Cloning and Expression of Alternative Transcripts of a Novel Neuroendocrine-specific Gene and Identification of Its 135-kDa Translational Product

(Received for publication, December 9, 1992, and in revised form, February 12, 1993)

Anton J. M. Roebroek,† Helgi J. K. van de Velde§, Adrie Van Bokhoven‡, Jos L. V. Broers,‡, Frans C. S. Ramaekers,‡ and Wim J. M. Van de Ven‡

From the †Laboratory for Molecular Oncology, Center for Human Genetics, University of Leuven, Herestraat 49, B-3000 Leuven, Belgium and the ‡Department of Molecular Cell Biology and Genetics, University of Limburg, Universiteitsring 50, 6229 ER Maastricht, The Netherlands

Monoclonal antibodies RNL-2 and RNL-3 were previously shown to react with four 35–45-kDa proteins, expressed only in small cell lung carcinoma (SCLC) NCI-H82 cells, but to stain a subset of neuroendocrine tissues and neoplasms (Broers, J. L. V., Mijnheere, E. P., Klein Rot, M., Schaart, G., Sijlmans, A., Boerman, O. C., and Ramaekers, F. C. S. (1991) Cancer 67, 619–633). We used RNL-2 and RNL-3 to isolate cDNA sequences that code for proteins containing the two corresponding epitopes and utilized such cDNAs to develop second generation antibodies. Using these antibodies, we identified a novel 135-kDa protein. The corresponding cDNAs were found to belong to a previously unknown gene with a neuroendocrine-specific expression pattern. Tentatively designated NSP gene. NSP transcription appeared to result in mRNAs of 3.4 and 1.8 kilobases (kb). In the NCI-H82 cells only, an apparently aberrant transcript of 2.3 kb was found. cDNAs containing coding sequences of the 3.4-, 2.3-, and 1.8-kb transcripts were isolated, and nucleotide sequence analysis revealed extensive sequence overlap and open reading frames for proteins of 776, 356, and 208 amino acids, respectively. The three deduced proteins all have a common carboxy-terminal part with two large hydrophobic regions. Transfection of the complete coding sequences of the 3.4-kb transcript resulted in the production of a protein with an apparent molecular mass of 135 kDa. This protein is predicted to be highly negatively charged (calculated pI of 4.35), to be rich in proline and serine, and to contain multiple potential phosphorylation sites.

On the basis of biochemical, histological and morphological characteristics, human lung carcinomas can be subdivided into two main subgroups, i.e. small cell lung carcinomas (SCLCs) and non-SCLCs. The latter groups include aequorin cell carcinomas, adenocarcinomas, adenosquamous carcinomas, and large cell carcinomas (1, 2). Moreover, the lung carcinoid tumors constitute a minor subgroup with typical overt neuroendocrine features. The SCLCs, the most aggressive pulmonary neoplasms, display a more restricted neuroendocrine differentiation pattern. Finally, in the subgroup of non-SCLCs, neuroendocrine differentiation is found in about 10% of the cases and seems to correlate with a more aggressive behavior (3). Neuroendocrine features of SCLC tumors are generally studied in the many cell lines that have been established from primary SCLC tumors. Such cell lines can be subdivided into two major categories, i.e. classical (about 70%) and variant (30%) type SCLC cell lines (4). Cell lines of the classical type possess the most pronounced neuroendocrine phenotype, whereas these neuroendocrine characteristics are lost to some extent in cell lines of the variant type. For instance, we recently showed that the 7B2 protein, which is selectively expressed in neurons and endocrine cells with secretory granules, is almost exclusively expressed in SCLC cell lines of the classical type (5). Of potential clinical interest is the observation that SCLC cell lines of the classical type are in general more sensitive to radiotherapy and chemotherapy (4) than variant SCLC cell lines.

In previous studies (6), monoclonal antibodies RNL-2 and RNL-3 were raised against whole cell preparations of the human lung carcinoma cell line NCI-H82, a SCLC cell line of the variant type. Both antibodies were shown to recognize distinct epitopes. In Western blot analysis performed so far, the two antibodies appeared to detect proteins only in NCI-H82 cells; the four proteins were a triplet ranging in molecular mass between 44 and 45 kDa and a protein of about 35 kDa. The interrelationship between these proteins remains to be established, but they are most likely unique to NCI-H82 cells. In contrast to the Western blot studies, immunohistochemical studies indicated that RNL-2 and RNL-3 could stain a subset of neuroendocrine tissues and neoplasms. With respect to lung carcinoma cells, both antibodies appeared to react only with some SCLCs and carcinoid tumors of the lung but in general not with non-neuroendocrine lung carcinomas. Altogether, these results suggest that both antibodies recognize epitopes in so far elusive neuroendocrine-specific proteins. Antibody RNL-2 has been submitted to the Second International Workshop on Small Cell Lung Cancer (7), and cluster analysis revealed that reactivity of the antibody does not

The abbreviations used are: SCLC, small cell lung carcinoma; non-SCLC, non-small cell lung carcinoma; PBS, phosphate-buffered saline; kb, kilobase(s); kbp, kilobase pair(s); bp, base pair(s); PAGE, polyacrylamide gel electrophoresis.
correspond to any of the other submitted antibodies.

To identify candidate neuroendocrine proteins that could have been the target proteins of both RNL-2 and RNL-3 in the previous immunohistochemical studies, we used second generation antibodies that were developed using as immunogen the translation product of newly isolated human cDNA sequences that contained the epitopes recognized by the two antibodies; such cDNA sequences were identified by screening a human expression cDNA library with RNL-2 and RNL-3. As a potential candidate, a 135-kDa protein was identified. To determine characteristics of this protein, the corresponding genetic sequences were isolated and their expression was studied in transfection experiments. The nucleotide sequence of these genetic sequences was determined, and, based on these data, characteristics of the 135-kDa protein could be predicted. Furthermore, the expression pattern of these genetic sequences was established by Northern blot analysis.

**Materials and Methods**

Cell Lines and Tissues—Cell lines used in this study included the SCLC cell lines NCI-H82 (8), SCLC-21H and SCLC-22H (9), GLC-1 and GLC-1-M13 (10), and the non-SCLC cell line HTB-182 (ATCC). Furthermore, human neuroblastoma cell lines CCL-127 and HTB11, human larynx carcinoma cell line CCI-23, human oral carcinoma cell line CCI-17, human pharynx carcinoma cell line CCL-138, mouse pituitary adenoma cell line AT-T20, rat pituitary adenoma cell line GH3, rat pheochromocytoma cell line PC-12, and rat medullary thyroid carcinoma cell line 6-23 were used and these cell lines were all obtained from the ATCC. Cell lines were grown in Dulbecco’s modified Eagle’s medium/Ham’s F-12 medium (1:1) supplemented with 10% fetal bovine serum or, in the case of cell line 6-23, with 10% calf serum POL-1 (lane 1, NCI-H82; lane 2, SCLC-21H; lane 3, SCLC-21H; lane 4, GLC-1-M13; lane 5, GLC-1. B, Western blot analysis of DNA of a Novel Neuroendocrine-specific Gene

**Construction and Screening of cDNA Libraries**—Purified mRNA was isolated from total RNA of cell lines NCI-H82 and GLC-1 and of a primary carcinoid tumor of the lung by oligo(dT)-cellulose affinity chromatography. Oligo(dT)-primed cDNA libraries were constructed using a cDNA synthesis kit of Promega (in case of GLC-1) or Pharmacia (in case of NCI-H82) and the cDNA synthesis kit of the lung. cDNAs were cloned into Agt11 using EcoRI linkers (in case of the GLC-1 cDNA library) or EcoRI adaptors (in case of the cDNA libraries of NCI-H82 and carcinoid tumor). About 3 x 10^6 plaques of the NCI-H82 cDNA library, obtained upon infection of Escherichia coli Y1096, were screened according to the procedure described by Young and Davis (14). In this screening procedure, monoclonal antibodies RNL-2 and RNL-3 were used, which were raised against NCI-H82 cells (6). Bound antibody was detected using ^125I-labeled sheep anti-mouse Ig (Fab)2 fragment. Subsequently, additional clones were isolated from the libraries using as molecular probes ^32P-labeled cDNA inserts of positive clones from the antibody screening. Hybridization was performed as previously described (15).

**Nucleotide Sequence Analysis**—For nucleotide sequence analysis, cDNA inserts were subcloned into M13mp18–19 or pUC18 and analyzed according to the deoxy chain termination method using standard primers or primers designed and synthesized on the basis of newly obtained sequence data. The nucleotide sequences were obtained from both strands and analyzed using the sequence analysis computer programs Genepro (Riverside Scientific), PC/Gene, and IntelliGenetics (IntelliGenetics, Inc.).

**Production of Bacterial Hybrid Proteins**—For production of bacterial hybrid proteins, isolated cDNA clones were subcloned into the bacterial expression vectors pUR290 and pUR292 (16). In these expression vectors, the open reading frame of three different cDNAs was cloned in frame near the 3’-end of the open reading frame of the β-galactosidase gene. A 1.2-kbp BamHI-Sall cDNA fragment of pAB49 (from BamHI site in pUC18 polylinker to an internal Sall site downstream of the stop codon) was cloned into pUR292 and will be referred to as pAB49/pUR292. Similarly, cloning of a 1.8-kbp BamHI-Sall cDNA fragment of pAB31 resulted in pAB31/pUR290. Cloning of an internal 2.8-kbp BamHI-Sall fragment of the cDNA insert of pAB124 resulted in pAB124/pUR290. For the production of hybrid proteins, standard protocols were used. (12). Rabbits were subcutaneously immunized and rechallenged twice with 200 μg of partially purified bacterial fusion protein of pAB124/pUR290. The generation of monoclonal antibodies MON-160, MON-161, and MON-162 and the detailed characterization of these will be published elsewhere.

**Western Blot and Immunoprecipitation Analysis**—Western blot analysis was performed as described by Van Duijnoven et al. (17). Briefly, cell pellets of the human SCLC cell lines NCI-H82, SCLC-21H, SCLC-22H, GLC-1, and GLC-1-M13 were lysed in lysis buffer (62.5 mM Tris- HCl (pH 6.8), 12.5% glycerol, 2% Nonidet P-40, containing as protease-inhibitors 2.5 mM phenylmethylsulfonyl fluoride, 1.25 mM EDTA, and 12.5 μg/ml leupeptin). Approximately equal protein amounts of cell lysates were loaded onto 10% polyacrylamide
cDNAs of a Novel Neuroendocrine-specific Gene

RESULTS AND DISCUSSION

Identification of a 135-kDa Protein Expressed in SCLC Cell Lines—In previous studies (6), it was reported that the monoclonal antibodies RNL-2 and RNL-3 recognize a triplet of proteins of about 45 kDa and a protein of about 35 kDa in immunoblotting experiments of the SCLC cell line NCI-H82 (see Fig. 1A, lane 1). It should be noted that these cells were also used for immunization of the mice that were used to generate the monoclonal antibodies. Expression levels of these proteins in the NCI-H82 cells were readily detectable. Similar Western blot analysis of other SCLC cell lines (Fig. 1A, lanes 2–5) failed to detect these proteins, suggesting that expression levels of the four proteins in these other SCLC cell lines are rather low and, therefore, escape detection or that these four proteins are unique to NCI-H82. However, using sensitive immunohistochemical techniques, staining of some primary SCLC tumors and neuroendocrine tissues was observed in the same studies (6). These observations suggested that antigens with epitopes recognized by RNL-2 or RNL-3 are more widely expressed but escaped detection by Western blot analysis. In an attempt to detect such antigens, second generation antibodies were developed using as immunogen the translational product of newly isolated (see below) human cDNA sequences, which were selected on the basis that they should encode the two epitopes recognized by the two monoclonal antibodies. Using a partially purified bacterially expressed hybrid protein that consisted of β-galactosidase and protein sequences containing both the epitope recognized by RNL-2 and the one recognized by RNL-3, a polyclonal rabbit serum (POL-1) and second generation mouse monoclonal antibodies (MON-160, MON-161, and MON-162) were obtained (details will be published elsewhere). Western blot analysis of the lung carcinosa cell line NCI-H82 using the polyclonal rabbit antiserum revealed the expression of the four 35–45-kDa proteins, detected by RNL-2 and RNL-3, and an additional 135-kDa protein (Fig. 1B, lane 1). Immunoprecipitation analysis of [35S]methionine-labeled proteins in lysates of cell line NCI-H82 with a mixture of monoclonal antibodies MON-160, MON-161, and MON-162 revealed the detection of only the 135-kDa protein (Fig. 1C, lane 1). The 135-kDa protein was also found expressed in other SCLC cell lines, such as for instance in SCLC-21H (Fig. 1C, lane 2). The protein could not be detected in GLC-1-M13 and the non-SCLC cell line HTB-182 (Fig. 1C, lanes 3 and 4). Retrospective immunoprecipitation analysis using RNL-2 and RNL-3 revealed that with these antibodies low amounts of a 135-kDa protein could be detected in NCI-H82 and SCLC-21H (data not shown). Altogether, these results support our assumption that expression of the 35–45-kDa proteins is most likely restricted to NCI-H82 cells. Furthermore, they make the 135-kDa protein a potential candidate to be the antigen recognized by RNL-2 and RNL-3 in the immunohistochemical studies of tissues that do not seem to express the low molecular mass (35–45-kDa) proteins. The observation that these 35–45-kDa NCI-H82 proteins and the 135-kDa protein all contain the two complete open reading frame of the 3.4-kb NSP transcript, was designated pSVL-NSP.4. In pSVL-NSP.4, expression of the NSP cDNA sequences is under control of the SV40 late promoter. Monkey kidney COS-1 cells were propagated in Iscove’s modified minimal medium, supplemented with fetal calf serum (10% v/v) and antibiotics (100 units/ml penicillin and 100 μg/ml streptomycin). Twenty-four hours after seeding, the semi-confluent cells were transfected with 20 μg of DNA in 2 ml of Iscove’s modified minimal medium, supplemented with 200 μg/ml DEAE-dextran. The transfection procedure used included a chloroquine shock (21). Thirty-six hours after transfection, cells were labeled with [35S]methionine, and immunoprecipitation analysis of NSP-encoded proteins was performed as described above.
epitopes recognized by RNL-2 and RNL-3 suggests a possible relationship between these proteins. Molecular characterization of the genetic sequences encoding these proteins resolved code proteins containing both the epitope recognized by RNL-2 and RNL-3 suggests a possible relationship between these proteins. Molecular characterization of the genetic sequences encoding these proteins resolved code proteins containing both the epitope recognized by RNL-2 and RNL-3 suggests a possible relationship between these proteins. Molecular characterization of the genetic sequences encoding these proteins resolved code proteins containing both the epitope recognized by RNL-2 and RNL-3 suggests a possible relationship between these proteins. Molecular ...
Fig. 4. A, nucleotide sequence and predicted amino acid sequence of a composite cDNA corresponding to the \( \text{NSP} \) transcript of 2.3 kb. The sequence is a composition of sequence data of pAB31 (from nucleotide 1 to 1889), pAB15 (from nucleotide 86 to 1795), and pAB37 (from nucleotide 1 to 1869). Sequence data of pAB49 (from nucleotide 1 to 2013) and pAB37 (from nucleotide 1 to 1795) is represented by arrows. The numbering between parentheses refers to the numbering of identical residues in the sequence of Fig. 3. The unique sequence at the 5'-end of the 1.8-kb transcript, pAB50 and pAB51, extend from nucleotide 147 to 1416 and from nucleotide 240 to 1223 (3021 in Fig. 3), respectively.

**Discussion**

Numbering of the amino acid residues of the deduced protein is indicated from the start codon ATG up to the stop codon TAA by an open box. In the composite cDNA corresponding to the 2.3-kb transcript, the open reading frame starts from the very 5'-end of the cDNA, but downstream of the unique 5'-end sequences (represented as a black box), the first ATG is found in the sequences shared with the 3.4-kb transcript. In the cDNA corresponding to the 1.8-kb transcript, the open reading frame starts also from the very 5'-end, and, in this case, the first ATG is found within the unique 5'-end sequences (represented as a shaded box).

As a Novel Neuroendocrine-specific Gene

using monoclonal antibody RNL-2 and monoclonal antibody RNL-3. Such screening resulted in the identification of 10 positive plaques, all showing immunoreactivity with both RNL-2 and RNL-3. After plaque purification, the cDNA inserts were subcloned into pUC18 and five of the resulting clones, designated and numbered pAB14 to pAB18, were used for further analysis. The insert sizes ranged from 1.7 to 1.8 kbp. Nucleotide sequence analysis revealed that three of these clones were identical, probably due to the fact that an amplified library was screened. Furthermore, the nucleotide sequence data suggested that all five cDNAs corresponded to the same as yet unknown gene, which was confirmed later (see below).

Northern blot analysis of NCI-H82 RNA with the largest cDNA (1.8-kbp insert of pAB15) as molecular probe revealed very high expression levels of transcripts exhibiting size heterogeneity in the range of 2.1–2.3 kb (hereafter referred to as 2.3-kb transcript), and lower levels of a transcript of about 3.4 kb (Fig. 2, lane 1). The highly expressed transcripts have not been found in any of the other SCLC cell lines tested (Fig. 2, lanes 2–5). Since these transcripts have only been detected in NCI-H82 cells to date, they are regarded as being unique for this cell line. Similar Northern blot analysis of total RNA isolated from a primary carcinoid tumor of the lung and various SCLC cell lines revealed expression of transcripts of about 3.4 kb and about 1.8 kb, the latter exhibiting size heterogeneity in the range of 1.6–1.8 kb (hereafter referred to as 1.8-kb transcript) (Fig. 2). Furthermore, expression levels appeared to vary (Fig. 2); expression in SCLC-22H (Fig. 2, lane 3) is beyond detection under the experimental conditions used (Fig. 2, lane 2); in SCLC-21H, both 3.4- and 1.8-kb transcripts are expressed (Fig. 2, lane 3); in GLC-1-M13, low levels of the 1.8-kb transcript (Fig. 2, lane 4); in GLC-1, only the 1.8-kb transcript seems to be expressed (Fig. 2, lane 5); and
Finally, in a primary carcinoid tumor of the lung, only the 3.4-kb transcript was detectable (Fig. 2, lane 6). Northern blot analysis with a 5'-end-PstI and a PstI-3'-end fragment of the human tumor cell lines resulted in the isolation of larger cDNAs belonging to the 2.3- and 2.1-kb transcripts. Because the NCI-H82 cDNA library could be used to isolate larger cDNAs corresponding to the 3.4- and 1.8-kb transcripts, respectively. The NCI-H82 cDNA library was subjected to agarose gel electrophoresis. Lane 1, neuroblastoma cell line CCI-127; lane 2, larynx carcinoma cell line CCI-23; lane 3, oral carcinoma cell line CCI-17; lane 4, neuroblastoma cell line HTB11; lane 5, pharynx carcinoma cell line CCI-138. Upper panel shows hybridization to the cDNA insert of pAB124 (30-h exposure). Lower panel shows hybridization to a hamster β-actin-specific probe (16-h exposure). Molecular weight markers are the same as in Fig. 2.

FIG. 6. Northern blot analysis of NSP gene expression in human tumor cell lines. Total RNA (15 µg) was subjected to agarose gel electrophoresis. Lane 1, neuroblastoma cell line CCI-127; lane 2, larynx carcinoma cell line CCI-23; lane 3, oral carcinoma cell line CCI-17; lane 4, neuroblastoma cell line HTB11; lane 5, pharynx carcinoma cell line CCI-138. Upper panel shows hybridization to the cDNA insert of pAB124 (30-h exposure). Lower panel shows hybridization to a hamster β-actin-specific probe (16-h exposure). Molecular weight markers are the same as in Fig. 2.

Cloning and Nucleotide Sequence Analysis of cDNAs Corresponding to the 3.4-, 2.3-, and 1.8-kb NSP Transcripts Reveal Extensive Overlap—Screening of the cDNA library of the carcinoid tumor of the lung with the insert of pAB15 as molecular probe resulted in the isolation of several overlapping cDNA clones. These clones are assumed to correspond to the 3.4-kb NSP transcript, since this tumor expressed only the 3.4-kb transcript (Fig. 2, lane 6). Only the three largest overlapping cDNA clones, C1, C14, and C20, were studied in detail; the 2.2-kbp insert of C1 and the 3.0-kbp insert of C14 were subcloned in pUC18 resulting in pAB44 and pAB124, respectively. The nucleotide sequences of the overlapping parts of the various cDNAs were identical. The cDNA sequence and the deduced amino acid sequence of the large open reading frame are shown in Fig. 3. It should be noted that the cDNA insert of pAB44, which contains the 3'-portion of the composite cDNA, did not contain a poly(dA) tail. It ended 1 bp after a putative polyadenylation signal. However, the cDNA insert of pAB49, which corresponds to the 1.8-kb NSP transcript and has a very extensive overlap with the insert of pAB44 (shown below), did contain a poly(dA) tail. Therefore, it is reasonable to deduce the sequences of the 3.4-kb transcript downstream of the polyadenylation signal from this cDNA. Altogether, the nucleotide sequence of the composite cDNA consists of 3214 bp. Taking a poly(A) tail into account, this might account for almost the complete 3.4-kb NSP transcript.

Rescreening of the NCI-H82 cDNA library with pAB15 resulted in the isolation of two cDNAs with larger inserts than the one of pAB15. These are most likely derived from the 2.3-kb NSP transcript (see data discussed below). Subcloning of these into pUC18 resulted in pAB31 and pAB37. Compared to pAB15, the insert of pAB31 contained an extension to the 5'-end, whereas the insert of pAB37 contained an extension to both the 5'- and the 3'-end. The nucleotide sequence of a composite cDNA of 2122 bp is shown in Fig. 4A. Comparison of this sequence to that of the composite cDNA corresponding to the 3.4-kb NSP transcript reveals that the sequences are identical except for 41 nucleotides at the 5'-end of the composite cDNA. We assumed that the cDNA inserts of pAB15, pAB31, and pAB37 are derived from the 2.3-kb NSP transcript, since there is an excess of this transcript in the NCI-H82 cells as compared to the 3.4-kb

FIG. 7. Northern blot analysis of NSP gene expression in mouse tissues. Total RNA (15 µg) was subjected to agarose gel electrophoresis. Lane 1, liver; lane 2, spleen; lane 3, heart; lane 4, kidney; lane 5, brain; lane 6, lung. Upper panel shows hybridization to the cDNA insert of pAB124 (24-h exposure). Lower panel shows hybridization to a hamster β-actin-specific probe (16-h exposure). Molecular weight markers are the same as in Fig. 2.
cDNAs of a Novel Neuroendocrine-specific Gene

The nucleotide sequence of the 2.3-kb transcript was not determined experimentally. However, the 3'-end of the three overlapping cDNAs, which were subcloned in pUC18, consisted of 2122 bp. Together with a poly(A) tail, this almost identical to a portion of the sequence (from nucleotide 14 (Fig. 4B) to nucleotide 2225 (Fig. 3)) of the 1.8-kb NSP transcript. The fact that the two independently isolated cDNAs both contain the same genetic sequences not represented in the 3.4-kb NSP transcript makes it highly unlikely that they are present in these cDNAs as a result of an artifact. The unique sequences in the 3.4- and 1.8-kb transcripts make it possible to obtain insight in future studies in the individual expression patterns of the transcripts by in situ hybridization analysis. This is of interest with respect to potential differences in the physiological roles of the two proteins assumed to be encoded by these transcripts.

The results described above clearly reveal the extensive overlap and close relationship between the various transcripts. It was of interest to find out whether the product of the cDNA corresponding to the 3.4-kb NSP transcript could be detected by RNL-2 and/or RNL-3. This appeared indeed to be the case as was demonstrated in Western blot analysis (data not shown) of a hybrid protein expressed in bacteria; this hybrid protein was encoded by a fusion gene which contained most of the coding sequences of the 3.4-kb NSP transcript fused in phase to β-galactosidase encoding sequences (see “Materials and Methods”). Neither RNL-2 nor RNL-3 showed any reactivity toward a hybrid protein containing sequences encoded by the cDNA corresponding to the 1.8-kb NSP transcript, which indicates that it does not contain the relevant epitopes.

Northern Blot Analysis Indicates That the NSP Gene Has a (Neuro)endocrine Expression Pattern—In previous immunohistochemical studies (6), expression of the antigen detected by RNL-2 and RNL-3 seemed restricted to neural and neuroendocrine tissues. To test whether expression of the NSP genetic sequences isolated with these antibodies displayed the same pattern, Northern blot analysis was performed on RNA samples of lung tumor cells. As shown in Fig. 2, four out of five neuroendocrine SCLC cell lines and a primary carcinoid lung tumor showed NSP expression. Northern blot analysis revealed expression of the 1.8-kb transcript as a weak expression of the 3.4-kb transcript in two human neuroblastoma cell lines, CCL-127 and HTB11 (Fig. 6, lanes 1 and 4), whereas in three non-neuroendocrine cell lines, i.e. the larynx carcinoma cell line CCI-23 (Fig. 6, lane 2), the oral carcinoma cell line CCI-17 (Fig. 6, lane 3), and the pharynx carcinoma cell line CCI-138 (Fig. 6, lane 5), no NSP expression was found. Similar analysis of various mouse tissue specimens showed expression in brain of two NSP transcripts of size similar to the human transcripts (Fig. 7, lane 5). Mouse liver, spleen, heart, kidney, and lung were negative (Fig. 7, lanes 1–4 and 6). Analysis of endocrine tissues revealed NSP expression in the neurointermediate lobe and anterior lobe of rat pituitary; transcripts of 3.4 and 1.8 kb were detected (Fig. 8, lanes 1 and 2). In human parathyroid, expression of only the 3.4-kb transcript could be detected (Fig. 8, lane 3); in human thyroid, expression of the 3.4-kb transcript. This was likely reflected in the relative amounts of corresponding cDNAs synthesized. Furthermore, the 5'-end of pAB31 contains 41 nucleotides that are not represented in the 3.4-kb NSP transcript (see also Fig. 5). However, the nucleotide sequence of the composite cDNA consists of 2122 bp. Together with a poly(A) tail, this almost identical to a portion of the sequence (from nucleotide 14 (Fig. 4B) to nucleotide 2225 (Fig. 3)) of the 1.8-kb NSP transcript. Finally, it should be noted that none of the cDNAs from the NCI-H82 library contained a poly(A) tail; therefore, the polyadenylation site of the 2.3-kb transcript was not determined experimentally. However, the 3'-end of the three overlapping cDNAs ends 2 bp before the putative polyadenylation signal mentioned. Furthermore, the observed size heterogeneity of the 2.3-kb transcript in NCI-H82 cells is probably due to alternative polyadenylation (see below).

To isolate cDNAs derived from the 1.8-kb NSP transcript, the GLC-1 cDNA library was screened since GLC-1 expresses only the 1.8-kb transcript. Screening resulted in the isolation of three overlapping cDNAs, which were subcloned in PUC18. The resulting clones were designated pAB49, pAB50, and pAB51, respectively. The nucleotide sequence of the largest cDNA (insert of pAB49) and the deduced amino acid sequence are shown in Fig. 4B. Comparison of this sequence to those corresponding to the 3.4- and 2.3-kb NSP transcripts revealed that the nucleotide sequence is identical to the former two except for its first 88 bp (see also Fig. 5). All three cDNAs corresponding to the 1.8-kb transcript contained a poly(A) tail. However, two distinct polyadenylation sites, about 200 bp apart (3021 and 3214 in Fig. 3), were apparently used. Both polyadenylation sites are preceded by at least one polyadenylation signal. The use of different polyadenylation signals (AATAAA and ATTAAA) (see Fig. 3) could explain the size heterogeneity of the 2.3- and 1.8-kb transcripts described above. The largest cDNA (pAB49) counted 1416 bp. Taking a poly(A) tail into account, this implies that probably a small 5'-end part of the 1.8-kb transcript is unaccounted for.

Fig. 8. Northern blot analysis of NSP gene expression in endocrine organs. Total RNA (15 μg) was subjected to agarose gel electrophoresis. Lane 1, rat neurointermediate lobe of the pituitary; lane 2, rat anterior lobe of the pituitary; lane 3, human parathyroid; lane 4, human thyroid; lane 5, mouse testis; lane 6, mouse ovary. Upper panel shows hybridization to the NSP cDNA insert of pAB124 (48-h exposure). Lower panel shows hybridization to a hamster β-actin-specific probe (16-h exposure). Molecular weight markers are the same as in Fig. 2.
cDNAs of a Novel Neuroendocrine-specific Gene

A, hydropathy analysis of the deduced protein encoded by the 3.4-kb NSP transcript (window: 7). B, distribution of charged amino acids in the deduced protein encoded by the 3.4-kb NSP transcript (window: 10). Numbering of the amino acids is given on the abscissa.

Fig. 10. Expression of the NSP cDNA corresponding to the 3.4-kb transcript in transfected COS-1 cells. [35S]Methionine-labeled lysates of NCI-H82 cells (lane 1), COS-1 cells transfected with pSVL-NSP3.4 DNA (lane 2), and untransfected COS-1 cells (lane 3) were subjected to immunoprecipitation and SDS-PAGE analysis as described under "Materials and Methods." A mixture of monoclonal antibodies MON-160, MON-161, and MON-162 was used. 35S-Labeled M, markers are indicated.

transcript was barely detectable (Fig. 8, lane 4). In mouse testis, NSP expression was found; however, the detected transcript appeared to be only 3.0 kb (Fig. 8, lane 5), the implications of which remain to be established. In mouse ovary, only the 3.4-kb transcript was found, but expression levels appeared to be very low (Fig. 8, lane 6). Finally, both the 3.4- and 1.8-kb NSP transcripts were found in pituitary adenoma cell lines AtT-20 and GH3, in medullary thyroid carcinoma cell line 6-23, and, to a lesser extent, in pheochromocytoma cell line PC-12 (data not shown). Altogether, we conclude that the NSP (neuro)endocrine expression pattern found in the Northern blot studies are concordant with the results of the immunohistochemical studies with the RNL-2 and RNL-3 monoclonal antibodies. RNL-2 and RNL-3 do not, however, cross-react with mouse and rat tissues.

The Deduced NSP Proteins Possess Common and Unique Features—On basis of the large open reading frame of the 3.4-kb NSP transcript, the amino acid sequence of a protein of 776 amino acid residues (NSP gene product A, calculated molecular mass of 84 kDa) was deduced (Fig. 3). The first ATG codon is found at position 123, and it is preceded by a stop codon in frame at position 21 (Fig. 3). The first ATG codon is found at position 123, and it is preceded by a stop codon in frame at position 21 (Fig. 3). The TAA stop codon is found at position 2451. Comparative analysis of the primary amino acid sequence of this deduced protein with known sequences (PIR release 33 and Swiss-Prot release 23) did not reveal any obvious sequence similarity to known amino acid sequences. However, some characteristic features could be defined. The protein does not have a signal peptide at its amino terminus but two large hydrophobic segments could be identified by hydropathy analysis in its carboxyl-terminal region (Fig. 9). SOAP and RAOARGOS analysis (23, 24) of these hydrophobic regions revealed that these could encompass both membrane-spanning domains. The region amino-terminal to these hydrophobic segments was found to be relatively rich in proline and serine residues. This region also contained a relatively high number of negatively charged residues (Fig. 9), leading to a calculated pI of 4.35 for the deduced protein. Finally, the deduced protein is predicted to contain multiple consensus sites for phosphorylation by various protein kinases, such as protein kinase C, casein kinase II, and tyrosine kinases.

The calculated molecular mass of the protein deduced from sequences of the 3.4-kb transcript differed significantly from
the observed electrophoretic mobility on SDS-PAGE. It is possible that the particular amino acid composition as described above is responsible for this. To confirm that the protein encoded by the 3.4-kb transcript runs indeed as a 135-kDa protein in SDS-PAGE analysis, an NSP cDNA containing the complete open reading frame of this transcript was expressed in COS-1 cells under control of the SV40 late promoter and the NSP-encoded protein was studied by immunoprecipitation and SDS-PAGE analysis; a mixture of monoclonal antibodies MON-160, MON-161, and MON-162 was used. As can be seen in Fig. 10, the protein expressed in the transfected COS-1 cells is 135 kDa (Fig. 10, lane 2) and has the same electrophoretic mobility as the protein immunoprecipitated from NCI-H82 cells (Fig. 10, lane 1). In non-transfected cells, the 135-kDa protein could not be detected (Fig. 10, lane 3). These results clearly establish the relationship between the 3.4-kb NSP transcript and this 135-kDa protein.

The deduced coding sequences of the 2.3- and 1.8-kb transcripts are present completely in the 3.4-kb transcript, except for a small stretch of 5′-coding sequences of the 1.8-kb transcript. The functional relevance of the possibly aberrant 2.3-kb NSP transcript remains to be established. From available cDNA sequence data, a large open reading frame starting at the 5′-end of the composite cDNA (Fig. 4A) is present. After the first 41 nucleotides, which are assumed to be unique for this transcript, the open reading frame is identical to the one of the 3.4-kb transcript (see Figs. 3, 4A, and 5). The first known ATG codon is found at position 291 in the part that is shared by the 3.4- and the 2.3-kb transcript. It is possible, however, that the true initiation codon remains to be found in still unknown sequences further upstream in the NSP gene. If translation starts at the ATG codon at position 291, which makes the product identical to a carboxy-terminal portion of the protein encoded by the 5.4-kb transcript, a protein of 356 amino acids (NSP gene product B) can be deduced with a calculated molecular mass of about 39 kDa. In this molecular mass range, four proteins (35–45 kDa) were found in NCI-H82 cells. It is not unreasonable to assume that these four proteins are post-translationally modified products of the deduced 39-kDa NSP protein.

As far as the open reading frame of the 1.8-kb NSP transcript is concerned, the deduced protein (NSP gene product C) is predicted to consist of at least 20 amino-terminal residues that are unique for the 1.8-kb transcript and 188 residues also present in the 3.4-kb transcript (see Figs. 3, 4B, and 5). We cannot exclude the possibility that the unique amino-terminal portion is even somewhat larger, since an upstream in frame ATG may be present in the presumed missing 5′-sequences of the 1.8-kb transcript.

Finally, although the postulated NSP proteins do not exhibit any apparent sequence similarity to known protein sequences, they possess structural features reminiscent of known neuroendocrine proteins, for instance the synaptic vesicle-associated protein amphiphysin (25). Like amphiphysin (predicted pI 4.39), the NSP proteins are very acidic (predicted pI of 4.35 for the protein deduced from the 3.4-kb transcript), are rich in proline residues, contain potential sites for phosphorylation by proline-directed protein kinases (see the many serine or threonine residues followed by a proline), protein kinase C (serines or threonines with a lysine or arginine residue in position +2), and casein kinase II (serines or threonines with an acidic residue in position +3 plus at least one other acidic residue between positions +1 and +5), and possess a hydrophobic stretch (potential membrane anchor) in the carboxy-terminal half. Both proteins lack a signal peptide. Abundant acidic amino acid stretches are also present in components of large dense-core neurosecretory vesicles, such as in chromogranin A and B (26). The neuroendocrine expression pattern of the NSP gene, together with the fact that the deduced NSP proteins have structural features similar to those of known neuroendocrine proteins, favor the possibility that the NSP proteins have a (neuro)endocrine-specific function. An important question that remains to be resolved in this context pertains to the expected functional differences between the highly related products of the 3.4- and the 1.8-kb NSP transcripts.

Acknowledgments—We thank M. Willems and M. Kranenberg for culturing the cell lines, I. Pauli for excellent technical assistance, G. Doucert for synthesizing primers, and P. Kools for the art work.

REFERENCES