A cDNA cloned from pregnant mouse uterus exhibits temporo-spatial expression and predicts a novel protein

John W. KASIK

Department of Pediatrics, Metro Health Medical Center, Case Western Reserve University, Room R-249, 2500 Metro Health Drive, Cleveland, 44109 OH, U.S.A.

A cDNA was cloned from a pregnant mouse uterus cDNA library. On conceptual translation, the cDNA has one long open reading frame that predicts a novel protein of 606 amino acids. This protein is principally composed of two CUB domains and a ZP domain; motifs found in proteins implicated in egg-sperm recognition. Probes derived from the cDNA were used to conduct Northern hybridizations. The expression of this mRNA is temporal; message first appears in the uterus 6 days prior to birth, it increases each subsequent day to attain maximal levels at 3 days prior to birth and then abruptly decreases during the last 3 days of pregnancy. The expression of this mRNA is restricted; message is abundant in the uterus during late pregnancy, but it is not found in non-pregnant uterus or in a variety of adult or fetal tissues. The temporo-spatial expression of this pregnant uterus specific mRNA and the consolidation in the predicted protein of two motifs implicated in early pregnancy events suggests that the product of the gene represented by this mRNA may play an important role in events that transpire during late pregnancy.

INTRODUCTION

To identify molecular tools that will facilitate the investigation of events surrounding the induction of parturition, a differential screen was conducted to isolate selectively expressed mRNAs using a cDNA library constructed from late pregnancy mouse uterus. The reported cDNA was selected for evaluation because the expression of this mRNA decreased during late pregnancy, suggesting that the product of the gene represented by this mRNA may possess an inhibitory influence on the induction of parturition.

Conceptual translation of the cDNA predicts a new and distinct protein that incorporates two recently defined motifs; the CUB motif (complement subcomponents C1r/C1s, Uegf protein and bone morphogenetic protein) [1] and the ZP motif (zona pellucida) [2]. The CUB motif has been associated with developmental phenomena [1] and is found in several developmentally associated proteins to include the neuronal A5 antigen [3], the spermadhesins [4–8] and two developmentally linked members of the transforming-growth-factor beta superfamily; tolloid and bone morphogenic protein [9,10]. However, it is also found in many other proteins not currently associated with development such as the tumour necrosis factor inducible protein TSG-6 (tumor necrosis factor stimulated gene-6) [11] or members of the complement family (C1s and C1r) [12–14]. The motif comprises 120 amino acids with four conserved cysteines; its function is currently unknown. The ZP motif is a peptide of 260 amino acids characteristically positioned near a carboxyl membrane-spanning region that includes eight conserved cysteines as well as invariant hydrophobic or aromatic amino acids suggesting a role in binding functions [2]. It is found in a heterogeneous group of proteins that includes the type III transforming-growth-factor beta receptor; betaglycan [15,16]. Binding studies conducted using a prokaryotic expression vector containing the betaglycan ZP domain confirm TGFβ binding to this motif [17]. Although these data were not corroborated by constructs employing a eukaryotic expression vector [18], the disparity may simply reflect differences resulting from post-translational modification.

These two motifs are present in two separate families of proteins implicated in events surrounding egg-sperm recognition; the spermadhesins and the ZP protein(s) [19,20]. The presence of two motifs implicated with early pregnancy events and the temporal and tissue restricted expression of this mRNA in the pregnant uterus suggest that this new protein may possess an important role in uterine events in late pregnancy.

EXPERIMENTAL

Materials

The study was approved by an animal use committee and conforms with established guidelines for the use of laboratory animals. The cDNA library was constructed from mouse uterus collected during the last 2 days of pregnancy (strain CF-1, birth = 20 days from mating) [21]. RNA used in Northern hybridizations was purified from reliably dated tissue (strain CD-1) by the acid guanidine technique [22]. A cDNA previously isolated and identified by this laboratory consisting of 800 bp of 3′ end of the mouse ubiquitin gene was used as a control for Northern blots. Probes for the differential screen were prepared using the same oligo(dT)-selected RNA employed in construction of the library and comparable RNA collected 3 days before birth [23]. A 1054 bp *Bam*HI fragment from the cDNA was used for probe preparation in Northern hybridizations as well as in a second screen of the library to isolate additional full-length cDNAs. Transcript size was determined using an RNA ladder (cat. no. 5620SA, Bethesda Research Laboratories, Gaithersburg, MD, U.S.A.).

Methods

Total RNA was purified from uterus as well as an assortment of adult and fetal tissues (strain CD-1); 20 μg aliquots were separated on 1% agarose gels and transferred to nylon membranes for use in Northern blotting. The quantity, integrity and

Abbreviations used: CUB, complement subcomponents; C1r/C1s, Uegf protein and bone morphogenetic protein; ZP, zona pellucida; TSG-G, tumour necrosis factor-stimulated gene-6.

efficacy of transfer of RNA was confirmed by viewing both the ethidium bromide-stained gels and membranes under UV light. [α-^{32}P]dCTP labelled cDNA probes were prepared by random priming (USB, Cleveland, OH U.S.A.) using standard protocols [24]. Northern hybridizations were performed under high stringency conditions (hybridization at 68 °C and final wash = 0.1 × SSC and 0.5 % SDS at 68 °C) and the blots exposed to Kodak XAR film for 2 days at −70 °C. One clone was completely sequenced (Sequenase, USB) in both directions using custom primers constructed at successive 250–300 bp intervals; two other full-length cDNAs were also sequenced extensively. The nucleotide sequence was conceptually translated (pc gene, Intelligenetics, San Jose, CA) and resultant nucleotide, as well as predicted amino acid sequence, compared to non-redundant nucleotide and protein databases in GenBank [25,26]. Linear full-length cDNA in Bluescript II KS (BSUTCZPfl, restriction digest with *Xho*I) as well as a truncated cDNA (BSUTCZPbgl, digest with *Bgl*II) were subjected to coupled *in vitro* transcription and translation with rabbit reticulocyte lysate in the presence of [^{35}S]methionine using a commercially available kit (TnT, Promega, Madison, WI, U.S.A.). The products were electrophoresed on an SDS/PAGE gel and the dried gel was exposed for 2 days to Kodak XAR film.

RESULTS

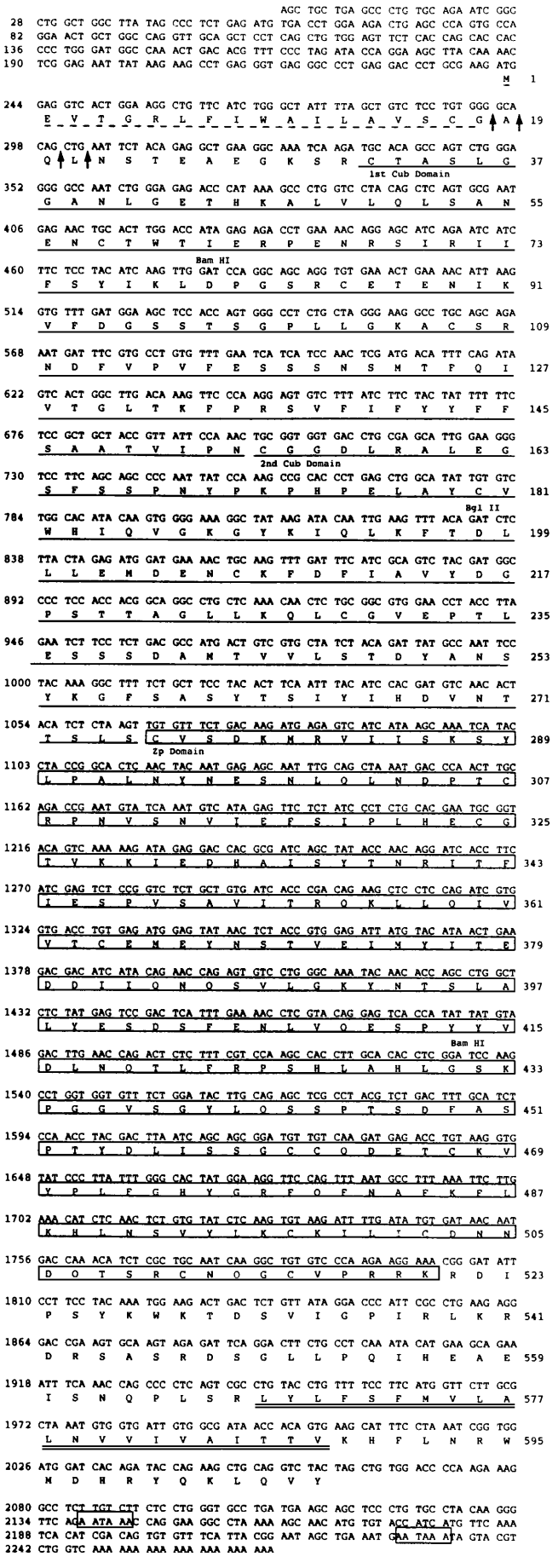
Comparison of the nucleotide sequence against the non-redundant nucleotide database of GenBank established the cDNA as novel. The cDNA (GenBank accession U69699) contains only one long open reading frame that conceptually translates a 606 amino acid protein with a predicted molecular weight of 67956 Da (Figure 1). Translation is presumed to start at the methionine-labelled amino acid 1. Consistent with the presence of a signal peptide, a majority of the first 20 amino acids are hydrophobic and potential cleavage sites are predicted between amino acids 18–19, 19–20, 20–21 and 21–22 [27]. The presence of this putative signal peptide reinforces the proposed location of the translational start site.

The protein was evaluated against the non-redundant protein database of GenBank [26]. Within the first 275 amino acids of the protein, two regions were identified that consistently displayed similarity to what was subsequently identified as the CUB motif. Analysis of these regions in the context of published CUB domains confirms that the protein contains two contiguous CUB domains; the first begins at Cys-32 and the second at Cys-154 (Figure 1).

At Cys-276, a region was identified that is similar to TGF betaglycan, uromodulin [28] and members of the ZP protein family; proteins conspicuous for the presence of a ZP motif [2]. Comparison of this region to published ZP domains confirms that this gene also contains a ZP domain (Figure 1).

The remainder of the protein displayed weak similarity to the protein ebnerin [29]. Analogous to ebnerin, there is a putative transmembrane domain present near the carboxyl end [30]. The presence of this putative transmembrane domain suggests that this protein may also function as an integral membrane-associated protein analogous to ebnerin or TGF betaglycan. However definitive evidence for this speculation requires confirmation from additional experiments. The protein terminates in a short 18 amino acid polypeptide presumably positioned within the cytoplasm.

To determine the pattern of expression of this gene in the uterus and at other potential sites in the mouse, Northern hybridizations were conducted against total RNA purified from late gestation uterus as well as a variety of other tissues. Message



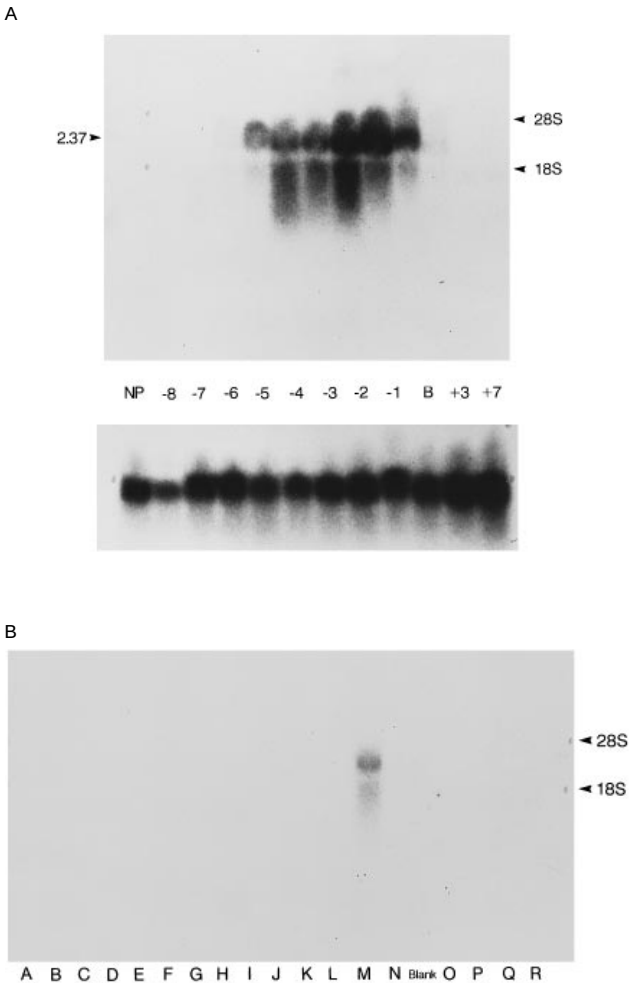


Figure 2 Northern hybridization

Northern hybridization using 20 μg (pooled from three mice) of total RNA obtained from various mouse tissues. (**A**) Data against RNA purified from mouse uterus collected during pregnancy. Days prior to (−) and following (+) birth (**B**) are indicated by numbers and NP = non-pregnant. Upper panel shows hybridization to probes from UTCZP and lower panel shows hybridization to control probes (ubiquitin). (**B**) Data from hybridization against RNA purified from a variety of adult and fetal mouse tissue. Lanes: A, brain; B, fetal brain; C, thymus; D, heart; E, lung; F, spleen; G, pregnant maternal liver; H, fetal liver; I, liver; J, pregnant maternal kidney; K, kidney; L, ovary; M, pregnant uterus; N, blank lane; O, placenta; P, skin; Q, fetal skin; R, gut.

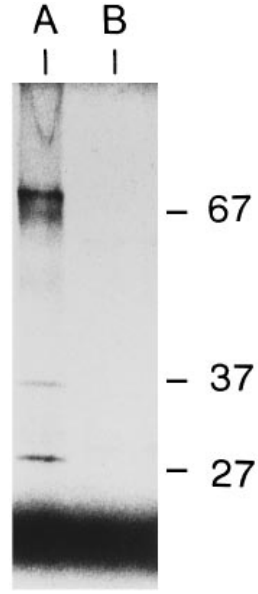


Figure 3 *In vitro* transcription and translation

Products of coupled *in vitro* transcription and translation from the plasmids BSUTCZPfl and BSUTCZPbgl are represented in lanes A and B. Three bands are prominent in lane A (BSUTCZPfl) located near the 67, 37 and 27 kDa markers representing transcription/translation from the full-length cDNA and transcription/translation from methionines carboxyl to the start site. As expected, all three bands are absent in products derived from the truncated BSUTCZPbgl plasmid (lane B).

first appears in the uterus at 6 days before birth (Figure 2A). The amount of mRNA (as reflected by signal strength) increases daily to reach a maximum at 3 days before birth. The level of mRNA than decreases on each subsequent day and by the first day following birth, it is almost undetectable. Signal is not detected in RNA collected from the uterus at 7 days before birth or in the absence of pregnancy to include days 3 and 7 following birth. Message was not found in the placenta or in any of the other fetal or adult tissues analysed (Figure 2B). The expression of this gene appears to be temporal and restricted to the gravid uterus. The RNA size marker on the Northern blot confirms the length of the cDNA. The Northern blot is also remarkable for the presence of prominent smears below the major band of hybridization. This does not represent degradation of the RNA as shown by the ubiquitin control; speculatively these smears may reflect rapid turnover of the message for this gene.

To confirm the presence of a functional open reading frame and the putative start site (methionine labelled 1), plasmids BSUTCZPfl and BSUTCZPbgl were subjected to coupled *in vitro* transcription and translation analysis. Three major bands near the 67, 37 and 27 kDa markers are evident (Figure 3) as well as a minor band in the vicinity of the 67 kDa band. The major 67 kDa band is consistent with translation from the putative start site and is compatible with the predicted molecular weight of the protein. The 37 and 27 kDa bands presumably represent translation from other methionines carboxyl to the start site. The minor band observed near the 67 kDa band conceivably represents transcriptional infidelity or post-translational modification of the protein. Translation from the truncated plasmid pBSUTCZPbgl results in loss of all bands as expected. The data confirm translation from the methionine labelled as 1.

Figure 1 Nucleotide and predicted amino acid sequence of UTCZP

Nucleotides are numbered on the left and amino acids are numbered on the right. The putative signal peptide is shown as a dashed underline (amino acid 1-M E V T . . V S C G-18). The potential cleavage sites for the signal peptide are indicated by the inverted arrows. The CUB domains are shown as a labelled single underline (first CUB domain amino acid 32-C T A S . . V I P N-153 and the second CUB domain amino acid 154-C G G D . . T S L S-275). The ZP domain is shown as a labelled boxed region (amino acid 276-C V S D . . P R R K-520). The putative transmembrane domain is shown by a double underline (amino acid 568-L Y L F . . I T T V-588). Boxed regions in the 3′ untranslated region indicate potential polyadenylation signals. *Bgl*II (nucleotide 831) and *Bam*HI (nucleotides 479 and 1533) restriction sites are as indicated.

DISCUSSION

In the absence of functional data, we have designated this cDNA as UTCZP (uterine cub motif zona pellucida motif). The mRNA encoded by this gene is relatively abundant within the uterus

during late pregnancy and expression is temporal and tissue restricted. The conceptually translated protein contains two motifs; the CUB motif and the ZP motif. Although these motifs are found in a variety of heterogeneous proteins, the presence of motifs that are prominent features of two distinct protein families implicated in egg-sperm recognition in a late pregnancy uterine-specific protein is intriguing. Collectively, the data suggest that this protein possesses an important role in late pregnancy uterine events.

The predicted protein is structurally similar to ebnerin; an integral membrane associated protein of unknown function that is found in von Ebner's glands [29] and ductin, a mouse protein found in the intestine, liver and pancreas [31]. Both of these proteins contain CUB and ZP motifs, however the arrangement of these domains in these proteins differs from the arrangement present in UTCZP. Analogous to these proteins, UTCZP has a putative transmembrane domain that is followed by a short cytoplasmic tail. Speculatively, TTCZP may also be an integral membrane-associated protein. To determine if a human counterpart exists, a search was made of the expressed sequence tag (EST) database (dbest) using peptide sequence derived from this protein. Several ESTs originating from human pancreas were found that exhibit a high degree of similarity to this protein. It is uncertain whether these ESTs represent the human counterpart to this protein or to ductin at present, but the peptides predicted from short open reading frames contained in these ESTs display more similarity to UTCZP than to ductin, suggesting that they may represent the human counterpart to UTCZP.

Bovine acidic seminal fluid protein is a member of the sperm-adhesin family that essentially consists of a single CUB domain [32–35]. *In vitro*, this protein functions as a mitogen, growth factor and it stimulates progesterone secretion in cultured ovarian cells [32]. If the CUB motifs found in UTZCP can be demonstrated experimentally to exhibit similar function, the presence of this protein in the uterus during late pregnancy has important implications in understanding the induction of parturition.

I would like to thank Dr H. J. Kung of CWRU for performing coupled *in vitro* transcription and translation.

REFERENCES

1 Bork, P. and Beckman, G. (1993) J. Mol. Biol. **231**, 539–545
2 Bork, P. and Sander, C. (1992) FEBS Lett. **300**, 237–240
3 Takagi, S., Hirata, T., Agata, K., Mochil, M., Eguchi, G. and Fujisawa, H. (1991) Neuron **7**, 295–307
4 Calvete, J. J., Solis, D., Sanz, L., Diaz-Maurino, T., Shafer, W., Mann, K. and Topfer-Petersen, E. (1993) Eur. J. Biochem. **218**, 719–725
5 Kwok, S. C., Yang, D., Dai, G., Soares, M. J., Chen, S. and McMurtry, J. P. (1993) DNA Cell Biol. **12**, 605–610
6 Sanz, L., Calvete, J. J., Mann, K., Gabius, H. J. and Topfer-Petersen, E. (1993) Mol. Reprod. Devel. **35**, 37–43
7 Calvete, J. J., Solis, D., Sanz, L., Diaz-Maurino, T. and Topfer-Petersen, E. (1994) Biol. Chem. Hoppe-Seyler **375**, 667–673
8 Dostalova, Z., Calvete, J. J., Sanz, L. and Topfer-Petersen, E. (1995) Eur. J. Biochem. **230**, 329–336
9 Shimell, M., Ferguson, E., Childs, S. and O'Connor, M. (1991) Cell **67**, 469–481
10 Wozney, J., Rosen, V., Celeste, A., Mitsock, L., Whitters, M., Kriz, R., Hewick, R. and Wang, E. (1988) Science **242**, 1528–1534
11 Lee, T., Wisniewski, H. and Vilcek, J. (1992) J. Cell Biol. **166**, 545–557
12 Tosi, M., Journet, A., Duponchel, C., Couture-Tosi, E. and Meo, T. (1989) Behring Institute Mitteilungen **84**, 65–71
13 Fothergill, J., Kemp, G., Paton, N., Carter, P. and Gray, P. (1989) Behring Institute Mitteilungen **84**, 72–79
14 Arlaud, G. J., Thielens, N. M. and Aude, C. (1989) Behring Institute Mitteilungen **84**, 56–64
15 Lopez-Casillas, F., Cheifetz, S., Doody, J., Andres, J., Lane, W. and Massague, J. (1991) Cell **67**, 785–795
16 Massague, J., Andres, J., Attisano, S., Cheifetz, S., Lopez-Casillas, F., Ohtsuki, M. and Wrana, J. (1992) Mol. Reprod. Dev. **32**, 99–104
17 Fukushima, D., Butzow, R., Hildebrand, A. and Ruoslahti, E. (1993) J. Biol. Chem. **268**, 22710–22715
18 Lopez-Casillas, F., Payne, H., Andres, J. and Massague, J. (1994) J. Cell Biol. **124**, 557–568
19 Dean, J. (1992) J. Clin. Invest. **89**, 1055–1059
20 Wassarman, P. (1990) Development **108**, 1–17
21 Kasik, J. and Rice, E. (1993) Mol. Cell. Endocrinol. **90**, 239–242
22 Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. **162**, 156–159
23 Sargent, T. (1987) Methods Enzymol. **152**, 423–432
24 Sambrook, J., Fritsch, E. and Maniatis, T. (1989) in Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
25 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) J. Mol. Biol. **215**, 403–410
26 Gish, W. and States, D. (1993) Nature Genetics **3**, 266–272
27 von Heijne, G. (1984) J. Mol. Biol. **173**, 243–251
28 Hession, C., Decker, J., Sherblom, A., Kumar, S., Yue, C., Mattaliano, R., Tizard, R., Kawashima, E., Schmeissner, U., Heletky, S. et al. (1987) Science **237**, 1479–1484
29 Li, X. and Snyder, S. H. (1995) J. Biol. Chem. **270**, 17674–17679
30 Rao, M. and Argos, P. (1986) Biochim. Biophys. Acta **869**, 197–214
31 Cheng, H., Bjerknes, M. and Chen, H. (1996) Anat. Rec. **244**, 327–343
32 Einspanier, R., Einspanier, A., Wempe, F. and Scheit, K. H. (1991) Biochem. Biophys. Res. Comm. **179**, 1006–1010
33 Wempe, F., Einspanier, R. and Scheit, K. (1992) Biochem. Biophys. Res. Commun. **183**, 232–237
34 Einspanier, R., Amselgruber, W., Sinowatz, F., Henle, T., Ropke, R. and Schams, D. (1993) J. Reprod. Fertil. **98**, 241–244
35 Einspanier, R., Krause, I., Calvete, J. J., Tofper-Petersen, E., Klostermeyer, H. and Karg, H. (1994) FEBS Lett. **344**, 61–64

Received 20 March 1997/3 November 1997; accepted 11 November 1997