Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 92–96, January 1995
Biochemistry

The importance of intramolecular ion pairing in intermediate filaments

(structure/α-helices/keratin/coiled-coil/ion pairs)

ANTHONY LETAI AND ELAINE FUCHS*

Howard Hughes Medical Institute, Departments of Molecular Genetics and Cell Biology, and of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL 60637

Communicated by Hewson Swift, The University of Chicago, Chicago, IL, September 21, 1994

ABSTRACT **Nuclear and cytoskeletal networks of 10-nm intermediate filaments (IFs) are probably ubiquitous in multicellular eukaryotes. They likely play a role in maintaining the mechanical integrity of a cell. With the exception of the nuclear lamins, IF proteins can form IFs *in vitro* in the absence of cofactors or associated proteins. Below we present data suggesting that the large α-helical "rod" domains of IF proteins are stabilized by large numbers (up to 50) of intrahelical ion pairs formed by residues of opposite charge situated four residues apart. These many ion pairs, sometimes involving up to 30% of the residues within a coiled-coil IF segment, can potentially contribute as much as 10–25 kcal/mol (1 kcal = 4.18 kJ) to the stability of a single α-helical rod. Such stabilization is likely to play a major role in the chemical and physical stability of IF networks *in vitro* and *in vivo*. An investigation of other coiled-coil proteins shows that selection for intrahelical ion pairing is not simply a property intrinsic to coiled-coil proteins. Rather, there is a correlation between the degree to which there is selection for intrahelical ion pairs and the extent to which a coiled-coil protein participates in highly ordered multimolecular interactions—e.g., as in IFs and myosin thick filaments. The propensity of putative ion pairs in some IF proteins—e.g., epidermal keratins—suggests that an underlying structural stability at the level of the monomer may play an important role in the extraordinary stability of dimers and higher ordered structures in cytoplasmic IFs.**

Given that intermediate filament (IF) proteins have not been crystallized, knowledge of their common structure has relied heavily on secondary structure predictions from primary sequence data (refs. 1–3; reviewed in refs. 4 and 5). IF proteins are subdivided into five distinct sequence types, all of which have a central 310- to 360-amino acid residue, α-helical rod domain interrupted by three short nonhelical linker segments (Fig. 1). The most highly conserved regions of IF proteins are ≈15 amino acids at each end of the rod. These regions are particularly sensitive to amino acid substitution or deletion (6–10).

The α-helical segments of IF rods contain heptad repeats of hydrophobic residues, where a and d of abcdefg sequences are frequently apolar (11). This creates a hydrophobic stripe that provides a surface for the intertwining of two IF polypeptides into a parallel, in-register coiled-coil dimer. This first step in IF polymerization (12, 13) is energetically so favorable for some IF proteins that it is stable even in buffers containing 8 M urea and a reducing agent (14). In solution, dimers align laterally in an antiparallel fashion to form tetramers (15) in staggered and unstaggered (14–23) arrangements. Lateral and longitudinal alignments of ≈10,000 tetramers give rise to

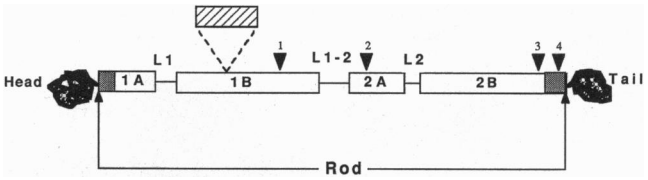

FIG. 1. Secondary structure of an IF protein. The structures of the amino-terminal head and carboxyl-terminal tail domains are poorly defined, and the configuration shown is meant only to indicate that they are nonhelical. In the central rod domain, there are four segments, denoted as 1A (37 amino acids), 1B (93 amino acids), 2A (39 amino acids), and 2B (98 amino acids), predicted to be α-helical and containing throughout the heptad repeat of hydrophobic residues involved in coiled-coil dimerization. Highly conserved rod end sequences are cross-hatched. The 42-amino acid insertion found in coil 1B of lamins and certain invertebrate IF proteins is represented by a hatched box. The rod is dissected by non-α-helical linker regions (lines), referred to as L1, L1-2, and L2. Numbered arrowheads mark the locations of four ion pairs that are highly conserved throughout the IF protein family.

≈four protofibrils intertwined to compose an IF (ref. 24 and references therein).

Sequence analyses reveal nonrandom distributions of alternating positive and negative charge in IF proteins (3, 12, 25, 26). Studies have focused on the possible role of ion pair formation in lateral interpolypeptide associations within IFs. One hypothesis is that the patterning of oppositely charged amino acid side chains across the interface of an IF dimer might aid in dimer formation (12, 25, 26). For such interactions to form within a coiled-coil, the charged side chains of an ion pair would occupy positions e and g in two opposing heptads, as outlined in Fig. 2. Relevant to IF proteins is the observation that acidic residues occur more frequently in position g than position e, whereas basic residues occur more frequently in position e than position g (4). Alignments based on optimal intermolecular ion pairing have led to the correct prediction that IF dimers are parallel and in register (27). Such ion pairing has also been found in the crystal structure of the coiled-coil dimer of yeast GCN4 protein (28), and two synthetic peptides, engineered to form this type of ion bridge when paired, assembled into a stable coiled-coil in solution (29). Such data suggest that intermolecular ion pairing is important in stabilizing coiled-coil IF dimers *in vivo*.

The remarkable stability of IFs and their subunits suggests that these structures are based upon a particularly stable α-helical conformation. In addition to α-helical enhancing features such as an abundance of helix-favoring residues and

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: IF, intermediate filament; IFAP, IF-associated protein; APC, adenomatous polyposis coli; MSP, macrophage scavenger receptor protein.

*To whom reprint requests should be addressed at: Howard Hughes Medical Institute, Department of Molecular Genetics and Cell Biology, The University of Chicago, 5841 South Maryland Avenue, Room N314, MC1028, Chicago, IL 60637.

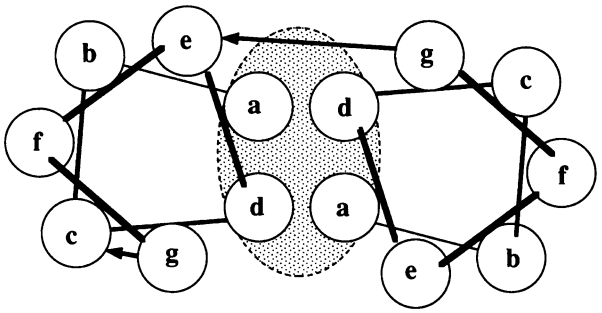

FIG. 2. Ion pairing in a coiled-coil dimer. The view is down the long axis of a dimer and shows one heptad repeat for each polypeptide chain. Residues at positions a and d are usually hydrophobic and stabilize dimer formation through the hydrophobic effect (stippled area). Interhelical ion pair is indicated by the long arrow. Intrahelical ion pairs can occur between acidic and basic residues spaced by exactly four residues, as shown, for example, by the short arrow. Such interactions can occur at any spacings of i, $i \pm 4$, which include the following positions in the coiled-coil sequences: a–e (i.e., a and e), b–f, c–g (shown), d–a, e–b, f–c, and g–d. Charged residues occurring on the outside surface of the dimer are potentially available not only for intrahelical ion pairing but also for higher-ordered interactions.

the ability to form coiled-coils, a solvent-exposed α-helix can also be stabilized by intrahelical ion pairs between oppositely charged residues. Surveys of crystallographically solved protein structures, including those of the coiled-coil proteins α-tropomyosin and myosin, indicate that in α-helical regions, there are a number of oppositely charged residues positioned exactly four residues apart (30, 31). The potential for intrahelical i, $i \pm 4$ ion pairing has also been noted in the crystal structure of the GCN4 coiled-coil (28). Biochemical studies demonstrate that i, $i \pm 4$ spacing of oppositely charged residues can indeed form ion pairs and that they are capable of stabilizing α-helices exposed to solvent (32–37). Estimates of the energetic contributions of these ion pairs to conformational stability range from 0.2 to 0.5 kcal/mol (1 kcal = 4.18 kJ) (33–36).

Thus far, the potential for intrahelical ion pairing within the IF superfamily has been examined for only one IF protein, lamin C (type V), and this protein displayed only a very modest nonrandom selection for putative ion pairing (31). At first glance, this observation might seem ominous for the importance of ion pairing in IF proteins. However, the known peculiarities of nuclear IF assembly may make lamins an exception in this regard. Given the extraordinary stability of cytoplasmic IFs, we wondered whether ion pairing might play a significant role in their assembly.

In the present study, we assessed the degree to which intrahelical ion pair formation may be involved in stabilizing α-helices in each of the five types of IF sequences and in species encompassing a broad range of the eukaryotic kingdom. Surprisingly, not only did we find a preponderance of i, $i \pm 4$ spacings characteristic of intramolecular ion pair formation but we also found that it is the most common form of charge arrangement that exists in cytoskeletal IF proteins. Additionally, the potential for this type of ion pairing has been evolutionarily conserved, underscoring the likelihood that intramolecular ion pairing plays a critical role in stabilizing the conformation of the rod domain of IF proteins.

MATERIALS AND METHODS

Protein Sequence Analysis. Nineteen IF protein sequences were chosen for the counting of basic residues (lysine and arginine), acidic residues (aspartic acid and glutamic acid), and i, $i \pm 4$ ion pairs, where i represents the position of a basic residue, and $i \pm 4$ represents the position of an acidic residue. Infrequently, putative ion pairs could form in a number of different ways and, in such cases, the arrangement was chosen to maximize the total number of ion pairs. The non-IF protein sequences chosen include sequences for (*a*) rat plectin, (*b*) human 230-kDa bullous pemphigoid antigen (BPA230), (*c*) human desmoplakin, (*d*) chicken smooth muscle myosin, (*e*) rat skeletal muscle myosin, (*f*) rat α-tropomyosin, (*g*) *Drosophila* paramyosin, (*h*) yeast GCN4, (*i*) human adenomatous polyposis coli (APC) gene product, (*j*) human α-, β-, and γ-fibrinogens, (*k*) human erythroid α-spectrin, (*l*) mouse laminins A, B1, and B2, (*m*) bacteriophage protein gp17, and (*n*) bovine type I macrophage scavenger receptor protein (MSP).

Monte Carlo Simulation. For each IF rod segment, a computer program was used to generate 1000 Monte Carlo simulations of random sequences that are the same length as the natural segment and that contain the same number of acidic and basic residues as the natural sequence. For instance, to simulate coil 1B of K14, 12 basic and 20 acidic residues were randomly distributed over a 93-amino acid array, and 1000 such randomly obtained sequences were analyzed. The program was then used to count how many putative ion pairs were formed as a consequence of random chance in each simulated sequence. A mean and standard deviation of the total number of ion pairs per simulated segment were calculated for each set of 1000 simulations. We then determined at a 95% confidence level how many standard deviations existed between the number of putative ion pairs found in the natural segment and the mean of the Monte Carlo simulated sequences. This number enabled us to calculate the probability that the number of ion pairs in the natural IF sequence might occur by mere chance positioning of the charged residues. For each IF protein, the simulations were performed two ways. In one way, charged residues were allowed to inhabit the usually hydrophobic positions a and d of each heptad. In the other way, charged residues were excluded from positions a and d. For the non-IF coiled-coil proteins, only the method where charged residues were excluded from positions a and d was used.

RESULTS AND DISCUSSION

A Remarkably Large Number of Putative Intra-α-Helical Ion Pairs in Type I–IV IF Proteins. Table 1 provides the total number of acidic and basic amino acids in each of the four α-helical rod segments of a particular IF polypeptide. Also given are the number of times a basic residue appears 4 residues away from an acidic residue in the primary structure. We refer to this occurrence as a putative ion pair or i, $i \pm 4$ spacing. Our focus is strictly on intramolecular ion pairing of this type.

For most IF proteins, the acidic and basic residues residing in the short coil 1A are not often in positions enabling participation in ion pairing. Those that are in such positions are generally not more than what could have occurred by chance. In contrast, the vast majority of basic residues in the long α-helical segments, coil 1B and coil 2B, are frequently in positions enabling participation in ion pairing. For the coil 1B and 2B sequences of most type I–IV IF proteins, the probability of a random chance occurrence of such richness in i, $i \pm 4$ spacing is $<3\%$. This calculated probability remains low whether the Monte Carlo simulations used for comparison allow or exclude charged residues from positions a and d in the heptad repeat (see Table 1). Ironically, human lamin C, the only IF that had been inspected previously for the presence of intrahelical ion pairs (31), is the IF protein demonstrating the least nonrandom selection for potential ion pairing in coils 1B and 2B. Although many ion pairs are present in these regions, the total number does not significantly differ from what would be expected from a random distribution of charged residues in lamin C. This perhaps reflects the very different assembly

Table 1. Occurrence of ion pairs within the four α-helical segments of IF rods

Protein segment	Coil 1A				Coil 1B				Coil 2A				Coil 2B			
	IP	B/A	P_{sim}	$P_{ex\text{-}hep}$	IP	B/A	P_{sim}	$P_{ex\text{-}hep}$	IP	B/A	P_{sim}	$P_{ex\text{-}hep}$	IP	B/A	P_{sim}	$P_{ex\text{-}hep}$
K1	0	6/5	1	1	9	14/19	<0.01	0.02	3	4/9	0.05	0.11	9	15/20	0.01	0.03
K5	1	6/6	0.74	0.83	11	15/18	<0.01	<0.01	3	4/9	0.07	0.10	10	17/19	<0.01	0.01
K8	1	7/4	0.63	0.71	12	14/24	<0.01	<0.01	3	4/9	0.05	0.10	10	15/21	<0.01	0.01
K10	3	6/8	0.20	0.32	9	11/21	<0.01	<0.01	3	5/8	0.10	0.17	6	8/17	<0.01	<0.01
K14	3	6/8	0.20	0.33	10	12/20	<0.01	<0.01	7	8/11	<0.01	0.01	7	11/19	<0.01	0.02
K18	2	8/8	0.74	0.84	11	14/20	<0.01	<0.01	5	6/9	<0.01	0.02	6	11/17	0.03	0.06
K19	2	6/7	0.43	0.58	10	13/20	<0.01	<0.01	5	6/9	<0.01	0.02	7	10/18	<0.01	<0.01
Ha 1	1	7/7	0.86	0.93	9	14/20	<0.01	0.02	5	5/9	<0.01	<0.01	7	10/17	<0.01	<0.01
Vimentin	2	5/8	0.42	0.56	11	16/26	0.02	0.07	3	5/8	0.10	0.19	8	13/22	0.03	0.06
Desmin	1	5/7	0.71	0.79	11	17/22	<0.01	0.03	3	5/7	0.06	0.12	10	14/22	<0.01	<0.01
NF-L	2	4/8	0.26	0.37	12	17/27	0.02	0.05	2	6/6	0.31	0.46	9	16/21	0.03	0.06
NF-M	1	4/8	0.70	0.79	12	15/28	<0.01	0.02	2	5/7	0.29	0.42	9	15/21	0.02	0.04
NF-H	1	6/6	0.74	0.82	9	16/20	0.03	0.07	2	5/7	0.28	0.44	9	15/21	0.01	0.04
Peripherin	1	4/6	0.54	0.63	12	18/27	0.02	0.07	4	7/7	0.04	0.09	9	14/21	<0.01	0.02
Nestin	2	4/7	0.20	0.29	8	16/20	0.08	0.20	4	7/8	0.06	0.14	2	11/16	0.81	0.86
Ascaris IFB	5	9/9	0.11	0.21	15	26/35	0.03	0.11	3	8/8	0.38	0.54	7	16/21	0.23	0.34
Lamin B	2	5/8	0.41	0.53	12	21/31	0.02	0.07	3	6/8	0.18	0.29	9	17/24	0.12	0.23
Lamin C	1	6/9	0.90	0.94	10	27/29	0.42	0.64	3	7/9	0.37	0.53	7	18/22	0.42	0.59
Lamin *Dr.*	1	5/7	0.75	0.82	18	27/33	<0.01	<0.01	3	6/8	0.17	0.30	9	17/25	0.14	0.28

Presented as a measure of statistical significance are estimates, made by Monte Carlo simulation, of the probability that the number of putative ion pairs (IP) could occur by chance (see text). A low probability for any given segment is a measure of whether there has been a statistically significant selection for putative ion pairs during evolution. B/A, bases/acids; P_{sim}, probabilities when Monte Carlo simulations used for comparison allow charged residues in the a and d positions of the heptad; $P_{ex\text{-}hep}$, probabilities when charged residues are excluded from the a and d positions of the Monte Carlo simulations. Such selection is observed in coils 1B and 2B of all type I–IV IF proteins surveyed and, to a lesser degree, in coil 2A of type I–III IF proteins. The IF protein sequences chosen can be obtained from GenBank and include sequences for (*i*) type I keratins: human K14, human K10, and human K19; mouse K18; a mouse hair keratin (Ha 1); (*ii*) type II keratins: human K1, human K5, and mouse K8; (*iii*) type III IF proteins: human vimentin, rat peripherin, chicken desmin, and *Ascaris* IF protein B (IFB); (*iv*) type IV IF proteins: human low molecular weight neurofilament protein (NF-L), human NF-M, and mouse NF-H; (*v*) type V IF proteins: human lamin C, mouse lamin B, and *Drosophila* lamin (lamin *Dr.*); and (*vi*) rat nestin, an IF protein that does not fall into one of the five types but that is most similar to the type IV IF class (for review, see ref. 2).

properties of the nuclear lamins compared with the cytoplasmic IFs. For the other 18 IF sequences, the number of putative intrahelical ion pairs in these domains is striking.

It is particularly revealing to examine putative ion pairing in keratins. In coil 1B, 9–12 putative ion pairs, or 20–30% of its amino acids, are found in each of eight different keratin sequences. The probabilities range from <1% to 2% that these favorable *i*, *i* ± 4 spacings between acidic and basic residues are a consequence of random chance. For coil 2B of each of these keratins, 6–10 *i*, *i* ± 4 spacings occur, again with a strong likelihood that the occurrence is not random. The numbers of putative ion pairs are comparable in the type III and type IV classes of IF proteins, with only slightly greater probabilities that they occur by random chance. Thus, throughout evolution, there seems to have been a significant nonrandom selection for such *i*, *i* ± 4 spacing in the longest α-helical segments of the type I–IV IF rods. A somewhat milder selective pressure appears to have been exerted for this spacing within the coil 2A segments of the type I–III IF polypeptides.

Coupled with previous studies revealing the importance of *i*, *i* ± 4 spacing in ion pairing (30, 31), the marked propensity of such spacings within the type I–IV IF rods makes it overwhelmingly likely that charged residues within the rod participate in intrahelical ion pairing. Solvent-exposed ion pairs contribute ≈0.2–0.5 kcal/mol toward the stabilization of an α-helical conformation (33–36). Hence, the effect of ion pairs on stabilization of the α-helical segments could be as much as 1.6–9.0 kcal/mol in coil 1B segments of IF proteins and 1.2–5.0 kcal/mol for coil 2B segments. For some IF proteins, intrahelical ion pairing could contribute as much as 10 kcal/mol to the formation of a stable α-helix, assuming that all *i*, *i* ± 4 spacings are involved in ion pairing of this sort.

IFs assembled from type I–IV IF proteins are stable structures, often resistant to disruption by heat or chemicals. It is expected that such a stable polymer could be made only if the building blocks themselves existed in stable conformations. X-ray fiber diffraction patterns confirm that the α-helical conformation persists in assembled IFs (38, 39). Thus, the higher ordered interactions that occur in an IF are likely to be facilitated by the "locking" of monomeric IF polypeptides into an α-helical conformation. In such a manner, intrahelical ion pairing would greatly promote IF assembly.

Conservation of Certain Putative Ion Pairs in IF Proteins. All type I–IV IF sequences possess the necessary features to promote substantial ion pairing in coils 1B and 2B. In several instances, the precise location of these pairs is strictly evolutionarily conserved (denoted in Fig. 1 by arrowheads). One highly conserved ion pair (arrowhead 1) is located in the second half of coil 1B. Surprisingly, the base of this pair is located at position a of the heptad repeat, a position typically occupied by a nonpolar residue. Yet in 56 of 56 IF sequences inspected, this residue is basic, and in 17 of the 19 sequences we analyzed, this basic side chain can form a putative ion pair with an acidic one at *i* ± 4. We predict that the presence of this conserved basic amino acid at a position expected to be hydrophobic derives from its ability to form an ion pair.

Another ion pair (arrowhead 4) is located in the highly conserved sequence, TYRKLLEGE, at the end of the rod domain. In 74 of 75 IF sequences searched, the basic nature of the R and acidic nature of the E residue are conserved. The exception is *Drosophila* lamin protein, where neither the base nor the acid is present. Such coordinate conservation (or lack thereof) of oppositely charged residues strengthens the notion that ion pairing plays a critical role in filament assembly. It seems likely that at sites where there is conservation of ion pair positioning, there may be a significant requirement for local stability of the α-helix.

Besides the four putative ion pairs noted in Fig. 1, the positioning of *i*, *i* ± 4 ion pairing is less conserved. Despite

variations in positioning, the number of potential ion pairs in IF proteins remains high.

Intrahelical Versus Interhelical Ion Pairing in IFs. Fourier transform analyses of the charged residues within IF rod domains have focused attention on a periodicity in the positioning of acidic and basic residues, where basic residues occur at intervals of ≈9.5 residues, and acidic residues have a similar periodicity that is ≈180° out of phase (3, 12, 25, 26). Although the structural significance of these periodicities has not yet been tested experimentally, it has been speculated that this periodicity plays a role in stabilizing higher-ordered interactions—i.e., interactions between subunits at least as large as the dimer and, most likely, in the lateral alignment of two dimers (3, 12, 26). It has also been suggested that this periodicity might exist as the third harmonic of a longer-range 28-residue charge repeat (3, 12, 26, 40), similar to that seen in the coiled-coil segment of myosin (41, 42). Such a periodicity has also been described for the desmoplakin family, strengthening the notion that this periodicity is reflective of intermolecular interactions involving coiled-coils (43).

The Fourier analyses that uncovered the 9.5-residue charge periodicity in IFs rested on the assumption that charged residues would otherwise be randomly distributed throughout the rod (3, 12, 25, 26). Yet, it has been known for some time that heptad repeats of hydrophobic residues place constraints on the a and d positions in the rod sequences. Coupled with our finding that i, $i \pm 4$ ion pairing is prevalent in IF rod segments, it is clear that there is considerable nonrandomness governing the positioning of charged residues. Future analyses should take this point into consideration.

Although simple comparisons of the maximum possible numbers of the two classes of ion pairing in type I–IV IF rods reveal that the potential for intramolecular ionic interactions (42–52 ion pairs) is substantially greater than that for intermolecular associations (9–18 ion pairs), it seems reasonable to expect that both types of interactions might play important roles in IF assembly. To some extent, these interactions might be nonoverlapping and distinct. An additional notion worth considering, however, is the possibility that conformational changes during IF assembly may in part be facilitated through the switching of acidic and basic interactions between helix-stabilizing intrachain salt bridges and interchain ionic associations.

It is particularly interesting that nearly half (49%) of the i, $i \pm 4$ spacings are located either at positions e–b (26%) or at positions c–g (23%), as there is evidence that charged residues at e and g are important in forming dimer-stabilizing intermolecular ion pairs (27) (Fig. 2). This is supportive of the notion that a switch from intramolecular to intermolecular ion pairing may help direct IF assembly. The crystal structure of GCN4 shows evidence for a similar switching, where many of the charged residues that were expected to participate in intrahelical ion pairs were instead involved in crystal contacts (28).

Intrahelical Ion Pairing in Non-IF Coiled-Coil Proteins. We wondered whether the selection for intrahelical ion pairs was specific to IF proteins or characteristic of all coiled-coil proteins. To answer this question, we performed Monte Carlo simulation analysis on triple-stranded and double-stranded coiled-coil proteins. Double-stranded proteins included a family of related IF-associated proteins (IFAPs): plectin, BP230, and desmoplakin; a family of myosin-related proteins: smooth muscle myosin, skeletal myosin, tropomyosin, and paramyosin; as well as GCN-4 of the *fos/jun* family, and APC, the product of the APC gene. The triple-stranded proteins included α-, β-, and γ-fibrinogen; spectrin; MSP; gp17, a protein from the tail fiber of bacteriophage T7; and laminins A, B1, and B2, which are a special case, as they form a dimer intermediate as described below.

The results shown in Table 2 indicate that a propensity for putative ion pairs is not a feature of all coiled-coil proteins, nor is it exclusive to the cytoplasmic IF proteins. It is especially notable that the entire family of IFAPs we studied showed highly significant nonrandom selection for putative ion pairs. This was the only family of proteins other than the cytoplasmic IFs that showed such a consistent selection for i, $i \pm 4$ ion pairing, a feature that could reflect an initial need for stabilizing these IFAPs as monomers and a later role in ion switching to form higher-ordered associations.

Selection for putative ion pairs was found in skeletal muscle myosin and paramyosin but not in smooth muscle myosin or tropomyosin. This may reflect an important functional dichotomy between the two pairs of proteins. Skeletal myosin assembles into highly ordered thick filaments that assemble to form a quasi-crystalline myofibril, and, at least in some invertebrates, paramyosin plays a critical role in thick filament assembly as a backbone protein that is necessary for normal filament elongation (44, 45). In contrast, filaments from smooth muscle myosin assemble into a loosely associated irregular arrangement that results in less well-coordinated contraction, and tropomyosin, though important for thick and

Table 2. Occurrence of ion pairs within coiled-coil segments of 18 non-IF proteins

Protein segment	IP	Rod length, aa	B/A	$P_{ex\text{-}hep}$
Plectin 1	74	617	111/140	<0.01
Plectin 2	29	325	57/72	0.05
Total	103	942	168/212	<0.01
BPA230-1	22	182	45/40	<0.01
BPA230-2	41	515	94/92	<0.01
Total	63	697	139/132	<0.01
Desmoplakin	87	889	185/182	<0.01
Mus. myosin				
Smooth	95	1039	205/250	0.20
Skeletal	105	1077	194/249	<0.01
α-Tropomyosin	23	284	50/79	0.74
Paramyosin	65	721	125/148	<0.01
GCN4	3	33	8/7	0.51
APC	8	298	30/40	0.54
α-Fibrinogen*	8	111	22/16	0.12
β-Fibrinogen*	5	112	13/17	0.23
γ-Fibrinogen*	5	111	14/17	0.30
Spectrin*	2–9	99–101	9–22/14–25	0.06–0.91
Laminin A*	24	563	83/78	0.35
Laminin B1†	39	575	75/110	<0.01
Laminin B2†	43	562	84/117	<0.01
gp17*	6	116	15/17	0.18
MSP*	6	163	18/27	0.47

Species information is in the text. Abbreviations as in Table 1. Certain proteins had divided coiled-coils that were considered separately—namely, plectin, BPA230, and spectrin. For spectrin, the range of probabilities for 17 coiled-coil segments studied is given, only one of which was <0.15; most were >0.50, suggesting little nonrandom selection for ion pairs. In the case of APC, 8 coiled-coil segments, the longest 56 residues in length, were combined to form one artificial segment for the analysis here. When each smaller segment was considered separately, no nonrandom selection for ion pairs was found (not shown). Proteins that are unmarked (no footnote symbol) form dimeric coiled-coils. The non-IF protein sequences chosen include sequences for (*a*) rat plectin, (*b*) human 230-kDa bullous pemphigoid antigen (BPA230), (*c*) human desmoplakin, (*d*) chicken smooth muscle myosin (Mus. myosin, smooth), (*e*) rat skeletal muscle myosin (Mus. myosin, skeletal), (*f*) rat α-tropomyosin, (*g*) *Drosophila* paramyosin, (*h*) yeast GCN4, (*i*) human APC gene product, (*j*) human α-, β-, and γ-fibrinogens, (*k*) human erythroid α-spectrin, (*l*) mouse laminins A, B1, and B2, (*m*) bacteriophage protein gp17, and (*n*) bovine type I MSP.

*Proteins that form triple-stranded coiled-coils.

†Laminins B1 and B2, which form a heterodimeric intermediate during trimer assembly.

thin filament functional interaction, is not a critical building block of the thick filament. Thus, as in cytoplasmic IFs, the selection for putative ion pairs in skeletal myosin and paramyosin may be related to their role as stable building blocks of higher-ordered structures. The absence of selection for ion pairs in dimeric coiled-coils such as APC and GCN4, which are relatively short and do not form filaments, is consistent with the notion that intrahelical ion pairing plays an important role in stabilizing subunits so that a well-ordered final structure may be obtained.

Of the proteins that form triple-stranded coiled-coils, none shows significantly nonrandom selection for putative ion pairs except for laminins B1 and B2. Intriguingly, laminins B1 and B2 form a stable heterodimer intermediate in laminin heterotrimer assembly, whereas laminin A does not seem to dimerize with itself or other laminins (46). When taken together with our analyses of other trimeric coiled-coil proteins, these results suggest that proteins that form triple-stranded coiled-coils have very little tendency to show nonrandom selection for putative ion pairs. This finding may be useful in predicting whether a protein bearing features of a coiled-coil will assemble into a two- or three-stranded rope. Based on the present data, such a sequence containing a preponderance of putative intrahelical ion pairs would be predicted to assemble into a double-stranded coiled-coil.

CONCLUSION

In conclusion, our present study has uncovered a hitherto unappreciated phenomenon—namely, that in the type I–IV IF α-helices, there is a strong propensity for nonrandom positioning of oppositely charged residues spaced by exactly four residues. The relative positioning of B ... A ion pairing coupled with the conservation of several ion pairs in key positions underscore the potential importance of intramolecular ion pairing in promoting α-helicity in the IF rod. This may be critical in optimizing the packing of polypeptides within an IF. Given that the majority of charged residues in the rods of many of these IF proteins have $i, i \pm 4$ spacing (this study), and yet many of these residues are also in proximity to be involved in interhelical ion pairing (3), it seems likely that switching from intrahelical to interhelical ion pairing serves as a driving force behind the early stages of filament assembly.

Furthermore, we have shown that a strong selection for putative ion pairs exists also in a family of coiled-coil IFAPs. Such selection in skeletal myosin and paramyosin, but not in smooth muscle myosin or tropomyosin, suggests that selection depends not only on a coiled-coil structure but also on the protein's functional role. We suggest that there is a general correlation between the extent to which the coiled-coil protein participates in highly ordered multimolecular aggregates, such as IFs and thick filaments, and the degree to which there is selection for intrahelical ion pairs. Lastly, the absence of such selection in solely trimeric coiled-coil proteins studied suggests a simple way for discriminating between assembly into two- or three-stranded coiled-coils based on amino acid sequence alone.

We thank Holgar Kley for providing us with a computer program of his own design for the purposes of performing our Monte Carlo sequence simulations. This work was supported by a grant from the National Institutes of Health and from the Howard Hughes Medical Institute. E.F. is an Investigator of the Howard Hughes Medical Institute. A.L. is funded by a Medical Scientist Training Program grant from the National Institutes of Health.

1. Geisler, N. & Weber, K. (1982) *EMBO J.* **1,** 1649–1656.
2. Hanukoglu, I. & Fuchs, E. (1982) *Cell* **31,** 243–252.
3. Conway, J. F. & Parry, D. A. D. (1988) *Int. J. Biol. Macromol.* **10,** 79–98.
4. Stewart, M. (1993) *Curr. Opin. Cell Biol.* **5,** 3–11.
5. Fuchs, E. & Weber, K. (1994) *Annu. Rev. Biochem.* **63,** 345–482.
6. Albers, K. & Fuchs, E. (1987) *J. Cell Biol.* **105,** 791–806.
7. Coulombe, P. A., Chan, Y.-M., Albers, K. & Fuchs, E. (1990) *J. Cell Biol.* **111,** 3049–3064.
8. Wong, P. C. & Cleveland, D. W. (1990) *J. Cell Biol.* **111,** 1987–2003.
9. Hatzfeld, M. & Weber, K. (1990) *J. Cell Biol.* **110,** 1199–1210.
10. Letai, A., Coulombe, P. A. & Fuchs, E. (1992) *J. Cell Biol.* **116,** 1181–1195.
11. McLachlan, A. D. (1978) *J. Mol. Biol.* **124,** 297–304.
12. Parry, D. A. D., Crewther, W. G., Fraser, R. D. B. & Macrae, T. P. (1977) *J. Mol. Biol.* **113,** 449–454.
13. Woods, E. F. & Inglis, A. S. (1984) *Int. J. Biol. Macromol.* **6,** 277–283.
14. Coulombe, P. A. & Fuchs, E. (1990) *J. Cell Biol.* **111,** 153–169.
15. Quinlan, R. A., Cohlberg, J. A., Schiller, D. L., Hatzfeld, M. & Franke, W. W. (1984) *J. Mol. Biol.* **178,** 365–388.
16. Aebi, U., Cohn, J., Buhle, L. & Gerace, L. (1986) *Nature (London)* **233,** 560–564.
17. Potschka, M., Nave, R., Weber, K. & Geisler, N. (1990) *Eur. J. Biochem.* **190,** 503–508.
18. Stewart, M., Quinlan, R. A. & Moir, R. D. (1989) *J. Cell Biol.* **109,** 225–234.
19. Steinert, P. M. (1991) *J. Struct. Biol.* **107,** 157–174.
20. Geisler, N., Kauffmann, E. & Weber, K. (1985) *J. Mol. Biol.* **182,** 173–177.
21. Geisler, N., Schuenemann, J. & Weber, K. (1992) *Eur. J. Biochem.* **206,** 841–852.
22. Steinert, P. M. & Parry, D. A. D. (1993) *J. Biol. Chem.* **268,** 2878–2887.
23. Steinert, P. M., Marekov, L. N., Fraser, R. D. B. & Parry, D. A. D. (1993) *J. Mol. Biol.* **230,** 436–452.
24. Heins, S., Wong, P. C., Muller, S., Goldie, K., Cleveland, D. W. & Aebi, U. (1993) *J. Cell Biol.* **123,** 1517–1534.
25. Conway, J. F. & Parry, D. A. D. (1990) *Int. J. Biol. Macromol.* **12,** 328–334.
26. Fraser, R. D. B., Macrae, T. P., Suzuki, E. & Parry, D. A. D. (1985) *Int. J. Biol. Macromol.* **7,** 257–274.
27. Parry, D. A. D., Steven, A. C. & Steinert, P. M. (1985) *Biochem. Biophys. Res. Commun.* **127,** 1012–1018.
28. O'Shea, E. K., Klemm, J. D., Kim, P. S. & Alber, T. (1991) *Science* **254,** 539–544.
29. O'Shea, E. K., Lumb, K. J. & Kim, P. S. (1993) *Curr. Biol.* **3,** 658–667.
30. Maxfield, F. R. & Scheraga, H. A. (1975) *Macromolecules* **8,** 491–493.
31. Sundaralingam, M., Sekharudu, Y. C., Yathindra, N. & Ravichandran, V. (1987) *Int. J. Quant. Chem. Quant. Biol. Symp.* **14,** 289–296.
32. Marqusee, S. & Baldwin, R. L. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 8898–8902.
33. Lyu, P. C., Gans, P. J. & Kallenbach, N. R. (1992) *J. Mol. Biol.* **223,** 343–350.
34. Serrano, L., Horvitz, A., Avron, B., Bycroft, M. & Fersht, A. R. (1990) *Biochemistry* **29,** 9343–9352.
35. Sail, D., Bycroft, M. & Fersht, A. R. (1991) *J. Mol. Biol.* **220,** 779–788.
36. Stellwagen, E., Park, S.-H., Shalongo, W. & Jain, A. (1992) *Biochemistry* **32,** 1193–1200.
37. Huyghues-Despointes, B. M. P., Scholtz, J. M. & Baldwin, R. L. (1993) *Protein Sci.* **2,** 80–85.
38. Crick, F. H. C. (1953) *Acta Crystallogr.* **6,** 689–697.
39. Pauling, L. & Corey, R. B. (1953) *Nature (London)* **171,** 59–61.
40. Parry, D. A. D., Conway, J. F. & Steinert, P. M. (1986) *Biochem. J.* **238,** 305–308.
41. Parry, D. A. D. (1981) *J. Mol. Biol.* **153,** 459–464.
42. McLachlan, A. D. & Karn, J. (1982) *Nature (London)* **299,** 226–231.
43. Green, K. J., Virata, M. L. A., Elgart, G. W., Stanley, J. R. & Parry, D. A. D. (1992) *Int. J. Biol. Macromol.* **14,** 145–153.
44. Epstein, H. F., Berliner, G. C., Casey, D. L. & Ortz, I. (1988) *J. Cell Biol.* **106,** 1985–1995.
45. MacKenzie, J. M. & Epstein, H. F. (1980) *Cell* **22,** 747–755.
46. Engel, J. (1993) in *Molecular and Cellular Aspects of Basement Membranes*, eds. Rohrbach, D. V. & Timpl, R. (Academic, San Diego), pp. 147–176.