Molecular modeling and structural characterization of a high glycinetyrosine hair keratin associated protein

Journal Name

ARTICLE

Molecular Modeling and Structural Characterization of a High Glycine-Tyrosine Hair Keratin Associated Protein

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Rakesh S. Singh,[a] Jeremy C. Palmer,[b] Paul D. A. Pudney,[c] Prem K. C. Paul,[d*] Christian Johannessen,[e] Pablo G. Debenedetti,[a] Janhavi Raut,[f] Ken Lee,[d] Massimo Noro,[d] David Tiemessen[c]

High glycine tyrosine (HGT) proteins are an important constituent of the keratin associated proteins (KAPs) present in human hair. The glassy state physics of hair fibers are thought to be largely regulated by KAPs, which exist in an amorphous state and are readily affected by environmental conditions. However, there are no studies characterizing the individual KAPs. In this paper, we present the first step to fill this gap by computational modeling and experimental studies on a HGT protein, KAP8.1. In particular, we have modeled the three-dimensional structure of this 63-residue protein using homology information from an anti-freeze protein in snow flea. The model for KAP8.1 is characterized by four strands of Poly-Proline II (or PPII) type helical secondary structures, held together by two Cysteine disulphide bridges. Computer simulations confirm the stability of the modelled structure and show that the protein largely samples the PPII and β-sheet conformations during the molecular dynamics simulations. Spectroscopic studies including Raman, IR and vibrational circular dichroism have also been performed on synthesized KAP8.1. The experimental studies suggest that KAP8.1 is characterised by β-sheet and PPII structures, largely consistent with the simulations studies. The model built in this work is a good starting point for further simulations to study in greater depth the glassy state physics of hair, including its water sorption isotherms, glass transition, and the effect of HGT proteins on KAP matrix plasticization. These results are a significant step towards our goal of understanding how the properties of hair can be affected and manipulated under different environmental conditions of temperature, humidity, ageing and small molecule additives.

Introduction

Human hair is a complex composite structure and is still not fully understood.[1-4] Proteins make up over 90% of the dry weight of human hair.[5] They exist in the cuticle and the cortex, and are responsible for almost all of the physical properties of hair fibers and consequent hair care benefits. Two of the most important protein classes present in the hair cortex are the intermediate filament proteins (IFs), and the keratin associated proteins (KAPs) also known as matrix proteins. The original two-phase model[2] of hair assumes that the rod-like IFs are embedded in a matrix of globular KAPs. Various refinements of the two-phase model that have been proposed in order to account for the mechanical properties of hair or wool fibres in terms of the properties of their IFs and KAPs, has been reviewed[1]. The relative merits of these models are still the subject of academic debate. It is therefore important to study the properties of isolated hair proteins in order to understand to what extent they can account for the physical properties of hair fibres. While much is known about the structure of keratin IFs,[6-7] there has been a distinct lack of structural information about the matrix of KAPs in which the IFs are embedded in the hair cortex. KAPs can be classed into three main groups: high-sulphur (HS) proteins, ultra high-sulphur (uHS) proteins and high glycine-tyrosine (HGT) proteins. Over 56 unique KAPs have been discovered,[8-9] however, very little is known about their properties. These proteins are found throughout the hair fiber. KAPs are thought to exist in an amorphous or glassy state and are readily affected by heat, humidity, ageing and the presence of small molecule actives, affecting the physical and mechanical properties of hair. Much of the rich physics of hair is thought to be regulated by the amorphous KAP matrix that embeds the IFs.

HGT proteins are present in different amounts in different keratinous substrates, giving rise to the observed differences in mechanical and physical properties of different keratinous materials.[10] This has led to the postulation of the internal plasticization theory[11-12] that states that the presence of differing amounts of hydrophobic HGT proteins in keratin substrates is responsible for the different glass transitions exhibited by these materials, and consequently their different physical and mechanical properties. As HGT proteins are in the amorphous matrix of keratin substrates that embed the IFs,

[a] Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ 08544
[b] Department of Chemical and Biomolecular Engineering, University of Houston, Houston, TX 77204
[c] Unilever R&D, Colworth Science Park, Sharnbrook, Bedfordshire, MK44 1LQ, UK
[d] Unilever R&D, Port Sunlight Laboratory, Quarry Road East, Wirral, CH63 3JW, UK
[e] Department of Chemistry, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium
[f] Unilever R&D, 64 Main Road, Whitefield, Bengaluru, 560 066, India

knowledge of their structural characteristics will be an important pre-requisite to understanding the glass-state physics of hair from a molecular standpoint. Here we have made a start by modelling the three-dimensional structure of the 63-residue HGT protein KAP8.1 using homology information from an anti-freeze protein in snow flea. We used molecular dynamics simulations, Raman spectroscopy, infrared spectroscopy (IR), and vibrational circular dichroism (VCD) to confirm the structural features of the modeled KAP8.1. Normally, to obtain a full 3D structure experimentally, X-Ray diffraction or NMR would have been the ideal techniques to use. However as KAP8.1 is insoluble in virtually all solvents, especially water, this was not possible and thus vibrational spectroscopic techniques were chosen to characterize KAP8.1. In the future we hope to build a model of KAP proteins that is representative of the human hair matrix, so as to make amenable to computational modeling the rich physics of hair, such as its glass transition as a function of moisture content (i.e. the Fox equation[11-12]), its ageing and internal plasticization, and the effects of various hair additives and actives, as well as its water sorption characteristics as a function of temperature.

Results and discussion

Computational modelling and molecular simulation details

Initial modelling was performed on Discover[13] Studio 4, using the CHARMM force field.[14] The sequence of KAP8.1 was obtained from SWISSPROT database[15] and the 3-dimensional structure of an anti-freeze protein[16] from snow-flea (3BOG) was obtained from the RCSB protein data bank.[17] The rationale behind this choice of template sequence for KAP8.1 modelling is explained in detail in the **Results** section.

The molecular dynamics (MD) simulations of KAP8.1 protein in bulk water were performed using GROMACS software package.[18-20] The protein was modeled using CHARM27 force field[21] with the compatible SPC/E water model.[22] All the bonds were constrained using the LINCS (linear constraint solver) algorithm.[23,24] To maintain constant temperature, the Nosé-Hoover thermostat[25,26] with a 0.2 picosecond (ps) relaxation time was used. The desired pressure was maintained using a Parrinello-Rahman barostat[27] with a 2 ps relaxation time. Periodic boundary conditions were applied in all the three dimensions. The short-range interactions were truncated at 10 Å and long range electrostatic terms were computed using the particle mesh Ewald[28] summation, with a grid spacing of 1.6 Å. The standard long-range dispersion corrections were used for the energy and pressure.[29] The leapfrog algorithm[29] was used to integrate the equations of motion with a time step of 2 femtoseconds (fs). The system was first equilibrated in the NPT ensemble at a temperature 300 K and 1 bar pressure for 400 ns, followed by a production run of 400 ns in the NVT ensemble.

Synthesis and structural characterization of KAP8.1

Synthesis of KAP8.1 protein: This 63-residue protein was synthesized by Severn Biotech Ltd (Worcs, DY11 6TJ, UK). The synthesized protein was supplied at a purity of 95% with analytical HPLC and MALDI-TOF validation.

Vibrational Spectroscopy (Raman, IR and VCD)

Raman spectra of the synthesized KAP8.1 samples were obtained using a WiTec Alpha 300R Confocal Raman Spectrometer. A laser wavelength of 785 nm and power of 50 mW were used, along with a 100 x Zeiss Objective. A total collection time of 15 minutes was used (60 s x 15 accumulations) to obtain high signal to noise ratio spectra. The spectra obtained were truncated to a range of 350 to 1800 cm^{-1} and then baselined. Peak fitting of the amide band was performed using GRAMS AI software. For peak fitting of the amide I band the spectra were truncated to a range of 1560 to 1735 cm^{-1} and baselined. The IR spectra were collected on a Biorad (now Agilent) FTS 6000 spectrometer with liquid nitrogen-cooled mercury cadmium telluride (MCT) detector at 4 cm^{-1} resolution and 500 scans. A micro-diamond attenuated total reflectance (ATR) set-up was used ('Golden Gate', Greasby Specac Ltd, BR5 4HE, UK). A background scan of the clean diamond was run, then the sample was put onto the diamond and slight pressure was applied by a screw device to ensure good contact with the diamond surface.

IR and VCD spectra of KAP8.1 in KBr pellets were obtained using a Biotools® ChiralIR-2X in transmission with a liquid nitrogen-cooled MCT detector (ATR is not possible with VCD). This was done with a rotating cell-holder and the spectra were collected at 4 cm^{-1} resolution and 20,000 scans. The KBr pellets were produced as is traditional in IR spectroscopy, as the pure powder would adsorb too much of the IR radiation.[30] This was done by mixing 0.5% (w/w) protein with spectroscopy-grade KBr, grinding the powder for 5 min. in an agate mortar for high macroscopic homogeneity. The dry mixture was subsequently loaded into a Greasby Specac Ltd micro-anvil and subjected to 10 Tonnes of pressure for 5 minutes in a Greasby Specac Ltd pellet press, before transferring the pellet to a standard pellet holder. The atmospheric background was subtracted from the IR spectrum during measurements and a VCD baseline was subtracted after measurement by recording a pure KBr pellet VCD spectrum under identical experimental conditions.

Results

Molecular model of KAP8.1

KAP8.1 is a 63-residue protein with a high glycine (G) and tyrosine (Y) content, the sequence of which is

MLCDNFPGAVFPGCYWGSYGYPLGYSVGCGYGSTYSPVGY
GFGYGYNGCGAFGYRRYSPFALY

This protein has 4 Cysteines (C) and is possibly held together by two disulfide linkages. The first step in the modelling was to find structures of proteins from the RCSB protein data bank with similar amino acid sequence to KAP8.1. No structure with greater than 50% homology was found. However, as KAP8.1 has high glycine content and only 4 cysteine residues, structures with high glycine content and very few cysteine were thought to be likely template structures for the target KAP8.1 protein. As will be shown later, a distinctive feature of this study is the cross-validation of the modelled structure with computer simulations and experiments performed on synthesized KAP8.1.

It is anticipated that a large amount of secondary structure would be dictated by the high glycine content of such proteins, which are normally characterized by polyproline II (PPII) or poly-glycine helices. A good match from the protein data bank was indeed found and this is the anti-freeze protein from snow flea called 3BOG, the crystal structure of which has been solved[16] and is characterized by PPII (or poly-glycine) helices with (φ-ψ) of (-70°, 140°) as shown in Fig. 1.

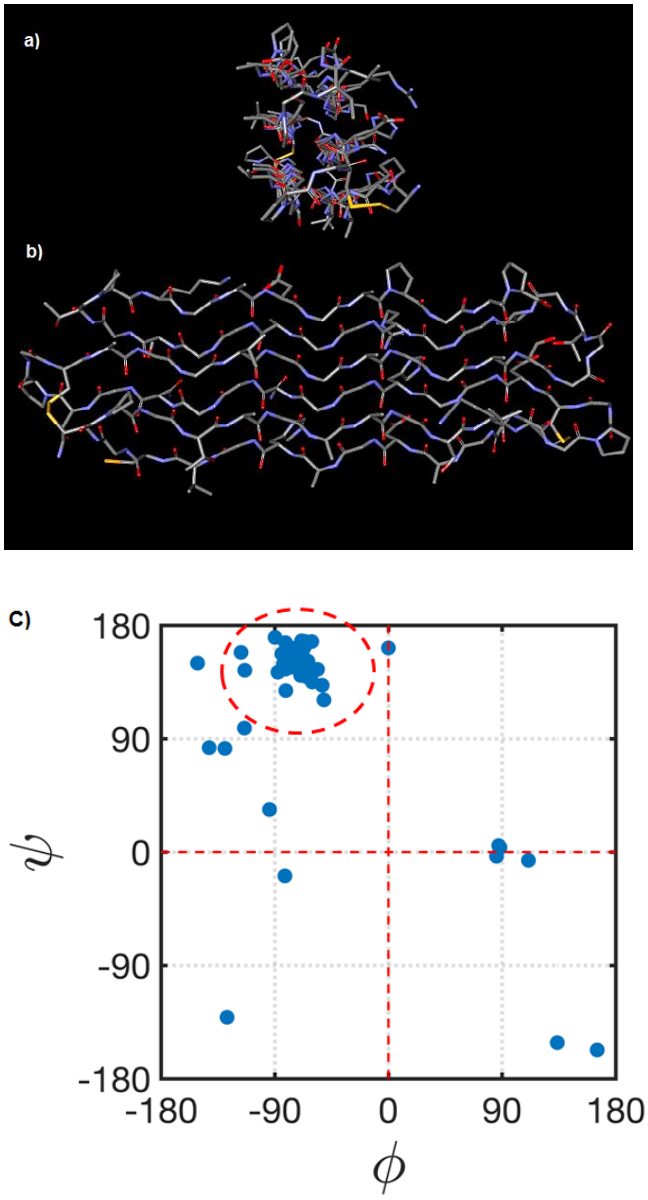

Figure 1. (a) end on and (b) side on views of snow flea anti-freeze protein16 3BOG showing clearly the PPII helices that this protein adopts. Colour code: carbon (grey), nitrogen (blue), oxygen (red) and Sulphur (yellow). (c) Ramachandran map[31] of 3BOG with the distribution of conformations is shown. The high concentration of points near (-70°, 140°) represents the PPII or poly-glycine helical conformation.

Though the two proteins, KAP8.1 and 3BOG, have only ~40% sequence homology, it is expected that the high glycine content and the consequent structural preference would make 3BOG a good template for the structure of KAP8.1. Moreover, as the cysteine residues are disposed in roughly similar areas of the protein, it is expected that the disulfide linkages for KAP8.1 would occur between similarly placed cysteine residues. This feature is illustrated in Fig. 2.

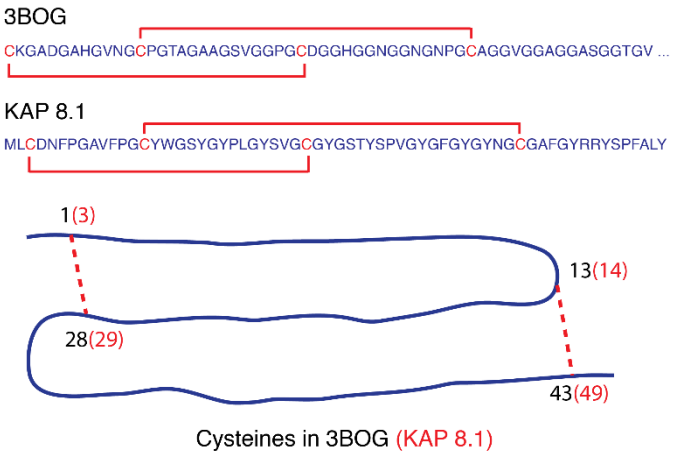

Figure 2. Amino acid sequences and disulfide linkages of 3BOG and KAP8.1. A schematic representation of KAP8.1 with the anticipated disulfide linkages (dashed red lines) is also shown.

In 3BOG the two disulfides are between residues 1-28 and 13-43. In KAP8.1 cysteines are present at positions 3, 14, 29 and 49. The first assumption made in the homology modelling was that the disulphide linkages in KAP8.1 would be between 3-29 and 14-49. This means that in the first cyclic part of the KAP8.1 structure there would be one less amino acid and in the second cyclic part there would be 5 more amino acid residues when compared with 3BOG (see Fig. 2).

Thus, the model for KAP8.1 using 3BOG as a template would have to account for the exact number of residues in the two cyclic structures with consistent β-turns and without introducing unstable short contacts. This was successfully done by employing the following steps:

(a) Starting with the structure of 3BOG, in-silico mutations, additions and deletions using Discovery Studio software were performed to obtain the final KAP8.1 structure.

(b) As KAP8.1 contains many large residues, in the first instance all residues facing away from the 3BOG protein structure were mapped and mutated to larger residues or aromatic residues such as tryptophan, tyrosine and phenylalanine. This was done to avoid unnecessary ring concatenations and spearing of the aromatic side chains within the protein bulk resulting in unnatural structures.

(c) The previous step led to a partial disruption of the original sequence. This was overcome by judiciously inserting and deleting small residues such as glycine or alanine to obtain the final correct sequence count. The structure was partially minimized after each insertion or deletion step.

(d) The inserted small residues such as glycine or alanine were mutated to the correct residue (serine, threonine, leucine etc.) as per the actual sequence of KAP8.1 to obtain the full KAP8.1 structure. Once again partial energy minimization of the structure was performed after each mutation.

(e) Finally, all residues of KAP8.1 model were checked to have the correct chirality (L) and all peptide bonds were ensured to have the trans configuration. This required some editing and subsequent minimizing of the KAP8.1 model structure. The final structure of modeled KAP8.1 is shown in Fig. 3 and was seen to have no unstable short contacts.

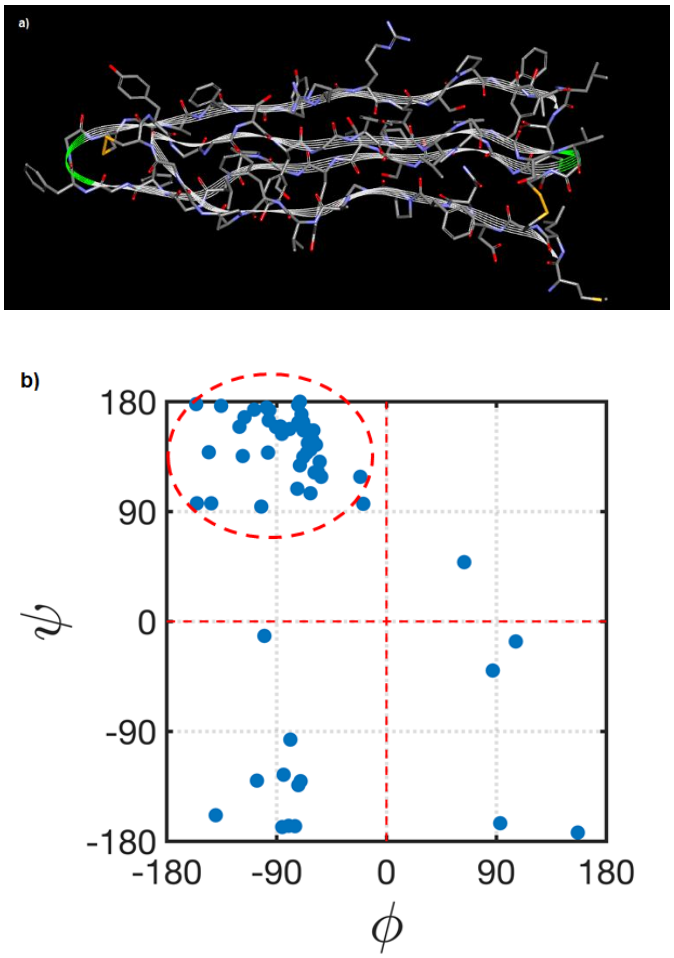

Figure 3. (a) Model of KAP8.1 starting from 3BOG as template. This structure is free of unstable short contacts. The ribbon depicts the protein backbone structure. (b) Ramachandran plot showing the secondary structure conformations largely in the PPII helix and β-sheet (indicated by the dotted red circle).

Computer simulations of the KAP8.1 model

In order to check the stability of this modeled 3D structure of KAP8.1 in aqueous environment, molecular dynamics simulations of a single protein unit in bulk water were performed. The root mean square deviation (RMSD) was computed during the course of the simulation. RMSD is a measure of the protein's structural change (usually restricted to the backbone atoms) with respect to a reference structure in the course of the simulation. In Fig. 4a we show the KAP8.1 backbone RMSD fluctuation during simulation for a period of 200 ns after initial equilibration. It can be seen that the RMSD of the KAP8.1 backbone is reasonably stable around 0.25 nm. This suggests that the modeled 3D structure of KAP8.1 is stable in aqueous environment.

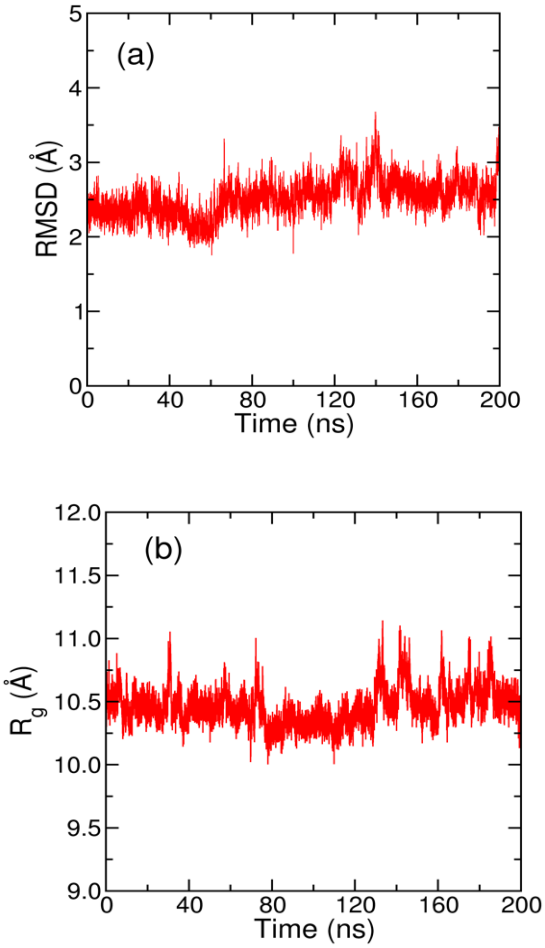

Figure 4: (a) Protein backbone root mean square deviation (RMSD) and (b) radius of gyration (R_g) of KAP8.1 protein during the course of molecular dynamics simulation at temperature 300K and 1 bar pressure.

The radius of gyration (R_g) of a protein is a measure of its compactness. If a protein structure is stable, it will likely maintain a relatively steady average value of R_g. If a protein folds or unfolds during the simulation, its R_g will change over time. In Fig. 4b, we show the radius of gyration for KAP8.1 during the simulation. We can see from the reasonably constant R_g values that the protein remains stable and maintains globular structure during the simulation. It should be noted that RMSD and R_g provide information about the global (not the local) structural change of a protein during the simulation.

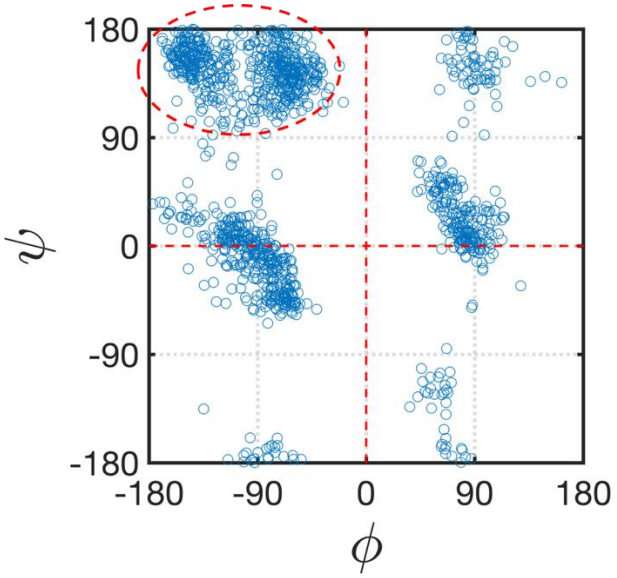

Figure 5: Ramachandran plot obtained from an equilibrated molecular dynamics trajectory of a single KAP8.1 in bulk water at temperature 300K and 1 bar pressure. The secondary structures are largely in the PPII helix and β-sheet regions (indicated by the red circle) of the Ramachandran plot.

To gain microscopic insight into the structural features of KAP8.1, in Fig. 5 we show the Ramachandran map[31] of equilibrated protein structures in bulk water. It is evident from the map that along with PPII helices (structural characteristics of the original model) more conformations are being explored in the β-sheet region with the inclusion of water in the simulations. We performed a clustering analysis to gain a deeper insight on whether each protein structure either shares the features of both PPII helix and β-sheet, or two distinct structures - one dominated by PPII helix and other by β-sheet, existing in dynamic equilibrium. Both of these scenarios will give rise to a Ramachandran map similar to the one shown in Fig. 5. Clustering analysis uses RMSD similarity as a criterion to partition individual protein structures into clusters. We used the g_cluster tool of GROMACS [18-20] with the GROMOS clustering algorithm,[32] with a 0.12 nm C^{α} RMSD cut-off to determine clusters (see Supplementary Materials for details). Ramachandran plots of the central protein structures of top three most populated clusters, that consist of > 70 % of the total protein structures, suggest that dominant protein structures in water share the characteristics of both PPII helix and β-sheet (see Supplementary Materials).

We have not used standard algorithms like DSSP[33] (Dictionary of Protein Secondary Structure) and STRIDE[34] (Structural Identification) to assign the secondary structure content of our modelled KAP8.1, as these algorithms rely on protein backbone hydrogen bond patterns and PPII helices do not consist any characteristic hydrogen bond patterns. We also must note that defining secondary structures from atomic resolution coordinates is not an exact process due to differences in the definition of secondary structures.

Assignment	KAP8.1 Peak (cm^{-1})
ν (SS) (gauche-gauche-gauche) Disulphide bond	505
Phenylalanine	619
Tyrosine	642
ν (CS) from methionine (trans form)	723
Tryptophan	758
δ (CCH) aliphatic (Tyrosine)	828
δ (CCH) aromatic (Tyrosine)	851
ν (CC) helix	933
ν (CC) aromatic ring/ Phenylalanine	1003
ν (CC) skeletal/cis conformation, Phenylalanine	1031
ν (CC) skeletal	1127
Tyrosine (CH)	1207
δ (NH) (NC) Amide III -sheet	1230
δ (NH) (NC) Amide III disordered	1244
δ (C H) Tryptophan	1340
δ CH_2	1446
Indole ring (Tryptophan)	1551
Aromatic (Tyrosine, Tryptophan, Phenylalanine)	1614
ν (CO) Amide I	1635-1700

Table 1: Assigned peaks of the Raman spectrum of KAP8.1 shown in Fig. 6. Symbols and represent (ν) stretching and (δ) bending modes.

Vibrational spectroscopic studies on KAP8.1

The Raman spectrum of KAP8.1 is shown in Fig. 6; the assignments of the corresponding peaks are listed in Table 1. The Raman spectrum for KAP8.1 shows a protein with a high tyrosine content (intense peaks at 642, 828, 851, 1207, 1614 cm^{-1}), which is consistent with the protein's sequence.

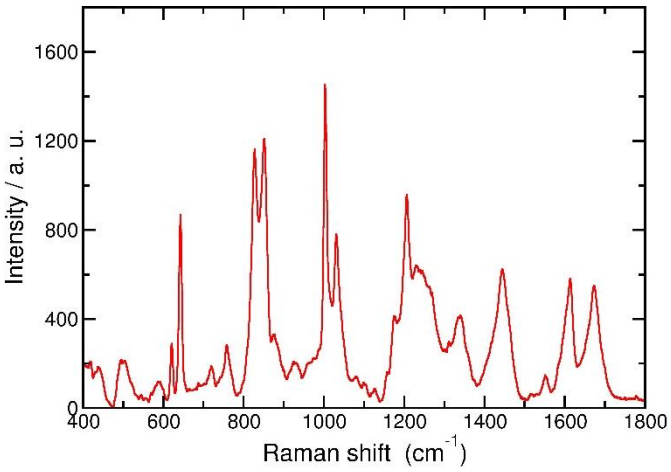

Figure 6: Raman spectrum of KAP8.1 powder obtained with a WiTec Alpha 300R Confocal Raman Spectrometer. A laser wavelength of 785nm and power of 50mW were used, along with a 100x Zeiss Objective.

What is particularly interesting in the spectrum is the Fermi doublet ratio (I_{850}/I_{830}): this ratio is greater than 1, which indicates[35] that the tyrosine residues are exposed. This implies that they would be pointing outwards, away from the centre of the protein. This is consistent with the predicted structure shown in Fig. 3, where the exposed tyrosine residues are clearly evident. The phenylalanine-glycine ratio can be estimated from the Raman spectrum using the following empirical formula[36]

$$1.25\frac{I_{624}}{I_{644}} = \frac{Phe}{Tyr}$$

For KAP8.1, the ratio calculated from the Raman spectrum is 0.44, and the actual ratio from the known sequence is 0.42 (12 Tyrosine and 5 Phenylalanine).

The amide I band is especially sensitive to different conformations.[37] This is because the band arises due to the stretch of the C=O bond, which is involved in hydrogen bonding, as well as a small contribution from an out-of-phase C-N stretch. The different hydrogen bonds formed as a result of the conformation cause the amide band to shift. Thus, the broad peaks are considered to be formed of smaller peaks, each assigned to a specific secondary structure conformation. There is a large literature on these assignments. [38-41] In order to determine the relative amounts of α-helix, β-sheet, PPII and other secondary structure elements present, the amide I band must be peak-fitted. The amide I band with underlying fitted peaks is shown in Fig. 7. The calculated amount of each secondary structure element is given in Table 2.

There is some uncertainty on the assignments shown in Table 2. The α-helix peak assignment for the amide I band is universally stated as being at 1650 cm^{-1}, and there is no debate in the literature regarding this peak assignment. This is because the structure is well defined and regular. The β-sheet band assignments depend on the various types of β-sheet present. There are parallel, anti-parallel and extended-strand β-sheets, which all have slightly different band assignments. Parallel β-sheet is assigned to 1635 cm^{-1} and anti-parallel and extended β-sheet conformations are assigned to 1670 cm^{-1}. The recent work by Anderson and coworkers[42] states that a strong band is only observed in the 1635 cm^{-1} region when there is large coupling between multiple β-strands. Most general β-sheet assignments in the literature[39,42-44] are in the 1670 cm^{-1} region of the amide I band.

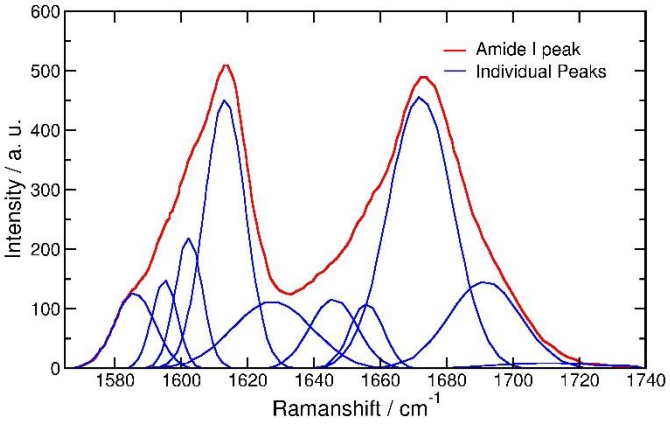

Figure 7: Peak fitting of amide I band of the Raman spectrum of KAP8.1 shown in Fig. 6

Table 2: Assigned peaks of the Raman amide I band shown in Fig. 7

Position (cm^{-1})	Area (%)	Assignment
1627	15	β-sheet
1646	10	α-helix
1656	6	Random Coils/PPII
1672	50	β-sheet* (or PPII)
1692	18	Disordered/turns

*This is confirmed to be β-sheet from the VCD results shown in Fig. 9

There is uncertainty around the amide I peak assignment regarding the PPII structure. There is very little experimental evidence for the PPII assignments. Most of the literature assignments are based on theoretical calculations. In the literature there are two regions of the amide I band associated with PPII, the 1660-1670 cm^{-1} and 1675-1685 cm^{-1} regions.[38,39,42] Therefore, the band centred at 1672 cm^{-1} could be from either β-sheet or PPII. Further evidence supporting one assignment or the other is provided in the next section.

IR and VCD studies of KAP8.1

In order to remove the uncertainty over assignment of β-sheet and PPII helices from the Raman spectrum additional spectroscopic experiments (IR and VCD) were performed on KAP8.1. The results are shown in Figs. 8 and 9. Fig. 8 and the top spectrum in Fig. 9

show that the IR spectrum of KAP8.1 powder taken in attenuated total reflection (ATR) mode, and in a KBr Pellet in transmission mode resemble each other closely. The amide I region of the IR spectrum, like that of the Raman spectrum, is sensitive to the different protein conformations. It is considered to consist of overlapping bands that come from the different secondary structures. Thus, resemblance of spectra taken in the two different modes clearly indicates that the protein retains its native conformation despite being subjected to pressure during pellet production. As with the Raman spectra, the assignment within the amide I region is not able to clearly differentiate between β-sheet and PPII structures. Vibrational circular dichroism (VCD) studies provide the extra level of sensitivity needed to distinguish these energetically similar structures. The VCD spectrum is spectrally rich in the amide I-III regions. The VCD spectrum of KAP8.1 (Fig. 9) clearly indicates a high proportion of β-sheet, as the observed pattern in the amide I and II regions, i.e., negative-positive-negative, constitutes the classical pattern found in β-sheet-rich proteins.[45]

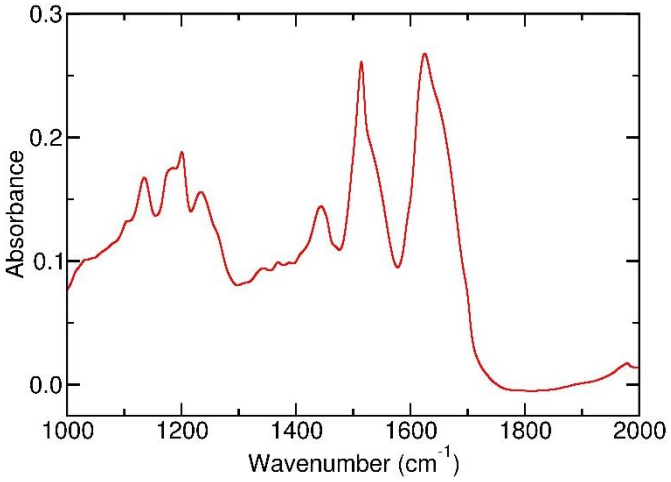

Figure 8. IR-ATR spectrum of KAP8.1 taken on a Biorad (now Agilent) FTS 6000 spectrometer with liquid nitrogen-cooled MCT detector at 4 cm^{-1} resolution and 500 scans. A micro diamond ATR set-up was used ('Golden Gate', Greasby Specac Ltd).

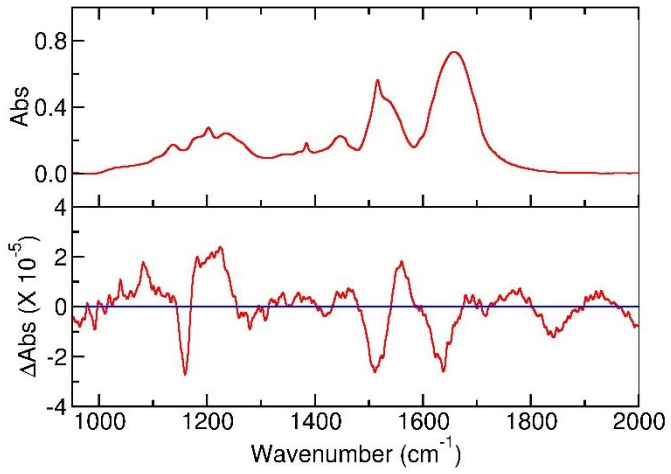

Figure 9. FT-IR (top) and VCD (bottom) spectra of KAP8.1 taken on a Biotools® Chiral IR-2X spectrometer using a liquid nitrogen-cooled MCT detector. They were taken in transmission mode with the KAP8.1 in a KBr pellet in a rotating cell holder. The spectra were collected at 4 cm^{-1} resolution and 20,000 scans

Furthermore, the amide III region of the VCD spectrum contains a band pattern (again negative-positive-negative) that is also found in β-sheet proteins.[45] The VCD bands in the amide III region are usually strong, which is probably due to the restrained dynamics of the protein under dry conditions. Although not quantitative, the VCD result does resolve the uncertainty of Raman and IR assignments due to overlapping band positions. Therefore, the band at 1672 cm^{-1} in Table 2 from the Raman spectrum is assigned to β-sheet as opposed to PPII. Thus, the proportions of secondary structures present can be taken as those given in Table 2 *i.e.* 65% β-sheet, 18% disordered/turn, 10%, α-helix, 6% random coil/PPII.

Discussion

This study uses computational modeling but is underpinned by verification with experimental measurements to explore the structure and properties of a particular keratin associated protein, KAP8.1, which is an important constituent of KAP matrix and about which very little is known. We have chosen to use vibrational spectroscopy for the verification of the structural features of the modelled KAP8.1, as it is an extremely powerful analytical technique for determining the molecular structure of simple and complex macromolecules. IR and Raman spectroscopy have been used for many years to examine the structure of proteins, especially their secondary structures.[40] These two related vibrational spectroscopic techniques possess many advantages for studying biological molecules, including having no molecular size limitations, being label-free, and possessing wide applicability to different sample conditions and types. As KAP8.1 is insoluble in virtually all solvents, especially water, vibrational spectroscopy is ideal as it can look at samples in all states and provides a large amount of structural information.[40,46,47] Here we use Raman spectroscopy, infrared spectroscopy (IR) and also vibrational circular dichroism (VCD) as it gives an extra level of differentiation among protein secondary structures.[45] Hence vibrational spectroscopy methods are particularly effective when crystals of proteins are hard to obtain[46] or when they are quite insoluble. They have been used to study keratin[48] and whole hairs,[43,49] including how their behavior changes under various conditions.[50] We have also used VCD, the chiral form of IR spectroscopy, as it gives an extra level of sensitivity that enables differentiation among secondary structures that are close in frequency in the linear forms of IR and Raman. To our knowledge, this is also the first time that VCD has been performed successfully on a solid protein sample in the native state.

The results of our computer simulations on the modeled protein and of concurrent spectroscopic studies on synthesized KAP8.1 suggest that the structure of this protein is characterized by either PPII helices and/or β-sheet-type structures with most of the Tyrosine residues facing outwards. While the modeling studies in vacuo and in water suggest a PPII structure with some β-sheet content (Figs. 3b and 5), the spectroscopic studies confirmed by VCD suggest predominantly β-sheet content. This slight discrepancy may be explained by the fact that these two regions are largely contiguous in the upper left hand corner of the Ramachandran map and conformational excursions between the two are possible. This is even more so if the sequence includes glycine residues and very few

proline residues. As is well known, proline is a cyclic amino acid and is constrained to have a ϕ value close to -65°. A β-sheet ϕ value of ~-140° is impossible for proline to adopt, and would result in a very high energy barrier between the PPII and the β-sheet conformations if the sequence contained a high proportion of proline residues. In KAP8.1, the sequence in the regions that we focus on has very few proline residues, and contains a high proportion of glycine. Moreover, the energy difference between PPII and β-sheet for many peptide sequences is vanishingly small[51] in water, and hence in a well equilibrated MD study we would expect both regions to be explored during the simulation. Indeed, this small energy difference is also the reason why these two secondary structural elements have similar frequencies in IR and Raman spectra. Several papers[52-55] discuss the propensity of small peptides to adopt PPII or β-sheet conformations that are influenced by the amino acid sequence and/or the presence of water. In any event, as the main aim of the homology model was to build a good starting structure for large scale computational studies, the model described in this paper with the accompanying spectroscopic support can be considered very good indeed.

Here for the first time a model of an important KAP protein - the HGT protein KAP8.1 has been described in such detail. This work has highlighted that while there are many physicochemical methods used to study and understand hair structure and function, computer simulations offer an extremely useful method, especially at molecular length scales. In the past, most of the modeling/simulations of hair proteins have been confined to the study of IFs – historically because the molecular structures were easy to build, and many template structures of coiled coils such as vimentin[56,57] are known. However, there have been no reported papers of modeling and simulations of the hair matrix proteins or KAPs and we hope that this will be the first of many papers in this area.

The model of KAP8.1 protein described here and its satisfactory confirmation from vibrational spectroscopy is an excellent starting point for further simulation studies on the KAP matrix and the inherent properties of hair. For example, it is postulated that the denaturation temperature of hair in the wet is influenced by the viscosity of the matrix proteins.[58] Also, as mentioned above, the glass transition of hair is a matrix property and is heavily influenced by water, ageing, and the presence of small additives. In follow-up studies we will investigate water sorption and glass transition phenomena of KAP8.1 using advanced simulation methods.

Finally, as mentioned previously, apart from the two-phase model, various refinements[1] have been proposed to describe the mechanical properties of hair or wool fibres in terms of the properties of their IFs and KAPs and this continues to be a subject of academic debate. All of these models involve α to β transitions (i.e. the unfolding of α helical regions and formation of β sheets) in the IFs above certain strain levels. For example, one model known as the series-zone model involves a second set of α to β transitions to explain the hardening of the fibre in the high strain regions found in typical stress-strain curves. This model treats the matrix as being thixotropic, affecting the ease of deformation, and opposing both folding and unfolding of the α helices. The series-zone model does not incorporate interactions between the IFs and the matrix proteins. However, other models involve the role of the matrix, and the role of interactions between the matrix proteins and the IFs. For instance, the Chapman-Hearle model[1] assumes the presence of disulphide bonds between KAPs, and between KAPs and IFs. These bonds are described as bearing a tensile stress at strains below the strain hardening region of the fibre stress/strain curve. This is claimed to originate from stress transfer between IFs and KAPs. Thus, understanding the structures of KAPs, and their contribution to the glassy state physics of hair will shed light on questions such as the nature of KAP-IF and KAP-KAP (both HGT-HS and HS-HS) interactions. For example, whether covalent KAP-IF and KAP-KAP interactions are required to explain the mechanical properties of hair fibres. Understanding the properties of the KAPs, and their interaction with each other and with IFs would be crucial for testing the validity of some models. In particular, this would help in understanding the role of the matrix in the plateau, strain hardening and recovery regions of the stress strain curve. Moreover, studying non-covalent KAP-IF interaction may provide information on the role of KAPs in determining IF assembly. A recent study[59] has demonstrated the effects of KAP8.1 on IF assembly in vitro. That study proposes that the KAP8.1-IF interaction is involved in the formation of ortho cortex cells rather than para cortex cells. Hence, understanding HGT KAPs and their interactions with IFs may be important in studying the morphology of keratin fibres, in addition to their mechanical properties. The approach of using modelling to develop our understanding of individual hair components and how they interact will contribute to the aim of providing a holistic understanding of the hair fibre, and the manipulation of its properties under different conditions.

Conclusions

In this work, a model of a HGT hair protein, KAP8.1, has been built using an anti-freeze protein 3BOG from snow flea as a template, which is characterized by four strands of PPII-type helical secondary structures held together by two Cysteine di-sulphide bridges. The initial modeled structure of KAP8.1 contained characteristics of both the PPII as well as **β**-sheet structures. We have used this as a starting point for further simulations and confirmed the stability of modeled structure by molecular dynamics studies. Raman, IR and VCD spectroscopic studies performed on a 63-residue synthesized KAP8.1 show that the protein is characterized by secondary structures largely consistent with the computational modeling.

The model built in this work can be a good starting point for simulations to study the hydration of KAP8.1, dehydration-induced glass transition, and the influence of small molecule additives on these processes and properties. It is hoped that this model will enable us to gain important insights into the glassy nature of hair matrix proteins, and how the properties of hair can be affected under different environmental conditions of temperature and humidity.

Acknowledgements

We thank Unilever Research & Development for funding.

References

1. J. W. Hearle, *Int. J. Biol. Macromolec*., 2000, **27**, 123.
2. M. Feughelman, *J. Appl. Polym. Sci*., 2001, **83**, 489.
3. C. Popescu and H. Hocker, *Chem. Soc. Rev*. 2007, **36**, 1282.
4. F. J. Wortmann, *Handbook of Textile Fiber Structure (Eicholm, S. et al. eds.)*, 2009, **2**, 108.
5. C. R. Robbins, *Chemical and physical behavior of human hair*, Springer, New York, USA, 2012

6. M. E. Rafik, J. Doucet and F. Briki, *Biophys. J.*, 2004, **86**, 3893.
7. R. D. B. Fraser, T. P. MacRae, D. A. D. Parry and E. Suzuki, *Proc. Natl. Acad. Sci. USA* ,1986, **83**, 1179.
8. M. A. Rogers, L. Langbein, S. Praetzel-Wunder, H. Winter and J. Schweizer, *Int. Rev. Cytol*., 2006, **251**, 209.
9. Y. Shimomura and M. Ito, J. Invest. Dermatol., 2005, **10**, 230.
10. E. G. Bendit and J. M. Gillespie, *Biopolymers* 1978, **17**, 2743.
11. F. J. Wortmann, M. Stapels, R. Elliott and L. Chandra, *Biopolymers*., 2006, **81**, 371.
12. I. Jinks, P. Paul and F. J. Wortmann, *Thermochimica Acta*., 2015, **614**, 33.
13. *Discover Studio 4 is a software product from: Biovia*, 5005 Wateridge Vista Drive, San Diego, CA 92121, USA
14. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comp. Chem.*, 1983, **4**, 187.
15. M. Magrane, *UniProt consortium Database(Oxford)*, 2011: bar009.
16. B. L. Pentelute, Z. P. Gates, V. Tereshko, J. L. Dashnau, J. M. Vanderkooi, A. A. Kossiakoff and S. B. Kent, *J. Am. Chem. Soc.*, 2008, **130**, 9695.
17. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235.
18. B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4**, 435.
19. D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. Berendsen, *J. Comput. Chem.*, 2005, **26**, 1701.
20. H. J. Berendsen, D. van der Spoel and R. van Drunen, *Comput. Phys. Commun.*, 1995, **91**, 43.
21. Jr. A. D. MacKerell, N. Banavali and N. Foloppe, *Biopolymers*, 2001, **56**, 257.
22. H. J. C. Berendsen, J. R. Grigera and T. P. Straatsma, *J. Phys. Chem.*, 1987, **91**, 6269.
23. B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463.
24. B. Hess, *J. Chem. Theory Comput.*, 2008, **4**, 116.
25. S. A. Nosé, *J. Chem. Phys.*, 1984, **81**, 511.
26. W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695.
27. M. Parrinello and A. Rahman, *J. Appl. Phys.* 1981, **52**, 7182.
28. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577.
29. M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*; Oxford University Press: New York, 1989.
30. J. M. Chalmers and G. Dent, *Industrial analysis with vibrational spectroscopy*, Royal Soc. Chem. ISBN 0-85404-565-1, 1999, P-136.
31. G. N. Ramachandran and V. Sasiskharan, *Adv. Protein Chem.*, 1968, **23**, 283.
32. X. Daura, K. Gademann, B. Jaun, D. Seebach, W.F. van Gunsteren, A.E. Mark, *Angew. Chem. Int. Ed. Engl*., 1999, **38**, 236.
33. W. Kabsch and C. Sander, *Biopolymers*, 1983, **22**, 2577.
34. D. Frishman and P. Argos, *Proteins: Structure, Function and Genetics*, 1995, **23**, 566.
35. P. G. Hildebrandt, R. A. Copeland, T. G. Spiro, J. Otlewski, M. Laskowski and F. G. Prendergast, *Biochemistry*, 1988, **27**, 5426.
36. W. S. Craig and B. P. Gaber, *J. Am. Chem. Soc.*, 1977, **99**, 4130.
37. J. Bandekar, *Biochim. Biophys. Acta* ,1992, **123**, 1120.
38. R. Tuma, *J. Raman Spectrosc.*, 2005, **36**, 307.
39. R. Schweitzer-Stenner, *J. Phys. Chem. B*, 2004, **108**, 16965.
40. A. Barth and C. Q. Zscherp, *Rev. Biophys.* 2002, **35**, 369.
41. R. Schweitzer-Stenner, *Vib. Spectrosc*., 2006, **42**, 98.
42. N. C. Maiti, M. M. Apetri, M. G. Zagorski, P. R. Carey and V. E. Anderson, *J. Am. Chem. Soc.*, 2004, **126**, 2399.
43. P. D. A. Pudney, E. Y. M. Bonnist, K. J. Mutch, R. Nicholls, H. Rieley and S. Stanfield, *Appl. Spectrosc.*, 2013, **67**, 1408.
44. A. Rygula, K. Majzner, K. M. Marzec, A. Kaczor, M. Pilarczyk and M. Baranska, *J. Raman Spectrosc.*, 2013, **44**, 1061.
45. T. A. Keiderling, *Curr. Opin. Chem. Biol*., 2002, **6**, 682.
46. L. Ashton, P. D. A. Pudney, E. W. Blanch and G. E. Yakubov, *Adv. Colloid Interface Sci.*, 2013, **199**, 66.
47. P. D. A. Pudney, S. L. Buckley, C. M. Sidebottom, S. N. Twigg, M. P. Sevilla, D. Roper, C.B. Holt, J. H. Telford, A. J. McArther and P. J. Lillford, *Arch. Biochem Biophys.*, 2003, **410**, 238.
48. A. C. Williams, H. G. M. Edwards and B. W. Barry, *J. Raman Spectrosc.*, 1994, **25**, 95.
49. K. L. A. Chan, S. G. Kazarian, A. Mavraki and D. R. Williams, *Applied Spec.*, 2005, **59**, 149.
50. F. I. Bell, R. Skinner, I. M. Tucker, Y. Leray, T. E. Lyons, K. Devine, P. Pudeny and T. Oikawa, *J. Cosmet. Sci.*, 2004, **55**, S19.
51. F. Avbelj and S. G. Grdadolnik, *Protein Sci.*, 2007, **16**, 273.
52. D. Meral, S. Toal, R. Schweitzer-Stenner and B. J. Urbanc, *J. Phys. Chem. B*, 2015, **119**, 13237.
53. N. V. Ilawe, A. E. Raeber, R. Schweitzer-Stenner, E. Siobhan, S. E. Toal and B. M. Wong, *Phys. Chem. Chem. Phys.*, 2015, **17**, 24917.
54. Z. Shi, K. Chen, Z. Liu, A. Ng, W. C. Bracken and N. R. Kallenbach, *Proc. Natl. Acad. Sci. USA* 2005, **102**, 17964.
55. P. J. Fleming, N. C. Fitzkee, M. Mezei, R. Srinivasan and G. D. Rose, *Protein Sci.*, 2005, **14**, 111.
56. S. V. Strelkov, H. Herrmann, N. Geisler, T. Wedig, R. Zimbelmann, U. Aebi and P. Burkhard, *EMBO J.*, 2002, **21**, 1255.
57. S. Nicolet, H. Herrmann, U. Aebi and S. V. Strelkov, *J. Struct. Biol.*, 2010, **170**, 369.
58. F. J. Wortmann, G. Sendelbach and C. Popescu, *J. Cosmet. Sci*., 2007, **58**, 311.
59. R. Matsunaga, R. Abe, D. Ishii, S. Watanabe, M. Kiyoshi, B. Nöcker, M. Tsuchiya, K. Tsumoto, *J. Struct. Biol*. 2013, **183**, 484.