Alternative splicing, a fortuitous or genetically programmed event to expand molecular diversity of milk proteins: Camel CSN1S2, a relevant model to try to provide some response elements

[1]Ryskaliyeva A., [1]Miranda, G., [2]Henry, C., [3]Faye B., [4]Konuspayeva G. and [1]Martin P.

[1]INRA, UMR GABI, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France
[2]INRA, UMR MICALIS, AgroParisTech, Plateforme d'Analyse Protéomique Paris Sud-Ouest (PAPPSO), Université Paris-Saclay, 78350 Jouy-en-Josas, France
[3]CIRAD, UMR SELMET, Montpellier, France
[4]Al-Farabi Kazakh State National University, Biology department, Almaty, Kazakhstan

Camelids are large even-toed ungulates, strictly herbivorous Mammals belonging to the order Artiodactyla including deer, giraffes, antelopes, sheep, goats and cattle, but differing from ruminants in a number of ways. They have three-chambered stomachs, and uniquely among Mammals, their red blood cells are elliptical. They also have a unique type of antibodies lacking the light chain, besides the normal antibodies found in other Mammals. Their genome is particularly "fragmented" with 74 pairs of chromosomes.

Camel's milk has proved or supposed therapeutic virtues. It is renowned for its ability to strengthen the immune system, prevent or alleviate autoimmune diseases, including ulcerative colitis and Crohn's disease. It would also have a prophylactic power on diabetes. In addition, camel milk is extremely rich in vitamin C, and the composition of its protein fraction is intriguing. Camel milk contains large amounts of an antimicrobial protein, the peptidoglycan recognition protein (PGRP), known as an intracellular component of neutrophils, which is present at a very low level in ruminant milks. It contains WAP, like rodents and lagomorphs, whereas Lysozyme C, which is an important component in mare's milk, is absent.

Up to now the composition of its casein fraction appeared to be relatively well established. However, analyzing milk of camelids originating from Kazakhstan, both in *Camelus dromedarius*, *Camelus bactrianus* and their hybrids, with powerful proteomic (LC-MS, LC-MS/MS), genomic (RNA sequencing) and bioinformatic tools, an unexpected complexity was observed, having probable consequences at the technological and nutritional levels. Indeed, a great diversity of molecular species, originating in genetic variants, post-translational modifications but also in the processing of primary transcripts, was highlighted. This situation is particularly conspicuous regarding αs2 casein for which 3 splicing variants were identified, including exon skipping and cryptic splice site usage, with phosphorylation levels for each of them ranging between 7 and 12 Phosphate groups. Such result provides useful novel information for understanding the evolution of casein genes and their expression across Mammals.

With the growing number of genes encoding milk proteins sequenced and displaying complex patterns of splicing, thus increasing the coding capacity of genes, the extreme protein isoform diversity generated from a single gene can no longer be considered as an epiphenomenon. Is it a fortuitous or a scheduled event to expand molecular diversity of milk proteins. Structural diversity and variability in expression level are both responsible for modifications in the organization and, consequently, changes in the physico-chemical properties of the casein micelle. A parsimonious vision of this issue addresses a major question: does this convey any biological significance? Important new insights are expected, in this field, in the near future