New consensus nomenclature for mammalian keratins

Jürgen Schweizer,1 Paul E. Bowden,3 Pierre A. Coulombe,4 Lutz Langbein,5 Thomas M. Magin,6 Lois Maltais,7 M. Bishr Omary,8 David A. D. Parry,9 Michael A. Rogers,9 and Mathew W. Wright10

Keratins (previously also called cytokeratins) are filament–forming proteins that provide mechanical support and fulfill a variety of additional functions in epithelial cells. In 1982, a nomenclature was devised to name the keratin proteins that were known at that point. The systematic sequencing of the human genome in recent years uncovered the existence of several novel keratin genes and their encoded proteins. Their naming could not be adequately handled in the context of the original system. We propose a new consensus nomenclature for keratin genes and proteins that relies upon and extends the 1982 system and adheres to the guidelines issued by the Human and Mouse Genome Nomenclature Committees. This revised nomenclature accommodates functional genes and pseudogenes, and although designed specifically for the full complement of human keratins, it offers the flexibility needed to incorporate additional keratins from other mammalian species.

What’s in a name? That which we call a rose, by any other name would smell as sweet (Romeo and Juliet, Act II, Sc. ii). Keratins (previously also called cytokeratins) are filament–forming proteins of epithelial cells and are essential for normal tissue structure and function. Keratin genes account for most of the intermediate filament genes in the human genome, making up the two largest sequence homology groups, type I and II, of this large multigene family. They are highly differentiation–specific in their expression patterns, implying functional differences. Mutations in most of them are now associated with specific tissue–fragility disorders, and antibodies to keratins are important markers of tissue differentiation and, therefore, tools in diagnostic pathology. Since the first keratins were sequenced and identified as type I and II intermediate filament proteins, the increasing numbers of keratins has provided an ongoing challenge for their clear identification and logical classification across species.

The first attempt at providing a comprehensive keratin nomenclature dates back to 1982. Moll et al. (1982) used 2D isoelectric focusing and SDS-PAGE to map the keratin profiles of a large number of normal human epithelia, tumors, and cultured cells. They grouped the basic–to–neutral type II keratins as K1–K8 and the acidic type I keratins as K9–K19 (Moll et al., 1982). Although not open–ended for type II keratins, this system has so far proven manageable, as the incorporation of a few novel type II keratins could be accomplished by the addition of discriminatory suffix letters to keratins exhibiting similar gel–electrophoretic properties (Collin et al., 1992a,b; Winter et al., 1998). Moreover, the Moll nomenclature has not been further challenged by the “hard” α–keratins of hair and nail (hair keratins), as these keratins were named Ha (acidic, type I) or Hb (basic to neutral, type II) followed by a number, with H standing for hair (Heid and Franke, 1986; Rogers et al., 1998, 2000). Overall, however, the present naming of keratins has not been systematic, and a reorganized and durable scheme is long overdue.

Genome analyses have recently demonstrated that humans possess a total of 54 functional keratin genes, i.e., 28 type I and 26 type II keratins, forming two clusters of 27 genes each on chromosomes 17q21.2 and 12q13.13 (the gene for the type I keratin K18 being located in the type II keratin gene domain; Hesse et al., 2001, 2004; Rogers et al., 2004, 2005; Table I). Recognition of the extent of this large mammalian gene family led to a suggested revised nomenclature (Hesse et al., 2004) based on an extended Moll system, K1–K8, and K9–K24 (Moll et al., 1982, 1990; Chandler et al., 1991; Zhang et al., 2001; Sprecher et al., 2002), and conceptually close to an earlier proposal (Rogers and Powell, 1993). In this nomenclature, all

Correspondence to Jürgen Schweizer: schweizer@dkfz.de
Abbreviation used in this paper: HGNC, Human Genome Nomenclature Committee.
human type I keratins were named Ka9 to KaX and all type II keratins were named Kb1 to KbY, thus, enabling type I and II keratins of other mammalian species to be added consecutively into this open-ended system. At the 2004 Gordon Conference on Intermediate Filaments in Oxford, an initiative to achieve international consensus led to the formation of a broad-based Keratin Nomenclature Committee that included active investigators in the keratin field and members of the Human Genome Nomenclature Committee (HGNC) and the Mouse Genome Nomenclature Committee. This committee evaluated several potential nomenclature schemes and, after extensive deliberation and consultation with other colleagues in the intermediate filament field, arrived at the consensus nomenclature system that is detailed in the following sections.

To structure the new nomenclature system, the 54 human keratins and their genes are divided into three categories: (1) epithelial keratins/genes, (2) hair keratins/genes, and (3) keratin pseudogenes. The nomenclature is also structured to allow for the inclusion of a fourth category of nonhuman epithelial and hair keratins of other mammalian species, whose genes are either absent or occur as pseudogenes in the human genome.

For both type I and II keratins, these four categories are numerically arranged in the following order (Table II): (1) human epithelial keratins, (2) human hair keratins, (3) nonhuman epithelial/hair keratins, and (4) human keratin pseudogenes.

Because of historical reasons and the extensive number of existing publications, the Moll designation for the epithelial keratins K1–K8 and K9–K24 (Moll et al., 1982, 1990; Chandler et al., 1991; Zhang et al., 2001; Sprecher et al., 2002) is retained, as is the existing HGNC gene designation scheme (i.e., KRT#).

### New nomenclature for type I keratins

#### Human epithelial keratins

After the classical type I epithelial keratins K9–K24, the numbers 25–28 were consecutively assigned to the four recently identified epithelial keratins K25irs1–4, which are differentially expressed in the inner root sheath of the human hair follicle (Bawden et al., 2001; Rogers et al., 2004). Thus, the numbers 9–28 represent the 17 human type I epithelial keratins and their genes (Tables I, II, and III [columns 4 and 5]). Numbers 11, 21, and 22 are unused for historical reasons. K11, for example, was originally thought to be a unique gene product (Moll et al., 1982), but was later confirmed as a polymorphic variant of K10 (Korge et al., 1992; Mischke, 1998). K21 (Chandler et al., 1991) has now been shown to be a rat orthologue for human and mouse K20 (Moll et al., 1993; Zhou et al., 2003), and it has therefore been removed from the category of human type I epithelial keratins. Position number 22 was previously reserved by the HGNC, but has never been used.

#### Hair keratin genes

The numbering of type II keratins

#### New nomenclature for type II keratins

#### Human epithelial keratins

Compared with the classical type I epithelial keratins, significantly more adjustments had to be made within the family of type II epithelial keratins K1–K8 (Table IV). The numbering of additional type II epithelial keratins begins with K71, after the nonhuman mammalian type I keratins (Table II). Thus, K71–K74 were assigned to the four type II inner root sheath keratins K6irs1–4 (Langbein et al., 2003; Table IV).

The number of distinct variants of the K6 gene has been a matter for some discussion, but sequencing of the human genome has revealed that there are, in fact, only three K6 variants, each encoded by their own gene (Rogers et al., 2005). These are now designated K6a (KRT6A), K6b (KRT6B), and K6c (KRT6C). Because of its lack of conformity to the rules of the HGNC, the hair follicle–specific keratin K6hf (Winter et al., 1998; Wang et al., 2003) has been repositioned after the last inner root sheath.
keratin K74 and so renamed K75 (Table IV). For similar reasons of identity and nomenclature conformity, the epidermal keratin K2e has been redesignated K2, and the palatal keratin K2p (Collin et al., 1992a,b) has been renamed K76 (Table IV, column 4). Although the completion of sequencing and analysis of the human type I keratin gene domain did not reveal any new keratin genes, that of the type II keratin gene domain led to the detection of four hitherto unknown genes whose encoded proteins had previously been designated K1b, K5b, K6l, and Kb20 (Hesse et al., 2001, 2004; Rogers et al., 2005; Table IV, column 1). Although there is only limited expression data available for K5b, K6l, and Kb20 (Rogers et al., 2005), keratin K1b has recently been demonstrated to be specifically expressed in eccrine sweat glands (Langbein et al., 2005). In the new nomenclature, these keratins are designated K77–K80 (Table IV, column 4). Collectively, K1–K8 and K71–K80 cover the twenty human type II epithelial keratins (Tables I and II).

Human hair keratins

Somewhat by design, the direct continuation of the numbering of the six human hair keratins (Table I), i.e., K81–K86, led to last digit matching between the old (Table IV, column 1) and the new nomenclature (Table IV, column 4), so that this kind of “aide-mémoire” exists for both types of human hair keratins.

Nonhuman epithelial/hair keratins

Unlike the situation for type I keratins, there are currently no type II keratins in this category. According to a recent study,
putative mouse and rat candidate genes might be located in a region syntenic to the KRT6B, KRT6C, and KRT6A subcluster in the human genome (Hesse et al., 2004). The numbering allocated to such potential keratins ranges from 87 to 120 (Table II), thus, covering approximately the same range as that assigned to their type I counterparts.

New nomenclature for human type I and II pseudogene

Human keratin pseudogenes of both types were also included in the new nomenclature system, to give all-inclusive schemes of the type I and II keratin gene chromosomal domains, as shown in Fig. 1. Whereas the slots reserved for type II pseudogenes extend from 121–220, type I pseudogenes have an open system starting with 221 (Table II), as this is the last category to be named.

Presently, there are eight keratin pseudogenes designated KRT121P–KRT128P located within the type II keratin gene domain on chromosome 12q13.13 (Table IV, column 5). Of these, KRT121P–KRT124P are hair keratin pseudogenes and KRT125P–KRT128P are epithelial keratin pseudogenes (Fig. 1). Should some of these pseudogenes have active counterparts in other species, the encoded keratins will retain the numbering of the respective pseudogene without the suffix “P,” and,
depending on their type, will be included in the respective category of nonhuman epithelial/hair keratins. The positions above 128 can be used to include keratin pseudogenes identified in other mammals.

The type I keratin gene domain contains two hair keratin and three epithelial keratin pseudogenes (Table III and Fig. 1). Two of them, the hair keratin pseudogene KRT41P and the epithelial keratin pseudogene KRT42P, possess active gene counterparts in other species and have already been named accordingly. The remaining three genes were designated KRT221P–KRT223P (Table III, column 5).

It should be emphasized that in addition to the aforementioned type I and II keratin pseudogenes in the keratin gene domains on chromosomes 17 and 121, there are at least 61 processed pseudogenes for the type II keratin K8 and 77 for the type I keratin K18, which are dispersed throughout the human genome. Moreover, there are five processed pseudogenes for the type I keratin K19, which are single pseudogenes located on chromosomes 4, 6, and 10, and with two pseudogenes located on chromosome 12. None of these contains an intact reading frame. Furthermore, the terminal segment on the human type I keratin gene domain spanning genes KRT14, KRT16, KRT17, and KRT42P (Fig. 1) is inserted four times into different regions of chromosome 17. This gives rise to three unprocessed pseudogenes for K14 and K16, and four for K17, as well as four KRT42P pseudogenes, which are all assumed to be nonfunctional (Hesse et al., 2001, 2004). The decision as to how these pseudogenes will be included into the respective lists of type I and II keratin pseudogenes will be left to the HGNC.

Concluding remarks

This modified and unifying nomenclature for mammalian keratins preserves the widely used and broadly referenced Moll designation system for the classical human epithelial keratins K1–K8 and K9–K24. The few changes that have been introduced reflect constraints to restrict the protein designation type K$nS$ ($n =$ a number; $S =$ a letter) to true keratin isoforms. Accordingly, the new nomenclature will have no significant impact on current textbooks and commercial catalogs. Major changes are nearly all restricted to recently identified epithelial keratins. However, laboratories working on hair keratins will have to, under the new system, part with the 20-year-old designations HaX and HbY and adopt new nomenclature. An effort has therefore been made to preserve the last digit between the old and the new designations, which is hoped will help researchers to adapt the new names. Although the new nomenclature assembles the human type I keratins into an almost uninterrupted series by preserving the original Moll nomenclature where possible, there is an unavoidable gap in the numbering of human type II keratins. Sufficient space has been left in the system for keratins occurring in other mammalian species, as well as for keratin pseudogenes. Moreover, we suggest that the term “keratin” rather than “cytokeratin” be used and that mammalian orthologues of human keratins be given the same naming system. As the revised nomenclature should facilitate communication and understanding within the community interested in keratins and their diseases, we advocate that this new system be used in all future studies.

The members of the Keratin nomenclature Committee are grateful to the following colleagues for their positive comments on the new keratin nomenclature: Bob Goldman, Kathy Green, Werner Franke, Elaine Sachs, Rudolf Leube, Jürgen Markl, Irwin McLean, Roland Moll, Bob Oshina, Jim Reinwald, George Rogers, Dennis Roop, and Klaus Weber.

Submitted: 29 March 2006
Accepted: 14 June 2006

References


