J. Biochem. 119, 898-905 (1996)

Isolation and Characterization of the Human Inter-α-Trypsin Inhibitor Family Heavy Chain-Related Protein (IHRP) Gene (ITIHL1)

Ken-ichi Saguchi, Takashi Tobe,[1] Ken Hashimoto, Yuta Nagasaki, Eiichi Oda, Yasuko Nakano, Nam-Ho Miura, and Motowo Tomita

Department of Physiological Chemistry, School of Pharmaceutical Sciences, Showa University, 1-5-8 Hatanodai, Shinagawa-ku, Tokyo 142

Received for publication, December 12, 1995

Inter-α-trypsin inhibitor (ITI) family heavy chain-related protein (IHRP) is a novel human glycoprotein that shows significant homology in amino acid sequence to proteins of the ITI family heavy chains from human plasma. Three overlapping clones that encode the human inter-α-trypsin inhibitor family heavy chain-related protein (IHRP) gene (ITIHL1) were isolated and characterized. The IHRP gene spans 15 kb and is composed of 24 exons from 27 to 207 bp in size with consensus splice sites. The gene codes for the precursor of IHRP, which is similar to inter-α-trypsin inhibitor (ITI) family heavy chains. Two major transcription initiation sites were identified in the 5′-flanking region. They contain putative promoter elements, but no typical TATA box. Some exons of this gene showed significant similarities to those of the ITI-H1 gene in nucleotide length and in intron phasing. The tissue-specific transcription of this gene may be due to the presence of binding sites for the hepatocyte nuclear factors LF-A1, HNF-5, NF-IL6, and C/EBP. This gene was found to be localized very close to another unknown gene related to EST (GenBank accession #: R54643, R50663, R50563, H27139, and R54913).

Key words: IHRP gene, TATA-less promoter.

Three members of the inter-α-trypsin inhibitor (ITI) family (serine protease inhibitors) have been purified from human plasma. The first is inter-α-trypsin inhibitor, which consists of heavy chain 1 (HC1), heavy chain 2 (HC2), and bikunin (*1*). The second is pre-α-trypsin inhibitor, which consists of heavy chain 3 (HC3) and bikunin (*2*). The third is a complex of HC2 and bikunin (*3*). In all of them, heavy chains (HC1, HC2, and HC3) and bikunin form complexes *via* a chondroitin sulfate chain (*3-6*). Bikunin, which has two Kunitz-type serine protease inhibitor domains, was first identified in urine as a trypsin inhibitor (*7, 8*). The functions of the various proteins belonging to the ITI family, including the ITI heavy chains, are not yet precisely known. One of them, a 150 kDa protein, has been reported to stabilize the cumulus extracellular matrix (*9, 10*). The members of the ITI family are thought to be plasma serine proteinase inhibitors in mammals, because bikunin has two Kunitz domains (*8*). The heavy chains of the ITI family members might act as regulators of bikunin activities in the ITI family because the serine protease inhibitor activity of bikunin alone is stronger than that of the ITI family members (*11*).

Inter-α-trypsin inhibitor (ITI) family heavy chain-related protein (IHRP) is a novel human glycoprotein that shows significant homology in amino acid sequence to the ITI family heavy chains from human plasma (*12, 13*). The cloning of the cDNA encoding IHRP has revealed that IHRP is identical to PK-120 reported by Nishimura *et al.* (*14*).

The human IHRP gene, which is named ITIHL1, is located in the p21-p14 region of chromosome 3, where the genes of heavy chain 1 and 3 of inter-α-trypsin inhibitor have been reported to be localized (*15*). The expression of IHRP in human tissues is restricted to liver (*15*). The heavy chains of ITI are also predominantly found in liver (*16*). Although IHRP has similarities in many respects to the heavy chains of the ITI family, IHRP does not form a complex with bikunin. Conversely, no ITI heavy chains have been reported to be present in free form in human plasma.

Based on the similarity of amino acid sequence, predominant expression in liver, and gene localization, we have proposed that IHRP is a member of the evolutionarily related gene family of ITI heavy chains (*15*). To clarify this hypothesis, we isolated and characterized the gene for IHRP. In the present study, we determined the exact number of exons and obtained a detailed map of the 5′-flanking sequence. The structure of IHRP gene (ITIHL1) was also compared with that of the ITI heavy chain H1 gene reported by Bost *et al.* (*17*).

MATERIALS AND METHODS

Reagents—Restriction endonucleases, T4 phage DNA ligase, recombinant Taq DNA polymerase, Molony murine leukemia virus reverse transcriptase, and pUC119 vector were purchased from Takara Shuzo (Otsu). Nitrocellulose

[1] To whom correspondence should be addressed. E-mail: ttobe@pharm.showa-u.ac.jp

Abbreviations: ITI, inter-α-trypsin inhibitor; ITI-H1 and H3, inter-α-trypsin inhibitor heavy chain 1 and heavy chain 3; IHRP, inter-α-trypsin inhibitor family heavy chain-related protein; EST, expression sequence tag.

filters (BA85) were purchased from Schleicher and Schull. [α-^{32}P]dCTP, [γ-^{32}P]ATP, and nylon filters were obtained from Amersham Japan. A random-primer labeling kit was obtained from Du Pont-New England Nuclear.

Screening of the Human Genomic Library—The human genomic DNA library obtained from Clontech (prepared by partial digestion of human placental DNA with *Sau*3A and inserted into the *Bam*HI site of λ phage EMBL3) was screened by *in situ* plaque hybridization with an IHRP cDNA probe labeled with ^{32}P by the random-primer method (*18*). Hybridization was carried out in a solution containing 5×SSC, 5×Denhardt's, 50 mM Tris-HCl, pH 7.5, 100 μg/ml of denatured salmon sperm DNA, 0.1% SDS, 10% dextran sulfate, and 50% formamide at 42°C. The membranes were washed with 2×SSC containing 0.1% SDS at room temperature and then with 0.1×SSC containing 0.1% SDS at 50°C (*19*). Several rounds of screening were carried out to obtain positive clones.

Characterization of the Positive Clones—Positive clones were analyzed by Southern blotting. Genomic clones were completely digested with a set of restriction enzymes, electrophoretically separated on a 0.7% agarose gel, and transferred onto nylon filters. The filters were hybridized with three different probes corresponding to the 5′, central, and 3′ regions of the IHRP cDNA sequence.

DNA Sequence Analysis—DNA inserts were characterized by partial restriction enzyme mapping and then subcloned into pUC119 vector. DNA sequencing was performed by using the dideoxy-chain-termination method with a Shimadzu model DSQ1000 DNA sequencer (*20*). The sequences were aligned and decoded with the Genetyx analysis program (SDC, Tokyo).

Primer Extension—A synthetic oligonucleotide (P1) corresponding to nucleotides 26 to 65 (5′-CTGCAGGTACGGACAGGCCTTGGGGGCTTCATCGTGGCTC-3′) in the 5′-end region of the IHRP cDNA was labeled with [γ-^{32}P]-ATP using polynucleotide kinase. The probe (2×10^6 cpm) was annealed to 10 μg human liver poly(A)-rich RNA from HepG2 cells for 16 h at 50°C in 15 μl of hybridization buffer containing 400 mM NaCl, 40 mM Pipes pH 6.4, 1 mM EDTA, and 80% formamide. The annealed RNA/primer mixture was precipitated and resuspended in 20 μl of buffer containing the following: 50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 10 mM dithiothreitol, actinomycin D (50 μg/ml), and 0.5 mM deoxyribonucleoside triphosphate, and extended with 100 units of Moloney murine leukemia virus reverse transcriptase at 42°C for 60 min. The RNA strand was then digested with 500 ng of DNase-free RNase A in the presence of 16 μM EDTA for 45 min at 37°C. Ammonium acetate was added to give a final concentration of 2 M, and the sample was extracted with phenol-chloroform-isoamyl alcohol (25 : 24 : 1). After precipitation with two volumes of ethanol and lyophilization, the sample was dissolved in 1 mM EDTA, 10 mM Tris-HCl, pH 8.0. The primer extension product was analyzed by acrylamide gel electrophoresis in parallel with DNA sequence reactions of the 5′ genomic subclone. These DNA sequence reactions were prepared using the same 40-bp oligonucleotide primer.

RESULTS

Approximately 5×10^5 clones from a human genomic DNA library were screened by *in situ* plaque hybridization using the full-length human IHRP cDNA as a probe. After successive purification of hybridization-positive plaques, 5 positive clones (C2, C3, C4, C7, and C24) were isolated. The radiolabeled cDNA fragments corresponding to the 5′, central, and 3′ regions of the IHRP cDNA sequence were used as probes to cross-hybridize with these clones. Three

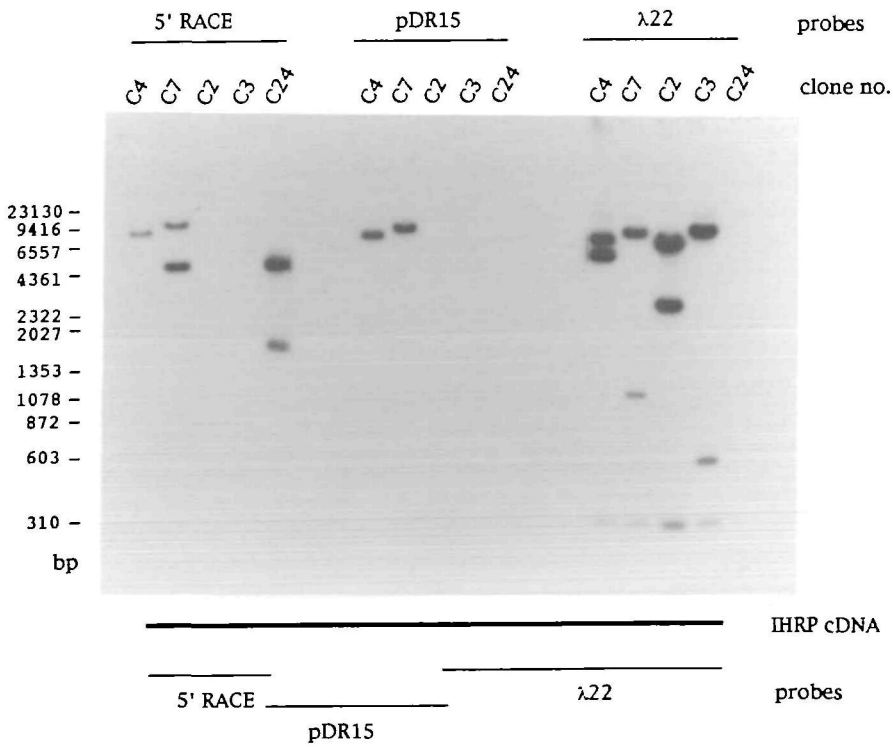

Fig. 1. **Southern blot analysis of genomic clones.** The isolated five genomic clones were hybridized with three different probes which correspond to the 5′, central, and 3′ regions of the IHRP cDNA. The phage DNAs were digested with *Bam*HI and *Sal*I, and separated by electrophoresis on a 1% agarose gel. The DNA fragments were transferred onto nylon filters and hybridized with the probes separately. The cDNA fragment generated by 5′ RACE, the inserted cDNA of clone pDR 15, and clone λ22 were used as probes for the 5′, central, and 3′ regions, respectively. The positional relationship of these probes is schematically presented at the bottom.

clones (C4, C7, and C24) were selected and further analyzed by restriction enzyme mapping and subcloning (Fig. 1). The results of this mapping and Southern blot hybridization demonstrated that the inserts of clones C24, C4, and C7 overlapped each other and that together they contained the complete coding sequence of IHRP (Fig. 2), plus additional genomic DNA extending 10.4 kb upstream and 3 kb downstream from the gene's exonic boundaries. The sequences of overlapped regions between these clones were confirmed to be identical. These clones span a region of 27 kb. A complete set of subclones was prepared to sequence each genomic clone. The exon-intron boundaries of the protein-coding region in the IHRP gene were identified by comparison with the previously reported cDNA sequence (*13*). The putative end of the last exon, exon 24, was defined as the last base of the human IHRP cDNA clone 22 previously described (*13*). All splice junction sequences conformed to the GT-AG rule for the nucleotides immediately flanking the exon borders (*21, 22*). The IHRP gene, approximately 15 kb in length, was organized into 24 exons interrupted by 23 introns (Fig. 2). In all, more than 90% of the IHRP gene was sequenced. A comparison with the cDNA sequence (3.0 kb) revealed that the overall coding capacity for this gene is about 20%. Comparison of the exon-intron organization with the protein structure revealed that the peptide-coding region of IHRP began at the beginning of exon 1 and ended at the beginning of exon 24 (Fig. 2). Exons range in size from 27 bp (exon 14 and 19) to 207 bp (exon 24). Introns vary in length from 80 to 2,130 nucleotides. The size of some introns was determined by restriction enzyme digestion or PCR; their average size was about 490 bp (Fig. 3).

To determine the transcription initiation site of the IHRP gene, primer extension experiments were performed using RNA from HepG2 cells and an oligonucleotide primer complementary to the IHRP mRNA. For the primer, we used a 40-base-long oligonucleotide complementary to the mRNA sequence that contained eight 5′ untranslated nucleotides followed by 32 nucleotides in the translated region. The position of the transcriptional initiation site was determined by analyzing the primer extension product in parallel with DNA sequence analysis of the 2.3-kb genomic DNA fragment upstream of the translation initiation site. The pattern of extension is shown in Fig. 4. Two major primer extension products of 62 and 56 bp (including primer) were detected (Fig. 4). Alignment of the longer one of the IHRP primer extension products with the sequence of the IHRP genomic DNA demonstrated that the human IHRP mRNA initiated at sequence AGAAGCCT, which is 30 nucleotides upstream from the mRNA translation initiation codon. This observation was consistent with the fact that gene transcription tends to initiate at a purine nucleotide (*23*). Therefore, we defined this adenine residue as the transcription initiation site, implying that it is the 5′ end of exon 1. The 22-nucleotide sequence in the IHRP mRNA determined by primer extension was further confirmed by sequence analysis of the cloned IHRP gene. This experiment indicated that the isolated genomic DNA fragment contained the promoter region of the IHRP gene.

To analyze the promoter region of the IHRP gene, we sequenced the 2.3 kb upstream 5′ flanking region that contained no consensus acceptor splice site. Analysis of the 5′ flanking sequence of the IHRP gene revealed that the gene had no TATA box (Fig. 5). The G+C content in the 5′ flanking region, calculated to be 57%, was not very high. This is different from other TATA-less promoters such as those found in the adenosine deaminase gene, the 3-hydroxy-3-methylglutaryl coenzyme A reductase gene, and the hypoxanthine-guanine phosphoribosyltransferase gene (*24-27*). By inspection, several regulatory element motifs were identified in this region of the IHRP gene (Fig. 5). The HNF-5 element was located at nucleotide −121, and four potential NF-IL6 elements 258, 515, 1,266, and 1,453 bp upstream of the putative cap site were identified. Consensus transcription factor LF-A1 sequence motifs were found at nucleotides −609, −1084, −1126, and −1774. In addition, there were two C/EBP motifs and one Sp1 site, as shown in Fig. 5.

Comparison of the organization of the IHRP gene with that of the human inter-α-trypsin inhibitor family heavy chain 1 (ITI-H1) gene revealed that the organization of the IHRP gene showed several striking similarities to the ITI-H1 gene (*17*). The structures of the genes of other ITI heavy chains (H2 and H3) are unknown so far. Like the ITI-H1 gene promoter, the IHRP gene did not have a typical TATA box in the putative promoter region. The twelve

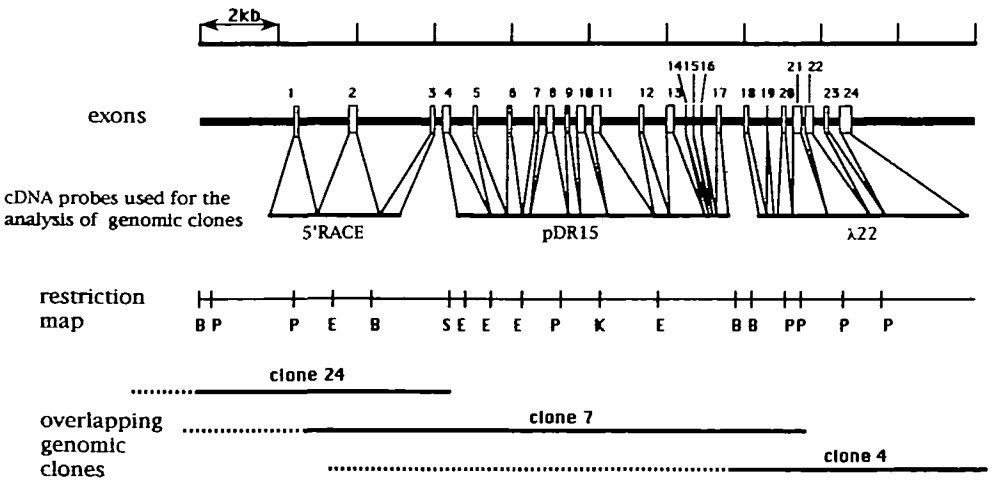

Fig. 2. **The genomic organization of the IHRP gene.** The boxes representing exons are numbered; filled-in boxes represent untranslated regions, whereas open boxes represent translated regions. The locations of recognition sites for several restriction enzymes are shown under the complete gene structure (B, *Bam*HI; P, *Pst*I; E, *Eco*RI; K, *Kpn*I). The corresponding exons to the cDNA probes (5′ RACE, pDR 15, and λ22) used for the analysis of genomic clones in Fig. 1 are schematically shown. The three overlapping phage clones comprising the IHRP gene are shown at the bottom.

EXON (bp)	5' Donnor Sequence	Intron (kb)	EXON	3' Acceptor Sequence
	120			121
1	ACTACTGCCGAAAAG gtagccagtg	I (1.40)	2	tcaagatag AATGGCATCGACATC
(123)	T T A E K			N G I D I
	281			282
2	CACCAACTTCTCCAT gtaggtgcct	II (2.13)	3	tgcccccag GAACATCGATGGCAT
(161)	T N F S M			N I D G M
	386			387
3	CGCTGGCCTCGTCAA gtgagctggt	III(0.13)	4	tctccccag GGCCACCGGGAGAAA
(105)	A G L V K			A T G R N
	549			550
4	GTCAAGCACCTGCAG gtacctgctg	IV (0.14)	5	tgcctccag ATGGACATTCACATC
(163)	V K H L Q			M D I H I
	660			661
5	CAGAATAAGACCAAG gtgggcctkt	V (0.35)	6	tccttccag GCTCACATCCGGTTC
(110)	Q N K T K			A H I R F
	789			790
6	GGGGGCTCCATTCAG gtgggtggac	VI (1.07)	7	tggccacag ATCGAGAACGGCTAC
(130)	G G S I Q			I E N G Y
	906			907
7	AGGAAAATCCAGCAG gtaggtcctg	VII (0.28)	8	cctgaccag ACCCGGGAAGCCCTA
(117)	R K I Q Q			T R E A L
	1075			1076
8	TCCAGGCCCTGGGAG gtaagtgtct	VIII (0.14)	9	gttcctcag GGACCAACATCAATG
(169)	Q A L G G			T N I N D
	1201			1202
9	GCGACCCCACTGTGG gtnagggccc	IX (0.18)	10	tgggtccag GGGAGACTAACCCCA
(126)	D P T V G			E T N P R
	1383			1384
10	GCCCTGCAGCTCCAG gtgccagttt	X (0.08)	11	accctgcag GACTTCTACCAGGAA
(182)	A L Q L Q			D F Y Q E
	1569			1570
11	GTCAGTGGGAAGCTG gtgagtgtgg	XI (1.30)	12	tctttccag CCTACACAGAACATC
(186)	V S G K L			P T Q N I
	1709			1710
12	GCTGCTGGAGCAAAC gtgagtcatc	XII (0.28)	13	aacctacag TGTCTCCGCATCCGA
(140)	L L E Q T			V S A S D
	1864			1865
13	AGCCCATGGAAGGCG gtgggtgtgg	XIII(0.31)	14	tttatccag AAAGTAGAAACAGGA
(155)	P M E G E			S R N R N
	1891			1892
14	GGAATGTCCACTCAG gtaagcagcg	XIV (0.52)	15	gatttccag GTTCCACTTTCTTCA
(27)	N V H S G			S T F F K
	1942			1943
15	AAATACCAAAACCAG gtgacaaagt	XV (0.15)	16	accacacag AGGCTTCCTTTTCTC
(51)	I P K P E			A S F S P
	1981			1982
16	GATGGAATAGACAAG gtagtttgag	XVI (0.23)	17	aaggctgag CTGGAGCTGCTGGCT
(39)	W N R Q A			G A A G S
	2107			2108
17	GCCGTCTGGCCATCT gtaagctccc	XVII(0.65)	18	tctccaaag TGCCTGCTTCAGCAC
(126)	R L A I L			P A S A P
	2182			2183
18	ATATGAAAATCGAAG gtgcccccc	XVIII(0.47)	19	ycctyymag AAACAACCATGACAA
(75)	M K I E E			T T M T T
	2209			2210
19	CAACCCAAACCCCAG gtaagtcccc	XIX (0.08)	20	tgttcccag CCCCCATACAGGCTC
(27)	T Q T P A			P I Q A P
	2326			2327
20	CAGACCCTGAGCAAG gtgagagggg	XX (0.49)	21	tcctcccag GGGTTGAGGTGACTG
(117)	D P E Q G			V E V T G
	2501			2502
21	CTCAGTGATGCCCGG gtaaggccca	XXI (0.47)	22	tttccacag CCTGAAGATGACCAT
(175)	S V M P G			L K M T M
	2656			2657
22	TTGGAGGGACCCTTG gtactgagcc	XXII(0.13)	23	gtatctcag GCCAGTTTTACCAGG
(155)	G G T L G			Q F Y Q E
	2753			2754
23	CCACTCTGCCACCAG gtgacttgga	XXIII(0.48)	24	ccatttcag AGAGCGCAGGCTGGA
(97)	H S A T R			E R R L D
				2962
24	----------ACTCATTACAAATAAAGAAAGGTGGTGTGAGCCTGGGAAGTGGGT			
(207)				

Fig. 3. **Exon/intron structure of human IHRP gene.** Intron (lowercase letters) and exon (capital letters) junction boundaries of the 24 exons and 23 introns are displayed. Approximate intron sizes were determined by restriction enzyme mapping, PCR, or direct sequencing. The polyadenylation signal is underlined, and the arrow indicates the site of poly(A) addition.

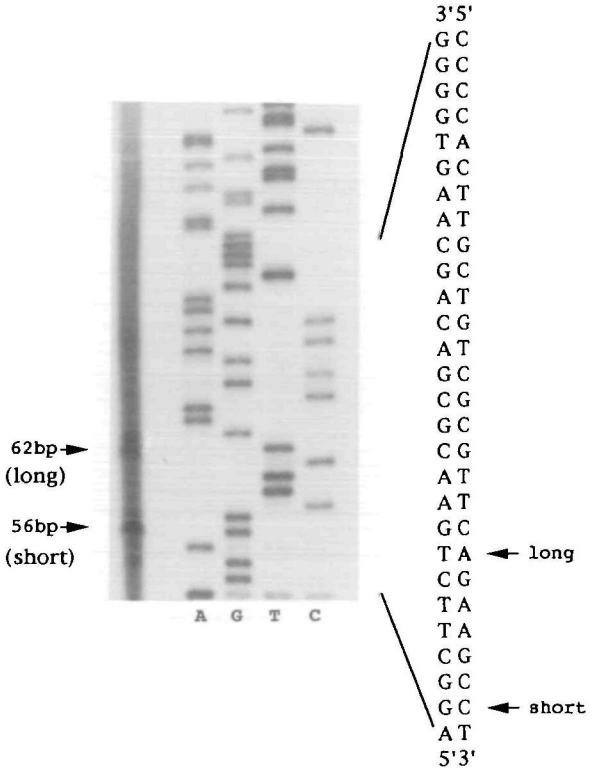

Fig. 4 **Mapping the 5′ end of the human IHRP mRNA.** A 40 bp end-labeled oligonucleotide primer (nucleotides 26-65 of the IHRP cDNA) was hybridized with HepG2 RNA. Hybrids were used as substrates for reverse transcriptase. Extension products were analyzed by electrophoresis in denaturing polyacrylamide gels along-side a sequence ladder of IHRP genomic DNA prepared using the same primer. A section of the nucleotide sequence of both strands is presented and the proposed transcription initiation sites (long and short) are marked with an arrow.

exons from 2 through 13 of the IHRP gene were very similar to those (from 3 through 14) of the ITI-H1 gene in nucleotide length and in intron phasing (Fig. 6).

DISCUSSION

We have previously reported the isolation and sequencing of cDNA clones encoding human IHRP (*13*). The amino acid sequence of IHRP predicted from the nucleotide sequence of its cDNA shows significant homology to those of the inter-α-trypsin inhibitor (ITI) family heavy chains. In addition, the IHRP gene is located in the p21-p14 region of chromosome 3, where the genes of heavy chains 1 and 3 of ITI have been reported to be localized. Because of the similarities of amino acid sequence, predominant expression in liver, and gene localization, we have proposed that IHRP is a member of the evolutionarily related gene family of ITI heavy chains (*15*).

In this study, we characterized the human IHRP gene. Three overlapping genomic clones, C4, C7, and C24, covered the entire IHRP gene. The human IHRP gene spans 15 kb and contains 24 exons from 27 to 207 bp in size, with the translational initiation codon residing in exon 1 and the stop codon residing in exon 24. The average size of the introns (ranging from 75 to 2,130 bp) is calculated to be 490 bp and is much smaller than that encountered in vertebrates (*28*); this is particularly true of introns 10 and 19, that are 79 and 75 nucleotides in length, respectively. This makes the genomic IHRP a compact gene, with a high coding/non-coding ratio. The coding capacity of this gene is calculated to be about 20%.

The primer extension experiment demonstrated that the transcription of the IHRP gene is initiated from the adenosine residue at nucleotide 1 in Fig. 5. As shown in Fig. 5, a short primer extension product was also detected, which is 56 bp in length, suggesting that the gene of IHRP contains two transcription initiation sites in its promoter region. The nucleotide sequence surrounding the putative transcription initiation site of the short product seems not to indicate the premature termination by forming secondary structure. Although the possibility that the short primer extension product is caused by premature termination can not be completely excluded, the transcription of the IHRP gene is thought to be mainly initiated from two different points. The transcription usually starts at several points in TATA-less promoters.

The nucleotide sequence of the 5′ end of IHRP cDNA, which was confirmed by sequencing the genomic clone, is different from that of PK-120 cDNA reported by Nishimura *et al.* (*14*). One possibility to explain this discrepancy is alternative splicing of the transcript from this gene, suggesting the existence of more upstream exon(s). However, we could not find this nucleotide sequence, which is identical to that of the 5′ end of PK-120 cDNA, within 2.3 kb upstream from the exon 1 proposed by us. At this time, the reason why the nucleotide sequences of the 5′ end of the two cDNA clones are different from each other is unknown. Comparison of the nucleotide sequence of IHRP cDNA with that of the cDNA for PK-120 showed single nucleotide differences at four positions (nt 105: C to T, nt 219: G to T, nt 284: A to T, and nt 371: A to G, in our numbering system). Two of these (nt 105 and nt 219), which are silent, might be due to polymorphisms. The other positions (nt 284 and nt 371) alter the predicted amino acid sequence, Asn to Ile and Asn to Ser, respectively. The analysis of the genomic clone revealed that nt 371 was G, suggesting a sequencing error in our cDNA sequence. Whereas nt 284 was A, identical to that of IHRP cDNA, nt 285 was different from that of the two cDNAs (C to A). From the codons containing these nucleotides, Asn is predicted from the IHRP cDNA, Ile from PK-120 cDNA, and Lys from the genomic clone. Further study is needed to define this codon.

Analysis of the 5′ flanking region of the IHRP gene revealed the absence of a typical TATA element. The promoters of genes that are regulated without the usual TATA or CCAAT sequences have been divided into two classes (*29*). The first has generally been associated with constitutively active housekeeping genes that have G+C-rich promoters with CpG islands, Sp1 binding sites, and often multiple start points (*24, 30*). The second group includes promoters that also lack a TATA element and are not G+C-rich. Smale and Baltimore have suggested that the second class of promoters is differentially or developmentally regulated; they describe a 17-bp motif, the initiator (Inr), that includes the transcription initiation site and is sufficient for basal transcription of the lymphocyte-specific terminal deoxyribonucleotidyl transferase gene (*29*). Another transcription initiator element (HIP1) has

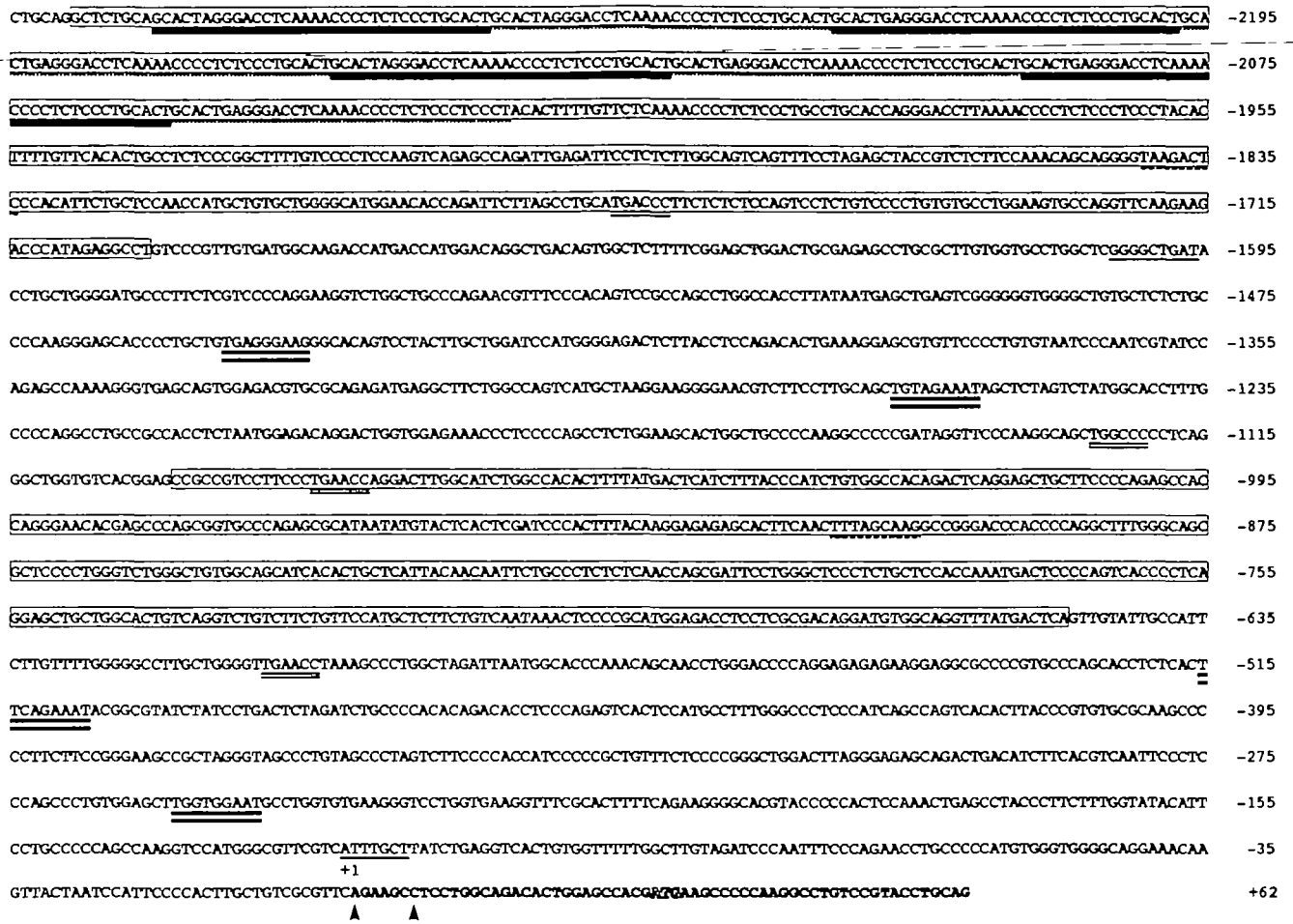

Fig. 5. **DNA sequence of the 5′ flanking region of the IHRP.** The transcribed sequence is in boldface type, and the translation initiation codon is shown by outline letters. The proposed mRNA cap sites are indicated by an arrowhead. The numbering uses the transcription start site upstream of exon 1 as the reference point. The putative Sp1 site (nt −1586) and HNF-5 element (nt −121) are underlined, while LF-A1 (nt −609, −1084, −1126, and −1774) sites are double-underlined. The NF-IL6 elements and C/EBP sites are indicated by a thick double underline and dotted underline. The tandem repeat of thick underline and dotted underline represents the 34 nucleotide repeat. The putative exons of the unknown gene are boxed.

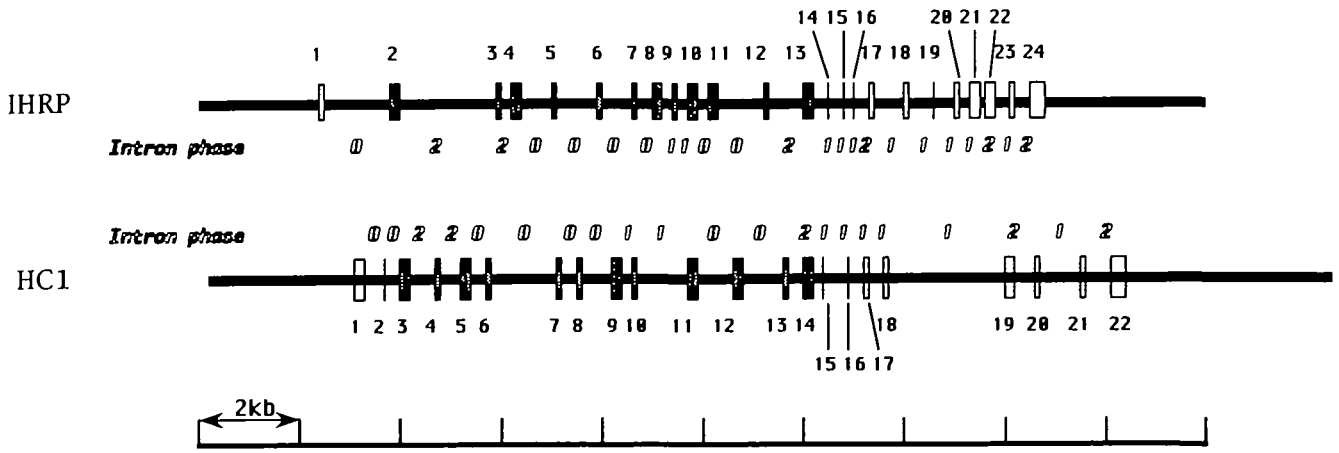

Fig. 6. **Comparison of human IHRP and ITI-H1 genes.** The structures of human IHRP and ITI-H1 genes are graphically shown. The shaded boxes represent the exons, which are similar in nucleotide length and intron phasing between the two genes. The open boxes represent the exons, which are not similar in nucleotide length and intron phasing between two genes. The exons are numbered and outline numbers represent the intron phasing. The nucleotide scale is shown at the bottom. HC1: ITI heavy chain 1.

been identified in the G+C-rich promoter of the dihydrofolate reductase gene (*31*). Although, the 5′ flanking region of the IHRP gene contains neither the consensus sequence for initiation nor the HIP1 element, the IHRP gene seems to be classified into the second class of TATA-less promoters. Furthermore, the reports that the pig IHRP gene was induced by acute-phase reactions supports this classification (*32, 33*). The transcription initiation sites in TATA-less promoters appear to be due to a conjunction of three types of factors. The first factor is local sequence elements or cap sequences which may in some cases be an initiator (*29, 34, 35*) and may require a bound protein factor (*36-38*). The second is the inducing effect of a factor like Sp1 at a defined distance, and the third is one or more activating factors which bind to the regulatory elements of the promoter and whose effect extends uniformly to the whole start site region. With regard to the second factor, it is possible that the local sequence −1586 to −1594 acts as an initiator, but it is very far from the transcription initiation site of the IHRP gene. When this activity is outside the locus, such factors can be initiators that have some activity independent of both Sp1 and TATA, but further analyses by mutation, footprinting, and *in vitro* transcription with purified factors will be needed to clarify this.

The restricted tissue-specific transcription of the IHRP gene might be due to the presence of binding sites for hepatocyte nuclear factor HNF-5 and to the oct-1 patterns, which are known to be involved in the specific expression of the genes in certain tissues by specifically fixing the OTF-2 transcription factor (*39-41*). The putative IL-6 responsive element identified in the 5′-flanking region of the ITI heavy chain H1 gene, NF-IL6 elements, C/EBP sites, and LF-A1 elements are also present in the 5′-flanking region of the IHRP gene (Fig. 5).

Interestingly, the 5′ flanking region of the IHRP gene reveals some characteristic features (Fig. 5). Thus, homology searches reveal that an unknown gene is located in this region. The nucleotide sequences of five cDNA clones (accession numbers R54643, R50663, R50563, H27139, and R54913) are completely included in this region. These nucleotide sequences have been registered as EST in GenBank, and were cloned from the Soares breast 2NbHBst cDNA library by Hillier *et al.* Because these nucleotide sequences are separated into two blocks on our genomic sequence, the two regions are predicted to be exons of a certain gene. The downstream exon contains a putative polyadenylation signal near the 3′ end. The eight direct repeats of 34 nucleotides are present in the upstream exon, which is about 2 kb upstream of the transcription initiation site of the human IHRP gene. The transcription initiation site of the human IHRP gene is only about 800 nucleotides distant from the 3′ end of the downstream exon containing the putative polyadenylation signal.

Although the exon number of the IHRP gene is different from that of the ITI-H1 gene, a structural comparison of the human IHRP gene with that of the human ITI-H1 gene demonstrates that these genes have many significant similarities to each other (Fig. 6). The sizes of exons 2 through 13 of the IHRP gene are almost identical to those of exons 3 through 14 of the ITI-H1 gene. Although the introns of the IHRP gene are quite different in nucleotide length from the corresponding introns of ITI-H1 gene, their phasings are identical. In both genes, the coding triplet is interrupted between the second and third nucleotides in introns 2, 3, and 12 of the IHRP gene, and in introns 3, 4, and 13 of the ITI-H1 gene. In introns 8 and 9 of the IHRP gene and the corresponding introns of ITI-H1 gene, the coding triplet is interrupted between the first and second nucleotides. Other introns in this region have no interruption in the coding triplet. The conserved amino acid sequences between IHRP and ITI-H1 are coded in these exons. The first exon of both genes corresponds to the 5′ untranslated region and a part of the signal peptide, while the last exon, the 24th exon of IHRP gene and the 22nd exon of ITI-H1 gene, contains the stop codon and the 3′ untranslated region. Both genes are relatively compact, because their coding capacity is about 20%. Although the exon-intron organization of IHRP gene is very similar to that of ITI-H1 gene, the nucleotide sequence of the 5′ flanking region of IHRP gene is quite different from that of ITI-H1 gene.

The similarities observed in the gene structures, together with the previous finding that they are located very close together on chromosome 3 (*15*), strongly indicate that at least the upstream halves of the genes are derived from the same ancestral gene by gene duplication.

REFERENCES

1. Jessen, T., Faarvang, K.L., and Plong, M. (1988) Carbohydrate as covalent crosslink in human inter-alpha-trypsin inhibitor: A novel plasma protein structure. *FEBS Lett.* **230**, 195-200
2. Enghild, J.J., Thogersen, I.B., Pizzo, S.V., and Salvesen, G. (1989) Analysis of inter-alpha-trypsin inhibitor and a novel trypsin inhibitor, pre-alpha-trypsin inhibitor, from human plasma. *J. Biol. Chem.* **264**, 15975-15981
3. Enghild, J.J., Salvesen, G., Thogersen, I.B., Valnickova, Z., Pizzo, S.V., and Hefta, S.A. (1993) Presence of the protein-glycosaminoglycan-protein covalent cross-link in the inter-alpha-inhibitor-related proteinase inhibitor heavy chain 2/bikunin. *J. Biol. Chem.* **268**, 8711-8716
4. Chirat, F., Balduyck, M., Mizon, C., Laroui, S., Sautiere, P., and Mizon, J. (1991) A chondroitin-sulfate chain is located on serine-10 of the urinary trypsin inhibitor. *Int. J. Biochem.* **23**, 1201-1203
5. Enghild, J.J., Salvesen, G., Hefta, S.A., Thogersen, I.B., Rutherfurd, S., and Pizzo, S.V. (1991) Chondroitin 4-sulfate covalently cross-links the chains of the human blood protein pre-α-inhibitor. *J. Biol. Chem.* **266**, 747-751
6. Morelle, W., Capon, C., Balduyck, M., Sautiere, P., Kouach, M., Michalski, C., Fournet, B., and Mizon, J. (1994) Chondroitin sulphate covalently cross-links the three polypeptide chains of inter-alpha-trypsin inhibitor. *Eur. J. Biochem.* **221**, 881-888
7. Balduyck, M., Mizon, C., Loutfi, H., Richet, C., Roussel, P., and Mizon, J. (1986) The major human urinary trypsin inhibitor is a proteoglycan. *Eur. J. Biochem.* **158**, 417-422
8. Hochstrasser, K., Schongerger, O.L., Rossmanith, I., and Wachter, E. (1981) Kunitz-type proteinase inhibitor derived by limited proteolysis of the inter-α-trypsin inhibitor. *Hoppe-Seyler's Physiol. Chem.* **362**, 1357-1362
9. Chen, L., Mao, S.J. T., and Larsen, W.J. (1992) Identification of a factor in fetal bovine serum that stabilizes the cumulus extracellular matrix: A role for a member of the inter-alpha-trypsin inhibitor family. *J. Biol. Chem.* **267**, 12380-12386
10. Castillo, G.M. and Templeton, D.M. (1993) Subunit structure of bovine ESF (extracellular-matrix stabilizing factor). *FEBS Lett.* **318**, 292-296
11. Yoshida, E., Sumi, H., and Mihara, H. (1990) Different molecular forms of acid-stable trypsin inhibitor in normal and uremic plasma. *Clin. Chim. Acta* **190**, 277-280

12. Choi-Miura, N.-H., Sano, Y., Oda, E., Nakano, Y., Tobe, T., Yanagishita, T., Taniyama, M., Katagiri, T., and Tomita, M. (1995) Purification and characterization of a novel glycoprotein which has significant homology to heavy chains of inter-α-trypsin inhibitor family from human plasma. *J. Biochem.* **117**, 400-407
13. Saguchi, K., Tobe, T., Hashimoto, K., Sano, Y., Nakano, Y., Miura, N.-H., and Tomita, M. (1995) Cloning and characterization of cDNA for inter-α-trypsin inhibitor family heavy chain-related protein (IHRP), a novel human plasma glycoprotein. *J. Biochem.* **117**, 14-18
14. Nishimura, H., Kakizaki, I., Muta, T., Sasaki, N., Pu, P.X., Yamashita, T., and Nagasawa, S. (1995) cDNA and deduced amino acid sequence of human PK-120, a plasma kallikrein-sensitive glycoprotein. *FEBS Lett.* **357**, 207-211
15. Tobe, T., Saguchi, K., Hashimoto, K., Miura, N.-H., Tomita, M., Li, F., Wang, Y., Minoshima, S., and Shimizu, N. (1995) Mapping of the newly identified human inter-α-trypsin inhibitor family heavy chain-related protein (IHRP) gene (ITIHL1) to human chromosome 3p21-p14. *Cytogenet. Cell Genet.* **71**, 296-298
16. Salier, J.P., Diarra-Mehrpour, M., Sesboue, R., Bourguignon, J., Benarous, R., Ohkubo, I., Kurachi, S., and Martin, J.P. (1987) Isolation and characterization of cDNA encoding the heavy chain of human inter-α-trypsin inhibitor (IαTI): Unambiguous evidence for multipolypeptide chain structure of IαTI. *Proc. Natl. Acad. Sci. USA* **84**, 8272-8276
17. Bost, F., Bourguignon, J., Martin, J.P., Sesboue, R., Thileberville, L., and Diarra-Mehrpour, M. (1993) Isolation and characterization of the human inter-α-trypsin inhibitor heavy-chain H1 gene. *Eur. J. Biochem.* **218**, 283-291
18. Feinberg, A.P. and Vogelstein, B. (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**, 6-13
19. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
20. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467
21. Mount, S.M. (1982) A catalogue of splice junction sequence. *Nucleic Acids Res.* **10**, 459-472
22. Breathnach, R. and Chambon, P. (1981) Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.* **50**, 349-383
23. Bucher, P. (1990) Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.* **212**, 563-578
24. Dynan, W.S. (1986) Promoters for housekeeping gene. *Trends Genet.* **6**, 196-197
25. Valerio, D., Duyvesteyn, M.G.C., Dekker, B.T.M., van der Voorn, L., van Ormondt, H., and van der Eb, A.J. (1985) Adenosine deaminase: Characterization and expression of a gene with a remarkable promoter. *EMBO J.* **4**, 437-443
26. Reynold, G.A., Basu, S.K., Osborne, T.S., Chim, D.J., Gil, G., Brown, M.S., Goldstein, J.L., and Luskey, K.I. (1984) HMG CoA reductase: A negatively regulated gene with unusual promoter and 5′ untranslated region. *Cell* **38**, 275-285
27. Patel, I., Framson, P.E., Caskey, C.T., and Chinault, A.C. (1986) Fine structure of the human hypoxanthine phosphoribosyltransferase gene. *Mol. Cell. Biol.* **6**, 393-403
28. Hawkins, J.D. (1988) A survey on intron and exon lengths. *Nucleic Acids Res.* **16**, 9893-9908
29. Smale, S.T. and Baltimore, D. (1989) The "initiator" as a transcription control element. *Cell* **57**, 103-113
30. Sehgal, A., Patil, N., and Chao, M. (1988) A constitutive promoter directs expression of the nerve growth factor receptor gene. *Mol. Cell. Biol.* **8**, 3160-3167
31. Means, A.L. and Farnham, P.J. (1990) Transcription initiation from the dihydrofolate reductase promoter is positioned by HIP1 binding at the initiation site. *Mol. Cell. Biol.* **10**, 653-661
32. Buchman, T.G., Cabin, D.E., Vickers, S., Deutschman, C.S., Delgado, E., Sussman, M.M., and Bulkley, G.B. (1990) Molecular biology of circulatory shock. Part II. Expression of four groups of hepatic genes is enhanced after resuscitation from cardiogenic shock. *Surgery* **108**, 560-566
33. Gonzalez-Ramon, N., Alava, M.A., Sarsa, J.A., Pineiro, M., Escartin, A., Garcia-Gil, A., Lampreave, F., and Pineiro, A. (1995) The major acute phase serum protein in pigs is homologous to human plasma kallikrein sensitive PK-120. *FEBS Lett.* **371**, 227-230
34. Roy, A.L., Meisterernst, M., Pognonec, P., and Roeder, R.G. (1991) Cooperative interaction of an initiator-binding transcription initiation factor and the helix-loop-helix activator USF. *Nature* **354**, 245-248
35. O'Shea-Greenfield, A. and Smale, S.T. (1992) Roles of TATA and initiator elements in determining the start site location and direction of RNA polymerase II transcription. *J. Biol. Chem.* **267**, 1391-1401
36. Hariharan, N. and Perry, R.P. (1990) Functional dissection of a mouse ribosomal protein promoter: Significance of the polypyrimidine initiator and an element in the TATA-box region. *Proc. Natl. Acad. Sci. USA* **87**, 1526-1530
37. Seto, E., Shi, Y., and Shenk, T. (1991) YY1 is an initiator sequence-binding protein that directs and activates transcription *in vitro*. *Nature* **354**, 241-245
38. Weis, L. and Reinberg, D. (1992) Transcription by RNA polymerase II: Initiator-directed formation of transcription-competent complexes. *FASEB J.* **6**, 3300-3309
39. Parslow, T.G., Blair, D.L., Murphy, W.J., and Granner, D.K. (1984) Structure of the 5′ ends of immunoglobulin genes: A novel conserved sequence. *Proc. Natl. Acad. Sci. USA* **81**, 2650-2654
40. Falker, F.G. and Zachan, H.G. (1984) Correct transcription of an immunoglobulin k gene requires an upstream fragment containing conserved sequence elements. *Nature* **310**, 71-74
41. Scheidereit, C., Cromlish, J.A., Gerster, T., Kawakami, K., Balmaceda, C.G., Alexander Currie, R., and Roeder, R.G. (1988) A human lymphoid-specific transcription factor that activates immunoglobulin genes is a homeobox protein. *Nature* **336**, 551-557