Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 5840–5843, August 1986
Biochemistry

Amino acid sequence of a prothoracicotropic hormone of the silkworm *Bombyx mori*

(insulin/relaxin)

HIROMICHI NAGASAWA*, HIROSHI KATAOKA*, AKIRA ISOGAI*, SABURO TAMURA*, AKINORI SUZUKI*, AKIRA MIZOGUCHI†, YUKO FUJIWARA†, ATSUSHI SUZUKI†, SUSUMU Y. TAKAHASHI†, AND HIRONORI ISHIZAKI†

*Department of Agricultural Chemistry, Faculty of Agriculture, The University of Tokyo, Tokyo 113, Japan; and †Biological Institute, Faculty of Science, Nagoya University, Chikusa-ku, Nagoya 464, Japan

Communicated by Gordon H. Sato, April 23, 1986

ABSTRACT **We have determined the complete amino acid sequence of 4K-PTTH-II, one of three forms of the M_r 4400 prothoracicotropic hormone of the silkworm *Bombyx mori*, active to brainless pupae of *Samia cynthia ricini*. Like vertebrate insulin, it consists of two nonidentical peptide chains (A and B chains). The A chain consists of 20 amino acid residues. The B chain is a mixture of four microheterogeneous peptides, two of which consist of 28 residues, and the other two, of 26 residues. 4K-PTTH-II has considerable sequence homology (40%) with human insulin, and it resembles porcine relaxin both in the carboxyl-terminal cysteine residue of the A chain and in the amino-terminal pyroglutamic acid residue of the B chain. The identical distribution of the six cysteine residues also indicates that 4K-PTTH-II belongs to the insulin family.**

Prothoracicotropic hormone (PTTH) is a brain peptide responsible for activation of prothoracic glands to produce ecdysone in insects (1). The hormonal activity has been recognized since 1922 (2), but only since 1982 has it been possible to isolate sufficient material for chemical characterization (3, 4). The adult heads of the silkworm *Bombyx mori* have two kinds of PTTHs: 4K-PTTH, which is active to brainless pupae of *Samia cynthia ricini* and whose molecular weight was estimated to be 4400, and 22K-PTTH, which is active to brainless pupae of *Bombyx mori* and whose molecular weight was estimated to be 22,000 (5). 4K-PTTH consists of at least three molecular species, 4K-PTTH-I, -II, and -III, whose first 19 amino acid residues were determined by automated Edman degradation, showing that they have considerable sequence homology with vertebrate insulin and its related peptides (6). In this paper, we describe the determination of the whole amino acid sequence of 4K-PTTH-II by chemical modification, enzymatic degradation, and microsequencing of the resulting peptide fragments and suggest that 4K-PTTH-II belongs to the insulin family of peptides.

MATERIALS AND METHODS

Isolation of 4K-PTTH-II. 4K-PTTH-II was purified from heads of male adult *B. mori* according to the method reported previously (6). The yield was about 300 μg per 10^6 *Bombyx* heads.

Reduction and Alkylation. 4K-PTTH-II (40 nmol) in 50 μl of 0.5 M Tris·HCl buffer (pH 8.5) containing 6 M urea was mixed with 120 μl of 0.04 M dithiothreitol in the same buffer, and the mixture was incubated under a nitrogen atmosphere at 37°C for 1 hr. Then 100 μl of 0.1 M iodoacetamide in distilled water was added to the reaction mixture, which was maintained in the dark under a nitrogen atmosphere at room temperature for 20 min. The reaction mixture was subjected to reversed-phase HPLC (RP-HPLC) on a column of Develosil 5-ODS (Nomura Kagaku) to give two peptides, carboxamidomethyl A and B chains (DA and DB). Linear gradient elution with 10–40% acetonitrile in 10 mM ammonium acetate was performed over 30 min. DB was applied to another RP-HPLC column (TSK ODS-120T, Toyo Soda, Tokyo) and eluted with a linear 30-min gradient of 25–32.5% acetonitrile in 0.1% trifluoroacetic acid to afford two peptides, DB1 and DB2.

Trypsin Digestion. The B-chain peptides DB1 (4 nmol) and DB2 (9 nmol) were dissolved separately in 40 μl of 0.1 M Tris·HCl, pH 7.8/0.02 M $CaCl_2$, to which 0.5 μg of trypsin (Sigma) in 1 μl of 1 mM HCl was added. The mixture was maintained at 32°C for 2 hr with occasional shaking. The resulting peptides were separated by RP-HPLC on a TSK ODS-120T column. A linear 50-min gradient of 5–40% acetonitrile in 0.1% trifluoroacetic acid was used for elution.

Carboxypeptidase A Digestion. A suspension of carboxypeptidase A (treated with diisopropyl fluorophosphate, Sigma) in water was centrifuged at 1000 × *g* for 5 min and the supernatant was removed to eliminate free amino acids. The precipitated carboxypeptidase A was dissolved in 0.1 M ammonium acetate (pH 9.0) at a concentration of 1 μg/μl, and 1 μl of this solution was added to 19 μl of the same buffer containing 4K-PTTH-II (0.7 nmol), DA (1.1 nmol), DB1 (0.6 nmol), or DB2 (1.0 nmol). The digestions were performed at 37°C. Aliquots (5 μl each) were taken at 0.5, 1, 2, and 4 hr after the initiation of incubation, and the released amino acids were quantified by RP-HPLC (Senshu Pak SEQ-4; Senshu Kagaku, Tokyo) after derivatization to phenylthiocarbamoyl amino acids (7).

Pyroglutamate Aminopeptidase Digestion. Tryptic peptides T2 (1.8 nmol) and T3 (5.0 nmol), and DB2 (0.5 nmol) were dissolved in 10 μl of 0.1 M phosphate buffer (pH 7.5) with 30 mM dithiothreitol. To these solutions, 4 μg of pyroglutamate aminopeptidase (Boehringer Mannheim) in 10 μl of the same buffer was added. After incubation at 35°C for 1 hr, the mixture was subjected to RP-HPLC on a TSK ODS-120T column developed with a gradient of 5–22.5% acetonitrile in 0.1% trifluoroacetic acid.

Chymotrypsin Digestion. DB1 (3.0 nmol) and DB2 (7.0 nmol) were dissolved in 20 μl of 0.1 M Tris·HCl buffer (pH 7.8). To these solutions, 0.5 μg of α-chymotrypsin (Sigma) in 1 μl of 1 mM HCl was added. After incubation at 25°C for 3 hr, the resulting peptides were separated by RP-HPLC on a Develosil 5-ODS column with a linear gradient of 5–40% acetonitrile in 10 mM ammonium acetate.

Amino Acid Analyses. Peptides (30–200 pmol) were hydrolyzed in constant-boiling HCl (Pierce) at 110°C for 20 hr.

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.

Abbreviations: PTTH, prothoracicotropic hormone; RP-HPLC, reversed-phase high-performance liquid chromatography.

After removal of HCl under reduced pressure, the hydrolysates were analyzed on a Shimazu HPLC system for amino acid analysis by post-column oxidation with sodium hypochlorite and successive derivatization with *o*-phthalaldehyde (8). This system is highly sensitive but far less accurate in quantification of proline than of the other amino acids.

Amino Acid Sequence Analyses. Some short peptides were subjected to manual Edman degradation, essentially according to the method of Edman and Henschen (9). All other peptides were sequenced by automated gas-phase sequencer (Applied Biosystems, Foster City, CA; model 470A). The phenylthiohydantoin derivatives of the amino acids obtained from each cycle of Edman degradation were identified by RP-HPLC on an Ultrasphere 5-ODS (Altex, Berkeley, CA) column at 40°C, using a gradient of 10–50% acetonitrile in 10 mM sodium acetate buffer (pH 4.5) for elution.

RESULTS

Fig. 1 summarizes the sequencing data of 4K-PTTH-II. Automated Edman degradation of intact 4K-PTTH-II afforded a single amino acid sequence through position 19, except that positions 6, 7, and 11 were not identifiable (6). On the other hand, incubation of intact 4K-PTTH-II with carboxypeptidase A resulted in a very slow release of approximately equimolar amounts of only aspartic acid and valine (16, 27, 39 and 96 pmol of aspartic acid and 18, 24, 35, and 102 pmol of valine at 0.5, 1, 2, and 4 hr, respectively). These data indicated that the carboxyl-terminal sequence was -Val-Asp rather than -Asp-Val, considering the substrate specificity of carboxypeptidase A, which releases carboxyl-terminal amino acids with aromatic or branched aliphatic groups more rapidly and removes only carboxyl-terminal amino acids with a free carboxyl group (10). Reductive alkylation of 4K-PTTH-II gave two peptides, carboxamidomethyl-A and -B (DA and DB), which were separated by RP-HPLC (Fig. 2*a*). Amino acid compositions of DA and DB were quite different, and the total number of residues of each amino acid of DA and DB was almost equal to that of intact 4K-PTTH-II (Table 1), indicating that 4K-PTTH-II consists of two peptide chains, A and B, connected to one another by disulfide bonds.

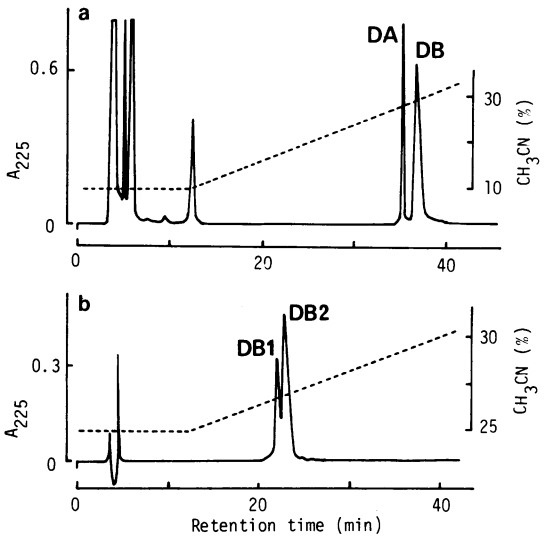

FIG. 2. Separation of A and B chains of 4K-PTTH-II. (*a*) HPLC on a Develosil 5-ODS column. Solvent A was 10% acetonitrile/10 mM ammonium acetate; solvent B was 40% acetonitrile/10 mM ammonium acetate. Elution was by a gradient of 0–100% solvent B over 40 min, at a flow rate of 1 ml/min. (*b*) HPLC on a TSK ODS-120T column. Solvent A was 10% acetonitrile/0.1% trifluoroacetic acid; solvent B was 40% acetonitrile in 0.1% trifluoroacetic acid. Elution was by a gradient of 50–75% solvent B over 30 min, at a flow rate of 0.8 ml/min.

Glycine was identified as the amino-terminal amino acid residue of DA, whereas no amino-terminal residue could be detected for DB, indicating that the amino terminus of the B chain is blocked. RP-HPLC elution with a slowly increasing concentration of acetonitrile (0.25% per min at a flow rate of 0.8 ml/min) resulted in separation of DB into two peaks, DB1 and DB2 (*ca.* 1:2 by area; Fig. 2*b*). Amino acid analyses of DB1 and DB2 suggested that the replacement of an alanine residue of DB2 with a glycine residue gives DB1.

Edman degradation of DA revealed carboxamidomethylcysteine [Cys(Cam)] residues at positions 6, 7, 11, and 20, in addition to the same sequence as obtained for intact 4K-

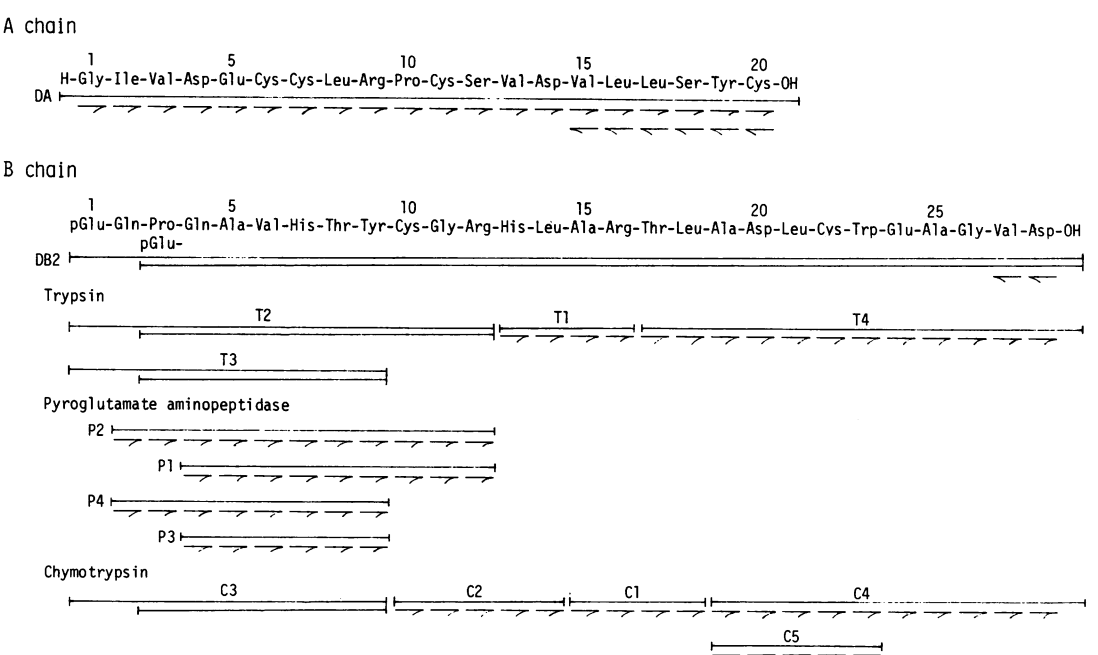

FIG. 1. Summary of sequencing data on A and B chains of 4K-PTTH-II. →, residue identified from the amino terminus by Edman degradation; ←, residue identified from the carboxyl terminus by carboxypeptidase A digestion. All cysteine residues were *S*-carboxamidomethylated. pGlu, pyroglutamic acid.

Table 1. Amino acid composition of 4K-PTTH-II and its A and B chains

Residue	4K-PTTH-II	DA	DB
Asx	3.8 (4)	2.0 (2)	2.2 (2)
Thr	2.0 (2)	0.1 (0)	2.1 (2)
Ser	1.8 (2)	1.9 (2)	0.4 (0)
Glx	5.1 (4.7)*	1.6 (1)	4.2 (3.7)*
Pro	2.7 (1.7)*	0.9 (1)	1.5 (0.7)*
Gly	3.6 (3.3)*	1.8 (1)	3.5 (2.3)*
Ala	3.8 (3.7)*	0.5 (0)	3.8 (3.7)*
Cys	3.6 (6)	ND (4)	ND (2)
Val	4.5 (5)	2.6 (3)	2.0 (2)
Met	0.0 (0)	0.0 (0)	0.0 (0)
Ile	0.8 (1)	0.7 (1)	0.1 (0)
Leu	6.1 (6)	3.2 (3)	2.6 (3)
Tyr	1.5 (2)	0.6 (1)	1.1 (1)
Phe	0.2 (0)	0.1 (0)	0.2 (0)
His	2.1 (2)	0.4 (0)	2.3 (2)
Lys	0.2 (0)	0.2 (0)	0.2 (0)
Arg	3.4 (3)	1.1 (1)	2.3 (2)
Trp	ND (1)	ND (0)	ND (1)

Numbers in parentheses are from amino acid sequence. ND, not done.
*Not an integer due to a mixture of four microheterogeneous peptides. For details, see text.

PTTH-II (Fig. 1). Fast release of carboxamidomethylcysteine, tyrosine, serine, and leucine was observed in carboxypeptidase A digestion of DA; the cleavage of Tyr[19]-Cys(Cam)[20] was nearly complete within 30 min, suggesting that the carboxyl-terminal residue is carboxamidomethylcysteine. Therefore, the A chain is composed of 20 amino acid residues.

A time-dependent, slow release of aspartic acid and valine by carboxypeptidase A digestion of DB2, as observed for intact 4K-PTTH-II, indicated the carboxyl-terminal sequence -Val-Asp. Since direct Edman degradation was not applicable to DB2 because of its blocked amino-terminus, DB2 was digested with trypsin. Four tryptic peptides, T1–T4, were obtained by HPLC (Fig. 3). Sequence analyses of these tryptic peptides showed that the amino termini of T2 and T3 are blocked, whereas those of T1 and T4 are free. T1 is a tetrapeptide with arginine at the carboxyl terminus, and T4 a dodecapeptide, whose carboxyl-terminal sequence, -Val-Asp, is identical to that obtained for DB2 and intact 4K-PTTH-II, indicating that T4 is a carboxyl-terminal fragment.

In consideration of the fact that the amino-terminal residue of most of the vertebrate peptide hormones with blocked amino termini is pyroglutamic acid (pGlu) and of the presence of about 3 mol of glutamic acid in both T2 and T3 by amino acid analyses (Table 2), T2 and T3 were digested with pyroglutamate aminopeptidase. As expected, the digestion yielded new peptides with free amino termini: P1 and P2 (1:2)

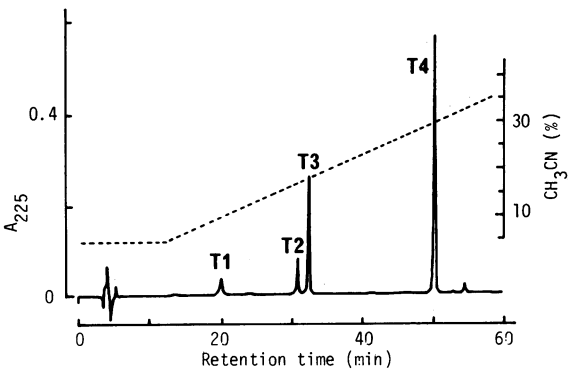

FIG. 3. Separation of tryptic peptides from *S*-carboxamidomethylated B chain by HPLC on a TSK ODS-120T column. Solvent A was 5% acetonitrile/0.1% trifluoroacetic acid; solvent B was 40% acetonitrile/0.1% trifluoroacetic acid. Elution was by a gradient of 0–100% solvent B over 50 min, at a flow rate of 1 ml/min.

were obtained from T2, and P3 and P4 (1:2) from T3. As shown in Fig. 1, P2 is longer than P1 by the amino-terminal dipeptide Glu-Pro-. Though pyroglutamic acid removed by pyroglutamate aminopeptidase could not be identified due to interference by the large amount of reagents originally contained in the pyroglutamate aminopeptidase preparation (sucrose, EDTA, dithiothreitol, and phosphate salts), the distinct decrease of the molar ratio of glutamic acid in the amino acid composition of P1 and P2 compared with that of T2 (Table 2) led to the conclusion that T2 is a mixture of pGlu-P1 and pGlu-P2. P3 and P4 are peptides that lack the carboxyl-terminal tripeptide -Cys(Cam)-Gly-Arg, from P1 and P2, respectively. Thus, T3 is a mixture of pGlu-P3 and pGlu-P4. The complete sequencing of the tryptic peptides allowed us to conclude that their sequence in DB2 is T2-T1-T4.

To confirm this conclusion, DB2 was digested with chymotrypsin. Five chymotryptic peptides, C1–C5, were obtained by HPLC, and their amino acid sequences are shown in Fig. 1. The sequences of C1 and C2 confirmed the connection of T1 to T4 and of T2 to T1, respectively. The connection was proved more directly by pyroglutamate aminopeptidase digestion of DB2, which generated two amino-terminal-free peptides, P5 and P6 (1:2). Edman degradation of P6 revealed the partial amino-terminal sequence Gln-Pro-Gln-Ala-Val-Xaa-Thr-Tyr-Cys(Cam)-Gly-Xaa-Xaa-Leu-Ala-Xaa-Thr-Leu-Ala-, though P6 could not be characterized fully. Therefore, the primary structure of DB2 was unambiguously determined to be a mixture of 28- and 26-residue peptides (Fig. 1).

Sequence analysis of DB1 was done in the same way as for DB2. The results showed that DB1 is also a mixture of 28- and 26-residue peptides (2:1) and that the only difference in the sequences between DB1 and DB2 exists at position 5: alanine

Table 2. Amino acid composition of peptide fragments from the B chain

Residue	T2	T3	P1	P2	P3	P4
Thr	0.9 (1)	1.0 (1)	0.7 (1)	0.8 (1)	0.9 (1)	0.9 (1)
Glx	2.8 (2.7)*	3.2 (2.7)*	1.1 (1)	2.1 (2)	1.3 (1)	2.0 (2)
Pro	1.0 (0.7)*	1.6 (0.7)*	0.0 (0)	0.9 (1)	0.0 (0)	1.3 (1)
Gly	1.1 (1)	0.2 (0)	1.5 (1)	2.1 (1)	0.4 (0)	0.3 (0)
Ala	1.0 (1)	1.0 (1)	0.9 (1)	1.2 (1)	0.8 (1)	1.0 (1)
Cys	ND (1)	ND (0)	ND (1)	ND (1)	ND (0)	ND (0)
Val	0.9 (1)	1.3 (1)	0.8 (1)	0.9 (1)	1.0 (1)	0.9 (1)
Tyr	0.5 (1)	0.4 (1)	0.5 (1)	0.5 (1)	0.6 (1)	0.7 (1)
His	0.8 (1)	1.0 (1)	0.9 (1)	0.9 (1)	1.1 (1)	1.1 (1)
Arg	1.0 (1)	0.0 (0)	0.7 (1)	0.9 (1)	0.0 (0)	0.0 (0)

Numbers in the parentheses are from amino acid sequence. ND, not done.
*Not an integer due to a mixture of two peptides. For details, see text.

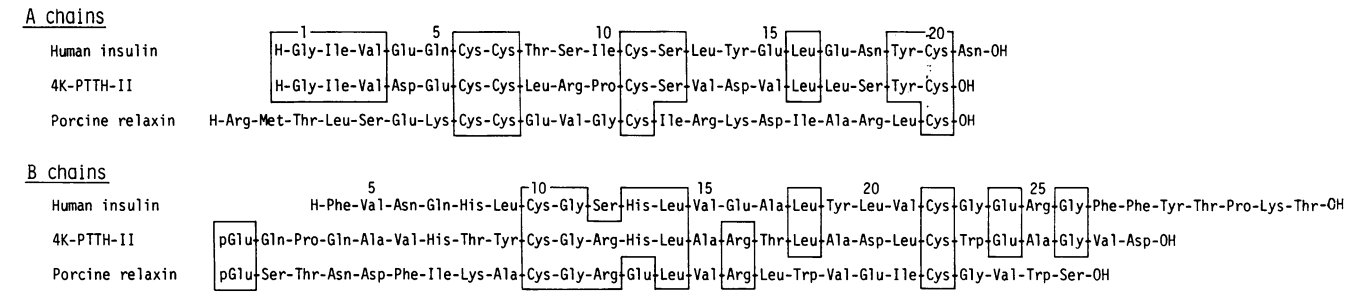

FIG. 4. Comparison of the amino acid sequence of 4K-PTTH-II with the sequences of human insulin and porcine relaxin. Identical residues are boxed. pGlu, pyroglutamic acid.

in DB2 is replaced by glycine in DB1. The approximate proportions of the four B-chain peptides [Ala^5]B28 (28-residue peptide with alanine at position 5 of the B chain), [Ala^5]B26, [Gly^5]B28, and [Gly^5]B26 are, respectively, 4:2:2:1.

DISCUSSION

Amino Acid Sequence of 4K-PTTH-II. The yields of 4K-PTTH-I, -II, and -III seem to vary from batch to batch. Since the yield of 4K-PTTH-II was highest in the present experiment, 4K-PTTH-II was used for sequence study.

The reductive alkylation of disulfide bonds of 4K-PTTH-II generated two nonidentical peptides, DA and DB. The amino acid sequence of DA was unambiguously determined by Edman degradation and carboxypeptidase A digestion. The amino acid sequence of DB was elucidated from the results of Edman degradation of peptide fragments obtained by digestion with various enzymes and of carboxypeptidase A digestion. DB is a mixture of four microheterogeneous peptides, whose amino termini are all blocked with pyroglutamic acid. Although four intact molecules of 4K-PTTH-II could not be separated under any HPLC conditions so far tested, DB was resolved into two peaks (DB1 and DB2, Fig. 2*b*) due to a difference of a residue only at position 5 (glycine or alanine). DB1 and DB2 were each a mixture of two peptides, of 28 and 26 residues, that were separated after pyroglutamate aminopeptidase digestion.

Carboxypeptidase A released only aspartic acid and valine slowly from intact 4K-PTTH-II, accounting for the inability of the enzyme to cleave Tyr^{19}-Cys^{20} in the A chain and Ala^{25}-Gly^{26} in the B chain. The former peptide bond, however, was cleaved quite easily by carboxypeptidase A after reduction of disulfide bonds. Since DB2 has two arginine residues, digestion of DB2 with trypsin was expected to give three peptide fragments. Unexpectedly, four tryptic peptides were obtained. T3 was generated by cleavage of Tyr^9-Cys(Cam)^{10} of the B chain, possibly due to chymotrypsin present as a contaminant in the trypsin used.

In summary, 4K-PTTH-II consists of A and B chains connected by disulfide bonds and is a mixture of four molecular species, whose structural differences are confined to the amino-terminal region of the B chain. A previous report (6) indicated the existence of at least three molecular species in 4K-PTTH–4K-PTTH-I, -II, and -III—that differed in the amino-terminal sequence of the A chain. Therefore, 4K-PTTH comprises many heterogeneous molecules, with amino acid substitutions or gaps in both A and B chains, although it is not apparent whether such heterogeneous 4K-PTTHs exist in each individual insect, because the starting material used for the extraction of 4K-PTTHs was a mixture of several races of the silkworm *B. mori*.

Homology with Insulin-Family Peptides. The complete amino acid sequence of a main component of 4K-PTTH-II is shown in Fig. 4 together with the sequences of human insulin and porcine relaxin (11) for comparison. As reported previously (6), considerable sequence homology was found in the amino-terminal 19 residues in the A chain between 4K-PTTH-II and human insulin and its related peptides. The B-chain sequence, determined in the present work, is less homologous (32%) with human insulin. In particular, no homology is present in the amino-terminal region of the B chain. Although 4K-PTTH-II is slightly homologous in amino acid sequence with relaxin, a peptide hormone also belonging to the insulin family, it still resembles porcine relaxin both in the carboxyl-terminal cysteine residue of the A chain and in the amino-terminal pyroglutamic acid residue of the B chain, both of which are important features of the relaxin molecule. The distribution of the six cysteine residues of 4K-PTTH-II is identical to the cysteine distribution of insulin, relaxin, and their related peptides, providing further evidence that 4K-PTTH-II belongs to the insulin family (12). These data suggest that the genes for insulin-family peptides have arisen from a common ancestral gene and that these genes must have been highly conserved during evolution to insects and mammals.

This work was supported by Grants-in-Aid for Scientific Research (nos. 58470106, 59360012, 60560133, and 61470134) from the Ministry of Education, Science, and Culture of Japan.

1. Ishizaki, H. & Suzuki, A. (1980) in *Neurohormonal Techniques in Insects*, ed. Miller, T. A. (Springer, New York), pp. 244–276.
2. Kopec, S. (1922) *Biol. Bull. (Woods Hole, Massachusetts)* **24,** 323–342.
3. Suzuki, A., Nagasawa, H., Kataoka, H., Hori, Y., Isogai, A., Tamura, S., Guo, F., Zhong, X.-C., Ishizaki, H., Fujishita, M. & Mizoguchi, A. (1982) *Agric. Biol. Chem.* **46,** 1107–1109.
4. Nagasawa, H., Kataoka, H., Hori, Y., Isogai, A., Tamura, S., Suzuki, A., Guo, F., Zhong, X.-C., Mizoguchi, A., Fujishita, M., Takahashi, S. Y., Ohnishi, E. & Ishizaki, H. (1984) *Gen. Comp. Endocrinol.* **53,** 143–152.
5. Ishizaki, H., Mizoguchi, A., Fujishita, M., Suzuki, A., Moriya, I., Ooka, H., Kataoka, H., Isogai, A., Nagasawa, H., Tamura, S. & Suzuki, A. (1983) *Dev. Growth Differ.* **25,** 593–600.
6. Nagasawa, H., Kataoka, H., Isogai, A., Tamura, S., Suzuki, A., Ishizaki, H., Mizoguchi, A., Fujiwara, Y. & Suzuki, A. (1984) *Science* **226,** 1344–1345.
7. Heinrikson, R. L. & Meredith, S. C. (1984) *Anal. Biochem.* **136,** 65–74.
8. Ishida, Y., Fujita, T. & Asai, K. (1981) *J. Chromatogr.* **204,** 143–148.
9. Edman, P. & Henschen, A. (1975) in *Protein Sequence Determination*, ed. Needleman, S. B. (Springer, Berlin), pp. 251–262.
10. Hartsuck, J. A. & Lipscomb, W. N. (1971) in *The Enzymes*, ed. Boyer, P. D. (Academic, New York), pp. 1–56.
11. Schwabe, C. & McDonald, J. K. (1977) *Science* **197,** 912–913.
12. Blundell, T. L. & Humbel, R. E. (1980) *Nature (London)* **287,** 781–787.