DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
47

STUDIES ON MAMMALIAN RIBOSOMAL PROTEIN S7

TARMO ANNILO

TARTU 1998

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS

47

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
47

STUDIES ON MAMMALIAN RIBOSOMAL PROTEIN S7

TARMO ANNILO

TARTU UNIVERSITY
PRESS

Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Estonia

Dissertation is accepted for the commencement of the degree of Doctor of Philosophy (in Molecular Biology) on November 19th, 1998 by the Council of the Institute of Molecular and Cell Biology, University of Tartu.

Opponent: Prof. Francesco Amaldi (University of Rome, Italy)

Commencement: January 15, 1999

The publication of this dissertation is granted by the University of Tartu

Tartu Ülikooli Kirjastuse trükikoda
Tiigi 78, Tartu 50410
Tellimus nr. 421

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

Current thesis is based on the following original publications which will be referred to by their Roman numerals:

I Annilo, T., Laan, M., Stahl, J. and Metspalu, A. (1995) The human ribosomal protein S7-encoding gene: isolation, structure and localization in 2p25. *Gene* **165**, 297–302.

II Annilo, T., Karis, A., Hoth, S., Rikk, T., Kruppa, J. and Metspalu, A. (1998) Nuclear import and nucleolar accumulation of the human ribosomal protein S7 depends on both a minimal nuclear localization sequence and an adjacent basic region. *Biochem. Biophys. Res. Commun.* **249**, 759–766.

III Annilo, T., Jelina, J., Pata, I. and Metspalu, A. (1998) Isolation and characterization of the mouse ribosomal protein S7 gene. *Biochem. Mol. Biol. Int.* **46**, 287–295.

LIST OF ABBREVIATIONS

5' TOP	—	5' terminal oligopyrimidine tract
aa	—	amino acid(s)
bp	—	basepair(s)
cDNA	—	complementary DNA
CpG	—	dinucleotide 5'-CG-3'
DNA	—	deoxyribonucleic acid
EST	—	expressed sequence tag
FISH	—	fluorescence *in situ* hybridization
kb	—	kilobasepair(s)
kDa	—	kilodalton(s)
mRNA	—	messenger RNA
NLS	—	nuclear localization signal
NPC	—	nuclear pore complex
PCR	—	polymerase chain reaction
RNA	—	ribonucleic acid
rp(s)	—	ribosomal protein(s)
rRNA	—	ribosomal RNA
snoRNA	—	small nucleolar RNA
TSP	—	transcription start point

1. INTRODUCTION

The organelle of protein synthesis, ribosome, consists of four RNA species and 80 different proteins in mammals. Ribosomal components make up about 15% of cellular protein and 85% of RNA. The functional principle and fundamental organization of the ribosome have been conserved throughout the evolution, although the ribosomes vary in primary structure and number of its constituents. In higher organisms, the machinery of the ribosome is more sophisticated and considerably less understood. Eukaryotic ribosomes are larger (up to 4,2 kDa in mammals compared to 2,5 kDa in bacteria), they contain more different RNA and protein molecules.

In vitro, the peptide bond can be synthesized without the ribosomal proteins, only by the *Escherichia coli* 23S rRNA (Nitta *et al.*, 1998). In a mature ribosome, however, indispensable protein components have been described (Uhlein *et al.*, 1998). Ribosomal proteins, that account only one-third of the ribosome mass, may have several functions. They could be structural elements required for stabilization of functional conformation of rRNA, serve regulatory functions in a translational process or even have activities outside the ribosome, not related to protein synthesis.

Assembly of the eukaryotic ribosome subunits occurs in the nucleolus. To ensure balanced supply of a large number of ribosome constituents in response to cellular growth rate and differentiation, their synthesis and nucleolar transport must be specifically coordinated.

The first fundamental step in elucidating the ribosome biosynthesis mechanism is isolation and characterization of genes for all ribosomal components. Determination of the amino acid sequences of mammalian ribosomal proteins has been completed recently, but genes for only 20 different proteins have been described so far. Genes encoding mammalian ribosomal proteins are scattered over the genome and belong to the families consisting of multiple copies of a given gene. As a rule, however, there is only one active copy for each ribosomal protein gene and the other 5...20 copies are pseudogenes.

The present work describes a detailed structural analysis of two mammalian — human and mouse — homologous genes encoding ribosomal small subunit protein S7 and regions important for its nuclear and nucleolar localization. The biological function of S7 is largely unknown. Experimental data indicate that S7, located at the ribosomal interface, is involved in the mRNA-40S subunit interaction (Mundus *et al.*, 1993) and in the eIF-3 attachment site formation (Westermann and Nygård, 1983). Molecular characterization of the mouse *S7* gene provides information for further gene targeting experiments to obtain information for consideration *S7* as a candidate gene for genetic disorders.

2. REVIEW OF LITERATURE

2.1. Mammalian ribosomal proteins

2.1.1. Organization and structure of ribosomal protein genes

In eukaryotic cells, rp genes are distributed throughout the genome. While in yeast most rp genes have two transcriptionally active copies (Planta and Mager, 1998), all mammalian ribosomal proteins are encoded by a single intron-containing functional gene (with an exception of human S4). Using intron-specific STS mapping, human genes for 75 rps have been localized recently; 51 of them for the first time (Kenmochi *et al.,* 1998). All chromosomes but 7 and 21 carry at least one rp gene. Despite success in genomic mapping and amino acid sequence determination, isolation of rp genes has been impeded by the large number of pseudogenes present in the genome. In addition to the single functional (transcriptionally active) gene, a typical rp gene family consists of up to 20 pseudogenes. Processed promoterless pseudogenes contain a poly(A) tail, are flanked by direct repeats and most probably originate from integration of reversed transcription of mRNA into the genome in germ-line (Vanin, 1985; Weiner *et al.,* 1986).

Despite their large number and dispersion in the genome, mammalian rp genes have similar architecture. This suggests that common structural elements contribute to the coordinate regulation of their transcription and translation. Typical features of rp genes are short 5' leader sequence, lack of canonical TATA box, small first exon and transcription start site located usually at the C residue within a polypyrimidine tract. All mammalian genes have one of the introns near the 5' end, very close to the initiation codon. 5' end of a gene is located in a CpG — rich island, feature which is especially associated with "housekeeping" genes; and a gene is a member of a multigene family. Among mammalian rp genes described so far, rat *S5* is the only one present in a single copy (Kuwano *et al.,* 1992). Although one can not rule out the possibility that another functional gene exists in some cases, only the human *S4* has been shown to have two active copies (Fisher *et al.,* 1990). These genes are located on the chromosomes X and Y and encode functionally interchangeable isoforms of S4, which differ at 19 of 263 positions. In contrast, two rp genes (*L10* and *L36a*) have been mapped solely on the X chromosome.

To date, about 20 different mammalian rp genes have been cloned and characterized to the different extent (Table I). In few cases, homologous genes from different organisms are available. Analysis of *S14* mammalian, insect, protozoan and two fungal genes has shown that exon-intron structure has been conserved only in vertebrates. Among lower eukaryotic species, intron locations are dramatically different (Rhoads and Roufa, 1991). In *S6*, human gene has an extra intron compared to the mouse homolog (Pata *et al.,* 1992; Pata and

Metspalu, 1996). In *S7,* lower eukaryotic homologs have less introns than vertebrates, but the positions of existing introns are similar.

L7a is located in a cluster of six unrelated housekeeping genes, covering about 45 kb (Huxley and Fried, 1990). The organization of this locus is conserved between human, mouse and chicken, but not in *Fugu* and invertebrates (Armes and Fried, 1996; Armes *et al.,* 1997). Whether such tight clustering in higher vertebrates is of functional significance or has been formed by random gene shuffling, is unclear.

Table I. Mammalian ribosomal protein genes. H, *Homo sapiens*; m, *Mus musculus*. The table with direct links to nucleotide data and Medline references can be accessed at http://www.ebc.ee/~tannilo/rp-genes.html

Gene	Accession No	Reference
rat P2	X55153	Chan and Wool, 1991
H L7	L16557	Seshadri *et al.,* 1993
m L7	M29015	Meyuhas and Klein, 1990
H L7a	X61923	Colombo *et al.,* 1991; De Falco *et al.,* 1993
m L7a	X54067	Huxley and Fried, 1990
H L9	U09954	Mazuruk *et al.,* 1996
H L19	X82201; AF003626	Davies and Fried, 1995
rat L19	X82202	
H L23a	AF001689	Fan *et al.,* 1997
m L30	K02928	Wiedemann and Perry, 1984
m L32	K02060	Dudov and Perry, 1984
H L41	AB01087	Go *et al.,* 1998 (direct submission)
H S3	L16016	Polakiewicz *et al.,* 1995
H S3a	X87373; Z83334	Nolte *et al.,* 1996
m S3a	Z83368	Rebane *et al.,* 1998
H S4	AF041428 (X); AF041427 (Y)	Zinn *et al.,* 1994
H S6	X67309; M77232	Pata *et al.,* 1992; Antoine and Fried, 1992
m S6	Z54209	Pata and Metspalu, 1996
H S7	Z25749	Annilo *et al.,* 1995
m S7	AF043285	Annilo *et al.,* 1998
H S8	X67247	Davies and Fried, 1993
H S14	M13934	Rhoads *et al.,* 1986
H S15 (rig)	M32405	Shiga *et al.,* 1990
rat S15	D11388	Takasawa *et al.,* 1992
m S16	M11408	Wagner and Perry, 1985
H S17	M18000	Chen and Roufa, 1988
m S24	X71972	Xu *et al.,* 1994
H S25	M64716	Imai *et al.,* 1994
H S26	U41448	Filipenko *et al.,* 1998
H S30 (fau1)	X65921	Kas *et al.,* 1992

2.1.2. Regulation of expression

Transcriptional regulation

In response to changes in the physiological conditions and during embryogenesis, the synthesis of ribosome components is coordinately regulated (Faliks and Meyuhas, 1982; Baum and Wormington, 1985; Larson *et al.,* 1991). Diverse mechanisms regulating transcription, transcript stability, splicing, translation, post-translational modification and protein stability are involved in modulation of rp production. Although rp genes are considered to belong to "housekeeping", ubiquitously expressed class of genes, the level of expression of rp mRNAs varies developmentally and from tissue to tissue, being highest in the tissues with elevated protein synthesis rate like liver and fetal brain (Mazuruk *et al.,* 1996). During *Xenopus* embryogenesis, rp genes display spatially distinct tissue-specific expression (Scholnik *et al.,* 1997). Decreased accumulation of rps during terminal differentiation of mouse myoblasts is regulated both at the transcriptional and translational level (Agrawal and Bowman, 1987). Expression of at least six rp genes is specifically down-regulated 5–10 fold when human fibroblasts become senescent (Seshadri *et al.,* 1993).

Unlike bacterial rp genes, genes encoding mammalian rps are not clustered in the genome and therefore their transcription rate is determined by the individual promoter strength. Approximately equal level of transcription of all rp genes could be achieved by sharing a similar promoter structure. Mammalian rp promoters are distinguished from other polymerase II promoters by having essential elements located over only a 200 bp region up- and downstream of a cap site including intragenic sequences in the first exon and the first intron (Hariharan *et al.,* 1989; Meyuhas and Klein, 1990). In addition, they lack a consensus TATA box, but in many cases, A+T rich region in positions –20 to –30 is present. Interestingly, in mouse *S16*, this A+T rich sequence functions as an essential functional promoter element (Hariharan and Perry, 1990).

To date, expression of the *L32* gene is one of the best studied. Four *cis*-acting sequence elements (two of them upstream and two downstream of the TSP) and binding of three *trans*-acting factors to these sequences have been confirmed experimentally (Hariharan *et al.,* 1989; Yoganathan *et al.,* 1992; Curcic *et al.,* 1997). In contrast to other rp genes, rat *L19* has an important promoter element farther upstream, between –210 and –250 (Davies and Fried, 1995).

A common feature of the rp genes and some other genes related to protein synthesis is a polypyrimidine tract surrounding TSP, which is involved in transcriptional as well as in translational regulation (Levy *et al.,* 1991; Avni *et al.,* 1994). Polypyrimidine tract usually consists of an uninterrupted sequence of about 7–13 pyrimidine nucleotides with a C residue at a cap site. It determines the position of the transcription start site; one or several purine substitutions in the 5' pyrimidine tract causes the displacement of the major TSP (Chung and Perry, 1991).

Regulation of human *S14* transcription involves two antisense RNAs from intron 1 as positive regulators and S14 protein itself as a negative regulator (Tasheva and Roufa, 1995). S14 binds to antisense regulatory RNAs and to its own mRNA at the 5' noncoding region. The exact molecular mechanism by which the antisense RNAs stimulate and S14 protein inhibits transcription of its own mRNA is unclear. Similar short antisense transcripts have been detected from introns of human *S17* and *S24* (Chen and Roufa, 1988; Xu and Roufa, 1996). In addition, although *S14* promoter contains the TATA motif 26 bp upstream of the first exon, its transcription does not require this sequence (Rhoads and Roufa, 1987).

Rp promoter structure seems to be modular; various rp promoters are not composed of exactly the same set of modules, but each combination creates a promoter that appears to have a similar efficiency and allows concerted expression of rp genes. Sequential deletion of regulatory factor-binding sites results in a progressive loss of promoter activity.

Best characterized factors regulating rp gene transcription are RFX-1 (α), GABP (β), γ and δ (YY-1/NF-E1/UCRBP). δ factor is a zinc-finger protein (Hariharan *et al.*, 1991). Factor γ binds to the A+T rich sequence in *L32* promoter and may play a role similar to TATA-binding protein in transcription initiation (Yoganathan *et al.*, 1993). α and β have been previously characterized as proteins regulating transcription of different genes (Safrany and Perry, 1995; Yoganathan *et al.*, 1992). β factor binds to a DNA element with a purine-rich core sequence 5'-GGAA-3'. Putative binding sites for these factors are present in the promoter region of many rp genes. It seems that the set of transcription factors collectively specifies transcription level and though the major determinants have been identified in some cases (Chung and Perry, 1993; Safrany and Perry, 1995), none of the factors has complete predominance. Therefore, even when the main factor is limiting, the other factors could insure sufficient activity of rp genes.

Translational regulation

In growth-arrested cells, about 80% of rp mRNAs are present in the form of translationally inactive ribonucleoprotein particles (mRNPs) (Meyuhas *et al.*, 1987). Translational control involves fast regulation of the translationally active fraction of rp mRNA according to protein synthesis rate in the cell. After mitogenic stimulation, majority of rp mRNAs becomes associated with polysomes (Kaspar *et al.*, 1990). Rp gene expression is regulated at the translational level in early development (Amaldi *et al.*, 1989), in response to various stimuli (Hammond and Bowman, 1988) and in differentiation (Agrawal and Bowman, 1987). Translational regulation is mediated primarily, but not only by a 5' terminal oligopyrimidine tract (TOP) of rp mRNAs (Mariottini and Amaldi, 1990; Levy *et al.*, 1991; Hammond *et al.*, 1991; Kaspar *et al.*, 1992; Avni *et al.*,

1994). This polypyrimidine tract may be of variable sequence and length, but there is only one mammalian rp gene described with nonpyrimidine TSP; namely mouse *S24*, which has a cap site at a G residue (Xu *et al.*, 1994). The 5' TOP mediated translational regulation mechanism is conserved between the animal and plant kingdom (Shama and Meyuhas, 1996).

Some *trans*-acting factors that bind sequences in the 5' UTR have been described, namely p56 in mouse (Severson *et al.*, 1995); La (Pellizzoni *et al.*, 1996) and cellular nucleic acid binding protein CNBP in *Xenopus* (Pellizzoni *et al.*, 1997). An additional regulatory factor Ro60 that is necessary for the binding of La and CNBP proteins was identified recently (Pellizzoni *et al.*, 1998). Despite these results, the mechanism of translational regulation is still poorly understood.

Regulation of rp gene expression could even be more complicated. For example, three different species of mouse *S24* mRNA are produced in a cell-specific manner through alternative splicing (Xu *et al.*, 1994). Alternative S24 proteins contain an extra lysine or tripeptide PKE at the C-terminus. It is supposed that this tripeptide is an essential part of a nuclear localization signal and its lack, destroying the NLS, leaves in differentiated cells S24 into the cytoplasm instead of being transported to the nucleus (Vladimirov *et al.*, 1996).

Certain rp mRNAs are overexpressed in different cancers (Barnard *et al.*, 1993). Systematic data whether this results in overproduction of rps and contributes to the malignant transformation, or is only reflection of abnormal transcriptional regulation, is lacking. In overexpression experiments with up to 34 extra copies of *S16* and *L32* in mouse myoblasts, accumulation of respective mRNAs increases proportionally to the gene copy number, but the normal level of rps is maintained by rapid degradation of overproduced protein (Bowman, 1987).

2.1.3. Ribosomal proteins and their function(s)

Mammalian ribosome has a mass of about 4,5 MDa, but 80 proteins constitute only one third of this. 40S subunit contains one molecule of 18S RNA and 30 proteins, 60S subunit contains 5S, 5,8S and 28S RNA and 50 proteins. Ribosomal proteins are small and basic, with average molecular weight about 18,5 kDa (ranging from 3,5 to 47). Rps are very ancient molecules and at the same time well conserved evolutionarily. For instance, *E. coli* L2 can be replaced by its human or archaebacterial homolog without affecting translational activity (Uhlein *et al.*, 1998).

More than 15 bacterial rp structures have been determined using NMR spectroscopy and X-ray crystallography (Ramakrishnan and White, 1998). For eukaryotic rps, no 3-dimensional structure information is available. However, even prokaryotic structures reveal very interesting information. None of the proteins studied contains a folded domain with a molecular weight over

17 kDa. Larger molecules contain two subdomains that are joined together. Genetic events like gene fusion (S5, S8 and L9), insertion (S4 and L1) and gene duplication (L6) have probably happened before the emergence of the present-day ribosome. The structure of L30 and L7/L12 is very similar, suggesting that these molecules are evolutionary related. A number of predicted DNA and RNA binding motifs have been also confirmed by structural analyses. S27, S29, L37, L37a, S27a and L40 have zinc finger domains and have a potential to bind to nucleic acids. It is not clear whether these motifs are functional today or reflect a former function of these proteins. For example, in S29 the zinc finger motif can be found in homologs from different species, however not always entirely intact (Chan *et al.,* 1993). Several rps (L7, L9, L12, L13, L35, L37, L37a,S2, S9) have motifs related to bZIP (basic region leucine zipper) element that mediates protein dimerization and nucleic acid binding (Wool *et al.,* 1995). Mammalian rps from different species are very similar, 32 of them are identical between rat and human. Most probably, rps from all eukaryotic species are homologs, because even yeast and mammalian proteins have average identity of 60% (Wool *et al.,* 1995; Mager *et al.,* 1997).

A number of experimental data indicate that individual rps could modulate different cellular processes (reviewed in Wool *et al.,* 1995; Wool, 1996). The list of rps with extraribosomal functions is extending and only few of them are mentioned here as examples. L7 is involved in translational control; the bZIP domain in L7 mediates the formation of homodimers, which inhibit translation by specifically binding distinct mRNAs (Neumann *et al.,* 1995). In addition, constitutive expression of this protein leads to cell cycle arrest in G_1 and induces apoptosis by a hypothetic pathway where L7 inhibits specifically translation of anti-apoptotic proteins (Neumann and Krawinkel, 1997). Suppression of previously enhanced S3a expression can trigger apoptosis (Naora *et al.,* 1998). S3 can act as an endonuclease in the DNA repair system. Evidence that rps can perform activities not related to protein synthesis has raised the question whether rps have been compiled from preexisting proteins with other cellular functions rather than evolved as specifically ribosomal components.

Drosophila Minute mutants can be considered as examples of multicellular organism with limited amount of individual rps. *Drosophila* heterozygous for *Minute* mutations (approximately 50 loci scattered throughout the genome) exhibit distinctive dominant complex phenotypes including delayed larval development, reduced viability and fertility, reduced body size and short, thin bristles. In a number of cases it has been confirmed that *Minute* phenotypes are due to mutations affecting rp genes, including *rp49/L32* (Kongsuwan *et al.,* 1985), *L9* (Schmidt *et al.,* 1996), *L19* (Hart *et al.,* 1993), *S3* (Andersson *et al.,* 1994), *S3a* (van Beest *et al.,* 1998), *S5* (McKim *et al.,* 1996), *S13* (Sæbøe-Larssen and Lambertsson, 1996). Similar but not identical phenotypes are associated with reduced levels of various components required for protein synthesis, for example inactivation of ribosomal RNA genes (Kay and Jacobs-Lorena,

1987). *Minutes* are non-additive, providing additional indirect evidence that affected genes code for proteins with similar functions that act in a concerted way. Investigating partial revertants of a strong *Minute* phenotype produced by P-element insertion in the promoter region of the gene encoding rp S3, a correlation between reduced S3 mRNA levels and severity of the *Minute* fenotype was shown (Sæbøe-Larssen *et al.,* 1998). It has been also demonstrated that *Minute* phenotype can be rescued by a cloned rp gene (Kongsuwan *et al.,* 1985; Schmidt *et al.,* 1996). In addition, it is possible to generate the *Minute* phenotype by expression of antisense rp mRNA (Qian *et al.,* 1988; Patel and Jacobs-Lorena, 1992). Any mutation affecting ribosome structure or assembly might result in a *Minute* phenotype, but in the same time it is unlikely that any mutation in any rp gene results in similar phenotype.

Human S4X and S4Y proteins function interchangeably in ribosomes, although the Y isoform makes up only ~10% of the total male S4 protein. *S4X* is one of seven housekeeping genes identified that has an Y-linked homolog and is not dosage compensated; it escapes X inactivation. It has been suggested that haploinsufficiency of S4 contributes to the Turner syndrome, a consequence of X-chromosome monosomy (Fisher *et al.,* 1990, Kenmochi *et al.,* 1998). The role of S4 in Turner syndrome is under suspicion but is not ruled out completely. Interestingly, rat S4 is identical to human S4X isoform. This means that the selective pressure has operated preferentially on S4X rather than Y isoform although they are most probably derived from a single common ancestor gene. Mouse *S4* is X-linked and undergoes inactivation. Y-linked *S4* homologs are present only in primates.

Large subunit contains three proteins that have distinct features: P0, P1 and P2. These proteins are acidic, phosphorylated and they are related to *E. coli* L10 and L12. P2 binds to the elongation factor 2 (Vard *et al.,* 1997).

In *Saccharomyces cerevisiae*, loss of most of the rp genes one by one is lethal. A database of yeast protein characteristics can be accessed at http://www.proteome.com/YPDhome.html. Effect of the lack of a single protein could be crucial in the ribosome assembly, where a copy of each of the proteins is required. Ribosome assembly involves a series of independent reactions and therefore deficiency of a single protein could lead to assembly termination. Qsr1p (L10 according to rat nomenclature), one of the few proteins added to 60S subunits in the cytoplasm, is essential for ribosome subunit joining, thereby controlling translational initiation (Dick *et al.,* 1997).

Determination of three-dimensional structure of the 80S rat ribosome at 25Å resolution (Dube *et al.,* 1998a) and rabbit reticulocyte ribosome at 21Å resolution (Dube *et al.,* 1998b) revealed that eukaryotic and prokaryotic ribosomes are very similar in the intersubunit cavity area but significant differences were observed in the outer surface. A larger mass and number of proteins in the mammalian ribosome results in a significant number of protrusions, extending in all directions. The large flat part on the outer surface of the large subunit

(FRS, flat ribosomal surface) appears to be an ideal surface for docking on the endoplasmic reticulum.

Although described poorly, it is known that the number and primary structure of mitochondrial rps is different from the cytoplasmic ones. Of 34 yeast mitoribosomal large subunit proteins identified, 13 show a high degree of similarity to *E. coli* rps and 19 are not related to any other rps (Kitakawa *et al.*, 1997; Graack and Wittmann-Liebold, 1998).

Knowledge of the role of post-translational processing of rps in mammalian cells is limited. Tissue-specific methylation of the rat L29 has been observed, but the function of this modification is not known (Williamson *et al.*, 1997). S6, which phosphorylation correlates with upregulated translation of especially 5' TOP mRNAs is most probably one of the targets of mitogenic signalling pathway (Peterson and Schreiber, 1998).

S7, the object of the present study, specifically interacts with 28S rRNA (Nygård and Nika, 1982) and large subunit protein L7 (Witte and Krawinkel 1997). Interaction of L7 and S7 is probably mediated by the bZIP domain of L7 and the putative leucine zipper domain of S7. Binding of S7 and L7 suggests that these proteins are important for 40S and 60S subunit interaction during the translation.

2.1.4. Ribosome assembly

Ribosomal subunits are assembled in the nucleolus. Therefore, 5S rRNA, which is synthesized outside the nucleolus and rps, synthesized in the cytoplasm, must be transported to the nucleoli. After a number of maturation steps, preribosomal subunits are exported to the cytoplasm where a few late-assembly proteins join them. Although the primary mechanism regulating ribosome synthesis is control of rRNA synthesis (Krauter *et al.*, 1979), diverse mechanisms regulate also the production of rps, as discussed in section 2.1.2.

During maturation and assembly with the 80 rps, pre-rRNA transiently associates with about 150 different snoRNA species that assist cleavage, nucleotide modification (methylation and pseudouridinylation) and formation of the correct structure of the pre-rRNA (for review see Tollervey and Kiss, 1997). While in plants and yeast most of snoRNAs are synthesized from polycistronic transcripts (Leader *et al.*, 1997), a large majority of vertebrate snoRNAs is encoded within introns of genes for the proteins that play a role in ribosome synthesis (as nucleolin) or rps (Cecconi *et al.*, 1996; Rebane *et al.*, 1998). In some cases, however, mammalian snoRNAs are encoded in introns of the genes which do not encode protein in their exonic sequences (Tycowski, 1996). Although the gene that houses human U19 snoRNA does not code for a protein in its exons, 5' terminus of its RNA is very similar to polypyrimidine stretch found in vertebrate rp genes (Bortolin and Kiss, 1998). This may

indicate that expression of snoRNA-coding genes and rp genes is coordinated by similar promoter elements.

The control of ribosomal synthesis involves two main problems: First, coordination of the production of four rRNAs and 80 proteins. Second, how cells regulate the number of ribosomes appropriate for the growth rate or nutritional conditions. Short-term regulation of ribosomal production is apparently achieved at levels of translation, assembly and turnover of rps, while long-term coordination depends on uniform promoter strength.

If the synthesis of a single rp or RNA is terminated, most other proteins and RNAs of the same subunit are degraded (Fabian and Hopper, 1987; Moritz *et al.*, 1990; Wittekind *et al.*, 1990). Therefore, turnover of ribosomal constituents results upon disassembly of incompletely assembled subunits. Yeast mutant ribosomes defective in rp L16 (one of rps highly conserved between eu- and prokaryotes) are defective in assembly rather than function, indicating the crucial role of assembly in ribosome biogenesis (Moritz *et al.*, 1991).

2.2. Nucleocytoplasmic transport

2.2.1. Mechanism of nucleocytoplasmic transport

Biogenesis of ribosomes involves transport across the nuclear envelope on both directions. The assembly of preribosomal subunits in the nucleolus requires initial import of rps to the nucleus. After that, subunits are exported to the cytoplasm.

The directed movement of macromolecules into and out of the nucleus occurs through the nuclear pore complexes (NPCs). NPC is highly dynamic, the largest supramolecular complex in the eukaryotic cell with a mass about 125 MDa (for recent reviews, see Doye and Hurt, 1997; Ohno *et al.*, 1998). NPC has a central channel with a functional diameter of up to 26 nm for active transport. In principle, small molecules up to 60 kDa can move through the NPC by passive diffusion, but actually, transport of most proteins including histones (Breeuwer and Golfarb, 1990) and rps (Rout *et al.*, 1997) and all RNA species is mediated by a highly selective, active transport mechanism.

Nuclear import is mediated by nuclear localization sequences within the imported proteins and transport factors known as karyopherins (in yeast) or importins. Import of NLS-bearing proteins can be divided into two steps: i) binding at the NPC proteins (nucleoporins), a process dependent on the presence of cytosolic components and ii) translocation across the central gated channel of the NPC, a step that requires energy and is assisted by additional factors like small GTPase Ran and Ran-interacting proteins (for detailed review, see Mattaj and Englmeyer, 1998).

The model of the first discovered and best characterized so called "classical" nuclear import pathway involves the importin α/β heterodimer receptor complex (Enenkel *et al.*, 1995). Importin α binds the import substrate with the "classical" NLS and β docks the trimeric complex to the NPC. In nucleoplasm, importin β is released due to its direct binding to Ran-GTP (Moroianu *et al.*, 1996). Since importin α alone has lower affinity for the NLS than does the α/β heterodimer, the import substrate is released from the complex.

Crystallographic analysis of the yeast karyopherin α in the absence and presence of a "classical" monopartite SV40 T antigen NLS has revealed that binding of the NLS peptide occurs at two conserved sites within the helical surface groove (Conti *et al.*, 1998). The larger binding site recognizes five basic residues, the smaller site binds specifically only two residues. These two binding sites match the bipartite NLS consensus. The distance that separates the two binding sites allows a 10 residue spacer between two basic clusters. This is in agreement with experimental data showing that the distance between two basic clusters can be increased but not decreased without affecting the nuclear transport (Robbins *et al.*, 1991). Aspartate and glutamate residues around the edges of the binding sites avoid the presence of amino acids other than lysine and arginine in the ligand. The crystal structure also explains the relative importance of each of the positions in the NLS and especially the requirement of the second lysine of the motif KKxK.

Pathways using different importin β homologs that bind to their NLSs directly rather than via importin-α like adaptor have been recently described (Ohno *et al.*, 1998). Several receptors of importin β family are involved in the nuclear transport of hnRNP and ribosomal proteins (Jakel and Görlich, 1998). Alternative import pathways use separate docking sites on the NPC (Mutvei *et al.*, 1992; Nehrbass *et al.*, 1993).

As the importin α contains an NLS by which it binds to importin β, importin α originally could have been a transport substrate itself. α-independent pathway seems to be evolutionary more ancient. During the course of evolution, several alternative pathways may have evolved to separate the import processes of major classes of import substrates. This reduces competition for import receptors between imported molecules.

Interestingly, nuclear export receptors, that mediate transport out of the nucleus, are also members of importin β family. The directionality of the transport is presumably achieved by assymmetric distribution of the GTPase Ran. Nuclear Ran is loaded with GTP, while cytoplasmic Ran is in the GDP-bound form (Richards *et al.* 1997).

2.2.2. Nuclear import signals

"Classical" NLS consists of 5–20 amino acids and contains several lysine and arginine residues but does not fit a tight consensus. Single cluster of positively charged residues is known as monopartite NLS [for example, import signal of SV40 large T antigen (Kalderon *et al.*, 1984)]. In bipartite NLS, two stretches of basic residues are separated by mutation-tolerant linker region of 10–12 amino acids. Best studied is a bipartite signal of the nucleolin (Robbins *et al.*, 1991). When fused to a normally cytoplasmic reporter protein, NLS is sufficient to direct the chimeric protein to the nucleus. Monopartite SV40 T antigen NLS mutant defective in nuclear import can be rescued by placing two basic amino acids 10 residues upstream of the NLS, thereby rendering defective monopartite NLS into functional bipartite signal (Makkerh *et al.*, 1996). The list of proteins that have different or more complex import signals than the "classical" one, is extending. For instance, the hnRNPA1 M9 domain rich in glycines and aromatic residues is capable to direct both import and export (Michael *et al.*, 1995).

Import signals for rps are very basic but in some cases have greater complexity than the classical NLS. Nuclear targeting sequences for three yeast rps, L3, L29 and L25 have been defined (Moreland *et al.*, 1985; Underwood and Fried, 1990; Schaap *et al.*, 1991). In L25, the NLS-containing region adds ~50% to the mass of the protein compared with its prokaryotic homologs. This suggests that the addition of NLSs may account in part for the larger mass of eukaryotic rps. These sequences are not very similar to the NLSs of substrates transported by the "classical" system. In contrast, three NLSs of human S6 are very similar to each other and to "classical" NLS (Schmidt *et al.*, 1995).

A number of studies have been described mechanisms controlling nuclear import by modulating NLS activity. The activity of NLS can be regulated in a tissue-specific manner or during development (Standiford and Richter, 1992). Nuclear import can be precisely regulated by such mechanisms as masking/unmasking of NLS (Zabel *et al.*, 1993), phosphorylation (Jans and Hubner, 1996) and cytoplasmic retention domains (Sittler *et al.*, 1996). The efficiency of the minimal NLS can be strongly affected by the protein context (Roberts *et al.*, 1987) or even by distant regions (Gao and Knipe, 1992). Multiple signals can cooperate to increase transport efficiency (Roberts *et al.*, 1987). Even the well-studied SV-40 T-antigen NLS requires domain N-terminal to the NLS for full activity (Rihs and Peters, 1989). As shown recently, neutral and even acidic amino acids can be essential for nuclear import (Makkerh *et al.*, 1996).

2.2.3. Import pathways for ribosomal proteins

It is estimated that about 100 rps are imported and 3 ribosomal subunits are exported through each (~5000) NPC per minute in a growing HeLa cell. In yeast, the maximum rate of rp nuclear import could be up to 15 molecules per NPC per second (Rout *et al.,* 1997). Rps are small enough to passively diffuse into the nucleus. However, free cytoplasmic rps are degraded rapidly within few minutes (Bowman, 1987) and they concentrate 50-fold in the nucleolus within 5 minutes of their synthesis. Therefore, they require active nuclear import what has been demonstrated in many cases (Moreland *et al.,* 1985; Underwood and Fried, 1990; Schaap *et al.,* 1991; Schmidt *et al.,* 1995; Russo *et al.,* 1997; Quaye *et al.,* 1996). Active and receptor-mediated transport makes nuclear entry more rapid and also prevents disadvantageous interactions in the cytoplasm. Some rps may be imported as complexes. For instance, nuclear localization of S3a has been confirmed with antibodies, although the protein itself does not have nuclear import ability (Hoth, 1993). Such "piggyback" mechanism is also involved in import of transcription factor E2F (De la Luna *et al.,* 1996) and yeast Cdc2 protein kinase (Booher *et al.,* 1989)

Nuclear import pathways used by rps have been studied in yeast and human cells. In a recent study, it was demonstrated that three human rps, L23a, S7 and L5 can each be imported alternatively by any of the four import receptors importin β, transportin, RanBP5 and RanBP7 (Jakel and Görlich, 1998). The classical α/β heterodimer pathway plays only a minor role. However, each of the rps behaves somewhat differently. S7 binds the importin α and apparently can also use the classical α/β heterodimer pathway. In addition, it was shown that L23a does not have different binding sites for each of the receptors. The nuclear import receptor binding is mediated by a 32 amino acid extremely basic domain (residues 32–74) which is absent in bacterial homologs of L23a but conserved among eukaryotes. C-terminal domain of L23a (residues 75–156) is conserved between eukaryotes, eu- and archaebacteria. One of rp import factors, transportin, mediates also import of hnRNP proteins, which have glycine-rich M9 domain import signal. It is interesting that transportin has distinct and nonoverlapping binding sites specific for very basic NLS-containing domain of L23a and M9 domain.

In yeast, nuclear import of rps is primarily mediated by a member of importin β family Kap123p/Yrb4p (Rout *et al.,* 1997). Although import receptors for human rps have been identified, mammalian homolog of this protein has not been described yet. Yeast but not *E. coli* rps bind with high affinity to purified Kap123. In the absence of Kap123p, another β importin Pse1p/Kap121p (the yeast homolog of human RanBP5), originally characterized as a factor affecting protein secretion (Chow *et al.,* 1992), substitutes it. Unlike Kap95p (yeast homolog of α-dependent importin β), Kap123p requires no partner but interacts

directly with its transport substrates. Kap123p does not associate with mature ribosomes, confirming the suggestion that to prevent recognition by the import factors, the NLSs are masked in the mature ribosome.

The fact that rps are imported by several transport factors suggests that nuclear import pathways have back-up mechanisms. Different import receptors act in concert as a partially redundant system, providing the flexibility necessary to coordinate the efficient nuclear import of rps under a variety of conditions. Interestingly, a single receptor can import very different substrates. Transportin has distinct binding sites for L23a and M9 domain and it mediates also import of other hnRNP proteins, which have no sequence similarity to M9 or L23a (Siomi *et al.*, 1997).

Export of ribosomal subunits is poorly understood. Kinetic competition analysis has revealed that a saturable factor in 40S subunit export is specific and is not limiting for tRNA export (Pokrywka and Goldfarb, 1995). However, 40S subunits can compete with tRNA for intranuclear RNA-binding retention sites. Cytoplasmic transport of microinjected ribosomal subunits from the *Xenopus* oocyte nucleus is a facilitated, saturable, energy-dependent process reaching a maximum rate of about 10^7 subunits per minute per oocyte (Bataille *et al.*, 1990). Surprisingly, ribosomal subunits from yeast and even *Escherichia coli* are efficiently exported from the *Xenopus* oocyte (Bataille *et al.*, 1990).

3. PRESENT INVESTIGATIONS AND DISCUSSION

3.1. Genomic structure of two mammalian ribosomal protein S7 genes (References I and III)

Isolation and primary sequence of the human S7 gene

At first, the human *S7* cDNA was isolated. Expression library screening with the rat S7-specific antisera yielded six positive clones. A complete sequencing of the largest clone revealed a single open reading frame of 577 nucleotides. Unfortunately, this human *S7* cDNA clone was missing the 5'-noncoding sequence, ATG start codon and following two nucleotides, but contained a canonical polyadenylation signal (AATAAA) 13 nt downstream of the TAA stop codon. To isolate a genomic clone, the human lymphocyte DNA library was screened using *S7* cDNA as a hybridization probe. Five clones were isolated. To exclude possible pseudogenes, restriction mapping and a hybridization using *S7* cDNA as the radioactive probe was carried out. Four clones were identified as pseudogenes, because their restriction pattern resembles that of the cDNA. One clone was suspected to contain the intron-containing *S7* and was selected for more detailed characterization. The restriction mapping with *Eco*RI, *Bam*HI, *Pst*I, *Hin*dIII, *Kpn*I and *Xma*I localized the *S7* into three *Hin*dIII fragments. Overlapping fragments covering the gene were subcloned and sequenced. Human *S7* gene spans about 5,5 kb and contains seven exons and six introns (Fig. 1 in reference I). Exons are 65–151 nt and introns 131–2318 nt in length. All splice donor and acceptor sites have perfect consensus (GT/AG) and the sequences immediately upstream of the 3' splice sites are pyrimidine-rich, which is common in mammalian introns (Mount, 1982). The initiation codon context GCCAUGU deviates a little from the optimal A/GCCAUGG (Kozak, 1986). The promoter region was analyzed for the presence of binding sites for common transcription factors and compared to other mammalian rp gene promoters. No consensus TATA or CAAT boxes were found, although a relatively A/T rich area is located between positions –50 and –36. Between nt –147... –142 and –116...–111 two binding sites for transcription factor GABP (CGGAAR) are present. Sequence CCTCGCGC at positions –19...–12 is similar to the conserved binding site for nuclear factor of the *L7a* promoter (TCTCGCGA) (Colombo and Fried, 1992).

The first exon of *S7* contains 72 nt of untranslated sequence and the ATG start codon is located in the second exon. While in most human rp genes described so far (*L7, L7a, S3, S4, S6, S8, S17* and *S24*) the starting Met and the following 1–3 aa are coded by the first exon, there are some cases in which the ATG codon resides in the second exon (*S14* and *S30*). As the first intron of the mouse *L32* has been demonstrated to contain a transcriptional regulatory

element (Chung and Perry, 1989), that, as suggested, must be located inside the approximately 200 bp promoter area, the notable shortness of the first exon of the mammalian rp genes may be important for transcriptional regulation.

Computer analysis for the distribution of CpG dinucleotides in *S7* revealed that the first three closely located exons are surrounded by a CpG rich region. Intron 4 includes two Alu-like sequences at nt 1480...1600 and 1780...2080; there is also an Alu-like sequence in the 5' flanking region at nt –580...–320 (Jurka and Milosavljevic, 1991).

Isolation and characterization of the mouse S7 gene

The gene encoding the mouse homolog of ribosomal protein S7 was isolated using PCR-based approach, assuming that the coding sequences of mammalian genes are conserved enough to allow amplification of mouse DNA with oligonucleotides designed according to human DNA sequence (Annilo *et al.,* 1998a). After testing several oligonucleotide combinations, primers T1 and N/P (see Fig. 1 in reference III), flanking intron 4 in human *S7* were chosen. PCR with these oligonucleotides generated two bands from mouse genomic DNA. The shorter PCR product was about 120 nucleotides in length, that corresponds to the distance predicted from the cDNA and was therefore assumed to be derived from the intronless pseudogenes. The larger product was about 500 nucleotides, indicating that the intron should exist between these two primers. This fragment was extracted, labelled and used to isolate the mouse *S7* gene from the genomic library.

The mouse *S7* gene is approximately 5 kb in length and exon sequences match the cDNA sequence perfectly (Fig. 1 in reference III). Vertebrate intron-containing *S7* genes from *Fugu rubripes* (Cecconi *et al.,* 1996), *Xenopus laevis* (Mariottini *et al.,* 1993), and *Homo sapiens* (Annilo *et al.,* 1995) have been described earlier. Introns in the vertebrate genes are located at identical positions. The largest *S7* gene is that of *X. laevis,* with introns about 2 to 3 times longer than in other vertebrates. In addition, genomic sequences from *Caenorhabditis elegans* (GenBank Z75714), *Saccharomyces cerevisiae* (GenBank Z71372 and X94335) and *Schizosaccharomyces pombe* (GenBank Z68198) are available. *C. elegans* gene is lacking introns at positions where vertebrates have the 2nd and 5th intron. Both copies of *C. cerevisiae S7* gene have one intron which position corresponds to the mammalian 3rd intron.

Interestingly, U17 snoRNA is processed from the introns of *Xenopus* (Cecconi *et al.,* 1996) and *Fugu* (Mariottini *et al.,* 1993) *S7* genes, but not from mammalian *S7* genes. Human and mouse U17 are encoded within the introns of U17 host gene *U17HG*, that have similar organization but do not have protein-coding potential (Pelczar and Filipowicz, 1998).

Mammalian ribosomal protein genes *S6*, *S14*, *S24* and *L7* have extensive homology in the first intron within each gene pair (Pata and Metspalu, 1996).

When we compared the introns of mouse and human *S7* genes, only a short homologous region within the first intron was detected (mouse nucleotides 163...189, 22 identities of 27), while mouse and human *S6* have 63% identity in a 350 nucleotide region. Whether this short homology reflects functional significance in *S7* remains unknown. The initiation codon in mouse *S7* is located at the second exon and has a context GCCAUGU which is similar to the optimal (Kozak, 1986). The polyadenylation signal is located 12 nucleotides downstream from the TAA stop codon.

The promoter region of mouse *S7*, like many other ribosomal protein gene promoters, lacks a canonical TATA box, but contains a number of consensus elements described in promoters of vertebrate ribosomal protein genes. Three GA-binding protein (GABP, β-factor) sites (consensus CGGAAR) in inverted orientation are located at positions –16...–21, -37...–42, and –78...–83 (Genuario and Perry, 1996). Sequences between nucleotides 55...65 (first untranslated exon) and 201...210 (antisense strand in the first intron) are very similar to the δ-factor binding element (consensus $^{C}/_{G}{}^{C}/_{G}C^{G}/_{T}$GCCATCT) described in the promoters of mouse *L30* and *L32* genes (Hariharan *et al.,* 1989). At positions 27...35 in the 5' untranslated region and 213...221 in the first intron are located the sequences with very high homology to *L7a* BoxA element, which in *L7a* is conserved between mammals and chicken (Colombo and Fried, 1992). The presence of transcription factor binding sequences supports the suggestion that the isolated gene is transcriptionally active. δ-elements, BoxA-like and GABP elements can be found in the promoter region of all vertebrate *S7* genes, suggesting that these are essential functional promoter elements. The region spanning first three exons of mouse *S7* gene is GC rich (G+C content 61% compared to 48% in the whole sequenced region). The ratio of CpG/GpC dinucleotides in this area is 0.96. These data indicate that, like in other rp genes, a CpG island at the 5' end of the *S7* gene is present.

Determination of the transcription start sites of human and mouse S7

The 5' end of the human and mouse mRNA was mapped by the primer extension assay. A single product obtained from human mRNA using primer E1 corresponds to a C residue (Fig. 2 in reference I), which is indicated as nt 1 in Fig. 1 in reference I. This cytidine is located within a pyrimidine tract of 12 nt in length, which, however, is interrupted by one G.

The major transcription initiation site of mouse *S7* gene is located at T residue 79 nucleotides upstream of the translation initiation codon. Minor bands at C nucleotides at 5 and 14 nucleotides farther upstream were also detected (Fig. 3 in reference III). These nucleotides are located within a pyrimidine-rich sequence; the main start point is embedded in a polypyrimidine tract of 13 nucleotides in length. Heterogeneity of start sites is not unique to mouse *S7*, it has also been observed in human *L19* (Davies and Fried, 1995) and *S14* (Rhoads *et al.,* 1986).

Genomic localization

Chromosomal localization of human *S7* was determined by PCR on the panel of the human/hamster somatic cell hybrid DNAs with intron-specific primers I6–5' (forward) and I6–3' (reverse). This experiment showed that *S7* maps on chromosome 2. The chromosomal position was specified by FISH. The locus for *S7* was assigned to 2p25, the most telomeric chromosomal band of the short arm of chromosome 2 (Fig. 4 in reference I). In addition to *S7*, at least three other rp genes are located on chromosome 2, namely *S27A*, *L31* and *L37A* (Kenmochi *et al.*, 1998). The chromosomal localization of the mouse *S7* gene has not been determined but most probably it is located on chromosome 12, within region homologous to human 2p25.

S7 gene family

The number of human *S7* gene copies was evaluated by Southern blot analysis. Hybridization of the radioactive cDNA probe to *Bcl*I, *Bst*EI, *Eco*RI and *Xba*I digests of HeLa DNA suggests that there are 8–10 copies of *S7* in the human genome.

Using fragment from intron 5 of the mouse *S7* as a probe, the bands with lengths as expected from the primary structure of the *S7* gene were detected. This demonstrates that intron-containing mouse *S7* gene is present in the single copy. The filter was stripped and then hybridized to mouse *S7* cDNA probe (Fig. 4 in reference III). Detection of multiple cDNA-hybridizing bands suggests that there are 10–12 additional copies (most likely processed pseudogenes) of *S7* in the mouse genome.

S7 protein

The protein predicted from the cDNA is rich in basic residues, contains 194 aa and has molecular weight of 22.1 kDa. The amino acid sequence of S7 was compared against the GenBank database using BLAST programs at http://www.ncbi.nlm.nih.gov/ (Altschul *et al.*, 1997). Predicted proteins for which complete cDNA or genomic sequence is known, are aligned in Fig. 2 in reference III. The human S7 is homologous to r-proteins isolated from other species: rat (Suzuki *et al.*, 1990), *Xenopus laevis* (Mariottini *et al.*, 1988), *Anopheles gambiae* (Salazar et al., 1993), *Manduca sexta* (EMBL L20096) and *Arabidopsis thaliana* (EMBL Z47607 and Z47625) (100%, 98%, 71%, 73% and 49% identity at an aa level, respectively). A large number of EST's potentially coding S7 homologs from different eukaryotic species is also available, but many of them are partial or possibly erroneous. No archaebacterial or eubacterial homolog was found by database similarity search. Two highly conserved regions of the protein (amino acids 98...109 and 115...118) are involved in the nuclear and nucleolar import of S7 (Annilo *et al.*, 1998b).

3.2. Analysis of nuclear and nucleolar localization of S7 (Reference II)

To identify the sequences responsible for nuclear and nucleolar targeting of ribosomal protein S7, constructs expressing fusion proteins of the wild-type or mutated S7 sequences with *E. coli* β-galactosidase were designed. Different cDNA fragments or oligonucleotides were cloned into the shuttle vector pKHlacZ (Figs. 1, 2A and 2C in reference II). SAOS-2 cells were electroporated with 2–4 μg of plasmid DNA and after 24 to 48 h post-transfection were stained with X-gal for β-galactosidase activity.

S7 accumulates the reporter protein within the nucleoli

Expression of the construct wtS7, lacking only the first two amino acids of S7 (starting Met and Phe), showed that S7 was able to transport the reporter protein to the nucleus, where it accumulates into the nucleoli (Fig. 3A). Partial nuclear staining of β-galactosidase in a control experiment (Fig. 3J) could be a technical artifact due to cytoplasmic staining above the nucleus (Russo *et al.*, 1997).

Deletions from the amino- or the carboxy-terminus were generated to investigate the role of different fragments of S7 in transport. Deletion of the first 80 amino acids (ΔN80) resulted in nuclear localization as in the case of the full-length S7 (Fig. 3B). The construct lacking amino-terminal 120 residues (ΔN120) or N-terminal 113 residues (Δ114C) showed reduced transport and staining was present in both the cytoplasm and nucleus (Fig. 3C and 3D). In addition, constructs ΔN120 and Δ114C were not capable of localizing into the nucleoli.

Requirements for nuclear targeting sequence in peptide-NLS and S7 protein are different

Sequence ^{115}Lys-Arg-Pro-Arg^{118} within the extremely conserved region of S7 is very similar to the nuclear import sequences in ribosomal protein S6 (Schmidt *et al.*, 1995). To verify its functionality, oligonucleotides encoding this peptide were inserted into the 5' end of lacZ gene. Indeed, efficient transport of β-galactosidase into the nucleus, but not to the nucleoli was observed (Fig. 3E). While the minimal S7 NLS ^{115}KRPR^{118} is very similar to the NLSs of ribosomal protein S6 (KRRR and KKPR), it differs from signals of yeast L29 (Underwood and Fried, 1990) and L3 (Moreland *et al.*, 1985) and rat L31 (Quaye *et al.*, 1995), which are longer and contain amino acids other than lysine, arginine and proline.

Impact of individual amino acids within the NLS-tetrapeptide and S7 protein to nuclear transport efficiency was studied as a next step. Every single amino

acid was substituted at a conservative or nonconservative manner (Fig 2C). As a conservative mutation we considered another basic amino acid, as a nonconsevative one, asparagine.

In a tetrapeptide, all mutations except for proline substitutions reduced nuclear transport (Figs. 3F-3H). Interestingly, impact of the mutations in the S7 protein context was different. Only proline[117] appeared to be critical, while substitution of any of the basic amino acids did not impair transport (Figs. 4A-C). Possible explanations are: 1) presence of a turn in a polypeptide chain due to proline is essential for nuclear targeting signal function within a mature S7, but not in the tetrapeptide; and 2) the context of the full-length protein can support and strengthen the partially defective NLS. Furthermore, none of the S7 mutations abolished nucleolar localization, suggesting that residues within the NLS that are essential for effective nuclear import are not critical for nucleolar accumulation once inside the nucleus.

Analysis of the nucleolar accumulation

To find out, which region besides the NLS controls the nucleolar targeting of S7, either the highly conserved region 78...86, which is rich in charged residues (S7-ΔI, see Fig. 2B), the arginine-lysine rich region 98...109 (S7-ΔII), the minimal NLS (S7-ΔIII), or the motif similar to NLS ^{142}KRIR^{145} (S7-ΔIV) was deleted. The results demonstrated that regions 78...86 (S7-ΔI) and 142...145 (S7-ΔVI) can be deleted without any impact upon nuclear or nucleolar location (Figs. 4D, 4G). Deletion of amino acids 98...109 (S7-ΔII) and 115...118 (minimal NLS, S7-ΔIII) prevented nucleolar accumulation and resulted in a distribution of the fusion protein between the nuclear and cytoplasmic compartment (Figs. 4E, 4F). Since none of these deletions resulted in completely or preferentially cytoplasmic location, combinations of deletions I and II, II and III, and II and IV were constructed. S7-ΔI/II and S7-ΔII/IV had cytoplasmic staining similar to S7-ΔII (Figs. 4H, 4J). Simultaneous deletion of amino acids 98...109 and 115...118 (S7-ΔII/III, containing 9 basic residues of 16) affected nuclear import most severely, resulting in preferentially cytoplasmic location (Fig. 4I).

These experiments demonstrate that effective nuclear and nucleolar accumulation of S7 requires both the minimal NLS and an adjacent basic region. To test whether the domains necessary for nucleolar targeting have functional significance outside of the S7 protein context, location of the chimeric protein S7/98-118 was studied (Fig. 3I). Indeed, these 21 amino acids were sufficient for nuclear transport and nucleolar targeting of the reporter protein. Within the full-length S7, unlike when out of the protein context, the tetrapeptide ^{115}KRPR^{118} alone is not sufficient for the nuclear import. Constructs that contained the minimal NLS but not the sequence ^{98}RRILPKPTRKSR^{109} (S7-ΔII, S7-ΔI/II and S7-ΔII/IV) showed, in addition to nuclear, also cytoplasmic

staining. Fusions Δ114C and S7-ΔIII, in which the minimal NLS was deleted, were not preferentially cytoplasmic, but also showed considerable nuclear accumulation. Only when the minimal NLS and region 98–109 were simultaneously deleted (S7-ΔII/III), localization of the resulting fusion protein was similar to the β-galactosidase staining of the control experiment.

Interestingly, point mutations within the S7 protein S7-M3 and S7-M5 (Figs. 4B, 4C) had stronger effect than deletion of the whole minimal signal S7-ΔIII (Fig. 4F). One possible explanation is that mutant NLS can bind to its receptor but can not promote nuclear targeting, thereby reducing the free cytoplasmic pool of receptors. Deletion of the NLS may also change the conformation of the protein and thereby expose otherwise masked sequence with nuclear targeting ability. In addition, import of rps is carried out by at least four different receptors of importin β family (Jäkel and Görlich, 1998). Different affinity of these receptors to different mutants can explain the fact that most of the deletions and point mutations resulted only in reduced transport but not in complete block of import.

CONCLUSIONS

To understand the mechanisms responsible for the stoichiometric synthesis of ribosomal components and flexible accommodation of the ribosome biosynthesis to the cell requirements for protein synthesis, determination of the primary structure and genomic organization of the ribosome components is fundamental.

1. The intron-containing genes for the human and mouse ribosomal protein S7 have been cloned and sequenced. As the sequenced genes contain introns and promoter region and their exons perfectly match cDNA, it can be concluded that the isolated genes are functional. The sequences can be obtained from GenBank, accession #Z25749 (human *S7* gene) and #AF043285 (mouse *S7* gene). These genes share the features typical to other mammalian ribosomal protein genes — lack of consensus TATA box on the promoter area, short untranslated leader, small first exon, CpG rich 5' region and presence of the family of the pseudogenes. Human *S7* was mapped to the short arm of the 2nd chromosome (2p25). The transcription start sites of *S7* genes are located at the pyrimidine residues within the oligopyrimidine tract.

2. S7 protein regions responsible for nuclear and nucleolar accumulation were identified. Despite the fact that S7 contains a tetrapeptide sequence ^{115}KRPR^{118}, which, when fused to the β-galactosidase, can function as an independent NLS, effective nuclear and nucleolar transport of the S7 protein requires the presence of another basic domain ^{98}RRILPKPTRKSR^{109}. In addition, a 21- amino acid S7 fragment containing these domains is sufficient for nuclear and nucleolar accumulation of the β-galactosidase fusion. The conservation of the nucleolar accumulation domain indicates that this region may control nuclear and nucleolar accumulation of all S7 homologs and that the mechanism leading to the nucleolar accumulation may also be conserved.

REFERENCES

Agrawal, M. G. and Bowman, L. H. (1987) Transcriptional and translational regulation of ribosomal protein formation during mouse myoblast differentiation. *J. Biol. Chem.* **262**, 4868–4875.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.

Amaldi, F., Bozzoni, I., Beccari, E. and Pierandrei-Amaldi, P. (1989) Expression of ribosomal protein genes and regulation of ribosome biosynthesis in *Xenopus* development. *Trends Biochem. Sci.* **14**, 175–178.

Andersson, S., Saeboe-Larssen, S., Lambertsson, A., Merriam, J. and Jacobs-Lorena, M. (1994) A *Drosophila* third chromosome *Minute* locus encodes a ribosomal protein. *Genetics* **137**, 513–520.

Annilo, T., Laan, M., Stahl, J. and Metspalu, A. (1995) The human ribosomal protein S7-encoding gene: isolation, structure and localization in 2p25. *Gene* **165**, 297–302.

Annilo, T., Jelina, J., Pata, I. and Metspalu, A. (1998a) Isolation and characterization of the mouse ribosomal protein S7 gene. *Biochem. Mol. Biol. Internat.* **46**, 287–295.

Annilo, T., Karis, A., Hoth, S., Rikk, T., Kruppa, J. and Metspalu, A. (1998b) Nuclear import and nucleolar accumulation of the human ribosomal protein S7 depends on both a minimal nuclear localization sequence and an adjacent basic region. *Biochem. Biophys. Res. Commun.* **249**, 759–766.

Antoine, M. and Fried, M. (1992) The organization of the intron-containing human S6 ribosomal protein (rpS6) gene and determination of its location at chromosome 9p21. *Hum. Mol. Genet.* **1**, 565–570.

Armes, N. and Fried, M. (1996) Surfeit locus gene homologs are widely distributed in invertebrate genomes. *Mol. Cell Biol.* **16**, 5591–5596.

Armes, N., Gilley, J. and Fried, M. (1997) The comparative genomic structure and sequence of the surfeit gene homologs in the puffer fish Fugu rubripes and their association with CpG-rich islands. *Genome Res.* **7**, 1138–1152.

Avni, D., Shama, S., Loreni, F. and Meyuhas, O. (1994) Vertebrate mRNAs with a 5'-terminal pyrimidine tract are candidates for translational repression in quiescent cells: characterization of the translational cis-regulatory element. *Mol. Cell. Biol.* **14**, 3822–3833.

Bataille, N., Helser, T. and Fried, H. M. (1990) Cytoplasmic transport of ribosomal subunits microinjected into the Xenopus laevis oocyte nucleus: a generalized, facilitated process. *J. Cell Biol.* **111**, 1571–1582.

Barnard, G. F., Staniunas, R. J., Mori, M., Puder, M., Jessup, M. J., Steele, G. D. Jr. and Chen, L. B. (1993) Gastric and hepatocellular carcinomas do not overexpress the same ribosomal protein messenger RNAs as colonic carcinoma. *Cancer Res.* **53**, 4048–4052.

Baum, E. Z. and Wormington, W. M. (1985) Coordinate expression of ribosomal protein genes during *Xenopus* development. *Dev. Biol.* **111**, 488–498.

Booher, R. N., Alfa, C. E., Hyams, J. S. and Beach, D. H. (1989) The fission yeast cdc2/cdc13/suc1 protein kinase: regulation of catalytic activity and nuclear localization. *Cell* **58**, 485–497.

Bortolin, M. L. and Kiss, T. (1998) Human U19 intron-encoded snoRNA is processed from a long primary transcript that possesses little potential for protein coding. *RNA* **4**, 445–454.

Bowman, L. H. (1987) The synthesis of ribosomal proteins S16 and L32 is not autogenously regulated during mouse myoblast differentiation. *Mol. Cell Biol.* **7**, 4464–4471.

Breeuwer, M. and Goldfarb, D. (1990) Facilitated nuclear transport of histone H1 and other small nucleophilic proteins. *Cell* **60**, 999–1008.

Cecconi, F., Crosio, C., Mariottini, P., Cesareni, G., Giorgi, M., Brenner, S. and Amaldi, F. (1996) A functional role for some *Fugu* introns larger than the typical short ones: the example of the gene coding for ribosomal protein S7 and snoRNA U17. *Nucleic Acids Res.* **24**, 3167–3172.

Chan, Y.L. and Wool, I.G. (1991) The structure of a gene containing introns and encoding rat ribosomal protein P2. *Nucleic Acids Res.* **19**, 4895–4900.

Chan, Y. L., Suzuki, K., Olvera, J. and Wool, I. G. (1993) Zinc finger-like motifs in rat ribosomal proteins S27 and S29. *Nucleic Acids Res.* **21**, 649–655.

Chen, I. T. and Roufa, D. J. (1988) The transcriptionally active human ribosomal protein S17 gene. Gene 70, 107–116.

Chow, T. Y., Ash, J. J., Dignard, D. and Thomas, D. Y. (1992) Screening and identification of a gene, PSE-1, that affects protein secretion in *Saccharomyces cerevisiae*. *J. Cell Sci.* **101**, 709–719.

Chung, S. and Perry, R. P. (1989) Importance of introns for expression of mouse ribosomal protein gene rpL32. *Mol. Cell Biol.* **9**, 2075–2082.

Chung, S. and Perry, R. P. (1991) Cell-free transcription of a mouse ribosomal-protein-encoding gene: the effects of promoter mutations. *Gene* **100**, 173–180.

Chung, S. and Perry, R. P. (1993) The importance of downstream delta-factor binding elements for the activity of the rpL32 promoter. *Nucleic Acids Res.* **21**, 3301–3308.

Colombo, P., Yon, J. and Fried, M. (1991) The organization and expression of the human L7a ribosomal protein gene. *Biochim. Biophys. Acta* **1129**, 93–95.

Colombo, P. and Fried, M. (1992) Functional elements of the ribosomal protein L7a (rpL7a) gene promoter region and their conservation between mammals and birds. *Nucleic Acids Res.* **20**, 3367–3373.

Conti, E., Uy, M., Leighton, L., Blobel, G. and Kuriyan, J. (1998) Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin α. *Cell* **94**, 193–204.

Curcic, D., Glibetic, M., Larson, D. E. and Sells, B. H. (1997) GA-binding protein is involved in altered expression of ribosomal protein L32 gene. *J. Cell. Biochem.* **65**, 287–307.

Davies, B. and Fried, M. (1993) The structure of the human intron-containing S8 ribosomal protein gene and determination of its chromosomal location at 1p32–p34.1. *Genomics* **15**, 68–75.

Davies, B. and Fried, M. (1995) The L19 ribosomal protein gene (RPL19): gene organization, chromosomal mapping, and novel promoter region. *Genomics* **25**, 372–380.

De Falco, S., Russo, G., Angiolillo, A. and Pietropaolo, C. (1993) Human L7a ribosomal protein: sequence, structural organization, and expression of a functional gene. *Gene* **126**, 227–235.

De la Luna, S., Burden, M. J., Lee, C. W. and La Thangue, N. B. (1996) Nuclear accumulation of the E2F heterodimer regulated by subunit composition and alternative splicing of a nuclear localization signal. *J. Cell Sci.* **109**, 2443–2452.

Dick, F. A., Eisinger, D. P. and Trumpower, B. L. (1997) Exchangeability of Qsr1p, a large ribosomal subunit protein required for subunit joining, suggests a novel translational regulatory mechanism. *FEBS Lett.* **419**, 1–3.

Doye, V. and Hurt, E. (1997) From nucleoporins to nuclear pore complexes. *Curr. Opin. Cell Biol.* **9**, 401–411.

Dube, P., Wieske, M., Stark, H., Schatz, M., Stahl, J., Zemlin, F., Lutsch, G. and van Heel, M. (1998a) The 80S rat liver ribosome at 25Å resolution by electron cryomicroscopy and angular reconstitution. *Structure* **6**, 389–399.

Dube, P., Bacher, G., Stark, H., Mueller, F., Zemlin, F., van Heel, M. and Brimacombe, R. (1998b) Correlation of the expansion segments in mammalian rRNA with the fine structure of the 80 S ribosome; a cryoelectron microscopic reconstruction of the rabbit reticulocyte ribosome at 21 A resolution. *J. Mol. Biol.* **279**, 403–421.

Dudov, K. P. and Perry, R. P. (1984) The gene family encoding the mouse ribosomal protein L32 contains a uniquely expressed intron-containing gene and an unmutated processed gene. *Cell* **37**, 457–468.

Enenkel, C., Blobel, G. and Rexach, M. (1995) Identification of a yeast karyopherin heterodimer that targets import substrate to mammalian nuclear pore complexes. *J. Biol. Chem.* **270**, 16499–16502.

Fabian, G. R. and Hopper, A. K. (1987) *RRP1*, a *Saccharomyces cerevisiae* gene affecting rRNA processing and production of mature ribosomal subunits. *J. Bacteriol.* **169**, 1571–1578.

Faliks, D. and Meyuhas, O. (1982) Coordinate regulation of ribosomal protein mRNA level in regenerating rat liver. Study with the corresponding mouse cloned cDNAs. *Nucleic Acids Res.* **10**, 789–801.

Fan, W., Christensen, M., Eichler, E., Zhang, X. and Lennon, G. (1997) Cloning, sequencing, gene organization, and localization of the human ribosomal protein RPL23A gene. *Genomics* **46**, 234–239.

Filipenko, M. L., Vinichenko, N. A., Karpova, G. G., Mertvetsov, N. P. and Amaldi, F. (1998) Isolation, structural analysis and mapping of the functional gene of human ribosomal protein S26. *Gene* **211**, 287–292.

Fisher, E. M. C., Beer-Romeo, P., Brown, L.G., Ridley, A., McNeil, J. A., Lawrence, J. B., Willard, H. F., Bieber, F. R. and Page, D. C. (1990) Homologous ribosomal protein genes on the human X and Y chromosomes: escape from X inactivation and possible implications for Turner syndrome. *Cell* 63, 1205–1218.

Gao, M. and Knipe, D. M. (1992) Distal protein sequences can affect the function of a nuclear localization signal. *Mol. Cell Biol.* **12**, 1330–1339.

Genuario, R. R. and Perry, R. P. (1996) The GA-binding protein can serve as both an activator and repressor of ribosomal protein gene transcription. *J. Biol. Chem.* **271**, 4388–4395.

Graack, H. R. and Wittmann-Liebold, B. (1998) Mitochondrial ribosomal proteins (MRPs) of yeast. *Biochem. J.* **329**, 433–448.

Hammond, M. L. and Bowman, L. H. (1988) Insulin stimulates the translation of ribosomal proteins and the transcription of rDNA in mouse myoblasts. *J. Biol. Chem.* **263**, 17785–17791.

Hammond, M. L., Merrick, W. and Bowman, L. H. (1991) Sequences mediating the translation of mouse S16 ribosomal protein mRNA during myoblast differentiation and in vitro and possible control points for the in vitro translation. *Genes. Dev.* **5**, 1723–1736.

Hariharan, N., Kelley, D. E. and Perry, R. P. (1989) Equipotent mouse ribosomal protein promoters have a similar architecture that includes internal sequence elements. *Genes Dev.* **3**, 1789–1800.

Hariharan, N. and Perry, R. P. (1990) Functional dissection of a mouse ribosomal protein promoter: significance of the polypyrimidine initiator and an element in the TATA-box region. *Proc. Natl. Acad. Sci. USA* **87**, 1526–1530.

Hariharan, N. Kelley, D. E. and Perry, R. P. (1991) Delta, a transcription factor that binds to downstream elements in several polymerase II promoters, is a functionally versatile zinc finger protein. *Proc. Natl. Acad. Sci. USA* **88**, 9799–9803.

Hart, K., Klein, T. and Wilcox, M. (1993) A *Minute* encoding a ribosomal protein enhances wing morphogenesis mutants. *Mech. Dev.* **43**, 101–110.

Hoth, S. (1993) Zur Biogenese von Ribosomen: S6-Genstruktur, S3a-Primärstruktur und -Kernimport. Dissertation, Universität Hamburg.

Huxley, C. and Fried, M. (1990) The mouse rpL7a gene is typical of other ribosomal protein genes in it's 5' region but differs in being located in a tight cluster of CpG-rich islands. *Nucl. Acids Res.* **18**, 5353–5357.

Imai, T., Sudo, K. and Miwa, T. (1994) Assignment of the human ribosomal protein S25 gene (RPS25) to chromosome 11q23.3 by sequence analysis of the marker D11S456. *Genomics* **20**, 142–143.

Jakel, S. and Görlich, D. (1998) Importin beta, transportin, RanBP5 and RanBP7 mediate nuclear import of ribosomal proteins in mammalian cells. *EMBO J.* **17**, 4491–4502.

Jans, D. A. and Hubner, S. (1996) Regulation of protein transport to the nucleus: central role of phosphorylation. *Physiol. Rev.* **76**, 651–685.

Jurka, J. and Milosavljevic, A. (1991) Reconstruction and analysis of human Alu genes. J. Mol. Evol. 32, 105–121.

Kalderon, D., Roberts, B. L., Richardson, W. D. and Smith, A. E. (1984) A short amino acid sequence able to specify nuclear location. *Cell* 39, 499–509.

Kas, K., Michiels, L. and Merregaert, J. (1992) Genomic structure and expression of the human fau gene: encoding the ribosomal protein S30 fused to a ubiquitin-like protein. *Biochem. Biophys. Res. Commun.* **187**, 927–933.

Kaspar, R. L., Rychlik, W., White, M. W., Rhoads, R. E. and Morris, D. R. (1990) Simultaneous cytoplasmic redistribution of ribosomal protein L32 mRNA and phosphorylation of eukaryotic initiation factor 4E after mitogenic stimulation of Swiss 3T3 cells. *J. Biol. Chem.* **265**, 3619–3622.

Kaspar, R. L., Kakegawa, T., Cranston, H., Morris, D. R. and White, M. W. (1992) A regulatory cis element and a specific binding factor involved in the mitogenic control of murine ribosomal protein L32 translation. *J. Biol. Chem.* **267**, 508–514.

Kay, M. A. and Jacobs-Lorena, M. (1987) Developmental genetics of ribosome synthesis in *Drosophila*. *Trends Genet.* **3**, 347–351.

Kenmochi, N., Kawaguchi, T., Rozen, S., Davis, E., Goodman, N., Hudson, T. J., Tanaka, T. and Page, D. C. (1998) A map of 75 human ribosomal protein genes. *Genome Res.* 8, 509–523.

Kitakawa, M., Graack, H. R., Grohmann, L., Goldschmidt-Reisin, S., Herfurth, E., Wittmann-Liebold, B., Nishimura, T. and Isono, K. (1997) Identification and characterization of the genes for mitochondrial ribosomal proteins of *Saccharomyces cerevisiae*. *Eur. J. Biochem.* 245, 449–456.

Kongsuwan, K., Yu, Q., Vincent, A., Frisardi, M. C., Rosbash, M., Lengyel, J. A. and Merriam, J. (1985) A *Drosophila Minute* gene encodes a ribosomal protein. *Nature* 317, 555–558.

Kozak, M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292.

Krauter, K. S., Soeiro, R. and Nadal-Ginard, B. (1979) Transcriptional regulation of ribosomal RNA accumulation during L6E9 myoblast differentiation. *J. Mol. Biol.* **134**, 727–741.

Kuwano, Y., Olvera, J. and Wool, I. G. (1992) The primary structure of rat ribosomal protein S5. A ribosomal protein present in the rat genome in a single copy. *J. Biol. Chem.* **267**, 25304–25308.

Larson, D. E., Zahradka, P. and Sells, B. H. (1991) Control points in eucaryotic ribosome biogenesis. *Biochem. Cell. Biol.* **69**, 5–22.

Leader, D. J., Clark, G. P., Watters, J., Beven, A. F., Shaw, P. J. and Brown, J. W. (1997) Clusters of multiple different small nucleolar RNA genes in plants are expressed as and processed from polycistronic pre-snoRNAs. *EMBO J.* **16**, 5742–5751.

Levy, S., Avni, D., Hariharan, N., Perry, R. P. and Meyuhas, O. (1991) Oligopyrimidine tract at the 5' end of mammalian ribosomal protein mRNAs is required for their translational control. *Proc. Natl. Acad. Sci. USA* **88**, 3319–3323.

Mager, W. H., Planta, R. J., Ballesta, J. G., Lee, J. C., Mizuta, K., Suzuki, K., Warner, J. R. and Woolford, J. (1997) A new nomenclature for the cytoplasmic ribosomal proteins of *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **25**, 4872–4875.

Makkerh, J. P. S., Dingwall, C. and Laskey, R. A. (1996) Comparative mutagenesis of nuclear localization signals reveals the importance of neutral and acidic amino acids. *Curr. Biol.* **6**, 1025–1027.

Mariottini, P., Bagni, C., Annesi, F. and Amaldi, F. (1988) Isolation and nucleotide sequences of cDNAs for Xenopus laevis ribosomal protein S8: similarities in the 5' and 3' untranslated regions of mRNAs for various r-proteins. *Gene* 67, 69–74.

Mariottini, P. and Amaldi, F. (1990) The 5' untranslated region of mRNA for ribosomal protein S19 is involved in its translational regulation during *Xenopus* development. *Mol. Cell Biol.* **10**, 816–822.

Mariottini, P., Bagni, C., Francesconi, A., Cecconi, F., Serra, M. J., Chen, Q. M., Loreni, F., Annesi, F. and Amaldi, F. (1993) Sequence of the gene coding for ribosomal protein S8 of *Xenopus laevis*. *Gene* 132, 255–260.

Mattaj, I. W. and Englmeier, L. (1998) Nucleocytoplasmic transport: the soluble phase. *Annu. Rev. Biochem.* **67**, 265–306.

Mazuruk, K., Scoen, T. J., Chader, G. J., Iwata, T. and Rodriguez, I. R. (1996) Structural organization and chromosomal localization of the human ribosomal protein L9 gene. *Biochem. Biophys. Acta* **1305**, 151–162.

McKim, K. S., Dahmus, J. B. and Hawley, R. S. (1996) Cloning of the *Drosophila melanogaster* meiotic recombination gene mei-218: a genetic and molecular analysis of interval 15E. *Genetics* **144**, 215–228.

Meyuhas, O., Thompson, E. A. Jr. and Perry, R. P. (1987) Glucocorticoids selectively inhibit translation of ribosomal protein mRNAs in P1798 lymphosarcoma cells. *Mol. Cell. Biol.* **7**, 2691–2699.

Meyuhas, O. and Klein, A. (1990) The mouse ribosomal protein L7 gene. Its primary structure and functional analysis of the promoter region. *J. Biol. Chem.* **265**, 11465–11473.

Michael, W. M., Choi, M. and Dreyfuss, G. (1995) A nuclear export signal in hnRNP A1: a signal-mediated, temperature-dependent nuclear protein export pathway. *Cell* **83**, 415–422.

Moreland, R. B., Nam, H. G., Hereford, L. M. and Fried, H. M. (1985) Identification of a nuclear localization signal of a yeast ribosomal protein. *Proc. Natl. Acad. Sci. USA* **82**, 6561–6565.

Moritz, M., Paulovich, A. G., Tsay, Y.F. and Woolford, J. L. Jr (1990) Depletion of yeast ribosomal proteins L16 or rp59 disrupts ribosome assembly. *J. Cell Biol.* **111**, 2261–2274.

Moritz, M., Pulaski, B. A. and Woolford, J. L. (1991) Assembly of 60S ribosomal subunits is perturbed in temperature-sensitive yeast mutants defective in ribosomal protein L16. *Mol. Cell. Biol.* **11**, 5681–5692.

Moroianu, J., Blobel, G. and Radu, A. (1996) Nuclear protein import: Ran-GTP dissociates the karyopherin alphabeta heterodimer by displacing alpha from an overlapping binding site on beta. *Proc. Natl. Acad. Sci. USA* **93**, 7059–7062.

Mount, S. M. (1982) A catalogue of splice junction sequences. *Nucleic Acids Res.* **10**, 459–472.

Mundus, D. A., Bulygin, K. N., Ven'yaminova, A. G., Vladimirov, S. N., Vratskikh, L. V., Repkova, M. N., Yamkovoi, V. S. and Karpova, G. G. (1993) Human placenta 40S ribosomal subunit affinity modification by oligoribonucleotide derivatives containing the AUG codon. Ribosomal proteins composing the codon-anticodon interaction site. *Mol. Biol.* 27, 91–95.

Mutvei, A., Dihlmann, S., Herth, W. and Hurt, E. C. (1992) NSP1 depletion in yeast affects nuclear pore formation and nuclear accumulation. *Eur. J. Cell Biol.*, **59**, 280–295.

Naora, H., Takai, I., Adachi, M. and Naora, H. (1998) Altered cellular responses by varying expression of a ribosomal protein gene: sequential coordination of enhancement and suppression of ribosomal protein S3a gene expression induces apoptosis. *J. Cell Biol.* **141**, 741–753.

Nehrbass, U., Fabre, E., Dihlmann, S., Herth, W. and Hurt, E. C. (1993) Analysis of nucleo-cytoplasmic transport in a thermosensitive mutant of nuclear pore protein NSP1. *Eur. J. Cell Biol.* **62**, 1–12.

Neumann, F., Hemmerich, P., von Mikecz, A., Peter, H. H. and Krawinkel, U. (1995) Human ribosomal protein L7 inhibits cell-free translation in reticulocyte lysates and affects the expression of nuclear proteins upon stable transfection into Jurkat T-lymphoma cells. *Nucleic Acids Res.* **23**, 195–202.

Neumann, F. and Krawinkel, U. (1997) Constitutive expression of human ribosomal protein L7 arrests the cell cycle in G_1 and induces apoptosis in Jurkat T-lymphoma cells. *Exp. Cell Res.* **230**, 252–261.

Nitta, I., Kamada, Y., Noda, H., Ueda, T. and Watanabe, K. (1998) Reconstitution of peptide bond formation with *Escherichia coli* 23S ribosomal RNA domains. *Science* **281**, 666–669.

Nolte, D., Taimor, G., Kalff-Suske, M. and Seifart, K. H. (1996) The human S3a ribosomal protein: sequence, location and cell-free transcription of the functional gene. *Gene* **169**, 179–851.

Nygård, O. and Nika, H. (1982) Identification by RNA-protein cross-linking of ribosomal proteins located at the interface between the small and the large subunits of mammalian ribosomes. *EMBO J.* **1**, 357–362.

Ohno, M., Fornerod, M. and Mattaj, I.W. (1998) Nucleocytoplasmic transport: the last 200 nanometers. *Cell* **92**, 327–336.

Pata, I., Hoth, S., Kruppa, J. and Metspalu, A. (1992) The human ribosomal protein S6 gene: isolation, primary structure and location in chromosome 9. *Gene* **121**, 387–392.

Pata, I. and Metspalu, A. (1996) Structural characterization of the mouse ribosomal protein S6-encoding gene. *Gene* **175**, 241–245.

Patel, R. and Jacobs-Lorena, M. (1992) Generation of *Minute* phenotypes by a transformed antisense ribosomal protein gene. *Dev. Genet.* **13**, 256–263.

Pelczar, P. and Filipowicz, W. (1998) The host gene for intronic U17 small nucleolar RNAs in mammals has no protein-coding potential and is a member of the 5'-terminal oligopyrimidine gene family. *Mol. Cell Biol.* **18**, 4509–4518.

Pellizzoni, L., Cardinali, B., Lin-Marq, N., Mercanti, D. and Pierandrei-Amaldi, P. (1996) A *Xenopus laevis* homologue of the La autoantigen binds the pyrimidine tract of the 5' UTR of ribosomal protein mRNAs in vitro: implication of a protein factor in complex formation. *J. Mol. Biol.* **259**, 904–915.

Pellizzoni, L., Lotti, F., Maras, B. and Pierandrei-Amaldi, P. (1997) Cellular nucleic acid binding protein binds a conserved region of the 5' UTR of *Xenopus laevis* ribosomal protein mRNAs. *J. Mol. Biol.* **267**, 264–275.

Pellizzoni, L., Lotti, F., Rutjes, S. A. and Pierandrei-Amaldi, P. (1998) Involvement of the *Xenopus laevis* Ro60 autoantigen in the alternative interaction of La and CNBP proteins with the 5'UTR of L4 ribosomal protein mRNA. *J. Mol. Biol.* **281**, 593–608.

Peterson, R. T. and Schreiber, S. L. (1998) Translation control: connecting mitogens and the ribosome. *Curr. Biol.* **8**, 248–250.

Planta, R. J. and Mager, W. H. (1998) The list of cytoplasmic ribosomal proteins of *Saccharomyces cerevisiae*. *Yeast* **14**, 471–477.

Pokrywka, N. J. and Goldfarb, D. S. (1995) Nuclear export pathways of tRNA and 40 S ribosomes include both common and specific intermediates. J. *Biol. Chem.* **270**, 3619–3624.

Polakiewicz, R. D., Munroe, D. J., Sait, S. N., Tycowski, K. T., Nowak, N. J., Shows, T. B., Housman, D. E. and Page, D. C. (1995) Mapping of ribosomal protein S3 and internally nested snoRNA U15A gene to human chromosome 11q13.3-q13.5. *Genomics* **25**, 577–580.

Qian, S., Hongo, S. and Jacobs-Lorena, M. (1988) Antisense ribosomal protein gene expression specifically disrupts oogenesis in *Drosophila melanogaster. Proc. Natl. Acad. Sci. USA* **85**, 9601–9605.

Quaye, I. K., Toku, S. and Tanaka, T. (1996) Sequence requirement for nucleolar localization of rat ribosomal protein L31. *Eur. J. Cell Biol.* **69**, 151–155.

Ramakrishnan, V. and White, S. W. (1998) Ribosomal protein structures: insights into the architecture, machinery and evolution of the ribosome. *Trends Biol. Sci.* **6**, 208–212.

Rebane, A., Tamme, R., Laan, M., Pata, I. and Metspalu, A. (1998) A novel snoRNA (U73) is encoded within the introns of the human and mouse ribosomal protein S3a genes. *Gene* **210**, 255–263.

Rhoads, D. D., Dixit, A. and Roufa, D. J. (1986) Primary structure of human ribosomal protein S14 and the gene that encodes it. *Mol. Cell Biol.* **6**, 2774–2783.

Rhoads, D. D. and Roufa, D. J. (1987) A cloned human ribosomal protein gene functions in rodent cells. *Mol. Cell. Biol.* **7**, 3767–3774.

Rhoads, D. D. and Roufa, D. J. (1991) Molecular evolution of the mammalian ribosomal protein gene, RPS14. *Mol. Biol. Evol.* **8**, 503–514.

Richards, S. A., Carey, K. L. and Macara, I. G. (1997) Requirement of guanosine triphosphate-bound ran for signal-mediated nuclear protein export. *Science* **276**, 1842–1844.

Rihs, H. P. and Peters, R. (1989) Nuclear transport kinetics depend on phosphorylation-site-containing sequences flanking the karyophilic signal of the Simian virus 40 T-antigen. *EMBO J.* **8**, 1479–1484.

Robbins, J., Dilworth, S. M., Laskey, R. A. and Dingwall, C. (1991) Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. *Cell* **64**, 615–623.

Roberts, B. L., Richardson, W. D. and Smith, A. E. (1987) The effect of protein context on nuclear location signal function. *Cell* **50**, 465–475.

Rout, M. P., Blobel, G. and Aitchison, J. D. (1997) A distinct nuclear import pathway used by ribosomal proteins. *Cell* **89**, 715–725.

Russo, G., Ricciardelli, G. and Pietropaolo, C. (1997) Different domains cooperate to target the human ribosomal L7a protein to the nucleus and to the nucleoli. *J. Biol. Chem.* **272**, 5229–5235.

Safrany, G. and Perry, R. P. (1995) The relative contributions of various transcription factors to the overall promoter strength of the mouse ribosomal protein L30 gene. *Eur. J. Biochem.* **230**, 1066–1072.

Salazar, C. E., Mills-Hamm, D., Kumar, V. and Collins, F. H. (1993) Sequence of a cDNA from the mosquito *Anopheles gambiae* encoding a homologue of human ribosomal protein S7. *Nucleic Acids Res.* **21**, 4147.

Schaap, P. J., van't Riet, J., Woldringh, C. L. and Raue, H. (1991) Identification and functional analysis of the nuclear localization signals of ribosomal protein L25 from *Saccharomyces cerevisiae*. *J. Mol. Biol.* **221**, 225–237.

Schmidt, C., Lipsius, E. and Kruppa, J. (1995) Nuclear and nucleolar targeting of human ribosomal protein S6. Mol. Biol. Cell 6, 1875–1885.

Schmidt, A., Hollmann, M. and Schafer, U. (1996) A newly identified *Minute* locus, M(2)32D, encodes the ribosomal protein L9 in *Drosophila melanogaster. Mol. Gen. Genet.* **251**, 381–387.

Scholnick, J., Sinor, C., Oakes, J., Outten, W. and Saha, M. (1997) Differential expression of *Xenopus* ribosomal protein gene XlrpS1c. *Biochim. Biophys. Acta* **1354**, 72–82.

Seshadri, T., Uzman, J. A., Oshima, J. and Campisi, J. (1993) Identification of a transcript that is down-regulated in senescent human fibroblasts. Cloning, sequence analysis, and regulation of the human L7 ribosomal protein gene. *J. Biol. Chem.* **268**, 18474–1880.

Severson, W. E., Mascolo, P. L. and White, M. W. (1995) Lymphocyte p56L32 is a RNA/DNA-binding protein which interacts with conserved elements of the murine L32 ribosomal protein mRNA. *Eur. J. Biochem.* **229**, 426–432.

Sæbøe-Larssen, S. and Lambertsson, A. (1996) A novel *Drosophila Minute* locus encodes ribosomal protein S13. *Genetics* **143**, 877–885.

Sæbøe-Larssen, S., Lyamouri, M., Merriam, J., Oksvold, M. P. and Lambertsson, A. (1998) Ribosomal protein insufficiency and the *Minute* syndrome in *Drosophila*: A dose-response relationship. *Genetics* **148**, 1215–1224.

Shama, S. and Meyuhas, O. (1996) The translational cis-regulatory element of mammalian ribosomal protein mRNAs is recognized by the plant translational apparatus. *Eur. J. Biochem.* **236**, 383–388.

Shiga, K., Yamamoto, H. and Okamoto, H. (1990) Isolation and characterization of the human homologue of rig and its pseudogenes: the functional gene has features characteristic of housekeeping genes. *Proc. Natl. Acad. Sci. USA* **87**, 3594–3598.

Siomi, M. C., Eder, P. S., Kataoka, N., Wan, L., Liu, Q. and Dreyfuss, G. (1997) Transportin-mediated nuclear import of heterogenous RNP proteins. *J. Cell Biol.* **138**, 1181–1192.

Sittler, A., Devys, D., Weber, C. and Mandel, J. L. (1996) Alternative splicing of exon 14 determines nuclear or cytoplasmic localisation of fmr1 protein isoforms. *Hum. Mol. Genet.* **5**, 95–102.

Standiford, D. M. and Richter, J. D. (1992) Analysis of a developmentally regulated nuclear localization signal in *Xenopus*. *J. Cell Biol.* **118**, 991–1002.

Suzuki, K., Olvera, J. and Wool, I. G. (1990) The primary structure of rat ribosomal protein S7. *FEBS Lett.* **271**, 51–53.

Takasawa, S., Tohgo, A., Unno, M., Shiga, K., Yonekura, H. and Okamoto, H. (1992) The primary structure of rat rig/ribosomal protein S15 gene. *Biochim. Biophys. Acta* **1132**, 228–230.

Tasheva, E. S. and Roufa, D. J. (1995) Regulation of human *RPS14* transcription by intronic antisense RNAs and ribosomal protein S14. *Genes. Dev.* **9**, 304–316.

Tollervey, D. and Kiss, T. (1997) Function and synthesis of small nucleolar RNAs. *Curr. Opin. Cell Biol.* **9**, 337–342.

Tycowski, K. T., Shu, M. D. and Steitz, J. A. (1996) A mammalian gene with introns instead of exons generating stable RNA products. *Nature* **379**, 464–466.

Uhlein, M., Weglohner, W., Urlaub, H. and Wittmann-Liebold, B. (1998) Functional implications of ribosomal protein L2 in protein biosynthesis as shown by in vivo replacement studies. *Biochem. J.* **331,** 423–430.

Underwood, M. R. and Fried, H. M. (1990) Characterization of nuclear localizing sequences derived from yeast ribosomal protein L29. *EMBO J.* **9**, 91–99.

van Beest, M., Mortin, M. and Clevers, H. (1998) *Drosophila* RpS3a, a novel minute gene situated between the segment polarity gene *scubitus interruptus* and dTCF. *Nucleic Acids Res.* **26**, 4471–4475.

Vanin, E. F. (1985) Processed pseudogenes: characteristics and evolution. *Annu. Rev. Genet.* **19**, 253–272.

Vard, C., Guillot, D., Bargis, P., Lavergne, J. P. and Reboud, J. P. (1997) A specific role for the phosphorylation of mammalian acidic ribosomal protein P2. *J. Biol. Chem.* **272**, 20259–20262.

Vladimirov, S. N., Ivanov, A. V., Karpova, G. G., Musolyamov, A. K., Egorov, T.A., Thiede, B., Wittmann-Liebold, B. and Otto, A. (1996) Characterization of the human small-ribosomal-subunit proteins by N-terminal and internal sequencing, and mass spectrometry. *Eur. J. Biochem.* 239, 144–149.

Wagner, M. and Perry, R. P. (1985) Characterization of the multigene family encoding the mouse S16 ribosomal protein: strategy for distinguishing an expressed gene from its processed pseudogene counterparts by an analysis of total genomic DNA. *Mol. Cell Biol.* **5**, 3560–3576.

Weiner, A. M., Deininger, P. L. and Efstratiadis, A. (1986) Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu. Rev. Biochem.* **55**, 631–661.

Westermann, P. and Nygård, O. (1983) The spatial arrangement of the complex between eukaryotic initiation factor eIF-3 and 40 S ribosomal subunit. Cross-linking between factor and ribosomal proteins. *Biochim. Biophys. Acta* **741**, 103–108.

Wiedemann, L. M. and Perry, R. P. (1984) Characterization of the expressed gene and several processed pseudogenes for the mouse ribosomal protein L30 gene family. *Mol. Cell Biol.* **4**, 2518–2528.

Williamson, N. A. Raliegh, J. Morrice, N. A. and Wettenhall, R. E. (1997) Post-translational processing of rat ribosomal proteins. Ubiquitous methylation of Lys22 within the zinc-finger motif of RL40 (carboxy-terminal extension protein 52) and tissue-specific methylation of Lys4 in RL29. *Eur. J. Biochem.* **246**, 786–793.

Witte, S. and Krawinkel, U. (1997) Specific interactions of the autoantigen L7 with multi-zinc finger protein ZNF7 and ribosomal protein S7. *J. Biol. Chem.* **272**, 22243–22247.

Wittekind, M., Kolb, J. M., Dodd, J., Yamagishi, M., Memet, S., Buhler, J. M. and Nomura, M. (1990) Conditional expression of RPA190, the gene encoding the largest subunit of yeast RNA polymerase I: effects of decreased rRNA synthesis on ribosomal protein synthesis. *Mol. Cell Biol.* **10**, 2049–2059.

Wool, I. G. (1979) The structure and function of eukaryotic ribosomes. *Ann. Rev. Biochem.* **48**, 719–754.

Wool, I. G. (1996) Extraribosomal functions of ribosomal proteins. *Trends Biochem. Sci.* **21**, 164–165.

Wool, I. G., Chan, Y.-L. and Glück, A. (1995) Structure and evolution of mammalian ribosomal proteins. *Biochem. Cell Biol.* **73**, 933–947.

Xu, L., He, G. P., Li, A. and Ro, H. S. (1994) Molecular characterization of the mouse ribosomal protein S24 multigene family: a uniquely expressed intron-containing gene with cell-specific expression of three alternatively spliced mRNAs. *Nucleic Acids Res.* **22**, 646–655.

Xu, W. B. and Roufa, D. J. (1996) The gene encoding human ribosomal protein S24 and tissue-specific expression of differentially spliced mRNAs. *Gene* **169**, 257–262.

Yoganathan, T., Bhat, N. K. and Sells, B. H. (1992) A positive regulator of the ribosomal protein gene, beta factor, belongs to the ETS oncoprotein family. *Biochem. J.* **287**, 349–353.

Yoganathan, T., Horikoshi, M., Roeder, R. G. and Sells, B. H. (1993) Direct binding of yeast transcription factor (TFIID) to the ribosomal protein L32 (rpL32) TATA-less promoter sequence. *FEBS Lett.* **326**, 163–166.

Zabel, U., Henkel, T., Silva, M. S. and Baeuerle, P. A. (1993) Nuclear uptake control of NF-kappa B by MAD-3, an I kappa B protein present in the nucleus. *EMBO J.* **12**, 201–211.

Zinn, A. R., Alagappan, R. K., Brown, L. G., Wool, I. and Page, D. C. (1994) Structure and function of ribosomal protein S4 genes on the human and mouse sex chromosomes. *Mol. Cell Biol.* **14**, 2485–2492.

IMETAJATE RIBOSOOMIVALK S7: HOMOLOOGSETE GEENIDE ISOLEERIMINE NING VALGU LOKALISATSIOON RAKUS

Kokkuvõte

Ribosoomide biogeneesi rakus reguleeritakse kõigil etappidel, alates ribosoomi komponentide geenide transkriptsioonist kuni subühikute ühinemiseni translatsiooni alustamiseks. Käesolevas töös on uuritud kahe imetaja — inimese ja hiire — ribosoomivalgu S7 geeni struktuuri ning identifitseeritud valgu piirkonnad, mis on vajalikud S7 transpordiks tuuma ning akumulatsiooniks tuumakesse, kus valgud assambleeruvad ribosomaalse RNAga.

1. Kirjeldati ribosoomivalgu S7 geeni inimesel ning hiirel. Hiire geeni isoleerimise kaugem eesmärk on ribosoomivalgu defitsiidi tõttu tekkida võivate häirete hindamine mudelhiire põhjal. Et *Drosophila*'l on kirjeldatud umbes 50 sarnase fenotüübiga *Minute* mutanti, millest seni molekulaarsel tasemel iseloomustatuist enamus on seotud ribosoomivalkudega, siis on alust arvata, et ka imetajatel võivad ribosoomivalkude ekspressiooni muutused olla suurte pärilike defektide põhjuseks.

Isoleeritud geenid on sarnase pikkusega, 5 kuni 5,5 tuhat aluspaari, koosnedes seitsmest eksonist ja kuuest intronist. *S7* geeni intronid paiknevad imetajatel identsetes positsioonides. Isoleeritud introneid sisaldav geen esineb nii inimese kui ka hiire genoomis ühe koopiana, protsessitud pseudogeene on mõlemal juhul kümmekond.

2. Punkt- ning deletsioonmutageneesi teel on identifitseeritud S7 piirkond, mis määrab valgu transpordi rakutuuma ja tuumakesse. S7 transport tuuma ja tuumakesse sõltub kaheosalisest signaalist (aminohapped 98...109 ja 115...118). Aminohappejärjestus 115...118 sisaldab nn. klassikalist tuumalokalisatsiooni signaali, mis on funktsionaalne ka iseseisvalt. See järjestus on võimeline küll importima valgu rakutuuma, kuid mitte tuumakesse. Järjestus 98...109 muudab NLSi toime tugevamaks ja kindlustab valgu liikumise tuumakesse, ribosoomide sünteesi kohta. Sama järjestus iseseisvalt toimib aga samuti ebaefektiivselt. Täispika S7 valgu puhul on seega efektiivseks akumulatsiooniks tuumakesse vajalikud mõlemad valgu piirkonnad.

Transport rakutuuma toimub vastavate retseptorite abil, mis tunnevad ära kindla aminohapete järjestuse imporditavas valgus — tuumalokalisatsiooni signaali. Edasine kogunemine tuumakesse sõltub tõenäoliselt funktsionaalsetest interaktsioonidest ribosoomi komponentide või ribosoomi assambleerimist assisteerivate faktoritega.

ACKNOWLEDGEMENTS

This work was carried out at the Institute of Molecular and Cell Biology, University of Tartu. I wish to express my gratitude to my supervisor Dr. Andres Metspalu for his advice, help and surpassing optimism.

To all my past and present colleagues in the lab, for fruitful cooperation and help, especially Dr. Illar Pata, Ana Rebane, Dr. Ants Kurg, Dr. Alar Karis, Tiina Rikk and Jelena Jelina.

To Dr. Maris Laan for performing fluorescence *in situ* hybridization.

To my German colleagues Prof. Dr. Joachim Kruppa and Dr. Stefan Hoth.

To Hiljar Sibul for critically reading the manuscript.

To the colleagues and teachers in Institute for friendly and creative atmosphere.

To Heidi Saulep, Viljo Soo, Marika Lauringson and Krista Liiv for technical help.

To my family for love and understanding.

PUBLICATIONS

I

Annilo, T., Laan, M., Stahl, J. and Metspalu, A. (1995)
The human ribosomal protein S7-encoding gene:
isolation, structure and localization in 2p25.
Gene **165**, 297–302.

Gene, 165 (1995) 297–302
 0378-1119/95/$09.50

GENE 09123

The human ribosomal protein S7-encoding gene: isolation, structure and localization in 2p25

(*Homo sapiens*; transcription initiation; fluorescence in situ hybridization; protein homology)

Tarmo Annilo[a], Maris Laan[b], Joachim Stahl[c], Andres Metspalu[a]

[a]*Institute of Molecular and Cell Biology, Tartu University, Estonian Biocentre, 23 Riia St., EE2400 Tartu, Estonia;* [b]*Department of Clinical Chemistry, University of Helsinki, Helsinki, Finland. Tel. (358-0) 471-4339;* [c] *Max-Delbrück Zentrum für Molekulare Medizin, Berlin, Germany. Tel. (49-30) 946-3233*

Received by J.L. Slightom: 4 April 1995; Accepted: 3 May 1995; Received at publishers: 7 June 1985

SUMMARY

We have identified a gene encoding the human ribosomal protein (r-protein) S7. The *S7* gene contains seven exons and six introns spanning about 6 kb. Organization of the gene is similar to that of *Xenopus laevis S8*, the only homologous intron-containing gene isolated so far. An mRNA transcribed from this gene has an open reading frame (ORF) of 582 nucleotides (nt), which encodes a protein of 194 amino acids (22.1 kDa). The transcription start point (*tsp*) was mapped by a primer extension assay to a C residue within a pyrimidine-rich tract. Human S7 (hS7) is identical to rat S7 (rS7) and exhibits significant similarity with the *X. laevis*, insect and plant homologs. We have used fluorescence in situ hybridization (FISH) to localize *S7* to chromosome 2p25.

INTRODUCTION

The eukaryotic ribosome consists of more than 80 different r-proteins and four RNA species. Mammalian *rp* genes are dispersed throughout the genome (Feo et al., 1992) and are present in multiple copies, from which only one contributes to the synthesis of mRNA, with the exception of human r-protein S4X/Y genes where both genes are transcribed and translated into proteins (Zinn et al., 1994). Each gene is a single functional unit containing all the essential regulatory elements and therefore the balanced production of the r-proteins requires a complicated mechanism involving certain common control points. A very high degree of similarity manifested by the r-proteins from different organisms is the quality that enables to uncover homologous r-proteins even from the distant species.

There is increasing ground of belief that some r-proteins may serve functions other than participation in protein synthesis. Mammalian S27, S29, L7, L37, L37a and a number of prokaryotic r-proteins have been reported to have zinc-finger-like or other DNA binding domains (Hemmerich et al., 1993; Chan et al., 1993). Apurinic/apyrimidinic endonuclease (AP) activity has been demonstrated in the case of *Drosophila melanogaster* S3 (Wilson et al., 1993) and AP3, homolog of the human P0 (Grabowski et al., 1991). At present the r-proteins are supposed to be a supplement to the earlier only RNA-containing ribosome (Wool, 1993) and these findings may be signs pointing to the former functions of these proteins.

A matter of great interest is the possible association of

Correspondence to: Dr. A. Metspalu, Institute of Molecular and Cell Biology, Tartu University, Estonian Biocentre, 23 Riia St., EE2400 Tartu, Estonia. Tel. (372-7) 420210; Fax (372-7) 420286; e-mail: andres@ebc.ee

Abbreviations: aa, amino acid(s); bp, base pair(s); DAPI, 4′-6′diamino-2-phenylindole; FISH, fluorescence in situ hybridization; hS7, human S7; kb, kilobase(s) or 1000 bp; *M.*, *Manduca*; nt, nucleotide(s); ORF, open reading frame; PCR, polymerase chain reaction; r-, ribosomal; rS7, rat S7; *rp*, gene encoding r-protein; *S7*, gene (DNA) encoding S7; S7, human r-protein S7; snoRNA, small nucleolar RNA; SSC, 0.15 M NaCl/0.015 M Na_3-citrate pH 7.6; *tsp*, transcription start point(s); *X.*, *Xenopus*.

SSDI 0378-1119(95)00405-X

r-proteins with the genetic diseases. For example, haploinsufficiency of human *S4*, which is unique among the human *rp* genes by means of its two distinct isoforms (*S4X* and *S4Y*, locating in chromosomes X and Y, respectively), is considered to play a critical role in Turner syndrome (Zinn et al., 1994; Fisher et al., 1990). It has been demonstrated in the case of *Drosophila Minute* mutants that the haploinsufficiency of *rp* genes may have dramatic results. *Minutes* are a class of over 50 genes, which produce in hemizygous condition a similar phenotype and are recessive lethal (Kay and Jacobs-Lorena, 1987). It is proposed, and in two cases proved, that *Minutes* encode the r-proteins (Andersson et al., 1994). The mutations affecting the *rp* genes may reduce the overall rate of protein synthesis, lengthen the cell cycle and result in severe developmental defects (Kay and Jacobs-Lorena, 1987). Another interesting r-protein gene is *S6*. Mutations in the *Drosophila melanogaster S6* result in tissue overgrowth, suggesting that *S6* can act as a tumor suppressor gene (Watson et al., 1992).

S7 is one of the proteins (besides S3 and/or S3a, S6 and S14 and/or S15) that has been reported to play an important role in the template-40S subunit interaction (Mundus et al., 1993). In this study we present the isolation and characterization of the functional *S7* gene and its mapping to 2p25.

EXPERIMENTAL AND DISCUSSION

(a) Isolation and characterization of cDNA and genomic clones

After screening the human cDNA expression library with the rS7-specific antisera, six positive clones were purified. The clone with the largest insert was subcloned and sequenced. A single 577-nt ORF was detected with

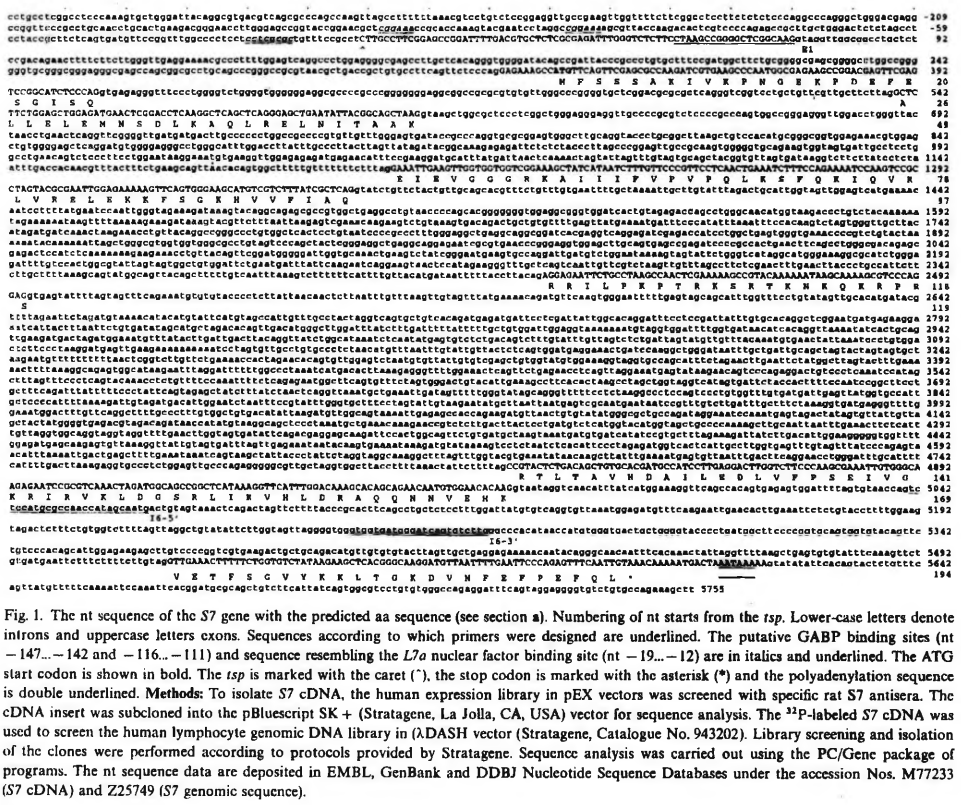
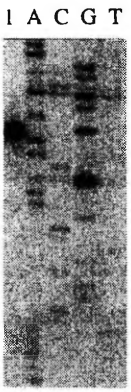


Fig. 1. The nt sequence of the *S7* gene with the predicted aa sequence (see section **a**). Numbering of nt starts from the *tsp*. Lower-case letters denote introns and uppercase letters exons. Sequences according to which primers were designed are underlined. The putative GABP binding sites (nt −147...−142 and −116...−111) and sequence resembling the *L7a* nuclear factor binding site (nt −19...−12) are in italics and underlined. The ATG start codon is shown in bold. The *tsp* is marked with the caret (^), the stop codon is marked with the asterisk (*) and the polyadenylation sequence is double underlined. **Methods:** To isolate *S7* cDNA, the human expression library in pEX vectors was screened with specific rat S7 antisera. The cDNA insert was subcloned into the pBluescript SK + (Stratagene, La Jolla, CA, USA) vector for sequence analysis. The ^{32}P-labeled *S7* cDNA was used to screen the human lymphocyte genomic DNA library in (λDASH vector (Stratagene, Catalogue No. 943202). Library screening and isolation of the clones were performed according to protocols provided by Stratagene. Sequence analysis was carried out using the PC/Gene package of programs. The nt sequence data are deposited in EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession Nos. M77233 (*S7* cDNA) and Z25749 (*S7* genomic sequence).

a poly(A)-tract preceded by a canonical polyadenylation signal (AATAAA) 13 nt downstream from the TAA stop codon. Comparison of the cDNA with database revealed homology with rat *S7* (89%), *X. laevis S8* (81%), mosquito *Anopheles gambiae S7* (66%) and tobacco hornworm *M. sexta S7* (65%) cDNAs. With the aim to isolate a genomic clone, the human lymphocyte DNA library was screened using *S7* cDNA as a hybridization probe. Five clones were analyzed at first. After the restriction mapping and the hybridization using *S7* cDNA as the radioactive probe, four of these were identified as pseudogenes, because their restriction pattern resembles that of the cDNA. One clone, which restriction with *Hin*dIII, the enzyme that does not cut the cDNA, generates multiple cDNA-hybridizing bands, was suspected to contain the intron-containing *S7* and was selected for the detailed characterization. The restriction mapping with *Eco*RI, *Bam*HI, *Pst*I, *Hin*dIII, *Kpn*I and *Xma*I localized the *S7* into three *Hin*dIII fragments. These fragments were subcloned and sequenced. Nucleotide sequence,



tttccgcctCTTGCTTCGGACGCCGGATTTT

Fig. 2. Primer extension analysis of the *tsp* (see section b). A primer extended product, shown in lane 1, corresponds to a C residue, which is indicated as nt 1 in Fig. 1. This cytidine is located within a pyrimidine tract, which, however, is interrupted by one G. The lanes ACGT contain sequencing reactions with the same primer to serve as markers. The genomic sequence and the *tsp*, indicated by an arrow head, are shown below. **Methods:** Primer extension was carried out using a 20-nt primer that is complementary to the first exon of the human *S7* gene between nt 53 and 73 (5'-CTTGCCGAGCCCCGGCTTAG). 20 ng of the primer was end-labeled as described (Sambrook et al., 1989) and hybridized to 50 µg of total RNA from HeLa cells. The annealed nucleic acids were incubated at 37°C for 2 h with 20 units of M-MuLV reverse transcriptase (NE Biolabs) in the presence of 1 mM each dNTP and 20 units of RNase inhibitor (Promega, Madison, WI, USA) in the buffer provided by Biolabs. The samples were analyzed on a 6% polyacrylamide/7 M urea gel.

```
hs MFSSSAKIVK PNGEKPDEFE SGISQALLEL -EMNSDLKAQL RELNITAAKE
rr ..........  ..........  ..........  -..........  ..........
xl ...T......  ..........  ..........  -..........  ..........
ag .VFG.KV.KA  ....-..A..  TQ.G..I...  -.......P..  .D.Y..R.R.
ms ----RPQ.L.  AG.TEA.S..  TS.....V..  =.T.........  ...Y..K...
at ...AQH..H.  EK.VELS.LD  EQVA..FFD.  .NT.QE..SE.  KD.YVNS.VQ

   IEVGGGRKAI  IIFVFVPQLK  SFQKIQVRLV  RELEKKFSGK  HVVFIAQRRI
   ..........  ..........  ..........  ..........  ..........
   ....A.....  ..........  ..........  ..........  ..........
   V.F-NNK...  ..Y....KQ.  A...V.T...  ..........  ......E...
   ..L-HNK.E.  ..Y..M.K..  A.....I...  ..........  .....CD.K.
   VDIS......  V.Y..FRLR.  A.R..HL...  ..........  D.I.V.T...

   LPKPTRKSRT  KNKQKRPRSR  TLTAVHDAIL  EDLVFPSEIV  GKRIRVKLDG
   ..........  ..........  ..........  ..........  ..........
   ..........  ..........  ..........  ....Y.....  .R........
   ....M.GR.D  P.........  P NV...Y...  ......A.V.  ..........
   ....SH.T.V  A.........  ...E.Y....  ......A...  ..........
   M-R.PK.GSA  ---VQ...N.  ...E..E.M.  .GGPGTQFAL

   SRLIKVHLDK  AQQNNVEHKV  ETFSGVYKKL  TGKDVNFEFP  EFQL
   ..........  ..........  ..........  ..........  ....
   ..........  ..........  ..........  .....V....  ....
   .Q........  N..TTI....  D..AS.....  ..R..T....  .NY.
   .Q........  N..TTI....  D..QS.....  ..RE.T....  .PY.
```

Fig. 3. Alignment of the proteins homologous to the hS7. The aa sequences were deduced from cDNAs. Dots (.) indicate identity, hyphens (-) indicate gaps that have been inserted to maximize identity. The **KRPR** nuclear localization signal of the hS7 is shown in bold and underlined. The following sequences are included in this alignment: hs-*Homo sapiens* S7, rr-*Rattus rattus* S7 (Suzuki et al., 1990), xl-*X. laevis* S8 (Mariottini et al., 1988), ag-*Anopheles gambiae* S7 (Salazar et al., 1993), ms-*M. sexta* S7 (Wang and Kanost, EMBL L20096)), at-*Arabidopsis thaliana* S7 (Philipps and Gigot, EMBL Z47607 and Z47625).

exon/intron structure and aa sequence deduced from cDNA are shown in Fig. 1. *S7* gene spans over 6 kb and contains seven exons and six introns. Exons are 65–151 nt and introns 131–2318 nt in length. All splice donor and acceptor sites have perfect consensus (GT/AG) and the sequences immediately upstream from the 3′ splice sites are pyrimidine-rich, which is common in mammalian introns (Mount, 1982). The initiation codon occurs in the context GCCAUGU which deviates a little the optimal A/GCCAUGG (Kozak, 1986). No consensus TATA or CAAT boxes were found, although a relatively A/T rich area is located between nt −50 and −36. Two binding sites for transcription factor GABP (CGGAAR) between nt −147...−142 and −116...−111 are present. Sequence CCTCGCGC between nt −19...−12 resembles the conserved binding site for nuclear factor of the *L7a* promoter (TCTCGCGA) (Colombo and Fried, 1992).

The first exon of *S7* contains 72 nt of untranslated sequence and the ATG start codon is located in the second exon. While in most human *rp* genes described so far (L7, L7a S3, S4, S6, S8, S17 and S24) the starting Met and following 1–3 aa are coded by the first exon, there are some cases in which the ATG codon residues in the second exon (S14 and S30). As the first intron of the mouse *L32* has been demonstrated to contain a transcriptional regulatory element (Chung and Perry, 1989), which, as suggested, must be located inside the approximately 200 bp promoter area, the notable shortness of

Fig. 4. Mapping of the human *S7* to a chromosome 2p25 using FISH (see section e). Based on the analysis of DAPI-banded metaphases and prometaphases, the locus for *S7* gene was assigned to 2p25, the most telomeric chromosomal band of the chromosome 2 (indicated by the arrows). **Methods:** Metaphase spreads were made from phytohemagglutinin-stimulated normal lymphocytes on a microscope slide by standard procedures. FISH was performed with biotin-11-dUTP labeled (Nick Translation Reagent Kit, BRL) *S7* genomic probe, essentially as described (Pinkel et al. 1986; Lichter et al., 1988). The hybridization mixture contained 50% formamide/10% (w/v) dextran sulfate/2 × SSC/7.5 μg biotinylated probe/0.27 mg of herring sperm DNA/0.075 mg sonicated human placenta DNA (Gibco, BRL) for competition (all per ml). For each slide, 150 ng of the labeled probe was applied. Detection of the hybridized probe was performed using avidin-FITC and amplified once with additional layers of biotinylated goat antiavidin (both 5 μg/ml, Vector Laboratories) and avidin-FITC. Nuclear DNA was counterstained with 1 μg/ml propidium iodide and stained with 5 μg/ml DAPI. Slides were mounted in antifade medium, examined with Zeiss Axiophot microscope and photographed using Ektachrome ASA 400 color film.

the first exon of the mammalian *rp* genes may be important for the transcriptional regulation.

Compared to the *X. laevis S8*, the only homologous intron-containing gene since described (Mariottini et al., 1993), only the positions but not the lengths of the introns were conserved. There is no obvious sequence similarity between the introns of human and *X. laevis* gene. It is somewhat surprising, because *U17XS8* snoRNA is coded in each of the six introns of the *X. laevis S8* (Cecconi et al., 1994), but the homologous human *U17* snoRNA is not encoded respectively by the introns of the *S7*, but by the cell cycle regulatory gene *RCC1* (Kiss and Filipowicz, 1993). The other vertebrate intron-coded snoRNAs have their genomic positions conserved (Prislei et al., 1993).

The computer analysis for the distribution of the CpG dinucleotides in the *S7* revealed that the first three closely located exons are surrounded by the CpG rich region. Intron 4 includes two *Alu*-like sequences at nt 1480 to 1600 and 1780 to 2080; there is also *Alu*-like sequence in the 5′ flanking region at nt −580 to −320 (Jurka and Milosavljevic, 1991). In respect of its gene size, *S7* is similar to *S14*. They are both almost 6 kb, while the other mammalian *rp* genes are smaller, about 3–4 kb.

The protein predicted from the cDNA contains 194 aa and (22.1 kDa). The hS7 is homologous to r-proteins isolated from other species: rS7 (Suzuki et al., 1990), *X. laevis* S8 (Mariottini et al., 1988), *Anopheles gambiae* S7 (Salazar et al., 1993), *Manduca sexta* S7 (Wang and Kanost, EMBL L20096) and *Arabidopsis thaliana* S7 (Philipps and Gigot, EMBL Z47607 and Z47625) (100%, 98%, 71%, 73% and 49% identity at an aa level, respectively; see Fig. 3). When aligned with rat, clawed frog and insect homologs, three most conserved regions of the S7

protein were detected. These include aa 78–95, 112–119, and 140–160. One of them matches the nuclear localization signal (^{115}Lys-Arg-Pro-Arg^{118}) of the protein (Annilo et al., 1994).

(b) Determination of the *tsp* of the *S7*

The 5′ end of the mRNA has been mapped by the primer extension assay using primer E1, which is complementary to the nt 53 to 73 in the first exon. A single product obtained corresponds to a C residue (Fig. 2), which is indicated as nt 1 in Fig. 1. This cytidine is located within a pyrimidine tract, which, however, is interrupted by one G. The *tsp* occurring at a C residue within a oligopyrimidine tract has been reported as a feature that is required for the transcriptional and translational control of the r-protein mRNAs (Levy et al., 1991). This polypyrimidine tract may be of variable sequence and length, but there is only one mammalian *rp* gene described whose 5′ *tsp* is not a C; namely mouse *S24*, which has a cap site at a G residue upstream from a pyrimidine tract (Xu et al., 1994). In the case of *S7*, this pyrimidine tract is 8 bp in length and is interrupted by one G. These observations suggest, that, at least in some cases, the presence of the G residue among the 5′ terminal pyrimidines enables the proper expression of the gene. Although, it has been demonstrated that a single A residue as a *tsp* preceding the general polypyrimidine tract can affect the translational regulation (Avni et al., 1994).

(c) Chromosomal localization and number of copies

We have shown earlier, using PCR with the intron-specific primers I6–5′ (forward) and I6–3′ (reverse) on the panel of the human/hamster somatic cell hybrid DNAs, that *S7* maps on the chromosome 2 (unpublished results). The chromosomal position was specified by FISH. Using fluorescence microscopy, 237 metaphase spreads hybridized with *S7* biotinylated genomic probe were examined. Hybridization signals at homologous sites on the short arm of chromosome 2 were seen in 203 (87.5%) metaphases; symmetrical doublets were present in 89 (37.5%) metaphases. No detectable signal or nonspecific binding was present in 24 spreads (14.3%). Based on the analysis of DAPI-banded metaphases and prometaphases, the locus for *S7* was assigned to 2p25, the most telomeric chromosomal band of the short arm of the chromosome 2 (Fig. 4). By now *S7* is the sole *rp* gene located on the 2 chromosome. To determine the size of the *S7* mRNA, we hybridized a radioactive cDNA probe against the poly(A)$^+$RNA from HeLa cells. A single band at 0.7 kb was present (data not shown). To detect members of the *S7* gene family, we carried out Southern blot analysis. Hybridization of the radioactive cDNA probe to *Bcl*I, *Bst*EI, *Eco*RI and *Xba*I digests of HeLa DNA suggests that there are 8–10 copies of the *S7* (data not shown).

(d) Conclusions

(*1*) The primary structure of the human *S7* gene is presented. This gene shares the features characterizing previously isolated *rp* genes – lack of consensus TATA box on its promoter area, short untranslated leader, small first exon, CpG rich 5′ region, presence of the family of the pseudogenes and location unlinked to other *rp* genes.

(*2*) The *tsp* of the gene is localized on a C residue within a pyrimidine-rich tract.

(*3*) Human *S7* maps on the short arm of the 2 chromosome (2p25).

ACKNOWLEDGEMENTS

We would like to thank Olev Kahre for synthesizing the oligodeoxynucleotide primers. The research described in this publication was made possible in part by Grant N LCV 000 from the International Science Foundation, Estonian Science Foundation and IECRF.

REFERENCES

Andersson, S., Sæbøe-Larssen, S., Lambertsson, A., Merriam, J. and Jacobs-Lorena, M.: A *Drosophila* third chromosome *Minute* locus encodes a ribosomal protein. Genetics 137 (1994) 513–520.

Annilo, T., Karis, A., Schmidt, C., Hoth, S., Lauringson, M., Pata, I., Kruppa, J. and Metspalu, A.: Characterization of the nuclear localization signals in human ribosomal proteins S3a, S6 and S7. Miami/Biotechnology Short Reports 5 (1994) 48.

Avni, D., Shama, S., Loreni, F. and Meyuhas, O.: Vertebrate mRNAs with a 5′-terminal pyrimidine tract are candidates for translational repression in quiescent cells: characterization of the translational *cis*-regulatory element. Mol. Cell. Biol. 14 (1994) 3822–3833.

Cecconi, F., Mariottini, P., Loreni, F., Pierandrei-Amaldi, P., Campioni, N. and Amaldi, F.: U17XS8, a small nucleolar RNA with a 12 nt complementarity to 18S rRNA and coded by a sequence repeated in the six introns of *Xenopus laevis* ribosomal protein S8 gene. Nucleic Acids Res. 22 (1994) 732–741.

Chan, Y-L., Suzuki, K., Olvera, J. and Wool, I.G.: Zinc finger-like motifs in rat ribosomal proteins S27 and S29. Nucleic Acids Res. 21 (1993) 649–655.

Chung, S. and Perry, R.P.: Importance of introns for expression of mouse ribosomal protein gene rpL32. Mol. Cell. Biol. 9 (1989) 2075–2082.

Colombo, P. and Fried, F.: Functional elements of the ribosomal protein L7a (rpL7a) gene promoter region and their conservation between mammals and birds. Nucleic Acids Res. 20 (1992) 3367–3373.

Feo, S., Davies, B. and Fried, M.: The mapping of seven intron-containing ribosomal protein genes shows they are unlinked in the human genome. Genomics 13 (1992) 201–207.

Fisher, E.M.C., Beer-Romeo, P., Brown, L.G., Ridley, A., McNeil, J.A., Lawrence, J.B., Willard, H.F., Bieber, F.R. and Page, D.C.: Homologous ribosomal protein genes on the human X and Y chromosomes: escape from X inactivation and possible implications for Turner syndrome. Cell 63 (1990) 1205–1218.

Grabowski, D.T., Deutsch, W.A., Derda, D. and Kelley, M.R.: *Drosophila* AP3, a presumptive DNA repair protein, is homologous to human ribosomal associated protein P0. Nucleic Acids Res. 19 (1991) 4297.

Hemmerich, P., von Mikecz, A., Neumann, F., Sözeri, O., Wolff-Vorbeck, G., Zoebelein, R. and Krawinkel, U.: Structural and functional properties of ribosomal protein L7 from humans and rodents. Nucleic Acids Res. 21 (1993) 223–231.

Jurka, J. and Milosavljevic, A.: Reconstruction and analysis of human Alu genes. J. Mol. Evol. 32 (1991) 105–121.

Kay, M.A. and Jacobs-Lorena, M.: Developmental genetics of ribosome synthesis in *Drosophila*. TIG 3 (1987) 347–351.

Kiss, T. and Filipowicz, W.: Small nucleolar RNAs encoded by introns of the human cell cycle regulatory gene *RCC1*. EMBO J. 12 (1993) 2913–2920.

Kozak, M.: Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eucaryotic ribosomes. Cell 44 (1986) 283–292.

Levy, S., Avni, D., Hariharan, N., Perry, R. and Meyuhas, O.: Oligopyrimidine tract at the 5′ end of mammalian ribosomal protein mRNAs is required for their translational control. Proc. Natl. Acad. Sci. USA 88 (1991) 3319–3323.

Lichter, P., Cremer, T., Borden, J., Manuelidis, L. and Ward, D.C.: Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. Hum. Genet. 80 (1988) 224–234.

Mariottini, P., Bagni, C., Annesi, F. and Amaldi, F.: Isolation and nucleotide sequences of cDNAs for *Xenopus laevis* ribosomal protein S8: similarities in the 5′ and 3′ untranslated regions of mRNAs for various r-proteins. Gene 67 (1988) 69–74.

Mariottini, P., Bagni, C., Francesconi, A., Cecconi, F., Serra, M.J., Chen, Q.-M., Loreni, F., Annesi, F. and Amaldi, F.: Sequence of the gene coding for ribosomal protein S8 of *Xenopus laevis*. Gene 132 (1993) 255–260.

Mount, S.M.: A catalogue of splice junction sequences. Nucleic Acids Res. 10 (1982) 459–472.

Mundus, D.A., Bulygin, K.N., Ven'yaminova, A.G., Vladimirov, S.N., Vratskikh, L.V., Repkova, M.N., Yamkovoi, V.S. and Karpova, G.G.: Human placenta 40S ribosomal subunit affinity modification by oligoribonucleotide derivatives containing the AUG codon. Ribosomal proteins composing the codon-anticodon interaction site. Mol. Biol. 27 (1993) 91–95.

Pinkel, D., Straume, T. and Gray, J.W.: Cytogenetic analyses using quantitive, high sensitive fluorescence hybridization. Proc. Natl. Acad. Sci. USA 83 (1986) 2934–2938.

Prislei, S., Michienzi, A., Presutti, C., Fragapane, P. and Bozzoni, I.: Two different snoRNAs are encoded in introns of amphibian and human L1 ribosomal protein genes. Nucleic Acids Res. 21 (1993) 5824–830.

Salazar, C.E., Mills-Hamm, D., Kumar, V. and Collins, F.H.: Sequence of a cDNA from the mosquito *Anopheles gambiae* encoding a homologue of human ribosomal protein S7. Nucleic Acids Res. 21 (1993) 4147.

Suzuki, K., Olvera, J. and Wool, I.G.: The primary structure of rat ribosomal protein S7. FEBS Lett. 271 (1990) 51–53.

Watson, K.L., Konrad, K.D., Woods, D.F. and Bryant, P.J.: *Drosophila* homolog of the human S6 ribosomal protein is required for tumor suppression in the hematopoietic system. Proc. Natl. Acad. Sci. USA 89 (1992) 11302–11306.

Wilson III, D.M., Deutsch, W.A. and Kelley, M.R.: Cloning of the *Drosophila* ribosomal protein S3: another multifunctional ribosomal protein with AP endonuclease DNA repair activity. Nucleic Acids Res. 21 (1993) 2516.

Wool, I.G.: The bifunctional nature of ribosomal proteins and speculations on their origins. In: Nierhaus, K.H. (Ed.), The Translational Apparatus. Plenum Press, New York, 1993, pp. 727–737.

Xu, L., He, G.-P., Li, A. and Ro, H.-S.: Molecular characterization of the mouse ribosomal protein S24 multigene family: a uniquely expressed intron-containing gene with cell-specific expression of three alternatively spliced mRNAs. Nucleic Acids Res. 22 (1994) 646–655.

Xu, W. and Roufa, D.J.: (EMBL U12202).

Zinn, A.R., Alagappan, R.K., Brown, L.G., Wool, I. and Page, D.C.: Structure and function of ribosomal protein S4 genes on the human and mouse sex chromosomes. Mol. Cell. Biol. 14 (1994) 2485–2492.

II

Annilo, T., Karis, A., Hoth, S., Rikk, T., Kruppa, J. and Metspalu, A. (1998)
Nuclear import and nucleolar accumulation of the human ribosomal protein S7
depends on both a minimal nuclear localization sequence and
an adjacent basic region.
Biochem. Biophys. Res. Commun. **249**, 759–766.

BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS 249, 759–766 (1998)
ARTICLE NO. RC989187

Nuclear Import and Nucleolar Accumulation of the Human Ribosomal Protein S7 Depends on both a Minimal Nuclear Localization Sequence and an Adjacent Basic Region

Tarmo Annilo, Alar Karis, Stefan Hoth,*† Tiina Rikk, Joachim Kruppa,* and Andres Metspalu[1]
Institute of Molecular and Cell Biology, Tartu University, Estonian Biocentre, Riia 23, Tartu, EE2400, Estonia;
**Department of Molecular Biology, University of Hamburg, 20146 Hamburg, Germany; and*
†*Gottfried Friedrich KG, 22765 Hamburg, Germany*

Received July 10, 1998

In the course of the eukaryotic ribosomal biogenesis, both the nuclear import and export are involved. We have studied the nuclear and nucleolar localization of the human ribosomal protein S7. We examined the subcellular distribution of the S7:β-galactosidase fusion proteins in SAOS-2 cells. We have identified two evolutionarily conserved domains, both of which are necessary for S7 nuclear and nucleolar targeting: amino acids 98 to 109 and 115 to 118. Out of the S7 protein context, a fragment 98...118, containing these domains, is sufficient for nuclear transport and nucleolar accumulation. Interestingly, a tetrapeptide ^{115}KRPR^{118}, which can act as an independent nuclear localization signal (NLS), is not sufficient for exclusively nuclear accumulation of the S7 protein if the adjacent region 98...109 is deleted. In addition, site-directed mutagenesis revealed that critical residues for nuclear targeting in this tetrapeptide and in the full-length S7 protein are different. While mutation of a Pro^{117} significantly impaired nuclear import of S7, similar substitution within the tetrapeptide-NLS had no effect on nuclear targeting. This suggests that to function perfectly, proper secondary structure of the S7 nuclear targeting domain is required. © 1998 Academic Press

Assembly of the eukaryotic ribosomal subunits takes place in the nucleolus where ribosomal proteins associate with four rRNA species (1). Ribosome biogenesis requires that approximately 80 ribosomal proteins, which are synthesized in the cytoplasm, have to be imported into the nucleus. After a number of maturation steps, preribosomal subunits exit the nucleus. Nucleo-cytoplasmic transport occurs through the nuclear pore complexes (NPC); nuclear targeting of proteins is specified by nuclear localization sequences (NLSs) within the imported proteins. NLSs are predominantly short basic amino acid sequences with only a limited homology in either a continuous or bipartite motif (2). NLSs are not proteolytically removed during transport and this allows nucleo-cytoplasmic shuttling of many nuclear proteins, like *Xenopus* heat shock related proteins (3) and nucleolar proteins nucleolin, B23 (4) and Nopp140 (5). The mechanism of import and export of macromolecules has been intensely studied in recent years and a number of soluble factors and NPC proteins involved in this process have been identified (6). Briefly, a complex of karyopherin α and karyopherin β (also termed importins) binds the NLS-bearing import substrate in the cytosol. The complex docks to the nuclear pore and then by the energy-dependent, Ran-mediated translocation step, the substrate and karyopherin α are translocated into the nucleus (7,8). Specific β-karyopherins for ribosomal proteins in yeast have been characterized (9).

While the primary structure of most of the eukaryotic ribosomal proteins is known (10), knowledge about their role in ribosome assembly and translational process is still largely missing. Nuclear transport of some yeast and mammalian ribosomal proteins has been studied before. In yeast L3, a single NLS resides within the first 21 amino acids (11); in L25, two regions responsible for nuclear transport were identified (12). Human L7a has three nuclear targeting domains (13). NLSs of yeast ribosomal protein L29, rat L31 and hu-

[1] Corresponding author: Dr. A. Metspalu, Institute of Molecular and Cell Biology, Tartu University, Estonian Biocentre, Riia 23, Tartu, EE2400, Estonia. E-mail: andres@ebc.ee.

Abbreviations: NLS, nuclear localization signal; NPC, nuclear pore complex.

0006-291X/98 $25.00

man S6 have been examined in detail. Yeast L29 has two NLSs which contain mainly basic residues and are identical in five of the seven amino acids (14). In ribosomal protein S6, three short NLSs and one region which is essential but not sufficient for nucleolar localization have been described (15). S3a contains a sequence that fits the consensus NLS proposed by Chelsky et al. (16), but experimental data show that S3a:β-galactosidase fusions do not have nuclear targeting ability (17). Therefore, nuclear targeting domains among ribosomal proteins are quite different and to understand ribosome biogenesis mechanism, every protein should be studied individually. For some NLSs, critical residues for their function have been identified. For example, mutation of any of the three arginines of the rat L31 NLS RLSRKR abolished nucleolar localization (18). NLS of SV40 large T antigen (19) and yeast histone H2B (20) have a unique lysine residue, mutation of which inhibits nuclear accumulation.

We have previously described the cDNA and gene encoding for human S7 (21). The S7 protein contains 194 amino acids and has been reported to play an important role in the template-40S subunit interaction (22). Here we describe a mutational analysis of S7 and demonstrate that although S7 contains a tetrapeptide with a nuclear targeting ability, for nucleolar accumulation an adjacent basic domain is also required.

MATERIALS AND METHODS

Plasmid constructs. The plasmid pKHlacZ was specifically designed for the nucleo-cytoplasmic transport studies (Fig. 1). Fusion proteins of the wild-type or mutated S7 sequences with *E. coli* β-galactosidase were constructed by cloning different cDNA fragments or oligonucleotides into the shuttle vector pKHlacZ at the NotI and HindIII sites. A schematic drawing of the fusion proteins is shown in Figure 2A. The constructs containing full-length S7 and deletion mutants lacking amino- or carboxy-terminal sequences were generated under standard conditions by PCR amplification of the corresponding cDNA fragment using Goldstar DNA Polymerase (EUROGENTEC, rue Bois Saint Jean, 14, 4102 Seraing, Belgium). For the constructs ΔN80, ΔN120 and Δ114C, sequences encoding amino acids 81...194, 121...194 and 3...113, respectively, were amplified. All PCR-derived constructs were generated using oligonucleotides with NotI(5') and HindIII(3') sites for insertion of the fragment.

For the peptide-NLS constructs, synthetic oligonucleotides encoding both strands of S7 wild-type (amino acids 115...118) or mutated NLS with a NotI site at the 5' end and a HindIII site at the 3' end were annealed and inserted into pKHlacZ. Site-directed mutagenesis of the S7 protein to alter the codons for Lys^{115}, Arg^{116}, Pro^{117} and Arg^{118} and internal deletions were carried out using Muta-Gene® M13 *in vitro* Mutagenesis Kit (BIO-RAD, Auhofstraße 78 D A-1130 Vienna Austria). As the plasmid pKHlacZ contains f1 phage replication origin, a single-stranded form of a construct containing full-length S7 cDNA (wtS7) was isolated and used as a template. All constructs were verified by dideoxy sequencing.

Cell culture and transfection. SAOS-2 cells were cultured in DMEM supplemented with 10% fetal calf serum at 37° C in 5% CO_2. Subconfluent cells were electroporated with 2-4 μg of plasmid DNA. After 24 to 48 h post-transfection, cells were fixed with 0,05% (v/v) glutaraldehyde in PBS, and stained with X-gal for β-galactosidase activity. Photomicrographs at 300× magnification were obtained with a Nikon Diaphot microscope.

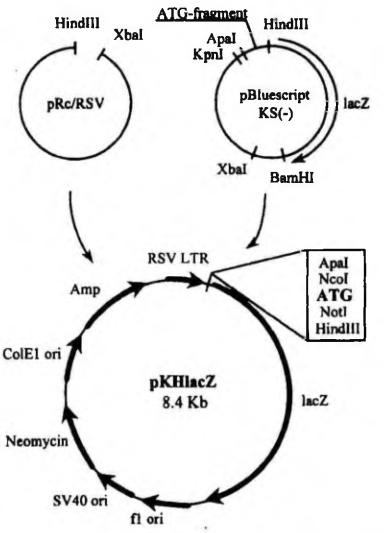


FIG. 1. Construction of the vector pKHlacZ used for testing the subcellular localization of the S7:β-galactosidase fusion proteins. Plasmid pPbxnlslacZ was used as a source for the *E. coli* lacZ gene. At first, the HindIII–BamHI fragment of pPbxnlslacZ, which contains the majority of the lacZ gene, was cloned into the same sites of pBluescript KS (−). In order to introduce a translation start signal, the pBluescript KS (−)-lacZ vector was digested with ApaI and HindIII and complementary synthetic oligonucleotides (5'CACCATGGCTAGCGGCCGCATA3' and 3'CCGGGTGGTACCGATCGCCGGCGTATTCGA5') providing the ATG start codon (underlined) and NcoI and NotI sites were annealed and inserted at the 5' end of the lacZ gene. The resulting construct was cut with KpnI and the recessed termini were filled using Klenow polymerase followed by cut with XbaI. ATG-lacZ fragment obtained was inserted into plasmid pRc/RSV (Invitrogen) which was digested with HindIII and after filling of recessed termini was cleaved with XbaI.

RESULTS

S7 Accumulates the Reporter Protein within the Nucleoli

In order to test whether S7 contains a targeting sequence for the nuclear and nucleolar import, we fused the almost full-length S7 sequence, lacking only the first two amino acids (starting Met and Phe), to the N-terminus of β-galactosidase (referred to as wtS7, see Fig. 2A). Transient expression of this fusion in SAOS-2 cells showed that S7 was able to transport the reporter protein to the nucleus, where it accumulates into the

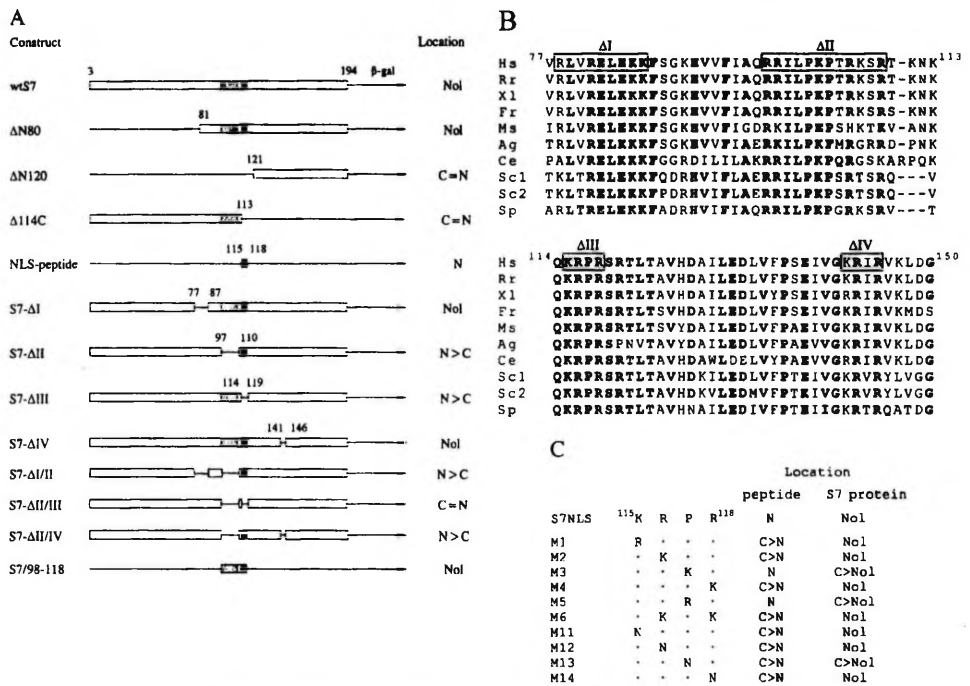


FIG. 2. Structure and subcellular distribution of the S7:β-galactosidase fusion proteins. (A) Blank boxes, S7 protein sequences; thin lines, deleted regions; black box, NLS; light stippling, basic region (amino acids 98 to 109) preceding the NLS; numbers indicate amino acids in the S7 sequence. (B) Partial alignment of S7 homologues. Sequences included in the alignment were: Hs, *Homo sapiens* (21); Rr, *Rattus rattus* (32); Xl, *Xenopus laevis* (33); Fr, *Fugu rubripes* (34); Ms, *Manduca sexta* (35); Ag, *Anopheles gambiae* (36); Ce, *Caenorhabditis elegans* (EMBL Z75714); Sc1 and Sc2, *Saccharomyces cerevisiae* (EMBL Z75004 and Z71372, respectively); and Sp, *Schizosaccharomyces pombe* (EMBL Z68198). Sequences in human S7 that were deleted in constructs ΔI–IV are denoted with boxes. Bold letters indicate the amino acids that are identical in nine out of ten sequences. (C) Effect of the point mutations on subcellular localization of the peptide-NLS- and full-length S7:β-galactosidase fusion proteins. The minimal nuclear targeting sequence is given in a single-letter code and amino acid positions are according to S7 sequence. N, nuclear; Nol, nuclear localization with nucleolar accumulation; C>N, C=N, and N>C indicate the predominant staining when the location was not totally cytoplasmic or nuclear.

nucleoli (Fig. 3A). In a control experiment we observed that β-galactosidase expressed from pKHlacZ was not exclusively cytoplasmic, but the staining was distributed between the nucleus and the cytoplasm (Fig. 3J). As shown recently, nuclear staining of β-galactosidase could be a technical artifact due to cytoplasmic staining above the nucleus (13). However, since the S7:β-galactosidase fusion was completely nuclear, this phenomenon does not interfere with the identification of nuclear targeting sequences.

Amino acid alignment of S7 homologues revealed the extremely conserved region between amino acids 114...123 (Fig. 2B). Sequence ^{115}Lys-Arg-Pro-Arg^{118} within this region is very similar to the nuclear import sequences in ribosomal protein S6. At first, deletions from the amino- or the carboxy-terminus of S7 were generated to explore the role of this sequence on transport. Deletion of the first 80 amino acids (ΔN80) resulted in the similar subcellular localization as the full-length S7 (Fig. 3B). The construct lacking the amino-terminal 120 residues (ΔN120) shows staining in both the cytoplasm and nucleus (Fig. 3C). When the N-terminal 113 residues were fused to the β-galactosidase (Δ114C), nuclear transport was also reduced, and staining was equally distributed in the nucleus and the cytoplasm (Fig. 3D). In addition, constructs ΔN120 and

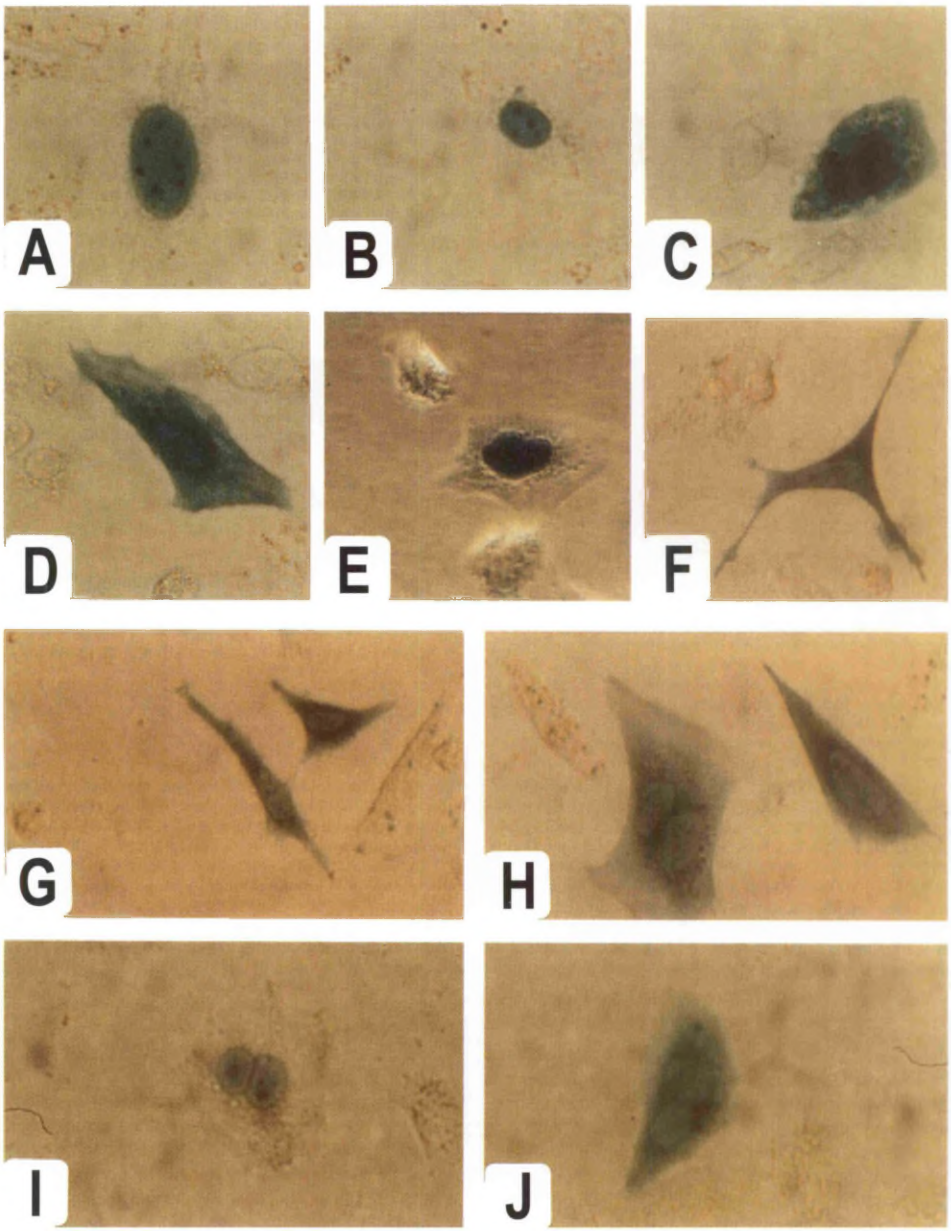


FIG. 3. Subcellular localization of the β-galactosidase fusion proteins. The plasmid DNAs encoding fusion proteins described in Fig. 2 were transfected into SAOS-2 cells. The location of fusions was visualised by X-gal staining 24 to 48 hours later. (A) wtS7, (B) ΔN80, (C) ΔN120, (D) Δ114C, (E) NLS-peptide, (F) NLS-M1, (G) NLS-M2, (H) NLS-M4, (I) S7/98-118, (J) pKHlacZ, expressing β-galactosidase alone. Original magnification 300×.

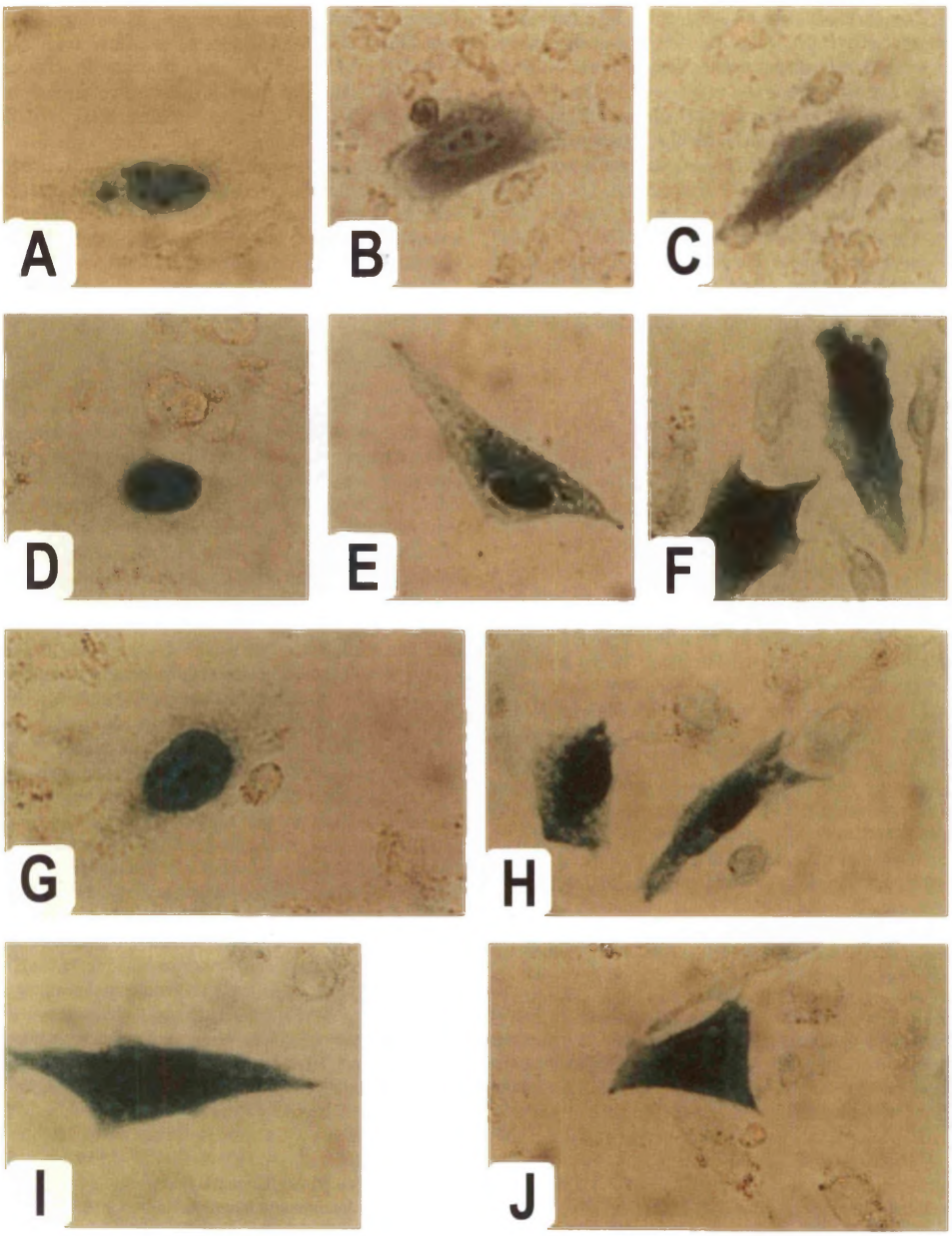


FIG. 4. Effect of point mutations and internal deletions on the subcellular localization of the S7:β-galactosidase fusion proteins expressed in SAOS-2 cells. Constructs were as described in Fig. 2. (A) S7-M1, (B) S7-M3, (C) S7-M5, (D) S7-ΔI, (E) S7-ΔII, (F) S7-ΔIII, (G) S7-ΔIV, (H) S7-ΔI/II, (I) S7-ΔII/III, (J) S7-ΔII/IV. Original magnification 300×.

Δ114C were not capable in localizing to the nucleoli. To verify whether tetrapeptide ^{115}Lys-Arg-Pro-Arg^{118} alone is sufficient to direct nuclear import, oligonucleotides encoding this sequence were synthesized and inserted into the 5' end of lacZ gene. Indeed, this peptide was able to transport β-galactosidase very efficiently into the nucleus, but no nucleolar accumulation could be detected (Fig. 3E).

Point Mutagenesis of the S7 NLS Revealed Differences in Nuclear Targeting Sequence Requirements between Peptide-NLS and S7 Protein

Having established that the tetrapeptide KRPR was sufficient to promote complete nuclear targeting when fused to the β-galactosidase, we next studied the impact of specific amino acids on nuclear transport efficiency. Every single amino acid was substituted within the NLS-tetrapeptide and within the S7 protein at a conservative or nonconservative manner (Fig 2C). As a conservative mutation we considered another basic amino acid, as a nonconservative one, asparagine. In a tetrapeptide, all mutations except for those which substituted proline reduced nuclear transport, as can be seen by the predominantly cytoplasmic staining of the cells (Figs. 3F-3H).

Impact of the mutations in the full-length S7 context was different. The Pro^{117}→Lys or Pro^{117}→Arg substitutions, which had no effect in the tetrapeptide, reduced considerably the nuclear targeting of the full-length protein (S7-M3 and S7-M5; Figs. 4B, 4C). In contrast, substitution of Lys^{115}, Arg^{116} or Arg^{118} by basic or neutral amino acid only slightly affected nuclear targeting (S7-M1, Fig. 4A). Furthermore, none of these mutations abolished nucleolar localization, suggesting that residues within the NLS that are essential for effective nuclear import are not critical for nucleolar accumulation.

Analysis of the Sequences Required for Nucleolar Accumulation

Since the point mutations affecting nuclear targeting did not prevent nucleolar accumulation of the imported protein, the question arose, which region besides the NLS controls the nucleolar targeting of S7. To answer this question, we deleted either the highly conserved region 78...86, which is rich in charged residues (S7-ΔI, see Fig. 2B), the arginine-lysine rich region 98...109 (S7-ΔII), the minimal NLS (S7-ΔIII), or the sequence similar to NLS ^{142}KRIR^{145} (S7-ΔIV). The subcellular location of S7-ΔI and S7-ΔIV was nuclear and nucleolar, similar to the wtS7 (Figs. 4D, 4G). In the deletions S7-ΔII and S7-ΔIII efficiency of nuclear transport was reduced and no nucleolar accumulation was detected (Figs. 4E, 4F). Since none of these deletions resulted in completely or preferentially cytoplasmic location, combinations of deletions I and II, II and III, and II and IV were constructed and expressed. S7-ΔI/II and S7-ΔII/IV had cytoplasmic staining similar to S7-ΔII (Figs. 4H, 4J). Simultaneous deletion of amino acids 98...109 and 115...118 (S7-ΔII/III, containing 9 basic residues of 16) affected nuclear import most strongly, resulting in preferentially cytoplasmic location (Fig. 4I). To test whether the domains necessary for nucleolar targeting have functional significance out of the S7 protein context, location of the chimeric protein S7/98-118 was studied (Fig. 3I). Indeed, these 21 amino acids were sufficient for nuclear transport and nucleolar targeting of the reporter protein.

DISCUSSION

In order to analyze the subcellular location and determine the sequence responsible for nuclear and nucleolar targeting of ribosomal protein S7, we have constructed a series of S7 fusion proteins with β-galactosidase. S7, like all ribosomal proteins, is a small basic protein that could theoretically enter the nucleus by diffusion, but it is reasonable to assume that active transport is exploited to ensure sufficient ribosome production in response to physiological changes.

In this study we demonstrate that despite the fact that S7 contains a tetrapeptide sequence ^{115}KRPR^{118}, which, when fused to the β-galactosidase, can function as an independent NLS, effective nuclear and nucleolar transport of the S7 protein requires the presence of another basic domain ^{98}RRILPKPTRKSR^{109}. In addition, a 21-amino acid S7 fragment containing these domains is sufficient for nuclear and nucleolar accumulation of the β-galactosidase fusion.

The minimal S7 NLS ^{115}KRPR^{118} is very similar to the NLSs of ribosomal protein S6 (KRRR and KKPR), but differs from signals of yeast L29 and L3 and rat L31, which are longer and contain amino acids other than lysine, arginine and proline. However, presence of a nuclear targeting consensus sequence within the protein not always reflects its nuclear targeting ability. For example, ribosomal protein S3a has a motif ^{218}KKPK^{221}, but it has been demonstrated that in β-galactosidase fusions neither the full-length S3a nor the N-terminus, containing this sequence, was able to promote nuclear import (17).

Mutation analysis revealed that substitution of any of the basic residues within the S7-derived NLS tetrapeptide reduced significantly nuclear localization. Surprisingly, the mutation results with the peptide did not apply to the full-length S7, where only proline^{117} appeared to be critical, while substitution of any of the basic amino acids did not impair transport. This suggests that: 1) presence of a turn in a polypeptide chain due to proline is essential for nuclear targeting signal function within a mature S7, but not in the tetra-

peptide; and 2) that the context of the full-length protein can support and strengthen the partially defective NLS.

Since mutations within the full-length S7 NLS did not affect nucleolar accumulation, we used deletion mutagenesis to identify the sequence responsible for nucleolar localization. As the possible domains we tested a cluster of conserved charged amino acids ^{78}RLVRELEKK^{86} (S7-ΔI); a sequence ^{98}RRILPKPTRKSR^{109} (S7-ΔII), which is also remarkably conserved and rich in basic residues; and a motif resembling the minimal NLS ^{142}KRIR^{145} (S7-ΔIV, Fig. 2B). We found that regions 78...86 and 142...145 can be deleted without any impact upon nuclear or nucleolar location. Deletion of amino acids 98...109 (S7-ΔII) and 115...118 (minimal NLS, S7-ΔIII) prevented nucleolar accumulation and resulted in a distribution of the fusion protein between the nuclear and cytoplasmic compartment (Figs. 4E, 4F).

Constructs which contained the minimal NLS but not the sequence ^{98}RRILPKPTRKSR^{109} (S7-ΔII, S7-ΔI/II and S7-ΔII/IV) showed in addition to nuclear also cytoplasmic staining. Therefore, in the context of the full-length S7, the tetrapeptide ^{115}KRPR^{118} alone is not sufficient for the nuclear accumulation of a β-galactosidase. In addition, fusions Δ114C and S7-ΔIII, in which the minimal NLS was deleted, were not preferentially cytoplasmic, but showed also considerable nuclear accumulation. Only when the minimal NLS and region 98-109 were simultaneously deleted (S7-ΔII/III), localization of the resulting fusion protein was similar to the β-galactosidase staining of the control experiment. Therefore, the results indicate that the region 98...109 is also capable to direct nuclear import to some extent.

Thus, it appears most likely that the tetrapeptide ^{115}KRPR^{118}, which alone can function as a nuclear targeting signal, may be supported by adjacent region (starting from amino acid 98) in the mature S7 protein, forming a bipartite NLS. Both these domains are highly conserved between yeast, insects and vertebrates. These sequence elements are also required for nucleolar accumulation, which most likely takes place via binding to other nucleolar proteins or to nucleic acids. It has been proposed that these interactions occur through the functional domains, not due to the specific nucleolar localization signal (23).

Interestingly, we found that point mutations within the S7 protein S7-M3 and S7-M5 (Fig. 4B, 4C) had stronger effect than deletion of the whole minimal signal S7-ΔIII. One possible explanation is that mutant NLS can bind to its receptor but cannot promote nuclear targeting, thereby reducing the free cytoplasmic pool of receptors. In the absence of NLS, another sequence with lower affinity for the receptor can direct nuclear import. Deletion of the NLS may also change the conformation of the protein and thereby uncover otherwise masked sequence with nuclear targeting ability.

The observation that most of the deletions and point mutations of S7 resulted only in reduced transport but not in complete block of import can be explained by the fact that nucleocytoplasmic transport is not just an "all-or-nothing" mechanism but is widely and precisely regulated (24). A number of studies have been described mechanisms controlling nuclear import by modulating NLS activity. The activity of NLS can be regulated both in a tissue-specific manner or during development (25). Nuclear import can be precisely regulated by such mechanisms as masking/unmasking of NLS (26), phosphorylation (24) and cytoplasmic retention domains (27). The efficiency of the minimal NLS can be strongly affected by the protein context (28) or even by distant regions (29). Multiple signals can cooperate to increase transport efficiency (28). Even the well-studied SV-40 T-antigen NLS requires domain N-terminal to the NLS for full activity (30). NLSs do not fit a tight consensus and as shown recently, neutral and even acidic amino acids can be important for nuclear transport (31).

The conservation of the NLS and a domain N-terminal to the NLS in S7 indicates that this region may control nuclear and nucleolar accumulation of all S7 homologues and that the mechanism leading to the nucleolar accumulation may be also conserved.

ACKNOWLEDGMENTS

We thank Marika Lauringson for her technical assistance at the beginning of the project and Olev Kahre for synthesizing the oligodeoxynucleotide primers. We are grateful to Carola Schmidt for the gift of the vector pPbxnlslacZ. This work was supported by Estonian Science Foundation Grant 2491.

REFERENCES

1. Warner, J. R. (1989) *Microbiol. Rev.* **53**, 256–271.
2. Hicks, G. R., and Raikhel, N. V. (1995) *Annu. Rev. Cell Dev. Biol.* **11**, 155–188.
3. Mandell, R. B., and Feldherr, C. M. (1990) *J. Cell Biol.* **111**, 1775–1783.
4. Borer, R. A., Lehner, C. F., Eppenberger, H. M., and Nigg, E. A. (1989) *Cell* **56**, 379–390.
5. Meier, U. T., and Blobel, G. (1992) *Cell* **70**, 127–138.
6. Melchior, F., and Gerace, L. (1995) *Curr. Opin. Cell Biol.* **7**, 319–324.
7. Izaurralde, E., and Adam, S. (1998) *RNA* **4**, 351–364.
8. Görlich, D. (1998) *EMBO J.* **17**, 2721–2727.
9. Rout, M. P., Blobel, G., and Aitchison, J. D. (1997) *Cell* **89**, 715–725.
10. Wool, I. G., Chan, Y.-L., and Glück, A. (1996) *in* Translational Control (Hershey, J. W. B., Mathews, M. B., and Sonenberg, N., Eds.), pp. 685–732, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
11. Moreland, R. B., Nam, H. G., Hereford, L. M., and Fried, H. M. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 6561–6565.

12. Schaap, P. J., van't Riet, J., Woldringh, C. L., and Raue, A. H. (1991) *J. Mol. Biol.* **221,** 225–237.
13. Russo, G., Ricciardelli, G., and Pietropaulo, C. (1997) *J. Biol. Chem.* **272,** 5229–5235.
14. Underwood, M. R., and Fried, H. M. (1990) *EMBO J.* **9,** 91–99.
15. Schmidt, C., Lipsius, E., and Kruppa, J. (1995) *Mol. Biol. Cell* **5,** 1875–1885.
16. Chelsky, D., Ralph, R., and Jonak, G. (1989) *Mol. Cell. Biol.* **9,** 2487–2492.
17. Hoth, S. (1993) Zur Biogenese von Ribosomen: S6-Genstruktur, S3a-Primärstruktur und -Kernimport. Dissertation, Universität Hamburg.
18. Quaye, I. K., Toku, S., and Tanaka, T. (1995) *Eur. J. Cell Biol.* **69,** 151–155.
19. Kalderon, D., Richardson, W. D., Markham, A. F., and Smith, A. E. (1984) *Nature* **311,** 33–38.
20. Moreland, R. B., Langevin, G. L., Singer, R. H., Garcea, R. L., and Hereford, L. M. (1987) *Mol. Cell. Biol.* **7,** 4048–4057.
21. Annilo, T., Laan, M., Stahl, J., and Metspalu, A. (1995) *Gene* **165,** 297–302.
22. Mundus, D. A., Bulygin, K. N., Ven'yaminova, A. G., Vladimirov, S. N., Vratskikh, L. V., Repkova, M. N., Yamkovoi, V. S., and Karpova, G. G. (1993) *Mol. Biol.* **27,** 91–95.
23. Mélese, T., and Xue, Z. (1995) *Curr. Op. Cell Biol.* **7,** 319–324.
24. Jans, D. A., and Hübner, S. (1996) *Physiol. Rev.* **76,** 651–685.
25. Standiford, D. M., and Richter, J. D. (1992) *J. Cell Biol.* **118,** 991–1002.
26. Zabel, U., Henkel, T., Silva, M. S., and Baeuerle, P. A. (1993) *EMBO J.* **12,** 201–211.
27. Sittler, A., Devys, D., Weber, C., and Mandel, J. (1996) *Hum. Mol. Gen.* **5,** 95–102.
28. Roberts, B. L., Richardson, W. D., and Smith, A. E. (1987) *Cell* **50,** 465–475.
29. Gao, M., and Knipe, D. M. (1992) *Mol. Cell. Biol.* **12,** 1330–1339.
30. Riha, H. P., and Peters, R. (1989) *EMBO J.* **8,** 1479–1484.
31. Makkerh, J. P. S., Dingwall, C., and Laskey, R. A. (1996) *Curr. Biol.* **6,** 1025–1027.
32. Suzuki, K., Olvera, J., and Wool, I. G. (1990) *FEBS Lett.* **271,** 51–53.
33. Mariottini, P., Bagni, C., Annesi, F., and Amaldi, F. (1988) *Gene* **67,** 69–74.
34. Cecconi, F., Crosio, C., Mariottini, P., Cesareni, G., Giorgi, M., Brenner, S., and Amaldi, F. (1996) *Nucleic Acids Res.* **24,** 3167–3172.
35. Jiang, H., Wang, Y., and Kanost, M. R. (1996) *Insect Mol. Biol.* **5,** 31–38.
36. Salazar, C. E., Mills-Hamm, D., Kumar, V., and Collins, F. H. (1993) *Nucleic Acids Res.* **21,** 4147.

III

Annilo, T., Jelina, J., Pata, I. and Metspalu, A. (1998)
Isolation and characterization of the mouse ribosomal protein S7 gene.
Biochem. Mol. Biol. Int. **46**, 287–295.

Vol. 46, No. 2, October 1998 BIOCHEMISTRY *and* MOLECULAR BIOLOGY INTERNATIONAL
Pages 287–295

ISOLATION AND CHARACTERIZATION OF THE MOUSE RIBOSOMAL PROTEIN S7 GENE

Tarmo Annilo, Jelena Jelina, Illar Pata, and Andres Metspalu*
Tartu University, Institute of Molecular and Cell Biology, Estonian Biocentre, 23 Riia St., EE2400 Tartu, Estonia

Received May 19, 1998

SUMMARY: We have isolated and characterized the single intron-containing gene encoding the mouse ribosomal protein S7. The mouse 129SV *S7* gene has seven exons and six introns spanning about 5000 nucleotides. The exon-intron structure of the gene is similar to other vertebrate homologues. Southern blot analysis showed that in addition to the isolated single-copy intron-containing gene, there are 10-12 members in the mouse S7 gene family, which are all most probably processed pseudogenes. The promoter region of *rpS7* contains several evolutionarily conserved putative regulatory elements. The main transcription start site was mapped to a T residue within a polypyrimidine tract, 79 nucleotides upstream from AUG codon.

Key words: ribosomal protein gene, S7

INTRODUCTION

The mammalian ribosome, consisting of four different RNAs and 80 proteins, is a highly complex organelle (1). Genes encoding individual ribosomal proteins are members of multigene families, which may contain more than 20 members (2). However, a single expressed intron-containing gene has been described for each of the mammalian ribosomal proteins analysed thus far (except for the human *S4*, for which there seems to be two active genes (3)). The other copies are believed to be processed pseudogenes. The presence of pseudogenes has hampered the isolation of functional genes and although the primary structure of most of the ribosomal proteins is known, only a small number of intron-containing genes have been described to date.

* Author for correspondence. Fax: 372-7-420-286; e-mail: andres@ebc.ee

1039–9712/98/140287–09$05.00/0

Using PCR-based approach (4), we have cloned and sequenced the gene encoding the mouse homologue of ribosomal protein S7. Molecular characterization of the mouse *S7* gene provides information for further gene targeting experiments. The biological function of S7 in protein synthesis is largely unknown, but experimental data indicate that S7 is located at the ribosomal interface, as shown by its crosslinking to ribosomal protein L7 (5) and 28S rRNA (6); it is also involved in the eIF-3 attachment site formation (7) and in the mRNA-40S subunit interaction (8).

MATERIALS AND METHODS

Isolation and analysis of DNA. Oligonucleotides T1 AAGTCCGGCTAGTTCGTGAATTGG and N/P CTTATTTTTCGTACGGCTTT were used to amplify the fourth intron of the mouse *S7* gene. This PCR product was purified, [^{32}P]-labeled and used as a probe for mouse 129SV genomic library (Stratagene, Cat. No. 946305) screening. Positive plaques were analysed by PCR using different primers to confirm the presence of the complete intron-containing *S7* gene. Clone λmS7-11 was chosen for further analysis. After restriction mapping, different overlapping fragments from this lambda clone were subcloned into the plasmid pBluescript SK+ (Stratagene). In parallel, four *S7* cDNA clones were isolated from the T-cell mouse cDNA library (Stratagene, Cat. No. 936303) using human *S7* cDNA as a probe.
DNA sequence analysis was performed by the dideoxy chain termination method with Thermo Sequenase^{TM} (Amersham) and ^{33}P labeled dideoxyribonucleoside triphosphate terminators.
Primer extension assay. Oligonucleotide H-29, complementary to nucleotides 51-70 was end-labeled with [γ-^{32}P] ATP and hybridised to 20μg of total RNA isolated from cultured mouse C127 cells. Primer extension was performed with 20 units of Moloney murine leukemia virus reverse transcriptase (Pharmacia Biotech) and the extension products were resolved on a 6% acrylamide-urea gel.
Southern analysis. 10 μg of genomic DNA from the mouse 129SV strain was digested with restriction endonucleases, resolved by electrophoresis, and transferred to Hybond-N+ membrane (Amersham). The filter was hybridised at first with a [^{32}P]-labeled HindIII fragment from the *S7* intron 5 and then to mouse *S7* cDNA.

RESULTS AND DISCUSSION

We expected that the coding sequences of mammalian genes are conserved enough to allow amplification of mouse DNA with the oligonucleotides designed according to human DNA sequence. Primers T1 and N/P (see Fig. 1), flanking intron 4 in human *rpS7* (9), generated two bands from mouse genomic DNA. The shorter PCR product was about 120 nucleotides, which corresponds to the distance predicted from the cDNA and was therefore assumed to derive from the intronless pseudogenes. The larger product, expectedly generated from the intron-containing gene, was about 500 nucleotides. This fragment was used to isolate the mouse *S7* gene from the 129SV genomic library.

The primary structure of the mouse *S7* intron-containing gene and the predicted amino acid sequence are shown in Fig. 1. Exon sequences are identical with the cDNA sequence. Gene architecture of mouse *S7* was compared with other eukaryotic homologues. Vertebrate intron-containing *S7* genes from *Fugu rubripes* (10), *Xenopus laevis* (11), and *Homo sapiens* (9) have been described earlier. Introns in vertebrate genes are located at identical positions. The largest is the *X. laevis* gene with the introns about 2 to 3 times longer than in other vertebrates. In addition, genomic sequences from *Caenorhabditis elegans* (GenBank Z75714), *Saccharomyces cerevisiae* (GenBank Z71372 and X94335) and *Schizosaccharomyces pombe* (GenBank Z68198) are available. *C. elegans* gene is lacking in introns at positions where vertebrates have the 2nd and 5th intron. Each of the two copies of *C. cerevisiae S7* gene have one intron which position corresponds to the mammalian 3rd intron.

Interestingly, U17 snoRNA is coded within the introns of the *Xenopus* and *Fugu* (10) *S7* gene, but not in the mouse or human *S7* gene. Human U17 is encoded within the cell cycle regulatory gene RCC1 (12); location of the mouse U17 gene is not known.

Mammalian ribosomal protein genes *S6*, *S14*, *S24* and *L7* have extensive homology in the first intron within each gene pair (13). When we compared the introns of mouse and human *S7* genes, only a short homologous region within the first intron was detected (mouse nucleotides 163...189, 22 identities of 27), while mouse and human *S6* have 63% identity in a 350 nucleotide region. Whether this homology reflects functional significance in *S7* remains unknown. All introns have perfect GT...AG consensus splice donor and acceptor sites and pyrimidine-rich sequence upstream of the 3' splice site (14). The initiation codon is located at the second exon and has a context GCCAUGU which is similar to optimal (15). The polyadenylation signal is located 12 nucleotides downstream from the TAA stop codon.

The promoter region of mouse *S7*, like many other ribosomal protein gene promoters, lacks a canonical TATA box, but contains a number of consensus elements described in promoters of vertebrate ribosomal protein genes. Three GA-binding protein (GABP, β-factor) sites (consensus CGGAAR) in inverted orientation are located at positions -16...-21, -37...-42, and -78...-83 (16). Sequences between nucleotides 55...65 (first untranslated exon) and 201...210 (antisense strand in the first intron) are very similar to the δ-factor binding element (consensus $^{C}/_{G}{}^{C}/_{G}C^{G}/_{T}$GCCATCT) described in the promoters of mouse *L30* and *L32* genes (17). At positions 27...35 in the 5' untranslated region and 213...221 in the first intron are

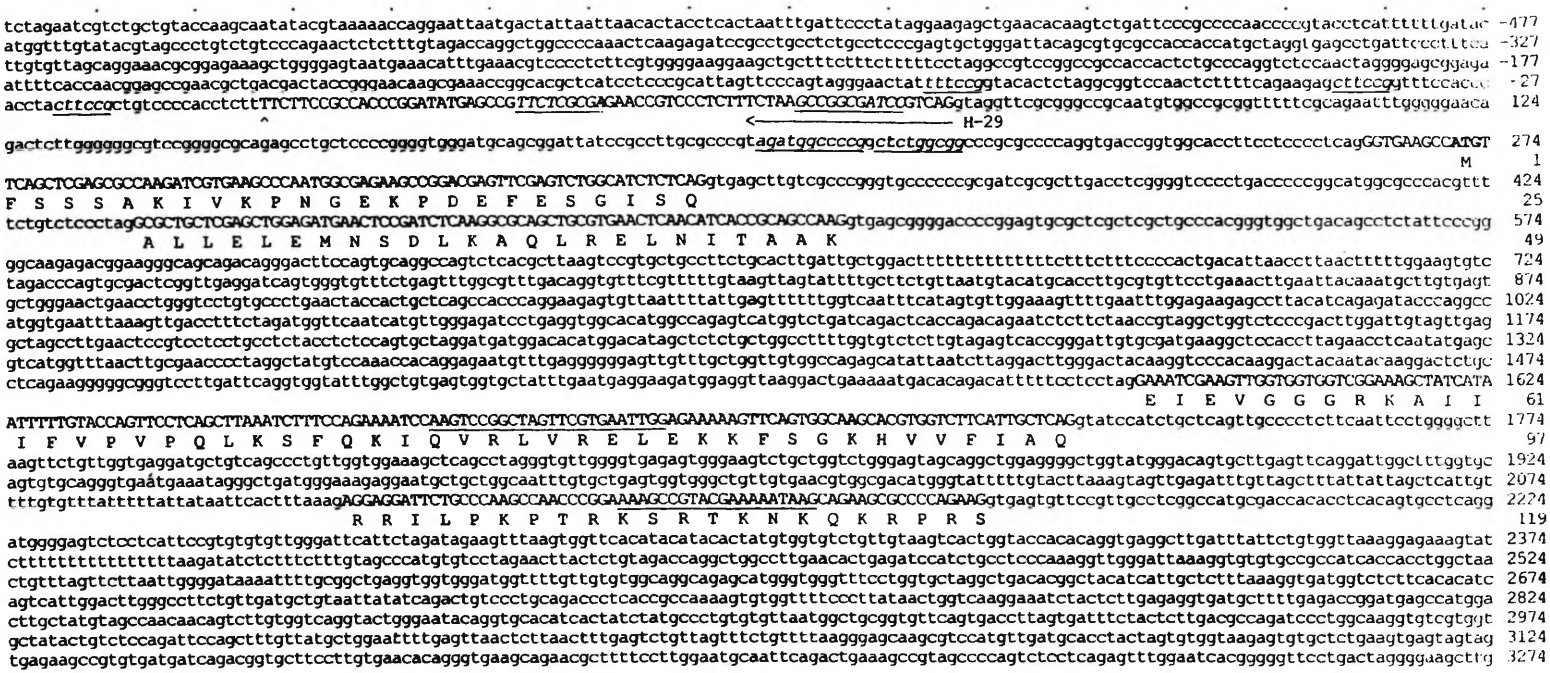

H-29

Fig. 1. The nucleotide sequence of the mouse *S7* gene. The numbering starts from the major transcription initiation site, indicated by the caret (^). The putative transcription factor binding sites (GABP, BoxA-like and δ-elements, see text for details) are in underlined italics. Exons are in uppercase letters, introns and flanking regions are in lowercase. Deduced amino acids are in a single letter code below the codons. Translation start codon and polyadenylation signal are shown in bold. Position of the oligonucleotide H-29 used for transcription start site mapping is indicated by arrowed line. Underlined sequences in the exons 4 and 5 indicate positions for two PCR primers used for amplification of the fourth intron.

located the sequences with very high homology to *L7a* BoxA element, which in *L7a* is conserved between mammals and chicken (18). δ-elements, BoxA-like and GABP elements can be found in the promoter region of all vertebrate *S7* genes, suggesting that these are essential for the *S7* promoter functioning. The region surrounding three first exons of mouse *S7* gene is GC rich (G+C content 61% compared to 48% in the whole sequenced region). The ratio of CpG/GpC dinucleotides in this area is 0.96. Presence of CpG island at the 5' region is also a common feature of ribosomal protein genes.

The amino acid sequence of S7 deduced from the cDNA was compared against the Genbank database using BLAST programs at http://www.ncbi.nlm.nih.gov/ (19). Predicted proteins for which complete cDNA or genomic sequence is known, are aligned in Fig. 2. A large number of EST's potentially coding S7 homologues from different eukaryotic species is also available, but as these sequences are partial or possibly erroneous, they were not included in the alignment. Mouse S7 protein is identical with human and rat S7 and highly homologous to other members of the eukaryotic S7 family. No archaebacterial or eubacterial homologue was found by database similarity search. Two highly conserved regions of the protein (amino acids 98...109 and 115...118) are involved in the nuclear and nucleolar import of S7 (Annilo et al., manuscript in preparation).

To determine the transcription initiation site for mouse S7 mRNA, we carried out a primer extension assay. The main extension product, corresponding to a T residue, 79 nucleotides upstream of the translation initiation codon, and minor bands at C nucleotides at 5 and 14 nucleotides farther upstream were detected (Fig. 3). These nucleotides are located within a pyrimidine-rich sequence; the main start point is embedded in a polypyrimidine tract of 13 nucleotides in length, which is usual among ribosomal protein genes.

Mouse *S7* gene family was examined by Southern blotting analysis. Using fragment from the *S7* intron 5 as a probe, the bands with the length as expected from the primary structure of the *S7* gene were detected. This demonstrates that intron-containing *S7* gene is present in the single copy. The same filter was then hybridised to mouse *S7* cDNA probe (Fig. 4). Detection of multiple cDNA-hybridising bands suggests that while the intron-containing *S7* is a single-copy gene, there are 10-12 additional copies (most likely processed pseudogenes) of *S7* in the mouse genome. All mammalian ribosomal protein genes studied to date are present in multiple copies; only S5 is present as a single copy in the rat genome (20).

```
mc MFSSSAKIVK -PNGEKPDEFE SGISQALLEL EMNS-DLKAQL RELNITAAKE
xl ...T..... -.......... .......... ....-..... ..........
fr ...T..... -.......... .......... ....-..... ..........
ag .VFG.-.VI. AG...-..A.. TQ.G..I... ....-...P.. .D.Y..R.R.
ms .---.T..L. AG-.TEA.S.. TS.....V.. .T..-..... ...Y..K...
ce .PEIIG.LL. -SD.KVVS.I. KQV....ID. .T.D-.VQS.. K..Y.VGV..
sp .SA-LN.... -RSSSQ.T.TD LLVA.C.YD. .SS.K.MAKE. .P.Q..S.R.
sA .SAPQ...L- ---SQA.T.L. LQVA..FV.. .NS.PE...E. .P.QFKSIR.
sB .S.VQS..L- ---SQA.S.L. LQVAKTFID. .SS.PE...D. .P.Q.KSIR.
lm .QPHLR.LR. -LKRG..SDL. ESVAK..F.. .ASHKT.RQE. PRFHVNTVR.

mc IEVGGGRKAI -IIFVPVPQLK SFQKIQVRLV RELEKKFSGK HVVFIAQRRI
xl ....A..... -.......... .......... .......... ..........
fr .....S.... -.......... .......... .......... ..........
ag V.F-NNK... -..Y....KQ. A...V.T... .......... ......E...
ms ..L-HNK.S. -..Y..M.K.. A.....I... .......... .....GD.K.
ce V.L.-NKS.. -..Y....... A.H..HPA.. .......G.R DILIL.K...
sp V.....K... -VV...Q.L.. A.H.C.A..T .......ADR ..I.......
sA .D.A..K..L -A.......S.A G.H.V.TK.T .......QDR ..I.L.E...
sB .D.T..K..L -VL.....A.S AYH.V.TK.T .......PDR ..I.L.E...
lm VK.PRSK.TV LVV.Y.LRF.M LVR.V.RA.T S....RHP.C L.MLV...Q.

mc LPKPTRKSRT K-NKQKRPRSR TLTAVHDAIL EDLVFPSEIV GKRIRVKLDG
xl .......... .-.......... .......... ....Y..... .R........
fr .........S .-.......... ...S...... .......... .......M.S
ag ....M.GR.D P-.........P NV...Y.... ......A.V. ..........
ms ....SH.T.V A-.......... ...S.Y.... ......A.... ..........
ce ....Q.G.KA RPQ......... ........W. DE..Y.A.V. .R........
sp ....G....V --T......... ......N... ..I...T..I ...T.QAT..
sA ....S.T..- --QV........ .......K.. ......T.... ...V.YLVG.
sB ....S.T..- --QV........ .......KV. ..M...T.... ...V.YLVG.
lm TKR..DVYKL -KQV.P.K..- -...FEN... N.MIY..DV. .R.W.CRT..

mc SRLIKVHLDK AQQNNVEHKV ETFSGVYKKL TGKDVNFEF- --PEFQL
xl .......... .......... .......... .....V...- --.....
fr .......... .......... .......... .....V...- --.....
ag .Q........ N..TTI.... D..AS..... ..R..T...- --..NY.
ms .Q........ N..TTI.... D..QS..... ..RE.T...- --..PY.
ce KKVY...... SH.T..G..I GV.AS..R.. .....T...- --.DPIF
sp RKT...F..N RDA.T.DY.L GS..S..H.. ...N.T...- --.VATGEGL
sA NKIQ..L..S KDVQQIDY.L .S.QA..N.. ...QIV..I- --.SETH
SB NKIQ..L..S KDVQQIDY.L .S.QA..N.. ...QIV..I- --.SQTN
lm .K.M..F..A RDRKRI.SRL RAVAY...Q. .HRR.S.G.M WN.KL.QVSTR
```

Fig. 2. Alignment of ten eukaryotic S7 amino acid sequences. Dots indicate identity to mammalian consensus, dashes indicate gaps to maximize homology. mc, mammalian consensus, based on 100% identity of mouse, rat and human S7; xl, *Xenopus laevis;* fr, *Fugu rubripes*; ag, *Anopheles gambiae;* ms, *Manduca sexta*; ce, *Caenorhabditis elegans*; sp, *Schizosaccharomyces pombe* (sequence starts from the amino acid 14); sA and sB, *Saccharomyces cerevisiae* sequences from chromosome XV and XIV, respectively; lm, *Leishmania major*.

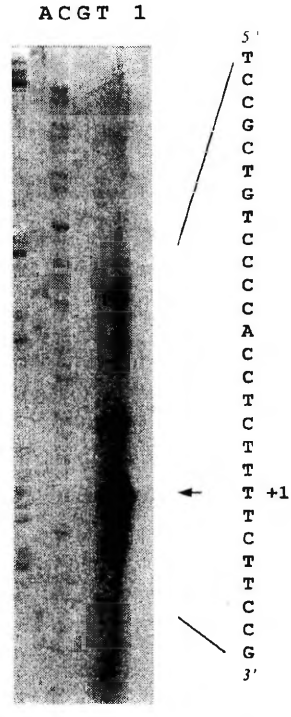


Fig. 3. Mouse S7 transcription start site mapping. Oligonucleotide H-29 (see Fig. 1) was end-labeled and used in a primer extension reaction with 20 μg of total mouse RNA. The extension products were analysed on a 6% acrylamide-urea gel. Lane 1 contains primer extension products, lanes ACGT show sequencing ladder of antisense DNA strand with the same primer. The main extension product is indicated by an arrow.

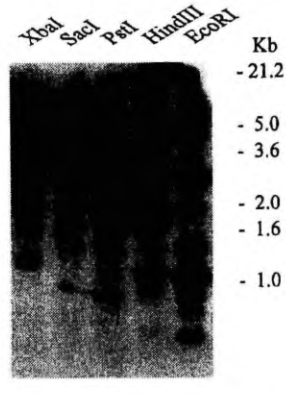


Fig. 4. Southern blot analysis of the mouse *S7* gene family. Genomic DNA from 129SV mouse was digested with restriction endonucleases EcoRI, HindIII, PstI, SacI and XbaI, resolved by electrophoresis in a 0.8% agarose gel, and transferred to Hybond-N+ membrane. Hybridization was performed with a mouse *S7* cDNA. DNA size marker is shown to the right.

The *S7* sequence has been deposited in the GenBank under the accession number AF043285.

ACKNOWLEDGMENTS

This study was supported by research grant 2491 from the Estonian Science Foundation.

REFERENCES

1. Wool, I.G., Chan, Y.-L. and Glück, A. (1995) Biochem. Cell Biol. 73, 933-947.
2. Johnson, K.R., Cook, S.A. and Davisson, M.T. (1994) Mamm. Genome 5, 670-687.
3. Fisher, E.M., Beer, R.P., Brown, L.G., Ridley, A., McNeil, J.A., Lawrence, J.B., Willard, H.F., Bieber, F.R. and Page, D.C. (1990) Cell 63, 1205-1218.
4. Davies, B., Feo, S., Heard, E. and Fried, M. (1989) Proc. Natl. Acad. Sci. USA 86, 6691-6695.
5. Witte, S. and Krawinkel, U. (1997) J. Biol. Chem. 272, 22243-22247.
6. Nygård, O. and Nika, H. (1982) EMBO J. 1, 357-362.
7. Westermann, P. and Nygård, O. (1983) Biochim. Biophys. Acta 741, 103-108.
8. Mundus, D.A., Bulygin, K.N., Ven'yaminova, A.G., Vladimirov, S.N., Vratskikh, L.V., Repkova, M.N., Yamkovoi, V.S. and Karpova, G.G. (1993) Mol. Biol. 27, 91-95.
9. Annilo, T., Laan, M., Stahl, J. and Metspalu, A. (1995) Gene 165, 297-302.
10. Cecconi, F., Crosio, C., Mariottini, P., Cesareni, G., Giorgi, M., Brenner, S. and Amaldi, F. (1996) Nucleic Acids Res. 24, 3167-3172.
11. Mariottini, P., Bagni, C., Francesconi, A., Cecconi, F., Serra, M.J., Chen, Q., Loreni, F., Annesi, F. and Amaldi, F. (1993) Gene 132, 255-260.
12. Kiss, T. and Filipowicz, W. (1993) EMBO J 12, 2913-2920.
13. Pata, I. and Metspalu, A. (1996) Gene 175, 241-245.
14. Breathnach, R. and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349-383.
15. Kozak, M. (1986) Cell 44, 283-292.
16. Genuario, R.R., Kelley, D.E. and Perry, R.P. (1993) Gene Expr. 3, 279-288.
17. Hariharan, N., Kelley, D.E. and Perry, R.P. (1989) Genes Dev. 3, 1789-1800.
18. Colombo, P. and Fried, M. (1992) Nucleic Acids Res. 20, 3367-3373.
19. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389-3402.
20. Kuwano, Y., Olvera, J. and Wool, I.G. (1992) J. Biol. Chem. 267, 25304-25308.

CURRICULUM VITAE

Tarmo Annilo

Date of birth:	June 16, 1965, Tartu
Citizenship:	Estonian
Personal status:	Married, two children
Address work:	University of Tartu, Institute of Molecular and Cell Biology, Riia 23, Tartu 51010, Estonia Tel: +372 7 375 034 Fax: +372 7 420 286 e-mail: tannilo@ebc.ee
Address home:	Tartu, Mõisavahe 35–31, Tartu 50708

Education

1983	finished Tartu Secondary School No 5
1991	graduated from University of Tartu, Estonia; Faculty of Biology and Geography, as molecular biologist

Scientific work

Since 1990 I have been working in Dr. Andres Metspalu's lab at University of Tartu, studying mammalian ribosomal protein genes and intracellular transport of ribosomal proteins.

CURRICULUM VITAE

Tarmo Annilo

Sünniaeg:	16. juuni 1965, Tartu
Kodakondsus:	Eesti
Perekonnaseis:	Abielus, kaks last
Aadress tööl:	Tartu Ülikooli Molekulaar- ja Rakubioloogia Instituut, Riia 23, Tartu 51010 Tel: (27) 375 034 Faks: (27) 420 286 e-mail: tannilo@ebc.ee
Aadress kodus:	Mõisavahe 35–31, Tartu 50708

Haridus

1983	lõpetanud Tartu 5. Keskkooli
1991	lõpetanud Tartu Ülikooli bioloogia-geograafia-teaduskonna

Teadustegevus

Olen uurinud imetajate ribosoomivalkude geene ning ribosoomivalkude rakusisest transporti. Praegu töötan ribosoomivalkude võimaliku seotusega pärilike haigustega ja multifunktsionaalsusega.

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS

1. **Toivo Maimets**. Studies of human oncoprotein p53. Tartu, 1991, 96 p.
2. **Enn K. Seppet**. Thyroid state control over energy metabolism, ion transport and contractile functions in rat heart. Tartu, 1991, 135 p.
3. **Kristjan Zobel**. Epifüütsete makrosamblike väärtus õhu saastuse indikaatoritena Hamar-Dobani boreaalsetes mägimetsades. Tartu, 1992, 131 lk.
4. **Andres Mäe**. Conjugal mobilization of catabolic plasmids by transposable elements in helper plasmids. Tartu, 1992, 91 p.
5. **Maia Kivisaar**. Studies on phenol degradation genes of *Pseudomonas* sp. strain EST 1001. Tartu, 1992, 61 p.
6. **Allan Nurk**. Nucleotide sequences of phenol degradative genes from *Pseudomonas sp.* strain EST 1001 and their transcriptional activation in *Pseudomonas putida*. Tartu, 1992, 72 p.
7. **Ülo Tamm**. The genus *Populus* L. in Estonia: variation of the species biology and introduction. Tartu, 1993, 91 p.
8. **Jaanus Remme**. Studies on the peptidyltransferase centre of the *E.coli* ribosome. Tartu, 1993, 68 p.
9. **Ülo Langel**. Galanin and galanin antagonists. Tartu, 1993, 97 p.
10. **Arvo Käärd**. The development of an automatic online dynamic fluorescense-based pH-dependent fiber optic penicillin flowthrought biosensor for the control of the benzylpenicillin hydrolysis. Tartu, 1993, 117 p.
11. **Lilian Järvekülg**. Antigenic analysis and development of sensitive immunoassay for potato viruses. Tartu, 1993, 147 p.
12. **Jaak Palumets**. Analysis of phytomass partition in Norway spruce. Tartu, 1993, 47 p.
13. **Arne Sellin**. Variation in hydraulic architecture of *Picea abies* (L.) Karst. trees grown under different enviromental conditions. Tartu, 1994, 119 p.
13. **Mati Reeben**. Regulation of light neurofilament gene expression. Tartu, 1994, 108 p.
14. **Urmas Tartes**. Respiration rhytms in insects. Tartu, 1995, 109 p.
15. **Ülo Puurand**. The complete nucleotide sequence and infections *in vitro* transcripts from cloned cDNA of a potato A potyvirus. Tartu, 1995, 96 p.
16. **Peeter Hõrak**. Pathways of selection in avian reproduction: a functional framework and its application in the population study of the great tit (*Parus major*). Tartu, 1995. 118 p.
17. **Erkki Truve**. Studies on specific and broad spectrum virus resistance in transgenic plants. Tartu, 1996. 158 p.
18. **Illar Pata**. Cloning and characterization of human and mouse ribosomal protein S6-encoding genes. Tartu, 1996. 60 p.

19. **Ülo Niinemets**. Importance of structural features of leaves and canopy in determining species shade-tolerance in temperature deciduous woody taxa. Tartu, 1996. 150 p.
20. **Ants Kurg**. Bovine leukemia virus: molecular studies on the packaging region and DNA diagnostics in cattle. Tartu, 1996. 104 p.
21. **Ene Ustav**. E2 as the modulator of the BPV1 DNA replication. Tartu 1996. 100 p.
22. **Aksel Soosaar**. Role of helix-loop-helix and nuclear hormone receptor transcription factors in neurogenesis. Tartu, 1996. 109 p.
23. **Maido Remm**. Human papillomavirus type 18: replication, transformation and gene expression. Tartu, 1997. 117 p.
24. **Tiiu Kull**. Population dynamics in *Cypripedium calceolus* L. Tartu, 1997. 124 p.
25. **Kalle Olli**. Evolutionary life-strategies of autotrophic planktonic microorganisms in the Baltic Sea. Tartu, 1997. 180 p.
26. **Meelis Pärtel**. Species diversity and community dynamics in calcareous grassland communities in Western Estonia. Tartu, 1997. 124 p.
27. **Malle Leht**. The Genus *Potentilla* L. in Estonia, Latvia and Lithuania: distribution, morphology and taxonomy. Tartu, 1997. 186 p.
28. **Tanel Tenson**. Ribosomes, peptides and antibiotic resistance. Tartu, 1997. 80 p.
29. **Arvo Tuvikene**. assessment of inland water pollution using biomarker responses in fish *in vivo* and *in vitro*. Tartu, 1997. 160 p.
30. **Urmas Saarma**. Tuning ribosomal elongation cycle by mutagenesis of 23S rRNA. Tartu, 1997. 134 p.
31. **Henn Ojaveer**. Composition and dynamics of fish stocks in the gulf of Riga ecosystem. Tartu, 1997. 138 p.
32. **Lembi Lõugas**. Post-glacial development of vertebrate fauna in Estonian water bodies. Tartu, 1997. 138 p.
33. **Margus Pooga**. Cell penetrating peptide, transportan, and its predecessors, galanin-based chimeric peptides. Tartu,1998. 110 p.
34. **Andres Saag**. Evolutionary relationships in some cetrarioid genera (Lichenized Ascomycota). Tartu, 1998. 196 p.
35. **Aivar Liiv**. Ribosomal large subunit assembly *in vivo*. Tartu, 1998. 158 p.
36. **Tatjana Oja**. Isoenzyme diversity and phylogenetic affinities among the eurasian annual bromes (*Bromus* L., Poaceae). Tartu, 1998. 92 p.
37. **Mari Moora**. The influence of arbuscular mycorrhizal (AM) symbiosis on the competition and coexistence of calcareous crassland plant species. Tartu, 1998. 78 p.
38. **Olavi Kurina**. Fungus gnats in Estonia (*Diptera: Bolitophilidae, Keroplatidae, Macroceridae, Ditomyiidae, Diadocidiidae, Mycetophilidae*). Tartu, 1998. 200 p.
39. **Andrus Tasa**. Biological leaching of shales: black shale and oil shale. Tartu, 1998. 98 p.
40. **Arnold Kristjuhan.** Studies on transcriptional activator properties of tumor suppressor protein p53. Tartu, 1998. 86 p.

41. **Sulev Ingerpuu.** Characterization of some human myeloid cell surface and nuclear differentiation antigens. Tartu, 1998. 163 p.
42. **Veljo Kisand.** Responses of planktonic bacteria to the abiotic and biotic factors in the shallow lake Võrtsjärv. Tartu, 1998. 118 p.
43. **Kadri Põldmaa.** Studies in the systematics of hypomyces and allied genera (Hypocreales, Ascomycota). Tartu, 1998. 178 p.
44. **Markus Vetemaa.** Reproduction parameters of fish as indicators in environmental monitoring. Tartu, 1998. 117 p.
45. **Heli Talvik.** Prepatent periods and species composition of different *Oesophagostomum* spp. populations in Estonia and Denmark. Tartu, 1998. 104 p.
46. **Katrin Heinsoo.** Cuticular and stomatal antechamber conductance to water vapour diffusion in *Picea abies* (L.) karst. Tartu, 1998.

ISSN 1024–6479
ISBN 9985–56–378–6