

Mammalian Mitochondrial Ribosomal Proteins (2)

AMINO ACID SEQUENCING, CHARACTERIZATION, AND IDENTIFICATION OF CORRESPONDING GENE SEQUENCES*

(Received for publication, June 9, 1999, and in revised form, September 19, 1999)

Thomas W. O'Brien‡, Scott E. Fiesler‡, Nancy D. Denslow‡, Bernd Thiede§,
Brigitte Wittmann-Liebold§, Edward B. Mougey¶, James E. Sylvester¶,
and Hanns-Rüdiger Graack||**

From the ‡Department of Biochemistry and Molecular Biology, College of Medicine, University of Florida, Gainesville, Florida 32610-0245, §Max-Delbrück-Center for Molecular Medicine, Robert-Rössle-Str. 10, D-13125 Berlin, Germany, ¶Nemours Children's Clinic, Research, Jacksonville, Florida 32207, and the ||Institute for Biology-Genetics, AG Kress, Free University of Berlin, Arnimallee 7, D-14195 Berlin, Germany

Four different classes of mammalian mitochondrial ribosomal proteins were identified and characterized. Mature proteins were purified from bovine liver and subjected to N-terminal or matrix-assisted laser-desorption mass spectroscopic amino acid sequencing after tryptic in-gel digestion and high pressure liquid chromatography separation of the resulting peptides. Peptide sequences obtained were used to virtually screen expressed sequence tag data bases from human, mouse, and rat. Consensus cDNAs were assembled *in silico* from various expressed sequence tag sequences identified. Deduced mammalian protein sequences were characterized and compared with ribosomal protein sequences of *Escherichia coli* and yeast mitochondria. Significant sequence similarities to ribosomal proteins of other sources were detected for three out of four different mammalian protein classes determined. However, the sequence conservation between mitochondrial ribosomal proteins of mammalian and yeast origin is much less than the sequence conservation between cytoplasmic ribosomal proteins of the same species. In particular, this is shown for the mammalian counterparts of the *E. coli* EcoL2 ribosomal protein (MRP-L14), that do not conserve the specific and functional highly important His²²⁹ residue of *E. coli* and the corresponding yeast mitochondrial Rml2p.

tent of mitochondrial ribosomes is higher than that of cytoplasmic ribosomes. It was estimated that mitochondrial ribosomes contain up to 100 different proteins (2, 3). Only a few MRPs have been characterized on the molecular level in the past (4–7). Recently, mammalian MRP sequences became available by direct N-terminal sequencing of purified proteins and screening of data bases for corresponding DNA sequences (8).² The direct approach applied by these authors is a versatile supplement to the laborious ways of MRP identification by mutational cloning or screening for similar proteins. Because MRPs are very heterologous among distantly related species, it is difficult to identify corresponding mammalian genes by screening with probes derived from yeast MRPs, for example. The direct characterization of MRPs from purified mitochondrial ribosomes offers the possibility to identify proteins that otherwise would remain unknown due to the lack of suitable reference proteins. Several “additional” MRPs have been identified in yeast and mammals (3, 8)^{2,3} that have no counterparts in bacterial or eukaryotic cytoplasmic ribosomes. Thus, it is our intention to characterize mammalian MRPs based on bovine MRPs as model proteins. The latter have been characterized by various biochemical and physical methods (2, 10–13). In providing molecular information on MRPs, we begin to understand the role of mammalian MRPs and their function in the maintenance of healthy mitochondria. In this paper, four different classes of mammalian MRPs are characterized and compared with their bacterial and yeast mitochondrial counterparts.

EXPERIMENTAL PROCEDURES

Analysis of Bovine MRPs—Isolation and purification of bovine MRPs has been described.² After two-dimensional PAGE and Coomassie Blue staining, individual spots of MRPs were cut from the gel. Individual proteins were termed in accordance to the established two-dimensional map of bovine MRPs (2). Proteins were subjected to in-gel digestion by trypsin according to the method of Otto *et al.* (14). Resulting peptides were isolated and concentrated, purified by reverse phase HPLC, and subjected to Edman sequencing (14) or tandem mass spectroscopic analysis with a Q-ToF (Micromass, Manchester, UK) equipped with a nanoflow Z-spray ion source. It should be noted that the latter method is not able to differentiate between leucine and isoleucine because of the identical MM of these two amino acids.

Computing—Virtual screening of public EST data bases was performed using the blast program of Altschul *et al.* (15) and the NCBI server. For screening of short peptides the advanced blast program was performed using the modified options “expected”: 1000 and “other options”: -e2. The analysis of the obtained sequences, the sequence com-

Mitochondrial ribosomal proteins (MRPs)¹ are the organellar counterparts of the cytoplasmic ribosomal proteins of the same species. MRPs are thought to be involved in the maintenance of the mitochondrial DNA, since mitochondria defective in protein biosynthesis lose their DNA successively (1). The protein con-

* This work was supported by United States Public Health Service Grant GM15438 (to T. W. O.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) S78756 (MRP-L5_{bov}), S78757 (MRP-L7_{bov}), S78758 (MRP-L14_{bov}), and S78759 (MRP-L26_{bov}), respectively.

** To whom correspondence should be addressed: Inst. for Biology-Genetics, AG Kress, Free University of Berlin, Arnimallee 7, D-14195 Berlin, Germany. Tel.: 49-30-838-2629; Fax: 49-30-838-3649; E-mail: graack@zedat.fu-berlin.de.

¹ The abbreviations used are: MRP, mitochondrial ribosomal protein; bp, base pair(s); MISP, mitochondrial import signal peptide; MM, molecular mass; ORF, open reading frame; PAGE, polyacrylamide gel electrophoresis; HPLC, high pressure liquid chromatography; EST, expressed sequence tag.

² Graack, H. R., Bryant, M. L., and O'Brien, T. W., (1999) *Biochemistry*, in press.

³ J. Anders, H.-R. Graack, J. Liu, and T. W. O'Brien, manuscript in preparation.

TABLE I

Peptide sequences of mature bovine mitochondrial ribosomal proteins obtained by amino acid sequencing

X represents amino acids that could not be identified, although the position in the respective peptide sequence in correspondence to the tryptic cleavage site is valid. MS sequence has been obtained by mass spectroscopy: leucine and isoleucine cannot be differentiated by this method. For amino acid residues in lowercase type, determination is uncertain. For MRP-L7_{bov} and MRP-L14_{bov}, respective sequences 2 and 3, amino acid sequences were obtained as a mixture, and individual sequences of peptides 2 and 3 were identified by comparison with the deduced amino acid sequence of MRP-L7_{human} and MRP-L14_{mouse}, respectively.

Protein name	Sequence	Molecular mass
		<i>kDa</i>
MRP-L5 _{bov}		
Peptide 1	XLPFELLEVEAK	35.5
Peptide 2	XLIYK	
MRP-L7 _{bov}		35.6
Peptide 1	VSLANLRPNPGSR	
Peptide 2	XYGVQLVEEGAD	
Peptide 3	XLPPEALVPYYTD	
MRP-L14 _{bov}		30
Peptide 1	ADIALVAGGNR	
Peptide 2	XVNGTAIIQ	
Peptide 3	XIIATENQKKP	
Peptide 4	mIDFK	
Peptide 5	(I/L)NNVESEGPR (MS)	
Peptide 6	(k/g)DT(I/L)(I/L)NSDHTGR (MS)	
Peptide 7	D(I/L)A(I/L)VAEVR (MS)	
MRP-L26 _{bov}	ISHGRVYRRLGLGPESRIHL	18.8

parison, and the assembly of virtual consensus cDNAs were performed as described (8) using several analytic programs of the GCG DNA analysis computer software package (16). Multiple alignments of sequences obtained were made using the "bestfit" program and by subsequent manual arrangements for optimal alignments to avoid domination of groups of, for example, the mammalian sequences. Blast searches for orthologous *Caenorhabditis elegans* protein sequences were performed on the Sanger Center's server. Calculations of N-terminal protein sequences for signal peptide properties were done with the Center for Biological Sequence Analysis server (17).

MRP L7 and L26 5'-Rapid Amplification of cDNA Ends—The 5' cDNA sequences of MRP L7 and L26 were obtained by performing rapid amplification of cDNA ends on cDNA from human fetus (CLONTECH, Marathon Ready cDNA) per the manufacturer's recommendations. Briefly, the 50- μ l first round amplification was 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 2 mM MgCl₂, 200 μ M dNTP, 200 nM adaptor primer (CLONTECH, AP1), and MRP-specific primer (L7c or L26c) and contained 500 pg of adaptor-ligated double-stranded cDNA, 2 units of *Taq* DNA polymerase (Life Technologies, Inc.), and 440 ng of TaqStart antibody (CLONTECH). Amplification was conducted using a step-down protocol as follows: 95 °C for 1 min, 95 °C for 30 s, and 72 °C for 2 min for five cycles; 95 °C for 30 s, 71 °C for 30 s, and 72 °C for 90 s for five cycles; and 95 °C for 30 s, 68 °C for 30 s, and 72 °C for 90 s for 25 cycles. The products of the first round amplification (4% of the total reaction) were subjected to a second round of amplification as above using the adaptor primer (AP1) and the appropriate MRP-specific primer (L7c or L26c) or one of several nested MRP-specific primers (L7b, L7a, L26b, or L26a). The single MRP-specific second round product from each reaction was isolated from an agarose gel, cloned (Invitrogen, pCRII-TOPO), and sequenced. The MRP-specific primers are as follows: L7a (5'-GGATGTAATGGAGTCTGGCCTCCCTC), L7b (5'-GGCTGGATGGTCACACCTCTCCC), L7c (5'-GGATCATAGAAG-GCTGAAGTAACAACACCACC), L26a (5'-CCCCCAGTTTGTATCTTT-GTACCGAGGG), L26b (5'-GCGAGGCAGAGGCAGGGGTG), and L26c (5'-CTCTTCAGATCCAGTTAAATCCCTGGTGTGTTGAGC).

RESULTS

Purification and Characterization of Individual MRPs—Ribosomal proteins of the bovine mitochondrial large ribosomal subunit have been purified previously.² We determined the MM of several individual MRPs by SDS-gel electrophoresis (Table I; see also Ref. 2). In parallel, proteins of the large mitoribosomal subunit were separated by two-dimensional

PAGE and individual MRPs subjected to in gel tryptic digestion. Peptides derived from individual proteins were separated by RP-HPLC and analyzed either by Edman N-terminal amino acid sequencing or by matrix-assisted laser-desorption mass spectroscopy. Several short peptide sequences were obtained (Table I), although some of them only as mixtures from two different peptides (see below). Peptides that terminate either with an arginine or lysine residue are considered to be sequenced completely, since these peptides result from a proteolytic cut C-terminal of their Lys or Arg residue.

Virtual Screening of Data Banks and Assembly of Consensus cDNA Sequences for Mammalian MRPs—The longest amino acid sequences (Table I) obtained for each of the analyzed MRPs were used to screen human, mouse, and other EST data banks. In all cases, several EST sequences were identified, and consensus cDNAs were assembled by repetitive screens and comparison of EST sequences of one species or among species to eliminate cDNA sequencing errors and to obtain continuous and conserved ORFs of human, mouse, and rat corresponding to the different bovine MRPs analyzed.

MRP-L5—Twelve and five "primary hits" were found for human and rat MRP-L5, respectively, by a preliminary screen of the EST data banks using MRP-L5_{bov} peptide 1 (Table I) as *in silico* screening probe. The MRP-L5_{human} consensus cDNA was assembled from four different ESTs, yielding a sequence of 1072 nucleotides encoding a complete (start codon to stop codon) ORF of 1017 bp (Table II). Additionally, two corresponding human genomic DNA clones (GenBankTM accession nos. AP000086 and AP000223) were identified. Sequence information from the genomic clone (AP000086) was used to confirm that in the 5' direction no further translational start codon is localized. Sequences from a putative TATA box in the genomic DNA downstream to the start codon of the most 5' EST were added to complete the consensus cDNA deduced from the ESTs (Table II). Two consecutive polyadenylation signals, AATAAA, were identified 3' from the stop codon. A continuous poly(A) stretch starts at bp 1451. Thus, the polyadenylation sites are considered to be the true signal sequences for this modification of the MRP-L5_{human} mRNA. The ORF of MRP-L5_{human} (Table II) encodes a protein of 338 amino acid residues (Fig. 1a). The MRP-L5_{bovine} peptides 1 and 2 (Table I) correspond to a sequence located in the center of the deduced MRP-L5_{human} sequence (Fig. 1a). The sequence is strongly conserved between bovine and human at these positions, with 14 out of 15 amino acid residues identical (Fig. 1a). For MRP-L5_{human} a putative cleavable mitochondrial import signal peptide (MISP) of 21 amino acid residues was calculated by the SignalP program (17). The mature MRP-L5_{human} after cleavage of the MISP has a calculated MM of 36,583 Da. This value is in good correspondence to the MM of the mature MRP-L5_{bovine} (*i.e.* after cleavage of a putative MISP) of 35.5 kDa (Table I), based on the proposal that mammalian MRPs are quite similar in their properties like MM, pI, and amino acid sequence (8).² For the processed form of MRP-L5_{human}, a pI of 7.36 was calculated. This value coincides with the characterization of MRP-L5_{bovine} and MRP-L5_{rat} as relatively less basic MRPs by two-dimensional PAGE (2) and comparative two-dimensional PAGE, respectively (18).

Comparison of the genomic clone and the assembled cDNA revealed the exon/intron structure of the MRP-L5_{human} gene (Fig. 2). The coding region of MRP-L5_{human} is spread out into 10 small exons separated by nine introns up to 3.5 kilobases in size. Exon A (bp 1–74 of the consensus cDNA; Table II) corresponds to bp 13,742–13,815 of the genomic DNA (AP000086) in reverse orientation, exon B (cDNA bp 75–280) corresponds to bp 12,789–12,994 (AP000086), exon C (cDNA bp 281–420) corresponds to bp 10,136–10,275 (AP000086), exon D (cDNA

TABLE II
 Assembly of identified EST sequences: consensus cDNAs in 5' to 3' direction

Determination of consensus cDNA sequences for deduction of mammalian MRP ORFs by assembly of EST sequences is shown. None of the EST sequences listed contains incompletely spliced products. Asterisks label incomplete ORFs. r.c, the nucleotide sequence was determined in reverse complement orientation. +, the assigned ORF ignores three frameshifts in accordance with the strong amino acid sequence conservation between the human 1, human 2, and human 3 sequences and the comparison with the CGI-22 sequence identified from the data bases. R-protein family, affiliation of deduced ORFs to existing families of similar ribosomal proteins.

Gene	bp Nos.	EST accession no.	Total length	ORF from bp no. to no.	Protein family
MRP-L5 _{human}	1-60	AA447882	1072	1-1017	None
	52-360	AA452180			
	1-480	AA426183			
MRP-L5 _{mouse}	256-479	AI051104r,c	1072	4-1014	None
	1-480	AI391121			
	330-531	AA124304			
MRP-L5 _{rat}	45-434	AI504375r,c	580	*1-525	None
	1-580	AI229051			
MRP-L7 _{human}	1-60	AA088577	1142	57-944	EcoL15
	28-120	AA460632			
	115-420	W00599			
	82-420	AA262046			
	147-490	AI436106r,c			
MRP-L7 _{mouse}	1-120	AA139966	1047	*2-922	EcoL15
	73-420	AA915116			
	16-300	AA266104			
	29-322	AA177229			
MRP-L7 _{rat}	1-362	H31581	362	*3-299*	EcoL15
	1-674	AI103129r,c			
MRP-L14 _{human1}	1-61	R72376	922	45-881 ⁺	EcoL2
	56-80	AA188450			
	1-250	AA479875			
	305-357	W68014			
	303-392	AA479875			
MRP-L14 _{human2}	64-506	AA513729r,c	517	*1-468	EcoL2
	1-240	AI038157r,c			
	240-298	AA513729r,c			
MRP-L14 _{human3}	301-518	AI038157r,c	603	*1-543	EcoL2
	1-240	AI038157r,c			
MRP-L14 _{mouse}	79-441	AI167786r,c	1012	50-970	EcoL2
	1-26	AA869812			
MRP-L14 _{rat}	1-552	AI043215	577	*2-535	EcoL2
	176-369	AA542368			
	121-360	AI324254r,c			
	1-577	AI171870r,c			
	1-7	AA316691			
MRP-L26 _{human}	1-52	AA307568	754	21-548	EcoL17
	36-406	AA398878			
	362-404	AA316691			
	434-458	AA312160			
	354-609	AA521444r,c			
MRP-L26 _{mouse}	1-479	AA0330585	1124	98-628	EcoL17
	109-409	AA560127			
	138-481	AA032939			
MRP-L26 _{rat}	1-17	H34212	666	14-307*	EcoL17
	4-652	AI231800r,c			
	1-543	AI103095r,c			

421-520) corresponds to bp 7708-7807 (AP000086), exon E (cDNA bp 521-588) corresponds to bp 6138-6207 (AP000068), exon F (cDNA bp 589-701) corresponds to bp 3642-3754 (AP000086), exon G (cDNA bp 702-767) corresponds to bp 233-298 (AP000086), exon H (cDNA bp 768-922) corresponds to bp 28,852-29,006 (AP000223, in reverse orientation), exon I (cDNA bp 923-969) corresponds to bp 24,876-24,922 (AP000223), and exon J (cDNA 970-1068) corresponds to bp 21,697-21,795 (AP000223). The total MRP-L5_{human} gene spans more than 22 kilobases. The genomic clones have been mapped to position 21q22.1.

For MRP-L5_{mouse}, no hits were found in the primary screen of mouse EST sequences using MRP-L5_{bovine} peptide 1 as screening probe. Instead, the complete MRP-L5_{human} protein sequence (Fig. 1a) was used to screen mouse ESTs for corresponding entries. A consensus cDNA of 1072 bp was assembled from three different ESTs (Table II) encoding a continuous

ORF of 1010 bp. However, comparison of the amino acid sequence of MRP-L5_{mouse} (Fig. 1a) deduced from the completely assembled consensus cDNA shows that only seven out of 11 amino acid residues are identical between MRP-L5s of mouse and bovine in this position. This value is below the threshold of the blast program used (15), explaining the inability to detect mouse ESTs by using the bovine sequences. Similar to the human cDNA, two consecutive polyadenylation signals, AT-TAAA and AATAAA, were found 3' of the stop codon. Only the last 3 nucleotides of the consensus cDNA (bp 1070-1072) may represent the 5'-end of the continuous poly(A) stretch, which can be deduced from the proximity of the two polyadenylation signals mentioned. No putative splice sites were determined in the center and the 3'-end of this consensus cDNA by sequence comparison of various EST sequences. A complete ORF encoding for 336 amino acid residues was determined (Fig. 1a). The deduced protein has a calculated MM of 38,548 Da and a

Alignment of deduced mammalian MRP amino acid sequences with MRP sequences of other sources

FIG. 1. Alignment of deduced mammalian MRP amino acid sequences with MRP sequences of other sources. Obtained deduced mammalian MRP amino acid sequences are aligned with bovine peptide sequences (Table I) and with corresponding ribosomal protein sequences of yeast MRPs and *E. coli*. Numbers give the respective amino acid position. Vertical lines mark identical amino acid positions; colons mark strongly conserved amino acid residues, and periods (.) mark weakly conserved amino acid residues. Dashes show N- or C-terminal ends of incomplete amino acid sequences, and asterisks mark stop codons. Amino acid residues in lowercase letters are uncertain in their determination by amino acid sequencing (see Table I). *x*, unidentified amino acids, although the positional number is valid in respect to the tryptic digest that created the respective peptides. *a*, alignment of mammalian MRP-L5 sequences. *b*, alignment of mammalian MRP-L7 sequences with the yeast YmL10 MRP sequence (accession no. Z71560) and the *E. coli* EcoL15 sequence (accession no. Sw:rl15_ecoli). *c*, alignment of the mammalian MRP-L14 protein sequences with the yeast RML2p MRP sequence (accession no. U18779) and the *E. coli* EcoL2 sequence (accession no. Sw:rl2_ecoli). Amino acid residues in boldface type indicate functional important and conserved amino acid positions according to Refs. 9 and 21. *d*, alignment of the mammalian MRP-L26 protein sequences with the yeast YmL8 MRP sequence (accession no. P22353) and with the *E. coli* EcoL17 sequence (accession no. Sw:rl17_ecoli).

a MRP-L5

mouse	MATAVGRILVLRFRGAGGQARWRFTATSPAAELSPTELTEMRNDLFNREKSRQLSLTPRTEKIEVHKVVKDTEGTFVMMNRNITSPYSCAMHLSWYC	97
human	MEALAMGRALRLWLVAPOGGIKWRFTATSPASQLSPTELTEMRNDLFNREKSRQLSLTPRTEKIEVHKVVKDTEGTFVMMNRNITSPYSCAMHLSWYC	100
rat		
mouse	SKSILALVDGQPMDMYKPLTKSCIEKFLTFKDFDPFKVNRKAWRSCAMMLCVIERAFKDDVYVSVLRAPEVVTAGAFYDVTLDKRLDEMPTRENLR	197
human	RKSILALVDGQPMDMYKPLTKSCIEKFLTFKDFDPFKVNRKAWRSCAMMLCVIERAFKDDVYVSVLRAPEVVTAGAFYDVTLDKRLDEMPTRENLR	200
bovine peptides	MLIYKMLPELLEVEAK	
rat	SPTKDAHALLYRDLDFPELTVDAARVALIEIQNNKYVDFIEEKASQNPRIYVKLHRIGDPIVISEGFLIPRTSUCFOYEVSAVINLPRSQPLVRRFQGL	135
mouse	SPTKDAHALLYRDLDFPELTVDAARVALIEIQNNKYVDFIEEKASQNPRIYVKLHRIGDPIVISEGFLIPRTSUCFOYEVSAVINLPRSQPLVRRFQGL	297
human	SPTKDAHALLYRDLDFPELTVDAARVALIEIQNNKYVDFIEEKASQNPRIYVKLHRIGDPIVISEGFLIPRTSUCFOYEVSAVINLPRSQPLVRRFQGL	300
rat	SLPIHLRAQFTINDKLLERSKMTVEITQTESTAPTO*	174
mouse	SLPIHLRAQFTINDKLLERSKMTVEITQTESTAPTO*	338
human	SLPIHLRAQFTINDKLLERSKMTVEITQTESTAPTO*	338

b MRP-L7

bovine peptides	VSLANLRNFGSR	
rat	-GNAFGRTSLDLISLPRVSLANLXPSLNSKRRERRPDRNRGRKCGRG.HKGEQRQTR.....PRLGFXG	66
mouse	-RFQGALEKTVGRCWVGARLRTSLDLISLPRVSLANLXPSLNSKRRERRPDRNRGRKCGRG.HKGEQRQTR.....PRLGFXG	79
human	MAGPLQGGGARLDLRLGPRVSLANLXPSLNSKRRERRPDRNRGRKCGRG.HKGEQRQTR.....PRLGFXG	87
YmL10	MKAERTGLRNSFTTIGRLINTFVSMMLTSLFKSPVLAQSVYVSLGQLKPSDGTAKSPKRLRQPSGLKTSGR.....VKSWEFG	88
<i>E. coli</i> L15	MRLNTPSPAEGSKAGKALGRITGSGLKTGCRGHKQKRSKGGVRRGEG	52
bovine peptides	xYGVQLVEEGAD	
rat	GOTPFYIRPKYGFNGHSHFQYQPLSLRRLQ- 99 -IPIKRYGVGVLEKADTFQAKVNIEMQASLAIATAEK	41
mouse	GOTPFYIRPKYGFNGHSHFQYQPLSLRRLQVLL.DLGRVDPQPIDLTQLVNGRGVITQPLKRDYGVVLEEGADTFQAKVNIEMQASLAIATAEK	179
human	GOTPFYIRPKYGFNGHSHFQYQPLSLRRLQVLL.DLGRVDPQPIDLTQLVNGRGVITQPLKRDYGVVLEEGADTFQAKVNIEMQASLAIATAEK	187
YmL10	GOTPFYIRPKYGFNGHSHFQYQPLSLRRLQVLL.DLGRVDPQPIDLTQLVNGRGVITQPLKRDYGVVLEEGADTFQAKVNIEMQASLAIATAEK	175
<i>EcoL15</i>	GMPLVRLRPQPTSKKA..ATTAEIRLSLAKVEGGVVL.....NTLKAANITGQIE.FAKVILAG...EVTTPVTVRGLRVTRGARAATA	137
bovine peptides	MLPPEALVPPYTD	
rat	NGGVVTFAYDPRSLIEILCKVFPFRLRQPIFKRMLPPEALVPPYTDKRNRYLADPAKFPPEARLELMKYGVVLDVTKDELPMSTRKDRPQIFFLGA	141
mouse	NGGVVTFAYDPRSLIEILCKVFPFRLRQPIFKRMLPPEALVPPYTDKRNRYLADPAKFPPEARLELMKYGVVLDVTKDELPMSTRKDRPQIFFLGA	279
human	NGGVVTFAYDPRSLIEILCKVFPFRLRQPIFKRMLPPEALVPPYTDKRNRYLADPAKFPPEARLELMKYGVVLDVTKDELPMSTRKDRPQIFFLGA	287
YmL10	AGLGLRAHLNPFQWLEKRGVPLQARPTKRRDIDFYSKEEKRYLVMEKDNKGRSHFLKQNVKSSLELEELSPKDVVNVVNSKVMKIKALD*	272
<i>EcoL15</i>	AGGKIE* 144	
rat	PGWVNMADKILKPTDENLKYYS*	168
mouse	PGWVNMADKILKPTDENLKYYS*	306
human	PGWVNMADKILKPTDENLKYYS*	314

calculated pI of 7.92. The deduced protein sequence of MRP-L5_{mouse} itself correspond very well to the sequence of MRP-L5_{human}, both sequences share 88.4% identical residues over a length of 320 amino acids. This high degree of sequence similarity is consistent with the general high similarities of mammalian MRPs (8).² However, by the SignalP program (17) no cleavage site for a MISP was proposed, since both proteins at their respective N termini are lacking a strong sequence similarity (Fig. 1*a*). This may result from the occurrence of a splice variant of the MRP-L5_{mouse} consensus cDNA found in the 5' direction from the sequence encoding 3 consecutive glycine residues (Fig. 1*a*). No mouse splice variant showing a high degree of sequence similarity to the MRP-L5_{human} was found. Another possibility is that the respective MISPs of MRP-L5s from human and mouse have been less strictly conserved as compared with the respective functional, mature proteins during evolution.

For MRP-L5_{rat} only ESTs corresponding to the C terminus were found. From the longest EST (Table II) a peptide of 174 amino acid residues was deduced that is in very good sequence correspondence to the mouse and human MRP-L5s (Fig. 1*a*). Further, the rat and mouse MRP-L5 cDNAs are quite similar in their respective 3'-untranslated trailer sequences. Like the mouse cDNA, the rat cDNA contains two consecutive polyadenylation signals, ATTAAA (bp 545–560) and AATAAA (bp 560–565), and a putative poly(A) track beginning at bp 577.

In searching the data bases for similar proteins, no corresponding sequences from eubacterial or mitochondrial sources were found. Mammalian MRP-L5s represent a new class of ribosomal proteins not found in ribosomes of bacterial or eukaryotic cytoplasmic origin. However, screening the *C. elegans* genomic data base, a contig (Y62F5.ctg01193) was identified that encodes a stretch of 238 amino acid residues showing 26% sequence identity to the MRP-L5_{human}. It remains speculative if this relatively low sequence similarity is significant.

MRP-L7—Fourteen, 15, and three primary hits of human, mouse, and rat origin, respectively, were found using bovine MRP-L7 peptide 1 (Table I) as virtual screening probe. For MRP-L7_{human}, a consensus cDNA of 1142 bp was assembled from five different EST sequences (Table II). Splice sites were localized at nucleotide positions 37/38, 52/53, and 392/393 by comparison of several completely and incompletely spliced EST sequences. However, to determine the true mature 5' splice form, we performed a 5'-rapid amplification of cDNA ends cloning of the corresponding cDNA from a fetal cDNA library and sequenced the obtained cDNA at its 5' terminus (data not shown). The obtained sequence perfectly matched one of the EST-derived 5' splice variants to provide the mature 5' terminus of the native MRP-L7 cDNA. A polyadenylation signal, ATTTAAA, was found at bp 1095–1100, and the poly(A) track begins at bp 1125. A continuous ORF of 296 amino acid residues was determined. The start codon at bp 57–59 is preceded

Alignment of deduced mammalian MRP amino acid sequences with MRP sequences of other sources

C MRP-L14

mouse		MALCALTSALRSLASAAITARVPTLLPAAQTSNVLQLPPALVSPSYRP	52
human1		MALCALTRALLVLNLAAPPVVAPAPSLPFAAQMMNGLLQPSALMLLPCRP	52
human CGI-22		MALCALTRALLVLNLAAPPVVAPAPSLPFAAQMMNGLLQPSALMLLPCRP	52
yeast RML2	MLVLGSLRSALSCSSTASLISKRNPCYPYIILCRITLSQSVMKQENTSKDDSSLNITPKLLKTIIPNDTIVLTKQDELKRRRKLSEVQPKRKLKPV		100
E. coli EcoL2		MAVVKCKPTS	10
bovine peptides		mIDFK	ADIALVAGGNR
rat			-VRVDXCRSADIALVAGGSRKR..
mouse	VHMSADRSKAFVSWKSRKTYITPVKMRKSGGRDHTGRIRVHGIGGGHKQNYRMIDFLFRPEPEKAPPEEKKVQVRYDPCRSADIALVAGGSRKR..		150
human1	VLTSVALNANFVSWKSRKTYITPVKMRKSGGRDHTGRIRVHGIGGGTKQRYRMIDFLFRPE. ETKSGPFEEKVQVRYDPCRSADIALVAGGSRKR..		149
human2		-RFKVGVDPRSDIALVAGGSRKR..	25
human3		-RFKVGVDPRSDIALVAGGSRKR..	25
CGI-22	VLTSVALNANFVSWKSRKTYITPVKMRKSGGRDHTGRIRVHGIGGGHKQNYRMIDFLFRPE. ETKSGPFEEKVQVRYDPCRSADIALVAGGSRKR..		149
RML2	PGLRMYSPIPYLXGRPVRLTVVRKHKGRNNSKTIYRHQGGHNRRTLIDF.....NRWEGGQVQRILEYDGRSSHIALKHNTPGELS		192
EcoL2	PGRRHVVKVNPFLHKGKPFAPLLEKNSKSGGRNNGRIITPHIGGGHKQAYRIVDF.....KRNKDGIPAVVERLEYDPRNSANIALVLY.KDGERR		102
bovine	xIIATENQK DITLNS.....DHTGR	INNVSESGPR	xV..NOTAIIQ
rat	WIIATENKAGDTILNS.....NHIGRMVAA.....QEGDAHPLGALPVGTLINNVSESGPRGAQYTRAACTCGVLLRKY..NGTAIIQLPSK		103
mouse	WIIATENKAGDTILNS.....NHIGRMVAA.....QEGDAHPLGALPVGTLINNVSESGPRGAQYTRAACTCGVLLRKY..NGTAIIQLPSK		232
human1	WIIATENKAGDTILNS.....NHIGRMVAA.....REGDAHPGLALPVGTLINNVSESGPRGAQYTRAAAGANNV.....		263
human2	WIIATENKAGDTILNS.....NHIGRMVAA.....REGDAHPGLALPVGTLINNVSESGPRGAQYTRAAAGANNV.....		91
human3	WIIATENKAGDTILNS.....NHIGRMVAA.....REGDAHPGLALPVGTLINNVSESGPRGAQYTRAACTCGVLLRKY..NGTAIIQLPSK		107
CGI-22	WIIATENKAGDTILNS.....NHIGRMVAA.....REGDAHPGLALPVGTLINNVSESGPRGAQYTRAACTCGVLLRKY..NGHSIIQLPSK		231
RML2	YIYACDLRPGDVFSPFRGIPQTLNEMGKVDPAILSVKTPQRGNCLPISMEIYPTIHNWGTTPVGGPKCRSAGTYARVLAKLPEKXAIYRLQSG		292
EcoL2	YIIAPKGLKAGD.....IQSGVDAAI.....KPGNTLPMRNI PVGTVINVMKPKGGQLRSAGTYVQIVA..RDGAYTILRLRSG		179
rat	RQMVLESCTATVGRVSNVNHNRVIGKAGRNRLGKRFNSLW.....QRKGGWAGRKIRPLPMKSYVKLPSAAAQ*		177
mouse	SKMVLSESCYVGRVSNVNHNRVIGKAGRNRLGKRFNSLW.....QRKGGWAGRKIRPLPMKSYVKLPSAAAQ*		306
human1VATVGRVSNVDHNRVIGKAGRNRLGKRFNSLW.....MRKGGWALLKIRPLPMKSYVKLPSASAQ*		328
human2VATVGRVSNVDHNRVIGKAGRNRLGKRFNSGRW.....MRKGGWAGRKIRPLPMKSYVKLPSASAQ*		156
human3	RQMVLETCTVATVGRVSNVDHNRVIGKAGRNRLGKRFNSGRW.....MRKGGWAGRKIRPLPMKSYVKLPSASAQ*		181
CGI-22	RQMVLETCTVATVGRVSNVDHNRVIGKAGRNRLGKRFNSGGTARGAGLPKDSATTPHEELREAAFCFCPKLISLVSXMKMPPNVILCYCGFVLGVFC		331
RML2	EHRVYSLAVATTGVVSNIDHQRSLGKAGRSRWLGIRPTVGVY.....AMNKCDEPHGGGGRKSKSNKLSMSPQQLAKGYKTRGKNQNRMKVDR		385
EcoL2	EMRKVEADCRATLGEVQNAEHLRVLGKAGAAARWGRVPRMGRVPTVVRGTANPVDMPHGGGEGRNKQKHP.VTQVGVQTKKKTRSKRTRKIVRR		271
CGI-22	FFCCFFVFAGEVMQTKRGEK*		353
RML2	PRGKDARL*		393
EcoL2	SK*		273

Fig.1—continued

Alignment of deduced mammalian MRP amino acid sequences with MRP sequences of other sources

d MRP-L26

bovine N-terminus	ISHGRVYRRLGLGPESRIHL		
rat	MRLSLAAAISHGRVYRRLGLGPESRFHLLRNLTLGLVHERIEATWARVDEMGRVAEKLIDYGLGDTNERAMRMDPWL.....TEKDLI		86
mouse	MRLSLAAAISHGRVYRRLGLGPESRFHLLRNLTLGLVHERIEATWARVDEMGRVAEKLIDYGLGDTNERAMRMDPWL.....TEKDLI		86
human	MRLSVAAAISHGRVFRMGLGPESRIHLRNLTLGLVHERIEATWARVDEMGRVAEKLIDYGLGDTNERAMRMDPWL.....TEKDLI		86
yeast YmL8	MTVGIAFRKDRDKAHRDALLAQALQFLGHESIVSTHAKCKEASRVAERITWTYKRAITTSNGVAQAEKLSQIQSOLFAGDNKRLM		87
E. coli EcoL17	MRRKSGQLNRNSHRQAMFRMAGSLVRHEIITKTLPKAKELRVVPLTTLAKTDSVNRRLAFART.....RDMEV		76
rat	PKLFLVLAAPRFQ- 98	-IEYGNCLPPLPMPHRDNLTLNQLLGLQODLHNGKASLHSSHTVQTPKT*	53
mouse	PKLFLVLAAPRFQGGQNYTRMLQIPNRKEQDRAKNAVIEYGNVLPPLPMPHRDNLTLNQLLGLQODLHNGKASLHSSHTVQTPKT*		176
human	PKLFLVLAAPRYDQTCGYTRMLQIPNR.SLDRAKNAVIEYGNCLPPLPMPHRDNLTLNQLLGLQODLHNGKASLHSSHTVQTPGI*		175
YmL8	KRLFSEIAPRYLERPGVTVRLRLRLEPR.ANDSAPQSVLELWDSVPMSESHTVNRGNLKMMLLVKSVINDANQLPHNPLTLQNLHKVAKFAEAQLH		183
EcoL17	AKLFNQLGPR.....GGYTRILKCGFR.AGDNAFMAVIELVDRSEKAFAAAE*		
YmL8	GETMLIKVLLREMSIPYDEALENERTQALLKEVYSSSLPKKTKKPSVYVMVPR*		238

by an in-frame stop codon at bp 33–35. The relative positions of the three different MRP-L7_{bov} peptides (Table I) were determined, although peptides 2 and 3 were obtained as a mixed sequence by Edman sequencing that could be resolved only by comparison with the deduced MRP-L7_{human} protein sequence (Fig. 1b). The sequences of the MRP-L7_{bov} peptides are in very good correspondence to the deduced MRP-L7_{human} sequence (Fig. 1b). We postulate an MISP of 21 amino acid residues by comparison of the peptide 1 sequence with the deduced MRP-L7_{human} ORF. However, this MISP of the MRP-L7_{human} remains speculative, since the N terminus of the corresponding bovine peptide 1 seems to be a product of an experimental

tryptic endoproteolytic digest rather than of an MISP cleavage during import into the mitochondrion (Fig. 1b). The SignalP program (17) failed to identify any putative cleavage site within the N-terminal 42 amino acid residues of the deduced MRP-L7_{human} ORF. The postulated mature MRP-L7_{human} protein has a calculated MM of 35,403 Da and a calculated pI of 10.93. The MM of MRP-L7_{human} is in good agreement with the MM of the mature MRP-L7_{bov} of 35.6 kDa determined by SDS-PAGE (2). The calculated pI seemed not to correspond to the electrophoretic migration properties of the corresponding MRP-L7_{bov} (2). The protein appears as a slow moving, streaking/tailing protein in the first dimension, suggesting that it is

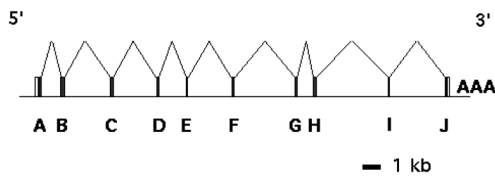


FIG. 2. Exon/intron structure of the human MRP-L5 gene. The genomic DNA is shown as a horizontal line. Boxes mark location of cDNA sequences; black boxes represent translated cDNA sequence, and open boxes show untranslated cDNA sequences. Exons are termed consecutively A, B, C, etc. AAA, location of poly(A) tract at the 3'-end of exon J.

significantly more acidic than the average MRPs (2). The strange migration properties result from the tight binding of RNA pieces of variable sizes when extracted by the urea/LiCl method (2). If bovine MRPs are extracted following an RNase treatment of the ribosomal subunits, MRP-L7_{bov} migrates along with MRPs of higher pI.⁴

For MRP-L7_{mouse} a consensus cDNA of 1047 bp was assembled. Putative splice sites were determined at bp 31/32, 245/246, 464/465, and 977/978. A polyadenylation signal, ATTAAA, was identified at bp 983–988, and a consecutive poly(A) stretch begins at bp 1006. The deduced MRP-L7_{mouse} ORF is incomplete at its N terminus. There was no mouse EST identified corresponding to the N terminus of the MRP-L7_{human}. The MRP-L7_{mouse} ORF encodes 306 amino acid residues and shows high sequence similarity compared with the MRP-L7_{human} protein (91.5% identity over 284 amino acid residues; see Fig. 1b). The deduced protein has a calculated MM of 34,995 Da and a pI of 11.01. Like MRP-L7_{human}, MRP-L7_{mouse} is a highly basic protein. For MRP-L7_{rat}, two different ESTs were identified that encode nonoverlapping fragments (Table II, Fig. 1b). The ORFs deduced from these ESTs show very high sequence similarities compared with the respective mouse and human MRP-L7 sequences (Fig. 1b). Thus, these ESTs can be anticipated to represent the corresponding rat MRP-L7 cDNA. No signal peptide cleavage sites were predicted by the SignalP program (17) within the 50 most N-terminal amino acid residues of the respective mouse and rat MRP-L7 sequences.

Sequence comparison with ribosomal proteins from other sources revealed that the mammalian MRP-L7s correspond to the yeast MRP YmL10 and the *E. coli* EcoL15 ribosomal protein (Fig. 1b). However, the conservation is not very high (e.g. 38.3% sequence identity and 48.2% similarity between MRP-L7_{human} and YmL10 over a stretch of 126 amino acid residues). Additionally, the mammalian MRP-L7 and, very unusually, the EcoL15 sequences contain insertions compared with the yeast YmL10 (Fig. 1b). In the *C. elegans* genomic data base, a contig (Y92H12.Contig307) containing corresponding DNA sequences was identified. Five different DNA stretches encoding 24–58 amino acid residues, respectively, show sequence identities between 29 and 41% to the mammalian MRP-L7 sequence. Putatively, the scattering of the short peptide-encoding DNA sequences reflects the exon/intron structure of the *C. elegans* gene ortholog to mammalian MRP-L7.

MRP-L14—From MRP-L14_{bov}, several different peptides were obtained after in-gel digestion of mature MRP-L14_{bov} by trypsin. Seven different peptide sequences were obtained by N-terminal Edman sequencing (Table I), although peptides 2 and 3 were provided as a mixed sequence that could be resolved only by the comparison to the deduced MRP-L14_{human} sequence (see below). The sequence information of peptide 1 was used for a primary screening of human, mouse, and rat EST data banks. Two, 14, and three primary hits were found, respectively. For

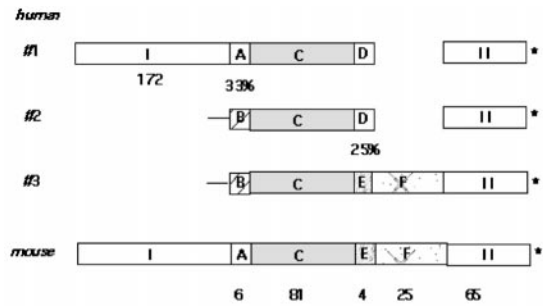


FIG. 3. Splice variants of mammalian MRP-L14 cDNAs. Three different human cDNAs (#1 to #3) and one mouse cDNA are compared. I, II, and C, identical exons found in all cDNAs as indicated. A, B, and D–F, variable exons. Numbers give extensions in amino acid residues encoded by the respective exons. Percentage numbers mark the proportion of identical amino acid residues between the exons above and below (A versus B and D versus E). Asterisks mark stop codons. Dashes mark incomplete cDNAs that are lacking 5' sequences that are unavailable from current EST data banks.

MRP-L14_{mouse}, a consensus cDNA of 1012 nucleotides was assembled (Table II). Furthermore, intron splice sites were localized at bp 139/140, 251/252, and 314/315, respectively, by comparison of incompletely spliced individual EST sequences with the assembled consensus cDNA sequence. A continuous ORF of 920 bp was determined (Table II). Between bp 985 and 990 a perfect polyadenylation signal, AATAAA, was identified. A continuous poly(A) track is found from bp 1006 up to the end of the consensus cDNA sequence. The nucleotide neighborhood of the first ATG codon fits the consensus sequence of eukaryotic start codons (19). Thus, we propose this ATG codon to be the true translation initiator of the MRP-L14_{mouse} sequence, although no 5' in frame stop codon was found. The deduced ORF encodes a protein of 306 amino acid residues (Fig. 1c). Compared with the MM of the mature MRP-L14_{bov} of 30 kDa (Table I), an MISF of approximately 30 amino acids may be proposed for the MRP-L14_{mouse}. By the SignalP program (17), a cleavage site between amino acid residues 18 (Ala) and 19 (Ala) is proposed. After cleavage of this proposed MISF, the mature MRP-L14_{mouse} has a calculated MM of 31,510 Da, which fits the MM of MRP-L14_{bov} deduced by SDS-PAGE (Table I). The calculated pI is 11.66. This is consistent with the finding of Pietromonaco *et al.* (18) that bovine and other mammalian MRPs have similar pIs. MRP-L14_{bov} has been characterized as the most basic protein in its class (18).

A single EST of rat origin was identified (Table II). From this EST sequence a 5' truncated ORF of 177 amino acid residues was deduced, which shows strong sequence similarity to all except one of the bovine peptide sequences (Fig. 1c).

Further, several EST sequences were identified for MRP-L14_{human}. They separate into three different classes by their overlapping sequence identities (Table II). Three differentially spliced versions of the MRP-L14_{human} cDNA were assembled (Figs. 1c and 3). Only cDNA 1 encodes for a complete MRP-L14_{human} ORF (Fig. 1c). Since the C-terminal part of the consensus cDNA 1 is deduced from a single EST sequence only (Table II), three frameshifts in the deduced amino acid sequence (*lowercase and underlined letters* in the human1 sequence, Fig. 1c) have to be ignored to obtain a highly conserved peptide sequence compared with the two other splicing variants of MRP-L14_{human} and MRP-L14_{mouse}. Additionally, a published human cDNA CGI-22 (accession no. AF132956) was found in the data bases. This cDNA was cloned by comparative gene cloning with the *C. elegans* proteome as template, but not further characterized. Also, it is not specified in the submission how accurate the sequencing was, e.g. from both strands or single reading only. However, it fits to the MRP-L14_{human}1

⁴ T. W. O'Brien, unpublished observations.

consensus cDNA overriding the primary two frameshifts in the MRP-L14_{human}1 consensus cDNA mentioned (Fig. 1c). It does follow the third mentioned frameshift, producing an elongated ORF that is, from the frameshift on, no longer similar to all other MRP-L14_{mammalian} or orthologous sequences from other sources (Fig. 1c). Comparing all mammalian MRP-L14s at these respective positions (Fig. 1c, C-terminal gap as compared with EcoL2), we conclude that the frameshifts in MRP-L14_{human}1 and CGI-22 are caused by sequencing errors due to the following reasons. (i) The frameshifts occur in a region that is rich in guanosine bases showing a high tendency for sequencing errors. (ii) If MRP-L14_{human}1 (frameshifted representation as compared with the consensus cDNA) and CGI-22 (no frameshift) are compared downstream from this particular position they show no sequence similarity anymore. (iii) However, the annotation of the MRP-L14s to the EcoL2 family of ribosomal proteins depends also on the conserved histidine respective glutamine residue (*boldface type* in Fig. 1c) that is not shown in the CGI-22 sequence. (iv) The CGI-22 protein sequence is elongated C-terminally as compared with the MRP-L14s. The (putatively) translated region covers many repetitive nucleotide stretches, e.g. a T-rich sequence (29 out of 32 nucleotides) translated as stretches of cysteines and phenylalanines (Fig. 1c). Similar stretches are typical for untranslated rather than for translated sequences.

cDNAs 2 and 3 are incomplete in their respective 5'-ends, however, since no ESTs extending further in the 5' direction were found (Figs. 1c and 3). Peptide sequences I, C, and II are shared by all three different cDNAs (Fig. 3). Peptide A of cDNA 1 is replaced in cDNA 2 and 3 by peptide B, which shares only 33% identical amino acid residues with peptide A. Peptide D of cDNAs 1 and 2 is replaced in cDNA 3 by peptide E. Additionally, cDNA 3 contains peptide F, which has no equivalent in cDNAs 1 and 2 (Fig. 3). However, if the human MRP-L14 cDNAs are compared with the corresponding mouse sequence (Fig. 1c), a fourth splicing version consisting of peptides I, A (as in cDNA 1), C, E, and F (as in cDNA 3), and II is found (Fig. 3). The MRP-L14_{rat} sequence follows the mouse version (Fig. 1c). The mature MRP-L14_{bov} contains the counterpart to the human peptide F, as far as it can be concluded by determination of the corresponding peptide (Table I; Fig. 1c).

By screening the data banks for proteins similar to the mammalian MRP-L14s, several bacterial and chloroplast ribosomal proteins of the L2 family were found. Correspondingly, MRP sequences of *Reclinomonas americana*, *Marchantia polymorpha*, *Oryza sativa*, *Acanthamoeba castellanii*, and *Paramecium tetraurelia* were identified. However, these corresponding MRPs are encoded by the respective mitochondrial DNAs. In the *C. elegans* genome data bases, a single contig (R08B6.ctg00286) was identified that contains several stretches of DNA encoding between 11 and 73 amino acid residues that are similar to the MRP-L14_{mammalian} sequences. From yeast, the Rml2p protein was identified, which has recently been proposed to be the mitochondrial counterpart of the bacterial L2 ribosomal proteins (21). The amino acid sequences of the yeast Rml2p and of the EcoL2 ribosomal protein were aligned to the mammalian MRP-L14s (Fig. 1c). Interestingly, the yeast Rml2p is longer than the respective mammalian counterparts. Several insertions and a C-terminal extension show that the functional similarities of the yeast and mammalian EcoL2 counterparts may be not as high as the conservation among eubacterial or eukaryotic cytoplasmic ribosomal proteins of the L2 family suggests (21). MRP-L14_{mouse} and Rml2p show a sequence identity of 31.7% over 207 amino acid residues only. This is pinpointed by the result that a specific histidine residue (*boldface H* in Fig. 1c) is *not* conserved among mam-

malian MRPs that is highly conserved in human cytoplasmic, yeast mitochondrial, and bacterial L2 class proteins (21, 22). In mammalian MRPs, this His is replaced by a glutamine in rat and mouse (Fig. 1c). Also, the insertions of 6 and 13 amino acid residues N-terminally of the conserved His residue in yeast Rml2p and EcoL2, respectively, are not conserved in mammalian MRP-L14s (Fig. 1c).

MRP-L26—MRP-L26_{bov} was not digested by trypsin prior to sequencing, but the mature protein was subjected to N-terminal peptide sequencing. A sequence of 20 amino acid residues was obtained (Table I). Several ESTs were identified of human, mouse, and rat origin, respectively, using this peptide information as virtual screening probe. For MRP-L26_{human}, a complete consensus cDNA of 754 nucleotides was assembled (Table II). An ORF of 527 nucleotides was identified encoding a basic protein of 176 amino acid residues (Fig. 1d). No further 5' extending ESTs were identified. To determine whether the EST-derived sequence represents the true mature form of the MRP-L26_{human} cDNA, we performed a 5'-rapid amplification of cDNA ends cloning the corresponding cDNA from a fetal cDNA library and sequenced the obtained cDNA at its 5' terminus (data not shown). The obtained sequence perfectly matched the 5'-end of the consensus cDNA derived from assembled ESTs; however, no additional sequence information extending the EST sequences in the 5' direction was obtained for MRP-L26_{human} cDNA. Therefore, we postulate the primary ATG codon at bp 21–23 to be the natural start codon for translation, since it fits the consensus site for eukaryotic translational start (19). Compared with the N terminus of the mature MRP-L26_{bov} determined by amino acid sequencing, we postulated the first 8 amino acids of the deduced MRP-L26_{human} sequence to serve as a MISP. This MISP does not belong to one of the specified classes of MISPs according to Ref. 20, R-2, R-3, or R-10, respectively. The postulated mature MRP-L26_{human} has a calculated MM of 19,391 Da and a pI of 10.8. The former value is in very good correspondence to the MM of the corresponding mature MRP-L26_{bov} of 18.8 kDa, which has been determined by SDS-PAGE (Table I). The calculated pI is consistent with the classification of MRP-L26 as the most basic protein of its size class in accordance with the result that MRPs of different mammals are very similar in their pI values (18). In the 3'-end of the assembled consensus cDNA, two adjacent polyadenylation signals, AATAGATAA, are found in positions 707–717 (overlapping nucleotides are underlined). From nucleotide 734 up to the 3'-end, a stretch of 21 consecutive Ala residues is found, which is considered to represent the mature poly(A) tail of the MRP-L26_{human} cDNA.

For MRP-L26_{mouse}, a consensus cDNA of 1124 nucleotides was assembled, encoding a protein of 177 amino acids (Table II). By comparison with incompletely spliced EST sequences, an intron was localized at the bp 580/581 boundary. The cDNA shows no polyadenylation site nor a 3' poly(A) track. Thus, compared with the shorter MRP-L26_{human} cDNA, we suspect putative intron sequences to be present in the 3'-end of the presented sequence. MRP-L26 of human and mouse are quite similar in sequence, although the MRP-L26_{mouse} contains a single amino acid insertion compared with the MRP-L26_{human} (Fig. 1d). We postulate a MISP of 8 amino acid residues for the same reasons as for MRP-L26_{human}. Additionally, the MRP-L26_{mouse} start codon is preceded by an in-frame TGA stop codon (bp 86–88), making the postulated translational start highly probable. For none of the determined N termini of mammalian MRP-L26s does the SignalP program (17) predict a cleavage site within the first 50 amino acid residues of the deduced ORFs. This feature characterizes the postulated MISPs to be rather unusual. However, the short sequences

contain a repetitive alanine motif, which has been postulated to be a common motif of mammalian MISPs for MRPs recently.³ The mature MRP-L26_{mouse} has a calculated MM of 19,435 Da and a pI of 10.53. These values are very similar to those of MRP-L26_{human}.

For MRP-L26_{rat} only fragmentary cDNA information was obtained (Table II). The first cDNA fragment shows an ORF of 98 amino acids very similar to the deduced N termini of the mouse and human MRP-L26 proteins (Fig. 1*d*). Beyond bp position 308 of this cDNA fragment, an intron sequence begins such that the continuing ORF no longer fits any more to the respective mouse and human sequences (see also Fig. 1*d*). A second cDNA fragment encodes the C-terminal 53 amino acid residues of the MRP-L26_{rat} (Table II, Fig. 1*d*). No polyadenylation signal nor poly(A) track are found in this EST fragment. However, the encoded peptide sequences of the fragmentary MRP-L26_{rat} cDNA sequences are very similar to the respective human and mouse sequences (Fig. 1*d*). Thus, the mammalian MRP-L26s can be characterized by these consensus cDNA sequences.

The MRP-L26_{bov} protein was identified as a putative early assembly protein by virtue of its topography, essentially buried in the subribosomal particle (10), and its strong RNA binding properties (11). These data are consistent with the identification of MRP-L26 as the counterpart of the *E. coli* EcoL17 ribosomal protein. EcoL17 has been described as an early assembly protein of the large subunit that is not involved in the catalytic function of the ribosome (23). Additionally, the yeast MRP YmL8 has been identified to be the corresponding yeast protein (24). The alignment of the MRPs with the *E. coli* EcoL17 is shown in Fig. 1*d*. YmL8 is an essential protein of the yeast mitoribosome. Gene disruption of the *MRP-L8* gene not only renders the mutant unable to grow on nonfermentable carbon sources but also hampers the growth on fermentable carbon sources (24). However, the overall similarities between EcoL17, YmL8, and the mammalian MRP-L26s are poor (Fig. 1*d*). MRP-L26_{human} and YmL8 share only 31.5% identical amino acid residues over a length of 121 amino acids. YmL8 contains an insert of several amino acid residues and a C-terminal elongation compared with the other ribosomal proteins of this family. These relatively weak similarities between yeast and mammalian MRPs coincide with the generally poor sequence conservation observed between MRPs of mammalian and yeast origin (8).² Similarly, in the *C. elegans* genomic DNA (Y54E10.Contig159), five separate DNA stretches are found encoding peptides between 24 and 58 amino acid residues in length that show sequence identities with the mammalian MRP-L26s between 29 and 41%. We suppose the scattering of these corresponding *C. elegans* DNA sequences to represent the exon/intron structure of the orthologous MRP-L26_{*C. elegans*} gene.

Altogether, four different classes of mammalian MRPs have been characterized on the molecular level. Three of them show significant sequence similarities to eubacterial and yeast mitochondrial ribosomal proteins. However, the conservation between yeast and mammalian MRPs is significantly lower than between cytoplasmic ribosomal proteins of the same species.

DISCUSSION

Characterization and Evolution of Mammalian MRPs—Mammalian MRPs have been characterized on the molecular level only by chance in the past (4–7). Recently, by combining protein purification and peptide sequencing efforts, genome project results and computational analysis, several mammalian MRP cDNA and protein sequences have become available (8).^{2,3} In general, mammalian MRPs of different origin are quite similar to each other in properties like amino acid sequence, MM, and pI. MM of corresponding MRPs calculated

from deduced amino acid sequences are similar to those determined by SDS-PAGE (2). However, corresponding MRPs of mammals and yeast are much less conserved than cytoplasmic ribosomal proteins of the same species. Cytoplasmic ribosomal proteins of yeast and rat show sequence identities ranging between 40 and 80% (25). Yeast and mammalian MRPs are much less conserved, between 31 and 38% identity (this work). This corresponds with the values obtained for other MRPs conserved among yeast and mammals (8).² Additionally, several mammalian MRPs have been identified that have no counterpart in the cytoplasmic ribosome (8).^{2,3} This result suggests that the evolutionary pressure on both the cytoplasmic and the mitochondrial protein biosynthetic apparatus must have been different during the adaptation of the premitochondrial endosymbiont to life within the hosting cell (24).

Structure and Function of MRPs—Three out of four mammalian MRPs described in this paper show significant sequence similarities to yeast mitochondrial and bacterial ribosomal proteins. Thus, similar functions of these MRPs, compared with the bacterial ribosomal proteins, can be inferred. MRP-L7_{bov} has been characterized as a strong RNA binding protein that is tightly bound to the ribosomal core particle under 4 M LiCl washing conditions (11). The protein is deeply buried in the 39 S subunit, and it is assumed to be involved in early assembly processes (11). EcoL15, the bacterial counterpart of the mammalian MRP-L7s, is a main assembly protein of the large ribosomal subunit, but it is not a true early assembly protein (18). However, the assembly function of MRPs may be different from that of their bacterial counterparts due to the higher protein content of mammalian mitochondrial ribosomes as compared with *E. coli* ribosomes (2). This might be true for the MRP-L26 protein too. MRP-L5_{bovine} was identified as a buried protein of mammalian mitochondrial ribosomes that can be stripped from the large ribosomal subunit only by incubation with 4 M LiCl (26). However, it is present in the large mitochondrial subunit in stoichiometric amounts (2). Since this protein lacks a bacterial or yeast mitochondrial counterpart, no further functional description can be given at this time.

MISPs of MRPs—Only in the minority of MRPs analyzed in this paper, a doubtless identification of MISPs was performed. Only in the case of MRP-L26_{bov}, a mature MRP was N-terminal sequenced, revealing the mature N terminus of the processed protein. However, the MISPs postulated by comparison of the mature N terminus of MRP-L26_{bov} with the deduced ORFs of human, mouse, and rat are unusually short (8 amino acid residues) and were not recognized by the SignalP prediction program (17). However, for other mammalian MRPs MISPs of comparable lengths have been determined.⁵ In yeast MISPs for MRPs of similar length have been detected repetitively (3). Therefore, we conclude these postulated MISPs to be the true import signals of the mammalian MRP-L26s.

In the cases of the three other groups of MRPs, only the computer predictions for MISPs can be evaluated. For MRP-L14_{human}1 and CGI-22, cleavage sites between amino acid residues 22 (Ala) and 23 (Pro) and amino acids 21 (Ala) and 22 (Ala) were predicted, respectively. For MRP-L5_{human} a MISP is predicted by the computer but not for the corresponding MRP-L5_{mouse}. No predictions were made for the mammalian MRP-L7s. In the latter case, the existence of a peptide sequence derived from the mature MRP-L7_{bov} close to the N termini of the mammalian MRP-L7s implies that putative MISPs do not extend further in the C-terminal direction than the N-terminal Arg residue of this peptide (Fig. 1*b*). However, the import properties of the MRP-L7s still to be elucidated seemed to be

rather unusual. Nevertheless, unusual MISPs have been also determined for other mammalian MRPs.⁵

It is astonishing to see that mammalian MRPs, whose yeast counterparts do not possess an N-terminal cleavable MISP, always have adopted additional N-terminal cleavable MISP sequences (YmL8/MRP-L26_{mammalian} (this work), YmL33/MRP-L28_{mammalian} (8), YmL38/MRP-L32_{mammalian} (8)). In the cases of the yeast MRPs without cleavable N-terminal MISPs, mitochondrial import signals can be inferred to reside in the interior of the protein sequence. YmL8, the yeast mitochondrial counterpart of mammalian MRP-L26, is transported into the mitochondria without cleavage of an N-terminal MISP (27). The signal sequence for mitochondrial import is localized internally in the middle of the protein. Interestingly, the N terminus of YmL8 (unprocessed) corresponds very well to the N termini of the mature mammalian MRP-L26s (Fig. 1d). The same phenomenon is observed in the case of YmL38/MRP-L32 (8), too. These findings suggest that mitochondrial import signaling by internal peptide sequences is either (i) restricted to yeast MRPs and/or (ii) the mammalian mitochondrial protein import mechanisms are strongly different from that of yeast or (iii) that mammalian MRPs have lost the ability to be imported by internal peptide signaling thus, which had to be compensated by the adoption of additional N-terminal MISPs. It should be noted at this point that mammalian MRPs seem to possess a mitochondrial import signaling pathway that is more strongly conserved than that of yeast MRPs that are imported by different import mechanisms. One hint is the conservation of a small alanine-rich sequence motif in MISPs of mammalian MRPs.⁵

Splice Variants and Functional Significance of MRP-L14—Splice variants and gene families for cytoplasmic ribosomal proteins are found in yeast as well as in mammals. MRP genes were considered to be present only once in the haploid genome in contrast to cytoplasmic ribosomal protein genes. Splicing itself occurs very rarely on yeast MRP pre-mRNAs, and no splice variants have been reported from yeast (3). However, this seems not to be true in mammals. For MRP-L28_{human} (8), two different cDNA sequences were assembled from the EST data banks. One cDNA (MRP-L28_{human}2) could be found in fetal liver spleen only, whereas the other version (MRP-L28_{human}1) is omnipresent during development and in different tissues (8). For MRP-L14, four different splice variants were determined from mouse and human (this work). The different MRP-L14_{human} cDNAs could not be assigned to a specific tissue or developmental stage. The exception is that cDNAs 2 and 3 (Fig. 2) derive from malignant tissues only. However, the occurrence of splice variants sheds light on a possible modification of the important function of mammalian MRP-L14. The bacterial and cytoplasmic counterparts of MRP-L14 are involved in the peptidyltransferase function of the ribosome and can replace each other in functional ribosomes (22). The functional relevance of individual amino acid residues has been shown (21, 22). The yeast mitochondrial counterpart of mam-

malian MRP-L14, Rml2p, is essential for mitochondrial function (21). Substitution of Rml2p His³⁴³ to Gln causes a conditional respiratory growth phenotype without detectable effect on the assembly or stability of the large mitoribosomal subunit (21). The corresponding His²²⁹ of EcoL2 has been shown to be involved in the peptidyltransferase activity of the ribosome (9, 22). The replacement of EcoL2 His²²⁹ by Gln does not affect the assembly of the ribosomes but abolishes completely the peptidyltransferase activity (9). Similar results have been reported for the substitution of His²²⁹ by glycine, arginine, and alanine (22). The authors claim that His³⁴³ is highly conserved throughout evolution to preserve this function. However, in the corresponding mammalian MRP-L14 proteins the corresponding His residue is not strictly conserved (Fig. 1c). In mouse and rat MRP-L14s, this His is substituted by Gln (Fig. 1c). Additionally, the N-terminal part of the 14-mer peptide sequence around Rml2p His³⁴³, which has been postulated to be very important for the L2 ribosomal protein function (21), is completely missing in all mammalian MRP-L14 sequences (Fig. 1c). Thus, the functional roles of His or Gln at this respective position as well as the conserved 14-mer amino acid sequence remain questionable.

REFERENCES

1. Myers, A. M., Pape, L. K., and Tzagoloff, A. (1985) *EMBO J.* **4**, 2087–2092
2. Matthews, D. E., Hessler, R. A., Denslow, N. D., Edwards, J. S., and O'Brien, T. W. (1982) *J. Biol. Chem.* **257**, 8788–8794
3. Graack, H.-R., and Wittmann-Liebold, B. (1998) *Biochem. J.* **329**, 433–448
4. Ou, J.-H., Yen, T. S. B., Wang, Y.-F., Kam, W. K., and Rutter, W. J. (1987) *Nucleic Acids Res.* **15**, 8919–8934
5. Shah, Z. H., O'Dell, K. M., Miller, S. C., An, X., and Jacobs, H. T. (1997) *Gene (Amst.)* **204**, 55–62
6. Tsang, P., Gilles, F., Yuan, L., Kuo, Y. H., Lupu, F., Samara, G., Moosikasawan, J., Goye, A., Zelenetz, A. D., and Tycko, B. (1995) *Hum. Mol. Genet.* **4**, 1499–1507
7. Marty, L., and Fort, P. (1996) *J. Biol. Chem.* **271**, 11468–11476
8. Goldschmidt-Reisin, S., Kitakawa, M., Herfurth, E., Wittmann-Liebold, B., Grohmann, L., and Graack, H.-R. (1998) *J. Biol. Chem.* **273**, 34828–34836
9. Cooperman, B. S., Wooten, T., Romero, D. P., and Traut, R. R. (1995) *Biochem. Cell Biol.* **73**, 1087–1094
10. Denslow, N. D., and O'Brien, T. W. (1984) *J. Biol. Chem.* **259**, 9867–9873
11. Piatyszek, M. A., Denslow, N. D., and O'Brien, T. W. (1988) *Nucleic Acids Res.* **16**, 2565–2583
12. O'Brien, T. W., Denslow, N. D., Harville, T. O., Hessler, R. A., and Matthews, D. E. (1980) in *The Organization and Expression of the Mitochondrial Genome* (Kroon, A. and Saccone, C., eds) pp. 301–305, Elsevier/North-Holland Biomedical Press, Amsterdam, The Netherlands
13. Denslow, N. D., Anders, J. C., and O'Brien, T. W. (1991) *J. Biol. Chem.* **266**, 9586–9590
14. Otto, A., Thiede, B., Müller, E.-C., Scheler, C., Wittmann-Liebold, B., and Jungblut, P. (1996) *Electrophoresis* **17**, 1643–1650
15. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389–3402
16. Genetics Computer Group (1997) *GCG Wisconsin Package*, version 9.1-Open-VMS, Genetics Computer Group, Madison, WI
17. Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997) *Protein Eng.* **10**, 1–6
18. Pietromonaco, S. F., Hessler, R. A., and O'Brien, T. W. (1986) *J. Mol. Evol.* **24**, 110–117
19. Kozak, M. (1981) *Nucleic Acids Res.* **9**, 5233–5252
20. Branda, S. S., and Isaya, G. (1995) *J. Biol. Chem.* **270**, 27366–27373
21. Pan, C., and Mason, T. L. (1997) *J. Biol. Chem.* **272**, 8165–8171
22. Ühlein, M., Weglöhner, W., Urlaub, H., and Wittmann-Liebold, B. (1998) *Biochem. J.* **331**, 423–430
23. Herold, M., and Nierhaus, K. H. (1987) *J. Biol. Chem.* **262**, 8826–8833
24. Kitakawa, M., Grohmann, L., Graack, H.-R., and Isono, K. (1990) *Nucleic Acids Res.* **18**, 1521–1529
25. Wool, I. G., Chan, Y.-L., and Glück, A. (1995) *Biochem. Cell Biol.* **73**, 933–947
26. Schieber, G. L., and O'Brien, T. W. (1982) *J. Biol. Chem.* **257**, 8781–8787
27. Matsushita, Y., and Isono, K. (1993) *Eur. J. Biochem.* **214**, 577–585

⁵ T. W. O'Brien, J. Liu, E. Sylvester, E. B. Mougey, N. Fischel-Ghodsian, B. Thiede, B. Wittmann-Liebold, and H.-R. Graack, submitted for publication.