

Nucleotide Sequence Evolution at the κ -Casein Locus: Evidence for Positive Selection Within the Family Bovidae

Todd J. Ward, Rodney L. Honeycutt* and James N. Derr*

Department of Veterinary Pathobiology, *Faculty of Genetics, Texas A & M University, College Station, Texas 77843-4467

Manuscript received July 2, 1997

Accepted for publication September 8, 1997

ABSTRACT

κ -Casein is a mammalian milk protein involved in a number of important physiological processes. In the gut, the ingested protein is split into an insoluble peptide (para κ -casein) and a soluble hydrophilic glycopeptide (caseinomacropeptide). Caseinomacropeptide is responsible for increased efficiency of digestion, prevention of neonate hypersensitivity to ingested proteins, and inhibition of gastric pathogens. Variation within this peptide has significant effects associated with important traits such as milk production. The nucleotide sequences for regions of κ -casein exon and intron four were determined for representatives of the artiodactyl family Bovidae. The pattern of nucleotide substitution in κ -casein sequences for distantly related bovid taxa demonstrates that positive selection has accelerated their divergence at the amino acid sequence level. This selection has differentially influenced the molecular evolution of the two κ -casein split peptides and is focused within a 34-codon region of caseinomacropeptide.

K-CASEIN is a protein in mammalian milk that determines the size and specific function of milk micelles (GUTIERREZ *et al.* 1996). These micelles increase the solubility of minerals and facilitate the transfer of nutrients from mother to offspring (DEV *et al.* 1994). The mature κ -casein protein has a labile peptide bond that is cleaved in the gut by the action of rennin to produce an insoluble peptide (para κ -casein or PKC) as well as a soluble hydrophilic glycopeptide (caseinomacropeptide or CMP) (QIAN *et al.* 1995). The function of PKC is not well known. However, CMP is responsible for clotting milk in the gut, which increases retention time and results in more efficient digestion (MERCIER *et al.* 1976). Caseinomacropeptide also reduces the immune response of neonates, preventing hypersensitivity reactions to ingested food proteins (OTANI and MONNAI 1993). In addition, species-specific glycosylation patterns of CMP can result in differential inhibition of gastric pathogens such as *Helicobacter pylori* (STRÖMQVIST *et al.* 1995). Interestingly, the different alleles described in domestic cattle for the CMP peptide of κ -casein induce significantly different amounts of glycosylation (LODES *et al.* 1996) and have been shown to produce significant differences in milk yield and protein percent (MARZALI and NG-KWAI-HANG 1986).

At the nucleotide sequence level, most mammalian protein encoding genes demonstrate a ratio of synonymous to nonsynonymous substitutions per site of ~5:1 (LI 1997). Exon four of κ -casein appears to be an excep-

tion to this rule. A recent phylogenetic study of several mammalian taxa revealed similar rates of synonymous and nonsynonymous substitutions per site within exon four of κ -casein, with the divergence at each codon position being roughly equivalent (GATESY *et al.* 1996). Although this pattern of divergence at κ -casein was interpreted by GATESY *et al.* (1996) as supporting a strictly neutral model of evolution for this gene, these authors did not compare the level of variation within exon four to that within intron four, and performed a limited analysis of their data. Because κ -casein plays a critical role in several important physiological processes, it seems unlikely that this gene would be completely free of selective constraint. Therefore, we examined the pattern of nucleotide substitution within the κ -casein gene of representative bovid taxa to determine if the molecular evolution of this gene is influenced by processes of selection.

MATERIALS AND METHODS

The nucleotide sequence of a 782-base pair (bp) segment of the κ -casein gene was determined for *Bison bison* ($n = 1$), *B. bonasus* ($n = 2$), *Bos javanicus* ($n = 3$), *B. grunniens* ($n = 3$), *B. gaurus* ($n = 1$), *Syncerus caffer caffer* ($n = 1$), *S. caffer nanus* ($n = 1$), *Tragelaphus imberbis* ($n = 1$), *Boselaphus tragocamelus* ($n = 1$), and *Capra hircus* ($n = 1$). Numbers in parenthesis refer to the number of individuals sequenced. These nucleotide sequences encompass 377-bp from exon four and 407-bp from intron four of the κ -casein gene. In addition, the corresponding *B. taurus* sequence was obtained from ALEXANDER *et al.* (1988).

Total genomic DNA was isolated from white blood cells by proteinase K treatment followed by phenol chloroform extraction (SAMBROOK *et al.* 1989). Previously published oligo-

Corresponding author: James N. Derr, Department of Veterinary Pathobiology, Texas A & M University, College Station, TX 77843-4467. E-mail: jderr@cvm.tamu.edu