of these cDNA clones revealed that they were derived most likely from the same mRNA. Using the largest cDNA clone (cHD2) as a probe in Northern blot analysis, only one mRNA of about 1.2 kb was detected in a number of SCLC cell lines, a primary SCLC tumor and the breast carcinoma cell line MCF7 (Fig. 3). Apparently, the cDNA clones obtained sofar were not full lenght and, therefore, a second screening was performed with cHD2 as a molecular probe. This resulted in the isolation of a large number of positive clones. Of these, 24 were studied by Southern blot analysis and the two largest cDNA clones (cHD4 and cHD24) were selected for nucleotide sequence analysis (Fig. 2A). In Fig. 2B, the complete nucleotide sequence of cHD4, which is 1.0 kbp long and thus about the same size as the corresponding mRNA, is shown. In this cDNA clone, the consensus sequence for a poly(A)-addition signal "AATAAA" is located at position 1001. Only one large open reading frame was found (Fig. 2A and B). The ATG start codon for this large open reading frame is located at position 64 and, interestingly, it is preceded by an in-frame TAA stop codon at position 45. This indicates that the ATG codon at position 64 is indeed the start codon. The stop codon of the large open reading frame is located at nucleotide position 949, thus giving an open reading frame for a protein of 295 amino acids with a calculated molecular weight of 33 kDa (Fig. 2B). This calculated molecular weight is in disagreement with the molecular weight of 40 kDa observed in Western blot analysis. This discrepancy is not likely due to glycosylation since the deduced amino acid sequence did not contain consensus sequences for N-linked glycosylation sites (NXS or NXT; 21). Furthermore, experiments with tunicamycin also indicated that the 40 kDa protein is not glycosylated.

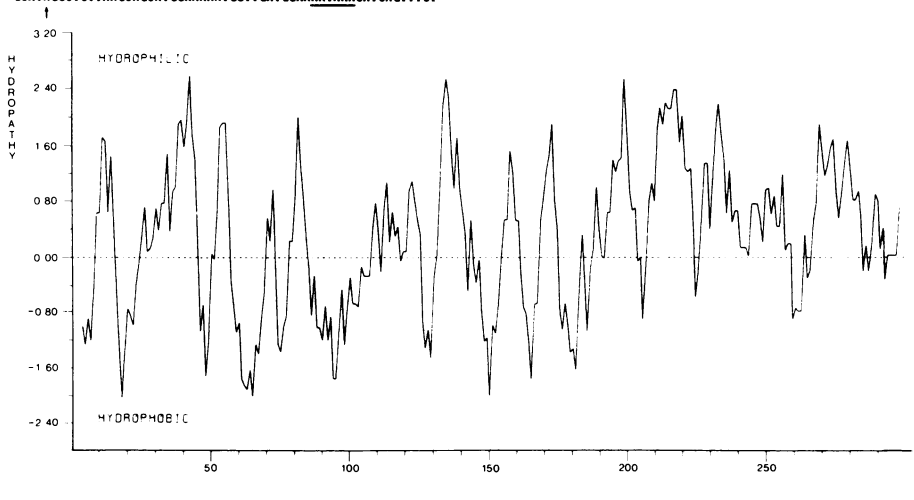
One of the isolated cDNA clones, clone cHD24, appeared to have a nucleotide sequence that was aberrant from that of clone cHD4. It contained an insertion of 15 bp. This insertion (GTTTGTGGAACAGTG) is located at



B

```
1                                                  61
TGCAGAGGGGTCCATACGGCGTTGTTCTGGATTCCCGTCGTAACTTAAAGGGAAACTTTCACAATGTCCGGAGCCCTTGATGTCCTGCAAATGAAGGAGGAGGATGTCCTTAAGTTCCTT
                                                   M  S  G  A  L  D  V  L  Q  M  K  E  E  D  V  L  K  F  L

121                                                181
GCAGCAGGAACCCACTTAGGTGGCACCAATCTTGACTTCCAGATGGAACAGTACATCTATAAAAGGAAAAGTGATGGCATCTATATCATAAATCTCAAGAGGACCTGGGAGAAGCTTCTG
 A  A  G  T  H  L  G  G  T  N  L  D  F  Q  M  E  Q  Y  I  Y  K  R  K  S  D  G  I  Y  I  I  N  L  K  R  T  W  E  K  L  L

241                                                301
CTGGCAGCTCGTGCAATTGTTGCCATTGAAAACCCTGCTGATGTCAGTGTTATATCCTCCAGGAATACTGGCCAGAGGGCTGTGCTGAAGTTTGCTGCTGCCACTGGAGCCACTCCAATT
 L  A  A  R  A  I  V  A  I  E  N  P  A  D  V  S  V  I  S  S  R  N  T  G  Q  R  A  V  L  K  F  A  A  A  T  G  A  T  P  I

361                                                421
GCTGGCCGCTTCACTCCTGGAACCTTCACTAACCAGATCCAGGCAGCCTTCCGGGACGCCACGGCTTCTTGTGGTTACTGACCCCAGGGCTGACCACCAGCCTCTCACGGAGGCATCTTAT
 A  G  R  F  T  P  G  T  F  T  N  Q  I  Q  A  A  F  R  E  P  R  L  L  V  V  T  D  P  R  A  D  H  Q  P  L  T  E  A  S  Y

481                                                541
GTTAACCTACCTACCATTGCGCTGTGTAACACAGATTCTCCTCTGCGCTATGTGGACATTGCCATCCCATGCAACAACAAGGGAGCTCACTCAGTGGGTTTAATGTGGTGGATGCTGGCT
 V  N  L  P  T  I  A  L  C  N  T  D  S  P  L  R  Y  V  D  I  A  I  P  C  N  N  K  G  A  H  S  V  G  L  M  W  W  M  L  A

601                                                661
CGGGAAGTTCTGCGCATGCGTGGCACCATTTCCCGTGAACACCCATGGGAGGTCATGCCTGATCTGTACTTCTACAGAGATCCTGAAGAGATTGAAAAAGAAGAGCAGGCTGCTGCTGAG
 R  E  V  L  R  M  R  G  T  I  S  R  E  H  P  W  E  V  M  P  D  L  Y  F  Y  R  D  P  E  E  I  E  K  E  E  Q  A  A  A  E

721                                                781
AAGGCAGTGACCAAGGAGGAATTTCAGGGTGAATGGACTGCTCCCGCTCCTGAGTTCACTGCTACTCAGCCTGAGGTTGCAGACTGGTCTGAAGGTGTACAGGTGCCCTCTGTGCCTATT
 K  A  V  T  K  E  E  F  Q  G  E  W  T  A  P  A  P  E  F  T  A  T  Q  P  E  V  A  D  W  S  E  G  V  Q  V  P  S  V  P  I

841                                                901
CAGCAATTCCCTACTGAAGACTGGAGCGCTCAGCCTGCCACGGAAGACTGGTCTGCAGCTCCCACTGCTCAGGCCACTGAATGGGTAGGAGCAACCACTGACTGGTCTTAAGCTGTTCTT
 Q  Q  F  P  T  E  D  W  S  A  Q  P  A  T  E  D  W  S  A  A  P  T  A  Q  A  T  E  W  V  G  A  T  T  D  W  S  ↑

961
GCATAGGCTCTTAAGCAGCATGGAAAAATGGTTGATGGAAAATAAACATCAGTTTCT
          ↑
```

C



position 316-330 and it does not result in a frame shift of the open reading frame. Furthermore, it should be noted that the 5'-end of this insertion resembles the concensus sequence of a 3' splice junction. The possibility therefore exists that cHD24 represents a not fully spliced precursor RNA or an alternatively spliced RNA.

Biochemical and immunological characterization of the protein encoded by cDNA cHD4.

To resolve the discrepancy between the calculated and observed molecular weight of the protein, an in vitro transcription and translation analysis was performed to characterize the protein that can be synthesized using cHD4. The insert of cHD4 was subcloned in pSP65 and RNA was synthesized in vitro. This RNA was translated in a reticulocyte lysate. The translation products were analyzed using SDS-polyacrylamide gel electrophoresis. Fig. 4, lane 3 shows that a major product of about 40 kDa and two minor products of about 38 and 30 kDa could be detected. Western blot analysis of the proteins synthesized in the reticulocyte lysate using MOC-32 did not result in the detection of any protein (Fig. 4, lane 1). Apparently, the amount of the 40 kDa protein synthesized under these conditions is not enough to allow detection. Mixing of the ^{35}S-labelled in vitro product with proteins of a lysate of GLC-1-M13 resulted in co-migration of the protein detected with MOC-32 and the protein synthesized in vitro with RNA derived from cHD4 (Fig. 4 lanes 2 and 4). This result showed that the protein with the calculated molecular weight of 33 kDa had an apparent electrophoretic mobility in SDS-PAGE of 40 kDa. As an alternative approach to provide more evidence that the epitope recognized by MOC-32 was indeed located in the amino acid sequence encoded by cHD4 and that, for instance, the β-galactosidase portion of the bacterial fusion protein is not involved, additional recombinant DNA constructs were made.

Fig. 2.
A. Restriction endonuclease map of the isolated cDNA clones. The open bar represents the large open reading frame and the arrows (↑) indicate the position of the stop codons in-frame with the ATG of the large open reading frame. E, EcoRI; H, HindIII; P, PstI; S, SacI. B. Nucleotide sequence of the cDNA clones encoding the 40 kDa protein. Arrows (↑) indicate the positions of the stop codons which are in-frame with the ATG of the large open reading frame. The poly(A)-addition signal is underlined. The amino acid sequence of the 40 kDa protein is shown in the conventional one letter code. A symmetrical sequence of an hexapeptide (LMWWML; position 580-597) and a number of repeats (e.g. TEDWSAQP position 853-876 and TEDWSAAP, position 880-903; AAATGA, position 334-351 and AAAEKA, position 709-726) are present in the nucleotide sequence. C. Hydropathy plot of the 40 kDa protein as determined by the computer program of Kyte and Doolittle (22).

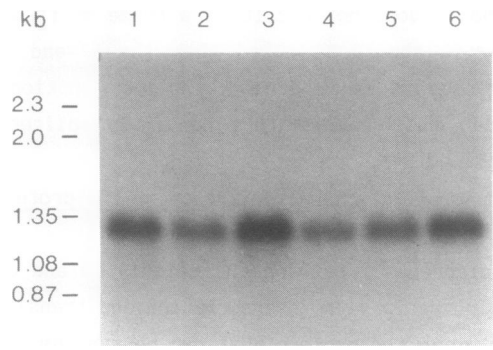


Fig. 3.
Identification of transcripts encoding the 40 kDa protein. Poly(A)-selected RNA was isolated from the cell lines GLC-1 (lane 1), GLC-1-M13 (lane 2), SCLC-16HV (lane 3), SCLC-16HC (lane 4), a primary human small cell lung tumor (A18022; lane 5) and the human breast carcinoma cell line MCF 7 (lane 6) and screened in Northern blot analysis with the insert of cHD4. Molecular weight markers included lambda DNA and PhiX174 DNA digested with the restriction endonucleases HindIII and HaeIII, respectively.

These included DNA constructs between cHD2, detected in the original screening of the expression library and the bacterial gene trpE or the gene for glutathion transferase. Also in these two cases, the fusion proteins were detectable in Western blot analysis by MOC-32 (Data not shown). Furthermore, indications were obtained as to where the epitope that was recognized by MOC-32 mapped in the protein. cDNA clone cCS3, which was isolated by immunological screening with MOC-32, started at nucleotide position 785. Therefore, the epitope recognized by MOC-32 is present in the 55 amino acids encoded by cCS3 and, thus, located at the carboxy-terminal portion of the 40 kDa protein. Using a computer program developed by Kyte and Doolittle (22), the hydropathy profile of the cHD4-encoded protein (Fig. 2C) was determined. This profile indicated that the region of amino acids 265-295 had a hydrophilic nature which is generally found for protein sequences constituting an epitope.

Identification of multiple processed pseudogenes corresponding to the gene encoding the 40 kDa protein.

In a first approach to characterize the genomic organization of the gene that encoded the 40 kDa protein, Southern blot analysis was performed (Fig. 5). Human genomic DNA was digested with the restriction endonucleases BamHI, EcoRI or HindIII (Fig. 5, lanes 1-3) and screened with cHD4 as a molecular

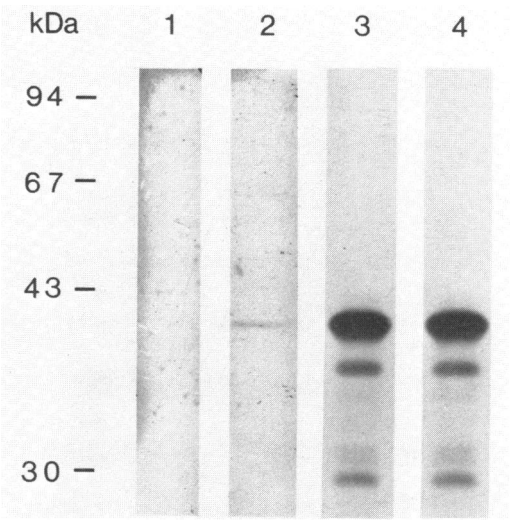


Fig. 4.
Western blot analysis of the in vivo and in vitro synthesized 40 kDa protein. Cell lysate of the in vitro synthesized ^{35}S-labelled 40 kDa protein (lanes 1-4) and of GLC-1-M13 (lanes 2 and 4) were analyzed by SDS-polyacrylamide gel electrophoresis. The Western blot was first incubated with MOC-32 (lanes 1 and 2) after which autoradiography of the same blot was performed (lanes 3 and 4).

probe. The results indicated that either the gene is composed of a great number of small exons or, more likely, that their exist multiple identical or highly related copies of it in the human genome. More insight was obtained by Southern blot analysis of human genomic DNA digested with a combination of the restriction endonucleases HindIII and PstI and hybridized with the 600 bp HindIII/PstI cDNA fragment from cHD4. Such analysis resulted in a strong hybridization signal with a DNA fragment of 600 bp. In addition, a number of weaker signals at various positions were observed (Fig. 5, lane 5). These data strongly indicated the presence in the human genome of multiple highly related copies of this gene. To study this further, a human genomic library (an amplified cosmid library) was screened with cHD4 and a number of positive clones were isolated. These clones could be divided into five different categories (Data not shown). Nucleotide sequence analysis revealed that each category represented a different pseudogene. The complete nucleotide sequence of two of these were established and compared to that of cDNA clone cHD4 (Fig. 6). The large open reading frame observed in cHD4 was no longer present

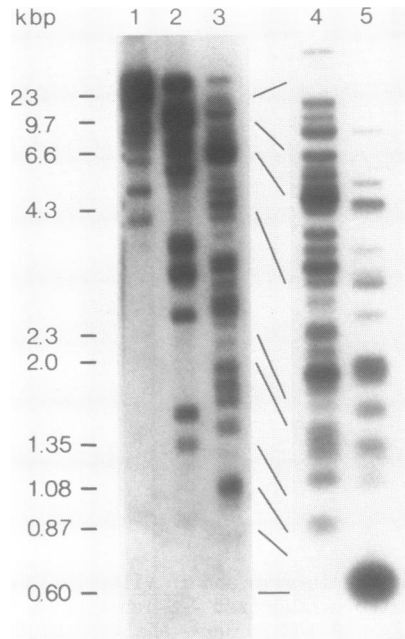


Fig. 5.
Identification of genetic sequences in human cellular DNA homologous to cDNA sequences encoding the 40 kDa protein. High molecular DNA was prepared from human blood cells and digested with restriction endonuclease BamHI (lane 1), EcoRI (lane 2), HindIII (lane 3 and 4), and HindIII/PstI (lane 5) and size-fractionated by agarose gel electrophoresis. Upon transfer of DNA to Hybond-N, hybridization analysis was performed under high-stringency conditions, using 50% formamide, with ^{32}P-nick-translated probes cHD4 (lanes 1-3) and the 600 bp HindIII/PstI insert of cHD4 (lane 4 and 5). Upon hybridization, filters were washed using 0.1xSSC + 0.1% SDS at 65°C in the final step. Molecular weight markers included are HindIII-digested lambda DNA fragments.

in these two fully sequenced pseudogenes. Recently, another pseudogene was described (23) so at least six pseudogenes are present in the human genome. The genomic sequence encoding the 40 kDa protein remains to be identified and the molecularly cloned 15 bp insertion in cDNA clone cHD24 could be instrumental in achieving this. Comparison of the restriction maps of the pseudogenes to the Southern blots presented in Fig. 5 indicated that the presence of more pseudogenes can be expected.

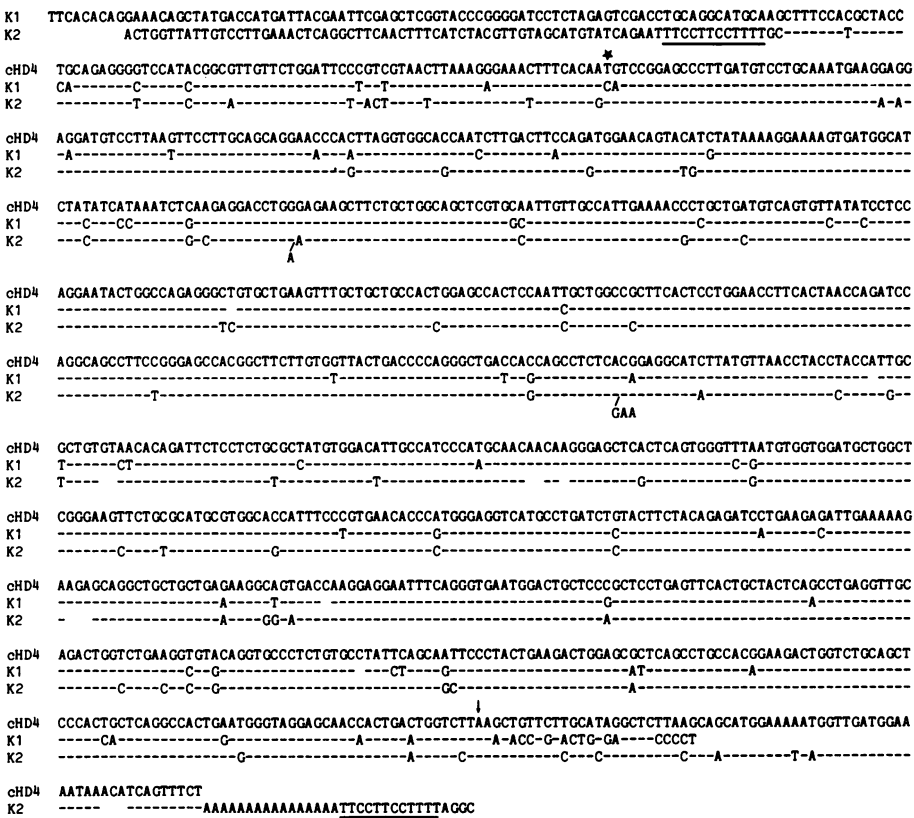


Fig. 6.
Comparison of cHD4 and two of its pseudogenes. The topline shows the nucleotide sequence of cHD4 and underneath the sequence of two different pseudogenes (K1 and K2). Bases which are the same as those in cHD4 are underneath (-). Additions and changes are marked by the relevant bases. The ATG of the large open reading frame as present in clone cHD4 is marked by an asterisk (*) and the stop codon by an arrow (↓). The 12 bp terminal repeats present in K2 are underlined.

DISCUSSION

Used as immunohistochemical tools, monoclonal antibodies are valuable to study pathodifferentiation in a rather direct and sensitive way. Such an approach may have limitations, since reactivity of a monoclonal antibody is based on a limited complexity of the one epitope against which it is directed. Therefore, on a larger antigen basis, a monoclonal antibody may not

always meet the required specificity. Independently derived data are needed in such a situation to confirm the antigen-specificity of a monoclonal antibody. In the present study, we report on the genetic sequences that encode the antigen recognized by monoclonal antibody MOC-32, which is presently used in clinical and experimental pathology of human lung cancer (6). Western blotting indicated that MOC-32 recognized as its only target a nonglycosylated protein of 40 kDa. Using immunological screening of an expression library, several cDNA clones, all corresponding to one single genetic locus, were isolated and from these the complete coding sequence of the gene was identified. The open reading frame contained the coding capacity for 295 amino acids with a calculated molecular weight of 33 kDa. From hybridization studies, it appeared that the protein is encoded by a 1.2 kb mRNA. The apparent discrepancy between the deduced molecular weight of 33 kDa and the observed molecular weight in Western blotting of 40 kDa appeared to be due to aberrant migration in SDS-PAGE analysis. This is so because it was found that the protein synthesized in vitro using the cloned coding sequences comigrated with the mature protein synthesized in a lung carcinoma cell line. Taken together, the biochemical and immunological data presented in this report are in agreement with the assumption that the molecularly cloned genetic sequences indeed correspond to the gene that encodes the 40 kDa protein recognized by MOC-32. To firmly establish the relationship between the 40 kDa protein produced in the lung carcinoma cell lines and the in vitro synthesized protein, a biochemical analysis such as V8 protease peptide mapping would be useful. However, this appeared impossible since MOC-32 can not be used to immunoprecipitate the 40 kDa protein (Unpublished observations). Therefore, such a proof awaits the further development of monoclonal antibodies that enable immunoprecipitation of the 40 kDa protein from cell lysates for comparative peptide mapping analysis.

Southern blotting analysis of human genomic DNA with the cHD4 cDNA clone as a molecular probe provided a first indications that this gene is present in the human genome (and also the mouse genome (data not shown)) in the form of multiple, highly related copies. Upon nucleotide sequence analysis of a number of cosmid clones isolated using cDNA cHD4, it appeared that the gene for the 40 kDa protein is a member of a multicopy gene family that predominantly consists of processed pseudogenes. It can not be excluded yet that there is more than one active mRNA producing gene for the 40 kDa protein. As can be seen in Fig. 6., the nucleotide sequence of pseudogene K2

is flanked by a repeat of 12 nucleotides. One could assume that the genetic sequence flanked by these repeats resembles the complete transcript of the gene for the 40 kDa protein.

Based upon the nucleotide sequence of cHD4, a number of features of the protein can be predicted. First of all, no consensus sequences for N-linked glycosylation were found. This is in agreement with the tunicamycin experiments which indicated that the 40 kDa protein is not glycosylated. Furthermore, the hydropathy plot in Fig. 2C revealed that the amino terminus of the protein is more hydrophobic which could indicate membrane anchoring of the protein via its amino terminal region. In immunohistochemical studies the protein was found to be localized in the cell membrane (L. de Leij, unpublished observation). Finally, a number of repeats and a symmetrical sequence of an hexapeptide are present (Fig. 2B). The implications of these sequences for the function of the protein remain to be established.

Comparison of the nucleotide sequence of cHD4 to already published sequence data revealed that the nucleotide sequence of a partial cDNA clone published by Wewer et al. (24) was identical to sequences in cHD4 except for one point mutation at position 582 (G instead of A; Fig. 2). In a recent report (25), additional sequence data of the same locus were presented and these were also identical to the sequences found in cHD4 except for one point mutation at position 56 (T instead of C; Fig. 2). Neither of the point mutations resulted in an amino acid substitution. Nucleotide sequence data of the mouse homolog are also known (26). Wewer et al. isolated their partial cDNA clone using a monoclonal antibody that detected the laminin receptor on immunoblots and reacted with purified laminin receptor in ELISA (24). A polyclonal antibody was prepared by Wewer et al. (27) using a synthetic polypeptide of twenty amino acids whose sequence was deduced from the nucleotide sequence of the cDNA. This polyclonal antibody reacted with the laminin receptor on immunoblots and affected laminin-mediated cell attachment (27). From their data, Wewer et al. concluded that they had molecularly cloned the laminin receptor. However, the laminin receptor they purified had a molecular weight of about 67 kDa, which exceeds the coding capacity of the now fully characterized locus. Glycosylation can also not account for the difference in molecular weight. This discrepancy needs to be resolved.

ACKNOWLEDGEMENTS

We thank G. Bepler for providing the lung cancer cell lines SCLC-16HV and SCLC-16HC, A.J.M. Roebroek for his advice on this work, C.H.H. Schoenmakers

and H. Spakman for their excellent technical assistance and J.A. Leunissen for computer analysis. This work was supported by the Netherlands Cancer Foundation, Contract No. NUKC Bioch. 87-13.

*EMBL accession nos X15005-X15007 (incl.)

REFERENCES

1 Schalken, J.A., Ebeling, S.B., Isaacs, J.T., Treiger, B., Bussemakers, M.J.G., de Jong, M.E.M. and Van de Ven, W.J.M. (1988) Cancer Res. 48, 2042-2046.
2 Matrisian, L.M., Glaichenhaus, N., Gesnel, M-C. and Breathnach, R. (1985) EMBO J. 4, 1435-1440.
3 De Leij, L., Poppema, S., Klein Nulend, J., ter Haar, A., Schwander, E., Ebbens, F., Postmus, P. and The, T.H. (1985) Cancer Res. 45, 2192-2200.
4 Waibel, R., O'Hara, C.J. and Stahel, R.A. (1987) Cancer Res. 47, 3766-3770.
5 De Leij, L., Postmus, P.E., Buys, C.H.C.M., Elema, J.D., Ramaekers, F., Poppema, S., Brouwer, M., van der Veen, A.Y., Mesander, G. and The, T.H. (1985) Cancer Res. 45, 6024-6033.
6 De Leij, L., Broers, J., Ramaekers, F., Berendsen, H. and Wagenaar, Sj.Sc. (1987) in Application of monoclonal antibodies in tumor pathology (Ruiter, D.J., Fleuren, G.J. and Warnaar, S.O. eds.), Martinus Nijhof, Dordrecht.
7 Bepler, G., Jaques, G., Koehler, A., Gropp, C. and Havemann, K. (1988) J. Cancer Res. Clin. Oncol. 114, 235-244.
8 Towbin, H., Staehelin, T. and Gordon, J. (1979) Proc. Natl. Acad. Sci. USA 76, 4350-4354.
9 Mulders, J.W.M., Voorter, C.E.M., Lamers, C., de HaardHoekman, W.A., Montecucco, C., van de Ven, W.J.M., Bloemendal, H. and de Jong, W. (1988) Curr. Eye Res. 7, 207-219.
10 Auffray, C. and Rougeon, F. (1980) Eur. J. Biochem. 107, 303-314.
11 McMaster, G.K. and Carmichael, G.G. (1977) Proc. Natl. Acad. Sci. USA 74, 4835-4838.
12 Van den Ouweland, A.M.W., Van Groningen, J.J.M., Schalken, J.A., Van Neck, H.W., Bloemers, H.P.J. and Van de Ven, W.J.M. (1987) Nucl. Acids Res. 15, 959-970.
13 Roebroek, A.J.M., Schalken, J.A., Bussemakers, M.J.G., van Heerikhuizen, H., Onnekink, C., Debruyne, F.M.J., Bloemers, H.P.J. and Van de Ven, W.J.M. (1986) Molec. Biol. Rep. 11, 117-125.
14 Huynh, T.V., Young, R.A. and Davis, R.W. (1985) in DNA cloning techniques, a practical approach (Glover, D. ed.), IRL Press, Oxford.
15 Young, R.A. and Davis, R.W. (1983) Proc. Natl. Acad. Sci. USA 80, 1194-1198.
16 Hanahan, D. and Meselson, M. (1980) Gene 10, 63-67.
17 Schalken, J.A., Van den Ouweland, A.M.W., Bloemers, H.P.J. and Van de Ven, W.J.M. (1985) Biochim. Biophys. Acta 824, 104-112.
18 Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33, 103-119.
19 Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
20 Staden, R. (1982) Nucl. Acids Res. 10, 2951-2961.
21 Yamamoto, T., Davis, C.G., Brown, M.S., Schneider, W.J., Casey, M.L., Goldstein, J.L. and Russell, D.W. (1984) Cell 39, 27-38.
22 Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105-132.

23 Segui-Real, B., Rhodes, C. and Yamada, Y. (1989) Nucl. Acids Res. 17, 1257.
24 Wewer, U.M., Liotta, L.A., Jaye, M., Ricca, G.A., Drohan, W.N., Claysmith, A.P., Rao, C.N., Wirth, P., Coligan, J.E., Albrechtsen, R., Mudryj, M. and Sobel, M.E. (1986) Proc. Natl. Acad. Sci. USA 83, 7137-7141.
25 Yow, H., Wong, J.M., Chen, H.S., Lee, C., Steele, G.D. and Chen, L.B. (1988) Proc. Natl. Acad. Sci. USA 85, 6394-6398.
26 Makrides, S., Chitpatima, S.T., Bandyopadhyay, R. and Brawerman, G. (1988) Nucl. Acids Res. 16, 2349.
27 Wewer, U.M., Taraboletti, G., Sobel, M.E., Albrechtsen, R. and Liotta, L.A. (1987) Cancer Res. 47, 5691-5698.