Determination of the Amino Acid Sequence of the Monkey, Sheep, and Dog Proinsulin C-Peptides by a Semi-micro Edman Degradation Procedure*

James D. Peterson, Soojah Nehrlich, Philip E. Oyer, and Donald F. Steiner

SUMMARY

Proinsulin C-peptides were isolated from the pancreas of the green monkey, sheep, and dog by established procedures. The amino acid sequences of these peptides have been determined using a semi-micro adaptation of the Edman degradation procedure. These sequences are compared with those of the human, bovine, porcine, and rat C-peptides. The bovine and ovine C-peptide sequences are identical, whereas the dog sequence appears to contain an 8 residue deletion near the NH2 terminus. Although the C-peptide sequences vary much more than the insulin A and B chain sequences among these species, several prominent features of the C-peptide structure have been conserved.

A major limitation of our previous sequence studies has been the large amount of C-peptide or proinsulin required to determine the interchain connecting peptide sequence. This has been due largely to cumulative losses due to the relatively low yields obtained at each successive step of the Edman degradation procedure employed, necessitating the sequencing of a large number of smaller peptide fragments. To establish the sequence of the human C-peptide, for example, 8 mg (2.7 μmoles) of material were required. Edman degradation of 1 mg (0.33 μmole) of intact human C-peptide yielded identifiable PTH derivatives only for the first 8 residues, and it was necessary to prepare and sequence a total of 15 peptide fragments from different enzymic digests to complete the structural determination. Essentially similar procedures have been used by other workers in this field (6-8). In order to utilize for sequence studies the limited amounts of pancreas available in certain species or from individual human patients where the C-peptide content may be less than 0.1 μmole per pancreas due to postmortem autolytic losses or pathological processes, a more efficient sequencing method was clearly desirable.

In this report, we describe a semi-micro adaptation of the Edman degradation procedure which gives high repetitive yields (approximately 95%) of the PTH derivatives, and thus allows a much larger number of residues to be successfully removed and identified before the yield of the PTH derivative drops to a prohibitively low level; an additional benefit is that it requires about half as much time per residue. The sequences of green monkey, canine, and ovine C-peptides, isolated from pancreas by established methods, have been determined by this procedure. These sequences are compared to the previously reported human, bovine (9), porcine (10), and rat (8) sequences with regard to their composition, general structure, and sequence homology.

MATERIALS AND METHODS

Fresh frozen lamb pancreas was obtained from the Chicago Lamb Packing Company, Chicago, Illinois. Frozen green monkey (Cercopithecus aethops) pancreas was a gift from the Novo Company, Copenhagen, Denmark. Fresh dog pancreas was obtained from the Medical Physiology Laboratory of the University of Chicago. Sequential grade N,N-dimethyl-N-allylamine and n-propyl alcohol were purchased from Pierce Chemical Company, Rockford, Illinois. Analytical reagent grade ethyl

* This work was supported by United States Public Health Service Grant AM-13994.
† Predoctoral Trainee, United States Public Health Service Training Grant NIH GM 424-12 Training.
§ Recipient of a United States Public Health Service Research Career Development Award.

1 The abbreviation used is: PTH, phenylthiohydantoin.
A procedure, which is a modification of previously described methods (10-12). Phenylisothiocyanate (5 μl) was added under a stream of N₂, stoppered 12-ml centrifuge tube; this was evacuated with a vacuum pump, mixed briefly, and then repressurized with N₂. 

A solution of dimethylallylamine in 1-propanol-H₂O (3:2 by volume), was extracted from this film (carefully, so as not to remove any solid peptide) with about 1 ml of peroxide-free ethyl ether containing 10⁻⁴ M dithiothreitol. The peptide, minus its NH₂-terminal residue, was then recycled. 

The conversion and identification of PTH derivatives—The ether extract, containing the thiazolinone, was dried under N₂, and conversion was carried out as described by Edman and Degg (13). PTH derivatives were identified by gas chromatography (14) on a Beckman model GC-45 or an F & M model 402 gas chromatograph, or by thin layer chromatography (15) using plates (6.7 x 10 cm) of Eastman 6000 silica gel sheets with solvent front migration distances of 5 to 6 cm for Solvent V and 2 to 3 cm for Solvent IV. 

Precautions—The ether used for extraction after cleavage was tested for peroxides before use by mixing a few milliliters of ether with an equal volume of 4% (w/v) aqueous potassium iodide for 5 to 10 min. The appearance of any noticeable yellow color in this time indicates the presence of peroxides. It is essential that this test be performed on every lot of ether used, regardless of its source; ether which gives a positive potassium iodide test will quickly ruin a degradation. 

All reagents and solvents used in the coupling and conversion steps (except the ether) were stored under N₂.

Isolation of C-Peptides from Pancreas

Isolation of monkey C-peptide from 500 g of monkey pancreas and of ovine C-peptide from 500 g of lamb pancreas was carried out essentially as described previously (2) except that the partition chromatography step was eliminated in the preparation of lamb C-peptide. For the dog pancreas (460 g), this procedure was abbreviated somewhat. The Rio-Gel P-30, carbosymethyl cellulose, and formylated cellulose column chromatography steps were eliminated, and the insulin-containing fractions from the Sephadex G-50 column were combined, lyophilized, and submitted to Edman degradation. After 22 steps, the material was than recycled. 

Accordingly, the reaction mixture, after 23 steps of Edman degradation, was treated with dimethylaminonaphthalene-5-sulfonyl chloride (dansyl chloride) (17) to irreversibly block any remaining amino groups. This material was then taken up in 100 μl of 0.1 M pyridine (pH 8.8), and 4 μg of α-chymotrypsin were added. After incubation for 2 hours at 37°, the reaction min. This mixture was then extracted once with about 500 μl of benzene (added, centrifuged, and removed under N₂), and the extract was discarded. The coupled peptide was dried, first at 50° under a stream of N₂, and then by evacuating at 50° with a vacuum pump. 

Concentration—About 100 μl of trifluoroacetic acid were added under N₂, and the mixture was incubated at 50° for 7 min. The trifluoroacetic acid was then evaporated with a stream of N₂, spreading the material in an open film over an area of 1 to 2 cm² at the bottom of the tube. The thiazolinone derivative was extracted from this film (carefully, so as not to remove any solid peptide) with about 1 ml of peroxide-free ethyl ether containing 10⁻⁴ M dithiothreitol. The peptide, minus its NH₂-terminal residue, was then recycled.

Conversion and Identification of PTH Derivatives—The ether extract, containing the thiazolinone, was dried under N₂, and conversion was carried out as described by Edman and Degg (13). PTH derivatives were identified by gas chromatography (14) on a Beckman model GC-45 or an F & M model 402 gas chromatograph, or by thin layer chromatography (15) using plates (6.2 x 10 cm) of Eastman 6000 silica gel sheets with solvent front migration distances of 5 to 6 cm for Solvent V and 2 to 3 cm for Solvent IV. 

Precautions—The ether used for extraction after cleavage was tested for peroxides before use by mixing a few milliliters of ether with an equal volume of 4% (w/v) aqueous potassium iodide for 5 to 10 min. The appearance of any noticeable yellow color in this time indicates the presence of peroxides. It is essential that this test be performed on every lot of ether used, regardless of its source; ether which gives a positive potassium iodide test will quickly ruin a degradation. 

All reagents and solvents used in the coupling and conversion steps (except the ether) were stored under N₂.

Amino acid compositions of dog, sheep, and monkey C-peptides*

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Dog</th>
<th>Sheep</th>
<th>Monkey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proline</td>
<td>1.05</td>
<td>2</td>
<td>2.06</td>
</tr>
<tr>
<td>Glycine</td>
<td>4.95</td>
<td>7</td>
<td>6.60</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Leucine</td>
<td>4.75</td>
<td>5</td>
<td>4.72</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lysine</td>
<td>0.88</td>
<td>1</td>
<td>1.80</td>
</tr>
<tr>
<td>Cysteine</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Valine</td>
<td>0.30</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Methionine</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td>23</td>
<td>23</td>
</tr>
</tbody>
</table>

* All values are expressed as residues per molecule; 24-hour hydrolysis in 6 N HCl at 110°.

Edman Degradation

Edman degradation was carried out according to the following procedure, which is a modification of previously described methods (10-12).

Coupling—Coupling buffer was prepared by titrating a 0.4 M solution of dimethylylamine in 1-propanol-H₂O (3:2 by volume) to pH 9.5 with trifluoroacetic acid. The peptide was dissolved in approximately 100 μl of coupling buffer in a glass-stopped 12-ml centrifuge tube; this was evacuated with a vacuum pump, mixed briefly, and then repressurized with N₂. Phenylisothiocyanate (5 μl) was added under a stream of N₂, and the tube was mixed thoroughly and incubated at 50° for 20
**LAMB C-PEPTIDE**

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
Glu-Val-Glu-Gly-Pro-Gln-Val-Gly-Ala-Leu-Glu-Leu-Ala-Gly-Gly-Pro-Gly-Ala-Gly-Gly-Leu-Glu-Gly-Pro-Pro-Gln

**INTACT**

P3

**DOG C-PEPTIDE**

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

**INTACT**

**MONKEY C-PEPTIDE**

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

**INTACT**

was stopped by addition of 2 drops of glacial acetic acid. Thin layer chromatography in the butanol-pyridine-acetic acid-water system (3) of 2-111 diquots taken before and after incubation revealed no ninhydrin-positive material before incubation and a single ninhydrin-positive component after incubation. This material (DC; see Fig. 1) without purification was dried and submitted to seven steps of Edman degradation; no significant heterogeneity was observed in the PTII derivative at any of these seven steps. Prior to the seventh step, 50% of the material was applied directly to the long column of the amino acid analyzer, as described above; 6 nmoles of free glutamine were found.

**Fig. 1.** Schematic summary of the amino acid sequence data on lamb, dog, and monkey pancreatic C-peptides and their chymotryptic fragments (see text for methodology). MTH, methylthiohydantoin.

**Lamb C-Peptide**

Intact lamb C-peptide (100 nmoles) was submitted to 25 steps of Edman degradation. The remaining material was then taken up in 100 μl of H2O, and 50 μl of this was applied directly to the long column of the amino acid analyzer; 10 nmoles of free glutamine were observed. The remaining material was dried and submitted to Edman degradation.

In addition, with procedures previously described for the bovine C-peptide (3), 150 nmoles of lamb C-peptide were digested with α-chymotrypsin; a two-dimensional peptide fingerprint, which was identical with that of the bovine C-peptide (3), was made with 10% of this digest, and the remainder was purified by thin layer electrophoresis (3). Three peptide fragments, corresponding to bovine C-peptide chymotryptic peptides P-1, P-2, and P-3 (3), were submitted to Edman degradation. With 15 nmoles of lamb P-3, identifiable PTH derivatives were obtained for the first 7 residues. Thirty nanomoles of lamb P-1 and 15 nmoles of lamb P-2 were sequenced for three and five steps, respectively, and then were combined (since it was clear by this point that P-2 was identical with P-1 except for the presence of 2 additional residues at its NH2 terminus) and sequenced to their COOH terminus. Prior to the final step, 50% of the free COOH-terminal residue was applied to the amino acid analyzer as described above; 1.5 nmoles of glutamine were observed.

**Monkey C-Peptide**

Approximately 100 nmoles of intact monkey C-peptide were submitted to Edman degradation, which proceeded well until position 10. At this point, methylisothiocyanate (18) was substituted for phenylisothiocyanate in the degradation procedure. The new reagent, unfortunately, was of poor quality; by position 24, the yield of the methylthiohydantoin derivative had decreased substantially and the carryover of residues from previous positions had markedly increased. Although phenylisothiocyanate was reinstated at position 24, residue 28 was the last residue that could be unambiguously identified. The last 3 residues of the peptide had already been sequenced with carboxypeptidase A, but it was still necessary to show that these 3 residues occupy...
FIG. 2. Alignment of the known proinsulin C-peptide sequences. Ex-cept where substitutions or deletions are indicated, each sequence is identical with the human sequence at each position. (-), deletion. With four exceptions, all substitutions at each position represent single nucleotide base changes; the Asn-Gly substitution at position 4, the Gly-Pro substitution at position 8, positions 29 to 31 of the intact peptide. This was possible since the monkey C-peptide, according to the data already accumulated, contained the same chymotrypsin-sensitive Leu-Ala bond as the human, porcine, and dog C-peptides. The COOH-terminal heptapeptide of the monkey C-peptide was liberated, as described under "Dog C-Peptide," by dansylation and chymotryptic digestion of the material remaining after 29 steps of Edman degradation. This material (MC; see Fig. 1) was then submitted to seven steps of Edman degradation.

**RESULTS AND DISCUSSION**

*Edman Degradation Procedure*—The most important limiting factor of the Edman procedure in our peptide studies has been the gradual decline in the yield of PTH derivatives as the degradation progresses. Edman and Begg, in their discussion of the factors which bring about the termination of a degradation (13), suggest two possible causes for this: oxidative desulfuration of the phenylthiocarbanil peptide and inadvertent extraction of some of the peptide after coupling or cleavage. The new procedure described here is designed to minimize oxidative desulfuration by reducing the amount of residual O2 in the coupling and cleavage media and by reducing the time the peptide is exposed to oxidative conditions. Extraction of the peptide is minimized by using a volatile coupling buffer (so that the phenylisothiocyanate, which is adequately removed by a single extraction, is the only component which must be extracted after coupling) and by using ether instead of 1-chlorobutane or 1,2-dichloroethane (13) to extract the thiazolinone. It has been found that other extracts substantially less peptide (determined by hydrolysis and amino acid analysis of the material remaining in the aqueous phase after the ethyl acetate extraction of the conversion step) and much more thiazolinone (as indicated by the yield of the PTH derivative) than an equal volume of 1-chlorobutane. Ether tends to suspend small flakes of solid peptide more readily than the other solvents, but these can be avoided when the ether is removed by pouring (rather than pipetting) it slowly from the tube which allows any free particles to settle near the bottom.

Using this new procedure, we have routinely sequenced all the way through peptides of 23 to 31 residues (dog, human, bovine, and lamb C-peptides), using 100 nmoles of peptide in each case (except 200 nmoles of human). Even when only 10 nmoles of lamb C-peptide were sequenced, PTH derivatives could be identified by thin layer chromatography through the first 10 positions. Since C-peptides are of rather restricted composition, it is possible that this procedure will not work as well with peptides of different compositions. Russell and Heinrickson, however, have successfully used this procedure to sequence a wide variety of tryptic and chymotryptic peptides from rhodanese, using 100 to 150 nmoles of starting material in each case; these peptides ranged in size from 2 to 20 residues and were not restricted in composition. The procedure has also been applied, with some success, to small amounts of intact proteins of molecular weight of about 24,000 (19).

**C-Peptide Sequences**—The sequence and a summary of the sequence data for each C-peptide studied are shown in Fig. 1, and a comparison of the eight C-peptide sequences now known

---

1. J. D. Peterson, unpublished data.
is given in Fig. 2. It can be seen that these C-peptide sequences are much more variable than the corresponding insulin A and B chain sequences (20). In amino acid composition, however, the C-peptides are remarkably similar; each contains 58 to 77% nonpolar residues, 4 or 5 acidic residues, and a few (2 to 6) polar residues. They contain no aromatic residues, no basic residues (except the rat peptide), and no cystine, methionine, or isoleucine.

In addition, the distributions of the acidic and polar side chains along the C-peptide molecules are quite similar, so that the general structures of these molecules are nearly identical (Fig. 3). It is noteworthy that this general structure is maintained even in the dog C-peptide, where 8 of the first 11 positions are not present, and in the rat C-peptides, which contain extra acidic and basic residues; thus a glutamic acid residue appears at position 10 of the dog C-peptide, while the homologous C-peptides (except rat) contain, in this region, only residues with uncharged side chains (cf. positions 12 to 26 in Fig. 2). The significance of this structural conservatism is not known. It is possible that the C-peptide interacts in some specific way with the rest of the proinsulin molecule, for example by helping to align the A and B chains and generate the correct disulfide bonds, during biosynthesis. (It should be noted, in this connection, that the extra acidic and basic residues in the rat C-peptide can be correlated with unusual features of the structure of the two rat insulins (20). Thus, both rat insulins have lysine, rather than asparagine, at position B-3, and rat insulin II, whose corresponding C-peptide lacks the glutamic acid at position 1/4, has a methionine for lysine substitution at position B-29. This correlation suggests the possibility of specific electrostatic interaction between the C-peptide and insulin regions of the proinsulin molecule. Furthermore, these substitutions tend to conserve the balance of positively and negatively charged side chains so that the two rat proinsulins have isoelectric pH values nearly identical with each other and with other proinsulins.) Another possibility is that the enzyme or enzyme system responsible for the in vivo conversion of proinsulin to insulin requires some of these general features of the C-peptide structure to direct its proteolytic attack. Since the C-peptide is secreted along with insulin into the serum (5, 21), it is also possible that some of these structural features are required for extrapancreatic functions of the C-peptide although no such functions are presently known.

It should be noted that there is no direct evidence that the peptides studied here represent the entire C-peptide regions of the proinsulins in question. It is possible that in some species the in vivo conversion mechanism splits the C-peptide into two or more fragments, or that autolytic cleavage has occurred at some sensitive bond in the C-peptide prior to, or during, extraction and purification. It is unlikely that this has occurred in the case of the lamb and monkey C-peptides because of the close homology between these and the bovine and human (respectively) C peptides, each of which has been shown to represent its entire proinsulin C-peptide segment (3, 4). In the case of the dog C-peptide, however, only a tentative conclusion that the entire C-peptide has been isolated is justified, based on the evidence of the general structure of the molecule (Fig. 3). In all three cases, final proof that these sequences represent the entire C-peptides must await the isolation of sufficient amounts of the proinsulins for compositional and structural studies.

Acknowledgments—We are indebted to William Montgomery for his assistance in the preparation of the dog and lamb C-peptides and to Roberta Erfurth for her assistance in the preparation of the manuscript.

REFERENCES

Determination of the Amino Acid Sequence of the Monkey, Sheep, and Dog Proinsulin C-Peptides by a Semi-micro Edman Degradation Procedure
James D. Peterson, Sooja Nehrlich, Philip E. Oyer and Donald F. Steiner


Access the most updated version of this article at http://www.jbc.org/content/247/15/4866

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/247/15/4866.full.html#ref-list-1