Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 2295–2297, April 1984
Biochemistry

Cloning and sequence analysis of cDNA for rat spleen thymosin β_4

(rat spleen mRNA/cDNA sequence/mRNA/peptide biosynthesis/peptide structure)

ALEKSANDRA WODNAR-FILIPOWICZ*, UELI GUBLER†, Y. FURUICHI*, MICHAEL RICHARDSON*, EUGENE F. NOWOSWIAT†, MOHINDAR S. POONIAN†, AND B. L. HORECKER*

*Roche Institute of Molecular Biology and †Department of Molecular Genetics, Roche Research Center, Hoffmann-La Roche Inc., Nutley, NJ 07110

Contributed by B. L. Horecker, December 23, 1983

ABSTRACT **Molecular cloning of a cDNA has established the sequence of the translated portion of the mRNA for rat spleen thymosin β_4. The presence of a methionyl initiator codon immediately preceding the codon for the first seryl residue of mature thymosin β_4 is consistent with previous results indicating the absence of a signal peptide in the product translated *in vitro* from rat spleen mRNA. The cDNA sequence analysis also established the presence of two terminator codons immediately following the codon for the COOH-terminal seryl residue. Thymosin β_4 is thus synthesized as a 5100-dalton peptide containing 44 amino acid residues. Removal of the initiator methionyl residue and acetylation of the NH_2-terminal serine residue would yield mature thymosin β_4 containing 43 amino acids. The absence of a signal peptide makes it unlikely that thymosin β_4 is a secreted peptide.**

Thymosin β_4 is a ubiquitous peptide, originally isolated from preparations of calf thymus fraction 5 (1). More recently, thymosin β_4 was shown to be widely distributed in cells and tissues of vertebrate species ranging from amphibia to mammals (refs. 2–5; for a review, see ref. 6). Its highly conserved structure suggests that it has an important biological function, but its proposed role as a thymic hormone (1) has been brought into question by the relatively large amounts found in tissues other than the thymus (2, 4) and the finding that it is actively synthesized by cells unrelated to the reticuloendothelial system (5).

In several cell-free translation systems, including reticulocyte lysates, wheat germ extracts, and a protein-synthesizing system from yeast, the product formed from rat spleen mRNA was shown to be identical to thymosin β_4, with no evidence for a signal or leader sequence or for the formation of a larger precursor polypeptide (7). This unusual pattern for the synthesis of a relatively small peptide, previously found in eukaryotes only for the protamines (8, 9), has now been confirmed by sequence analysis of a cloned cDNA derived from a rat spleen mRNA fraction. The translated mRNA sequence reported here consists of 44 codons, including the initiator methionyl codon plus 43 codons for mature thymosin β_4.

MATERIALS AND METHODS

Rat spleen poly(A)$^+$ RNA was isolated and the mRNA for thymosin β_4 was partially purified by sizing as described (7). Double-stranded cDNA was synthesized and cloned in *Escherichia coli* RR1 as described (10, 11). Two oligonucleotide probes were synthesized by the solid-state phosphite method (ref 12; see Fig. 1). The probes were labeled with ^{32}P at the 5′ termini (13) and used for colony sceening (14). Colony hybridizations were carried out in 0.75 M NaCl/0.075 M sodium citrate/0.01% $NaDodSO_4$/0.2% bovine serum albumin/0.2% polyvinylpyrrolidone/0.2% Ficoll/partially hydrolyzed yeast RNA (100 μg/ml) at 30°C for 2 hr using 1 pmol of labeled probe per ml. The filters were washed with 0.6 M NaCl/0.06 M sodium citrate for 60 min at 40°C, and the filters were exposed to x-ray film with an intensifying screen overnight.

RNA and Southern blot analysis, using thymosin β_4-specific DNA labeled by nick-translation, were carried out according to standard procedures (ref. 15 and methods cited in refs. 10 and 11). Determination of the nucleotide sequence of cloned cDNA was carried out by the Maxam–Gilbert method (13).

RESULTS

Preparation of a Clone Bank and Screening for Recombinants Containing cDNA Inserts Coding for Thymosin β_4. The synthesis of double-stranded cDNA was carried out using 8.7 μg of the partially purified rat spleen poly(A)$^+$ RNA as template (7). The yield in the first strand synthesis was 16.5%. An aliquot of this cDNA was rendered double-stranded, tailed with dGTP, and annealed to oligo(dC)-tailed *Eco*RV-cut pBR322 plasmid DNA. A portion of this recombinant plasmid corresponding to 80 ng of double-stranded cDNA was used for transformation of *E. coli* RR1, yielding 2 $\times$ 10^5 colonies. This corresponds to an overall cloning efficiency of 8 $\times$ 10^5 ampicillin-resistant transformants per μg of input mRNA.

Plates containing 20,000 colonies were screened by hybridization with the NH_2-terminal oligodeoxynucleotide probe (probe A, Fig. 1) and 12 well-separated strongly hybridizing colonies were selected for further analysis. The sizes of the cDNA inserts in the recombinant plasmids recovered from clones B and L were estimated by agarose gel electrophoresis to be 350 to 600 base pairs, after correcting for the presence of the oligo(C) and oligo(G) tails and the pBR322 sequences included in the restriction fragment. Southern blot analysis (Fig. 2) showed that each insert hybridized strongly to both the NH_2-terminal and the COOH-terminal oligodeoxynucleotide probes.

Sequence Analysis of the cDNA Insert in Clone B. The plas-

```
                    5                  10                  15
AcSer-Asp-Lys-Pro-Asp-Met-Ala-Glu- Ile-Glu-Lys-Phe-Asp-Lys-Ser-
                3'-CTP TAC CGN CTQ TA

                    20                  25                  30
Lys-Leu-Lys-Lys-Thr-Glu-Thr-Gln-Glu-Lys-Asn-Pro-Leu-Pro-Ser-

                    35                  40          43
Lys-Glu-Thr-Ile-Glu-Gln-Glu-Lys-Gln-Ala-Gly-Glu-SerOH
                3'-CTQ GTQ CTQ TTQ GTQ GC
```

Fig. 1. Structures of thymosin β_4 and the synthetic oligodeoxynucleotide probes. Probe A, corresponding to amino acid residues 5–9, was synthesized as a mixture of 16 different 14-mers. Probe B, corresponding to amino acid residues 34–40, was synthesized as a mixture of 32 different 17-mers. The letters N, Q, and P represent the presence of all four nucleotides, of both pyrimidines, or of both purines, respectively.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

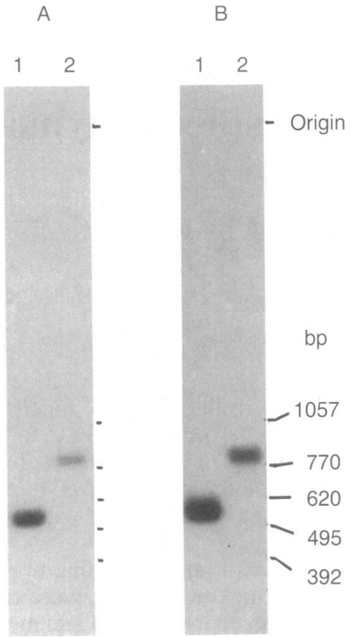

Fig. 2. Southern blot analysis of cloned thymosin β_4 cDNA. The recombinant plasmids (clones B and L), containing thymosin-specific DNA, were digested with the restriction enzyme *Bam*HI. The resulting cDNA inserts were electrophoresed on a 1% agarose gel, blotted onto nitrocellulose filters (14), and hybridized to the 5′-^{32}P-labeled NH_2-terminal and COOH-terminal oligonucleotide probes. The hybridization was carried out at 37°C overnight. The filter was washed twice, at 37°C, before autoradiography. The positions of the *Hinc*II digestion products of ϕX174 replicative form DNA are indicated for size comparison. (*A*) Hybridization to the NH_2-terminal probe (probe A). (*B*) Hybridization to the COOH-terminal probe (probe B). Lanes: 1, cDNA insert from clone B; 2, cDNA insert from clone L. bp, Base pairs.

mid from clone B was linearized with *Hind*III and the sequence was partially determined by the chain-terminator method (16), using probe B as the primer. This confirmed the presence of thymosin β_4 sequences in this plasmid.

The complete sequence of the cDNA was determined for clone B by the method of Maxam and Gilbert (ref. 13; Figs. 3 and 4). The insert was found to contain the entire coding sequence for thymosin β_4 with an initiator codon immediately preceding the first seryl residue and a pair of terminator codons immediately following the COOH-terminal seryl residue (Fig. 4). These findings confirm the results of cell-free translation experiments carried out with partially purified thymosin β_4 mRNA (7), which yielded a product identical in size and structure to thymosin β_4.

Size of the mRNA. The size of the thymosin β_4 mRNA was established by RNA blot analysis of total spleen poly(A)$^+$ RNA. Nick-translated clone B was used as a probe. A prominent RNA of ≈800 nucleotides was detected (Fig. 5).

Structural Features of the Amino Acid Sequence of Thymosin β_4. A computer calculation of the secondary structure of thymosin β_4, based on the Chou and Fasman algorithm (17), predicted α-helical structure for most of the molecule, except for a potential turn site located between residues 27 and 33. Hydrophilicity predicted for hexapeptide sequences of thymosin β_4 by the method of Hopp and Woods (18) was maximum for sequences 1–7, 11–24, and 33–41. No hexapeptide sequences yielded negative (hydrophobic) values. The major antigenic epitopes in thymosin β_4 crosslinked to keyhole limpet hemocyanin have previously been reported to include residues 1–8 and 22–32 (19).

DISCUSSION

Molecular cloning and sequence analysis of a cDNA have identified the complete translated sequence for the 43 amino acid residues of thymosin β_4. The locations of the initiator methionyl codon and of two terminator codons support the conclusion that thymosin β_4 is synthesized as a peptide containing 44 amino acids, with processing limited to removal of the first methionyl residue and acetylation of the exposed NH_2-terminal seryl residue. The absence of a larger polypeptide precursor containing a signal peptide confirms earlier evidence, based on its wide tissue distribution (2), high tissue content (4), and synthesis by a variety of established cell lines (5), that thymosin β_4 is not a secreted thymic hormone. The structural analysis reported here supports the earlier suggestion (5) that it may serve as a component of the cytoskeletal system.

The direct synthesis of a relatively small polypeptide, only 43 amino acid residues long, is an unusual finding for eukaryotic systems. To our knowledge, the only other example of this mode of synthesis is that of the protamines, which contain 31–33 amino acid residues and which have been shown to be synthesized from a small (4S–6S) mRNA (8, 9). The mRNA for thymosin β_4 appears to be larger, sedimenting at 8S–9S (7). Our analysis shows that the thymosin β_4 mRNA is 800 nucleotides long. Only 302 of the 800 nucleotides were present in clone B, which does not represent a full-length copy of the mRNA. Partial sequence analysis of the insert in clone L (unpublished data) showed it to contain the poly(A) segment of the mRNA.

The availability of the cloned cDNA for thymosin β_4 makes possible the isolation of the gene for this ubiquitous peptide, and perhaps also of the genes for the peptides related to thymosin β_4 (20). Analysis of the structure of these genes should provide valuable clues to their function.

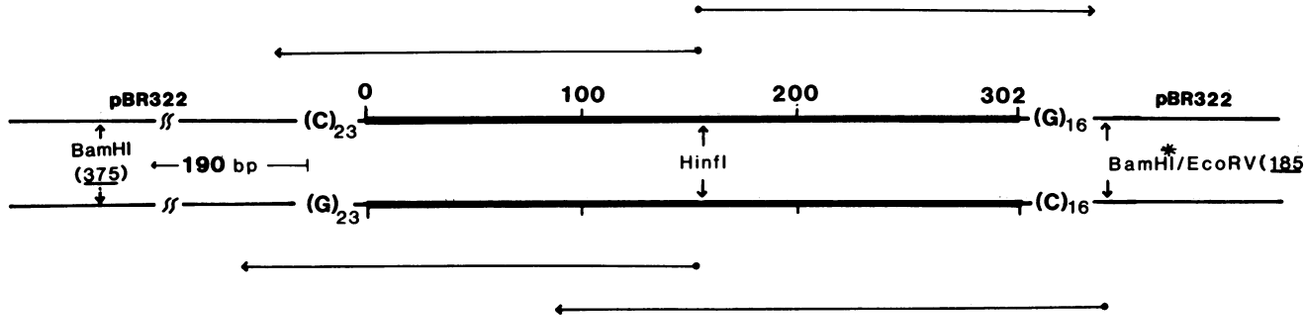

Fig. 3 Strategy for determining the nucleotide sequence of the cloned DNA copy of thymosin β_4 mRNA. The arrows indicate the direction of sequence analysis, and the solid circles identify the labeled sites. The numbers refer to the distance in nucleotides. $(C)_n$ and $(G)_n$ are the oligo(C) and oligo(G) residues added to form the chimeric plasmids for transformation. *Bam*HI* identifies a new *Bam*HI site (10) created in the preparation of the chimeric plasmid from the *Eco*RV site (residue 185 in pBR322).

```
                                  30                          54
GCCCAGCTCGCTCAGCTCCTTCCAGCAACCATG TCC GAC AAA CCC GAT ATG GTC
                                   1
                                 Met Ser Asp Lys Pro Asp Met Ala

                                 81                 100     108
GAG ATC GAG AAA TTC GAT AAG TCG AAG TTG AAG AAG ACA GAA ACA CAA GAG AAA
Glu Ile Glu Lys Phe Asp Lys Ser Lys Leu Lys Lys Thr Glu Thr Gln Glu Lys

                                 135                            162
AAT CCT CTG CCT TCA AAA GAA ACA ATT GAA CAA GAG AAG CAA GCT GGC GAA TCG
                                                                 43
Asn Pro Leu Pro Ser Lys Glu Thr Ile Glu Gln Glu Lys Gln Ala Gly Glu Ser

                                     200
TAA TGA GGCGAGCGCCGCCAATATGCACTGTACATTCCACGAGCATTGCCTTCTTATTTTAC
 -   -

TTCTTTTAGCTGTTTAACTTCGTAAGATGCAAAGAGGTTGGATCAAGTTTAAATGACTGTGCT

      300
GCCCCTTTCA
```

Fig. 4. Cloned cDNA from thymosin β_4 mRNA showing the translated sequence. The numbers 1 and 43 identify the NH_2-terminal and COOH-terminal amino acid residues, respectively, of thymosin β_4. The dashes identify termination codons.

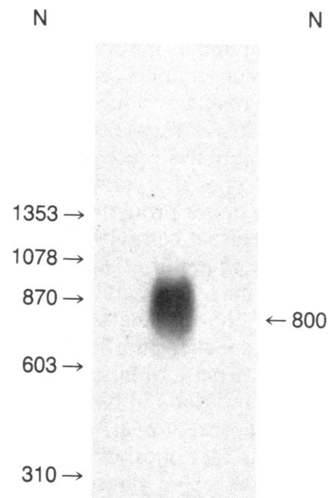

Fig. 5. The size of the mRNA for thymosin β_4 as determined by RNA blot analysis. Total poly(A)$^+$ RNA (5 μg) from rat spleen was denatured with glyoxal and analyzed by electrophoresis on a 1.4% agarose gel. The RNA was transferred to a nylon membrane and hybridized to the plasmid from clone B, labeled by nick-translation. The hybridized mRNA band was visualized by autoradiography. ϕX174 DNA restriction fragments were visualized by staining with ethidium bromide. N, nucleotide(s).

1. Low, T. L. K., Hu, S.-K. & Goldstein, A. L. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 1162–1166.
2. Hannappel, E., Xu, G.-J., Morgan, J., Hempstead, J. & Horecker, B. L. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 2172–2175.
3. Xu, G.-J., Hannappel, E., Morgan, J., Hempstead, J. & Horecker, B. L. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4006–4009.
4. Erickson-Viitanen, S., Ruggieri, S., Natalini, P. & Horecker, B. L. (1983) *Arch. Biochem. Biophys.* **221**, 570–576.
5. Goodall, G. J., Morgan, J. I. & Horecker, B. L. (1983) *Arch. Biochem. Biophys.* **221**, 598–601.
6. Horecker, B. L. & Morgan, J. (1984) in *Lymphokines*, ed. Pick, E. (Academic, New York), pp. 15–35.
7. Wodnar-Filipowicz, A. & Horecker, B. L. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1811–1815.
8. Gedamu, L. & Dixon, G. H. (1976) *J. Biol. Chem.* **251**, 1446–1454.
9. Gedamu, L. & Dixon, G. H. (1976) *J. Biol. Chem.* **251**, 1455–1463.
10. Gubler, U. & Hoffman, B. J. (1983) *Gene* **25**, 263–269.
11. Gubler, U., Monahan, J. H., Lomedico, P. T., Bhatt, R. S., Collier, K. J., Hoffman, B. J., Böhlen, P., Esch, F., Ling, N., Zeytin, F., Brazeau, P., Poonian, M. S. & Gage, L. P. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 4311–4314.
12. Matteucci, M. H. & Caruthers, M. H. (1981) *J. Am. Chem. Soc.* **103**, 3185–3191.
13. Maxam, A. M. & Gilbert W. (1980) *Methods Enzymol.* **65**, 499–560.
14. Hanahan, D. & Meselson, M. (1980) *Gene* **10**, 63–67.
15. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
16. Smith J. A. H. (1980) *Methods Enzymol.* **65**, 560–580.
17. Chou, P. Y. & Fasman, G. D. (1978) *Adv. Enzymol.* **47**, 45–148.
18. Hopp, T. P. & Woods, K. R. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 3824–3828.
19. Goodall, G. J., Hempstead, J. L. & Morgan, J. I. (1983) *J. Immunol.* **131**, 821–825.
20. Erickson-Viitanen, S., Ruggieri, S., Natalini, P. & Horecker, B. L. (1983) *Arch. Biochem. Biophys.* **225**, 407–413.