

A Unifying Nomenclature for Keratin Associated Proteins

Authors: Jon Hickford, Gong Hua, Zidong Yu, Jeffrey Plowman, Hiutong Zhou, Grant McKenzie

Most protein in hair and wool is of two broad types: keratin intermediate filament proteins (KIF) and keratin associated proteins (KAP). KIF nomenclature has recently been reviewed and a unifying system proposed. By comparison KAP nomenclature has not been reviewed recently and despite a large increase in the number of KAP genes (currently called KRTAP) identified in both humans and other species.

KAP genes were originally first identified and categorised in sheep, reflecting both the economic importance of wool and the preponderance of wool protein biochemistry undertaken around the mid-twentieth century. Accordingly KAP proteins were originally categorised on the basis of their amino acid composition into high sulphur, ultra-high sulphur and high glycine/tyrosine categories, a nomenclature that is still useful, but now appears to ineffective in describing the diversity of KAP's.

With the advent of large-scale whole genome sequencing over 80 human KAP genes have been identified and placed in 27 families, each comprising 1-12 members. This basic categorisation would appear to be quite useful for characterising wool KAP genes given the homology now being reported between the human and sheep genomes.

In addition to this, extensive polymorphism is now being described in both human and ovine KAP genes. This polymorphism is probably an underlying factor in the observed variation in hair and wool traits. Because of the economic importance surrounding the potential influence of polymorphism on wool traits there is good reason to accommodate this genetic variation in the KAP nomenclature scheme. We believe the current KAP nomenclature can be easily modified to become a consensus system for all mammalian species. Given the higher levels of diversity among KAP genes from species of more distant phylogenetic relationship, this system needs to be both flexible to accommodate variation and informative.

We suggest a KAP protein or gene name should have the form:

“*sp*-KAP m - npl * x ” where:

Sp is a unique abbreviation of the species, for example, “*hosa*” for *Homo sapiens*, “*ovar*” for *Ovis aries*, “*bota*” for *Bos taurus* and “*Mumu*” for *mus musculus*, for both genes and proteins the name becomes **KAP** (mirroring the change from KRT to K), **M** is a number identifying the family, **n** is a constituent of that family, **p** signifies a pseudogene if present or if there is an obvious fault in the gene (e.g. presence of an unexpected stop codon), **l** if present, signifies “like” and refers to a temporary “place-holder” until the family (or constituency within that family) is confirmed and **x** is an alphabetical letter signifying a variant or allele, but preferably at the level of a haplotype encompassing the promoter, 5' UTR, exons and introns and 3'UTR regions. To simplify the nomenclature, we propose using the same text for the protein and the gene, with non-italicised text for the protein and *italicised text* for the gene.

The establishment and maintenance of such a consensus nomenclature system should facilitate more efficient storage and retrieval of data and define a common language for these genes and proteins. However, in common usage in literature and reflecting the importance of the history of

gene and protein discovery, we suggest that for some time historic names such as B2A or KRTAP 1-1 are bracketed after the first use of the description of the gene or protein using this new system.