STRUCTURAL STUDIES ON THE NUCLEAR LAMINS AND OTHER INTERMEDIATE FILAMENT PROTEINS

A thesis presented in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy in Biophysics at
Massey University.

James Frederick CONWAY

October, 1989

Massey University Library. Thesis Copyright Form

Title of thesis: Structural Studies on the Nuclear Lamins and other Intermediate Filament Proteins

(1) (a) I give permission for my thesis to be made available to readers in the Massey University Library under conditions determined by the Librarian.

~~(b)~~ I do not wish my thesis to be made available to readers without my written consent for _____ months.

(2) (a) I agree that my thesis, or a copy, may be sent to another institution under conditions determined by the Librarian.

~~(b)~~ I do not wish my thesis, or a copy, to be sent to another institution without my written consent for _____ months.

(3) (a) I agree that my thesis may be copied for Library use.

~~(b)~~ I do not wish my thesis to be copied for Library use for _____ months.

Signed J F Conway

Date 10/10/89

The copyright of this thesis belongs to the author. Readers must sign their name in the space below to show that they recognise this. They are asked to add their permanent address.

NAME AND ADDRESS	DATE

ABSTRACT

A number of aspects of IF chain and molecular structure, as well as molecular aggregation, have been examined. These include the delineation of periodicities in the sequences of structural domains of IF proteins, the distribution of amino acid residues within the heptad substructure, the flexibility of the peptide backbone, the extent of homology among the IF proteins, the packing of chains in the dimeric molecule, and the axial packing of molecules in the IF. Particular focus has been placed on the newly sequenced type V IF proteins (the nuclear lamins and the *Helix pomotia* B protein) and on a type III IF protein (peripherin).

A parallel in-register arrangement of chains in the molecule is predicted for peripherin and the type V chains from a consideration of interchain ionic interactions. Also, periodicities in the linear distribution of charged residues in the rod domains of these proteins are shown to be comparable with periods in other IF chains. Ionic interactions between lamin molecules have been used to assess the likely modes of molecular aggregation in an *in vitro* assembly and a model is presented which also satisfies the constraints imposed by electron microscope data. In this model, antiparallel arrays of molecules are half-staggered and an extended conformation for the carboxy-terminal domains is predicted. Simple explanations are given for the transition between paracrystalline and lattice structures and for the disassembly of the lamin meshwork concommitant with hyperphosphorylation. The method of calculating intermolecular ionic interaction profiles is enhanced and a new, three-dimensional method is developed.

The inhomogeneous distribution of residues in the heptad substructure can be correlated to the coiled-coil structure and chain packing in the molecule. In particular, the ~75% occupancy rate of apolar residues in the internal a and d heptad positions is shown to be a general feature of α-fibrous proteins. Variability of residues in the outer b, c and f positions indicates that structural or functional specificity in the rod domain may be determined by these parts of the sequence. The predicted flexibilities of IF chains have been compared to the underlying structure for the chains.

Evidence from sequence homology studies suggests that several new subtypes are appropriate in the classification scheme. For the hard keratins the terms types Ia and IIa are proposed and for the soft keratins, types Ib and IIb; the need to separate the neurofilaments into the type IV class separate from the type III IF chains is confirmed; and division of the type V chains into cytoskeletal and karyoskeletal groups is indicated. A more detailed delineation is made of regions within the amino- and

carboxy-terminal domains than has been possible previously. Periodic features of the homology profiles for the rod domain are examined and found to be similar to those in the linear distribution of residues in the amino acid sequences. Comparison between amphibian and mammalian keratins, and also between hard and soft keratins reveals that type II chains are maintained at a higher level of fidelity than type I chains. Consensus rod domain sequences are derived for the various IF subtypes: absolutely conserved regions of primary structure identify types or subtypes.

ACKNOWLEDGEMENTS

It is with great pleasure that I acknowledge my chief supervisor, Prof. D.A.D. Parry, for his assistance and commitment to this work. His indefatigable energy and high standards have contributed largely to the work described herein and I look forward to working with him in the future. Prof. P.T Callaghan and the staff of the Physics and Biophysics Department at Massey University have been very supportive during the course of my studies and I appreciate their efforts on my behalf. The other doctoral and graduate students have provided a friendly atmosphere within which to work and I wish them every success in their endeavours. Mr P.M. Ngan, Director of the Massey University Image Analysis Unit, has been of great assistance in providing access to the Unit's facilities and in many interesting discussions on a variety of topics.

I much appreciate the kind hospitality provided by Drs. P.M. Steinert and A.C. Steven during the course of a visit to the National Institutes of Health in 1986 and also for the extra impetus they have provided in the completion of this work. I look forward to continued collaboration with them. In addition, I acknowledge the valuable contributions towards the cost of that trip from the Royal Society of New Zealand, the Dean of the Science Faculty at Massey, Prof. G.N. Malcolm, and the Fibrous Protein Merit Award.

Thanks also to my parents for their encouragement and assistance during the earlier period of my studies. Finally, special thanks to Sharon Wards for her support and encouragement during the course of my PhD studies. I am indebted to her for all her time spent directly and indirectly in the preparation of this manuscript.

Some amino acid sequences were kindly provided prior to publication by:

Dr RA. Lazzarini (human NF-M chain)
Dr SSM. Chin (rat NF-M chain)
Professor GE. Rogers and Dr BC. Powell (a type II keratin chain from sheep)
Dr PM. Steinert and Dr DR. Roop (mouse M50k and M55k chains)

Electron micrographs were kindly provided by:

Dr PM. Steinert (Figure 1-1 of reconstituted IF)
Professor GE. Rogers (Figure 1-2 of wool filaments in cross-section)
Dr AE. Goldman and Dr RD Goldman (Figure 4-2 of lamin paracrystals)
Dr C. Cohen (Figure 4-7 of a tropomyosin lattice/paracrystal structure)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. AN OVERVIEW OF INTERMEDIATE FILAMENT STRUCTURE

Intermediate filaments (IF) are long, unbranched structures found in a diverse range of cell types (Figure 1-1). They form one of three major classes of cytoskeletal networks and are so-called because their diameters (~10 nm) lie intermediate in size between the smaller actin-containing microfilaments and the larger diameter microtubules and myosin-containing thick filaments. The recent addition of the nuclear lamin proteins into the IF family has shown that IF networks are found not only throughout the cytoplasm but also within the cell nucleus where they are thought to provide at least part of the structural integrity of the nuclear envelope.

The IF proteins were initially classified according to the cell types with which they were associated. The first of these classes comprised the relatively large group of keratins from epithelial cells; the second class contained the desmin protein found in myogenic cells; glial fibrillary acidic protein (GFAP) in astroglial cells, and vimentin in cells of mesenchymal origin, made up the third and fourth classes respectively.

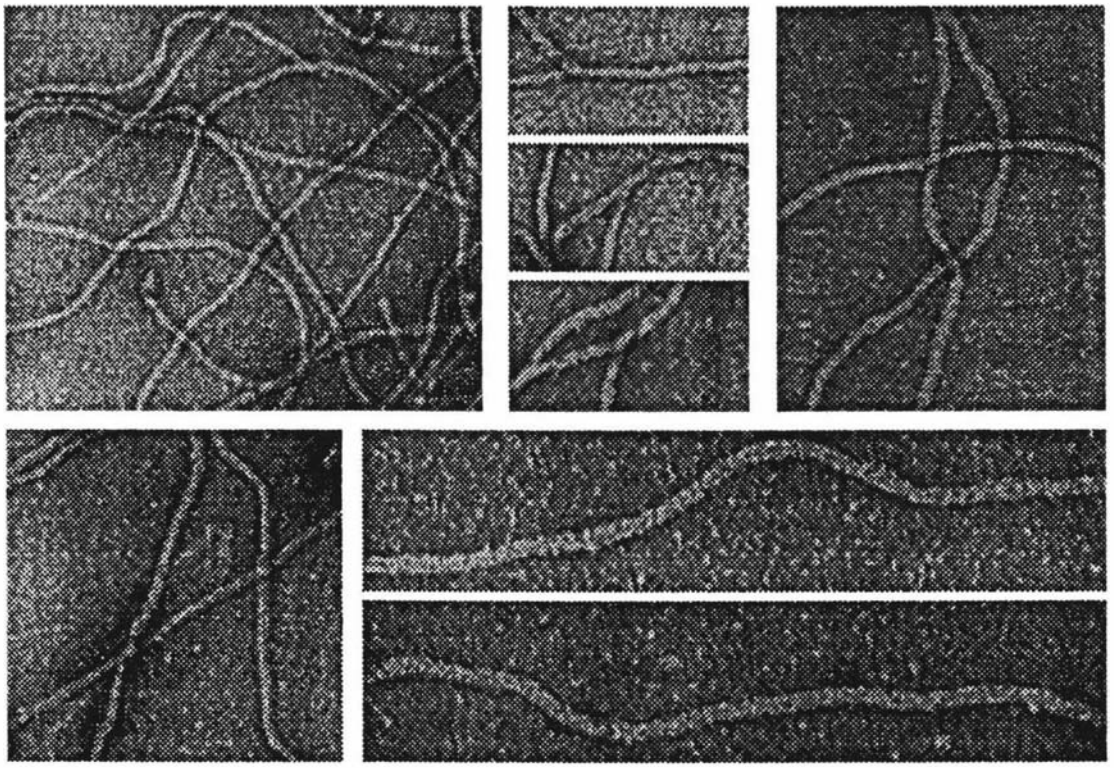


Figure 1-1 Examples of the 10 nm-diameter intermediate filament (IF) structures *in vitro* (separate images are not to scale). Branching filaments are evident as are filaments crossing over each other. Figure courtesy of Dr P.M. Steinert.

(Vimentin, however, was also found in cells that expressed other IF). The fifth class encompassed the neurofilaments from neuronal cells and was further divisible into three sub-classes according to the relative molecular weights of the chains.

As amino acid sequence data became available for the various IF chains, an alternative classification system based on sequence homology present in the major α-helix-containing fraction of the IF proteins largely replaced the cell-specific nomenclature. The early sequence studies of Crewther and co-workers on wool keratins revealed two kinds of α-helix-rich segment: type I had a net acidic character and type II was neutral-basic. Subsequent keratin sequences, primarily from the epidermal keratins, were found to fall into these same two groups and consequently the chains from which these characteristic sequences were derived were termed type I and type II IF chains. The equivalent α-helical regions in desmin, vimentin and GFAP showed greater sequence homology with one another than with the keratins and were classified together as type III IF chains. Although the neurofilament proteins were initially included within the type III grouping, increasing evidence has indicated that they are members of a distinct group and they are now referred to as type IV IF chains. Several recent additions to the IF family of proteins have been made: peripherin, a type III IF protein; the nuclear lamin proteins which are now established as a new IF type-protein - type V; and the *Helix pomotia* B protein which shows structural homology with the nuclear lamins.

Since the presence of cell-specific IF indicates a diversity of function, it is pertinent to ascertain whether this translates into a number of variants on the basic model for IF structure. The low-resolution substructures of some IF have been partially defined in the last five years or so and this will form the basis of the work described here. This thesis, therefore, represents a collection of studies on the structure and aggregation of IF molecules with special reference to the more recently characterized members of the IF family – the nuclear lamin proteins in particular but also peripherin and the *Helix pomotia* B protein.

1.1 Early Structural Work on Keratin IF

The structure of wool keratin has been a focus for research over the past 50 years due largely to the commercial significance of the wool fibre. The keratin IF (originally termed microfibrils) in wool, hair and quill are about 10 nm in diameter and are embedded in a matrix of proteins of high-sulphur and of high-glycine-tyrosine content (Figure 1-2). Keratins from these sources are termed 'hard' to distinguish them from the 'soft' (or epidermal) keratins found in the stratum corneum of skin. IF in hard keratin are preferentially oriented parallel to the axis of the fibre and have proven to be ideal subjects for X-ray diffraction studies. In contrast the orientation of IF in the soft

keratins is much poorer and consequently few X-ray diffraction data have been recorded. Pioneering work in this area was carried out by Astbury and co-workers who studied hard keratinized tissues from a variety of sources (Astbury and Street, 1931; Astbury and Marwick, 1932; Astbury and Woods, 1933). They noted a number of distinctive X-ray diffraction patterns: α (from wool keratin), β (from stretched mammalian hard keratin), the feather-pattern (from hard avian and reptilian tissue), and the amorphous pattern (from the cuticle of animal hair). The terms 'α-keratin' and 'α-fibrous protein' are derived from this nomenclature and refer to proteins that give rise to the α-pattern. Astbury and Woods (1933) proposed that the α-pattern was generated by a folded structure and that the β-form of keratin arose from stretching the molecular chains into an almost fully extended conformation. Current descriptions of the β-structure are based on the pleated sheet models of Pauling and Corey (1951b, 1953b) and will not be discussed further here (see, for example, Fraser *et al*, 1972 and Fraser and MacRae, 1973b).

A number of polypeptide chain structures were proposed to explain the α-pattern (see for example, Astbury and Bell, 1941; Huggins, 1943; Astbury *et al*, 1948; Ambrose *et al*, 1949; Bragg et al, 1950; Pauling and Corey, 1950). The model of Pauling and Corey (elaborated in Pauling and Corey, 1951a and Pauling *et al*, 1951) placed the residues in a helical arrangement, and two variants were described: one had 3.7 residues per turn of helix and the other had 5.1 residues per turn. These were termed the α-helix and the γ-helix respectively (Pauling and Corey, 1951a) and featured intrachain hydrogen bonds and planar amide groups. Perutz (1951a,b) eliminated from consideration all models but the Pauling and Corey α-helix by observation of a meridional reflection of spacing 0.149 nm in the X-ray patterns of poly-γ-benzyl-L-glutamate, keratin, and hæmoglobin. Only the α-helix was expected to give rise to this spacing which corresponds to the axial rise per residue. (The 0.149 nm meridional

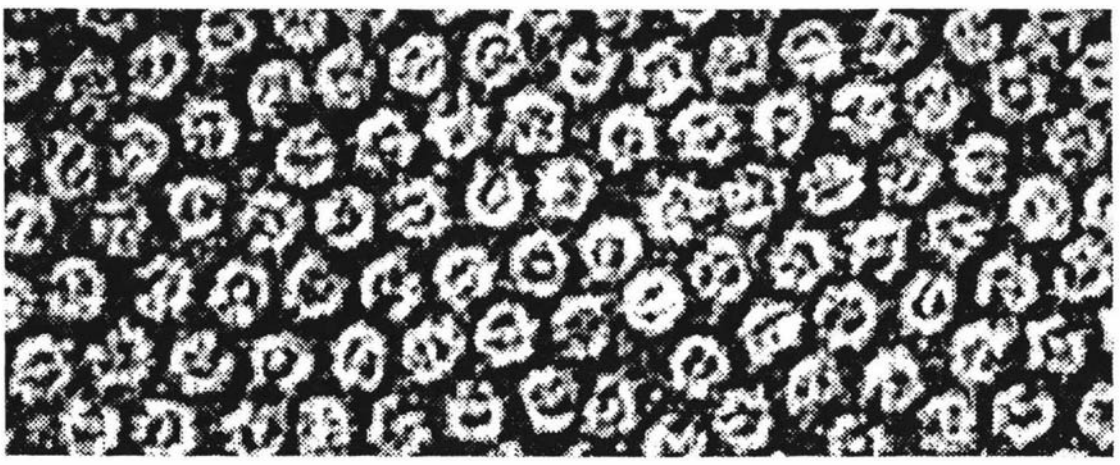


Figure 1-2 Cross-section of intermediate filaments in wool. Figure courtesy of Professor G.E. Rogers.

reflection was originally noted by MacArthur, 1943, in the X-ray diffraction pattern from African porcupine quill tip but was interpreted in terms of repeating sidechains along the polypeptide chain. By using the synthetic polypeptide, poly-γ-benzyl-L-glutamate, Perutz demonstrated that the 0.149 nm reflection was independent of the specific natures of the sidechains). Cochran *et al* (1952) calculated the Fourier transform of an α-helix and quantitative agreement for this conformation was found from X-ray diffraction studies on the synthetic polypeptides poly-γ-methyl-L-glutamate (Bamford *et al*, 1952) and poly-L-alanine (Brown and Trotter, 1956).

The screw sense of the α-helix *in vivo* was not determined for a number of years. Pauling *et al* (1951) observed that for residues with the L-configuration (which is predominantly the case *in vivo*) the position of the sidechains differed between left- and right-handed α-helices (except for glycine which has a sidechain consisting of a single hydrogen atom). Huggins (1952) showed that for L-amino acids, significant steric hindrance was probable between the β-carbon and the carbonyl-oxygen in the same residue for a left-handed α-helix whereas no such hindrance would occur for a right-handed one. Subsequent studies on α-helix-forming synthetic polypeptides (for example poly-L-alanine: Elliott and Malcolm, 1956, 1959) showed that the α-helices were right-handed, as is the general case, although several left-handed α-helices of marginal stability have subsequently been reported (see Fraser *et al*, 1972). The first direct confirmation of the presence of α-helices in proteins was obtained from the 0.2 nm Fourier synthesis of myoglobin (Kendrew *et al*, 1960). In addition, the screw sense of all the α-helical segments in the molecule was shown to be right-handed.

A prominent feature on the meridian of the X-ray diffraction patterns from α-keratin and other α-fibrous proteins (for example α-tropomyosin) is a strong reflection at a spacing of 0.515 nm (Astbury and Street, 1931). This is not predicted by an undistorted α-helical structure which should instead show off-meridional layer line diffraction at an axial spacing of 0.54 nm (Crick, 1952). An explanation of the strong 0.515 nm reflection and the missing layer line was made independently by Crick (1952) and Pauling and Corey (1953a): in their models the α-helices were distorted so as to wrap around one another in a supercoil of opposite sense to that of the α-helix and this structure was termed a 'coiled-coil rope'. The 0.515 nm reflection was shown to arise from the axial rise per turn of the supercoil. Crick (1953) described a seven residue substructure, later termed the heptapeptide or heptad repeat, in which sidechains from two of every seven residues could be made to fit neatly into the spaces between the sidechains on the other α-helix or α-helices. This arrangement of side-chains along the line of contact was termed 'knob-hole' packing and the nature of these internalized sidechains was suggested to be hydrophobic. Confirmation of this

suggestion was provided many years later by Hodges *et al* (1972), Parry (1974) and Stone *et al* (1974) who demonstrated such a substructure in the amino acid sequence of tropomyosin - apolar residues were indeed successively three and four residues apart. McLachlan and Stewart (1975) introduced a nomenclature for the positions of the residues in the heptad, a, b, c, d, e, f and g, where a and d are the internal positions of the coiled-coil and are filled predominantly by apolar residues. In IF proteins, about 75% of residues in the a and d positions are apolar (Parry and Fraser, 1985).

Hard α-keratin, although ideal for X-ray diffraction studies, is not readily amenable to chemical analysis. The high content of covalent disulphide bonds present in wool, for example, effectively welds the proteins into a mechanically inert structure (Fraser *et al*, 1972). Even when the disulphide bonds are reduced, a complex mixture of related proteins is revealed (Crewther *et al*, 1965). The problems associated with sequencing the keratin proteins in wool (using the classic protein sequencing techniques) were summarized by Hogg *et al* (1978) as follows: the protein chains were relatively large at 40-50 kDa; the number of similar chains within the relatively small range of molecular weights was also large; and identification of particular chains was difficult. Nonetheless, Crewther and Dowling (1971) and Dowling and Crewther (1972) have shown by amino acid analysis, peptide mapping and optical rotatory dispersion measurements that two types of chain segment, termed I and II, were present in subfractions of helical fragments of S-carboxymethylkerateine-A. Crewther and co-workers (Crewther, 1976; Crewther *et al*, 1978a; Elleman *et al*, 1978; Gough *et al*, 1978; Hogg *et al*, 1978) sequenced a type I and a type II segment of lengths 103 and 109 residues respectively and showed that a heptad substructure existed with apolar residues common in the a and d positions. Consequently, in a helix of 3.6 residues per turn these apolar residues generate a stripe running around the outside of the undistorted α-helix - a result in accord with the coiled-coil model of Crick (1953) and the tropomyosin model (see above). They also noted that 32% of the residues were identical between the type I and the type II segments when the sequences were aligned to maximize homology. Crewther *et al* (1978b) extended these results by comparing the sequences of five fragments (~30 residues each) from different wool keratin proteins: of those, three were type I segments and two were type II segments. As before, the identity between the groups was about 30% but within each type there were few differences. More recently, Crewther *et al* (1985) have fully sequenced the rod domains of two type I wool keratin proteins (components 8a and 8c-1) and two type II proteins (components 7 and 7c) and have reported that the identity within the type I and II groupings is 90-95%.

Long standing problems concerning the structure of the coiled-coil in α-keratin proteins have been to determine the pitch length of the coiled-coil and the number and relative orientation of polypeptide chains that comprise the rope. Crick (1953) described coiled-coil ropes with two and three α-helical strands and suggested that the three-stranded rope would be an appropriate model for α-keratin. Pauling and Corey (1953a) originally suggested that α-keratin comprised 'α-cables' of six α-helices coiled around a seventh with single α-helices in the interstices between the cables. Evidence for a two-strand rope was found by Cohen and Holmes (1963) in the X-ray diffraction pattern of a highly oriented specimen of paramyosin and they concluded that the pitch length of the left-handed coiled-coil was 17.8±1 nm. The degree of orientation in the X-ray diffraction pattern from α-keratin was insufficient to allow a clear choice to be made between two and three strands (Fraser *et al*, 1964a, 1964b, 1965) and in addition the X-ray diffraction pattern was overlaid with an elaborate interference pattern associated with higher levels of structure (Fraser *et al*, 1972). Estimates of the pitch length for α-keratin ranged from 14 to 17 nm and 21 to 26 nm for two- and three-stranded ropes respectively (Fraser *et al*, 1965) but it was suggested that the degree of similarity between the diffraction patterns from tropomyosin, myosin, paramyosin (all of which were known to contain two-stranded ropes from physico-chemical data) and α-keratin would tend to favour a two-strand rope model for α-keratin.

In principle, two-stranded coiled-coils can adopt one of two forms: in the first both α-helices are similarly directed, or parallel (ie, the polypeptide mainchain sequences, -NH-αC-C'-, were oriented in the same direction), whereas in the other the chains are oppositely directed, or antiparallel. Parry and Suzuki (1969) calculated the free energies of the parallel and antiparallel forms for a two-stranded model of poly-L-alanine and compared them with the energies of the analogous pairs of straight α-helices. The coiled-coils were shown to be significantly more stable than the straight α-helix analogues and the antiparallel coiled-coil was favoured over the parallel. A feature not considered at that time, however, was the role played by charged sidechains in specifying the orientation of the chains in the coiled-coil (Parry, 1974, 1975; McLachlan and Stewart, 1975, 1976). McLachlan and Stewart (1975) and Parry (1975) showed that the distribution of charged residues in α-tropomyosin resulted in a largely acidic stripe in the e position of the heptad and a largely basic stripe in the g position where both stripes were adjacent to, but on opposite sides of, the 'internal' apolar stripe. Parallel chains allowed salt bridges (or ionic interactions) to be made between a pair of helices but antiparallel chains placed similarly charged residues in close proximity and caused destabilization. Interchain ionic interactions were studied for the type I and type II chain segments of Crewther *et al* (described above) in a two-

stranded rope (Parry *et al*, 1977) and a three-stranded rope (Parry, 1979) and all combinations of chains, polarity (ie, parallel or antiparallel) and relative stagger were examined. While the number of chains in the molecule could not be determined, a parallel, in-register arrangement of a type I chain and a type II chain was considered among the more likely candidates. A related study by McLachlan (1978) suggested that the coiled-coil in α-keratin was more likely to be two-stranded than three-stranded.

The chemical studies on wool keratin originally appeared to favour a three chain subunit (Crewther and Harrap, 1967; Crewther *et al*, 1968, Crewther and Dowling, 1971; Skerrow *et al*, 1973; Lee and Baden, 1976; Lotay and Speakman, 1977; Steinert, 1978; Steinert *et al*, 1980a) although in some cases a two chain structure could not be excluded. Subsequent studies by Crewther *et al* (1980) showed that type I proteins interact specifically with type II proteins in a 1:1 molar ratio although there was no evidence to show whether single or multiple chains of each species were involved in the molecular structure. However, crosslinking studies by Ahmadi and Speakman (1978) and Ahmadi *et al* (1980) provided strong evidence for a four-chain subunit. Further studies by Woods and Gruen (1981) and Gruen and Woods (1983) showed that the wool microfibril contained equal numbers of type I and II chains and that the fundamental sub-unit was a tetramer of chains configured as a pair of coiled-coils. More detailed studies by Inglis *et al* (1983) and Woods and Inglis (1984) confirmed that the dimer of chains in the coiled-coil molecule from wool keratin was indeed comprised of a type I chain and a type II chain - a heterodimeric structure - and that the chains were parallel. Their data also indicated that the molecules were antiparallel and approximately half-staggered so that the N-terminal regions of the rod domain were overlapped.

1.2 Soft Keratin and Other IF

Until relatively recently, the so-called 'soft' or epidermal keratins from the *stratum corneum* of skin had not been studied as extensively as the hard keratins from wool or quill. Certainly the filaments found in the *stratum corneum* were known to be similar in diameter to those from hard keratin-containing tissue (~10 nm). However, the degree of alignment was so much poorer than in the hard keratins that the use of physical techniques such as X-ray diffraction and electron microscopy had been very limited although specimens prepared from mammalian epidermal extracts had yielded an α-pattern of poor quality. A comparison of the amino acid content of hard and soft keratins had shown much similarity. Major chemical differences did exist, however, between the hard and soft keratins, mainly in the relative contents of the cystine and glycine residues (see Fraser *et al*, 1972).

The hard keratins are very insoluble and are embedded in a matrix of high-sulphur proteins and high glycine-tyrosine proteins. A high degree of disulphide crosslinks between hard-keratin chains gives rise to this insolubility as well as the 'hardness' of the filament. It also gives rise to the degree of orientation so necessary for useful X-ray diffraction study. The soft keratins, however, have few disulphide cross-links and are much more amenable to chemical investigation.

Chemical studies on all members of the IF family revealed that the proteins were more closely related to one another than had been thought initially (for example, Steinert *et al*, 1978, 1980a,b). Studies on the proteolytic digests of epidermal keratins (Woods, 1983; Parry *et al*, 1985) produced four-chain and two-chain particles analogous to those produced from hard keratins (Woods and Inglis, 1984; and see above) and it was concluded that epidermal keratin IF was also a type I-type II heterodimer composed of parallel, two-stranded coiled-coils. Quinlan and Franke (1982, 1983) induced crosslinks in type III IF chains using 1,10-phenanthroline cupric ion complexes and subsequently isolated crosslinked homo- and hetero-dimers: they suggested that the crosslink between the single cysteine residues in these sequences was inter-molecular. Parry *et al* (1985) showed that the crosslink could instead be made intra-molecularly between parallel, in-register chains and used this as further evidence for a common molecular structure for all IF proteins. Sequence studies of the wool keratins, the epidermal keratins, and the other IF have also shown that these proteins, originating from a wide variety of cell types, exhibit a high degree of homology in the central, largely α-helical (rod domain) portion of the chains (Geisler and Weber, 1982; Weber and Geisler, 1982; Crewther *et al*, 1983; Dowling *et al*, 1983).

Little was known about the terminal domains of IF proteins before amino acid sequence data became available. Optical rotatory dispersion experiments had indicated that the keratin molecule was not all α-helical: Harrap (1963), for example, showed that the helical content was only 50-60% (see also Crewther *et al*, 1966, 1968). Early sequencing studies on wool keratin by O'Donnell and co-workers (for example, O'Donnell, 1969) had revealed fragments with no structure that could be related to the regular α-pattern deduced from the X-ray diffraction data. Later studies by Crewther *et al* (1983) on more complete hard α-keratin sequences showed that the N- and C-terminal portions of the chains were not likely to be α-helical in conformation and in fact β-bends were largely predicted.

It is now clear that all IF chains share a common structural plan (a central coiled-coil domain bracketed by terminal domains of non-repetitive secondary structure - see

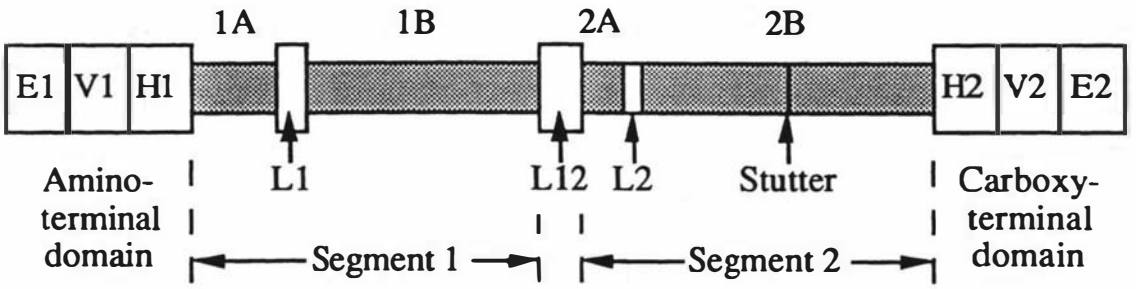


Figure 1-3 Schematic representation of the IF protein chain. The central rod region is largely of coiled-coil structure (shaded) with breaks at the linker regions, L1 and L12, neither of which are predicted to be coiled-coil or α-helical in conformation, and at link L2 which is predicted to be α-helical but not a coiled-coil. The lengths of segments 1 and 2 are each about 20.5 nm and the combined length of the rod is predicted to be about 47 nm. The terminal domains vary greatly in size among the IF proteins and are largely non-α-helical.

Figure 1–3: Geisler *et al*, 1982a; Crewther *et al*, 1983). However, a large range of molecular weights is apparent in the IF group. This arises almost entirely from variations in the sizes of the terminal domains: extreme examples are the cytokeratin 19 protein (mol. wt. 44 kDa: Bader *et al*, 1986; Eckert, 1988) which is almost devoid of a C-terminal domain (only nine residues in length from the generally accepted boundary between the rod and C-terminal domains) and the neurofilament NF-H protein (mol. wt. 140 kDa: for example the mouse NF-H chain, Shneidman *et al*, 1988) which has a C-terminal domain of length 658 residues.

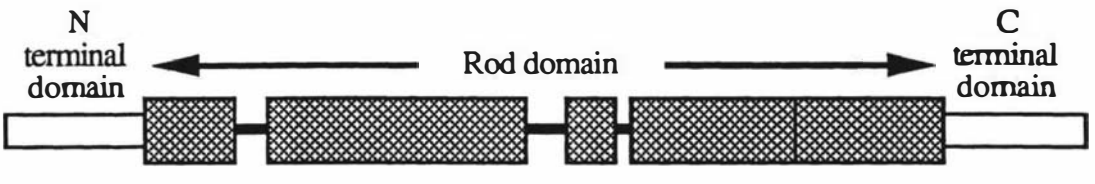
The structural role of the terminal domains is not altogether clear although the amino-terminal domain does appear to be involved in the formation of at least some IF. Certainly the carboxy-terminal domain is not essential for keratin IF formation, as exemplified by the tailless cytokeratin 19 protein. Also, removal of a portion of the carboxy-terminal domain does not inhibit the development of normal type III IF but removal of part of the amino-terminal domain from desmin (Geisler *et al*, 1982a: Kaufmann *et al*, 1985), vimentin (Traub and Vorgias, 1983) and keratin (Sauk *et al*, 1984) does appear to limit the filament-forming ability of these proteins. (Lu and Johnson, 1983, have disputed the involvement of the amino-terminal domain in filament formation but no corroboration of their evidence has been forthcoming). However, Steinert *et al* (1983a) have shown that pieces of the non-α-helical terminal domains from epidermal keratin chains could be removed enzymatically from intact filaments without affecting the structural integrity of the filament, as judged by its appearance in the electron microscope. Hence, although portions of the terminal domains may be necessary for the assembly of IF, the mature structure seems to be stabilized largely by the rod domains and does not necessarily rely on the terminal domains for maintenance of the structure. The accessibility of the terminal domains to enzymatic cleaving implies that they are at least partly, if not fully, on the exterior

surface of the filament.

Examination of the N- and C-terminal domain sequences enables classification into subdomains which show homology within each IF type (Steinert and Parry, 1985; Steinert *et al*, 1985b; see Figure 1-3). Adjacent to the rod domain are regions of high sequence homology, H1 and H2. Distal to these are domains which are variable in content and size, termed V1 and V2. At the ends of the chains are the E1 and E2 domains which usually have a high charge content. Keratin chains, for example, often exhibit glycine-serine rich motifs in the V1 and V2 domains and are basic in the E1 and E2 domains as are the type III chains. This net basic character is in contrast to the major coiled-coil regions of the rod domain which are acidic (Steinert and Parry, 1985). Some of the N- and C-terminal subdomains are not evident in particular IF groups: type I soft keratins are missing the H2 domain for example, whereas the C-terminal domain of type III chains is all H2 with no V2 or E2 regions.

1.3 Current Models of Intermediate Filament Structure

Aspects of the currently held model for the structure of the IF molecule have already been outlined: a pair of parallel right-handed α-helices aligned in axial register to form a left-handed coiled-coil configuration (the rod domain) and bracketed by terminal domains of non-repetitive secondary structure. Predictive schemes have indicated that the rod domain of the IF molecule is comprised of four coiled-coil regions (segments 1A: 35 residues, 1B: 101 residues, 2A: 19 residues, and 2B: 121 residues) separated by short links (link L1: 7-14 residues; L12: 16-17 residues and L2: 8 residues) of largely unknown conformation (Crewther *et al*, 1983; Steinert *et al*, 1983a; Dowling *et al*, 1983; Steinert and Parry, 1985; see Figure 1-3 and Table 1-1). The coiled-coil regions are readily grouped into two segments, 1 (1A-L1-1B) and 2 (2A-L2-2B), each about 20.5 nm in length (assuming the average axial rise per residue of 0.1485 nm measured from X-ray diffraction studies of α-keratin). This is very similar to a structural repeat deduced from the X-ray diffraction patterns of hard keratin (see Fraser *et al*, 1980) and visualized in electron microscopy studies of keratins and other IF (Milam and Erickson, 1982; Sauk *et al*, 1984). If a similar average axial rise per residue is assumed for the linker regions as for the coiled-coil segments, the combined length of the rod domain would be about 47 nm. The linker regions, however, could be arranged as loops external to the coiled-coil and this could result either in a minimum length of 20-21 nm if the rod domain was folded back on itself around the L12 linker region or of 41 nm for a linear rod structure. Alternatively, the links could be fully extended with an axial rise per residue of 0.334 nm leading to a rod domain as long as 53 nm. High-resolution X-ray diffraction studies have revealed an axial period



	1A	L1	1B	L12	2A	L2	2B
I	35	11-14	101	16	19	8	121
II	35	10-14	101	17	19	8	121
III	35	8	101	16	19	8	121
IV	35	10	101	17,22	19	8	121

Table 1-1 Comparison of the lengths of structural domains within the rod domain of IF proteins.

of 47 nm (Fraser *et al*, 1976) and measurements of rotary shadowed IF molecules from a variety of sources show that the length of the molecule is about 45-50 nm (Steinert, 1981; Geisler *et al*, 1982a, 1985b; Quinlan *et al*, 1984; Ip *et al*, 1985; see also Steven *et al*, 1989) and this is consistent with a partially extended structure of the link regions in a linear rod domain of the type described above.

The regular coiled-coil structure of the rod domain is correlated with the regular distribution of apolar residues in a heptad substructure - about 75% of the residues in the a and d positions of the heptad are apolar (Parry and Fraser, 1985). Aggregation of pairs of coiled-coil chains into a parallel, in-register molecule is largely due to the basic and acidic residues (lysine, arginine and aspartic and glutamic acids) which are also distributed in a non-random manner within the rod domain. The e position of the heptad has a net basic character while the g position is net acidic and hence ionic interactions (ie, salt bridges) can be made between residues in the e and g positions of adjacent chains in the molecule. This implies a parallel rather than antiparallel arrangement of chains (see Figure 1-4). Ionic interactions are maximised for an in-register arrangement of parallel chains.

In addition to regularities involving the e and g positions, highly significant long range periods have been found in the linear distributions of the acidic and basic residues in the coiled-coil domains (Parry *et al*, 1977; Parry and Fraser, 1985). In segment 1B

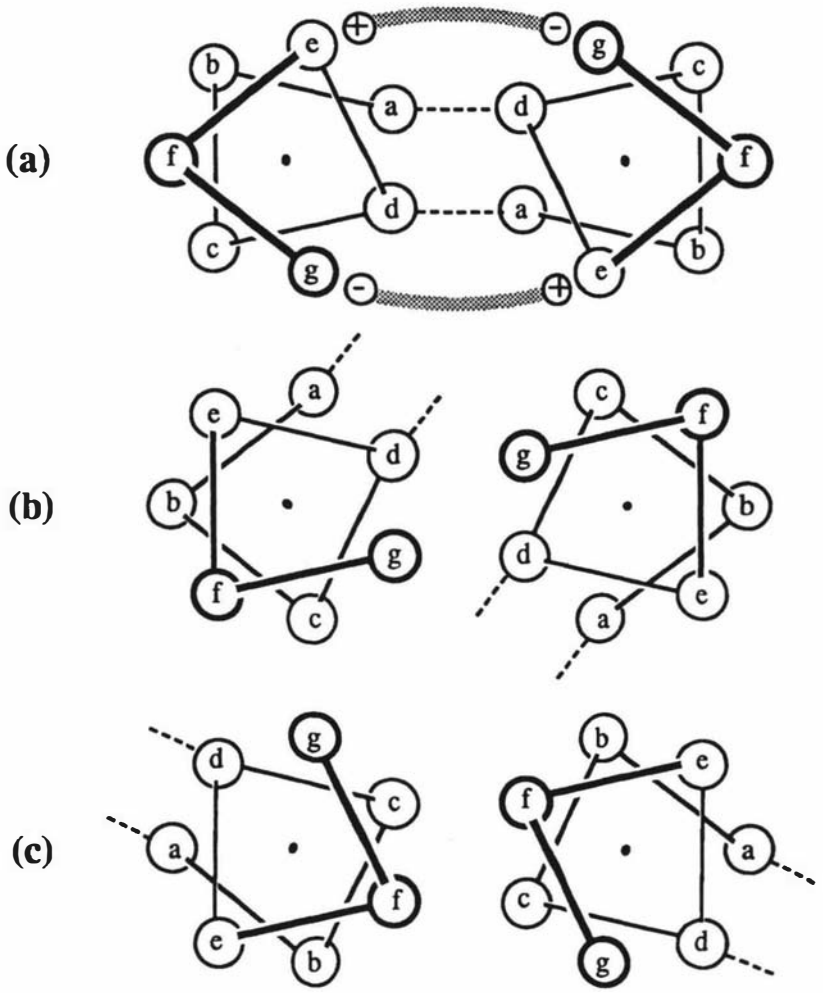


Figure 1-4 The heptad pattern in coiled-coils can be represented as $(\underline{a}\text{-}\underline{b}\text{-}\underline{c}\text{-}\underline{d}\text{-}\underline{e}\text{-}\underline{f}\text{-}\underline{g})_n$ where a and d are typically apolar. The coiled-coil structure is stabilized by the knob-hole packing of these apolar residues, as shown in (a) where the a and d positions joined by dashed lines are axially staggered relative to each other to optimize the meshing of the hydrophobic sidechains. Oppositely charged residues in the e and g positions are also able to interact by forming salt bridges (ionic interactions) and, in doing so, specify both the parallel orientations of the chains and the in-register alignment. Two stutters are apparent in the heptad substructure of the rod domains. Continuation of the undistorted coiled-coil geometry through these points are shown in (b) for the stutter near the midpoint of segment 2B and in (c) for the stutter within the link, L2. Figure adapted from Parry and Fraser (1985).

(101 residues) the periods for the acidic and the basic residues are both ≈9.6 residues (≡1.42 nm) and in segment 2 (=2A-L2-2B) they are ≈9.8 residues (≡1.46 nm). The periods of the acidic and basic residues differ in phase by about 180° in each case. Despite the heptad stutters in segment 2, the period in the charged residues is continuous throughout this section of the rod domain and indicates that segment 2 is a single structural unit (Parry, 1989). The slight, but significant difference in periods between segments 1B and 2 could prevent the aggregation of these segments via ionic interactions: only 1B-1B and 2-2 combinations might be possible. However, supercoiling of segments 1B and 2 about each other might permit an alignment of the

axial periods of the charged residues and allow 1B-2 aggregation to be achieved (Parry and Fraser, 1985).

The possible arrangements of a pair of two-stranded coiled-coils to form the four-chain structural unit (tetramer) was investigated by Crewther *et al* (1983). Only alignments that maximised the overlap of the two major coiled-coil domains, segments 1B and 2, were considered as these provided the maximum opportunity for ionic interactions to be made between the molecules. Five classes of arrangements were considered possible: parallel molecules either in-register or approximately half-staggered, or anti-parallel molecules that are either completely overlapped or half-staggered. The favoured model was a half-staggered, antiparallel arrangement with the N-terminal segments of the rod overlapped by about 28 nm giving a combined length (excluding the terminal domains) of about 60 nm. Support for this model was found in the tryptic digestion study of Woods and Inglis (1984) on hard α-keratin in which two types of helical particles were found: a four-chain fragment from the N-terminal segment of the rod domains (previously described by Woods and Gruen, 1981 and Gruen and Woods, 1983) and a heterodimeric fragment from the C-terminal segment of the rod domains. Since no fragment with both N- and C–terminal segments was found, the data strongly implied an antiparallel arrangement rather than a parallel one. (Note that an in-register, parallel arrangement of coiled-coils dimers is possible in theory but would not give rise to a filament structure). Further evidence for the antiparallel arrangement of dimers has come from limited chymotryptic digestion experiments on reduced carboxymethylated wool IF which resulted in covalently linked N-terminal segments of the rod domain (Sparrow *et al*, 1989). Cross-linking peptides were isolated and characterized and their locations were compatible only with the dimers being approximately half-staggered and antiparallel.

Support for a half-staggered arrangement of molecules is provided by the electron microscopy studies of Steven *et al* (1989) who visualized particles with a variety of lengths: 20-25 nm rods; 45-50 nm rods, some of which were kinked near their centres; and 70-80 nm rods which were kinked about one-third along from one end. All of these structures are compatible with the ~20 nm dimensions of the major coiled-coil domains in the IF chain. In addition they show at least two modes of molecular aggregation: fully overlapped and half-overlapped.

The conformations of the terminal domains are currently unknown. The enzyme cleavage studies on epidermal keratin IF by Steinert *et al* (1983a, described above) indicate that at least part of the terminal domains are located on the outside of the filament. Solid state Nuclear Magnetic Resonance (NMR) studies on epidermal keratin

IF (Mack *et al*, 1988; Steinert *et al*, 1989) and prekeratin IF (Steven *et al*, 1989) revealed little order in the structures of the N- and C-terminal domains but showed high flexibility about the ubiquitous glycyl peptide bonds, and by implication, of the polypeptide backbones in these domains. Properties of the V1 and V2 domains bear some similarities to those of the proposed Ω-loop (Zhou *et al*, 1988; Steven *et al*, 1989) but as yet there is no direct evidence for this element of secondary structure in IF chains.

The reported diameter of the IF has varied from the earlier estimates of 7 nm (electron microscopy, Birbeck and Mercer, 1957; Rogers, 1959; X-ray diffraction, Fraser and MacRae, 1959; Fraser *et al*, 1959, 1973) to 14-15 nm (Steven *et al*, 1982, 1985; Steven, 1989) and depends to some extent on the criteria used for determining the filament's 'edge'. Protofilaments of about 2 nm diameter were described by Fraser *et al* (1962) and a '9+2' arrangement of protofilaments, reminiscent in part of the organization of tubules in cilia and flagella, was suggested for the filament by Filshie and Rogers (1961) on the basis of evidence from thin-section electron microscopy. Subsequent evidence (for example, Fraser *et al*, 1972) suggested instead a ring-core substructure. Scanning transmission electron microscopy (STEM) measurements of the radial density of unstained freeze-dried IF showed that the filament was composed of a uniform density-core of diameter 9-10 nm surrounded by a diffuse periphery that extended to 15-16 nm (Steinert *et al*, 1983b; Steven *et al*, 1985, 1989). The ratios of core mass to peripheral mass for vimentin and an epidermal keratin were similar to those for the rod domain to terminal domain masses, implying that the terminal domains are largely on the external surface of the filament. Mass-per-length measurements of native unstained filaments of vimentin using STEM have also indicated that IF consist of polymorphs: the density of the major component IF was consistent with 16 molecules in cross-section (≡32 chains) and a minor component contained about 11 molecules (Steven *et al*, 1982). The lower mass variant was suggested to be be a breakdown product or an immature filament (Steven, 1989). Reconstituted keratin IF also showed several mass variants and in each case the axial density of the filament scaled with the average density of its constituent subunits (Steven *et al*, 1983a,b). Neurofilaments reconstituted *in vitro* also contained 16 molecules in cross-section although polymorphs with fewer or greater numbers of chains were present to a lesser degree (Troncoso *et al*, 1989).

X-ray diffraction studies on well-ordered hard keratin tissue have revealed an axial period of 47 nm in which there are seven or eight quasi-equivalent units on a helix of pitch 22 nm (Fraser and MacRae, 1973a, 1983, 1985; Fraser *et al*, 1985, 1986). A discontinuous surface lattice model has been generated from the X-ray diffraction data,

the STEM mass data and ionic interaction calculations (Fraser *et al*, 1985, 1986, 1988). Each lattice point is associated with a single tetramer (two antiparallel molecules approximately half-staggered). Some studies have suggested that the IF structure is organized by increasing levels of aggregation of subfilaments: a single molecule forms the 2 nm-diameter protofilament; pairs of these protofilaments form a 4.5 nm-diameter protofibril (the tetramer); and protofibrils aggregate in groups of four (typically) as variants of the intact IF (Aebi *et al*, 1983; see also Steinert and Roop, 1988).

1.4 Conservation of Amino Acid and Gene Sequences

A comparison of the amino acid sequences of the rod domains of IF chains reveals several features common to all IF. Two regions of very highly conserved sequence occur in the last 4-5 heptads at the C-terminal end of segment 2B (Geisler and Weber, 1982; Hanukoglu and Fuchs, 1983) and in heptads 3-4 of segment 1A (Weber and Geisler, 1982, 1984; Steinert *et al*, 1984a; Parry and Fraser, 1985; Steinert and Parry, 1985). In contrast, the linker segment L1 is variable in length and shows little or no homology amongst the various IF chains. Segment L12 is more regular in length (generally 16 or 17 residues depending on the chain type) and in homology (Crewther *et al*, 1983). The secondary structure of these two link segments is not predicted to be α-helical and they do not contain the heptad substructure or regular distribution of ionic residues apparent in the coiled-coil regions. The third linker segment, L2, is absolutely conserved in length (8 residues) and shows high homology across the IF chains. In addition, the regularities in the distributions of ionic residues are maintained across L2 although the heptad substructure is disturbed somewhat by the irregular step between consecutive apolar residues at that point (see below).

The heptad substructure of the coiled-coil undergoes a conserved stutter (ie, the heptad sequence a-b-c-d-e-f-g is broken by the insertion or deletion of several residues) near the midpoint of segment 2B (an insertion of four residues: Geisler *et al*, 1982a; Dowling *et al*, 1983; Steinert *et al*, 1985c). This general feature has been observed in the coiled-coil sequences of all α-fibrous proteins except tropomyosin (see Cohen and Parry, 1989). The consequences of this disruption of the regular heptad structure are unclear although possible distortions of the coiled-coil have been theorized (McLachlan and Karn, 1983; Parry and Fraser, 1985). The effect of this particular stutter, however, is to cause the internal a and d positions to be rotated by 360°/7 (≈54°) relative to the axis of the α-helix (see Figure 1-4b). Assuming that the coiled-coil is undistorted at some distance remote from the stutter on both sides, the stutter may be accommodated by a gradual change in the pitch length of the coiled-coil over an extensive region (McLachlan and Karn, 1983). Alternatively, there may be a highly

localized discontinuity in structure which could result in a kink in the axis of the coiled-coil (Parry and Fraser, 1985). A similar feature occurs in the link region, L2, where an insertion of five residues (or deletion of two) occurs and as a result the a and d positions are rotated by ≈154° (Figure 1-4c). The requirement for such distortion of the regular coiled-coil structure is not known but its importance is reflected in the absolute conservation of these stutters in all IF protein sequences.

Regularities in the dispositions of acidic and basic residues in the rod domain of IF proteins have already been described: the charged stripes evident in the e and g positions of the heptad specify in part the alignment of chains to form the dimeric IF molecule and the next level of structure appears to be determined by the 9.6-residue period of the charged residues. It is possible that higher levels of IF structure are also dependent to some degree on other regularities in the positioning of these residues. The self-assembly of IF *in vitro*, and possibly *in vivo* also, must be some function of the protein sequences. The dispositions of acidic and basic residues, as well as the apolar residues, appear to play an important part in this process.

Conserved features are also apparent in the nucleotide sequences of the genes that encode IF proteins. In particular, the positions of six introns are generally conserved within the rod domain (Quax *et al*, 1983, 1985; Lehnert *et al*, 1984; Marchuk *et al*, 1984; Krieg *et al*, 1985; Rieger *et al*, 1985; Tyner *et al*, 1985; Lewis and Cowan, 1986). Desmin, vimentin and GFAP have, in addition to these six introns, two others at common positions within the C-terminal domain and several keratin protein genes also exhibit one of these (Marchuk *et al*, 1984; Tyner *et al*, 1985; Lewis and Cowan, 1986).

An interesting exception to this pattern is the bovine cytokeratin 19 protein (Bader *et al*, 1986) which was found to be lacking the intron near the junction of the C-terminal and rod domains. This protein is also unusual in that the C-terminal domain is small (9 residues in length) and effectively constitutes a continuation of the heptad substructure from the rod domain: apparently it is more an extension to the rod than a distinct non-α-helical domain. The corresponding human cytokeratin 19 protein (Eckert, 1988) also has the same intron structure and the same number of amino acids. Furthermore, the two protein sequences are 89% identical.

Other exceptions to the general IF intron pattern are the neurofilament proteins which have only two introns within the rod domain, both of which are conserved among the NF proteins but which do not correspond to intron positions in the other IF proteins (NF-L: Lewis and Cowan, 1986; NF-M: Myers *et al*, 1987; NF-H: Lees *et al*, 1988). Genes for mouse and human NF-L proteins show 90% homology within the exons,

conservation of the positions and sizes of the introns but considerable differences in the nucleotide sequences of the introns (Julien *et al*, 1987).

The purpose of introns has not been established and their positions do not seem to coincide with structural domains or functional properties of the proteins in any regular way. They do, however, allow some conclusions to be drawn on the evolution of the genetic material of which they are a part. The conservation of introns among a protein family expressed in such diverse cell types is an important indication of a common ancestor gene. In addition, the difference between the intron positions for the types I-III and type IV groups provides evidence on how long ago they diverged (see, for example, Steinert and Roop, 1988).

1.5 Lamin, Peripherin and the *Helix* A and B proteins

The nuclear envelope is a double membrane composed of two lipid bilayers separated by approximately 20 to 40 nm. Between the external and internal membranes is a region known as the perinuclear space. The outer nuclear membrane is continuous with the membrane of the endoplasmic reticulum, as is the perinuclear space with the lumen of the endoplasmic reticulum. The ribosomes that 'stud' the rough endoplasmic reticulum are also apparent on the outer nuclear membrane. The inner and outer nuclear membranes are connected at the nuclear pores which allow exchange of material between the cytoplasm and nucleus.

A complex layer of filaments lining the nucleoplasmic surface of the inner nuclear membrane has been observed in invertebrate cells, and a similar layer of filaments, 15-20 nm thick, has also been seen in ultra-thin sections of vertebrate cells (Fawcett, 1966; Scheer *et al*, 1976). The diameters of the filaments, called the nuclear lamina, have been reported as 5-10 nm (Scheer *et al*, 1976). The native lamin meshwork of *Xenopus* oocytes was shown to be an orthogonal network with sides of about 52 nm (Aebi *et al*, 1986). This filamentous meshwork comprises a scaffolding associated with the pore complexes, the inner nuclear membrane and interphase chromatin. It also appears to assist in the structural organization of the nucleus. Gerace *et al* (1978) showed that the nuclear lamina from rat liver cells was comprised mainly of three proteins (termed lamins A, B and C by Gerace and Blobel, 1980) with molecular weights of 60-70 kDa and which were located at the periphery of the normal interphase nucleus. In addition, lamin B was found to be significantly different from lamins A and C on the basis of peptide mapping and immunological studies (Gerace and Blobel, 1980; Shelton *et al*, 1980). However, coincident with the disassembly of the nuclear envelope during prophase, antigens to the three nuclear lamin proteins became distributed evenly throughout the cell until telophase, at which stage they became

localized around the daughter chromosomes. It was proposed that the lamina was reversibly disassembled concomitant with disintegration of the nuclear envelope during mitosis.

The nuclear lamina was shown to be interposed between the nuclear envelope and the chromatin (Fawcett, 1966). Lamin B was more resistant to extraction from membranes than lamins A and C (Gerace and Blobel, 1982) which suggests that it has a special membrane binding role and indeed a lamin B receptor protein has been described recently in turkey erythrocytes (Worman *et al*, 1988) and in yeast cells (Georgatos *et al*, 1989). During mitosis, lamins A and C become diffuse throughout the cytoplasm but lamin B remains associated with membranes (Gerace and Blobel, 1980; Burke and Gerace, 1986). It has been suggested that nuclear membrane fragments may be 'labelled' by lamin B for use in re-establishment of the envelope (see Stick *et al*, 1988). A study on chicken nuclear lamin proteins, lamins A, B1 and B2, also found that lamin A was dispersed throughout the cytoplasm during mitosis and that lamins B1 and B2 were membrane-bound and hence were considered functional analogues of the mammalian lamin B (Stick *et al*, 1988). In addition, lamin B2 was found to be associated with the endoplasmic reticulum and, during the mitotic disassembly of the nuclear envelope, the lamin B2 proteins may become reversibly concentrated in the endoplasmic reticulum. Characterization of the different biochemical and structural features of the A and B lamins has been described recently by Peter *et al* (1989).

The integrity of the lamin meshwork is thought to be regulated by the degree of phosphorylation of the lamin proteins (Gerace and Blobel, 1980). Sites for phosphorylation in keratin are specific serine and threonine residues in the terminal domains (Steinert, 1988). Such residues are abundant in the C-terminal domain of the lamin molecules: for example, examination of the protein sequences shows that this domain of the human lamins A and C contains 44 and 23 serine residues respectively and 20 and 14 threonine residues respectively. However, newly synthesized lamin proteins require some alternative to phosphorylation in order to avoid the possibility of premature oligomerization while in the cytoplasm. A higher molecular mass precursor (71 kDa) to lamin A (68 kDa) was studied by Lehner *et al* (1986a) who suggested that the precursor form could allow migration of lamin A protein into the nucleus where the precursor would be processed into the mature protein and integrated into the lamina meshwork in some unspecified manner. Lehner *et al* (1986a) also identified two variants of the lamin B protein: a mammalian form, lamin B1, and an avian form, lamin B2. An apparent higher molecular mass precursor for lamin B2 was also synthesized but *in vitro* translation produced equal amounts of the two forms of the protein. The reason for this is unclear. No precursor was found for lamin B1. (No

investigation was made into precursors for lamin C. However, if the lamin proteins copolymerize, a non-aggregating form of one protein could be sufficient to prevent filament formation).

Zackroff *et al* (1984) isolated 'keratin-like' proteins located near the periphery of the nucleus and suggested a relationship with the nuclear lamin proteins. This was confirmed in further studies by Goldman *et al* (1986) who also proposed that the nuclear lamins formed an IF-like network in the nuclear lamina and that this was connected with the cytoplasmic IF network, possibly through the nuclear pores or by some trans-membrane mechanism. The possibility of the filamentous networks being unrelated was discounted for two reasons: the component proteins were related biochemically and the cytoplasmic IF network was known to be closely associated with the nuclear surface as well as the plasma membrane. The protein sequences for human lamins A and C were determined by McKeon *et al* (1986) and Fisher *et al* (1986), who showed conclusively that the nuclear lamins were indeed IF-type proteins. Amphibian lamin proteins from *Xenopus* laevis have been described and named lamins A, L_I, L_{II}, L_{III}, L_{IV} (for review, see Wolin *et al*, 1987). Invertebrate lamins from the surf clam and Drosophilia have also been reported as have minor lamins (Lehner *et al*, 1986b).

The primary structures of human lamins A and C bear a remarkable resemblance to those of the other IF proteins and common structural properties have been postulated for both lamins and IF (Gerace, 1985; McKeon *et al*, 1986; Fisher *et al*, 1986). McKeon *et al* (1986) demonstrated that human lamins A and C shared the general structural plan of IF chains – an α-helix-rich rod domain bracketed by largely non-α-helical N- and C-terminal domains – and that the rod domain of lamins may be subdivided into segments in much the same manner as for IF. In addition, the human A and C lamins were seen to be identical from the N-terminus through the rod domain and into the C-terminal domain and McKeon *et al* (1986) suggested that both proteins arose from the same gene by differential processing – this may be related to the unusual repeat of four histidine residues after which the two polypeptide sequences diverge.

An important difference noted between the rod domains of IF and lamin was a six-heptad insert (ie, 42 residues) in segment 1B of the lamin chains. Weber (1986) also noted that the start of the insert occurred at the position of a conserved intron (see also Steinert *et al*, 1985b). This extension is predicted to result in a length of ~52 nm for the rod domain of lamin molecules. Dimers of lamins A and C from rat liver reveal a ~52 nm rod with two globular heads at one end (Aebi *et al*, 1986), presumably the

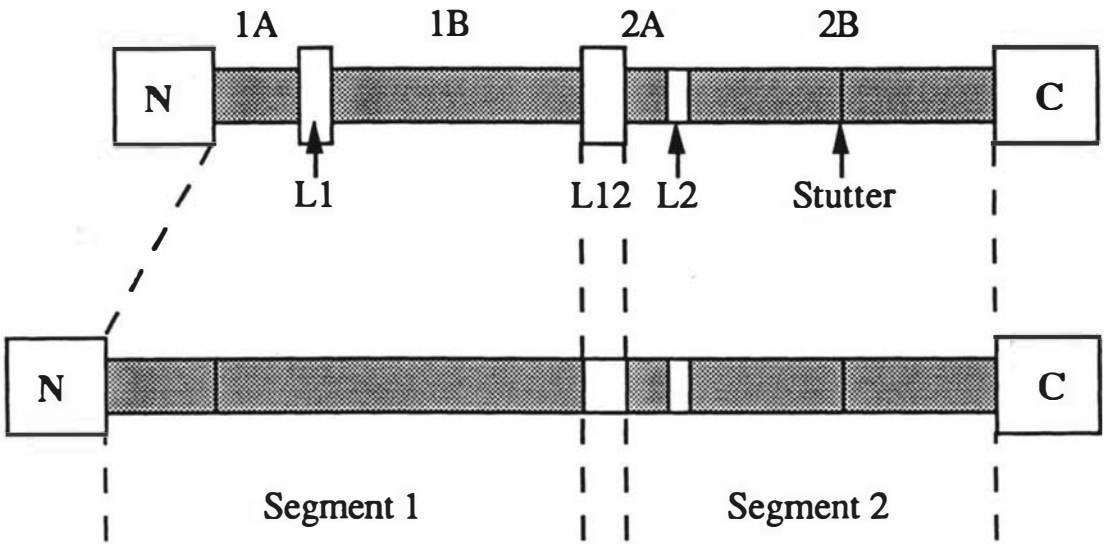


Figure 1-5 Schematic representations of the chain structures for IF types I-IV (top) and lamin. Linker segments L1 and L12 in types I-IV IF are not predicted to be α-helical in structure, as indicated by boxes wider than the shaded coiled-coil regions (top). In the type V (lamin) chain, however, the entire rod domain is predicted to be α-helical (bottom). Segment 1B is extended by a 42-residue insertion and linker segment L1 has been replaced by a stutter in the heptad substructure.Figure adapted from Parry *et al* (1986).

larger C-terminal domains. Dimers of lamin B were similar in form although typically only one globular head was apparent. Other points noted by Parry *et al* (1986) in their analysis of the sequence data were (i) the 42-residue insert is part of a 70-residue piece in human lamins A and C that replaces residues 43-70 inclusive in segment 1B of IF chains (numbering from the amino-terminal end of segment 1B), (ii) segment L12 is predicted to be α-helical in lamin chains but non-α-helical in other IF chains, (iii) segment L1 in lamin has a heptad substructure and is also predicted to be α-helical and indeed it can be reduced to a single stutter in the heptad phasing for segment 1, and (iv) unlike IF type I-IV chains, lamins do not appear to form long filamentous structures *in vitro* but instead form tactoids or paracrystalline arrays which have alternating dark- and light-staining bands when negatively stained and observed by electron microscopy (Zackroff *et al*, 1984; Goldman *et al*, 1986). The resultant model for the lamin chain is shown in Figure 1-5 alongside that for IF chains (Parry *et al*, 1986). The rod domain of human lamins A and C is predicted to be entirely α-helical with a heptad substructure present throughout segment 1 and in segments 2A and 2B. In addition, the ionic and heptad regularities in the rod domain were found to be maintained across the linker region L12. Parry *et al* (1986) suggested that the lamins be termed type V IF proteins since they showed a high degree of homology with IF overall but did not fall naturally into the type I-IV classification previously established. Indeed, the novel extension to segment 1B and the intranuclear location of these proteins together imply that the lamins are fundamentally different from the cytoplasmic IF proteins.

An interesting parallel to the description of the lamin protein structure is the recent characterization of other proteins that have the same longer 1B segment seen first in the lamin A and C proteins. Weber *et al* (1988) have sequenced two epithelial proteins from the snail *Helix pomotia* of molecular weights 66 kDa (*Helix* A) and 52 kDa (*Helix* B). Both form monocomponent IF *in vitro* indicating that they are not analogous to the heteropolymeric keratin IF despite their occurrence in a near equal molar ratios in epithelial tissue. Although the homology in structure and sequence between the nuclear lamin and *Helix* proteins might have been taken as evidence for a similar location and function, this is not supported by Weber *et al* (1988) who note that a karyophilic motif in the C-terminal domains of the nuclear lamin proteins (the sequence Lys-Lys-Arg-Lys-Leu-Glu, Fisher *et al*, 1986) is not apparent in the sequence of the *Helix* proteins. However, the departures from the type I-IV IF structure described above for human lamin A and C are shared also by the *Helix* B protein. No other IF proteins have demonstrated this variation to the structure of the rod domain and so it is appropriate to designate the *Helix* B protein a type V IF chain. The sequence of the *Helix* A reported by Weber *et al* (1988) was incomplete and no assignment is yet possible for this chain.

Studies on a neuronal protein of molecular weight ~58 kDa by Liem *et al* (1978), Portier *et al* (1984a,b), Franke *et al* (1986) and Parysek and Goldman (1987) had shown that it was related to the IF family but was not one of the neurofilament triplet proteins, NF-L, NF-M or NF-H. This protein was termed peripherin by Portier *et al* (1984a). The sequence (Leonard *et al*, 1988; Parysek *et al* , 1988) was shown to have greatest homology with the type III IF sequences, desmin and vimentin. Although its expression in neuronal tissue might suggest that peripherin is a type IV IF protein, the gene structure (Thompson and Ziff, 1989) follows the pattern for the type I-III IF proteins (six conserved introns in the rod domain) rather that for the neurofilament proteins (two conserved introns in the rod domain) and peripherin is therefore conclusively not a type IV IF protein.

The high content of serine residues in the N-terminal domain was described as being "novel" (Leonard *et al*, 1988). However the peripherin protein is only average in this respect amongst the type III IF proteins which have 23 (hamster desmin), 25 (hamster vimentin), 7 (mouse GFAP) and 24 (peripherin) serines in this domain (sequences as cited by Leonard *et al*, 1988). Indeed, it is only the GFAP protein that is unusual in having a lower serine content in its N-terminal domain. The serine residues are potential sites for phosphorylation, a process which appears to have importance as a regulatory mechanism for the polymerization of IF proteins in general (see later). Phosphorylated peripherin, however, is found almost exclusively in the insoluble

cytoskeleton (Aletta *et al*, 1989), indicating that it may still be an integral part of the filament network. Aletta *et al* (1989) suggest that this is consistent with phosphorylation-promoted assembly, an opposite effect to that reported for other IF proteins. Vimentin (Inagaki *et al*, 1987) and desmin (Geisler and Weber, 1988), for example, have a reduced ability to polymerize when phosphorylated and the nuclear lamin network disassembles concomitant with phosphorylation of the lamin proteins (Gerace and Blobel, 1980). Aletta *et al* (1989) also report that non-detectably phosphorylated peripherin is present in the cytoskeleton and so no firm conclusion can be drawn from these data as to the effect of phosphorylation on the *in situ* proteins.

A truly novel feature of the peripherin protein is its expression in neuronal tissue, previously thought to be the exclusive domain of the neurofilament proteins. Leonard *et al* (1988) reported that the distributions of the NF proteins and peripherin mRNA within the nervous system are different, although there is some overlap, and that the regions which express peripherin are all evolutionarily old. The term "peripherin" originated from the location of this protein in neurons peripheral to the nervous system (Portier *et al*, 1984a) although it is also expressed in certain of the central nervous system neurons (Leonard *et al*, 1988).

1.6 Function of Intermediate Filaments

Intermediate filaments form a highly insoluble skeleton that extends throughout the cytoplasm and is also found in the nucleus. The functions of the IF networks are unclear as are the mechanisms by which they are controlled and their associations with other elements of the cytoplasm. Their ubiquity and underlying homology imply some fundamental importance for cell physiology that transcends species and cell types, although some cultured cell lines apparently grow normally in the absence of IF (Venetianer *et al*, 1983). The nuclear lamin proteins appear to be especially widespread in eukaryotic cells: lamin A and B analogues have been reported in yeast cells (Georgatos *et al*, 1989) in addition to mammals, amphibia, insects and birds. The first steps towards understanding the role of IF have been made by determining the anchorage sites of some IF proteins to membranes and it has become increasingly apparent that IF constitute some form of connection between the cell surface and the nucleus.

Vimentin IF were shown to be associated with the cell membrane, possibly via head-on interactions between the non-α-helical amino-terminal domain and at least one membrane-anchoring protein, ankyrin (Georgatos *et al*, 1985; Georgatos and Marchesi, 1985). Studies on desmin IF located the plasma membrane binding site to within the amino-terminal domain (Georgatos *et al*, 1987). Georgatos *et al* (1985)

showed that ankyrin inhibited IF formation of vimentin beyond a protofilament stage (probably the tetrameric subunit) and, since the *in vitro* assembly process for IF appears to require intact amino-terminal domains (vimentin: Traub and Vorgias, 1983; desmin: Kaufmann *et al*, 1985), this suggests that ankyrin may serve *in vivo* as a "capping site" that effectively terminates the filament at the membrane (Georgatos *et al*, 1987; Steinert and Roop, 1988). An alternative attachment method was suggested by Georgatos *et al* (1985) whereby the filaments bind side-on to the membrane via short branches and loop back into the cytoplasm. However, the branches might themselves abut the membrane in an end-on fashion. In addition to vimentin and desmin, keratin IF also appear to attach to the cell periphery, possibly to desmoplakin proteins which are found in complex regions of cell-to-cell adhesion called desmosomes (see, for example, Goldman and Dessev, 1989; Green *et al*, 1989).

IF are also localized around the exterior of the nuclear surface. *In vitro* binding of desmin to lamin B has been described (Georgatos *et al*, 1987) and this association apparently involves a portion of the carboxy-terminal domain of desmin adjacent to the rod domain. Vimentin has also been observed to associate with lamin B in a similar manner to desmin (Georgatos and Blobel, 1987). Indeed, lamin B has been suggested as a nucleating centre for IF (see, for example, Gerace and Burke, 1988) since it is clear that the ankyrin-mediated attachment to the plasma membrane is not capable of performing such a role (Georgatos *et al*, 1985). The method by which lamin B might interact with IF proteins through the nuclear envelope, if indeed it does so at all, is unclear. A connection through the nuclear pores is possible as is a trans-membrane interaction mediated by some membrane process. Georgatos and Blobel (1987) have suggested that IF networks may be directly anchored to the nuclear lamina at distinct locations coinciding with the nuclear pores.

A feature of the direct connection model for IF to membrane-binding proteins is that a certain amount of local polarity is implied for the IF: the amino-terminal domains of the IF chains link to the plasma membrane whereas the carboxy-terminals bind at the nuclear membrane. However, the structures reconstituted *in vitro* show no axial polarity (see, for example, the GFAP paracrystals of Stewart *et al*, 1989a,b and Quinlan *et al*, 1989). If the filaments are axially apolar (ie, equal numbers of oppositely directed chains), why should the amino- and carboxy-terminal domains of the molecule associate with separate membranes in the cell? What role then is played by the amino- and carboxy-terminal domains in the main body of the filaments? Do they bind to some other membranes in the cytoplasmic space?

In addition to the locale of IF attachment, recent work has examined the role of phosphorylation as a post-synthetic modifier of IF structure. Phosphorylation of the nuclear lamina has already been described as has the general oligamerization-blocking effect of phosphorylation on vimentin and desmin (above) and evidence is now accumulating that the cytoplasmic IF networks also undergo structural changes as a result of phosphorylation. Major sites for phosphorylation have been determined for the human epidermal keratin 1 protein as well as the turnover rates of phosphate isomers of this protein (Steinert, 1988). Sites in the more mobile parts of the terminal domains showed the highest rate of turnover and the rate dropped in sites closer to the relatively stiff rod domain. This may be due in part to the accessibility of different regions of the chain to external phosphorylating agents. As Steinert (1988) points out, there is no indication why the keratin IF examined should undergo post-synthetic disassembly and reassembly and so the purpose of this phosphorylation remains unclear. Other cellular activities may cause the proximity of phosphorylating agents to the IF network leading to opportunistic interactions.

Disassembly of IF networks during mitosis has been studied recently by Chou *et al* (1989). The amount and sites of phosphorylation on the IF chains (desmin and vimentin) were stated to be different during mitosis than during interphase and complete disassembly of the *in vitro* IF network by phosphorylation was described. These results indicate that phosphorylation is a regulator of IF structure: phosphorylation can disassemble IF networks (and block oligamerization of the subunits: Geisler and Weber, 1988) and conversely dephosphorylation will allow self-assembly of the molecules into IF.

An interesting development has been the recent discovery of lamins A and B in yeast (Georgatos *et al*, 1989). These cells divide by closed mitosis where the nuclear envelope is not dismantled as in higher eukaryotic cells but instead develops a cytoplasmic extension that 'buds off' as a daughter nucleus with a separate membrane. There is no apparent requirement for the disassembly/reassembly of the nuclear lamina and the role of the lamina must be directed entirely towards maintaining the structural integrity of the nuclear envelope (as described above). An investigation into the primary structure of such lamins would be of use in identifying regions of the chain than are involved in specific functions. Also, the consequence of hyper-phosphorylation of the yeast lamins, which presumably does not take place during closed mitosis, would be of interest.

1.7 Structural Form of the Thesis

The detailed structure of IF cannot be determined from the currently available X-ray diffraction data or from electron microscopy studies. However, there are some attributes of their conformations that may be elucidated by non-physical techniques. The studies described in this thesis are a collection of such indirect methods which are nonetheless evaluated on the basis of their ability to conform with such physical data as are available. Protein sequence data are the raw material on which these studies are based: the ever increasing body of sequences available allows new insights to be made into the structural hierarchy of the IF.

Periodicities are examined in the linear distribution of charged residues within the rod domains of a type III IF protein (peripherin) and of various type V IF chains (the nuclear lamins and the *Helix pomotia* B protein). It is important to establish whether differences occur between these proteins and other IF chains, especially with regard to the extended 1B segment characteristic of the type V chains.

Studies on the primary sequences of IF in general include investigation of residue occupancy in the heptad substructure, calculation of chain flexibilities (Chapter 2) and quantification of the degree of homology present amongst the IF proteins (Chapter 3). For the latter work, amino acid sequences of the rod domains of IF chains are compared in order to describe more accurately the features that are common to IF chains as well as to distinguish sub-groups. Consensus sequences for the subtypes and for all IF chains are derived and, in addition, homology among residues in the heptad positions of the coiled-coil are determined. The current data base is sufficiently large to allow some interesting conclusions to be drawn that aid in further elucidation of IF structure and function.

The packing of IF chains into the molecule and of the molecules into larger scale structures is also studied (Chapter 4). Prediction schemes based on the calculation of potential ionic interactions enables the most favourable axial alignments to be found. The simple methods employed in investigating the alignment of molecules are expanded to include more of the three-dimensional information in the molecular structure.

Models for the aggregation of lamin molecules based on electron microscopy data and ionic interaction studies are constructed. These are used to explain some of the more unusual features of the nuclear lamina as well as indicating how phosphorylation can result in depolymerization of the lamina filaments.

2. PRIMARY STRUCTURE

The number of IF protein sequences available has increased dramatically over the seven or so years since the first full sequence - chicken gizzard desmin - was determined by Geisler and Weber (1982). At the present time over 40 sequences have been fully or partially completed. Examination of these sequences has yielded valuable clues to the likely molecular and filament structure of the proteins. For example, secondary structure can be predicted fairly reliably for fibrous proteins where extensive regions of fixed conformation exist. On this basis, a model for the structure of the IF molecule has been proposed in which there is a central rod-like domain with a regular coiled-coil structure that is interrupted by well conserved breaks. Higher levels of organization, however, are less easy to determine due to the complexity of the information encoded in the protein sequences. Regularities in the disposition of residues having a certain character (for example, the mutually attractive acidic and basic residues) may be indicators of special modes of aggregation of specific IF chains. The likelihood of this feature arising purely by chance is greatly reduced where these regularities are shared by many other protein chains. The characterization of regular patterns in the distribution of residues is not sufficient to deduce the packing order of the protein chains but does indicate that regular packing of the proteins does occur. A list of the proteins used in various studies in this work are given in Table 2-1. Several other IF protein sequences have been published recently but are not included in any of the analyses detailed in this thesis. These include the human epidermal type II keratin K5 (Lersch and Fuchs, 1988) and the rat NF-H protein (Dautigny *et al*, 1988).

In this Chapter, regularities in the dispositions of the charged residues are examined for the rod domains of the newly characterized proteins: the human and *Xenopus* lamin proteins, the *Helix pomotia* B protein and peripherin. Fourier transforms of the distributions of these residues have already been undertaken on a number of type I to IV IF proteins and these are compared with the new data presented here. Secondly, this study looks at the distribution of residues in the heptad substructure and how it relates to the predicted structures of the coiled-coil segments in the rod domain. Finally, the flexibility profiles of representatives of each of the classes of IF proteins are calculated with a view to comparing known structural features of the chains with their predicted mobility.

Type I		
Abbrev.	Full name	Source
B40K	Cow (Bovine) 40K	Bader *et al* (1986)
B50K	Cow (Bovine) 50K	Jorcano *et al* (1984)
B54K	Cow (Bovine) 54K	Jorcano *et al* (1984), Rieger *et al* (1985)
H46K	Human 46K	Raychaudury *et al* (1986)
H50K	Human 50K	Hanukoglu and Fuchs (1982), Marchuk *et al* (1984, 1985)
H56.5K	Human 56.5K	Steinert *et al* (private communication)
M47K	Mouse 47K	Singer *et al* (1986)
M50K	Mouse 50K	Steinert and Roop (private communication)
M50K	Mouse M pkSCC 50K	Knapp *et al* (1987)
M52K	Mouse M pkSCC 52K	Knapp *et al* (1987)
M55K	Mouse	Steinert and Roop (private communication)
M59K	Mouse	Krieg *et al* (1985), Steinert *et al* (1983a)
8a	Sheep Component 8a	Crewther *et al* (1985)
8c-1	Sheep Component 8c-1	Crewther *et al* (1983), Dowling *et al* (1983, 1986)
XL51	*Xenopus*	Hoffmann and Franz (1984)
XL70	*Xenopus*	Winkles *et al* (1985)
XL81	*Xenopus*	Jonas *et al* (1985)

Table 2-1 IF amino acid sequences.

Type II		
Abbrev.	Full name	Source
H55K	Human 55K	Glass *et al* (1985)
H56K	Human 56K	Hanukoglu and Fuchs (1983), Tyner *et al* (1985)
H67K	Human 67K	Steinert *et al* (1985a)
M60K	Mouse 60K	Steinert *et al* (1984a)
M67K	Mouse 67K	Steinert *et al* (1985a)
5	Sheep Component 5	Crewther *et al* (1985)
7c	Sheep Component 7c	Sparrow and Inglis (1980), Crewther *et al* (1983), Dowling *et al* (1983), Rogers (1984)
7x	Sheep Component 7x (unnamed)	Powell and Rogers (private communication)
XL64 23	*Xenopus* XL 64 (pUF23)	Hoffmann *et al* (1985)
XL64 164	*Xenopus* XL 64 (pUF164)	Hoffmann *et al* (1985)
Type III		
Abbrev.	Full name	Source
CGD	Chicken Gizzard Desmin	Geisler and Weber (1982, 1983), Geisler *et al* (1982a)
HD	Hamster Desmin	Quax *et al* (1984)
PSD	Pig Stomach Desmin	Geisler and Weber (1981), Geisler *et al* (1982b)
CV	Chicken Vimentin	Zehner and Paterson (1985)
HELV	Hamster Eye Lens Vimentin	Quax *et al* (1983), Quax-Jeuken *et al* (1983)
PELV	Pig Eye Lens Vimentin	Geisler and Weber (1981), Geisler *et al* (1982b)
MGFAP	Mouse Glial Fibrillary Acidic Protein	Lewis *et al* (1984)
PGFAP	Pig Glial Fibrillary Acidic Protein	Geisler and Weber (1982, 1983)
RP	Rat Peripherin	Leonard *et al* (1988), Parysek *et al* (1988), Thompson and Ziff (1989)

Table 2-1 *continued*

Type IV		
Abbrev.	Full name	Source
HNF-L	Human Neurofilament Light	Julien *et al* (1987)
MNF-L	Mouse Neurofilament Light	Lewis and Cowan (1985, 1986)
PNF-L	Pig Neurofilament Light	Geisler *et al* (1982b, 1983, 1985c)
RNF-L	Rat Neurofilament Light	Julien *et al* (1985)
HNF-M	Human Neurofilament Medium	Myers *et al* (1987)
MNF-M	Mouse Neurofilament Medium	Levy *et al* (1987)
PNF-M	Pig Neurofilament Medium	Geisler *et al* (1984)
RNF-M	Rat Neurofilament Medium	Napolitano *et al* (1987)
HNF-H	Human Neurofilament Heavy	Lees *et al* (1988)
MNF-H	Mouse Neurofilament Heavy	Shneidman *et al* (1988)
PNF-H	Pig Neurofilament Heavy	Geisler *et al* (1985a)
Type V		
Abbrev.	Full name	Source
HLA	Human Lamin A	McKeon *et al* (1986), Fisher *et al* (1986)
XLA	*Xenopus* Lamin A	Krohne *et al* (1987)
XLB	*Xenopus* Lamin B	Wolin *et al* (1987)
Hlx B	*Helix pomotia* B	Weber *et al* (1988)

Table 2-1 *continued*

2.1 Fourier Analysis

This section will be concerned with an investigation of the regularities in the linear disposition of the acidic and the basic residues in the amino acid sequences comprising the major heptad-containing segments of human lamins A and C, *Xenopus* lamins A and B, *Helix pomotia* B protein and peripherin (see Table 2-1 for sources). Each of these proteins manifests some unique extension to the basic model for IF protein structure or location: the lamin proteins are special amongst IF in that they occur within the cell nucleus (other than during mitosis); the type III peripherin protein occurs in neuronal tissue, previously thought to be the exclusive domain of the type IV neurofilament proteins; the lamin and *Helix pomotia* proteins have a novel extension to the 1B segment of the rod domain that would allow the coiled-coil structure to be extended, however unlike lamin the *Helix pomotia* proteins do not appear to be located within the cell nucleus (Weber *et al* , 1988).

Regularities in amino acid sequences are associated either with the structural or the

functional features of proteins, and their evaluation gives important clues in model building studies. The heptad substructure characteristic of regions of coiled-coil conformation, for example, is largely defined by the apolar residues spaced at intervals of three and four residues successively (described in more detail in Chapter 1). In addition, the charged residues are often found in specific positions within the heptad which facilitate either interchain or intermolecular interactions. Patterns such as these involving the e and g heptad positions (defined in Chapter 1) allow pairs of such coiled-coils to pack together in an optimal way (the 'knob-into-hole' packing postulated by Crick, 1952), shielding hydrophobic residues from the aqueous environment and bringing oppositely charged residues from both chains into close proximity where they can form attractive electrostatic interactions. IF proteins express these sequence features clearly and together they are largely responsible for the self-assembly of chains into the dimeric molecule. Other regular periods are also evident in IF protein sequences and these are thought to direct higher levels of aggregation from the molecule to the tetrameric unit, and the tetrameric unit to the 10 nm filament. Parry and Fraser (1985) have investigated the rod domain sequences of a number of IF proteins for regularities in the placement of acidic and basic residues. They described a feature common to the major coiled-coil domain sequences of all IF proteins in the type I to IV categories: segment 1B shows a ~9.54 residue period for these residues and segment 2 has a slightly longer period at ~9.84 residues. In addition, the periods of the acidic and basic residues were found to lie approximately 180° out of phase indicating that charged residues were grouped along the chain in clumps of alternating charge.

2.1.1 Fourier Analysis - Method

Fourier transform techniques applied to primary sequence data from proteins have been described by McLachlan and Stewart (1976) and are used here after minor modification. The original method is as follows: residues of interest in the amino acid sequence (for example, acidic or basic residues) are represented by unity and all other residues are replaced by zero. The data set is then zero-filled to improve resolution of the resulting Fourier data and also to meet the modulo 2 size constraints imposed on the data set by the Fast Fourier Transform (FFT) algorithm used. Protein sequences studied have lengths of either 101 residues (segment 1B for IF types 1-IV), 143 residues (segment 1B for IF type V) or 148 residues (segment 2 for all IF types) and all are zero-filled to 2048. The discrete Fourier transform D(f) of the zero-filled data set d(x) is then calculated:

$$D(f) = \mathbf{F}\{d(x)\} = \frac{1}{N}\int_{0}^{N} d(x).\exp\left(\frac{-2\pi i x f}{N}\right)dx$$

The variable x is the residue position in the linear sequence and f is its analogue in the Fourier domain. The Fourier intensities are scaled by

$$\frac{N_r C_r}{N_r - C_r} \frac{1}{I_0}$$ where N_r = size of the unfilled residue data set
C_r = count of residues of interest
I_0 = Fourier intensity at zero frequency

The scaled Fourier intensities will thus be given by

$$\mathbf{FI}\{d(x)\} = \frac{N_r C_r}{N_r - C_r} \frac{1}{I_0} |D(f)|^2$$

where $|D(f)|^2$ are the Fourier intensities of D(f).

The method outlined above is modified by the introduction of an extra operation prior to zero-filling: the linear sequence is baseline corrected, ie, the average value of the sequence is subtracted from all data points so that the new average is zero. This simple procedure avoids the large zero-frequency spike and its associated side lobes that are otherwise evident in the Fourier transforms. A comparison between the original and the baseline correction methods is shown in Figure 2-1 together with the difference between the transforms. The large zero-frequency peak (Figure 2-1b) results from the non-zero average value of the data set and would not generally be considered a problem. However, in combination with the zero-filling operation, the large spike is spread by a significant amount through the Fourier domain and interferes to varying degrees with other peaks as well as producing 'spurious' peaks at low frequencies (see Appendix A). These 'spurious' peaks, or side lobes, are only apparent in the Fourier data when the central peak is relatively large (the ratio of the central lobe of the sinc function to the first side lobe is approximately 20:1). This is generally the case for the zero-frequency peak if the baseline is not corrected. As can be seen from Figure 2-1c, the effect of the baseline correction is small beyond the zero-frequency sinc function.

The scaling factor described above from McLachlan and Stewart (1976) makes use of the zero-frequency intensity I_0 produced by their method. However, the baseline correction method results in I_0 being zero and so an alternative derivation for the scaling factor must be used. The value of I_0 for the original method is also equal to the square of the average value of the digitized sequence prior to the baseline correction and may be used in place of I_0 in the calculation of the scaling factor.

The relationship between frequency in Fourier space and period in the residue sequence is:

$$\text{period} = \frac{N_f}{\text{frequency}}$$ where N_f = size of the filled data set.

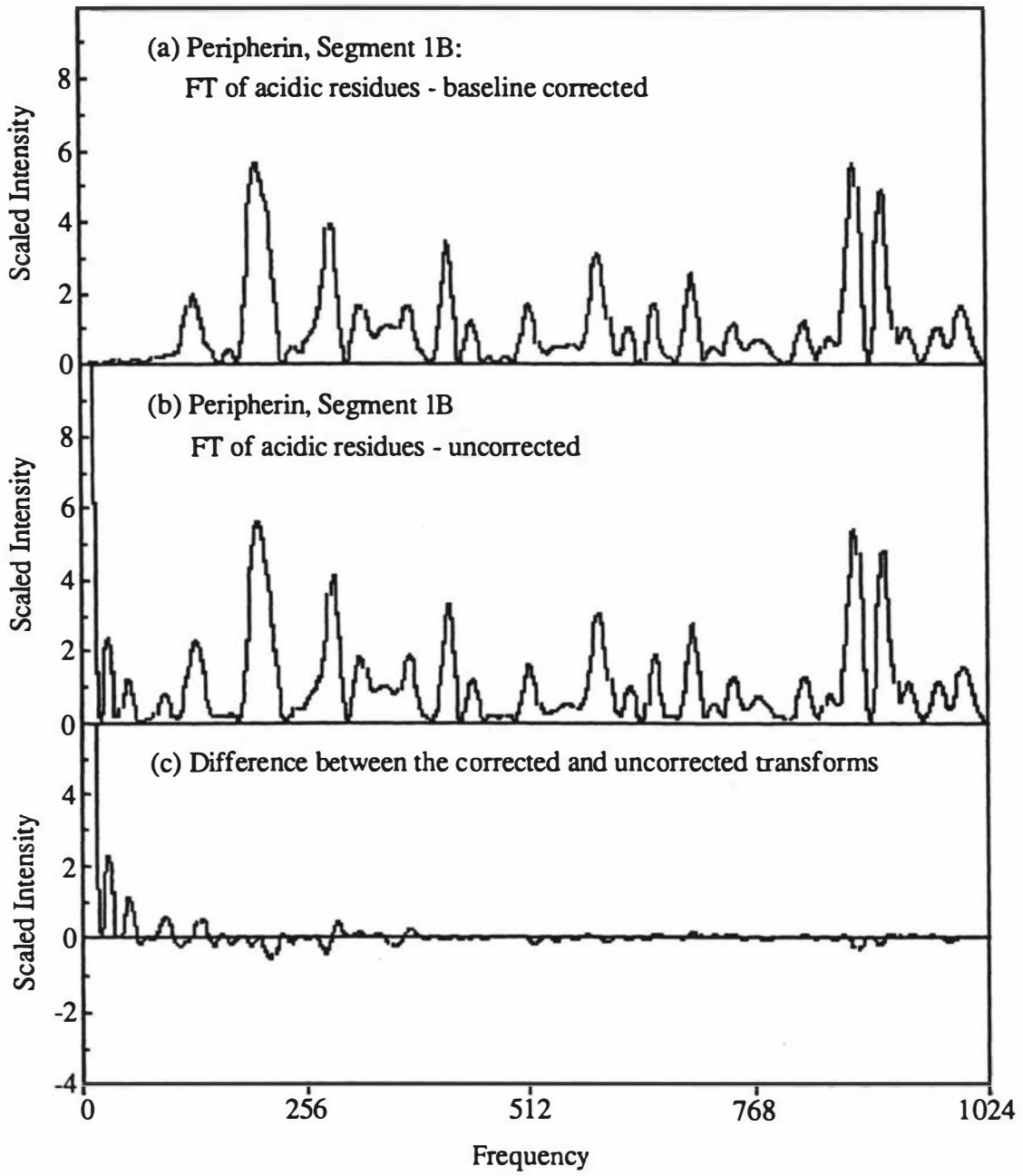


Figure 2-1 Comparison of the scaled Fourier intensities of (a) the baseline-corrected and (b) raw data. The differences between the Fourier intensities of (a) and (b) are also shown (c).

2.1.2 Fourier Analysis - Results

The scaled Fourier intensities for acidic and basic residues in rod domain segments of peripherin, *Xenopus* lamins A and B and the *Helix pomotia* B protein are shown in Appendix B, Figures B-1 to B-5, and a selection of the peaks are listed in Tables B-1 to B-5. The transform for human lamin A, which has been analysed previously by Parry *et al* (1986), has been included to provide a comparison with the other sequences showing an extended 1B segment.

The peripherin data display strong peaks corresponding to a 9.6–9.9 residue period in both the acidic and the basic residues in segments 1B and 2. In the case of the 1B segment, however, the period for the acidic residues appears to have merged with a nearby peak of slightly longer period to produce a single, wider peak (Appendix B, Figure B-1). Comparison with the results obtained for some other type III proteins (chicken gizzard desmin, hamster eye lens vimentin and mouse GFAP - Parry and Fraser, 1985) shows a high degree of similarity (Table 2-2) and provides further support for the assignment of peripherin as a type III IF protein. The acidic residues in segments 1B also reveal the second and third orders of the heptad period (3.47 and 2.33 residues respectively) which is to be expected since charged residues are located predominantly in the b, c, e, f and g positions of the heptad but rarely in positions a and d. The heptad repeat is interrupted near the middle of segment 2B and also in link segment L2 and consequently no significant orders of this period are observed in the transforms of either the acidic or basic residues in segment 2. This is generally true of all short-range (but not necessarily long-range) periods in sequences containing heptad stutters in otherwise regular patterns.

	Peripherin		Other Type III IF	
	Period	Intensity	Period	Intensity
Segment 1B				
Acidics	–	–	9.63 ± 0.21	5.62 ± 1.41
Basics	9.62	5.64	9.62 ± 0.05	4.39 ± 0.40
Segment 2				
Acidics	9.75	7.39	9.76 ± 0.05	4.40 ± 2.10
Basics	9.89	7.36	9.82 ± 0.02	6.05 ± 0.34

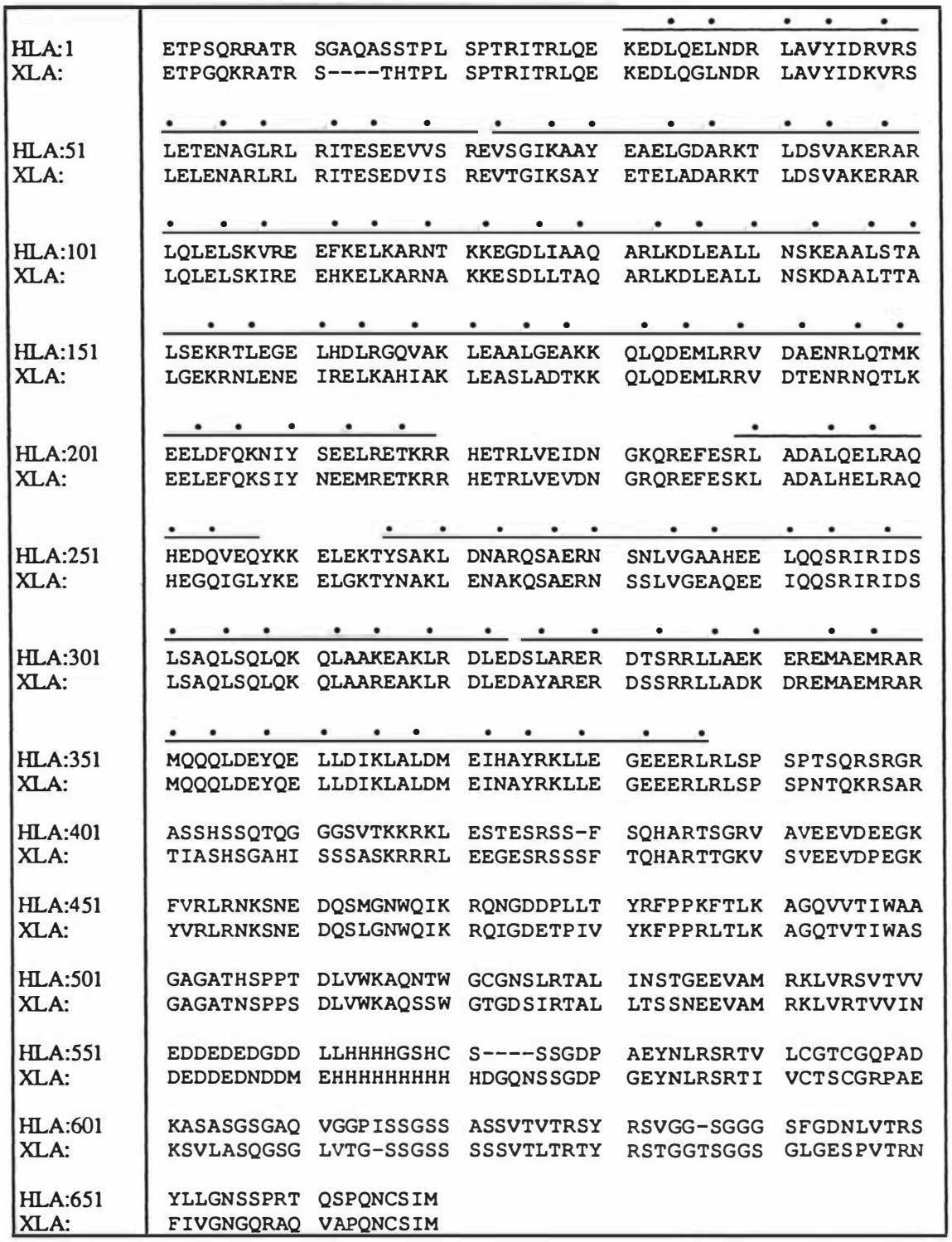
Table 2-2 Comparison of the dominant period in the linear distribution of the acidic and the basic residues for segments 1B and 2 of rat peripherin and some other type III IF proteins (Parry and Fraser, 1985). The ~9.6 residue period for the acidic residues in segment 1B of peripherin has merged with a nearby peak to produce a single broad peak (Appendix B, Figure B-1) and so no exact value can be determined for it.

Xenopus lamin A and B and *Helix pomotia* B proteins share the extended 1B segment which was first observed in the human lamin A and C proteins. The Fourier transforms of charged residues from the rod domain segments of these proteins are shown in Figures B-3 and B-4 with the human lamin A protein providing a basis for comparison (Figure B-2; also Parry *et al*, 1986). The similarities between the transforms of the human lamin A and *Xenopus* lamin A proteins are striking. Examination of the dispositions of the charged residues in the rod segments of the two proteins shows that they are nearly identical although some difference does occur in the distribution of the acidic residues in segment 2. Extensive homology is apparent throughout the primary sequences (Krohne *et al*, 1987): the subdomains of the rod are identical in length and 83% of the residues are identical within the rod with a further 7% conserved in character (see Figure 2-2).

An interesting feature of the transforms of human lamin A that was not noted previously (Parry *et al*, 1986) is an apparent ~19.86 period which can be inferred from peaks corresponding to periods of 9.94 residues (≈19.86÷2) and 6.61 residues (≈19.86÷3). This pair is particularly strong for the acidic residues in segment 1 and in the rod domain as a whole but is notably absent from the basic residues in segment 2. The *Xenopus* lamin A protein also shows this feature but not the *Xenopus* lamin B protein. No evidence for a ~19.86 period in the distributions of charged residues has been reported for other IF proteins in either of the two major rod domain segments (Parry and Fraser, 1985). The significance (or otherwise) of this observation is not apparent.

Several of the transforms of charged residues in both human and *Xenopus* lamin A show peaks corresponding to orders of the heptad substructure (7, 3.5, and 2.33 residue periods). In addition, the basic residues of segment 2 and the rod domain as a whole reveal two strong peaks corresponding to the first two orders of a ~5.06 residue repeat. The segment 1 transforms also show strong peaks corresponding to periods of 3.10 residues (acidics) and 3.02 residues (basics). Neither the ~5.06 or 3.10 and 3.02 periods falls on any orders of the ~19.86 repeat (19.86÷4=4.97 and 19.86÷6=3.33) and no explanation of the significance of these periods is currently possible.

Xenopus lamin B does not reveal a ~19.86 repeat but shows the 9-10 residue repeat common to all IF. In the acidic residues this period is 9.85 residues and in the basic residues it varies from 9.23 residues (segment 1) to 9.85 (segment 2). The 3.03 residue repeat apparent in the segment 1 basics of the lamin A proteins is also strongly represented in *Xenopus* lamin B and falls close to the third order of the 9.23 residue repeat (9.23÷3=3.08).



```
                                                                  . .   . . .
HLA:1     ETPSQRRATR SGAQASSTPL SPTRITRLQE KEDLQELNDR LAVYIDRVRS
XLA:      ETPGQKRATR S----THTPL SPTRITRLQE KEDLQGLNDR LAVYIDKVRS

          . . .  . . .  . . .  . . .  . . .  . . .  . . .
HLA:51    LETENAGLRL RITESEEVVS REVSGIKAAY EAELGDARKT LDSVAKERAR
XLA:      LELENARLRL RITESEDVIS REVTGIKSAY ETELADARKT LDSVAKERAR

          . . .  . . .  . . .  . . .  . . .  . . .  . . .
HLA:101   LQLELSKVRE EFKELKARNT KKEGDLIAAQ ARLKDLEALL NSKEAALSTA
XLA:      LQLELSKIRE EHKELKARNA KKESDLLTAQ ARLKDLEALL NSKDAALTTA

            . .  . . .  . . .  . . .  . . .  . . .  . . .
HLA:151   LSEKRTLEGE LHDLRGQVAK LEAALGEAKK QLQDEMLRRV DAENRLQTMK
XLA:      LGEKRNLENE IRELKAHIAK LEASLADTKK QLQDEMLRRV DTENRNQTLK

          . . .  . . .                                 .  . . .
HLA:201   EELDFQKNIY SEELRETKRR HETRLVEIDN GKQREFESRL ADALQELRAQ
XLA:      EELEFQKSIY NEEMRETKRR HETRLVEVDN GRQREFESKL ADALHELRAQ

          . .       .  . . .  . . .  . . .  . . .  . . .
HLA:251   HEDQVEQYKK ELEKTYSAKL DNARQSAERN SNLVGAAHEE LQQSRIRIDS
XLA:      HEGQIGLYKE ELGKTYNAKL ENAKQSAERN SSLVGEAQEE IQQSRIRIDS

          . . .  . . .  . . .  . . .  . . .  . . .  . . .
HLA:301   LSAQLSQLQK QLAAKEAKLR DLEDSLARER DTSRRLLAEK EREMAEMRAR
XLA:      LSAQLSQLQK QLAAREAKLR DLEDAYARER DSSRRLLADK DREMAEMRAR

          . . .  . . .  . . .  . . .  . . .
HLA:351   MQQQLDEYQE LLDIKLALDM EIHAYRKLLE GEEERLRLSP SPTSQRSRGR
XLA:      MQQQLDEYQE LLDIKLALDM EINAYRKLLE GEEERLRLSP SPNTQKRSAR

HLA:401   ASSHSSQTQG GGSVTKKRKL ESTESRSS-F SQHARTSGRV AVEEVDEEGK
XLA:      TIASHSGAHI SSSASKRRRL EEGESRSSSF TQHARTTGKV SVEEVDPEGK

HLA:451   FVRLRNKSNE DQSMGNWQIK RQNGDDPLLT YRFPPKFTLK AGQVVTIWAA
XLA:      YVRLRNKSNE DQSLGNWQIK RQIGDETPIV YKFPPRLTLK AGQTVTIWAS

HLA:501   GAGATHSPPT DLVWKAQNTW GCGNSLRTAL INSTGEEVAM RKLVRSVTVV
XLA:      GAGATNSPPS DLVWKAQSSW GTGDSIRTAL LTSSNEEVAM RKLVRTVVIN

HLA:551   EDDEDEDGDD LLHHHHGSHC S----SSGDP AEYNLRSRTV LCGTCGQPAD
XLA:      DEDDEDNDDM EHHHHHHHHH HDGQNSSGDP GEYNLRSRTI VCTSCGRPAE

HLA:601   KASASGSGAQ VGGPISSGSS ASSVTVTRSY RSVGG-SGGG SFGDNLVTRS
XLA:      KSVLASQGSG LVTG-SSGSS SSSVTLTRTY RSTGGTSGGS GLGESPVTRN

HLA:651   YLLGNSSPRT QSPQNCSIM
XLA:      FIVGNGQRAQ VAPQNCSIM
```

Figure 2-2 Protein sequences for human lamin A (HLA) and *Xenopus* lamin A (XLA) – for sources, see Table 2-1. The heptad-containing regions of the rod domain are indicated above the sequences by a line. Several blanks have been inserted to maintain optimal alignment of the sequences. The internal a and d positions of the heptad are marked above the sequences by '•'.

The *Helix pomotia* B protein reveals a very strong peak corresponding to a 9.23 residue repeat in the disposition of basic residues in the rod domain (Figure B-5). This is slightly less than the 9.31 and 9.35 residue periods of segments 1 and 2 respectively. The acidic residues of the rod domain also show a dominant peak corresponding to a 9.27 residue period. Another strongly represented period in the basic residues is a 12.19 residue repeat which is apparent in both segment 1 and the whole rod domain.

The dominant 9.23-9.35 residue repeats reported above for the *Xenopus* lamin B and *Helix* B proteins are similar to the period resolved in the acidic residues of segment 1B of the type I keratins (9.28±0.02: Parry and Fraser, 1985). These periods are significantly less than the 9.5-10 residue periods found in all other transforms undertaken here and by Parry and Fraser (1985). Weber *et al* (1988) located the *Helix pomotia* A and B proteins in cells that also expressed keratin IF and suggested that the A and B proteins were not localized in the nucleus as are the structurally similar lamin proteins. If the *Helix* B protein coexists in the cytoplasm with keratin, then it is possible that the common ~9.23 period in charged residues may allow coaggregation although there is no evidence yet to support this idea. Similarly, the lamin B protein, although sited within the nuclear envelope, is postulated as a nuclear anchorage site for the cytoplasmic IF network (Gerace and Burke, 1988). It may be of significance that the ~9.25 residue period is present in the distribution of acidic residues of segment 1B in keratins and in the basic residues of segment 1 in the *Xenopus* lamin B protein.

The six transforms for each of the type V proteins have been multiplied together to highlight periods that are commonly represented in each. The results are summarized in Table 2-3. A low intensity period of 11.64 residues is revealed for the human and *Xenopus* lamin A proteins in addition to the dominant 9.94 residue period already discussed in the human and *Xenopus* lamins. A similar operation on the four

Human lamin A	*Xenopus* lamin A	*Xenopus* lamin B	*Helix pomotia* B
11.64 (452)	11.64 (136)		12.19 (166)
			10.14 (228)
9.94 (2938)	9.94 (1702)	9.89 (3198)	9.25 (11932)
	2.77 (117)		2.84 (113)
2.52 (142)	2.54 (103)		
	2.47 (290)		
	2.022 (100)	2.022 (320)	

Table 2-3 Periods corresponding to major peaks (ie, scaled intensities greater than 100) resulting from multiplying the six Fourier transforms for each protein together. The intensities are shown in brackets.

transforms of peripherin yields a single peak (intensity 1130) corresponding to a repeat of 9.85 residues.

2.2 Residue Distribution in the Heptad

The seven residue quasi-repeat, or heptad substructure, that is found in major portions of the primary sequences of α-fibrous proteins is fundamental to the coiled-coil structure. Apolar residues are commonly found at spacings of three and four residues successively and give rise to a hydrophobic stripe that winds in a left-handed sense around the axis of a right-handed α-helix. Pairs of α-helices suffer a small distortion that allows the stripes to be aligned along the interior axis of a supercoil thus shielding them from the aqueous environment. The apolar residues interlock in a 'knob-into-hole' packing that provides much stability to the coiled-coil structure (see Chapter 1). These internal positions of the coiled-coil in IF are, on average, occupied at a rate of 75% by apolar residues (Parry and Fraser, 1985). The adjacent e and g positions are frequently occupied by charged residues: in IF proteins the e positions are more commonly basic, or positively charged and the g positions are often acidic, or negatively charged. This arrangement specifies the relative alignment of α-helices prior to the precise docking along the apolar stripes: the oppositely charged e and g stripes from different α-helices are adjacent when the chains are parallel.

The distribution of residues in the major heptad-containing portions of IF protein sequences is presented in Table 2-4a. The data reveal that (i) acidic residues are very uncommon in position a and relatively uncommon in d, and that such residues occur with approximately equal frequency in positions b, c, e and g, (ii) the basic residues occur very rarely in d though they may occur in a, and are favoured in e relative to g, and (iii) the apolar residues are most highly maintained in positions a and d, are found comparably in positions e, f and g and are least common in positions b and c. Table 2-4a shows other points of interest, some of which have been noted previously (Parry, 1982; Parry and Fraser, 1985). For example, certain residues occur with greatest frequency in the outer positions of the coiled-coil (b, c and f); these include glycine (80%), serine (76%), aspartic acid (73%), cysteine (63%), alanine (61%) and asparagine (56%). In general, such residues (excluding aspartic acid) are not normally associated with the ability to specify intermolecular association.

Residues occurring with greatest frequency in the innermost a and d positions are tyrosine (86%), leucine (74%), isoleucine (72%), tryptophan (70%), valine (52%), methionine (50%) and histidine (48%). In some cases there is an asymmetric distribution between the a and d positions thus emphasizing the structural uniqueness of each position. Examples include isoleucine (a-81%; d-19%), valine (a-64%;

	Residue count									Percentage Occurence						
	a	b	c	d	e	f	g	Totals		a	b	c	d	e	f	g
Ala	76	263	171	123	66	200	143	1042	Ala	4.8	17.0	10.4	7.5	4.0	12.2	8.9
Cys	9	0	23	8	13	34	3	90	Cys	0.6	0.0	1.4	0.5	0.8	2.1	0.2
Asp	5	148	198	2	66	154	108	681	Asp	0.3	9.5	12.1	0.1	4.0	9.4	6.7
Glu	9	294	282	118	348	200	382	1633	Glu	0.6	19.0	17.2	7.2	21.2	12.2	23.9
Phe	60	1	47	79	13	0	4	204	Phe	3.8	0.1	2.9	4.8	0.8	0.0	0.2
Gly	16	38	121	17	20	78	6	296	Gly	1.0	2.5	7.4	1.0	1.2	4.8	0.4
His	27	17	24	42	3	18	12	143	His	1.7	1.1	1.5	2.6	0.2	1.1	0.7
Ile	334	14	12	80	55	51	31	577	Ile	20.9	0.9	0.7	4.9	3.4	3.1	1.9
Lys	95	79	50	8	226	116	181	755	Lys	6.0	5.1	3.1	0.5	13.8	7.1	11.3
Leu	406	41	33	717	117	86	112	1512	Leu	25.5	2.6	2.0	43.6	7.1	5.2	7.0
Met	76	31	8	48	34	30	23	250	Met	4.8	2.0	0.5	2.9	2.1	1.8	1.4
Asn	63	69	111	31	101	117	35	527	Asn	3.9	4.5	6.8	1.9	6.2	7.1	2.2
Pro	2	9	3	0	0	0	1	15	Pro	0.1	0.6	0.2	0.0	0.0	0.0	0.1
Gln	8	230	91	43	115	121	228	836	Gln	0.5	14.8	5.6	2.6	7.0	7.4	14.2
Arg	86	137	142	9	262	167	114	917	Arg	5.4	8.8	8.7	0.5	16.0	10.2	7.1
Ser	24	95	193	19	54	150	44	579	Ser	1.5	6.1	11.8	1.2	3.3	9.1	2.7
Thr	15	51	81	37	104	61	95	444	Thr	0.9	3.3	4.9	2.3	6.3	3.7	5.9
Val	147	30	30	81	29	51	67	435	Val	9.2	1.9	1.8	4.9	1.8	3.1	4.2
Trp	6	0	0	8	0	6	0	20	Trp	0.4	0.0	0.0	0.5	0.0	0.4	0.0
Tyr	131	3	19	174	15	2	12	356	Tyr	8.2	0.2	1.2	10.6	0.9	0.1	0.7
Totals	1595	1550	1639	1644	1641	1642	1601	11312	Totals	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Apolar	1154	120	149	1179	263	220	249	3334	Apolar	72.4	7.7	9.1	71.7	16.0	13.4	15.6
Basic	181	216	192	17	488	283	295	1672	Basic	11.3	13.9	11.7	1.0	29.7	17.2	18.4
Acidic	14	442	480	120	414	354	490	2314	Acidic	0.9	28.5	29.3	7.3	25.2	21.6	30.6

Table 2-4a The distribution of residues in the heptad-containing regions of IF proteins combined. The symbols a-g represent residue positions within the heptad. Large apolar (Phe, Ile, Leu, Met, Val, Trp, Tyr), basic (Lys, Arg) and acidic (Asp, Glu) residues are summarized at the bottom of the table.

	Residue count									Percentage Occurence						
	a	b	c	d	e	f	g	Totals		a	b	c	d	e	f	g
Ala	32	36	22	75	17	36	35	253	Ala	10.2	11.9	7.0	24.0	5.4	11.5	11.2
Cys	2	0	0	1	0	0	0	3	Cys	0.6	0.0	0.0	0.3	0.0	0.0	0.0
Asp	0	42	45	3	14	27	23	154	Asp	0.0	13.9	14.2	1.0	4.5	8.6	7.3
Glu	1	63	69	21	91	43	60	348	Glu	0.3	20.8	21.8	6.7	29.1	13.7	19.2
Phe	6	2	1	5	1	2	0	17	Phe	1.9	0.7	0.3	1.6	0.3	0.6	0.0
Gly	2	6	8	4	6	14	4	44	Gly	0.6	2.0	2.5	1.3	1.9	4.5	1.3
His	4	12	5	3	4	10	3	41	His	1.3	4.0	1.6	1.0	1.3	3.2	1.0
Ile	35	3	6	16	8	5	7	80	Ile	11.1	1.0	1.9	5.1	2.6	1.6	2.2
Lys	31	50	48	3	20	37	36	225	Lys	9.9	16.5	15.2	1.0	6.4	11.8	11.5
Leu	104	5	11	109	17	9	20	275	Leu	33.1	1.7	3.5	34.8	5.4	2.9	6.4
Met	8	2	2	8	2	3	0	25	Met	2.5	0.7	0.6	2.6	0.6	1.0	0.0
Asn	14	16	13	4	20	11	10	88	Asn	4.5	5.3	4.1	1.3	6.4	3.5	3.2
Pro	0	0	0	0	0	0	0	0	Pro	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Gln	4	20	29	14	59	18	48	192	Gln	1.3	6.6	9.2	4.5	18.8	5.8	15.3
Arg	22	22	27	4	15	47	32	169	Arg	7.0	7.3	8.5	1.3	4.8	15.0	10.2
Ser	7	10	19	6	14	27	16	99	Ser	2.2	3.3	6.0	1.9	4.5	8.6	5.1
Thr	5	9	8	7	17	11	10	67	Thr	1.6	3.0	2.5	2.2	5.4	3.5	3.2
Val	31	3	3	18	8	12	9	84	Val	9.9	1.0	0.9	5.8	2.6	3.8	2.9
Trp	0	0	0	3	0	0	0	3	Trp	0.0	0.0	0.0	1.0	0.0	0.0	0.0
Tyr	6	2	0	9	0	1	0	18	Tyr	1.9	0.7	0.0	2.9	0.0	0.3	0.0
Totals	314	303	316	313	313	313	313	2185	Totals	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Apolar	190	17	23	165	36	32	36	499	Apolar	60.5	5.6	7.3	52.7	11.5	10.2	11.5
Basic	53	72	75	7	35	84	68	394	Basic	16.9	23.8	23.7	2.2	11.2	26.8	21.7
Acidic	1	105	114	24	105	70	83	502	Acidic	0.3	34.7	36.1	7.7	33.5	22.4	26.5

Table 2-4b The distribution of residues in the heptad-containing regions of paramyosin, myosin and tropomyosin combined. Large apolar (Phe, Ile, Leu, Met, Val, Trp, Tyr), basic (Lys, Arg) and acidic (Asp, Glu) residues are summarized at the bottom of the table.

	Total Counts									Average Percentage Occurance						
	a	b	c	d	e	f	g	Totals		a	b	c	d	e	f	g
Ala	108	299	193	198	83	236	178	1295	Ala	7.5	14.4	8.7	15.7	4.7	11.8	10.1
Cys	11	0	23	9	13	34	3	93	Cys	0.6	0.0	0.7	0.4	0.4	1.0	0.1
Asp	5	190	243	5	80	181	131	835	Asp	0.2	11.7	13.2	0.5	4.2	9.0	7.0
Glu	10	357	351	139	439	243	442	1981	Glu	0.4	19.9	19.5	6.9	25.1	13.0	21.5
Phe	66	3	48	84	14	2	4	221	Phe	2.8	0.4	1.6	3.2	0.6	0.3	0.1
Gly	18	44	129	21	26	92	10	340	Gly	0.8	2.2	5.0	1.2	1.6	4.6	0.8
His	31	29	29	45	7	28	15	184	His	1.5	2.5	1.5	1.8	0.7	2.1	0.9
Ile	369	17	18	96	63	56	38	657	Ile	16.0	0.9	1.3	5.0	3.0	2.4	2.1
Lys	126	129	98	11	246	153	217	980	Lys	7.9	10.8	9.1	0.7	10.1	9.4	11.4
Leu	510	46	44	826	134	95	132	1787	Leu	29.3	2.1	2.7	39.2	6.3	4.1	6.7
Met	84	33	10	56	36	33	23	275	Met	3.7	1.3	0.6	2.7	1.4	1.4	0.7
Asn	77	85	124	35	121	128	45	615	Asn	4.2	4.9	5.4	1.6	6.3	5.3	2.7
Pro	2	9	3	0	0	0	1	15	Pro	0.1	0.3	0.1	0.0	0.0	0.0	0.0
Gln	12	250	120	57	174	139	276	1028	Gln	0.9	10.7	7.4	3.5	12.9	6.6	14.8
Arg	108	159	169	13	277	214	146	1086	Arg	6.2	8.0	8.6	0.9	10.4	12.6	8.7
Ser	31	105	212	25	68	177	60	678	Ser	1.9	4.7	8.9	1.5	3.9	8.9	3.9
Thr	20	60	89	44	121	72	105	511	Thr	1.3	3.1	3.7	2.2	5.9	3.6	4.6
Val	178	33	33	99	37	63	76	519	Val	9.5	1.5	1.4	5.3	2.2	3.5	3.5
Trp	6	0	0	11	0	6	0	23	Trp	0.2	0.0	0.0	0.7	0.0	0.2	0.0
Tyr	137	5	19	183	15	3	12	374	Tyr	5.1	0.4	0.6	6.7	0.5	0.2	0.4
Totals	1909	1853	1955	1957	1954	1955	1914	13497	Totals	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Apolar	1344	137	172	1344	299	252	285	3833	Apolar	66.4	6.7	8.2	62.2	13.8	11.8	13.5
Basic	234	288	267	24	523	367	363	2066	Basic	14.1	18.8	17.7	1.6	20.5	22.0	20.1
Acidic	15	547	594	144	519	424	573	2816	Acidic	0.6	31.6	32.7	7.5	29.4	22.0	28.6

Table 2-4c (Left) The distribution of residues in the heptad containing regions of the IF and myosin-type proteins used in Tables 2-1a and 2-1b. (Right) Average percentage distribution for the IF and myosin-type proteins used in Tables 2-1a and 2-1b. Large apolar (Phe, Ile, Leu, Met, Val, Trp, Tyr), basic (Lys, Arg) and acidic (Asp, Glu) residues are summarized at the bottom of the table.

d-36%), methionine (a-61%; d-39%), histidine (a-39%; d-61%) and leucine (a-36%; d-64%).

Another feature of Table 2-4a is the preponderance of leucine residues in the d position where leucine occurs four times as often as any other residue and makes up 44% of the total. Leucine and isoleucine contribute 26% and 21% respectively to the residues found in the a position. The average occurrence of large apolar residues in the a and d positions is 72%, a figure in close agreement with that determined by Parry and Fraser (1985). Including the alanines in the apolar group raises this figure to 78%.

The coiled-coil rod regions of several myosin-type sequences (myosin, paramyosin and tropomyosin) have also been analysed (Table 2-4b) for comparison with the IF proteins. The occurrence of large apolar residues in the a and d positions is not as great as for IF proteins (57%) but this figure becomes comparable when alanines are included (74%). Interestingly, acidic residues clearly dominate the e position for the myosin-type proteins (33.5% acidic: 11.2% basic) but the g position for the IF proteins (29.8%: 18.6%).

When the "myosins" and IF proteins are combined into a single table (Table 2-4c) the dominance of leucine is again apparent in the d position of the heptad and together with isoleucine it dominates the a position. Acidics and basics are now more evenly matched in e but acidics remain more common in g. It is interesting to note that some of the less common residues are excluded from certain positions within the heptad: cysteine is never found in b; proline (an α-helix-disrupting residue) is never found in d, e, or f; and tryptophan is only found in a, d, and f.

This study emphasizes the importance of apolar residues in specifying the heptad substructure and hence the stability of the coiled-coil. In particular, isoleucine and leucine are especially common in the a position of the heptad and leucine is dominant in the d position. Charged residues are found fairly evenly amongst the remaining, more external positions: those in the e and g positions are involved in specifying the relative orientations of the chains that form the dimeric molecule while the others are on the exterior of the coiled-coil and specify higher orders of aggregation.

2.3 Flexibility

Prediction schemes are used to characterize the attributes of protein chains where no direct method of measurement is readily available. These schemes generally apply a statistically weighted scoring system to a region of protein sequence in an effort to establish the potential for a given characteristic such as hydrophobicity (Rose, 1978;

Kyte and Doolittle, 1982), antigenicity (Hopp *et al*, 1981) or the probability of adopting a particular element of secondary structure (Chou and Fasman, 1974; Garnier *et al*, 1978). A scheme for calculating the relative axial stagger of coiled-coil proteins has met with particular success in predicting the stagger of collagen molecules (and also myosin and paramyosin) and is used in Chapter 4 in a modified form.

Karplus and Schulz (1985) introduced a method for predicting the flexibility of protein chains by using temperature factor data deduced from refined crystallographic structures. Their prime aim was to provide an improved tool for selecting peptide antigens and cross-reacting peptides based on the link between segmental flexibility and antigenic determinacy (Westhof *et al*, 1984; Tainer *et al*, 1984). This method was used by Conway *et al* (1989) and Conway and Parry (1989) to compare the flexibilities of different IF chain types, and also the different segments of the IF chains. However, some care must be taken in interpreting the results from IF protein chains as the method of Karplus and Schulz is based on data collected from globular proteins - the rod domain of IF is clearly not globular although the N- and C-terminal domains may be.

The data base of Karplus and Schulz comprised the temperature factors of the Cα atoms of residues from 31 globular proteins. The averages and ranges of these factors were noted to vary between proteins and this was assumed to be a result of differences in structure refinement methods rather than natural variances. As a consequence, the temperature factors were normalized and the root-mean-square deviation was made constant. The average normalized temperature factor was determined for each residue and the residues were grouped into two classes: average normalized temperature factor >1 ('flexible') or <1 ('rigid'). Separate flexibility indices were determined from the average normalized temperature factors according to whether a residue had zero, one or two 'rigid' neighbours. These indices were used to generate a profile of chain flexibility which was then smoothed with a triangular weighting profile (ie weights $^{1}/_{16}$, $^{2}/_{16}$, $^{3}/_{16}$, $^{4}/_{16}$, $^{3}/_{16}$, $^{2}/_{16}$, $^{1}/_{16}$).

Conway *et al* (1989) and Conway and Parry (1989) assumed that the same method could be applied to the IF chains even though they are not globular proteins. The weighting scheme was modified in order not to disguise effects arising from the relatively short 7 and 10 residue periods known to be present in the rod domain sequences (ie, the heptad substructure and the charged residue periods respectively) and the weights used were $^{1}/_{4}$, $^{2}/_{4}$, $^{1}/_{4}$. Flexibility profiles from a selection of IF chains from all chain types are shown in Figure 2-3 and the average scores (± s.d.) for the chain segments are listed in Table 2-5.

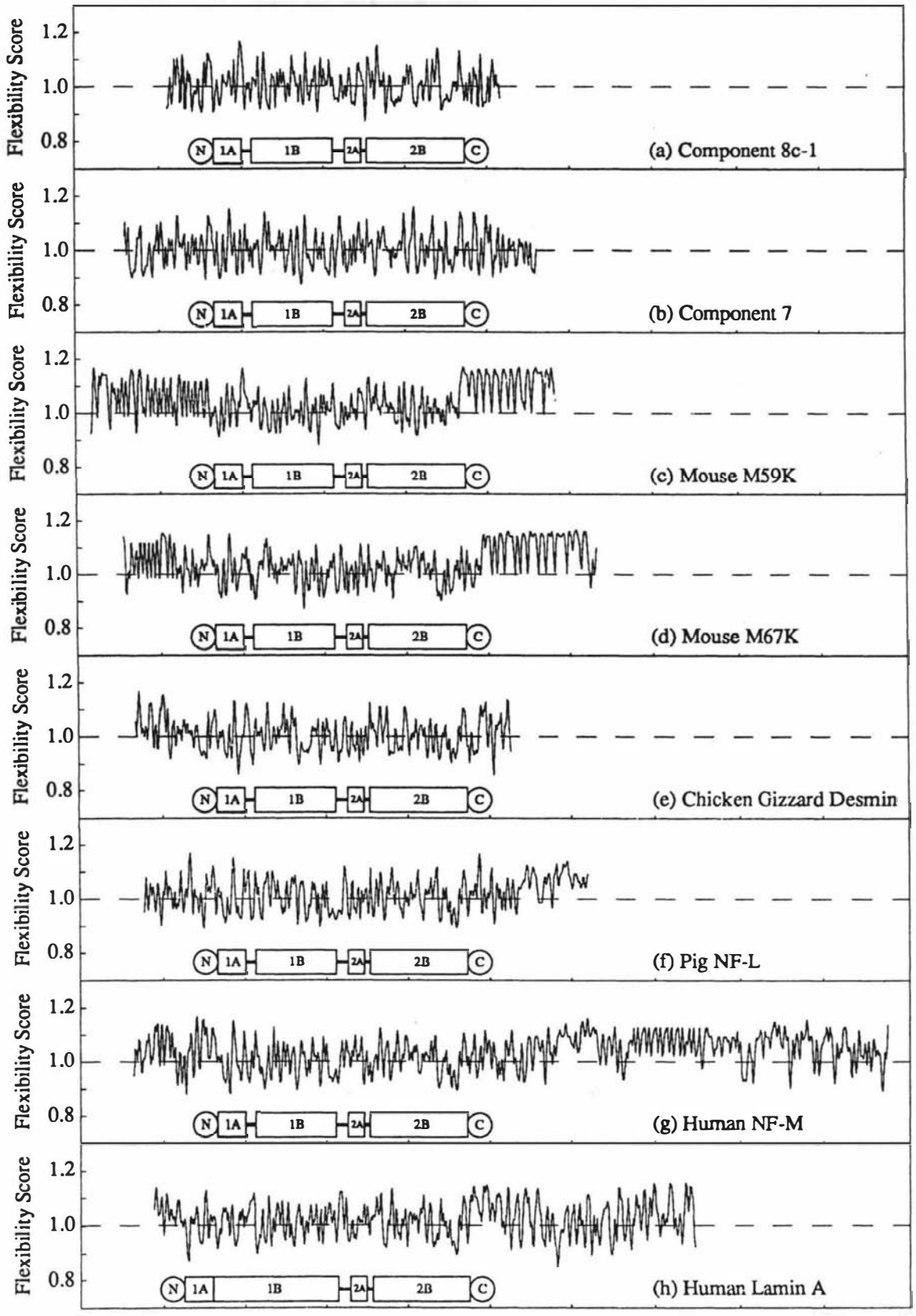


Figure 2-3 Flexibility profiles for a selection of IF chains. The horizontal axis is divided into units of 100 residues. Scores above 1.0 on the vertical axis indicate greater than average flexibility and scores below 1.0 indicate more rigid regions of sequence. Sequences are aligned by the N-terminus of segment 2A.

Segment	Type Ia Comp. 8c-1	Type IIa Comp. 7	Type Ib M59K	Type IIb M67K	Type III CGD	Type IV-L PNF-L	Type IV-M HNF-M	Type V Lamin A
E1 and V1	?	?	1.064±0.082	1.061±0.081	1.037±0.074	—	—	—
H1	?	?	1.060±0.078	1.031±0.071	0.995±0.080	1.016±0.073	1.044±0.083	1.053±0.061
1A	1.004±0.064	1.002±0.087	1.000±0.070	1.009±0.078	0.982±0.074	0.989±0.077	0.998±0.080	1.007±0.073
L1	1.062±0.107	1.014±0.092	1.072±0.070	1.050±0.053	1.064±0.074	1.058±0.074	1.003±0.085	—
1B	1.010±0.073	0.999±0.074	1.003±0.068	1.012±0.068	1.001±0.068	1.005±0.072	1.000±0.071	1.011±0.067
L12	0.990±0.049	1.011±0.089	1.002±0.068	1.014±0.070	0.983±0.070	0.976±0.079	0.988±0.074	1.041±0.075
2A	0.987±0.061	0.984±0.071	1.005±0.052	0.989±0.074	0.979±0.072	0.986±0.074	1.038±0.059	0.987±0.050
L2	1.032±0.061	1.027±0.077	1.046±0.061	1.033±0.081	0.976±0.060	1.002±0.046	0.967±0.073	1.040±0.052
2B	1.003±0.074	1.008±0.075	1.024±0.069	1.015±0.069	1.001±0.070	1.006±0.072	1.007±0.068	1.007±0.071
H2	?	?	—	1.017±0.056	1.015±0.085	1.030±0.067	1.021±0.069	1.025±0.083
V2 and E2	?	?	1.110±0.064	1.105±0.067	—	1.079±0.044	1.062±0.066	1.041±0.081
N	1.003±0.084	1.002±0.076	1.063±0.082	1.051±0.079	1.028±0.077	1.016±0.073	1.044±0.083	1.053±0.061
Rod	1.007±0.072	1.004±0.077	1.015±0.069	1.013±0.072	0.998±0.070	1.003±0.073	1.003±0.071	1.010±0.069
C	0.998±0.079	1.000±0.074	1.110±0.064	1.094±0.072	1.015±0.085	1.047±0.064	1.055±0.068	1.025±0.082

Table 2-5 Mean flexibility indices (± s.d.) for chain segments from all IF Types. Sources are shown in Table 3-1. Abbreviations are as follows: Comp. 8c-1, Component 8c-1; Comp. 7, Component 7; M59K, Mouse 59K; M67K, Mouse 67K; CGD, Chicken Gizzard Desmin; PNF-L, Pig neurofilament light chain; HNF-M, Human neurofilament medium chain; ?, extent of the segment has not been determined (see Chapter 3); —, segment is not present in the sequence.

A striking feature of the flexibility profiles is the regular pattern shown in the terminal domains of the epidermal keratins M59K and M67K and also in parts of the C-terminal domain of human NF-M. The calculated flexibility values are particularly high for these 'soft' keratin chains in the V1 and V2 subdomains and this is consistent with these regions being flexible, interactive and having an external location on the surface of the IF. Indeed a high degree of flexibility as well as strength is required in the epidermis which forms the barrier between an animal and its environment. The C-terminal domains of neurofilament chains are part of an extended bridge between IF and microtubules and may also be seen to require a degree of flexibility.

The periodicities present in the flexibility profiles of the V1 and V2 subdomains of the epidermal keratins include a 5-residue period and a quasi-halved 9-residue period in the V1 subdomain of M59K, a 10-residue period in the V2 subdomain of M59K, a 5-residue period in the V1 subdomain of M67K, and a 28-residue period in the V2 subdomain of M67K. These periods and others (see Steinert *et al*, 1983a, 1985a) are in each case associated with a glycine-serine rich structural motif. The regular pattern observed in the middle of the C-terminal domain of human NF-M is due to a 13-residue repeat (KSPVEEKGKSPVP).

Average flexibility scores for the segments are shown in Table 2-5. Segments 1A, 1B, and 2B are typically close to unity as is the rod domain as a whole. The H1 and H2 domains are generally more flexible than the rod domains (with the exception of H1 in the chicken gizzard desmin sequence) and the End and Variable domains combined (ie E1 and V1, E2 and V2) are the most flexible. The hard α-keratin chains, however, have mean flexibility indices close to unity over their entire N- and C-terminal domains and this may be an indication that these regions display a considerable degree of intra-subtype homology (see also Crewther *et al*, 1985, and Sparrow *et al*, 1989). Furthermore, their particular scores are comparable to those for the entire rod domains of all the chains studied and suggest that the terminal domains of the hard α-keratins are better defined structurally than most other IF chains.

The rod domain segment predicted to have the highest mean flexibility is generally the link, L1. This is consistent with the variable length of this link which can differ even within a single chain type and which is actually non-existent in type V chains. Almost certainly the structure of L1 is poorly defined.

Segment 2A is predicted to have (marginally) the lowest flexibility while the adjacent link segment L2 is predicted to have the second highest within the rod domain (except for component 7 from wool α-keratin where it is the highest). Interestingly, the pig NF-L sequence shows the reverse of this trend: segment 2A has the highest mean

flexibility score and segment L2 has the lowest. These segments are of particular interest as they represent the major portion of a region where the coiled-coil undergoes a rearrangement caused by the discontinuity, or stutter, in the heptad phasing that occurs in segment L2. This latter segment, although predicted to be α-helical (see, for example, Geisler and Weber, 1982; and Parry and Fraser, 1985), lacks the heptad repeat characteristic of coiled-coil conformations found in the α-fibrous protein class of structures. The flexibility scores indicate that both chains of the IF molecule are flexible in the region of L2 and it is possible that such flexibility may provide a means by which the integrity of the coiled-coil structure in segments 2A and 2B can be maintained whilst allowing the coherence of segment 2 across the stutter to be re-established in a gradual manner. The differences between flexibility profiles in this region may indicate alternative conformations of the two chains around the stutter.

The flexibility profiles of chains comprising a heterodimeric keratin molecule are not in phase in the vicinity of the second heptad stutter found close to the centre of segment 2B in all IF chains. This may indicate that the heptad phasing is re-established at this point by a relatively sharp discontinuity in the molecular structure, possibly involving a kink in the axis of the coiled-coil. The nature of the discontinuity suggested here is in contrast to that postulated above for the stutter in segment L2.

A feature of the flexibility profiles is that regions in the rod domain of high flexibility are frequently bounded by regions of particularly low flexibility. It follows that a high local flexibility score should not be interpreted in terms of a breakdown in coiled-coil structure but instead as a short region of structure with a greater propensity for interaction with other IF molecules or perhaps intermediate filament associated proteins (IFAPs). Indeed many of the peaks can be correlated with the presence of basic residues (predominantly lysine) and acidic residues. These charged residues are intimately involved in the interactions which specify both molecular assembly and aggregation *in vivo* (see Chapter 1).

The lamin chain shows a higher flexibility score for the N-terminal domain than the C-terminal domain, unlike most of the other sequences studied. The N-terminal domain of lamin is amongst the smallest of any of the IF chains whereas the C-terminal domain of lamin A is one of the largest after the NF-H chains.

In summary, some of the features of the flexibility profiles can be related to the sequences of the chains, such as the glycine-serine rich area in the terminal domains of the epidermal keratins, and a tentative link to function can also be proposed, viz. the flexibility of skin. Generalizations of the profile data made above always have exceptions and these may be the result of a link between predicted flexibility and the

function of individual IF. Indeed, although IF share a common plan for the rod domain, they are found in a wide range of cell types and hence it is to be expected that the variability present in the terminal domains in particular, as well as in the less well defined regions of the rod domain, is a result of differing function.

2.4 Summary

A variety of structural studies based on IF protein sequences have been presented in this Chapter. These include an examination of the periodicities in the rod domains, the distribution of amino acid residues in the heptad substructure and the flexibility of the peptide backbone. Fourier techniques were used to investigate regularities in the disposition of charged residues in the rod domains of peripherin, the nuclear lamin proteins, and the *Helix pomotia* B protein. The Fourier data for peripherin showed a high degree of similarity with that from other type III IF chains and confirm the assignment of peripherin to the type III grouping. The close homology between the human and *Xenopus* lamin A proteins is reflected in the Fourier transforms of their charged residues. A novel feature of these transforms is an apparent doubling of the 9-10 residue period found commonly in IF chains; this is associated with the larger scale packing of the molecules (compared to the short range heptad structure which is related directly to the mode of chain packing).

The transforms of sequences from the *Xenopus* lamin B rod domain segments reveal a period of ~9.3 residues in the disposition of basic residues in segment 1B. This is a little shorter than generally found in the same segments in other IF chains although it is characteristic of the period found in acidic residues in segment 1B of type I keratins. The *Helix* protein also reveals high intensity peaks corresponding to a similar period. Lamin B proteins are thought to be nuclear anchorage sites for the cytoplasmic IF network (Georgatos and Blobel, 1987; Gerace and Burke, 1988) and this common shorter period allows the possibility of a direct linkage between the cytoplasmic and nuclear IF networks. *Helix* B type V IF are apparently coexpressed with keratin IF in the cytoplasm (Weber *et al*, 1988) and it is possible that the common period might also allow coaggregation of the two networks although it must be emphasized that there is no evidence yet available to suggest that this actually takes place.

An analysis of the distribution of residues within the heptad substructure shows the preference of certain types of residues in certain positions. Apolar residues are confirmed to occupy ~75% of the a and d positions though the distribution of residues between the two positions is highly non-uniform. For example, leucine is found much more frequently in the d position than the a position but for isoleucine the a position is more commonly occupied than the d. This emphasizes that, although the a and d

positions are similar in that they are internal to the coiled-coil structure and crucial for the formation of the hydrophobic core that stabilizes the molecule, they are nonetheless stereochemically non-equivalent. This difference had also been observed by Phillips *et al* (1986) who correlated the occurrence of apolar residues having branched sidechains in the d position with more flexible regions of coiled-coil or even "hinge" regions in the case of myosin (see also Cohen and Parry, 1989).

The charged residues are also unevenly distributed between the a and d and the e and g positions although of course there are relatively few non-apolar residues in the internal a and d heptad positions. The asymmetric distribution found, however, does enable basic residues in the a and e positions of the same chain to interact with acidic residues in the d and g positions of an adjacent chain (see Figure 1-4) to stabilize the two-chain molecular structure.

An interesting comparison between two families of α-fibrous proteins reveals that for IF proteins the acidic residues are more common than basic residues in the g position of the heptad substructure but that for myosins the acidics dominate in the structurally similar e position. These two heptad positions are important for specifying the (parallel) orientation of the chains in the molecule. In the case of tropomyosin, which has a more pronounced division of charged residues among the two positions (Parry, 1975), the e and g stripes on adjacent chains will attract if the chains are parallel or repel if antiparallel. For IF and the myosin-type proteins in general, the net charge of each position is more even.

The flexibility profiles of representative IF chains has revealed some features that may be identified with structural elements of the molecule. The terminal domains are generally more flexible than the rod domains, especially for the soft keratins and the neurofilament chains studied. This flexibility may be a functional requirement of these domains (or at least parts of them). The epidermis, for example, is tough yet yielding and this 'soft' character may be attributable in part to the terminal domains of the constituent IF chains and the interactions these make with other IF or IF-associated proteins. In contrast, the hard keratin chains studied are predicted to be less flexible than their epidermal counterparts. Regular patterns of residues in parts of the amino acid sequence are also reflected in the flexibility profiles.

For most IF chains the link region L1 is generally predicted to have the greatest flexibility of the rod domain segments: this is also one of the rod domain regions with the lowest degree of sequence homology. Another interesting observation is the flexibility apparent in the link L2 which also contains a stutter in the phasing of the heptad substructure. The apparent freedom allowed in the conformations of the chains

in this region may allow the integrity of the coiled-coil structures to be maintained over the discontinuity. The stutter towards the centre of segment 2B, however, appears to be accommodated by a more abrupt conformational rearrangement of the coiled-coils.

3. SEQUENCE HOMOLOGY

In general terms, homology may be described as a measure of correspondence or sameness among objects. This concept is useful for identifying groupings of objects with high degrees of homology and may also show the features that are common to a grouping as well as those that differentiate one group from another. In this chapter, homology is defined in terms of the degree of similarity between the linear distribution of the amino acid residues constituting the coiled-coil rod domain of IF protein molecules, although the techniques used here may be applied to the sequences of any family of proteins. The structure of the coiled-coil rod domain is relatively well defined and constant in length compared to the terminal domains of the IF proteins and so sequence comparisons amongst the proteins are more easily carried out in this domain.

Sequence data from a selection of IF proteins (Table 2-1) are used to generate three homology statistics (h_a, h_r, h_s) that describe, in large part, the degree of correspondence amongst groupings of the proteins. In general, mammalian sequences are examined although some parts of this study will compare non-mammalian (usually amphibian proteins) and mammalian proteins. The homology statistics are also used to generate a consensus, or archetypal, sequence for the various IF chain types I-V. The alignment of incomplete rod domain sequences is maintained by using blank positions which are otherwise ignored in the calculation of homology. Blanks are also used in the variable length segments L1 and L12 to maximize homology within these segments as well as to maintain alignment. The positions and numbers of blanks are somewhat arbitrary in L1 and L12 as the degree of homology is relatively low. The total number of residue positions for the rod domain of types I-IV IF has been set to 327; for type V IF chains this has been increased to 359 as a result of the 42 residue insert in segment 1B and the replacement of the link L1 by extensions to segments 1A and 1B for which no padding is necessary.

Two sequences exhibit anomalous lengths for some rod domain segments. The Endo B cytokeratin (M47K - Singer *et al*, 1986) contains 17 residues in link L12 instead of the usual 16 for type I IF chains (Table 1-1). Also, an extra residue (acidic) is inserted within the last heptad of segment 2B which is strictly conserved in length in all other IF chains. For these reasons, this protein has been excluded from the analyses described in this section. The human NF-L chain (Julien *et al*, 1987) has a length of 103 residues for segment 1B, unlike all other type I-IV IF chains. The two extra

residues (a glutamic acid and a threonine residue) are easily detected by comparison with other NF-L chains and in both cases represent a repeat of the residue immediately preceding it. Also, segment L12 is only 16 residues in length unlike all other NF-L and NF-M chains for which the length is 17 residues. Because relatively few NF-L sequences available, a modified sequence has been used where these two repeated residues in segment 1B have been deleted.

3.1 Homology Statistics

Suppose that N sequences are to be compared and that the padded lengths of the sequences is n (n=327 for type I-IV chains; n=369 for type V chains) where the origin is at the amino-terminus of segment 1A. Let $A(i,j)$ denote the amino acid within protein sequence i ($1 \le i \le N$) at a linear position j ($1 \le j \le n$). The various segments of the rod domain for types I-IV IF are defined as follows:

segment 1A	$A(i,1)$	to	$A(i,35)$	35 residues
segment L1	$A(i,36)$	to	$A(i,56)$	21 residues
segment 1B	$A(i,57)$	to	$A(i,157)$	101 residues
segment L12	$A(i,158)$	to	$A(i,179)$	22 residues
segment 2A	$A(i,180)$	to	$A(i,198)$	19 residues
segment L2	$A(i,199)$	to	$A(i,206)$	8 residues
segment 2B	$A(i,207)$	to	$A(i,327)$	121 residues

For the type V sequences, the segment boundaries are:

segment 1A	$A(i,1)$	to	$A(i,41)$	41 residues
segment 1B	$A(i,43)$	to	$A(i,189)$	147 residues
segment L12	$A(i,190)$	to	$A(i,211)$	22 residues
segment 2A	$A(i,212)$	to	$A(i,230)$	19 residues
segment L2	$A(i,231)$	to	$A(i,238)$	8 residues
segment 2B	$A(i,239)$	to	$A(i,359)$	121 residues

The 42-residue insert in segment 1B is part of a larger 70-residue piece at positions 90 to 159 inclusive in the type V chains which replaces residues 99 to 126 of the types I-IV chains (Parry *et al*, 1986). Although this causes misalignment of the type V chains with the type I-IV chains, only homology within chain types is calculated and so the relative alignment of chains of different types is not important.

The stutter in the phasing of the heptad substructure that occurs near the middle of segment 2B necessitates further subdivision of this segment into 2B1 and 2B2 which are bounded for types I-IV IF as follows:

segment 2B1	$A(i,207)$	to	$A(i,264)$	58 residues
segment 2B2	$A(i,266)$	to	$A(i,327)$	62 residues

and for type V:

segment 2B1	$A(i,239)$	to	$A(i,296)$	58 residues
segment 2B2	$A(i,298)$	to	$A(i,359)$	62 residues

A single residue has been designated as the 'locus' of this heptad stutter (residue 265 in types I-IV IF and residue 297 in type V IF). This is a somewhat arbitrary choice as in fact the heptad pattern can be continued through the 'locus' from either direction for a short way and indeed it is not at all apparent whether the stutter is manifested as a sharp discontinuity or as a more extensive but gradual re-ordering of the coiled-coil structure. A similar situation arises for the link L1 in type V chains which is reduced to a stutter in the heptad phasing: in type I-IV chains it is a short non-α-helical segment. As with the segment 2B stutter, a single residue position has been chosen in the type V chains as the centre of the stutter – position 42.

The homology statistics h_a, h_r and h_s are defined below. Note that the homology scores are given as percentages where 100% homology indicates identical sequences.

3.1.1 Amino Acid Homology Score, h_a

Each amino acid is compared in turn with each amino acid in the N sequences sharing a common value of j, including itself. For every comparison a look-up table (for example, Table 3-1) is used to score the degree of homology between the pairs of residues. The basis of the scoring system is as follows: residue pairs having no relationship are scored zero; residue pairs of conservative size and character (eg. Leu-Val, Met-Phe, etc.) are scored 0.5; residue pairs of conservative charge (eg. Lys-Arg or Asp-Glu) are scored 0.75; pairs of identical residues are scored unity. Other look-up tables are used for scoring homology among specific types of residues and are listed in Table 3-2. For each comparison made the scores are added and normalized to generate the homology statistic $h_a(i,j)$. This may be formulated mathematically as:

$$h_a(i,j) = \frac{1}{N} \sum_{k=1}^{N} S[A(i,j):A(k,j)]$$

where $S[A(i,j):A(k,j)]$ is the look-up score for the residue pair $A(i,j)$ and $A(k,j)$. Note that this definition implies a minimum homology score of

$$\frac{1}{N} S[A(i,j):A(i,j)]$$

3.1.2 Residue Homology Score, h_r

This is defined as the largest value of $h_a(i,j)$ at some residue position j, ie.

$$h_r(j) = \max[h_a(i,j)] \text{ for } 1 \le i \le n$$

	Asp	Glu	Arg	Lys	His	Ile	Leu	Met	Phe	Tyr	Val	Ala	Asn	Cys	Gln	Gly	Pro	Ser	Thr	Trp	
Asp	1																				Asp
Glu	0.75	1																			Glu
Arg			1																		Arg
Lys			0.75	1																	Lys
His			0.5	0.5	1																His
Ile						1															Ile
Leu						0.5	1														Leu
Met						0.5	0.5	1													Met
Phe						0.5	0.5	0.5	1												Phe
Tyr						0.5	0.5	0.5	0.75	1											Tyr
Val						0.5	0.5	0.5	0.5	0.5	1										Val
Ala						0.25	0.25	0.25	0.25	0.25	0.25	1									Ala
Asn	0.5												1								Asn
Cys														1							Cys
Gln		0.5													1						Gln
Gly																1					Gly
Pro																	1				Pro
Ser																		1			Ser
Thr																		0.5	1		Thr
Trp									0.5	0.5										1	Trp
	Asp	Glu	Arg	Lys	His	Ile	Leu	Met	Phe	Tyr	Val	Ala	Asn	Cys	Gln	Gly	Pro	Ser	Thr	Trp	

Table 3-1 Look-up table for mixed homology scores S[*A(i,j):A(k.j)*]

		Asp	Glu	Arg	Lys	Ile	Leu	Met	Phe	Tyr	Val
Acidic Homology Table	Asp	1									
	Glu	1	1								
Basic Homology Table	Arg			1							
	Lys			1	1						
Large Apolar Homology Table	Ile					1					
	Leu					1	1				
	Met					1	1	1			
	Phe					1	1	1	1		
	Tyr					1	1	1	1	1	
	Val					1	1	1	1	1	1
		Asp	Glu	Arg	Lys	Ile	Leu	Met	Phe	Tyr	Val

Table 3-2 Look-up tables for Acidic, Basic and Large Apolar homology

and is a reflection of the most common feature of the various sequences being compared at that position. It leads naturally to the idea of a consensus residue and, by extension, to a consensus sequence.

3.1.3 Segment Homology Score, h_s

The value of $h_s(i)$ is the average $h_a(i,j)$ value within an entire sequence or some subsequence (often a coiled-coil or link segment). If m_1 and m_2 are the positions of the first and last residues respectively in some sequence of interest then the $h_s(i)$ score for that subsequence is defined as

$$h_s(i) = \frac{1}{(m_2 - m_1 + 1)} \sum_{j=m_1}^{m_2} h_a(i,j)$$

where $1 \le m_1 < m_2 \le n$. The value $(m_2 - m_1 + 1)$ is the number of residues over which the average was taken. This statistic provides an indication of the divergence of IF proteins and of their appropriate classification within the types I-V categories. This may also be assessed using other homology techniques (see, for example, Crewther *et al*, 1983 and Parry and Fraser, 1985).

3.2 Homology within the Rod Domain

The residue homology scores (h_r) have been calculated using the look-up table specified in Table 3-1 for the mammalian type I-IV sequences and all the available type V sequences (Table 2-1). Homology profiles for each type are shown in Figure 3-1. Prior to graphing, the raw scores have been smoothed by convolution with a Gaussian function of half width two residues (see Appendix C). Regions of these smoothed curves where three or more consecutive residues have residue homology scores

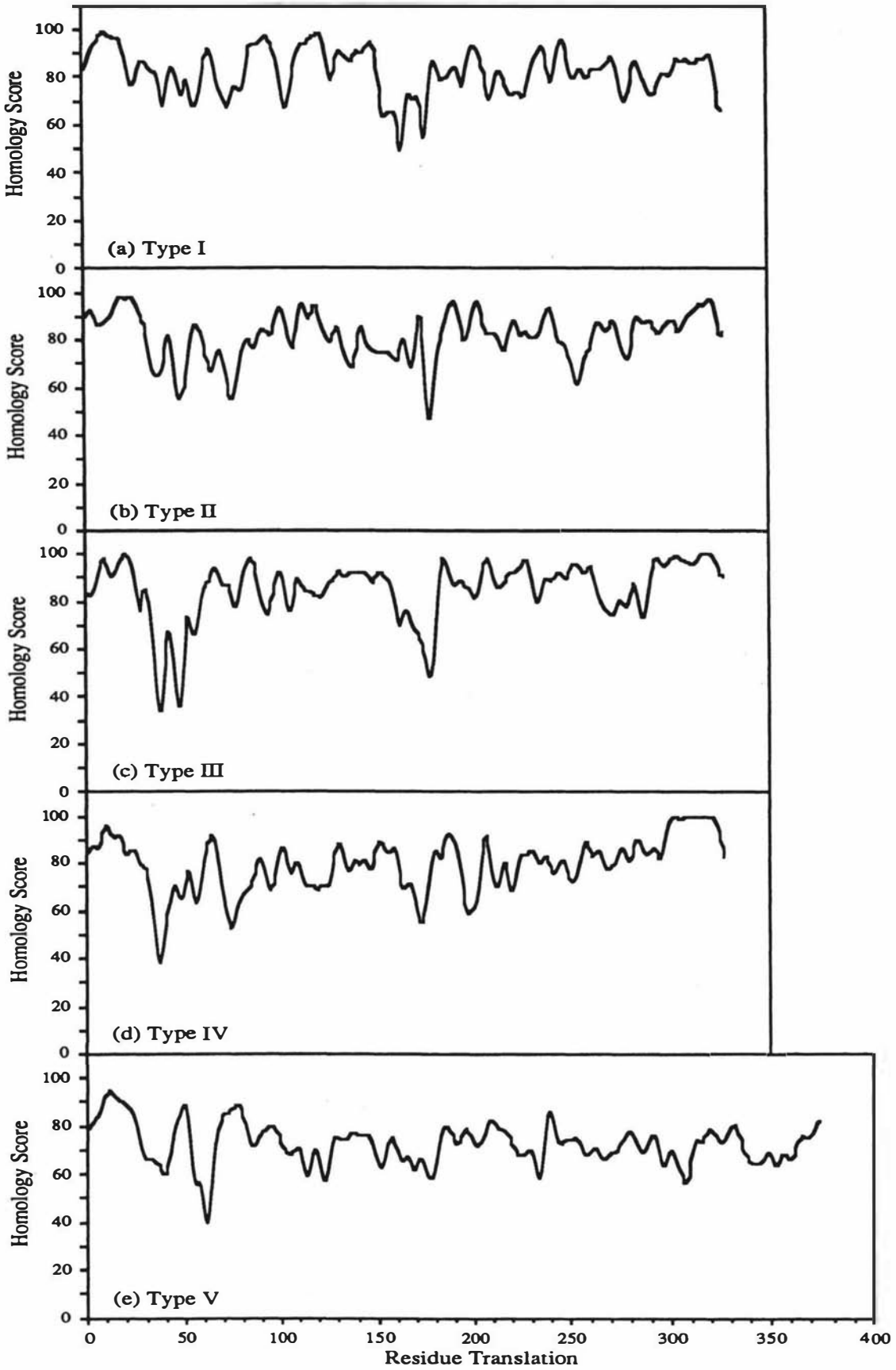


Figure 3-1 Homology profiles for mammalian types I-IV chains and all available type V IF chains. The residue homology scores (h_r) have been smoothed with a Gaussian function of half width two residues. Regions of highly conserved sequence are apparent in segment 1A and at the C-terminal end of segment 2B in the type I-IV profiles.

Segment	Type I	Type II	Type III	Type IV	Type V
1A	4–20	1–4	7–12	8-17	8–15
L1	–	–	–	–	–
1B	63–65	99–101	64–68	63–65	–
	85–97	111–114	80–87	72–76	
	110–124	117–120	129–132		
	140–150		135–145		
			150–154		
L12	–	–	–	–	–
2A	–	187–192	182–187	186–189	–
L2	199–202	200–204	204–206	–	–
2B	234–236	237–241	207–210	298–324	359–367
	244–248	310–323	217–229		
			242–246		
			248–262		
			290–327		

Table 3-3 Regions in the homology profiles (h_r scores) where the scores are greater than or equal to 90%. The regions listed are at least three residues in length and indicate highly conserved primary structure.

$h_r \geq 90\%$ or $h_r < 60\%$ are listed in Tables 3-3 and 3-4 respectively. In addition, the segment homology scores (h_s) have been calculated for the 7 subdomains of the rod region and for the rod domain as a whole. The scores for the chains belonging to the various types were then averaged and are shown in Table 3-5. The lamins have also been analysed as a subgroup of type V IF to highlight the greater degree of homology present than when they are taken in combination with the *Helix pomotia* B protein.

3.2.1 Coiled-Coil Segments

The homology data indicate that segment 1A is generally highly conserved within each chain type (Figure 3-1 and Table 3-4) and also in the type I-IV IF chains as a whole (Table 3-5) compared to the other coiled-coil segments. The observation of highly conserved sequences within segment 1A and at the C-terminal end of segment 2B has been noted previously (see Chapter 1) and is confirmed by this analysis. Indeed the degree of homology is very high in the C-terminal portion of segment 2B for the type III and type IV chains and over the last 30 residues it is close to 100% (Figure 3-1; Table 3-3). The type V chains reveal a similar highly conserved region in segment 1A but no corresponding region at the C-terminus of segment 2B. This is due to the overall low homology of the *Helix pomotia* B protein with the other type V rod

Segment	Type I	Type II	Type III	Type IV	Type V
1A	–	–	35	34–35	35-41
L1	–	47–50	36–40 45–50	36–41	–
1B	–	74–77	–	72-76	43–55 154–156
L12	161–165 174–176	176–179	174–179	171–174	–
2A	–	–	–	–	–
L2	–	–	–	–	–
2B	–	–	–	–	267–270

Table 3-4 Regions in the homology profiles (h_r scores) where the scores are less than 60%. The regions listed are at least three residues in length and indicate variability in the conservation of primary structure.

domains despite other obvious structural similarities (see Table 3-5): the lamins taken alone do exhibit this feature.

The mean h_s scores (Table 3-5) show an interesting difference between coiled-coil segments for the types I-IV IF. The mean h_s score for segment 1B of type I chains is about 5% higher than that for segment 2B (and indeed for segment 2 as a whole). For types II, III and IV chains the mean values of h_s for segment 1B are lower by 6, 9 and 12% respectively than those for segment 2B. The type V lamin chains reveal similar mean scores for the 1B and 2B segments (82.3% and 86.7% respectively). Consequently for keratin molecules, which are believed to be obligate heteropolymers containing a type I chain and a type II chain (see Chapter 1), the sequences of molecular segments 1B and 2 are preserved at comparable levels of significance whereas for the homopolymeric type III and type IV molecules, the sequence of segment 2B has been maintained with greater fidelity than has segment 1B. This indicates that segment 2 in the type III and the type IV molecules plays a slightly different structural role than in the keratin molecules and again raises the possibility that IF may exhibit more than one structural form.

3.2.2 Link Segments

The α-helical link segment L2 exhibits a higher degree of homology within and across all chain types than the other link segments (Table 3-3, Table 3-5, Figure 3-1, Figure 3-2). The h_s scores for L2 range from 83-89% for types I, II and III chains and these values are very similar to those noted previously for coiled-coil segment 1A. The type IV and V h_s scores for segment L2 are a little lower at 68.4% and 66.3% respectively

Segment	Type I	Type II	Types I and II
1A	85.8±4.4 (8)	86.3±1.9 (8)	69.9±2.2 (16)
L1	42.1±6.8 (10)	42.5±3.6 (8)	37.6±5.4 (18)
1B	80.0±3.7 (10)	72.5±1.8 (8)	58.8±4.8 (18)
L12	54.4±4.8 (12)	62.9±1.7 (8)	48.9±4.1 (20)
2A	74.2±4.7 (13)	78.4±4.3 (8)	63.8±4.7 (21)
L2	88.2±3.1 (13)	89.6±3.8 (8)	72.1±5.5 (21)
2B	74.8±4.2 (13)	78.3±2.4 (8)	59.7±4.2 (21)
Rod	70.4±4.3 (8)	72.5±1.8 (8)	57.5±3.3 (16)

Segment	Type III	Type IV	Types III and IV
1A	85.0±2.7 (5)	81.8±3.3 (9)	76.4±3.4 (14)
L1	31.8±1.5 (5)	38.7±2.4 (9)	34.6±4.6 (14)
1B	78.0±4.1 (4)	70.4±3.6 (9)	65.5±2.9 (13)
L12	56.3±3.7 (6)	51.3±0.9 (10)	46.8±4.4 (16)
2A	88.6±4.9 (7)	75.2±0.6 (11)	73.4±4.9 (18)
L2	86.0±3.7 (7)	68.8±7.6 (11)	71.1±7.3 (18)
2B	86.9±2.5 (8)	82.0±2.2 (10)	75.4±1.8 (18)
Rod	71.8±2.9 (4)	66.7±2.5 (9)	62.5±2.9 (13)

Segment	Types I, II, III and IV	Type V (Lamins only)	Type V (all)
1A	66.3±2.3 (30)	86.7±3.2 (3)	72.5±8.5 (4)
L1	32.6±6.0 (32)		
1B	52.7±2.9 (31)	82.3±2.9 (3)	64.5±11 (4)
L12	38.7±5.1 (36)	78.0±2.6 (3)	57.3±19 (4)
2A	60.2±5.1 (39)	75.7±4.0 (3)	60.0±8.9 (4)
L2	61.3±7.3 (39)	83.0±0.0 (3)	66.3±10 (4)
2B	56.7±3.8 (39)	86.7±4.0 (3)	67.3±12 (4)
Rod	51.7±2.9 (29)	70.0±2.6 (3)	55.0±10 (4)

Table 3-5 Mean segment homology scores (h_S) for the rod domain segments in IF proteins (see Figure 3-1). Values are quoted as mean ± s.d. (number of complete sequences in group). Segments L1 and L12 are padded to allow optimum alignment of the sequences - the values correspond to the maximum possible score for those segments and cannot be directly compared to other entries in the table. The type V sequences have been examined as two groups: one without the *Helix pomotia* B protein (lamins only) and one with it (all).

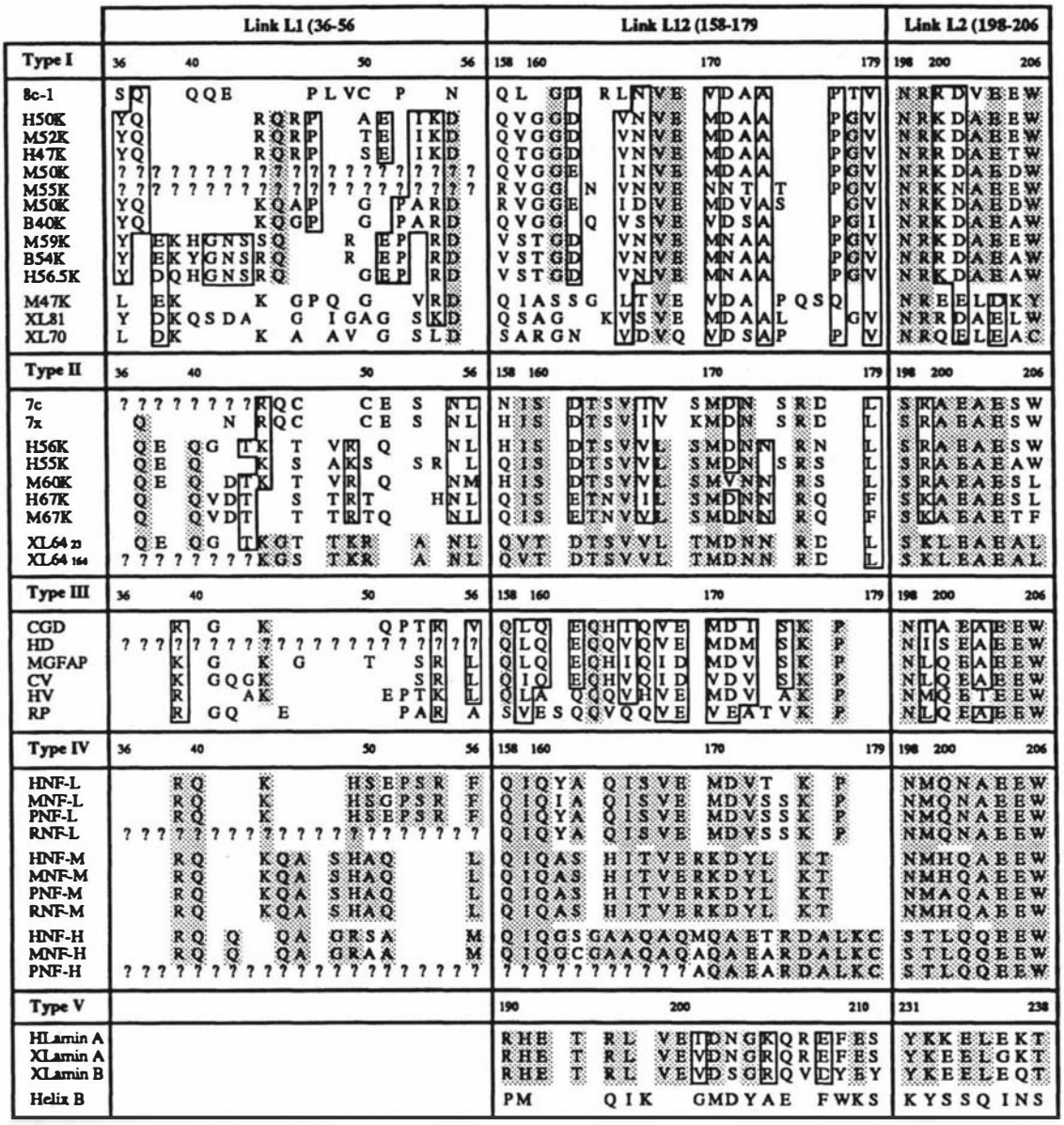


Figure 3-2 Comparison of the amino acid sequences for IF link segments L1, L12 and L2. The sequence abbreviations are defined in Table 2-1. L1 and L12 have both been padded with blanks in order to maximize the homology present and to maintain alignment within the rod domains. Those residues which are conserved in character (ie, conservatively substituted) are boxed and those that are absolutely conserved within a chain type or subtype are shaded. The symbol '?' is used to represent an incomplete sequence. No link L1 segment is defined for the type V chains.

although the lamins alone have a rather higher mean score of 83.0%. The differences in the sequence of segment L2 between IF types are also of significance as they clearly distinguish one type from another. For example, the first four residues in type I and type II chains are **N-R-basic-D** and **S-R-A-E** respectively and each is maintained virtually without exception (Figure 3-2). For type III, IV-L and IV-M chains, the second residue in segment L2 is apolar (cf. arginine or lysine in type I and II chains) and for type IV-H chains it is threonine. The fourth residue is non-acidic in type IV chains, unlike all other IF chains except the type V *Helix pomotia* B chain. Type III

and IV chains are also similar in that link L2 ends with the three residue motif **E-E-W**. The lamins can be distinguished by their sequence in this region which may be generalised as **Y-K-••-E-L-E-••-T**. The *Helix* B protein does not easily fit into this or any other IF pattern and remains somewhat anomalous.

The Glu-6 in segment L2 is conserved in all IF chains studied except for M47K and *Xenopus* lamin A (Figure 3-2) and appears to represent a residue crucial to structure or function. The fact that it occurs in a region where the heptad regularity is temporarily disrupted is almost certainly significant. The implication of highly conserved features in segment L2 is that the IF chains have a highly specific conformation here that cannot readily tolerate sequence changes and the effect of this may be to allow the integrity of the coiled-coil structure to be maintained either side of the discontinuity. It is interesting to compare the homology here with that of the other heptad discontinuity that occurs in a region close to the centre of the coiled-coil segment 2B (residue 265; see Figure 3-1). In this case the homology scores vary somewhat amongst the types (Figure 3-1) but are neither very high (≥90%) nor very low (<60%). The M47K sequence is somewhat different from those of other type I keratins, as detailed in the introduction to this Chapter, and is only listed in Figure 3-2 for completeness. The *Xenopus* lamin B chain has a glycine residue in this position and may well adopt an alternative structure around the discontinuity.

Segments L1 and L12 present a problem in the calculation of their homology statistics due to their variable lengths even within types. Some homology clearly exists within both segments as can be seen in Figure 3-2 where they have been padded to allow alignment of the sequences. Overall, however, the L1 and L12 segments show the least conserved primary structure of any rod domain segment (Table 3-3, Table 3-4 and Table 3-5). Segment L12, which can be from 15 to 22 residues in length, has been padded with up to seven blank residues to maximize the reasonable degree of homology in the amino acid sequences of all IF chains (Figure 3-2). Segment L1 is even more variable in length, and requires about twice as many blanks as segment L12: even then the maximization of homology shown in Figure 3-2 is relatively poor. It is clearly inappropriate that the h_s homology scores from these padded segments be compared with those from non-padded segments. As this and other studies have shown that IF chains have evolved from a common ancestor (Section 1.4), it is tempting to speculate that the precursors to segments L1 and L12 were both somewhat longer than they now appear and that their conformations have permitted variations (primarily deletions) to be tolerated with only limited structural restrictions. For the case of the type V proteins, the link L1 appears to be replaced by small extensions to the coiled-coil segments 1A and 1B.

3.2.3 Type V Chains

The type V chains require special considerations since their structure is unique compared with other IF proteins. For example, the link L1 is effectively replaced by extensions of the coiled-coil segments 1A and 1B: consequently it does not require any padding. Also, the relatively recent establishment of this IF type has prevented any sensible comparisons of the chains that constitute the group. Until recently only the sequences of human lamins A and C were available and these shared a common rod domain sequence. A number of type V sequences from non-mammalian sources have now been completed and some comparison amongst the chains is possible.

An observation made in Chapter 2 was that the *Xenopus* and human lamin A chains were very similar (83% of residues in the rod domains were identical). The mean homology scores of Table 3-5 for the lamins (human lamin A and *Xenopus* lamins A and B) reflect this although the *Xenopus* lamin B sequence, which is not so similar as the others, tends to reduce the scores somewhat. The *Helix pomotia* B protein was included in a separate calculation and the resulting scores were significantly lower than for the lamin chains alone. This raises the possibility of further sub-typing within the type V classification although more sequence data from a variety of sources would be required to establish the validity of this suggestion.

3.3 Subtyping on the basis of Homology

3.3.1 Hard and Soft Keratin IF Chains

The division of keratin protein chains into the type I and type II categories is straightforward on the basis of homology within the rod domain segments and also the N- and C- terminal domain segments (see Chapter 1). However, analysis of the hard α-keratin sequences available (Table 2-1) show that significant differences occur between each of them and the epidermal keratin chains from the same type classification. For example, the hard keratin chains have a higher cysteine content than their epidermal counterparts and the majority of these residues form disulphide linkages which render the keratin complex highly insoluble (Chapter 1); the type II hard α-keratin chains have 10 residues in segment L1 (Fraser *et al* , 1988) compared to 11 or 12 for the epidermal type II chains and these are the shortest L1 segments among the protein chains included in this study; when combinations of hard and soft keratin are compared directly, the numbers of identical residues are notably lower than between combinations of the soft epidermal chains or between combinations of the hard keratin chains. Furthermore, Crewther *et al* (1985) have reported that there is a

Segment	Type I		Type II	
	Hard	Soft	Hard	Soft
1A	79.0**	88.0±1.7	85.3*	86.8±1.1
L1	32.0**	44.6±4.9	39.0*	44.6±2.9
1B	74.0**	81.5±1.9	72.7†	72.4±2.3
L12	49.0*	55.5±4.5	63.3†	62.6±2.2
2A	66.0**	75.7±3.3	79.3†	77.8±5.5
L2	84.0*	89.0±2.6	93.0*†	87.6±3.4
2B	68.0**	76.1±3.2	79.0†	77.8±3.1
Rod	64.0**	72.5±2.0	72.7	72.4±2.3

Table 3-6 Mean segment homology scores (h_s) for the rod domain segments in hard and soft mammalian keratins. Those mean h_s scores for the hard keratins that differ by at least 1 s.d. and 2 s.d. from the soft keratin scores are indicated by * and ** respectively. † marks the hard keratin h_s scores that are higher than the epidermal equivalents.

90-95% identity between the rod domains of components 8c-1 and 8a (both type I hard chains) and also between the rod domains of components 7c and 5 (both type II hard chains).

The h_s homology scores for the hard and soft keratins are listed in Table 3-6. The h_a scores, from which these h_s values were derived, originate from an analysis encompassing all mammalian keratins of a particular type. The mean h_s score (±s.d.) for the epidermal keratin chains was then calculated and, because the contribution from the hard and soft keratins has been separated, the values quoted differ a little from those listed in Table 3-5. The results indicate that the h_s scores for the type I hard keratins are generally at least two standard deviations less than those derived for the type I soft keratins. The differences between the type II hard and soft keratins are less marked, indicating that the sequence changes which have occurred have been of a more conservative nature.

As proposed in Conway and Parry (1988), these differences suggest that it is appropriate for type I and type II IF chains to be subdivided into type Ia and IIa (hard α-keratins) and type Ib and IIb (epidermal keratins). This classification implies that type Ia and type IIa chains could coaggregate and that type Ib and type IIb chains could coaggregate but that cross-combinations of hard and soft are not likely. Further data is still required to substantiate this proposal.

3.3.2 Type IV IF Chains

The type IV IF neurofilament chains can be subdivided into three groups on the basis

of molecular weight: NF-L (light), NF-M (medium) and NF-H (heavy). Analysis of the rod domains of the neurofilament chains included in this study shows that the chain designations L, M and H may also be assigned on the basis of sequence homology. Indeed, the NF-L rod domain sequences show a remarkable 98% identity within the rod domain. The NF-M and NF-H sequences also show high levels of identity at 95% and 96% respectively (see also Myers *et al*, 1987 and Napolitano *et al*, 1987 for identity throughout the NF-M chains). In almost all cases the non-identical residues are conservative substitutes. In contrast, the degree of identity between the groups is much lower: for NF-L and NF-M rod domain sequences it is about 53%; NF-L and NF-H are 42%; NF-M and NF-H are 53%; and all NF rod domain sequences together show about 35% identity. It is interesting that the heavy and light chains are least identical but show similar conservation with the medium chains. This is an indication that the three neurofilament subtypes may have evolved from a common IF-type protein. The greatest evolutionary changes to this group are apparent in the differing extents of the carboxy-terminal domains but the rod domains also reveal the effects of divergence. Section 3.4 will show that the sequences comprising the H1 and H2 subdomains in the NF-L and NF-M chains are maintained at a similar level to those in the rod domain.

The extraordinarily high degree of sequence identity within the NF-L, NF-M and NF-H subtypes is a strong indication that only minor sequence variations can be tolerated without losing structural and/or functional viability. The amino acid sequences of neurofilament chains clearly provide an alternative means of chain classification (L, M, or H) without reference to molecular weight data.

3.4 Homology among the 'H' subdomains

Analysis of the sequences of the N- and C-terminal domains in keratin IF chains has shown that both may be subdivided into segments of special character (Steinert *et al*, 1984b; Steinert and Parry, 1985; Steinert *et al*, 1985a; see Figure 1-3). These subdomains exhibit bilateral symmetry with respect to the IF coiled-coil rod domain. Segments H1 and H2 are defined as those subdomains with a high degree of sequence homology within a chain type grouping which are located immediately adjacent to the ends of the rod domain (ie. at the N- and C-terminal ends respectively). Distal to H1 and H2 are the so-called V1 and V2 'variable' subdomains. These differ greatly in both sequence and length and frequently have high contents of glycine and serine residues. Finally, distal to V1 and V2 and hence at the N- and C-terminal ends of the chain are the E1 and E2 'end' domains respectively (Parry *et al*, 1987b). The increased sequence data available has allowed the character and size of the H1 and H2 subdomains to be reassessed. It should be noted that the precise N-terminal boundary

for H1 and the C-terminal boundary for H2 are often not well-defined and the limits may lie a few residues displaced from those stated. The presence of the H1 and H2 subdomains implies directly that these regions adopt a specific conformation *in vivo*, though their role has yet to be elucidated.

(i) Type I. Steinert and Parry (1985) and Steinert *et al* (1985a) have reported that the H1 and H2 subdomains are absent in type I chains. This study confirms that H2 is indeed absent for type Ib chains but no conclusions can be made for the type Ia chains as there is only one sequence, component 8c-1, available for this subdomain. There is no evidence of an H2 subdomain in the non-mammalian type I chains (*Xenopus*) nor is there sign of any significant homology between the individual *Xenopus* chains in this region.

For subdomain H1 there is evidence of a seven residue region adjacent to the rod domain that is homologous in all mammalian type I IF chains (Table 3-7 and see also Figure 3-4). A longer H1 subdomain may exist for the type Ia chains but, as above, this cannot be determined without further sequence data. The limited *Xenopus* sequences do not appear to have an H1 subdomain nor do they exhibit any significant interchain homology.

Some sequences included in this analysis exhibit extensive homology between themselves far in excess of that specified by the lengths of H1 and H2 (see, for

Chain Type	H1 (residues)	H2 (residues)
I (all)	7	0
Ia	≥7	≥0
Ib	7	0
II (all)	36	20
IIa	≥32	≥10
IIb	36	20
III (all)	20	58
IV (all)	9	13
IV-L	89	160
IV-M	102	83
IV-H	98	75
V (all)	5	0
Lamins	13	6

Table 3-7 Lengths of the homologous subdomains H1 and H2 for mammalian type I-IV chains and for all type V chains. Greater than symbols (≥) are used where there is insufficient sequence data to delineate precisely the extent of a subdomain.

example, H1: H46K, H50K and H56.5K; H2: H50K and M50K and also B50K and B54K). It may be concluded that such sequences have similar conformations and functions *in vivo*.

(ii) Type II. The extents of H1 and H2 in type II chains as a whole are 36 and 20 residues respectively as originally reported by Steinert and Parry (1985) and Steinert *et al* (1985a) (Table 3-7). The IIb grouping also follows this pattern but as there is only limited data available for the IIa subtype no definitive sizes for the H1 and H2 subdomains can be established beyond the minimum lengths of 32 and 10 residues respectively. There are no *Xenopus* data pertaining to H1 but in H2 the extent of homology for *Xenopus* chains appears to be 20 residues, the same as that noted for mammalian type IIb chains. No homology beyond this limit is apparent among the *Xenopus* chains in this region.

As noted for the type I analysis, individual combinations of type II chains can reveal a higher degree of homology over a much longer piece of the sequence than that specified by the H1 and H2 subdomains (see, for example, H1: M60K and M67K; H2: M67K and H67K and also M60K and H56K).

(iii) Type III. The sizes of the H1 and H2 subdomains as reported by Steinert and Parry (1985) and Steinert *et al* (1985a) are confirmed for type III chains as being 20 and 58 residues respectively (Table 3-7). The latter value represents the entire C-terminal domain of type III chains and includes three blanks inserted to maximize homology among the chains.

(iv) Type IV. Insufficient neurofilament sequence data have prevented an assessment of the lengths of the H1 and H2 subdomains prior to Conway and Parry (1988). Sequences of four NF-L, four NF-M and three NF-H chains are available and the results of the analysis are listed in Table 3-7. As a whole, the type IV chains reveal short H1 and H2 subdomains of 9 and 13 residues in length respectively. However an extensive pattern of homology is evident within each of the subtypes and it is apparent that the H, V, and E subdivisions do not apply as well to the type IV chains as to the type I and II chains. The entire N-terminal domains of of NF-L chains, NF-M chains and of NF-H chains (89, 102 and 98 residues respectively, with inserted blanks where appropriate) are highly homologous and may be designated H1-L, H1-M and H1-H respectively. The proportions of identical residues in this domain are 92%, 98% and 86% for the light, medium and heavy chains respectively. These values are comparable to the identity within the rod domain of these groups (Section 3.3.2). However, in the C-terminal domain two homologous regions are evident which are separated by a variable region. The first homologous region, which is adjacent to the rod domain, is

160 residues long for NF-L chains, 83 residues for NF-M chains and for NF-H chains it is 75 residues. This region encompasses all of the C-terminal domain for NF-L chains. The second homologous region is 111 residues for the NF-M chains and 75 for the NF-H chains. The intervening region in the NF-M and NF-H chains contains a glutamic acid-rich region followed by a repeating run of between 6 to 13 residues that commonly includes the sequence **LYS-SER-PRO**. (These repeats can be observed, for example, as regular patterns in the flexibility profile for the human NF-M chain in Figure 2-3: see Geisler *et al*, 1987, and Shneidman *et al*, 1988, for discussions on the structure and function of this terminal domain). The small data base of neurofilament proteins limits the reliability of the analysis and the extents of the subdomains may need revising as further data becomes available.

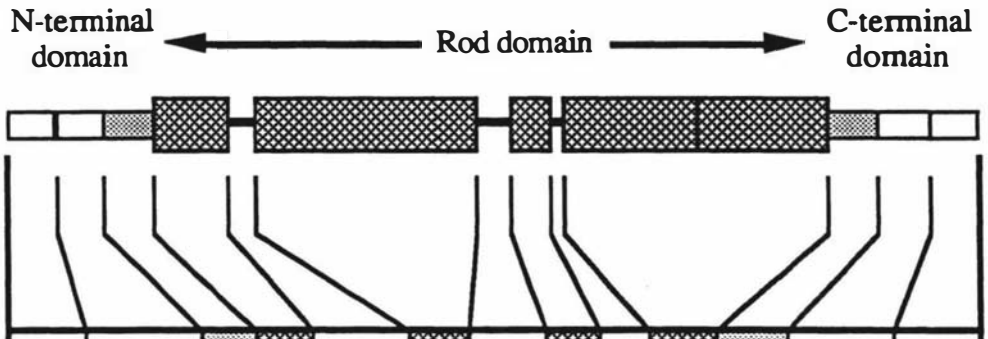
(iv) Type V. The lamin H1 and H2 subdomains have lengths different to those found in type I-IV IF chains (Table 3-7). Extensive homology is present in the N- and C-terminal domains of the human lamins A and C and the *Xenopus* lamin A sequence (see Figure 2-2) but inclusion of the *Xenopus* lamin B chain reduces the lengths of H1 and H2 to 13 and 6 residues respectively. The *Helix pomotia* B chain further reduces the extents of these regions. As for the type IV chains, the subdivision scheme is not entirely appropriate for the type V chains. More protein sequences are required to define the nature and boundaries of the H1 and H2 subdomains and also to determine whether further subtyping of the type V proteins is appropriate.

Table 3-8 shows the extents of all the IF chain domains and represents an update from that given in Chapter 1 (Table 1-1). It incorporates the new data presented in this section on the H1 and H2 domains and on the type IV and V chains in general. Several chain types still require further sequence data to allow a complete analysis of their terminal domain boundaries including the hard keratins, the neurofilaments and the type V chains.

3.5 Comparison of Amphibian and Mammalian Scores

Complete or partial sequences of three type I, two type II and two type V amphibian (*Xenopus*) IF proteins have been used in this study (Table 2-1). Much of the analysis of homologous features has focused on mammalian proteins but it is also important to assess the degree of relatedness that exists between IF chains from these two diverse sources.

The classification of the amphibian keratin chains in types I and II is unequivocal using the standard criterion of sequence homology in the rod domain where the differences between the mammalian and amphibian chains are generally small. In particular, the



	E1	V1	H1	1A	L1	1B	L12	2A	L2	2B	H2	V2	E2
Ia	?	?	≥7	35	11	101	16	19	8	121	≥0	?	?
Ib	0-10	65-140	7	35	11-14	101	16	19	8	121	0	0-110	5-30
IIa	?	?	≥32	35	10	101	17	19	8	121	≥10	?	?
IIb	5-70	20-160	36	35	11-14	101	17	19	8	121	20	25-125	21
III	19-81		20	35	8	101	16	19	8	121	58	0	
IV-L	0		89	35	10	101	17	19	8	121	160	0	
IV-M	0		102	35	10	101	17	19	8	121	83	237	111
IV-H	0		98	35	10	101	22	19	8	121	75	≥493	200
V	25-64		5	41	0	148	15,19	19	8	121	0	185-287	

Table 3-8 Extents of the structural domains for IF chains. '?' indicates insufficient sequence data are available for the boundaries to be evaluated. In the case of the type III and V chains no clear division into E and V domains is possible on the same basis as for type I and II chains. The same is true for the type IV chains although a domain of variable length does exist in the carboxy-terminal of the chains. This domain is bracketed by regions which show a high degree of homology within each subtype.

type II comparisons show that the segment homology scores for the Xenopus chains are comparable with those for the mammalian epidermal chains (Table 3-9), ie, the *Xenopus* sequences show little variation from the consensus. A point of minor divergence lies in segment L2 (which is generally highly conserved, see Section 3.2.2) where the *Xenopus* sequences **SKLEAEAL** differ in several respects from the mammalian consensus sequence **SRAEAES**$^{\mathbf{L}}_{\mathbf{W}}$ (Figure 3-2). The lengths of segments L2 (8 residues) and L12 (17 residues) are strictly conserved in the mammalian and amphibian chains, However, the length of segment L1, which is either 10 or 12 residues long in mammalian chains, is 14 residues long in the amphibian chains. In general there is little homology between any of the type II sequences in this segment.

In contrast to the type II comparisons, the mean segment homology scores for the three type I *Xenopus* keratins are less than those for the equivalent segments in the

Segment	Type I		Type II	
	Xenopus[a]	Mammalian	*Xenopus*[b]	Mammalian
1A	81.0**	89.2±1.8	–	91.4±1.1
L1	32.5**	44.3±5.9	55.0*†	47.4±4.2
1B	75.0**	81.9±1.8	77.5	77.8±3.3
L12	48.0*	56.0±4.9	65.0†	64.4±2.6
2A	67.0**	76.8±3.2	82.0†	80.8±5.8
L2	66.5**	86.2±2.2	80.0*	82.8±2.3
2B	66.5**	75.7±3.0	75.0*	78.0±2.3
Rod	61.5**	72.7±1.5	–	74.0±2.0

Table 3-9 Mean segment homology scores (h_s) for the rod domain segments in amphibian (*Xenopus*) and mammalian epidermal keratins. Those mean h_s scores for the *Xenopus* keratins that differ by at least 1 s.d. and 2 s.d. form the mammalian scores are indicated by * and ** respectively. † marks the *Xenopus* h_s scores that are higher than the epidermal equivalents. (a) *Xenopus* XL51, XL70 and XL81. (b) *Xenopus* XL64 23 and XL64 164 (See Table 2-1 for sources).

mammalian chains, usually by at least one standard deviation. In this respect the *Xenopus* type I and type II keratins parallel the hard type I and type II α-keratins in that the variations in homology between the type I chains (*Xenopus*, hard, and epidermal keratins) are greater than between the type II chains from the same sources (Tables 3-6 and 3-9). Apparently type II chains as a whole seem less able to permit variation in primary structure than type I chains. This feature can also be seen in the smaller standard deviations for the segment homology scores for type II chains compared with those for type I chains (Table 3-5).

The lengths of segments L12 (17 residues) and L2 (8 residues) for *Xenopus* type I chains are the same as for the mammalian type I chains but segment L1, which is either 11 or 14 residues long in mammalian chains, is 11 or 15 residues long in *Xenopus* chains. Once again, relatively little homology is present in this segment when mammalian and amphibian chains are compared. Segment L2 in the *Xenopus* chains shows the most variation in any segment when compared to the consensus type I sequence: **NR•AcApAc••** in *Xenopus* and **NRKDAEAcW** in mammals (Figure 3-2).

Insufficient sequence data for type V IF chains are currently available to allow a similar analysis of the degree of homology amongst mammalian and amphibian lamin sequences. However a preliminary examination of their sequences reveals some pronounced similarities between several mammalian and amphibian members of this group. The human lamins A and C have identical amino-terminal and rod domains, and a portion of their carboxy-terminal domains are also identical. The *Xenopus* lamin

A rod domain sequence is 83% identical with these two human lamin chains and a further 7% of the sequences are maintained in chemical character (see Figure 2-2). The similarities extend throughout the protein sequences of the three lamin proteins: 23 of the 30 residues are identical in the N-terminal domains (which include 4 blanks to maintain alignment) and 166 of the 283 residues are identical in the C-terminal domains of the lamin A sequences (which also include 4 blanks to maintain alignment; human lamin C is excluded because it has a shorter C-terminal domain). A comparison of human and *Xenopus* lamin B chains would be of great interest but as yet no sequence data are available for the human lamin B protein.

3.6 Periodic Features in the Homology Score Distributions

Fourier transforms of the residue homology scores h_r derived separately for the large apolar residues, the acidic residues and basic residues have been calculated for the coiled-coil segments 1B and 2 using the look-up tables in Table 3-2. The method of performing the transform is similar to that described in Chapter 2 (Section 2.1.1) where the input data record is baseline corrected and zero-filled, although for this study the filled sequences were only 1024 elements in length and the resulting Fourier intensities have not been scaled. The Fourier intensities for types I to V chains are shown in Figure 3-3 and a selection of the most significant periods are listed in Table 3-10. Note that the stutter in heptad phasing located towards the middle of segment 2B (residue 265) was taken into account by calculating separately the Fourier transforms of the apolar homology profiles either side of the stutter, ie. segments 2B1 and 2B2 (defined in Section 3.1).

The distribution of apolar residues in segment 2B shows no periodicity longer than the heptad repeat. In segment 1B some less well defined periods are revealed in addition to the heptad repeat. These other periods are also seen in the acidic and basic homology distributions described below.

Periods common in the homology profiles of the acidic and the basic residues (orders marked with [a] in Table 3-10) are almost identical to those previously reported for the distribution of the same residue types in individual IF chains (~9.55 in segment 1B and ~9.85 in segment 2 – see Chapters 1 and 2). This conservation of charged residues is a further indication of their importance in specifying modes of molecular aggregation in the IF. Another period revealed by the transforms for segment 1B is equal to three times that of the -9.55 period. This longer period (28-30 residues or 4.3 nm) can be considered as arising from the ~9.5 residue repeat beating with the heptad period, ie. $3 \times 9.5 \approx 4 \times 7.0$ residues, and is an indication of the fact that charged residues occur frequently in the outer b, c and f positions of the coiled-coil structure

(Parry, 1982; Parry and Fraser, 1985; also see Section 2.2 and Table 2-4a). Such positions have an inherent seven residue repeat. In addition to the 28-30 residue period in segment 1B there is a less well defined period of about 10.5-11.0 residues in the distribution of homologous acidic residues in all but type II chains. This period is also present in the distribution of the homologous basic residues in type V chains. The ~11 and ~28 residue periods lie almost exactly in the simple ratio 3:8 (ie, 0.375: type I-0.372; type III-0.377; type IV-0.375) which implies that they are both suborders of a longer regularity. For type I chains this segment 1B period is 84 residues (ie. 3x28.2 or 8x10.5) and for type III, IV and V chains it is 88 residues (ie. 3x29.2 or 8x11.0). Note that the 84 residue period (type I) is an integral multiple of heptads whereas the 88 residue period (types III, IV and V) is not. The significance of this observation is presently unclear.

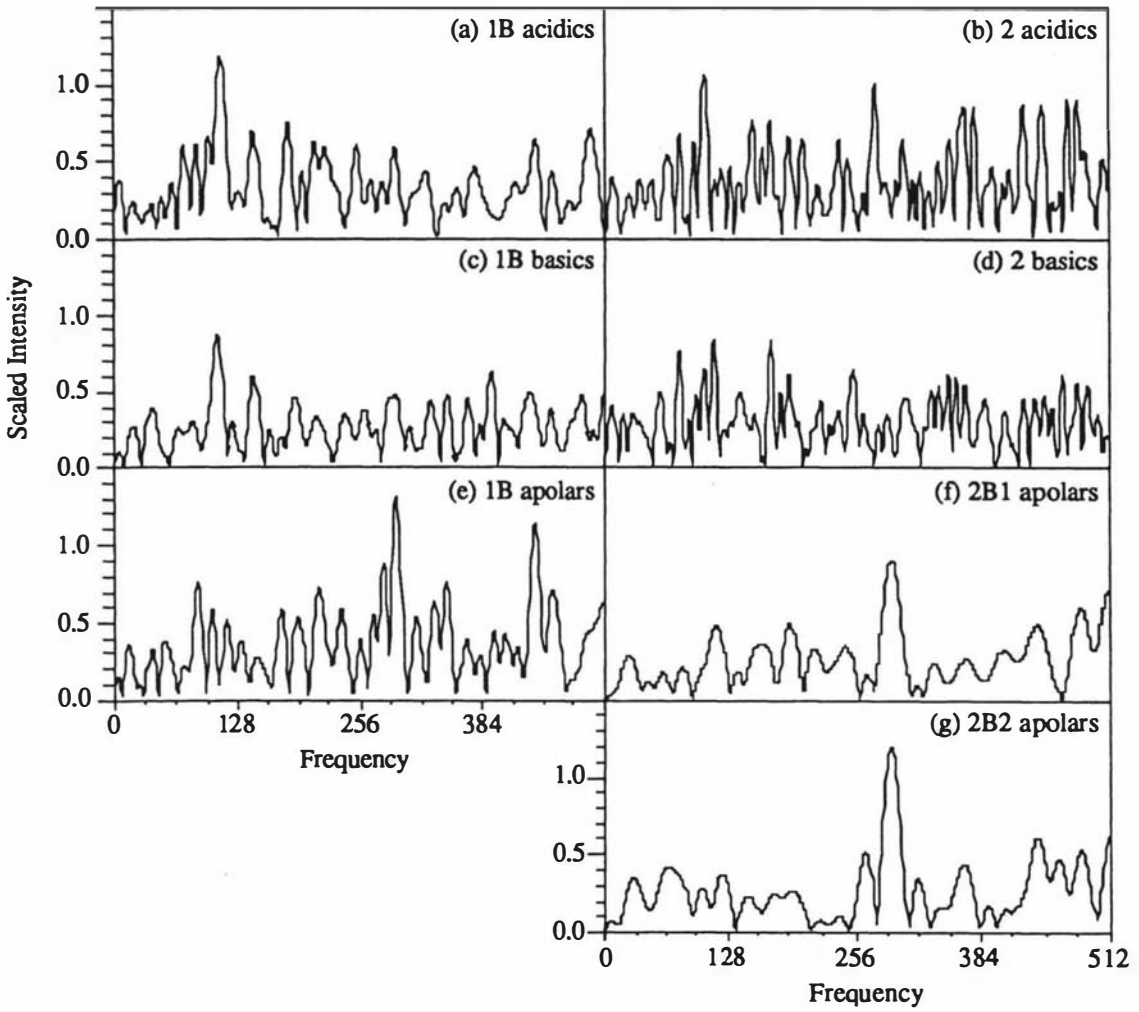


Figure 3-3a Fourier transforms of the homology arrays for type I IF proteins derived from the acidic residues in (a) segment 1B and (b) segment 2, the basic residues in (c) segment 1B and (d) segment 2, and the large apolar residues in (e) segment 1B, (f) segment 2B1 and (g) segment 2B2. The frequency axes are related to periods in the homology profiles by the expression:

$$\text{period} = \frac{1024}{\text{frequency}}$$

As for the ~9.5 residue period in segment 1B, the ~9.8 period in segment 2 represents the third order of a longer repeat that is consistent among the chain types (29.5-29.8 residues). The profiles of the acidic homologies in segment 2 for type I-IV chains show a period of ~10.9 residues as do the type II, III and IV basic homology profiles. This period is related to the ~29.7 period in the ratio 4:11. The type I basic homology curve shows a slightly higher period of 11.7 residues which is related to 29.7 in the ratio 4:10. Neither of the ~10.9 or 11.7 periods occur in the type V basic homology distribution of segment 2.

The three main periods found can beat together to show a longer range period of ~118 residues (ie. 4x29.7 = 118.8; 10x11.7 = 117.0; 11x10.8 = 119.4). This period is considerably less than the length of segment 2 (143 residues) and thus cannot be attributed to an artefact of the edge effects arising from computation.

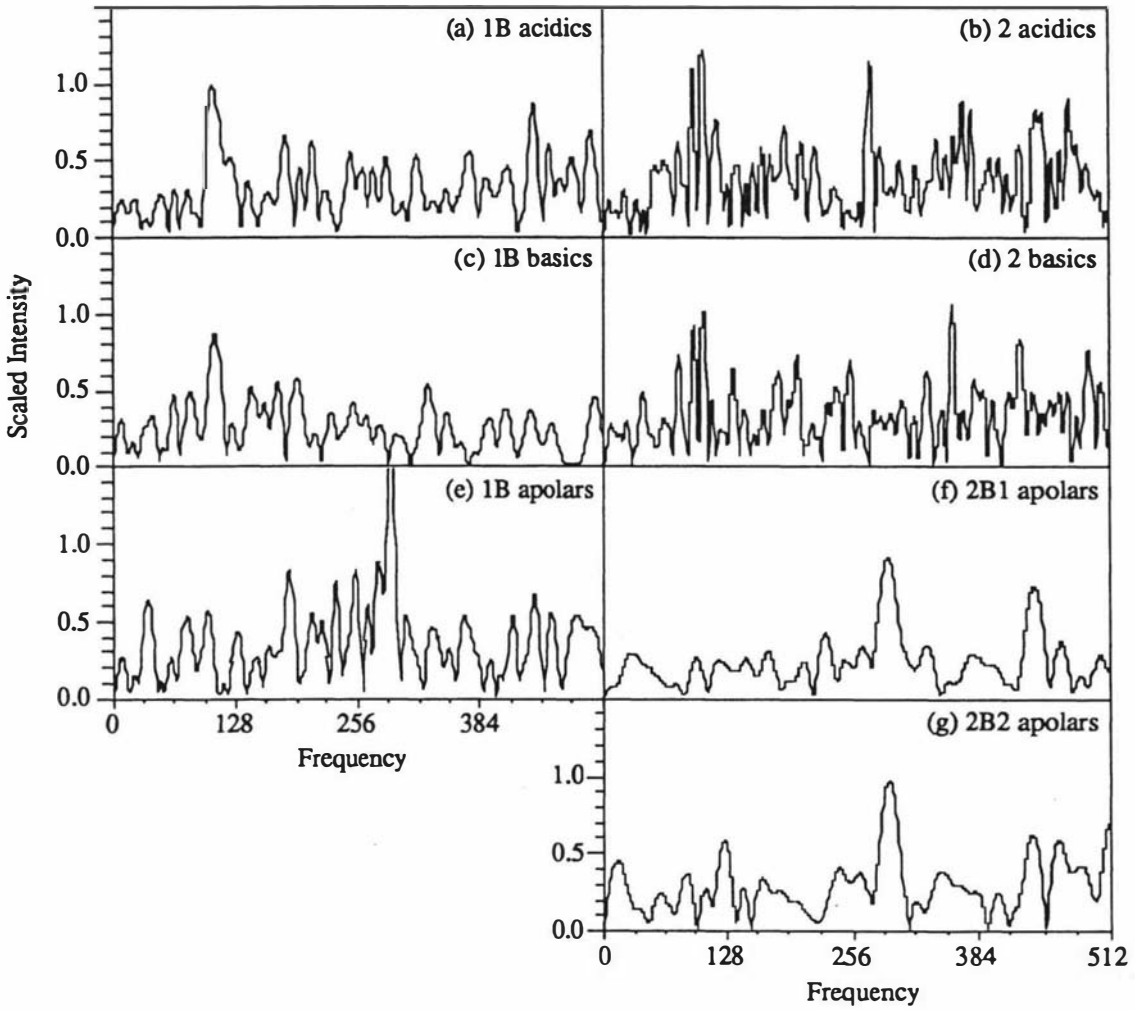


Figure 3-3b Fourier transforms of the homology arrays for type II IF proteins derived from the acidic residues in (a) segment 1B and (b) segment 2, the basic residues in (c) segment 1B and (d) segment 2, and the large apolar residues in (e) segment 1B, (f) segment 2B1 and (g) segment 2B2. The frequency axes are related to periods in the homology profiles by the expression:

$$\text{period} = \frac{1024}{\text{frequency}}$$

The main conclusions from this section are summarized as follows:

(1) The homology profiles for large apolar residues in segment 1B and segment 2 in all chain types are dominated by the heptad repeat. This emphasizes the importance of these residues in specifying the heptad repeat and, by implication, the integrity of the coiled-coil structure of the IF molecule.

(2) The highly conserved periodic distribution of acidic and basic residues in both segment 1B and segment 2 implies that these residues play an important role in the self-assembly of IF molecules into IF through the formation of specific ionic interactions.

(3) The degree of regularity in the distribution of the homologous ionic residues, as determined by the magnitude of the Fourier intensities and the numbers of orders present, varies between chain types. This indicates the possibility that alternative modes of assembly via ionic interactions may occur occur in the IF

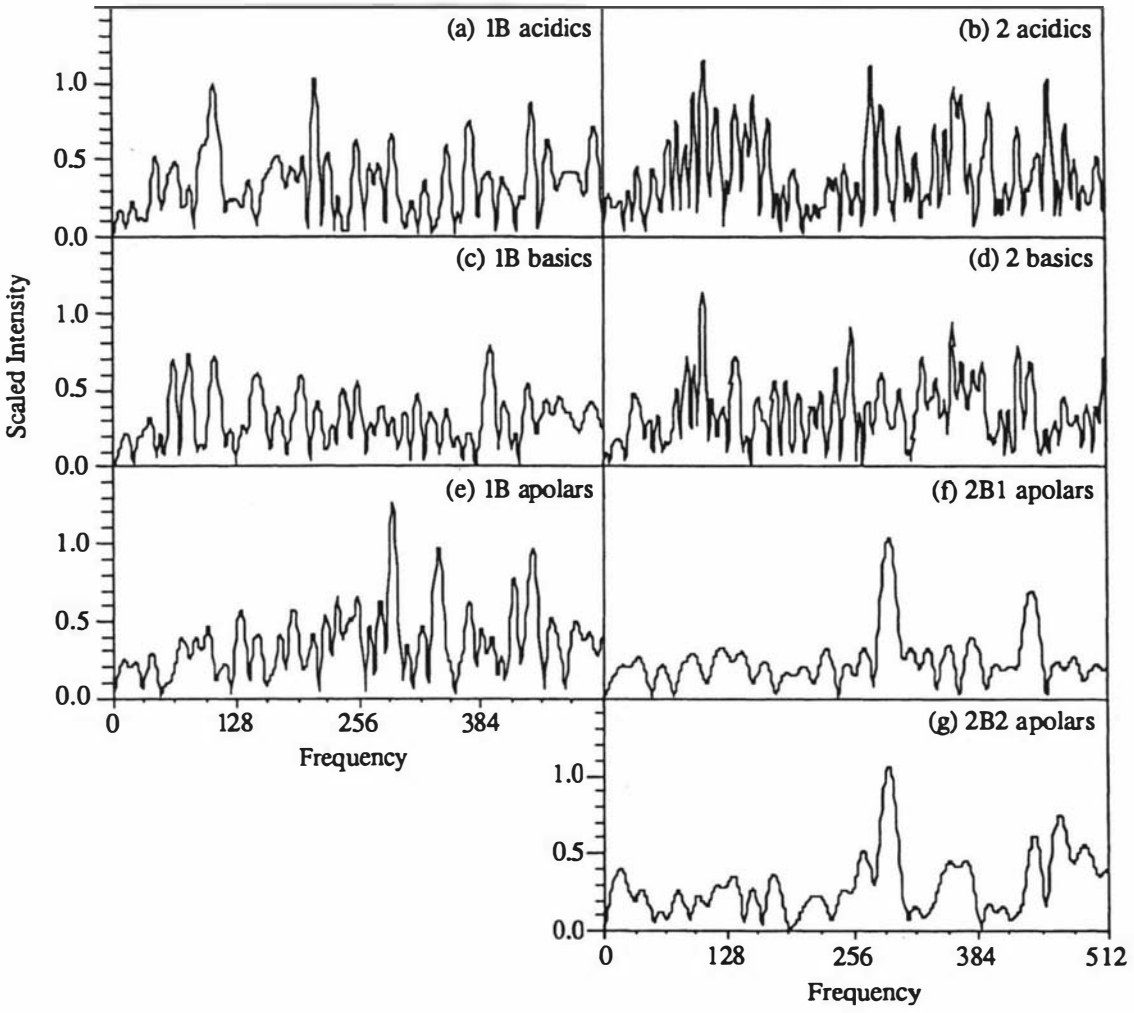


Figure 3-3c Fourier transforms of the homology arrays for type III IF proteins derived from the acidic residues in (a) segment 1B and (b) segment 2, the basic residues in (c) segment 1B and (d) segment 2, and the large apolar residues in (e) segment 1B, (f) segment 2B1 and (g) segment 2B2. The frequency axes are related to periods in the homology profiles by the expression:

$$\text{period} = \frac{1024}{\text{frequency}}$$

corresponding to the different chain types present.

(4) Several periodicities occur in the distribution of homologous ionic residues in addition to the 9.3-9.9 residue repeat and such periods can be related to a repeat that is comparable (although shorter) than the segment length in both segment 1B and segment 2.

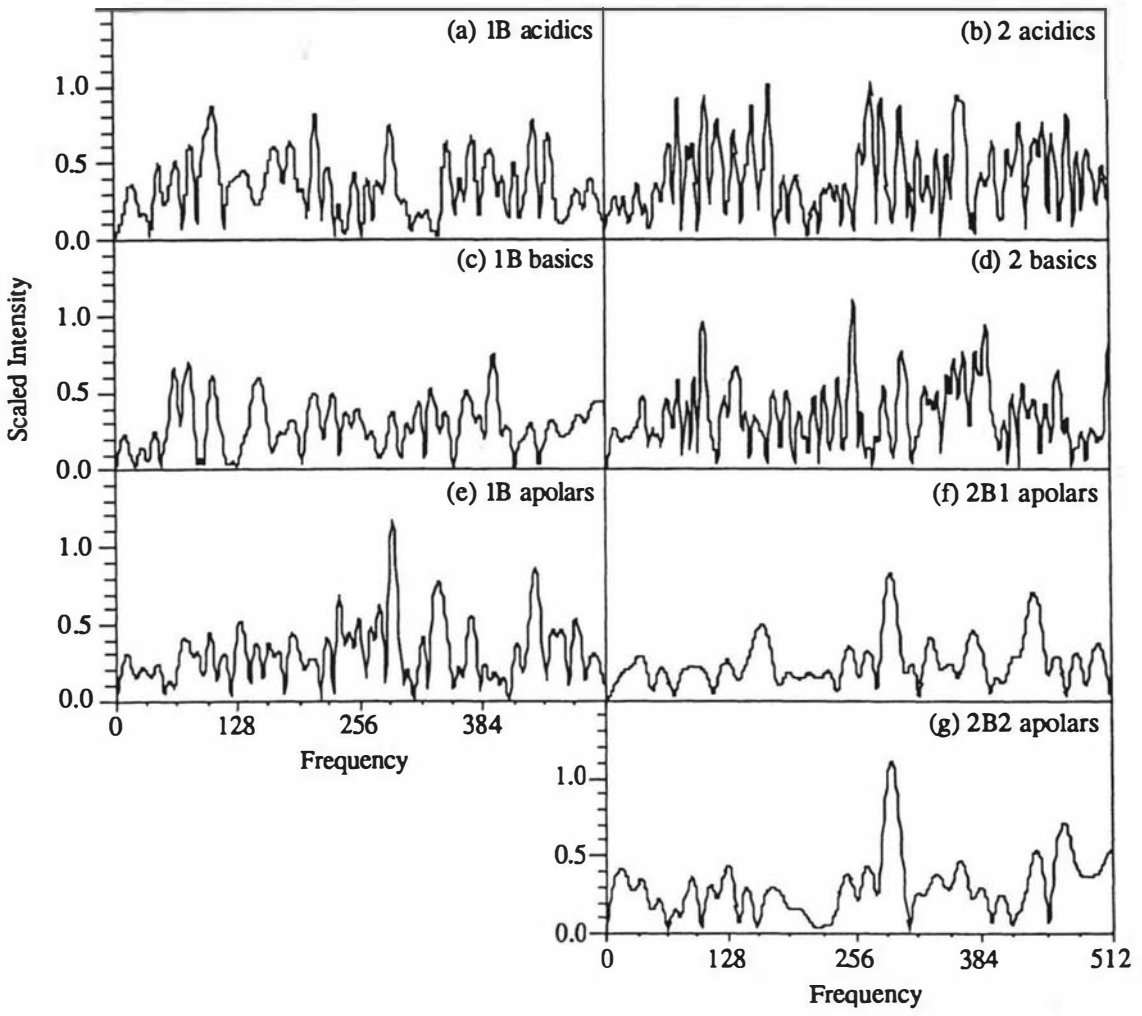


Figure 3-3d Fourier transforms of the homology arrays for type IV IF proteins derived from the acidic residues in (a) segment 1B and (b) segment 2, the basic residues in (c) segment 1B and (d) segment 2, and the large apolar residues in (e) segment 1B, (f) segment 2B1 and (g) segment 2B2. The frequency axes are related to periods in the homology profiles by the expression:

$$\text{period} = \frac{1024}{\text{frequency}}$$

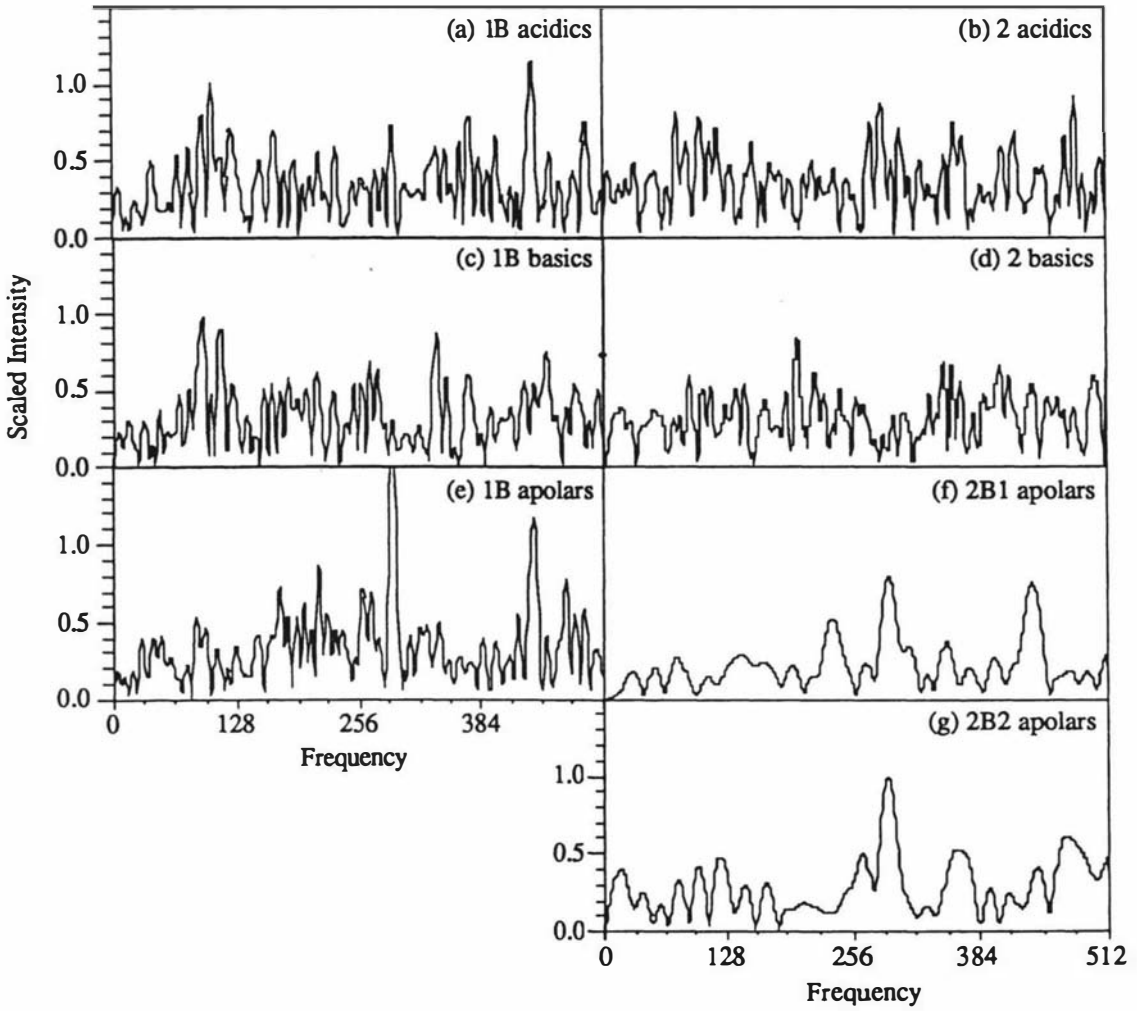


Figure 3-3e Fourier transforms of the homology arrays for type V IF proteins derived from the acidic residues in (a) segment 1B and (b) segment 2, the basic residues in (c) segment 1B and (d) segment 2, and the large apolar residues in (e) segment 1B, (f) segment 2B1 and (g) segment 2B2. The frequency axes are related to periods in the homology profiles by the expression:

$$\text{period} = \frac{1024}{\text{frequency}}$$

	Type I		Type II		Type III		Type IV		Type V	
	Period	Intensity	Period	Intensity	Period	Intensity	Period	Intensity	Period	Intensity
1B Acidic	10.56 (P'/1)	0.66	9.66 (P/3) [a]	0.99	11.01 (P'/1)	0.62	9.85 (P/3) [a]	0.87	14.95 (P/2)	0.54
	9.31 (P/3) [a]	1.18	7.21 (P/4)	0.37	9.66 (P/3) [a]	1.01	5.54 (P'/2)	0.64	11.01 (P'/1)	0.80
	7.11 (P/4)	0.68	5.66 (P/5)	0.67	5.51 (P'/2)	0.42	4.88 (P/6)	0.82	9.94 (P/3) [a]	1.01
	5.69 (P/5)	0.74	4.10 (P/7)	0.55	4.85 (P/6)	1.09	3.53 [b]	0.74	3.51 [b]	0.73
	4.06 (P'/2)	0.43	3.57 (P/8)	0.52	3.68 (P/8)	0.50	2.93 (P/10)	0.63	2.75 (P'/4)	0.79
	3.51 (P/8, P'/3) [b]	0.58	3.22 (P/9)	0.53	3.52 [b]	0.68	2.74 (P'/4)	0.67	2.33 [b]	1.15
	2.33 (P/12) [b]	0.64	2.62 (P/11)	0.39	2.93 (P/10)	0.56	2.45 (P/12)	0.50		
			2.32 [b]	0.87	2.75 (P'/4)	0.82	2.33 [b]	0.79		
			2.19 (P/13)	0.39	2.33 [b]	0.83	2.26 (P/13)	0.69		
	P = 28.2 res P'= 10.5 res		P = 28.7 res		P = 29.2 res P'= 11.0 res		P = 29.3 res P'= 11.0 res		P = 29.9 res P'= 11.0 res	
1B Basic	9.57 (P/3) [a]	0.88	9.57 (P/3) [a]	0.87	9.66 (P/3) [a]	0.69	9.94 (P/3) [a]	0.62	11.01 (P'/1)	0.98
	7.06 (P/4)	0.60			4.79 (P/6)	0.45	4.92 (P/6)	0.50	9.23 (P/3) [a]	0.89
	2.61 (P/11)	0.63			3.23 (P/9)	0.49			3.95 (P/7)	0.54
									2.76 (P/10,P'/4)	0.61
									2.32 [b]	0.54
	P = 28.5 res		(P = 28.7 res)		(P = 28.9 res)		(P = 29.8 res)		P = 27.7 res P'= 11.0 res	

Table 3-10 A selection of the most significant periodicities in the homology scores of IF coiled-coil segments. (a) indicates the domimant 9.3-9.9 residue period in the homology distributions of the acidic and basic residues and is directly comparable to the period found in the linear distribution of the same residues in individual sequences. (b) indicates either the second or third order of the fundamental heptad repeat in coiled-coil structures. A period in parenthesies, eg. (P=29.6 res), indicates that only one order is present but that by analogy with other data the true period is shown.

	Type I		Type II		Type III		Type IV		Type V	
	Period	Intensity	Period	Intensity	Period	Intensity	Period	Intensity	Period	Intensity
1B Apolar	11.91 (P'/1)	0.75	27.68 (P/1)	0.63	11.91 (P'/1)	0.44	4.00 P'/3)	0.54	3.51[b]	1.85
	5.89 (P'/2)	0.58	5.57 (P/5)	0.82	5.92 (P'/2)	0.49	3.52[b]	1.17	2.33[b]	1.18
	3.50[b]	1.31	4.66 (P/6)	0.50	4.02 (P'/3)	0.70	3.02 (P'/4)	0.77		
	2.95	0.76	4.02 (P/7)	0.83	3.51[b]	1.31	2.33[b]	0.86		
	2.33[b]	1.14	3.52 (P/8)[b]	1.59	3.01 (P'/4)	0.99				
			3.06 (P/9)	0.49	2.33[b]	0.98				
			2.78 (P/10)	0.54						
			2.32 (P/12)[b]	0.66						
			2.12 (P/13)	0.53						
	heptad P'= 11.8 res		heptad P = 27.8 res		heptad P'= 12.0 res		heptad P'= 12.0 res		heptad	
2 Acidic	11.01 (P'/1)	0.62	10.89 (P'/1)	1.08	10.89 (P'/1)	0.94	15.06	0.62	9.94 (P/3)[a]	0.79
	9.85 (P/3)[a]	1.06	9.85 (P/3)[a]	1.21	9.75 (P/3)[a]	1.13	11.01 (P'/1)	0.62	3.72 (P/8)	0.76
	5.92 (P/5)	0.76	5.89 (P/5)	0.53	7.47 (P/4)	0.85	9.66 (P/3)[a]	0.94	2.50 (P/12)	0.59
	4.27 (P/7)	0.64	5.45 (P'/2)	0.74	5.95 (P/5)	0.77	7.53 (P/4)	0.71		
	3.71 (P/8)	1.00	3.72 (P/8)	1.15	5.45 (P'/2)	0.33	5.95 (P/5)	1.01		
	2.70 (P/11, P'/4)	0.85	2.98 (P/10)	0.64	4.25 (P/7)	0.39	3.74 (P/8)	1.02		
	2.17 (P'/5)	0.90	2.71 (P/11, P'/4)	0.83	3.71 (P/8)	1.13	3.36 (P/9)	0.86		
			2.29 (P/13)	0.81	2.99 (P/10)	0.71	2.99 (P/10)	0.53		
			2.17 (P'/5)	0.90	2.70 (P/11, P'/4)	0.44	2.33[b]	0.66		
					2.47 (P/12)	0.35				
					2.29 (P/13)	0.54				
					2.12 (P/14)	0.51				
	P = 29.7 res P'= 10.9 res		P = 29.5 res P'= 10.9 res		P = 29.7 res P'= 10.9 res		P = 29.8 res (P'= 11.0 res)		P = 29.8 res	

Table 3-10 *continued*

	Type I		Type II		Type III		Type IV		Type V	
	Period	Intensity	Period	Intensity	Period	Intensity	Period	Intensity	Period	Intensity
2 Basic	11.77 (P"/1)	0.31	10.78 (P'/1)	0.92	10.78 (P'/1)	0.64	10.89 (P'/1)	0.59	9.85 (P/3)[a]	0.51
	9.85 (P/3)[a]	0.65	9.85 (P/3)[a]	1.00	9.85 (P/3)[a]	1.13	9.94 (P/3)[a]	0.96	2.33[b]	0.54
	5.89 (P/5, P"/2)	0.83	2.33[b]	0.48	7.47 (P/4)	0.73	7.42 (P/4)	0.67		
	2.92 (P"/4)	0.62			5.39 (P'/2)	0.57	3.35 (P/9)	0.77		
	2.33 (P"/5)[b]	0.46			4.27 (P/7)	0.66	2.69 (P/11)	0.77		
					3.58 (P'/3)	0.60	2.46 (P/12)	0.48		
					2.99 (P/10)	0.55	2.33[b]	0.45		
					2.69 (P'/4)	0.65				
					2.33[b]	0.70				
	P = 29.5 res P"= 11.7 res		(P = 29.5 res) (P'= 10.8 res)		P = 29.8 res P'= 10.8 res		P = 29.8 res (P'= 10.9 res)		(P = 29.6 res)	
2B1 Apolar	3.50[b]	0.90	3.50[b]	0.91	3.51[b]	1.03	3.49[b]	0.83	3.49[b]	0.79
	2.32[b]	0.49	2.33[b]	0.72	2.33[b]	0.70	2.34[b]	0.71	2.34[b]	0.74
	heptad		heptad		heptad		heptad		heptad	
2B2 Apolar	3.50[b]	1.20	3.48[b]	0.98	3.51[b]	1.07	3.49[b]	1.10	3.51[b]	0.99
	2.32[b]	0.61	2.34[b]	0.62	2.32[b]	0.61	2.34[b]	0.53	2.32[b]	0.41
	heptad		heptad		heptad		heptad		heptad	

Table 3-10 *continued*

3.7 Consensus Sequences for IF Chain Types

Consensus sequences for the rod domain and for the H1 and H2 segments of the IF types Ib, IIa, IIb, III, IV-L, IV-M, IV-H and V are shown in Figure 3-4, as is the single type Ia sequence available. An overall consensus IF sequence is also shown. In those situations where several residues occur with approximately equal probability then each residue is listed. In some case, merely the residue type is listed, ie. apolar - Ap; acidic - Ac; basic - Ba. Residues that are conserved within a chain type but differ between the types may be considered as characteristic markers of these chains types. In addition, they clearly indicate probable and potentially important variations in molecular function and local structure among the types.

The homology study permits an analysis of the most conserved positions within the heptad substructure or residue types within each position. The results of this analysis are listed in Table 3-11 and confirm an earlier report by Parry and Fraser (1985) that the 'inner' positions a and d of the heptad were those most highly conserved, that the next most conserved positions were the e and g residues (ie, those immediately adjacent to a and d in axial projection) and finally those least conserved lay in the 'outer' positions of the coiled-coil, ie. b, c and f. This emphasizes that residues in position a and d, which are predominantly apolar, are crucial to the integrity of the coiled-coil structure and that residues in the e and g positions, which are commonly charged, have considerable importance in providing further stability to the structure and also in specifying the relative polarity of the constituent chains through the maximization of ionic interactions. Residues in positions b, c and f have particular importance in the process of molecular aggregation but are more easily able to tolerate sequence variations without losing their ability to specify precise modes of linear and lateral assembly. This is especially true for the outermost position f.

	Heptad Positions						
	a	b	c	d	e	f	g
Mixed homology table[a]	74.9	59.5	56.6	78.0	71.4	52.7	66.3
Identity homology table[b]	60.1	51.3	48.1	67.4	60.1	41.3	57.4
Acidic homology table[c]	1.0	28.5	29.4	7.3	25.3	22.0	30.2
Basic homology table[c]	12.2	15.1	11.9	2.2	29.7	17.6	19.7
Apolar homology table[c]	71.9	7.8	9.8	72.4	16.4	13.5	15.1

Table 3-11 Mean residue homology scores (h_r) calculated for each position in the heptad substructure using a selection of homology criteria. [a] See Table 3-1. [b] As for Table 3-1 but with zeros everywhere except along the diagonal. [c] See Table 3-2.

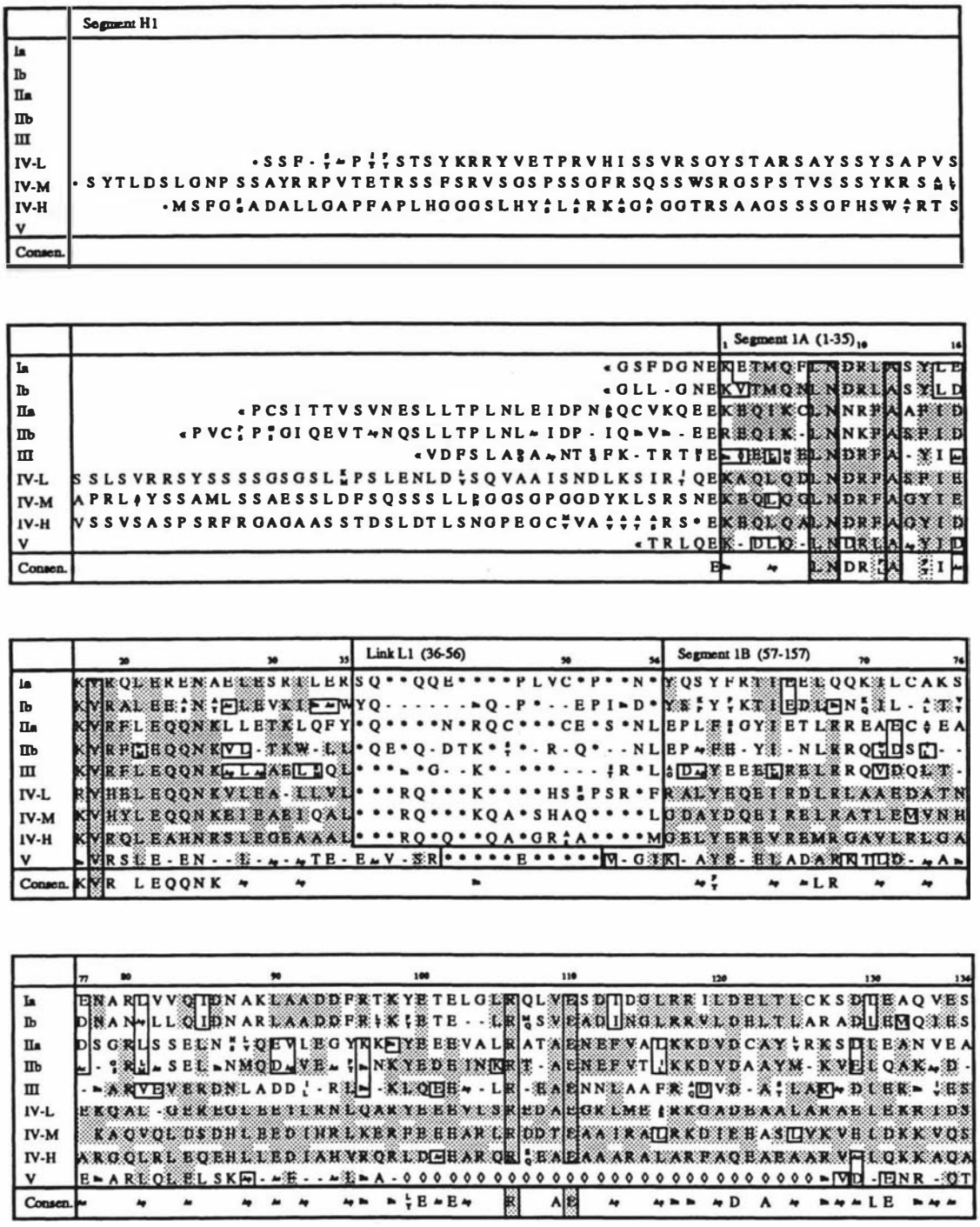


Figure 3-4 The consensus sequences for each of the IF chain types and subtypes are listed and aligned to aid comparison. The segment boundaries are also marked and extend from segment H1 through the rod domain to segment H2 (where present). Those residues that are conservatively substituted are shown boxed and those that are absolutely maintained in a chain type or subtype are shown shaded. Residues which are identical across all chain types are shaded and boxed. The overall consensus sequence (Consen.) is also listed. The symbols used in this figure have the following meanings: *, inserted blank; ?, incomplete sequence; •, chain termination; «, continuation towards the N-terminus; », continuation towards the C-terminus; –, position of low homology where a consensus cannot be assigned; 0, the site of a 70 residue replacement for residues 99-126 (ie. a net 42 residue insertion); Ac, acidic residues aspartic acid and glutamic acid; Ba, basic residues arginine and lysine; Ap, apolar residues alanine, isoleucine, leucine, methionine, phenylalanine, tyrosine and valine.

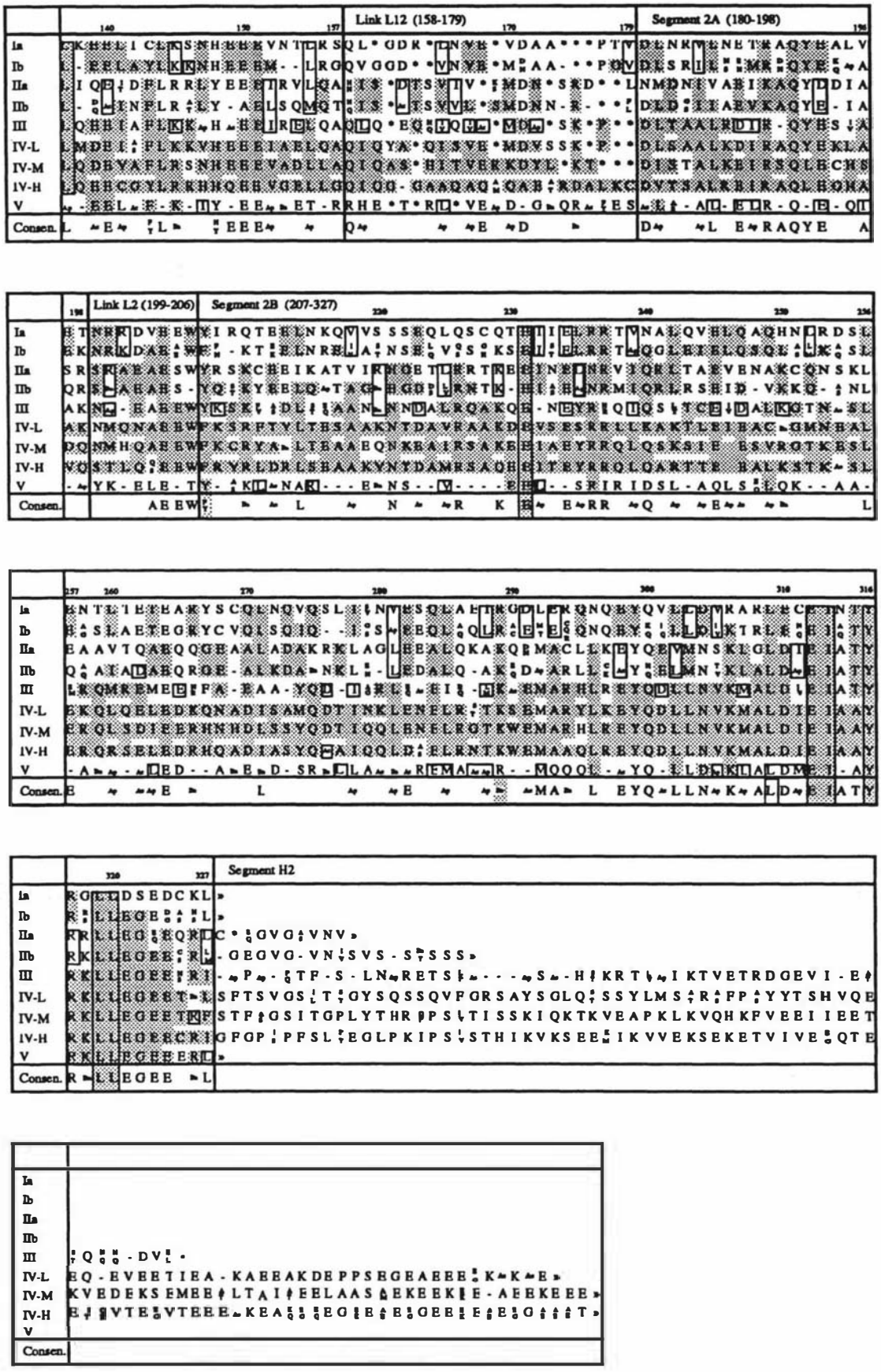


Figure 3-4 *continued*

Type I		Type II		Type III		Type IV		Type V	
Trp	65%	Phe	76%	Cys	90%	Trp	65%	Tyr	44%
Tyr	44%	Glu	54%	Tyr	42%	Glu	53%	Phe	33%
Glu	41%	Lys	53%	Gly	39%	Leu	47%	Leu	30%
Pro	41%	Tyr	45%	His	31%	Arg	36%	Glu	26%
Leu	38%	His	42%	Leu	30%	Ile	35%	Ile	21%
Gln	36%	Ile	42%	Arg	26%	Tyr	26%		
Arg	33%	Arg	41%	Glu	24%	Gln	24%		
		Ala	40%						
		Leu	38%						
		Gly	38%						

Table 3-12 Percentage occurrence of highly conserved residues in the rod domain ($h_r \geq 90\%$)

Specific amino acids that are highly conserved in some position in the rod domain sequence ($h_r \geq 90\%$) are listed in Table 3-12. Note the generally low values for type V. This is due to the relatively low homology between the lamins and the *Helix pomotia* B protein. Generally, the observations are consistent with the pattern of residues conserved in other families of related proteins. For example, tryptophan, tyrosine and phenylalanine are generally conserved at a high level, possibly as a consequence of their known role in providing regions of structure with a high stability (Huber, 1979; Burley and Petsko, 1985). Features more specific to individual IF chain types include the high percentage conservation of glutamic acid and proline residues in 'acidic' type I chains, of glutamic acid, lysine, arginine and histidine in 'neutral-basic' type II chains, and of cysteine and glycine in type III chains.

In conclusion the data again illustrate the probability that IF containing different chain types will exhibit polymorphism.

3.8 Summary

The homology study has revealed several novel aspects of IF structure as well as confirming other, established ideas. The general classification of IF proteins into several types is reflected in the difference in homology scores between sequences belonging to the same type and sequences from different types. In particular, neurofilament IF chains are confirmed to be distinct from the type III chains and have been classified as type IV. On the basis of homology, the neurofilaments have also been shown to comprise at least three sub-types, 'light' (NF-L), 'medium' (NF-M) and 'heavy' (NF-H). The delineation of keratins into two sub-types ('hard' and 'soft') on the basis of cell source and chemical character is also apparent in the segment

homology scores (h_s) for the rod domain segments (Section 3.3.1). Also, the relatively low homology between the type V lamin chains and the *Helix pomotia* B chain suggests that further subdivision of this class may be necessary.

An important conclusion that can be made from the differing degrees of homology exhibited by the various chain types is that IF composed of type I/type II chains, type III chains, type IV chains or type V chains may have slightly different structures (ie, exhibit polymorphism). This may be manifested in minor variations of molecular packing or in the precise geometry of the surface lattice of the IF on which the tetrameric structural unit is known to be positioned (see Chapter 1). Type V IF molecules have already been shown to form a variety of related polymorphic forms (Zackroff *et al*, 1984; Aebi *et al*, 1986; Parry *et al*, 1987a; also see Section 4.3).

There are highly significant peaks in the Fourier transforms of the arrays of the conserved acidic, the conserved basic and the conserved large apolar residues within the rod domains of a particular chain type. These periods are related to one another and indicate a long-range period comparable (although not related) to the length of the segment from which the array originated. Primarily, however, the most significant and common peaks (~9.55 and ~9.85 residues in segments 1B and 2 respectively) parallel exactly those previously discovered in the linear distribution of all acidic, basic and apolar residues (considered separately). The evidence is now overwhelming that the conservation of these residues is crucial to specifying the structure of the coiled-coil and the modes of aggregation of IF molecules through ionic interactions.

Subdomains within the amino- and carboxy-terminal domains have been more fully delineated than has been possible previously. The original scheme, which was developed for the keratins, is not as relevant for the other IF chains although the homologous regions, H1 and H2 can be assigned. New evidence has been presented for the existence of a small H1 subdomain in type Ib mammalian chains. In addition, both H1 and H2 subdomains are present in the type IV chains and in the type V lamins.

The keratin chains from amphibia are readily classified according to the type I/II nomenclature derived initially for mammalian keratins. However, the type I *Xenopus* chains show less homology to their mammalian counterparts than do the type II chains. This is also true of the hard mammalian keratins when compared to the epidermal keratins and may be an indication of a greater conservation in the primary structure of the rod domains in the type II chains compared to that in the type I chains.

Consensus sequences for the various IF subtypes have been presented. Highly

conserved residues imply positions crucial either to the structure or function of IF in general; other residues that are less well conserved may nonetheless represent important elements of specialized structure or function. Residues within the a and d positions are shown to be conserved at a higher level than those in the other positions: this reflects their importance in specifying the coiled-coil structure. The e and g positions, important for the alignment of chains in the molecule, are the next best conserved followed by the 'outer' positions, b, c and f. It is apparent that functional and structural diversity can be built into the rod domain without upsetting the fundamental organization of the IF molecule.

More data, especially the sequences of hard α-keratins, neurofilaments and type V proteins, are required to verify some of the conclusions presented in this Chapter and to extend others.

4. SECONDARY AND TERTIARY STRUCTURE

Ionic interactions have been shown to play an important but somewhat different role from apolar and hydrogen-bond interactions in the aggregation of fibrous biopolymers (see, for example, Hulmes *et al*, 1973; Parry, 1975, 1981; McLachlan and Stewart, 1976). The range of influence for ionic interactions is much greater than that for apolar interactions - the electrostatic force which comprises the ionic interaction is inversely related to the separation of the participating charged atoms (ignoring the effects of the intervening medium) whereas the van der Waals attraction that mediates apolar interactions attenuates by the sixth power of the separation. Also, charged residues generally have longer and more flexible sidechains than the apolar residues and these sidechains are often situated on the periphery of the molecule. In contrast, apolar residues tend to be 'internalized' to shield them from the polar, aqueous environment of a biological system. Hence, the ionic interactions cause adjacent molecules to come to a close and suitable conjunction prior to the final 'docking' provided by apolar and hydrogen-bond interactions. It follows then that the ionic interactions are often crucial in specifying the mode of chain and molecular assembly and that the apolar residues serve chiefly to stabilize the molecules and maintain them in a suitable configuration once they have been assembled. For this reason, theoretical studies on the hierarchical structure of IF and other fibrous proteins have paid particular attention to the positions of the charged residues in the amino acid sequences in order to determine the most likely mode of assembly based on mutual attraction between oppositely charged residues.

Interchain ionic interactions are calculated for rat peripherin, the lamin proteins and the *Helix pomotia* B protein in order to determine the likely modes of chain aggregation. These are compared with previous studies on other IF chains. One-dimensional ionic interactions between molecules are also studied for these proteins. In particular, the assembly of chains into the molecule and the determination of the likely mode of molecular aggregation for the nuclear lamin proteins is examined. Maxima in the ionic interaction curves for lamin molecules are used to find likely axial staggers between the molecules *in vivo* and *in vitro*, and which in turn lead to a model that is consistent with the electron microscope data obtained from negatively-stained lamin paracrystals. The models are also compared with data from filamentous structures of lamin molecules and the *in vivo* lamin meshwork. Suitable physical data, such as the axial

periodicities that may be derived from paracrystalline structures, are currently very scarce for IF chains and similar studies have only been undertaken on a GFAP protein fragment (Stewart *et al*, 1989b).

Finally, an attempt is made to extend further the one-dimensional techniques for calculating inter-molecular ionic interactions to a more general, three-dimensional method. Preliminary results of this study are reported and show that appreciable progress has been made towards refining the technique into a useful tool for predicting molecular structure based on amino-acid sequence information.

4.1 1D Ionic Interactions Between Chains

Interchain ionic interactions have been studied previously for a selection of IF proteins (for example Parry *et al*, 1986). In all cases a parallel in-register arrangement of chains has been preferred, a result which is in agreement with the physico-chemical data described in Chapter 1. In this section, the interchain ionic interactions which stabilize homodimeric molecules are examined for the recently sequenced IF proteins, peripherin, human lamins A and C, *Xenopus* lamins A and B and the *Helix pomotia* B protein.

Residues in heptad positions a and e on one chain may interact with residues on the other chain in the axially adjacent g position, and similarly for the d and g positions with the axially adjacent e position. Only segments of unbroken heptad substructure are examined in this study and hence the rod domain of the IF chains is analysed piecemeal and the results for each coiled-coil segment are summed. Segment 2B, for example, has been subdivided into two sections of continuous coiled-coil since a discontinuity in heptad phasing occurs near the centre of the segment: the amino-terminal piece, 2B1, is 58 residues in length and the carboxy-terminal piece, 2B2, is 62 residues. The linker L1 of the lamin and *Helix* B proteins is reduced to a discontinuity in the coiled-coil structure of segments 1A and 1B and so the lengths of these segments are 41 and 147 residues respectively. In types I-IV IF chains segment 1A is 35 residues in length and segment 1B is 101 residues.

Chain segments are staggered axially in integral heptad steps so as to maintain the internal apolar core of the coiled-coil dimers. Ionic interactions between residues in the appropriate heptad positions of adjacent chains are scored according to their potential to attract (+1) or repel (-1) each other, or are scored zero if they are considered non-interacting. Scores are made for each stagger; in addition a total score is calculated. The results are given in Table 4-1. In all cases the highest positive score is found for parallel in-register chains. This is consistent with the generalized model for the IF

Stagger (heptad)	Peripherin: Parallel Chains 1A	1B	2A	2B1	2B2	Total
0	2	10	0	4	2	18
±1	2	4	0	0	0	6
±2	-4	3	-1	0	0	-2
Stagger (heptad)	**Peripherin: Antiparallel Chains 1A**	**1B**	**2A**	**2B1**	**2B2**	**Total**
-2	4	3	0	0	-2	5
-1	-1	-4	-1	0	-5	-11
0	-2	-2	0	-2	-2	-6
1	-4	-7	0	-5	-5	-21
2	4	-6	0	2	-2	-2

Table 4-1a Interchain ionic interactions for rat peripherin. The chains are staggered by integral multiples of the heptad (ie, seven residues) in order to maintain the integrity of the coiled-coil structure. Segment 2B has been analysed in the two pieces of continuous heptad structure either side of the discontinuity near the centre of 2B. The length of 2B1 and 2B2 is 58 and 62 residues respectively.

Stagger (heptad)	Human Lamins A and C: Parallel Chains 1A	1B	2A	2B1	2B2	Total
0	6	18	0	4	8	36
±1	-2	1	0	1	-4	-4
±2	-1	4	0	1	-4	8
Stagger (heptad)	**Human Lamins A and C: Antiparallel Chains 1A**	**1B**	**2A**	**2B1**	**2B2**	**Total**
-2	-1	-2	0	0	6	-3
-1	-2	1	-1	-2	-9	-13
0	2	-6	0	2	8	6
1	2	6	0	-1	-5	2
2	-6	0	0	4	0	-2

Table 4-1b Interchain ionic interactions for human lamins A and C which have a common rod domain sequence. See the caption to Table 4-1a for details. Also, due to the continuation of heptad structure into the Link region L1 the lengths of segments 1A and 1B are 41 and 147 residues respectively.

Stagger (heptad)	*Xenopus* lamin A: Parallel Chains 1A	1B	2A	2B1	2B2	Total
0	6	18	0	4	8	36
±1	-2	1	0	1	-4	-4
±2	-1	4	0	1	-4	8
Stagger (heptad)	*Xenopus* lamin A: Antiparallel Chains 1A	1B	2A	2B1	2B2	Total
-2	-1	-2	0	0	6	-3
-1	-2	1	-1	-2	-9	-13
0	2	-6	0	2	8	6
1	2	6	0	-1	-5	2
2	-6	0	0	4	0	-2

Table 4-1c Interchain ionic interactions for *Xenopus* lamin A. See the caption to Table 4-1b.

Stagger (heptad)	*Xenopus* lamin B: Parallel Chains 1A	1B	2A	2B1	2B2	Total
0	2	10	0	2	8	22
±1	0	2	0	-1	-3	-2
±2	0	7	0	2	3	8
Stagger (heptad)	*Xenopus* lamin B: Antiparallel Chains 1A	1B	2A	2B1	2B2	Total
-2	0	-6	0	-2	4	-4
-1	0	0	-1	0	-8	-9
0	1	0	0	2	6	9
1	-2	4	0	-1	-5	-4
2	-3	2	0	0	-2	-3

Table 4-1d Interchain ionic interactions for *Xenopus* lamin B. See the caption to Table 4-1b.

Stagger (heptad)	*Helix pomotia* B: Parallel Chains 1A	1B	2A	2B1	2B2	Total
0	2	12	0	4	0	18
±1	2	-8	0	-1	2	-5
±2	-2	7	0	1	-3	3
Stagger (heptad)	*Helix pomotia* B: Antiparallel Chains 1A	1B	2A	2B1	2B2	Total
-2	3	-6	0	0	-4	-7
-1	-3	7	-1	-2	5	-6
0	-7	-4	0	2	-6	-15
1	2	0	0	-2	0	0
2	1	2	0	2	-4	1

Table 4-1e Interchain ionic interactions for *Helix pomotia* B. See the caption to Table 4-1b.

molecule and provides further support for it. Furthermore, the interchain ionic interactions for the human and *Xenopus* lamin A proteins are identical. Examination of the charged residues in the a, d, e and g positions of the heptad-containing regions for both sequences reveals that only six substitutions are made and in all cases the charge character is conserved (ie, the basic residues arginine and lysine may be interchanged as can aspartic acid and glutamic acid; see Figure 2-2).

The total scores for parallel chains are summarized in Table 4-2 along with those for several other IF proteins. Peripherin (Table 4-2a) is very similar to other type III IF. The lamin A chains (Table 4-2b) reveal particularly high scores for the in-register arrangement: they are at least twice as great as for the type I to IV chains. This is a reflection in part of the greater number of charged residues in the rod domain of lamin A (for example, human lamin A has 118 charged residues in segments 1B and 2 combined, compared to an average of 90 for types I-IV IF: Parry *et al*, 1986) but also indicates that a greater proportion of these charged residues are in a position to make interchain ionic bonds than for other IF. Table 4-3 lists the total ionic interactions per heptad pair for the coiled-coil domains of segment 1, segment 2 and the rod. This may be viewed as an "interaction density". Only the scores for the parallel in-register arrangement of chains are included. Segment 1 generally has a much greater density of interchain ionic interactions than segment 2 although in the case of *Xenopus* lamin B the scores are similar. The lamin A proteins show the greatest interaction density (0.89 in segment 1) and the *Helix pomotia* B protein has the lowest (0.20 in segment 2). One conclusion is that lamin A molecules inherently have much more stability than those of any type I-IV IF.

Stagger (heptad)	H50K	H56K	CGD	HELV	MGFAP	PNF-L	Peripherin
0	8	12	14	16	18	12	18
±1	5	6	2	5	1	-1	6
±2	4	3	-4	0	-8	1	-2

Table 4-2a Interchain ionic interactions for parallel chains from a selection of type I-IV IF proteins (Parry *et al*, 1986) including peripherin. H50K and H56K are type I and type II IF chains respectively from human epidermal keratin; CGD - chicken gizzard desmin, HELV - hamster eye lens vimentin, MGFAP - mouse glial fibrillary acidic protein and peripherin are type III IF chains; and PNF-L - porcine neurofilament light chain is a type IV IF chain. Sources are listed in Table 2-1.

Stagger (heptad)	Human lamins A and C	*Xenopus* lamin A	*Xenopus* lamin B	*Helix pomotia* B
0	36	36	22	18
±1	-4	-4	-2	-5
±2	8	8	8	3

Table 4-2b Interchain ionic interactions for parallel chains from V IF. Human lamins A and C share a common rod domain sequence and hence have identical interaction scores. The rod domain sequence for *Xenopus* lamin A is also very similar to the human lamins A and C.

Protein	1A+1B	2A+2B1+2B2	Total
Peripherin	12 in $19^{3}/7 = 0.62$	6 in $19^{6}/7 = 0.30$	18 in $39^{2}/7 = 0.46$
Human Lamins A&C	24 in $26^{6}/7 = 0.89$	12 in $19^{6}/7 = 0.60$	36 in $46^{5}/7 = 0.77$
Xenopus Lamin A	24 in $26^{6}/7 = 0.89$	12 in $19^{6}/7 = 0.60$	36 in $46^{5}/7 = 0.77$
Xenopus Lamin B	12 in $26^{6}/7 = 0.45$	10 in $19^{6}/7 = 0.50$	22 in $46^{5}/7 = 0.47$
Helix pomotia B	14 in $26^{6}/7 = 0.52$	4 in $19^{6}/7 = 0.20$	18 in $46^{5}/7 = 0.39$

Table 4-3 Interchain ionic interactions per heptad pair for the proteins listed in Table 4-1. Only interactions for parallel in-register chains are included. The heptad regions in segments 1 and 2 have been combined separately (1A+1B and 2A+2B1+2B2 respectively) and together (Total). The count of heptads for each combined group is given in fractions of seven for conciseness.

4.2 1D Ionic Interactions Between Molecules

A very simple model of fibrous protein structure has been used successfully to relate the amino acid sequence of collagen to the relative axial positions of type I collagen molecules in the fibrils found *in vivo* (Hulmes *et al*, 1973). The mechanism on which the preferred mode of assembly was based was found to be attributable largely to ionic interactions with a smaller contribution from apolar interactions; hydrogen bond interactions between hydrophilic residues were not studied. In the model of Hulmes *et al* (1973), the amino acid residues in an α-chain were spaced evenly in a linear, one-dimensional array. Two such arrays, each representing a single chain in adjacent molecules, were aligned either mutually parallel or antiparallel. Starting from a position of maximum relative stagger (ie, minimum overlap) the arrays were moved past each other and a score of possible apolar and ionic interactions between residues on the different chains was generated for each position of relative axial displacement (or stagger). Charged residues lying within three linear residue translations of each other were assumed to be able to interact; for apolar residues the interaction range was reduced to two. The scores revealed a general trend toward greater potential for interactions as the molecules become increasingly overlapped, as was expected. However, overlaid on this background was a regular pattern of peaks with notable maxima occurring at intervals of 234 residues (ie, ~67 nm), a value previously determined from X-ray diffraction and electron microscope data and which corresponds to the axial stagger between molecules in the fibril (Bear, 1944; Hodge and Petruska, 1963).

The method of Hulmes *et al* (1973) for calculating the potential number of ionic interactions between charged residues from different molecules of collagen has also been applied to several of the α-fibrous proteins such as tropomyosin (Parry, 1975; McLachlan and Stewart, 1976), myosin (Parry, 1981; McLachlan and Karn, 1983), paramyosin (Cohen *et al*, 1987) and IF (Crewther *et al*, 1983; Fraser *et al*, 1985, 1986; Parry *et al*, 1987a). Fraser *et al* (1985) studied all possible combinations of the IF rod domain segments 1B and 2: 1BU-1BU, 1BU-1BD, 2U-2U, 2U-2D, 1BU-2U, 1BU-2D where 1B and 2 are the major rod domain segments and U (up) and D (down) designate the chain directions with regard to the progression from the N- to the C-terminal end of the segment. Note that U-U and D-D are equivalent as are U-D and D-U. Due to the large number of maxima in the resulting interaction curves, a technique was employed for estimating the quantiles of scores from 1000 randomized sequences at each stagger of the chains. Only those maxima above some specified cutoff were deemed to be significant. The conclusion of this study was that significant

maxima in the interaction curves were indeed consistent with the dimensions of the surface lattice of α-keratin as determined by X-ray diffraction analysis, and favoured models were proposed as a basis for further studies.

The one-dimensional ionic interaction study on IF molecules was extended by Parry *et al* (1987a) to include the recently sequenced human lamin proteins (McKeon *et al*, 1986; Fisher *et al*, 1986). Lamin presented a simpler case than other IF in that the entire rod domain was predicted to be α-helical and hence the spacing between residues in the entire rod domain could be assumed constant. This meant that the analysis need only be performed for two cases - the rod domains of parallel (U-U) and antiparallel (U-D) molecules - rather than for combinations of smaller segments of the rod domains of the molecules as has proved necessary for other IF types. Also, the lamin molecules were assumed to be homodimeric in the rod domain as the sequences of human lamins A and C are identical from the N-terminus right through to a point in the non-helical C-terminal domain. The ionic interaction curves for lamin molecules are shown in Figure 4-1.

The general trend for the two interaction curves is an approximately linear increase in interaction score with increasing overlap - an indication of the relative evenness in the distribution of charged residues within the rod domain. The interaction curve for parallel homodimeric molecules (for example, Figure 4-1, U-U) shows a substantial dip in the score at zero stagger. This is due to the exact alignment of charged residues of the same sign from both homodimers which effectively reduces the chances of finding an oppositely charged residue within the interaction range - in this case the interaction range is ±1 residue translation and hence the reduction is by a factor of approximately one in three.

The relationship between ionic interaction score and overlap is not sufficiently regular to allow significance to be assigned to peaks in the curve on an analytical basis. This is especially apparent in the top graph of Figure 4-1 at zero stagger (as described above) and in the lower graph of Figure 4-1 where there are several wobbles in the general trend for positive staggers. The empirical strategy employed by Fraser *et al* (1985) was also used by Parry *et al* (1987a) in analysing the ionic interaction curves of lamin although it was revised to take into account one of the grosser non-uniformities in the distribution of residues in the heptad substructure (described below). Also, unlike keratin, the lamin molecules were assumed to be homodimeric and consequently, when calculating significance levels, the same randomized sequence was used for both chains of both molecules.

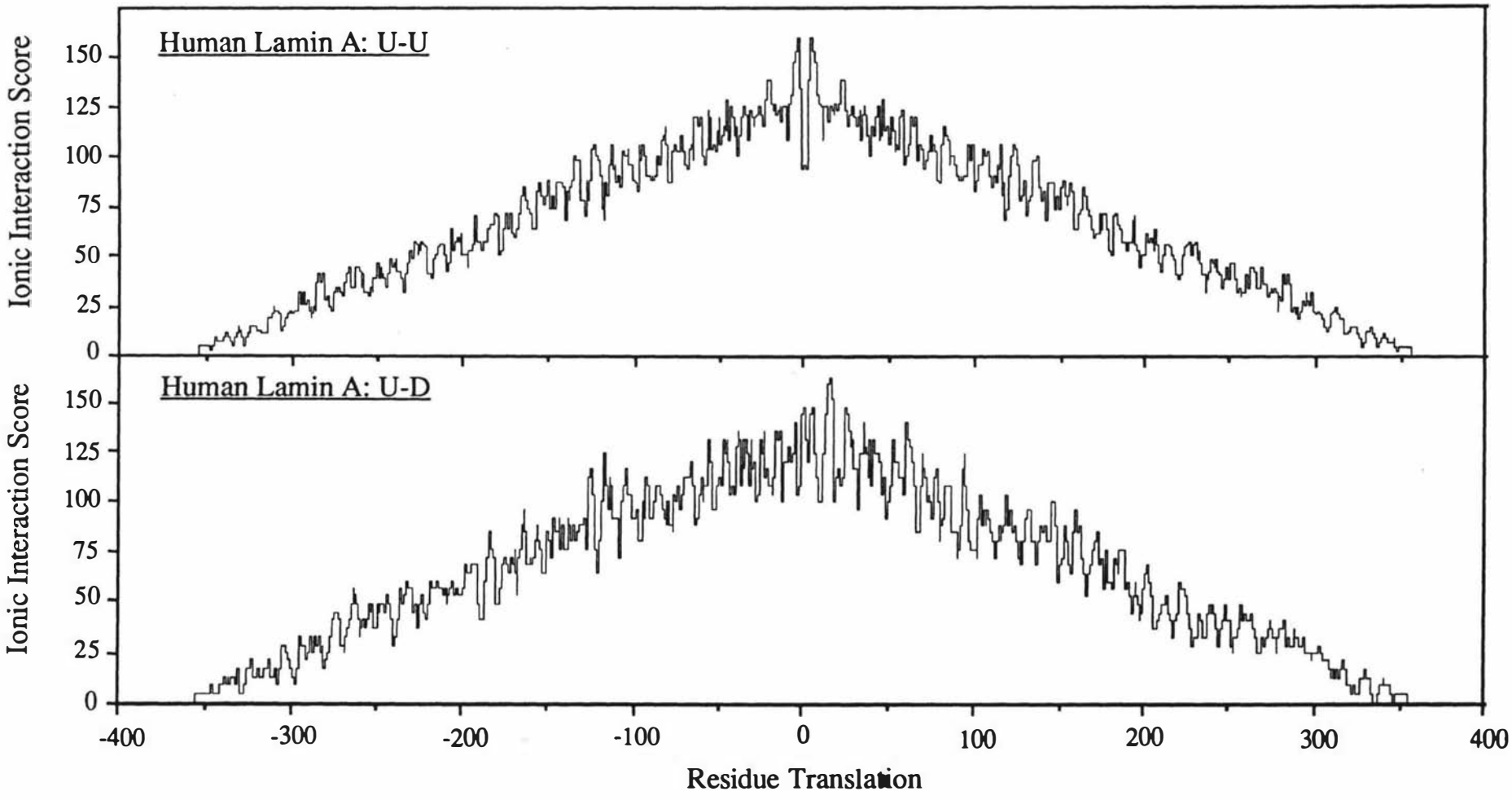


Figure 4-1 Ionic interaction curves for parallel (U-U) and antiparallel (U-D) rod domains of human lamin A. The curves for human lamin C are the same as these because the rod domain sequences are identical. Figure adapted from Parry *et al* (1986).

The original method of Trajstman and Lucas for identifying interaction scores that were above some predetermined limit of significance was described by Fraser *et al* (1985). In this procedure the residues of each of the four chains were randomly reassigned to positions within the sequence and a new interaction curve was calculated in the same manner as for the originals. This randomization and recalculation was carried out 1000 times and an empirical distribution of scores was generated for each relative stagger of the molecules. From each distribution a quantile for the particular stagger was determined (S_0) and then compared to the actual ionic interaction score of the rod domain sequences (S) at the same stagger in order to determine if the actual score was significant or not. The cutoff level used by Fraser *et al* (1985) was 0.975.

A modified version of this method was developed by Parry *et al* (1987a) in order to take into account the high occupancy rate (~75%) of the a and d positions in the heptad by apolar residues (an essential feature of the coiled-coil structure, Parry and Fraser, 1985). The new procedure involved grouping the residues in the a and d positions separately from those in the other positions. Reassignment on a random basis was made independently within the two groups so that the a and d positions remained ~75% apolar. An additional refinement to the method was to generate only one randomized chain for each of the 1000 iterations and to infer the other three chains from the first rather than to generate four separately randomized chains as per Fraser *et al* (1985). This is appropriate for the type III-V chains where two homodimeric molecules (ie, four identical chains) are being examined. As before, a level of 0.975 was used as a test of significance.

Other refinements along similar structural lines are possible for this analysis. However, the gross simplification of the chain structure from a supercoiling of α-helices to a pair of linear arrays probably means that such efforts would not significantly improve the accuracy of predicting the packing modes available to the molecules *in vivo*. Another approach which could improve the accuracy of the method is to weight differently the four combinations of ionic interactions: Asp-Lys, Asp-Arg, Glu-Lys, Glu-Arg. This could be done according to the frequency of such combinations in a database of proteins with known structure. However, insufficient data are available at present to warrant this degree of sophistication.

The modelling of a three-dimensional coiled-coil as a simple linear array of residues does, of course, ignore the radial and azimuthal information about the residues and only utilizes their axial positions. This simplification of the chain structure has been made primarily to render the analysis computationally tractable. However, the assumption that the axial coordinate of the residues is of greater relevance for this

analysis than the other coordinates does have some justification especially with regard to the range of influence of the relatively long and flexible sidechains of the charged residues. These particular sidechains may extend for a distance comparable to the diameter of the coiled-coil (and the equivalent structure in collagen) and it is conceivable that such arm-like sidechains on the outside of the cylinder-like structure of the coiled-coil will have more freedom of movement in planes transverse to the cylinder axis than in a direction parallel to the axis. Good evidence to support this view has been obtained from nuclear magnetic resonance on the peptide backbone of collagen (Torchia and Van der Hart, 1976). A restriction on the axial motion of sidechains would imply that precision in specifying the axial coordinate of a residue is of greater importance than for the other coordinates. This is especially true for determining the axial range of influence of the sidechains. For this reason, the one-dimensional simplification has justification beyond that of computational tractability.

Maxima from the interaction curves for the rod domains of human lamin A molecules (Figure 4-1) that are above the 0.975 quantile are listed in Table 4-4. One of the highest peaks occurs for a stagger of Δz(U-U) equal to ± 4 residues. Although this is unimportant for the purposes of building a model to describe the axial aggregation of molecules it provides an important mechanism for the lateral accumulation of molecules to generate width for the final structure. The rates of lateral and axial accretion of molecules are not known, and neither is the diameter-limiting mechanism understood. It is nonetheless worthwhile to point out that a mechanism does exist for lateral assembly and that this mechanism, based as it is on ionic interactions, is similar in principle to that proposed for the axial aggregation of the lamin molecules.

Ionic interaction curves have also been calculated for the rod domains of rat peripherin molecules and significant scores are listed in Table 4-5. This type III IF protein is predicted to have the "classical" IF rod domain structure which includes the usual interruptions in the coiled-coil structure and the ionic regularities, namely the link regions L1 and L12. Hence the sequence has been subdivided into the two major coiled-coil domains, segments 1B and 2, and all combinations of segments and relative orientations have been analysed. The scores may be compared to those for the other type III IF proteins, chicken gizzard desmin and hamster eye lens vimentin, which have been studied in a similar manner by Fraser *et al* (1985).

The scores (S) and quantile scores (S_0) are generally higher for peripherin than the other type III chains in interactions involving segment 1B but are similar for segment 2-segment 2 arrangements. The differences in the numbers of charged residues between these sequences appears to be sufficient to explain this discrepancy:

Parallel molecules			Antiparallel molecules		
Δz(U–U)	S	S_0	Δz(U–D)	S	S_0
±345	8	8	-264	56	52
±311	24	24	-184	84	84
±297	32	30	-164	96	92
±294	32	32	-125	116	112
±284	40	36	-117	124	112
±281	40	38	-114	112	112
±267	44	42	-105	116	116
±262	44	44	15	164	148
±229	58	58	24	148	144
±193	70	70	60	140	132
±164	84	84	94	124	124
±136	100	96	145	100	100
±123	106	102	160	96	96
±114	106	104	282	40	40
±4	158	146	288	36	36
			340	12	12

Table 4-4 Significant ionic interactions between the rod domains of human lamin A molecules (or lamin C molecules) as a function of relative axial stagger (see Figure 4-1). Δz(U-U) and Δz(U-D) signify the relative staggers between parallel and antiparallel chains respectively. Only those scores (S) greater than or equal to the 0.975 quantile (S_0) are listed.

peripherin has 50 charged residues in segment 1B and 52 in segment 2 (1B: 30 acidic – 20 basic, 2: 30 acidic – 22 basic); chicken gizzard desmin has 42 and 52 charged residues in segments 1B and 2 respectively (1B: 25-17, 2: 30-22); and hamster eye lens vimentin has 45 and 52 charged residues in segments 1B and 2 respectively (1B: 29-16, 2: 31-21). The total counts for the segment 2 sequences are identical amongst the three proteins and hence the similarity in scores. However, segment 1B of peripherin contains more charged residues and the chances of interactions being scored are greater.

Ionic interaction curves have also been calculated for *Xenopus* lamins A and B and the *Helix pomotia* B proteins. These are shown in Appendix D along with tables of the significant peaks. Until such time as new structural data have been obtained for these proteins no further progress can be made in establishing a model for their packing (such as that described in the next section for the human lamin proteins). However, in the case of *Xenopus* lamin A, which is highly homologous to the human lamins A and C in sequence, the following study may be expected to apply equally well to its

Parallel molecules				Antiparallel molecules			
	Δz(U–U)	S	S_0		Δz(U–D)	S	S_0
1BU-1BU	±54	32	30	**1BU-1BD**	3*	68	60
	±4*	58	58				
2U-2U	±138	6	6	**2U-2D**	-65	44	36
	±113	20	18		-34*	48	48
	±103	24	20		4*	60	56
	±93	24	24		13*	64	52
	±72*	32	30		34*	44	44
	±63*	34	42		43*	56	40
	±34*	44	42		53*	40	40
	±26*	46	44		64*	36	36
	±5*	64	50		74*	32	32
					100	28	24
					110	24	20
					121	16	16
					132	12	12
					135	8	8
					143	4	4
1BU-2U	-133	10	10	**1BU-2D**	-85*	32	32
	-94*	34	28		-34*	50	48
	-82*	32	32		-25*	46	46
	-71*	38	38		4*	48	46
	-55*	46	44		33*	36	34
	-43*	48	48		44*	32	30
	-14	48	48		53*	26	26

Table 4-5 Significant ionic interactions between rod domain segments of peripherin molecules as a function of relative axial stagger (see Appendix D, Figure D-1). Δz(U-U) and Δz(U-D) signify the relative staggers between parallel and antiparallel chains respectively. Only those scores (S) greater than or equal to the 0.975 quantile (S_0) are listed. Those staggers that correspond to maxima in the ionic interaction curves for the type III proteins studied by Fraser *et al* (1985) are marked with an asterisk.

molecular aggregation as to that of the human lamins.

4.3 Modelling Lamin

A number of studies on *in vivo* and *in vitro* assemblies of lamin IF have been reported, but data that prove of particular relevance to this study are as follows:

1) Lamins prepared under conditions of low ionic strength from isolated nuclei or

from cytoskeletal preparations form paracrystals with a well defined banding pattern. The period originally reported for this pattern was 37-40 nm (Zackroff *et al*, 1984; Goldman *et al*, 1986). However, Parry *et al* (1987a) repeated these measurements on a number of negatively-stained paracrystals and showed that the true axial periodicity was 48±2 nm - a value somewhat greater than that previously stated but which nonetheless appeared to be consistent with the original figures of Zackroff *et al* (1984).

2) Aebi *et al* (1986) studied the nuclear lamina from isolated nuclear envelopes of *Xenopus* oocytes and revealed the presence of an orthogonal lattice of side 52 nm. In addition, a 25 nm beading was observed in Triton-treated preparations of the *Xenopus* lamin meshwork which had been washed thoroughly in low ionic strength buffer prior to freeze-drying and metal shadowing. This 25 nm beading was also observed in filamentous assemblies formed *in vitro* from 1:1 mixtures of purified rat liver lamins A and C dialysed against low ionic strength buffer. The 25 nm period of the beading is almost exactly half the period of the lamina meshwork (52 nm) and the negatively-stained paracrystals (48 nm).

Parry *et al* (1987a) suggested that the packing of the molecules in the paracrystalline structures, the *in vivo* lattice and the *in vitro* filamentous structures were related. In particular they suggested that the *in vitro* filamentous assemblies were laterally compressed versions of the *in vivo* lattice. In order to investigate this proposal, model building studies were performed which incorporated the observed banding structure data and the likely relative staggers of lamin molecules deduced from the ionic interaction studies (Table 4-4). The observed electron microscope data from measurements made on the negatively stained paracrystals (Parry *et al*, 1987a) were:

1) The axial period of the alternating light-staining and dark-staining bands was found to be 48±2 nm. This value is a little lower than the lattice spacing of 52 nm observed by Aebi *et al* (1986, see above) in the nuclear lamina of *Xenopus* oocytes. The difference is probably attributable to shrinkage effects from sample preparation methods used in electron microscopy and also to periodic variations in the structural regularity and stability of the assembly rather than to alternate modes of packing.
2) The relative widths of the light- and dark-staining bands were 0.49±0.04: 0.51±0.04 respectively.

In addition, four other observations of relevance were noted (see Figure 4-2 and also Figures 1 and 2 of Parry *et al*, 1987a):

3) The dark-staining band corresponded to a region where the molecular packing was of relatively low density. This was particularly apparent in electron micrographs of small, newly-forming lamin paracrystals where individual rod-like

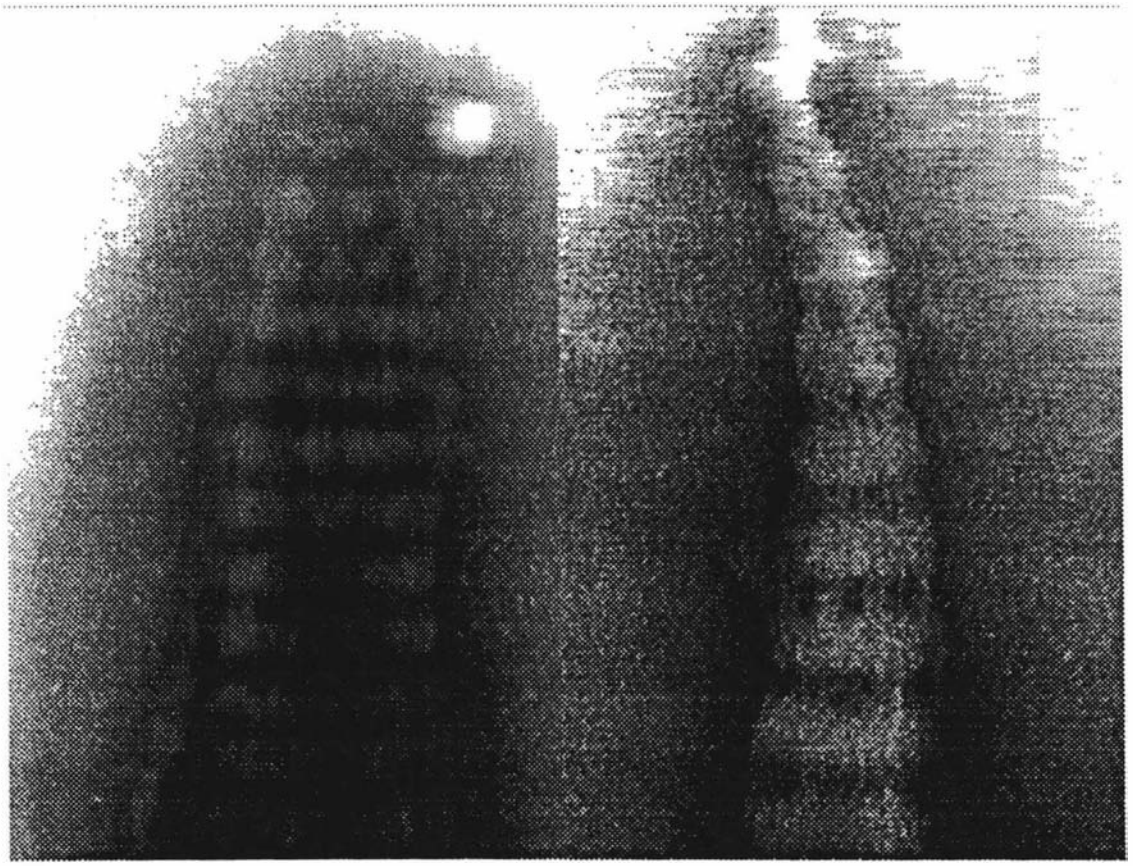


Figure 4-2 Electron micrographs of negatively-stained human lamin A paracrystals. Globular domains are apparent in the paracrystal on the left towards the end of the structure. These are assumed to correspond to the large C-terminal domains of the molecule. A smaller, newly forming paracrystal is shown on the right (not to scale) and reveals rod-like threads traversing the dark bands. Figures courtesy of Drs A.E. and R.D. Goldman.

threads could be seen. These threads were interpreted as being the rod domains of the lamin molecules. A greater density of proteinaceous material was evident in the light-staining bands of the paracrystals and this was probably due in part to the large C-terminal domains of the molecules.

4) The small, newly forming paracrystals revealed evidence of dihedral symmetry. This was manifested by the symmetry of the banding pattern and by the light-staining bands terminating both ends of the structures. This latter point also indicated that nuclear lamin molecules must terminate at, or close to, the boundaries of these bands.

5) Arrays of molecules in larger paracrystals also showed evidence of dihedral symmetry through the appearance of globular attachments at both ends of the structures. The globules were believed to be the large C-terminal domains of the lamin molecules and were confined to the light-staining bands (see, for example, Figure 3 of Zackroff *et al*, 1984).

6) No axial substructure is apparent in either the light- or dark-staining bands and it is reasonable to assume that the packing within both bands is uniform at the low

resolution available.

The evidence above suggests that the molecules in the paracrystals are arranged in an antiparallel manner and that they contain equal numbers of oppositely directed molecules. These data were used in conjunction with data from the intermolecular ionic interaction studies to generate a model for the molecular aggregation of human lamin A.

Several parameters used in the model-building process must be defined. Distances along the axis of the paracrystal are measured in terms of 'residue translations' (rather than nanometres) where a residue translation is defined as the average axial rise per residue in an α-helical coiled-coil conformation. The value used here is 0.1485 nm (see, for example, Fraser *et al*, 1972 and Fraser and MacRae, 1973b). Other parameters are defined as follows (see Figures 4-3 and 4-4) :

l = length of the rod domain of the lamin molecules. There are 356 residues in the rod domains of human lamins A and C (McKeon *et al*, 1986; Fisher *et al*, 1986; Parry *et al*, 1986).

p = periodicity of the paracrystals. In this case the period is taken as 48 nm, or in terms of residue translations:

$$\frac{48 \text{ nm}}{0.1485 \text{ nm res}^{-1}} = 323 \text{ residues}$$

This value is also considered equivalent to the lattice spacing observed in *Xenopus* oocytes.

s = smallest relative axial stagger between a dimer of antiparallel molecules that generates the putative four chain asymmetric unit

$\mathbf{d_n}$ = effective axial length of the N-terminal non-α-helical domain of the lamin molecule in the negatively-stained paracrystal.

$\mathbf{d_c}$ = effective axial length of the C-terminal non-α-helical domain of the lamin molecule in the negatively-stained paracrystal.

Several intermolecular interactions are made by the asymmetric unit and these may be associated with molecules having the following relative staggers (see Figures 4-3 and 4-4):

(i) $\Delta z(U\text{-}U) = \mathbf{p}$ (twice)

(ii) $\Delta z(U\text{-}D) = \mathbf{s}$

(iii) $\Delta z^{*}(U\text{-}D) = \mathbf{s} - \mathbf{p}$

(iv) $\Delta z^{\dagger}(U\text{-}D) = \mathbf{s} + \mathbf{p}$ provided $\mathbf{s} + \mathbf{p} < \mathbf{l}$

Δz refers to the difference in the axial coordinate between rod domains of different molecules and is measured in residue translations (Fraser *et al*, 1985). U and D refer to the 'Up' and 'Down' orientations of the molecules as defined in Section 4.2. The

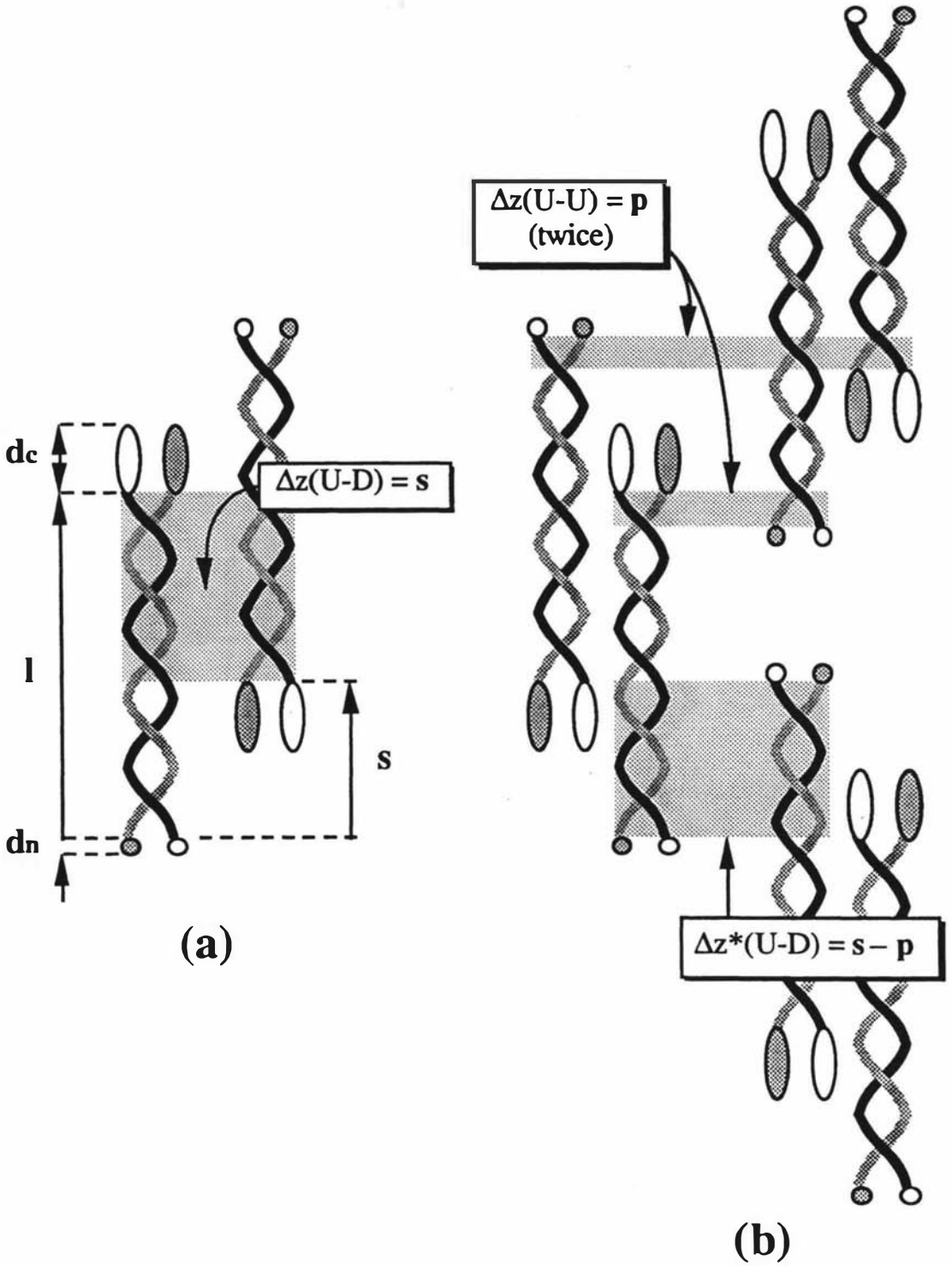


Figure 4-3 Molecular structure for Model A of lamin paracrystals. (a) Structure of the asymmetric unit which comprises a dimer of molecules staggered by an axial distance s. This distance is specified by three possible sets of ionic interactions, one of which is shown as a shaded rectangle between the molecules. (b) The two other types of ionic interactions possible for Model A are shown shaded between dimers of molecules. These specify the larger scale structure of the paracrystal. The light-staining bands observed in electron micrographs of the negatively-stained paracrystals correspond to regions containing the C-terminal domains.

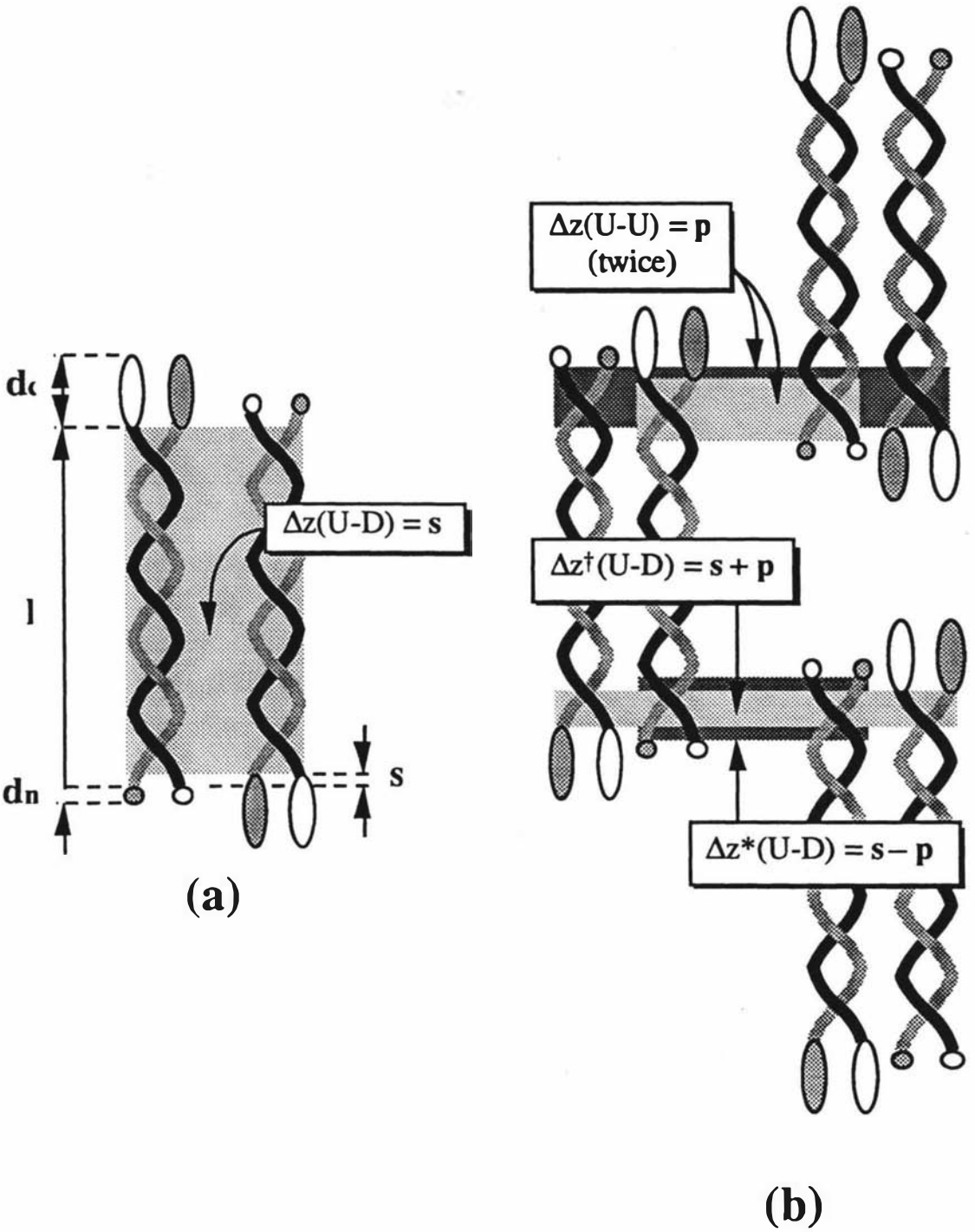


Figure 4-4 Molecular structure for Model B of lamin paracrystals. (a) Structure of the asymmetric unit which comprises a dimer of molecules staggered by a distance s. This distance is specified by four possible sets of ionic interactions, one of which is shown as a shaded rectangle between the molecules. (b) The three other types of ionic interactions possible for Model B are shown shaded between dimers of molecules. These specify the larger scale structure of the paracrystal. The light-staining bands observed in electron micrographs of the negatively-stained paracrystals correspond to regions containing the C-terminal domains.

asterisk (*) and dagger (†) symbols indicate alternative values for Δz(U-D).

The assumption of uniform packing within each of the light- and dark-staining bands of the paracrystal restricts model-building to two general classes of structure (Figures 4-3 and 4-4). In Model A, the antiparallel molecules in the dimer are approximately half-staggered and in Model B the molecules in the dimer are largely overlapped. Thus in Model A, the distal ends of the C-terminal domains are in near axial proximity, whereas in Model B it is the proximal ends that are in near axial register. Note that the large C-terminal domains are likely to have a fairly extended conformation so as to span half the width of a light-staining band and maintain a near uniform density of packing. This will be discussed in greater detail later but it is worth pointing out that uniform packing could also be achieved in Model A by further extending the C-terminal domains so that they span the full width of a light-staining band.

The ionic interaction data from Section 4.2 can now be used to find suitable values for **p**, s, $\mathbf{d_n}$ and $\mathbf{d_c}$. Consider Model A (Figure 4-3). A suitable maxima for **p** from the interaction curve for parallel molecules is 324 residues (48.1 nm) which has a score of 14 ionic interactions. The length of the region containing the terminal domains (ie, the light-staining band in the paracrystals) is:

$$(0.49\pm0.04)\mathbf{p} = 159\pm12 \text{ residues}$$

and this may also be described by the expression:

$$(0.49\pm0.04)\mathbf{p} = \mathbf{l} + \mathbf{s} - \mathbf{p} + 2\mathbf{d_n} \quad \text{where } \mathbf{l} = 356 \text{ and } \mathbf{p} = 324$$

and hence

$$2\mathbf{d_n} + \mathbf{s} = 127\pm12 \text{ residues}$$

The interactions between antiparallel molecules (U-D) occur at relative staggers of **s** and **s** – **p** and it follows that the appropriate interaction curve should be examined for pairs of interactions with a relative separation of **p** (324 residues). One such pair would occur when

$$\mathbf{s} = 160 \text{ residues (96 interactions)}$$

and

$$\mathbf{s} - \mathbf{p} = -164 \text{ residues (96 interactions)}$$

Both of these staggers correspond to interactions above the 0.975 quantile level. By substitution:

$$\mathbf{d_n} = -16\pm6 \text{ residues} \approx -2 \text{ nm}$$

This implies that the N-terminal domains (which contain only 30 residues) are so small that they are not visualized in the electron micrographs of the paracrystals.

Taking the length of the C-terminal domains as approximately half the width of the light-staining bands as discussed previously, the likely value of $\mathbf{d_c}$ is given by the

expression:

$$d_c \approx \frac{p + s - l}{2}$$

$$\approx 64 \text{ residues (9.5 nm)}$$

This length implies a highly elongated C-terminal domain and suggests that there must be important structural and functional reasons for such a conformation. As mentioned above, it is also possible that the C-terminal domains extend across the width of the light-staining bands in which case the expression for d_c becomes:

$$d_c \approx p + s - l$$

In other words, d_c could be twice as long as before.

The apparent globular form of the C-terminal domains observed for isolated nuclear lamin molecules from rat by Aebi *et al* (1986) does not discount the possibility of an elongated structure. Myosin, for example, was originally believed to have spherical heads but under better preparative procedures was later shown to have elongated heads with a high aspect ratio (Elliott and Offer, 1978). Certainly the volume of the pair of C-terminal domains in the lamin molecule is large (Table 4-6). These domains are also clearly apparent in the micrographs of Zackroff *et al* (1984). The volume of a single C-terminal domain of human lamin A calculated from the amino acid sequence is about 37 nm^3 (Table 4-6). As this could lead to a structure of length 9.5 nm and approximate diameter 2.2 nm, the volume data are clearly not inconsistent with a structure extending over half the width of a light-staining band of the lamin paracrystal.

Analysis of the amino acid sequence of the C-terminal domains of human lamins A and C has revealed extensive but localized runs of charged residues. Furthermore, the relatively high serine content of regions of this domain suggests that they may have importance as potential sites for hyperphosphorylation (Fisher *et al*, 1986). Both of these observations indicate a high charge count and hence an elongated structure for the C-terminal domain is again predicted. It is reasonable to assume that lamins from other sources will have similar structures (cf. Types I-IV IF) although additional physico-chemical data will be necessary to provide proof of this. The *Xenopus* lamin A sequence is highly homologous to human lamin A, suggesting that it will share a common structure with its human analogue.

In Model B (Figure 4-4), the smallest but highest scoring value of Δz(U-D) occurs when **s** is 15 residues (164 interactions). Using the same value of **p** as for Model A (**p** = 324 residues), the value for Δz^*(U-D) will be -309 residues and corresponds to 16 interactions. For Model B,

$$s + p = 339 \text{ residues} < l$$

	Residue Volume (nm^3)	Human Lamin A Numbers of Residues	Human Lamin A Total Volume (nm^3)	Human Lamin C Numbers of Residues	Human Lamin C Total Volume (nm^3)
Ala	0.089	17	1.513	11	0.979
Arg	0.182	21	3.822	17	3.094
Asn	0.123	11	1.353	7	0.861
Asp	0.117	14	1.638	11	1.287
Cys	0.107	5	0.535	1	0.107
Gln	0.149	13	1.937	9	1.341
Glu	0.145	13	1.885	12	1.740
Gly	0.063	31	1.953	16	1.008
His	0.153	8	1.224	7	1.071
Ile	0.168	5	0.840	3	0.504
Leu	0.168	17	2.856	12	2.016
Lys	0.174	11	1.914	10	1.740
Met	0.168	3	0.504	2	0.336
Phe	0.197	5	0.985	4	0.788
Pro	0.130	12	1.560	7	0.910
Ser	0.094	44	4.136	23	2.162
Thr	0.199	20	3.980	14	2.786
Trp	0.235	4	0.940	4	0.940
Tyr	0.199	4	0.796	1	0.199
Val	0.141	19	2.679	14	1.974
Totals		277	37.050	185	25.843

Table 4-6 Volume calculation for the C-terminal domains of human lamins A and C based on the residue volume data collected by Parry and Baker (1984) and the lamin sequence data of Fisher *et al* (1986). The sequences of human lamins A and C are identical from the N-terminal domain to the 179^{th} residue of the C-terminal domain. Thereafter, lamin A continues for 98 residues and lamin C for 6.

and so interaction (iv) above will be viable. The nearest maxima is at

$$\Delta z^{\dagger}(U\text{-}D) = 340 \text{ residues (12 interactions)}$$

The width of the light-staining bands for Model B is (as before)

$$(0.49\pm0.04)\mathbf{p} = 159\pm12$$

and this may also be described by the expression:

$$(0.49\pm0.04)\mathbf{p} = 2\mathbf{d_c} - \mathbf{p} - \mathbf{s} + \mathbf{l} \qquad \text{where } \mathbf{l} = 356 \text{ and } \mathbf{p} = 324$$

Hence

$$\mathbf{d_c} = 71\pm6 \text{ residues (11 nm)}$$

Interaction	Model A		Model B	
	Stagger (residue translations)	Score	Stagger (residue translations)	Score
Δz(U-U) = p (twice)	324	14+14	324	14+14
Δz(U-D) = s	160	96	15	164
Δz^*(U-D) = s – p	-164	96	-309	16
$\Delta z^\dagger$(U-D) = s + p (s + p < l)			340	12
	Total	220	Total	220

Table 4-7 The sets of ionic interactions used in Models A and B (Figures 4-3 and 4-4).

A summary of the ionic interactions in the asymmetric unit for the two models is listed in Table 4-7. The total interactions for Models A and B (220) are identical. Important differences, however, do exist between the two models. For example, in Model A it is the rod-rod ionic interactions that specify both the structure of the dimer and the large scale axial coherence of the molecular assembly whereas in Model B the rod-rod interactions specify the dimer and it is predominantly the C-terminal domain to rod interactions that specify the axial coherence of the structure. An interesting observation is that the link L2 and the 'stutter' in segment 2B are in approximate axial alignment in Model A but are not so related in Model B. The significance (or otherwise) of this point has yet to be ascertained.

Both models use the same stagger between the dimer of molecules (Δz(U-U) = p = 324 residues) which results in the axial alignment of the first and last 32 residues of the rod domain. Although this stagger does not represent a major maximum in the ionic interaction curve, it does correspond closely to the observed period of the banding pattern in the negatively-stained paracrystal and in addition there are several points of note about the 32 residue sub-sequences (Figure 4-5). For example, a number of identical residues lie in exact axial register. There is also a small region which is highly hydrophobic in character within each sub-sequence and these regions are axially adjacent at this stagger. These regions exhibit the highest concentration of hydrophobicity within the entire rod domain of the lamin sequence and appear to indicate that localized apolar interactions may also be important in specifying the axial assembly of lamin molecules. No apolar interaction curves have been calculated, however, as the distribution of apolar residues outside of the a and d positions of the heptad has been found to be random (Parry *et al*, 1977; McLachlan and Stewart, 1982; Dowling *et al*, 1983).

Another feature of the 32-residue sub-sequences is that each includes one of the two

Figure 4-5 The sub-sequences that comprise the interacting portions of the rod domains of the lamin sequences staggered by $\Delta z(U\text{-}U) = p$ where p is 324 residues for both Models A and B. Residues are numbered from the start of the N-terminal domain which is 30 residues in length. The top segment begins at the N-terminus of the rod domain and the lower segment finishes at the C-terminus of the rod domain. Identical residues in the two sequences are boxed and conservative substitutions are underlined. The shaded box shows two regions which have the highest hydrophobicity in the rod domain of lamin molecules.

highly conserved sequences found in all IF chains in segment 1A and towards the end of segment 2B (see, for example, Parry and Fraser, 1985). The features observed within these two interacting regions of the rod domain suggest that one of their major roles is to specify the axial aggregation of parallel lamin molecules and, by implication, the axial assembly of other IF molecules since these chains also contain similar highly conserved sequences.

Two perpendicular diads are seen in the negatively stained paracrystals and these occur at the centre of both the light- and dark-staining bands. Using the parameters defined before, the positions may be expressed as:

Dark-staining band: $\frac{l + s - p}{2}$ = 96.0 residues (Model A)

or 23.5 residues (Model B)

Light-staining band: $\frac{l + s}{2}$ = 258.0 residues (Model A)

or 185.5 residues (Model B)

The possible diad position at residue 96 lies close to the position of the start of the 42 residue insert in segment 1B of type V IF molecules which is not present in other IF. Also, the possible diad at residue 23.5 lies within the highly conserved sequence in segment 1A. The relevance of these observations is unclear at this stage.

No preference has yet been stated for one of the proposed Models over the other. Some evidence is available, however, to suggest that Model A is more likely to be valid than Model B. As detailed in the start of this Section, a 'beading' of period 24-25 nm has been observed in filamentous structures prepared under conditions of low ionic strength (Aebi *et al*, 1986). Assuming that the molecular packing in the

paracrystals is identical to that in the filamentous structures, examination of the Models shows that a 25 nm period can be generated from Model A by realigning the large C-terminal domains perpendicular to that axis of the molecule (Figure 4-6). The proximal ends of the C-terminal domains would occur on average at alternate intervals of **p** - **s** and **s** (164 and 160 residues respectively, or 24.4 and 23.7 nm). The difference in orientations of the terminal domains using different staining and visualization procedures may be related to ionic conditions since beading becomes more apparent as ionic strength is lowered (Aebi *et al*, 1986). In Model B the C-terminal domains are spaced by the period of the paracrystal (48 nm) and so could not generate a 25 nm period in such a simple way.

Aebi *et al* (1986) also noted that washing the Triton-extracted nuclear lamin meshwork from *Xenopus* oocytes with low ionic strength buffer resulted in a 24-25 nm beading in the orthogonal lattice. This suggests that the packing of lamin molecules in the lattice is similar to that in the filamentous structures. As before, assuming that the molecular packing in the lattice is the same as that suggested in Model A for the paracrystal, then the beading can be explained in the same manner as for the filamentous structures. Model A can also explain in a straightforward way how the lattice structure may be disassembled by hyperphosphorylation. For reasons already discussed it is likely that many (but not necessarily all) of the sites of phosphorylation in lamin molecules will be located in the C-terminal domain (see also Fisher *et al*, 1986). The proximity of the rod and C-terminal domains means that the marked change of the charge profile of the C-terminal domain on phosphorylation would be expected to destabilize the ionic interactions between them. It could possibly also have a similar effect on adjacent rod-rod interactions and hence cause dissociation. Conversely, dephosphorylation would permit self-assembly of the lattice structure.

The lattice nodes would be expected to correspond to one of the two perpendicular diads observed in the negatively-stained paracrystals by consideration of symmetry. In either case, the large C-terminal domains would appear to make only a marginal contribution to the stability of the cross-over regions. The nodes would therefore consist essentially of intersecting rod domains. If the lattice is collapsed by rotating the intersecting rod domains relative to each other a paracrystalline structure such as described by Model A could be produced (Figure 4-7). Another α-fibrous protein, tropomyosin, reveals two such packing modes coexisting in the same structure (Figure 4-8) which provides support for a similar duality in lamin structure.

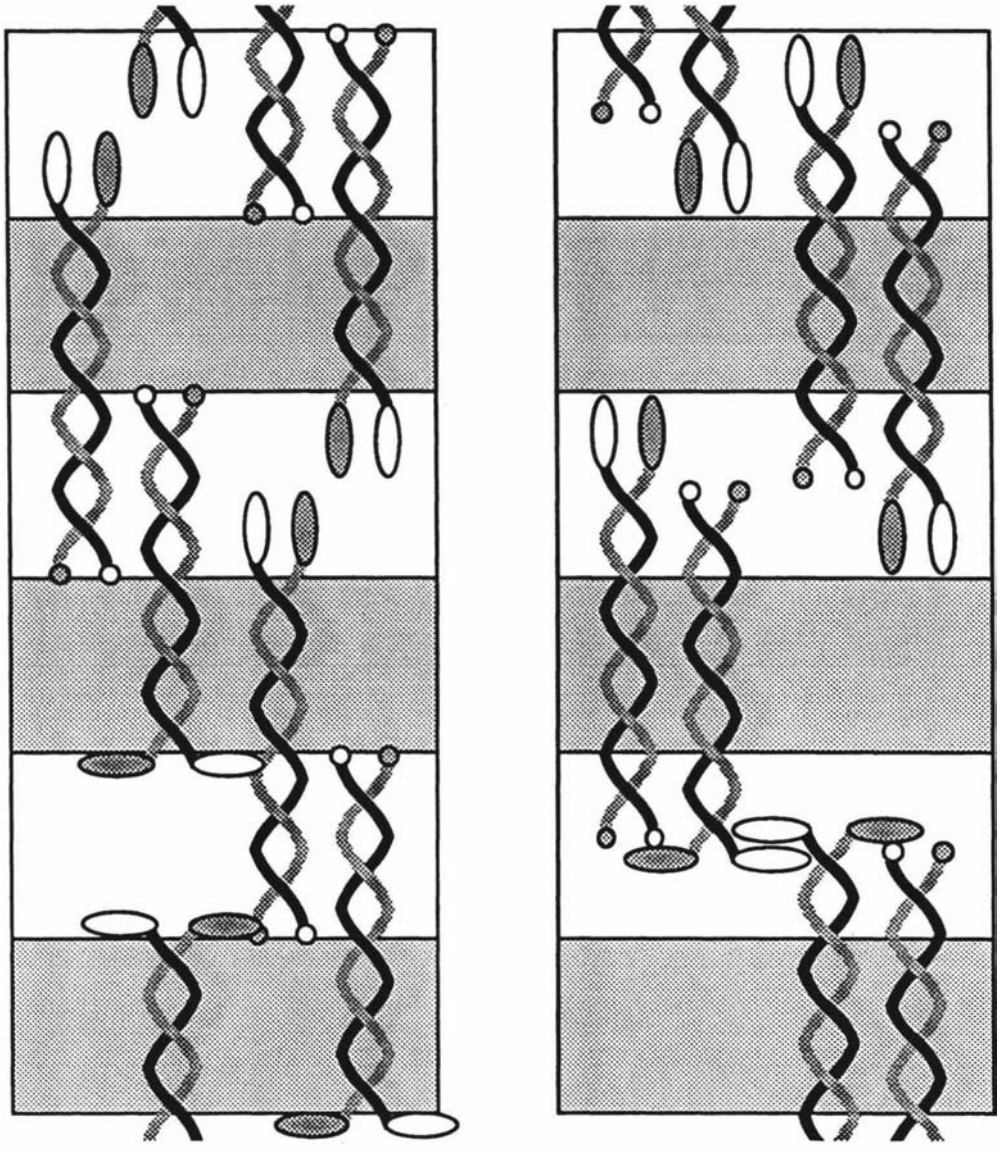


Figure 4-6 Schematic representation of the two modes of molecular assembly capable of explaining the stain distribution in the electron micrographs of negatively-stained lamin paracrystals. (a) In Model A the molecules comprising the dimer are antiparallel and approximately half- staggered whereas in (b) the molecules are antiparallel and almost completely overlapped. In the lower parts of both diagrams the large C-terminal domains have been oriented perpendicular to the axis of the paracrystal. In (a) this generates a quasi-repeat of ≈25 nm but in (b) the repeat is twice that of (a).

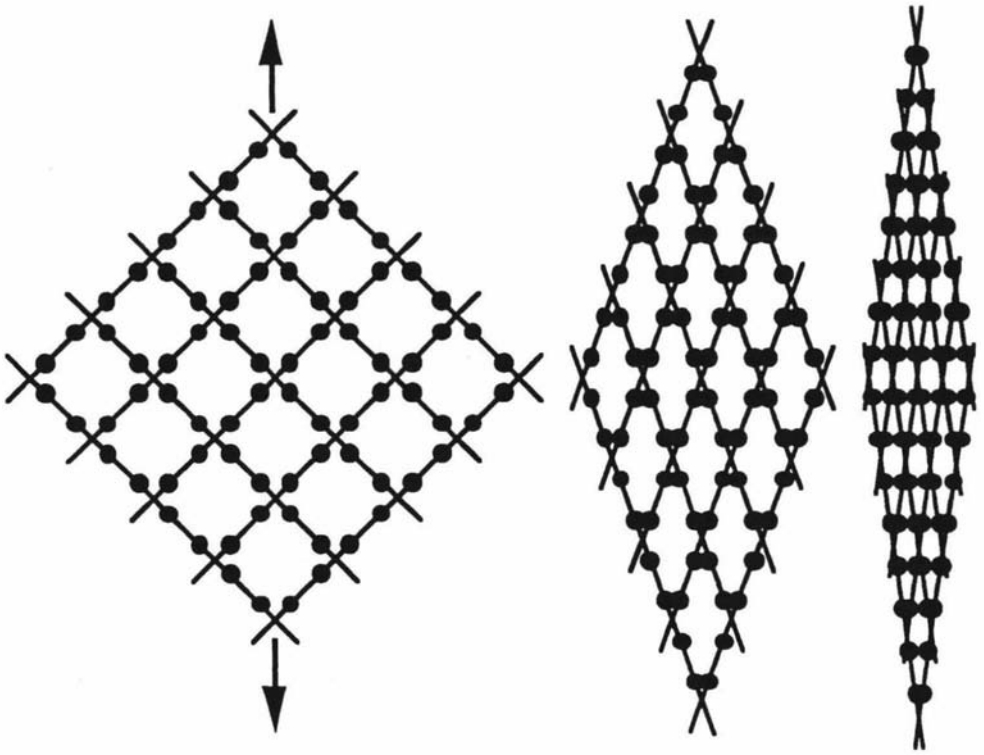


Figure 4-7 Possible transformation between the lattice and paracrystalline forms of lamin structure. The large C-terminal domains are drawn as filled circles and, in the collapsed lattice (to the right), indicate the approximate positions of the light-staining bands in the paracrystal. See also Figure 4-8.

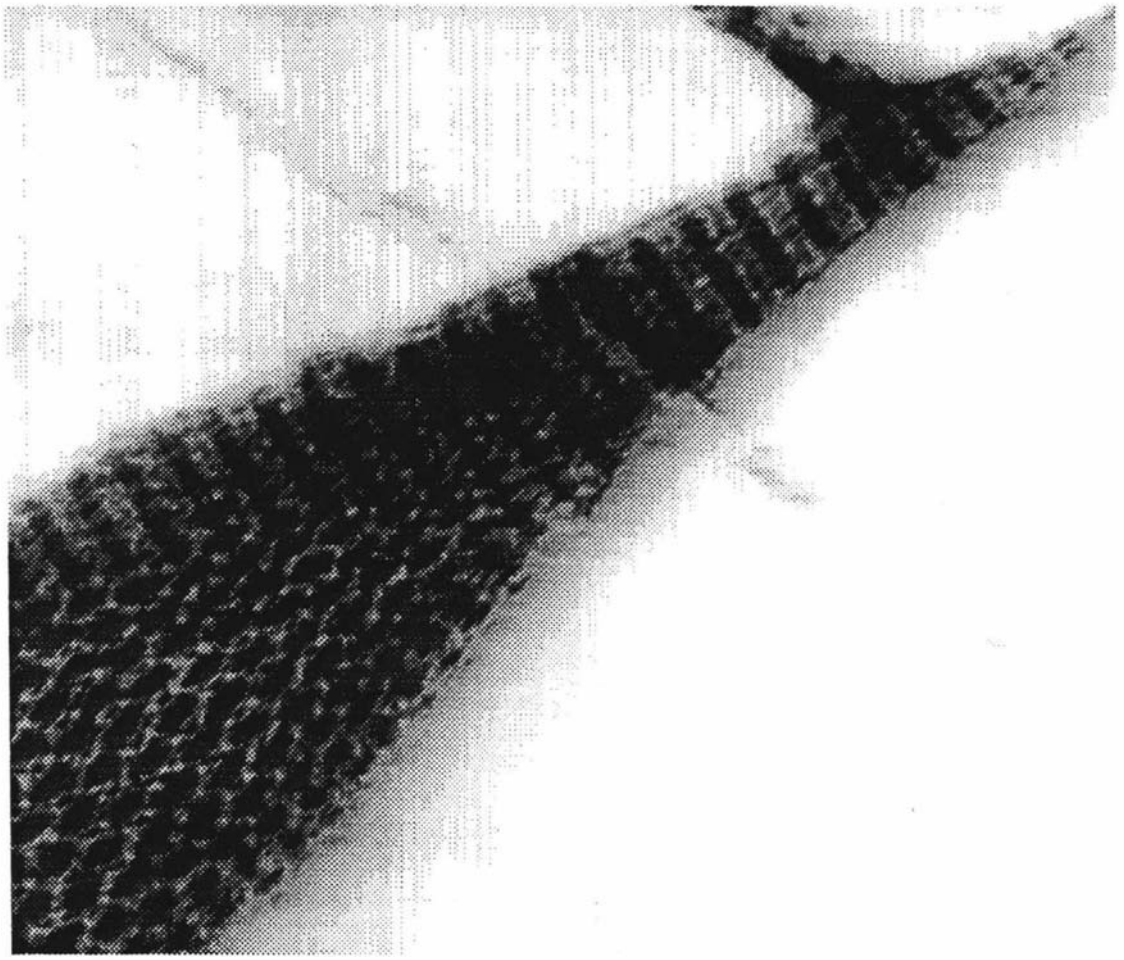


Figure 4-8 Electron micrograph of tropomyosin revealing two packing modes coexisting in the same structure. The crystalline lattice form at the lower left merges with the banded tactoid at the upper right. This lattice-filament duality may be similar to that proposed for lamin (Figure 4-7). Figure courtesy of Dr C. Cohen.

4.4 3D Ionic Interactions Between Molecules

Difficulties do arise with the linear array model of the coiled-coil used in the 1D method of calculating interactions because of the loss of azimuthal and radial information about each residue. The linear model would allow the possibility of interactions between residues in different molecules which the constraints of distance and orientation (ie, steric hindrance) would normally disallow. In particular, residues internal to the dimer of coiled-coils (ie, those in the a and d positions of the heptad) would rarely be in a position to allow interactions to occur with those in other molecules. Also, residues with identical axial coordinates but on opposite sides of a two-molecule assembly (for example) would be physically unable to interact with each other because of their separation. An advantage of a three-dimensional study is that several criteria previously not included can be built into the model and other criteria can be defined more realistically. For example, the distance limit previously imposed

N_0	N_1	M	P (nm)	N
1	24	84	12.12	3.6522
1	26	91	13.21	3.6400
1	28	98	14.29	3.6296
1	30	105	15.37	3.6207
1	32	112	16.44	3.6129
1	36	126	18.58*	3.6000

Table 4-8 Values used to derive a set of five pitch lengths for the 3D interaction study. N_0 is the number of turns of major helix; N_1 is the number of turns of minor helix; P is the pitch length of the coiled-coil; M is the number of turns in the repeat distance, c, where $c=N_0P$. Also included are the values that correspond closely to the pitch length (*) used by Parry and Suzuki (1969) in their 3D study of intrachain potential energy of α-helices and coiled-coil ropes.

between interacting residues was defined in terms of a range of linear residue translations. However, the cummulative azimuthal displacements of residues on an α-helix are approximately 103°, 206°, 309° (≡ -51°), 412° (≡ 52°) and so on. Hence, it is possible that while one residue may be in a position to interact with a particular site on an adjacent molecule, its nearest neighbour on the α-helix may be rotated too far towards the inside of the coiled-coil to do so. However, the third or fourth neighbour along (at azimuths of -51° and 52° and axial displacements of 0.45 nm and 0.59 nm respectively) may be in a better position to participate in interactions at the site and may even compete with the original residue in making a bond between the molecules.

The pitch length of the IF coiled-coil must also be included as a variable in the three-dimensional analysis since it has not yet been accurately determined. Five values were chosen that cover a range of 'reasonable' pitch lengths suggested by physical data (Cohen and Holmes, 1963; Elliott *et al*, 1968; Fraser *et al*, 1971; Parry, 1975; Phillips *et al*, 1986). These values are listed in Table 4-8 along with other parameters for determining the pitch length of the coiled-coil.

An assumption made to simplify the three-dimensional study is that the two molecules are related either by a parallel diad (parallel molecules) or by a perpendicular diad (antiparallel molecules). This has the effect of matching the azimuthal coordinates for the molecules (see Figure 4-9). The independent variables are thus reduced from four (pitch length of the coiled-coil, relative axial stagger between molecules and two azimuthal coordinates, one for each molecule) to three and makes the study computationally manageable. Note that the sense of rotation of the molecules is different for the parallel and antiparallel cases.

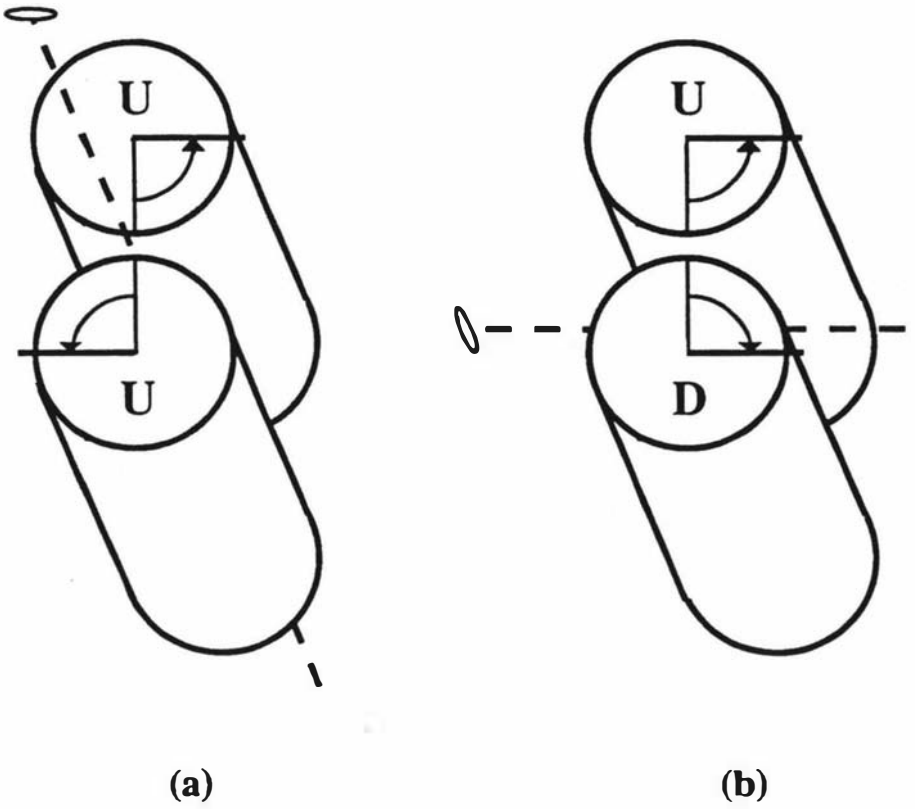


Figure 4-9 Relative directions of rotation for the three-dimensional modelling of molecules. (a) Parallel (U-U) and (b) antiparallel (U-D) molecules. The dashed lines indicate the axes of symmetry used to derive one molecule from the other.

4.4.1 Generation of Coordinates for a Coiled-Coil Molecule

A coordinate system is used whereby the z-axis is aligned along the axis of the coiled-coil and increases with the residue sequence from the amino- to carboxy-terminus of the protein chain. The x-axis is arbitrarily aligned so that the axes of the two α-helices at their starting positions lie on it: the two α-helices are equidistant from the origin but are displaced in opposite directions along the x-axis. A starting set of coordinates for the atoms comprising a single residue on an α-helix with 3.6 residues per turn is given in Table 4-9. Using these data a coiled-coil with radius r_0 and pitch length P may be generated using the methods described by Crick (1953) and Fraser *et al* (1964c). The coordinates of residues in a second, parallel strand coiled-coil are determined by the application of a parallel diad. Coiled-coils consisting of antiparallel chains have not been studied as there is no experimental evidence to suggest that such an arrangement exists in any α-fibrous protein.

The coiled-coil pitch length is specified in part by the number of residues per turn in the undistorted α-helix. Hence it would appear necessary to use a different starting set of coordinate data for coiled-coils of different pitch lengths. However, the analysis of

Atom	i	x_{i0} (nm)	y_{i0} (nm)	z_{i0} (nm)	r_{i0} (nm)	$\emptyset_{i0}$ (°)
N	1	0.1051	-0.1139	-0.0043	0.1550	-47.3
H	2	0.1227	-0.1005	-0.1017	0.1586	-39.3
Cα	3	0.2159	-0.0731	0.0830	0.2279	-18.7
Hα	4	0.2608	-0.1617	0.1297	0.3069	-31.8
C'	5	0.1632	0.0206	0.1919	0.1645	7.2
O	6	0.1858	-0.0016	0.3118	0.1858	-0.5
Cβ	7	0.3218	0.0000	0.0000	0.3218	0.0

Table 4-9 Starting set coordinates of a residue in an undistorted α-helix (Parry and Suzuki, 1969). The α-helix is right-handed with a unit rise of 0.15 nm. There are 18 residues in 5 turns of α-helix and hence the unit twist is +100°. Coordinates of an extended section of α-helix can be generated by application of the following expressions:

$$r_{im} = r_{i0}$$
$$\emptyset_{im} = \emptyset_{i0} + 100m$$
$$z_{im} = z_{i0} + 0.15m$$

The symbols i and m refer to the number of the atom in the reference set and to the residue number respectively. The standard set of coordinates in the Table above refer to the zeroth residue, ie m=0.

the packing of the coiled-coil molecules described in this section is of low resolution and is not sensitive to small changes in coordinate data. The coordinates for the coiled-coils comprising a pair of 1B segments with a pitch length of 12.1 nm are given in Table 4-10 and plotted in Figure 4-10. It is important to point out that by arranging for the β-carbon atom of the starting coordinate data set to lie on the x-axis, the sidechains of the apolar residues in the a and d positions of the two strands are approximately enmeshed to generate close to optimal knob-hole packing.

Having generated the coordinates of the two strands of a single coiled-coil molecule, the molecule is then translated a distance of 1 nm along the x-axis from the origin. The coordinates of residues in a second coiled-coil rope are calculated by the application of either a perpendicular diad, ie a two-fold axis parallel to the y-axis (antiparallel molecules) or a parallel diad, ie a two-fold axis parallel to the z-axis (parallel molecules) (see Figure 4-9). The exact separation of the molecules has been chosen in a somewhat arbitrary fashion and indeed some of the outermost β-carbon atoms from the two molecules appear to overlap. Although this arrangement is not feasible from an energetic viewpoint, the relatively large range allowed for interactions avoids any difficulties that might arise from the imprecise location of the molecules.

4.4.2 Determination of Interactions

The sidechains of the amino acids originate at the β-carbon atoms, the coordinates of which are readily calculable for a regular coiled-coil structure. However, limited freedom of rotation about each of the single bonds in the sidechains prevents a

	Molecule 1						Molecule 2					
	Coiled-Coil 1			Coiled-Coil 2			Coiled-Coil 1			Coiled-Coil 2		
Residue	x	y	z	x	y	z	x	y	z	x	y	z
144-b E	1.93	-0.11	0.00	0.07	0.11	0.00	-1.93	-0.11	14.49	-0.07	0.11	14.49
145-c M	1.76	0.56	0.36	0.24	-0.56	0.36	-1.76	0.56	14.13	-0.24	-0.56	14.13
147-e E	1.47	-0.43	0.41	0.53	0.43	0.41	-1.47	-0.43	14.09	-0.53	0.43	14.09
148-f E	2.05	0.00	0.68	-0.05	0.00	0.68	-2.05	0.00	13.81	0.05	0.00	13.81
149-g E	1.47	0.43	0.96	0.53	-0.43	0.96	-1.47	0.43	13.53	-0.53	-0.43	13.53
151-b R	1.76	-0.56	1.01	0.24	0.56	1.01	-1.76	-0.56	13.48	-0.24	0.56	13.48
152-c E	1.93	0.11	1.37	0.07	-0.11	1.37	-1.93	0.11	13.12	-0.07	-0.11	13.12
154-e R	1.19	-0.61	1.42	0.81	0.61	1.42	-1.19	-0.61	13.08	-0.81	0.61	13.08
155-f R	1.91	-0.53	1.69	0.09	0.53	1.69	-1.91	-0.53	12.80	-0.09	0.53	12.80
156-g Q	1.63	0.14	1.97	0.37	-0.14	1.97	-1.63	0.14	12.52	-0.37	-0.14	12.52
158-b D	1.38	-0.86	2.02	0.62	0.86	2.02	-1.38	-0.86	12.47	-0.62	0.86	12.47
159-c A	1.86	-0.38	2.38	0.14	0.38	2.38	-1.86	-0.38	12.11	-0.14	0.38	12.11
161-e T	0.86	-0.63	2.43	1.14	0.63	2.43	-0.86	-0.63	12.07	-1.14	0.63	12.07
162-f G	1.53	-0.91	2.70	0.48	0.91	2.70	-1.53	-0.91	11.79	-0.48	0.91	11.79
163-g Q	1.61	-0.19	2.98	0.39	0.19	2.98	-1.61	-0.19	11.51	-0.39	0.19	11.51
165-b A	0.89	-0.93	3.03	1.11	0.93	3.03	-0.89	-0.93	11.46	-1.11	0.93	11.46
166-c R	1.56	-0.76	3.39	0.44	0.76	3.39	-1.56	-0.76	11.11	-0.44	0.76	11.11
168-e E	0.57	-0.47	3.43	1.43	0.47	3.43	-0.57	-0.47	11.06	-1.43	0.47	11.06
169-f V	1.00	-1.05	3.71	1.00	1.05	3.71	-1.00	-1.05	10.78	-1.00	1.05	10.78
170-g E	1.43	-0.47	3.99	0.57	0.47	3.99	-1.43	-0.47	10.50	-0.57	0.47	10.50
172-b D	0.44	-0.76	4.04	1.56	0.76	4.04	-0.44	-0.76	10.45	-1.56	0.76	10.45
173-c N	1.11	-0.93	4.40	0.89	0.93	4.40	-1.11	-0.93	10.10	-0.89	0.93	10.10
175-e L	0.39	-0.19	4.44	1.61	0.19	4.44	-0.39	-0.19	10.05	-1.61	0.19	10.05
176-f D	0.48	-0.91	4.72	1.53	0.91	4.72	-0.48	-0.91	9.77	-1.53	0.91	9.77
177-g N	1.14	-0.63	5.00	0.86	0.63	5.00	-1.14	-0.63	9.49	-0.86	0.63	9.49
179-b Q	0.14	-0.38	5.05	1.86	0.38	5.05	-0.14	-0.38	9.44	-1.86	0.38	9.44
180-c K	0.62	-0.86	5.41	1.38	0.86	5.41	-0.62	-0.86	9.09	-1.38	0.86	9.09
182-e K	0.37	0.14	5.45	1.63	-0.14	5.45	-0.37	0.14	9.04	-1.63	-0.14	9.04
183-f Q	0.09	-0.53	5.73	1.91	0.53	5.73	-0.09	-0.53	8.76	-1.91	0.53	8.76
184-g K	0.81	-0.61	6.01	1.19	0.61	6.01	-0.81	-0.61	8.48	-1.19	0.61	8.48
186-b Q	0.07	0.11	6.06	1.93	-0.11	6.06	-0.07	0.11	8.44	-1.93	-0.11	8.44
187-c E	0.24	-0.56	6.42	1.76	0.56	6.42	-0.24	-0.56	8.08	-1.76	0.56	8.08
189-e I	0.53	0.43	6.46	1.47	-0.43	6.46	-0.53	0.43	8.03	-1.47	-0.43	8.03
190-f Q	-0.05	0.00	6.74	2.05	0.00	6.74	0.05	0.00	7.75	-2.05	0.00	7.75
191-g L	0.53	-0.43	7.02	1.47	0.43	7.02	-0.53	-0.43	7.47	-1.47	0.43	7.47
193-b Q	0.24	0.56	7.07	1.76	-0.56	7.07	-0.24	0.56	7.43	-1.76	-0.56	7.43
194-c E	0.07	-0.11	7.43	1.93	0.11	7.43	-0.07	-0.11	7.07	-1.93	0.11	7.07
196-e E	0.81	0.61	7.47	1.19	-0.61	7.47	-0.81	0.61	7.02	-1.19	-0.61	7.02
197-f N	0.09	0.53	7.75	1.91	-0.53	7.75	-0.09	0.53	6.74	-1.91	-0.53	6.74
198-g N	0.37	-0.14	8.03	1.63	0.14	8.03	-0.37	-0.14	6.46	-1.63	0.14	6.46
200-b A	0.62	0.86	8.08	1.38	-0.86	8.08	-0.62	0.86	6.42	-1.38	-0.86	6.42
201-c A	0.14	0.38	8.44	1.86	-0.38	8.44	-0.14	0.38	6.06	-1.86	-0.38	6.06
203-e R	1.14	0.63	8.48	0.86	-0.63	8.48	-1.14	0.63	6.01	-0.86	-0.63	6.01
204-f A	0.48	0.91	8.76	1.53	-0.91	8.76	-0.48	0.91	5.73	-1.53	-0.91	5.73
205-g D	0.39	0.19	9.04	1.61	-0.19	9.04	-0.39	0.19	5.45	-1.61	-0.19	5.45
207-b D	1.11	0.93	9.09	0.89	-0.93	9.09	-1.11	0.93	5.41	-0.89	-0.93	5.41
208-c A	0.44	0.76	9.44	1.56	-0.76	9.44	-0.44	0.76	5.05	-1.56	-0.76	5.05

Table 4-10 Coordinates of a pair of antiparallel 1B segments for chicken gizzard desmin. The pitch length of the coiled-coils is 12.1 nm. Residue numbers listed are relative to the amino-terminus of the chain and are followed by the heptad positions of the residues and their single letter code. Residues in the innermost positions of the heptad, a and d, have not been included as they are assumed to be incapable of forming intermolecular interactions. The x,y,z coordinates are measured in nanometres relative to an origin midway between the starting residues of the molecules: the z-axis is directed parallel to the axes of the molecules and the x-axis is colinear with the first residues from the amino-terminal end of the segments occupying the outermost (f) position of the heptad. These coordinates are plotted in Figure 4-10.

	Molecule 1						Molecule 2					
	Coiled-Coil 1			Coiled-Coil 2			Coiled-Coil 1			Coiled-Coil 2		
Residue	x	y	z	x	y	z	x	y	z	x	y	z
210-e T	1.43	0.47	9.49	0.57	-0.47	9.49	-1.43	0.47	5.00	-0.57	-0.47	5.00
211-f L	1.00	1.05	9.77	1.00	-1.05	9.77	-1.00	1.05	4.72	-1.00	-1.05	4.72
212-g A	0.57	0.47	10.05	1.43	-0.47	10.05	-0.57	0.47	4.44	-1.43	-0.47	4.44
214-b I	1.56	0.76	10.10	0.44	-0.76	10.10	-1.56	0.76	4.40	-0.44	-0.76	4.40
215-c D	0.89	0.93	10.45	1.11	-0.93	10.45	-0.89	0.93	4.04	-1.11	-0.93	4.04
217-e E	1.61	0.19	10.50	0.39	-0.19	10.50	-1.61	0.19	3.99	-0.39	-0.19	3.99
218-f R	1.53	0.91	10.78	0.48	-0.91	10.78	-1.53	0.91	3.71	-0.48	-0.91	3.71
219-g R	0.86	0.63	11.06	1.14	-0.63	11.06	-0.86	0.63	3.43	-1.14	-0.63	3.43
221-b E	1.86	0.38	11.11	0.14	-0.38	11.11	-1.86	0.38	3.39	-0.14	-0.38	3.39
222-c S	1.38	0.86	11.46	0.62	-0.86	11.46	-1.38	0.86	3.03	-0.62	-0.86	3.03
224-e Q	1.63	-0.14	11.51	0.37	0.14	11.51	-1.63	-0.14	2.98	-0.37	0.14	2.98
225-f E	1.91	0.53	11.79	0.09	-0.53	11.79	-1.91	0.53	2.70	-0.09	-0.53	2.70
226-g E	1.19	0.61	12.07	0.81	-0.61	12.07	-1.19	0.61	2.43	-0.81	-0.61	2.43
228-b A	1.93	-0.11	12.11	0.07	0.11	12.11	-1.93	-0.11	2.38	-0.07	0.11	2.38
229-c F	1.76	0.56	12.47	0.24	-0.56	12.47	-1.76	0.56	2.02	-0.24	-0.56	2.02
231-e K	1.47	-0.43	12.52	0.53	0.43	12.52	-1.47	-0.43	1.97	-0.53	0.43	1.97
232-f K	2.05	0.00	12.80	-0.05	0.00	12.80	-2.05	0.00	1.69	0.05	0.00	1.69
233-g V	1.47	0.43	13.08	0.53	-0.43	13.08	-1.47	0.43	1.42	-0.53	-0.43	1.42
235-b E	1.76	-0.56	13.12	0.24	0.56	13.12	-1.76	-0.56	1.37	-0.24	0.56	1.37
236-c E	1.93	0.11	13.48	0.07	-0.11	13.48	-1.93	0.11	1.01	-0.07	-0.11	1.01
238-e I	1.19	-0.61	13.53	0.81	0.61	13.53	-1.19	-0.61	0.96	-0.81	0.61	0.96
239-f R	1.91	-0.53	13.81	0.09	0.53	13.81	-1.91	-0.53	0.68	-0.09	0.53	0.68
240-g E	1.63	0.14	14.09	0.37	-0.14	14.09	-1.63	0.14	0.41	-0.37	-0.14	0.41
242-b Q	1.38	-0.86	14.13	0.62	0.86	14.13	-1.38	-0.86	0.36	-0.62	0.86	0.36
243-c A	1.86	-0.38	14.49	0.14	0.38	14.49	-1.86	-0.38	0.00	-0.14	0.38	0.00

Table 4-10 *continued*

meaningful prediction of their atomic coordinates. In order to reflect the mobility of the sidechains, a "volume of influence" is calculated: if the β-carbon atoms of a pair of residues from adjacent molecules lie within the volume, then the residues have the potential to interact. The volume chosen is an ellipsoid where

$$\frac{(\Delta x)^2}{0.8^2} + \frac{(\Delta y)^2}{0.8^2} + \frac{(\Delta z)^2}{0.3^2} < 1$$

and Δx, Δy and Δz are the differences in the x, y and z coordinates respectively of the β-carbon atoms of a pair of residues. This volume implies that interacting residues may be separated by up to 0.8 nm azimuthally (ie, in the x-y plane) but by only 0.3 nm (~2 residues) axially. Justification for this restriction on the axial freedom of the sidechains has already been discussed (Section 4.2).

Interactions studied include those between charged residues (lysine or arginine with aspartic acid or glutamic acid), hydrophilic residues (serine, threonine, glutamine, asparagine, arginine and tyrosine) and apolar residues (leucine, valine, isoleucine, methionine, phenylalanine, tyrosine and alanine). Some residues may exhibit multiple characteristics: for example, tyrosine has both apolar and hydrophilic properties and arginine has both charged and hydrophilic properties. In addition to the three types of

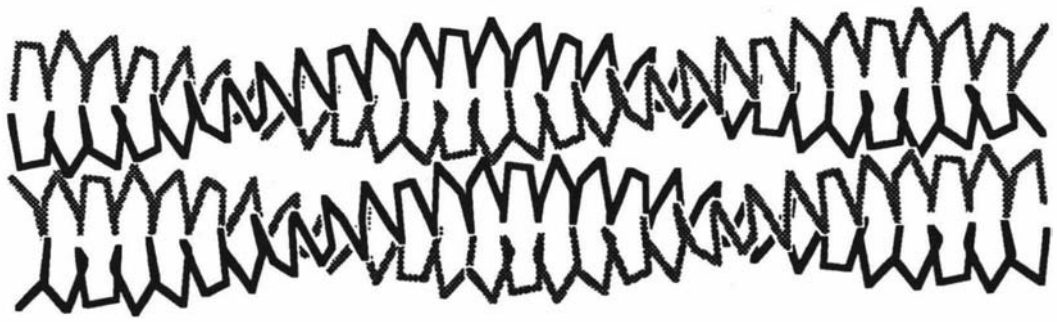


Figure 4-10 Schematic view of two 1B segments, each a dimer of coiled-coil chains. The segments have been rotated slightly out of the plane and some perspective has been added to enhance the view. The line segments shown connect the positions calculated for the Cβ atoms in Table 4-10, although the residues internal to the coiled-coils have been included here.

interaction calculated directly, a sum of the three is also made.

At this relatively early stage in the development of the three-dimensional technique, all the interactions have been weighted similarly with the exception of those involving alanine. Few physical data are available to allow useful modification of the weighting scheme and it is worth considering the success of the one-dimensional study which is based on a very simple model of protein structure. Down-weighting the interactions involving alanine, however, is based on its relative smallness compared to the other hydrophobic residues in its group. The interaction of alanine with a large hydrophobic residue is weighted by 0.25 and alanine with itself is weighted by 0.25^2. In addition, the criterion used for interaction distance is simple: no allowance is made for the different flexibilities or lengths of the sidechains or for the different rates at which the interactions attenuate over distance.

With three independent variables (pitch length, relative axial stagger and a common azimuth for the molecules) and four dependent variables (ionic, hydrogen-bond, apolar and total interaction scores) the study generates a large amount of data. For example, analysis of segment 1B involves varying the axial stagger between ±15 nm in 0.2 nm steps and the relative rotation from 0° to 180° in 2° steps (because the molecules are homodimeric and their rotations are equal, rotations from 180° to 360° are equivalent to those from 0° to 180°) and results in 27000 scores being generated for each pitch length and interaction type. These data can be represented compactly by an interaction map such as that shown in Figure 4-11 where rotation is represented along the horizontal axis, stagger is along the vertical axis and the interaction score is represented by a grey level on the graph - darker areas indicate higher interaction scores. The four different types of interaction are shown together as a mosaic in Figure 4-11 and a separate mosaic is calculated for each pitch length.

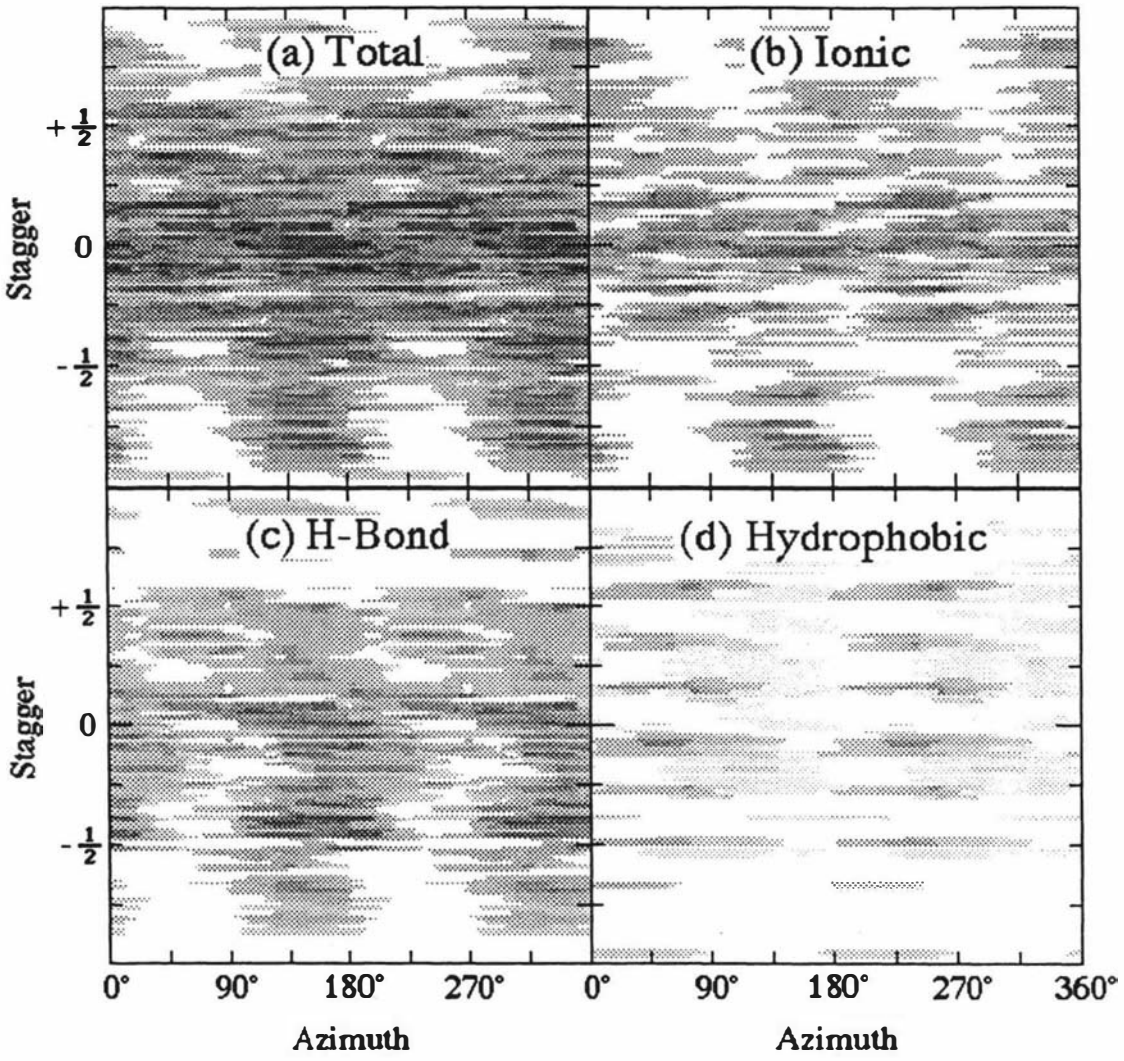


Figure 4-11 Interaction map for parallel 1B segments of chicken gizzard desmin. The pitch length of the coiled-coil is 12.1 nm. Higher scores are represented by darker shading. Stagger is displayed in units of the segment length. In this case, segment 1B is 101 residues long corresponding to 15 nm (assuming a constant axial rise per residue of 0.1485 nm).

4.4.3 Analysis of the Interaction Maps

A selection of the highest scoring total interactions for segment 1B of chicken gizzard desmin are listed in Table 4-11 (antiparallel segments) and 4-12 (parallel segments). No attempt at generating statistical levels of significance have been made. Such calculations are straightforward in principle and would follow the method detailed in Section 4.2. However, the enormous computational effort involved disallows any attempt at this at the present time.

The scores recorded in Tables 4-11 and 4-12 are typically about 25% of the maximum scores calculated by Fraser *et al* (1985) in their one-dimensional analysis of ionic residues from the same protein. This difference arises primarily from the distance constraints placed on the interaction parameters in the three-dimensional study. A

Pitch (nm)	Azimuth (°)	Stagger (residues)	Total Score
16.4	132	-12.00	16.00
12.1	6	-9.33	13.00
14.2	108	-9.33	14.50
15.3	108	-9.33	14.50
13.2	14	-4.00	12.00
14.2	150	-4.00	14.00
15.3	178	-4.00	14.25
16.4	174	-4.00	15.25
12.1	146	0.00	15.63
13.2	134	4.00	13.50
14.2	82	33.33	13.00
15.3	100	33.33	13.00

Table 4-11 A selection of the highest three-dimensional interaction scores for antiparallel 1B segments of molecules of chicken gizzard desmin. The total score is the sum of the ionic, hydrogen bonding and apolar interaction scores. Due to the symmetry of the homodimeric molecules, the same scores are found for azimuths offset by 180° from those listed. The stagger is recorded in units of residue translation in axial projection (=0.1485 nm).

simple example illustrates this point. In the one-dimensional study, ionic interactions between two 1B segments are scored in pairs. This occurs because the molecule is assumed to consist of a pair of parallel, in-register chains and when modelled as a single linear sequence there is effectively two identical residues in each position. These pairs of residues are assumed to interact in unison. However, in the three-dimensional model, the previously 'colinear' residues are now radially opposed about the axis of the molecule and are unlikely to contribute to the interaction score simultaneously. This implication alone of the three-dimensional structure of the molecules inherently reduces the scores by a factor of about two.

Higher interaction scores are revealed between antiparallel rather than parallel segments although the difference is not great. Also, higher scores are apparent at the extremes of pitch length studied (12.1 and 16.4 nm): however this may be an artifact of the geometry of the structures studied and the fixed relationship between the azimuths of each molecule. The axial length of segment 1B is approximately 15 nm (101 residues each with an axial rise of ~0.1485 nm), a value which is of the same order as the pitch lengths studied. The 16.4 nm pitch length arranges the residues of segment 1B on less than one pitch of coiled-coil. As the pitch length is reduced, the angle of the pitch becomes shallower (ie, closer to a plane transverse to the axis of the molecule) and the number of residues within interaction range of each other decreases

Pitch (nm)	Azimuth (°)	Stagger (residues)	Total Score
12.1	~150 (= -30°)	~±2.00	14.19
13.2	0	±1.33	14.06
14.2	~103	0.00	12.06
15.3	124	0.00	13.00
16.4	140	±2.67	14.00
13.2	~91	±21.33	12.00
14.2	~103	±21.33	13.00
15.3	~115	±21.33	12.00

Table 4-12 A selection of the highest three-dimensional interaction scores for parallel 1B segments of molecules of chicken gizzard desmin. The total score is the sum of the ionic, hydrogen bonding and apolar interaction scores. Due to the symmetry of the homodimeric molecules, the same scores are found for azimuths offset by 180° from those listed. The stagger is recorded in units of residue translation in axial projection (=0.1485 nm). The '~' symbol indicates that the score was common to a range of azimuths or staggers of which only the central one is listed here.

slightly. Further reduction of the pitch length to below the axial length of segment 1B causes the residues to be arranged on more than one pitch and hence the numbers of residues within interaction range will increase.

A feature of Table 4-12 (parallel segments) is that the azimuths of the high scoring orientations appear to increase with increasing pitch length. Such a trend is not apparent in the antiparallel case. This result can also be attributed in part to the geometry of the structure. In the case of parallel segments, the azimuth increases with the pitch length in order to "maintain" optimal interactions whereas for antiparallel segments no adjustment in relative azimuth is possible with the implementation of the perpendicular diad axis assumed.

The results presented here do not reveal any clear preference for either the pitch length of the coiled-coil or for the relative orientations of the molecules. Rather than a direct failing of the model, the problem probably lies in the lack of refinement of some parameters. For example, some weighting scheme for the different kinds of interaction as well as separate criteria for their ranges would be preferable to the arbitrary methods used here. However, until some physical evidence is presented no detailed specifications of these parameters can be made realistically. The excessive computational load also prevents a systematic examination of different combinations of feasible parameters.

A limitation of this molecular model is that the coiled-coil is assumed to be a regular structure. However, the helix geometry suffers local distortions which are likely to affect both the positions and orientations of the amino acid sidechains as well as the

pitch length. In addition to being irregular, the molecular structure is also dynamic and there is evidence that the non-covalent bonds between residues are frequently made and remade. Nuclear magnetic resonance studies on IF proteins by Mack *et al* (1988) suggest that the peptide backbone in the rod domain is relatively stiff but that the sidechains are very mobile: leucine sidechains revealed two states that were related by rotations of ~120° about both the Cα-Cβ and Cβ-Cγ bonds. Also, the flexibility of the protein backbone (which was examined in Chapter 2) is further evidence that IF proteins do not have a 'fixed' conformation but instead experience constant motion of the residues relative to each other. Both of these points are currently incorporated in the relatively loose definition of the interaction distance and it is possible that a more exact specification may not be feasible.

4.5 Summary

Interchain ionic interactions have been calculated for the lamin proteins, peripherin and the *Helix pomotia* B protein. In each case, a parallel in-register arrangement of chains is clearly preferred in accordance with previous studies on other IF and α-fibrous protein chains. Generally, segment 1 contributes a greater proportion of ionic interactions to the total than segment 2. Of special interest is the unusual intensity of interactions predicted for the human and *Xenopus* lamin A chains: the scores of these proteins (36) are fully twice that of peripherin (18) for example. These results indicate

1) that the proteins studied follow the general plan of IF structure in so far as the molecules are comprised of parallel, in-register chains,
2) that within the molecule, segment 1 is relatively more stable than segment 2, and
3) that the lamin A molecules are particularly stable compared to the other IF molecules studied.

This last point may indicate that lamin molecules require greater structural integrity than the cytoplasmic IF possibly to meet the special demands made on the lamin meshwork in supporting the nuclear envelope.

Intermolecular ionic interactions have also been studied for this same group of proteins. A brief discussion of the original method for calculating one-dimensional ionic interactions for collagen and α-fibrous proteins is presented in Section 4.2. This method has been used as the starting point for the development of a more sophisticated approach which takes into account the uneven distribution of apolar residues in coiled-coil segments. The positions of maxima in the interaction profiles for peripherin are found to compare well with those from a previous study on other type III proteins and provide further confirmation that peripherin is indeed a type III IF protein. The results

for human lamins A and C are also of significance in that they can be combined with physical data on lamin paracrystal and lattice structures and used to generate a plausible model for the structures. The interactions for the remaining proteins are included for completeness but no further conclusions are possible without additional physical data.

The model building study on human lamins A and C attempts to rationalize a number of observations made on the various lamin polymorphs by explaining the disparate data with a common model of molecular packing. The one-dimensional ionic interaction data are used to supply constraints on the modes of aggregation of the molecules and a single preferred model is presented. The model is simple in form, comprising a half-staggered array of antiparallel lamin molecules with each molecule consisting of a pair of parallel, in-register chains. It also predicts that the large C-terminal domains are highly extended. The axial aggregation of lamin molecules is proposed to be specified largely by ionic interactions. Apolar interactions between similarly directed segments corresponding to the highly conserved sequences found in all IF proteins in segment 1A and at the C-terminal of segment 2B are also considered to be of importance.

An additional feature of the proposed model for the packing of lamin molecules is the possibility of dual lattice-paracrystal structures. This is illustrated diagrammatically with the large C-terminal domains giving rise to the banding pattern observed in the paracrystal on the collapse of the lattice form. Evidence for such a duality is given in an electron micrograph for another α-fibrous protein where two different forms of packing (crystalline lattice and paracrystal) are apparent in the same structure: one packing mode merges smoothly with the other.

The mode of packing proposed here for lamin molecules is closely analogous to that suggested previously for other IF molecules (Geisler *et al*, 1985b; Parry and Fraser, 1985; Steinert and Parry, 1985). However lamin molecules are capable of adopting a variety of structural forms while IF molecules generally form only filaments of 'native' or 'near-native' dimensions (≈ 10 nm diameter) although paracrystals formed from fragments of a type III IF protein have been reported recently (Quinlan *et al*, 1989; Stewart *et al*, 1989a,b). The polymorphism of lamin may be attributable to the fundamental differences in lamin structure compared to the other IF molecules. One such difference is the relative simplicity of the structure for lamin which is predicted to be α-helical throughout the rod domain and which has fewer breaks in the heptad and ionic substructures than other IF molecules. Consequently, lamin molecules are more akin to the α-fibrous proteins in muscle such as tropomyosin, paramyosin and

myosin, all of which exhibit considerable polymorphism *in vitro*.

The role of phosphorylation as a control mechanism for the assembly of the nuclear lamina is well established. The nuclear lamina is known to disassemble concomitant with phosphorylation. Serine and threonine residues are generally considered to be potential sites for phosphorylation and this has been observed in the lamin proteins by Ottaviano and Gerace (1985). The model for aggregates of lamin molecules outlined above also provides a basis for understanding the dissociation of the lamin network. The phosphorylation sites within the terminal domains would lie in close proximity to the rod domains. The axial cohesion of the lamin structure could then be destabilized by the addition of extra charged groups. Presumably the intermolecular ionic interactions will allow reassembly of the structure once the charged moieties have been removed.

An interesting observation by Ottaviano and Gerace (1985) is that low concentrations of phosphate were associated with assembled lamins in growing interphase cells. Gerace (1986) suggested that this might allow local weakening of the lamin meshwork so that newly synthesized lamins could be incorporated. This idea is also consistent with the model above although the nature of the insertion of new material into the orthogonal lattice is not clear. In particular, the method of extending the lattice while maintaining a constant cross-over spacing is difficult to visualize. It is possible that some temporary distortion to the orthogonal lattice is required: the lattice spacing might expand between two parallel filaments to allow a third, preassembled filament to be inserted.

A refinement to the interaction method is offered in Section 4.4. The one-dimensional approximation to the molecular structure is modified to a three-dimensional model. This allows a more realistic consideration of the potential for interactions to be made between residues on adjacent molecules although at this stage there are insufficient data for proper specification of the limits and relative weights associated with different types of interaction. Preliminary results are presented for a type III IF protein, chicken gizzard desmin, but unfortunately no clear conclusions can be made on the likely modes of molecular packing.

5. SUMMARY

A number of aspects of IF chain and molecular structure, as well as molecular aggregation, have been examined in this thesis. These include the extent of homology among the IF proteins, the delineation of periodicities in the sequences of structural domains of IF proteins, the distribution of amino acid residues within the heptad substructure, the flexibility of the peptide backbone, the packing of chains in the dimeric molecule, and the axial packing of molecules in the IF. To a large degree the studies have been based on the amino acid sequences of the IF proteins. This approach has proved remarkably successful in that it has allowed much of the structural hierarchy to be specified from the primary structure right up to the level of molecular organization in the IF.

Several new conclusions can now be made about IF chain classification. In addition, other established ideas can be confirmed.

1) Subtyping of the type I and II keratin IF chains has been proposed on the basis of sequence homology within the rod domain. The 'hard' keratins are termed types Ia and IIa and the 'soft' or epidermal keratins are termed Ib and IIb.
2) The peripherin chain is confirmed as a type III IF chain in spite of its location in neuronal tissue and its coexpression with the type IV neurofilament chains. Examination of the gene structure by others had already indicated that peripherin shared a greater homology with type III IF proteins than type IV but the Fourier, homology and ionic interaction studies presented here all provide confirmatory evidence for this classification.
3) The classification of neurofilament chains as a type IV group separate from the type III IF proteins is also confirmed by the homology studies. In addition, the NF-L, NF-M and NF-H subtyping, which were assigned initially on the basis of the molecular weights of the proteins, is shown to be reflected in the homology within the rod domains and indeed throughout the entire amino acid sequences.
4) The type V chains may require subtyping based on the evidence of the sequence disparities between the karyoskeletal and cytoskeletal members of the group.

Although simple nomenclature is clearly desirable, the more detailed specification of IF groupings described above is a response to the increasing structural diversity apparent from the large body of IF protein sequence information now available. A greater understanding of the protein structure is reflected in the new subclasses. In some cases the finer divisions are revealed on a macroscopic scale as, for example, between the

'hard' and 'soft' keratins.

Some of the results obtained in this work are in good agreement with previously established features of IF protein structure. The Fourier transform studies on the newly sequenced type V IF proteins and the type III peripherin protein reveal periodicities common to all IF. In particular, a 9-10 residue period in the linear distribution of acidic and basic residues is present in the transforms. This period is also evident in Fourier transforms of the homology curves for acidic and basic residues in each of the IF types, indicating that it is conserved among all IF proteins. The ubiquity of this period provides overwhelming evidence of its importance in specifying the modes of aggregation of molecules in the IF.

An interesting feature of some of the transforms for type V IF chains is the ~9.23 residue period which has been observed previously only in the type I keratins. This period raises some new possibilities for the interaction of coexpressed IF networks in the cytoplasm (ie, keratin-containing IF and the *Helix pomotia* B-containing IF) although it may only reflect a common structural organization adopted by the separate networks. In addition, the role of lamin B as a possible nuclear anchorage site for the cytoplasmic IF has been raised (Georgatos and Blobel, 1987) and again the ~9.23 residue period common in lamin B and the type I keratins indicates that a means for direct interaction between the IF networks does exist. The problem of how the cytoskeletal and karyoskeletal IF interact across the nuclear membrane has yet to be resolved. Some evidence does exist, however, that parts of both networks are localized around the nuclear pores and hence may maintain connectivity through the pore structure.

Another structural feature of the coiled-coil regions that is confirmed here is the dominance of apolar residues in the internal a and d positions of the heptad substructure. The examination of residue occupancy of the heptad positions presented in Chapter 2 is based on a greatly expanded protein sequence database from that of Parry and Fraser (1985). Nonetheless, it confirms their measurement of an occupancy rate of ~75% for apolar residues in the internal positions of the heptad. The average homology scores of residues in these positions is also higher than for residues in other positions of the heptad. However the different numbers of individual apolar residues in the structurally similar a and d positions emphasize that they are not stereochemically identical. Residues in the adjacent e and g positions reveal the second highest mean homology scores and in the remaining positions, the homology is significantly lower. It is apparent that functional and structural diversity can be built into the 'outer' heptad positions, b, c and f, domain without upsetting the molecular

structure.

The conclusion that the lamin A molecule is stabilized to a greater degree by interchain ionic interactions than other IF molecules is interesting. The nuclear lamina provides much of the structural integrity of the nuclear envelope and, in addition, provides a scaffold for the nuclear pore structures and an anchor for interphase chromatin. Presumably these structural demands on the lamina require a greater stability for the lamin filaments than is needed for the cytoplasmic IF.

An area of interest that has not been addressed in this thesis is the characterization of the portions of protein sequences that specify either homodimeric or heterodimeric molecules. Reconstituted keratin filaments are formed from mixtures of type I and II proteins whereas the other IF readily re-form monocomponent filaments (Steinert, 1981a) although the various type III chains are capable of copolymerization *in vitro* (Steinert *et al*, 1981, 1982). The reason for this dichotomy of dimer structure is currently unknown: neither are the sequence-specific features that give rise to heterodimeric molecules on one hand (types I and II) and homodimers on the other (types III, IV and V). Some progress has been made, however, in identifying the broad regions of keratin rod domains that recognise the complementary polypeptide in the heterodimer although no explanation of the mechanism is yet possible (Magin *et al*, 1987). The reason for this extra complexity in the structure of keratin molecules is also obscure. One possibility is that the rod domain sequences largely dictate the unique pairing of particular type I chains with their type II partners (in addition to specifying the axial coherence of the filament) and that the terminal domains have some special functional role that could not be achieved by a homodimeric structure.

A further elaboration on this theme is possible. Cohen and Parry (1986) have made an interesting observation that intracellular α-helical coiled-coil proteins are two-stranded (for example, IF proteins, tropomyosin, paramyosin and myosin) whereas all three-strand coiled-coil proteins are extracellular and include fibrinogen (Doolittle *et al*, 1978; Parry, 1978), laminin and, more recently, the gp17 tail fibre of the T7 bacteriophage (Steven *et al*, 1988). In addition they noted that all two-stranded coiled-coil molecules have the same parallel-chain, in-register structure (Cohen and Parry, 1989). These protein sequences also contain the information which specify whether the molecule is a two or three stranded rope and again it is of interest to determine how and where this is specified by the amino acid sequences.

The rod domain of the IF chain is predicted to be less flexible on average than the terminal domains. This is in agreement with the perceived role of the rod as specifying the axial integrity of the filament (ie, the 'core') and with the terminal domains being

located at least partly on the filament periphery. Some portions of the rod of highest flexibility are located near the heptad stutters and may facilitate distortion of the coiled-coils to accommodate the break. The structural implications of the disruptions to the structure of the coiled-coil, however, remain obscure.

Several of the studies presented here are hindered somewhat by insufficient sequence data. Additional information from the 'hard' keratins is required to allow homology comparisons to be made and to more fully delineate the boundaries between the terminal sub-domains. Few type V sequences are available: in particular, sequences from lamin B proteins and cytoplasmic type V IF (analogous to the *Helix pomotia* B protein) are required. Both of these proteins are of special interest: the lamin B because of its suggested role in connecting the karyo- and cyto-skeletal IF networks and the *Helix* protein which appears to represent a new class or sub-class of IF chain. This latter protein shares an extended 1B segment with the lamin proteins but otherwise the homology between it and the lamins is not as great as that amongst the lamins alone. The location of the *Helix* protein within the cell is also different from that of the lamins and hence it is not unlikely that their functions are dissimilar. The potential for interaction between keratins and the *Helix* protein has been discussed but evidence to substantiate the occurrence of such aggregation is required.

Differences between the amphibian and mammalian keratins and also between the hard and soft keratins indicate that the rod domains of type II chains have retained a greater degree of homology overall than have the type I chains. This conclusion raises the question of why one partner of the obligate heterodimeric molecule should change at a greater rate than the other. It seems possible that the structural or functional roles of the rod domains are slightly different for the type I and II chains.

The method for calculating intermolecular ionic interactions based on a one-dimensional model of molecular structure remains a useful tool for predicting likely modes of aggregation. Some improvement on the technique is still possible, especially with regard to establishing a weighting scheme for the four possible types of interaction (Asp-Lys, Asp-Arg, Glu-Lys, Glu-Arg). Some physical basis for determining the appropriate scale factors is required, however. An empirical basis is also possible whereby potential ionic interactions measured for some protein of known structure (preferably similar to the coiled-coil of IF) provide the relative probabilities in much the same manner as the temperature factor data of a group of proteins allow chain flexibility to be predicted (Section 2.3).

The three-dimensional method also shows promise although greater attention to the detail of the structure and the parameters of the interactions is required. A problem

inherent in the method is determining the positions of the amino acid sidechains. The model used for the rod domain of the IF molecule assumes a regular coiled-coil structure but it is apparent that *in vivo* the coiled-coils will suffer local distortions (Phillips *et al*, 1986) in much the same way as α-helices in crystalline globular proteins have geometries noticeably different from the classical α-helix. This problem is compounded further by the mobility of the peptide backbone and the sidechains as indicated by NMR studies (Torchia and van der Hart, 1976; Mack *et al*, 1988). The uncertainty in sidechain position is currently overcome by specifying a volume within which the β-carbons of the two structures must exist if an interaction is to take place. However, the results presented here imply that the method is not yet selective enough to yield unequivocal results. Unfortunately it is not clear how the selection criteria may be improved at the present time.

Polymorphism has been observed in aggregates of the nuclear lamin proteins and has provided extra sources of structural information about molecular packing. These data have been used in combination with intermolecular ionic interaction studies to build a model for the axial packing of lamin molecules. This model can explain in a simple way the effect of phosphorylation in dismantling the lamin meshwork during mitosis and the subsequent reassembly on dephosphorylation. In addition, the transition between the observed lattice and paracrystalline forms is outlined. Predictions in this work and in Conway and Parry (1988) of polymorphism in the type I-IV IF have been confirmed by the mouse GFAP paracrystals described recently (Quinlan *et al*, 1989; Stewart *et al*, 1989a,b). Variations in structure among IF from different chain types are also predicted based on the differing homology amongst the rod domains of the type I-IV chains.

Stewart *et al* (1989b) have developed a model for the packing of rod domain fragments of mouse GFAP in paracrystalline structures. This model bears some similarities to that proposed for lamin paracrystals in Chapter 4 although the relative axial staggers of molecules are somewhat different. More axial substructure is apparent in the GFAP paracrystals than in lamin and this may result from the absence of the terminal domains in the GFAP fragments. [As an aside, it may be posited that if the terminal domains were cleaved from lamin molecules, the rod domains would aggregate in very well-ordered paracrystals]. However, an identical packing mode for both GFAP and lamin molecules is unlikely due to the differences in the lengths of the 1B segments in the chains.

The terminal domains are probably the least well understood structural elements of IF proteins. In many instances, phosphorylation sites have been identified within these

domains but little progress on their tertiary conformation or their functions has been made. Certainly the terminal domains do not appear to be so crucial as the rod domain for maintaining the axial structure of the filament and it is probable that the functions of different IF are specified largely by the terminal domains. In some cases, the length of the terminal domain far outruns that of the rod domain (the NF-H chains are an extreme example): in effect this implies a larger potential for function than for those IF chains where the terminal domains are short. The location of at least parts of these domains on the exterior of the filament suggests that their role is one of interacting with other proteins in the cell. The structure of the terminal domains, however, does not lend them to X-ray diffraction study in the same manner as the rod domains and hence future work on this aspect of the IF protein structure must rely on electron microscopy, chemical studies and predictive structural techniques. Alternatively, it may be possible to crystallize the terminal domains individually.

The special features of the recently characterized type V IF group raise many questions about IF proteins in general. For instance, is the extended 1B segment a remnant of the postulated ancestor to all IF proteins or a development from it? Does the molecular structure containing the extended 1B segment differ from that for the type I-IV IF molecule? Where these type V IF coexist with other IF, how do the different networks interact? The protein sequence studies here do not answer these questions but hopefully add to the ever-increasing body of knowledge about the IF proteins that will someday allow a fuller explanation of their structure and function.

In addition to IF structure, the role of IF in the cell has not yet been fully determined. The proposed connections between the cell nucleus via the cytoplasmic IF to the cell membrane suggests several possibilities: positioning of the nucleus, maintenance of cell shape and signal transmission or sensing between the nucleus and the cell surface. The further connection between IF of neighbouring cells through the desmosomes carries these ideas onto a larger scale and allows the possibility of physical cell-to-cell communication via the IF networks.

An area of direct application for this kind of work is in the investigation of cell dysfunction in so far as the IF are involved. For instance, Osborn and Weber (1982, 1983, 1986) have outlined the advantages for tumour diagnosis by IF typing. Certain types of tumour express particular IF types (eg, carcinomas are characterized by specific cytokeratins) and in addition the tumour will generally continue to express the same IF as the original cells. The use of IF antibodies may aid in the classification of tumours as an alternative to the usual histological procedures.

Skin disorders related to keratin IF have been under investigation by several groups.

Steinert *et al* (1980c) have determined that some skin disorders can be associated with abnormalities in the keratin chain, specifically in the non-α-helical terminal domains. Tyner and Fuchs (1986) describe keratin chains from human epidermis that are expressed during periods of rapid growth such as wound-healing. Some chronic skin diseases are related to uncontrolled proliferation of these chains and hence the regulatory mechanisms involved are a subject of interest as well as the structural differences between the keratins expressed during normal tissue growth and during rapid regrowth. Trachtenberg *et al* (1986) and Trachtenberg (1987) have investigated the structure of keratin IF in a specific chronic skin condition - psoriasis. The keratin IF contained in psoriatic cells is reduced in amount and in addition appears to be proteolytically degraded. The importance of defining the structural hierarchy and function of IF is clear: abnormalities on a microscopic level can and do result in tangible and serious defects occurring.

The advances made in recent years in determining the structure and spatial arrangement of IF proteins have been great. Acknowledgement of the homology between the keratins and the non-epithelial IF has allowed a coalescence of these previously separate fields of study and has resulted in many valuable insights into the structure and function of IF. The recent inclusion of the type V proteins into the IF family has revealed further variations on the common structural framework of the IF molecule and suggested that IF form part not only of the cytoplasmic skeleton but also of the karyoskeleton. Other work on potential IF-binding protein components of the desmosomes, such as the desmoplakins, indicates the possibility of an intercellular connection between the IF networks. It is clear that IF is an element of the cell structure that has come of age and can no longer be regarded as merely the cell debris in centrifuge experiments.

High quality X-ray diffraction and electron microscope data will be needed if the structure of IF is to be determined in greater detail than at present. It is to be hoped that fragments of IF or of IF molecules may soon be available for crystallization in a form suitable for X-ray analysis. In particular, the pitch length of the of the supercoil and the nature of the deformations in the supercoil structure at the heptad discontinuities are aspects of the structure that could be revealed by such studies.

In addition to the established techniques of X-ray diffraction and electron microscopy, new methods are now being used to investigate IF structure, and indeed protein structure in general. Site-directed mutagenesis is beginning to reveal new aspects of the structure of IF chains such as the nuclear localizational signal for the nuclear lamin proteins (Heald *et al*, 1988; Loewinger and McKeon, 1988). This technique has great

potential for probing specific features of IF molecules and their modes of assembly. Nuclear magnetic resonance studies are being used to determine information about the dynamics of the IF molecule and have shown that the positions of the amino acid residues of these proteins and their sidechains are not as fixed as was once thought. Advances in preparative techniques for electron microscopy are also likely to yield finer details of the molecular and IF structure by allowing samples to be preserved at a higher level of fidelity than is currently possible. Image enhancement and analysis techniques used in conjunction with electron microscopy will surely result in improved resolution of structural detail. Scanning transmission electron microscopy has been used for determining mass-per-unit-length data on IF and shows potential for investigating the early stages of IF assembly.

The rate of progress in advancing our knowledge about the IF proteins has been remarkable in the past decade. However, the emphasis in the field is now moving away from IF structure *per se* and towards the IF-associated proteins such as desmoplakin, fibronectin, and filaggrin. Only when the complex interplay between the IF and its associated proteins is understood will the full function of the cytoskeleton become apparent. We may expect significant developments in this exiting area of cell biology will occur during the 1990's.

APPENDICES

Appendix A : Zero-Filling prior to Fourier Transformation

The operation of zero-filling prior to Fourier transformation has the effect of improving the resolution in Fourier space. Generally, zero filling is performed by extending the data record by the addition of data points to the end of the record. For example, the Fourier transforms described in Chapter 2 involve sequence arrays of 101 to 356 data points which are zero-filled to a total length of 2048 points. The improved resolution in the Fourier space is achieved at the cost of introducing small side lobes for each peak. The consequences of the zero-filling operation will be described here.

Zero-filling may be described as the truncation of a long, hypothetical data record, d(x), by multiplication with a rectangle function, $\Pi(x)$. This rectangle function is defined as

$$\Pi(x) = \begin{cases} A & 0 \le x < W \\ 0 & W \le x < P \end{cases}$$

where W is the width of the non-zero data segment and P is the period, or full width of the filled record. For the purposes of zero-filling, A is typically set equal to unity. The discrete Fourier transform of $\Pi(x)$ is defined as

$$\begin{aligned} F\{\Pi(x)\} &= \frac{1}{P}\int_{-\infty}^{\infty}\Pi(x).\exp(\tfrac{-2\pi ixf}{P})dx \\ &= \frac{1}{P}\int_{0}^{W}A\exp(\tfrac{-2\pi ixf}{P})dx \\ &= \frac{-A}{2\pi if}\left[\exp(\tfrac{-2\pi ixf}{P})\right]_0^W \\ &= \frac{-A}{2\pi if}\left[\exp(\tfrac{-2\pi iWf}{P}) - 1\right] \\ &= \frac{A}{2\pi if}\left[\exp(\tfrac{\pi iWf}{P}) - \exp(\tfrac{-\pi iWf}{P})\right]\exp(\tfrac{-\pi iWf}{P}) \\ &= \frac{A}{\pi f}\sin(\tfrac{\pi Wf}{P})\exp(\tfrac{-\pi iWf}{P}) \end{aligned}$$

$$\begin{aligned} |F\{\Pi(x)\}| &= A\frac{W}{P}\,\frac{\sin(\pi\frac{W}{P}f)}{\pi\frac{W}{P}f} \\ &= A\frac{W}{P}\,\mathrm{sinc}(\pi\frac{W}{P}f) \qquad \text{where } \mathrm{sinc}(x) = \frac{\sin(x)}{x} \\ &= A\frac{W}{P}\,\mathrm{sinc}(\alpha) \qquad \text{where } \alpha = \pi\frac{W}{P}f \end{aligned}$$

The intensity function is simply the square of the Fourier transform,

$$|F\{\Pi(x)\}|^2 = \left[A \frac{W}{P}\right]^2 \text{sinc}^2(\alpha)$$

See Figure A-1 for a graphical representation of these functions. Minima occur for

$$\sin(\alpha) = 0 \qquad \alpha \neq 0$$

$$\text{or, } \sin(\pi \frac{W}{P} f) = 0$$

$$\text{which implies that } \frac{W}{P} f = n \qquad \text{where } n = 1, 2, 3, \ldots$$

$$\text{or, } f = n \frac{P}{W}$$

Maxima occur when the slope of the sinc function is zero, ie

$$\frac{d}{d\alpha} \text{sinc}(\alpha) = 0$$

$$\frac{\cos(\alpha)}{\alpha} - \frac{\sin(\alpha)}{\alpha^2} = 0$$

$$\frac{1}{\alpha^2}[\alpha\cos(\alpha) - \sin(\alpha)] = 0$$

$$\alpha\cos(\alpha) - \sin(\alpha) = 0$$

$$\alpha = \frac{\sin(\alpha)}{\cos(\alpha)}$$

$$\alpha = \tan(\alpha)$$

$$\text{ie, } \alpha \approx 0, \pm 1.4303\pi, \pm 2.4590\pi, \pm 3.4707\pi, \ldots$$

$$\text{or } f \approx 0, \pm 1.4303 \frac{P}{W}, \pm 2.4590 \frac{P}{W}, \pm 3.4707 \frac{P}{W}, \ldots$$

The half width limits of the main intensity peak are given by:

$$\left[\frac{\sin(\alpha)}{\alpha}\right]^2 = \frac{1}{2}$$

$$\alpha = \pm \sqrt{2} \sin(\alpha)$$

$$\text{ie, } \alpha \approx \pm 0.4429\pi$$

$$\text{or, } f \approx \pm 0.4429 \frac{P}{W}$$

The half width is then

$$\Delta\alpha \approx 0.8858\pi$$

$$\text{or, } \Delta f \approx 0.8858 \frac{P}{W}$$

The ratios of the main intensity peak to the first few side lobes are:

$$\text{at } \alpha = 0, \quad \text{sinc}^2(\alpha) = 1$$

$$\text{at } \alpha = 1.4303\pi, \quad \text{sinc}^2(\alpha) \approx 0.04719 \approx \frac{1}{21.19}$$

$$\text{at } \alpha = 2.4590\pi, \quad \text{sinc}^2(\alpha) \approx 0.01648 \approx \frac{1}{60.68}$$

$$\text{at } \alpha = 3.4707\pi, \quad \text{sinc}^2(\alpha) \approx 0.008340 \approx \frac{1}{119.90}$$

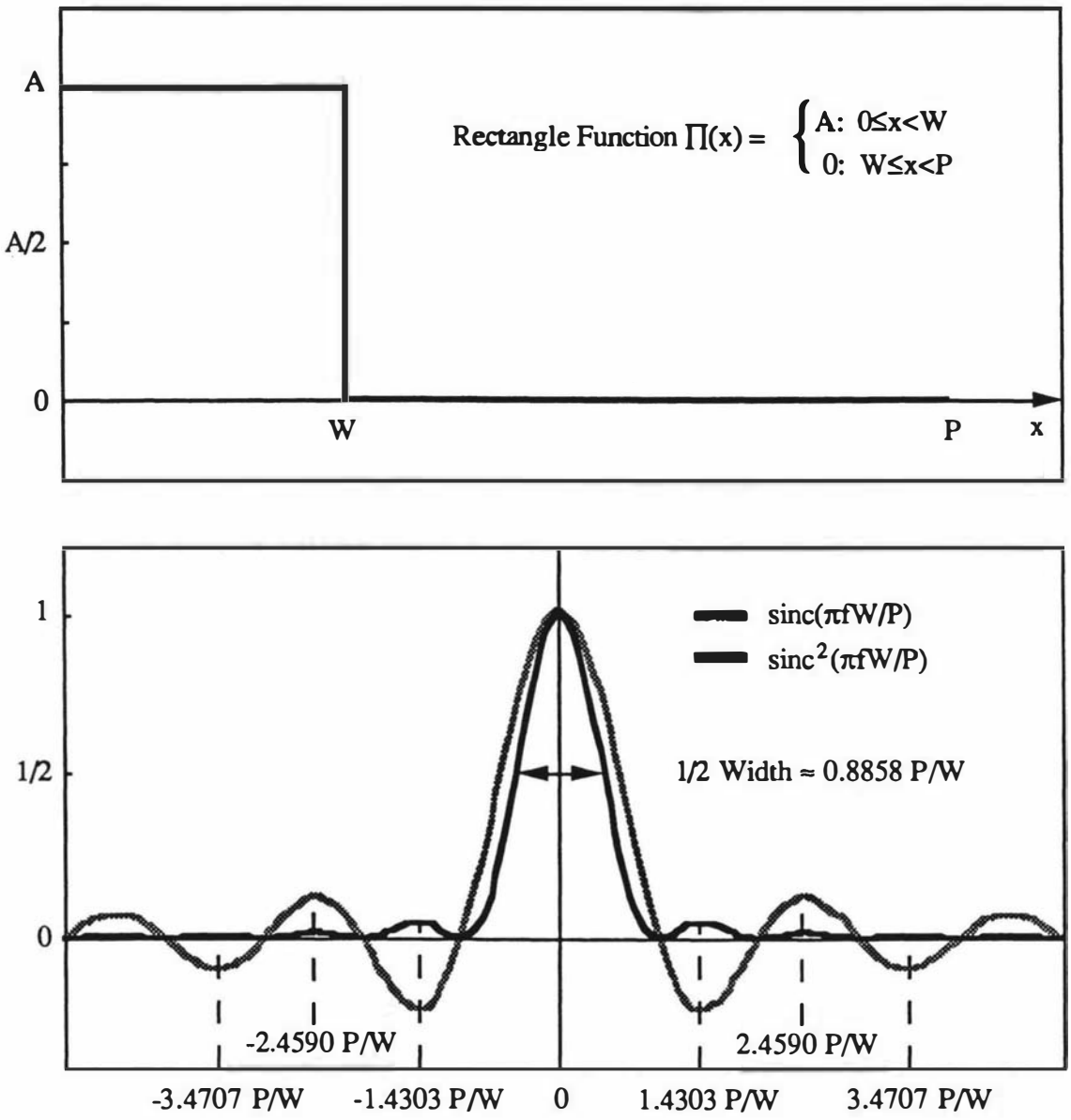


Figure A-1 The rectangle function ∏ and its Fourier transform pair, the sinc function. The intensity of the sinc function is also shown along with the positions of the first three side lobes and the half width of the central peak in the intensity distribution.

Using the length of segment 1B as an example and zero-filling to 2048 points,

$$W = 101$$
$$P = 2048$$
$$A = 1$$

then the rectangle function is defined as

$$\Pi(x) = \begin{cases} 1 & 0 \le x < 101 \\ 0 & 101 \le x < 2048 \end{cases}$$

and the intensity of its Fourier transform is

$$F\{\Pi(x)\}^2 = \left[A\frac{W}{P}\right]^2 \operatorname{sinc}^2(\alpha) \qquad \text{where } \alpha = \pi\frac{W}{P}f$$

$$= \left[\frac{101}{2048}\right]^2 \operatorname{sinc}^2(\pi\frac{101}{2048}f)$$

The half width of the main intensity peak is

$$f \approx 0.8858\frac{P}{W}$$

$$\approx 17.96$$

and the first sidelobes occur at frequencies of

$$f \approx \pm 1.4303\frac{P}{W}$$

$$\approx \pm 29.00$$

The effect of multiplying the data record by a second function (ie, the Π function) is to convolute the Fourier transform of the data by the Fourier transform of the second function. Convolution by the sinc function causes peaks in the data transform to be broadened by the width of the central peak of the sinc function as well as allowing "leakage" of peaks into the areas defined by the sidelobes of the sinc function. The relative strengths of the side lobes to the main peak are not large (less than 5%) and generally their effect on the Fourier plane can be ignored. However in the case of large peaks, such as the zero-frequency spike produced when the baseline has not been corrected for, the side lobes become significant relative to the rest of the Fourier data. The zero-frequency peak may be eliminated by performing a baseline correction operation prior to the Fourier transform.

Appendix B : Fourier Transforms

The Fourier transforms of charged residues in the rod domains of peripherin, human lamins A and C, *Xenopus* lamins A and B and the *Helix pomotia* B protein are presented in this appendix. These curves are discussed in Chapter 2. The relationship between the frequency axis and periods in the sequence is:

$$\text{period} = \frac{N_f}{\text{frequency}} \qquad \text{where } N_f = \text{size of the data set.}$$

The value of N_f is 2048 for all of the transforms included here. Some periods of structural significance with their corresponding frequencies are:

1) the 9 to 10 residue period (≡ frequencies from 228 to 205),

2) the heptad period (1st order ≡ 293; 2nd order ≡ 585; and 3rd order ≡ 878).

Although the heptad is defined mainly by apolar residues, nonetheless the charged residues can reveal some regularities of this order due to their rarity in the a and d positions and their conservation in the e and g positions.

	Peripherin			
Period	Segment 1B		Segment 2	
(residues)	acidics	basics	acidics	basics
13.045		4.712		
10.343	5.576			
9.894				7.359
9.752			7.391	
9.615		5.639		
6.919		6.099		
4.008				5.292
3.710			4.530	
3.471	3.078			
2.852				5.022
2.833			4.353	
2.330	5.580			
2.248	4.873			
2.241			4.268	

Table B-1 A selection of peaks in the Fourier transforms of acidic and basic residues in the major coiled-coil segments of peripherin (Figure B-1). Their scaled intensities are shown together with the period to which the peaks correspond. All peaks above an arbitrarily chosen cutoff intensity of 4.0 are listed although lower peaks are included where they coincide with periods that may be related to known structural features.

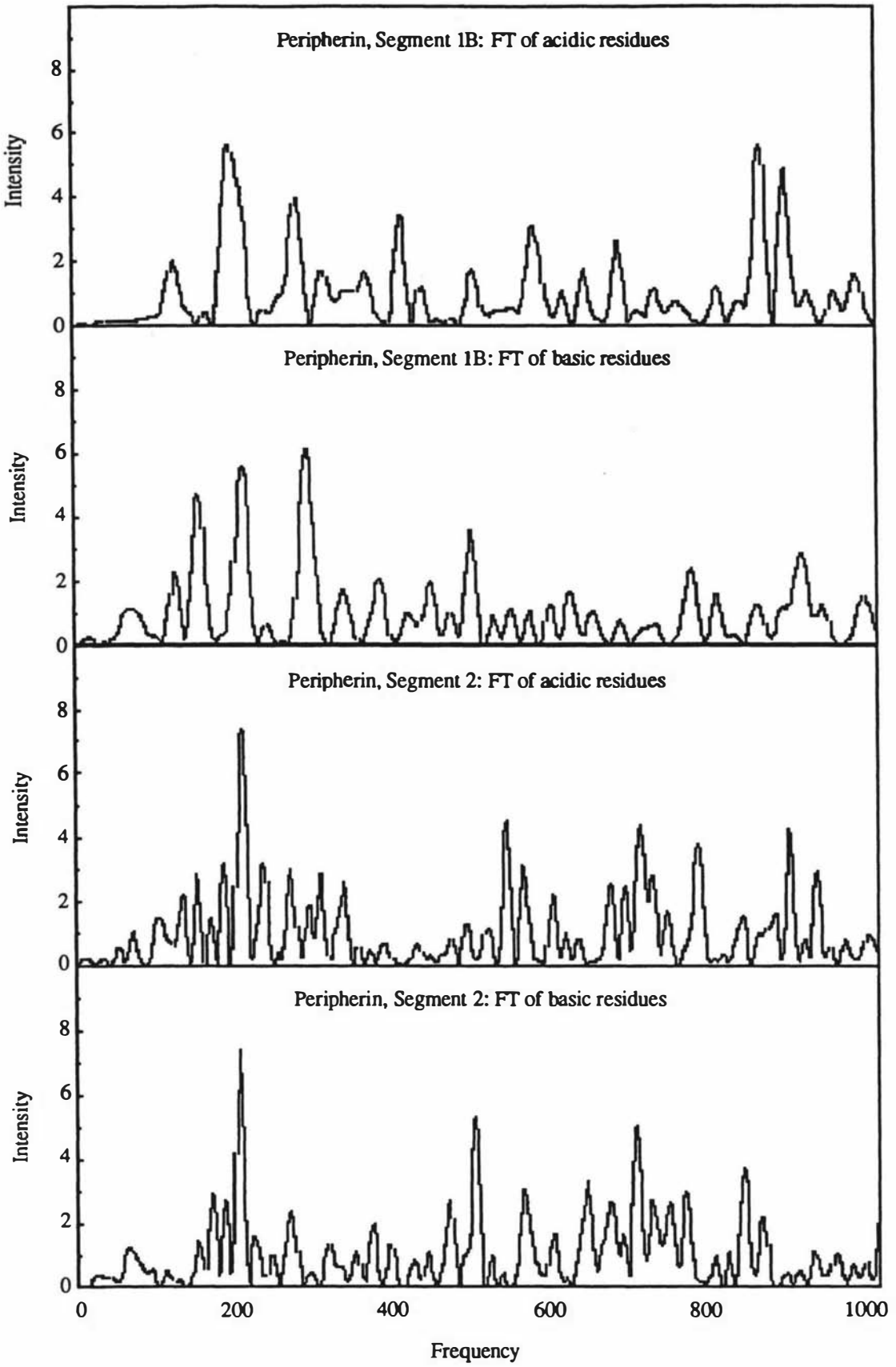


Figure B-1 Fourier transforms of the dispositions of acidic residues (top graphs) and basic residues (lower graphs) in rod domain segments of rat peripherin. A selection of peaks are listed in Table B-1. Period is related to frequency by the expression:

$$\text{period} = \frac{2048}{\text{frequency}}$$

Period (residues)	Human Lamins A and C					
	Segment 1B		Segment 2		Rod domain	
	acidics	basics	acidics	basics	acidics	basics
12.962					4.504	
11.838		4.018				
11.636						6.136
9.942	6.232	3.476				4.417
9.894					6.272	
9.309		6.245				3.654
8.569			4.235		5.874	
6.606	4.526				6.053	
5.069				7.352		
5.044						6.123
3.568			4.045			
3.374					4.420	
3.098	5.248				4.113	
3.025						5.521
3.021		8.346				
2.825					4.237	
2.768					4.805	
2.760	4.153					
2.528				6.067		
2.465						5.621
2.343					4.869	
2.338	6.979					
2.317						4.298
2.278		4.733				
2.122			6.071		5.129	
2.048				4.204		
2.036						5.349

Table B-2 A selection of peaks in the Fourier transforms of acidic and basic residues in the major coiled-coil segments of human lamins A and C (as for Table B-1). See Figure B-2.

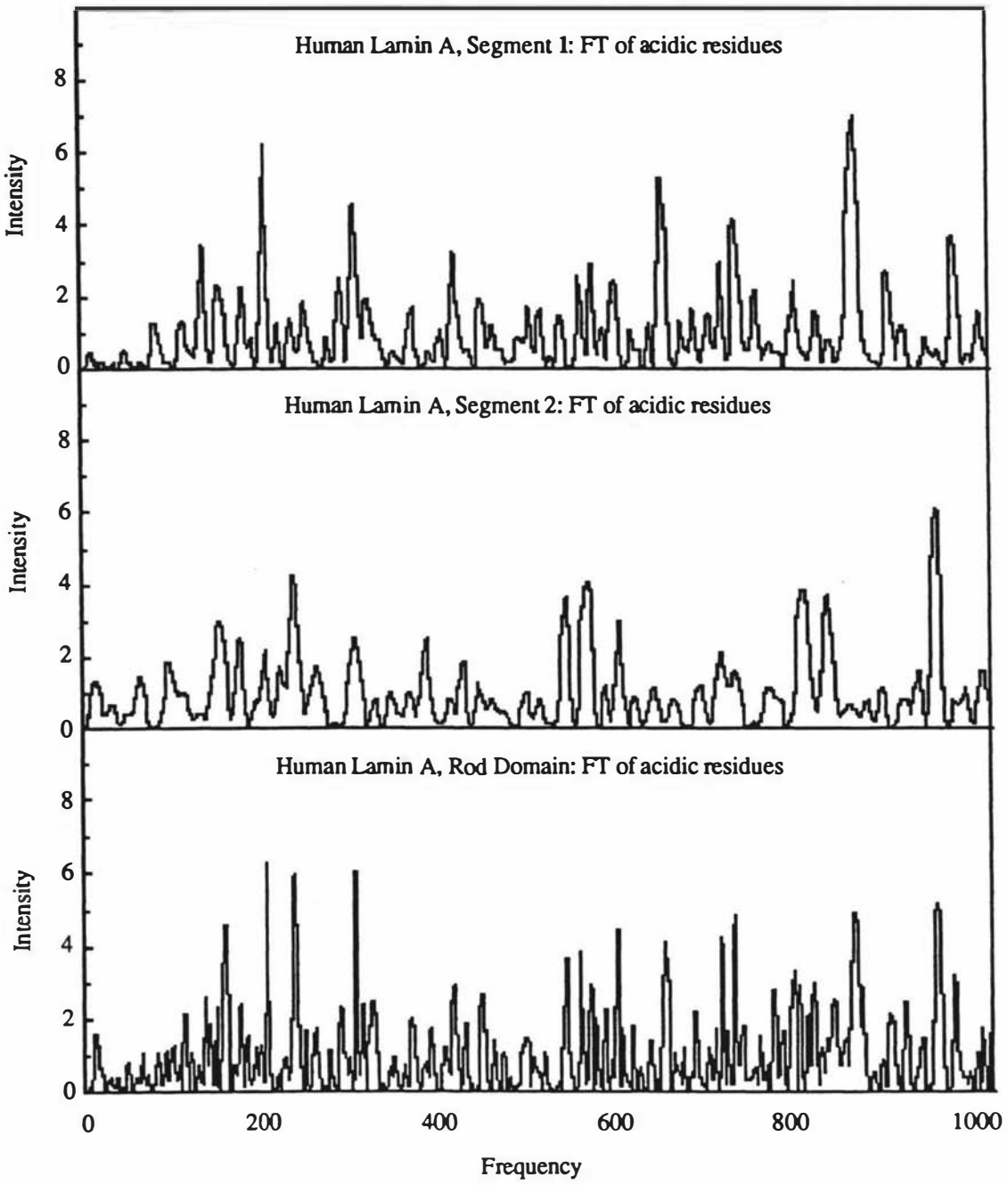


Figure B-2a Fourier transforms of the dispositions of acidic residues in rod domain segments of the human lamin A protein. These transforms are also equivalent to those for the human lamin C protein which has a primary rod domain sequence identical to that of lamin A. A selection of peaks are listed in Table B-2. Period is related to frequency by the expression:

$$\text{period} = \frac{2048}{\text{frequency}}$$

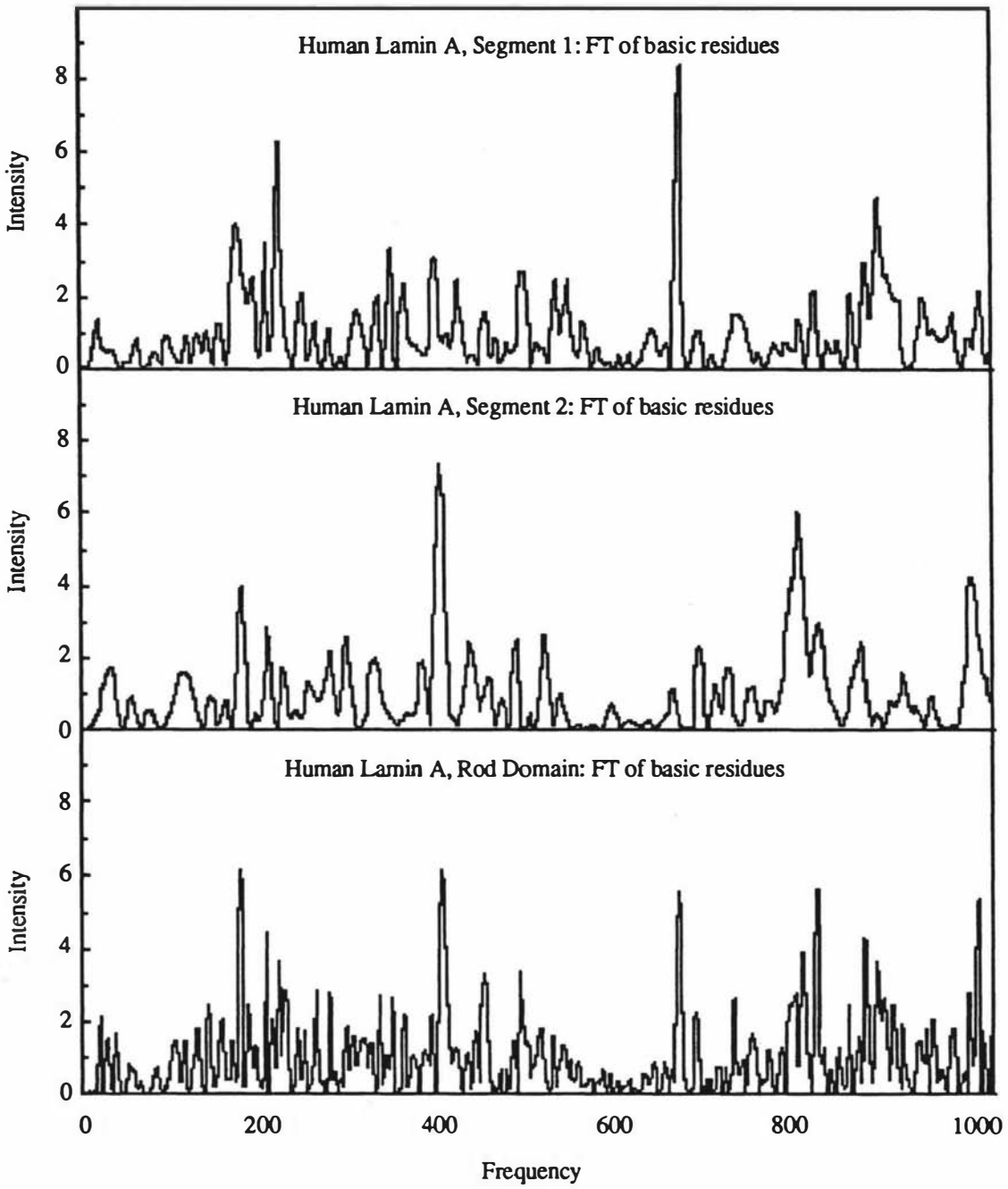


Figure B-2b Fourier transforms of the dispositions of basic residues in rod domain segments of the human lamin A protein. These transforms are also equivalent to those for the human lamin C protein which has a primary rod domain sequence identical to that of lamin A. A selection of peaks are listed in Table B-2. Period is related to frequency by the expression:

$$\text{period} = \frac{2048}{\text{frequency}}$$

Period (residues)	*Xenopus* Lamin A Segment 1B acidics	Segment 1B basics	Segment 2 acidics	Segment 2 basics	Rod domain acidics	Rod domain basics
13.653			4.708			
11.636						4.383
9.942	6.089	3.117				3.979
9.894					5.271	
9.309		6.903				4.638
8.569					5.449	
6.966	3.066					
6.606					5.259	
6.585	5.369					
6.461					4.018	
5.069				7.368		
5.057						6.254
3.374					4.123	
3.098	4.560					
3.025		6.924				5.163
2.768					5.056	
2.528				6.956		
2.507						4.020
2.465						6.620
2.335	6.812					
2.278		4.384				
2.127					4.968	
2.124			5.731			
2.086					4.002	
2.079	4.145					
2.038						5.544

Table B-3 A selection of peaks in the Fourier transforms of acidic and basic residues in the major coiled-coil segments of *Xenopus* lamin A (as for Table B-1). See Figure B-3.

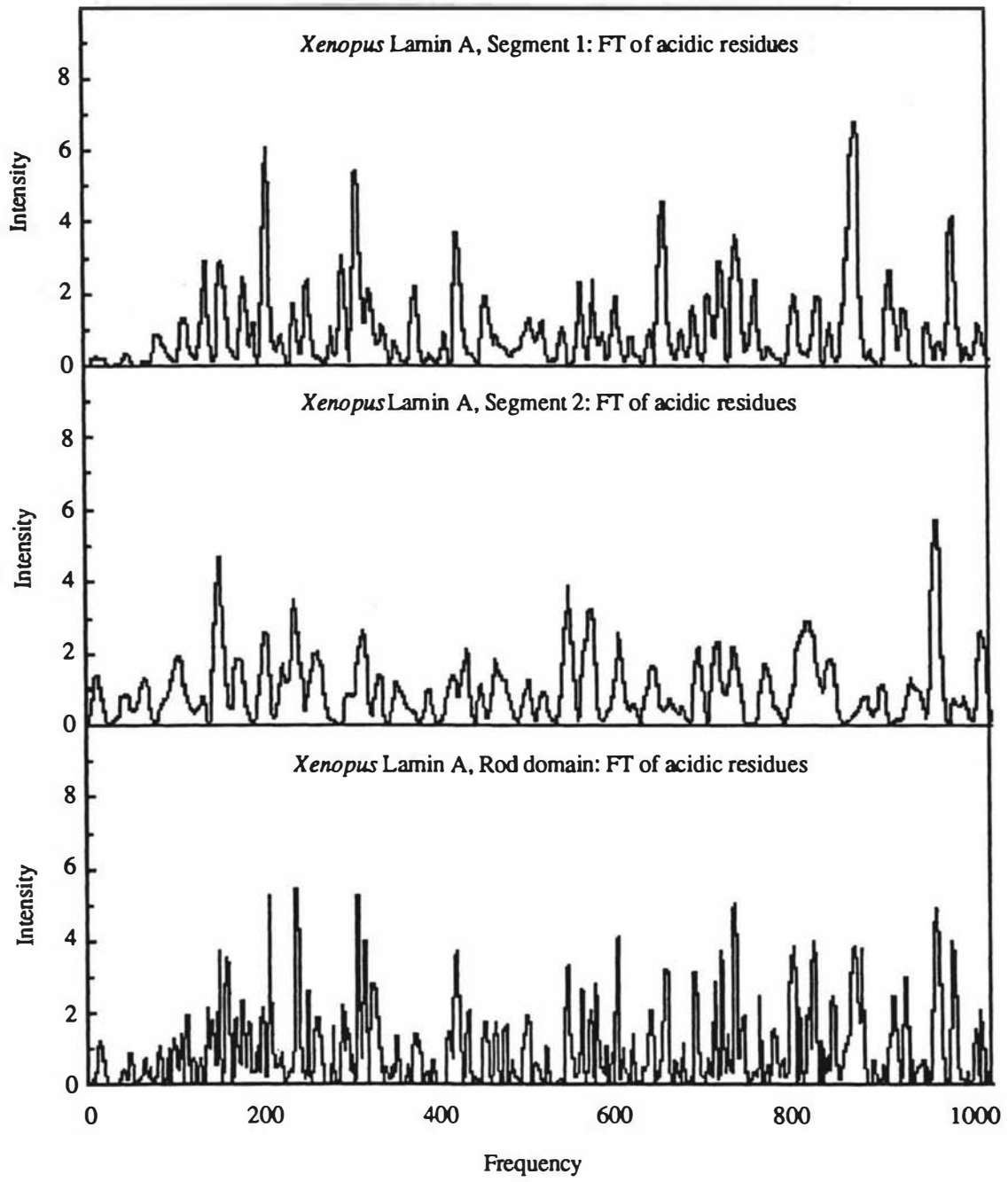


Figure B-3a Fourier transforms of the dispositions of acidic residues in rod domain segments of the *Xenopus* lamin A protein. A selection of peaks are listed in Table B-3. Period is related to frequency by the expression:

$$\text{period} = \frac{2048}{\text{frequency}}$$

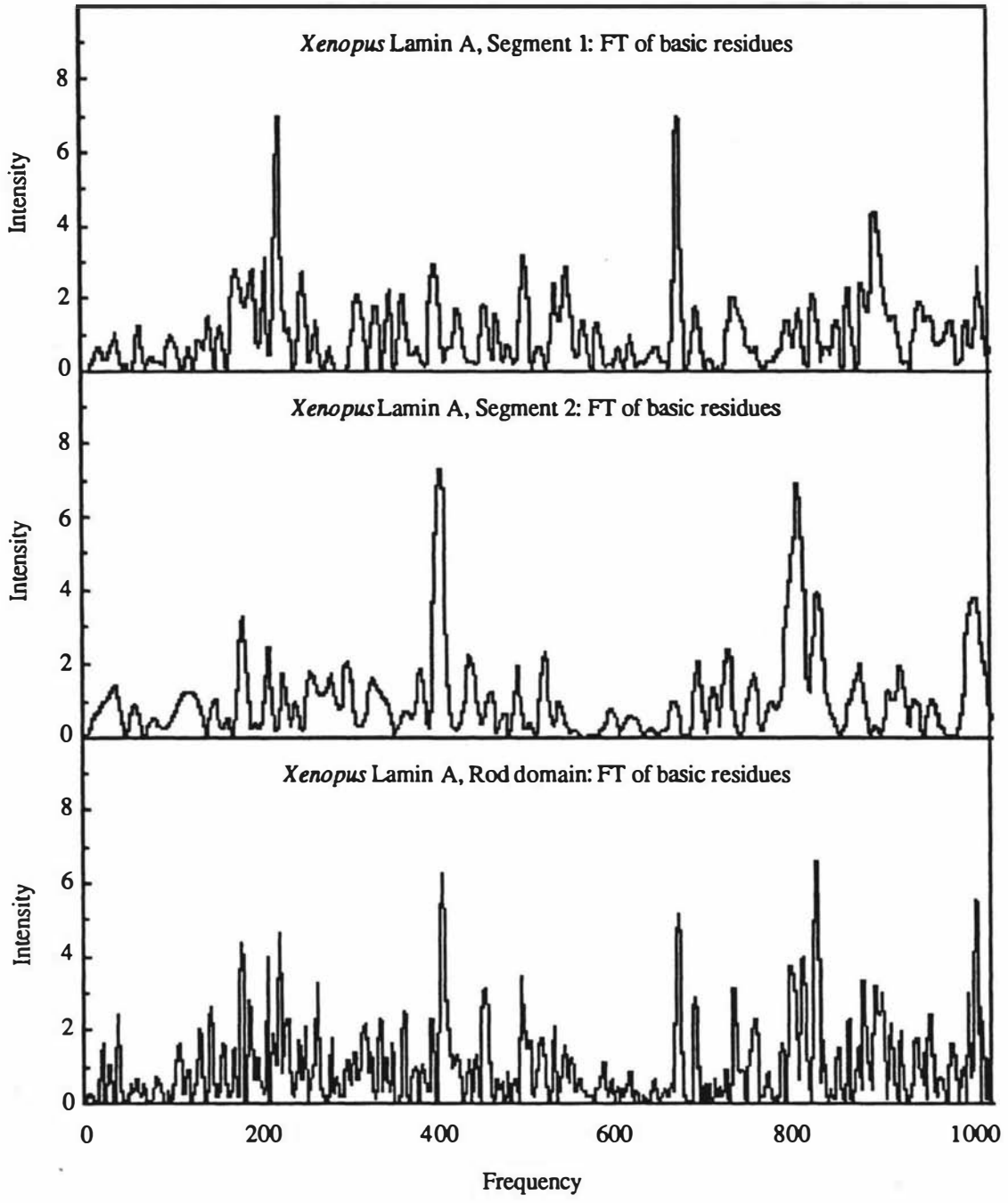


Figure B-3b Fourier transforms of the dispositions of basic residues in rod domain segments of the *Xenopus* lamin A protein. A selection of peaks are listed in Table B-3. Period is related to frequency by the expression:

$$\text{period} = \frac{2048}{\text{frequency}}$$

Period (residues)	*Xenopus* Lamin B Segment 1B acidics	Segment 1B basics	Segment 2 acidics	Segment 2 basics	Rod domain acidics	Rod domain basics
14.629	4.507					
12.190					4.509	
11.703		4.441				
11.571						5.504
10.089			4.403			
9.942						3.183
9.846	7.405			3.139	7.964	
9.225		4.603				
8.291		4.505				
7.341						4.928
4.267						5.141
3.568			4.025			
3.525	4.004					
3.513					4.981	
3.030		6.207				
3.025						4.490
2.817	4.227					
2.775			4.052			
2.768					5.903	
2.333	6.131					
2.243		4.246				
2.238						3.585
2.038						4.889
2.022					4.863	

Table B-4 Peaks in the Fourier transforms of acidic and basic residues in the major coiled-coil segments of *Xenopus* lamin B (as for Table B-1). See Figure B-4.

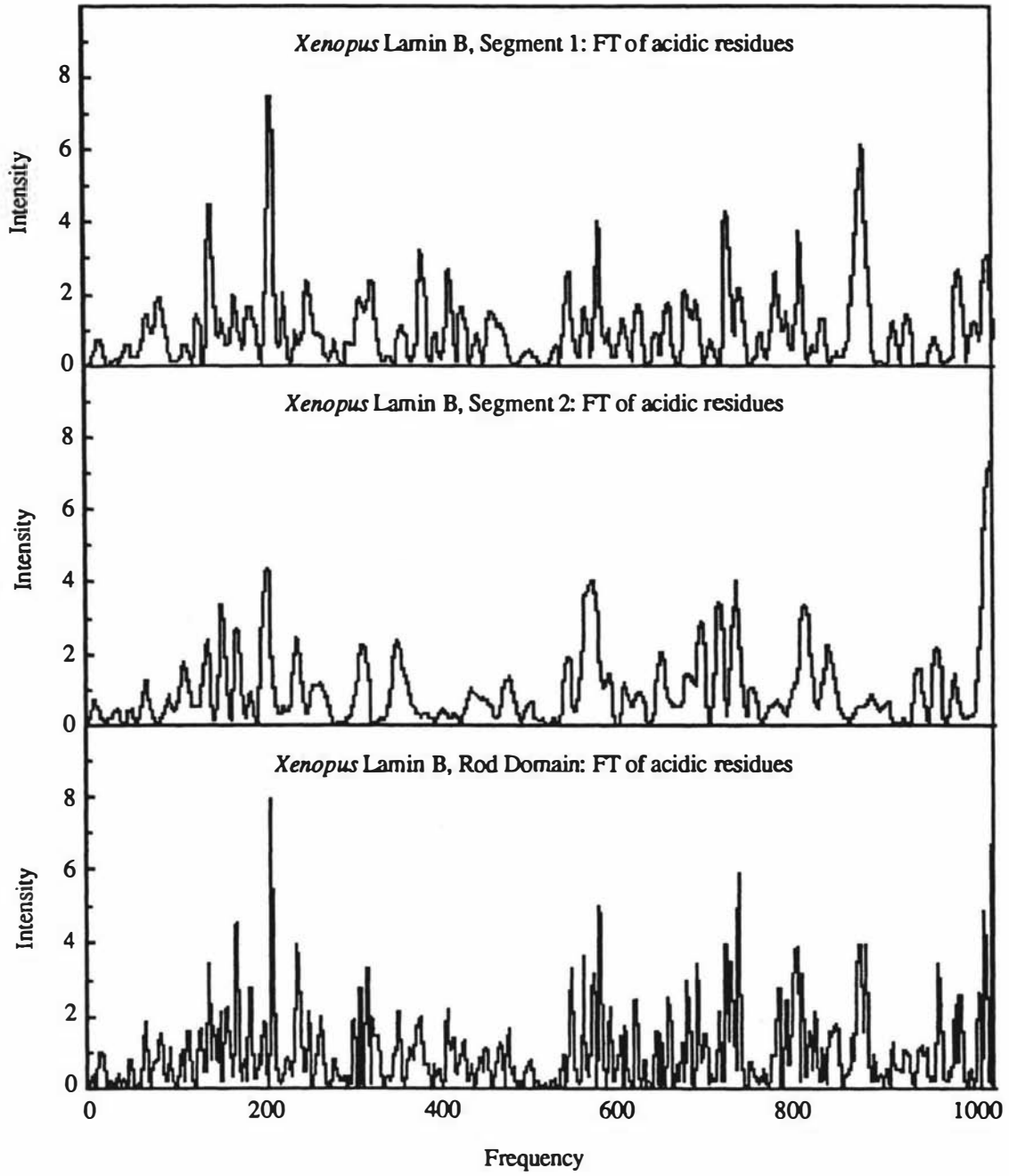


Figure B-4a Fourier transforms of the dispositions of acidic residues in rod domain segments of the *Xenopus* lamin B protein. A selection of peaks are listed in Table B-4. Period is related to frequency by the expression:

$$\text{period} = \frac{2048}{\text{frequency}}$$

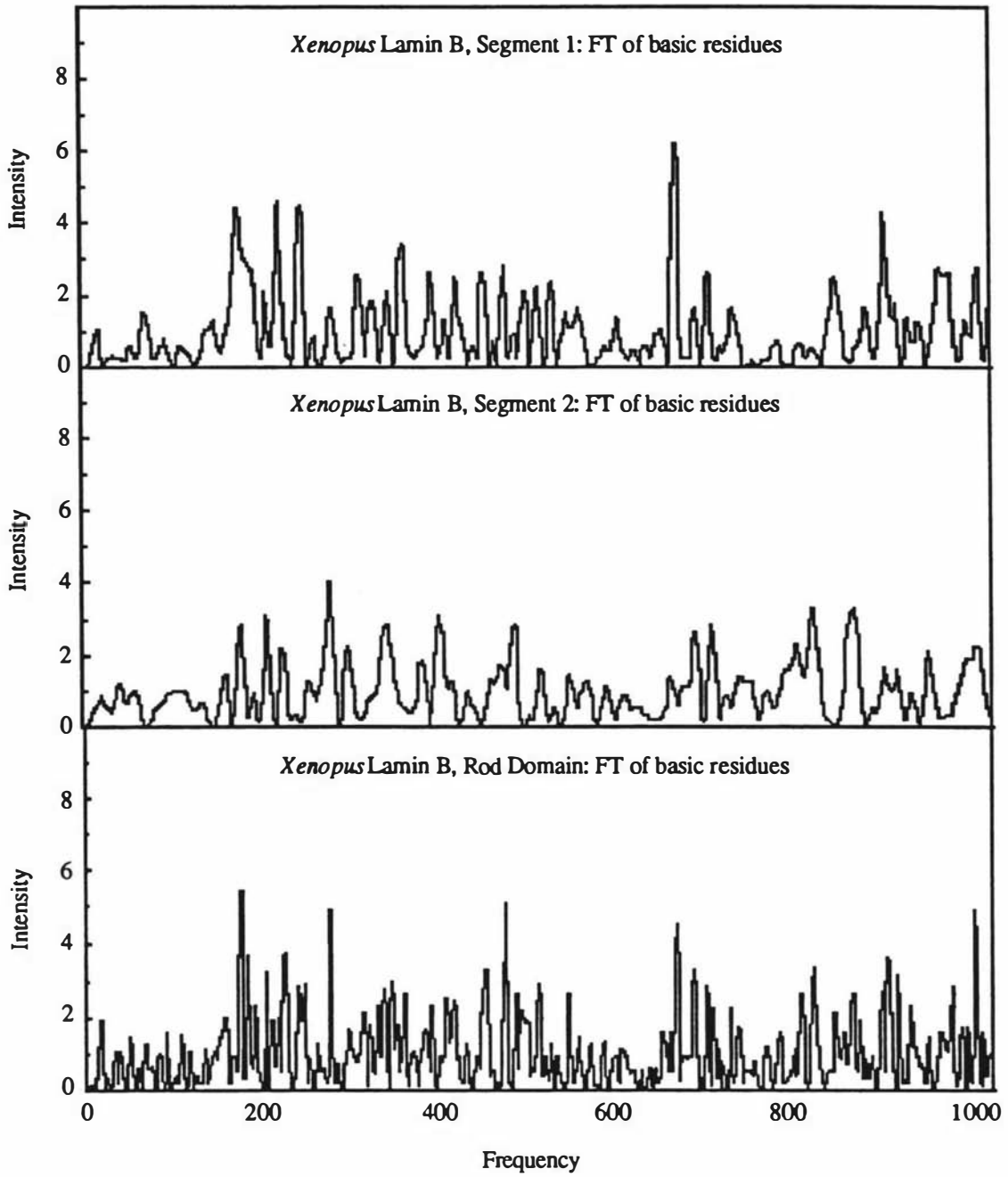


Figure B-4b Fourier transforms of the dispositions of basic residues in rod domain segments of the *Xenopus* lamin A protein. A selection of peaks are listed in Table B-4. Period is related to frequency by the expression:

$$\text{period} = \frac{2048}{\text{frequency}}$$

Period (residues)	*Helix pomotia* B Segment 1B acidics	Segment 1B basics	Segment 2 acidics	Segment 2 basics	Rod domain acidics	Rod domain basics
12.190		5.146				7.362
10.894		4.345				4.273
10.039	4.310					
9.846					3.028	
9.526						3.469
9.352				6.310		
9.309		5.673				
9.267					5.540	
9.225						9.860
9.062			5.574			
8.225	4.057					
6.502				4.739		
5.347						4.530
4.112						5.108
3.513	4.614					
2.951	4.339					
2.926						4.120
2.840						4.325
2.447					4.279	
2.263				4.509		
2.167			5.088			
2.158					4.177	
2.077	4.166					

Table B-5 Peaks in the Fourier transforms of acidic and basic residues in the major coiled-coil segments of the *Helix pomotia* B protein (as for Table B-1). See Figure B-5.

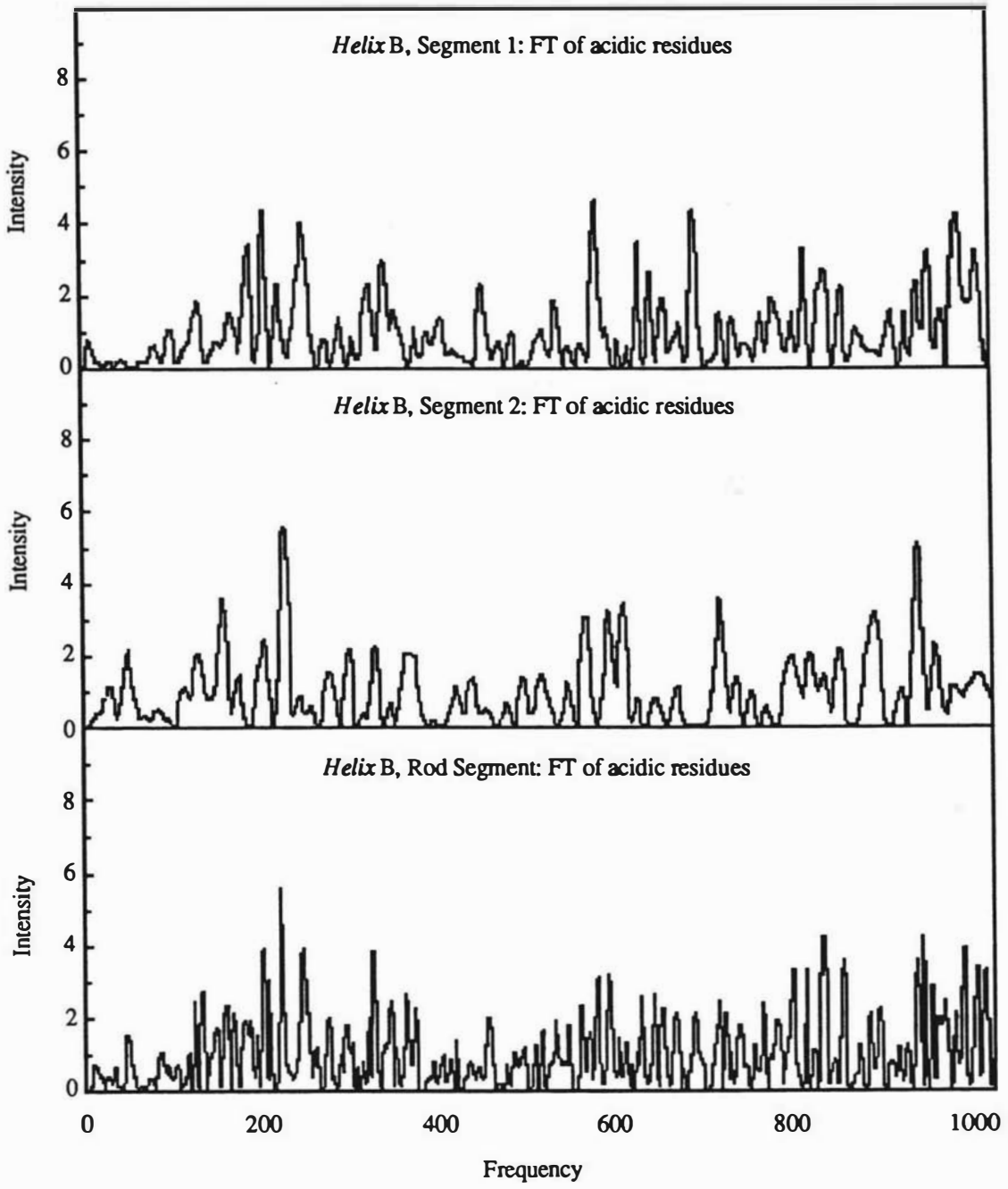


Figure B-5a Fourier transforms of the dispositions of acidic residues in rod domain segments of the *Helix pomotia* B protein. A selection of peaks are listed in Table B-5. Period is related to frequency by the expression:

$$\text{period} = \frac{2048}{\text{frequency}}$$

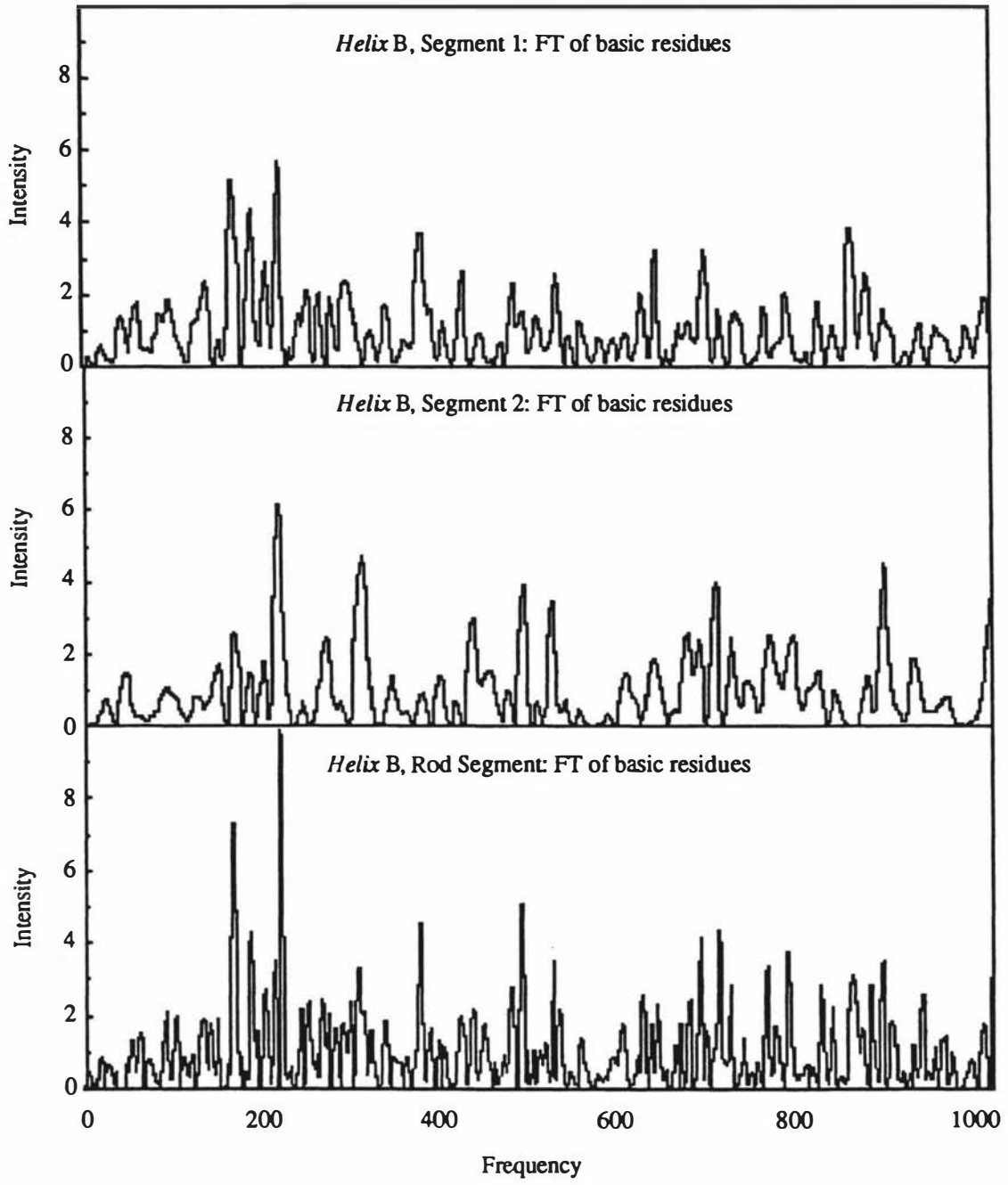


Figure B-5b Fourier transforms of the dispositions of basic residues in rod domain segments of the *Helix pomotia* B protein. A selection of peaks are listed in Table B-5. Period is related to frequency by the expression:

$$\text{period} = \frac{2048}{\text{frequency}}$$

Appendix C : Curve Smoothing

Curve smoothing is performed by convolution of the raw data with a Gaussian function of the form :

$$g(x) = \exp\left(- \frac{4 \ln 2 \, x^2}{\Delta^2} \right)$$

where $-\infty<x<\infty$ and Δ is the half width of the Gaussian function. The Fourier transform G(X) of the Gaussian function g(x) is given by

$$G(X) = \mathbf{F}\{g(x)\} = \exp\left(- \frac{\pi^2 \Delta^2 X^2}{4 \ln 2} \right)$$

where X is the Fourier (or reciprocal) space coordinate. For ease of computation, the smoothing operation is performed in Fourier space in the following manner :

$$h(x) * g(x) \overset{FT}{\Rightarrow} H(X) \cdot G(X) = S(X) \overset{invFT}{\Rightarrow} s(x)$$

where H(X) and S(X) are the Fourier transforms of the unsmoothed h(x) and the smoothed s(x) respectively. The * is the symbol for convolution and the • symbolizes multiplication. Although this series of operations looks more complicated than directly convoluting h(x) and g(x), it is in fact computationally faster because it involves fewer basic arithmetic operations.

Appendix D : Intermolecular Ionic Interactions

Ionic interactions for the sequences studied in Section 4.2 are listed here. The scores for human lamins A and C and for peripherin, which are also given in Section 4.2, are reproduced here for completeness. In addition, the interaction curves corresponding to the tables are shown in this Appendix.

Parallel molecules				Antiparallel molecules			
	Δz(U–U)	S	S_0		Δz(U–D)	S	S_0
1BU-1BU	±54	32	30	**1BU-1BD**	3	68	60
	±4	58	58				
2U-2U	±138	6	6	**2U-2D**	-65	44	36
	±113	20	18		-34	48	48
	±103	24	20		4	60	56
	±93	24	24		13	64	52
	±72	32	30		34	44	44
	±63	34	42		43	56	40
	±34	44	42		53	40	40
	±26	46	44		64	36	36
	±5	64	50		74	32	32
					100	28	24
					110	24	20
					121	16	16
					132	12	12
					135	8	8
					143	4	4
1BU-2U	-133	10	10	**1BU-2D**	-85	32	32
	-94	34	28		-34	50	48
	-82	32	32		-25	46	46
	-71	38	38		4	48	46
	-55	46	44		33	36	34
	-43	48	48		44	32	30
	-14	48	48		53	26	26

Table D-1 Significant ionic interactions between rod domain segments of peripherin molecules as a function of relative axial stagger (see Figure D-1). Δz(U-U) and Δz(U-D) signify the relative staggers between parallel and antiparallel molecules respectively. Only those scores (S) greater than or equal to the 0.975 quantile (S_0) are listed.

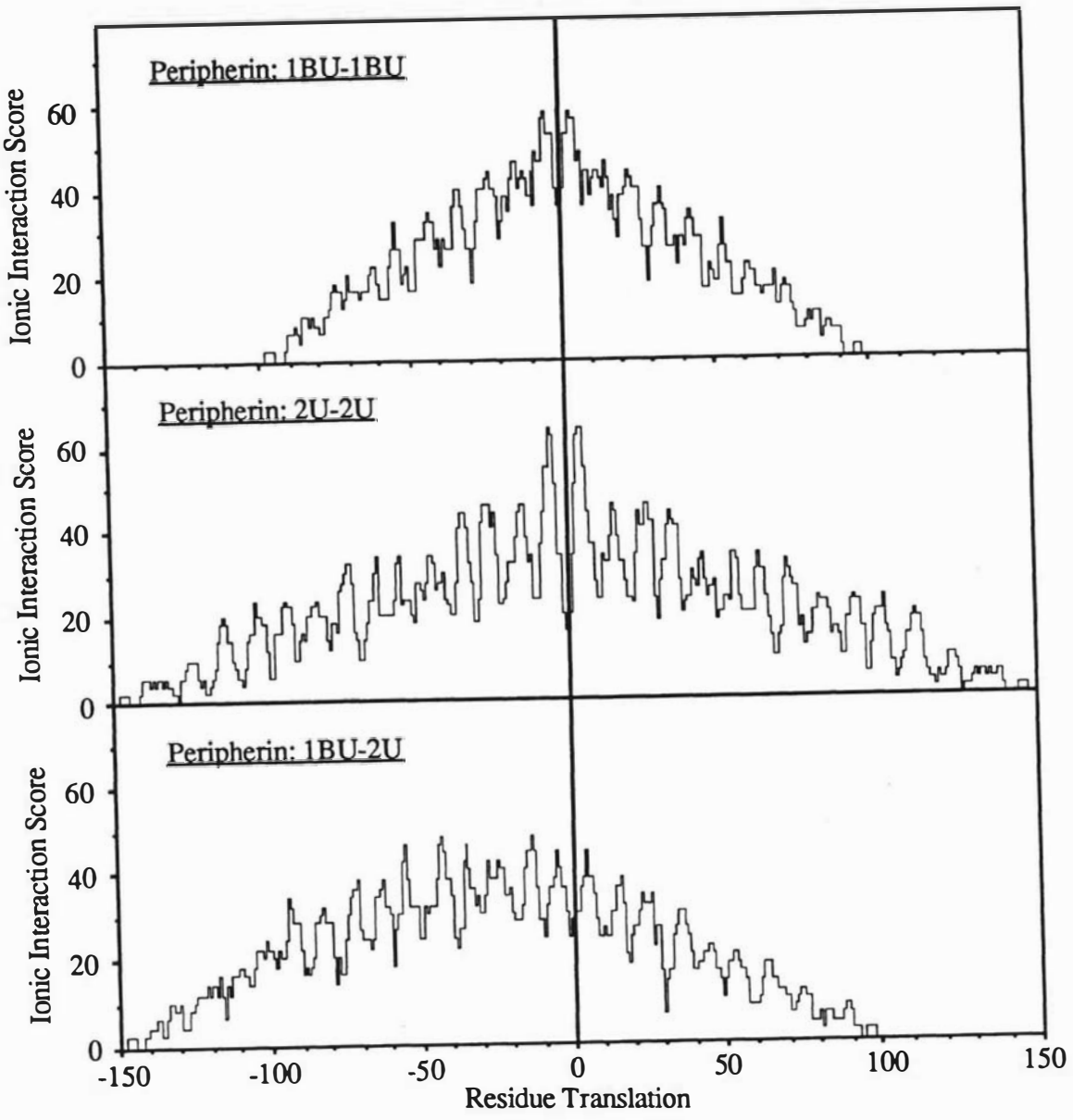


Figure D-1a Ionic interaction curves for parallel segments of rat peripherin.

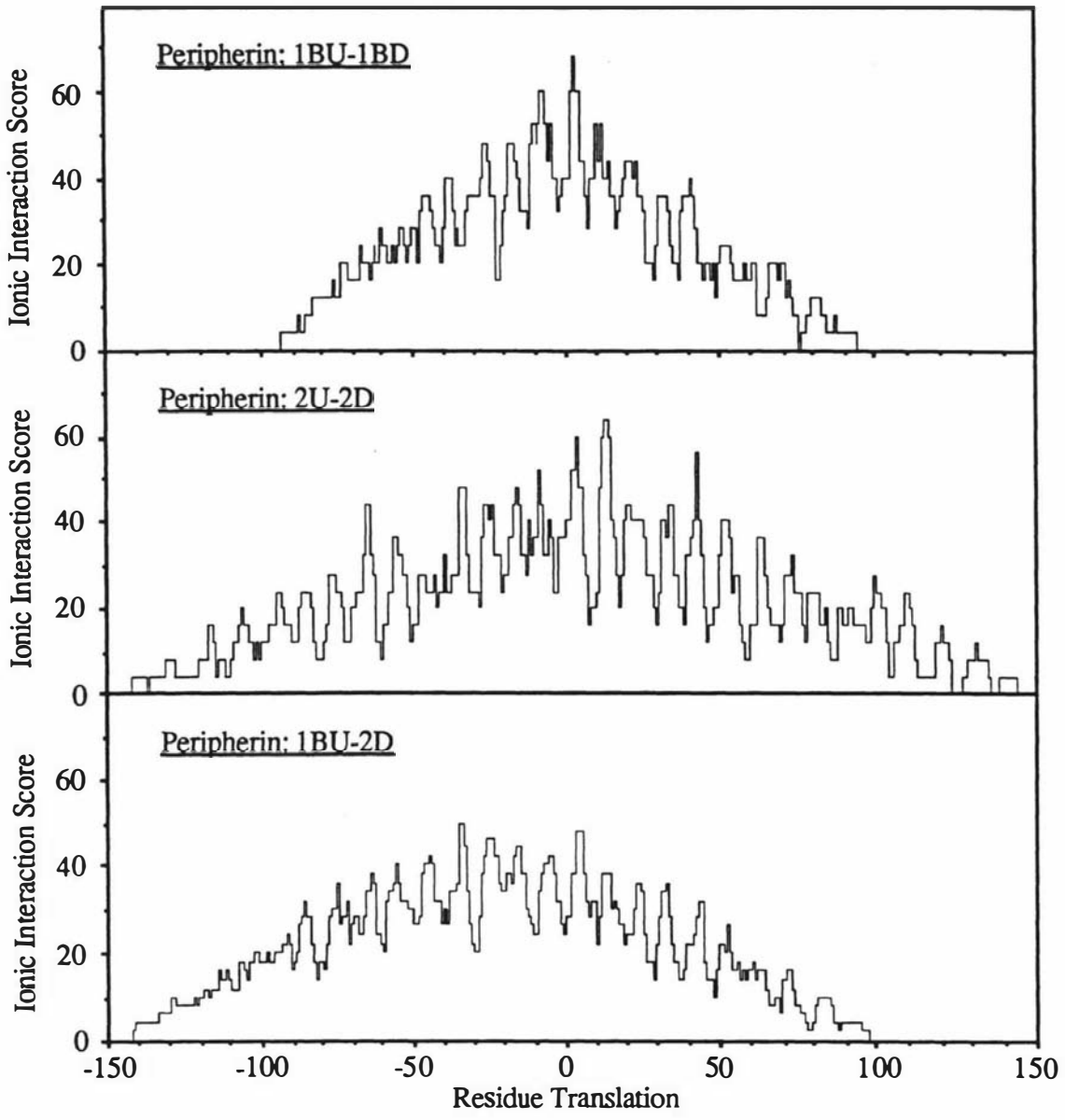


Figure D-1b Ionic interaction curves for antiparallel segments of rat peripherin.

Parallel molecules			Antiparallel molecules		
Δz(U–U)	S	S_0	Δz(U–D)	S	S_0
±345	8	8	-264	56	52
±311	24	24	-184	84	84
±297	32	30	-164	96	92
±294	32	32	-125	116	112
±284	40	36	-117	124	112
±281	40	38	-114	112	112
±267	44	42	-105	116	116
±262	44	44	15	164	148
±229	58	58	24	148	144
±193	70	70	60	140	132
±164	84	84	94	124	124
±136	100	96	145	100	100
±123	106	102	160	96	96
±114	106	104	282	40	40
±4	158	146	288	36	36
			340	12	12

Table D-2 Significant ionic interactions between the rod domains of human lamin A molecules (or lamin C molecules) as a function of relative axial stagger (see Figure D-2). Δz(U-U) and Δz(U-D) signify the relative staggers between parallel and antiparallel molecules respectively. Only those scores (S) greater than or equal to the 0.975 quantile (S_0) are listed.

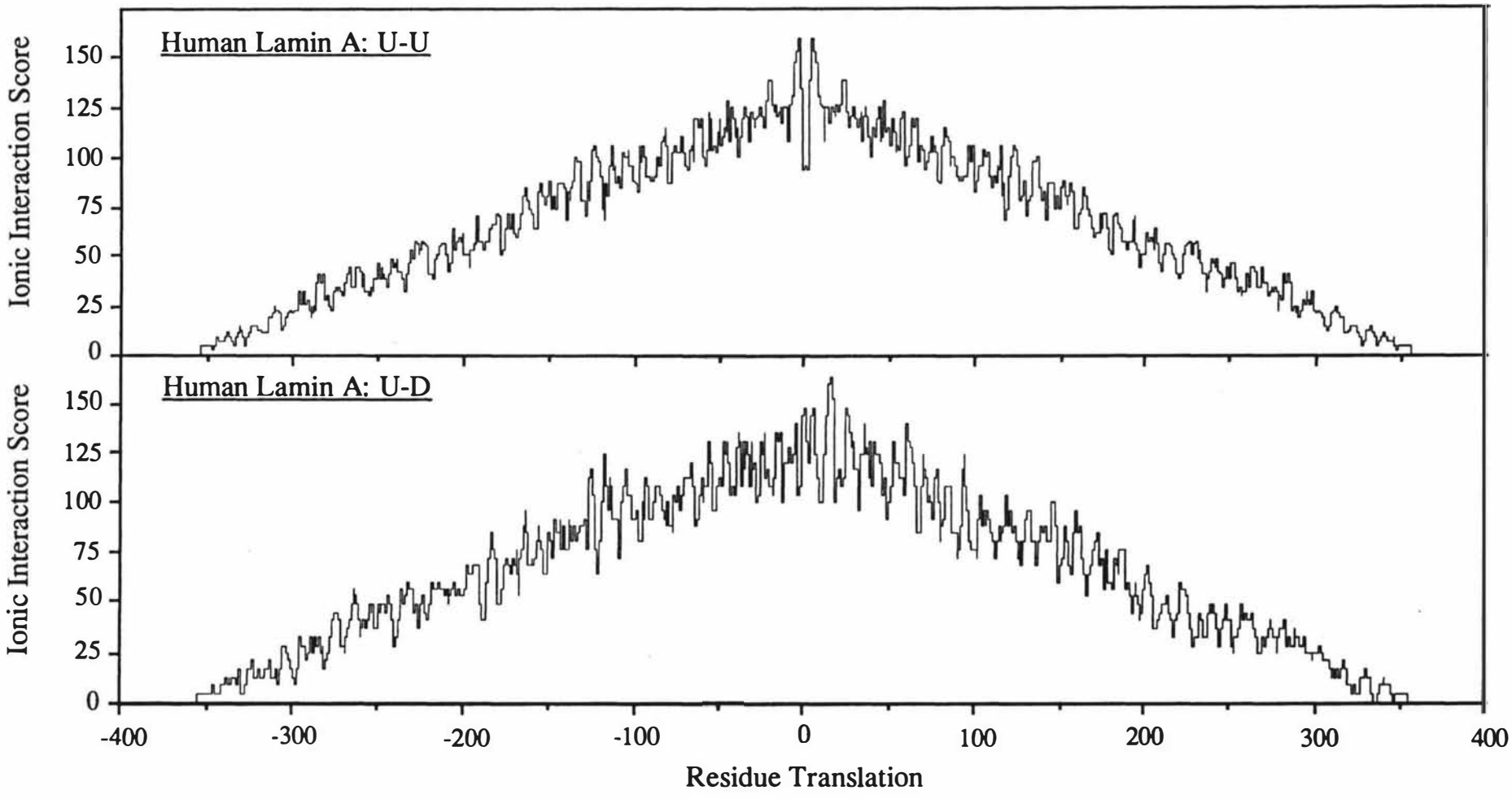


Figure D-2 Ionic interaction curves for parallel (U-U) and antiparallel (U-D) rod domains of human lamin A. The curves for human lamin C are the same as these because the rod domain sequences are identical. Figure adapted from Parry *et al* (1986).

Parallel molecules			Antiparallel molecules		
Δz(U–U)	S	S_0	Δz(U–D)	S	S_0
±345	8	8	-295	36	36
±297	30	30	-264	56	52
±282	40	36	-184	92	84
±267	46	42	-164	96	88
±263	46	44	-144	100	100
±229	56	56	-125	116	108
±136	98	94	-117	116	112
±123	102	100	-69	128	128
±115	104	102	-36	140	140
±21	140	138	-24	144	144
±4	154	146	0	152	152
			15	164	148
			24	156	144
			60	140	132
			69	128	128
			94	120	120
			146	100	100
			283	40	40
			288	36	36
			340	12	12

Table D-3 Significant ionic interactions between the rod domains of *Xenopus* lamin A molecules as a function of relative axial stagger (see Figure D-3). Δz(U-U) and Δz(U-D) signify the relative staggers between parallel and antiparallel molecules respectively. Only those scores (S) greater than or equal to the 0.975 quantile (S_0) are listed.

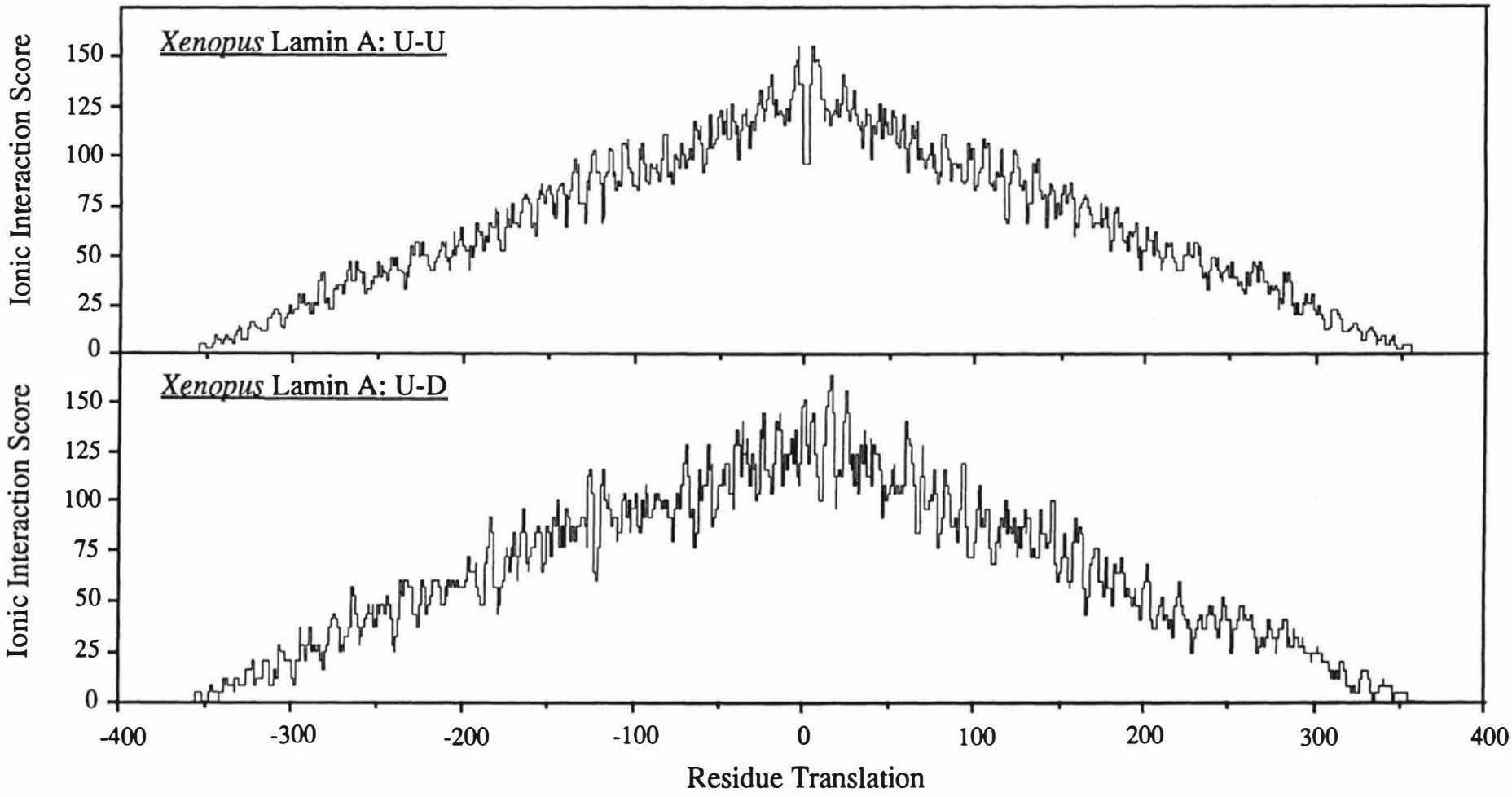


Figure D-3 Ionic interaction curves for parallel (U-U) and antiparallel (U-D) rod domains of *Xenopus* lamin A.

Parallel molecules			Antiparallel molecules		
Δz(U–U)	S	S_0	Δz(U–D)	S	S_0
±352	4	4	-353	4	4
±339	10	10	-313	24	24
±294	32	28	-283	36	36
±281	38	32	-281	36	36
±262	46	38	-264	44	44
±259	40	38	-158	84	80
±251	50	42	-36	132	120
±242	46	44	-6	128	128
±151	80	76	4	128	128
±114	90	88	53	112	112
±4	134	124	86	104	104
			130	92	92
			160	92	84
			171	84	76
			221	64	60
			239	56	52
			260	48	44
			281	40	36
			330	16	16
			340	12	12
			352	4	4

Table D-4 Significant ionic interactions between the rod domains of *Xenopus* lamin B molecules as a function of relative axial stagger (see Figure D-4). Δz(U-U) and Δz(U-D) signify the relative staggers between parallel and antiparallel molecules respectively. Only those scores (S) greater than or equal to the 0.975 quantile (S_0) are listed.

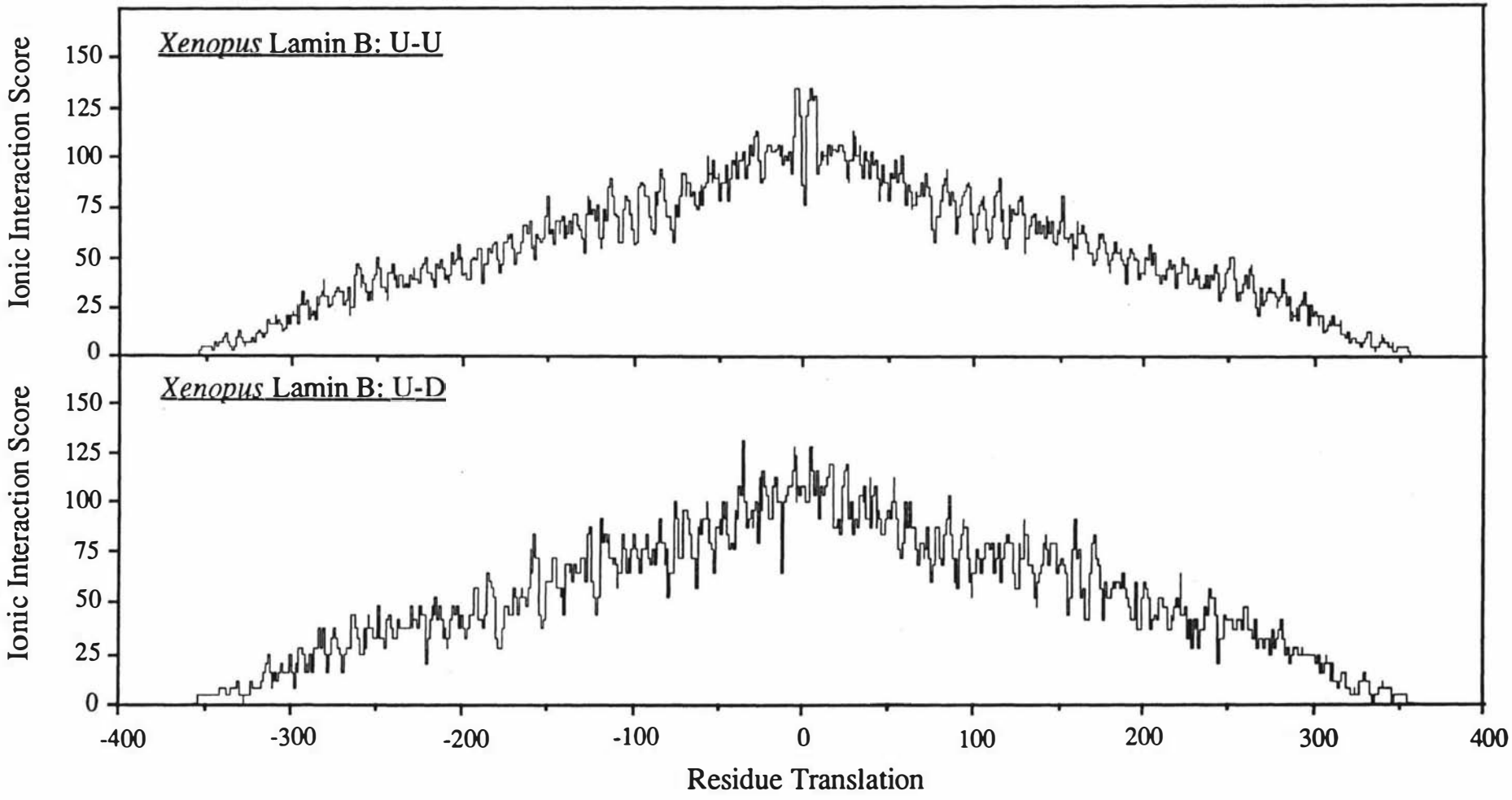


Figure D-4 Ionic interaction curves for parallel (U-U) and antiparallel (U-D) rod domains of *Xenopus* lamin B.

Parallel molecules			Antiparallel molecules		
Δz(U–U)	S	S_0	Δz(U–D)	S	S_0
±348	6	4	-349	4	4
±335	10	10	-341	8	8
±310	20	20	-334	12	12
±300	22	22	-322	16	16
±198	56	52	-264	52	40
±180	58	58	-214	56	56
±160	68	66	-199	64	60
±153	72	68	-164	72	72
±125	80	76	-155	76	76
±114	82	80	-142	84	80
±105	84	82	-125	92	84
±77	94	92	-105	88	88
±6	124	112	-96	92	92
			-79	104	100
			-40	108	108
			-34	120	112
			133	80	80
			326	16	16
			336	12	12
			348	4	4

Table D-5 Significant ionic interactions between the rod domains of *Helix pomotia* B molecules as a function of relative axial stagger (see Figure D-5). Δz(U-U) and Δz(U-D) signify the relative staggers between parallel and antiparallel molecules respectively. Only those scores (S) greater than or equal to the 0.975 quantile (S_0) are listed.

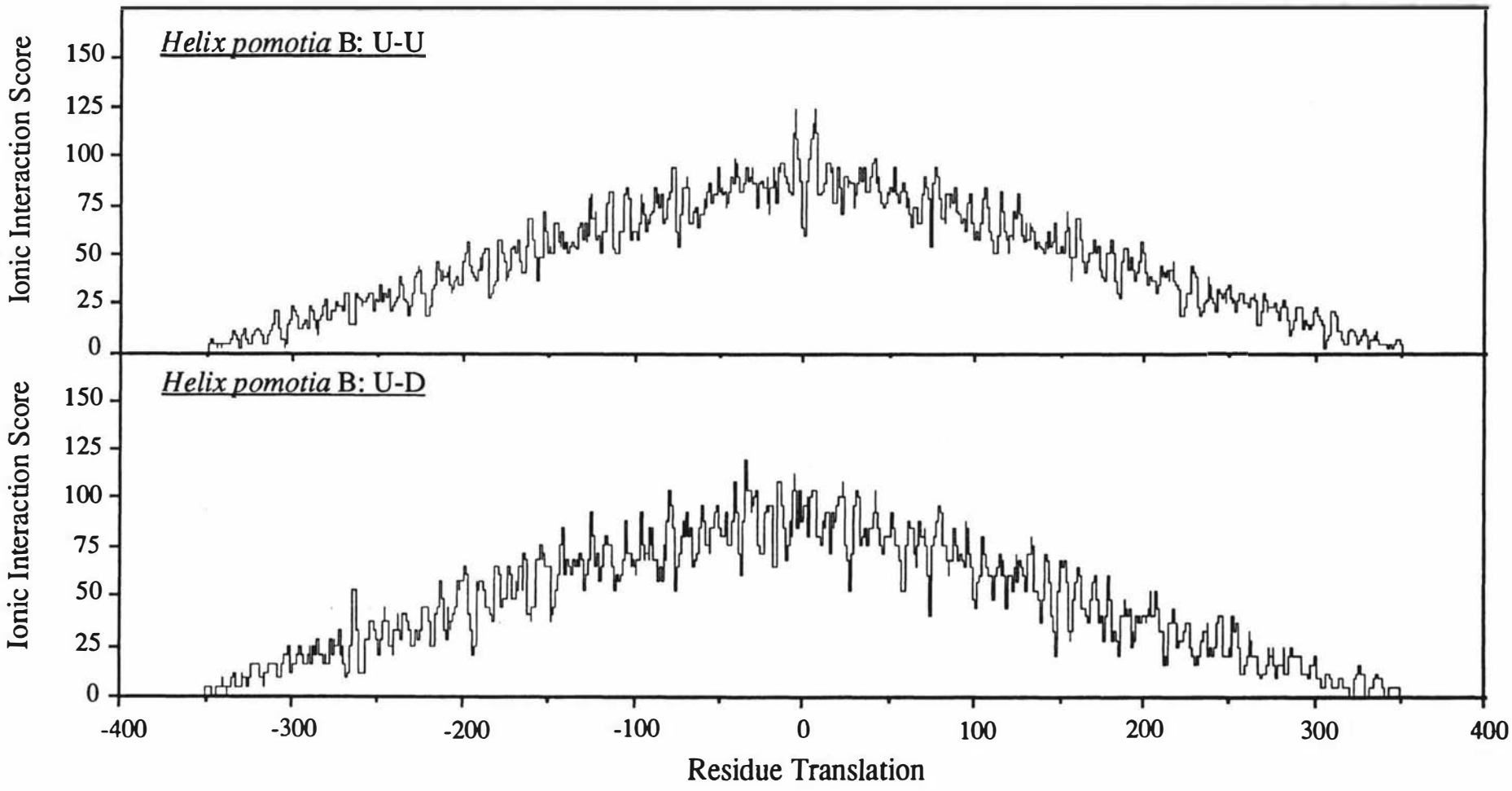


Figure D-5 Ionic interaction curves for parallel (U-U) and antiparallel (U-D) rod domains of the *Helix pomotia* B protein.

BIBLIOGRAPHY

Aebi, U., Fowler, W.E., Rew, P. and Sun, T.-T. (1983) J. Cell Biol., **97**, 1131-1143
The fibrillar substructure of keratin filaments unravelled

Aebi, U., Cohn, J., Buhle, L. and Gerace, L. (1986) Nature *(London)*, **323**, 560-564
The nuclear lamina is a meshwork of intermediate-type filaments

Ahmadi, B. and Speakman, P.T. (1978) FEBS Lett., **94**, 365-367
Suberimidate crosslinking shows that a rod-shaped, low cystine, high helix protein prepared by limited proteolysis of reduced wool has four protein chains

Ahmadi, B., Boston, N.M., Dobb, M.G. and Speakman, P.T. (1980) in "Fibrous Proteins: Scientific, Industrial and Medical Aspects", eds D.A.D. Parry and L.K.Creamer, Academic Press, London, **2**, 161-166
Possible four-chain repeating unit in the microfibril of wool

Aletta, J.M., Shelanski, M.L. and Greene, L.A. (1989) J. Biol. Chem., **264**, 4619-4627
Phosphorylation of the peripherin 58-kDa neuronal intermediate filament protein: Regulation by nerve growth factor and other agents

Ambrose, E.J., Elliott, A. and Temple, R.B. (1949) Nature *(London)*, **163**, 859-862
New evidence on the structure of some proteins from measurements with polarized infrared radiation

Astbury, W.T. and Bell, F.O. (1941) Nature *(London)*, **147**, 696-699
Nature of the intramolecular fold in α-keratin and α-myosin

Astbury, W.T. and Marwick, T.C. (1932) Nature *(London)*, **130**, 309-310
X-ray interpretation of the molecular structure of feather keratin

Astbury, W.T. and Street, A. (1931) Phil. Trans. Roy. Soc. London, **230A**, 75-101
X-ray studies on the structure of hair, wool and related fibres I: General

Astbury, W.T. and Woods, H.J. (1933) Phil. Trans. Roy. Soc. London, **232A**, 333-394
X-ray studies on the structure of hair, wool and related fibres II: The molecular structure and elastic properties of hair keratin

Astbury, W.T., Reed, R. and Spark, L.C. (1948) Biochem. J., **43**, 282-287
An X-ray and electron microscope study of tropomyosin

Bader, B.L., Magin, T.M., Hatzfeld, M. and Franke, W.W. (1986) EMBO J., **5**, 1865-1875
Amino acid sequence and gene organization of cytokeratin no. 19, an exceptional tail-less intermediate filament protein

Bamford, C.H., Brown, L., Elliott, A., Hanby, W.E. and Trotter, I.F. (1952) Nature *(London)*, **169**, 357-358
Structure of synthetic polypeptides

Bear, R.S. (1944) J. Amer. Chem. Soc., **66**, 1297-1305
X-ray diffraction studies on protein fibres: I. The large fibre-axis period of collagen

Birbeck, M.S.C. and Mercer, E.H. (1957) J. Biophys. Biochem. Cytol., **3**, 203-214
The electron microscopy of the human hair follicle: I. Introduction and the hair cortex

Bragg, W.L., Kendrew, J.C. and Perutz, M.F. (1950) Proc. Roy. Soc. London, **203A**, 321-357
Polypeptide chain configurations in crystalline proteins

Brown, L. and Trotter, I.F. (1956) Trans. Faraday Soc., **52**, 537-548
X-ray studies of poly-L-alanine

Burke, B. and Gerace, L. (1986) Cell, **44**, 639-652
A cell free system to study reassembly of the nuclear envelope at the end of mitosis

Burley, S.K. and Petsko, G.A. (1985) Science, **229**, 23-28
Aromatic-aromatic interaction: A mechanism of protein structure stabilization

Chou, P.Y. and Fasman, G.D. (1974) Biochemistry, **13**, 222-245
Prediction of protein conformation

Chou, Y.-H., Rosevear, E and Goldman, R.D. (1989) Proc. Natl. Acad. Sci. USA, **86**, 1885-1889
Phosphorylation and disassembly of intermediate filaments in mitotic cells

Cochran, W., Crick, F.H.C. and Vand, V. (1952) Acta Crystallogr., **5**, 581-586
The structure of synthetic polypeptides 1. The transforms of atoms on a helix

Cohen, C. and Holmes, K.C. (1963) J. Mol. Biol., **6**, 423-432
X-ray diffraction evidence of α-helical coiled-coils in native muscle

Cohen, C. and Parry, D.A.D. (1986) Trends Biochem. Sci., **11**, 245-248.
α-Helical coiled-coils - a widespread motif in proteins

Cohen and Parry, D.A.D. (1989) Proteins: Structure, Function and Genetics, **6**, in press.
α-Helical coiled coils and bundles: How to design an α-helical protein

Cohen, C., Lanar, D.E. and Parry, D.A.D. (1987) Biosci. Rep., **7**, 11-16
Amino acid sequence and structural repeats in schistosome paramyosin match those of myosin

Conway, J.F. and Parry, D.A.D. (1988) Int. J. Biol. Macromol., **10**, 79-98
Intermediate filament structure: 3. Analysis of sequence homologies

Conway, J.F. and Parry, D.A.D. (1989) in "Cytoskeletal and Extracellular Proteins: Structure, Interactions and Assembly", eds U. Aebi and J. Engel, Springer Series in Biophysics, Springer-Verlag, Berlin, **3**, 140-149
Structural and spatial organization of intermediate filament and nuclear lamin proteins

Conway, J.F., Fraser, R.D.B., MacRae, T.P. and Parry, D.A.D. (1989) in "The Biology of Wool and Hair", eds G.E. Rogers, P.J. Reis, K.A. Ward and R.C. Marshall, Chapman and Hall, London and New York, 127-144
Protein chains in wool and epidermal keratin IF: Structural features and spatial arrangement

Crewther, W.G. (1976) Proc. 5th Int. Wool Text. Res. Conf. Aachen, **1**, 1-101
Primary structure and chemical properties of wool

Crewther, W.G. and Harrap, B.S. (1967) J. Biol. Chem., **242**, 4310-4319
The preparation and properties of a helix-rich fraction obtained by partial proteolysis of low sulphur S-carboxymethylkerateine from wool

Crewther, W.G. and Dowling, L.M. (1971) Appl. Polym. Symp., **18**, 1-20
Preparation and properties of large peptides from the helical regions of the low-sulphur proteins of wool

Crewther, W.G., Fraser, R.D.B., Lennox, F.G. and Lindley, H. (1965) Adv. Protein Chem., **20**, 191-346
The chemistry of keratins

Crewther, W.G., Gillespie, J.M., Harrap, B.S. and Inglis, A.S. (1966) Biopolymers, **4**, 905-916
Low-sulphur proteins from α-keratins. Interrelations between their amino acid compositions, α-helix contents, and the supercontraction of the parent keratin

Crewther, W.G., Dobb, M.G., Dowling, L.M. and Harrap, B.S. (1968) in "Symposium on Fibrous Proteins", ed W.G. Crewther, Butterworths, 329-340
The structure and aggregation of low-sulphur proteins derived from alpha keratins

Crewther, W.G., Inglis, A.S. and McKern, N.M. (1978a) Biochem. J., **173**, 365-371
Amino-acid sequences of α-helical segments from S-carboxymethylkerateine A: Complete sequence of a type-II segment

Crewther, W.G., Gough, K.H., Inglis, A.S. and McKern, N.M. (1978b) Text. Res. J., **48**, 160-162
Sequence homologies in helical segments from α-keratin

Crewther, W.G., Dowling, L.M., Gruen, L.C., Sparrow, L.G. and Woods, E.F. (1980) Proc. 6th Int. Wool Text. Res. Conf. Pretoria, **2**, 1-12
Interactions between the microfibrillar proteins of wool

Crewther, W.G., Dowling, L.M., Steinert, P.M. and Parry, D.A.D. (1983) Int. J. Biol. Macromol., **5**, 267-274
Structure of intermediate filaments

Crewther, W.G., Dowling, L.M., Inglis, A.S., Sparrow, L.G., Strike, P.M. and Woods, E.F. (1985) Proc. 7th Int. Wool Text. Res. Conf. Tokyo, **1**, 85-94
Wool microfibrils and intermediate filaments have common structural features

Crick, F.H.C. (1952) Nature *(London)*, **170**, 882-883
Is α-keratin a coiled-coil?

Crick, F.H.C. (1953) Acta Crystallogr, **6**, 689-697
The packing of α-helices: Simple coiled-coils

Dautigny, A., Pham-Dinh, D., Roussel, C., Felix, J.M., Nussbaum, J.L. and Jolles, P. (1988) Biochem, Biophys. Res. Commun., **154**, 1099-1106
The large neurofilament subunit (NF-H) of the rat: cDNA cloning and in situ *detection*

Doolittle, R.F., Goldbaum, D.M. and Doolittle, L.R. (1978) J. Mol. Biol., **120**, 311-325
Designation of sequences involved in the "coiled-coil" interdomainal connections in fibrinogen: Construction of an atomic scale model

Dowling, L.M. and Crewther, W.G. (1972) Proc. Aust. Biochem. Soc., **5**, 3
Distribution of two types of α-helical units in the low-sulphur protein chains from α-keratin

Dowling, L.M., Parry, D.A.D. and Sparrow, L.G. (1983) Biosci. Rep., **3**, 73-78
Structural homology between hard α-keratin and the intermediate filament proteins desmin and vimentin

Dowling, L.M., Crewther, W.G. and Inglis, A.S. (1986) Biochem. J., **236**, 695-703
The primary structure of component 8c-1, a subunit protein of intermediate filaments in wool keratin

Eckert, R.L. (1988) Proc. Natl. Acad. Sci. USA, **85**, 1114-1118
Sequence of the human 40-kDa keratin reveals an unusual structure with very high sequence identity to the corresponding bovine keratin

Elleman, T.C., Crewther, W.G. and van der Touw, J. (1978) Biochem. J., **173**, 387-391
Amino-acid sequence of α-helical segments from S-carboxymethylkerateine A: Statistical analysis

Elliott, A. and Malcolm, B.R. (1956) Nature *(London)*, **178**, 192
Absolute configuration and optical rotation of folded (α) polypeptides

Elliott, A. and Malcolm, B.R. (1959) Proc. Roy. Soc. (London.), **249A**, 30-41
Chain arrangement and sense of the α-helix in poly-L-alanine fibres

Elliott, A. and Offer, G. (1978) J. Mol. Biol., **123**, 505-519
Shape and flexibility of the myosin molecule

Elliott, A., Lowy, J., Parry, D.A.D. and Vibert, P.J. (1968) Nature *(London)*, **218**, 656-659
Puzzle of the coiled-coils in the paramyosin filament of molluscan muscles

Fawcett, D.W. (1966) Amer. J. Anat., **199**, 129-146
On the occurrence of a fibrous lamina on the inner aspect of the nuclear envelope in certain cells of vertebrates

Filshie, B.K. and Rogers, G.E. (1961) J. Mol. Biol., **3**, 784-786
The fine structure of α-keratin

Fisher, D.Z., Chaudhary, N. and Blobel, G. (1986) Proc. Natl. Acad. Sci. USA, **83**, 6450-6454
cDNA sequencing of nuclear lamins A and C reveals primary and secondary structural homology to intermediate filament proteins

Franke, W.W., Grund, C. and Achtstätter, T. (1986) J. Cell Biol., **103**, 1933-1943
Co-expression of cytokeratins and neurofilament proteins in a permanent cell line: cultured rat PC12 cells combine neuronal and epithelial features

Fraser, R.D.B. and MacRae, T.P. (1959) Biochem. Biophys. Acta, **29**, 229-240
Structural implications of the equatorial X-ray diffraction pattern of α-keratin

Fraser, R.D.B. and MacRae, T.P. (1973a) Polymer, **14**, 61-67
The structure of α-keratin

Fraser, R.D.B. and MacRae, T.P. (1973b) "Conformation in Fibrous Proteins and Related Synthetic Polypeptides", Academic Press, New York and London

Fraser, R.D.B. and MacRae, T.P. (1983) Biosci. Rep., **3**, 517-525
The structure of the α-keratin microfibril

Fraser, R.D.B. and MacRae, T.P. (1985) Biosci. Rep., **5**, 573-579
Intermediate filament structure

Fraser, R.D.B. and MacRae, T.P. (1988) Int. J. Biol. Macromol., **10**, 178-184
Surface lattice in α-keratin filaments

Fraser, R.D.B., MacRae, T.P. and Rogers, G.E. (1959) Nature *(London)*, **183**, 592-594
Structure of α-keratin

Fraser, R.D.B., MacRae, T.P. and Rogers, G.E. (1962) Nature *(London)*, **193**, 1052-1055
Molecular organization of α-keratin

Fraser, R.D.B., MacRae, T.P. and Miller, A. (1964a) Nature *(London)*, **203**, 1231-1233
Molecular structure of α-keratin

Fraser, R.D.B., MacRae, T.P. and Miller, A. (1964b) J. Mol. Biol., **10**, 147-156
The coiled-coil model of α-keratin structure

Fraser, R.D.B., MacRae, T.P. and Miller, A. (1964c) Acta Crystallogr., **17**, 813-816
The Fourier transform of the coiled-coil model for α-keratin

Fraser, R.D.B., MacRae, T.P. and Miller, A. (1965) J. Mol. Biol., **14**, 432-442
X-ray diffraction patterns of α-fibrous proteins

Fraser, R.D.B., MacRae, T.P. Millward, G.R., Parry, D.A.D., Suzuki, E. and Tulloch, P.A. (1971) Appl. Polym. Symp., **18**, 65-83
The molecular structure of keratins

Fraser, R.D.B., MacRae, T.P. and Rogers, G.E. (1972) "Keratins: Their composition, structure and biosynthesis", Charles C. Thomas, Springfield, USA

Fraser, R.D.B., Gillespie, J.M. and MacRae, T.P. (1973) Comp. Biochem. Physiol., **44B**, 943-947
Tyrosine-rich proteins in keratins

Fraser, R.D.B., MacRae, T.P. and Suzuki, E. (1976) J. Mol. Biol., **108**, 435-452
Structure of the α-keratin microfibril

Fraser, R.D.B., Jones, L.N., MacRae, T.P., Suzuki, E. and Tulloch, P.A. (1980) Proc. 6th Int. Wool Text. Res. Conf. Pretoria, **1**, 1-33
The fine structure of the wool fibre

Fraser, R.D.B., MacRae, T.P., Suzuki, E. and Parry, D.A.D. (1985) Int. J. Biol. Macromol., **7**, 258-274
Intermediate filament structure: 2. Molecular interactions in the filament

Fraser, R.D.B., MacRae, T.P., Parry, D.A.D. and Suzuki, E. (1986) Proc. Natl. Acad. Sci. USA, **83**, 1179-1183
Intermediate filaments in α-keratins

Fraser, R.D.B., MacRae, T.P., Sparrow, L.G. and Parry, D.A.D. (1988) Int. J. Biol. Macromol., **10**, 106-112
Disulphide bonding in α-keratin

Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol., **120**, 97-120
Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins

Geisler, N. and Weber, K. (1981) Proc. Natl. Acad. Sci. USA, **78**, 4120-4123
Comparison of the proteins of two immunologically distinct intermediate-sized filaments by amino acid sequence analysis: Desmin and vimentin

Geisler, N. and Weber, K. (1982) EMBO J., **1**, 1649-1656
The amino acid sequence of chicken muscle desmin provides a common structural model for intermediate filament proteins

Geisler, N. and Weber, K. (1983) EMBO J., **2**, 2059-2063
Amino acid sequence data on glial fibrillary acidic protein (GFA); Implications for the subdivision of intermediate filaments into epithelial and non-epithelial members

Geisler, N. and Weber, K. (1988) EMBO J., **7**, 15-20
Phosphorylation of desmin in vitro *inhibits formation of intermediate filaments; identification of three kinase A sites in the aminoterminal head domain*

Geisler, N., Kaufmann, E. and Weber, K. (1982a) Cell, **30**, 277-286
Proteinchemical characterization of three structurally distinct domains along the protofilament unit of desmin 10 nm filaments

Geisler, N., Plessmann, U. and Weber, K. (1982b) Nature *(London)*, **296**, 448-450
Related amino acid sequences in neurofilaments and non-neuronal intermediate filaments

Geisler, N., Kaufmann, E., Fischer, S., Plessmann, U. and Weber, K. (1983) EMBO J., **2**, 1295-1302
Neurofilament architecture combines structural principles of intermediate filaments with carboxy-terminal extensions increasing in size between triplet proteins

Geisler, N., Fischer, S., Vandekerckhove, J., Plessmann, U. and Weber, K. (1984) EMBO J., **3**, 2701-2706
Hybrid character of a large neurofilament protein (NF-M): Intermediate filament type sequence followed by a long and acidic carboxy-terminal extension

Geisler, N., Fischer, S., Vandekerckhove, J., Van Damme, J., Plessmann, U. and Weber, K. (1985a) EMBO J., **4**, 57-63
Protein-chemical characterization of NF-H, the largest mammalian neurofilament component; intermediate filament-type sequences followed by a unique carboxy-terminal extension

Geisler, N., Kaufmann, E.. and Weber, K. (1985b) J. Mol. Biol., **182**, 173-177
Antiparallel orientation of the two double-stranded coiled-coils in the tetrameric protofilament unit of intermediate filaments

Geisler, N., Plessmann, U. and Weber, K. (1985c) FEBS Lett., **182**, 475-478
The complete amino acid sequence of the major mammalian neurofilament protein (NF-L)

Geisler, N., Vandekerckhove, J. and Weber, K. (1987) FEBS Lett., **221**, 403-407
Location and sequence characterization of the major phosphorylation sites of the high molecular mass neurofilament proteins M and H

Georgatos, S.D. and Marchesi, V.T. (1985) J. Cell Biol., **100**, 1955-1961
The binding of vimentin to human erythrocyte membranes: A model system for the study of intermediate filament-membrane interactions

Georgatos, S.D. and Blobel, G. (1987) J. Cell Biol., **105**, 117-125
Lamin B constitutes an intermediate filament attachment site at the nuclear envelope

Georgatos, S.D., Weaver, D.C. and Marchesi, V.T. (1985) J. Cell Biol., **100**, 1962-1967
Site specificity in vimentin-membrane interactions: Intermediate filament subunits associate with the plasma membrane via their head domains

Georgatos, S.D., Weber, K., Geisler, N. and Blobel, G. (1987) Proc. Natl. Acad. Sci. USA, **84**, 6780-6784
Binding of two desmin derivatives to the plasma membrane and the nuclear envelope of avian erythrocytes: Evidence for a conserved site-specificity in intermediate filament-membrane interactions

Georgatos, S.D., Maroulakou, I. and Blobel, G. (1989) J. Cell Biol., **108**, 2069-2082
Lamin A, lamin B, and lamin B receptor analogues in yeast

Gerace, L. (1985) Nature *(London)*, **318**, 508-509
Traffic control and structural proteins in the eukaryotic nucleus

Gerace, L. (1986) Trends Biochem. Sci., **11**, 443-446
Nuclear lamina and organization of nuclear architecture

Gerace, L. and Blobel, G. (1980) Cell, **19**, 277-287
The nuclear envelope lamina is reversibly depolymerized during mitosis

Gerace, L. and Blobel, G. (1982) Cold Spring Harbor Symp. Quant Biol., **46**, 967-978
Nuclear lamina and the structural organization of the nuclear envelope

Gerace, L. and Burke, B. (1988) Ann. Rev. Cell Biol., **4**, 335-374
Functional organization of the nuclear envelope

Gerace, L. Blum, A. and Blobel, G. (1978) J. Cell Biol., **79**, 546-566
Immunocytochemical localization of the major polypeptides of the nuclear pore complex-lamina fraction

Glass, C., Kim, H.K. and Fuchs, E. (1985) J. Cell Biol., **101**, 2366-2373
Sequence and expression of a human type II mesothelial keratin

Goldman, A.E., Maul, G., Steinert, P.M., Yang, H.-Y. and Goldman, R.D. (1986) Proc. Natl. Acad. Sci. USA, **83**, 3839-3843
Keratin-like proteins that coisolate with intermediate filaments of BHK-21 cells are nuclear lamins

Goldman, R.D. and Dessev, G.N. (1989) in "The Biology of Wool and Hair", eds G.E. Rogers, P.J. Reis, K.A. Ward and R.C. Marshall, Chapman and Hall, London and New York, 87-95
Intermediate filaments: Problems and perspectives

Gough, K.H., Inglis, A.S. and Crewther, W.G. (1978) Biochem. J., **173**, 373-385
Amino-acid sequences of α-helical segments from S-carboxymethylkerateine A: Complete sequence of a type-I segment

Green, K.J., Parry, D.A.D., Steinert, P.M., Virata, M.L.A., Wagner, R.M., Angst, B.D. and Nilles, L.A. (1989) J. Biol. Chem., in press
Structure of the human desmoplakins: Implications for function in the desmosomal plaque

Gruen, L.C. and Woods, E.F. (1983) Biochem. J., **209**, 587-595
Structural studies on the microfibrillar proteins of wool

Hanukoglu, I. and Fuchs, E. (1982) Cell, **31**, 243-252
The cDNA sequence of a human epidermal keratin: Divergence of sequence but conservation of structure among intermediate filament proteins

Hanukoglu, I. and Fuchs, E. (1983) Cell, **33**, 915-924
The cDNA sequence of a type II cytoskeletal keratin reveals constant and variable structural domains among keratins

Harrap, B.S. (1963) Aust. J. Biol. Sci., **16**, 231-240
The conformation of a soluble wool keratin derivative

Heald, R., Loewinger, L., McKeon, F. (1988) J. Cell Biol., **107**, 661a
Using mutants expressed in vitro *to define domains in the nuclear lamin proteins required for the assembly of the nuclear envelope*

Hodge, A.J. and Petruska, J.A. (1963) in "Aspects of Protein Structure", ed G.N. Ramachandran, Academic Press, London, 289-300
Recent studies with the electron microscope on ordered aggregates of the tropocollagen molecule

Hodges, R.S., Sodek, J., Smillie, L.B. and Jurasek, L. (1972) Cold Spring Harbor Symp. Quant. Biol., **37**, 399-410
Tropomyosin: Amino acid sequence and coiled-coil structure

Hoffmann, W. and Franz, J.K. (1984) EMBO J., **3**, 1301-1306
Amino acid sequence of the carboxyl-terminal part of an acidic type I cytokeratin of molecular weight 51000 from Xenopus laevis *epidermis as predicted from the cDNA sequence*

Hoffmann, W., Franz, J.K., and Franke, W.W. (1985) J. Mol. Biol., **184**, 713-724
Amino acid sequence microheterogeneities of basic (type II) cytokeratins of Xenopus laevis *epidermis and evolutionary conservativity of helical and non-helical domains*

Hogg, D.McC., Dowling, L.M. and Crewther, W.G. (1978) Biochem. J., **173**, 353-363
Amino-acid sequences of α-helical segments from S-carboxymethylkerateine A: Tryptic and chymotryptic peptides from a type-II segment

Hopp, T.P. and Woods, K.R. (1981) Proc. Natl. Acad. Sci. USA, **78**, 3824-3828
Prediction of protein antigenic determinants from amino acid sequences

Huber, R. (1979) Trends Biochem. Sci., **4**, 271-276
Conformational flexibility and its functional significance in some protein molecules

Huggins, M.L. (1943) Chem. Rev., **32**, 195-218
The structure of fibrous proteins

Huggins, M.L. (1952) J. Amer. Chem. Soc., **74**, 3963
Polypeptide helices in proteins

Hulmes, D.J.S., Miller, A., Parry, D.A.D., Piez, K.A. and Woodhead-Galloway, J. (1973) J. Mol. Biol., **79**, 137-148
Analysis of the primary structure of collagen for the origins of molecular packing

Inagaki, M., Nishi, Y., Nishizawa, K., Matsuyama, M. and Sato, C. (1987) Nature *(London)*, **328**, 649-652
Site-specific phosphorylation induces disassembly of vimentin filaments in vitro

Inglis, A.S., Sutherland, W.J. and Woods, E.F. (1983) Proc. Aust. Biochem. Soc., **15**, 31
Coiled-coil particles from wool: Sequenator analysis using a stationary cup procedure

Ip, W., Hartzer, M.K., Pang, Y.-Y.S. and Robson, R.M. (1985) J. Mol. Biol., **183**, 365-375
Assembly of vimentin in vitro *and its implications concerning the structure of intermediate filaments*

Jonas, E., Sargent, T.D. and Dawid, I.B. (1985) Proc. Natl. Acad. Sci. USA, **82**, 5413-5417
Epidermal keratin gene expressed in embryos of Xenopus laevis

Jorcano, J.L., Rieger, M., Franz, J.K., Schiller, D.L., Moll, R. and Franke, W.W. (1984) J. Mol. Biol., **179**, 257-281
Identification of two types of keratin polypeptides within the acidic cytokeratin subfamily I

Julien, J.-P., Ramachandran, K. and Grosveld, F. (1985) Biochem. Biophys. Acta, **825**, 398-404
Cloning of a cDNA encoding the smallest neurofilament protein from the rat

Julien, J.-P., Grosveld, F., Yazdanbaksh, K., Flavell, D., Meijer, D. and Mushynski, W. (1987) Biochem. Biophys. Acta, **909**, 10-20
The structure of a human neurofilament gene (NF-L): A unique exon-intron organization in the intermediate filament gene family

Julien, J.-P., Cote, F., Beaudet, L., Sidkey, M., Flavell, D., Grosveld, F.G. and Mushynski, W.E. (1988) Gene, **68**, 307-314
Sequence and structure of the mouse gene coding for the largest neurofilament subunit

Karplus, P.A. and Schulz, G.E. (1985) Naturwissenschaften, **72**, 212-213
Prediction of chain flexibility in proteins

Kaufmann, E., Weber, K. and Geisler, N. (1985) J. Mol. Biol., **185**, 733-742
Intermediate filament forming ability of desmin derivatives lacking the amino-terminal 67 or the carboxy-terminal 27 residues

Kendrew, J.C., Dickerson, R.E., Strandberg, B.E., Hart, R.G., Davies, D.R., Phillips, D.C. and Shore, V.C. (1960) Nature *(London)*, **185**, 422-427
Structure of myoglobin: Three-dimensional Fourier synthesis at 2 Å resolution

Knapp, B., Rentrop, M., Schweizer, J. and Winter, H. (1987) J. Biol. Chem., **262**, 938-945
Three cDNA sequences of mouse type I keratins: Cellular localization of the mRNAs in normal and hyperproliferate tissues

Krieg, T.M., Schafer, M.P., Cheng, C.K., Filpula, D., Flaherty, P., Steinert, P.M. and Roop, D.R. (1985) J. Biol. Chem., **260**, 5867-5870
Organization of a type I keratin gene: Evidence for evolution of intermediate filaments from a common ancestral gene

Krohne, G., Wolin, S.L., McKeon, F.D., Franke, W.W. and Kirschner, M.W. (1987) EMBO J., **6**, 3801-3808
Nuclear lamin L_1 of Xenopus laevis*: cDNA cloning, amino acid sequence and binding specificity of a member of the lamin B subfamily*

Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol., **157**, 105-132
A simple method for displaying the hydropathic character of a protein

Lersch, R. and Fuchs, E. (1988) Mol. Cell. Biol., 8, 489-493
Sequence and expression of a type II keratin, K5, in human epidermal cells

Lee, L.D. and Baden, H.P. (1976) Nature *(London)*, **264**, 377-379
Organisation of the polypeptide chains in mammalian keratin

Lees, J.F., Shneidman, P.S., Skuntz, S.F., Carden, M.J. and Lazzarini, R.A. (1988) EMBO J., **7**, 1947-1955
The structure and organization of the human heavy neurofilament subunit (NF-H) and the gene encoding it

Lehner, C.F., Fürstenberger, G., Eppenberger, H.M. and Nigg, E.A. (1986a) Proc. Natl. Acad. Sci. USA, **83**, 2096-2099
Biogenesis of the nuclear lamina: In vivo *synthesis and processing of nuclear protein precursors.*

Lehner, C.F., Kurer, V., Eppenberger, H.M. and Nigg, E.A. (1986b) J. Biol. Chem., **261**, 13293-13301
The nuclear lamin protein family in higher vertebrates: Identification of quantitively minor lamin proteins by monoclonal antibodies

Lehnert, M.E., Jorcano, J.L., Zentgraf, H., Blessing, M., Franz, J.K. and Franke, W.W. (1984) EMBO J., **3**, 3279-3287
Characterization of bovine keratin genes: Similarities of exon patterns in genes coding for different keratins

Leonard, D.G.B., Gorham, J.D., Cole, P., Greene, L.A. and Ziff, E.B. (1988) J. Cell Biol., **106**, 181-193
A nerve growth factor-regulated messenger RNA encodes a new intermediate filament protein

Levy, E., Liem, R.K.H., D'Eustachio, P. and Cowan, N.J. (1987) Eur. J. Biochem., **166**, 71-77
Structure and evolutionary origin of the gene encoding mouse NF-M, the middle-molecular-mass neurofilament protein

Lewis, S.A. and Cowan, N.J. (1985) J. Cell Biol., **100**, 843-850
Genetics, evolution, and expression of the 68,000-mol-wt neurofilament protein: Isolation of a cloned cDNA probe

Lewis, S.A. and Cowan, N.J. (1986) Mol. Cell. Biol., **6**, 1529-1534
Anomalous placement of introns in a member of the intermediate filament multigene family: An evolutionary conundrum

Lewis, S.A., Balcarek, J.M., Krek, V., Shelanski, M. and Cowan, N.J. (1984) Proc. Natl. Acad. Sci. USA, **81**, 2743-2746
Sequence of a cDNA clone encoding mouse glial fibrillary acidic protein: Structural conservation of intermediate filaments

Liem, R.K.H., Yen, S.-H., Salomon, G.D. and Shelanski, M.L. (1978) J. Cell Biol., **79**, 637-645
Intermediate filaments in nervous tissue

Loewinger, L. and McKeon, F.D. (1988) EMBO J., **7**, 2301-2309
Mutations in the nuclear lamin proteins resulting in their aberrant assembly in the cytoplasm

Lotay, S.S. and Speakman, P.T. (1977) Nature *(London)*, **265**, 274-276
Three-chain merokeratin from wool may be a fragment of the microfibril component macromolecule

Lu, Y.-J, and Johnson, P. (1983) Int. J. Biol. Macromol., **5**, 347-350
The N-terminal domain of desmin is not involved in intermediate filament formation: Evidence from thrombic digestion studies

MacArthur, I. (1943) Nature *(London)*, **152**, 38-41
Structure of α-keratin

Mack, J.W., Torchia, D.A. and Steinert, P.M. (1988) Biochem., **27**, 5418-5426
Solid-state NMR studies of the dynamics and structure of mouse keratin intermediate filaments

Magin, T.M., Hatzfeld, N. and Franke, W.W. (1987) EMBO J., **6**, 2607-2615
Analysis of cytokeratin domains by cloning and expression of intact and deleted polypeptides in Escherichia coli

Mandelkow, E., Thomas, J. and Cohen, C. (1977) Proc. Natl. Acad. Sci. USA, **74**, 3370-3374
Microtubule structure at low resolution by x-ray diffraction

Marchuk, D., McCrohan, S. and Fuchs, E. (1984) Cell, **39**, 491-498
Remarkable conservation of structure among intermediate filament genes

Marchuk, D., McCrohan, S. and Fuchs, E. (1985) Proc. Natl. Acad. Sci. USA, **82**, 1609-1613
Complete sequence of a gene encoding a human type I keratin: Sequences homologous to enhancer elements in the regulatory region of the gene

McKeon, F.D., Kirschner, M.W. and Caput, D. (1986) Nature, **319**, 463-468
Homologies in both primary and secondary structure between nuclear envelope and intermediate filament proteins

McLachlan, A.D (1978) J. Mol. Biol., **124**, 297-304
Coiled coil formation and sequence regularities in the helical regions of α-keratin

McLachlan, A.D. and Stewart, M. (1975) J. Mol. Biol., **98**, 293-304
Tropomyosin coiled-coil interactions: Evidence for an unstaggered structure

McLachlan, A.D. and Stewart, M. (1976) J. Mol. Biol., **103**, 271-298
The 14-fold periodicity in α-tropomyosin and the interaction with actin

McLachlan, A.D. and Stewart, M. (1982) J. Mol. Biol., **162**, 693-698
Periodic charge distribution in the intermediate filament proteins desmin and vimentin

McLachlan, A.D. and Kahn, J. (1983) J. Mol. Biol., **164**, 605-626
Periodic features in the amino acid sequence of nematode myosin rod

Milam, L. and Erickson, H.P. (1982) J. Cell Biol., **94**, 592-596
Visualization of a 21-nm axial periodicity in shadowed keratin filaments and neurofilaments

Myers, M.W., Lazzarini, R.A., Lee, V.M.-Y., Schlaepfer, W.W. and Nelson, D.L. (1987) EMBO J., **6**, 1617-1626
The human mid-size neurofilament subunit: A repeated protein sequence and the relationship of its gene to the intermediate filament gene family

Napolitano, E.W., Chin, S.S.M., Colman, D.R. and Liem, R.K.H. (1987) J. Neurosci., **7**, 2590-2599
Complete amino acid sequence and in vitro *expression of rat NF-M, the middle molecular weight neurofilament protein*

O'Donnell, (1969) Aust. J. Biol. Sci., **22**, 471-488
Studies on reduced wool: IX. The N-terminal sequence of a fragment produced by cleavage of component 8 with cyanogen bromide

Osborn, M. and Weber, K. (1982) Cell, **31**, 303-306
Intermediate filaments: Cell-type-specific markers in differentiation and pathology

Osborn, M. and Weber, K. (1983) Lab. Invest., **48**, 372-394
Biology of disease. Tumour diagnosis by intermediate filament typing: A novel tool for surgical pathology

Osborn, M. and Weber, K. (1986) Trends Biochem. Sci., **11**, 469-472
Intermediate filament proteins: A multigene family distinguishing major cell lineages

Ottaviano, Y. and Gerace, L. (1985) J. Biol. Chem., **260**, 624-632
Phosphorylation of the nuclear lamins during interphase and mitosis

Parry, D.A.D. (1974) Biochem. Biophys. Res. Commun., **57**, 216-224
Structural studies on the tropomyosin/troponin complex of vertebrate skeletal muscle

Parry, D.A.D. (1975) J.Mol.Biol., **98**, 519-535
Analysis of the primary sequence of α-tropomyosin from rabbit skeletal muscle

Parry, D.A.D. (1978) J.Mol.Biol., **120**, 545-551
Fibrinogen: A preliminary analysis of the amino acid sequences of the portions of the α, β and γ-chain postulated to form the interdomainal link between globular regions of the molecule

Parry, D.A.D. (1979) in "Fibrous Proteins: Scientific, Industrial and Medical Aspects", eds D.A.D. Parry and L.K.Creamer, Academic Press, London, **1**, 393-427
Determination of structural information from the amino acid sequences of fibrous proteins

Parry, D.A.D. (1981) J.Mol.Biol., **153**, 459-464
Structure of rabbit skeletal myosin: Analysis of the amino acid sequences of two fragments from the rod region

Parry, D.A.D. (1982) Biosci. Rep., **2**, 1017-1024
Coiled-coils in α-helix-containing proteins: Analysis of the residue types within the heptad repeat and the use of these data in the prediction of coiled-coils in other proteins

Parry, D.A.D. (1989) in "Cellular and Molecular Biology of Intermediate Filaments", eds R.D. Goldman and P.M. Steinert, Plenum Press (in press)
Primary and secondary structure of IF protein chains and modes of molecular aggregation

Parry, D.A.D. and Suzuki, E. (1969) Biopolymers, **7**, 199-206
Interchain packing energies of multistranded α-helical ropes

Parry, D.A.D. and Baker, E.N. (1984) Rep. Progr. Phys., **47**, 1133-1232
Biopolymers

Parry, D.A.D. and Fraser, R.D.B. (1985) Int. J. Biol. Macromol., **7**, 203-213
Intermediate filament structure: 1. Analysis of IF protein sequence data

Parry, D.A.D., Crewther, W.G., Fraser, R.D.B. and MacRae, T.P. (1977) J. Mol. Biol., **113**, 449-454
Structure of α-keratin: Structural implication of the amino acid sequences of the type I and type II chain segments

Parry, D.A.D., Steven, A.C. and Steinert, P.M. (1985) Biochem. Biophys. Res. Commun., **127**, 1012-1018
The coiled-coil molecules of intermediate filaments consist of two parallel chains in exact axial register

Parry, D.A.D., Conway, J.F. and Steinert, P.M. (1986) Biochem. J., **238**, 305-308
Structural studies on lamin: Similarities and differences between lamin and intermediate-filament proteins

Parry, D.A.D., Conway, J.F., Goldman, A.E., Goldman, R.D. and Steinert, P.M. (1987a) Int. J. Biol. Macromol., **9**, 137-145
Nuclear lamin proteins: Common structures for paracrystalline, filamentous and lattice forms

Parry, D.A.D., Fraser, R.D.B., MacRae, T.P. and Suzuki, E. (1987b) in "Fibrous Protein Structure", eds J.M. Squire and P.J. Vibert, Academic Press, 193-214
Intermediate filaments

Parysek, L.M. and Goldman, R.D. (1987) J. Neurosci., **7**, 781-791
Characterization of intermediate filaments in PC12 cells

Parysek, L.M., Chisholm, R.L., Ley, C.A. and Goldman, R.D. (1988) Neuron, **1**, 395-401
A type III intermediate filament gene is encoded in mature neurons

Pauling, L. and Corey, R.B. (1950) J. Amer. Chem. Soc, **72**, 5349
Two hydrogen-bonded spiral configurations of the polypeptide chain

Pauling, L. and Corey, R.B. (1951a) Proc. Natl. Acad. Sci. USA, **37**, 235-240
Atomic coordinates and structure factors for two helical configurations of polypeptide chains

Pauling, L. and Corey, R.B. (1951b) Proc. Natl. Acad. Sci. USA, **37**, 729-740
Configurations of polypeptide chains with favoured orientations around single bonds: Two new pleated sheets

Pauling, L. and Corey, R.B. (1953a) Nature *(London)*, **171**, 59-61
Compound helical configurations of polypeptide chains: Structure of proteins of the α-keratin type

Pauling, L. and Corey, R.B. (1953b) Proc. Natl. Acad. Sci. USA, **39**, 253-256
Two rippled-sheet configurations of polypeptide chains, and a note about the pleated sheets

Pauling, L., Corey, R.B. and Branson, H.R. (1951) Proc. Natl. Acad. Sci. USA, **37**, 205-211
The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain

Perutz, M.F. (1951a) Nature *(London)*, **167**, 1053-1054
New x-ray evidence on the configuration of polypeptide chains: Polypeptide chains in poly-γ-benzyl-L-glutamate, keratin and hæmoglobin

Perutz, M.F. (1951b) Nature *(London)*, **168**, 653-654
The 1.5-Å reflection from proteins and polypeptides

Peter, M., Kitten, G.T., Lehner, C.F., Vorburger, K., Bailer, S.M., Maridor, G. and Nigg, E.A. (1989) J. Mol. Biol., **208**, 393-404
Cloning and sequencing of cDNA clones encoding chicken lamins A and B_1 and comparison of the primary structures of vertebrate A- and B-type lamins

Phillips, G.N., Fillers, J.P. and Cohen, C. (1986) J. Mol. Biol., **192**, 111-131
Tropomyosin crystal structure and muscle regulation

Portier, M.-M., Brachet, P., Croizat, B. and Gros, F. (1984a) Dev. Neurosci., **6**, 215-226
Regulation of peripherin in mouse neuroblastoma and rat PC12 pheochromocytoma cell lines

Portier, M.-M., de Nechaud, B. and Gros, F. (1984b) Dev. Neurosci., **6**, 335-344
Peripherin, a new member of the intermediate filament protein family

Quax, W., Egberts, W.V., Hendricks, W., Quax-Jeuken, Y. and Bloemendal, H. (1983) Cell, **35**, 215-223
The structure of the vimentin gene

Quax, W., van der Heuvel, R., Egberts, W.V., Quax-Jeuken, Y. and Bloemendal, H. (1984) Proc. Natl. Acad. Sci. USA, **81**, 5970-5974
Intermediate filament cDNAs from BHK-21 cells: Demonstration of distinct genes for desmin and vimentin in all vertebrate classes

Quax, W., van den Broek, L., Egberts, W.V., Ramaekers, F. and Bloemendal, H. (1985) Cell, **43**, 327-338
Characterization of the hamster desmin gene: Expression and formation of desmin filaments in nonmuscle cells after gene transfer

Quax-Jeuken, Y.E.F.M., Quax, W.J., and Bloemendal, H. (1983) Proc. Natl. Acad. Sci. USA, **80**, 3548-3552
Primary and secondary structure of hamster vimentin predicted from the nucleotide sequence

Quinlan, R.A. and Franke, W.W. (1982) Proc. Natl. Acad. Sci. USA, **79**, 3452-3456
Heteropolymer filaments of vimentin and desmin in vascular smooth muscle tissue and cultured baby hamster kidney cells demonstrated by chemical crosslinking

Quinlan, R.A. and Franke, W.W. (1983) Europ. J. Biochem., **132**, 477-484
Molecular interactions in intermediate-sized filaments revealed by chemical crosslinking: Heteropolymers of vimentin and glial filament protein in cultured human glioma cells

Quinlan, R.A., Cohlberg, J.A., Schiller, D.L., Hatzfeld, M. and Franke, W.W. (1984) J. Mol. Biol., **178**, 365-388
Heterotypic tetramer (A_2D_2) complexes of non-epidermal keratins isolated from cytoskeletons of rat hepatocytes and hepatoma cells

Quinlan, R.A., Moir, R.D. and Stewart, M. (1989) J. Cell Sci., **93**, 71-83
Expression in Escherichia coli *of fragments of glial fibrillary acidic protein: characterization, assembly properties and paracrystal formation*

Raychaudury, A., Marchuk, D., Lindhurst, M. and Fuchs, E. (1986) Mol. Cell. Biol., **6**, 539-548
Three tightly linked genes encoding human type I keratins: Conservation of sequence in the 5'-untranslated leader and the 5'-upstream regions of coexpressed keratin genes

Rieger, M., Jorcano, J.L. and Franke, W.W. (1985) EMBO J., **4**, 2261-2267
Complete sequence of a bovine type I cytokeratin gene: conserved and variable intron positions in genes of polypeptides of the same cytokeratin family

Rogers, G.E. (1959) J. Ultrastruct. Res., **2**, 309-330
Electron microscopy of wool

Rogers, G.E. (1984) Biochem. Soc. Symp., **49**, 85-108
Studies on keratin multigene families

Rose, G.D. (1978) Nature *(London)*, **272**, 586-590
Prediction of chain turns in globular proteins on a hydrophobic basis

Sauk, J.J., Krunweide, M., Cocking-Johnson, D. and White, J.G. (1984) J. Cell Biol., **99**, 1590-1597
Reconstitution of cytokeratin filaments in vitro*: Further evidence for the role of nonhelical peptides in filament assembly*

Scheer, U., Kartenbeck, J., Trendelenburg, M.F., Stadler, J. and Franke, W.W. (1976) J. Cell Biol., **69**, 1-18
Experimental disintegration of the nuclear envelope: Evidence for pore-connecting fibrils

Shelton, K.R., Higgins, L.L., Cochran, D.L., Ruffolo, J.J. and Egle, P.M. (1980) J. Biol. Chem., **255**, 10978-10983
Nuclear lamins of erythrocyte and liver

Shneidman, P.S., Carden, M.J., Lees, J.F. and Lazzarini, R.A. (1988) Molec. Brain Res., **4**, 217-232
The structure of the largest murine neurofilament protein (NF-H) as revealed by cDNA and genomic sequences

Singer, P.A., Trevor, K. and Oshima, R.G. (1986) J. Biol. Chem., **261**, 538-547
Molecular cloning and characterization of the Endo B cytokeratin expressed in preimplantation mouse embryos

Skerrow, D., Matoltsy, A.G. and Matoltsy, M.N. (1973) J. Biol. Chem., **248**, 4820-4826
Isolation and characterization of the α-helical regions of epidermal prekeratin

Sparrow, L.G. and Inglis, A.S. (1980) Proc. 6th Int. Wool Text. Res. Conf. Pretoria, **2**, 237-246
Characterization of the cyanogen bromide peptides of component 7c, a major microfibrillar protein from wool

Sparrow, L.G., Dowling, L.M., Loke, V.Y. and Strike, P.M. (1989) in "The Biology of Wool and Hair", eds G.E. Rogers, P.J. Reis, K.A. Ward and R.C. Marshall, Chapman and Hall, London and New York, 145-155
Amino acid sequences of wool keratin IF proteins

Steinert, P.M. (1978) J. Mol. Biol., **123**, 49-70
Structure of the three-chain unit of the bovine epidermal keratin filament

Steinert, P.M. (1981) in "Microscopy of Proteins", ed J.R. Harris, Academic Press, London, **1**, 125-166
Intermediate filaments

Steinert, P.M. (1988) J. Biol. Chem., **263**, 13333-13339
The dynamic phosphorylation of the human intermediate filament keratin 1 chain

Steinert, P.M. and Parry, D.A.D. (1985) Ann. Rev. Cell Biol., **1**, 41-65
Intermediate filaments: Conformity and diversity of expression and structure

Steinert, P.M. and Roop, D.R. (1988) Ann. Rev. Biochem., **57**, 593-625
Molecular and cellular biology of intermediate filaments

Steinert, P.M., Zimmerman, S.B., Starger, J.M. and Goldman, R.D. (1978) Proc. Natl. Acad. Sci. USA, **75**, 6098-6101
Ten-nanometer filaments of hamster BHK-21 cells and epidermal keratin filaments have similar structures

Steinert, P.M., Idler, W.W. and Goldman, R.D. (1980a) Proc. Natl. Acad. Sci. USA, **77**, 4534-4538
Intermediate filaments of baby hamster kidney (BHK-21) cells and bovine epidermal keratinocytes have similar ultrastructures and subunit domain structures

Steinert, P.M., Starger, J.M. and Goldman, R.D. (1980b) in "Fibrous Proteins: Scientific, Industrial and Medical Aspects", eds D.A.D. Parry and L.K.Creamer, Academic Press, London, **2**, 227-236
Homologous structures of 100 Å filaments and mammalian keratin filaments

Steinert, P.M., Peck, G.L. and Idler, W.W. (1980c) in "Biochemistry of Normal and Abnormal Epidermal Differentiation", eds. I.A. Bernstein and M. Seiji, University of Tokyo Press, Tokyo, 391-406
Structural changes of human epidermal α-keratin in disorders of keratinization

Steinert, P.M., Idler, W.W., Aynardi-Whitman, M., Zackroff, R.V. and Goldman, R.D. (1982) Cold Spring Harbor Symp. Quant. Biol., **46**, 465-474
Heterogeneity of intermediate filaments assembled in vitro

Steinert, P.M., Rice, R.H., Roop, D.R., Trus, B.L. and Steven, A.C. (1983a) Nature *(London)*, **302**, 794-800
Complete amino acid sequence of a mouse epidermal keratin subunit and implications for the structure of intermediate filaments

Steinert, P.M., Steven, A.C. and Roop, D.R. (1983b) J. Invest. Dermatology, **81**, 1-14
Structural features of epidermal keratin filaments reassembled in vitro

Steinert, P.M., Parry, D.A.D., Racoosin, E.L., Idler, W.W., Steven, A.C., Trus, B.L. and Roop, D.R. (1984a) Proc. Natl. Acad. Sci. USA, **81**, 5709-5713
The complete cDNA and deduced amino acid sequence of a type II mouse epidermal keratin of 60,000 Da: Analysis of sequence differences between type I and type II keratins

Steinert, P.M., Jones, J.C.R. and Goldman, R.D. (1984b) J. Cell Biol., **99**, 22s-27s
Intermediate filaments

Steinert, P.M., Parry, D.A.D., Idler, W.W., Johnson, L.D., Steven, A.C. and Roop, D.R. (1985a) J. Biol. Chem., **260**, 7142-7149
Amino acid sequences of mouse and human epidermal type II keratins of Mr 67,000 provide a systematic basis for the structural and functional diversity of the end domains of keratin intermediate filament subunits

Steinert, P.M., Steven, A.C. and Roop, D.R. (1985b) Cell, **42**, 411-419
The molecular biology of intermediate filaments

Steinert, P.M., Idler, W.W., Zhou, X.-M., Johnson, L.D., Parry, D.A.D., Steven, A.C. and Roop, D.R. (1985c) Ann. N. Y. Acad. Sci., **455**, 451-461
Structural and functional implications of amino acid sequences of keratin intermediate filament subunits

Steinert, P.M., Torchia, D.A. and Mack, J.W. (1989) in "The Biology of Wool and Hair", eds G.E. Rogers, P.J. Reis, K.A. Ward and R.C. Marshall, Chapman and Hall, London and New York, 157-167
Structural features of keratin intermediate filaments

Steven, A.C. (1989) in "Cellular and Molecular Biology of Intermediate Filaments", eds R.D. Goldman and P.M. Steinert, Plenum Press (in press)
Intermediate filament structure: Diversity, polymorphism, and analogy to myosin

Steven, A.C., Wall, J., Hainfeld, J. and Steinert, P.M. (1982) Proc. Natl. Acad. Sci. USA, **79**, 3101-3105
Structure of fibroblastic intermediate filaments: Analysis by scanning transmission electron microscopy

Steven, A.C., Hainfeld, J.F., Trus, B.L., Wall, J.S. and Steinert, P.M. (1983a) J. Biol. Chem., **258**, 8323-8329
The distribution of mass in heteropolymer intermediate filaments assembled in vitro

Steven, A.C., Hainfeld, J.F., Trus, B.L., Wall, J.S. and Steinert, P.M. (1983b) J. Cell Biol., **97**, 1939-1944
Epidermal keratin filaments assembled in vitro *have masses-per-unit-length that scale according to average subunit mass: Structural basis for homologous packing of subunits in intermediate filaments*

Steven, A.C., Trus, B.L., Hainfeld, J.F. Wall, J.S. and Steinert, P.M. (1985) Ann. N.Y. Acad. Sci., **455**, 371-380
Conformity and diversity in the structures of intermediate filaments

Steven, A.C. Trus, B.L., Maizel, J.V., Unser, M., Parry, D.A.D., Wall, J.S., Hainfeld, J.F. and Studier, F.W. (1988) J. Mol. Biol., **200**, 351-365
Molecular substructure of a viral receptor-recognition protein. The GP17 tail fibre of bacteriophage T7

Steven, A.C., Mack, J.W., Trus, B.L., Bisher, M.E. and Steinert, P.M. (1989) in "Cytoskeletal and Extracellular Proteins: Structure, Interactions and Assembly", eds U. Aebi and J. Engel, Springer Series in Biophysics, Springer-Verlag, Berlin, **3**, 15-26
Structure and assembly of intermediate filaments: Multi-faceted, myosin-like (but non-motile) cytoskeletal polymers

Stewart, M., Quinlan, R.A., Moir, R.D., Clarke, S.R. and Atkinson, S.J. (1989a) in "Cytoskeletal and Extracellular Proteins: Structure, Interactions and Assembly", eds U. Aebi and J. Engel, Springer Series in Biophysics, Springer-Verlag, Berlin, **3**, 150-159
The role of repeating sequence motifs in interactions between α-helical coiled-coils such as myosin, tropomyosin and intermediate-filament proteins

Stewart, M., Quinlan, R.A. and Moir, R.D. (1989b) J. Cell Biol., **109**, 225-234
Molecular interactions in paracrystals of a fragment corresponding to the α-helical coiled-coil rod portion of glial fibrillary acidic protein: Evidence for an antiparallel packing of the molecules and polymorphism related to intermediate structure

Stick, R., Angres, B., Lehner, C.F. and Nigg, E.A. (1988) J. Cell Biol., **107**, 397-406
The fates of chicken nuclear lamin proteins during mitosis: Evidence for a reversible redistribution of lamin B_2 protein between inner nuclear membrane and elements of the endoplasmic reticulum

Stone, D., Sodek, J., Johnson, P. and Smillie, L.B. (1974) Proc. 9th FEBS Meeting (Budapest), **31**, 125-136
Tropomyosin: Correlation of amino acid sequence and structure

Tainer, J.A., Getzoff, E.D., Alexander, H., Houghten, R.A., Olson, A.J., Lerner, R.A. and Hendrickson, W.A. (1984) Nature *(London)*, **312**, 127-134
The reactivity of anti-peptide antibodies is a function of the atomic mobility of sites in a protein

Thompson, M.A. and Ziff, E.B. (1989) Neuron, **2**, 1043-1053
Structure of the gene encoding peripherin, an NGF-regulated neuronal-specific type III intermediate filament protein

Torchia, D.A. and van der Hart, D.L. (1976) J. Mol. Biol., **104**, 315-321
Carbon-13 magnetic resonance evidence for anisotropic molecular motion in collagen fibrils

Traub, P. and Vorgias, C.E. (1983) J. Cell Sci., **63**, 43-67
Involvement of the N-terminal polypeptide of vimentin in the formation of intermediate filaments

Trachtenberg, S. (1987) Biochem. Biophys. Acta, **923**, 327-332
Assemblies of psoriatic keratin and their relation to normal intermediate filament structure

Trachtenberg, S., Steinert, P.M., Trus, B.L. and Steven, A.C. (1986) in "Electron Microscopy and Alzheimer's Disease", ed J. Metuzals, San Francisco Press, 56-59
Electron microscopy and image analysis of paracrystalline filament bundles obtained from psoriatic epidermal keratin

Trinick, J., Cooper, J., Seymour, J. and Egelman, E.H. (1986) J. Microsc. (Oxford), **141**, 349-360
Cryo-electron microscopy and three-dimensional reconstruction of actin filaments

Troncoso, J.C., Häner, M., March, J.L., Reichelt, R., Engle, A. and Aebi, U. (1989) in "Cytoskeletal and Extracellular Proteins: Structure, Interactions and Assembly", eds U. Aebi and J. Engel, Springer Series in Biophysics, Springer-Verlag, Berlin, **3**, 33-38
Structure and assembly of neurofilaments (NF): Analysis of native NF and of specific NF subunit combinations

Tyner, A.L. and Fuchs, E. (1986) J. Cell Biol., **103**, 1945-1955
Evidence for posttranscriptional regulation of the keratins expressed during hyperproliferation and malignant transformation in human epidermis

Tyner, A.L., Eichman, M.J. and Fuchs, E. (1985) Proc. Natl. Acad. Sci. USA, **82**, 4683-4687
The sequence of a type II keratin gene expressed in human skin: Conservation of structure among all intermediate filament genes

Venetianer, A., Schiller, D.L., Magin, T. and Franke W.W. (1983) Nature *(London)*, **305**, 730-733
Cessation of cytokeratin expression in a rat hepatoma cell line lacking differentiated functions

Weber, K. (1986) Nature *(London)*, **320**, 402
Link between lamins and intermediate filaments

Weber, K. and Geisler, N. (1982) EMBO J., **1**, 1155-1160
The structural relation between intermediate filament proteins in living cells and the α-keratins of sheep wool

Weber, K. and Geisler, N. (1984) in "Cancer Cells 1, The Transformed Phenotype", eds A.J. Levine, G.F. Vande Woude, W.C. Topp and J.D. Watson, Cold Spring Harbour, New York, **1**, 153-159
Intermediate filaments – from wool α-keratins to neurofilaments: A structural overview

Weber, K. Plessmann, U., Dodemont, H. and Kossmagk-Stephan, K. (1988) EMBO J., **7**, 2995-3001
Amino acid sequences and homopolymer-forming ability of the intermediate filament proteins from an invertebrate epithelium

Westhof, E., Altschuh, D., Moras, D., Bloomer, A.C., Mondragon, A., Klug, A. and Van Regenmortel, M.H.V. (1984) Nature *(London)*, **311**, 123-126
Correlation between segmental mobility and the location of antigenic determinants in proteins

Winkles, J.A., Sargent, T.D., Parry, D.A.D., Jonas, E. and Dawid, I.B. (1985) Mol. Cell. Biol., **5**, 2575-2581
Developmentally regulated cytokeratin gene in Xenopus laevis

Wolin, S.L., Krohne, G. and Kirschner, M.W. (1987) EMBO J., **6**, 3809-3818
A new lamin in Xenopus *somatic tissue displays strong homology to human lamin A*

Woods, E.F. (1983) Biochem. Int., **7**, 769-774
Structural studies on the microfibrillar proteins of wool: Characterization of the α-helix-rich particle produced by chymotryptic digestion

Woods, E.F. and Gruen, L.C. (1981) Aust. J. Biol. Sci., **34**, 515-526
The number of polypeptide chains in the rod domain of bovine epidermal keratin

Woods, E.F. and Inglis, A.S. (1984) Int. J. Biol. Macromol., **6**, 277-283
Organization of the coiled-coils in the wool microfibril

Worman, H.J., Yuan, J., Blobel. G. and Georgatos, S.D. (1988) Proc. Natl. Acad. Sci. USA, **85**, 8531-8534
A lamin B receptor in the nuclear envelope

Zackroff, R.V., Goldman, A.E., Jones, J.C.R., Steinert, P.M. and Goldman, R.D. (1984) J. Cell Biol., **98**, 1231-1237
Isolation and characterization of keratin-like proteins from cultured cells with fibroblastic morphology

Zehner, Z.E. and Paterson, B.M. (1985) Ann. N.Y. Acad. Sci., **455**, 79-94
The chicken vimentin gene: Aspects of organization and transcription during myogenesis

Zhou, X.-M., Idler, W.W., Steven, A.C., Roop, D.R. and Steinert, P.M. (1988) J. Biol. Chem., **263**, 15584-15589
The complete sequence of the human intermediate filament chain keratin 10: Subdomainal divisions and model for folding of end domain sequences

PUBLICATIONS

The research detailed in the following publications forms much of the basis for this thesis. Papers included (in chronological order) are:

1) Parry, D.A.D., Conway, J.F. and Steinert, P.M. (1986) Biochem. J., **238**, 305-308
Structural studies on lamin: Similarities and differences between lamin and intermediate-filament proteins

2) Parry, D.A.D., Conway, J.F., Goldman, A.E., Goldman, R.D. and Steinert, P.M. (1987a) Int. J. Biol. Macromol., **9**, 137-145
Nuclear lamin proteins: Common structures for paracrystalline, filamentous and lattice forms

3) Conway, J.F. and Parry, D.A.D. (1988) Int. J. Biol. Macromol., **10**, 79-98
Intermediate filament structure: 3. Analysis of sequence homologies

4) Conway, J.F. and Parry, D.A.D. (1989) in "Cytoskeletal and Extracellular Proteins: Structure, Interactions and Assembly", eds U. Aebi and J. Engel, Springer Series in Biophysics, Springer-Verlag, Berlin, **3**, 140-149
Structural and spatial organization of intermediate filament and nuclear lamin proteins

5) Conway, J.F., Fraser, R.D.B., MacRae, T.P. and Parry, D.A.D. (1989) in "The Biology of Wool and Hair", eds G.E. Rogers, P.J. Reis, K.A. Ward and R.C. Marshall, Chapman and Hall, London and New York, 127-144
Protein chains in wool and epidermal keratin IF: Structural features and spatial arrangement