

Structure of the human κ -casein gene

Anders Edlund, Thore Johansson, Brith Leidvik, Lennart Hansson *

Symbicom AB, S-90124, Sweden. Present address: Molecular Biology, Preclinical Research and Development, Astra Hässle AB, Tvistevägen 48, S-90736 Umeå, Sweden

Received 30 January 1996; revised 10 April 1996; accepted 11 June 1996

Abstract

The human κ -casein-encoding gene, *Kca*, was cloned and sequenced. The structural gene consists of five exons ranging from 33 to 496 nucleotides (nt) separated by introns ranging from 1146 to 2942 nt, and extends over 8821 nt. All intron/exon splice junctions conform to the GT/AG rule. The gene organization is similar to that of the bovine gene. The 5'-flanking region contains an A+T-rich sequence; TTTAATT, close to where the TATA motif is found in most other genes, a CAAT box, and an AP-1 consensus sequence. In addition, one *Alu* repetitive element was found in the second intron.

Keywords: Milk protein gene; Nucleotide sequence; Gene organization; Regulatory sequences; *Alu* sequence

1. Introduction

The caseins are the predominant proteins in the milk of most mammalian species. The stage and tissue specific expression pattern of the casein genes, e.g. during lactation in mammary glands, is regulated by various hormones (Devinoy et al., 1988; Jahn et al., 1989; Groenen and Van der Poel, 1994). Human milk casein is composed of two casein species, β - and κ -casein, of which κ -casein represents less than 10% of the casein fraction. Human κ -casein with an apparent molecular weight of 35 kDa, is heavily glycosylated. The carbohydrates constitutes about 50–60% of its molecular mass (Kunz and Lönnerdal, 1989).

Sequences of κ -casein, *Kca*, cDNAs have been reported for rat (Nakashima et al., 1984), cow (Stewart et al., 1984), guinea-pig (Hall, 1990), pig (Levine et al., 1992), human (Bergström et al., 1992), goat (Coll et al., 1993) and rabbit (Bösze et al., 1993). In addition, the bovine *Kca* gene (Alexander et al., 1988), and 5'-flanking regions from the goat (Coll et al., 1995) and sheep (Spira et al.: Genbank No. L31372) are described. The accumulation of knowledge about *Kca* gene organization and

sequence information from various species, intraspecies variants (Schild et al., 1994) and expression studies of recombinant *Kca* genes in transgenic models (Persuy et al., 1995), are all prerequisites for the dissection of the complex regulation of *Kca* gene activity. In this study, the human *Kca* gene is analyzed and described.

2. Experimental and discussion

2.1. Postulated structure of the human *kca* gene

By comparison of the human *Kca* cDNA sequence (Bergström et al., 1992) to the bovine *Kca* cDNA sequence (Stewart et al., 1984) and to the genomic sequence and organization of the bovine *Kca* gene (Alexander et al., 1988), putative exon/intron boundaries were postulated for the human *Kca* gene. In order to map the positions of the introns, a set of oligonucleotides to be used as PCR primers, complementary to sequences of the different exons (Table 1), were designed. The results obtained in these PCR experiments, using human DNA as template, indicated that the size of intron-1, -3, and -4 is about 2.1 kb, 1.1 kb, and 1.8 kb, respectively. The size of intron 2 could not be determined in these experiments.

* Corresponding author. Tel. +46 90 175320; Fax +46 90 175399.

Abbreviations: bp, base pair(s); cDNA, DNA complementary to RNA; kb, kilobase(s) or 1000 bp; *Kca*, gene encoding κ -casein; kDa, kilodalton; nt, nucleotide(s); PCR, polymerase chain reaction; *tsp*, transcriptional start site; UTR, untranslated region.

Table 1
PCR primers used to map the introns

Primer	Location	Sequence	Direction	Nucleotides		
SYM 2707	Exon 1	GAGAGAAGATCTGACTGGCACGAGGAAAGG	5'-3'	8	-	36
SYM 2921	Exon 2	AGGCAGGGTTAATGCCAGGG	3'-5'	92	-	73
SYM 3273	Exon 2	CAATAATGAAGAGTTTTCTTC	5'-3'	40	-	60
SYM 2922	Exon 3	TGCTGGTTGTTTCTGGTTT	3'-5'	121	-	113
SYM 3255	Exon 3	CTGTGGAGGTTCAAAACC	5'-3'	100	-	117
SYM 2430	Exon 4	ACATAATACATTGGGACATA	3'-5'	196	-	177
SYM 2767	Exon 4	CACCAAGGAAATATCAAAG	5'-3'	597	-	615
SYM 2272	Exon 5	AAGGAGAGTGTGAAGTAGTA	3'-5'	665	-	646

2.2. Isolation of the human *Kca* gene

The human *Kca* gene was cloned from two unrelated human genomic libraries, one constructed from human placenta DNA and the second from human female leukocyte DNA. Both libraries were cloned into λ EMBL-3 (Clontech, Palo Alto, CA, USA). From the human placenta library two genomic clones containing transcribed sequences of the human *Kca* gene were isolated using the previously cloned *Kca* cDNA as a probe (Bergström et al., 1992). The two plaques were subjected to several re-screening experiments and recombinant phage DNA was purified (Sambrook et al., 1989), digested with *SalI*, and fragments of about 18 kb and 17 kb were isolated by agarose electrophoresis and cloned into *SalI* linearized pUC19 plasmid, resulting in plasmids pS457 and pS459. A third clone was isolated using the 2.1-kb intron-1 fragment obtained from the PCR experiment described above as a probe. As a result, DNA fragment of about 15 kb was isolated and cloned as above and the resulting plasmid was designated pS767.

The genomic leukocyte DNA library was screened as described and four positive plaques were isolated. One plaque was obtained with the 2.1-kb intron-1 probe and three plaques were obtained with the full length cDNA probe. Inserts derived from the isolated plaques were cloned into pUC19 plasmid as above and designated pS768, pS809, pS810, and pS811. The DNA fragments were about 12 kb, 14 kb, 17 kb and 17 kb, respectively.

2.3. Characterization and sequence analysis of *Kca* genomic fragments

Restriction mapping showed that the isolated fragments overlapped and that they covered about 35 kb of human genomic DNA (Fig. 1). Two plasmids, pS457 and pS459, showed a partly different restriction pattern suggesting that there were some rearrangements during the cloning of these two fragments.

In order to facilitate nt sequence analysis restriction fragments were derived from the plasmids and the inserts were subjected to nt sequence analysis. The complete

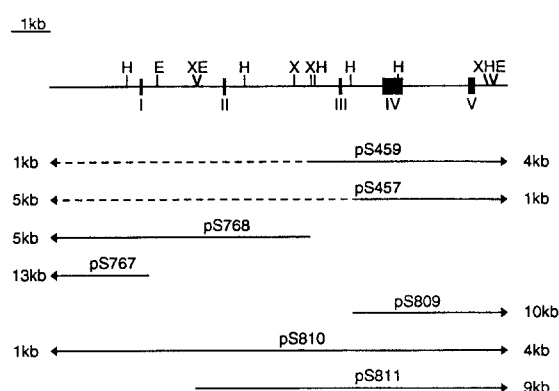


Fig. 1. Gene organization and restriction map of the human *Kca* gene. Exons are shown as solid boxes. Seven overlapping *Kca* genomic clones, pS459, pS457, pS768, pS767, pS809, pS810, and pS811 are shown in the 5'-to-3' orientation. Numbers given beside the arrows indicate the length of additional DNA in the genomic clones. The dotted lines show the rearranged regions observed in two clones. Restriction enzymes used are indicated as follows: E, *EcoRI*; H, *HindIII*; X, *XbaI*.

sequence of all subclones was determined using T7 sequencing kit (Pharmacia, Uppsala, Sweden) on double-stranded templates. The analyzed human *Kca* genomic nt sequence consists of 9844 bp, and contains the complete exon and intron structure of the gene (Fig. 2). The human *Kca* gene consist of 5 exons separated by 4 introns and the transcribed region extends over 8821 nt. The translational start is localized in exon-2, and the translational stop in exon-4.

For unknown reasons experiments performed in order to define the *tsp* of the human *Kca* gene have not been successful. However, several observations indicate the position of the *tsp*. First, upon comparison with the bovine *Kca* gene (Alexander et al., 1988) extensive homology in this region is observed. Second, sequence information from a number of independent cDNA clones does not indicate the existence of a longer 5'-UTR. Third, the size of *Kca* mRNA is in agreement with the proposed *tsp*. Fourth, a TATA like A + T-rich sequence is found at positions -32 to -26.

The exons are relatively small, size range is between 33 and 496 bp, and the introns range in size from 1146

Fig. 2. Nucleotide sequence of the human *Kca* gene. 9844 nt are shown including 361 bp of 5'-flanking region, 8821 bp exon and intron sequences, and 662 bp of 3'-flanking region. Exon sequences are shown in upper case letters, intron sequences are shown in lower case letters, and the amino acid sequence are listed below the translated exons using the one letter code under the second nt of each codon. For numbering of the *Kca* sequence the putative *tsp* was assigned nt 1. For amino acid numbering (in italics), the first Met in the signal peptide was assigned to position 1. The putative regulatory elements found in the 5'-flanking region are marked by boxes. The dotted line below the sequence indicates the location of the *Alu* element in intron 2. The sequence reported in this article has been assigned the GenBank accession No. U51899.

Table 2
Intron splice junctions of human *Kca* gene

Intron	Conserved nt ^a	Exon	Intron	Exon	Conserved nt ^a
1	7/9	AAAG	GTAATG...TAAATTTATCTTTAG	GTGC	11/15
2	7/9	TTTG	GTAAGT...TCTGGAACCTCCCCAG	GCTG	11/15
3	6/9	AGCA	GTAAGT...TTATTTTTTTTGCAG	TGCC	13/15
4	8/9	GCAG	GTAAAT...CATTATTTTTTATAG	GACT	13/15
Consensus ^b		MAG	GTRAGT...YYYYYYYYYYYNYAG	G	

^aFraction of nucleotides conserved relative to Mount consensus sequence.

^bM=A or C; R=A or G; Y=C or T; N=A or C or G or T.

to 2942 bp. The results obtained in the initial PCR experiments (see Section 2.1)) were confirmed by DNA hybridization experiments and sequence data. Moreover, at this stage the size of intron-2 was also confirmed by PCR experiments. All exon/intron boundaries conform to the Breatnach (GT/AG) rule (Breatnach et al., 1978) and all splice junctions conform well to the consensus sequence suggested by Mount (1982) (Table 2). No coding triplets are disrupted by splicing sites similar to what was observed for the human β -casein gene (Hansson et al., 1994).

A search of the *Kca* sequence for the presence of repeat elements, revealed one sequence in intron 2 which share homology with the *Alu* element.

2.4. Comparison of the human and bovine *Kca* genes

The bovine *Kca* gene organization (Alexander et al., 1988) was compared with that of the human *Kca* gene. The comparison showed that the exon/intron organization of these two genes was highly conserved (Fig. 3). Both genes have five exons and four introns, the translational start is localized in exon-2, and the translational stop in exon-4. At nt level, the highest degree of homology is observed for exon 2: 87%; for exon 1: 81%.

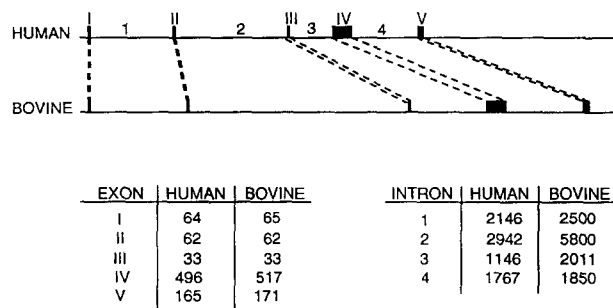


Fig. 3. Comparison of the human and bovine *Kca* genes. The *Kca* genes are aligned from the 5'-end of exon I; introns are numbered 1–4, and exons are numbered I–V on the human gene. Tables below the genes list the sizes (bp) of the introns and exons in the two genes.

2.5. Sequence analysis of putative regulatory elements

A computer-assisted sequence search (MacVector, International Biotechnologies Inc.) of the 361-bp 5'-flanking region of the human *Kca* gene identified some sequences similar to putative promoter and regulatory elements which have been characterized as important in the expression and regulation of eukaryotic genes. The A+T-rich sequence, TTTAATT, is found at positions –32 to –26, close to where the TATA motif is found in most other genes. No CCAAT sequence was found in the region approximately 80 bp upstream from the transcriptional start. However, a CCAAT element was found in the forward direction at –160. Another sequence which could be involved in regulation of expression of the gene is the sequence TGACTCA which begins at –53. This is a 7/7 match with the sequence which has been shown to be recognized by the transcription factor AP-1 (Lee et al., 1987). AP-1 recognition sequences are also present in the 5'-flanking region of K-casein gene from cow, goat and sheep.

References

- Alexander, L.J., Stewart, A.F., Mackinlay, A.G., Kapelinskaya, T.V., Tkach, T.M. and Gorodetsky, S.I. (1988) Isolation and characterization of the bovine κ -casein gene. *Eur. J. Biochem.* 178, 395–401.
- Bergström, S., Hansson, L., Hernell, O., Lönnnerdal, B., Nilsson, A.-K. and Strömquist, M. (1992) Cloning and sequencing of human κ -casein cDNA. *J. DNA sequencing and mapping* 3, 245–246.
- Bösze, Zs., Devinoy, E., Puissant, C., Fontaine, M.L. and Houdebine, L.M. (1993) Characterization of rabbit K-casein cDNA: control of K-casein gene expression in vivo and in vitro. *J. Mol. Endocrinol.* 11, 9–17.
- Breatnach, R., Benoist, C., O'Hare, K., Gannon, F. and Chambon, P. (1978) Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequence at the exon-intron boundaries. *Proc. Natl. Acad. Sci. USA* 75, 4853–4857.
- Coll, A., Folch, J.M. and Sanchez, A. (1993) Nucleotide sequence of the goat K-casein cDNA. *J. Anim. Sci.* 71, 2833.
- Coll, A., Folch, J.M. and Sanchez, A. (1995) Structural features of the 5'-flanking region of the caprine K-casein gene. *J. Dairy Sci.* 78, 973–977.
- Devinoy, E., Hubert, C., Jolivet, G., Thépot, D., Clerque, N., Desaleux, M., Dion, M., Servely, J.L. and Houdebine, L.M. (1988) Recent data

- on the structure of rabbit milk protein genes and on hormonal control of their expression. *Reprod. Nutr. Dev.* 28, 1145–1164.
- Groenen, M.A.M. and Van der Poel, J.J. (1994) Regulation of expression of milk protein genes: a review. *Livestock Prod. Sci.* 38, 61–78.
- Jahn, G.A., Djiane, J. and Houdebine, L.M. (1989) Inhibition of casein synthesis by progestagens in vitro: modulation in relation to concentration of hormones that synergize with prolactin. *J. Steroid Biochem.* 32, 373–379.
- Hall, L. (1990) Nucleotide sequence of guinea-pig K-casein cDNA. *Nucleic Acids Res.* 18, 6129.
- Hansson, L., Edlund, A., Johansson, T., Hernell, O., Strömqvist, M., Lindquist, S., Lönnerdal, B. and Bergström, S. (1994) Structure of the human β -casein encoding gene. *Gene* 139, 193–199.
- Kunz, C. and Lönnerdal, B. (1989) Casein micelles and casein subunits in milk. In: Atkinson S.A. and Lönnerdal B. (Eds.), *Protein and Non-protein Nitrogen in Human Milk*. CRC Press, Boca Raton, FL, USA, pp. 10–27.
- Lee, W., Mitchell, P. and Tjian, R. (1987) Purified transcription factor AP-1 interacts with TPA-inducible enhancer elements. *Cell* 49, 741–752.
- Levine, W.B., Alexander, L.J., Hoganson, G.E. and Beattie, C.W. (1992) Cloning and sequencing of the porcine K-casein cDNA. *Anim. Genet.* 23, 361.
- Mount, S.M. (1982) A catalogue of splice junction sequences. *Nucleic Acids Res.* 10, 459–471.
- Nakashi, H.L., Grantham, F.H. and Gullino, P.M. (1984) Expression of K-casein in normal and neoplastic rat mammary gland is under the control of prolactin. *J. Biol. Chem.* 259, 14894–14898.
- Persuy, M.A., Legrain, S., Printz, C., Stinnakre, M.G., Lepourry, L., Brignon, G. and Mercier, J.C. (1995) High-level, stage- and mammary-tissue-specific expression of a caprine K-casein-encoding mini-gene driven by a β -casein promoter in transgenic mice. *Gene* 165, 291–296.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Schild, T.A., Wagner, V. and Geldermann, H. (1994) Variants within the 5'-flanking regions of bovine milk protein genes: I. K-casein-encoding gene. *Theor. Appl. Genet.* 89, 116–120.
- Stewart, A.F., Willis, I.M. and Mackinlay, A.G. (1984) Nucleotide sequences of bovine alpha s1- and kappa-casein cDNAs. *Nucleic Acids Res.* 12, 3895–3907.