The characterization and localization of the mouse thymopoietin/lamina-associated polypeptide 2 gene and its alternatively spliced products.

R Berger, L Theodor, J Shoham, et al.

Genome Res. 1996 6: 361-370
Access the most recent version at doi:10.1101/gr.6.5.361

RESEARCH

The Characterization and Localization of the Mouse Thymopoietin/Lamina-associated Polypeptide 2 Gene and its Alternatively Spliced Products

Raanan Berger,[1,4] Livia Theodor,[1,4] Jacob Shoham,[2] Ellen Gokkel,[1] Frida Brok-Simoni,[1] Karen B. Avraham,[3] Neal G. Copeland,[3] Nancy A. Jenkins,[3] Gideon Rechavi,[1,5] and Amos J. Simon[1]

[1]Institute of Hematology, The Chaim Sheba Medical Center, Tel-Hashomer, affiliated with the Sackler School of Medicine, Tel-Aviv University, Israel; [2]Life Sciences Faculty, Bar-Ilan University, Ramat-Gan, Israel; [3]Mammalian Genetics Laboratory, ABL–Basic Research Program, National Cancer Institute–Frederick Cancer Research and Development Center, Frederick, Maryland 21702.

Thymopoietins (Tmpos) are a group of ubiquitously expressed nuclear proteins, with sequence homology to lamina-associated polypeptide 2 (LAP2). Here we report the isolation and characterization of seven mouse *Tmpo* mRNA transcripts named *Tmpo* α, β, β′, γ, ϵ, δ, and ζ. The α, β, and γ *Tmpo* cDNA clones are the mouse homologs of the previously characterized human α, β, and γ TMPOs, respectively, whereas *Tmpo* ϵ, δ, and ζ are novel cDNAs. Additionally, the mouse *Tmpo* gene was cloned and characterized. It is a single-copy gene organized in 10 exons spanning ~22 kb, which encodes all of the described *Tmpo* cDNA sequences, located in the central region of mouse chromosome 10. The almost identical genomic organization between the human and mouse genes, and the novel alternatively spliced mouse transcripts, led us to reanalyze the human TMPO gene. The human β-specific domain was found to be encoded by 3 exons designated 6a, 6b, and 6c and not by a single exon as described previously. These findings suggest that there may be more human transcripts than currently recognized. The possible involvement of the new growing family of Tmpo proteins in nuclear architecture and cell cycle control is discussed.

Thymopoietin (Tmpo) was originally isolated from bovine thymic extracts (Goldstein 1974) as a 49-amino-acid polypeptide (Schlesinger and Goldstein 1975). The immunomodulating effects attributed to Tmpo and its putative active domain thymopentin (amino acids 32–36, Arg-Lys-Asp-Val-Tyr) led to clinical trials using thymopentin as a drug in several diseases such as rheumatoid arthritis (Kantharia et al. 1989) and human immunodeficiency virus infection (Conant et al. 1992). Characterization of the amino acid sequence of the polypeptide from a variety of tissues (Audhya et al. 1981; Audhya and Goldstein 1988) revealed several amino acid substitutions, which suggested a tissue-specific expression pattern of various isoforms.

Isolation of a bovine thymopoietin cDNA (Zevin-Sonkin et al. 1992) and subsequently human cDNAs, encoding three related but distinct TMPOs (TMPO α, β, and γ) (Harris et al. 1994), expanded our knowledge about Tmpo. Human TMPO α, β, and γ share an identical amino-terminal domain of 187 amino acids, which is followed in TMPOα by a specific domain (506 amino acids). TMPOs β and γ are closely related structurally, with TMPOβ differing from TMPOγ only by the insertion of a β-specific domain of 109 amino acids after amino acid 220.

Recently, a single human TMPO gene was isolated and characterized (Harris et al. 1995). The gene spans over ~35 kb of genomic DNA, containing 8 exons that encode the three TMPO-spliced mRNAs. TMPOβ was found to be the human homolog of the rat lamina-associated poly-

[4]These authors contributed equally to this work.
[5]Corresponding author.
E-MAIL wsamos@dapsas1.weizmann.ac.il; FAX 972-3-530-3506.

peptide 2 (LAP2) (Furukawa et al. 1995; Harris et al. 1995), an integral protein of the inner nuclear membrane.

In this study we report the isolation and molecular characterization of seven distinct mouse *Tmpo* (locus designation) cDNAs that encode for six putative mouse Tmpo proteins. In addition, the genomic structure and chromosomal localization of the mouse *Tmpo* gene is elucidated. The differences in genomic organization between the human and mouse genes, and the novel alternatively spliced mouse transcripts, led us to reanalyze the human TMPO gene.

RESULTS

Analysis of the Mouse *Tmpo* α, β, β′, ϵ, δ, γ, and ζ Sequences

The mouse *Tmpo*ζ cDNA clone was isolated from thymus cDNA library and characterized, using a 126-bp fragment, encoding Tmpo amino acids 1–42, from the bovine clone cDNA 113 (Zevin-Sonkin et al. 1992), as a probe. The same library was subsequently screened with a probe derived from the amino-terminal 790 bp of the *Tmpo*ζ cDNA. One hundred fifty-five positive clones were obtained. Repetitive screenings and restriction enzyme analysis revealed at least seven distinct *Tmpo* transcripts. A representative clone from each of them was chosen for further sequence analysis. The nucleotide and predicted amino acid sequences of mouse *Tmpo* α, β, β′, ϵ, δ, γ, and ζ are shown in Figure 1. Examination of mouse *Tmpo*α sequence (Fig. 1A) reveals a short region of basic amino acids (amino acids 188–194) suggestive of a nuclear localization domain (Kalderon et al. 1984), and a possible tyrosine phosphorylation site (amino acids 618–625) (Patschinsky et al. 1982). The sequence S/T-P-X-X, a potential recognition sequence for cdc2-related kinases (Nigg 1993) is found 10 times throughout the *Tmpo*α sequence. Figure 1B presents the β, β′, ϵ, δ, and γ *Tmpo* sequences. The *Tmpo* β and β′ clones are identical in their open reading frame (ORF) sequence but differ in an additional 3′-untranslated region (3′ UTR) sequence starting in A^{1715} of the β′ clone, probably because of an alternative polyadenylation signal. Hydropathy analysis revealed that similar to the human isoforms, all of the mouse *Tmpo* clones lack an amino-terminal hydrophobic signal sequence typical of secreted polypeptides (not shown). However, this analysis revealed that like the human TMPO β and γ, *Tmpo* β, ϵ, δ, and γ contain a possible hydrophobic transmembrane domain near their carboxyl termini (amino acids 409–432 of *Tmpo*β), suggesting a possible association of these proteins with cellular membranes. The sequence S/T-P-X-X is found seven times in the *Tmpo*β sequence. Two of these, T-P-R-K at residues 255–258 and T-P-K-K at residues 319–322, are especially likely to be cdc2 kinase substrates because of the basic residues in the X positions (Nigg 1993). *Tmpo*ζ is identical to *Tmpo* β, ϵ, δ, and γ in its amino-terminal domain through Gln^{219}. Its ORF, however, stops 5 amino acids downstream of this point followed by a distinct 1619-bp 3′ UTR domain (Fig. 1C).

Figure 2 demonstrates a schematic presentation of the different mouse cDNA clones. Mouse *Tmpo* α, β, δ, ϵ, γ, and ζ share an identical amino-terminal 186-amino-acid domain. Like the human and the bovine Tmpo isoforms, amino acids 1–49 are highly homologous to the originally purified 49-amino-acid bovine Tmpo (Schlesinger and Goldstein 1975). After Glu^{186}, *Tmpo*α diverges from the other *Tmpo*. *Tmpo*ϵ differs from *Tmpo*β only because it lacks the β-specific residues 220–259. *Tmpo*δ lacks both the β and the ϵ/β-specific domains contained within residues 220–291 of *Tmpo*β. *Tmpo*γ is missing the β, ϵ/β, and δ/ϵ/β-specific domains contained within amino acids 220–328 of *Tmpo*β. RT–PCR analysis of mouse thymus total RNA confirmed the presence of the novel *Tmpo* mRNA transcripts (not shown). Mouse *Tmpo* α, β, and γ ORF sequences are 78%, 90%, and 91% identical to the previously published human TMPO α, β, and γ sequences (Harris et al. 1994), respectively, suggesting interspecies conserved functions.

Genomic Organization of the *Tmpo* Gene

Overlapping clones covering the entire *Tmpo* gene were isolated from a BALB/c liver λ genomic library (Fig. 3). Restriction mapping and partial sequencing of the clones suggested that *Tmpo* α, β, ϵ, δ, γ, and ζ are produced via alternative mRNA splicing from a single gene. DNA sequence of the relevant regions was obtained by sequencing of genomic subclones and by using internal primers. This made it possible to define the precise location of all the exons, the sequences of all the intron–exon junctions, and the 5′-flanking region of the *Tmpo* gene. Figure 3 schematically presents the organization of the *Tmpo* gene and the overlapping genomic clones used for mapping and sequencing. The gene contains 10 exons spanning ~22 kb genomic DNA.

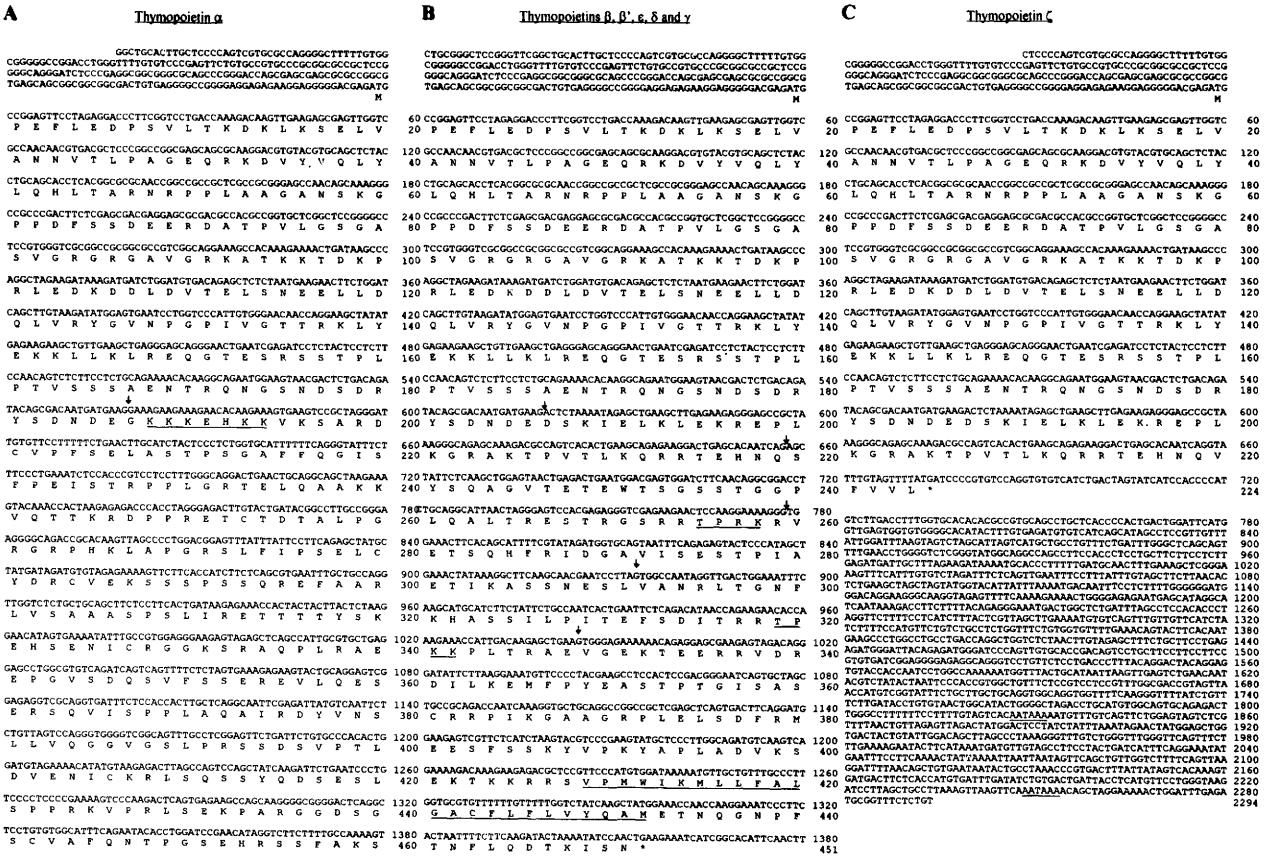

Figure 1 Nucleotide and predicted amino acid sequences of mouse *Tmpo*α (*A*), *Tmpo* β, β′, ϵ, δ, and γ (*B*), and *Tmpo*ζ (*C*). The sequences are numbered so that amino acid +1 is the amino-terminal proline of mature *Tmpo* and nucleotide +1 is the first nucleotide of the proline codon. The predicted amino acid (single-letter code) is depicted under the middle nucleotide of the corresponding codon. (*A*) The arrow (G^{187}) indicates the beinning of the unique α-domain. The two underlined sequences are a short region of basic amino acids (residues 188–194) suggestive of a nuclear localization domain and a consensus sequence for tyrosine phosphorylation. (*B*) Nucleotide and amino acid numbers are for *Tmpo*β. Arrows above amino acids D^{187}, S^{220}, V^{260}, V^{292}, and V^{329} indicate the beginning of β/ϵ/δ/γ/ζ, β, β/ϵ, β/ϵ/δ, and β/ϵ/δ/γ domains, respectively. The underlined sequences are the cdc2 kinase consensus sequences (residues 255–258 and 319–322), a hydrophobic domain in *Tmpo* β, ϵ, δ, and γ (residues 409–432), and one AATAAA (nucleotides 1655–1660) polyadenylation site. The sequence downstream to G^{1716} is unique to *Tmpo*β′. (*C*) The two possible polyadenylation sites are underlined. Nucleotide sequences for the mouse *Tmpo* mRNA have been deposited with the GenBank data base under accession nos. U39073 (Tmpoζ), U39074 (Tmpoβ), U39075 (Tmpoϵ), U39076 (Tmpoδ), U39077 (Tmpoγ), and U39078 (Tmpoα).

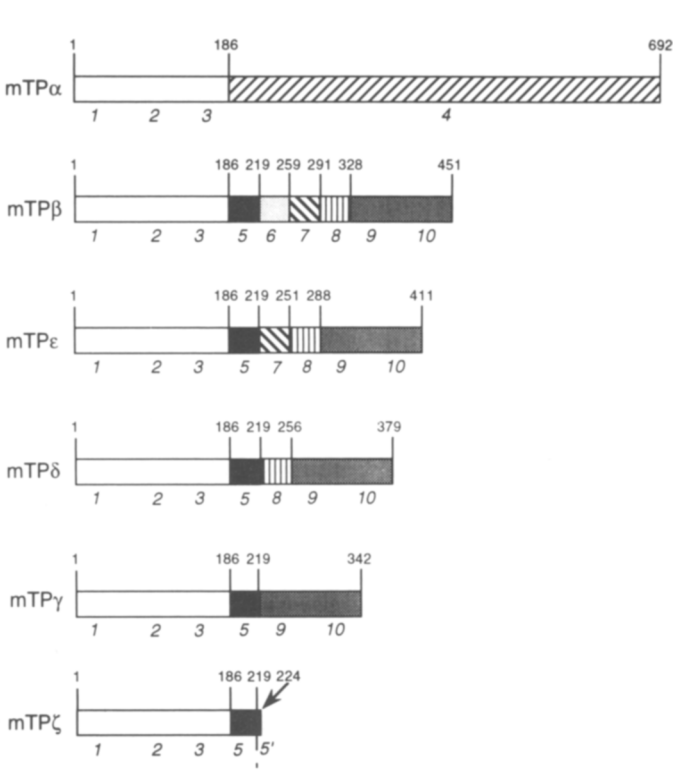

Figure 2 Schematic diagram of the various mouse *Tmpo* transcripts. The numbers above the bars indicate amino acid position, and the numbers below indicate the coding exons. The open bar (1–186) depicts the identical amino-terminal domain. The hatched bar (186–692) in *Tmpo*α is the unique α-domain. The solid bar (189–219) is the β, ϵ, δ, γ, and ζ domain. The lightly stippled bar (219–259) is the β-specific domain. The hatched bar (259–291 of *Tmpo*β) is the ϵ/β-specific domain. The vertically striped bar (291–328 of *Tmpo*β) is the δ/ϵ/β-specific domain, and the heavily stippled bar (328–451 of *Tmpo*β) is the γ/δ/ϵ/β-specific domain. The arrow indicates the last 5 amino acids of *Tmpo*ζ.

Analysis of sequences of exon borders revealed that exon 1 encodes the 5′-untranslated region and amino acids 1–91, exon 2 encodes amino acids 92–133, and exon 3 encodes amino acids 134–186. These exons are spliced to form the common sequences present in all seven mouse *Tmpo* genes. The α-specific domain (residues 187–692) and its 3′ UTR are encoded by the large exon 4. Exon 5 encodes amino acids 187–219 of *Tmpo* β, ϵ, δ, γ, and ζ. Exon 5′ is the 3′ UTR of *Tmpo*ζ and extends downstream from exon 5 without an intron between them. Hence, the 5–5′ border region is functioning as a donor-splicing site for the alternatively spliced *Tmpo* β, ϵ, δ, and γ. Exon 6 is the β-specific domain and encodes amino acids 220–259 of *Tmpo*β. Exon 7 is the ϵ/β-specific domain encoding amino acids 260–291, and exon 8 is the δ/ϵ/β-specific domain encoding amino acids 292–328 of *Tmpo*β. Exons 6, 7, and 8 are relatively close together (Fig. 3), encoding amino acids 220–328, which are homologous to the human β-specific domain (residues 221–329) encoded by exon 6 (Harris et al. 1995). Exons 9 and 10 are γ/δ/ϵ/β-specific domains. Exon 9 encodes amino acids 329–358 of *Tmpo*β, and exon 10 encodes amino acids 359–451 of *Tmpo*β and further encodes the 3′ UTR of *Tmpo* β, β′, ϵ, δ, and γ.

Exon 6 of the Human TMPO Gene Contains Two Small Introns

The difference between mouse and human Tmpo gene organization regarding the β-specific domain (residues 220–328), coupled with the lack of the entire human exon 6 sequence in the GenBank data base (accession nos. U18269 and U18271), led us to characterize this region in the human TMPO gene. Human genomic DNA was amplified by PCR (see Methods). A ~750-bp PCR fragment was cloned and characterized. Figure 4 presents the genomic sequence of the human TMPO exon 6 region. This region contains two intronic sequences, dividing exon 6 into three smaller exons, termed 6a, 6b, and 6c. These exons are organized in the same pattern as exons 6, 7, and 8 of the mouse *Tmpo*, respectively. These data strongly suggest that like the mouse *Tmpo* gene, the human TMPO gene contains 10 exons.

Sequence Analysis of Exon/Intron Borders

All splice sites for the distinct mouse *Tmpo* mRNAs contain the canonical GT and AG dinucleotides at the intron borders (Table 1). The splice sites match consensus splice site sequences to varying extents. The mouse and the human 3′-splice sites share a significant homology (Table 2). The polypyrimidine tracts of the mouse exon 4, 6, 7, and 8 3′-splice sites are highly identical to the polypyrimidine tract of the human exon 4,

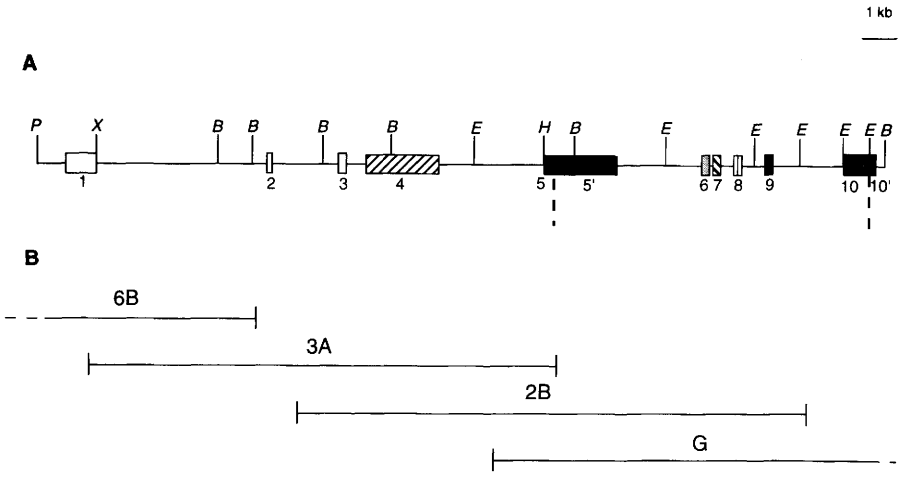

Figure 3 Physical map of the mouse *Tmpo* gene. (*A*) Exons are numbered and indicated as boxes. Introns are indicated by a thin line. Sites for restriction enzymes *Bam*HI (B), *Eco*RI (E), *Hin*dIII (H), *Pvu*II (P), and *Xho*I (X) are shown. The various depictions in the exon bars (e.g., open, solid, hatched, etc.) correspond to the encoded regions in Fig. 2. Exons 5′ and 10′ encode TPζ 3′ UTR and TPβ′ unique sequence, respectively. (*B*) Overlapping genomic clones used for mapping and sequencing.

6a, 6b, and 6c 3′-splice sites, respectively. Interestingly, unlike the other, less conserved 3′-splice sites, these 3′-splice sites are participants in alternative splicing events.

Analysis of the 5′-flanking Region of the *Tmpo* Gene

The TSSG program analysis (Prestridge 1995) of the 5′-flanking sequence of the *Tmpo* gene revealed a predicted transcription start site located 378 bp upstream of the translation initiation codon (Fig. 5). The mouse and the human Tmpo 5′-flanking regions share conserved promoter sequences around the predicted transcription start site. Several potential binding sites for known transcription factors were analyzed (Fig. 5). Like the human gene, no obvious TATAAA sequence, a known binding site for the general transcription factor TFIID–TBP, could be found in the usual position ~30 bp 5′ to the transcription start sites. This absence is characteristic of some other genes expressed in many tissues. However, a closely related sequence (TTTAAA) is present 29 bp upstream of the predicted transcription start site. Recognition sequences for the transcription factor Sp1 (Kadonaga et al. 1986) are found at positions + 65, + 27, −52, −83, −119 and −198. Two GCCAAT boxes, potential binding sites for members of the CTF/NF-1 family of transcription factors (Santoro et al. 1988), are present within direct repeat sequences in positions + 35 and + 68. Several potential binding sites for the Even-skipped (Eve) homeo box protein are present at positions −14, −78, −143, and −221.

Tmpoβ and LAP2 are Homologous Proteins

Data base comparison revealed that all human and mouse Tmpo cDNA sequences share a remarkable identity and homology with the nuclear LAP2, isolated from rat. Figure 6 shows the comparison between the predicted amino acid sequence of the mouse Tmpoβ, the LAP2 (GenBank accession no. U18314), and the human TMPOβ (accession no. U09087). The mouse Tmpoβ and the rat LAP2 are 96% identical (Fig.

```
661-AGCTATTCTCAAGCTGGAATAACTGAGACTGAATGGACAAGTGGATCTTCAAAAGGCGGA
    Exon 6a
    CCTCTGCAGGCATTAACTAGGGAATCTACAAGAGGGTCAAGAAGAACTCCAAGGAAAAGG-780/

    gtgatgcaaggcttattccttgggttttcagatttgtagggtttagtattatttatattt

    attgtttttgttttgtttcaaactaacag/781-GTGGAAACTTCAGAACATTTTCGTAT
                                    Exon 6b
    AGATGGTCCAGTAATTTCAGAGAGTACTCCCATAGCTGAAACTATAATGGCTTCAAGCAA

    CGAATCCTTA-872/gtaaatatgtttcgtaaactatacaagtggtattctttgtaaatt

    ----acccttaattggaaatcggggag... ~300 bp of intronic sequence ...

    taagtgtctgtgttatgtttggataattctgagtctgaataatttgaatcttggcag/873-

    GTTGTCAATAGGGTGACTGGAAATTTCAAGCATGCATCTCCTATTCTGCCAATCACTGAA
    Exon 6c
    TTCTCAGACATACCCAGAAGAGCACCAAAGAAACCATTGACAAGAGCTGAA-987
```

Figure 4 Nucleotide sequence of exon 6 region of the human TMPO gene. The exonic sequences are underlined and depicted as uppercase, designated 6a, 6b, and 6c. Nucleotides of the intronic sequences are in lowercase. The numbers that flank the exonic sequences correspond to the numbered base pairs of the human TMPOβ gene (Harris et al. 1994).

Table 1. Sequences of the exon/intron boundaries of the mouse *Tmpo* gene

5′ INTRON	EXON	3′ INTRON
	1 -GGCAGG	GTAAGGCGAACCCCCCCGGA
TAATTGTACTTTGTTTGCAG	AAAGCC- 2 -TTGTGG	GTAATGGGTTTTATTTGTTT
GAACTTCTCCCTTAACCCAG	GAACAA- 3 -ATGAAG	GTAACATTTTAACTGCTCTT
TGCCTCTTTTGCCTCTACAG	GAAAGA- 4 α-domain	
TTCCTCGATGTTATTTCCAG	ACTCTA- 5 -AATCAG	GTATTTGTAGTTTTATGATC
ATGTGTCGATGCTTGACTAG	AGCTAT- 6 -AAAAGG	GTGACGCAGGGCTTGCCGCT
TTCTGTTTTCAAACTAACAG	GTGGAA- 7 -TCCTTA	GTAAATATGTCTTATAATCT
GATAATTTGAATCTTGGCAG	GTGGCC- 8 -GCTGAA	GTAAATGGATACCATTTAGC
GTTTCTTTTTTTCTCACTAG	GTGGGA- 9 -AATCAG	GTACCTGGATGGATAAAACT
TTCTGTTTCACTTTGAACAG	TGCTAG-10	

(1) The consensus (AG) and GT) dinucleotides at the intron borders are in boldface type.
(2) The intronic sequence downstream of exon 5 is the amino-terminal sequence of exon 5′.

6), suggesting that Tmpoβ is the mouse homolog of the rat LAP2.

The *Tmpo* Gene is Located in the Central Region of Mouse Chromosome 10

The mouse chromosomal location of Tmpo (locus designation, *Tmpo*) was determined by interspecific backcross analysis using progeny derived from matings of (C57BL/6J × *Mus spretus*)F_1 × C57BL/6J mice. This interspecific backcross mapping panel has been typed for >2000 loci that are well distributed amoung all the autosomes as well as the X chromosome (Copeland and Jenkins 1991). C57BL/6J and *M. spretus* DNAs were digested with several enzymes and analyzed by Southern blot hybridization for informative restriction fragment length polymorphisms (RFLPs) using a probe derived from the mouse cDNA. The mapping results indicated that *Tmpo* is located in the central region of mouse chromosome 10 linked to insulin-like growth factor-1 (*Igf1*) and mast cell growth factor (*Mgf*; Fig. 7). Although 178 mice were analyzed for every marker and are shown in the segregation analysis (Fig. 7), up to 196 mice were typed for some pairs of markers. Each locus was analyzed in pair-wise combinations for recombination frequencies using the additional data. The ratios of the total number of mice exhibiting recombinant chromosomes to the total number of mice analyzed for each pair of loci and the recombination frequencies between the loci are shown in Figure 7.

Comparative gene mapping in mouse and human has revealed numerous regions of homology between the two species, and this is clearly demonstrated between the central to distal portion of mouse chromosome 10 and human chromosome 12. The human homologs of *Tmpo* and *Mgf* map to 12q22, and the human *Igf1* locus maps to 12q22–q23.

DISCUSSION

In this study the genomic structure and the various alternatively spliced transcripts of mouse *Tmpo* are reported. The mouse *Tmpo* gene, which gives rise to at least seven alternatively spliced mRNA transcripts, is encoded by a single gene containing 10 exons. The novel mouse *Tmpo* mRNA, together with the similar genomic organization between the mouse and the revised human TMPO gene, and their 3′-splice site homology (Table 2), suggest a possible existence of more human alternatively spliced *Tmpo* transcripts. This was confirmed by RT–PCR on human thymic mRNA (not shown). Moreover, Harris et al. (1994) demonstrated the cross-reaction of their anti-human TMPO antibodies with the 75, 51, and 39 kD, of α, β, and γ isoforms, respectively. However, these antibodies also cross-reacted with another band, of ~43–44 kD [Fig. 2 in Harris et al. (1994)], which we suggest to be the human TMPOδ gene.

The absolute and relative abundances of *Tmpo* α, β, and γ mRNAs appear to vary in different tissues and cell lines (Harris et al. 1994; Berger et al. 1995), suggesting that both expression and alternative splicing of *Tmpo* may be regulated in a tissue-specific manner. One possible mechanism for control of the formation of *Tmpo* α and

Table 2. Comparison between the mouse and the human 3′-splice sites

Mouse 3′-splice site	Human 3′-splice site
TAATTGT**ACTTTGTTT**GCAG /Exon 2	TTACTGG**ACTTTGTTT**ACAG /Exon 2
GAAC**TTCT**C**GCTTAA**CCCAG /Exon 3	CAAG**TTCT**G**GCTTAA**TCCAG /Exon 3
TGCCTCTTTTGCCTCTACAG /Exon 4	**TGCCTCTTTTGCCTCTACAG** /Exon 4
TTCCTCG**ATGTTATTTCGAG** /Exon 5	TTCTCCA**ATGTTATTTCGAG** /Exon 5
ATGTGTCGATGCTTGACTAG /Exon 6	**ATGTGTTGATGCTTGAATAG** /Exon 6a
TTCTGT**TTTCAAACTAACAG** /Exon 7	GTTTTG**TTTCAAACTAACAG** /Exon 6b
AATAATTTGAATCTTGGCAG /Exon 8	**AATAATTTGAATCTTGGCAG** /Exon 6c
GTTTCTTTTTTTCTCACTAG /Exon 9	GTTTGTCTGTTTCTTATTAG /Exon 7
TTCTG**TTTCACT**TTGAACAG /Exon 10	CCTCC**TTTCACT**CCCAACAG /Exon 8

Identical sequences between the murine and the human 3′-splice sites are in boldface type and underlined.

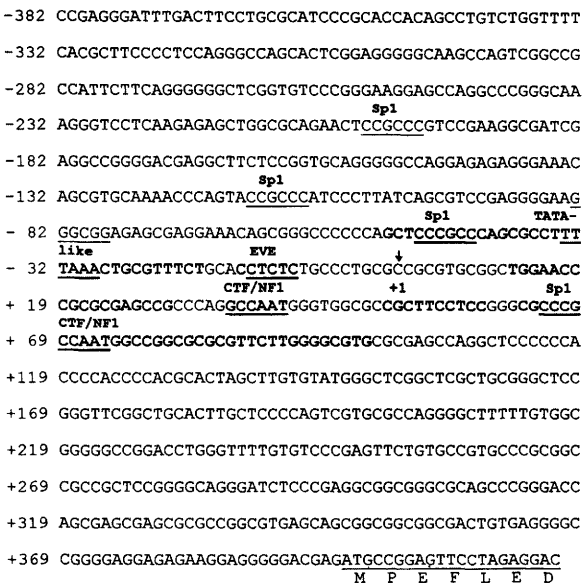

Figure 5 Sequence of the 5′ end of the mouse *Tmpo* gene. (↓) The transcription initiation site (nucleotide + 1). The conserved promoter sequences between mouse and human are in boldface type. Sequences identical to binding site sequences for known transcription factors are underlined, and the names of the transcription factor are given above. The sequence has been deposited with GenBank under accession no. U38185.

ζ mRNA is via control of its 3′-end cleavage and polyadenylation site before splicing occurs (McKeown 1992). This would eliminate all of the downstream exons encoding *Tmpo* β, ϵ, δ, and γ sequences and remove them as competitors for splicing of the exon 3 to the α-specific exon 4. Harris et al. (1995) demonstrated a conserved sequence downstream of the two alternative polyadenylation signals for human TMPOα mRNA, suggesting that this sequence may be a binding site for a factor that regulates TMPOα mRNA 3′-end formation, perhaps in a tissue-specific manner (Keller 1992). The mouse *Tmpo*ζ mRNA sequence (Fig. 1C) shows two further polyadenylation sites, but the regulation mechanism of the *Tmpo*ζ mRNA 3′-end formation has not yet been identified.

The 5′ end of the *Tmpo* gene is GC-rich, a characteristic feature of many genes with a wide range of tissue expression. Comparison of mouse and human Tmpo promoter sequences reveals several common conserved sequences, around the predicted transcription start site, in a similar order. These conserved sequences contain potential binding sites for known transcription factors, such as the Sp1, Eve, CTF/NF1, and TTTAAA domains. The functional significance of these motifs has not yet been studied; however, the significant evolutionary conservation further supports the essential role of Tmpo in diverse cellular functions.

The *Tmpo* gene is located in the central region of mouse chromosome 10. We have compared our chromosome 10 interspecific map with a composite mouse linkage map that reports the map location of many uncloned mouse mutations, provided from Mouse Genome Database (The Jackson Laboratory, Bar Harbor, ME). Although several mutations lie in the region of *Tmpo*, none have a phenotype consistent with what might be predicted for a mutation in *Tmpo*.

LAP2 is an integral membrane protein of the inner nuclear membrane, which binds directly to both lamin B1 and chromosomes in a mitotic phosphorylation-regulated manner (Foisner and Gerace 1993). The biochemical and physiological properties of LAP2 suggest an important role in nuclear envelope reassembly at the end of mitosis and/or anchoring of the nuclear lamina and in-

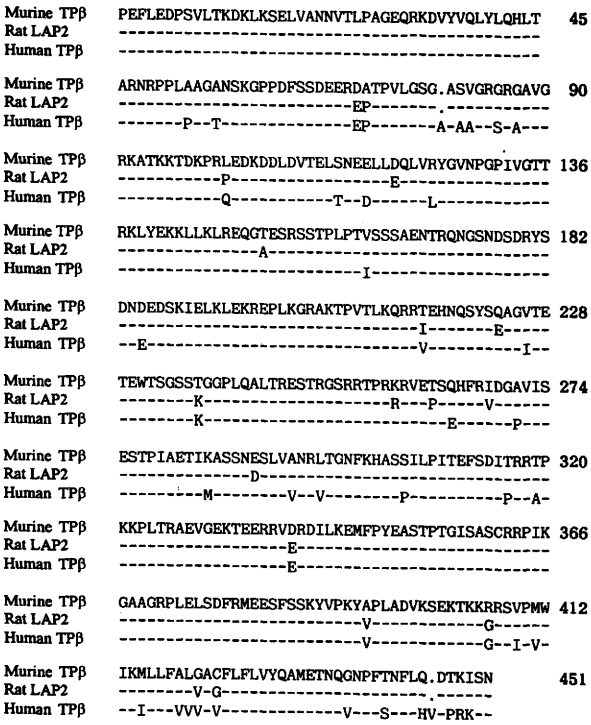

Figure 6 Comparison of the full-length predicted amino acid sequences of the mouse Tmpoβ, rat LAP2, and human TMPOβ. (Dashes represent identical amino acids). Numbers represent the mouse TPβ amino acids.

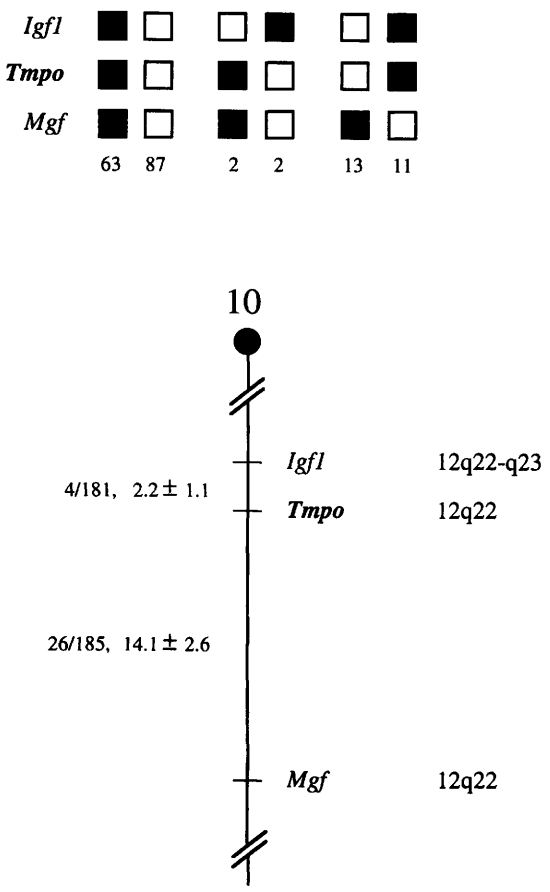

Figure 7 Chromosomal location of *Tmpo* in the mouse genome. The locus was mapped by interspecific backcross analysis. The segregation patterns of *Tmpo* and flanking genes are shown at the *top*. Each column represents the chromosome identified in the backcross progeny that was inherited from the (C57BL/6J × *M. spretus*) F_1 parent. The solid boxes represent the presence of a C57BL/6J allele; open boxes represent the presence of *M. spretus* allele. The number of offspring inheriting each type of chromosome is listed beneath each column. A partial chromosome 10 linkage map showing the location of *Tmpo* in relation to linked genes is shown at the *bottom*. The number of recombinant N_2 animals over the total number of N_2 animals typed plus the recombination frequencies, expressed as genetic distance (in cM) (±1 S. E.) is shown for each pair of loci (*left*). The positions of loci in human chromosomes are shown at *right*. References for the human map positions of loci cited in this study can be obtained from Genome Data Base (GDB), a computerized data base of human linkage information maintained by the William H. Welch Medical Library of The Johns Hopkins University (Baltimore, MD).

terphase chromosomes to the nuclear envelope. The amino acid sequence of the rat LAP2 is 96% identical to the mouse Tmpoβ and 91% identical to the human TMPOβ, indicating that they are structurally and functionally homologous proteins. Hence, Tmpo α, ϵ, δ, γ,and ζ are the LAP2 gene alternatively spliced forms. That assumption is supported further by Western blot analysis using LAP2-specific antibodies, which yielded bands at 53/43/41 kD (Konstantinov et al. 1995). The 53-kD polypeptide corresponds to the predicted molecular mass of the Tmpoβ/LAP2 protein. Based on the calculated molecular mass of the various Tmpo isoforms, we suggest that the 43/41-kD bands are the alternatively spliced δ and γ forms, respectively. Like Tmpoβ/LAP2, the ϵ, δ, and γ Tmpo contain a single putative membrane-spanning sequence (Fig. 1B) (Furakawa et al. 1995), which suggests that these alternatively spliced isoforms are three additional integral membrane proteins of the inner nuclear membrane.

By expressing deletion mutants of LAP2 in cultured cells, it was found that the smallest nucleoplasmic fragment of LAP2 that can specify binding to components associated with the nuclear envelope are residues 244–398 (Furakawa et al. 1995). However, mouse Tmpo ϵ, δ, and γ are identical to the Tmpoβ/LAP2 except that they lack residues 220–259, 220–291, and 221–328, respectively. Therefore, Tmpoβ/LAP2 and Tmpo ϵ, δ, and γ may have different binding potencies to the various components associated with the nuclear envelope.

Two of the most favored cdc2 kinase sites in LAP2 (Tmpoβ) are found at residues 256–259 and 320–323 (Furakawa et al. 1995). Phosphorylation of these sites by cdc2 kinase may be involved in modulating the interactions of LAP2 with chromatin and lamin during mitosis. However, Tmpo ϵ and δ lack the cdc2 kinase site positioned at residues 256–259, whereas Tmpo γ is missing both sites. These findings propose a possible role for the alternative splicing mechanism in nuclear events and cell cycle regulation.

METHODS

Isolation of Mouse *Tmpo* cDNA Clones

A cDNA clone, designated *Tmpoζ* was isolated from a mouse (B6/CBAFJ females, 6–8 weeks old) thymic cDNA library (commercially purchased from Strategene). The initial probe used was a 126-bp fragment encoding Tmpo amino acids 1–42 from the bovine cDNA clone 113 (Zevin-Sonkin et al. 1992). Hybridization was performed with 35% formamide, 5× SSPE (Sambrook et al. 1989) at 42°C, with the highest stringency wash in 0.1× SSPE at 50°C. In the subsequent round of screening, a probe derived from

the initial mouse clone was used. The hybridization was in 0.5 M $NaHPO_4$ (pH 7.2), 7% SDS, at 65°C, and the highest stringency wash was in 0.04 M $NaHPO_4$ (pH 7.2), 1% SDS, at 65°C. All sequences reported were determined on both strands by the Sanger technique using Sequenase version 2.0 kit (Amersham).

Isolation and Analysis of Genomic Clones

Four overlapping partial genomic clones were isolated from a library, prepared from a BALB/c liver DNA *Sau*3A-digested, and cloned into the EMBL 3A vector. The library was screened with the different cloned mouse *Tmpo* cDNAs. Clones were characterized initially by partial restriction mapping and by hybridization with defined regions of *Tmpo* cDNAs. Fragments of interest were subcloned into pBluescript II SK(+) (Stratagene) for DNA sequencing. The intron–exon boundaries were determined by direct comparison of the nucleotide sequences of *Tmpo* cDNAs clones and genomic sequences, using sense and antisense primers within the different cDNAs regions. Some of the sequencing primers were designed considering the alternatively spliced pattern. The sizes of introns were confirmed by restriction endonuclease digestion.

Isolation of the Human Exon 6 and its Intronic Sequences

PCR was performed using a DNA thermal cycler. The following human TMPOβ oligonucleotide primers were synthesized: 5′-AGCTATTCTCAAGCTGGAA-3′, sense nucleotides 661–680 and 5′-TTCAGCTCTTGTCAATGG-3′, antisense nucleotides 987–970 (Harris et al. 1994). A 50-μl reaction mixture containing 500–1000 ng of placental genomic human DNA in 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 1.5 mM $MgCl_2$, 0.001% gelatin, 250 mM (each) dNTPs: dATP, dCTP, dGTP, dTTP, 20 pmoles of each of the primers, and 1 unit of *Taq* DNA polymerase was subjected to 35 cycles of amplification. PCR conditions was follows: 1 min at 94°C, 1 min at 50°C, and 1.5 min at 72°C with a final elongation step at 72°C for 10 min. The human sequences were obtained using the ALFexpress automatic DNA sequencer (Pharmacia Biotech).

Sequence Analysis

Computer analysis of DNA and protein sequences were done using the Genetics Computer Groups (GCG) software package (Genetics computer group 1991). Sequences were analyzed for homology to nucleic acid and protein data bases using the Blast program (Altschul et al. 1990) of the National Center for Biotechnology Information via the Internet.

Analysis of the 5′-flanking region of the gene for candidate sequences similar to known binding sites for transcriptional regulatory proteins, and for the recognition of the start of transcription site, was done with the TSSG program (Prestridge 1995) using the TFD transcription factor data base (Ghosh 1991).

The sequences described here have been deposited in the GenBank data base (accession nos. U38185 and U39073–U39078).

Interspecific Mouse Backcross Mapping

Interspecific backcross progeny were generated by mating (C57BL/6J × *M. spretus*) F_1 females and C57BL/6J males as described (Copeland and Jenkins 1991). A total of 205 N_2 mice were used to map the *Tmpo* locus (see text for details). Southern blot analysis was performed as described (Jenkins et al. 1982). All blots were prepared with Hybond-N^+ membrane (Amersham). The *Tmpo* probe, a PCR-amplified fragment from the *Tmpo* mouse cDNA, was labeled with [α^{32}P]dCTP using a random priming labeling kit (Stratagene); washing was done to a final stringency of 0.5 × SSCP, 0.1% SDS at 65°C. A fragment of 4.8 kb was detected in *Xba*I-digested C57BL/6J DNA and a fragment of 4.0 kb was detected in *Xba*I-digested *M. spretus* DNA. The presence or absence of the 4.0-kb *M. spretus*-specific fragment was followed in the backcross mice.

A description of the probes and RFLPs for the loci linked to *Tmpo*, including insulin-like growth factor-1 (*Igf1*) and mast cell growth factor (*Mgf*) has been reported previously (Copeland et al. 1990). Recombination distances were calculated as described (Green 1981) using the computer program SPRETUS MADNESS. Gene order was determined by minimizing the number of recombination events required to explain the allele distribution patterns.

ACKNOWLEDGMENTS

We thank Dr. Eitan Friedman for his helpful and critical comments regarding this work, and Biana Freits for excellent technical assistance. We thank D. Swing and D.J. Gilbert for excellent assistance. This research was supported, in part, by the National Cancer Institute, Department of Health and Human Services, under contract with ABL (N.G.C., N.A.J., and K.B.A.) and by a grant from the National Institutes of Health, National Research Service Award 5F32GM15909-02 from the National Institute of General Medical Sciences (K.B.A.). The work reported is part of the Ph.D. thesis to be submitted by R.B. to the Sackler Faculty of Medicine, Tel Aviv University. The sequence data described in this paper have been submitted to the GenBank data library under accession nos. U39073–U39078 and U38185.

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

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Audhya, T. and G. Goldstein. 1988. Amino acid sequence of thymopoietin isolated from skin. *Ann. N.Y. Acad. Sci.* **548:** 233–240.

Audhya, T., D.H. Schlesinger, and G. Goldstein. 1981. Complete amino acids sequences of bovine Thymopoietins I, II, and III: Closely homologous polypeptides. *Biochemistry* **20:** 6195–6200.

Berger, R., L. Theodor, F. Brok-Simoni, H. Ben-Bassat, L. Trakhtenbrot, J. Shoham, and G. Rechavi. 1995.

Demonstration of Thymopoietin transcripts in different hematopoietic cell lines. *Acta Haematol.* **93:** 62–66.

Conant, M.A., L.H. Calabrese, S.E. Thompson, B.J. Poiesz, S. Rasheed, R.L. Hirsch, L.A. Meyerson, A.B. Kremer, C.C. Wang, and G. Goldstein. 1992. Maintenance of $CD4^+$ cells by thymopentin in asymptomatic HIV-infected subjects: Results of a double-blind, placebo-controlled study. *AIDS* **6:** 1335–1339.

Copeland, N.G. and N.A. Jenkins. 1991. Development and applications of a molecular genetic linkage map of the mouse genome. *Trends Genet.* **7:** 113–118.

Copeland, N.G., D.J. Gilbert, B.C. Cho, P.J. Donovan, N.A. Jenkins, D. Cosman, D. Anderson, S.D. Lyman, and D.E. Williams. 1990. Mast cell growth factor maps near the *Steel* locus on mouse chromosome 10 and is structurally altered in a number of *Steel* alleles. *Cell* **63:** 175–183.

Foisner, R. and L. Gerace. 1993. Integral membrane proteins of the nuclear envelope interact with lamins and chromosomes, and binding is modulated by mitotic phosphorylation. *Cell* **73:** 1267–1279.

Furukawa, K., N. Pante, U. Aebi, and L. Gerace. 1995. Cloning of a cDNA for lamina-associated polypeptide 2 (LAP2) and identification of regions that specify targeting to the nuclear envelope. *EMBO J.* **14:** 1626–1636.

Genetics Computer Group. 1991. Program manual for the GCG package, version 7, April, 1991, Madison, WI.

Ghosh, D. 1991. New developments of a transcription factors data-base. *Trends Biochem. Sci.* **16:** 445–447.

Goldstein, G. 1974. Isolation of bovine thymin: A polypeptide hormone of the thymus. *Nature* **247:** 11–14.

Green, E.L. 1981. Linkage, recombination and mapping. In *Genetics and probability in animal breeding experiments* pp. 77–113. Oxford University Press, New York, NY.

Harris, C.A., P.J. Andryuk, S. Cline, H.K. Chan, A. Natarajan, J.J. Siekierka, and G. Goldstein. 1994. Three distinct human thymopoietins are derived from alternatively spliced mRNAs. *Proc. Natl. Acad. Sci.* **91:** 6283–6287.

Harris, C.A., P.J. Andryuk, S. Cline, S. Mathew, J.J. Siekierka, and G. Goldstein. 1995. Structure and mapping of the human Thymopoietin (Tmpo) gene and relationship of human Tmpo β to rat lamin-associated polypeptide 2. *Genomics* **28:** 198–205.

Jenkins, N.A., N.G. Copeland, B.A. Taylor, and B.K. Lee. 1982. Organization, distribution and stability of endogenous ecotropic murine leukemia virus DNA sequences in chromosomes of *Mus. musculus*. *J. Virol.* **43:** 26–36.

Kadonaga, J.T., K.A. Jones, and R. Tjian. 1986. Promoter-specific activation of RNA polymerase II transcription by Sp1. *Trends Biochem. Sci.* **11:** 20–23.

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

Kantharia, B.K., N.J. Goulding, N.D. Hall, J. Davies, P.J. Maddison, P.A. Bacon, M. Farr, J.A. Wojtulewski, K.M. Englehart, S.P. Liyanage, and N.L. Cox. 1989. Thymopentin (Tmpo-5) in the treatment of rheumatoid arthritis. *Br. J. Rheumatol.* **28:** 118–123.

Keller, W. 1992. The biochemistry of 3′-end cleavage and polyadenylation of messenger RNA precursors. *Annu. Rev. Biochem.* **61:** 419–440.

Konstantinov, K., R. Foisner, D. Byrd, F.-T Liu, W.-M Tsai, A. Wiik, and L. Gerace. 1995. Integral membrane proteins associated with the nuclear lamina are novel autoimmune antigens of the nuclear envelope. *Clin. Imunnol. Immunopathol.* **74:** 89–99.

McKeown, M. 1992. Alternative mRNA splicing. *Annu. Rev. Cell Biol.* **8:** 133–155.

Nigg, E.A. 1993. Cellular substrates of $p34^{cdc2}$ and its companion cyclin dependent kinases. *Trends Cell Biol.* **3:** 296–301.

Patschinsky, T., T. Hunter, F.S. Esch, J.A. Coo, and B.M. Seft. 1982. Analysis of the sequence of amino acids surrounding sites of tyrosine phosphorylation. *Proc. Natl. Acad. Sci.* **79:** 973–977.

Prestridge, D.S. 1995. Predicting Pol II promoter sequences using transcription factor binding sites. *J. Mol. Biol.* **249:** 923–932.

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

Santoro, C., N. Mermod, P.C. Andrews, and R. Tjian. 1988. A family of human CCAAT-box-binding proteins active in transcription and DNA replication: Cloning and expression of multiple cDNAs. *Nature* **334:** 218–224.

Schlesinger, D.H. and G. Goldstein. 1975. The amino acid sequence of thymopoietin II. *Cell* **5:** 361–365.

Zevin-Sonkin, D., E. Ilan, D. Guthmann, J. Riss, L. Theodor, and J. Shoham. 1992. Molecular cloning of the bovine thymopoietin gene and its expression in different calf tissues: Evidence for a predominant expression in thymocytes. *Immunol. Lett.* **31:** 301–310.

Received January 12, 1996; accepted in revised form March 29, 1996.