

SHORT COMMUNICATION

Cloning and sequencing of human κ -casein cDNA

SVEN BERGSTRÖM, LENNART HANSSON*, OLLE HERNELL†, BO LÖNNERDAL‡, ANNA KARIN NILSSON* and MATS STRÖMQVIST*

Department of Microbiology, University of Umeå, S-90187 Umeå, Sweden,

*Symbicom AB, P.O. Box 1451, S-90124 Umeå, Sweden,

†Department of Pediatrics, University of Umeå, S-90187 Umeå, Sweden,

‡Department of Nutrition, University of California, Davis, CA 95616, USA

EMBL X66417

A cDNA encoding κ -casein of human milk was cloned and sequenced. The κ -casein cDNA was isolated from a λ gt11 library generated from mRNA prepared from a mammary gland biopsy obtained from a lactating woman. The library was screened with polyclonal rabbit antibodies raised against purified native κ -casein.

The obtained nucleotide sequence contained an ORF sufficient to encode the entire amino acid sequence of a κ -casein precursor protein consisting of 182 amino acids. This includes a tentative signal peptide of 20 amino acids and a processed protein of 162 amino acids.

KEY WORDS: cDNA cloning, human milk protein, κ -protein, κ -casein, nucleotide sequence

Several human milk proteins have been shown to perform physiological functions in the newborn (Lönnérda, 1985). Some of them have been demonstrated to have immune function or antibacterial activity, such as secretory IgA and lysozyme. Others have more nutritional roles, such as assisting in lipid digestion (bile-salt stimulated lipase) or in iron absorption (lactoferrin). A common characteristic of the above described proteins is that they all belong to the class of whey proteins; i.e. the soluble protein fraction remaining after precipitation of the caseins (Lönnérda, 1989). Considerably less is known about the physiological functions of the different caseins. Human casein consists largely of the β -casein (>80%) and only of a small fraction of

κ -casein, while α -casein, a major subunit of milk from most species, is absent (Kunz and Lönnérda, 1990). Native β -casein is a 25 kDa protein (Lönnérda *et al.*, 1990). β -casein may vary in degree of posttranslational phosphorylation; from 0 to 5, on serine and threonine residues located within the first 10 amino acid residues from the N-terminal end (Kunz and Lönnérda, 1989). In contrast, κ -casein with an apparent molecular weight of 35 kDa, is heavily glycosylated. The carbohydrate constitutes about 50%–60% of its molecular mass (Kunz and Lönnérda, 1989). The carbohydrate-dominant part can be cleaved off by chymosin resulting in a so-called glycomacropptide and the non-glycosylated para- κ -casein (Kunz and Lönnérda, 1989). Little is known about the biological function of κ -casein. To get a better understanding of the structure and function of human κ -casein and to allow studies of factors affecting regulation of its synthesis and secretion we have cloned and sequenced the cDNA coding for this protein.

mRNA was isolated and purified from a mammary gland biopsy obtained from a lactating woman. A λ gt11 cDNA library was generated using this mRNA, and subsequently screened with polyclonal rabbit antibodies raised against purified native κ -casein. Several positive plaques were obtained by immunodetection and one positive plaque was purified and digested with *Eco*RI. The isolated *Eco*RI fragment was subcloned into pUC18 and subjected to nucleotide sequence analysis. The analysis revealed that the isolated fragment did

Address correspondence to: Dr L. Hansson.

```

      20      40      60
*   *   *   *   *
GAATTCGAGAGAAAGACCTCACTGGCAGGAGAAAGTGAATAATGAAGAGTTTCTTCTAGTTGTCAA
      80      100     120     140
*   *   *   *   *
TGGCTTGGCAATTAACCTTCTTTTGGCTGTGGAGTTTCAAAACAGAAAGAACAGCATGCCATGAG
A L A L T L P F L A V E V Q N Q K Q P A C H E >
      160     180     200
*   *   *   *   *
AATGATGAAAGACCATCTCTATCAGAAAACAGCTCCATATGTCCCAATGTATATGTGCCAATAGCTATC
N D E R P F Y Q K T A P Y V P M Y Y V P N S Y>
      220     240     260     280
*   *   *   *   *
CTTATTTGGAACCAATTGTACCAAGTAGACAGCTATAGCAATTAATTAATCCATATGTGCTCGCAC
P Y Y G T N L Y Q R R P A I A I N N P Y V P R T>
      300     320     340
*   *   *   *   *
ATATTATGCAACCCAGCTGTAGTTAGGCCATGCCCAATTCCTCAGCGGCAATACCTGCCAATAGC
Y Y A N P A V V R P H A Q I P Q R Q Y L P N S>
      360     380     400     420
*   *   *   *   *
CAGCTACCACTGTGGTACGTGGCCAAAGCTGCATCATTTATTGGCCATCCCCCAAGAAATTC
H P P C V V R R P N L H P S F I A I P P K K I>
      440     460     480
*   *   *   *   *
AGGATAAAATATCATCTCTACCATCAATACCATCTGTACTGTGAACCTACACAGCTCTGCCACTGA
Q D K I I I P T I N T I A T V E P T P A P A T E>
      500     520     540     560
*   *   *   *   *
ACCAATGGTGGACAGTGTAGTCACTCCAGAGCTTTTTCAGAGTCCATCATCAGAGCACCCTGAGACA
P T V D S V V T P E A F S E S I I T S T P E T>
      580     600     620
*   *   *   *   *
ACCAAGTTCAGTTATCTCCACCTACGGCATAAAACACCAAGGAATATCAAGAACACACGACGAGAC
T T V A V T P P T A * >
      640     660     680     700
*   *   *   *   *
TTGCTTAAAGCAAAATTACTACTTTCACACTCTCTCTCAGCCATTGTCTGCTTTCAGTCAACAGAAATGT
      720     740     760
*   *   *   *   *
GATTTTCAGATGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
      780     800     820     840
*   *   *   *   *
CACAAATAAGCAACIGATTGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAGGAATTTC

```

Figure 1 Complete nucleotide sequence and deduced amino acid sequence of the human κ -casein cDNA (EMBL Data library accession number X66417). The tentative signal peptidase cleavage site is indicated by an arrow. The *Eco*RI recognition sequence in the 5' end and the nucleotides downstream of the polyA tail in the 3' end which includes another *Eco*RI site, are vector sequences derived from the cloning procedure.

not contain the entire κ -casein cDNA. This 440 bp fragment was used as a nucleotide probe in an additional plaque hybridization screening of the mammary gland cDNA library. One positive plaque containing a 857 bp *Eco*RI insert was isolated, purified and digested with *Eco*RI. The resulting fragment was cloned into *Eco*RI cleaved

pUC18. The complete nucleotide sequence of both strands of the cDNA region was determined using a T7 sequencing kit (Pharmacia), on double stranded templates. Specific oligonucleotides complementary to pUC18 or κ -casein sequences were used. The 857 bp nucleotide sequence contains an ORF sufficient to encode the entire amino acid sequence of κ -casein precursor protein consisting of 182 amino acids. This includes a tentative signal peptide of 20 amino acids and a processed protein of 162 amino acids. The deduced amino acid sequence is different at eight positions compared to the published sequence determined by amino acid analysis (Brignon *et al.*, 1985). Based on the proposed cleavage site we suggest that the peptide chain contain four additional amino acid residues than previously published (Brignon *et al.*, 1985). If so, the amino acid sequence of the N-terminal is identical to that of rat and very similar to those reported for other species, with the following amino acid sequence; E-V-Q-N.

ACKNOWLEDGEMENTS

This work is supported by grants from Symbicom AB, and the Swedish Medical Research Council (19X-05708; 16X-07922).

REFERENCES

- Brignon, G., Chtourou, A. and Ribadeau-Dumas, B. (1985). Preparation and amino acid sequence of human κ -casein. *FEBS Lett.* **188**, 48–54.
- Kunz, C. and Lönnerdal, B. (1989). Casein micelles and casein subunits in human 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.
- Kunz, C. and Lönnerdal B. (1990). Casein and casein subunits in preterm milk, colostrum and mature human milk. *J. Pediatr. Gastroenterol. Nutr.* **10**, 454–461.
- Lönnerdal, B. (1985). Biochemistry and physiological function of human milk proteins. *Am. J. Clin. Nutr.* **42**, 1299–1317.
- Lönnerdal, B. (1989). Whey protein in human milk. In Atkinson S.A. and Lönnerdal, B. (eds.), *Protein and Non-protein Nitrogen in Human Milk* (CRC Press, Boca Raton, FL, USA), pp. 29–42.
- Lönnerdal, B., Bergström, S., Andersson, Y., Hjalmarsson, K., Sundqvist, A.K. and Hernell, O. (1990). Cloning and sequencing of a cDNA encoding human milk β -casein. *FEBS Lett.* **269**, 153–156.