# Porcine Proinsulin: Characterization and Amino Acid Sequence

Abstract. *Proinsulin in nearly homogeneous form has been isolated from a preparation of porcine insulin. A molecular weight close to 9100 was calculated from the amino acid composition and from sedimentation-equilibrium studies. Through the action of trypsin this single-chain protein is transformed to desalanine insulin by cleavage of a polypeptide chain connecting the carboxy-terminus of the B chain to the amino-terminus of the A chain of insulin. The amino acid sequence of this connecting peptide was found to be Arg-Arg-Glu-Ala-Gln-Asn-Pro-Gln-Ala-Gly-Ala-Val-Glu-Leu-Gly-Gly-Gly-Leu-Gly-Gly-Leu-Gln-Ala-Leu-Ala-Leu-Glu-Gly-Pro-Pro-Gln-Lys-Arg.*

Steiner and co-workers have identified a possible insulin precursor from human beta-cell tumor (*1*, *2*) and from rat islet tissue (*2*) and proposed that these pancreatic proteins be designated "proinsulins." The data suggested that proinsulin is a single-chain protein that can be transformed to an insulin-like component through tryptic cleavage of a polypeptide chain connecting the COOH-terminus of the B chain to the $NH_2$-terminus of the A chain (*3*). Thus, evidence was provided that insulin may be derived from a single-chain protein rather than from the combination of two independently synthesized chains (*4*).

When these studies were first reported (*1*), it became clear that the properties of proinsulin were similar, if not identical, to those of a protein that had been isolated in our laboratory as a minor component of a crystalline porcine insulin preparation (Fig. 1). The purposes of this report are to describe the properties and the primary structure of this proinsulin-like protein, and thus to confirm and extend the data of Steiner and co-workers.

This proinsulin preparation was isolated during the chromatography of porcine insulin on columns of DEAE-cellulose (*5*) in urea-containing buffers according to methods described (*6*). A high degree of homogeneity was demonstrated by rechromatography, by electrophoresis in polyacrylamide gels (Fig. 1), and by ultracentrifugal studies (*7*). These sedimentation-equilibrium analyses established a molecular weight of $9100 \pm 300$, a value which agrees very well with the molecular weight of 9082 as calculated from amino acid analyses (Table 1). The amino acid composition of this porcine proinsulin preparation differs considerably from that of a partially purified bovine proinsulin preparation reported recently by Yip and Lin (*8*). This difference may be due to a species difference as well as to the degree of purity.

Phenylalanine was found to be the sole $NH_2$-terminal residue of proinsulin (Fig. 2) by both the Edman degradation (*9*) and the dansylation method (*10*). Using a combination of successive Edman reactions followed by dansylation of each newly released $NH_2$-group, we found that the $NH_2$-terminal tetrapeptide sequence was identical to that in the B chain of porcine insulin (*11*; Fig. 2). The single COOH-terminal residue was found to be asparagine, as determined both by hydrazinolysis (*12*) and by digestion with carboxypeptidase A. Further evidence that this proinsulin is a single-chain protein was provided by electrophoresis on both cellulose acetate strips (*13*) and polyacrylamide gels, after conversion to either the corresponding *S*-sulfonate (*14*) or the reduced-carboxymethylated form (*15*).

Proinsulin exhibited insulin-like activity to the extent of about 3 I.U./mg (international units) by the mouse-convulsion assay (*16*) and about 6 I.U./mg by the radioimmunoassay (*17*). Although this activity may or may not be intrinsic to proinsulin, it does not appear to be due to contamination with insulin. A contamination of this magnitude (10 to 25 percent, considering that insulin has a potency of 25 I.U./mg) would have been readily detected by disc electrophoresis (compare Fig. 1). Furthermore, insulin chains were not detected by cellulose-acetate electrophoresis after proinsulin was converted to the *S*-sulfonate form as described above. The hypoglycemic activity of proinsulin was increased significantly after a limited digestion with TPCK-trypsin (*18*). Judged by the mouse-convulsion assay such treatment (*19*) of proinsulin increased the activity of the digest mixture from 3 to about 12 I.U./mg; as judged by the radioimmunoassay, the activity increased from 6 to 12 I.U./mg. The disc-electrophoresis patterns of this digest resembled closely the 50-minute pattern shown in

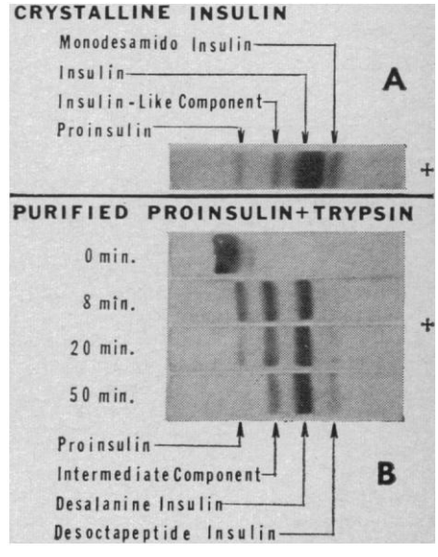


Fig. 1 (right). Comparison of polyacrylamide disc-gel electropherograms of crystalline porcine insulin, purified proinsulin isolated from the insulin, and proinsulin after mild tryptic digestions. The gels were prepared from 20 percent acrylamide and 0.2 percent *N,N'*-methylene-bisacrylamide. The electrophoretic fractionations were conducted essentially by the method of Davis (*24*), and protein components were stained with Coomassie blue (*25*). Tryptic digestions (enzyme : substrate ratio, 1 : 100 by weight) were performed at 37°C in 0.05*M* tris-HCl, 0.02*M* $CaCl_2$, *p*H 7.5. Reactions were stopped by addition of an equal volume of fresh, deionized 7*M* urea adjusted to *p*H 2.0 with concentrated HCl. All samples (100 $\mu$g in 100 $\mu$l) were applied directly to gels and were subjected to electrophoresis at 1 ma per gel tube for 15 minutes and then at 2 ma per tube for 105 minutes at 25°C. The most anodic component in crystalline insulin is desamidoinsulin (*26*), whereas the most anodic component in the 50-minute digest of proinsulin is desoctapeptide insulin (*6*). Insulin and desalanine insulin have the same net charge and migrate identically in this system. The intermediate component formed during the tryptic transformation of proinsulin is still undefined and may or may not be identical to the insulin-like component in crystalline insulin. This insulin-like component has been isolated from insulin by DEAE-cellulose chromatography and has essentially the same amino acid composition as insulin and a hypoglycemic potency of about 22 I.U./mg by mouse-convulsion assay.

Table 1. Molar amino acid ratios of insulin, proinsulin, and the isolated products from a tryptic digest of proinsulin. The molar ratios are based on aspartic acid as unity. All samples were hydrolyzed with 6*N* HCl at 105°C for 30 hours in sealed, evacuated tubes and analyzed on a Spinco 120C amino acid analyzer (*23*). Triplicate determinations were made on insulin and proinsulin and single determinations on all other samples.

| Amino acid | Porcine insulin | | Pro-insulin | Products from a tryptic digest of proinsulin* | | | |
|---|---|---|---|---|---|---|---|
| | Theory | Determined | | Desalanine insulin | Glutamyl-peptide | Ala-Arg | Arg |
| Tryptophan | | | 0.00† | | | | |
| Lysine | 1 | 0.96 | 1.95 | 1.00 | 0.92 | | |
| Histidine | 2 | 1.97 | 1.87 | 2.00 | | | |
| Arginine | 1 | 0.97 | 3.76 | 1.04 | 1.01 | 1.00 | 1.00 |
| Aspartic acid | 3 | 3.00 | 4.00 | 3.00 | 1.00 | | |
| Threonine | 2 | 1.90 | 1.86 | 1.86 | | | |
| Serine | 3 | 2.66 | 2.93 | 2.52 | | | |
| Glutamic acid | 7 | 7.01 | 13.50 | 7.19 | 7.13 | | |
| Proline | 1 | 0.98 | 4.00 | 0.97 | 2.75 | | |
| half-Cystine‡ | 6 | 5.65 | 5.25 | 5.37 | | | |
| Glycine | 4 | 4.04 | 10.59 | 3.98 | 7.13 | | |
| Alanine | 2 | 2.06 | 6.54 | 1.15 | 5.17 | 1.34 | |
| Valine | 4 | 3.60 | 4.53 | 3.89 | 1.01 | | |
| Isoleucine | 2 | 1.59 | 1.52 | 1.81 | | | |
| Leucine | 6 | 6.03 | 10.49 | 6.14 | 4.82 | | |
| Tyrosine | 4 | 3.80 | 3.60 | 3.49 | | | |
| Phenylalanine | 3 | 3.07 | 3.02 | 3.00 | | | |
| Total number of residues | 51 | 51 | 84§ | 50 | 31 | 2 | 1 |

* Desalanine insulin and the glutamyl peptide isolated by DEAE-cellulose chromatography. Ala-Arg and free arginine isolated by Sephadex G-25 chromatography. † No tryptophan was found in a 2*N* $Ba(OH)_2$ hydrolyzate, when separation was made on a column (0.9 by 50 cm) of Aminex A-4 resin (Bio-Rad) at 50°C and a 0.7*N* sodium citrate buffer, *p*H 4.26. ‡ Includes small amounts of cysteic acid. § The next higher integers were assumed for glutamic acid, glycine, alanine, and leucine.

Fig. 1, although the experimental conditions differed slightly. This increase in activity was the result of a transformation of the single-chain proinsulin to the two-chain desalanine insulin (*20*) through cleavage of a peptide fragment connecting the COOH-terminus of the B chain to the $NH_2$-terminus of the A chain (Fig. 1). Desalanine insulin was isolated from the 50-minute tryptic digest (Fig. 1) by chromatography on DEAE-cellulose and had normal biological potency (23 I.U./mg, by mouse-convulsion assay; 26 I.U./mg, by radioimmunoassay), the same amino acid composition as porcine insulin (Table 1) with the exception of one less alanine residue, and the same $NH_2$-terminal sequences ($A_1 - A_5$, $B_1 - B_6$) as porcine insulin (*11*; Fig. 2). These data provided strong evidence that the primary structure of the isolated desalanine insulin was identical to that of porcine insulin (*11*) except that it lacked the $B_{30}$ alanyl residue (*20*).

The amino acid compositions of the other products of trypsin digestion are listed in Table 1. In addition to desalanine insulin, the mild tryptic hydrolysis yielded a 31-residue peptide with

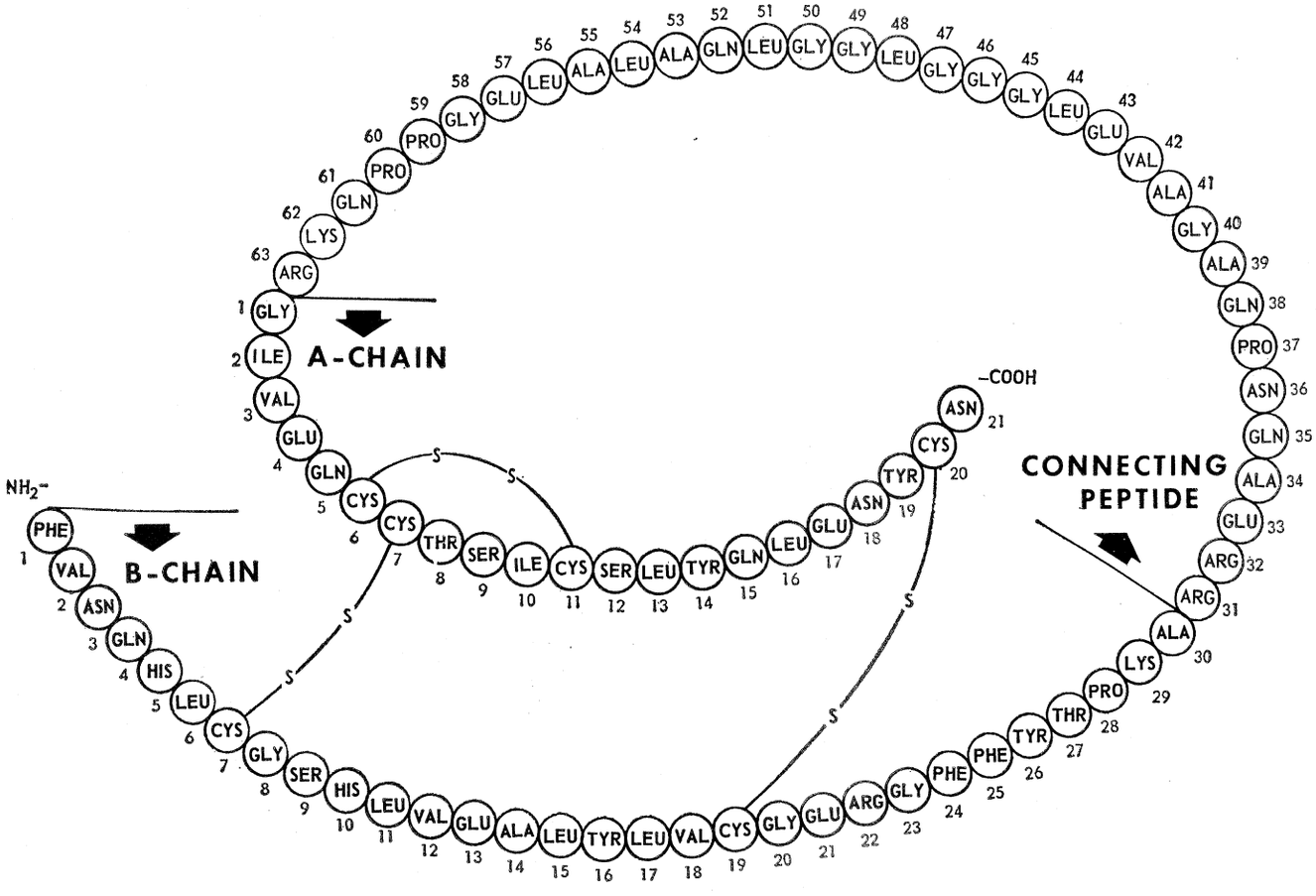


Fig. 2. Proposed primary structure of porcine proinsulin.

an $NH_2$-terminal glutamic acid (called glutamyl peptide), a dipeptide identified as Ala-Arg, and free arginine. Since trypsin cleaved the Lys-Ala ($B_{29}$-$B_{30}$) bond to give desalanine insulin and since the expected molar ratios of alanine were accounted for in the glutamyl peptide, it followed that the peptide Ala-Arg represented positions $B_{30}$-$B_{31}$ (*21*). Two arginine residues were found in a chymotryptic fragment ($B_{27}$-$B_{54}$) isolated from a digest of intact proinsulin by a combination of G-50F Sephadex and Dowex 50-X4 chromatography; hence, the position of the free arginine was established as $B_{32}$. The amino acid sequence (Fig. 2) of the 31-residue glutamyl peptide ($B_{33}$–$B_{63}$) was determined by the combined Edman degradation-dansylation procedure with chymotryptic peptides isolated from a Dowex 50-X4 column eluted with pyridine-acetate buffers. Chymotryptic digests yielded peptides $B_{33}$–$B_{54}$, $B_{55}$–$B_{63}$, $B_{57}$–$B_{63}$, $B_{33}$–$B_{44}$, $B_{45}$–$B_{51}$, $B_{45}$–$B_{48}$, $B_{49}$–$B_{51}$, $B_{52}$–$B_{54}$, and $B_{55}$–$B_{56}$. Carboxypeptidase A digests were used to confirm the Glu-Leu ($B_{43}$–$B_{44}$) and the Leu-Gln-Ala-Leu ($B_{51}$–$B_{54}$) sequences. Treatment of the glutamyl peptide with trypsin and carboxypeptidase B established that the COOH-terminal dipeptide was Lys-Arg ($B_{62}$–$B_{63}$). Thus, the terminal connection between the fragment and the A chain was deduced to be Arg-Gly ($B_{63}$–$A_1$). The complete amino acid sequence of the connecting peptide ($B_{31}$–$B_{63}$) is shown in Fig. 2 as it is incorporated into the proposed primary structure of porcine proinsulin.

The physiologic significance of proinsulin and the mechanism responsible for the transformation of proinsulin to insulin are not known. Our study shows that trypsin, if involved at all, cannot be the sole releasing enzyme since trypsin digestion in vitro leads to the formation of desalanine insulin, not insulin. This finding was to be expected in view of the known specificity requirements for trypsin. Possibly trypsin or a trypsin-like enzyme is involved in the cleavage of the $B_{63}$–$A_1$ bond, whereas another enzyme (or enzymes) may hydrolyze the basic residues around the terminal $B_{30}$–$B_{31}$ connection to give insulin. In support of this possibility, we have in fact, found that the $B_{63}$–$A_1$ bond (Arg-Gly) is hydrolyzed first. This was observed during carefully controlled tryptic digestions in which a significant amount of $NH_2$-terminal glycine was liberated.

Our data along with the work of Steiner and co-workers provide strong evidence that insulin is synthesized as a single-chain precursor which is subsequently converted in a zymogen-like manner as hypothesized (*22*). However, the exact nature and location of such a transformation and the physiologic importance of this precursor concept remain to be determined.

RONALD E. CHANCE
ROBERT M. ELLIS
WILLIAM W. BROMER
*Lilly Research Laboratories, Indianapolis, Indiana 46206*

### References and Notes

1. D. F. Steiner and P. E. Oyer, *Proc. Nat. Acad. Sci. U.S.* **57**, 473 (1967).
2. D. F. Steiner, D. Cunningham, L. Spigelman, B. Aten, *Science* **157**, 697 (1967).
3. The COOH-terminal amino acid of the B chain is alanine in porcine and bovine insulin, and it is threonine in human insulin. The $NH_2$-terminal residue of the A chain is glycine for all species determined [L. F. Smith, *Amer. J. Med.* **40**, 662 (1966)].
4. R. E. Humbel, *Proc. Nat. Acad. Sci. U.S.* **53**, 853 (1965); S. A. Morenkova and M. A. Zaidenberg, *Dokl. Akad. Nauk SSSR* **173**, No. 4-6B, 1218 (1967).
5. Abbreviations: Arg, arginine; Ala, alanine; Asn, asparagine; Cys, half-cystine; Gln, glutamine; Glu, glutamic acid; Gly, glycine; His, histidine; Ile, isoleucine; Lys, lysine; Pro, proline; Phe, phenylalanine; Ser, serine; Tyr, tyrosine; Thr, threonine; Val, valine; DEAE, diethylaminoethyl; tris, tris (hydroxymethyl) amino methane; I.U., international unit; chain position of fragments is designated by letter and subscript numerals.
6. Modification of the method of E. O. P. Thompson and I. J. O'Donnell, *Australian J. Biol. Sci.* **13**, 393 (1960), as used by W. W. Bromer and R. E. Chance, *Biochim. Biophys. Acta* **133**, 219 (1967).
7. Ultracentrifugal studies are under investigation by Dr. B. H. Frank of this laboratory.
8. C. C. Yip and B. J. Lin, *Biochem. Biophys. Res. Commun.* **29**, 382 (1967).
9. *Coupling*: S. Eriksson and J. Sjöquist, *Biochim. Biophys. Acta* **45**, 290 (1960); *Cleavage:* B. Blombäck, M. Blombäck, P. Edman, B. Hessel, *ibid.* **115**, 371 (1966); *Conversion*: D. Ilse and P. Edman, *Australian J. Chem.* **16**, 411 (1963).
10. "Dansyl" is an acronym for 1-dimethylaminonaphthalene-5-sulfonyl; the method is described by W. R. Gray and B. S. Hartley, *Biochem. J.* **89**, 379 (1963).
11. H. Brown, F. Sanger, R. Kitai, *ibid.* **60**, 556 (1955).
12. Y. Kawanishi, K. Iwai, T. Ando, *J. Biochem.* **56**, 314 (1964).
13. Beckman Microzone cellulose acetate electrophoresis, with a 0.03*M* phosphate–7*M* urea buffer at *p*H 6.5.
14. J. L. Bailey and R. D. Cole, *J. Biol. Chem.* **234**, 1733 (1959).
15. Disulfide bonds reduced with dithiothreitol. Newly formed sulfhydryls reacted with recrystallized iodoacetic acid to form the corresponding carboxymethyl derivative.
16. K. Culhane, H. P. Marks, D. A. Scott, J. W. Trevan, *Biochem. J.* **23**, 397 (1929).
17. Carried out by Dr. W. O. Storvick of this laboratory according to the method of G. W. Probst, W. F. Brown, H. J. Henry, *J. Pharm. Sci.* **55**, 1408 (1966).
18. L(1-tosylamido-2-phenyl) ethyl chloromethyl ketone (TPCK), V. Kostka and F. H. Carpenter, *J. Biol Chem.* **239**, 1799 (1964).
19. Enzyme : substrate ratio was 1:250; 3 hours at 37°C in 0.05*M* tris-HCl, 0.02*M* $CaCl_2$, *p*H 7.5.
20. Desalanine insulin refers to insulin with the COOH-terminal alanine removed from the B chain. Since desalanine insulin and insulin are indistinguishable by electrophoresis, chromatography, bioactivity, or immunochemical specificity, only amino acid analysis will distinguish between the two molecules (Table 1).
21. The $NH_2$-terminal residue of the connecting peptide is numbered $B_{31}$ in a continuation of the 30-residue B chain sequence.
22. D. Givol, F. De Lorenzo, R. F. Goldberger, C. B. Anfinsen, *Proc. Nat. Acad. Sci. U.S.* **53**, 676 (1965).
23. Amino acid analyses carried out under the supervision of E. E. Logsdon of this laboratory.
24. B. J. Davis, *Ann. N.Y. Acad. Sci.* **121**, 404 (1964).
25. A. Chrambach, R. A. Reisfeld, M. Wyckoff, J. Zaccari, *Anal. Biochem.* **20**, 150 (1967).
26. F. H. Carpenter and S. L. Hayes, *Biochemistry* **2**, 1272 (1963).
27. We thank M. Boucher, J. Patterson, R. P. Perkins, G. Schertz for technical assistance.

25 March 1968

## Abnormal Water Balance in a Mutant Strain of Chickens

Abstract. *Polydipsia and polyuria are pronounced in chickens of a selected strain and this diabetes insipidus is inherited. The kidneys of such birds are capable of an antidiuretic response when lysine vasopressin or arginine vasotocin is injected. Osmotic pressure and sodium concentration of the plasmas of normal and mutant chickens are identical. Chicks predicted to have diabetes insipidus on the basis of parental pedigree are polydipsic.*

Excessive drinking (polydipsia) and urinary output (polyuria) are present in many commercial chicken flocks. These conditions have been recognized among birds placed in individual cages for various biological studies and for commercial egg production. Watery droppings were discovered in a strain of chickens maintained at Pennsylvania State University in 1957 (*1*). During 5 years this condition was associated with polydipsia and was characteristic of a few pedigreed female birds in only one of two strains which were hatched and reared in the same way. Results of extensive breeding experiments started in 1963 support the idea that polydipsia and polyuria in this strain of chickens are inherited (*2*). The condition can thus be characterized as hereditary diabetes insipidus (DI). For experimental purposes, two inbred strains of white Leghorn chickens were developed, one normal and the other possessing DI. As judged from the amount of water drunk or from the ratio of the amount of water drunk to the amount of food eaten, inbreeding of the DI line since 1963 has resulted in no increase in the severity of the disease. Similar measures of the normal line have also remained fairly constant. Progeny ($F_1$) of reciprocal crosses of the two lines in 1966 were normal. Segregation of normal and DI traits