Cloning and sequencing of a cDNA encoding human milk β-casein

Bo Lönnerdal[1,3], Sv Bergström[2], Yvonne Andersson[3], Karin Hjalmarsson[4], Anna Karin Sundqvist[4] and Olle Hernell[3]

[1]*Department of Nutrition, University of California, Davis, CA 95616, USA,* [2]*Department of Microbiology and* [3]*Department of Pediatrics, University of Umeå, Sweden* and [4]*Symbicon AB, Umeå, Sweden*

Received 1 June 1990

A cDNA of 1065 bp encoding the human milk β-casein was cloned and sequenced using a synthetic oligodeoxyribonucleotide probe and a human mammary gland library. The nucleotide (nt) sequence contained an open reading frame sufficient to encode the entire amino-acid (aa) sequence of a β-casein precursor protein consisting of 210 aa and a signal peptide of 15 aa. The nt sequence shows 45–62% homology to those of bovine, ovine, rat, and mouse β-caseins. The highly phosphorylated site, which is responsible for the calcium-binding capacity of β-casein, the signal peptide, and a sequence encoding for an inhibitor to the angiotensin-converting enzyme seem highly conserved among the β-caseins with known sequences.

Nucleotide sequence; Human milk protein; Calcium-binding site; Sequence comparison; Angiotensin-converting enzyme inhibitor

1. INTRODUCTION

Breast-feeding is known to be superior to formula-feeding for newborn infants [1]. Not only does breast milk provide a well-balanced supply of nutrients, but also several biologically active components that are known or likely to have physiological functions in the infant [2]. Such components include secretory IgA, lactoferrin, and lysozyme, which together act in the defense against infection. Other components facilitate the uptake of nutrients from breast-milk, e.g. lipases that enhance fat digestion [3] and mineral- and vitamin-binding proteins that promote the absorption of these nutrients [4].

β-Casein is a phosphorylated protein which is present in milk of several species [5]. This protein – or its digested fragments – is thought to enhance calcium absorption by chelating calcium to its phosphorylated residues and thereby keeping it in an absorbable form [6,7]. It is also known that parts of the β-casein molecule, the so-called β-casomorphins, have opioid activity and may be involved in sleeping patterns of newborns [8]. β-Casein is the major casein subunit in human milk [9] and has a molecular mass (M_w of about 25 kDa. The polypeptide chain is phosphorylated at 0–5 serine/threonine residues close to the N-terminus, and the amino acid sequence has been determined [10].

Several milk proteins genes have been cloned and sequenced; these cloned genes have primarily been from rodents or dairy animals [11–21]. Similarly, the information gained on factors that regulate milk protein gene expression has come from these species [22–24]. Knowledge on genes coding for human milk proteins is very limited and has mostly focused on α-lactalbumin and the expression of its gene, particularly in mammary carcinomas [12]. In this study, we have cloned and sequenced the gene for human milk β-casein. The information gained also enhances our knowledge on the structure of human milk protein genes and how they are related to similar genes from other species. It also provides a foundation for studies on factors regulating the expression of genes for human milk proteins.

Correspondence address: S. Bergström, Department of Microbiology, University of Umeå, S-901 87 Umeå, Sweden

2. MATERIALS AND METHODS

A λ-gt11 human mammary gland cDNA library prepared from lactationally competent adult human mammary gland was obtained from Clontech Lab., Palo Alto, CA. The clones of the human β-casein were screened by plaque hybridization using *E. coli* Y 1090 [25]. A synthetic 42-mer oligonucleotide probe (5′-GAGCAAGGG AAGAGGCAAATGAAGATTTTCAAGATCAGTCAA-3′) corresponding to amino acids 117–130 in the β-casein sequence was synthesized [10]. The oligonucleotide probe was [γ-^{32}P] dATP-labeled using T4 polynucleotide kinase (New England Biolabs, Beverly, MA) [26]. Hybridization was carried out for 12–15 h at 40°C, and the membranes were washed and autoradiographed on X-ray film (Amersham, UK) [27]. Six positive plaques were identified in the primary screening.

Following secondary screening, phage DNA of purified clones were isolated from plate lysates [25]. Restriction mapping was performed and the β-casein gene was localized in the cloned fragment by Southern blotting [25].

One of the λ-gt11 clones carrying an insert hybridizing to the β-casein 42-mer probe was digested with the restriction endonuclease *Eco*RI, and the cDNA insert was separated from DNA by electrophoresis in 1% Sea Kem GTG Agarose (FMC BioProducts, ME, USA). The cDNA fragment was isolated by electroelution and ethanol precipitation [25]. Thereafter, the fragment was ligated to *Eco*RI-digested alkaline phosphatase-treated pUC18 DNA and

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

transformed into *E. coli* TG1. Transformants were selected on plates containing 100 μg/ml of carbenicillin, 40 μg/ml of 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) and 1 mM isopropyl-β-D-thiogalactoside (IPTG; Sigma, St. Louis, MO, USA). A recombinant plasmid carrying the cDNA insert was identified and designated pAKB1. Plasmid pAKB1 DNA was subjected to restriction endonuclease analysis. The complete nucleotide sequence of both strands of the region encoding the β-casein gene was determined, using T7 sequencing kit (Pharmacia, Uppsala, Sweden), on double stranded templates as described by the vendor. As primers for sequencing reactions, specific oligonucleotides complementary to pUC18 or β-casein sequences were used.

3. RESULTS AND DISCUSSION

3.1. *Cloning and sequencing of β-casein*

By plaque hybridization, using the β-casein oligonucleotide probe, several positive λ-gt11 clones were isolated. One clone, with an 1.1 kb *Eco*RI insert, was selected and the β-casein cDNA insert was recloned into the plasmid pUC18. The complete nucleotide sequence of the selected clone is shown in Fig. 1. The overall length of the cDNA was 1065 bp, including the poly(A) tail. The open reading frame begins from the first nucleotide 'A' at the 5′-end and codes for a signal peptide of 15 amino acids and the mature casein consisting of 210 amino acid residues. The coding cDNA of human β-casein is flanked by one 390 bp non-coding region at the 3′-terminal end and one of 48 bp at the 5′-terminus. The size of our human β-casein cDNA (1065 bp) is similar to that recently reported for ovine β-casein [19].

3.2. *Deduced amino acid sequence of β-casein*

The amino acid sequence deduced from the nucleotide sequence shows a large degree of homology with that derived from protein sequencing (Fig. 2). There are, however, some discrepancies. In particular, we found the peptide chain to consist of 210 amino acid residues. Moreover, no codons for the amino acids reported at positions 19 and 207 in the 212 amino acid sequence by Greenberg et al. [10] were found. The same investigators have, however, earlier reported the polypeptide chain to be 210 amino acid residues [28]. Other discrepancies were found at positions 15 (Thr instead of Pro), 32 (Gly vs Thr), 34 (Glu vs Gln), 104 (Ser vs Gln), 133 (Leu vs Ser), 158 (Gln vs Glu), 167 (Gln vs Glu), 169 (Val vs Leu), 173 (Gln vs Val), 192 (Thr vs Pro), 198 (Thr vs Pro), 199 (Gln vs Glu), 201–206 (Leu-Ala-Pro-Val-His-Asn vs Ser-Thr-Thr-Glx-Ala-

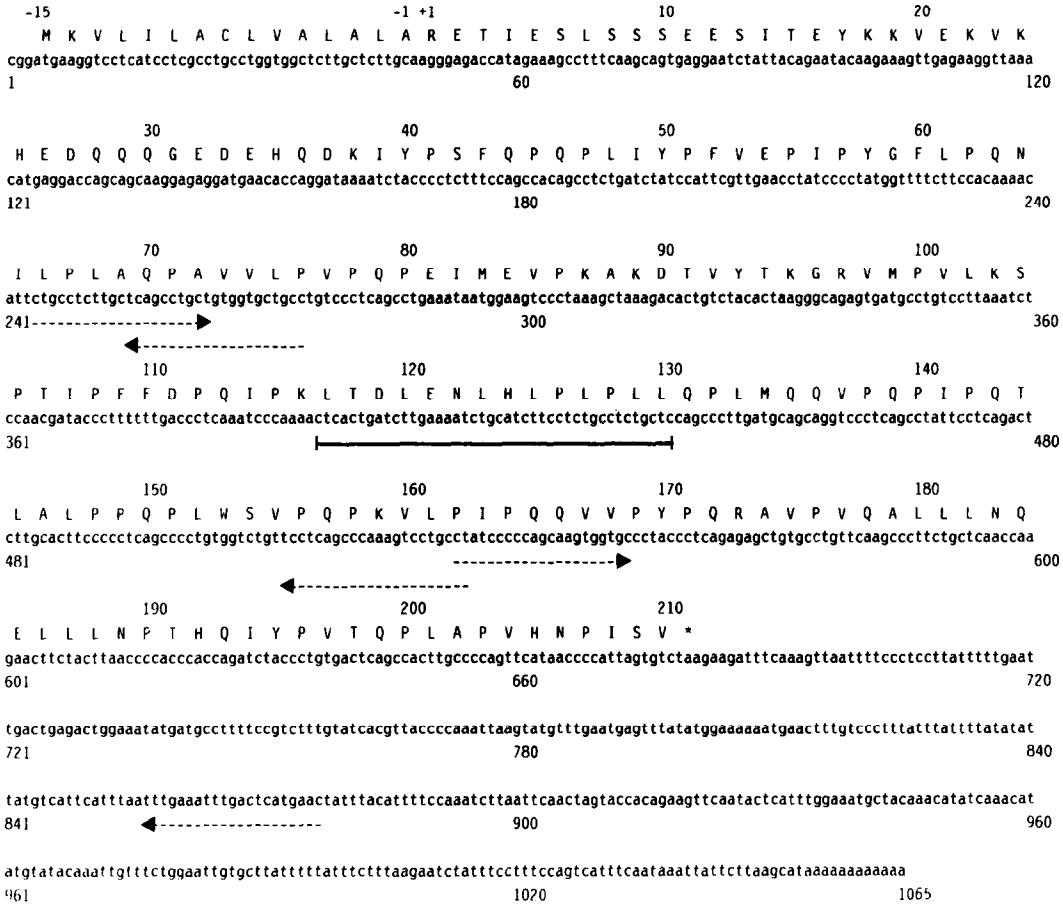

Fig. 1. Complete nucleotide and deduced amino acid sequences of the human β-casein cDNA fragment. The position of the oligonucleotide probe used for screening of the cDNA library is shown by underlining. The broken arrows show the positions of the different oligodeoxyribonucleotide primers used for sequencing of the cDNA fragment.

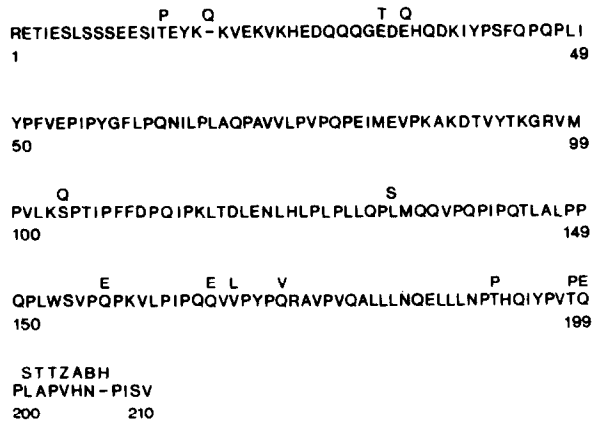

Fig. 2. Comparison of the amino acid sequence as deduced from the nucleotide sequence of the human β-casein cDNA (bottom) with the sequence obtained by β-casein protein sequencing (top, [10]).

Asz-His). Menon and Ham [29] recently described part of the human β-casein gene and also reported similar discrepancies at the C-terminal end of the protein. They suggested that this could result from genetic polymorphism. The 15 residue signal peptide of human β-casein is identical to those of bovine, ovine and rabbit β-caseins and, except for 1 residue, identical to those of rat and mouse β-caseins. As can be seen in Fig. 3 there are some single base substitutions between the species, which usually occur at the wobble position and thus code for the same amino acid. A consensus polyadenylation recognition signal AAUAAA, is located 16 nucleotides upstream from the poly(A) tail. An 11 nucleotides long motif (bp 823–833; TTTATTTATTT), which might be involved in the stabilization of the mRNA, has earlier been described in other β-casein genes [19] and is also found in human β-casein. Recently, a β-casein derived heptapeptide which inhibits bovine angiotensin 1-converting enzyme (ACE) was described [30]. A nucleotide sequence coding for a similar heptapeptide (6 of 7 residues identical) is found in human β-casein at amino acid residues 167–173, which is equivalent to the positions in bovine β-casein. Whether this peptide can inhibit human ACE is not yet fully known. Recently, however, Kohmura et al. [31] described such an activity for synthetic peptides of human β-casein, but the peptides with highest activity apparently reside in another part of the protein, i.e. at residues 39–52. A pentapeptide from positions 168–172 had moderate activity, however, and was similar to the same peptide from bovine β-caseins after ACE had cleaved off a dipeptide from the heptapeptide. Thus, it is possible that there are several sequences with such an inhibitory activity, or that only one is actually formed in vivo.

As mentioned, human β-casein is highly phosphorylated at serine and threonine residues close to the N-terminal end [10]. The nucleotide sequence found for this region has a large degree of identity to

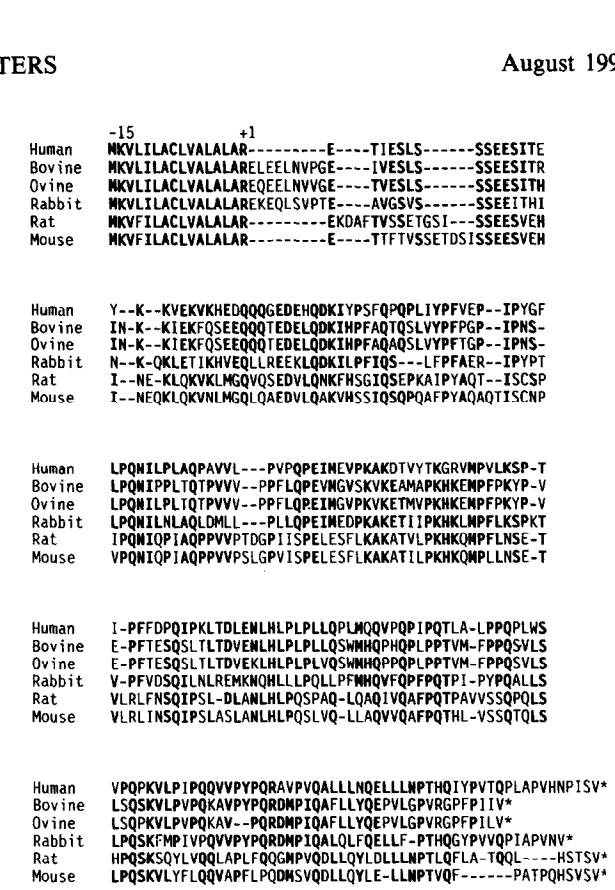

Fig. 3. Alignments of amino acid sequences derived from human β-casein (this paper), bovine β-casein [18], ovine β-casein [19], rabbit β-casein [21], rat β-casein [13], and mouse β-casein [15]. Homologous amino acids are shown in bold letters.

that of bovine β-casein (Fig. 3). It is commonly believed that this phosphorylated part of β-casein gives the molecule its capacity to bind calcium and thus, to participate in micelle formation [9]. The biological importance of this part of the molecule may explain the high degree of conservation among species.

Bovine β-casein contains a sequence of amino acids (bp 283–315) which during proteolysis form β-casomorphins and peptides that are immunostimulatory [32]. Similar peptides were not apparent in the human β-casein sequence.

Comparison between the published amino acid sequences deduced from cDNA sequences of β-caseins from several species (bovine, ovine, rat, mouse, rabbit) can be made in Fig. 3. A relatively large degree of homology, 45–62%, is found between human β-casein and the corresponding proteins from other species. A much larger degree of sequence similarity is found, however, for β-caseins from species that are more closely related such as bovine–ovine and rat–mouse.

Acknowledgements: This work was supported by grants from the Swedish Medical Research Council (19X-05708; 16X-07922), and by Symbicom AB, Umeå, Sweden. B.L. was supported as a visiting scientist by grant 19V-08905 from the Swedish Medical Research Council. The authors are grateful to Dr Staffan Normark for constructive support during this work.

REFERENCES

[1] American Academy of Pediatrics, Committee on Nutrition (1980) Pediatrics 65, 657–658.
[2] Lönnerdal, B. (1985) Am. J. Clin. Nutr. 42, 1299–1317.
[3] Hernell, O., Bläckberg, L. and Bernbäck (1989) in: Protein and Non-Protein Nitrogen in Human Milk (Atkinson, S.A. and Lönnerdal, B. eds) CRC Press, Boca Raton, FL, pp. 221–236.
[4] Lönnerdal, B. (1989) in: Milk Proteins: Nutritional, Clinical, Functional, and Technical Aspects (Barth, C.A. and Schlimme, E. eds) Steinkopff, Darmstadt, pp. 87–96.
[5] Jenness, R. and Sloan, R.E. (1970) Dairy Sci. Abstr. 32, 599–612.
[6] Mykkänen, H.W. and Wasserman, R.H. (1980) J. Nutr. 110, 2141–2148.
[7] Naito, H. and Suzuki, H. (1974) Agric. Biol. Chem. 38, 1543–1545.
[8] Brantl, V. (1984) Eur. J. Pharmacol. 106, 213–218.
[9] Kunz, C. and Lönnerdal, B. (1990) Am. J. Clin. Nutr. 51, 37–46.
[10] Greenberg, R., Groves, M.L. and Dower, H.J. (1984) J. Biol. Chem. 259, 5132–5138.
[11] Bonsing, J. and MacKinlay, A.G. (1987) J. Dairy Res. 54, 447–461.
[12] Hall, L., Emery, D.C., Davies, M.S., Parker, D. and Craig, R.K. (1987) Biochem. J. 242, 735–742.
[13] Blackburn, D.E., Hobbs, A.A. and Rosen, J.M. (1982) Nucleic Acids Res. 10, 2295–2307.
[14] Jones, W.K., Yu-Lee, L.-Y., Clift, S.M., Brown, T.L. and Rosen, J.M. (1985) J. Biol. Chem. 260, 7042–7050.
[15] Yoshimura, M., Banerjee, M.R. and Oka, T. (1986) Nucleic Acids Res. 14, 8224.
[16] Yoshimura, M. and Oka, T. (1989) Gene 78, 267–275.
[17] Jiminez-Flores, R., Kang, Y.C. and Richardson, T. (1987) Biochem. Biophys. Res. Commun. 142, 617–621.
[18] Stewart, A.F., Bonsing, J., Beattie, C.W., Shah, F., Willis, I.M. and MacKinley, A.G. (1987) Mol. Biol. Evol. 4, 231–241.
[19] Provot, C., Persuy, M.-A. and Mercier, J.-C. (1989) Biochimie 71, 827–832.
[20] Devinoy, E., Schaerer, E., Jolivet, G., Fontaine, M.-L., Kraehenbuhl, J.-P. and Houdebine, L.-M. (1988) Nucleic Acids Res. 16, 11813.
[21] Schaerer, E., Devinoy, E., Kraehenbuhl, J.-P. and Houdebine, L.-M. (1988) Nucleic Acids Res. 16, 11814.
[22] Ganguly, R., Ganguly, N., Mehta, N.M. and Banerjee, M.R. (1980) Proc. Natl. Acad. Sci. USA 77, 6003–6006.
[23] Lee, K.F., DeMayo, F.J., Atiee, S.H. and Rosen, J.M. (1988) Nucleic Acids Res. 16, 1027–1041.
[24] Rosen, J.M., Rodgers, J.R., Couch, C.A., Bisbee, C.A., David-Inouye, Y., Campbell, S.M. and Yu-Lee, L.Y. (1986) Ann. NY Acad. Sci. 478, 63–76.
[25] Maniatis, T.E., Fritsch, E.F. and Sambrook, J. (1982) in: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[26] Bergström, S., Robbins, K., Koomey, J.M. and Swanson, J. (1986) Proc. Natl. Acad. Sci. USA 83, 3890–3894.
[27] Hjalmarsson, K., Marklund, S., Engström, Å. and Edlund, T. (1987) Proc. Natl. Acad. Sci. USA 84, 6340–6344.
[28] Greenberg, R., Groves, M.-L. and Peterson, R.F. (1976) J. Dairy Sci. 59, 1016–1018.
[29] Menon, R.S. and Ham, R.G. (1989) Nucleic Acids Res. 17, 2869.
[30] Maruyama, S., Nakagomi, K., Tomizuka, N. and Suzuki, H. (1985) Agric. Biol. Chem. 49, 1405–1409.
[31] Kohmura, M., Nio, N., Kubo, K., Minoshima, Y., Munekata, E. and Ariyoshi, Y. (1989) Agric. Biol. Chem. 53, 2107–2114.
[32] Migliori-Samour, D. and Jolles, P. (1988) Experientia 44, 188–193.