Gene, 123 (1993) 277–281

GENE 06868

Organization and sequence of the gene encoding the human acrosin-trypsin inhibitor (HUSI-II)*

(Kazal-type inhibitor; acrosin inhibitor; exon-intron mapping; nucleotide sequence analysis; transcription start point; chromosomal localization; glycoprotein hormones; chorionic gonadotropin)

Annemarie Möritz[a]**, Karl-Heinz Grzeschik[b], Edgar Wingender[c] and Edwin Fink[a]

[a] Department of Clinical Chemistry and Clinical Biochemistry, University of Munich, Munich, Germany; [b] Department of Human Genetics, University of Marburg, Marburg, Germany. Tel. (49-6421)284-080; and [c] Department of Genetics, Gesellschaft für Biotechnologische Forschung mbH, Braunschweig, Germany. Tel. (49-531)6181-203

Received by H.M. Krisch: 24 February 1992; Revised/Accepted: 21 August/10 September 1992; Received at publishers: 29 September 1992

SUMMARY

A complete cDNA encoding the acrosin-trypsin inhibitor, HUSI-II, was used as a probe to isolate genomic clones from a human placenta library. Three clones which cover the entire *HUSI-II* gene were isolated and characterized. The exon-intron organization of the gene was determined and found to be identical to other known Kazal-type inhibitor-encoding genes. The striking similarity in the amino acid sequences which was found previously in HUSI-II and glycoprotein hormone β-subunits, is neither reflected in codon usage nor in the exon-intron arrangement of the genes. A 1.8-kb segment 5′ of the gene was sequenced. The analysis of this sequence showed that *HUSI-II* contains a G+C-rich region upstream from the transcription start point (*tsp*) which fulfills the criteria for a CpG island. Furthermore, in the first intron, a potential glucocorticoid-responsive element was found as a half-palindrome flanked by two CACCC elements. Determination of the *tsp* by S1 mapping revealed that *HUSI-II* has multiple *tsp*. Genomic Southern hybridization was used to show that *HUSI-II* is a single-copy gene. The localization of the gene to chromosome 4 was determined by hybridization of a 5′ genomic fragment to the DNA of a panel of somatic hybrids between human and rodent cells.

INTRODUCTION

The acrosin-trypsin inhibitor HUSI-II isolated from human seminal plasma (Fink et al., 1971), seems to be synthesized exclusively in the male genital tract (Hehlein-Fink, 1990). Its primary structure (Fink et al., 1990) identified it as a member of the Kazal-type inhibitor family (Laskowski and Kato, 1980).

Kazal-type inhibitors, serine proteinase inhibitors, contain one or more Kazal-type domains, which are protein domains with primary structures similar to that of the bovine pancreatic secretory trypsin inhibitor (Kazal et al., 1948; Greene et al., 1969). Multiheaded Kazal-type inhibitors contain up to seven domains, each having a distinct aa sequence and different inhibitory properties (Laskowski and Kato, 1980).

Correspondence to: Dr. E. Fink, Department of Clinical Chemistry and Clinical Biochemistry, University of Munich, Nussbaumstrasse 20, D-8000 Munich 2, Germany. Tel. (49-89)5160-2601; Fax (49-89)5160-4735.

*A compilation of regulatory and protein-interacting elements and a programme for localizing these regions is available on request from Dr. E. Wingender, Gesellschaft für Biotechnologische Forschung, Mascheroder Weg 1, D-3300 Braunschweig, Germany. Tel. (49-531) 6181-203; Fax (49-531) 6181-202; e-mail:EWI@V.GBF-Braunschweig.DBP.DE.

**Present address: Pharmacia Biosystems GmbH, Munzingerstraße 9, D-7800Freiburg/Br., Germany. Tel. (49-761)4903-294.

Abbreviations: aa, amino acid(s); bp, base pair(s); cpm, counts/min; *GRE*, glucocorticoid-responsive element; hCGβ, human chorionic gonadotropin, β-subunit; HUSI-II, human seminal plasma inhibitor; *HUSI-II*, gene encoding HUSI-II; kb, kilobase(s) or 1000 bp; nt, nucleotide(s); oligo, oligodeoxyribonucleotide(s); PSTI, pancreatic secretory trypsin inhibitor; *tsp*, transcription start point(s).

To date, two human Kazal-type inhibitors have been isolated and characterized by aa sequencing. They are the PSTI (Greene et al., 1976), and the HUSI-II (Fink et al., 1991). An oligo deduced from the aa sequence of HUSI-II was employed to isolate and analyse cDNA-clones from testis and seminal vesicle libraries (Möritz et al., 1990). One of these cDNA clones was used to clone the *HUSI-II* gene in order to investigate the gene structure and regulation of expression of the inhibitor.

EXPERIMENTAL AND DISCUSSION

(a) Evidence that *HUSI-II* is a single-copy gene

Genomic placenta DNA was digested with several restriction enzymes and a genomic Southern blot of the DNA was hybridized using a *HUSI-II* cDNA fragment (Möritz et al., 1991) as a probe (Fig. 1). The cDNA fragment hybridized at high stringency to a maximum of three different restriction fragments of human placenta DNA. Such a low number of hybridizing fragments is generally taken as a strong indication that only a single copy of the gene occurs in the genome (Baumann et al., 1988; Stühmer et al., 1989).

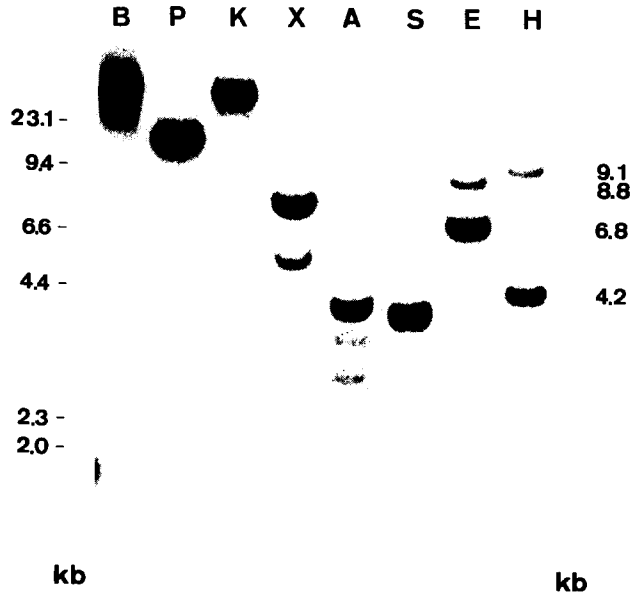

Fig. 1. Genomic Southern blot analysis of the *HUSI-II* gene. Genomic placenta DNA (10 µg) was digested overnight at 37°C with: A, *Acc*I; B, *Bam*HI; E, *Eco*RI; H, *Hind*III; K, *Kpn*I; P, *Pst*I; S, *Sac*I; X, *Xba*I. The digested DNA was resolved on a 0.8% agarose gel together with a DNA marker (kb on the left). Transfer onto a BiodyneA membrane (Pall, Dreieich, Germany), probing with a full-length *HUSI-II* cDNA fragment which was labeled by random priming using [^{32}P]dCTP, and washing of the membrane were performed as described by Sambrook et al. (1989). The sizes of the genomic *Eco*RI- and *Hind*III-fragments are as indicated on the right. They correspond to the genomic fragments isolated from digested genomic clones (Fig. 2).

(b) Structure of the *HUSI-II* gene

The gene structures of three Kazal-type inhibitors have been published: the two multiheaded inhibitors of chicken, ovomucoid with three domains (Stein et al., 1980) and ovoinhibitor with seven domains (Scott et al., 1987), and the human PSTI with one domain (Horii et al., 1987; Tan et al., 1988). The exon-intron arrangement is identical in all these inhibitor genes (Tan et al., 1988): one intron is localized in the sequence encoding the leader peptide, 14 nt upstream from the N-terminus of the mature protein, and two introns are present within each of the eleven inhibitor domains: the first intron located 6 nt upstream from the first Cys codon, the second intron 11 nt downstream from the 5th Cys codon.

Presuming an identical organisation of the *HUSI-II* gene, the exon-intron structure was determined by using exonic oligos as primers for sequencing the exon-intron boundaries. The complete sequence of exons was confirmed by sequencing the complementary strand with intron primers obtained from the new sequence data. Compared to the cDNA sequence (Möritz et al., 1991) one difference was detected: a C instead of the T^{1169} (Fig. 3) of the cDNA. This nt exchange represents a silent mutation of the codon of Val^{9} (Fig. 3) and probably reflects the fact that the genomic and cDNA libraries were obtained from specimens of different individuals. The

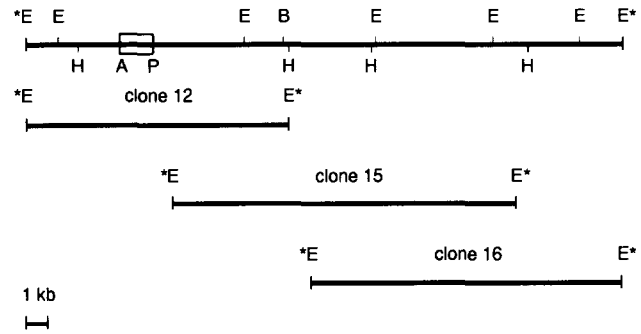

Fig. 2. Restriction map of the genomic region containing the *HUSI-II* gene. 2×10^6 recombinant λDash phages (Stratagene, La Jolla, CA) from a genomic placental library were screened with full-length *HUSI-II* cDNA (Möritz et al., 1991) as a probe according to Sambrook et al. (1989). Three hybridizing clones numbered 12, 15 and 16 were isolated. Full and partial digests of phage DNA of the three clones were prepared (Sambrook et al., 1989) to construct a restriction map using the enzymes: *Eco*RI (E), *Bam*HI (B), *Hind*III (H). Clone 12 was also digested with *Acc*I (A), *Xba*I and *Pst*I (P) to obtain shorter fragments for the 5′ region sequence analysis; the resulting 1.8-kb *Acc*I-*Pst*I restriction fragment used for sequencing (cf. Fig. 3) is boxed. The characterized region of about 30 kb is flanked by artificial *Eco*RI cloning sites, E*, which originate from the λDash cloning vector. The three clones overlap as indicated below the map. Exon containing fragments were identified in hybridization experiments using oligos deduced from the cDNA sequence. No hybridization was observed for the 5′ end *Eco*RI fragment (1.4 kb) and for the three *Eco*RI fragments (12.4 kb) representing the 3′ part of the covered genomic region. Suitable fragments were subcloned into the vector pTZ19R (Mead et al., 1986) for further analysis.

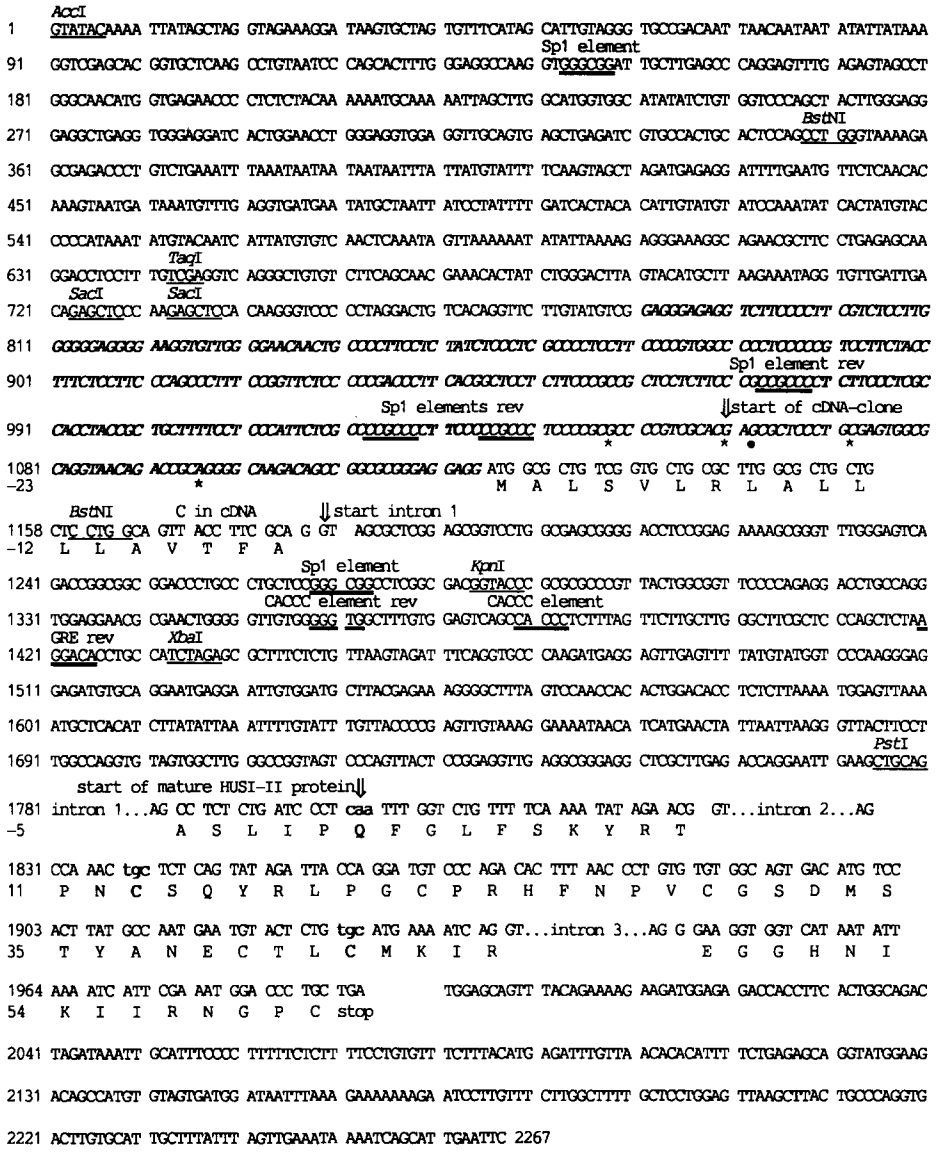

Fig. 3. Organization and partial sequence of the *HUSI-II* gene. Insert DNA of subclones was sequenced using the T7 Polymerase kit (Pharmacia Biosystems, Freiburg, Germany) with suitable primers (see section **b**). The 1.8-kb *Acc*I-*Pst*I fragment (containing 1.1-kb 5′ region, exon 1 and 0.6 kb of intron 1), exons 2, 3 and 4 including the exon-intron boundaries were sequenced. Genomic clones 12, 15 and 16 (Fig. 2) overlap in intron 2 as was deduced by sequence and mapping data. The start codon of the mature HUSI-II protein and the Cys codons, to which the intron positions are referred in section **b**, are marked by boldface, lowercase letters. Numbering of the nt sequence is arbitrary and corresponds only to the nt sequenced. The putative CpG island is italicized and marked by bold type, promoter elements are doubly underlined. rev, element located in the reverse orientation on the lower strand. The major *tsp* of the gene are marked by asterisks, the minor *tsp* is marked by a heavy dot. Only those restriction sites of *Bst*NI and *Taq*I are indicated which are important for S1 mapping (see section **e**). The sequence data were analysed using the program Microgenie, version 5 (Beckman Instruments). The sequence data have been deposited with the GenBank Data Library under accession Nos. M84967 and M91438.

exon-intron arrangement of the *HUSI-II* gene is identical to that of the other published Kazal-type inhibitor genes and the same holds true for the intron phases: intron 1, phase 1; intron 2, phase 0; intron 3, phase 2. Furthermore, all introns obey the 5′GT...AG3′-rule (Abelson, 1979).

(c) Comparison to the gene of human chorionic gonadotropin

The aa sequence of HUSI-II shows a striking similarity to the β-subunits of the glycoprotein hormones (Fink et al., 1990) which is most striking between aa 17 and 28 (aa numbering of Fig. 3): 10 of the 12 aa are identical to a sequence of hCGβ. In β-chains of human luteinizing hormone, follicle-stimulating hormone and thyroid-stimulating hormone the corresponding homology is only 9, 6 and 5 aa, respectively. In Kazal-type proteinase inhibitors this segment includes the region of primary contact between the inhibitor and the proteinase and also the reactive inhibitor site (Arg^{23}-His^{24} in Fig. 3; discussed in detail by Laskowski, 1986).

At the DNA level the sequence similarity is less pronounced. In *HUSI-II* and *hCGβ* (Policastro et al., 1983)

only 63% identity is observed for the corresponding nt sequences and only for 2 aa are the identical codons used. In addition, no similarity exists in the exon-intron organizations. This low similarity at the genomic level in contrast to the protein level suggests that after evolving from a common ancestor, the aa sequences have been conserved in spite of the divergence because this segment is functionally important. It is tempting to speculate that, like in Kazal-type inhibitors (cf, above), the respective sequence segment in glycoprotein hormones will also be involved in complex formation, for example with either the hormone α-chain or the receptor.

(d) Analysis of the 5′-flanking region of the *HUSI-II* gene

The search for regulatory elements in the 5′-flanking sequence (Fig. 3) showed that no TATA- or CCAAT-boxes are located upstream from the putative start codon of the HUSI-II protein. The 5′ regions of TATA-less genes are often G+C rich and have been shown to contain the so-called 'CpG islands' (Bird, 1986) which are thought to be involved in gene regulation (Weih et al., 1991). Indeed, a sequence segment fulfilling the criteria of a CpG island (Bird, 1986) is found between nt 781 and 1134 (Fig. 3): The G+C content of this region is 69% and the amount of the dinucleotides GpC and CpG is nearly identical (30:28).

CpG island genes often contain Sp1-motifs 'GGGCGG' which are active in either direction (La-Thangue and Rigby, 1988). In the *HUSI-II* gene, five Sp1 boxes are found at nt 143, 973, 1023, 1035 and 1268; three of them are located within the CpG island.

A half palindrome of the glucocorticoid-responsive element 'GGTACAN$_3$TGTTCT', to which androgen, progesterone and mineralocorticoid receptors can bind (Beato, 1989), is located at nt 1420 in the first intron of the *HUSI-II* gene. It is flanked by two CACCC elements at nt 1358 and 1379 which act synergistically with glucocorticoid regulatory elements (Strähle et al., 1988). Half palindromic sequences of glucocorticoid elements can be functional (Tora et al., 1988) and can also be located in the first intron of genes (Strähle et al., 1988).

Taken together, the analysis of the region upstream from the translation start point suggests a constitutive regulation of the *HUSI-II* gene since it contains a CpG island with Sp1-elements and no TATA- or CCAAT-boxes (Dynan, 1986). In contrast, however, the analysis of 0.6 kb of intron 1 is consistent with the *HUSI-II* gene being hormone-regulated as was indicated by previous experiments at the protein level (Haendle et al., 1965; Hehlein-Fink, 1990).

(e) Transcription start points (*tsp*) of the *HUSI-II* gene

If a TATA-box is absent or deleted experimentally, multiple *tsp* are often found (Reynolds et al., 1984). S1 analysis of the TATA-less *HUSI-II* gene confirms this finding: one minor and four major *tsp* at nt 1048, 1060, 1062, 1072 and 1096 (nt numbering of Fig. 3) were detected (Fig. 4). Furthermore, the *tsp* at nt 1060 corresponds to the start of the full-length cDNA clone used for isolating the *HUSI-II* gene (Fig. 3).

(f) Chromosomal assignment of the *HUSI-II* gene

A 551-bp *Sac*I-*Kpn*I-fragment (nt 738–1288, Fig. 3), which contains no repetitive sequences and, therefore, hy-

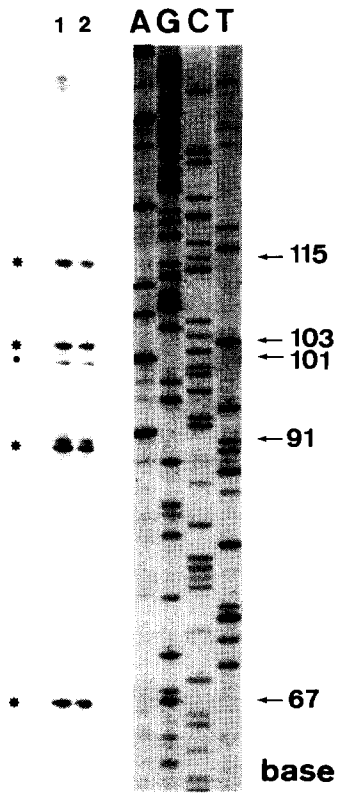

Fig. 4. S1 map of the *HUSI-II* gene. The 1.8-kb *Acc*I-*Pst*I-fragment containing the 5′ region of the *HUSI-II* gene (Fig. 3) was digested with *Bst*NI. The resulting 812-bp *Bst*NI fragment (nt 350–1161, Fig. 3) was dephosphorylated (Sambrook et al., 1989) and further digested with *Taq*I yielding a 517-bp *Taq*I-*Bst*NI fragment (nt 644–1161, Fig. 3) which is dephosphorylated only at the *Bst*NI site in exon 1 (nt 1461, Fig. 3). It was separated from the 299-bp *Bst*NI-*Taq*I fragment (nt 350–643, Fig. 3) and labeled with [γ-^{32}P]ATP (5000 Ci/mmol, Amersham, Braunschweig; Sambrook et al., 1989). For hybridization, 50 000 cpm of the fragment nt 644–1161 (Fig. 3) and 20 μg total testis RNA (lane **1**) or 2 μg poly(A)$^+$ testis RNA (lane **2**) were coprecipitated with ethanol. The quality of RNAs was confirmed by Northern analysis prior to S1 mapping. DNA and RNA were resuspended in 20 μl hybridization buffer (40 mM PIPES-KOH pH 6.4/400 mM NaCl/1 mM EDTA/80% formamide). After 30 min at 85°C, the samples were incubated at 52°C for 8–12 h submersed in a water bath. S1 digestion of single-stranded nucleic acids was performed (Sambrook et al., 1989) and, after phenol extraction, the protected fragments were separated on a 6% sequencing gel and visualized by autoradiography. A sequencing ladder was included to enable accurate determination of the product sizes (lanes A, G, C and T; these are overexposed because the sequencing reaction was not optimal). The four major *tsp* are marked by asterisks, the minor *tsp* is marked by a heavy dot. The numbering on the right refers to the lengths of protected fragments found by this analysis.

bridizes only with the 9.1-kb genomic *Eco*RI-fragment of the *HUSI-II* gene, was used to probe a Southern blot of *Eco*RI-restricted genomic DNA from a panel of human-rodent hybrid cell lines (Luerssen et al., 1990). The fragment segregated with chromosome 4. In one clone, which cytogenetically shows no intact chromosome 4, it segregated with a marker known to be localized close to the tip of the short arm of chromosome 4, but not with a marker previously assigned to the long arm. This might indicate that the *HUSI-II* gene is located on 4p.

ACKNOWLEDGMENTS

We thank Dr. H. Gram, Sandoz Pharma Ltd., Basel, Switzerland for providing the placenta library and Dr. F. Tan, University of Illinois, Chicago, IL for providing genomic placenta DNA. This work was supported by the Sonderforschungsbereich 207 of the University of Munich and grant Gr 373/16-2.

REFERENCES

Abelson, J.: RNA processing and the intervening sequence problem. Annu. Rev. Biochem. 48 (1979) 1035–1069.

Baumann, A., Grupe, A., Ackermann, A. and Pongs, O.: Structure of the voltage-dependent potassium channel is highly conserved from *Drosophila* to vertebrate central nervous system. EMBO J. 7 (1988) 2457–2453.

Beato, M.: Gene regulation by steroid hormones. Cell 56 (1989) 335–344.

Bird, A.P.: CpG rich islands and the function of DNA methylation. Nature 321 (1986) 209–213.

Dynan, W.S.: Promoters for housekeeping genes. Trends Genet. 2 (1986) 196–197.

Fink, E., Jaumann, E., Fritz, H., Ingrisch, H. and Werle, E.: Protease-Inhibitoren im menschlichen Spermaplasma: Isolierung durch Affinitätschromatographie und Hemmverhalten. Hoppe-Seyler's Z. Physiol. Chem. 352 (1971) 1591–1594.

Fink, E., Hehlein-Fink, C. and Eulitz, M.: Amino acid sequence elucidation of human acrosin-trypsin inhibitor (HUSI-II) reveals that Kazal-type proteinase inhibitors are structurally related to beta-subunits of glycoprotein hormones. FEBS Lett. 270 (1990) 222–224.

Greene, L.J. and Bartelt, D.C.: The structure of the bovine pancreatic secretory trypsin inhibitor – Kazal's inhibitor. II. The order of the tryptic peptides. J. Biol. Chem. 244 (1969) 2646–2657.

Greene, L.J., Pubols, M.H. and Bartelt, D.C.: Human pancreatic secretory trypsin inhibitor. Methods Enzymol. 45 (1976) 813–825.

Haendle, H., Fritz, H., Trautschold, I. and Werle, E.: Über einen hormonabhängigen Inhibitor für proteolytische Enzyme in männlichen accessorischen Geschlechtsdrüsen und im Sperma. Hoppe Seyler's Z. Physiol. Chem. 343 (1965) 185–188.

Hehlein-Fink, C.: Aufklärung der Primärstruktur des humanen Akrosininhibitors HUSI-II und Untersuchungen zu seiner physiologischen Funktion. Thesis, Technical University Munich, 1990.

Horii, A., Kobayashi, T., Tomita, N., Yamamoto, T., Fukushige, S., Murotsu, T., Ogawa, M., Mori, T. and Matsubara, K.: Primary structure of the human pancreatic secretory trypsin inhibitor (PSTI) gene. Biochem. Biophys. Res. Commun. 149 (1987) 635–641.

Kazal, L.A., Spicer, D.S. and Brahinski, R.A.: Isolation of a crystalline trypsin inhibitor-anticoagulant protein from pancreas. J. Am. Chem. Soc. 70 (1948) 3034–3040.

Laskowski Jr., M.: Protein inhibitors of serine proteinases – mechanism and classification. In: M. Friedman (Ed.), Nutritional and Toxicological Significance of Enzyme Inhibitors in Foods. Plenum, New York, 1986, pp. 1–17.

Laskowski Jr., M. and Kato, I.: Protein inhibitors of proteinases. Annu. Rev. Biochem. 49 (1980) 593–626.

LaThangue, N.B. and Rigby, P.W.J.: Trans-acting protein factors and the regulation of eukaryotic transcription. In: Hames, B.D. and Glover, D.M. (Eds.), Transcription and Splicing. IRL Press, Oxford, 1988, pp. 1–42.

Luerssen, H., Mattei, M.-G., Schröter, M., Grzeschik, K.-H., Adham, J.M. and Engel, W.: Nucleotide sequence of the gene for human transition protein 1 and its chromosomal localization on chromosome 2. Genomics 8 (1990) 324–330.

Mead, D.A., Szczesna-Skorupa, E. and Kemper, B.: Single-stranded DNA 'blue' T7 promoter plasmids: a versatile tandem promoter system for cloning and protein engineering. Prot. Eng. 1 (1986) 67–74.

Möritz, A., Lilja, H. and Fink, E.: Molecular cloning and sequence analysis of the cDNA encoding the human acrosin-trypsin inhibitor (HUSI-II). FEBS Lett. 278 (1991) 127–130.

Policastro, P., Ovitt, C.E., Hoshina, M., Fukuoka, H., Boothby, M.R. and Boime, I.: The β-subunit of human chorionic gonadotropin is encoded by multiple genes. J. Biol. Chem. 258 (1983) 11492–11499.

Reynolds, G.A., Basu, S.K., Osborne, T.F., Chin, D.J., Gil, G., Brown, M.S., Goldstein, J.L. and Luskey, K.L.: HMG CoA reductase: a negatively regulated gene with unusual promoter and 5′ untranslated region. Cell 38 (1984) 275–285.

Sambrook, J., Fritsch, E.F. and Maniatis, T.: Molecular Cloning. A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

Scott, M.J., Huckaby, C.S., Kato, I., Kohr, W.J., Laskowski Jr., M., Tsai, M.-J. and O'Malley, B.W.: Ovoinhibitor introns specify functional domains as in the related and linked ovomucoid gene. J. Biol. Chem. 262 (1987) 5899–5907.

Stein, J.P., Catterall, J.F., Kristo, P., Means, A.R. and O'Malley, B.W.: Ovomucoid intervening sequences specify functional domains and generate protein polymorphism. Cell 21 (1980) 681–687.

Strähle, U., Schmid, W. and Schütz, G.: Synergistic action of the glucocorticoid receptor with transcription factors. EMBO J. 7 (1988) 3389–3395.

Stühmer, W., Ruppersberg, J.P., Schröter, K.H., Sakmann, B., Stocker, M., Giese, K.P., Perschke, A. and Pongs, O.: Molecular basis of functional diversity of voltage-gated potassium channels in mammalian brain. EMBO J. 8 (1989) 3235–3244.

Tan, F., Gippner, C. and Fink, E.: On the genomic organization of the human pancreatic secretory trypsin inhibitor gene. Biol. Chem. Hoppe Seyler's 369, Suppl. 5 (1988) 51–54.

Tora, L., Gaub, M.P., Mader, S., Dierich, A., Bellard, M. and Chambon, P.: Cell-specific activity of a GGTCA half-palindromic oestrogen-responsive element in the chicken ovalbumin gene promoter. EMBO J. 7 (1988) 3771–3778.

Weih, F., Nitsch, D., Reik, A., Schütz, G. and Becker, P.B.: Analysis of CpG methylation and genomic footprinting at the tyrosine aminotransferase gene: DNA methylation alone is not sufficient to prevent protein binding in vivo. EMBO J. 10 (1991) 2559–2567.