

Heart-specific splice-variant of a human mitochondrial ribosomal protein (mRNA processing; tissue specific splicing)

O. Spirina^a, Y. Bykhovskaya^a, A.V. Kajava^b, T.W. O'Brien^c, D.P. Nierlich^d, E.B. Mougey^e,
J.E. Sylvester^e, H.-R. Graack^f, B. Wittmann-Liebold^f, N. Fischel-Ghodsian^{a,*}

^aAhmanson Department of Pediatrics, Steven Spielberg Pediatric Research Center, Medical Genetics Birth Defects Center, The Cedars-Sinai Medical Center Burns and Allen Research Institute and UCLA School of Medicine, Los Angeles, CA, USA

^bCenter for Molecular Modeling, CIT, NIH, Bethesda, MD, USA

^cUniversity of Florida, Gainesville, FL, USA

^dUCLA, Los Angeles, CA, USA

^eNemours Children's Clinic, Jacksonville, FL, USA

^fMax-Delbruck Center, Berlin, Germany

Received 31 July 2000; received in revised form 9 September 2000; accepted 10 October 2000

Received by S. Salzberg

Abstract

It has been proposed that splice-variants of proteins involved in mitochondrial RNA processing and translation may be involved in the tissue specificity of mitochondrial DNA disease mutations (Fischel-Ghodsian, 1998. *Mol. Genet. Metab.* 65, 97–104). To identify and characterize the structural components of mitochondrial RNA processing and translation, the Mammalian Mitochondrial Ribosomal Consortium has been formed. The 338 amino acid (aa) residues long MRP-L5 was identified (O'Brien et al., 1999. *J. Biol. Chem.* 274, 36043–36051), and its transcript was screened for tissue specific splice-variants. Screening of the EST databases revealed a single putative splice-variant, due to the insertion of an exon consisting of 89 nucleotides prior to the last exon. Screening of multiple cDNA libraries revealed this inserted exon to be present only in heart tissue, in addition to the predominant MRP-L5 transcript. Sequencing of this region confirmed the EST sequence, and showed in the splice-variant a termination triplet at the beginning of the last exon. Thus the inserted exon replaces the coding sequence of the regular last exon, and creates a new 353 aa long protein (MRP-L5V1). Sequence analysis and 3D modeling reveal similarity between MRP-L5 and threonyl-t-RNA synthetases, and a likely RNA binding site within MRP-L5, with the C-terminus in proximity to the RNA binding site. Sequence analysis of MRP-L5V1 also suggests a likely transmembrane domain at the C-terminus. Thus it is possible that the MRP-L5V1 C-terminus could interfere with RNA binding and may have gained a transmembrane domain. Further studies will be required to elucidate the functional significance of MRP-L5V1. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Heart-specific splice-variant; Human mitochondrial ribosomal protein; mRNA Processing; Tissue specific splicing

1. Introduction

Cellular viability is dependent upon adequate levels of energy, the majority of which is produced through oxidative phosphorylation in the mitochondria. The process of oxidative phosphorylation requires proteins coded for by both the mitochondrial and nuclear genomes. The thirteen mitochondrial mRNAs are translated by the genetically distinct mitochondrial ribosome. Mitochondrial malfunction that leads to human disease could thus be caused by mutations in either nuclear-encoded genes or mitochondrial-encoded genes. Since many mitochondrial disorders are caused by mutations in genes for mitochondrial encoded rRNA and tRNA, components involved in mitochondrial translation, we hypothesize that errors in nuclear encoded genes for

Abbreviations: aa, amino acid(s); bp, base pair(s); BCM, Baylor College of Medicine; BLAST, Basic Local Alignment Search Tool; cDNA, DNA complementary to RNA; dNTP, deoxyribonucleoside triphosphate; ds, double strand(ed); E. coli, Escherichia coli; EST, Expressed Sequence Tag; kb, kilobase(s) or 1000 bp; kDa, kilodalton(s); μ l, microliter(s); mM, millimole(s); MRP, mitochondrial ribosomal protein; MRP-L5, mitochondrial ribosomal protein L5; MRP-L5V1, mitochondrial ribosomal protein L5 variant 1; nt, nucleotide(s); oligo, oligodeoxyribonucleotide; ORF, open reading frame; PCR, polymerase chain reaction; rpm, revolutions per minute; TBE, Tris-borate buffer (9 mM Tris-borate, 20 mM EDTA); u, unit(s)

* Corresponding author. Department of Pediatrics, WT#1165, Cedars-Sinai Medical Center, 8700 Beverly Blvd., Los Angeles, CA 90048, USA. Tel.: +1-310-423-6392; fax: +1-310-423-1402.

E-mail address: nfischel@xchg.peds.csmc.edu (N. Fischel-Ghodsian).

mitochondrial ribosomal proteins will also contribute to human pathology.

In order to identify and characterize the nuclear genes involved in mitochondrial translation, the Mammalian Mitochondrial Ribosomal Consortium has been formed. Biochemical purification of mitochondrial ribosomal proteins (MRPs), followed by microsequencing and EST database searches, has led so far to the sequence of 28 MRP genes, 13 of which have no recognizable homologue in any other organism (Sylvester et al., 1999; Graack et al., 2000; O'Brien et al., 2000). Since many of the mitochondrial DNA disorders involving rRNA and tRNA mutations have tissue-specific clinical signs, and since mutations in cytoplasmic ribosomal proteins have similarly been shown to cause tissue-specific diseases (Draptchinskaia et al., 1999), we have speculated that tissue-specific splice-variants or isoforms of some of these MRPs should exist. They could exert their effect either by their role in the mitochondrial ribosome, or alternatively by an entire different function. Interestingly, secondary functions of cytoplasmic ribosomal proteins have been described (Wool, 1996) and examples of secondary tissue-specific functions for mitochondrial genes exist: the mitochondrial large ribosomal RNA gene in *Drosophila melanogaster* can in a few cells be processed for export into the cytoplasm where it induces pole cell formation in embryos, a key event in the determination of the germ line (Kobayashi et al., 1993); similarly, in the mouse the ND1 protein can be processed so that part of it is being presented on the cell membrane with a minor histocompatibility protein (Wang et al., 1991). We have thus initiated a systematic screen for such tissue-specific expression of MRPs. In this manuscript we describe the first finding supporting the existence of tissue-specific variants for human MRPs with possible separate function. MRP-L5 appears to have heart-specific splicing, and this variant was deposited to GenBank with accession number AF270511.

2. Materials and methods

2.1. Web based search for putative splice-variants

MRP-L5 cDNA sequence was identified as described (O'Brien et al., 1999), and used for identification of different cDNA variants. For this purpose, BLAST and TBLASTX searches (Altschul et al., 1997) were used against the human dbEST (Expressed Sequence Tags) database at the GenBank Web server (www.ncbi.nlm.nih.gov/BLAST/).

2.2. cDNA and cDNA libraries

Oligo dT primed cDNA libraries were obtained from different sources: Human brain, fetal brain, liver, heart and testis SuperScript™ cDNA libraries were purchased from Life Technologies (Long Island, NY), a human leukocyte Matchmaker™ cDNA library was acquired from Clon-

tech (Palo Alto, CA), and a fetal cochlea cDNA library was received from Cynthia Morton (Robertson et al., 1994). In addition, lung and muscle Multiple Choice™ cDNAs were obtained from OriGene Technologies, Inc (Rockville, MD).

The SuperScript™ cDNA libraries were grown and plasmid DNA was purified as described in the manufacturer's protocol (Life Technologies). The Matchmaker human leukocyte cDNA library was grown according to the manufacturer's protocol and DNA was extracted with YEAST-MAKER™ Yeast plasmid isolation kit (Clontech). The fetal cochlea library in the Uni-ZAP®XR vector (Stratagene, La Jolla, CA) was amplified according to the manufacturer's protocol (Stratagene). DNA was extracted according to the standard method for purification and DNA extraction of bacteriophage λ (Sambrook et al., 1989).

2.3. PCR analysis and sequencing

Synthetic oligonucleotides for MRP-L5, based on GenBank accession AF239727, and described as 'Homo sapiens chromosome 21 unknown mRNA', were designed using Oligo 6.3 program (Molecular Biology Insights Inc., Cascade, CO). Primer names include the number of the starting oligonucleotide from the start codon of the ORF and the letter for the orientation (F-forward, R-reverse), and are shown in the 5' to 3' orientation:

741F: CAGAACCCTGAGAGAATAGTC
in exon G,

1014R: TGTACATTCCTCTGTTGCTT
in exon J,

ADD1.1004R:
TGCCTAGTAAATAGAAGGAGA
in add-in exon ADD1,

ADD1.984F:
TTCCTTCTATTTACTACGCA in exon ADD1.

PCR reactions contained 100 ng of template DNA (corresponding to approximately 5×10^8 cfu/μl), 0.2 mM dNTPs, 10 pmol of each primer, 1.5 mM MgCl₂, 10 mM Tris-HCl, pH 8.3; 50 mM KCl, and 1.25 units Taq DNA polymerase (Display Systems Biotech, Vista, CA) in a total volume of 40 μl. The DNA was initially denatured at 94°C for 3 min, followed by 35 cycles of denaturing at 94°C for 20 s, annealing at 51°C for 45 s, and extension at 72°C for 3 min, and final 10 min extension in GeneAmp™ PCR system 9700 (PE Applied Biosystems, Foster City, CA).

For sequencing reactions with the dsDNA cycle sequencing system (Life Technologies), the amplification products were extracted from the gel using Concert™ Rapid Gel Extraction System (Life Technologies). Sequencing reactions were performed by end labeling primers with $\gamma^{33}\text{P}$ ATP, following by cycle sequencing, using primers

previously used for PCR amplifications. Sequencing products were resolved on 6% denaturing acrylamide gel.

2.4. cDNA and protein sequence analysis, alignment and modeling

Alignment of DNA was performed using CLUSTALW program (Thompson et al., 1994). For splice sites prediction the RNASPL program (Solovyev et al., 1994) at the Baylor College of Medicine (BCM) Web server was used (www.hgsc.bcm.tmc.edu/SearchLauncher/).

Protein homologies were evaluated by PSI-BLAST against non-redundant, and TBLASTN against human EST databases in NCBI.

Pattern matches were performed using the ProfileScan server (www.isrec.isb-sib.ch/software/PFSCAN) at ISREC (Swiss Institute for Experimental Cancer Research) to search against PROSITE database (Bairoch et al., 1997). Conserved protein regions were evaluated against the BLOCKS database (www.blocks.fhrc.org) (Henikoff and Henikoff, 1994). Transmembrane regions were evaluated by Web-based TMPred server (www.ch.embnet.org/software/TMPRED) (Hofmann and Stoffel, 1993).

A multiple protein sequence alignment was constructed using ClustalW program (Higgins et al., 1996) and manual alignment corrections were applied whenever necessary based on the known 3D structure of threonyl-tRNA synthetase from *E. coli*. The structural model of MRP-L5 was built using HOMOLGY module of Insight II program (Dayring et al., 1986).

3. Results and discussion

3.1. Identification of MRP-L5 splice-variant

The cDNA sequence for human MRP-L5 was initially assembled by homology search of human EST databases with peptide sequence obtained by microsequencing of bovine MRP-L5 (O'Brien et al., 1999). We compared this published sequence in the THCD database (Quackenbush et al., 2000) to Unigene clusters (Boguski and Schuler, 1995)

and the non-redundant database. This showed that the MRP-L5 mRNA sequence completely corresponds to an unknown human mRNA on chromosome 21 (GenBank accession number AF239727).

To screen for putative splice-variants BLAST homology search of the MRP-L5 cDNA against the human EST database was performed. All resulting sequences were individually compared, and revealed one EST (AA972680 - derived from three human normalized pooled libraries - fetal lung NbHL19W, testis NHT, and B cell NCI_CGAP_GCB1) containing the MRP-L5 sequence with an additional 89 nucleotides close to the 3' end of the gene.

In order to determine whether this insertion could be a real variant, rather than an artifact, the genomic sequence of the gene was analyzed. The complete genomic sequence of MRP-L5 (AP000223) is 22 kb long and consists of ten exons ranging from 51 to 194 bp, and nine introns ranging from 747 to 3127 bp (O'Brien et al., 1999). The newly identified variant EST contained exons H, I and J, and an additional 89 nucleotides inserted between exons I and J (Fig. 1, dark vertical box ADD1). The splicing consensus sequence was predicted by the RNASPL program at BCM (Solovyev et al., 1994) on both sides of the insertion.

The known cDNA sequence of human MRP-L5 consists of 1072 nt (O'Brien et al., 1999), starting with the start signal of the open reading frame (ORF). The ORF finishes with the stop codon at 1017 nt position. The cDNA has two polyadenylation signals at bp 1044 and bp 1049. The predicted length of the corresponding protein is 338 amino acid residues with a calculated molecular weight of 36 kDa (O'Brien et al., 1999). Sequence analysis predicts that the addition of ADD1 changes the 3' end of MRP-L5 by adding 30 new amino acids and creating a new termination codon at the very beginning of exon J, leading to a replacement of the C-terminus in the splice-variant, denoted MRP-L5V1 (Fig. 2). MRP-L5V1 is predicted to be 353 aa long.

3.2. Expression pattern of the MRP-L5 splice-variant

In order to characterize the expression pattern of MRP-L5V1, PCR amplifications from cDNA of a variety of

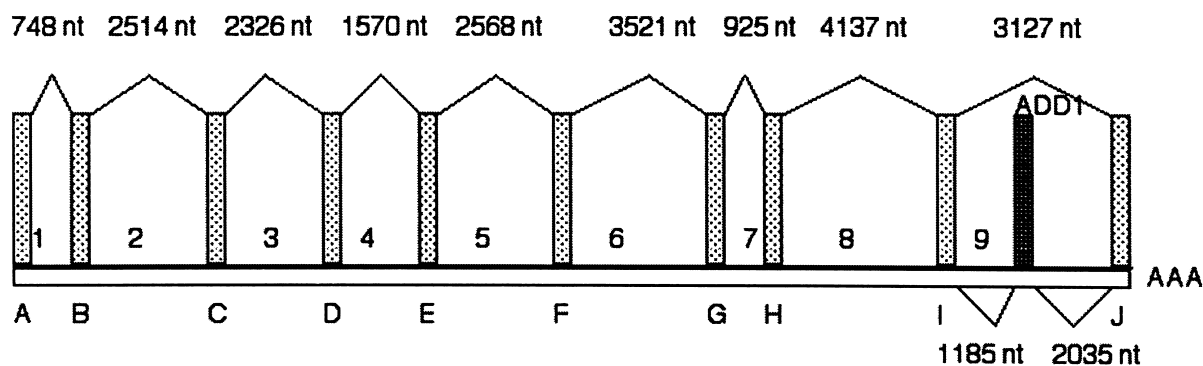


Fig. 1. Genomic structure of MRP-L5. Letters correspond to exons, numbers are introns. Exons: A from 1–74; B, 75–280; C, 281–420; D, 421–520; E, 521–588; F, 589–701; G, 702–767; H, 768–922; I, 923–969; J, 970–1068. ADD1 consists of 89 bp.

```

MRP-L5:  CGGAAAATG-----969
          R  K  M
MRP-L5V1: CGGAAAATGACTCCATTCCATTCTCCTTCTATTACTACGCAGTCATTCTTCACTACC 1020
          R  K  M  T  P  F  P  I  L  L  L  F  T  T  Q  S  F  F  T  T
MRP-L5:  -----970-----GTAAGTGAAGATCAAAGTAAAG 991
          V  T  E  D  Q  S  K
MRP-L5V1: TCGCCTGAGTCGTACCTCCTCCATGGAACAGTCTCAGAGTAACTGAAGATCAAAGTAAAG 1080
          S  P  E  S  Y  L  L  H  G  T  V  S  E
          1058
MRP-L5:  CAACAGAGGAATGTACATCTACCTAATAAC 1021
          A  T  E  E  C  T  S  T
MRP-L5V1: CAACAGAGGAATGTACATCTACCTAATAAC 1110
    
```

Fig. 2. cDNA and protein sequence alignment for MRP-L5 and MRP-L5V1. Underlined letters code for terminal codons.

human tissues were performed. Fig. 3 shows amplification between exons G and J. The predominant band in all tissues is a 273 bp band corresponding to the cDNA for MRP-L5. Only in cDNA from heart tissue a small quantity of an additional 364 bp band, corresponding to the expected size for MRP-L5V1, can be seen. To more directly and

more sensitively detect the presence of this splice-variant, PCR amplification using primers in ADD1 and exon G were used (741F and 1004R). Again a 263 bp band was seen only in heart tissue (Fig. 4). Interestingly no band was seen in lung, lymphocytes or testis, which was the tissue of origin for the sole EST containing this splice variation. To assess

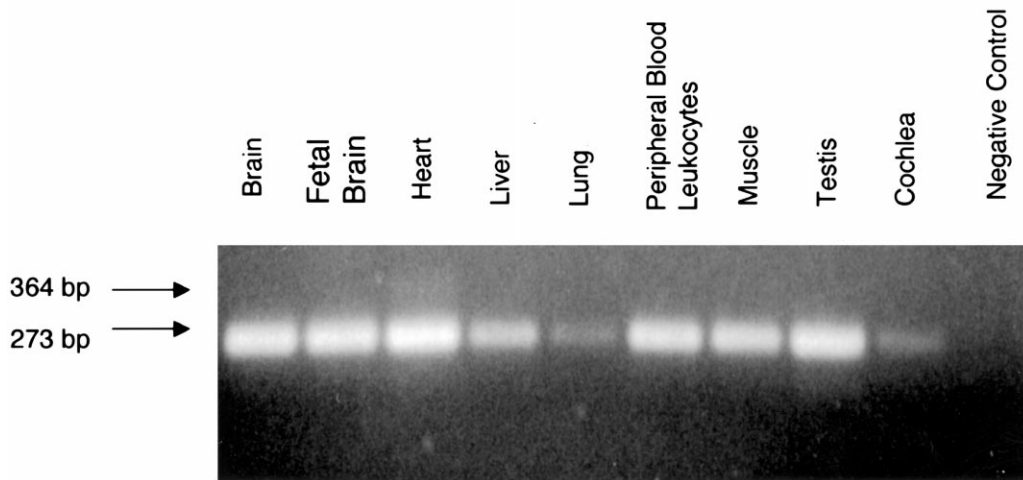


Fig. 3. PCR results for primers, located in exons G and J (741 F and 1014 R).

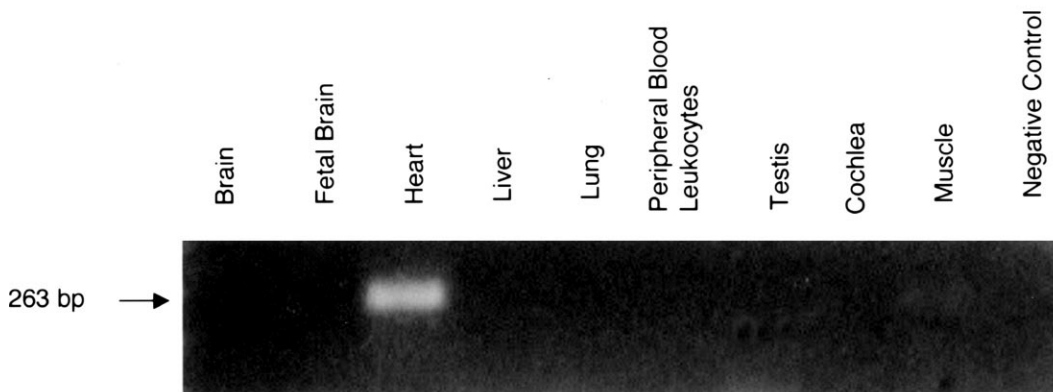


Fig. 4. PCR results for primers, located in exons G and ADD1 (741 F and ADD1.1004 R).

possible contamination of this library with heart-specific sequences we screened this library for the presence of other known heart specific ESTs by BLAST search. We were able to detect sequences for human myosin-binding protein C (cardiac-type), muscle and heart-specific isoform cytochrome c oxidase subunit (COX VIIa-M) and a cardiac-specific homeobox gene (CSX) in this library (data not shown), confirming the presence of heart cDNAs in the original library.

3.3. Sequence confirmation of MRP-L5V1

In order to confirm the sequence of MRP-L5V1, cDNA from the heart cDNA library was PCR amplified with PCR primers 741F and ADD1.1004R and primers ADD1.984F and 1014R. The resulting fragments were directly sequenced from both ends, and were identical to the sequence contained in EST AA972680.

3.4. Amino acid sequence analysis and homology-based modeling

To evaluate putative changes in the function and structure of MRP-L5V1 due to the change in the C-terminus, sequence similarity searches were executed. BLAST and BLOCKS server search of both C-termini against different protein databases did not produce any primary sequence homology. The search against the PROSITE database revealed two putative casein kinase II phosphorylation sites for MRP-L5V1. In addition, and perhaps more strikingly, the TMpred server (Hofmann and Stoffel, 1993) predicts a transmembrane domain at the C-terminal part of MRP-L5V1 with a score of 914 (in this program a score over 500 is considered significant).

In order to evaluate whether the changed C-terminus affects the structure and function of the entire protein, a three dimensional model of the protein was generated as follows: PSI-BLAST search using the MRP-L5 and MRP-L5V1 sequences resulted in highly significant matches to the N-terminal domains of several threonyl-tRNA synthetases with E-values (the number of hits one can expect to see by chance when searching a database of particular size) equal to 5×10^{-18} for the synthetase from yeast and 5×10^{-7} for the synthetase from *E. coli*. The synthetase from *E. coli* has been crystallized in complex with the tRNA molecule (Sankaranarayanan et al., 1999). The availability of the X-ray structure provided us the possibility to generate a structural model of MRP-L5, which suggests that it can bind RNA in a similar manner (Fig. 5). In Fig. 5 the C-terminal end of MRP-L5 is shown in the α -helical conformation taken from the threonyl-tRNA synthetase structure. The C-terminus end is located near the putative RNA-binding site, therefore, the difference between the C-termini of MRP-L5 and MRP-L5V1 may affect the RNA binding. Thus, we can speculate that the RNA binding function of MRP-L5V1 could be lost, and instead a different function requiring transmembraneous signaling and phosphorylation is substituted. It remains entirely unknown what this function could be, and why it is only present in cardiac tissue.

Acknowledgements

This work was done as part of the Mammalian Mitochondrial Ribosomal Consortium and is supported by NIH/NIDCD grant RO1DC04092. JES acknowledges also support from the Nemours Research Programs.

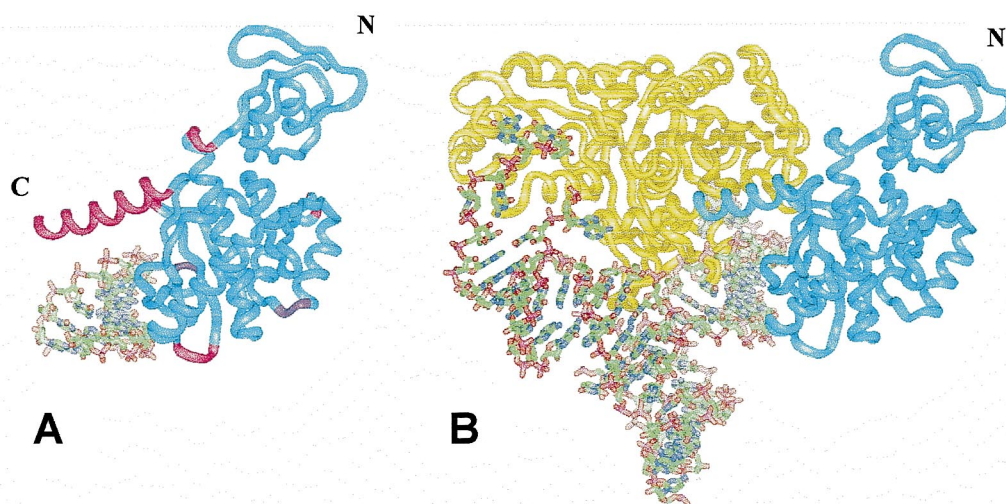


Fig. 5. Structural comparison between (A) complex of MRP-L5 and short fragment of A-form RNA helix and (B) complex of threonyl-tRNA synthetase from *E. coli* with tRNA (Sankaranarayanan et al., 1999). The RNA is shown in green and red. Fragments in blue indicate regions of similar structure. The fragment in yellow at the C-terminus of *E. coli* synthetase indicates the region not present in MRP-L5. Fragments of MRP-L5 in purple and in violet indicate insertions and deletions, respectively (compared with threonyl-tRNA synthetase). The C-terminal end of MRP-L5 (purple) is shown in the α -helical conformation taken from the threonyl-tRNA synthetase structure. The figure was generated using the Insight II program (Dayring et al., 1986).

References

- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-Blast: a new generation of protein database search programs. *Nucleic Acid Res.* 25, 3389–3402.
- Bairoch, A., Bucher, P., Hofmann, K., 1997. The PROSITE database, its status in 1997. *Nucleic Acids Res.* 25, 217–221.
- Boguski, M.S., Schuler, G.D., 1995. ESTablishing a human transcript map. *Nat. Genet.* 10, 369–371.
- Dayring, H.E., Tramonato, A., Sprang, S.R., Fletterick, R.J., 1986. Interactive program for visualization and modeling of proteins, nucleic acids and small molecules. *J. Mol. Graph.* 4, 82–87.
- Draptchinskaia, N., Gustavsson, P., Andersson, B., Petersson, M., Willig, T.N., Dianzani, I., Ball, S., Tchernia, G., Klar, J., Mattsson, H., Tentler, D., Mohandas, N., Carlsson, B., Dahl, N., 1999. The gene encoding ribosomal protein S19 is mutated in Diamond-Blackfan anaemia. *Nat. Genet.* 21, 169–175.
- Fischel-Ghodsian, N., 1998. Mitochondrial RNA processing and translation: link between mitochondrial mutations and hearing loss? *Mol. Genet. Metab.* 65, 97–104.
- Graack, H.-R., Sylvester, J.E., Mougey, E.B., Fischel-Ghodsian, N., Wittmann-Liebold, B., O'Brien, T.W., 2000. Mammalian mitochondrial ribosomal proteins (MRPs) and their corresponding genes: identification, characterization, and genetics. *Eur. J. Med. Res.* 5, 61.
- Henikoff, S., Henikoff, J.G., 1994. Protein family classification based on searching a database of blocks. *Genomics* 19, 97–107.
- Higgins, D.G., Thompson, J.D., Gibson, T.J., 1996. Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.* 266, 383–402.
- Hofmann, K., Stoffel, W., 1993. TMbase A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler* 347, 166.
- Kobayashi, S., Amikura, R., Okada, M., 1993. Presence of mitochondrial large ribosomal RNA outside mitochondria in germ plasm of *Drosophila melanogaster*. *Science* 260, 1521–1524.
- O'Brien, T.W., Fiesler, S.E., Denslow, N.D., Thiede, D., Wittmann-Liebold, B., Mougey, E.B., Sylvester, J.E., Graack, H.R., 1999. Mammalian mitochondrial ribosomal proteins (2). *J. Biol. Chem.* 274, 36043–36051.
- O'Brien, T.W., Liu, J., Sylvester, J.E., Mougey, E.B., Fischel-Ghodsian, N., Thiede, B., Wittmann-Liebold, B., Graack, H.R., 2000. Mammalian mitochondrial ribosomal proteins (4). Amino acid sequencing, characterization, and identification of corresponding gene sequences. *J. Biol. Chem.* 275, 18153–18159.
- Quackenbush, J., Liang, F., Holt, I., Pertea, G., Upton, J., The, T.I.G.R., 2000. Gene Indices: reconstruction and representation of expressed gene sequences. *Nucleic Acid Res.* 28, 141–145.
- Robertson, N.G., Khetarpal, U., Gutiérrez-Espeleta, G.A., Bieber, F.R., Morton, C.C., 1994. Isolation of novel and known genes from a human fetal cochlea cDNA library using subtractive hybridization and differential screening. *Genomics* 23, 42–50.
- Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. *Molecular Cloning: A Laboratory Manual*, 2nd Edition. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- Sankaranarayanan, R., Dock-Bregeon, A.C., Romby, P., Caillet, J., Springer, M., Rees, B., Ehresmann, C., Ehresmann, B., Moras, D., 1999. The structure of threonyl-tRNA synthetase-tRNA(Thr) complex enlightens its repressor activity and reveals an essential zinc ion in the active site. *Cell* 97, 371–381.
- Solovyev, V.V., Salamov, A.A., Lawrence, C.B., 1994. Predicting internal exons by composition and discriminant analysis of spliceable open reading frames. *Nucleic Acids Res.* 22, 5156–5163.
- Sylvester, J.E., Graack, H., Liu, J., Mougey, E.B., Maguire, B.A., Fischel-Ghodsian, N., Wittman-Liebold, B., O'Brien, T.W., 1999. Consortium to clone, map, and study human mitochondrial ribosomal protein genes. *Am. J. Hum. Genet.* 65, A421.
- Thompson, J.D., Higgins, D.G., Gibson, T.J., CLUSTAL, W., 1994. Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680.
- Wang, C.R., Loveland, B.E., Fischer-Lindahl, K., 1991. H-2M3 encodes the MHC class I molecule presenting the maternally transmitted antigen of the mouse. *Cell* 66, 35–345.
- Wool, I.G., 1996. Extraribosomal functions of ribosomal proteins. *Trends Biochem. Sci.* 21, 164–165.