Evidence for a Menkes-like protein with a nuclear targeting sequence

Manchi C. M. REDDY, Sudeep MAJUMDAR and Edward D. HARRIS¹

Department of Biochemistry and Biophysics and the Faculty of Nutrition, Texas A&M University, 2128 TAMUS, College Station, TX 77843-2128, U.S.A.

Extracts from three human cell lines were found to contain abridged Menkes disease gene transcripts with novel insertion sequences. The transcript variant that is the focus of the present study codes for a 103-residue protein containing the first heavy-metal-binding domain (Hmb1) of ATP7A, the Cu-ATPase associated with Menkes disease. This transcript variant has a 45-bp nucleotide insert interposed between exons 1 and 2 of ATP7A that starts with a 5' ATG that is in-frame with the downstream ATG translation start site of ATP7A. We report here that the 66-bp nucleotides positioned between the upstream and downstream ATG sites encode 22 amino acid residues whose primary structure in part meets the criteria for a nuclear-localization sequence (NLS). We have referred to the transcript as nuclear Menkes-like

(NML) 45. A green fluorescent protein (GFP) construct with NML45 when transfected in Chinese hamster ovary cells localized to the cell nucleus. A similar construct without the 66-bp segment exhibited a random dispersed fluorescent pattern in the cytosol. GFP constructs encoding ATP7A exons likewise failed to direct GFP into the cell nucleus, suggesting the nuclear determinant is not in an internal domain of the protein. The data suggest that the 22-residue segment contains an NLS for an 11.2-kDa protein with one Cu-binding site that may function as a chaperone to transport Cu into the nucleus of mammalian cells.

Key words: alternative splicing, ATP7A mRNA, copper transport, Menkes disease gene, nuclear-localization sequence.

INTRODUCTION

Menkes disease is an X-linked, often fatal disorder of Cu metabolism that primarily afflicts male infants. The disease results from a failure to transport Cu across the intestinal and blood–brain barriers, which leads to arrested growth, mental retardation and overt symptoms of severe Cu deficiency [1]. The gene for Menkes disease (*MNK*) encodes a membrane-bound Cu-transporting ATPase (ATP7A) that preserves cellular Cu homoeostasis by expelling Cu ions from a cell's interior [2]. ATP7A mRNA identified by Northern-blot analysis is 8.5 kb and is comprised of 23 exons [3,4]. Recent data from other laboratories and ours have shown that ATP7A mRNA appears in different spliced forms, depending on cell type [5,6]. The function of alternative transcripts is unknown, but their abundant presence in normal cells disconnects them from the pathogenesis of Menkes disease.

To investigate their function, we have isolated and sequenced alternative transcripts from poly(A)⁺-containing mRNAs of human cell lines and human fibroblasts. The PCR products generated from mRNAs from five separate cell lines have remarkably reproducible consistency in their electrophoretic patterns. This increases the likelihood that the PCR products were spawned from alternatively spliced mRNAs as opposed to random events during amplification. Previously, we identified a transcript in BeWo cells and human fibroblasts that had a novel 45-bp insert joining exons 1 and 2 [7]. The 3' end of the 45 bp joined exon 2 at a position 21 bp upstream of the published ATG start codon of ATP7A [8]. Further, the 45 bp contained an additional ATG at the extreme 5' end that was in-frame with this downstream start-site ATG of ATP7A [8]. Assuming the upstream ATG was an alternative translation start site, the 45-bp

insert extends the coding region by 66 bp and designates a 22-amino-acid-residue extension that would appear at the N-terminus of the ATP7A protein. Sequencing the cDNA containing the 45-bp insert, however, revealed that the 66-bp extension was within the reading frame of an mRNA that coded for a 103-residue protein and was not part of the full-length ATP7A. Here, we confirm that mRNA from SY5Y neuroblastoma cells gives rise to a transcript with a 45-bp insert and the potential to code for a 103-residue, 11.2-kDa protein that binds copper. We have determined further that residues within the peptide coded by the 66 bp meet the structural and functional criteria of a nuclear-localization sequence (NLS).

EXPERIMENTAL

Cells

SH-SY5Y human neuroblastoma cells were obtained from the A.T.C.C. (CRL-2266) and grown to confluence in a 1:1 mixture of minimum essential medium with non-essential amino acids, 90% Ham's F12 medium and 10% fetal bovine serum. Chinese hamster ovary CHO-K1 cells (A.T.C.C. CC1-61) were grown in F-12K medium (Sigma) with 2 mM L-glutamine, 1.5 g/l NaHCO₃ and 10% fetal bovine serum (Summit Biotechnology) with 1% Fungizone/streptomycin/penicillin (BioWhittaker) to approx. 80% confluence.

Reverse transcriptase PCR (RT-PCR)

RNA was extracted from confluent SY5Y neuroblastoma cells grown in plastic flasks (RNeasy, Qiagen). Total RNA (5 µg) was

Abbreviations used: ATP7A, Menkes Cu-transporting ATPase; GFP, green fluorescent protein; *MNK*, Menkes disease gene; NLS, nuclear-localization sequence; NML, nuclear Menkes-like; ORF, open reading frame; RT-PCR, reverse transcriptase PCR.

¹ To whom correspondence should be addressed (e-mail eharris@tamu.edu).

Figure 1 Diagram of pEGFP-N3 multiple cloning site (MCS) showing *NheI* restriction sites

Numbers indicate position sites on the vector. The lower panel shows the end nucleotide sequences that were added to accommodate *NheI* restriction sites on intact NML45 and NML45 lacking the 66-bp NLS.

reverse-transcribed (Gibco-BRL Superscript Preamplification System with Moloney-murine-leukaemia virus) using 2 h for the longer mRNA species. PCR used the eLONGase enzyme (Life Technologies) for amplification of longer fragments and Promega *Taq* DNA polymerase amplification system (Promega, Madison, WI, U.S.A.) for shorter DNA templates.

Primers were designed as follows: forward primer, 27-mer (13–39) 5'-GCTACTGTGACTTCTCCGATTGTGTGA-3' covering the exon 1 region, and reverse primer, 27-mer (4794–4768) 5'-GCCAACAGATAAACGCAACTCAAATAG-3' covering exon 23. The cDNA encompassing the whole open reading frame (ORF) was predicted to be 4.7 kb. Amplification of cDNAs was carried out using 2 μ l of reverse-transcribed products in the presence of 1.0 μ M MNK-specific primers, 2 units of eLONGase, and the provided buffer containing 1.8 mM Mg^{2+} . All PCR operations were performed in a Perkin-Elmer Cetus Geneamp PCR System 2400 at 94 °C with the programmed specifications as follows: 1 min of amplification/denaturation, 25–30 cycles at 94 °C for 15 s, 58–65 °C for 30 s, 68 °C for 45–60 s/kb of fragment. Expand Long Template *Taq* (2 units; Boehringer Mannheim) was added followed by a 12-min incubation time at 72 °C to allow TA cloning of the longer templates generated by the eLONGase. The amplified product was applied to 0.8% agarose gels in 0.045 M Tris/0.045 M borate/0.001 M EDTA buffer, pH 8.0, using λ -phage genomic DNA marker cut with *EcoRI* and *HindIII* (Promega) to determine the position and size of fragments.

RNase protection assay

The existence of an mRNA corresponding to the RT-PCR product called nuclear Menkes-like (NML) 45 was validated by an RNase protection assay using the RPA111 kit (Ambion). *In vitro* synthesis of RNA transcripts from DNA templates attached to a T7 promoter was utilized to prepare the probes, employing a MAXIScript In Vitro Transcription Kit (Ambion). Two anti-sense RNA probes (599 and 495 bp) to NML45 were generated from 1 μ g of each purified DNA template. The two DNA templates were obtained by PCR amplification using two sense

primers covering positions 10–29 and 114–133 of NML45 as follows: 5'-ACTTCTCCGATTGTGTGAGC-3' and 5'-TGAG-GAAACTGAGCATCAGA-3', respectively. The common anti-sense primer, 5'-CGCGCGTAATACGACTCACTATAGCAG-ACAGTAGTGAGTTTATAGAGGGC-3', corresponded to positions 609–584. The primer was designed to incorporate a T7 promoter site (underlined) into the sequence. Mouse β -actin (304 bp) was used as an anti-sense control probe. Radiolabelled RNA was prepared by adding [α - 32 P]UTP (800 Ci/mmol) to a final concentration of 5.25 μ M for each probe and 3.12 μ M for the mouse β -actin control. The mixture was incubated at 37 °C for 90 min, 10 units of DNase were added and incubation was continued for 20 min at 37 °C. SY5Y RNA and yeast RNA (control) (20 μ g each) were mixed with the radioactive probes. Probes for 599 and 495 bp were present at a concentration of 1.8 ng/20 μ g of RNA; β -actin had 450 pg/20 μ g of RNA. After an overnight incubation at 42 °C to effect hybridization, each sample was digested with RNaseA/RNaseT1 (0.37 units/15 units/ μ l), in 150 μ l of buffer according to Ambion's protocol. RNase inactivation was for 1 h. Electrophoresis was performed on a 5% acrylamide/8 M urea denaturing gel for 2 h at 17 V/cm and, after removing surface moisture, the gel was exposed to X-ray film (Kodak, Biomax-MS) for 30 min at –80 °C.

Cloning of SY5Y mRNA variants

RT-PCR-generated fragments were cloned into a TOPO-XL autoligating cloning vector (TA Cloning Kit, Invitrogen). The reaction used 2 μ l (60 ng) of TOPO-XL, 5 μ l of purified PCR product (approx. 120 ng) and sterile water to 10 μ l. The ligation reaction was performed at room temperature for 5 min after which 2 μ l of the reaction solution was added directly to 2-ml plastic screw-cap vials containing 50 μ l of TOP 10 cells, stirred gently and placed on ice for 30 min. The reaction mixture was then heat-shocked for 30 s in a 42 °C water bath and placed immediately on ice for 2 min. To each vial was added 250 μ l of SOC medium (2% tryptone/0.05% yeast extract/0.05% NaCl/2.5 mM KCl/10 mM $MgCl_2$ /20 mM glucose) and the vials were incubated at 37 °C for 30 min on a rotary-shaking incubator. From each vial, 50 μ l were spread on Luria-Bertani agar plates containing 50 μ g/ml kanamycin. Colonies were randomly selected and screened by PCR. Insert orientation was determined by extracting the purified plasmid DNA from the positive clones and digesting for 90 min with 2 units of *EcoRI*/*Bam*HI per 2 μ g of plasmid containing the insert. After screening, properly oriented positive clones were sequenced using plasmid-specific primers.

DNA sequence determination

RT-PCR products separated on agarose gels were extracted (QIAEX Gel Extraction kit, Qiagen) and cloned into a general cloning vector. Cycle-fluorescence sequencing of the cloned DNA used the ABI Dye Terminator Kit containing AmpliTaq DNA polymerase with primers designed for the individual fragments. Sephadex G-50 (Pharmacia) was used to remove unbound nucleotides. The purified products were suspended in loading buffer and applied to an Applied Biosystems 377 DNA sequencer.

Cloning the 45-bp cDNA

Cloning the cDNA generated from an mRNA template followed the procedure described previously [6]. Primers were selected on the basis of a T_m that closely matched the T_m of the oligonucleotide 5'-TGAGGAACTGAGCATCAGAAA-3', a 22 mer that

complemented the 45-bp sequence. An exon-1 forward primer (5'-ACTTCTCCGATTGTGTGAGC-3') covered positions 22–41. Primers to both exon 1 and the 45-bp sequence were paired with a 23-mer reverse primer (5'-TCACATTAAGGTATCAC-TGGAAG-3') that covered exon-23 positions 4901–4882. The PCR fragments generated were cloned into a pCR3.1 vector.

A ligation reaction was prepared containing 2 µl (60 ng) of vector, 1 µl of 10× ligation buffer, 1 µl of T4 DNA ligase, 5 µl of purified PCR product (about 120 ng) and sterile water to 10 µl. The ligation reactions were incubated at 14 °C overnight and aliquots of the ligation mixture were added directly to the TOP 10 cells, propagated and screened as described above.

Green-fluorescent-protein (GFP) constructs

GFP constructs were made using the pEGFP-N3 vector (Clontech). cDNA inserts containing the ORF of each transcript were isolated from cloned pCR3.1 plasmid by PCR utilizing the following anchor primers: the upper primer 5'-ATAAGCTAG-CATGAGGAACTGAGCATCA-3' complementing bases specific to the 45-bp insert, and the reverse primer 5'-ATAAG-CTAGCCTGTAAAGTTAAAGGAAGA-3'. These primers were designed to incorporate *NheI* sites at the 5' and 3' ends with no intervening in-frame stop codons (Figure 1). A primer that excluded the 45 bp was designed using the ATP7A initiation codon with a *NheI* anchor: 5'-ATAGGCTAGCATCAAAAT-GGATCCAAGTAT-3'. Subsequently the PCR products were digested with *NheI* restriction enzyme and cloned into the *NheI*-cut eukaryotic expression vector pEGFP-N. The two GFP constructs were isolated and purified from *NheI* restriction fragments and ligated into the mammalian expression vector pEGFP-N, which contains a construct of GFP at the C-terminus of a multiple cloning site.

Cell culture, transfection and selection of stable cell lines

CHO-K1 cells were transfected with pEGFP-N vector using a Genetronics Electro Cell manipulator according to the instructions of the manufacturer. After the transfection the cells were cultured for 48 h in growth media supplemented with 10% fetal calf serum. The medium was renewed 24 h after transfection, and 48 h after transfection the medium was changed to one containing 400 µg/ml Geneticin (G-418, Life Technologies). Cells surviving G-418 treatment were examined for green fluorescence under a fluorescence microscope using a filter that allowed excitation at 488 nm and broad-band emission.

Fluorescence microscopy

CHO-K1-transfected cells selected against 400 µg/ml G-418 antibiotic were grown for 24–48 h on chambered glass slides (Lab-Tek Chambered Coverglass, Nunc) in CHO-K1 growth medium with 10% Fungizone/streptomycin/penicillin. The cells were examined under a Zeiss Axiovert 135 fluorescence microscope with a 100× objective and field size of 40–50 µm. Images were captured using a Photometrics Star-1 CCD camera.

RESULTS

Alternative ATP7A transcripts

Performing RT-PCR with poly(A)⁺ templates and primers designed to amplify the exon 1–23 region of ATP7A mRNA generated at least four cDNA bands whose electrophoretic patterns among three different human cell lines were basically identical (Figure 2, left-hand panel). Bands of 4.7, 1.8 and 0.8 kb

were prominent in amplified products from all the cells. The 0.8-kb band, however, was split into two components, designated 0.8a and 0.8b. SY5Y cells showed a second band that appeared below the 4.7-kb band (labelled 4.7b in Figure 2, left-hand panel). This band was not seen in fibroblasts or Caco-2 cDNAs. When the 4.7-kb cDNA was used as a template to repeat the amplification, only a single 4.7-kb band was generated (Figure 2, right-hand panel, lanes 3 and 4). This single band appeared when either plasmid-containing or free 4.7-kb cDNA was used as the template. These observations lessen the likelihood that bands below the 4.7-kb band were spawned from the 4.7-kb band.

Table 1 summarizes the composition data for cDNAs generated from SY5Y cells. Although only SY5Y data are reported, human fibroblasts and Caco-2 cells had base sequences that were practically indistinguishable from SY5Y. Each showed the cDNA labelled 4.7a to be a full-length ATP7A transcript, i.e. with all 23 exons intact and in sequence as originally reported [8]. This band appeared in all the cells. The band labelled 4.7b was expressed strongly in SY5Y and had earlier been seen in undifferentiated human placental (BeWo) cells (results not shown). It is absent from all other cells examined thus far. Complete sequence data revealed that the 4.7b cDNA resembles 4.7a but lacked exon 10. The band labelled 1.8 lacked exons 3–15 and had an in-frame link between exons 2 and 16 as reported earlier [6].

As noted by their lengths, the split bands labelled 0.8a and 0.8b were missing a sizable number of exons. Both, however, had novel inserts in their respective 5' untranslated regions. Band 0.8b had a 192-bp insert in the region between the non-coding exon 1 and exon 2, the first coding exon. The base and derived amino acid sequences of the insert have been reported previously [7]. Band 0.8a had a 45-bp insertion also interposed between exons 1 and 2. This insert, however, had an ATG at the extreme 5' end that was in-frame with the downstream ATG identified as the translation start site for ATP7A (Figure 3). The position of the ATG, therefore, was consistent with an alternative translation start codon. The band with the 192-bp insert had no upstream ATG codon.

Characterization of the 45-bp-containing transcript

Of the various cDNA fragments seen in Figure 2, only 0.8a contained a 45-bp insert. Our goal was to investigate its possible function. First, to verify the existence of 0.8a mRNA, we performed an RNase protection assay using two separate anti-sense probes that covered the sequences between exons 1 and 23. More specifically, one probe covered exon 1–45-bp–exon 2–exon 3–exon 22–exon 23. The second probe covered the same sequence order, but began at the 45-bp insert. The two probes were predicted to complement segments of 599 and 495 nucleotides, respectively. Figure 4 shows the results of the protection assay. Figure 4(A) shows the cDNA templates used to generate the 599 (lane 2) and 495 (lane 3) RNA probes. In Figure 4(B), lanes 1 and 2 represent the 599 and 495 anti-sense probe fragments, respectively, observed following RNase digestion in the presence of extracts of SY5Y RNA. The β -actin control probe (304 bp) was included with the two samples. Lanes 3, 4 and 5 show the digests of 599, 495 and β -actin probes using yeast RNA as a protectorate. Only SY5Y RNA protected each probe from nuclease digestion. Moreover, only one protected segment was seen in each lane and its size was close to the size predicted by the PCR data, suggesting the existence of an mRNA with the expected structure indicated by PCR analysis.

The 3' end of the 45-bp-containing transcript revealed an intact exon 2 and an exon 3–22 junction that was created by

Figure 2 Gel-electrophoresis pattern of amplified cDNAs from different human cell lines

(Left-hand panel) RT-PCR amplification. Lanes 1–3, 50 cycles; 4–6, 25 cycles. Lane 1, SY5Y neuroblastoma cells; lane 2, Caco-2 cells; lane 3, normal fibroblasts; lane 4, purified fragments of 0.8a and 0.8b bands; lanes 5 and 6, 0.8b and 0.8a fragments amplified (25 cycles) from clone. (Right-hand panel) Lane 1, TOPO-XL alone (3.7 kb); lane 2, TOPO-XL with 4.7-kb fragment; lane 3, 4.7-kb fragment amplified from plasmid clone; lane 4, 4.7-kb fragment amplified from an unbound 4.7-kb cDNA template. MW, *HindIII/EcoRI* markers.

Table 1 Summary of SY5Y cDNAs and potential protein products

Band	Exons present	Feature	Estimated molecular mass (kDa)
4.7a	1–23	Intact ATP7A	170
4.7b	1–9, 11–23*	Lacking exon 10	145
1.8	1–2, 16–23*	Lacking exons 3–15	55
0.8a	1–3, 22†	45-bp insert at 5' end	11.2
0.8b	Unknown	192-bp insert at 5' end	Unknown

* In-frame.

† Out-of-frame stop codon.

Figure 3 Position of the 45-bp insert in the 5' untranslated region of ATP7A

(A) The 45 bp in position between exons 1 and 2. (B) The 5' end of the 45-bp insert contains an ATG sequence that is in-frame with an ATG downstream (underlined), identified as the start site for ATP7A (arrow).

linking base 37 in exon 3 with base 12 of exon 22 (Figure 5). Examining the sequence of bases at the fusion of the two exons, it was not possible to determine those contributed by exon 3 from those of exon 22. The fusion, however, resulted in a one-base shift in the reading frame, which generated a TAA stop codon at nucleotide positions 85–87 in exon 22 (Figure 5). The 28

Figure 4 Determination of NML45 mRNA by RNase protection assay

(A) NML45 cDNA (lane 1) and cDNA templates for generating 599 (lane 2) and 495 (lane 3) anti-sense RNA probes. (B) SY5Y RNA (20 µg) with ³²P-labelled 599 (lane 1) or 495 (lane 2) nucleotide anti-sense RNA. A β-actin control probe (304 bp) was present in lanes 2 and 3. Lanes 3–5, 20 µg of yeast RNA added to probes of 599, 495 and β-actin, respectively. Lanes 6 and 7 have the β-actin probe in the presence of 10 µg of SY5Y RNA. M, Ambion's RNA Century™ Marker Template Plus, transcribed in the presence of [³²P]UTP.

Forward primer

1 GCTACTGTGACTTCTCCGATTGTGTGAGCTTTGTTGGAGCCTGCGTACGTGGATT-NML45
 13 GCTACTGTGACTTCTCCGATTGTGTGAGCTTTGTTGGAGCCTGCGTACGTGGATT-ATP7A

57 ATCGCTGCCACGGTCTGCGTAGCTCCAGAGGTTTAACCATAGGATAGAGAAACCAG-NML45
 68 ATCGCTGCCACGGTCTGCGTAGTTCAGAGGTTTAACCATAGGATAGAGAAACCAG-ATP7A

→NML45 start

113 **M R K L S I R K R P N N L L K**
 ATG AGG AAA CTG AGC ATC AGA AAG AGA CCA AAC AAT TTG CTG AAG-NML45
 125 *** *** *** *** *** *** *** *** *** *** *** *** *** ***-ATP7A

→ATP7A start

158 E C N E E I K M D P S M G V N
 GAA TGT AAT GAG GAA ATC AAA ATG GAT CCA AGT ATG GGT GTG AAT-NML45
 125 GAA TGT AAT GAG GAA ATC AAA ATG GAT CCA AGT ATG GGT GTG AAT-ATP7A

Metal Binding Motif

203 S V T I S* V E G M T C N S Q V
 TCT GTT ACC ATT TCT GTT GAG GGT ATG ACT TGC AAT TCC TGT GTT-NML45
 172 TCT GTT ACC ATT TCT GTT GAG GGT ATG ACT TGC AAT TCC TGT GTT-ATP7A

248 W T I E Q Q I G K V N G V H H
 TGG ACC ATT GAG CAG CAG ATT GGA AAA GTG AAT GGT GTG CAT CAC-NML45
 215 TGG ACC ATT GAG CAG CAG ATT GGA AAA GTG AAT GGT GTG CAT CAC-ATP7A

I K V S L E E K N A T I I Y A

293 ATT AAG GTA TCA CTG GAA GAA AAA AAT GCA ACT ATT ATT TAT GCC-NML45
 259 ATT AAG GTA TCA CTG GAA GAA AAA AAT GCA ACT ATT ATT TAT GAC-ATP7A

Exon 2 ← Exon 3

338 H W F G F A A L D G I C S N G
 CAT TGG TTT GGT TTT GCA GCC CTG GAT GGG ATC TGC AGC AAT GGC-NML45
 4283 I G L V L Q P W M G S A A M A
 ATT GGT TTG GTT TTG CAG CCC TGG ATG GGA TCT GCA GCA ATG GCT-ATP7A

→ Exon 22

383 C F I C F C S T F F S S L ***
 TGC TTC ATC TGT TTC TGT AGT ACT TTC TTC TCT CTT CCT TAA -NML45
 4327 A S S V S V V L S S L F L K
 GCT TCA TCT GTT TCT GTA GTA CTT TCT TCT CTC TTC CTT AAA-ATP7A

424 ACTTTACAGGAAACCAACTTACGAGAGTTATGAAGTGCCTGCCCGGAGCCAGATAGGACAG-NML45
 4368 ACTTTACAGGAAACCAACTTACGAGAGTTATGAAGTGCCTGCCCGGAGCCAGATAGGACAG-ATP7A

485 AAGAGTCCTTCAGAAATCAGCGTTCATGTTGGAATAGATGATACCTCAAGGAATTCTCCTA-NML45
 4432 AAGAGTCCTTCAGAAATCAGCGTTCATGTTGGAATAGATGATACCTCAAGGAATTCTCCTA-ATP7A

546 AACTGGGTTTGCTGGACCGGATTGTTAATTATAGCAGAGCCTCTATAAACTCACTACTGTC-NML45
 4491 AACTGGGTTTGCTGGACCGGATTGTTAATTATAGCAGAGCCTCTATAAACTCACTACTGTC-ATP7A

607 TGATAAACGCTCCCTAAACAGTGTGTTACCAGTGAACCTGACAAGCACTCACTCCTGGTG-NML45
 4552 TGATAAACGCTCCCTAAACAGTGTGTTACCAGTGAACCTGACAAGCACTCACTCCTGGTG-ATP7A

668 GGAGACTTCAGGGAAGATGATGACACTGCATTATAAAAGGCCATGGAGAGTGCTGCCAGTT-NML45
 4613 GGAGACTTCAGGGAAGATGATGACACTGCATTATAAAAGGCCATGGAGAGTGCTGCCAGTT-ATP7A

729 TAACCTTGTCATGCACTGACACAGCATTCATGATGTACCTTCACCTTTTCAAATATTGTAG-NML45
 4674 TAACCTTGTCATGCACTGACACAGCATTCATGATGTACCTTCACCTTTTCAAATATTGTAG-ATP7A

Reverse Primer

790 AAGGATTTTTCTCATGCTCTTATATTAGGGATTCTATTTGAGTTGCGTTTATCTGTTGGC-NML45

4735 AAGGATTTTTCTCATGCTCTTATATTAGGGATTCTATTTGAGTTGCGTTTATCTGTTGGC-ATP7A

Figure 5 Nucleotide sequence of NML45 protein from SY5Y neuroblastoma cells

Comparison with ATP7A. The sequence corresponding to the 45-bp insert is in bold. The box shows the residues that make up the first heavy-metal-binding site of ATP7A. *, Shows the position where a single bp substitution C → T (TTT in place of TCT) in BeWo cells would lead to phenylalanine in place of serine. NML45 fusion point between exon 3 and 22 gives rise to an out-of-frame sequence terminating with a TAA at positions 4365–4367 (***). The ATP7A start site and numbering of ATP7A is based on the original sequence data [8]. Primer sequences are underlined.

1 ATG AGG AAA CTG AGC ATC AGA AAG AGA CCA AAC AAT
 37 TTG CTG AAG GAA TGT AAT GAG GAA ATC AAA ATG GAT
 73 CCA AGT ATG GGT GTG AAT TCT GTT ACC ATT TCT GTT
 109 GAG GGT ATGA CT TGC AAT TCC TGT GTT TGG ACC ATT
 145 GAG CAG CAG ATT GGA AAA GTG AAT GGT GTG CAT CAC
 181 ATT AAG GTA TCA CTG GAA GAA AAA AAT GCA ACT ATT
 217 ATT TAT GCC CAT TGG TTT GGT TTT GCA GCC CTG GAT
 253 GGG ATC TGC AGC AAT GGC TGC TTC ATC TGT TTC TGT
 289 AGT ACT TTC TTC TCT CTT CCT

Figure 6 Base composition and sequence data of NML45

The N-terminal region encompassing the NLS is shown in bold. The bent arrow shows the consensus start site of ATP7A [8]. Boxes refer to C-terminal amino acid residues in the putative NML45 protein that deviate from the ATP7A protein sequence.

bases beyond the fusion point predictably could not give rise to Menkes residues even though they followed an ATP7A base order. The nucleotide sequences that were out-of-frame near the C-terminus would give rise to a cysteine-rich motif characterized by a single CXXXCXXCXC. As noted above, the 5' ATG of the 45-bp insert was exactly 66 bp upstream from the start site of ATP7A. Assuming this ATG was the translation start site, the protein encoded would span 103 amino acids and the bases between the upstream and downstream ATGs would constitute a coding segment that corresponded to a 22-amino acid sequence on the extreme N-terminus of the 103-residue protein. Based on the ORF, the protein has a predicted mass of 11.2 kDa. Because exon 2 is intact and in-frame, the protein would also contain the first heavy-metal-binding motif (Hmb1) of ATP7A.

Evidence for an NLS

Further analysis revealed that the first 22 residues on the N-terminus of the 103-residue protein, if decoded, would correspond to a polypeptide with the structure MRKLSIRKRPNNLLKECNEEIK (Figure 6). The RK...RKR in the sequence spans eight residues and is flanked by a proline residue. Using PSORT-sort analysis, one is able to predict with 47.8% probability that the eight-residue span meets the sequence criteria for an NLS [9]. These same analyses reported that a plasma membrane-, endoplasmic reticulum- or cytosolic-locating signal was far less likely. Based on the analysis, we have named the 103-residue protein NML45.

To determine if the 22 residues in the N-terminus of NML45 function as an NLS, constructs were prepared that contained the

coding sequences for the 103-residue protein attached via its C-terminus to GFP. Control constructs of GFP alone and NML45 lacking the 66 bp were also prepared. GFP constructs incorporating the 1.8-kb band and exons 2–10 were included in the study to determine if internal exons had an NLS property. All constructs were transfected into CHO cells and direct fluorescence analysis was performed on cells surviving G-418 treatment. Figure 7 shows confocal and fluorescent images of CHO-K1 cells transfected with GFP constructs. In Figure 7(b) there is a clear indication that the fluorescence of the NML45 lacking the 22-residue presequence was more diffuse and did not localize in any one area. Figure 7(d) shows that the fluorescence from intact NML45, i.e. with the 66 bp, gave a stronger indication of a centralized location in a region that confocal image analysis suggests is the nucleus of the cell. Figures 7(e) and 7(f) show a repeat of the above analysis using a different GFP vector and higher magnification. Again, cells transfected with NML45 minus the 22 residues showed fluorescence confined mostly to the cytosol. The fluorescent molecules seemed to outline, but not penetrate, the nucleus. This same nuclear-exclusion pattern was seen with GFP alone (results not shown). In sharp contrast, an intact NML45 construct showed a clear fluorescence within the cell nuclei and not at the periphery (arrows in Figures 7e and 7f). The nuclear boundary was not sharply profiled and there was no indication of aggregation at the nuclear membrane. Figure 7(e) further shows that the fluorescence did not penetrate into the nucleolus (dark circle). Constructs composed of exons 2 and 16–23 (band 1.8, Figure 2, left-hand panel) and of exons 1–10 exhibited a pattern very similar to NML45 lacking the 22 residues (results not shown). These data suggest that a signal for nuclear penetration lies within the 22 residues at the N-terminus and is not associated with other exon-coded regions of the protein. They further support NML45 as a nuclear-targeting protein.

DISCUSSION

It is yet to be determined whether small poly(A)⁺ RNAs with Menkes-like sequences have functional significance in Cu transport and homeostasis. These mRNA variants are not necessarily connected with disease mutations, but could arise during the normal processing of the mRNA for ATP7A. As seen in Figure 2 (left-hand panel), three such alternative transcripts may be common to all human cells. The transcript appearing as a 1.8-kb cDNA has been found in RT-PCR-amplified mRNA from HepG₂ cells and developmentally arrested BeWo cells. Since expression of a full-length ATP7A mRNA transcript is not seen in liver [10,11] and is seen in BeWo cells only when the cells are induced to differentiate [12], one may question whether the 1.8-kb fragment is synthesized as a means to regulate the expression of full-length ATP7A mRNA.

Skin fibroblasts and lymphoblasts from a male patient diagnosed with occipital horn syndrome, a mild form of Menkes disease, synthesize an ATP7A variant that lacks exon 10 [13]. ATP7A protein missing exon 10, which codes for two transmembrane domains, remains resident in the endoplasmic reticulum and does not enter the Golgi network [13,14]. Using exon-10-specific nucleotides, Dierick et al. [3] showed that the expression of transcripts lacking exon 10 is common in human lymphoblasts and fibroblasts, whereas the cerebrum or cortex give only slight evidence for the loss of this 234-bp exon. Here, we showed that a pure cell line, SY5Y, routinely expresses a transcript lacking the exon. That skin fibroblasts do not strongly

Figure 7 Confocal and fluorescence analysis of transfected CHO-K1 cells

CHO-K1 cells transfected with GFP constructs were selected against G-418 and identical fields were analysed by confocal microscopy and fluorescence imaging. Confocal (**a**) and fluorescence (**b**) images of cells transfected with GFP-containing vector alone. Confocal (**c**) and fluorescence (**d**) images of GFP-containing vector + NML45 insert. Fluorescence images of GFP vector + NML45 that omitted (**e**) or retained (**f**) the 66-bp extension. Arrows show nuclear boundaries. (**a–d**) GFP vector was pQBI63 from Quantum Biotechnology; (**e, f**) GFP vector was pEGFP-N3 from Clontech (see the Experimental section).

express an ATP7A variant lacking exon 10 agrees with Qi and Byers [13]. The evidence, therefore, suggests that some cells remove exon 10 during routine processing. Thus one must consider that the lesion that gives rise to occipital horn syndrome could be due to a failure to suppress the exon-10 splicing factor and not necessarily a mutation in the ORF of the gene. This issue needs to be resolved. A further generalization is to consider that cells that do not express a full-length ATP7A are likely to possess nuclear splicing factors capable of modifying ATP7A mRNA before it is transported to the cytosol. By comparing Caco-2 cells or normal fibroblasts with SY5Y cells and BeWo cells it seems clear that the expression of ATP7A transcript variants depends on the cell type or developmental stage. This is compelling evidence that modifications to ATP7A mRNA occur by post-transcriptional processes associated with cell-specific mechanisms.

A first inclination is to consider transcripts missing sizable numbers of exons or containing novel inserts as artifacts of the highly sensitive PCR procedure. Several lines of evidence speak against this conclusion. First, numerous human cell lines clearly show a limited and reproducible pattern of the cDNAs displayed; 4.7 kb, 1.8 kb and 0.8 (0.8a and 0.8b) kb being the most common. cDNA artifacts would have given a more varied pattern. Second, mRNA from SY5Y and BeWo cells gave cell-type-specific displays. Third, a 4.7a cDNA when used as a plasmid-bound template or as an unbound cDNA in a repeat of the amplification procedure generated only a single 4.7-kb cDNA and not smaller cDNA fragments (Figure 2, right-hand panel). Finally, RNase protection assay confirmed the presence of an RNA fragment with the sequence structure predicted by the PCR sequence data. These data show that a 0.8-kb cDNA is likely to be produced from a pre-existing mRNA template and is not an aberrant

Figure 8 Structural similarity between exon 3, exon 22 and NML45

The upper part of the figure shows G³⁷ in exon 3 bonding to C¹² of exon 22. The lower figure compares the sequence data of NML45 and exons 3 and 22. In the fusion event, the A at position 38 is omitted from NML45.

product derived from ATP7A mRNA during amplification. Moreover, the band labelled 0.8b contains a 192-bp insert that MNK genomic sequence analysis has located to be approx. 12 kb upstream from exon 2 in what is considered the intron 1 region of the gene (S. Majumdar and E. D. Harris, unpublished work). This sequence has been shown by others (T. Glover, personal communication) and by us [7] to appear in some ATP7A cDNAs. Neither the 192- nor the 45-bp inserts were reported in the original sequence analysis of ATP7A cDNA by Vulpe et al. [8]. With probes designed from sequence information for a full-size ATP7A mRNA, there is the possibility that smaller transcripts containing the 45- and 192-bp sequences were either not detected or were assigned to the 60-kb intron 1 region of the unprocessed transcript.

In this report we provide evidence that human cells synthesize an mRNA with a 45-bp insertion sequence that codes for an 11.2-kDa protein. The putative protein, NML45, is basically chimaerial, incorporating the first heavy-metal-binding site of ATP7A and featuring a cysteine-rich CXXXXCXXC motif suggestive of a second heavy-metal-binding site. Bases beyond the exon 3–22 junction provide a template for 29 non-Menkes residues that comprise the C-terminus of NML45. The 3–22 junction is not a splice site by the usual Chambon criteria [15]. Figure 8 shows that exons 3 and 22 share six identical bases, i.e. TTTATG. Sequence overlap could be a structural feature that led to the union of the G in exon 3 with the C in exon 22. Although this is non-conventional splicing, there have been instances reported in the literature where non-conventional sites come into play as a result of positive effectors in the nuclei that regulate non-conventional splicing [16]. The *tsg101* gene that confers tumour sensitivity in breast and uterine cancers is subject to non-conventional alternative splicing. Because of their low abundance, these alternative transcripts are generally not detected in Northern-blot analysis [17]. Truncated transcripts of the gene appear to occur due to 3–6 base sequence similarities at the deletion junction [18]. Moreover, the alternatively spliced transcript has been associated with isoforms that encode transcriptional regulators that are capable of interacting with transactivating factors which recognize nuclear receptors [19–21]. A key question to be decided is whether cells have transacting factors that access cryptic splicing sites, and whether these factors act on ATP7A mRNA to create an exon 3–22 union.

The N-terminus of NML45 has what appears to be a bipartite NLS, i.e. two runs of basic amino acids separated by a spacer

region and flanked by a proline that is positioned to disrupt helix formation [22]. Cytosolic proteins bearing a mono- or bipartite NLS recognize the α component of karyopherin, a heterodimeric receptor. A second component, karyopherin β , targets the import complex to the nuclear pore for movement into the nucleus. A protein that lacks an NLS is effectively barred from entering.

Because the NML45 transcript has been found in all cells examined in our laboratory and because NML45 protein generated *in vitro* appears to penetrate the nucleus, it is conceivable that NML45 mRNA codes for a mammalian nuclear copper-transport peptide. Alternatively, NML45 protein could regulate the expression of the *ATP7A* gene. The latter is based on structural similarities between the predicted NML45 protein and the transcription factor Mac1p of the yeast *Saccharomyces cerevisiae*. Both share a cysteine-rich C-terminal domain and both have nuclear targeting sequences [23]. Another parallel can be drawn between NML45 and copZ. Strausak and Solioz [24] have shown that the expression of copAB, two Cu-transporting ATPases in *Enterococcus hirae*, is regulated through a suppressor regulatory protein (copY). copZ reverses copY suppression by transporting Cu to copY and forcing its disassociation from the DNA. It is of interest that the 103-residue NML45 shares 45% sequence identity with the 69-residue copZ. One may consider NML45 to be a mammalian analogue of copZ, but with an NLS to allow access into the nucleus. The data, therefore, forecast a potential for overlap between yeast, bacterial and mammalian systems with regard to regulating the expression of Cu-ATPases. They also forecast nuclear transport of Cu as a requirement for maintaining Cu homeostasis. It remains to be determined whether intracellular Cu concentrations in any way influence NML45 targeting to the nucleus and whether nuclear Cu plays a role in determining the different isoforms of ATP7A mRNA.

This study was supported in part by U.S. Department of Agriculture grant 98-35200-6055 and grant P30-ES09106 awarded to the Center for Environmental and Rural Health at Texas A&M University. We also thank Dr Tom Glover for providing genomic clones used in the assays and for stimulating and insightful discussions.

REFERENCES

- Danks, D. M. (1995) Disorders of copper transport, in *The Metabolic and Molecular Basis of Inherited Disease* (Scriver, C. R., Baudet, A. L., Sly, W. M. and Valle, D., eds), pp. 2211–2235, McGraw-Hill, New York
- Vulpe, C. D. and Packman, S. (1995) Cellular copper transport. *Annu. Rev. Nutr.* **15**, 293–322
- Dierick, H., Ambrosini, L., Spencer, J., Glover, T. W. and Mercer, J. F. B. (1995) Molecular structure of the Menkes disease gene (*ATP7A*). *Genomics* **28**, 462–469
- Tümer, Z., Vural, B., Tonnesen, T., Chelly, J., Monaco, A. P. and Horn, N. (1995) Characterization of the exon structure of the Menkes disease gene using vectorette PCR. *Genomics* **26**, 437–442
- Mercer, J. F., Grimes, A., Ambrosini, L., Lockhart, P., Paynter, J. A., Dierick, H. and Glover, T. W. (1994) Mutations in the murine homologue of the Menkes gene in dappled and blotchy mice. *Nat. Genet.* **6**, 374–378
- Reddy, M. C. M. and Harris, E. D. (1998) Multiple transcripts coding for the Menkes gene: evidence for alternative splicing of Menkes mRNA. *Biochem. J.* **334**, 71–77
- Harris, E. D., Reddy, M. C. M., Qian, Y., Tiffany-Castiglioni, E., Majumdar, S. and Nelson, J. (1999) Multiple forms of the Menkes Cu-ATPase. *Adv. Exp. Med. Biol.* **448**, 39–51
- Vulpe, C., Levinson, B., Whitney, S., Packman, S. and Gitlschier, J. (1993) Isolation of a candidate gene for Menkes disease and evidence that it encodes a copper-transporting ATPase. *Nat. Genet.* **3**, 7–13
- Nakai, K. and Horton, P. (1999) PSORT: a program for detecting sorting signals in proteins and predicting their cellular localization. *Trends Biochem. Sci.* **24**, 34–36
- Mercer, J. F. B., Livingston, J., Hall, B., Paynter, J. A., Begy, C., Chandrasekharappa, S., Lockhart, P., Grimes, A., Bhawe, M., Siemieniak, D. and Glover, T. W. (1993) Isolation of a candidate gene for Menkes disease and evidence that it encodes a copper-transporting ATPase. *Nat. Genet.* **3**, 20–25
- Chelly, J., Tümer, Z., Tonnesen, T., Petterson, A., Ishikawa-Brush, Y., Tommerup, N., Horn, N. and Monaco, A. P. (1993) Isolation of a candidate gene for Menkes disease that encodes a potential heavy metal binding protein. *Nat. Genet.* **3**, 14–19

- 12 Qian, Y., Majumdar, S., Reddy, M. C. M. and Harris, E. D. (1996) Coincident expression of Menkes gene with copper efflux in human placental cells. *Am. J. Physiol.* **270**, C1880–C1884
- 13 Qi, M. and Byers, P. H. (1998) Constitutive skipping of alternatively spliced exon 10 in the ATP7A gene abolishes Golgi localization of the Menkes protein and produces the occipital horn syndrome. *Hum. Mol. Genet.* **7**, 465–469
- 14 Francis, M. J., Jones, E. E., Levy, E. R., Ponnambalam, S., Chelly, J. and Monaco, A. P. (1998) A Golgi localization signal identified in the Menkes recombinant protein. *Hum. Mol. Genet.* **7**, 1245–1252
- 15 Breatnach, R., Benoish, C., O'Hare, K., Gannon, F. and Chambon, P. (1978) Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the intron–exon boundaries. *Proc. Natl. Acad. Sci. U.S.A.* **75**, 4853–4857
- 16 Calkhoven, C. F. and Ab, G. (1996) Multiple steps in the regulation of transcription factor level and activity. *Biochem. J.* **317**, 329–342
- 17 Klaes, R., Kloor, M., Willeke, F., Melsheimer, P., von Knebel, D. and Ridder, R. (1999) Significant increase of a specific variant TSG101 transcript during the progression of cervical neoplasia. *Eur. J. Cancer* **35**, 733–737
- 18 Lin, S. Y., Chen, Y. J. and Chang, J. G. (1998) Multiple truncated transcripts of TSG101 in gastrointestinal cancers. *J. Gastroenterol. Hepatol.* **13**, 1111–1114
- 19 Ferrer, M., Lopez-Borges, S. and Lazo, P. A. (1999) Expression of a new isoform of the tumor susceptibility TSG101 protein lacking a leucine zipper domain in Burkitt lymphoma cell lines. *Oncogene* **18**, 2253–2259
- 20 Hittelman, A. B., Burakov, D., Iniguez-Lluhi, J. A., Freedman, L. P. and Garabedian, M. J. (1999) Differential regulation of glucocorticoid receptor transcriptional activation via AF-1-associated proteins. *EMBO J.* **18**, 5380–5388
- 21 Sun, Z., Pan, J., Hope, W. X., Cohen, S. N. and Balk, S. P. (1999) Tumor susceptibility gene 101 protein represses androgen receptor transactivation and interacts with p300. *Cancer* **86**, 689–696
- 22 Moore, M. S. (1998) Ran and nuclear transport. *J. Biol. Chem.* **273**, 22857–22860
- 23 Serpe, M., Joshi, A. and Kosman, D. J. (1999) Structure-function analysis of the protein-binding domains of Mac1p, a copper-dependent transcriptional activator of copper uptake in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **274**, 29211–29219
- 24 Strausak, D. and Solioz, M. (1997) CopY is a copper-inducible repressor of the *Enterococcus hirae* copper ATPases. *J. Biol. Chem.* **272**, 8932–8936

Received 29 February 2000/9 June 2000; accepted 13 July 2000