Mammalian Mitochondrial Ribosomal Proteins

N-TERMINAL AMINO ACID SEQUENCING, CHARACTERIZATION, AND IDENTIFICATION OF CORRESPONDING GENE SEQUENCES

Sabine Goldschmidt-Reisini, Madoka Kitakawa, Elrike Herfurth, Brigitte Wittmann-Liebold, Lutz Grohmann, and Hanns-Rüdiger Graack

From the Max-Planck-Institute for Molecular Genetics, Ihnestrasse 63, D-14195 Berlin, Germany and the Institute for Genetics, Free University of Berlin, Arnimallee 7, D-14195 Berlin, Germany

The integrity of healthy mitochondria is supposed to depend largely on proper mitochondrial protein biogenesis. Mitochondrial ribosomal proteins (MRPs) are directly involved in this process. To identify mammalian mitochondrial ribosomal proteins and their corresponding genes, we purified mature rat MRPs and determined 12 different N-terminal amino acid sequences. Using this peptide information, data banks were screened for corresponding DNA sequences to identify the genes or to establish consensus cDNAs and to characterize the deduced MRP open reading frames. Eight different groups of corresponding mammalian MRPs constituted from human, mouse, and rat origin were identified. Five of them show significant sequence similarities to bacterial and/or yeast mitochondrial ribosomal proteins. However, MRPs are much less conserved in respect to the amino acid sequence among species than cytoplasmic ribosomal proteins of eukaryotes and bacteria.

Intact mitochondrial protein biosynthesis has been shown to be indispensable for the maintenance of mitochondrial DNA in yeast (1). Nearly all of the mitochondrial ribosomal proteins (MRPs) investigated so far are essential for proper mt protein synthesis (2). Knock-out mutants of yeast MRP genes lose their respiratory capacity and change to rho- or rho+ mt genetic status by successive losses of mt DNA (1). In higher eukaryotes, the knowledge about comparable functions of MRPs is only rudimentary, since only a few MRPs have been characterized on the molecular level. The protein composition of mammalian mitochondrial ribosomes has been studied extensively (3–5). Some properties of mt ribosomes such as structure (6), binding of nucleotides and RNA (7–11), and interaction with different factors have been studied (12–14). However, only 3 of the approximately 80–100 different human MRPs have been described at the molecular level so far. MRL3, which is the EcoL3 counterpart in human mt ribosomes, was identified as an overexpressed r-protein in Mahlavu hepatocytic cells (15). Later, it was postulated to be a true MRP by virtue of its sequence similarity to the corresponding yeast MRP YmL9 (16). MRP12 was identified as a delayed-early response gene similar in sequence to the Escherichia coli L7/L12 r-protein (17). The metazoan mitochondrial counterpart of Ecos12 has been characterized in Drosophila, human, and mouse (18, 19). In Drosophila a mutation of mt S12 causes abnormal behavior. This is the first case reported so far of affection of the status of an animal by an MRP mutation (18). Diseases affecting mitochondria are known in humans, and are caused by nuclear mutations responsible for the loss of mt DNA as a secondary effect by a so far unknown mechanism (20, 21). Mutations of MRPs are good candidates affecting mt genetic and/or physiological status. To characterize mammalian MRPs and to compare their biochemical properties with those of their (essential) counterparts, e.g., of yeasts, we identified several mammalian MRPs and their corresponding gene sequences. We used N-terminal sequence information obtained from purified rat MRPs to screen DNA data banks and to characterize identified MRPs of rat, human, and mouse.

EXPERIMENTAL PROCEDURES

Determination of Rat N-terminal MRP Peptide Sequences—Preparation of mitochondrial ribosomal proteins from rat liver was accomplished according to Ref. 3. Proteins were extracted with acetic acid and lyophilized (22), and were separated by two-dimensional PAGE as described (23). Proteins from the second dimension gels were transferred to polyvinylidine difluoride membranes by Western blotting (24). N-terminal sequencing of blotted proteins was done as described (24).

Computing—Sequence searches were performed with the BCM Search Launcher program (25). First, the rat MRP N-terminal peptide sequences were compared with EST sequences with the “general protein sequence/pattern searches” and the “TBLASTN/blast query versus 6-frame translation of dbest with Entrez & SRS links (NCBI/BMC)” subprogram. Nucleic acid sequences obtained were selected for overlapping and extension, and a consensus sequence was established using the “pileup” program (26). The putative ORF (open reading frame) was localized using the “translate” program (26). The resulting peptide sequence was aligned to the rat N-terminal peptide sequence by the “bestfit” program (26). In the case of ORFs open at their 5’ and/or 3’ end, the first and the last 50 nucleic acids, respectively, of the established consensus sequence were subjected to a second search using the “nucleic acid sequence search” (25). The search was performed with the “BLAST/N/blast with Repeat Masker and Entrez & SRS links (NCBI/BMC)” program (27). The abbreviations used are: MRP, mitochondrial ribosomal protein; EST, expressed sequence tag; MISP, mitochondrial import signal peptide; mt, mitochondrial; ORF, open reading frame; PAGE, polyacrylamide gel electrophoresis; r-protein, ribosomal protein.

* 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 N-terminal peptide sequences reported in this paper have been deposited in the MRPS Data Base with accession nos. S78411 (MRP-L8rat), S78412 (MRP-L22rat), S78413 (MRP-L24rat), S78414 (MRP-L25rat), S78415 (MRP-L27rat), S78416 (MRP-L28rat), S78417 (MRP-L32rat), S78418 (MRP-L33rat), S78419 (MRP-L34rat), S78420 (MRP-L35rat), S78421 (MRP-S13rat), and S78422 (MRP-S20rat), respectively.

† Present address: Dept. of Biochemistry, Tel Aviv University, 69978 Ramat-Aviv, Tel Aviv, Israel.
‡ Present address: Faculty of Science, Dept. of Biology, University of Kobe, Nada-Ku, 657 Kobe, Japan.
¶ Present address: Max-Delbrück-Center for Molecular Medicine, Robert-Rösle-Strasse 10, D-13125 Berlin-Buch, Germany.
† Present address: BioInside GmbH, Wartheallee 21, D-14513 Tempel, Germany.
‡‡ To whom correspondence should be addressed. Tel.: 49-30-838-2629; Fax: 49-30-838-3649; E-mail: graack@zedat-fu-berlin.de.

Received for publication, June 10, 1998, and in revised form, September 15, 1998.

43828 This paper is available on line at http://www.jbc.org
RESULTS

N-terminal Sequencing of Purified Rat MRPs and Computational Analysis—Several rat MRPs were purified from mt r-protein mixtures. The proteins were identified according to the two-dimensional map of rat MRPs (27), and their isoelectric points were determined (Table I). The proteins were blotted from two-dimensional gels onto polyvinylidene difluoride membranes and N-terminally sequenced. Twelve different N-terminal sequences were obtained (Table I). The sequences of MRPL22 rat and MRPL24 rat are almost identical, and thus they are the only ones to be picked up by the computer. Accordingly, we assign the mammalian MRP-L8 proteins as members of the L10 family of r-proteins.

MRP-L8—in the primary search using the obtained MRPL8rat peptide sequence as screening probe, 12 and 5 “primary hits” of mouse and human ESTs were found, respectively. The consensus cDNA sequence of MRPL8mouse was assembled by multiple searches from many different EST sequences (Table II). A complete ORF of 261 amino acid residues was identified. Positions 28–46 of this ORF correspond to the MRP-L8rat N-terminal peptide (Table I, Fig. 1a, present report). A cleavable mitochondrial import signal peptide (MISP) of 27 amino acid residues (MISP) of 27 amino acid residues was deduced from this cDNA, for which the N-terminal sequence of the mature MRP-L22rat was used as target sequence (Table II). The incomplete ORF deduced for MRP-L22rat and MRP-L24rat are almost identical, and thus they are the only ones to be picked up by the computer. Accordingly, we assign the mammalian MRP-L8 proteins as members of the L10 family of r-proteins.

MRP-L22/24—Fourteen and 14 primary hits for MRP-L22human and MRP-L22mouse were found, respectively. Additionally, a single EST sequence of rat origin was identified (Table II). The incomplete ORF deduced for MRP-L22rat matches the N-terminal sequence of the mature MRP-L22rat from positions 42 to 70 (Fig. 1b, present report). An N-terminal MISP of 41 amino acids is postulated for MRP-L22rat. For MRP-L22mouse, a complete cDNA consensus sequence of 826 base pairs was determined (Table II). An ORF of 201 amino acid residues was deduced from this cDNA, for which the N-terminal 80 amino acid residues are almost identical to the MRP-L22rat protein. Further, MRP-L22mouse shows approximately 80% identity to the deduced MRP-L22human sequence over its entire length (Fig. 1b; see below). An MISP of 41 amino acid residues is postulated for MRP-L22mouse. The ORF of MRP-L22human was deduced from a consensus cDNA of 775 base pairs (Table II). MRP-L22human consists of 206 amino acid residues, which are highly conserved as compared with the rat and mouse MRP-L22 sequences (Fig. 1b, present report). By comparison with the rat mature peptide sequence, an MISP of sequence identity to each other over their entire length. An MISP of 28 amino acid residues is postulated in comparison to the rat N-terminal L8 peptide and in accordance with (28, 29). Both the mouse and the human putative MISPs are highly hydrophobic; they contain only positively charged and no negatively charged amino acid residues, and very few hydroxylated amino acid residues. The arginine (R) residues in position −2 (MRP-L8mouse) and −10 and −2 (MRP-L8human) classify the peptides as R−2 (MRP-L8mouse) and R−10/R−2 (MRP-L8human), respectively, according to Ref. 29.

Significant sequence similarities of the mammalian L8 MRPs were detected by computer search with the yeast MRPL8 and bacterial r-proteins of the L10 family (Fig. 1a, Table III). However, the percentages of sequence similarities and sequence identities are low (Table III) and might be below the threshold. The family of the bacterial L10 r-proteins itself is very heterogeneous, and a comparison of the citrus greening disease-associated bacterial L10 with the E. coli L10 r-protein shows only 45% similarity and 32.4% identity over a stretch of 153 amino acid residues, respectively. These values are very low, but nonetheless the membership of similar proteins from different bacterial species to the L10 r-protein family is supported, e.g. by the similar location of the corresponding genes within the same operons, respectively. However, the similarities of the mammalian MRP-L8s to the L10 r-proteins class are the only ones to be picked up by the computer. Accordingly, we assign the mammalian MRP-L8 proteins as members of the L10 family of r-proteins.

| Protein name | IP |
|--------------|----|
| MRP-L8rat    | 8.6|
| MRP-L22rat   | 7.7|
| MRP-L24rat   | 8.2|
| MRP-L25rat   | 9.0|
| MRP-L27rat   | 8.9|
| MRP-L28rat   | 9.1|
| MRP-L31rat   | 8.3|
| MRP-L32rat   | 9.4|
| MRP-L33rat   | 9.2|
| MRP-L41rat   | 9.6|
| MRP-S13rat   | 8.1|
| MRP-S20rat   | 8.7|

| Protein name | 10 | 20 | 30 | 40 | 50 |
|--------------|----|----|----|----|----|
| MRP-L8rat    |    |    |    |    |    |
| MRP-L22rat   |    |    |    |    |    |
| MRP-L24rat   |    |    |    |    |    |
| MRP-L25rat   |    |    |    |    |    |
| MRP-L27rat   |    |    |    |    |    |
| MRP-L28rat   |    |    |    |    |    |
| MRP-L31rat   |    |    |    |    |    |
| MRP-L32rat   |    |    |    |    |    |
| MRP-L33rat   |    |    |    |    |    |
| MRP-L41rat   |    |    |    |    |    |
| MRP-S13rat   |    |    |    |    |    |
| MRP-S20rat   |    |    |    |    |    |

Downloaded from http://www.jbc.org/ by guest on July 22, 2018
FIG. 1. Alignment of deduced mammalian MRP amino acid sequences with rat MRP N-terminal peptide sequences. *Numbers give the respective amino acid position. Lines (−) mark identical amino acid positions, colons (:) mark strongly conserved amino acid residues, and points (.) 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 written in lowercase letters are uncertain in their determination by amino acid sequencing (see Table I). x marks unidentified amino acids although the positional number is valid. a, alignment of mammalian MRP-L8 protein sequences with the citrus greening disease-associated bacterium L10 (accession no. M94319, rpIJ gene). b, alignment of mammalian MRP-L22 protein sequences. c, alignment of MRP-L25 protein sequences with the *E. coli* EcoL22 sequence (accession no. sw:rl22_ecoli). Section marks (§) in the alignment of MRP-L25 human and EcoL22 mark amino acid residues that are absolutely conserved among nearly all members of the EcoL22 r-protein family. d, alignment of mammalian MRP-L27 protein sequences. e, alignment of the mammalian MRP-L28 sequences with the yeast YmL33 sequence (accession no. D90217), and the *E. coli* EcoL30 sequence (accession no. sw:rl30_ecoli). f, alignment of MRP-L31 sequences. g, alignment of the mammalian MRP-L32 sequences with the *E. coli* EcoL14 sequence (accession no. sw:rl14_ecoli), and the yeast YmL38 sequence (accession no. SS8000). h, alignment of the mammalian MRP-S13 sequences.
46 amino acid residues is postulated for MRP-L22 human. According to Ref. 29, the putative MISPs of rat, mouse, and human MRP-L22 belong to the R-none class of MISPs. All three show the typical properties of MISPs such as many hydrophobic, positively charged, and hydroxylated amino acid residues. In addition, genomic sequences were identified for MRP-L22 human in the data banks. The gene for MRP-L22 human was located on chromosome 22q11 cosmid clone 102 g9 (accession...
| Gene          | bp nos. | EST accession no. | Total length | ORF from bp no. to no. | r-Protein family |
|--------------|---------|------------------|--------------|------------------------|-----------------|
| MRP-L8mouse | 1–30    | AA011755         | 1065         | 19–804                 | L10             |
|              | 11–21   | AA008842         |              |                        |                 |
|              | 41–115  | AA011755         |              |                        |                 |
|              | 100–109 | AA008842         |              |                        |                 |
|              | 125–495 | AA011755         |              |                        |                 |
|              | 149–253 | AA060996         |              |                        |                 |
|              | 105–204 | AA895789         |              |                        |                 |
|              | 107–469 | AA607351         |              |                        |                 |
|              | 2–473   | R56369           | 1195         | 3–788                  | L10             |
|              | 181–350 | AA316625         |              |                        |                 |
|              | 90–260  | AA325435         |              |                        |                 |
|              | 2–473   | R80333           |              |                        |                 |
| MRP-L8human | 1–382   | AA011755         | 1065         | 19–804                 | L10             |
|              | 11–21   | AA008842         |              |                        |                 |
|              | 41–115  | AA011755         |              |                        |                 |
|              | 100–109 | AA008842         |              |                        |                 |
|              | 125–495 | AA011755         |              |                        |                 |
|              | 149–253 | AA060996         |              |                        |                 |
|              | 105–204 | AA895789         |              |                        |                 |
|              | 107–469 | AA607351         |              |                        |                 |
|              | 2–473   | R56369           | 1195         | 3–788                  | L10             |
|              | 181–350 | AA316625         |              |                        |                 |
|              | 90–260  | AA325435         |              |                        |                 |
| MRP-L22rat   | 1–296   | H31290           | 296          | 6–248*                 | New             |
| MRP-L22mouse | 1–150   | AA240860         | 826          | 127–732                | New             |
|              | 36–450  | AA238619         |              |                        |                 |
|              | 428–550 | AA237833         |              |                        |                 |
|              | 212–349 | AA438136         |              |                        |                 |
| MRP-L22human | 1–25    | W49836           | 775          | 25–645                 | New             |
|              | 1–506   | AA307896         |              |                        |                 |
|              | 244–488 | W95859           |              |                        |                 |
| MRP-L25rat   | 1–279   | H313219          | 279          | 6–279*                 | L22             |
| MRP-L25mouse | 1–70    | AA522192         | 644          | 8–644*                 | L22             |
|              | 54–72   | AA543344         |              |                        |                 |
|              | 90–249  | AA522192         |              |                        |                 |
|              | 237–423 | AA058062         |              |                        |                 |
|              | 419–541 | AA543344         |              |                        |                 |
|              | 559–642 | AA522192         |              |                        |                 |
| MRP-L25human | 1–359   | H87659           | 824          | 5–676                  | L22             |
|              | 230–464 | AA101598         |              |                        |                 |
|              | 231–462 | N47515 r,c       |              |                        |                 |
| MRP-L27rat   | 11–489  | W61748           | 691          | 150–557                | YmL27           |
| MRP-L27mouse | 493–622 | AA230667         |              |                        |                 |
|              | 221–322 | AA612051         |              |                        |                 |
| MRP-L27human | 1–213   | A056473          | 583          | 106–519                | YmL27           |
|              | 10–260  | AA431468 r,c     |              |                        |                 |
|              | 447–462 | W07096           |              |                        |                 |
|              | 277–379 | AA431468 r,c     |              |                        |                 |
| MRP-L28rat   | 1–189   | H33611           | 189          | 2–189*                 | L30             |
| MRP-L28mouse | 1–215   | W44996           | 594          | 20–520                 | L30             |
|              | 296–582 | AA217953         |              |                        |                 |
|              | 195–286 | AA286206         |              |                        |                 |
| MRP-L28human | 1–50    | AA005080         | 657          | 81–566                 | L30             |
|              | 17–175  | N87046           |              |                        |                 |
|              | 181–409 | AA282351         |              |                        |                 |
|              | 479–657 | AA772463 r,c     |              |                        |                 |
| MRP-L28mouse2| 1–520   | AA703185         | 520          | 254–505*               | L30             |
| MRP-L31rat   | 1–505   | AA859601 r,c     | 505          | 122–505*               | New             |
| MRP-L31mouse | 11–244  | AA270023         | 647          | 92–520                 | New             |
|              | 88–500  | AA592070         |              |                        |                 |
|              | 25–450  | W25509           |              |                        |                 |
| MRP-L31human | 1–50    | AA592070         | 566          | 115–501                | New             |
|              | 27–542  | N46796           |              |                        |                 |
| MRP-L32rat   | 1–443   | AA009076 r,c     | 443          | 3–352*                 | L14             |
| MRP-L32mouse | 1–100   | AA245437         | 981          | 71–508                 | L14             |
|              | 100–450 | W24195           |              |                        |                 |
|              | 449–509 | AA245437         |              |                        |                 |
|              | 478–848 | W13027           |              |                        |                 |
| MRP-L32human | 1–190   | AA434296 r,c     | 780          | 76–615                 | L14             |
|              | 20–108  | C02170           |              |                        |                 |
|              | 279–288 | AA434296 r,c     |              |                        |                 |
|              | 119–168 | C02170           |              |                        |                 |
|              | 218–439 | AA128346         |              |                        |                 |
|              | 308–531 | AA639660 r,c     |              |                        |                 |
| MRP-S13rat   | 1–480   | A1011965         | 886          | 19–552*                | New             |
|              | 212–250 | H31115           |              |                        |                 |
|              | 8–374   | A1045711 r,c     |              |                        |                 |
| MRP-S13mouse | 1–60    | AA104183         | 977          | 5–607                  | New             |
|              | 35–58   | AA073181         |              |                        |                 |
|              | 86–284  | AA104183         |              |                        |                 |
|              | 80–88   | AA245399 r,c     |              |                        |                 |
|              | 195–254 | AA080247         |              |                        |                 |
|              | 149–217 | AA245399         |              |                        |                 |
|              | 323–329 | AA080247         |              |                        |                 |
|              | 225–229 | AA245399         |              |                        |                 |
|              | 207–242 | AA790446         |              |                        |                 |
|              | 197–522 | AA007787 r,c     |              |                        |                 |
Proteins in horizontal lines are compared to proteins in vertical columns. aa, extension of similar sequences in numbers of amino acid residues. %, numbers before the slash (/) give the similarity of the compared amino acid stretches in percentage, numbers after the slash (/) give the identity of the compared amino acid stretches in percentage, as calculated by the “bestfit” computer program (26). Ecol, E. coli r-protein of the large (L) subunit. YmL, yeast MRP of the large (L) subunit. The species of the citrus greening disease-associated bacterium is not further specified in the database (accession no. M94319, rpL28 gene).

| Gene                        | bp nos.     | EST accession no. | Total length | ORF from bp no. to no. | r-Protein family |
|-----------------------------|-------------|-------------------|--------------|------------------------|------------------|
| MRP-S13human                | 1–4498      | AF036329          | 4498         | 3755–3966              | New              |
| YmL27                       | 389–454     | AI057633          | 514          | 1–135                  |                  |
| MRP-L25human                | 195–368     | AA421393          |              |                        |                  |
| MRP-L25mouse                | 273–546     | AA477685          |              |                        |                  |
| MRP-L8mouse                 | 76 aa, 44.4/26.3% |            | 151 aa, 35.8/22% |                  |                  |
| Citrus gr. bact. L10        | 62 aa, 37/22.5% |              |              |                        |                  |
| MRP-L25human                | 84 aa, 42/31.8% |              |              |                        |                  |
| YmL27                       |              |                  |              |                        |                  |
| MRP-L27mouse                | 68 aa, 44.1/33.8% |              |              |                        |                  |
| YmL27                       |              |                  |              |                        |                  |
| MRP-L28mouse                | 52 aa, 48/24.6% |            | 52 aa, 40/21% |                  |                  |
| EcoL30                      | 53 aa, 49/29.9% |              |              |                        |                  |
| MRP-L32mouse                | 58 aa, 46/23.5/1.0% |        | 82 aa, 53.4/38.3% |                  |                  |
| EcoL14                      | 138 aa, 59/48% |              |              |                        |                  |

**TABLE III—continued**

**Sequence comparison of similar ribosomal proteins from mitochondria, and bacteria**

Proteins in horizontal lines are compared to proteins in vertical columns. aa, extension of similar sequences in numbers of amino acid residues. %, numbers before the slash (/) give the similarity of the compared amino acid stretches in percentage, numbers after the slash (/) give the identity of the compared amino acid stretches in percentage, as calculated by the “bestfit” computer program (26). Ecol, E. coli r-protein of the large (L) subunit. YmL, yeast MRP of the large (L) subunit. The species of the citrus greening disease-associated bacterium is not further specified in the database (accession no. M94319, rpL28 gene).

**TABLE III**

Proteins in horizontal lines are compared to proteins in vertical columns. aa, extension of similar sequences in numbers of amino acid residues. %, numbers before the slash (/) give the similarity of the compared amino acid stretches in percentage, numbers after the slash (/) give the identity of the compared amino acid stretches in percentage, as calculated by the “bestfit” computer program (26). Ecol, E. coli r-protein of the large (L) subunit. YmL, yeast MRP of the large (L) subunit. The species of the citrus greening disease-associated bacterium is not further specified in the database (accession no. M94319, rpL28 gene).

**Fig. 2. Genomic map of the exon/intron structure of the MRP-L22human gene.** The genomic DNA is shown as a horizontal line, and restriction sites are represented by vertical lines. Boxes mark location of cDNA sequences; black boxes represent translated cDNA sequence, and open boxes show untranslated cDNA sequences. mRNA was assembled from three different ESTs (Table II). To avoid a frameshift as compared with the deduced amino acid sequence of MRP-L25mouse base pair 225 of EST H87659, which is an “n,” was omitted. The complete ORF deduced spans 223 amino acid residues. A polyadenylation signal AATAAA is found 3’ to the stop codon. MRP-L25human and MRP-L25mouse are closely related proteins except for their respective C termini (Fig. 1c). MISP’s 33 amino acids in length are postulated for all three rat, mouse, and human MRP-L22s. The signal peptides belong to the R-none class according to Ref. 29.

Similarity of the human and mouse proteins to the E. coli L22 r-protein has been postulated (EST accession no. AA101598). More than 30 different L22 r-proteins were picked out from the data bases using human or mouse MRP-L25 amino acid sequence as a screening probe. When MRP-L25human is compared with EcoL22 at the amino acid residue level, a sequence similarity of only 42%, and a sequence identity of 31.8% was detected, covering a region of 84 amino acid residues (Table III). However, when the complete MRP-L25human sequence was compared with different members of the L22 protein family, several amino acids were identified that are identical among all of them (Fig. 1c). Thus, the affiliation of the mammalian MRP-L25 proteins to the EcoL22 r-protein family is confirmed.
son to the mature N terminus of rat MRP-L27 for MRP-L27\textsubscript{human} and MRP-L27\textsubscript{mouse}, MISPs of 13 amino acid residues are postulated, respectively (see Fig. 1d). Both MISPs are highly hydrophobic with few (two in human, one in mouse) positively charged and two hydroxylated (mouse) amino acid residues. Both signal peptides belong to the R-none class of signal peptides according to Ref. 29. The consensus cDNAs of human and mouse show polyadenylation signals AATAAA closely located downstream of the respective stop codons. The nucleotide environments of these signals are highly conserved between the MRP-L27\textsubscript{mouse} and MRP-L27\textsubscript{human} consensus cDNAs.

Sequence comparison revealed weak but significant sequence similarities of the mammalian MRP-L27 to the N-terminal portion of the yeast YmL27 MRP (Fig. 1d, Table III). Although the sequence similarities are not high, they are comparable to the values obtained for other mammalian MRPs similar to bacterial and yeast mitochondrial r-proteins (Table III). The yeast YmL27 shows no sequence similarity to any known r-protein (31), and thus the discovery of the mammalian counterparts of YmL27 defines the first MRP family that is not similar to known r-proteins from other sources.

**MRP-L28**—One, 6, and 48 primary hits were found for rat, human, and mouse MRP-L28, respectively. For MRP-L28\textsubscript{rat}, an incomplete ORF was identified. The last 11 amino acid residues of the ORF deduced are in complete agreement with the first 11 amino acid residues of the mature MRP-L28\textsubscript{rat} protein, as determined by amino acid sequencing (Table I, Fig. 1e, present report). Furthermore, most of this ORF shows strong sequence similarities to the ORFs deduced for MRP-L28\textsubscript{human} and MRP-L28\textsubscript{mouse} respectively (Fig. 1e). Interestingly, among the human MRP-L28 ESTs, two different groups of consensus cDNA sequences were identified and assembled. MRP-L28\textsubscript{human}1 (Table II, Fig. 1e) encodes an ORF of 162 amino acid residues. The consensus cDNA of MRP-L28\textsubscript{human}2, which is supported by four independent ESTs, is not complete (Fig. 1e). Strikingly, it contains a stop codon (*) caused by a missing nucleotide in the cDNA coding for the putative signal peptide (see Fig. 1e). The stop codon as well as the frameshift are found in all of the four detected ESTs. Further, the four ESTs of MRP-L28\textsubscript{human}2 are all of fetal liver spleen origin, whereas the ESTs identified for MRP-L28\textsubscript{human}1 are products of mRNA isolated from different tissues such as retina, melanocyte, fetal liver spleen and fetal heart, and parathyroid tumor. Thus, irrespective of whether the MRP-L28\textsubscript{human}2 mRNA is a product of a pseudogene not being translated \textit{in vivo}, it seems that for MRP-L28\textsubscript{human} at least two different genes do exist that are transcribed in a tissue-specific manner.

The MRP-L28\textsubscript{mouse} consensus cDNA codes for an ORF, which is quite similar in sequence to the MRP-L28\textsubscript{rat}, MRP-L28\textsubscript{human}1, and MRP-L28\textsubscript{human}2, respectively (Fig. 1e). However, within the 5’ part of the consensus cDNA (Table II), a single frameshift is found. This frameshift is supported by six out of six ESTs identified for this region. The frameshift causes an alternative N terminus of MRP-L28\textsubscript{mouse} without a translational start codon, which is not similar to the corresponding N termini of rat and human MRP-L28 (Fig. 1e). If the frameshift is not taken into account, the N terminus of MRP-L28\textsubscript{mouse} corresponds well to the respective rat and human sequences (Fig. 1e). It might be speculated that the apparent frameshift is caused by sequencing errors, or that the MRP-L28\textsubscript{mouse} cDNA is the MRP-L28\textsubscript{human}2 "homologue" rather than the MRP-L28\textsubscript{human}1 homologue. For all MRP-L28s (neglecting the frameshifts), MISPs of 34 amino acid residues are postulated. For MRP-L28\textsubscript{human}1 and MRP-L28\textsubscript{human}2, stop codons 5’ to the translational start support this assignment. In the case of rat and mouse MRP-L28s, there are no stop codons preceding the starting methionine, due to the short 5’ sequences of the respective consensus cDNAs (Table II). The 5’ encoded amino acid sequence is shown in the case of MRP-L28\textsubscript{rat} (Fig. 1e), but is very unlikely to be a true part of the MISP. All four MISPs are characterized by a high proportion of hydrophobic and positively charged amino acid residues. With the exception of MRP-L28\textsubscript{human}2, they belong to the R–2 class of cleavable signal sequences according to Ref. 29. In MRP-L28\textsubscript{human}2, the arginine in position –2 is replaced by a histidine (Fig. 1e).

The MRP-L28\textsubscript{human} and MRP-L28\textsubscript{mouse}1 sequences showed a weak but significant sequence similarity to the yeast YmL33 MRP and the EcoL30 r-protein, respectively. The regions of sequence similarity span the total length of the EcoL30 protein, and 49 amino acid residues of the N terminus of YmL33, and the middle part of MRP-L28\textsubscript{mouse} respectively (Fig. 1e, Table I). Interestingly, the overall size of the yeast MRP is considerably less than that of the mammalian MRPs. The EcoL30 protein itself is only two-third the size of the yeast YmL33. Thus, the putative “core” of these different r-proteins is much smaller than the mitochondrial representatives of unicellular and multicellular eukaryotes.

**MRP-L31**—Three, 11, and 28 primary hits were found for rat, human, and mouse MRP-L31, respectively. For MRP-L31\textsubscript{rat}, an incomplete ORF of 127 amino acid residues was deduced from the cDNA lacking some amino acid residues of the putative MISP (Fig. 1f, Table III). Amino acid residues 18–40 correspond to the N-terminal amino acid sequence of the mature MRP-L31\textsubscript{rat} obtained by amino acid sequencing. The MRP-L31\textsubscript{rat} sequence is very similar to the deduced amino acid sequences of MRP-L31\textsubscript{human} and MRP-L31\textsubscript{mouse} (Table II, Fig. 1f, present report). MISPs of 31 and 32 amino acid residues, respectively, are postulated (Fig. 1f). The MISPs are highly hydrophobic and positively charged, and contain hydroxylated but no negatively charged amino acid residues. The (truncated) MRP-L31\textsubscript{rat} MISP belongs to the R–2 class, whereas the mouse and human MRP-L31 MISPs belong to the R-none class of mt import peptides according to Ref. 29. No significant sequence similarity to any known protein was found. Thus, the mammalian L31 MRPs define a new class of MRPs.

**MRP-L32**—For the MRP-L32 proteins 2, 10, and 38, primary hits were identified for rat, human, and mouse in the EST data bases, respectively. An incomplete ORF of 101 amino acid residues was deduced from a single rat EST sequence (Table III). From amino acid residues 2 to 59, this ORF corresponds to the amino acid sequence of the mature MRP-L32\textsubscript{rat} determined by amino acid sequencing (Fig. 1g). The complete ORFs deduced for both the MRP-L32\textsubscript{mouse} and MRP-L32\textsubscript{human} proteins are quite similar except for their extreme N termini (Fig. 1g, present report). Both proteins show a very good sequence correspondence to the MRP-L32\textsubscript{rat} sequence (Fig. 1g). MISPs of 30 and 64 amino acid residues are postulated for MRP-L32\textsubscript{mouse} and MRP-L32\textsubscript{human} respectively. However, the elongated form of MRP-L32\textsubscript{human} seems to be unlikely, since the surrounding of the second ATG codon perfectly matches the consensus sequence for the start of eukaryotic translation (as the mouse ATG does), whereas the first ATG does not. The MRP-L32\textsubscript{mouse} start codon is preceded by an in-frame stop codon. Both the mouse and human N termini show general properties of MISPs, such as a high content of hydrophobic, positively charged, and hydroxylated amino acid residues, but no negatively charged residues. The mouse MISP belongs to the R–10 class according to Ref. 29. In the human sequence, the arginine (R) at position –10 is replaced by a histidine (H). However, it has not been shown so far that histidine can functionally replace arginine in MISP processing.
The mammalian L32 MRPs show significant sequence similarities as compared with E. coli r-protein L14 and the corresponding YmL38 MRP of yeast (Fig. 1g, Table III). The N termini of the latter two correspond to the postulated mature N termini of the mammalian L32 MRPs. Interestingly, the YmL38 lacks a cleavable MISP (2). Thus, the N terminus of the mature YmL38 corresponds to the N terminus of the mature (i.e., after mt import and processing) mammalian MRP-L32s (Fig. 1g).

**MRP-S13**—Astonishingly, only three and one ESTs were found in the primary search for rat and mouse MRP-S13, respectively. Human ESTs corresponding to this MRP were identified using the mouse cDNA sequences as “screening probe.” For MRP-S13(mouse), an ORF was deduced that lacks an N-terminal start codon and an N-terminal extension as compared with the mouse and human MRP-S13s (Table II, Fig. 1a, present report). Although 50 different ESTs of rat origin were identified by using the extreme 5’ end of EST A10455711r,c (see Table II) as screening probe, none of them extends more than 15 base pairs in the 5’ direction. Because the MRP-S13(mouse) ORF presented in Fig. 1h is preceded by three in-frame stop codons, and because the deducible amino acid residues at the very N terminus do not correspond to the respective MRP-S13(mouse) amino acid sequence in the same position, we assume that the 5’ end nucleotide sequence of EST A10455711r,c represents an intron that was not reverse transcribed during EST creation and/or sequenced. The mature MRP-S13(mouse) as compared with the N-terminal sequence obtained by direct amino acid sequencing is highly conserved, as compared with the respective mouse and human MRP-S13 sequences (Fig. 1h). For MRP-S13(mouse) an ORF of 200 amino acid residues was deduced from a consensus cDNA of 977 nucleotides (Table II). The pre-mRNA of MRP-S13(mouse) contains at least three introns, two at positions 107/108 and 154/155, and a third intron of 80 nucleotides at position 363/364 of the consensus cDNA. This was deduced by comparison of EST sequences derived from incompletely spliced mRNA molecules with the mature consensus cDNA. At position 925–930, a polyadenylation signal AATAAA was found, and 28 consecutive adenine residues mark the location of the poly(A) tail from position 943 onward. The deduced ORF is highly conserved as compared with the mature MRP-S13(mouse) sequence (Fig. 1h, present report). For MRP-S13(human), a consensus cDNA was assembled, and the corresponding genomic DNA was localized 3’ of the GnRH-II gene (Ref. 32; accession no. AF036329). The EST consensus cDNA sequence corresponds to that of the genomic DNA from nucleotide 3763 to nucleotide 4424 of the latter. Two short introns are covered by this region (data not shown). However, the genomic DNA sequence does not cover the entire MRP-S13(human) sequence. Accordingly, the C terminus was completed by adding EST derived consensus cDNA sequences (Table II).

For both the MRP-S13(mouse) and the MRP-S13(human), an MISP of 27 amino acid residues is postulated, respectively. Although not identical in sequence, both peptides are quite similar in specific properties such as a high content of hydrophobic amino acid residues and a net positive charge. The MRP-S13(mouse) MISP belongs to the R–2 class, and the MRP-S13(human) MISP belongs to the R–3 class of signal peptides according to Ref. 29. In general, the MISPs of the mammalian MRPs presented in this study are quite similar to each other. This conclusion is not based on the primary amino acid sequences, which are heterogeneous; instead, an analysis of the properties shows common features. The MISPs of the MRPs are between 27 and 46 amino acids in length. The only exceptions are the MRP-L27 MISPs of 13 amino acid residues. Nearly all of them show a structure prediction of an N-terminal α-helix joined to a C-terminal β-sheet. The putative MISP of MRP-L32(human) of 64 amino acid residues shows the α-helix-β-sheet motif twice in a row. All MRP MISPs are highly hydrophobic, they have a net positive charge, and they contain less hydroxylated amino acid residues than is common for other MISPs (28).

Altogether, sequences of 23 different mammalian MRPs have been identified by comparison with the N-terminal peptide sequences of purified rat MRPs determined by biochemical methods (Table I). The significance of the deduced ORF sequences is influenced by the inaccurate EST sequencing results. However, although all the ORF sequences deduced need further confirmation by classical cDNA isolation and sequencing, the deduced amino acid sequences are reliable in terms of consensus cDNA assembling and sequence comparison with similar proteins of the same r-protein family. Thus, eight classes of mammalian MRPs were characterized in this work.

**DISCUSSION**

Our understanding of mt genetics linked to mutations in nuclear genes suffers from a lack of knowledge of the influence of nuclear encoded proteins on mt maintenance and function. This is also a crucial point in the elucidation of molecular mechanisms for nuclear-inherited mitochondrial diseases. Nuclear-encoded MRPs are good candidates for proteins involved in mt genetics, as has been shown in yeast (1, 2). In the present report, we describe the identification of eight groups of mammalian MRPs from rat, mouse, and human.

It might seem surprising that the numbers of ESTs picked from the data bases at least as primary hits vary so much among the different MRPs investigated. One could assume that ESTs respective mRNAs of different gene products, which are present in stoichiometric amounts within the cell (the mt ribosome), would be represented in the data bases in similar ratios. In yeast the expression of MRP genes is differentially regulated at the mRNA, translational, and protein stability levels. Thus, equal amounts of MRp mRNAs are not present (for a review, see Ref. 2). Additionally, it should be noted that ESTs derive from different healthy and tumor tissues. The degree of respiration and the mitochondrial content of different tissues differs remarkably, causing a different level of expression of nuclear genes for mt proteins. The automatic processing of mRNAs yielding ESTs may also influence the occurrence of individual ESTs by selective preferences for the oligonucleotides and reverse transcriptases used. Therefore, it seems rather unlikely that the relative amounts of MRP ESTs detected would mirror any (assumed) molecular ratios within an “ideal” cell.

By comparison of the deduced MRp ORFs with the rat N-terminal amino acid sequences of the mature MRPs isolated, we postulated several MISPs. These postulated peptides were further analyzed according to the criteria of common features for such import sequences (28, 29). However, no consensus sequence in the sense of a canonical amino acid sequence has been found for MISPs, and it is unlikely that such a consensus will emerge. Thus, in general, the mammalian MRP MISPs do indeed fit the postulated properties such as a highly hydrophobic character, a net positive charge, and hydroxylated amino acid residues (28). The assignment of the mammalian MRP MISPs to various classes of substrates for the mt processing proteases is common to MRPs. This has been shown for more than 30 different yeast MRPs (2). Although the MISPs presented in this study are on average 30–40 amino acid residues in size, shorter ones such as the MISPs of the mammalian MRP-L27s (13 amino acid residues) are not unusual. In yeast short MISPs (between 14 and 7 amino acid residues) have been identified for MRPs (2). The lack of the mammalian MRP MISPs for hydroxylated amino acid residues may be a specific property of this class of mammalian proteins imported into
mitochondria.

Five of the different mammalian MRP classes show significant sequence similarities to r-proteins from other mt and/or bacterial sources. However, the calculated values for sequence similarities and identities are low as compared with most of the homogeneous classes of r-proteins from other sources (33). In this context it should be noted that a comparison of yeast MRPs to r-proteins of defined classes reveals that MRPs in general seem to be much more divergent (for a review, see Ref. 2). MRPs show N- and C-terminal and/or internal sequence elongations as compared with eubacterial/chloroplast/eukaryotic cytoplasmic r-proteins. MRPs are less similar to other members of the same r-protein family and among species (2). Mammalian and yeast MRPs are as much divergent as mammalian MRPs and E. coli r-proteins. This is in sharp contrast to rat and yeast cytoplasmic r-proteins, which on average share 60% identical amino acid residues (34). On the other hand, the specific large subunit r-proteins picked out from the data bases by comparison to mammalian large subunit MRPs point to a reduced but still reliable sequence conservation of mammalian MRPs and bacterial r-proteins. It should be mentioned that, in certain families of bacterial r-proteins, some of the members are so divergent that they are not assigned as similar by a direct sequence comparison, but only by “intermediates” from other sources (see “MRP-L5”) under “Results”). Thus, mammalian MRPs also fit to these divergent classes of r-proteins.

Since mt ribosomes contain many more proteins than cytoplasmic ribosomes, it is not surprising that several MRPs were identified for which no similar proteins, e.g. in E. coli ribosomes, exist. These “new” MRPs define new classes of r-proteins. They may represent the “molecular excess” of proteins as compared with cytoplasmic ribosomes. In yeast the majority of MRPs are not similar to any other r-protein (2). However, the fact that only one of the mammalian MRPs which are not similar to E. coli r-proteins shares sequence similarity with yeast MRPs (Yml27/MPR-L27 mammalian) (i) may result from the still incomplete characterization of all yeast MRPs, and/or (ii) may be the consequence of further reaching divergences of yeast and mammal mt ribosomes in their respective protein compositions. It should be noted at this point that yeast and mammal mt ribosomes differ strongly in their protein/RNA ratio, although they seem to possess the same total molecular mass (35). The mammalian MRP-L32s, MRP-L31s, and MRP-S13s seem to represent such additional proteins, as compared with E. coli and yeast mt r-proteins. Due to the stringent washing and purification methods applied, it can be excluded that these additional proteins represent mt translational factors loosely attached to the mt ribosome rather than true MRPs. The purification methods for yeast and rat MRPs are comparable, and no mt translational factor has so far been found as a contaminant of mt ribosome preparations (36). Only two-dimensional PAGE spots, which appear reproducibly in stoichiometric amounts in repetitive experiments, are counted as true MRPs (27).

In general, MRPs are much less conserved from one species to another than cytoplasmic r-proteins. This finding raises questions concerning the molecular mechanisms that have allowed MRPs such a divergent evolution while keeping the mitochondrial protein biosynthesis machinery intact. Furthermore, is it possible to generalize results about MRPs to the same extent as has been done for bacterial and eukaryotic cytoplasmic r-proteins? Obviously, the use of yeast MRPs as a model system for mammalian MRPs suffers from the low conservation of sequences and proposed functions. A practical consequence is that we will not be able to identify most of the mammalian MRP genes simply by comparison of yeast MRP sequences with the increasing number of unknown mammalian EST and genomic DNA sequences. Peptide sequences obtained from mammalian MRPs are much more helpful for this purpose. Short peptide sequence data correspond to DNA sequence data from the same species or close relatives strongly enough to identify the corresponding genes unambiguously. This method has been successfully applied to yeast MRPs (24, 36). Further, the correspondence of human cytoplasmic r-proteins separated by two-dimensional PAGE to sequenced r-protein genes has been proven (30). In this work, we have applied this approach for the first time to mammalian MRPs and their corresponding genes.

Acknowledgment—We thank R. Brimacombe for carefully reading the manuscript.

REFERENCES

1. Myers, A. M., Pape, L. K., and Tzagoloff, A. (1985) EMBO J. 4, 2087–2092
2. Graack, H.-R., and Wittmann-Liebold, B. (1998) Biochem. J. 329, 433–448
3. Czempiel, W., Klose, J., and Bass, R. (1976) FEBS Lett. 62, 259–262
4. Pietromonaco, S. F., Denisow, N. D., and O’Brien, T. W. (1991) Biochim. Biophys. Acta (Paris) 1127, 287–335
5. Calhoun, A., Biao, D. L., and Cunningham, C. C. (1995) Anal. Biochem. 232, 437–455
6. Denisow, N. D., and Stoﬄer-Meilicke, M. (1989) Endocytob. Cell Res. 6, 183–192
7. Piatyszek, M. A., Denisow, N. D., and O’Brien, T. W. (1988) Nucleic Acids Res. 16, 2565–2583
8. Denisow, N. D., Michaels, G. S., Monotya, J., Attard, G., and O’Brien, T. W. (1989) J. Biol. Chem. 264, 8328–8338
9. Denisow, N. D., Anders, J. C., and O’Brien, T. W. (1991) J. Biol. Chem. 266, 9586–9590
10. Liao, H. X., and Spremuli, L. L. (1989) J. Biol. Chem. 264, 7518–7522
11. Farwell, M. A., Schirawski, J., Hager, P. W., and Spremuli, L. L. (1996) Biochim. Biophys. Acta 1309, 122–130
12. Denisow, N. D., LiCata, V. J., Gualerzi, C., and O’Brien, T. W. (1988) Biochemistry 27, 3521–3527
13. Kumazawa, Y., Schwartzbach, C. J., Liao, H. X., Mizumoto, K., Kazuyo, Y., Miura, K., Watanabe, K., and Spremuli, L. L. (1991) Biochim. Biophys. Acta 1090, 167–172
14. Schwartzbach, C. J., and Spremuli, L. L. (1991) J. Biol. Chem. 266, 16324–16330
15. Ou, J.-H., Yin, T. S. B., Wang, Y.-F., Kam, W. K., and Rutter, W. J. (1987) Nucleic Acids Res. 15, 8919–8934
16. Graack, H.-R., Grohmann, L., Kitakawa, M., Schafer, K.-L., and Kruft, V. (1992) Eur. J. Biochem. 206, 373–380
17. Marti, L. and Fort, P. (1996) J. Biol. Chem. 271, 11468–11476
18. Royden, C. S., Pirrotta, V., and Jan, L. Y. (1987) J. Biol. Chem. 262, 9952–9958
19. Grossmann, L. I., and Shoubridge, E. A. (1996) BioEssays 18, 983–991
20. Zeviani, M., Petruzzella, V., and Carrozzone, R. (1997) Bioenerg. Biomembr. 29, 121–130
21. Hardy, S. J., and Kurland, C. G. (1999) Biochemistry 28, 2907–2905
22. Ulbrich, B., Czempiel, W., and Bass, R. (1989) Eur. J. Biochem. 180, 337–343
23. Grohmann, L., Graack, H.-R., and Kitakawa, M. (1991) FEBS Lett. 284, 51–56
24. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang. Z., Miller, W., and Lipman, D. J. (1997) Nucleic Acids Res. 25, 3383–3402
25. GCG (1997) Wisconsin Package. Version 8.1 OpenVMS, University of Wisconsin, Madison, Wisconsin
26. Czempiel, W. (1980) Untersuchungen zur Struktur und Funktion mitochondrialer Ribosomen aus Warmbluetern. Ph.D. thesis, Free University of Berlin, Berlin
27. von Heijne, G., Glatz, K., and Herrmann, R. G. (1989) Eur. J. Biochem. 180, 535–545
28. Branda, S. S., and Isaya, G. (1995) J. Biol. Chem. 270, 27366–27373
29. Vladimirnov, S. N., Ivanov, A. V., Karpova, G. G., Musolyanov, A. K., Egorov, T. A., Thiede, B., Wittmann-Liebold, B., and Otto, A. (1996) Eur. J. Biochem. 239, 144–149
30. Graack, H.-R., Grohmann, L., and Kitakawa, M. (1991) Biochimie (Paris) 73, 837–844
31. White, R. B., Eisen, J. A., Kasten, T. L., and Fernald, R. D. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 305–309
32. Muller, E.-C., and Wittmann-Liebold, B. (1997) Cell. Mol. Life Sci. 53, 34–50
33. Wool, I. G., Chan, Y.-L., and Glick, A. (1995) Biochem. Cell Biol. 73, 933–947
34. Kitakawa, M., and Isono, K. (1991) Biochimie (Paris) 73, 813–825
35. Kitakawa, M., Graack, H.-R., Grohmann, L., Geldschmidt-Resin, S., Herfurth, E., Wittmann-Liebold, B., Nishimura, T., and Isono, K. (1997) Eur. J. Biochem. 245, 449–456
