Heidemann Lecture:

PROBING COLLAGEN STRUCTURE AND FUNCTION

John A.M. Ramshaw

Melbourne, Australia

INTRODUCTION

Collagen, the main protein of connective tissues, is of considerable economic importance, being the material from which leather and gelatin, and more recently a range of medical products, are derived. Overall, these products have a value to the world economy of between 4 to 5 USD billion each year. Research into all aspects of collagen chemistry and biology has been a key part in securing full value in the various collagen based industries, and a considerable understanding of the structure and function of collagen has now emerged. Empirical information allowed people in pre-history to use hides and skin more effectively. For example, it has been suggested that as early as around 5,300 years before present a variety of identifiable properties and applications for different sources of collagen products were recognised and considered, with specific choices of material being made (O'Sullivan *et al.,* 2016). Later, but still around 2 millennia ago, it was known that that the quality of gelatines varied between the sources of raw materials (Pliny, *c.* 50). But clearly, it was during the 20th century that all major steps in understanding collagen structure and function have emerged.

Initially, the availability of soluble forms of collagen assisted in studies of collagen structure and function. Possibly the first report of soluble collagen was that acetic acid in histology methods caused swelling and solubility (Zachariades, 1900). Later, Nageotte made solutions of collagen in acetic acid, and was able to manipulate these, for example to reform fibrils (Nageotte, 1927a,b). Later, small amounts of collagens were also purified using citric acid (Orekhovich *et al*., 1948), and then by using neutral NaCl solutions (Gross *et al*., 1955; Jackson & Fessler, 1955). However, the major step forward was the introduction of enzyme solubilisation of tissues (Nishihara & Miyata, 1962). Various enzymes were examined, including trypsin, chymotrypsin, papain, ficin and pepsin. This approach relies on the collagen remaining intact during the proteolysis. Over time, the use of pepsin in acetic acid, where the chaotropic activity of the acetic acid assists in bringing collagen into solution, has proved the most frequently adopted method and gives excellent yields of soluble collagen from juvenile tissues.

Collagen structure

Structural studies started from compositional data and later used protein sequencing methods so that an appreciation of the primary structure of collagen emerged during the 1960's (Hannig, 1960; ChandraRajan & Bose, 1967; Kang & Gross, 1970). These data also confirmed the presence of the characteristic (Gly-Xaa-Yaa) repeating triplet motif of collagens, as first suggested by Astbury (1933). Interestingly, it was only at the end of the 1960's that it was also realised that more than one genetically distinct collagen existed, when a distinct structure, collagen type II, was identified (Miller & Matukas, 1969). Nowadays, particularly due to the successful human genome analyses a wider variety of collagen genes has been discovered and their sequences established, such that some 28 genetically distinct collagen types have now been identified (Ricard-Blum, 2011).

Other structural studies were concerned with secondary and particularly the tertiary structure of collagen. The first X-ray diffraction patterns for collagen were published after the Great War (for example, Herzog & Jancke, 1921). Subsequently, various researchers, including Astbury (1933) attempted to interpret these and similar patterns. It was noted by Astbury (1940) that improved images that could be obtained by stretching. Subsequently useful patterns were reported using collagen kept moist and stretched by 8 - 10% (Cowan *et al.*, 1953) that allowed structural interpretations. It was Ramachandran, who first proposed the triple-helical structure for collagen (Ramachandran & Kartha, 1954). These authors (Ramachandran & Kartha, 1955) and others (Rich & Crick, 1960) further refined the structure. Later, quantitative X-ray diffraction data, collected from stretched kangaroo tail tendon, was used to test models for the conformation of the polypeptide chains in collagen (Fraser *et al.*, 1979). These data suggested that the average unit twist of the molecular helix was close to that expected for a 3/10 helix. A linked atom least-squares refinement of structures based on two possible models (Rich & Crick, 1961) strongly supported one of these models, the RCII structure (Rich & Crick, 1961) was a very good fit to the experimental data (Fraser *et al.*, 1979).

These, and other data, have all helped towards building an ever expanding picture of collagen, including its structure, its biosynthesis, its biochemistry, its interactions, and its catabolism. However, it is of interest that the majority of these data relate to the properties of individual molecules (Figure 1a,b). Thus there are still many areas that are less well or poorly understood, for example in the processes involved in fibril and fibre bundle formation, the organisation of collagen in tissues and the relationships to other molecules in the extracellular matrix. Essentially, the smaller the scale, the more that is known, while as the scale and dimensions get larger, considerably less is known about the structures, interactions and processes involved.

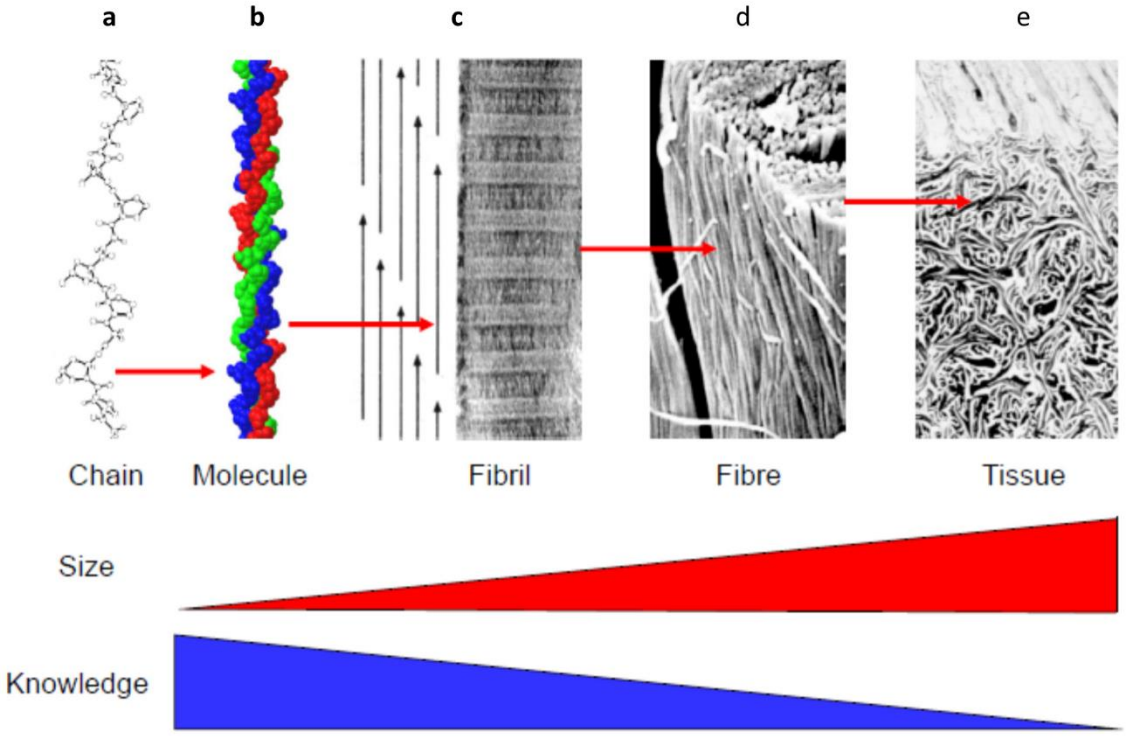

Figure 1: The hierarchical assembly of collagenous tissue (skin, e) from individual molecules (a,b) through fibrils (c) and fib re bundles (d).

As noted above, the triple-helix conformation of collagen is now well defined, and every native collagen type, from whatever source, not just mammalian tissues, is characterised by the collagen triple-helical motif. It comprises three individual chains (Figure 1a) that are each in a left-handed polyproline II-like helix. These three chains are wound into a right-handed supercoiled, rope-like structure (Figure 1b), the collagen triple-helix.

A consequence of this tightly coiled structure is that every third residue is positioned within the inside of the structure. Only glycine (Gly), which is the smallest amino acid, is able to fit into the limited available space. This results in collagen chains typically having a repeating $(Gly\text{-}Xaa\text{-}Yaa)_n$ repeating structure (Brodsky & Ramshaw, 1997). Hence, potential collagen sequences can be identified by this repeating amino acid sequence. This can be useful particularly for collagens from more primitive species where a range of human and mammalian homologs are not found. For example, a collagen-like sequence that forms the silk of the cocoon of various sawflies can be seen as collagenous from the repeating amino acid sequence (Figure 2). This recent structural data (Sutherland *et al.*, 2013) confirmed the expectation from X-ray diffraction data more than 50 years earlier (Rudall, 1962). Within this context, however, very short sequences are unlikely to fold into a stable triple–helix so n probably needs to be 20 or more in most instances.

Chain-A: VPVATPSKGSKSGHGGESGNYGHGGRGGDGSDGGAGGVGGGRSGGSGWA
GPQGPRGADGKIGPAGPQGPSGPAGPTGPVGPRGDAGRPGATGATGPDGPKGEFGPQGPSGPRGAPGP
QGPAGPTGRD
GPKGAAGPAGAAGPAGATGPDGPKGEFGPQGPSGPRGAPGPQGPAGPTGRDGPKGAAGPAGAAGPA
GSPGAQGETGDR
GDRGLKGDVGAQGGKGIPGPAGPRGQTGPNGLPGAKGETGPKGAQGPAGPAGPKGEDGATGETGPRG
PAGPAGAAG
KDIIIWKGQKGWRSPSERKSY

Chain-B: VPVPGDDQGRAESHASASSSASAGNGGNKGKADSYAEADSYASAGNDGKTGNAGSYAAAGSSASAG
ND
GNAGSYASGNSYANAGNDGNRGNRGDDGGRRGQGW
GRVGPAGEQGRQGPAGPPGPVGDRGQTGRSGQTGPQGPQGPSGPAGRQGKAGSVGETGKAGPAGPV
GATGPVGPAGAQ
GQTGAAGPRGAQGPQGPAGRQGKAGSVGETGKAGPAGPVGATGPVGPAGAQGQTGAAGPRGAQGP
QGPKGAMGPKGDQ
GPKGEQGQRGEQGAVGENGAPGPAGAKGATGAPGPKGAQGDTGARGEDGVRGPAGPKGLNGQKGAT
GPVGPAGPKGRKG
RVIFMDTPYKGDNKGDNKGGNKYDNKYDNKYDNKYDNKGDNKGQSY

Chain-C: RSVKGGKSGKGSGGKSGGYDGRNNGGNDGGNDGGDGALAWMQ
GPIGKEGPIGERGAAGPKGPVGQTGAVGDRGAVGAPGPNGRTGSVGVKGAQGQQGPSGPRGATGAAG
PAGPAGPTGEQ
GATGAPGAAGPTGRTGSVGVKGAQGQQGPSGPRGATGAAGPAGPAGPTGEQGATGAPGAAGPTGPP
GLQGPRGAKGVQ
GEKGPQGAQGNRGAPGVSGPRGPTGDRGQQGAKGDVGAKGATGAPGAAGPRGPSGLKGATGNQGPA
GPAGQAG

EDAVYPWDNKGSDNKGYDNKGYDNKGSDNKGYDNKGYDNKGYDNKGYDNKGYDNKGYDNKGYDNK
GYVNKDYDNKGYD
NKGESY

Figure 2: The amino acid sequences of the three chains of the novel collagen silk found in the willow sawfly, **Nematus oligospilus, showing the Gly-Xaa-Yaa repeating sequence. It does not contain any Hyp.** (The N- and C-terminal non-collagenous domains are in blue, and the collagenous domains are underlined and in red, with Gly residues in black.)

While all collagens comprise three chains, these are not necessarily three identical chains (Figure 1b). In some cases, the three chains are the same, providing homotrimers. These include type II and type III collagens, $\alpha1[II]_3$, $\alpha1[III]_3$. In other collagens, more than one chain type may be present, giving heterotrimers. These include type I collagen that normally has the form of $\alpha1[I]_2\alpha2[I]$, and type V collagen, $\alpha1[V]_2\alpha2[V]$. In some cases, three different chains can be found, as in the minor form of type V collagen, $\alpha1[V]\alpha2[V]\alpha3[V]$.

A key structural element that contributes significantly to thermal stability in mammalian collagens (Figure 1b), and indeed most animal collagens, is the hydroxylation of specific proline residues. The enzyme prolyl-4- hydroxylase modifies proline residues in the Yaa position only in the (Gly-Xaa-Yaa) repeat to give 4-hydroxyproline (Hyp) (Myllyharju, 2003). In cell culture, where the extent of hydroxylation of type I collagen can be controlled; reducing the hydroxylation also reduces the collagen stability (Rosenbloom *et al.*, 1973). Across the animal kingdom, the degree of hydroxylation is correlated to the environmental temperature that the animal experiences. Thus, homeotherms usually have full hydroxylation, while poikilotherms, such as fishes, frequently have incomplete hydroxylation (Rigby & Robinson, 1975).

The different human (mammalian) collagen types fall into a selection of structural and functional sub-groups (Ricard-Blum, 2011). The most abundant, and hence the most commercially important, are the five fibril forming, interstitial collagens (Figure 1c). The main one of these interstitial collagens is type I collagen, which comprises around 90% of the collagen in the body. Most remaining collagen types are non-fibrillar. Typically these are minor components, but they may be in high concentration in their particularly specific locations, for example type IV in basement membranes and type X collagen in hypertrophic cartilage. Some of these collagens are associated with fibrils, others form networks, while others are membrane associated (Ricard-Blum, 2011).

It is important to remember that for a given collagen type, the collagens properties may differ between species. An example of this is with type III collagen. This collagen can be more sensitive to proteolysis than other interstitial collagens, where human, dog and cat type III collagens were cleaved more rapidly by trypsin than that from guinea pig, while type III collagen from chicken was particularly resistant to trypsin (Welgus *et al.,* 1985). Complete susceptibility to pepsin has been shown for kangaroo type III collagen, where the digestion to solubilise type I collagen from mature skin was sufficient to entirely digest the type III collagen (Stephens *et al.,* 1988). As type III collagen is a significant component of scar tissue, an enzymatic bating treatment during kangaroo leather production can lead to dissolution of the natural scar marks leaving gaps in the grain surface.

Collagen packing

The packing of collagen into fibrils (Figure 1c) through the staggered overlap structure has been the subject of various reviews (eg: Hulmes, 2002) and accounts for a range of X-ray diffraction and transmission electron microscopy (TEM) observations. TEM images show an alternating dark/light repeating pattern, the 'D-period' repeat, due to differential uptake of stain between and the 'gap' and 'overlap' regions of the fibril (Figure 1c). The best approach to determining the dimension of the D-period repeat is X-ray diffraction with appropriate calibration controls, as the dimension is not effected in the same way as the various processing steps used in TEM. Generally, tendon tissues such as the standard rat tail tendon and reconstituted fibrils have a D-period on 67 nm, while more extensible tissues such as skin, where type III collagen is a co-component, a short D-period of about 65 nm is found (Brodsky *et al*., 1980).

The processes of fibril formation and regulation are not yet fully understood. The initiation of fibril formation has been studied in solution through specific proteolytic activity on the precursor, procollagen (Kadler *et al.*, 1990). In this *in vitro* system, collagen monomers are formed from the procollagen and form into fibrils at 37 °C. These newly formed fibrils were tapered, with pointed tips at both ends, and showed cross striations in TEM. Growth of fibrils was from the tips only with one end becoming more pointed and showing the principal growth, while the other became less sharp and blunter. The needle-like format, however, is distinct from fibrils isolated from tissues. Interestingly, both tips start with the N-terminal of the collagen and there is an inversion zone where the two parts meet (Figure 3) (Kadler *et al.*, 1990; White *et al.*, 2002).

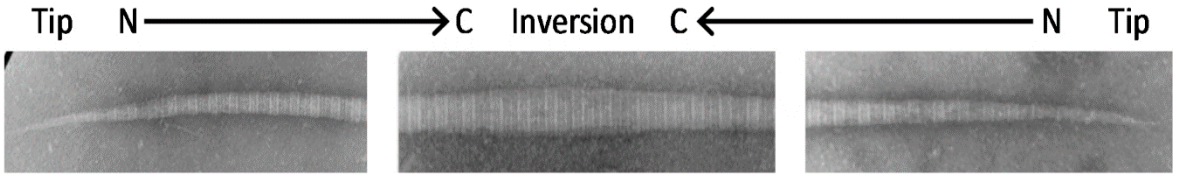

Figure 3: A newly formed fibril (in vivo) showing the tips and the inversion zone, with the N-terminal to C-terminal orientation of molecules indicated, N → C.

Establishing the fibril formation process *in vivo* is more difficult, but it is possible to isolate newly formed fibrils from tissue. For example, new fibrils from 14-day chicken embryo tendons were asymmetric, having a short and a long tapered end and were centrosymmetric with respect to molecular packing (Birk *et al.*, 1995), consistent with previous *in vitro* observations (Kadler *et al.*, 1990). An alternative approach, which provides new fibrils from an adult animals, was to present an isolated zone inside an ePTFE tube into which new tissue could grow (White *et al.*, 2002). This allowed isolation of fibrils of up to a defined age. These too showed characteristics, including asymmetric banding and an inversion zone (Figure 3) that had been seen in earlier studies (White *et al.*, 2002). The mechanisms by which these neofibrils grow into larger fibrils is not fully clear but data suggests that longer fibrils may be formed by linear fusion of smaller segments (Birk *et al.,* 1995), while lateral associations may lead to thicker fibrils (Parkinson *et al.,* 1994).

The approach of inserting a porous ePTFE tube (White *et al.*, 2002) also allows the tracking of new collagen infiltration into the porous matrix, for example by using *in situ* hybridisation or immunohistology to follow different collagen types (White *et al.*, 2014). Immunohistology (Figure 4) shows that the deposition of different collagen types happened at different rates with type V and VI collagens being the first to be deposited followed by the major interstitial collagens. Little type XII

collagen could be seen within the porous mesh, although it was seen in the surrounding tissues, while type XIV was seen throughout the porous structure of the implanted mesh (White *et al.*, 2014).

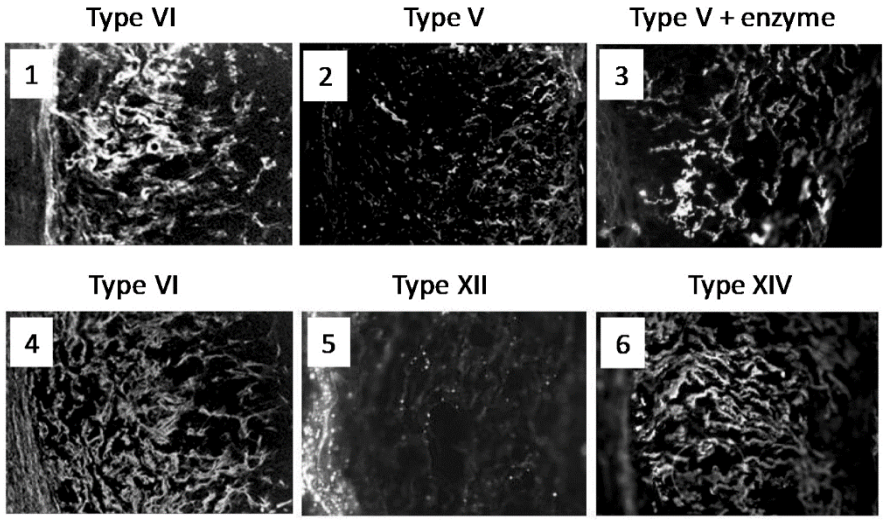

Figure 4: Immunohistochemistry examination of ePTFE explant samples after 14 days (1-4) and after 28 days (5,6). The collagen types detected (shown as white is positive) are indicated. Type V collagen can be masked by other interstitial collagens but can be visualised after enzyme treatment.

Data on the mechanisms of fibre bundle formation (Figure 1d) and how these bundles assemble into functional tissue are even more limited. Interestingly, the leather industry has given an example of genetic control of tissue structure (Figure 1e), with recessive inheritance (Dufty & Peters, 1984). This is 'vertical fibre defect' that has been observed in Hereford cattle (but not in poll Herefords) which effects the strength and manufacturing of shoe and upholstery leather splits (Dufty & Peters, 1984). In this case, the fibre bundles at the top of the corium from 'vertical fibre' animals is more open arranged perpendicular to the surface as opposed to the tight interwoven structure of 'normal' hides (Figure 5). This means that in a split any underlying interwoven structure is removed and the top of the corium lacks strength as there are no interwoven fibre bundles to provide strength (Figure 5).

Changes in tissue structure during tanning

Professor Eckhart Heidemann was involved in a wide variety of chemical and biophysical studies on collagen. One of these topics included changes to tissue and collagen during hide processing and tanning. Various approaches are possible. These include immunohistology where the presence or absence of various extracellular matrix components, not just collagens, can be followed during processing (Stephens *et al*., 1993). This shows, for example, that after liming, bating and pickling the type IV and type VII collagens associated with the basal lamina remain and contribute to a quality grain layer as methods that damage the grain also led to loss of these collagens and laminin (Stephens *et al*., 1993).

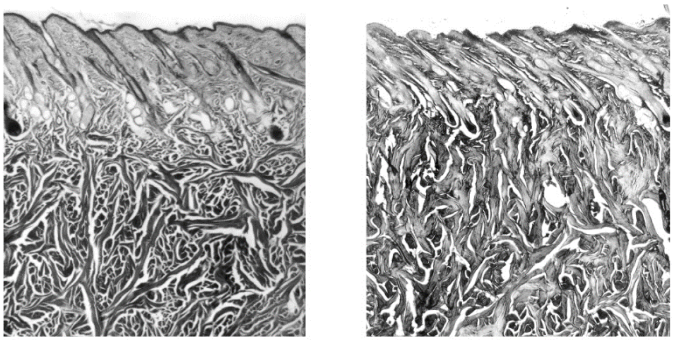

Figure 5: Microscopy of normal and 'vertical fibre' hide samples showing the distinct difference in the collagen fibre bundle organisation in these two types.

Diffraction changes during tanning were followed by Heidemann (Heidemann & Keller, 1970). They provided pioneering studies on this topic in an era when interpretations of the data were possible and made significant progress compared to the initial studies in the 1920's (see Heidemann & Keller, 1970). They used hide based samples in their studies. This contrasted with the work of others who were looking at molecular structure where very few collagen materials, but most notably rat tail tendon, gave a well resolved, quasi-crystalline diffraction pattern that was amenable to molecular analysis (for example, Fraser *et al*., 1983, and prior references cited). These researchers were well aware that this detailed diffraction pattern could be disrupted by some chemical treatments, such as with glutaraldehyde.

Early results on the effects of tanning showed variability in the results, possibly due to variations in water content. However, alternative data collection and careful control of water content and subsequent conditioning led to reproducible data (Heidemann & Keller, 1970). These data showed that the side chain (molecular) spacing was dependent on water content and that various tanning agents, including chrome and vegetable tannins can enter and penetrate into collagen fibril structures (Heidemann & Keller, 1970). Present day studies, for example on the effects of fat-liquoring (Sizeland *et al*., 2015) have the advantage of advanced synchrotron technology for data collection. Nevertheless, the experimental principles defined earlier (Heidemann & Keller, 1970), including taking account of and defining water content, salt and additive content, and the extent of conditioning are still key to obtaining quality data.

Peptide Models of Collagen

Another topic of collagen research where Professor Heidemann provided leadership was in the use of peptides and polypeptides to study the stability of the triple helix. This is also an area of research that has been studied by others subsequently, expanding on the initial concepts.

The stability of the triple helix is determined by the Xaa and Yaa residues in the amino acid sequence (Persikov *et al.,* 2005a). Changes that alter the Gly residue are frequently associated with disease (Myllyharju & Kivirikko, 2001), but are also found as natural interruptions in some collagens (Brazel *et al.,* 1987).

Heidemann examined a range of constructs, including synthetic polypeptides (Heidemann & Bernhardt, 1968; Heidemann *et al.*, 1973) such as $(Pro\text{-}Ala\text{-}Gly)_n$ where it was shown that the length of the construct, preferably with a molecular weight of 11,000 or more, was important to give maximal, reproducible melting data. Data was also reported that indicated that the residues in the Xaa and Yaa position in the repeat structure affected triple helical stability (Heidemann *et al.*, 1973).

This system was examined by a range of researchers (see Heidemann & Roth, 1982). Subsequently, Heidemann examined shorter collagen-like peptides with defined structure (Roth & Heidemann, 1981). A block-polymer, Boc-(Gly-Pro-Pro)$_5$-(Gly-Pro-Leu)$_5$-(Gly-Pro-Pro)$_5$ was examined and showed that small structures could form into a triple-helical conformation, although in this instance a methanol/water solvent system was necessary (Roth & Heidemann, 1981). Later, Heidemann studied more complex peptide systems (Thakur *et al.*, 1986; Germann & Heidemann, 1988). The key to these studies was the development of a covalently crosslinked trimer and selection of the collagen triplets based on a frequency analysis of calf skin collagen (Dölz & Heidemann, 1986). The trimer was made starting with a Lys-Lys dipeptide. Unblocking of the one α- and two ε-amino groups then gives three amino groups onto which the polypeptide chains can be built (Thakur *et al.*, 1986). Differential unblocking allows the potential production of heterotrimers. These, and subsequent data obtained with this approach (Thakur *et al.*, 1986; Germann & Heidemann, 1988) showed that the identity of the particular triplets included in the polypeptide structure clearly changed the triple-helical stability, as shown by the melting temperatures.

A wide range of other peptide model studies of collagen have followed (see Koide, 2007). Some designs have proved of particular interest as they have allowed a crystal structure determination at the molecular level, and have given significant detail on the triple-helical structure and the effect of imino acid content on the helical parameters (Bella, 2016). Other model peptide designs have been used for studying binding sites, for example the binding of a monoclonal antibody to type III collagen (Glattauer *et al.*, 1997). The use of differential disufide bond formation has been used to generate model heterotrimer peptide structures (Renner *et al*., 2004), structures which are more appropriate for certain collagen types, such as type I collagen. One chain contains a Cys-Cys sequence while the others contain a single Cys residue positioned to allow proper triple helix association relative to the Cys-Cys containing peptide (Renner *et al*., 2004).

Another design has been the 'host-guest' system that was used to get a better understanding of how substitutions of individual amino acids in either the Xaa or Yaa positions in a triple helix ('guest' residues) destabilise the molecule relative to the 'host' structure of acetyl-(Gly-Pro-Hyp)$_8$-Gly Gly-amide (Shah *et al*., 1996; Persikov *et al.*, 2000; Persikov *et al.*, 2005a). An interesting outcome was that a single Arg substitution in the Yaa position was not destabilising and was as good as having Hyp in that position (Yang *et al*., 1997). As well as single amino acid substitutions, data for two and multiple substitutions was collected. Most interesting was the stability provide by ion pair formation, such that Lys containing sequences such as Gly-Pro-Lys-Gly-Asp/Glu-Hyp provided significant thermal stability to the triple helix (Persikov *et al.*, 2005b). Together, all these data can be combined and allow the derivation of a formula that allows for the prediction of the stability of collagen sequences (Figure 6) (Persikov *et al*., 2005a).

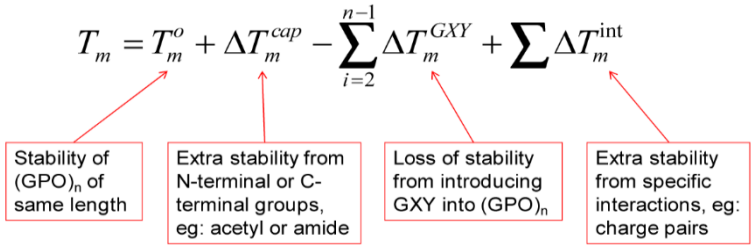

Figure 6: Prediction of collagen stability based on a compilation of 'host-guest' peptide data. This is available for use at http://compbio.cs.princeton.edu/csc/ .

Recombinant collagens

Synthetic peptide chemistry can have limitations. For example, repetitive yields at each addition step need to be very high, preferably 99% or better, in order to get a sufficient yield of the desired product. Purification can prove difficult, especially for larger peptides, especially co-synthesised trimers, where the difference between the desired product and one with a single deletion can be extremely small, making chromatographic separation during purification very difficult.

An alternative approach to study structure and function in collagens that has emerged more recently is the use of recombinant systems. Initially these were used to produce human collagens and to provide structural variations for study of selected functional sites. More recently, recombinant products have been used to allow study of many minor collagen types that had been inferred from DNA data. For example, DNA data has been used to synthesise a protein fragment that can then be used for monoclonal antibody production. These antibodies can then be used for identifying the tissue location of the new collagen and can potentially then be used for isolation from tissue.

An example of the use of a recombinant system to study collagen function was for examining how sequence changes changed the extent of Matrix Metalloproteinases (MMP) cleavage (Williams & Olsen, 2009). MMP's are essential for the normal turnover of collagen in tissues. The enzyme MMP-1 cleaves a specific glycine–isoleucine bond in type III collagen, about three quarters along the molecule. Since similar bonds occur elsewhere in collagen the suggestion was that this particular region had a more relaxed helical structure. The cleavage reaction was studies using a panel of modified collagens with sequence changes proximal to the cleavage site (Figure 7). The changes did not significantly change the helix melting temperatures (Figure 7). A single amino acid change at the Ile785 cleavage site to proline resulted in partial MMP-1 resistance, with cleavage found at other novel sites in the cleavage region (Williams & Olsen, 2009). Replacement of multiple Yaa-position residues by proline (Figure 7) led to complete resistance to MMP-1 (Williams & Olsen, 2009).

Construct															Collagenase Cleavage	Tm (°C)
	Residue	781- P_4	782- P_3	783- P_2	784- P_1	↓	785- P_1'	786- P_2'	787- P_3'	788- P_4'	789- P_5'	790- P_6'	791- P_7'	792- P_8'	793- P_9'	794- P_{10}'	795- P_{11}'	
FG-5016	(Wild type)	Gly-	Pro-	Leu-	Gly-	↓	Ile-	Ala-	Gly-	Ile-	Thr-	Gly-	Ala-	Arg-	Gly-	Leu-	Ala-	+++	42.1
FG-5015						↓	Pro-											+	41.4
FG-5029				Pro-		↓		Pro-			Pro-			Pro-			Pro-	0	42.3
FG-5021				Pro-		↓	Pro-	Pro-			Pro-			Pro-			Pro-	0	43.3

Figure 7. Examination of the effects of recombinant sequence variations in the type III collagen MMP cleavage region on the cleavage efficiency.

A further recombinant system that is proving of interest is the production of non-animal, bacterial collagens. This recently discovered diversity in collagens provides structures that are typically stable around 35 - 39 °C (Yu *et al*., 2014), yet do not contain any Hyp, an essential residue for the stability of almost all animal collagens. This lack of secondary modification, and the ready availability of commercially made synthetic genes, means that these collagens can be readily produced in large amounts by recombinant technology using *E. coli* as the host. Purification is also readily achieved, either using a sequence tag such as His_6, (Yoshizumi *et al.*, 2009) or by precipitation and proteolytic removal of contaminant proteins (Peng *et al.*, 2014).

These new designed constructs have been used, for example, to further define biologically active sites in collagen chains, including integrin, fibronectin, and heparin binding domains (Yu *et al.*, 2014). As with mammalian recombinant collagen, the non-animal bacterial collagen system has also been used to study MMP cleavage (Yu *et al*., 2012). In this system, varying numbers of Gly-Xaa-Yaa triplets from human type III collagen, from the region of the collagenase cleavage site, were inserted between two triple helix domains of the Scl2 collagen-like protein from *Streptococcus pyogenes* (Figure 8). It was found that the minimum type III sequence necessary for cleavage by the two collagenases was 5 triplets (Figure 8). The rate of cleavage by MMP-1 of the chimera containing six triplets (P7-P11') of type III collagen was similar to that of native type III collagen (Figure 8).

Construct	778- P7	779- P6	780- P5	781- P4	782- P3	783- P2	784- P1	↓	785- P1'	786- P2'	787- P3'	788- P4'	789- P5'	790- P6'	791- P7'	792- P8'	793- P9'	794- P10'	795- P11'	Collagenase Cleavage	Tm (°C)
P4-P8'	*Gly-*	*Lys-*	*Pro-*	Gly-	Pro-	Leu-	Gly-	↓	Ile-	Ala-	Gly-	Ile-	Thr-	Gly-	Ala-	Arg-	*Gly-*	*Pro-*	*Arg-*	0	36.9
P7-P11'	*Gly-*	*Lys-*	*Pro-*	Gly-	Pro-	Leu-	Gly-	↓	Ile-	Ala-	Gly-	Ile-	Thr-	Gly-	Ala-	Arg-	Gly-	Leu-	Ala-	+	36.4
P4-P11'	Gly-	Ala-	Pro-	Gly-	Pro-	Leu-	Gly-	↓	Ile-	Ala-	Gly-	Ile-	Thr-	Gly-	Ala-	Arg-	*Gly-*	*Pro-*	*Arg-*	+	36.5
P7-P8'	Gly-	Ala-	Pro-	Gly-	Pro-	Leu-	Gly-	↓	Ile-	Ala-	Gly-	Ile-	Thr-	Gly-	Ala-	Arg-	Gly-	Leu-	Ala-	0	36.5
P1-P11'	*Gly*	*Gln-*	*Pro-*	*Gly-*	*Lys-*	*Pro-*	Gly-	↓	Ile-	Ala	Gly-	Ile-	Thr-	Gly-	Ala-	Arg-	Gly-	Leu-	Ala-	0	36.3

Figure 8. Examination of the effects of recombinant sequence variations in the type III collagen MMP cleavage region on the cleavage efficiency using type III collagen segments sandwiched between two S. pyogenes collagen sequences. The 'host' bacterial collagen sequences are shown in italics in the Figure.

The other interesting option for these non-animal (bacterial) collagens and their functional variants is as potential biomedical materials, as they are readily produced, reproducibly in good yield, easily purified and are non-cytotoxic and non-immunogenic (Peng *et al*., 2010; Cosgriff-Hernandez *et al.*, 2010; Yu *et al.*, 2014).

CONCLUSIONS

On-going research into collagen structure and function will continue to bring benefits to collagen-based industries, through enhancing the value of products and providing ideas and technology for new product opportunities. For example, understanding collagen interactions with solvents and other compounds in new processing systems will help to address environmental pressures. Also, production of recombinant products may provide new industrial opportunities. In particular, in contrast to the present knowledge at a molecular level, research into collagen at the scale of fibre bundle and tissue formation, which are presently very under-represented in our knowledge base, will bring further commercial benefits.

References

Astbury WT. Some problems in the X-ray analysis of the structure of animal hairs and other protein fibres. Trans Faraday Soc. 1933; 29:193-205.

Astbury WT. The molecular structure of the fibers of the collagen group. First Procter memorial lecture. J Int Soc Leather Trades Chem. 1940; 24:69-92.

Bella J. Collagen structure: new tricks from a very old dog. Biochem J. 2016; 473:1001-1025.

Birk DE, Nurminskaya MV, Zycband EI. Collagen fibrillogenesis in situ: fibril segments undergo post-depositional modifications resulting in linear and lateral growth during matrix development. Dev Dyn. 1995; 202:229-243.

Brazel D, Oberbäumer I, Dieringer H, Babel W, Glanville RW, Deutzmann R, Kühn K. Completion of the amino acid sequence of the α1 chain of human basement membrane collagen (type IV) reveals 21 non-triplet interruptions located within the collagenous domain. Eur J Biochem. 1987; 168:529-536.

Brodsky B, Eikenberry EF, Cassidy K. An unusual collagen periodicity in skin. Biochim Biophys Acta. 1980; 621:162-166.

Brodsky B, Ramshaw JAM. The collagen triple-helix structure. Matrix Biol. 1997; 15:545-554.

ChandraRajan J, Bose SM. Amino acid sequences of N- and C-terminal and lysine-containing peptides of collagen and elastin. Indian J Biochem. 1967; 4:120-125.

Cosgriff-Hernandez E, Hahn MS, Russell B, Wilems T, Munoz-Pinto D, Browning MB, Rivera J, Höök M. Bioactive hydrogels based on Designer Collagens. Acta Biomater. 2010; 6:3969-3977.

Cowan PM, North ATC, Randall JT. High-angle X-ray diffraction of collagen fibres. In: The Nature and Structure of Collagen, ed. Randall JT, Butterworths, London, 1953; 241-249.

Dölz R, Heidemann E. Influence of different tripeptides on the stability of the collagen triple helix. I. Analysis of the collagen sequence and identification of typical tripeptides. Biopolymers. 1986; 25:1069-1080.

Dufty JH, Peters DE. Confirmation of the recessive mode of inheritance of the vertical fiber hide structure in Hereford cattle. J Soc Leather Technol Chem. 1984; 68:35.

Fraser RD, MacRae TP, Miller A, Suzuki E. Molecular conformation and packing in collagen fibrils. J Mol Biol. 1983; 167:497-521.

Fraser RDB, MacRae TP, Suzuki E. Chain conformation in the collagen molecule. J Mol Biol. 1979; 129:463-481.

Germann HP, Heidemann E. A synthetic model of collagen: an experimental investigation of the triple-helix stability. Biopolymers. 1988; 27:157-163.

Glattauer V, Werkmeister JA, Kirkpatrick A, Ramshaw JAM. Identification of the epitope for a monoclonal antibody that blocks platelet aggregation induced by type III collagen. Biochem J. 1997; 323:45-49.

Gross J, Highberger JH, Schmitt FO. Extraction of collagen from connective tissue by neutral salt solutions. Proc Natl Acad Sci U S A. 1955; 41:1-7.

Hannig K. Investigations on the sequence of amino acid residues in collagen. Fed Proc. 1960; 19:1.

Heidemann E, Bernhardt HW. Synthetic polypeptides as models for collagen. Nature. 1968; 220:1326-1327.

Heidemann E, Roth W. Synthesis and investigation of collagen model peptides. Adv Polymer Sci. 1982; 43:143-203.

Heidemann E. X-ray studies of tanned collagen. J Amer Leather Chem Assoc. 1970; 65:512-536.

Heidemann ER, Harrap BS, Schiele HD. Hybrid formation between collagen and synthetic polypeptides. Biochemistry. 1973; 12:2958-2963.

Herzog RO, Jancke W. Verwendung von Röntgenstrahlen zur Untersuchhung metamikroskopischer biologischer Strukturen. Festschrift der Kaiser Wilhelm Gesellschraft. Verlag von Julius Springer, Berlin, 1921, 118-121.

Hulmes DJ. Building collagen molecules, fibrils, and suprafibrillar structures. J Struct Biol. 2002; 137:2-10.

Jackson DS, Fessler JH. Isolation and properties of a collagen soluble in salt solution at neutral pH. Nature. 1955; 176:69-70.

Kadler KE, Hojima Y, Prockop DJ. Collagen fibrils in vitro grow from pointed tips in the C- to N-terminal direction. Biochem J. 1990; 268:339-343.

Kang AH, Gross J. Amino acid sequence of cyanogen bromide peptides from the amino-terminal region of chick skin collagen. Biochemistry. 1970; 9:796-804.

Koide T. Designed triple-helical peptides as tools for collagen biochemistry and matrix engineering. Philos Trans R Soc Lond B Biol Sci. 2007; 362:1281-1291.

Miller EJ, Matukas VJ. Chick cartilage collagen: a new type of α1 chain not present in bone or skin of the species. Proc Natl Acad Sci U S A. 1969; 64:1264-1268.

Myllyharju J, Kivirikko KI. Collagens and collagen-related diseases. Ann Med. 2001; 33:7-21.

Myllyharju J. Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. Matrix Biol. 2003; 22:15-24.

Nageotte J. Coagulation fibrillaire in vitro du collagene dissous dans un acide dilue. Comptes rendus de l'Académie des sciences. 1927a; 184:115-117.

Nageotte J. Sur le caillott artificial de collagène; signification, morphologie générale et technique. Comptes Rendus des Seances de la Societe de Biologie et de ses Filiales. 1927b; 96:172-174.

Nishihara T, Miyata T. The effects of proteases on the soluble and insoluble collagens and the structures of insoluble collagen fiber. In: Collagen Symposium, Vol. 3. 1962; 66-93.

Orekhovich VN, Tustanovskii AA, Orekhovich KD, Plotnikova NE. O prokollagene kozhi. Biokhimiya, 1948; 13:55-60.

O'Sullivan NJ, Teasdale MD, Mattiangeli V, Maixner F, Pinhasi R, Bradley DG, Zink A. A whole mitochondria analysis of the Tyrolean Iceman's leather provides insights into the animal sources of Copper Age clothing. Sci Rep. 2016; 6:31279.

Parkinson J, Kadler KE, Brass A. Self-assembly of rodlike particles in two dimensions: A simple model for collagen fibrillogenesis. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics. 1994; 50:2963-2966.

Peng YY, Stoichevska V, Madsen S, Howell L, Dumsday GJ, Werkmeister JA, Ramshaw JAM. A simple cost-effective methodology for large-scale purification of recombinant non-animal collagens. Appl Microbiol Biotechnol. 2014; 98:1807-1815.

Peng YY, Yoshizumi A, Danon SJ, Glattauer V, Prokopenko O, Mirochnitchenko O, Yu Z, Inouye M, Werkmeister JA, Brodsky B, Ramshaw JAM. A *Streptococcus pyogenes* derived collagen-like protein as a non-cytotoxic and non-immunogenic cross-linkable biomaterial. Biomaterials. 2010; 31:2755-2761.

Persikov AV, Ramshaw JAM, Kirkpatrick A, Brodsky B. Amino acid propensities for the collagen triple-helix. Biochemistry. 2000; 39:14960-14967.

Persikov AV, Ramshaw JAM, Brodsky B. Prediction of collagen stability from amino acid sequence. J Biol Chem. 2005a; 280:19343-19349.

Persikov AV, Ramshaw JAM, Kirkpatrick A, Brodsky B. Electrostatic interactions involving lysine make major contributions to collagen triple-helix stability. Biochemistry. 2005b; 44:1414-1422.

Pliny G. Natural History XI. ~50; 39, 231.

Ramachandran GN, Kartha G. Structure of collagen. Nature. 1954; 174:269-270.

Ramachandran GN, Kartha G. Structure of collagen. Nature. 1955; 176:593-595.

Renner C, Saccà B, Moroder L. Synthetic heterotrimeric collagen peptides as mimics of cell adhesion sites of the basement membrane. Biopolymers. 2004; 76:34-47.

Ricard-Blum S. The collagen family. Cold Spring Harb Perspect Biol. 2011; 3:a004978.

Rich A, Crick FHC. The molecular structure of collagen. J Mol Biol. 1961; 3:483-506.

Rigby BJ, Robinson MS. Thermal transitions in collagen and the preferred temperature range of animals. Nature. 1975; 253:277-279.

Rosenbloom J, Harsch M, Jimenez S. Hydroxyproline content determines the denaturation temperature of chick tendon collagen. Arch Biochem Biophys. 1973; 158:478-484.

Roth W, Heidemann E. Triple helix coil transition of a blockpolymer with the sequence Boc-(Gly-Pro-Pro)$_5$-(Gly-Pro-Leu)$_5$-(Gly-Pro-Pro)$_5$-NH_2. Int J Pept Protein Res. 1981; 17:527-530.

Rudall KM. Silk and other cocoon proteins. Comp Biochem. 1962; 4B:397–433.

Shah NK, Ramshaw JAM, Kirkpatrick A, Shah C, Brodsky B. A host-guest set of triple-helical peptides: stability of Gly-X-Y triplets containing common nonpolar residues. Biochemistry. 1996; 35:10262-10268.

Sizeland KH, Wells HC, Norris GE, Edmonds RL, Kirby N, Hawley A, Mudie ST, Haverkamp RG. Collagen D-spacing and the effect of fat liquor addition. Stephens LJ, Werkmeister JA, Caine JM, Ramshaw JAM. Pepsin soluble collagens from kangaroo skin. Leder. 1988; 39:88-93.

Stephens LJ, Werkmeister JA, Ramshaw JAM. Changes to bovine hides during leather processing. J Soc Leather Technol Chem. 1993; 77:71-74.

Thakur S, Vadolas D, Germann HP, Heidemann E. Influence of different tripeptides on the stability of the collagen triple helix. II. An experimental approach with appropriate variations of a trimer model oligotripeptide. Biopolymers. 1986; 25:1081-1086.

Welgus HG, Burgeson RE, Wootton JA, Minor RR, Fliszar C, Jeffrey JJ. Degradation of monomeric and fibrillar type III collagens by human skin collagenase. Kinetic constants using different animal substrates. J Biol Chem. 1985; 260:1052-1059.

White JF, Werkmeister JA, Darby IA, Bisucci T, Birk DE, Ramshaw JAM. Collagen fibril formation in a wound healing model. J Struct Biol. 2002; 137:23-30.

White JF, Werkmeister JA, Bisucci T, Darby IA, Ramshaw JAM. Temporal variation in the deposition of different types of collagen within a porous biomaterial implant. J Biomed Mater Res A. 2014; 102:3550-3555.

Williams KE, Olsen DR. Matrix metalloproteinase-1 cleavage site recognition and binding in full-length human type III collagen. Matrix Biol. 2009; 28:373-379.

Yang W, Chan VC, Kirkpatrick A, Ramshaw JAM, Brodsky B. Gly-Pro-Arg confers stability similar to Gly-Pro-Hyp in the collagen triple-helix of host-guest peptides. J Biol Chem. 1997; 272:28837-28840.

Yoshizumi A, Yu Z, Silva T, Thiagarajan G, Ramshaw JAM, Inouye M, Brodsky B. Self-association of *Streptococcus pyogenes* collagen-like constructs into higher order structures. Protein Sci. 2009; 18:1241-1251.

Yu Z, An B, Ramshaw JAM, Brodsky B. Bacterial collagen-like proteins that form triple-helical structures. J Struct Biol. 2014; 186:451-461.

Yu Z, Visse R, Inouye M, Nagase H, Brodsky B. Defining requirements for collagenase cleavage in collagen type III using a bacterial collagen system. J Biol Chem. 2012; 287:22988-22997.

Zachariades PA. Recherches sur la structure du tissu conjonctif, sensibilité du tendon aux acides. Comptes Rendus des Seances de la Societe de Biologie et de ses Filiales. 1900; 52:182-184.