Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 343–346, January 1985
Biochemistry

Primary structure of rat thymus prothymosin α

(acidic polypeptide/thymic polypeptide/amino acid sequence/secondary structure)

A. A. HARITOS*, R. BLACHER†, STANLEY STEIN, J. CALDARELLA, AND B. L. HORECKER

Roche Institute of Molecular Biology and Biopolymer Research Department, Roche Research Center, Nutley, NJ 07110

Contributed by B. L. Horecker, September 19, 1984

ABSTRACT **The primary structure of prothymosin α from rat thymus, containing 113 amino acid residues, is reported as follows: AcSer-Asp-Ala-Ala-Val-Asp-Thr-Ser-Ser-Glu-Ile-Thr-Thr-Lys-Asp-Leu-Lys-Glu-Lys-Lys(20)-Glu-Val-Val-Glu-Glu-Ala-Glu-Asn-Gly-Arg(30)-Asp-Ala-Pro-Ala-Asn-Gly-Asn-Ala-Gln-Asn(40)-Glu-Glu-Asn-Gly-Glu-Gln-Glu-Ala-Asp-Asn(50)-Glu-Val-Asp-Glu-Glu-Glu-Glu-Glu-Gly-Gly(60)-Gly-Glu-Glu(Asx,Gly,Gly,Glx,Glx,Glx,Glx(70),Glx,Glx,Glx,Glx,Glx,Glx)Asn-Gly-Asp-Glu(80)-Asp-Glu-Glu-Ala-Glu-Ala-Pro-Thr-Gly-Lys(90)-Arg-Val-Ala-Glu-Asp-Asp-Glu-Asp-Asp-Asp(100)-Val-Glu-Thr-Lys-Lys-Gln-Lys-Lys-Thr-Asp(110)-Glu-Asp-AspOH. The sequence of the first 28 amino acids at the NH_2 terminus is identical to that of calf thymosin α_1. The dicarboxylic amino acids, which account for nearly half of the total residues in prothymosin α, are largely clustered in the central portion of the polypeptide chain. The polypeptide contains no aromatic or sulfur-containing amino acids. A computer analysis of the three-dimensional structure based on the primary sequence suggests that the molecule is composed of at least five α-helical regions interrupted by one short extended chain and three short random coils.**

A mixture of peptides from calf thymus identified as thymosin fraction 5 (1–3) has been reported to restore immune function in thymectomized mice (1) and to yield positive results in other *in vivo* and *in vitro* assays for the induction and maintenance of immune function (ref. 4; for reviews of the earlier literature, see refs. 3, 5, and 6). Thymosin fraction 5 also has shown promise in clinical trials with children with immunodeficiency diseases (7) and with cancer patients (8).

Thymosin fraction 5 was shown to be a mixture of peptides ranging in molecular weight from 1000 to 15,000 and in isoelectric point from 4.0 to 7.0 (9). The first component of thymosin fraction 5 to be isolated in pure form and characterized was thymosin α_1, an acidic peptide containing 28 amino acid residues (9). This peptide was reported to show many of the biological activities of thymosin fraction 5 (refs. 9–12; for a review, see ref. 6). Both thymosin fraction 5 and thymosin α_1 protected immunodeficient mice against challenge with *Candida albicans* (13–15) and other opportunistic microorganisms (14). Two peptides closely related to thymosin α_1 have recently been isolated from preparations of calf thymosin fraction 5 (16). They were found to be identical to α_1 in amino acid sequence, except that one, des-(25–28)-thymosin α_1, was four amino acid residues shorter and the other, thymosin α_{11}, contained seven additional amino acids at the COOH terminus. The presence in thymosin fraction 5 of three peptides differing only in length at the COOH terminus supported our earlier suggestion (17) that they were fragments derived by proteolysis of a larger native polypeptide.

We recently have described the isolation of a polypeptide, named prothymosin α, that meets this expectation (18). It was detected by using a radioimmunoassay for thymosin α_1 and was the only immunoreactive substance detected in the thymus extracts. For its isolation we used a procedure that precluded the possibility of proteolytic modification. Prothymosin α was characterized as a highly acidic polypeptide (pI = 3.55) containing approximately 113 amino acid residues with the thymosin α_1 sequence at its NH_2 terminus (18). In the mouse protection test and in assays for lymphokine release, prothymosin α was found to be 10–20 times as potent as thymosin α_1 (unpublished observations). Its primary sequence and secondary structure deduced from the primary sequence are reported here.

MATERIALS AND METHODS

Materials. All chemicals and solvents used were chromatography grade. Trypsin [L-1-tosylamino-2-phenylmethyl chloromethyl ketone (TPCK)-treated] was purchased from Worthington; *Staphylococcus aureus* V8 protease, from Miles; and thermolysin and carboxypeptidase Y, from Sigma. Hydroxylamine was from Mallinckrodt. Prothymosin α was isolated from rat thymus (18).

Methods. Digestion with *S. aureus* V8 protease was carried out by adding 1.9 μg of the enzyme to 75 μg of prothymosin α dissolved in 554 μl of 50 mM ammonium acetate (pH 4.0). After incubation for 15 hr at room temperature, the sample was lyophilized.

For digestion with thermolysin, a solution containing 10 μg of the enzyme and 267 μg of prothymosin α in 25 μl of 0.12 M ammonium bicarbonate was incubated for 2 hr at 37°C and lyophilized.

Digestion with trypsin was carried out with 10 μg of L-1-tosylamino-2-phenylmethyl chloromethyl-treated trypsin dissolved in 4 μl of 1 mM HCl added to 200 μg of prothymosin α in 52 μl of 0.4 M pyridine. After incubation for 15 hr at room temperature, the reaction was terminated by the addition of 4.0 μl of formic acid, followed by 60 μl of water and 40 μl of buffer A (1 M formic acid/0.2 M pyridine, pH 2.8).

For cleavage with hydroxylamine (19), 400 μg of prothymosin α was dissolved in 60 μl of 2 M hydroxylamine solution adjusted to pH 9.0 with potassium carbonate. After incubation for 14 hr at 50°C, the reaction was terminated by the addition of 26 μl of formic acid and 64 μl of buffer A.

For digestion with carboxypeptidase Y (20), 150 μg (12.6 pmol) of prothymosin α was dissolved in 96 μl of 50 mM sodium acetate (pH 5.5). To this buffer, 10-μg aliquots of carboxypeptidase Y were added at 0, 4, and 8 hr. At timed intervals 5–10% of the total volume was removed, diluted to 100 μl with ice-cold buffer of 67 mM sodium citrate (pH 2.0),

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.

*Present address: Zoological Laboratory, Faculty of Science, University of Athens, Athens 621, Greece.
†Present address: Chiron Corporation, 4560 Horton Street, Emeryville, CA 94608.

FIG. 1. The amino acid sequences of thymosins α_1 and α_{11} isolated from calf thymus fraction 5 and of prothymosin α isolated from rat thymus. The peptides used to establish the primary structure of prothymosin α are shown by the bars and letters, designated according to the cleavage method used as follows: H, cleavage by hydroxylamine; S, *S. aureus* V8 protease; T, trypsin; and Th, thermolysin. ▬, Sequences established by automated Edman degradation. The amino terminus is acetylated for all three peptides. The composition of the segment including residues 64–76 based on the results shown in Table 1 is shown in parentheses. The sequence of the first 20 residues is based on the composition of the peptides T1, T2, Th1, and Th2 and the published sequences of thymosin α_1 (1) and thymosin α_{11} (16). The COOH-terminal sequence Glu-Asp-AspOH was confirmed by the order of release of aspartic and glutamic acids by carboxypeptidase Y. The recommended one-letter notation for amino acids (23) is used.

1.5% (vol/vol) 1-propanol, and norleucine as internal standard, and frozen for subsequent amino acid analysis.

The peptide mixtures generated by enzymatic digestion and chemical hydrolysis were dissolved in buffer A, isolated by reversed-phase HPLC, and analyzed for amino acid composition as described (16).

Automated amino acid sequence analysis was performed on 0.5–1.5 nmol of peptides by using a gas-phase sequenator [Applied Biosystems (Foster City, CA), Model 470A] as described by Hewick *et al.* (21). Phenylthiohydantoin derivatives of amino acids were separated on HPLC by using an Ultrasphere ODS column (Altex) as described by Hawke *et al.* (22).

RESULTS

Determination of the Amino Acid Sequence. Automated sequence analysis could not be applied directly to prothymosin α because of the blocked NH_2 terminus (18). The primary structure (Fig. 1) was established by sequence analysis of peptides generated by digestion with trypsin (peptides designated "T" in Fig. 1 and Table 2), *S. aureus* V8 protease ("S"), and thermolysin ("Th") and by chemical cleavage with hydroxylamine ("H") specific for Asn-Gly bonds (19).

The sequence of the first 20 amino acid residues was deduced from the amino acid compositions of peptides T1, T2, Th1, and Th2 (Table 2) and the expected identity of this segment with thymosin α_1 (9) and thymosin α_{11} (16). Peptide T1 was clearly derived from the blocked NH_2 terminus based on the following observations. (*i*) It yielded negative results in the automated sequence analysis. (*ii*) Its composition was identical to that of residues 1–10 of thymosin α_1. (*iii*) It was detected in the HPLC fractions by its reaction with the anti-thymosin α_1 antiserum, which had been shown to be directed against an epitope including residues 1–10 of thymosin α_1 (unpublished results). This evidence also supports the identification of the blocking group as an acetyl residue.

Two tryptic peptides (T3 and T4), having identical amino acid compositions (Table 2) corresponding to residues 21–30 of thymosin α_{11} (Fig. 1) were recovered from the HPLC separations in the ratio of 2 equivalents of T4 to 1 of T3. Sequence analysis showed that they differed only at position 28, T3 yielding aspartic acid and T4 yielding asparagine. The presence of either amino acid at this position had been reported earlier in the sequence analysis of thymosin α_{11} (ref. 16; see Fig. 1). Peptide T5 was assigned to the position shown on the basis of the overlap with the known sequence of thymosin α_{11} (Fig. 1). Although this large peptide emerged as a single peak on HPLC, it also yielded both aspartic acid (38%) and asparagine (62%) at the position corresponding to residue 35 of prothymosin α. The first amino acid in this peptide was aspartic acid, replacing glutamic acid in calf thymosin α_{11}. This represents the only difference in the NH_2-terminal sequence of prothymosin α from rat thymus and the corresponding sequence of thymosin α_{11} from calf thymus.

Digestion of prothymosin α with carboxypeptidase Y indi-

Table 1. Amino acid compositions of prothymosin α and of the sequenced and unsequenced segments

	Prothymosin*	Sequenced segments†	Unsequenced segments‡
Asx	24.2 ± 1.1 (24)	23	1
Thr	6.3 ± 0.3 (6)	6	0
Ser	3.5 ± 0.3 (3)	3	0
Glx	39.0 ± 1.8 (39)	29	10
Gly	10.5 ± 0.3 (10)	8	2
Ala	10.1 ± 0.1 (10)	10	0
Val	5.6 ± 0.2 (6)	6	0
Ile	1.0 ± 0.0 (1)	1	0
Leu	1.1 ± 0.1 (1)	1	0
Lys	8.9 ± 0.3 (9)	9	0
Arg	2.1 ± 0.1 (2)	2	0
Pro	2.3 ± 0.5 (2)	2	0
Total	(113)	100	13

*The numbers in parentheses represent the nearest integral values. In the previous publication (18) the number of glycine residues was reported to be five. This was later found to be due to an error in the glycine content in the amino acid mixture used to standardize the analysis.

†The sum of residues 1–63 and 77–113.

‡Calculated from the difference between the total composition (column 2) and the values for the sequenced portions shown in column 3.

Table 2. Amino acid compositions of peptide fragments derived from prothymosin α

	T1 (1–14)	T2 (15–17)	T3 (21–30)	T4 (21–30)	T5 (31–90)	Th1 (11–15)	Th2 (16–21)	Th3 (92–113)	H1 (44–113)	H2 (77–113)	S1 (83–94)
Asx	2.0	1.0	1.3	1.0	10.1	1.1		8.1	15.7	10.5	0.9
Thr	2.9				1.4	1.9		2.2	2.7	3.4	0.9
Ser	2.8										
Glx	1.0		4.0	4.0	23.4		2.1	4.7	32.1	10.8	3.5
Gly			1.0	1.0	6.9				7.3	3.0	1.2
Ala	2.0		1.0	1.0	5.6			0.9	4.6	3.0	2.7
Val	1.0		1.5	1.5	1.0			2.0	3.0	2.3	1.0
Ile	0.9					1.0					
Leu		1.0					1.0				
Lys	1.1	1.0	0.3	0.2	1.0	1.4	3.3	4.0	4.7	4.9	1.2
Arg			0.9	1.0					1.4	1.1	1.0

The numbers in parentheses represent the location of the peptide in the primary sequence (Fig. 1). Proline was not determined in these analyses.

cated a COOH-terminal sequence of Glu-Asp-AspOH. This sequence also was found at the COOH terminus of peptide Th3, which established the position of that peptide (Fig. 1).

Partial sequence analysis of peptides derived by cleavage with hydroxylamine (peptides H1 and H2) provided the overlaps between peptides T5 and the peptides derived from the COOH terminus.

The segment following residue 63 could not be sequenced unambiguously because it was composed almost entirely of glutamic acid residues and followed a sequence that was also rich in glutamic acid. The composition of this peptide was estimated by subtracting the residues in the known sequences from the total amino acid composition (Table 1). The residue at position 77 was assumed to be asparagine as required by the specificity of cleavage by hydroxylamine.

Prediction of the Secondary Structure. A computer calculation of the secondary structure of prothymosin α based on the predictive model of Chou and Fasman (24, 25) indicated an extended structure composed of five helical segments, including residues 1–7, 14–30, 37–60, 78–86, and 90–113, with an extended chain and two short random coils separating these segments (Fig. 2). The unsequenced segment from residue 60 to residue 78 also would have a high probability of α-helical structure because of its high content of glutamic acid or glutamine.

DISCUSSION

Prothymosin α is an unusual polypeptide with respect to its high content of acidic amino acids, its low content of hydrophobic amino acids, and the absence of aromatic and sulfur-containing amino acids. Two-thirds of the acidic residues are located in the middle half of the sequence, between the two proline residues at positions 33 and 87. This segment contains no basic amino acids. Despite its high content in glutamic acid, it was resistant to digestion with *S. aureus* V8 proteinase.

The presence of both aspartic acid and asparagine at positions 28 and 35 may reflect partial deamidation during the isolation procedure, perhaps as a result of the preliminary heating to inactivate proteinases (18). However, the fact that the ratio of asparagine to aspartic acid was found to be almost constant from preparation to preparation (unpublished results) and the fact that the same heterogeneity is found in thymosin α_{11} (16) isolated by a completely different procedure suggests that this structural heterogeneity may reflect genetic polymorphism. The resolution of this question and the sequence of the short missing segment must await the results of cloning of the cDNA for the prothymosin α mRNA.

The fragments thymosin α_1 and thymosin α_{11} are both generated by the hydrolysis of Asn-Gly bonds at the positions showing structural heterogeneity, presumably by a proteinase with specificity for this bond. On the other hand, both des-(24–28)-thymosin α_1 and thymosin β_8, a fragment of thymosin β_9 present in preparations of calf thymosin fraction 5 (26), would arise from the hydrolysis of glutamyl bonds. Neither of these proteolytic specificities has yet been detected in animal tissues.

A computer-assisted search failed to identify any proteins with significant homology to prothymosin α. Its presence in highest concentrations in thymus and spleen and its potent activity in protecting mice against opportunistic infections (details to be published elsewhere) are consistent with a role for this polypeptide in the maintenance of immune function.

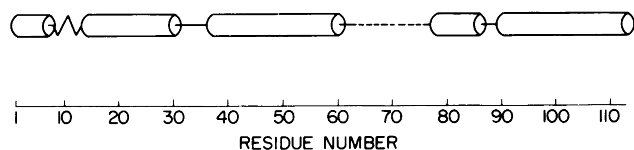

FIG. 2. Secondary structure of prothymosin α predicted from the amino acid sequence. Residues were scored for their relative tendencies to exist in three possible states—i.e., α-helix (barrels), extended chain (β-sheet, zigzags) and coil (straight lines). The unsequenced segment is shown as a broken line.

The authors are indebted to Dr. Gary L. Gilliland of the Department of Physical Chemistry for the computer analysis of the secondary structure based on the predictive model of Chou and Fasman.

1. Goldstein, A. L., Guha, A., Zatz, M. M., Hardy, M. A. & White, A. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 1800–1803.
2. Hooper, J. A., McDaniel, M. C., Thurman, G. B., Cohen, G. H., Schulof, R. S. & Goldstein, A. L. (1975) *Ann. N.Y. Acad. Sci.* **249,** 125–144.
3. Low, T. L. K., Thurman, G. B., Chincarini, C., McClure J. E., Marshall, G. C., Hu, S.-K. & Goldstein, A. L. (1979) *Ann. N.Y. Acad. Sci.* **332,** 33–48.
4. Zatz, M. M., Oliver, J., Samuels, C., Skotnicki, A. B., Sztein, M. B. & Goldstein, A. L. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 2882–2885.
5. White, A. (1980) in *Biochemical Actions of Hormones*, ed. Litwack, G. (Academic, New York), Vol. 7, pp. 1–46.
6. Schulof, R. S., Low, T. L. K., Thurman, G. B. & Goldstein, A. L. (1981) in *The Lymphocyte*, eds. Sell, K. W. & Miller, W. V. (Liss, New York), pp. 191–215.
7. Wara, D. W., Barrett, D. J., Ammann, A. J. & Cowan, M. J. (1979) *Ann. N.Y. Acad. Sci.* **332,** 128–134.
8. Chretien, P. B., Lipson, S. D., Makuch, R. W. & Kenady, D. E. (1979) *Ann. N.Y. Acad. Sci.* **332,** 135–147.
9. Goldstein, A. L., Low, T. L. K., McAdoo, M., McClure, J., Thurman, G. B., Rossio, J., Lai, C.-Y., Chang, D., Wang, S.-S., Harvey, C., Ramel, A. H. & Meienhofer, J. (1977) *Proc.*

Natl. Acad. Sci. USA **74,** 725–729.
10. Goldschneider, I., Ahmed, A., Bollum, F. J. & Goldstein, A. L. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 2469–2473.
11. Huang, K.-Y., Kind, P. D., Jagoda, E. M. & Goldstein, A. L. (1981) *J. Interferon Res.* **1,** 411–420.
12. Hu, S.-K., Low, T. L. K. & Goldstein, A. L. (1981) *Mol. Cell. Biochem.* **41,** 49–58.
13. Bistoni, F., Marconi, P., Frati, L., Bonmassar, E. & Garaci, E. (1982) *Infect. Immun.* **36,** 609–614.
14. Ishitsuka, H., Umeda, Y., Nakamura, J. & Yagi, Y. (1983) *Cancer Immunol. Immunother.* **14,** 145–150.
15. Salvin, S. B. & Neta, R. (1983) *Cell. Immunol.* **75,** 160–172.
16. Caldarella, J., Goodall, G. J., Felix, A. M., Heimer, E. P., Salvin, S. B. & Horecker, B. L. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 7424–7427.
17. Hannappel, E., Davoust, S. & Horecker, B. L. (1982) *Biochem. Biophys. Res. Commun.* **104,** 266–271.
18. Haritos, A. A., Goodall, G. J. & Horecker, B. L. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 1008–1011.
19. Bornstein, P. & Balian, G. (1977) *Methods Enzymol.,* **47,** 132–145.
20. Jones, B. N., Paabo, S. & Stein, S. (1981) *J. Liq. Chromatogr.* **4,** 565–586.
21. Hewick, R. M., Hunkapiller, M. W., Hood, L. E. & Dreyer, W. J. (1981) *J. Biol. Chem.* **256,** 7990–7997.
22. Hawke, D., Yuan, P.-M. & Shively, J. E. (1982) *Anal. Biochem.* **120,** 302–311.
23. IUPAC-IUB Commission on Biochemical Nomenclature (1968) *Eur. J. Biochem.* **5,** 151–153.
24. Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13,** 211–222.
25. Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13,** 222–245.
26. Hannappel, E., Davoust, S. & Horecker, B. L. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 1708–1711.