Fugu intron oversize reveals the presence of U15 snoRNA coding sequences in some introns of the ribosomal protein S3 gene.

C Crosio, F Cecconi, P Mariottini, et al.

Genome Res. 1996 6: 1227-1231
Access the most recent version at doi:10.1101/gr.6.12.1227

LETTER

Fugu Intron Oversize Reveals the Presence of U15 snoRNA Coding Sequences in Some Introns of the Ribosomal Protein S3 Gene

Claudia Crosio,[1] Francesco Cecconi,[1] Paolo Mariottini,[2] Gianni Cesareni,[1] Sydney Brenner,[3] and Francesco Amaldi[1,4]

[1]Department of Biology, University of Rome, "Tor Vergata," 00133 Rome, Italy; [2]Department of Biology, Terza University of Rome, 00154 Rome, Italy; [3]King's College, Cambridge CB2 1ST, United Kingdom

We present here the analysis of the genomic organization of the *Fugu* gene coding for ribosomal protein S3 and its intron encoded U15 RNA, and compare it with the homologous human and *Xenopus* genes. Only two of the six *Fugu* S3 gene introns do not contain the U15 sequence and are in fact shorter than 100 nucleotides, as most *Fugu* introns. The other four introns are somewhat longer and contain sequences homologous to U15 RNA; two of these represent functional copies, as shown by microinjections of *Fugu* transcripts into *Xenopus* oocytes, whereas the other two appear to be nonfunctional pseudocopies. Thus *Fugu* turns out to be ideal for the study of intron encoded snoRNAs, partly because of the reduced cloning and sequencing workload, and partly because the intron length per se can be an indication of the presence of a snoRNA coding sequence.

[The sequence data described in this paper have been submitted to the EMBL data library under accession no. X97794.]

The fish *Fugu rubripes* has, among vertebrates, a particularly compact genome, which is approximately eight times smaller than that of mammals (Brenner et al. 1993). This is in part due to the small size of most introns, which have a modal length of less than 100 nucleotides. This seems to be particularly favorable when studying the organization of those genes that host the coding sequences for small nucleolar RNAs in their introns (Seraphin 1993; Sollner-Webb 1993; Maxwell and Fournier 1995; Steitz and Tycowski 1995).

Ever since the intron localization has been revealed for mouse U14 RNA gene (Liu and Maxwell 1990), the number of intron-encoded small nucleolar RNAs (snoRNAs) in various vertebrates has grown to about fifty and is still increasing fast. In general, the host genes code for ribosomal proteins (r-proteins) (Fragapane et al. 1993; Kiss and Filipowicz 1993; Prislei et al. 1993; Cecconi et al. 1994; Qu et al. 1994; Nicoloso et al. 1996) or for other proteins involved in the production and function of the translation apparatus (Liu and Maxwell 1990; Nag et al. 1993; Nicoloso et al. 1994), or appear to have lost function (Tycowski et al. 1996). All these snoRNAs are produced by processing of the host gene pre-mRNA and present one or two regions complementary to the rRNA (Bachellerie et al. 1995) that have been shown to be implicated, in the case of the fibrillarin-associated snoRNAs, in the site-specific 2′-O-methylation of pre-rRNA (Kiss-László et al. 1996). In particular, U15 RNA has been found to be encoded at least in the first intron of the human *rpS3* gene (Tycowski et al. 1993) and in three of the six introns of the *Xenopus* gene for the same r-protein (Pellizzoni et al. 1994) (formerly *rpS1*, here referred to as *rpS3*, adopting the rat nomenclature as the unified system). The multiple copies of U15 RNA coding sequences are quite divergent, conserving well only the two ends of the molecules containing the C and D boxes, the sequences complementary to the 28S rRNA, and a seven-nucleotide loop in the middle of the molecule. The remaining part of the sequence is very divergent, although it is always able to acquire the same secondary structure (Pellizzoni et al. 1994).

We present here the cloning and analysis of the *rpS3* gene of *Fugu* and of the U15 RNA coding sequences hosted in its introns, and a comparison with the *Xenopus* and human counterparts.

[4]Corresponding author.
E-MAIL amaldi@utovrm.it; FAX 39-6-72594316.

RESULTS AND DISCUSSION

Fugu genomic fragment hybridizing to two *Xenopus* probes, for rpS3 and for U15 RNA, has been isolated and sequenced as described in Methods. The resulting nucleotide sequence with its flanking regions is shown in Figure 1. Exon/intron boundaries and the 5′ and 3′ ends of the *Fugu* gene have been identified through comparison with the *Xenopus* r-protein S3 cDNA and gene. Transcription start site is located within a pyrimidine tract, as in all vertebrate r-protein genes (Hariharan et al. 1989; Pierandrei-Amaldi and Amaldi 1994; Meyuhas et al. 1996). The gene is made up by seven exons as its *Xenopus* homologs. The position of the six introns is conserved perfectly in the two species and, at least, with that of the first three introns of the human gene. The size difference between the *Fugu* gene (3593 bp) and the *Xenopus* counterpart (12,691 bp) is entirely a result of the smaller size of the introns. Although these are all rather short, only two of them, the second (90 nucleotides) and the third (84 nucleotides) have the typically very small size of most *Fugu* introns. The remaining four introns range between 341 and 891 nucleotides. Two of these (introns 4 and 6) contain a U15 RNA sequence, 76% homologous with respect to each other, and somewhat less (63–71%) with respect to the *Xenopus* and human U15 sequences. The regions corresponding to the C and D boxes, the sequences complementary to 28S RNA and the 7-nucleotide loop mentioned above, are all very well conserved. In spite of the considerable divergence of the rest of the molecule, the computer-derived secondary structure of these *Fugu* U15 RNA (not shown) is substantially identical to the one proposed for the human and *Xenopus* RNAs (Pellizzoni et al. 1994).

Careful inspection of introns 1 and 5, which are also longer than most *Fugu* introns, revealed the presence of sequences (broken underline in Fig. 1) that, though degenerated, can still be recognized as related to the U15 sequence, particularly in the region corresponding to the C and D boxes and to the sequence complementary to 28S rRNA. Attempts to derive a computer secondary structure for these U15 pseudocopies comparable to the canonical one have failed. To verify if these were in fact pseudocopies, we tested their capacity to be properly processed by microinjecting various in vitro transcribed radioactive RNAs, corresponding to different regions of the *Fugu rpS3* gene, into *Xenopus* oocytes, a system that also works interspecifically (Cecconi et al. 1995). While the transcripts of introns 4 and 6 are correctly and efficiently processed to produce mature U15 RNA, the transcripts of introns 1 and 5

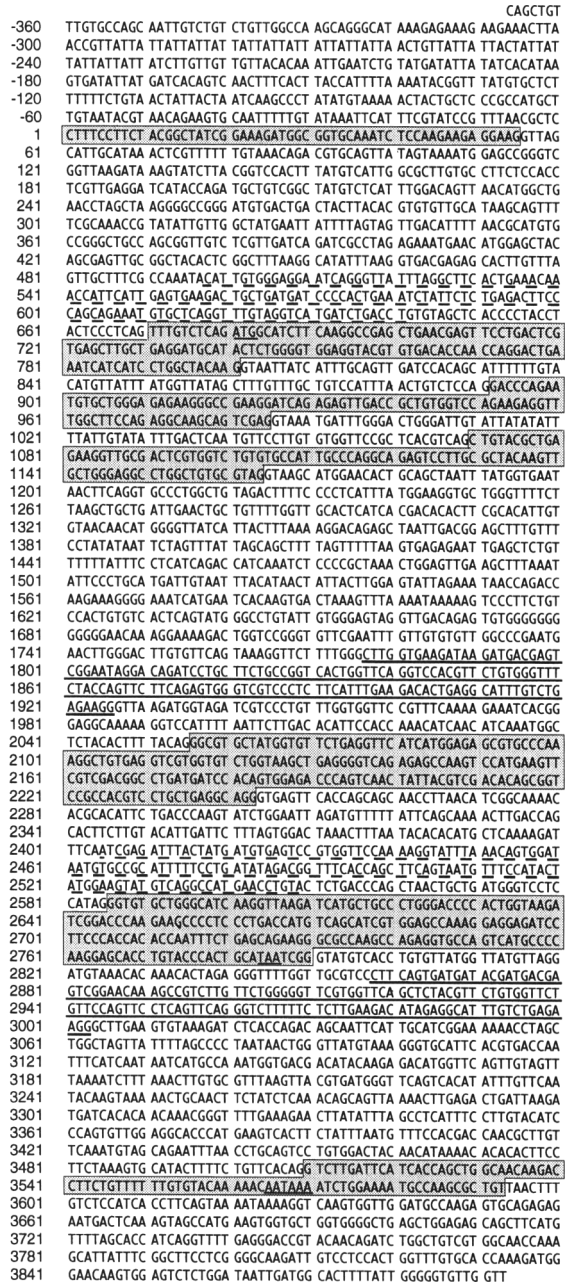

Figure 1 Nucleotide sequence of the *Fugu S3/U15* gene and its flanking regions. The sequence is numbered relative to the transcription start point. The seven exons are boxed and shaded. Pseudo and canonical U15 sequence copies are underlined (solid and broken lines, respectively) within the introns. The start codon is underlined within the second exon; the stop codon and the polyadenylation signal are underlined within the sixth and the last exon, respectively.

are completely degraded, confirming that the U15 sequences present in these two introns are nonfunctional pseudocopies (examples of these experiments are shown in Fig. 2).

Recently, we have found a similar situation for *Fugu* U17 snoRNA, which is encoded in the introns of the *rpS7* gene. In this case no "empty" intron was found. In fact all six introns were relatively long: four contained canonical U17 sequence copies and two contained pseudocopies (Cecconi et al. 1996).

Figure 3 compares the overall organization of the *Fugu rpS3/U15* gene with that of the *Xenopus* counterpart and part of the human gene. Although the coding sequences for U15 RNA have remained hosted in the same *rpS3* gene during the evolution of vertebrates, they have moved from one to another intron. This observation, together with the presence of U15 sequence pseudocopies in other introns, indicates a highly mobile situation for intron nested snoRNA coding sequences, possibly because of sequence duplications followed by divergence.

Interestingly, the two introns that do not contain the U15 RNA sequence, neither canonical nor pseudocopy, present the very reduced size

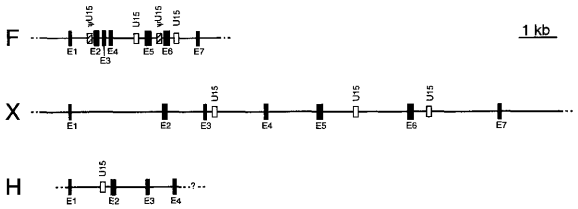

Figure 3 Comparison of the *S3/U15* gene organization of *Fugu* (F) with that of *Xenopus* (X) (Pellizzoni et al. 1994) and with the available portion of the human gene (H) (Tycowski et al. 1993). Exons are represented by solid boxes. Open and hatched boxes represent, respectively, the canonical and pseudocopies of U15 sequences.

(<100 nucleotides) typical of most *Fugu* introns. This indicates that the presence of U15 sequences is the reason for the relatively large size of the other four introns, and supports the assumption that a specific structural or functional role of intronic sequences can explain the other described exceptions of long introns in *Fugu* genes (Aparicio et al. 1995; Baxendale et al. 1995; Macrae and Brenner 1995; Mason et al. 1995).

The finding that the length of the snoRNA containing introns exceeds the one expected from simple addition of the snoRNA sequence to an average short *Fugu* intron seems to exclude a transposition mechanism neatly involving the repeated sequences, suggesting duplications of more extensive regions that include the snoRNA coding sequences.

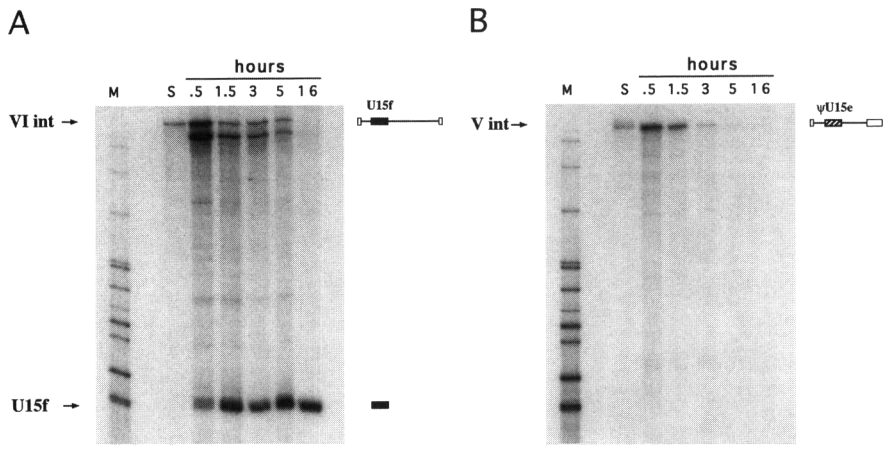

Figure 2 Analysis of *Fugu* U15 snoRNA production by RNA microinjections into *Xenopus* oocyte nuclei. (*A*) An in vitro synthesized 800-nucleotide radioactive RNA (VIi), corresponding to part of the sixth exon, the entire sixth intron bearing the U15f copy, and part of the seventh exon of the *Fugu S3* gene, was injected in *Xenopus* oocytes. After incubation, at increasing time intervals (30 min to 5 hr), total RNA was extracted and analyzed by gel electrophoresis and autoradiography. Arrows point to the intact injected transcript and to the mature product. A schematic representation of the RNA molecules is shown at *right*. (*B*) An in vitro synthesized 537-nucleotide radioactive RNA (Vi), corresponding to part of the fifth exon, the entire fifth intron bearing the ΨU15e copy and part of the sixth exon of the *Fugu S3* gene, was used and analyzed as in *A*. (M) size markers; (S) transcript substrate.

METHODS

Cloning and Sequencing

F. rubripes cosmid library (Baxendale et al. 1995) has been probed with a *Xenopus* cDNA specific for r-protein S3 (Di Cristina et al. 1991) and a *Xenopus* genomic fragment containing a U15 snoRNA gene copy (Pellizzoni et al. 1995). The finding of clones hybridizing to both probes suggested that the U15 RNA coding sequence is hosted in the S3 gene of *Fugu* as it occurs in *Xenopus* and in man. The selected cosmid 136B1 has been digested with *Hin*dIII to generate a 2500-bp

fragment containing the 5′region and the central portion of the gene and with *Hin*dIII–*Xba*I to generate a 2400-bp fragment containing the 3′ region of the gene. These fragments have been cloned in the pEMBL18 and pBluescript KS(+) vectors, respectively (clones pF-S3.1 and pF-S3.2). A PCR amplification on the 136B1 cosmid was performed to check whether the two clones were contiguous. For sequencing, plasmids pF-S3.1 and pF-S3.2 were digested with various restriction enzymes to generate overlapping fragments and subcloned in the Bluescript KS(+) vector. Analysis and manipulation of DNA and RNA were performed according to standard laboratory manuals (Sambrook et al. 1989). Sequencing was performed by the dideoxy chain-termination method (Sanger et al. 1977) on both strands of overlapping fragments.

In Vitro Synthesis of Radioactive Transcripts

To prepare the transcripts to be used as processing substrates in microinjected oocytes, various genomic fragments encompassing the U15 sequences, derived from plasmids pF-S3.1 and pF-S3.2, were cloned under the T7 promoter in pBluescript: pF-S3.10 and pF-S3.11 (nucleotides 2195–2732 and 2732–3532, respectively, in Fig. 1). For both pF-S3.10 and pF-S3.11 plasmids, 1 μg of DNA was digested with *Eco*RI and transcribed with T7 RNA polymerase, in the presence of 50 μCi of [α-^{32}P]UTP, as described by Melton et al. (1984). The 537-nucleotide long (Vi) and 800-nucleotide long (VIi) transcripts were obtained from pF-S3.10 and pF-S3.11 plasmids, respectively. After transcription and DNase digestion, the RNAs were purified by phenol-chloroform-isoamyl alcohol (50:50:1) extraction and ethanol precipitation, and resuspended in H_2O for microinjection.

RNA Microinjections in *Xenopus* Oocytes

Isolation of *Xenopus* stage V-VI oocytes, microinjection of RNA into the germinal vesicle, oocyte incubation, manual isolation of germinal vesicle, RNA extraction, and polyacrylamide electrophoresis analysis were all carried out as described previously (Cecconi et al. 1995).

ACKNOWLEDGMENTS

We thank M. Giorgi for expert technical assistance. This work was supported by grants from Progetto Finalizzato Ingegneria Genetica, C.N.R., and from Ministero Università e Ricerca Scientifica e Tecnologica.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Aparicio, S., A. Morrison, A. Gould, J. Gilthorpe, C. Chaudhuri, P. Rigby, R. Krumlauf, and S. Brenner. 1995. Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, *Fugu rubripes*. *Proc. Natl. Acad. Sci.* **92:** 1684–1688.

Bachellerie, J.-P., M. Michot, A. Nicoloso, J. Balakin, J. Ni, and M.J. Fournier. 1995. Antisense snoRNAs: A family of nucleolar RNAs with long complementarities to rRNA. *Trends Biochem. Sci.* **20:** 261–264.

Baxendale, S., S. Abdulla, G. Elgar, D. Buck, M. Borcks, G. Micklem, R. Durbin, G. Bates, S. Brenner, S. Beck, and H. Lehrach. 1995. Comparative sequence analysis of the human and pufferfish Huntington's disease genes. *Nature Genet.* **10:** 67–75.

Brenner, S., G. Elgar, R. Sandford, A. Macrae, B Venkatesh, and S. Aparicio. 1993. Characterization of the pufferfish (*Fugu*) genome as a compact model vertebrate genome. *Nature* **366:** 265–268.

Cecconi, F., P. Mariottini, F. Loreni, P. Pierandrei-Amaldi, N. Campioni, and F. Amaldi. 1994. U17^{XS8}, 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:** 732–741.

Cecconi, F., P. Mariottini, and F. Amaldi. 1995. The *Xenopus* intron-encoded U17 snoRNA is produced by exonucleolytic processing of its precur in oocytes. *Nucleic Acids Res.* **23:** 467-0-4676.

Cecconi, F., C. Crosio, P. Mariottini, G. Cesareni, M. Giorgi, S. Brenner, and F. Amaldi. 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.

Di Cristina, M., R. Menard, and P. Pierandrei-Amaldi. 1991. *Xenopus laevis* ribosomal protein S1a cDNA sequence. *Nucleic Acids Res.* **19:** 1943.

Fragapane, P., S. Prislei, A. Michienzi, E. Caffarelli, and I. Bozzoni. 1993. A novel small nucleolar RNA (U16) is encoded inside a ribosomal protein intron and originates by processing of pre-mRNA. *EMBO J.* **12:** 2921–2928.

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

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

Kiss-László, Z., Y. Henry, J.P. Bachellerie, M. Caizergues-Ferrer, and T. Kiss. 1996. Site-specific methylation of preribosomal RNA: A novel function for small nucleolar RNAs. *Cell* **85:** 1077–1088.

Liu, J. and E.S. Maxwell. 1990. Mouse U14 snRNA is encoded in an intron of the mouse cognate hsc 70 heat shock gene. *Nucleic Acids Res.* **18:** 6565–6571.

Macrae, A.D. and S. Brenner. 1995. Analysis of the

dopamine receptor family in the compact genome of the puffer fish *Fugu rubripes*. *Genomics* **25:** 436–446.

Mason, P.J., D.J. Stevens, L. Luzzatto, S. Brenner, and S. Aparicio. 1995. Genomic structure and sequence of the *Fugu rubripes* glucose-6-phosphate dehydrogenase gene (G6PD). *Genomics* **26:** 587–591.

Maxwell, E.S. and M.J. Fournier. 1995. The small nucleolar RNAs. *Annu. Rev. Biochem.* **35:** 897–934.

Melton, D.A., P.A. Krieg, M.R. Rebagliati, T. Maniatis, K. Zinn, and M.R. Green. 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmid containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* **12:** 7035–7056.

Meyuhas, O., D. Avni, and S. Shama. 1996. Translational control of ribosomal protein mRNAs in eukaryotes. In *Translational control* (ed. J.W.B. Hershey, M.B. Mathews, and N. Sonenberg), pp. 363–388. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Nag, M.K., T.T. Thai, E.A. Ruff, N. Selvamurugan, M. Kunnimalaiyaan, and G.L. Eliceiri. 1993. Genes for E1, E2, and E3 small nucleolar RNAs. *Proc. Natl. Acad. Sci.* **90:** 9001–9005.

Nicoloso, M., M. Caizergues-Ferrer, B. Michot, M.C. Azum, and J.P. Bachellerie. 1994. U20, a novel small nucleolar RNA, is encoded in an intron of the nucleolin gene in mammals. *Mol. Cell. Biol.* **14:** 5766–5776.

Nicoloso, M., L.H. Qu, B. Michot, and J.P. Bachellerie. 1996. Intron-encoded, antisense small nucleolar RNAs: The characterization of nine novel species points to their direct role as guides for the 2′-O-ribose methylation of rRNAs. *J. Mol. Biol.* **260:** 178-195.

Pellizzoni, L., C. Crosio, N. Campioni, F. Loreni, and P. Pierandrei-Amaldi. 1994. Different forms of U15 snoRNA are encoded in the introns of the ribosomal protein S1 gene of *Xenopus laevis*. *Nucleic Acids Res.* **22:** 4607–4613.

Pellizzoni, L., C. Crosio, and P. Pierandrei-Amaldi. 1995. Sequence of the cDNA and gene coding for ribosomal protein S1 of *Xenopus laevis*. *Gene* **154:** 145–151.

Pierandrei-Amaldi, P. and F. Amaldi. 1994. Aspects of regulation of ribosomal protein synthesis in *Xenopus laevis*: Review. *Genetica* **94:** 181–193.

Prislei, S., C. Michienzi, C. Presutti, P. Fragapane, and I. Bozzoni. 1993. Two different snoRNAs are encoded in introns of amphibian and human L1 ribosomal protein genes. *Nucleic Acids Res.* **21:** 5824–5830.

Qu, L.H., M. Nicoloso, B. Michot, M.C. Azum, M. Caizergues-Ferrer, M.H. Renalier, and J.P. Bachellerie. 1994. U21, a novel small nuclolar RNA with a 13 nt. complementarity to 28S rRNA, is encoded in an intron of ribosomal protein L5 gene in chicken and mammals. *Nucleic Acids Res.* **22:** 4073–4081.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **74:** 5463–5467.

Seraphin, B. 1993. How many intronic snRNAs? *Trends Biochem. Sci.* **18:** 330–331.

Sollner-Webb, B. 1993. Novel intron-encoded small nucleolar RNAs. *Cell* **75:** 403–405.

Steitz, J.A. and K.T. Tycowski. 1995. Small RNA chaperones for ribosome biogenesis. *Science* **270:** 1626–1627.

Tycowski, K.T., M.-D. Shu, and J.A. Steitz. 1993. A small nucleolar RNA is processed from an intron of the human gene encoding ribosomal protein S3. *Genes & Dev.* **78:** 1176–1190.

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

Received August 2, 1996; accepted in revised form September 9, 1996.