KU Leuven
Biomedical Sciences Group
Department of Pharmaceutical
and Pharmacological Sciences
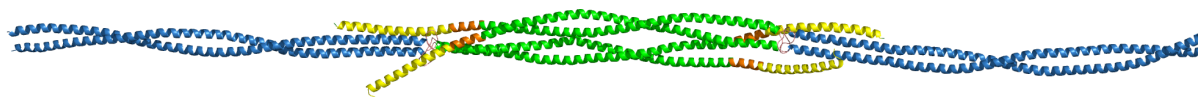Laboratory for Biocrystallography

CRYSTALLOGRAPHIC STUDIES
OF VIMENTIN AND LAMIN A

Anastasia CHERNYATINA

Promoter: Prof. Dr. Sergei V. Strelkov
Jury members:
Prof. Dr. Ann Gils (Chair)
Prof. Dr. Harald Herrmann-Lerdon
Prof. Dr. Kristof van Hecke
Prof. Dr. Eveline Lescrinier
Prof. Dr. Chris Ulens

Dissertation presented in partial fulfilment

of the requirements for the degree of

Doctor of Pharmaceutical Sciences

November 2013

An X-ray beam - the brightest light,
Born in the dance of whirling grains,
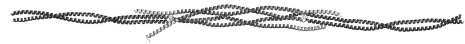At its full speed it rushes
Against the rock of crystal smashes.

*

Lucky witnesses of the crush,
Happy we are to collect the flash.
Gathering splinters of this collapse
Leads us to electron density maps.

*

Inside the rock in silent cores
A shining fortune is concealed,
The turns of life:
          a strand of mine,
                    a strand of yours,
The coiled coil is revealed.

November, 2013

Acknowledgements

Probably, the biggest crystal I have grown in Belgium is a crystal of friendship. Here in Leuven, I had many wonderful moments shared with my friends: Tatiana Afrikanova, Adrien Combaz, Jaime Undurraga, Veronica Méndez Muñoz, Pierpaolo Spinelli, Ivan Lavenia, Dima Kaliukhovich, Masha and Nikolay Chumerin, Natasha Dyubankova, Kolya Manyakov, Max Mayorov, Julia Gorkina, Arturo Cn, Alina Remeeva, Kubik and Sasha Zharikova, Denitza Denkova, Stefan Antonov, Daniel Szőcs, Matias Timmermans, Joffre Gutierrez Royo, Manisha Bisht, Hiwa Modarresi, Richard Stahl, Aneta Markova, Roland Pastorino, Carolina Fernandez Ramon and especially Vera Lazenka. The distance between Belgium and my home did not seem big thanks to my friends from Russia: Dina Gainullina, Anastasia Shebanova, Anton Kuzmin, Nikolay Murzov, Dima Morozov, Anton Usachev, Danya Ilchenko, Zoya Romanenko, Masha Kolotvina, Pavel Gavrilushkin, Jenya Averkina, Ilya Murashkin, Kostya Baranov, Jenya Shachnev, Anya Kuroschenkova, Katya Shmulevich, Lena Vorozhtsova and departed this life but alive in my heart Shurik Kishkinskiy. Thank you, my friends for all your support and frankness, adventures, trips, evenings - the life we had together. Thank you for many years of your wonderful and so important for me friendship.

My special thanks go to Mia Marshall for her hospitality and kindness. Also I would like to acknowledge my favourite language teachers: Bérénice Castillo, Marijke Erauw and Jackie Clare for their interesting, joyful and useful classes; as well as dance classes of jazz-modern by Jac Delsing and salsa by Carl Hourcau, which kept my muse awake.

In the end I would like to warmly thank my family: my aunties, uncles, cousins, nephews, nieces, my brother and, of course, my parents, Nina and Aleksandr. My dear, thank you for your support, for teaching me, for being proud of me and believing in my success, for your advice when it was needed, your company in many-many trips we did together through these years and for your big parental love.

Мои дорогие родители, в этой части своего тезиса я хочу сказать вам обоим большое спасибо за ваше тепло и поддержку, за то, что гордитесь мной и переживаете за меня, за вашу веру в мой успех, вашу компанию во многих увлекательных путешествиях и, конечно, за вашу большую родительскую любовь.

Table of Contents

List of abbreviations

Å	Ångström
ADC	autosomal dominant cataract
ALS	amyotrophic lateral sclerosis
AxD	Alexander disease
C. elegans	*Caenorhabditis elegans*
CC	coiled coil
CCD	charge-coupled device
CCP4	Collaborative Computational Project number 4
CM	carboxymethyl (ion-exchange group)
DEAE	diethylaminoethyl (ion-exchange group)
DTT	dithiothreitol
E. coli	*Escherichia coli*
EBS	Epidermolysis bullosa simplex
EDTA	ethylene diamine tetraacetic acid
EGTA	ethylene glycol tetraacetic acid
EM	electron microscopy
EPMR/EPR	electron paramagnetic resonance
ESI-MS	electrospray ionization mass spectrometry
GABA	gamma-aminobutyric acid
GFAP	glial fibrillary acidic protein
HA	heavy atom
IBD	inflammatory bowel disease
IF	intermediate filament
Ig	immunoglobulin
IPTG	isopropyl β-D-1-thiogalactopyranoside
K / KIF	keratin / keratin intermediate filament
MAD	multiple -wavelength anomalous dispersion
MES	2-(N-morpholino)ethanesulfonic acid
MF	microfilament
MIR	multiple -wavelength isomorphous replacement
MPD	2-methyl-2,4-pentanediol
MR	molecular replacement
MT	microtubule
NF-L/M/H	neurofilament – light/middle/heavy chain
NLS	nuclear localization site
PDB (ID)	Protein Data Bank (Identifier)
PEG	polyethylene glycol
SAD	single -wavelength anomalous dispersion
SAXS	small angle X-ray scattering
SDSL	site-directed spin labelling
SDS-PAGE	sodium dodecyl sulphatepolyacrylamide gel electrophoresis
SeMet	selenomethionine
SIR	single-wavelength isomorphous replacement
SLS	Swiss Light Source
Tris	tris(hydroxymethyl)aminomethane
ULF	unit-length filament
WT	wild-type

Chapter 1. Introduction

1.1 X-ray crystallography and coiled coils

Diffraction of X-rays scattered by electrons applied to protein crystals allows one to get an insight into the atomic resolution structure of these biomolecules. This is the most widespread and successfully used technique for structure investigation as seen from the number of determined structures to date (Protein Data Bank (PDB), www.rcsb.org). Protein structure in atomic details is indispensable for a detailed understanding of an organization and mechanism of action of proteins as well as of causes of many human diseases.

Interestingly, the first molecules studied structurally with the help of X-rays were keratins and DNA, both containing helical structures. The properties of keratin and collagen fibres were actively studied in the 20th century alongside with the development of textile industry. Britain was especially interested in these studies, as it was one of the main producers of wool since the Industrial Revolution in the middle of the 18th century. It was no surprise that it was an English biophysicist and lecturer in the Textile Physics at the University of Leeds, William Astbury, who discovered α-helical repeat in unstretched wool fibres in early 1930s. He also demonstrated using an X-ray diffraction on stretched wool fibres that the helical structure uncoils, forming an extended state [1]. This state was called by Astbury a β-form [1]. Later, in 1951, this terminology was kept when two main protein secondary structure motifs, α-helices and β-sheets, were described by Pauling *et al.* [2, 3]. A helix with 3.7-residues per turn was defined as the main fold of α-keratin, α-myosin, and similar fibrous proteins [2]. It took one more year for Astbury's student Francis Crick to describe a specific complex of several intertwined α-helices. These structures were called coiled coils (CC) [4, 5]. Thus, the rise of the scientific interest to the CC proteins fell on the rise of the use of X-rays for protein structure determination.

An α-helical CC is a wide-spread structural motif found in 1.5-3.3 % of all proteins [6, 7]. A typical CC is formed by two α-helices twisted around each other in order to shield their hydrophobic residues (Fig.2, B-D). This hydrophobic core defines an architecture of the CC. Due to an elongated structure, CCs often serve as 'cables' for transmitting signals or organising filamentous network of the cell. For example, CCs are the main structural components of transcription factor proteins forming Leu-zippers - GCN4, Fos/Jun [8, 9], in membrane proteins such as G-protein coupled receptors (e.g., rhodopsin), cytoskeletal intermediate filament (IF) proteins [10] and many others [11].

In this work we focus on IF proteins, which among all cytoskeletal proteins are the least understood. Their complete structures remain unresolved crystallographically to date.

1.2 Historical overview of intermediate filament proteins

IFs together with microtubules (MTs) and microfilaments (MFs) form three filamentous systems of the cytoskeleton of metazoan cells (Fig. 1) [10, 12]. The MTs (~25 nm in diameter) and MFs (~7 nm in diameter) were studied in atomic detail [13, 14]. They consist of globular proteins (tubulins and actin respectively) limited in number and diversity among different species. In contrast to MTs and MFs, IFs correspond to a big heterogeneous system consisting of fibrous proteins, structural information of which is obscure.
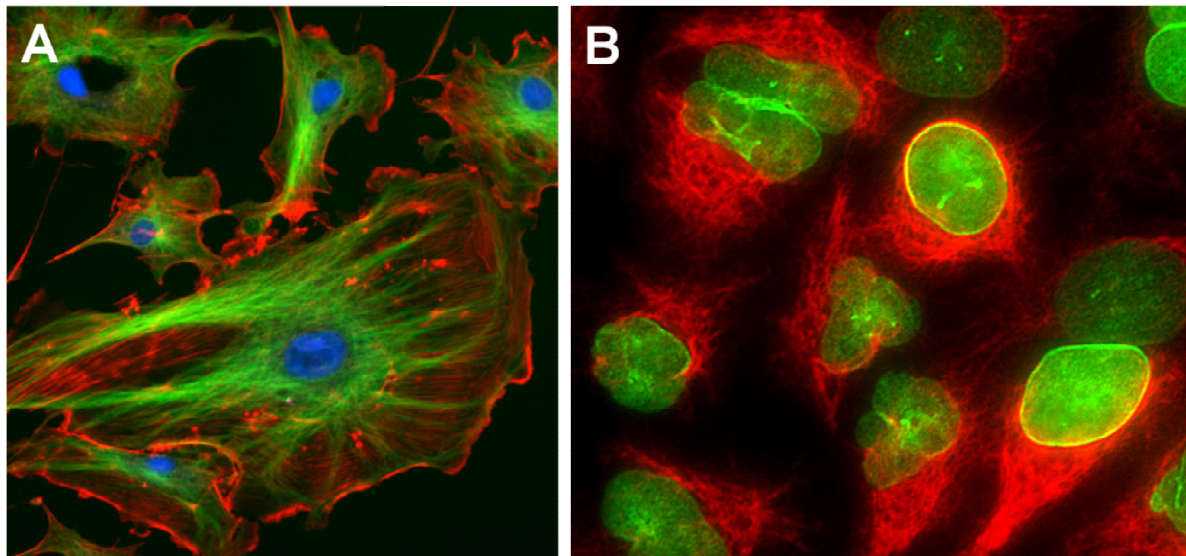
*Fig.1. Immunofluorescence microscopy with immunostained DNA, microtubules, actin filaments, lamin A and keratin 18. **A.** Endothelial cells from bovine pulmonary artery. Blue staining (4,6-diamidino-2-phenylindole) indicates DNA in nuclei; microtubules are marked green (an antibody bound to fluorescein isothiocyanate) and actin filaments are labelled red (phalloidin bound to* tetramethylrhodamine *isothiocyanate). Copied from www.wikipedia.org (public domain). **B.** Immunofluorescence microscopy using Thermo Scientific dyes: cytokeratin 18 labelled red (DyLight 594) and lamin A labelled green (DyLight 594) in adenocarcinomic human alveolar basal epithelial cells A549. Copied from www.piercenet.com with permission of Thermo Fisher Scientific Inc.*

IFs found in muscle cells (cultured from chick embryos) were first characterized as a separate class in 1968 by Ishikawa and colleagues [15]. Their research showed that the diameter of IFs is about 10 nm, intermediate between MTs and MFs. IFs were assigned to a new class of cytoskeletal proteins [15]. After this discovery, the class of IF proteins was expanding: new IFs were found in many different tissues in both, cyto- and nucleoplasm of the cells. Purification and *in vitro* assembly of intact IFs became a big step towards better understanding of IF architecture [16-20]. It was noted early on that purified in denaturing conditions (using 8 M urea) keratins from bovine epidermis self-assemble into ~ 8 nm wide filaments under low-salt conditions: 5 mM Tris-HCl (in the range of pH from 7 to 8) [16, 21]. These filaments could be found in the pellet after centrifugation for 1 h at 250000*g. However

this pellet could again be resuspended in the presence of denaturing agents such as 8M urea, 6M guanidine or 0.1% SDS [16, 21]. The most interesting discovery was that these proteins are capable of refolding into filaments again when dialysed against low-salt buffers without denaturing agents [21]. A mechanism of the filament assembly was completely unknown. At that time Steinert and colleagues described seven different species of keratins extracted from bovine epidermis [16]. They showed that none of these proteins formed filaments when isolated alone [16]. Different pair-wise combinations of keratins necessary for the filament formation were found and appeared to have a molar ratio of 1:2 in each case. Hence, the building block of keratins was initially erroneously considered to be a trimer [16]. However, subsequent experiments showed that the building block of all IFs is a dimer [22-24]: the initial assumption concerning the trimerization of keratins derived from the fact that keratin filaments re-assembled *in vitro* contain both hetero- and homo-dimers [22].

An estimation of the α-helical content of different regions of IF proteins (using optical rotary dispersion and circular dichroism) revealed similarities between desmin and vimentin extracted from baby hamster kidney as well as between vimentin and two keratins from bovine epidermis [25]. Afterwards, sequences of the other IF proteins became available. A comparison of the sequences of human epidermal keratin, two sheep α-keratins, chicken muscle desmin and parts of the sequences of porcine vimentin, neurofilament and bovine glial fibrillary acidic protein (GFAP) that were available, revealed tripartite structure characteristic of all IF proteins. They all contained presumably α-helical rod domain and non-α-helical head and tail domains (Fig. 2, A) [24, 26]. Moreover, Geisler and Weber analysed the amino-acid distribution in the rod domains and showed high amino-acid sequence similarities between different IF proteins [24]. For keratins and desmin a homology of 66-69% within the regions corresponding to the beginning of the coil1 and the end of the coil2, respectively, was observed. In addition about 75% identity along the whole coil2 (and 86% homology between the second halves of coil2 regions) of vimentin and desmin proteins was demonstrated [24]. Geisler and Weber used Crick's finding that seven residue repeat indicates the possibility of CC formation [5] (Fig. 2 B, C). They demonstrated characteristic for CC distribution of apolar amino acids within the rod domains of studied IF proteins [24]. In addition, it was found that these proteins have a specific distribution of charged amino acids on the surface of the CC [24, 27] (Fig. 2, B).
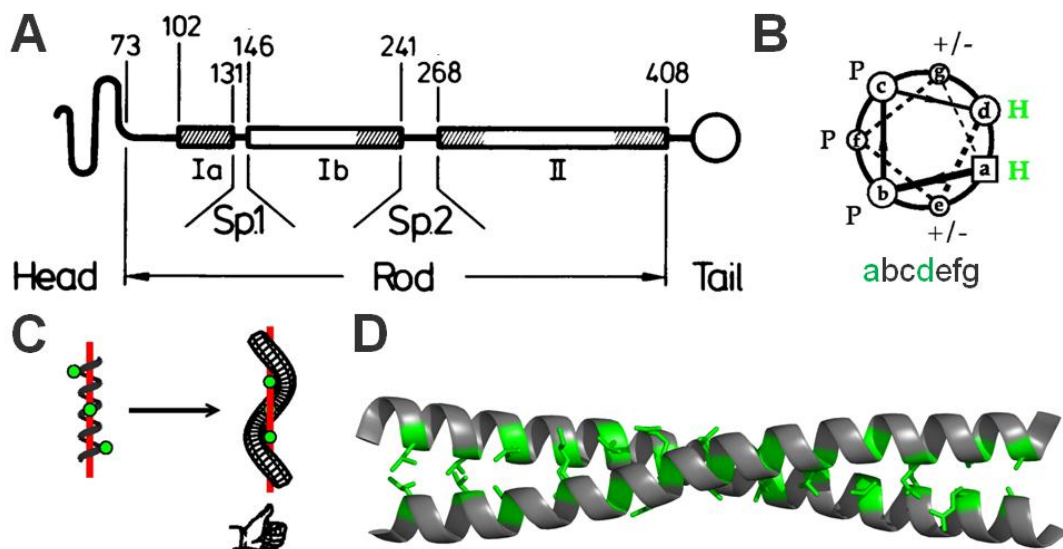
*Fig.2. Tripartite structure of IF proteins: flexible head and tail domains flanking an α-helical rod domain, leading to formation of a CC dimer. **A**. Sequencing analysis of chicken desmin resulted in the tripartite model, suggested by Geiser et al.[24] (figure is copied from [24]). Regions, shown in rectangles are α-helical, Ia, Ib and II correspond to coil1A, coil1B and coil2, respectively. Hatched areas indicate highly conserved regions among chicken gizzard desmin, human epidermal keratin 14, sheep wool keratins 8c-1 and 7c, porcine vimentin and NF-L, as well as bovine GFAP. Sp1 and Sp2 stand for non-α-helical spacers, linkers L1 and L12. **B**. A scheme of the cross-section of an α-helix with heptad distribution of hydrophobic amino acids with repeating pattern of positions abcdefg, where a and d are occupied by hydrophobic (H), b, c, f – by polar (P) and e, g - by charged (+/-) amino acids. **C.** positions a and d form a hydrophobic thread on the surface of an α-helix, leading to the formation of the classical left-handed coiled-coil, shown in **D**. Figures B and C are reproduced from [28], with permission of The American Association for the Advancement of Science, licence agreement number 3259411262768.*

To shed light on the whole class of IF proteins, it was necessary not only to know the amino acid sequences, but also to characterise their distribution in the organism (tissue specificity, localization in the cell) as well as their functions and assembly process. In the late 1970's only a small part of all IF proteins was described and no full sequence of any IF protein was known. However, different IF proteins could be already distinguished with the help of immunofluorescence microscopy. Using antibodies to vimentin, keratin, desmin and neurofilaments Franke *et al.* described the tissues where each IF protein was found [29]. This led to the knowledge of the unique pattern of the localization in tissues for each of them, and to the initial outdated classification of all IF proteins [29]. According to this classification an IF protein found in mesenchyme-derived cells (*e.g.* endothelium, fibroblasts, blood vessels, connective tissues) and in some brain- tumour derived cells was ascribed to the first type and named vimentin. This name came from Latin *vimentum* - flexible branches of willow trees used for wicker-work [29]. The second type consisted of IF proteins found in epithelia-derived cells (prekeratin). The third type included IF proteins characteristic of muscle cells (desmin). Neurofilament proteins were assigned to the fourth type [29]. Later glial fibrillary acidic protein (GFAP), detected in glial cells and cells of glial origin, was described as one

more type [30]. Interestingly, at that time it was believed that biochemical and immunological taxonomety of IF proteins was close to complete [31], though discoveries of new IF proteins are still ongoing even nowadays.

As soon as the sequences of most cytoplasmic IF proteins became available, their classification was revised. The main criteria to distinguish different types of IF proteins became the similarity between central rod domains sequences [32]. Thus, acidic and basic keratins were classified as type 1 and 2 IF proteins. Keratin IFs are obligate heteropolymers each consisting of both types [20, 33], while most of other IF proteins form homodimers. Vimentin, desmin and GFAP were observed to have a high similarity of the rod domain sequence and therefore were combined into the type 3 of IFs. Neurofilaments showing rather different rod domain sequence were classified into type 4 [32].

In parallel with cytoplasmic, another class of IF proteins was found in the nuclei. Three nuclear IF proteins (named later lamin A, B and C [34]) were characterised immunocytochemically using specific antibodies [35]. Proteins with masses 60, 67 and 70 kDa were found specifically in the periphery of the interphase nucleus [35] and shown later to be also present in non-membrane associated nucleoplasm during mitosis [34]. The fact that lamina dissociates from the membrane during mitosis was explained by depolymerization of its fibrous proteins [34]. A few years later the nucleotide sequence of the coding region of lamin A/C showed a high homology between the α-helical rod domains with the cytoplasmic IF proteins [36, 37]. Lamins were placed in the IF family also due to the organization of the native nuclear lamina as well as the structure and assembly properties of purified lamins [38]. In a commonly used system they are classified as a separate fifth class of IF proteins [39] (see Table 1).

Later two more IF proteins were found uniquely in the eye lens, named filensin (formerly called CP94, CP95, or CPII5) and phakinin (formerly called CP49 or CP47) and classified as the sixth class of IF proteins [40]. The sequencing of the full human genome and the genomes of new species is a reason for discoveries of new IFs. The new IF protein crescentin (from bacteria *Caulobacter crescentus* [41, 42]), the IF-like protein isomin (from an arthropod *Isotomurus maculates* [43]) as well as the new types of keratins [44, 45] have been discovered recently.

Due to many parameters (tissue specificity, cell specificity or sequence similarities) the classification of IF is rather complex. For example, different IFs may co-express in one cell type: vimentin and desmin in baby hamster kidneys [17], keratin and vimentin in carcinoma cells [46]. In addition different IFs may express in the same cells at different stages of tissue development. So, ectodermal cells expressing keratins differentiate into vimentin-expressing mesenchymal cells of neural tube [47]; lamin A is only expressed during gastrulation and later is substituted by lamin B [48]). Here we use the system where more

than 70 members of IF proteins in higher metazoa are classified into 6 classes (Table 1). This is an extended classification based on [39].

Table 1. Classification of IF proteins.

Type	Protein	№ of genes	MW	Cell type	Disease
Type I	Acidic keratins	28 (17 epithelial and 11 hair KIF)	42-64 kDa	Epithelia	EBS(K5/K14) Keratoderma disorders Hair and nail defects (for details see Fig. 3)
Type II	Basic keratins	26 (20 epithelial and 6 hair KIF)	50-66 kDa		
Type III	Vimentin Desmin GFAP Peripherin	1 1 1 1	54 kDa 53 kDa 50 kDa 54 kDa	Mesenchyme Muscle Glia, astrocytes PNS neurons	vimentin-related ADC Desmin-related myopathy Alexander disease ALS (for details see Table 2)
Type IV	NF-L NF-M NF-H α-internexin	1 1 1 1	62 kDa 102 kDa 110 kDa 55 kDa	CNS neurons	Neurodegenerative disorders Charcot-Marie-Tooth disease Parkinson
Type V	Lamins A / C Lamins B_1 Lamins B_2	1 1 1	70/63 kDa 67 kDa 72 kDa	Mature cells Developing cells Germ-line cells	Laminopathies (for details see Table 3)
Type VI	Phakinin Filensin	1 1	49 kDa 95 kDa	Lens	ADC (violated lens transparency)

Due to their inherent mechanical and (de)polymerization properties, IFs are involved in shaping of the cell organelles, positioning and structural integrity of cells and tissues [49, 50]. Therefore, IFs were thought to be purely 'mechanical integrators of cellular space' [31, 50]. Subsequently, other functions of IFs were discovered: interaction with other proteins, DNA and organelles, as well as up- and down-regulation of many cell processes (for overview see [50]). IFs are involved in cell adhesion, migration and signalling pathways [51]. Recent studies demonstrate that IFs are stress proteins as they are highly dynamic and in most cases are regulated under stress conditions. Characteristically for stress proteins they inhibit apoptosis, interact with many heat shock proteins and are major substrates for stress kinases [52].

All functions of IF proteins depend on the organelle, cell and tissue type they are expressed in. The functions may vary during the cell cycle. Phosphorylation is the main post-translational modification used by cells to alter subcellular organization and dynamics of the IF network [53].

1.3 IF family – classification, functions and related diseases

1.3.1 Type 1 and 2

Type 1 and 2 IF proteins consist of acidic and basic keratins, respectively. Keratins are major IF proteins uniquely expressed in epithelial cells of vertebrates [54]. They are ubiquitous in the skin of mammals, forming ~30% of the total protein of basal and ~85% of the total protein of outermost layer of the epidermis [54]. All keratins form hetero-dimers composed of acidic (pKa 4.5-5.5) and basic (pKa 6.5-7.5) keratins, different in their masses and isoelectric points [55].

Keratins play a big role in human health as well as in our everyday life. For example, horns, leather, fur, feathers are used in clothing, weaving, knitting, in bedding materials, interior decorations and medicine. Consequently, health of keratinized and cornified tissues of the livestock is one of the primary goals of veterinary medicine. For human population a well-attended look and health of skin/hair/nail has an extremely important role. Keratins are a target for many investigations in the field of cosmetics and dermatology as well as in treatment of keratin-associated diseases [56]. Up to date 54 keratin genes were described in human [57] (see Table 1). The expression of keratins in different epithelial cells at different stages of cellular differentiation depends on the cell type (see Fig. 3) [55].

In cells keratins play different roles. The main role of keratins is mechanical: maintaining the stability and integrity of epithelial cells and tissue resilience [58]. In addition, keratins have many other functions: they are involved in signalling pathways, activated under mechanical stress such as wound healing [59] as well as in apoptosis and cornification - a specific type of the cell death [60]. Due to a very specific pattern of expression, keratins serve as markers of human tumours of different origin and widely used in immunohistochemical diagnosis of carcinomas [57]. Mutations in keratins are linked to a big number of disorders. In the human IF protein database [61] the diseases caused by mutations in keratins are subdivided onto five types.

1. Epidermolysis bullosa simplex (EBS) is the first disease reported to be linked to a point mutation (R125C) in keratin 14 [62]. EBS is characterised by blistering upon mechanical trauma due to the keratin aggregation in keratinocytes of *stratum basale* [62]. After this discovery many other mutations in the keratins 14 and 5 were shown to be related to EBS. About 1 out of 20,000 people suffer from some form of EBS and in most cases it is an autosomal dominant, inherited disease [61].

2. Keratoderma disorders are linked to autosomal dominant mutations in different keratins. The most widespread keratodermas are related to mutations in K1 or K10. These are the two major keratins expressed in keratinocytes of s*tratum spinosum* and other body sites, where cornifying stratified squamous epithelium is present – hair, nails, epidermis of palm and sole. The other types of keratoderma are associated with K2, K6 isomer (K6c), K9,

K10, K14 and K16 abnormalities (for details see [61]). The symptoms of keratodermas are similar to EBS, but the histological analysis reveals different affected tissues.

3. Hair and nail defects are characterised by hypertrophic nail dystrophy related to mutations in K6a/K16 or K6b/K17. These keratins are normally expressed during wound healing or as a response to stress. However, if the latter are overexpressed in normal tissues, it leads to painful thickened finger and toe nails, different severity focal keratodermas and pelosebaceous (in the hair follicles and sebaceous glands) cysts. In addition, many mutations in trichocyte keratins (K74, K81, K83, K85 and K86) lead to hair defects/loss and nail abnormalities [61].

4. Keratin disorders affecting other tissues are related to different keratins. Oral, anal and genital epithelia can be affected when s*tratum spinosum* keratins K4 and K13 are mutated. Mutations K3 or K12 in *stratum corneum* are related to occurrence of corneal microcysts causing erosions and photosensitivity (for details see [61]).

5. Simple epithelial keratin mutations cause widely spread inflammatory bowel diseases (IBDs). IBDs are a heterogeneous group of diseases, including Crohn disease and ulcerative colitis. According to available statistics, IBDs affect intestinal epithelium of ~1/400 Americans and as much as ~1/20 of Scottish population. Despite the fact that the mechanism of IBD development is not understood there are links made to mutations in K8/K18 [63]. In addition, many liver disorders are caused by mutations in K8/K18 [64].
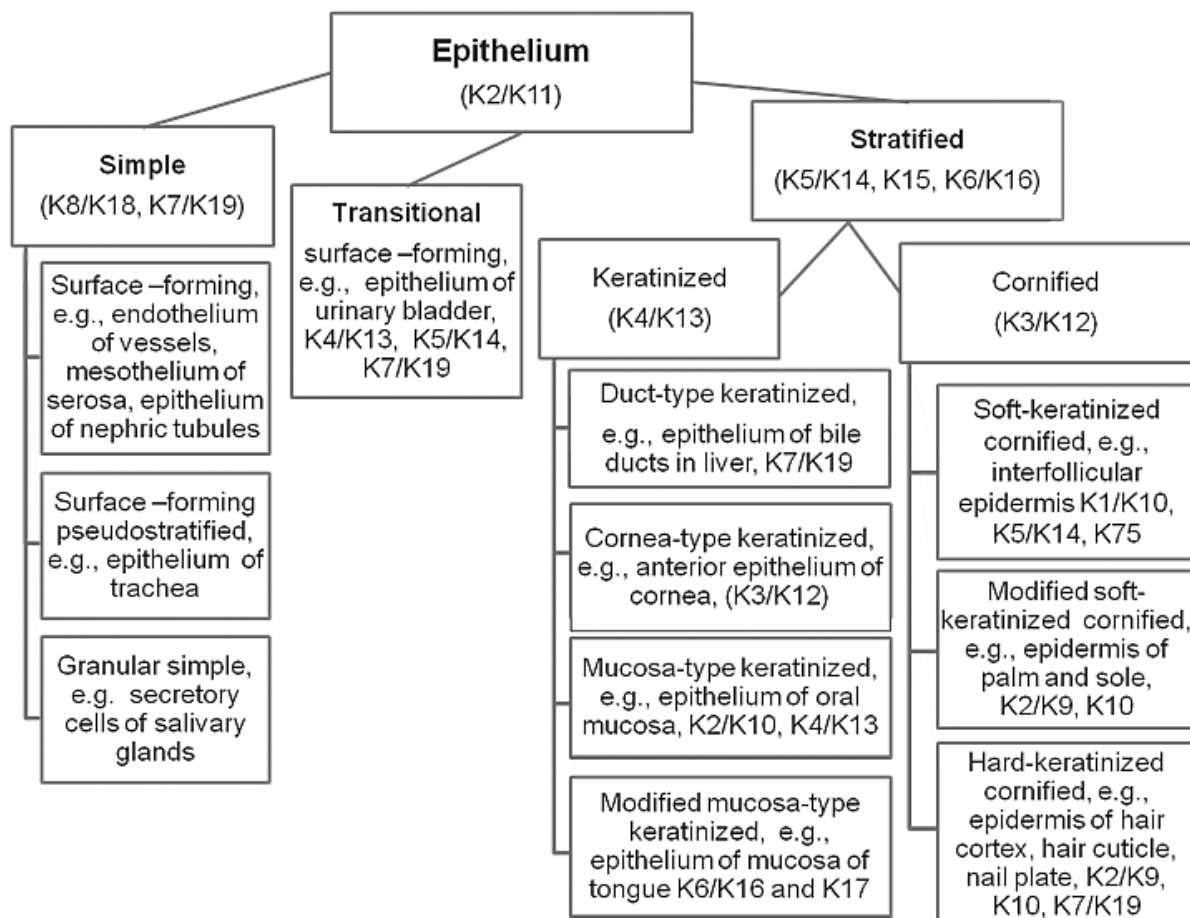
Fig. 3. Types of epithelium and epithelial tissues where different keratins are expressed [56].

1.3.2 Type 3

Type three IF proteins are combined in one group due to the similarity of the rod domain sequences [32, 39]. Despite such similarity, it is a heterogeneous group, including different proteins characteristic for different tissues. It includes vimentin found mainly in the cells of mesenchymal origin, desmin and syncoilin characteristic of the muscular cells, GFAP in the cells of glial origin, and peripherin found in peripheral neurons. Consequently, the functions of these proteins differ and depend significantly on the cell type they are expressed in.

It was noticed early on that different IF proteins of the third type may co-express. For example, along with GFAP vimentin can be expressed in the glial cells [29], desmin and vimentin are co-expressed in muscle cells and fibroblasts [18, 65]. Later it was also shown that type 3 IF proteins can co-polymerise *in vitro*, for example, desmin and vimentin [66]. Probably, the property of IF proteins to replace each other serves an important role in the case when, for example, one of the IF proteins is damaged or expressed in a low amount and hence may be partially substituted.

Vimentin

Vimentin is found in many tissues, from eye lens to the muscle and sperm cells but not independently from other IFs, suggesting it acts in combination with those IFs in many aspects of cell function. As all IF proteins, vimentin supports and anchors the position of the organelles in the cytosol (it is attached to the nucleus, endoplasmic reticulum and mitochondria, either laterally or terminally), as well as provides cells with resilience. In addition to mechanical functions, vimentin is involved into many other processes. It plays an important role in human health. It is shown to regulate leukocyte migration and cell adhesion [51]. Moreover, vimentin is overexpressed in different cancers tumour cells (for review see [67]). Its presence in the cytoplasm is shown to be necessary for metastatic spread of many cancers, for instance in breast tumour cell lines [68]. It was demonstrated that vimentin is required not for growth but for the late stage elongation of invadopodia – a specialised protrusion of invasive cancer cells, typical for breast and colon cancers [69]. Vimentin is used for detecting colon cancer, as it is differentially methylated in tissues from colon cancer and unmethylated in normal human tissues (patent [70]). It is known that vimentin is highly expressed in lens fibre cells, forming there a membrane-connected cytoskeleton. Correspondingly, mutations in vimentin can affect the eyesight, *e.g.*, the E151K mutation causes a dominant, pulverulent cataract [71]. Surprisingly, despite the presence and the high conservation of vimentin in different species, vimentin null-mice show no abnormalities under normal conditions [72]. However the phenotypic differences occur under stress conditions, such as CNS injury - the absence of vimentin or GFAP causes reduced scar formation and axonal regeneration as well as skin wound healing becomes slower due to delayed migration of fibroblasts into the wound site [73]. Furthermore, counterbalancing on the rotating platform

becomes very difficult for vimentin null-mice, which points to affected motor coordination [74]. In addition, vimentin is shown to interact with viruses (*e.g.*, foot-and-mouth disease [75] and Bluetongue virus [76], affecting cattle, pig, sheep, goats, etc.) and a disruption of the vimentin structures results in a reduction of the amount of the virus released from infected cells [75, 76]. In addition, recently a new vimentin sub-type was found, which is secreted from both blood endothelial cells and activated macrophages; its role is not clear yet [77]. Many functions of vimentin working in a tight coordination with other proteins remain to be identified.

GFAP

Glial fibrillary acidic protein was first found in glial cells (i.e. microglia, astrocytes, oligodendrocytes, ependymal cells [31]) and later in other tissues. It is the main IF protein in astroglia where vimentin, nestin and synemin can also be found. This protein demonstrates a high level of variability: at least eight different isoforms of GFAP have been identified [78]. It is also known that GFAP expression is enhanced upon brain damage, CNS degeneration as well as in the aged brain [78].

A GFAP-related Alexander disease (AxD) is caused by the biggest number of various mutations (Table 2 [61]), despite the fact that only about 500 cases have been described. This progressive neurological degenerative disease shows such symptoms as rapid head growth and clinical decline, usually fatal within the first decade of life. Patients suffer seizures, atonia and a range of cognitive and developmental defects. This disease is characterized by the degeneration of fibres and cell bodies of the fibrillary neuroglia, where GFAP, αB-crystallin, heat shock protein 27 and other components aggregate. It was shown that GFAP mutations, causing AxD, decrease the solubility of this protein [79], which might be explained by jeopardized filament assembly.

Peripherin

Peripherin is mainly found in neurons of the peripheral nervous system. As typical for the type 3 IF proteins, it forms homodimers at the first step of assembly. However, in some neurons it was found in heterodimers with neurofilaments or α-internexin. In dorsal root ganglia cells of rats it forms an IF network together with α-internexin and neurofilaments (NF, including H, M and L subunits) [80]. Peripherin is shown to be involved in axonal elongation, axonal regeneration after an injury and development of sensory neurons, however many aspects of its functioning are not understood completely [81]. A number of known mutations in peripherin and NF proteins cause amyotrophic lateral sclerosis 1 (ALS1). According to available statistics, it is one of the major neuropathologies found in ~1/50000 Americans. Upon this disease, upper and lower motor neurons degenerate. As a consequence, muscles do not get the nervous signals, weaken and undergo atrophy [61].

Desmin

Desmin is the main IF protein of muscle cells including skeletal, cardiac and smooth muscles. Its name derives from the Greek "desmos" – a link. Indeed, as the name implies, it plays an important role in anchoring the organelles in the cytosol. It links Z-discs of neighbouring sarcomeres (laterally and longitudinally – see Fig. 4), the nucleus, mitochondria and cell membrane. In addition desmin serves as a connection between neighbouring muscle cells *via* desmosomes (in cardiac muscles) or costamers and myotendinous junctions (in skeletal muscles) (Fig. 4 [82]). Consequently, it provides a mechanical and structural integrity of the muscles. In mammalian skeletal muscles it can be detected at the early developmental stage (in somites and myoblasts of the embryo). It participates in myogenesis - but at that stage it is not the major IF protein in muscles, which instead is vimentin [83]. This may explain why desmin knock-out mice develop normally, with normally looking muscles, but after birth cardiomyopathy, skeletal myopathy and smooth muscle dysfunction begin to develop [84, 85]. In these mice the positioning, distribution and function of mitochondria are strongly affected. The anchorage of the myofibrils to the plasma membrane at the costameres is easily lost and regeneration of the skeletal muscles is delayed [84].

Human desmin is 53.5 kDa in size. It is related to the largest percent of known diseases among the type 3 of IF proteins (Table 2). Desmin-related myopathy manifests itself by the structural disintegration of the muscle Z-line and accumulations of protein in the muscle fibres. Clinical features include progressive weakness in skeletal muscles, cardiac arrhythmia and other irregularities of the heart function [61].
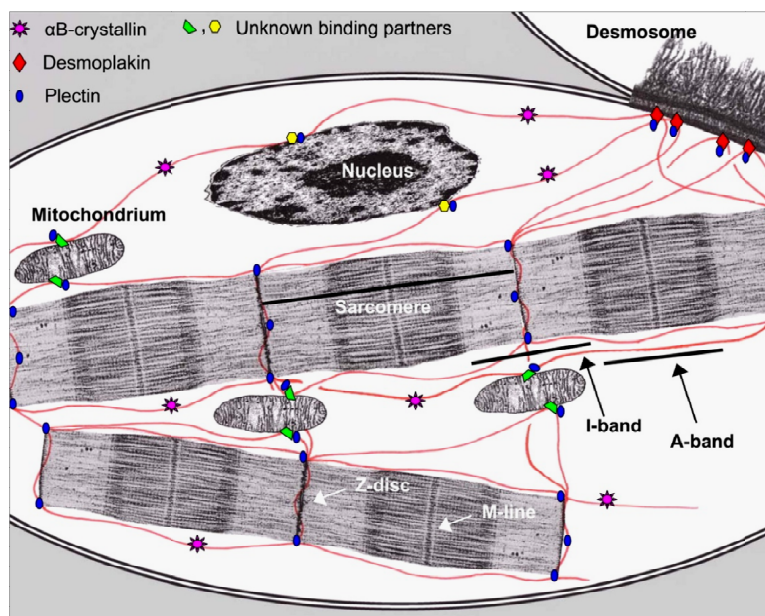
Fig. 4. Schematic representation of a myocyte. Desmin (shown as red filaments) interconnects Z-discs of neighbour sarcomeres, the nucleus, mitochondria and the cell membrane. Individual myocytes are connected via desmosomes (in cardiac muscles) or costamers and myotendinous junctions (in skeletal muscles), where desmin is found in the interaction with desmoplakin and plectin (Reproduced from [82] with permission of Elsevier, licence agreement number 3259430110476).

Nestin, syncoilin and synemin

Due to sequence homology, nestin, syncoilin and synemin are assigned to the type 3, although they have a lot of sequence similarity with the type 4 IF proteins [82]. Nestin is found mainly in the cells of central and peripheral nervous system at the early stages of development. Upon development nestin degrades. As it exists only in early developing neurons and glia, nestin serves as a marker of neurogenesis. Syncoilin and synemin are mainly found in the muscle cells. None of these three proteins are able to self-assemble into IFs on their own. Instead, they form heteropolymers. Nestin is shown to form *in vitro* heteropolymers with vimentin or α-internexin [86]. Syncoilin and synemin form heterodimers with vimentin or desmin *in vitro* and are included in the desmin aggregates in muscle (for review see [82, 87]). So far, no mutations related to diseases have been found in these proteins [61].

Table 2. Diseases associated with mutations in the type 3 IF proteins [61].

Disease	№ of variation record	Protein
Acquired partial lipodystrophy	1	GFAP
Alexander disease	205	GFAP
Amyotrophic lateral sclerosis 1	9	peripherin
Autosomal dominant cataract	1	vimentin
Cardiomyopathy (different types)	17	desmin
Desmin-related limb-girdle muscular dystrophy	2	desmin
Desmin-related myopathy	119	desmin
Distal myopathy	2	desmin
Scapuloperoneal syndrome type Kaeser	1	desmin

1.3.3 Type 4

The type 4 includes neurofilament proteins and α-internexin. Neurofilaments together with the MTs are the main structural elements in the cytoplasm of neurons. NF family is represented by the three types: light (NF-L, 62 kDa in size), medium (NF-M, 102 kDa) and heavy (NF-H, 112 kDa) chains. NFs are expressed in high concentrations in axons of neuronal cells, providing axon the mechanical strength and regulating its diameter. Among the proteins of the fourth type, α-internexin is expressed first and later found co-polymerized with NFs in most of mature neurons. NF proteins as well as keratins at the first step of assembly form heterodimers, NF-L /NF-M, NF-L/NF-H as well as any of NFs with vimentin or α-internexin [88]. Among the fourth type of IFs, α-internexin can form homodimers in a specific manner, similar to lamins, forming first longitudinally annealed dimers which then assemble laterally [88, 89].

Mutations in NFs are shown to be involved into several neurodegenerative disorders. Charcot-Marie-Tooth group of diseases (characterised by progressive loss of muscles and

touch sensation, one of the most common inherited neurological disorders affecting 1/2,500 people), ALS (discussed earlier), Parkinson (characterised by shaking, rigidity, difficulty of movement and depressions, results in accumulation of α-synuclein into inclusion bodies in neurons) and neuronal intermediate filament inclusion disease (characterised by behavioural and personality changes, memory loss, cognitive impairment, language deficits and motor weakness) [61]. Due to a great variety of proteins involved in these diseases, its heterogeneity and complexity, the exact mechanism of most of them remains to be discovered.

1.3.4 Type 5

Lamins belong to the fifth type of IF proteins, based on their intracellular localization (nucleus) and the assembly properties [38]. Being uniquely nuclear IF proteins, they show a number of differences with cytoplasmic IF on the different levels.

At the sequence level, there are two features characteristic for lamins. First, their coil1B is 42 amino acids (six heptads) longer than the one of cytoplasmic IF proteins. In addition, their rod, according to the secondary structure predictions is capable of forming a contiguously α-helical rod domain [90] (for sequence see Fig. 11). All lamins include a nuclear localization site (NLS) and compact domain with immunoglobulin (Ig) fold as a part of the tail domain. In addition, most of lamins have a conserved CAAX box at the C-end, which undergoes farnesylation during lamina maturation.

At the level of assembly, nuclear lamins show a behaviour which is different from that of most of the cytoplasmic IFs, similar only with α-internexin (Fig. 7 C). The assembly process starts with the formation of longitudinally annealed dimers and tetramers where the beginning of the rod overlaps with the end of the rod (in so called A_{CN} mode, see Fig. 6, [91, 92]). Some experiments show that this overlap is crucial for assembly [93]. Thereafter, long, tetrameric in a cross-section units, assemble laterally in ~10 nm wide IFs *in vivo* [38, 91, 92, 94-96]. However *in vitro* assembly results in paracrystals – a layer of laterally assembled long units with the major axial repeat of stained bands 48–50 nm from each other [94]. Interestingly, only one type of lamin found in *Caenorhabditis elegans* (*C. elegans*) could be assembled into filaments *in vitro* [95], but showed 12-16 *versus* 32-48 monomers per cross-section in cytoplasmic vimentin [97].

At the cellular level, lamin network looks differently from the vimentin, desmin and other cytoplasmic IF networks. While, in general, IFs in the cytoplasm are distributed throughout the whole space (Fig. 1A), lamins mainly form a nuclear lamina attached to the perinuclear membrane (Fig. 1B). Thus, the concentration of lamins in the nuclear periphery is much higher than in the non-membrane associated nucleoplasm, where, however, some lamins are present - depending on the lamin/cell type [98] and the moment of the cell cycle [34]).

It is believed that evolutionary cytoplasmic IF proteins derive from lamins. This assumption is based on two facts. First, there are organisms having either both, nuclear and cytoplasmic IFs, or only lamin-like IFs [98]. Second, cytoplasmic IFs found in the most primitive chordates (*e.g.*, tunicate *Styela plicata* has IFs close by homology to desmin and keratin 8), molluscs (*e.g.*, IFs in oesophagus epithelium of the snail *Helix pomatia)* and in nematodes (*e.g.*, IFs from a muscle tissue of the *Ascaris suum),* all have additional 42 amino acid domain within the coil1B, characteristic for lamins [99, 100]. Proper nuclear lamins have been found only in metazoans of the animal kingdom [98].

There is also a hypothesis that lamins evolved as an adaptation to mobility in animals. First, lamins are not found in plants and fungi (except some IFs, *e.g.* found in the pea *Pisum sativum* [101]). Second, the symptoms of many laminopathies include muscle defects [102], hence lamins are important for mechanical integrity of the physically stressed cells. Lamins play an important role for the nuclear integrity and maintenance of the chromatin structure. Lamin IFs form a network for proper functioning of the transcription apparatus: lamina matrix allows different proteins to bind to chromatin for DNA regulation, transcription, replication and repair [103, 104].

Based on the structure and expression patterns, all lamins are divided into two types: A and B. The A-type lamins are mainly expressed following the gastrulation phase of tissue development. Lamin B is expressed in all somatic cells of mature organisms.

In mammals, there are three lamin genes: LMNA, LMNB1 and LMNB2, expressing seven lamin isoforms. LMNA is a matrix for expression of the two major lamins, A and C; and two minor lamins. Lamin C is a product of two post-translational modifications of the prelamin A – 1) truncation of 98 amino acids (including the CAAX box) and 2) binding of six unique amino acids at the C-terminus. LMNB1 and LMNB2 encode B-type lamins (B1, B2 and B3) [98]. B1 and B2 lamins are found in all somatic cells, while B3 type is characteristic for the germ-line cells (for details see Table 3) [105].

Interestingly, the C-terminus of lamins is the most variable region, defining the functional differences between lamins. It is also the main binding domain for lamin-binding proteins [103].

Table 3. Classification and characterization of human lamins

Gene	type of lamin	where expressed	Diseases
LMNA	A	in differentiating cells [36, 104]	>750 sequence variants of different mutations in *LMNA* causing a vast majority of laminopathies [61]
	C		
	AΔ10		
	C2	germ-line cells [106]	
LMNB1	B1	all cells, including embryonic stem cells [104]	Autosomal dominant leukodystrophy [107]
LMNB2	B2		Acquired partial lipodystrophy [108]
	B3	germ-line cells [105]	

Strikingly, different mutations of more than 30% of the 664 lamin A amino acids are shown to cause diseases [61, 98]. A common name of this group of diseases is laminopathies. The pathogenic mutations are located throughout the whole lamin sequence and not specifically restricted to one region (for an overview see Fig. 6 in [98], and Fig. 5 in this thesis adapted from [64]). Due to such a complexity, there is still a lot of research needed to understand the mechanisms of most of laminopathies. One of the most severe types of this disease is progeria. It causes premature ageing, short stature, tight atrophied skin, hair greying and it is usually lethal in the second decade of life. The mechanism of this disease is connected to the jeopardized post-translational modification of prelamin A. In the normal conditions prelamin A is farnesylated at the Cys of the CAAX box, then the C-terminal AAX amino acids are removed and the terminal Cys is methylated. This farnesylated prelamin A is then incorporated into the perinuclear envelope as a part of nuclear lamina. To dissociate from the lamina, prelamin A has to be modified into lamin A *via* the proteolysis of the 15 C-terminal amino acids including the farnesylated site. When the last step is affected due to mutations, there is no lamin A released in the nucleoplasm, the nuclear IF network weakens, resulting in abnormal nuclear shape and affected gene expression [109].

1.3.5 Connection between the localization of mutations and disease phenotype

A tremendous number of reports on IF protein mutations causing human diseases has been released in the last few years ([61, 64] Fig. 5). Correspondingly, "IF-pathies" are a hot topic and molecular information is urgently needed for the understanding of potential pathomechanisms. Interestingly, many pathogenic mutations are found in conservative region of the rod domain of IF proteins – coil 1A and the end of coil2 [61]. For example, mutations in coil1A of K14 [62] as well as in the end of coil 2 in K5 and K14 cause EBS [57, 61]. All reported mutations in K9 causing epidermolytic palmoplantar keratoderma (a common type of keratoderma, *e.g.,* found in 4.4 per 100,000 men in Northern Ireland) affecting palms and soles are within coil1A [110]). Mutations in coil1A of K17 lead to hypertrophic nail dystrophy (pachyonychia congenita type 2) [111]. Mutations in coil 1A of GFAP cause severe developmental neurological disorder named Alexander disease [79]. Likewise, mutations mainly found in coil1B and in the end of coil2 of desmin give rise to muscular dystrophies and cardiomyopathies. In contrast, for nuclear IF proteins a correlation between the localization of mutations and disease phenotype is less pronounced. More than 400 mutations located all along the molecule of nuclear lamin A, expressed in differentiating cells of the human body, lead to more than 10 distinct disease entities [98].

IF-related disease mechanisms are tightly connected to their structure, and its knowledge is indispensible for developing treatment strategies. Structures at atomic

resolution may help to understand how the interactions within particular regions of IF proteins influence the functioning of cells, tissues and organs.
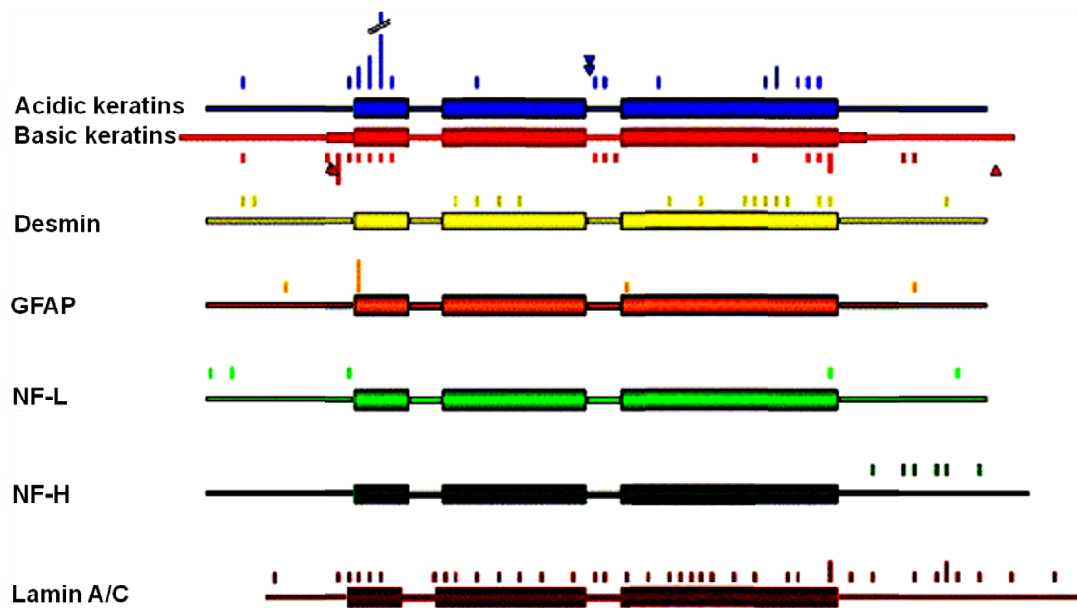
Fig. 5. Distribution of mutations within different domains of IF proteins leading to human diseases. Seven IF proteins belonging to five different types are shown schematically with their head, coil1A, coil1B, L12, coil2 and tail domains respectively. The length of vertical bars corresponds to the frequency of mutations, triangles indicate widely spread IBD, liver and pancreatic disorders, caused by mutations in K8/K18. It is easy to see that the mutations in cytoplasmic IFs are more domain-specific, often affecting specifically head, tail and conservative regions (the beginning and the end of the rod domain), while in lamins mutations are found all along its sequence. Figure is adapted from [64] (with permission of Elsevier, licence agreement number 3259421251614), where a correction to the coil2 domain structure was added.

1.4 IF structure and assembly mechanism

1.4.1 Sequence analysis
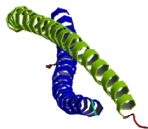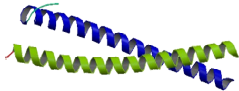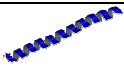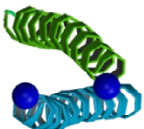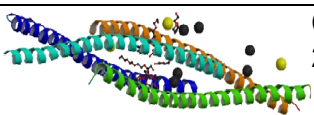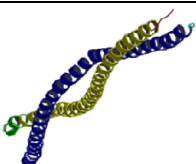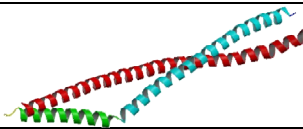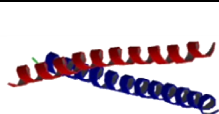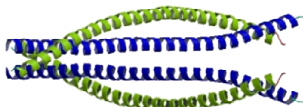
As it was described earlier, circular dichroism experiments [25] and the sequence analysis [24, 26] revealed a tripartite structure, an assigned hallmark of all IF proteins: a central mostly α-helical rod domain is flanked by non-α-helical head and tail domains. While the rod domain has a uniform size and includes two conservative domains in the beginning and the end [24] (Fig. 2, A), head and tail domains have no defined secondary structure and highly variable in both, amino acid composition and length. It was noticed early on that a specific distribution of hydrophobic amino acids within the rod allows formation of a CC [24, 27] (Fig. 2, B-D). A further detailed analysis of the amino acid sequences of the rod domains gave different results. Some sequence-based predictions suggested that the rod should contain three contiguous α-helical regions, relying only on the propensity of different amino acids to form an α-helix [24, 112]. Another analysis taking into account only a heptad-based distribution of apolar amino acids, suggested the existence of four CC segments (coils 1A,
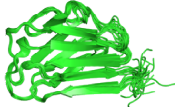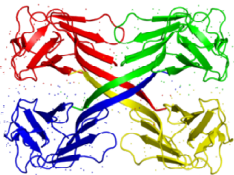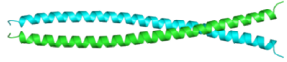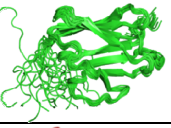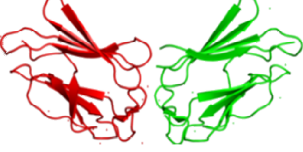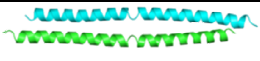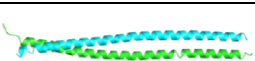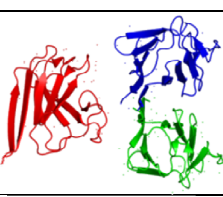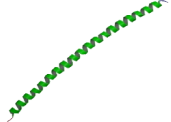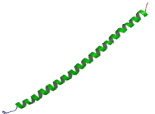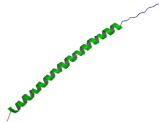
1B, 2A and 2B) interconnected by linkers L1, L12 and L2 (reviewed in [113]). However, more recent analyses suggested that there may be deviations from the heptad pattern [113, 114] These deviations are called stutters (four-residue inserts) and stammers (three-residue inserts) [115]. A combination of the heptad and one stutter leads to eleven residues; a heptad with two stutters results in 15 residues. Therefore, known repeat patterns of the core residues in CCs are represented by heptad (abcdefg), hendecad (abcdefghijk) or quindecad (abcdefghijklmno) repeats, where positions a, d, h, l are the core positions. Correspondingly, classical left-handed, parallel or right-handed CCs can be formed, respectively [28, 116, 117]. According to the study of Harbury *et al.*, depending on the composition of residues in the hydrophobic core, CC dimers, trimers or tetramers may occur [118]. CC core positions are usually occupied by Leu, Ala, Glu, Val, Ile and Lys residues [118, 119].

1.4.2 Structural studies of IF fragments

The assembly process of IF proteins is tightly connected to the CC formation. This is a driving force for dimerization of two monomers containing α-helical rods into a parallel left-handed CC [120]. Such dimer is a very stable complex. In particular, cytoplasmic vimentin reveals dimers even in 6 M urea [121]. Consequently, the smallest unit which can reveal an IF protein structure is its dimeric "building block". Two techniques are used to get an atomic resolution structures: nuclear magnetic resonance (NMR) and X-ray crystallography. Both techniques are only partially applicable for IF protein studies. This is due to the fact that the elongated full length IF dimer (~45 nm – cytoplasmic, ~50 nm - nuclear) is very flexible and has a tendency to self-assemble. Both properties are known to jeopardize crystallization, and NMR can be used only for small molecules. In order to overcome this obstacles, a "divide-and-conquer" approach was applied [122]. It is based on the study of many short fragments of IF dimer with subsequent merging of the overlapping structures. Using this approach, a number of CC fragments of vimentin, keratin and lamin proteins has been resolved crystallographically as well as using the NMR technique and deposited in the PDB (www.rcsb.org, Table 4).

Table 4. Summary of solved X-ray / NMR structures of IF protein fragments (www.rcsb.org).

Human vimentin fragment	PDB ID	Structure	Description (reference)
328-411, called cys2	1GK4		Crystal structure of human vimentin coil 2B fragment [123]
385 – 412, called Z2B	1GK6		Crystal structure of human vimentin coil 2B fragment linked to GCN4 Leu zipper [123]
102-138	1GK7		Crystal structure of human vimentin coil 1A fragment [123]
101-139	3G1E		Crystal structure of coil 1A of human vimentin with a stabilizing mutation Y117L [124]
263-334	3KLT		Crystal structure of the beginning of coil 2 fragment of human vimentin [125]
144-251	3UF1		Crystal structure of human vimentin linkerL1-coil1B fragment [126]
261-335	**3TRT***		Crystal structure of stabilized vimentin coil2 fragment with mutations L265C, L269(MSE), C328(MSE) [127]
153-238	**3SWK***		Crystal structure of vimentin coil1B fragment [128]
99-189	**3S4R***		Crystal structure of vimentin coil1A/1B fragment with a stabilizing mutation Y117L [128]
crystallized: 99-189, visible: 144-189	**3SSU***		Crystal structure of vimentin coil1A/1B fragment [128]

Keratin fragment	PDB ID	Structure	Description (reference)
Human K14 (295-422) human K5 (350-477)	3TNU		Crystal structure of heterocomplex of coil 2B domains of human keratin 5 and keratin 14 [129]

Lamin fragments	PDB ID	Structure	Description (reference)
Mouse lamin A (408 - 545)	3UFG		NMR structure of Ig-like domain of mouse nuclear lamin (Kobayashi, *et al.*, unpublished)
Human lamin A/C (428-549)	1IVT		NMR structures of the C-terminal globular domain of human lam a/c [130]
Human lamin A/C (436-552)	1IFR		Crystal structure of lamin a/c globular domain [131]
Human lamin A/C (435-552)	3GEF		Crystal structure of the r482w mutant of lamin a/c [132]
Human lamin A/C (305-387)	1X8Y		Crystal structure of human lamin coil 2b [92]
Human lamin B2 (449-569)	2LLL		NMR structure of c-terminal globular domain of human lamin B2 (A.Lemak *et al.*, unpublished)
Human lamin B1 (426-558)	3JT0		Crystal structure of the C-terminal fragment (426-558) of human lamin B1 (Kuzin *et al.*, unpublished)
Human lamin A/C (328-398)	2XV5		Crystal structure of human lamin A, coil 2b fragment [133]
Human lamin B1 (311-388)	3TYY		Crystal structure of human lamin B1, coil 2 segment [134]
Human lamin B1 (428-550)	3UMN		Crystal structure of lamin B1: globular tail domain, residues 428-550 [134].
Human lamin B1 (439-549)	2KPW		NMR structure of lamin B1 (439-549), Swapna *et al.*, unpublished
Human lamin A/C (313-386)	3V4Q		Crystal structure of R335W mutant of human lamin A [135]
Human lamin A/C (313-386)	3V4W		Crystal structure of E347K mutant of lamin A [135]
Human lamin A/C (313-386)	3V5B		Crystal structure of coil 2B of human lamin A [135]

Fragments solved in this thesis are also shown; corresponding PDB IDs are bold and underlined.

1.4.3 IF assembly process

Typically, IF proteins are purified in denaturing conditions. It was shown that IFs are insoluble in the aqueous solution, but can be solubilised in the presence of denaturing agents such as 8M urea, 6M guanidine or 0.1% SDS. Thereafter, by performing a step-wise dialysis of the IF protein solution against buffers with decreasing denaturing agent concentrations, we can reconstitute different stages of the assembly process going in the cell, and reassemble the filaments *in vitro* [16, 21]. First, in 6 M urea dimers are obtained [121]. Afterwards, in 5 M urea, two dimers associate into a tetramer, that stays such upon further dialysis into low salt buffer, *e.g.,* 5mM Tris-HCl (pH8.4) [121, 136]. These 60 nm long tetramers result from the association of dimers in a half-staggered, antiparallel fashion.

Cross-linking studies of vimentin and keratin tetramers where Lys residues were stitched with sulfo-DST showed interdimeric cross-links between coils 1 (alignment mode A_{11}) and coils 2 (mode A_{22}, Fig. 6) [137, 138]. In particular, for vimentin the central residue in A_{11} mode was 188 [138]. Two other modes do not exist at the tetrameric stage of assembly, but only at the full length filament: A_{12} – where two antiparallel dimers are in register and A_{CN} – where the beginning of the rod domain of one dimer overlaps with the end of the rod domain of another dimer. Interestingly, A_{CN} mode does not exist in mature cytoplasmic IFs and the calculation of Steinert et al. [137, 138] is most probably erroneous. This is opposite to lamins, where A_{CN} mode is shown to be indispensible for the assembly process and found also in the crystal structures [91, 93].

Another approach used to reveal the arrangement of the molecules in the tetramer was site-directed spin labelling electron-paramagnetic resonance (SDSL-EPR). In EPR experiments the signals from nitroxide spin labels attached to Cys residues are used (Cys can be incorporated by point mutagenesis in a designed position). A series of experiments on vimentin confirmed the existence of alignment modes A_{11} and A_{22}. The central residues of A_{11} and A_{22} modes were shown to be 191 and 348, respectively [139].
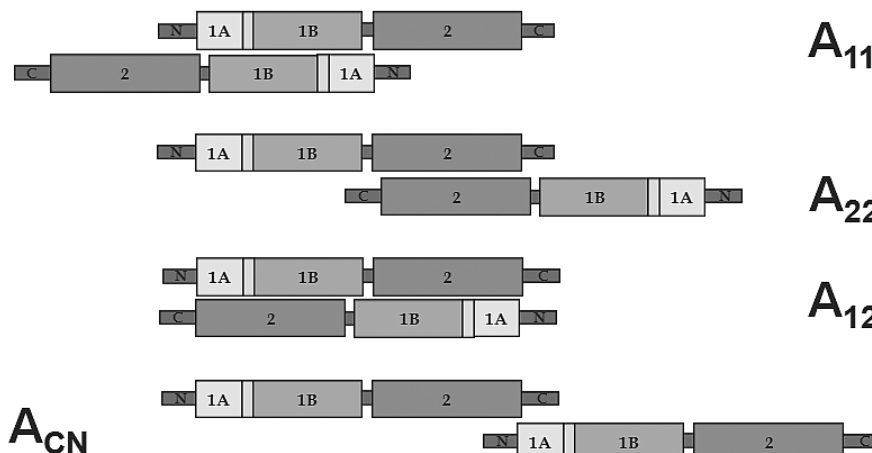
Fig. 6. Alignment mode of the IF protein dimers, based on cross-linking experiments of Steinert et al. [137, 138].

Modes A_{11}, A_{22} and A_{12} coexist in the mature filament [138], but, however, in solution mainly the mode A_{11} can be observed. A tetramer in A_{11}-mode [121, 140] is the building element of cytoplasmic IFs. Once the ionic strength is raised, rapid lateral association of about 8 A_{11}-type tetramers results in the formation of so-called unit-length filaments (ULFs), 16 nm in diameter [97]. Subsequently, the ULFs gradually anneal in the longitudinal direction, ultimately yielding long compacted filaments of 10-12 nm in diameter [121]. Such a mechanism of assembly is a feature of almost all cytoplasmic IFs (Fig. 7A). Nevertheless among IFs following this classical mode of assembly there are exclusions. For example, at the first step of the assembly, keratins, neurofilaments and few other IF proteins form heterodimers [33, 80, 88] (Fig. 7B). In addition, nuclear lamins also follow a different scheme of assembly (Fig. 7C). Its mechanism is less understood. The assembly process of lamins is believed to start with the parallel, unstaggered dimers assembling head-to-tail with a few nm overlap to yield linear polymers with a ~48 nm repeat distance [91, 93]. Thereafter long dimeric structures align in an anti-parallel way to form protofilaments – longitudinal tetramers [94]. *In vitro* such longitudinally annealed dimers/tetramers associate in paracrystals [38, 91, 92, 94-96]. Possibly *in vivo* at the same dimeric or tetrameric stage these protofilaments form mature filaments. It was shown that lamin filaments of *C. elegans* assembled *in vitro* are composed of three or four laterally associated protofilaments [95]. However for human lamins this number varies both: along the same filament and between different filaments, and is hardly measurable, due to the fact that *in vitro* no filaments could be assembled. Nevertheless, in all cases mature lamin filaments in different species are ~10nm in diameter.
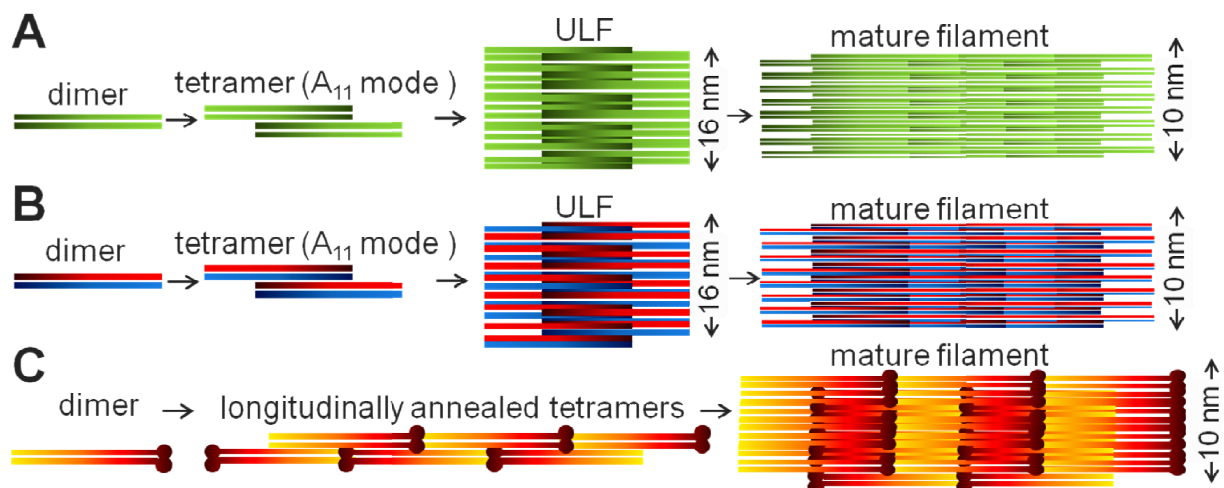
*Fig. 7. Schematic representation of the assembly steps of IFs. **A.** Classical assembly, characteristic for most of the cytoplasmic IFs including vimentin. The assembly starts with the formation of a homodimer. Thereafter two antiparallel dimers anneal laterally in A_{11} mode into the tetramer, then ULF (~16 nm in diameter), which in turn longitudinally anneal into filaments, following the last step of filament maturation – i.e. compactization. Mature filament is about 10 nm wide. **B.** Filament assembly characteristic for keratins and NFs, starting from the formation of a heterodimer (on the left), further steps of assembly are the same as described in A. **C.** Filament assembly of lamins. When homodimers are formed, they first anneal longitudinally. Thereafter the longitudinal tetrameric (in a cross-section) constructs are formed, following lateral annealing and compactization stages, resulting in mature ~10 nm in diameter filaments.*

1.4.4 Role of the head and tail domains

Head and tail domains are flexible in comparison with the CC rod domain, and represent a high variability across the IF family. These domains are shown to be involved in inter- or intra-dimeric interaction. In contrast to the rod domain (which is 308 residues long in vimentin and 359 residues in lamin A, Fig.11), the head and tail domains are variable in size and amino acid composition. Correspondingly, the difference in the mass of different IF proteins, listed in Table 1, is mainly due to the variable head/tail domains. Probably, the head and tail domains are the triggers defining the main differences between different IF proteins, leading to different filament assembly pathways and different mature filaments, maintaining different functions in the cell.

While the similarity between head as well as tail regions of different IF proteins is low, the same IF protein typically has common features in different species. For instance, the head domain of vimentin in several fish, frog, chicken and humans has two conserved regions (9 and 12 residues long), a specific pattern of aromatic amino acids as well as 12 Arg residues scattered along its length. The latter residues are important for the filament assembly (for details see [141]). As described before, wild-type (WT) vimentin forms A_{11} mode tetramers in the low salt buffer (5mM Tris-HCl pH 8.4) [140] and assembles into filaments under high salt conditions (50-160 mM NaCl, 25 mM Tris-HCl pH 7.5) [121]. The vimentin head deletion (up to the residue 82 in human and the residue 72 in *Xenopus*) leads to the assembly arrest at the level of a dimer in a low salt concentration buffer. Furthermore, in the presence of a high salt concentration the assembly stops at the level of the tetramers in A_{22} mode [121, 140]. Probably, by preventing interactions between coils2 of two neighbour dimers, the head domain plays a crucial role for longitudinal annealing of ULFs involving interactions in the mode A_{22}. It was also demonstrated for other IF proteins that the presence of the head domain is crucial for the formation of *bona fide* filaments [142].

The role of the tail domain is less understood. Among different species the vimentin and desmin tail domains are well conserved as well as have three similar motifs with keratins 8/18 (for review see [141]). Surprisingly, a deletion of the vimentin tail (in particular experiment, starting from residue 412 for human and 404 for *Xenopus*) did not affect the assembly strongly. However, it led to the formation of filaments that were slightly wider than normal [121]. Some IF proteins lack the tail domain or have a very short one. *E.g.*, eye lens IF protein phakinin has no tail while its obligatory for dimerization partner filensin has a very long one [40]. Same is true for NF-L, which has a very short tail domain and forms heterodimers with long-tailed NF-M or NF-H [88]. Interestingly, it was demonstrated that tailless NF-M and NF-H retained ability to form mentioned heterodimers [88]. In contrast to cytoplasmic IF proteins, lamin A and B tail domains contain a conservative 105 residues block. Its multiple structures were solved crystallographically and using NMR (for details see

Table 4) and its structure represents immunoglobulin-like domain (for review see [42]). A number of mutations found within this block cause laminopathies [61].

Several techniques have been used to study the structure of the head and tail domains of cytoplasmic IF proteins. Cross-linking studies revealed several intermolecular cross-links between the head and the middle of coil2 as well as intramolecular cross-links pointing at the head stretched along coil1A and a half of coil1B (Wedig et al., in preparation; Fig. 8 in [42]). It was suggested that when the salt concentration is raised, different ions interfere with salt bridges between acidic coil1 and basic head domains. This liberates heads which in turn is necessary for interactions with coil2 for the later stages of the filament assembly [42].

SDSL-EPR experiments gave similar results with more details [143, 144]. Head domain was shown to fold back on the coil 1 to form a symmetric structure within the dimer. Two heads are aligned parallel to each other, in register and span across approximately the first half of the coil1. The heads interact with one side of the dimer and make it asymmetric, which in turn must help the alignment of dimers into the tetramer [143].

Moreover, SDSL-EPR was used to see the effect of phosphorylation of vimentin with protein kinase A. *In vivo* phosphorylation is the main regulatory mechanism for IF network formation [53]. *E.g.,* phosphorylation drives the disassembly of vimentin network and division of the cell during mitosis. Moreover in general phosphorylation is a cell signal for proper IF subcellular compartmentalization, turnover, interactions with other proteins, cellular response to stress and disease pathogenesis (for review see [53]). As it was demonstrated by SDSL-EPR experiments, *in vitro* phosphorylation resulted in disassembly of the tetramers (weakening of A_{11} mode and complete extinction of the A_{22} mode); however, dimers stayed intact [145]. Moreover, phosphorylation caused a repulsion of the two heads from each other (see Fig. 5 in [143]) while an interaction with the rod domains was not lost [143]. Apparently, close proximity of the two parallel heads is more favourable for the tetramer and is regulated by phosphorylation.

In contrast to the head domain, the tail domain was shown using SDSL-EPR method to be mainly disordered and flexible in tetramers. However, in intact filaments a short region of the tail (residues 445-452) becomes opposed to the same region of the neighbour tail (see Fig. 5 in [146]). Interestingly, deletions of residues 444-452 as well as a point mutation G405V caused formation of irregular fibrils which aggregated laterally [147]. From this it was concluded that the tail plays its role during longitudinal annealing and compactization stage of the filament maturation [146], which is in agreement with the data showing that the tail domain is partially responsible for the filament diameter [121].

Chapter 2. Research objectives

We focused on X-ray crystallographic studies of human vimentin and lamin A as 'model' IF proteins. To obtain the structure of the dimer, the 'building block' of all IFs, we are using a 'divide-and-conquer' approach [122, 123]. It is based on preparation and crystallization of multiple shorter fragments of IF dimer with subsequent merging of the overlapping structures. This strategy is necessary in order to overcome the flexibility and self-assembly of the full-length protein, as it makes crystallization impossible.

As shown in Table 4, this approach has been successfully used before. For vimentin, several fragments of coil 1 and coil 2 segments were resolved (for the PDB IDs and references see Table 4) but the model of the rod domain remained incomplete. The structures of the linkers L1 and L12 and coil 1B were not resolved. For lamins several structures corresponding to the tail domain as well as the end of the second half of the coil 2 were resolved (for the PDB IDs and references see Table 4) but the major part of the rod domain remained to be investigated.

Relying on similarities between the sequences of vimentin and lamin A rod domains we are able to use the experience with vimentin for the design of lamin fragments. In addition, the aforementioned sequence similarity of different IF proteins should allow us to use one model protein for structure prediction of the others. Structural data on a model IF protein dimer should shed light on the effect of different mutations as well as give an insight into the process of the filament assembly. These key aspects would lead us to better understanding of IF proteins and related diseases. To reach this goal we focused our work on the following issues:

- obtaining X-ray crystallographic structures of vimentin and lamin A fragments;
- modelling the dimer and the tetramer of the IF rod domain based on the fragment structures;
- studying the ULF formation and IF assembly using complementary biophysical methods.

Chapter 3. Methods and general procedures

High-resolution structural information of various regions of vimentin and lamin A can be obtained using protein crystallography in combination with the discussed earlier "divide-and-conquer" approach [122]. In addition, intelligent mutagenesis can be used to minimize the possible artifacts caused by the relatively short length of the fragments.

For this we used the following steps:

1) design of suitable fragments with accurately chosen mutations (if necessary),

2) fragment amplification and cloning into the chosen expression vector,

3) protein expression and purification,

4) protein crystallization,

5) crystallographic diffraction data collection,

6) structure determination,

7) structure interpretation and modelling.

3.1 Design of the suitable fragments

Fragments of the rod domain were designed using data of the cross-linking [138] and site-directed spin labelling electron paramagnetic resonance (SDSL-EPR) experiments [126, 139, 143, 144, 146, 148-150], ultra-centrifugation, EM [121], small-angle X-ray scattering (SAXS) [151, 152], computer modelling, available NMR and crystal structures (see Table 4). Jpred3 [153] and NetSurfP [154] algorithms were employed for the prediction of the α-helical regions and estimation of the probability of a particular residue to be buried within the hydrophobic core. The aim of our design was to get a stable CC dimer, either purely α-helical or containing a linker region. In some cases we had to use mutations in order to preserve a dimeric state and to stabilize the structure.

The success of every crystallographic experiment depends on the possibility to overcome several main 'hurdles' such as protein purification, crystallization in a biologically relevant conformation and phasing. Therefore, the 'trial-and-error' method was extensively used. To determine the structure of any particular region we created a set of fragments of different length with varying N- and C-ends (see Tables 5, 6). In our hands only about one structure out of ten purified proteins could be resolved.

3.2 Molecular cloning

Vimentin fragments were amplified from the full-length human vimentin cDNA. All mutations, listed in the Tables 5, 6, including the mutation Y117L in the *1AB* fragment as well as L265C, L269M, C328M mutations in the *D3st* fragment, were introduced into the corresponding expression vectors using the QuikChange site-directed mutagenesis protocol

(Stratagene). All plasmid sequences were verified by DNA sequencing starting with the T7 promoter or terminator.

Three different vectors were used for the expression of the protein fragments (Tables 5 and 6): 1) the vector pPEP-TEV that allows overexpression of a fusion construct including an N-terminal His$_6$-tag, a 5kDa laminin spacer, a TEV protease cleavage site and the sequence of interest; 2) the pETHSUK/pETHSUL vector [155], resulting in a fusion product containing a His$_6$-tag, the SUMO sequence, the SUMO protease cleavage site and the vimentin fragment; and 3) pET15-NESG vector, allowing us to get a construct carrying His$_6$-SHM- sequence at the N-end [126]. Primers were designed using ApE plasmid editor (written by M. Wayne Davis) and synthesized by LGC Genomics Company (UK).

3.3 Protein expression and purification

Expression constructs were transformed into BL21(DE3) or BL21(DE3) Rosetta *E. coli* competent cells (Stratagene, USA and NEB, England) using the heat-shock method. Induction of the protein expression was done by either of the two methods:

1) Freshly transformed bacteria were cultivated at 37°C in LB medium with 100 µg/ml ampicillin until OD$_{600}$ reached 0.6, and then 3h more in the presence of 1 mM IPTG.

2) Bacteria were cultivated in auto-induction media ZYP-5052 at 24°C till OD$_{600}$ reached 6, and then for 7h more at 18°C [156].

In general, a significant portion of the expressed protein was in the inclusion bodies. Therefore to dissolve the target protein, cells were collected, resuspended in 8M urea, 10mM Tris pH 8, 0.1 M NaH$_2$PO$_4$ and 5mM β-mercaptoethanol, lysed using sonication (Branson Digital Sonifier), and the cellular dust was removed by centrifugation (60 min, 13000 x g). The fusion products were purified on a pre-equilibrated Ni-chelating column (Sigma His-select) and eluted with 8 M urea, 10mM Tris HCl, 0.1 M NaH$_2$PO$_4$, pH 6.3. Pure fractions were dialysed into the 'cleavage buffer' (40mM Tris-HCl pH 8, 150mM NaCl) using 3.5 kDa cut-off membrane (Spectrum) and then digested with the TEV-protease (fragments *1AB* and *1ABL*) or SUMO-protease (coil1B fragments) at 4°C. In the case of NESG constructs no digestion was needed. Where applicable, a second Ni-chelating column was then used to remove the His-tag-containing fragments. Pure protein fractions were concentrated by ultrafiltration (Amicon, 3.5kDa cut-off) and gel-filtered on a Superdex 75, 200 or 300 16/60 columns in the 'cleavage buffer' using an Äkta chromatographic system (GE Healthcare, see Fig. S1).
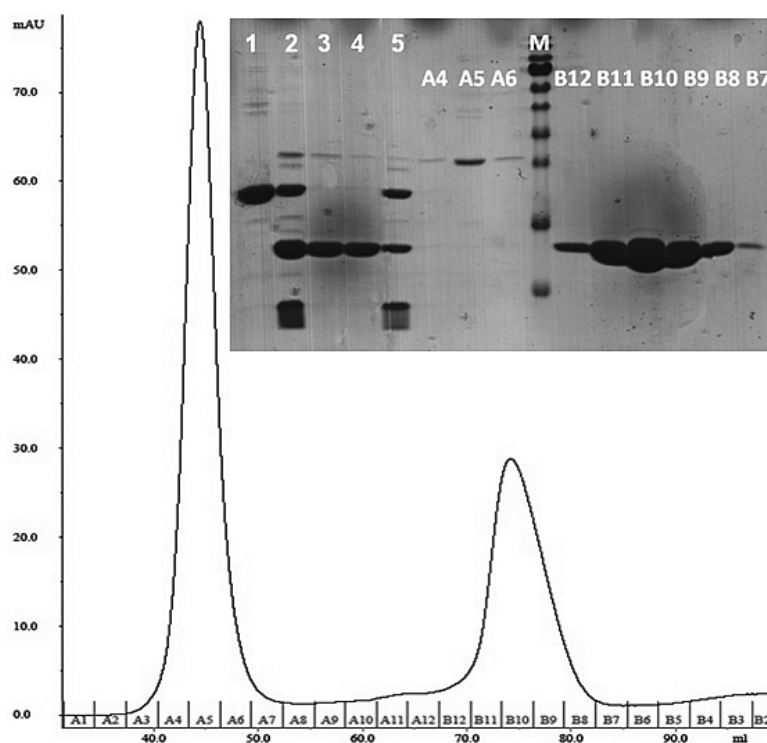
Fig. S1. Purification of lamin A fragment, residues 93-214. Estimated masses of the fusion product and the final protein is 20.2 kDa and 14 kDa, respectively. SDS-PAGE. Lines 1-5 correspond to the protein 1) eluted from the first Ni-chelating column, 2) digested with TEV protease, 3, 4) flow-through from the second Ni-chelating column (cleaved lamin A fragment), 5) elution from the second Ni-chelating column (mixture of cleaved, uncleaved lamin A fragments and His-tag). Lines A4-A6 correspond to the TEV protease eluted from the size-exclusion column. M stands for the protein marker, with corresponding masses 10, 15, 25, 35 kDa, from the bottom to the top, respectively. Lines B12-B7 correspond to the eluted lamin A fragment (14 kDa). Size-exclusion chromatogram of combined protein fractions - lines 3 and 4 on SDS-PAGE. The two main peaks correspond to the eluted TEV protease (fractions A4-A6) and laminA fragment (fractions B12-B8), respectively.

Expression of the selenomethionine (SeMet) containing protein fragments followed either of the two protocols: Cowie and Cohen's (using IPTG induction) [157] or Studier's (using autoinductive PASM 5052 media) [156]. Expression of SeMet-containing protein fragments *1AB*, *1ABL* and *D3st* followed the protocol of Cowie and Cohen [157]. Briefly, the transformed BL21(DE3) cells were grown on a minimal medium (38 g/l $Na_2HPO_4*2H_2O$, 15 g/l KH_2PO_4, 2.5 g NaCl, 5 g/l NH_4Cl) at 37°C until OD_{600}=0.6. At this point SeMet (50 mg/l, Sigma) complemented by natural amino acids (50-100 mg/l) excluding methionine was added. After 15 min the expression was induced by 1 mM IPTG addition. Culture growth was continued until OD_{600} reached 2 (13 h). Expression of the SeMet-containing *1B* fragment was done in the autoinductive PASM 5052 media [156] containing 200 mg/ml of each of 17 amino acids (excluding Cys, Tyr, and Met), 10 mg/ml Met, 125 mg/ml SeMet, and 100 nM vitamin B_{12} [156]. The purification procedures for the SeMet labeled proteins were the same as for the unlabelled. As detected by mass-spectrometry analysis, the substitution of Met by SeMet in the *1B* fragment was about 66% (Fig. 16).

3.4 Full-length vimentin purification

Three Cys mutants were engineered using a QuickChange mutagenesis and DNA ligation techniques. All sequences were verified by GCTA sequencing company. Mutant proteins as well as the WT vimentin were expressed recombinantly using pDS5 vector transformed into TG1 *E. coli* cells. Further purification consisted of the following three steps. Firstly, inclusion bodies were purified and dissolved in 9.5 M urea buffer. Thereafter the proteins were separated from impurities using two consequent anion and cation exchange chromatography columns (respectively, DEAE and CM sepharose, GE Healthcare, Sweden). The fractions were analysed using SDS-PAGE.

3.5 *In vitro* IF assembly and electron microscopy

Pure proteins were used for the filament assembly experiments. Firstly, a step-wise dialysis was performed in 8, 6, 4, 2, 1, 0 M urea buffers until the protein was in the final solution (1mM EDTA, 1mM EGTA, 1 mM DTT, 5mM Tris pH 8.4) at the concentration 0.3 mg/ml. Afterwards an assembly buffer (20mM Tris pH 7, 50 mM NaCl) was added at 37°C and different states of assembly (after 10sec and 1h) were fixed by an addition of a glutaraldehyde solution.

Electron microscopy (EM) grids covered with a layer of carbon were used for preparation of EM samples. Assembled protein was kept on the grid for 15 sec to bind and then negatively stained with uranyl acetate. Dried grids were observed with the electron microscope (DKFZ, Heidelberg, Germany) at the magnification of 6.3-20k.

3.6 Crystallization and crystal mounting

To obtain crystals, a vapour diffusion method was used (Fig. 8 C). A droplet with a mixture of protein and precipitant solutions was brought to equilibrium with the initial precipitant solution in a closed vial, using either a hanging or sitting drop (Fig. 8A). During the evaporation of the water from the protein droplet, both the protein and the precipitant concentrations are increasing. When the solubility limit of the protein is reached, it either precipitates or crystallizes (Fig. 8B).

Before crystallization, pure proteins were concentrated to 1-10 mg/ml (depending on the solubility and the overall yield) in 10-12.5 mM Tris-HCl pH 8, 10-40 mM NaCl. Robotic crystallization screening was performed with commercially available kits including Crystal Screen 1+2, Index, AmSO$_4$ (Hampton Research), JCSG+, PACT (Qiagen) and Wizard I + II (Emerald) crystallization screens. Sitting drops were set up by mixing 0.1-0.3 µl of a protein solution with 0.1-0.3 µl of a reservoir solution in 96-well MRC plates (Molecular Dimensions) at 20°C or 4°C using the Mosquito robot (TTP LabTech) and monitored with Rock Imager (Formulatrix). The crystallization conditions were manually optimized using hanging drops of

typically 1-3 µl in volume. All crystals were briefly soaked in the mother liquor supplemented with 30% glycerol or PEG 400, mounted on the loops (Hampton Research) and flash-frozen in the liquid nitrogen.
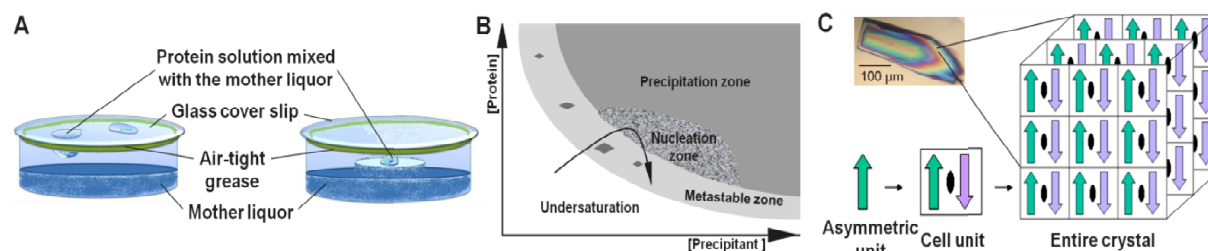
*Fig. 8. Protein crystallization. **A.** Hanging and sitting drops, vapour diffusion technique representation. **B.** Phase diagram, illustrating the change in protein behaviour during the equilibration of the droplet solution with the mother-liquor, adapted from [158]. **C.** An illustrtion of the crystal components (adapted from www.rcsb.org). In the crystal millions copies of the molecule are arranged in a highly ordered way - in a crystallographic lattice. One or several symmetry-related asymmetric units form a building block of the lattice – called cell unit, which, multiplied by symmetry operators, represents the arrangement of all the molecules in the crystal.*

3.7 Data collection

The process of data collection is one of the most important as its quality is often a limiting factor for phasing and also the quality of the final structure.

3.7.1 X-ray sources and detectors

The presence and quality of crystal diffraction was assessed in our laboratory using MicroMax 007HF machine, consisting of a high-intensity microfocus rotating copper anode X-ray generator (Rigaku, Japan, Fig. 9A) and a Mar345 image plate detector (MARresearch, Germany). The generator produces an X-ray flux of ~ 10^9 photons per second, with beam diameter of 150µm at the sample position. Exposures of about 10 min/frame were used.

To achieve higher intensity of the X-ray beam and to be able to tune the wavelength, the synchrotron facilities are used (Fig. 9B). Thanks to higher intensity of the X-ray beam, data collection at the synchrotron may lead to higher resolution data and its collection requires less time (of the order of seconds). In addition, the possibility to vary the wavelength is necessary for optimal data collection for crystals with heavy atoms (HA) where at a particular wavelength an effect of anomalous dispersion takes place. In our case diffraction patterns were collected at the Soleil synchrotron (France) at the beamline Proxima1 using an ADSC Q315r and a pixel detector PILATUS 6M (DECTRIS, Baden, Switzerland) and at the SLS synchrotron (Switzerland) at the beamline X06DA, using a Mar225 CCD detector. For comparison, at the Soleil synchrotron the beam size at the sample position can be set between 100 -250 µm in diameter, and has an electron flux > $5x10^{12}$ photons per second over the entire energy range for a stored beam current of 500 mA [159].

A photon counting detector PILATUS 6M [160] consists of 60 sensor plates, which record X-rays in single-photon counting mode and the data are stored immediately with a very small processing time (5 ms). This detector is also much faster than majority of the other detectors as it is not sensitive to X-rays during readout and therefore no mechanical shutter is required [160]. In addition almost no detector noise is added to the signal as opposed to the integrating detectors (http://pilatus.web.psi.ch/pilatus.htm).
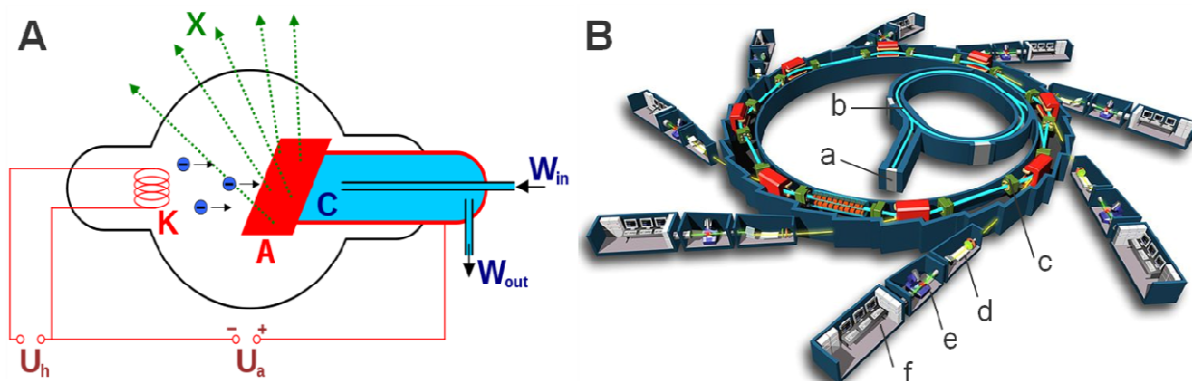
*Fig. 9. X-ray generators. **A.** Schematic representation of an X-ray tube. K/A – cathode and anode, U_h – cathode heater voltage, U_a – anode voltage, $W_{in/out}$ – water circulating system of the anode water cooler (C), X - X-rays (copied from www.wikipedia.org with permission of the copyright holder Hmlich). The X-rays are emitted by the Cu anode after a stream of electrons hits its surface. The electrons are directed by a cathode and accelerated in a vacuum tube. As the X-rays are emitted by the Cu atoms, the emitted wave has a constant length (λ=1.5418 Å). **B.** General scheme of the Soleil synchrotron. Electrons (a flux shown in cyan) are produced by the electron canon (a) and directed into the small accelerating ring, called booster (b), afterwards the accelerated electrons are directed into the big synchrotron ring (c) for further acceleration with the help of particle accelerators (shown in green) and magnets (shown in red). When the speed of the electrons is high enough (close to the speed of light), they emit X-rays (shown in yellow) directed into beamlines. Each beamline is equipped with the optic room (d), the room for X-ray experiments (e) and the room for remote operations of the experiments (copied from www.wikipedia.org with permission of the copyright holder EPSIM 3D/JF Santarelli, Synchrotron Soleil).*

3.7.2 Choice of the X-ray wavelength

To solve the phase problem, we used SeMet derivatives of the proteins, expressed in *E.coli* in SeMet rich media [156, 157]. Se has its predicted K absorption (resonance) edge at 12.658 keV. To confirm the presence of the Se atoms in the crystals and to determine the optimal resonance ('peak') wavelength for SAD experiments, a fluorescence scan was performed (Fig. 10). This scan measures an integral intensity of fluorescence as a function of the incident X-ray energy. Fluorescence is a function of an anomalous component of the atomic form factor (*f''*). The maximum of the f" is at the resonance energy of the HA.
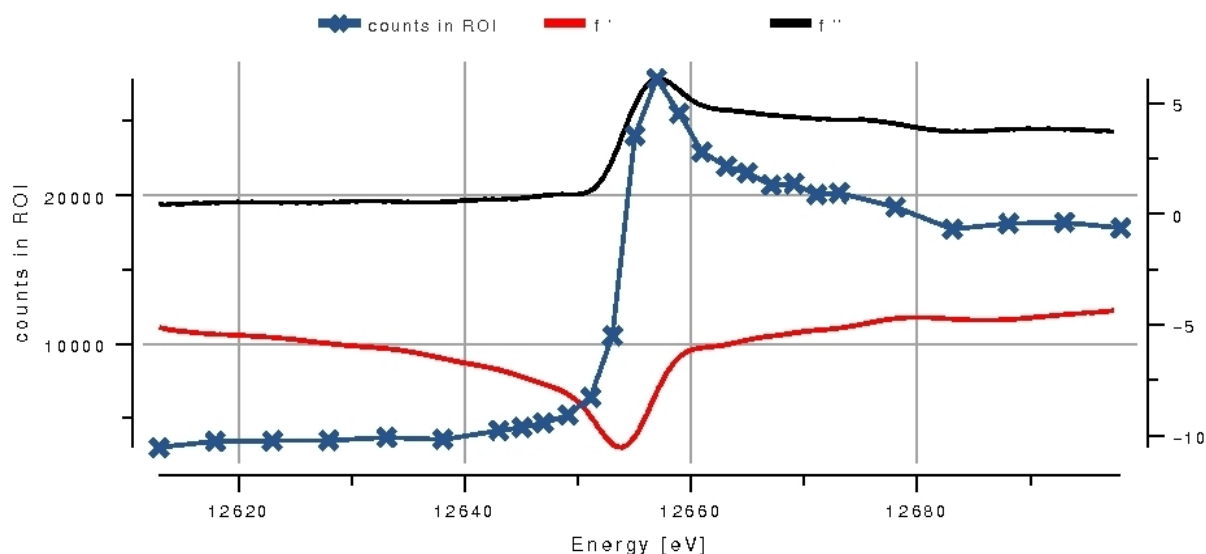
Fig. 10. A fluorescence scan for Se K edge from the crystal of 1ABL vimentin fragment made at Swiss Light Source, beamline PXIII. Theoretical peak of spectral lines of Se are at the energy 12.658 keV. The peak of fluorescence at 12.657 keV (f'=-7.9, f''=6.1) confirms the incorporation of the Se atoms in the crystal. Calculated inflection energy is 12.654 keV (f'=-10.5, f''=3.5). Red line corresponds to the dispersive component f' and black line corresponds to the anomalous component f'' of the atomic form factor, blue line with crosses corresponds to fluorescence.

3.8 Crystal structure determination and analysis

3.8.1 Determination of vimentin coil1 structure: fragments *1AB, 1ABL* and *1B*

Diffraction data were processed using the HKL package [161] (fragments *1AB* and *1ABL)* or XDS and Xscale [162] (fragment *1B,* see Table 7). Initial phasing, density modification and model building was done using an Auto-Rickshaw server [163]. For the fragments *1AB* and *1B,* a SAS protocol of Auto-Rickshaw was used. For the fragment *1ABL* SIRAS protocol of Auto-Rickshaw (a native and SeMet data sets) gave the best results. HA sites were found using the program SHELX [164]. All three structures contained two chains per asymmetric unit, with two (*1AB* and *1ABL* fragments) or three (*1B* fragment) SeMet residues per chain. All Se atoms could be located in each structure. Initial phases were calculated using the program MLPHARE [165]. Density modification, phase extension and NCS averaging (for the *1AB* structure) were performed using the programs RESOLVE [166] and DM [167]. Initial model building was done using ARP/wARP [168] and SHELXE [164]. Manual model completion was performed using Coot [169]. *Phenix.refine* [170] was used for the model refinement. The final refinement of the *1AB* and *1ABL* structures was against the native data sets. The program MolProbity [171] was used to analyze the quality of the final models. Interactions within the structure were analyzed with the program PISA v1.18 [172].

3.8.2 Structure determination of the first half of vimentin coil2: *D3st* fragment

Initially, a dataset was collected for the *D3* (L265C/C328A) crystals, but phasing of these data using MR failed. Thereafter a MAD dataset was collected for SeMet-containing *D3st* crystals, with the presence of selenium confirmed by a fluorescence scan. Diffraction data were processed using the *HKL2000* package [161]. The initial low-resolution (3 Å) two-wavelength MAD dataset (Table 7) was submitted to the Auto-Rickshaw on-line server [163]. All four HAs were found using the SHELX suite, followed by phase calculation, density modification and building of the initial polyalanine model [164, 173]. Thereafter PHENIX [170, 174] and Coot [169] were used for model completion and refinement against the high resolution (2.3 Å) dataset (Table 7). The *D3st* structure revealed one dimer per asymmetric unit.

The program MolProbity [171] was used to analyze the quality of the final models (Table 7). Interactions within the structure were analyzed with the program PISA v1.18 [172].

Chapter 4. Results

As a starting point for crystallographic studies, we have re-visited the prediction of the CC structure within the IF rod based on its amino-acid sequence. We argue that the CC regions should comply with two criteria. Firstly, they should have a high probability of α-helix formation, as can be evaluated using standard secondary structure prediction algorithms, such as Jpred3 [153]. Secondly, they should exhibit regularly spaced apolar residues forming the CC core. To this end, we have used the NetSurfP algorithm [154] that estimates the probability for a particular residue to be buried within the hydrophobic core (Fig. 11). While well-established CC prediction algorithms, such as COILS [175], only consider the heptad periodicity typical for left-handed CCs, one should also look for other possibilities, including the 11-residue (hendecad) and 15-residue (quindecad) periodicities that drive the formation of a parallel α-helical bundle and a right-handed CC, respectively [116, 176].

We found this generalized approach to be highly accurate towards localizing the CC regions of the human vimentin dimer. In particular, Jpred3 algorithm suggests that the first α-helical segment (coil1A) should include residues 97 to 139 (Fig. 11). Moreover, within the whole 1A region the buried residues indicated by the NetSurfP algorithm are in a perfect agreement with the heptad pattern. This very pattern is found in the crystal structure of coil1A (Y117L mutant) dimer *1AL* (PDB entry *3G1E* [124]). At the linker L1 region an interruption of the α-helical structure was predicted for eight residues, Gly140 to Gly147 (Fig. 2A). From analysis of the heptad patterns within coil1A and coil1B one can readily conclude that the linker L1 causes a phase shift in the heptad periodicity that corresponds to an insert of five residues (or a deletion of two residues). This is distinct from either one of the two phase perturbations commonly seen within continuous CC structures, the stutter (equivalent to an insert of four residues, converting a heptad into a hendecad) or the stammer (an insert of three residues) [115, 176]. Moreover, this is also distinct from either of the inserts (stutter or stammer) repeated twice. A continuous CC running from coil1A into coil1B is therefore highly unlikely. The length of the linker L1 and, correspondingly, the phase shift at the linker are conserved throughout all type 3 IF proteins (Fig. 11), while in other IF types L1 has a different length and sequence (Fig. 18) [114].

Interestingly, for both vimentin and lamin A, NetSurfP algorithm showed a possibility for the end of the coil1 and the beginning of the coil2 to form a parallel α-helical bundle (Fig. 11). Subsequently, for vimentin this structure was confirmed experimentally [125-128].

```
<<<      Head        97 --------------- Coil 1A ------------- 139  L1  148 -------------
                    -----------------HHH HHHHHHHHHHH-HHHHHHHHHHHHHHHHHHHHHHHHHHH--------HHHHHHHHHHHHHHHHHH
                     efgabcdefgabcdefgabcdefgabcdefgabcdefgabcde          bcdefgabcdefgabcd
Vim  PGVRLLQDSVDFSLADAINTEFKNTRTNEKVELQELNDRFANYIDKVRFLEQQNKILLAELEQLKGQGKSRLGDLYEEMRELRRQVDQL
Des  PSSYGAGELLDFSLADAVNQEFLTTRTNEKVELQELNDRFANYIEKVRFLEQQNAALAAEVNRLKGREPTRVAELYEEELRELRRQVEVL
GFAP -----LPTRVDFSLAGALNAGFKETRASERAEMMELNDRFASYIEKVRFLEQQNKALAAELNQLRAKEPTKLADVYQAELRELRLRLDQL
Per  ALLRLPSERLDFSMAEALNQEFLATRSNEKQELQELNDRFANFIEKVRFLEQQNAALRGELSQARGQEPARADQLCQQELRELRRELELL
                                                                                 EVV
LamA PSQRRATR-SGAQASSTPLSPTRITRLQEKEDLQELNDRLAVYIDRVRSLETENAGLRLRITESE SREVSGIKAAYEAELGDARKTLDSV
                    ------------------HHHHHHHHHHHHHHHHHHHHHH-HHHHHHHHHHHHHHHHHHHHHH--------HHHHHHHHHHHHHHHH
<<<      Head        23------------------ Coil 1A -------------70  L1  79--------------

                    ------------------------ Coil 1B ---------------------------------------------- 249
                    HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH---EE
                    efgabcdefgabcdefgabcdefgabcdefgabcdefgabcdefgabcdefgabcdefgabcdefghijkabcdefga
Vim  TNDKARVEVERDNLAEDIMRLREKLQEEMLQREEAENTLQSFRQDVDNASLARLDLERKVESLQEEIAFLKKLHEEEIQELQAQIQEQHV
Des  TNQRARVDVERDNLLDDLQRLKAKLQEEIQLKEEAENNLAAFRADVDAATLARIDLERRIESLNEEIAFLKKVHEEEIRELQAQLQEQQV
GFAP TANSARLEVERDNLAQDLATVRQKLQDETNLRLEAENNLAAYRQEADEATLARLDLERKIESLEEEIRFLRKIHEEEVRELQEQLARQQV
Per  GRERDRVQVERDGLAEDLAALKQRLEEETRKREDAEHNLVLFRKDVDDATLSRLELERKIESLMDEIEFLKKLHEEELRDLQVSVESQQV
       NTKKEGDLIAAQARLKDLEALLNSKEAALSTALSEKRTLEGE
LamA AKERARLQLELSKVREEFKELKARLHDLRGQVAKLEAALGEAKKQLQDEMLRRVDAENRLQTMKEELDFQKNIYSEELRETKRRH-ETRL
                    HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH----EEE
                    ------------------------ Coil 1B ----------------------------------------------220

                        265-------------------------- Coil 2 -----------------------------------
                    E-EE--------HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH-HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
                    hijkabcdefghijkabcdefghijkabcdefghijkabcdefgabcdefgabcdefgabcdefgabcdefgabcdef
Vim  Q-IDVDVS-KPDLTAALRDVRQQYESVAAKNLQEAEEWYKSKFADLSEAANRNNDALRQAKQESTEYRRQVQSLTCEVDALKGTNESLER
Des  Q-VEMDMS-KPDLTAALRDIRAQYETIAAKNISEAEEWYKSKVSDLTQAANKNNDALRQAKQEMMEYRHQIQSYTCEIDALKGTNDSLMR
GFAP H-VELDVA-KPDLTAALKEIRTQYEAMASSNMHEAEEWYRSKFADLTDAAARNAELLRQAKHEANDYRRQLQSLTCDLESLRGTNESLER
Per  QQVEVEATVKPELTAALRDIRAQYESIAAKNLQEAEEWYKSKYADLSDAANRNHEALRQAKQEMNESRRQIQSLTCEVDGLRGTNEALLR
       QREFE
LamA --VEIDNG-KSRLADALQELRAQHEDQVEQYKKELEKTYSAKLDNARQSAERNSNLVGAAHEELQQSRIRIDSLSAQLSQLQKQLAAKEA
                    --EEE-----HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH---HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
                    L12     237-------------------------------- Coil 2 ----------------------------------

                    --------------------------- Coil 2 ---------------------404       Tail       >>>
                    HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH------------------EEEEE------
                    gabcdefghijkabcdefgabcdefgabcdefgabcdefgabcdefgabcdefgabcdefga
Vim  QMREMEENFAVEAANYQDTIGRLQDEIQNMKEEMARHLREYQDLLNVKMALDIEIATYRKLLEGEESRISLPLPNFSSLNLRETNLDSLP
Des  QMRELEDRFASEASGYQDNIARLEEEIRHLKDEMARHLREYQDLLNVKMALDVEIATYRKLLEGEESRINLPIQTYSALNFRETSPEQRG
GFAP QMREQEERHVREAASYQEALARLEEEGQSLKDEMARHLQEYQDLLNVKLALDIEIATYRKLLEGEENRITIPVQTFSNLQIRETSLDTKS
Per  QLRELEEQFALEAGGYQAGAARLEEELRQLKEEMARHLREYQELLNVKMALDIEIATYRKLLEGEESRISVPVHSFASLNIKTTVPEVEP
LamA KLRDLEDSLARERDTSRRLLAEKEREMAEMRARMQQQLDEYQELLDIKLALDMEIHAYRKLLEGEEERLRLSPSPTSQRSRGRASSHSSQ
                    HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH----EE--------EEEEEEE----E
                    L12     237----------------------------- Coil 2 --------381      Tail       >>>
```

Fig. 11. Primary structure analysis of the IF rod domain. Sequence alignment for the rod domains of human type 3 IF proteins including vimentin, desmin, GFAP and peripherin and type 5 IF protein - nuclear lamin A. Residues predicted by the NetSurfP algorithm [154] to be buried inside the structure are highlighted in yellow. Heptad assignment is indicated. The secondary structure prediction for vimentin and lamin A sequences by the Jpred3 algorithm [153] is shown on the top and the bottom respectively ('H' indicates α-helix, 'E' indicates β-sheet). Stutter inserts resulting in 11-residue repeats are highlighted with cyan. Proline residues are highlighted in gray. Positions corresponding to the residues conservative among all IF proteins are highlighted in green.

Based on the primary sequence analysis we were able to proceed with the engineering of the constructs suitable for crystallization. Fragments corresponding to several regions of the IF protein rod domain were designed. According to the part of IF protein to be resolved this work consisted of several directions:

1) Vimentin coil1A/L1/coil1B structure (constructs 1-6, 21-26, Table 5)

2) Vimentin coil2 structure (constructs 7-10, Table 5)

3) Vimentin L12 structure (constructs 11-20, 27-36, Table 5)

4) Coil1B - tetrameric organization of vimentin (constructs 37-42, Table 5)

5) Lamin A structure and its tetrameric organization (constructs 43-51, Table 6)
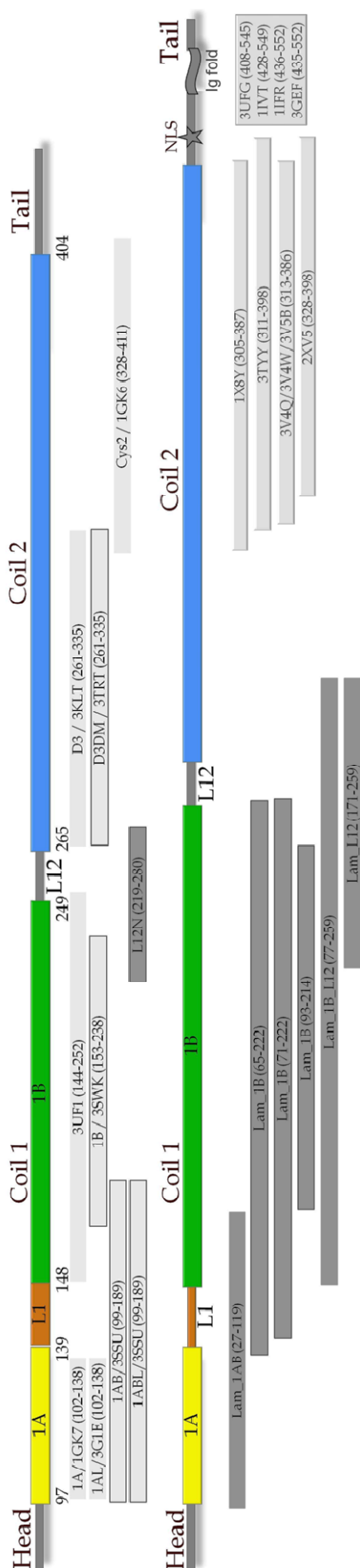
Fig. 12. Schematic representation of human vimentin and lamin A proteins with corresponding fragments used for crystallographic analysis. The head, tail, α-helical rod domain (including coil1A, linker L1, coil1B, linker L12 and coil 2) are indicated with corresponding amino acid numbers. White rectangles correspond to crystallographically solved structures (for references see Table 5). Outlined rectangles are solved in frames of this work. Dark grey rectangles correspond to the crystallized fragments, structures of which have not been solved yet. Upper drawing corresponds to vimentin and its fragments. Lower drawing corresponds to lamin A and its fragments.

Table 5. Vimentin fragments and crystals

Number	Vimentin fragment	vector	crystals	diffraction / structure (ID)
1	79-252		-	
2	99-189 (fragment *1AB*)		+ (1)	2.6 Å/ *3SSU*
3	99-189, SeMet		+	
4	99-189, Y117L (fragment *1ABL*)		+ (2)	2.45 Å/ *3S4R*
5	99-189, Y117L, SeMet		+ (3)	
6	99-189, E103C	pPEP-TEV (S-)	-	
7	261-335, L269C, C328A		-	
8	261-335, L265C, C328A		+ (4)	1.7 Å/ phasing
9	261-335, L265C, L269(MSE), C328(MSE), SeMet (fragment *D3st*)	pPEP-TEV(S-)	+ (5)	2.3 Å / *3TRT*
10	261-335, Y276C, L269M, C328M		-	
11	167-280, A280C		-	
12	174-291, Y291C		-	
13	167-276, Y276C		-	
14	153-265, L265C		-	
15	153-327		-	
16	224-330		-	
17	230-330		-	
18	220-330		-	
19	180-327		-	
20	180-313		-	
21	161-238		+ (6)	trimer
22	161-238, SeMet		+ (7)	trimer
23	161-243		+ (8)	trimer
24	161-243, SeMet		+ (9)	
25	153-238		+ (10)	
26	153-238, SeMet (fragment *1B*)	pETSUK	+ (11)	1.7 Å/ *3SWK*
27	219-276, F233M, Y276C		-	
28	212-276, F233M, Y276C		-	
29	219-280, F233M, Y276M, A280C		+ (12)	2 Å / phasing
30	219-280, F233(MSE), Y276(MSE), A280C, SeMet		+ (13)	2.3 Å/phasing
31	219-280, F233M, Y276M, A280L		-	
32	212-280, F233M, Y276M, A280C		+ (14)	
33	212-280, F233(MSE), Y276(MSE), A280C, SeMet		+ (14)	
34	219-292, Y276M, W290A, Y291M		-	
35	219-304, Y276M, W290A, Y291M, F295M		-	
36	202-276, F206M, F233M, Y276C	pETSUL	-	
37	144-283, N283C, Y276M			
38	144-272, N272C		not pur	
39	110-283, Y276M, N283C			
40	99-283, Y276M, N283C		-	
41	99-289, Y276M		not pur	
42	87-340, Y276M, C328M, L340C	pET15-NESG (His$_6$-SHM-)		

 "not pur" stays for not purified fragments; "phasing" – for not yet solved structures; column "vector": in the brackets an additional residue or fused construct occurring in the final fragment is indicated; column "crystals": number in brackets indicates the number of the corresponding crystal picture shown after the Table 6.

Table 6. Lamin A fragments and crystals

Number	Lamin A fragment	vector	crystals	diffraction / structure
43	27-119, L85M, L104M		+ (15)	
44	65-222, L85M		+ (16)	2.7 Å/ phasing
45	71-222, L85M		+ (17)	
46	71-222, L85(MSE), SeMet		+ (18)	2.9 Å/ phasing
47	77-214, L85M	pPEP-TEV (S-)	not pur	
48	93-214		+ (19)	2.4 Å/ phasing
49	93-214, SeMet		+ (20)	2.35Å/phasing
50	77-259, L85M, Y259C		+ (21)	
51	171-259, Y259C		+ (22)	

"not pur" stays for not purified fragments; "phasing" – for not yet solved structures; column "vector": in the brackets an additional residue or fused construct occurring in the final fragment is indicated; column "crystals": number in brackets indicates the number of the corresponding crystal picture shown below.
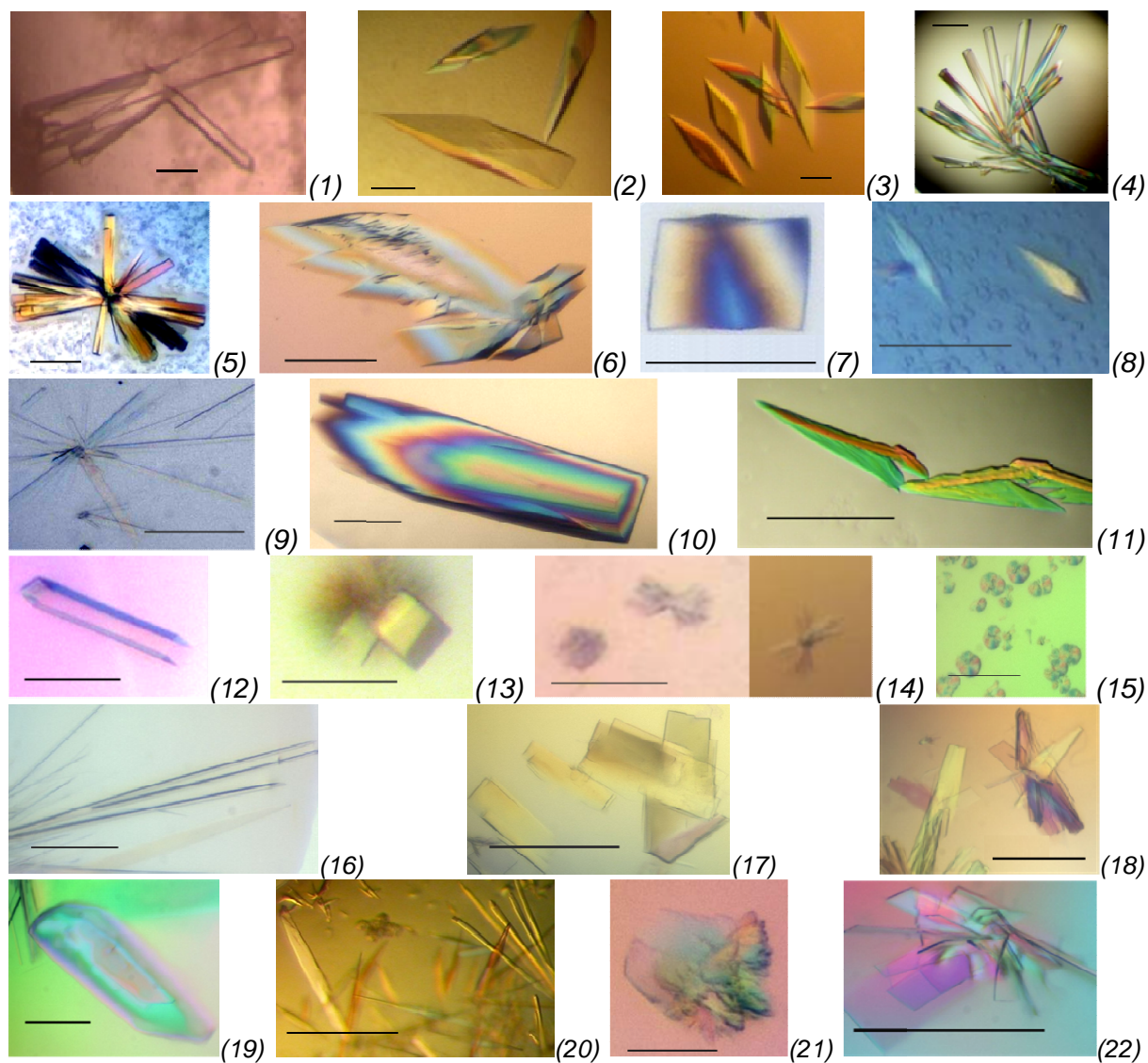
Scale bar - 100 µm

Table 7. Crystallographic statistics

Vimentin fragment		*1AB* (98-189)		*1ABL* (98-189)		*1B* (153-238)	D3st (261-335)
		Native	**SeMet**	**Native**	**SeMet**	**SeMet**	**SeMet**
Crystallization conditions		0.04M Ca acetate, 0.1M MES pH 6, 6% isopropanol, 4˚C	0.06M Ca acetate, 0.1M MES pH5.5 6% isopropanol, 4˚C	0.45 M Mg acetate, 0.1M Na cacodylate pH 6.5, 20% MPD, 4˚C	0.35M Li_2SO_4, 0.1 M MES pH 6.0, 35% MPD, 4˚C	0.2M NH_4 acetate, 0.1 M Bis-Tris pH 6.5, 25% w/v PEG 3350, 20˚C	2M $(NH_4)_2SO_4$, 0.1M Tris pH 8.5, 20˚C
Data collection							
Wavelength (Å)		0.9763	0.9786 (peak)	1.0000	0.9798 (inflection)	0.9790 (peak)	0.9793 (peak)
Space group		$I2_12_12_1$		$R32$		$P2_1$	$I222$
Unit cell parameters (Å)		45.6, 83.5, 118.6	45.7, 82.8, 119.2	65.3, 65.3, 377.3	65.0, 65.0, 378.5	40.2, 54.5, 44.1	53.4, 75.5, 86.9
Resolution (Å)		24.3-**2.6** (2.64-2.6)	50-**3.0** (3.05-3.0)	48-**2.45**(2.49-2.45)	50-**2.8** (2.85-2.8)	34.3-**1.7** (1.8-1.7)	28-**2.3** (2.34-2.30)
Rsym ** **(%)**		0.072 (0.6)	0.096 (0.277)	0.071(0.420)	0.090 (0.295)	0.068 (0.778)	0.088 (0.352)
I/σ		8.9 (1.1)	12.3 (5.2)	14.5 (1.7)	7.8 (2)	13.26 (2.6)	10.7 (2)
Unique reflections		7261	4745*	11209	6867*	38320*	14768 (695)*
Completeness (%)		99.9 (100)	99.6 (96.9)*	93.7 (55.9)	82.1 (34.9)*	98.9 (96.1)*	97.8 (87.1)*
Multiplicity		5.5 (5.6)	8.1 (6.5)*	10.9 (8.5)	4.8 (3.2)*	4.9 (4.6)*	3.4 (2.5)*
Mosaicity (°)		0.55	0.6	0.85	0.25	0.44	0.8
Wilson B-factor ($Å^2$)		70.7		44.0		30.6	37.6
Refinement							
Ordered residues	**chain A**	144-189		102-189		153-238	261-335
	chain B	149-187		99-189		153-237	262-334
R value***	**R** $_{work}$	0.24		0.29		0.25	0.22
	R $_{free}$	0.28		0.32		0.28	0.30
R.m.s. deviations	**Bond lengths (Å)**	0.007		0.011		0.004	0.008
	Bond angles (°)	1.1		1.22		0.64	1.034
Number of atoms	**total**	730		1521		1694	1260
	ligand/ water	0		16		251	68
Average B-factor	**total**	93.8		78.3		38.8	40.1
	ligand/ water	-		80.1		39.9	58.1
PDB ID		*3SSU*		*3S4R*		*3SWK*	*3TRT*

* Reflections F+ and F- were treated as different; data in parentheses are for the highest resolution bin;

** $R_{sym} = \Sigma_h\Sigma_i|I_{ih}-<I_h>|/\Sigma_h\Sigma_i\,I_{ih}$, where I_{ih} is the i-th observation of I_h and $<I_h>$ is the mean of N_h measurements;

*** $R = \Sigma|F_{obs}-F_{calc}|/\Sigma F_{obs}$, with the 'free' value calculated with 5% of reflections.

4.1 Vimentin coil1

The main results described in this part are published in: **Atomic structure of the vimentin central alpha-helical domain and its implications for intermediate filament assembly.** <u>*Chernyatina A. A.,*</u> *Nicolet S., Aebi U., Herrmann H., Strelkov S. V. (2012) Proc Natl Acad Sci USA 109: 13620-13625.*

4.1.1 Crystal structure of the *1AB* fragment

To obtain experimental evidence towards the L1 structure, we have crystallized vimentin fragment *1AB* containing residues 99-189 and therefore encompassing the coil1A, linker L1 and the first half of coil1B (Fig. 13 A). The calculated Matthews volume of the crystals assuming two 11.1 kDa chains per asymmetric unit is 2.55 A^3/Da corresponding to 52 % solvent, which implies a rather loose packing. The structure was phased experimentally by single-wavelength anomalous scattering collected from the crystals of SeMet-containing protein (see Table 7 for diffraction data statistics). The final structure refined against native data at 2.6Å resolution reveals a parallel CC dimer (Fig. 13 A). Surprisingly, a clear electron density is only observed starting from residues 144 and 149 in chains A and B, respectively. Moreover, SDS-PAGE analysis of the crystals confirmed that the fragment was intact (mass 11.1kDa). In addition, first few traceable residues in each chain are mobile, as evident from increased B-factors. Both the coil1A and most of the linker L1 are therefore completely disordered in this crystal structure. Tyr150 is the first residue forming the hydrophobic core of coil1B, which has a fairly regular left-handed CC structure. The hydrophobic core is formed by the residues in *a* and *d* positions of the heptads in a perfect agreement with the sequence-based prediction (Fig. 11). Notably, the still visible part of the linker L1 in chain A (residues 144 to 149) is also α-helical (Fig. 13A, orange colouring). The crystal packing arrangement is such that the 1B parts alone are sufficient to form an interconnected lattice, albeit having large voids that should be accommodating the poorly ordered 1A parts.
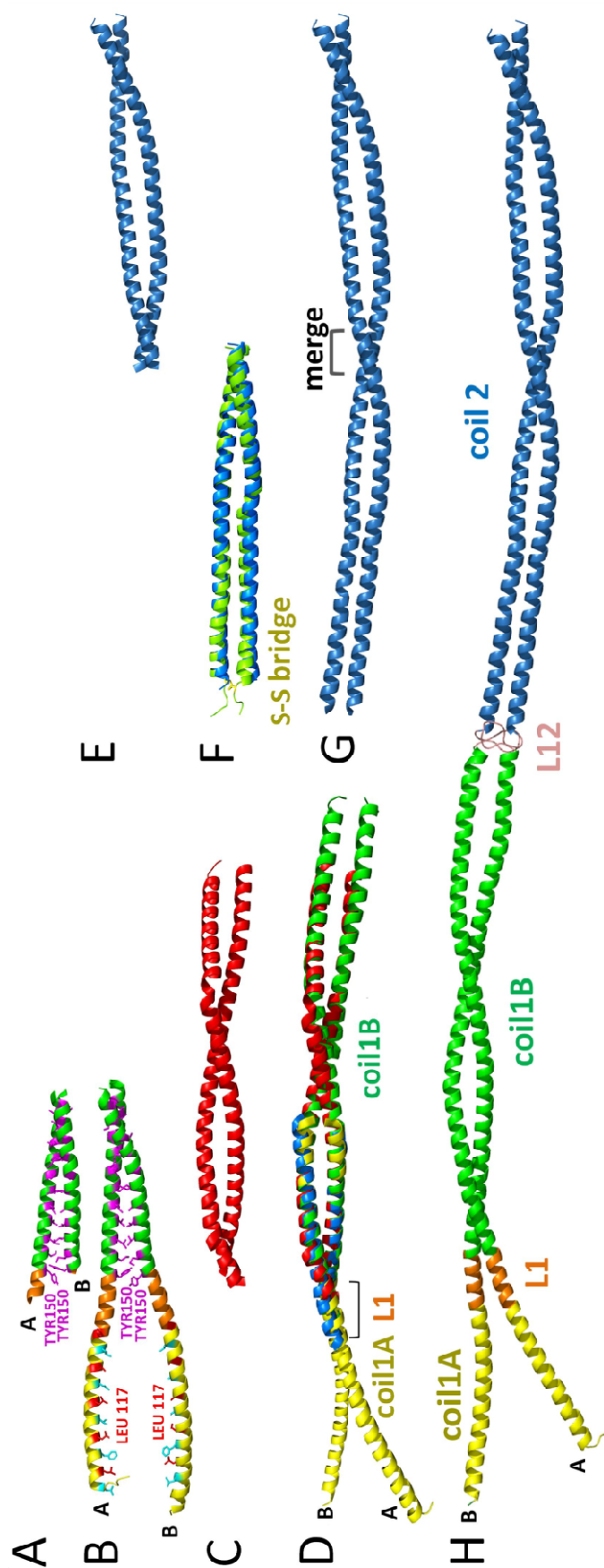
*Fig. 13. Crystal structures of vimentin fragments. **A.** Ribbon diagram of the 1AB structure. The side chains of residues in heptad positions a and d are shown in magenta. With the exception of residues 144 to 149 (shown in orange) in chain A, the rest of the linker L1 and coil1A is disordered in the crystals. Here and further all figures are drawn using the program Pymol (177). **B.** Ribbon diagram of the 1ABL structure showing two chains A and B present in the crystallographic asymmetric unit. The 1A, L1 and 1B parts are coloured yellow, orange and green, respectively. Within coil1A, the residues in predicted heptad positions a and d (see Fig. 11) are shown in cyan and red, respectively. Within coil1B, the residues in both heptad positions a and d are shown in magenta. The mutated Leu117 residue is labelled. **C.** Ribbon diagram of the 1B structure showing two chains A and B present in the crystallographic asymmetric unit. **D.** Least-squares structural superposition of the 1AB, 1ABL and 1B fragments (128) as well as the PDB entry 3UF1 (126). The four structures are shown in blue, yellow, red and green, respectively. **E.** Ribbon diagram of the crystal structure of the fragment Cys2 that corresponds to the second part of coil2 (123). F. Least-squares structural superposition of the D3 modelled dimer (125) and D3st (127) fragment, shown in blue and green, respectively. Disulfide link of the D3st structure is shown in yellow. **G.** The ribbon representation of vimentin coil2 model. **H.** 3D structure of vimentin dimer. The coil1 structure (with coil1A, L1 and coil1B shown in yellow, orange and green, respectively) is based on the superposition of the crystal structures shown in A-D. Similarly, the coil2 structure (blue) is a superposition of D3, D3st and Cys2 fragment structures (125, 126, 127). The linker L12 for which no crystallographic data are available yet is shown schematically in pink. In the full dimer, the rod domain is flanked by the flexible head and tail domains, for which no crystallographic information is available yet.*

4.1.2 Crystal structure of the modified 1AB fragment

In order to stabilize the coil1A part of the fragment *1AB*, a specific mutation Y117L was introduced (Fig. 13 B, fragment *1ABL*). This mutation had been found before to drastically increase melting temperature of the dimer of a coil1A fragment (from 32 °C to 71 °C at the protein concentration 1mg/ml) [124]. This discovery led to the crystallization of the mutant *1AL* protein in a dimeric conformation (Fig. 12, PDB ID *3G1E* [124]) while the WT coil1A being a very weak dimer in the solution was crystallized as a monomer (denoted as *1A* in the Fig. 12; PDB ID *1GK7* [122]). Therefore, to stabilize the *1AB* dimer, the same mutation *Y117L* was introduced and the mutated protein *1ABL* was successfully crystallized. The crystals revealed the whole fragment structure, corresponding to the first half of coil 1. Surprisingly, both chains were found to contain exclusively α-helical structure that continues from the coil1A part into L1 and coil1B. This led to the fact that while the 1B part formed the predicted dimer, the two 1A α-helices splayed apart upstream of Tyr 150 (heptad position *d*) due to the rigid α-helical structure of L1 (Fig. 13 B) resulting in a 25 Å separation between the C_α-atoms of residues Leu138 in chains A and B [128].

Crystallographic packing of the *1ABL* fragments

An explanation of how the single mutation Y117L could stabilise the coil1A within the crystal is based on the formation of specific contacts between the fragments *1ABL*. As it is described above, the arrangement of the 1A parts within the *1ABL* structure is radically different from a parallel CC dimer, even though such a dimer should be very stable and indeed was found in the crystals of the *1A* fragment with the Y117L mutation [124]. However, in the crystals of the fragment 1ABL a tetrameric left-handed CC assembly involving 1A helices from crystal symmetry-related molecules is found instead. This tetramer includes chains *A* and *B'* (being a symmetry equivalent of *B*) running in one direction, as well as chains *A'* and *B''* (being further symmetry equivalents of *A* and *B* respectively) running in the opposite direction (Fig. 14 A). A common hydrophobic core involving residues from all the four chains stabilizes the tetramer. This stabilization is due to three different types of hydrophobic contacts: between two parallel chains *A* and *B'*, between antiparallel chains *A* and *A'* as well as between antiparallel chains *B'* and *B''* (Fig. 14 A). Unexpectedly, these contacts involve residues not only in *a* and *d* positions of the predicted heptads but also additional residues in positions *e* and *g*, as schematically represented in Fig. 14 B.

Interestingly, altogether the interactions of the 1A and 1B parts lead to the formation of a dodecamer featuring six 1B dimers and three 1A tetramers (Fig. 15). The dodecamer has the point symmetry 32, which corresponds to the space group R32 of the crystals. Here the antiparallel overlap of the 1A parts (such as, *e.g.,* represented by chains *A* and *A'* in Fig. 14 A) measures ~7nm. This does not correspond to any of the antiparallel dimer-dimer alignments (A_{11}, A_{12} or A_{22}) found in the filaments [138, 177]; in particular the A_{11} contact

features the complete overlap of coil1 segments and measures ~23 nm [138]. Moreover, the tetramer of the 1A parts was only observed for the fragment with the mutated Leu117 residue, while the native sequence is likely to be refractive to such tetramer formation due to the clash of the Tyr117 residues of antiparallel chains (see Fig. 14 A). At the same time, this residue is fully compatible with a parallel dimer formation, even though it negatively affects its stability.

*Fig. 14. Crystal contacts of the fragment 1ABL lead to the formation of an unnatural tetramer. **A**. An antiparallel CC tetramer formed by the 1A parts in the 1ABL crystals. Chain A' is a crystal symmetry equivalent of chain A. Chains B' and B'' are crystal symmetry equivalents of chain B. For clarity, other symmetry-related chains forming the CC within the 1B part are not shown here (see Fig. 14 B). The residues involved in the hydrophobic core of the tetramer are shown in red, cyan and gray according to the pairing indicated in panel E. In particular, the 'knobs-to-knobs' interface of the antiparallel chains A and A' is formed by five Leu residues on each side (all in positions d) and is shown with their side chains in red. The interface of antiparallel chains B' and B' is of a classical (antiparallel) 'knobs-into-holes' type; it involves residues in the a and e positions of either chain. **B**. Alignment of the sequences of the four 1A parts forming the antiparallel tetramer. The heptad assignment is indicated below each chain. The opposing pairs of residues forming the hydrophobic interactions ('knobs-into-holes') between chains A and B' and chains A' and B'' (parallel, one residue register shift) are shown in cyan. Similar interactions formed by chains B and B' (antiparallel) are highlighted with gray. The 'knob-to-knob' interface of antiparallel chains A and A' is highlighted with red.*

*Fig. 15. Dodecamers found in the 1ABL crystals, shown in stereo. The view is rotated by 90°
compared to Fig. 14 A. Coil11A, L1 and the first half of coil11B part are shown in yellow, orange
and green, respectively.*

4.1.3 Crystal structure of the *1B* fragment corresponding to the coil 1B

For the vimentin coil1B, an α-helical structure is predicted between residues 148 and 249 (Fig. 11). The expected heptad pattern is also reliably confirmed by the NetSurfP algorithm with the exception of a single discontinuity, corresponding to a stutter, which appears to be present close to the C-terminus of the coil1B (Fig. 11). This is a new feature that was not detected by the heptad-filtering algorithms before. We have prepared a new series of vimentin coil1B fragments with varying N- and C-termini, yielding diffracting crystals for three fragments 161-238, 161-243 and 153-238, respectively (denoted as 21-23, Table 5). All three structures could be phased using SeMet SAD. However, the crystal structures of the two former fragments revealed CC trimers rather than the dimers expected for the full-length protein. Apparently, the first eleven residues of coil1B are important for the formation of the CC with the correct number of α-helices. The last fragment including residues 153-238 (denoted *1B* in the Fig. 12) formed the expected dimers. Its crystal structure at 1.7 Å resolution reveals a regular left-handed CC structure all along its length (Fig. 13 C) in a full agreement with the predicted heptad pattern (Fig. 11) [128].

SeMet-derivatization of *1B* fragment

For unknown reasons crystals of the native protein did not diffract. However, we were able to obtain the structure of *1B* fragment using its SeMet derivative. As it was confirmed by electrospray ionization mass spectrometry (ESI-MS) analysis (kindly performed by Prof. Rozenski), the substitution of Met by SeMet was not complete (Fig. 16). Deconvoluted mass spectrum obtained from ESI-MS analysis of purified vimentin fragment *B1* in 10 mM Tris-HCl pH 8, 12 mM NaCl buffer gave a set of peaks. The peaks at 10373, 10395, 10416 Da and others correspond to the fragment *1B* mass (MW_{calc} = 10374 Da) and same fragments with one, two or more Na^+ ions attached (change in MWcalc = 22 Da), respectively. Deconvoluted mass spectrum obtained from ESI-MS analysis of purified vimentin SeMet fragment *B1* in the same buffer gave a new set of peaks with shifts in masses. There are three sites for substitution in each monomer. The population of the protein where the substitution is full corresponds to the peaks at 10514, 10535, 10557, *etc.* Da (MWcalc = 10515 Da and 22 Da change of mass is due to interaction with Na+ ions). The population with only two S substituted by Se corresponds to the mass 10467 Da (MWcalc = 10468 Da). The population with one S substituted by Se corresponds to the peak at 10420 Da (MWcalc = 10421 Da). The population with non-substituted S corresponds to the peaks 10373, 10395 Da (MWcalc = 10374 Da and 10396 Da for 1B interacting with Na+). The efficiency of Se substitution was estimated as the ratio of deconvolution peak areas with ImageJ program (rsbweb.nih.gov/ij) showed about 66% of fully Se-substituted protein in the final solution.

Despite the fact that there were more Se atoms than S atoms, the electron density map of corresponding region did not show big peaks characteristic for Se. This can be explained either by low occupancy or flexibility of the anomalous scatterer (SeMet may have several alternative conformations, among which we see only major one). In the final structure (PDB ID *3SWK*) normal Met residues were modelled in.
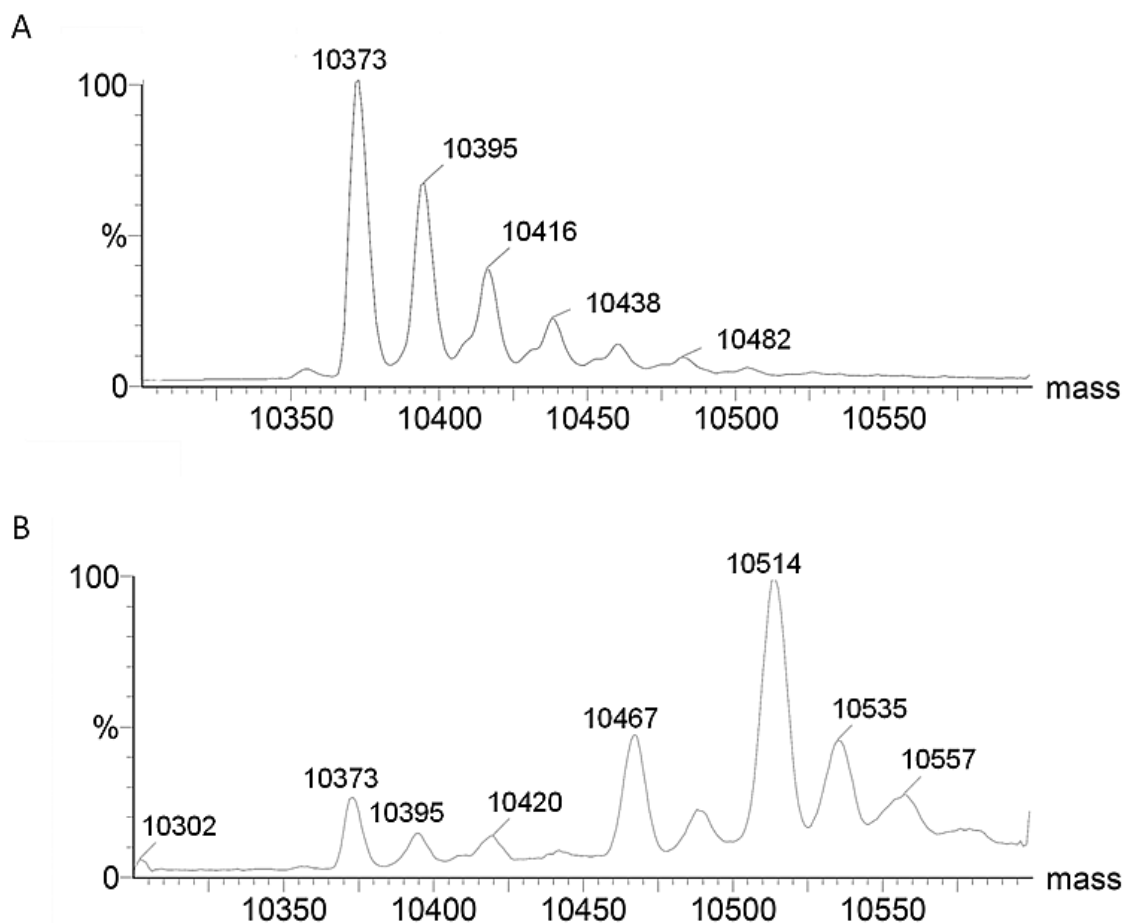
Fig. 16. Results of the mass-spectrometry analysis of the 1B protein sample. **A.** *Deconvoluted mass spectrum obtained from ESI-MS analysis of purified vimentin fragment B1 in Tris-NaCl buffer.* **B.** *Deconvoluted mass spectrum obtained from ESI-MS analysis of purified vimentin fragment B1 with sulphur (S) substituted by selenium (Se) in Tris-NaCl buffer.*

4.1.4 Complete vimentin coil 1 structure

Independently of our work, Aziz *et al.* recently solved the structure of a longer vimentin *1B* fragment forming tetramers in crystals (denoted as *1Btet*, Fig. 12, PDB entry *3UF1* [126]). The structure of this fragment can be superimposed well with our *1B* fragment. In addition it revealed possible contacts leading to the tetramer formation as a second step of the filament assembly. Here it should be noted that the *1Btet* structure confirms the primary sequence analysis and the NetSurfP prediction of a parallel α-helical bundle at the very C-terminus of coil1B, resulting from a stutter at the position 238 (Fig. 11). Finally, the *1Btet* structure reveals an α-helix starting already at the residue 144, *i.e.* for a few residues of linker L1, which supports our observation of an α-helical linker L1 in the *1ABL* fragment.

Taken together, the four crystal structures provide a redundant view of the coil1. The superposition of all these structures produces a reasonable match (Fig. 17 A), confirmed by an analysis of the CC radius and the pitch (Fig. 17 B).
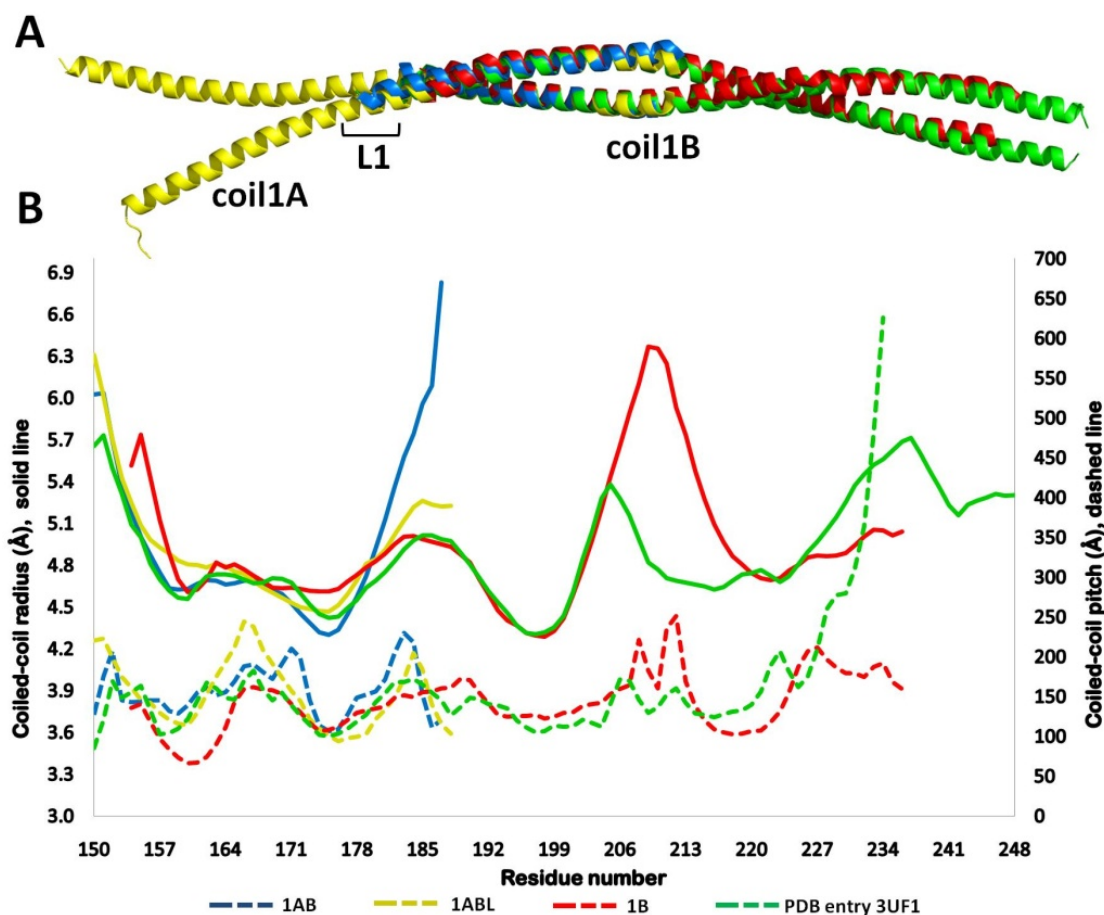
*Fig. 17. Complete structure of the vimentin coil1 based on the crystal structures. **A.** Least-squares structural superposition of the 1AB, 1ABL and 1B fragments as well as the PDB entry 3UF1. The four structures are shown in blue, yellow, red and green, respectively. **B.** Plots of the CC radius (solid lines) and the pitch (dash) as a function of residue number calculated using the program Twister [178]. The data for the 1ABL, 1AB, 1B and 3UF1 structures are shown in the same colours as in panel A.*

4.1.5 Linker L1 in IF proteins

Our crystallographic data on the *1ABL* fragment reveal that the vimentin linker L1 can accommodate a fully α-helical conformation [128]. This unexpected observation deviates from secondary structure predictions that indicated an interruption of the α-helical structure at L1. In addition, unlike human vimentin, desmin, GFAP and peripherin all contain a proline residue within L1 (Fig. 11). The Pro residue in the same position is also found in vimentin of electric ray *Torpedo californica* [179]. When embedded in an α-helical structure, Pro residues result in a destabilizing kink [180]. In lamin A and B, however, the residual assignment allows coil1 to be a complete CC: if the L1 had an α-helical conformation, it would include a stutter insert [115, 176]. Similarly, in both keratin chain types and also in NF proteins, the shift corresponds to an insert of six residues, *i.e.* a double stammer (Fig. 18). It has been shown experimentally that such a situation can be tolerated within a continuous CC if its handedness is locally switched from the left to the right [178]. So far, an α-helix formation within the linker L1 region should be therefore best considered a possibility rather than a rule. Importantly, there must be interplay between the coil1A and linker L1 dynamics. Indeed, if linker L1 stays fully α-helical, this will not allow the 1A parts to zip into a CC within the same full-length vimentin dimer (see Fig. 13 H). At the dimer stage, a tendency of the α-helix to extend into the linker L1 may overweigh a weak propensity of coil1A to zip into (intradimeric) CC, which are mutually exclusive possibilities.

```
<<<                      coil 1A                        L1          coil 1B    >>>
        abcdefgabcdefgabcdefgabcdefgabcdefgabcde_____abcdefgabcdefgabcdefg
K18   42  IQNEKETMQSLNDRLASYLDRVRSLETENRRLESKIREHLEKKG-PQVRDWSHYFKIIEDLRAQIFANTVD 145
K8    87  RTQEKEQIKTLNNKFASFIDKVRFLEQQNKMLETKWSLLQQQKT-ARSNMDNMFESYINNLRRQLETLGQE 156
Vim  100  RTNEKVELQELNDRFANYIDKVRFLEQQNKILLAELEQLKGQ---GKSRLGDLYEEEMRELRRQVDQLTND 167
NFL   85  RTQEKAQLQDLNDRFASFIERVHELEQQNKVLEAELLVLRQKHS-EPSRFRALYEQEIRDLRLAAEDATNE 154
LaA   28  RLQEKEDLQELNDRLAVYIDRVRSLETENAGLRLRITESEEVVSREVSGIKAAYEAELGDARKTLDSVAKE  98
```

Fig. 18. Primary structure analysis of the linker L1 with adjacent parts of coil 1A and coil 1B of IF rod domain among the five IF classes. Sequence alignment is shown for the rod domains of human keratins K1 and K8, vimentin, neurofilament and nuclear lamin A respectively. Residues that are predicted by the NetSurfP algorithm [154] to be buried inside the structure are highlighted in yellow. Heptad assignment is indicated above the sequences; core positions are highlighted with yellow. Proline residues are highlighted with gray.

4.2 Vimentin coil2

The main results described in this part are published in: **Stabilization of vimentin coil2 fragment via an engineered disulfide.** <u>Chernyatina A. A.</u>, *Strelkov S. V., (2012) J Struct Biol 177: 46-53.*

4.2.1 Summary

Contrary to a prediction that there must be a linker within the coil2 [181], it was later suggested that the coil2 is purely α-helical [182]. More precisely, a possibility for a parallel bundle structure of contiguous α-helices based on a hendecad (11-residue) repeat was suggested for the part corresponding to both segment 2A and linker L2 [182]. The only structural evidence was obtained in the form of a crystal structure of the *D3* fragment of human vimentin (residues 261 to 335) which corresponds to the first half of the coil2 domain [125]. However, this fragment stayed monomeric in solution but formed non-native tetramers within the crystals (Fig. 19 C). Formation of such tetramers is not compatible with the full-length dimer due to topological reasons. At that point, the native dimeric state of this region was modelled by excising two parallel chains out of the tetramer and bringing their N-terminal halves closer so that their core residues would come into contact. However, a direct evidence of the parallel bundle at the N- end of the coil2 and the absence of the linker L2 in the native dimer state was still missing.

To overcome this problem, a Cys residue was brought in the core positions at the N-end of the *D3* structure for dimer stabilization; and the mutant *D3st* bearing L265C mutation was crystallized. The structure of the *D3st* mutant revealed a dimer, confirming previous cross-link experiments [138] and EPR analysis [148] as well as fitting well with the proposed model of the *D3* dimer [125] (Fig. 13 F, 19 A). The structure of *D3st* fragment contains residues 261-335 in the chain A and 262-334 in the chain B, *i.e.,* practically the complete fragment. The final coordinates and experimental structure factors have been deposited in the PDB with accession code *3TRT* [127]. An important conclusion of these studies is that coil2 forms a contiguous CC structure, *i.e.* the linker L2 does not exist.

Previously, a structure of the second half of the coil2 was resolved crystallographically. Two fragments corresponding to this region called *Cys2* (Fig. 13 E) and a fusion fragment *Z2B* demonstrated classical left-handed CCs [123], with a number of salt bridges. Such extremely strong CCs are characteristic for this highly conservative region among all IF proteins. In addition, C*ys2* fragment demonstrates that the stutter (residues 351-354 insert into the heptad periodicity) is fully compatible with the CC geometry. A superposition of the structures of the fragment *Cys2* [123], the modeled dimer of *D3* fragment [125] and the fragment *D3st* [127] resulted in a full model of vimentin coil2 (Fig. 13, E-G).

4.2.2 Design of a stabilizing disulfide within a coiled coil

Our goal was to induce a dimeric state of the previously studied vimentin *D3* fragment that corresponds to the N-terminal half of the coil2 - by introducing *Cys* residues in specific core positions. We focussed on creating a disulfide near the N-terminus of the structure, as this part failed to dimerize properly with the original *D3* fragment [125]. Assuming the traditional heptad-based possibility and a short left-handed segment 2A, its first two core positions would be residues Leu265 - *d* and Leu269 - *a* (see Fig.11). Alternatively, we have examined the tetrameric parallel bundle structure observed for the *D3* fragment [125] (Fig. 19C).

A position of a side chain in a CC structure with respect to its axis is defined by the so-called Crick angle. This angle is the main parameter that determines the suitability of a particular (core) position for disulfide introduction. In heptad-based CCs, the Crick angle is about +20˚ for an *a* position (the less favourable position for disulfide introduction according to [183]) and about -35˚ for a *d* position (the most favourable position) [178]. In the *D3* structure, the first core position (position *h* of the hendecad repeat) is Leu265 with α_{Crick}=-19.9˚ when calculated for the two parallel chains of the tetramer (Fig.20 A). This Crick angle value is relatively close to the one for a *d* position of a heptad-based CC. This suggests that if the continuous α-helices seen in the *D3* tetramer are brought together in a proper dimer, the residues in position 265 would be suitable for a disulfide link (Fig. 220 A). The second core position in *D3* is Leu269 (position *a* of the hendecad, α_{Crick}=20.9˚). Altogether, for the N-terminus of coil2, both the traditional heptad assignment and the structure derived from the *D3* tetramer consistently suggest that Leu265 and Leu269 are involved in the core and that either of them could be potentially suitable for the disulfide introduction.

Correspondingly, we have prepared two modified variants of the original *D3* fragment that contained mutations to Cys in the positions 265 and 269, respectively. Additionally, in each case the naturally occurring residue Cys328 (*f* position of the heptad repeat exposed on the outside of the CC) was replaced by an Ala to exclude the possibility of additional disulfide bridge formation. For both modified fragments, a complete disulfide oxidation occurred spontaneously in the course of purification (in Tris-HCl buffer pH 8 without reducing agents) as evident from the SDS-PAGE (Fig. 19 B).

4.2.3 Crystal structure of *D3* fragment stabilized as a dimer

Of the two prepared *D3* modifications, L265C/C328A and L269C/C328A, only the former one could be crystallized even after extensive screening. Possibly, this fact may indicate that a disulfide link in a position *a* (residue 269) which was suggested to be suboptimal for the CC stability [183] results in a poorly folded, flexible structure refractive to crystal formation. As explained in the Methods section, to phase the crystal structure of *D3*

with disulfide at the residue 265, SeMet-based phasing was necessary. Since the sequence of the *D3* fragment does not include any Met residues, two such residues were additionally introduced. The first methionine was inserted in place of Leu269 in a core position *a*, as natural methionines are commonly found in the CC core positions and stabilize it; moreover some other IF proteins contain methionines in a homologous position (e.g., human keratin 8). The second methionine replaced Cys328. The resulting triple mutant *D3* (L265C/L269M/C328M) used for structure determination will be henceforth referred to as *D3st* (stabilized).
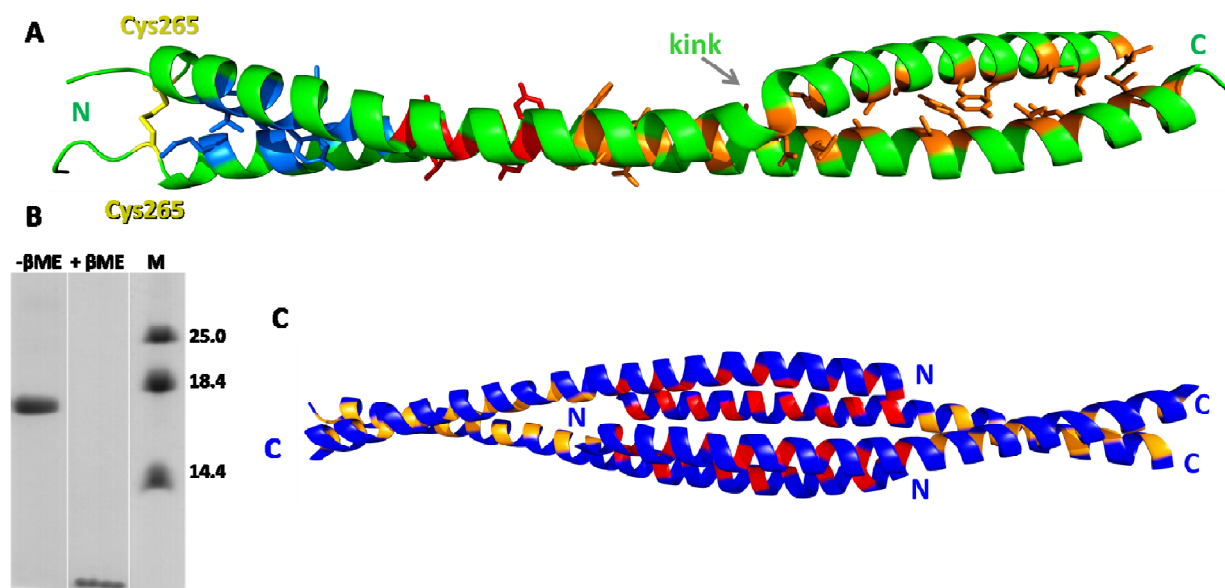
*Fig. 19. The D3st dimer structure compared with the D3 tetramer. Here and elsewhere molecular graphics was done with Pymol (DeLano, 2002) [184]. **A.** Ribbon diagram of D3st dimer structure. Disulfide link is shown in yellow. Side chains forming the hydrophobic core are shown in blue and red (N-terminal part with 15 and 11-residue repeats respectively) and orange (C-terminal part with heptad repeats). **B.** SDS-PAGE of D3 (L265C/C328A) sample without reducing agents (first lane) and in the presence of β-mercaptoethanol (second lane). The result for the D3 (L269C/C328A) mutant was identical. Molecular weight marker (kDa) is shown on the right. **C.** Ribbon diagram of the D3 structure. Core positions within the N-terminal half, featuring a hendecad repeat and forming a tetrameric bundle, are shown in red. Core positions within the C-terminal part forming a left-handed CCs are in orange.*

The structure of the *D3st* fragment reveals an expected CC dimer with the α-helices starting immediately after residue Cys265 (Fig. 19A). As seen from the CC geometry analysis with program Twister [178] (Fig. 20 A, B), the C-terminal part of the structure (starting at about residue 300) is a classical left-handed CC, as expected. This structure superimposes well with the corresponding region previously resolved in the *D3* fragment structure and, correspondingly, both the CC radius and the pitch correlate well for the two regions (Fig. 20 B). Unexpectedly, however, there is a 'kink' interrupting the normal α-helical structure for residues Asn303, Arg304, Asn305, in one of the two chains of the *D3st* structure (Fig. 19A, 25). These residues are more 'stretched' than they should be within the normal α-helix. As a result, four characteristic main-chain hydrogen bonds, *i* to *i+4,* are missing. The

main-chain angles for both Asn residues are within few degrees of φ=-105˚, ψ=+17˚. For Arg304 and all surrounding residues they are close to φ=-60˚, ψ=-45˚ which are the typical values for an α-helix. While this kink appears to be a crystallographic artefact caused by crystal contacts (Fig. 25), we note that Asn residues are among the least favoured in the α-helical conformation [185] and should therefore cause a local weakening of the structure. However, the introduced disulfide link near the N-terminus of *D3st* fulfills its designed goal, as the two α-helices stay in contact with each other all along the fragment's length, mediated by the hydrophobic core.

A

Res	D3			D3st	
265 L	h	-19.9		l	16.8	
266 T	i	81.9		m	110.7	
267 A	j	-168.3		n	-158.3	
268 A	k	-81.8		o	-66.3	
269 L	a	20.9		a	27.7	
270 R	b	116.7		b	121.9	
271 D	c	-140.9		c	-143.5	
272 V	d	-48.6		d	-48.7	
273 R	e	55.4		e	45.4	
274 Q	f	152.5	11	f	138.2	
275 Q	g	-106.6		g	-127.5	
276 Y	h	-11.2		h	-36.3	15
277 E	i	90.7		i	65.5	
278 S	j	-163.8		j	158.5	
279 V	k	-82.7		k	-108.1	
280 A	a	21.2		l	-14.1	
281 A	b	119.2		m	84.4	
282 K	c	-145.0		n	-178.0	
283 N	d	-45.8		o	-81.7	
284 L	e	56.3		a	16.5	
285 Q	f	147.2	11	b	113.2	
286 E	g	-112.3		c	-146.3	
287 A	h	-16.9		d	-49.8	
288 E	i	83.2		e	47.8	
289 E	j	176.5		f	147.5	11
290 W	k	-87.7		g	-113.6	
291 Y	a	10.1		h	-18.9	
292 K	b	108.4		i	81.7	
293 S	c	-154.4		j	-171.7	
294 K	d	-59.5		k	-77.4	
295 F	e	42.6		a	27.1	
296 A	f	142.6	11	b	126.1	
297 D	g	-117.5		c	-131.6	
298 L	h	-19.6		d	-34.4	7
299 S	i	82.6		e	69.9	
300 E	j	177.1		f	165.3	
301 A	k	-79.2		g	-90.0	
302 A	a	23.6		a	17.3	
303 N	b	122.8		b	106.6	
304 R	c	-137.1		c	-134.4	
305 N	d	-38.0		d	...	

B

Fig. 20. Coiled-coil geometry in D3 tetramer and D3st dimer. For the analysis of D3, only two parallel chains A and B were included. **A.** *Crick angles for residues 265 to 305 in both structures as calculated with program Twister [178]. The remaining C-terminal residues follow typical left-handed coiled-coil geometry in both structures. Core positions and the corresponding Crick angles are highlighted in bold. Square brackets indicate the heptad (orange), hendecad (red) and quindecad (blue) repeats.* **B.** *Plots of the coiled-coil radius and pitch as a function of residue number. The data for the D3 and D3st structures are shown in blue and green, respectively. The radius and pitch are shown with solid and dashed lines, respectively.*

4.2.4 N-terminal coiled-coil geometry and disulfide link conformation

The middle part of the *D3st* structure (residues from ~280 to ~300) has a parallel-bundle geometry seen already with the *D3* fragment. At the same time, the very N-terminal part of *D3st* was found to have a right-handed geometry not seen previously for the *D3* fragment. The right-handed twist becomes pronounced upstream of residue 280, and the CC pitch drops below 100Å near the N-terminus (Cys265 residue, Fig. 20 B). In line with this, the analysis of the Crick angles reveals minor differences in the hydrophobic core structure

between the *D3* and *D3st* fragments (Fig. 20 A). In the latter, the span between residues 269 and 302 (both in *a* positions) contains three hendecads (*abcdefghijk*, with core residues underlined), while in the *D3* structure there is a single quindecad (*abcdefghijklmno*), followed by one hendecad and one heptad. Each of these patterns corresponds to a different local supercoiling (see Fig. 19 A, 20 B), with a quindecad repeat in particular matching somewhat a right-handed geometry [116].

Average CC radius, *i.e.* one half of the distance between the α-helical axes, for residues 265 to 300 is 4.97Å. This is not significantly different from the C-terminal part with its average radius of 4.65Å being close to the typical values seen for left-handed CC dimers. These observations indicate that the CC core packing of *D3st* dimer is reasonably tight. At the same time, the core positions within the N-terminal half feature many residues deviating from the most favoured set (Leu, Val, Ile and Met) [186]. In particular, there are two alanines (Ala287 and Ala302) which are too small for optimal core packing and known to destabilize the core [187], and three aromatic residues (Tyr276, Tyr291 and Phe295) which are relatively bulky. These factors are likely to contribute to the fact that the *D3* fragment failed to dimerise.

Within the *D3* structure, the Leu265 was in a position *h* of the hendecad-based CC. While the theoretical value of Crick angle for this position is close to zero [116], the observed value in the *D3* tetramer was $\alpha_{Crick}=-19.9°$ (Fig. 20 A). However, in line with the differences between the *D3* and *D3st* dimers just described, in the latter structure, Cys265 has $\alpha_{Crick}=16.8°$, *i.e.* quite similar to an *a* position. While this core geometry was suggested to be suboptimal for disulfide formation [183], certain features of the N-terminal conformation of *D3st* chains appear to compensate for this.

The main-chain conformation of the two Cys265 residues ($\varphi=-102°$, $\psi=+17°$ and $\varphi=+81°$, $\psi=+1°$, respectively) deviates significantly from α-helical geometry. The few N-terminal residues preceding Cys265 are also non-α-helical (Fig. 21 C, D). As a consequence, the C_β atoms of Cys265 are located slightly 'downstream' N-C direction of the α-helix in comparison with the corresponding C_α atoms (Fig. 21 D). This is opposite to the normal α-helical geometry such as observed for the Cys33 residue in the structure with PDB ID *1KD9* [188] (Fig. 21 F). Furthermore, in *D3st* structure the disulfide link is located 'upstream' of the C_α positions and the C, C_α, C_β and S_γ atoms are located roughly in one plane (Fig. 21 D). In the *1KD9* structure the link is 'downstream' of the C_α positions, with the N, C_α, C_β and S_γ atoms of Cys33 being in one plane (Fig. 21 F). Interestingly, Zhou et al reported 'upstream' orientation for disulfides modelled in both *a* and *d* positions of GCN4 leucine zipper (Fig. 6 in [183]). Finally, detailed analysis of the disulfide link geometry (Table 8) suggests against a strained conformation of either Cys33 link in the structure with PDB ID *1KD9* or Cys265 link in *D3st*.
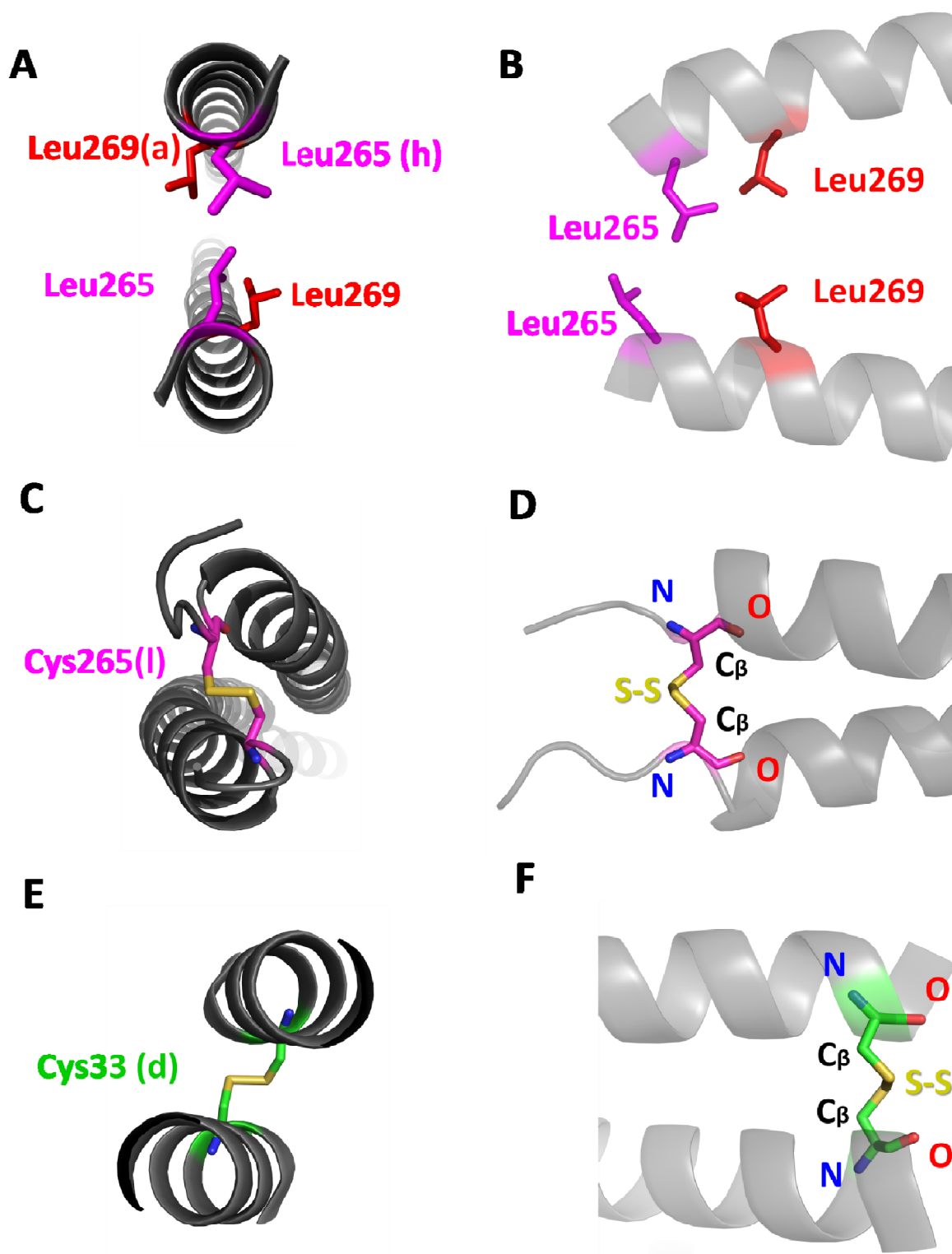
*Fig. 21. Geometry of the core side chains and disulfide links in dimeric CC. Left panels show the view along the N->C direction of the amino acid chains. Right panels show a perpendicular view with the N->C direction from the left to the right. **A, B.** The first two core residues Leu265 (magenta) and Leu269 (red) within the two parallel chains of D3 tetramer. The two other chains are not shown. **C, D.** Disulfide link between residues Cys265 in the D3st structure. **E, F.** Disulfide link between residues Cys33 (d position) in the 1KD9 structure [188].*

Table 8. Disulfide link geometry

Structure	*1KD9*	*D3st*
Residue	Cys33	Cys265
Crick angle (^{o}) / core position	-36.4 / d	16.8 / 'a-like'
Distance between C_α atoms	6.2	5.6
Distance between C_β atoms*	4.0	3.7
Angle C_β-S-S **	104.9	104.0
Dihedral angle at the disulfide ***	-95.7	-85.3

Ideal distance between C_β of two Cys to form a disulphide bond is 4.08 Å [189].

*** Average value for both symmetric residues is given; ideal value is 103˚ [183].*

**** Ideal value +/-90˚ [183].*

In summary, while earlier analyses suggested a strictly parallel bundle structure for the first 35 residues of the coil2, the *D3st* structure reveals a right-handed geometry near its N-terminus. We argue that such structure is justified by a slightly different hydrophobic core structure than predicted before, and that the observed '*a*-layer-like' geometry of Cys265 residues appears to be mainly the consequence of that. As a caveat, we note that disulfide formation could in turn also affect the chain conformation and thus the CC twists in its immediate vicinity. Finally, local crystal contacts could also play some role in modulating the CC twist, as the N-terminal parts (residues 261 to 290) of both chains A and B of the dimer come into a relatively close but yet predominantly hydrophilic contact with their symmetry equivalents due to a crystallographic two-fold axis.

4.2.5 Disulfide as a stabilising tool: advantages and drawbacks

This work demonstrates the use of a rationally designed disulfide introduction to a relatively short CC fragment that fails to dimerise on its own. Previously, a fusion with the 31-residue GCN4 leucine zipper to promote the correct assembly of a 28-residue fragment corresponding to the C-terminal part of the vimentin coil2 was used [190]. The GCN4 zipper is known to form a very stable, parallel CC dimer [9]. The idea behind the fusion was that the CC structure would continue smoothly from the GCN4 part into the target fragment. Therefore, a design of such a fusion (just like the introduction of the disulfide link described here) requires a reasonable starting prediction of the CC structure of the target fragment (core position assignment) at the fusion point. In the simplest case of a heptad-containing fragment, the heptad pattern should retain the phase across the fusion with the GCN4 zipper. A further limitation of this fusion approach is that the GCN4 part adds additional length of about 45Å to the already elongated CC fragment, which may have a negative effect on crystallizability. In addition, while a point mutation to Cys can be rapidly introduced into an existing expression construct using standard methods, preparation of a GCN4 fusion requires a considerably more elaborate molecular cloning.

Crystal structure of the *D3st* fragment reveals, unexpectedly, that the linked Cys265 residues are arranged roughly in the same way as the residues in an *a*-layer of a classical

left-handed CC. In correlation with theoretical predictions [183], such arrangement demands a structural distortion, as the Cys residues deviate from an α-helical conformation. Near the N-terminus of the fragment, such distortion can however be easily tolerated. In comparison, a disulfide in a *d* position of a regular left-handed CC seen in the crystal structure with PDB ID *1KD9*, does not distort it. In line with this, introduction of a disulfide in the cytokeratin K8/K18 heterodimer by mutating a *d* position (equivalent to vimenin residue 372) within the coil2 to Cys in both chains was shown to have no effect on the keratin filament assembly [33]. Furthermore, cross-linking by disulfides in core positions was used to stabilize hydrogels formed from synthetic CC proteins [191]. Interestingly, the long CC dimer of αα-tropomyosin containing a Cys in an *a* position appeared to assemble in the same fashion in both, oxidised and reduced state of the disulfide [192].

In conclusion, introduction of disulfide links *via* engineered cysteines is a powerful tool to stabilize dimeric CCs and hence promote the correct assembly. Critical to this method is the ability to correctly place Cys into a core position, on hand of a sequence-based prediction or other information. The best strategy is to create disulfides possibly close to either termini of the CC. In such cases cysteines may be introduced in either '*a*-like' or '*d*-like' core positions, *i.e.* comparable to those observed in *a* or *d* positions of heptad-based CCs, respectively. Indeed, our crystallographic analysis indicates that a disulfide in '*a*-like' position requires a local distortion of α-helical geometry. However, near the chain terminus, such a distortion can easily be accommodated without affecting the rest of the structure.

4.3 The linker L12 structure

The only remaining structurally unsolved part of the vimentin rod domain is the linker L12. Our goal was to get the crystallographic structure of this linker. For this, we designed a number of fragments, corresponding to a part of the coil 1B, linker L1 and a part of the coil 2 (Table 5, fragments 27-36, Fig. 22). We were able to obtain diffracting crystals of a short fragment (residues 219-280) carrying a stabilizing mutation A280C in addition to mutations F233M, Y276M (used for SeMet derivatisation). The best crystal diffracted to 2 Å resolution. To overcome the phase problem we needed HAs incorporation in the crystal. For this, we tried SeMet derivatization. A single obtained crystal (50*50*50 µm^3 size, Fig. 22) of SeMet derivative protein diffracted to 2.3 Å resolution. However, poor strength of the anomalous signal did not allow us to solve the structure. Attempts to reproduce this crystal or to grow bigger crystals were not successful until now. In addition, MIR technique could not be used as the space groups of the native and derivative crystals differ. Structure solution of the L12 linker is still in progress.
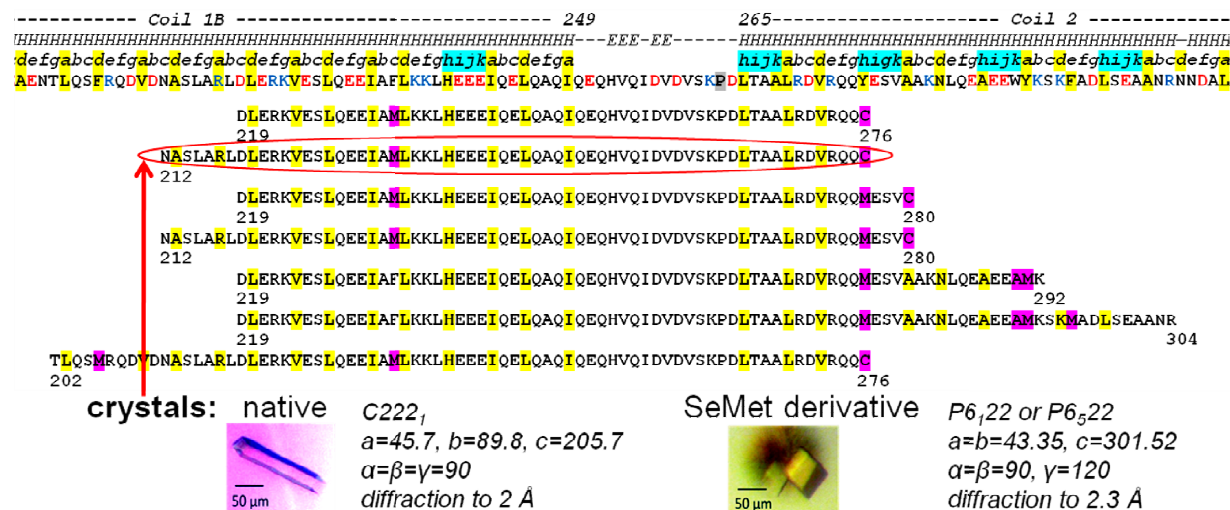
Fig. 22. Schematic representation of the designed fragments for the vimentin linker L12 structure determination. Primary structure analysis of the IF rod domain corresponding to the end of the coil 1, L12 and the beginning of the coil2 is shown on the top. The same way as in the Fig. 11, residues that are predicted by the NetSurfP algorithm [154] to be buried inside the structure are highlighted in yellow. Heptad assignment is indicated. The secondary structure prediction for vimentin sequence by the Jpred3 algorithm [153] is shown on the top ('H' indicates α-helix, 'E' indicates β-sheet). Stutter inserts resulting in 11-residue repeats are highlighted with cyan. Proline residues are highlighted with gray. Fragments corresponding to the linker L12 region, tried for crystallization are shown in the middle. The fragment (residues 219-280) shown in red circle gave crystals in the native and SeMet derivative form, shown at the bottom of the figure with a corresponding crystallographic statistics.

As SeMet derivatization did not give useful phase information, a synthetic peptide (residues 219-280) was produced in the laboratory of Dr. Rackwitz (Heidelberg, Germany). This fragment carried one Cys point mutation (highlighted in violet) and the single phenylalanine (highlighted in red) was replaced by p-Br-L-Phe during the synthesis:

DLERKVESLQEEIA**F**LKKLHEEEIQELQAQIQEQHVQIDVDVSKPDLTAALRDVRQQYESV**C**

Br could be used later for heavy-atom phasing and Cys should stabilize CC dimer at the C-end. So far, no diffracting crystals of the synthetic peptide could be obtained.

4.4 Kinks

In three of the recently obtained vimentin fragment structures [126-128] we noted α-helical kinks, *i.e.* sharp bends of the α-helix with a loss of several main-chain hydrogen bonds pertinent to α-helical geometry. Such a kink apparently helps to accommodate local distortions, needed to either form a proper CC or due to crystallographic contacts. It is known that an α-helix is rather rigid structure, stabilised by hydrogen bonds between the O atom of a residue with the number i and the hydrogen attached to the N atom of the i+4 residue. However, CC structure requires bending of the α-helices. This already creates a certain tension in α-helices. If local forces are applied to CCs, the tension increases. Therefore, occurring kinks may be useful in IF proteins to compensate for local small deviations upon higher oligomer and filament assembly.

4.4.1 Kink in *1B* fragment

In the *1B* structure [128], a proper CC formation is preserved by the following mechanism: one α-helix is kinked and twists around the other one which remains almost straight (Fig. 23 A, B). In the region of the kink the normal i to i+4 hydrogen bonding pattern typical for the α-helical backbone is interrupted. Four H-bonds are missing: Phe206(O) – Val210(N), Arg207 (O) - Asp211(N), Gln208 (O)-Asn212 (N) and Asp209 (O)-Ala213 (N).
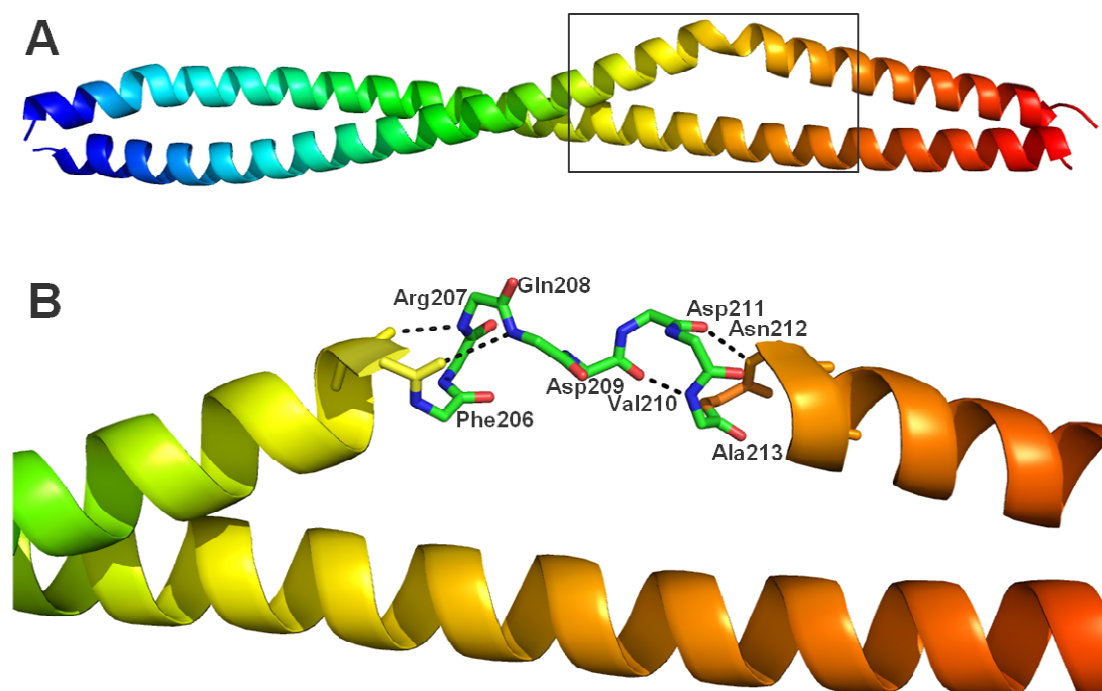

*Fig. 23. 'Kink' within one of the α-helical chains of the fragment 1B. **A**. Ribbon diagram of 1B fragment in rainbow colouring (from N terminus in blue to C terminus in red). **B**. Close-up of the kink region from the A plane, marked in rectangle.*

4.4.2 Kink in *1Btet* fragment

Interestingly, in *1Btet* fragment corresponding to coil1B of vimentin rod domain [126], the kink occurs very close to the site of the kink in *1B* fragment described before (Fig. 24 A, B). It represents a bend of one α-helix in each dimer missing two H-bonds: Leu203(O) – Arg207(N), Gln204(O) – Gln208(N).
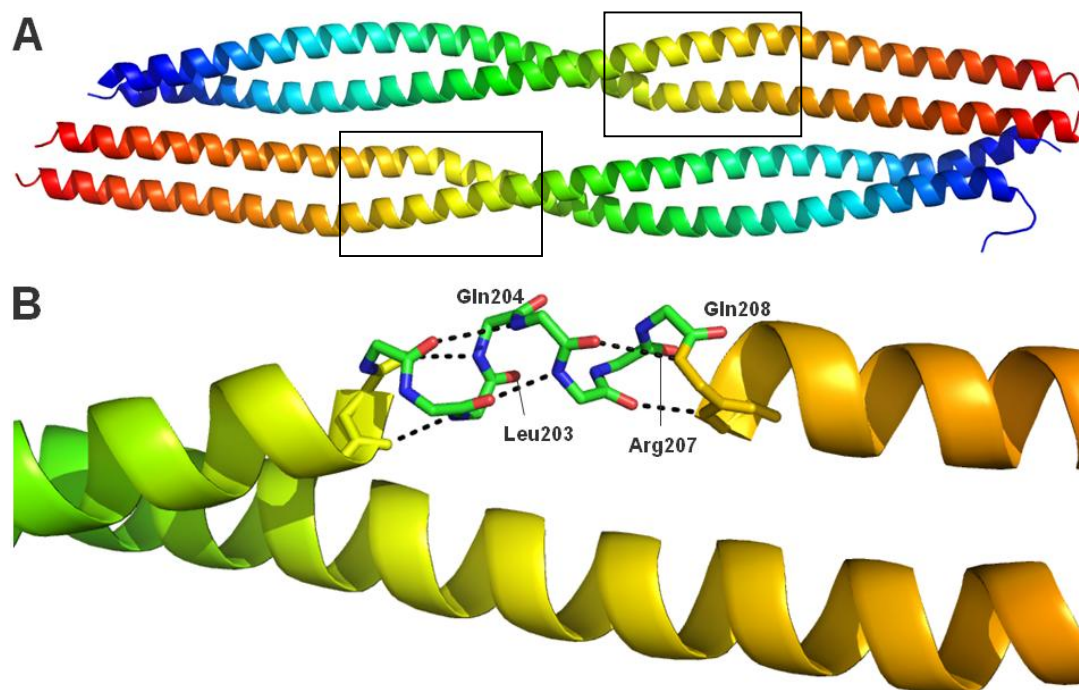
*Fig. 24. 'Kink' within the two α-helical chains of the tetramer 1Btet (PDB ID: 3UF1) [126]. **A.** Ribbon diagram of 1Btet structure: two CC in an antiparallel arrangement in rainbow colouring (from the N-terminus in blue to the C-terminus in red). Kink regions are shown in rectangles. **B.** A zoom-in to the kink region of one of the dimers.*

4.4.3 Kink in *D3st* fragment

Interestingly, the kink region separates a classical left-handed CC from a parallel untwisted CC region in the middle of *D3st* structure [127]. The kink links the middle (parallel) and the C-terminal (left-handed twist) regions of the structure (Fig. 25, A-C). It represents a short loop knocked out from the α-helical pattern of the B chain of *D3st* fragment. Within the kink there are four broken hydrogen bonds: Ala301(O) – Asn305(HN), Ala302(O) - Asn306(HN), Asn303(O)-Asp307(HN), Arg304(O) – Ala308(HN) (Fig. 25 C). In addition, one hydrogen bond not compatible with α-helical geometry occurs, namely Asn303(O) – Asn306(HN). These residues are more 'stretched' than they should be within a normal α-helix. As a result, four characteristic i to i+4 main-chain hydrogen bonds are missing. The main-chain angles for both Asn303, 305 and 306 residues are within few degrees of $\varphi=-105°$, $\psi=+17°$. For Arg304 and all surrounding residues they are close to $\varphi=-60°$, $\psi=-45°$ which are typical values for an α-helix. This kink appears to be caused by crystal contacts (Fig. 25 A). A C-terminus of one dimer points to another symmetry-related dimer, causing formation of the kink (Fig. 25 B).
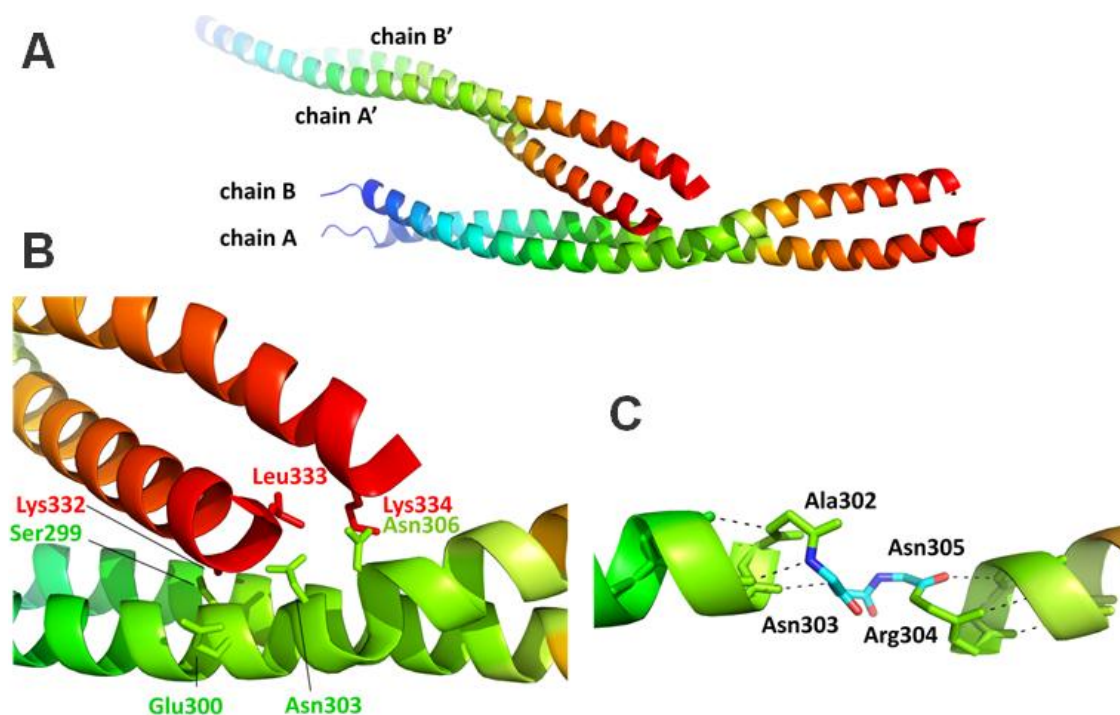
Fig. 25. 'Kink' within one of the α-helical chains of D3st. **A***. Crystal lattice arrangement likely causing formation of the kink. Ribbon diagram of D3st and its symmetry related neighbour in rainbow colouring (from N terminus in blue to C terminus in red). C-termini of a symmetry-related dimer (A'B') clash with chain B of the dimer AB. A further consequence of this clash is that the Cα atoms of residues Ala302 ("a" position) come unusually close to each other, as evident from the CC radius plot (Fig. 20 B).* **B***. Close-up on the crystal contact.* **C***. Detailed view on the kink region including residues Asn303, Arg304 and Asn305 of the chain B. These residues are more 'stretched' than they would be within α-helix.*

Notably, all three examples of the kinks within vimentin CC rod occur when the hydrophobic residues pattern changes. Consecutively, CC changes its pitch upon a transition from the classical left-handed CC to the parallel untwisted CC. The sequences of the two regions, showing three kinks are: **L**Q**Δ**SF**R**QD**Δ**VDN**A** (residues 203-213, **Δ** stands for the kinks in *1Btet* and *1B* fragments, respectively) and **AA**NR**N**ND**A** (residues 301-308, kink-forming residues are underlined).

Both regions are rich in the residues with high (Ala, Leu and Arg, shown in red) and very low (Val, Asn, Asp, shown in blue) α-helical propensity defined by Pace and Scholtz [185]. Apparently, such a pattern of amino acid residues allows accommodation of both an α-helical as well as a kink conformations.

4.5 Vimentin tetramer and higher assembly stages

4.5.1 Modelling of vimentin tetramer based on crystallographic data

With the new crystallographic data, an atomic structure of the vimentin rod becomes essentially complete. This allowed us to construct an atomic model of the vimentin tetramer (Fig. 26 C) and also to obtain important insights into the mechanism of subsequent IF assembly.

As it was recently confirmed by spin-labelling and EPMR studies [138, 177], a formation of the A_{11}-type solution tetramers brings the residues Lys188 or Glu191 of the two dimers in register (Fig. 26 A, B). Interestingly, the crystal structure of the vimentin fragment corresponding to the complete coil1B (PDB entry *3UF1*) contains two antiparallel dimers per asymmetric unit centred at the position 191. Therefore we used this crystal structure as a template to build the full tetramer - save the non-α-helical 'head' and 'tail' domains (Fig. 26 C). The overlapping crystal structures were merged together with subsequent energy minimization in Refmac [165] that removed bad van der Waals contacts.

Within our tetramer model, the α-helical linkers L1 of one dimer are roughly aligned with the C-terminal part of coil1B of the antiparallel dimer. Our analysis showed that the leading force for the tetramer formation is a distinct complementarity in the charge distribution on adjacent surfaces. As the most striking feature, the last coil1A residue Lys139 and the linker L1 (140-GQGKSRLG-147, positively charged) of one dimer are found aligned with the C-terminus of coil1B of the antiparallel dimer (238-HEEEIQELQAQI-249, negatively charged) [128].
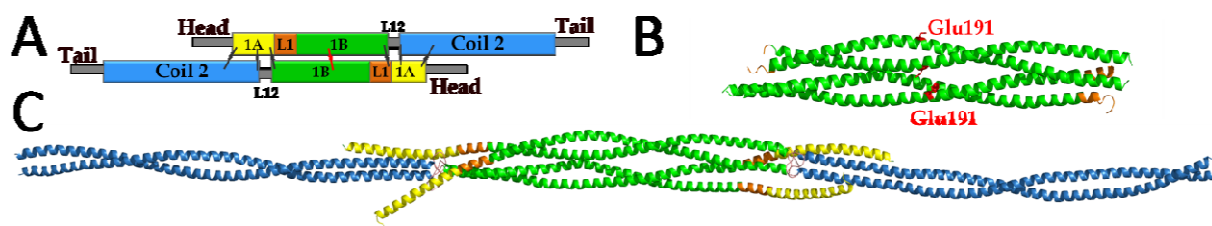
*Fig. 26. Modelling of the vimentin tetramer. **A.** Scheme of an A_{11} alignment mode according to Steinert et al. [138] and Hess et al. [139]. Black and red sticks indicate intermolecular cross-links of Lys residues with sulpho-DST: 143-235, 139-235, 188-188, 120-262, 139-236, 104-282, where the tetramer is centred at the residue 188 [138]. The same alignment mode A_{11} was confirmed by SDSL EPR. The interactions were postulated on the basis of EPR spectra broadening of the frozen vimentin mutants labelled at corresponding positions: 188-197, 188-193 and 191-191 in the centre of the tetramer [139]. **B.** Ribbon diagram of the 1Btet structure showing an antiparallel tetramer present in the crystallographic asymmetric unit. The tetramer is in A_{11} alignment mode, centred at the position 191, in a good agreement with the results of the experiments described in panel A. **C.** Ribbon representation of the rod domain tetrameric structure of vimentin. The modelling is based on the structure of the dimer shown in Fig. 13 H and the structure 1Btet [126] as a core for the tetramer contacts.*

Furthermore, it was suggested earlier that coil1A may serve as a dynamic 'switch' with the functionally significant property to alternate between a monomeric helix and a CC dimer depending on the particular IF assembly stage [124, 193]. Our modelling shows that once A_{11}-type tetramers stack laterally, a CC formation by the 1A helices coming from two different dimers becomes feasible (Fig. 27 C). We hypothesize that this situation, which we here denote as the 'cross-coil', indeed could take place at the ULF stage and/or in mature filaments. The 'cross-coil' formation, just like an intradimeric CC, would correspond to the proximity of spin-labelled residues in *a* and *d* positions within coil1A of full-length vimentin, as suggested from the broadening of the corresponding EPMR spectra [144] and shown in the crystal structure *1AL* [124] (for the superposition of the 'cross-coil' and *1AL* structure see Fig. 27 B). This possibility may explain the lack of longitudinal assembly of the *Y117L* mutant, since the formation of 'cross-coils' should work as a topological hurdle.
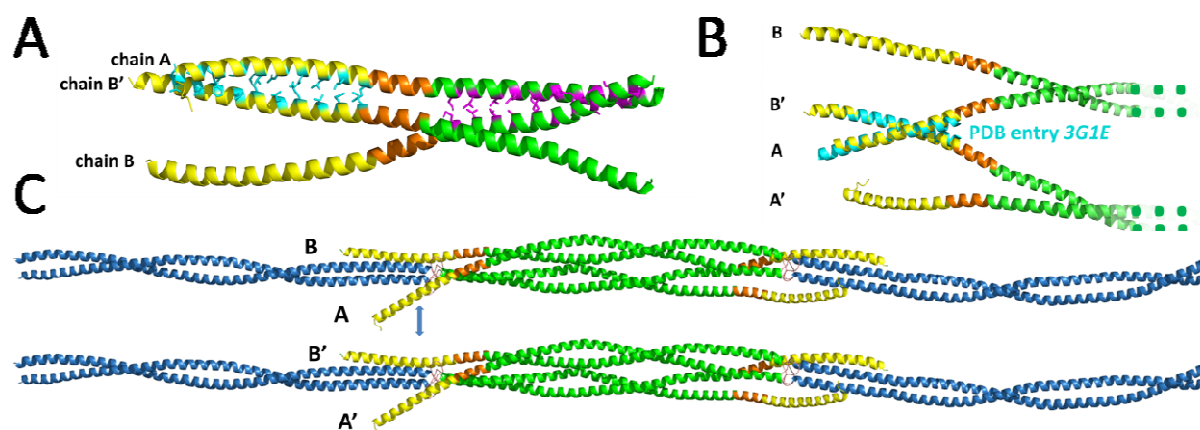
Fig. 27. Putative mechanism of the vimentin lateral assembly. ***A.*** *Ribbon diagram of the 1ABL structure showing chains A and B as well as a crystal symmetry-related copy (B') of the latter. Chains B and B' make contacts with chain A within coil1B and coil1A parts, respectively. The residues that are involved in the interaction of the coil1A parts of the chains A and B' are shown in cyan (the same colouring is used in Fig.13 B); these correspond to heptad positions a and e in the chain A and g and d in the chain B'. Correspondingly, there is a register shift of one residue between parallel chains A and B'. Within coil1B, the residues in both heptad positions a and d are shown in magenta.* ***B.*** *The structure of the 1ABL fragment allowing two dimers to form inter-dimeric CC 1A, a 'cross-coil', well fitting with 1AL dimeric structure [124], shown in cyan. The colouring of the 1ABL structure is the same as used in Fig. 13 B. Possible involvement of parallel coil1A segments in higher lateral assembly (octamers and beyond). If two tetramers (panel C) are laterally aligned, the dimers that run parallel to each other (AB and A'B') may interact via a 'cross-coil' formation of the coil1A segments.*

4.5.2 Solution SAXS of the heavy atom labelled vimentin

While being a diffraction-based technique like crystallography, SAXS is capable of providing 3D structural information at the nanometre level. Unlike crystallography SAXS can provide structural information on flexible and polydisperse systems. As opposed to NMR technique, SAXS is not limited by the size of a macromolecule. In general, SAXS provides accurate information about the shape of a macromolecule under various solution conditions, which could be readily derived from buffer subtracted scattering pattern (radially averaged diffraction intensity *versus* the scattering angle).

To understand the alignment of the dimers within the tetramers and ULFs as well as to estimate the distances within the rod domain (the length of L12) we are planning to do a recently revamped variant of SAXS experiments with HA labelling (HA-SAXS). HA-SAXS provides the pairwise distance information between HA labels, as it was demonstrated in few studies [194]. The HA signal is extracted from the difference scattering pattern of the labelled and unlabeled proteins. In 1980 it was estimated that this is possible if the overall amount of electrons of HAs is >0.004 of that of the protein [195, 196]. In our case, the full-length vimentin contains ~29000 electrons. Consequently, we need at least two Hg atoms (each contains 80 electrons) to be attached to the protein to give a HA signal on the SAXS curve. This can be achieved using such components as tetrakis-(acetoxy)-mercury-methane or mercarbide marker (each containing four Hg atoms).

In addition, it was shown that a 65% aqueous sucrose solution can be used to match the average electron density of the protein (contrast matching). If HA labels are used, only the SAXS signal from the HAs is collected, while the protein itself becomes 'invisible' [194]. With contrast matching, even single HA labels may be used regardless to the size of the protein. However, the presence of sucrose could potentially influence the assembly of vimentin.

The success of the SAXS experiments relies on the outcome of the following steps:

- construction of the desired mutations to introduce Cys
- purification of the corresponding proteins
- *in vitro* assembly and assessment of the filaments with electron microscopy
- identification of the appropriate compound for HA labelling
- analysis of the effect of sucrose on vimentin assembly
- SAXS measurements at the synchrotron

Design of vimentin mutants for HA-SAXS experiments

Three mutants each bearing Cys residues in a *d* core position of a CC were designed (Fig. 28). Two clusters at the positions 171 within the tetramer should confirm the existing alignment of the dimers within the tetramer. The same is true for the position 227. In low molarity buffer, vimentin is present as tetramers [121]. If we can measure HA-SAXS for such labelled tetramers, we can extract the pair-wise distances between positions 171-171, 227-227, 283-283 as well as 328-328. These distances should allow us to establish the exact structure of the tetramer, as well as to calculate the length of the linker L12. The three sites for mutagenesis (171, 227, 283) are chosen at the *d* positions according to the NetSurfP algorithm prediction [154]. In addition, residues 171 and 283 have been shown to be core positions using EPR experiments [139, 148]. The distance between two HAs attached to the core residues will be about several Å, which is much smaller than the target measured distance. Therefore with a good approximation the few Å distance can be neglected, and two close HAs will behave as a cluster, providing stronger HA signal.
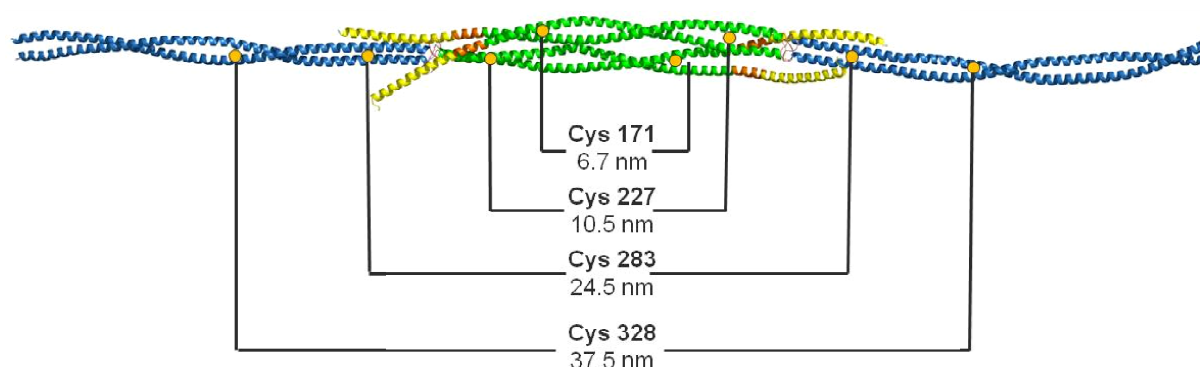
Fig. 28. Schematic representation of the designed mutations for the HA-SAXS experiments. Each vimentin mutant (Cys171, Cys227, Cys283) contains one Cys per monomer in the d position. The WT vimentin has one Cys328 in the f position. The arrangement of the mutated sites should lead to the formation of the HA clusters within the dimer (shown as orange circlets), two clusters per each tetramer. The distances between the HA clusters, measured in PyMol [184] using the tetramer model [128], are indicated.

Assessment of vimentin assembly *in vitro* using EM

The four proteins: WT full-length vimentin together with the three mutants (Cys171, Cys227, Cys283) were expressed in *E. coli* and purified under denaturing conditions. Renatured proteins were assembled in a standard assembly buffer for 1h at 37˚C. Thereafter, EM images of the assembled proteins were obtained. While the assembly of the WT protein along with the two mutants (Cys171, Cys227) showed normal filaments, the third mutation (Cys283) affected the assembly process and only ULFs could be detected (Fig. 29).

This finding along with the previously detected mutations causing similar effect [140, 197] may not only help to investigate the vimentin assembly in future but possibly points at new specific interactions necessary for the proper assembly of IFs, discussed in the next chapter.
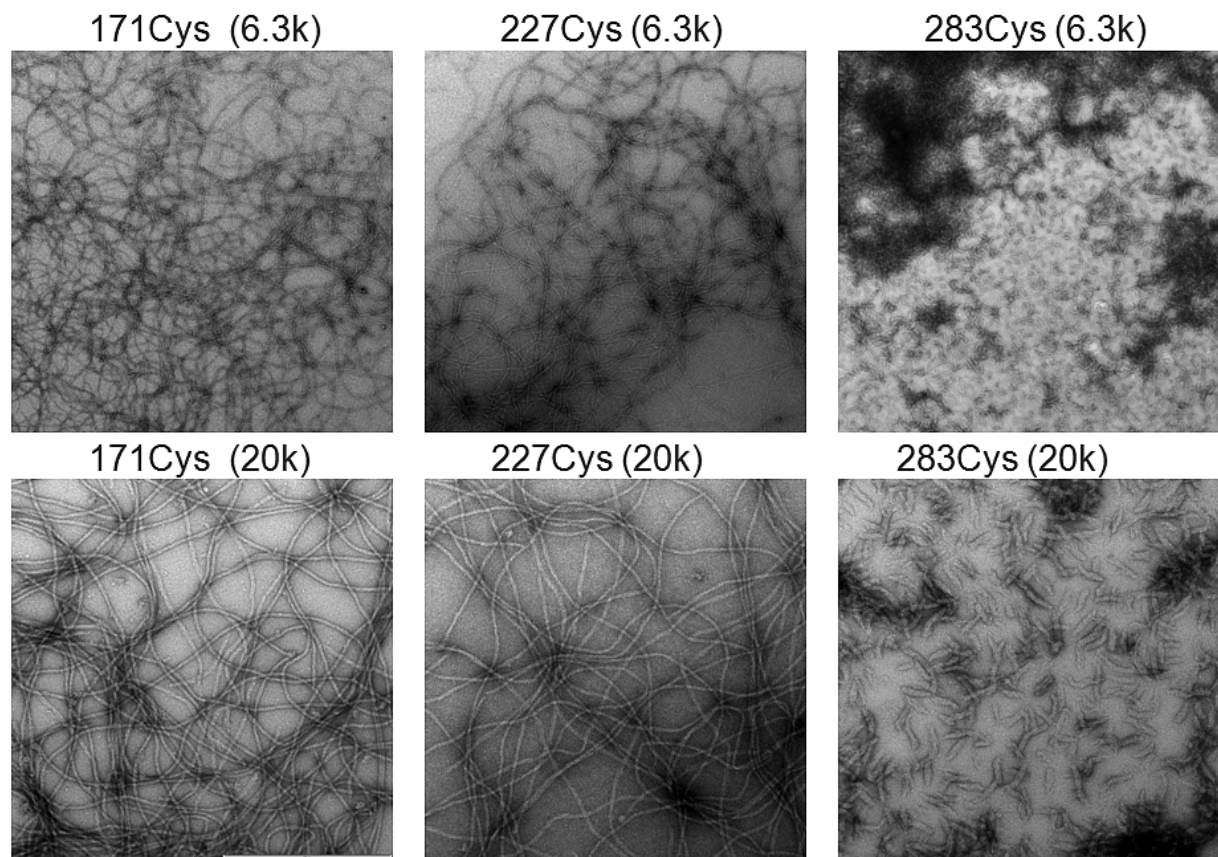
Fig. 29. EM observation of the filaments of the human vimentin mutants at the magnification 6300 and 20000. Samples are assembled at 37˚C for 1h. Corresponding mutations are indicated above each picture. Scale bar corresponds to 500 nm.

Chapter 5. General conclusions and perspectives

5.1 Structure of the IF rod dimer

IFs are an important component of animal cytoskeleton and their studies are of a big industrial and medical interest. Not only products of our daily use including pharmacological and pharmaceutical goods derive from the animal tissues where main building components are the IF proteins in particular but also a number of different diseases are related to altered IF functioning. A vast amount of mutations found in different domains of IF proteins cause IF-pathies [61]. In addition, vimentin is overexpressed in tumour cells of different origin and enhances their motility and invasive activity [67-69]. To understand underlying mechanisms and to be able to manipulate development or progress of diseases, a knowledge of IF structure is necessary.

The main goal of this PhD work was to get an insight into the IF structure in atomic detail. This is the first step towards understanding of how different mutations cause IF-related diseases in humans. However a gap between the knowledge of the structure and ability to treat related diseases is huge. Even if the structure becomes available, there will be a long way before a certain drug can be produced. At first the design of active compounds compensating for the lost functions or preventing aberrant interactions may be needed. Thereafter, an efficiency of the interaction (between the drug and the target protein) has to be proved by other experiments including crystallography. When the complex is resolved, an interaction with a drug may be improved or assessed further with other *in vitro* and *in vivo* experiments, using cell biological, biochemical and biophysical approaches.

From the sequence analysis and other studies (*i.e.*, immunological, biophysical, *etc.*) we can estimate how structurally close two IF proteins are. All IF proteins have similar rod domains and overall tripartite organization. Therefore, the structure of any IF protein can be used to a certain extent to understand the structure of another IF protein. To get broader understanding of the structural properties of IFs in this work we studied one cytoplasmic (vimentin) and one nuclear (lamin A) IF protein.

The main challenge in obtaining the structure of the IF proteins is their flexibility and natural tendency to self-assemble, which disturb crystallization. In addition, IF proteins are too big for NMR studies. Therefore, a 'divide-and-conquer' fragment-based approach was used [122].

As a result, three new structures of human vimentin fragments that span the first half of the rod domain, including coil1A, linker L1 and coil1B (PDB IDs: *3SSU, 3S4R* and *3SWK* as well as one structure corresponding to the coil2 (PDB ID: *3TRT)* were deposited to the PDB [127, 128].

It should be noted that all attempts to solve the target crystal CC structures using molecular replacement (MR) failed. MR search is based on an already known structure. The main condition to be able to use a known structure for MR is its structural similarity with the final model. As a rule of thumb for majority of the proteins, a sequence similarity higher than 40% indicates a good candidate for a successful MR. However, in the case of CC this rule does not work. It was demonstrated that the MR solution cannot be found for an IF protein fragment, if the root-mean square deviation of the superimposed target structure and the model is more than 2 Å (personal communications with Dmytro Guzenko, unpublished). The main difficulty to find a correct position of the model protein is the symmetry of an α-helix as well as of the CC. A-helicity leads to the periodically repeating peaks of the Patterson map in the direction coinciding with the axis of the α-helix. During the search of the model position many similar solutions correspond to the Patterson map and the correct solution could not be distinguished. Therefore, all four structures obtained during this thesis work were solved using SeMet-containing proteins. Se was incorporated during protein expression using media containing SeMet. Even though the substitution rate of S by Se was only about 66%, it was enough to solve the phase problem.

Another problem tackled in this work is a formation of CCs with wrong multiplicity. This is due to the fact that instead of a whole protein short fragments were used for crystallization. For example, trimeric structures corresponding to vimentin coil1B (see chapter 4.1.3) or tetrameric *D3* structure corresponding to a first half of vimentin coil2 [125] were observed. To overcome this problem we developed an approach to stabilize dimeric state of the corresponding fragments. An intelligently incorporated mutation at preferably *d* [183] or *h* [127] positions of the CC core to Cys allowed us to bring two α-helices in a proximity characteristic for CC using an S-S bridge. This technique led to a successfully resolved dimeric *D3st* structure of the first half of vimentin coil2 (PDB ID: *3TRT*) [127]. Combining the determined structures with other available in the PDB IF protein fragments structures (see Table 4) allowed us to model dimeric and tetrameric stages of the vimentin assembly (Fig.13 and 26). These models help to understand the mechanism of the filament assembly.

5.2 Driving force of the tetramer formation

The tetramer formation is made possible by a specific interaction of the two dimers. It was shown that both the 'headless' vimentin and isolated vimentin rod stayed dimeric in the standard low molarity buffer (5mM Tris-HCl, pH 8.4) [121]. The head's involvement into tetramer formation appears to be essential and based on the attraction between eleven arginines of the head domain and the acidic region around the first half of the coil2 domain [42]. Notably, these Arg are the only charged residues within the first 80 residues of vimentin sequence (*i.e.*, the head domain).

The tetrameric model revealed specific salt bridges. These bridges occur between the end of the coil1A of one dimer and the end of the coil 1B of the other dimer. They align two dimers in an antiparallel fashion, the A_{11} mode (Fig. 6 and 26, for details see chapter "Vimentin tetramer and higher oligomer states") as well as stabilise a compact tetramer. We hypothesize that the complimentary charge patterns of the rod domains, while being insufficient to drive the stable tetramer formation alone, may be important for defining the correct register of the A_{11} tetramer.

5.3 Driving force of the filament elongation

Our tetramer models can be assembled into putative ULFs. We show that the N283C mutation yields vimentin assembling to ULFs only, while further annealing of ULFs does not occur. Apparently an interaction important for assembly is affected by this mutation. This implies that the position 283 is directly involved in this interaction. On one hand, the beginning of the coil 2, where the residue Asn283 is located (heptad *d* position, Fig. 11), is believed to interact with the neighbouring parallel α-helix exposing Asn283 for such an interaction to form a parallel CC dimer. On the other hand, our data may suggest yet another possibility – an interaction with coil 1A and the head. There are several facts in support of this hypothesis.

1. Within the tetramer coil 1A is aligned with the first half of coil2 (including the position 283) (Fig. 26).

2. The mutation Y117L within the coil1A leads to a similar effect as mutation N283C: the assembly stops at the ULF level [197]. Coil1A, bearing a native Tyr117, is a very weak dimer in solution [122], stabilized by the Y117L mutation [124]. From this we can conclude that coil 1A has to be monomeric in order to allow longitudinal annealing of the ULFs into the filaments. The split of coil1A into two monomers within the rod dimer was confirmed by the structure of the fragment *1ABL* [128]. However, coil 1A has a typical hydrophobic core, which points at a possibility of a contact with another close nearby α-helix, for example, the beginning of the coil2.

3. The coil2 has a tendency to unpair its intrinsic CC at its N-end (up to residue 303), as it was demonstrated in *D3* structure [125].

4. It was shown that an addition of the coil1 fragments to the full-length vimentin tetramers results in complexes not capable of ULF formation [140]. The same could be expected when coil2 fragment is added. However, an addition of coil2 does not jeopardize normal assembly process [140]. The latter can be explained by that fact that an interaction coil1A-coil1A is much stronger than coil2-coil2 and the effect is more pronounced (nevertheless, an influence of the addition of the short fragment corresponding to the parallel CC of the beginning of coil 2 on the filament assembly has to be checked).

5. As demonstrated by cross-linking experiments [42] the heads are liberated from the interaction with coil1 at the longitudinal stage of the filament assembly. This may lead to the new interaction between the head domains and the coils 2 of the neighbouring ULF. Moreover, it was demonstrated with SDSL-EPR experiments [143, 144] that at the early stages of assembly the head domain is attached to the coil 1. Phosphorylation results in disassembly of the filaments in cells [53], while *in vitro* causes disassembly of the tetramers (weakening of the A_{11} mode and complete extinction of the A_{22} mode) till the dimers [145]. Under phosphorylation, a repulsion of the two heads from each other was demonstrated in the dimers (see Fig. 5 in [143]), at the same time an interaction of the head domain with the rod was not lost [143]. These experiments confirm that at the dimeric level the head interacts with the coil1 of the same molecule.

6. For several IF proteins the presence of the head domain was shown indispensable for the formation of *bona fide* filaments [142]. In particular, the headless vimentin forms only dimers in a low-salt "tetrameric" buffer and the tetramers of the A_{22} mode in a high-salt "assembly" buffer [121, 140]. This points at the fact that the head domain prevents coils 2 from interdimeric interactions (*i.e.* A_{22} mode) during the lateral assembly, saving this contacts to be formed at the later assembly stage – during the longitudinal annealing of the ULFs.

All these experiments taken together may indicate that the interdimeric CC between coils1A occurs during assembly of the ULFs. Afterwards, this contact weakens due to rearrangement of the ionic interactions. As a result, structural changes occur, leading to the interdimeric interaction of the coil1A and the beginning of the coil2. Probably, such an interaction liberates the head domains to interact with the neighbouring ULF, in order to bring the ends of the coil2 in register to form antiparallel tetramers of the mode A_{22}. These contacts are necessary for the filament elongation. However, the hypothesis about the switch of interdimeric interactions coil1A-coil1A to coil1A-beginning of the coil2 still remains to be investigated.

5.4 Unsolved structures – vimentin linker L12 and lamin A fragments

So far no fragments containing significant parts of the head and tail domains of any IF protein could be crystallized, with the exception of an Ig-fold present within the tail domain of nuclear lamins (Table 4) [131]. None of the long fragments including coil1-L12-coil2 parts of vimentin (believed to self-assemble into tetramers) could be crystallized to date.

Well-diffracting crystals of the vimentin fragments including the linker L12 were obtained, in native, SeMet and synthetic forms (fragments 29, 30, Table 5). However, the phase problem has not been overcome yet. This is due to the low quality of anomalous signal, possibly linked to poorly ordered Se atoms. With the current data, a correct HA substructure could not be determined. In addition, no crystals containing the synthetic

fragment alone or in a complex with the other fragments could be observed so far. Consequently, the structure determination of the linker L12 is still in progress.

Based on the results with vimentin rod fragments, we designed multiple fragments of lamin A, covering coil1, linker L12 and a part of coil2. Some of these fragments gave well-diffracting crystals. Two constructs were crystallized in both native and SeMet forms (fragments 45, 46, 48, 49, Table 6), the phase problem still remains to be solved.

5.5 Future perspectives

The results of this PhD work contribute to the structural knowledge of the IF proteins. Many fragments, corresponding to various regions of vimentin and lamin A proteins have been expressed, purified and studied with respect to solubility as well as crystallizability. The crystallographically resolved structures were carefully analysed and compared with results of the other experiments, such as SAXS, cross-linking studies, SDSL-EPR studies, EM, ultracentrifugation and others.

As the next major step, we propose SAXS experiments with the HA labelled vimentin and its mutants. These experiments will provide a set of pairwise distances within vimentin teramer (in A_{11} mode). The information will lead to the accurate positioning of coil1 and coil2 relatively to each other in the dimer as well as the positioning of the two dimers in the tetramer. In turn this will allow ULF modelling with a high degree of accuracy and will shed light on the filament structure.

Overall, the vimentin structure still has to be completed. This knowledge will be used for modelling of other IF proteins. In addition, crystallographic studies of other various IF proteins can give new information. The crystallographic studies of the lamin A structure are at a rather advanced stage. A number of its fragments has been crystallized and the diffraction patterns have been collected. With the results of this work at hand, we hope that one day the full IF protein structure will be available and will be used in medicine to help to improve the quality of our life.

English Summary

Cytoskeleton of eukaryotic cells contains three main filament systems including microfilaments, intermediate filaments (IFs) and microtubules. Among these three systems, IFs are the least understood due to their heterogeneity and complexity. IF family consists of more than 70 members (Szeverenyi et al. 2008), some of which are strictly tissue- or cell-specific, while others are present in almost all mammalian cells. Besides the mechanical role (shaping of the cell organelles, positioning and providing structural integrity of cells and tissues (Goldman et al. 1996; Toivola et al. 2005)), IFs are involved in up- and down-regulation of the cell processes including adhesion, migration and signalling pathways (Toivola et al. 2005; Ivaska et al. 2007). In addition IFs may act as stress proteins (Toivola et al. 2010).

IFs are an important component of animal cytoskeleton and their studies are of a big industrial and medical interest. Firstly, products of our daily use derive from the animal tissues where main building components are the IF proteins. Secondly, a number of different diseases are related to mutations in IF proteins (Szeverenyi et al. 2008). In addition, vimentin is overexpressed in tumour cells of different origin and enhances their motility and invasive activity (Whipple et al. 2008; Schoumacher et al. 2010; Satelli and Li 2011). To understand the underlying mechanisms and to be able to manipulate development or progress of these diseases, the three-dimensional structural knowledge of IFs is indispensable.

All IFs share a similar tripartite primary structure and common principles of assembly. A filament assembly process starts with the dimer formation, which is the elementary building block of all IF proteins. It is the three-dimensional structure of the dimer which defines the further assembly pathway and the properties of the assembled filaments. The elementary dimer consists of a central α-helical rod domain flanked by the N- and C-terminal domains without any clear secondary structure. One of the first sequence analysis gave an IF protein model where the rod domain contained three main CC segments, coil1 and coil2, separated by the linker L12, where coil1 in turn is subdivided into coil1A and coil1B by the linker L1.

In this work we focused on X-ray crystallographic studies of human vimentin and lamin A as the 'model' IF proteins. Due to the sequence similarity, the models of one cytoplasmic and one nuclear IF proteins can be used for structure prediction of other IF proteins. In order to obtain the structure of the dimer we are using a 'divide-and-conquer' approach (Strelkov et al. 2001). It is based on the preparation and crystallization of multiple shorter fragments of the IF dimer with subsequent merging of the overlapping structures. This strategy helps to overcome flexibility and self-assembly of the full-length IF proteins, both making crystallization impossible. While the attempts to crystallize a full-length IF protein have failed, X-ray structures could be determined for a number of IF protein

fragments (see Table 4). For vimentin, previously resolved structures covered the beginning of the coil1 and the whole coil2 (Strelkov et al. 2002; Meier et al. 2009; Nicolet et al. 2010).

In this thesis four new crystal structures were obtained. They cover the coil1 of the vimentin rod domain (Chernyatina et al. 2012) as well as to the beginning of the coil2 (Chernyatina and Strelkov 2012). At the same time, one more structure corresponding to the second half of the coil1 was obtained independently by Aziz et al. (Aziz et al. 2012). The knowledge of all available structures allowed us to produce a nearly complete atomic model of vimentin rod domain which includes the three α-helical segments (coil1A, coil1B and coil2) interconnected by the linkers (L1 and L12). In particular we show that contrary to earlier predictions the complete coil1 is capable of forming an uninterrupted α-helical structure.

We confirmed a fully α-helical structure of the coil2 by crystallizing a dimeric fragment corresponding to its first half. In addition, with this structure we demonstrated a strategy to promote correct dimerization of short CC fragments *via* introducing disulfide links (Chernyatina and Strelkov 2012).

An analysis of the available X-ray structures of vimentin revealed regularly appearing hydrogen bonds breaks in the α-helices. These breaks, called kinks, may be characteristic for all IF proteins and contribute to the flexibility of IFs. This hypothesis, however, has to be investigated further.

We were able to expand the structural knowledge of the vimentin fragments for modelling dimeric and tetrameric stages of the vimentin assembly (Fig.13 and 26). Proposed models shed light on the mechanism of the consequtive assembly stages. We hypothesise that two alternative and consecutive interdimeric interactions involving coil 1A may occur. We propose a two-step conformational change from the interdimeric interaction coil1A-coil1A within one ULF to the intratetrameric interaction coil1A-beginning of coil2, when the assembly switches from the lateral to the longitudinal annealing of the subunits. This scheme is in an agreement with the results of other experiments, in particular, the cross-linking (Herrmann and Aebi 2004) and SDSL-EPR studies (Pittenger 2008; Aziz et al. 2009; Aziz et al. 2010). However, this conjecture remains to be investigated further.

In addition we propose a new approach for studying IF dimers and tetramers using HA-SAXS. This technique is based on a possibility to extract pair-wise distances between HAs attached to Cys residues in a contrast-matched protein/sucrose solution (Grishaev et al. 2012). To evaluate an effect of the implemented mutations electron microscopy observation of the assembled filaments of the WT and three mutants vimentin was performed. To our surprise the mutant N283C abolished its assembly at the ULF level. This mutation therefore may be used as a tool for a further IFs research. Indirectly, the effect of this mutation suggests a possibility of the CC contact between coil1A and the beginning of the coil2, discussed above.

Finally, a number of lamin A fragments was purified and screened for crystallization. Despite the fact that no new lamin A structures have been obtained in the frame of this work, the crystallographic studies are at a rather advanced stage, as some fragments have been crystallized and their diffraction data collected. The phase determination is in progress.

Nederlandse Samenvatting

Het cytoskelet van eukaryote cellen bestaat uit drie groepen van filament systemen, de microfilamenten, intermediare filamenten (IFs) en microtubuli. Van deze drie groepen, zijn de IF het minst gekarakteriseerd omwille van hun complexiteit en heterogeniteit. De familie van IF eiwitten bevat meer dan 70 leden (Szeverenyi et al. 2008), waarvan sommigen strikt weefsel- of celspecifiek zijn, terwijl andere leden te vinden zijn in nagenoeg alle cellen bij zoogdieren. IFs hebben niet alleen een mechanische functie – die bestaat uit vormgeving aan cel-organellen, positionering en voorzien van structurele integriteit aan cellen en weefsels – maar zijn ook betrokken in de regulatie van cellulaire processen zoals adhesie, migratie en signaaltransductie (Toivola et al. 2005; Ivaska et al. 2007). Bovendien kunnen IF optreden als stress eiwitten (Toivola et al. 2010).

IFs zijn een belangrijke component van het cytoskelet, en de studie ervan is van groot belang voor zowel de industriële als medische wereld. Veel producten uit het dagelijkse leven (zoals kledij, voeding en geneesmiddelen onder andere) zijn afhankelijk van gezond vee in het algemeen en dan voornamelijk hun vacht, nagels en horens (weefsels die voornamelijk opgebouwd zijn uit IFs). Verder vinden verschillende ziekten hun oorsprong bij mutaties in intermediaire filamenten (Szeverenyi et al. 2008). Daarenboven is vimentin zelfs overmatig aanwezig in verscheidene tumorcellen waar het de beweging en metastase-activiteit verhoogt (Whipple et al. 2008; Schoumacher et al. 2010; Satelli and Li 2011). Om de onderliggende mechanisme te begrijpen en om het ontstaan en de ontwikkeling van ziekten te manipuleren, is kennis over hun drie-dimensionale structuur van cruciaal belang.

Alle IF hebben een gelijkaardige drieledige primaire structuur en opbouwmechanisme. Het assembleren van intermediaire filamenten start met de vorming van een dimeer, dat de basis eenheid vormt voor alle IF eiwitten. De drie-dimensionale structuur van het dimeer bepaalt de verdere opbouw en zo ook de uiteindelijke eigenschappen van het filament. Het dimeer bestaat uit een centraal α-helicaal "staaf" domein dat geflankeerd wordt door N- en C-terminale domeinen zonder duidelijke secundaire structuur. Eén van de eerste sequentie-analyses resulteerde in een model aan waarin het centrale staaf-domein bestaat uit drie coiled-coil segmenten: coil1 en coil2, die van elkaar gescheiden worden door een linker-domein genaamd linker L12. Coil1 is op zijn beurt opgedeeld in coil1A en coil1B die op zich dan weer gescheiden worden door linker L1.

Deze studie omvat kristallografische studies van humaan vimentin en lamin A als model-eiwitten voor intermediaire filamenten. Vanwege gelijkenissen in eiwitsequentie, kan vimentin gebruikt worden als model voor cytoplasmatische IF en lamin A voor nucleaire IF. Om tot de 3D structuur te komen, werd een techniek genaamd 'verdeel-en-heers' gebruikt Strelkov et al. 2001). Dit is gebaseerd op de aanmaak en kristallisatie van verscheidene

kortere fragmenten van het IF-dimeer waarvan de uiteinden overlappen. Door samenvoegen van de 3D structuren van deze kleinere fragmenten, kan zo een model bekomen worden voor het gehele eiwit. Deze strategie werd toegepast om de flexibiliteit en polymerisatie tot filamenten van het wild type eiwit te omzeilen, dewelke kristallisatie onmogelijk maken. Integenstelling tot de onsuccesvolle pogingen om de volledige IF eiwitten te kristalliseren, werden de structuren bepaald voor een aantal IF eiwit fragmenten (zie Tabel 4). Voor vimentin was de structuur van een deel van coil1 en de volledige coil2 reeds bepaald in vorige studies (Strelkov et al. 2002; Meier et al. 2009; Nicolet et al. 2010).

Tijdens dit proefschrift werden vier nieuwe kristalstructuren verkregen. Deze omvatten coil1 van het centrale staaf-domein van vimentin (Chernyatina et al. 2012) alsook het N-terminale deel van coil2 (Chernyatina and Strelkov 2012). Terzelfdertijd, werd onafhankelijk van deze studie ook de structuur van het C-terminale deel van coil1 bepaald door Aziz *et al.* (Aziz et al. 2012). Met deze beschikbare structuren werd een nagenoeg volledig moleculair model gebouwd van het vimentin staaf-domein dat de drie helicale segmenten omvat (coil1A, coil1B en coil2) erbonden door hun linkers (L1 en L12, respectievelijk). In het bijzonder konden we aantonen dat de volledige coil1 bestaat uit een aangesloten α-helicale structuur, in tegenstelling tot eerdere voorspellingen.

We bevestigden ook een volledige α-helicale structuur voor coil2, door het kristalliseren van een dimeer fragment dat de eerste helft van coil2 omvat. Daarnaast konden we met behulp van deze structuur een strategie voorstellen om de juiste dimerisatie van korte coiled-coil fragmenten te promoten via het introduceren van disulfide-bruggen (Chernyatina and Strelkov 2012).

Analyze van de beschikbare X-stralen kristalstructuren van vimentin, onthulde regelmatig voorkomende breuken in de waterstofbruggen in de α-helices. Deze breuken, die kinks worden genoemd, kunnen karakteristiek zijn voor IF eiwitten en bijdragen tot hun flexibiliteit. Maar deze hypothese vereist nog verder onderzoek.

We zijn erin geslaagd om de structurele kennis uit fragmenten van vimenten aan te wenden voor het opbouwen van een modelstructuur voor de dimerisatie en tetramerisatie van vimentin (Fig.13 en 26). Deze modellen onthullen het mogelijke mechanisme van achtereenvolgende stappen in opbouw. We stellen een uit twee stappen bestaande verandering voor van de interdimerische interactie tussen coil 1A en coil 1A naar coil 1A met het begin van coil 2 wanneer de montage verandert van een laterale naar een longitudinale aaneenschakeling van de subunits. Dit schema is een goede overeenkomst met resultaten uit andere experimenten, en in het bijzonder met cross-linking (Herrmann and Aebi 2004) en SDSL-EPR studies (Pittenger 2008; Aziz et al. 2009; Aziz et al. 2010). Desalniettemin, moet dit model nog verder onderzocht worden.

Daarnaast stellen we een nieuwe strategie voor voor de studie van IF dimeren en tetrameren gebruikmakende van HA-SAXS. Deze techniek is gebaseerd op de mogelijkheid

om afstanden tussen gekoppelde zware metalen verbonden aan Cys residu's te meten in een contrast-gepaarde sucrose/eiwit oplossing (Grishaev et al. 2012). Om het effect van aangebrachte mutaties te evalueren, werden elektronen microscopie van de geassembleerde filamenten van de wild type en drie mutanten uitgevoerd. Tot onze verbazing, werd in de N283C mutant de opbouw van filamenten aan het niveau van ULF verhinderd. Deze mutatie kan dus worden gebruikt als hulpmiddel in verdere studies. Het effect van deze mutatie impliceert op indirecte wijze de mogelijkheid van CC contact tussen coil1A en het begin van coil2 zoals hierboven beschreven.

Tot slot werden ook verscheidene fragmenten van lamin A opgezuiverd en gekristalliseerd. Desondanks het feit dat er geen nieuwe lamin A structuren bekomen werden in dit proefschrift, zijn de kristallografische studies verder geavanceerd, daar dat sommige fragmenten gekristalliseerd zijn en X-stralen diffractie patronen gecollecteerd zijn. Het oplossen van het fase-probleem is lopende.

References

1. Astbury, W. T., and Woods, S. J. (1930) *Nature* **126**, 913-914
2. Pauling, L., Corey, R. B., and Branson, H. R. (1951) *Proceedings of the National Academy of Sciences of the United States of America* **37**(4), 205-211
3. Pauling, L., and Corey, R. B. (1951) *Proceedings of the National Academy of Sciences of the United States of America* **37**(11), 729-740
4. Crick, F. H. C. (1952) *Nature* **170**, 882-883
5. Crick, F. H. C. (1953) *Acta crystallographica* **6**, 689-697
6. Wolf, E., Kim, P. S., and Berger, B. (1997) *Protein Sci* **6**(6), 1179-1189
7. Lupas, A., Van Dyke, M., and Stock, J. (1991) *Science (New York, N.Y* **252**(5009), 1162-1164
8. Landschulz, W. H., Johnson, P. F., and McKnight, S. L. (1988) *Science (New York, N.Y* **240**(4860), 1759-1764
9. O'Shea, E. K., Klemm, J. D., Kim, P. S., and Alber, T. (1991) *Science (New York, N.Y* **254**(5031), 539-544
10. Herrmann, H., Bar, H., Kreplak, L., Strelkov, S. V., and Aebi, U. (2007) *Nature reviews* **8**(7), 562-573
11. Kohn, W. D., Mant, C. T., and Hodges, R. S. (1997) *The Journal of biological chemistry* **272**(5), 2583-2586
12. Fuchs, E., and Weber, K. (1994) *Annual review of biochemistry* **63**, 345-382
13. Nogales, E., Wolf, S. G., and Downing, K. H. (1998) *Nature* **391**(6663), 199-203
14. Kabsch, W., and Vandekerckhove, J. (1992) *Annual review of biophysics and biomolecular structure* **21**, 49-76
15. Ishikawa, H., Bischoff, R., and Holtzer, H. (1968) *The Journal of cell biology* **38**(3), 538-555
16. Steinert, P. M., Idler, W. W., and Zimmerman, S. B. (1976) *Journal of molecular biology* **108**(3), 547-567
17. Zackroff, R. V., and Goldman, R. D. (1979) *Proceedings of the National Academy of Sciences of the United States of America* **76**(12), 6226-6230
18. Gard, D. L., and Lazarides, E. (1980) *Cell* **19**(1), 263-275
19. Zackroff, R. V., and Goldman, R. D. (1980) *Science (New York, N.Y* **208**(4448), 1152-1155
20. Steinert, P., Idler, W., Aynardi-Whitman, M., Zackroff, R., and Goldman, R. D. (1982) *Cold Spring Harbor symposia on quantitative biology* **46 Pt 1**, 465-474
21. Steinert, P. M., and Gullino, M. I. (1976) *Biochemical and biophysical research communications* **70**(1), 221-227
22. Steinert, P. M. (1990) *The Journal of biological chemistry* **265**(15), 8766-8774
23. Quinlan, R. A., Hatzfeld, M., Franke, W. W., Lustig, A., Schulthess, T., and Engel, J. (1986) *Journal of molecular biology* **192**(2), 337-349
24. Geisler, N., and Weber, K. (1982) *The EMBO journal* **1**(12), 1649-1656
25. Steinert, P. M., Idler, W. W., and Goldman, R. D. (1980) *Proceedings of the National Academy of Sciences of the United States of America* **77**(8), 4534-4538
26. Hanukoglu, I., and Fuchs, E. (1982) *Cell* **31**(1), 243-252
27. Parry, D. A., Crewther, W. G., Fraser, R. D., and MacRae, T. P. (1977) *Journal of molecular biology* **113**(2), 449-454
28. Harbury, P. B., Plecs, J. J., Tidor, B., Alber, T., and Kim, P. S. (1998) *Science (New York, N.Y* **282**(5393), 1462-1467
29. Franke, W. W., Schmid, E., Osborn, M., and Weber, K. (1978) *Proceedings of the National Academy of Sciences of the United States of America* **75**(10), 5034-5038
30. Lazarides, E. (1982) *Annual review of biochemistry* **51**, 219-250
31. Lazarides, E. (1980) *Nature* **283**(5744), 249-256
32. Steinert, P. M., and Parry, D. A. (1985) *Annu Rev Cell Biol* **1**, 41-65
33. Hatzfeld, M., and Weber, K. (1990) *The Journal of cell biology* **110**(4), 1199-1210
34. Gerace, L., and Blobel, G. (1980) *Cell* **19**(1), 277-287

35. Gerace, L., Blum, A., and Blobel, G. (1978) *The Journal of cell biology* **79**(2 Pt 1), 546-566

36. Fisher, D. Z., Chaudhary, N., and Blobel, G. (1986) *Proceedings of the National Academy of Sciences of the United States of America* **83**(17), 6450-6454

37. McKeon, F. D., Kirschner, M. W., and Caput, D. (1986) *Nature* **319**(6053), 463-468

38. Aebi, U., Cohn, J., Buhle, L., and Gerace, L. (1986) *Nature* **323**(6088), 560-564

39. Steinert, P. M., and Roop, D. R. (1988) *Annual review of biochemistry* **57**, 593-625

40. Merdes, A., Gounari, F., and Georgatos, S. D. (1993) *The Journal of cell biology* **123**(6 Pt 1), 1507-1516

41. Ausmees, N., Kuhn, J. R., and Jacobs-Wagner, C. (2003) *Cell* **115**(6), 705-713

42. Herrmann, H., and Aebi, U. (2004) *Annual review of biochemistry* **73**, 749-789

43. Mencarelli, C., Ciolfi, S., Caroti, D., Lupetti, P., and Dallai, R. (2011) *BMC biology* **9**, 17

44. Rogers, M. A., Edler, L., Winter, H., Langbein, L., Beckmann, I., and Schweizer, J. (2005) *The Journal of investigative dermatology* **124**(3), 536-544

45. Rogers, M. A., Winter, H., Langbein, L., Bleiler, R., and Schweizer, J. (2004) *Differentiation; research in biological diversity* **72**(9-10), 527-540

46. Eriksson, J. E., Dechat, T., Grin, B., Helfand, B., Mendez, M., Pallari, H. M., and Goldman, R. D. (2009) *The Journal of clinical investigation* **119**(7), 1763-1771

47. Moll, R., Moll, I., and Franke, W. W. (1984) *Archives of dermatological research* **276**(6), 349-363

48. Stick, R., and Hausen, P. (1985) *Cell* **41**(1), 191-200

49. Goldman, R. D., Khuon, S., Chou, Y. H., Opal, P., and Steinert, P. M. (1996) *The Journal of cell biology* **134**(4), 971-983.

50. Toivola, D. M., Tao, G. Z., Habtezion, A., Liao, J., and Omary, M. B. (2005) *Trends in cell biology* **15**(11), 608-617

51. Ivaska, J., Pallari, H. M., Nevo, J., and Eriksson, J. E. (2007) *Experimental cell research* **313**(10), 2050-2062

52. Toivola, D. M., Strnad, P., Habtezion, A., and Omary, M. B. (2010) *Trends in cell biology* **20**(2), 79-91

53. Omary, M. B., Ku, N. O., Tao, G. Z., Toivola, D. M., and Liao, J. (2006) *Trends in biochemical sciences* **31**(7), 383-394

54. Fuchs, E., and Marchuk, D. (1983) *Proceedings of the National Academy of Sciences of the United States of America* **80**(19), 5857-5861

55. Moll, R., Franke, W. W., Schiller, D. L., Geiger, B., and Krepler, R. (1982) *Cell* **31**(1), 11-24

56. Bragulla, H. H., and Homberger, D. G. (2009) *Journal of anatomy* **214**(4), 516-559

57. Moll, R., Divo, M., and Langbein, L. (2008) *Histochemistry and cell biology* **129**(6), 705-733

58. Pekny, M., and Lane, E. B. (2007) *Experimental cell research* **313**(10), 2244-2254

59. Navarro, J. M., Casatorres, J., and Jorcano, J. L. (1995) *The Journal of biological chemistry* **270**(36), 21362-21367

60. Weil, M., Raff, M. C., and Braga, V. M. (1999) *Curr Biol* **9**(7), 361-364

61. Szeverenyi, I., Cassidy, A. J., Chung, C. W., Lee, B. T., Common, J. E., Ogg, S. C., Chen, H., Sim, S. Y., Goh, W. L., Ng, K. W., Simpson, J. A., Chee, L. L., Eng, G. H., Li, B., Lunny, D. P., Chuon, D., Venkatesh, A., Khoo, K. H., McLean, W. H., Lim, Y. P., and Lane, E. B. (2008) *Hum Mutat* **29**(3), 351-360

62. Coulombe, P. A., Hutton, M. E., Letai, A., Hebert, A., Paller, A. S., and Fuchs, E. (1991) *Cell* **66**(6), 1301-1311

63. Owens, D. W., Wilson, N. J., Hill, A. J., Rugg, E. L., Porter, R. M., Hutcheson, A. M., Quinlan, R. A., van Heel, D., Parkes, M., Jewell, D. P., Campbell, S. S., Ghosh, S., Satsangi, J., and Lane, E. B. (2004) *Journal of cell science* **117**(Pt 10), 1989-1999

64. Magin, T. M., Reichelt, J., and Hatzfeld, M. (2004) *Experimental cell research* **301**(1), 91-102

65. Granger, B. L., and Lazarides, E. (1979) *Cell* **18**(4), 1053-1063

66. Steinert, P. M., Idler, W. W., Cabral, F., Gottesman, M. M., and Goldman, R. D. (1981) *Proceedings of the National Academy of Sciences of the United States of America* **78**(6), 3692-3696

67. Satelli, A., and Li, S. (2011) *Cell Mol Life Sci* **68**(18), 3033-3046

68. Whipple, R. A., Balzer, E. M., Cho, E. H., Matrone, M. A., Yoon, J. R., and Martin, S. S. (2008) *Cancer research* **68**(14), 5678-5688

69. Schoumacher, M., Goldman, R. D., Louvard, D., and Vignjevic, D. M. (2010) *The Journal of cell biology* **189**(3), 541-556

70. Markowitz, S. D. (2012) Methods and compositions for detecting colon cancers. In., Case Western Reserve University

71. Muller, M., Bhattacharya, S. S., Moore, T., Prescott, Q., Wedig, T., Herrmann, H., and Magin, T. M. (2009) *Hum Mol Genet* **18**(6), 1052-1057

72. Colucci-Guyon, E., Portier, M. M., Dunia, I., Paulin, D., Pournin, S., and Babinet, C. (1994) *Cell* **79**(4), 679-694

73. Stover, J. F., Sakowitz, O. W., and Unterberg, A. W. (2004) *Acta Neurochir Suppl* **89**, 81-85

74. Colucci-Guyon, E., Gimenez, Y. R. M., Maurice, T., Babinet, C., and Privat, A. (1999) *Glia* **25**(1), 33-43

75. Gladue, D. P., O'Donnell, V., Baker-Branstetter, R., Holinka, L. G., Pacheco, J. M., Fernandez Sainz, I., Lu, Z., Ambroggio, X., Rodriguez, L., and Borca, M. V. (2013) *Journal of virology* **87**(12), 6794-6803

76. Bhattacharya, B., Noad, R. J., and Roy, P. (2007) *Virol J* **4**, 7

77. Xu, B., deWaal, R. M., Mor-Vaknin, N., Hibbard, C., Markovitz, D. M., and Kahn, M. L. (2004) *Molecular and cellular biology* **24**(20), 9198-9206

78. Middeldorp, J., and Hol, E. M. (2011) *Progress in neurobiology* **93**(3), 421-443

79. Hsiao, V. C., Tian, R., Long, H., Der Perng, M., Brenner, M., Quinlan, R. A., and Goldman, J. E. (2005) *Journal of cell science* **118**(Pt 9), 2057-2065

80. Athlan, E. S., and Mushynski, W. E. (1997) *The Journal of biological chemistry* **272**(49), 31073-31078

81. Lariviere, R. C., Nguyen, M. D., Ribeiro-da-Silva, A., and Julien, J. P. (2002) *Journal of neurochemistry* **81**(3), 525-532

82. Bar, H., Strelkov, S. V., Sjoberg, G., Aebi, U., and Herrmann, H. (2004) *Journal of structural biology* **148**(2), 137-152

83. Herrmann, H., Fouquet, B., and Franke, W. W. (1989) *Development (Cambridge, England)* **105**(2), 279-298

84. Milner, D. J., Weitzer, G., Tran, D., Bradley, A., and Capetanaki, Y. (1996) *The Journal of cell biology* **134**(5), 1255-1270

85. Li, Z., Colucci-Guyon, E., Pincon-Raymond, M., Mericskay, M., Pournin, S., Paulin, D., and Babinet, C. (1996) *Developmental biology* **175**(2), 362-366

86. Steinert, P. M., Chou, Y. H., Prahlad, V., Parry, D. A., Marekov, L. N., Wu, K. C., Jang, S. I., and Goldman, R. D. (1999) *The Journal of biological chemistry* **274**(14), 9881-9890

87. Herrmann, H., and Aebi, U. (2000) *Current opinion in cell biology* **12**(1), 79-90

88. Ching, G. Y., and Liem, R. K. (1993) *The Journal of cell biology* **122**(6), 1323-1335

89. Abumuhor, I. A., Spencer, P. H., and Cohlberg, J. A. (1998) *Journal of structural biology* **123**(3), 187-198

90. Parry, D. A., Conway, J. F., and Steinert, P. M. (1986) *The Biochemical journal* **238**(1), 305-308

91. Kapinos, L. E., Schumacher, J., Mucke, N., Machaidze, G., Burkhard, P., Aebi, U., Strelkov, S. V., and Herrmann, H. (2010) *Journal of molecular biology* **396**(3), 719-731

92. Strelkov, S. V., Schumacher, J., Burkhard, P., Aebi, U., and Herrmann, H. (2004) *Journal of molecular biology* **343**(4), 1067-1080

93. Sasse, B., Aebi, U., and Stuurman, N. (1998) *Journal of structural biology* **123**(1), 56-66

94. Stuurman, N., Heins, S., and Aebi, U. (1998) *Journal of structural biology* **122**(1-2), 42-66

95. Ben-Harush, K., Wiesel, N., Frenkiel-Krispin, D., Moeller, D., Soreq, E., Aebi, U., Herrmann, H., Gruenbaum, Y., and Medalia, O. (2009) *Journal of molecular biology* **386**(5), 1392-1402

96. Grossman, E., Dahan, I., Stick, R., Goldberg, M. W., Gruenbaum, Y., and Medalia, O. (2012) *Journal of Structural Biology* **177**(1), 113-118

97. Herrmann, H., Haner, M., Brettel, M., Ku, N. O., and Aebi, U. (1999) *Journal of molecular biology* **286**(5), 1403-1420

98. Dittmer, T. A., and Misteli, T. (2011) *Genome biology* **12**(5), 222

99. Riemer, D., and Weber, K. (1998) *Journal of cell science* **111**, 2967-2975

100. Geisler, N., Schunemann, J., Weber, K., Haner, M., and Aebi, U. (1998) *Journal of molecular biology* **282**(3), 601-617.

101. Blumenthal, S. S., Clark, G. B., and Roux, S. J. (2004) *Planta* **218**(6), 965-975

102. Houben, F., Ramaekers, F. C., Snoeckx, L. H., and Broers, J. L. (2007) *Biochimica et biophysica acta* **1773**(5), 675-686

103. Goldman, R. D., Gruenbaum, Y., Moir, R. D., Shumaker, D. K., and Spann, T. P. (2002) *Genes & development* **16**(5), 533-547

104. Dechat, T., Pfleghaar, K., Sengupta, K., Shimi, T., Shumaker, D. K., Solimando, L., and Goldman, R. D. (2008) *Genes & development* **22**(7), 832-853

105. Furukawa, K., and Hotta, Y. (1993) *The EMBO journal* **12**(1), 97-106

106. Furukawa, K., Inagaki, H., and Hotta, Y. (1994) *Experimental cell research* **212**(2), 426-430

107. Padiath, Q. S., Saigoh, K., Schiffmann, R., Asahara, H., Yamada, T., Koeppen, A., Hogan, K., Ptacek, L. J., and Fu, Y. H. (2006) *Nature genetics* **38**(10), 1114-1123

108. Hegele, R. A., Cao, H., Liu, D. M., Costain, G. A., Charlton-Menys, V., Rodger, N. W., and Durrington, P. N. (2006) *American journal of human genetics* **79**(2), 383-389

109. Eriksson, M., Brown, W. T., Gordon, L. B., Glynn, M. W., Singer, J., Scott, L., Erdos, M. R., Robbins, C. M., Moses, T. Y., Berglund, P., Dutra, A., Pak, E., Durkin, S., Csoka, A. B., Boehnke, M., Glover, T. W., and Collins, F. S. (2003) *Nature* **423**(6937), 293-298

110. Reis, A., Hennies, H. C., Langbein, L., Digweed, M., Mischke, D., Drechsler, M., Schrock, E., Royer-Pokora, B., Franke, W. W., Sperling, K., and et al. (1994) *Nature genetics* **6**(2), 174-179

111. Smith, F. J., Corden, L. D., Rugg, E. L., Ratnavel, R., Leigh, I. M., Moss, C., Tidman, M. J., Hohl, D., Huber, M., Kunkeler, L., Munro, C. S., Lane, E. B., and McLean, W. H. (1997) *The Journal of investigative dermatology* **108**(2), 220-223

112. Quax-Jeuken, Y. E., Quax, W. J., and Bloemendal, H. (1983) *Proceedings of the National Academy of Sciences of the United States of America* **80**(12), 3548-3552

113. Parry, D. A. D., and Steinert, P. M. (1999) *Quarterly reviews of biophysics* **32**(2), 99-187

114. Smith, T. A., Strelkov, S. V., Burkhard, P., Aebi, U., and Parry, D. A. (2002) *Journal of structural biology* **137**(1-2), 128-145

115. Brown, J. H., Cohen, C., and Parry, D. A. (1996) *Proteins* **26**(2), 134-145

116. Kuhnel, K., Jarchau, T., Wolf, E., Schlichting, I., Walter, U., Wittinghofer, A., and Strelkov, S. V. (2004) *Proceedings of the National Academy of Sciences of the United States of America* **101**(49), 17027-17032

117. Woolfson, D. N. (2005) *Advances in protein chemistry* **70**, 79-112

118. Harbury, P. B., Zhang, T., Kim, P. S., and Alber, T. (1993) *Science (New York, N.Y* **262**(5138), 1401-1407

119. Parry, D. A., Fraser, R. D., and Squire, J. M. (2008) *Journal of structural biology* **163**(3), 258-269

120. Parry, D. A. (1982) *Bioscience reports* **2**(12), 1017-1024

121. Herrmann, H., Haner, M., Brettel, M., Muller, S. A., Goldie, K. N., Fedtke, B., Lustig, A., Franke, W. W., and Aebi, U. (1996) *Journal of molecular biology* **264**(5), 933-953

122. Strelkov, S. V., Herrmann, H., Geisler, N., Lustig, A., Ivaninskii, S., Zimbelmann, R., Burkhard, P., and Aebi, U. (2001) *Journal of molecular biology* **306**(4), 773-781.

123. Strelkov, S. V., Herrmann, H., Geisler, N., Wedig, T., Zimbelmann, R., Aebi, U., and Burkhard, P. (2002) *The EMBO journal* **21**(6), 1255-1266.

124. Meier, M., Padilla, G. P., Herrmann, H., Wedig, T., Hergt, M., Patel, T. R., Stetefeld, J., Aebi, U., and Burkhard, P. (2009) *Journal of molecular biology* **390**(2), 245-261

125. Nicolet, S., Herrmann, H., Aebi, U., and Strelkov, S. V. (2010) *Journal of structural biology* **170**(2), 369-376

126. Aziz, A., Hess, J. F., Budamagunta, M. S., Voss, J. C., Kuzin, A. P., Huang, Y. J., Xiao, R., Montelione, G. T., Fitzgerald, P. G., and Hunt, J. F. (2012) *The Journal of biological chemistry* **287**(34), 28349-28361

127. Chernyatina, A. A., and Strelkov, S. V. (2012) *Journal of structural biology* **177**(1), 46-53

128. Chernyatina, A. A., Nicolet, S., Aebi, U., Herrmann, H., and Strelkov, S. V. (2012) *Proceedings of the National Academy of Sciences of the United States of America* **109**(34), 13620-13625

129. Lee, C. H., Kim, M. S., Chung, B. M., Leahy, D. J., and Coulombe, P. A. (2012) *Nature structural & molecular biology* **19**(7), 707-715

130. Krimm, I., Ostlund, C., Gilquin, B., Couprie, J., Hossenlopp, P., Mornon, J. P., Bonne, G., Courvalin, J. C., Worman, H. J., and Zinn-Justin, S. (2002) *Structure (Camb)* **10**(6), 811-823.

131. Dhe-Paganon, S., Werner, E. D., Chi, Y. I., and Shoelson, S. E. (2002) *The Journal of biological chemistry* **18**, 18

132. Magracheva, E., Kozlov, S., Stewart, C. L., Wlodawer, A., and Zdanov, A. (2009) *Acta Crystallogr Sect F Struct Biol Cryst Commun* **65**(Pt 7), 665-670

133. Kapinos, L. E., Burkhard, P., Herrmann, H., Aebi, U., and Strelkov, S. V. (2011) *Journal of molecular biology* **408**(1), 135-146

134. Ruan, J., Xu, C., Bian, C., Lam, R., Wang, J. P., Kania, J., Min, J., and Zang, J. (2012) *FEBS letters* **586**(4), 314-318

135. Bollati, M., Barbiroli, A., Favalli, V., Arbustini, E., Charron, P., and Bolognesi, M. (2012) *Biochemical and biophysical research communications* **418**(2), 217-221

136. Lichtenstern, T., Mucke, N., Aebi, U., Mauermann, M., and Herrmann, H. (2012) *Journal of structural biology* **177**(1), 54-62

137. Steinert, P. M., Marekov, L. N., Fraser, R. D., and Parry, D. A. (1993) *Journal of molecular biology* **230**(2), 436-452

138. Steinert, P. M., Marekov, L. N., and Parry, D. A. (1993) *The Journal of biological chemistry* **268**(33), 24916-24925.

139. Hess, J. F., Budamagunta, M. S., Voss, J. C., and FitzGerald, P. G. (2004) *The Journal of biological chemistry* **279**(43), 44841-44846

140. Mucke, N., Wedig, T., Burer, A., Marekov, L. N., Steinert, P. M., Langowski, J., Aebi, U., and Herrmann, H. (2004) *Journal of molecular biology* **340**(1), 97-114

141. Herrmann, H., and Aebi, U. (1998) *Sub-cellular biochemistry* **31**, 319-362

142. Traub, P., and Vorgias, C. E. (1983) *Journal of cell science* **63**, 43-67

143. Aziz, A., Hess, J. F., Budamagunta, M. S., Voss, J. C., and Fitzgerald, P. G. (2010) *The Journal of biological chemistry* **285**(20), 15278-15285

144. Aziz, A., Hess, J. F., Budamagunta, M. S., FitzGerald, P. G., and Voss, J. C. (2009) *The Journal of biological chemistry* **284**(11), 7330-7338

145. Pittenger, J. T., Hess, J.F., Budamagunta, M.S., Voss, J.C., Fitzgerald, P.G. (2008) *Biochemistry* (Oct 14), 47(41):10863-10870

146. Hess, J. F., Budamagunta, M. S., Aziz, A., FitzGerald, P. G., and Voss, J. C. (2012) *Protein Sci* **22**(1), 47-55

147. Kouklis, P. D., Hatzfeld, M., Brunkener, M., Weber, K., and Georgatos, S. D. (1993) *Journal of cell science* **106 (Pt 3)**, 919-928

148. Hess, J. F., Budamagunta, M. S., Shipman, R. L., FitzGerald, P. G., and Voss, J. C. (2006) *Biochemistry* **45**(39), 11737-11743

149. Hess, J. F., Budamagunta, M. S., FitzGerald, P. G., and Voss, J. C. (2005) *The Journal of biological chemistry* **280**(3), 2141-2146

150. Budamagunta, M., Hess, J., Fitzgerald, P., and Voss, J. (2007) *Cell biochemistry and biophysics* **48**(1), 45-53

151. Sokolova, A. V., Kreplak, L., Wedig, T., Mucke, N., Svergun, D. I., Herrmann, H., Aebi, U., and Strelkov, S. V. (2006) *Proceedings of the National Academy of Sciences of the United States of America* **103**(44), 16206-16211

152. Brennich, M. E., Nolting, J. F., Dammann, C., Noding, B., Bauch, S., Herrmann, H., Pfohl, T., and Koster, S. (2011) *Lab on a chip* **11**(4), 708-716

153. Cole, C., Barber, J. D., and Barton, G. J. (2008) *Nucleic acids research* **36**(Web Server issue), W197-201

154. Petersen, B., Petersen, T. N., Andersen, P., Nielsen, M., and Lundegaard, C. (2009) *BMC structural biology* **9**, 51

155. Weeks, S. D., Drinker, M., and Loll, P. J. (2007) *Protein expression and purification* **53**(1), 40-50

156. Studier, F. W. (2005) *Protein expression and purification* **41**(1), 207-234

157. Cowie, D. B., and Cohen, G. N. (1957) *Biochimica et biophysica acta* **26**(2), 252-261

158. Chayen, N. E. (2004) *Current opinion in structural biology* **14**(5), 577-583

159. Ascone, I., Girard, E., Gourhant, P., Legrand, P., Roudenko, O., Roussier, L., and Thompson, A. W. (2006) Proxima 1, a new beamline on the third generation SR source SOLEIL combining PX and single-crystal BioXAS. In. *13th International Conference on X-ray absorption fine structure (XAFS13)* American Institute of Physics, Stanford, California, USA

160. Huelsen, G., Broennimann, C., Eikenberry, E. F., and Wagner, A. (2006) *Journal of Applied Crystallography* **39**, 550-557

161. Otwinowski, Z., and Minor, W. (1997) *Methods in enzymology* **277 Part A**, 307-325

162. Kabsch, W. (2010) *Acta crystallographica* **66**(Pt 2), 125-132

163. Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S., and Tucker, P. A. (2005) *Acta crystallographica* **61**(Pt 4), 449-457

164. Sheldrick, G. M. (2010) *Acta crystallographica* **66**(Pt 4), 479-485

165. Collaborative_Computational_Project_number_4. (1994) The CCP4 suite: programs for protein crystallography. In. *Acta crystallographica*

166. Terwilliger, T. C. (2000) *Acta crystallographica* **56**(Pt 8), 965-972

167. Cowtan, K. (1994) *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* (31), 34-38

168. Perrakis, A., Sixma, T. K., Wilson, K. S., and Lamzin, V. S. (1997) *Acta crystallographica* **55**, 448-455

169. Emsley, P., and Cowtan, K. (2004) *Acta crystallographica* **60**(Pt 12 Pt 1), 2126-2132

170. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) *Acta crystallographica* **66**(Pt 2), 213-221

171. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) *Nucleic acids research* **35**(Web Server issue), W375-383

172. Krissinel, E., and Henrick, K. (2007) *Journal of molecular biology* **372**(3), 774-797

173. Schneider, T. R., and Sheldrick, G. M. (2002) *Acta crystallographica* **58**(Pt 10 Pt 2), 1772-1779

174. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) *J Appl Crystallogr* **40**(Pt 4), 658-674

175. Lupas, A. (1997) *Current opinion in structural biology* **7**(3), 388-393

176. Gruber, M., and Lupas, A. N. (2003) *Trends in biochemical sciences* **28**(12), 679-685

177. Hess, J. F., Budamagunta, M. S., Voss, J. C., and FitzGerald, P. G. (2004) *The Journal of biological chemistry*

178. Strelkov, S. V., and Burkhard, P. (2002) *Journal of structural biology* **137**(1-2), 54-64.

179. Schaffeld, M., Herrmann, H., Schultess, J., and Markl, J. (2001) *European journal of cell biology* **80**(11), 692-702

180. Barlow, D. J., and Thornton, J. M. (1988) *Journal of molecular biology* **201**(3), 601-619

181. North, A. C., Steinert, P. M., and Parry, D. A. (1994) *Proteins* **20**(2), 174-184

182. Parry, D. A. (2006) *Journal of structural biology* **155**(2), 370-374

183. Zhou, N. E., Kay, C. M., and Hodges, R. S. (1993) *Biochemistry* **32**(12), 3178-3187

184. DeLano, W. L. (2002) The PyMOL Molecular Graphics System. In. *DeLano Scientific*

185. Pace, C. N., and Scholtz, J. M. (1998) *Biophysical journal* **75**(1), 422-427

186. Walshaw, J., and Woolfson, D. N. (2001) *Journal of molecular biology* **307**(5), 1427-1450.

187. Kwok, S. C., and Hodges, R. S. (2004) *The Journal of biological chemistry* **279**(20), 21576-21588

188. Keating, A. E., Malashkevich, V. N., Tidor, B., and Kim, P. S. (2001) *Proceedings of the National Academy of Sciences of the United States of America* **98**(26), 14825-14830

189. Porta, J. M., Ros, L., Thomas, F., Corcho, F., Canto, J., and Perez, J. J. (2007) *Journal of computational chemistry* **28**(13), 2170-2189

190. Herrmann, H., Strelkov, S. V., Feja, B., Rogers, K. R., Brettel, M., Lustig, A., Haner, M., Parry, D. A., Steinert, P. M., Burkhard, P., and Aebi, U. (2000) *Journal of molecular biology* **298**(5), 817-832

191. Shen, W. (2005) Structure, dynamics, and properties of artificial protein hydrogels assembled through coiled-coil domains. In., Caltech Library, California Institute of Technology

192. Mo, J. M., Holtzer, M. E., and Holtzer, A. (1991) *Proceedings of the National Academy of Sciences of the United States of America* **88**(3), 916-920

193. Smith, T. A., Strelkov, S. V., Burkhard, P., Aebi, U., and Parry, D. A. (2002) *Journal of structural biology* **137**(1-2), 128-145.

194. Grishaev, A., Anthis, N. J., and Clore, G. M. (2012) *Journal of the American Chemical Society*

195. Feigin, L. A., and Svergun, D. I. (1987) Structure Analysis by Small-angle X-ray and Neutron Scattering. In., Plenum Press, New York

196. Vainshtein, B. K., Feigin, L. A., Lvov, Y. M., Gvozdev, R. I., Marakushev, S. A., and Likhtenshtein, G. I. (1980) *FEBS letters* **116**(1), 107-110

197. Georgakopoulou, S., Moller, D., Sachs, N., Herrmann, H., and Aebi, U. (2009) *Journal of molecular biology* **386**(2), 544-553

Curriculum Vitae

Anastasia A. Chernyatina

Date of birth: 21.06.1986
Place of birth: Lobnya, Moscow region, Russia
Email: Anastasia.Chernyatina@gmail.com

Education

November 2009 - December 2013: PhD student at the Laboratory for Biocrystallography, Department of Pharmaceutical and Pharmacological Sciences, KU Leuven, Belgium.
January 2009 - November 2009: Pre-doctoral student at the KU Leuven.
September 2003 - May 2008: BSc/MSc degree in Biophysics at the Lomonosov Moscow State University, School of Biology, Russia. State Academic scholarship. Diploma with honours.
September 2001- May 2003: Biophysics-mathematics-oriented courses at the Kolmogorov Advanced Educational Scientific Centre affiliated with the Moscow State University, Russia.

Research experience

KULeuven, Belgium, 2009-2013, Predoctoral and PhD student

- Intermediate filament structure determination using X-ray crystallography, Laboratory for Biocrystallography, supervisor Prof. S. Strelkov
    - cloning, purification and crystallization of intermediate filament proteins
    - crystallographic data collection using home rotating anode X-ray source (Rigaku) and at the synchrotrons in France (Soleil, ESRF), Switzerland (SLS) and Germany (DESY)
    - data acquisition and processing using Mar345, HKL 2000, XDS, Mosflm crystallographic software
    - structure determination using CCP4, Phenix, Shelx C/D/E, Phaser, ARP/wARP, autoSharp software
- Supervision of master students since 2011
- Professional training courses
    - Advances in automated analysis and validation of protein NMR structures, Hokkaido, Japan, 2013
    - EMBO Practical Course on Exploiting Anomalous Scattering in Macromolecular Structure Determination, Grenoble, France, 2011. Oral and poster presentations.
    - Crystallographic software Phenix workshop, Utrecht, The Netherlands, 2011
    - Nanobody and protein production training, Gent, Belgium, 2010
    - Courses on effective scientific presentation, writing and publishing, Leuven, Belgium
    - Mass spectrometry, NMR and molecular modelling courses, Leuven, Belgium

Moscow State University, Russia, 2003-2008

- Institute for Genetics and Selection of Industrial Microorganisms, 2007-2008
    - Studies of bacterial regulatory elements riboswitches, supervisor Prof. A. Mironov
- Ajinomoto-Genetika Research Institute, 2006-2007
    - Development of a regulatory system for gene expression, supervisor Prof. S. Mashko
- Faculty of Bioengineering and Bioinformatics, MSU, 2003-2004
    - Technical assistant in the laboratory of Prof. P. Vrzhesch

Languages

- mother tongue Russian;
- fluent spoken and written English;
- intermediate level of spoken and written French;
- basic knowledge of Spanish and Dutch.

List of publications

Journal publications

1. **Chernyatina AA**, Nicolet S, Aebi U, Herrmann H, Strelkov SV (2012) Atomic structure of the vimentin central alpha-helical domain and its implications for intermediate filament assembly. Proc Natl Acad Sci U S A 109: 13620-13625
2. **Chernyatina AA**, Strelkov SV (2012) Stabilization of vimentin coil2 fragment *via* an engineered disulfide. J Struct Biol 177: 46-53

Conference proceedings

1. **Chernyatina, A.**, Herrmann, H., Strelkov, S.; Crystallographic Studies of Vimentin and Lamin A Rod Domains. International Conference on Structural Genomics - Structural Life Science, Sapporo, Japan, July 29-August 1, 2013.
2. Strelkov, S., **Chernyatina, A.**, Aebi, U., Herrmann, H.; Atomic structure of IF dimer and its implications for filament assembly. Gordon Research Conference on Intermediate Filaments. Bates College, Lewiston MA, USA, June 17-22, 2012.
3. **Chernyatina, A.**, Nicolet, S., Strelkov, S.; Crystallographic Studies of Vimentin Rod Domain. European Conference on Intermediate Filaments in Health and Disease. Mykonos, Greece, 16-19 June 2011.
4. Strelkov, S., **Chernyatina, A.**, Nicolet, S., Herrmann, H.; Crystallographic studies of IFs: new structures and new surprises. Gordon Research Conference on Intermediate Filaments. Tilton, USA, 20-25 June 2010.
5. Nicolet, S., **Chernyatina, A.**, Herrmann, H., Aebi, U., Strelkov, S.; Atomic structure of intermediate filament dimer: recent advances. Alpbach Meeting on Coiled Coils, Collagen and Co-proteins. Alpbach, Austria, 6-12 September 2009.

Protein Data Bank (PDB) entries

3TRT	Chernyatina AA, Strelkov SV. Crystal structure of stabilized vimentin coil2 fragment with mutations L265C, L269(MSE), C328(MSE).
3SWK	Chernyatina AA, Strelkov SV. Crystal structure of vimentin coil1B fragment.
3S4R	Chernyatina AA, Strelkov SV. Crystal structure of vimentin coil1A/1B fragment with a stabilizing mutation Y117L.
3SSU	Nicolet S, Chernyatina AA, Strelkov SV. Crystal structure of vimentin coil1A/1B fragment.