Chapter 7

X-Ray Diffraction Detects D-Periodic Location of Native Collagen Crosslinks *In Situ* and Those Resulting from Non-Enzymatic Glycation

Rama Sashank Madhurapantula and
Joseph P.R.O. Orgel

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.71022

Abstract

Synchrotron based X-ray diffraction experiments can be highly effective in the study of mammalian connective tissues and related disease. It has been employed here to observe changes in the structure of Extra-Cellular Matrix (ECM), induced in an *ex vivo* tissue based model of the disease process underlying diabetes. Pathological changes to the structure and organization of the fibrillar collagens within the ECM, such as the formation of non-enzymatic crosslinks in diabetes and normal aging, have been shown to play an important role in the progression of such maladies. However, without direct, quantified and specific knowledge of where in the molecular packing these changes occur, development of therapeutic interventions has been impeded. *In vivo*, the result of non-enzymatic glycosylation i.e. glycation, is the formation of sugar-mediated crosslinks, aka advanced glycation end-products (AGEs), within the native D-periodic structure of type I collagen. The locations for the formation of these crosslinks have, until now, been inferred from indirect or comparatively low resolution data under conditions likely to induce experimental artifacts. We present here X-ray diffraction derived data, collected from whole hydrated and intact isomorphously derivatized tendons, that indicate the location of both native (existing) and AGE crosslinks *in situ* of D-periodic fibrillar collagen.

Keywords: type I collagen, diabetes, crosslinking, advanced glycation end-products, X-ray diffraction

1. Introduction

The extracellular matrix (ECM) is a complex network of biomolecules that provides structural and functional support to cells. Both within and with the ECM, cells proliferate and

form three-dimensional tissues in multicellular organisms. As such, the ECM is important to homeostasis, function and survival in higher vertebrates. Varying composition of the ECM per tissue, supports widely different biological functions. For instance, blood vessels, organ systems, bones, tendons, skin and so forth although significantly different in material properties and cellular-tissue functions, contain many common ECM elements. The collagens (there are nearly 30 different types) account for a major part of nearly all vertebrate ECMs. Type I collagen is the most predominant collagen found in animals (and also found in some insects and plants). It is the major constituent of tendons, ligaments, skin, bone, teeth, body organ frameworks and so forth. While type I collagen is surpassed in some specific connective tissues (such as in cartilage), major structural features of type I collagen appear to be largely common to the other fibril-forming types [1]. Therefore, the research presented here, conducted on type I collagen rich tendons, can be used to extend our structural understanding of other fibrillar collagens and tissues with fairly reasonable extrapolation. This facility is enhanced by the fact that the structure of type I collagen is one of the best-defined structures in the collagen family. The *in situ* packing structure of type I collagen from rat-tail tendons has been determined with an associated and publicly accessible molecular model resulting from this work [2] (**Figure 1**). With this structural model and the data resulting from this study, the effects of molecular changes in collagen

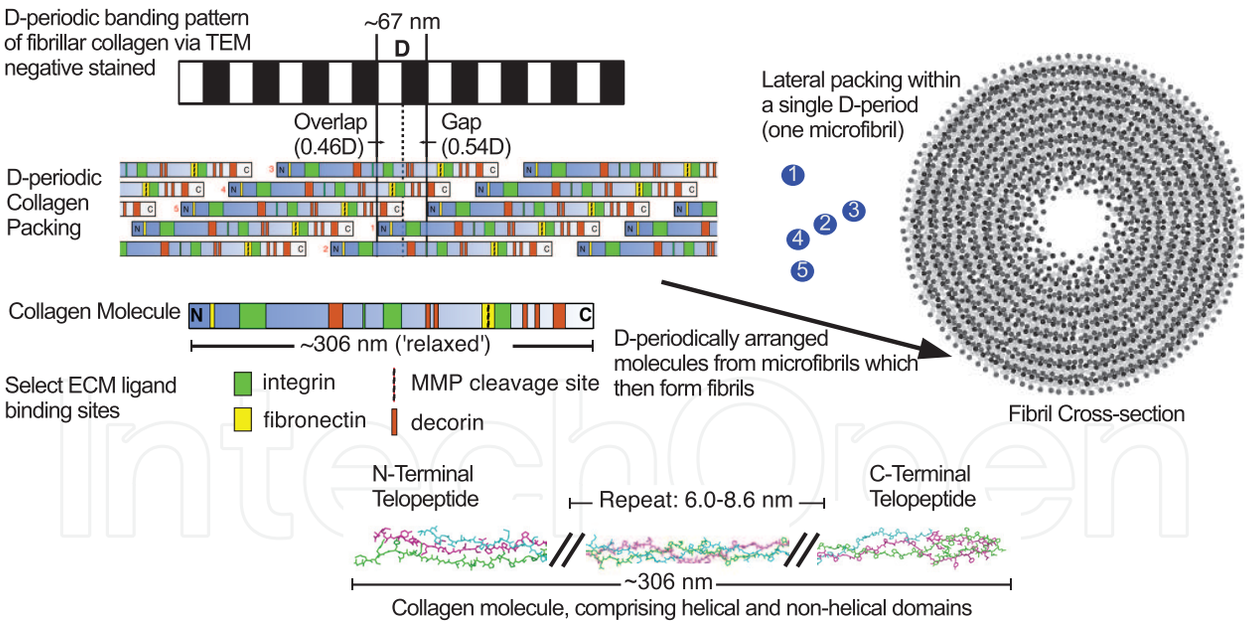

Figure 1. Structural hierarchy of type I collagen. The principal components of collagen fibrils are the triple helices (~306 nm) that are primarily composed of peptides with a residue sequence of G-X-Y (where X and Y are commonly Proline and Hydroxyproline). The triple helical regions of the monomers are flanked on the N and C termini by non-helical telopeptides. Neighboring collagen molecules are staggered from one another to form the collagen microfibril. This is the three-dimensional structure formed by the 67 nm D-periodic repeat comprising an overlap region (where 5 monomer molecular segments align) and the gap region (4 monomer molecular segments align). Several important ligand binding sites are indicated (colored boxes within collagen molecules) on the monomers which adopt a network of potentially related sites in the microfibrillar packing of collagen.

structure and organization brought about by disease may be understood and thus treatments formulated to address them.

1.1. Type I collagen – molecular packing and stabilization

Type I collagen is the major component of rat-tail tendons, the sample tissue of this study. These tendons samples are not single crystals; their supramolecular arrays are very large in cellular terms and are insoluble. Therefore, normal (single-crystal) crystallographic methods cannot be applied without adaptation. Nor is it necessary to form artificial crystals of collagen as the collagen molecules *in situ* of their tissues are *crystalline*. They are arranged in a quasi-hexagonal packing scheme radially and D-periodic staggered packing scheme longitudinally that form supramolecular structures [1], the microfibril and fibril. The fundamental biological units of this ECM protein in animals are the collagen fibrils, which also are the crystalline units that diffract X-rays. These fibrils align nearly parallel to each other within tendons (**Figure 1**). As alluded to above, the ~300 nm long collagen molecules align axially to form the 67 nm staggered repeat (with 234 amino acids per molecular segment within this 67 nm repeat) as observed by electron microscopy and X-ray diffraction. This 67 nm repeat is referred to as the D-period where fractions of D provide an axial location originating from the N-terminal end of the D-period or referring to how much of the D-period is gap or overlap. Within these 67 nm repeats, are the overlap region which is 0.46D and the gap region which is 0.54D.

The axial (or longitudinal) arrangement of collagen molecules within a fibril is significantly more crystalline than the lateral (or radial) packing discussed above, giving a series of well-ordered Bragg X-ray reflections along the meridian of the X-ray fiber diffraction pattern [3]. Putting the information from *both* the lateral and axial X-ray diffraction, Orgel et al. [2] reported a full 3D structure and accompanying model that is well-accepted and regularly forms the basis of structural-function interpretation from the molecular scale upwards.

The hierarchical packing structure explained above is maintained *in vivo* by means of post-translational modifications to the collagen fibril and through the attachment of stabilizing ligands to both the fibril and fibril-bundle(s) [4, 5]. In this present study, we look at the role of chemical crosslinks on the axial packing structure of collagen and the specific placement within the D-period of these cross-links using X-ray diffraction methodology.

1.2. X-ray fiber diffraction

The principals of X-ray crystallography have been previously applied to study the structure of types I and II collagen and how they are adapted to their functional role [1, 2, 6, 7].

As is well known, X-rays are a form of electromagnetic radiation. Much as when visible light is incident on an object, the object scatters these rays. For visible light, the scattered rays can then be refocused using a lens and these focused rays form an image. For scattered X-rays however although theoretically possible, image focus, is in practice not yet achievable for the

study of biological samples on the molecular scale. However, as a matter of almost common course at this time, the diffracted (scattered) X-rays can be recorded using X-ray detectors such as charge coupled devices (CCDs).

Rat-tail tendons diffract focused X-rays to give rise to a series of well-ordered Bragg reflections in alignment with the long axis of the tendon, which can be recorded on CCDs. As mentioned earlier, these reflections are a result of the crystalline D-periodic packing of collagen along the axis [7, 8]. The amplitudes (the square root of recorded intensity) of the Bragg reflections extracted from these images can be used with appropriate scaling, along with the missing 'phase' information to calculate electron density maps, along one crystallographic unit cell [9, 10]. One such method, is to use the known phase information of another known structure; one that is isomorphous to the structure under investigation is particularly advantageous. This was the method employed for the analyses presented in this study.

Native phases from the previously determined structure of type I collagen were combined with the amplitude information obtained from XRD patterns of isomorphous derivatives of rat-tail tendons created by incubating them in sugar solutions to create non-enzymatically formed chemical crosslinks (analogous to those formed in diabetes and normal aging) and further derivatizing them with microscopy dyes specific to these crosslinks. Appropriate scaling of native and isomorphous derivative data can be used to calculate difference Fourier and Patterson maps that may be used to locate those differences induced in the experimental tendons (glycated tendons) and those that are native (representative of normal, non-diseased state). This is based on the same method that was used to determine the structure of type I collagen [1, 2, 7]. The phase information from this structure may then be used in creating difference Fourier maps that indicate the locations of chemical cross-links to a level of accuracy that is similar to the resolution of the data used to obtain the maps.

1.3. Fourier and Patterson maps

The diffraction pattern of an object is equivalent to its own Fourier transform. Hence a reverse Fourier transform calculated using the amplitudes of the Bragg reflections obtained from the XRD patterns and phases would result in a continuous electron density profile along one D-period (the axial crystallographic unit cell) of type I collagen. A "difference Fourier" can be calculated by using the subtracted difference between derivative (for instance glycated tendon) and native (unmodified tendon) diffraction amplitudes and the phases for the native structure. This difference Fourier provides a map of electron densities added to the native structure and thus indicate regions in the D-period where chemical modifications have occurred.

In addition, calculating the Patterson and difference Patterson function from glycated, dyed and unmodified rat-tail tendons gives data on the distribution and the magnitude of electron dense regions along the D-period of collagen. The use of the Patterson function provides an important cross-verification of the accuracy of the interpretation based on the difference Fourier map/s. The key supporting value of the Patterson function, is that it does not rely on phase information and is a 'pure' observation of key data. This also makes it much more complex in terms of

the structure of the molecule, than the electron density maps reconstructed by Fourier maps. A molecule with N atoms will give rise to N(N−1)+1 peaks in a Patterson map. However, with the proper amino acid sequence information as we have for type I collagen and some knowledge of its structure (we have substantial detailed knowledge of its native structure), the Patterson map becomes a powerful instrument of model validation for fibrillar collagen structures based on X-ray diffraction data [6, 11].

A method of cross-verification between Difference Fourier and Difference Patterson maps has been established previously, by using the observed locations of changes in electron densities from the difference Fourier to synthesize a "model" Patterson map of the difference between derivative and native (i.e. where the crosslinks occur and have been labeled). This is then compared to the observed difference Patterson derived from the amplitudes of the derivative and native diffraction patterns. Since the Patterson map is completely unaffected by assumptions about phase information, a difference Patterson of derivative and native data should represent the standard to which a model Patterson should conform. The model Patterson should be similar to that of the (model independent) difference Patterson if the Difference Fourier and resulting interpretive model are correct (the interpretive model from the Fourier map being used to generate the model Patterson function). This approach was applied to the study of glycated tendons presented here.

1.4. Enzymatic crosslinks in type I collagen

Enzymatic crosslinking is the formation of stabilizing covalent linkages between collagen monomers in the molecular packing of collagen [12]. The mechanism of crosslinking is effectively one of the final steps in the building of fibrils, thereby functionalizing collagen [13]. This happens through a specific set of enzymatic reactions at determined locations. A post-translational enzymatic oxidative deamination of lysine or hydroxylysine residues by the enzyme lysyl oxidase results in the formation of intra and intermolecular crosslinks with other lysine or hydroxylysine residues (**Figure 2**). The activity of this enzyme is highly regulated by steric requirements and by the amino acid sequences upstream and downstream of the target [14]. The quantification of these crosslinks has been established using chromatography [15].

The locations of these enzymatic crosslinks constrain the possible conformation of the C-telopeptides of the two α1 chains to form sharp hairpin turns at around residues 13 and 14 of the 25-residue telopeptides [16]. The formation of intermolecular crosslinks at specific sites is a stabilizing mechanism for the supramolecular packing of collagen molecules. Impairments of crosslinks, either destabilization or hyper stabilization can lead to pathological conditions [17–19].

1.5. Glycation

Glycation, also known as Maillard reaction, is a non-enzymatic reaction between reducing sugars and amino groups in proteins, lipids and nucleic acids. It is a process by which pathologies

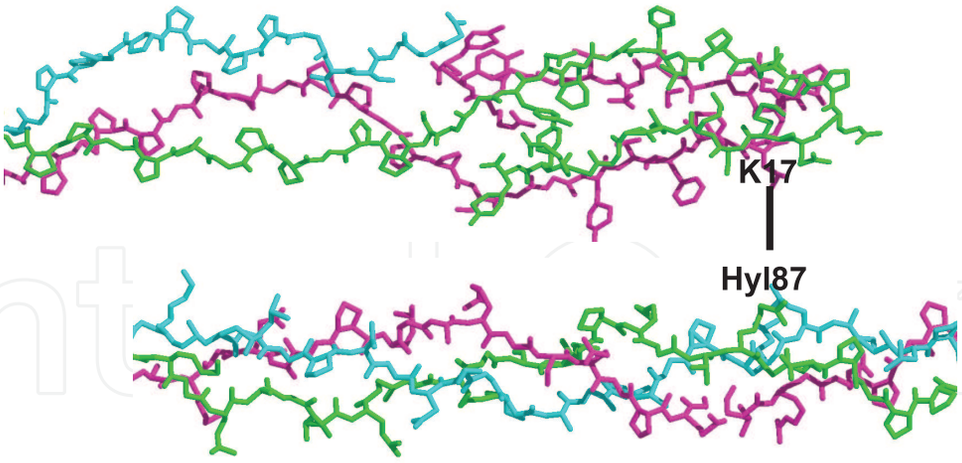

Figure 2. Crosslinks at the C-terminal telopeptide formed as a result of lysyl oxidase activity. A crosslink between lysine (residue 17) on the C-telopeptide is shown with a hydroxyproline (residue 87) on one of the α1 chains within the triple helical region.

associated with diabetes progress. The process of glycation involves the generation of a series of intermediates including Schiff bases and Amadori products, and ultimately results in the formation of advanced glycation end-products (AGE). This is a slow process, spread over a span of a few weeks, making biochemical species with relatively long half-lives, such as collagen, prime targets for these reactions [20]. Glycation is a major complication in diabetic patients and is common in both afflicted animal models and humans [21]. The availability of high concentrations of glucose in the blood stream leads to (as yet to be determined) changes in the packing structure of collagen which effects and probably impairs, both the functional and mechanical aspects of collagen's role in the ECM (**Figure 3**). If considerable changes are made to its mechanical properties, those alterations may in turn, induce or further aid in the progression of pathological conditions. Considering the role of fibrillar collagens in the structure of vasculature, as a sole target discussed among several possibilities, the significance of these changes becomes readily apparent. AGEs are heterogeneous and complex groups of compounds, strongly implicated in the pathobiology of diabetes [21]. The initial chemistry behind the formation of AGEs has been known since the early 1900s, although a great deal of important information with regards to the contribution of AGEs in disease has been uncovered in the last 30 years, even while the specific structural sites effected by these changes were yet to be determined.

The process of glycation is a Maillard reaction that can be divided into three broad stages [22]. The first stage of glycation is a result of a nucleophilic attack by the amine group, most commonly the ε-amino group (side chain) in amino acids such as lysine, on the carbonyl (—CHO) group of the sugar in open chain confirmation to form a Schiff base. This is a reversible step [23]. Both pentoses and hexoses can take part in this reaction, although only in open chain conformation which gives the otherwise closed carbonyl group access to react with the amines from the protein. This Schiff base is then rearranged into a relatively more stable Amadori product [22, 24].

The Amadori product formed in the first stage of the glycation reaction breaks down to give rise to several compounds, such as glyoxal, methylglyoxal, and 3-deoxyglucosone and various other smaller intermediates [22, 24, 25].

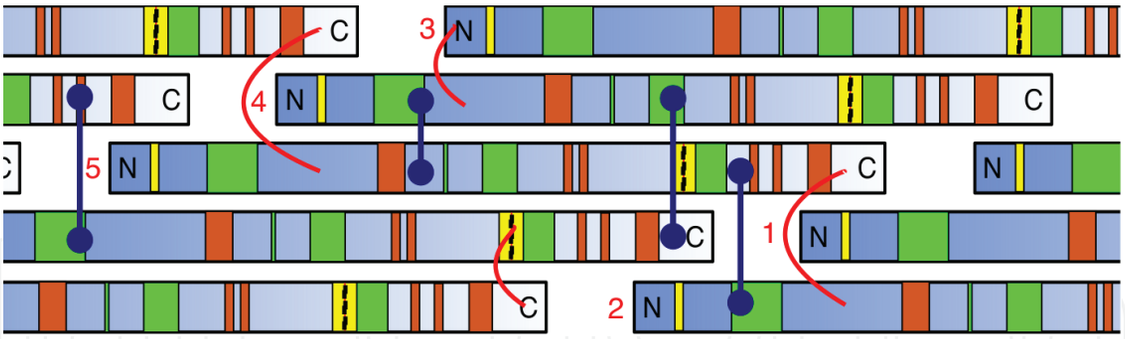

Figure 3. A magnified representation of the microfibrillar arrangement of collagen monomers. Mature enzyme-mediated (lysyl oxidase) crosslinks are marked in orange and possible, candidate, sugar-mediated (glycation) crosslinks are marked in blue. The colored bands in the collagen monomers show ligand binding sites as depicted in **Figure 1**.

The final/late stage of glycation is a result of the highly reactive α-dicarbonyls from the intermediate stage, reacting with amine and other groups from basic amino acids. After further rearrangement of the chemical structures, a large set of potentially varied and stable AGEs are formed which permanently alter the structure of the proteins that participated in the reaction. These steps are represented in a flow diagram in **Figure 4**. For a detailed discussion on the chemical reaction and the intermediates, see above referenced works.

The structural nature of AGEs is varied. Some AGEs, such as carboxymethyllysine (CML) and carboxymethylarginine (CMA), exhibit no special structural characteristics. Some examples of lysine-lysine AGEs include glyoxal-lysine dimer (GOLD), methylglyoxal-lysine dimer (MOLD), while pentosidine, glucosepane, DOGDIC, MODIC, GODIC are lysine-arginine crosslinks [24]. As already mentioned, the overall kinetics of the non-enzymatic process of glycation are rather slow. The availability of reducing sugars, mainly glucose, in open chain form is fairly limited in the human body but this becomes a significant issue for some common human conditions. Poor glucose control, such as that seen in diabetic patients, can lead to formation and deposition of AGEs in various tissues [26]. Metabolic products of glucose, such as glucose-6-phosphate, tend to be faster glycating agents and to react more quickly with

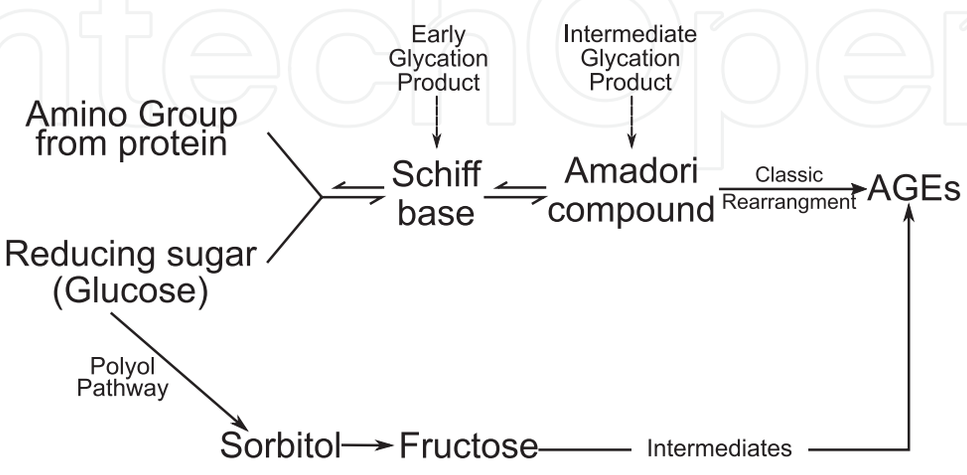

Figure 4. A schematic representation of the process of glycation *in vivo*.

proteins than this sugar in its native form [20, 27]. The pathological effects of glycation of collagens has been described in various works cited in **Table 1**.

1.6. Detection of AGEs

The nature of glycation in collagen has been a point of discussion and scientific inquiry for at least 30 years. Historically, some researchers have indicated that the process and the location of these crosslinks are more or less random and uncontrolled. However, Wess et al. [28] speculated based on their X-ray diffraction studies that a few small peaks in their difference Fourier maps along with a major peak at the lysine 434 in the gap region of the α1 chain, may originate from glycation, and that the process of glycation may be much more specific than it was (at that time) assumed. The ability to find the locations of these changes has further been fundamentally confounded by the rather small electron density changes brought about by the crosslink products, i.e. the AGEs. This makes their detailed study impractical with X-ray diffraction, unless additional X-ray diffraction contrast is provided, i.e. added electron density to sites specific to these glycation events. This study addresses that need.

Boronic acid derivatives have been used to bind with AGEs and other chemical moieties of this class to reproducibly create fluorescent contrast [29–31] in previous experimental work, unrelated to X-ray diffraction. Danslyated boronic acids have been used in determining the concentration of glycation mediated crosslinks in globular and fibrillar proteins and fats have been developed and commonly adopted [32]. These fluorescent contrast agents also happen to have, intrinsic to their nature, some significant electron density. Thus, although their structures are too large to be functionally useful for high-resolution single crystal crystallography, they may be exceedingly useful at lower resolution

Publication and year	Glycation associated pathology
Stitt et al., 2005, 1998	Destabilization of the structure of the vitreous gel within the eye
Albon et al., 1995	Age-related changes in the human lamina cribrosa, and the prevalence of chronic open angle glaucoma (COAG)
Dyer et al., 1993	Changes in mechanical properties – stiffer and more brittle collagen
Brownlee, 2000	Altered cell proliferation, motility, gene expression and response to cytokines
Gugliucci and Bendayan, 1996	Increased activation of cytokine (NF-kB) activity resulting in activation of Matrix Metalloproteinases (MMP) and increased collagen digestion
Murillo et al., 2008	Impaired wound healing as a result of altered periodontal cell behavior
Thomas & Lascelles, 1966	Accumulation of collagen in the endoneurial interstitium leading to impaired neuronal function
Sternberg et al., 1995	End-stage renal disease from glycation of basement membrane in glomeruli
Avendano et al., 1999	Impaired myocardial function from stiffness in collagen in the ventricles

Table 1. Evidence of collagen and AGE related pathologies.

(nanometer scale) X-ray diffraction analyses; studies akin to those reported here, where they effectively form isomorphous derivatives, beyond the upper resolution limit of the data collected. Based on these first principles, we present here a method and preliminary results we use to identify the location of AGEs within the fibrillar packing of type I collagen in rat-tail tendons.

2. Materials and methods

All X-ray diffraction experiments were conducted *in situ* on rat-tail tendons (RTT). Rat-tail tendons were carefully dissected out from the tails of healthy adult rats aged 3–6 months. These animals were euthanized, at the time, for other animal studies and the tails were generously donated. The tails were immediately stored at −40°C until the time of dissection for tendons. After dissection, all tendons were stored at 4°C in 10–50 mM Tris Buffered Saline (TBS) (Sigma Aldrich St. Louis MO; Cat no. T5912; 10x stock, diluted as needed). 0.04% (w/v) sodium azide (NaN_3) was added to all storage buffers and sample incubations to prevent microbial contamination.

2.1. Sample setup for glycation

0.2 M solution of glucose was made by slowly mixing the appropriate amount of D-glucose (Sigma Aldrich St. Louis MO; Cat no. 158968) and the serially diluting the stock to make the desired concentrations. All sugar solutions were made in 10 mM TBS. Tendons were deposited in Falcon tubes containing each of the sugar solutions and incubated in a 37°C water bath for 20–40 days.

2.2. Treatment with fluorescent dyes

3-(Dansylamino)phenylboronic acid (DPBA) (Sigma Aldrich (St. Louis MO; Cat No. 30423-) is a fluorescent dye that binds to the AGEs [32]. A 30 μM stock of DPBA was made in methanol. A working standard DPBA solution at 3 μM was made by diluting the using Phosphate Buffered Solution (PBS), at pH 7.4. Non-glycated and glycated tendons were incubated in the working standard solution of DPBA for 20 minutes and then washed in PBS for 2 minutes, prior to XRD data collection.

2.3. Small and wide angle X-ray diffraction

RTT diffraction patterns were obtained at Biophysics Collaborative Access Team (BioCAT) facility at Argonne National Laboratory, Chicago IL. The Mar 165 CCD detector was used to collect all small (1–15 Bragg orders) and wide-angle (3–22 Bragg orders) diffraction patterns. The sample to detector distances for small angle (SAX) and wide angle (WAX) X-ray diffraction were from 1.5 to 2.5 m and from 25 to 70 cm, respectively. Beamline optics were setup to deliver focused X-ray beams at sizes between 50 and 100 μm in width and

height (at an average flux of 10^{13} photons/s) for SAX studies. KB mirrors were used to deliver focused beam at 10–60 μm for WAX and micro-WAX studies at a flux of ~1×10^{-12} photon/s. The sample mount was positioned so that the sample was held at the focal point of the beam. The samples were mounted onto the sample compartment, shown in **Figure 5**, immediately before data collection. The sample compartment was mounted on an XY positioner to move the sample into the beam path. Multiple patterns from each sample were collected and the sample was moved, vertically and/or horizontally to limit radiation damage and improve the quality of data. Each point on the tendons was exposed to no more than 1 second to the X-rays. Based on the flux of the available beam, if saturation of pixels was observed on the diffraction pattern, additional attenuator foils were deployed. The same data collection strategies were used to collect data from control, glycated and iodine treated samples.

Amplitudes of each reflection from the diffraction patterns were obtained by azimuthally integrating the data from a sector covering the meridional reflection series (**Figure 6**). Difference Fourier and Patterson analyses was performed using the amplitudes and phases from native type I collagen as described previously [6, 11, 33]. After appropriate scaling, the data from SAX and WAX measurements were combined to improve the resolution of the interpretations.

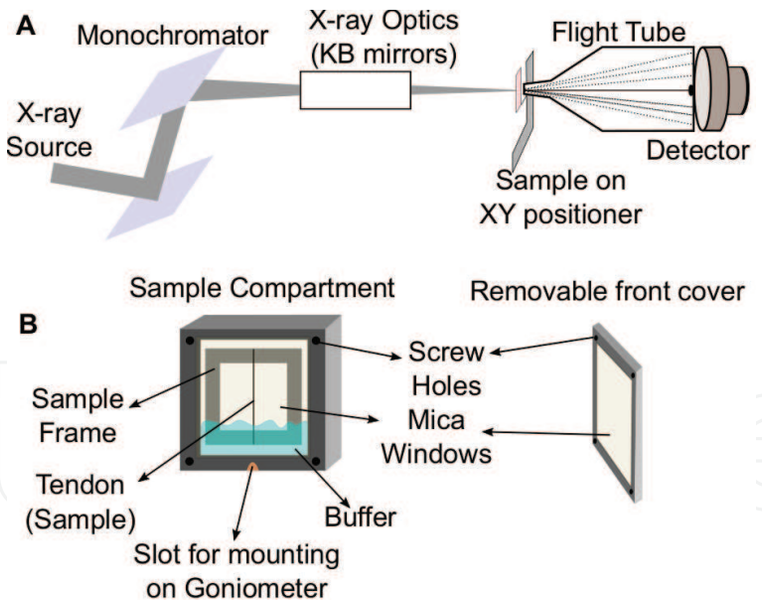

Figure 5. XRD beamline setup and a standard sample cell for fiber XRD data collection. (A) KB mirrors were used to deliver a focused beam at the sample position. The sample compartment was placed on a motorized XY-positioner to move the sample at the time of recording data. An evacuated flight tube is used to deliver the diffracted (scattered) X-rays to the marCCD detector at the end of the tube. The length of the tube can be varied for SAX and WAX experiments. (B) Research grade mica windows (thickness 0.15–0.21 mm) are used to cover the windows on either sides of the chamber to minimize data loss from any diffracted ray absorption from the compartment itself. An inner sample frame is used to tie the tendon onto and then be placed inside the compartment. A cylindrical slot in the bottom enables the entire rig to be mounted onto a goniometer head, which can then be controlled remotely. The buffer at the bottom of the compartment allows for hydration of tendon by capillary action.

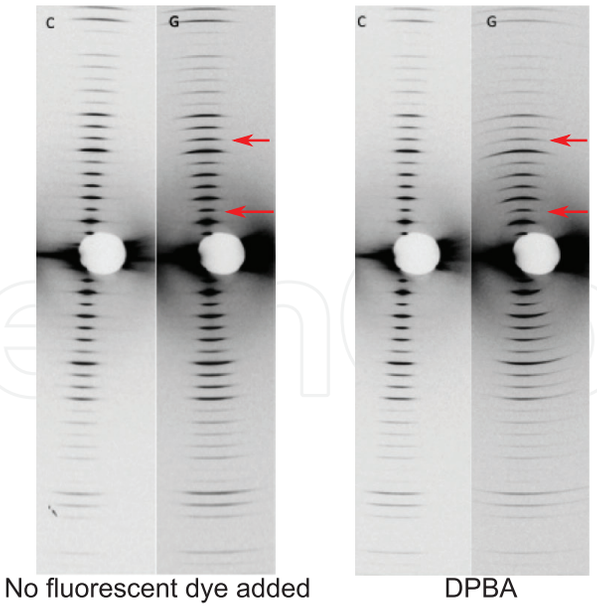

Figure 6. WAX diffraction patterns from non-treated (C), glucose- (G) and treated tendons. Red arrows show orders where an increase in the intensity of reflection was observed.

3. Results and discussion

X-ray diffraction data were interpreted to determine the location of both native crosslinks (those formed normally in non-diseased tissue) and those formed through the process of sugar incubation *in vitro*. As stated before, the latter sugar incubation induced cross-links simulate the *in vivo* process of formation of crosslinks through glycation/normal aging. Difference Fourier/electron density maps, to a resolution of ~35 Å (19 meridional orders) were calculated from these X-ray data. Interpretation of the D-periodic location of existing and newly formed crosslinks was then validated by comparison with difference Patterson maps. This verified the correct scaling of derivative to native data, the isomorphous nature of the derivative data and finally, the correct interpretation of where these crosslinks are from comparison of model Patterson functions based on the locations determined for crosslinks [6, 11, 33]. The peaks in difference Patterson maps are a result of addition of electron densities to specific regions within the unit cell, one of the hallmarks in processing of isomorphous data [9]. The cross verification of the labeling positions determined from difference Fourier maps with the relative periodicity spacing indicated by the difference Patterson maps provides information on the location (and plausible identity) of these crosslinks. This is important, since the Patterson maps derived from the experimental data, are unbiased by experimenter assumptions. Therefore, the close match between difference Patterson functions calculated from observed data and model Patterson functions calculated from the structure factors derived from a D-periodic model of crosslink locations, provides robust verification of the correct determination of the crosslink locations; all the more so, since the model of crosslink location is derived directly from difference Fourier maps.

3.1. Preliminary diffraction experiments with glycation

Data collected from glycated tendons showed minor changes in relative intensities (**Figure 6**). Considering the very small change in electron density brought about by the freshly formed, sugar-mediated crosslinks, this is to be expected. The small changes detected in intensities, however, indicate that there may be isomorphous changes that can be better characterized with increased electron density contrast. Contrast agent that binds (also isomorphously) to the reaction products (AGEs) where used to make these crosslinks more "visible" to X-rays.

3.2. Use of fluorescent dyes to detect crosslinks

As indicated, the use of fluorescent dyes was mandated to add electron density contrast, via isomorphous addition of electron density (contrast) to the crosslinks. Compounds known to bind to AGE-derived crosslinks for collagen were selected as potential candidates for multiple isomorphous replacement (MIR). 3-(Dansylamino) phenylboronic acid (DPBA) has been demonstrated to increase fluorescence signal by binding to the reaction product from glycation of bovine serum albumin (BSA) [32]. This property was used to increase the contrast of crosslinks (native and AGEs) in glycated and non-glycated collagen samples to reflect in the difference Fourier and Patterson maps.

3.3. Difference electron density analysis

As described earlier, difference Fourier and Patterson analysis was performed using the amplitudes of the Bragg reflections extracted from the diffraction patterns of glycated and non-glycated (control) tendons after treatment with DPBA. **Table 2** shows the matrix of difference Fourier and Patterson analyses performed.

The addition of electron densities due to glycation can be observed in the comparison between the difference Fourier maps displayed in **Figure 7**. Difference Patterson maps, generated from these diffraction amplitudes is show in **Figure 8**. The series of added densities can then be

	Non-glycated	**Non-glycated + DPBA**	**Glycated**	**Glycated + DPBA**
Non-glycated	–	Native Crosslinks (with contrast)	Sugar Crosslinks (without contrast)	Native + Sugar Crosslinks (with contrast)
Non-glycated + DPBA		–	Invalid calculation	Sugar Crosslinks (with contrast)
Glycated			–	Native Crosslinks
Glycated + DPBA				–

Table 2. Table showing difference Fourier and Patterson calculated from observed data. A difference Fourier/Patterson is simply put, the difference series calculation (the Fourier or the Patterson) between two data sets of structure functions (amplitudes). The red and blue text represent the respective series of the same colors shown in **Figure 7**.

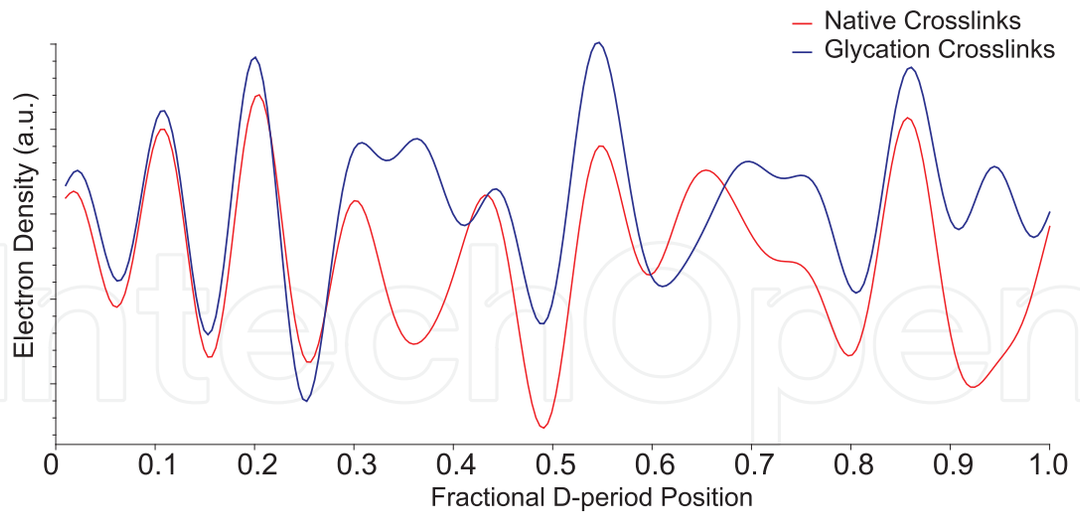

Figure 7. Difference Fourier map from glycated tendons treated with DPBA (blue) and non-glycated (native) tendons treated with DPBA (red) along 1D period represented fractional positions. The increase in electron densities or the introduction of new peaks in the glycated series, in comparison to the non-glycated, are a result of AGE crosslinks formed from the process of glycation.

used to synthesize a model difference Patterson to compare to that obtained from the experimental data to confirm scaling and the application of the right phase information (**Figure 8**). The specific sites of increased electron densities are better demonstrated in **Figure 9**.

3.4. Interpretation of sites of glycation

Lysine (K), arginine (R) and hydroxylysine (U) residues are the principal residues for crosslink (both enzymatic and non-enzymatic) formation. Close proximity of these residues to each

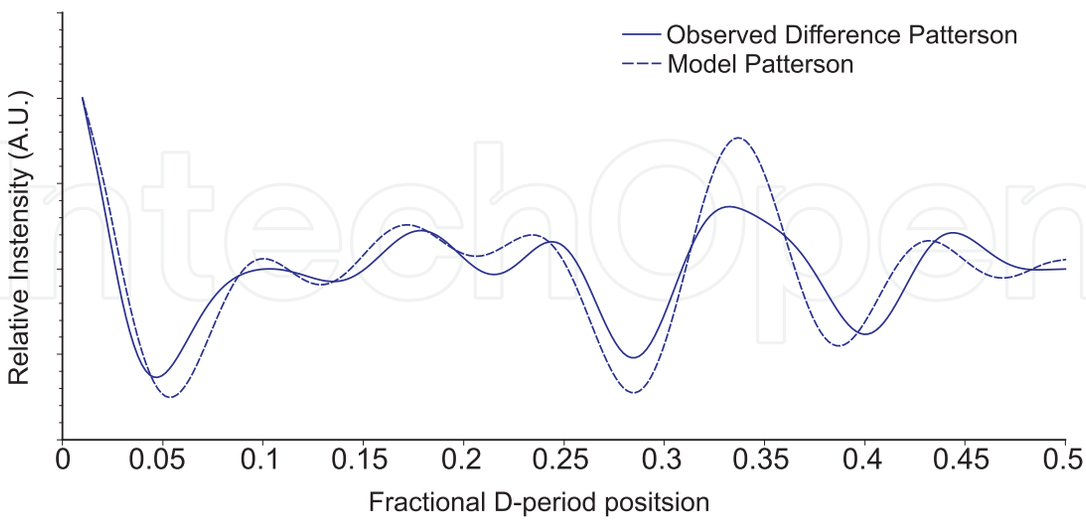

Figure 8. Model vs. observed difference Patterson maps. Difference Patterson map between glycated and non-glycated tendons with DPBA (solid blue line) along 1D period represented in fractional positions. The peak positions show the distances between electron dense regions seen in the difference Fourier map (**Figure 7**). The similarity model Patterson (dashed blue line) confirms that the scaling of derivative amplitudes from the XRD patterns and the application of native phases is appropriate, thus validating the sequence and structure specific interpretations.

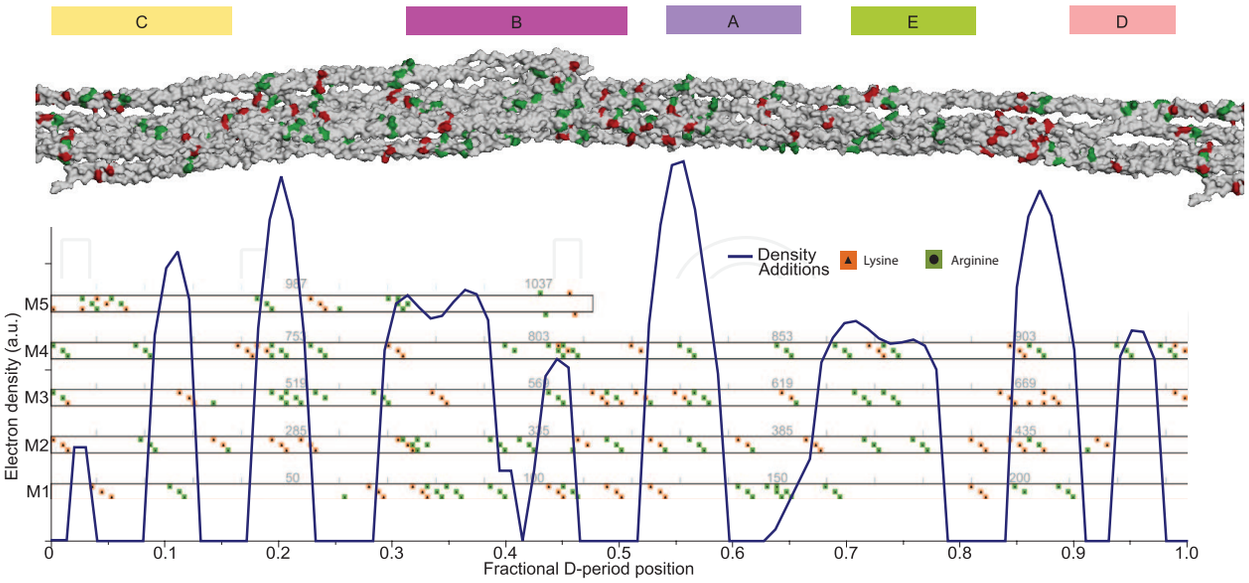

Figure 9. Candidate sites for glycation and regions of increased electron densities from crosslinking within a collagen D-period. A linear arrangement of amino acid sequences is shown with lysines (K) and hydroxylysines (U) marked in orange and arginines (R) residues marked in green. These positions are aligned with a 3D surface rendered model of the collagen microfibril. The linear and 3D representations are also presented in the context of the difference Fourier map (**Figure 7**) with the sites showing regions of increased electron density. An increase in electron density is observed around clusters of candidate residues in both representation, thereby supporting the validity of the results. See the discussion on the amino acid residues in each of these regions and their relevance to physiology and possible pathologies.

other within the three-dimensional structure is also a prerequisite to crosslink formation. Two very distant candidate residues are unlikely to be involved in crosslink formation, when compared to closely located residues, which cannot be reliably determined without reference to the three-dimensional structure. Candidate sites for glycation are spread out throughout the sequence (more specifically, the D-period) of collagen (**Figure 9**). These sites are of particular importance for the formation of mature enzymatic crosslinks between the non-helical telopeptides.

In the N-terminal telopeptide (0.025D) there are clusters of K and U residues which participate in the lysyl oxidase-mediated intermolecular crosslink formations. Furthermore, a series of R residues in this locale can be involved in AGE formation and hyperstabilize the enzymatic formed crosslinks at the N-terminus. Based on our data, these AGEs could potentially compete for the candidate K and U residues involved in the formation of enzymatic crosslinks, thus pathologically changing the nature of these crosslinks. Similarly, our data indicate at the C-terminal telopeptide (0.43D), a cluster of R residues participate in AGE crosslinking [16]. Candidate K and U crosslinking residues on the C-terminal telopeptide are normally the sites for enzymatic-intermolecular crosslinking. Together the N- and the C- terminus enzymatic crosslinks are prominently involved in the stabilization of the packing structure of collagen into fibrils. A possible site for glycation crosslink, adjacent to the N-terminal telopeptides was reported by Sweeney et al. [34].

Glycation-mediated crosslinks and possible sites for the formation of these crosslinks are described here. The results reported here now provide us with a context for interpretation

of the 3D structure of type I collagen with appropriate AGE-related crosslinks and study their effects in physiology and disease. For instance, our data indicate the formation of AGE crosslinks near the location of the high affinity integrin and von Willebrand Factor (vWF) sites. This may have great significance towards explaining the poor wound healing properties of diabetic tissues. If these crosslinks obscure these high-affinity cell-binding domains in the ECM, the tissue is likely significantly impacted in its ability to respond to both wound management and the repair of normal wear and tear. Potentially, these data can be used in the design of improved therapies, now that a molecular cause is indicated (see following discussion).

Notable changes are indicated by increased electron density of isomorphous derivatives, evident in difference Fourier maps, along the D-period (67 nm) of type I collagen fibrils in those derivative RTT tissues incubated with high concentrations of glucose. This is indicative of significant structural changes, i.e. crosslinking, and is observed from the experimental data. That these changes are crosslinks is further supported by the presence of crosslink labile amino acid sequences found in the regions of the D-period in which the additional electron density is observed, thereby confirming that the process of glycation is site-specific. The fact that these crosslinks alter the packing structure and functional (as in cell interactive for instance) properties of collagen is indicated by the proximity of these clusters of crosslinks to highly significant functional sites and regions in the collagen packing structure.

The peak indicating additional crosslinking, that appears at 0.1D confirms the glycation site proposed by Sweeney et al. [34], with a crosslink between monomers 3 and 4 of the microfibril. Additionally, our data indicate possible AGE crosslinking between monomers 1, 2 and 4. The peak at around 0.22D is observed, for which a cluster of principal sites is responsible. This location is of particular importance as it is located around the region where monomer 4 of one microfibril interacts with another neighboring collagen. The additional crosslinks around this location might hyper-stabilize and/or obscure the resulting interaction. This region is also closely placed with one of the $\alpha_1\beta_1$ and $\alpha_2\beta_1$ integrin binding sites (GFOGER) [35]. It is worth noting that one of the residues in this recognition site is an arginine, which might directly be involved in the AGE crosslink. This may well alter integrin-mediated cell adhesion through either direct interference (steric hindrance of integrin binding) or by obstructing/delaying the limited proteolytic activity of neighboring collagen molecules needed before integrin binds at this site.

The crosslinking electron density peaks at 0.3D, 0.38D and 0.46D surround a cluster of key cell interaction sites. For example, these AGEs might be significant with respect to the fibronectin and MMP binding sites. Alterations in this physiological phenomenon can severely change the accessibility and deposition of collagen in wound healing. Also within close proximity of this crosslink are several RGD sequences, which are principal recognition sites for 8 of the 24 integrins and a possible explanation for advanced rate of aging and poor wound healing ability in diabetic patients [36]. The peaks indicating crosslinks at 0.38D and the 0.46D are in the region of the $(GPO)_5$ site, which is a potent site for platelet binding and the process of clotting [4]. This could constitute part of a plausible explanation for

the suppressed blood clotting in diabetic patients. This is also of great importance in the development of successful surgical implantation materials, such a heart valves and capillaries for anastomoses where the ability of the implant material to support clotting is vital. Furthermore, a crosslink at 0.46D (gap-overlap interface) can interfere with the fibrillogenesis control sequence (FCS) in this region. This site controls the formation of fibrils by forming mature crosslinks to stabilize the molecular packing of collagen monomers into fibrils and fibers. The crosslinks at the gap-overlap interface could also alter the packing in such a manner that the MMP interaction site becomes more accessible, thereby leading to increased proteolysis. Our data also suggest that the crosslink at 0.46D can also interfere with the GMOGER site in this region. This has been identified as an important site for integrin-mediated cell binding.

The 0.55D crosslink sites appears immediately adjacent to the C-terminal telopeptides and at the beginning of the gap region. At this location, monomer 2 of one microfibril can form a crosslink with neighboring collagen monomers to stabilize the packing structure. The significance of 0.55D crosslinks are not obvious until it is considered in the context of the 0.38, 0.46 and the aforementioned 0.55D clusters of crosslinks. Together, they surround the enzymatic cleavage-site of collagenase for fibrillar collagen around 0.44D on monomer 4. If there were significant glycation induced crosslinking on either side of the collagenase cleavage site, then it is possible that even if the enzyme is able to access its substrates-site, which will not be a given, it might have little or no impact on the proteolysis of fibrillar collagen as the microfibril and fibril will remain intact due to the pathological crosslinks bridging the enzymatic break in the collagen molecule. In other words, extensive AGE crosslinking can be seen in this molecular model, to prevent or hinder wound healing and removal of damaged collagenous tissue, which coincides with clinical observations also [21, 37–39].

The peaks indicating crosslinks at the 0.7–0.77D (and a separate peak at 0.88D) are due to the high concentration of principal sites of glycation in these regions. In physiology, this region is responsible for the binding of small leucine-rich repeat proteoglycans (sLRRP PGs), which, using their long glycosaminoglycan chains (GAG) help in bundling fibrils into stable structures [5]. The crosslinks formed at these sites could result in inherent detachment of the PGs leading to disintegration of fibrils. Similar disintegration of fibrils in type II collagen in human articular cartilage, as a result of treatment with the biglycan (PG) antibody has been exhibited as a plausible mechanism for rheumatoid arthritis [33]. Similar conclusions can be drawn with the disintegration of type I collagen fibrils resulting in disease. This region is also the principal recognition site for von Willebrand factor binding and crosslinks here could lead to faulty coagulation in diabetics. These crosslinks are a possible explanation to how the process of thrombosis is affected in diabetic patients [40]. The 0.95D crosslink forms a link between the N-terminal of the neighboring collagen molecule. This region is the principal binding region for amyloid precursor protein (APP) and is implicated in the pathology of Alzheimer's disease.

Although all potentially functionally significant, we draw attention again to the significance of the additional glycation induced crosslinking that is indicated to occur at 0.38, 0.46 and

0.55D. As these locations surround the collagenase cleavage-site, these crosslinks could be implicated to interfere with the normal and healthy removal and repair of damaged collagenous tissues. Crosslinks effect these processes, firstly by limiting the availability of the collagen's cleavage site and interaction region at 0.44D; secondly, by hyperstabilizing the packing structure, even after the enzymatic cleavage occurs, thereby defeating the release of unhealthy collagen fibrils. The 0.7–0.77D region is a second area indicating significant glycation induced crosslinking, which directly interferes with and/or hypostabilizes principal matrix-matrix association regions (via proteoglycan mediated inter-fibrillar crosslinking GAG chains) and notable, the von Willebrand factor interaction site in the fibrillar collagen's gap region.

4. Conclusions

There is a strong correlation between the availability of sites for glycation, namely lysine, arginine and hydroxylysine residues, their placement in the microfibrillar arrangement and the electron density peaks indicating crosslinking in sugar incubated tendons (**Figure 9**). The regions of increased electron density, as a result of glycation induced crosslinks, directly correspond with candidate amino acid crosslinking sites along the collagen microfibril. This data appears to experimentally confirm several previously estimated locations for crosslinking [34, 41]. The significance of these data to our understanding of the fundamental structure of connective tissues and the effects of AGE induced cross-links are immediately apparent as discussed above and as follows.

The structure and packing of type I collagen have been objects of study for many decades, with the resolution of packing coming to a better level of understanding. There are various structural features that contribute to the final fibrillar packing, resulting in the classic 67 nm repeating D-period. Crosslinking is one of the most important features to stabilize their packing into fibrils and in turn, the organization of these fibrils into fibers and the bases of tissues and organogenesis. The three-dimensional structure and its derived model for collagen, obtained from high-resolution X-ray diffraction studies [2, 7, 16] accounts for enzyme-mediated crosslinks, formed as a result of post-translational modification. However, molecular dynamics simulations of the packing of type I collagen predict a 11–19% shrinkage of the 67 nm D-period – which is not what is observed in nature! Therefore, there are additional parameters required for molecular dynamic simulations to resemble the observed natural structure. Some of this can be attributed to the formation of AGEs as a result of normal physiological aging and/or sugar-mediated crosslinking [42]. The data presented here, give us insight into the location of these crosslinks including which amino acid residues are involved and can be used to refine the packing structure and associated simulations of type I collagen. These simulations can be used to study plausible collagen related disease mechanisms, to support experimental and clinical data and aid with the development of appropriate therapeutics.

Author details

Rama Sashank Madhurapantula[1] and Joseph P.R.O. Orgel[2]*

*Address all correspondence to: orgel@iit.edu

1 Department of Biology, Pritzker Institute of Biomedical Science and Engineering, Illinois Institute of Technology, Chicago IL, USA

2 Departments of Biology, Physics and Biomedical Engineering, Pritzker Institute of Biomedical Science and Engineering, Illinois Institute of Technology, Chicago IL, USA

References

[1] Orgel JPR, Miller A, Irving TC, Fischetti RF, Hammersley AP, Wess TJ. The in situ supermolecular structure of type I collagen. Structure. 2001;**9**(11):1061-1069. DOI: 10.1016/S0969-2126(01)00669-4

[2] Orgel JPRO, Irving TC, Miller A, Wess TJ. Microfibrillar structure of type I collagen in situ. Proceedings of the National Academy of Sciences. 2006;**103**(24):9001-9005. DOI: 10.1073/pnas.0502718103

[3] Orgel JPRO, Irving TC. Advances in fiber diffraction of macromolecular assembles. In: Meyers RA, editor. Encyclopedia of Analytical Chemistry. Chichester, UK: John Wiley & Sons, Ltd; 2014. pp. 1-26

[4] Orgel JPRO, San Antonio JD, Antipova O. Molecular and structural mapping of collagen fibril interactions. Connective Tissue Research. 2011;**52**(1):2-17. DOI: 10.3109/03008207.2010.511353

[5] Orgel JPRO, Eid A, Antipova O, Bella J, Scott JE. Decorin core protein (decoron) shape complements collagen fibril surface structure and mediates its binding. PLoS One. 2009;**4**(9):e7028. DOI: 10.1371/journal.pone.0007028

[6] Antipova O, Orgel JPRO. In situ D-periodic molecular structure of type II collagen. Journal of Biological Chemistry. 2010;**285**(10):7087-7096. DOI: 10.1074/jbc.M109.060400

[7] Orgel JP. The molecular structure of collagen [thesis]. University of Stirling; 2000

[8] Madhurapantula RS. Studies on connective and neurological tissues in relation to disease [thesis]. Ann Arbor: Illinois Institute of Technology; 2015

[9] Blundell TL, Johnson LN. Protein Crystallography 1st ed. United States: Elsevier Science Publishing Co Inc; 1976. ISBN: 9780121083502

[10] Orgel JPRO, Irving TC. Advances in fiber diffraction of macromolecular assembles. In: Encyclopedia of Analytical Chemistry. John Wiley & Sons, Ltd; 2006. DOI: 10.1002/9780470027318.a9420

[11] Orgel JPRO, Persikov AV, Antipova O. Variation in the helical structure of native collagen. PLoS One. 2014;**9**(2):e89519. DOI: 10.1371/journal.pone.0089519

[12] Eyre DR, Paz MA, Gallop PM. Cross-linking in collagen and elastin. Annual Review of Biochemistry. 1984;**53**(1):717-748. DOI: 10.1146/annurev.bi.53.070184.003441

[13] Henkel W, Glanville RW. Covalent crosslinking between molecules of type I and type III collagen. European Journal of Biochemistry. 1982;**122**(1):205-213. DOI: 10.1111/j.1432-1033.1982.tb05868.x

[14] Last JA, Armstrong LG, Reiser KM. Biosynthesis of collagen crosslinks. International Journal of Biochemistry. 1990;**22**(6):559-564. DOI: 10.1016/0020-711X(90)90031-W

[15] Sims TJ, Avery NC, Bailey AJ. Quantitative Determination of Collagen Crosslinks. In: Streuli CH, Grant ME. (eds) Extracellular Matrix Protocols. Methods in Molecular Biology, vol 139. Humana Press; 2000. p. 11-26. DOI: 10.1385/1-59259-063-2:11

[16] Orgel JP, Wess TJ, Miller A. The in situ conformation and axial location of the intermolecular cross-linked non-helical telopeptides of type I collagen. Structure. 2000;**8**(2):137-142. DOI: 10.1016/S0969-2126(00)00089-7

[17] Mitome J, Yamamoto H, Saito M, Yokoyama K, Marumo K, Hosoya T. Nonenzymatic cross-linking pentosidine increase in bone collagen and are associated with disorders of bone mineralization in dialysis patients. Calcified Tissue International. 2011;**88**(6):521-529. DOI: 10.1007/s00223-011-9488-y

[18] Tomkins O, Garzozi HJ. Collagen cross-linking: Strengthening the unstable cornea. Clinical Ophthalmology (Auckland, N.Z.). 2008;**2**(4):863-867

[19] Alhayek A, Lu P-R. Corneal collagen crosslinking in keratoconus and other eye disease. International Journal of Ophthalmology. 2015;**8**(2):407-418. DOI: 10.3980/j.issn.2222-3959.2015.02.35

[20] Singh R, Barden A, Mori T, Beilin L. Advanced glycation end-products: A review. Diabetologia. 2001;**44**(2):129-146. DOI: 10.1007/s001250051591

[21] Peppa M, Uribarri J, Vlassara H. Glucose, advanced glycation end products, and diabetes complications: What is new and what works. Clinical Diabetes. 2003;**21**(4):186-187. DOI: 10.2337/diaclin.21.4.186

[22] Priego Capote F, Sanchez J-C. Strategies for proteomic analysis of non-enzymatically glycated proteins. Mass Spectrometry Reviews. 2009;**28**(1):135-146. DOI: 10.1002/mas.20187

[23] Brownlee M, Vlassara H, Cerami A. Nonenzymatic glycosylation products on collagen covalently trap low-density lipoprotein. Diabetes. 1985;**34**(9):938-941

[24] Cho S-J, Roman G, Yeboah F, Konishi Y. The road to advanced glycation end products: A mechanistic perspective. Current Medicinal Chemistry. 2007;**14**(15):1653-1671. DOI: 10.2174/092986707780830989

[25] Thornalley PJ, Langborg A, Minhas HS. Formation of glyoxal, methylglyoxal and 3-deoxyglucosone in the glycation of proteins by glucose. The Biochemical Journal. 1999; **344**(Pt 1):109-116

[26] Snedeker JG, Gautieri A. The role of collagen crosslinks in ageing and diabetes - The good, the bad, and the ugly. Muscles, Ligaments and Tendons Journal. 2014;**4**(3):303-308

[27] Pischetsrieder M. Chemistry of glucose and biochemical pathways of biological interest. Peritoneal Dialysis International: Journal of the International Society for Peritoneal Dialysis. 2000;**20**(Suppl 2):S26-S30

[28] Wess TJ, Miller A, Bradshaw JP. Cross-linkage sites in type I collagen fibrils studied by neutron diffraction. Journal of Molecular Biology. 1990;**213**(1):1-5. DOI: 10.1016/S0022-2836(05)80115-9

[29] Mallia AK, Hermanson GT, Krohn RI, Fujimoto EK, Smith PK. Preparation and use of a boronic acid affinity support for separation and quantitation of glycosylated hemoglobins. Analytical Letters. 1981;**14**(8):649-661. DOI: 10.1080/00032718108055476

[30] Hayashi Y, Makino M. Fluorometric measurement of glycosylated albumin in human serum. Clinica Chimica Acta. 1985;**149**(1):13-19

[31] Guan Y, Zhang Y. Boronic acid-containing hydrogels: Synthesis and their applications. Chemical Society Reviews. 2013;**42**(20):8106. DOI: 10.1039/c3cs60152h

[32] Sattarahmady N, Moosavi-Movahedi AA, Ahmad F, et al. Formation of the molten globule-like state during prolonged glycation of human serum albumin. Biochimica et Biophysica Acta. 2007;**1770**(6):933-942. DOI: 10.1016/j.bbagen.2007.02.001

[33] Antipova O, Orgel JPRO. Non-enzymatic decomposition of collagen fibers by a biglycan antibody and a plausible mechanism for rheumatoid arthritis. PLoS One. 2012; **7**(3):e32241. DOI: 10.1371/journal.pone.0032241

[34] Sweeney SM, Orgel JP, Fertala A, et al. Candidate cell and matrix interaction domains on the collagen fibril, the predominant protein of vertebrates. Journal of Biological Chemistry. 2008;**283**(30):21187-21197. DOI: 10.1074/jbc.M709319200

[35] Knight CG, Morton LF, Peachey AR, Tuckwell DS, Farndale RW, Barnes MJ. The collagen-binding A-domains of integrins α1β1 and α2β1 recognize the same specific amino acid sequence, GFOGER, in native (triple-helical) collagens. Journal of Biological Chemistry. 2000;**275**(1):35-40. DOI: 10.1074/jbc.275.1.35

[36] Kim JP, Zhang K, Kramer RH, Schall TJ, Woodley DT. Integrin receptors and RGD sequences in human keratinocyte migration: Unique anti-migratory function of alpha 3 beta 1 epiligrin receptor. The Journal of Investigative Dermatology. 1992;**98**(5):764-770

[37] Hennessey PJ, Ford EG, Black CT, Andrassy RJ. Wound collagenase activity correlates directly with collagen glycosylation in diabetic rats. Journal of Pediatric Surgery. 1990;**25**(1):75-78. DOI: 10.1016/S0022-3468(05)80167-8

[38] Peppa M, Stavroulakis P, Raptis SA. Advanced glycoxidation products and impaired diabetic wound healing. Wound Repair and Regeneration. 2009;**17**(4):461-472. DOI: 10.1111/j.1524-475X.2009.00518.x

[39] Peppa M, Brem H, Ehrlich P, et al. Adverse effects of dietary glycotoxins on wound healing in genetically diabetic mice. Diabetes. 2003;**52**(11):2805-2813. DOI: 10.2337/diabetes.52.11.2805

[40] Kessler L, Wiesel ML, Attali P, Mossard JM, Cazenave JP, Pinget M. Von Willebrand factor in diabetic angiopathy. Diabetes & Metabolism. 1998;**24**(4):327-336

[41] Gautieri A, Redaelli A, Buehler MJ, Vesentini S. Age- and diabetes-related nonenzymatic crosslinks in collagen fibrils: Candidate amino acids involved in Advanced Glycation End-products. Matrix Biology. 2014;**34**:89-95. DOI: 10.1016/j.matbio.2013.09.004

[42] Varma S, Botlani M, Hammond JR, Scott HL, Orgel JPRO, Schieber JD. Effect of intrinsic and extrinsic factors on the simulated D-band length of type I collagen. Proteins: Structure, Function, and Bioinformatics. 2015;**83**:1800-1812. DOI: 10.1002/prot.24864

IntechOpen