



# Age- and diabetes-related nonenzymatic crosslinks in collagen fibrils: Candidate amino acids involved in Advanced Glycation End-products



Alfonso Gautieri<sup>a,b,\*</sup>, Alberto Redaelli<sup>a</sup>, Markus J. Buehler<sup>b</sup>, Simone Vesentini<sup>a</sup>

<sup>a</sup> Biomechanics Group, Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy

<sup>b</sup> Laboratory for Atomistic and Molecular Mechanics, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave. Cambridge, MA, USA

## ARTICLE INFO

### Article history:

Received 27 June 2013

Received in revised form 16 September 2013

Accepted 16 September 2013

Available online 21 September 2013

### Keywords:

Collagen

Atomistic simulations

Advanced glycation end-products

Crosslinks

Ageing

Diabetes

## ABSTRACT

Ageing and diabetes share a common deleterious phenomenon, the formation of Advanced Glycation Endproducts (AGEs), which accumulate predominantly in collagen due to its low turnover. Though the general picture of glycation has been identified, the detailed knowledge of which collagen amino acids are involved in AGEs is still missing. In this work we use an atomistic model of a collagen fibril to pinpoint, for the first time, the precise location of amino acids involved in the most relevant AGE, glucosepane. The results show that there are 14 specific lysine–arginine pairs that, due to their relative position and configuration, are likely to form glucosepane. We find that several residues involved in AGE crosslinks are within key collagen domains, such as binding sites for integrins, proteoglycans and collagenase, hence providing molecular-level explanations of previous experimental results showing decreased collagen affinity for key molecules. Altogether, these findings reveal the molecular mechanism by which glycation affects the biological properties of collagen tissues, which in turn contribute to age- and diabetes-related pathological states.

© 2013 Elsevier B.V. Open access under [CC BY-NC-ND license](http://creativecommons.org/licenses/by-nc-nd/3.0/).

## 1. Introduction

There is a wealth of clinical data available in the literature on age-related alterations in the functioning of all human tissue, including susceptibility to injury and to reduce healing capacity (Semba et al., 2010). Noticeably, these deleterious effects of ageing all involve, to a greater or lesser degree, the most important structural protein, namely, collagen. Collagen molecules make up 30% of total protein in the body and form the basis of many vital organs such as the heart, bone, kidney and skin (Kadler et al., 2007). The collagen molecule is synthesized as a trimeric molecule containing two  $\alpha 1$ , and one  $\alpha 2$  chains, each of about 1000 amino acids. Upon secretion from the cell, collagen molecules assemble into fibrils and are enzymatically crosslinked (Sweeney et al., 2008).

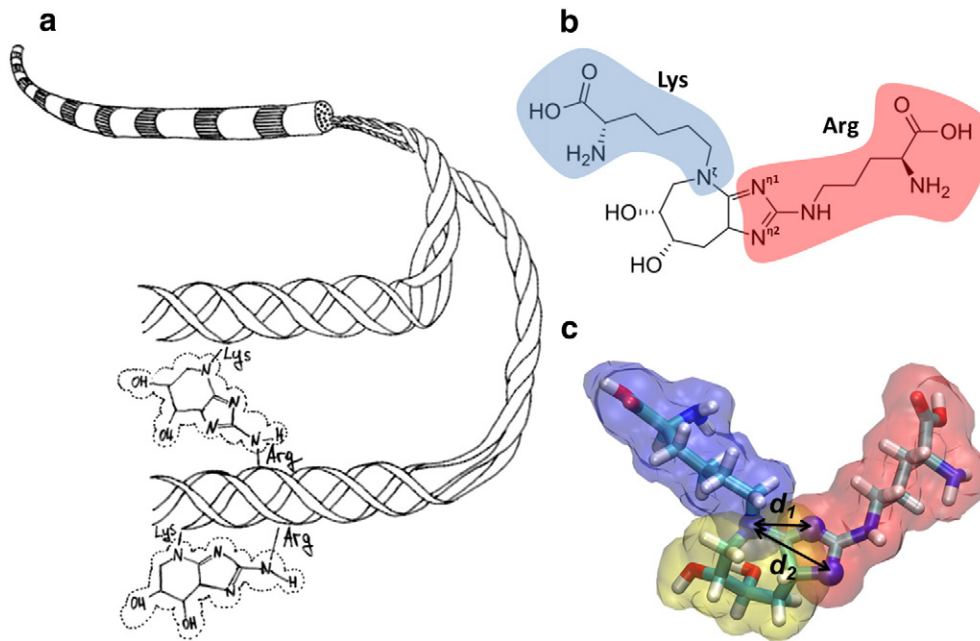
In aging, type I collagen becomes less flexible and more acid insoluble, which correlates with the accumulation of nonenzymatic crosslinks known as Advanced Glycation End-products (AGEs) (Brennan, 1989; Bailey, 2001). Glycation and AGEs form *in vivo* on collagen via nonenzymatic reactions that covalently add a sugar moiety onto the protein (Paul and Bailey, 1996) and their accumulation is particularly high in long-lived proteins, such as collagen. Indeed, collagen half-life varies between tissues but remains generally large, from 1 to 2 years for bone collagen to about 10 years for type I in skin (Avery and Bailey, 2006). The low biological turnover of collagen makes it therefore susceptible to

interaction with metabolites, primarily glucose. Besides the elderly, people who suffer with type II diabetes are particularly badly affected by AGE cross-linking (Andreassen et al., 1981; Schneider and Kohn, 1982). Whilst elder people have abundance of long-lived proteins that have slowly accumulated AGE crosslinks, type II diabetics have abnormally high levels of glucose in their system leaving a surplus which is available to glycate proteins.

Several glycation crosslinks have been proposed (Avery and Bailey, 2006) but are all present in minute quantities except glucosepane (Sell et al., 2005), which is a lysine–arginine crosslinking AGE which could make a significant change to the biomechanics and biological activity of fibrillar collagen (Fig. 1a). Nonetheless, despite the recognized importance of AGEs, there are still several important open questions about their role in the onset of pathological conditions. In particular, there are two different aspects of collagen glycation that need to be further understood. One is the biomechanical effects of nonenzymatic intermolecular crosslinking: glucose reaction with the amino acid side-chains, and subsequent further reaction to form a crosslink with an adjacent collagen molecule, result in a modification of the physical properties of the collagen (e.g.: elasticity), but the detailed effects of AGEs on collagen nanomechanics are still unknown. The second aspect is related to collagen interactions: glycation of specific amino acid involved in intermolecular recognition could lead to the dramatic modification of the interaction of collagen with other molecules such as proteoglycans (PG), enzymes (e.g., collagenase) and cell integrins. However, it is still not clear which collagen amino acids – and thus which collagen binding sites – are mostly affected by AGEs.

\* Corresponding author at: Biomechanics Group, Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy. Tel.: +39 022399 4142.

E-mail address: [alfonso.gautieri@polimi.it](mailto:alfonso.gautieri@polimi.it) (A. Gautieri).



**Fig. 1.** Collagen fibril and glucosepane. Panel (a) shows a schematic of collagen fibril and the formation of glucosepane, which covalently links Lysine and Arginine sidechains. The two amino acids can belong to separate molecules (top), forming an intermolecular crosslink, or to the same molecule (bottom). Panel (b) shows the chemical structure of glucosepane, which crosslinks lysine (blue) to arginine (red). Panel (c) shows the molecular model of glucosepane, used to evaluate the distances between the terminal atoms of the two residues ( $d_1 \approx 2.6$  Å,  $d_2 \approx 3.8$  Å).

Here we use a three-dimensional full atomistic molecular model of the collagen microfibril (Gautieri et al., 2011) to pinpoint which specific amino acids are involved in the formation of the most important AGE crosslinks (i.e., glucosepane) within the collagen fibril. The molecular details of this model allow to infer about the positions of likely lysine–arginine crosslinking sites and, when coupled with the mapping of collagen binding sites (Sweeney et al., 2008), it allows to relate the glycation sites with biological malfunctions.

## 2. Results

We use the atomistic model of collagen type I microfibril, based on X-ray crystallography (Orgel et al., 2006) and described in detail in previous works (Gautieri et al., 2011; Nair et al., 2013), to identify the most likely position of AGE crosslinks. Our investigation is limited to a single specific AGE, namely glucosepane, because its levels are found to be up to 1000 times higher than any other nonenzymatic crosslink (Monnier et al., 2005; Sell et al., 2005). The molecular model of glucosepane (a lysine–arginine crosslink) shows that in the compound the distance between the terminal atoms of the two amino acids involved is between 2.6 Å and 3.8 Å (see Fig. 1b–c). The larger distance (increased by 30%) is used as a cutoff to identify, within the atomistic three-dimensional collagen microfibril model, lysine–arginine pairs that are close enough to form glucosepane crosslinks. Results show that there are 14 lysine–arginine pairs that spend at least 25% of the time closer than the cutoff distance (see Table 1), and thus are located in a relative position such that, in the presence of glucose, they are able to form a glucosepane crosslink. Within the most frequently observed pairs there are 6 lysine residues from  $\alpha 2$  chain (8 from  $\alpha 1$  chain) and 5 arginine residues from  $\alpha 2$  chain (9 from  $\alpha 1$  chain). Out of 14 crosslinks, six are intramolecular, whereas 8 involve lysine and arginine residues from different collagen molecules. The full list of lysine–arginine pairs (86) that are found closer than the cutoff distance in at least one frame of the molecular dynamics trajectory is given in Table S1.

### 2.1. Lysine residues

Two lysine residues most often found close to an arginine residue (LYS  $\alpha 1$ -434 and LYS  $\alpha 2$ -453) are also known to be two of the four preferentially glycosylated residues (Reiser et al., 1992). The other two preferred glycation sites, i.e. LYS  $\alpha 2$ -479 and LYS  $\alpha 2$ -924, are found close to arginine with a much lower frequency (see full list in Table S1) and thus are less likely to form glucosepane crosslink (even if glycosylated). Interestingly, two lysine–arginine pairs that most likely form nonenzymatic crosslinks involve LYS  $\alpha 1$ -87, which is known to form enzyme-mediated crosslinks (Eyre et al., 2008). Since physiological enzyme-mediated crosslinks are formed at young age and then their number remains constant (Monnier et al., 2005) whereas nonenzymatic crosslinks appear at older age, the LYS  $\alpha 1$ -87 residues already involved in physiological crosslinks are most likely not involved in AGE crosslinks. However, the LYS  $\alpha 1$ -87 (there are two for each collagen molecule) that are not involved in physiological crosslinks could be glycosylated and involved in nonenzymatic crosslinks. Concerning the collagen binding sites, many of the lysines involved in glucosepane formation are within or close to integrin, heparin or proteoglycan (keratan sulphate) binding sites. Notably, lysine  $\alpha 1$ -531 is within the cell interaction domain, which features sequence GFPGER<sub>502–507</sub>, which is considered to be critical for collagen–cell interaction (Sweeney et al., 2008).

### 2.2. Arginine residues

Several arginines involved in AGE crosslinks are found within or close to integrin, heparin or proteoglycan (keratan sulphate or dermatan sulphate) binding sites (see Table 1). In particular, ARG  $\alpha 1$ -501 is located right next to the GFPGER<sub>502–507</sub> sequence, which is a key collagen–cell interaction domain. Another crucial residues involved in glucosepane formation is ARG  $\alpha 1$ -789, which is close to the binding site of matrix metalloproteinase-1 (MMP-1), the most important collagenase enzyme (Sweeney et al., 2008).

**Table 1**

Glucosepane-forming lysine–arginine pairs. Each amino acid is identified by the helical number and chain (notation  $\alpha 1$ -a and  $\alpha 1$ -b are used to distinguish the two  $\alpha 1$  chains). For each pair it is reported the time in which the terminal atoms of the side chains are closer than the cutoff of 5 Å, thus in a configuration promoting the formation of glucosepane crosslink. Involvement of major intermolecular interactions or binding sites (BS) is reported in the notes (Sweeney et al., 2008). Intermolecular AGEs (noted with \*) are likely to affect the mechanics of the tissue, providing load transfer between molecules in addition to enzymatic crosslinks.

Residue	Chain	Residue	Chain	Time closer than cutoff (%)	Intermolecular crosslinks	Notes		
LYS	434	$\alpha 1$ -a	ARG	906	$\alpha 2$	96.84%	*	LYS: preferred glycation site, close to $\alpha 1\beta 2$ integrin BS.
LYS	564	$\alpha 1$ -b	ARG	567	$\alpha 1$ -a	93.97%		LYS: close to keratan sulphate BS. ARG: within keratan sulphate BS.
LYS	87	$\alpha 1$ -a	ARG	789	$\alpha 1$ -a	87.64%	*	LYS: involved in enzymatic x-link, within heparin BS. ARG: close to MMP-1 cleavage site.
LYS	531	$\alpha 1$ -a	ARG	291	$\alpha 2$	83.62%	*	LYS: within the cell interaction domain, $\alpha 1\beta 1/\alpha 2\beta 1$ integrin BS.
LYS	453	$\alpha 2$	ARG	453	$\alpha 1$ -b	73.28%		LYS: preferred glycation site. ARG: within keratan sulphate BS.
LYS	420	$\alpha 2$	ARG	888	$\alpha 1$ -b	69.25%	*	ARG: within dermatan sulphate BS.
LYS	264	$\alpha 2$	ARG	501	$\alpha 1$ -b	69.25%	*	LYS: within $\alpha 2\beta 1$ integrin BS. ARG: within cell interaction domain, close to $\alpha 1\beta 1/\alpha 2\beta 1/\alpha 1\beta 1$ integrin BS.
LYS	87	$\alpha 1$ -a	ARG	90	$\alpha 1$ -a	50.57%		LYS: involved in enzymatic x-link, within heparin BS. ARG: within $\alpha 1\beta 1$ integrin BS, within heparin BS.
LYS	657	$\alpha 1$ -a	ARG	420	$\alpha 1$ -b	45.69%	*	
LYS	564	$\alpha 2$	ARG	90	$\alpha 1$ -b	32.18%	*	LYS: close to keratan sulphate BS. ARG: within $\alpha 1\beta 1$ integrin BS, within heparin BS.
LYS	252	$\alpha 1$ -a	ARG	252	$\alpha 2$	28.74%		
LYS	290	$\alpha 2$	ARG	291	$\alpha 1$ -a	27.87%		
LYS	108	$\alpha 1$ -b	ARG	342	$\alpha 2$	25.57%	*	LYS: within $\alpha 1\beta 1$ integrin BS.
LYS	974	$\alpha 2$	ARG	977	$\alpha 2$	25.57%		

### 3. Discussion

Previous works have shown that the major nonenzymatic crosslink in collagen is glucosepane (Monnier et al., 2005; Sell et al., 2005), an AGE crosslink involving lysine and arginine residues. It has been shown that the presence of AGEs affects the biological properties of collagen, and in particular its interaction with a series of extracellular matrix molecules (Andreassen et al., 1981; Monnier and Cerami, 1981; Kent et al., 1985; Reigle et al., 2008), suggesting that glycation affects certain specific collagen binding sites. However, the identification of which exact amino acids, along the triple helix, are involved in the formation of AGEs have remained elusive. The novelty of our work is that we pinpoint, for the first time, the exact position of 14 lysine–arginine pairs that are located, within the collagen fibril, close enough to form nonenzymatic crosslinks associated with the ageing process and diabetes (see Table 1).

Among these proposed crosslinks, 8 are intermolecular crosslinks, whereas 6 involve amino acids from the same molecule. In addition to these 14 crosslinks, there are several other lysine–arginine pairs (72) that are close enough only for a minor part of the time and thus less likely – but still able – to form AGE crosslinks (see full list in Table S1).

During its biosynthesis, collagen is subject to several post-translational modifications, where the most important are proline hydroxylation, lysine hydroxylation and hydroxylysine glycosylation. The last two modifications could potentially affect AGE formation, since they involve lysine, but their effect is likely negligible. Lysine hydroxylation leaves intact the amine group which is required for glucosepane formation and thus hydroxylysines can participate in AGE formation (Saito and Marumo, 2010). In the case of glycosylated hydroxylysine the amine group is preserved but the bulky sugar chains could prevent the AGE formation by steric hindrance. However, glycosylation in type I collagen – the main focus of our work and the most common type – is rather low ( $\approx 1\%$ ) (Kivirikko and Myllyla, 1979; Shinkai and Yonemasu, 1979), thus not affecting the overall AGE formation.

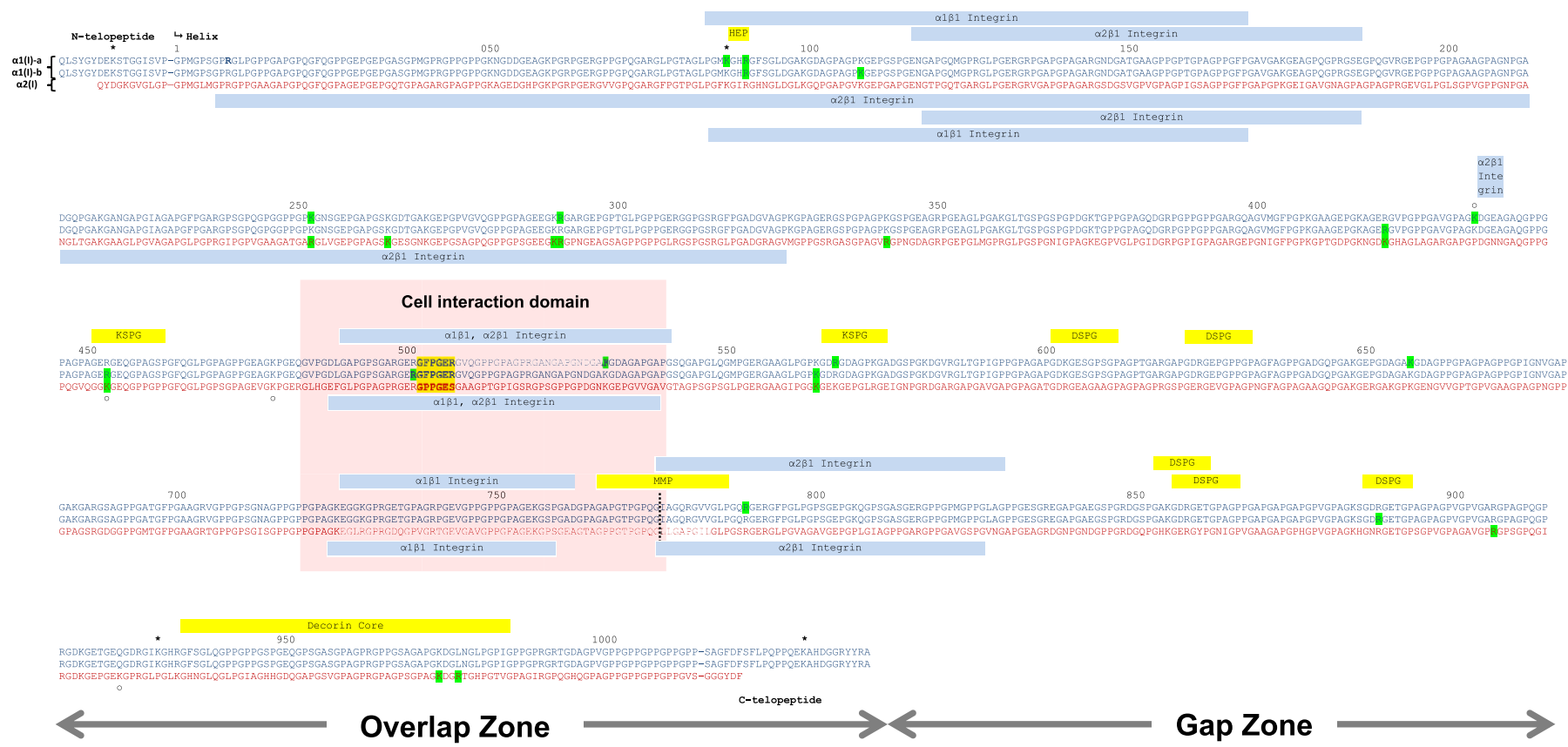
The effects of AGEs are twofold: 1) mechanical: intermolecular crosslinking may result in major alterations of the physical properties (e.g., increased fibril stiffness), which are observed to change slowly with age and at higher rate in diabetes mellitus due to high glucose levels; 2) intermolecular interactions: AGEs alter the biochemical profile of the collagen molecule and if they occur at specific sites can affect the intermolecular and cell–collagen interactions.

#### 3.1. Mechanical effects

The mechanical effects of AGEs on collagenous tissues are long known and include stiffening, increased failure load and decreased viscoelasticity of tendons (Galeski et al., 1977; Andreassen et al., 1981; Danielsen and Andreassen, 1988; Li et al., 2013), loss of bone plasticity and toughness (Tang and Vashishth, 2011; Zimmermann et al., 2011), stiffening and increased fragility of cartilage (Verzijl et al., 2002). We find that there are 8 lysine residues for each collagen molecule that are often very close to an arginine residue of an adjacent molecule and thus, in the presence of glucose, they are likely to form an intermolecular crosslink. Simplified mechanical models of tendon fibrils (Buehler, 2008) and bone (mineralized) fibrils (Siegmund et al., 2008) have previously investigated in a parametric way the possible effects of increased crosslinking. These works have shown that already a single nonenzymatic crosslink (in addition to the physiological enzymatic ones) is able to drastically change the mechanical properties of collagen tissues, inducing increased stiffness and decreased toughness. In this work we show that the number of intermolecular crosslinks is likely  $\approx 10$  per molecule and thus undoubtedly responsible for the observed stiffening of collagen tissues.

#### 3.2. Effects on molecular recognition

In Fig. 2 we have represented the amino acids involved in nonenzymatic crosslinks on the map of the major collagen domains, to help find the possible influence of AGE on collagen interaction with other molecules. Glycated collagen has been found to show reduced affinity for heparin and keratan sulphate proteoglycan, but not for dermatan sulphate and decorin proteoglycans, alongside with impaired endothelial cell migration (Reigle et al., 2008). Our results provide strong evidences that glucosepane crosslinks are close or directly involve the two proposed binding sites for keratan sulphate (residues 451–461 and 565–575) and for heparin (residues 87–90), but do not involve the decorin core binding site nor the five proposed dermatan sulphate binding sites (except for arginine  $\alpha 1$ -888) (see Fig. 2). Moreover, several amino acids proposed to form AGE crosslinks lie within the cell-interaction domain. In particular, ARG  $\alpha 1$ -501 is located right next to the binding sites of  $\alpha 1\beta 1/\alpha 2\beta 1$  integrins (GFPGER<sub>502–507</sub>), which are the predominant type I collagen receptors on endothelial cells (Di Lullo et al., 2002). We also found that arginine  $\alpha 1$ -789 is very close to the cleavage site of MMP-1, the most important enzyme



Green: residues involved in glucosepane crosslink  
 Pink background: major cell interaction domain  
 Yellow: major binding sites for proteoglycans, collagenases, heparin and integrin (GFPGER)  
 Cyan: secondary integrin binding sites  
 \* Enzymatic cross-links

○ Preferred glycation sites  
 --- MMP-1 Cleavage site  
 HEP: Heparin  
 KSPG: Keratan sulphate proteoglycan  
 DSPG: Dermatan sulphate proteoglycan

**Fig. 2.** Human collagen type I fibril domains and major binding sites. The collagen triple helix sequence is shown (source Pubmed,  $\alpha 1(1)$  accession #NP000079.2 and  $\alpha 2(1)$  NP000080.2) and the proposed location of glucosepane-forming amino acids is highlighted in green. Most of these amino acids fall close or within major binding sites for proteoglycans, integrins and enzymes (yellow) or within secondary binding sites for integrins (blue). Collagen domains are obtained from (Sweeney et al., 2008). Green: residues involved in glucosepane crosslink. Pink background: major cell interaction domain. Yellow: major binding sites for proteoglycans, collagenases, heparin and integrin (GFPGER). Cyan: secondary integrin binding sites. \* Enzymatic cross-links. ○ Preferred glycation sites. --- MMP-1 Cleavage site. HEP: Heparin. KSPG: Keratan sulphate proteoglycan. DSPG: Dermatan sulphate proteoglycan.

responsible for the degradation and remodeling of extracellular matrix. This modification close to the enzyme binding site due to glycation and crosslinking could explain the observed resistance to collagenases of glycated collagen (Andreassen et al., 1981; Monnier and Cerami, 1981; Kent et al., 1985).

### 3.3. Other AGEs

Glucosepane has been found to be largely more common than other AGEs, with levels 10–1000 times higher than other AGEs (Sell et al., 2005). The second most important AGE crosslink, named MOLD, is a Lysine–Lysine crosslink (Monnier et al., 2005). Hence it is interesting to investigate if this is due to a geometric probability. For this we investigate the average number of Lys–Arg pairs in close proximity (which could lead to glucosepane formation) and the number of Lys–Lys pairs (which could lead to MOLD). As it can be anticipated from the sequence of the three chains, our analysis of Lys–Lys pairs shows that there are hundreds of Lys–Lys pairs for which the time spent closer than cutoff is higher than 80%. All these pairs are formed by same-position Lysines, arising from the fact that the two  $\alpha 1$  chains are identical and the  $\alpha 2$  chain has a high homology. Apart from same-position Lysines, there are a few pairs with high (>25%) geometric propensity for AGE formation, reported in Table 2. These results suggest that Lys–Lys AGEs are favored from the geometric point of view, particularly thanks to the proximity of same-position Lysines. This in turn suggests that the chemical pathway must be much more favorable for glucosepane, in order to explain its experimentally observed predominance. Furthermore, it is important to observe that all AGE cross links are either Lys–Arg or Lys–Lys (Monnier et al., 2005), hence the differences in the levels of different AGEs cannot be due to the proximity of involved amino acids but must be dictated by the availability of cross-linking agent (e.g., glucose, fructose, methylglyoxal, ...) and on the specific differences in formation kinetics.

## 4. Conclusions

Although the age- and diabetes-related changes in collagen physical properties have been known for several decades, the molecular mechanisms are only now being unraveled. In this work, for the first time, we identify the specific amino-acids involved in nonenzymatic crosslinking of collagen, and discuss the mechanical and biological consequences, which include stiffening of the tissue and impaired interaction with proteoglycans, collagenases and cell's integrins.

Our approach is meant to be complementary to biochemical assays and aims to identify glycation points with single-residue precision. This possibility has been unlocked by the recent development of the atomistic model of collagen fibril (Gautieri et al., 2011). The identification of glycation sites is based on a geometric basis: indeed, LYS and ARG residues to form a covalent bond must be close enough for the glucose-mediated cross-link to be formed. The major limitation of the work is related to the fact that vicinity in space represents a necessary but not sufficient condition for cross-link to occur. In this view, the list of LYS/ARG pairs must be read as a likelihood of AGE crosslinks formation. A further limitation is that our model does

not take in consideration post-translational modification and stable interaction with extracellular matrix biomolecules that could protect binding sites from glycation.

Glycation is believed to provide changes in mechanical properties, it would be extremely interesting to investigate the changes in mechanics in light of the discoveries put forth in this work. The fully atomistic modeling of AGEs mechanical effects is impractical for several reasons: the atomistic model used in this work is based on the microfibril unit cell and employs periodic boundary conditions to model the entire fibril, which hinders the possibility to add mechanical crosslinks that involve the periodic images. On the other hand, considering a non-periodic atomistic model of a collagen fibril would result in a model size with prohibitive computational costs. A feasible strategy could be the use of simplified models similar to those developed in previous works (Buehler, 2008; Siegmund et al., 2008) that would greatly reduce the particles of the system. In this case, the modeling of finite size fibrils would result in high but treatable computational costs. However, this strategy involves the careful parameterization of the coarse-grain model and in particular of the AGE crosslinks. This would allow investigating in a parametric way the influence of the amount AGEs crosslinks, which is largely uncertain: it has been proposed (Sell et al., 2005) that there is one AGE every 2 molecules in diabetic collagen and one AGE every 5 molecules in aged collagen. However, these data do not distinguish between intra-molecular crosslinks (which are anticipated to not contribute to mechanical properties) and inter-molecular crosslinks (which would stiffen the fibril). Furthermore, it is still unknown how AGEs are distributed through the fibril cross-section (i.e., if they are homogeneously distributed or only on the surface) and it could be very interesting to test the different hypotheses.

Life-expectancy has increased substantially in recent years and this is a general trend in western countries. This, coupled to a decrease in birth rate, means that society now is becoming year by year older. The every-day effects of aging are clear for all to see: the increased incidence of arthritis and diabetes, the reduced heart and kidney function, the reduced mobility etc. Most of these put a strain on global health services, as well as decreasing of independence and the burden of medical care. But what if we could moderate the aging process? The most promising approach likely involves the use of deglycating enzymes, however available enzymes are either ATP-dependent (thus not working in the extracellular environment) or are only able to deglycate small substrates (and thus not suitable for large proteins such as collagen) (Sell and Monnier, 2012). Hence, the *de novo* engineering of deglycation enzymes is mandatory for the successful reversion of AGEs and, to this aim, knowing the precise substrate (i.e., the local collagen sequence) where the engineered enzyme should act represents a necessary preliminary step. In this view our work could pave the way for the development of advanced strategies for the mitigation of ageing and diabetes molecular effects.

## 5. Computational procedures

The goal is to exploit full-atomistic modeling of collagen that enables us to perform a systematic study of collagen glycation from a fundamental, molecular point of view. This allows us to answer questions such as “what are the amino acids involved in AGE crosslinks?” and to provide molecular-level explanations for the experimentally-observed modification of collagen biological and mechanical properties.

### 5.1. Collagen atomistic model

The geometry and composition of the molecular model are fully described in previous works (Gautieri et al., 2011; Nair et al., 2013). Briefly, the collagen microfibril model used here is based on Protein Data Bank entry 3HR2 (Orgel et al., 2006). Since the structure reported in (Orgel et al., 2006) includes only backbone alpha carbons and the

**Table 2**  
List of AGE-forming lysine–lysine pairs, excluding same-position lysines.

Residue	Chain	Residue	Chain	Time closer than cutoff (%)
LYS	408	$\alpha 2$	LYS 174	$\alpha 2$ 67.7%
LYS	936	$\alpha 1$ -a	LYS 933	$\alpha 2$ 61.3%
LYS	290	$\alpha 1$ -b	LYS 59	$\alpha 2$ 31.9%
LYS	884	$\alpha 2$	LYS 648	$\alpha 2$ 30.6%
LYS	974	$\alpha 1$ -b	LYS 270	$\alpha 2$ 25.0%

primary sequence of *rattus norvegicus*, we used homology modelling to obtain a full-atom structure with the human collagen type I sequence. The fibril model is solvated using the solvate plugin of VMD (Humphrey et al., 1996) by adding TIP3P water molecules. Since the molecule at physiological pH includes a net charge (positive net charge +34), counterions ( $\text{Cl}^-$ ) are added in order to keep the system neutral. The final solvated all-atom system contains  $\approx 75,000$  atoms, including  $\approx 36,000$  water atoms. The first step of energy minimization is performed by a steepest descent algorithm using the NAMD 2.9 code and the CHARMM force field (MacKerell et al., 1998), which includes parameters for hydroxyproline amino acid (HYP) found in collagen (Park et al., 2005). This force field has been widely validated for a variety of biochemical models of proteins including collagen (Gautieri et al., 2009, 2012a, 2012b; Qin et al., 2012).

### 5.2. Fibril equilibration

Extensive full atomistic simulations are carried out using the NAMD 2.9 code. Rigid bonds are used to constrain covalent bond lengths involving hydrogen atoms, thus allowing an integration time step of 2 fs. Nonbonding interactions are computed using a cut-off for neighbor list at 1.35 nm, with a switching function between 1.0 and 1.2 nm for van der Waals interactions, while the Particle-Mesh Ewald sums (PME) method is applied to describe electrostatic interactions. The fibril model is equilibrated through 100 ns molecular dynamics simulations at a temperature of 310 K (37 °C). The structural convergence of the structure is reached within the first 30 ns, hence the remaining 70 ns are used for analysis.

### 5.3. Identification of nonenzymatic crosslinks

Nonenzymatic crosslinks derived from AGEs are identified on a geometric basis. The principal crosslinking AGE is glucosepane (Sell et al., 2005), which is lysine–arginine crosslinks derived from glucose. The molecular model of glucosepane is built and simulated for 100 ps using LAMMPS code (Plimpton, 1995) and the ReaxFF force field (Duin et al., 2001). The trajectory of the simulation is used to define the average distances between the lysine  $\text{N}^\zeta$  atom and the two arginine  $\text{N}^{\eta 1}$  atoms, involved in the formation of the crosslink (see Fig. 1b–c). The average distances are found to be 2.6 Å (between lysine  $\text{N}^\zeta$  and arginine  $\text{N}^{\eta 1}$  atom) and 3.8 Å (between lysine  $\text{N}^\zeta$  and arginine  $\text{N}^{\eta 2}$  atom). Glucosepane crosslink can form in the presence of glucose only for lysine–arginine pairs that are close enough for the reaction to occur. Hence, the 100 ns molecular dynamics trajectory of the collagen microfibril is analyzed through an in-house VMD script in order to find lysine–arginine pairs for which the distance between lysine  $\text{N}^\zeta$  and any one arginine  $\text{N}^{\eta 1}$  is below a 5 Å cutoff, chosen as 130% the larger of the two calculated distances within glucosepane) and the basis of previous studies suggesting that a proximity of  $\approx 5$  Å is a good predictor of AGE formation (Acosta et al., 2000).

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.matbio.2013.09.004>.

### Acknowledgments

The authors would like to thank Prof. Jess G. Snedeker and Gion Fessel from University and ETH Zurich for the many fruitful discussions concerning this work. This work has been supported by Fondazione Cariplo, grant n°2011-0270. High-performance computing resources have been provided by CINECA Consortium through the ISCRA initiative and by DEISA Consortium through the PRACE initiative. MJB acknowledges support from the Office of Naval Research (N000141010562). The authors declare no competing financial interest.

### References

- Acosta, J., Hettinga, J., Fluckiger, R., Krumrei, N., Goldfine, A., Angarita, L., Halperin, J., 2000. Molecular basis for a link between complement and the vascular complications of diabetes. *Proc. Natl. Acad. Sci. U. S. A.* 97, 5450–5455.
- Andreassen, T.T., Seyerhansen, K., Bailey, A.J., 1981. Thermal-stability, mechanical-properties and reducible cross-links of rat tail tendon in experimental diabetes. *Biochim. Biophys. Acta* 677, 313–317.
- Avery, N.C., Bailey, A.J., 2006. The effects of the Maillard reaction on the physical properties and cell interactions of collagen. *Pathol. Biol. (Paris)* 54, 387–395.
- Bailey, A.J., 2001. Molecular mechanisms of ageing in connective tissues. *Mech. Ageing Dev.* 122, 735–755.
- Brennan, M., 1989. Changes in solubility, non-enzymatic glycation, and fluorescence of collagen in tail tendons from diabetic rats. *J. Biol. Chem.* 264, 20947–20952.
- Buehler, M.J., 2008. Nanomechanics of collagen fibrils under varying cross-link densities: atomistic and continuum studies. *J. Mech. Behav. Biomed.* 1, 59–67.
- Danielsen, C.C., Andreassen, T.T., 1988. Mechanical-properties of rat tail tendon in relation to proximal distal sampling position and age. *J. Biomech.* 21, 207–212.
- Di Lullo, G.A., Sweeney, S.M., Korkko, J., Ala-Kokko, L., San Antonio, J.D., 2002. Mapping the ligand-binding sites and disease-associated mutations on the most abundant protein in the human, type I collagen. *J. Biol. Chem.* 277, 4223–4231.
- Eyre, D.R., Weis, M.A., Wu, J.J., 2008. Advances in collagen cross-link analysis. *Methods* 45, 65–74.
- Galeski, A., Kastelic, J., Baer, E., Kohn, R.R., 1977. Mechanical and structural changes in rat tail tendon induced by alloxan diabetes and aging. *J. Biomech.* 10, 775–782.
- Gautieri, A., Vesentini, S., Redaelli, A., Buehler, M.J., 2009. Inter-molecular slip mechanism in tropocollagen nanofibrils. *Int. J. Mater. Res.* 100, 921–925.
- Gautieri, A., Vesentini, S., Redaelli, A., Buehler, M.J., 2011. Hierarchical structure and nanomechanics of collagen microfibrils from the atomistic scale up. *Nano Lett.* 11, 757–766.
- Gautieri, A., Pate, M.I., Vesentini, S., Redaelli, A., Buehler, M.J., 2012a. Hydration and distance dependence of intermolecular shearing between collagen molecules in a model microfibril. *J. Biomech.* 45, 2079–2083.
- Gautieri, A., Vesentini, S., Redaelli, A., Buehler, M.J., 2012b. Viscoelastic properties of model segments of collagen molecules. *Matrix Biol.* 31, 141–149.
- Humphrey, W., Dalke, A., Schulten, K., 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33.
- Kadler, K.E., Baldock, C., Bella, J., Boot-Handford, R.P., 2007. Collagens at a glance. *J. Cell Sci.* 120, 1955–1958.
- Kent, M.J.C., Light, N.D., Bailey, A.J., 1985. Evidence for glucose-mediated covalent cross-linking of collagen after glycosylation invitro. *Biochem. J.* 225, 745–752.
- Kivirikko, K.I., Myllylä, R., 1979. Collagen glycosyltransferases. *Int. Rev. Connect. Tissue Res.* 8, 23–72.
- Li, Y., Fessel, G., Georgiadis, M., Snedeker, J.G., 2013. Advanced glycation end-products diminish tendon collagen fiber sliding. *Matrix Biol.* 32 (3–4), 169–177.
- MacKerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, W.E., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., Karplus, M., 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102, 3586–3616.
- Monnier, V.M., Cerami, A., 1981. Nonenzymatic browning in vivo: possible process for aging of long-lived proteins. *Science* 211, 491–493.
- Monnier, V.M., Mustata, G.T., Biemel, K.L., Reihl, O., Lederer, M.O., Dai, Z.Y., Sell, D.R., 2005. Cross-linking of the extracellular matrix by the Maillard reaction in aging and diabetes – an update on “a puzzle nearing resolution”. *Ann. N. Y. Acad. Sci.* 1043, 533–544.
- Nair, A.K., Gautieri, A., Chang, S.W., Buehler, M.J., 2013. Molecular mechanics of mineralized collagen fibrils in bone. *Nat. Commun.* 4.
- Orgel, J.P.R.O., Irving, T.C., Miller, A., Wess, T.J., 2006. Microfibrillar structure of type I collagen in situ. *Proc. Natl. Acad. Sci. U. S. A.* 103, 9001–9005.
- Park, S., Radmer, R.J., Klein, T.E., Pande, V.S., 2005. A new set of molecular mechanics parameters for hydroxyproline and its use in molecular dynamics simulations of collagen-like peptides. *J. Comput. Chem.* 26, 1612–1616.
- Paul, R.G., Bailey, A.J., 1996. Glycation of collagen: the basis of its central role in the late complications of ageing and diabetes. *Int. J. Biochem. Cell Biol.* 28, 1297–1310.
- Plimpton, S., 1995. Fast parallel algorithms for short-range molecular-dynamics. *J. Comput. Phys.* 117, 1–19.
- Qin, Z., Gautieri, A., Nair, A.K., Inbar, H., Buehler, M.J., 2012. Thickness of hydroxyapatite nanocrystal controls mechanical properties of the collagen-hydroxyapatite interface. *Langmuir* 28, 1982–1992.
- Reigle, K.L., Di Lullo, G., Turner, K.R., Last, J.A., Chervoneva, I., Birk, D.E., Funderburgh, J.L., Elrod, E., Germann, M.W., Surber, C., Sanderson, R.D., Antonio, J.D.S., 2008. Non-enzymatic glycation of type I collagen diminishes collagen-proteoglycan binding and weakens cell adhesion. *J. Cell. Biochem.* 104, 1684–1698.
- Reiser, K.M., Amigable, M.A., Last, J.A., 1992. Nonenzymatic glycation of type-I collagen – the effects of aging on preferential glycation sites. *J. Biol. Chem.* 267, 24207–24216.
- Saito, M., Marumo, K., 2010. Collagen cross-links as a determinant of bone quality: a possible explanation for bone fragility in aging, osteoporosis, and diabetes mellitus. *Osteoporos. Int.* 21, 195–214.
- Schnider, S.L., Kohn, R.R., 1982. Effects of age and diabetes-mellitus on the solubility of collagen from human-skin, tracheal cartilage and dura mater. *Exp. Gerontol.* 17, 185–194.
- Sell, D.R., Monnier, V.M., 2012. Molecular basis of arterial stiffening: role of glycation – a mini-review. *Gerontology* 58, 227–237.

- Sell, D.R., Biemel, K.M., Reihl, O., Lederer, M.O., Strauch, C.M., Monnier, V.M., 2005. Glucosepane is a major protein cross-link of the senescent human extracellular matrix. Relationship with diabetes. *J. Biol. Chem.* 280, 12310–12315.
- Semba, R.D., Nicklett, E.J., Ferrucci, L., 2010. Does accumulation of advanced glycation end products contribute to the aging phenotype? *J. Gerontol. A Biol. Sci. Med. Sci.* 65, 963–975.
- Shinkai, H., Yonemasu, K., 1979. Hydroxylysine-linked glycosides of human complement subcomponent C1q and various collagens. *Biochem. J.* 177, 847–852.
- Siegmund, T., Allen, M.R., Burr, D.B., 2008. Failure of mineralized collagen fibrils: modeling the role of collagen cross-linking. *J. Biomech.* 41, 1427–1435.
- Sweeney, S.M., Orgel, J.P., Fertala, A., McAuliffe, J.D., Turner, K.R., Di Lullo, G.A., Chen, S., Antipova, O., Perumal, S., Ala-Kokko, L., Forlino, A., Cabral, W.A., Barnes, A.M., Marini, J.C., San Antonio, J.D., 2008. Candidate cell and matrix interaction domains on the collagen fibril, the predominant protein of vertebrates. *J. Biol. Chem.* 283, 21187–21197.
- Tang, S.Y., Vashishth, D., 2011. The relative contributions of non-enzymatic glycation and cortical porosity on the fracture toughness of aging bone. *J. Biomech.* 44, 330–336.
- van Duin, A.C.T., Dasgupta, S., Lorant, F., Goddard, W.A., 2001. ReaxFF: a reactive force field for hydrocarbons. *J. Phys. Chem. A* 105, 9396–9409.
- Verzijl, N., DeGroot, J., Ben Zaken, C., Braun-Benjamin, O., Maroudas, A., Bank, R.A., Mizrahi, J., Schalkwijk, C.G., Thorpe, S.R., Baynes, J.W., Bijlsma, J.W.J., Lafeber, F.P.J.G., TeKoppele, J.M., 2002. Crosslinking by advanced glycation end products increases the stiffness of the collagen network in human articular cartilage – a possible mechanism through which age is a risk factor for osteoarthritis. *Arthritis Rheum.* 46, 114–123.
- Zimmermann, E.A., Schaible, E., Bale, H., Barth, H.D., Tang, S.Y., Reichert, P., Busse, B., Alliston, T., Ager, J.W., Ritchie, R.O., 2011. Age-related changes in the plasticity and toughness of human cortical bone at multiple length scales. *Proc. Natl. Acad. Sci. U. S. A.* 108, 14416–14421.