Molecular mechanisms underlying the assembly of intermediate filaments

Laurent Kreplaka, Ueli Aebi, Harald Herrmann

M.E. Müller Institute for Structural Biology, Biozentrum, University of Basel, 4056 Basel, Switzerland
Division Cell Biology, German Cancer Research Center, D-69120 Heidelberg, Germany

Received 22 July 2004
Available online 16 September 2004

Keywords: Intermediate filaments; Microtubules; Cytoskeleton

Introduction

The cytoskeleton of animal cells exhibits a dynamic architecture that relies on the interplay of three filament systems, that is, microtubules (MTs), microfilaments (MFs), and intermediate filaments (IFs). They are integrated into a complex, regulated network by associated and cytolinker proteins [1–3]. MTs and MFs are assemblies of globular subunits whose dynamics are controlled by nucleotide hydrolysis. In contrast, IFs assemble without the need of any cofactors from fibrous, coiled-coil forming proteins. These coiled coils are only dissociated at high concentrations of urea. Due to their intrinsic polarity, MTs and MFs serve as tracks for motor proteins to move cargoes, including IFs and their precursors [4]. In contrast, IFs are believed to be structurally nonpolar [5,6], and the integrity of the IF network depends strongly on MTs. This has been demonstrated by the massive rearrangement of IFs within the cytoplasm after disruption of MTs by colcemid [7].

Another key feature of IFs is their resistance against extraction with buffers containing nonionic detergents and high concentrations of salt [8]. This property was originally and is still used to isolate them from cells and tissues. The human body harbors more than 65 different IF proteins that are differentially expressed in complex patterns during embryonic development and that are characteristic for specific tissues or cell compartments [9]. An extreme case is the complex expression of at least 20 different hair keratins during the growth of this appendage [10,11]. This variety has to be further related to the more than 200 different cell types of the human body whose various functions and mechanical performances are linked to specific cytoskeletal compositions.

In particular, the resistance of cells to mechanical stress is probably predominantly revealed by the IF network [12]. This is because IFs are highly stable and insoluble components of human cells [13], be it for the nuclear lamins or the highly cross-linked keratins in epidermis and its appendages. Other IFs such as vimentin filaments exhibit a highly dynamic behavior as revealed by live-cell imaging techniques suggesting that IFs can be as dynamic as MTs and MFs [14,15].

The reason for the stability of IFs resides in the general structural principles of their constituting proteins that, in turn, specify their complex self-assembly pathways. Like other fibrous proteins, IF proteins are highly charged extended α-helical molecules forming dimeric coiled coils [16]. Short double-stranded α-helical coiled coils are a very common structural motif among globular proteins mediating dimerization [17–19]. In the case of IF proteins, however, four coiled-coil segments, interrupted by short linkers, generate 40- to 50-nm-long rod-like dimers with an average diameter of 2 nm. The structural information amassed over the years on various IFs in vitro and within “hard tissues” such as wool, hair, and quill now offers the opportunity to unravel the structural design underlying IF assembly pathways.
From sequence homology classes to assembly groups

Originally, IFs were classified according to their tissue-specific expression: keratins in epithelia, desmin in muscle, vimentin in cells of mesenchymal origin, glial fibrillary acidic protein (GFAP) in glial cells, and the neurofilament-triplet proteins (NF-L, NF-M, and NF-H) in neuronal cells [2,20]. With the availability of cDNA libraries from various organisms and the improvement of cloning techniques, the common nature of the constituent proteins of intermediate filaments in various tissues, including the lamins, was recognized [21,22]. On the grounds of extensive sequence comparisons, the IF family was divided into five sequence homology classes (SHC) [23]: type I constitutes the acidic keratins, type II contains the basic keratins, type III encompasses vimentin, desmin, GFAP, and peripherin, type IV represents the neurofilament triplet proteins NF-L, NF-M, and NF-H together with α-internexin and syncoilin. Last but not least, type V comprises A- and B-type nuclear lamins.

All IF proteins share a common tripartite secondary structure plan: a central, largely α-helical “rod” domain that is flanked by non-α-helical N- and C-terminal end domains “forcing” the molecules into double-stranded parallel and in-register coiled coils (see for review, see Refs. [24–26]). The dimerization of the “rod” is driven by four α-helical segments, named 1A, 1B, 2A, and 2B, exhibiting a heptad-repeat substructure (abcdefg)n in which the a and d positions are commonly occupied by hydrophobic amino acids [5]. This type of repeat generates a left-handed “hydrophobic seam” on the surface of the right-handed α-helix so that two such α-helices interact via their hydrophobic seams to coil around each other in a left-handed manner [27]. Fragments of the vimentin and the lamin A “rod” have been recently crystallized and their atomic structures solved by X-ray crystallography [28–31]. Two interesting features were observed in the sense that they differed significantly from simple prediction. The first one was the atomic structure of the end of coil 2B where the coiled-coil domain ends 10 amino acids earlier than predicted previously and reveals an “opening” of the coiled coil in the IF consensus motif exposing a cluster of glutamate residues (EGEE) [28,30]. The second feature was the detailed structure of the “stutter” of segment 2B that is a strictly conserved insert of four amino acids with respect to a continuous heptad pattern. The stutter results in a local unwinding of the coiled coil [30]. Establishing an atomic structure of the entire “rod” domain for both vimentin and lamin A is now becoming feasible [29], and meaningful models can already be proposed to explain the mechanism of IF assembly (Figs. 1A and 2A).

In vitro assembly obviously starts with the formation of a parallel, unstacked coiled-coil dimer as soon as the urea concentration drops below 8 M, although some keratins, such as K5 and K14, form heterodimers already at 9.5 M urea [32]. When the urea concentration is lowered to about 5 M, cytoplasmic vertebrate IF proteins of SHC I–IV form antiparallel, approximately half-staggered tetramers that remain stable when the urea is completely removed by dialysis into buffers of low ionic strength (2–5 mM Tris–HCl) and high pH (8.4–9.0) [33–35] (Fig. 1). Some invertebrate cytoplasmic IF proteins can also form stable tetramers at high ionic strength (150 mM NaCl) and high pH (above 8.0) [36]. Furthermore, vertebrate and invertebrate lamins (type V) remain at the dimer stage even at very high ionic strength (250–500 mM NaCl) and high pH (8.5) [37,38]. Nevertheless, these proteins form soluble complexes under the low-salt/high-pH conditions published for cytoplasmic IF proteins (Lotsch, Mücke, and Herrmann, unpublished observations).

The following assembly pathway can be observed for vertebrate cytoplasmic IF proteins upon increase of the ionic strength and decrease of the pH to 7.5:

1. A very fast lateral aggregation of approximately eight tetramers into unit-length filaments (ULFs) of roughly 60 nm in length and up to 20 nm in width is observed by electron microscopy (EM) [39] and atomic force microscopy (AFM) in buffer of rapidly fixed samples (Figs. 1B and C).
2. Longitudinal annealing of ULFs leads to the formation of up to 300-nm-long fibers within the first minute (Fig. 1D). This elongation process is associated with a significant, temperature-dependent increase of the specific viscosity [40].
3. From 3 min on, individual filaments start to reduce their diameter and adopt a much smoother surface. The viscosity reaches a plateau at about 15–20 min. At this point, filaments are about 0.5–1 µm long and “open” ones occur side by side with “compacted” ones. The assembly process is not finished at this point, because IFs “mature” during the next hour of incubation, that is, they develop into fully compacted, 10 nm wide very long filaments such that free ends are hardly observed (Fig. 1E). This compaction occurs under various assembly regimes as described early on for desmin [41]. Interestingly, dialyzing the vimentin tetramers into the final assembly buffer yields filaments with a more homogeneous mass-per-length distribution compared to that of the filaments assembled by buffer addition as determined by scanning transmission electron microscopy (STEM) [35].

For the assembly of lamins, dimers (Fig. 2B) are subjected to dialysis against physiological buffers for several hours. Thereby, three main assembly states have been revealed (Figs. 2C, D, and E):

1. Strikingly, the first aggregation phase is a longitudinal head-to-tail association of dimers yielding very flexible polar polymers (Fig. 2C; Ref. [38]).
The head-to-tail polymers associate laterally, probably in an antiparallel, half-staggered fashion to form IF-like assemblies with a distinct 25 nm “beading” repeat (Fig. 2D; Ref. [42]).

The filaments are not stable over time and further associate laterally into large paracrystalline arrays (Fig. 2E). Such paracrystals were also observed in vivo in the nucleus and the cytoplasm of insect cells after overexpression of Xenopus lamin A and Drosophila lamin C [43].

According to their differential assembly behavior, vertebrate IF proteins can be assigned to three independent assembly groups: groups 1 and 2 contain all the cytoplasmic IFs and are characterized by the formation of tetramers and ULFs; group 3 contains the nuclear IFs that have the distinct
tendency to form extended head-to-tail polymers. Group 1 is constituted by the keratins that do not significantly coassemble with members of group 2 such as vimentin and NF-L, despite rather similar molecular organization. Importantly, one should note that some cytoplasmic IF proteins such as nestin or synemin need specific coassembly partners to integrate into filaments [2].

The rod is not enough

Now that the existence of sequence homologies and assembly classes among IF proteins has been amply documented, what is the structural blueprint that can explain such precise and hierarchical assembly pathways for IFs? The first structural studies on IFs were focusing on hard keratin filaments within wool and hair, employing EM and X-ray diffraction analysis [44,45]. The main outcome was the definition of a two-dimensional lattice model precisely specifying the lateral positioning of the coiled-coil dimers along the filament axis [45]. Thirty years later, cross-linking experiments performed with various IF proteins were used to derive a concept for the three lateral modes of dimer–dimer association within mature IFs. This was taken as further proof of the surface lattice model [46,47]. These lateral association modes correspond to three different antiparallel tetramers named A11, A22, and A12 in reference to the coiled-coil segments that are interacting: A11 for the antiparallel, approximately half-staggered overlap through coil 1, A22 for the antiparallel, approximately half-staggered overlap through coil 2, and A12 for the antiparallel overlap of the full dimers.

A recent study combining analytical ultracentrifugation with cross-linking has now provided evidence how the basic non-α-helical head domain may mediate A11 interaction ([48]; see also Fig. 1A). A temperature-sensitive mutant that is arrested at the ULF stage when induced to form filaments at room temperature provided further proof that ULFs are prominently assembled from A11 type tetramers and that permanent A22 and A12 configurations are probably only generated upon elongation and compaction on the route to mature IFs [48].

To understand these findings one has to keep in mind that the “rod” domain of all cytoplasmic IF proteins is highly negatively charged at neutral pH. For example, vimentin exhibits a surplus of 24 acidic amino acids over basic ones in the “rod” domain. In contrast, the N-terminal head domain contains generally no acidic residues but a large number of positively charged arginine residues (12 in vimentin). This interesting distribution of charges within these two domains is not accidental because deletion of the human vimentin N-terminal head domain abolishes filament formation and stops the assembly at a tetrameric stage after salt addition [48]. In fact, it appears that the folding of the N-terminal head domain back onto segment 1A may be an important feature of IF protein dimers that could explain why segment 1A on its own is a monomeric α-helix in solution [29]. As a matter of fact, the N-terminal head domain dynamically controls the first lateral aggregation phase, that is, ULFs formation, by joining adjacent tetramers (Tatjana Wedig and Harald Herrmann, unpublished observations). In this context, the role of the C-terminal tail domain has remained more elusive, but the in vitro assembly of deletion mutants has shown that this domain is crucial to specify the IF diameter [35]. Hence, at least for the cytoplasmic IF proteins, we can conclude that the two end domains are modulating or controlling the assembly of the “rod”, which on its own has the tendency to aggregate or to form paracrystalline arrays in the presence of divalent cations [49].

This statement appears to be also true for nuclear lamins in a very different context. The dominating feature of lamin assembly is a head-to-tail association of dimers that is also suppressed by deletion of the N-terminal head domain [50]. This lamin domain is the shortest N-terminal head domain among all IF proteins but it also exhibits a pattern of positively charged arginines (Fig. 2A), though it contains fewer such residues than vimentin. In fact, when vimentin and lamins are aligned by superimposing the common coil 1A sequence, the lamin head is so short that it corresponds to a head truncated vimentin that is dimeric at low ionic strength and halts at the tetrameric state under assembly conditions [30,35]. Hence, the first dimer–dimer interaction of lamins follows entirely different principles than those derived for cytoplasmic IF proteins. It appears that for the nuclear lamins the short non-α-helical head is mediating head-to-tail association instead of lateral aggregation. A recent crystallographic study of a 2B fragment from human lamin A has revealed that the reason for this head-to-tail overlap may be caused by the electrostatic interaction that involves the negatively charged tip of coil 1A, a pattern of alternating positively and negatively charged segments along coil 2B, and a positively charged head domain (Fig. 2A; Ref. [31]). In this scenario too the role of the C-terminal end domain is unclear, but it is definitely engaged in the head-to-tail polymer formation because tailless Drosophila lamin Dm0 reveals lateral association of head-to-tail dimers [51]. These two distinct examples emphasize how the atomic structure of “rod” domain coiled-coil fragments can be combined with the wealth of biochemical and EM information available to produce plausible molecular explanations of in vitro IF assembly pathways. Nevertheless, our knowledge on IF protein assembly is still rather rudimentary and calls for more crystallographic work to be done.

As yet, there have been until now no direct observations of the elongation and compaction stages of the cytoplasmic IFs. For the lamins, there is still the need for a robust assembly assay in which the various subsequent stages can be followed in real time. Moreover, lamins are isoprenylated in vivo, most likely to enhance their targeting to the nuclear membrane [52,53]. Hence, it would be interesting to design an assay to study the influence of lipid membranes and
isoprenylation on the lamin assembly pathway. It has furthermore to be established if A- and B-type lamins follow the same assembly mechanisms. Their coassembly within the nuclear lamina may indeed be mediated by associated factors such as HP1, LBR, LAP2, and emerin [54]. One possibility to tackle these questions is the use of AFM as a tool to follow IF assembly in buffer solution in real time [55,56].

**IF assembly in living cells**

One might wonder to what extent in vitro assembly pathways relate to the in vivo situation. Concerning the lamins, very few data are available because live-cell imaging methods are not well suited to study interfacial mechanisms such as nuclear envelope and lamina formation. However, in the case of cytoplasmic IFs, a plausible scenario can already be proposed. Everything starts with the production of the coiled-coil dimer, and this step is universal for all IF proteins including the nuclear lamins. We still do not know if the interaction between the two chains occurs cotranslationally, thereby preventing the misfolding of individual molecules due to their pronounced hydrophobic seam, or if chaperones are engaged to stabilize the monomers and only later, during or after synthesis, allow their dimerization. However, the smallest IF subunit to be traced in the cell is, at least for vimentin, not the coiled-coil dimer but a species with biochemical properties expected for a tetramer [57]. This correlates well with the observation that the association of two coiled coils to form a tetramer already occurs under denaturing conditions in vitro [58]. Interestingly, the tetramers described by Soellner et al. [57] are generally considered to represent the soluble pool of IF proteins within the cytoplasm that is in equilibrium with tetramers already being integrated in the IF network (for a discussion, see Ref. [5]). It is most likely that the first stable assembly intermediates in the cytoplasm may indeed be short half-lived soluble complexes [59] that are actively and rapidly transported to and integrated into the cellular IF system [60]. Taking into consideration the available in vitro data, a reasonable candidate would be the ULF that, in turn, would have to be stabilized and thereby prevented from unscheduled longitudinal annealing to other ULFs by binding, for instance, to acceptor complexes of motor proteins in a “capped” form. Alternatively, ULFs could grow on such complexes and, after obtaining a critical length, be transported in an MT-dependent fashion to IF ends. This type of scenario is in total agreement with the behavior of GFP-tagged vimentin obtained by live cell imaging [14]. A similar behavior has been observed for keratin IFs except that the distribution within the cell follows different principles [15,61]. Short vimentin IFs are moved from the cell center to the periphery whereas equivalent keratin IF particles appear first at the cell periphery before they are transported inward in an energy- and MT-dependent manner. Notably, IFs do not reveal an obvious organizing center of filament formation such as, for example, the centrosome for MTs. Instead, cytoplasmic IF proteins appear to first assemble into ULFs and short filament segments in situ close to the sites of protein synthesis. This targeting of IFs may not be related to their assembly pathway because vimentin mutants that are unable to assemble are delivered to a restricted number of foci in the cytoplasm too [62]. Most interestingly, copolymerization of different IF proteins in vivo may involve specific “assembly assistants” like NUDEL, as demonstrated recently for the neurofilament-triplet proteins [63].

One way to validate this general framework was to express mutated and truncated IF proteins in cells. A straightforward example was given by headless human vimentin that does not form filaments in vitro, and was diffusely distributed throughout the entire cytoplasm when expressed in stably transfected cells [64]. Working with a tailless IF protein appeared much more complex, in line with the far less elusive role of the tail domain in IF formation in vitro. Nevertheless, transfection of tailless keratins 8 and 18 into keratin-free fibroblasts revealed that the C-terminal end domain is not essential for proper assembly and network formation [65]. However, within one culture of stably transfected cells, individual cells will organize, maybe depending on the cell cycle, tailless keratin IFs exclusively in the cytoplasm, in the nucleus, or in both the cytoplasm and the nucleus. The cause for this ectopic localization appeared to reside within the IF consensus motif YRKLEGEE that is located at the end of segment 2B, that is, immediately flanking the C-terminal end domain. To explain nuclear targeting, the two basic residues in e and f positions of the α-helix of this motif were assumed to form, in conjunction with the corresponding residues on the other chain of the coiled-coil dimer, a “conformation-type” discontinuous nuclear localization signal [64]. This was formally proven by expression of tailless *Xenopus* vimentin. This truncated protein is assembly incompetent at 37°C and will not be targeted to the nucleus any more when this motif has been removed in addition to the tail domain [64]. The necessary part of the tail for cytoplasmic retention has indeed been very precisely mapped to a highly conserved sequence motif residing within the tail domain [64]. Hence, the C-terminus is either containing a cellular “retention signal” or it is reacting with the end of coil 2 thereby masking the consensus motif and preventing its interaction with karyopherin-type molecules.

**Conclusion**

For a long time, IF proteins were considered as “dull molecules” with no important cellular function except for structural support, and difficult to characterize in vitro due to their fibrous nature. Slowly but definitely, this view is now clearly outdated because the handling of long-coiled-
coil proteins in vitro has dramatically improved, in conjunction with the increasing amount of mutations giving rise to complex genetic human diseases including such complex disorders as premature ageing [66,67]. In this context, the structure-based analysis of IF assembly pathways, first in vitro and later in vivo, is becoming the new challenge of cytoskeleton research. Interestingly, the molecular description of IF assembly intermediates is pushing classical structural methods, such as X-ray crystallography and electron microscopy, to their limits thereby calling for methodological improvements [68] and novel interdisciplinary approaches. Yet another challenge will be to define what biological role the huge surface of IFs plays in the multitude of physiological processes within the various cell types [69].

Acknowledgments

We thank Sergei V. Strelkov for constructive discussions, critical reading of the manuscript, and for his help with the figures. H.H. acknowledges support from the German Research Foundation (DFG). U.A. acknowledges support from the Swiss National Science Foundation (SNF), the M. E. Müller Foundation of Switzerland, and the Canton Basel Stadt. L.K. acknowledges support from the Swiss Society for Research on Muscular Diseases.

References


