ZOOLOGICAL SCIENCE **15**: 931–937 (1998)

Two Molecular Forms of Insulin from Barfin Flounder, *Verasper moseri*, are Derived from a Single Gene

Tadashi Andoh[1]* and Hiromichi Nagasawa[2]

[1]*Hokkaido National Fisheries Research Institute, 116 Katsurakoi, Kushiro 085-0802, Japan*
[2]*Department of Applied Biochemistry, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan*

ABSTRACT—We have purified and characterized two molecular forms of insulin from the Brockmann bodies of barfin flounder, *Verasper moseri*: a normal type of insulin (insulin-I), which consisted of 21 amino acid residues for the A-chain and 30 residues for the B-chain, and a novel type of insulin (insulin-II), which had an extension of two amino acid residues at the N-terminus of the B-chain. The additional two residues at the N-terminus of B-chain of insulin-II were Pyr-Ala which had not yet been reported in vertebrate insulins. Except for these two residues, the amino acid sequence of insulin-II was completely identical with that of insulin-I. Each Brockmann body extract from five individuals contained both insulins, indicating that insulin-I and -II were the products of non-allelic expression. By polymerase chain reaction, only one nucleotide sequence of preproinsulin gene encoding insulin-I and -II was obtained, and the amino acid sequence of A- and B-chains deduced from the nucleotide sequence was identical with that of insulin-I and also insulin-II established by Edman degradation. Furthermore, genomic Southern blot analysis using a part of nucleotide sequence of the preproinsulin gene as a probe showed a single positive band in all cases of genomic DNA digested with each of three restriction endonucleases. These results indicate that insulin-I and -II of barfin flounder arise from a single preproinsulin by proteolytic cleavage at different sites of the signal peptide region.

INTRODUCTION

Insulin is a key hormone for anabolism in fish. It stimulates the uptake of glucose and amino acids by the skeletal muscle and liver, and increases the rate of protein synthesis in these tissues (Mommsen and Plisetskaya, 1991). It also shows growth-promoting effect in fish. Duan and Hirano (1992) reported that insulin at physiological concentrations may directly exert growth-promoting actions on the branchial cartilage of the eel. Others have also reported the possibility that insulin accelerated growth in fish (Ludwig *et al*., 1977; Ablett *et al*., 1981; Plisetskaya *et al*., 1991; Sundby *et al*., 1991). Furthermore, it is thought that insulin is involved in smoltification (Plisetskaya *et al*., 1988) and spawning migration in salmon (Plisetskaya *et al*., 1987). Thus, insulin is a multi-functional hormone that regulates several physiological phenomena in fish.

For the last three decades, insulins have been purified and characterized from over 60 species of vertebrates. Most of these species contain a single molecular form of insulin. However, it was reported that some species of fish, such as bonito (*Katsuwonus pelamis*; Yamamoto *et al*., 1960; Kotaki *et al*., 1962), toadfish (Smith, 1966), Pacific ratfish (*Hydrolagus colliei*; Conlon *et al*., 1989) and paddlefish (*Polyodon spathula*; Nguyen *et al*., 1994), contain multiple molecular forms of insulin.

In the present study, we examined by purification and amino acid sequence determination of insulins whether a species of flounder, *Verasper moseri* (barfin flounder), contained multiple molecular forms of insulin or not. In addition, we analyzed nucleotide sequence of a preproinsulin gene.

MATERIALS AND METHODS

Extraction and purification of flounder insulins

Insulins were extracted essentially according to Conlon *et al*. (1987) with slight modification. The Brockmann bodies (550 mg) on the gall bladder were taken from cultured barfin flounder (body weight: 4-6 kg; N = 4) and stored at −40°C until extraction. The frozen Brockmann bodies were minced by scissors and homogenized in 10 ml of ethanol/0.7 M HCl (3:1, v/v) using a Teflon-glass homogenizer. The homogenate was stirred overnight at 4°C and centrifuged at 10,000 × g for 20 min at 4°C. The ethanol was removed from the supernatant by vacuum centrifugation. The resulting precipitate was removed by centrifugation at 10,000 × g for 10 min at 4°C. The supernatant was taken out, added with trifluoroacetic acid (TFA) to the final concentration of 0.1%, and applied to two Sep-Pak C18 cartridges (Millipore, Milford, MA, USA) connected in tandem. After washing with 0.1% TFA, the cartridges were eluted with 80% acetonitrile in 0.1% TFA. This eluate was concentrated to 500 µl by vacuum centrifugation and added with TFA to the final concentration of 0.1%. The resulting solution was injected onto an ODS-120T column (2.0 × 150

* Corresponding author: Tel. +81-154-91-9136; FAX. +81-154-91-9355.

mm, Tosoh, Tokyo, Japan), and eluted with a linear gradient formed from water/TFA (100:0.1, v/v) to acetonitrile/water/TFA (50:50:0.1, v/v) at a flow rate of 100 μl/min for a 115-min and at column temperature of 30°C. The collected fractions were subjected to mass spectral analysis using a Kompact MALDI III time-of-flight mass spectrometer (Shimadzu, Kyoto, Japan) with 2-(4-hydroxyphenyl-azo) benzoic acid as the matrix in the positive ion and linear modes. The accuracy of this machine is specified to be 0.2%. The fractions which showed a prominent peak around *m*/*z* 6000 were dried by vacuum centrifugation, and were separately injected onto a μ RPC C2/C18 SC column (2.1 × 100 mm, SMART system, Pharmacia, Uppsala, Sweden) under the same conditions as in the previous HPLC except for the acetonitrile concentration which was increased from 0 to 22.5% (v/v) over 5 min, to 42.5% (v/v) over 70 min and to 50% (v/v) over 10 min. Absorbance was measured at 280 nm.

Enzyme immunoassay

Insulin-like immunoreactivity of the fractions from the chromatographies were measured by an enzyme immunoassay (EIA) system using anti-bovine insulin antiserum and horseradish peroxidase-bovine insulin complex as labeled hormone (Andoh and Nagasawa, 1997).

Structural analysis

The peptides in the fractions which showed a prominent peak around *m*/*z* 6000 in the mass spectrum were reduced and carboxymethylated. The carboxymethylated derivatives were separated according to Andoh and Nagasawa (1997). Amino acid sequences were determined by automated Edman degradation using a protein sequencer, Applied Biosystems 476A (Perkin Elmer, Norwalk, CT, USA) in the pulsed-liquid mode.

Digestion with pyroglutamate aminopeptidase

The eluate from HPLC containing S-carboxymethylated B-chain was concentrated to 10 μl by vacuum centrifugation and incubated at 50°C for 48 hr with 1 mU of *Pfu* pyroglutamate aminopeptidase (Takara, Otsu, Japan) in 100 μl of 50 mM sodium phosphate buffer (pH 7.0) containing 1 mM EDTA and 10 mM DTT. To stop the digestion, TFA was added to the final concentration of 1% (v/v) and the solution was injected onto a μ RPC C2/C18 SC column. The elution condition was the same as those in the first HPLC step except for the gradient program. Elution was carried out with a 100-min linear gradient.

Extraction and purification of insulins from a single individual

Brockmann bodies (51-124 mg, N = 5) was used and extracted individually essentially according to Cutfield *et al*. (1986). The precipitates of the extracts by diethylether-ethanol treatment were air-dried and dissolved in 3 M acetic acid. After centrifugation at 10,000 × g for 10 min at room temperature, each supernatant was injected onto an ODS-120T column (2.0 × 150 mm, Tosoh). The elution conditions for HPLC were the same as those in the first HPLC step. The materials eluted near the retention time of the fraction collected at the first HPLC step were collected peak by peak manually by monitoring absorbance at 280 nm. An aliquot of each fraction was subjected to N-terminal amino acid sequence analysis up to the fifth residue. Mass spectrum of each fraction was also measured.

Nucleotide sequence analysis of preproinsulin gene from genomic DNA

Genomic DNA was extracted from the liver (0.1 g) of barfin flounder using Catrimox-14 (Iowa Biotechnology Corp., Oakdate, IA, USA) or GenTLEkun (Takara), according to the manufacturally protocol. Three steps of polymerase chain reaction (PCR) were performed for amplification of the preproinsulin gene(s). The first PCR was performed using genomic DNA and two degenerate oligonucleotide primers, 5'-CGCCCTCTATCTGGT(CG)TG-3' and 5'-GCACTGCTCTAC(AG)AT(AG)CC-3', which correspond to the conserved internal region of B-chain and N-terminal region of A-chain of preproinsulin of fish. The PCR was carried out on a DNA thermal cycler, model 9600 (Perkin Elmer) for 30 cycles; each cycle consists of reactions at 94°C for 30 sec, 50°C for 30 sec, and 72°C for 60 sec. The amplified PCR product was separated by electrophoresis on 1% agarose gel. A band was excised, purified, and cloned into the pT7Blue T-Vector (Novagen Inc., Madison, WI, USA). The nucleotide sequences of both strands of each of four clones selected randomly were analyzed by the dideoxynucleotide chain termination method (Sanger *et al*., 1977).

The second PCR was performed using LA PCR *in vitro* cloning kit (Takara) according to the manufacturally protocol. This kit is designed for amplification of 5'-upstream region or 3'-downstream region of already-known sequences by nested PCR using two sets of primers from genomic DNA. *Eco*RI and *Xba*I were used in this experiment. Degenerate oligonucleotide primers which corresponded to the C-terminal region of B-chain, the N-terminal region of C-peptide and the region connecting C-peptide and A-chain were designed from the sequence established by the first PCR. In amplification of 3'-downstream region of A-chain, 5'-GAGACGTGGACCCTCTGCTG-3' and C1 primer from the kit , and 5'-GATGATGGTGAAGCGAGGCA-3' and C2 primer also from the kit, were used in primary and secondary PCR, respectively. The primers for amplification of 5'-upstream region of B-chain were 5'-TGCCTCGCTTCACCATCATC-3' and C1 primer for the primary PCR, and 5'-TTGGGGGTGTAGAAGAAGCC-3' and C2 primer for the secondary PCR. Each PCR was carried out for 30 cycles each consisting of reactions at 94°C for 30 sec, 55°C for 120 sec, and 72°C for 60 sec. Nucleotide sequence analysis was carried out by the same method as described above.

The third PCR was performed using degenerate oligonucleotide primers which corresponded to 5'-upstream region of signal peptide including the initiation codon (ATG) and C-terminal region of A-chain established by the second PCR. The primers used were 5'-TTTCAGTTTTCCTCCAAGATG-3' and 5'-TCAGTTGCAGTAGTTC-TGCA-3'. The PCR conditions were the same as those of the second PCR.

Nucleotide sequence analysis of preproinsulin cDNA

Total RNA was extracted from the Brockmann body (20 mg) of barfin flounder by using a kit (RNeasy; Quiagen, Hilden, Germany), and was transcribed into cDNA by a reverse transcriptase (M-MLV reverse transcriptase; Promega Inc., Madison, WI, USA) at 42°C for 1 hr. Nested PCR was performed by using the obtained cDNA and two sets of degenerate oligonucleotide primers. The primers for the primary PCR were 5'-ATGGCGGCGCTGTGGCTCCAGTCT-3' corresponding to the N-terminus of signal peptide and 5'-CTAACGGAC-TCCGGGCTGAGCGGC-3' corresponding to 3'-downstream region of A-chain. The primers for the secondary PCR were 5'-TCTCTGCT-CGTCCTAATGCTCGTG-3' corresponding to the central region of signal peptide and 5'-TCAGTTGCAGTAGTTCTGCAGGTC-3' corresponding to the C-terminus of A-chain. The PCR conditions were the same as those in the second PCR step for genomic DNA. Nucleotide sequence analysis was carried out by the same method as above except for a cloning vector. The pGEM T-Vector (Promega Inc.) was used in this analysis.

Southern blot analysis

Genomic DNA (10 μg) was digested separately with *Bam*HI, *Eco*RI and *Hin*dIII, electrophoresed in a 0.8% agarose gel, and blotted onto a nylon membrane (Qiabrane; Qiagen). As a probe for Southern hybridization, a part of genomic DNA encoding preproinsulin was amplified by nested PCR as described in preproinsulin cDNA sequence analysis, and labeled with alkaline phosphatase by a labeling kit (AlkPhos Direct; Amersham, Buckinghamshire, UK). The membrane was hybridized with this probe in AlkPhos Direct hybridization buffer containing 0.5 M NaCl and 4% Blocking reagent (Amersham) at 55°C overnight. The membrane was washed twice in 50 mM phosphate

buffer (pH 7.0) containing 2 M urea, 0.1% SDS, 150 mM NaCl, 10mM $MgCl_2$ and 0.2% Blocking reagent at 60°C for 10 min, and once in 1 M Tris-HCl (pH 10.0) containing 2 M NaCl at room temperature for 5 min. Hybridization signals were detected by using chemiluminescence detection solution (CDP-star; Amersham) and X-ray film (Hyperfilm-ECL; Amersham).

RESULTS

Purification of insulins

The elution profile of the Brockmann body extract from barfin flounder is shown in Fig. 1. The fraction that elutes at 77-79 min showed insulin-like immunoreactivity by the EIA. This fraction was injected onto a μ RPC C2/C18 SC column for further purification (Fig. 2). Two major peaks (material-I and -II) were obtained, and these were found to be insulins of barfin flounder. The final yields of both materials from 550 mg of the Brockmann bodies were estimated to be 5 nmol porcine insulin equivalents based on its absorbance (1.06) at 276 nm at the concentration of 1 mg/ml water for convenience.

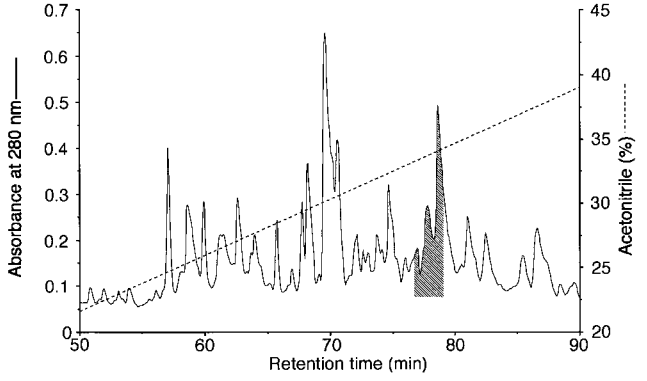

Fig. 1. Fractionation of an extract from barfin flounder Brockmann bodies by HPLC. An extract after partial purification on Sep-pak cartridges was chromatographed on an ODS-120T column. The fraction of the shaded area was subjected to further purification.

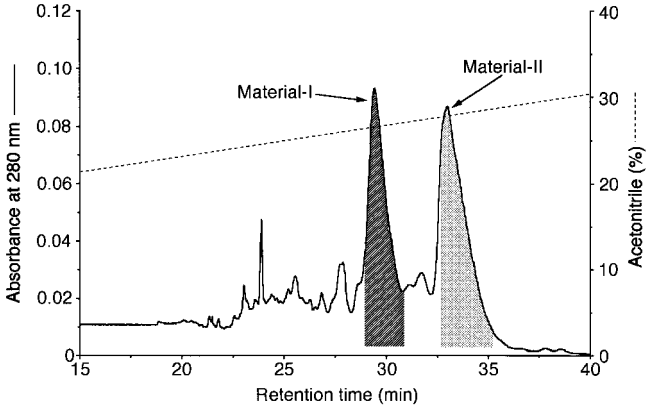

Fig. 2. Purification of candidates of barfin flounder insulins by HPLC. Material-I and material-II from the first HPLC step were purified to homogeneity on a μ RPC C2/C18 SC column. The shaded areas of each material were subjected to amino acid sequence analysis after reductive carboxymethylation.

After reductive carboxymethylation, putative A- and B-chains were obtained from each of the materials by reversed-phase HPLC (profiles not shown).

Characterization of barfin flounder insulins

Automated Edman degradation was successful for both chains of material-I and the putative A-chain of material-II. The amino acid sequences (Fig. 3) were in good agreement with the mass spectral data. Measured and calculated values for $(M + H)^+$ were *m/z* 2672 and *m/z* 2674 for putative A-chain of material-I, *m/z* 3457 and *m/z* 3461 for putative B-chain of material-I, *m/z* 2674 and *m/z* 2674 for putative A-chain of

A-chain

```
                          1    5    10   15   20
Barfin flounder-I(1)      GIVEQCCHKPCNIFDLQNYCN
Barfin flounder-II(1)     ---------------------
Flounder(2)               ---------------------
Anglerfish(3)             --------R------------
Carp(4)                   ----------E----------
Chum salmon(5)            ---------------------
Ray(6)                    ----H---NT-SL---EG---
Hagfish(7)                ---------R-S-YD-E----
Lamprey(8)                --------RK-S-Y-ME----
```

B-chain

```
                          1    5    10   15   20   25   30
Barfin flounder-I(1)        VLPPQHLCGAHLVDALYLVCGERGFFYTPK
Barfin flounder-II(1)     pQA----------------------------
Flounder(2)                 -V----------------------------
Anglerfish(3)               -A-A-----S-----------D-----N--
Carp(4)                   NAGAN-----S---E-------P-----N--
Chum salmon(5)              AAA-----S------------K----N--
Ray(6)                      LPS---------E---F---PK--Y-L--A
Hagfish(7)                  RTTG----KD--N---IA--V-----D-TKM
Lamprey(8)              SALTGAGGT----S---E---V---D-------SKT
```

Signal peptide

```
                          1    5    10   15   20   25
Barfin flounder(1)*       MAALWLQSV SLLVLMLVSWSGSQA
Anglerfish(9)*            --------F -----LV--------
Carp(10)*                 -- V-I-AG A--F-LA--SVNA
Chum salmon(11)*          -- F---AA -----LAL-PGVDA
Hagfish(12)*              --LSPFLAAVIP---L-SRAPP-ADT
```

C-peptide

```
                          1    5    10   15   20   25   30   35
Barfin flounder(1)*       KRDVDPLLGFLPAKSGGAAAGG ENEVAEFAFKDQMEMMVKR
Anglerfish(9)*            -----Q------P--------A-AD-------------------
Carp(10)*                 ------P-----P--AQ       -T---D-----HA-VIR--
Chum salmon(11)*          -------I---SP--AK       ---  -YP----T------
Eel(13)                     --E------SP---Q       ----DD-PY-G-G-L
Hagfish(12)*              ---TGA-AA---LAYAE      DNESQDDESIGINEVLKS--
```

Fig. 3. Comparison of the primary structures of barfin flounder insulins to insulins of other vertebrates. The sequences of A- and B-chains were established by amino acid sequence analysis, while those of signal peptides and C-peptides were deduced from the nucleotide sequence. Pyroglutamic acid residue is represented as pQ. The sequences were cited from (1) the present work; (2) *Platichthys flesus*, Conlon *et al.* (1987); (3) *Lophius americanus*, Neumann *et al.* (1969); (4) *Cyprinus carpio*, Makower *et al.* (1982); (5) *Oncorhynchus keta*, Rusakov *et al.* (1990); (6) *Torpedo marmorata*, Conlon and Thim (1986); (7) *Myxine glutinosa*, Peterson and Steiner (1975); (8) *Lampetra fluviatilis*, Conlon *et al.* (1995); (9) *L. americanus*, Hobart *et al.* (1980); (10) *C. carpio*, Hahn *et al.* (1983); (11) *O. keta*, Koval *et al.* (1989); (12) *M. glutinosa*, Chan *et al.* (1981); and (13) *Anguilla anguilla*, Conlon and Thim (1989). N-Terminal Lys residues of C-peptides in each species except for hagfish are C-terminal residues of B-chain. It is possible that Arg of N-terminus of the eel C-peptides is cleaved by proteolysis. Asterisks indicate amino acid sequences deduced from nucleotide sequences.

material-II, respectively. On the other hand, significant phenylthiohydantoin (PTH) -amino acid was not obtained for the putative B-chain of material-II, suggesting a blocked N-terminus. Pyroglutamate aminopeptidase digestion of this peptide removed the blocked N-terminus, indicating that the N-terminus of the putative B-chain of material-II was a pyroglutamate. Protein sequencing of the deblocked peptide identified 31 amino acid residues (Fig. 3). This was confirmed by comparison of mass spectra of the peptides before and after pyroglutamate aminopeptidase digestion, *m*/*z* 3639 and *m*/*z* 3527, respectively. The difference between undigested and digested peptides was 112, which corresponded to the molecular mass of a Pyr residue. The primary structure of the putative A-chains of these two materials was identical and also identical with that of *P. flesus* insulin. Therefore, we concluded that material-I and material-II were insulins of barfin flounder, and designated as barfin flounder insulin-I and -II, and were compared with insulins from other vertebrates (Fig. 3). The amino acid sequence of B-chain of insulin-II consisted of 32 residues, and is completely identical with that of insulin-I except for an extension of two amino acid residues at the N-terminus.

Production of insulin-I and -II by a single individual

The elution patterns of Brockmann body extracts from five individuals in an ODS-120T column were quite similar to one another (three profiles from five individuals are shown in Fig. 4), and two peaks were observed at nearly the same retention times as those for the barfin flounder insulins. The prominent peaks in the mass spectra of these two fractions were observed at *m*/*z* 5779 and *m*/*z* 5964, whose values were almost equal to the calculated ones, *m*/*z* 5781 and *m*/*z* 5964, of insulin-I and -II, respectively. Furthermore, the N-terminal amino acid sequences of these two peak materials from an individual were analyzed up to the fifth residue without reductive carboxymethylation. The first peak gave a pair of PTH-amino acids, (GV)(IL)(VP)(EP)(Q), and the latter peak gave a single PTH-amino acid, GIVEQ, at each cycle. These indicated that these two materials were insulin-I and -II and that both insulins were produced by a single barfin flounder.

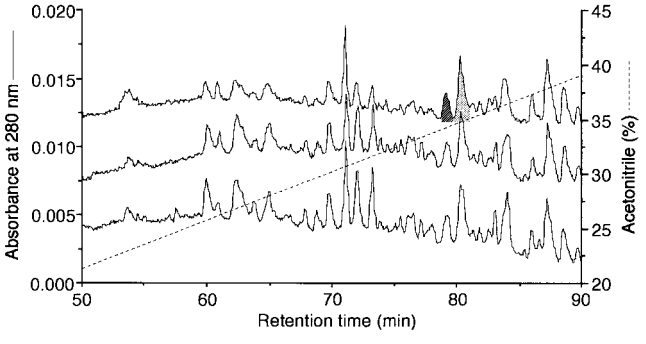

Fig. 4. Profiles of individual extracts from Brockmann bodies by reversed phase HPLC. Brockmann body extracts from each of three individuals of barfin flounder were fractionated on an ODS-120T column. The materials from shaded areas which were eluted near the time of insulin-I and -II were directly subjected to N-terminal amino acid sequence and mass spectral analyses.

Nucleotide sequence of a genomic DNA encoding preproinsulin

Genomic DNA encoding preproinsulin was amplified by three steps of PCR. Positions of the amplified DNA segments in preproinsulin gene are shown in Fig. 5 and the nucleotide sequences determined are shown Fig. 6. Only one product (0.28 kbp) was amplified in the first PCR and the nucleotide sequence contained a region encoding C-terminus of B-chain and C-peptide.

In the second PCR, a nucleotide sequence (0.54 kbp) encoding the signal peptide and the N-terminal region of B-chain was amplified from the genomic DNA digested with *Xba*I.

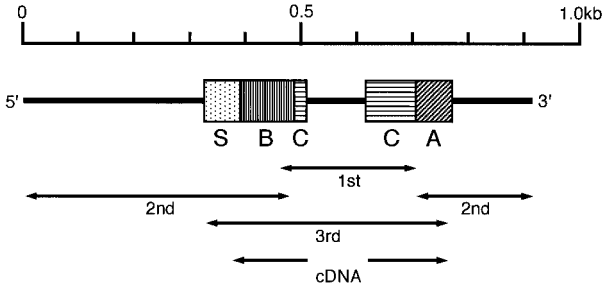

Fig. 5. Structure of a barfin flounder preproinsulin gene and strategy for sequence determination. Coding regions are indicated by the boxes and non-coding regions containing intron are indicated by the thin line. S, B, C and A indicate a signal peptide, B-chain, C-peptide and A-chain, respectively. Arrows with “1st”, “2nd”, “3rd” and “cDNA” indicate amplified segments by each step of PCR.

```
 ac ttc ttc acc tcc tcc tcc ttt ttg tgt gtt tct agt tct ttc cct atg ttc tca ctc    59
cat ttt cat cct ttg ctt ctc ctc gtc ttc agt tat ttt cag atc cgt cat ctc tta agg   119
taa aac gat ttg aaa cct ctt cac tga tgt ctt cat gtc aaa act gaa agc tga ttt agt   179
tta gaa aaa taa tat gta taa aga aag act tca gat gaa gaa tcg gtc tca ttg tcg gac   239
ggt gcc aac att taa aaa atc tat ttt ttc caa gct ttc att taa aaa atg ata aga taa   299

aag cgt gtg ttt cag ttt tcc tcc aag ATG GCG GCG CTG TGG CTC CAG TCT GTC TCT CTG   359
                                     M   A   A   L   W   L   Q   S   V   S   L    11

CTC GTC CTA ATG CTC GTG TCA TGG TCG GGC TCC CAG GCC GTG CTC CCC CCA CAG CAC CTC   419
                          TCA TGG TCG GGC TCC CAG GCC GTG CTC CCC CCA CAG CAC CTC    42
 L   V   L   M   L   V   S   W   S   G   S   Q   A   V   L   P   P   Q   H   L    31

TGT GGC GCT CAC CTG GTC GAC GCC CTC TAC CTG GTC TGT GGC GAG AGA GGC TTC TTC TAC   479
TGT GGC GCT CAC CTG GTC GAC GCC CTC TAC CTG GTC TGT GGC GAG AGA GGC TTC TTC TAC   102
 C   G   A   H   L   V   D   A   L   Y   L   V   C   G   E   R   G   F   F   Y    51

ACC CCC AAG AGA GAC GTG GAC CCT CTG CTG Ggt gag aac atg aca gct tta aac aca gaa   539
ACC CCC AAG AGA GAC GTG GAC CCT CTG CTG G                                       133
 T   P   K   R   D   V   D   P   L   L   G                                      62

cct tga aat gaa cca gaa ctt cct caa cag tca ctt gga ctc ggt ctc atg aga aac tcg   599

tgt ctc tca gGT TTC CTT CCT GCG AAG TCG GGC GGA GCT GCA GCG GGC GGC GAG AAC GAG   659
            GT TTC CTT CCT GCG AAG TCG GGC GGA GCT GCA GCG GGC GGC GAG AAC GAG   183
               F   L   P   A   K   S   G   G   A   A   A   G   G   E   N   E    78

GTG GCC GAG TTC GCC TTC AAG GAC CAG ATG GAG ATG ATG GTG AAG CGA GGC ATC GTG GAG   719
GTG GCC GAG TTC GCC TTC AAG GAC CAG ATG GAG ATG ATG GTG AAG CGA GGC ATC GTG GAG   243
 V   A   E   F   A   F   K   D   Q   M   E   M   M   V   K   R   G   I   V   E    98

CAG TGC TGC CAC AAG CCC TGT AAC ATC TTT GAC CTG CAG AAC TAC TGC AAC TGA acg ccg   779
CAG TGC TGC CAC AAG CCC TGT AAC ATC TTT                                         273
 Q   C   C   H   K   P   C   N   I   F   D   L   Q   N   Y   C   N   *          115

ctc agc ccg gag tcc gtt agc ccc ccg acc cat cac ccc cct cac gcc ccg gcc ctt tgc   839
gtc aga gga acg gca ctg ctg tga aat gga tga aat tat ttt tcc tag aag gtg ggg ttt   899
tgt gaa ttc                                                                    908
```

Fig. 6. Nucleotide sequence of the genomic DNA (regular letters) and cDNA (italic letters) for barfin flounder preproinsulin and deduced amino acid sequence. The coding regions are indicated by capital letters and non-coding regions and an intron are indicated by small letters. The deduced amino acid sequence is represented by the single letter code. The asterisk indicates termination codon. TATA box is indicated by solid underline and polyadenylation signal sequence is indicated by dotted underline.

This sequence contained 5'-CAGGCC-3' which encodes for Gln-Ala at B-chain N-terminus, indicating that this nucleotide sequence was a segment encoding insulin-II. A nucleotide sequence (0.25 kbp) encoding 3'-downstream region of N-terminal sequence of A-chain was also amplified from the genomic DNA digested with *Eco*RI.

In the third PCR, a nucleotide sequence encoding a region between the N-terminus of the signal peptide and the C-terminus of A-chain was amplified. The amplified product showed one band (0.48 kbp) on gel electrophoresis as in the first and second PCR, and its sequence matched sequences that were established in the first and second PCR. In each step of PCR, no other bands were detected on gel electrophoresis. This results indicate that nucleotide sequence obtained in each PCR originated from a single preproinsulin gene.

Structure of preproinsulin gene

The cDNA amplified by PCR is 273 bp, and corresponded to the sequence encoding from position 19 (Ser) of the signal peptide to position 14 (Phe) of A-chain except for an intron which split the C-peptide region (Figs. 5 and 6). The split position by this intron in preproinsulin gene was determined by comparison of nucleotide sequences between genomic DNA and cDNA. The intron consisted of 99 bp and split at position 7 (Gly) of C-peptide, and was flanked by GT-AG sequence. The amplified segment of cDNA and the nucleotide sequence of preproinsulin gene from genomic DNA were completely identical (Fig. 6).

The coding region of preproinsulin gene consisted of 348 bp, and a TATA box and a polyadenylation signal sequence were found at 124 bp 5'-upstream from the initiation codon and 116 bp 3'-downstream from termination codon, respectively (Fig. 6).

The amino acid sequences of A- and B-chains deduced from nucleotide sequence of genomic DNA was identical with those of insulin-II and also insulin-I established by Edman degradation. Amino acid sequences of the signal peptide and C-peptide were also deduced from nucleotide sequence and compared with those from other species (Fig. 3). The signal peptide of barfin flounder preproinsulin consists of 22 amino acid residues and contains a hydrophobic region, Leu-Leu-X-Leu, in the central region like other vertebrate insulins. C-Peptide consists of 37 amino acid residues with dibasic amino acids at both N- and C-termini (Fig. 3).

Southern blot analysis

Genomic Southern blot analysis showed a single positive band in all cases of genomic DNA digested with each of three restriction endonucleases (Fig. 7). Bands were observed at 3.0 kb (*Eco*RI), 3.3 kb (*Bam*HI) and 2.9 kb (*Hin*dIII). These indicate that preproinsulin gene of barfin flounder is a single copy gene.

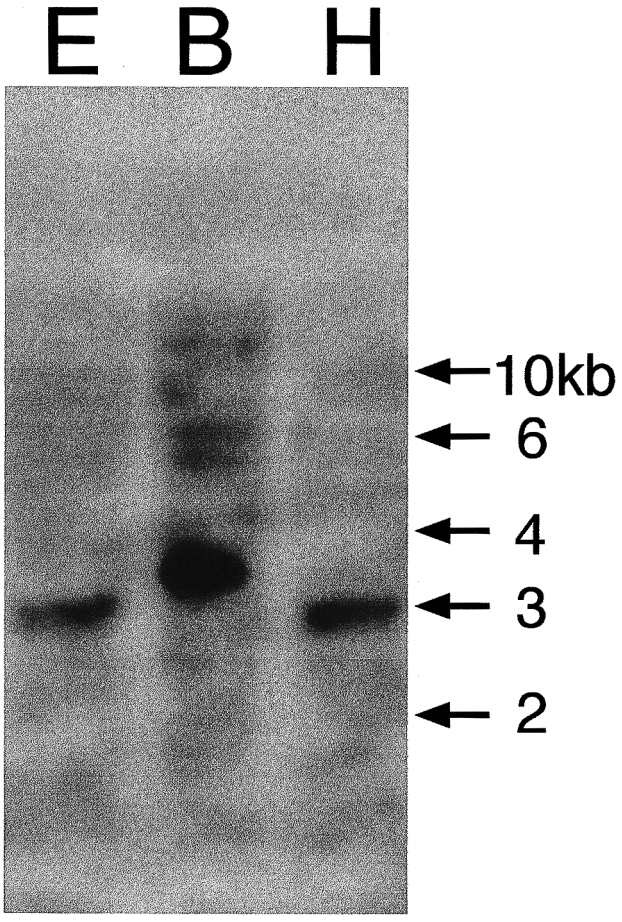

Fig. 7. Genomic Southern blot of barfin flounder preproinsulin gene. Genomic DNA was digested with *Eco*RI (lane E), *Bam*HI (lane B), and *Hin*dIII (lane H).

DISCUSSION

We identified two molecular forms of insulin, insulin-I and -II in barfin flounder. The existence of two molecular forms of insulin has previously been reported in toadfish (Smith, 1966), bonito (Yamamoto *et al*., 1960; Kotaki *et al*., 1962) and paddlefish (Nguyen *et al*., 1994). In these species, there are substitutions of internal amino acid residues between insulins of each respective species, indicating that these species contain two genes encoding each insulin molecule. However, the complete primary structures of insulins have not been determined in these species. In chum salmon (*Oncorhynchus keta*), several DNA clones encoding preproinsulin with sequence substitutions have been obtained (Kashuba *et al*., 1986; Koval *et al*., 1989; Kavsan *et al*., 1993), but only one form of insulin was purified as a peptide (Rusakov *et al*. 1990). The existence of two insulin genes was also reported in mouse (Buenzli *et al*., 1972; Wentworth *et al*., 1986), rat (Smith, 1966; Soares *et al*., 1985) and *Xenopus laevis* (Shuldiner *et al*., 1989a, b), and the two insulin in these species have already been purified. Conlon *et al*. (1989) isolated four molecular forms of insulin from the pancreas of *H. colliei*. Amino acid sequence analyses suggested that these four insulins were generated from a single

proinsulin by proteolytic cleavages at different sites at the flanking regions of C-peptide. In contrast, in barfin flounder, the results of amino acid sequence analyses suggested that insulin-I and -II arose from a single preproinsulin by proteolytic cleavage at different sites of the signal peptide region. This is supported by the facts that any other preproinsulin genes were not amplified in the three PCRs using different sets of primers, and only a single band was observed in genomic Southern blot analysis using three different kinds of restriction endonucleases. Thus, the generation system of multiple molecular forms of insulin in barfin flounder is different from those of other vertebrates.

Although insulins have been purified from over 60 vertebrate species and their amino acid sequences analyzed, insulin which contains Pyr at the N-terminus of B-chain has never been reported. Thus, insulin-II of barfin flounder is a novel type of vertebrate insulin. However, relaxins which also belong to the insulin superfamily contain Pyr residues at the N-terminus of A-chain in rat (Marie *et al*., 1981) and of B-chain in porcine (James *et al*., 1977; Schwabe *et al*., 1977) and shark (*Odontaspis taurus*; Gowan *et al*., 1981). Insulin-like peptides containing Pyr have also been reported in invertebrate, for example, in B-chains of bombyxins (Nagasawa *et al*., 1990) and in A- and B-chains of molluscan insulin-related peptides (MIPs; Li *et al*., 1992). Li *et al*. (1992) indicated the possibility that prolonged storage induces cyclization of a Gln residue at the N-terminus of A-chain of MIP-I.

In this study, we determined the complete nucleotide sequence of preproinsulin gene from genomic DNA of barfin flounder. The nucleotide sequences of preproinsulin genes were analyzed in only three species of teleosts; chum salmon (Sorokin *et al*., 1983; Kashuba *et al*., 1986; Koval *et al*., 1989; Kavsan *et al*., 1993), carp (Hahn *et al*., 1983), anglerfish (Hobart *et al*., 1980), and only in chum salmon, where the preproinsulin gene was analyzed from genomic DNA. In chum salmon, the complete nucleotide sequence of preproinsulin-I gene has been determined (Koval *et al*., 1989), whereas the nucleotide sequence analysis of preproinsulin-II gene is restricted to the regions encoding C-peptide and A-chain (Kavsan *et al*., 1993). The lengths of intron which splits the C-peptide region in chum salmon preproinsulin-I and -II genes are 265 and 287 bp, respectively, (Kavsan *et al*., 1993), which are almost three times as long as that (99 bp) of barfin flounder. However, the position of the intron in the barfin flounder is in common with those in preproinsulin-I and -II genes of chum salmon, and was flanked by GT-AG sequence according to the Chambon's rule (Breathnach and Chambon, 1981) like salmon preproinsulin-II gene. From the results of insulin gene analyses of mammals (e.g., human; Bell *et al*., 1980) and salmon (Koval *et al*., 1989), there is a possibility that another intron is present between TATA box and nucleotide sequence encoding signal peptide. However, this possible intron in barfin flounder has not been analyzed yet.

The signal peptide and C-peptide were much more variable than A- and B-chains from the deduced amino acid sequences from four fishes. Especially, length of the C-peptides is considerably different from one another, and besides, the sequence homology of the C-peptide central region is very low among these four fish, suggesting that the C-peptides do not have important function.

ACKNOWLEDGMENTS

We thank Mr. S. Imamura, a director of Japan Sea-Farming Association (JASFA), and Mr. K. Watanabe, a staff of JASFA, for supplying us with barfin flounders. We also thank Prof. F. Yamazaki, Hokkaido University, and Dr. T. Matsubara, Hokkaido National Fisheries Research Institute, for their helpful suggestions for this study. This work was supported in part by a grant (Development of fundamental technologies for effective improvement of aquatic organisms) from the Ministry of Agriculture, Forestry and Fisheries of Japan and by funds from the Cooperative Program provided by Ocean Research Institute, The University of Tokyo.

REFERENCES

Ablett RF, Sinnhuber RO, Holmes RM, Selivonchick DP (1981) The effect of prolonged administration of bovine insulin in rainbow trout (*Salmo gairdneri* R.). Gen Comp Endocrinol 43: 211–217

Andoh T, Nagasawa H (1997) Competitive enzyme immunoassay using antiserum against mammalian insulin for monitoring the purification of fish insulins. Bull Hokkaido Natl Fish Res Inst 61: 17–26

Bell GI, Pictet RL, Rutter WJ, Cordell B, Tischer E, Goodman HM (1980) Sequence of human insulin gene. Nature 284: 26–32

Breathnach R, Chambon P (1981) Organization and expression of eukaryotic split genes coding for protein. Annu Rev Biochem 50: 349–383

Buenzli HF, Glatthaar B, Kunz P, Muelhaupt E, Humbel RE (1972) Amino acid sequence of the two insulins from mouse (*Mus musculus*). Hoppe-Seyler's Z Physiol Chem 353: 451–458

Chan SJ, Emdin SO, Kwok SC, Kramer JM, Falkmer S, Steiner DF (1981) Messenger RNA sequence and primary structure of preinsulin in a primitive vertebrate, Atlantic hagfish. J Biol Chem 256: 7595–7602

Conlon JM, Thim L (1986) Primary structure of insulin and a truncated C-peptide from an elasmobranchian fish, *Torpedo marmorata*. Gen Comp Endocrinol 64: 199–205

Conlon JM, Davis M, Thim L (1987) Primary structure of insulin and glucagon from the flounder (*Platichthys flesus*). Gen Comp Endocrinol 66: 203–209

Conlon JM, Goke R, Andrews PC, Thim L (1989) Multiple molecular form of insulin and glucagon-like peptides from the Pacific ratfish (*Hydrolagus colliei*). Gen Comp Endocrinol 73: 136–146

Conlon JM, Thim L (1989) Isolation and primary structure of the C-peptide of proinsulin from the European eel (*Anguilla anguilla*). Comp Biochem Physiol 93B: 359–362

Conlon JM, Bondareva V, Rusakov Y, Plisetskaya EM, Mynarcik DC, Whittaker J (1995) Characterization of insulin, glucagon, and somatostatin from the river lamprey, *Lampetra fluviatilis*. Gen Comp Endocrinol 100: 96–105

Cutfield JF, Cutfield SM, Carne A, Emdin SO, Falkmer S (1986) The isolation, purification and amino-acid sequence of insulin from the teleost fish *Cottus scorpius* (daddy sculpin). Eur J Biochem 158: 117–123

Duan C, Hirano T (1992) Effects of insulin-like growth factor-I and insulin on the *in vitro* uptake of sulphate by eel branchial cartilage: evidence for the presence of independent hepatic and pancreatic sulphation factors. J Endocrinol 133: 211–219

Gowan LK, Reinig JW, Schwabe C, Bedarkar S, Blundell TL (1981) On the primary and tertiary structure of relaxin from the sand

tiger shark (*Odontaspis taurus*). FEBS Lett 129: 80–82
Hahn V, Winkler J, Rapoport TA, Liebscher DH, Coutelle C, Rosenthal S (1983) Carp preproinsulin cDNA sequence and evolution of insulin genes. Nucleic Acids Res 11: 4541–4552
Hobart PM, Shen LP, Crawford R, Pictet RL, Rutter WJ (1980) Comparison of the nucleic acid sequence of anglerfish and mammalian insulin mRNAs from cloned cDNAs. Science 210: 1360–1363
James R, Niall HD, Kwok S, Bryant-Greenwood GD (1977) Primary structure of porcine relaxin: homology with insulin and related growth factor. Nature 267: 544–546
Kashuba VI, Kavsan VM, Petrenko AI, Dmitrenko VV, Koval AP (1986) Allelic polymorphism of the salmon preproinsulin gene. Mol Biol 20: 845–852
Kavsan V, Koval A, Petrenko O, Roberts Jr CT, LeRoith D (1993) Two insulin gene are present in the salmon genome. Biochem Biophys Res Comm 191: 1373–1378
Kotaki A, Kurioka U, Satake K (1962) Studies on insulin. IV. On the amino acid composition of the glycyl chain of bonito insulin-I. J Biochem 51: 375–379
Koval AP, Petrenko AI, Demetrenko KGM (1989) Nucleotide sequence of the chum salmon preproinsulin gene. Mol Biol 23: 473–480
Li KW, Geraerts WPM, Ebberink RHM, Joose J (1992) Purification and sequencing of molluscan insulin-related peptide I (MIP I) from the neuroendocrine light green cells of *Lymnaea stagnalis*. Mol Cell Endocrinol 85: 141–150
Ludwig B, Higgs DA, Fagerlund UHM, McBride JA (1977) A preliminary study of insulin participation in the growth regulation of coho salmon (*Oncorhynchus kisutch*). Can J Zool 55: 1756–1758
Makower A, Dettmer R, Rapoport TA, Knospe S, Behlke J, Prehn S, Franke P, Etzold G, Rosenthal S (1982) Carp insulin: Amino acid sequence, biological activity and structural properties. Eur J Biochem 122: 339–345
Marie JJ, Borjesson BW, Walsh JR, Niall HD (1981) Limited sequence homology between porcine and rat relaxin: implications for physiological studies. Endocrinology 108: 726–729
Mommsen TP, Plisetskaya EM (1991) Insulin in fishes and agnathans: History, structure, and metabolic regulation. Rev Aquat Sci 4: 225–259
Nagasawa H, Kataoka H, Suzuki A (1990) Chemistry of *Bombyx* prothoracicotropic hormone and bombyxin. In "Molting and Metamorphosis" Ed by E Ohnishi and H Ishizaki, Japan Scientific Societies Press, Tokyo and Springer-Verlag, Tokyo, pp 33–48
Neumann PA, Koldenhof M, Humbel RE (1969) Amino acid sequence of insulin from the anglerfish (*Lophius piscatorius*). Hoppe-Seyler's Z Physiol Chem Bd 350: 1286–1288
Nguyen TM, Mommsen TP, Mims SM, Conlon JM (1994) Characterization of insulins and proglucagon-derived peptides from a phylogenetically ancient fish, the paddlefish (*Polyodon spathula*). Biochem J 300: 339–345
Peterson JD, Steiner DF (1975) The amino acid sequence of the insulin from a primitive vertebrate, the Atlantic hagfish (*Myxine glutinosa*). J Biol Chem 250: 5183–5191
Plisetskaya EM, Donaldson EM, Dye HM (1987) Plasma insulin levels during the spawning migration of the pink salmon, *Oncorhynchus gorbuscha*. J Fish Biol 31: 21–23
Plisetskaya EM, Swanson P, Bernard MG, Dickhoff WW (1988) Insulin in coho salmon (*Oncorhynchus kisutch*) during the parr to smolt transformation. Aquaculture 72: 151–164
Plisetskaya EM, Buchelli-Narvaez LI, Hardy RW, Dickhoff WW (1991) Effects of injected and dietary arginine on plasma insulin levels and growth of pacific salmon and rainbow trout. Comp Biochem Physiol 98A: 165–170
Rusakov YI, Karasev VS, Bondareva VM, Pertseva MN, Pankov YA (1990) Isolation primary structure, and biological and immunological properties of pink and chum salmon insulins. Comp Biochem Physiol 95B: 477–481
Sanger FS, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 75: 5463–5467
Schwabe C, McDonald JK, Steinetz BG (1977) Primary structure of the B-chain of porcine relaxin. Biochem Biophys Res Commum 75: 503–510
Shuldiner AR, Bennett C, Robinson ER, Roth J (1989a) Isolation and characterization of two different insulins from an amphibian, *Xenopus laevis*. Endocrinology 125: 469–477
Shuldiner AR, Phillips S, Roberts Jr CT, LeRoith D, Roth J (1989b) *Xenopus laevis* contains two nonallelic preproinsulin genes. J Biol Chem 264: 9428–9432
Smith LF (1966) Species variation in the amino acid sequence of insulin. Am J Med 40: 662–666
Soares MB, Schon E, Henderson A, Karathanasis SK, Cate R, Zeitlin S, Chirgwin J, Efstratiadis A (1985). RNA-mediated gene duplication: the rat preproinsulin I gene is a functional retroposon. Mol Cell Biol 5: 2090–2103
Sorokin AV, Skoblov YS, Mironov AA, Zlochevskii ML, Dem' yanova NG, Rebentish BA, Kozlov YI, Kavsan VM (1983) Nucleotide sequence of cloned insulin cDNA of *Oncorhynchus keta*. Dokl Biochem 269: 80–83
Sundby A, Eliassen K, Refsti T, Plisetskaya E (1991). Plasma levels of insulin, glucagon and glucagon-like peptide in salmonids of different weights. Fish Physiol Biochem 9: 223–230
Wentworth BM, Schaefer IM, Villa-Komaroff L, Chirgwin JM (1986) Characterization for two non-allelic genes encoding mouse preproinsulin. J Mol Evol 23: 305–312
Yamamoto M, Kotaki A, Okuyama T, Satake K (1960) Studies on insulin. I. Two different insulins from Langerhans islet of bonito fish. J Biochem 48: 84–92

(Received June 2, 1998 / Accepted August 18, 1998)