INVESTIGATIONS OF LYSINE AND LYSINE-DERIVED CROSSLINKS IN ELASTIN
VIA SOLID-STATE NMR SPECTROSCOPY

A THESIS SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF
HAWAI'I AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF

MASTER OF SCIENCE

IN

CHEMISTRY

DECEMBER 2019

By

Paulina Panek

Thesis Committee:

Kristin Kumashiro, Chairperson
Philip Williams
Rui Sun

DEDICATION

I dedicate this thesis to the people who made my journey possible. The first scientist I have ever met whose old university textbooks sparked my curiosity – my mother Grażyna Panek. My two closest friends, Patrycja Polak and Dilan Ustek, whose encouragement and support meant more than I could ever express with words. Finally, I would like to thank my very first chemistry teacher Ms. Maria Biczel, whose uplifting words were what I needed to hear to finish this thesis.

ACKNOWLEDGEMENTS

ABSTRACT

Elastin is responsible for elasticity and resiliency of force-bearing tissues in vertebrates. This polymeric protein is produced from a ~70 kDa precursor, tropoelastin, During the post-translational modification, a network of inter- and intramolecular covalent crosslinks is formed from lysine residues. As many as eleven different bi-, tri-, and tetrafunctional modifications are present in mature elastin, which are critical for normal tissue function. These products have only been characterized in hydrolysates and enzymatically-cleaved peptides, but never observed directly in the intact protein.

Two isotopes of lysine, [U-^{13}C,^{15}N-Lys] and [^{13}Cε-Lys], were independently incorporated into the neonatal rat smooth muscle cell (NRSMC) elastin. Samples with high levels of enrichment (~90%) were used for high-resolution solid-state nuclear magnetic resonance (ssNMR) studies. One- and two-dimensional measurements allowed for analysis of the secondary structures of lysines. Random coil and α-helices were the predominant conformations. The helical content is higher in a frozen sample than at the physiological temperature. Experiments based on cross-polarization allowed for quantification of unmodified lysine as well as desmosine and isodesmosine, crosslinks unique to elastin. The presence of rare bi- and trifunctional crosslinks was confirmed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

^{1}H	proton
^{13}C	carbon-13 isotope
^{15}N	nitrogen-15 isotope
1D	one-dimensional
2D	two-dimensional
Å	Angstrom
AA	allysine aldol
Ab	ion abundance
ADES	allodesmosine
Ala, A	alanine
Arg, R	arginine
Asn, N	asparagine
Asp, D	aspartic acid
B_0	external magnetic field
B_1	transverse magnetic field
BB	backbone
BMRB	biological magnetic resonance data bank
BLN	bovine ligamentum nuchae
CD	circular dichroism
CNBr	cyanogen bromide
CO	carbonyl carbon
ComPmultiCP	composite-pulse multi cross polarization
COPD	chronic obstructive pulmonary disease
CP	cross-polarization

CPN	cyclopentasine
CSA	chemical shift anisotropy
CW	continuous wave
Cα	carbon-alpha
Cβ	carbon-beta
Cγ	carbon-gamma
Cδ	carbon-delta
Cε	carbon-epsilon
d_1	recycle delay
Da	Dalton
DARR	dipolar assisted rotational resonance
Des	desmosine
DLNL	dehydrolysinonorleucine
DMDES	dehydromerodesmosine
DMEM	Dulbecco's modified Eagle's medium
DP	direct polarization
DSS	sodium 2,2-dimethyl-2-silapentane-5-sulfonate
ECM	extracellular matrix
ESI	electrospray ionization
EX	exon
exp	exponential
f_1	indirect dimension in the frequency domain
f_2	direct dimension in the frequency domain
FA	formic acid
FBS	fetal bovine serum
FDAA	1-fluoro-2,4-dinitrophenyl-5-L-alanine amide
FDLA	1-fluoro-2-4-dinitrophenyl-5-L-leucine amide

FID	free-induction decay
FWHM	full width half maximum
GHz	gigahertz
Gln, Q	glutamine
Glu, E	glutamic acid
Gly, G	glycine
h	Planck's constant
HH	Hartmann-Hahn
HCl	hydrochloric acid
HOSE	hierarchically ordered spherical environment
HPLC	high-performance liquid chromatography
Hz	hertz
IDES	isodesmosine
IDP	intrinsically disordered protein
IR	infrared
I_z	z-component of the angular spin operator I
k	force constant
K	Kelvin
kDa	kilodalton
kHz	kilohertz
LC	liquid chromatography
Leu, L	leucine
LNL	lysinonorleucine
LOX	lysyl oxidase
Lya	allysine
Lys, K	lysine
Lys equiv.	lysine equivalents

$(Lys)_n$	polylysine with n number of residues
MAS	magic-angle spinning
MDES	merodesmosine
mg	milligram
MH	protonated molecular ion
MHz	megahertz
mL	milliliter
mm	millimeter
mM	millimolar
MS	mass spectrometry
ms	millisecond
$NaHCO_3$	sodium bicarbonate
NDES	neodesmosine
NEAA	non-essential amino acids
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NRE	neighboring residue effects
NRSMC	neonatal rat smooth muscle cells
ns	nanosecond
OH	hydroxyl
PDB	Protein Data Bank
PEN	pentasine
Phe, F	phenylalanine
ppm	parts per million
Pro, P	proline
QUCP	quantitative cross polarization
R_c	ratio of isotopic abundance

R_e	ratio of isotopic abundance in enriched analytes
rf	radio frequency
rINEPT	refocused insensitive nuclei enhanced by polarization transfer
rSNOB	refocused selective excitation for biochemical applications
s	second
S/N	signal-to-noise ratio
Ser, S	serine
ssb	spinning sideband
ssNMR	solid-state nuclear magnetic resonance
S_z	z-component of the angular spin operator S
T	temperature
t	time
T_1	longitudinal spin-lattice relaxation time constant
$T_{1\rho}$	longitudinal spin-lattice relaxation time constant in the rotating frame
T_2	transverse relaxation time constant
$T_2{}'$	coherence lifetimes during spin-echo
T_{CH}	carbon-proton cross-polarization time constant
T-75	tissue culture flask with 75 cm^2 surface area
TE	tropoelastin
TFA	Tetrafluoroacetic acid
T_g	glass transition temperature
TOF	time-of-flight
TOSS	total sideband supression
TPPM	two-pulse phase modulation
Tyr, Y	tyrosine
UV	ultraviolet
Val, V	valine

VT	variable-temperature
°C	degree Celsius
ΔG	relative free energy
ΔG^*	relative free energy representing the bottom of a free energy basin
$\Delta \delta$	difference in chemical shifts
γ	gyromagnetic ratio
η	noe enhancement factor
μL	microliter
μm	micrometer
μs	microsecond
τ, τ'	time delay
τ_r	rotor period
ω_r	rotor spinning speed

CHAPTER 1

INTRODUCTION

1.1 Elastin

Elastin is a protein found in the extracellular matrix of all vertebrates except primitive cyclostomes. It gives elasticity and resilience to blood vessels, skin, lungs, aorta, ligaments, and many more, as it is found in almost every organ. This protein undergoes billions of stretch-recoil cycles without significant turnover.[1] In fact, the elastic fiber is normally a metabolically stable unit over the human lifespan.[2]

Elastin is a polymeric protein stabilized through covalent crosslinks derived from lysine (Lys). These post-translational modifications differentiate mature elastin from its soluble precursor, tropoelastin (TE). The amino acid composition of the TE monomer varies, depending on the species. However, it generally consists of about 800 residues, with glycine (Gly), alanine (Ala), proline (Pro), and valine (Val) being the main constituents.[3]

1.2 Elasticity models

Up to this day, both the mechanism of elasticity and the three-dimensional structure of elastin remain unclear. The extensive network of crosslinks makes elastin insoluble in common solvents. Therefore, solution nuclear magnetic resolution (NMR) spectroscopy cannot be used to resolve the structure of the mature protein. In addition, elastin does not crystallize, which excludes X-ray crystallography from the pool of available characterization methods. The postulated models of elasticity are based on the limited experimental data, as no high-resolution technique is applicable for this protein.

In the network model (figure 1a), elastin is considered to be a rubber-like elastomer; the chains are permanently crosslinked but remain kinetically free.[4] Entropy is assumed to drive the recoil after stretching. This model agrees with some early studies.[5,6] However, unlike rubbers, elastin is not self-lubricating and has the elastic properties only when swollen in water.[7] In addition, the network model does not account for the coacervation of elastin[8], that is, the process of monomer self-association during the elastic fiber assembly.[9] This phenomenon occurs in solutions of TE heated to 37°C and results in the proper alignment of monomers for crosslink formation.[3]
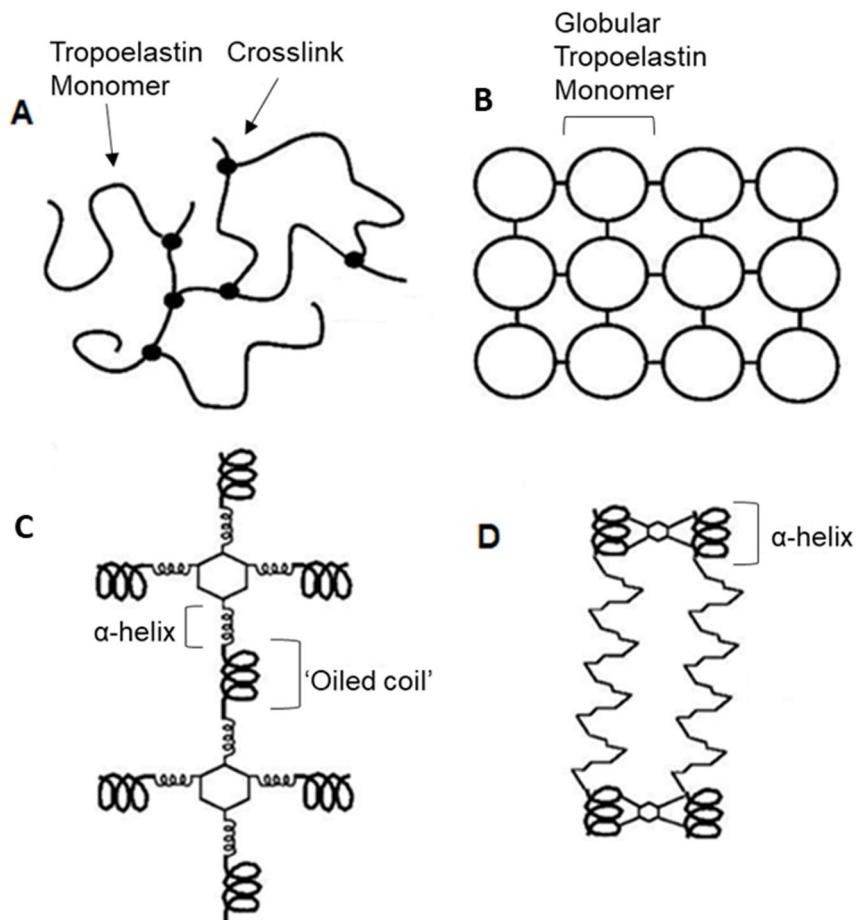
Figure 1.1: Models of elasticity.
a) Network model,[4] (b) liquid drop model,[10] (c) oiled coil model,[11] (d) fibrillar model.[12] Figures are reproduced from ref. 3.

The liquid drop (figure 1.1b)[10] and oiled coil (figure 1.1c)[11] are similar models; they assume stretching exposes the hydrophobic domains of elastin to the solvent. As a result, the entropy of the system decreases. In the liquid drop model[10], monomers form globules with hydrophilic surfaces and hydrophobic interiors. In the other model[11], the hydrophobic Pro and Val residues are inside of the 'oiled coil', which alternates with α-helical crosslinking domains. Both models were rejected due to high mobility of the hydrophobic regions confirmed by optical spectroscopies[13] and ^{13}C NMR studies.[14]

In the fibrillar model (figure 1.1d), hydrophobic domains form β-spirals that alternate with α-helical crosslinking domains.[12] The chains are additionally held together by hydrophobic interactions. The segments between the β-turns are dynamic and undergo librations, i.e., low-amplitude and high-frequency rocking motions. Stretching of the protein lowers the amplitude of those librations, introducing more order. Therefore, entropy drives the recoil. This model was rejected due to inconsistency with circular dichroism (CD)[15] and molecular dynamics (MD) simulations.[16] The experimental data suggest β-spirals are not found in the structure of elastin[15], whereas computational studies support the hydrophobic (rather than librational) mechanism of elasticity.[16]

Tamburro et. al. developed a conformational ensemble model that assumes an equilibrium between polyproline II helix (PPII), β-turns, and unordered coils (figure 1.2).[17] Each of these conformations has its own, additional internal fluctuations. Stretching shifts the conformational equilibria from folded to extended structures. The recoil is driven by high entropy of the protein in the relaxed state.
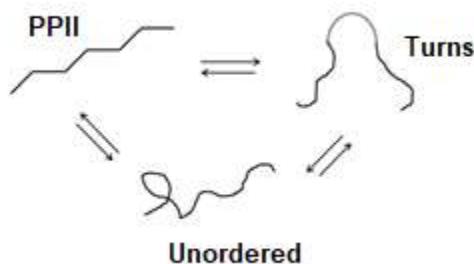
Figure 1.2: Tamburro conformational ensemble model.[17]
Polyproline II helix (PPII), β-turns, and unordered coil structures are in equilibrium. Figure reproduced from Tamburro et. al.

The conformational ensemble model was based on NMR and CD studies of peptides encoded by individual exons of TE (figure 1.3), rather than the complete protein. The organization of elastin has been postulated to have fractal properties, i.e., to have repeating sequences found at different scales in the structure.[18,19] In Tamburro's analysis, each exon corresponds to a self-contained structure that has a distinct function and is autonomous of the other parts of the protein. This premise is supported by coacervation of the domains encoded by exons 18, 20, and 24 (EX18, EX20, EX24) in a similar way to the intact TE.[17]
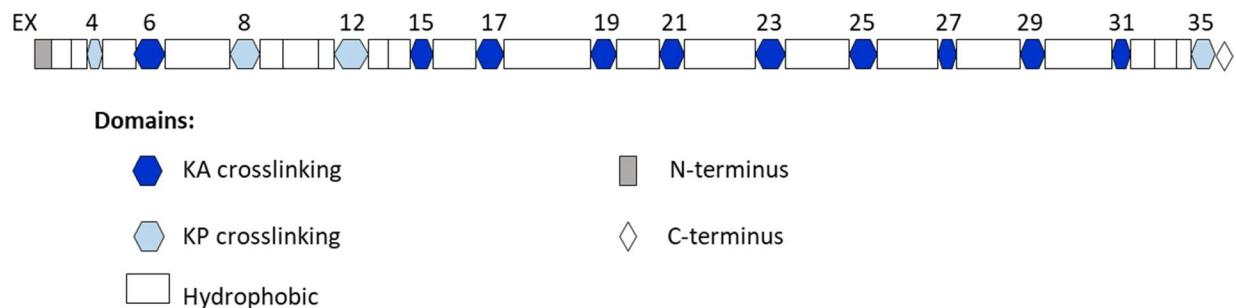
Figure 1.3: Organization of tropoelastin gene based on the cDNA structure.[20]
Figure adapted from Tamburro et. al.[17] Exon 1 is the signaling sequence. Exon 36 is a hydrophilic peptide responsible for interaction with other elastic fiber proteins. Hydrophobic domains are represented with rectangles. Crosslinking domains are indicated with dark blue (alanine-rich) and light blue (proline-rich) hexagons.

1.3 Biosynthesis of elastin

The biosynthesis of elastin starts with the conversion of Lys residues in TE to allysine (Lya), a reactive aldehyde (figure 1.4A).[21] This oxidative deamination of the ε-amino group of Lys is performed by lysyl oxidase (LOX).[22] Lya then reacts further by nonenzymatic condensation to form allysine aldol (AA) via aldol condensation with another Lya[23], or dehydrolysinonorleucine (DLNL) through a Schiff base reaction with Lys.[24] DLNL, similarly to dehydromerodesmosine (DMDES), serves as both an intermediate for more complex structures and an independent crosslink. In the latter case, DLNL and DMDES are eventually reduced to lysinonorleucine (LNL)[25] and merodesmosine (MDES)[26], respectively. Some crosslinks have been reported in the literature less frequently and have not been quantified (figure 1.4B). Allodesmosine (ADES)[27], cyclopentenosine (CPN)[28], and pentasine (PEN)[29] are products of alternative reactions of Lya and Lys residues. Neodesmosine (NDES)[30], in contrast, is a product of the dealkylation of isodesmosine (IDES).
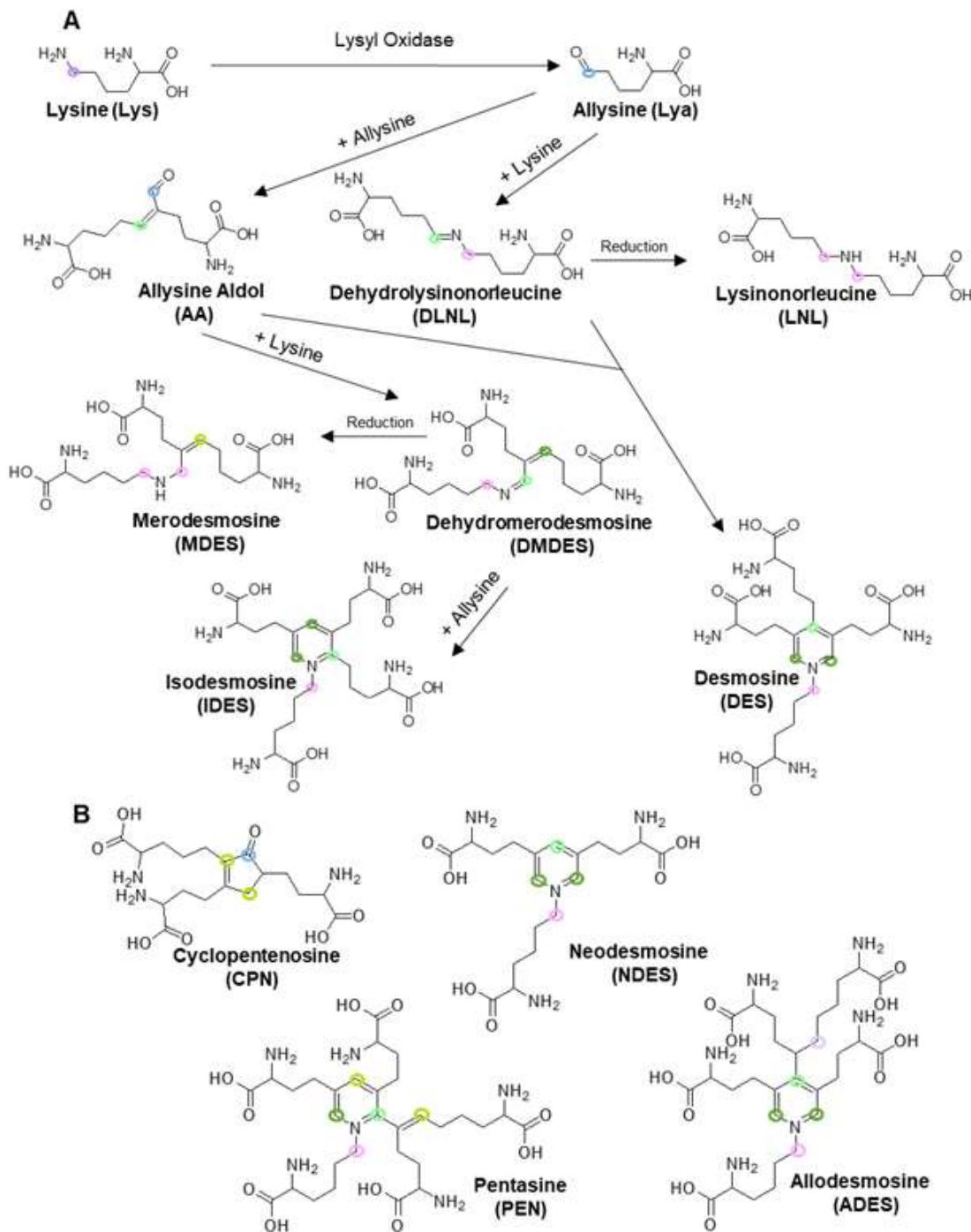
Figure 1.4: Post-translational modification of lysine and the resulting crosslinks.[31]
The synthesis of the most common crosslinks (A) and the less frequently reported species (B) are shown. Colored circles indicate ^{13}C originating from $^{13}C\varepsilon$ in [ε-^{13}C-Lys]: aliphatic (pink, purple), heteroaromatic (green), and carbonyl (blue).

1.4 Crosslinking in health and disease

Elastin's crosslinks play an important role in the health and disease of mammalian connective tissue. Undercrosslinked elastin does not have the properties of elasticity and resilience. Pigs deprived of copper, an essential cofactor to LOX, suffer from impaired crosslink synthesis, resulting in ruptures of major vessels.[32] Similar effects are observed in the genetic disorders of elastin. In Menkes disease, incomplete crosslinking occurs due to compromised absorption of copper. High mortality is observed among patients, most of whom do not live past the age of 3.[33] X-linked cutis laxa, on the other hand, is caused by defects in LOX itself. The prognosis depends on the severity of elastin impairment. If skin is the only affected tissue, life expectancy is normal. However, if the internal organs are impaired and pulmonary damage is present, the prognosis is poor.[34]

Diseases associated with the pathology of elastin are also acquired. Smoking tobacco and air pollution cause chronic obstructive pulmonary disease (COPD). This disorder is associated with lower concentrations of the two crosslinks unique to elastin, desmosine (DES) and isodesmosine (IDES), in the alveolar tissue.[35,36] As a result of elastolytic proteases degrading the elastic fibers, patients with COPD have higher concentrations of these two amino acids in their blood and urine. A sensitive analytical technique allowing for the detection of DES and IDES is needed for diagnostic and prognostic purposes.[37]

Degradation of elastin is also related to aging processes. The rapid synthesis rates in young animals decrease with age, and this protein is characterized by a slow turnover. Both the amount of elastin and the quantity of crosslinks in elastin are lower in tissues collected from older animals.[38,39] In a study of human aortas, the quantity of nearly every kind of crosslink was lower in the samples collected from older subjects. Interestingly, the ratio of NDES to the sum of DES and IDES showed a positive correlation with age after adolescence. The weakening of the elastic fibers is attributed to this relative increase in the NDES content.[40]

1.5 Quantification of crosslinks

Crosslinks have been isolated, characterized, and quantified by multiple research groups.[24,41,42,43] However, these lysine-derivatives have never been observed in native, intact elastin. Since this protein is relatively stable, strong acids and bases, as well as high temperature and pressure, were often used in extraction and purification procedures.[44] Moreover, crosslinks are typically isolated from hydrolysates, further limiting the accuracy of the quantitative studies.[45]

The quantities of crosslinks reported for a single sample of elastin depend on the purification protocol used (table 1.1).[21] Harsh conditions affect the number of crosslinks and yield a more fragmented protein. Thus, the isolated elastin may not resemble the one found *in vivo*.[46] Values for AA are the most reliable, varying only by 5%. Discrepancies for tetrafunctional crosslinks, DES and IDES, are about 15%. LNL has the least reproducible results, differing by over 50%. Furthermore, each of the compared approaches accounts only for 25 – 29 out of the 38 lysine equivalents in BLN elastin.[47]

Table 1.1: Quantities of crosslinks in BLN elastin after different purification protocols.[21]
Quantities are reported in Lys equiv. per 1000 residues.

Compound	Isolation Method			
	Alkali	Autoclaved	Collagenase	Formic Acid
IDES	4.24	4.36	4.24	3.56
DES	5.48	5.20	5.16	4.60
DMDES	2.73	2.37	2.55	2.43
MDES	0.66	0.78	0.66	0.69
DLNL	0.20	0.14	0.30	0.50
LNL	2.28	1.66	1.38	1.22
AA	5.84	5.52	5.80	5.86
Lys (unmodified)	6.10	7.00	8.50	5.90
Lys equivalents (sum)	27.53	27.03	28.59	24.76

1.6 Crosslinking sites

High-resolution mass spectrometry (MS) data provide information on positions and patterns of crosslinking in elastin (figure 1.5).[43] Both KP and KA domains contain some unmodified Lys. KP domains are almost exclusively crosslinked by LNL, whereas KA domains are linked by LNL, AA, DES, and IDES. About half of the bifunctional crosslinks do not undergo further condensation due to steric hindrance. Similarly, Lys residues followed by bulky amino acids (leucine, isoleucine, phenylalanine, or tyrosine) remain unmodified, as LOX cannot access them. A large number of intramolecular crosslinks between Lys residues in close proximity was found. Overall, elastin is heterogeneously crosslinked; i.e., two domains are not always connected with each other. Similarly, a Lys residue within the TE sequence may be simultaneously found unmodified and in different types of crosslinks.
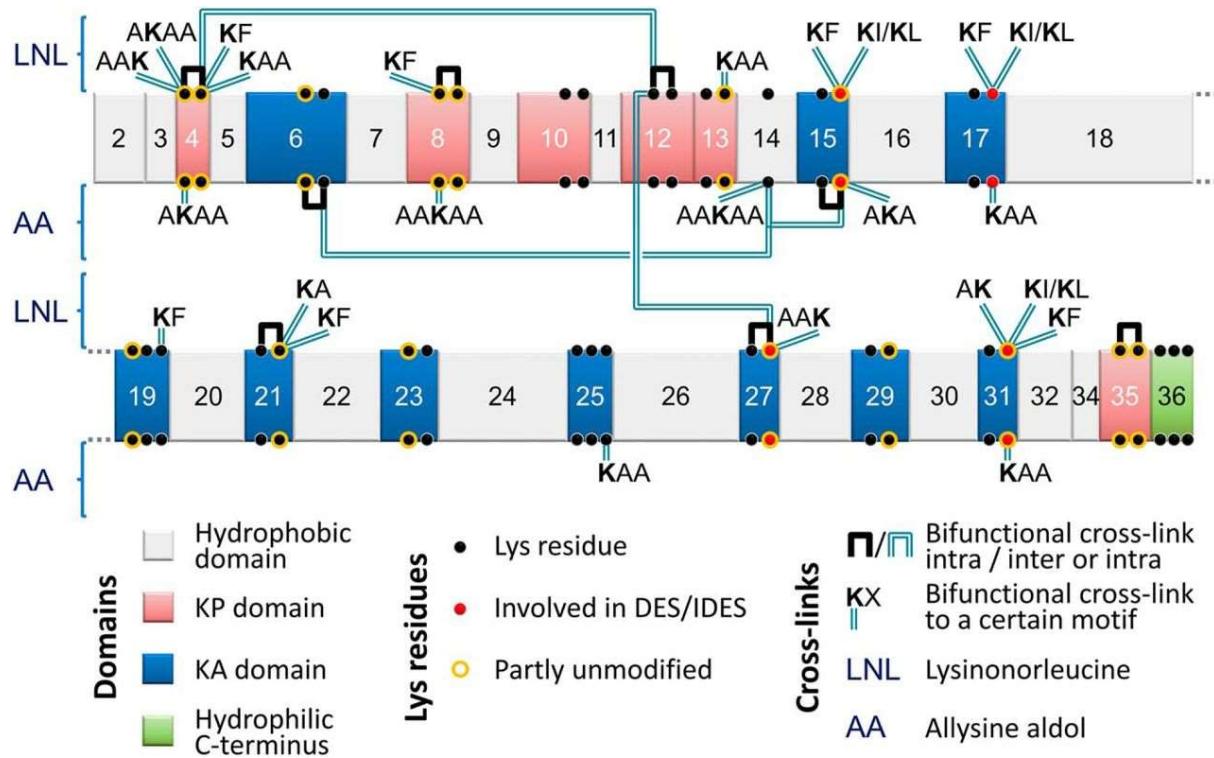
Figure 1.5: Crosslinking sites in elastin.
Figure reproduced from Schmeltzer et. al.[43] Gray boxes represent hydrophobic domains. Hydrophilic domains with KP and KA motifs are shown in red and blue, respectively. The width of the boxes corresponds to the relative length of the domains. Numbering corresponds to exon assignment. For clarity, all links are shown as intramolecular crosslinks.

The high-resolution MS data provide important information about the organization of elastin, but this approach has limitations. For example, no residues of highly reactive Lya were detected. Elastin was enzymatically cleaved, and soluble peptides were studied rather than the intact protein. To understand crosslinking in native elastin as well as secondary structures and dynamics, a complementary analytical technique is needed.

1.7 Solid-state NMR

NMR is a powerful method that has been widely used to characterize the structure and dynamics of biomolecules. Only soluble samples that are available at sufficient concentrations can benefit from solution NMR studies. Solid-state NMR (ssNMR) does not have as many limitations. Because crystalline structure is not required, this technique allows for studies of amorphous proteins that are inaccessible to X-ray diffraction.

Biological samples are commonly characterized with NMR measurements of spin-1/2 nuclei, such as ^{1}H and ^{13}C. ^{1}H nuclei offer high natural abundance and sensitivity (table 1.2). However, only relatively small molecules allow for sufficient resolution in the 0 – 12 ppm chemical shift range of ^{1}H. The spectral range of ^{13}C is greater (0 – 200 ppm), and the ^{13}C chemical shift can identify secondary structures in the protein backbone.[48] However, sensitivity is only a quarter of that of ^{1}H, and the natural abundance of ^{13}C is very low (1.1%). Selective isotopic enrichment is often necessary to observe the targeted signals. The ^{15}N nucleus is even more problematic, as very low natural abundance is accompanied by poor sensitivity. Even with isotopic enrichment, detection of rare residues may not be possible.

Table 1.2: Natural abundance levels and relative sensitivities based on gyromagnetic ratios (γ/γ_H) of common nuclei.[49]

Nucleus	Natural Abundance (%)	Relative Sensitivity
^{1}H	100.00	100
^{13}C	1.10	25
^{15}N	0.37	10

SsNMR spectra are characterized by broad signals resulting from chemical shift anisotropies (CSA) and dipolar couplings. The external magnetic field, **B$_0$**, affects the electrons surrounding the nucleus, which produce small magnetic fields. These fields either add or subtract from the magnetic field experienced by the nucleus. Electron clouds are rarely symmetrically distributed. Therefore, the effects of the electron density on the resonance frequency are orientation-dependent. Similarly, the interactions of nuclear magnetic moments of different spins, i.e., heteronuclear dipolar couplings, also depend on orientation. In solution, molecular tumbling averages the effects of CSA and dipolar couplings, but such rapid motions are not present in solids. **M**agic **A**ngle **S**pinning (MAS) is applied to reduce the orientation-dependent interactions (figure 1.6).
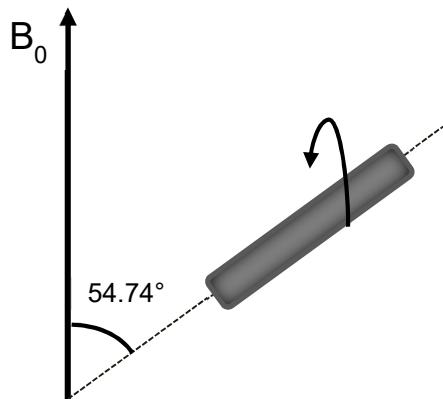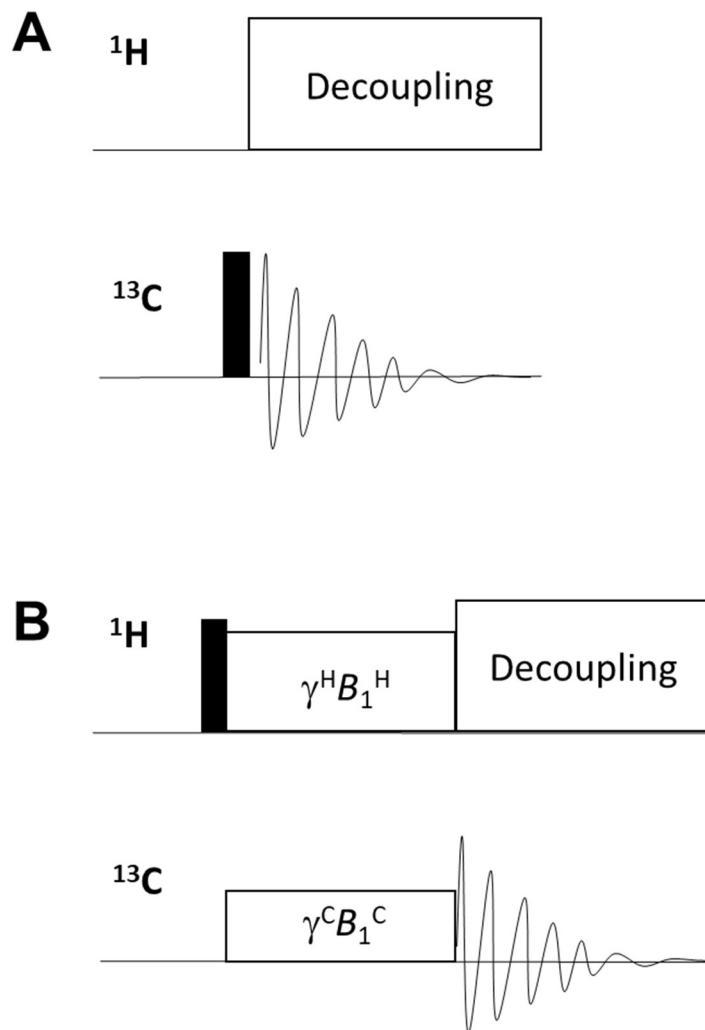
Figure 1.6: Spinning a rotor with a sample at a magic angle.

MAS refers to the spinning of the rotor containing a sample at 54.74° relative to the external magnetic field B_0. The mechanical spinning results in averaging of the internuclear interactions, reducing the effects of CSA and dipolar couplings. If the spinning rate is smaller than the magnitude of anisotropy, spinning sidebands are observed at the multiples of the spinning speed. Spinning sidebands may introduce ambiguity in assignments, especially in spectra of large molecules.

Direct **P**olarization (DP) is an excitation scheme traditionally used for solution ^1H NMR of organic compounds. However, ^{13}C DP has been used to characterize elastin.[50,51] This excitation scheme provides quantitative NMR spectra, i.e., with peak areas proportional to the number of ^{13}C spins. In this single-pulse measurement, a 90° pulse is applied on the ^{13}C channel. An acquisition period follows with high-power ^1H decoupling, which removes the broadening caused by the heteronuclear dipolar couplings (figure 1.7A).

Figure 1.7: Pulse sequences for (A) ^{13}C DP and (B) ^{13}C CP experiments.
Black rectangle represents a 90° pulse.

The main drawback of the DP measurement is the long experimental time needed for acceptable signal-to-noise ratios (S/N) in the spectra of biological solids. The recycle delay, or time needed for recovery of the equilibrium magnetization after each scan, depends on the spin-lattice relaxation time constant ($T_1{}^C$), which tends to have large values in biological solids. To obtain >99% recovery, the recycle delay needs to be $5 \times T_1^C$.

Cross Polarization Magic Angle Spinning (CPMAS) is a common ssNMR experiment that has better S/N than DP in a given time period.[52] In this measurement, a 90° ^{1}H pulse rotates the magnetization to the transverse plane, which is then spin-locked using continuous radio-frequency (rf) irradiation, B_1^H (figure 1.7B). At the same time, an rf field, B_1^C, is applied on the ^{13}C channel, such that the Hartmann-Hahn (HH) condition[53] is met:

$$\frac{\gamma^C B_1^C}{2\pi} = \frac{\gamma^H B_1^H}{2\pi} + n\nu_r \qquad \text{(Equation 1.1)}$$

where ν_r is a spinning rate of the rotor, and n = 0, ± 1, ± 2.

The dipolar-coupling-mediated transfer of magnetization from ^{1}H to ^{13}C spins follows the ^{1}H 90° pulse. This process results in an enhancement of signal intensity (compared to DP), which is proportional to the differences in the gyromagnetic ratios $\left(\frac{\gamma^H}{\gamma^C} = 4\right)$.[49] Since the carbon excitation arises entirely from the contact with protons, the T_1^H constant determines the delay between successive scans rather than T_1^C. ^{1}H T_1 relaxation times are shorter than those of ^{13}C, decreasing the experimental time.

If spinning sidebands are observed in a CP/MAS spectrum, Total Sideband Suppression (TOSS) sequence can declutter a complex spectrum without altering the spinning rate.[54] Without TOSS, the amplitudes and phases of the sideband components are dispersed, due to different orientations of spins. After the four 180° ^{13}C pulses (figure 1.8) of TOSS, the phases become evenly distributed in the complex plane, i.e., cancelled when averaged.[55] The periodic variation at the spinning frequency is eliminated, resulting in spectra without spinning sidebands.
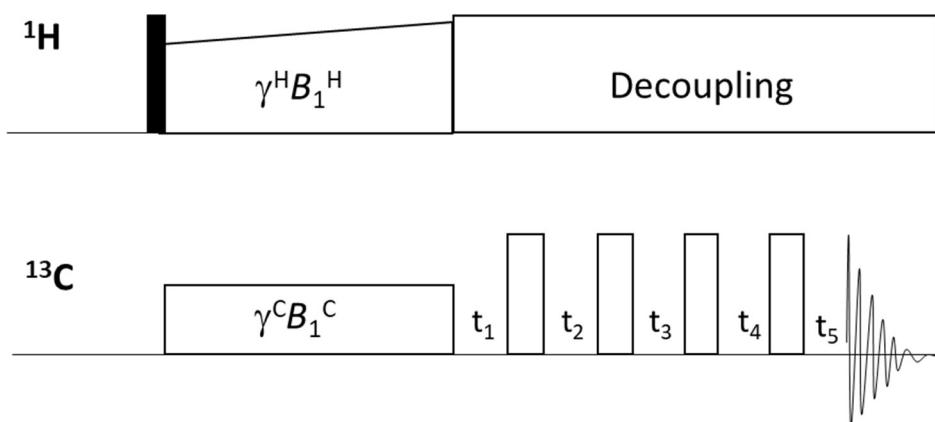
Figure 1.8: Pulse sequence for ^{13}C CP TOSS experiment.[54]
Black and white rectangles represent 90° and 180° pulses, respectively. TOSS delays, t_1, t_2, t_3, t_4, and t_5, separate ^{13}C 180° pulses.

1.8 Scope of thesis

This thesis presents studies of lysine and lysine-derived crosslinks in mature, intact elastin via ssNMR spectroscopy. Two isotopes, [U-^{13}C,^{15}N-Lys] and [^{13}Cε-Lys], were independently incorporated into the neonatal rat smooth muscle cell (NRSMC) elastin. Secondary structures are assigned, based on one- and two- dimensional experiments. The dynamics was probed by spin-lattice relaxation measurements. Modified cross-polarization excitation schemes allowed for detection and characterization of crosslinks in the intact protein. Quantities of unmodified lysine as well as desmosine and isodesmosine were determined, and the rare bi- and trifunctional crosslinks were detected.

1.9 References

[1] Davis, E. Stability of elastin in the developing mouse aorta: a quantitative radioautographic study. *Histochem.* **1993**, *100*, 17-26.

[2] Shapiro, S. D.; Endicott, S. K.; Province, M. A.; Pierce, J. A.; Campbell, E. J. Marked longevity of human lung parenchymal elastic fibers deduced from prevalence of D-aspartate and nuclear weapons-related radiocarbon. *J. of Clin. Invest.* **1991**, *87*, 1828-1834.

[3] Vrhovski, B.; Weiss, A. S. Biochemistry of tropoelastin. *Eur. J. Biochem.* **1998**, *258*, 1-18.

[4] Hoeve, C. A. J.; Flory, P. J. The elastic properties of elastin. *Biopolymers* **1974**, *13*, 677-686.

[5] Aaron, B. B.; Gosline, J. M. Optical properties of single elastin fibres indicate random protein conformation. *Nature* **1980**, *287*, 865-867.

[6] Dorrington, K.; Grut, W.; Mccrum, N. G. Mechanical state of elastin. *Nature* **1975**, *255*, 476-478.

[7] Gotte, L.; Mammi, M.; Pezzin, G. Some structural aspects of elastin revealed by X-ray diffraction and other physical methods. In Symposium on Fibrous Proteins. Butterworths, Sydney, Australia. 236-245.

[8] Cox, B. A.; Starcher, B. C.; Urry, D. W. Coacervation of tropoelastin results in fiber formation. *J. Biol. Chem*. **1974**, *249*, (3), 997-998.

[9] Kozel, B. A.; Mecham, R. P.; Rosenbloom, J. Elastin. In The extracellular matrix: an overview. R.P. Mecham; New York, 2011; 267-303.

[10] Weis-Fogh, T.; Anderson, S. O. New molecular model for the long-range elasticity of elastin. *Nature* **1970**, *227*, 718-721.

[11] Gray, W. R.; Sandberg, L. B.; Foster, J. A. Molecular model for elastin structure and function. *Nature* **1973**, *246*, 461-466.

[12] Venkatachalam, C. M.; Urry, D. W. Development of a linear helical conformation from its cyclic correlate. β-Spiral model of the elastin poly(pentapeptide) (VPGVG)n. *Macromolecules* **1981**, *14*, 1225-1229.

[13] Debelle, L.; Alix, A. J. P.; Jacob, M.; Huvenne, J.; Berjot, M.; Sombret, B.; Legrand, P. Bovine Elastin and κ-Elastin Secondary Structure Determination by Optical Spectroscopies. *J. Biol. Chem.* **1995**, *270*, 26099-26103.

[14] Torchia, D. A.; Piez, K. A., Mobility of elastin chains as determined by ^{13}C nuclear magnetic resonance. *J. Mol. Biol.* **1973**, *76*, 419-424.

[15] Gross, P. C.; Possart, W.; Zeppezauer, M. An alternative structure model for the polypentapeptide in elastin. *Z. Naturforsch. C* **2003**, *58*, 873-878.

[16] Li, B.; Alonso, D. O. V.; Bennion, B. J.; Daggett, V. Hydrophobic hydration is an important source of elasticity in elastin-based biopolymers. *J. Am. Chem. Soc.* **2001**, *123*, 11991-11998.

[17] Tamburro, A. M.; Bochicchio, B.; Pepe, A. Dissection of human tropoelastin: Exon-by-exon chemical synthesis and related conformational studies. *Biochemistry* **2003**, *42*(45), 13347-13362.

[18] Tamburro, A. M.; De Stradis, A.; D'Alessio, L. Fractal aspects of Elastin supramolecular organization. *J. Biomol. Struct. Dyn.* **1995**, *12*, 1161-1172.

[19] Tamburro, A. M.; Daga-Gordini, D.; Guantieri V.; De Stradis A. On the molecular and the supramolecular structure of elastin. In *Chemistry and Properties of Biomolecular Systems*; Russo, N., Anastassopoulou, S., Barone, G., Eds.; Topics in Molecular Organization and Engineering; Springer: Dordrecht, 1994; vol. 11, pp 389−403.

[20] Pierce, R. A.; Deak, S. B.; Stolle, C. A.; Boyd, C. D. Heterogeneity of rat tropoelastin mRNA revealed by cDNA cloning. *Biochemistry* **1990**, *29*, 9677-9683.

[21] Francis, G.; John, R.; Thomas, J. Biosynthetic pathway of desmosines in elastin. *Biochem. J.* **1973**, *136*, 45-55.

[22] Pinnell, S. R.; Martin, G. R. The Cross-Linking of Collagen and Elastin: Enzymatic Conversion of Lysine in Peptide Linkage to α -Aminoadipic-δ -Semialdehyde (Allysine) by an Extract from Bone. *PNAS* **1968**, *61*, 708-716.

[23] Lent, R. W.; Smith, B.; Salcedo, L. L.; Faris, B.; Franzblau, C. Reduction of elastin. II. Evidence for the presence of α-aminoadipic acid delta.-semialdehyde and its aldol condensation product. *Biochemistry* **1969**, *8*, 2837-2845.

24 Franzblau, C.; Sinex, F. M.; Faris, B.; Lampidis, R. Identification of a new crosslinking amino acid in elastin. *Biochem. Biophys. Res. Commun.* **1965**, *21*, 575-581.

25 Davril, M.; Han, K. Purification of lysinonorleucine cross-linked peptide fraction from porcine aorta elastin. *Int. J. Pept. Protein Res*. **1976**, *8*(2), 177-181.

26 Paz, M. A.; Henson, E.; Blumenfeld, O. O.; Seifter, S.; Gallop, P. M. Dehydromerodesmosine and merodesmosine in elastin. *Biochem. Biophys. Res. Commun.* **1971**, *44*, 1518-1523.

27 Suyama, K.; Nakamura, F. Isolation and characterization of new cross-linking amino acid, allodesmosine, from the acid hydrolysate of elastin. *Biochem. Biophys. Res. Commun*. **1990**, *170*(2), 713-718.

28 Akagawa, M.; Yamazaki, K.; Suyama, K. Cyclopentenosine, Major Trifunctional Crosslinking Amino Acid Isolated from Acid Hydrolysate of Elastin. *Arch. Biochem. Biophys.* **1999**, *372*(1), 112-120.

29 Starcher, B.; Cook, G.; Gallop, P.; Hensen, E.; Shoulders, B. Isolation and Characterization of a Pentameric Amino Acid from Elastin. *Connect. Tissue Res*. **1987**, *16*(1), 15-25.

30 Watanabe, M.; Sawai, T.; Nagura, H.; Suyama, K. Age-related alteration of cross-linking amino acids of elastin in human aorta. *Tohoku J. of Exp. Med.* **1996**, *180*, 115-130.

31 Heinz, A.; Ruttkies, C. K. H.; Jahreis, G.; Schräder, C. U.; Wichapong, K.; Sippl, W.; Keeley, F. W.; Neubert, R. H. H.; Schmelzer, C. E. H. In vitro cross-linking of elastin peptides and molecular characterization of the resultant biomaterials. *Biochim. Biophys. Acta* **2013**, *1830*, 2994-3004.

32 Shields, G. S.; Coulson, W. F.; Kimball, D. A.; Carnes, W. H.; Cartwright, G. E.; Wintrobe, M. M. Studies on copper metabolism. 32. Cardiovascular lesions in copper-deficient swine. *Am. J. Pathol.* **1962**, *41*, 603-621.

33 Tumer, Z.; Moller, L.B. Menkes disease. *Eur. J. Hum. Genet*. **2010**, *18*, 511-518.

34 Gara S,; Litaiem N. Cutis Laxa (Elastolysis) [Updated 2018 Dec 2]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2018 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK532944/

35 Eyre, D. R.; Paz, M. A.; Gallop, P. M. Cross-linking in collagen and elastin. *Am. J. Clin. Nutr*. **1978**, *53*, 717-748.

36 Deslee, G.; Woods, J. C.; Moore, C. M.; Liu, L.; Conradi, S. H.; Milne, M.; Gierada, D. S.; Pierce, J.; Patterson, A.; Lewit, R. A.; Battaile, J. T.; Holtzman, M. J.; Hogg, J. C.; Pierce, R. A. Elastin expression in very severe human COPD. *Eur. Respir. J.* **2009**, *34*, 324-331.

37 Annovazzi, L.; Viglio, S.; Perani, E.; Luisetti, M.; Baraniuk, J.; Casado, B.; Cetta, G.; Iadarola, P.; El Rassi, Z. Capillary electrophoresis with laser-induced fluorescence detection as a novel sensitive approach for the analysis of desmosines in real samples. *Electrophoresis* **2004**, *25*, 683-691.

38 John, R.; Thomas, J. Chemical compositions of elastins isolated from aortas and pulmonary tissues of humans of different ages. *Biochem. J.* **1972**, *127*(1), 261-269.

39 Fujimoto, D. Aging and cross-linking in human aorta. *Biochem. Biophys. Res. Commun.* **1982**, *109*(4), 1264-1269.

40 Watanabe, M.; Sawai, T.; Nagura, H.; Suyama, K. Age-related alteration of cross-linking amino acids of elastin in human aorta. *TJEM* **1996**, *180*(2), 115-30.

41 Miller, E. J.; Martin, G. R.; Piaz, K. A. The utilization of lysine in the biosynthesis of elastin crosslinks. *Biochem. Biophys. Res. Commun.* **1964**, *17*, 248-253.

42 Snider, R.; Faris, B.; Verbitzki, V.; Moscaritolo, R.; Salcedo, L. L.; Franzblau, C. Elastin biosynthesis and cross-link formation in rabbit aortic smooth muscle cell cultures. *Biochemistry* **1981**, *20*, 2614-2618.

43 Schräder, C.; Heinz, A.; Majovsky, P.; Karaman Mayack, B.; Brinckmann, J.; Sippl, W.; Schmelzer, C. Elastin is heterogeneously cross-linked. *J. Biol. Chem.* **2018**, *293*(39), 15107-15119.

44 John, R.; Thomas, J. Localization and chemical composition of elastin in ox lung. *Intl. J. Biochem.* **1971**, 529-536.

45 Guay, M.; Lamy, F. The troublesome crosslinks of elastin. Trends Biochem. Sci. **1979**, *4*, 160-164.

46 Soskel, N. T.; Sandburg, L. B. A comparison of six methods of extracting elastin residue from hamster lungs. *Exp. Lung Res.* **1983**, *4*(2), 109-119.

47 Raju K.; Anwar R. A. Primary structures of bovine elastin a, b, and c deduced from the sequences of cDNA clones. *J. Biol. Chem.* **1987**, *262*(12), 5755-5762.

[48] Wishart, D. S.; Sykes, B. D.; Richards, F. M. Relationship Between Nuclear-Magnetic-Resonance Chemical-Shift and Protein Secondary Structure. *J. Mol. Biol.* **1991**, *222*, 311-333.

[49] Harris, R*. Nuclear Magnetic Resonance Spectroscopy*; Longman Scientific and Technical: Essex, 1986.

[50] Perry, A.; Stypa, M. P.; Foster, J. A.; Kumashiro, K. K. Observation of the glycine in elastin using ^{13}C and ^{15}N solid-state NMR spectroscopy and isotopic labeling. *J. Am. Chem. Soc.* **2002**, *124*, 6832-6833.

[51] Kumashiro, K. K. *Solid-State NMR Studies of Elastin and Elastin Peptides*. Dordrecht: Springer Netherlands: Dordrecht, 2006; p 93-99.

[52] Pines, A.; Gibby, M. G.; Waugh, J. S. Proton-enhanced NMR of dilute spins in solids. *J. Chem. Phys.* **1973**, *59*, 569-590.

[53] Hartmann, S. R.; Hahn, E. L. Nuclear double resonance in the rotating frame. *Phys. Rev.* **1962**, *128*, (5), 2042-2053.

[54] Dixon, W. T.; Schaefer, J.; Sefcik, M. D.; Stejskal, E. O.; McKay, R. A. Total suppression of sidebands in CPMAS C-13 NMR. *J. Magn. Reson.* **1982**, *49*, 341-345.

[55] Antzutkin, O. N.; Song, Z.; Feng, X.; Levitt, M. H. Suppression of sidebands in magic-angle-spinning nuclear magnetic resonance: General principles and analytical solutions. *J. Chem. Phys.* **1994**, *100*, 130-140.

CHAPTER 2

ISOTOPIC ENRICHMENT OF LYSINES IN ELASTIN

2.1 Isotopic enrichment via the neonatal rat smooth muscle cell (NRSMC) line

Selective labeling of amino acid increases the intensity of targeted signals. The NMR-active ^{13}C nuclei has low natural abundance, constituting only 1.1% of carbons in unenriched samples.[1] Moreover, lysine is a minor component of elastin. Without isotopic enrichment, the sensitivity and resolution are insufficient for NMR studies. Signals of rare crosslinks are not observed at all, and unmodified lysines are indistinguishable from the rest of the protein.

The ideal elastin expression system achieves high isotopic enrichment, allows for easy manipulation of the conditions, and does not raise ethical concerns. In the early days of the elastin research, ^{13}C nuclei were introduced to samples by incubation of tissues in labeled media.[2,3] The produced elastin carried the label in only 20% of its lysine, 10% of valine, and 6% of alanine residues. This approach required large quantities of the expensive isotope and substantial sacrifice of vertebrates. Thus, modern NMR studies rely on expression through cell cultures.

Labeled elastin-like peptides[4,5] and a recombinant human tropoelastin[6] have been prepared with bacterial hosts, but this method is not well-suited for studies of lysine and crosslinks. Lysine is an essential amino acid in animals, but it is synthesized *de novo* in bacteria.[7] Direct substitution, i.e., supplementing the cell culture media with labeled lysine, would not result in high enrichment of elastin in bacterial hosts. Unenriched lysine is still synthesized by *E. coli*, diluting the labeled residues. Moreover, prokaryotes lack the enzymes for the post-translational modification of tropoelastin. Therefore, treatment of the bacterial expression product with chemical crosslinkers would be needed, and the resulting crosslinks would differ from those in native elastin.[6,8]

The neonatal rat smooth muscle cell (NRSMC) line allows for production of large quantities of labeled, native-like elastin in a short amount of time. Mammalian expression systems are equipped with the lysyl oxidases needed for crosslinking. The protein obtained from NRSMC culture has the same profile as the one extracted from a rat aorta *in vivo*, as confirmed by electron microscopy and amino acid analysis.[9] Sufficient quantities for ssNMR spectroscopy are obtained within 6 to 8 weeks.[10]

High incorporation of labeled glycine, alanine, proline, and valine into elastin have been achieved with the NRSMC line.[10,11,12] For essential amino acids, direct substitution results in enrichment above 80%. If a nonessential amino acid is labeled, this well-understood system allows for manipulation of the conditions. For example, improvement of incorporation levels has been achieved by inhibition of in-cell synthesis of alanine.[12] The NRSMC line is a well-established protocol, suitable for ssNMR studies of lysine and crosslinks in elastin.

2.2 Targeting lysine and crosslinks

This thesis explores two labeling schemes of elastin. [U-^{13}C,^{15}N-Lys] elastin is grown in media with uniformly-^{13}C- and -^{15}N-enriched lysine (figure 2.1A). Separately, the medium is supplemented with lysine that has a single ^{13}C label at the epsilon (ε) position to obtain [ε-^{13}C-Lys] elastin (figure 2.1B). The NMR-active nuclei are also incorporated into the numerous crosslinks that are derived from this amino acid during post-translational modification.
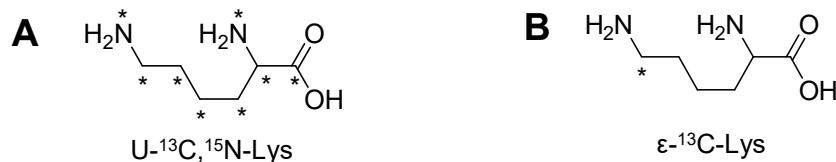
Figure 2.1: Structures of enriched lysine.
U-^{13}C,^{15}N-Lys (A) and ε-^{13}C-Lys (B) are incorporated into elastin. Asterisk (*) indicates a labeled nucleus.

Different types of covalent crosslinks are present in mature elastin (figure 2.2). During elastogenesis, the lysyl oxidase (LOX) enzyme converts a lysine residue in tropoelastin to a reactive allysine (Lya), which further reacts to form bi-, tri-, and tetrafunctional crosslinks.[13] The most abundant modification products are desmosine (DES) and isodesmosine (IDES).[14] Allysine aldol (AA) and lysinonorleucine (LNL) are also commonly reported, but unlike DES and IDES, they are not unique to elastin.[15] The details of the biosynthetic pathways are discussed in Chapter 1.
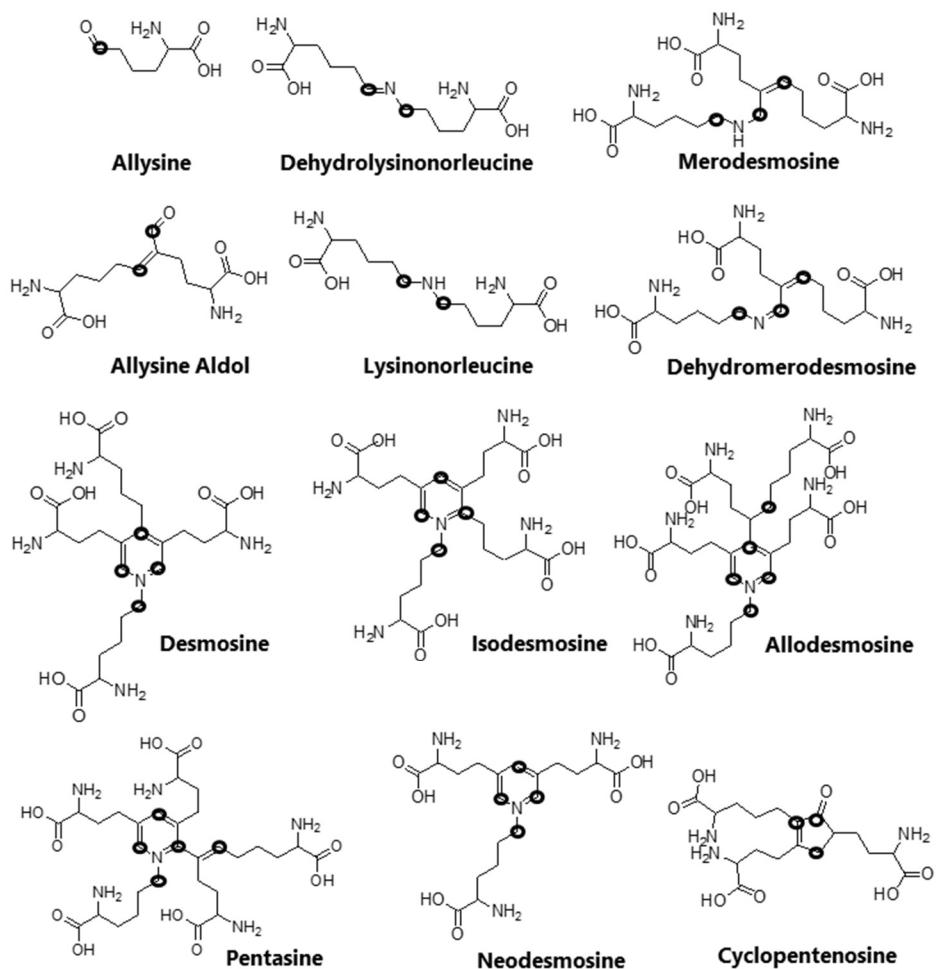
Figure 2.2. Structures of the lysine-derived crosslinks in elastin.
NRSMC cultures grown in media with U-^{13}C,^{15}N-Lys have ^{13}C and ^{15}N labels at all carbon and nitrogen sites of the crosslinks. Circles show locations of ^{13}C in [ε-^{13}C-Lys] elastin.

[U-^{13}C,^{15}N-Lys] elastin has the NMR-active isotope at all carbon and nitrogen atoms in the pictured crosslinks (figure 2.2). Thus, they have correspondingly higher molecular weights (MW) than their counterparts in the unenriched protein. For example, enriched Lya, the smallest of the modification products, is 8 g/mol heavier than its unlabeled equivalent. The difference in MW of labeled and unlabeled tetrafunctional crosslinks, DES and IDES, is 29 g/mol.

The mechanisms of crosslink formation determine the label-bearing sites in [ε-^{13}C-Lys] elastin. With the exception of neodesmosine (NDES), the number of lysine equivalents corresponding to each crosslinking moiety is the same as the increase of the molecular weight. Therefore, the difference in the molecular weight of even the largest pentafunctional species, such as the rare pentasine (PEN) or allodesmosine (ADES), does not exceed 5 g/mol.

Neodesmosine appears to be a trifunctional crosslink, which would correspond to three lysine equivalents. However, in the [ε-^{13}C-Lys] elastin sample, the expected increase in the molecular weight is 4 g/mol rather than 3 g/mol. Unlike other crosslinks, NDES is formed during the decomposition of DES and IDES, rather than synthesis from three lysine equivalents (figure 2.3). The bond between Cδ and Cε in DES or IDES is broken, and the Cε remains in the pyridinium ring. The exact details of the reaction mechanism have not been elucidated, but this process is associated with aging. Studies of human aortas showed a gradual increase of the $\frac{\text{NDES}}{\text{DES+IDES}}$ ratio with age.[16]
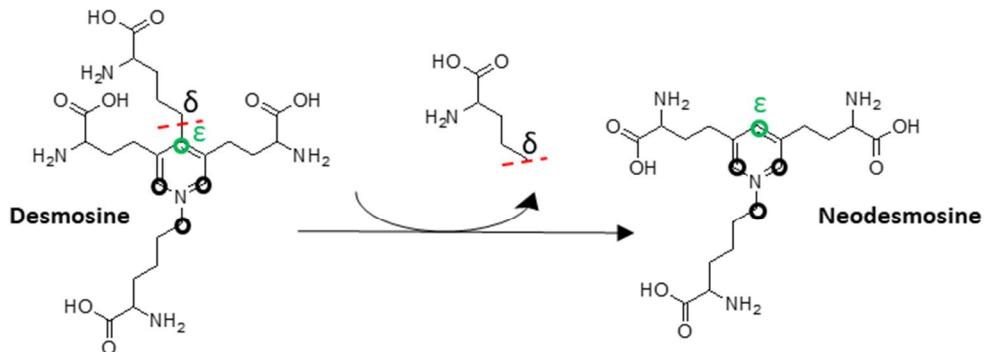
Figure 2.3. Formation of neodesmosine from partial decomposition of desmosine.
Circles indicate ^{13}Cε in [ε-^{13}C-Lys] elastin. The green circle is a remnant of the fourth lysine equivalent that was present in desmosine.

2.3 Metabolism of enriched lysine

Incorporation of an enriched, essential amino acid, such as lysine, into a protein is straightforward. Since the targeted residue cannot be synthesized within the cells from other molecules, direct supplementation is sufficient, without other manipulation of the metabolic pathways. The desired isotope is added to lysine-free cell culture media, serving as the only source of this amino acid. As the labeled lysines undergo post-translational modification, the NMR-active nuclei are incorporated into the resulting crosslinks.

For the targeted enrichment to be effective, it is crucial that the isotopic labels are introduced only at the desired sites. If an enriched residue is a precursor for synthesis of other amino acids, the NMR-active nuclei are passed to other locations in the protein. This process is known as *scrambling*. The possible destinations of the labels in enriched lysine are predicted, based on the metabolic pathways.

In mammalian cells, lysine is catabolized through saccharopine or pipecolic acid intermediate (figure 2.4).[17] These two routes converge into a common degradative pathway once the allysine intermediate (Lya) is formed. In the steps that follow, carbon dioxide (originating from CO- and Cε-Lys) and two molecules of acetyl CoA (one from Cα-, Cβ-Lys and one from Cδ-, Cγ-Lys) are produced. Therefore, catabolism of both [ε-^{13}C-Lys] and [U-^{13}C,^{15}N-Lys] produces $^{13}CO_2$, but only the uniformly labeled lysine leads to enriched acetyl CoA.
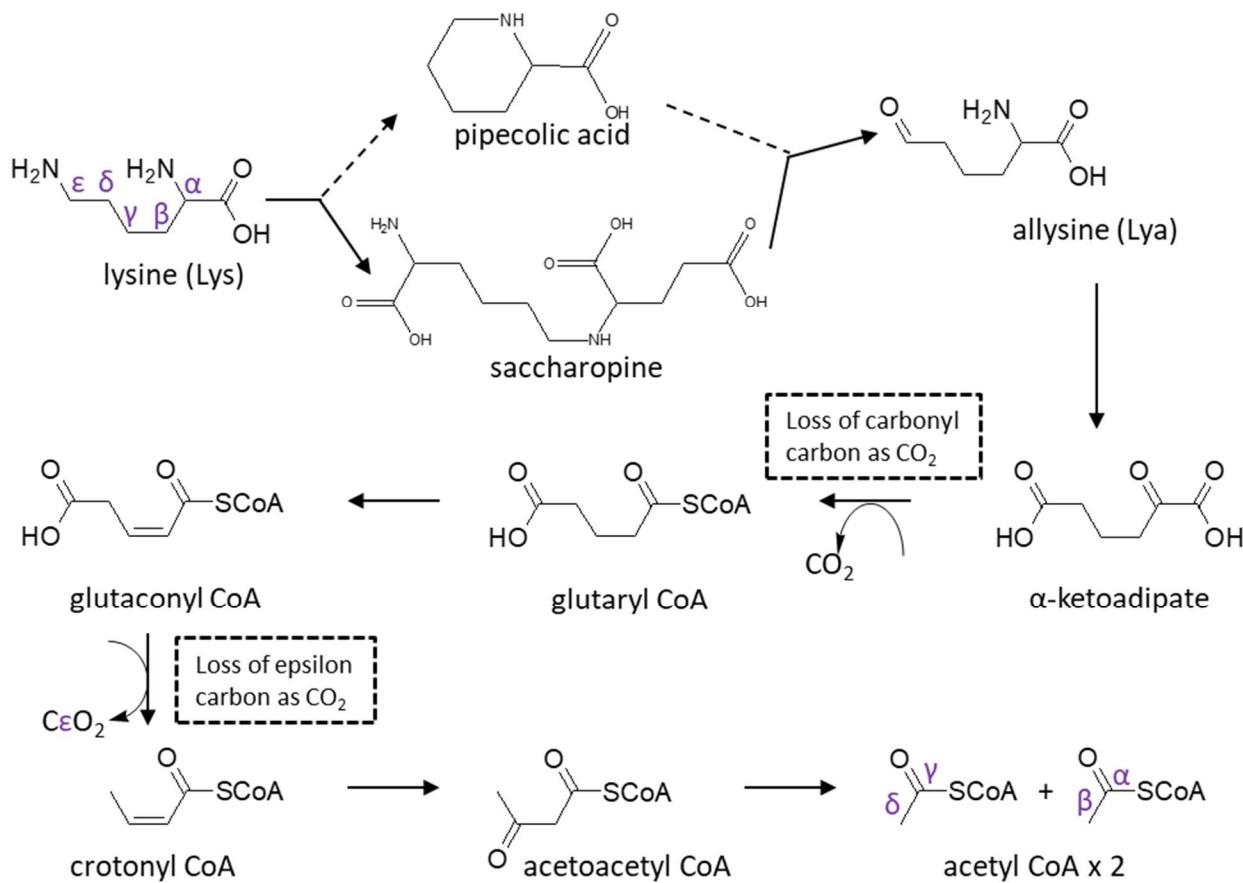
Figure 2.4: Catabolic pathway of lysine.[18]
Structures of the most significant intermediates are shown. Allysine is formed via pipecolic acid pathway (dashed arrows) or saccharopine pathway (solid line). Once it is produced, there is a single pathway of decomposition. CO and Cε are lost as CO_2 molecules. The resulting two molecules of acetyl CoA are derived from Cα, Cβ and Cδ, Cγ, respectively.

Acetyl CoA enters the Krebs cycle and serves as a precursor to three amino acids present in elastin: glutamine (Gln), aspartate (Asp), and arginine (Arg) (figure 2.5). The tropoelastin (TE) sequence only contains trace amounts of these residues (< 1.1%). However, if scrambling is substantial, it could still be observed in the NMR data. For comparison, lysine constitutes 4.5% of the TE monomer before crosslinking, and ~1% of the monomer remains unmodified in mature elastin.
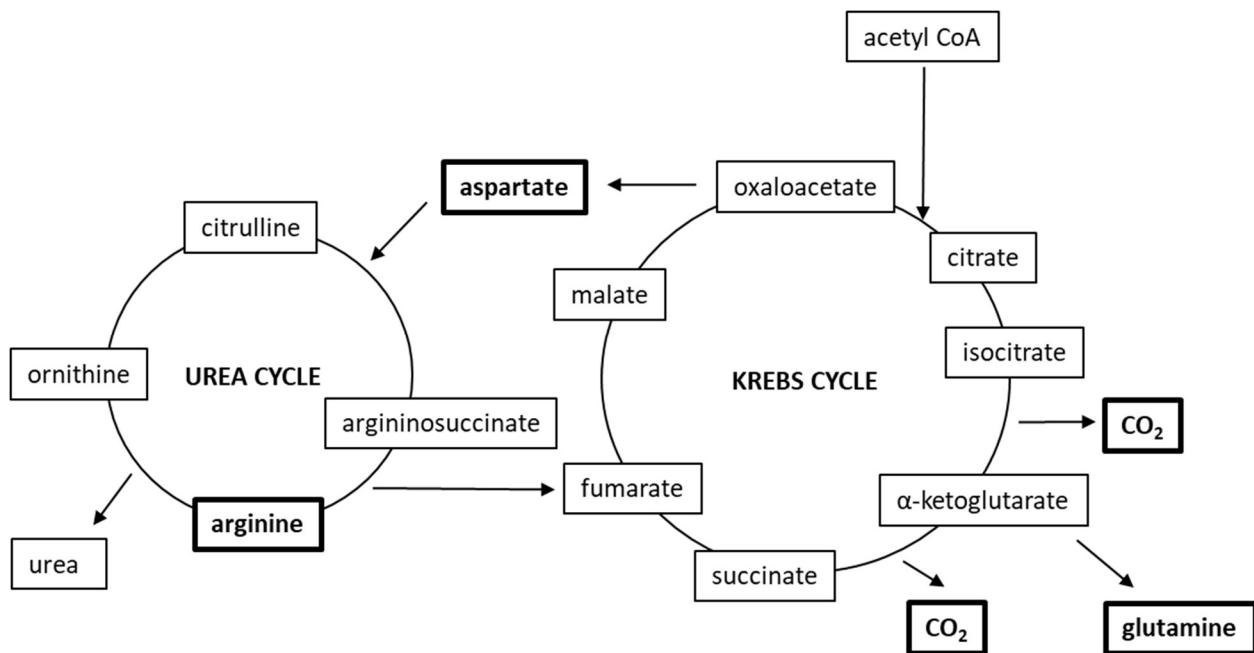
Figure 2.5: Scrambling pathways in the Krebs and urea cycles.[4]
Aspartate, glutamate, arginine, and CO_2 are emphasized with bold font and frames, as they could lead to scrambling in elastin. Carbon dioxide is a precursor to glycine.

Carbon dioxide is produced in several steps of lysine catabolism and leads to scrambling to glycine (Gly). Both [ε-^{13}C-Lys] and [U-^{13}C,^{15}N-Lys] release $^{13}CO_2$ originating from ^{13}Cε when glutaconyl CoA is degraded to crotonyl CoA (figure 2.4). In addition, the breakdown of [U-^{13}C,^{15}N-Lys] produces three more molecules of $^{13}CO_2$; one from lysine carbonyl, released when α-ketoadipate is converted to glutaryl CoA (figure 2.4), and two more in the Krebs cycle (figure 2.5). Gly is the most abundant amino acid of elastin (37.2% of TE monomer); hence even minor enrichment of this residue is observed in the spectroscopic data. Additionally, Ser (1.3% of the TE monomer), is synthesized from Gly, adding another possible scrambling pathway.
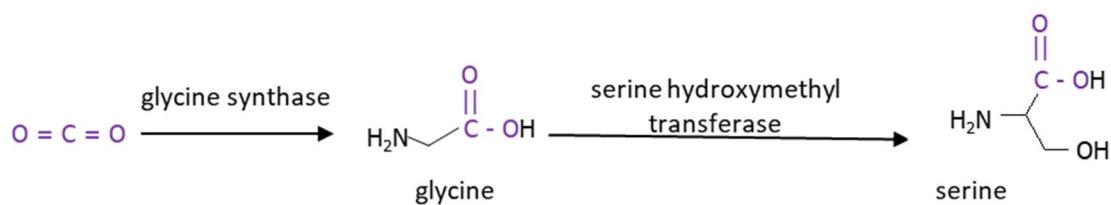
Figure 2.6: Biosynthesis of glycine and serine from carbon dioxide.
Carbon from CO_2 gas (purple) becomes the carbonyl of glycine and later the carbonyl of serine.

Preliminary analysis of the various pathways in the metabolism of lysine and other amino acids provides a short list of targets for the HPLC-MS assay. These theoretical and experimental approaches complement each other, ensuring viability and accuracy of the labeling protocol. High isotopic enrichment of lysine and the crosslinks is anticipated. Low levels of enrichment at ^{13}CO-Gly are also expected. The HPLC-MS assay will confirm whether these moieties are labeled and will provide a quantitative measure of the enrichment levels.

2.4 Enrichment evaluation via Marfey's chemistry

The objective of the assay is quantification of the enrichment at the lysines and the crosslinks as well as other residues. High incorporation of labeled lysine via direct substitution is expected, given this amino acid is essential. Therefore, enrichment levels of crosslinks formed in the post-translational modification of Lys are also expected to be substantial. The catabolism products of Lys serve as substrates for synthesis of other amino acids. If any scrambling occurred at Gly, Ser, or other residues, the enrichment needs to be quantified for an accurate analysis of the NMR spectra.

The enrichment assay is based on the Marfey's method, a protocol that has been widely used in the analysis of protein hydrolysate composition.[19,20,21,22] In this approach, amino acids are derivatized with 1-fluoro-2,4-dinitrophenyl-5-L-alanine amide (FDAA). The derivatization procedure is simple and inexpensive. The products are stable for up to a week in ambient temperatures, if protected from light. Sufficient resolution of amino acid derivatives is achieved using a common octadecyl carbon chain (C18) column, with acetonitrile-phosphate buffer as a mobile phase and UV detection.[20]

To determine isotopic enrichment with Marfey's chemistry, HPLC is combined with mass spectroscopy (MS). This setup is referred to as *the advanced Marfey's method*.[23] Minor modifications of the original protocol include replacement of the mobile phase with volatile acetonitrile-tetrafluoroacetic acid (TFA). In addition, to improve sensitivity of electrospray ionization (ESI), amino acids are derivatized with 2,4-dinitrophenyl-5-L-leucine amide (FDLA) instead of FDAA (figure 2.7).[24,25]

1-fluoro-2,4-dinitrophenyl-5-L-alanine amide
(FDAA)

1-fluoro-2,4-dinitrophenyl-5-L-leucine amide
(FDLA)
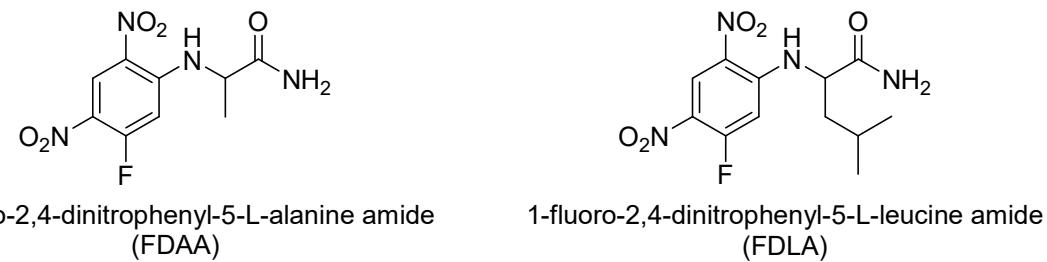
Figure 2.7: The derivatization reagents of the original and advanced Marfey's method.
FDAA is known as the Marfey's reagent. FDLA is its analog used in the enrichment assay for elastin.

The elastin samples are hydrolyzed, and the reaction with FDLA in mild alkaline conditions follows (10 min, 85°C). The free amine group on an amino acid reacts via nucleophilic substitution with the aromatic fluorine of FDLA (figure 2.8). If more than one amine is present, as in lysine or desmosine, multiple substitutions (and products) are possible. The derivatized amino acids are separated on a reverse-phase column. The retention times depend on the hydrophobicity of the species. The more hydrophobic derivatives interact more strongly with the C18 column, resulting in longer elution times.
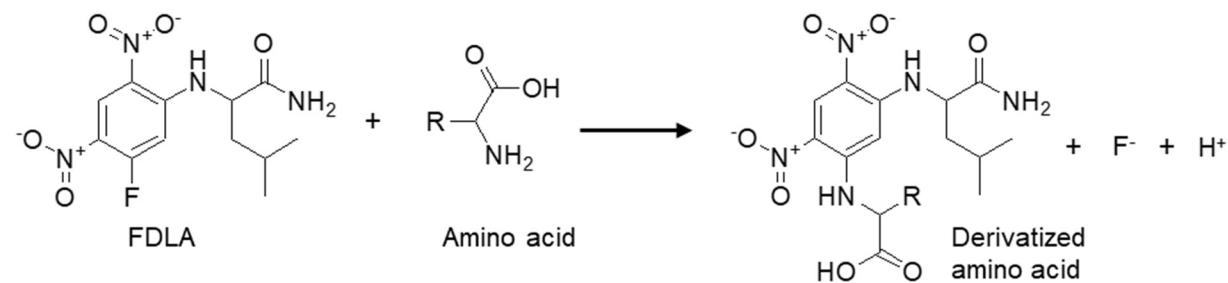
Figure 2.8: Reaction between FDLA and an amino acid.
All available amine groups react with the aromatic fluorine.

The relative abundances of the isotopically enriched (Ab[MH + n] [+]) and unlabeled (Ab[MH][+]) amino acids were used to quantify enrichment (%) in each residue, where n is the number of labeled nuclei in each structure. Wolfe's equation is utilized as follows[26]:

$$\text{isotopic enrichment (\%)} = \frac{100 \times (\text{Re} - \text{Rc})}{1 + \text{Re} - \text{Rc}} \hspace{3cm} \text{(Equation 2.1)}$$

where:

$$\text{Re} = \frac{\text{Ab}[\text{MH}+\text{n}]^+}{\text{Ab}[\text{MH}]^+} \quad \text{for the enriched sample}$$

$$\text{Rc} = \frac{\text{Ab}[\text{MH}+\text{n}]^+}{\text{Ab}[\text{MH}]^+} \quad \text{for the unenriched sample}$$

The analysis of crosslinks with multiple incorporation sites is complex. For example, n = 29 for desmosine in the [U-^{13}C,^{15}N-Lys] elastin sample. Fully enriched desmosine does not exist in nature (Ab[MH + 29]$^+$ = 0), therefore Rc = 0, as

$$\text{Rc} = \frac{\text{Ab}[\text{MH} + \text{n}]^+}{\text{Ab}[\text{MH}]^+} = \frac{0}{\text{Ab}[\text{MH}]^+} = 0$$

If a sample is highly enriched, it is likely the molecular ion of the unenriched DES ([MH]$^+$) will not be detected(Ab[MH]$^+$ = 0). In that case:

$$\text{Re} = \frac{\text{Ab}[\text{MH} + \text{n}]^+}{\text{Ab}[\text{MH}]^+} = \frac{\text{Ab}[\text{MH} + \text{n}]^+}{0}$$

In highly enriched elastin samples, the denominator of Re, Ab[MH]$^+$, is zero for DES and the enrichment (%) cannot be determined.

2.5 Materials and methods

This section provides the detailed procedure for the preparation of elastin samples through NRSMC culture, introduction of isotopically labeled lysine, and the evaluation of the isotope incorporation. The detailed descriptions of the standard growth media for unlabeled samples, and lysine-free media for isotopic labeling is found in Appendix I.

2.5.1 Preparation of elastin with isotopically enriched lysine via NRSMC cultures

Labeled and unlabeled elastin samples were prepared from cultures of NRSMC. Previous reports provide details of the extraction of these cells from neonatal rats and the preparation for long-term storage in liquid nitrogen.[10,27,28]

The frozen pellet of NRSMC was quickly thawed in a 37°C water bath and then resuspended in 10 mL of standard culture media at 37°C in a T-75 flask. The flask was incubated overnight at 5% CO_2 and 37°C to imitate physiological conditions. The media was replaced with 20 mL of standard culture media the next day and every other day thereafter (i.e., day 3 and day 5). After 7 days, the cells were 90% confluent. At that point, the cells were washed twice, each time with 10 mL of 1 x PBS solution. Two mL of trypsin were added and the flask was incubated for 5 - 6 min at 37°C. Four mL of standard culture media were added to deactivate trypsin. The resulting suspension was spun in a centrifuge (Thermo Electron Corporation, Waltham, MA) at 1000 rpm for 10 min. The supernatant was aspirated, and the pellet was resuspended in 3 mL of standard culture media. One mL of this suspension was added to each T-75 flask, which was filled with 9 mL of standard culture media at 37°C. This procedure results in three seeded T-75 flasks. Two of these flasks were used to grow [U-^{13}C,^{15}N-Lys] elastin, and the third one was used for unlabeled elastin sample (figure 2.9, yellow background). The entire procedure was repeated to obtain two flasks for growth of [ε-^{13}C-Lys] elastin and another with unlabeled protein (figure 2.9, pink background).
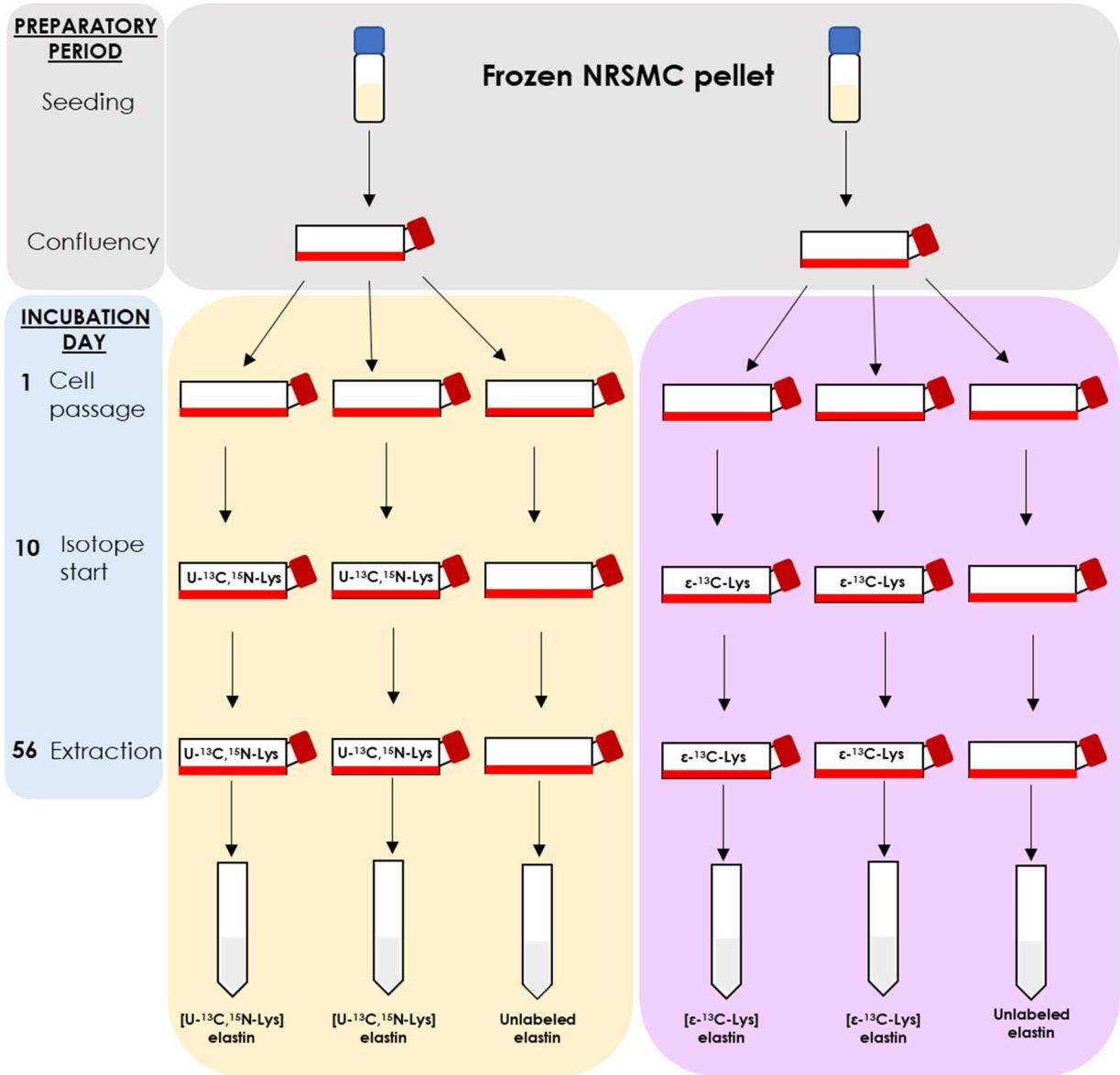
Figure 2.9: Preparation of elastin samples.
Preparation of [U-^{13}C,^{15}N-Lys] elastin (yellow background) and [ε-^{13}C-Lys] elastin (pink background). Two labeled and one unlabeled elastin sample was prepared for each labeling scheme. After a 7-day preparatory period, NRSMC were passed to three T-75 flasks (Day 1). The isotope was introduced to 2 out of 3 cell flasks (Flask 3 is control, natural-abundance) on Day 10. This supplementation continued until extraction on Day 56.

Elastin is synthesized at the highest rate when cells reach confluency, which occurred on Day 10 of the incubation. At this point, the labeled amino acids were introduced to the two out of the three flask. DMEM stock without lysine, glucose, arginine, glutamine, and phenol red (Gibco, New York, NY) was purchased. Before preparing the cell culture media, the commercial formulation was supplemented with D-glucose (1.000 g/L), L-arginine • HCl (0.084 g/L), L-glutamine (0.584 g/L), and phenol red • Na (0.11 g/L). L-Lys • HCl was supplemented to the control flask intended for growth of unlabeled elastin at a concentration of 0.146 g/L. [U-^{13}C,^{15}N, 99%] and [ε-^{13}C, 99%] lysine (Cambridge Isotope Laboratories, Tewksbury, MA) were used for the labeled flasks in the respective labeling schemes. [U-^{13}C,^{15}N-Lys] • 2 HCl was supplemented at a concentration of 0.181 g/L. The concentration of [ε-^{13}C-Lys] • 2 HCl was 0.175 g/L. Details of the custom media formulation and supplementation are provided in Appendix I.

The incubation was continued until the cultures reached Day 56 (or 8 weeks). After that time, the ECM layer was scraped from the three flasks and transferred to three Falcon tubes. The elastin samples were then purified using the cyanogen bromide (CNBr) method.[29] Briefly, 20 mL of CNBr solution (50 mg/mL in 70% formic acid) were added to each Falcon tube containing the collected cell matrix. The tubes were capped and left in a shaker overnight. Then, 12 mL of water were added to each tube. The mixtures were left uncapped in a hood for 4 hours to allow evaporation of any remaining toxic HCN. The tubes were centrifuged, and the supernatant was disposed. Twenty ml of guanidinium hydrochloride solution (5 M) and 0.2 mL of β-mercaptoethanol (1% w/w) were added to each tube and left overnight. A series of five centrifugations, each followed by washings with water, was conducted to remove any salts. Purified protein was stored in ultrapure water in a refrigerator at 4°C. Each T-75 flask yields 55-60 mg of purified elastin after removal of visible water (70-80% w/w) or ~20 mg protein (dry weight).

2.5.2 Quantification of isotopic incorporation

The following procedure was performed in triplicate for [U-^{13}C,^{15}N-Lys] elastin, [ε-^{13}C-Lys] elastin, and the control (unenriched) elastin. Approximately 2 mg of elastin are blotted on a KimWipe (Kimberley-Clark, Irving, TX) to remove visible water and transferred to a test tube with a lid. One half-mL of H_2O and 0.5 mL of concentrated HCl were added to the sample. The test tube was capped and left in a 100°C oil bath for 24 hours. The resulting hydrolysate was transferred to a 10 mL beaker and gently heated to dryness in a hood. The residue was then dissolved in 400 μL of an aqueous $NaHCO_3$ solution (84 mg/mL) and placed in a test tube with a flea. A bright yellow solution of FDLA in acetone (10 mg/mL) was added in the quantity of 200 μL. The biphasic yellow-colorless mixture was agitated until homogenous and then placed in an oil bath at 85°C for 5 – 10 min. After the solution cooled to room temperature, three extractions with ethyl ether followed. An aliquot of 250 μL of the purified aqueous solution was transferred to a 10 mL beaker and mixed with 150 μL of 2N HCl solution. Addition of the acid causes the red solution to form a yellow precipitate. It was confirmed with pH strips that pH ≈ 5. The precipitate was dissolved in 600 μL of acetonitrile, and this solution was run through a 20 μm filter before injection onto the HPLC column.

Each of the following three HPLC standards were prepared for reaction with 200 μL of FDLA solution. The derivatization procedure was the same as described for elastin hydrolysates.

(1) A derivatized mixed standard of glycine, alanine, proline, and lysine was used to optimize the chromatography conditions, and confirm the retention times. The solution was prepared from 5 mg of Lys, 19 mg of Gly, 12 mg of Ala, and 9 mg of Pro, which were dissolved in 10 mL of MiliQ water. The proportions mimic those naturally occurring in NRSMC elastin. Two hundred twenty μL of this standard were used in the derivatization reaction with FDLA.

(2) The second standard was an equimolar mixture of desmosine and isodesmosine (Cat. No. MD687, Elastin Products Co., Inc., Owensville, MO). It confirmed the retention times of the multiple substitution products obtained in the derivatization of these crosslinks. Five mg of the purchased sample were dissolved in 1 mL of water. One hundred seventy-six μL of this standard were used in the derivatization reaction with FDLA.

(3) A glutamine (Gln) standard was prepared to confirm its retention time. Gln is a potential site of scrambling, and it has similar molecular weight to Lys (MW_{Gln} = 146.14 g/mol, MW_{Lys} = 146.19 g/mol). Moreover, both amino acids form mono- and disubstitution products in the reaction with FDLA. A 50 μM aqueous Gln solution was prepared, and 220 μL were used in the derivatization reaction with FDLA.

All high-resolution MS data were acquired using an Agilent 6545 Q-ToF with attached Infinity 1290 UHPLC (Agilent Technologies, Santa Clara, CA). All MS data was collected using ESI in the positive mode. Chromatographic separations for LC-MS analysis were performed using a reverse-phase column with a gradient of two solvents: water with 0.1% formic acid as a modifier (mobile phase A) and acetonitrile (mobile phase B). A 2.1 x 50 mm, 1.8 µm Zorbax Eclipse Plus C18 column (Agilent Technologies, Santa Clara, CA) was used. One µL of the sample was injected onto the column. The initial solvent composition was 85% mobile phase A : 15% mobile phase B. After 0.75 minutes, mobile phase B was ramped to 65% over 9 minutes prior to the final wash at 100% mobile phase B. The flow was set to 0.5 mL/min. Chromatographic entities were identified using Agilent MassHunter Qualitative Analysis B.07.00 (Agilent Technologies, Santa Clara, CA).

To evaluate the precision of the approach, and ensure its accuracy, three samples of [ε-^{13}C-Lys] elastin are derivatized (E1, E2, E3) (figure 2.10) and then analyzed by LC/MS. The Re values for E1, E2, E3 are then paired with the Rc values of each of the four control (unlabeled) samples (N1, N2, N3, N4) to calculate twelve enrichment data points (Equation 2.1). Three of the unlabeled samples (N1-N3) were taken from the unenriched protein grown concurrently with [U-^{13}C,^{15}N-Lys] elastin. N4 was obtained from the protein that was grown concurrently with [ε-^{13}C-Lys] elastin. The four Rc values were sufficient for the quantification; all control samples were consistent with each other and reflected the natural-abundance levels of common isotopes.
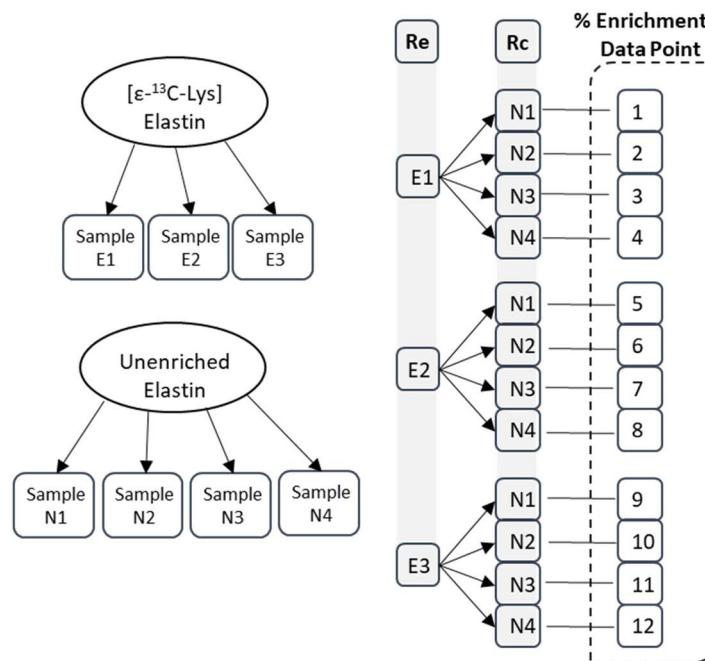
Figure 2.10: Mathematical analysis of LC-MS data for the [ε-^{13}C-Lys] elastin.
Each labeled sample (E1, E2, E3) was paired with each of the four unlabeled ones (N1, N2, N3, N4) to calculate a total of 12 enrichment values (enclosed by dashed lines).

A statistical analysis for each amino acid is conducted on the resulting data set. To evaluate the significance of the calculated enrichment levels, one-sided t-test was performed at the 95% confidence interval. If the mean enrichment value failed the significance test, 0.00% was listed for the given amino acid. Otherwise, the standard deviation of the mean, or standard error, σ_x, was calculated from the following formula:

$$\sigma_x = \frac{s}{\sqrt{N}}$$ (Equation 2.2)

Where s is the standard deviation and N is the number of data points.

Five samples of [U-^{13}C,^{15}N-Lys] elastin (20 data points, E1-E5 paired with N1-N4) were used to determine its enrichment. Initially, three samples of [U-^{13}C,^{15}N-Lys] elastin (12 data points) yielded a low calculated t-value for Gly (table 2.1) and, thus, the results appeared inconclusive. It was predicted that the enrichment levels of Gly in the [U-^{13}C,^{15}N-Lys] elastin would exceed or be the same as those in [ε-^{13}C-Lys] elastin. Glycine enrichment results from $^{13}CO_2$ formed in lysine catabolism, and [U-^{13}C,^{15}N-Lys] has more pathways leading to the $^{13}CO_2$ precursor than [ε-^{13}C-Lys] (figures 2.4 - 2.5). Scrambling to glycine is evident in [ε-^{13}C-Lys] elastin. To verify the results, two additional enriched samples were derivatized and included in the analysis. With the data for E1-E5 and N1-N4, the calculated t-value doubled, confirming the statistical significance of the obtained enrichment values.

Table 2.1: Statistical analysis of enrichment of glycine (%).

Elastin sample	Enrichment (%)	Number of data points	Calculated t-value	Critical t-value at 95% confidence level	Statistical significance of enrichment?
[ε-^{13}C-Lys]	1.4	12	6.535	1.796	Yes
[U-^{13}C,^{15}N-Lys]	0.4	12	1.567	1.796	No
[U-^{13}C,^{15}N-Lys]	0.5	20	3.109	1.729	Yes

2.6 Challenges specific to lysine enrichment evaluation

The LC-MS assay for Marfey's derivatives of elastin with isotopically enriched lysines poses several challenges that were not present in the analogous evaluations of glycines[11] or alanines[12]. The unique obstacles result from the low lysine content in tropoelastin, the variety of post-translational modification products, and multiple products of the FDLA derivatization. Each of the listed issues adds to the overall dilution of the crosslinking compounds, and in some cases, pushes the amount below the limits of detection for this method.

2.6.1 Low quantity of lysine and crosslinks in elastin

Lysine constitutes fewer than 5% of the residues in the tropoelastin monomer. This number is several times lower than the content of glycine and alanine, which make up 37% and 20% of the protein (or 311 and 164 residues), respectively. In contrast, there are only 38 lysines in the 836 residues of the monomer. Details of the amino acid sequence and composition of NRSMC tropoelastin are included in Appendix II.

This low number of lysine residues further decreases upon post-translational modification. Tropoelastin is a precursor to the mature insoluble elastin, and, once the crosslinks are formed, the content of unmodified lysine drops from 38 to 5[30], corresponding to < 1% of residues in the monomer. To make matters even more complicated, modified lysines form twelve different products (figure 2.2), out of which eight have been quantified (table 2.2). Variations in the reported quantities are attributed to tissue source, age, protein processing, and the experimental approach. The content of each crosslink never exceeds a few residues per tropoelastin monomer.

Table 2.2: Quantities of crosslinks in elastin from different tissues.

Crosslink	Quantity (Lys equiv. per 1000 residues)	Tissue type
Allysine	2, 9.5, <1	Rat aorta, 1, 2, 8-15 weeks, respectively[31]
Lysinonorleucine	2.2	BLN[32]
	1.0 – 2.2	BLN (ranging from foetal to 12 years old)[33]
	2	Pig aorta[34]
Dehydrolysinonorleucine	0 - 0.3	BLN (ranging from foetal to 12 years old)[33]
Allysine Aldol	4-5	BLN[35]
	5 - 7	Chick embryo aorta tissue culture[35]
	1.5 - 3.4	BLN (ranging from foetal to 12 years old)[33]
Merodesmosine	0.5	BLN[32]
	0.2 - 0.3	BLN (ranging from foetal to 12 years old)[33]
Dehydromerodesmosine	0.2	BLN[32]
	0 - 0.9	BLN (ranging from foetal to 12 years old)[33]
Desmosine	10.1	BLN[32]
	0.8 - 1.8	BLN (ranging from foetal to 12 years old)[33]
	4.4	NRSMC cell culture[30]
Isodesmosine	5.4	BLN[32]
	0.8 - 1.3	BLN (ranging from foetal to 12 years old)[33]
	2.8	NRSMC cell culture[30]
Pentasine	Trace	BLN[36]
Cyclopentasine	Not quantified	BLN[37]
Allodesmosine	Not quantified	BLN[38]
Neodesmosine	Not quantified	BLN[39]

2.6.2 Multiple Marfey's derivatives

Lysine and crosslinks each form multiple products in the reaction with the Marfey's reagent. For example, reaction of FDLA with Ala results in a single product with one retention time. In the same reaction with Lys, di-substitution (FDLA-Lys-FDLA) and mono-substitution (FDLA-Lys 1 and FDLA-Lys 2) products are formed (figure 2.11A). Instead of observing one signal resulting from the five unmodified Lys residues per TE monomer, the intensity is split between three peaks (figure 2.11b). The obtained product ratios are not reproducible.

Relative abundance of Lys derivatives in MS spectra

Batch	FDLA-Lys 1	FDLA-Lys 2	FDLA-Lys-FDLA	Total
Unenriched 04/2018	1.25	2.75	1.00	5.00
Unenriched 12/2018	0.75	1.35	2.90	5.00
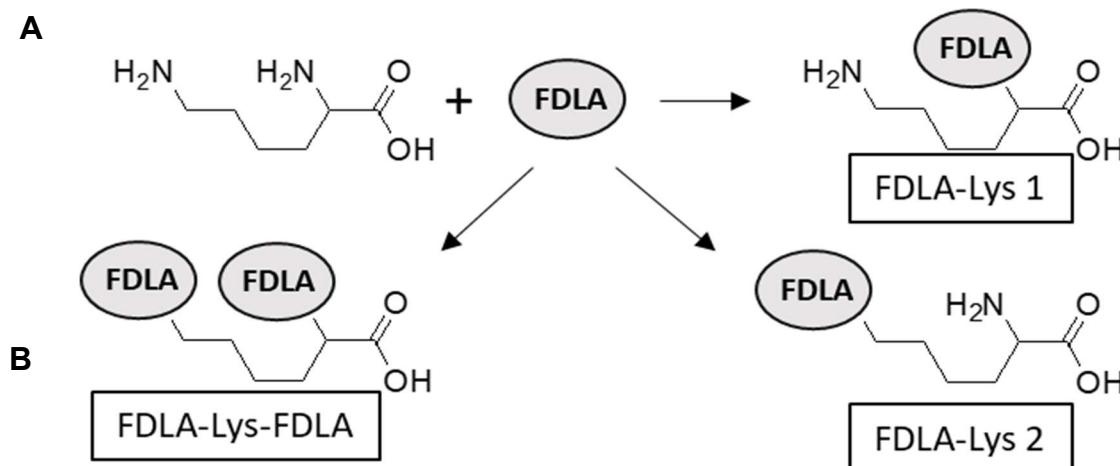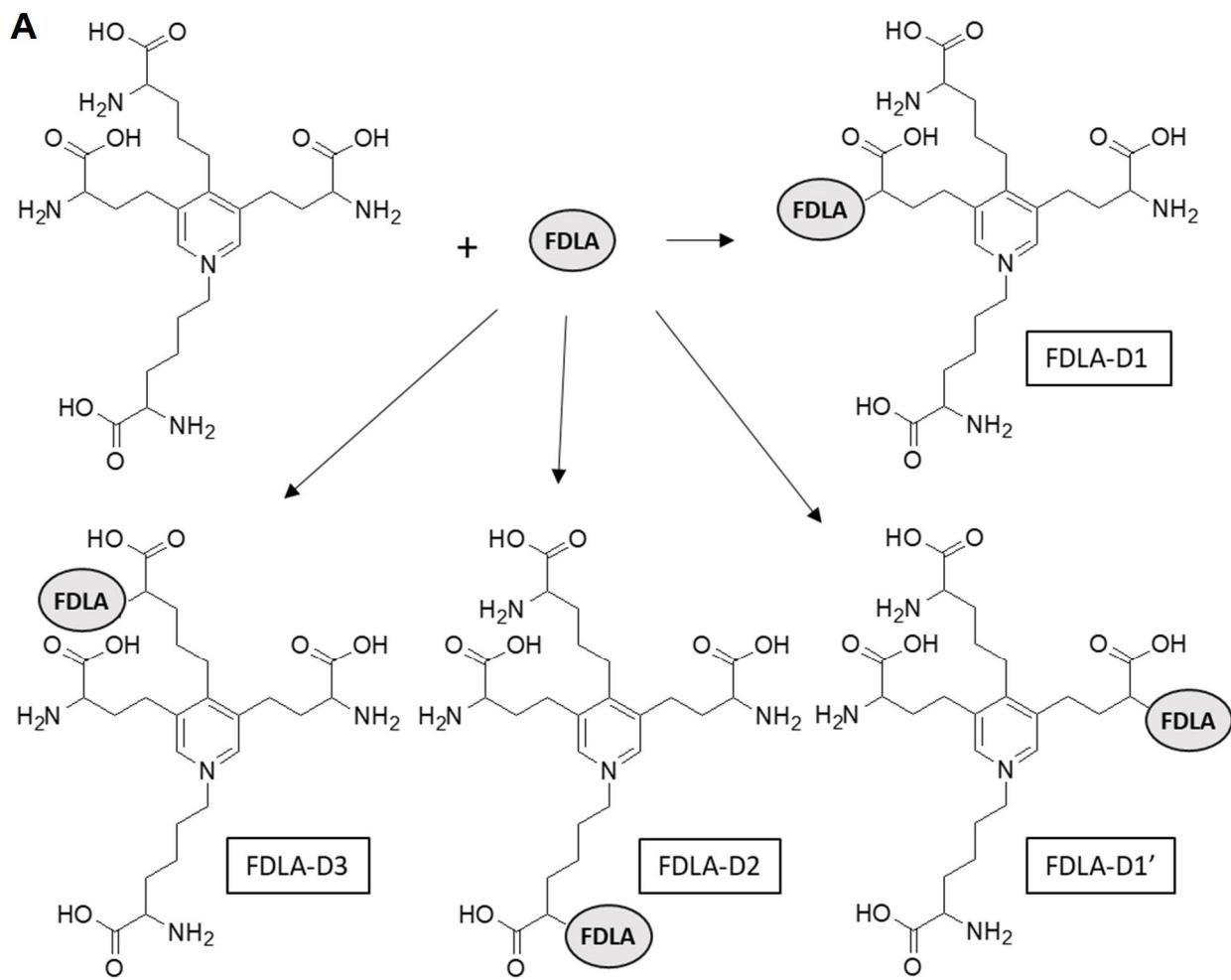
Figure 2.11: FDLA derivatives of lysine.
A: Products of derivatization of lysine with FDLA.
B: Relative abundance of each lysine derivative in two batches of unenriched elastin. Total (sum) of abundance values is normalized to 5. FDLA-Lys 1 and FDLA-Lys 2 are assigned to the structures in (A) arbitrarily.

The problem of multiple Marfey's derivatives becomes even more complex as the number of lysine equivalents in a crosslink increases. In tetrafunctional desmosine, four amine groups are available for substitution. Three chromatographically distinct products are formed in the monosubstitution (figure 2.12a). Di-, tri-, and tetra- substitution takes place as well. The derivatization time was extended from 10 min to 20 min to maximize the quantity of the tetra-FDLA-(I)DES. However, the results were similar to the previous runs.

The HPLC-MS assay is a sensitive method, which resolves the isomers formed in monosubstitution of DES (FDLA-D1, FDLA-D2, FDLA-D3) as well as the di- and tri-substitution. Therefore, a total of twelve distinct DES derivatization products are formed (figure 2.12b). IDES is present in comparable quantities to DES, but it does not have any structural symmetry. Thus, it yields sixteen distinct products.

Figure 2.12: Possible products of the reaction between desmosine and FDLA.
A: Reaction scheme showing mono-substitution products. FDLA-D1 and FDLA-D1' are identical, due to symmetry of the molecule.
B: Summary of distinguishable products for desmosine and isodesmosine.

2.6.3 Multiple isotope incorporation patterns

Incomplete isotope incorporation into crosslinks results in varied molecular weights. For example, if 90% of Lys residues in elastin is enriched, one in every ten residues is unenriched (figure 2.13). Given one DES consists of four Lys equivalents, the aforementioned 10 Lys produce 2.5 DES residues. One of these crosslinks has a lower molecular weight than the remaining 1.5 molecules, as it incorporated the single unenriched Lys residue. The partially enriched crosslinks in both [U-^{13}C,^{15}N-Lys] elastin and [ε-^{13}C-Lys] elastin are not detected with MS.
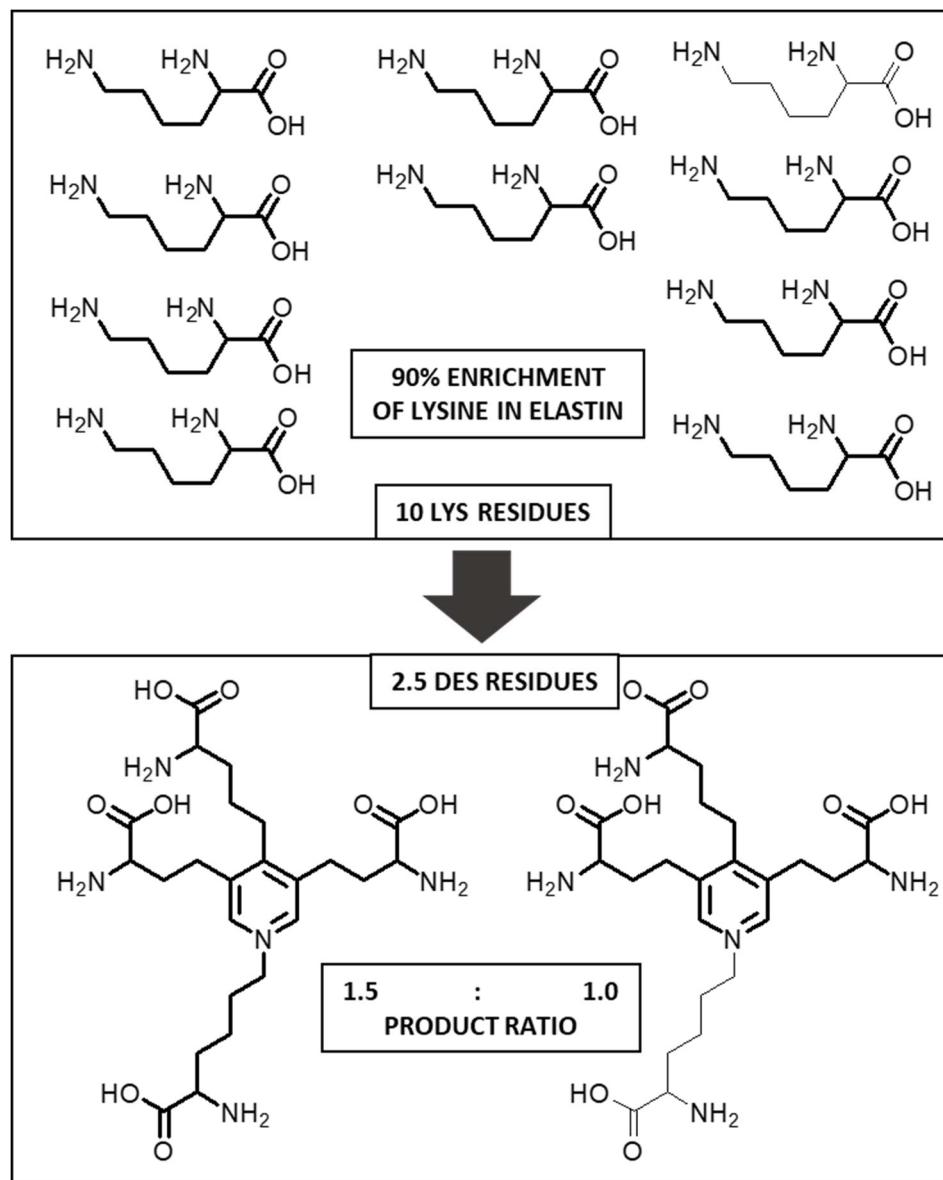
Figure 2.13: Example of incomplete isotope incorporation in desmosine in [U-^{13}C,^{15}N-Lys] elastin.
Bold lines symbolize fragments carrying ^{13}C and ^{15}N labels.

Some other, less likely, combinations are possible as well, such as two enriched and two unenriched or one enriched and three unenriched Lys equivalents per desmosine molecule. However, even assuming only the discussed scenario, the most intense desmosine peak recognized by the MassHunter algorithm of the Agilent software will correspond to the molecular weight of only 1.5 fully enriched desmosine molecules out of every 2.5 desmosine residues (figure 2.13). The larger the number of Lys equivalents per crosslink, the larger the effect of incomplete isotope incorporation will be.

2.6.4 Synergy of the challenges

The described challenges (Sections 2.6.1 - 2.6.3) all contribute to detection of multiple signals arising from a single amino acid. The challenges produce a synergistic effect. For example, desmosine has 5 potential incorporation patterns (0 - 4 enriched lysine equivalents in the structure). Given 12 possible products of the reaction with FDLA, technically 60 distinct outcomes are possible. Some of these products are more prevalent than others, but all of them contribute to dilution, in some cases, resulting in concentrations below the detection limits. As a result, the rare crosslinks are not observed at all with MS. The detection of the more common DES and IDES is inconsistent. The limitations of the Marfey's method and Wolfe's equation need to be acknowledged in the analysis of this complex system.

2.7 Results and discussion

An HPLC-MS assay evaluating enrichment of [U-^{13}C,^{15}N-Lys] and [ε-^{13}C-Lys] elastin was conducted. The protein samples were hydrolyzed and then derivatized with FDLA. MS data for these derivatives were used to calculate the enrichment (%) of each amino acid. The reported values reflect the percentage of the residues that carry a labeled nucleus (^{13}C or ^{15}N), exceeding natural-abundance levels.

2.7.1 Enrichment of lysine

High incorporation of labeled lysine into elastin was achieved via direct substitution. The enrichment of [U-^{13}C,^{15}N-Lys] elastin was 87.3 ± 0.4% (table 2.3). The calculations for [U-^{13}C,^{15}N-Lys] elastin were based on data for FDLA-Lys-FDLA, as the other substitution products, i.e. FDLA-Lys 1 and FDLA-Lys 2, were not detected in all of the HPLC-MS trials.

The enrichment of lysine in [ε-^{13}C-Lys] elastin was estimated to be ~87%, or equivalent to that of U-^{13}C,^{15}N-Lys] elastin. NMR data confirm successful incorporation of [ε-^{13}C-Lys] (Chapter 4). The relative abundance of the enriched lysine was detected in each run. However, the unenriched lysine was not detected in any of the three trials for [ε-^{13}C-Lys] elastin. Therefore, the Wolfe equation was not applicable, as Re would have 0 in the denominator (equation 2.1). Both labeling protocols followed the same schedule for, e.g., isotope introduction and protein extraction. Thus, the enrichment levels are expected to be comparable.

Table 2.3: Lysine enrichment in [U-^{13}C,^{15}N-Lys] elastin and [ε-^{13}C-Lys] elastin.
Number of residues is given per mature TE monomer after post-translational modification.

Amino Acid	Residues per Monomer	Enrichment (%)	
		[U-^{13}C,^{15}N-Lys] Elastin	[ε-^{13}C-Lys] Elastin
Lys	5[a]	87.3 ± 0.4	~ 87 (estimated)
	7[b]		

[a] amino acid analysis of elastin from 8-week long NRSMC cell culture[30]
[b] NMR studies, see Chapter 4

2.7.2 Enrichment of desmosine and isodesmosine

Desmosine and isodesmosine, hereafter defined as (I)DES, are indistinguishable by MS. The assay results for these crosslinks are inconsistent. To calculate enrichment (%), labeled and unlabeled (I)DES needs to be detected in a single enriched sample. Only one HPLC-MS run for [U-^{13}C,^{15}N-Lys] elastin met this condition (table 2.4). In the other seven trials for enriched elastin (of either labeling scheme), only the labeled (I)DES was observed; the unenriched crosslink was not detected. Similarly, not all the unlabeled elastin runs provided data on unenriched (I)DES.

Table 2.4: HPLC-MS results for (I)DES in [U-^{13}C,^{15}N-Lys] elastin.
Tetra-FDLA-(I)DES refers to desmosine and isodesmosine with all four amino groups substituted with FDLA.

Compound	Retention Time (min)	M/z	Relative Abundance (Ab)
Tetra-FDLA-(I)DES [MH]$^+$	9.572	1702.6740	4520.32
Tetra-FDLA-(I)DES (labeled) [MH+29]$^+$	9.572	1731.7385	64742.80

Wolfe's equation does not capture the complexity of the tetrafunctional crosslinks. This approach neglects the multiple isotopic incorporation patterns, that is, labeled (I)DES which contains at least one unenriched lysine equivalent. With 87% enrichment of Lys, partially labeled (I)DES is non-negligible; approximately 1 out of every 2.5 (I)DES residues is partially labeled (figure 2.13).

Converting the relative abundances to ratios, rather than percentages, is a more informative way of interpreting the data. For every unenriched (I)DES (m/z = 1703) there are ~14 fully enriched ones (m/z = 1732). The HPLC-MS assay did not provide data on partially labeled species. Thus, their exact contribution was not determined experimentally. However, following the ratio determined in Section 2.6.3, it is estimated that for every 14 fully enriched (I)DES residues (m/z = 1732), there is ~9 partially labeled residues (m/z = 1723) (figure 2.13).

2.7.3 Scrambling to glycine and aspartic acid

Quantification of the enrichment of ^{13}CO-Gly allows for an accurate interpretation of NMR spectra. Enrichment of glycine in [U-^{13}C,^{15}N-Lys] and [ε-^{13}C-Lys] elastin was 0.5 ± 0.2% and 1.4 ± 0.2%, respectively (table 2.5). These percentages translate to 1.6 and 4.4 glycine residues per TE monomer. In both labeled proteins, the ^{13}CO$_2$ released during decomposition of Lys (figures 2.4 – 2.5) was incorporated at the carbonyl position in Gly (figure 2.6).

Scrambling to aspartic acid was confirmed in [U-^{13}C,^{15}N-Lys] elastin. Enrichment of 1.9 ± 0.4% was found. Asp is a minor component of elastin (1 residue per 836 in a TE monomer). Therefore, the contribution of this enriched amino acid is equivalent to 0.02 residues of ^{13}CO-Asp per monomer. This relatively small quantity of ^{13}C is negligible in NMR spectra. No scrambling to [ε-^{13}C-Lys] elastin was observed, which is consistent with the predictions based on metabolic pathways.

41

Table 2.5: Scrambling in [U-^{13}C,^{15}N-Lys] elastin and [ε-^{13}C-Lys] elastin.
Number of residues is given per TE monomer, that is, 836 amino acids.

Amino Acid	Residues per Monomer	[U-^{13}C,^{15}N-Lys] Elastin		[ε-^{13}C-Lys] Elastin	
		Enrichment (%)	Number of Enriched Residues per Monomer	Enrichment (%)	Number of Enriched Residues per Monomer
Gly	311	0.5 ± 0.2	1.56 ± 0.62	1.4 ± 0.2	4.35 ± 0.62
Asp	1	1.9 ± 0.4	0.02 ± 0.01	0.0	0.00

The assay detects isotopic enrichment of an amino acid but does not distinguish between ^{15}N and ^{13}C incorporation. Both nuclei were introduced to [U-^{13}C,^{15}N-Lys] elastin. However, no metabolic pathway that that transfers ^{15}N from Lys to Gly or Asp was identified, whereas such routes exist for ^{13}C (figure 2.5). Thus, the enrichment of these two amino acids is attributed to the ^{13}C nuclei. NMR studies are consistent with ^{13}CO-Gly scrambling (Chapter 3).

2.7.4 Results for remaining amino acids

Glycine, aspartic acid, serine, glutamine, and arginine were predicted sites of scrambling. Enrichment of Gly, which serves as a precursor to Ser, was confirmed in both [U-^{13}C,^{15}N-Lys] and [ε-^{13}C-Lys] elastin. However, no enrichment of Ser was found. Gln, Arg, and Asp are synthesized from acetyl CoA, a product of lysine catabolism, which enters the Krebs cycle (figure 2.5). Scrambling to Asp in the U-^{13}C,^{15}N-Lys] elastin was confirmed. However, Gln and Arg were not enriched in [U-^{13}C,^{15}N-Lys] elastin, and neither of these three amino acids was enriched in [ε-^{13}C-Lys] elastin. These experimental results agree with the predictions based on the metabolic pathways.

2.8 Summary and conclusions

High incorporation of labeled lysine into elastin (~87%) was achieved with the NRSMC line. This expression system produces NMR-scale quantities of native, crosslinked elastin. The direct substitution approach was used, i.e., unlabeled lysine was replaced with enriched lysine in the culture media. Targeted enrichment of valine[11] and glycine[10] was achieved previously with this method.

Evaluation of labeled lysines in elastin posed unique challenges. Unlike previously studied amino acids[10,11,12], Lys is a minor component of this protein. Moreover, its low concentration is decreased by post-translational modification, which results in formation of multiple (Lys-derived) crosslinks. These products consist of several lysine equivalents, which introduces additional complexity to the studied system. A variety of species with different molecular weights is derived from a single crosslink due to incomplete substitution with FDLA and partial incorporation of the labeled lysine. Understanding of the intricate system of lysine-derived crosslinks is necessary for an accurate interpretation of the HPLC-MS data.

The label was successfully passed to desmosine and isodesmosine, but their enrichment (%) could not be accurately determined. However, the ratio of fully enriched (I)DES to unenriched (I)DES was found to be 14 : 1. The rare bifunctional crosslinks, such as LNL or AA, were not detected by MS, most likely due to low concentrations. These modification products are, however, observed in NMR data. Details of the quantification of Lys and Lys-derived residues by ssNMR spectroscopy are provided in Chapter 4.

Minor scrambling to ^{13}CO-Gly (< 2%) was detected in both [U-^{13}C,^{15}N-Lys] and [ε-^{13}C-Lys] elastin. Considering this amino acid is the main component of elastin, ^{13}CO-Gly is non-negligible in the NMR data. In [U-^{13}C,^{15}N-Lys] elastin, 2% enrichment of Asp was found. Only trace quantities of this amino acid are present in elastin, therefore ^{13}CO-Asp has no impact on the NMR spectra. No other sites of scrambling were identified at 95% confidence interval.

Highly enriched samples [U-^{13}C,^{15}N-Lys] elastin and [ε-^{13}C-Lys] elastin were obtained for ssNMR studies. Advanced Marfey's method adjusted to the complex system of crosslinks allowed for the assessment of enrichment. The determined values are consistent with ssNMR spectra (Chapter 4). Details of characterization of Lys residues and crosslinks by ssNMR spectroscopy are provided in Chapters 3 and 4.

2.9 References

[1] Harris, R. K. *Nuclear magnetic resonance spectroscopy*. John Wiley and Sons Inc. New York, NY: 1986.

[2] Fleming, W. W.; Sullivan, C. E.; Torchia, D. A. Characterization of molecular motions in carbon-13-labeled aortic elastin by carbon-13-proton magnetic double resonance. *Biopolymers* **1980**, *19*, 597-617.

[3] Torchia, D. A.; Batchelder, L. S.; Fleming, W. W.; Jelinski, L. W.; Sarkar, S. K.; Sullivan, C. E. Mobility and function in elastin and collagen. *Ciba Found. Symp.* **1983**, *93*, 98-115.

[4] Keeley, F. W.; Bellingham, C. M.; Woodhouse, K. A., Elastin as a self-organizing biomaterial: use of recombinantly expressed human elastin polypeptides as a model for investigations of structure and self-assembly of elastin. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2002**, *357*, 185-189.

[5] Hong, M.; Isailovic, D.; McMillan, R. A.; Conticello, V. P. Structure of an Elastin-Mimetic Polypeptide by Solid-State NMR Chemical Shift Analysis. *Biopolymers* **2003**, *70*, 158-168.

[6] Mithieux, S. M.; Rasko, J. E. J.; Weiss, A. S. Synthetic elastin hydrogels derived from massive elastic assemblies of self-organized human protein monomers. *Biomaterials* **2004**, *25*, 4921-4927.

[7] Velasco, A. M.; Leguina, J. I.; Lazcano, A. Molecular Evolution of the Lysine Biosynthetic Pathways. *J. Mol. Evol.* **2002**, *55*, 445-449.

[8] Lim, D. W.; Nettles, D. L.; Setton, L. A.; Chilkoti, A. Rapid cross-linking of elastin-like polypeptides with (hydroxymethyl)phosphines in aqueous solution. *Biomacromolecules* **2007**, *8*, 1463-1470.

[9] Oakes, B. W; Batty, A. C. The synthesis of elastin, collagen, and glycosaminoglycans by high density primary cultures of neonatal rat aortic smooth muscle. An ultrastructural and biochemical study. *Eur. J. Cell Biol.* **1982**, 27, 34-46.

[10] Perry, A.; Stypa, M. P.; Foster, J. A.; Kumashiro, K. K. Observation of the glycines in elastin using ^{13}C and ^{15}N solid-state NMR spectroscopy and isotopic labeling. *J. Am. Chem. Soc.* **2002**, *124*, 6832-6833.

[11] Ohgo, K.; Dabalos, C. L.; Kumashiro, K. K. Solid-State NMR Spectroscopy and Isotopic Labeling Target Abundant Dipeptide Sequences in Elastin's Hydrophobic Domains. *Macromolecules* **2018**, *51*, 2145-2156.

[12] Djajamuliadi, J.; Ohgo, K.; Kumashiro, K. K. Targeting Alanines in the Hydrophobic and Cross-Linking Domains of Native Elastin with Isotopic Enrichment and Solid-State NMR Spectroscopy. *Macromolecules* **2018**, *51*, 2157-2168.

[13] Pinnell, S. R.; Martin, G. R. The Cross-Linking of Collagen and Elastin: Enzymatic Conversion of Lysine in Peptide Linkage to α -Aminoadipic-δ -Semialdehyde (Allysine) by an Extract from Bone. *PNAS* **1968**, *61*, 708-716.

[14] Rucker, R. B.; Murray, J. Cross-linking amino acids in collagen and elastin. *AJCN* **1978**, *31*, 1221-1236.

[15] Eyre, D. R.; Paz, M. A.; Gallop, P. M. Cross-linking in collagen and elastin. *Annu. Rev. Biochem.* **1984**, *53*, 717-748.

[16] Watanabe, M.; Sawai, T.; Nagura, H.; Suyama, K. Age-related alteration of cross-linking amino acids of elastin in human aorta. *Tohoku J. Exp. Med.* **1996**, *180*, 115-130.

[17] Bender, D. A. Amino acid metabolism, 3rd ed., Wiley-Blackwell: Chichester, 2012.

[18] Salway, J. G. Metabolism at glance, 3rd ed., Blackwell Pub.: Malden, 2004.

[19] Marfey, P. Determination of D -amino acids. II. Use of a bifunctional reagent, 1,5-difluoro-2,4-dinitrobenzene. *Carlsberg Res. Commun.* **1984**, *49*, 591-596.

[20] Bhushan, R.; Brückner, H. Use of Marfey's reagent and analogs for chiral amino acid analysis: Assessment and applications to natural products and biological systems. *J. Chromatogr. B* **2011**, *879*, 3148-3161.

[21] Szokan, G.; Mezo, G.; Hudecz, F. Application of Marfey's reagent in racemization studies of amino acids and peptides. *J. Chromatogr.* **1988**, *444*, 115-122.

[22] B'Hymer, C.; Montes-Bayon, M.; Caruso, J. A. Marfey's reagent: Past, present, and future uses of 1-fluoro-2,4-dinitrophenyl-5-L-alanine amide. *J. Sep. Sci.* **2003**, *26*, 7-19.

[23] Fujii, K.; Ikai, Y.; Mayumi, T.; Oka, H.; Suzuki, M.; Harada, K. A nonempirical method using LC/MS for determination of the absolute configuration of constituent amino acids in a peptide: elucidation of limitations of Marfey's method and of its separation mechanism. *Anal. Chem.* **1997**, *69*, 3346-3352.

[24] Harada, K.; Fujii, K.; Mayumi, T.; Hibino, Y.; Suzuki, M.; Ikai, Y.; Oka, H. A method using LC/MS for determination of absolute configuration of constituent amino acids in peptide - advanced Marfey's method. *Tetrahedron Lett.* **1995**, *36*, 1515-1518.

[25] Fujii, K.; Ikai, Y.; Oka, H.; Suzuki, M.; Harada, K. A nonempirical method using LC/MS for determination of the absolute configuration of constituent amino acids in a peptide: combination of Marfey's method with mass spectrometry and its practical application. (liquid chromatography/mass spectrometry). *Anal. Chem.* **1997**, *69*, 5146-5151.

[26] Wolfe, R. R., *Tracers in Metabolic Research: Radioisotope and Stable Isotope/mass Spectometry Methods*. A.R. Liss: 1984.

[27] Dabalos, C., Characterization of the proline residues in elastin. Ph.D. Dissertation, University of Hawai'i, Honolulu, HI, 2016.

[28] Djajamuliadi, J., Investigations of the unique role of alanines in the "elastin puzzle" by solid-state NMR spectroscopy and molecular dynamics simulations. Ph.D. Dissertation, University of Hawai'i, Honolulu, HI, 2017.

[29] Perry, A.; Stypa, M. P.; Tenn, B. K.; Kumashiro, K. K. Solid-State ^{13}C NMR Reveals Effects of Temperature and Hydration on Elastin. *Biophys. J.* **2002**, *82*(2), 1086-1095.

[30] Barone, L. M.; Faris, B.; Chipman, S. D.; Toselli, P.; Oakes, B. W.; Franzblau, C. Alteration of the extracellular matrix of smooth muscle cells by ascorbate treatment. *Biochim. Biophys. Acta* **1985**, *840*(2), 245-254.

[31] Umeda, H.; Kawamorita, K.; Suyama, K. High-performance liquid chromatographic quantification of allysine as bis- p -cresol derivative in elastin. *Amino Acids* **2001**, *20*(2), 187-199.

[32] Field, J.M.; Rodger, G.W.; Hunter, J.C.; Serafini-Fracassini, A.; Spina, M. Isolation of elastin from bovine auricular cartilage. *Arch. Biochem. Biophys*. **1978**, *191*(2), 705-713.

[33] Francis, G.; John, R.; Thomas, J. Biosynthetic pathway of desmosines in elastin. *Biochem. J.* **1973**, *136*(1), 45-55.

[34] Davril, M.; Han, K. Purification of lysinonorleucine cross-linked peptide fraction from porcine aorta elastin. *Int. J. Pept. Protein Res*. **1976**, *8*(2), 177-181.

[35] Lent, R. W.; Smith, B.; Salcedo, L. L.; Faris, B.; Franzblau, C. Studies on the reduction of elastin. II. Evidence for the presence of alpha-aminoadipic acid delta-semialdehyde and its aldol condensation product. *Biochem.* **1969**, 8(7), 2837-2845.

[36] Starcher, B.; Cook, G.; Gallop, P.; Hensen, E.; Shoulders, B. Isolation and Characterization of a Pentameric Amino Acid from Elastin. *Connect. Tissue Res*. **1987**, *16*(1), 15-25.

[37] Akagawa, M.; Yamazaki, K.; Suyama, K. Cyclopentenosine, Major Trifunctional Crosslinking Amino Acid Isolated from Acid Hydrolysate of Elastin. *Arch. Biochem. Biophys.* **1999**, *372*(1), 112-120.

[38] Suyama, K.; Nakamura, F. Isolation and characterization of new cross-linking amino acid, allodesmosine, from the acid hydrolysate of elastin. *Biochem. Biophys. Res. Commun*. **1990**, *170*(2), 713-718.

[39] Usuki, T.; Yanuma, H. Synthesis of neodesmosine, a crosslinking pyridinium amino acid of elastin, via a Negishi cross-cneofoupling. *Heterocycles* **2013**, *87*(1), 55-63.

CHAPTER 3

SECONDARY STRUCTURES AND DYNAMICS OF LYSINE RESIDUES IN ELASTIN

3.1 Introduction

SsNMR is the only high-resolution technique that provides structural information about elastin. Since this protein does not crystallize and does not dissolve in any standard solvent, X-ray diffraction and solution NMR studies cannot be performed. This chapter focuses on secondary structures and dynamics in [U-^{13}C,^{15}N-Lys] elastin, which allows for observation of all the ^{13}C sites originating from lysine.

One- and two-dimensional ^{13}C NMR spectra are compared to the data in the Re-**Ref**erenced Protein Chemical Shift **D**ata**B**ase (RefDB) to assign secondary structures and evaluate the neighboring residue effects (Section 3.3.1 – 3.3.3). The influence of temperature on the conformations of Lys is probed via a variable-temperature (VT) ssNMR study (section 3.3.4). The spin-lattice relaxation times, T_1, are measured with a modified inversion recovery pulse sequence (Section 3.3.5).

3.1.1 Secondary structures and chemical shifts of amino acids

Determination of the three-dimensional (3D) structures of proteins is a common research objective. Up to this date (10/24/2019), spatial arrangements of 157,145 biological macromolecules have been described in the Protein Data Bank (PDB).[1] Over 131,000 of these structures were elucidated with X-ray diffraction, whereas 11,000 were based on NMR data. Elastin is not found in the PDB database, as it cannot be studied with traditional approaches.

Proteins have four levels of structure: primary, secondary, tertiary, and quaternary. Primary refers to the sequence of amino acids, encoded by DNA, and held together by peptide bonds. Secondary structure is defined by the conformation of local segments and depends on the hydrogen-bonding patterns. The final result of the folding process, i.e., the 3D shape of the entire protein, is the tertiary structure. Sometimes multiple polypeptide chains operate as a single functional unit. Organization of such subunits is described by quaternary structure.

The most prevalent secondary structure in proteins is the α-helix. Alanine has the highest known helical propensity and, accordingly, a helical penalty of 0.00 kJ/mol (table 3.1). Proline disrupts α-helices, because it does not have a protonated nitrogen to donate to a hydrogen bond[2]; hence it has the highest penalty (13.22 kJ/mol). The conformational flexibility of glycine (4.18 kJ/mol) makes the relatively constrained α-helices undesirable. Lysine has a low penalty (1.09 kJ/mol) and is commonly found in helices.

Table 3.1: Relative α-helical propensities of amino acids.[3]
Helical penalty for alanine is arbitrarily set to 0.00, and the relative helical penalties for the remaining amino acids are listed in ascending order. (The role of neighboring residues is neglected.)

Amino Acid	Helical penalty (kJ/mol)
Ala	0.00
Arg	0.88
Leu	0.88
Met	1.00
Lys	1.09
Gln	1.63
Glu	1.67
Ile	1.72
Trp	2.05
Ser	2.09
Tyr	2.22
Phe	2.26
His	2.55
Val	2.55
Asn	2.72
Thr	2.76
Cys	2.85
Asp	2.89
Gly	4.18
Pro	13.22

Both sequence and conformation affect the isotropic chemical shifts of an amino acid in a protein. The influence of neighboring residues[4,5] and secondary structure[6] have been investigated with statistical analyses of chemical shift databases. The sequence-dependent effects are the most pronounced at ^{15}N and ^{13}CO sites of amino acids in proteins.[4] In the three-amino-acid sequence, $R_{-1} - Lys - R_{+1}$, the chemical shifts of ^{13}CO-Lys and $^{13}C\alpha$-Lys are influenced by the residue that follows lysine (R_{+1}).

The data for lysine in different conformations show some resolution (table 3.2), but a definitive assignment is not possible based on a single ^{13}C-Lys site. For example, $^{13}C\alpha$-Lys in α-helix (59.11 ppm) is well resolved from its counterpart in a β-sheet (55.01 ppm). However, the same site in a random coil (56.40 ppm) overlaps with $^{13}C\alpha$-Lys in β-sheet. This ambiguity can be resolved in two-dimensional ssNMR correlations.

Table 3.2: ^{13}C chemical shifts of lysine categorized by secondary structure.[6]
Averaged values are based on a dataset for 36 distinct proteins with > 6100 amino acids. Standard deviation is shown in parentheses.

	δ (^{13}C) (ppm)		
	α-helix	**β-sheet**	**random coil**
^{13}CO-Lys	177.79 (2.22)	174.93 (1.25)	176.15 (1.40)
$^{13}C\alpha$-Lys	59.11 (1.19)	55.01 (1.00)	56.40 (1.80)
$^{13}C\beta$-Lys	32.31 (1.08)	34.86 (1.79)	32.57 (1.30)

Predominant amino acids of elastin, glycine (Gly), alanine (Ala), and proline (Pro), have been isotopically labeled and characterized by ssNMR spectroscopy previously.[7,8,9,10] Findings for Ala are particularly relevant to the conformational analysis of the Lys residues. These two amino acids are found in the crosslinking (CL) domains in tandem repeats, such as KAAK or KAAAK. Evidence for α-helical and random coil populations was found for Ala residues in the CL domains. No β-sheet conformations were observed (table 3.3).

Table 3.3: Alanine populations in NRSMC elastin.[7]
Table reproduced from Djajamuliadi et. al.[7] Ala residues are found in α-helices or random coils. Lys residues are emphasized in a bold font. Cross-linking (CL) and hydrophobic (HP) domains are indicated. Z represents an amino acid other than Ala or Gly.

Ala population	Predominant Structure	R_{-1} – Ala – R_{+1} Group	Representative Sequence	Domain type
1	α-helix	R_{-1} – A – A R_{-1} – A – Z	AAA, **K**AA, AA**K**	CL
2	random coil	R_{-1} – A – G	GAG, AAG, VAG	HP
3	random coil	R_{-1} – A – Z	GAV, GAL, AA**K**	HP + CL
4	random coil	R_{-1} – A – A	AAA, **K**AA, PAA	CL

3.1.2 Selection of ssNMR experiments for structural studies of elastin

To explore different populations of lysine in elastin, various excitation schemes are explored. DP reflects all ^{13}C sites, regardless of conformation and dynamics, and is quantitative. CP selectively detects rigid regions, and **r**efocused **I**nsensitive **N**uclei **E**nhanced by **P**olarization **T**ransfer (rINEPT) enhances signals from mobile domains. These experiments provide insights into the secondary structures of lysine in rigid and mobile sections of the protein.

DP with **s**teady-**s**tate **N**uclear **O**verhauser **E**ffect (ssNOE/DP) is used to obtain a higher S/N in a shorter experimental time than in DP. SsNOE improves the ^{13}C intensities of protonated carbons by a factor of η, according to Equation 3.2. The ^{13}C populations with high mobility are more effectively enhanced by NOE than the rigid ones.

$$1 + \eta = \frac{Intensity\ (with\ {}^{1}H\ irradiation)}{Intensity\ (without\ irradiation)}$$ (Equation 3.1)

In ssNOE/DP, the ^{1}H spin populations are saturated using low-power rf irradiation, before a 90° pulse on the ^{13}C channel (figure 3.1). As the high-energy ^{1}H spins return to equilibrium, the z-magnetization is transferred to ^{13}C spins through the process of cross-relaxation.
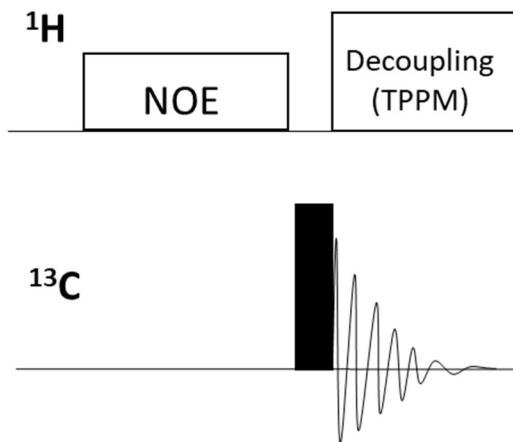
Figure 3.1: Pulse sequence for ^{13}C ssNOE/DP experiment.
Black rectangle represents a 90° pulse.

The rINEPT experiment[11,12] targets mobile regions of hydrated elastin that undergo "liquid-like" motions (figure 3.2).[13] The fast internal dynamics of the hydrated protein attenuate the dipolar couplings. Thus, the CP excitation scheme, which is based on the ^1H-^{13}C dipolar coupling, is not effective in this system. In rINEPT measurements, the magnetization transfer efficiency is governed by the coherence lifetime, T_2'. Mobile systems have longer T_2' values than the rigid ones, resulting in higher efficiency of the rINEPT transfer. Only protonated carbons are detected with this experiment, because polarization is transferred from ^1H to directly bonded ^{13}C nuclei.

In the rINEPT experiment, the initial 90° pulse on the ^1H channel creates transverse magnetization. Relaxation results in dephasing during the first τ-delay at a rate of $e^{-2\tau/T_2'}$. Thus, after 2τ, only the ^1H coherences with long lifetimes survive, i.e., protons in mobile domains, which typically have large-amplitude oscillations and long T_2' values. A 180° pulse is applied to both spins, and offsets are refocused after 2τ, but the J_{CH} coupling continues to evolve during the entire 2τ period. At the end of 2τ, proton magnetization components are anti-phase, and the simultaneous 90° pulses on ^1H and ^{13}C channel cause a polarization transfer. The final spin-echo sequence (τ' - 180° - τ') refocuses the anti-phase magnetization, so that broadband decoupling can be applied during acquisition. Since magnetization of the spins decays according to $e^{-2\tau'/T_2'}$, only the highly mobile ^{13}C spins with long T_2' values are detected.
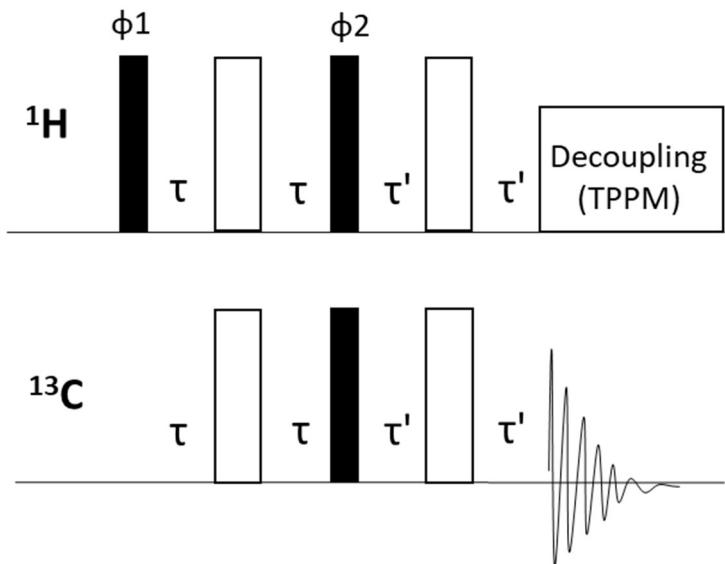
Figure 3.2: Pulse sequence for ^{13}C-{^{1}H} rINEPT experiment.
Black and white rectangles represent 90° and 180° pulses, respectively. The phases were as follows: Φ_1 = {x}; Φ_2 = {y}.

Signal overlap is unavoidable in the 1D NMR of elastin. For example, studies of [ϵ-^{13}C-Lys] elastin confirm a contribution from crosslinked $^{13}C\epsilon$ in the $^{13}C\alpha$ region of the spectrum (Chapter 4). Moreover, scrambling to ^{13}CO-Gly is confirmed with the HPLC-MS assay (Chapter 2) and NMR data (Chapter 4). The ^{13}CO-Gly intensity is not resolved from ^{13}CO-Lys. These overlapping signals are resolved in a two-dimensional (2D) ^{13}CO-$^{13}C\alpha$-Lys correlation spectrum.

Dipolar **A**ssisted **R**otational **R**esonance (DARR) is a popular method of characterization of solid proteins, allowing for ^{13}CO-$^{13}C\alpha$ correlation. This technique is appropriate for samples with adjacent ^{13}C nuclei, which allow for ^{13}C spin diffusion. After the initial CP excitation, a selective, r-SNOB shaped[14] pulse is applied to ^{13}CO with ^{1}H continuous wave (CW) decoupling. Since only ^{13}CO signals are detected in the indirect dimension, the spectral window is narrowed to 164 – 184 ppm, which reduces the total experimental time. A spectrum is recorded for each increment of t_1, until the signal decays. The phase of the r-SNOB pulse is alternated between x and y, and the phase of the receiver shifts between x and -x. This two-step phase cycling detects t_1-modulated signals of ^{13}CO, but not others. A 90° ^{13}C pulse is applied, and DARR mixing follows to obtain ^{13}CO-$^{13}C\alpha$ correlations. Since the natural abundance of ^{13}C is low (~1.1 %), the correlations detected by this pulse sequence originate from ^{13}C - ^{13}C pairs in enriched Lys residues only. A final 90° ^{13}C pulse flips the magnetization for detection, and high-power TPPM decoupling is applied on the ^{1}H channel during acquisition.

51
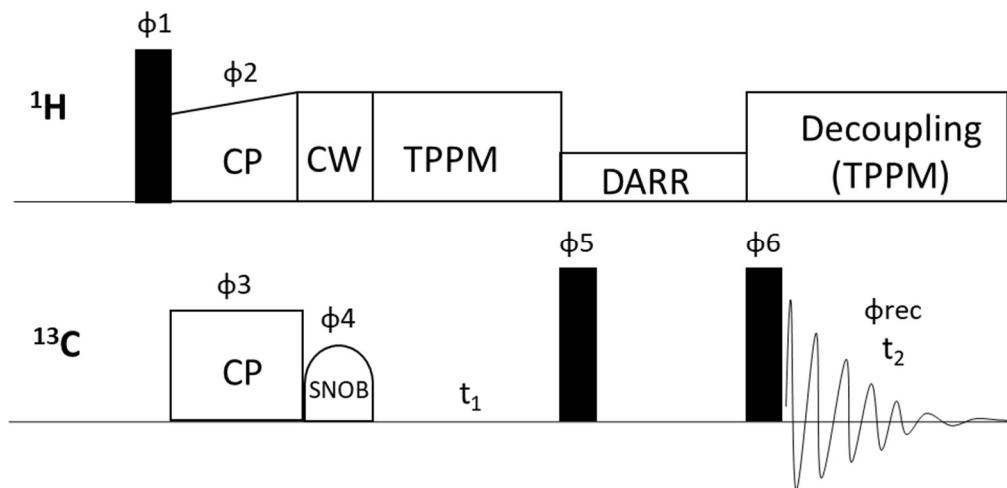
Figure 3.3: Pulse sequence for 2D ^{13}CO-^{13}Cα DARR correlation.[15]
Black rectangle represents a 90° pulse. The phase cycling was as follows: Φ_1 = {y, y, y, y, -y, -y, -y, -y};
Φ_2 = {x}; Φ_3 = {-y}; Φ_4 = {x}8 {y}8; Φ_5 = {x}; Φ_6 = {x, y, -x, -y}; Φrec = {x, y, -x, -y, -x, -y, x, y, -x, -y, x, y, x, y, -x, -y}.

3.1.3 Interpretation of NMR spectra with RefDB data

Due to a range of solvent conditions, reference standards, and measurement protocols, chemical shifts are prone to inconsistent referencing. RefDB is a secondary database developed to address this problem by validating and correcting the ^{1}H, ^{13}C, and ^{15}N data from the **B**iological **M**agnetic **R**esonance **B**ank (BMRB). The SHIFTX program[16] uses X-ray or NMR atomic coordinates to calculate NMR resonance frequencies. These calculated values are then compared to their counterparts in BMRB. A variety of statistical evaluations is performed, and any misassignments, typographical, or chemical referencing errors are corrected.

Chemical shifts of 2162 proteins (October 2019) are categorized as HELIX, SHEET, or COIL by the **V**olume, **A**rea, and **D**ihedral **A**ngle **R**eporter (VADAR) server.[17] The HELIX and SHEET assignments are based on dihedral angles[18], hydrogen bonding patterns[19], and geometrical constraints[20]. If a structure meets criteria for neither of these two conformations (for example, β-turn), it is assigned as COIL.

The RefDB data for lysine residues are plotted in histograms to illustrate the distribution of the chemical shifts in each conformation. These plots are compared to 1D spectra of [U-^{13}C,^{15}N-Lys] elastin, and the secondary structures are assigned using ^{13}CO-Lys and ^{13}Cα-Lys sites. The amino acids that follow Lys, i.e., R_{+1}, are identified in the tropoelastin sequence, and the neighboring residues effects are calculated. The chemical shifts of Lys bonded to a given R_{+1} are then compared to ^{13}CO-^{13}Cα-Lys correlations obtained from a 2D DARR experiment.

3.1.4 Variable-temperature ^{13}C MAS NMR study

Temperature affects both the secondary structure of an amino acid sequence and the dynamics of a protein. The ^{13}C CP signal intensities increase when hydrated BLN elastin is cooled below the freezing point of water.[21] These differences in the lineshapes may be attributed to the *glass transition*, i.e., the reversible change from a molten, rubber-like state to a brittle, glassy state.

The temperature of the glass transition, T_g, depends on the hydration level of elastin.[22] As the motions of the hydration water become restricted, the mobility of the protein is greatly reduced.[23] In the dehydrated state, T_g is 200°C, whereas elastin at 60% hydration has $T_g \sim 0$°C.[24]

Previous VT studies of alanines in the CL domains showed significant reductions in dipolar couplings and an increase in α-helical content below T_g.[25] Generally, fast and large-amplitude motions in warm temperatures promote the formation of short helices. Longer helical chains are favored as the dynamics of the protein is attenuated below T_g. Since the CL domains primarily consist of Ala and Lys residues, similar dependencies are expected in the VT study of [U-^{13}C,^{15}N-Lys] elastin.

3.1.5 Spin-lattice relaxation measurements

Nuclear spins in a static field B_0 show a net alignment of the magnetic moments along the field direction, typically defined as the z-axis. This equilibrium state is perturbed by an rf pulse, which rotates the net magnetization towards the transverse plane. Following a pulse, the bulk z-magnetization re-establishes its equilibrium value through spin-lattice relaxation.

Spin-lattice relaxation is an incoherent process that occurs when spins experience *local fields*; the magnitudes and orientations of these fields are not the same in different parts of the sample. If the local field oscillates at the Larmor frequency and has transverse components, the magnetic moment which experiences that field is rotated to a new direction. Alignment along the z-axis is favored, as it is the low energy arrangement, and the net magnetization eventually returns to its equilibrium value. In NMR of spin-½ nuclei, local fields are predominantly generated by the magnetic moments of other spins, i.e., *the dipolar mechanism*.

Spin-lattice relaxation is characterized by a time constant, T_1, which describes motion in the MHz regime, as opposed to $T_{1\rho}$ that reflects the rotating frame and typically occurs in kHz range. T_1 measurements are used to probe overall and site-specific dynamics in polymers and proteins.[26,27,28] The value of T_1 is also crucial in NMR spectroscopy, because it determines the necessary wait time between data acquisitions for signal averaging.

T_1, can be determined in experiments relying on DP[29], CP[30], and rINEPT[31]. In the original inversion recovery experiment[29], a ^{13}C 180° pulse is followed by an array of delay periods, τ (figure 3.4A). The ^{13}C 90° pulse flips the magnetization to the transverse plane for detection, and an acquisition period with high-power ^1H decoupling follows. To determine T_1 constants, peak intensities are plotted as a function of τ and fitted to the following equation:

$$I(\tau) = M_0 \times \left(1 - 2e^{\left(-\frac{\tau}{T_1}\right)}\right) \qquad\qquad \text{(Equation 3.2)}$$

where $I(\tau)$ is intensity at time τ and M_0 is the z-magnetization at equilibrium.
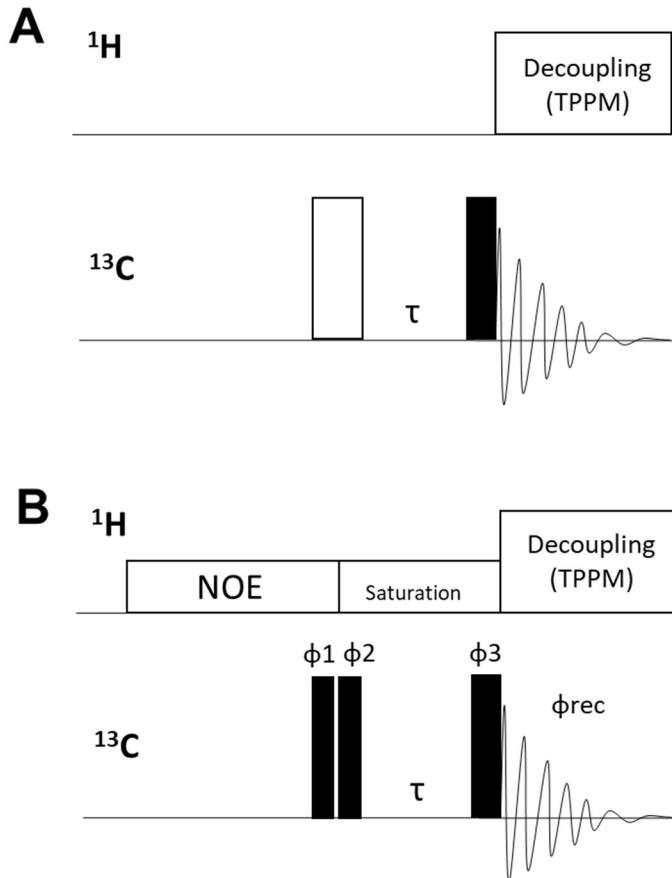
Figure 3.4: Pulse sequence for (A) ^{13}C T_1 inversion recovery experiment and (B) modified inversion recovery with ssNOE and DP excitation.[32]
Black and white rectangles represent 90° and 180° pulses, respectively. Phase cycling in B) is as follows: Φ_1 = {x, x, y, y, -x, -x, -y, -y}; Φ_2 = {x, -x, y, -y, -x, x, -y, y}; Φ_3 = Φ_1; Φrec = {x, -x, y, -y, -x, x, -y, y}.

For determination of T_1 constants in enriched elastin, the inversion recovery sequence is modified with ssNOE and phase alternation (figure 3.4B).[32] The ssNOE enhancement maximizes signal intensities in the shortest experimental times for mobile polymers, such as hydrated elastin. Phase alternation results in single-exponential decay curves that end at zero intensity at $\tau = \infty$. In the odd scans (n = 1, 3, 5...), the first two pulses flip the magnetization from +z-direction to the -z-direction, i.e., they are equivalent to the 180° pulse in the classic inversion recovery experiment. In the even scans (n = 2, 4, 6...), however, they are equivalent to a 0° pulse. Since the receiver phase is 180° in the even scans, intensities $I^{odd}(\tau)$ are subtracted from $I^{even}(\tau)$ during the experiment. Therefore, the signal, $I(\tau)$, over two scans is described by Equation 3.3:

$$I(\tau) = I(\infty) \times (1 - \alpha) \times e^{\left(\frac{-\tau}{T_1}\right)}$$
<div align="right">Equation 3.3</div>

where $\alpha = e^{\frac{-d_1}{T_1}}$, d_1 stands for recycle delay, and $I(\infty)$ is the equilibrium magnetization at $\tau = \infty$ under ^{1}H saturation.

Unlike in the traditional inversion recovery experiment, positive intensities are obtained for all signals during the decays. Moreover, full recovery of magnetization is not necessary, as $I(\infty) = 0$, not M_0, i.e., the resulting single-exponential decay relaxation curve does not depend on the recycle delay, which can be shorter than $5 \times T_1$. Fitting model $A \times e^{\frac{-\tau}{T_1}}$ is used where A is the amplitude of the exponential function.

3.2 Materials and methods

Solid-state NMR spectroscopy

About 60 mg of hydrated elastin samples, prepared as described in Section 2.5, were blotted on a KimWipe and packed in a 4.0 mm rotor. Kel-F top and bottom spacers (Revolution NMR, Fort Collins CO) fitted with fluorosilicone micro o-rings (Apple Rubber Products, Lancaster NY) were used to seal the rotor and maintain hydration levels. No loss of water was ensured by weighing the rotors before and after each experiment. The packed rotors were stored in a freezer. The [ε-^{13}C-Lys] elastin sample was not used for any of the experiments discussed in this chapter.

The Agilent DD2 NMR console (Agilent Technologies, Santa Clara, CA) with 9.4 T wide-bore (89 mm) superconducting magnet (Oxford Instruments, Oxford, UK) was used for solid-state NMR measurements. The ^{1}H Larmor frequency was 399.935 MHz, and a 4 mm triple-resonance (HXY) T3 magic angle spinning probe (Chemagnetics/Varian NMR, Ft. Collins, CO) was used. The temperature of the sample was calibrated using Pb(NO$_3$)$_2$[33] under MAS with 8 kHz spinning rate. The 8 kHz MAS rate was used in all experiments, unless stated otherwise. Sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) in D$_2$O [δ(^{1}H) = 0.0 ppm at 25°C] was used as an external reference. The tetramethylsilane (TMS) scale was applied to reference ^{13}C chemical shifts with hexamethylbenzene serving as an external standard [δ(^{13}C) = 17.0 ppm at 25°C].

For ^{13}C DP experiments performed on [U-^{13}C,^{15}N-Lys] and unenriched elastin samples, 4.5 μs ^{13}C 90° pulse was used with a 10 s recycle delay. A Hahn echo (τ_{echo}-π-τ_{echo}) block was utilized with τ_{echo} = 125 μs. To remove dipolar (^{1}H-^{13}C) interactions, two-pulse phase modulation (TPPM)[34] ^{1}H decoupling was applied for 30 ms with an applied field strength of $\gamma^H B_1{}^H/2\pi \sim 50$ kHz.

In the variable-temperature study, [U-^{13}C,^{15}N-Lys] elastin was observed at 37°C, 20°C, 10°C, and 0°C via ssNOE/DP (figure 3.1). An initial 4.7 µs ^{13}C 90° pulse was followed by acquisition under TPPM ^{1}H decoupling.[34] The ssNOE enhancement was achieved by applying ^{1}H CW irradiation ($\gamma^{H}B_{1}^{H}/2\pi \sim 1$ kHz) throughout the 1.5 s recycle delay. Each spectrum is a result of 4096 scans.

^{13}C CP experiments were performed on [U-^{13}C,^{15}N-Lys] elastin and unenriched elastin samples at 37°C and -20°C. A 4.7-µs ^{1}H 90° pulse was followed by a 1 ms contact time with recycle delay of 5 s. The B$_1$ field strengths for CP were set to $\gamma^{C}B_{1}^{C}/2\pi = 50$ kHz and $\gamma^{H}B_{1}^{H}/2\pi = 42$ kHz, so that the Hartmann-Hahn sideband condition is met and $\gamma^{C}B_{1}^{C}/2\pi = \gamma^{H}B_{1}^{H}/2\pi + n^{*}\nu_{r}$, where ν_{r} is the spinning rate of the rotor, and n = 1. CP field strength for ^{1}H was ramped from 95% to 105%. A Hahn echo (τ_{echo}-π-τ_{echo}) block was utilized for CP with $\tau_{echo} = 125$ µs. TPPM ^{1}H decoupling was applied during the 20 ms acquisition time with an applied field strength of $\gamma^{H}B_{1}^{H}/2\pi \sim 50$ kHz.

In the ^{13}C-{^{1}H} rINEPT, 90° pulse lengths for ^{1}H and ^{13}C were 5.0 µs and 4.6 µs, respectively (figure 3.2). The recycle delay was set to 1.5 s. The initial pulse is followed by a spin-echo sequence. Typically, the length of τ is in the order of $1/{|4 * J_{CH}|}$. Delays for the first (- τ - 180° - τ -) and second (- τ' - 180° - τ' -) spin-echo blocks were set to $\tau = 1.25$ ms and $\tau' = 1.0$ ms, respectively. TPPM ^{1}H decoupling was used during the 30 ms acquisition time with an applied field strength of $\gamma^{H}B_{1}^{H}/2\pi \sim 50$ kHz.

In 2D ^{13}CO-^{13}Cα DARR, [U-^{13}C,^{15}N-Lys] elastin was frozen at -20°C and spun at this temperature for about 40 min for equilibration (figure 3.3). An initial 4.2-µs ^{1}H 90° pulse was followed by a 1 ms contact time. The CP field strengths were set to $\gamma^{C}B_{1}^{C}/2\pi = 50$ kHz and $\gamma^{H}B_{1}^{H}/2\pi = 42$ kHz, so that the Hartmann-Hahn sideband condition is met and $\gamma^{C}B_{1}^{C}/2\pi = \gamma^{H}B_{1}^{H}/2\pi + n^{*}\nu_{r}$, where n = 1. The CP field strength for ^{1}H was ramped from 95% to 105%. For the ^{13}CO-selective 180° pulse, **r**efocused **S**elective excitatio**N** f**O**r **B**iochemical applications (r-SNOB)[35] pulse (pulse width: 1063 µs = $8.5^{*}\tau_{r}$)[36] was applied with ^{1}H CW decoupling ($\gamma^{H}B_{1}^{H}/2\pi = 63$ kHz). The 5-ms DARR mixing time was applied to obtain mainly directly-bound ^{13}CO-^{13}Cα correlations. Two 4.5-µs ^{13}C 90° pulses were applied at the beginning and end of the mixing time. TPPM ^{1}H decoupling was used during the 30 ms acquisition time with applied field strength $\gamma^{H}B_{1}^{H}/2\pi \sim 50$ kHz. Quadrature detection in the t_1 dimension was achieved by using States-TPPI with phase shifts of the ^{13}C CP contact pulse and the receiver. The spectral width in the indirect dimension was set to 2000 Hz (= $\frac{1}{4} \times \nu_{r}$) with the transmitter frequency at 174 ppm, for a spectral window of 164 ppm – 184 ppm. Nine complex points were collected in the t_1 dimension corresponding to $t_{1max} = 4$ ms. Two hundred fifty-six scans were accumulated per t$_1$ point. A 3.0-s recycle delay was used, and the total experimental time was 4 hours.

^{13}C T_1 modified inversion recovery with ssNOE/DP begins with two ^{13}C 90° pulses, each 4.7 μs long (figure 3.4B). In odd scans (scan 1, 3, etc.), these two pulses are equivalent to a 180° pulse, and the z-magnetization is flipped to the (-z)-direction. In the even scans (scan 2, 4, etc.), the effects on the magnetization are equivalent to a 0° pulse, i.e., magnetization is stored along the (+z)-direction. The lengths of delay time, τ, that follows were [0, 0.3, 0.5, 0.9, 1.5, 2.0, 3.0, 4.0, 5.0] s. A 4.7 μs ^{13}C 90° read pulse is applied to tip the spins into the transverse plane, to allow for their detection. The 180° receiver phase is on even scans results in subtraction of the (2*n)th scan from the (2*n+1)th scan. TPPM ^1H decoupling (strength $\gamma^H B_1^H/2\pi \sim 50$ kHz) takes place during the 30 ms acquisition time. ^1H low-level CW irradiation ($\gamma^H B_1^H/2\pi \sim 1$kHz) is applied for the length of the recycle delay (3 s).

Analysis of RefDB data

RefDB[38] is a secondary database that contains re-referenced ^1H, ^{13}C, and ^{15}N chemical shift data from the **B**iological **M**agnetic **R**esonance **B**ank (BMRB). Briefly, the **V**olume, **A**rea, and **D**ihedral **A**ngle **R**eporter (VADAR) server[17] assigns HELIX and SHEET secondary structures to atoms in 2,162 proteins in RefDB, according to dihedral angles[18], hydrogen bonding patterns[19], and geometrical constraints[20]. Structures that do not match the requirements for HELIX or SHEET are assigned as COIL.

RefDB ^{13}C chemical shift data for ^{13}CO- and $^{13}C\alpha$-Lys are sorted and plotted in histograms. Each interval (or bin) spans 0.2 ppm. The height of each bar corresponding to a given bin represents the relative number of observations, normalized over the total number of data points in a given histogram. HELIX histograms are based on 3,315 (^{13}CO-Lys) and 4,536 ($^{13}C\alpha$-Lys) data points. COIL histograms include 3,082 (^{13}CO-Lys) and 4,510 ($^{13}C\alpha$-Lys) data points. SHEET histograms include 2,191 occurrences of ^{13}CO-Lys and 3,066 of $^{13}C\alpha$-Lys.

Intra-residue ^{13}CO-$^{13}C\alpha$ correlations for lysine in HELIX, COIL, and SHEET were based on 2869, 2573, and 1859 available data points in RefDB, respectively. Gaussian function was created for each ^{13}CO-$^{13}C\alpha$ correlation, and the final surface results from addition of these functions. The same exponential scale is used for the 2D DARR contour and the RefDB correlations. In both, lines represent 20, 24, 29, 35, 41, 50, 60, 72, and 86% (going from innermost to outermost), where 100% represents the total enclosed volume.

Adjustment for neighboring residue effects

Chemical shifts accounting for secondary structure and neighboring residue effects are calculated for lysine in elastin. Briefly, ^{13}C chemical shifts of Lys in random coil, δ_{coil}(Lys), is obtained from a study of a model peptide, Ac-GGKGG-NH$_2$, in denaturing conditions.[37] The correction factor for each amino acid[4], $\Delta\delta(R_{+1})$, is then added to calculate the chemical shift of ***coil*** that accounts for the neighboring residue effects (NRE), i.e., $\delta_{coil,NRE}$(Lys) = δ_{coil}(Lys) + $\Delta\delta(R_{+1})$. Similarly, $\delta_{helix,NRE}$(Lys) represents the chemical

shifts of lysine in α-helix adjusted for NRE. **Helix** is calculated from the values for **coil** and secondary shifts from RefDB[38] as follows: $\delta_{helix,NRE}(^{13}CO\text{-Lys}) = \delta_{coil,NRE}(^{13}CO\text{-Lys}) + 2.06$ ppm and $\delta_{helix,NRE}(^{13}C\alpha\text{-Lys}) = \delta_{coil,NRE}(^{13}C\alpha\text{-Lys}) + 2.34$ ppm. The obtained $^{13}CO\text{-}^{13}C\alpha$ correlations for each KR_{+1} pair are plotted, using Gaussian functions to reflect the relative number of occurrences in the NRSMC tropoelastin monomer. The volume under each crosspeak is used as a reference (100%) for the contour lines that follow exponential scale and represent 10, 20, 40, and 80% of the volume (going from innermost to outermost).

3.3 Results and discussion

Solid-state ^{13}C NMR measurements of [U-^{13}C,^{15}N-Lys] elastin give insights into conformations and dynamics of lysine residues in the hydrated protein. Secondary structures are tentatively assigned, based on chemical shifts in RefDB. The neighboring residue effects on the isotropic chemical shifts are also used to assign the crosspeaks in a 2D ^{13}CO - $^{13}C\alpha$-Lys correlation experiment. A modified inversion recovery measurement determines the ^{13}C spin-lattice relaxation constants for the labeled sites in lysines and the crosslinks.

3.3.1 One-dimensional ^{13}C MAS NMR spectra

The targeted ^{13}C sites are observed in a DP spectrum of [U-^{13}C,^{15}N-Lys] elastin (figure 3.5A). However, lysine is a minor residue in elastin; it constitutes about 5% of the amino acids in NRSMC tropoelastin[39] and less than 1% in the mature, crosslinked protein. Even with isotopic enrichment approaching 90%, the signals of labeled nuclei are obscured by the remaining amino acids.

The natural-abundance ^{13}C signals (figure 3.5B) are subtracted from the spectrum of the enriched protein (figure 3.5A) to produce a difference spectrum (figure 3.5C), reflecting the destinations of the incorporated ^{13}C labels. All one-dimensional spectra presented in this chapter are processed in this way, unless stated otherwise. The scaling factors of the enriched and unenriched spectra are chosen to cancel the methyl signals (10 – 20 ppm) of unenriched amino acids in the difference spectrum. Neither lysine nor any of the crosslinks have methyl groups. Moreover, the identified scrambling pathways lead to enrichment of carbonyl of other amino acids, not the aliphatic carbons (Chapter 2). Therefore, no ^{13}C labels are expected in CH_3.
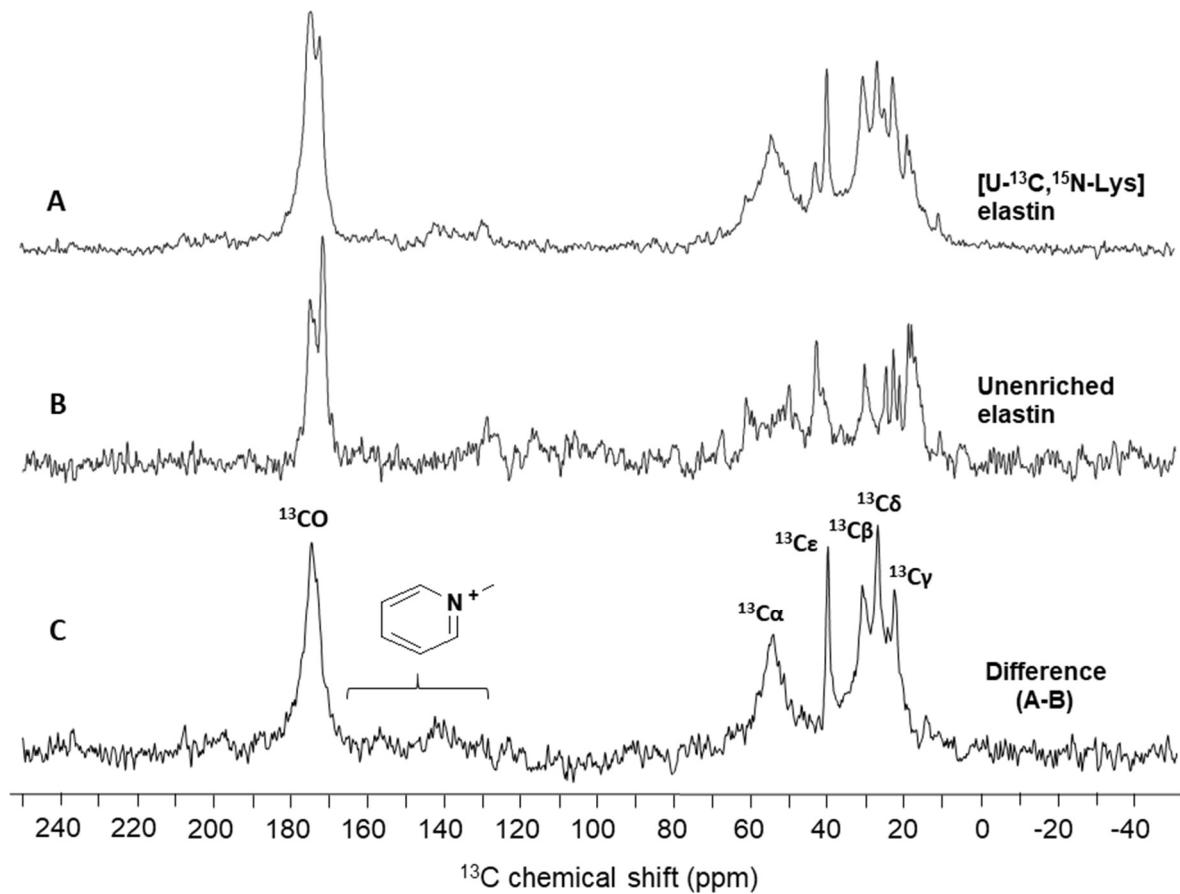
Figure 3.5: ^{13}C DP/MAS NMR spectra of labeled and unlabeled elastin at 37°C.
(A) Spectrum of [U-^{13}C,^{15}N-Lys] elastin, (B) spectrum of unenriched elastin, and (C) difference spectrum
(A - B), displaying signals originating from [U-^{13}C,^{15}N-Lys]. Spectra (A) and (B) were each acquired with a
recycle delay of 10 s and 1024 scans, then processed with 40 Hz line broadening. Each spectrum is
scaled to its tallest peak. The resolved intensities in the difference spectrum (C) are assigned to the sites
in lysine and the crosslinks.

The peaks in the ^{13}C DP/MAS NMR difference spectrum of [U-^{13}C,^{15}N-Lys] elastin at 37°C (figures 3.5C and 3.6A) are assigned to the sites in Lys, as follows. The carbonyl feature has its highest intensity at 174.6 ppm. A weak signal is observed in the 140 - 160 ppm range, which is expected for the pyridinium ring of desmosine[40], isodesmosine[41], and other heteroaromatic carbons. The peak at 54.5 ppm is assigned as ^{13}Cα-Lys. The characteristic, narrow intensity at 40.1 ppm corresponds to ^{13}Cε of unmodified lysine. Its lineshape is attributed to high mobility at the terminal carbon of unmodified lysine. The remaining aliphatic peaks at 30.9, 27.1, and 22.9 ppm are assigned to ^{13}Cβ-, ^{13}Cδ-, and ^{13}Cγ-Lys, respectively.

Protonated carbons in mobile domains are enhanced by rINEPT (figure 3.6B). No carbonyl signal is observed, because polarization transfer only occurs at carbons that are bonded to protons. In comparison to DP and CP spectra (figure 3.6A-B), the sharp ^{13}Cε-Lys peak is much taller than the other aliphatic signals in rINEPT, confirming high mobility at the end of the lysine sidechain. The ^{13}Cα-Lys signal is weak, which is attributed to the rigid nature of the backbone carbons.

CP selectively detects rigid regions. In the spectrum obtained at 37°C (figure 3.6C), the carbonyl feature is shorter than in the DP spectrum and has height comparable to ^{13}Cα-Lys. The peak assigned as ^{13}Cβ-Lys has lower intensity than in DP, suggesting this site does not have as many rigid populations as other aliphatic carbons. The lineshapes of the remaining aliphatic carbons resemble those in DP.

Due to rigidification of the protein, signal intensities greatly improve for a frozen sample of hydrated [U-^{13}C,^{15}N-Lys] elastin. The lineshapes in the CP spectrum recorded at -20°C are different from those in experiments conducted at the physiological temperature. All peaks are broader; ^{13}Cβ-, ^{13}Cδ-, and ^{13}Cγ-Lys are not resolved. ^{13}Cε-Lys signal is weak, confirming that the lysine sidechain loses mobility upon freezing. The broad, asymmetric ^{13}Cα-Lys peak appears to consist of at least two overlapping populations. The signals in the heteroaromatic region of the spectrum, assigned as ^{13}C sites in the rigid pyridinium ring, are clearly distinguished from noise. The backbone carbonyl feature is broader and taller than in CP at 37°C. The assignment of CP spectra at -20ºC are discussed in detail in Chapter 4.
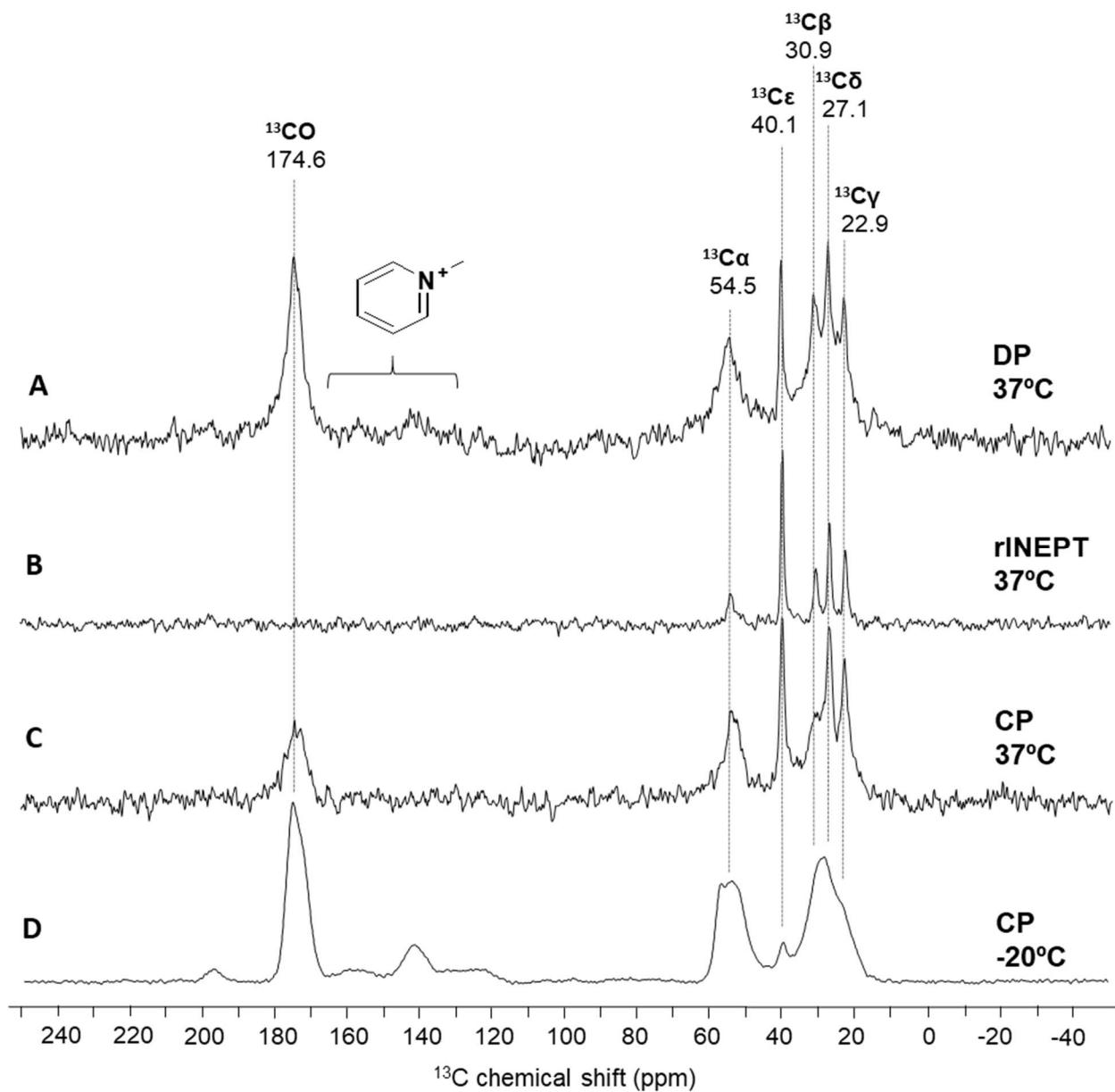
Figure 3.6: ^{13}C difference spectra of [U-^{13}C,^{15}N-Lys] elastin.
(A) DP at 37°C acquired with recycle delay of 10 s and 1024 scans (figure 3.5C), (B) rINEPT measurement at 37°C with recycle delay of 1.5 s and 6144 scans, (C) CP at 37°C obtained with recycle delay of 5 s and 6144 scans, and (D) CP TOSS acquired at -20°C with recycle delay of 5 s, and 2048 scans. The acquired spectra were processed with 40 Hz line broadening, and each was scaled to its tallest peak. The dashed lines correspond to the highest position of each feature in the spectrum (A).

3.3.2 Secondary structure assignments in 1D ^{13}C NMR spectra based on RefDB data

Solid-state NMR is a powerful and versatile technique that provides valuable information on the secondary structure of proteins. One-dimensional ^{13}C NMR spectra are compared against the data in RefDB to predict the conformations of lysine in elastin at -20°C and 37°C. The results of the analysis are later used in the analysis of two-dimensional spectra.

The chemical shifts of ^{13}CO-Lys and ^{13}Cα-Lys in RefDB are plotted in histograms (figure 3.7). Width of each interval (or bin) corresponds to a 0.2 ppm range. The height of each bar corresponds to the relative number of observations of each bin, normalized over the total number of occurrences in a given structure (HELIX, SHEET, or COIL). Each histogram was based on 2,000 – 4,500 data points with assigned secondary structures in RefDB (details are provided in Section 3.2). The bins with the highest occurrences in the histograms for ^{13}CO-Lys and ^{13}Cα-Lys in HELIX (176.6 and 57.2 ppm, respectively) are downfield from those of the SHEET (173.0 and 53.0 ppm) and COIL (174.2 and 54.2 ppm). The COIL histograms are generally broader and overlap significantly with those for SHEET. However, HELIX is resolved from the other two conformations.
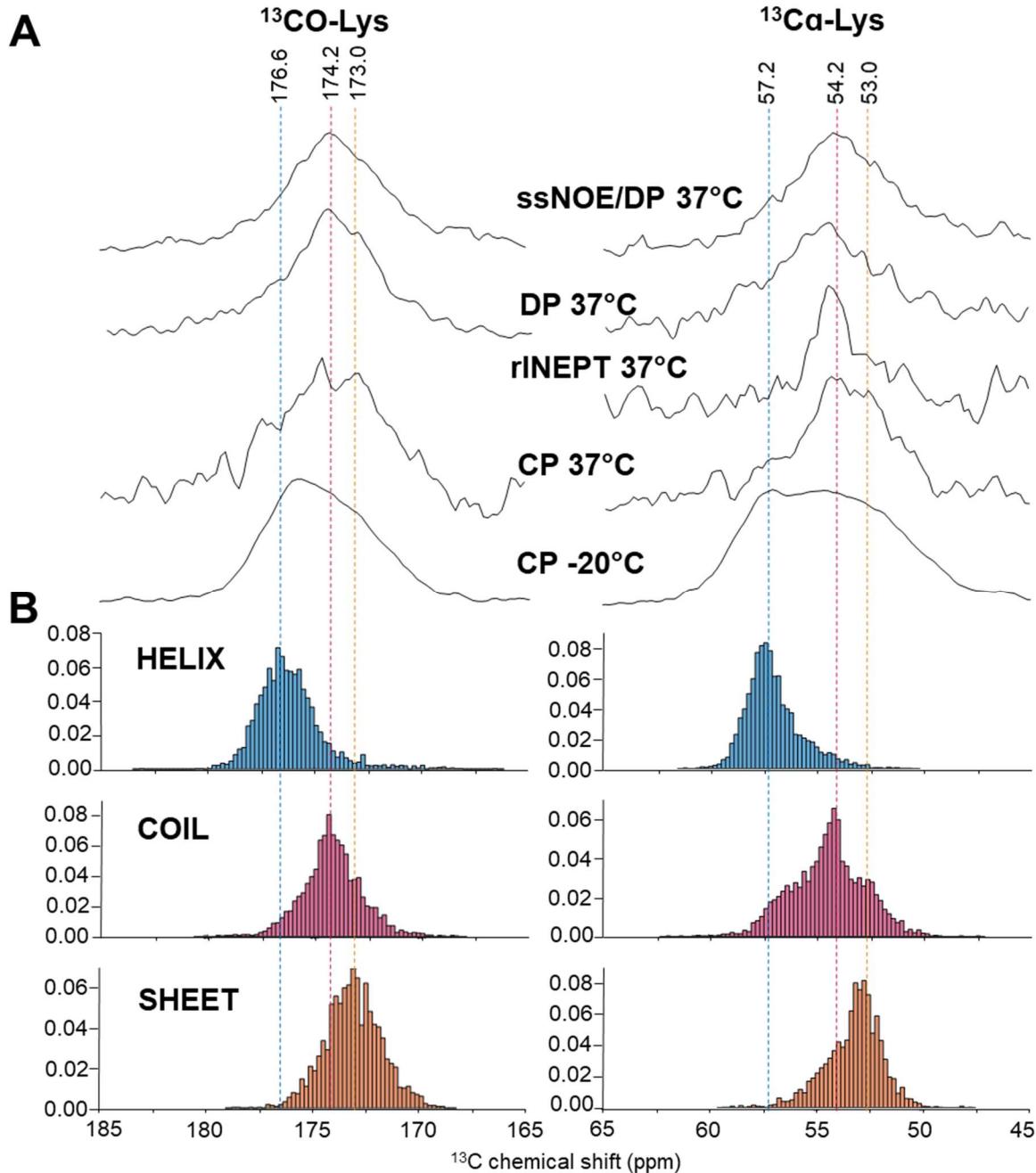
Figure 3.7: Chemical shifts of ^{13}CO-Lys and ^{13}Cα-Lys in [U-^{13}C,^{15}N-Lys] elastin spectra and RefDB database.

(A) ^{13}C difference spectra of [U-^{13}C,^{15}N-Lys] elastin acquired at 37°C with ssNOE/DP, DP, rINEPT, and CP (37°C and -20°C). (B) Distributions of ^{13}C chemical shifts of ^{13}CO-Lys (left) and ^{13}Cα-Lys (right) categorized as HELIX, COIL, and SHEET in RefDB. ^{13}C-Lys resonances are sorted and categorized into intervals (or bins) spanning 0.2 ppm. The height of each bar reflects the relative number of observations of a given bin, normalized over the total number of occurrences. The sum of the bar heights in each histogram is 1 (or 100%). Dashed lines correspond to the highest number of occurrences in each histogram for HELIX (blue), COIL (pink), and SHEET (orange).

64

The DP experiment at 37°C reveals ^{13}CO- and ^{13}Cα-Lys peaks with the highest intensities at 174.6 and 54.6 ppm, respectively. A sharp peak at 54.6 ppm is observed in rINEPT. The signals in CP at 37°C are more upfield; a broad ^{13}CO-Lys intensity is centered at ~174 ppm and has a shoulder at ~177 ppm. ^{13}Cα-Lys is centered at ~ 53.5 ppm and has a downfield shoulder at ~57 ppm. At 37°C, weak intensities consistent with HELIX are observed in ssNOE/DP, DP, and CP at 37°C, but not rINEPT. Spectra recorded at this temperature are more consistent with the upfield chemical shifts of SHEET (173.0 and 53.0 ppm) and COIL (174.2 and 54.2 ppm) than the downfield HELIX (176.6 and 57.2 ppm).

COIL is the most likely predominant structure at 37°C. SHEET and COIL cannot be distinguished, based on the ^{13}C chemical shifts from 1D experiments, as indicated by the histograms. However, alanines in crosslinking domains are mostly found in HELIX with contributions from COIL populations.[7] Given Lys in the CL domains is typically surrounded by Ala residues, some consistency in the secondary structures of these two amino acids (KA) is expected. And, as no alanine population in CL domain was found in SHEET, it is unlikely for lysine to take that conformation.

The CP lineshapes are significantly different at -20°C compared to those at 37°C. At -20°C, the ^{13}CO-Lys feature is centered at ~ 175 ppm, and ^{13}Cα-Lys is a broad peak centered at ~54.5 ppm with a distinct shoulder at ~57 ppm. The broad ^{13}CO-Lys and ^{13}Cα-Lys features overlap with all three histograms (HELIX, COIL, and SHEET), indicating more than one population of lysine. The downfield signals are aligned with HELIX, implying an increase in the helical content as the sample is frozen. HELIX is the dominant secondary structure only in CP spectra at -20°C.

In summary, COIL appears to be dominant at the physiological temperature. Weak signals consistent with HELIX are also detected. According to the CP experiments at -20°C and 37°C, the helical content is higher at the negative temperature. SHEET and COIL chemical shifts in RefDB are not resolved, but the previous studies of alanines in elastin make SHEET an unlikely candidate.

3.3.3 Secondary structures in 2D ^{13}CO-^{13}Cα DARR correlation

The multiple populations observed in 1D spectra of [U-^{13}C,^{15}N-Lys] elastin (figure 3.7) are resolved with a 2D DARR experiment. Two-dimensional data allows for analysis of secondary structures and neighboring residue effects. The observed signals originate from ^{13}CO-^{13}Cα pairs in the enriched lysine residues only. Thus, scrambling to ^{13}CO-Gly is not observed in the 2D DARR spectra.

The ^{13}CO-$^{13}C\alpha$ DARR spectrum reveals at least three populations of lysine (figure 3.8A). Skyline projections reflect the tallest points of the contour crosspeaks in the spectrum. The dashed lines indicate the peak positions of the major crosspeaks at $(\delta(^{13}CO), \delta(^{13}C\alpha))$ = (176.0, 57.1), (174.8, 54.5), and (172.7, 52.3) ppm, which are denoted **I**, **II**, and **III**, respectively. Populations **I** and **II** are clearly resolved, but **III** is relatively broad.

To assign secondary structures, the 2D spectrum is compared to the correlated ^{13}CO-$^{13}C\alpha$ chemical shifts for HELIX (blue), COIL (pink), and SHEET (orange) from RefDB (figure 3.8B). The data used for the 1D analysis (figure 3.7) is represented in form of a 2D plot. HELIX is well resolved, but COIL and SHEET overlap, with the latter being a broad distribution. The major population in the 2D spectrum (**I**) matches HELIX (blue). Population (**II**) overlaps with the RefDB correlations for a COIL (pink). The broad crosspeak (**III**) is consistent with the unresolved SHEET (orange) and COIL (pink).

Chemical shifts of lysine that account for both sequence- and structure-dependent effects are predicted with a semi-empirical approach. The chemical shift of lysine in random coil is obtained from a model peptide in denaturing conditions.[37] This base value is then adjusted with the correction factor for each R_{+1}.[4] The resulting chemical shifts for **coil** account for both conformation and neighboring residues effects. **Helix** is calculated from the **coil** values, by adding the secondary shifts from RefDB (+2.06 ppm for ^{13}CO and +2.34 ppm for $^{13}C\alpha$).[38]

The predictions for **helix** and **coil** are plotted in a 2D contour plot with a weighed Gaussian function that reflects for the number of occurrences of each KR_{+1} pair in the TE sequence (figure 3.8C and D). The most common amino acid following lysine in the NRSMC TE sequence is alanine (R_{+1} = A); KA is found 19 times. The predicted crosspeaks of KX (where X = Y, V, L, S, or R) are not resolved from those of KA. However, KP occurs 5 times in the TE monomer and is resolved from KA and KX by 2 ppm (both ^{13}CO-Lys and $^{13}C\alpha$-Lys) in each secondary structure.

Predicted chemical shifts of Lys in **helix** and **coil** match the populations observed in the 2D DARR spectrum (figure 3.8C). Population (**I**) in the experimental spectrum overlaps with KA and KX in **helix**. Crosspeak (**II**) matches KA and KX in **coil** as well as KP in **helix**. KP in **coil** is clearly resolved from the remaining predictions and is consistent with population (**III**).

Coil is the most likely conformation of KP pairs. Even though KP in *helix* overlaps with population (**II**), proline has high helical penalty (table 3.1). This amino acid does not have a protonated nitrogen, which is needed for hydrogen bonding in α-helix. Moreover, the predicted chemical shifts of both ^{13}CO-Lys and ^{13}Cα-Lys in *sheet* are in the upfield regions of the spectrum, i.e. a crosspeak at (171, 52) ppm, and no such signal was detected.

Overall, the presence of HELIX is unambiguous. This finding is consistent with the CP results for the frozen sample (figure 3.7) and the previous studies of alanines.[7] Signals matching COIL are also observed. No signal for KP in *sheet* confirms the assignment of COIL as a predominant secondary structure in 1D experiments (figure 3.7).
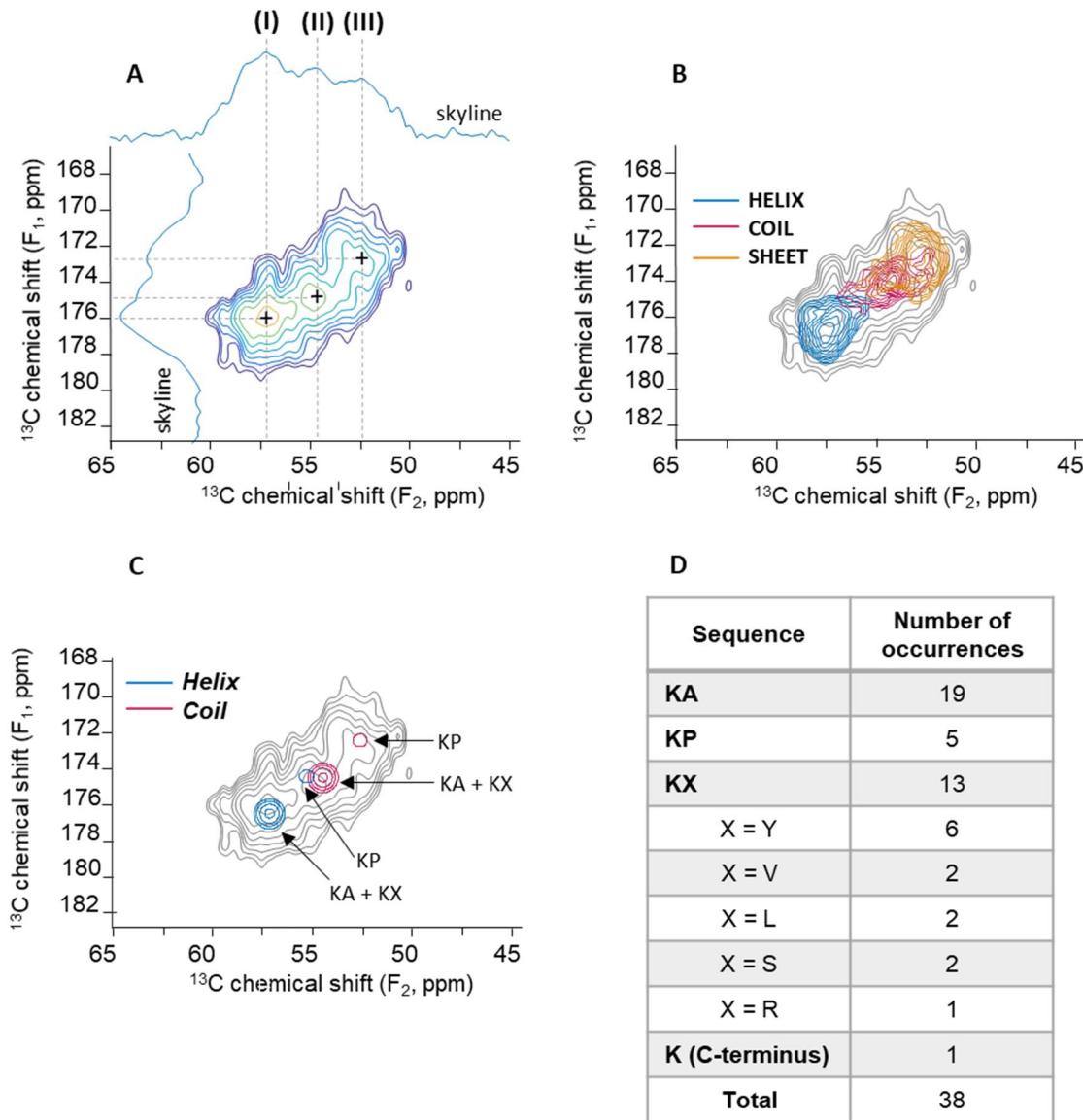
Figure 3.8: 2D selective ^{13}CO-^{13}Cα DARR spectrum of [U-^{13}C,^{15}N-Lys] elastin at -20°C and predicted chemical shifts of lysine.

(A) ^{13}CO-^{13}Cα correlations of Lys in [U-^{13}C,^{15}N-Lys] elastin. Skyline projections of both dimensions are shown. The three major crosspeaks are denoted (**I**), (**II**), and (**III**); dashed lines mark their positions.

(B) Observed spectrum (A, gray contours) with plots of ^{13}CO-^{13}Cα correlations for HELIX (blue), COIL (pink), and SHEET (orange) from RefDB.

(C) Observed spectrum (A, gray contours) with predicted ^{13}CO-^{13}Cα correlations. *Coil* (pink) is calculated from the reported neighboring residue effects. *Helix* (blue) is calculated from *coil* with the secondary shifts from RefDB: +2.06 ppm for ^{13}CO and +2.34 ppm for ^{13}Cα. The 2D plots are generated with Gaussian functions and reflect the relative number of occurrences of KR$_{+1}$ pairs in the tropoelastin sequence.

(D) Two-amino-acid sequences (Lys – R$_{+1}$) in NRSMC tropoelastin.[39]

68

3.3.4 Variable-temperature study

SsNOE/DP and CP are used in the variable-temperature study of [U-^{13}C,^{15}N-Lys] elastin at positive and negative temperatures, respectively (figure 3.9). The two different excitation schemes are chosen due to low intensities of CP at 37°C (figure 3.6) and broad lines in DP at negative temperatures.[25] Data for unenriched elastin were not recorded to reduce the experimental time. Therefore, the presented spectra reflect all populations of ^{13}C in [U-^{13}C,^{15}N-Lys] elastin and are not difference spectra.

In ssNOE/DP (37°C to 0°C), peaks of unenriched amino acids are observed in addition to ^{13}C-Lys signals. The asymmetric lineshape of ^{13}CO is attributed to the contribution from ^{13}CO-Gly, which at 37°C has a peak at 172.0 ppm.[42] Overlapping peaks between 14 and 20 ppm correspond to methyl carbons of alanine, valine, and leucine. Nonetheless, intensities originating from labeled lysine are sufficiently resolved to conduct a variable-temperature study.
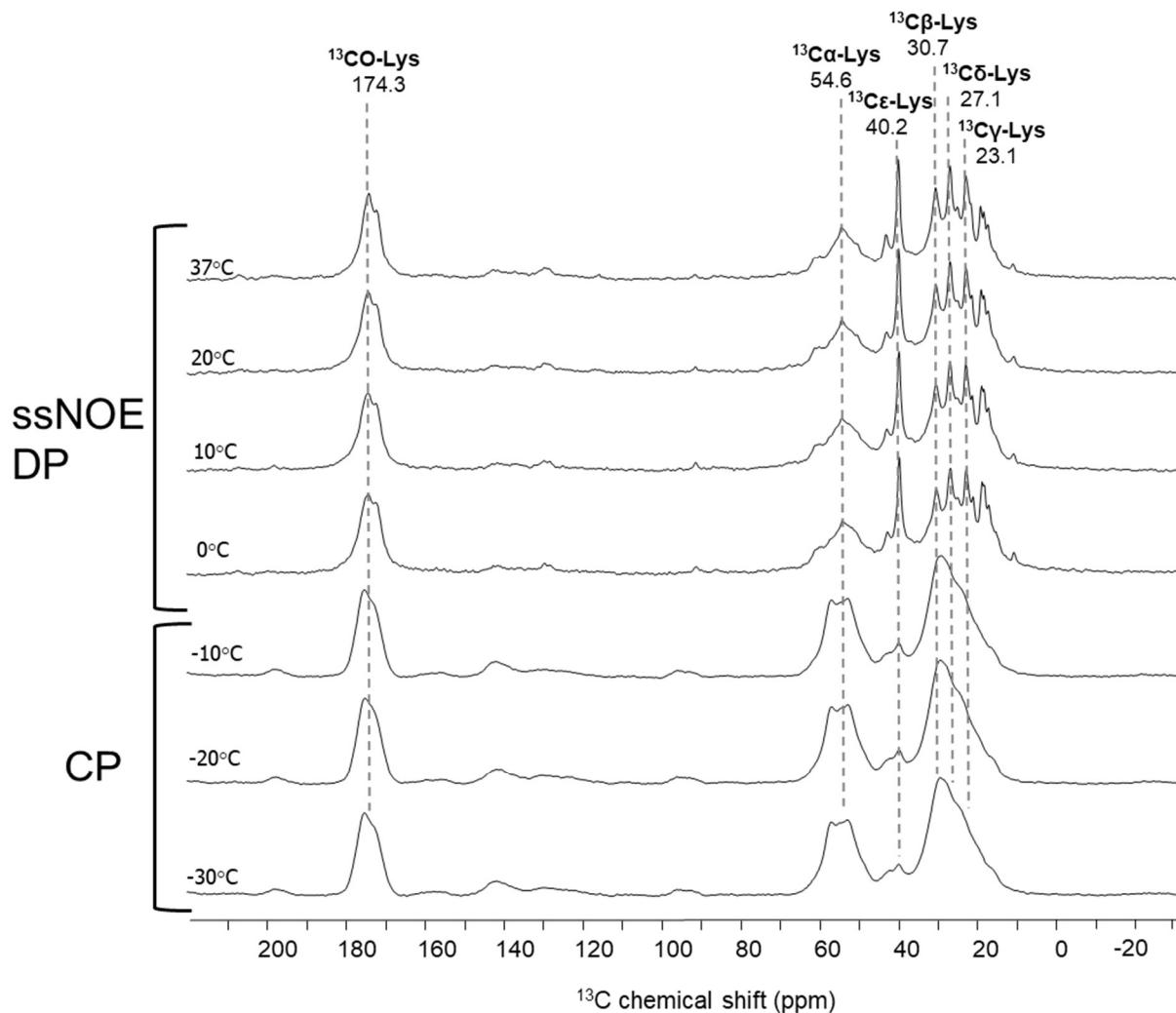
Figure 3.9: Variable-temperature ^{13}C NMR spectra of [U-^{13}C,^{15}N-Lys] elastin.
SsNOE/DP and CP experiments acquired 4096 and 1024 scans, respectively. The dashed lines correspond to the highest position of each feature at 37°C. Each spectrum was processed with 40 Hz line broadening and scaled to the ^{13}CO-Lys peak.

The changes in chemical shifts of ^{13}CO-Lys and the aliphatic carbons at positive temperatures are negligible (table 3.4). The biggest change in the chemical shift occurs for ^{13}CO (0.3 ppm), ^{13}Cε (0.3 ppm), and ^{13}Cα (0.2 ppm). The least affected sites are ^{13}Cδ (0.1 ppm) and ^{13}Cγ (0.1 ppm).

Table 3.4: ^{13}C chemical shifts of lysine in [U-^{13}C,^{15}N-Lys] elastin above 0˚C.
Chemical shift of the tallest peak position for ^{13}C ssNOE/DP spectra (figure 3.9) are given. The row designated as |Δ| shows the absolute difference in chemical shift between 37°C and 0°C for each site.

δ (^{13}C) (ppm)

Temperature	^{13}CO	^{13}Cα	^{13}Cβ	^{13}Cγ	^{13}Cδ	^{13}Cε
37°C	174.3	54.6	30.7	23.1	27.1	40.2
20°C	174.5	54.5	30.7	23.0	27.0	40.1
10°C	174.6	54.5	30.6	23.0	27.0	40.0
0°C	174.6	54.4	30.5	23.0	27.0	39.9
|Δ| (37°C - 0°C)	0.3	0.2	0.2	0.1	0.1	0.3

Studies of poly-L-lysine ([Lys]$_n$) in helix and coil are an additional reference for resonances in different secondary structures (table 3.5).[43] The chemical shift of ^{13}CO is the least affected by the conformation; the difference between helix and coil is only 0.3 ppm. The chemical shift of ^{13}Cδ is the most sensitive to conformation, having |Δ| > 4 ppm. The remaining sites differ by 1 – 2 ppm.

Table 3.5: ^{13}C chemical shifts of [Lys]$_n$.[43]
^{13}C solution NMR chemical shifts of [Lys]$_n$ in D$_2$O are shown. The row designated |Δ| (helix - coil) is the difference in the chemical shift of the two conformations.

δ (^{13}C) (ppm)

	^{13}CO	^{13}Cα	^{13}Cβ	^{13}Cγ	^{13}Cδ	^{13}Cε
[Lys]$_n$ α-helix	174.8	55.3	31.6	23.8	30.7	41.4
[Lys]$_n$ coil	174.5	53.6	30.7	22.4	26.5	39.5
|Δ| (α-helix - coil)	0.3	1.7	0.9	1.4	4.2	1.9

The chemical shifts from RefDB (section 3.3.2) and the literature [Lys]$_n$ values (table 3.5) are consistent; the shifts of aliphatic ^{13}C-Lys typically vary by 1 – 4 ppm when helix and coil conformation are compared. In contrast, the chemical shifts for these sites in elastin at 37°C and 0°C vary by ≤ 0.3 ppm, i.e., the ^{13}C-Lys chemical shifts do not significantly change as the temperature is lowered to 0°C. This observation is consistent with the findings for the Ala residues in elastin[25] as well as glass transition temperatures reported for the hydrated protein.[22]

The chemical shift of $^{13}C\alpha$ at 37°C (54.6 ppm) appears to be in between that of helix (55.3 ppm) and coil (53.6 ppm) in [Lys]$_n$ (table 3.5). However, the value determined in the variable-temperature study is consistent with population (**II**) in the 2D DARR spectrum (figure 3.8), which is assigned as KA and KX in *coil*.

The lineshapes of the CP spectra recorded at negative temperatures (-10°C, -20°C, -30°C) are not significantly different from each other. A single asymmetric peak is observed for $^{13}C\beta$, $^{13}C\gamma$, and $^{13}C\delta$. $^{13}C\epsilon$, which is narrow and well-resolved in the DP experiments at positive temperatures, has a significantly lower intensity in the CP experiments of a frozen sample. The lineshape of $^{13}C\alpha$ implies the presence of at least two populations with peaks at ~53.0 ppm and 57.2 ppm. The asymmetric ^{13}CO feature consists of ^{13}CO-Lys centered at 175.4 ppm and the natural-abundance backbone ^{13}CO, typically observed at 173 ppm in CP of hydrated elastin at -20°C.[21]

The chemical shifts of ^{13}CO and the aliphatic carbons of lysine do not significantly change as the temperature is lowered from 0°C to -30°C (table 3.6). The |Δ| for each of these sites does not exceed 0.2 ppm. Intensities in the heteroaromatic region (120 – 165 ppm) are detected in each of the CP spectra. However, the S/N is not sufficient to accurately report changes in response to temperature.

Table 3.6: ^{13}C chemical shifts of lysine in [U-^{13}C,^{15}N-Lys] elastin below 0˚C.
Chemical shift of the tallest peak position for ^{13}C CP spectra (figure 3.9) are given. The row designated as |Δ| shows the absolute difference in chemical shift between 0°C and -20°C for each site.

<table>
<tr><td colspan="6" align="center">δ (^{13}C) (ppm)</td></tr>
<tr><td>Temperature</td><td>^{13}CO</td><td>$^{13}C\alpha$ (helix)</td><td>$^{13}C\alpha$ (coil)</td><td>$^{13}C\beta$, $^{13}C\gamma$, $^{13}C\delta$,</td><td>$^{13}C\epsilon$</td></tr>
<tr><td>- 10°C</td><td>175.5</td><td>57.1</td><td>53.0</td><td>29.3</td><td>40.0</td></tr>
<tr><td>- 20°C</td><td>175.4</td><td>57.0</td><td>52.9</td><td>29.5</td><td>40.0</td></tr>
<tr><td>- 30°C</td><td>175.4</td><td>57.2</td><td>53.0</td><td>29.5</td><td>40.1</td></tr>
<tr><td>|Δ| (-30°C - (-10°C))</td><td>0.1</td><td>0.2</td><td>0.1</td><td>0.2</td><td>0.1</td></tr>
</table>

The results below 0°C are consistent with the analysis of 1D and 2D data in the previous sections. In RefDB, the chemical shift with highest number of occurrences for $^{13}C\alpha$ -Lys is 57.2 ppm for HELIX and 53.0 ppm for COIL, which closely matches the two peaks assigned as $^{13}C\alpha$ (helix) and $^{13}C\alpha$ (coil) (table 3.6). The latter two also overlap with populations (**I**) and (**III**) in the 2D spectrum obtained at -20°C (figure 3.8), which were assigned as KA and KX in *helix* and KP in *coil*, respectively.

Overall, α-helical content of lysine is higher in the frozen sample than at the physiological temperature. This change may be attributed to the glass transition. CP at -20°C is more efficient than DP at 37°C, but the latter experiment provides more detailed information about the conformations of the aliphatic carbons. Since $^{13}C\beta$, $^{13}C\gamma$, and $^{13}C\delta$ are not resolved in CP at -20°C, analysis of their secondary structures below the T_g is not possible.

3.3.5 ^{13}C T_1 measurement

The T_1 constants are obtained from the modified inversion recovery experiment (figure 3.4) conducted at 37°C (table 3.7). Signal intensities are plotted against time τ to give exponential decay curves. The T_1 value of each resolved site is extracted from the fitted trendline. The exponential decay fittings are included in Appendix IV. The determined T_1 values were used to optimize the recycle delay in experiments described in Chapters 3 and 4.

The relaxation constants range from 0.41 s for the $^{13}C\varepsilon$ to 1.37 s for ^{13}CO. The results for [U-^{13}C,^{15}N-Lys] elastin are consistent with the studies of hydrated [Lys]$_n$, which showed fast motion increases with the distance from the backbone.[44] The T_1 value for the rigid pyridinium ring at 142.0 ppm is estimated to be 1.40 s. However, an accurate analysis was not possible for this site, because the detected signal is too broad and weak. No R^2 value is provided for this particular site, and the reported constant should be treated as a rough estimate. The quality of data for the remaining ^{13}C-Lys sites was satisfactory ($R^2 >$ 0.980).

Table 3.7: $T_1{}^C$ values for lysine in [U-^{13}C,^{15}N-Lys] elastin.
T_1 is estimated for the broad feature at 142 ppm (*). The uncertainties of the T_1 values represent the standard error of regression for a given ^{13}C site.

δ (^{13}C) (ppm)	Assignment	T_1 (s)	R^2
23.1	$^{13}C\gamma$-Lys	0.56 ± 0.07	0.981
27.1	$^{13}C\delta$-Lys	0.48 ± 0.03	0.994
30.9	$^{13}C\beta$-Lys	0.53 ± 0.05	0.990
40.3	$^{13}C\varepsilon$-Lys	0.41 ± 0.05	0.982
54.7	$^{13}C\alpha$-Lys	1.06 ± 0.15	0.983
142.0	Pyridinium ring	1.4 (*)	N/A
175.0	^{13}CO-Lys	1.37 ± 0.23	0.982

The relaxation times of elastin vary greatly for different sites in this protein and are influenced by the hydration levels and temperature.[21,45] Typically, long ^{13}C T_1 constants indicate carbons in rigid environments and, conversely, short T_1 values are characteristic of mobile domains. The T_1 values for hydrated [U-^{13}C,^{15}N-Lys] elastin at 37°C are similar to previously reported work on elastin isolated from bovine ligamentum nuchae (BLN).[21] The aliphatic region has nearly liquid-like time constants (< 1 s), consistent with the high mobility of the protein.

3.4 Summary and conclusions

The [U-^{13}C,^{15}N-Lys] elastin sample differs from the previously explored amino acids in elastin. Due to the low content of Lys in the protein, difference spectra are obtained for nearly all 1D experiments. Otherwise, the signals of the highly enriched lysines overlap with the natural-abundance ^{13}C populations of other amino acids, such as glycine, alanine, or proline.

Analysis of the 1D chemical shifts of ^{13}CO and $^{13}C\alpha$ suggests COIL and HELIX are the dominant secondary structures of lysine in elastin. At 37ºC, COIL appears to be the most common conformation, and measurements at -20°C shows an increase in helical content as the sample is frozen. The 2D DARR experiment allows for selective observation of ^{13}CO - $^{13}C\alpha$ correlations.

The 2D DARR at -20°C allows for resolution of population observed in 1D experiments. The increase of HELIX in the frozen sample is confirmed. Semi-empirical approach allows for predictions of chemical shifts that account for both structure- and sequence-dependent effects. The populations corresponding to KA and KX pairs are assigned as both *helix* and *coil*. The dominant conformation of KP is *coil*.

A variable-temperature study (-30°C to 37°C) is consistent with the conformational analysis of 1D and 2D spectra. An increase in the helical content is evident when temperature is lowered from 0°C to -10°C. This finding is consistent with the studies of Ala in the crosslinking domains.[25] At warmer temperatures, shorter helices are formed due to the fast, large-amplitude motions of elastin. As the dynamics of the protein is attenuated, longer helices form in the crosslinking domains.

The spin-lattice relaxation measurements used the modified inversion recovery pulse sequence to determine the ^{13}C T_1 values in the lysine residues of elastin. This experiment is chosen as it effectively enhances the signals of low-γ nuclei in rapid motion and provides single-exponential decays in a shorter time. The determined values ranged from 0.41 s for the mobile $^{13}C\varepsilon$ to 1.37 s for the rigid ^{13}CO-Lys. A rough estimate for the pyridinium ring signal at 142 ppm is 1.4 s. The results are consistent with high mobility of elastin, as ^{13}C time constants below 1 s are similar to those of liquids.

Solid-state NMR is a powerful, versatile technique that allows for studies of secondary structures and dynamics of the lysine in elastin. The post-translational modification products, i.e., crosslinks, introduce a layer of complexity absent in the previous studies of other amino acids in this protein. Nonetheless, a great deal can be learned from the analysis of spectra recorded with different excitation schemes and at different temperatures.

3.5 References

[1] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I.N.; Bourne. P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235-242.

[2] Richardson, J. S. The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* **1981**, *34*, 167-339.

[3] Pace, C. N; Scholtz, J. M. A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins. *Biophys. J.* **1998**, *75*, 422-427.

[4] Wang, Y. J.; Jardetzky, O. Investigation of the neighboring residue effects on protein chemical shifts. *J. Am. Chem. Soc.* **2002**, *124*, 14075-14084.

[5] Wang, Y. J.; Jardetzky, O. Predicting ^{15}N chemical shifts in proteins using the preceding residue-specific individual shielding surfaces from φ, ψ i−1, and χ 1 torsion angles. *J. Biomol. NMR* **2004**, *28*, 327-340.

[6] Wang, Y. J.; Jardetzky, O. Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci.* **2002**, *11*, 852-861.

[7] Djajamuliadi, J.; Ohgo, K.; Kumashiro, K. K. Targeting Alanines in the Hydrophobic and Cross-Linking Domains of Native Elastin with Isotopic Enrichment and Solid-State NMR Spectroscopy. *Macromolecules* **2018**, *51*, 2157-2168.

[8] Ohgo, K.; Dabalos, C. L.; Kumashiro, K. K. Solid-State NMR Spectroscopy and Isotopic Labeling Target Abundant Dipeptide Sequences in Elastin's Hydrophobic Domains. *Macromolecules* **2018**, *51*, 2145-2156.

[9] Dabalos, C. Characterization of the proline residues in elastin. Ph.D. Dissertation, University of Hawai'i, Honolulu, HI, 2016.

[10] Dabalos, C. L.; Ohgo, K.; Kumashiro, K. K. Detection of Labile Conformations of Elastin's Prolines by Solid-State Nuclear Magnetic Resonance and Fourier Transform Infrared Techniques. *Biochemistry* **2019**, *58*, 3848-3860.

[11] Morris, G. A.; Freeman, R. Enhancement of nuclear magnetic resonance signals by polarization transfer. *J. Am. Chem. Soc.* **1979**, *101*, 760-762.

[12] Burum, D. P.; Ernst, R. R. Net Polarization Transfer via a J-Ordered State for Signal Enhancement of Low-Sensitivity Nuclei. *J. Magn. Reson.* **1980**, *39*, 163-168.

[13] Ohgo, K.; Niemczura, W. P.; Seacat, B. C.; Wise, S. G.; Weiss, A. S.; Kumashiro, K. K. Resolving Nitrogen-15 and Proton Chemical Shifts for Mobile Segments of Elastin with Two-dimensional NMR Spectroscopy. *J. Biol. Chem.* **2012**, *287*, 18201-18209.

[14] Kupce, E.; Boyd, J.; Campbell, I. D. Short Selective Pulses for Biochemical Applications. *J. Magn. Reson. Ser. B* **1995**, *106*, 300-303.

[15] Takegoshi, K.; Nakamura, S.; Terao, T. ^{13}C-^{1}H dipolar-assisted rotational resonance in magic-angle spinning NMR. *Chem. Phys. Lett.* **2001**, *344*, 631-637.

[16] Neal, S.; Nip, A. M.; Zhang, H.; Wishart, D. S. Rapid and accurate calculation of protein ^{1}H, ^{13}C, ^{15}N chemical shifts. *J. Biomol. NMR* **2003**, *26*, 215-240.

[17] Willard, L.; Ranjan, A.; Zhang, H. Y.; Monzavi, H.; Boyko, R. F.; Sykes, B. D.; Wishart, D. S. VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res.* **2003**, *31*, 3316-3319.

[18] Levitt, M.; Greer, J. Automatic identification of secondary structure in globular proteins. *J. Mol. Biol.* **1977**, *114*, 181-239.

[19] Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577-2637.

[20] Richards, F. M.; Kundrot, C. E. Identification of structural motifs from protein coordinate data: secondary structure and first-level supersecondary structure. *Proteins* **1988**, *3*, 71-84.

[21] Perry, A.; Stypa, M. P.; Tenn, B. K.; Kumashiro, K. K. Solid-state ^{13}C NMR reveals effects of temperature and hydration on elastin. *Biophys. J.* **2002**, *82*, 1086-1095.

[22] Kakivaya, S. R.; Hoeve, C. A. J. The Glass Point of Elastin. *PNAS* **1975**, *72*, 3505-3507.

[23] Samouillan, V.; Tintar, D.; Lacabanne, C. Hydrated elastin: Dynamics of water and protein followed by dielectric spectroscopies. *Chem. Phys.* **2011**, *385*, 19-26.

24 Samouillan, V.; André, C.; Dandurand, J.; Lacabanne, C. Effect of water on the molecular mobility of elastin. *Biomacromolecules* **2004**, *5*, 958-964.

25 Djajamuliadi, J. Investigations of the unique role of alanines in the "elastin puzzle" by solid-state NMR spectroscopy and molecular dynamics simulations. Ph.D. Dissertation, University of Hawai'i, Honolulu, HI, 2017.

26 Palmer, A. G. NMR characterization of the dynamics of biomacromolecules. *Chem. Rev.* **2004**, *104*, 3623-3640.

27 Palmer, A. G.; Kroenke, C. D.; Loria, J. P. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol.* **2001**, *339*, 204-238.

28 Mirau, P. A. A practical guide to understanding the NMR of polymers. Wiley: Hoboken, N.J., 2005.

29 Vold, R. L.; Waugh, J. S.; Klein, M. P.; Phelps, D. E. Measurement of spin relaxation in complex systems. *J. Chem. Phys.* **1968**, *48*, 3831-3832.

30 Torchia, D. A. The measurement of proton-enhanced carbon-13 T1 values by a method which suppresses artifacts. *J. Magn. Reson.* **1978**, *30*, 613-616.

31 Kowalewski, J.; Morris, G. A., A Rapid Method for Spin-Lattice Relaxation-Time Measurements On Low Magnetogyric Ratio Nuclei - Inept Signal Enhancement. *J. Magn. Reson.* **1982**, *47*, 331-338.

32 Ohgo, K.; Kumashiro, K. K. Efficient and accurate determination of ^{13}C T1 and T1ρ relaxation time constants of high-mobility polymers from limited-resolution spectra with optimized pulse sequences and data fitting. *J. Magn. Reson.* **2018**, *297*, 161-171.

33 Bielecki, A.; Burum, D. P. Temperature-Dependence of Pb-207 Mas Spectra of Solid Lead Nitrate - An Accurate, Sensitive Thermometer For Variable-Temperature MAS. *J. Magn. Reson. A* **1995**, *116*, 215-220.

34 Bennett, A. E.; Rienstra, C. M.; Auger, M.; Lakshmi, K. V.; Griffin, R. G. Heteronuclear Decoupling in Rotating Solids. *J. Chem. Phys.* **1995**, *103*, 6951-6958.

35 Kupce, E.; Boyd, J.; Campbell, I. D. Short Selective Pulses for Biochemical Applications. *J. Magn. Reson. B* **1995**, *106*, 300-303.

36 Li, Y.; Wylie, B. J.; Rienstra, C. M. Selective refocusing pulses in magic-angle spinning NMR: Characterization and applications to multi-dimensional protein spectroscopy. *J. Magn. Reson.* **2006**, *179*, 206-216.

37 Schwarzinger, S.; Kroon, G. J. A.; Foss, T. R.; Chung, J.; Wright, P. E.; Dyson, H. J. Sequence-dependent correction of random coil NMR chemical shifts. *J. Am. Chem. Soc.* **2001**, *123*, 2970-2978.

38 Zhang, H. Y.; Neal, S.; Wishart, D. S. RefDB: A database of uniformly referenced protein chemical shifts. *J. Biomol. NMR* **2003**, *25*, 173-195.

39 Pierce, R. A.; Deak, S. B.; Stolle, C. A.; Boyd, C. D. Heterogeneity of rat tropoelastin messenger-RNA revealed by cDNA cloning. *Biochemistry* **1990**, *29*, 9677-9683.

40 Usuki, T.; Yamada, H.; Hayashi, T.; Yanuma, H.; Koseki, Y.; Suzuki, N.; Masuyama, Y.; Lin, Y. Y. Total synthesis of COPD biomarker desmosine that crosslinks elastin. *Chem. Commun.* **2012**, *48*, 3233-3235.

41 Sugimura, T.; Komatsu, A.; Koseki, Y.; Usuki, T. Pr(OTf)3-promoted Chichibabin pyridine synthesis of isodesmosine in H2O/MeOH. *Tetrahedron Lett.* **2014**, *55*, 6343-6346.

42 Perry, A.; Stypa, M. P.; Foster, J. A.; Kumashiro, K. K. Observation of the glycine in elastin using ^{13}C and ^{13}N solid-state NMR spectroscopy and isotopic labeling. *J. Am. Chem. Soc.* **2002**, *124*, 6832-6833.

43 Saito, H.; Smith, I. C. Carbon-13 nuclear magnetic resonance studies of polyamino acids: the helix-coil transition of poly-L-lysine. *Arch. Biochem. Biophys.* **1973**, *158*, 154-163.

44 Krushelnitsky, A.; Faizullin, D.; Reichert, D. Hydration dependence of backbone and side chain polylysine dynamics: A ^{13}C solid-state NMR and IR spectroscopy study. *Biopolymers* **2004**, *73*, 1-15.

45 Kumashiro, K. K.; Ohgo, K.; Elliott, D. W.; Kagawa, T. F.; Niemczura, W. P. Backbone motion in elastin's hydrophobic domains as detected by ^{2}H NMR spectroscopy. *Biopolymers* **2012**, *97*, 882-888.

CHAPTER 4

QUALITATIVE AND QUANTITATIVE ANALYSIS OF CROSSLINKS IN ELASTIN VIA SOLID-STATE NMR SPECTROSCOPY

4.1 Introduction

Selective isotopic labeling of lysines and lysine-derived crosslinks in elastin improves sensitivity and resolution in the ssNMR experiments. Without incorporation of NMR-active nuclei, signals of these rare residues are too weak to be distinguished from the rest of the protein or are not detected at all. Targeted enrichment overcomes these obstacles, allowing for characterization of crosslinks.

SsNMR spectroscopy is not limited by protein size, solubility, or ability to crystalize. Insoluble elastin, close to the one found *in vivo*, is studied, rather than a hydrolysate or enzymatically cleaved peptides. Crosslinks are observed in their native state, following a gentle purification protocol, minimizing the bias from side reactions. Since NMR spectroscopy is non-destructive, a single labeled sample is sufficient for an unlimited number of experiments.

In this Chapter, crosslinks are detected in native, enriched NRSMC elastin. Various NMR experiments are conducted to find the optimal conditions for observation of all the labeled sites. The ^{13}C signals are assigned, based on solution NMR data and predicted chemical shifts (Section 4.3.1). Lysine and its derivatives are quantified in two enriched samples of elastin (Section 4.3.2). Attempts to characterize crosslinks via ^{15}N NMR are also presented (Section 4.3.3).

4.1.1 CP-based quantitative ssNMR measurements

Traditional CP experiments are non-quantitative for samples with carbons in different chemical environments.[1] The competing effects of the carbon-proton cross-polarization time constant (T_{CH}) and the proton spin-lattice relaxation time constant in the rotating frame ($T_{1\rho}{}^H$) govern the dynamics in this excitation scheme. In typical solids, T_{CH} is determined by the strength of the ^{1}H - ^{13}C coupling. Spin diffusion among protons dictates the $T_{1\rho}{}^H$, which characterizes the ^{1}H polarization decay. Carbons that are bonded to hydrogen generally have stronger dipolar couplings than the nonprotonated ones, and hence, short T_{CH} values. Conversely, carbons not attached to hydrogen or in structural groups with significant mobility (e.g. methyl rotor) have weaker couplings and longer T_{CH}. Thus, signal intensities for sites with different T_{CH} constants are unevenly enhanced in CP.

Modified CP sequences benefit from the magnetization transfer from ^{1}H to ^{13}C while allowing for quantitative measurements.[2,3,4] To observe and quantify the rare crosslinks in intact elastin, the enhancement of S/N provided by CP is necessary. In the resulting spectra, peaks areas correspond to the number of spins in the respective sites, allowing for quantification of the resolved features.

Quantitative **C**ross **P**olarization (QUCP)[4] allows for quantitative measurements of enriched samples with adjacent ^{13}C-^{13}C sites, such as [U-^{13}C,^{15}N-Lys] elastin. The pulse sequence begins with a 90° ^{1}H pulse followed by a CP mixing time (figure 4.1). The magnetization transfer from ^{1}H to ^{13}C spins favors carbons with short T_{CH} values. Then, a 90° ^{13}C pulse rotates the magnetization to the z-direction, and a dipolar assisted rotational resonance (DARR) mixing period follows. The polarization is redistributed through ^{13}C-^{13}C spin diffusion and flipped for detection with another ^{13}C 90° pulse. The resulting ^{13}C QUCP spectrum provides quantitative data.
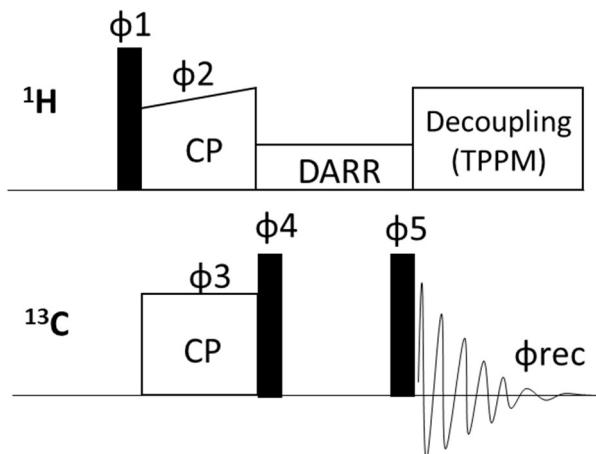
Figure 4.1: Pulse sequence for ^{13}C-{^{1}H} QUCP experiment.[4]
The phase cycling was modified from the original paper as follows: Φ_1 = {x, -x}; Φ_2 = {y}; Φ_3 = {y, -y, y, -y, x, -x, x, -x}; Φ_4 = {-x, x, -x, x, y, -y, y, -y}; Φ_5 = {x, -x, -y, y, x, -x, -y, y}; Φrec = {y, -y, x, -x, y, -y, x, -x}.

Composite-**P**ulse **Multi**ple **C**ross **P**olarization (ComPmultiCP) is a quantitative experiment optimized for samples with dilute labels, such as [ε-^{13}C-Lys] elastin.[2,3] It begins with successive loops of an excitation pulse followed by a mixing time and a storage pulse (figure 4.2). The composite ^{13}C pulses improve accuracy of the quantitative data as the method becomes less sensitive to Hartmann-Hahn mismatch, flip-angle error, and resonance offset.[3] Short CP blocks minimize the non-uniform signal loss due to $T_{1\rho}{}^H$ relaxation. The t_z period allows ^{1}H magnetization to recover and reverses the loss due to proton spin diffusion. The ^{13}C magnetization increases at different rates, depending on T_{CH}, but reaches a steady-state at all ^{13}C sites at the end of the loop. The final (sixth) CP contact time is shorter to eliminate any residual effects of non-uniform $T_{1\rho}{}^H$ relaxation. Lastly, a spin echo sequence follows on the ^{13}C channel, removing the effects of inhomogeneities in the B_0 field by refocusing the magnetization, which reaches a maximum after the second delay τ_r.
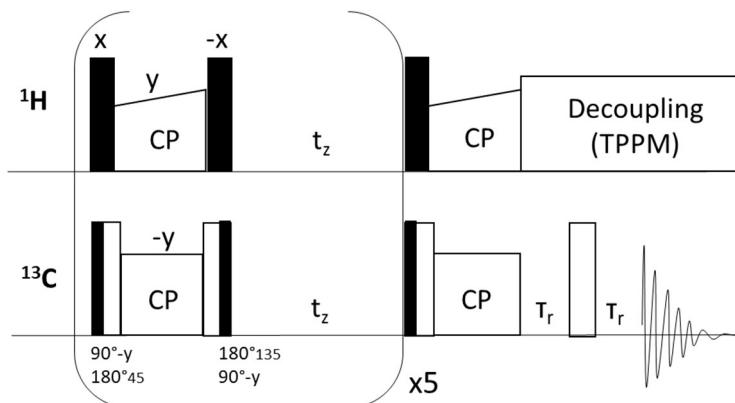
Figure 4.2: Pulse sequence for ^{13}C-{^{1}H} ComPmultiCP experiment.[3]
Black and white rectangles represent 90° and 180° pulses, respectively. A Hahn echo block is utilized before the acquisition with delays τ equal to the rotor period, τ_r.

A single, CP-based, quantitative measurement is not applicable to both [U-^{13}C,^{15}N-Lys] elastin and [ε-^{13}C-Lys] elastin. The first sample has consecutive ^{13}C-^{13}C spins, whereas ^{13}C nuclei is bonded to NMR-inactive ^{12}C in [ε-^{13}C-Lys] elastin. Thus, ^{13}C spin diffusion occurs in [U-^{13}C,^{15}N-Lys] elastin but not in [ε-^{13}C-Lys] elastin. In QUCP, the unevenly enhanced magnetizations are redistributed during DARR mixing time. This equilibration is too slow without uniform enrichment of lysine, and the ^{13}C spins in [ε-^{13}C-Lys] elastin are quantified with ComPmultiCP instead. In this experiment, the effects of differential $T_{1\rho}$ relaxation are overcame with multiple CP blocks followed by ^{1}H repolarization periods, t_z. Spin diffusion in [U-^{13}C,^{15}N-Lys] elastin would contribute to (uneven) equilibration of the magnetization of protonated and nonprotonated carbons during t_z, resulting in a non-quantitative spectrum.

4.2 Materials and methods

[U-^{13}C,^{15}N-Lys] elastin, [ε-^{13}C-Lys] elastin, and the natural-abundance protein were prepared as described in Chapter 2. The typical sample mass, rotor inserts, and other parameters for ssNMR experiments are included in Chapter 3. Conditions and parameters for ^{13}C DP and CP experiments are the same as in Section 3.2, unless stated otherwise.

For CP TOSS[5] experiment at -20°C, samples of [U-^{13}C,^{15}N-Lys], [ε-^{13}C-Lys], and unenriched elastin were frozen at -20°C and spun at this temperature for about 40 min for equilibration. A 4.2-μs ^{1}H 90° pulse was followed by a 1 ms contact time. Spinning sidebands were suppressed by the TOSS-4 sequence[5] with 9.2 μs ^{13}C 180° pulses (figure 1.8). The delays between these pulses were: $t_1 = 0.1226\tau_r$, $t_2 = 0.0773\tau_r$, $t_3 = 0.2236\tau_r$, $t_4 = 1.0433\tau_r$, $t_5 = 0.7744\tau_r$ (where τ_r is the rotor period). After t_5, TPPM ^{1}H decoupling was applied during the 20 ms acquisition time with an applied field strength of $\gamma^H B_1^H/2\pi \sim 60$ kHz.

The ^{13}C QUCP[4] experiment was performed on [U-^{13}C,^{15}N-Lys] elastin and the unenriched elastin at -20°C (figure 4.1). A 3.5-μs ^{1}H 90° pulse was followed by a 1-ms contact time with a 5-s recycle delay. DARR[6] mixing was applied for 500ms with B_1 field strength of $\gamma^H B_1^H/2\pi = 1*\nu_r = 8$ kHz. The two ^{13}C pulses (90°$_{\Phi4}$, 90°$_{\Phi5}$) had width of 4.7 μs. TPPM ^{1}H decoupling was applied during the 20 ms acquisition time with an applied field strength of $\gamma^H B_1^H/2\pi \sim 60$ kHz.

^{13}C ComPmultiCP[3] experiment was conducted on [ε-^{13}C-Lys] elastin and unenriched elastin at -20°C (figure 4.2). A composite ^{13}C 90°-180° excitation pulse was followed by CP with 1 ms contact time and then a ^{13}C 180°-90° storage pulse. After time tz (800 ms), this sequence is repeated four more times. After the final (sixth) CP block, with contact time of 500 μs, a Hahn echo block follows ($\tau_{echo} = \tau_r = 125$ μs). TPPM ^{1}H decoupling was applied during the 20 ms acquisition time with an applied field strength of $\gamma^H B_1^H/2\pi \sim 60$ kHz. A recycle delay of 5 s was used. The complete phase cycling can be found at http://www.ksrlab.org/nmrksr/.

The Agilent 600 MHz DD2 NMR spectrometer equipped with a 3 mm Direct Detect™ NMR probe was used to obtain ^{13}C solution NMR spectra of a commercially available mixture of DES and IDES (Cat. No. MD687, Elastin Products Co., Inc., Owensville, MO). Five mg of the mixture were dissolved in 1 mL of water. One hundred seventy-six μL of this standard were used in the derivatization reaction with FDLA (Chapter 2). The remaining solution was lyophilized overnight, yielding about 4 mg of dry sample. The powder was dissolved in MeOH-d$_4$ and placed in a 3 mm NMR tube. The spectrum (14976 scans, 3 s recycle delay) was processed in Mnova software with Whittaker-Smoother baseline correction.

The ^{15}N DP experiment was performed on [U-^{13}C,^{15}N-Lys] elastin sample at 37°C. A 9.3 μs ^{15}N 90° pulse was followed by a Hahn echo (τ_{echo}-π-τ_{echo}) block with τ_{echo} = 20 μs. To remove dipolar (^{1}H-^{15}N) interactions, ^{1}H TPPM decoupling was used. The recycle delay was 10 s.

The ^{15}N CP experiment was performed on [U-^{13}C,^{15}N-Lys] elastin sample at -20°C. A 4.2 μs ^{1}H 90° pulse was followed by a 3.5 ms contact time with recycle delay of 5 s. The CP field strengths were set to $\gamma^N B_1^N/2\pi$ = 28 kHz and $\gamma^H B_1^H/2\pi$ = 36 kHz, so that the Hartmann-Hahn sideband condition is met and $\gamma^N B_1^N/2\pi = \gamma^H B_1^H/2\pi + n^* \nu_r$, where ν_r is a spinning rate of the rotor, and n = 1. The CP field strength for ^{1}H was ramped from 95% to 105%. A Hahn echo (τ_{echo}-π-τ_{echo}) block was utilized for CP with τ_{echo} = 125 μs.

4.3 Results and discussion

Samples of isotopically labeled elastin were isolated and purified in mild conditions, allowing for quantitative ssNMR studies of an intact protein. Briefly, a digest with cyanogen bromide (CNBr) in formic acid was used. CNBr cleaves proteins at methionine, which is not found in the NRSMC elastin (Appendix II). Then, extraction with guanidinium hydrochloride and mercaptoethanol followed. Guanidinium salt is a chaotropic agent that disrupts hydrogen bonds, whereas mercaptoethanol reduces disulfide bridges. Cysteine, the only residue forming the latter, does not appear in the elastin sequence. This step is intended to remove cysteine-containing proteins, such as collagen.

Lysine and lysine-derived crosslinks are observed and quantified in intact elastin with ssNMR spectroscopy. ^{13}C signals are assigned based on solution NMR data for the well-characterized DES and IDES, and predictions for the rare bi- and tri-functional crosslinks. Quantitative CP-based measurements allow for analysis of crosslinking content in the samples of enriched elastin. ^{15}N NMR spectra do not provide sufficient S/N for studies of modified lysines.

4.3.1 Predictions of ^{13}C chemical shifts of crosslinks

To interpret and assign the complex NMR spectra, external ^{13}C chemical shift references for the crosslinks are needed. DES and IDES have been extensively characterized, due to their potential to serve as biomarkers for COPD.[7,8,9] However, chemical shift predictions are the only available option for the remaining lysine derivatives.

Mnova Best algorithm was used to predict the ^{13}C chemical shifts for the crosslinks (Appendix III). This method combines two different algorithms, hierarchically ordered spherical environment (HOSE) and neural networks. HOSE is suitable for query structures that are well-represented in the literature. The neural network tends to be more accurate for novel compounds that are not commonly found in databases. Combining the two approaches results in more accurate and precise predictions.

Plots of the predicted ^{13}C chemical shifts of crosslinks reveal resolved regions in the spectra (figure 4.3). The aldehyde carbonyl (^{13}CO$_{CHO}$) signals around 200 ppm are unique to Lya, AA, and CPN. Peaks in the 135 – 165 ppm range arise from bi-, tri-, and tetrafunctional crosslinks. The pyridinium ring carbons have predicted chemical shifts of 141, 144, and 151 ppm in DES and 137, 139, 146, and 151 ppm in IDES. Bifunctional AA also has signals in this region, at 153 ppm (both labeling schemes) and 142 ppm (only [U-^{13}C,^{15}N-Lys] elastin). DLNL also overlaps with the pyridinium ring signals (160 ppm). MDES has a well-resolved signal at 128 ppm, which can be used to confirm presence of this trifunctional crosslink in elastin.

The remaining regions are too crowded in [U-^{13}C,^{15}N-Lys] elastin spectra for assignments, but are informative in [ε-^{13}C-Lys] sample. For example, the crosslinking aliphatic signals (20 – 60 ppm) are not resolved in [U-^{13}C,^{15}N-Lys] elastin spectra, but [ε-^{13}C-Lys] sample allows for observation of unmodified lysine at 40 ppm. The ^{13}Cε site in modified Lys residues is shifted downfield to the ^{13}Cα region (48 – 60 ppm) and indicate presence of crosslinks, e.g., LNL (48 ppm), DLNL (57 ppm), MDES (49 ppm, 52 ppm), DES (64 ppm), or IDES (62 ppm). The chemical shifts of ^{13}CO$_{BB}$ (178 – 174 ppm) are not resolved in [U-^{13}C,^{15}N-Lys] elastin, but this region can be used to quantify scrambling to ^{13}CO-Gly (Chapter 2) in [ε-^{13}C-Lys] elastin.
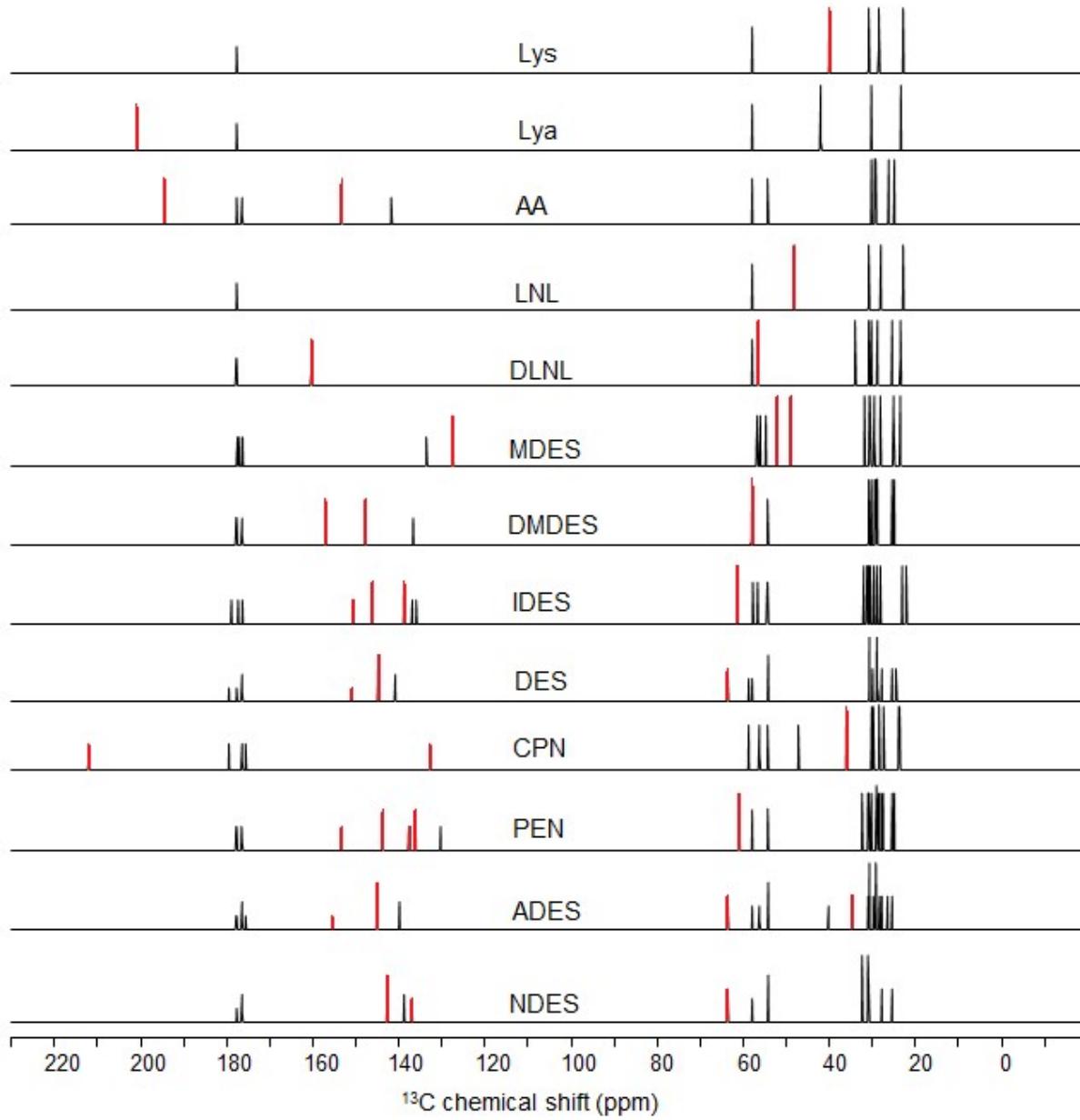
Figure 4.3: Predicted ^{13}C chemical shifts of lysine and lysine-derived crosslinks in [U-^{13}C,^{15}N-Lys] elastin and [ε-^{13}C-Lys] elastin.

Red denotes signals originating from ^{13}Cε in [ε-^{13}C-Lys] elastin. Both red and black signals are expected in the spectra of [U-^{13}C,^{15}N-Lys] elastin. Predictions for lysine and crosslinks (figure 1.4) were obtained using Mnova Best algorithm. Peak height reflects the number of protons bonded to the corresponding site.

To validate the predictions, a solution ^{13}C NMR spectrum of a commercially available DES and IDES mixture was obtained (figure 4.4). ^{13}CO$_{BB}$ peaks are observed at 176 and 177 ppm. Three signals at 142, 144, and 161 ppm are consistent with the pyridinium ring predictions. Peak at 63 ppm is assigned as ^{13}Cε. The overlapping ^{13}Cα signals are observed in the 56 – 57 ppm range. The upfield signals (23 – 33 ppm) arise from aliphatic ^{13}Cβ, ^{13}Cγ, and ^{13}Cδ. A peak at 49.5 ppm corresponds to methanol.

An unidentified peak at 41.5 ppm is likely a contaminant. The purity of the sample reported by the manufacturer is 99.5%. The commercially available DES and IDES mixture is obtained from BLN elastin and is assessed via amino acid analysis. Non-amino acid contaminants could be present in the commercial standard.
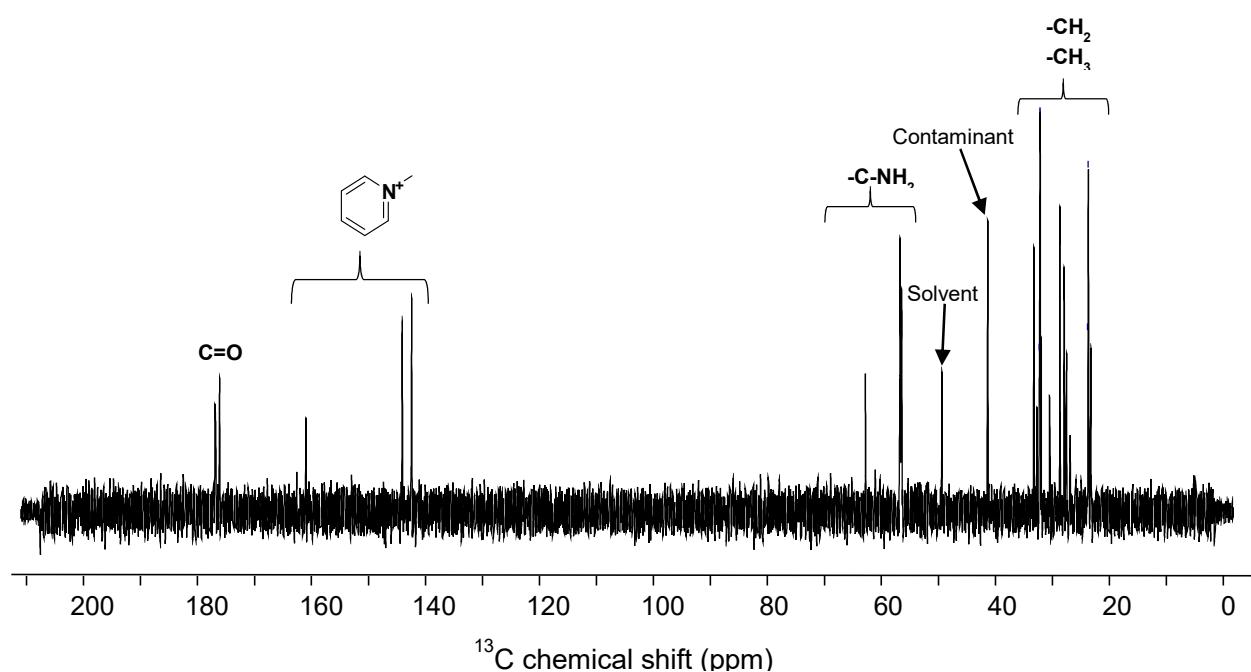
Figure 4.4: Solution ^{13}C NMR spectrum of DES and IDES mixture.
Spectrum for a 4 mg sample in MeOH-d$_4$ was obtained from 14976 scans. Baseline correction (Whittaker Smoother) was applied.

Overall, the predicted chemical shifts are consistent with the experimental data, but minor discrepancies are observed, in particular in the pyridinium ring data (tables 4.1 - 4.2). Mnova provides significantly lower values for C6$_{DES}$, C6$_{IDES}$, and C5$_{IDES}$ than reported for synthetic products[8,9] and observed for the commercial standard (figure 4.4). The predicted value for C6$_{DES}$ (150.9 ppm) is upfield compared to those of the synthetic (161.0 ppm) and the commercial (161.0 ppm) crosslinks. The calculated chemical shift for C6$_{IDES}$ (138.9 ppm) is also more upfield than the experimental data for the synthetic (146.7 ppm) and the commercial (144.1 ppm) products. The predicted shift of C5$_{IDES}$ (136.0 ppm) is also underestimated compared to those of the synthetic (141.3 ppm) and the commercial (142.5 ppm) product. The experimental values are generally consistent with each other, i.e., Mnova predictions for C6 and C5 sites are incorrect.

Additional inconsistencies are found in the Mnova predictions for an aliphatic carbon (C8$_{DES}$) and carbonyl carbons (C24$_{DES}$ and C31$_{IDES}$). C8$_{DES}$ has a predicted chemical shift of 27.9 ppm; the value reported for this site is 32.7 ppm for the synthetic product. Both C24$_{DES}$ and C31$_{IDES}$ have calculated shifts of 174.7 and 178.8 ppm, respectively, whereas 174.7 ppm is reported for both of these two sites in the literature. The values measured for the commercial sample are in between, at 176.9 and 176.1 ppm, respectively.
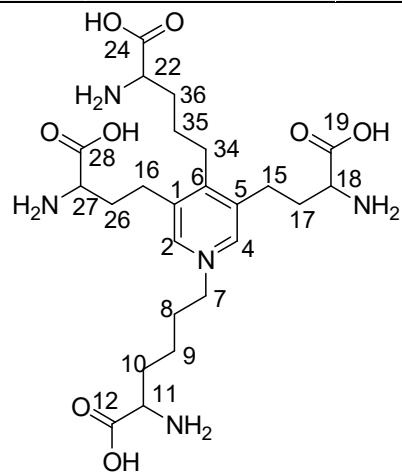
The chemical shifts IDES in the commercial mixture do not match the predictions or the synthetic product values.[9] The Mnova values for C1 and C4 are 136.9 and 150.7 ppm, respectively, and match the chemical shifts of the synthetic IDES (139.4 and 154.2 ppm). However, the closest peaks in the NMR spectrum of the commercial product (figure 4.4) are at 142.5 and 161.0 ppm. This finding implies the commercial sample actually consisted mostly of DES.

In summary, Mnova chemical shift predictions are useful tools in interpretation of the ^{13}C spectra of labeled elastin. Some of signals originating from bi-, tri-, and tetrafunctional crosslinks are expected to be sufficiently resolved to allow for their unambiguous assignment. The predictions for the pyridinium ring carbons are not consistent with the solution NMR data for commercial and synthetic product. Thus, experimental values are used in assignments of DES and IDES.

Table 4.1: Predicted and experimental ^{13}C chemical shifts of DES.
Difference exceeding 4ppm in **red**.

Carbon Atom	^{13}C Chemical Shift (ppm)				
	A	**B**	**C**	**D**	**E**
	Mnova Prediction	**Synthetic DES[i]**	**Δ_1 (A - B)**	**Commercial (I)DES[ii]**	**Δ_2 (A - D)**
1 C	140.8	142.3	-1.5	142.5	-1.7
2 CH	144.8	144.2	0.6	144.1	0.7
4 CH	144.8	144.2	0.6	144.1	0.7
5 C	140.8	142.3	-1.5	142.5	-1.7
6 C	150.9	161.0	**-10.1**	161.0	**-10.1**
7 CH2	63.7	63.0	0.7	62.9	0.8
8 CH2	27.9	32.7	**-4.8**	27.6	0.3
9 CH2	25.5	24.0	1.5	23.8	1.7
10 CH2	31.0	32.4	-1.4	32.4	-1.4
11 CH	58.0	55.3	2.7	56.6	1.4
12 C	177.7	174.7	3.0	176.1	1.6
15 CH2	29.1	28.1	1.0	28.1	1.0
16 CH2	29.1	28.1	1.0	28.1	1.0
17 CH2	30.8	32.4	-1.6	32.3	-1.5
18 CH	54.3	55.2	-0.9	56.6	-2.3
19 C	176.4	174.6	1.8	176.1	0.3
22 CH	58.9	55.2	3.7	56.6	2.3
24 C	179.4	174.7	**4.7**	176.9	2.5
26 CH2	30.8	31.6	-0.8	32.1	-1.3
27 CH	54.3	55.2	-0.9	56.8	-2.5
28 C	176.4	174.6	1.8	176.1	0.3
34 CH2	28.9	28.1	0.8	28.8	0.1
35 CH2	24.6	23.6	1.0	23.4	1.2
36 CH2	30.1	32.4	-2.3	33.4	-3.3
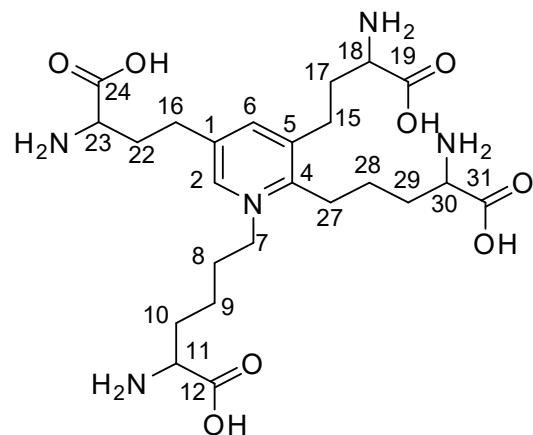
[i] From Ref.8: *Total synthesis of COPD biomarker desmosine that crosslinks elastin.*
[ii] From solution NMR of commercial product (figure 4.4).

Table 4.2: Predicted and experimental ^{13}C chemical shifts of IDES.

Difference exceeding 4ppm in **red**.

Carbon Atom	^{13}C Chemical Shift (ppm)				
	F	**G**	**H**	**I**	**J**
	Mnova Prediction	**Synthetic IDES**[iii]	**Δ₃ (F - G)**	**Commercial (I)DES**[iv]	**Δ₄ (F - I)**
1 C	136.9	139.4	-2.5	142.5	**-5.6**
2 CH	146.1	143.8	2.3	144.1	2.0
4 C	150.7	154.2	-3.5	161.0	**-10.3**
5 C	136.0	141.3	**-5.3**	142.5	**-6.5**
6 CH	138.9	146.7	**-7.8**	144.1	**-5.2**
7 CH2	61.5	58.6	2.9	62.9	-1.4
8 CH2	29.1	31.0	-1.9	28.8	0.3
9 CH2	23.2	21.9	1.3	23.4	-0.2
10 CH2	31.3	30.5	0.8	30.6	0.7
11 CH	56.7	55.0	1.7	56.6	0.1
12 C	177.3	175.0	2.3	176.9	0.4
15 CH2	29.8	27.9	1.9	27.6	2.2
16 CH2	31.0	27.9	3.1	27.6	3.4
17 CH2	31.2	31.5	-0.3	32.1	-0.9
18 CH	54.4	54.8	-0.4	56.8	-2.4
19 C	176.2	174.4	1.8	176.1	0.1
22 CH2	32.1	31.4	0.7	32.1	0.0
23 CH	54.5	54.7	-0.2	56.8	-2.3
24 C	176.3	174.4	1.9	176.1	0.2
27 CH2	28.2	28.7	-0.5	28.1	0.1
28 CH2	22.2	24.3	-2.1	23.4	-1.2
29 CH2	30.6	30.8	-0.2	32.1	-1.5
30 CH	57.8	54.8	3.0	56.8	1.0
31 C	178.8	174.7	**4.1**	176.1	2.7

[iii] From ref. 9: *Pr(OTf)3-promoted Chichibabin pyridine synthesis of isodesmosine in H2O/MeOH.*
[iv] From solution NMR of commercial product (figure 4.4).

4.3.2 Assignment of ^{13}C chemical shifts

Observation of the rare crosslinks requires selection of an appropriate excitation scheme and experimental conditions. DP at 37°C offers quantitative spectra, but the signal intensities in the heteroaromatic region (135 – 165 ppm) are insufficient for analysis (figure 3.6). However, signals originating from crosslinks are detected in CP spectra of hydrated elastin sample cooled to -20°C (figure 4.6).

Lysine and crosslinking signals are not observed in the CP TOSS spectrum of unlabeled elastin (figure 4.6 A). The $^{13}CO_{BB}$ peak has the highest intensity at 173 ppm. The broad feature centered around 129 ppm arises from the heteroaromatic rings of phenylalanine and tyrosine. The peaks at $\delta(^{13}C)$ = 43, 49, 53, 57, 60 ppm originate from the natural-abundance $^{13}C\alpha$ of amino acids, with the most upfield one assigned as glycine. Peaks at 30 and 25 ppm arise predominantly from the aliphatic carbons of valine and proline. The methyl signals ($\delta(^{13}C)$ = 19, 16 ppm) originate from valine and alanine.[10]

The CP TOSS spectrum of [U-^{13}C,^{15}N-Lys] elastin confirms high isotopic enrichment (figure 4.6B). The lineshape is significantly different compared to the spectra of unenriched elastin. Two broad intensities centered at 198 and 142 ppm are observed. The aliphatic region has peaks at 57 and 53 ppm, which are consistent with the spectra of the unlabeled sample, and two additional features with the tallest points at 40 and 29 ppm.

Even with ~90% isotope incorporation, natural-abundance signals need to be subtracted from those of the enriched proteins. Scaling factors are chosen to cancel the signals originating from methyl groups of valine and alanine in the upfield region of the spectrum (figure 4.6C). Similarly, the aliphatic signals upfield from the narrow $^{13}C\varepsilon$-Lys feature in the [ε-^{13}C-Lys] elastin spectrum (figure 4.6D) are cancelled in the difference spectrum (figure 4.6E).
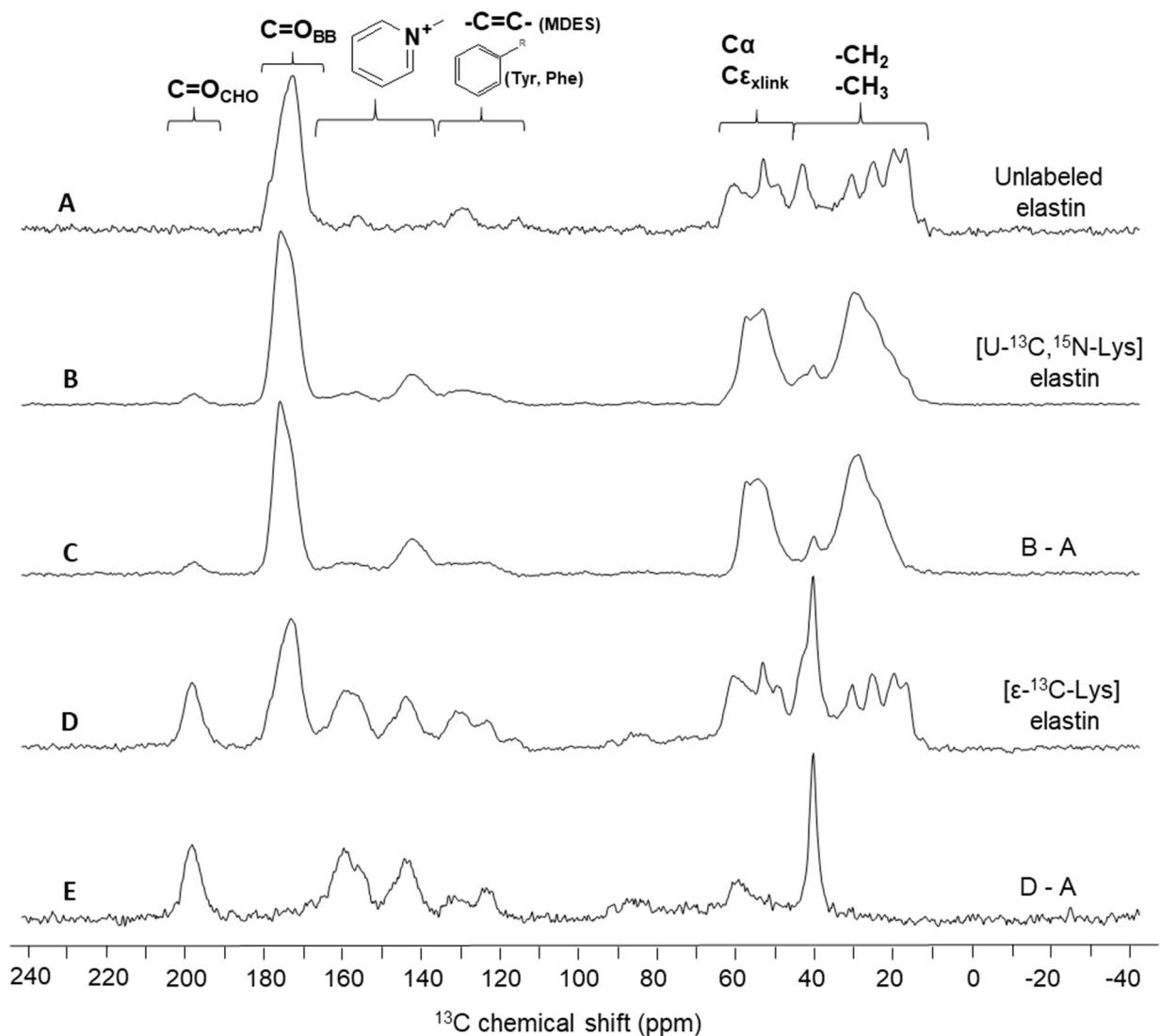
Figure 4.5: ^{13}C CP TOSS NMR spectra of labeled and unlabeled elastin at -20°C.
(A) Spectrum of unenriched elastin, (B) spectrum of [U-^{13}C,^{15}N-Lys] elastin, and (C) difference spectrum (B - A) showing signals originating from [U-^{13}C,^{15}N-Lys] (figure 4.5C). (D) Spectrum of [ε-^{13}C-Lys] elastin and (E) difference spectrum (D - A) displaying signals originating from [ε-^{13}C-Lys]. Spectra (A), (B), and (D) were acquired with recycle delay of 5 s and 2048 scans, then processed with 40 Hz line broadening. Each spectrum is scaled to its tallest peak.

90

The predicted and experimental chemical shifts were used to assign the difference spectra of [U-^{13}C,^{15}N-Lys] elastin (figure 4.5C). Aldehyde carbonyl peak (^{13}CO$_{CHO}$) at 198 ppm could correspond to Lya, AA, and CPN. Lya is highly reactive; it is either present in very low quantities[11] or not found at all.[12] CPN is so rare, it has never been quantified.[13] Therefore, AA is the most likely source of any signal in this region. The backbone carbonyl peak (^{13}CO$_{BB}$) is observed at 176 ppm. The broad features ranging from 135 to 166 ppm arise from the pyridinium rings of DES, IDES, and possibly less common crosslinks that contain this moiety. Some ^{13}Cε of unsaturated bifunctional crosslinks, such as AA, DMDES, or DLNL, possibly contribute to this feature as well. A broad signal between 115 and 136 ppm is consistent with the predictions for MDES (128 and 134 ppm) and other rare crosslinks such as CPN (133 ppm) and PEN (130 ppm). Heteroaliphatics ranging from 45 to 60 ppm include both the ^{13}Cα populations and ^{13}Cε in modified Lys. The broad feature in the saturated aliphatics region between centered at 29 ppm originates from the remaining carbon atoms (^{13}Cβ, ^{13}Cγ, ^{13}Cδ) of lysine and crosslinks. A sharp peak at exactly 40.0 ppm corresponds to ^{13}Cε of unmodified Lys.

Due to the natural linewidths of solid proteins, the peaks for the aliphatic ^{13}C sites in [U-^{13}C,^{15}N-Lys] elastin are not resolved (figure 4.5C, 20 – 38 ppm). Additional complexity arises from the overlap of peaks from the crosslinks, many of which have very similar structures. A spinning sideband of ^{13}Cα has intensity comparable to those of the crosslinking peaks at the spinning rate of 8 kHz, and its expected chemical shift is at ~135 ppm, complicating the interpretation of the crowded, heteroaromatic region.

The CP TOSS spectrum of [ε-^{13}C-Lys] elastin is consistent with high levels of isotopic enrichment determined via HPLC-MS assay (figure 4.5D). The aliphatic region of the spectrum agrees with the lineshape of the unenriched elastin (figure 4.5A), with the exception of an additional, narrow peak at 40 ppm, assigned as ^{13}Cε of unmodified lysine. The downfield region of the spectrum reveals peaks with δ(^{13}C) = 123, 131, 144, 159, and 198 ppm. Carbonyl of elastin is centered around 173 ppm.

CP TOSS difference spectrum of [ε-^{13}C-Lys] complements the results for [U-^{13}C,^{15}N-Lys] elastin. Assignments are based on the Mnova Best predictions and the chemical shifts reported for synthetic DES and IDES (figure 4.6). The distinct features and the possible corresponding ^{13}Cε sites in the crosslinks are highlighted with color. The weak, broad feature between 80 and 93 ppm is denoted as "‡" and does not match any ^{13}C chemical shift predicted for the crosslinks in elastin.
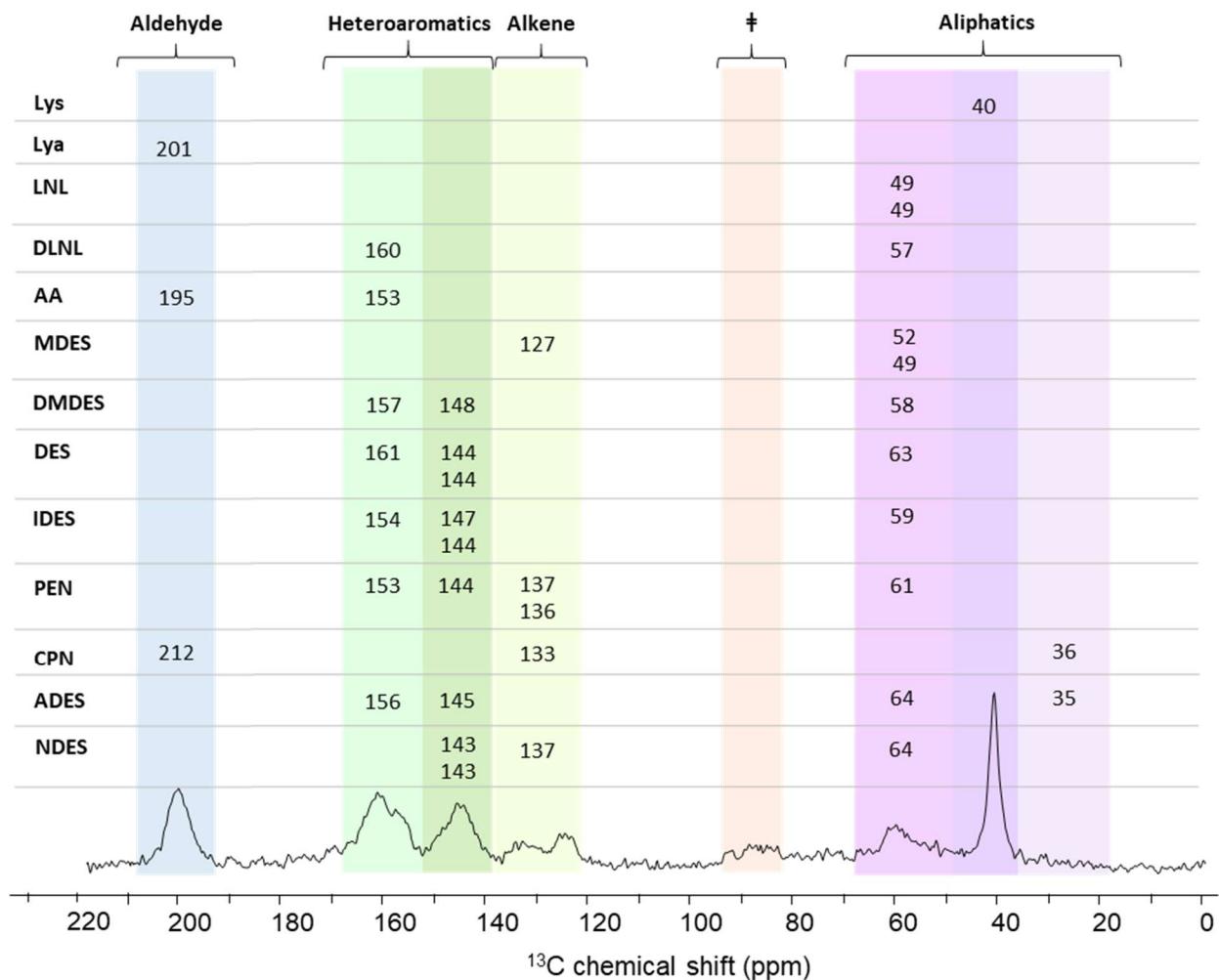
Figure 4.6: 13C CP TOSS NMR difference spectrum of [ε-13C-Lys] elastin and 13Cε chemical shifts of lysine and crosslinks.

	Aldehyde	Heteroaromatics		Alkene	‡	Aliphatics	
Lys						40	
Lya	201						
LNL						49 49	
DLNL		160				57	
AA	195	153					
MDES				127		52 49	
DMDES		157	148			58	
DES		161	144 144			63	
IDES		154	147 144			59	
PEN		153	144	137 136		61	
CPN	212			133			36
ADES		156	145			64	35
NDES			143 143	137		64	

Figure 4.6: ^{13}C CP TOSS NMR difference spectrum of [ε-^{13}C-Lys] elastin and ^{13}Cε chemical shifts of lysine and crosslinks.
The ^{13}Cε chemical shifts of lysine and crosslinks are compared to the CP TOSS results (figure 4.5E). DES[8] and IDES[9] list experimental chemical shifts, remaining entries reflect predictions (figure 4.4). Color scheme corresponds to the different types of chemical environments (figure 1.4).

The putative aldehyde peak centered at 198 ppm was observed in the spectrum of [ε-^{13}C-Lys] elastin spectra (figure 4.6). This feature also appears in the spectrum of [U-^{13}C,^{15}N-Lys] elastin (figure 4.5C), but not the unenriched protein (figure 4.5A). Therefore, this signal originates from a crosslink, rather than contaminant or a side product of the protein purification. Lya, AA, and CPN, the possible sources of this signal, all have the aldehyde or ketone attached to a carbon derived from ^{13}Cε-Lys (figure 1.4).

The lineshapes in the 149 – 166 ppm region are different in the spectra of the [ε-^{13}C-Lys] elastin compared to [U-^{13}C,^{15}N-Lys] elastin (figure 4.5C), as only three out of five pyridinium carbons in DES and IDES originate from Cε-Lys. Two resolved features with tallest positions at 144 and 159 ppm are observed. Similarly, the lineshape of the broad signal in the 120 – 136 ppm range, assigned as alkenes, looks different in the [ε-^{13}C-Lys] elastin spectrum. A more distinct peak at 123 ppm, which is the closest match for MDES, is detected. The weak, broad signal; with the highest point at 59 ppm is the ^{13}Cε that shifted downfield as a result of crosslink formation. This spectrum has only one clearly defined peak upfield, the ^{13}Cε of unmodified Lys centered at 40.0 ppm. No peak around 35 – 36 ppm, as predicted for ADES and CPN, is observed, implying these crosslinks were absent from the NRSMC elastin sample.

An unassigned weak, broad feature is observed between 80 and 93 ppm (figure 4.6). This signal is not present in the spectra of unenriched elastin (figure 4.5A), which implies that it originates from ^{13}Cε-Lys. To exclude the possibility of incomplete sideband suppression by TOSS, another spectrum was obtained at 6 kHz (figure 4.8). The signal is still present, and its overall lineshape has not changed. This feature most likely corresponds to a hydroxyl group of a minor lysine modification product that either has not been isolated yet or was a by-product of the extraction and purification procedures.
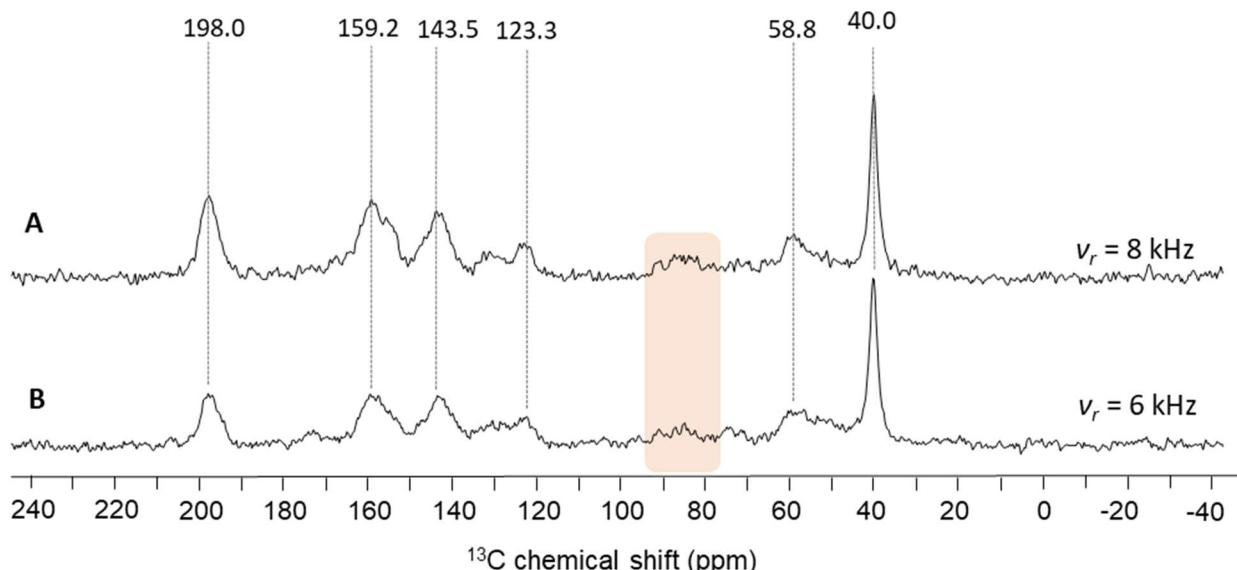
Figure 4.7: ^{13}C CP TOSS NMR difference spectra of [ε-^{13}C-Lys] elastin at varied MAS rates.
Spectra were obtained for [ε-^{13}C-Lys] elastin and natural-abundance protein at (A) 8 kHz (figure 4.5E) and (B) 6 kHz spinning rate and 1232 scans. A weak, broad feature at 80 – 95 ppm is highlighted in orange. Intensities are adjusted to the tallest peak at 40 ppm. Dashed lines correspond to the highest position of each feature in the spectrum (A).

Overall, the lineshapes of the remaining sites in the spectra do not appear significantly changed at the lower spinning rate. The only exception is the shoulder at 154 ppm (figure 4.7A), which is not as distinct at 6 kHz (figure 4.7B). However, results are inconclusive as this deviation may be due to different S/N in the two experiments.

4.3.3 Quantification of crosslinks

Two different CP-based experiments were needed, as the two the isotopically labeled samples are not compatible with a single approach. [U-^{13}C,^{15}N-Lys] elastin is uniformly enriched, whereas [ε-^{13}C-Lys] elastin has a single ^{13}C label per lysine equivalent. In QUCP of [U-^{13}C,^{15}N-Lys] elastin, a broadband homonuclear recoupling technique redistributes the unevenly enhanced magnetizations of ^{13}C spins. This method relies on ^{13}C spin diffusion, which is too slow without uniform ^{13}C enrichment. Therefore, ComPmultiCP is instead used for [ε-^{13}C-Lys] elastin. This method combines multiple CP steps separated by ^{1}H repolarization periods that erase the uneven loss of ^{1}H magnetization due to the differential $T_{1\rho}$ relaxation. Polarization transfer from ^{1}H to ^{13}C results in a local magnetization loss of ^{1}H spins, but this effect is cancelled by ^{1}H spin diffusion. Surrounding ^{1}H spins constitute a large reservoir of ^{1}H magnetization in samples that are not uniformly enriched. In addition, ComPmultiCP spectrum of [U-^{13}C,^{15}N-Lys] elastin would not be quantitative due to ^{13}C spin diffusion that would contribute to equilibration of the magnetization of carbons

QUCP quantifies the labeled sites in [U-^{13}C,^{15}N-Lys] elastin (figure 4.8A). The experiment was first optimized with a sample of U-^{13}C,^{15}N-Lys (10% w/w in unenriched lysine). After 500 ms mixing time, the enhancement factors for all the ^{13}C sites were almost uniform (within 3% error). This error is consistent with the one reported for U-^{13}C,^{15}N-Tyr.[4] To further evaluate the accuracy of the measurement, the number of ^{13}C nuclei in the backbone carbonyl (^{13}CO$_{BB}$) is compared to the remaining labeled sites. ^{13}CO$_{BB}$ centerband and its sidebands are adjusted to a total area of 1.04 (red), which accounts for the intensities from 167 to 182 ppm and includes the ± 1 spinning sidebands from 87 to 115 ppm and 247 to 262 ppm. Due to scrambling (Chapter 2), about 0.04 of ^{13}CO-Gly residues are present for each ^{13}CO-Lys. These two ^{13}CO populations are not resolved; hence, the carbonyl feature is normalized to 1.04. The rest of the spectrum (gray) consists of signals coming from the remaining five labeled carbons of lysine (^{13}Cα, ^{13}Cβ, ^{13}Cγ, ^{13}Cσ, ^{13}Cε), and their products of the post-translational modification. The obtained ratio (1.00 : 4.89) is close to the theoretical one (1.00 : 5.00). The unaccounted difference (5.00 - 4.89 = 0.11) implies experimental error of ~2%.
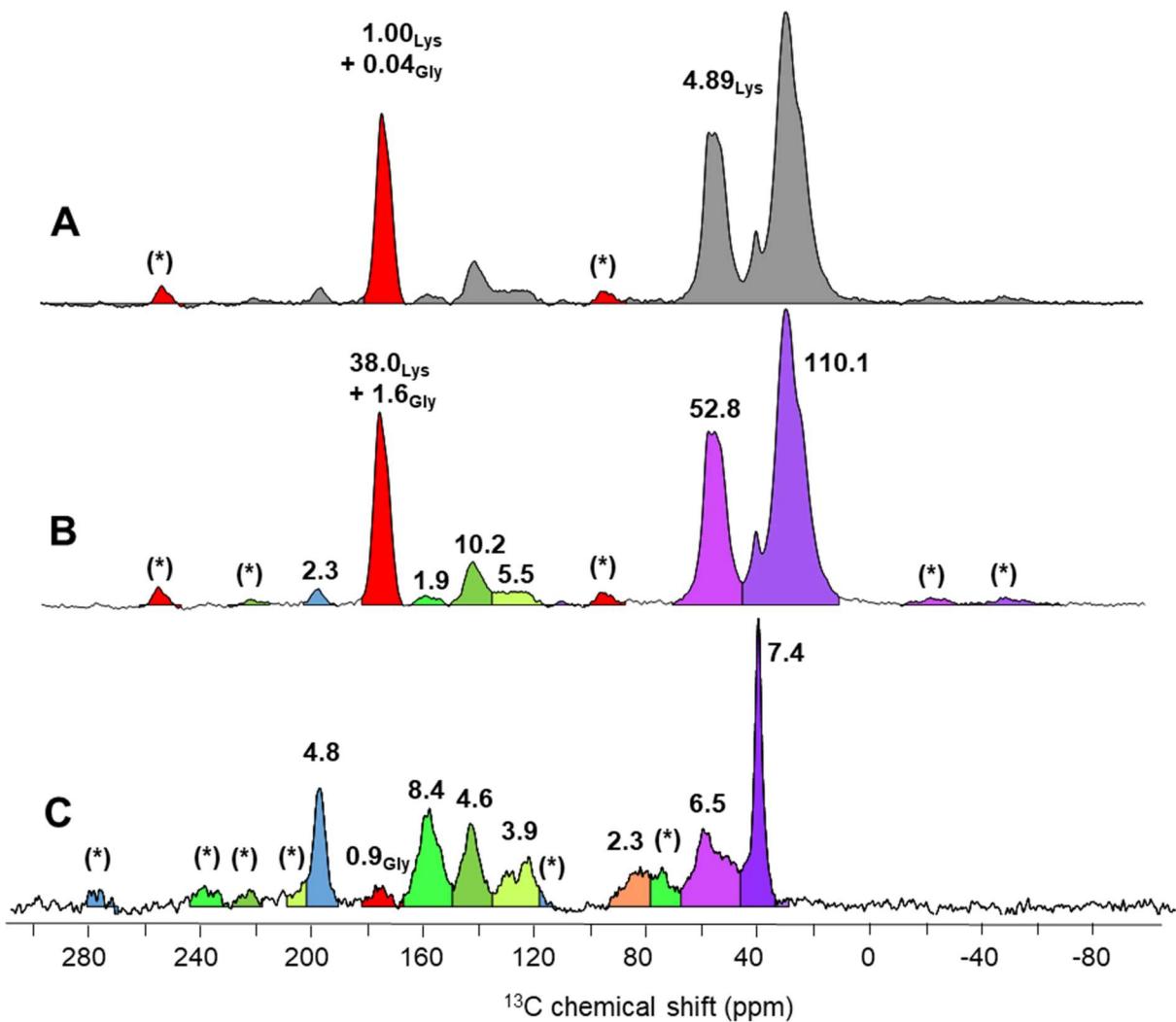
Figure 4.8: Integrated difference spectra of quantitative ^{13}C-$\{^1$H$\}$ CP/MAS NMR experiments.
(A) QUCP (difference) spectrum of [U- ^{13}C,^{15}N-Lys] elastin. The total area of the backbone carbonyls (red, 1.04) reflects the contributions of ^{13}CO-Lys (1.00) and ^{13}CO-Gly (0.04). All other sites (gray) have a total area of 4.89.
(B) Alternate illustration of integrated QUCP spectrum of [U-^{13}C,^{15}N-Lys] elastin (figure 4.8A). The total (centerband + ssbs) intensity of the backbone carbonyls (red) is scaled to 39.6 (^{13}CO-Lys, 38.0; ^{13}CO-Gly, 1.6).
(C) ComPmultiCP spectrum of [ε-^{13}C-Lys] elastin showing integration of resolved regions. Sum of the integrated areas (excluding ^{13}CO-Gly) is adjusted to 38.0.
Spectra were obtained at -20°C with 1024 scans (A, B) and (C) 1536 scans, and then processed with 40 Hz line broadening. Integrated areas are shown above respective peaks and include the area of any corresponding spinning sideband (*). Each spectrum is scaled to its tallest peak. Color scheme in B and C denotes different chemical environments of ^{13}C (figure 4.6).

QUCP of [U-^{13}C,^{15}N-Lys] elastin and ComPmultiCP of [ε-^{13}C-Lys] elastin spectra provide quantitative information about the destinations of the labels from the enriched lysine, i.e., unmodified lysines and the crosslinks (figure 4.8A-C). Areas under each resolved site are integrated to reflect the relative populations of ^{13}C in each chemical environment. ^{13}CO-Lys feature is normalized to 38.0 to reflect the number of lysine equivalents per 836 residues in the NRSMC tropoelastin (TE) monomer (figure 4.8B). Similarly, the sum of all the integrated areas (excluding ^{13}CO-Gly) is adjusted to 38.0 in the ComPmultiCP spectrum of [ε-^{13}C-Lys] elastin (figure 4.8C).

The experimental results for both labeling schemes are summarized and compared to the theoretical values for elastin with no crosslinks and with all Lys residues modified to (I)DES or AA (tables 4.3 - 4.4). Each lysine equivalent consists of 6 carbons, hence a total of $38 \times 6 = 228$ ^{13}C sites is expected in the spectrum of [U-^{13}C,^{15}N-Lys] elastin. The presented approach accounts for 221 of these (table 4.3 column I), which corresponds to ~3% experimental error, and this value is used to determine the uncertainty of the integrated areas in QUCP. The uncertainty of the ComPmultiCP results for [ε-^{13}C-Lys] elastin (3%) is based on the reported enhancement factors.[3]

Table 4.3: Integrated areas in QUCP spectrum of [U-^{13}C,^{15}N-Lys] elastin.
^{13}C chemical shift range used in QUCP spectrum (figure 4.8B) includes spinning sidebands for the respective centerbands (shown in parenthesis). The area of ^{13}CO$_{BB}$-Lys (marked with *) is normalized to 38.0. Color denotes chemical environments (figure 4.8).

^{13}C chemical shift range (ppm)	Integrated areas			
	I	Theoretical		
		II	III	IV
	Experimental (figure 4.8B)	No crosslinks	All Lys modified to (I)DES	All Lys modified to AA
10 - 45 (-70 - (-35), 103 - 115)	110.1 ± 3.3	152	95.0	95
45 - 70 (-35 - (-12))	52.8 ± 1.6	38	47.5	38
115 - 135	5.5 ± 0.2	0	0.0	0
135 - 150 (215 - 230)	10.2 ± 0.3	0	38.0	19
150 - 165	1.9 ± 0.1	0	9.5	19
167 - 182 (87 - 102, 247 - 262)	38.0* ± 1.1	38	38.0	38
191 - 203	2.3 ± 0.1	0	0.0	19
TOTAL	**221** ± 6.6	**228**	**228**	**228**

Table 4.4: Integrated areas in ComPmultiCP spectrum of [ε-^{13}C-Lys] elastin.
^{13}C chemical shift range used in QUCP spectrum (figure 4.8C) includes spinning sidebands for the respective centerbands (shown in parenthesis). Color denotes chemical environments (figures 4.6 and 4.8). The contribution of ^{13}CO-Gly (red) is not included in the sum (column I).

^{13}C chemical shift range (ppm)	Integrated areas			
	I	Theoretical		
		II	III	IV
	Experimental (figure 4.8C)	No crosslinks	All Lys modified to (I)DES	All Lys modified to AA
34 - 46	7.4 ± 0.2	38	0.0	0
46 - 68	6.5 ± 0.2	0	9.5	0
79 - 95	2.3 ± 0.1	0	0.0	0
119 - 136 (203 - 210)	3.9 ± 0.1	0	0.0	0
135 - 150 (219 – 230)	4.6 ± 0.1	0	19.0	0
150 - 165 (68 - 79, 231 – 245)	8.4 ± 0.3	0	9.5	19
167 - 183	0.9 ± 0.1	0	0.0	0
191 - 203 (113 - 119, 271 – 282)	4.8 ± 0.1	0	0.0	19
TOTAL	**38.1** ± 1.1	**38**	**38**	**38**

The broad feature with the point of highest intensity at 29 ppm, has an area of 110 and originates from the aliphatic carbons (spectrum 4.8B, purple). ^{13}Cε and ^{13}Cδ carbons of lysine are shifted downfield when in a crosslink. Without crosslinking, the total area of this region is 4 times that of ^{13}CO-Lys, to account for four aliphatic carbons: ^{13}Cβ, ^{13}Cγ, ^{13}Cδ, and ^{13}Cε (table 4.3). The experimental value of 110 is lower than the expected 152 if no crosslinks were present, which confirms post-translational modification. This aliphatic region in the spectrum of [ε-^{13}C-Lys] elastin (figure 4.8C) only has one sharp peak at 40 ppm, assigned as ^{13}Cε in unmodified lysine. According to the spectral data, 7.4 Lys residues per TE monomer do not undergo post-translational modification. This value is higher than the reported 5 unmodified lysine residues per TE monomer after 8 weeks of NRSMC cell culture.[14]

The ^{13}Cα-Lys lineshape (45 – 70 ppm) is broad, with 2 peaks at 55 and 57 ppm (figure 4.8B only, pink). It includes an unresolved contribution from ^{13}Cε-Lys in crosslinks. The latter population is clearly observed in [ε-^{13}C-Lys] elastin spectra at 60 ppm (4.8C, pink). Without crosslinking, the ratio of the areas ^{13}CO-Lys to ^{13}Cα-Lys is 38 : 38, or 1 : 1 (table 4.3). Due to ^{13}Cε contributions to the ^{13}Cα region, the experimental ratio is 38 : 53 (~1 : 1.4).

A broad feature between 115 and 135 ppm is observed in the spectrum of [U-^{13}C,^{15}N-Lys] elastin (figure 4.8B, light green). It is assigned to MDES (as confirmed by CP TOSS) as well as the spinning sidebands of the aldehyde (blue) and ^{13}Cα (purple). The integrated area is 5.5, but due to overlap with the sidebands, MDES content cannot be established. In the [ε-^{13}C-Lys] spectrum, this feature has a peak at 123 ppm, and the integrated area is 3.9 (figure 4.8C, light green). The smaller area is explained by the loss of contribution from ^{13}Cα spinning sideband.

The peaks at 142 and 159 ppm of the QUCP spectrum of [U-^{13}C,^{15}N-Lys] elastin (figure 4.8B, dark green) are primarily assigned to the pyridinium ring carbons of DES and IDES. In [ε-^{13}C-Lys] elastin, these two populations are present in a 8.4 : 4.6 (~2 : 1) ratio, whereas the upfield component is greater in the [U-^{13}C,^{15}N-Lys] sample (1.9 : 10.2, or ~ 1 : 5). If the only crosslinks formed were DES and IDES, then the ratio of the two peaks would be 1 : 4 in [U-^{13}C,^{15}N-Lys] elastin and 1 : 2 in [ε-^{13}C-Lys] elastin. Therefore, the uniformly labeled sample contains the tetrafunctional crosslinks almost exclusively. In contrast, the integrated areas of the two lineshapes of the [ε-^{13}C-Lys] sample reflect additional contributions from other crosslinks, likely DLNL, AA, DMDES, PEN, or NDES. The differences between samples are expected, as others have shown that crosslinking in elastin varies from sample to sample, with some exhibiting, e.g., greater DES (over IDES) or greater bifunctional (vs. tetrafunctional) moieties.[12,15,16]

The integrated areas of the (mostly) pyridinium carbons are used to quantify the crosslinks in elastin (table 4.5). If all carbons in the 150 - 165 ppm region of [U-^{13}C,^{15}N-Lys] elastin originate from DES and IDES, the normalized area of 1.9 corresponds to 7.6 Lys equivalents of (I)DES per TE monomer. If, instead, the 135 - 150 ppm region is used for the analysis, again assuming no contribution from bi- and trifunctional crosslinks, then there are 2.5 (I)DES residues (or 10.2 Lys equivalents) per monomer. Similarly, the peak at 143 ppm in the [ε-^{13}C-Lys] spectrum has area of 4.6 (figure 4.8C), which translates to 9.2 Lys equivalents of (I)DES per TE monomer. The feature between 150 and 165 ppm has implies 8.4 (I)DES residues (33.6 Lys equivalents) per TE monomer, which is not probable. This overrepresentation cannot be attributed to an experimental error as ^{13}Cε in the (I)DES ring that is assigned to this peak is nonprotonated. The magnetization is initially transferred from ^{1}H nuclei in the ComPmultiCP experiment, meaning this ^{13}C population is likely to be underrepresented if the experiment was not truly quantitative. The most likely cause of the discrepancy between the two labeled samples is a larger number of bifunctional crosslinks in [ε-^{13}C-Lys] elastin than in the [U-^{13}C,^{15}N-Lys] elastin.

Table 4.5: Quantification of unmodified Lys, (I)DES, and AA in elastin.
Chemical shift range used for the integrals is provided in the first column. Color denotes different chemical environments (figure 4.6, 4.8).

δ (^{13}C) (ppm)	Integrated area[v]		Assignment	Number of Lys equiv. per TE monomer	
	[U-^{13}C,^{15}N-Lys] elastin	[ε-^{13}C-Lys] elastin		[U-^{13}C,^{15}N-Lys] elastin	[ε-^{13}C-Lys] elastin
34 – 46	N/A	7.4 ± 0.2	Lys	N/A	7.4 ± 0.2
135 – 150	10.2 ± 0.3	4.6 ± 0.1	(I)DES	10.2 ± 0.3	9.2 ± 0.2
150 – 165	1.9 ± 0.1	8.4 ± 0.3	(I)DES	7.6 ± 0.4	33.6 ± 1.0
191 – 203	2.3 ± 0.1	4.8 ± 0.1	AA	4.6 ± 0.2	9.6 ± 0.2

The reported quantity of (I)DES ranges from 1.7 to 12.5 Lys equivalents per TE monomer, depending on the quantification method, tissue type, species, and age of the animal.[17,18] The results of NMR quantification are within this range (table 4.5), except for the discussed outlier. Amino acid analysis of NRSMC elastin determined 6 Lys equivalents per TE monomer.[14] The values obtained in ssNMR may be higher as a result of the non-destructive methods used, and signals originating from bifunctional crosslinks contributing to the heteroaromatic region of the spectrum.

The peaks at 198 ppm in both spectra (figure 4.8B and 4.8C, blue) are assigned to an aldehyde. The interpretation of the area under these features depends on the assignment. The content of the highly reactive Lya is considerable only during the early stages of growth and is generally low otherwise. Therefore, AA is more likely source of this peak, and the area corresponds to 4.6 lysine equivalents in [U-^{13}C,^{15}N-Lys] elastin and 9.6 in [ε-^{13}C -Lys] elastin (table 4.5). The quantity is larger in the latter, providing further evidence for higher number of bifunctional crosslinks in that sample. The reported values for AA range from 1.2 to 5.8 Lys equivalents.[17,19]

---

[v] From Figure 4.8.

The small broad peak between 79 and 95 ppm in [ε-^{13}C-Lys] elastin spectrum is not a spinning sideband. Presence in CP TOSS experiments, as well as spectra recorded at a different spinning rate (figure 4.7), reject this hypothesis. None of the reported crosslinking compounds in the [ε-^{13}C-Lys] elastin sample has a chemical shift in this region. The area (2.3) is included in the total normalized to 38.0. The chemical shift is consistent with a hydroxyl group attached to ^{13}Cε, which may be a result of a side reaction during the protein purification procedures.

The ComPmultiCP spectrum, unlike QUCP of the uniformly labeled sample, allows for a clear observation of scrambling to ^{13}CO-Gly (figure 4.8C). When the difference is generated to cancel out the signals of the natural-abundance aliphatics (^{13}Cα, ^{13}Cβ, ^{13}Cγ), a small signal remains in the ^{13}CO$_{BB}$ region. This feature corresponds to 0.9 glycine residues per TE monomer, which is lower than the enrichment level determined by HPLC-MS assay, but within the same order of magnitude. The results are consistent with scrambling to ^{13}CO-Gly through the glycine synthase pathway.

The exact identities and quantities of bi- and trifunctional crosslinks are ambiguous, given spectral overlap and the various possible products of the post-translational modification. However, presence of a multitude of lysine derivatives is undeniable. Theoretical considerations of crosslinked tropoelastin with only one type of modification (either (I)DES or AA) prove that a single compound cannot account for the number of the detected ^{13}C populations (tables 4.3 – 4.4).

Overall, it is demonstrated that ssNMR can be used for observation and quantification of crosslinks in elastin without the use of destructive methods. Unlike the previous crosslinking studies, the reported values do not have the bias resulting from hydrolysis, and other harsh processing techniques. The main drawback of the presented approach is the spectral overlap. However, the data from the two labeling schemes gives insights into the complex system of crosslinks.

4.3.4 ^{15}N NMR spectra of [U-^{13}C,^{15}N-Lys] elastin

Similar to ^{13}C, the ^{15}N chemical shifts are sensitive to the environment, and can be used to assign the secondary structures. ^{15}Nε -Lys and ^{15}NH-Lys features are detected in CP echo at -20°C and DP echo at 37°C of [U-^{13}C,^{15}N-Lys] elastin (figure 4.9). In both spectra, the peaks are centered at 32 and 118 ppm, respectively. The ^{15}Nε -Lys feature is narrow in the DP spectrum (FWHM = 68 Hz) and has a much lower intensity in the CP experiment, confirming high mobility of this site.
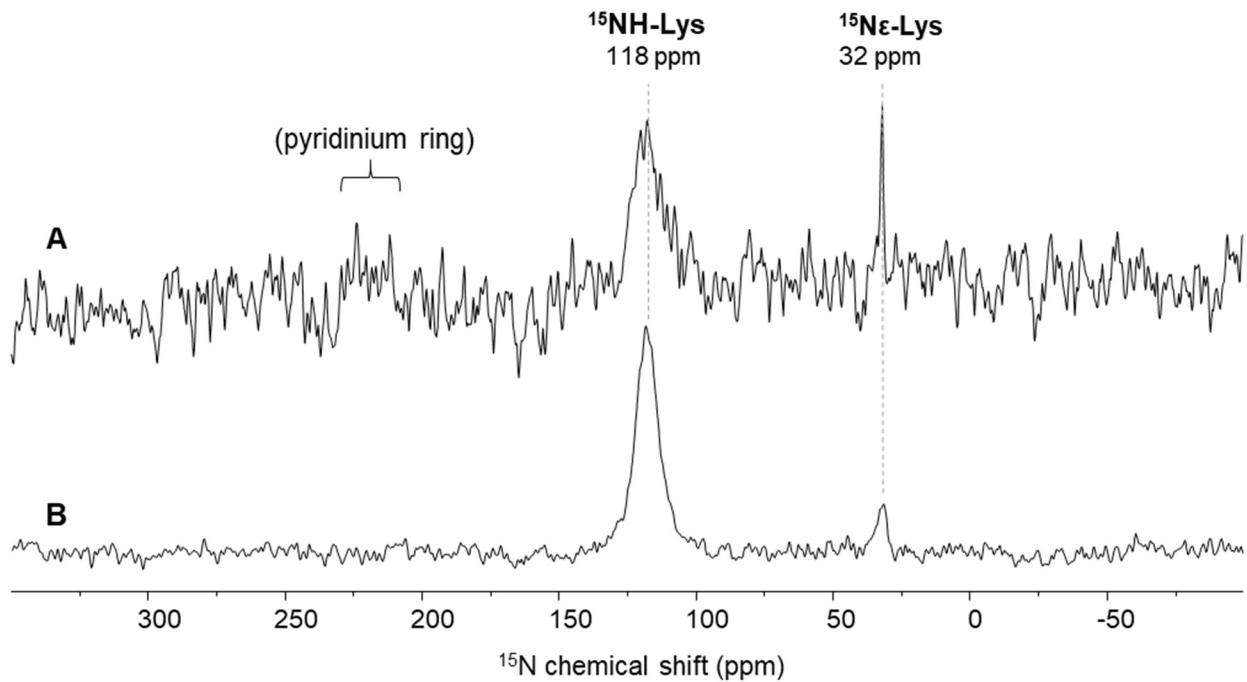
Figure 4.9: ^{15}N NMR spectra of [U-^{13}C,^{15}N-Lys] elastin.
(A) DP echo spectrum acquired at 37°C and 1382 scans. (B) CP echo spectrum obtained at -20°C and 4096 scans. Dashed lines correspond to the highest peak positions in (A). Both spectra were processed with 40 Hz line broadening, and their intensities were normalized to the tallest peak.

The ^{15}N chemical shifts of crosslinks are resolved from the amino group and the ^{15}Nε of unmodified Lys. The expected ^{15}N chemical shift range for the pyridinium ring of DES and IDES is 210 – 230 ppm, based on the calculated chemical shift for a model compound with this moiety, N-ethylpyridine (figure 4.10).[20] No signal in this range is detected with CP. Results for DP are inconclusive due to poor S/N.
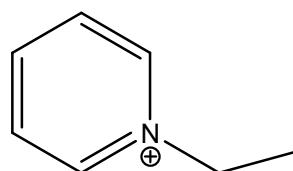
Figure 4.10: N-ethylpyridine used for prediction of ^{15}N chemical shift.

The reason for poor signal intensities is the low sensitivity of the ^{15}N nuclei (table 1.2). The relative sensitivity of ^{15}N is only 40% of that of ^{13}C. Given that ^{13}C signals for most of the crosslinks are weak, the poor S/N in ^{15}N spectra is expected. The ^{15}N signals are too weak to be detected, even with uniform isotopic enrichment.

4.4 Summary and conclusions

Bi-, tri-, and tetrafunctional crosslinks are observed and quantified in ssNMR spectra of intact, native elastin. Quantification of (I)DES, AA, and unmodified lysine is performed with CP-based measurements. The unmodified lysine content in [ε-^{13}C-Lys] elastin is 7.4 residues per TE monomer. Both QUCP and ComPmultiCP detect an aldehyde peak, tentatively assigned as AA (4.6 - 9.6 Lys equivalents). The quantity of (I)DES was 7.6 - 10.2 Lys equiv. per TE monomer. About 6.4 Lys equivalents of ^{13}Cε is shifted downfield when in crosslink, overlaying with the signals of ^{13}Cα. The presence of MDES is confirmed by CP TOSS, but this crosslink is not quantified due to overlapping signals with spinning sidebands in the quantitative experiments. No characteristic aliphatic ^{13}Cε signals for ADES and CPN were observed, implying these crosslinks were not present in the NRSMC sample.

Scrambling to ^{13}CO-Gly in [ε-^{13}C-Lys] elastin, which was determined by the HPLC-MS assay in Chapter 2, is confirmed by the ssNMR data. According to the ComPmultiCP spectra, about 0.9 Gly residues per TE monomer are enriched. Scrambling to ^{13}CO-Gly in [U-^{13}C,^{15}N-Lys] elastin could not be determined, as ^{13}CO-Gly is not resolved from ^{13}CO$_{BB}$-Lys.

^{15}N ssNMR spectra of [U-^{13}C,^{15}N-Lys] elastin do not provide sufficient intensities for studies of the rare crosslinks. The naturally low relative sensitivity of ^{15}N nuclei does not allow for a sufficient signal-to-noise ratio for conclusive results. Peaks corresponding to ^{15}N in unmodified Lys, both in the backbone and ^{15}Nε are observed. ^{15}N CP Echo at -20°C offers better signal intensities than the ^{15}N DP Echo experiment at 37°C.

Overall, it is demonstrated that ssNMR can be used for quantification of (I)DES in native elastin without employing destructive methods. The exact identity and quantity of bi- and trifunctional crosslinks is ambiguous, given the overlapping signals and multiple possible compounds. However, the presence of these modification products indisputable.

4.5 References

[1] Fu, R.; Hu, J.; Cross, T. A. Towards quantitative measurements in solid-state CPMAS NMR: A Lee–Goldburg frequency modulated cross-polarization scheme. *J. Magn. Reson.* **2004**, *168*, 8-17.

[2] Johnson, R. L.; Schmidt-Rohr, K. Quantitative solid-state C-13 NMR with signal enhancement by multiple cross polarization. *J. Magn. Reson.* **2014**, *239*, 44-49.

[3] Duan, P.; Schmidt-Rohr, K. Composite-pulse and partially dipolar dephased ComPmultiCP for improved quantitative solid-state C-13 NMR. *J.Mag. Res.* **2017**, *285*, 68-78.

[4] Hou, G. J.; Deng, F.; Ding, S. W.; Fu, R. Q.; Yang, J.; Ye, C. H. Quantitative cross-polarization NMR spectroscopy in uniformly C-13-labeled solids. *Chem. Phys. Lett.* **2006**, *421*, 356-360.

[5] Dixon, W. T.; Schaefer, J.; Sefcik, M. D.; Stejskal, E. O.; McKay, R. A. Total suppression of sidebands in CPMAS C-13 NMR. *J. Magn. Reson.* **1982**, *49*, 341-345.

[6] Takegoshi, K.; Nakamura, S.; Terao, T. ^{13}C-^{1}H dipolar-assisted rotational resonance in magic-angle spinning NMR. *Chem. Phys. Lett.* **2001**, *344*, 631-637.

[7] Turino, G. M.; Ma, S.; Lin, Y. Y.; Cantor, J. O.; Luisetti, M. Matrix elastin: a promising biomarker for chronic obstructive pulmonary disease. *Am. J. Res. Crit. Care Med.* **2011**, *184*, 637-641.

[8] Usuki, T.; Yamada, H.; Hayashi, T.; Yanuma, H.; Koseki, Y.; Suzuki, N.; Masuyama, Y.; Lin, Y. Y. Total synthesis of COPD biomarker desmosine that crosslinks elastin. *Chem. Commun.* **2012**, *48*, 3233-3235.

[9] Sugimura, T.; Komatsu, A.; Koseki, Y.; Usuki, T. Pr(OTf)3-promoted Chichibabin pyridine synthesis of isodesmosine in H2O/MeOH. *Tetrahedron Lett.* **2014**, *55*, 6343-6346.

[10] Perry, A.; Stypa, M. P.; Tenn, B. K.; Kumashiro, K. K. Solid-state ^{13}C NMR reveals effects of temperature and hydration on elastin. *Biophys. J.* **2002**, *82*, 1086-1095.

[11] Umeda, H.; Kawamorita, K.; Suyama, K. High-performance liquid chromatographic quantification of allysine as bis- p -cresol derivative in elastin. *Amino Acids* **2001**, *20*, 187-199.

[12] Schräder, C.; Heinz, A.; Majovsky, P.; Karaman Mayack, B.; Brinckmann, J.; Sippl, W.; Schmelzer, C. Elastin is heterogeneously cross-linked. *J. Biol. Chem.* **2018**, *293*(39), 15107-15119.

[13] Akagawa, M.; Yamazaki, K.; Suyama, K. Cyclopentenosine, Major Trifunctional Crosslinking Amino Acid Isolated from Acid Hydrolysate of Elastin. *Arch. Biochem. Biophys.* **1999**, *372*(1), 112-120.

[14] Barone, L. M.; Faris, B.; Chipman, S. D.; Toselli, P.; Oakes, B. W.; Franzblau, C. Alteration of the extracellular matrix of smooth muscle cells by ascorbate treatment. *Biochim. Biophys. Acta* **1985**, *840*(2), 245-254.

[15] Watanabe, M.; Sawai, T.; Nagura, H.; Suyama, K. Age-related alteration of cross-linking amino acids of elastin in human aorta. *Tohoku J. Exp. Med.* **1996**, *180*, 115-130.

[16] Umeda, H.; Kawamorita, K.; Suyama, K. High-performance liquid chromatographic quantification of allysine as bis- p -cresol derivative in elastin. *Amino Acids* **2001**, *20*(2), 187-199.

[17] Francis, G.; John, R.; Thomas, J. Biosynthetic pathway of desmosines in elastin. *Biochem. J.* **1973**, *136*(1), 45-55.

[18] Field, J. M.; Rodger, G. W.; Hunter, J. C.; Serafini-Fracassini, A.; Spina, M., Isolation of elastin from bovine auricular cartilage. *Arch. Biochem. Biophys.* **1978**, *191*, 705-713.

[19] Lent, R. W.; Smith, B.; Salcedo, L. L.; Faris, B.; Franzblau, C. Studies on the reduction of elastin. II. Evidence for the presence of alpha-aminoadipic acid delta-semialdehyde and its aldol condensation product. *Biochemistry* **1969**, *8*(7), 2837-2845.

[20] W. Robien, NMRPREDICT-Server, entry ID: NNMR_UWRH000907; http://nmrpredict.orc.univie.ac.at/, (accessed 10/9/2019).

CHAPTER 5

SUMMARY AND CONCLUDING REMARKS

The elasticity and resilience of elastin are crucial for the proper function of vertebrate tissue. In particular, the importance of lysine-derived crosslinks is clear. Undercrosslinked elastin – associated with genetic disorders[1], induced copper-deficiency[2], or inhibition of LOX activity[3] – lacks the elasticity that is critical for various life processes.[4,5] To better understand the role of crosslinks, various studies focused on isolation, characterization, and quantification of these lysine derivatives.[6,7,8,9,10] However, none of these approaches utilized intact protein, and the analysis of crosslinks after hydrolysis or enzymatic cleavage may have limited accuracy.[11]

Strategic isotopic labeling paired with ssNMR is a versatile approach that provides quantitative, structural, and dynamical information on native elastin's lysines and lysine-derived crosslinks. The four major amino acids of this protein have been enriched using the neonatal rat smooth muscle cell (NRSMC) line and then characterized by ssNMR spectroscopy.[12,13,14] However, unlike the lysines in this study, the glycines, alanines, prolines, and valines of elastin do not undergo post-translational modification. The enrichment of lysine is straightforward since this amino acid is essential; direct substitution results in high incorporation of the isotopically labeled residues (>87%). However, the assessment of the labeling – how much of the lysines and their derivatives are enriched, and if the other residues were affected – required a modified approach. The analysis of the MS results had to account for partially enriched crosslinks and products of incomplete Marfey's derivatization. Also, the possible sites of scrambling were identified based on the catabolic pathways of lysine, but only minor enrichment of glycine (<2%) was detected.

^{13}C chemical shifts from one- and two-dimensional (1D) measurements led to tentative assignments of lysines in the contexts of sequence and secondary structure (Chapter 3). In the simplest 1D experiments, CP selects rigid regions, rINEPT enhances mobile domains, and DP reflects all ^{13}C sites, regardless of mobility. The lineshapes of the major features indicate that there are at least two dominant populations, assigned as random coil and α-helix. The helical content increases as the temperature is lowered to -20°C. These results are consistent with the previous studies of alanines, which reported helix and coil to be the predominant conformations of the crosslinking domains and higher helical content at -20°C.[13] Alanine is the most common neighboring residue of lysine, hence structural similarity is expected.

Two-dimensional (2D) correlation spectroscopy of the crowded ^{13}CO-^{13}Cα regions reveals at least three populations of lysine. Comparison to the RefDB[15] chemical shifts of ^{13}CO- and ^{13}Cα-Lys in different secondary structures confirms the conclusions drawn from the 1D experiments. The observed populations are assigned based on the semi-empirical approach that accounts for both sequence- and structure-dependent effects. The chemical shifts of KA and KX (where X = Y, V, L, S, and R) are resolved from those of KP. The predicted values for KA and KX in both α-helix and random coil as well as KP in coil match the three major populations detected in the 2D DARR spectrum. No KP pairs in β-sheet conformation were observed. The 2D experiment complements the 1D measurements as it reflects ^{13}CO-^{13}Cα of lysine only and provides more resolution.

The complex system of lysine-derived crosslinks was characterized by quantitative experiments performed on [U-^{13}C,^{15}N-Lys] elastin and [ε-^{13}C-Lys] elastin. In quantitative cross-polarization (QUCP)[16] and composite pulse multiple CP (ComPmultiCP)[17] spectra, the integrated peak areas are proportional to the number of spins in each chemical environment, unlike "conventional" CP (Chapter 3). (CP at -20°C is an optimal excitation scheme for observation of the rare crosslinks, but these spectra over-represent protonated carbons.)

The number(s) of unmodified lysine, allysine aldol, and (iso)desmosine per tropoelastin (TE) monomer in the labeled samples of native elastin were determined, and other bi- and trifunctional crosslinks were detected. Elastin's unique desmosine and isodesmosine constituted 7.6 – 10.2 Lys equivalents per TE monomer. About 7 residues of unmodified lysine were detected. Spectra of [ε-^{13}C-Lys] elastin confirmed that the signal at 198 ppm originates from ^{13}Cε and that the most likely crosslink was allysine aldol (4.6 – 9.6 Lys equivalents).

In summary, solid-state NMR offers valuable insights into the structure of native elastin, focusing here on the lysines and the crosslinking domains in which they are found. Quantification of lysines and the crosslinks in native elastin by ssNMR yielded results that were consistent with the previously reported values for hydrolysates[18,19,20,21]; notably, it is the first known quantitative report for intact elastin. Quantification of Lys and the crosslinks by LC/MS and ssNMR may be extended to more elastin samples with more complex labeling schemes involving this amino acid, which will further enhance our understanding of these regions that are critical for normal elastic fiber function.

5.1 References

[1] Gara S,; Litaiem N. Cutis Laxa (Elastolysis) [Updated 2018 Dec 2]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2018 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK532944/

[2] Shields, G. S.; Coulson, W. F.; Kimball, D. A.; Carnes, W. H.; Cartwright, G. E.; Wintrobe, M. M. Studies on copper metabolism. 32. Cardiovascular lesions in copper-deficient swine. *Am. J. Pathol.* **1962**, *41*, 603-621.

[3] Schmelzer, C. E. H.; Heinz, A.; Troilo, H.; Lockhart-Cairns, M. P.; Jowitt, T. A.; Marchand, M. F.; Bidault, L.; Bignon, M.; Hedtke, T.; Barret, A.; McConnell, J. C.; Sherratt, M. J.; Germain, S.; Hulmes, D. J. S.; Baldock, C.; Muller, L. Lysyl oxidase-like 2 (LOXL2)-mediated cross-linking of tropoelastin. *FASEB* **2019**, *33*, 5468-5481.

[4] Starcher, B.; Hill, C. H.; Matrone, G. Importance of dietary copper in the formation of aortic elastin, *J. Nutrition* **1964**, *82*, 318-322.

[5] Miller, E. J.; Martin, G. R.; Mecca, C. E.; Piez, K. A. The biosynthesis of elastin crosslinks. The effect of copper deficiency and a lathyrogen, *J. Biol. Chem.* **1965**, *240*, 3623-3627.

[6] Miller, E. J.; Martin, G. R.; Piaz, K. A. The utilization of lysine in the biosynthesis of elastin crosslinks. *Biochem. Biophys. Res. Commun.* **1964**, *17*, 248-253.

[7] Snider, R.; Faris, B.; Verbitzki, V.; Moscaritolo, R.; Salcedo, L. L.; Franzblau, C. Elastin biosynthesis and cross-link formation in rabbit aortic smooth muscle cell cultures. *Biochemistry* **1981**, *20*, 2614-2618.

[8] Franzblau, C.; Sinex, F. M.; Faris, B.; Lampidis, R. Identification of a new crosslinking amino acid in elastin. *Biochem. Biophys. Res. Commun.* **1965**, *21*, 575-581.

[9] Schräder, C.; Heinz, A.; Majovsky, P.; Karaman Mayack, B.; Brinckmann, J.; Sippl, W.; Schmelzer, C. Elastin is heterogeneously cross-linked. *J. Biol. Chem.* **2018**, *293*(39), 15107-15119.

[10] Umeda, H.; Kawamorita, K.; Suyama, K. High-performance liquid chromatographic quantification of allysine as bis- p -cresol derivative in elastin. *Amino Acids* **2001**, *20*(2), 187-199.

[11] Guay, M.; Lamy, F. The troublesome crosslinks of elastin. Trends Biochem. Sci. **1979**, *4*, 160-164.

[12] Ohgo, K.; Dabalos, C. L.; Kumashiro, K. K. Solid-State NMR Spectroscopy and Isotopic Labeling Target Abundant Dipeptide Sequences in Elastin's Hydrophobic Domains. *Macromolecules* **2018**, *51*, 2145-2156.

[13] Djajamuliadi, J.; Ohgo, K.; Kumashiro, K. K. Targeting Alanines in the Hydrophobic and Cross-Linking Domains of Native Elastin with Isotopic Enrichment and Solid-State NMR Spectroscopy. *Macromolecules* **2018**, *51*, 2157-2168.

[14] Perry, A.; Stypa, M. P.; Foster, J. A.; Kumashiro, K. K. Observation of the glycines in elastin using ^{13}C and ^{15}N solid-state NMR spectroscopy and isotopic labeling. *J. Am. Chem. Soc.* **2002**, *124*, 6832-6833.

[15] Neal, S.; Nip, A. M.; Zhang, H.; Wishart, D. S. Rapid and accurate calculation of protein ^{1}H, ^{13}C, ^{15}N chemical shifts. *J. Biomol. NMR* **2003**, *26*, 215-240.

[16] Hou, G. J.; Deng, F.; Ding, S. W.; Fu, R. Q.; Yang, J.; Ye, C. H. Quantitative cross-polarization NMR spectroscopy in uniformly C-13-labeled solids. *Chem. Phys. Lett.* **2006**, *421*, 356-360.

[17] Johnson, R. L.; Schmidt-Rohr, K. Quantitative solid-state C-13 NMR with signal enhancement by multiple cross polarization. *J. Magn. Reson.* **2014**, *239*, 44-49.

[18] Francis, G.; John, R.; Thomas, J. Biosynthetic pathway of desmosines in elastin. *Biochem. J.* **1973**, *136*(1), 45-55.

[19] Field, J. M.; Rodger, G. W.; Hunter, J. C.; Serafini-Fracassini, A.; Spina, M. Isolation of elastin from bovine auricular cartilage. *Arch. Biochem. Biophys.* **1978**, *191*, 705-713.

[20] Barone, L. M.; Faris, B.; Chipman, S. D.; Toselli, P.; Oakes, B. W.; Franzblau, C. Alteration of the extracellular matrix of smooth muscle cells by ascorbate treatment. *Biochim. Biophys. Acta* **1985**, *840*(2), 245-254.

[21] Lent, R. W.; Smith, B.; Salcedo, L. L.; Faris, B.; Franzblau, C. Studies on the reduction of elastin. II. Evidence for the presence of alpha-aminoadipic acid delta-semialdehyde and its aldol condensation product. *Biochemistry* **1969**, *8*(7), 2837-2845.

APPENDIX I

COMPOSITION OF NRSMC GROWTH MEDIA

Table AI.1: Composition of Growth Media for NRSMC Cultures.

Component	Manufacturer	Catalog Number	% by Volume
Dulbecco's Modified Eagle's Medium (DMEM) – low glucose	Sigma – Aldrich	D6046	87
Fetal Bovine Serum	GIBCO	26140	10
Sodium Pyruvate, 100mM (100x)	GIBCO	11360	1
Antibiotic – Antimycotic (100x)	GIBCO	15240	1
Non-Essential Amino Acids (NEAA) (100x)	GIBCO	11140	1

Table AI.2: Composition of Non-Essential Amino Acids (NEAA) (100x) (Gibco, #11140).

Amino Acid	Concentration (mg/L)
Glycine	750
L-alanine	890
L-asparagine	1320
L-aspartic acid	1330
L-glutamic acid	1470
L-proline	11150
L-serine	1050

Table AI.3: Composition of Dulbecco's Modified Eagle's Medium (DMEM)

Component	Concentration (g/L)	
	Standard DMEM (Sigma - Aldrich, #D6046)	**Lysine-free DMEM for enriched samples** (Gibco, #A2493901)
Inorganic Salts		
$CaCl_2$	0.2	0.2
$Fe(NO_3)_3 \cdot 9\ H_2O$	0.0001	0.0001
$MgSO_4$	0.09767	0.09767
KCl	0.4	0.4
$NaHCO_3$	3.7	3.7
NaCl	6.4	6.4
NaH_2PO_4	0.109	0.109
Amino Acids		
L-arginine • HCl	0.084	-
L-cysteine • 2 HCl	0.0626	0.0626
L-glutamine	0.584	-
Glycine	0.03	0.03
L-histidine • HCl • H_2O	0.042	0.042
L-isoleucine	0.105	0.105
L-leucine	0.105	0.105
L-lysine • HCl	0.146	-
L-methionine	0.03	0.03
L-phenylalanine	0.066	0.066
L-serine	0.042	0.042
L-threonine	0.095	0.095
L-tryptophan	0.016	0.016
L-tyrosine • 2 Na • 2 H_2O	0.10379	0.10379
L-valine	0.094	0.094
Vitamins		
Choline chloride	0.004	0.004
Folic acid	0.004	0.004
Myo-Inositol	0.0072	0.0072
Niacinamide	0.004	0.004
D-pantothenic acid • ½ Ca	0.004	0.004
Pyridoxine • HCl	0.00404	0.00404
Riboflavin	0.0004	0.0004
Thiamine • HCl	0.004	0.004
Other		
D-glucose	1.0	-
Phenol red • Na	0.0159	-
Pyruvic acid • Na	0.11	-

Table AI.4: Lysine-free DMEM Supplementation

Component	Concentration (g/L)	Cell Culture		
		$[U-^{13}C,^{15}N-Lys]$	$[\varepsilon-^{13}C-Lys]$	Control (Unenriched)
L-lysine • HCl	0.146	No	No	Yes
$[U-^{13}C,^{15}N-Lys]$ • 2 HCl	0.181*	Yes	No	No
$[\varepsilon-^{13}C-Lys]$ • 2 HCl	0.175*	No	Yes	No
L-arginine • HCl	0.084	Yes	Yes	Yes
L-glutamine	0.584	Yes	Yes	Yes
D-glucose	1.0	Yes	Yes	Yes
Phenol red • Na	0.11	Yes	Yes	Yes

* Adjusted for molar concentration of lysine

APPENDIX II

AMINO ACID COMPOSITION OF TROPOELASTIN FROM CULTURED NEONATAL RAT SMOOTH MUSCLE CELLS (NRSMC)

Figure AII.1: Tropoelastin Sequence.

GVPGAVPGGVPGGLPGGVPGGVYYPGAGIGGGLGGGALGPGGKPPKPGAGLLGAFGAGPGGLGGAG
PGAGGVLVPGGGAGAAAAYKAAAKAGAGLGGIGGVPGGVGVGGVPGAVGVGGVPGAVGGIGGIGGLG
VSTGAVVPQLGAGVGAGGKPGKVPGVGLPGVYPGGVLPGTGARFPGVGVLPGVPTGTGVKAKVPGGG
GGAFSGIPGVGPFGGQQPGVPLGYPIKAPKLPGGYGLPYTNGKLPYGVAGAGGKAGYPTGTGVGSQAA
VAAAKAAKYGAGGGGVLPGVGGGGIPGGAGAIPGIGGITGAGTPAAAAAAKAAAKAAKYGAAGGLVPGG
PGVRVPGAGIPGVGIPGVGGIPGVGGIPGVGGPGIGGPGIVGGPGAVSPAAAAKAAAKAAKYGA
RGGVGIPTYGVGAGGFPGYGVGAGAGLGGASQAAAAAAAAKAAKYGAGGAGTLGGLVPGAVPGALPG
AVPGALPGAVPGALPGAVPGVPGTGGVPGTPAAAAAAAAAKAAAKAGQYGLGPGVGGVPGGVGVGGL
PGGVGPGGVTGIGTGPGTGLVPGDLGGAGTPAAAKSAAKAAAKAQYGAAAGLGAGVPGLGVGAGVPG
FGAGAGGFGAGAGVPGFGAGAVPGSLAASKAAKYGAAGGLGGPGGLGGPGGLGGPGGLGGPGGLGG
PGGLGGVPGGVAGAPAAAAAAKAAAKAAQYGLGGAGGLGAGGLGAGGLGAGGLGAGGLGAGGLGAG
GVIPGAVGLGGVSPAAAAKAAKYGAAGLGGVLGARPFPGGGVAARPGFGLSPIYPGGGAGGLGVGGKP
PKPYGGALGALGYQGGGCFGKSCGRKRK

Table AII.1: Amino Acid Composition of NRSMC tropoelastin.[i]

Amino Acid	Number of Residues	% of Residues
Glycine	311	37.2
Alanine	164	19.6
Proline	93	11.1
Valine	75	9.0
Leucine	52	6.2
Lysine	38	4.5
Tyrosine	23	2.8
Isoleucine	21	2.5
Threonine	17	2.0
Serine	11	1.3
Phenylalanine	11	1.3
Glutamine	9	1.1
Arginine	7	0.8
Cysteine	2	0.2
Aspartic Acid	1	0.1
Asparagine	1	0.1
TOTAL:	836	100

[i] Pierce, R. A.; Deak, S. B.; Stolle, C. A.; Boyd, C. D., Heterogeneity of rat tropoelastin messenger-RNA revealed by cDNA cloning. *Biochemistry* **1990**, 29, 9677-9683.

APPENDIX III

PREDICTED ^{13}C CHEMICAL SHIFTS OF LYSINE-DERIVED CROSSLINKS

Mnova Best algorithm was used for all predicted chemical shift. Error reported by the software is included.

Table AIII.1: Predicted ^{13}C chemical shifts of lysine.

Atom	Shift (ppm)	Error (ppm)
1 C	177.7	3.0
3 CH	58.0	5.3
4 CH2	31.0	5.8
5 CH2	23.0	4.8
6 CH2	28.6	4.0
7 CH2	40.1	3.6
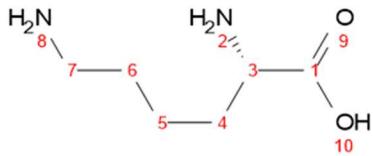
Table AIII.2: Predicted ^{13}C chemical shifts of allysine.

Atom	Shift (ppm)	Error (ppm)
1 C	177.7	3.0
2 CH	58.0	5.3
5 CH2	30.3	5.8
6 CH2	23.5	5.8
7 CH2	42.1	6.1
8 CH	200.9	4.4
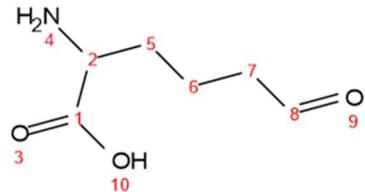
Table AIII.3: Predicted ^{13}C chemical shifts of allysine aldol.

Atom	Shift (ppm)	Error (ppm)
1 C	177.7	3.0
2 CH	58.0	5.3
5 CH2	30.3	5.8
6 CH2	25.0	5.8
7 CH2	29.3	5.8
8 CH	153.2	8.8
9 C	141.8	10.9
10 CH	194.6	4.6
12 CH2	26.3	5.5
13 CH2	29.5	5.8
14 CH	54.4	5.3
16 C	176.4	3.0
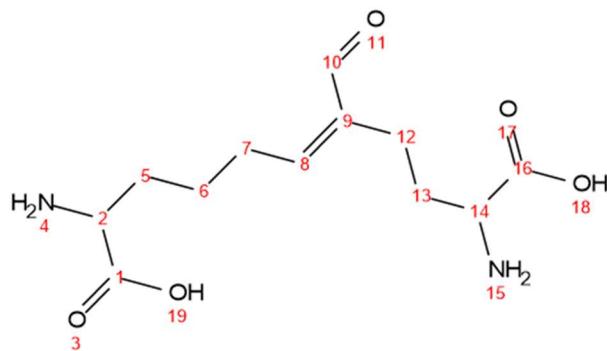
Table AIII.4: Predicted ^{13}C chemical shifts of dehydrolysinonorleucine.

Atom	Shift (ppm)	Error (ppm)
1 CH2	34.1	5.5
2 CH	160.3	7.7
4 CH2	56.6	6.2
5 CH2	28.9	5.8
6 CH2	23.6	5.8
7 CH2	30.3	5.8
8 CH	58.0	5.3
10 C	177.7	3.0
13 CH2	25.5	5.8
14 CH2	31.0	5.8
15 CH	58.1	5.3
17 C	177.8	3.0
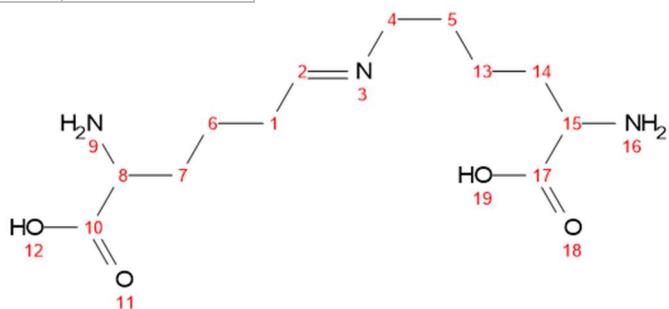
Table AIII.5: Predicted ^{13}C chemical shifts of lysinonorleucine.

Atom	Shift (ppm)	Error (ppm)
1 C	177.7	3.0
2 CH	58.0	5.3
3 CH2	31.0	5.8
4 CH2	23.0	4.8
5 CH2	28.1	5.8
6 CH2	48.3	5.3
10 CH2	48.3	5.3
11 CH2	28.1	5.8
12 CH2	23.0	4.8
13 CH2	31.0	5.8
14 CH	58.0	5.3
15 C	177.7	3.0

112
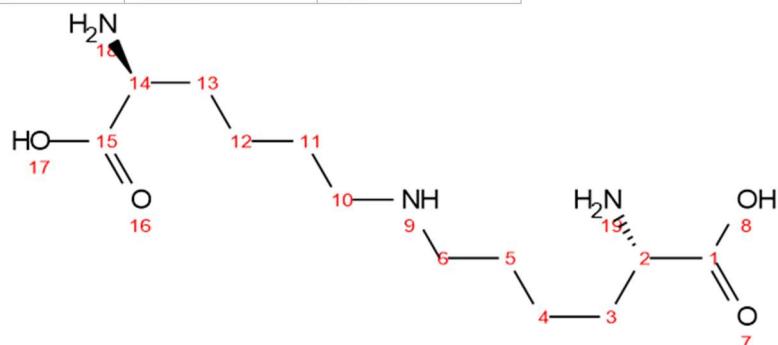
Table AIII.6: Predicted ^{13}C chemical shifts of merodesmosine.

Atom	Shift (ppm)	Error (ppm)
1 CH2	28.2	4.6
2 CH	127.6	6.9
3 C	133.6	11.6
4 CH2	52.2	4.8
6 CH2	49.0	6.2
7 CH2	28.3	4.8
8 CH2	23.6	3.7
9 CH2	30.7	5.9
10 CH	56.1	5.6
12 C	177.0	4.9
15 CH2	32.0	3.6
16 CH2	29.7	5.8
17 CH	54.9	5.6
19 C	176.3	4.9
22 CH2	25.2	4.3
23 CH2	30.8	5.8
24 CH	56.9	5.6
26 C	177.5	4.9
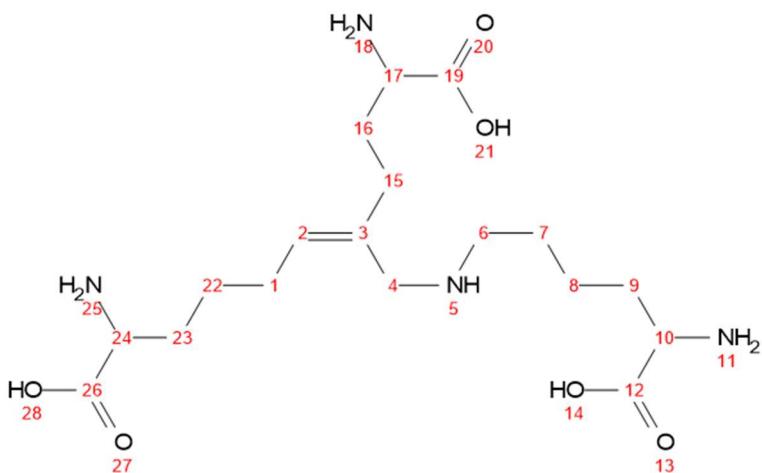
Table AIII.7: Predicted ^{13}C chemical shifts of dehydromerodesmosine.

Atom	Shift (ppm)	Error (ppm)
1 CH2	29.3	5.8
2 CH	147.7	10.2
3 C	136.7	7.5
4 CH	157.1	10.9
6 CH2	58.1	6.2
7 CH2	28.9	5.8
8 CH2	25.5	5.8
9 CH2	31.0	5.8
10 CH	58.0	5.3
12 C	177.7	3.0
15 CH2	29.0	5.5
16 CH2	29.5	5.8
17 CH	54.4	5.3
19 C	176.4	3.0
22 CH2	25.0	5.8
23 CH2	30.3	5.8
24 CH	58.1	5.3
26 C	177.8	3.0
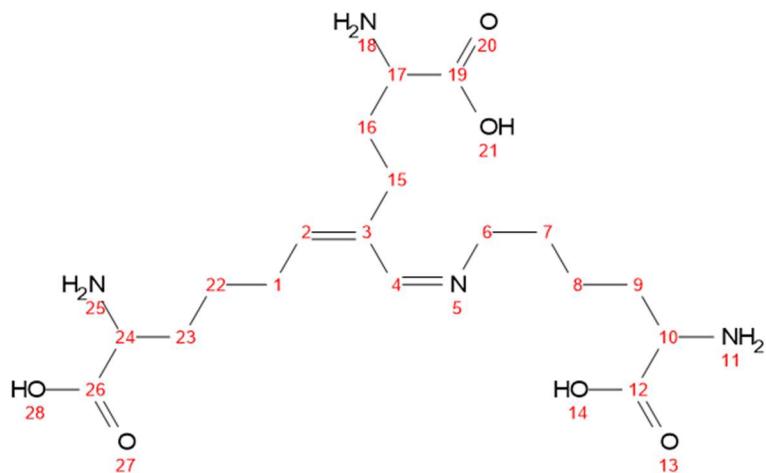
Table AIII.8: Predicted ^{13}C chemical shifts of desmosine.

Atom	Shift (ppm)	Error (ppm)
1 C	140.8	10.9
2 CH	144.8	10.9
4 CH	144.8	10.9
5 C	140.8	10.9
6 C	150.9	10.9
7 CH2	63.7	6.6
8 CH2	27.9	5.8
9 CH2	25.5	5.8
10 CH2	31.0	5.8
11 CH	58.0	5.3
12 C	177.7	3.0
15 CH2	29.1	5.8
16 CH2	29.1	5.8
17 CH2	30.8	5.8
18 CH	54.3	5.3
19 C	176.4	3.0
22 CH	58.9	4.0
24 C	179.4	3.1
26 CH2	30.8	5.8
27 CH	54.3	5.3
28 C	176.4	3.0
34 CH2	28.9	5.8
35 CH2	24.6	5.8
36 CH2	30.1	5.8
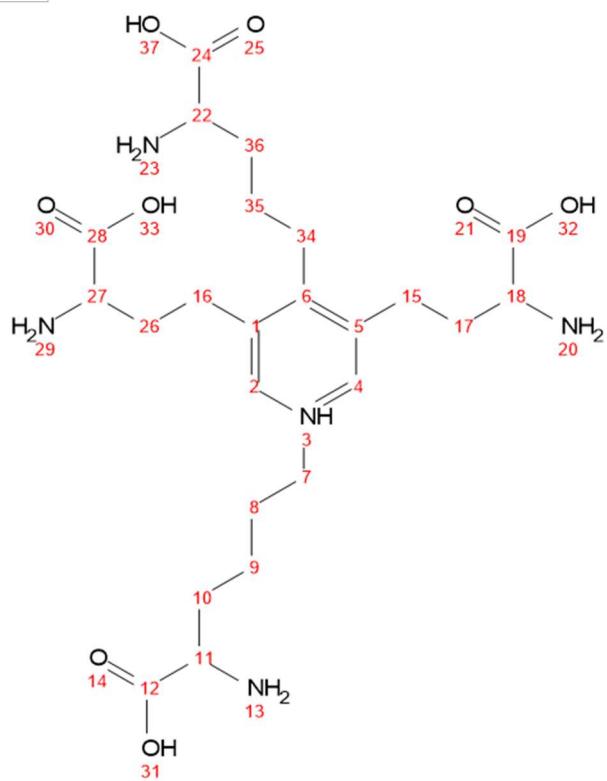
Table AIII.9: Predicted ^{13}C chemical shifts of isodesmosine.

Atom	Shift (ppm)	Error (ppm)
1 C	136.9	6
2 CH	146.1	7.4
4 C	150.7	9.1
5 C	136.0	8.2
6 CH	138.9	7.0
7 CH2	61.5	9.0
8 CH2	29.1	4.6
9 CH2	23.2	4.2
10 CH2	31.3	5.9
11 CH	56.7	5.6
12 C	177.3	4.9
15 CH2	29.8	5.2
16 CH2	31.0	3.8
17 CH2	31.2	5.8
18 CH	54.4	5.6
19 C	176.2	4.9
22 CH2	32.1	5.9
23 CH	54.5	5.6
24 C	176.3	4.9
27 CH2	28.2	5.2
28 CH2	22.2	4.6
29 CH2	30.6	5.9
30 CH	57.8	4.9
31 C	178.8	5.0

116
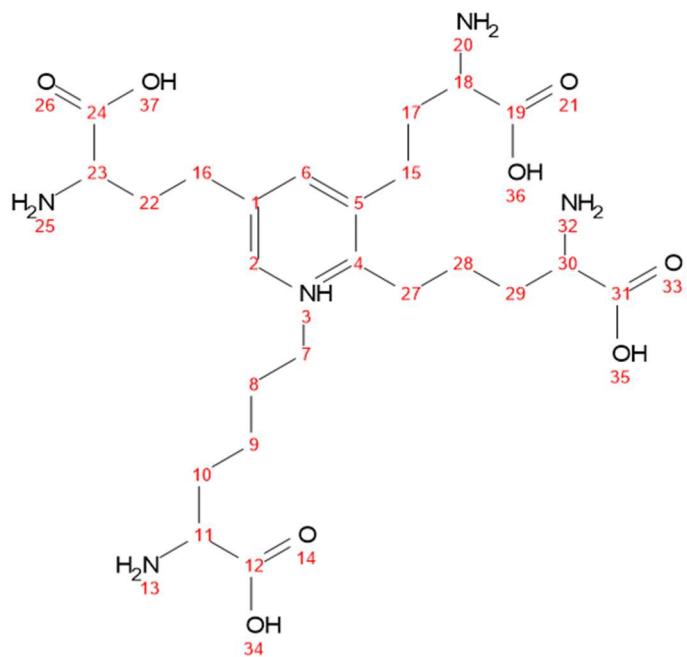
Table AIII.10: Predicted ^{13}C chemical shifts of cyclopentenosine.

Atom	Shift (ppm)	Error (ppm)
1 C	132.8	7.5
2 C	175.4	7.7
3 CH2	36.1	5.5
4 CH	47.3	6.2
5 C	212.0	5.2
7 CH2	24.0	5.8
8 CH2	23.8	5.8
9 CH2	30.3	5.8
10 CH	58.9	4.0
12 C	179.4	3.1
15 CH2	27.5	5.5
16 CH2	28.6	5.5
17 CH2	28.5	5.5
18 CH2	29.9	5.8
19 CH	54.4	5.3
21 C	176.4	3.0
24 CH	56.3	6.1
26 C	175.6	5.8
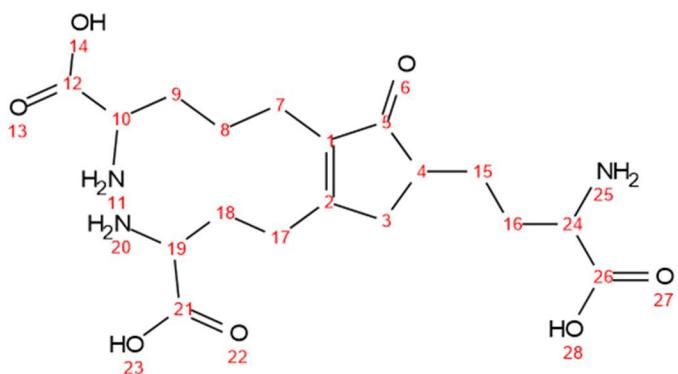
Table AIII.11: Predicted ¹³C chemical shifts of neodesmosine.

Atom	Shift (ppm)	Error (ppm)
1 C	138.8	7.5
2 CH	142.7	10.9
4 CH	142.7	10.9
5 C	138.8	7.5
6 CH	137.0	7.8
7 CH2	63.7	6.6
8 CH2	27.9	5.8
9 CH2	25.5	5.8
10 CH2	31.0	5.8
11 CH	58.0	5.3
13 C	177.7	3.0
16 CH2	32.5	5.8
17 CH2	31.2	5.8
18 CH	54.3	5.3
20 C	176.4	3.0
23 CH2	32.5	5.8
24 CH2	31.2	5.8
25 CH	54.3	5.3
27 C	176.4	3.0
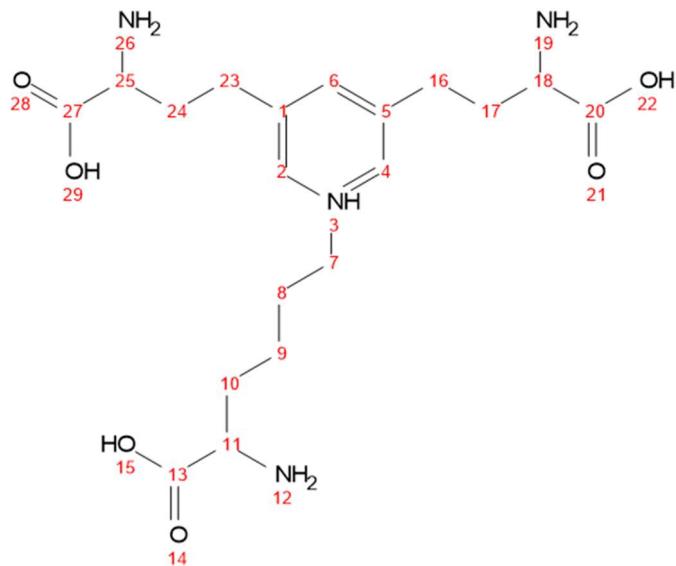
Table AIII.12: Predicted ^{13}C chemical shifts of allodesmosine.

| Atom	Shift (ppm)	Error (ppm)
1 C	139.9	7.5
2 CH	145.0	10.9
4 CH	145.0	10.9
5 C	139.9	7.5
6 C	155.5	10.9
7 CH2	63.7	6.6
8 CH2	27.9	5.8
9 CH2	25.5	5.8
10 CH2	31.0	5.8
11 CH	58.0	5.3
13 C	177.7	3.0
16 CH2	29.3	5.8
17 CH2	30.8	5.8
18 CH	54.3	5.3
20 C	176.4	3.0
23 CH2	29.3	5.8
24 CH2	30.8	5.8
25 CH	54.3	5.3
27 C	176.4	3.0
30 CH	40.3	6.1
31 CH2	34.8	4.8
32 CH2	29.9	4.8
33 CH2	29.4	5.8
34 CH2	28.5	5.5
35 CH	56.3	6.1
37 C	175.6	5.8
40 CH2	26.6	5.8
41 CH2	31.1	5.8
42 CH	58.1	5.3
44 C	177.8	3.0
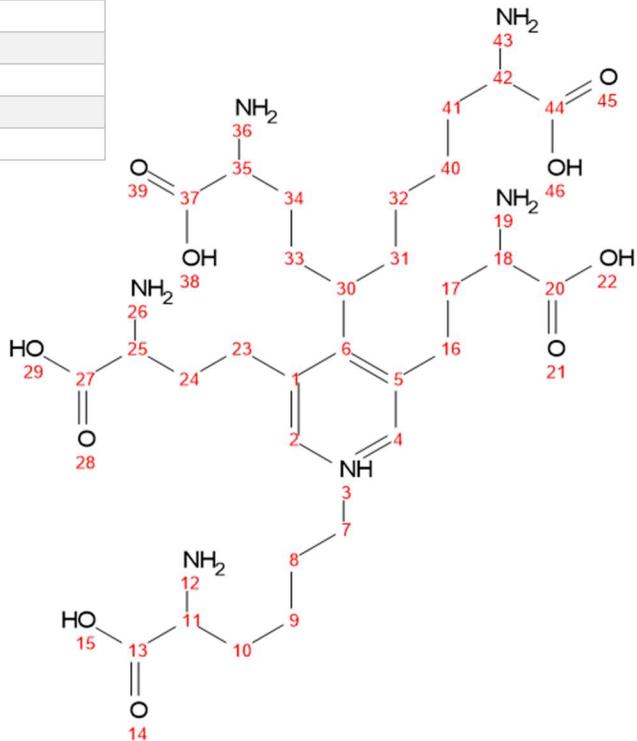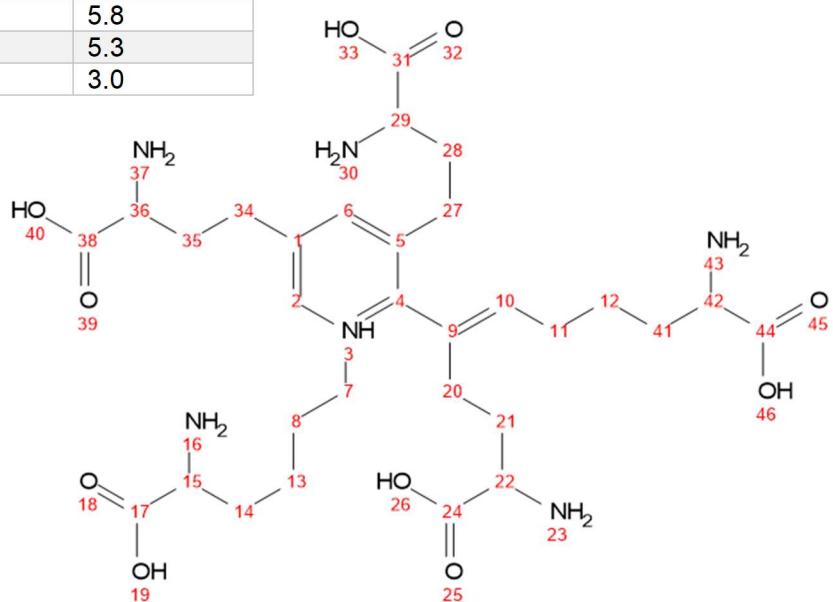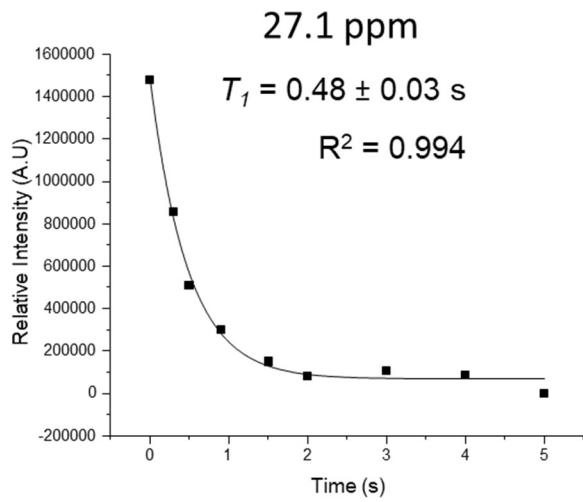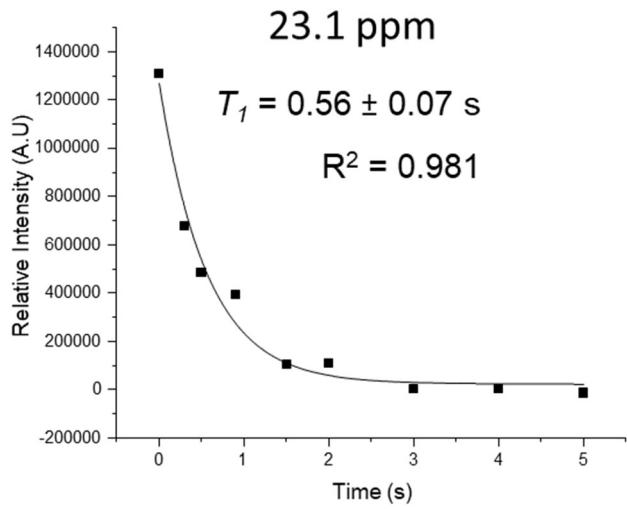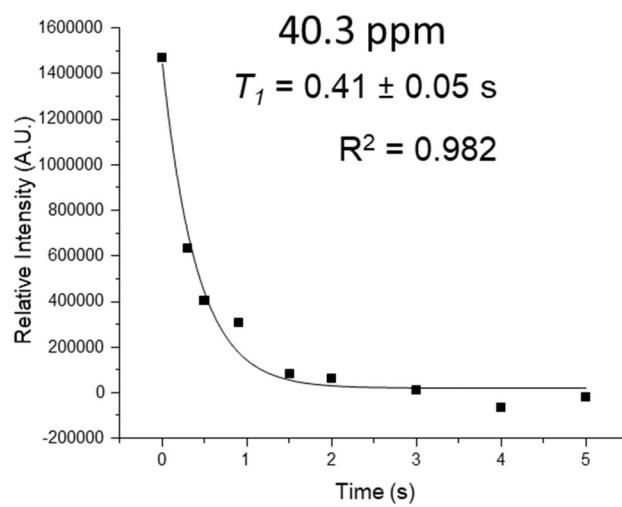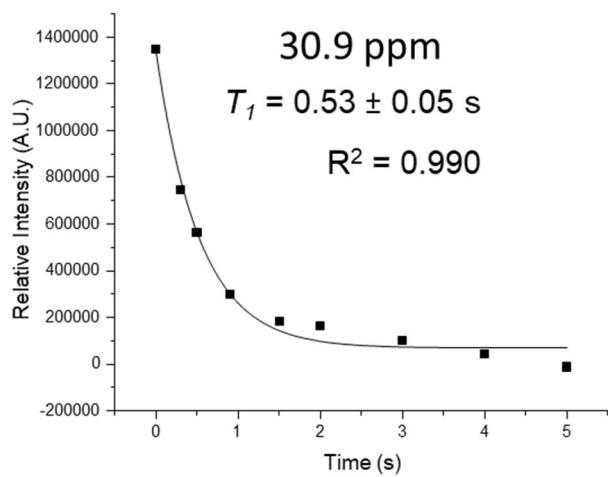
Table AIII.13: Predicted ^{13}C chemical shifts of pentasine.

| Atom	Shift (ppm)	Error (ppm)
1 C	137.4	7.5
2 CH	143.7	10.9
4 C	153.3	10.9
5 C	137.7	7.5
6 CH	136.2	7.8
7 CH2	61.1	7.6
8 CH2	28.0	5.8
9 C	130.3	7.5
10 CH	136.7	7.8
11 CH2	29.2	5.8
12 CH2	25.0	5.8
13 CH2	25.5	5.8
14 CH2	31.0	5.8
15 CH	58.0	5.3
17 C	177.7	3.0
20 CH2	27.6	5.8
21 CH2	28.6	5.8
22 CH	54.4	5.3
24 C	176.4	3.0
27 CH2	29.1	5.8
28 CH2	30.8	5.8
29 CH	54.3	5.3
31 C	176.5	3.0
34 CH2	32.5	5.8
35 CH2	31.2	5.8
36 CH	54.4	5.3
38 C	176.6	3.0
41 CH2	30.3	5.8
42 CH	58.1	5.3
44 C	177.8	3.0

APPENDIX IV

DETERMINATION OF ^{13}C T_1 VALUES FROM MODIFIED INVERSION RECOVERY MEASUREMENTS
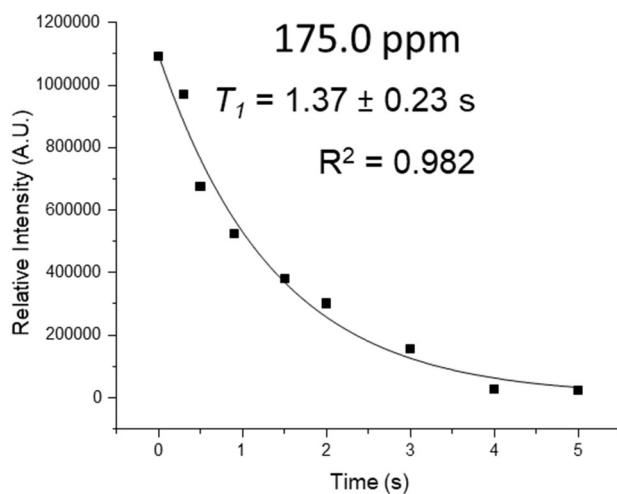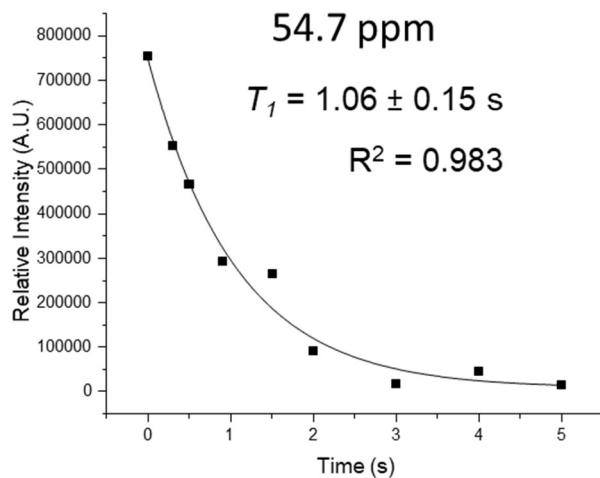
**Figure AIV.1: Exponential decay plots for each ^{13}C-Lys in [U-^{13}C,^{15}N-Lys] elastin.**
OriginPro 2018b software was used for the exponential fittings. Output is shown on the right of each graph.