Structure, Function, and Dynamics of Keratin Intermediate Filaments

Peter M. Steinert
Skin Biology Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, Maryland, U.S.A.

All mammalian cells contain a complex cytoskeleton composed of three principal structural proteins and their associated proteins: actin-containing microfilaments, tubulin-containing microtubules, and intermediate filaments (IF). Of these, IF are the most complex in terms of the numbers of protein chains [1]. Recent work has described a total of six distinct types of IF, including approximately 15 acidic type I and 15 neutral-basic type II keratins of epithelia; four different type III proteins, such as vimentin of many mesenchymal cell types; four type IV neurofilament chains; one to three type V lamins of all eukaryote cells; and nestin, a single type VI protein of neuroectodermal stem cells. Type I/type II keratin intermediate filaments (KIF) are the major differentiation products of the epidermis and its derivatives. The KIF adopt a complex array in cells, forming from an elaborate cage-like network around the nucleus and indeed are in intimate contact with the nuclear karyoskeleton lamina complex. Emanating from the nuclear region, cables of KIF (tonofilbr bundles) course throughout the cell and impact the cell periphery at specialized junctions, such as desmosomes and hemidesmosomes [2]. The structural continuity of the KIF within cells has led to the suggestion that they are involved in many aspects of cellular behavior [1-4]. Their apparent continuity into neighboring cells in three dimensions implies a critical role in the maintenance of the structural integrity of an entire epithelium such as the epidermis.

In recent years, a much clearer image of the dynamic organization and function of KIF in cells has emerged. This has occurred concomitantly with newer data on the structure of KIF and the discovery that simple point mutations in KIF chains can cause several types of pathology. The purpose of this review is to summarize recent spectacular advances in our understanding of the biology of KIF and their role in diseases.

KIF DYNAMICS IN LIVING CELLS

Although current textbooks infer that IF in general and KIF in particular are relatively static entities that serve merely to support the nucleus and provide tensile strength to a cell, several recent types of data show that in fact IF are highly dynamic structures as well. In mitosis, for example, the nuclear lamina complex, built largely of the IF lamin proteins, is reversibly disassembled and reassembled at cytokinesis as the chromosomes condense. This process is tightly controlled by phosphorylation of the lamin protein chains by a series of specific protein kinases that cause disassembly into soluble tetramers. Subsequently, at metaphase, the lamins are dephosphorylated, allowing reassembly of the lamina complex [4-8]. Many fibroblasts reversibly disassemble their vimentin IF during mitosis concomitantly with the lamins, also under the influence of phosphorylation/dephosphorylation [6,9,10]. Certain immortalized epithelial cell lines, such as HeLa cells, likewise transiently disassemble their KIF [11]. However, most epithelial cells, such as primary epidermal keratinocytes, do not. Rather, whereas their nuclear lamins are disassembled, the cytoplasmic KIF networks are retained essentially intact into late cytokinesis, at which time the networks are divided into two [11,12]. Nevertheless, the KIF proteins of post-mitotic terminally differentiating epidermal cells also undergo rapid phosphate exchange [13], the function of which was not understood until more recent microinjection experiments were performed. When injected into living cells, single keratin proteins are rapidly assimilated onto the endogenous KIF cytoplasmic network [10,14]. This occurs at innumerable sites simultaneously throughout the KIF network and is complete within 30 min. These experiments establish the existence of "soluble" unpolymerized pools of KIF protein, probably consisting of small oligomers, that exist in rapid equilibrium with the KIF networks. If a large excess of protein is microinjected, the networks are seriously disrupted because these small pools are overwhelmed. This highly dynamic exchange of protein is under the control of phosphorylation/dephosphorylation cycling. Phosphatase inhibitors that poison the reversible process promote disassembly of the networks and accumulation of "soluble" hyperphosphorylated protein [10,15]. Overall, therefore, the data establish that KIF are highly dynamic structures, continuously exchanging protein throughout their length in response to different phases of the cell cycle, cell movement, and differentiation [11,16]. As discussed below, mutations in the keratin chains can drastically alter the dynamic behavior of the KIF, resulting in different types of pathology.

HOW TO BUILD A KIF

Understanding the mechanism and functional importance of KIF dynamics will require a detailed understanding of KIF structure. All IF protein chains possess common structural features: a central α-helical rod domain of conserved secondary structure and flanking amino- and carboxyl-terminal end domains. The particular sequence properties of these three domains, and the structures of the genes that encode them, is the basis for their classification into six different sequence types, including the type I (acidic) and type II (neutral-basic) keratins [2,16,17]. I showed a long time ago that KIF are obligate heteropolymers, requiring at least two different protein chains for assembly in vitro [18]. More rigorous analysis of the expression patterns and distribution of the 30 or so keratin chains has now established that most epithelial cell types express a particular

Reprint requests to: Dr. Peter M. Steinert, NIAMS, Building 6, Room 425, NIH, Bethesda, MD 20892.

Abbreviations:
CE: cornified cell envelope
EBS: epidermolysis bullosa simplex
EH: epidermolytic hyperkeratosis
IF: intermediate filament(s)
KIF: keratin intermediate filament(s)
KIFAPs: keratin intermediate filament associated proteins
NMR: nuclear magnetic resonance

0022-202X/93/$06.00 Copyright © 1993 by The Society for Investigative Dermatology, Inc.
pair of type I and type II keratins [19]. For example, basal epidermal cells express keratin 5 (K5) (type II) and K14 (type I), and suprabasal cells committed to terminal differentiation express largely K1 (type II) and K10 (type I) [19], with smaller amounts of K2e,2p (type II) and K9 (type I) in certain body locations [20].

Current views suggest that the central rod domain in large part specifies the way in which the keratin chains associate to form KIF. Although various studies indicate that at least some of the end-domain sequences are required for KIF assembly in vitro and in vivo, it is generally thought, but not yet rigorously proved, that the end domains are important in describing the particular function of the KIF in cells [17,21]. The rod domain consists of four α-helical segments that possess a repeating heptad amino acid residue peptide motif $(a-b-c-d-e-f-g)_h$, that has the potential to form a two-chain coiled coil with a similar sequence (Fig 1A). These are interspersed with non-α-helical linkers that presumably confer flexibility to the rod. The heptad regularity of the coiled-coil structure reverses or “stutters” in the middle of the larger 2B rod domain segment (Fig 1). The end-domain sequences likewise have been divided into subdomains on the basis of sequence comparisons [17,22,23] and, in the case of keratins, these consist of H1 (both type I and type II chains) or H2 (type II chains only) subdomains immediately flanking the beginning and ends of the rod domains, V1 and V2 subdomains that contain special peptide repeats (see below), and terminal E1 and E2 subdomains (Fig 1A).

The first step in building a KIF is the formation of a two-chain coiled-coil molecule. Two keratin protein chains become aligned in parallel and in exact axial register and are stabilized by interactions between hydrophobic residues in the $a$ and $d$ positions of the heptad, which fit together in a knob-in-hole configuration [24–26]. The other heptad positions are often occupied by hydrophilic residues that specify higher orders of KIF structure. In the case of KIF, this dimer molecule is a heterodimer consisting of one type I and one type II chain. Although KIF containing K1/K14 or K5/K10 chains can assemble in vitro, nature uses the K5/K14 and K1/K10 partners, presumably to fulfill the peculiar tensile properties required for the different layers of the epidermis and because of their peculiar end domains (see below).

The next step involves the correct alignment of two, three, and/or four molecules [27,28]. Kinetic studies reveal this to be the rate-limiting step in KIF assembly in vitro, but once it is accomplished, the small oligomers serve as nuclei for further very rapid assembly [28]. In fact, it seems probable that these small oligomers represent the size of the “soluble” forms that hyperphosphorylation produces and that are involved in dynamic exchanges with the KIF networks in living cells. Recent extensive cross-linking experiments [21,29] have now established the precise alignments of the two, three, and four molecules with respect to each other (Fig 1B). Two neighboring molecules are aligned antiparallel in three possible modes: A11, in which two molecules are staggered so that their 1B rod-domain segments are approximately aligned; A22, in which the 2B segments are aligned; and A12, in which two molecules are aligned in register. A11 and A22 form at the two- and four-molecule level of assembly; A12 forms at the three- and four-molecule level. The cross-linking data confirm earlier predictions for these three modes of alignment on the basis of a variety of indirect data from electron microscopy, X-ray diffraction, and theoretical calculations of interactions [30–32]. Thus, KIF consists of alternating rows of antiparallel in-register and antiparallel staggered molecules (Fig 1C). Close inspection of alignments A11 and A22 reveal that two similarly directed molecules overlap each other by 1.6 nm (alignment A22) so that the last 10–11 residues of the 2B rod domain segment of one molecule overlap with the first 10–11 residues of the 1A rod domain segment of the next molecule. On building a two-dimensional model for the organization of neighboring molecules in a KIF with these new rules (Fig 1C), it is apparent that several key sequences overlap each other five times per unit molecule length of 46 nm (shaded in Fig 1B).

These include the beginning of 1A and the end of 2B (in modes A12 and A11); 1A with the middle of the rod, in the vicinity of linker L2 (A11); and 2B with the L2 region (A22). The fourth and fifth involve...
the end-domain sequences H1 and H2 immediately adjacent to the beginning and end of the rod domain, respectively, which also overlap with 1A, 2B, and L2 regions. Of interest, these five-sequencer regions—H1, beginning of 1A, L2, end of 2B, and H2—represent the five most highly conserved sequences throughout the entire family of IF chains (H1, H2 specific for type II keratins only) [50]. It should also be pointed out that a sixth highly conserved region common to all IF chains is the stutter region of the 2B rod domain sequence. Recent work [21] has shown that the H1 (36 residues long) and H2 (20 residues long) sequences are likely to possess a defined conformation, including a β-strand motif of the same length so that these may dock with each other to form a β-sheet in the A2 alignment mode. The H1 region is also basic in charge, so that it might dock by simple ionic interactions with the very acidic L2 and the end of the 2B rod-domain sequences (in alignments A1 or A2 and A12, respectively).

A series of experiments was designed to test directly the hypothesis that these sequences are all critically involved in organizing and/or stabilizing nearest-neighbor molecule alignments in KIF. Synthetic peptides corresponding to the H1, 1A, and 2B sequence regions specifically prevent the assembly of keratin chains into KIF in vitro and even disassemble pre-formed KIF in vitro [21,29,33]. The 1A and 2B peptides at molar ratios of only 1:1 dissociate the fundamental coiled-coil molecule, the 10-11 residue overlap window of the H1 peptide disassembles KIF down only as far as two- to four-molecule oligomers [21,29]. Apparently, these peptides compete with and displace the same sequence regions of the intact protein chains in the KIF, promoting collapse of filament integrity. Furthermore, as expected from their highly conserved sequences and structures, we found that synthetic peptides with altered charge or conformation properties are much less effective at interfering with KIF structure [21,29,34]. In contrast, whereas an L2 peptide does not interfere with KIF integrity, it does protect KIF from disassembly in equimolar mixtures by the other three; that is, the L2 sequence region normally interacts with the other sequence regions in the maintenance of KIF structure. Finally, all of these new data provide a molecular explanation as to why KIF require both a type I and type II chain for assembly in vitro and in vivo: the H1 and H2 sequences of the type II chain specify the precise molecular alignments of nearest-neighbor molecules [21]. Whereas a type II homodimer could theoretically afford the same structural requirement, a type I–type II heterodimer is nevertheless more thermodynamically stable [26,27].

However, the higher order levels of KIF structure still remain uncertain. We know from quantitative mass measurements by use of scanning transmission electron microscopy that KIF are highly polymeric with respect to their masses per unit length, ranging from approximately 12 to 24 molecules wide per molecule length of 46 nm [35]. By use of solid-state nuclear magnetic resonance (NMR), we also know that KIF are highly flexible [36], consistent with their appearance both in vitro and in vivo, which suggests a somewhat loose molecular packing. X-ray diffraction analyses of the particularly well-organized KIF of the porcupine quill tip have suggested several possible ways in which the rows of molecules may be arranged in a compact ropelike conformation [31,32], but rigorous tests of these models must still be devised. Electron micrographs of the earliest stages of KIF disassembly reveal unrolling of rows of molecules into three to six protofibrils, each of which unfolds into two to three protofilaments [34,37]. Attempts at model building show that the rows of molecules of Fig 1C do not fold together in a uniform pattern but leave a discontinuity or seam spiraling along the axis of the KIF [29,32]. Thus, KIF may unravel and even engage in dynamic molecular exchanges along this seam. Although the various types of experiments discussed have imposed important constraints on any three-dimensional model for KIF structure, there is currently a lack of information on the exact disposition of the side chains of the amino acids of the coiled coils and how these interact by both ionic and hydrogen bonds to stabilize molecular packing in folded KIF. Likewise, precise information on the role, if any, of the end-domain sequences in molecular packing is still required.

Solution NMR experiments on model synthetic peptides and protein fragments are in progress to address these questions.

**HERITABLE SKIN DISORDERS OF KIF**

One of the more exciting recent advances in keratin biology has been the discovery that single-point mutations in keratin genes, resulting in the substitutions of inappropriate amino acids, are the proximal causes of at least two different types of autosomal dominant diseases. Epidermolysis bullosa simplex (EBS), in which the physical integrity of the basal layer of the epidermis is compromised, is attributable to mutations in the K5 and K14 chains in all probands of the disease thus far reported [38,39]. Similarly, in epidermolytic hyperkeratosis (EH), which affects the structural integrity of the spinous and granular layers of the epidermis, mutations in the K1 and K10 chains are causative [34,39–41]. To date, the mutations in these two diseases have been located in the H1 region of the K1 chain [34], the 1A region of K10 [42,43] or K14 [44], the stutter region of K14 [45], and the end of 2B of K1 [43], K5 [46], and K10 [43]. Remarkably, these locations constitute four of the six highly conserved and critical molecular overlap regions for KIF structure described above. The H1 and stutter mutations involve substitutions from leucine to proline residues, which are likely to cause dramatic changes in protein conformation. The first breaks the β-strand motif in H1 and thus may interfere with the docking of neighboring molecules; the second involves a key d helical position of the 2B rod domain segment and thus may introduce an unacceptable bend in the molecule. Each of the other identified rod mutations involve net changes in charged residues within the 10–11 residue overlap window of the molecular alignment A′N (Fig 1B). Because we still do not know the three-dimensional structure of KIF, it is not clear how such changes in charged residues alter structure. Although comparisons of the many available keratin-chain sequences have revealed numerous sequence polymorphisms [22,34], most residue positions in the A′N window never vary [30], implying critical roles for KIF structure. Presumably, therefore, any changes in charge due to mutations alter key ionic and/or hydrogen bonds. Furthermore, it is known that net insertions or deletions of charged residues do affect KIF structure, both in in vitro assembly reactions [21] and as a result of phosphorylation [6,7,15]. Nevertheless, despite the lack of an exact molecular understanding at this time, it is clear that the inappropriately substituted amino acids resulting from mutations alter KIF structure. This then interferes with the normal dynamic behavior of KIF in the living epidermal cells, promoting weakened filament structure and resulting in epidermal cell lysis and tissue disintegration.

As noted above, the several identified mutations in separate probands of EBS and EH have been found in K5/K14 and K1/K10, respectively, in or near the 1A or 2B segments of the rod domains. However, several groups reportedly have been unable to locate mutations in these regions of K1 or K10 in several other cases of EH [34,42,47]. It is therefore possible that some cases of EH may be due to mutations in other regions of the chains. For example, our structural analyses suggest the L2 and H2 regions are also likely “hot spots” for damaging mutations (Fig 1B), and perhaps other end-domain sequences should not be overlooked. Alternatively, some cases may be due to mutations in the very recently discovered K2 or K9 genes that are expressed in specific regional patterns [20] or even in other genes expressed during late stages of differentiation in the epidermis [40].

One feature common to both EBS and EH is wide variation in both clinical presentation and severity of disease. In some families, the diseases are very severe whereas in others they are relatively mild or involve the epidermis in only selected anatomic sites. Within a given family, the diseases are usually homogeneous. Because K5/K14 and K1/K10 KIF are expressed in the epidermis of the entire body, it is possible that some variations may be due to epigenetic factors, such as mild physical trauma or the environment [38], or to modifying genes. Can these variations be correlated with specific types of mutations in the type I or type II keratins or even with specific sequence domains of the keratin chains? The idea here
is that the variability in severity or disease distribution or presentation might be due to amino acid substitutions that have differing consequences for KIF structure and thus cell or epidermal tissue integrity. Such a correlation would be of enormous benefit for clinical diagnosis and genetic counseling and perhaps future treatment paradigms. Two approaches are being applied to explore this issue. The first is to identify the causative mutation in each of a large number of separate probands with EBS or EH in an attempt to establish a catalog, a particularly labor-intensive undertaking. An important spin-off will be an improved understanding of KIF structure and function that may lead to novel therapeutic strategies. An alternative approach involves creation of cell lines or transgenic animals that express keratins with specific defined mutations [45, 48, 49]. The advantage of this approach is the potential use of the model systems to explore chemotherapeutic or gene therapy strategies.

ROLE OF END DOMAINS IN KIF

As we have seen, the H1 and H2 subdomains of keratin chains are important in defining molecular alignments in KIF structure. The precise dispositions and roles of the remainder of the end-domain sequences are less clear. Some early data indicated that large tracts of such sequences on the K1 and K10 chains are located at the periphery of the KIF [50]. Coupled with the observation that these sequences have been conserved in form between the K1 and K10 chains of different species [22, 23], yet are highly variable between different keratin (and other) IF chains [30], we have suggested that these end-domain sequences are important in defining the function of the KIF characteristic of a particular epithelial cell type [1, 2, 17]. If so, what is the role of the very glycine-rich sequences of K1/K10 KIF in terminally differentiating epidermis? The E1, V1, and V2 sequences of these keratins in particular are configured as a series of hypervariable peptide repeats of the form (x, y)n, where x is usually a long-chain aliphatic or aromatic residue; y is usually glycine, with an occasional hydrophilic residue such as serine; and n may vary from 3 to 32. We have proposed that these sequences adopt a novel protein conformation, termed the glycine loop [51], in which the recurrent hydrophilic residues associate, forcing the intervening glycines into a loop configuration, so that the entire V1 or V2 subdomain may adopt a compact rosette-like conformation. Our attempts to resolve the structure of this motif by use of solid-state NMR have revealed a highly flexible conformation, devoid of a precise three-dimensional shape [36]. Furthermore, the V2 but not V1 regions are polymorphic with respect to amino acid sequence, there being at least eight different alleles for K10 [52] and two for K1 [53]. Each allele differs by omission of one or more entire glycine loops. Presumably, some variability in the size of these sequences has been tolerated during evolution because it does not interfere with function, unlike the H1, H2, and rod domain sequences. But why? A potential clue to this question lies in the discovery of the protein loricrin, which constitutes up to 70% of the cornified cell envelope (CE) of terminally differentiated epidermis. We found that loricrin also consists largely of three glycine-loop motifs, interspersed by short glutamine- and lysine-rich domains that are targets for cross-linking by N²-(γ-glutamyl)lysine isopeptide cross-links catalyzed by epidermal transglutaminases [54]. Because loricrin is a late differentiation product, we envisage that the cytoplasmic surface of the CE bristles with glycine loops. We have suggested the "Velcro hypothesis" [51], in which the glycine loops of loricrin and the K1/K10 KIF interact by weak hydrophobic and hydrogen-bonding interactions, thereby stabilizing and integrating the CE with the KIF cytoskeleton within the cornified cells. The unique flexibility of the glycine-loop motifs has the particular advantage of conferring essential flexibility characteristics to the entire epidermis. These glycine-loop sequences are highly insoluble, thus rendering the K1/K10 the most insoluble KIF known. On the other hand, the end-domain sequences of the K5 and K14 chains, expressed in the inner basal cells, contain only limited glycine-loop sequences and are enriched in polar residues instead, so that these KIF are more soluble. Thus, the end-domain sequences of the K1 and K10 pair of chains have evolved to contribute to the flexible yet insoluble barrier function of the outermost layers of the epidermis.

We note that many hyperproliferative and cornification disorders of the epidermis, in which late differentiation products of the epidermis, including the K1/K10 keratins and loricrin, are expressed in low amounts, often result in defective, thickened, rigid scales with diminished flexibility and barrier function [38]. A number of biophysical and molecular biologic experiments are in progress to explore these ideas. For example, what would be the expected phenotype of transfected epidermal cells in culture or in the epidermis of a transgenic mouse that express keratins or loricrins with mutated glycine-loop motifs?

KIF-ASSOCIATED PROTEINS

This proposed interaction of the KIF with the loricrin component of the cell periphery reminds us that KIF are also involved in interactions with numerous other molecules in cells, so-called KIF-associated proteins (KIFAPs), thereby integrating tissue structure, function, and dynamics. A recent appraisal [17] has shown there are several different types of KIFAPs of at least three classes. Certain class 1 KIFAPs, such as filagrin of the epidermis [55, 63] and the high-sulfur proteins of the hair-follicle cortical cells [56], typical of terminally differentiated cells, are of low molecular weight and serve as KIF matrix proteins by binding the KIF in tight arrays. Class 2 KIFAPs, such as trichohyalin [57] and plectin [58, 59], are of high molecular weight and seem to bind KIF in loose network arrays. Certain class 3 KIFAPs, such as loricrin, desmoplakin I and II of desmosomes [60] and their hemidesmosomal equivalent, the bulbous pemphigoid antigen [61], and perhaps even the nuclear laminas [4], are thought to bind to the ends of KIF, although there is still no direct evidence that desmoplakins or bulbous pemphigoid antigen-bind keratins. The possible modes of several of these KIF–KIFAP interactions are now becoming understood. The high-sulfur proteins form numerous disulfide bond cross-links with the cysteine-rich end-domain sequences of the hair KIF [56], thereby contributing to a rigid structure. Like the high-sulfur proteins, filaggrins are also basic proteins but bind to the negatively charged rod domain portions of KIF by way of ionic interactions instead [63], thus retaining a more flexible character of the epidermis. Newer data have suggested that trichohyalin may function as a scaffold protein within the CE and also bind to KIF by ionic interactions [57], thereby contributing to the three-dimensional organization of the KIF network within the cornified cells. Plectin and desmoplakins probably bind to the rod domains of KIF through ionic interactions with peculiar 38-residue repeat motifs [58–62], although more rigorous experiments will be needed to prove this. Not surprisingly, in the case of plectin, these interactions are also reversibly altered by phosphorylation cycling [59], in concert with the dynamic properties of the KIF themselves.

SUMMARY

The recent widespread application of modern methods of structural biology, molecular biology, and molecular genetics has provided a wealth of new information on the structure and function of the KIF of the epidermis. One of the more surprising aspects of this work has been the realization of the dynamic behavior of the KIF in living cells. Perhaps one of the more exciting aspects has been the discovery and understanding of how simple, single-nucleotide–point mutations in the keratin proteins can cause defects in the KIF that in turn cause serious pathology in the epidermis. The serendipitous and coincident nature of these studies shows us how an integrated, multifaceted approach will be necessary to solve further fundamental questions and to devise useful therapeutic approaches for the management of diseases of cornification. I fully expect that these issues will advance rapidly in the near future.
REFERENCES


21. Steinert PM, Parry DAD: The conserved H1 domain of the type II keratin 1 chain plays an essential role in the alignment of nearest neighbor molecules in mouse and human keratin 1/keratin 10 intermediate filaments at the two- to four-molecule level of structure. J Biol Chem 2878–2887, 1993


basis of epidermolytic hyperkeratosis; a disorder of differentiation-specific keratin genes. Cell 70:811-819, 1992


