Divergence of insulin-like growth factors I and II in the elasmobranch, *Squalus acanthias*

Stephen J. Duguay[a,*], Shu Jin Chan[a], Thomas P. Mommsen[b], Donald F. Steiner[a]

[a]*Howard Hughes Medical Institute and Department of Biochemistry and Molecular Biology, University of Chicago, 5841 S. Maryland Avenue, Chicago, IL 60637, USA*

[b]*Department of Biochemistry and Microbiology, University of Victoria, Victoria, B.C., Canada V8W 3P6*

Received 21 July 1995

Abstract **Recent studies have shown that vertebrates, including teleostean fishes, amphibians, birds and mammals, contain two distinct insulin-like growth factor (IGF) genes. In contrast agnathans, represented by hagfish, apparently have only one IGF that has features characteristic of both IGF-I and IGF-II. Between these groups the elasmobranchs occupy a critical position in terms of the phylogeny of IGFs. We sought to determine if gene duplication and divergence of IGF-I and IGF-II occurred before or after divergence of elasmobranchs from other vertebrates by cloning IGF-like molecules from *Squalus acanthias*. Our analysis shows that *Squalus* liver produces two distinct IGF-like molecules. One has greater sequence identity to, and conserved features characteristic of, known IGF-I molecules, while the other is more IGF-II like. These results suggest that the prototypical IGF molecule duplicated and diverged in an ancestor of the extant gnathostomes.**

Key words: Insulin-like growth factor; *Squalus acanthias* (spiny dogfish shark)

1. Introduction

Insulin-like growth factors (IGFs) are polypeptides that play important roles in development and growth [1]. IGFs were initially identified as 'sulfation factors' because of their ability to stimulate the incorporation of sulfate by cartilage [2], 'non-suppressible insulin-like activity' (NSILA) due to insulin-like effects on adipose tissue metabolism [3], and 'multiplication stimulating activity' (MSA) because they were able to stimulate multiplication of chicken embryo fibroblasts [4]. As these factors were characterized in greater detail it became evident that sulfation factor, NSILA, and MSA were the same or very similar molecules. Purification of NSILA from human serum and amino acid sequencing revealed two peptides, named insulin-like growth factor-I and -II, that had significant homology to each other as well as proinsulin [5,6]. Like proinsulin, IGF-I and IGF-II have a B-, C-, and A-domain. The B- and A-domains of proinsulin and IGFs show a high degree of sequence identity while the C-domains have little sequence identity. The C-domain of IGFs is not removed by proteolytic processing, as is the case in proinsulin, and in addition, the IGFs contain carboxy-terminal sequences that have been named the D- and E- domains. The E-domain is removed as a post-translational modification to produce mature IGF [7].

The development of recombinant DNA technology has stimulated much work on the evolution of insulin and IGF genes. An insulin-like peptide (ILP) cDNA has been cloned from the protoochordate *Amphioxus californiensis* that has characteristics of both insulin and IGF. ILP contains a B-, C-, and A-domain similar to those of proinsulin and IGF. The C-domain is flanked by paired basic residues and might thus be removed by proteolytic processing as is the case with proinsulin. However, the A-domain of ILP is followed by sequences reminiscent of the D- and E-domain of IGF. *Amphioxus* ILP might therefore resemble a prototype insulin/IGF molecule [8]. In contrast, the hagfish belongs to the class agnatha (jawless vertebrates) and is considered to be one of the most primitive vertebrates. Two members of the insulin family have been cloned from the Atlantic hagfish [9,10]. One of them is clearly an insulin, while the other is an IGF with nearly equal sequence identity to either human IGF-I or human IGF-II. This suggests that hagfish IGF might represent a prototype IGF molecule. Recently, molecules with obvious sequence identity to either IGF-I or IGF-II have been sequenced from teleosts, amphibians, birds, and mammals (see legend to Fig. 2 for specific references). Therefore, an important question regarding the evolution of IGF-I and IGF-II is when did the gene duplication of the ancestral IGF molecule occur to give rise to two distinct IGFs? Elasmobranchs are a major group of vertebrates that lie between agnathans and teleosts on the phylogenetic tree. We sought to address the question of IGF divergence by cloning IGF-like molecules from an elasmobranch, *Squalus acanthias*, the spiny dogfish shark.

2. Materials and methods

Total RNA was prepared from *Squalus* liver using the guanidine thiocyanate procedure [17]. Complementary DNA was synthesized from total RNA using random primers and MMLV reverse transcriptase (Gibco/BRL). PCR amplification was performed for 40 cycles at 94°C for 1 min, 50°C for 1 min 30 s, and 65°C for 1 min 30 s with Taq DNA polymerase (Perkin Elmer). Primers used were IGF-9: 5′-GCPCAPTACATQTCG/TAG, IGF-10: 5′-GCGCAGTAG/TGTQTCG/TAG, IGF-14: 5′-GAA/GACNCTNTCT/CGGNGG, IGF-15: 5′-GAA/GACNCTNTGT/CGGNTC, and IGF-17: 5′-GAA/GACNCTNTGT/CGGNGC, where P = A or G, Q = C or T, and N = A, G, C, or T. PCR products were separated by electrophoresis on 8% polyacrylamide gels. Bands of the appropriate size were excised, eluted, and cloned into the *Eco*RV site of pBluescript (Stratagene). The 5′ and 3′ ends of the cDNAs were amplified by RACE [18], gel purified, and

*Corresponding author. Fax: (1) (312) 702 4292.

Abbreviations: IGF, insulin-like growth factor; ILP, insulin-like peptide; PCR, polymerase chain reaction; RT-PCR, reverse transcriptase PCR; RACE, rapid amplification of cDNA ends; bp, basepair.

The nucleotide sequence data reported in this paper have been submitted to the EMBL Data Library under Accession Numbers Z50081 and Z50082.

SSDI 0014-5793(95)00879-9

blunt-end ligated to pBluescript. DNA sequencing was performed using the dideoxynucleotide method with Sequenase version 2.0 (United States Biochemical). DNA and protein sequence analysis was performed with GeneWorks version 2.4 from IntelliGenetics.

3. Results and discussion

3.1. Cloning of shark IGF-I and IGF-II

Amino acid sequence alignments of IGF-I and IGF-II sequences from mammals, birds, amphibians, and fish revealed regions in the B- and A-domains that were conserved among either IGF-I or IGF-II sequences, but not both IGF-I and IGF-II. This suggested that degenerate PCR primers made to these areas might allow preferential amplification of either IGF-I or IGF-II from shark liver cDNA, if both sequences existed. Primers IGF-9 and IGF-10 were designed to anneal to the LeuGluMetTyrCysAla sequence of the IGF-I A-domain and the LeuGluThrTyrCysAla sequence of the IGF-II A-domain, respectively. Primers IGF-17, IGF-14, and IGF-15 were designed to anneal to the B-chain sequence GluThrLeuCysGlyAla of IGF-I, GluThrLeuCysGlyGly of IGF-II, and GluThrLeuCysGlySer of hagfish IGF, respectively. Various combinations of these primers were used for PCR of shark liver cDNA. Primer pairs IGF-9/IGF-17 and IGF-10/IGF-15 gave PCR products of the expected size of 180 bp. These PCR products were cloned and sequenced and found to encode IGF sequences. RACE was used to obtain 5′ and 3′ cDNA sequences. Finally, the entire coding region of each IGF was then amplified by PCR with two different primer pairs, cloned and sequenced. Each IGF has therefore been amplified by three independent PCR reactions in order to verify that the sequence is correct.

3.2. Analysis of shark IGF-I and IGF-II sequences

We propose that the two distinct IGF sequences that were cloned from shark liver cDNA are homologues of the vertebrate IGF-I and IGF-II genes based on overall sequence identity with mammalian, avian, amphibian, and piscine IGF sequences, as well as conservation of specific residues which appear to be unique to either IGF-I or IGF-II.

Shark IGF-I (Fig. 1A) contains a 24 residue signal peptide. Most other IGF-I sequences contain much longer signal peptides (Fig. 2A). There were no in-frame stop codons found upstream of the shark IGF-I signal peptide in the longest 5′-RACE product and it is therefore possible that an additional initiator methionine residue does exist further upstream. Mature shark IGF-I consists of 70 amino acids and has 66% and 62% sequence identity to human IGF-I and IGF-II, respectively (Tables 1 and 2). Conserved residues common to all IGF-I sequences, including shark, but not IGF-II sequences are Pro^{2} and Ala^{8} of the B- domain, $Gly^{30}Tyr^{31}$ of the C-domain, and Met^{59} of the A-domain (Fig. 2A). We also note that all IGF-I D-domains are 8 or 9 amino acids long and all IGF-II D-domains are 6 amino acids in length.

Table 1
Percentage amino acid sequence identity of mature IGF-I vs. IGF-I (upper right) and IGF-II vs. IGF-II (lower left)

IGF-I

IGF-II	human	chicken	frog	trout	shark	hagfish	IGF-I
human		89	84	80	66	55	human
chicken	84		89	84	68	56	chicken
---	-	-		84	63	55	frog
trout	79	73	_		62	55	trout
shark	66	59	_	60		54	shark
hagfish	62	56	_	58	59		hagfish
	human	chicken	---	trout	shark	hagfish	

IGF-II

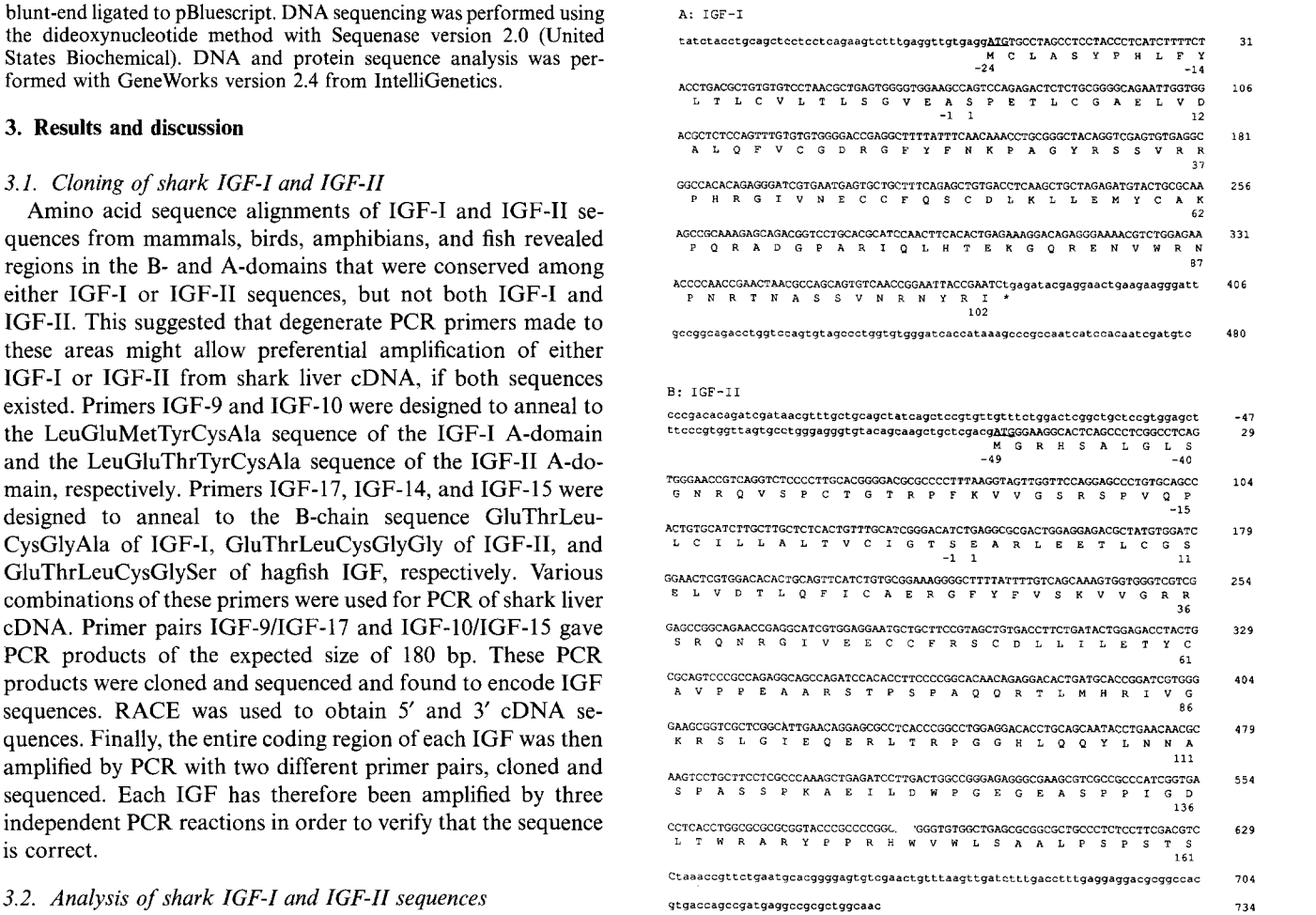

Fig. 1. Nucleotide and deduced amino acid sequence of shark IGF-I (A) and IGF-II (B). The cDNA sequence is shown on top and numbered at right. 5′ and 3′ untranslated sequences are indicated by lower case letters and the coding region is shown in upper case letters. The initiator methionine codon is underlined and the stop codon is indicated by an asterisk. The deduced amino acid sequence is represented by the single letter code and numbered below.

The shark IGF-I E-domain contains 32 amino acids. IGF-I E-domains from other species are 35 amino acids long but longer forms can be generated by alternative splicing of mRNA [19,20]. PCR amplification of the E-domain region of shark IGF-I yielded a single band (data not shown), suggesting that alternative splicing does not occur in shark liver. All IGF-I E-domains begin after arginines, which have been shown to be necessary for proteolytic removal of the E-domain during prohormone processing. The lysine residue at position 68 is also required for processing [21]. It is interesting that shark IGF-I contains a glycine in position 68. We have previously shown that substitution of Lys^{68} in human IGF-I with glycine completely inhibits processing at Arg^{71} [21]. This suggests that shark proIGF-I might not be processed to mature IGF-I, or a com-

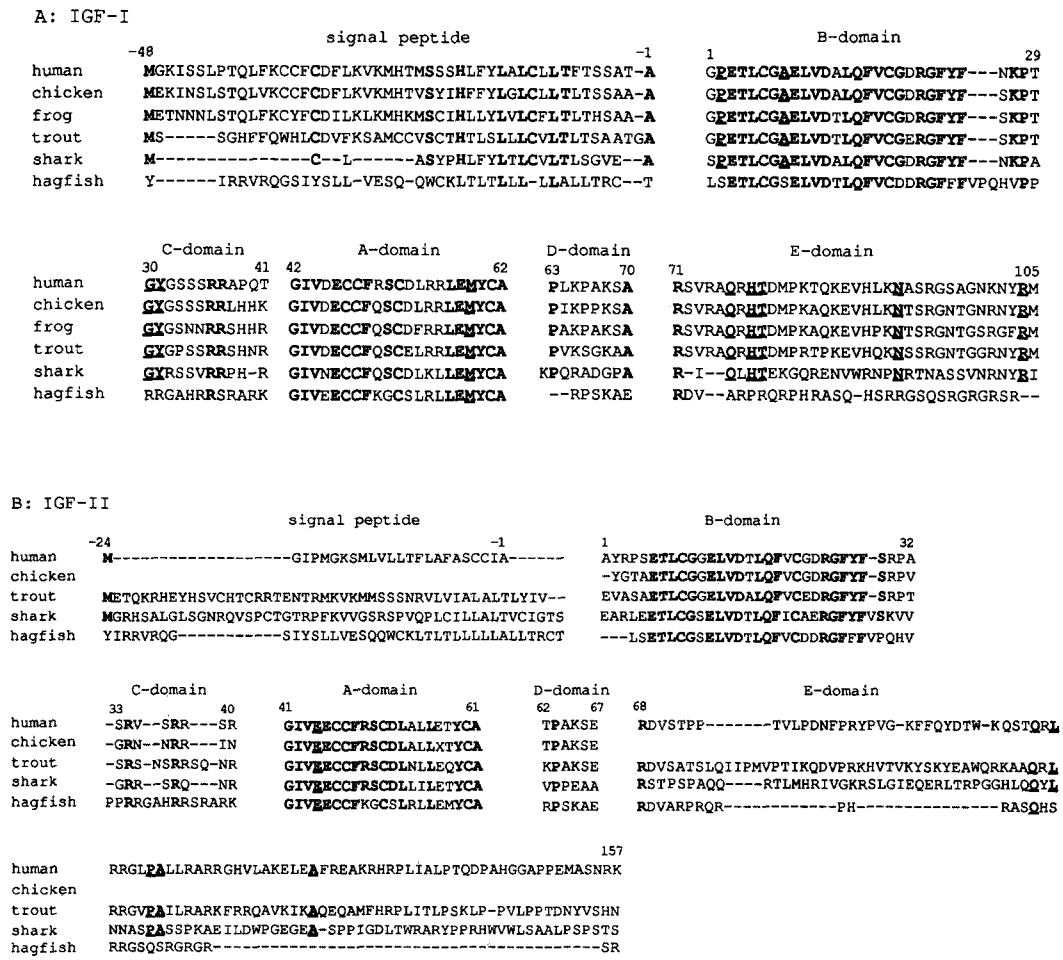

Fig. 2. Amino acid alignment of IGF-I (A) and IGF-II (B) sequences. Amino acids are represented by the single letter code and numbering corresponds to human IGF-I residues. Residues conserved in all species are indicated by bold-faced type and residues that are unique to either IGF-I or IGF-II are underlined. References for sequences are as follows: human IGF-I and -II [11,12], chicken IGF-I and -II [13,14], frog IGF-I [15], trout IGF-I and -II [16], and hagfish IGF [10].

pensatory mutation may have occurred in the shark proIGF-I processing enzyme such that it still cleaves the substrate. It is also possible that shark proIGF-I is processed at an alternative LysXaaXaaArg site, such as LysProGlnArg[65] or LysGlyGlnArg[81]. Residues that are conserved in all IGF-I sequences but not found in the IGF-II E-domains are Gln[76], His[78]Thr[79], the Asn[92]XaaSer/Thr *N*-glycosylation site, and Arg[104]. Despite low sequence identity with E-domains of IGF-I from other species, the basic nature (calculated pI = 11.81) of the shark IGF-I E-domain is conserved.

Shark IGF-II has a signal peptide that is 49 amino acids long and preceded by an in-frame stop codon 36 nucleotides upstream of the initiator methionine codon (Fig. 1B). Mature shark IGF-II is 68 amino acids long and has 66% and 55% sequence identity with human IGF-II and IGF-I, respectively (Tables 1 and 2). The only residue that is conserved and unique to all IGF-II sequences, compared to IGF-I, is Glu[44] of the A-domain (Fig. 2B). However, shark IGF-II contains several other features that are hallmarks of IGF-II sequences. The GluThrLeuCysGly sequence found in all IGF B-domains is preceded by two amino acids in IGF-I sequences but by 4 or 5 residues in IGF-II sequences. Comparison of IGF-I and IGF-II A-domain sequences reveals a Leu/PheArgArgLeu[57] motif in IGF-I. The homologous region in IGF-II contains hydropho-

Table 2
Percentage amino acid sequence identity of mature IGF-I vs. mature IGF-II

		IGF-I					
		human	chicken	frog	trout	shark	hagfish
IGF-II	human	63	62	64	60	62	62
	chicken	63	61	65	57	58	56
	trout	67	66	67	62	62	58
	shark	55	55	53	59	55	59
	hagfish	55	56	55	55	54	–

bic residues in place of the basic residues found in IGF-I. Shark IGF-II contains two hydrophobic residues in this motif. Shark IGF-II also contains a 6 amino acid D-domain whereas IGF-I D-domains are 8 or 9 amino acids long (Fig. 2).

IGF-II E-domains tend to be longer than IGF-I E-domains and have not been found to be alternatively spliced. The shark IGF-II E-domain follows this trend. It is 84 amino acids long and no evidence for alternative splicing was obtained using RT-PCR analysis (data not shown). As noted previously [20], both IGF-I and IGF-II E-domains from all species are extremely basic. The predicted isoelectric point of the shark IGF-II E-domain is 11.47. The LysXaaXaaArg motif identified as required for processing of proIGF-I [21] has been conserved in the D/E-domain junction of all IGF-II sequences except shark IGF-II, where Glu has been substituted for Lys. Given the importance of basic residues for prohormone processing, it is surprising to find that an acidic residue has been substituted for Lys. However, it is possible that the shark proIGF-II processing enzyme may still cleave this precursor, or that shark proIGF-II is secreted unprocessed but still retains appropriate biological activity. Residues that are conserved in all IGF-II sequences but not found in the IGF-I E-domains are Gln^{100}, Leu^{102}, $Pro^{107}Ala^{108}$, and Ala^{124}.

Our results show that shark liver contains two distinct IGFs, one homologous to IGF-I and the other homologous to IGF-II. This indicates that the ancestral IGF gene duplicated before chondrichthyes diverged from the tetrapod lineage, more than 440 million years ago. Furthermore, sequence comparisons demonstrate that IGF-I and IGF-II have evolved slowly during vertebrate evolution. Thus, gnathostome IGF-I sequences are more similar to each other than to IGF-II, and gnathostome IGF-II sequences are more similar to each other than to IGF-I (Tables 1 and 2). On the other hand, hagfish IGF has characteristics of gnathostome IGF-I and IGF-II. Like IGF-I sequences, hagfish IGF has a longer C-domain, a methionine residue at position 59 of the A-domain, and a short E-domain. However, hagfish IGF lacks the GlyTyr sequence found at the beginning of all IGF-I C-domains. It also contains the invariant glutamate found at position 44 of the A-domain of all IGF-II sequences, has the 6 amino acid D-domain characteristic of IGF-II molecules, and shows slightly higher sequence identity to human, trout, and shark IGF-II than IGF-I. As noted above, IGF-I sequences contain a $Leu/PheArgArgLeu^{57}$ motif in the A-chain while IGF-II sequences do not have basic residues in this area but tend to be hydrophobic. The homologous region in the hagfish A-domain is intermediate: LeuArgLeuLeu. This analysis indicates that hagfish IGF contains features characteristic of and distinct to both IGF-I and IGF-II, while elasmobranchs, teleosts, amphibian, birds and mammals contain IGFs which are clearly either IGF-I or IGF-II molecules, thus supporting the hypothesis that hagfish IGF may predate the ancestral molecule that duplicated and diverged to generate gnathostome IGF-I and IGF-II [22].

Acknowledgements: We would like to thank Paul Gardner for oligonucleotide synthesis and Jie Lai-Zhang for technical assistance. This work was supported by the Howard Hughes Medical Institute and in part by NIH Grants DK13914 and DK20595.

References

[1] Jones, J.I. and Clemmons, D.R. (1995) Endocr. Rev. 16, 3–34.
[2] Salmon Jr., W.D. and Daughaday, W.M. (1957) J. Lab. Clin. Med. 49, 825–836.
[3] Froesch, E.R., Burgi, H., Ranseier, E.B., Bally, P. and Labhart, A. (1963) J. Clin. Invest. 42, 1816–1834.
[4] Temin, H.M. (1966) J. Nat. Cancer Inst. 37, 167–175.
[5] Rinderknecht, E. and Humbel, R.E. (1978) J. Biol. Chem. 253, 2769–2776.
[6] Rinderknecht, E. and Humbel, R.E. (1978) FEBS Lett. 89, 283–286.
[7] Daughaday, W.H. and Rotwein, P. (1989) Endocr. Rev. 10, 68–91.
[8] Chan, S.J., Cao, Q.-P. and Steiner, D.F. (1990) Proc. Natl. Acad. Sci. USA 87, 9319- 9323.
[9] Chan, S.J., Emdin, S.O., Kwok, S.C.M., Kramer, J.M., Falkmer, S. and Steiner, D.F. (1981) J. Biol. Chem. 256, 7595–7602.
[10] Nagamatsu, S., Chan, S.J., Falkmer, S. and Steiner, D.F. (1991) J. Biol. Chem. 266, 2397–2402.
[11] Jansen, M., van Schaik, F.M.A., Ricker, A.T., Bullock, B., Woods, D.E., Gabbay, K.H., Nussbaum, A.L., Sussenbach, J.S. and Van den Brande, J.L. (1983) Nature 306, 609–611.
[12] Bell, G.I., Merryweather, J.P., Sanchez-Pescador, R., Stempien, M.M., Priestley, L., Scott, J. and Rall, L.B. (1984) Nature, 310, 775–777.
[13] Kajimoto, Y. and Rotwein, P. (1989) Mol. Endocrinol. 3, 1907–1913.
[14] Kallincos, N.C., Wallace, J.C., Francis, G.L. and Ballard F.J. (1990) J. Endocrinol. 124, 89–97.
[15] Kajimoto, Y. and Rotwein, P. (1990) Mol. Endocrinol. 4, 217–226.
[16] Shamblott, M.J. and Chen, T.T. (1993) Proc. Natl. Acad. Sci. USA 89, 8913–8917.
[17] Chirgwin, J.M., Przbyla, A.E., McDonald, R.J. and Rutter W.J. (1979) Biochemistry 18, 5294–5299.
[18] Frohman, M.A., Dush, M.K. and Martin G.R. (1988) Proc. Natl. Acad. Sci. USA 85, 8998–9002.
[19] Rotwein, P. (1986) Proc. Natl. Acad. Sci. USA 83, 77–81.
[20] Duguay, S.J., Park, L.K., Samadpour, M. and Dickhoff, W.W. (1992) Mol. Endocrinol. 6, 1202–1210.
[21] Duguay, S.J., Lai-Zhang, J. and Steiner, D.F. (1995) J. Biol. Chem. in press.
[22] Chan, S.J., Cao, Q.P., Nagamatsu, S. and Steiner, D.F. (1993) Biochemistry and Molecular Biology of Fishes (P.W. Hochachka and T.P. Mommsen, eds.) vol. 2, pp. 407–417, Elsevier Science Publishers, New York.