Evidence of Introduction by Molecular Cloning of Artificial Inverted Sequence at the 5' Terminus of the Sense Strand of Rat Insulin-like Growth Factor-I cDNA

Hisanori KATO, Asako TAKENAKA, Yutaka MIURA, Makoto NISHIYAMA
and Tadashi NOGUCHI*

*Department of Agricultural Chemistry, Faculty of Agriculture,
The University of Tokyo, Bunkyo-ku, Tokyo 113, Japan*

Received February 1, 1990

A cDNA of insulin-like growth factor (IGF)-I mRNA was obtained from a rat liver cDNA library. This cDNA contained a 5' terminal sequence of 48 base pairs complementary to a sequence in the 3' terminal region (an inverted repeat sequence) except for one nucleotide insertion in the 3' region. Base sequence analysis of this cDNA, northern blot analysis of rat liver mRNA using a probe specific to 5' region of cDNA, base sequence studies of the 5' upstream region of a clone of the rat IGF-I gene, and some other studies suggested strongly that the inverted repeat sequence of the IGF-I cDNA used was introduced artificially during preparation of the cDNA library.

Insulin-like growth factor-I (IGF-I) is a hormone which has been assumed to mediate the activity of growth hormone (GH) on target organs.¹⁾ Since IGF-I shows insulin-like activities on various tissues, the insulin-like activity of GH may be explained by the action of IGF-I.²⁾

cDNAs of human,^{3,4)} rat,^{5,6)} mouse,⁷⁾ and porcine⁸⁾ IGF-I have been cloned. The structures of these cDNAs are heterogeneous in the base sequences corresponding to the N- or C-terminal region of prepro IGF-I or the 5'- or 3'-untranslated region of mRNA,^{4,5,9)} although the animals including humans have a single IGF-I gene.^{10,11)} Particularly, in rat cDNA, Shimatsu and Rotwein⁶⁾ reported two different cDNAs in the 5'-untranslated region and Roberts *et al.*⁹⁾ showed 3 different cDNAs in the region. Each one of the cDNAs obtained by them showed the same structure in the 5'-untranslated region. Class A cDNA reported by Roberts *et al.*⁹⁾ had a special feature; its 5'-untranslated region contains a 40-bp inverted repeat sequence of a sequence found in the 3'-untranslated region. Using a solution

hybridization/RNase protection assay method, Lowe *et al.*¹²⁾ showed that mRNA with this inverted repeat sequence is contained in rat liver, kidney, lung, testis, and stomach. They further claimed that the difference of the mRNA structure is closely related to the translation activity of the mRNA.

We obtained cDNAs of IGF-I from the cDNA library of rat liver. One of them contained the inverted repeat sequence that had been reported by Roberts *et al.*⁹⁾ However, we got several lines of evidence suggesting that this inverted repeat sequence was artificially produced during preparation of the cDNA library.

Materials and Methods

Materials. Enzymes including restriction endonucleases, T4 DNA ligase, and T4 polynucleotide kinase were purchased from commercial suppliers (Takara Shuzo, Boehringer Mannheim, and Toyobo). Plasmids pUC118 and 119 and bacteriophage M13 mp 18 and 19 were purchased from Takara Shuzo. All radionucleotides were obtained from New England Nuclear. Nitrocellulose filters were from Schleicher and Schuell.

* To whom all correspondence should be addressed.

Hybridization probes. Oligonucleotide probes were synthesized using a Beckman System 1 Plus DNA Synthesizer and labeled at the 5'-end using [γ - 32 P]ATP and T4 polynucleotide kinase. A 42-base oligonucleotide corresponding to the DNA sequence encoding amino acids 10-23 of human IGF-I was used for cDNA screening. The sequence of the probe is the same as that used by Rotwein.⁴⁾ Two other oligonucleotide probes were used in Northern blot hybridizations. Probe A (5' TCAGAGTGGG-CATTTCCCTCAATGAAATACACAAGTAAA 3') and B (5' TGCCTCATTTTGGTACGTCCCGGGTCGT-TTACACAGCA 3') are 39-base oligomers that are complementary to different portions of the 5' upstream sequence of Class A rat IGF-I cDNA. Probe A is complementary to the Class A-specific region and probe B is to the common region of Class A and B cDNA (Fig. 1). Probe B was also used in screening the rat genomic DNA library.

Screening of cDNA and genomic library. A rat liver cDNA library in λ gt11 and a rat genomic library in Charon 4A (liver, partial *Hae*III digest) were obtained from Clontech. According to the manufacturer's catalogue, the cDNA was synthesized by the method of Young.¹⁹⁾ Approximately 2×10^6 bacteriophage plaques from the cDNA library and 1×10^6 plaques from the genomic library were screened by standard methods.¹³⁾

DNA sequence analysis. Restriction fragments were subcloned into M13 mp 18 and 19 bacteriophages. DNA was sequenced using an M13 sequence kit (Takara Shuzo) or SequenaseTM (United States Biochemical) according to the manufacturer's protocol. A λ mapping system (Takara Shuzo) was used for restriction mapping of genomic DNA

clones.

Northern blot analysis. Total RNA was prepared from the liver of adult 200-250 g male Wistar rats by the guanidine thiocyanate method of Glisin *et al.*¹⁴⁾ Poly (A)⁺ RNA was prepared with oligo d(T) cellulose (Collaborative Research Inc.) according to the manufacturer's instructions. RNA gel electrophoresis and RNA blot analysis were done by the method of Lowe *et al.*^{12,15)} Briefly, 5 μ g of poly (A)⁺ RNA was denatured in formaldehyde, size-fractionated on a 1.5% agarose gel, and transferred to a nylon membrane (Gene Screen, New England Nuclear) using 1 M ammonium acetate as the transfer buffer. RNA was fixed to the membrane by baking *in vacuo* at 80°C for 2 hr. The blots were prehybridized in $5 \times$ SSPE¹³⁾/20% formamide/10 $\times$ Denhardt's solution¹³⁾/1% NaDodSO₄/100 μ g of denatured salmon sperm DNA per ml for 2 hr at 50°C. Hybridization was done in $5 \times$ SSPE/20% formamide/1% NaDodSO₄ with 1×10^6 cpm per ml of labeled oligonucleotide for 20 hr at 50°C. The blots were washed twice in $2 \times$ SSPE and 0.2% NaDodSO₄ at 40°C for 30 min and then for 15 min and 5 min in $0.5 \times$ SSPE at 40°C. The autoradiograms were exposed for 4 days.

Results

Structure of cDNA

We obtained two cDNA clones containing 5'-untranslated regions corresponding to the Classes A and C named by Roberts *et al.*⁹⁾ Figure 1 shows the nucleotide sequence of Class

Fig. 1. Sequence of Class A Rat cDNA Encoding Prepro-IGF-I.

The predicted amino acid sequence is numbered from the amino terminus of IGF-I. The region corresponding to mature IGF-I is boxed. Shaded portions at the 5' and that at the 3' end of the cDNA comprise mutually inverted repeat sequence. The sequence flanking the shaded portion (GAATTCGG) is that of the *Eco*RI linker. Oligonucleotide probes A and B were designed to hybridize to the underlined sequences of the IGF-I mRNA. The point from which the class B sequence diverges from the class A sequence is indicated by ▲.

Fig. 2. Autoradiograms of Northern Blots of Adult Rat Liver RNA with Oligonucleotide Probes.

Probes A and B (described in Materials and Methods and in Fig. 1) were labeled and hybridized to RNA blots of 5 μ g of rat liver poly (A)⁺ RNA. Arrows represent different mRNA species of IGF-I.

A cDNA. The boxed region is the mature IGF-I-coding region. The structure of Class B cDNA was different from Class A in its structure upstream of the nucleotide indicated by ($\blacktriangle$). Shaded areas are the inverted repeat portion. In our case, the length of the inverted repeat was 48 bases which is longer than that reported by Roberts *et al.*⁹⁾ Among the 48 nucleotides, only one nucleotide was different between the two complementary regions, one A indicated by a star in Fig. 1 was inserted in the 3'-untranslated region.

Northern blot analyses

We found 4 species of mRNAs in poly (A)⁺ RNA extracted from rat liver, the sizes of which were 0.8–1.2, 2.0, 6.0, and 7.4 kb. This northern blot analysis (Fig. 2) was done using a probe of synthetic DNA corresponding to the common region of Class A and B (probe B shown by the underline B in Fig. 1). On the contrary, no mRNA was found when we used a probe specific to Class A mRNA (probe A shown by the underline A in Fig. 1). These results suggest that rat liver does not contain mRNA corresponding to Class A cDNA.

Fig. 3. A Map of a Genomic Clone Containing 5' Untranslated Region of Rat IGF-I Gene and Partial Sequence of the Clone.

Exons are depicted by boxes, introns and flanking regions by the thin lines. *Eco*RI (E) sites are indicated on the line. Exon 2 corresponds to the first 25 amino acids of the B domain of IGF-I and 9 amino acids that precede the B domain. Exon 1 contains the region that corresponds to the 5' untranslated region of Class C cDNA, and exon 1' that of Class B cDNA. The sequence of the region corresponding to probe B and its adjacent region is shown. The point at which Classes A and B diverge is indicated by ($\blacktriangle$). The shaded region of 10 bp is the inverted repeat of the 11 bp region at the 3' untranslated region.

Partial sequence of the upstream region of rat IGF-I gene

Using probe B as a hybridization probe, we obtained from a rat genomic library some λ phages that contained a region corresponding to the 5' region of IGF-I gene. The exon corresponding to Class C cDNA has been sequenced by Shimatsu and Rotwein¹¹⁾ and recently that of Class B by Bucci *et al.*¹⁶⁾ One of our clones (λ IGF 5-2) contained an approximately 6 kb longer nucleotide sequence in the 5' upstream region than those of Shimatsu and Rotwein¹¹⁾ or Bucci *et al.*¹⁶⁾ (Fig. 3). Dot hybridization and Southern blot hybridization showed that λ IGF 5-2 does not

contain any region that hybridizes with probe A. The nucleotide sequence of the region that hybridizes with probe B coincided with that reported by Bucci *et al.*¹⁶⁾ The point from the upstream of which Class A and B diverges ($\blacktriangle$ in Fig. 3), did not show any canonical consensus sequence for an acceptor splice site (intron-exon boundary).

Discussion

These results strongly suggest that the inverted repeat of our Class A cDNA is an artificial product. The following observations give evidence for this hypothesis. 1. The

Fig. 4. A Possible Mechanism of Formation of an Artificial Inverted Repeat Sequence during Cloning.

The hypothesis of the formation of an artificial inverted repeat sequence was deduced from the following assumptions. 1) After the synthesis of the first strand, the original mRNA is degraded by the intrinsic RNase H activity in the reverse transcriptase.²⁰⁾ 2) The complementary region of the cDNA is annealed (fold back), and the whole length of the inverted repeat is synthesized by the intrinsic DNA dependent DNA polymerase activity of reverse transcriptase.²⁰⁾ 3) This hairpin-like single strand DNA is denatured at 100°C (boiling) before second strand synthesis.

inverted repeat sequence starts exactly from both the 5' and 3' termini of cDNA, although this is not a prerequisite for the artificial formation of an inverted repeat sequence (Weaver *et al.*).¹⁷⁾ 2. All nucleotides of such long regions as 48 bp are complementary with an exception of only one inserted nucleotide in the 3'-terminal region, which is common to Class A and B. 3. No Class A mRNA was found by Northern blot analysis. 4. No exon which corresponds to Class A was found in the 16 kb upstream from exon 2, although those corresponding to Class B and C were found within 7 kb upstream from exon 2. 5. The structure of genomic DNA showed that the point at which Classes A and B diverge did not show the consensus sequence of a splicing point (GT-AG pair).

Figure 4 shows a way to explain the formation of the artificial inverted repeat sequence. The figure shows that after the synthesis of the first strand, 10 nucleotides in the 5' region and 11 nucleotides in the 3' region annealed and a new sequence was synthesized using these regions as the primer. This mechanism of artificial cDNA sequence production is also reported in bovine parathyroid hormone.¹⁷⁾

From these observations, we conclude that the inverted repeat sequence of our Class A-like cDNA was produced during preparation of the cDNA library. This does not necessarily mean that the inverted repeat of cDNA of Roberts *et al.*⁹⁾ was also synthesized artificially. Although the 3'-terminal sequence of their Class A cDNA is not reported, it is probable enough to assume that the inverted repeat of their cDNA was also introduced during preparation of the library, because the library of theirs and ours is the same. However, in subsequent articles,^{12,18)} they proved the occurrence of mRNA of Class A cDNA by the solution hybridization/RNase protection assay method. The discrepancy between their results and ours may be explained by the following possibilities. 1. The Class A sequence is far upstream from the 5'-terminal region. We could not find the sequence in the 5'-untranslated

region up to 16 kb. But there is still a possibility that the sequence is further upstream. 2. The Class A sequence is an artificial product. Because the size difference in the solution hybridization/RNase protection assay is small between Class A and B mRNAs, there is the possibility that Class B mRNA was erroneously identified as Class A mRNA. Lowe *et al.*¹²⁾ also reported that they could not find Class A mRNA by Northern blot analysis. 3. The DNA sequence for Class A mRNA is not present in genomic DNA and Class A mRNA was produced after transcription by some unknown mechanism. We assume that this possibility is small.

Our results suggest strongly that the 5'-untranslated region of Class A cDNA was introduced artificially. This means that analysis of the structure of cDNA does not necessarily give the true structure of mRNA.

Acknowledgment. The authors are grateful to Dr. Teruhiko Beppu of the Department of Agricultural Chemistry, Faculty of Agriculture, The University of Tokyo for his valuable suggestions.

References

- 1) E. M. Spencer (ed.), "Insulin-like Growth Factors/Somatomedins," Walter de Gruyter, Berlin and New York, 1983.
- 2) E. R. Froesch, C. Schmid, J. Schwander and J. Zapf, *Ann. Rev. Physiol.*, **47**, 443 (1985).
- 3) M. Jansen, F. M. A. van Schaik, A. T. Ricker, B. Bullock, D. E. Woods, K. H. Gabbay, A. L. Nussbaum, J. S. Sussenbach and J. L. Van den Brande, *Nature*, **306**, 609 (1983).
- 4) P. Rotwein, *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 77 (1986).
- 5) C. T. Roberts, Jr., S. R. Lasky, W. L. Lowe, Jr., W. T. Seaman and D. LeRoith, *Mol. Endocrinol.*, **1**, 243 (1987).
- 6) A. Shimatsu and P. Rotwein, *Nucl. Acids Res.*, **15**, 7196 (1987).
- 7) G. I. Bell, M. M. Stempien, N. M. Fong and L. B. Rall, *Nucl. Acids Res.*, **14**, 7872 (1986).
- 8) A. Tavakkol, F. A. Simmen and R. C. M. Simmen, *Mol. Endocrinol.*, **2**, 674 (1988).
- 9) C. T. Roberts, Jr., S. R. Lasky, W. L. Lowe, Jr. and D. LeRoith, *Biochem. Biophys. Res. Commun.*, **146**, 1154 (1987).
- 10) P. Rotwein, K. M. Pollock, D. K. Didier and G. G.

- Krivi, *J. Biol. Chem.*, **261**, 4828 (1986).
- 11) A. Shimatsu and P. Rotwein, *J. Biol. Chem.*, **262**, 7894 (1987).
 - 12) W. L. Lowe, Jr., C. T. Roberts, Jr., S. R. Lasky and D. LeRoith, *Proc. Natl. Acad. Sci. U.S.A.*, **84**, 8946 (1987).
 - 13) T. Maniatis, E. F. Fritsch and J. Sambrook, in "Molecular Cloning: A Laboratory Manual," Cold Spring Harbor, New York, 1982.
 - 14) V. Glisin, R. Crkvenjakov and C. Byus, *Biochemistry*, **13**, 2633 (1974).
 - 15) W. L. Lowe, Jr., A. E. Schaffner, C. T. Roberts, Jr. and D. LeRoith, *Mol. Endocrinol.*, **1**, 181 (1987).
 - 16) C. Bucci, P. Mallucci, C. T. Roberts, R. Frunzio and C. B. Bruni, *Nucl. Acids Res.*, **17**, 3596 (1989).
 - 17) C. A. Weaver, D. F. Gordon and B. Kemper, *Proc. Natl. Acad. Sci. U.S.A.*, **78**, 4073 (1981).
 - 18) M. Adamo, W. L. Lowe, Jr., D. LeRoith and C. T. Roberts, Jr., *Endocrinology*, **124**, 2737 (1989).
 - 19) T. V. Huynh, R. A. Young and R. W. Davis, in "DNA Cloning Techniques, A Practical Approach," Vol. II, ed. by D. M. Glover, JRL Press, Oxford, 1985, pp. 49-78.
 - 20) I. M. Verma, in "The Enzymes," Vol. 14, ed. by P. D. Boyer, Academic Press, 1981, pp. 87-103.
-