INVESTIGATIONS OF THE UNIQUE ROLE OF ALANINES IN THE “ELASTIN PUZZLE” BY SOLID-STATE NMR SPECTROSCOPY AND MOLECULAR DYNAMICS SIMULATIONS

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAIʻI AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

CHEMISTRY

MAY 2017

By

Jhonsen Djajamuliadi

Dissertation Committee:

Kristin Kumashiro, Chairperson
John Head
Joseph Jarrett
Thomas Hemscheidt
David Jameson

DEDICATION

I dedicate this dissertation to my family. Special thanks go to my mother who has been a constant support throughout my educational journey. I give my deepest gratitude and warmest affection to my wife, Sachi, my source of encouragement, discernment and courage.

POST TENEBRAS LUX

"After darkness, light"

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to my advisor, Professor Kristin Kumashiro, for her continuous support throughout my graduate studies. Her supervision has helped direct my journey as a scientist, researcher and scholar. I am especially grateful for the numerous opportunities to attend scientific conferences and academic workshops, which contributed to my scholastic career. In addition, her assistance while I was writing this dissertation is much appreciated.

I would like to thank the members of my dissertation committee, Professors Thomas Hemscheidt, John Head, Joseph Jarrett, and David Jameson for their helpful feedback and suggestions that propelled my project forward.

I appreciate my lab colleagues for their support during my graduate career. I am thankful for the countless technical help provided by Dr. Kosuke Ohgo. His attention to details and big-picture outlook have positively influenced my research philosophy. I am appreciative for the proofreading and editing assistance provided by Dr. Walt Niemczura, Dr. Chester Dabalos and, David Shinsato during the writing of this dissertation. Most of all, I am grateful for the friendship and camaraderie with all my colleagues.

I would like to acknowledge the financial support that I received during my graduate program. I am thankful for Professor Kristin Kumashiro's support in the form of research assistantship, which is partially funded by a grant from the National Science Foundation (MCB-1022526). A portion of this research project used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. I appreciate the startup allocations provided through XSEDE, which allowed me to perform most of my MD simulations. Lastly, I am thankful for the teaching assistantship opportunities provided by the Department of Chemistry at the University of Hawai'i at Mānoa.

ABSTRACT

Elastin is the main protein component of the elastic fiber that provides extensibility and resilience to vertebrate tissues. Despite its importance to tissue functionality, elastin's high-resolution molecular structure remains elusive. Structural determination using solution-state nuclear magnetic resonance (NMR) spectroscopy or X-ray diffraction is not applicable to this polymer, due to its insolubility and non-crystallinity. Elastin is the mature crosslinked aggregates of tropoelastin, a ~70 kDa soluble monomeric protein consisting of alternating hydrophobic and crosslinking domains. Tropoelastin's crosslinking domains are characterized by two or three lysines in alanine-rich sequences, such as AAAAKAAKAA and AAAKAAAKAA; the hydrophobic regions are identified by tandem repeats of nonpolar amino acids such as PGVGVA. The abundance of alanines in the two domain types provides an opportunity to probe structural and dynamical microenvironments throughout the polymer. Hence, the characterization of alanines in this protein serves as a significant step to solving the "elastin puzzle".

New insights into the structure and dynamics of elastin are provided by solid-state NMR (ssNMR) spectroscopy and molecular dynamics (MD) simulations. Elastin samples with high levels of isotopic enrichment (~80%) at the alanines were expressed by neonatal rat smooth muscle cells (NRSMC) culture. The results of one- and two-dimensional ssNMR experiments indicate a heterogeneity of conformation and dynamics of alanines in this crosslinked polymer, elastin. Variable-temperature ssNMR and MD studies suggest a conformational equilibrium between α-helix and random coil populations in the crosslinking domains. The application of helix-coil transition theory to the interpretation of ssNMR lineshapes provides new insights to understanding elastin's molecular structure. The Ala-rich crosslinking domains undergo rapid helix-coil interconversions, whereas the hydrophobic domains are characterized by random coil, which exhibits fast large-amplitude motions. Based on these results, an updated model is proposed to describe elastin's molecular organization, reflecting the dynamic characteristics of this protein.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

^{13}C	carbon-13 isotope
^{15}N	nitrogen-15 isotope
$^{15}N \rightarrow {}^{13}CO$	magnetization transfer from nitrogen to carbonyl carbon
1D	one-dimensional
^{1}H	proton
$^{1}H \rightarrow {}^{15}N$	magnetization transfer from proton to nitrogen
$^{1}J_{C\alpha CO}$	one-bond (through-bond) *J*-couplings between carbon-alpha and carbonyl carbon
$^{1}J_{C\alpha C\beta}$	one-bond (through-bond) *J*-couplings between carbon-alpha and carbon-beta
2D	two-dimensional
Å	Angstrom
Ab	ion abundance
a_i	acceleration of particle i
Ala, A	alanine
$(Ala)_n$	polyalanine with n number of residues
ALDL	allysine aldol
ALT	alanine transaminase enzyme
AMBER	assisted model building with energy refinement
AOA	aminooxyacetic acid
Arg, R	arginine
Asn, N	asparagine
Asp, D	aspartic acid
B_0	external magnetic field
B_1	transverse magnetic field
BMRB	biological magnetic resonance data bank

BNL	bovine nuchal ligament
CD	circular dichroism
CHARMM	chemistry at Harvard macromolecular mechanics
cMD	conventional molecular dynamics
CNBr	cyanogen bromide
CO	carbonyl carbon
COSY	correlation spectroscopy
CP	cross-polarization
CSA	chemical shift anisotropy
CTUC-COSY	constant-time uniform-sign cross-peak correlation spectroscopy
CW	continuous wave
Cα	carbon-alpha
Cβ	carbon-beta
Cγ	carbon-gamma
Cδ	carbon-delta
d_1	recycle delay
Da	dalton
DCP	double cross-polarization
DD	dipole-dipole coupling
DLNL	dehydrolysinonorleucine
DMEM	Dulbecco’s modified Eagle’s medium
DP	direct polarization
DQF	double quantum filter
ECM	extracellular matrix
ELP	elastin-like peptide(s)
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
FDLA	1-fluoro-2-4-dinitrophenyl-5-L-leucinamide
FEL	free energy landscape
F_i	force exerted on particle i
FID	free-induction decay
fs	femtosecond
FT	Fourier transformation
FWHM	full-width at half maximum or linewidth
GB	generalized Born
GBSA	generalized Born-surface area
GHz	gigahertz
Gln, Q	glutamine
Glu, E	glutamic acid
Gly, G	glycine
GOR	Garnier-Osguthorpe-Robson
GROMACS	Groningen machine for chemical simulations
GUI	graphical user interface
h	Planck’s constant
HAN	hydroxylamine
HB, H-Bond	hydrogen-bond
HCl	hydrochloric acid

H_{CS}	Hamiltonian for the chemical shift interaction
HFA	hexafluoro-1-propanol
H_{IS}	Hamiltonian for the dipole-dipole coupling between abundant spin I and rare spin S
HPLC	high-performance liquid chromatography
Hz	hertz
IDP	intrinsically disordered protein
IMD	interactive molecular dynamics
IR	infrared
I_z	z-component of the angular spin operator I
k	force constant
K	Kelvin
k_B	Boltzmann's constant
kDa	kilodalton
kHz	kilohertz
L/G	Lorentzian/Gaussian
LC	liquid chromatography
Leu, L	leucine
LOW-BASHD	long-observation-window band-selective homonuclear decoupling
LOX	lysyl oxidase
LOXL	lysyl oxidase-like
Lys, K	lysine
MAS	magic-angle spinning
MC	Monte Carlo
MD	molecular dynamics
mg	milligram
MH	protonated molecular ion

MHz	megahertz
m_i	mass of particle i
mL	milliliter
mm	millimeter
mM	millimolar
MS	mass spectrometry
ms	millisecond
$NaHCO_3$	sodium bicarbonate
NAMD	nanoscale molecular dynamics
NEAA	non-essential amino acids
NMR	nuclear magnetic resonance
nOe	nuclear Overhauser effect
NRSMC	neonatal rat smooth muscle cells
ns	nanosecond
OH	hydroxyl
OPLS	optimized potential for liquid simulation
P(r)	probability distribution of a molecular system along the reaction coordinate, r
PALD	propionaldehyde
PB	Poisson-Boltzmann equation
PDB	protein data bank
Phe, F	phenylalanine
PLP	pyridoxal phosphate
PP-II	polyproline II
ppm	parts per million
Pro, P	proline
ps	picosecond

PSF	protein structure file
q	atomic charge
QQQ	triple-quadrupole
R_c	ratio of isotopic abundance
R_e	ratio of isotopic abundance in enriched analytes
REMD	replica-exchange molecular dynamics
RF	radio frequency
r_i	atomic coordinate of particle i
r_{ij}	interatomic distance between particle i and j
rINEPT	refocused insensitive nuclei enhanced by polarization transfer
r_{IS}	internuclear distance between spin I and spin S
RMSD	root-means-square deviation
RMSF	root-means-square fluctuation
rSNOB	refocused selective excitation for biochemical applications
s	second
S/N	signal-to-noise ratio
SA	solvent accessibility
SASA	solvent-accessible surface area
SCH-1	Schiff base compound-1
Ser, S	serine
SMD	steered molecular dynamics
SPPS	solid-phase peptide synthesis
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-75	tissue culture flask with 75 cm^2 surface area
TACC	Texas advanced computing center
TCA	tricarboxylic acid cycle
TFE	trifluoroethanol
T_g	glass transition temperature
TOBSY	total bond spectroscopy
TOCSY	total correlation spectroscopy
TOF	time-of-flight
TPPM	two-pulse phase modulation
Tyr, Y	tyrosine
U	potential energy
UCDQF-COSY	uniform-sign cross-peak double-quantum-filtered correlation spectroscopy
UV	ultraviolet
Val, V	valine
vdW	van der Waals interactions
VMD	visual molecular dynamics
VT	variable-temperature
wSEDUCE-1	windowed detection with selective decoupling using crafted excitation
°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

$\Delta\delta(R_{+1})$	correction factor for the residue following alanine in R_{-1}-Ala-R_{+1} motif
$\Delta\delta(R_{-1})$	correction factor for the residue preceding alanine in R_{-1}-Ala-R_{+1} motif
δ_{iso}	isotropic chemical shift
δ_{CSA}	magnitude of chemical shift anisotropy
δ_{11}, δ_{22}, δ_{33}	principal components of the chemical shift tensor
δ_{rc}(Ala)	random coil chemical shift for alanine
Δt	molecular dynamics timestep
γ	gyromagnetic ratio
γ^*	phase shift in dihedral potential function
ε	dielectric constant
ε_0	vacuum permittivity
ε^{LJ}	depth of well in Lennard-Jones potential
η	nOe enhancement factor
θ	angle with respect to the external magnetic field B_0
θ^*	angle formed by three bonded atoms
μ_0	permeability of free space
μL	microliter
μm	micrometer
μs	microsecond
σ	zeroth distance in Lennard-Jones potential
τ, τ', τ_1'	time delay
τ_C	correlation time
Φ	phi torsion angle
Ψ	psi torsion angle
ω	omega torsion angle
ω_r	rotor spinning speed

CHAPTER 1. INTRODUCTION

A. ELASTIC FIBER, ELASTIN, AND TROPOELASTIN

Elastic fibers provide flexibility and resilience to vertebrate tissues, like arteries, lungs, and skin, that are subject to stretch-and-recoil cycles. Although these fibers are assembled during early development, their durability extends through the entire lifetime of the animal [1-3]. Damage to adult tissues typically result in improperly repaired fibers, which often lead to pulmonary or cardiovascular disease.

Elastin is the protein component that confers elasticity to the elastic fiber. This protein is highly crosslinked, and it is typically organized into sheets or lamellae in the extracellular matrix (ECM). Elastin provides the flexibility and resilience to vertebrate tissues and organs to undergo billions of stretch-recoil cycles throughout the animal's lifespan without significant turnover [1-3]. Apart from its biomechanical property to withstand repetitive forces, elastin also plays a role in various physiological processes [4, 5]. Despite extensive studies on this polymer, the three-dimensional structure of elastin remains elusive, and the mechanism for elasticity is still uncertain.

A.1 Elastic fiber assembly

The assembly of elastic fibers or elastogenesis is a multistep process that involves complex interactions among many ECM proteins and enzymes. Although detailed interactions and assembly steps are not fully understood, the overall process generally occurs in two stages, microfibril formation and elastin deposition. Microfibrils are strand-like fibers that are composed of several different glycoproteins, and they serve as the scaffold for the accumulation of elastin. Elastin assembly begins with the secretion of its soluble monomer, tropoelastin (**Figure 1.1a**) [6]. On the cell surface, these monomers undergo a self-association process called coacervation and form globular aggregates [7]. The size of these aggregates is carefully controlled by a family of glycoproteins that include the fibulins and fibrillins [8]. Lysyl oxidase (LOX) and LOX-like (LOXL) enzymes perform an oxidative deamination of the lysine sidechains of tropoelastin [8, 9], followed by condensation reactions that form intra- and intermolecular crosslinks. These crosslinked aggregates are subsequently deposited onto the microfibrils and added to the developing elastic fiber.

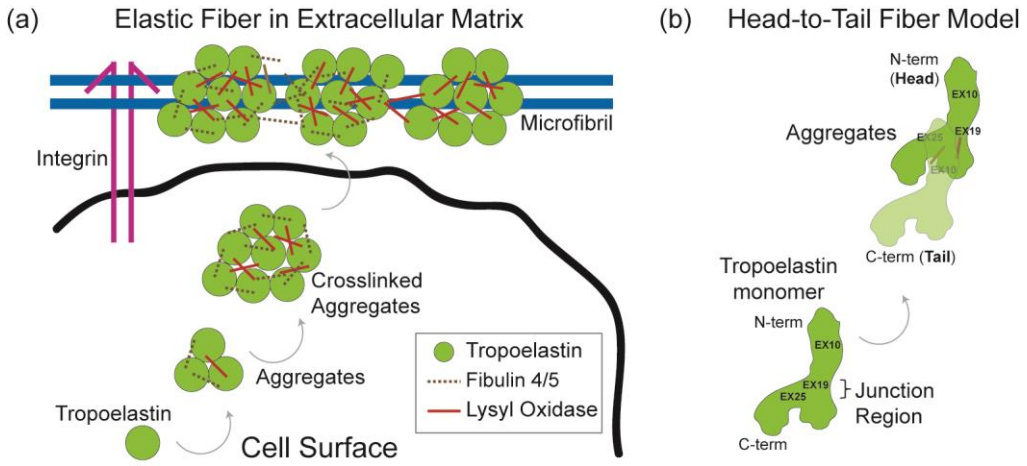


Figure 1.1 Models for (a) elastic fiber assembly and (b) head-to-tail crosslinking of tropoelastin monomers. Figures are adapted from [6] and [10]. **Green** represents tropoelastin monomers. Curved arrows illustrate the direction of fiber assembly.

Recent studies shed some light to the orientation of tropoelastin aggregation during assembly. Tropoelastin monomers are proposed to align in a 'head-to-tail' fashion (**Figure 1.1b**) [10], which designates the domains encoded by exons 19 and 25 (EX19 and EX25) as the intermonomer junction. In mature elastin, lysines in this region are crosslinked with another pair of Lys in the Pro-rich domain encoded by exon 10, located in another monomer [10, 11]. These studies imply a specificity and order in the macromolecular assembly of elastin.

A.2 Models of elastin and tropoelastin

Several models have been proposed to account for elastin's flexibility, including the random chain [12], the liquid drop [13], the oiled coil [14], and the fibrillar network [15]. These representations differ in their descriptions of order within the molecule. The oldest model, random chain [12], describes elastin as a network of long unstructured chains that are interconnected by elastin's crosslinks (**Figure 1.2a**). The protein is considered to be a rubber-like elastomer that exhibits free kinetic motions. Elasticity is attributed to the decrease of entropy upon stretching, and the tendency to restore randomness provides the recoiling force. However, this model fails to take account of the necessity of water as a plasticizer in elastin [16]. Additionally, the absence of local order cannot explain the self-aggregation phenomenon or inverse-temperature phase transitions in elastin and elastin peptides [17].

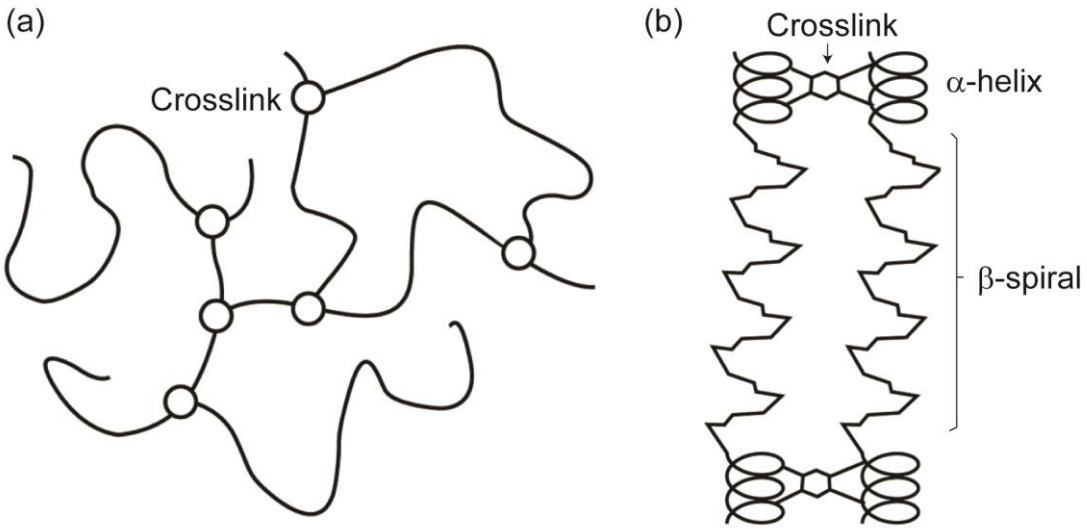


Figure 1.2 Classical models of elastin. (a) Random chain network [12] and (b) fibrillar model [15]. In (a), curly lines and circles represent random unstructured chains and elastin's crosslinks, respectively. In (b), hexagons and springs represent crosslinks and α-helices; jagged lines illustrate the β-spiral motif.

In contrast, other models like the oiled coil [14] and the fibrillar network [15] suggest the existence of local order within the polymer. In particular, the fibrillar model proposes the prominence of ⟨-helices and β-spirals in elastin's crosslinking and hydrophobic domains, respectively. The β-spiral consists of repeats of VPGVG in consecutive β-turns forming a loosely bound helix (**Figure 1.2b**). Elasticity is explained by the librational entropy mechanism, by which distortion of β-spirals upon stretching decreases entropy and provides the restoring force for recoil [18]. However, no experimental evidence supports this mechanism, and recent molecular dynamics (MD) simulations refute the proposed elastomeric properties of the β-spirals [19, 20]. Overall, questions regarding the structural character in elastin and the basis for elasticity remain unanswered by these models. Hence, the search for a reasonable representation of elastin structure continues.

Elastin studies have concomitantly been focused on the characterization of its precursor, tropoelastin. This protein is a ~70 kDa soluble monomer that is rich in Gly, Ala, Val, and Pro. The gene structure of tropoelastin is generally composed of two alternating domain types, hydrophobic and crosslinking (**Figure 1.3**). The hydrophobic domains are characterized by repeats of nonpolar residues in common motifs such as VPGVG, and VGGVG. In contrast, the crosslinking domains feature pairs of Lys residues that are interspersed by two or three amino acids such as KAAK or KAAAK.

Figure 1.3 Representation of tropoelastin's alternating domains based on the cDNA structure [21]. Gray (⬡) and white hexagons (⬡) represent the KP- and KA-type crosslinking domains, respectively. Black rectangles (▮) represent hydrophobic domains. Gray rectangle (▮) and white diamond (◇) represent the N- and C-term domains, respectively. Figure is adapted from [22].

Secondary structures in each of tropoelastin's domains have been extensively investigated by Tamburro and co-workers [23, 24]. An ensemble of conformations is suggested from the results of exon-by-exon studies of elastin-like peptides in a mixture of water and trifluoroethanol, a solvent that induces the formation of secondary structures. The crosslinking domains are characterized by α-helices [23], whereas the hydrophobic regions are described by labile structures that interconvert among PP-II, β-turns, and an unordered state [25] (**Figure 1.4**).

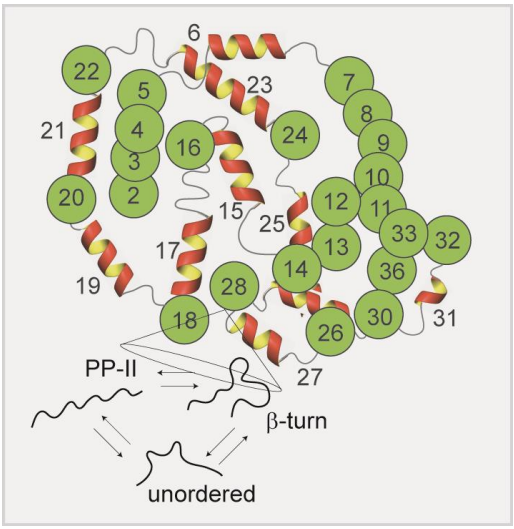


Figure 1. 4 Tamburro's idealized tropoelastin model [24]. **Green** represents hydrophobic domains; structural interconversions among PP-II, β-turn and unordered state are indicated by double arrows. **Red**/**yellow** represents α-helices in crosslinking domains. Figure adapted from [24].

Unlike tropoelastin, elastin is abundant in crosslink moieties [26]. *In vivo*, the crosslinking reaction is catalyzed by LOX and LOXL enzymes, which perform an oxidative deamination of a lysine sidechain to form an aldehyde called allysine (*panel 1*, **Figure 1.5a**). The spontaneous condensation of two intrachain allysines yields an allysine aldol (ALDL), and a similar nonenzymatic reaction between an allysine and an unmodified Lys forms a dehydrolysinonorleucine (DLNL) (*panel 2*) [27, 28]. These two moieties can undergo further interchain crosslinking reaction that yields a desmosine (*panel 3*), a tetrafunctional crosslink that is unique to elastin. Other biosynthetic pathways may also form an isomer of the

tetrafunctional moiety called isodesmosine [13, 29], as well as tri- and pentafunctional crosslinks called dehydromerodesmosine and cyclopentenosine, respectively [30-32].

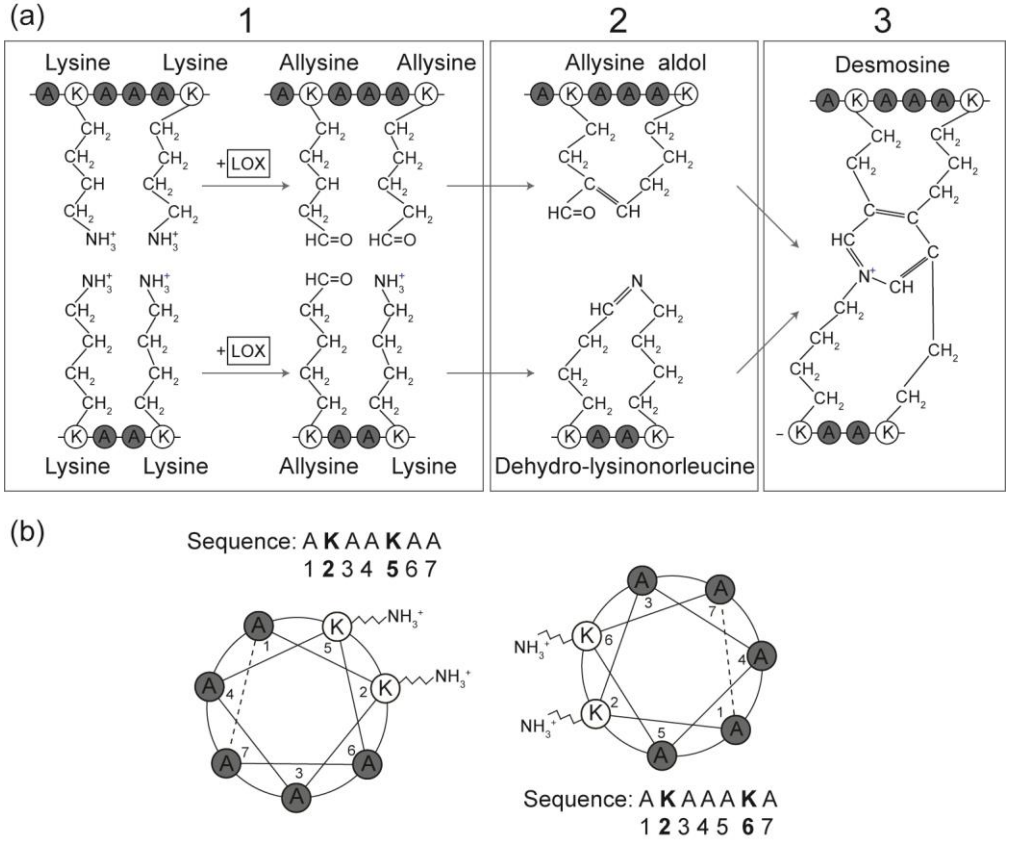


Figure 1.5 Lys-derived crosslinks in elastin. (a) Structures and synthetic routes of bi- and tetrafunctional crosslinks; (b) spatial juxtaposition of four lysines that leads to desmosine formation. In (a), *panel 1*, lysyl oxidase (LOX) performs an oxidative deamination on one or two lysine(s) to form allysine(s). *Panels 2 and 3*, spontaneous condensation reactions of nearby moieties form bi- and tetrafunctional crosslinks, respectively. *Top*, intrachain allysine aldol (ALDL) formation takes place in lysine pairs that are interspersed by 3 alanines (**A**). *Bottom*, intrachain allysine aldol (ALDL) formation takes place in lysine pairs that are interspersed by only 2 alanines (**A**). Arrows indicate the direction of biosynthetic routes. In (b), large circles represent Edmundson's helical wheels [33], which help visualize the cross-sections of α-helices and the placements of each residue relative to each other in the helix. *Left*, Lys (K) residues at positions **2** and **5** in, e.g., A**K**AA**K**AA sequence are located on the same side of the helix. *Right*, Lys sidechains at positions **2** and **6** in, e.g., A**K**AAA**K**A sequence are also proximal to each other.

Crosslink formations in elastin's Ala-rich domains are facilitated by α-helices. A pair of Lys (**K**) that are interspersed by two or three amino acids, e.g., **K**AA**K** or **K**AAA**K**, have their sidechains oriented on the same side of the α-helix (**Figure 1.5b**) [11, 34]. This spatial arrangement allows for the nonenzymatic condensation of allysines forming an intrachain crosslink such as ALDL. This crosslink and a nearby DLNL may subsequently form an interchain crosslink such as desmosine [35, 36].

Although the biosynthetic pathways of crosslinks are well understood, the function of Lys-derived moieties in the α-helical domains is still uncertain. Additionally, the roles of ALDL, DLNL or desmosine in the elastomeric properties of mature elastin have not been investigated. Therefore, extensive studies are necessary to identify the significance of these crosslinks in the structural integrity of elastin.

A.3 The unique role of alanines in elastin

Each of the four major amino acids in elastin (i.e., Gly, Pro, Val, and Ala) has a different biochemical property that influences the structure and function of elastin. Glycines have no sidechain groups and are known to be helix-breakers [37, 38]. Hence, they accommodate higher degrees of motional freedom than other residues. Valines have branched chain residues that are known to participate in hydrophobic interactions during elastogenesis [39]. Prolines in elastin have liquid-like mobility that enhances segmental motions in the backbone [40]. These amino acids also occur in tandem repeats forming common motifs like VPGVG, which was shown to form consecutive β-turns and proposed to provide the recoiling force [18].

In contrast to Gly, Val, and Pro, the alanines are prominent in both crosslinking and hydrophobic domains of tropoelastin (**Table 1.1.**). In the crosslinking domains, they are generally found in tandem with two or more repeating residues. However, hydrophobic domains are not found to contain neighboring alanines. Rather, each Ala in the amino acid sequence is often adjacent to or flanked by a pair of glycines. A slightly larger distribution of alanines is observed in the crosslinking domains (57%), relative to those in the hydrophobic regions (43%).

The abundance of alanines in the domains of elastin provides a promising opportunity to study the microenvironments throughout the protein. Presumably, the structural characters of Ala in the hydrophobic domains are significantly different from those in the crosslinking domains. Therefore, the characterization of alanines in elastin will provide insights into the structure and dynamics of the different domains in the protein.

Table 1.1. Amino acid sequences of rat tropoelastin's hydrophobic and crosslinking domains. The primary structure is obtained from ElastoDB [41]. Alanines are highlighted in red. White and gray shading represent hydrophobic and crosslinking domains, respectively. The two respective regions have 73 and 95 alanines, which account for 43% and 57% of the total number of Ala in the entire protein. (*Domains encoded by exon 12, 13, 14, 15 and 33 are subject to alternative splicing [42].)

Exon	Amino Acid Sequence
1	MAGLTAAVPQPGVLLILLLNLLHPAQPG
2	GVPGAVPGGVPGGLPGGVPGGVYYP
3	GAGIGGGLGGG
4	ALGPGGKPPKP
4A	GAGLLGAFGA
5	GPGGLGGAGPGA
6	GGVLVPGGGAGAAAAYKAAAKA
7	GAGLGGIGGVPGGVGVGGVPGAVGVGGVPGAVGGIGGIGGLGVST
8	GAVVPQLGAGVGAGGKPGKVP
9	GVGLPGVYPGGVLPGT
10	GARFPGVGVLPGVPTGTGVKAKVP
11	GGGGGAFSGIP
12*	GVGPFGGQQPGVPLGYPIKAPKLP
13*	GGYGLPYTNGKLPY
14*	GVAGAGGKAGYPTGT
15*	GVGSQAAVAAAKAAKY
16	GAGGGGVLPGVGGGGIPGGAGAIPGIGGIT
17	GAGTPAAAAAAKAAAKAAKY
18	GAAGGLVPGGPGVRVPGAGIPGVGIPGVGGIPGVGGIPGVGGIPGVGGPGIGGPGIVGGP
19	GAVSPAAAAKAAAKAAKY
20	GARGGVGIPTYGVGAGGFPGYGVGAGAGLG
21	GASQAAAAAAAAKAAKY
22	GAGGAGTLGGLVPGAVPGALPGAVPGALPGAVPGALPGAVPGVPGTGGVP
23	GTPAAAAAAAAAKAAAKAGQY
24	GLGPGVGGVPGGVGVGGLPGGVGPGGVTGIGTGPGTGLVPGDLG
25	GAGTPAAAKSAAKAAAKAQY
26	RAAAGLGAGVPGLGVGAGVPGFGAGAGGFGAGAGVPGFGAGA
27	VPGSLAASKAAKY
28	GAAGGLGGPGGLGGPGGLGGPGGLGGPGGLGGPGGLGGVPGGVA
29	GAPAAAAAAKAAAKAAQY
30	GLGGAGGLGAGGLGAGGLGAGGLGAGGLGAGGLGAGGVIPGAVGLG
31	GVSPAAAAKAAKY
32	GAAGLGGVLGARPFPGG
33*	GVAARPGFGLSPIYP
34	GGGAGGLGVG
35	GKPPKPYGGALGALGYQ
36	GGGCFGKSCGRKRK

A.4 Common secondary structures in proteins and random coil in unfolded peptides

Each protein secondary structure is typically associated with a particular set of torsion angles, i.e., phi (ϕ), psi (ψ), and omega (ω). By convention, the angles of rotation about the covalent bonds, N–Cα (ϕ) and Cα–CO (ψ), range from -180° to 0° and from 0° to 180° (**Figure 1.6a**). These combination of torsion angles are typically illustrated on a (ϕ vs. ψ) map, called the Ramachandran plot [43]. Certain combinations are considered disallowed (white, **Figure 1.6b**), because they result in steric hindrances. The "fully allowed" regions (green) correspond to configurations that have no atomic clashes, and the "partially allowed" regions (gray) reflect conformations with modest contacts, but frequently observed in many proteins. The amide partial double-bond character limits the ω angle to 180° in all amino acids except for prolines.

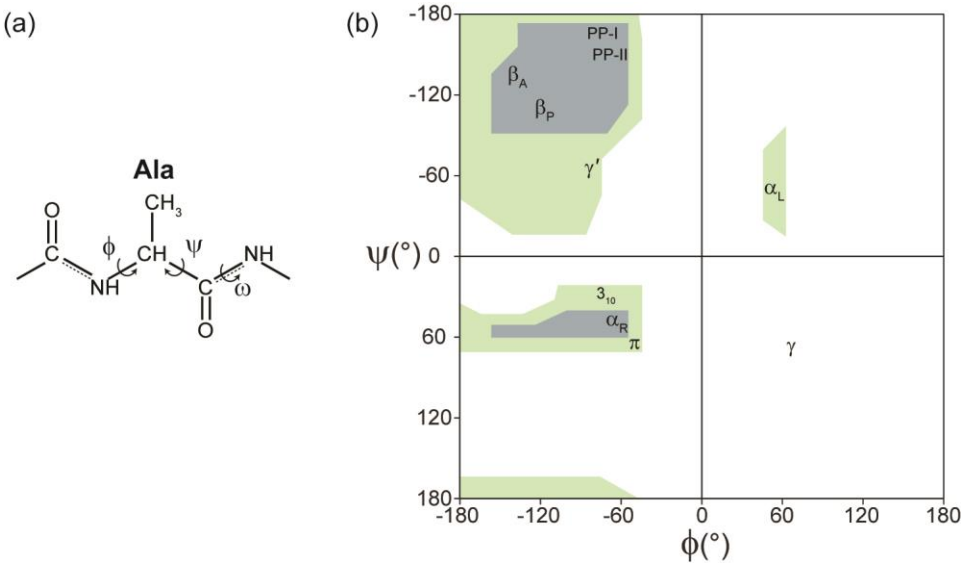


Figure 1.6 Torsion angles of common secondary structures in peptides and proteins. (a) The (ϕ, ψ) angles of an alanine; (b) Ramachandran plot showing regions of common structural motifs. In (a), dotted lines indicate the partial double-bond character in the C-N. In (b), **green** and **gray** shading correspond to the partially and fully allowed regions in the Ramachandran space [43]. Torsion angles of common structural motifs are indicated on the map. α_R, α_L, 3_{10} and π represent right-handed α-, left-handed α-, 3_{10}- and π-helices, respectively. PP-I and PP-II correspond to polyproline type I and II, respectively. β_A and β_P are anti-parallel and parallel β-sheets, respectively. γ and γ' describe gamma- and inverse gamma-turns, respectively.

Common secondary structures in peptides and proteins occupy distinct regions of the Ramachandran plot. For example, the right-handed α-helix and polyproline-II are identified at (ϕ, ψ) = (-60°, -45°) and (-78°, 149°), respectively. **Table 1.2.** lists the torsion angles of common structural motifs [44].

Table 1.2. Torsion angles of common secondary structures. (ϕ, ψ) angles obtained from [44] and [45].

Secondary Structure	Abbreviation	(ϕ, ψ)
Right-handed α_R-helix	α_R	(-57°, -47°)
Right-handed α_L-helix	α_L	(57°, 47°)
Parallel β-sheet	β_P	(-119°, 113°)
Anti-parallel β-sheet	β_A	(-139°, 135°)
Polyproline I	PP-I	(-83°, 158°)
Polyproline II	PP-II	(-78°, 149°)
γ-turn	γ	(75°, -65°)
Inverse γ-turn	γ'	(-79°, 69°)

Peptide segments lacking defined secondary structures have been traditionally described as random coil. However, this term has created much confusion about the peptide's actual conformation, or lack thereof. In the context of protein-based polymers, random coil is defined as a state that has no preferences for a particular secondary structure [12, 46], and its corresponding torsion angles occupy all regions of the Ramachandran plot with similar probabilities. Although such a fully unstructured state may be present at extremely high temperatures [47], recent studies of unfolded peptides and intrinsically disordered proteins (IDP) indicate that swift interconversions of secondary structures are more likely found in the so-called random coil. Moreover, common motifs corresponding to the α- and/or β-regions of the Ramachandran plot are identified [48-50], which makes the term *statistical coil* [49-52] a better description for the structural character of unfolded peptides and IDP like elastin. Nevertheless, the term random coil is used throughout this dissertation to uphold the consistency with previous literature [53].

B. CHARACTERIZATION METHODS

A high-resolution three-dimensional structure of elastin remains elusive to this day. Conventional characterization techniques such as solution-state Nuclear Magnetic Resonance (NMR) spectroscopy and X-ray crystallography are not applicable, because elastin is an insoluble, amorphous solid, and it cannot be crystallized. Although early studies by infrared (IR) and circular dichroism (CD) identify common secondary structures in this protein [54], the exact regions that correspond to these motifs are unknown. Hence, other characterization techniques are necessary to identify the location of these secondary structures in elastin's domains.

Solid-state NMR (SSNMR) spectroscopy and molecular dynamics (MD) simulations are applicable to the characterization of native elastin. ssNMR spectroscopy has been successfully implemented in the study of fibrillar proteins such as amyloids [55, 56], collagen [57, 58], and silk [59, 60]. NMR-based approaches are effective for obtaining detailed molecular information such as torsion angles, interatomic-distances and dynamics in large, biological molecules [61, 62]. Similarly, MD simulations offer atomic-level insights into the underlying mechanisms occurring within a particular protein domain, complementary to experiments. Thus, solid-state NMR spectroscopy and MD simulations are expected to resolve some uncertainties regarding the molecular structure of elastin.

B.1 Brief introduction of solid-state NMR spectroscopy

NMR spectroscopy of spin-1/2 nuclei (e.g., ^{1}H and ^{13}C) is widely used to characterize biological molecules. Although ^{1}H spectra are used ubiquitously for soluble proteins, its application to solids is limited. Strong dipolar couplings among protons in the solid lattice cause severe line broadening, rendering the assignment of ^{1}H chemical shifts unfeasible. The narrow range of ^{1}H chemical shifts (0-12 ppm) also makes peak assignments of broad signals difficult. However, ^{13}C nuclei have a significantly broader spectral range (0-200 ppm), and the isotropic chemical shifts are known to be conformation-dependent [63, 64], which makes assignments of secondary structures for proteins and polypeptides achievable.

NMR lineshapes of solid samples are significantly broader than those of liquids. The broad solid-state signals primarily arise from chemical shift anisotropy (CSA) and heteronuclear dipole-dipole coupling (DD) interactions. In the absence of isotropic liquid-like molecular tumbling, the spectrum displays a broad lineshape called the powder pattern (**Figure 1.7**). This pattern corresponds to the sum of all possible molecular orientations in the sample.

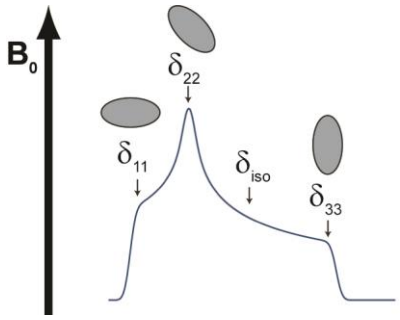


Figure 1.7 Powder pattern in solid-state NMR spectrum and schematic representation of the principal values of the CSA. Line represents a static powder pattern. Ellipsoids represent electron densities around nuclei; their orientations with respect to external magnetic field (**B_0**) give rise to different resonance frequencies, which are represented by principal values δ_{11}, δ_{22}, and δ_{33}.

The anisotropy, or orientation dependence, of chemical shifts is traced to the effect of **B_0** on the electrons surrounding a nucleus. Electron densities are generally thought to be ellipsoid in shape, and the extent of which these electron clouds affect the resonance frequency of a nucleus depends on their orientations with respect to **B_0**. The largest effect (most deshielded) corresponds to the principal value δ_{11}, which occurs in an ellipsoid that has its narrowest diameter oriented along **B_0**. The smallest effect (least deshielded) corresponds to the principal value δ_{33}, which occurs in an electron cloud that has its widest diameter oriented along **B_0**. Furthermore, δ_{11} and δ_{33} correspond to left and right edges on the powder pattern, respectively. The principal value δ_{22} represents an ellipsoid orientation that is perpendicular to δ_{11} and δ_{33}, and its position corresponds to the highest intensity in the powder pattern. In the special case δ_{11}=δ_{22}, the chemical shift Hamiltonian is described by,

$$H_{cs} = \gamma B_0 I_z[\delta_{iso} + \frac{1}{2}\delta_{CSA}(3cos^2\theta - 1)] \qquad \text{(eqn. 1.1)}$$

where the average of the three principal values, also known as the isotropic chemical shift, δ_{iso}, is described by

$$\delta_{iso} = 1/3(\delta_{11} + \delta_{22} + \delta_{33}) \qquad \text{(eqn. 1.2)}$$

and the magnitude of the CSA, δ_{CSA}, is

$$\delta_{CSA} = \delta_{33} - \delta_{iso} \qquad \text{(eqn. 1.3)}$$

In proteins and polypeptides, the extent of chemical shielding effect is related to torsion angles and secondary structures [65, 66].

Interactions of nuclear magnetic moments in two different spins such as ^{1}H and ^{13}C gives rise to heteronuclear DD (**Figure 1.8**). The Hamiltonian for this coupling is given by

$$H_{IS} = -\left(\frac{\mu_0}{4\pi}\right)\frac{\hbar\gamma_I\gamma_S}{r_{IS}^2}(3cos^2\theta - 1)I_zS_z \qquad \text{(eqn. 1.4)}$$

where r_{IS} is the internuclear distance, μ_0 is the permeability of free space, the angle θ describes the orientation of the internuclear vector IS with respect to $\mathbf{B_0}$, γ_I and γ_S are gyromagnetic ratios of the respective abundant spin I (such as ^{1}H) and rare spin S (such as ^{13}C), and I_z and S_z are the z-component of the angular spin momentum operators I and S, respectively. The magnitude of heteronuclear DD is, as indicated by eqn. 1.4, influenced by the internuclear distance (r_{IS}), gyromagnetic ratios (γ_I and γ_S), and the orientation of the internuclear vector IS with respect to $\mathbf{B_0}$.

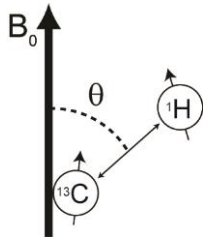


Figure 1.8 Dipolar coupling between ^{13}C and ^{1}H nuclei. Angle θ describes the orientation of the (^{13}C-^{1}H) vector with respect to the external magnetic field, $\mathbf{B_0}$.

To obtain high-resolution spectra that are comparable to the liquid-state spectra, a few techniques have been developed for solids. The standard method today involves a mechanical spinning of the sample at an angle, θ, of 54.74°, relative to the magnetic field, $\mathbf{B_0}$ (**Figure 1.9**). This method, called magic-angle spinning (MAS), significantly reduces the effect of spectral broadening, as the spatial terms ($3\cos^2\theta - 1$) in the CSA and DD interactions are equal to zero with the magic-angle (θ=54.74°).

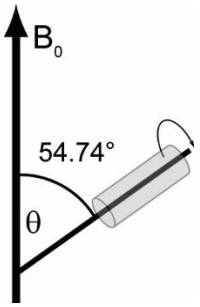


Figure 1.9 Magic angle spinning (MAS) technique in solid-state NMR spectroscopy. The rotor is spun at the angle (θ) = 54.74° relative to external magnetic field ($\mathbf{B_0}$).

The applied MAS rate for typical solids must be equal to or larger than the magnitude of the anisotropy. Otherwise, the CSA is not completely averaged, and spinning sidebands are observed about the isotropic chemical shifts. **Figure 1.10** shows the effect of increasing MAS rate on the ^{13}C signals of

glycine powder. A higher spinning speed is required for the ^{13}CO region than the ^{13}Cα, because the anisotropy of the former is larger (~140 kHz) than the latter (~40 kHz) [67].

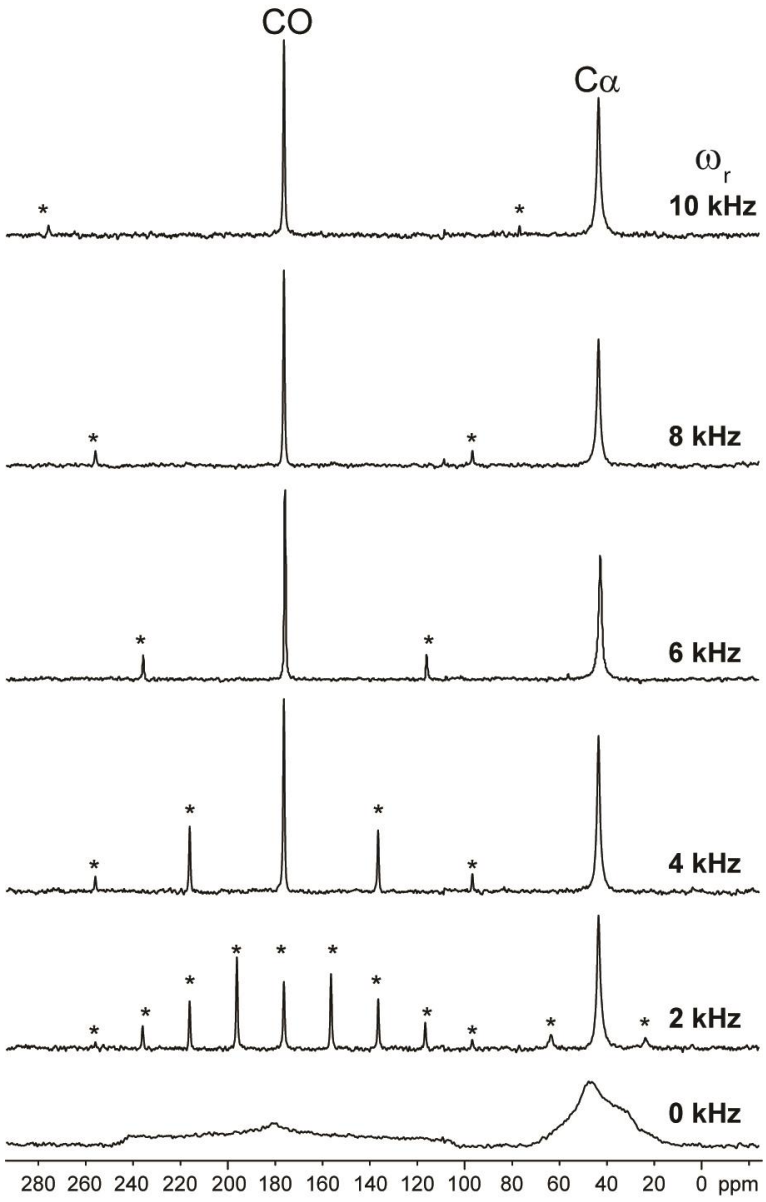


Figure 1.10 Effect of variable MAS frequencies on the ^{13}C spectra of glycine powder. Asterisks (*) indicate spinning sidebands; ω_r represents rotor spinning speed. The powder pattern is observed in the absence of MAS. Spectra were acquired with 400 MHz spectrometer and processed with 40 Hz of exponential broadening.

In addition to MAS, the decoupling of heteronuclear DD interactions, e.g., between ^{13}C and ^{1}H narrows the lineshapes for typical solids. This decoupling technique is conducted by applying a high-power ^{1}H irradiation during ^{13}C acquisition, with **C**ontinuous-**W**ave (CW) or **T**wo-**P**ulse **P**hase **M**odulation

(TPPM) [68]. For an effective removal of a directly attached ^{13}C-^{1}H DD interactions, the applied decoupling power should be > 66 kHz [67].

In summary, ssNMR is the suitable tool to characterize the alanines in insoluble elastin. Secondary structures of the alanines in the protein and the motional regimes in the Ala-rich domains can be probed by this method. The ssNMR results may be useful for refining the molecular models of elastin and providing explanations to the polymer's elastomeric properties.

B.2 Brief introduction of molecular dynamics simulations

Molecular dynamics simulations have emerged as a valuable computational tool to study the structure and dynamics of macromolecules, such as proteins and nucleic acids. All-atom simulations provide insights into the molecular structure of a protein or peptide, which can complement experimental approaches. Recent advancement in computing power allows the calculation of large and complex macromolecules to be achieved in days or weeks.

The typical algorithm utilizes Newton's equation of motion to solve for the acceleration of a set of particles, with respect to the applied forces, as described by

$$\boldsymbol{F_i = m_i a_i} \qquad \text{(eqn. 1.1)}$$

where F_i is the force exerted on particle i, with masses m and acceleration a_i. The forces acted upon the particle are derived from the potential energy, U, of the system,

$$-\frac{\boldsymbol{dU}}{\boldsymbol{dr_i}} = \boldsymbol{m_i}\frac{\boldsymbol{d^2 r_i}}{\boldsymbol{dt^2}} \qquad \text{(eqn. 1.2)}$$

and the changes of particle coordinates as a function of time, t.

The potential energies are generally represented by a set of bonded (U_{bonded}) and non-bonded ($U_{non\text{-}bonded}$) interactions, called force fields. Common force fields such as CHARMM [69] and AMBER [70] contain energy functions that account for these two types of atomic interactions. Bonded interactions generally describe stretching, bending and rotating motions of dihedral angles, as described by

$$U_{bond} = \sum_{bond\ i} k_i^{\ bond}(r_i - r_{0i})^2 \qquad \text{(eqn. 1.3)}$$

$$U_{angle} = \sum_{angle\ i} k_i^{\ angle}(\theta_i^* - \theta_{0i}^*)^2 \qquad \text{(eqn. 1.4)}$$

$$U_{dihedral} = \sum_{dihedral\ i} \begin{cases} k_i^{\ dihedral}[1 + \cos(n_i\phi_i - \gamma_i^*)],\ n_i \neq 0 \\ k_i^{\ dihedral}(\phi_i - \gamma_i^*)^2,\ \ n = 0 \end{cases} \qquad \text{(eqn. 1.5)}$$

The stretching interaction (U_{bond}) between two atoms is modeled by a harmonic oscillator system with the bond force constant (k^{bond}) and the displacement from the equilibrium positions of the two atoms (r-r_o). Bending motions involve three covalently bonded atoms (U_{angle}), which is calculated using the angular force constant (k^{angle}) and the displacement from an equilibrium angle (θ^*-θ_o^*). The dihedral rotations formed by four covalently bonded atoms are calculated the multiplicity of the dihedral function (n), the angle (ϕ), and the phase shift (γ^*).

Non-bonded interactions between a pair of atoms, i and j, are described by the van der Waals (vdW) and the Coulomb electrostatic energies,

$$U_{vdW} = \sum_i \sum_{j>i} 4\pi\varepsilon_{ij}^{LJ}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \quad \text{(eqn. 1.6)}$$

$$U_{Coulomb} = \sum_i \sum_{j>i} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \quad \text{(eqn. 1.7)}$$

The vdW energies (U_{vdW}) are calculated using parameters from the Lennard-Jones potentials, including the distance at zeroth potential (σ) and the depth of potential ($\mathcal{E}^{LJ}$) and the distance within the pair of interacting atoms (r). The Coulombic potential ($U_{Coulomb}$) includes parameters such as the electrical permittivity ($\mathcal{E}_0$), charge (q) and the interatomic distance (r_{ij}). These energy functions represent the sum of forces that are exerted on each atom of a given molecule.

Common force fields like CHARMM [71], AMBER [70], and OPLS [72] contain the base parameters for the calculation of (previously described) energy functions. For instance, the force constants (k_{bond}) for all pairs of bonded atoms in the 20 naturally occurring amino acids are included in, e.g., the CHARMM22 [71] force field. However, parameters for novel residues like the isomer of valine, (i.e., norvaline) are not included in the force field. Hence, initial parameterization is necessary to perform a simulation of peptides containing this unusual amino acid. Such force field parameterizations are performed for elastin's crosslinks, for example, as described in Chapter 4.

A typical MD simulation generates a time-dependent trajectory from iterated integrations of Newton's equations of motion (**Figure 1.11a**). The atomic coordinates at a given time point, $x_i(t_0+\Delta t)$, are determined from the integration of acceleration, d^2r_i/dt^2, which is derived from the gradient of potential energy function, U, and the initial coordinates of the molecule, $x_i(t_0)$. This aspect of MD is often called a deterministic method, i.e., the atomic coordinates and velocities of any given time points, t, can be determined, provided the quantities at the initial time point, t_0. During MD simulations, each integration cycle is done iteratively until a certain simulation end-point is reached, where a selected thermodynamic property converges to a constant value in the simulation. The length of the simulation is often selected

based on the time scale of common biological processes (**Figure 1.11b**). For instance, the "hinge bending" of a given protein segment is typically observed in the time scale of 10^{-9} - 1 seconds.

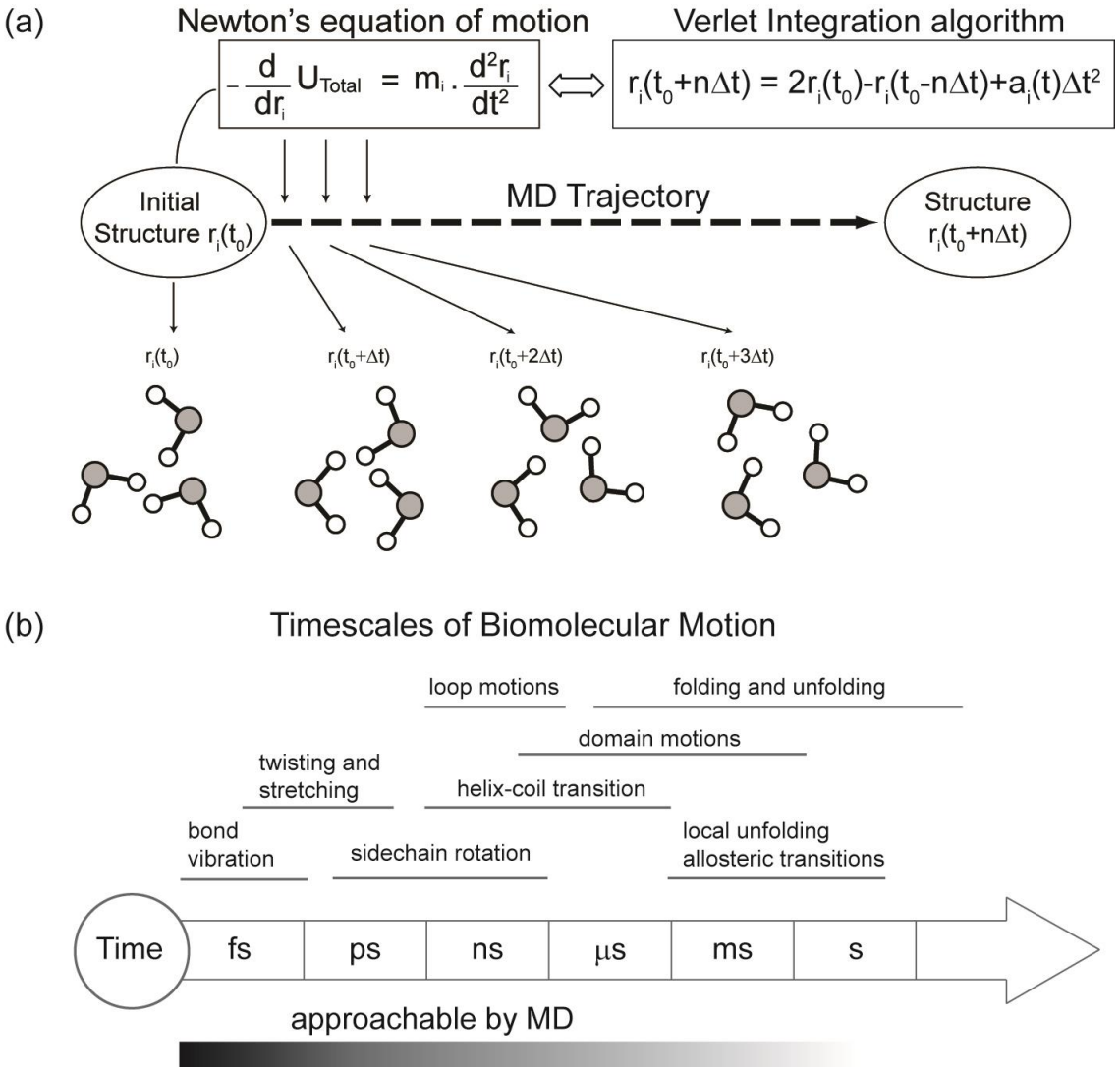


Figure 1.11 (a) Generalized concept of an MD simulation and (b) timescales for biomolecular motion. In (a), typical MD algorithm calculates the atomic coordinates of a given structure from its initial time point (t_0) to latter time points ($t_0+n\Delta t$) Newton's equation of motion. The forces acting upon each atom (r_i) are computed from the gradient of potential energy (U). New atomic coordinates are obtained from the integration of acceleration at each time step (**–**), which is iterated using Verlet algorithm [73].

Water plays an important role in protein folding simulations. Two types of solvent models are applicable for MD simulations, explicit and implicit. Explicit solvation typically provides a more accurate representation for solute hydration, as it implements all-atom representations in the computation of electrostatic interactions. In contrast, implicit solvation simulates the effect of solvents on the solute

without atomistic representations, which offers less computing time as fewer atoms are included in the computation.

In explicit solvation, water molecules are directly in contact with the surface of the solute (**Figure 1.12**). The solute-solvent electrostatic interactions are calculated using the Coulomb potential function described in the force field. Many water models like TIP3P [74], SPC [75], and F3C [76] have been developed for simulations of biological molecules. The TIP3P [74] is most commonly used for protein folding simulations, as it yields fairly accurate thermodynamic properties of water. This model also uses simple potential functions, which are computationally efficient for MD.

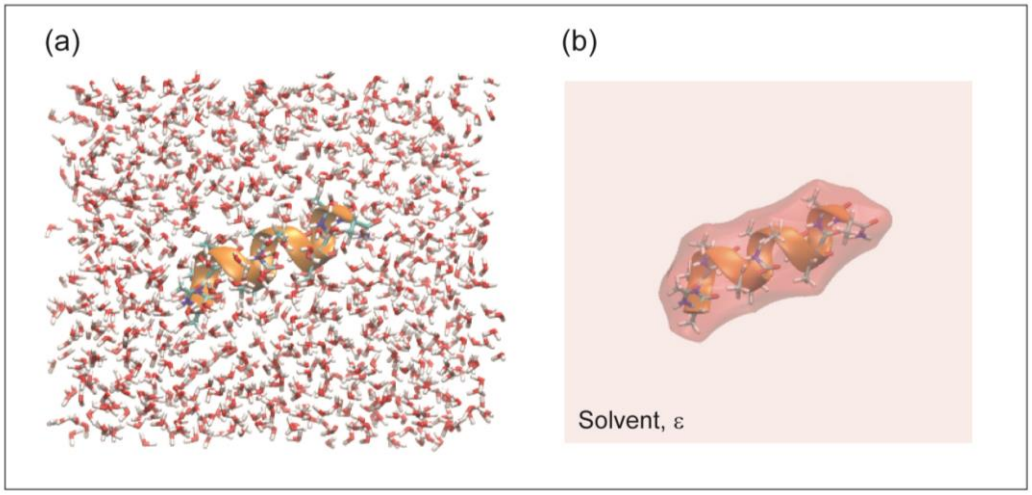


Figure 1.12 Solvation of an α-helical peptide, $(Ala)_{11}$, using (a) explicit and (b) implicit water models. In (a), hydration is treated explicitly, as shown by water molecules (licorice) surrounding the peptide. In (b), hydration is handled implicitly by introducing a continuous medium with a high dielectric constant, ε.

Implicit solvation considers the surrounding environment of the solute as a continuous medium with a high dielectric constant, ε (**Figure 1.12b**). The average solvent effects on the solute is derived from the calculation of free energy of solvation, which is composed nonpolar and polar electrostatic interactions. The polar interactions are typically estimated using the generalized Born (GB) theory [77] or calculated rigorously with the Poisson-Boltzmann (PB) equation [78]. The GB model provides a better computational efficiency than PB, and it offers an analytical expression to evaluate the Coulomb interactions of water molecules. Conversely, nonpolar interactions, or hydrophobic effects, are calculated using the free energy of solvent accessibility (SA), which is a term that simulates the vdW interactions. The combination of GB and SA model has recently become the popular choice for implicit solvent simulations of biological molecules [77], as it simulates both electrostatic and hydrophobic effects.

In summary, MD simulations and ssNMR spectroscopy are effective tools to characterize biological macromolecules like elastin. The combination of these two techniques offers insights into local structures and the mobility in select domains, which are otherwise difficult to obtain with other approaches.

C. GOAL AND OBJECTIVES

The development of molecular models to describe the structure and function of elastin is ongoing. Experimental and computational findings [79-82] challenge the validity of classical models, such as random chain [12], liquid drop [13], fibrillar [15] representations, etc. Tamburro's tropoelastin model mainly describes the secondary structures occurring within the soluble monomer. Moreover, the tropoelastin representation was constructed based on CD and NMR measurements of elastin peptides in a trifluoroethanol/water (80:20) mixture, a solvent-system that induces the formation of secondary structures [24, 83]. For these reasons, a molecular representation of mature, crosslinked elastin is still desired to help understand the structure-function relationship of this protein in its native ECM environment.

The end goal of this research project is to obtain a molecular model for native elastin, an insoluble and biosynthetically crosslinked polymer. This project is primarily focused on the characterization of alanines, the amino acid found abundantly in both crosslinking and hydrophobic domains. Solid-state NMR spectroscopy and MD simulations are the primary tools used for this investigation.

One of the objectives for this project is the expression of elastin samples that are isotopically-labeled at the alanines. Neonatal rat smooth muscle cell (NRSMC) culture has been shown to produce native elastin that is suitable for ssNMR characterization [84]. Multiple enrichment strategies are investigated to obtain samples with isotopically-labeled alanines.

Another research objective is the identification of local secondary structures in elastin's hydrophobic and crosslinking domains. Although α-helices are often reported in the Ala-rich domains of tropoelastin [24], their existence (or lack thereof) in native elastin has not been validated. The cause of low α-helical content in the crosslinking domains is not fully understood. Moreover, the role of crosslinking moieties in these helical domains has not been elucidated. For these reasons, ssNMR spectroscopy is implemented for the characterization of alanines in NRSMC elastin. MD simulations are also performed to gain insights into the structure and dynamics of the Ala-rich crosslinked regions.

The last objective of this study is the determination of temperature effects on the conformational preferences of the alanines in elastin. Like other polymers, elastin undergoes a temperature-dependent glass transition. This transition was previously recognized as a change of protein dynamics from largely "liquid-like" to "solid-like", as elastin was cooled below the freezing point of water. However, this temperature-dependent property has not been studied in the context of the conformations and/or

dynamics of Ala. Therefore, variable-temperature ssNMR experiments are performed on NRSMC elastin samples containing isotopically-labeled alanines, and temperature-replica exchange MD simulations are conducted on the Ala-rich crosslinking domains.

D. SCOPE OF DISSERTATION

D.1 Chapter 2 – Expression of elastin with isotopically-enriched alanines for solid-state NMR measurements

The use of isotopically enriched samples is advantageous for solid-state NMR studies. The NRSMC express insoluble and biosynthetically crosslinked elastin. Multiple strategies to incorporate [U-^{13}C]Ala and [U-^{13}C,^{15}N]Ala in NRSMC elastin were investigated. Liquid chromatography-mass spectrometry (LC-MS) and ssNMR spectroscopy were used to evaluate the enrichment level in [U-^{13}C]Ala NRSMC elastin. The combination of an enzymatic inhibition scheme and a multifold boost of isotopes in the culture media yield samples with high isotopic enrichment (~80%) and without isotopic scrambling.

D.2 Chapter 3 – Investigations of structure and dynamics of alanine residues in elastin by solid-state NMR spectroscopy at physiological temperature

The structure and dynamics of alanines in hydrated NRSMC elastin were investigated at 37° C. Conformation- and sequence-dependence of ^{13}C chemical shifts were considered in the characterization of alanines. The random coil population is prominent in the hydrophobic domains, whereas both α-helices and random coil are present in the crosslinking regions. Selective detection experiments were used to distinguish the mobility of Ala in the crosslinking from those in the hydrophobic domains. Measurements of NMR relaxation parameters support the assignment of Ala's conformations and the assessment of Ala's mobility in NRSMC elastin.

D.3 Chapter 4 – Elucidation of structure and dynamics of elastin's crosslinking domains by molecular dynamics simulations

The random coil and α-helical populations of alanines in elastin's crosslinking domains were further investigated by MD simulations. Computations were performed on the two Ala-rich domains that are critical to elastin assembly, i.e., EX19 and EX25. To mimic the native ECM environment, DLNL and ALDL (bifunctional) crosslinks were also included in the simulation of these two domains. The analyses were focused on identifying helix- and coil-like structures in MD trajectories from the calculation of root-mean-square deviations (RMSD) and the quantification of %-helical content. The free-energy landscapes of peptide folding also provide insights into possible helix-to-coil or coil-to-helix transitions in the α-helical domains. The roles of backbone hydration and Lys-derived crosslinks in stabilizing the α-helices of EX25

and EX19 were also investigated. MD results indicate the existence of helix-coil equilibrium in the crosslinking domains of elastin.

D.4 Chapter 5 – Variable-temperature solid-state NMR studies of alanines in hydrated elastin

The helix-coil equilibrium in Ala-rich crosslinking domains was investigated using ssNMR spectroscopy. Variable-temperature experiments were performed on hydrated NRSMC elastin to determine the effect of temperature on the structure and dynamics of the alanines. Significant differences in the α-helical content in elastin's crosslinking domains were observed above and below the glass transition temperature (T_g). The helix-coil transition theory was considered in the interpretation of the corresponding ssNMR spectra. The statistical mechanical formulation of the helix-coil transition and simulations of NMR lineshapes provided the means to understand the temperature-dependent helical propensities in Ala-rich crosslinking domains in NRSMC elastin.

D.5 Chapter 6 – Variable-temperature studies of elastin's crosslinking domains molecular dynamics simulations

Temperature-replica exchange MD simulations were used to examine the helix-coil equilibrium in elastin's crosslinking domains. Computations were performed on EX19 and EX25 domains containing DLNL and ALDL crosslinks, respectively. The α-helical content in these peptides was evaluated as a function of temperature. Populations of helix- and coil-like structures in the trajectories were identified with RMSD calculations and hydrogen-bond evaluations. The free-energy landscapes of peptide folding indicated a helix-coil transition with the increase of temperature. Torsion angle and root-mean-square fluctuations provided insights into the α-helical and random coil Ala populations in these domains.

D.6 Chapter 7 – Summary and concluding remarks

The conclusion chapter highlights the ssNMR and MD results and their significance to elastin research. A new model for elastin's structure and function is also proposed based on these results.

E. REFERENCES

1. Davis, E.C., *Stability of elastin in the developing mouse aorta: a quantitative radioautographic study.* Histochemistry, 1993. **100**(1): p. 17-26.
2. Shapiro, S.D., S.K. Endicott, M.A. Province, J.A. Pierce, and E.J. Campbell, *Marked longevity of human lung parenchymal elastic fibers deduced from prevalence of d-aspartate and nuclear weapons-related radiocarbon.* Journal of Clinical Investigation, 1991. **87**(5): p. 1828-34.
3. Keeley, F.W., C.M. Bellingham, and K.A. Woodhouse, *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.* Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences, 2002. **357**(1418): p. 185-189.
4. Debelle, L., and A.J.P. Alix, *The structures of elastins and their function.* Biochimie, 1999. **81**(10): p. 981-994.
5. Debelle, L., and A.M. Tamburro, *Elastin: molecular description and function.* International Journal of Biochemistry & Cell Biology, 1999. **31**(2): p. 261-272.
6. Wagenseil, J.E., and R.P. Mecham, *New insights into elastic fiber assembly.* Birth Defects Research Part C: Embryo Today: Reviews, 2007. **81**(4): p. 229-240.
7. Clarke, A.W., E.C. Arnspang, S.M. Mithieux, E. Korkmaz, F. Braet, and A.S. Weiss, *Tropoelastin massively associates during coacervation to form quantized protein spheres.* Biochemistry, 2006. **45**(33): p. 9989-9996.
8. Kozel, B.A., B.J. Rongish, A. Czirok, J. Zach, C.D. Little, E.C. Davis, R.H. Knutsen, J.E. Wagenseil, M.A. Levy, and R.P. Mecham, *Elastic fiber formation: a dynamic view of extracellular matrix assembly using timer reporters.* Journal of Cellular Physiology, 2006. **207**(1): p. 87-96.
9. Vrhovski, B., S. Jensen, and A.S. Weiss, *Coacervation characteristics of recombinant human tropoelastin.* European Journal of Biochemistry, 1997. **250**(1): p. 92-98.
10. Baldock, C., A.F. Oberhauser, L. Ma, D. Lammie, V. Siegler, S.M. Mithieux, Y. Tu, J.Y.H. Chow, F. Suleman, M. Malfois, S. Rogers, L. Guo, T.C. Irving, T.J. Wess, and A.S. Weiss, *Shape of tropoelastin, the highly extensible protein that controls human tissue elasticity.* Proceedings of the National Academy of Sciences, 2011. **108**(11): p. 4322-4327.
11. Brown-Augsburger, P., C. Tisdale, T. Broekelmann, C. Sloan, and R.P. Mecham, *Identification of an elastin cross-linking domain that joins three peptide chains: possible role in nucleated assembly.* Journal of Biological Chemistry, 1995. **270**(30): p. 17778-17783.
12. Hoeve, C.A.J., and P.J. Flory, *The elastic properties of elastin.* Biopolymers, 1974. **13**(4): p. 677-686.
13. Partridge, S.M., D.F. Elsden, and J. Thomas, *Constitution of the cross-linkages in elastin.* Nature, 1963. **197**(4874): p. 1297-1298.

14. Weis-Fogh, T., and S.O. Andersen, *New molecular model for the long-range elasticity of elastin.* Nature, 1970. **227**(5259): p. 718-721.
15. Urry, D.W., *What is elastin; what is not.* Ultrastructural Pathology, 1983. **4**(2-3): p. 227-51.
16. Partridge, S., *Elastin.* Adv Protein Chem, 1963. **17**: p. 227-302.
17. Cox, B.A., B.C. Starcher, and D.W. Urry, *Communication: Coacervation of tropoelastin results in fiber formation.* Journal of Biological Chemistry, 1974. **249**(3): p. 997-8.
18. Urry, D.W., *Arterial mesenchyme and arteriosclerosis. Studies on the conformation and interaction of elastin.* Advances in Experimental Medicine and Biology, 1974. **43**(0): p. 211-43.
19. Gross, P.C., W. Possart, and M. Zeppezauer, *An alternative structure model for the polypentapeptide in elastin.* Zeitschrift für Naturforschung C, 2003. **58**(11-12): p. 873-878.
20. Li, B., and V. Daggett, *Molecular basis for the extensibility of elastin.* Journal of Muscle Research & Cell Motility, 2003. **23**(5): p. 561-573.
21. Bashir, M.M., Z. Indik, H. Yeh, N. Ornstein-Goldstein, J.C. Rosenbloom, W. Abrams, M. Fazio, J. Uitto, and J. Rosenbloom, *Characterization of the complete human elastin gene. Delineation of unusual features in the 5'-flanking region.* Journal of Biological Chemistry, 1989. **264**(15): p. 8887-91.
22. Vrhovski, B., and A.S. Weiss, *Biochemistry of tropoelastin.* European Journal of Biochemistry, 1998. **258**(1): p. 1-18.
23. Tamburro, A.M., B. Bochicchio, and A. Pepe, *Dissection of human tropoelastin: exon-by-exon chemical synthesis and related conformational studies.* Biochemistry, 2003. **42**(45): p. 13347-13362.
24. Tamburro, A.M., A. Pepe, and B. Bochicchio, *Localizing α-helices in human tropoelastin: assembly of the elastin "puzzle".* Biochemistry, 2006. **45**(31): p. 9518-9530.
25. Tamburro, A.M., V. Guantieri, L. Pandolfo, and A. Scopa, *Synthetic fragments and analogues of elastin. II. Conformational studies.* Biopolymers, 1990. **29**(4-5): p. 855-70.
26. Eyre, D.R., M.A. Paz, and P.M. Gallop, *Cross-linking in collagen and elastin.* Annual Review of Biochemistry, 1984. **53**(1): p. 717-748.
27. Guay, M., and F. Lamy, *The troublesome crosslinks of elastin.* Trends in Biochemical Sciences, 1979. **4**(7): p. 160-164.
28. Lent, R.W., B. Smith, L.L. Salcedo, B. Faris, and C. Franzblau, *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): p. 2837-45.
29. Akagawa, M., and K. Suyama, *Mechanism of formation of elastin crosslinks.* Connective Tissue Research, 2000. **41**(2): p. 131-41.

30. Francis, G., R. John, and J. Thomas, *Biosynthetic pathway of desmosines in elastin.* Biochemical Journal, 1973. **136**(1): p. 45-55.
31. Nakamura, F., K. Yamazaki, and K. Suyama, *Isolation and structural characterization of a new crosslinking amino acid, cyclopentenosine, from the acid hydrolysate of elastin.* Biochemical and Biophysical Research Communications, 1992. **186**(3): p. 1533-1538.
32. Akagawa, M., K. Yamazaki, and K. Suyama, *Cyclopentenosine, major trifunctional crosslinking amino acid isolated from acid hydrolysate of elastin.* Archives of Biochemistry and Biophysics, 1999. **372**(1): p. 112-120.
33. Schiffer, M., and A.B. Edmundson, *Use of helical wheels to represent the structures of proteins and to identify segments with helical potential.* Biophysical Journal, 1967. **7**(2): p. 121-35.
34. Gray, W.R., L.B. Sandberg, and J.A. Foster, *Molecular model for elastin structure and function.* Nature, 1973. **246**(5434): p. 461-6.
35. Baig, K.M., M. Vlaovic, and R.A. Anwar, *Amino acid sequences C-terminal to the cross-links in bovine elastin.* Biochemistry Journal, 1980. **185**(3): p. 611-616.
36. Foster, J.A., L. Rubin, H.M. Kagan, C. Franzblau, E. Bruenger, and L.B. Sandberg, *Isolation and characterization of cross-linked peptides from elastin.* Journal of Biological Chemistry, 1974. **249**(19): p. 6191-6.
37. Chou, P.Y., and G.D. Fasman, *Conformational parameters for amino acids in helical, β-sheet, and random coil regions calculated from proteins.* Biochemistry, 1974. **13**(2): p. 211-222.
38. P Y Chou, a., and G.D. Fasman, *Empirical predictions of protein conformation.* Annual Review of Biochemistry, 1978. **47**(1): p. 251-276.
39. Toonkool, P., S.A. Jensen, A.L. Maxwell, and A.S. Weiss, *Hydrophobic domains of human tropoelastin interact in a context-dependent manner.* Journal of Biological Chemistry, 2001. **276**(48): p. 44575-44580.
40. Dabalos, C.L., *Characterization of the proline residues in elastin*. 2016, University of Hawaii at Manoa: Ann Arbor. p. 161.
41. He, D., M. Chung, E. Chan, T. Alleyne, K.C. Ha, M. Miao, R.J. Stahl, F.W. Keeley, and J. Parkinson, *Comparative genomics of elastin: sequence analysis of a highly repetitive protein.* Matrix Biology, 2007. **26**(7): p. 524-40.
42. Pierce, R.A., A. Alatawi, S.B. Deak, and C.D. Boyd, *Elements of the rat tropoelastin gene associated with alternative splicing.* Genomics, 1992. **12**(4): p. 651-658.
43. Ramachandran, G.N., C. Ramakrishnan, and V. Sasisekharan, *Stereochemistry of polypeptide chain configurations.* Journal of Molecular Biology, 1963. **7**: p. 95-9.

44. *IUPAC-IUB commission on biochemical nomenclature. Abbreviations and symbols for the description of the conformation of polypeptide chains. Tentative rules (1969).* Biochemical Journal, 1971. **121**(4): p. 577-585.
45. Rose, G.D., L.M. Glerasch, and J.A. Smith, *Turns in peptides and proteins.* Advances in Protein Chemistry, 1985. **37**: p. 1-109.
46. Flory, P.J., *Statistical thermodynamics of mixtures of rodlike particles. 5. Mixtures with random coils.* Macromolecules, 1978. **11**(6): p. 1138-1141.
47. Flory, P.J., *Principles of polymer chemistry*. 1953: Cornell University Press.
48. Smith, L.J., K.M. Fiebig, H. Schwalbe, and C.M. Dobson, *The concept of a random coil: residual structure in peptides and denatured proteins.* Folding Design, 1996. **1**(5): p. R95-R106.
49. Toal, S., and R. Schweitzer-Stenner, *Local order in the unfolded state: conformational biases and nearest neighbor interactions.* Biomolecules, 2014. **4**(3): p. 725-773.
50. Jha, A.K., A. Colubri, K.F. Freed, and T.R. Sosnick, *Statistical coil model of the unfolded state: resolving the reconciliation problem.* Proceeding of the National Academy of Sciences of the United States of America, 2005. **102**(37): p. 13099-13104.
51. Vila, J., H. Baldoni, D. Ripoll, and H. Scheraga, *Unblocked statistical-coil tetrapeptides in aqueous solution: quantum-chemical computation of the carbon-13 NMR chemical shifts.* Journal of Biomolecular NMR, 2003. **26**(2): p. 113-130.
52. Vila, J., D. Ripoll, H. Baldoni, and H. Scheraga, *Unblocked statistical-coil tetrapeptides and pentapeptides in aqueous solution: A theoretical study.* Journal of Biomolecular NMR, 2002. **24**(3): p. 245-262.
53. Roberts, S., M. Dzuricky, and A. Chilkoti, *Elastin-like polypeptides as models of intrinsically disordered proteins.* FEBS Letter, 2015. **589**(19): p. 2477-2486.
54. Debelle, L., A.J. Alix, S.M. Wei, M.P. Jacob, J.P. Huvenne, M. Berjot, and P. Legrand, *The secondary structure and architecture of human elastin.* European Journal of Biochemistry, 1998. **258**(2): p. 533-9.
55. Tycko, R., *Solid-state NMR studies of amyloid fibril structure.* Annual Review of Physical Chemistry, 2011. **62**: p. 279-299.
56. Petkova, A.T., Y. Ishii, J.J. Balbach, O.N. Antzutkin, R.D. Leapman, F. Delaglio, and R. Tycko, *A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR.* Proceedings of the National Academy of Sciences, 2002. **99**(26): p. 16742-16747.
57. Huster, D., *Chapter 4 Solid-state NMR studies of collagen structure and dynamics in isolated fibrils and in biological tissues*, in *Annual Reports on NMR Spectroscopy*, A.W. Graham, Editor. 2008, Academic Press. p. 127-159.

58. Naito, A., S. Tuzi, and H. Saitô, *A high-resolution ^{15}N solid-state NMR study of collagen and related polypeptides.* European Journal of Biochemistry, 1994. **224**(2): p. 729-734.
59. Simmons, A., E. Ray, and L.W. Jelinski, *Solid-state ^{13}C NMR of nephila clavipes dragline silk establishes structure and identity of crystalline regions.* Macromolecules, 1994. **27**(18): p. 5235-5237.
60. van Beek, J.D., L. Beaulieu, H. Schafer, M. Demura, T. Asakura, and B.H. Meier, *Solid-state NMR determination of the secondary structure of Samia cynthia ricini silk.* Nature, 2000. **405**(6790): p. 1077-1079.
61. Separovic, F., and A. Naito, *Advances in biological solid-state NMR: proteins and membrane-active peptides*. 2014: Royal Society of Chemistry.
62. Saitô, H., I. Ando, and A. Naito, *Solid state NMR spectroscopy for biopolymers: principles and applications*. 2006: Springer Netherlands.
63. Saito, H., R. Tabeta, A. Shoji, T. Ozaki, and I. Ando, *Conformational characterization of polypeptides in the solid state as viewed from the conformation-dependent carbon-13 chemical shifts determined by the carbon-13 cross polarization/magic angle spinning method: oligo(L-alanine), poly(L-alanine), copolymers of L- and D-alanines, and copolymers of L-alanine with N-methyl- or N-benzyl-L-alanine.* Macromolecules, 1983. **16**(7): p. 1050-1057.
64. Saito, H., *Conformation-dependent ^{13}C chemical shifts: a new means of conformational characterization as obtained by high-resolution solid-state ^{13}C NMR.* Magnetic Resonance in Chemistry, 1986. **24**(10): p. 835-852.
65. Havlin, R.H., H. Le, D.D. Laws, A.C. deDios, and E. Oldfield, *An ab initio quantum chemical investigation of carbon-13 NMR shielding tensors in glycine, alanine, valine, isoleucine, serine, and threonine: comparisons between helical and sheet tensors, and the effects of χ1 on shielding.* Journal of the American Chemical Society, 1997. **119**(49): p. 11951-11958.
66. Sun, H., L.K. Sanders, and E. Oldfield, *Carbon-13 NMR shielding in the twenty common amino acids: comparisons with experimental results in proteins.* Journal of the American Chemical Society, 2002. **124**(19): p. 5486-5495.
67. Duer, M.J., *Introduction to solid-state NMR spectroscopy*. 2005: Wiley.
68. Bennett, A.E., C.M. Rienstra, M. Auger, K.V. Lakshmi, and R.G. Griffin, *Heteronuclear decoupling in rotating solids.* Journal of Chemical Physics, 1995. **103**(16): p. 6951-6958.
69. MacKerell, A.D., D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, *All-atom*

empirical potential for molecular modeling and dynamics studies of proteins. Journal of Physical Chemistry B, 1998. **102**(18): p. 3586-3616.

70. Cornell, W.D., P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, and P.A. Kollman, *A second generation force field for the simulation of proteins, nucleic acids, and organic molecules.* Journal of the American Chemical Society, 1995. **117**(19): p. 5179-5197.
71. Brooks, B.R., R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, and M. Karplus, *CHARMM: A program for macromolecular energy, minimization, and dynamics calculations.* Journal of Computational Chemistry, 1983. **4**(2): p. 187-217.
72. Jorgensen, W.L., and J. Tirado-Rives, *The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin.* Journal of the American Chemical Society, 1988. **110**(6): p. 1657-1666.
73. Verlet, L., *Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules.* Physical Review, 1967. **159**(1): p. 98-103.
74. Jorgensen, W.L., J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein, *Comparison of simple potential functions for simulating liquid water.* Journal of Chemical Physics, 1983. **79**(2): p. 926-935.
75. Berendsen, H.J.C., J.P.M. Postma, W.F. van Gunsteren, and J. Hermans, *Interaction models for water in relation to protein hydration*, in *Intermolecular Forces: Proceedings of the Fourteenth Jerusalem Symposium on Quantum Chemistry and Biochemistry Held in Jerusalem, Israel, April 13–16, 1981*, B. Pullman, Editor. 1981, Springer Netherlands: Dordrecht. p. 331-342.
76. Levitt, M., M. Hirshberg, R. Sharon, K.E. Laidig, and V. Daggett, *Calibration and testing of a water model for simulation of the molecular dynamics of proteins and nucleic acids in solution.* Journal of Physical Chemistry B, 1997. **101**(25): p. 5051-5061.
77. Tsui, V., and D.A. Case, *Theory and applications of the generalized born solvation model in macromolecular simulations.* Biopolymers, 2000. **56**(4): p. 275-291.
78. Fogolari, F., A. Brigo, and H. Molinari, *The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology.* Journal of Molecular Recognition, 2002. **15**(6): p. 377-92.
79. Debelle, L., A.J. Alix, M.P. Jacob, J.P. Huvenne, M. Berjot, B. Sombret, and P. Legrand, *Bovine elastin and kappa-elastin secondary structure determination by optical spectroscopies.* Journal of Biological Chemistry, 1995. **270**(44): p. 26099-103.
80. Li, B., D.O.V. Alonso, B.J. Bennion, and V. Daggett, *Hydrophobic hydration is an important source of elasticity in elastin-based biopolymers.* Journal of the American Chemical Society, 2001. **123**(48): p. 11991-11998.

81. Li, B., D.O.V. Alonso, and V. Daggett, *The molecular basis for the inverse temperature transition of elastin.* Journal of Molecular Biology, 2001. **305**(3): p. 581-592.
82. Li, B., and V. Daggett, *Molecular basis for the extensibility of elastin.* Journal of Muscle Research & Cell Motility, 2002. **23**(5): p. 561-573.
83. Tamburro, A.M., B. Bochicchio, and A. Pepe, *The dissection of human tropoelastin: from the molecular structure to the self-assembly to the elasticity mechanism.* Pathologie Biologie, 2005. **53**(7): p. 383-389.
84. Perry, A., M.P. Stypa, J.A. Foster, and K.K. Kumashiro, *Observation of the glycines in elastin using ^{13}C and ^{15}N solid-state NMR spectroscopy and isotopic labeling.* Journal of the American Chemical Society, 2002. **124**(24): p. 6832-6833.

CHAPTER 2. EXPRESSION OF ELASTIN WITH ISOTOPICALLY-ENRICHED ALANINES FOR SOLID-STATE NMR MEASUREMENTS

A. INTRODUCTION

Solid-state NMR (SSNMR) characterization of elastin provides insight into the protein's molecular structure in the elastic fiber. Unlike the other amino acids, alanines are prominently found in both the hydrophobic and the crosslinking domains of elastin. High-resolution ssNMR studies of Ala are achievable using samples with isotopically enriched samples. However, the isotopic enrichment of alanines in native elastin had not been successfully implemented until this study.

Insoluble elastin with ^{13}C enrichment at its alanines is produced using a mammalian cell expression system (section A.1). Several schemes designed to incorporate isotopically-enriched Ala into elastin were investigated (section A.2). Mass spectrometry and one-dimensional ssNMR spectroscopy were used to evaluate the efficacy of the implemented strategies (sections B-C). Details on the structure and dynamics of elastin with isotopically-enriched Ala residues will be described in Chapter 3.

A.1 Mammalian cell-based expression system to produce elastin with isotopic enrichment of targeted amino acids

The holistic approach of elastin studies includes samples from mammalian tissues, such as bovine nuchal ligament (BNL) [1, 2]. Tissue-derived elastin is insoluble and biosynthetically crosslinked. However, it contains only naturally-occurring NMR-active isotopes, often at low abundance, and the characterization of backbone conformation using ^{13}C NMR chemical shifts is often hampered by significant peak overlap. For instance, the ^{13}CO signals of Ala in BNL are poorly resolved from those of the prolines, making it difficult to isolate the NMR properties of the targeted residue type.

The difficulty of high-resolution ssNMR studies of elastin is the lack of a suitable expression system that allows for isotopic labeling. Solid-phase peptide synthesis (SPPS) is not feasible due to the polymer size, so the production of a full-length tropoelastin has been previously accomplished by the *Escherichia coli* gene expression system [4]. However, bacterial expression also presents challenges, as these systems lack the necessary enzymes and cellular chaperones to polymerize tropoelastin (monomers) into a highly crosslinked polymer that mimics the native system.

The reductionist approach utilizes model polypeptides that are less complex than native elastin. These peptides are obtained either by solution peptide chemistry, as in single-domain studies [5-7], or by

bacterial expression, as demonstrated with mimetics with alternating crosslinking and hydrophobic domains [8, 9]. Chemical crosslinking has also been introduced in elastin-like peptides (ELP) to produce samples with elastic moduli and other physicochemical properties that are similar (but not identical) to elastin [9-11]. However, chemical crosslinks in these ELPs do not replicate desmosine and related crosslinking moieties found in native elastin.

Insoluble, isotopically enriched, and biosynthetically crosslinked elastin is expressed by mammalian vascular smooth muscle cells [12]. Insoluble elastin is harvested from the extracellular matrix (ECM) of neonatal rat smooth muscle cells (NRSMC) cultures, where tropoelastin is polymerized and crosslinked by cellular chaperones and enzymes [12-14]. Previously, elastin samples with ^{15}N-Gly and ^{13}C-Gly enrichment were expressed by NRSMC's for ssNMR characterization of the glycines [3]. In this study, the Gly that is normally supplied in the growth media was replaced by isotopically-labeled Gly. This approach (*direct substitution*) was effective, because glycine's endogenous production in mammalian neonates and fetuses is inadequate for optimal growth [15, 16]. Hence, the demand for this amino acid in NRSMC cultures is supplied via uptake from the culture media. The substitution of Gly with ^{13}C-enriched Gly resulted in elastin samples with adequate enrichment levels (35-40%) for the ssNMR studies [3].

However, direct substitution is not effective for the expression of isotopically-enriched alanines in NRSMC elastin. Attempts to implement this scheme for ^{13}C-Ala were previously made by the Kumashiro lab (unpublished data). Briefly, [1-^{13}C]Ala was supplied in the media at lower concentrations than with Gly; i.e., the typical concentration of Ala in the media is 9 mg/L, in contrast to the 34 mg/L for Gly. As expected, little enhancement of the ^{13}C NMR signals for alanines in elastin was observed, reflecting low ^{13}C-Ala enrichment levels.

The low incorporation level of enriched Ala that results from *direct substitution* was attributed to the presence of or competition with intracellular (unenriched) Ala. In mammalian cells, the endogenous supply of Ala comes from the conversion of pyruvate by the alanine transaminase (ALT) enzyme (**Figure 2.1**), and the intracellular pyruvate is obtained from the breakdown of glucose, a supplied nutrient in the media. Therefore, the low ^{13}C enrichment level in the protein with direct substitution was likely due to the utilization of (unenriched) intracellular Ala during elastin synthesis, as the amount of glucose (1000 mg/L) was significantly greater than the 9 mg per liter of media that was used for ^{13}C-enriched Ala.

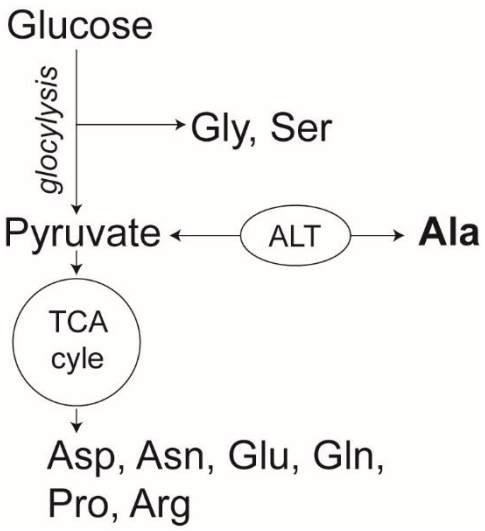


Figure 2.1 Endogenous synthesis of alanine and other amino acids in NRSMC. The production of **Ala** from pyruvate is facilitated by alanine transaminase (ALT) enzyme. Pyruvate is (also) obtained from the breakdown of glucose (glycolysis). Pyruvate enters the tricarboxylic acid (TCA) cycle, where Asp, Asn, Glu, Gln, Pro, and Arg are produced. Arrows indicate the direction of the metabolic pathway.

Although unsuccessful, results of direct substitution provided clues to enhancing the level of isotopic enrichment of Ala in NRSMC elastin. Namely, the effective strategy must target the intra and extracellular sources of this amino acid. Several methods to incorporate isotopes into the two cellular pools of Ala will be described in the next section.

A.2 Strategies to incorporate ^{13}C-Ala into elastin

An effective strategy to produce elastin with high levels of isotopically-enriched Ala targets the pathways by which isotopes enter the system. The direct sources of Ala for protein syntheses in NRSMC are the metabolic by-product of the (endogenous) breakdown of glucose and the (exogenous) supply in the culture media. The following subsections describe several strategies to incorporate ^{13}C-Ala into elastin by exploiting the two delivery routes.

A.2.1 Direct substitution of ^{13}C-Ala and ^{13}C enrichment of alanine's metabolic precursors. Initial experiments focused on refinement of direct substitution. From a practical point of view, the direct substitution strategy is simple and straightforward to use, because it requires no modification of Dulbecco's Modified Eagle's Medium (DMEM), a primary ingredient in NRSMC culture media (details in Appendix 1). The ^{13}C- or ^{15}N-Ala isotope is instead provided in the non-essential amino acids (NEAA) solution.

Unlike direct substitution which supplies isotopes extracellularly, the second strategy uses ^{13}C-enriched glucose and ^{13}C-pyruvate, which are the endogenous sources of ^{13}C-Ala. The use of ^{13}C-glucose in NRMSC culture facilitates the intracellular synthesis of ^{13}C-enriched alanine. Each glucose molecule is converted into two molecules of Ala via its precursor, pyruvate. As illustrated on **Figure 2.2a**, the supply of ^{13}C-labeled glucose in the media will result in the production of intracellular ^{13}C-enriched Ala. The cellular concentration of ^{13}C-Ala is expected be high in NRSMC, because the amount of glucose provided in the media is large (1000 mg/L), relative to the amount of Ala in the NEAA solution (9 mg/L). For this reason, the supply of ^{13}C-glucose is expected to significantly increase the incorporation of ^{13}C-Ala in elastin. However, the disadvantage with ^{13}C-glucose is the metabolic scrambling effect. Namely, isotopes will be distributed into other amino acids downstream of glycolysis (Gly, Ser) and the TCA cycle (Asp, Pro, etc.). Alternatively, the use of ^{13}C-enriched pyruvate is expected to yield ^{13}C-Ala (**Figure 2.2b**), without scrambling into Gly and Ser.

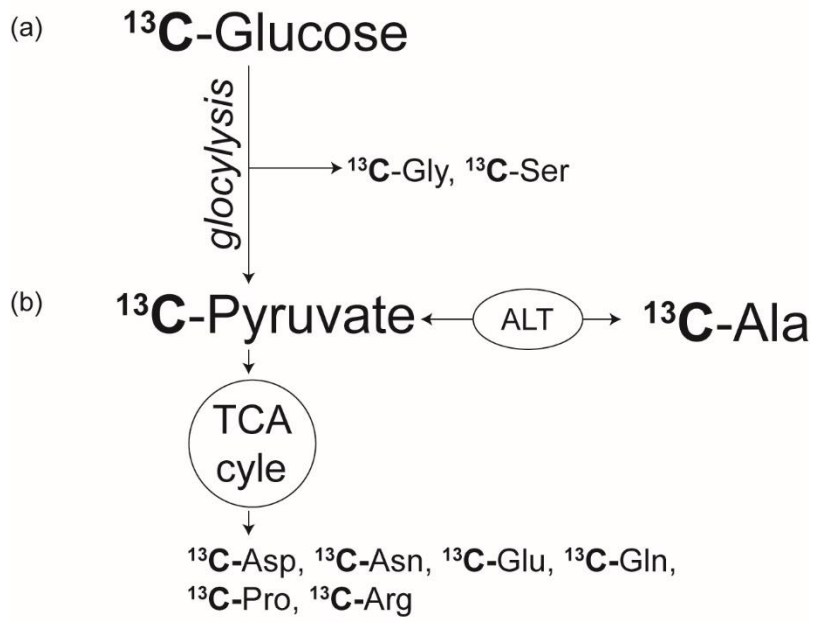


Figure 2.2 Intracellullar synthesis of ^{13}C-Ala in NRSMC by the metabolism of (a) ^{13}C-enriched glucose and (b) ^{13}C-enriched pyruvate. Alanine transaminase (ALT) enzyme converts ^{13}C-pyruvate to ^{13}C-Ala. ^{13}C-pyruvate is obtained either from the breakdown of ^{13}C-glucose (intracellularly) or from the culture media (extracellularly). ^{13}C-pyruvate enters the TCA cycle, where ^{13}C-enriched Asp, Asn, Glu, Gln, Pro and Arg are produced.

The third approach targets both the endogenous and exogenous sources of isotopically-enriched Ala. This strategy combines direct substitution of (extracellular) ^{13}C-Ala with the use of sodium ^{13}C-pyruvate solution to produce the intracellular alanine. The use of two sources is expected to increase the total concentration of ^{13}C-enriched alanine during elastin synthesis.

A.2.2 Enzymatic inhibition of synthesis of intracellular, unenriched Ala, with supplementation of extracellular ^{13}C-enriched Ala. The incorporation of isotopes into elastin may also be accomplished by inhibition of the intracellular production of Ala, accompanied by the supply of an extracellular source via the culture media. This enrichment scheme focuses on inhibition of the enzyme that converts pyruvate to Ala (**Figure 2.3a,b**). In addition, isotopically-enriched Ala is supplemented in the culture media at a higher concentration than the typical formulation for NRSMC culture. The excess is necessary, because a complete inhibition of ALT shifts the cellular demand for alanine to an exogenous source. Hence, the requirement for this amino acid must be met through the culture media to maintain normal metabolism in the cells.

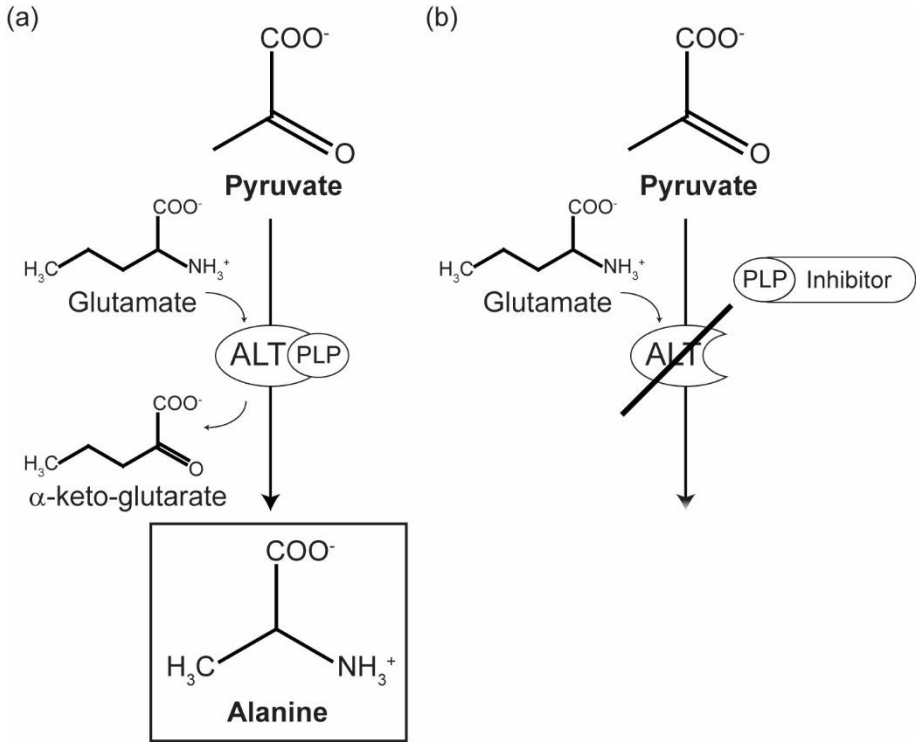


Figure 2.3 (a) Normal and (b) inhibited intracellular synthesis of alanine. The conversion of pyruvate is facilitated by alanine transaminase (ALT). ALT is inactive when its cofactor, pyridoxal phosphate (PLP), is bound to an inhibitor.

The combination of ALT inhibition and a multifold increase of isotope concentration in the media has two distinct advantages. First, isotopic scrambling is minimal. I.e., Ala is the only enriched amino acid. Specifically, the inhibition of ALT activity should prevent the conversion of ^{13}C-Ala into ^{13}C-pyruvate, which enters the TCA cycle to produce amino acids such as Asp, Glu, Pro, etc. Second, this method expands the range of isotopically-enriched Ala that can be incorporated into elastin. For instance, the expression of elastin with ^{15}N-Ala enrichment is feasible by this strategy; however, ^{15}N-labeling is not

achievable by the supplementation of metabolic precursors, because ^{15}N sites are not present in the backbone of glucose or pyruvate molecules.

The NRSMC line is a suitable expression system to produce native elastin, i.e., an insoluble and biosynthetically crosslinked sample, with isotopic enrichment at the alanines. The high level of Ala enrichment in NRSMC elastin will facilitate the investigation of structure and dynamics by ssNMR spectroscopy. The characterization of Ala in both hydrophobic and crosslinking domains will elucidate the molecular structure of elastin in the elastic fiber (Chapter 3-6).

B. MATERIALS AND METHODS

B.1 Elastin synthesis with NRSMC cultures and elastin purification

The production of insoluble elastin using NRSMC culture has been reported previously [12-14]. The composition of the growth media, the non-essential amino acid mixture, and the DMEM are shown in **Appendix 1**.

Briefly, 10-14-day timed-pregnant Sprague-Dawley rats (Charles River Laboratories, Wilmington, MA) were purchased, and the aortae from the newborn (2-3 day old) pups were extracted with aseptic technique. Following extraction, the tissue was digested with collagenase and elastase in a serum-free culture media at 37 °C for 30-45 minutes. The digestion period was ended by the addition of media containing 10% fetal bovine serum. The mixture was transferred into a centrifuge tube and spun at 1000 rpm for 10 minutes. The pellets were suspended with (fresh) media in the centrifuge tube. This suspension was subsequently seeded in a T-75 culture flask (Corning Inc, Corning, NY) and then incubated in the growth media for 5-7 days, until they reached 70-90% confluency. The cells were then digested using trypsin for 5-6 minutes at 37 °C. Following this digestion, the solution was centrifuged at 1000 rpm for 10 minutes, and the pellets were suspended in a solvent containing 90% fetal bovine serum (FBS) and 10% dimethyl sulfoxide for long-term storage in liquid nitrogen.

Each culture began with the resuspension of the frozen pellet in a standard culture media at 37 °C. These cells were transferred into one or two T-75 flask(s) supplied with either the standard culture media or the isotopically-enriched culture media. For the assessment of cell viability and the determination of isotopic enrichment levels, cultures were grown in a 6-well plate for 6-8 weeks. For the production of ssNMR samples, live cells were grown in two T-75 flasks, yielding roughly 300-400 mg of unpurified wet sample. At the end of the incubation period, the ECM layer was collected using cell scrapers (Fisher Scientific, Hampton, NH) for subsequent purification of elastin.

The methods for elastin purification have been reported elsewhere [17]. Here, overnight digestion with cyanogen bromide (CNBr) was employed [8]. After 6-8 weeks of culture, the collected cell matrix was digested in a 20 mL solution of CNBr (50 mg/mL in 70% formic acid) in a 50-mL centrifuge tube for 18-24 hours. Then, 20-40 mL of water were added to the mixture and left under the hood for 4 hours, as excess HCN gas evolves during this time. The suspension was centrifuged, and the supernatant was removed. A solution containing 5 M guanidinium hydrochloride (20 mL) and 1% β-mercaptoethanol (0.2 mL) was added to the tube, and the mixture was incubated overnight. This mixture was then centrifuged, and the supernatant was removed. After 4-5 water washes, the remaining insoluble elastin (~120 mg) was kept in ultrapure water. The typical weight of hydrated elastin for a 4.0 mm ssNMR rotor is 50-60 mg.

B.2 Isotopic enrichment strategies

Samples were prepared using several isotopic enrichment strategies described in the *Introduction* (section **A.2**). A typical NRSMC culture media is comprised of DMEM (87%), FBS (10%), sodium pyruvate (1%), NEAA solution (1%) and antibiotic-antimycotic solution (1%). The formulations are shown in **Appendices 1 and 2**.

B.2.1 Replacement with [1-^{13}C]Ala in the culture media. The direct substitution strategy replaces Ala, one of the constituents in the NEAA solution, with isotopically-enriched Ala. Initially, to assess the isotopic enrichment level with this approach, [1-^{13}C, 99%]Ala (Cambridge Isotope Laboratories, Tewksbury, MA) was used. For these cultures, the NEAA solution was prepared by adding enriched alanine to autoclaved, ultrapure water. Two isotope concentrations (30 mg/L, 1×; 750 mg/L, 25×). were prepared for the assessment of the efficacy of this method.

B.2.2 Utilization of metabolic precursors to Ala, [1-^{13}C]pyruvate and (or) [U-^{13}C]glucose. The media was prepared using custom formulations of DMEM and sodium pyruvate.

Glucose: DMEM stock was purchased without glucose (Gibco, New York, NY). Uniformly-^{13}C-enriched glucose, [U-^{13}C, 99%]glucose (Cambridge Isotope Laboratories, Tewksbury, MA), was supplied to the cells at a concentration of 1000 mg/L.

Pyruvate: Custom DMEM without Gln, Val, Gly and sodium pyruvate was purchased (Gibco, New York, NY). This solution was also used for other applications such as the isotopic enrichment of Gly and Val residues in elastin. Sodium [1-^{13}C, 99%]pyruvate powder (Cambridge Isotope Laboratories, Tewksbury, MA) was dissolved in autoclaved ultrapure water; its concentration in the growth media that was added to cells was 96 mg/L.

B.2.3 Incorporation via [1-^{13}C]Ala and [1-^{13}C]pyruvate. Custom DMEM without Gln, Gly, Val and sodium pyruvate was used. Unenriched Gln, Gly, and Val (Sigma-Aldrich, St. Louis, MO) were supplied in the NEAA solution. [1-^{13}C, 99%]Ala and sodium [1-^{13}C, 99%]pyruvate were supplied in the culture media at 30 mg/L and 96 mg/L, respectively, as described above.

B.2.4 Incorporation with enzymatic inhibition. A database search on BRENDA [17] identified two inhibitors, hydroxylamine (HAN) and aminooxyacetic acid (AOA), of alanine transaminase (ALT) in rat livers [18] (**Figure 2.4**).

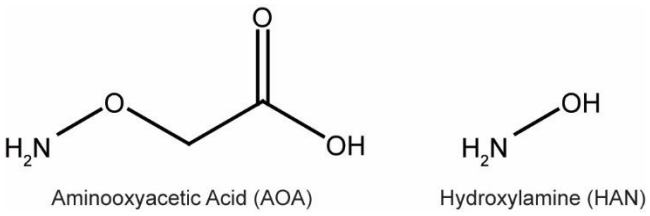


Figure 2.4 Structures of aminooxyacetic acid (AOA) and hydroxylamine (HAN), inhibitors of the ALT enzyme [18].

Inhibitor concentrations were first optimized, using three HAN and AOA (Sigma-Aldrich, St. Louis, MO) concentrations (0.1, 1.0, and 2.5 mM) in the culture media. During the 6-8 week incubation period, the viability of the cells was monitored by eye and by light microscopy for any signs of matrix detachment and excessive cell death. The optimal inhibitor condition was determined by the visual assessment of the condition of the NRSMC's at the end of the incubation time.

After the ideal concentration of the inhibitor was identified, another round of optimization was needed to determine the concentration of ^{13}C-Ala for high isotopic enrichment. The media solution was prepared with 0.5, 1.0, 2.0 and 3.0 mM ^{13}C-Ala (i.e., 5-fold, 10-fold, 20-fold and 30-fold of the initial condition of 0.1 mM). The cells were grown in a 6-well plate for 6-8 weeks. Cell matrices were collected, and elastin was isolated. The levels of ^{13}C enrichment were evaluated by LC-MS (next section).

NRSMC cultures with the optimal inhibitor concentration and isotope concentration were grown in two T-75 flasks to produce ~200 mg (wet weight) insoluble elastin for subsequent NMR measurements. Elastin samples were produced in ssNMR-scale quantities with two isotopes of alanines, [U-^{13}C, 99%]Ala and [U-^{13}C, 99%,^{15}N, 98%]Ala.

B.3 Assessment of isotopic enrichment level

B.3.1 Sample preparation. Purified elastin was hydrolyzed in 6 M hydrochloric acid (HCl) at 100 °C for 24 hours. Then, the hydrolysate was transferred into a 5 mL beaker and heated on a hot plate, until a white, dry residue remained. The dry hydrolysate was dissolved with 400 μL of aqueous sodium bicarbonate ($NaHCO_3$) solution (84 mg/mL).

To facilitate the LC-MS analysis, the elastin hydrolysate was derivatized with 1-fluoro-2-4-dinitrophenyl-5-L-leucinamide (FDLA) [19, 20], a variant of the commercially available Marfey's reagent [21]. The FDLA is synthesized from a mixture of 202 mg of dinitrofluorobenzene, 30 mL of acetonitrile, and 174 μL of N,N-diisopropylethylamine. A suspension of L-leucinamide hydrochloride (166.5 mg in 60 mL of acetonitrile) was added dropwise to the mixture, followed by constant stirring for 2 hours. The

solution was then concentrated to half its volume using a rotary evaporator, followed by the addition of 100 mL of deionized water. After another cycle of evaporation, the FDLA precipitate was collected and lyophilized for 24 hours, yielding ~215 mg of dry residue.

The derivatization of amino acids precedes the LC-MS analysis. The elastin hydrolysate was added into a 200 μL suspension of FDLA in acetone (10 mg/mL). The reaction was maintained at 85 °C, in an oil bath, for 5-10 minutes until a yellow-orange color was observed. Then, the mixture was cooled to room temperature, followed by three extractions with diethyl ether (2 mL each). The 250 μL aliquot was neutralized with 150 μL of 2 M HCL and diluted with 600 μL of acetonitrile. The solution was transferred through a 20-μm disposable filter into a sample vial for LC-MS analysis. **Figure 2.5** shows the derivatization reaction of an amino acid, i.e., Ala, with FDLA to produce the FDLA-derivative of alanine, i.e., Ala-FDLA.

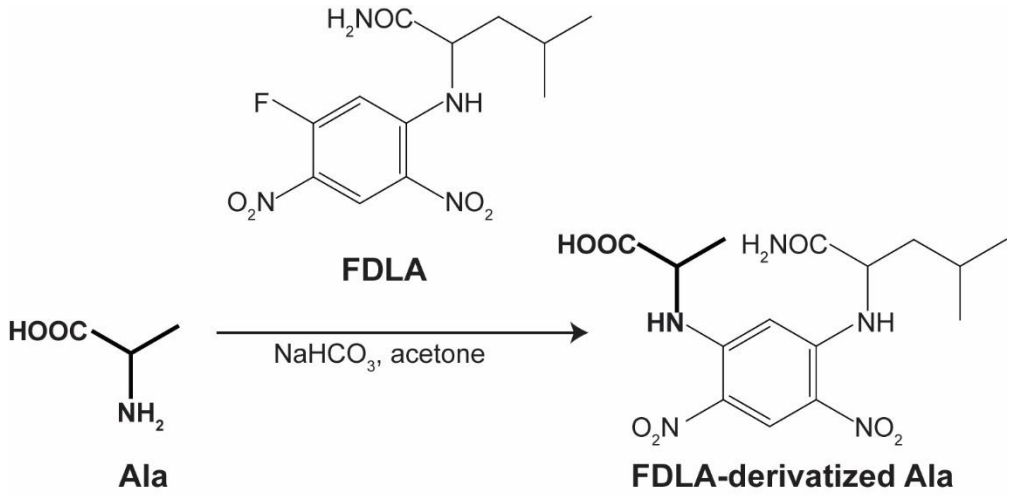


Figure 2.5 Derivatization reaction between Ala and FDLA. The reaction occurs at 85 °C in sodium bicarbonate and acetone. FDLA is functionalized by Ala (**bold**) to form FDLA-derivatized Ala.

B.3.2 LC-MS. The evaluation of isotopic enrichment in alanines and detection of isotopic scrambling were performed using a tandem configuration of high-performance liquid chromatography (HPLC) and mass spectrometry (MS) with electrospray ionization (ESI) in positive-ion mode.

The separation of FDLA-derivatized elastin hydrolysate was performed on a Luna 5 μm C-18 LC column (Phenomenex, Torrance, CA). The column was regulated at room temperature, 22 °C, and 1 μL of injection volume was used. The mobile phase consisted of a mixture of 0.1% formic acid (FA) in water (solvent A) and 0.1% FA in acetonitrile (solvent B). The flow rate was maintained at 0.7 mL/min

throughout the experiment. Amino acid derivatives were detected by UV at 340 nm. The gradient of solvent concentrations and elution times are listed in **Table 2.1**.

Triple-Quadrupole (QQQ) Detection: FDLA-derivatives were separated using Agilent 1200 series (Agilent Technologies, Santa Clara, CA) equipped with an autosampler, and the mass spectrometry was conducted using Agilent 6410 Triple Quad LC/MSD (Agilent Technologies, Santa Clara, CA). The ESI voltage was 4.0 kV with the gas nitrogen pressure set at 30 psi, respectively, and the capillary temperature was set to 325 °C. Data were collected in the profile mode, and a mass range of 105-1500 m/z was covered with a scan time of 500 ms.

Time-Of-Flight (TOF) Detection: Some experiments were run on an Agilent 1100 series HPLC configured with the Agilent 6210 LC/MSD-TOF (Agilent Technologies, Santa Clara, CA). The same ESI configuration as the QQQ (above) was used. The spectrometer is controlled by Agilent MassHunter Workstation Software, version B.02.00 (Agilent Technologies, Santa Clara, CA).

Table 2.1 LC elution mode for the separation of FDLA-derivatized elastin hydrolysate in the Q-TOF system. A linear gradient is used in each time increment.

Elution Time (min)	% Solvent B
0.0	30
35.0	30
38.0	50

B.3.3 Quantification of isotopic enrichment. In the positive ion mode of ESI, a protonated molecular ion $[M+H]^+$, or simply MH, is observed in the mass spectrum, where M is the molecular mass of the species of interest with the most abundant, naturally occurring isotope(s). The MH+1 and MH+2 peaks correspond to less abundant isotopes in the analyte. For instance, in **Figure 2.6a**, the peaks at 384.1541 m/z (MH) and 385.1575 m/z (MH+1) represent the ^{12}C-, ^{13}C-Ala-FDLA, respectively. The abundance of each species is reflected by the peak height. With 1-^{13}C enrichment of the analyte, the MH+1 peak is higher than the MH peak (**Figure 2.6b**).

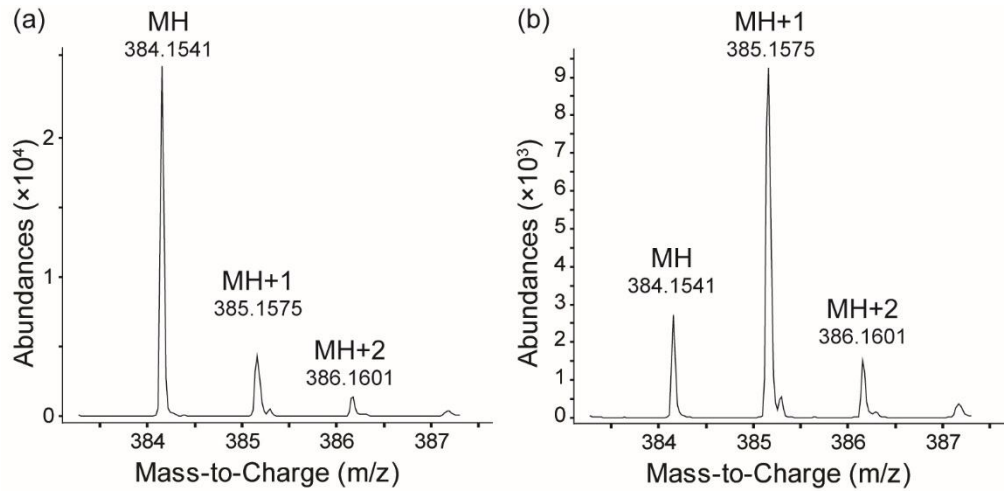


Figure 2.6 Mass spectra of (a) unenriched and (b) 1-^{13}C-enriched Ala-FDLA. Peak heights of ($M+H^+$) and ($M+H^+$)+1 indicate the relative abundance of unenriched and 1-^{13}C-enriched analyte, respectively. Spectra were acquired using the TOF method described in section B.3.3.

In the MS of unenriched elastin hydrolysates (**Figure 2.6a**), the ratio of isotopic abundance, R_c, is calculated, according to equation 2.1,

$$\boldsymbol{R_c = \frac{Ab[MH+1]}{Ab[MH]}} \qquad \text{(eqn. 2. 1)}$$

where $\mathrm{Ab[MH+1]}$ and $\mathrm{Ab[MH]}$ are the relative abundances for 1-^{13}C-Ala-FDLA and 1-^{12}C-Ala-FDLA, respectively.

In the MS of ^{13}C-enriched hydrolysates (**Figure 2.6b**), the peak height for the 1-^{13}C-Ala-FDLA at 385.1575 m/z is higher than that for the unenriched analyte, which reflects isotopic enrichment. The ratio of abundance in the 1-^{13}C-enriched analyte is given by R_e, according to equation 2.2,

$$\boldsymbol{R_e = \frac{Ab[MH+1]}{Ab[(MH)]}} \qquad \text{(eqn. 2. 2)}$$

where $\mathrm{Ab[MH+1]}$ and $\mathrm{Ab[MH]}$ correspond to the abundances of 1-^{13}C-Ala-FDLA and 1-^{12}C-Ala-FDLA species, respectively. The isotopic enrichment level of 1-^{13}C-Ala in the analyte was then quantified using Wolfe's method [22] in equation 2.3,

$$\%\,\boldsymbol{Enrichment = \frac{100\times(R_e-R_c)}{1+(R_e-R_c)}} \qquad \text{(eqn. 2. 3)}$$

B.3.4 Solid-state NMR experiments. Hydrated elastin samples were packed into 4.0 mm rotors; a typical sample size is 50-60 mg. The hydration level was maintained by sealing the sample with custom machined Kel-F spacers (Revolution NMR, Fort Collins, CO) fitted with fluorosilicone micro o-rings (Apple

Rubber Products, Lancaster, NY). SSNMR experiments were conducted on a Varian Unity Inova spectrometer (Varian, Palo Alto, CA) equipped with a wide-bore superconducting magnet with a ^{1}H resonance frequency of 399.976 MHz. ^{13}C spectra were acquired using a 4.0 mm triple resonance (HXY) T3 MAS probe (Chemagnetics/Varian, Fort Collins, CO). The MAS rate was 8 kHz. ^{13}C chemical shifts were referenced to the tetramethylsilane (TMS) scale, using hexamethylbenzene (HMB) as an external standard ($\delta(^{13}CH_3)$ = 17.0 ppm) at room temperature. ^{13}C spectra were acquired at 37 °C using the direct polarization method, with full details in Chapter 3. Briefly, a 4.0 μs ^{13}C 90° pulse was used with a 10-s recycle delay, and TPPM decoupling was applied during acquisition with a field strength ($\gamma.B_1^H/2\pi$) of ~60 kHz. Spectra were processed using MestReNova software, version 9 (Mestrelab Research, Escondido, CA), and 10 Hz of exponential line broadening was used in all spectra.

C. RESULTS AND DISCUSSIONS

Isotopically-enriched, biosynthetically crosslinked elastin is available from NRSMC cultures [12-14]. With direct substitution, this expression system is effective for the enrichment of amino acids like Gly [3], Val, and other residues in elastin. However, the approach is not optimal for the expression of isotopically-enriched alanines. Therefore, several strategies that may enhance the level of isotopic incorporation are investigated. The following sections describe the results of the study.

C.1 Direct substitution strategy showed minimal incorporation of [1-^{13}C]Ala in elastin

The direct substitution method is a simple and straightforward way to incorporate isotopes into elastin. This technique employs the replacement of unenriched Ala in NEAA solution with 1-^{13}C-enriched Ala.

Two concentrations of [1-^{13}C]Ala were prepared. First, the same concentration as the previous work was used (9 mg/L, 0.1 mM), i.e., the concentration in the commercially available NEAA solution. Second, a 25-fold increase of [1-^{13}C]Ala in the media (i.e., 225 mg/L, 2.5 mM) was used with the intention of increasing the utilization of exogenous Ala in elastin synthesis.

In the first scheme, 9 mg/L of [1-^{13}C]Ala was supplied in the culture media via NEAA solution. The LC-MS analysis showed ~4% isotopic enrichment of Ala in elastin. The low level of incorporation is due to the utilization of endogenously synthesized, unenriched Ala during protein synthesis. The intracellular Ala is the product of the glycolysis reaction of glucose, which is present at concentrations that are >100-fold higher. Hence, the amount of available ^{13}C-enriched Ala for elastin production is very small, compared to the unenriched amino acid.

In the second scheme, 225 mg/L of [1-^{13}C]Ala was provided in the media. The LC-MS analysis showed that ~48% of isotopic enrichment level was achieved by this method. The result indicates that a boost of (extracellular) [1-^{13}C]Ala concentration directly influences the level of ^{13}C enrichment in the protein. Despite the improvement on the level of isotopic enrichment, there are several drawbacks to utilization of a multifold increase of ^{13}C-enriched Ala in the media to produce samples at ssNMR-scale quantities. First, this method requires a large supply of isotope in the media to increase the availability of ^{13}C-enriched Ala for protein synthesis, making this strategy cost-ineffective and impractical. Second, the increase of extracellular isotopically-enriched alanines such as [U-^{13}C]Ala may lead to the metabolic scrambling effect. As described earlier, ALT not only facilitates pyruvate-to-Ala but also Ala-to-pyruvate conversion. The abundance of extracellular [U-^{13}C]Ala in the media can promote the intracellular production of [U-^{13}C]pyruvate, subsequently used to produce isotopically-enriched amino acids such as

^{13}C-Asp, ^{13}C-Glu, ^{13}C-Pro, etc. Third, the hypertonic condition caused by the excessive supply of isotopes in the media can be detrimental to the cells [23]. For these reasons, the direct substitution method was not chosen for scaling up the production of elastin for ssNMR measurements.

C.2 Use of enriched precursors to Ala resulted in isotopic scrambling

Intracellular biosynthesis of Ala in mammalian cells begins with glycolysis and ends with the transamination of pyruvate. Two sources of pyruvate are present in the typical NRSMC growth media, glucose (in the DMEM) and the sodium pyruvate solution. To produce intracellular ^{13}C-enriched Ala, [U-^{13}C]glucose (1000 mg/L) or [1-^{13}C]sodium pyruvate (96 mg/L) were used. After 8 weeks of culture, cell matrices were harvested, and insoluble elastin was purified (*Methods* section B.2).

The incorporation of ^{13}C-Ala into elastin differed with the precursors used in the media. The supply of [U-^{13}C]glucose and (sodium) [1-^{13}C]pyruvate resulted in, respectively, ~50% and ~6% isotopic enrichment at the alanines,. This result indicates that the strategy with [U-^{13}C]glucose was more effective at producing intracellular ^{13}C-enriched Ala than that with the sodium [1-^{13}C]pyruvate solution.

The advantage of [U-^{13}C]glucose supplementation is its cost-efficiency. Isotopically-enriched glucose is inexpensive, compared to [1-^{13}C]Ala that is used in excess in the direct substitution method (section C.1). However, this improvement is offset by the effect of isotopic scrambling to other amino acids in elastin. Namely, ^{13}C isotopes were distributed into Ser and Gly, as well as the byproducts of the TCA cycle (e.g., Gly, Pro, etc.), shown in **Appendix 3**. This scrambling effect hampers the ssNMR investigation, because the ^{13}C Pro signals overlap with those of Ala, complicating the interpretation of the spectra, among other considerations. Furthermore, this method is not suitable for ^{15}N enrichment, because no nitrogen sites are present in glucose and pyruvate molecules.

C.3 Combination of ^{13}C-pyruvate and ^{13}C-Ala in the media resulted in low isotopic enrichment

Methods that involve a single source of either intra- or extracellular ^{13}C-Ala yield ~4% to ~50% isotopic enrichment (section C.1-C.2). Briefly, the direct substitution of ^{13}C-Ala at 9 mg/L resulted in ~4% of enrichment level, and the supply of sodium [1-^{13}C]pyruvate in the media produced ~6% 1-^{13}C-enriched alanines in elastin. Although utilization of [U-^{13}C]glucose resulted in a fairly high level of U-^{13}C enrichment on alanines (~50%), the isotopic scrambling into glycines and prolines was substantial.

To enhance isotopic enrichment and minimize isotopic scrambling, the incorporation of isotopically-enriched alanines by both intra- and extracellular pathways is considered. Namely, this

strategy supplies extracellular [1-^{13}C]Ala via direct substitution (9 mg/L) and sodium [1-^{13}C]pyruvate (96 mg/L) in the culture media. [1-^{13}C]pyruvate was chosen instead of [U-^{13}C]glucose, so isotopes are not distributed into Gly and Ser. After 8 weeks of culture, cell matrices were harvested, and insoluble elastin was purified using the procedure described in the *Methods* (section B).

The LC-MS analysis indicated that the level of ^{13}C enrichment in the sample was ~10%. Although this strategy was aimed at incorporating isotopes via both intra- and extracellular pathways, the ^{13}C concentrations provided in the media (9 and 96 mg/L) were very small, compared to the amount of (unenriched) Ala, which is produced primarily from the breakdown of glucose at 1000 mg/L.

C.4 Inhibition of Ala biosynthesis facilitates the utilization of extracellular, isotopically-enriched Ala for elastin synthesis

With the previously described approaches, low isotopic enrichment of alanines in NRSMC elastin was due to the competitive effects of intracellular Ala synthesis. The approach with [1-^{13}C]Ala via direct substitution and sodium [1-^{13}C]pyruvate was ineffective, because the extracellular concentration of [1-^{13}C]Ala (9 mg/L) was significantly lower than that obtained intracellularly via glycolysis. The use of [U-^{13}C]glucose to obtain ^{13}C-enriched Ala endogenously resulted in isotopic scrambling. Optimally, the biosynthesis of (unenriched) alanine must be minimized, whereas the supply of ^{13}C-Ala via direct substitution must be maximized.

To minimize the production of intracellular Ala, the activity of ALT must be inhibited. Two compounds were identified on the BRENDA database [18] as potent ALT inhibitors, HAN and AOA. Previously, 2.5 mM solution of either HAN or AOA fully inhibited the activity of two ALT isoforms in the livers of starved rats after 30 minutes of incubation [24]. However, the activity of such compounds in the NRSMC line had not yet been reported. Furthermore, cell viability in the presence of ALT inhibitors for the entire duration of a 6-8 week incubation period was not previously investigated.

To assess cell viability and elastin synthesis under ALT inhibition, NRSMC cultures were grown in the presence of HAN and AOA, using three inhibitor concentrations (0.1, 1.0 and 2.5 mM). In 2.5 mM HAN or AOA, portions of the matrix layer detached from the surface of the flask after 2-3 weeks of culture. Additionally, less cell matrix was observed in 1.0 mM of HAN. Optimal conditions involve an ALT inhibitor at the highest possible concentration(s) that would still allow normal cell growth and matrix deposition. Therefore, 1.0 mM of AOA was chosen as the optimal ALT inhibitor concentration for NRSMC.

In addition to the ALT inhibitor, an excess of [1-^{13}C]Ala was used in the growth media to further increase isotopic enrichment levels. A surplus of [1-^{13}C]Ala in the media increased the level of ^{13}C

enrichment in elastin (section C.1), but, on its own, this approach is costly and inefficient. However, paired with the ALT inhibition, more of the supplied isotope is incorporated into the protein, as shown below.

LC-MS was used to quantify isotopic enrichment with the combination of ALT inhibition (1.0 mM AOA) and excess [1-^{13}C]Ala. The level of ^{13}C enrichment was examined as a function of ^{13}C-Ala concentrations in the media, i.e., 0.5, 1.0, 2.0 and 3.0 mM. As illustrated in **Figure 2.7**, the optimal level of excess [1-^{13}C]Ala was identified at 2.0 mM, which resulted in ~80% enrichment. The highest concentration of [1-^{13}C]Ala, 3.0 mM, did not result in a notable change of enrichment levels. Below these concentrations, a dependence between the [1-^{13}C]Ala concentration and the ^{13}C enrichment was observed, as the latter decreases to ~40% at 0.5 mM [1-^{13}C]Ala.

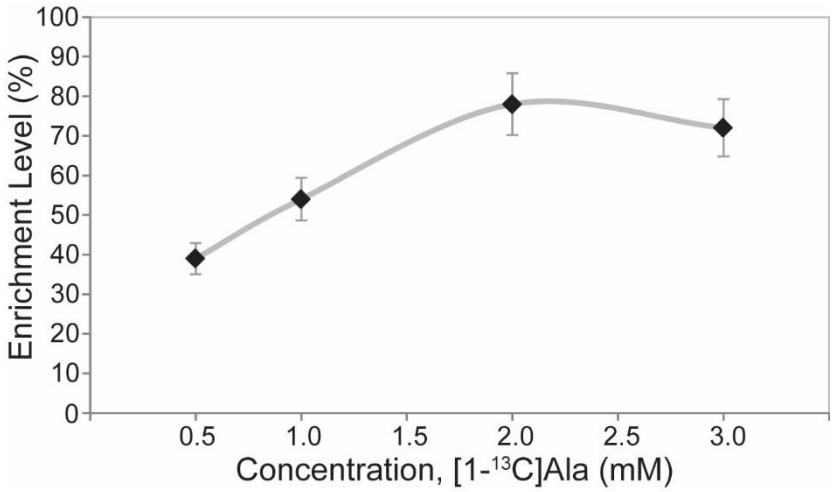


Figure 2.7 Isotopic enrichment levels in elastin hydrolysates as a function of ^{13}C-Ala concentration in the culture media, in the presence of 1.0 mM AOA. Each marker (♦) represents the average value from three NRSMC cultures grown concurrently. The error bars indicate the typical uncertainty of the LC-MS analysis (± 10%).

The combination of ALT inhibition and the excess of isotopically-enriched Ala (in media) has distinct advantages. First, isotopic scrambling is minimal, unlike the ^{13}C-glucose pathway (**Appendix 3**). Although traces of ^{13}C were detected in Asp and Glu (< 2%) and other byproducts of the TCA cycle, their contributions are insignificant for the study of alanines. Second, this strategy expands the possibility of isotopic incorporation to other sites of alanine, such as ^{15}N-Ala, [2-^{13}C]Ala, and [3-^{13}C]Ala, which are useful for a greater range of ssNMR applications. **Table 2.2**. summarizes the levels of ^{13}C enrichment using the four schemes of ^{13}C-Ala incorporation in NRSMC elastin.

Table 2.2. ^{13}C enrichment levels for 4 strategies for the incorporation of ^{13}C-Ala in elastin. Enrichment levels were determined by LC-MS. (AOA is **a**mino**o**xyacetic **a**cid)

No.	Isotopic Enrichment Method	Concentrations of Isotopes and AOA in Culture Media	^{13}C-Ala Source		Enrichment Levels
			Endo	Exo	
1	Direct substitution	(a) 9 mg/L [1-^{13}C]Ala		+	4% ±1%
		(b) 225 mg/L [1-^{13}C]Ala		+	48% ±5%
2	Metabolic precursors	(a) 96 mg/L [1-^{13}C]pyruvate	+		6% ±1%
		(b) 1000mg/L [U-^{13}C]glucose	+		50% ±5%
3	Direct substitution & metabolic precursors	96 mg/L [1-^{13}C]pyruvate & 9 mg/L [1-^{13}C]Ala	+	+	10% ±1%
4	Direct substitution & enzymatic inhibition	180 mg/L [1-^{13}C]Ala & 1.0 mM AOA	-	+	80% ±8%

The optimal strategy involves the intra- and extracellular supplies of ^{13}C-enriched alanines (**Table 2.2.**). Direct substitution of [1-^{13}C]Ala at 9 mg/L and the sodium [1-^{13}C]pyruvate supplementation at 96 mg/L (**No.** 1a, 2a) are the least successful strategies, with ~5% ^{13}C enrichment. The combination of direct substitution and metabolic precursors of [1-^{13}C]Ala (**No.** 3) was not effective, as only ~10% of isotopically-enriched alanines was quantified. A high level of [U-^{13}C]Ala, ~50%, was determined for elastin hydrolysates from the NRSMC cultures grown in [U-^{13}C]glucose (**No.** 2b); however, isotopes were also distributed into other amino acids, such as Gly and Pro (section C.2). In contrast, the combination of enzymatic (ALT) inhibition with the excess [1-^{13}C]Ala (**No.** 4) resulted in the highest level for isotopically-enriched alanines, ~80%, with minimal isotopic scrambling (section C.4).

C.5 1D solid-state NMR spectra strongly suggest conformational heterogeneity in Ala residues of elastin

Two methods are effective for the expression of NRSMC elastin with high levels of isotopic enrichment. Namely, strategies that incorporate the supply of [U-^{13}C]glucose and excess [1-^{13}C]Ala with ALT inhibition showed ~50% and ~80% enrichment levels, respectively. These techniques were chosen to produce insoluble elastin in ssNMR-scale quantities. In the scaled-up cultures, the combination of AOA and excess [U-^{13}C]Ala produced samples with isotopic enrichment at the carbon sites of alanines. The combination of AOA and [U-^{13}C,^{15}N]Ala was used for the isotopic enrichment at both ^{13}C and ^{15}N sites of Ala.

The high isotopic enrichment via simultaneous extra- and intracellular pathways is reflected by the ^{13}C intensities in the ssNMR measurements (**Figure 2.8a-c**). NRSMC elastin that was expressed in 1.0 mM of AOA incubation and 2 mM of [U-^{13}C]Ala supplementation showed the highest signal intensities (**Figure 2.8a**). The sample produced by [U-^{13}C]glucose supplementation shows prominent ^{13}C peaks for

Ala, Pro and Gly residues, due to scrambling (**Figure 2.8b**). Elastin that was expressed in the standard (unenriched) media formulation shows lowest relative intensities (**Figure 2.8c**). Therefore, the combination of ALT inhibitors with the excess ^{13}C-Ala is the optimal condition to produce elastin with distinctly high ^{13}C intensities for alanines in elastin. A detailed examination on the conformational and dynamical states of Ala is described in the next Chapter.

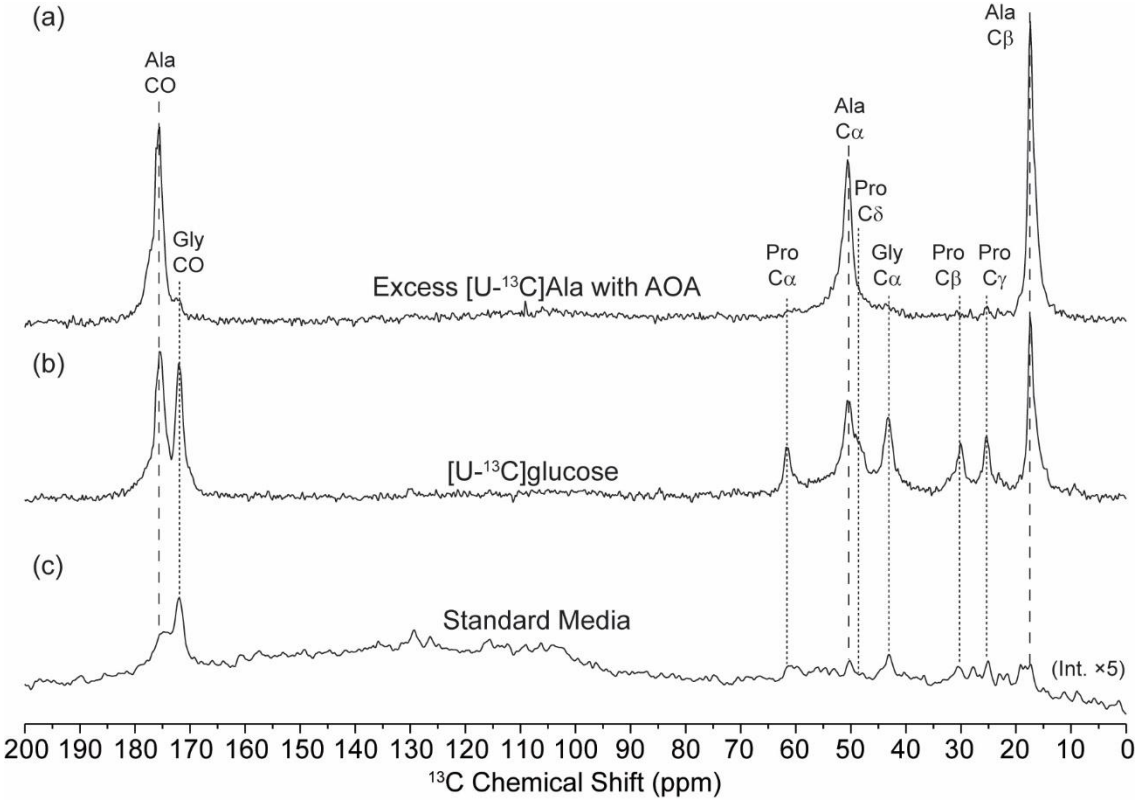


Figure 2.8 ^{13}C DPMAS spectra of hydrated elastin samples that were obtained from NRSMC cultures supplied with (a) excess [1-^{13}C]Ala (180 mg/L) and 1.0 mM AOA to inhibit ALT, (b) [U-^{13}C]glucose, and (c) standard (unenriched) culture media. Dotted lines (⋮) indicate signals from ^{13}C-Pro and ^{13}C-Gly, whereas dashed lines (⋮) indicate ^{13}C-Ala. Spectrum from the unenriched elastin reflects the natural-abundance ^{13}C signals from all amino acids. The number of acquisition scans for (a), (b) and (c) were 1024, 256 and 1024 respectively.

D. SUMMARY AND CONCLUSIONS

The NRSMC line is known to produce native elastin in milligram quantities [12-14]. The sample expressed by NRSMC cultures is insoluble and biosynthetically crosslinked by enzymes and molecular chaperones. Isotopic enrichment of NRMSC elastin is feasible by the simple replacement of Gly with isotopically-enriched Gly in the culture media (i.e., direct substitution). By this method, a fairly high level of isotopic enrichment of glycines in elastin has been obtained for subsequent high-resolution ssNMR studies [3].

In contrast to Gly, the enrichment of alanines has not been successful in NRSMC until this study. The major challenge in the incorporation of isotopically-enriched Ala by direct substitution is the endogenous synthesis of alanine. In the cell, ALT converts pyruvate into Ala. The presence of intracellular Ala hampers the incorporation of extracellular ^{13}C-Ala into elastin. Therefore, several isotopic enrichment strategies were explored.

The preliminary work was focused on addressing the supply of either intracellular or extracellular sources of Ala in the system. A strategy that exploits the supply of excess [1-^{13}C]Ala (25-fold) in the media was first investigated. This method resulted in the production of elastin with ~48% of ^{13}C enrichment level. However, this method was not cost-effective and unsuitable for the production of ssNMR-scale quantities of elastin.

Another strategy that is based on the endogenous production of ^{13}C-Ala with the supply of [U-^{13}C]glucose in the media was also explored. Although this method yielded a fairly high ^{13}C enrichment level (~50%), isotopic scrambling was identified. ^{13}C isotopes were distributed into Ala and other amino acids, such as Gly, Pro, Ser, etc. This resultant elastin hampers the characterization of alanines by ssNMR, because the ^{13}C signals for Ala overlap with those of Pro, leading to ambiguities with assignment.

The optimal isotopic enrichment scheme was found to involve the inhibition of ALT to minimize the intracellular concentration of unenriched Ala, along with the excess of ^{13}C-Ala in the media. To inhibit ALT, a 1.0 mM AOA solution was used. To endogenously produce isotopically-enriched alanines, 180 mg/L of [1-^{13}C]Ala was provided in the culture media. This combination resulted in a sample that has high isotopic enrichment (~80%) with minimal isotopic scrambling. Details on the characterization of Ala residues by ssNMR spectroscopy will be provided in the next Chapter.

E. REFERENCES

1. Perry, A., M.P. Stypa, B.K. Tenn, and K.K. Kumashiro, *Solid-State ^{13}C NMR reveals effects of temperature and hydration on elastin.* Biophysical Journal, 2002. **82**(2): p. 1086-1095.
2. Pometun, M.S., E.Y. Chekmenev, and R.J. Wittebort, *Quantitative observation of backbone disorder in native elastin.* Journal of Biological Chemistry, 2004. **279**(9): p. 7982-7987.
3. Martin, S.L., B. Vrhovski, and A.S. Weiss, *Total synthesis and expression in Escherichia coli of a gene encoding human tropoelastin.* Gene, 1995. **154**(2): p. 159-166.
4. Tamburro, A.M., B. Bochicchio, and A. Pepe, *Dissection of Human Tropoelastin: exon-by-exon chemical synthesis and related conformational studies.* Biochemistry, 2003. **42**(45): p. 13347-13362.
5. Tamburro, A.M., B. Bochicchio, and A. Pepe, *The dissection of human tropoelastin: from the molecular structure to the self-assembly to the elasticity mechanism.* Pathologie Biologie, 2005. **53**(7): p. 383-389.
6. Tamburro, A.M., A. Pepe, and B. Bochicchio, *Localizing α-helices in human tropoelastin: assembly of the elastin "puzzle".* Biochemistry, 2006. **45**(31): p. 9518-9530.
7. Mithieux, S.M., S.G. Wise, and A.S. Weiss, *Tropoelastin — A multifaceted naturally smart material.* Advanced Drug Delivery Reviews, 2013. **65**(4): p. 421-428.
8. Keeley, F.W., C.M. Bellingham, and K.A. Woodhouse, *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.* Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences, 2002. **357**(1418): p. 185-189.
9. Bellingham, C.M., M.A. Lillie, J.M. Gosline, G.M. Wright, B.C. Starcher, A.J. Bailey, K.A. Woodhouse, and F.W. Keeley, *Recombinant human elastin polypeptides self-assemble into biomaterials with elastin-like properties.* Biopolymers, 2003. **70**(4): p. 445-455.
10. Vieth, S., C.M. Bellingham, F.W. Keeley, S.M. Hodge, and D. Rousseau, *Microstructural and tensile properties of elastin-based polypeptides crosslinked with genipin and pyrroloquinoline quinone.* Biopolymers, 2007. **85**(3): p. 199-206.
11. Oakes, B.W., A.C. Batty, C.J. Handley, and L.B. Sandberg, *The synthesis of elastin, collagen, and glycosaminoglycans by high density primary cultures of neonatal rat aortic smooth muscle. An ultrastructural and biochemical study.* European Journal of Cell Biology, 1982. **27**(1): p. 34-46.
12. Jones, P.A., T. Scott-Burden, and W. Gevers, *Glycoprotein, elastin, and collagen secretion by rat smooth muscle cells.* Proceedings of the National Academy of Sciences of the United States of America, 1979. **76**(1): p. 353-357.

13. Barone, L.M., B. Faris, S.D. Chipman, P. Toselli, B.W. Oakes, and C. Franzblau, *Alteration of the extracellular matrix of smooth muscle cells by ascorbate treatment.* Biochimica et Biophysica Acta, 1985. **840**(2): p. 245-54.
14. Perry, A., M.P. Stypa, J.A. Foster, and K.K. Kumashiro, *Observation of the glycines in elastin using ^{13}C and ^{15}N solid-state NMR spectroscopy and isotopic labeling.* Journal of the American Chemical Society, 2002. **124**(24): p. 6832-6833.
15. Cetin, I., A.M. Marconi, A.M. Baggiani, M. Buscaglia, G. Pardi, P.V. Fennessey, and F.C. Battaglia, *In vivo placental transport of glycine and leucine in human pregnancies.* Pediatric Research, 1995. **37**(5): p. 571-575.
16. Paolini, C., A.M. Marconi, S. Ronzoni, M. Di Noio, P. Fennessey, G. Pardi, and F. Battaglia, *Placental transport of leucine, phenylalanine, glycine, and proline in intrauterine growth-restricted pregnancies.* Journal of Clinical Endocrinology and Metabolism, 2001. **86**(11): p. 5427-5432.
17. Daamen, W.F., T. Hafmans, J.H. Veerkamp, and T.H. van Kuppevelt, *Comparison of five procedures for the purification of insoluble elastin.* Biomaterials, 2001. **22**(14): p. 1997-2005.
18. Schomburg, I., A. Chang, O. Hofmann, C. Ebeling, F. Ehrentreich, and D. Schomburg, *BRENDA: a resource for enzyme data and metabolic information.* Trends in Biochemical Sciences, 2002. **27**(1): p. 54-56.
19. Fujii, K., Y. Ikai, H. Oka, M. Suzuki, and K.-i. Harada, *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.* Analytical Chemistry, 1997. **69**(24): p. 5146-5151.
20. Fujii, K., Y. Ikai, T. Mayumi, H. Oka, M. Suzuki, and K.-i. Harada, *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.* Analytical Chemistry, 1997. **69**(16): p. 3346-3352.
21. Marfey, P., *Determination of d-amino acids. II. Use of a bifunctional reagent, 1,5-difluoro-2,4-dinitrobenzene.* Carlsberg Research Communications, 1984. **49**(6): p. 591.
22. Wolfe, R.R., *Tracers in metabolic research: radioisotope and stable isotope/mass spectrometry methods.* Laboritory and Research Methods in Biology and Medicine, 1984. **9**: p. 1-287.
23. Wheatley, D.N., and B. Angus, *Inhibition of cell division in mammalian cell cultures by hypertonic medium.* Experientia, 1973. **29**(11): p. 1393-1395.
24. Vedavathi, M., K.S. Girish, and M.K. Kumar, *Isolation and characterization of cytosolic alanine aminotransferase isoforms from starved rat liver.* Molecular and Cellular Biochemistry, 2004. **267**(1): p. 13-23.

CHAPTER 3. INVESTIGATIONS OF STRUCTURE AND DYNAMICS OF ALANINE RESIDUES IN ELASTIN BY SOLID-STATE NMR SPECTROSCOPY AT PHYSIOLOGICAL TEMPERATURE

A. INTRODUCTION

The neonatal rat smooth muscle cell is a suitable expression system for the production of insoluble, hydrated, and biosynthetically crosslinked elastin for high-resolution solid-state NMR (ssNMR) studies, and it allows for isotopic labeling of select amino acids of interest. Alanines are suitable candidates for isotopic enrichment and ssNMR characterization, because they are abundantly found in both hydrophobic and crosslinking domains.

In this Chapter, ssNMR spectroscopy is used to characterize hydrated NRSMC elastin at 37 °C, which is isotopically-labeled at the alanines. The conformation- and sequence-dependent effects on the ^{13}C chemical shifts of alanines were investigated using one-dimensional (sections C.1 and C.2) and two-dimensional experiments (section C.2). NMR relaxation studies were also performed to support the conformational assignments of alanines in hydrophobic and crosslinking domains (section C.3). The results provide a framework for understanding the structure and dynamics of alanines in elastin's crosslinking domains, which provide motivation for further investigations with all-atom molecular dynamics (MD) simulations (Chapter 4), variable-temperature ssNMR experiments (Chapter 5), and temperature-replica exchange MD computation (Chapter 6).

A.1 Conformation- and sequence-dependence in ^{13}C chemical shifts of alanines in proteins and polypeptides

The isotropic ^{13}C chemical shifts of alanines in proteins and polypeptides are dependent on two factors, conformation and sequence. In general, the conformation-dependence is observed for all carbon sites of alanines. The respective ^{13}CO-, ^{13}Cα-, and ^{13}Cβ-Ala chemical shifts in a polyAla peptide adopting the α-helical structure (176.4, 52.4, and 14.9 ppm) are distinctly different from those with the β-sheet (171.8, 48.2, and 19.9) [1]. However, the chemical shift for a given carbon site may correspond to two or more common structural motifs, which makes structural assignments using a single ^{13}C resonance difficult. For example, as shown in **Table 3.1**, the ^{13}Cβ chemical shift for alanines in the polyAla peptide with the β-turn motif is observed at 17.4 ppm, identical to that for the polyproline II (PP-II). Therefore, ^{13}Cα- and ^{13}CO-Ala resonances for the respective conformations are used to resolve the ambiguity.

Table 3.1. Conformation-dependent ^{13}C chemical shifts of alanines in polypeptides adopting α_R-helix, β-sheet, random coil, β-turn type II, and 3_1-helix (PP-II). $(Ala)_n$* represents poly(L-alanine) peptide with molecular weights ranging from 1,000 to 200,000 Da. Hexafluoro-2-propanol (HFA) is a solvent that induces the formation of secondary structures. The chemical shift for random coil is obtained from the average of all ^{13}C-Ala resonances of 20 proteins, selected in the referenced study.

^{13}C	α_R-helix		β-sheet	random coil	β-turn (type II)	3_1-helix (or PP-II)
Sites	$(Ala)_n$*	Silk Fibroin in HFA#	$(Ala)_n$*	Statistical Average**	$(Ala\text{-}Gly)_{15}$ ##	$(Ala\text{-}Gly\text{-}Gly)_{10}$[$]
CO	176.4	not reported	171.8	175.2	177.5	174.6
Cα	52.4	51.5	48.2	50.6	51.5	48.9
Cβ	14.9	16.5	19.9	17.1	17.4	17.4

References: *[1], **[2, 3], #[4], ##[5], $[6]

The sequence-dependence is most evident in the ^{15}N and ^{13}CO chemical shifts of a given amino acid in proteins and polypeptides in random coil [7, 8]. ^{15}N and ^{13}CO resonances of a given residue **X** in the three-amino acid sequence motifs, R_{-1}-**X**-R_{+1}, are mainly influenced by the preceding (R_{-1}) and following (R_{+1}) amino acid variants, respectively (**Figure 3.1**).

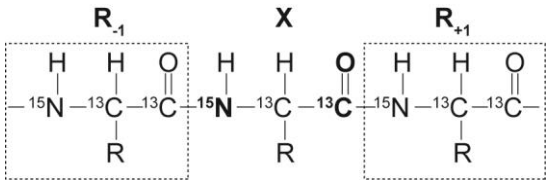


Figure 3.1 Chemical structure of a representative amino acid X and its neighboring residues in the three-amino acid sequence motif, R_{-1}-X-R_{+1}. ^{15}N and **^{13}C** resonances of residue X are influenced by variation in the preceding, R_{-1}, and following amino acids, R_{+1}.

This sequence-dependence, or *neighboring residue* effect, was recently demonstrated by ssNMR studies of synthetic ^{15}N-enriched elastin [9]. The ^{15}N chemical shifts for random coil glycines are influenced by the preceding (R_{-1}) residues. Namely, four distinct ^{15}N chemical shifts are observed, corresponding to Val-**Gly**, Pro/Leu-**Gly**, Gly-**Gly** and Ala-**Gly** sequences in the primary structure of tropoelastin.

Sequence-dependence in the ^{13}C chemical shifts of alanines in NRSMC elastin is effected primarily on the carbonyl site. The ^{13}CO resonance of **Ala** in the three-amino-acid sequence, R_{-1}-**Ala**-R_{+1}, is influenced by the amino acid that follows (R_{+1}). In the primary structure of rat tropoelastin, the neighboring residue (R_{+1}) of alanines are categorized into three motifs, R_{-1}-**Ala**-Gly, R_{-1}-**Ala**-Ala, and R_{-1}-**Ala**-Z, where Z is neither Ala nor Gly (**Table 3.2**). The R_{-1}-**Ala**-Ala sequences are primarily found in the crosslinking domains, and the R_{-1}-**Ala**-Gly motifs are prominent in the hydrophobic domains. In contrast, the R_{-1}-**Ala**-Z sequences are distributed across both hydrophobic and crosslinking domains.

Table 3.2. The sequence designs of alanines using the three-amino-acid motif, i.e., R_{-1}-Ala-R_{+1}. R_{-1}-**Ala**-Gly and R_{-1}-**Ala**-Ala sequence motifs are mainly found in the hydrophobic and the crosslinking domains, respectively; R_{-1}-Ala-Z motif (Z ≠ Ala, Gly) is present in both domains. The primary structure of rat tropoelastin is obtained from ELASTO-DB [10].

R_{-1}-Ala-Gly	R_{-1}-Ala-Z	R_{-1}-Ala-Ala
Gly-Ala-Gly	Ala-Ala-Lys	Ala-Ala-Ala
Ala-Ala-Gly	Gly-Ala-Val	Lys-Ala-Ala
Lys-Ala-Gly	Gly-Ala-Leu	Pro-Ala-Ala
Val-Ala-Gly	Gly-Ala-Arg	Gly-Ala-Ala
	Gly-Ala-Phe	Gln-Ala-Ala
	Gly-Ala-Ile	Val-Ala-Ala
	Gly-Ala-Pro	Leu-Ala-Ala
	Gly-Ala-Ser	Arg-Ala-Ala
	Tyr-Ala-Ser	Ser-Ala-Ala
	Ala-Ala-Ser	
	Ala-Ala-Arg	
	Ala-Ala-Val	
	Ala-Ala-Tyr	
	Lys-Ala-Pro	
	Lys-Ala-Lys	
Total 45 (27%)	Total 55 (33%)	Total 66 (40%)

In essence, isotropic ^{13}C chemical shifts facilitate the characterization of alanines in NRSMC elastin. The effects of secondary structures and *neighboring residues* on these chemical shifts are considered in the assignment of alanine populations.

A.2 One-dimensional (1D) ssNMR measurements provide ^{13}C chemical shifts of alanines

The molecular structure of tropoelastin's crosslinking domains is different from that of hydrophobic domains. The crosslinking domains are characterized by Ala-rich sequences with propensities for α-helices [11], whereas the hydrophobic domains are Gly-rich regions adopting several conformations, in equilibria, including PP-II, β-turn and random coil [11-14]. This structural organization traditionally describes the early stage of elastin assembly, in which self-aggregation of tropoelastin monomers occur [15]. However, this model has not been used to describe mature elastin, i.e., a hydrated, insoluble, and highly crosslinked polymer.

Solid-state NMR studies of mature elastin tend to be hampered by several challenges. First elastin is significantly more complex than that of elastin-like peptides (ELP), because the full-length

polymer is larger and more diverse. Hence, the corresponding spectrum include overlap of broad signals, which makes the interpretation more difficult. Second, common solid-state NMR techniques such as **C**ross-**P**olarization (CP) exploit the dipolar coupling interactions in solids with negligible motions. However, hydrated NRSMC elastin is highly mobile, and the attenuation of dipolar couplings in this system significantly reduces CP efficiencies. For these reasons, several ssNMR methods are explored for the characterization of hydrated NRMSC elastin at 37 °C.

A.2.1 Direct Polarization (DP) method. A single pulse excitation technique, or DP, is used in ssNMR characterization of alanines in hydrated NRSMC elastin. Traditionally, this experiment acquires ^{1}H spectra for organic molecules in solutions. Recently, ^{13}C DP was shown effective in solid-state NMR studies of hydrated elastin [16-19]. In the timing diagram used in ssNMR, the DP is achieved by a 90° pulse on the channel of interest, e.g., ^{13}C (or ^{15}N), followed by an acquisition period with high-power ^{1}H decoupling such as TPPM [20] (**Figure 3.2a**). The ^{1}H decoupling removes the heteronuclear (^{1}H-^{13}C) dipolar couplings, which typically broaden the ^{13}C lineshapes. This experiment is also performed under MAS, which removes the chemical shift anisotropy in ^{13}C spectra of solids (Chapter 1).

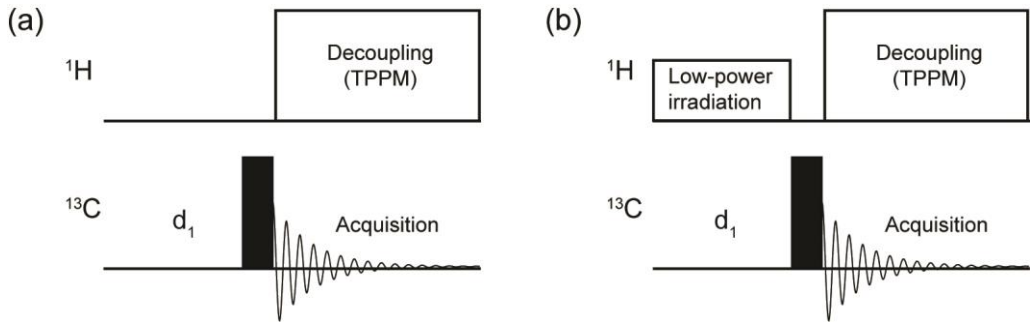


Figure 3.2 Pulse sequences for ^{13}C (a) DP and (b) DP with steady-state nOe. Black rectangles represent 90° pulses; d_1 represents recycle delay. **M**agic **A**ngle **S**pinning (MAS) is applied in the experiments. TPPM decoupling is used to remove dipolar heteronuclear (^{1}H-^{13}C) interactions. LOW-BASHD [21] and wSEDUCE-1 [22] decoupling may be applied during acquisition to detect ^{13}CO and $^{13}C\beta$ spectra without $^{1}J_{COC\alpha}$ and $^{1}J_{C\alpha C\beta}$ couplings, respectively.

The use of DP for solid samples has both advantages and disadvantages. The advantages of DP are in the simplicity of the experiment and in the quantitative aspect of the acquired signal. A typical 1D ^{13}C DP spectrum reflects the total carbon populations in the sample, regardless of structure and relative mobility. Hence, the DP spectrum includes signals that originate from the mobile environment in the system, typically not accessible by CP due to the attenuation of dipolar couplings. However, the acquisition of ^{13}C DP spectra with high signal-to-noise (S/N) ratio requires long experimental times. The total time is influenced by the length of the recycle delay, d_1 (**Figure 3.2a**). This period allows for all carbon magnetizations to return to equilibrium after each scan. This equilibration period typically occurs

within 5 times of the spin-lattice relaxation time constants (T_1's), which are typically long for ^{13}C nuclei in solids [23, 24].

The implementation of steady-state nuclear Overhauser effect (nOe) in a ^{13}C DP experiment, i.e., DP-nOe, enhances ^{13}C signal intensities by a factor of η,

$$\mathbf{1} + \boldsymbol{\eta} = \frac{\boldsymbol{Intensity\ (with\ ^1H\ irradiation)}}{\boldsymbol{Intensity\ (without\ irradiation)}} \qquad \text{(eqn. 3. 1)}$$

This nOe enhancement is accomplished by the saturation of ^1H spin populations using low-power RF irradiation, prior to the 90° pulse on the ^{13}C channel (**Figure 3.2b**). This procedure allows for the cross-relaxation mechanism to occur, as the population of high-energy ^1H spins return to equilibrium and the ^{13}C spin populations are increased. The nOe enhancement is biased by molecular motion, as its effectivity mostly affects ^{13}C populations with high mobility. The DP-nOe excitation is implemented in the two-dimensional ssNMR experiments (section A.3).

A.2.2 Detection of the rigid region by CP. CP is a standard ssNMR experiment to transfer the magnetization of abundant, high-γ nuclear spins like ^1H to rare, low-γ spins like ^{13}C or ^{15}N. In a standard ^{13}C CP experiment, the initial magnetization is acquired by a 90° pulse on the ^1H channel (**Figure 3.3**). Then, the magnetizations are spin-locked in the rotating frame using a continuous RF field on ^1H channel, $\mathbf{B_1^H}$. Concurrently, another RF field on the ^{13}C channel, $\mathbf{B_1^C}$, is applied such that the Hartmann-Hahn matching condition ($\gamma^H.B_1^H = \gamma^C.B_1^C$) is met, in which B_1^H and B_1^C correspond to the RF field strengths of $\mathbf{B_1^H}$ and $\mathbf{B_1^C}$ channels, respectively [25]. Then, the energy gaps between ^1H and ^{13}C spin states in the respective rotating frames are equal at this condition, allowing for cross-polarization. I.e., the dipolar-coupling-mediated transfer of magnetizations from the ^1H to the ^{13}C spins.

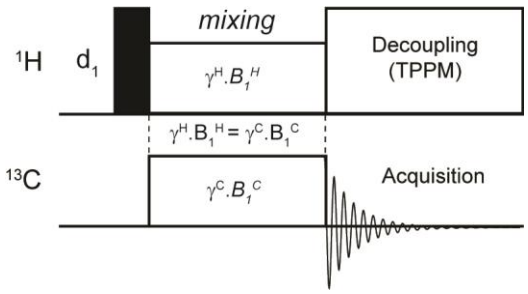


Figure 3.3 Pulse sequence for ^{13}C CP experiment. MAS is applied during the experiment. Black rectangle represents 90° pulse; d_1 represents recycle delay. Cross-polarization occurs during *mixing* under Hartmann Hahn matching condition ($\gamma^C.B_1^C=\gamma^H.B_1^H$).

For typical biological solids, the acquisition of ^{13}C spectra using CP with MAS, i.e., CPMAS, is more effective and more efficient than that using DPMAS. First, the ^{13}C CPMAS intensity is typically enhanced by a factor of 4, relative to the ^{13}C DPMAS, due to the differences in the gyromagnetic ratio (~

γ_H/γ_C). Second, the d_1 period set for the cross-polarization experiment is shorter than that for ^{13}C direct polarization, because the T_1 values of ^1H nuclei are generally shorter than those of ^{13}C. This NMR property allows for more scans to be acquired with the ^{13}C CP than the ^{13}C DP, resulting in improved signal-to-noise (S/N) ratio in the corresponding spectrum at a given experimental time.

The CP efficiency for hydrated NRSMC elastin is low, because the dipolar couplings are significantly attenuated in this mobile system. Hence, this method is used to select for regions in hydrated NRSMC elastin where the dipolar couplings are not completely removed (section C.3 and Chapter 5).

A.2.3 Detection of the mobile region by rINEPT. A method that has been useful for the detection of signals from hydrated elastin is **r**efocused **I**nsensitive **N**uclei **E**nhanced by **P**olarization **T**ransfer (rINEPT) [9]. The rINEPT was originally developed to enhance the observed signals of nuclei with low γ (^{13}C or ^{15}N) by a polarization transfer from the nuclei with high γ (^1H) for organic molecules in solutions [26, 27]. In general, ^{13}C rINEPT signals are enhanced by a factor of 4 (~ γ_H/γ_C) relative to those obtained by DP.

In rINEPT, the transfer of magnetizations occurs through *J*-coupling interactions instead of dipolar couplings. **Figure 3.4** illustrates the pulse sequence for a 1D {^1H}-^{13}C rINEPT experiment. The initial magnetization is prepared by a 90° pulse on the ^1H channel, followed by a spin-echo sequence, –τ–180°–τ– (*spin-echo-1*). The length of τ is typically in the order of $1/|4*J_{CH}|$, which is ~2 ms for $^1J_{CH}$ of ~125 Hz. During the first τ-delay, the precession of ^1H spins at different rates result in the loss of transverse magnetizations, or dephasing. However, the 180° pulse inverts the magnetization vectors, and the ^1H coherence is reestablished in the second τ-delay. During the spin-echo, transverse magnetizations in the rotating frame decay as a function of $\exp(-2\tau'/T_2'^H)$, in which $T_2'^H$ corresponds to the time constants for proton coherences during spin-echo (–τ–180°–τ–). The mobile regions in the sample, which have large ^1H oscillations, typically have longer $T_2'^H$ values (e.g., ~10 ms) than those in the rigid domains (e.g., 1 ms). Hence, only ^1H coherences with long lifetimes, or the mobile components, survive after the 2τ delays (4 ms).

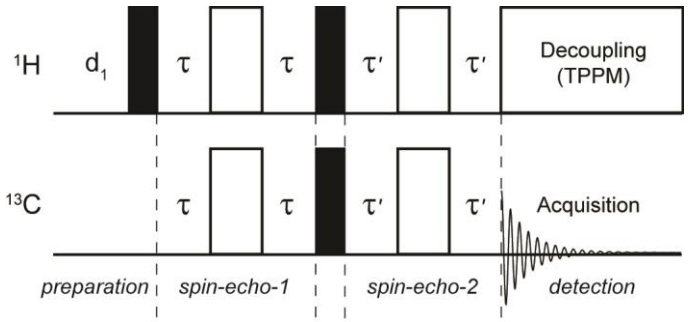


Figure 3.4 Pulse sequence for {^{1}H}-^{13}C rINEPT experiment. Black and white rectangles indicate 90° and 180° pulses, respectively. d_1 indicates recycle delay; τ and τ′ are spin-echo delays. MAS is applied during the experiment.

The remaining ^{1}H magnetization is transferred to ^{13}C by simultaneous 90° pulses on both ^{1}H and ^{13}C channels. During the spin-echoes that follows (*spin-echo-2*), ^{13}C magnetizations decay as a function of exp(-2τ′/T_2′^{C}), in which T_2′^{C} corresponds to the time constants for carbon magnetizations during –τ′–180°–τ′–. After 2τ′ delays (i.e., 4 ms for CH group), ^{13}C coherences with long lifetimes (i.e., long T_2′^{C}) survive, else diminish. Therefore, only the mobile components in the protein are detected during *acquisition*.

A.2.4 ^{1}H→^{15}N→^{13}CO relays of coherence transfers. An inter-residue coherence transfer is achievable by consecutive rINEPT's on ^{1}H-^{15}N and ^{15}N-^{13}C spin pair systems. Specifically, the detection of ^{13}CO signals from a given residue, *i*, originating from the ^{1}H magnetization of the adjacent residue, *i+1,* is achievable via a two-step transfer:^{1}H→^{15}N (**I**) and ^{15}N→^{13}CO (**II**) (**Figure 3.5a**). This experiment acquires ^{13}CO signals from **Ala** in the R$_{-1}$-**Ala**-Ala motif, which is found mostly in the crosslinking domains of hydrated NRSMC elastin (previously shown in **Table 3.2**). Additionally, this rINEPT-based experiment primarily detects the mobile component in the corresponding region. The inter-residue coherence transfer is facilitated by a sample that is isotopically-labeled at both ^{15}N and ^{13}CO sites of alanines (Chapter 2).

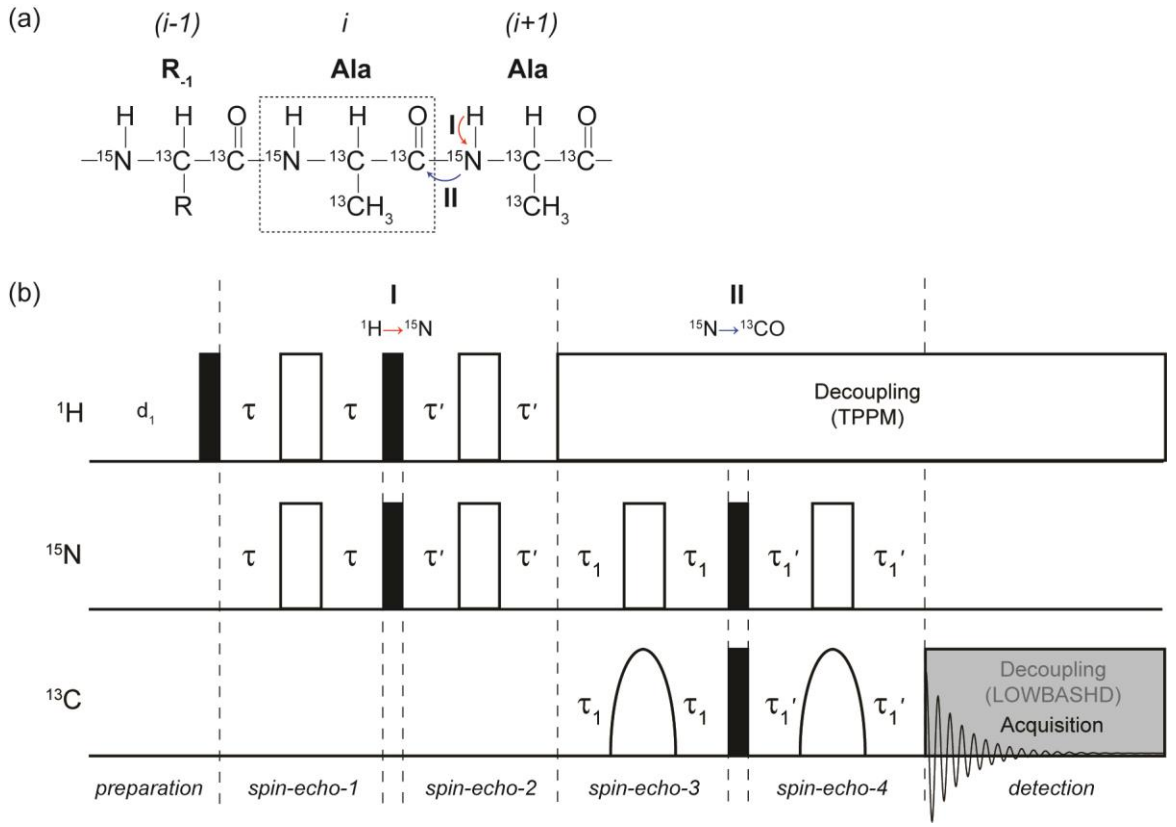


Figure 3.5 (a) A two-step rINEPT to select for paired alanines with the R_{-1}-Ala-Ala motif, and (b) the corresponding pulse sequence. MAS is applied during the experiment. Two-step rINEPT includes $^{1}H \rightarrow ^{15}N$ and $^{15}N \rightarrow ^{13}CO$ transfers. Black and white rectangles indicate 90° and 180° pulses, respectively. White parabolas represent selective 180° pulses on ^{13}CO. d_1 indicates recycle delay; τ, τ', τ_1 and τ_1' are spin-echo delays. Gray shading indicates ^{13}CO detection with $^{1}J_{CaCO}$ decoupling using LOW-BASHD [21].

The timing diagram for $^{1}H \rightarrow ^{15}N \rightarrow ^{13}CO$ transfer is essentially composed of two rINEPT sequences (**Figure 3.5b**). The initial magnetization is prepared by a 90° pulse on the ^{1}H channel, followed by a relay of coherence from the amide proton to the nitrogen, i.e., the first rINEPT step ($^{1}H \rightarrow ^{15}N$ (**I**)). As described earlier, ^{1}H magnetizations decay as a function of $\exp(-2\tau/T_2'^{H})$ in the first period (*spin-echo-1*), whereas ^{15}N coherences decay as a function of $\exp(-2\tau'/T_2'^{N})$ in the latter period (*spin-echo-2*). Coherences with long T_2' survive through these echoes, and the remaining ^{15}N magnetizations are transferred to the carbonyl nuclei by the second rINEPT sequence.

A coherence transfer from the amide nitrogen to the carbonyl carbon, i.e., $^{15}N \rightarrow ^{13}CO$, is achieved by the second rINEPT block (**II**), in **Figure 3.5b**. This transfer is accomplished by means of simultaneous 90° pulses that are 'sandwiched' by spin-echo sequences, on both ^{15}N and ^{13}C channels. However, selective 180° pulses are used in the ^{13}C channel to prevent the $^{15}N \rightarrow ^{13}C\alpha$ coherence transfer. Similar to the first rINEPT block (**I**), the two echoes (*spin-echo-3* and *spin-echo-4*) function as mobility-based filters.

Namely, coherences with short T_2' values remain at the ends of the $2\tau_1$ and $2\tau_1'$ periods, whereas others diminish. This rINEPT-based measurement is effective for samples with high mobility such as elastin, because they typically have long T_2'.

The *acquisition* period involves the detection of ^{13}CO signals using **L**ong-**O**bservation-**W**indow **B**and-**S**elective **H**omonuclear **D**ecoupling (LOW-BASHD) technique [21], to remove ^{13}CO-$^{13}C\alpha$ *J*-couplings that broaden the carbonyl lineshapes. ^{1}H (TPPM) decoupling is also applied during ($^{15}N \rightarrow ^{13}CO$) rINEPT and *acquisition* periods.

A.2.5 NMR relaxation measurements and the timescales of molecular motions. Insights into the molecular dynamics of proteins and polymers are obtained from the study of nuclear spin relaxations. NMR relaxation parameters probe molecular motions in the milli- to nanosecond regimes. Fast motions in the nano- to picosecond timescale are typically investigated by longitudinal, T_1, and transverse, T_2, relaxations. Motions in the micro- to millisecond regimes are assessed with spin-lattice relaxations in the rotating frame, $T_1\rho$.

^{13}C T_1 and $T_1\rho$ measurements are effective for the characterization of elastin's mobility in the crosslinking and hydrophobic regions [17, 19, 28], which are complimentary to the DP, CP and rINEPT (sections A.2.1-A.2.3). The pulse sequence for ^{13}C T_1 inversion recovery (**Figure 3.6a**) is described by a 180° ^{13}C pulse, followed by an array of delay periods, τ. A 90° ^{13}C pulse brings the magnetization to the transverse plane, which is observed during *acquisition* with TPPM decoupling [20]. The result of this experiment is an array of 1D ^{13}C spectra for the different τ delays. The relaxation rate constant is then obtained from an exponential fitting of the peak intensities as a function of τ. The T_1 values (usually in s) reflect molecular motions in the MHz timescale, often used to assess segmental motions in protein and polymers.

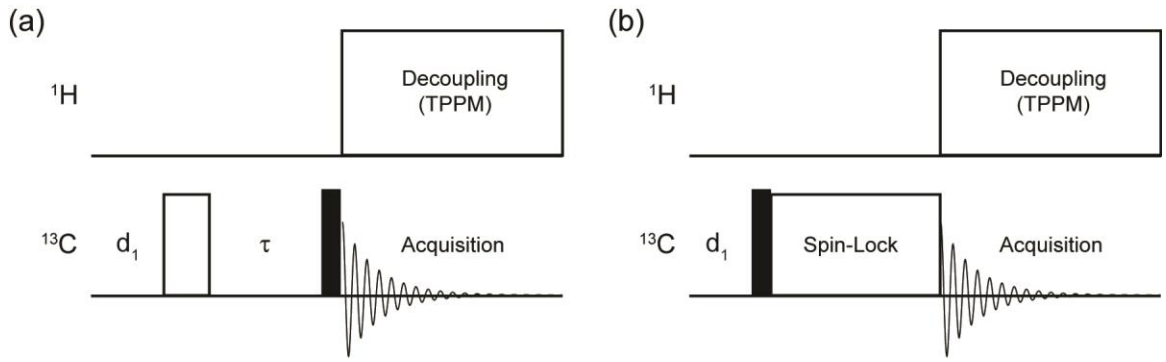


Figure 3.6 Pulse sequence to measure ^{13}C (a) T_1 and (b) $T_1\rho$ relaxations using DP excitation. MAS is applied during the experiment. Black and white rectangles indicate 90° and 180° pulses respectively; d_1 indicates recycle delay. Delay (τ) and Spin-Lock periods are arrayed in the respective experiments.

The DP-based ^{13}C $T_1\rho$ measurements are performed using a single 90° ^{13}C pulse excitation, followed by an array of spin-locking times (**Figure 3.6b**). In the rotating frame, magnetizations are spin-locked in the transverse plane with a given $\mathbf{B_1}^C$ field. ^{13}C signals are detected during *acquisition* with TPPM decoupling[20]. The result of the ^{13}C $T_1\rho$ experiment is an array of 1D ^{13}C spectra for these spin-locking times. The relaxation rate constant ($T_1\rho$ value) is obtained from the exponential fitting of peak intensities as a function of spin-locking times. The T_1 values (usually in ms) reflect molecular motions in the kHz timescale, relevant for biopolymer studies.

A.3 Two-dimensional (2D) ssNMR experiments yield homonuclear ^{13}C-^{13}C chemical shift correlations

Two-dimensional ssNMR experiments are effective in resolving overlapping signals that typically occur in one-dimensional spectra. A generalized 2D experiment has four periods of pulses and/or time delays, *preparation*, *evolution (t_1)*, *mixing*, and *acquisition (t_2)* (**Figure 3.7**). During *preparation*, excitation methods such as DP, DP-nOe, rINEPT and CP are used to target select populations. DP quantitatively obtains all NMR-active populations of a specific nucleus of interest, regardless of structure and dynamics. DP-nOe and rINEPT select the mobile regions in elastin, whereas CP focuses on the rigid components (section C.3.1) and for measurements below the freezing point of water (Chapter 5). During *evolution*, a train of RF pulses and/or time delays is applied to allow for NMR interactions to evolve and the spin systems to be frequency-labeled, such as with the first chemical shift offset (Ω_1), e.g., in the case of homonuclear correlation spectroscopy. Another set of RF pulses and/or delays is applied during *mixing* to facilitate the transfer of magnetization from one spin system to another. This transfer is typically accomplished by either a through-bond (*J*-couplings) or a through-space (dipolar couplings) interaction.

During *acquisition*, transverse magnetizations are detected as FID's, and the spin systems are generally frequency-labeled with the second chemical shift offset (Ω_2) and homonuclear *J*-coupling.

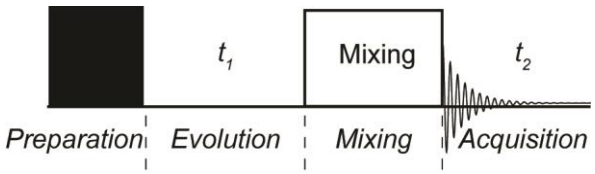


Figure 3.7 Generalized pulse sequence for acquiring two-dimensional NMR spectrum. One of the excitation pulses (DP, DP-nOe, rINEPT, and CP) is used in the *preparation* period. A series of pulses and/or delays are implemented in *evolution* and *mixing* periods. Signals are detected during *acquisition*. In (b), *left*, incremented t_1 points produce an array of FID's in t_2 domain. *Middle*, first Fourier transformation (FT) converts each FID from the time domain, t_2, into the frequency domain, f_2. *Right*, second FT converts the time domain, t_1, into, the frequency domain, f_1. The result is typically shown as a contour plot.

A 2D experiment is essentially an array of 1D experiments for which FID's are obtained during t_2. An FID is acquired for each increment of t_1 points. Two Fourier transformations (FT) generate the 2D spectrum in frequency domains. First, the FT converts signals from the time domain, t_2, into the frequency domain, ω_2. Subsequently, another FT yields the frequency domain, ω_1. The result is a 2D contour plot that describes the two frequency domains, ω_1 and ω_2, often called the indirect (f_1) and direct (f_2) dimensions.

A.3.1 CTUC-COSY Experiment. One of the earliest developed multi-dimensional NMR experiments, **CO**rrelation **S**pectroscop**Y** (COSY) [29], yields 2D spectra with homonuclear (^{1}H-^{1}H) correlations via the through-bond interactions (*J*-couplings). The **D**ouble-**Q**uantum-**F**iltered **COSY** (DQF-COSY) was later developed to acquire only in-phase lineshapes in both diagonal- and cross-peaks, which makes spectral processing and interpretation easier. The DQF-COSY experiment differs from the COSY only by the addition of another 90° pulse, which forms the double quantum filter (DQF) (**Figure 3.8**). The initial transverse magnetization is obtained using a direct polarization, 90° pulse, on the ^{1}H channel. During t_1, transverse magnetization evolves under chemical shift offset (Ω_1) and *J*-coupling. Then, the double-quantum filter selects the double-quantum coherences and converts them into observable single-quantum coherences, which are detected during *acquisition*.

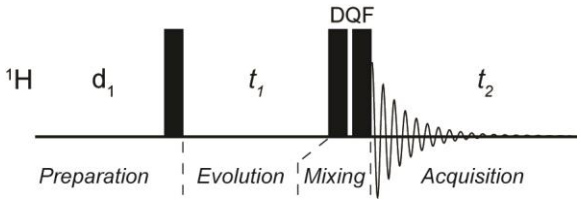


Figure 3.8 Pulse sequence for the DQF-COSY experiment. Black rectangle indicates 90° pulse; d_1 represents recycle delay. DQF represents a double-quantum filter; the two respective 90° pulses have x and y phases. Zero-, single- and double-quantum coherences are present after the first 90° pulse. However, only the double-quantum coherence is converted into observable magnetization after the second 90° pulse.

The 2D COSY and DQF-COSY experiments are not suitable for the acquisition of ^{13}C-^{13}C correlation spectra for solids. These COSY-based spectra typically contain absorptive and dispersive lineshapes in the diagonal- and cross-peaks, due to the in-phase and anti-phase terms arising from the *J*-coupling-mediated transfer. In ^{13}C spectra of solid samples, the ^{13}C lineshapes are relatively broad, compared to $^1J_{C\alpha C\beta}$ or $^1J_{C\alpha CO}$ couplings. Hence, the presence of dispersive lineshapes (anti-phase terms) in the COSY spectrum leads to the cancellation of cross-peak intensities [30], which makes this experiment ineffective for solid samples including elastin.

Improved cross-peak resolution has been achieved by recent developments in COSY-based experiments (**Figure 3.9a**). A modified version of the DQF-COSY, i.e., **U**niform-Sign **C**ross-Peak **D**ouble-**Q**uantum-**F**iltered **COSY** (UCDQF-COSY) was shown to produce in-phase lineshapes in both diagonal- and cross-peaks in the 2D spectra of disordered solids [31]. More recently, improved spectral resolution is achieved by the **C**onstant-**T**ime **U**niform-Sign **C**ross-Peak **COSY** (CTUC-COSY) experiment [32], which eliminates $^1J_{CC}$ in the indirect dimension (f_1). The combination of CTUC-COSY and LOW-BASHD is effective for the acquisition of 2D spectra without the $^1J_{CO\text{-}C\alpha}$ couplings, in both f_1 and f_2 dimensions, respectively [31].

Figure 3.9b illustrates the pulse sequence for the 2D CTUC-COSY experiment to acquire ^{13}CO-$^{13}C\alpha$ correlations. The *preparation* period in this 2D experiment includes a ^{13}C DP excitation to acquire the initial magnetizations, followed by a constant-time *evolution* period. This period is kept constant (for $2*\tau$) by the implementation of τ-$t_1/2$ and τ+$t_1/2$ delays, respectively, prior and subsequent to the 180° pulse. During constant-time evolution, (τ-$t_1/2$)–180°–(τ+$t_1/2$), only the chemical shift offset (Ω_1) is refocused. The $^1J_{C\alpha CO}$ is not refocused, because non-selective 180° pulses are set on both ^{13}CO and $^{13}C\alpha$. For the remaining time, i.e., $2*\tau$ - $2*(\tau$-$t_1/2)$ = t_1, the chemical shift offset is effective, and the magnetization evolves under the chemical shift offset (Ω_1); *J*-coupling remains effective for $2*\tau$. As t_1 is increased (by changing the position of the 180° pulse), the resulting signal intensity is modulated only by the chemical shift offset, not *J*-coupling. Therefore, the resulting spectrum shows no $^1J_{COC\alpha}$ couplings in the f_1 dimension.

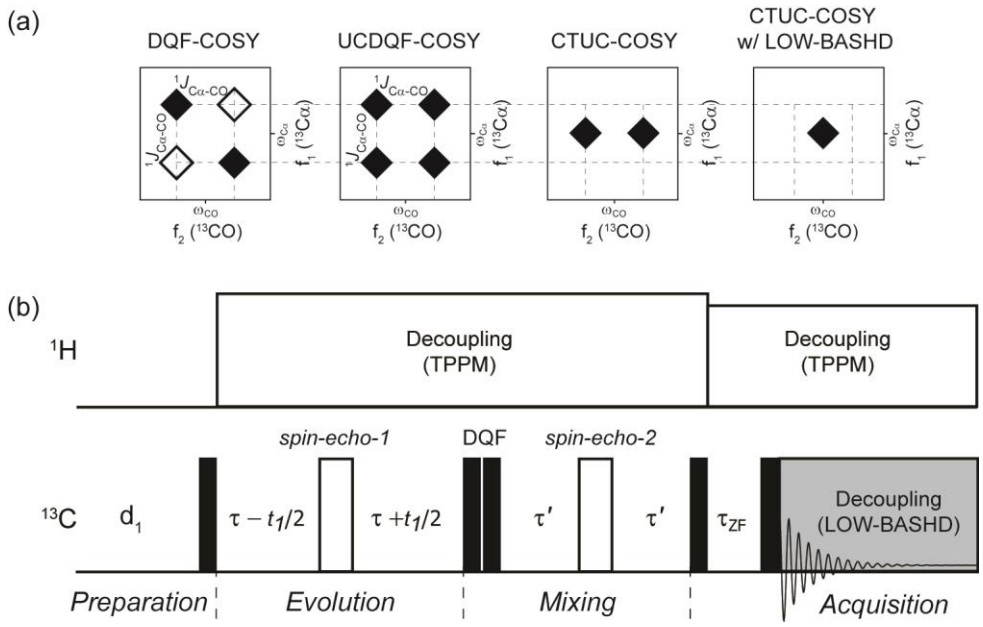


Figure 3.9 (a) Lineshapes in 2D cross-peaks of COSY-based experiments; (b) pulse sequence for 2D CTUC-COSY experiment with LOW-BASHD. MAS is applied during the experiment. In (a), anti-phase and in-phase cross-peaks are represented by ( ) and ( ), respectively; scalar couplings ($^1J_{C\alpha CO}$) are represented by vertical lines in both f_1 and f_2 dimensions. DQF-COSY yields anti-phase cross-peaks and retains *J*-couplings. UCDQF-COSY yields in-phase cross-peaks, and retains *J*-couplings. CTUC-COSY yields in-phase cross-peaks with no *J*-couplings in f_1 dimension. CTUC-COSY with LOW-BASHD yields in-phase cross-peaks without *J*-couplings, neither in f_1 nor f_2 dimension. In (b), black and white rectangles indicate 90° and 180° pulses, respectively. The d_1 represents recycle delay; τ and τ' are delay periods in *spin-echo-1* and *spin-echo-2,* respectively. Constant-time evolution is represented by τ-t_1/2 and τ+t_1/2 delays. Gray shading represents ^{13}CO detection with LOW-BASHD [21].

The *mixing* period consists of a double-quantum filter and a spin-echo. The double-quantum filter (DQF) is composed of a pair of 90° pulses. The *spin-echo-2* period refocuses chemical shifts, and it converts the anti-phase term to in-phase under the *J*-coupling evolution. The z-filter is composed of two 90° pulses that are interspersed by a short delay (τ_{ZF}), and it removes unwanted anti-phase terms and to select for the in-phase terms that give rise to pure absorptive lineshapes in the cross-peaks [33]. Finally, the ^{13}CO signals are detected during *acquisition* using the LOW-BASHD [21], which removes $^1J_{COC\alpha}$ couplings in the f_2 dimension. ^{1}H TPPM decoupling [20] is also applied to remove heteronuclear dipolar couplings.

Cross-peak intensities in CTUC-COSY experiment are limited by the coherence lifetimes of the carbon nuclei ($T_2'^C$). ^{13}C populations with short T_2' decay significantly after *mixing*, resulting in the loss of

intensities. In contrast, nuclei with mobile regions have long T_2' and are detected during *acquisition*. For this reason, this experiment is most effective for the characterization of hydrated NRSMC elastin at 37 °C.

A.3.2 R-TOBSY Experiment. TOtal **C**orrelation **S**pectroscop**Y** (TOCSY) is useful for the assignment of proton chemical shifts of organic molecules in solution. Similar to COSY, TOCSY acquires ^{1}H-^{1}H correlations via through-bond interactions (*J*-couplings). The pulse sequence for TOCSY is different from that for COSY by the addition of an isotropic mixing period (**Figure 3.10**).

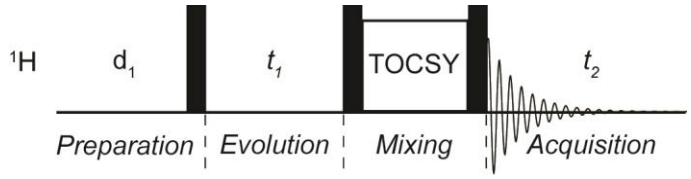


Figure 3.10 Pulse sequence for 2D TOCSY experiment. Black rectangle represents 90° pulses, and d_1 represents recycle delay. A train of pulses is applied during isotropic *mixing*, such as WALTZ-16 [34] or MLEV-16 [35].

During *mixing*, a train of pulses allows for coherence transfer from one spin system to another via *J*-couplings [36]. The advantage of using TOCSY over COSY is the absence of anti-phase terms that generate dispersive lineshapes in the cross-peaks [33]. Hence, the cross-peak intensities are not diminished by the severe overlap of broad dispersive lineshapes (section A.3.1).

Despite the benefits that TOCSY offers, its application is limited to organic molecules in solution. The solid-state analogue of TOCSY, **TO**tal **B**ond **S**pectroscop**Y** (TOBSY) incorporates a *J*-coupling mediated magnetization transfer via isotropic mixing to obtain homonuclear correlations [37]. The 2D (^{13}C-^{13}C) TOBSY experiment is effective for a variety of biological molecules ranging from a simple decapeptide [38] to a complex protein such as bacteriochlorophyll a [39] and hydrated NRSMC elastin [40].

The pulse sequence for 2D R-TOBSY experiment is similar, but not identical, to that of TOCSY (**Figure 3.11**). The initial magnetization is *prepared* by a DP-nOe excitation, i.e., a low-power RF irradiation on the ^{1}H channel, prior to the ^{13}C 90° pulse. This steady-state nOe condition is used to maximize the non-equilibrium ^{13}C populations by cross-relaxation. During t_1 *evolution*, carbon magnetization evolves under the chemical shift offset of site 1 (Ω_1) and the *J*-coupling(s). The *mixing* period consists of the R-TOBSY sequence that is sandwiched by two ^{13}C 90° pulses. The $R30_6$[14] cycle suppresses dipolar and anisotropic interactions, but it does not affect *J*-couplings during coherence transfer, as recently demonstrated for proline characterization in NRSMC elastin [40]. TPPM decoupling is applied during t_1, *mixing* and t_2 periods.

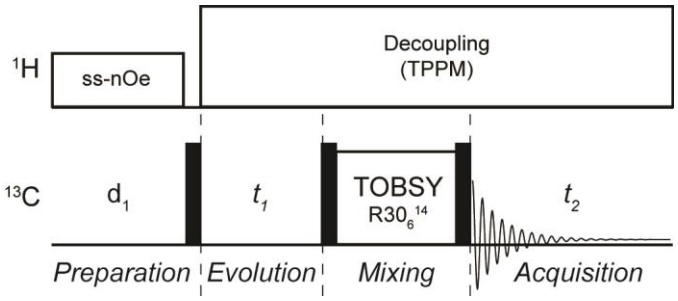


Figure 3.11. Pulse sequence for 2D R-TOBSY experiment. MAS is applied during the experiment. Steady-state nOe (ss-nOe) indicates low-power ^{1}H irradiation during d_1, which is the recycle delay. Black rectangles indicate 90° pulses. The $R30_6^{14}$ sequence [41] is used during *mixing* period.

Solid-state NMR spectroscopy provides the means to characterize alanines in NRSMC elastin. One-dimensional DP, rINEPT and CP experiments yield ^{13}C-Ala spectra with isotropic chemical shifts indicate secondary structures. The two-dimensional experiments such as R-TOBSY and CTUC-COSY provide homonuclear ^{13}C-^{13}C correlation spectra, which are useful for distinguishing alanine populations in the crosslinking and hydrophobic domains. NMR relaxation measurements assess the MHz and kHz-scale motions of alanines in elastin, which support tentative conformational assignments through ^{13}C chemical shifts.

B. MATERIALS AND METHODS

B.1 Preparation of [U-^{13}C]Ala- and [^{15}N,U-^{13}C]Ala elastin samples for ssNMR measurements

Elastin enriched with [U-^{13}C]Ala and [^{15}N, U-^{13}C]Ala were prepared using the method described in **Chapter 2**. Both samples were hydrated. For MAS, ~50-60 mg of hydrated samples were packed into a 4.0-mm rotors. Hydration level was maintained by sealing the sample with Kel-F (top & bottom) spacers, fitted with fluorosilicone micro o-rings (Apple Rubber Products, Lancaster NY). Masses of rotors were measured before and after each experiment to monitor the hydration level.

B.2 Solid-state NMR experiments

Solid-state NMR experiments were performed on an Agilent DD2 console, running at a ^{1}H resonance frequency of 399.976 MHz. ^{13}C spectra were acquired using a 4.0-mm triple resonance (HXY) T3 MAS probe (Varian-Chemagnetics, Fort Collins, CO). The MAS rate was 8 kHz for all experiments. The ^{13}C chemical shifts were referenced to the tetramethylsilane (TMS) scale, using a hexamethylbenzene (HMB) as an external standard ($\delta(^{13}CH_3)$ = 17.0 ppm), at room temperature. All ssNMR experiments were conducted at 37 °C. The two-step rINEPT experiment was performed using the [U-^{13}C, ^{15}N]Ala elastin sample. Other experiments were performed using [U-^{13}C]Ala elastin.

For direct polarization (DP), a 4.0-4.5 μs ^{13}C 90° pulse was used with 6-10 s recycle delay. For cross-polarization (CP) [23], 4.5-5.0 μs ^{1}H 90° pulse was used with a 1 ms contact time and 5 s recycle delay. The CP transfer was set to the field strength ($\gamma^H.B_1^H/2\pi = \gamma^C.B_1^C/2\pi$) of ~50 kHz. For 1D rINEPT [27], the recycle delay was set to 1.5 s, and the ^{1}H and ^{13}C 90° pulses were 4.5 μs and 5.0 μs, respectively. Echo delays τ and τ' in **Figure 3.4** were set to 1.2 and 1.0 ms, respectively. TPPM (^{1}H) decoupling was applied during acquisition, using the field strength ($\gamma^H.B_1^H/2\pi$) of ~40-60 kHz.

For ^{13}CO detection with $^{1}J_{CaCO}$ decoupling, the **L**ong-**O**bservation-**W**indow **BA**nd-**S**elective **H**omonuclear **D**ecoupling (LOW-BASHD) was used [21]. The ^{13}C transmitter frequency was set at 175.6 ppm (on resonance with ^{13}CO-Ala). Selective 200 μs Gaussian 180° (π) pulses with cosine amplitude-modulation were applied every 8 ms. Additional 4 ms delays were included before and after the first and last π pulses. The total acquisition time was 40 ms for each scan, corresponding to five cycles of π pulses. The amplitude of modulation frequency was 12,600 Hz, which is the difference between ^{13}CO and ^{13}Cα-Ala resonance frequencies. The spectral width was set to 2,500 Hz, and it corresponds to a 400 μs dwell time. In this dwell time, the receiver was gated off for 200 μs to accommodate for the π pulse; data points

were consequently oversampled in the remaining 200 μs. This oversampling results in the attenuation of FID intensities where the π pulses were applied, and the attenuation gives rise to decoupling sidebands in the acquired ^{13}CO spectrum. The sideband intensities were negligible in this study; hence, they are ignored during spectral processing.

Windowed detection with **SE**lective **D**ecoupling **U**sing **C**rafted **E**xcitation (wSEDUCE-1) was used to observe ^{13}Cβ DP spectrum without $^{1}J_{C\alpha C\beta}$. SEDUCE-1 is a homonuclear decoupling method that uses amplitude-modulated, frequency-selective, shaped pulses [22]. The length of wSEDUCE-1 waveform was 638.4 μs, and it is placed on-resonance with ^{13}Cα–Ala. Phase cycling was performed using MLEV-16 supercycle [35]. The dwell time (33.6 μs) was shared between the windowed acquisition and the frequency-selective irradiation. The length of the ^{13}C irradiation was 38% of the dwell time (12.6 μs), and the highest B_1 field strength ($\gamma.B_1^C/2\pi$) among the applied pulses was 1.1 kHz. CW (^{1}H) decoupling with the field strength ($\gamma.B_1^H/2\pi$) of ~37 kHz was used during detection.

For the 1D two-step rINEPT experiment, the initial magnetization was prepared using a 5.1 μs ^{1}H 90° pulse. The ^{1}H→^{15}N transfer (**I**) (**Figure 3.5b**) was performed with 5.1 μs ^{1}H 90° and 9.0 μs ^{15}N 90° pulses. The delays in *spin-echo-1* and *spin-echo-2* periods, τ′ and τ′, were 2.5 and 2.2 ms, respectively. The ^{15}N→^{13}CO rINEPT (**II**) was performed with 9.0 μs ^{15}N 90° and 4.5 μs ^{13}C 90° pulses. The delays in *spin-echo-1* and *-2* periods, τ_1' and τ_1', were 14 and 10 ms, respectively. For the carbonyl-selective π pulses during *spin-echo-3* and *-4*, **R**efocused **S**elective excitatio**N** f**O**r **B**iochemical Applications (r-SNOB) pulses (pulse width: 438 μs and maximum field strength of 6.0 kHz) were used [42]. The ^{13}C transmitter frequency was set at 175.6 ppm (on resonance with ^{13}CO-Ala). TPPM decoupling was used throughout the ^{15}N-^{13}C rINEPT transfer and acquisition period. LOW-BASHD decoupling was used during ^{13}C acquisition.

For the 2D CTUC-COSY experiment, the initial magnetization was obtained by a 4.5-μs ^{13}C 90° pulse. The constant-time evolution period was used during t_1 evolution (**Figure 3.9b**). The delays in *spin-echo-1* and *spin-echo-2* periods, τ and τ′, were 4.0 and 2.5 ms, respectively. Two ^{13}C carrier frequencies were used for the 2D measurements; 60.0 ppm during t_1 and 176.6 ppm during detection. The z-filter in this experiment has a length of 10 msec. ^{1}H (TPPM) decoupling was used throughout the experiment, and ^{13}C LOW-BASHD decoupling was used during acquisition. The spectrum was processed using 5 Hz of line broadening on both f_1 and f_2 dimensions.

For the 2D R-TOBSY experiment, a DP excitation with steady-state nOe was used to obtain the initial magnetization. The nOe was achieved using a low-power ^{1}H irradiation (0.7 kHz) during the 3s recycle delay. Then, a 3.9 μs ^{13}C 90° pulse was applied, followed by TOBSY mixing (**Figure 3.11**). To maximize the transfer efficiency for ^{13}Cα-^{13}Cβ correlations, the ^{13}C carrier frequency was placed in the

middle of $^{13}C\alpha$ and $^{13}C\beta$ signals (34.4 ppm). The spectral width in the direct dimension (f_2) was 5,600 Hz. Sixty-four scans were acquired for each t_1 point, and the number of t_1 points was 40. The total experimental time was 4 hours. To maximize the transfer efficiency for ^{13}CO-$^{13}C\alpha$ correlations, the ^{13}C carrier frequency was placed at the middle of ^{13}CO and $^{13}C\alpha$ signals (113 ppm). The spectral width in the direct dimension was 17,500 Hz. Sixty-four scans were acquired on each t_1 point, and the number of t_1 points applied was 96. The total experimental time was 10 hours.

The TOBSY mixing period has the symmetry of $R30_6$[14] [41]. The mixing time (τ_{mix}=6 ms) is comprised of 8 cycles. Each cycle contains 30 R elements that span 6 rotor periods. Two successive R elements are phase-shifted by 84° and -84°. A ^1H CW decoupling was used during mixing ($\gamma^H B_1^H/2\pi$ ~ 50 kHz) with an offset of 60 kHz. TPPM decoupling ($\gamma^H B_1^H / 2\pi$ = 40-60 kHz) was applied during the 10 ms acquisition time, and LOW-BASHD was used for the ^{13}CO detection.

The ^{13}C T_1 experiment was performed with the inversion recovery method. The ^{13}C 180° and 90° pulse widths (**Figure 3.6a**) were 10 μs and 5 μs, respectively. The ^{13}C $T_{1\rho}$ experiment used a DP-based excitation (**Figure 3.6b**). A single 90° ^{13}C pulse is followed by a spin-locking period at $\gamma^C B_1^C/2\pi$ = 50 kHz. To obtain T_1, ^{13}C peak intensities were fit to $I(t)$= A*exp(-τ/T_1)+B. To obtain $T_{1\rho}$, the ^{13}C peak intensities were fit to a single exponential decay function, $I(t)$= A*exp(-τ/T_1).

C. RESULTS AND DISCUSSIONS

Solid-state NMR spectroscopy characterizes alanines in hydrated NRSMC elastin at 37 °C. Conformation-dependent isotropic ^{13}C chemical shifts were used to assign the secondary structures of alanines (section C.1), and two-dimensional ssNMR experiments were performed to resolve signal overlap and to validate the conformational assignments for alanines (sections C.2.1-C.2.2). The two-step rINEPT experiment detected the mobile region in the crosslinking domains (section C.2.3). NMR relaxation parameters were measured to assess MHz- and kHz-scale motions of alanines in NRSMC elastin (section C.3).

C.1 Multiple alanine populations exist in NRSMC elastin

The ^{13}C DP spectrum reflects all carbon populations, regardless of conformation and dynamics (section A.2). Preliminary structural assignments were obtained from the ^{13}C DP spectrum of hydrated [U-^{13}C]Ala NRSMC elastin at 37 °C (**Figure 3.12**). The spectrum is characterized by asymmetric lineshapes for the ^{13}CO, $^{13}C\alpha$ and $^{13}C\beta$ sites of alanines. The ^{13}CO signal has a narrow lineshape (FWHM ~160 Hz) with two peaks, 176.1 and 175.7 ppm, and a downfield shoulder at ~177 ppm. The $^{13}C\alpha$ signal appears homogeneous, as only one broad lineshape (FWHM ~90 Hz) is observed with the tallest point at 50.6 ppm. The highest $^{13}C\beta$ intensity is identified at 17.4 ppm with an upfield shoulder at ~16 ppm. ^{13}CO, $^{13}C\alpha$ and $^{13}C\beta$ resonances at 176.7, 50.6, and 17.4 ppm indicate random coil alanines [2, 3]. In contrast, alanines in α-helices are found at ~177 ppm (^{13}CO) and ~16 ppm ($^{13}C\beta$) [1, 43]. This result indicates the presence of multiple populations of alanines in hydrated NRSMC elastin at 37 °C. These populations include α-helical and random coil populations, which were previously identified in elastin-based peptides [11, 44].

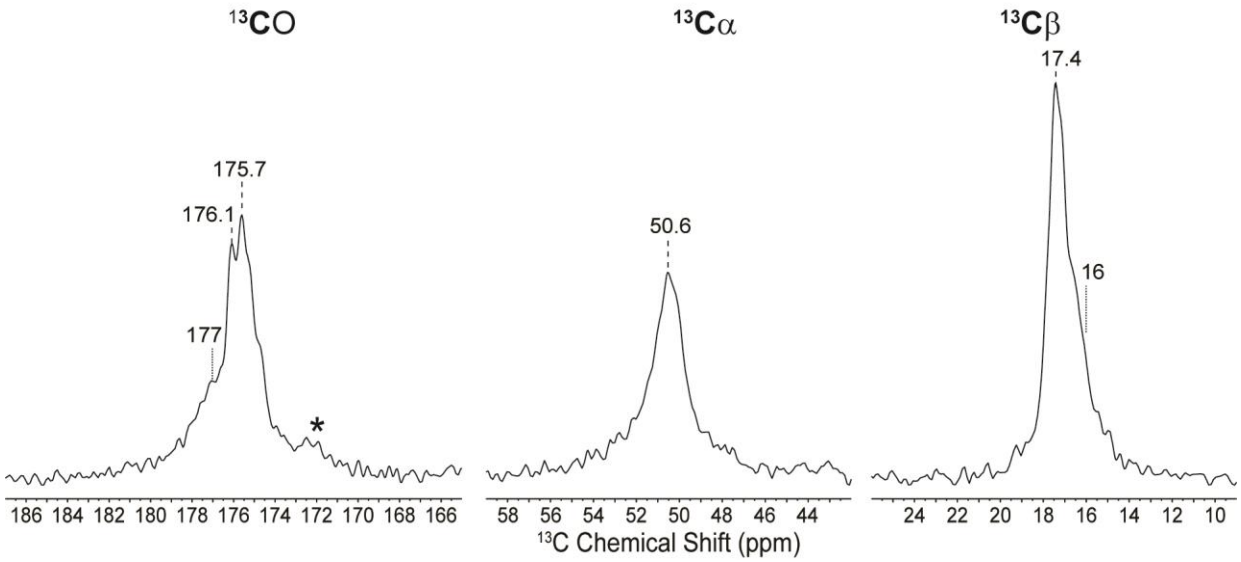


Figure 3.12 ^{13}C DP spectrum of hydrated [U-^{13}C]Ala NRSMC elastin at 37 °C. Dashed lines (┆) indicate resolved peaks, and dotted lines (⋮) indicate poorly resolved features. Asterisk (*) in ^{13}CO lineshape indicate the natural-abundance ^{13}CO-Gly signal.

Broadened lineshapes due to homonuclear (^{13}C-^{13}C) *J*-couplings are observed in the ^{13}C DP spectrum of [U-^{13}C]Ala elastin (**Figure 3.12**). This effect makes conformational assignments based on isotropic chemical shifts difficult. To remove $^{1}J_{COC\alpha}$ and $^{1}J_{C\beta C\alpha}$ effects, ^{13}CO- and ^{13}Cβ-Ala spectra were acquired using DP with LOW-BASHD [21] and wSEDUCE-1 [22] decoupling. ^{13}C DP spectra with LOW-BASHD and wSEDUCE-1 decoupling have different linewidths than those acquired simply by DP. The linewidth for the ^{13}CO-Ala signal decreased from ~165 to ~125 Hz, whereas that for ^{13}Cβ-Ala increased slightly from ~100 to ~110 Hz (**Appendix 4**). The ^{13}CO-Ala spectrum (**Figure 3.13a**) is characterized by a narrow lineshape with features at 175.9 and 175.0 ppm and a downfield shoulder approximately at 177 ppm. The highest intensity of the ^{13}Cβ-Ala signal is at 17.1 ppm and a shoulder feature is identified at ~16 ppm. The asymmetry in these lineshapes suggests the presence of multiple alanine populations.

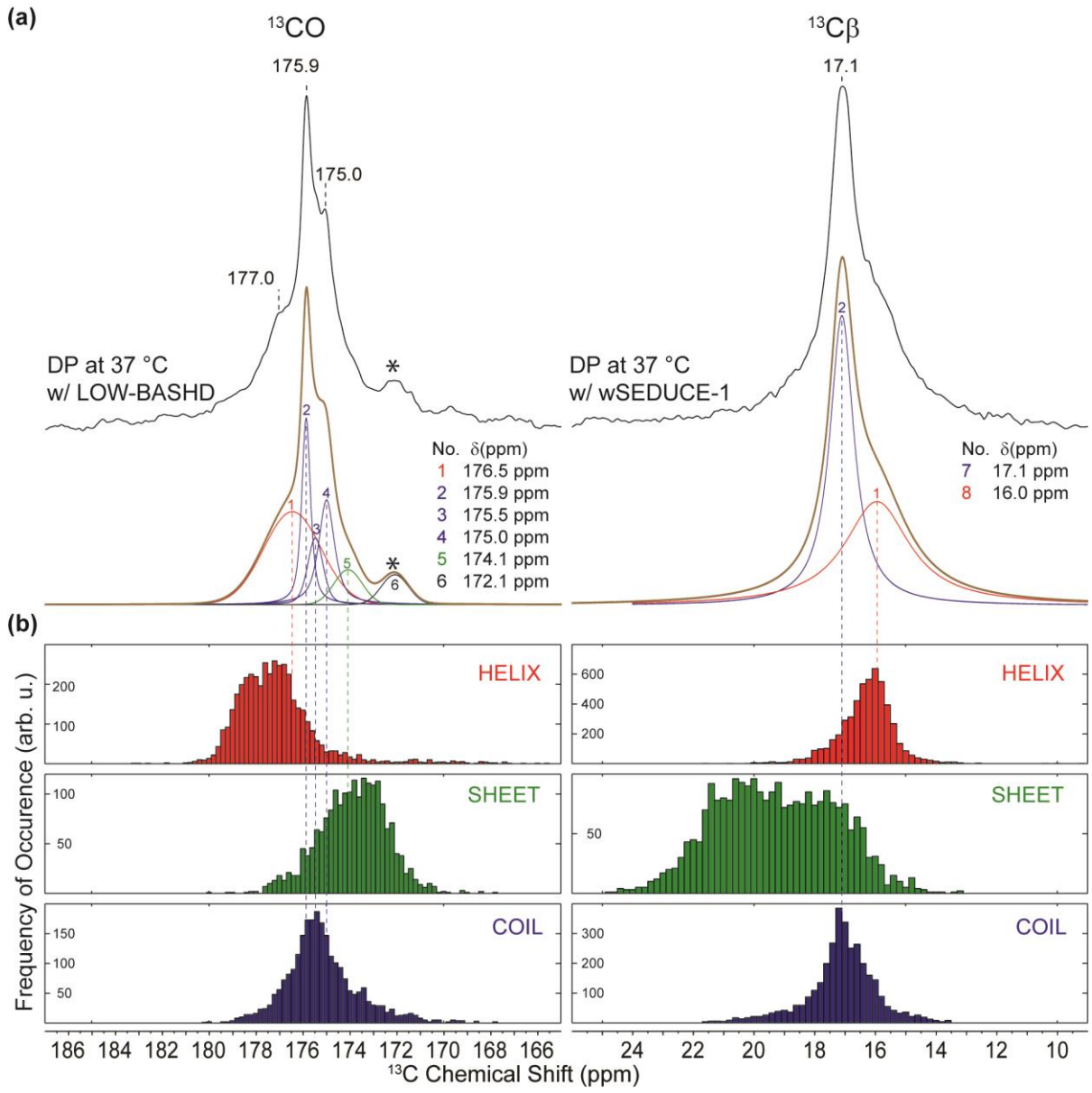


Figure 3.13 (a) ^{13}C DP spectra of hydrated [U-^{13}C]Ala NRSMC elastin at 37 °C and their deconvolution results; (b) distributions of ^{13}C chemical shifts of alanines in RefDB [45], categorized as HELIX, SHEET, and COIL. MAS is applied during the experiment. In Fig 3.13(a), *top*, decoupled ^{13}CO- and ^{13}Cβ-Ala lineshapes were acquired using DP with LOW-BASHD [21] and DP with wSEDUCE-1 [22], respectively. Dashed lines (¦) indicate resolved features in ^{13}C DP spectra, and asterisk (*) marks the natural-abundance ^{13}CO-Gly signal. In Fig 3.13(a), *bottom*, deconvolution subroutine is performed with Lorentzian-Gaussian (L/G) components, colored according to the corresponding structures. The sum all L/G shapes is shown in **brown**. Histograms in Fig 3.13(b) reflect the distributions of ^{13}C chemical shifts of alanines in the corresponding structures; the height of each bar (in arb. u.) reflects the frequency of occurrence of a given ^{13}C chemical shift in 2162 proteins in RefDB, as of September 2016. Colored vertical dashed lines in Fig 3.13(b) and (c) reflect the agreement of ^{13}C chemical shifts between the center-of-mass of each L/G component and the given populations.

To identify distinct populations of alanines, a deconvolution subroutine was performed on each spectrum (**Figure 3.13b**). Each lineshape was fitted with Lorentzian-Gaussian (L/G) functions, and the results were interpreted with ^{13}C chemical shift distributions of alanines in a number of proteins that are archived in RefDB [45]. (RefDB is a database of protein chemical shifts derived from the Biological Magnetic Resonance Data Bank (BMRB) [46], containing ^{1}H, ^{13}C, and ^{15}N chemical shifts of 2162 proteins (as of September 2016). These chemical shifts have been re-referenced and corrected using SHIFTX [47], a program that calculates NMR resonance frequencies from the X-ray or NMR atomic coordinates.) The distributions of ^{13}C chemical shifts for alanines are categorized into HELIX, SHEET, and COIL (**Figure 3.13b**), which correspond to the α-helix, β-sheet and other motifs. The ^{13}CO signal requires six components to reproduce the experimental result. The major contribution **1** (45%) at 176.5 ppm likely arises from alanines in an α-helix, which are most the frequently observed structure among other motifs. Components **2**, **3**, and **4** at 175.9, 175.5, and 175.0 ppm contribute to the total Ala population by 18%, 10%, and 19%, respectively. These resonances are consistent with COIL ^{13}C chemical shift distributions. Additionally, these three components reflect the *neighboring residue* effects in the ^{13}CO-Ala chemical shifts (further discussed in section C.2.2). The minor component **5** (8%) at 174.1 ppm is tentatively assigned as β-sheet, as its center-of-mass coincides with SHEET. The $^{13}C\beta$-Ala lineshape was adequately fitted using two components at 17.1 ppm (**7**) and 16.0 ppm (**8**). The broader component (FWHM ~260 Hz) at 16.0 ppm (49%) agrees with the ^{13}C resonance of alanines found most frequently as HELIX in RefDB. The narrower component (FWHM ~100Hz) at 17.1 ppm (51%) is consistent with ^{13}C-Ala resonances of both SHEET and COIL. These tentative assignments are summarized on **Table 3.3**.

The deconvolution of ^{13}CO- and $^{13}C\beta$-Ala lineshapes identifies lower α-helical content than expected from theoretical values. In theory, the amount of a given secondary structure in a protein can be calculated from its primary structure using prediction algorithms like GOR IV [48] and PSIPRED [49]. GOR IV [48] predicts that 58% of tropoelastin's amino acids are found in the α-helix (**Appendix 5**). This prediction is consistent with the total number of alanines in the crosslinking domains of rat tropoelastin (57%), implying that all Ala in the corresponding regions are α-helical. However, this theoretical estimation is inconsistent with the ssNMR results of [U-^{13}C]Ala NRSMC elastin. Namely, the major components in ^{13}CO- and $^{13}C\beta$-Ala lineshapes, i.e., at 176.5 ppm (**1**) and 16.0 ppm (**8**), respectively, only account for 45-49% of the total Ala populations. Hence, the deconvolution shows that fewer alanines are found in α-helices, consistent with the literature findings [50].

Table 3.3. Results of deconvolution subroutines. ^{13}CO- and ^{13}Cβ-Ala chemical shifts correspond to the center-of-masses of L/G components in the deconvolution of DP spectra with LOW-BASHD and wSEDUCE-1 decoupling, respectively. Tentative assignments of each component is made based on the ^{13}C chemical shift distributions of alanines in 2162 proteins in RefDB [45]. HELIX and SHEET represent the α-helix and β-sheet structures, respectively. COIL indicates other motifs, which are not α-helix nor β-sheet.

Sites	No.	δ (ppm)	Width (Hz)	Area (%)	Structural Motif
^{13}CO	1	176.5	290	45	HELIX
	2	175.9	44	18	COIL, SHEET
	3	175.5	77	10	COIL, SHEET
	4	175.0	80	19	COIL, SHEET
	5	174.1	140	8	SHEET
	6	172.1	140	7	^{13}C-Gly
^{13}Cβ	7	17.1	97	51	COIL, SHEET
	8	16.0	260	49	HELIX

Multiple conformations of alanines are observed from the 1D ssNMR study of hydrated NRSMC elastin at 37 °C. The ^{13}C chemical shifts of alanines indicate at least two populations are present, α-helix and random coil. The α-helical content in elastin's crosslinking domains seems lower than the predicted, which is consistent with other studies [11, 14].

C.2 Alanines in NRSMC elastin are found primarily in α-helix and random coil populations

As shown in the previous sections, alanines in [U-^{13}C]Ala NRSMC elastin at 37 °C exist in multiple populations. Deconvolution subroutine of the 1D ^{13}C DP spectrum suggests the presence of common motifs such as α-helix, β-sheet and random coil. However, a precise identification of these motifs is hampered by the overlap ^{13}C-Ala signals. Two-dimensional ssNMR spectroscopy addresses the spectral overlap in a 1D spectrum and facilitates conformational assignments. The 2D R-TOBSY experiments yield ^{13}CO-^{13}Cα and ^{13}Cα-^{13}Cβ correlations in hydrated NRSMC elastin at 37 °C (section C.2.1), which is useful for identifying alanine populations in the system. The CTUC-COSY with LOW-BASHD experiment obtains ^{13}CO-^{13}Cα correlations without *J*-couplings (section C.2.2), which simplifies the lineshape in the carbonyl dimension (f_2).

C.2.1 Alanines in the crosslinking domains have different conformations from those in the hydrophobic domains. Two-dimensional R-TOBSY obtains ^{13}CO-^{13}Cα and ^{13}Cα-^{13}Cβ correlations in hydrated NRSMC elastin at 37 °C. The DP-nOe excitation in this experiment enhances signals from

mobile regions in the sample (section A.3.1). This method yields a complete set of ^{13}C-^{13}C chemical shifts correlations, useful for the characterization of alanine populations.

Asymmetrical lineshapes were observed in the 2D cross-peaks of ^{13}CO-$^{13}C\alpha$ correlation spectrum (**Figure 3.14**). The highest intensities span broad distributions of ^{13}CO resonances from 174.5 to 176.5 ppm and $^{13}C\alpha$ resonances from 49 to 51.5 ppm. Two components are observed at (175.6, 50.5) (*green line*) and (175.0, 50.1) (*blue line*) in the (^{13}CO, $^{13}C\alpha$) correlations. A minor peak is observed at 176.7 ppm (^{13}CO) and 51.5 ppm ($^{13}C\alpha$) (*red line*).

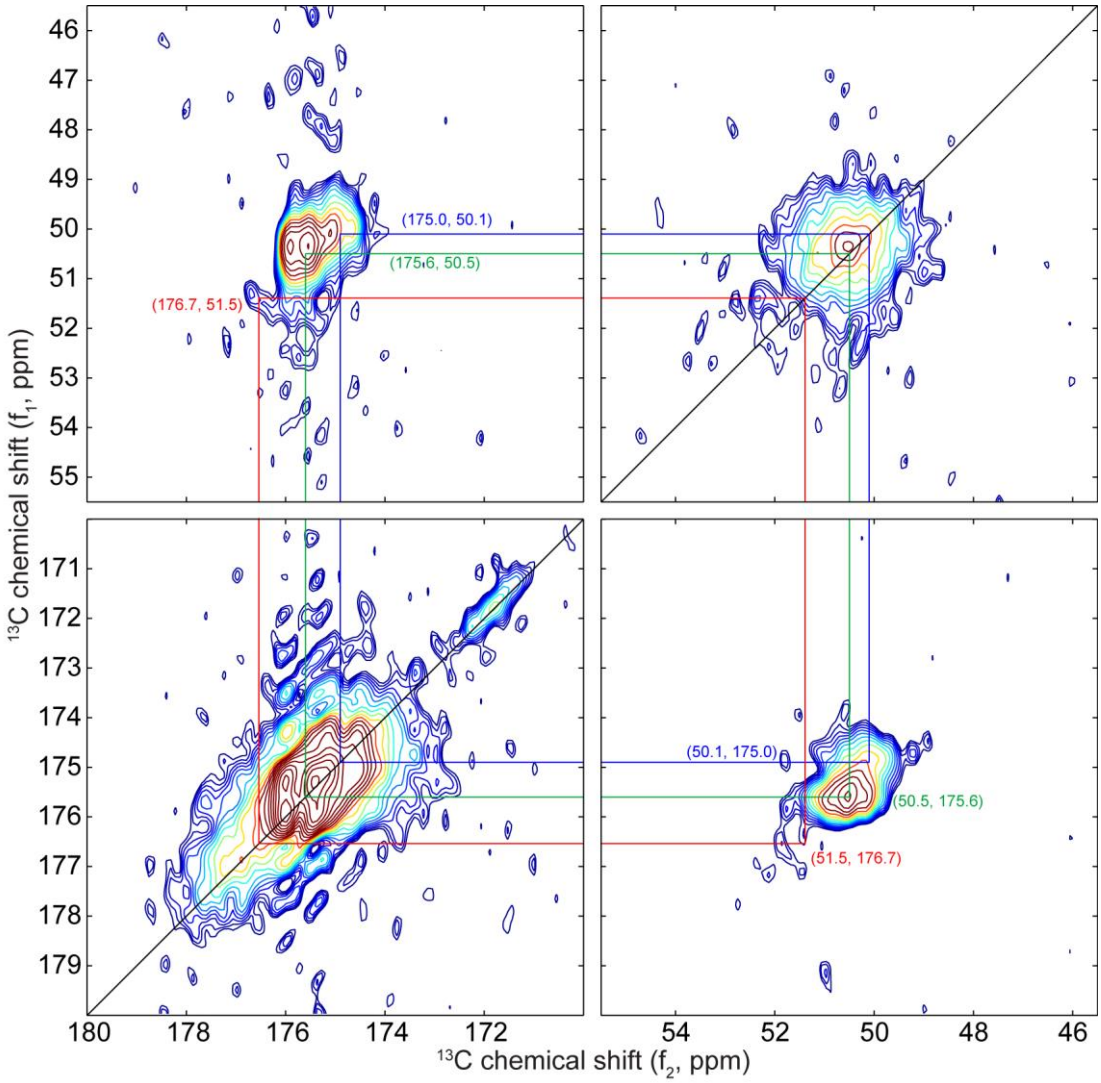


Figure 3.14. ^{13}CO-$^{13}C\alpha$ correlations in the 2D R-TOBSY spectrum of hydrated [U-^{13}C]Ala elastin at 37 °C. The spectrum is processed with 20 Hz line broadening in f_1 and f_2 dimensions. Colored lines approximate positions for alanine populations.

Multiple Ala populations in the ^{13}CO-^{13}Cα correlations are consistent with those in the ^{13}Cα-^{13}Cβ spectrum (**Figure 3.15**). The cross-peaks cover broad ranges of ^{13}CO and ^{13}Cβ resonances, i.e., ~49-52 ppm and ~16.5-18.5 ppm, respectively. The prominent peak at (^{13}CO, ^{13}Cα) = (50.5, 17.1) (*green line*) overlaps with another signal at (50.1, 17.4) (*blue line*). A minor cross-peak at (51.5, 16.6) (*red line*) is also observed.

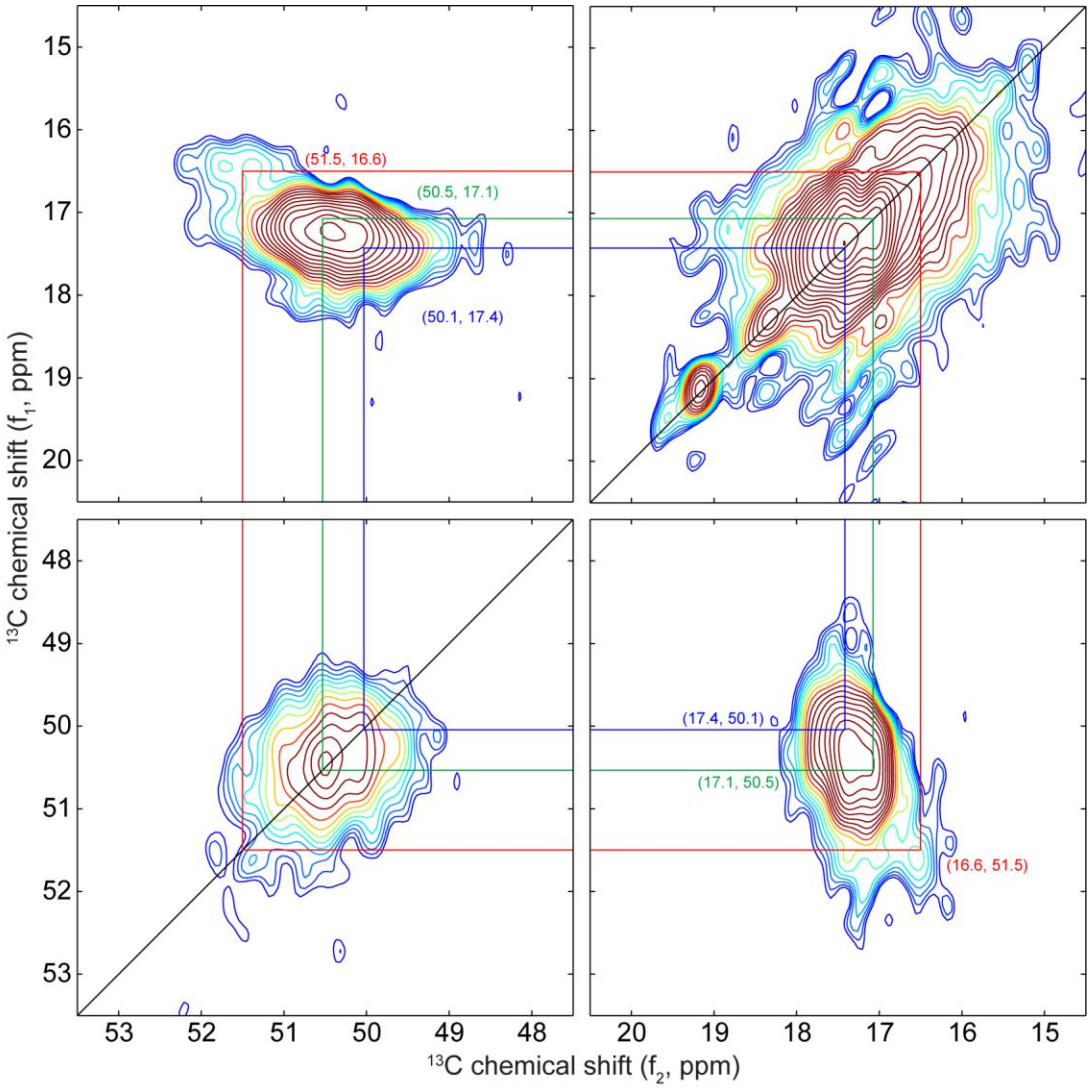


Figure 3.15 ^{13}Cα-^{13}Cβ correlations in the 2D R-TOBSY spectrum of hydrated [U-^{13}C]Ala elastin at 37 °C. The spectrum is processed with 20 Hz line broadening in f_1 and f_2 dimensions. Colored lines indicate approximate positions for alanine populations.

The conformations of alanines were assigned using on the ^{13}CO-^{13}Cα and ^{13}Cα-^{13}Cβ correlations (above). Alanine populations showing the highest intensities in the 2D spectrum correspond to those found in random coil [2, 3], as indicated the ^{13}CO, ^{13}Cα and ^{13}Cβ intensities at 175.6, 50.5, and 17.1 ppm,

respectively. The second population is centered at 175.0 (^{13}CO), 50.1 (^{13}Cα), and 17.4 ppm (^{13}Cβ), also corresponds to random coil alanines due to the *neighboring residue* effect (section C.2.2). The third population is assigned as alanines in the α-helical conformation, as indicated by the cross-peaks located at (176.7, 51.5, 16.6) in the (^{13}CO, ^{13}Cα, ^{13}Cβ) resonances. These chemical shift correlations are summarized in **Table 3.4**. The 2D R-TOBSY spectra provide evidence for the conformational heterogeneity of alanines, which support previous 1D assignment by deconvolution subroutines (section C.1).

Table 3.4. Chemical shifts of alanines in elastin from 2D R-TOBSY measurements at 37 °C with tentative structural assignments.

^{13}C δ (ppm)			Elastin Domain	Conformations
CO	Cα	Cβ		
176.7	51.5	16.6	Crosslinking	α-helix [1, 43]
175.6	50.5	17.1	Hydrophobic	Random coil [2, 3]
175.0	50.1	17.4	Crosslinking	Random coil [2, 3]

Broad features due to ^{13}C-^{13}C *J*-couplings are observed in the 2D R-TOBSY spectra (**Figure 3.14** and **3.15**). The cross-peaks indicate that two ^{13}CO-Ala resonances (175.6 and 175.0 ppm) correspond to the random coil population (**Table 3.4**). However, the deconvolution of 1D ^{13}CO DP spectrum suggests three ^{13}CO-Ala signals, which are centered at 175.9, 175.5, and 175.0 ppm (section C.1). The discrepancy is resolved by the acquisition of 2D spectra without *J*-couplings (next section).

C.2.2 Random coil alanines are present in both hydrophobic and crosslinking domains of NRSMC elastin. The CTUC-COSY with LOW-BASHD experiment yields 2D ^{13}CO-^{13}Cα correlations of alanines in the mobile region of hydrated NRSMC elastin. This method removes $^{1}J_{COC\alpha}$ in both f_1 and f_2 dimensions using constant-time evolution and homonuclear decoupling during t_2, respectively (section A.3.2). The ^{13}CO-^{13}Cα correlation spectrum reflects three populations of alanines in the mobile region of hydrated NRSMC elastin at 37 °C (**Figure 3.16a**). The major peak is observed at (^{13}CO, ^{13}Cα) = (175.9, 50.6) ppm. Two minor components are identified at (175.3, 50.3) and (175.0, 50.2). These ^{13}CO resonances are fairly consistent with those corresponding to components **2**, **3**, and **4** (**Table 3.3**, section C.1), identified by the deconvolution of ^{13}CO-Ala DP lineshape. The 0.3-0.9 ppm differences in carbonyl chemical shifts, $\Delta\delta$ (^{13}CO) are smaller than the 1-3 ppm (differences) used to distinguish secondary structures [7]. Hence, the three features did not arise from the variation of secondary structures but most likely from *neighboring residue* effects.

.

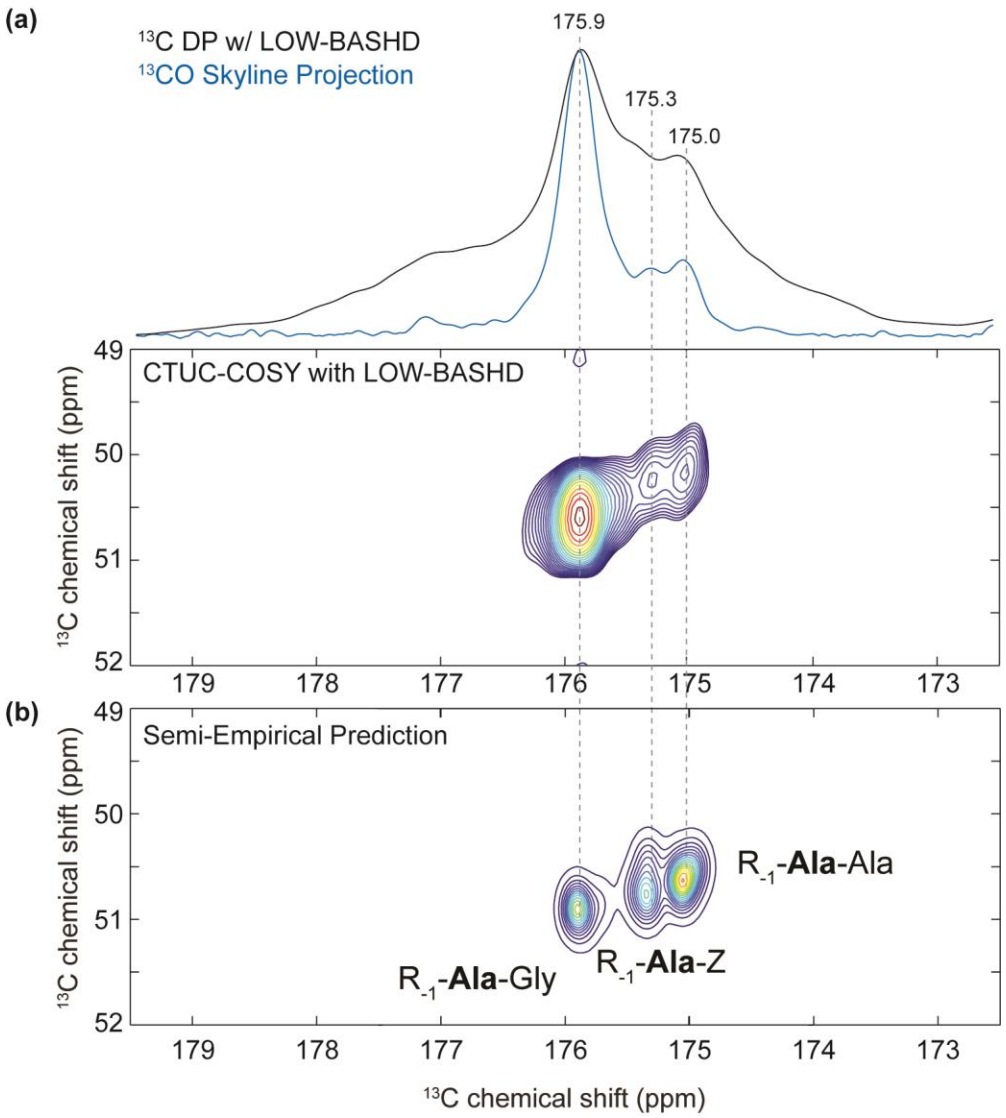


Figure 3.16 Comparison of experimental and predicted spectra for alanines in NRSMC elastin. The 2D ^{13}CO-^{13}Cα correlations obtained by (a) CTUC-COSY with LOW-BASHD experiment, and (b) semi-empirical prediction based on tropoelastin's primary structure [9]. MAS is applied during the experiment. Part (a) shows the skyline projection of the 2D cross-peak (**blue**) and ^{13}CO-Ala DP with LOW-BASHD spectrum (**black**), reproduced from **Figure 3.13a** (section C.1). Skyline projection reflects the tallest intensities in the 2D cross-peak. The 2D CTUC-COSY cross-peaks reflect alanines in the mobile region of elastin. Part (b) reflects ^{13}CO-^{13}Cα correlations that are expected of alanines in a fully random coil state. Vertical dashed lines (¦) indicate resolved carbonyl peaks in the corresponding populations.

The semi-empirical approach [9, 51] was employed to predict the spectrum for fully (100%) random coil alanines in NRSMC elastin. Briefly, the calculation uses the ^{13}C chemical shifts of random coil **Ala**, obtained from the solution NMR study of the model peptide Ac-GG**Ala**GG-NH_2 in 8 M urea [52].

Then, the corrected chemical shifts which considers the *neighboring residue* effects on **Ala** [51] was computed using

$$\delta_{\boldsymbol{Ala}}\,(corrected) = \delta_{rc}(\boldsymbol{Ala}) + \Delta\delta(R_{-1}) + \Delta\delta(R_{+1}) \qquad \text{(eqn. 3.2)}$$

where δ_{rc}(**Ala**) is the base value for random coil **Ala** [52]; $\Delta\delta(\mathrm{R_{-1}})$ and $\Delta\delta(\mathrm{R_{+1}})$ are correction factors for the respective preceding and following residues in $\mathrm{R_{-1}}$-**Ala**-$\mathrm{R_{+1}}$ motifs (**Figure 3.1**, section A.1). These were calculated systematically for each of the three-amino-acid sequences found in tropoelastin, described in **Table 3.2** (section A.1). **Appendix 6** shows an example calculation for the ^{13}CO chemical shift of **Ala** in the Lys-**Ala**-Tyr motif.

Corrected chemical shifts for random coil alanines were then plotted as a 2D ^{13}CO-^{13}Cα correlation spectrum (**Figure 3.16b**), utilizing Gaussian functions (with FWHM of 0.10 ppm in both dimensions) to reflect the relative number of occurrences of the given sequence in tropoelastin. The preceding residue ($\mathrm{R_{-1}}$) has very little *neighboring residue* effect on the ^{13}CO chemical shift of **Ala**. Hence, a contour peak labeled as, e.g., $\mathrm{R_{-1}}$-**Ala**-Gly means that all alanines followed by glycines have similar ^{13}CO chemical shifts, with little effect from the $\mathrm{R_{-1}}$ residue. **Appendix 7** shows the influence of $\mathrm{R_{-1}}$ residues on the central alanines. This approach predicted three contour peaks for the fully random coil population of alanines, as indicated by $\mathrm{R_{-1}}$-**Ala**-Gly, $\mathrm{R_{-1}}$-**Ala**-Ala, and $\mathrm{R_{-1}}$-**Ala**-Z (Z ≠ Ala, Gly). The greatest intensity at (^{13}CO, ^{13}Cα)=(175.9, 50.6) corresponds to the $\mathrm{R_{-1}}$-**Ala**-Ala sequences, mostly found in the crosslinking domains. The smaller peak at (175.9, 50.9) arises from the $\mathrm{R_{-1}}$-**Ala**-Gly motifs, which are abundant in the hydrophobic domains. The smallest contour peak is observed at (175.3, 50.8), corresponding to $\mathrm{R_{-1}}$-**Ala**-Z sequences (Z≠Ala, Gly), which are found in both hydrophobic and crosslinking domains.

Peak positions observed from the ^{13}CO-^{13}Cα correlation experiment are consistent with those obtained from the semi-empirical prediction, but the intensities are not comparable. Specifically, the CTUC-COSY peaks with ^{13}CO-Ala resonances of 175.9, 175.3, and 175.0 ppm (**Figure 3.16a**, *bottom*) are in good agreement with the predicted $\mathrm{R_{-1}}$-**Ala**-Gly, $\mathrm{R_{-1}}$-**Ala**-Z, and $\mathrm{R_{-1}}$-**Ala**-Ala populations, respectively (**Figure 3.16b**). However, the peak intensity at (175.0, 50.2) in the 2D spectrum is significantly lower than the predicted contour for $\mathrm{R_{-1}}$-**Ala**-Ala sequences, indicating that <100% of total alanines are found in this population. This result indicates that only a minor portion of alanines in the crosslinking domains are found in the mobile region of NRSMC elastin, which is random coil. Presumably, the majority of the alanines with $\mathrm{R_{-1}}$-**Ala**-Ala motifs is found in the rigid α-helical regions at ~177 ppm, not detected by the CTUC-COSY experiment.

C.2.3 Random coil population of alanines are abundant in the crosslinking domains of denatured elastin. The above described results suggest that the most upfield signal in both 1D ^{13}CO-Ala DP and 2D ^{13}CO-^{13}Cα correlation spectra (**Figure 3.16a**) correspond to the alanines in the crosslinking domains. To confirm this assignment, the ^{13}CO-Ala spectrum corresponding to the R$_{-1}$-**Ala**-Ala sequence was obtained using the two-step rINEPT, an inter-residue (^{1}H→^{15}N→^{13}CO) coherence transfer experiment (section A.2.4). The [^{15}N, U-^{13}C]Ala elastin was solvated in 8 M urea, a denaturant. DP and two-step rINEPT experiments were performed at room temperature (20-21 °C) to avoid the formation of cyanates [53], which may alter elastin's molecular structure.

Samples of hydrated (**Figure 3.17a**) and denatured elastin (**Figure 3.17b-c**) were used to identify the ^{13}CO chemical shift of random coil alanines. The ^{13}CO-Ala lineshape of hydrated elastin at 20 °C (**Figure 3.17a**) is relatively consistent with that at 37 °C (section C, **Figure 3.13a**); minor differences in ^{13}CO chemical shifts are attributed to temperature effects. The ^{13}CO signal at 20 °C has a narrow lineshape (~145 Hz) with two features, 176.1 and 175.3 ppm, and a downfield shoulder at ~177 ppm. The deconvolution subroutine shows three L/G components (*green*) corresponding to random coil alanines, which arise from the *neighboring residue* effect and represents the R$_{-1}$-**Ala**-Gly, R$_{-1}$-**Ala**-Z (Z≠Ala, Gly), and R$_{-1}$-**Ala**-Ala motifs in the primary structure. The ^{13}CO-Ala signal for the denatured elastin is characterized by two narrower lineshapes at 176.0 ppm (~40 Hz) and 175.5 ppm (~48 Hz) (**Figure 3.17b**). This sample shows no downfield ^{13}CO signals corresponding to alanines in α-helical domains (**Appendix 8**). Moreover, the lineshape for denatured elastin is adequately reproduced by three L/G components (*green*) centered at 176.0, 175.5, and 175.3 ppm, which reflects the *neighboring residue* effect. Hence, this result suggests that random coil alanines are present in the previously described hydrated sample.

The two-step rINEPT selectively detects ^{13}CO-**Ala** signals corresponding to the R$_{-1}$-**Ala**-Ala motif in (denatured) NRSMC elastin in urea (**Figure 3.17c**). The spectrum is characterized by a narrow peak (~50 Hz) at 175.3 ppm. The deconvolution subroutine indicates that one component at 175.3 ppm is adequate to reproduce the experimental lineshape. The major population at 175.3 ppm represents the signal for random coil alanines, which is consistent with the ^{13}CO DP spectrum with LOW-BASHD (**Figure 3.17b**). Therefore, the upfield signal in the ^{13}CO-Ala lineshape corresponds to the R$_{-1}$-**Ala**-Ala motif, which is relatively consistent with the previously described feature in hydrated elastin.

The two-step rINEPT experiment did not yield a ^{13}CO-**Ala** spectrum for hydrated NRSMC elastin. As described in section A.2.4, this experiment is primarily effective for samples that have high mobility and, correspondingly, long coherence lifetimes, T_2'. However, the T_2's for carbons in hydrated elastin are much shorter than those in urea, resulting in no detection of ^{13}C signals.

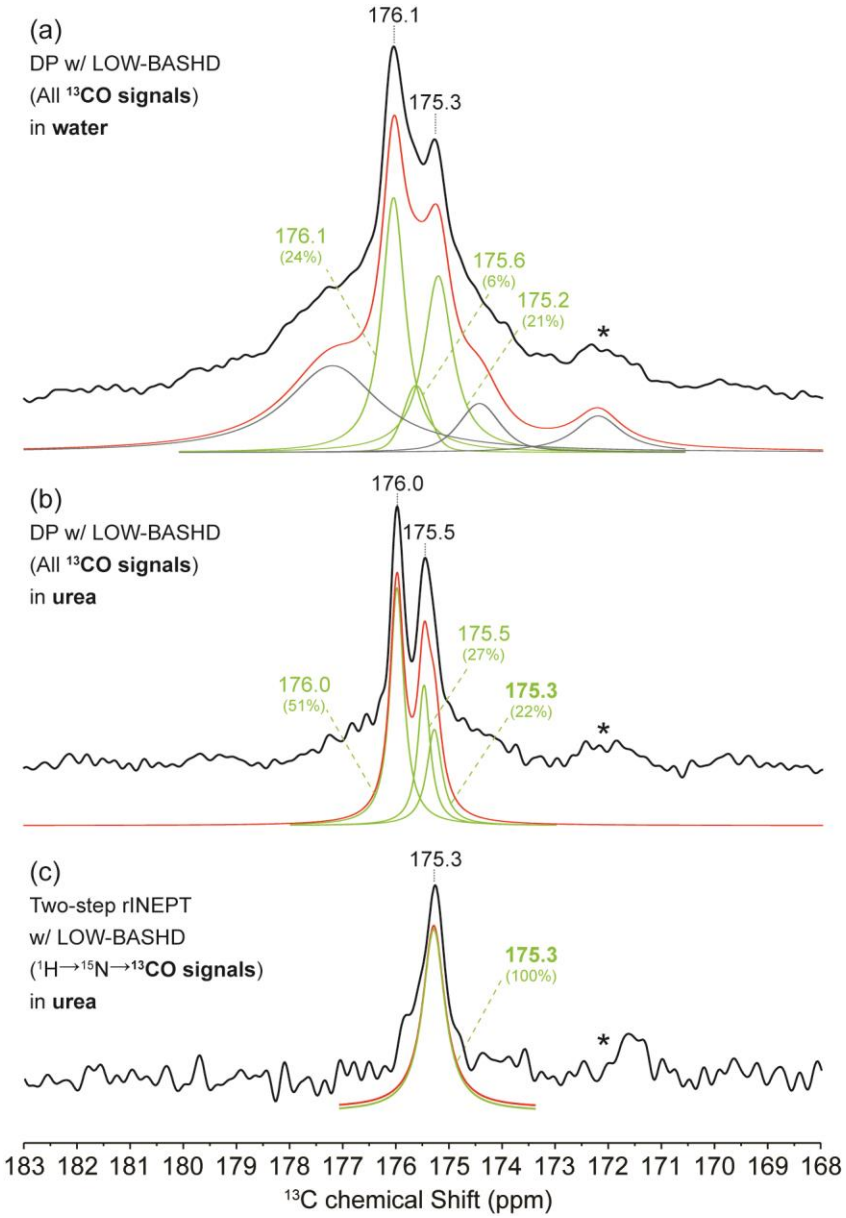


Figure 3.17 ^{13}CO-Ala lineshapes at 20 °C. DP with LOW-BASHD spectra of (a) hydrated and (b) urea-solvated samples; (c) two-step rINEPT spectrum of urea-solvated elastin. LOW-BASHD decouples $^{1}J_{COC\alpha}$. Two-step rINEPT selects ^{13}CO signals from **Ala**-Ala pairs in [^{15}N, U-^{13}C]Ala NRSMC elastin (section A.2.4). Dotted lines indicate tallest points of resolved peaks. Deconvolution subroutine is performed with Lorentzian-Gaussian (L/G) lines. Six-component lines are used for (a), based on the result for ^{13}CO-Ala of elastin in water at 37 °C (section C.1). Random coil components are represented by ***green*** L/G lines, and other contributions are shown in ***gray***. The sum of L/G lines is shown in ***red***. Asterisks (*) indicate natural-abundance ^{13}CO-Gly signals.

C.3 Random coil population of alanines has a different mobility from α-helices

Based on isotropic ^{13}C-Ala chemical shifts, two discrete populations are identified, α-helix and random coil. The former is characterized by a set of torsion angles (ϕ, ψ = -57°, -45°) with backbone hydrogen-bond stabilization. The latter represents rapidly interconverting structures (including α-helix and β-sheet) that sample a broad range of torsion angles [54-58].

The conformational assignment of alanines in hydrated NRSMC elastin at 37 °C is confirmed with mobility-based ssNMR experiments. The α-helix and random coil populations have different timescales of motion that are distinguishable by CP and rINEPT experiments. Regions with high mobility ($\tau_c < 10^{-9}$ s) are identified by rINEPT, whereas rigid domains ($\tau_c > 10^{-5}$ s) are observed by CP [59, 60] (section C.3.1). Furthermore, the MHz- and kHz-scale motions in the elastin domains are probed by ^{13}C T_1 and $T_{1\rho}$ measurements (section C.3.2).

C.3.1 Elastin domains are distinguished by their timescales of motion. Figure 3.19a-c illustrates ^{13}C DP, CP and rINEPT spectra of hydrated [U-^{13}C]Ala elastin at 37 °C. As described in section C.1, ^{13}C DP spectra reflect all carbon populations, regardless of conformation and dynamics. The tallest points in the ^{13}CO, ^{13}Cα, and ^{13}Cβ-Ala DP spectra at 175.6, 50.5, and 17.4 ppm, respectively, correspond to the random coil population. In contrast, the respective shoulder features at 177, ~51, and 16 ppm arise from the α-helical domains (**Figure 3.18a**). The ^{13}C DP lineshapes also reflect the sum rINEPT and CP spectra (**Appendix 9**).

The ^{13}C refocused INEPT spectrum reflects the mobile regions in elastin (**Figure 3.18b**). ^{13}Cα- and ^{13}Cβ-Ala peaks are observed by rINEPT; no ^{13}CO-Ala signal was detected due to the absence of directly bonded protons. The ^{13}Cα peak has a narrow lineshape (~170 Hz) at 50.5 ppm, and it is adequately reproduced by a single L/G component in the deconvolution. The ^{13}Cβ linewidth is narrower (~90 Hz), and the tallest intensity at 17.4 ppm corresponds to the random coil population. The deconvolution subroutine indicates two populations, i.e., the major component is centered at 17.4 ppm (87%), and the minor one is at 16.6 ppm (~13%). The signals at 50.5 ppm (^{13}Cα) and 17.4 ppm (^{13}Cβ) correspond to random coil, as previously assigned (**Table 3.4**). This result supports the tentative assignments of alanine populations using ^{13}C isotropic chemical shifts, described previously in sections C.1-C.2. Additionally, the spectra reflect two mobile populations of alanines, which have motional timescales observable by rINEPT ($\tau_c < 10^{-9}$ s) [59, 60]. The random coil component presumably exhibits large-amplitude fluctuations that are similar to the structural interconversions of *statistical coil* [54-58].

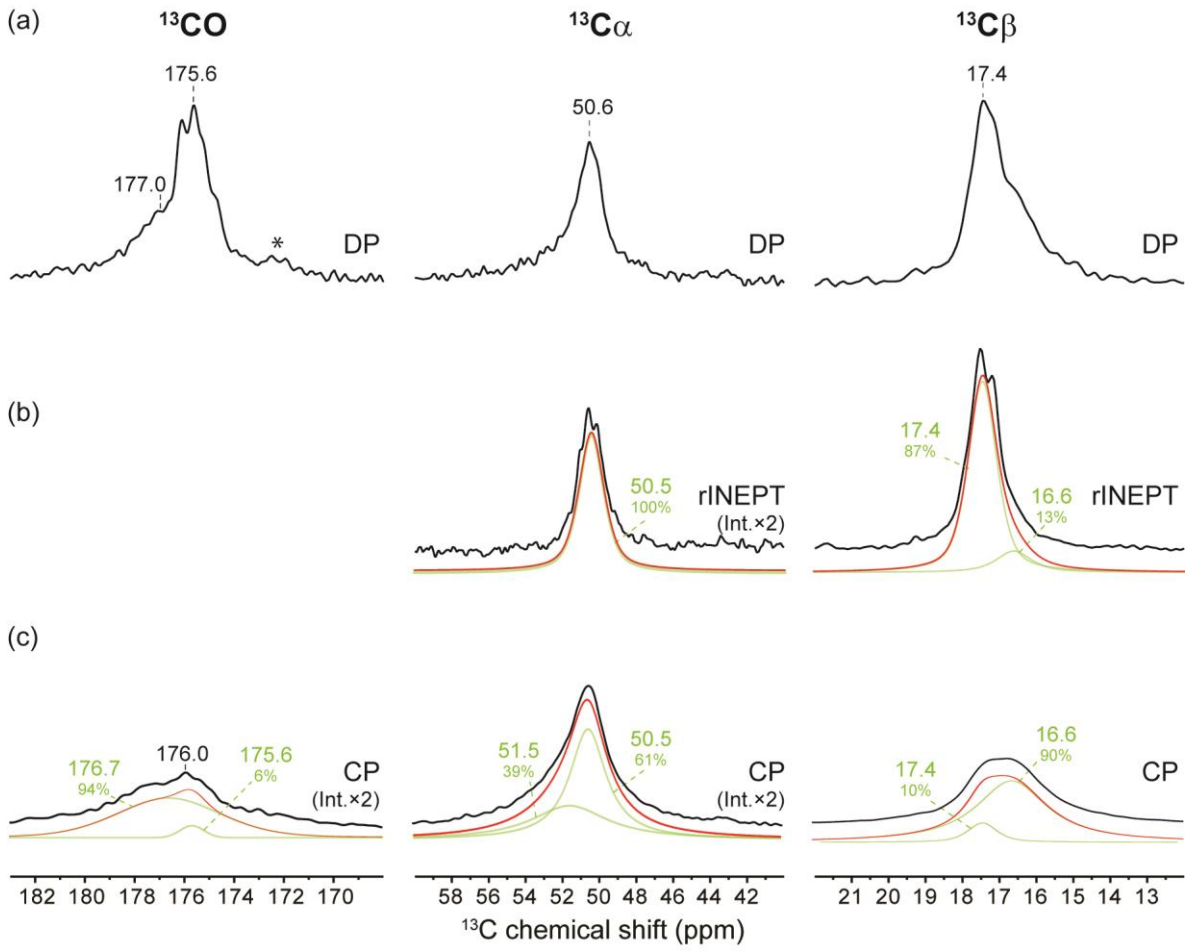


Figure 3.18 ^{13}CO, ^{13}Cα, and ^{13}Cβ-Ala signals acquired by (a) DP, (b) rINEPT, and (c) DP spectra of hydrated [U-^{13}C]Ala elastin at 37 °C. MAS is applied during the experiment. In (a), Dashed lines (┆) indicate the tallest peaks in the ^{13}C DPMAS spectrum, which is reproduced from **Figure 3.12**. Asterisk (*) represents the natural-abundance ^{13}CO-Gly signal. In (b) and (c), Lorentzian-Gaussian (L/G) components (**green**) are used in the deconvolution subroutine. The center-of-mass of each component is fixed to the ^{13}C-Ala chemical shift corresponding to random coil or α-helix, listed on **Table 3.4** (section C.2). The sum of L/G shapes (**red**) are offset vertically from the experimental result (**black**) for clarity. No rINEPT signal is detected in the ^{13}CO region due to the absence of directly bonded protons. The ^{13}C DP, rINEPT and CP spectra were acquired with 256, 1024 and 6144 transients, respectively. Each signal was processed with 10 Hz of line broadening, and the height of each peak is scaled to the maximum intensity of the DP spectrum.

The ^{13}C CP spectrum reflects the relatively rigid regions in elastin (**Figure 3.18c**). The absolute CP intensities are significantly lower than the rINEPT and DP spectra, due to reduced dipolar couplings in the hydrated sample. The carbonyl lineshape is broad (~400 Hz), and its tallest point is observed at 176.0 ppm. The simplest deconvolution of the ^{13}CO lineshape features a major population (94%) at 176.7 ppm and a minor one (6%) at 175.6 ppm. The asymmetric ^{13}Cα lineshape has its tallest intensity at 50.5 ppm, and it is composed of two L/G components in the deconvolution, 51.5 ppm (39%) and 50.5 ppm (61%).

These two populations are relatively consistent with those identified in the 2D correlation spectra (section C.2). The ^{13}Cβ region shows a broad feature with intensities over 14-19 ppm; its center-of-mass is roughly at ~17 ppm. The deconvolution of this lineshape yields two populations, centered at 16.5 ppm (90%) and 17.4 ppm (10%). These results reflect the presence of two relatively rigid populations of alanines, which have motional timescales that are identifiable by CP ($\tau_c > 10^{-5}$ s) [59, 60]. The major component at 176.6, 51.1, and 16.5 ppm corresponds to alanines in α-helices, which presumably exhibit low-amplitude fluctuations such as helix-bending and -breathing [61]. In contrast, the minor population at 17.4, 50.4, and 175.7 ppm presumably corresponds to terminal residues that are unfolded.

C.3.2 Alanines in α-helical domains have different kHz-dynamics from the random coil population. NMR relaxation measurements assess protein dynamics over a wide range of motional timescales, 10^{-9}-10^{-1} s [62, 63]. Spin-lattice relaxation (T_1) and spin-lattice relaxation in the rotating frame ($T_{1\rho}$) are effective for probing segmental dynamics in elastin and elastin mimetics [16-19, 28, 64]. The ^{13}C T_1 measurement is sensitive to molecular motions in the nano- to picosecond timescale, such as sidechain rotation or loop reorientation in proteins [65, 66]. In contrast, the ^{13}C $T_{1\rho}$ reflects chain dynamics in the micro- to millisecond regimes [67], which often describe domain motions in polymers and the folding/unfolding of secondary structures in proteins [62, 63].

^{13}C T_1 measurements investigate fast motions in the MHz timescale. Larger values correspond to regions with higher mobility, whereas smaller ^{13}C T_1's reflect segments with low-amplitude fluctuations. Negligible differences in ^{13}C T_1's are observed between the α-helical and random coil alanines (**Table 3.5**). The α-helical population indicated by ^{13}CO-, ^{13}Cα-, and ^{13}Cβ-Ala resonances at 176.7, 51.5, and 16.6 ppm has ^{13}C T_1's of 1.1, 0.7, and 0.8 s, respectively. These values are similar to those for random coil alanines at 175.0, 50.5, and 17.4 ppm, i.e., 1.1, 0.6, and 0.6 s. In general, little variation in ^{13}C T_1's is observed over the entire ranges of ^{13}CO-, ^{13}Cα-, and ^{13}Cβ-Ala chemical shifts (**Figure 3.19**). The ^{13}CO T_1's range from 0.9 to 1.2 s between 174.5 and 178 ppm. The ^{13}Cα T_1's range from 0.6 to 0.8 s for signals at 50-52 ppm, and the ^{13}Cβ T_1's range from 0.6 to 1.0 s for intensities at 16-18 ppm. These results suggest that the MHz-scale motions of alanines found in random coil and α-helical regions are indistinguishable.

Table 3.5. T_1 and $T_{1\rho}$ values for ^{13}CO, ^{13}Cα, and ^{13}Cβ regions corresponding to random coil and α-helical alanines in elastin. The three ^{13}CO chemical shifts for random coil Ala reflect the *neighboring residue* effect described in section C.2. ^{13}Cβ and ^{13}Cα-Ala resonances corresponding to α-helix and random coil are derived from 2D ^{13}C-^{13}C correlation spectra, previously shown in Table 3.4. The uncertainties represent the standard error of regression for a given ^{13}C chemical shift.

Conformations	Domains	^{13}C Sites	δ (ppm)	^{13}C T_1 (s)	^{13}C $T_{1\rho}$ (ms)
random coil	Hydrophobic	CO	175.9	1.1 ±0.1	9.7 ±1.2
	Hydrophobic	CO	175.3	1.1 ±0.1	11.4 ±0.9
	Crosslinking	CO	175.0	1.1 ±0.2	10.2 ±1.5
	Hydrophobic	Cα	50.5	0.6 ±0.1	4.7 ±1.3
	Hydrophobic	Cβ	17.4	0.6 ±0.1	12.9 ±3.0
α-helix	Crosslinking	CO	176.7	1.1 ±0.1	3.6 ±0.4
	Crosslinking	Cα	51.5	0.7 ±0.1	1.8 ±0.5
	Crosslinking	Cβ	16.6	0.8 ±0.1	4.4 ±1.3

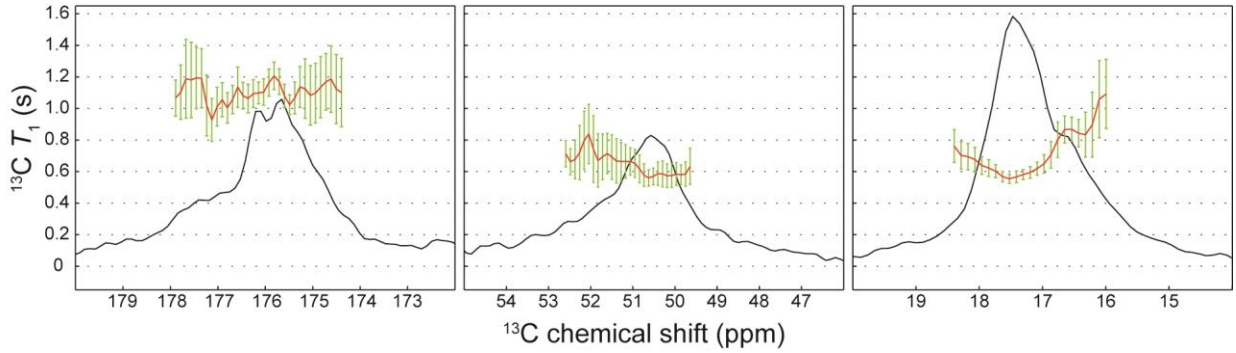


Figure 3.19 ^{13}C T_1's for alanines in α-helical and random coil populations of elastin. ***Black*** represents ^{13}CO-, ^{13}Cα- and ^{13}Cβ-Ala DP lineshapes at 37 °C. **Red** represents the rate constant $T_{1\rho}$ for a given ^{13}C chemical shift in each lineshape, which was fitted using a single exponential function. **Green** represents the standard error of the regression, which is the average distance that a given T_1 value deviates from the regression line; the length of a given error bar corresponds to the 95% confidence interval.

^{13}C $T_{1\rho}$ values reflect the differences in domain mobility. The high values generally represent faster kHz-scale dynamics, whereas low ^{13}C $T_{1\rho}$'s correspond to slower motions. In general, a wide range of ^{13}C $T_{1\rho}$ values, ~2 to ~12 ms, is observed for each region (**Figure 3.20**). The ^{13}C $T_{1\rho}$'s for the random coil population differ from those for the α-helix. Alanines in random coil found at 175.0 and 175.9 ppm have ^{13}CO $T_{1\rho}$'s of 10.2 ±1.5 and 9.7 ±1.2 ms, respectively, indicating fast domain motions. However, lower ^{13}C $T_{1\rho}$ values are observed in the α-helical region at 176.7 ppm, i.e., 3.6 ±0.4 ms, indicating slower kHz-scale motions. The ^{13}Cα $T_{1\rho}$'s for random coil at 50.5 ppm and α-helix at 51.5 ppm are 4.7 ±1.3 ms and 1.8 ±0.5 ms, respectively. In addition, the coil population at 17.4 ppm has ^{13}Cβ $T_{1\rho}$ of 12.9 ±3.0 ms, which is significantly higher than the α-helix (4.4 ±1.3 ms). These values are summarized in **Table 3.5**.

These results indicate that alanines in random coil have faster large-amplitude motions relative to those in α-helical domains.

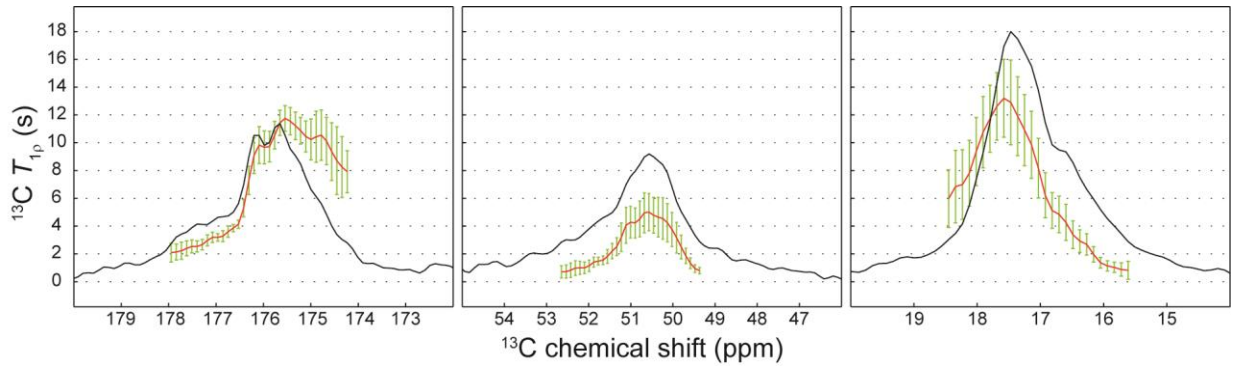


Figure 3.20 ^{13}C $T_{1\rho}$'s for alanines in α-helical and random coil populations of elastin. ***Black*** represents ^{13}CO-, ^{13}Cα- and ^{13}Cβ-Ala DP lineshapes at 37 °C. **Red** represents the rate constant $T_{1\rho}$ for a given ^{13}C chemical shift in each lineshape, which was fitted using a single exponential function. **Green** represents the standard error of the regression, which is the average distance that a given T_1 value deviates from the regression line; the length of a given error bar corresponds to the 95% confidence interval.

The conformational assignment of alanines in NRSMC elastin is affirmed by the results of selective detection and NMR relaxation measurements. Alanines in the crosslinking domains are found in both random coil and α-helical regions, whereas those in the hydrophobic domains are observed mostly in the random coil. The relatively rigid α-helical domains have low ^{13}C $T_{1\rho}$'s and are observed by CP. In contrast, the random coil population has high ^{13}C $T_{1\rho}$ values and is selectively detected by rINEPT. The results indicate that alanines in the mobile region undergo large-amplitude fluctuations, whereas those in the relatively rigid domains undergo slower motions.

D. SUMMARY AND CONCLUSIONS

The α-helix is known to provide resilience and mechanical stiffness against deformations, stresses, and strains in proteins [68, 69] . In the early stages of elastin assembly, α-helices are formed in the crosslinking domains of tropoelastin to facilitate the spatial alignment of lysines prior to crosslinking [70, 71]. In mature elastin, the abundance of Lys-derived crosslinks in the Ala-rich domains gives rise to the polymer's resilience. However, recent reports show that the α-helical content in elastin peptides and mimetics is less than expected [50], which gives rise to many speculations regarding the molecular structure of the protein.

Solid-state NMR spectroscopy was used to characterize hydrated NRSMC elastin at 37 °C. The [U-^{13}C]Ala elastin provides insights into structure and dynamics of the alanines, which are found abundantly in the hydrophobic and crosslinking domains. Secondary structures were identified based on isotropic ^{13}C-Ala chemical shifts, obtained via one- and two-dimensional ssNMR experiments. Selective detection experiments and measurements of NMR relaxation parameters probe domain mobility in the protein.

Tentative assignments of alanine populations were made using one-dimensional ssNMR measurements. The asymmetric ^{13}C-Ala lineshapes suggests the presence of at least two populations of alanines in hydrated NRSMC elastin, α-helix and random coil.

Support for multiple conformations of alanines was also obtained from the 2D homonuclear (^{13}C-^{13}C) correlation spectra. Alanines in the hydrophobic domains are primarily found in random coil, whereas those in the crosslinking regions are characterized by α-helix and random coil. The existence of random coil alanines in the α-helical crosslinking domains was confirmed by the sequence-dependence of ^{13}CO-Ala chemical shifts. The *neighboring residue* effect is calculated using a semi-empirical prediction, and the result indicates that random coil alanines in rat tropoleastin give rise to three ^{13}CO-**Ala** signals corresponding to the R$_{-1}$-**Ala**-Gly, R$_{-1}$-**Ala**-Z (Z ≠ Ala, Gly), and R$_{-1}$-**Ala**-Ala motifs. The predicted ^{13}CO-Ala chemical shifts are found in good agreement with those observed in the 2D correlation spectrum. However, the peak intensity corresponding to the R$_{-1}$-**Ala**-Ala in the experimental spectrum was significantly lower than expected, suggesting that only a small portion of Ala in the crosslinking domains is found in random coil. This result was further confirmed by an inter-residue coherence transfer experiment, i.e., two-step rINEPT, which selectively detected the ^{13}CO-**Ala** signal corresponding to the R$_{-1}$-**Ala**-Ala sequence found in the crosslinking domains of NRSMC elastin.

Selective detection experiments and measurements of NMR relaxation parameters provide insights into the dynamics of alanines in elastin. The population of α-helical alanines was observed primarily by the CP experiment, which is sensitive to slower motions with $\tau_c > 10^{-5}$ s [59, 60]. This population also has the lowest ^{13}C $T_{1\rho}$ values among other regions, indicating slow kHz-scale motions such as helix-breathing and -bending [61] are exhibited in these segments. In contrast, the random coil population of alanines was selectively detected by rINEPT, which is sensitive to faster motions ($\tau_c < 10^{-9}$ s) [59, 72] . This coil population also has significantly higher ^{13}C $T_{1\rho}$ values than the α-helical domains, which supports the rapid interconversions of structural motifs exhibited by random coil, or *statistical coil* [54-58]. Such large-amplitude fluctuations are also found in the flexible loop regions of globular proteins [73, 74], and the large-amplitude motions in NRSMC elastin are consistent with elastin mimetic studies [13, 14, 75].

In conclusion, ssNMR studies provide insights into the structure and dynamics in hydrated NRSMC elastin at 37 °C. Multiple conformations are identified for alanines in hydrophobic and crosslinking domains. The former is predominantly random coil, whereas the latter is characterized by α-helices and random coil. The α-helical domains are rich Lys-derived crosslinks, which give rise to elastin's resilience. The random coil, or better called *statistical coil* [54-58], provides a framework for rapid structural interconversions that may occur during stretch-recoil cycles.

E. REFERENCES

1. Saito, H., *Conformation-dependent ^{13}C chemical shifts: a new means of conformational characterization as obtained by high-resolution solid-state ^{13}C NMR.* Magnetic Resonance in Chemistry, 1986. **24**(10): p. 835-852.
2. Wishart, D.S., C.G. Bigam, J. Yao, F. Abildgaard, H.J. Dyson, E. Oldfield, J.L. Markley, and B.D. Sykes, *^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR.* Journal of Biomolecular NMR, 1995. **6**(2): p. 135-140.
3. Wishart, D.S., and B.D. Sykes, *The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data.* Journal of Biomolecular NMR, 1994. **4**(2): p. 171-180.
4. Ha, S.-W., T. Asakura, and R. Kishore, *Distinctive influence of two hexafluoro solvents on the structural stabilization of Bombyx mori silk fibroin protein and its derived peptides: ^{13}C NMR and CD studies.* Biomacromolecules, 2006. **7**(1): p. 18-23.
5. Asakura, T., J. Ashida, T. Yamane, T. Kameda, Y. Nakazawa, K. Ohgo, and K. Komatsu, *A repeated β-turn structure in poly(Ala-Gly) as a model for silk I of Bombyx mori silk fibroin studied with two-dimensional spin-diffusion NMR under off magic angle spinning and rotational echo double resonance.* Journal of Molecular Biology, 2001. **306**(2): p. 291-305.
6. Ashida, J., K. Ohgo, K. Komatsu, A. Kubota, and T. Asakura, *Determination of the torsion angles of alanine and glycine residues of model compounds of spider silk $(AGG)_{10}$ using solid-state NMR methods.* Journal of Biomolecular NMR, 2003. **25**(2): p. 91-103.
7. Wang, Y., *Secondary structural effects on protein NMR chemical shifts.* Journal of Biomolecular NMR, 2004. **30**(3): p. 233-244.
8. Wang, Y., and O. Jardetzky, *Investigation of the neighboring residue effects on protein chemical shifts.* Journal of the American Chemical Society, 2002. **124**(47): p. 14075-14084.
9. Ohgo, K., W.P. Niemczura, B.C. Seacat, S.G. Wise, A.S. Weiss, and K.K. Kumashiro, *Resolving nitrogen-15 and proton chemical shifts for mobile segments of elastin with two-dimensional NMR spectroscopy.* Journal of Biological Chemistry, 2012. **287**(22): p. 18201-18209.
10. He, D., M. Chung, E. Chan, T. Alleyne, K.C. Ha, M. Miao, R.J. Stahl, F.W. Keeley, and J. Parkinson, *Comparative genomics of elastin: sequence analysis of a highly repetitive protein.* Matrix Biology, 2007. **26**(7): p. 524-40.
11. Tamburro, A.M., A. Pepe, and B. Bochicchio, *Localizing α-helices in human tropoelastin: assembly of the elastin "puzzle".* Biochemistry, 2006. **45**(31): p. 9518-9530.

12. Martino, M., A. Bavoso, V. Guantieri, A. Coviello, and A.M. Tamburro, *On the occurrence of polyproline II structure in elastin.* Journal of Molecular Structure, 2000. **519**(1–3): p. 173-189.
13. Tamburro, A.M., B. Bochicchio, and A. Pepe, *Dissection of human tropoelastin: exon-by-exon chemical synthesis and related conformational studies.* Biochemistry, 2003. **42**(45): p. 13347-13362.
14. Tamburro, A.M., B. Bochicchio, and A. Pepe, *The dissection of human tropoelastin: from the molecular structure to the self-assembly to the elasticity mechanism.* Pathologie Biologie, 2005. **53**(7): p. 383-389.
15. Narayanan, A.S., R.C. Page, F. Kuzan, and C.G. Cooper, *Elastin cross-linking in vitro. Studies on factors influencing the formation of desmosines by lysyl oxidase action on tropoelastin.* Biochemical Journal, 1978. **173**(3): p. 857-862.
16. Kumashiro, K.K., *Solid-state NMR studies of elastin and elastin peptides*, in *Modern Magnetic Resonance*, G.A. Webb, Editor. 2006, Springer Netherlands: Dordrecht. p. 93-99.
17. Kumashiro, K.K., J.P. Ho, W.P. Niemczura, and F.W. Keeley, *Cooperativity between the hydrophobic and cross-linking domains of elastin.* Journal of Biological Chemistry, 2006. **281**(33): p. 23757-23765.
18. Perry, A., M.P. Stypa, J.A. Foster, and K.K. Kumashiro, *Observation of the glycines in elastin using ^{13}C and ^{15}N solid-state NMR spectroscopy and isotopic labeling.* Journal of the American Chemical Society, 2002. **124**(24): p. 6832-6833.
19. Perry, A., M.P. Stypa, B.K. Tenn, and K.K. Kumashiro, *Solid-state ^{13}C NMR reveals effects of temperature and hydration on elastin.* Biophysical Journal, 2002. **82**(2): p. 1086-1095.
20. Bennett, A.E., C.M. Rienstra, M. Auger, K.V. Lakshmi, and R.G. Griffin, *Heteronuclear decoupling in rotating solids.* Journal of Chemical Physics, 1995. **103**(16): p. 6951-6958.
21. Struppe, J.O., C. Yang, Y. Wang, R.V. Hernandez, L.M. Shamansky, and L.J. Mueller, *Long-observation-window band-selective homonuclear decoupling: increased sensitivity and resolution in solid-state NMR spectroscopy of proteins.* Journal of magnetic resonance (San Diego, Calif. : 1997), 2013. **236**: p. 89-94.
22. McCoy, M.A., and L. Mueller, *Selective shaped pulse decoupling in NMR: homonuclear [carbon-13]carbonyl decoupling.* Journal of the American Chemical Society, 1992. **114**(6): p. 2108-2112.
23. Pines, A., M.G. Gibby, and J.S. Waugh, *Proton-enhanced NMR of dilute spins in solids.* Journal of Chemical Physics, 1973. **59**(2): p. 569-590.
24. Schaefer, J., E.O. Stejskal, and R. Buchdahl, *High-resolution carbon-13 nuclear magnetic resonance study of some solid, glassy polymers.* Macromolecules, 1975. **8**(3): p. 291-296.
25. Hartmann, S.R., and E.L. Hahn, *Nuclear double resonance in the rotating frame.* Physical Review, 1962. **128**(5): p. 2042-2053.

26. Morris, G.A., *Sensitivity enhancement in nitrogen-15 NMR: polarization transfer using the INEPT pulse sequence.* Journal of the American Chemical Society, 1980. **102**(1): p. 428-429.
27. Morris, G.A., and R. Freeman, *Enhancement of nuclear magnetic resonance signals by polarization transfer.* Journal of the American Chemical Society, 1979. **101**(3): p. 760-762.
28. Kumashiro, K.K., M.S. Kim, S.E. Kaczmarek, L.B. Sandberg, and C.D. Boyd, *^{13}C cross-polarization/magic angle spinning NMR studies of α-elastin preparations show retention of overall structure and reduction of mobility with a decreased number of cross-links.* Biopolymers, 2001. **59**(4): p. 266-275.
29. Aue, W.P., E. Bartholdi, and R.R. Ernst, *Two‐dimensional spectroscopy. Application to nuclear magnetic resonance.* Journal of Chemical Physics, 1976. **64**(5): p. 2229-2246.
30. Bovey, F.A., and P.A. Mirau, *NMR of polymers*. 1996: Elsevier Science.
31. Mueller, L.J., D.W. Elliott, G.M. Leskowitz, J. Struppe, R.A. Olsen, K.-C. Kim, and C.A. Reed, *Uniform-sign cross-peak double-quantum-filtered correlation spectroscopy.* Journal of Magnetic Resonance, 2004. **168**(2): p. 327-335.
32. Chen, L., R.A. Olsen, D.W. Elliott, J.M. Boettcher, D.H. Zhou, C.M. Rienstra, and L.J. Mueller, *Constant-time through-bond ^{13}C correlation spectroscopy for assigning protein resonances with solid-state NMR spectroscopy.* Journal of the American Chemical Society, 2006. **128**(31): p. 9992-9993.
33. Keeler, J., *Understanding NMR spectroscopy*. 2011: Wiley.
34. Shaka, A.J., J. Keeler, T. Frenkiel, and R. Freeman, *An improved sequence for broadband decoupling: WALTZ-16.* Journal of Magnetic Resonance (1969), 1983. **52**(2): p. 335-338.
35. Levitt, M.H., R. Freeman, and T. Frenkiel, *Broadband heteronuclear decoupling.* Journal of Magnetic Resonance (1969), 1982. **47**(2): p. 328-330.
36. McDermott, A.E., and T. Polenova, *Solid state NMR studies of biopolymers*. 2012: Wiley.
37. Baldus, M., and B.H. Meier, *Total correlation spectroscopy in the solid state. The use of scalar couplings to determine the through-bond connectivity.* Journal of Magnetic Resonance, Series A, 1996. **121**(1): p. 65-69.
38. Hardy, E.H., R. Verel, and B.H. Meier, *Fast MAS total through-bond correlation spectroscopy.* Journal of Magnetic Resonance, 2001. **148**(2): p. 459-464.
39. Heindrichs, A.S.D., H. Geen, C. Giordani, and J.J. Titman, *Improved scalar shift correlation NMR spectroscopy in solids.* Chemical Physics Letters, 2001. **335**(1–2): p. 89-96.
40. Dabalos, C.L., *Characterization of the proline residues in elastin*. 2016, University of Hawaii at Manoa: Ann Arbor. p. 161.

41. Chan, J.C., and G. Brunklaus, *R sequences for the scalar-coupling mediated homonuclear correlation spectroscopy under fast magic-angle spinning.* Chemical physics letters, 2001. **349**(1): p. 104-112.
42. Kupce, E., J. Boyd, and I.D. Campbell, *Short selective pulses for biochemical applications.* Journal of Magnetic Resonance, Series B, 1995. **106**(3): p. 300-303.
43. Saito, H., R. Tabeta, A. Shoji, T. Ozaki, and I. Ando, *Conformational characterization of polypeptides in the solid state as viewed from the conformation-dependent carbon-13 chemical shifts determined by the carbon-13 cross polarization/magic angle spinning method: oligo(L-alanine), poly(L-alanine), copolymers of L- and D-alanines, and copolymers of L-alanine with N-methyl- or N-benzyl-L-alanine.* Macromolecules, 1983. **16**(7): p. 1050-1057.
44. Hong, M., D. Isailovic, R.A. McMillan, and V.P. Conticello, *Structure of an elastin-mimetic polypeptide by solid-state NMR chemical shift analysis.* Biopolymers, 2003. **70**(2): p. 158-168.
45. Zhang, H., S. Neal, and D. Wishart, *RefDB: a database of uniformly referenced protein chemical shifts.* Journal of Biomolecular NMR, 2003. **25**(3): p. 173-195.
46. Ulrich, E.L., H. Akutsu, J.F. Doreleijers, Y. Harano, Y.E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk, Z. Miller, E. Nakatani, C.F. Schulte, D.E. Tolmie, R. Kent Wenger, H. Yao, and J.L. Markley, *BioMagResBank.* Nucleic Acids Research, 2008. **36**(Database issue): p. D402-D408.
47. Neal, S., A.M. Nip, H. Zhang, and D.S. Wishart, *Rapid and accurate calculation of protein ^{1}H, ^{13}C and ^{15}N chemical shifts.* Journal of Biomolecular NMR, 2003. **26**(3): p. 215-240.
48. Garnier, J., J.-F. Gibrat, and B. Robson, *GOR method for predicting protein secondary structure from amino acid sequence*, in *Methods in Enzymology*. 1996, Academic Press. p. 540-553.
49. Jones, D.T., *Protein secondary structure prediction based on position-specific scoring matrices.* Journal of Molecular Biology, 1999. **292**(2): p. 195-202.
50. Miao, M., J.T. Cirulis, S. Lee, and F.W. Keeley, *Structural determinants of cross-linking and hydrophobic domains for self-assembly of elastin-like polypeptides.* Biochemistry, 2005. **44**(43): p. 14367-14375.
51. Schwarzinger, S., G.J.A. Kroon, T.R. Foss, J. Chung, P.E. Wright, and H.J. Dyson, *Sequence-dependent correction of random coil NMR chemical shifts.* Journal of the American Chemical Society, 2001. **123**(13): p. 2970-2978.
52. Schwarzinger, S., G.J. Kroon, T.R. Foss, P.E. Wright, and H.J. Dyson, *Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView.* Journal of Biomolecular NMR, 2000. **18**(1): p. 43-8.
53. Lippincott, J., and I. Apostol, *Carbamylation of cysteine: a potential artifact in peptide mapping of hemoglobins in the presence of urea.* Analytical Biochemistry, 1999. **267**(1): p. 57-64.

54. Smith, L.J., K.M. Fiebig, H. Schwalbe, and C.M. Dobson, *The concept of a random coil: residual structure in peptides and denatured proteins.* Folding and Design, 1996. **1**(5): p. R95-R106.
55. Jha, A.K., A. Colubri, K.F. Freed, and T.R. Sosnick, *Statistical coil model of the unfolded state: resolving the reconciliation problem.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(37): p. 13099-13104.
56. Vila, J., H. Baldoni, D. Ripoll, and H. Scheraga, *Unblocked statistical-coil tetrapeptides in aqueous solution: quantum-chemical computation of the carbon-13 NMR chemical shifts.* Journal of Biomolecular NMR, 2003. **26**(2): p. 113-130.
57. Toal, S., and R. Schweitzer-Stenner, *Local order in the unfolded state: conformational biases and nearest neighbor interactions.* Biomolecules, 2014. **4**(3): p. 725-773.
58. Vila, J., D. Ripoll, H. Baldoni, and H. Scheraga, *Unblocked statistical-coil tetrapeptides and pentapeptides in aqueous solution: A theoretical study.* Journal of Biomolecular NMR, 2002. **24**(3): p. 245-262.
59. Nowacka, A., N.A. Bongartz, O.H.S. Ollila, T. Nylander, and D. Topgaard, *Signal intensities in ^{1}H–^{13}C CP and INEPT MAS NMR of liquid crystals.* Journal of Magnetic Resonance, 2013. **230**: p. 165-175.
60. Nowacka, A., P.C. Mohr, J. Norrman, R.W. Martin, and D. Topgaard, *Polarization transfer solid-state NMR for studying surfactant phase behavior.* Langmuir, 2010. **26**(22): p. 16848-16856.
61. Itoh, K., and T. Shimanouchi, *Breathing vibration of poly-L-alanine α-helix.* Biopolymers, 1971. **10**(8): p. 1419-1420.
62. Krushelnitsky, A., and D. Reichert, *Solid-state NMR and protein dynamics.* Progress in Nuclear Magnetic Resonance Spectroscopy, 2005. **47**(1–2): p. 1-25.
63. Krushelnitsky, A., D. Reichert, and K. Saalwächter, *Solid-state NMR approaches to internal dynamics of proteins: from picoseconds to microseconds and seconds.* Accounts of Chemical Research, 2013. **46**(9): p. 2028-2036.
64. Kumashiro, K.K., T.L. Kurano, W.P. Niemczura, M. Martino, and A.M. Tamburro, *^{13}C CPMAS NMR studies of the elastin-like polypeptide (LGGVG)n.* Biopolymers, 2003. **70**(2): p. 221-226.
65. Kleckner, I.R., and M.P. Foster, *An introduction to NMR-based approaches for measuring protein dynamics.* Biochimica et biophysica acta, 2011. **1814**(8): p. 942-968.
66. Akke, M., and A.G. Palmer, *Monitoring macromolecular motions on microsecond to millisecond time scales by R1ρ−R1 constant relaxation time NMR spectroscopy.* Journal of the American Chemical Society, 1996. **118**(4): p. 911-912.
67. Schaefer, J., E.O. Stejskal, and R. Buchdahl, *Magic-angle ^{13}C NMR analysis of motion in solid glassy polymers.* Macromolecules, 1977. **10**(2): p. 384-405.

68. Buehler, M.J., and S. Keten, *Elasticity, strength and resilience: a comparative study on mechanical signatures of α-helix, β-sheet and tropocollagen domains.* Nano Research, 2008. **1**(1): p. 63-71.
69. Shahbazi, Z., *Mechanical model of hydrogen bonds in protein molecules.* American Journal of Mechanical Engineering, 2015. **3**(2): p. 47-54.
70. Keeley, F.W., C.M. Bellingham, and K.A. Woodhouse, *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.* Philosophical Transactions of the Royal Society of London, series B: Biological Sciences, 2002. **357**(1418): p. 185-189.
71. Brown-Augsburger, P., C. Tisdale, T. Broekelmann, C. Sloan, and R.P. Mecham, *Identification of an elastin cross-linking domain that joins three peptide chains: possible role in nucleated assembly.* Journal of Biological Chemistry, 1995. **270**(30): p. 17778-17783.
72. Fischer, M.W.F., L. Zeng, Y. Pang, W. Hu, A. Majumdar, and E.R.P. Zuiderweg, *Experimental characterization of models for backbone picosecond dynamics in proteins. Quantification of NMR auto- and cross-correlation relaxation mechanisms involving different nuclei of the peptide plane.* Journal of the American Chemical Society, 1997. **119**(51): p. 12629-12642.
73. Krieger, F., A. Moglich, and T. Kiefhaber, *Effect of proline and glycine residues on dynamics and barriers of loop formation in polypeptide chains.* Journal of the American Chemical Society, 2005. **127**(10): p. 3346-52.
74. Steinert, P.M., J.W. Mack, B.P. Korge, S.Q. Gan, S.R. Haynes, and A.C. Steven, *Glycine loops in proteins: their occurrence in certain intermediate filament chains, loricrins and single-stranded RNA binding proteins.* International Journal of Biological Macromolecules, 1991. **13**(3): p. 130-9.
75. Lelj, F., A.M. Tamburro, V. Villan, P. Grimaldi, and V. Guantieri, *Molecular dynamics study of the conformational behavior of a representative elastin building block: Boc-Gly-Val-Gly-Gly-Leu-Ome.* Biopolymers, 1992. **32**(2): p. 161-172.

CHAPTER 4. ELUCIDATION OF STRUCTURE AND DYNAMICS OF ELASTIN'S CROSSLINKING DOMAINS BY MOLECULAR DYNAMICS SIMULATIONS

A. INTRODUCTION

Solid-state NMR (SSNMR) studies provide insights into conformational and dynamical heterogeneity of alanines in hydrated NRSMC elastin (Chapter 3). In the crosslinking domains, alanines are found in both α-helix and random coil. However, the presence of random coil is unexpected in these Ala-rich regions, because these domains have high α-helical propensities. For this reason, all-atom molecular dynamics (MD) simulations are utilized to further investigate the secondary structures of alanines in the crosslinking domains.

MD simulations were performed on Ala-rich peptides that represent the crosslinking domains of native elastin. Computations were done on two peptides that are encoded by exons 19 and 25 in human tropoelastin, with the Lys residues modified into dehydrolysinonorleucine and allysine aldol to model the native crosslinks. Analyses of the MD trajectories provide insights into the heterogeneity of alanine conformations, the stability (or lack thereof) of α-helices, and the origin of random coil in the crosslinking domains.

A.1 MD simulations provide atomic-level detail on elastin's hydrophobic and crosslinking domains

All-atom MD simulations on the full-length crosslinked elastin (polymer) are not feasible due to the size of the system. In addition, the computation of a single monomer is hampered by the lack of a high-resolution tropoelastin structure. For these reasons, many *in silico* studies are conducted primarily on smaller systems, such as a single hydrophobic, or crosslinking, domain of tropoelastin [1].

Over the past few decades, elastin studies using MD simulations have been focused on the characterization of mimetic peptides with amino acid sequences from the hydrophobic domains, such as GVGGL [2], VPGVG [3, 4], and GVG(VPGVG) [1], which are traditionally believed to be the source of elasticity. In combination with experimental techniques, MD simulations have confirmed the presence of

secondary structures in these peptides such as PP-II, β-strands, α-turns type I, II, and VIII, as well as random coil [2, 5, 6]. In fact, a model that describes a dynamic interconversion of structures in elastin, or 'conformational equilibria', was proposed by these studies [7-9].

MD simulations have been useful for examining the mechanism for elasticity and the validity of elastin's molecular models. In particular, MD studies show the restoring force of stretched elastin mimetics, $(VPGVG)_{18}$, to be due to the increase of hydration entropy within the polypeptide's hydrophobic interior [10]. Furthermore, this entropic increase contributes to the hydrophobic collapse in elastin and elastin mimetics, resulting in the inverse-temperature transition, i.e., elastin's biochemical property of gaining order upon heating. The utility of MD in elastin research was further recognized with the rebuttal of the β-spiral model, once thought be responsible for flexibility [11]. The computational study was later affirmed by experimental measurements using CD, FTIR and NMR spectroscopy [12], and the model was refuted.

Although much attention has been focused on the molecular properties of elastin's hydrophobic domains, less has been reported on the crosslinking domains. Within the last ten years, only two publications demonstrated the use of computational techniques to study select crosslinking domains of human tropoelastin [13, 14]. One study reported the result of Monte Carlo calculations on the peptide structure of exon 36 in human tropoelastin [13]. This investigation provided insights into the role of the C-terminus, its interactions with other extracellular matrix proteins, and the presence of β-turns and random coil in the peptide. The other study reported the result of MD simulations on the role of a putative 'hinge' region in Ala-rich crosslinking domains encoded by exons 21 and 23 [14]. Specifically, this peptide contains a GVGTP sequence that facilitates the adoption of 'open' and 'closed' conformations.

The predominance of α-helices in the Ala-rich crosslinking domains of tropoelastin has recently been scrutinized, as this secondary structure is found less prevalent than predicted [5, 15]. Both CD and solution-state NMR studies found low α-helical content in the Ala-rich regions in tropoelastin in water [5, 16]. Furthermore, the α-helical structure was identified only in a mixture of TFE and water, a solvent that induces the formation of secondary structures [16]. In elastin-like peptides (ELP), the lack of helical conformation is typically accompanied by random coil [5, 7], consistent with the ssNMR results of Chapter 3. However, none of the previous reports [8, 16] have utilized all-atom MD simulations to investigate the lack of stability in the α-helical structures of elastin's crosslinking domains, until this study.

MD simulations are used to obtain biophysical insights into the structures and dynamics of elastin's crosslinking domains. To complement the ssNMR results of Chapter 3, possible secondary structures of alanines in these regions are investigated. The backbone mobility of alanines in α-helix and

random coil is investigated, and the role of Lys-derived crosslinks in the Ala-rich domains is examined. The results of this study may be useful for understanding the molecular structures of the crosslinking domains and their macromolecular organization in elastin.

A.2. Modeling the native environment of elastin's crosslinking domains

The crosslinking domains encoded by exons 19 and 25 of human tropoelastin (i.e., EX19 and EX25) are selected for the current MD study. These regions are believed to be critical for elastin assembly, because they contain a site for a putative inter-monomer junction, i.e., a region where one tropoelastin monomer is crosslinked to another in a *'head-to-tail'* fashion [17]. In addition, EX19 and EX25 have Lys-derived crosslinks[18], which as characteristics of native elastin and are essential for the polymer's resilience [19].

In native tissues, the EX19 is connected to EX25 by a tetrafunctional crosslink, desmosine, formed from the four Lys residues in the two regions [18, 20]. The biosynthesis of desmosine is preceded by a condensation reaction of two bifunctional crosslinks, allysine aldol (ALDL) and dehydrolysinonorleucine (DLNL), which are intermediates that are also prevalent in native tissues [20, 21]. The formation of a bifunctional crosslink occurs by oxidative deamination of Lys sidechains, followed by a spontaneous condensation reaction [22] (Chapter 1). Furthermore, such reactions occur in Lys residues that are spatially aligned in the α-helical structure [23].

A peptide containing bifunctional crosslink is a suitable candidate to represent elastin's crosslinking domains. EX19 domain with DLNL (i.e., EX19-DLNL) and EX25 domain with ALDL (i.e., EX25-ALDL) are appropriate models for MD investigations. Each of these peptides is relatively small, and, therefore, requires less computational time than a larger system, such as a desmosine crosslink that joins both EX19 and EX25 together.

The ALDL, or DLNL, crosslink is formed from two Lys that are interspersed either by two or by three other residues [18]. As shown in **Figure 4.1**, the DLNL in EX19 is formed by two Lys residues, K353 and K356, that are interspersed by two alanines, A354 and A355. In contrast, the ALDL in EX25 is formed by two Lys, K503 and K507, that are separated by a serine, S504, and two alanines, A504 and A505.

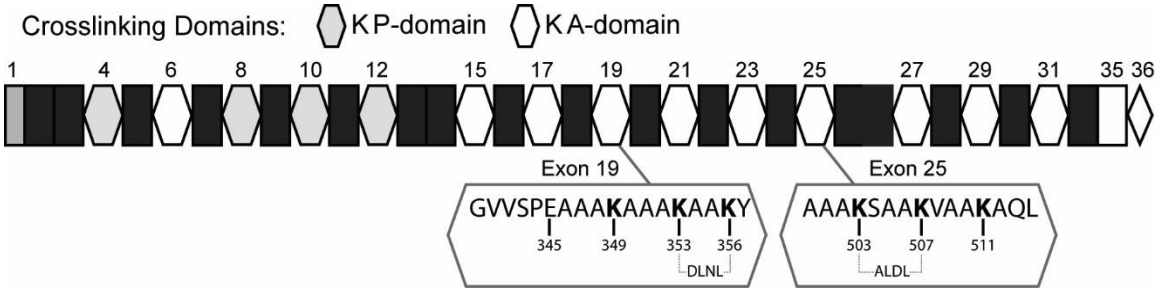


Figure 4.1 Representation of human tropoelastin sequence from Weiss and coworkers [24]. Gray and white hexagons represent the KP- and KA-type crosslinking domains, respectively. Black rectangles represent hydrophobic domains. Gray rectangle and white diamond represent the N- and C-term domains, respectively. Insets are amino acid sequences for EX19 and EX25, and the Lys residues involved in DLNL and ALDL crosslinking, respectively, are indicated in bold.

The molecular modeling study of EX19-DLNL and EX25-ALDL follows a multi-step process (**Figure 4.2**). The initial modeling of the crosslinked peptides requires two courses of action, i.e., force field parametrization of ALDL and DLNL crosslinks (section B.1) and construction of EX25-ALDL and EX19-DLNL peptides (section B.2). The force fields for ALDL and DLNL moieties must be created, because they are unavailable in common force field libraries like CHARMM [25, 26] and GROMACS [27]. Once the parametrization of crosslinks is completed, the EX19-DLNL and EX25-ALDL peptides are constructed and solvated for subsequent MD simulations.

The MD simulations of EX19-DLNL and EX25-ALDL are performed in several steps (section B.3). Briefly, each starting structure is first energy-minimized to remove the steric contacts among atoms. Then, a short period of MD is run to unfold the α-helical structure, followed by a longer simulation of peptide folding to allow for the (initial) α-helical conformation to be restored, or other secondary structures to be adopted. The trajectory structures in each production period of MD are analyzed (section B.4). The analysis was focused on the identification of secondary structures, stabilizing interactions in α-helices, and energetics of peptide folding.

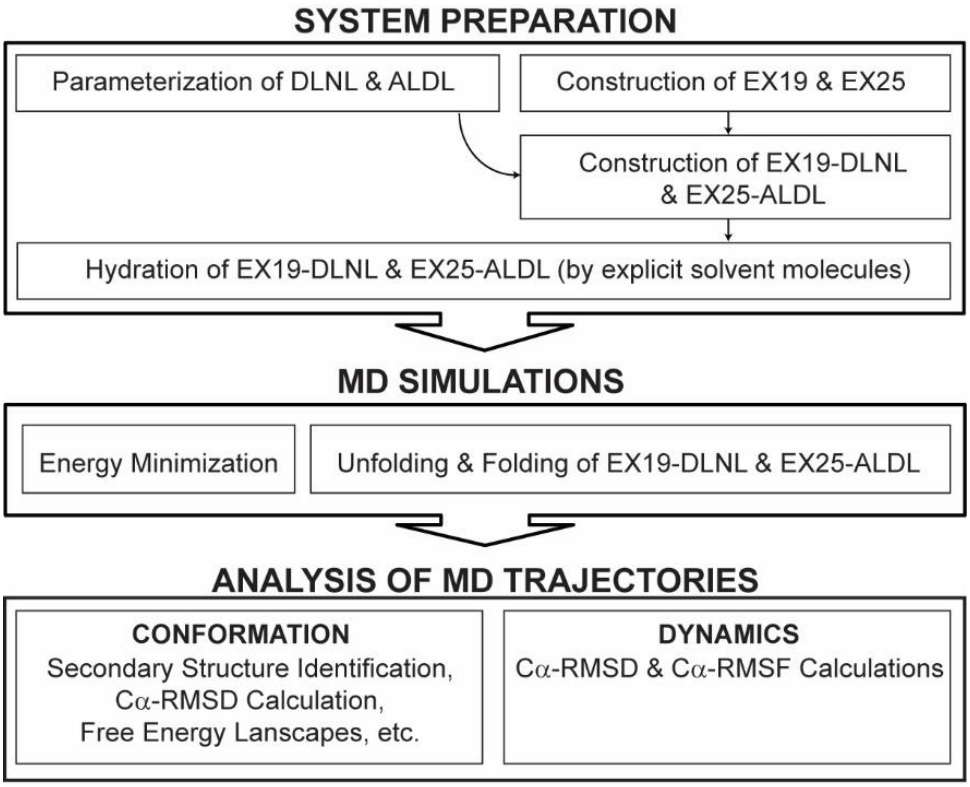


Figure 4.2 Step-by-step process of the molecular modeling of hydrated EX25-ALDL and EX19-DLNL. Arrows indicate the sequential order of action. System preparation involves a three-step process that is described in the methods sections B.1-B.2. MD simulations encompass minimization, unfolding, and folding periods (section B.3). Methods of analysis are focused on the conformational and dynamical study of the peptides (section B.4).

B. PARAMETRIZATION, SIMULATION AND ANALYSIS METHODS

B.1 Force field parametrization of ALDL and DLNL crosslinks

A molecular structure that is used for MD simulations is constructed, based on the geometrical constraints that are described in the topology and parameter files. The topology file contains information about partial charges and atom connectivity within a molecule. The parameter file includes all constants that are used to calculate the molecular interactions in bonded and non-bonded atoms, which include the equilibrium bond lengths and angles, bond and angle force constants, as well as dihedrals and improper force constants (Chapter 1). There were no previously existing such terms for ALDL and DLNL. Therefore, the development of topology and parameter files to describe ALDL and DLNL is crucial for the construction of crosslinked EX25 and EX19, respectively.

Force field parametrization of elastin's native bifunctional crosslinks is the first preparation step to take prior to MD simulations of EX25-ALDL and EX19-DLNL. Unlike any underivatized amino acid like Lys, these crosslinks contain functional groups that are not typically described in common force field libraries such as CHARMM [25, 26] and GROMACS [27]. Specifically, ALDL and DLNL contain aldehyde and Schiff base groups, respectively.

MD calculates atomic interactions in each crosslink as the combination of two chemical groups. Specifically, the ALDL is considered as a molecule containing methylene chains and an aldehyde group, whereas the DLNL is treated as one containing methylene groups and a Schiff base (**Figure 4.3**). The geometries and chemical properties of each of these components resemble those of other molecules with similar features. For instance, the rotation of dihedral angles in the methylene groups of ALDL is similar to that in Lys sidechains. Additionally, the C=O bond distance in the aldehyde group of ALDL should be similar to that of small molecules containing identical functional group, such as propionaldehyde.

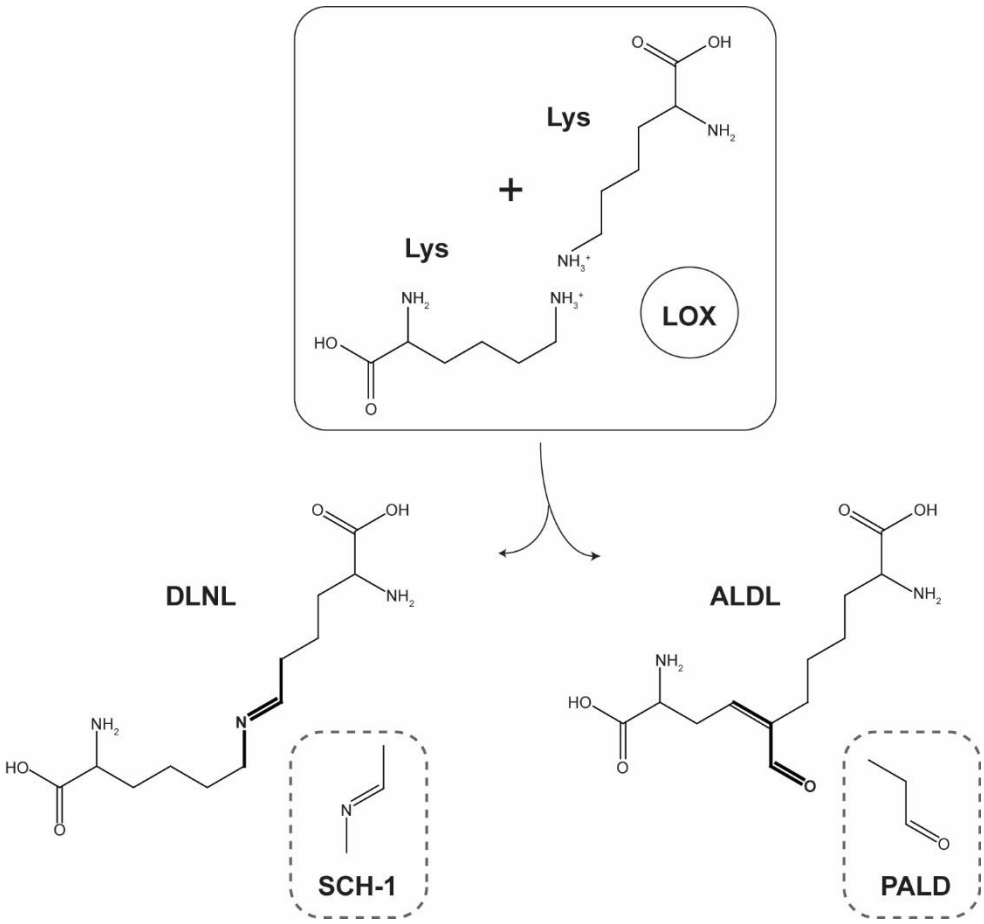


Figure 4.3 Lysine-derived bifunctional crosslinks in elastin and analogous molecules used to parametrize their functional groups. *Top*, chemical reaction describing the modification of two **Lys** to form a crosslink, as facilitated by the lysyl oxidase enzyme (LOX). *Bottom*, Topology and parameter file entries for **DLNL**'s and **ALDL**'s functional groups (in bold) are obtained from **SCH-1** and **PALD**, respectively, which are small molecules stored in the CHARMM GUI's library [25, 28, 29].

The force field development of Lys-derived crosslinks was achieved using the minimal approach. This strategy allows for the integration of two pre-existing topology and parameter files to describe each crosslink molecule [25]. File entries from propionaldehyde (PALD) are used in conjunction with those from Lys to create the ALDL crosslink (bottom of **Figure 4.3**). Analogously, parameters from the Schiff base model compound-1 (SCH-1) are combined with those from Lys to construct the DLNL. The topology and parameter files for both PALD and SCH-1 are obtained from the CHARMM GUI's library for small molecules [25, 28, 29], also included in **Appendix 10**.

B.2 Construction of starting geometries for EX19-DLNL and EX25-ALDL peptides

The construction of starting geometries for MD simulations of EX25-ALDL and EX19-DLNL involves three steps (**Figure 4.4**). First, fully α-helical peptides, EX25 and EX19, were created. Second, Lys sidechains were modified into ALDL and DLNL crosslinks, respectively, using the approach described above. Third, solvent molecules were included explicitly around each crosslinked peptide for subsequent MD simulations.

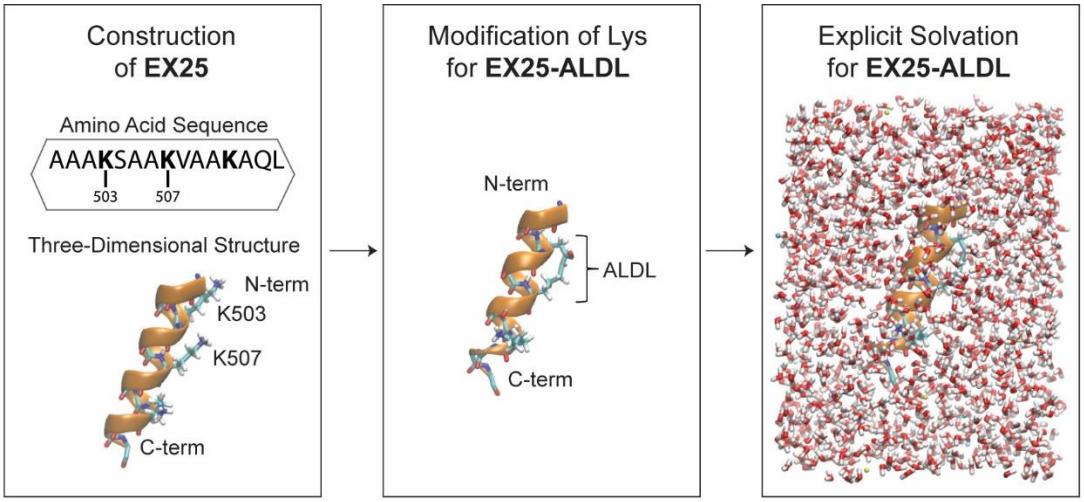


Figure 4. 4 Three stages of model preparation procedure for the MD simulation of representative EX25-ALDL. *Left*, the construction of EX25 is done using VMD's built-in *molefacture* plugin [30]. *Center*, Lys modification to form DLNL crosslink using the *PSFgen* module. *Right*, explicit solvation of EX25-ALDL is done using the *solvate* module. Ribbon representations are used to display EX25 and EX25-ALDL; licorice representation is used to display its backbone atoms, sidechains, and water molecules. Arrows (⟶) reflect the order of system preparation.

The construction of EX25-DLNL and EX19-ALDL was done using VMD modeling package [30]. The *molefacture* module [30] was used to generate the three-dimensional structures for EX25 and EX19, by the input of amino acid sequences (**Figure 4.4**, *left*). The initial geometry of each peptide was an α-helix, i.e., all torsion angles were set to the (ϕ, ψ) angles of (-57°, -47°). The N- and C-termini were acetylated and amidated, respectively. Once the molecular coordinates were set, the *PSFgen* module of VMD was used to create PDB and PSF files, which are necessary for the MD simulations.

Sidechain modification of lysines was conducted by incorporating the, newly parametrized, topology file entries for ALDL and DLNL into, respectively, EX25 and EX19 peptides. The *PSFgen* module was used for the addition of crosslink moieties into the initial α-helical structures (**Figure 4.4**, *center*). Geometrical constraints for the aldol and the Schiff base groups were added into EX25 and

EX19, respectively, via the PATCH option in the module. The PATCH feature allows for the atom connectivity of Lys sidechains to be reconfigured into ALDL or DLNL crosslinks. Once the modification of Lys was completed, a short energy-minimization period (1000 steps) was applied to relieve any steric clashes.

Explicit solvents for EX25-ALDL and EX19-DLNL were prepared using the *solvate* module of VMD (**Figure 4.4***, right*). An asymmetric solvent box containing TIP3P water [31] was built around each peptide with 10 Å of solvent padding. The *autoionize* plugin was used to neutralize the charged molecule at pH 7 with 0.15 M of Na^+ and Cl^- ions. The total numbers of atoms for each simulation of EX25-ALDL and EX19-DLNL were ~4300 and ~5400 atoms respectively.

B.3 MD Simulations of hydrated and crosslinked elastin peptides

All-atom MD simulations were performed using the NAMD 2.9 simulation package [32] on a MacPro desktop running a 2.26 GHz dual quad-core Intel processors. The CHARMM22 force field [33] was used, with additional topology and parameter files for the crosslinks.

The unfolding of α-helical structures in EX19-DLNL and EX25-ALDL prepared the system for subsequent conformational sampling by MD. Each peptide underwent a 2000 step energy minimization process. Then, 200 ps of interactive MD (IMD) simulations [34, 35] were performed to unfold the starting helical structures into an unstructured coil (**Figure 4.5**). Within this unfolding period, a force was applied on the carbonyl atom of residue A354 in EX19-DLNL, and that of residue A505 in EX25-ALDL, perpendicular to the helical axis. The procedure is done interactively on the VMD console, i.e., by click and drag of the cursor on the atom of interest. The applied force was constant for the first 100 ps, until the α-helical structure appeared to be completely unfolded. Then, the peptide was allowed to equilibrate for another 100 ps.

A 200 ns MD simulation was performed after the equilibration period. The isothermal-isobaric ensemble is used, in which a fixed number of atoms (4300 and 5300 for EX25-ALDL and EX19-DLNL, respectively), constant pressure (1 atm), and constant temperature (300 K) were used throughout the entire simulation. Each simulation was repeated five times for a total of 1 μs trajectory.

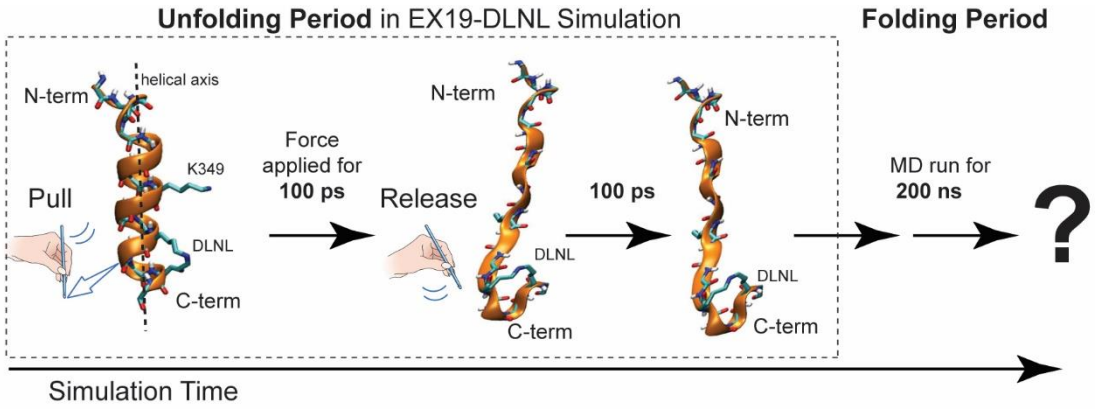


Figure 4.5 Representative unfolding and folding periods in MD simulations of EX19-DLNL. A few N-terminal residues deviate from the α-helical conformation after initial minimization. The unfolding of initial (minimized) structure is performed by interactive MD (IMD) simulations on the VMD console. Within the first 100 ps of unfolding period, a steering force (pull) is applied on the CO atom of residue, A354, perpendicular to the helical axis ( ); No forces (release) are applied during the second 100-ps period. To illustrate the unfolding of α-helix, the solvent molecules are not shown; the backbone atoms and Lys sidechains are shown without hydrogen atoms. The folding period is performed for 200 ns, in the background using NAMD.

The integration time steps for the bonded, short-range non-bonded, and long-range non-bonded (electrostatic) interactions were set to 2 fs, 1 fs, and 2 fs, respectively. A periodic boundary condition was implemented with a 10 Å cutoff, and the trajectory coordinates were recorded every 10 ps for analysis. Frequent inspection of the trajectories was done to ensure that each peptide was not interacting with its own neighboring copies. A sample configuration file that is used to run a NAMD simulation is shown in **Appendix 10**.

B.4 Analysis of MD trajectories for structural and dynamical characteristics

MD trajectories were analyzed using VMD's built-in modules [30] and tk/tcl scripts. Tcl scripts were obtained from the mailing list archive on the VMD-developer's webpage <http://www.ks.uiuc.edu/Research/vmd/mailing_list/>, and some of these scripts were modified for the analyses used in the current study. All script files (e.g., tcl, csh-shell, rtf) are included in **Appendix 11**. OriginPro (OriginLab Corporation, Northhampton, MA) and Microsoft Excel (Microsoft Corporation, Redmond, WA) were used for data sorting, processing and plotting.

In VMD, secondary structures are identified using the STRIDE algorithm [36, 37]. This algorithm defines common motifs such as α-helix or β-sheet based on the weighted contributions of backbone hydrogen-bond energies and torsion angles. In particular, if at least two hydrogen-bonds are formed

between the donor (NH) and acceptor (CO) atoms of paired residues (*i, i+4)* and (*i+1*, *i+5*), then the four central residues, *i+1, i+2, i+3, and i+4*, are considered to be in the α-helical conformation. The recognition pattern for other structural motifs, such as β-sheets, 3_{10}-helix, π-helix, and turns, are described in detail elsewhere [37-41].

Total populations of secondary structures (such as α-helix, 3_{10}-helix, turns and coil) in each peptide trajectory were quantified using STRUCTcount.tcl. Briefly, this script determines the frequency of a given secondary structure occurring at each of the peptide's amino acids, at a given trajectory time point. Two steps were involved in the quantification process. First, the structural motif of each residue, at a given trajectory time point, was determined using the STRIDE algorithm. Second, the number of occurrences of each motif was divided by the total number of configurations in the trajectory, expressed as percentage. For instance, if an alanine, A501, is found in an α-helix 396 times, over the total 1800 configurations, then the normalized population for this secondary structure is 22% (i.e., 396/1800*100%).

The 'timeline plots' were used to display structural information about a peptide as a function of simulation time. These plots were created using VMD's built-in *timeline* module. The output of this analysis was a postscript file, which was converted into a JPEG (image file) using Adobe illustrator (Adobe Systems, San Jose, CA). The *timeline* module was also used to quantify the number of hydrogen-bonding interactions in the peptide backbone, as a function of simulation time. To utilize this feature, angle and distance constraints that describe a hydrogen-bond interaction between the donor and the acceptor atoms were set to 40° and 3.5 Å, respectively. These constraints were used to find a list of atomic index-numbers describing peptide backbone that are involved in hydrogen-bonding, i.e., acceptor (O), hydrogen (H), and donor (N) atoms (HBdefineindex.rtf). The output of this procedure was then stored into a text file, and the data was sorted using a csh-shell script (PickHB.csh) and plotted using OriginPro.

The α-helical content in each peptide within a given MD trajectory was determined using percentahelix.tcl. The calculation for helical content involves several steps. First, the secondary structures adopted by each residue at a given trajectory time point were determined using the STRIDE algorithm. Second, the number of residues that were classified as an α-helix was quantified and expressed as percentages. For instance, if 6 of 18 residues are each defined as an α-helix, then the helical content is equal to 33.3% (i.e., 6/18*100%) at the given trajectory time point. These percentages were then averaged over the total number of trajectory time points (1800) to obtain the total α-helical content.

Backbone hydration was assessed using the average water coordination number (countOwater.tcl). This value describes the average number of water molecules that are proximal to the peptide's carbonyl atoms. A water molecule is coordinated to a carbonyl atom if its oxygen atom is

positioned within the first hydration shell, i.e., ≤ 3.6 Å from the peptide backbone. Then, the number of water molecules coordinated to each residue is averaged over the total number of trajectory time points, 1800.

The root-mean-squared deviation (RMSD) measures the displacement of a particular atom, r_i, or groups of atoms, at a given trajectory time point, t, with respect to a reference timepoint, t_0, which is normalized by the total number of atoms in the molecule, N, according to equation 4.1.

$$\boldsymbol{RMSD}(\boldsymbol{t}) = \sqrt{\frac{\mathbf{1}}{\boldsymbol{N}}\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{N}}(\boldsymbol{r_i}(\boldsymbol{t}) - \boldsymbol{r_i}(\boldsymbol{t_0}))^{\mathbf{2}}} \qquad \text{(eqn. 4. 1)}$$

VMD's *RMSD Trajectory Tool* module was used to perform the RMSD calculations. The initial α-helical structures (of EX19-DLNL or EX25-ALDL) were chosen as the set of reference coordinates, by which RMSD values were computed. Prior to each calculation, an alignment of trajectory structures was performed with respect to the initial structure to remove the effect of the peptide's translational motion. The output of the analysis was stored as a text file and imported into OriginPro for processing and plotting. Histograms of RMSD values were generated to obtain a distribution of structures for which a Kernel-smooth function was used for visualization.

To examine the conformational sampling of alanines in EX25-ALDL and EX19-DLNL trajectories, Ramachandran plots were generated for low-RMSD (Cα-RMSD < 2.5 Å) and high-RMSD structures (Cα-RMSD > 2.5 Å). Torsion angles of a representative structure in a given trajectory time point (section C.2.3) were obtained using VMD's *Ramaplot* module. Additionally, the torsion angles of each residue in structures recorded in all trajectory time points were acquired using phipsi.tcl. This script collects the (ϕ, ψ) angles of every residue within a peptide and combine them into a single text file. A csh-shell script (phipsitotext.csh) was used to sort and categorize the (ϕ, ψ) angles based on the type of amino acids. The values were then plotted using Excel or OriginPro.

The root-mean-squared fluctuation (RMSF) calculates the deviation of an atomic coordinate, r_i, with respect to a reference coordinate, r_i^{ref}, which is time-averaged over a period of T.

$$\boldsymbol{RMSF}(\boldsymbol{r}) = \sqrt{\frac{\mathbf{1}}{\boldsymbol{T}}\sum_{\boldsymbol{t_j}=\mathbf{1}}^{\boldsymbol{T}}\left(\boldsymbol{r_i}\left(\boldsymbol{t_j}\right) - \boldsymbol{r_i}^{\boldsymbol{ref}}\right)^{\mathbf{2}}} \qquad \text{(eqn. 4. 2)}$$

The RMSF values of Cα atoms (i.e., Cα-RMSF) in EX19-DLNL and EX25-ALDL were calculated using RMSF.tcl, which reflecting the time-averaged backbone dynamics per-residue (RMSF.tcl).

The free energy landscape (FEL) is determined by calculating the normalized probability distribution function, $\mathrm{P(r)}$, from a histogram analysis on the conformations sampled during MD, using

$$P(r) = Z^{-1}exp(-\beta W(r)) \quad \text{(eqn. 4. 3)}$$

, where r is a specified set of reaction coordinates and Z is the partition function [42, 43]. The reaction coordinates typically describe the molecular properties of the system such as Cα-RMSD, radius of gyration, atomic contacts, etc. The relative free energy, ΔG, also known as the potential mean force (PMF), $W(r)$, is then calculated by

$$\Delta G(r) = W(r_2) - W(r_1) = -k_B T.ln\, P(r) \quad \text{(eqn. 4. 4)}$$

, where k_B is the Boltzmann's constant and T is the temperature. Hence, the free energy surface is a two-dimensional contour map that reflects relative free energies of trajectory structures in the simulation, as a function of reaction coordinates 1, r_1, and 2, r_2. An FEL can be generated using, for instance, Cα-RMSD (r_1) and radius of gyration (r_2). Such FEL's are typically characterized by free energy minima, or basins, which correspond to an accessible thermodynamic state in the system, each of which represents a cluster of conformations with closely related r_1 and r_2 values.

The free energy landscapes for EX25-ALDL and EX19-DLNL trajectories were generated using the g_sham module on the GROMACS package [27, 44, 45]. Several tcl-scripts and csh-shell scripts (StripTermCatPdbGmx.csh, createFEL1.csh, createFEL2.csh) were used to convert the trajectory file produced by NAMD into a compatible format used by GROMACS. Two types of FEL's were generated for the analysis of EX25-ALDL and EX19-DLNL trajectories. First, the FEL was projected on the number of backbone hydrogen-bonds and Cα-RMSD values, as the two axes. Second, another contour map was plotted using %-helical content and Cα-RMSD values. The csh-shell script (createFEL1.csh) was used to sort the GROMACS output of relative free energy calculations, and the data was converted into a text file (createFEL2.csh) for subsequent processing and graphing using OriginPro.

C. RESULTS AND DISCUSSIONS

All-atom MD simulations were used to study the structure and dynamics of the Ala-rich EX25-ALDL and EX19-DLNL peptides, the representatives for native elastin's crosslinking domains. The secondary structures of the two crosslinked peptides during the simulations were quantified (section C.1.1). The mobility in each residue was evaluated using root-mean-square fluctuations (section C.1.2). The role of Lys-modified crosslinks and backbone hydration in the two hydrated peptides were investigated (sections C.1.3-C.1.4). Relative free energy calculation was used to examine the transition pathways of peptide folding and their kinetics in EX25-ALDL and EX19-DLNL (section C.2). Lastly, the torsion angles that define structural motifs of the alanines in EX25-ALDL and EX19-DLNL studies were analyzed (section C.3).

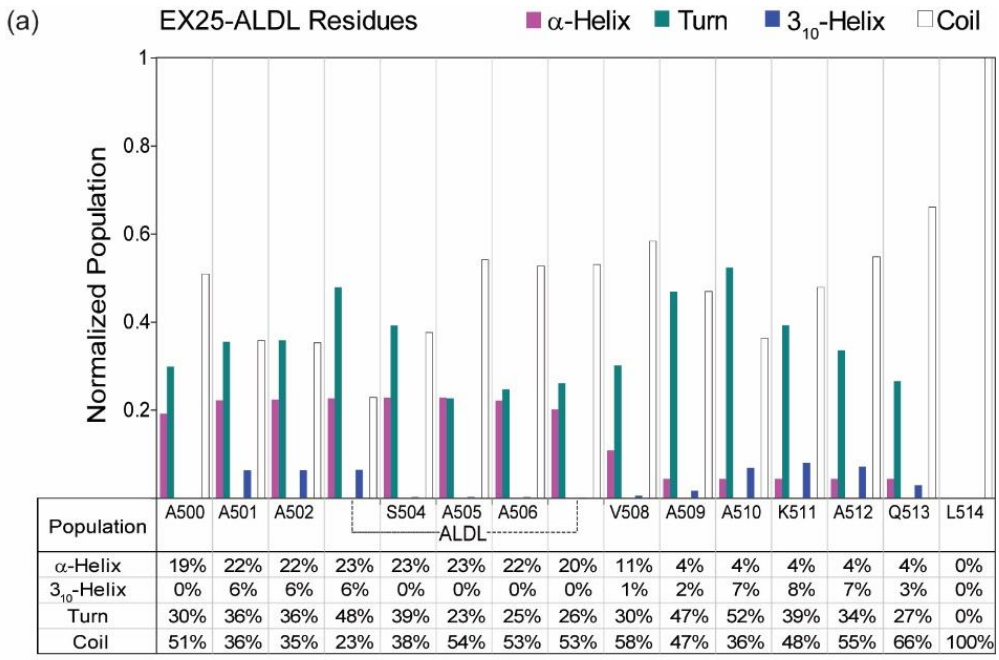
C.1 MD simulations probe the role of Lys-derived crosslinks and backbone hydration in the formation of α-helical segments in EX25-ALDL and EX19-DLNL

C.1.1. MD samples a large conformational space, and lysine-derived crosslinks facilitate the restoration of α-helices. The low α-helical content that is typically observed in elastin's crosslinking domains called for better understanding of the stability of this secondary structure. MD simulations were used to examine the conformational and dynamical features in the two Ala-rich crosslinking domains that are involved in elastin assembly. Specifically, calculations were run on EX25-ALDL and EX19-DLNL, the two Ala-rich regions containing Lys-derived crosslink moieties (ALDL and DLNL).

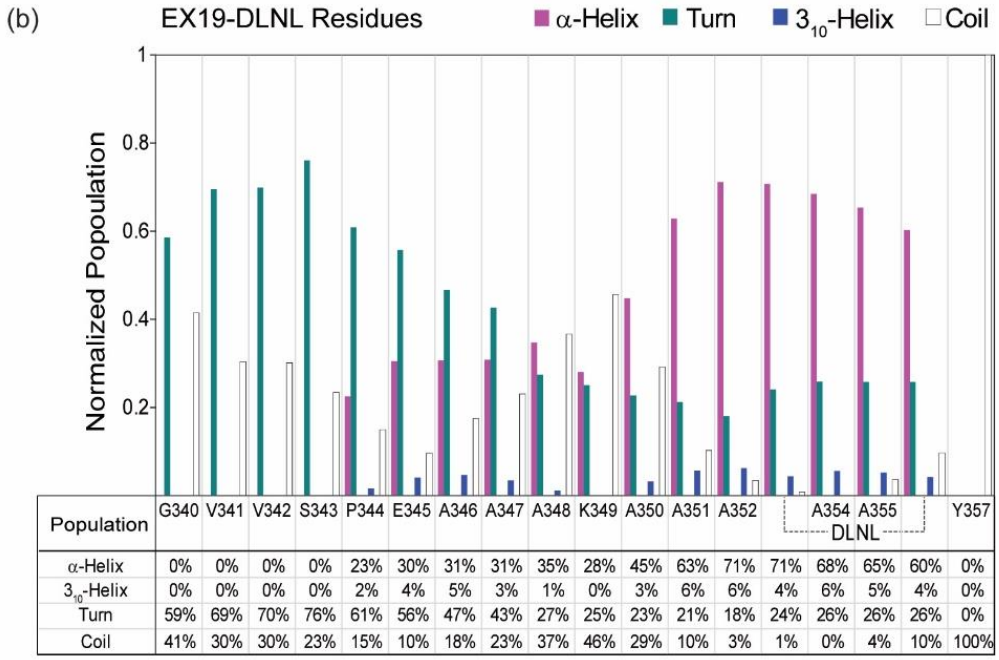
Five simulations were performed for each of the two covalently modified elastin domains, with each run analyzed individually and in aggregate. After initial minimizations, each simulation underwent a 200 ps preparation period that involves the unfolding of the initial α-helical structure to an unstructured state. Then, 200 ns of MD simulations were performed to allow each peptide to restore the helical conformation, adopt other secondary structures, or remain unstructured. The quantification of secondary structures is performed for the aggregate of 1 μs, i.e., encompassing five 200 ns simulations.

EX25-ALDL adopts at least two or three secondary structures during the simulations. As shown in **Figure 4.6a**, each peptide constituent, except for L514, spent the most time in three conformations, α-helix, turn, and coil. The α-helix (magenta bars) is identified in the N-terminal portion of the chain, from A500 to V508, for 11-20% of the trajectory time. However, the α-helical structure is identified for <5% of the simulations in the seven residues that follow, i.e., V509-L513. Overall, the turn motif (green bars) is present ≥23% throughout the entire sequence. Similarly, coil (white bars) is observed for all residues,

ranging from 23% for the ALDL crosslink to 100% for the C-terminus L514. Furthermore, L514 in EX25-ALDL have 100% (1.0) of population in coil, reflecting that none of the trajectories show these two sites to be a part of a defined conformation. Minor populations of 3_{10}-helix and bridge conformations are also observed at several sequential residues, such as A502-K504(ALDL) and V508-Q513. Moreover, negligible populations of β-sheet and π-helix were sampled during the simulation (not shown). The variety of structural motifs explored appears to be consistent with the result of CD measurements [46], in which α-helix is not the only conformation that is present in this domain. Furthermore, the presence of multiple structural motifs is consistent with the idea of conformational ensemble that was previously suggested in studies of elastin-like peptides [47].



Population	A500	A501	A502	S504	A505	A506	V508	A509	A510	K511	A512	Q513	L514
				ALDL									
α-Helix	19%	22%	22%	23%	23%	23%	22%	20%	11%	4%	4%	4%	4%	4%	0%
3$_{10}$-Helix	0%	6%	6%	6%	0%	0%	0%	0%	1%	2%	7%	8%	7%	3%	0%
Turn	30%	36%	36%	48%	39%	23%	25%	26%	30%	47%	52%	39%	34%	27%	0%
Coil	51%	36%	35%	23%	38%	54%	53%	53%	58%	47%	36%	48%	55%	66%	100%



Population	G340	V341	V342	S343	P344	E345	A346	A347	A348	K349	A350	A351	A352	A354	A355	Y357
														DLNL		
α-Helix	0%	0%	0%	0%	23%	30%	31%	31%	35%	28%	45%	63%	71%	71%	68%	65%	60%	0%
3$_{10}$-Helix	0%	0%	0%	0%	2%	4%	5%	3%	1%	0%	3%	6%	6%	4%	6%	5%	4%	0%
Turn	59%	69%	70%	76%	61%	56%	47%	43%	27%	25%	23%	21%	18%	24%	26%	26%	26%	0%
Coil	41%	30%	30%	23%	15%	10%	18%	23%	37%	46%	29%	10%	3%	1%	0%	4%	10%	100%

Figure 4.6 Relative population of adopted secondary structures during MD simulations of (a) EX25-ALDL and (b) EX19-DLNL. Colors in the graph represent the conformations identified by the STRIDE algorithm [34], and the height of the bars reflects the percentage of the total run time (1 μs) that a given residue resides in the α-helix (magenta), 3_{10}-helix (blue), turn (green), or coil (white). STRIDE also identifies isolated bridge, extended, and π-helix motifs, but their negligible populations are not shown in this figure.

The distribution of populations of secondary structures in the EX19-DLNL (**Figure 4.6b**) strikes an interesting contrast to that of the EX25-ALDL. In EX19-DLNL, α-helix and turn structures are the two major populations observed during the folding period. The α-helix populates the C-terminal portion of the peptide, whereas the turn is prevalent in the N-terminal residues. Unlike in EX25-ALDL, the helical and turn structures are observed to a significant extent (~30%-70%) at all residues. A minor population of the 3_{10}-helix is also observed (~2%-6%), but only for residues that have a significant residence time in the α-helical conformation. The 3_{10}-helices have previously been proposed as the intermediates in the folding of α-helices [48], especially in conformational transition between α-helix and the unfolded state [49].

Timeline plots provide a variety of thermodynamic quantities from an MD trajectory of EX25-ALDL and EX19-DLNL. Two features were investigated as a function of time, i.e., secondary structures and backbone hydrogen bonds. First, the timeline plot provides an insight into the evolution of secondary structures during the folding simulation of an unstructured peptide. Second, the quantification of backbone H-bond interactions, accompanying the exploration of conformational space, was also conducted. Ribbon structures that represent multiple time points in a 200 ns MD trajectory are displayed, to illustrate the restoration of α-helix and/or the adoption of other structures during the simulation. The evolution of secondary structures occurring in a (200 ns) MD trajectory was investigated; plots for other trajectories are provided in **Appendix 12**. The starting structure of each crosslinked peptide, after energy minimization, is an α-helix. After the unfolding period, the peptide's coordinates are monitored. The most prevalent structures are the turn (turquoise), α-helix (magenta), and coil (uncolored).

A representative MD trajectory of EX25-ALDL illustrates the partial refolding of the α-helix, especially in the region that is tethered by the crosslinks. At the beginning of the production period, the N-terminal half of the chain adopts the turn conformation, whereas those in the C-terminal half are present as coil (**Figure 4.7a,b**). Several constituents in the turn conformation also adopt 3_{10}-helix and α-helix intermittently within the first 40 ns. However, at ~40 ns, the adoption of turn conformation is also extended to the latter half of the chain until a possible equilibrium structure is reached after 70 ns. As the α-helix is formed in the N-terminal region, the rest of the chain (A509 to Q513) fluctuates between 3_{10}-helix and turn. By ~140 ns, all such fluctuations subside, and the C-terminal region remains as turn through the next ~20 ns of the simulation. At the last time point (200 ns), about 67% of peptide constituents are assigned as α-helix by the STRIDE algorithm, and the remainder, 37%, are determined as coil.

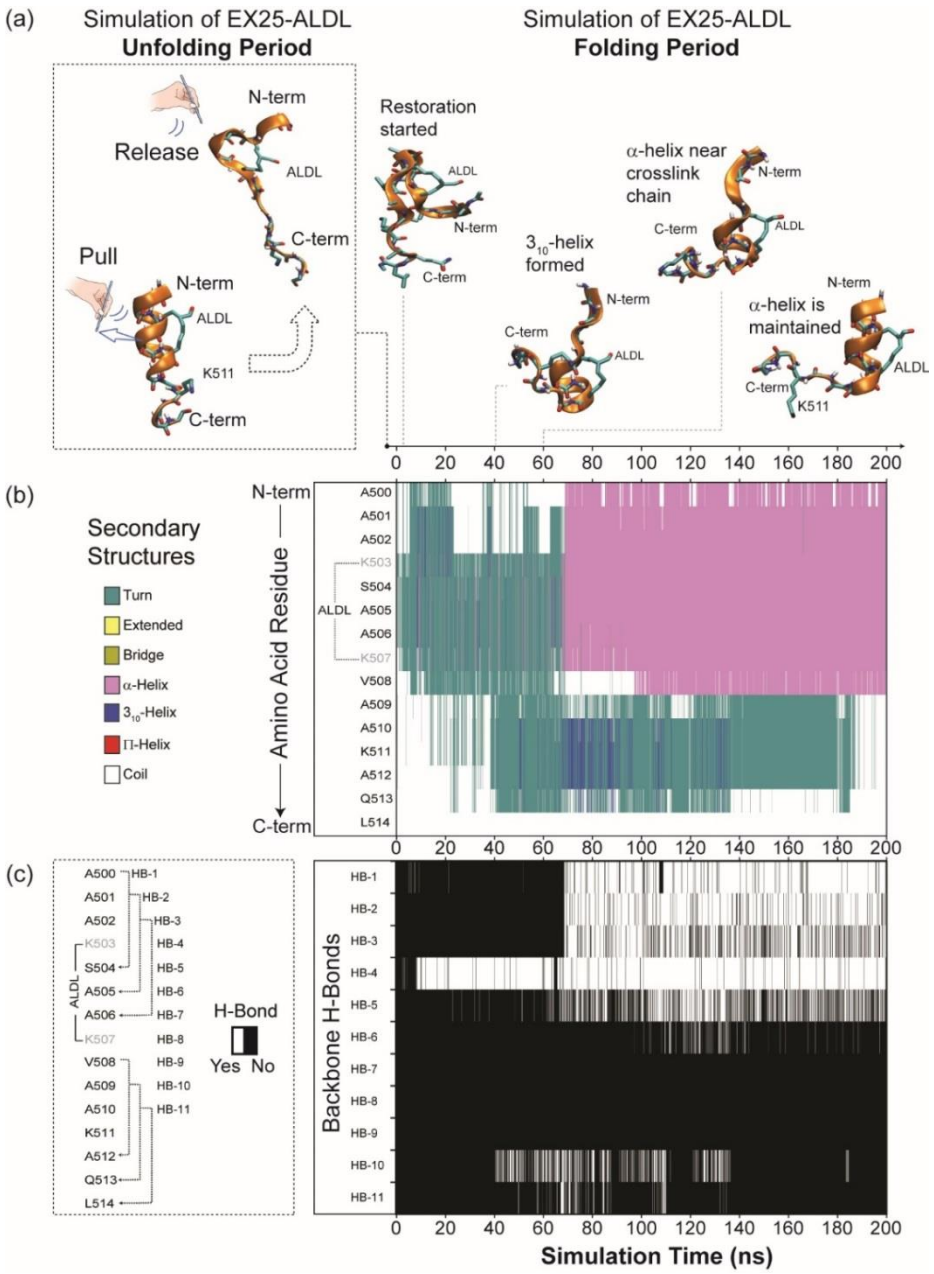


Figure 4.7 (a) Peptide representations of EX25-ALDL simulations and the timeline plots of (b) secondary structures and (c) backbone hydrogen-bonds. Horizontal axes indicate the length of a single MD trajectory (200 ns). Structures in (a) represent various time points in the trajectory. Colors in (b) indicate the various conformations identified by STRIDE [34]. Black/white lines in (c) indicate whether hydrogen-bond (HB) involving residues *i* and *i+4* is formed or not. The first hydrogen-bond interaction is defined between the CO of A500 and NH of S504, i.e., HB-1. A total of eleven hydrogen-bonds are present in a fully helical EX25-ALDL, which are indicated in succession from N-to C-terminus by HB-1, HB-2, HB-3, etc.

The restoration of the α-helical structure in EX25-ALDL is initiated by the hydrogen-bond formation of residues involved in ALDL crosslinking (**Figure 4.7c**)**.** The formation of a hydrogen bond (HB) occurs between the *ith* and *(i+4)th* residues, and eleven hydrogen-bonds are present in a (initial) fully α-helical EX25-ALDL. Within the first ~8 ns of the folding period, HB-1 and HB-4 are formed intermittently. These hydrogen-bonds correspond to the (*i/i+4*) interactions in A500/S504 and K503/K507, respectively. Then, HB-4 persists for ~40 ns, followed by HB-10 and HB-11. At 65 ns, four additional stabilizations are propagated from HB-4, i.e., HB-1, HB-2, HB-3 and HB-5, which extends this interaction through the N-terminus. This result shows that the hydrogen-bonds forming the first α-helical turn is achieved between the two modified Lys, K503 and K507, i.e., the ALDL crosslink.

The MD trajectory of EX19-DLNL shows refolding of the α-helix, except for the four N-terminal residues. Structural features from early time points (i.e., 0-40 ns) are displayed, such as the 3_{10}-helix and turn in the C-terminal residues (**Figure 4.8a,b**). These residues spend progressively more time in the α-helical conformation, until the latter form and then persist from the C-terminus to the P344. However, the N-terminal residues, G340-S343, did not adopt any helical structures but alternated between the turn and coil through the entire simulation period.

The timeline plot of backbone hydrogen-bonds (H-bonds) illustrates the formation of hydrogen-bonds that stabilize the α-helical conformation, as a function of time. In an α-helix, a single helical turn is formed by an H-bond between the CO and NH atoms of *ith* and *(i+4)th* residues, respectively. The plot illustrates the early time point formation of the α-helical structure in the C-terminal region (**Figure 4.8c**). The first backbone stabilization is formed at ~5 ns, involving the H-bonding between residues K353 and Y357 (HB-14). Within the next 20 ns, several interactions occur in a sequential fashion from HB-13 to HB-11. At ~40 ns trajectory point, another series of consecutive H-bonding formation occurs from HB-6 to HB-9, and these interactions are sustained through the end of the trajectory. Transient perturbations to HB-12, HB-13 and HB-14 are observed at 110 ns, indicating a 40 ns period of helical unfolding. These H-bond interactions are recovered at ~170 ns and they persisted to the end of the simulation.

ALDL and DLNL are the site of helix nucleation, i.e., the region where the first turn of an α-helix is formed. In both EX25-ALDL and EX19-DLNL, the initial formation of α-helical segments begins at the backbone atoms of ALDL and DLNL moieties, respectively. These moieties covalently tether two Lys residues together and limit the conformational space that is explored by the backbone atoms. Hence, this crosslinking increases the probability of the corresponding and nearby residues to readopt the α-helical structure.

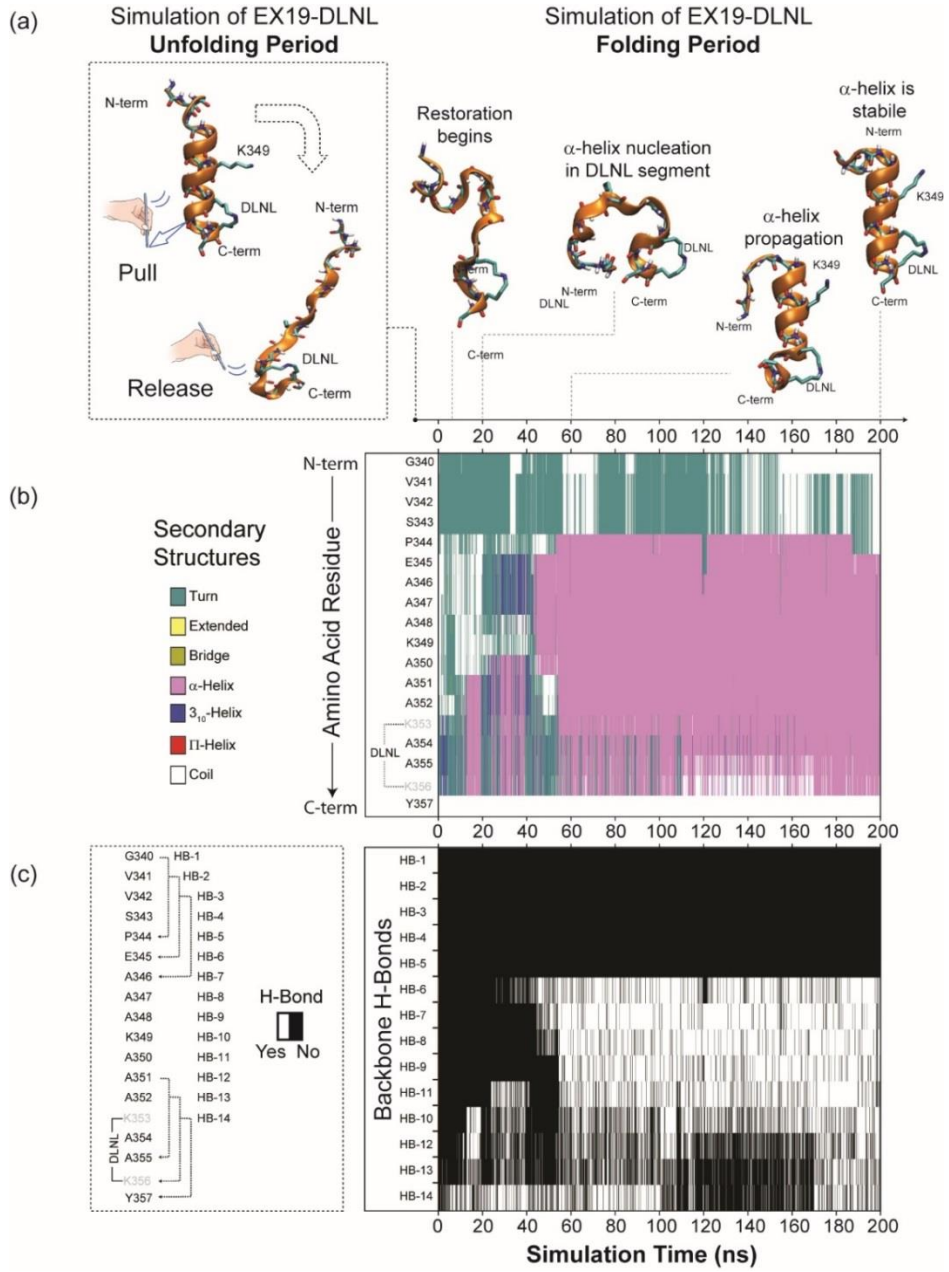


Figure 4.8 (a) Peptide representation of EX19-DLNL simulations and the timeline plots of (b) secondary structures and (c) backbone hydrogen-bonds. Horizontal axes indicate the length of a single MD trajectory (200 ns). Structures in (a) represent various time points in the trajectory. Colors in (b) indicate the various conformations identified by STRIDE [34]. Black/white lines in (c) indicate whether a hydrogen-bond (HB) involving residues *i* and *i+4* is formed or not. The first hydrogen-bond interaction is defined between the CO of G340 and NH of P344, i.e., HB-1. A total of fourteen hydrogen-bonds are present in a fully helical EX19-DLNL, which are indicated in succession from N- to C-terminus by HB-1, HB-2, HB-3, etc.

C.1.2 Terminal residues of EX25-ALDL and EX19-ALDL exhibit large-amplitude motions. To examine the backbone mobility of each residue in EX25-ALDL and EX19-DLNL peptides, the root-mean-square fluctuations (RMSF) of the Cα atoms were calculated [50-52]. The Cα-RMSF plot provides a time-averaged Cα position of each amino acid in EX25-ALDL and EX19-DLNL, over the last 100 ns of each trajectory. High Cα-RMSF values reflect large motional amplitude, whereas low Cα-RMSF values imply restricted backbone fluctuations. **Figure 4.9** shows the Cα-RMSF plots for representative trajectories of EX25-ALDL and EX19-DLNL. Others are shown in **Appendix 13**.

The C-terminal residues in EX25-ALDL have higher mobility than others (**Figure 4.9a**). L515 and K512 have the highest Cα-RMSF's among other amino acids, i.e., 4.7 and 3.6 Å, respectively. He A513, Q513, and A511 also show large backbone fluctuations with Cα-RMSF's that range between 2.4 Å and 2.7 Å. The remainder have the average atomic fluctuations of ≤ 2 Å, except for the ALDL moiety derived from K508. The RMSF's for N-terminal alanines, A501-A503, are observed between 1.8 Å and 2 Å. Alanines that are flanked by ALDL, A506 and A507, have 1.1 Å and 1.8 Å, respectively. The A510 has the Cα-RMSF of 1.3 Å, markedly lower than those in the C-terminus. The high RMSF's of C-terminal residues suggest that large-amplitude motions are exhibited within the last 100 ns of the simulation. During this period, the C-terminal region did not adopt the α-helical conformation. In contrast, this peptide segment explores a large conformational space, which includes regions with common structural motifs such as 3_{10}-helix and turns (section C.1.1).

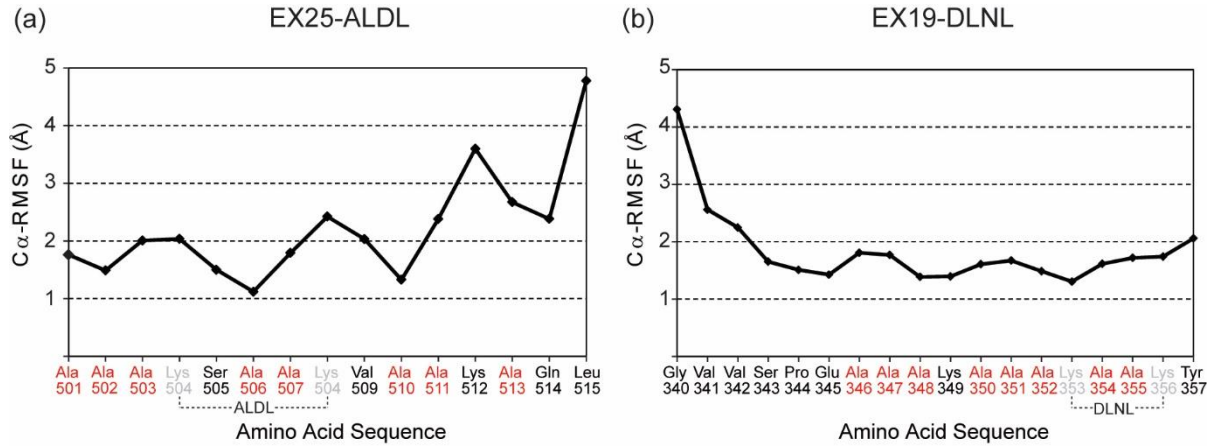


Figure 4.9 Representative Cα-RMSF plots for (a) EX25-ALDL and (b) EX19-DLNL trajectories at 300 K. The RMSF values are calculated with respect to the average Cα-coordinates within the last 100 ns of each 200 ns trajectory. Amino acid sequences show the relative position of each alanine (A) in the chain.

Alanines in the α-helical segment of EX19-DLNL have small ranges of motions (**Figure 4.9b**). All alanines in EX19-DLNL have Cα-RMSF's below 2 Å, typical for residues in the α-helix. For example,

A354 and 355, which are flanked by DLNL, are both observed with Cα-RMSF's of ~1.7 Å. The A346 and A347 also have similar values, i.e., ~1.8 Å. The remainder have RMSF's that range between 1.3 Å and 1.6 Å. This result indicates a restriction of backbone motion in the α-helical Ala population of EX19-DLNL, which occurs within the last 100 ns of the trajectory. In contrast, non-helical residues such as the N-terminal glycine and valines have broader ranges of motions, as indicated by Cα-RMSF's > 2 Å.

In summary, the Cα-RMSF evaluates the motional feature in each backbone residue which distinguishes the central from the terminal residues. Terminal alanines that are typically found in non-helical conformation have high backbone fluctuations, whereas central helical residues have limited conformational space. Additionally, alanines that are flanked by ALDL or DLNL are found with low RMSF values, suggesting little backbone fluctuation in the corresponding region.

C.1.3 Lysine-derived crosslinks and their neighboring residues have higher α-helical content than other sites in the elastin peptides. To validate the role of Lys-derived crosslinks in stabilizing the α-helical structure, the helical content in EX25-ALDL and EX19-DLNL was quantified. The percentage of time over the 200 ns MD simulation (%-helical content) reflects the frequency that each peptide constituent is defined as α-helix by STRIDE [36, 37].

Representative ribbon structures show the α-helical segments in EX25-ALDL (**Figure 4.10a**). The five EX25-ALDL trajectories are labeled as ALDL-1, 2, 3, 4 and 5. Structures for ALDL-1, -3, -4, and -5 are observed at the last time point of the trajectory. Two structures, ALDL-1 and ALDL-5, have α-helical segments on the N-termini. However, ALDL-3 and ALDL-4 were unstructured, as no helices were at any time during the simulation. A short helical segment was identified in ALDL-2 at an early time point but not retained through the end of the simulation.

The %-helical content varies within each EX25-ALDL trajectory (**Figure 4.10b**). Both ALDL-1 and ALDL-5 show ≥ 30% α-helical content in the N-terminal residues spanning A501 to A506. ALDL-2 has no helical segment in the N-terminus, but it shows a slight preference (~20%) for this conformation in the C-terminal residues A509-Q513. In contrast, ALDL-3 and ALDL-4 did not adopt the α-helical structure within the given simulation time (200 ns).

The average %-helical content of the five trajectories was calculated for each amino acid in EX25-ALDL (**Figure 4.10c**). The highest helical content in EX25-ALDL, ~20%, is found within the N-terminal half of the chain, which includes residues A500-K507. Valine V508 has an average helical content of 11%. C-terminal residues, A509-Q513, are found in the α-helical structure about 4% of the trajectory time, whereas the end residue, L514, is never found with this secondary structure. This map indicates that higher helical propensity is observed in the N-terminal region of EX25-ALDL, which

contains the crosslink moiety (ALDL). This moiety is formed by the covalent bonding of Lys residues at the *ith* (K503) and *(i+4)th* positions (K507). The *i*-to-*(i+4)* interaction is a common target for sidechain crosslinking in drug-fused α-helical peptides for therapeutic agents, as the stability of the α-helix is associated with enhanced cell permeability, protease-resistance and protein-protein interactions [53, 54].

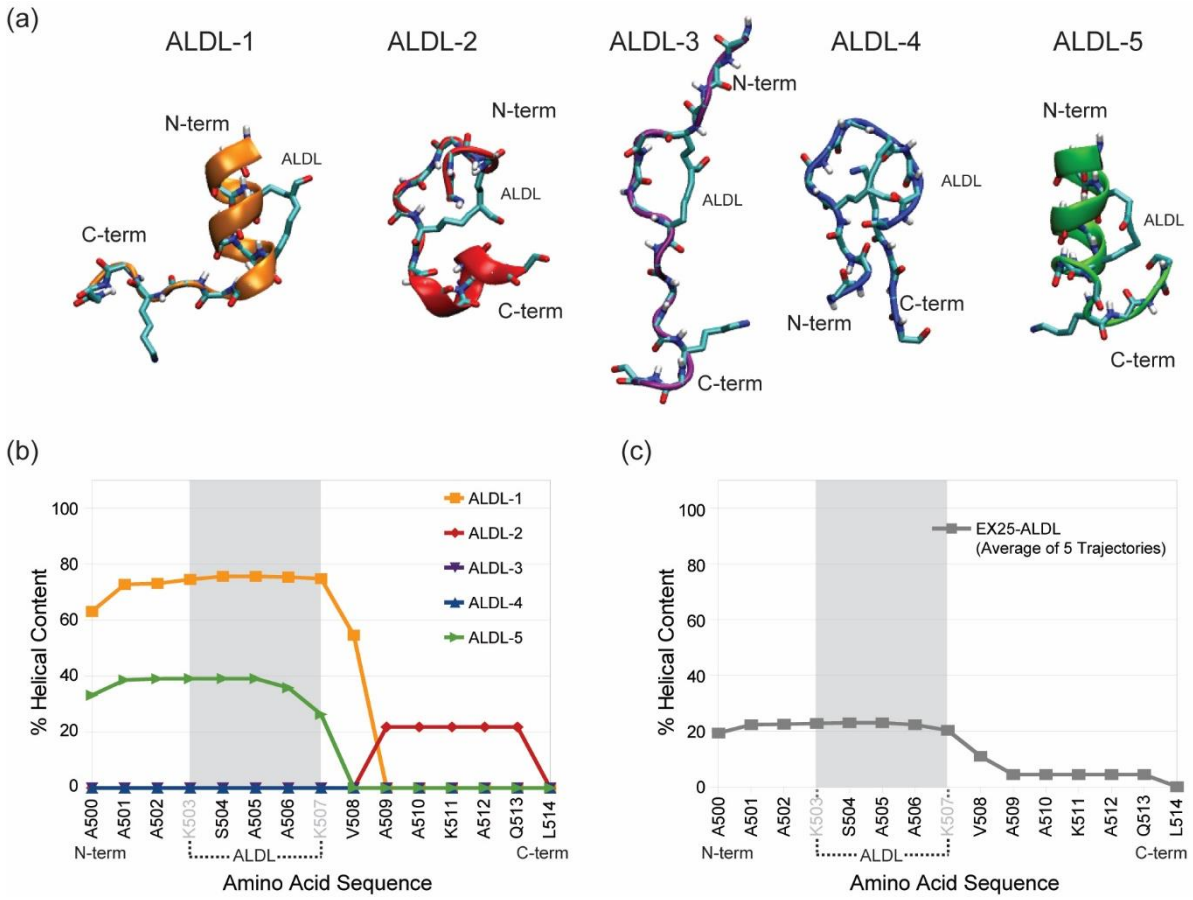


Figure 4.10 (a) Representative structures and (b) %-helical content of each of the five EX25-ALDL trajectories; (c) average %-helical content over the five EX25-ALDL trajectories. The five EX25-ALDL trajectories are labeled as ALDL-1, 2, 3, 4 and 5. All ribbon structures in (a) reflect the last time point in the trajectories except for ALDL-2, showing an early time point when a single turn of α-helix is formed. Colors in (b) represent the %-helical content in five EX25-ALDL trajectories. Several trajectories depict no helical content (0%) in the N- and/or C-terminal regions. The %-helical content in (c) reflects the average over five EX25-ADL trajectories (**gray** line). Gray shades in (b) and (c) correspond to the amino acids localized between the ALDL crosslink.

Each of EX19-DLDL representative structures has a distinct profile of α-helical segments (**Figure 4.11a**). DLNL-1 has the longest α-helical chain among other EX19-DLNL trajectories, i.e., DNL-2, -3, -4, and -5. In contrast, ~2-3 helical turns are formed in the C-terminal region of DLNL-2 and DLNL-3. The

representative structure for DLNL-4 has two α-helical segments that are interspersed by the Lys residue, K349.

The nucleation site for the α-helical structure in EX19-DLNL occurs on the C-terminal residues containing the DLNL crosslink. The map of helical content differs from one trajectory to the other, with the common points at the four N-terminal residues (G340-S343) and C-terminal Y357, where this structure's occurrence is negligible (**Figure 4.11b**). The C-terminal residues spanning from A350 to K356 have at least 40% α-helical content. The central residues, S343-A348, in DLNL-1 and DLNL-4 have > 70%, whereas those in DLNL-2, -3 and -5 have < 30% α-helical content.

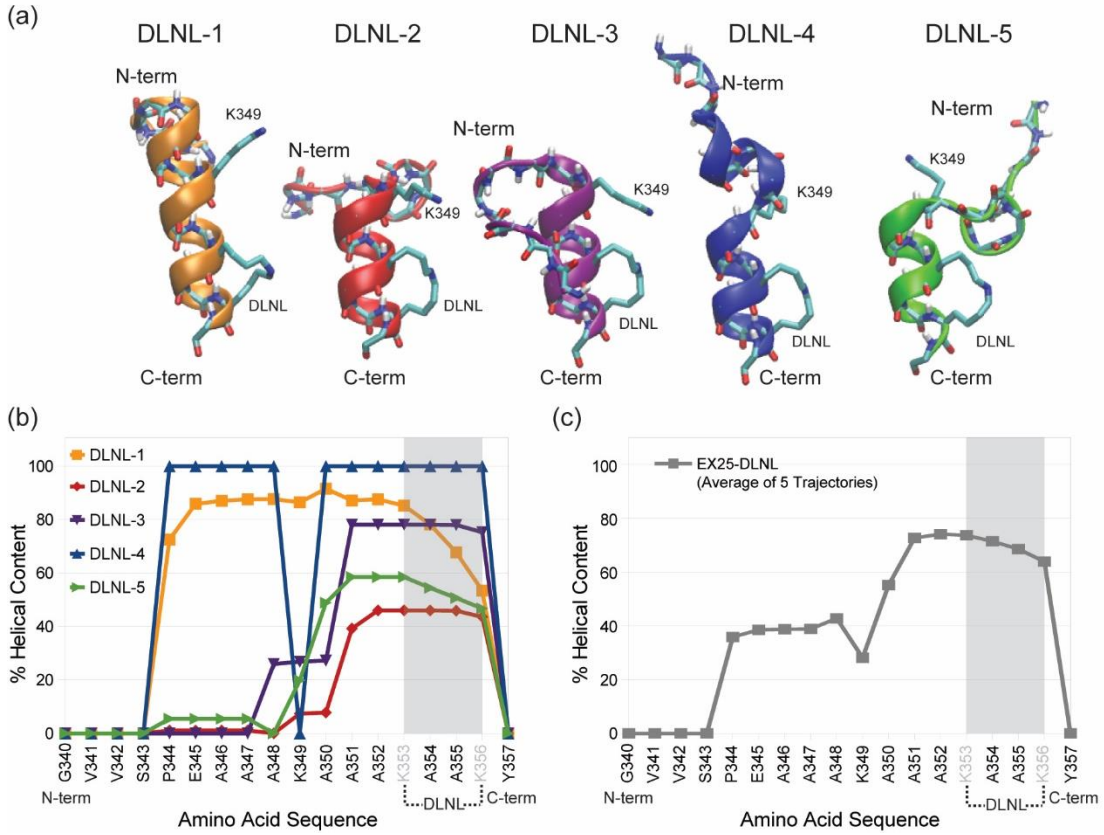


Figure 4.11 (a) Representative structures and (b) %-helical content in each of the five EX19-DLNL trajectories; (c) average %-helical content over the five EX19-DLNL trajectories. The five EX19-DLNL trajectories are labeled as DLNL-1, 2, 3, 4 and 5. Each ribbon structure in (a) is obtained from the last trajectory time point. Colors in (b) represent the %-helical content in each 200 ns trajectory. No helical content (0%) is identified in the first four residues, i.e., G340-S343. The %-helical content in (c) reflects the average over five EX19-DLNL trajectories (**gray** line). Gray shades in (b) and (c) correspond to the amino acids localized between the DLNL crosslink.

The average helical content in each amino acid in EX19-DLNL was calculated (**Figure 4.11c**). The N-terminal residues including G340-S343 and C-terminal end residue, Y357, are never found in the α-helical structure during the simulation. Residues ranging from P344 to A348 are found in the α-helical structure ~40% of the total trajectory time. The unmodified lysine, K349, has an average helical content of 28%, whereas its neighbor (A350) nearly has twice as much, 55%. C-terminal residues spanning from A351 to K357 are most often found in the α-helical conformation, i.e., 73%, of the time. This map indicates that higher helical propensities are found in the C-terminal region of EX19-DLNL, where the DLNL crosslink is located. Hence, the result suggests that the stability of the α-helical structure is enhanced in the presence of crosslink moieties, which is consistent with the experimental studies on crosslinked elastin mimetics (EP20-24) [55].

C.1.4 Low backbone hydration promotes the stability of α-helical segments in the crosslinking domains. Backbone hydration in EX25-ALDL and EX19-DLNL was evaluated using the average coordination number of water. This value provides a measure of solvent accessibility, which is obtained by calculating the average number of water molecules that are found in the first hydration shell, over the last 100 ns of the trajectory. The first hydration shell is defined as the region where the oxygen atoms of water are either in contact with or found within 3.6 Å of the peptide's carbonyl oxygen. An average of 1 water molecule is typically found near each carbonyl oxygen of an α-helical segment. A low coordination number (≤1) indicates that water molecules are blocked from accessing the backbone atoms of a given residue. In contrast, a high coordination number (>1) reflects high solvent accessibility, and hydrogen-bonds are likely to formed between the protein and the water molecules. The average coordination number of water was computed for a single, representative (200 ns) trajectory of EX25-ALDL and of EX19-DLNL. Other trajectories are included in **Appendix 14**.

The average coordination number identifies EX25-ALDL residues that are frequently in contact with water molecules (**Figure 4.12a**). Except for K503, N-terminal residues from A500 to A506 show a low average, 0.8 ± 1, indicating the sequestration of water molecules from the corresponding amino acids. The modified Lys, K503, has a significantly low coordination number, 0, reflecting dehydration of the corresponding backbone region.

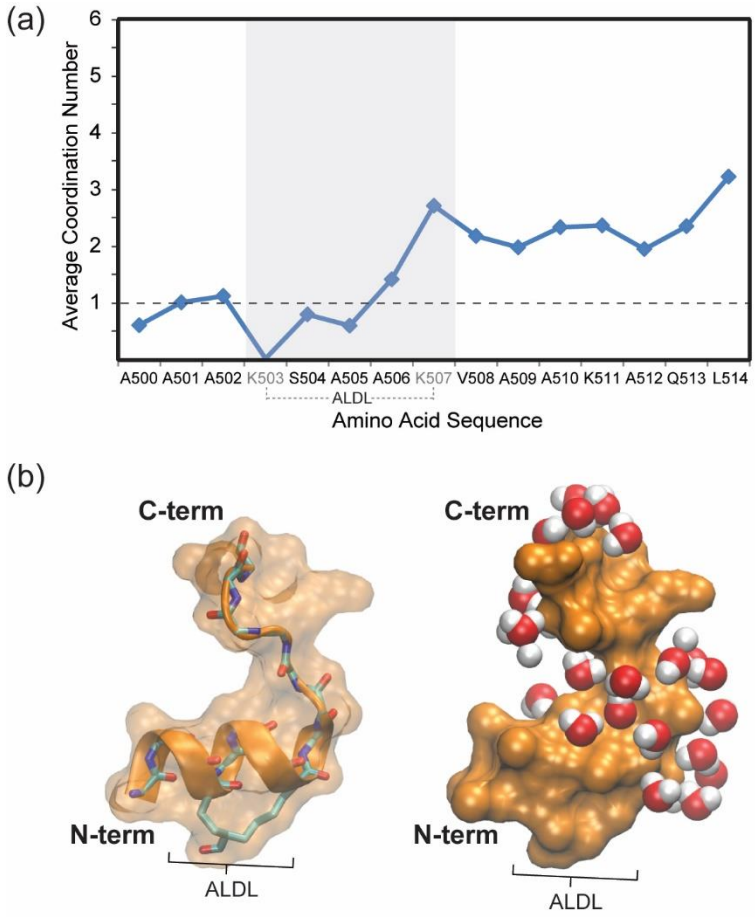


Figure 4.12 Hydration in EX25-ALDL. (a) Average coordination number of water for each amino acid in EX25-ALDL; (b) a representative structure and water molecules in the first hydration shell. The first hydration shell is defined as the region where water molecules are found within 3.6 Å of the peptide's backbone. The average coordination number of water in (a) is calculated over the last 100 ns of the MD trajectory, during which a substantial α-helical segment is adopted. Gray shade indicates amino acids that are tethered by ALDL crosslink (i.e., modified K503 and K507). Horizontal dashed line indicates the average coordination number of 1, typical for carbonyl oxygens found in the α-helical structure. In Fig.4.11 (b), *left*, representative EX25-ALDL structure is taken from the last trajectory time point. Backbone atoms without hydrogens and ALDL moiety are shown in licorice and ribbon representations. Transparent surface (**orange**) representations show the three-dimensional outline of the molecule. *Right*, representative EX25-ALDL and its hydration water are shown respectively in surface and Van der Waals representations. Only water molecules with oxygen atoms found within 3.6 Å of the peptide's carbonyl oxygens are shown.

Amino acids that are located near the ALDL crosslink, i.e., K503, S504 and A505, show the least number of water molecules among others, <1 Å, as no water molecules are found on the van der Waals surface of the corresponding region (**Figure 4.12b**). The low backbone hydration is due to the presence

of the bulky ALDL moiety which prevents the access of water to the backbone atoms of the corresponding residues. This water sequestration, or backbone-shielding, effect characterizes the N-terminal region of EX25-ALDL, which is α-helical. This result is consistent with other studies on α-helical Ala-rich peptides containing large sidechains like Arg and Lys [56, 57]. In these systems, the solute-solvent hydrogen-bonding interactions are reduced in number, whereas protein-protein interactions are prominent, resulting in the higher stability of the α-helical segments [56, 57]. In contrast, each of C-terminal residues from K507 to L514 have the average coordination number of 2.0 and higher, which reflect the solvent-exposed region in EX25-ALDL. Hence, the C-terminal residues are likely to form hydrogen-bonds with water molecules, resulting in the low propensity for the α-helical conformation.

Terminal residues in the representative EX19-DLNL structure are accessible by water (**Figure 4.13a**). The average coordination number of each N- and C-terminal residue encompassing G340-V342 and A355-Y357 is 2.5 ± 0.2. The remainder have an average of ~1 water molecule that is in proximity with each carbonyl oxygen, except for S343, E345, K349, and K353 that each has a value of <1. The result shows that the non-helical N-terminal region is frequently found in contact with water molecules, reflecting a high number of protein-water interactions. At the C-terminus, the carbonyl oxygens of the peptide are fully exposed to solvents, resulting in the high coordination number of water (~2.5). However, the α-helical structure is maintained in this C-terminal region, due to the presence of DLNL crosslink.

The effect of backbone-shielding is evident in the four residues that have <1 water coordination number, i.e., S343, E345, K349 and K353. As shown in **Figure 4.13b**, the low backbone hydration for K353 (~0.7) is due to the presence of the bulky DLNL moiety that blocks the solvent access to the backbone atoms of this residue. Analogously, the shielding effect is also identified for K349 and E345, which are charged residues that form a salt-bridge. This salt-bridge blocks the access of water molecules to the peptide backbone, which enhances the stability of the α-helical structure. Finally, the ~0.6 coordination number for S343 is due to the proximity of Ser's carbonyl oxygen atom to the hydroxyl (OH) group in its sidechain, as opposed to the hydroxyl group (OH) of water

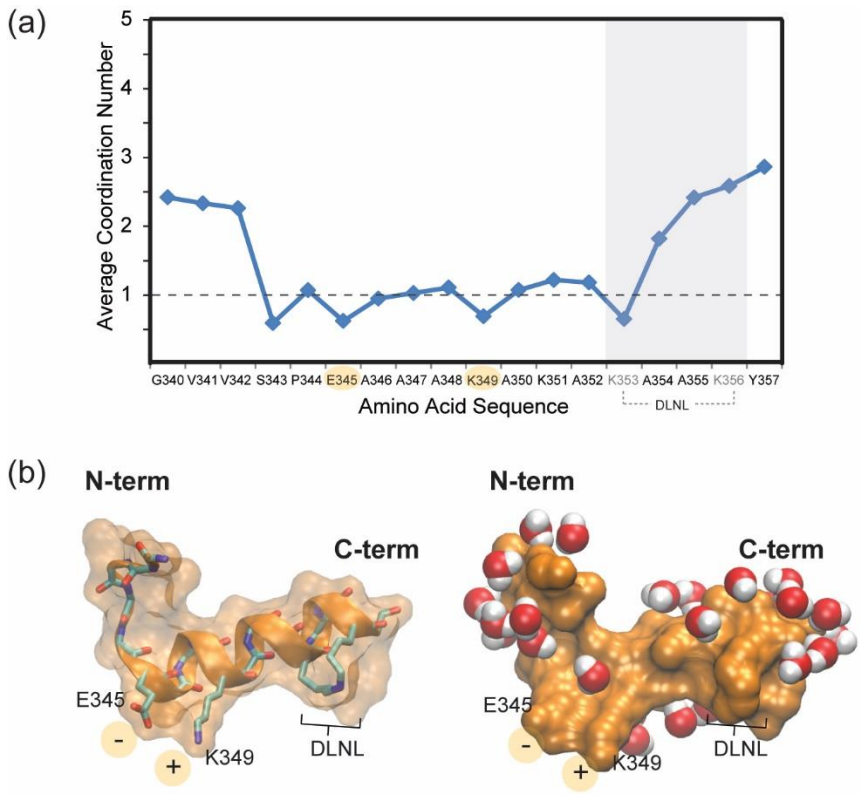


Figure 4.13 Hydration in EX19-DLNL. (a) Average coordination number of water for each amino acid in EX19-DLNL; (b) a representative structure and water molecules in the first hydration shell. The first hydration shell is defined as the region where water molecules are found within 3.6 Å of the peptide's backbone. The average coordination number of water in (a) is calculated over the last 100 ns of the MD trajectory, during which a substantial α-helical segment is adopted. Gray shade indicates amino acids that are tethered by DLNL crosslink (i.e., modified K353 and K356). Horizontal dashed line indicates the average coordination number of 1, typical for carbonyl oxygens found in the α-helical structure. In Fig.4.11 (b), *left*, representative EX19-DLNL structure is taken from the last trajectory time point. Backbone atoms without hydrogens, DLNL moiety, and sidechains of E345 and K349 are shown in licorice and ribbon representations. The transparent surface (**orange**) representations show the three-dimensional outline of the molecule. *Right*, representative EX19-DLNL and its hydration water are shown respectively in surface and Van der Waals representations. Only water molecules with oxygen atoms found within 3.6 Å of the peptide's carbonyl oxygens are shown. Positive and negative charges in K349 and E345, respectively, (**yellow** shaded) indicate the salt-bridge interaction in EX19-DLNL.

In summary, hydration analyses of EX25-ALDL and EX19-DLNL trajectories provide insights into role of elastin's crosslinks and the presence of a salt-bridge in the stabilization of α-helical structures. First, Lys sidechains and Lys-derived crosslinks, i.e., ALDL and DLNL, sequester water molecules from the backbone atoms. This backbone-shielding effect promotes protein-protein interactions via hydrogen-bonds, which enhances the stability of the α-helix. Second, the salt-bridge formed between Glu and Lys

sidechains in EX19-DLNL functions as a bulky moiety that induces backbone dehydration, which enhances the stability of the α-helix.

C.2 A two-state equilibrium provides a framework for understanding the conformational heterogeneity in EX25-ALDL and EX19-DLNL

C.2.1. Root-mean-square deviations define helix- and coil-like structures of EX25-ALDL and EX19-DLNL peptides. The heterogeneity of alanine conformations in EX25-ALDL and EX19-DLNL trajectories was studied in the context of a two-state system, i.e., α-helix and random coil (helix and coil for short). Possible interconversions of structures from one state to another, helix-to-coil or coil-to-helix transitions, were also investigated.

In MD studies, conformational changes are typically probed using the root-mean-square deviation (RMSD), which reflects the relative differences of atomic coordinates between a trajectory structure at a given time point and a reference set of coordinates, such as the initial configuration, as in this study. The RMSD values of the alpha carbons, Cα-RMSD, are calculated for EX25-ALDL and EX19-DLNL with respect to the fully folded, α-helical configuration (**Figure 4.14a,b**, *left*). The plot of Cα-RMSD's over trajectory time depicts structural conversions that occur in the course of an MD simulation, analogous to the timeline plots of section C.1.1. A low Cα-RMSD value reflects a helical structure that is relatively similar to the initial configuration. In contrast, a high Cα-RMSD value indicates a conformation that deviates from the fully folded α-helix, i.e., coil-like.

To distinguish the helical from the coil-like structures of EX25-ALDL and EX19-DLNL, histograms of the Cα-RMSD values were generated (**Figure 4.14a,b**, *right*). Kernel-smooth functions [58] were used to assist in the visualization of RMSD traces, and the distributions of RMSD's provide a distinct marker for the helical and coil-like structures. Namely, populations with Cα-RMSD values of < 2.5 Å were defined as helical, otherwise coil. Traces for all EX25-ALDL trajectories generally show local maxima corresponding to coil-like structures with Cα-RMSD's of >2.5 Å. The largest population is identified for a trace that shows two components (*red*), centered at 4.5 Å (major) and at 5 Å (minor). Other traces also feature multiple local maxima, such as one centered at 3.4 Å and 4.7 Å (*green*) and another at 3.3 Å and 3.8 Å (*orange*). Each of the remaining plots show only a single peak in the histogram, centered at 4.7 (*blue*) and at 4.8 Å (*purple*). These features suggest that the coil-like structures of EX25-ALDL are not fully unstructured, but they are partially unfolded and have unique structural features.

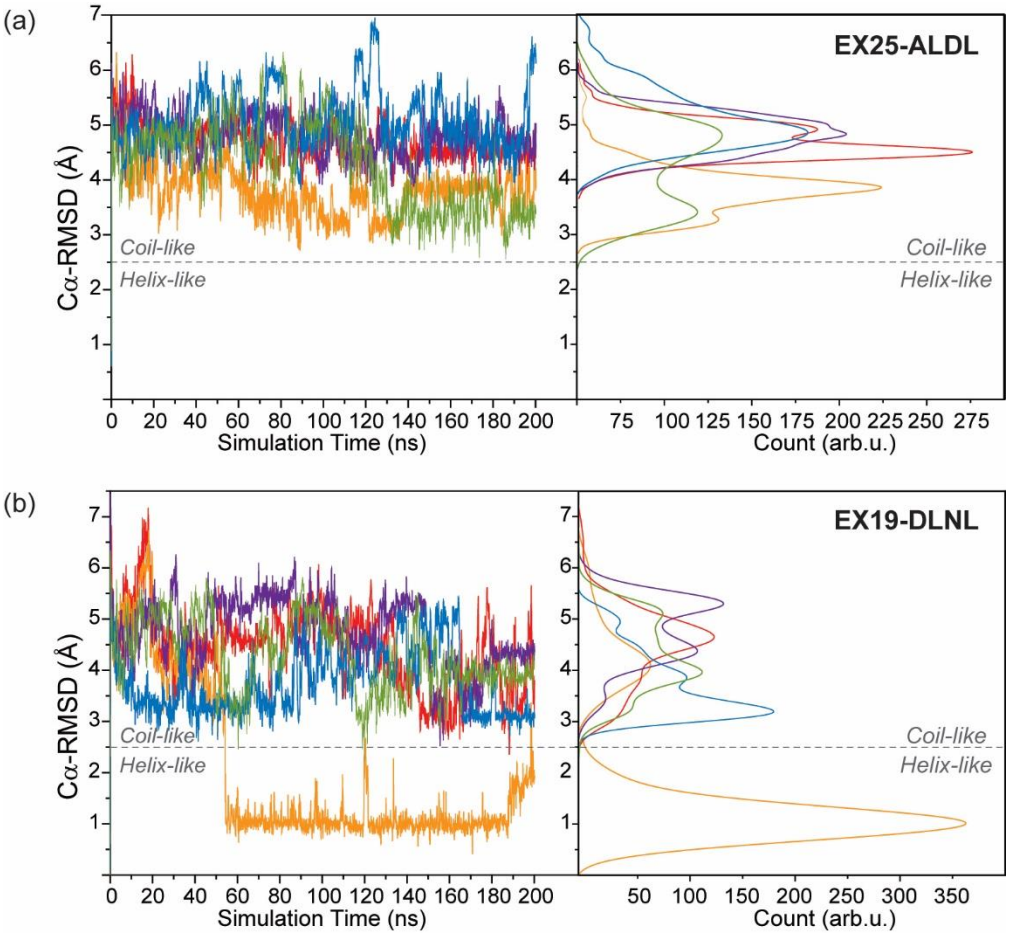


Figure 4.14 Superposition of Cα-RMSD plots (left) and Kernel-smoothened histograms (right) in (a) EX25-ALDL and (b) EX19-DLNL trajectories. *Left*, RMSD plots of five, 200 ns, trajectories of EX25-ALDL and EX19-DLNL, with each represented by a colored trace. The horizontal axis depicts the time points in the MD trajectory. *Right*, histograms show the relative population of a given trajectory structure with a given RMSD value. Horizontal dashed lines at 2.5 Å separates the helical and coil-like structures.

Both high- and low-RSMD maxima are identified in the histograms for EX19-DLNL trajectories (**Figure 4.14b**, *right*). Most Cα-RMSD traces feature multiple local maxima with high-RMSD values, with the major peaks centered at 3.2 Å (*blue*), 3.9 Å (*green*), 4.7 Å (*red*), and 5.3 Å (*purple*). These populations correspond to coil-like structures with unique structural features. Only one histogram shows maxima that are located in both high- and low-RMSD regions, with peaks identified at 4.1 Å and 0.8 Å (*orange*).

All structures in EX25-ALDL trajectories are considered to be coil(s). As shown in **Figure 4.14a** (*left*), the RMSD plot of each trajectory increases sharply at the beginning of the simulation, indicating

large conformational differences between the unfolded peptide and the (initial) α-helix. During the 200 ns simulation, the RMSD plot of each trajectory fluctuates between 3 Å and 7 Å. However, none of them decreases below 2.5 Å at any particular time point; hence, none of the structures is considered to be helix-like.

In contrast, structural transitions are observed in the MD simulation of EX19-DLNL. In one of the five trajectories (*orange*) (**Figure 4.14b**, *left*), the RMSD plot is shown to fluctuate between 3 Å and 7 Å within the first 40 ns of the MD simulation, indicating the prominence of coil-like structures at the beginning of the simulation. Then, the Cα-RMSD decreases sharply at ~45 ns and settles to a value of ~1 Å by ~60 ns, indicating a helical conformation is observed. However, the plot escalates back to a value of ~3 Å within the last 15 ns, prior to termination. These features indicate that the coil-to-helix and helix-to-coil conversions are observed at the 45 ns and 185 ns time points, respectively.

Although an EX19-DLNL trajectory indicates the presence of the helix state, the (four) others show the predominance of the coil state. This disagreement suggests that the 200 ns period of MD simulation is inadequate for a thorough conformational sampling that includes both α-helix and random coil, as the typical helix-coil transitions in polypeptides occur in the ns to μs timescale [59].

In summary, high-RMSD populations of EX25-ALDL and EX19-DLNL peptides include a mixture of structural motifs. As shown in the previous section, 3_{10}-, α-helix and turns are found in these domains (section C.1.1). In addition, short α-helical segments are adopted by amino acids that are proximal to ALDL and DLNL crosslinks, i.e., helix nucleation sites, and longer α-helices propagate from these regions (sections C.1.1-C.1.3). Moreover, the existence of local order in the coil-like structures in EX25-ALDL and EX19-DLNL is consistent with the concept of *statistical coil* in unfolded peptides and intrinsically disordered proteins [60-62].

C.2.2 Energetics of peptide folding provide a framework for defining low-energy states in the crosslinked peptides. The free energy landscape (FEL) is a two-dimensional contour map of the relative free energies of all trajectory structures in a simulation as a function of reaction coordinates, r_1 and r_2. These coordinates typically describe the molecular properties of the system such as Cα-RMSD, radius of gyration, atomic contacts, etc. An energy minimum, or free energy basin, corresponds to an accessible thermodynamic state in the system, each of which represents a cluster of conformations with closely related properties. The depth of a given basin, ΔG^* in kJ/mol, reflects the relative size of a given population of structures.

Free energy landscapes were generated for the aggregate 1 μs trajectories, or five 200-ns trajectories, of EX25-ALDL and of EX19-DLNL. Each landscape (**Figures 4.15 and 4.16**) is a contour

map showing the relative free energies of trajectory structures as described by %-helical content and Cα-RMSD. The distribution of Cα-RMSD values (vertical axis) is used to distinguish helix- and coil-like structures, whereas the %-helical content (horizontal axis) reflects the (α-helical) propensities in the corresponding populations.

The FEL for EX25-ALDL features three free energy minima, **1**, **2**, and **3** (**Figure 4.15a**). This topology reflects multiple populations of unfolded and partially folded structures. A free energy basin **1** describes a wide range of Cα-RMSD values from 2.5 Å to 7 Å and no helical content with ΔG^* ~9.9 kJ/mol, and it is centered at 0% helical content and RMSD~5 Å, or (%-helix, RMSD) = (0%, 5 Å). This basin corresponds to the largest population of structures in the MD trajectory, as represented by an unfolded EX25-ADL containing no α-helical segments. Another basin, **2**, covers a narrower range of RMSD values, 3.5 Å - 5.5 Å, and higher α-helical content, 20%-30%, which is centered at (27%, 4.5 Å) with ΔG^* ~15.8 kJ/mol. This basin represents a set of structures with several residues forming a short α-helical segment. The representative structure has a short helical segment in the C-terminal region that is formed by four residues, the least number of amino acids necessary to form a single turn of an α-helix. A third free energy minimum (**3**) is centered at (42%, 3.5 Å) with ΔG^* ~13.6 kJ/mol. It represents structures with RMSD~3-4 Å and covers a broader range of helical content, 30%-55%. Its representative structure shows a short helical segment that is formed by seven N-terminal residues. Features in this FEL indicate the presence of coil-like structures with high RSMD values, >2.5 Å, and with varying α-helical propensities. Intermediary states are also present between populations **2** and **3**, as indicated by the shallow free energy region with ΔG^* ~17 kJ/ mol, localized at (30%, 3.7 Å). These intermediate structures also imply the presence of structural conversions, as EX25-ALDL residues adopt progressively higher α-helical content from 27% (**2**) to 47% (**3**) during the course of the folding simulation.

The absence of intermediary states between region **1** and **2** reflects the high energetics in the helix nucleation step. Namely, these two populations are separated by a high free energy region, ΔG^* ~19 kJ/mol, corresponding to ~10-15%-helical content. Unlike the non-helical structures in population **1**, those in region **2** have at least four residues in the helical configuration. This helix nucleation process involves the hydrogen-bond formation of four consecutive residues to adopt the first turn of an α-helix, which is energetically more costly than the elongation or propagation process. The high energetics of helix nucleation is consistent with the helix-coil transition theory [63, 64].

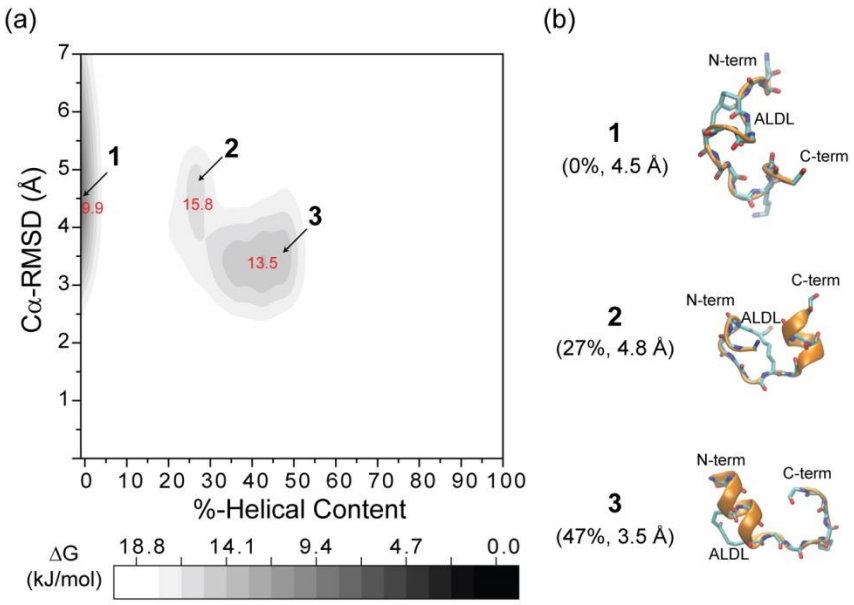


Figure 4.15 (a) Free energy landscape and (b) representative structures for EX25-ALDL trajectories. The contour map is generated from the PMF calculation of structures in five (200 ns) MD trajectories at 300 K, described as a function of %-helical content (horizontal) and Cα-RMSD (vertical). The ΔG* at the bottom of each basin is shown in red. The grayscale ( ) reflects the differences in relative free energy in kJ/mol. Representative structures for regions **1**, **2**, and **3** are obtained from the indicated (%-helical content, RMSD) coordinate in the FEL.

The FEL for EX19-DLNL trajectories shows five free energy basins, **1**, **2**, **3**, **4**, and **5** (**Figure 4.16**) A free energy basin with ΔG* ~15.9 kJ/mol (**1**) is observed over a wide range of Cα-RMSD values, 2.5 Å - 7 Å, and no helical content. This basin is centered at (0%, 4.5 Å) and it consists of the population of unfolded EX19-DLNL structures, as shown by the representative structure. Similar to the FEL for EX25-ALDL, this basin is isolated from the rest by a high free energy region with ΔG* ~24 kJ/mol, which reflects the folding energetics of the helix nucleation step. Another free energy basin (**2**) describing RMSD values ranging from 2.5 Å to 6 Å and helical content ranging from 20% to 40% has a ΔG* of ~20.4 kJ/mol. This basin is centered at (32%, 4 Å), and it corresponds to the partially helical EX19-DLNL structures. Its representative structure consists of an α-helical segment at the C-terminus of the peptide. A third free energy minimum at (56%, 3.5 Å) and ΔG* ~18.9 kJ/mol covers narrower ranges of RMSD values, 2.5 Å - 5 Å, and helical content, 45%-65%. This basin (**3**) also represents EX25-ALDL structures that are partially unfolded, and these structures have a single or multiple α-helical segment(s). The representative structure has 55% helical content and it contains two short helical segments in each end of the chain.

The two remaining basins are found in the low-RSMD region of the FEL, corresponding to structures with the most resemblance to the initial α-helical EX19-DLNL. A free energy basin, **4**, is identified with a narrow range of RMSD values, 0-2.5 Å and a relatively higher helical content, 50%-65%. This basin, centered at (61%, 1 Å) with ΔG* ~18.9 kJ/mol, represents a population that is mostly α-helical, as illustrated by the representative peptide with twelve residues adopting this secondary structure, 63% helical content. The other basin at (72%, 0.8 Å) and ΔG* ~17 kJ/mol (**5**) corresponds to structures with longer helical segments, as shown by the representative peptide.

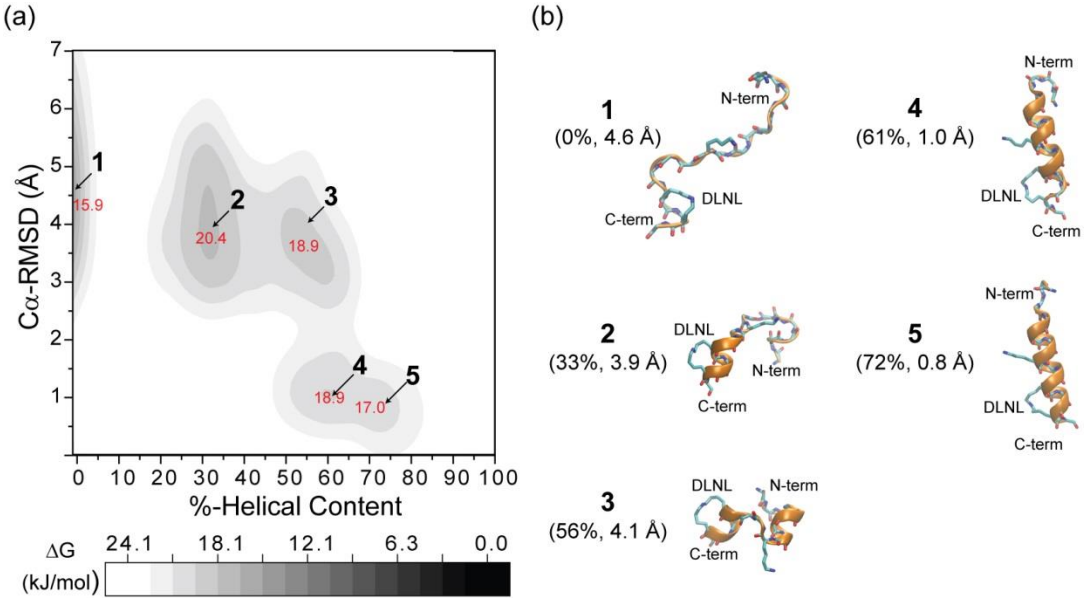


Figure 4.16 Free energy landscape and (b) representative structures for EX19-DLNL trajectories. The contour map is generated from the PMF calculation of structures in five (200 ns) MD trajectories at 300 K, described as a function of %-helical content (horizontal) and Cα-RMSD (vertical). The ΔG* at the bottom of each basin is shown in red. The grayscale ( ) reflects the differences in relative free energy in kJ/mol. Representative structures for regions **1**, **2**, **3**, **4**, and **5** are obtained from the indicated (%-helical content, RMSD) coordinate in the FEL.

The topology of the FEL for EX19-DLNL suggests the folding pathway of an unfolded structure to a folded peptide. The largest population corresponds to non-helical structures (**1**) in the MD trajectory. The peptide progressively gains higher α-helical content over the course of the simulation. A single α-helical turn is nucleated in either N- or C-terminus, as shown by representative structure **2**. This nucleation is followed by the helix elongation process that forms short helical segments in both N- and C-termini, as illustrated by structure **3**. Then, a divergence in the folding pathway is observed, as the helical segments are presumably formed in the middle of EX19-DLNL chain (**4**). This event is energetically

favorable, because the DLNL region in the C-terminus plays a role in stabilizing the helical segment in EX19-DLNL (section C.1). Correspondingly, most of the peptide residues adopt the α-helical conformation, except for the N-terminal region, as shown by representative **5**.

In summary, features in the FEL for EX25-ALDL and EX19-DLNL trajectories indicate the presence of low- and high-RMSD structures. The low-RMSD population includes those with high α-helical content, ≥ 60%, and with features that are similar to the initial, fully α-helical, configuration. In contrast, the high-RMSD populations are comprised of partially and fully unfolded structures. The partially unfolded structures have short segments of α-helices, which are formed by at least four residues. Hence, the FEL analyses indicate that both helix- and coil-like structures exist in the free energy landscapes for EX25-ALDL and EX19-DLNL peptides. However, the coil-like populations are not fully unstructured. Rather, they include common secondary structures, such as α-helical segments and other motifs described in section C.1.

C.2.3. Existence of multiple basins in the free energy landscapes supports the conformational ensemble model. Features in low- and high-RMSD structures of EX25-ALDL and EX19-DLNL trajectories were examined using additional FEL's. Each free energy surface in **Figure 4.17** and **4.18** is generated from the relative free energies of structures in the combined (1 μs) trajectories, as a function of Cα-RMSD and the number of backbone H-bonds. The number of H-bonds (vertical axis) reflects the degree of backbone stabilization in structures that are related by their Cα-RMSD values (horizontal axis).

Multiple well-defined free energy basins are identified in the FEL for EX25-ALDL (**Figure 4.17**). All free energy minima are observed with Cα-RMSD >2.5 Å; each basin differs in depth and breadth. The span of RMSD's for each basin decreases gradually with each additional H-bond. The shallowest minimum, ΔG^* ~14.4 kJ/mol, is correlated with the narrowest range of Cα-RMSD values (3-3.5 Å). This basin represents a small population of partially unfolded structures containing the greatest number of hydrogen-bonds. The representative structure **1** (**Figure 4.17**) corresponding to (3.4 Å, 9 H-bonds) in the FEL illustrates an EX25-ALDL peptide with two α-helical segments at the termini. The Ramachandran plot for this representative structure shows that most of the peptide constituents are found with torsion angles concentrated in the α_R-helix region (-57°, -45°). However, only a few residues populate other non-helical regions such as turn at (-75°, 0°) and α_L-helix at (57°, 45°).

A free energy basin covers a slightly wider range of Cα-RMSD's, 3-5 Å, and fewer H-bonds, 6-7, which represents partially unfolded structures and those with shorter helical segments. For instance, the representative structure **2** only contains one α-helical segment in the N-terminal region of EX25-ALDL, which is shown by the cluster of points in the α_R-helix region (-57°, -45°). The C-terminal non-helical

residues have torsion angles outside of the α_R-helix population; they are found in regions representing other motifs, such as (-75° ± 15°, 0° ± 15°) and (-150°, 45°) for turns, (-50°, 150°) for PP-II, and (75°, 45°) for α_L-helix. The population of partially unfolded structures increase with fewer number of backbone hydrogen-bonds, as indicated by two deeper basins (ΔG^* ~13.1-13.4 kJ/mol) centered at (4 Å, 4 H-bonds) and at (4.5 Å, 2 H-bonds) in the FEL.

Last, a free energy basin (ΔG^* ~9.7 kJ/mol) that covers the widest range of Cα-RMSD values, 3 Å- 6.6 Å, and no hydrogen-bonds characterizes the largest population of EX25-ALDL structures. This population represents fully unfolded structures, with no α-helical content. The representative structure (**3**) illustrates an EX25-ALDL peptide that is extended with (ϕ, ψ) angles that are broadly distributed over multiple regions in the Ramachandran plot. Hence, the result shows the large conformational sampling that is explored by the majority of EX25-ALDL structures in the trajectory.

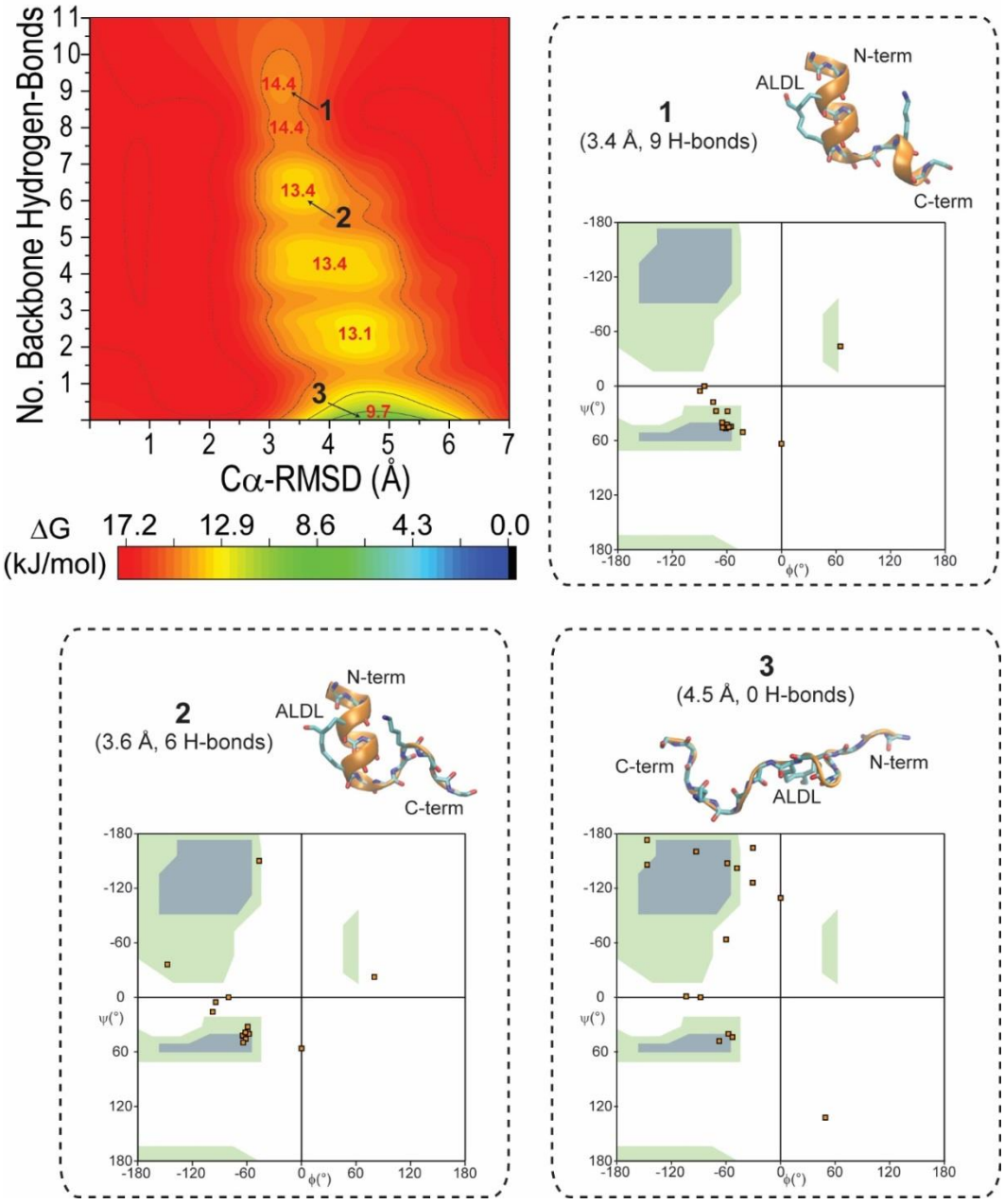


Figure 4.17 Free energy landscape, representative structures, and torsion angles for EX25-ALDL. FEL shows a contour map of ΔG*'s for all (five 200-ns) trajectory structures, as a function of backbone hydrogen-bonds and Cα-RMSD values. Insets show representative structures and Ramachandran plots corresponding to basins **1**, **2**, and **3**, which are obtained from the indicated (RMSD, No. H-Bonds) coordinate. Each Ramachandran plot contains 15 points (■), representing the total number of amino acids in EX25-ALDL. Green and blue shades in each plot correspond to the partially and fully allowed regions in the Ramachandran space [65].

The FEL for EX19-DLNL shows multiple free energy basins corresponding to high- and low-RMSD structures (**Figure 4.18**). A free energy basin with ΔG* ~8.7 kJ/mol is identified with RMSD values ranging from 0.5 Å to 2 Å with 6 to 12 H-bonds. This population corresponds to low-RMSD structures that are partially folded, which is relatively similar to the initial α-helical configuration. The representative structure (**1**) shows an EX19-DLNL peptide from (0.8 Å, 9 H-bonds) in the FEL with a continuous α-helical segment, except for the unwound N-terminal residues. Torsion angles that are primarily found in the α_R-helical region (-57°, -45°), with very slight deviations (± 15°). The non-helical residues have (ϕ, ψ) = (-60°, 115°) and (ϕ, ψ) = (-120°, 118°), which are generally considered to be in the β-sheet region of the Ramachandran space.

The FEL for EX19-DLNL trajectories also show two free energy basins that represent high-RMSD structures. First, a free energy basin with ΔG* ~7.8 kJ/mol is observed, describing RMSD's from 2.5 Å to 5.5 Å and H-bonds from 7 to 11. This basin includes partially unfolded structures that are extensively hydrogen-bonded, corresponding to high α-helical content. However, these high-RMSD structures do not have a single continuous helical segment like the initial configuration of EX19-DLNL. For example, the representative structure **2** (3.5 Å, 9 H-bonds) include two α-helical segments. The Ramachandran plot also shows several residues with torsion angles outside of the α_R-helical region; (-50°, 120°), (-75°, 125°), and (-100°, 15°) are near the PP-II and turn regions.

The largest population of EX19-DLNL structures are identified in the free energy basin (ΔG* ~6.3 kJ/mol) that covers the widest ranges of RMSD's, 2.5 Å – 6 Å, and H-bonds, 0 to 7. This population primarily represents partially unfolded structures with a short α-helical segment. For instance, the representative structure (**3**) contains one helical segment that is formed by five residues in the C-terminal region where the DLNL is present, which supports the previous discussions regarding the site for helix nucleation (sections C.1.1-C.1.2). The torsion angles for this representative structure are broadly distributed across multiple regions in the Ramachandran plot, including the α_R-helix, turn, and β-sheet regions. In addition to partially unfolded structures, this free energy basin also includes structures that do not have any backbone H-bonds, such as those located at (4.5 Å, 0 H-bonds) in the FEL. Overall, the FEL shows that the high-RMSD population of EX19-DLNL is a mixture of structures with diverse extent of backbone stabilization and α-helical content.

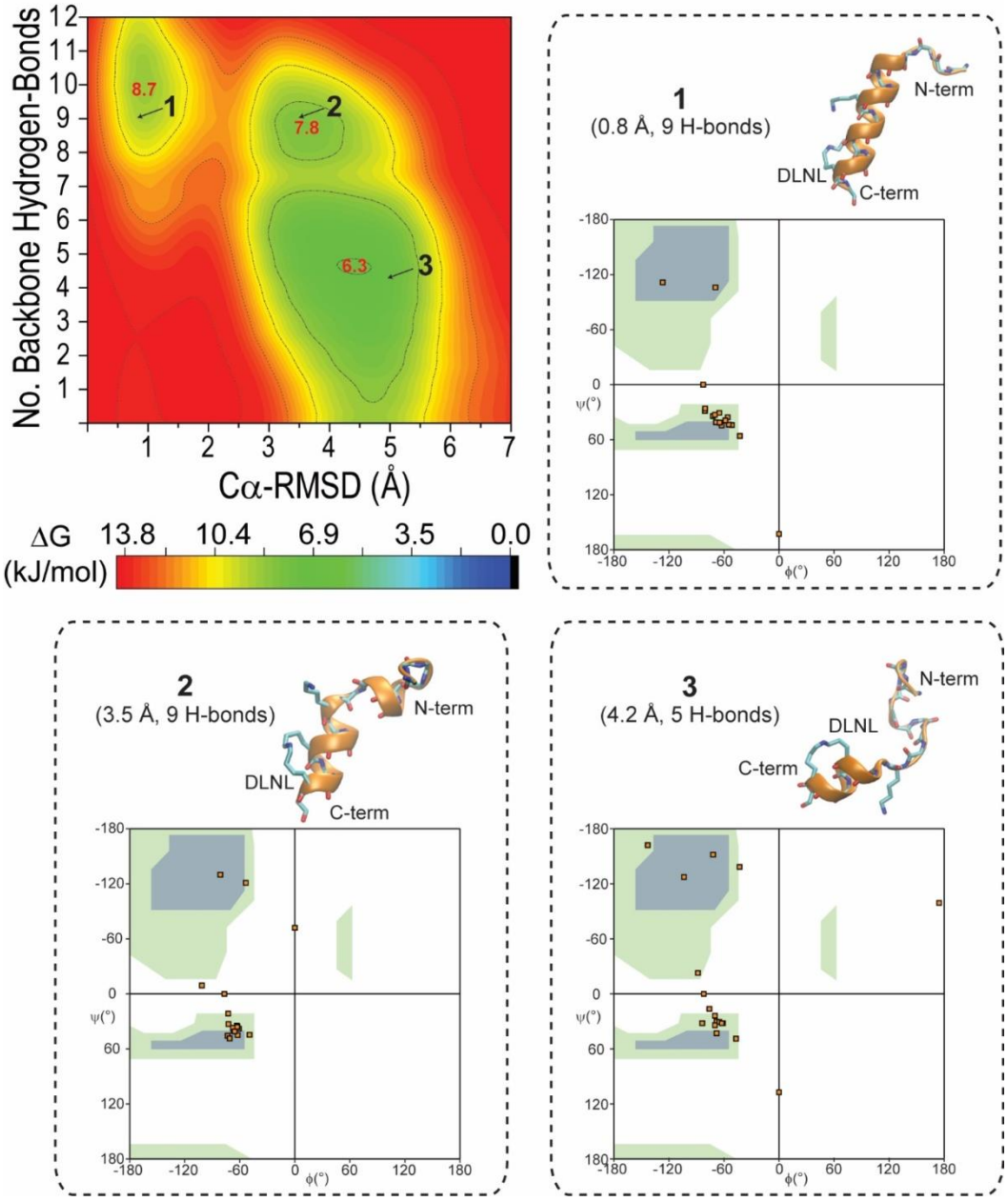


Figure 4.18 Free energy landscape, representative structures, and torsion angles for EX19-DLNL. FEL shows a contour map of ΔG*'s for all (five 200-ns) trajectory structures, as a function of backbone hydrogen-bonds and Cα-RMSD values. Insets show representative structures and Ramachandran plots corresponding to basins **1**, **2**, and **3**, which are obtained from the indicated (RMSD, No. H-Bonds) coordinate. Each Ramachandran plot contains 18 points (■), representing the total number of amino acids in EX19-DLNL. Green and blue shades in each plot correspond to the partially and fully allowed regions in the Ramachandran space [65].

In summary, multiple low-energy states exist in the folding pathways of EX25-ALDL and EX19-DLNL. These states are characterized by trajectory structures that have low and high Cα-RMSD values, relative to the (initial) helical configuration. Only high-RMSD structures are prevalent in both EX25-ALDL and EX19-DLNL trajectories. Although low-RMSD structures exist in EX19-DLNL trajectory, an accurate description of the FEL requires an exhaustive conformational sampling that is done, for instance, through variable-temperature studies (Chapter 6). Nevertheless, the existence of multiple minima in the free energy landscapes of EX25-ALDL and EX19-DLNL is consistent with the conformational equilibrium in elastin peptides [5], and this feature is also a characteristic of the intrinsically disordered proteins [66-68], such as elastin [69].

C.3 Torsion angle distributions of alanines in low-RMSD structures are different from those in high-RSMD structures

To complement the ssNMR studies described in Chapter 3, structural investigation of the crosslinking domains was focused on the alanines of EX25-ALDL and EX19-DLNL. In particular, the torsion angles of each alanine in the two corresponding peptide chains were extracted from the trajectory structures. The map of (ϕ, ψ) angles reflects the structural motif that is adopted by a given residue, over a 1 μs simulation time, i.e., five 200 ns trajectories. The Ramachandran plots display the (ϕ, ψ) angles of alanines from each trajectory structure of EX25-ALDL and EX19-DLNL (**Figures 4.19** and **4.20**). Each point (black) corresponds to a set of (ϕ, ψ) angles that is adopted by a given structure at a given simulation time point. These plots reflect the populations of low-RMSD (**Figure 4.19**, *top*) and high-RMSD (**Figure 4.19** and **Figure 4.20**, *bottom*) structures for each Ala in the respective domains. Common structural motifs such as α_R-helix and PP-II are typically associated with a particular set of torsion angles, i.e., (ϕ, ψ = -65°, -45°) and (-65°, 145°), respectively. A cluster of points in the α_R-helix region, for example, indicates the adoption of this secondary structure by a given alanine in either the low- or high-RMSD population.

The Ramachandran plots illustrate a conformational heterogeneity of alanines in EX25-ALDL (**Figure 4.19**). Only high-RMSD structures were explored during the MD simulations of EX25-ALDL, as discussed in section C.1.1. Each alanine samples two or three regions in the Ramachandran plots, including α_R-helix (-57°, -45°), α_L-helix (57°, 45°), β-strand (-145°, 145°), and PP-II (-65°, 145°). Each residue also has distinct torsion angle populations, and each population covers a wide span of values (ϕ ±30°, ψ ±30°). A501 and A502 are primarily found in the α_R-helix region (72-74%), with some distributions in β-sheet and PP-II (20-23%), as well as the α_L-helix (3-8%) region. Alanines that are flanked by the

ALDL moiety, A505 and A506, have very few points in the α_L-helix region (<1%). This lack of α_L-helix component implies a restriction of torsion angle sampling due to ALDL-crosslinking. The A506 also has a minor population (~1%) at (-60°, 60°), the inverse γ-turn region [65, 70]. The A510 has (ϕ, ψ) angles in α_R-helix (34%) and β-sheet (4%) regions, but the major population is found in the α_L-helix (72%). A509 and A512 have identical α_R-helix component (54%), but different α_L-helix distributions, i.e., 18% and 2%, respectively.

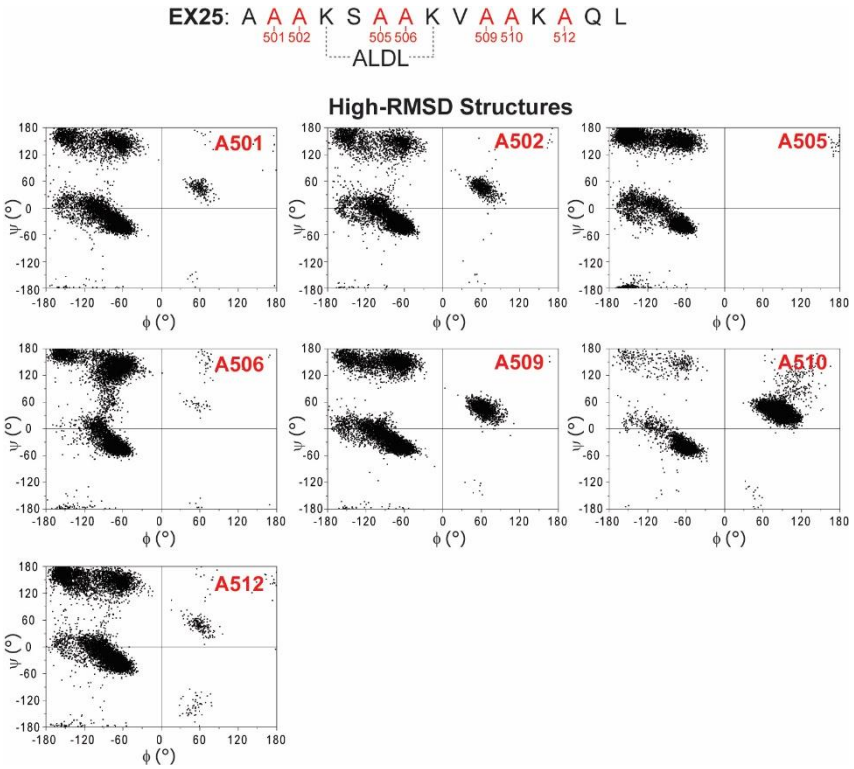


Figure 4.19 Torsion angles of alanines in high-RMSD structures of EX25-ALDL trajectories. *Top*, Amino acid sequence of EX25, highlighting positions of all Ala (**A**). Each point (•) represents the (ϕ, ψ) angles for a given alanine at each time point over all five trajectories. No low-RMSD structures are present.

The Ramachandran plots for alanines in EX19-DLNL show tight distributions of (ϕ, ψ) angles in the low-RMSD population and a broad scatter in high-RMSD structures (**Figure 4.20**). Torsion angles for all low-RMSD alanines are concentrated in the α_R-helix region (-57 ±15°, -45 ±15°), except for A354. The (ϕ, ψ) distribution of A354 extends to the bridge region, centered at (-100 ±15°, 0 ±15°). In contrast, the torsion angles of all high-RMSD alanines are scattered across multiple regions in the Ramachandran plots, including α_R-, α_L-helix, β-strand, and PP-II motifs. A broader span (ϕ, ψ) angles, i.e., (±30°, ±30°), is also observed from the center of each population. Central residues A346-A350 occupy all of the described regions in distinct ratio. For instance, the α_R- and α_R-helical populations for A346 are 72% and

7%, respectively, whereas their respective components for A350 are 56% and 12%. Several C-terminal alanines (A351-A355) are found with less than 2% of their populations in the α_L-helix region. More strikingly, the (ϕ, ψ) angles for A354 and A355 are mostly concentrated in the α_R-helix region (97-99%), with minor populations scattered in other areas ($\leq$3%). This result suggests that the DLNL moiety limits the torsion angle sampling of the two flanked alanines (A354 and A355), which is consistent with the effect of the ALDL crosslink on the neighboring alanines.

Torsion angle distributions of EX25-ALDL and EX19-DLNL high-RMSD structures display the heterogeneity of Ala conformations in elastin's crosslinking domains. Namely, alanines do not solely adopt the α-helix but also other structural motifs. The Ramachandran plots of high-RMSD structures show a large sampling of conformational space for the α_R-, α_L-helix, β-strand, and PP-II regions, consistent with those of *statistical coil* in IDP's [60-62, 71, 72]. In addition, ALDL and DLNL moieties limit the torsion angle sampling of neighboring or flanked alanines. This reduction of motional amplitude decreases the conformational space or, more precisely, the entropy in the free energy of peptide folding, which allows the enthalpy to take prominence in the formation of α-helices. This result supports ALDL- and DLNL crosslinks as the helix nucleation sites in EX25-ALDL and EX19-DLNL, respectively, which are consistent with previously described analyses (section C.1).

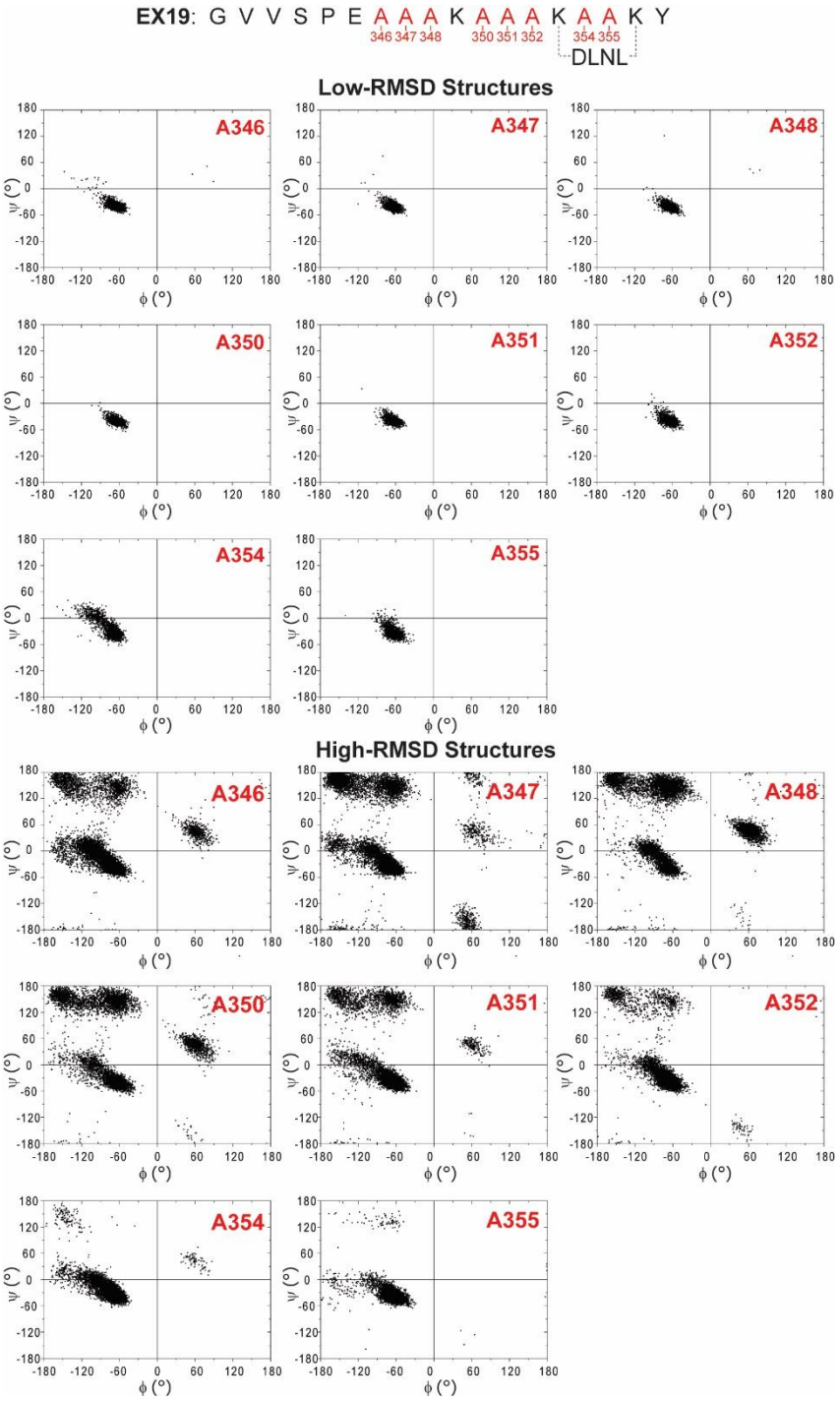


Figure 4.20 Torsion angles of alanines in the high- and low-RMSD structures of EX19-DLNL trajectories at 300 K. *Top*, Amino acid sequence of EX19, highlighting positions of all Ala (**A**). Each point (•) represents the (ϕ, ψ) angles for a given alanine at each time point over all five trajectories.

D. SUMMARY AND CONCLUSIONS

Solid-state NMR studies of hydrated NRSMC elastin identify the conformational heterogeneity of alanines in the crosslinking domains (Chapter 3). Although alanines in these regions are found in α-helices, a small portion is found in random coil. However, the exact structural character of the regions sampled by alanines is inaccessible by ssNMR. Hence, MD simulations were used to obtain insights into the conformational and dynamical features of these crosslinking regions.

MD simulations provide insights into the molecular-scale organizations of crosslinking domains in native elastin. Computations were performed on EX25-ALDL and EX19-DLNL peptides, which are models for Ala-rich domains Lys-derived crosslinks. MD shows a diverse population of secondary structures that are sampled by the peptides. Alanines in the central regions of the peptides are often found in α-helices, whereas those in the termini are identified as random coil. Furthermore, peptide termini undergo rapid fluctuations among common motifs such as α_R-, α_L-, and 3_{10}-helix, and turns, consistent with the description of *statistical coil* in unfolded peptides and IDP [60-62, 71, 72].

MD also provides insights into the role of ALDL and DLNL crosslinks in EX25-ALDL and EX19-DLNL, respectively. The Lys-derived crosslinks facilitate the restoration of α-helices from unfolded EX25-ALDL and EX19-DLNL peptides. The first turn of an α-helix in each chain is formed by hydrogen-bond interactions in the backbone atoms of amino acids that are adjacent to the ALDL or DLNL crosslink. Hence, each crosslink and its neighboring residues are called the helix nucleation site. Additional turns are then propagated from this region, resulting in a longer helical segment or higher α-helical content.

Backbone dehydration also plays a role in stabilizing the α-helical structure in EX25-ALDL and EX19-DLNL. The MD simulations show that the backbone-shielding effect decreases peptide-solvent interactions and increases peptide-peptide stabilizations via hydrogen-bonds, which enhances the stability of the α-helical segment. This shielding effect arises from the presence of bulky moieties that blocks the access of water molecules to the peptide's backbone atoms. ALDL and DLNL crosslinks are the primary factors of backbone shielding in EX25-ALDL and EX19-DLNL, respectively. This effect has also been reported previously in other helix-forming peptides containing large polar sidechains, such as Lys and Arg [56, 57, 73].

Electrostatic interactions in charged sidechains of EX19-DLNL also enhance the stability of the α-helix. The E345 and K349 in this peptide form a salt-bridge, which blocks water molecules from coming in contact with the backbone atoms. This electrostatic interaction helps maintain the backbone hydrogen-

bonding between the *ith* (E345) and *(i+4)th* positions (K349), which is favorable for the stability of an α-helix [57]. The *i-to-i+4* stabilizations are often targeted as the model for sidechain crosslinking that increases the stability of an α-helical segment, commonly used in designs of drug-fused helical peptides [53, 54].

The energetics of peptide folding reveal the existence of multiple low-energy states for EX25-ALDL and EX19-DLNL peptides. These states correspond to helix- and coil-like structures, which are categorized by the Cα-RMSD values. The low-RMSD population represents structures with high α-helical content, whereas high-RMSD structures are those with lower α-helical content. However, MD also reveals a more complex feature to the high-RMSD population. Namely, this population represents partially unfolded structures that contain short helical segments and other structural motifs, consistent with the idea of 'conformational ensemble' that was previously proposed for elastin peptides [5]. Additionally, swift interconversions occur among these low-energy states, indicating a conformational equilibrium in the crosslinking domains [5].

The existence of α-helices and random coil (or *statistical coil*) in the crosslinking domains offers insights into elastin's elastomeric properties. Elastin's resilience is partly contributed by the presence of α-helices in the Ala-rich regions, as this secondary structure is known to provide mechanical stiffness in cellular proteins [74]. As the elastic fiber is stretched, α-helical segments may be unfolded and fewer backbone hydrogen-bonds are retained. Correspondingly, Ala-rich domains in random coil undergo rapid fluctuations and sample a wide range of conformations. As the elastic fiber is allowed to recoil, hydrogen-bond formation begins at the regions where Lys-derived crosslinks are found (i.e., helix nucleation site), followed by the helix elongation that proceeds through the rest of the corresponding domains.

E. REFERENCES

1. Rousseau, R., E. Schreiner, A. Kohlmeyer, and D. Marx, *Temperature-dependent conformational transitions and hydrogen-bond dynamics of the elastin-like octapeptide GVG(VPGVG): a molecular-dynamics study.* Biophysical Journal, 2004. **86**(3): p. 1393-407.
2. Lelj, F., A.M. Tamburro, V. Villan, P. Grimaldi, and V. Guantieri, *Molecular dynamics study of the conformational behavior of a representative elastin building block: Boc-Gly-Val-Gly-Gly-Leu-Ome.* Biopolymers, 1992. **32**(2): p. 161-172.
3. Wasserman, Z.R., and F.R. Salemme, *A molecular dynamics investigation of the elastomeric restoring force in elastin.* Biopolymers, 1990. **29**(12-13): p. 1613-1631.
4. Li, B., D.O.V. Alonso, and V. Daggett, *The molecular basis for the inverse temperature transition of elastin.* Journal of Molecular Biology, 2001. **305**(3): p. 581-592.
5. Tamburro, A.M., B. Bochicchio, and A. Pepe, *The dissection of human tropoelastin: from the molecular structure to the self-assembly to the elasticity mechanism.* Pathologie Biologie, 2005. **53**(7): p. 383-389.
6. Zhao, B., N.K. Li, Y.G. Yingling, and C.K. Hall, *LCST behavior is manifested in a single molecule: elastin-like polypeptide (VPGVG)n.* Biomacromolecules, 2016. **17**(1): p. 111-118.
7. Martino, M., A. Bavoso, M. Saviano, B. Di Blasio, and A. Tamburro, *Structure and dynamics of elastin building blocks. Boc-LG-OEt, Boc-VG-OEt, Boc-VGG-OH.* Journal of Biomolecular Structure and Dynamics, 1998. **15**(5): p. 861-875.
8. Tamburro, A.M., B. Bochicchio, and A. Pepe, *Dissection of human tropoelastin: exon-by-exon chemical synthesis and related conformational studies.* Biochemistry, 2003. **42**(45): p. 13347-13362.
9. Tamburro, A.M., V. Guantieri, L. Pandolfo, and A. Scopa, *Synthetic fragments and analogues of elastin. II. Conformational studies.* Biopolymers, 1990. **29**(4-5): p. 855-70.
10. Li, B., D.O.V. Alonso, B.J. Bennion, and V. Daggett, *Hydrophobic hydration is an important source of elasticity in elastin-based biopolymers.* Journal of the American Chemical Society, 2001. **123**(48): p. 11991-11998.
11. Li, B., and V. Daggett, *Molecular basis for the extensibility of elastin.* Journal of Muscle Research & Cell Motility, 2003. **23**(5): p. 561-573.
12. Gross, P.C., W. Possart, and M. Zeppezauer, *An alternative structure model for the polypentapeptide in elastin.* Zeitschrift für Naturforschung C, 2003. **58**(11-12): p. 873-878.

13. Floquet, N., A. Pepe, M. Dauchez, B. Bochicchio, A.M. Tamburro, and A.J.P. Alix, *Structure and modeling studies of the carboxy-terminus region of human tropoelastin.* Matrix Biology, 2005. **24**(4): p. 271-282.
14. Djajamuliadi, J., T.F. Kagawa, K. Ohgo, and K.K. Kumashiro, *Insights into a putative hinge region in elastin using molecular dynamics simulations.* Matrix Biology, 2009. **28**(2): p. 92-100.
15. Miao, M., J.T. Cirulis, S. Lee, and F.W. Keeley, *Structural determinants of cross-linking and hydrophobic domains for self-assembly of elastin-like polypeptides.* Biochemistry, 2005. **44**(43): p. 14367-14375.
16. Tamburro, A.M., A. Pepe, and B. Bochicchio, *Localizing α-helices in human tropoelastin: assembly of the elastin "puzzle".* Biochemistry, 2006. **45**(31): p. 9518-9530.
17. Baldock, C., A.F. Oberhauser, L. Ma, D. Lammie, V. Siegler, S.M. Mithieux, Y. Tu, J.Y.H. Chow, F. Suleman, M. Malfois, S. Rogers, L. Guo, T.C. Irving, T.J. Wess, and A.S. Weiss, *Shape of tropoelastin, the highly extensible protein that controls human tissue elasticity.* Proceedings of the National Academy of Sciences, 2011. **108**(11): p. 4322-4327.
18. Brown-Augsburger, P., C. Tisdale, T. Broekelmann, C. Sloan, and R.P. Mecham, *Identification of an elastin cross-linking domain that joins three peptide chains: possible role in nucleated assembly.* Journal of Biological Chemistry, 1995. **270**(30): p. 17778-17783.
19. Eyre, D.R., M.A. Paz, and P.M. Gallop, *Cross-linking in collagen and elastin.* Annual Review of Biochemistry, 1984. **53**(1): p. 717-748.
20. Mecham, R.P., and J.A. Foster, *A structural model for desmosine cross-linked peptides.* Biochemical Journal, 1978. **173**(2): p. 617-625.
21. Francis, G., R. John, and J. Thomas, *Biosynthetic pathway of desmosines in elastin.* Biochemical Journal, 1973. **136**(1): p. 45-55.
22. Davis, N.R., and R.A. Anwar, *Mechanism of formation of desmosine and isodesmosine cross-links of elastin.* Journal of the American Chemical Society, 1970. **92**(12): p. 3778-3782.
23. Gray, W.R., L.B. Sandberg, and J.A. Foster, *Molecular model for elastin structure and function.* Nature, 1973. **246**(5434): p. 461-6.
24. Vrhovski, B., and A.S. Weiss, *Biochemistry of tropoelastin.* European Journal of Biochemistry, 1998. **258**(1): p. 1-18.
25. Brooks, B.R., R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, and M. Karplus, *CHARMM: A program for macromolecular energy, minimization, and dynamics calculations.* Journal of Computational Chemistry, 1983. **4**(2): p. 187-217.
26. Brooks, B.R., C.L. Brooks, A.D. Mackerell, L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer,

J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R.M. Venable, H.L. Woodcock, X. Wu, W. Yang, D.M. York, and M. Karplus, *CHARMM: The biomolecular simulation program.* Journal of Computational Chemistry, 2009. **30**(10): p. 1545-1614.

27. Van Der Spoel, D., E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, and H.J.C. Berendsen, *GROMACS: Fast, flexible, and free.* Journal of Computational Chemistry, 2005. **26**(16): p. 1701-1718.
28. Lee, J., X. Cheng, J.M. Swails, M.S. Yeom, P.K. Eastman, J.A. Lemkul, S. Wei, J. Buckner, J.C. Jeong, Y. Qi, S. Jo, V.S. Pande, D.A. Case, C.L. Brooks, A.D. MacKerell, J.B. Klauda, and W. Im, *CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field.* Journal of Chemical Theory and Computation, 2016. **12**(1): p. 405-413.
29. Jo, S., T. Kim, V.G. Iyer, and W. Im, *CHARMM-GUI: A web-based graphical user interface for CHARMM.* Journal of Computational Chemistry, 2008. **29**(11): p. 1859-1865.
30. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* Journal of Molecular Graphics, 1996. **14**(1): p. 33-8, 27-8.
31. Jorgensen, W.L., J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein, *Comparison of simple potential functions for simulating liquid water.* Journal of Chemical Physics, 1983. **79**(2): p. 926-935.
32. Phillips, J.C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kalé, and K. Schulten, *Scalable molecular dynamics with NAMD.* Journal of Computational Chemistry, 2005. **26**(16): p. 1781-1802.
33. MacKerell, A.D., D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, *All-atom empirical potential for molecular modeling and dynamics studies of proteins.* Journal of Physical Chemistry B, 1998. **102**(18): p. 3586-3616.
34. Stone, J.E., J. Gullingsrud, and K. Schulten, *A system for interactive molecular dynamics simulation*, in *Proceedings of the 2001 symposium on Interactive 3D graphics*. 2001, ACM. p. 191-194.
35. Stone, J.E., A. Kohlmeyer, K.L. Vandivort, and K. Schulten, *Immersive molecular visualization and interactive modeling with commodity hardware*, in *Advances in Visual Computing: 6th International Symposium, ISVC 2010, Las Vegas, NV, USA, November 29 – December 1, 2010,*

Proceedings, Part II, G. Bebis, et al., Editors. 2010, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 382-393.

36. Heinig, M., and D. Frishman, *STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins.* Nucleic Acids Research, 2004. **32**(Web Server issue): p. W500-W502.
37. Frishman, D., and P. Argos, *Knowledge-based protein secondary structure assignment.* Proteins: Structure, Function, and Bioinformatics, 1995. **23**(4): p. 566-579.
38. Kabsch, W., and C. Sander, *Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features.* Biopolymers, 1983. **22**(12): p. 2577-2637.
39. Richardson, J.S., *The anatomy and taxonomy of protein structure.* Advances in Protein Chemistry, 1981. **34**: p. 167-339.
40. Wilmot, C.M., and J.M. Thornton, *β-Turns and their distortions: a proposed new nomenclature.* Protein Engineering, 1990. **3**(6): p. 479-493.
41. Stickle, D.F., L.G. Presta, K.A. Dill, and G.D. Rose, *Hydrogen bonding in globular proteins.* Journal of Molecular Biology, 1992. **226**(4): p. 1143-59.
42. García, A.E., and K.Y. Sanbonmatsu, *Exploring the energy landscape of a β hairpin in explicit solvent.* Proteins: Structure, Function, and Bioinformatics, 2001. **42**(3): p. 345-354.
43. Stock, G., A. Jain, L. Riccardi, and P.H. Nguyen, *Exploring the Energy Landscape of Small Peptides and Proteins by Molecular Dynamics Simulations*, in *Protein and Peptide Folding, Misfolding, and Non-Folding*. 2012, John Wiley & Sons, Inc. p. 55-77.
44. Berendsen, H.J.C., D. van der Spoel, and R. van Drunen, *GROMACS: A message-passing parallel molecular dynamics implementation.* Computer Physics Communications, 1995. **91**(1): p. 43-56.
45. Lindahl, E., B. Hess, and D. van der Spoel, *GROMACS 3.0: a package for molecular simulation and trajectory analysis.* Molecular modeling annual, 2001. **7**(8): p. 306-317.
46. He, D., M. Chung, E. Chan, T. Alleyne, K.C. Ha, M. Miao, R.J. Stahl, F.W. Keeley, and J. Parkinson, *Comparative genomics of elastin: sequence analysis of a highly repetitive protein.* Matrix Biology, 2007. **26**(7): p. 524-40.
47. Martino, M., A. Bavoso, V. Guantieri, A. Coviello, and A.M. Tamburro, *On the occurrence of polyproline II structure in elastin.* Journal of Molecular Structure, 2000. **519**(1–3): p. 173-189.
48. Millhauser, G.L., *Views of helical peptides: a proposal for the position of 3(10)-helix along the thermodynamic folding pathway.* Biochemistry, 1995. **34**(12): p. 3873-7.
49. Daggett, V., and M. Levitt, *Molecular dynamics simulations of helix denaturation.* Journal of Molecular Biology, 1992. **223**(4): p. 1121-1138.

50. Papaleo, E., L. Riccardi, C. Villa, P. Fantucci, and L. De Gioia, *Flexibility and enzymatic cold-adaptation: a comparative molecular dynamics investigation of the elastase family.* Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 2006. **1764**(8): p. 1397-1406.
51. Paul, M., M. Hazra, A. Barman, and S. Hazra, *Comparative molecular dynamics simulation studies for determining factors contributing to the thermostability of chemotaxis protein "CheY".* Journal of Biomolecular Structure and Dynamics, 2014. **32**(6): p. 928-949.
52. Keskin, O., R.L. Jernigan, and I. Bahar, *Proteins with similar architecture exhibit similar large-scale dynamic behavior.* Biophysical Journal, 2000. **78**(4): p. 2093-2106.
53. Henchey, L.K., A.L. Jochim, and P.S. Arora, *Contemporary strategies for the stabilization of peptides in the α-helical conformation.* Current Opinion in Chemical Biology, 2008. **12**(6): p. 692-697.
54. Zhang, H., Q. Zhao, S. Bhattacharya, A.A. Waheed, X. Tong, A. Hong, S. Heck, F. Curreli, M. Goger, D. Cowburn, E.O. Freed, and A.K. Debnath, *A cell-penetrating helical peptide as a potential HIV-1 inhibitor.* Journal of Molecular Biology, 2008. **378**(3): p. 565-580.
55. Reichheld, S.E., L.D. Muiznieks, R. Stahl, K. Simonetti, S. Sharpe, and F.W. Keeley, *Conformational transitions of the cross-linking domains of elastin during self-assembly.* Journal of Biological Chemistry, 2014. **289**(14): p. 10057-10068.
56. García, A.E., and K.Y. Sanbonmatsu, *α-Helical stabilization by side chain shielding of backbone hydrogen bonds.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(5): p. 2782-2787.
57. Ghosh, T., S. Garde, and A.E. García, *Role of backbone hydration and salt-bridge formation in stability of alpha-helix in solution.* Biophysical Journal, 2003. **85**(5): p. 3187-3193.
58. Hastie, T., R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction, second edition*. 2009: Springer New York.
59. McCammon, J.A., S.H. Northrup, M. Karplus, and R.M. Levy, *Helix–coil transitions in a simple polypeptide model.* Biopolymers, 1980. **19**(11): p. 2033-2045.
60. Toal, S., and R. Schweitzer-Stenner, *Local order in the unfolded state: conformational biases and nearest neighbor interactions.* Biomolecules, 2014. **4**(3): p. 725-773.
61. Smith, L.J., K.M. Fiebig, H. Schwalbe, and C.M. Dobson, *The concept of a random coil: residual structure in peptides and denatured proteins.* Folding and Design, 1996. **1**(5): p. R95-R106.
62. Jha, A.K., A. Colubri, K.F. Freed, and T.R. Sosnick, *Statistical coil model of the unfolded state: resolving the reconciliation problem.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(37): p. 13099-13104.

63. Chen, Y., Y. Zhou, and J. Ding, *The helix–coil transition revisited.* Proteins: Structure, Function, and Bioinformatics, 2007. **69**(1): p. 58-68.
64. Scholtz, J.M., and R.L. Baldwin, *The mechanism of alpha-helix formation by peptides.* Annual Review of Biophysical and Biomolecular Structure, 1992. **21**: p. 95-118.
65. Ramachandran, G.N., C. Ramakrishnan, and V. Sasisekharan, *Stereochemistry of polypeptide chain configurations.* Journal of Molecular Biology, 1963. **7**: p. 95-9.
66. Babu, M.M., *Intrinsically disordered proteins.* Molecular BioSystems, 2012. **8**(1): p. 21-21.
67. Oldfield, C.J., and A.K. Dunker, *Intrinsically disordered proteins and intrinsically disordered protein regions.* Annual Review of Biochemistry, 2014. **83**(1): p. 553-584.
68. Dunker, A.K., J.D. Lawson, C.J. Brown, R.M. Williams, P. Romero, J.S. Oh, C.J. Oldfield, A.M. Campen, C.M. Ratliff, K.W. Hipps, J. Ausio, M.S. Nissen, R. Reeves, C. Kang, C.R. Kissinger, R.W. Bailey, M.D. Griswold, W. Chiu, E.C. Garner, and Z. Obradovic, *Intrinsically disordered protein.* Journal of Molecular Graphics and Modelling, 2001. **19**(1): p. 26-59.
69. Roberts, S., M. Dzuricky, and A. Chilkoti, *Elastin-like polypeptides as models of intrinsically disordered proteins.* FEBS Letters, 2015. **589**(19): p. 2477-2486.
70. Rose, G.D., L.M. Glerasch, and J.A. Smith, *Turns in peptides and proteins.* Advances in Protein Chemistry, 1985. **37**: p. 1-109.
71. Vila, J., H. Baldoni, D. Ripoll, and H. Scheraga, *Unblocked statistical-coil tetrapeptides in aqueous solution: quantum-chemical computation of the carbon-13 NMR chemical shifts.* Journal of Biomolecular NMR, 2003. **26**(2): p. 113-130.
72. Vila, J., D. Ripoll, H. Baldoni, and H. Scheraga, *Unblocked statistical-coil tetrapeptides and pentapeptides in aqueous solution: A theoretical study.* Journal of Biomolecular NMR, 2002. **24**(3): p. 245-262.
73. Vila, J.A., D.R. Ripoll, and H.A. Scheraga, *Physical reasons for the unusual α-helix stabilization afforded by charged or neutral polar residues in alanine-rich peptides.* Proceedings of the National Academy of Sciences, 2000. **97**(24): p. 13075-13079.
74. Shahbazi, Z., *Mechanical model of hydrogen bonds in protein molecules.* American Journal of Mechanical Engineering, 2015. **3**(2): p. 47-54.

CHAPTER 5. VARIABLE-TEMPERATURE SOLID-STATE NMR STUDIES OF ALANINES IN HYDRATED ELASTIN

A. INTRODUCTION

Conformational heterogeneity is identified at the alanines of hydrated neonatal rat smooth muscle cells (NRSMC) elastin at the physiological temperature, 37 °C (Chapter 3). Briefly, Ala in hydrophobic domains are found primarily as random coil. In contrast, alanines in the Ala-rich crosslinking domains are found in both α-helices and random coil, with an equilibrium position that is dependent on temperature.

To probe the effect of temperature on the conformations of Ala in hydrated NRSMC elastin, variable-temperature solid-state NMR (ssNMR) studies were conducted. A statistical mechanical expression from the helix-coil theory was used to determine the probability of Ala-rich sequences in the crosslinking domains to adopt the α-helical conformation. This theoretical prediction was used to simulate the ^{13}CO lineshapes for alanines, and the results were used to help interpret the ssNMR spectra acquired above and below the glass transition temperature.

A.1 Previous variable-temperature NMR studies for the characterization of hydrated elastin and other biopolymers

The variable-temperature (VT) approach is useful for the thermodynamic investigation of biopolymers. For instance, conformational transitions between α-helix and random coil in the poly(Ala) sequences of the silk fibroin protein of wild silkworm, *Samia cynthia ricini*, were discovered by VT NMR studies [1-3]. Insights into the relationship between solute (protein) and solvent (water) interactions in hydrated proteins have also been provided by other VT-based methods [4-6].

The effect of temperature on the ^{13}C ssNMR lineshapes of hydrated BNL elastin was previously demonstrated [7]. As shown in **Figure 5.1,** a notable increase of signal intensities was observed in select aliphatic regions, due to the increase of CP efficiency as the sample was cooled below the freezing point of water (-10 °C). In the report, the lineshape modification was attributed to the rigidification process that involves the transition of elastin from a “liquid-like” to a “solid-like” material. This phenomenon is often called the glass transition, i.e., a transition of an amorphous polymer from a molten rubber-like state to a brittle, “glassy” state with the decrease of temperature.

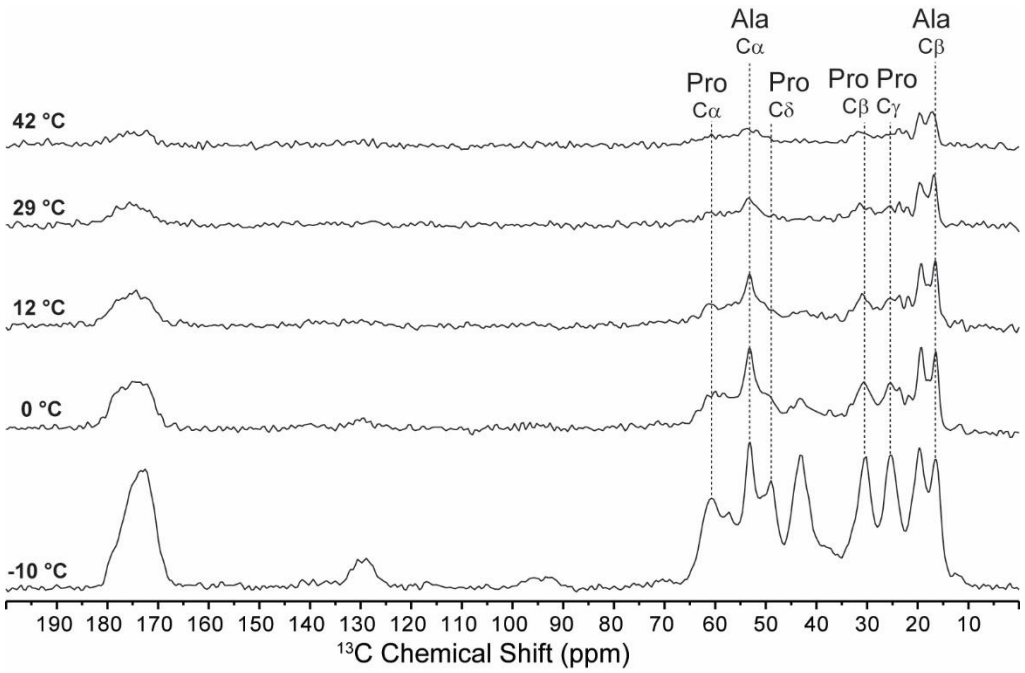


Figure 5.1 ^{13}C CPMAS spectra of hydrated BNL elastin acquired at temperatures ranging from 42 to -10 °C. Dotted lines (⋮) indicate signals from the aliphatic regions of Ala and Pro residues. ^{13}C spectra are reprocessed from the previously published data set [7].

VT NMR studies have been used to investigate the molecular properties of other biopolymer systems. For instance, ^{13}C signals were previously used to determine the conformational states of Ala in the silk fibroin protein of the wild silkworm, *Samia cynthia ricini* [1, 3]. Analysis of the ^{13}C-Ala lineshapes corresponding to the silk's poly(Ala) sequences identifies the presence of α-helix and random coil. Populations in this two-state conformational equilibrium (between helix and coil) varied with temperature. Namely, as the temperature was increased from -5 to 50 °C, the gradual changes of the ^{13}CO and ^{13}Cα lineshapes and resonance frequencies reflected a transition from helix to coil.

A.2 Helix-coil transition theory and elastin's Ala-rich crosslinking domains

Polymers undergo a reversible conversion from α-helix to random coil, also known as the helix-coil transition. This phenomenon was first observed in poly(Glu) and poly(benzyl-Glu) [8-10], where the polymer is identified to be in a conformational equilibrium between α-helix and random coil. Since its discovery, the helix-coil transition has been extensively studied both experimentally and theoretically. The experimental approach typically focused on CD measurements of α-helical content in various synthetic polymers [11, 12] and globular proteins [13]. In contrast, the theoretical approach was primarily directed

towards defining statistical thermodynamic equations that are used to calculate the polymer's temperature-dependent α-helical propensities.

The theory of helix-coil transition has undergone much development over the past few decades. Lifson-Roig's early mathematical treatment of this phenomenon assumed backbone hydrogen-bonding to be the only stabilizing interaction in the α-helix [14]. However, the approach has been progressively refined with the addition of several contributions that could stabilize the polymer's conformation. For example, Bixon, Scheraga, and Lifson [15] introduced the role of hydrophobic contacts in the α-helix. Scheraga and Poland also incorporated this interaction in the random coil model [16]. Other contributions from the helix-helix interface [16], sequence-specific, N- and C-capping effects [17-19] have also been implemented in recent theories. These modifications are applicable to both homopolymers (i.e., polymer with a single repeating unit) and copolymers (i.e., polymer with multiple repeating units) [17, 20-22].

The helix-coil theory also finds utility in the theory of protein folding [23]. In solution, polypeptides adopt a variety of conformations. The aim of the statistical mechanical approach of the helix-coil transition theory is to calculate the probability for a particular segment of the chain to adopt the α-helical conformation. Correspondingly, this approach also provides a rationale for the lack of α-helical content and the resulting presence of random coil in polymers. In recent studies of disordered and unfolded proteins, the random coil (better described as *statistical coil*) does not indicate an absence of structure(s). On the contrary, the coil is described by an ensemble of rapidly interconverting structures, which include motifs that are commonly found in globular proteins such as β-strand, turns, etc. [24-26]. More succinctly, in helix-coil theory, the coil is defined simply as the absence of α-helix.

Helix-coil theory provides an suitable basis for the simulation of ^{13}CO lineshapes of Ala-rich peptides [27], as shown in **Figure 5.2**. Namely, the spectral features are primarily influenced by the lengths of peptides; i.e., short chains of $(Ala)_n$ with $n < 12$ have low helical propensities with upfield carbonyl NMR frequencies that are typical for random coil. In contrast, long chains have high helical propensities and are observed downfield, in the regions typically observed for the chemical shifts for α-helical ^{13}CO-Ala. These trends were demonstrated in the NMR study of protein-based polymers. For instance, the helix-coil transition theory was used in conjunction with NMR experiments to interpret the ^{13}C lineshapes of silk fibroin from the wild silkworm, *Samia cynthia ricini* (described earlier) [3]. The α-helical propensities of the fibroin's poly(Ala) sequences were calculated using Bixon-Scheraga-Lifson's refinement of the helix-coil theory [15]. In turn, the propensities were used to simulate the ^{13}CO lineshape of the poly(Ala) regions [3]. The study concluded that 22 residues ($n = 22$) were required to reproduce the acquired ^{13}CO-Ala signals for the $(Ala)_n$ sequences in silk fibroin.

In addition to the length-dependence, the ^{13}CO-Ala lineshapes are influenced by the relative position of each alanine in $(Ala)_n$, with respect to the termini. Residues that are in the middle of the chain

have higher α-helical propensities than those at or near the termini. Hence, they resonate downfield from terminal alanines, which are typically in random coil. These properties are reflected in the ^{13}CO lineshapes of $(Ala)_n$ with n = 14 and 16 (**Figure 5.2**). The asymmetric "tailing" lineshapes of the longer chains reflect the populations of central Ala in an α-helix and terminal Ala in random coil, whereas short-chain alanines, i.e., $(Ala)_6$, $(Ala)_8$ and $(Ala)_{10}$, have roughly symmetric Gaussian lineshapes.

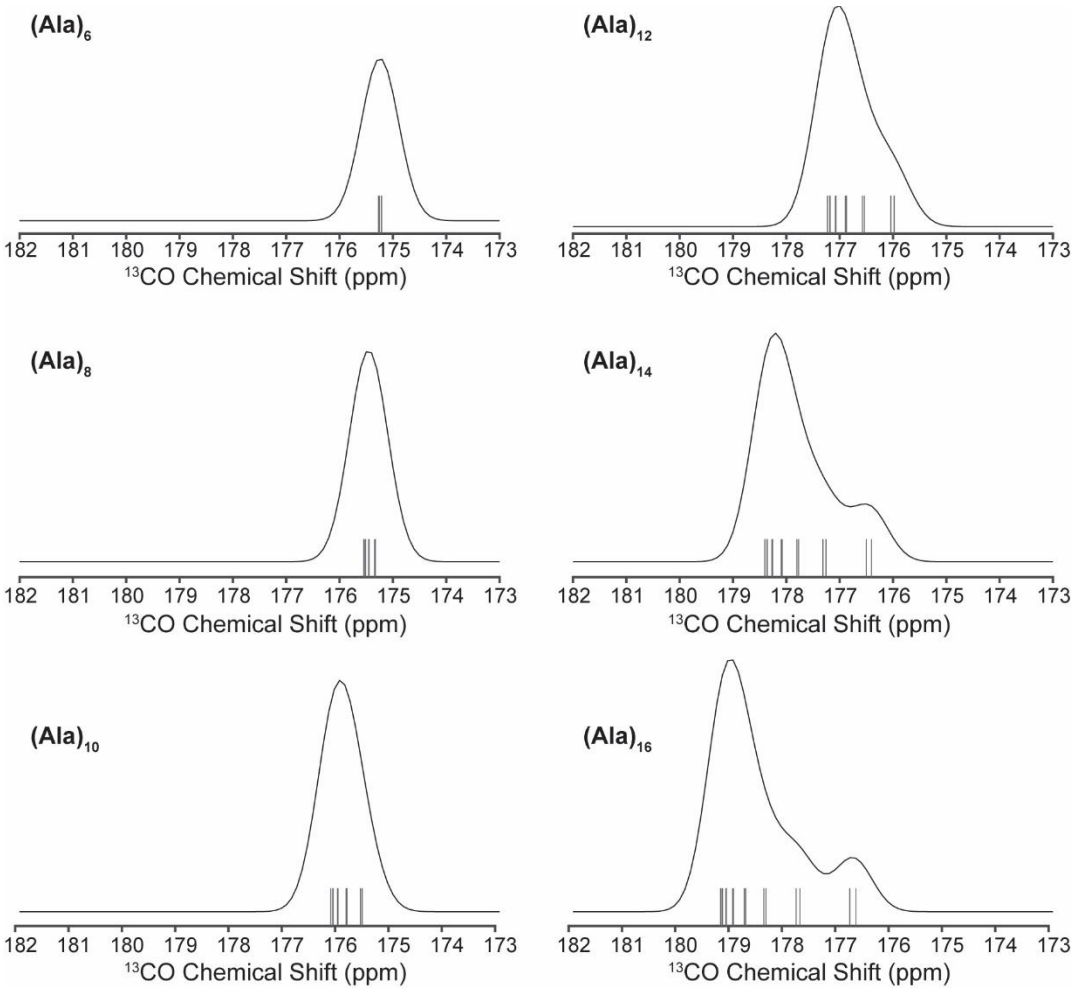


Figure 5.2 Simulations of ^{13}CO NMR lineshapes corresponding to $(Ala)_n$ with n = 6-16. Bixon-Scheraga-Lifson's approach [15] was used with the statistical weights, w, v, and σ of 1, 0.03, and 1.7, respectively. This figure is reproduced from the NMR study of wild silkworm's silk fibroin [3]. Ullman's method [27] was used to create the ^{13}CO lineshape for each $(Ala)_n$ peptide, which is the sum of Gaussian lines that are positioned at the resonance frequencies marked by the vertical lines (|). The vertical lines correspond to the predicted chemical shifts for the alanines in $(Ala)_n$. For the simulation, the highest probability for α-helix is assigned to the ^{13}CO resonance of 178.0 ppm, whereas the lowest probability (i.e., coil) is set to 175.0 ppm, which are estimates of experimental values [28, 29]

The predicted ^{13}CO chemical shifts for alanines in a longer peptide are found over a greater range, whereas those in a shorter peptide are found over a narrower region. For example, the ^{13}CO chemical shifts for Ala in $(Ala)_{19}$ range from 176.1 to 177.8 ppm, whereas for those in $(Ala)_9$ are expected from 175.2 to 175.5 ppm (**Figure 5.3**).

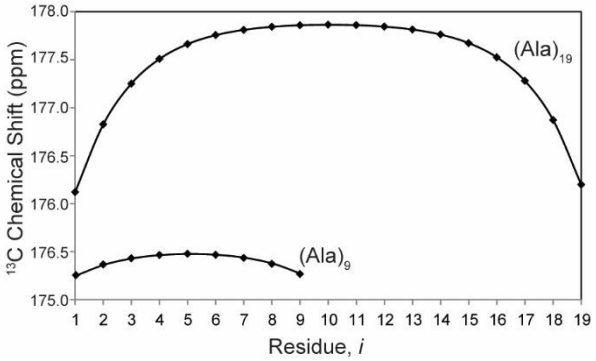


Figure 5.3 Predicted ^{13}CO chemical shifts of alanine residues in $(Ala)_{19}$ and $(Ala)_9$. The ^{13}CO resonances are calculated using Bixon-Scheraga-Lifson's approach of the helix-coil theory [15]. The horizontal axis indicates the *ith* position of alanine in the sequence. Central residues of $(Ala)_{19}$ are located within ~177.0-177.8 ppm, the typical downfield resonances for α-helices [30, 31]. The terminal residues are expected to resonate upfield, ~176.1 ppm in the coil region. Central and terminal alanines in $(Ala)_9$ range from 175.2 to 175.5 ppm, reflecting coil [32, 33].

In summary, helix-coil theory provided a means to interpret the result of NMR measurements on the silk fibroin's poly(Ala) sequences [3]. The helical propensity for each Ala residue was used as a basis for successful simulations of ^{13}CO lineshapes. This approach is next applied to the interpretation of ssNMR results of hydrated elastin at various temperatures (section C.2).

A.3 CP-based experiments for the detection of ^{13}C signals of Ala acquired at -20 °C

A VT ssNMR investigation of hydrated elastin requires multiple approaches. In Chapter 3, experiments were conducted with DP-nOe excitation to acquire ^{13}C spectra of elastin at the physiological temperature. This method is effective for the study of a hydrated elastin sample at 37 °C due to its high mobility. However, protein dynamics are largely reduced at low temperatures, especially below the freezing point of water. Therefore, another excitation method that is suitable for low temperature measurement is necessary.

Excitation with cross-polarization (CP) [34] is optimal for hydrated elastin at -20 °C. As described previously, the mobility of elastin is largely reduced below the freezing point of water, so CP is effective for detection of the relatively rigid regions in the sample, because.

A.3.1 Selective double CP experiment to observe Ala-Ala pairs in elastin. A technique that allows for the selective detection of ^{13}CO signals from the Ala-rich regions of strategically enriched elastin is described in Chapter 3, section A.2.5. Briefly, the selection of alanines with the R_{-1}-Ala-Ala motif at 37 °C was accomplished using two-step rINEPT, $^{1}H \rightarrow ^{15}N$ and $^{15}N \rightarrow ^{13}CO$ (**I** and **II** on **Figure 5.4a**). At -20 °C, an analogous two-step transfer via the dipolar coupling was achieved with double CP (DCP) [35].

The timing diagram for the $^{1}H \rightarrow ^{15}N \rightarrow ^{13}CO$ transfer includes two cross-polarization sequences (**Figure 5.4b**). The initial transverse magnetization is prepared by a 90° pulse on the ^{1}H channel and is followed by the first step of CP transfer (**I**), during which both ^{1}H and ^{15}N channels are irradiated at B_1 fields that satisfy the Hartmann-Hahn condition ($\gamma^H.B_1^H = \gamma^N.B_1^N$). The ($^{1}H \rightarrow ^{15}N$) coherence transfer occurs as the energy levels of the two spins are matched.

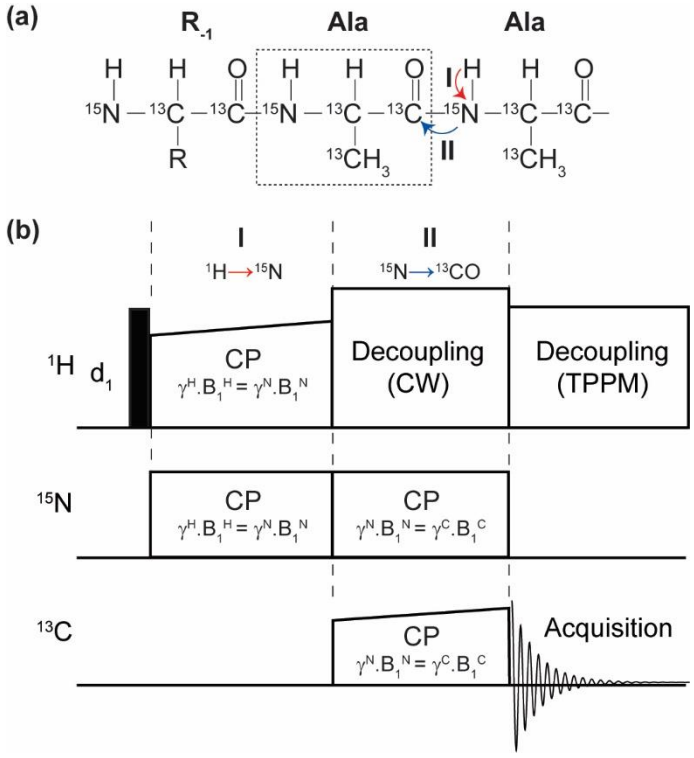


Figure 5.4 (a) $^{1}H \rightarrow ^{15}N \rightarrow ^{13}CO$ coherence transfer to select paired alanines, and (b) pulse sequence for the DCP experiment. Cross-polarization is used for $^{1}H \rightarrow ^{15}N$ transfer (step **I**), with ^{1}H RF field strength ramped from 95% to 105% of B_1^H. SPECIFIC-CP [36] condition is set for $^{15}N \rightarrow ^{13}CO$ transfer (step **II**), with ^{13}C RF field was ramped from 98% to 102% of B_1^C. Black represents a 90° pulse on the ^{1}H channel. Selective detection of signals from paired Ala is facilitated by the high levels of ^{13}C and ^{15}N enrichment in the sample (~80%).

In the second CP-based transfer (**II**), the magnetization is transferred from ^{15}N to ^{13}C, by applying B_1 fields that satisfy the $\gamma^N.B_1^N = \gamma^C.B_1^C$. The SPECIFIC-CP condition [36] is set by placing the carrier frequencies of ^{13}C and ^{15}N at 170 ppm and 120 ppm, respectively. This setup allows for the magnetization

to be transferred from ^{15}N to ^{13}CO, but not to $^{13}C\alpha$, so that only ^{13}CO signals are detected during *acquisition*. High-power ^{1}H decoupling is applied during both ($^{15}N \rightarrow ^{13}CO$) CP and *acquisition* periods.

A.3.2 CP-based 2D R-TOBSY experiment. Chapter 3 described the ^{13}CO-$^{13}C\alpha$ and $^{13}C\alpha$-$^{13}C\beta$ homonuclear correlation spectroscopy of hydrated elastin at 37 °C using 2D R-TOBSY. This technique also employed DP-nOe excitation, which exploits the high mobility of hydrated elastin at the physiological temperature. At -20 °C, the 2D R-TOBSY experiment requires CP-based excitation, as the dynamics of the protein and its waters of hydration are diminished below the freezing point of water. During the *preparation* period, the spins are irradiated under Hartmann-Hahn matching conditions ($\gamma^H B_1^H = \gamma^C B_1^C$). Otherwise, the pulse sequence for the experiment at -20°C is the same as the one used with DP-nOe at the higher temperatures (**Figure 5.5**).

The t_1 *evolution* period allows for the transverse magnetization to evolve under the chemical shift offset and the *J*-coupling interactions. The R-TOBSY isotropic *mixing* period ensures that coherence transfer occurs via *J*-couplings, while other interactions such as dipolar couplings and CSA are suppressed. During *acquisition* under t_2, the ^{13}C signals are detected with high-level decoupling.

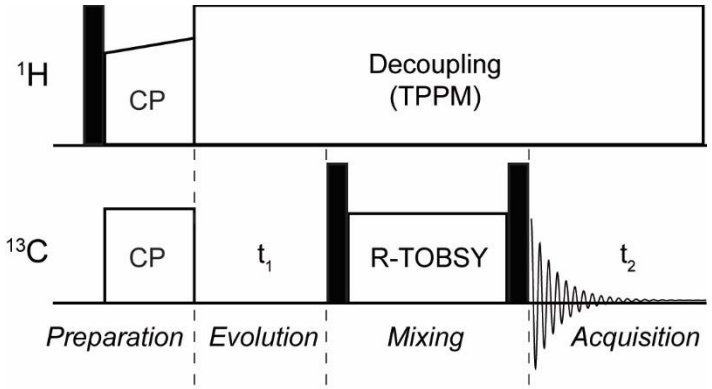


Figure 5.5 Pulse sequence for the CP-based 2D-RTOBSY experiment. Black represents 90° pulses. During cross-polarization, the ^{1}H RF field was ramped from 96% to 104% of B_1^H.

Variable-temperature ssNMR studies provide a tool to investigate the effect of temperature on the structures of biopolymers. For instance, VT studies of the silk fibroin peptide of *Samia cynthia ricini* revealed temperature-dependent helix-coil transitions of alanines in the poly(Ala) sequences [1-3]. Using VT ssNMR measurements, the investigation of alanines in elastin's hydrophobic and crosslinking domains will provide insight into the molecular-level organization of the protein in the elastic fiber.

B. EXPERIMENTAL METHODS

B.1 Preparation of elastin samples for ssNMR measurements

Isotopically-enriched elastin was prepared using the method described in **Chapter 2**, using a combination of the 1.0 mM of AOA inhibitor in the growth media that is supplemented with [37]Ala or [U-^{13}C,^{15}N]Ala at 180 mg/L. The typical sample mass, rotor inserts, and other parameters for sample preparation for ssNMR measurements were described in **Chapter 3**.

B.2 Optimization of ssNMR parameters for VT studies

One-dimensional ssNMR experiments were performed on a Varian Unity Inova spectrometer (Palo Alto, CA) operating at a ^{1}H resonance frequency of 399.976 MHz. 2D R-TOBSY and 1D DCP experiments were conducted with an Agilent DD2 console (Agilent Technologies, Santa Clara, CA) operating at a ^{1}H resonance frequency of 399.976 MHz. All ^{13}C spectra were acquired using a 4.0-mm triple resonance (HXY) T3 MAS probe (Chemagnetics/Varian, Fort Collins, CO). The MAS rate was 8 kHz for all experiments. The ^{13}C chemical shifts were referenced to the tetramethylsilane (TMS) scale, using hexamethylbenzene (HMB) as an external standard ($\delta(^{13}CH_3)$ = 17.0 ppm), at room temperature.

The [U-^{13}C]Ala NRSMC elastin was observed using 1D and 2D NMR experiments over a range of temperatures from -20 °C to 37 °C. DP, CP, and R-TOBSY pulse sequences were employed. The [^{15}N, U-^{13}C]Ala NRSMC elastin was used for the double CP experiment (i.e., $^{1}H \rightarrow {}^{15}N \rightarrow {}^{13}CO$ coherence transfer) at -20 °C.

For direct polarization (DP), a 3.3 μs ^{13}C 90° pulse was used with a 10 s recycle delay. For cross-polarization (CP), a 5.1 μs ^{1}H 90° pulse was followed by a 1 ms contact time, with a 5 s recycle delay. The field strengths for CP transfer were set to $\gamma^{H}B_1^{H}/2\pi = \gamma^{C}.B_1^{C}/2\pi$ ~50 kHz. TPPM ^{1}H decoupling [37] was applied during acquisition, with a B_1 field strength of $\gamma.B_1^{H}/2\pi$ ~40-60 kHz. The ^{13}CO detection was achieved using LOW-BASHD [38], as described in Chapter 3.

For the DCP experiment, a 5 s recycle delay was used. The first CP-based magnetization (^{1}H-^{15}N) was obtained using a 4.6 μs ^{1}H 90° pulse, with a 700 μs contact time. The ^{1}H RF field was ramped from 95% to 105% of ($\gamma^{H}B_1^{H}$ / 2π) ~36 kHz. SPECIFIC-CP [36] was used during the second (^{15}N-^{13}C) cross-polarization transfer with 4 ms contact time. The ^{13}C RF field was ramped from 98% to 102% of ($\gamma^{C}B_1^{C}$ / 2π) ~20 kHz. CW ($\gamma^{H}B_1^{H}$ / 2π = 77 kHz) and TPPM ($\gamma^{H}B_1^{H}$ / 2π = 57 kHz) decoupling were applied

on the ^{1}H channel during the second CP and acquisition, respectively. A 20 ms acquisition time was used. 1024 scans were acquired, for a total experimental time of ~1.5 hours.

For the CP-based 2D R-TOBSY experiment [39], a 3 s recycle delay was used. The cross-polarization transfer was prepared by a 3.8-3.9 μs ^{1}H 90° pulse, followed by 500-800 μs mixing times. The t_1 evolution took place under TPPM ^{1}H decoupling ($\gamma^{H}B_1^{H}/2\pi$ ~ 50 kHz). Then, a 3.9 μs ^{13}C 90° pulse was applied, followed by TOBSY mixing. To obtain ^{13}Cα-^{13}Cβ correlations, the ^{13}C carrier frequency was set to 34.4 ppm, such that it is exactly between ^{13}Cα and ^{13}Cβ during TOBSY mixing. The spectral width in the direct dimension (f_2) was 5600 Hz. 64 scans were acquired for each t_1 point, and the number of t_1 points was 40. The total experimental time was 4 hours. To obtain ^{13}CO-^{13}Cα correlations, the ^{13}C carrier frequency was placed at the middle of ^{13}CO and ^{13}Cα (113 ppm). The spectral width in the direct dimension was 17,500 Hz. 64 scans were acquired on each t_1 point, and the number of t_1 points applied was 96. The total experimental time was 10 hours.

The TOBSY mixing period has the symmetry of $R30_6^{14}$ [40]. The mixing time (τ_{mix}=6 ms) is comprised of 8 cycles. Each cycle contains 30 R elements that span across 6 rotor periods. Each R element is phase-shifted by 84° and -84°. A ^{1}H CW decoupling was used during mixing ($\gamma^{H}B_1^{H}/2\pi$ ~ 50 kHz) with an offset of 60 kHz. TPPM decoupling ($\gamma^{H}B_1^{H}$ / 2π = kHz) was applied during the 10 ms acquisition time.

B.3 Simulation of carbonyl lineshape using helix-coil transition theory

Helix-coil theory according to Bixon, Scheraga and Lifson [15] was used to calculate α-helical propensities and the corresponding ^{13}C NMR spectra [27] (**Appendix 15)**. All calculations were performed on MATLAB, version 2012b (Mathworks, MA) (**Appendix 16** for script) on a Windows-based PC with an Intel Core 2 Duo processor running at 2.8 GHz.

Bixon-Scheraga-Lifson's theory explains the stability of an Ala-rich sequence in an α-helix using two interactions, hydrogen-bonding and hydrophobic (via the sidechains), illustrated in **Figure 5.6.** The hydrogen-bond is formed between the amide hydrogen at the *ith* position and the carbonyl oxygen at (*i+4)*, and a distinctive hydrophobic interaction occurs between the Cβ of the *ith* residue and the Cα of the (*i+3)th*. As illustrated in **Figure 5.6**, a representative hydrogen-bond in $(Ala)_5$ is formed if, and only if, the three alanines, i=1, 2, and 3, are found in the α-helical conformation, in which (ϕ_1, ψ_1), (ϕ_2, ψ_2), and (ϕ_3, ψ_3) have the values of (-57°, -45°). The hydrophobic interaction between i=1 and i=4 is present only if the ψ_1, ϕ_2, ψ_2, ϕ_3, and ψ_3 angles are all found in the α-helical region of the Ramachandran plot; the ϕ_1 angle does not have to adopt -57° for this interaction [15].

In Bixon-Scheraga-Lifson's study [15], the poly(Ala) chain was flanked by poly(Glu) to render the entire polymer soluble in water. This arrangement allows for the first and last alanines in $(Ala)_n$ to be involved in hydrogen-bonding or hydrophobic interactions with the flanking polymers. For instance, the amide hydrogen of Ala at the N-terminus, $i=5$, can form a hydrogen-bond with the carbonyl oxygen of a Glu in the neighboring poly(Glu). Analogously, each Ala-rich sequence of elastin's crosslinking domains is flanked by the hydrophobic domains. Hence, the terminal alanines in the crosslinking regions also have non-zero probabilities to form a hydrogen-bond or hydrophobic interaction with residues from the flanking (hydrophobic) domains.

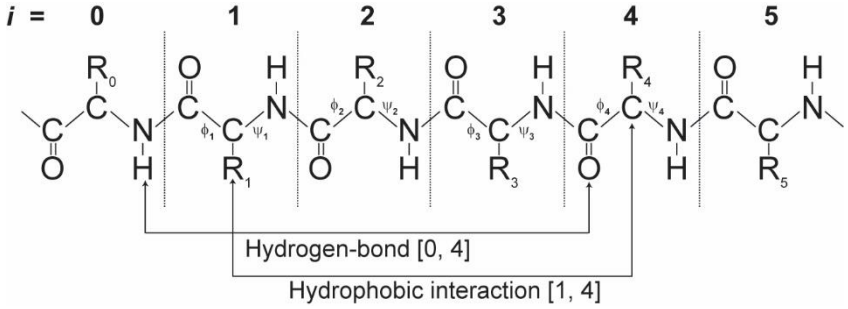


Figure 5.6 Schematic representation of a polypeptide chain with residues *i* = 0 to 5. A representative hydrogen-bond [0,4] is formed between the NH group at i = 0 and the CO group at i = 4, if and only if the (ϕ_1, ψ_1), (ϕ_2, ψ_2), and (ϕ_3, ψ_3) have the values corresponding to those of an α-helix (-57°, -45°). A representative hydrophobic interaction [1,4] occurs between the Cβ of residue i = 1 (R_1) and the Cα of i = 4, if the ψ_1, ϕ_2, ψ_2, ϕ_3, and ψ_3 are found in the α-helical conformation.

To formulate the partition function for the poly(Ala) sequence in an α-helix, the generalized nearest-neighbor model is used [41]. The statistical weights for an alanine at the *ith* position is denoted by w, or v, or $w\sigma$, or u, or αu, depending on the helix or coil assignment of the neighboring $(i+1)th$ and $(i+2)th$ residues. An alanine that is found in an α-helical segment that is stabilized by a hydrogen-bond contributes to the partition function by w. The statistical weight for a residue with a hydrophobic interaction is denoted by σ. Hence, residues containing both hydrogen-bond and hydrophobic interactions contribute to the partition function by $w\sigma$. On the other hand, the last two residues in a segment of an α-helix have the nonzero factor of v, due to the absence of hydrogen-bonds. The statistical weight for an Ala in random coil is denoted by u. However, a coil residue that is preceded by two (or more) alanines in the helical conformation is denoted by αu, where α is a scaling factor that reflects the strength or absence of the hydrophobic interaction in these residues. This factor α is defined by $\alpha=1+v^{1/2}(\sigma-1)$, in which σ approaches 1 (and α approaches 1, as a result) in the absence of a hydrophobic bond. The statistical weights for alanines in the partition function of a poly(Ala) are assigned using these definitions. For example, alanines in the $(Ala)_{12}$ chain with a particular combination of helix (h) and coil (c) states, described by

$$h\ h\ h\ h\ h\ c\ c\ h\ h\ c\ c\ c \qquad \text{(eqn. 5. 1)}$$

contribute to the partition function as

$$w\sigma\ w\sigma\ w\sigma\ v\ v\ u\ \alpha u\ v\ v\ u\ u\ u \qquad \text{(eqn. 5. 2)}$$

The probability that the *ith* residue in the $(Ala)_n$ chain is found in the α-helical conformation, *P(n, i)*, is given by

$$P(n,i) = F(n,i)/Z(n) \qquad \text{(eqn. 5. 3)}$$

where *Z(n)* is the partition function (eqn. 5.4) and *F(n, i)* is the conditional partition function (eqn. 5.8). The partition function *Z(n)* is the sum over all combinations of h and c states of alanines in the $(Ala)_n$ system, given by

$$Z(n) = eW^n e^+ \qquad \text{(eqn. 5. 4)}$$

where e and e^+ are, respectively, row and column vectors that represent terminal residues of the helical chains (eqn. 5.5-5.6),

$$e = [1 \quad 1 \quad 1] \qquad \text{(eqn. 5. 5)}$$

$$e^+ = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \qquad \text{(eqn. 5. 6)}$$

and each alanine is represented by the matrix, *W*. Furthermore, W_i gives the statistical weight of the *ith* residue in the $(Ala)_n$ sequence when the helix (h) or coil (c) states of the neighboring alanines at the *i+1* and *i+2* positions are known, as described earlier, by

$$W_i = \begin{array}{cc|ccc} & i+2 & h & c & c\cup h \\ i & i+1 & h & h & c \\ \hline h & h & w\sigma & v & 0 \\ h & c & 0 & 0 & v \\ c & c\cup h & \alpha u & u & u \end{array} \qquad \text{(eqn. 5. 7)}$$

, where the notation $c \cup h$ means c or h.

The evaluation of *Z(n)* is facilitated by a similarity transformation that diagonalizes the matrix *W*, such that the conditional partition function, *F(n, i)*, is

$$F(n,i) = eW^{i-1}HW^{n-i}e^+ \qquad \text{(eqn. 5. 8)}$$

where

$$H = \begin{bmatrix} w\sigma & v & 0 \\ 0 & 0 & v \\ 0 & 0 & 0 \end{bmatrix} \qquad \text{(eqn. 5. 9)}$$

for the *ith* residue in a sequence of $(Ala)_n$.

Equations (5.3-5.9) are used to calculate the probability of each residue in the $(Ala)_n$ chain to be in the α-helical conformation. Alanines that are in the middle of the chain have higher helical propensities than those in the termini; hence, their statistical values approach 1 (or 100%). In contrast, terminal

residues have low tendencies to be in the α-helical structure, and their probabilities approach 0, corresponding to random coil. For example, the helical probability for each of the 14 alanines in $(Ala)_{14}$ is dependent on its relative position in the chain (**Figure 5.7**). Specifically, the probabilities for central residues, i = 3-12, to reside in the helical conformation are > 50%, whereas terminal alanines, i = 1, 2, 13, and 14, are found in this structure only 25-45% of the time.

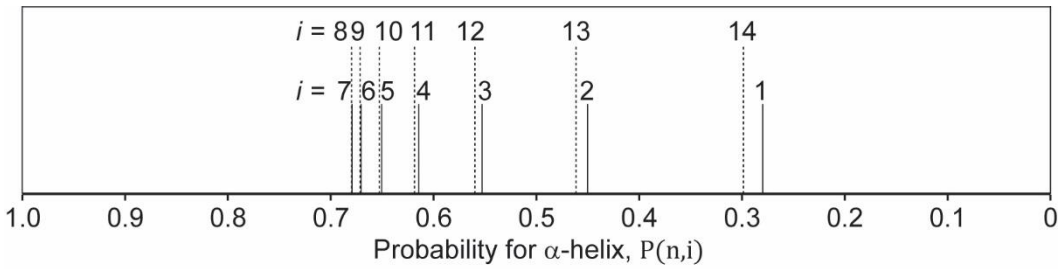


Figure 5.7 Probability of each alanine in $(Ala)_{14}$ to be in a helical conformation, as calculated by Bixon-Scheraga-Lifson's theory [15]. Horizontal axis represents the probability of each residue at the *ith* position of $(Ala)_{14}$ to be in α-helix. The highest probability corresponds to 1.0 ($P(14, i)$=100%), and the lowest probability corresponds to 0 ($P(14,i)$=0). Solid vertical lines (|) mark the helical probabilities for N-terminal residues with i = 1 to 7, whereas dotted vertical lines (¦) indicate those for C-terminal residues with i = 8 to 14. The height of vertical line is selected arbitrarily.

By Ullman's method [27], the ^{13}C NMR peak of a particular poly(Ala) sequence can be simulated from the result of the helical probability calculation (above). The simulated lineshape reflects the propensity for each residue in the peptide chain to be found in an α-helix. The chemical shift of the *ith* residue, δ_{ni}, in the $(Ala)_n$ is given by

$$\delta ni = \delta h\, P(n,i) + \delta c\, (1 - P(n,i)) \qquad \text{(eqn. 5. 10)}$$

with resonance frequencies δ_h and δ_c for the helix and coil, respectively. The intensity of the peak $I(n, i)\delta$ corresponding to the *ith* residue at a frequency δ is then given by

$$I(n,i)\delta = A^{-1}\, exp(-(\delta - \delta ni)^2 / 2A^2) \qquad \text{(eqn. 5. 11)}$$

where A corresponds to the Gaussian linewidth. The spectral intensity, $I(n)\delta$, is therefore given by

$$I(n)\delta = \sum_i I(n,i)\delta \qquad \text{(eqn. 5. 12)}$$

The ^{13}CO lineshape for a poly(Ala) chain can be calculated using equations 5.9-5.11. The carbonyl lineshape for $(Ala)_{14}$, for instance, is characterized by a peak centered at ~178.3 ppm and a shoulder at ~176.5 ppm (**Figure 5.8**). This spectrum reflects the sum of the Gaussian lines that are predicted for alanines in the sequence.

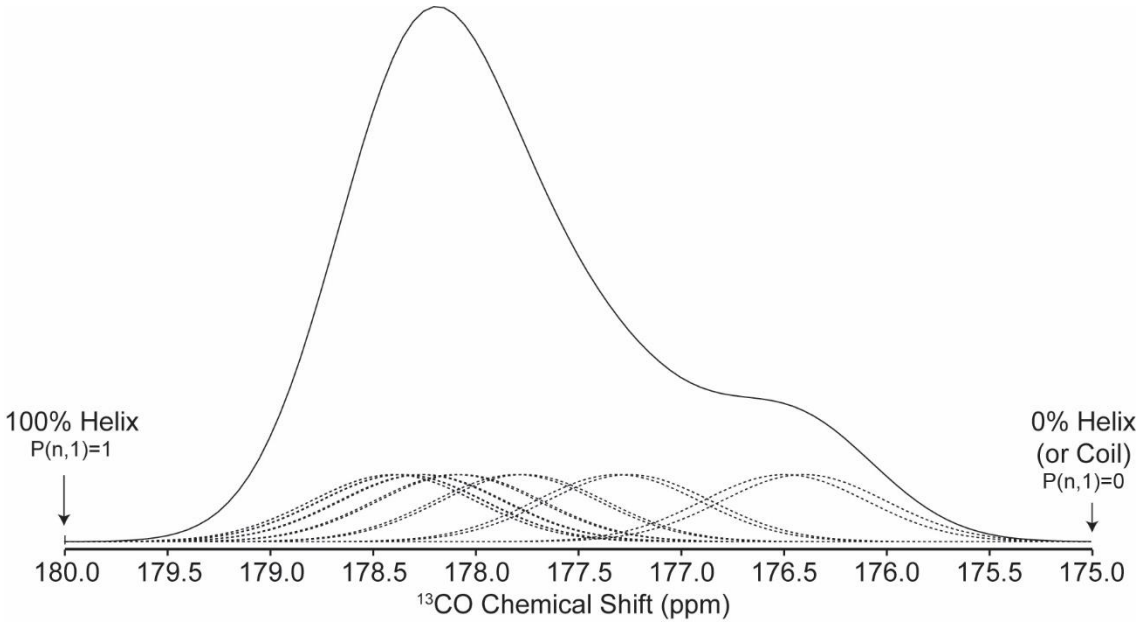


Figure 5.8 ^{13}CO lineshape for $(Ala)_{14}$ sequence. The predicted lineshape (*solid black*) is obtained from the sum of Gaussian lines (*dashed*), each of which corresponds to an alanine in $(Ala)_{14}$, as calculated using equation 5.11. The position of each Gaussian line also reflects the helical probability of each residue at the *ith* position, determined using equation 5.9. The highest probability for α-helix 100% ($P(n, i)=1$) and the lowest probability ($P(n,i)=0$), coil, would correspond to ^{13}CO resonances of 180 and 175 ppm, respectively, which are based on experimental values [31, 42]. Bixon-Scheraga-Lifson's approach [15] was used with the statistical weights, w, v, and σ of 1, 0.03, and 1.7, respectively. The parameter for Gaussian linewidth, A, is 0.08.

To account for polymer polydispersity, i.e., a mixture of polypeptide sequences with various lengths, the total intensity is calculated by

$$I\delta = \sum_n G(n) I(n) \delta \qquad \text{(eqn. 5. 13)}$$

where $G(n)$ is the weight fraction of the n-mer. For instance, the simulated carbonyl signal for a 1:1:1 mixture of $(Ala)_{10}$, $(Ala)_{14}$, and $(Ala)_{20}$ is reflected by the sum of each of the corresponding ^{13}CO lineshapes (**Figure 5.9**).

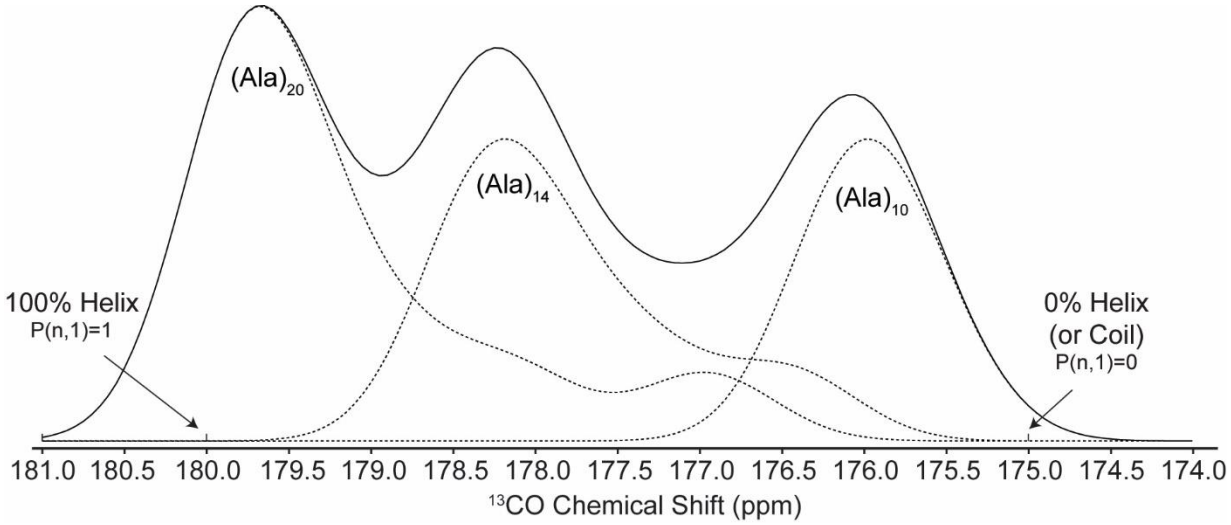


Figure 5.9 ^{13}CO lineshape representing the mixture of Ala-rich sequences, in which $(Ala)_{10}$, $(Ala)_{14}$, and $(Ala)_{20}$ are present at the 1:1:1 ratio. The predicted lineshape (*solid black*) is obtained from the sum of each line (*dashed*) corresponding to $(Ala)_n$ sequences with n=10, 14, and 20. The highest probability for α-helix ($P(n, i)=1$) is assigned to 180 ppm, and the lowest probability ($P(n,i)=0$), coil, is 175 ppm [28, 29]. Bixon-Scheraga-Lifson's approach [15] was used with the statistical weights, w, v, and σ of 1, 0.03, and 1.7, respectively. The parameter for each Gaussian linewidth, A, is 0.08.

Simulations of the ^{13}CO lineshape that represents elastin's Ala-rich crosslinking domains require the polydisperse case of $(Ala)_n$ sequences. The ^{13}CO lineshape above T_g was reproduced by utilizing short and long chains with n ranging from 2 to 14. In contrast, the carbonyl lineshape above T_g was reproduced with only long sequences, n=8-18. The statistical weights for the partition function, v and σ, were set to 0.03 and 1.7 respectively, as described elsewhere [3, 43, 44]. The statistical weights for both random coil (v) and hydrogen-bonded helical state (w) were both 1.0 [3]. The Gaussian linewidth parameter (A) was set to 0.30 and 0.35 for the simulation of lineshapes above and below T_g, respectively. The ^{13}CO resonances of alanines in α-helix, δh, and coil, δc, were 180 and 175 ppm, respectively, which are estimates of the experimental values [31-33, 42].

C. RESULTS AND DISCUSSIONS

Solid-state NMR studies on hydrated NRSMC elastin with isotopic enrichment at its alanines were performed at temperatures ranging from 37 to -20 °C. The effect of temperature on the structure and dynamics of Ala was investigated (section C.1). The detection of ^{13}CO-Ala signals that correspond to the Ala-rich crosslinking domains was done using a DCP experiment at -20 °C (section C.2). Two-dimensional ssNMR correlation spectroscopy was used to identify the structural motifs of Ala in the hydrophobic domains (section C.3). The helix-coil transition theory was used to calculate the α-helical propensity of alanines in the Ala-rich crosslinking domains and to account for the ^{13}CO ssNMR lineshapes above and below T_g (section C.4).

C.1 Temperature-dependent features in ^{13}C DPMAS and CPMAS spectra indicate conformational and dynamical transitions in alanines

^{13}C-Ala resonances were previously used to identify helix-coil interconversions in wild silkworm silk fibroin studies [1, 3]. Namely, ^{13}CO-Ala chemical shifts are typical of an α-helix at lower temperatures. Upon warming, the peaks move upfield to the region expected for a random coil.

To determine whether a similar temperature-dependent transition is present in elastin, ^{13}C DPMAS spectra of [U-^{13}C]Ala NRSMC elastin were acquired at temperatures ranging from 37 to -20 °C. As introduced in Chapter 3, the DP experiment is effective for detection of all ^{13}C signals, regardless of structure and relative mobility.

Distinctly different ^{13}C lineshapes are observed above and below the freezing point of water (**Figure 5.10**). At 37 °C, the ^{13}CO lineshape has its tallest point at 175.6 ppm (FWHM~160 Hz), with a shoulder at ~177 ppm. For ^{13}Cβ, similar features are observed at 17.3 and ~16.6 ppm. ^{13}Cα appears to be more homogeneous, with its peak at 50.5 ppm (FWHM~90 Hz) with no clearly visible shoulder. These features are consistent across the spectra acquired at 20, 10, and 0 °C. However, all ^{13}C peaks are drastically altered below the freezing point of water. At -10 and -20 °C, the ^{13}CO lineshape is broad (FWHM~500-600 Hz). ^{13}Cα has a resolved feature at 52.6 ppm, with a shoulder at 50.5 ppm, and the lineshape is also broader. The asymmetric ^{13}Cβ lineshape has greater intensity in the upfield feature at ~16.0 ppm, in contrast to the spectrum at 37 °C.

The ^{13}C-Ala spectra reflect increased α-helical and decreased random coil content with sample cooling. I.e., as the temperature was reduced from 0 to -10 °C, signal intensities that correspond to alanines in α-helix, i.e., ~178, 52.6 and 16.0 ppm, increased. However, those expected for Ala in random

coil, i.e., 175.6, 50.5 and 17.4 ppm, decreased. The changes in the helix/coil content, as indicated by the lineshapes, are attributed to the coil-to-helix (conformational) transition between 0 and -10 °C.

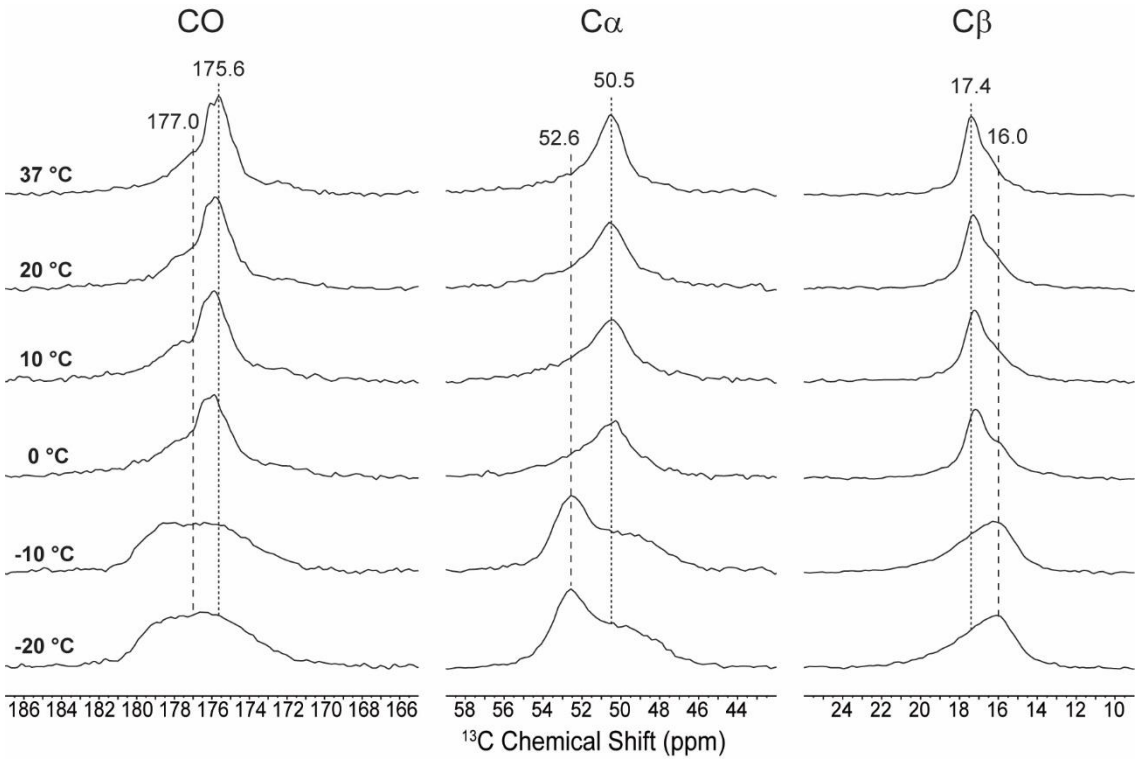


Figure 5.10 ^{13}C DPMAS spectra of [U-^{13}C]Ala hydrated elastin at 37, 20, 10, 0, -10 and -20 °C. Dotted lines (⋮) mark the peaks of the spectra acquired from 37 to 0 °C, and they represent the approximate ^{13}C-Ala resonances for random coil [32, 33]. Dashed lines (¦) mark the peaks of ^{13}C signals at -10 and -20 °C; 177 ppm is tentatively assigned to α-helical ^{13}CO Ala [30, 31].

The conformational transition in hydrated NRSMC elastin that occurs between 0 and -10 °C is correlated with the glass transition. This event is typically described as a morphological change of an amorphous polymer from molten rubber-like to a brittle “glassy” state as the sample is cooled below T_g [45-47]. The cross-polarization experiment is effective for the detection of motional changes from the “liquid-like” to “solid-like” (Chapter 3). Similar to the BNL study [7], the glass transition in NRSMC elastin is indicated by the sharp increase of CP signal intensities as the sample was cooled below 0 °C (**Figure 5.11**). ^{13}C CPMAS spectra acquired at 37, 20, 10 and 0 °C have low signal intensities, in contrast to those acquired at -10 and -20 °C. The stark contrast in CP intensities demonstrates that a glass transition occurs between 0 and -10 °C, consistent with DSC studies on hydrated BNL [48]. During this transition, the network of hydration water is immobilized [49], and the protein’s motion is greatly reduced, resulting in the enhancement of cross-polarization efficiencies. This result also shows two motional regimes in hydrated NRSMC elastin, similar to those observed in hydrated BNL [7].

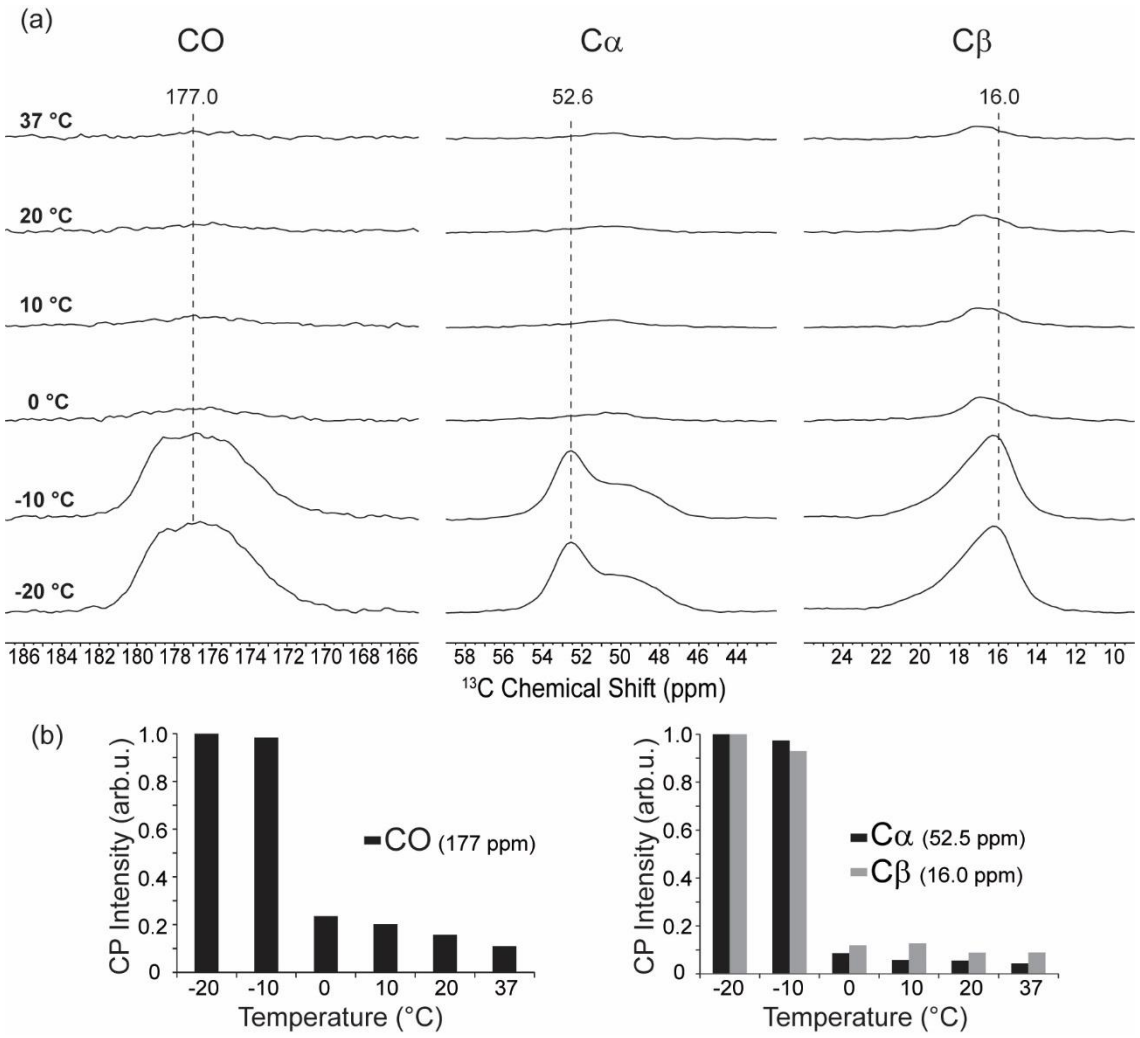


Figure 5.11 Variable-temperature ^{13}C CPMAS experiments of hydrated [U-^{13}C]Ala elastin. (a) ^{13}C CPMAS spectra at 37, 20, 10, 0, -10 and -20 °C. (b) Signal intensities of the carbonyl (left) and aliphatic peaks (right) as a function of temperature. Dashed lines (┆) mark the peaks of ^{13}C signals at -10 and -20 °C; 177 ppm is tentatively assigned to α-helical ^{13}CO Ala [30, 31].

The isotropic ^{13}C chemical shifts are useful for identifying common secondary structures in proteins and polypeptides [30, 31]. For instance, the ^{13}CO, ^{13}Cα, and ^{13}Cβ resonances for alanines were used to distinguish the α-helix from the β-sheet conformation in a poly(Ala) peptide [30]. These resonances were also used to characterize the α-helix and the random coil Ala in hydrated NRSMC elastin at 37 °C (Chapter 3). Furthermore, temperature-dependence in the ^{13}C chemical shifts of Ala has been used as an indicator for a conformational transition in the poly(Ala) region in wild silkworm's silk fibroin [1]. Specifically, the ^{13}CO resonance corresponding to the α-helix at -5 °C, i.e., 177.0 ppm, was shown to move upfield to 175.7 ppm at 40 °C, which was the carbonyl chemical shift for random coil. This difference of ^{13}CO-Ala chemical shifts ($\Delta\delta$) was 1.3 ppm, which was consistent with the range of

secondary shift for alanines in α-helix, i.e., +1-3 ppm [50] (Chapter 3). Hence, the helix-to-coil transition of alanines was concluded from the reported VT study [3].

To identify a possible conformational transition occurring in hydrated NRSMC elastin above T_g, the ^{13}CO-Ala resonance was investigated at several temperatures, i.e., 37, 20, 10, 0, and -4 °C (**Figure A.19**, **Appendix 17**). At 37 °C, the carbonyl signal is characterized by a broad lineshape with multiple overlapping components, including a downfield intensity (at ~177 ppm) and upfield peaks at 175.9, 175.6 and 175.0 ppm. The broad feature at 177 ppm corresponds to the Ala-rich crosslinking domains, whereas the three overlapping (upfield) peaks correspond to alanines with R_{-1}-**Ala**-Gly, R_{-1}-**A**la-Z and R_{-1}-**Ala**-Ala sequences (Chapter 3). These features are generally conserved in the other spectra, acquired at 20, 10, 0 and -4 °C. However, the linewidth increases and the center of mass in each ^{13}CO intensity moves downfield with the decrease of temperature. Specifically, the peaks corresponding to Ala with both R_{-1}-**Ala**-Gly and R_{-1}-**Ala**-Ala motifs move slightly downfield, $\Delta\delta$=0.4-0.6 ppm, as the sample was cooled from 37 to -4 °C.

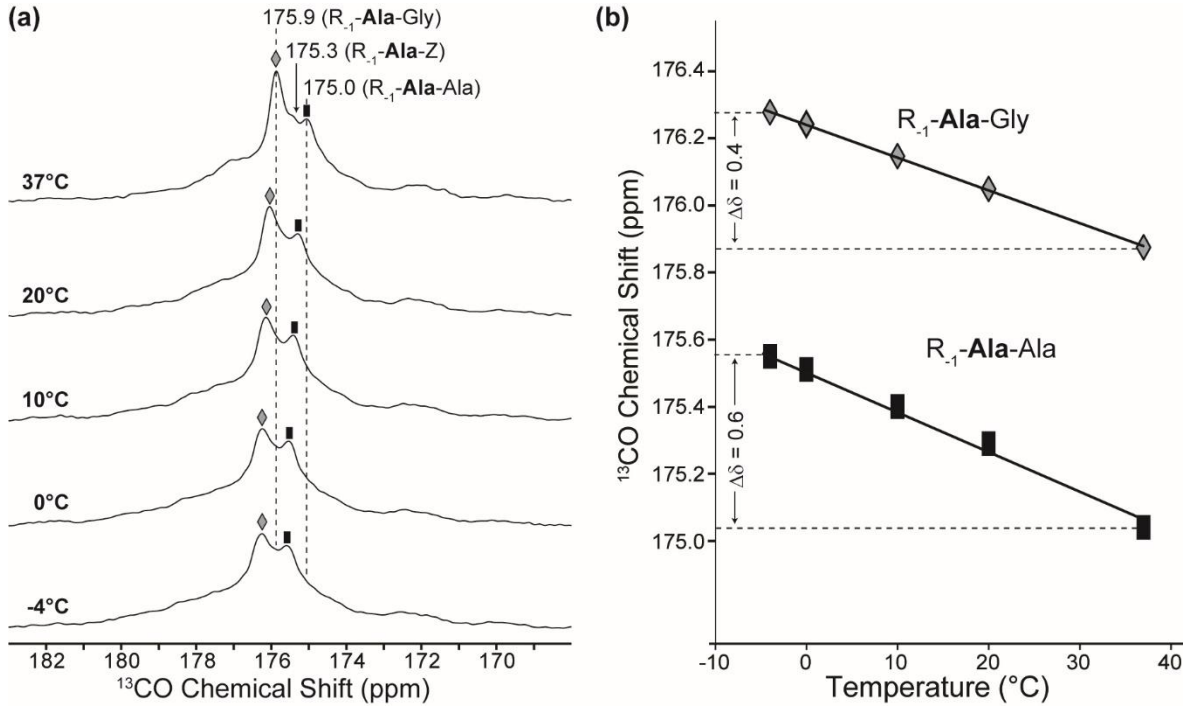


Figure 5.12 (a) VT spectra of ^{13}CO DP with LOW-BASHD homonuclear decoupling, and (b) ^{13}CO chemical shifts of alanines, R_{-1}-Ala-Gly (◊) and R_{-1}-Ala-Ala (▮), as a function of temperature. LOW-BASHD decoupling was implemented to remove the line broadening effect from *J*-couplings and to simplify peak assignments. Dotted lines (┆) on the VT spectra indicate the ^{13}C chemical shifts assigned for Ala in the R_{-1}-**Ala**-Gly, R_{-1}-**Ala**-Z (Z≠Ala,Gly) and R_{-1}-**Ala**-Ala motifs. The $\Delta\delta$ values in the plot of temperature-dependent ^{13}CO resonances (5.12b) represent the difference in chemical shifts between the spectrum acquired at 37 °C and that at -4 °C.

The temperature-dependence for alanines with the R_{-1}-**Ala**-Ala motif suggests a coil-to-helix transition with sample cooling (**Figure 5.12b**). However, $\Delta\delta$=0.6 ppm, is smaller than the 1-3 ppm secondary shift typically observed for an α-helix [50], indicating that the random coil Ala at 37 °C did not fully convert to the α-helical structure at -4 °C. This result supports the high mobility of Ala above T_g. Furthermore, residues at the termini of Ala-rich sequences in the crosslinking domains are the ones most likely undergoing such motion (Chapter 4). Alanines have undergone the full extent of the helix-coil transition as the temperature is reduced below T_g (section C.2).

C.2 Features in the ^{13}CO-Ala spectra reflect the high α-helical content in alanines below T_g

To validate the conformational transition occurring above and below T_g, the ^{13}CO-Ala lineshapes acquired at 37 and -20 °C were examined (**Figure 5.13a,b**). As described earlier, at 37 °C (above T_g), alanines are found in the α-helices and random coil, as shown by the ^{13}C DP intensities at 177.0 and ~175-176 ppm, respectively. In addition, signal intensities for the random coil alanines are higher than those for the α-helix. However, at -20 °C (below T_g), the ^{13}CO CP lineshape shows the relative decrease of random coil intensity and the appearance of a downfield signal with a peak at 178.6 ppm, also corresponding to the α-helix [29]. This CP lineshape reflects the population of all Ala in both hydrophobic and crosslinking domains.

The CP-based ^{1}H→^{15}N→^{13}CO transfer at -20 °C is used to acquire carbonyl intensities of alanines with the R_{-1}-**Ala**-Ala motif (section A.3.1), which are primarily found in the crosslinking regions. The SPECIFIC-CP experiment is selective for ^{13}CO, so ^{13}Cα is suppressed [36]. The ^{13}CO-Ala DCP spectrum is characterized by an asymmetric lineshape with intensities that are distributed over a range of ^{13}CO resonances from ~173 to ~180 ppm and a downfield center-of-mass (**Figure 5.13c**). The peak is observed at 178.6 ppm, which is assigned to Ala in the α-helical conformation [29]. Hence, the overall DCP lineshape shows that the α-helix is the primary conformation adopted by alanines in the crosslinking domains at -20 °C (below T_g), in contrast to the ^{13}C DP spectrum at 37 °C (above T_g). Furthermore, the DCP lineshape suggests that a coil-to-helix conversion of Ala was accomplished with the cooling of sample below T_g, resulting in an increase of α-helical content in the protein.

Temperature-dependence in the α-helical content of elastin's crosslinking domains are supported by the helix-coil transition theory (section A.2). This theory explains that longer $(Ala)_n$ chains are more likely to adopt α-helices, and each residue in the chain is found in the helical conformation. In addition, the high helical propensity of Ala in the Ala-rich sequences corresponds to a downfield resonance in the ^{13}CO NMR spectrum [27]. Furthermore, the downfield ^{13}CO signal also corresponds to the alanines that

are located in the central region of the helical chains [29]. For these reasons, the ^{13}CO DCP spectrum acquired below T_g (-20 °C) demonstrates the prevalence of α-helices at the Ala-Ala pairs. In contrast, the ^{13}CO DP lineshape above T_g (37 °C) suggests the presence of some α-helices and, mostly, random coil in the Ala-rich sequences of elastin's crosslinking domains. To validate these findings, simulations based on the helix-coil theory are used to reproduce the features of ^{13}CO-Ala lineshapes acquired above and below T_g (section C.4).

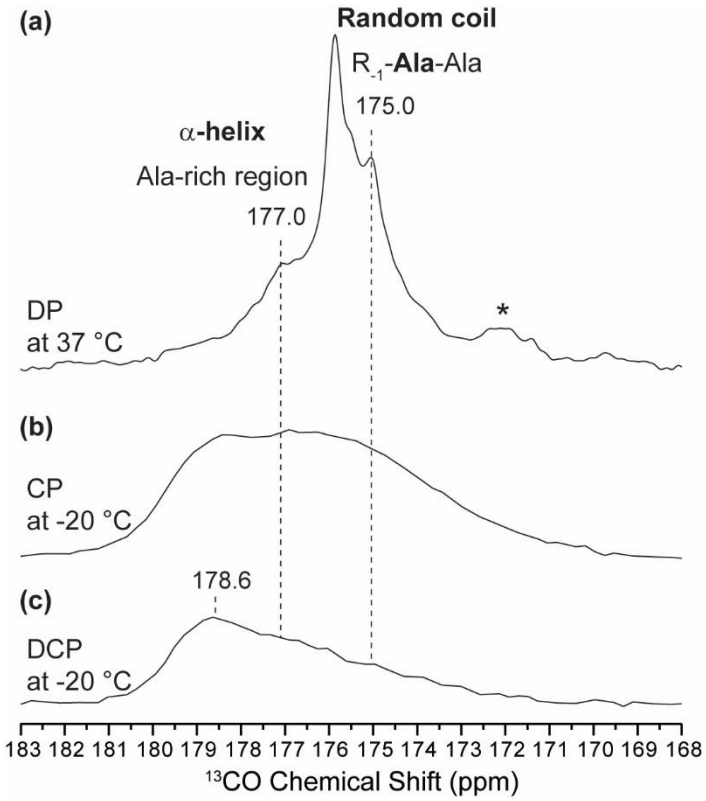


Figure 5.13 ^{13}CO-Ala lineshapes acquired by (a) DP at 37 °C and (b) CP, and (c) DCP at -20 °C. Dashed lines indicate carbonyl resonances for alanines in the Ala-rich region (α-**helix**) and in the R_{-1}-**Ala**-Ala motif (**random coil**) [29-33]; 178.6 ppm marks the peak for ^{13}CO signal acquired by DCP transfer, which primarily selects for alanines in the croslinking domains. SPECIFIC-CP [36] was also used to selectively detect ^{13}CO-Ala signal. Asterisk (*) indicates ^{13}C-Gly signal at natural abundance.

C.3 Homonuclear (^{13}C-^{13}C) correlation spectroscopy illustrate the conformational heterogeneity of alanines below T_g

Little has been known about the conformation of Ala in the hydrophobic domains of hydrated NRSMC elastin below T_g. The overlap of signals in the 1D ^{13}C CPMAS spectrum acquired at -20 °C makes structural assignments for Ala difficult via ^{13}C isotropic chemical shifts (section C.2). For this

reason, a two-dimensional approach was implemented in the study of hydrated elastin below T_g. Specifically, the CP-based 2D R-TOBSY experiment was performed at -20 °C to obtain ^{13}CO-$^{13}C\alpha$ and $^{13}C\alpha$-$^{13}C\beta$ chemical shift correlations. Such correlations are useful for distinguishing ^{13}C signals of alanines in the crosslinking domains from those in the hydrophobic regions.

The cross-peaks on the ^{13}CO-$^{13}C\alpha$ correlation spectrum include two populations of alanines, major and minor (**Figure 5.14**). The major population (~60%) has its highest point at 178.8 ppm (^{13}CO) and 52.5 ppm ($^{13}C\alpha$) (*red line*), but its intensity is distributed over a broad ^{13}CO range of ~175 to ~180 ppm. This component originates from Ala in the crosslinking domains, which composes ~58% of the total alanines in tropoelastin's primary structure. The minor population (~40%) is characterized by an overlap of intensities with ^{13}CO resonances ranging from ~173 to ~177 ppm and $^{13}C\alpha$ resonances ranging from ~48 to ~52 ppm. The highest intensities for the minor population are identified by two peaks, one with (^{13}CO, $^{13}C\alpha$) resonances of (175.9 ppm, 50.3 ppm) (*green line*) and another at (175.3 ppm, 51.3 ppm) (*magenta line*). An additional cross-peak is identified at 174.0 ppm (^{13}CO) and 49.0 ppm ($^{13}C\alpha$) (*blue line*).

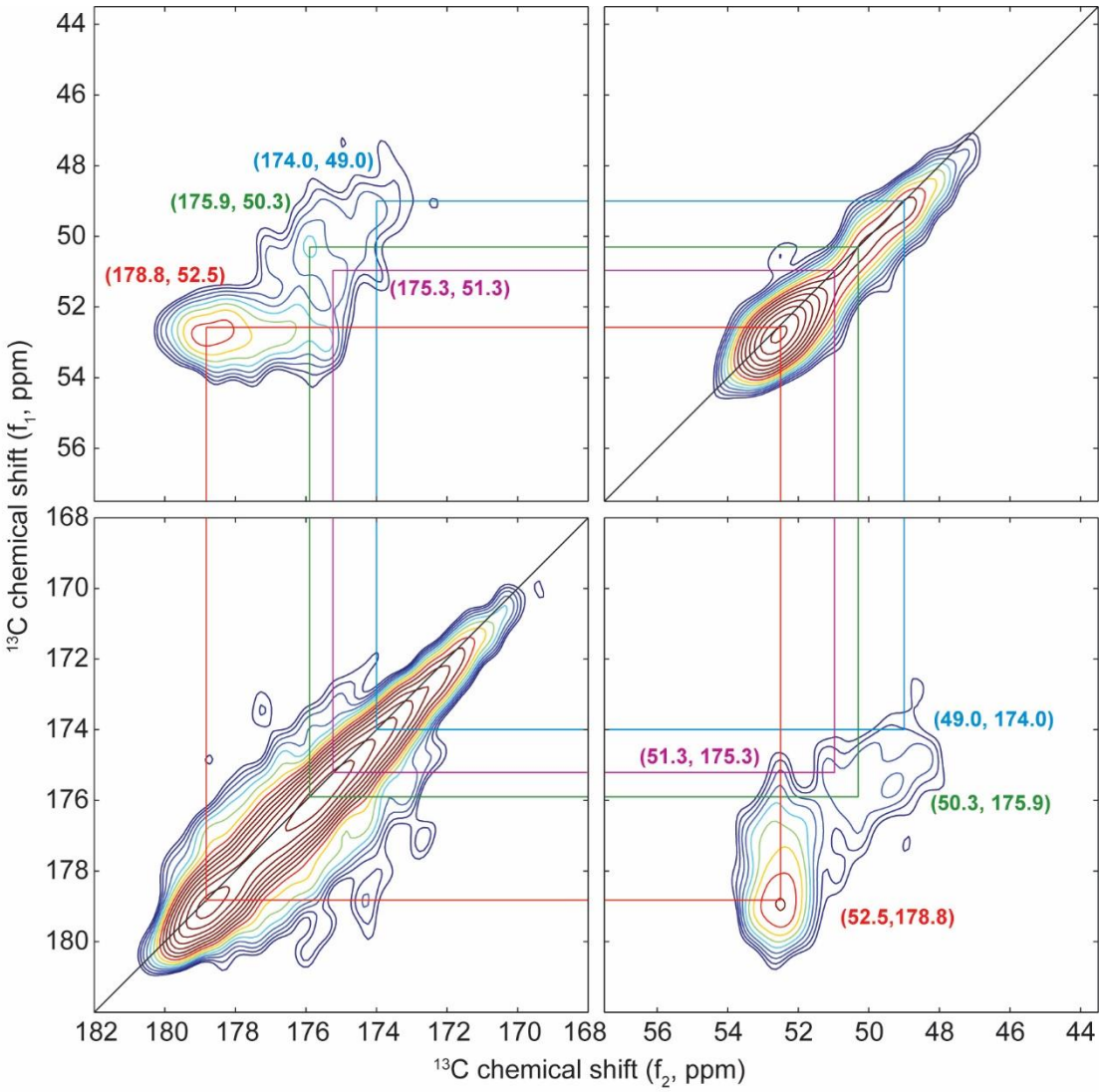


Figure 5.14 ^{13}CO-^{13}Cα correlations in the 2D R-TOBSY spectrum of [U-^{13}C]Ala elastin acquired at -20 °C. The experiment utilizes a CP-based excitation. The spectrum is processed with 20 Hz of line broadening in f_1 and f_2 dimensions. Colored lines highlight correlations for resolved alanine populations.

The cross-peaks in the ^{13}Cα-^{13}Cβ spectrum also reflect the presence of two alanine populations, i.e., major and minor (**Figure 5.15**), consistent with the ^{13}CO-^{13}Cα correlations. The major component is observed as concentric contours with a peak at 52.5 ppm (^{13}Cα) and 16.0 ppm (^{13}Cβ) (*red line*). The minor population of alanines is distributed over ^{13}Cα and ^{13}Cβ chemical shifts, respectively, ranging from ~48 to ~54 ppm and from ~14 to ~18 ppm. This population has at least three apparent peaks, with centers-of-masses located at the (^{13}Cα, ^{13}Cβ) resonances of (50.3, 17.1) (*green line*), (51.3, 16.5) (*magenta line*), and (49.0, 17.4) (*blue line*).

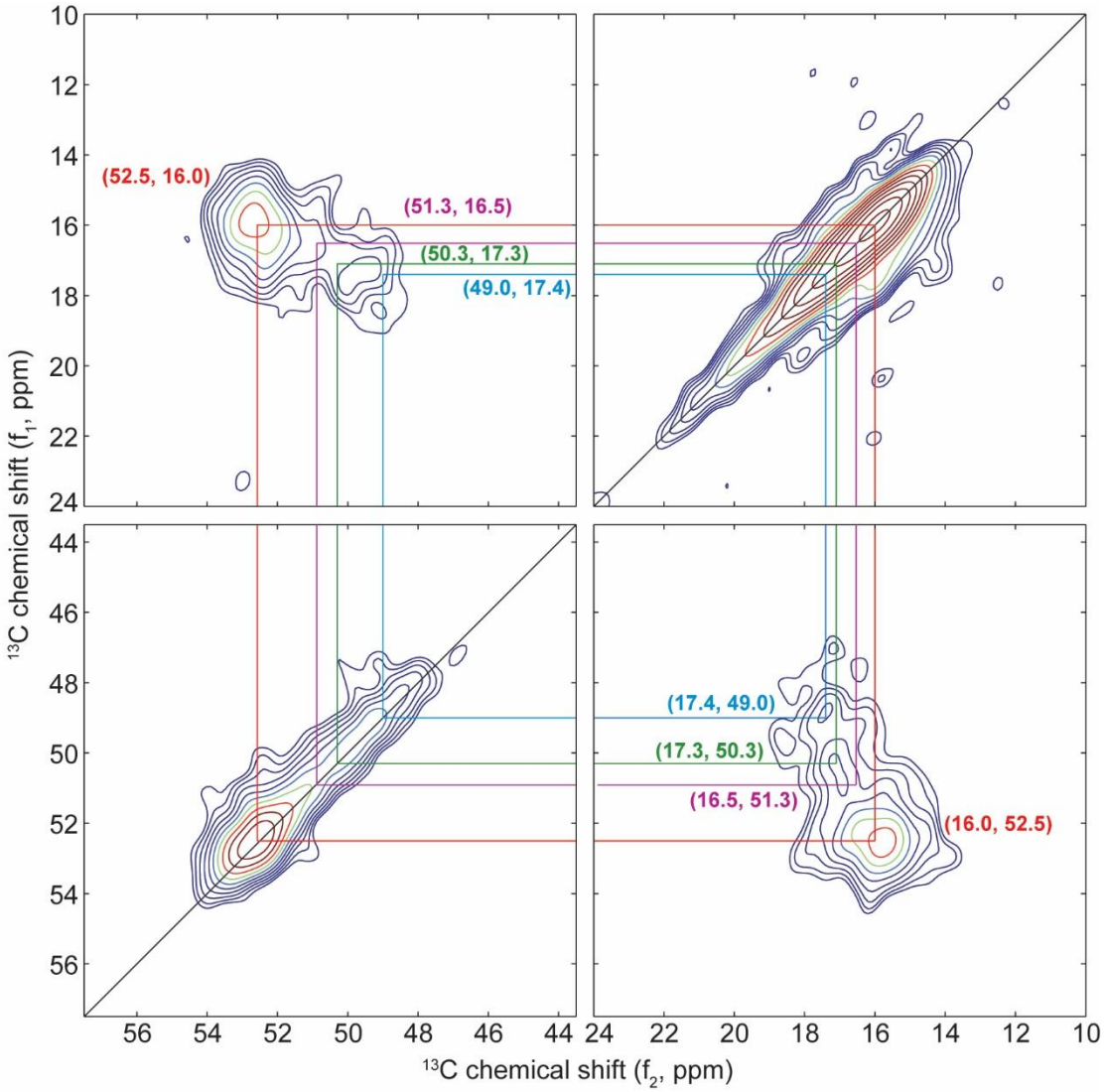


Figure 5.15 ^{13}Cα-^{13}Cβ correlations in the 2D R-TOBSY spectrum of [U-^{13}C]Ala elastin acquired at -20 °C. The experiment utilizes a CP-based excitation. The 2D spectrum is processed with 20 Hz of line broadening in f_1 and f_2 dimensions. Colored lines highlight correlations for resolved alanine populations.

The conformations of alanines below T_g were assessed based on the ^{13}CO-^{13}Cα and ^{13}Cα-^{13}Cβ correlations (above). The major secondary structure adopted by Ala in the crosslinking domains is the α-helix, as indicated by the ^{13}CO, ^{13}Cα, and ^{13}Cβ resonances of 178.8, 52.5, and 16.0 ppm, respectively. [30, 31]. In contrast, alanines in the hydrophobic domains have at least two structural motifs, random coil and PP-II. Random coil alanines are indicated by the cross-peak at 175.9 (^{13}CO), 50.3 (^{13}Cα) and 17.1 ppm (^{13}Cα) [32, 33], whereas the PP-II structure is identified by the intensity at 174.0 (^{13}CO), 49.0 (^{13}Cα) and 17.4 ppm (^{13}Cα) [51]. As summarized in **Table 5.1**, the three sets of ^{13}C-Ala in the hydrophobic

domains are assigned as the three motifs of alanines, R_{-1}-**Ala**-Gly, R_{-1}-**Ala**-Ala, and R_{-1}-**Ala**-Z (where Z ≠ Ala, Gly), occurring in tropoelastin's primary structure (Chapter 3).

Table 5.1. Chemical shifts of resolved alanines in elastin from 2D R-TOBSY measurements at -20 °C with tentative structural assignments. Alanines in the primary structure of tropoelastin are categorized into R_{-1}-**Ala**-Gly, R_{-1}-**Ala**-Ala, and R_{-1}-**Ala**-Z (where Z ≠ Ala, Gly) sequence motifs.

No	^{13}C δ (ppm)			Elastin Domain	Conformational Assignment	Sequence Motif
	CO	Cα	Cβ			
1	178.8	52.5	16.0	Crosslinking	α-helix [30, 31]	R_{-1}-**Ala**-Ala
2	175.9	50.3	17.1	Hydrophobic	Random coil [32, 33]	R_{-1}-**Ala**-Gly
3	175.3	51.3	16.5	Hydrophobic	Random coil [32, 33]	R_{-1}-**Ala**-Z
4	174.0	49.0	17.4	Hydrophobic	PP-II [51]	R_{-1}-**Ala**-Ala

C.4 Lineshape simulations of ^{13}CO-Ala support the temperature-dependence of α-helices in the crosslinking domains

Helix-coil transition theory was used to perform simulations of ^{13}CO lineshapes of alanines in the crosslinking domains. Specifically, the probability of each alanine in an Ala-rich sequence to be found in an α-helix is calculated using Bixon-Scheraga-Lifson's approach [15], and the corresponding NMR lineshapes are simulated using Ullman's method [27]. The crosslinking domains of elastin are characterized by tandem repeats of two or three Ala that are interspersed by other residues such as Lys or Tyr, i.e., KAAK or KAAAK. To simplify the calculation, such motifs were treated as continuous strands of alanines, $(Ala)_n$, where n represents the number of residues or the length of the chain.

The poly(Ala) sequence is a simplified yet effective model for simulations of the ^{13}CO-Ala lineshapes of elastin's crosslinking domains. The substitution of a non-Ala residue with Ala is justified by a few factors. First, the majority of Lys residues in native or a biosynthetically produced (NRSMC) elastin are modified into elastin's crosslinks [52]. These Lys-derived moieties, including other amino acids adjacent to them, promote the formation of α-helix (Chapter 4). Hence, the backbone conformation of these residues is very likely an α-helix, especially at -20 °C. Second, the statistical weights (i.e., w, v, σ and u) for Tyr and Lys typically used for helix-coil formulations are similar to those for Ala [44]. In other words, these amino acids have comparable helix-forming tendencies as alanine. Furthermore, a cation-pi interaction [53] between the sidechain of Lys (**K**) at the *ith* position and Tyr (**Y**) at the (*i*+*4*) in **K**AAK**Y** motif is believed to form prior to crosslinking, which stabilizes the α-helical structure in regions containing such sequence.

Elastin's crosslinking domains are characterized by the polydispersity of Ala-rich sequences, i.e., a mixture of $(Ala)_n$ sequences with assorted number of residues, n. To reproduce the acquired ^{13}CO-Ala lineshapes by simulations, numerous permutations of n (eqn. 5.12, section B.3) were considered in the

calculation. **Table 5.2** and **5.3** show the optimal combination of $(Ala)_n$ sequences to simulate the ^{13}CO-Ala lineshapes above and below T_g.

The domains encoded by exons 19, 21, 23 and 25 were considered as long chains of alanines, $(Ala)_n$ with n=12-18, for the simulations. These Ala-rich regions have distinctly high helical content [54]. Specifically, the stability of the α-helix in the domains encoded by EX21/23 was verified in computation [54] and experiment [55]. Additionally, EX19 and EX25 domains are known to be involved in elastin assembly, and they must both assume α-helical conformations for crosslinking [56].

A distribution of long $(Ala)_n$ sequences best reproduce the ^{13}CO-Ala lineshape below T_g. As shown in **Table 5.2**, non-Ala residues in the primary structures, e.g., Tyr (Y) and Lys (K), were substituted with Ala (A). This approach allowed for the use of a single continuous chain of alanines to represent each of elastin's crosslinking domain. For instance, Y and K in EX6 were replaced with A to form a strand of $(Ala)_{13}$. Such residue swaps, e.g., Y-to-A, K-to-A, etc., were performed for every sequence with at least 2 alanines repeated in tandem (AA). As a result, $(Ala)_n$ sequences with n ranging from 8 to 18 were used to represent alanines in the crosslinking domains and to simulate the carbonyl lineshape below T_g.

Table 5.2. Primary structures of tropoelastin's crosslinking domains and corresponding $(Ala)_n$ sequences for the simulation of ^{13}CO-Ala lineshape below T_g. Underlined amino acids in each domain represent a single continuous chain, in which non-Ala residues are substituted to alanine to form $(Ala)_n$ that ranges from 8 to 18 residues (n).

EX	Tropoelastin's Primary Structures	$(Ala)_n$	n
1	MAGLTAAVPQPGVLLILLLNLLHPAQPG	-	
4	ALGPGGKPPKP	-	
6	GGVLVPGGGAGAAAAYKAAAKA	$(Ala)_{13}$	13
8	GAVVPQLGAGVGAGGKPGKVP	-	
10	GARFPGVGVLPGVPTGTGVKAKVP	-	
12*	GVGPFGGQQPGVPLGYPIKAPKLP	-	
15*	GVGSQAAVAAAKAAKY	$(Ala)_{11}$	11
17	GAGTPAAAAAAKAAAKAAKY	$(Ala)_{15}$	15
19	GAVSPAAAAKAAAKAAKY	$(Ala)_{13}$	13
21	GASQAAAAAAAAKAAKY	$(Ala)_{16}$	16
23	GTPAAAAAAAAAKAAAKAGQY	$(Ala)_{18}$	18
25	GAGTPAAAKSAAKAAAKAQY	$(Ala)_{15}$	15
27	VPGSLAASKAAKY	$(Ala)_8$	8
29	GAPAAAAAAKAAAKAAQY	$(Ala)_{15}$	15
31	GVSPAAAAKAAKY	$(Ala)_9$	9
35	GKPPKPYGGALGALGYQ	-	
36	GGGCFGKSCGRKRK	-	

* Subject to alternative splicing

Skyline projections were generated as the representative experimental ^{13}CO-Ala lineshapes for alanines (**Figure 5.16a**). Each projection reflects the tallest points of the contour cross-peaks in the 2D ^{13}CO-^{13}Cα correlation spectrum. Two regions, 1 and 2 (red), were selected to represent alanines in the crosslinking and hydrophobic domains, respectively (section C.3).

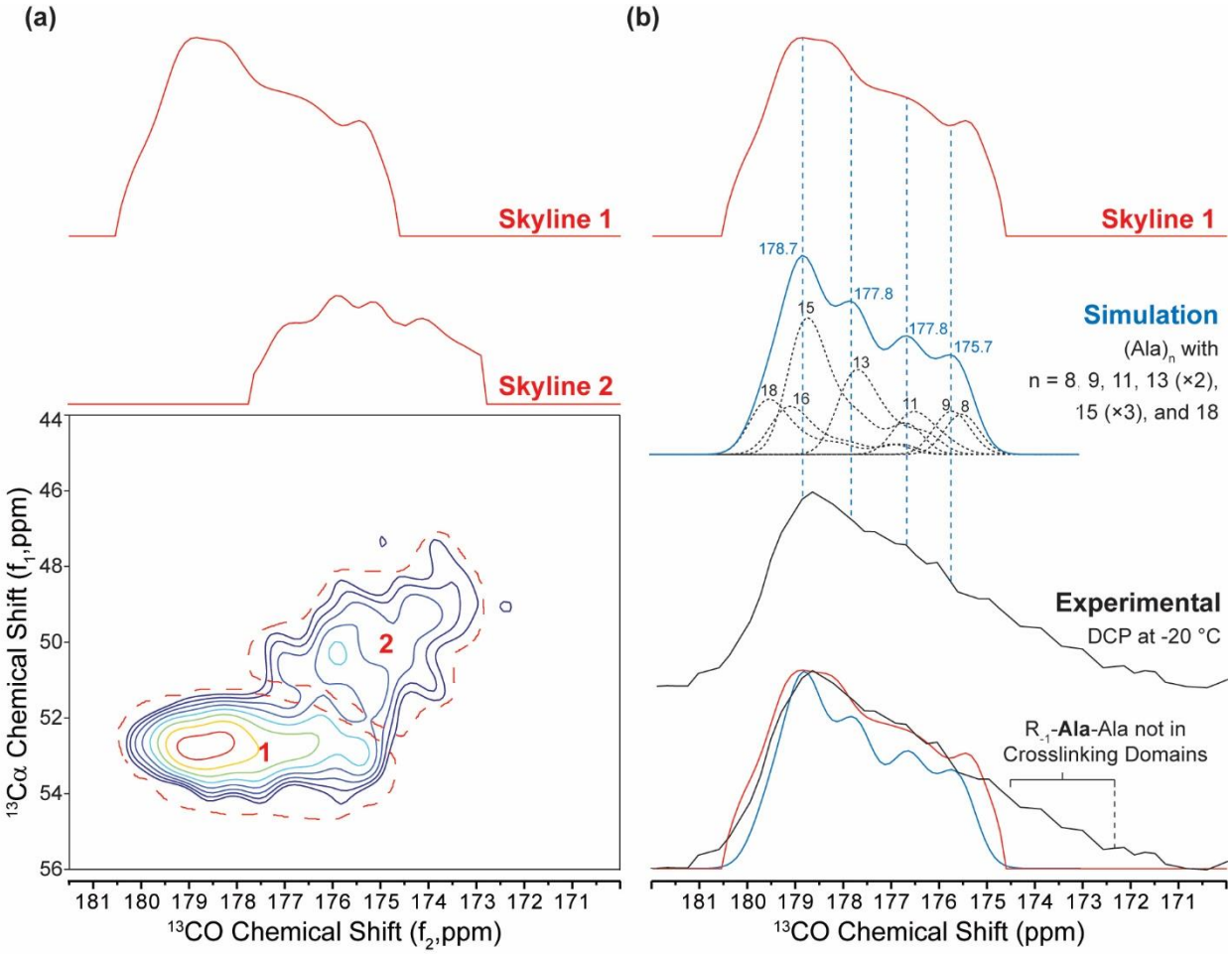


Figure 5.16 (a) Cross-peak from the 2D R-TOBSY spectrum at -20 °C and the skyline projections of the ^{13}CO intensities calculated from the two (red) outlined regions, 1 and 2; (b) comparison of the skyline projection of region 1 with the simulated ^{13}CO-Ala lineshape and the SPECIFIC-CP spectrum acquired at -20 °C (Fig. 5.13c). The skyline projection of each outlined region is calculated by multiplying the cross-peak intensities with a non-zero value (i.e., 1) and those outside of the area are suppressed to zero. Simulated lineshape in (b) was performed using Bixon-Scheraga-Lifson's helix-coil theory [15] and Ullman's method [27]. *n* values of 8, 9, 11, 13 (2×), 15 (3×), and 18 reflect the probable lengths of the $(Ala)_n$ sequence in elastin below T_g. Individual contributions for the designated values *n* for $(Ala)_n$ (Table 5.2) is illustrated by dashed lines (black) and the sum is shown by the solid line (blue). Vertical dashed lines (¦) indicate peak positions in the simulated lineshape. The bottom of Figure 5.16(b) shows the superposition of the ^{13}CO-Ala signals from the skyline projection (red), the simulation (blue), and the SPECIFIC-CP experiment (black); the tallest peak is scaled to the maximum intensity of the experimental spectrum, at ~178.6 ppm.

The simulated ^{13}CO lineshape for long $(Ala)_n$ sequences roughly match the skyline projection of region 1 (red), corresponding to alanines in the α-helical conformation. As shown in **Figure 5.16b**, the highest peak is located at ~178.7 ppm, which arises from the sum of ^{13}CO signals of long Ala-rich sequences such as $(Ala)_{15}$, $(Ala)_{16}$, and $(Ala)_{18}$. Contributions from other $(Ala)_n$ sequences with $n \leq 13$ are shown by three discernable peaks that are located at 177.8, 176.8 and 175.7 ppm, respectively, in successively decreasing intensities.

The lineshape simulation also helps identify ^{13}CO-Ala signals that are found outside of the crosslinking domains. The bottom of **Figure 5.16b** shows the superimposition of ^{13}CO-Ala lineshapes acquired from the skyline projection of region 1 (red), the simulation (blue), and the DCP experiment at -20 °C (black). A discrepancy in signal intensities is observed at ~173-174 ppm, in which the simulated lineshape and the skyline projection do not show any signals at such frequencies. This upfield intensity is attributed to the minor population of alanines with the R_{-1}-**Ala**-Ala motif that are found outside of the crosslinking regions. Specifically, the signal originates from tandem repeats of Ala, i.e., AA, in the hydrophobic domains (Chapter 1), which are also detected by the DCP experiment.

Short sequences best reproduce the ^{13}CO-Ala lineshape above T_g. With the exception of exons 19, 21, 23 and 25 (described earlier), all Ala-rich crosslinking domains each contain two or three $(Ala)_n$ sequences (**Table 5.3**). For instance, EX6 is comprised of two short strands of alanines, i.e., $(Ala)_4$ and $(Ala)_3$. Overall, $(Ala)_n$ sequences with n ranging from 2 to 14 were used to represent the tandem repeats of alanines in the crosslinking domains and to simulate the carbonyl lineshape above T_g.

Table 5.3. Primary structures of tropoelastin's crosslinking domains and corresponding $(Ala)_n$ sequences for the simulation of ^{13}CO-Ala lineshape above T_g. Underlined amino acids in each domain represent a single continuous chain, in which non-Ala residues are substituted to alanine to form $(Ala)_n$ that ranges from 2 to 14 residues (n).

EX	Tropoelastin's Primary Structures	$(Ala)_n$	n
1	MAGLTAAVPQPGVLLILLLNLLHPAQPG	-	
4	ALGPGGKPPKP	-	
6	GGVLVPGGGAGAAAAYKAAAKA	$(Ala)_4$, $(Ala)_3$	4,3
8	GAVVPQLGAGVGAGGKPGKVP	-	
10	GARFPGVGVLPGVPTGTGVKAKVP	-	
12*	GVGPFGGQQPGVPLGYPIKAPKLP	-	
15*	GVGSQAAVAAAKAAKY	$(Ala)_2$, $(Ala)_3$, $(Ala)_2$	2(×2), 3
17	GAGTPAAAAAAKAAAKAAKY	$(Ala)_{11}$, $(Ala)_2$	11, 2
19	GAVSPAAAAKAAAKAAKY	$(Ala)_{12}$	12
21	GASQAAAAAAAAKAAKY	$(Ala)_{12}$	12
23	GTPAAAAAAAAAKAAAKAGQY	$(Ala)_{14}$	14
25	GAGTPAAAKSAAKAAAKAQY	$(Ala)_{13}$	13
27	VPGSLAASKAAKY	$(Ala)_2$, $(Ala)_2$	2(×2)
29	GAPAAAAAAKAAAKAAQY	$(Ala)_6$, $(Ala)_3$, $(Ala)_2$	6,3,2
31	GVSPAAAAKAAKY	$(Ala)_4$, $(Ala)_2$	4,2
35	GKPPKPYGGALGALGYQ	-	
36	GGGCFGKSCGRKRK	-	

* Subject to alternative splicing

The simulated lineshape for short $(Ala)_n$ sequences is comparable with the carbonyl spectrum above T_g. Two major features characterize the simulated lineshape, i.e., a broad downfield lineshape with a narrower upfield peak (**Figure 5.17a**). The narrow component at ~175 ppm arises from short alanine chains such as $(Ala)_2$, $(Ala)_3$, and $(Ala)_4$, which forms random coil. The broad component with a center-of-mass at ~177 ppm encompasses lineshapes for $(Ala)_n$ sequences where n=11-14, which indicates α-helical alanines. As illustrated in **Figure 5.17b**, the experimental ^{13}CO-Ala lineshape (black) roughly corresponds to the simulated one (blue). The broad ^{13}C signal at ~177 ppm is comparable to the major feature in the simulation, and the resolved intensity at 175.0 ppm which corresponds to the R_{-1}-**Ala**-Ala motif is also similar to the narrow peak in the simulation. The peak at ~176 ppm in the experimental ^{13}CO spectrum corresponds to alanines in the hydrophobic regions with R_{-1}-**Ala**-Gly and R_{-1}-**Ala**-Z motifs in random coil (Chapter 3), which is excluded from the lineshape simulation. A discrepancy in signal intensities was also observed at other resonances, such as ~174 ppm, because the ^{13}C spectrum above T_g was obtained using DP instead of the selective SPECIFIC-CP (below T_g). Hence the ^{13}C DP lineshape includes all alanine signals, not only those from the R_{-1}-**Ala**-Ala motif. In conclusion, the agreement between the simulated and the experimental lineshapes demonstrates that short $(Ala)_n$ sequences were

necessary to reproduce the ^{13}CO signal for alanines above T_g, which represent the random coil component.

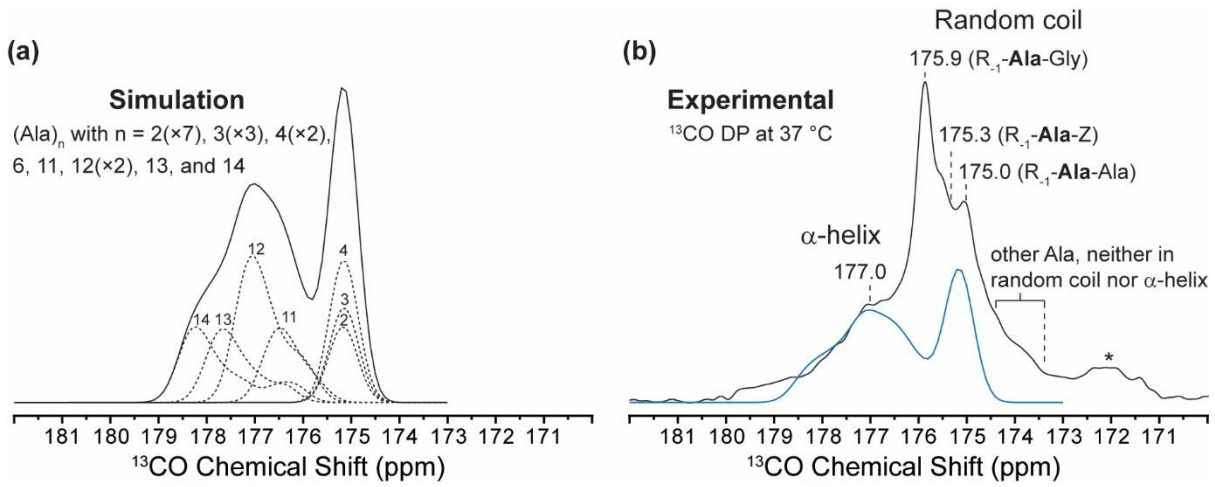


Figure 5.17 (a) Simulated ^{13}CO-Ala lineshapes and (b) comparison of simulation with the ^{13}CO DP spectrum acquired at 37 °C, above T_g. Bixon-Scheraga-Lifson's helix-coil theory [15] and Ullman's method [27] were used to perform the simulation of (a). *n* values of 2 (7×), 3 (3×), 4 (3×), 6, 11, 12 (3×), 13, and 14 reflect the probable lengths of the $(Ala)_n$ sequence in elastin above T_g. In (a), dashed lines represent the contribution of each $(Ala)_n$, and the solid line is their sum. Figure 5.15 (b) gives the experimental (black solid line) and simulated (blue) lineshapes, scaled to the intensity of the feature at 177.0 ppm. Asterisk (*) indicates ^{13}C-Gly signal at natural abundance.

Variable-temperature ssNMR measurements of hydrated NRSMC elastin, above and below T_g, demonstrate the conformational heterogeneity of the alanines. Above T_g, a temperature-dependent conformational equilibrium between α-helix and random coil exists in the alanines. Higher α-helical content and reduced coil populations are observed at lower temperatures, i.e., below T_g. The heterogeneity of structural motifs is also observed for alanines in the hydrophobic domains. Both random coil and PP-II conformations are identified in these regions, which are consistent with previous reports [57-59].

D. SUMMARY AND CONCLUSIONS

Solid-state NMR spectroscopy and MD simulations indicate the existence of conformational equilibria in native elastin (Chapter 4 & 5). Briefly, alanines in the crosslinking domains are found in α-helical and random coil populations, whereas those in the hydrophobic regions are prominently found as random coil. The effects of temperature on the structure and dynamics of alanines in elastin's domains were investigated by ssNMR.

Variable-temperature ssNMR spectroscopy sheds light into the effect of temperature on the dynamics of alanines in hydrated NRSMC elastin. The ^{13}C-Ala CP intensities below T_g are markedly higher than those above T_g, indicating a significant reduction of dipolar couplings. This reduction is likely due to the decrease of protein mobility, as the dynamical feature of the protein changes from a "liquid-like" to a "solid-like" upon cooling [7]. Hence, the ssNMR results indicate that large-amplitude motions in elastin are largely attenuated, as the protein is cooled below the glass transition temperature.

The VT study also provides insights into the effect of temperature on the conformation of alanines in hydrated NRSMC elastin. Above T_g, the ^{13}CO-Ala DP spectrum shows a narrow lineshape with a broad downfield feature, which indicates random coil and α-helical populations, respectively. However, below T_g, the intensity of the downfield feature is significantly enhanced, suggesting the increase of α-helical content upon cooling. Additionally, the ^{13}CO-Ala DCP lineshape and 2D ^{13}C-^{13}C correlation spectra show that most of the alanines in the crosslinking domains are α-helical. In contrast, the hydrophobic domains of elastin below T_g feature a mixture of conformations, as indicated by ^{13}C signals for alanines in random coil and PP-II motif.

The helix-coil transitions theory provides an explanation to the variation of α-helical content in the crosslinking domains of native elastin. In general, this theory predicts the α-helical propensity for a given poly(Ala) peptide, and this propensity corresponds to a particular chemical shift in the NMR spectrum. Simulations of NMR lineshapes employing Bixon-Scheraga-Lifson's helix-coil theory [15] and Ullman's approach [27] were used to predict the ^{13}CO-Ala lineshapes above and below T_g. A mixture of short and long $(Ala)_n$ sequences best reproduce the ^{13}CO-Ala lineshape above T_g. In contrast, below T_g, the carbonyl lineshape is best reproduced by a distribution of longer $(Ala)_n$ sequences. Therefore, the results indicate that the helical content in elastin's crosslinking domains depends on the length of the α-helical chains. Shorter helices are formed at warmer temperature, because elastin undergoes fast and large-amplitude motions. However, longer helical segments are adopted below T_g, because the dynamics of the protein are largely attenuated, which is presumably due to the absence of translational diffusion of water below T_g [60, 61].

E. REFERENCES

1. Asakura, T., and T. Murakami, *NMR of silk fibroin. 4. Temperature- and urea-induced helix-coil transitions of the -(Ala)n- sequence in Philosamia cynthia ricini silk fibroin protein monitored by carbon-13 NMR spectroscopy.* Macromolecules, 1985. **18**(12): p. 2614-2619.
2. Asakura, T., H. Yoshimizu, and F. Yoshizawa, *NMR of silk fibroin. 9. Sequence and conformation analyses of the silk fibroins from Bombyx mori and Philosamia cynthia ricini by 15N NMR spectroscopy.* Macromolecules, 1988. **21**(7): p. 2038-2041.
3. Asakura, T., H. Kashiba, and H. Yoshimizu, *NMR of silk fibroin. 8. Carbon-13 NMR analysis of the conformation and the conformational transition of Philosamia cynthia ricini silk fibroin protein on the basis of Bixon-Scheraga-Lifson theory.* Macromolecules, 1988. **21**(3): p. 644-648.
4. Doster, W., *The dynamical transition of proteins, concepts and misconceptions.* European Biophysics Journal, 2008. **37**(5): p. 591-602.
5. Doster, W., S. Busch, A.M. Gaspar, M.S. Appavou, J. Wuttke, and H. Scheer, *Dynamical transition of protein-hydration water.* Phys Rev Lett, 2010. **104**(9): p. 098101.
6. Doster, W., *The protein-solvent glass transition.* Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 2010. **1804**(1): p. 3-14.
7. Perry, A., M.P. Stypa, B.K. Tenn, and K.K. Kumashiro, *Solid-state ^{13}C NMR reveals effects of temperature and hydration on elastin.* Biophysical Journal, 2002. **82**(2): p. 1086-1095.
8. Doty, P., A.M. Holtzer, J.H. Bradbury, and E.R. Blout, *Polypeptides. II. The configuration of polymers of γ-benzyl-L-glutamate in solution* Journal of the American Chemical Society, 1954. **76**(17): p. 4493-4494.
9. Doty, P., and J.T. Yang, *Polypeptides . VII. Poly-γ-benzyl-L-glutamate: the helix-coil transition in solution* Journal of the American Chemical Society, 1956. **78**(2): p. 498-500.
10. Doty, P., A. Wada, J.T. Yang, and E.R. Blout, *Polypeptides. VIII. Molecular configurations of poly-L-glutamic acid in water-dioxane solution.* Journal of Polymer Science, 1957. **23**(104): p. 851-861.
11. Blout, E.R., and R.H. Karlson, *Polypeptides. III. The synthesis of high molecular weight poly-γ-benzyl-L-glutamates.* Journal of the American Chemical Society, 1956. **78**(5): p. 941-946.
12. Katchalski, E., and M. Sela, *Synthesis and chemical properties of poly-α-amino acids.* Advances in Protein Chemistry, 1958. **13**: p. 243-492.
13. Matthews, B.W., *Structural basis of protein stability and DNA-protein interaction.* Harvey Lectures, 1985. **81**: p. 33-51.
14. Lifson, S., and A. Roig, *On the Theory of helix-coil transition in polypeptides.* Journal of Chemical Physics, 1961. **34**(6): p. 1963-1974.

15. Bixon, M., H.A. Scheraga, and S. Lifson, *Effect of hydrophobic bonding on the stability of poly-L-alanine helices in water.* Biopolymers, 1963. **1**(5): p. 419-429.
16. Poland, D.C., and H.A. Scheraga, *Statistical mechanics of noncovalent bonds in polyamino acids. III. Interhelical hydrophobic bonds in short chains.* Biopolymers, 1965. **3**(3): p. 305-313.
17. Vásquez, M., and H.A. Scheraga, *Effect of sequence-specific interactions on the stability of helical conformations in polypeptides.* Biopolymers, 1988. **27**(1): p. 41-58.
18. Doig, A.J., A. Chakrabartty, T.M. Klingler, and R.L. Baldwin, *Determination of free energies of N-capping in alpha-helixes by modification of the Lifson-Roig helix-coil theory to include N- and C-capping.* Biochemistry, 1994. **33**(11): p. 3396-3403.
19. Serrano, L., J.-L. Neira, J. Sancho, and A.R. Fersht, *Effect of alanine versus glycine in [alpha]-helices on protein stability.* Nature, 1992. **356**(6368): p. 453-455.
20. Marqusee, S., V.H. Robbins, and R.L. Baldwin, *Unusually stable helix formation in short alanine-based peptides.* Proceedings of the National Academy of Sciences, 1989. **86**(14): p. 5286-5290.
21. Williams, L., K. Kather, and D.S. Kemp, *High helicities of Lys-containing, Ala-rich peptides are primarily attributable to a large, context-dependent Lys stabilization.* Journal of the American Chemical Society, 1998. **120**(43): p. 11033-11043.
22. Lewis, P.N., N. Go, M. Go, D. Kotelchuck, and H.A. Scheraga, *Helix probability profiles of denatured proteins and their correlation with native structures.* Proceedings of the National Academy of Sciences of the United States of America, 1970. **65**(4): p. 810-815.
23. Dobson, C.M., and A.R. Fersht, *Protein Folding*. 1996: Cambridge University Press.
24. Smith, L.J., K.M. Fiebig, H. Schwalbe, and C.M. Dobson, *The concept of a random coil: residual structure in peptides and denatured proteins.* Folding Design, 1996. **1**(5): p. R95-R106.
25. Toal, S., and R. Schweitzer-Stenner, *Local order in the unfolded state: conformational biases and nearest neighbor interactions.* Biomolecules, 2014. **4**(3): p. 725-773.
26. Jha, A.K., A. Colubri, K.F. Freed, and T.R. Sosnick, *Statistical coil model of the unfolded state: resolving the reconciliation problem.* Proceedings of the National Academy of Sciiences of the United States of America, 2005. **102**(37): p. 13099-13104.
27. Ullman, R., *On the controversy over "fast" and "slow" helix–coil transition rates in polypeptides.* Biopolymers, 1970. **9**(4): p. 471-487.
28. Job, G.E., R.J. Kennedy, B. Heitmann, J.S. Miller, S.M. Walker, and D.S. Kemp, *Temperature- and length-dependent energetics of formation for polyalanine helices in water: assignment of wAla(n,T) and temperature-dependent CD ellipticity standards.* Journal of the American Chemical Society, 2006. **128**(25): p. 8227-8233.

29. Kennedy, R.J., S.M. Walker, and D.S. Kemp, *Energetic characterization of short helical polyalanine peptides in water: analysis of ^{13}CO chemical shift data.* Journal of the American Chemical Society, 2005. **127**(48): p. 16961-16968.
30. Saito, H., R. Tabeta, A. Shoji, T. Ozaki, and I. Ando, *Conformational characterization of polypeptides in the solid state as viewed from the conformation-dependent carbon-13 chemical shifts determined by the carbon-13 cross polarization/magic angle spinning method: oligo(L-alanine), poly(L-alanine), copolymers of L- and D-alanines, and copolymers of L-alanine with N-methyl- or N-benzyl-L-alanine.* Macromolecules, 1983. **16**(7): p. 1050-1057.
31. Saito, H., *Conformation-dependent ^{13}C chemical shifts: a new means of conformational characterization as obtained by high-resolution solid-state ^{13}C NMR.* Magnetic Resonance in Chemistry, 1986. **24**(10): p. 835-852.
32. Wishart, D.S., and B.D. Sykes, *The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data.* Journal of Biomolecular NMR, 1994. **4**(2): p. 171-180.
33. Wishart, D.S., C.G. Bigam, J. Yao, F. Abildgaard, H.J. Dyson, E. Oldfield, J.L. Markley, and B.D. Sykes, *^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR.* Journal of Biomolecular NMR, 1995. **6**(2): p. 135-140.
34. Pines, A., M.G. Gibby, and J.S. Waugh, *Proton-enhanced NMR of dilute spins in solids.* Journal of Chemical Physics, 1973. **59**(2): p. 569-590.
35. Schaefer, J., R.A. McKay, and E.O. Stejskal, *Double-cross-polarization NMR of solids.* Journal of Magnetic Resonance (1969), 1979. **34**(2): p. 443-447.
36. Baldus, M., A.T. Petkova, J. Herzfeld, and R.G. Griffin, *Cross polarization in the tilted frame: assignment and spectral simplification in heteronuclear spin systems.* Molecular Physics, 1998. **95**(6): p. 1197-1207.
37. Bennett, A.E., C.M. Rienstra, M. Auger, K.V. Lakshmi, and R.G. Griffin, *Heteronuclear decoupling in rotating solids.* Journal of Chemical Physics, 1995. **103**(16): p. 6951-6958.
38. Struppe, J.O., C. Yang, Y. Wang, R.V. Hernandez, L.M. Shamansky, and L.J. Mueller, *Long-observation-window band-selective homonuclear decoupling: increased sensitivity and resolution in solid-state NMR spectroscopy of proteins.* Journal of Magnetic Resonance (San Diego, Calif. : 1997), 2013. **236**: p. 89-94.
39. Baldus, M., R. Iuliucci, and B. Meier, *Probing through-bond connectivities and through-space distances in solids by magic-angle-spinning nuclear magnetic resonance.* Journal of the American Chemical Society, 1997. **119**(5): p. 1121-1124.

40. Chan, J.C., and G. Brunklaus, *R sequences for the scalar-coupling mediated homonuclear correlation spectroscopy under fast magic-angle spinning.* Chemical Physics Letters, 2001. **349**(1): p. 104-112.
41. Suzuki, M., *Generalized exact formula for the correlations of the Ising model and other classical systems.* Physics Letters, 1965. **19**(4): p. 267-268.
42. Merutka, G., H. Jane Dyson, and P.E. Wright, *'Random coil' ^{1}H chemical shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series GGXGG.* Journal of Biomolecular NMR, 1995. **5**(1): p. 14-24.
43. Platzer, K.E.B., V.S. Ananthanarayanan, R.H. Andreatta, and H.A. Scheraga, *Helix-coil stability constants for the naturally occurring amino acids in water. IV. Alanine parameters from random poly(hydroxypropylglutamine-co-L-alanine).* Macromolecules, 1972. **5**(2): p. 177-187.
44. Matheson, R.R., and H.A. Scheraga, *Calculation of the Zimm-Bragg cooperativity parameter σ from a simple model of the nucleation process.* Macromolecules, 1983. **16**(7): p. 1037-1043.
45. Kakivaya, S.R., and C.A. Hoeve, *The glass point of elastin.* Proceedings of the National Academy of Sciences of the United States of America, 1975. **72**(9): p. 3505-3507.
46. Vitkup, D., D. Ringe, G.A. Petsko, and M. Karplus, *Solvent mobility and the protein 'glass' transition.* Nature Structural and Molecular Biology, 2000. **7**(1): p. 34-38.
47. Ringe, D., and G.A. Petsko, *The 'glass transition' in protein dynamics: what it is, why it occurs, and how to exploit it.* Biophysical Chemistry, 2003. **105**(2–3): p. 667-680.
48. Pometun, M.S., E.Y. Chekmenev, and R.J. Wittebort, *Quantitative observation of backbone disorder in native elastin.* Journal of Biological Chemistry, 2004. **279**(9): p. 7982-7987.
49. Samouillan, V., D. Tintar, and C. Lacabanne, *Hydrated elastin: Dynamics of water and protein followed by dielectric spectroscopies.* Chem Phys, 2011. **385**(1–3): p. 19-26.
50. Wang, Y., *Secondary structural effects on protein NMR chemical shifts.* Journal of Biomolecular NMR, 2004. **30**(3): p. 233-244.
51. Ashida, J., K. Ohgo, K. Komatsu, A. Kubota, and T. Asakura, *Determination of the torsion angles of alanine and glycine residues of model compounds of spider silk $(AGG)_{10}$ using solid-state NMR methods.* Journal of Biomolecular NMR, 2003. **25**(2): p. 91-103.
52. Eyre, D.R., M.A. Paz, and P.M. Gallop, *Cross-linking in collagen and elastin.* Annual Review of Biochemistry, 1984. **53**(1): p. 717-748.
53. Crowley, P.B., and A. Golovin, *Cation-pi interactions in protein-protein interfaces.* Proteins, 2005. **59**(2): p. 231-9.
54. Djajamuliadi, J., T.F. Kagawa, K. Ohgo, and K.K. Kumashiro, *Insights into a putative hinge region in elastin using molecular dynamics simulations.* Matrix Biology, 2009. **28**(2): p. 92-100.

55. Kumashiro, K.K., J.P. Ho, W.P. Niemczura, and F.W. Keeley, *Cooperativity between the hydrophobic and cross-linking domains of elastin.* Journal of Biological Chemistry, 2006. **281**(33): p. 23757-23765.
56. Brown-Augsburger, P., C. Tisdale, T. Broekelmann, C. Sloan, and R.P. Mecham, *Identification of an elastin cross-linking domain that joins three peptide chains: possible role in nucleated assembly.* Journal of Biological Chemistry, 1995. **270**(30): p. 17778-17783.
57. Tamburro, A.M., B. Bochicchio, and A. Pepe, *Dissection of human tropoelastin: exon-by-exon chemical synthesis and related conformational studies.* Biochemistry, 2003. **42**(45): p. 13347-13362.
58. Tamburro, A.M., B. Bochicchio, and A. Pepe, *The dissection of human tropoelastin: from the molecular structure to the self-assembly to the elasticity mechanism.* Pathologie Biologie, 2005. **53**(7): p. 383-389.
59. Bochicchio, B., and A. Pepe, *Role of polyproline II conformation in human tropoelastin structure.* Chirality, 2011. **23**(9): p. 694-702.
60. Schirò, G., Y. Fichou, F.-X. Gallat, K. Wood, F. Gabel, M. Moulin, M. Härtlein, M. Heyden, J.-P. Colletier, A. Orecchini, A. Paciaroni, J. Wuttke, D.J. Tobias, and M. Weik, *Translational diffusion of hydration water correlates with functional motions in folded and intrinsically disordered proteins.* Nature Communications, 2015. **6**.
61. Tarek, M., and D.J. Tobias, *Role of protein-water hydrogen bond dynamics in the protein dynamical transition.* Physical Review Letters, 2002. **88**(13): p. 138101.

CHAPTER 6. VARIABLE-TEMPERATURE STUDIES OF ELASTIN'S CROSSLINKING DOMAINS BY MOLECULAR DYNAMICS SIMULATIONS

A. INTRODUCTION

The distribution of alanine conformations in the crosslinking domains is governed by a thermodynamic equilibrium that is dependent on temperature (Chapter 5). At the physiological temperature, 37 °C, alanines exist in both α-helix and random coil. However, upon cooling, an increase of α-helical content is identified in the Ala-rich regions. Specifically, the conversion of random coil alanines to the helical conformation is observed, as the temperature is lowered from 37 °C to -20 °C. Our observations of the temperature-dependent coil-to-helix transition provided the motivation for the computational studies of Ala-rich sequences in elastin's crosslinking regions.

Computations that exploit the effects of temperature on the secondary structures and backbone mobility of elastin's crosslinking domains were performed. A prediction algorithm based on helix-coil transition theory was used to evaluate the α-helical content in these regions at a series of variable temperatures. Temperature-replica exchange molecular dynamics simulations were also used to study EX25-ALDL and EX19-DLNL, which are representative crosslinking domains in native elastin.

A.1 Estimation of α-helical content by AGADIR

Bioinformatics tools are useful for coarse assessments of protein structures. Commonly used prediction algorithms such as GOR IV [1] and PSIPRED [2] predict the secondary structures of a peptide from its amino acid sequence. Chou and Fasman's early predictive algorithm calculates the probability of each residue in a peptide sequence to be found as α-helix, β-sheet, or coil (neither α-helix nor β-sheet), using statistical weights that are derived from X-ray solved structures in a database [3]. Although these algorithms are effective for evaluating globular proteins, they overestimate the relative amounts of the above-described structures in disordered proteins like elastin and, therefore, have limited accuracy.

This chapter gives results obtained with a predictive algorithm that accounts for the helix-coil equilibrium in elastin's Ala-rich crosslinking domains. Specifically, AGADIR is a tool that estimates the α-helical content of monomeric peptides, based on helix-coil transition theory [4-13]. This algorithm accurately predicts the helical propensities for hundreds of monomeric peptides at various temperatures, ionic strengths, and pH's, as confirmed by NMR and CD spectroscopy [5, 6, 10]. Unlike the Chou-Fasman

algorithm, AGADIR considers the energetics involved in the formation of α-helix, and the molecular partition function that encompasses statistical weights for both α-helix and random coil is then computed at residue-level. The average helical propensity for the peptide and/or for each residue in the chain is calculated from the input of primary structure (amino acid sequence) and the selection of several modifiable parameters such as temperature, pH, etc. AGADIR also generates the α-helical content as a function of temperature, which is useful for the evaluation of the helical stability.

A.2 Variable-temperature studies by replica-exchange MD (REMD) simulations

Previous chapters showed that a temperature-dependent helix-to-coil transition occurs in the Ala-rich regions of elastin. Long chains of α-helices are observed primarily at low temperatures like -10 °C or -20 °C. Upon warming, the α-helical content is reduced, and the amount of random coil is increased. To examine the molecular aspects of alanines during the helix-coil interconversions, variable-temperature computational studies are explored.

A simple approach that runs conventional MD (cMD) simulations at multiple temperatures is impractical and computationally costly. As discussed in Chapter 4, a 1 μsec simulation at 300 K was inadequate for the full recovery of EX25-ALDL's helical conformation, and only the coil state was sampled. Furthermore, a simulation at low temperature samples a narrow conformational space, as the system is often trapped in a local free energy minimum. Numerous computing processors are required to perform variable-temperature (VT) simulations in series, with each run sampling a different temperature configuration from another. Therefore, a method that requires relatively short computational time to perform parallel MD simulations over a broad distribution of temperatures is desired.

Replica-exchange MD (REMD) simulations are chosen for the VT investigation of elastin's crosslinking domains. The REMD algorithm was initially developed to obtain a global minimum-energy state from a single peptide simulation, by increasing the rate of conformational sampling in the computation [14, 15]. Parallel, simultaneous calculations of the system's replicas (i.e., molecular coordinate copies of the original structure) are run over a broad range of temperatures. The merit of REMD lies in the exchange of temperatures among replicas, thereby overcoming local energy barriers in the reaction coordinates of protein folding. The effectiveness of this algorithm was first demonstrated in the folding simulation of the pentapeptide met-enkephalin [14]. The distribution of potential energies from the REMD simulations of met-enkephalin agrees with that obtained from a series of independent cMD simulations performed at many temperatures. In addition, the conformational space that was explored by REMD was broader than that by cMD, indicating a more efficient computational route in the former than in

the latter. Since the initial development of REMD, many successful applications to biological systems have been reported [16-19].

The REMD algorithm utilizes MD and Monte Carlo (MC) approaches to efficiently sample the configurational space of a system. In a typical procedure, the coordinates for several independent replicas are calculated simultaneously over a broad distribution of temperatures (*left of* **Figure 6.1**). Each replica undergoes iterated cycles of short MD and MC processes. The MD determines the molecular coordinates of each replica at a particular temperature; the MC evaluates $P(T_1 \leftrightarrow T_2)$, the probability of the *temperature exchange* between a replica at T_1 and its "neighbor" at T_2, i.e., the replica at the next-higher or next-lower temperature. The temperature exchange is accepted or rejected based on the Metropolis acceptance criterion (eqn. 6.1),

$$\boldsymbol{P(T_1 \leftrightarrow T_2) = Min(1, e^{\Delta})} \qquad \text{(eqn. 6. 1)}$$

in which Δ is given by

$$\Delta = (\boldsymbol{B_1 - B_2})\,(\boldsymbol{U_1 - U_2}) \qquad \text{(eqn. 6. 2)}$$

where $(B_1\text{-}B_2)$ is the difference in reciprocal temperatures and $(U_1\text{-}U_2)$ is the difference in potential energies between a replica (1) and its neighbor (2). This process is iterated until the REMD simulation reaches convergence.

The results of REMD simulations are typically analyzed with methods that are used for cMD computations. Thermodynamic properties are extracted from each REMD trajectory and presented as a function of temperature.

Computational studies via AGADIR prediction and REMD simulations are used to probe the conformational equilibrium of alanines and other residues in EX25-ALDL and EX19-DLNL peptides. The AGADIR algorithm evaluates the α-helical content in these systems, and the REMD simulations provide biophysical insights into the helix-coil transitions.

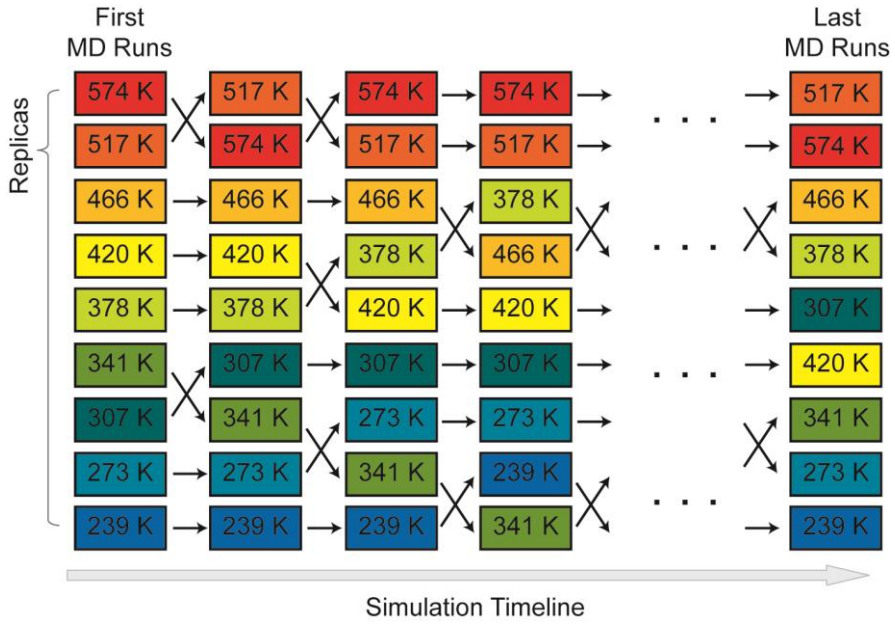


Figure 6.1 Procedure for replica-exchange MD (REMD) simulations using temperature exchanges. Colored boxes represent short MD runs of each replica. Crossed arrows indicate a temperature exchange of neighboring replicas, i.e., systems with closely related temperatures, as described by equation 6.1. Horizontal arrows reflect no exchanges. Three sequential black dots represent repeated cycles of MD runs and exchanges.

B. SIMULATION PROCEDURE AND ANALYSIS METHODS

B.1 AGADIR predictions of tropoelastin's peptide sequences

AGADIR [4-13] calculation of the helical propensities of each hydrated monomeric peptide requires several parameters. The primary structures of the crosslinking domains encoded by EX25 and EX19 were input into the search algorithm. Each peptide was acetylated and amidated on the N- and the C-termini, respectively. Conditions were set at pH 7 and an ionic strength of 0.15 M, with temperatures ranging from 274 to 374 K. The output of the online calculation was saved as a text file. Excel spreadsheets were then used to plot the melting curve, i.e., the predicted α-helical content as a function of temperature.

B.2 Starting geometries for REMD simulations

The procedure for creating the starting structures of EX25-ALDL and EX19-DLNL peptides for REMD simulations is identical to the one used for cMD simulations described in Chapter 4. Briefly, α-helical structures were created on VMD using the molefacture module. The N- and C-termini were modified by acetylation and amidation, respectively. The ALDL and DLNL crosslinks were modifications of Lys sidechain in EX25 and EX19 peptides, respectively, which were accomplished using the PSFgen module of VMD via patch files (**Appendix 10**).

The Generalized Born-Surface Area (GBSA) implicit solvent system was used on NAMD [20] to simulate the effects of water around the peptides. The Generalized Born implicit solvent, recently implemented on the NAMD package, is effective for parallel computing of large molecules, as it accelerates the electrostatic computations and reduces the total computational cost [21].

The electrostatic cutoff distances in a GBSA system were typically longer than those in explicit solvent systems. The cutoff distance for the short-range non-bonded interactions was set to 16.0 Å. The switching distance for the vdW smoothing function was set to 15 Å, and the pair list distance for atom inclusions was 18.0 Å. To represent the electrostatic screening of water, the ion concentration was specified as 0.15 M, and the dielectric constant was set to 80. The cutoff distance for the calculation of the Born radius (alphaCutoff) was 14 Å. The hydrophobic energy contribution from implicit solvent was calculated using the solvent-accessible surface area (SASA) energy of 0.006 kcal/mol/Å^2. Particle mesh Ewald summation and periodic boundary conditions were not used in the GBSA implicit solvent. The

integration time steps for the bonded, short-range and long-range non-bonded electrostatic interactions were 1, 2, and 4 fs, respectively.

B.3 Protocols for running temperature-varied REMD simulations

Replica-exchange MD simulations were performed using the NAMD simulation package [20]. A tutorial on running a replica-exchange simulation on NAMD is available on the developer's webpage <http://www.ks.uiuc.edu/Research/namd/2.9/ug/node66.html>. REMD simulations were performed on the STAMPEDE cluster, at the Texas Advanced Computing Center (TACC) at the University of Texas at Austin. Sample configuration files and batch scripts are included in **Appendix 18**.

The REMD simulations of EX25-ALDL and EX19-DLNL peptides were performed using 16 replicas with temperatures ranging from 263 to 574 K. The probability acceptance ratio was 40-45%. The initial velocities for the molecular dynamics vary with the temperature of the system. The probability exchange was calculated every 1.0 ps of MD using 1 fs integration time step. A total of 20 ns of REMD simulations were performed per replica.

B.4 Analysis of REMD trajectories for structural and dynamical characteristics

REMD trajectories were analyzed using VMD's built-in modules and modified tcl-scripts, previously described in Chapter 4 (**Appendix 19**). Data processing, sorting, and plotting were performed on OriginPro and Microsoft Excel.

As a qualitative measure of peptide dynamics, the Cα-RMSD values were computed as a function of simulation time, with VMD's RMSD Trajectory Tool module. The output of each computation was stored as a text file, and the data was plotted with line graphs using OriginPro. The histogram for each Cα-RMSD plot was created, and a Kernel-smooth function was implemented in the visualization. These plots were used to define either a helix or a coil state for each trajectory structure. Peptide coordinates with high RMSD values were considered as coil, whereas those with low RMSD values were categorized as helix. The boundary line (i.e., RMSD values) that separated the helix from the coil state was determined from the shapes of the Kernel-smooth functions. For EX25-ALDL, structures with RMSD values of <1.5 Å were categorized as the helix state, otherwise coil. For EX19-DLNL, the RMSD value of <2.0 Å was used as the boundary line between the helix and the coil. The Cα-RMSF plots were also calculated for REMD trajectories to illustrate the time-averaged backbone dynamics per-residue (*RMSF.tcl*).

In VMD, secondary structures are defined using the STRIDE algorithm [22, 23]. The population of secondary structures (such as α-helix, 3_{10}-helix, turns and coil) in each peptide trajectory was computed

using *STRUCTcount.tcl*. The helical content was calculated using the ratio between the number of α-helical residues and the total number of amino acids in the peptide. The average ratio was then computed over the total number of trajectory frames, and the value was expressed in percentage.

The free energy landscapes were generated using the g_sham module on GROMACS package [24-26], as described in Chapter 4. Several tcl-scripts and csh-shell scripts (*StripTermCatPdbGmx.csh*, *createFEL1.csh*, *createFEL2.csh*) were used to convert the trajectory file produced by NAMD into a compatible format used by GROMACS. Contour plots were generated using OriginPro to help visualize the free-energy surfaces.

The *phipsi.tcl* was used to collect all (ϕ, ψ) angles in EX25-ALDL and EX19-DLNL, and the *phipsitotext.csh* was used to sort these torsion angles by residue.

C. RESULTS AND DISCUSSIONS

Computational methods were used to study the structure and dynamics of the Ala-rich EX25-ALDL and EX19-DLNL crosslinking domains in native elastin. The α-helical content of the representative peptides was predicted using the AGADIR algorithm and REMD simulations (section C.1). Additionally, hydrogen-bond interactions in the backbone atoms of EX25-ALDL and EX19-DLNL were examined, and their role in stabilizing the α-helical structure was investigated. REMD trajectory structures were analyzed in the context of the helix-coil transitions (section C.2). Temperature effects on the equilibrium between helix and coil states were explored through free energy landscape analyses (section C.3). Lastly, structural motifs and relative mobility of the alanines in EX25-ALDL and EX19-DLNL studies were examined (section C.4).

C.1 AGADIR predictions and REMD simulations probe temperature effects on the α-helical structures in elastin's crosslinking domains

AGADIR is a prediction algorithm that calculates the average α-helical content of monomeric peptides in solution using the helix-coil transition theory. Unlike GOR IV and PSIPRED which predict common secondary structures like α-helix and β-sheet through sequence alignments [1, 2], AGADIR quantifies the α-helical content of a given sequence by considering the energetics involved in its formation [27]. AGADIR also accounts for the conditions of the peptides in solution, such as pH, ionic strength and temperature [4-13]. The accuracy of AGADIR's prediction in calculating the α-helical content of monomeric peptides has been validated by CD measurements [5, 6, 10].

The AGADIR algorithm, based on helix-coil transition theory, predicts that the α-helical content of tropoelastin's KA domains, encoded by exons 15, 17, 19, 21, 23, 25, 27, 29, and 31, depends on temperature. **Figure 6.2** illustrates the α-helical content of tropoelastin's Ala-rich domains at temperatures between 274 and 373 K, as predicted by AGADIR. The AGADIR results are consistent with the trend observed by CD and NMR measurements [28]. Namely, the amount of α-helix increases as temperature decreases. EX21 and EX23 domains have the highest relative helical content at all temperatures, with roughly 55% at 274 K and 2% at 374 K. The α-helical structures in these two domains were previously investigated using MD simulations [29], and the Lys in EX21 and EX23 have also been identified as 'hot residues', which are the active sites for crosslinking reactions in the recombinant tropoelastin [30-32]. With the exception of EX21 and EX23, all KA and KP crosslinking domains have a unique profile. For instance, at 300 K, the α-helical content of EX19 and EX25 domains was estimated at 15% and 5%, respectively. These values are different from the others, and they also reflect unequal rates of change,

with respect to temperature. Furthermore, the difference in α-helical content in these two domains is greater at 274 K (22%) than at 373 K (1%).

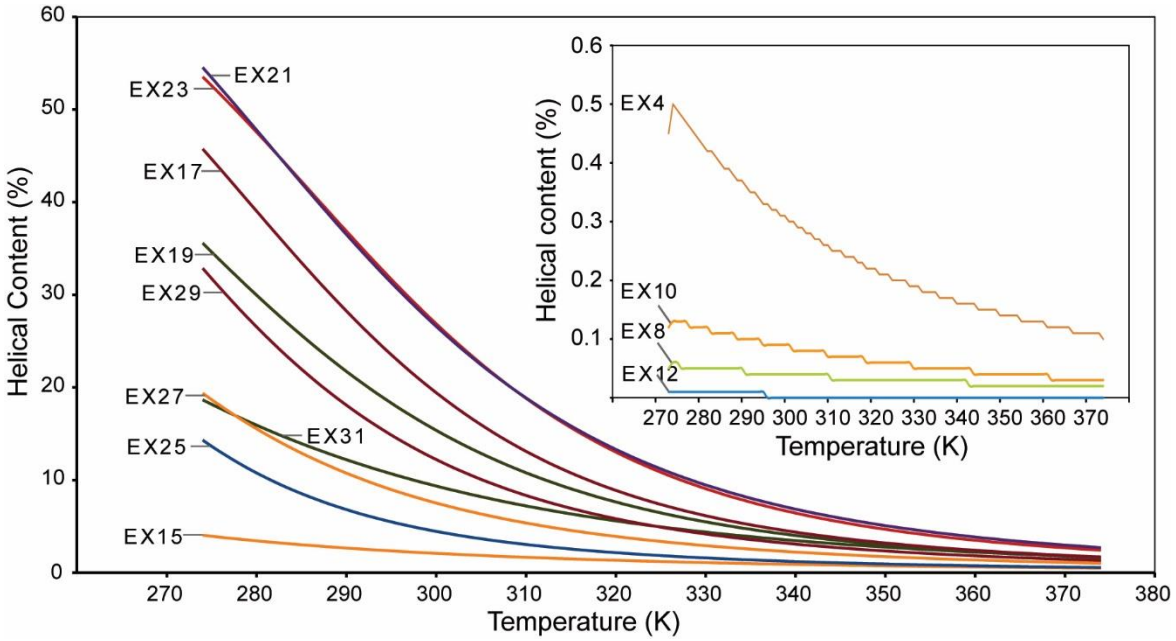


Figure 6.2 AGADIR prediction of the helical content in tropoelastin's KA (main) and KP (inset) crosslinking domains as a function of temperature. Smooth (╰) and jagged (↿) curves reflect helical propensities for the KA- and KP-type crosslinking domains.

Some crosslinking domains have notably low helical propensities. The regions containing KP motifs (i.e., EX4, EX8, EX10, and EX12) have < 0.5% of α-helical content (inset in **Figure 6.2**), which may be associated with structural disorder in the corresponding regions. Namely, the KP domains are located in the N-terminal 'coil' region in the head-to-tail fiber axis model of tropoelastin [33], which is believed to be responsible for flexibility. Among the KA motifs, EX15 has the lowest helical propensities. This domain is estimated to have 19% and 1% of α-helical content at 274 K and 374 K, respectively. The low helical propensity in EX15 is consistent to the lack of crosslinking activity in this region [30].

Generally, AGADIR predicts lower α-helical content than conventional algorithms like GOR IV [1] and PSIPRED [2]. For example, AGADIR estimates EX19 and EX25 to have 15% and 5% α-helical content at 300 K. In contrast, GOR IV predicts the two domains to have 67% and 56% α-helix, which are higher than observed (**Appendix 19**).

Temperature-dependent helical propensities of elastin's Ala-rich crosslinking domains were also investigated using REMD simulations. The simulations were performed on EX25-ALDL, EX19-DLNL, EX25, and EX19, which represent two of these domains with and without crosslink moieties. To complement the result of AGADIR predictions, REMD simulations of EX25 and EX19 domains without

crosslinks were run over a broad distribution of temperatures ranging from 263 K to 574 K, and the α-helical content was measured from the trajectory structures.

REMD simulations show temperature-dependence of α-helical content for EX25 and EX19 (**Figure 6.3**). The temperature-dependence of this plot is similar to the one predicted by AGADIR (section C.1), and the trend is also consistent with CD and NMR measurements [28]. The loss of α-helical structures, i.e., helix melting, is observed in the two peptides with the increase of temperature. The rate of helix melting in EX25 is different from that in EX19. At 263 K, EX25 and EX19 has 71% and 66% helical content, respectively, which progressively decreases at higher temperatures. The two melting curves intersect at 333 K, which is also the helix melting temperature (T_m) [34]. T_m is the temperature at which two peptides have 50% helical content. Below T_m, the helical content of EX25 drops at a faster rate than that of EX19. At 573 K, negligible α-helical content (< 5%) is observed in the two peptides. However, there are some differences between REMD and AGADIR. For instance, the REMD simulations show higher α-helical content than the AGADIR predictions over all temperatures. The high helical content in the trajectory structures may be due to a couple of factors. The CHARMM forcefield, used in the REMD simulations, has been shown to overestimate the α-helical content in several peptides [35, 36]. The GBSA implicit solvent was reported to have a large nucleation parameter for helix formation, such that the overall α-helical content in a peptide is higher and the T_m is overestimated [34, 37].

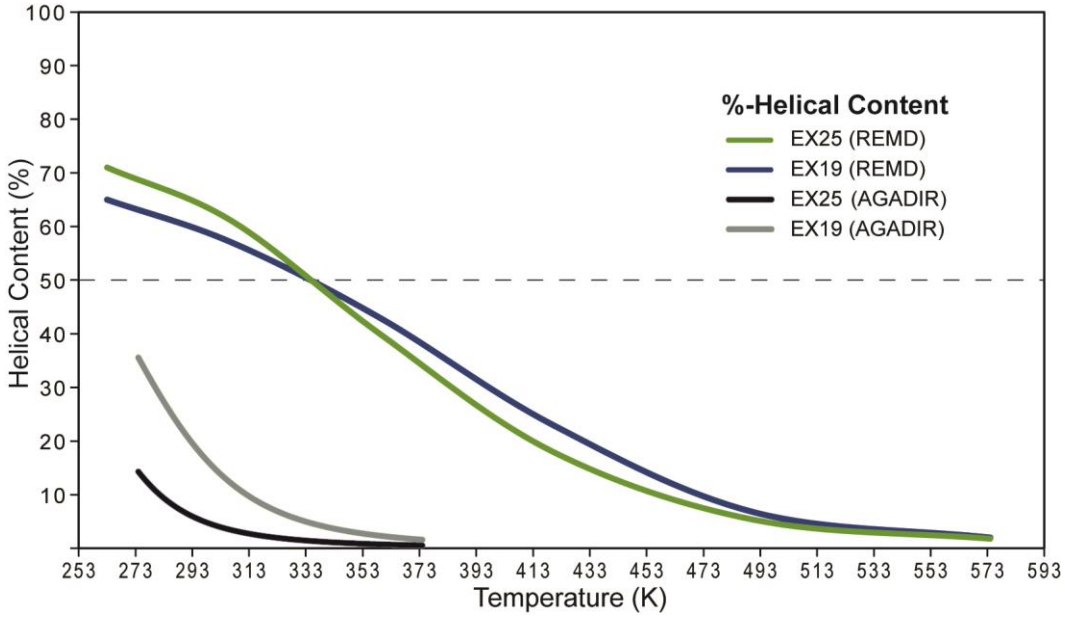


Figure 6.3 Helical content in EX25 and EX19 domains as a function of temperature. **Green** and **blue** curves represent the helical content in REMD trajectory structures of EX25 and EX19 peptides, whereas **black** and **gray** curves represent AGADIR predictions of the α-helical propensities in EX25 and EX19 domains without crosslinks. The dashed horizontal line (- - -) marks the level at which peptides have 50% helical content, termed as the helix melting temperature (T_m).

REMD simulations of EX25-ALDL and EX19-DLNL were also run over a broad distribution of temperatures ranging from 263 K to 574 K. The α-helical content and the corresponding number of hydrogen-bonds (H-bonds) were measured from the trajectory structures. The stability of α-helices in EX25-ALDL and EX19-DLNL peptides is dependent on the temperature (**Figure 6.4**). The highest α-helical content in each peptide, ~50-55%, is observed at 263 K. At this temperature, the average number of H-bonds, 5 and 7, is also the greatest for EX25-ALDL and EX19-DLNL, respectively. However, the helical content and the number of H-bonds decrease progressively with the increase of temperature. For example, at 373 K, the α-helical content is ~25-30%, and there are 3 and 4 H-bonds in EX25-ALDL and EX19-DLNL, respectively. At 573 K, negligible helical content (< 5%) is observed with only 1 H-bond formation. This helix melting profile shows the correlation between the number of backbone H-bonds and the α-helical content, the characteristic interaction that was used in the early development of helix-coil theory [27].

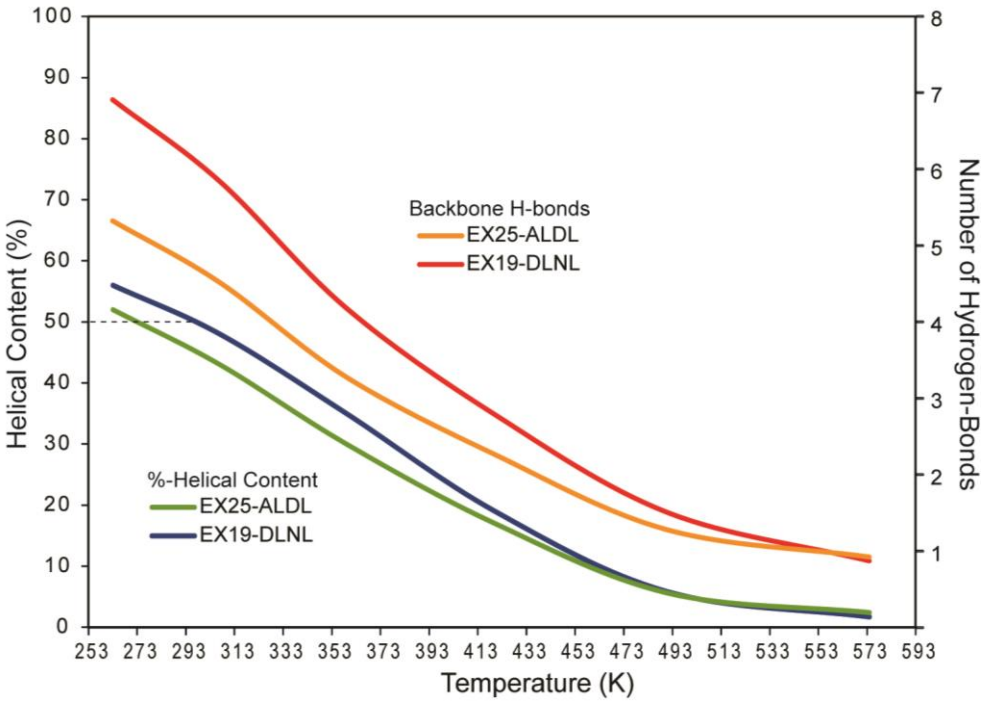


Figure 6.4 Helical content and the number of hydrogen-bonds in EX25-ALDL and EX19-DLNL peptides as a function of temperature. Orange and red curves reflect the total number of hydrogen-bonds in the backbone of EX25-ALDL and EX19-DLNL, respectively. Blue and green curves represent the percentage of the peptide that adopts α-helix, an average calculated for all structures in the REMD simulations of EX25-ALDL and EX19-DLNL, respectively. The dashed horizontal line (----) marks the level at which peptides have 50% helical content, termed as the helix melting temperature (T_m) [34].

C.2 Root-mean-square deviations define the helix and coil states of EX25-ALDL and EX19-DLNL peptides

Each peptide in the REMD trajectory might be considered to be helix- or coil-like on the basis of the root-mean-square deviation of the Cα atoms, Cα-RMSD. The deviations are computed with respect to the initial, minimized α-helical structures of EX25-ALDL and EX19-DLNL. Hence, the Cα-RMSD values reflect the distortion in or from the peptide's (initial) α-helical structure at a given trajectory time point. A low RMSD value indicates a structure that is similar to the initial α-helical conformation, whereas a high RMSD value reflects significant deviation, or may be simply termed coil. Furthermore, a high RMSD may also correspond to a partially unfolded structure with short segments of α-helices. In previously described cMD simulations (Chapter 4), the RMSD values were plotted as a function of simulation time to indicate time-dependent changes in atomic coordinates. However, this feature is not applicable for REMD simulations, because the peptide coordinates are gathered from non-continuous trajectories of multiple replicas. For this reason, the Cα-RMSD values for EX25-ALDL and EX19-DLNL are plotted as histograms (**Figure 6.5**). A point on the histogram gives the relative number of structures observed over the entire trajectory with a particular RMSD. Kernel-smooth functions [38] were used to assist in the visualization.

The determination of EX25-ALDL and EX19-DLNL as helical and/or coil-like is based, in part, on this quantity (RMSD), which describes the deviation of peptides from their common starting structures, i.e., the α-helix. As shown in **Figure 6.5**, there are relatively high occurrences of structures with Cα-RMSD <1.5 Å and <2.5 Å for EX25-ALDL and EX19-DLNL, respectively, with the remainder of the structures with higher Cα-RMSD values with populations that increase at higher temperatures. At 263 K, the traces show two local maxima of coil-like structures with RMSD values of ~3 Å and ~4.5 Å for EX25-ALDL and ~4 Å and ~5 Å for EX19-DLNL. However, the distribution of structures narrows with the increase of temperature, as only one maximum is observed at ~4.8 Å in both EX25-ALDL and EX19-DLNL at 574 K.

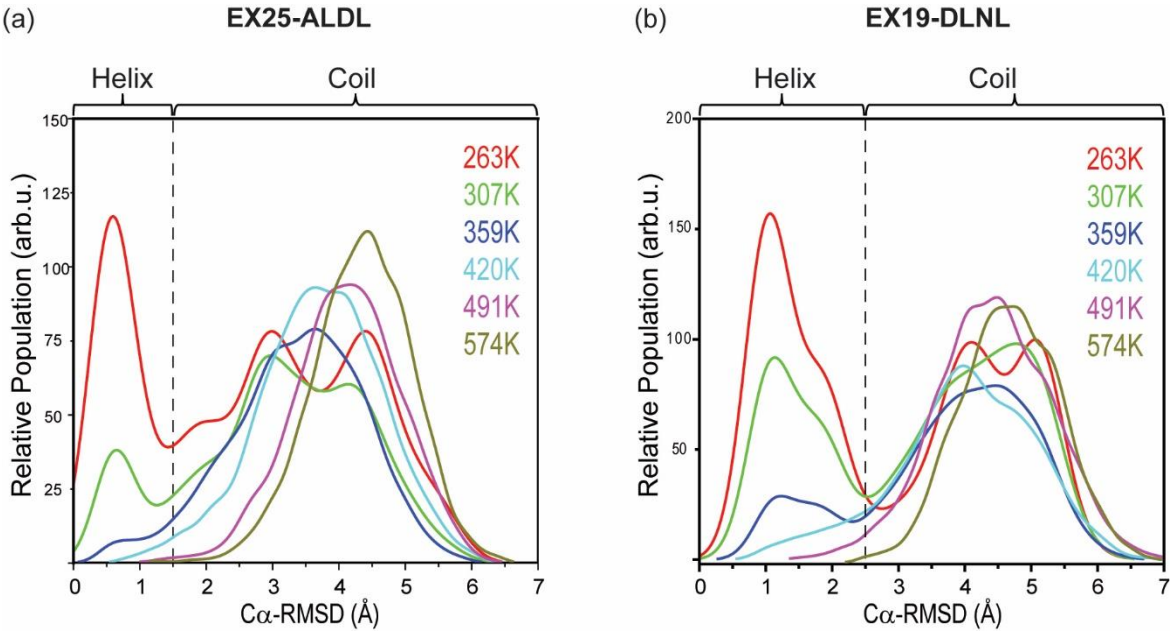


Figure 6.5 Superposition of Cα-RMSD values in the REMD simulations of (a) EX25-ALDL and (b) EX19-DLNL. Traces represent the histograms of RMSD values at 263 K (red), 307 K (green), 359 K (blue), 420 K (cyan), 491 K (maroon) and 574 K (brown), fitted by Kernel-smooth function. The arbitrary unit (arb. u.) in the vertical axis corresponds to the number of structures identified with certain RMSD values, relative to the total number of trajectory structures. Vertical dashed lines (¦) indicate the boundary between helix- and coil-like structures.

The RMSD values illustrate the temperature dependence of the structural distributions. The helix-like structures of EX25-ALDL and EX19-DLNL are extracted primarily from the low-temperature trajectories including 263 K, 307 K, and 359 K, although minor populations also exist in the simulations at 420 K and 491 K. Furthermore, the intensity of RMSD traces for helix-like structures decreases with heating, which reflects the helix melting process. In contrast, the coil-like structures are prominent across all temperatures, and the RMSD features vary with temperature. For instance, two local maxima in the RMSD plot at 263 K are observed in the coil state, indicating two related groups of partially unfolded peptides or coil-like structures. However, the two maxima converge as the temperature is increased, which implies that the helical segments in these structures are lost at higher temperatures. These features support the AGADIR predictions, i.e., the stability of α-helices decreases with the increase of temperature.

C.3 Free energy landscape analyses for characterization of the molecular structures of EX25-ALDL and EX19-DLNL

C.3.1 Energetics provide a framework for defining multiple low-energy states of the crosslinked domains. The free energy landscape (FEL) is a two-dimensional contour map of relative free energies of all trajectory structures in a simulation as a function of specified reaction coordinates. These coordinates typically describe the molecular properties of the system such as Cα-RMSD, radius of gyration, atomic contacts, etc. An energy minimum, or free energy basin, corresponds to an accessible thermodynamic state in the system, each of which represents a cluster of conformations with closely related properties. The depth of a given basin, ΔG^* in kJ/mol, reflects the relative size of a given population.

The FEL for EX25-ALDL and EX19-DLNL trajectories were projected on two coordinates, %-helical content (vertical axis) and Cα-RMSD values (horizontal axis) (**Figure 6.6** and **6.7**). As previously described, the helix- and coil-like structures were defined using low and high Cα-RMSD values, respectively. However, a structure with a high Cα-RMSD that is nominally "coil-like" may also contain a sizable number of residues adopting the α-helical conformation.

The FEL for EX25-ALDL trajectories shows well-defined energy minima (**Figure 6.6a**). A free energy basin that represents a population of structures with the highest α-helical content has Cα-RMSD values from 0 Å to 1.5 Å and helical content ranging from 60% to 90%. The bottom of this basin, ΔG^* ~11.4 kJ/mol, is centered at Cα-RMSD~0.5 Å and 75% helical content, or (RMSD, %-helical content) = (0.5 Å, 75%), and the ΔG^* value increases with each concentric contour line. Basins with low RMSD's represent mostly α-helical structures with features that are similar to the initial configuration of EX25-ALDL. These low-RMSD populations include those with structures containing slightly lower α-helical content. For example, the representative structure (**1**) (**Figure 6.6b**) corresponding to (0.8 Å, 67%) in the FEL illustrates an EX25-ALDL peptide with its C-terminal residues unwound from the fully helical conformation. In contrast, a free energy basin that covers narrow ranges of RMSD, 3.5 Å – 4 Å, and helical content, 60%-70%, is identified with a ΔG^* of 14.7 kJ/mol, representing a minor population of structures containing some α-helical segments. The representative EX25-ALDL for this basin (**2**), centered at (3.8 Å, 67%), shows two contiguous α-helical segments that are separated by a turn. This peptide has ten amino acids that are considered to be in the α-helical conformation, equating to 66% α-helical content. Although this peptide (**2**), incidentally, has the same helical content as the previously described structure (**1**), its RMSD value (3.8 Å) differs from the previous one (0.8), which indicates a significant deviation from the starting peptide, which is a single continuous α-helical chain.

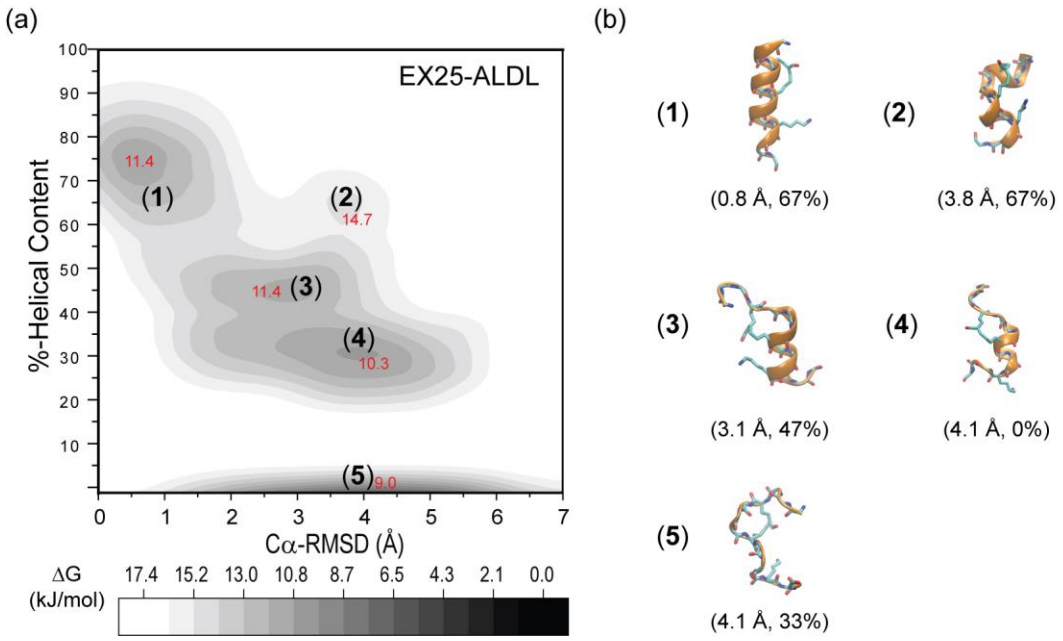


Figure 6.6 (a) Free energy landscape and (b) representative structures for EX25-ALDL trajectories. The contour map of relative free energy is plotted as a function of Cα-RMSD (horizontal) and %-helical content (vertical). The plot is generated from the PMF calculation of all REMD trajectory structures that are recorded for all temperatures, 263-574 K. The ΔG* at the bottom of each basin is shown in red. The grayscale ( ) reflects the differences in relative free energy in kJ/mol. In Figure 6.6(b), representatives (**1**), (**2**), (**3**), (**4**), and (**5**) are obtained from the indicated (RMSD, %-helix) coordinate in the FEL.

A free energy basin with ΔG* ~11.4 kJ/mol is observed with a lower range of helical content, 40%-50%, but higher Cα-RMSD values, 1.5 Å - 4 Å. This population corresponds to structures that are partially unfolded from the starting fully helical conformation. As illustrated by the representative EX25-ALDL (**3**) at (3.1 Å, 47%), the peptide has seven residues forming the α-helical structure. This structure is also related to another representative (**3**) that only has five residues in a shorter helical segment, which corresponds to (4.1 Å, 33%) in the basin with ΔG* ~10.3 kJ/mol describing Cα-RMSD's ranging from 1.5 Å to 5.5 Å and helical content ranging from 20% to 40%.

Last, non-α-helical structures are represented by the free energy basin covering the widest range of Cα-RMSD values from 1.5 Å to 6.5 Å and 0% helical content. This free energy basin has the ΔG* of ~9 kJ/mol, reflecting the largest population of structures in all combined REMD trajectories. In addition, this basin is isolated from the others by a free energy barrier with ΔG* ~17.4 kJ/mol, centered roughly at (3.5 Å, 15%) in the FEL. This free energy barrier exists in the topology, because trajectory structures with <5 residues appearing in sequence with (ϕ, ψ) angles corresponding to the α-helix (-57°, -45°) cannot form the shortest helical segment, i.e., a single turn of an α-helix by *i*-to-*(i+4)th* hydrogen-bonding. Hence, the

representative structure (**5**), at (4.1 Å, 0%), shows no α-helical segments in the chain, even though a few residues were found with the (ϕ, ψ) angles of an α-helix.

The FEL profile for EX19-DLNL is comparable to that for EX25-ALDL (**Figure 6.7a**). Namely, multiple free energy basins are distributed over the low and high RMSD values, representing helix- and coil-like structures, respectively. Two connected basins with ΔG* ~19.4 kJ/mol are observed, covering narrow ranges of low RMSD values, 0.5 Å – 2.2 Å, and high helical content, 50%-80%. These basins correspond to a population of trajectory structures with the highest α-helical content, which are relatively similar to the starting EX19-DLNL. Representative structures for the corresponding basins at (1.1 Å, 72%) (**1**) (**Figure 6.7b**) and at (1.5 Å, 55%) (**2**) have α-helical segments that slightly differ in lengths. These two representatives are also related to the partially unfolded, coil-like, structures that populate the free energy basin with RMSD values ranging from 2.5 Å to 4.5 Å, helical content ranging from 45% to 65%, and ΔG* of ~19.4 kJ/mol. The representative EX19-DLNL (**3**) at (3.5 Å, 50%) shows an α-helical segment formed by nine central residues. A free energy basin that comprises wider ranges of Cα-RMSD values, 1.5-5.5 Å, and helical content, 20-40%, is identified with a ΔG* of 17.8 kJ/mol. This population corresponds to a set of structures with a short α-helical segment, as shown in the representative peptide at (4.7 Å, 28%) (**4**).

Lastly, the largest population of EX19-DLNL structures comprises those with no α-helical content. This population is represented by the free energy basin that covers a wide range of RMSD values from 1.5 Å to 7 Å, with the ΔG* ~16.3 kJ centered at (4.5 Å, 0%). Furthermore, this basin is isolated from other minima, similar the feature observed in the FEL for EX25-ALDL. As described earlier, non-helical structures populating this basin have less than five residues with (ϕ, ψ) angles corresponding to the α-helix, as shown in the representative peptide (**5**).

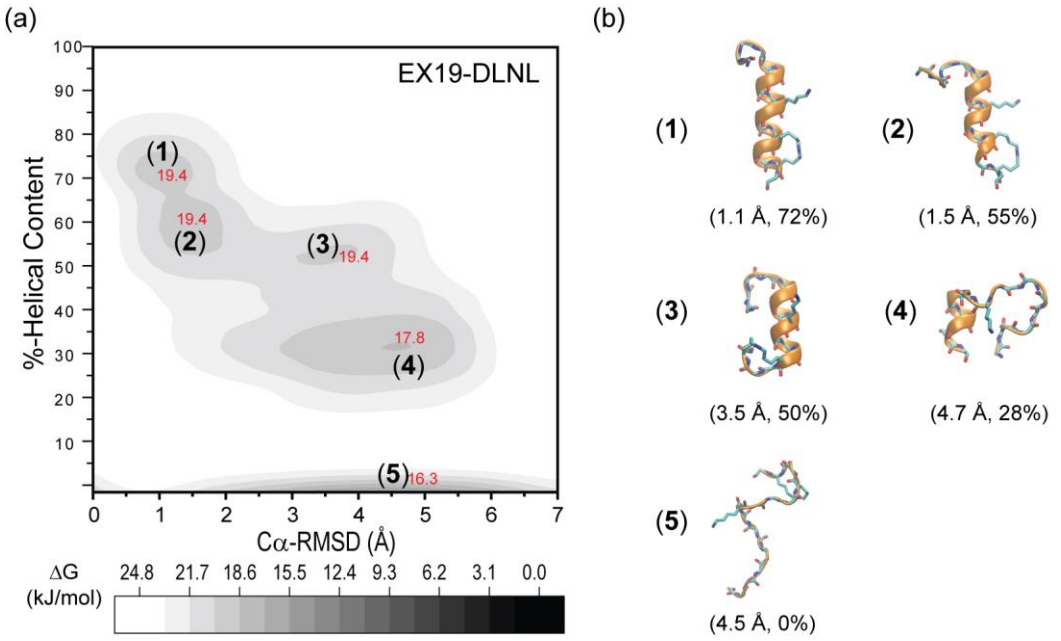


Figure 6.7 (a) Free energy landscape and (b) representative structures for EX19-DLNL trajectories. The contour map of relative free energy is plotted as a function of Cα-RMSD (horizontal) and %-helical content (vertical). The plot is generated from the PMF calculation of all REMD trajectory structures that are recorded for all temperatures, 263-574 K. The ΔG* at the bottom of each basin is shown in red. The grayscale ( ) reflects the differences in relative free energy in kJ/mol. In Figure 6.6(b), representatives (**1**), (**2**), (**3**), (**4**), and (**5**) are obtained from the indicated (RMSD, %-helix) coordinate in the FEL.

In summary, features in the FEL for EX25-ALDL and EX19-DLNL trajectories indicate the presence of at least two major populations, with low- and high-RMSD structures, in equilibrium. The low-RMSD population is characterized by helix-like structures with at least 50% α-helical content and with Cα-atom coordinates approximating those in the fully folded peptide. The high-RMSD population tends to have a lower helical content, and the high RSMD in each corresponding structure reflects a deviation from the α-helical conformation. The low- and high-RMSD structures are also separated by shallow regions, corresponding to the population of other intermediate structures. This result suggests that the helix- and coil-like structures exist in equilibrium, and the two populations are highly accessible in the folding of EX25-ALDL and EX19-DLNL.

C.3.2 Energetics of peptide folding reveal temperature-dependent shifts of equilibrium positions between the low- and high-RMSD structures. Features in the low- and high-RMSD structures of EX25-ALDL and EX19-DLNL were further investigated using additional FEL's. The contour maps were projected on two axes, Cα-RMSD and hydrogen-bonds (H-bonds), and the analysis was performed for each temperature investigated by REMD. As shown in previous sections, both low- and high-RMSD structures exist in the free energy landscapes. The number of H-bonds is therefore used to examine the backbone stabilization in the two populations.

Multiple low energy minima are observed in the free energy landscapes for EX25-ALDL trajectories, and the landscape topology at each temperature is unique. **Figure 6.8** shows FEL, torsion angles, and representative structures at 263, 307, 359, 420, and 491 K. The free energy surface at 263 K is characterized by multiple well-defined basins. First, a series of deep free energy minima are observed, covering large ranges of H-bonds, from 2 to 10, and Cα-RMSD values, from 3.5 to 6 Å. The two deepest minima in this series are identified at (4.8 Å, 2 H-bonds) with ΔG* ~6.3 kJ/mol and at (4.8 Å, 9 H-bonds) with ΔG* ~6.6 kJ/mol, respectively, representing the largest populations of structures containing the lowest and highest degree of backbone stabilization, respectively. Equal-depth minima are distributed evenly between the two corresponding basins, with ΔG* ranging from 7.0 to 7.2 kJ/mol. These basins are also connected to a series of shallower free energy basins, with ΔG*'s ~7.6-7.9 kJ/mol, centered at ~2.9 Å (RMSD) and distributed over a wide range of H-bonds, 2-8. Last, a succession of free energy basins is observed with a wide range of H-bonds that vary from 2 to 10, centered at the Cα-RMSD value of 0.7 Å. The deepest basin in this series is identified at (0.7 Å, 8 H-bonds), with ΔG* ~7.4 kJ/mol, representing a cluster of structures containing the longest, continuous, α-helical segments. This FEL shows the equilibrium of helix- and coil-like structures containing varying degree of backbone stabilization. The major population is comprised of the high-RMSD structures (~4.8 Å), which are tied to the minor populations of structures with lower RMSD's, 2.9 Å and 0.7 Å.

The majority of EX25-ALDL structures are found with high RMSD values at higher temperatures. For example, in the FEL at 307 K, a few shallow basins with ΔG* ~11-12 kJ/mol are observed, covering H-bonds that range from 3 to 6 and centered at the RMSD of 0.7 Å. Unlike the FEL at 263 K, this landscape does not feature a free energy minimum corresponding to 10 H-bonds. At 307 K, well-defined free energy basins are centered at the Cα-RMSD value of ~5 Å, with evenly distributed number of H-bonds ranging from 2 to 8. In addition, the span of RMSD's describing these basins decreases gradually with each additional backbone H-bond. Specifically, the RMSD values describing the basin with 2 H-bonds ranges from 3 Å to 5.5 Å, whereas those covering the basin with 8 H-bonds only ranges from 4.5 Å to 5 Å. The deepest basin is centered at (~5 Å, 2 H-bonds) with ΔG* ~8.9 kJ/mol, representing the greatest population of coil-like structures with the fewest backbone H-bond interactions.

Fewer hydrogen-bonded structures are observed at higher temperatures. As shown in the FEL at 359 K, free energy basins with high RMSD values are found with ≤ 6 H-bonds. Although a minor population is observed at (4.5 Å, 6 H-bonds), with ΔG* ~9.7 kJ/mol, the remainder is distributed almost-equally in deeper basins, with ΔG*'s of ~8-9 kJ/mol, corresponding to H-bonds ranging from 0 to 4. The decrease of hydrogen-bonding is progressive with temperature, as shown by the greater reduction of populations corresponding to ≤ 6 H-bonds. For example, the FEL at 420 K does feature a minor population at (4.5 Å, 6 H-bonds) that was previously identified in that at 359 K. Correspondingly, the

majority of coil-like structures are found without hydrogen-bonding, as indicated by the basin centered at (4 Å, 0 H-bonds) with a ΔG^* of 8.6 kJ/mol. At this temperature (420 K), a few minima are observed with H-bonds ranging from 2 to 4, and ΔG^* values are roughly 9-10 kJ/mol. The abundance of non-hydrogen-bonded structures are mostly evident in the FEL at 491 K, as the pronounced free energy basin centered at (4 Å, 0 H-bonds) is found with the lowest ΔG^*, i.e., 8 kJ/mol. At this temperature, very few structures containing $\geq$ 2 H-bonds are observed.

The loss of α-helical EX25-ALDL structures at higher temperatures is shown by the expansion of torsion angle distributions from the α_R-helix to other regions. As shown in **Figure 6.8**, at 263 K, the low-RMSD representative structure has most of its amino acids adopting the α_R-helical conformation with (ϕ, ψ) angles of (-57±15°, -45±15°); the two sets of angles centered at (0°, -45°) and at (60°, 0°) correspond to the N- and C-terminal residues, respectively. At 307 K, the high-RSMD representative structure has its N-terminal helical segment unwound, and the torsion angles of the corresponding residues are distributed over the β-sheet region. For example, the A502 is found with (ϕ, ψ = -120°, 150°), and the remaining three are found within ±15° nearby. Central residues, such as V508, have torsion angles that are concentrated in the α_R-helix and β-sheet regions, and the C-terminal A512 have (ϕ, ψ) angles that are near the α_L-helix region. At 420 K, the peptide appears fully unfolded from its initial α-helical structure, and a greater conformational space is explored, including the PP-II region at (-75°, -150°).

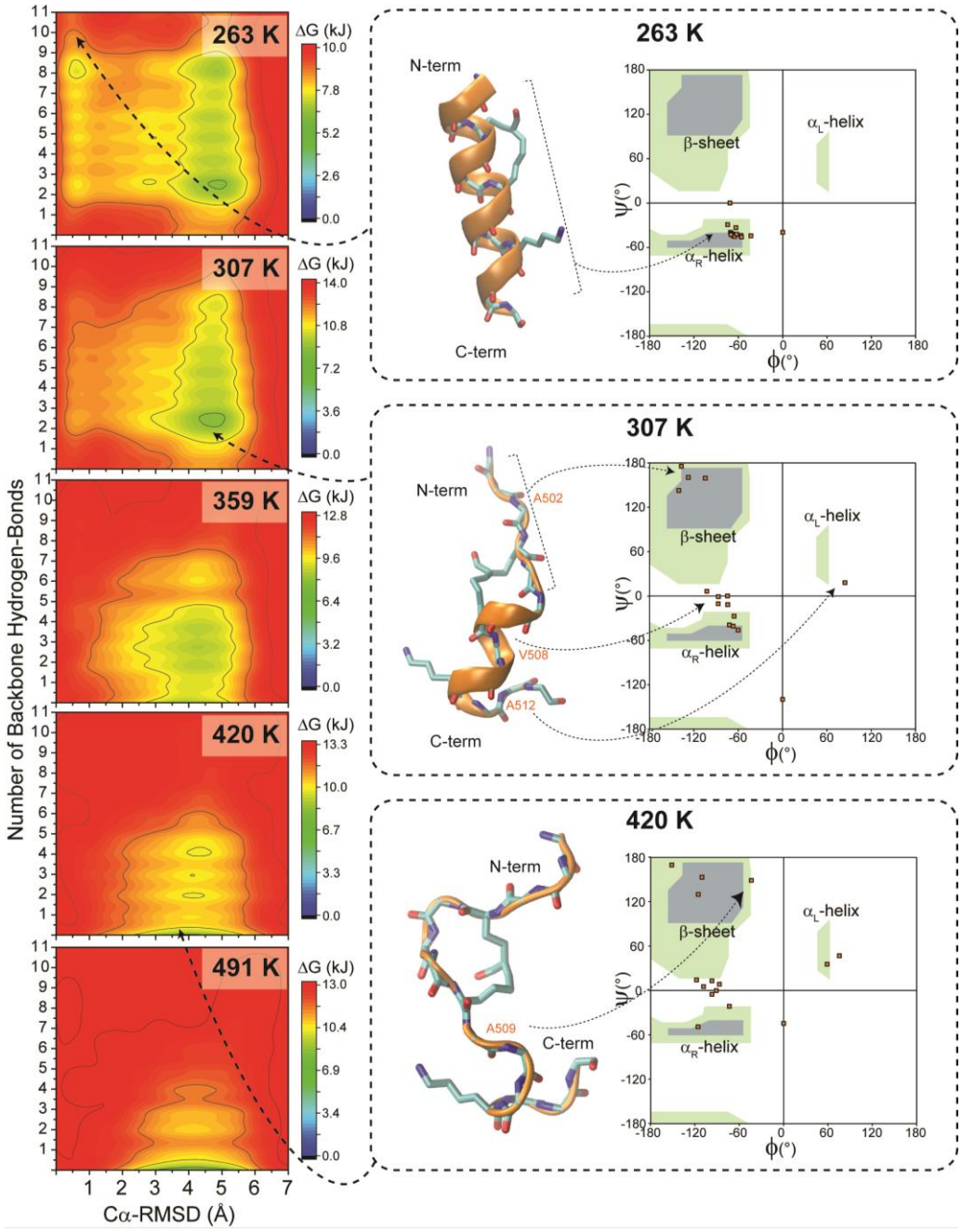


Figure 6.8 Representative free energy landscapes, structures, and torsion angles for EX25-ALDL. *Left*, FEL shows a contour map of ΔG*'s for all trajectory structures at a given REMD temperature, which are projected on the number of backbone hydrogen-bonds and Cα-RMSD values. *Center and right*, representative structures and Ramachandran plots for the corresponding free energy basins in the FEL's at 263 K, 307 K and 420 K. Each plot contains 15 points ( ), representing the total number of amino acids in EX25-ALDL. Amino acid indicated in orange point to regions of unique structural motifs in the Ramachandran plot. **Green** and **blue** shades in each plot correspond to the partially and fully allowed regions in the Ramachandran space [39].

The FEL's for EX19-DLNL also show multiple free energy minima (**Figure 6.9**). In contrast to those in EX25-ALDL, deep free energy basins corresponding to low-RMSD structures are observed at low temperatures. At 263 K, three free energy basins with RMSD <2.5 Å are observed in the FEL. First, a deep basin with ΔG* ~7.8 kJ/mol is centered at (1.2 Å, 8 H-bonds), which represents the major population of α-helical structures in the trajectory. Two shallower basins with ΔG* ~9.4 kJ/mol and ΔG* ~9.2 kJ/mol are also identified, corresponding to structures with 10 and 5 H-bonds, respectively. Additionally, several minima are observed with Cα-RMSD values ranging from 3.5 Å to 5.5 Å with 3 to 8 H-bonds. The deep basin with ΔG* ~8.1 kJ/mol is centered at (4.5 Å, 5 H-bonds) and represents the largest population of high-RMSD, or coil-like, structures. Another basin is also found at (4 Å, 7 H-bonds) with ΔG* ~8.7 kJ/mol, which is a smaller set of structures that have a narrower range of Cα-RMSD values, 3.5-4.2 Å. Apart from these two minima, another basin is found with fewer number of H-bonds, 3, and a smaller range of RMSD's, 4.3-4.7 Å. This FEL shows that EX19-DLNL structures equally populate the free energy minima corresponding to both high- and low-RMSD values, at 263 K.

A redistribution of populations occurs at higher temperatures. Three basins describing both low- and high-RMSD values are observed in the FEL at 307 K, and these basins are observed at the same (RMSD, H-bonds) coordinates as those at 263 K. However, the depth and width of each basin in the FEL vary with temperature. For example, the free energy minimum centered at (1.2 Å, 5 H-bonds) appears deeper at 307 K with ΔG* ~6.8 kJ/mol, relative to that at 263 K with ΔG* ~9.4 kJ/mol. Additionally, the basin corresponding to structures with 5 H-bonds covers a wider range of Cα-RMSD values at 307 K, i.e., 2.5-5.5 Å, relative to the range at 263 K, i.e., 3.5-5.5 Å. This basin also reflects a greater population of structures at 307 K, as indicated by the lower ΔG* value, 5.5 kJ/mol, than the one at 263 K, 8.1 kJ/mol. The population of high-RMSD structures with fewer number of H-bonds are also abundant at 307 K, as illustrated by the deep basin at (4.5 Å, 3 H-bonds) with ΔG* ~7.1 kJ/mol and a shallow basin at (4.5 Å, 1 H-bonds) with ΔG* ~7.1 kJ/mol. In summary, the FEL at 307 K reflects the abundance of both high- and low-RSMD structures, but a greater abundance is apparent in those containing fewer H-bonds.

A greater distribution over the high-RMSD structures are observed at 359 K. Multiple well-defined free energy minima are evenly distributed over a wide range of H-bonds from 2 to 9 and a narrow range of Cα-RMSD values from ~1 Å to ~1.5 Å. These basins are relatively similar in depth, with the ΔG* of each basin ranging from 7.7 to 8.1 kJ/mol. Multiple basins are distributed over wide range of H-bonds, 1-7 Å and higher range of RMSD values, 2.5-5 Å. However, these basins are lower in ΔG* values, 6.1-7.1 kJ/mol, indicating that the largest population in the trajectory is primarily consisted high-RMSD, coil-like, structures.

Successively fewer H-bonds are identified at higher temperatures. For example, a series of free energy basins is observed at 420 K, covering a range of RMSD values from 3 to 5.5 Å and backbone H-

bonds from 0 to 6. However, at 491 K, this series runs through a narrower range of H-bonds, from 0 to 4. Furthermore, the number of deep basins progressively decrease at the higher temperature, i.e., only one deep basin (ΔG^* ~7.3 kJ/mol) is observed at 491 K, whereas several equally deep basins (ΔG^* ~7.5 kJ/mol) are identified at 420 K. This result indicates that non-hydrogen-bonded coil-like structures are more prevalent at 491 K than at 420 K, implying extensive losses of α-helical segments at higher temperatures.

The loss of α-helical segments in EX19-DLNL is shown by the increase of torsion angle sampling (**Figure 6.9**). At 263 K, the N-terminal serine, S343, in the representative EX19-DLNL has torsion angles in β-sheet region at (-150°, 145°), whereas other residues primarily explore the α_R-helix region on the Ramachandran plot. The expansion of dihedral angles is more apparent at 307 K, as the long continuous helical segment is interspersed by an alanine (A352) that samples the left-handed α-helical region, at (58°, 45°). In addition, the N-terminal valine, V342, has the torsion angles of (-90°, -45°) in the region known as inverse γ-turn [40]. At 420 K, torsion angles of alanines, A346 and A348, are found in the β-sheet and PP-II regions, respectively, at (-140°, 121°) and (-65°, 145°).

In conclusion, the FEL analyses for EX25-ALDL and EX19-DLNL reveal the effect of temperature on the equilibrium of low-RMSD and high-RMSD structures. At low temperatures such as 263 K, highly hydrogen-bonded structures are observed. However, at higher temperatures such as 491 K, high-RMSD structures with no hydrogen-bonds are prominent. Hence, the two-state (helix-coil) equilibrium position is shifted towards the coil-like structures at higher temperatures, which is consistent with ssNMR study of alanines in the crosslinking domains of hydrated elastin (Chapter 5), AGADIR predictions of elastin's crosslinking domains (section C.1), and CD measurements on soluble elastin peptides [28].

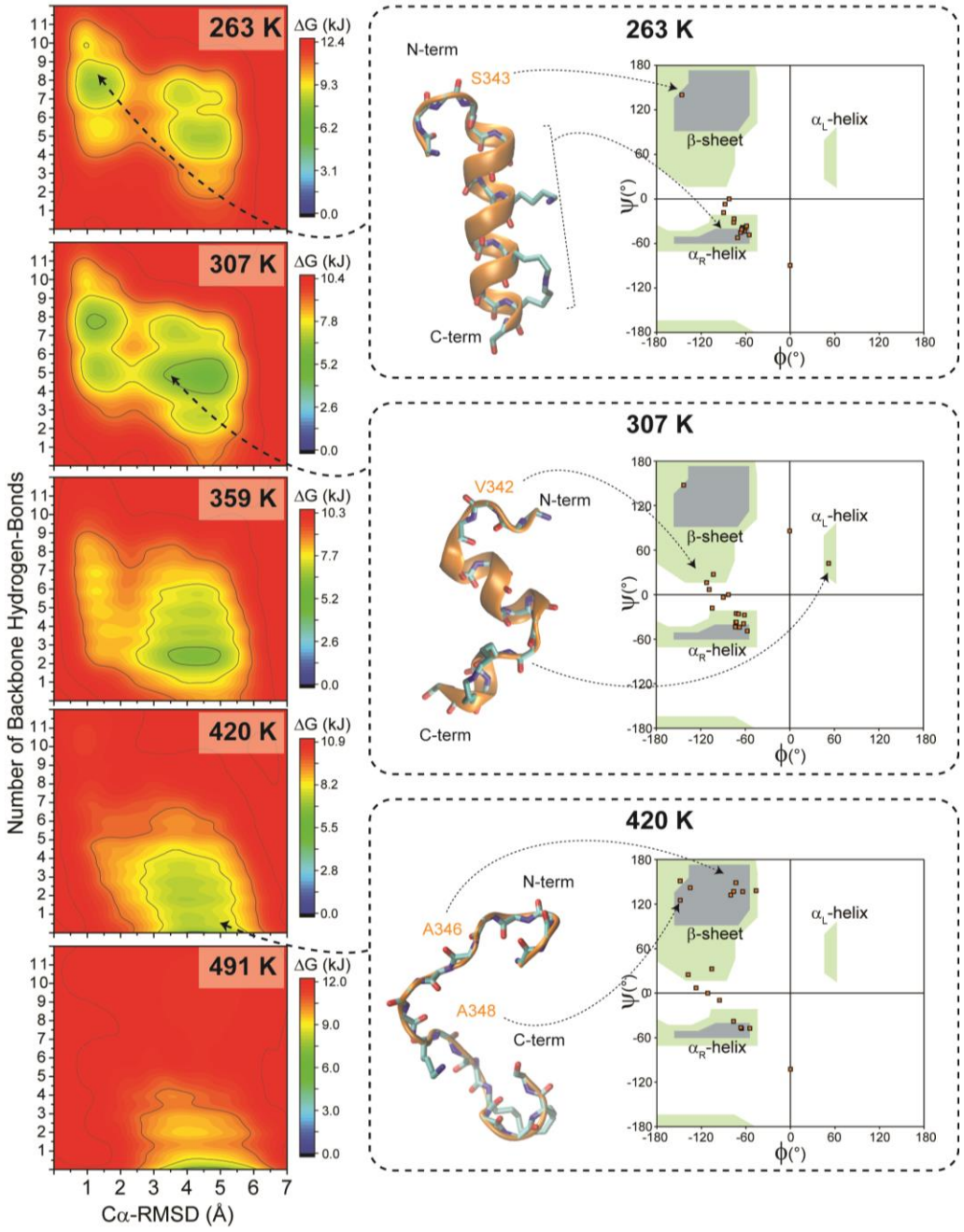


Figure 6.9 Representative free energy landscapes, structures, and torsion angles for EX19-DLNL. *Left*, FEL shows a contour map of ΔG*'s for all trajectory structures at a given REMD temperature, which are projected on the number of backbone hydrogen-bonds and Cα-RMSD values. *Center and right*, representative structures and Ramachandran plots for the corresponding free energy basins in the FEL's at 263 K, 307 K and 420 K. Each plot contains 18 points ( ), representing the total number of amino acids in EX19-DLNL. Amino acid indicated in orange point to regions of unique structural motifs in the Ramachandran plot. **Green** and **blue** shades in each plot correspond to the partially and fully allowed regions in the Ramachandran space [39].

C.4 Torsion angles and root-mean-square fluctuation analyses probe secondary structures and backbone mobility of alanines in EX25-ALDL and EX19-DLNL

C.4.1 Torsion angle distributions of alanines in high-RMSD structures are different from those in low-RMSD structures. Each structure in every trajectory in the REMD simulations was roughly defined as either helix- or coil-like, based on its Cα-RMSD value (section C.3), which, in turn, reflects the average atomic deviations of Cα atoms with respect to the initial coordinates. However, such analyses did not provide structural information about each residue in EX25-ALDL and EX19-DLNL peptides. Hence, torsion angles were extracted from the REMD trajectory's structures to examine the backbone conformation of alanines. The Ramachandran plots in **Figures 6.10** and **6.11** illustrate the (ϕ, ψ) angles of all Ala in the EX25-ALDL and EX19-DLNL trajectories at three representative temperatures, 263 K, 359 K, and 491 K. I.e., each point corresponds to the given Ala's torsion angles at each time point in the trajectory. These plots reflect the relative populations of the low-RMSD (**Figures 6.10-6.11**, *top*) and high-RMSD (**Figures 6.10-6.11**, *bottom*) structures for each Ala in the respective elastin domain. For example, no points are observed in the α_R-helical region of the Ramachandran plot for an alanine that never folds into this structure. In contrast, a cluster of points in the α_R-helical region indicates the adoption of this secondary structure by a given alanine in either the high- or low-RMSD population.

The Ramachandran plots for alanines in EX25-ALDL show tight distributions of torsion angles in the low-RMSD structures and a diversity of structural motifs in the high-RMSD structures. As shown in **Figure 6.10** (*top*), at 263 K (*black points*), the low-RMSD A502 has (ϕ, ψ) angles that are focused mostly at (-60°, -45°), the α_R-helix region, with very slight deviation, ±15°. In contrast, the high-RMSD A502 are scattered across multiple regions, including α_R-helix, α_L-helix (60°, -45°), PP-II (-65°, 145°), and β-strand (-150°, 155°), in **Figure 6.10** (*bottom*). The major population is observed in the α_R-helix region, as seen for low-RMSD. However, a broader distribution of (ϕ, ψ) angles are sampled for the high-RMSD structures; i.e., ϕ ranges from -45° to -120°, and ψ ranges from -60° to 30°. In other words, a more pronounced deviation, ±30°, is observed for the α_R-helical A502 with higher RMSD's.

A redistribution of (ϕ, ψ) angles is also identified with temperature variation. As illustrated in **Figure 6.10** (*top*), the low-RMSD A502 population at 263 K (*black points*) decreases with temperature, with fewer structures at 359 K (*green points*) and at 491 K (*red points*). In contrast, high-RMSD A502 has a more diffused cluster of torsion angles in the α_R-helical region at 263 K (*black points*), which expands into the regions characterizing β-sheet and α_L-helix at higher temperatures like 359 K (*green points*) and 491 K (*red points*). These features are generally observed in the Ramachandran plots for other alanines in EX25-ALDL. This result reflects a decrease of α-helical content in the peptide at higher temperatures, consistent with ssNMR results (Chapter 5) and other experimental reports [28].

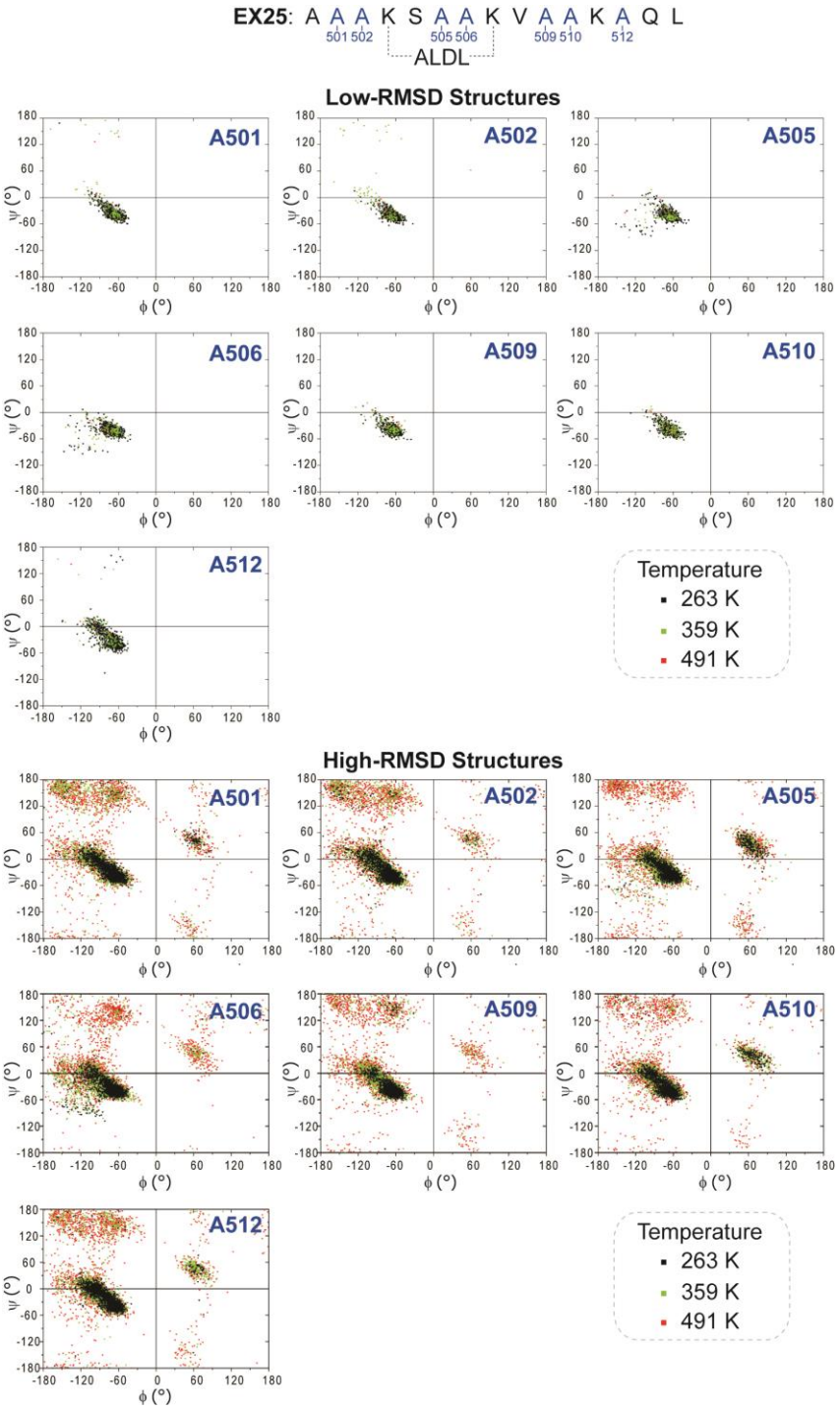


Figure 6.10 Torsion angles of alanines in low- and high-RMSD structures of EX25-ALDL from three representative temperatures, 263 K (black), 359 K (green) and 491 K (red). Amino acid sequence of EX19 shows the relative position of each alanine (A) in the chain. Each point (•) in the Ramachandran plot represents the (ϕ, ψ) angles of a given alanine in EX25-ALDL at a given trajectory time point. The total number of points, 2000, is distributed over the low-RMSD (*top*) and high-RMSD (*bottom*) populations.

The Ramachandran plots for alanines in EX19-DLNL show similar features to those in EX25-ALDL (**Figure 6.11**). Namely, alanines in the low-RMSD structures sample a narrow distribution of torsion angles, whereas those in the high-RMSD structures explore a wider range of structural motifs. For example, at 263 K, the sole population of low-RMSD A351 (*black points*) is identified with (ϕ, ψ) angles that are concentrated at (-60°±15°, -45°±15°), the α_R-helix region. In contrast, two major populations of high-RMSD A502 are found in the α_R- and α_L-helix regions, which are observed at (-60°±30°, -45°±30°) and (60°±30°, -45°±30°), respectively. In addition, minor populations are scattered across multiple regions of the Ramachandran plot, including PP-II at (-65°±30°, 145°±30°), and β-strand at (-150°±30°, 155°±30°).

A temperature-dependent redistribution of torsion angles is also observed in each plot. As illustrated in **Figure 6.11** (*top*), the helix-state A351 population at 263 K (*black points*) decreases at higher temperatures, as shown by the lesser numbers of structures at 359 K (*green points*) and at 491 K (*red points*). The high-RMSD A502 samples a broader torsion angle distribution at the two corresponding higher temperatures. For example, at 359 K (*green points*), the alanine population is scattered around the outskirts of the α_R- and α_L-helix regions. At 491 K (*red points*), a large population is observed in the entire β-sheet region, covering the ϕ angle that ranges from -180° to -30° and the ψ angle ranging from 120° to 180°. Sparsely distributed A502 populations are also observed at (-120°±60°, -150°±30°) and (60°±30°, -150°±30°). In general, the Ramachandran plots reflect the increase of torsion angle sampling at higher temperatures, which reduces the number of helix-like structures but increases the coil-like population.

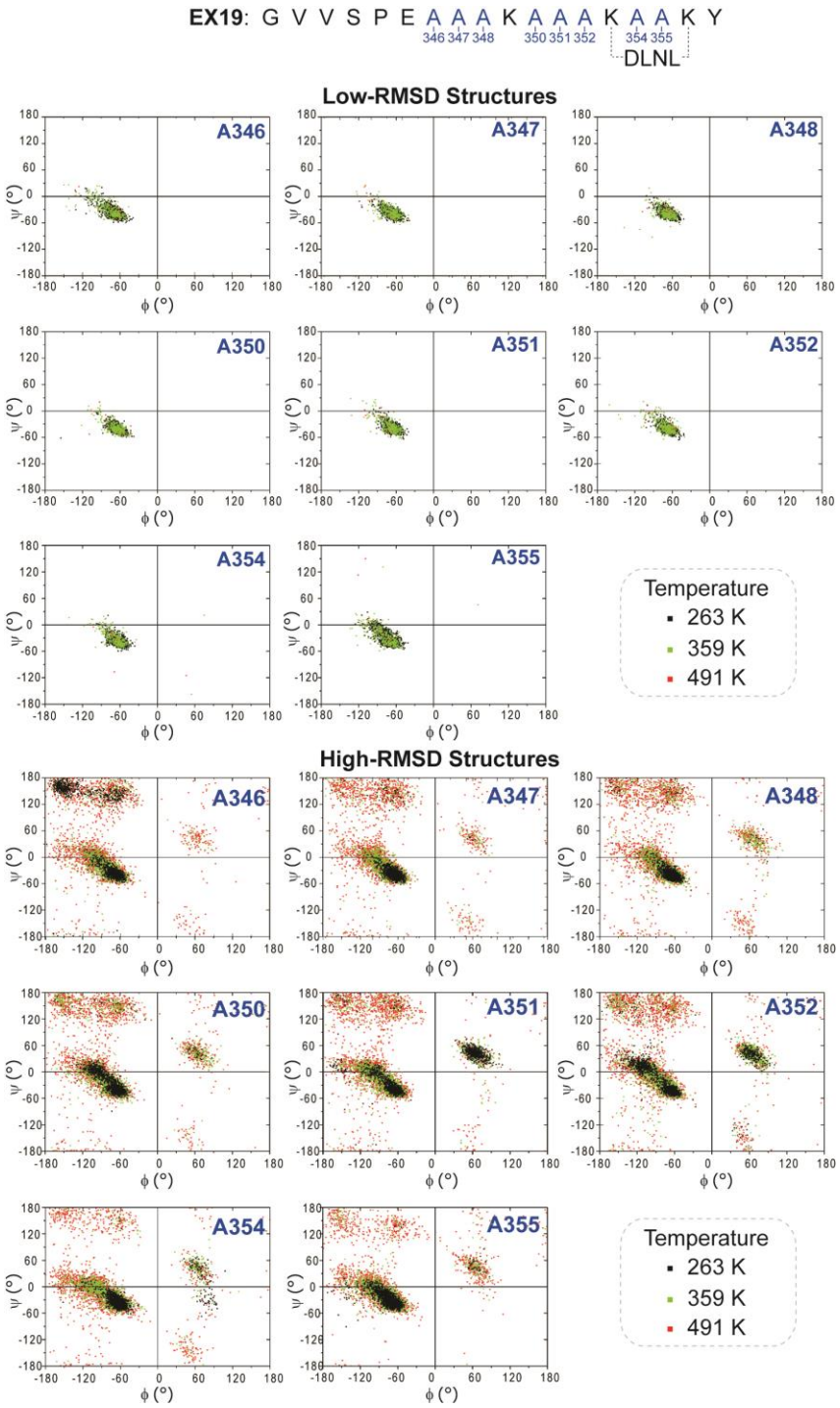


Figure 6.11 Torsion angles of alanines in low- and high-RMSD structures of EX19-DLNL extracted from three representative temperatures, 263 K (black), 359 K (green) and 491 K (red). Amino acid sequence of EX19 shows the relative position of each alanine (A) in the chain. Each point (•) in the Ramachandran plot represents the (ϕ, ψ) angles of a given alanine in EX19-DLNL at a given trajectory time point. The total number of points, 2000, is distributed over the low-RMSD (*top*) and high-RMSD (*bottom*) populations.

C.4.2 Backbone mobility of alanines varies with temperature. The effect of temperature on the backbone mobility of alanines in EX25-ALDL and EX19-DLNL is examined using the root-mean-square fluctuation (RMSF) analysis. The RMSF calculates the atomic displacement of a particular atom, or groups of atoms, with respect to a set of reference coordinates (Chapter 3). The reference structure is calculated from the average atomic coordinates of EX25-ALDL and EX19-DLNL trajectories. Hence, the Cα-RMSF values reflect the atomic fluctuations of the Cα atoms during the simulation, which is a qualitative measure of backbone mobility. High Cα-RMSF values imply high mobility, whereas low Cα-RMSF values reflect restricted backbone fluctuations.

In general, terminal alanines in representative EX25-ALDL trajectories have higher mobility than the central residues (**Figure 6.12a**). At 263 K, the alanines have Cα-RMSF's that range between 2 and 3 Å, except for the N-terminal A500 and the central A505. The A500 has the highest RMSF (4 Å) among others, indicating a large motional-amplitude. In contrast, the A505 is found with the lowest Cα-RMSF value (~1.6 Å), reflecting a limited conformational space due to the presence of ALDL moiety (Chapter 4). This plot suggests that a relatively mobile backbone is observed in the N-terminus, as the corresponding residue (A500) is unwound form the α-helical structure. This alanine is likely found in random coil, as it samples a large distribution of torsion angles and exhibits rapid interconversions of structural motifs [41-43].

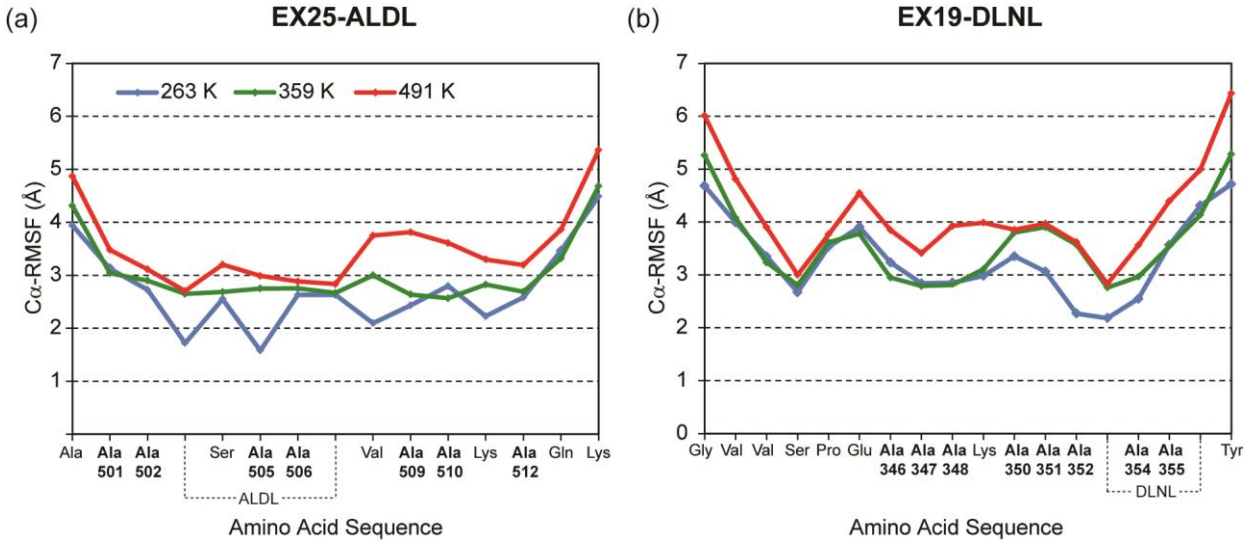


Figure 6.12 Representative Cα-RMSF values of each amino acid in (a) EX25-ALDL and (b) EX19-DLNL at three representative REMD temperatures, 263 K (blue), 359 K (green) and 491 K (red).

Higher backbone fluctuations are observed at higher temperatures (**Figure 6.12a**). At 359 K, most of the alanines in EX25-ALDL sample a narrower range of conformational space, as shown by little

variation of Cα-RMSF's, i.e., 2.5 ±0.5 Å. At 491 K, the alanines exhibit faster motions with larger amplitudes, as a larger range of RMSF values (~3-4 Å) is observed. The N-terminal A500 is relatively more mobile than other residues at 359 K and 491 K, with Cα-RSMF's of 4.2 Å and 4.8 Å, respectively. The plot corresponding to simulations at 491 K shows a striking contrast to the other plots, as residues A509 and A510 have significantly higher fluctuations, with RMSF~3.8 Å and ~3.7 Å, respectively. These results imply that a faster motion is exhibited and a larger conformational sampling is achieved in simulations that are performed at higher temperatures.

Terminal alanines in the representative EX19-DLNL trajectories also have higher mobility than the central residues (**Figure 6.12b**). At 263 K, the Cα-RMSF for each alanine spanning A346-A351 varies slightly, ±0.1 Å, indicating that the corresponding residues have similar motional features with each other. Smaller ranges of motions are sampled by A352 and A354, as shown by the RMSF values of 2.3 Å and 2.5 Å, respectively. The C-terminal A355 has the highest Cα-RMSF among others, 3.5 Å, reflecting the largest motional amplitude among other alanines.

A greater conformational sampling for EX19-DLNL residues is achieved at higher temperatures. At 359 K, Cα-RMSF values for central alanines A346-A348 and C-terminal A355 are increased slightly (<0.2 Å), relative to those at 263 K. In contrast, significantly higher RMSF's (>0.6 Å) at 359 K are observed for A350-A352 and A354, relative to those at 263 K. Additionally, at 491 K, the Cα-RMSF values for central residues A346-A38 and C-terminal A354-A355 are higher than those at 359 K. These results indicate that all regions of EX19-DLNL generally access greater ranges of motions at higher temperatures.

Variable-temperature studies shed light on the effect of temperature on the structure and dynamics of alanines in elastin's crosslinking domains. The REMD results support a temperature-dependent conformational equilibrium in these Ala-rich domains, consistent with the ssNMR studies (Chapter 5). The α-helical structures in EX25-ALDL and EX19-DLNL trajectories are mostly identified at lower temperatures, whereas random coil populations are observed at all temperatures. Backbone interactions influence the stability of α-helices, as the decrease of hydrogen-bond formations results in the reduction of the α-helical content. Torsion angle and Cα-RMSF analyses identified the increase of conformational space at higher temperatures, which increases the likelihood for random coil populations in the crosslinking regions.

D. SUMMARY AND CONCLUSIONS

The predominance of α-helical structures in tropoelastin's Ala-rich crosslinking domains has recently been scrutinized, as this secondary structure is often measured to be lower than predicted [44]. With CD measurements, the α-helical content in these regions are shown to increase with cooling, indicating the temperature-dependence of this conformation. However, the detailed structural characterization of alanines in elastin and elastin peptides is not feasible by CD spectroscopy.

Solid-state NMR spectroscopy and MD simulations offers valuable insight into the molecular structure of alanines in hydrated NRSMC elastin. Alanines in the crosslinking domains are found in a conformational equilibrium between α-helix and random coil (Chapter 3 & 4). Additionally, the equilibrium position shifts towards the α-helix at cooler temperatures, reflecting a coil-to-helix transition of the alanines (Chapter 5). This temperature-dependence of α-helices has been extensively studied in peptide-based polymers using helix-coil transition theory [45-47]. However, the application of this helix-coil transition theory to the examination of α-helical structures in elastin's crosslinking domains has not been done until this study.

Temperature-replica exchange MD was used to perform all-atom simulations on EX25-ALDL and EX19-DLNL peptides, which are representative of native Ala-rich crosslinking domains in elastin. The results show that the α-helical content in these regions is inversely related to temperature, as fewer α-helical structures were observed at higher temperatures. The α-helical content in EX25-ALDL and EX19-DLNL is correlated with the degree of backbone stabilization. Lower helical content corresponds to fewer numbers of hydrogen-bonds, and the loss of hydrogen-bond interactions shifts the conformational equilibrium towards random coil. The REMD results reveal the temperature-dependence of helical propensities in EX25-ALDL and EX19-DLNL, consistent with previous CD and solution-state NMR results on tropoelastin's crosslinking domains [28, 48].

The energetics of peptide folding were used to study the effect of temperature on the helix-coil equilibrium in EX25-ALDL and EX19-ALDL. The relative free energies corresponding to REMD structures were calculated, and the results were plotted as free energy landscapes. The FEL of a given temperature features multiple well-defined low energy minima (basins) that are distributed across both low-RMSD and high-RMSD structures. At low temperatures, multiple minima exist in both low- and high-RMSD structures, indicating the accessibility of helix- and coil-like features during folding simulations. At high temperatures, the FEL reveal the predominance of high-RMSD, coil-like, structures. The presence of multiple free energy basins in the folding pathways for EX25-ALDL and EX19-DLNL indicates the likelihood of finding helix- or coil-like structures in the folding simulation, consistent with the

conformational feature of intrinsically disordered proteins [41-43] and in agreement with the conformational ensemble proposed for elastin peptides [48]

The dynamical features of alanines in EX25-ALDL and EX19-DLNL were analyzed using root-mean-square fluctuation calculation of Cα atoms. Large-amplitude motions were observed in terminal residues of EX25-ALDL and EX19-DLNL, which are often associated with the fraying of α-helix [49, 50] due to the absence of backbone H-bonding interactions. The helical fraying allows terminal residues to sample a larger conformational space. Hence, terminal alanines in elastin's Ala-rich crosslinking domains are most likely found in random coil, consistent with the helix-coil transition theory [51] and ssNMR results of hydrated NRSMC elastin (Chapter 5).

In conclusion, REMD simulations provide insights into the effect of temperature on the structure and dynamics of elastin's crosslinking domains. REMD simulations indicate that the α-helical content in EX25-ALDL and EX19-DLNL is dependent on temperature. Namely, a higher helical content is observed at lower temperatures, consistent with CD and NMR studies [28]. The simulations also identified large amplitude-motions in terminal alanines, which explains the origin of the random coil population in the crosslinking domains. At low temperatures, backbone fluctuations of the alanines are significantly reduced, and a smaller torsion angle space is sampled. Hence, longer α-helices are likely to be formed at low temperatures, as previously demonstrated by ssNMR lineshapes (Chapter 5).

E. REFERENCES

1. Garnier, J., J.-F. Gibrat, and B. Robson, *GOR method for predicting protein secondary structure from amino acid sequence*, in *Methods in Enzymology*. 1996, Academic Press. p. 540-553.
2. Jones, D.T., *Protein secondary structure prediction based on position-specific scoring matrices.* Journal of Molecular Biology, 1999. **292**(2): p. 195-202.
3. P Y Chou, a., and G.D. Fasman, *Empirical predictions of protein conformation.* Annual Review of Biochemistry, 1978. **47**(1): p. 251-276.
4. Muñoz, V., and L. Serrano, *Elucidating the folding problem of helical peptides using empirical parameters.* Nature Structural & Molecular Biology, 1994. **1**(6): p. 399-409.
5. Muñoz, V., and L. Serrano, *Elucidating the folding problem of helical peptides using empirical parameters. II. Helix macrodipole effects and rational modification of the helical content of natural peptides.* Journal of Molecular Biology, 1995. **245**(3): p. 275-296.
6. Muñoz, V., and L. Serrano, *Elucidating the folding problem of helical peptides using empirical parameters. III. Temperature and pH dependence.* Journal of Molecular Biology, 1995. **245**(3): p. 297-308.
7. Muñoz, V., and L. Serrano, *Local versus nonlocal interactions in protein folding and stability – an experimentalist's point of view.* Folding and Design, 1996. **1**(4): p. R71-R77.
8. Muñoz, V., and L. Serrano, *Development of the multiple sequence approximation within the AGADIR model of α-helix formation: Comparison with Zimm-Bragg and Lifson-Roig formalisms.* Biopolymers, 1997. **41**(5): p. 495-509.
9. López-Hernández, E., P. Cronet, L. Serrano, and V. Muñoz, *Folding kinetics of Che Y mutants with enhanced native α-helix propensities.* Journal of Molecular Biology, 1997. **266**(3): p. 610-620.
10. Lacroix, E., A.R. Viguera, and L. Serrano, *Elucidating the folding problem of α-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters1.* Journal of Molecular Biology, 1998. **284**(1): p. 173-191.
11. Viguera, A.R., V. Villegas, F.X. Avilés, and L. Serrano, *Favourable native-like helical local interactions can accelerate protein folding.* Folding and Design, 1997. **2**(1): p. 23-33.
12. Serrano, L., J.-L. Neira, J. Sancho, and A.R. Fersht, *Effect of alanine versus glycine in [alpha]-helices on protein stability.* Nature, 1992. **356**(6368): p. 453-455.
13. Villegas, V., A.R. Viguera, F.X. Avilés, and L. Serrano, *Stabilization of proteins by rational design of α-helix stability using helix/coil transition theory.* Folding and Design, 1996. **1**(1): p. 29-34.

14. Sugita, Y., and Y. Okamoto, *Replica-exchange molecular dynamics method for protein folding.* Chemical Physics Letters, 1999. **314**(1–2): p. 141-151.
15. Sugita, Y., and Y. Okamoto, *Replica-exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscape.* Chemical Physics Letters, 2000. **329**(3–4): p. 261-270.
16. Zhou, R., *Replica exchange molecular dynamics method for protein folding simulation*, in *Protein Folding Protocols*, Y. Bai and R. Nussinov, Editors. 2006, Humana Press: Totowa, NJ. p. 205-223.
17. Mori, T., N. Miyashita, W. Im, M. Feig, and Y. Sugita, *Molecular dynamics simulations of biological membranes and membrane proteins using enhanced conformational sampling algorithms.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2016. **1858**(7, Part B): p. 1635-1651.
18. Baldissera, G., M.P. dos Santos Cabrera, J. Chahine, and J.R. Ruggiero, *Role of peptide–peptide interactions in aggregation: protonectins observed in equilibrium and replica exchange molecular dynamics simulations.* Biochemistry, 2015. **54**(13): p. 2262-2269.
19. Gnanakaran, S., H. Nymeyer, J. Portman, K.Y. Sanbonmatsu, and A.E. García, *Peptide folding simulations.* Current Opinion in Structural Biology, 2003. **13**(2): p. 168-174.
20. Phillips, J.C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kalé, and K. Schulten, *Scalable molecular dynamics with NAMD.* Journal of Computational Chemistry, 2005. **26**(16): p. 1781-1802.
21. Tanner, D.E., K.-Y. Chan, J.C. Phillips, and K. Schulten, *Parallel generalized Born implicit solvent calculations with NAMD.* Journal of chemical theory and computation, 2011. **7**(11): p. 3635-3642.
22. Frishman, D., and P. Argos, *Knowledge-based protein secondary structure assignment.* Proteins: Structure, Function, and Bioinformatics, 1995. **23**(4): p. 566-579.
23. Heinig, M., and D. Frishman, *STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins.* Nucleic Acids Research, 2004. **32**(Web Server issue): p. W500-W502.
24. Lindahl, E., B. Hess, and D. van der Spoel, *GROMACS 3.0: a package for molecular simulation and trajectory analysis.* Molecular Modeling Annual, 2001. **7**(8): p. 306-317.
25. Berendsen, H.J.C., D. van der Spoel, and R. van Drunen, *GROMACS: A message-passing parallel molecular dynamics implementation.* Computer Physics Communications, 1995. **91**(1): p. 43-56.
26. Van Der Spoel, D., E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, and H.J.C. Berendsen, *GROMACS: Fast, flexible, and free.* Journal of Computational Chemistry, 2005. **26**(16): p. 1701-1718.

27. Lifson, S., and A. Roig, *On the Theory of helix-coil transition in polypeptides.* Journal of Chemical Physics, 1961. **34**(6): p. 1963-1974.
28. Tamburro, A.M., A. Pepe, and B. Bochicchio, *Localizing α-helices in human tropoelastin: assembly of the elastin "puzzle".* Biochemistry, 2006. **45**(31): p. 9518-9530.
29. Djajamuliadi, J., T.F. Kagawa, K. Ohgo, and K.K. Kumashiro, *Insights into a putative hinge region in elastin using molecular dynamics simulations.* Matrix Biology, 2009. **28**(2): p. 92-100.
30. Dyksterhuis, L.B., C. Baldock, D. Lammie, T.J. Wess, and A.S. Weiss, *Domains 17–27 of tropoelastin contain key regions of contact for coacervation and contain an unusual turn-containing crosslinking domain.* Matrix Biology, 2007. **26**(2): p. 125-135.
31. Mithieux, S.M., and A.S. Weiss, *Elastin*, in *Advances in Protein Chemistry*. 2005, Academic Press. p. 437-461.
32. Mithieux, S.M., S.G. Wise, M.J. Raftery, B. Starcher, and A.S. Weiss, *A model two-component system for studying the architecture of elastin assembly in vitro.* Journal of Structural Biology, 2005. **149**(3): p. 282-289.
33. Baldock, C., A.F. Oberhauser, L. Ma, D. Lammie, V. Siegler, S.M. Mithieux, Y. Tu, J.Y.H. Chow, F. Suleman, M. Malfois, S. Rogers, L. Guo, T.C. Irving, T.J. Wess, and A.S. Weiss, *Shape of tropoelastin, the highly extensible protein that controls human tissue elasticity.* Proceedings of the National Academy of Sciences, 2011. **108**(11): p. 4322-4327.
34. García, A.E., and K.Y. Sanbonmatsu, *α-Helical stabilization by side chain shielding of backbone hydrogen bonds.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(5): p. 2782-2787.
35. Best, R.B., N.V. Buchete, and G. Hummer, *Are current molecular dynamics force fields too helical?* Biophysical Journal, 2008. **95**(1): p. L07-9.
36. Tanizaki, S., J. Clifford, B.D. Connelly, and M. Feig, *Conformational sampling of peptides in cellular environments.* Biophysical Journal, 2008. **94**(3): p. 747-759.
37. Nymeyer, H., and A.E. García, *Simulation of the folding equilibrium of α-helical peptides: a comparison of the generalized Born approximation with explicit solvent.* Proceedings of the National Academy of Sciences, 2003. **100**(24): p. 13934-13939.
38. Hastie, T., R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction, second edition*. 2009: Springer New York.
39. Ramachandran, G.N., C. Ramakrishnan, and V. Sasisekharan, *Stereochemistry of polypeptide chain configurations.* Journal of Molecular Biology, 1963. **7**: p. 95-9.
40. Rose, G.D., L.M. Glerasch, and J.A. Smith, *Turns in peptides and proteins.* Advances in Protein Chemistry, 1985. **37**: p. 1-109.

41. Jha, A.K., A. Colubri, K.F. Freed, and T.R. Sosnick, *Statistical coil model of the unfolded state: resolving the reconciliation problem.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(37): p. 13099-13104.
42. Smith, L.J., K.M. Fiebig, H. Schwalbe, and C.M. Dobson, *The concept of a random coil: residual structure in peptides and denatured proteins.* Folding and Design, 1996. **1**(5): p. R95-R106.
43. Toal, S., and R. Schweitzer-Stenner, *Local order in the unfolded state: conformational biases and nearest neighbor interactions.* Biomolecules, 2014. **4**(3): p. 725-773.
44. Miao, M., J.T. Cirulis, S. Lee, and F.W. Keeley, *Structural determinants of cross-linking and hydrophobic domains for self-assembly of elastin-like polypeptides.* Biochemistry, 2005. **44**(43): p. 14367-14375.
45. Doty, P., A.M. Holtzer, J.H. Bradbury, and E.R. Blout, *Polypeptides. II. The configuration of polymers of γ-benzyl-L-glutamate in solution.* Journal of the American Chemical Society, 1954. **76**(17): p. 4493-4494.
46. Doty, P., and J.T. Yang, *Polypeptides . VII. Poly-γ-benzyl-L-glutamate: the helix-coil transition in solution.* Journal of the American Chemical Society, 1956. **78**(2): p. 498-500.
47. Doty, P., A. Wada, J.T. Yang, and E.R. Blout, *Polypeptides. VIII. Molecular configurations of poly-L-glutamic acid in water-dioxane solution.* Journal of Polymer Science, 1957. **23**(104): p. 851-861.
48. Tamburro, A.M., B. Bochicchio, and A. Pepe, *The dissection of human tropoelastin: from the molecular structure to the self-assembly to the elasticity mechanism.* Pathologie Biologie, 2005. **53**(7): p. 383-389.
49. Chakrabartty, A., A.J. Doig, and R.L. Baldwin, *Helix capping propensities in peptides parallel those in proteins.* Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(23): p. 11332-11336.
50. Chakrabartty, A., and R.L. Baldwin, *Stability of alpha-helices.* Advances in Protein Chemistry, 1995. **46**: p. 141-76.
51. Matheson, R.R., and H.A. Scheraga, *Calculation of the Zimm-Bragg cooperativity parameter σ from a simple model of the nucleation process.* Macromolecules, 1983. **16**(7): p. 1037-1043.

CHAPTER 7. SUMMARY AND CONCLUDING REMARKS

A. CHARACTERIZATION OF ALANINES IN NATIVE ELASTIN

Elastin is the principal protein component of the elastic fiber, known to confer elasticity and resilience in vertebrate tissues. Many models have been proposed to explain the mechanism for elasticity and the macromolecular organization of elastin. Most of these studies agree on the existence of local order within this protein [1-3]. In particular, Tamburro's tropoelastin model identifies the crosslinking domains as α-helices and the hydrophobic domains as labile structures that interconvert among PP-II, β-turns and random coil [2]. However, this model describes the soluble (tropoelastin) monomer, not the crosslinked insoluble polymer (mature elastin). Additionally, this representation was constructed based on exon-by-exon studies of elastin-like peptides in a mixture of trifluoroethanol and water, which induces the formation of secondary structures [3]. Therefore, the search for a suitable model for mature, crosslinked, elastin continues.

Solid-state NMR spectroscopy is effective for the investigation of structure and dynamics in native elastin. The characterization of alanines in elastin offers insights into the protein's molecular organization, as this amino acid is abundantly found in both hydrophobic and crosslinking domains. However, high-resolution ssNMR studies of the alanines necessitate a sample that contains a high level of isotopic enrichment. Previously, isotopically-labeled elastin samples were obtained from neonatal rat smooth muscle cells (NRSMC) cultures via *direct substitution* [4]. Its efficacy was demonstrated for essential amino acids such as Val and Lys, as well as the special case of Gly [4]. However, its application to Ala resulted in a negligible isotopic enrichment level (~4%). For this reason, several enrichment schemes were explored to increase the level of isotopes that are incorporated in the protein (Chapter 2). Among several investigated approaches, the combination of enzymatic inhibition and excess supply of isotopically-labeled Ala in the culture media produces elastin with a high enrichment level (~80%).

Solid-state NMR measurements at 37 °C identifies the structural diversity of alanines in hydrated [U-^{13}C]Ala and [U-^{13}C,^{15}N]Ala NRSMC elastin (Chapter 3). Conformation-dependent ^{13}C chemical shifts indicate the presence of random coil population in the hydrophobic domains, which reflects rapid interconversions of multiple structures [5]. Fast and large-amplitude motions of alanines in these regions are confirmed by the result of ^{13}C $T_{1\rho}$ measurements, and this mobile population is also detected by the refocused INEPT experiment. The random coil population is also supported by the sequence-dependence in ^{13}C-Ala chemical shifts. The calculated values for fully random coil alanines in rat tropoelastin show good agreement with experiment. Namely, the three resolved peaks observed in the 2D ^{13}CO-^{13}Cα

correlation spectrum are consistent with the predicted ^{13}CO-**Ala** features corresponding to for R_{-1}-**Ala**-Gly, R_{-1}-**Ala**-Z (Z ≠ Ala, Gly), and R_{-1}-**Ala**-Ala motifs in random coil. Hence, this *neighboring residue* effect validates the coil assignment of alanines in hydrated NRSMC elastin.

A conformational equilibrium is identified between α-helix and random coil alanines in the crosslinking domains of NRSMC elastin. Conformation-dependent ^{13}C chemical shifts indicate that alanines in these regions are found in the α-helix. This population has relatively low mobility and is selectively detected by the CP experiment. In addition, shorter ^{13}C $T_{1\rho}$'s were observed, suggesting that slower kHz-scale motions are exhibited in these domains. However, the deconvolution of ^{13}C-Ala lineshape indicated a lower α-helical content than theoretically predicted. Namely, only 45-49% helical component is observed by ssNMR, whereas a conventional structural algorithm like GOR IV [6] predicts 58% helical content in tropoelastin, which originates from Ala in the crosslinking domains. This result indicates that a small portion (9-13%) of alanines in the crosslinking domains are found in random coil, implying the instability of α-helices at 37 °C, consistent with previous elastin studies [7].

Computational studies were performed to trace the origin of random coil population in the crosslinking domains and its relationship with the α-helical structure (Chapter 4). Molecular dynamics (MD) simulations were conducted on EX25-ALDL and EX19-DLNL peptides, which are the representatives for the crosslinking regions in native elastin. Based on the Cα-RMSD values with respect to the initial α-helical configuration, MD simulations identified helix- and coil-like structures. These two respective populations correspond to low- and high-Cα-RMSD structures in the trajectories. The low-RMSD structures are characterized by ≥60% helical content. This high propensity for α-helices was, partly, due to the effect of backbone dehydration that is imposed by elastin's bifunctional crosslinks. Specifically, large ALDL and DLNL moieties inhibit the access of water to the backbone atoms, increasing the propensity for protein-protein (intramolecular) hydrogen-bonding that stabilizes the α-helix. In contrast, the high-RMSD, coil-like, structures have ≤60% helical content. This population corresponds to peptides with segments that exhibit large-amplitude fluctuations, sampling multiple structural motifs like turns, β-strand and PP-II. These mobile residues correspond to the helical fraying [8, 9] of EX25-ALDL and EX19-DLNL peptides. Furthermore, a coil-to-helix or helix-to-coil transition occurs in these peptides, as characterized by multiple minima in the free energy landscapes of peptide folding, indicating a conformational equilibrium in the crosslinking domains [2].

The effect of temperature on the conformational equilibrium of alanines in hydrated NRSMC elastin was probed using variable-temperature ssNMR measurements (Chapter 5). The ^{13}C-Ala CP intensities below T_g is markedly increased, due to the reduction of dipolar couplings. This increase of CP efficiency upon cooling is attributed to the change of protein dynamics from "liquid-like" to "solid-like" [10].

The ^{13}C-Ala DP lineshapes acquired above T_g is significantly different from those obtained below T_g, indicating a change of conformations upon cooling. Helix-coil transitions theory and simulations of NMR peaks were used to help interpret the ^{13}CO-Ala lineshapes above and below T_g. The results indicated that an increase of α-helical content is observed in the crosslinking domains as the temperature is reduced. A mixture of short and long helical segments is present in the Ala-rich crosslinking domains above T_g, whereas mostly longer helices are identified below T_g. These results indicated that conformational equilibrium for alanines in the crosslinking domains is shifted towards the α-helix with the decrease of temperature.

The temperature-dependent helix-coil equilibrium in elastin's crosslinking domains was further examined using computational methods (Chapter 6). AGADIR predictions provided temperature-dependent profiles of α-helical propensities in elastin's crosslinking domains. Temperature-replica-exchange simulations were performed on EX25-ALDL and EX19-DLNL peptides over a broad distribution of temperatures, which allows for a thorough conformational sampling. The REMD simulations confirmed the abundance of low-RMSD, helix-like structures at low temperatures, which supports the helix-melting profile predicted by AGADIR. Higher helical content and greater hydrogen-bond interactions at the backbone indicate increased stability of α-helices at low temperature. In contrast, the high-RMSD, coil-like, population is predominant at high temperatures. This population corresponds to structures with common motifs, such as PP-II, β-strands, turns, etc. The free energy landscapes for the EX25-ALDL and EX19-DLNL show multiple low energy states, which is consistent with the features of intrinsically disordered proteins [11, 12]. The FEL's indicate that conformational equilibrium positions are shifted towards the α-helix upon cooling, consistent with the ssNMR results.

In summary, solid-state NMR and MD simulations offer insights into the molecular structure of native elastin. SSNMR characterization of alanines in hydrated NRSMC elastin provides a framework for understanding the structure and dynamics in the protein's hydrophobic and crosslinking domains. MD simulations offer a complementary result to the ssNMR, elucidating the structure-function relationship in elastin's crosslinking domains.

B. A MOLECULAR MODEL FOR ELASTIN'S STRUCTURE

A refined model is constructed with the results of our investigations of alanines in NRSMC elastin by experimental and computational methods (inset of **Figure 7.1**). Above T_g or at the physiological condition, elastin's hydrophobic domains are largely characterized by random coil, featuring large-amplitude fluctuations of the backbone chains and preferences for common structural motifs such as PP-II, β-strands and turns. Large-amplitude kHz-scale motions are also exhibited by the alanines in these regions. Conversely, the crosslinking domains are best described by conformational equilibria between random coil and short α-helical segments. Helix-coil transitions occur in each of these domains, resulting in a lower α-helical content than typically predicted by structural prediction algorithms. However, the regions where elastin's crosslinks are formed such as exons 19 and 25 have higher α-helical propensities. Alanines in these regions sample a smaller range of conformational space, because the backbone motion is limited by the crosslink moieties. Elastin's molecular structure above T_g corresponds to a system that has a free energy landscape featuring multiple low energy states, and facile conformational transitions occur among the corresponding populations, which are common in common in IDP [11, 12].

In contrast, below T_g, elastin's crosslinking domains are characterized by the polydispersity of α-helical chains, and the hydrophobic domains feature a mixture of common structural motifs. Below the freezing point of water, backbone motions in both hydrophobic and crosslinking regions are significantly reduced. Helix-coil transitions in the crosslinking domains are suspended, resulting in the formation of longer α-helices. The free energy landscapes corresponding to these domains have fewer low energy states and deeper free energy basins, suggesting a lower probability for conformational transitions below T_g.

Numerous implications follow, regarding the relationships among Ala's secondary structures, the role of hydration, and the protein's elastomeric property. Elastin's flexibility is traced to the random coil or *statistical coil* population in the hydrophobic domains above T_g, which represent swiftly interconverting structures, consistent with Tamburro's concept of conformational equilibria in tropoelastin [2]. Hydration also facilitates such interconversions, as water is known to be the plasticizer in elastin [15]. Conversely, the resilience in native elastin is traced to the stability of α-helices in the regions where bi- or tetrafunctional crosslinks are formed. Although helix-coil transitions occur in the crosslinking domains, Lys-modified moieties limit the access of water to the backbone atoms. For instance, the DLNL crosslink in EX19 sequesters water molecules from hydrating the corresponding region, thereby enhancing the stability of the α-helix. Local sidechain interactions also contribute to the α-helical stability in elastin's

domains. For instance, Lys and Glu residues in EX19 form a salt-bridge, which blocks hydration water. Correspondingly, the stability of the α-helix contributes to elastin's resilience.

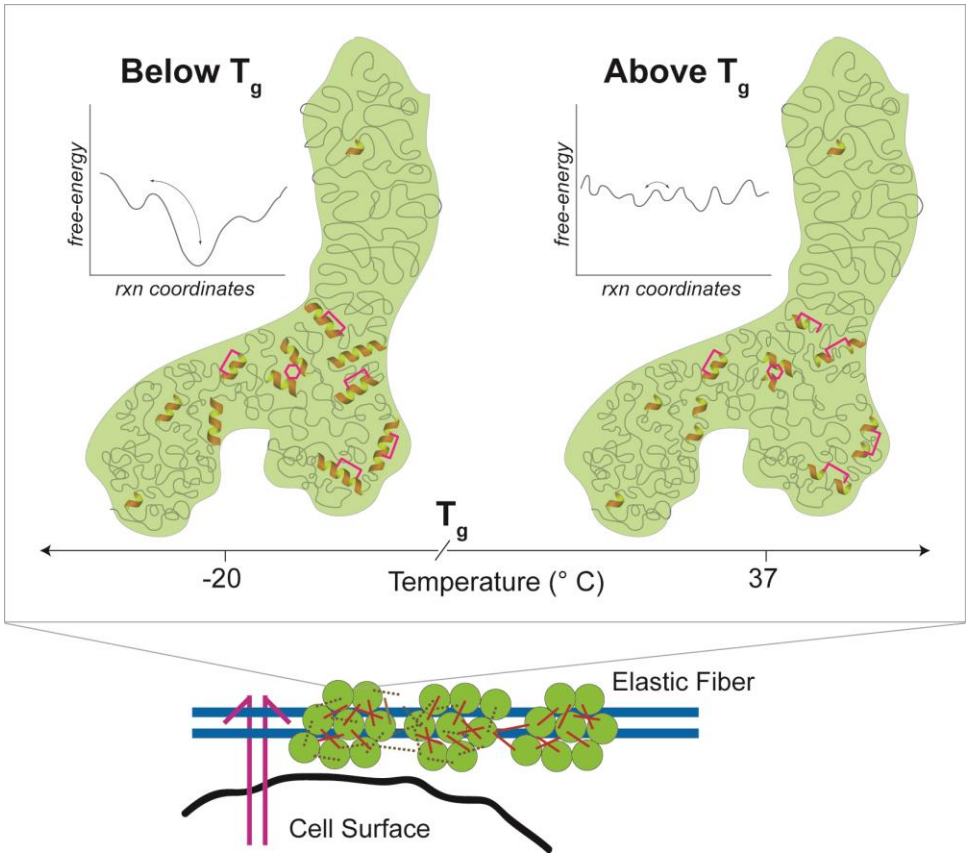


Figure 7.1 Proposed molecular structure for native elastin. *Top*, molecular structure for each tropoelastin monomer below T_g (*left*) and above T_g (*right*), as illustrated using Weiss's SAXS model [13] (**green**); free energy landscapes illustrate the topology of folding pathways in the corresponding structures. In each model, curly lines represent random coil, which corresponds to swift fluctuations of several common motifs, such as PP-II, β-strand, turns, etc. Ala-rich crosslinking domains are indicated by α-helices (**brown**/**yellow**). Intramonomer crosslinking is represented by brackets and hexagons (**magenta**). *Bottom*, diagram of the elastic fiber is reproduced from **Figure 1.2** (Chapter 1 and [14]).

The molecular structure for native elastin below T_g serves as a model for a dehydrated sample, which is known to be brittle and inflexible [15]. Previously, a fully dehydrated BNL elastin at 37 °C was shown to be dynamically identical with one that is hydrated and cooled below T_g [10]. Elastin's flexibility is curbed by the decrease of hydrogen-bond interactions between the protein and hydration water. Furthermore, the brittleness in dehydrated elastin may correspond to the absence of conformational transitions, as reflected by the features of the free energy landscapes. Namely, there are fewer low energy states in the FEL, which decreases the likelihood of facile conformational interconversions.

In conclusion, the results of this research project mark an important milestone in the study of native elastin. Despite recent discussions regarding elastin as a model for intrinsically disordered proteins [16-18], ssNMR measurements confirm the existence of local order within hydrated NRSMC elastin. In particular, α-helical segments with different chain lengths are observed in the crosslinking domains, in thermodynamic equilibrium with the random coil. The helix-coil transitions theory offers a plausible explanation for the significant contrast in elastin's α-helical content, i.e., above and below T_g. MD simulations affirm the temperature-dependence of the α-helical propensities in elastin's crosslinking domains. This study shows that elastin's crosslinks help stabilize α-helices in the Ala-rich crosslinking domains, which shed light on the source of resilience in this polymer.

C. REFERENCES

1. Tamburro, A.M., B. Bochicchio, and A. Pepe, *Dissection of human tropoelastin: exon-by-exon chemical synthesis and related conformational studies.* Biochemistry, 2003. **42**(45): p. 13347-13362.
2. Tamburro, A.M., B. Bochicchio, and A. Pepe, *The dissection of human tropoelastin: from the molecular structure to the self-assembly to the elasticity mechanism.* Pathologie Biologie, 2005. **53**(7): p. 383-389.
3. Tamburro, A.M., A. Pepe, and B. Bochicchio, *Localizing α-helices in human tropoelastin: assembly of the elastin "puzzle".* Biochemistry, 2006. **45**(31): p. 9518-9530.
4. Perry, A., M.P. Stypa, J.A. Foster, and K.K. Kumashiro, *Observation of the glycines in elastin using ^{13}C and ^{15}N solid-state NMR spectroscopy and isotopic labeling.* Journal of the American Chemical Society, 2002. **124**(24): p. 6832-6833.
5. Smith, L.J., K.M. Fiebig, H. Schwalbe, and C.M. Dobson, *The concept of a random coil: residual structure in peptides and denatured proteins.* Folding and Design, 1996. **1**(5): p. R95-R106.
6. Garnier, J., J.-F. Gibrat, and B. Robson, *GOR method for predicting protein secondary structure from amino acid sequence*, in *Methods in Enzymology*. 1996, Academic Press. p. 540-553.
7. Miao, M., J.T. Cirulis, S. Lee, and F.W. Keeley, *Structural determinants of cross-linking and hydrophobic domains for self-assembly of elastin-like polypeptides.* Biochemistry, 2005. **44**(43): p. 14367-14375.
8. Chakrabartty, A., A.J. Doig, and R.L. Baldwin, *Helix capping propensities in peptides parallel those in proteins.* Proc Natl Acad Sci U S A, 1993. **90**(23): p. 11332-11336.
9. Chakrabartty, A., and R.L. Baldwin, *Stability of alpha-helices.* Adv Protein Chem, 1995. **46**: p. 141-76.
10. Perry, A., M.P. Stypa, B.K. Tenn, and K.K. Kumashiro, *Solid-state ^{13}C NMR reveals effects of temperature and hydration on elastin.* Biophys J, 2002. **82**(2): p. 1086-1095.
11. Bálint, M., S. István, and D. Zsuzsanna, *The expanding view of protein–protein interactions: complexes involving intrinsically disordered proteins.* Physical Biology, 2011. **8**(3): p. 035003.
12. Burger, V., T. Gurry, and C. Stultz, *Intrinsically Disordered Proteins: Where Computation Meets Experiment.* Polymers, 2014. **6**(10): p. 2684.
13. Baldock, C., A.F. Oberhauser, L. Ma, D. Lammie, V. Siegler, S.M. Mithieux, Y. Tu, J.Y.H. Chow, F. Suleman, M. Malfois, S. Rogers, L. Guo, T.C. Irving, T.J. Wess, and A.S. Weiss, *Shape of tropoelastin, the highly extensible protein that controls human tissue elasticity.* Proceedings of the National Academy of Sciences, 2011. **108**(11): p. 4322-4327.

14. Wagenseil, J.E., and R.P. Mecham, *New insights into elastic fiber assembly.* Birth Defects Research Part C: Embryo Today: Reviews, 2007. **81**(4): p. 229-240.
15. Partridge, S., *Elastin.* Advances in Protein Chemistry, 1963. **17**: p. 227-302.
16. Tompa, P., and M. Fuxreiter, *Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions.* Trends in Biochemical Sciences, 2008. **33**(1): p. 2-8.
17. Roberts, S., M. Dzuricky, and A. Chilkoti, *Elastin-like polypeptides as models of intrinsically disordered proteins.* FEBS Letters, 2015. **589**(19): p. 2477-2486.
18. Rauscher, S., and R. Pomès, *Structural Disorder and Protein Elasticity*, in *Fuzziness: Structural Disorder in Protein Complexes*, M. Fuxreiter and P. Tompa, Editors. 2012, Springer US: New York, NY. p. 159-183.

APPENDIX 1. COMPOSITION OF NRSMC GROWTH MEDIA

Table A.1. Standard Growth Media for NRSMC Cultures

NRSMC Culture Media Composition	% Volume[1]
Dulbecco's Modified Eagle's Medium (DMEM)[2]	87
Fetal Bovine Serum, Catalog No. 26140	10
Sodium Pyruvate, 100 mM (100×), Catalog No. 11360	1
Antibiotic-Antimycotic (100×), Catalog No. 15240	1
Non-Essential Amino Acids (NEAA)[3]	1

Notes:

[1] Growth media are prepared from stock solutions with either (1×) or (100×) concentration. Each media component is purchased from Sigma-Aldrich (Sigma-Aldrich, St. Louis MO), unless stated otherwise.

[2] DMEM (1×), Catalog No.11885, is used for standard NRSMC cultures (without isotopic enrichment), detailed in **Table A.3**. The DMEM composition can be customized to fit the desired isotopic enrichment scheme (**Appendix 2**). For instance, DMEM without glucose (Catalog No. 11966) is available. A custom DMEM solution can also be purchased without select ingredients (Chapter 2, sections B.2.2-B.2.3).

[3] NEAA solution, 10 mM (100×), Catalog No.11140, is purchased from GIBCO (GIBCO, New York, NY), its composition is shown in **Table A.2**. For isotopic enrichment with NRSMC, a homemade NEAA solution is reconstituted from the corresponding amino acid powders, which are dissolved in autoclaved Milli-Q Ultrapure water (**Appendix 2.2**).

Table A.2. Composition of Non-Essential Amino Acid Solution, 10 mM (100×)

Amino Acids	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 A.3. Composition of DMEM (1×) from Sigma-Aldrich, Cat. No. 11885

Amino Acids	Concentration (mg/L)
L-Arginine hydrochloride	84.00
L-Cysteine dihydrochloride	62.57
L-Glutamine	54.00
Glycine	30.00
L-Histidine hydrochloride monohydrate	42.00
L-Isoleucine	104.80
L-Leucine	104.80
L-Lysine hydrochloride	146.20
L-Methionine	30.00
L-Phenylalanine	66.00
L-Serine	42.00
L-Threonine	95.20
L-Tryptophan	16.00
L-Tyrosine disodium salt dihydrate	103.79
L-Valine	93.60
Vitamins	**Concentration (mg/L)**
Choline chloride	4.00
D-calcium pantothenate	4.00
Folic acid	4.00
Niacinamide	4.00
Pyridoxine hydrochloride	4.00
Riboflavin	0.40
Thiamine hydrochloride	4.00
L-Inositol	7.20
Inorganic Salts	**Concentration (mg/L)**
Calcium chloride anhydrous	200.00
Ferric nitrate nonahydrate	0.10
Magnesium sulfate anhydrous	97.67
Potassium chloride	400.00
Sodium bicarbonate	3700.00
Sodium chloride	6400.00
Sodium phosphate monobasic monohydrate	125.00
Other	**Concentration (mg/L)**
D-Glucose (dextrose)	1000.00
Phenol red	15.00
Sodium pyruvate	110.00

APPENDIX 2. AMINO ACID CONCENTRATIONS IN CUSTOM DMEM AND NEAA SOLUTIONS

Table A.4. Supplementation of Custom DMEM[1] for Isotopic Enrichment with NRSMC Cultures

Component		DMEM, 87%	NEAA, 1%	Total, 88%	Total, 88%	Hm-NEAA[2]
		(1×), mg/L	(1×), mg/L	(1×), mg/L	(100×), mg/L	Mass, mg
NEAA Composition	L-Glycine[3]	26.1	7.5	33.6	3360	21.5
	L-Alanine[4]	**0.0**	**8.9**	**8.9**	**890**	**5.7*[5]**
	L-Asparagine	0.0	13.2	13.2	1320	8.4
	L-Aspartic Acid	0.0	13.3	13.3	1330	8.5
	L-Glutamic Acid	0.0	14.7	14.7	1470	9.4
	L-Proline	0.0	11.5	11.5	1150	7.4
	L-Serine	36.5	10.5	47.0	4704	30.1
Supplied to DMEM	L-Valine[3]	81.4	0.0	81.4	8143	52.1
	L-Glutamine[3]	508.1	0.0	508.1	-	40.6
	Sodium pyruvate[3]	95.7	1.1	96.8	9680	62.0
	Glucose[6]	870.0	0.0	870.0	87000	556.8

Notes:

[1] All DMEM solutions are purchased from Sigma-Aldrich (Sigma-Aldrich, St. Louis, MO). DMEM (Catalog No. 11185) is used for the enrichment strategy of combined ALT inhibitor and excess alanines (Chapter 2, section B.2.4). Custom DMEM without L-glycine, L-valine, and L-glutamine is purchased for the [1-^{13}C, 99%]pyruvate enrichment, and DMEM without glucose (Catalog No. 11966) is used in the [U-^{13}C, 99%]glucose enrichment (Chapter 2, sections B.2.2-B.2.3).

[2] Homemade NEAA **(Hm-NEAA)** solution is prepared by dissolving powders of amino acids in autoclaved Milli-Q Ultrapure water. The masses of powders (in mg) are calculated for 2 T-75 flasks.

[3] L-Glycine, L-valine, L-glutamine, and sodium pyruvate (which are absent in custom DMEM[1]) are dissolved into the homemade stock NEAA solution. However, the L-glutamine is supplemented into the culture media weekly (as needed), because it degrades in solution over short periods.

[4] **L-Alanine** is substituted with isotopically-enriched L-Ala (e.g., [U-^{13}C,99%]Ala, [U-^{13}C,99%, ^{15}N, 98%]Ala, etc.), which is purchased from Cambridge Isotope Laboratories (Cambridge Isotope Laboratories, Tewksbury, MA).

***[5]** For the enrichment strategy utilizing ALT inhibition and excess alanines, isotopically-enriched L-Ala is supplied in multifold concentrations of the standard requirement, **5.7** mg. For instance,

57 mg and 114 mg of alanines are required to prepare **Hm-NEAA** solutions containing 10-fold and 20-fold L-Ala concentrations, respectively.

[6] [U-^{13}C, 99%]Glucose is purchased from Cambridge Isotope Laboratories (Cambridge Isotope Laboratories, Tewksbury, MA).

APPENDIX 3. BIOSYNTHETIC PATHWAYS IN NRSMC AND THE EXTENT OF ISOTOPIC SCRAMBLING

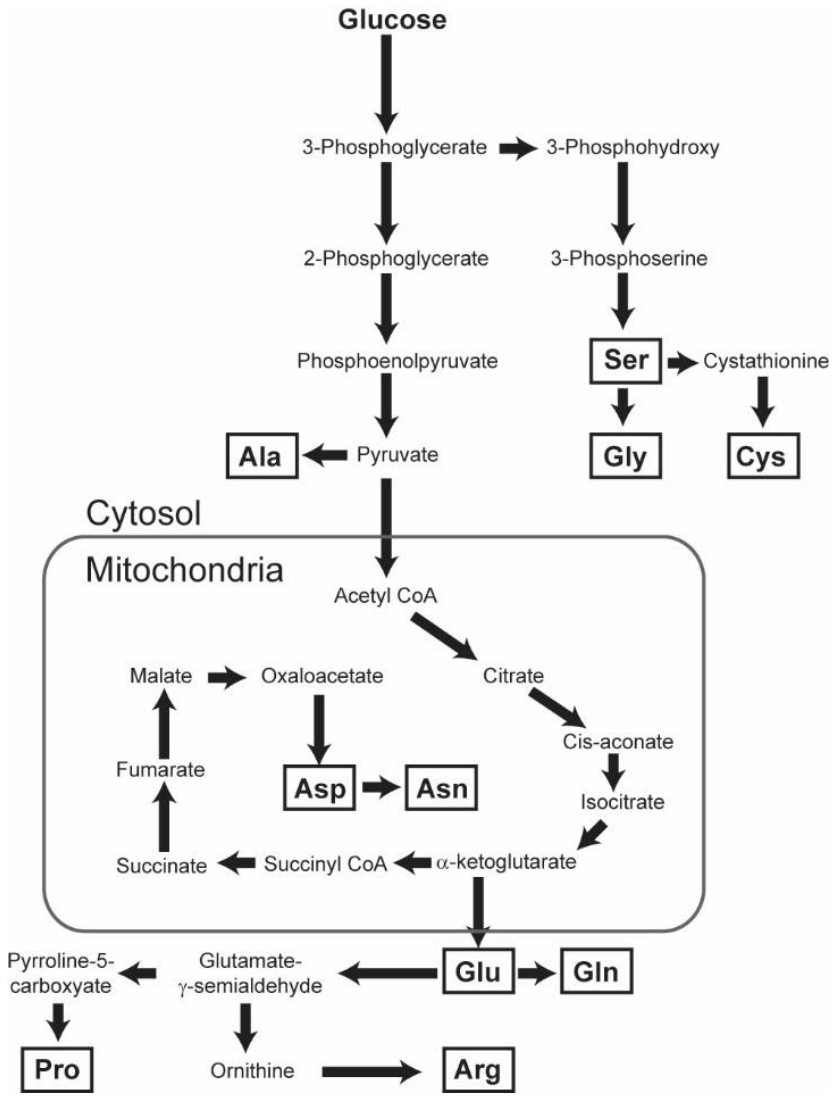


Figure A.1 Biosynthetic pathways of essential amino acids in mammalian cells. This figure is adopted and modified from [1].

Note: Isotopes from [U- ^{13}C]glucose could be distributed into several amino acids, including Gly, Ser, Cys, Ala, Asp, Asn, Arg, Pro, Glu, and Gln

[1] Salway, J.G.; Biosynthesis of the non-essential amino acids. In *Metabolism at a Glance*, edition no.3; John Wiley & Sons: 2013; p. 79.

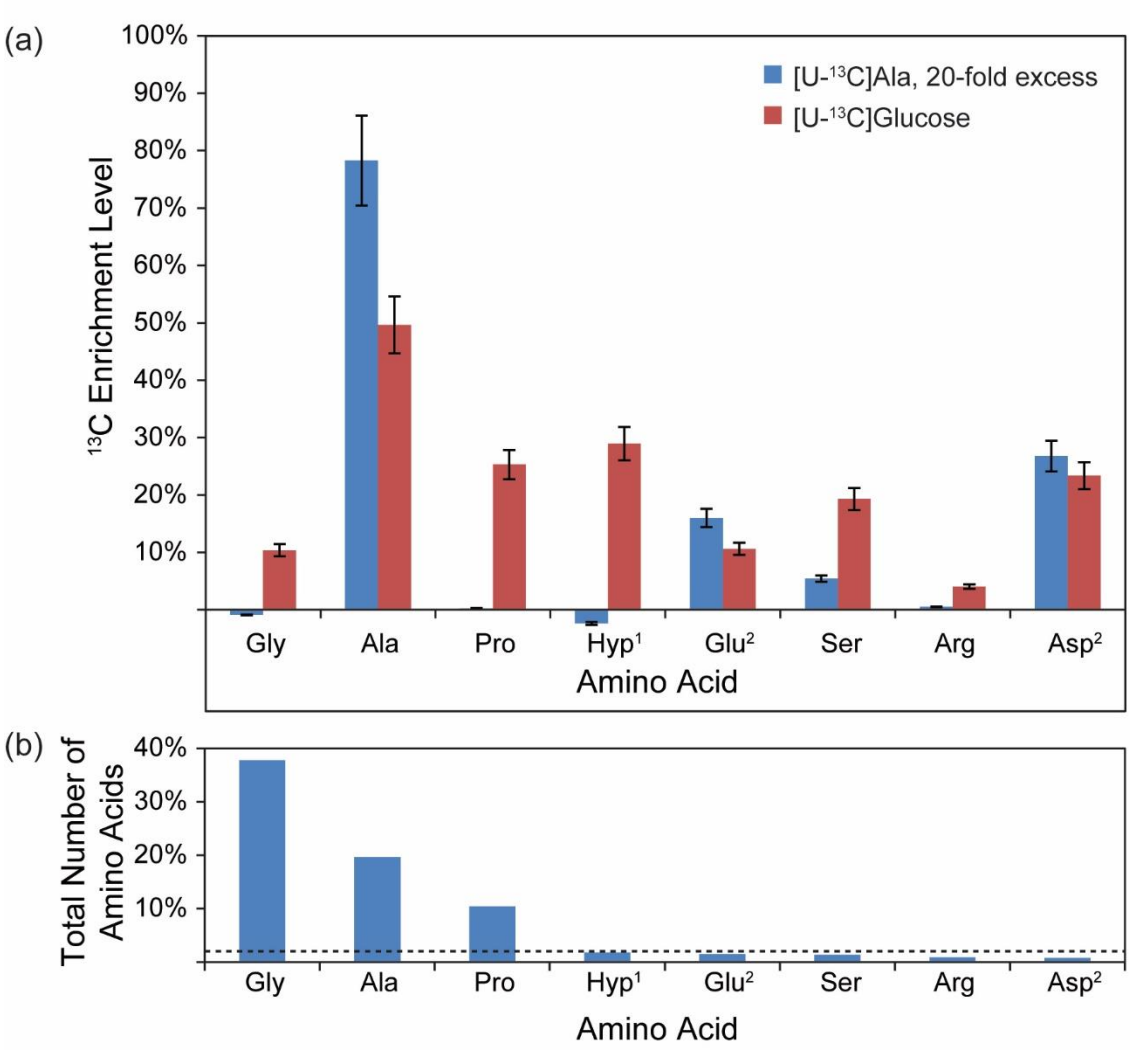


Figure A.2 (a) Degree of isotopic enrichment in some essential amino acids in NRSMC elastin, grown in [U-^{13}C]glucose (■) and 20-fold excess [U-^{13}C]Ala (■); (b) percent essential amino acids in the primary structure of rat tropoelastin. In (a), bar intensities reflect the ^{13}C enrichment levels, and the error bars (⌶) represent the typical uncertainty in the LC-MS analysis, ±10%. In (b), bar intensities reflect the total number of essential amino acids, based on the primary structure of rat tropoelastin, and the horizontal dashed line (- - -) indicates the 2% level of the total content.

Notes:

[1] Hydroxyproline is a hydroxylated Pro, modified post-translationally.

[2] During hydrolysis, Asn and Gln are converted into Asp and Glu, respectively. Hence, the ^{13}C enrichment level for Asp reflects those of Asp and hydrolyzed Asn. Analogously, the ^{13}C enrichment level for Glu reflects those of Glu and hydrolyzed Gln.

APPENDIX 4. ^{13}C SPECTRA OF ALA IN NRSMC ELASTIN AND OF L-ALA IN SOLUTION

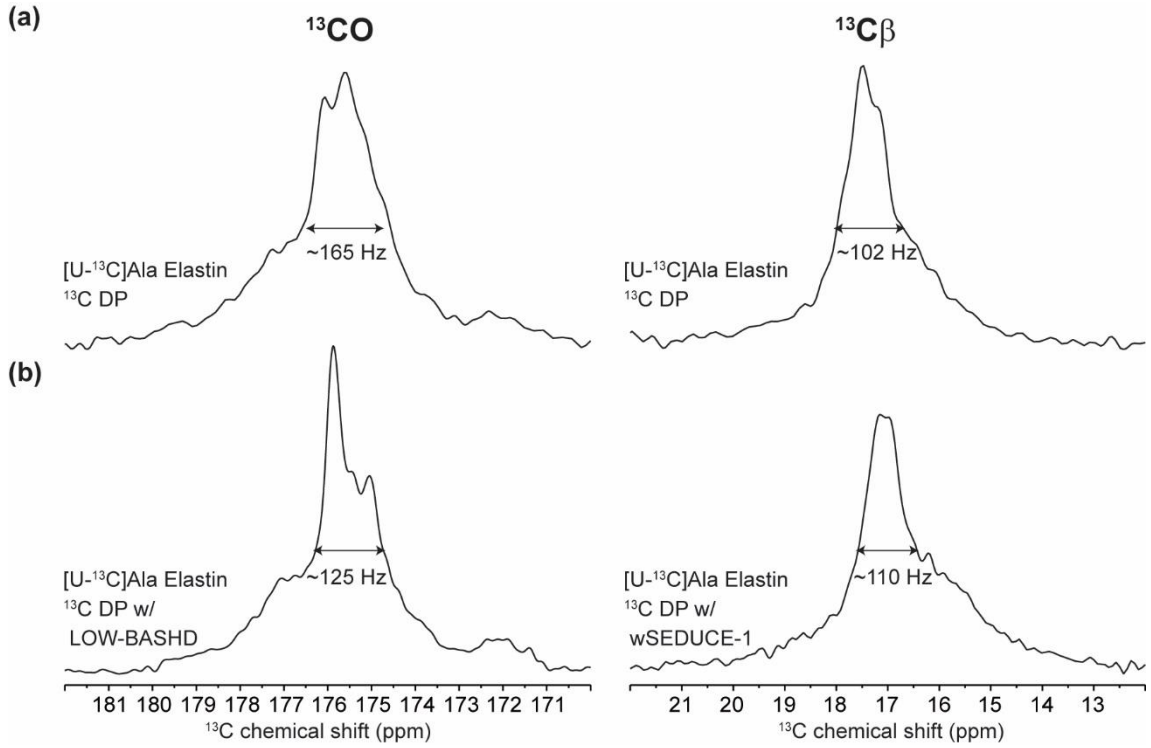


Figure A.3. ^{13}CO- and ^{13}Cβ-Ala lineshapes of hydrated [U-^{13}C]Ala elastin at 37° C, acquired using (a) DP and (b) DP with homonuclear decoupling. LOW-BASHD decouples $^{1}J_{C\alpha CO}$ and wSEDUCE-1 decouples $^{1}J_{C\alpha C\beta}$. Double-headed arrow represents full widths at half maximum (FWHM) in each lineshape.

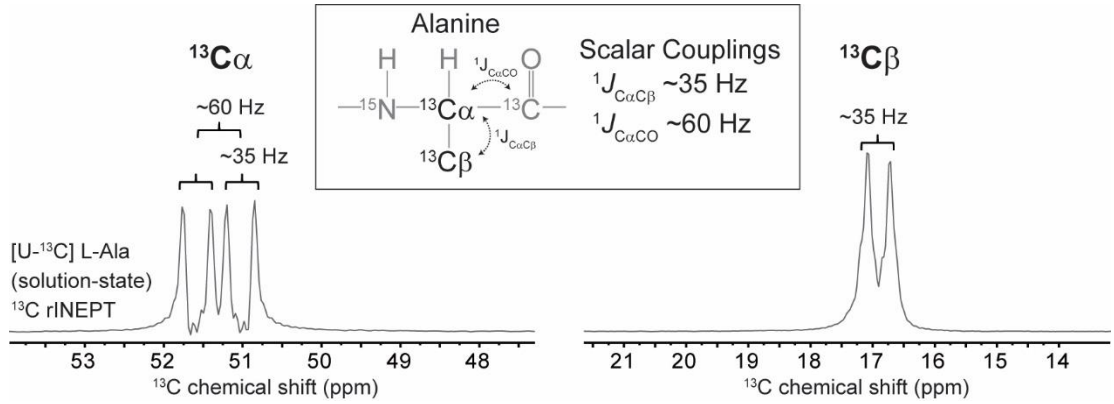


Figure A.4 ^{13}Cα- and ^{13}Cβ-Ala lineshapes of [U-^{13}C]L-Ala powder in water. ^{13}C spectra were obtained using rINEPT at room temperature, 22° C.

APPENDIX 5. RESULT OF GOR IV PREDICTION FROM RAT TROPOELASTIN'S PRIMARY STRUCTURE

```
Amino Acid Residues
Secondary Structures
MAGLTAAVPQPGVLLILLLNLLHPAQPGGVPGAVPGGVPGGLPGGVPGGVYYPGAGIGGGLGGGALGPGG
ccccccccccchhhhhhhhccccccccccceeccccccccccccceeeeccccccccccccccccc
KPPKPGAGLLGAFGAGPGGLGGAGPGAGGVLVPGGGAGAAAAYKAAAKAGAGLGGIGGVPGGVGVGGVPG
ccccccceeeeeccccccccccccccceeeecccchhhhhhhhhhhhcccccccccccceeecccccc
AVGVGGVPGAVGGIGGIGGLGVSTGAVVPQLGAGVGAGGKPGKVPGVGLPGVYPGGVLPGTGARFPGVGV
eeeecccccceeceeceeccceeccccceeeeeccccccccccccceeecccccccccccccccccccee
LPGVPTGTGVKAKVPGGGGGAFSGIPGVGPFGGQQPGVPLGYPIKAPKLPGGYGLPYTNGKLPYGVAGAG
ecccccceeeeeeccccccceecccccccccccccccccccccccccccccceeecccccceeeeccc
GKAGYPTGTGVGSQAAVAAAKAAKYGAGGGGVLPGVGGGGIPGGAGAIPGIGGITGAGTPAAAAAAKAAA
ccccccccccchhhhhhhhhhhhcccccceeeecccccccccccccceeeecccccchhhhhhhhhhh
KAAKYGAAGGLVPGGPGVRVPGAGIPGVGIPGVGGIPGVGGIPGVGGIPGVGGPGIGGPGIVGGPGAVSP
hhhhhhccccceecccccceecccccccceecccccccccccccccccccccccccccccceecccccch
AAAAKAAAKAAKYGARGGVGIPTYGVGAGGFPGYGVGAGAGLGGASQAAAAAAAAAKAAKYGAGGAGTLGG
hhhhhhhhhhhhhccccceeeeeeeccccccccceeecccccchhhhhhhhhhhhhhhhhcccccccccc
LVPGAVPGALPGAVPGALPGAVPGALPGAVPGVPGTGGVPGTPAAAAAAAAAKAAAKAGQYGLGPGVGGV
ccccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhcccccccccccc
PGGVGVGGLPGGVGPGGVTGIGTGPGTGLVPGDLGGAGTPAAAKSAAKAAAKAQYRAAAGLGAGVPGLGV
cceeeecccccccccccceeeeeccccceeeecccccccchhhhhhhhhhhhhhhhhhhhhcccccccceee
GAGVPGFGAGAGGFGAGAGVPGFGAGAVPGSLAASKAAKYGAAGGLGGPGGLGGPGGLGGPGGLGGPGGL
ecccccccccccccccccccccccccccchhhhhhhhhhhcccccccccccccccccccccccccccc
GGPGGLGGVPGGVAGAPAAAAAAKAAAKAAQYGLGGAGGLGAGGLGAGGLGAGGLGAGGLGAGGLGAGGV
ccccccccccccccchhhhhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccce
IPGAVGLGGVSPAAAAKAAKYGAAGLGGVLGARPFPGGGVAARPGFGLSPIYPGGGAGGLGVGGKPPKPY
eeceeccccchhhhhhhhhhhhhcccccccccccccccceecccccccceeecccccccceecccccccc
GGALGALGYQGGGCFGKSCGRKRK
cceeeeeecccccccccccceec
Sequence length :   864 (AA Sequence from RefDB [1]
```

GOR 4

Alpha helix	(Hh) :	159	is	18.40%
3_{10} helix	(Gg) :	0	is	0.00%
Pi helix	(Ii) :	0	is	0.00%
Beta bridge	(Bb) :	0	is	0.00%
Extended strand	(Ee) :	119	is	13.77%
Beta turn	(Tt) :	0	is	0.00%
Bend region	(Ss) :	0	is	0.00%
Random coil	(Cc) :	586	is	67.82%
Ambiguous states	(?) :	0	is	0.00%
Other states	:	0	is	0.00%

Ala in Hh : 95 out of 168 Total Ala (57%)

[1] Zhang, H., S. Neal, and D. Wishart, *RefDB: A database of uniformly referenced protein chemical shifts.* Journal of Biomolecular NMR, 2003. 25(3): p. 173-195.

APPENDIX 6. CORRECTED ^{13}CO CHEMICAL SHIFTS OF ALA IN THE LYS-ALA-TYR SEQUENCE

The procedure to calculate the corrected chemical shift for **Ala**, $\delta_{\mathbf{Ala}}$, in the three-amino-acid sequence, Lys-**Ala**-Tyr, is shown in **Figure A.3.3**. In general, the correction factor for a given residue **X** is obtained from the difference of observed Gly (*G*) chemical shifts in the Ac-G*G***X***G*G-NH_2 peptide, relative to those in the reference peptide Ac-G*G***G***G*G-NH_2. Hence, the correction factor for Lys, $\Delta\delta$(**K**), is obtained from the difference of observed ^{13}CO-*Gly* chemical shift in Ac-GG**K***G*G-NH_2, relative to that in the Ac-GG**G***G*G-NH_2 peptide (inset of **step 1**). Analogously, the correction factor for **Tyr**, $\Delta\delta$(**Y**), is computed based on the comparison of ^{13}C-*Gly* resonance in Ac-G*G***Y**GG-NH_2 with respect to the Ac-G*G***G**GG-NH_2 (inset of **step 2**). The chemical shift of the central **Ala**, δ_{rc}(**Ala**)=178.5 ppm, is obtained from the Ac-GG**A**GG-NH_2 peptide in 8 M urea [1] (**step 3**). Thus, the *neighboring residue* effects [2] in the ^{13}CO-**Ala** chemical shift, $\delta_{\mathbf{Ala}}$(corrected), is calculated using eqn. 3.2.

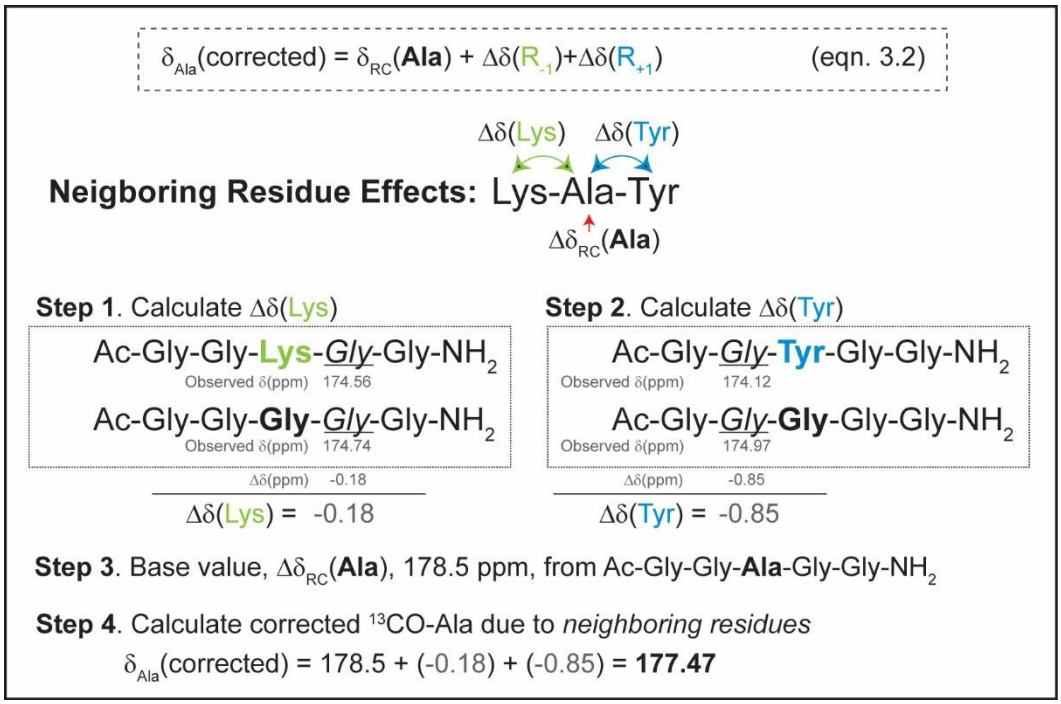


Figure A.5 Procedure for the calculation of *neighboring residue* effects on the ^{13}CO chemical shift of Ala in the Lys-Ala-Tyr sequence. The base value for residue **Ala** in Ac-GG**A**GG-NH_2, $\Delta\delta_{RC}$(**Ala**) is obtained from literature [1].

[1] Schwarzinger, S., G.J. Kroon, T.R. Foss, P.E. Wright, and H.J. Dyson, *Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView.* J Biomol NMR, 2000. **18**(1): p. 43-8.

[2] Schwarzinger, S., G.J.A. Kroon, T.R. Foss, J. Chung, P.E. Wright, and H.J. Dyson, Sequence-dependent correction of random coil NMR chemical shifts. Journal of the American Chemical Society, 2001. 123(13): p. 2970-2978.

APPENDIX 7. ^{13}CO-^{13}Cα CORRELATIONS OF RANDOM COIL ALA

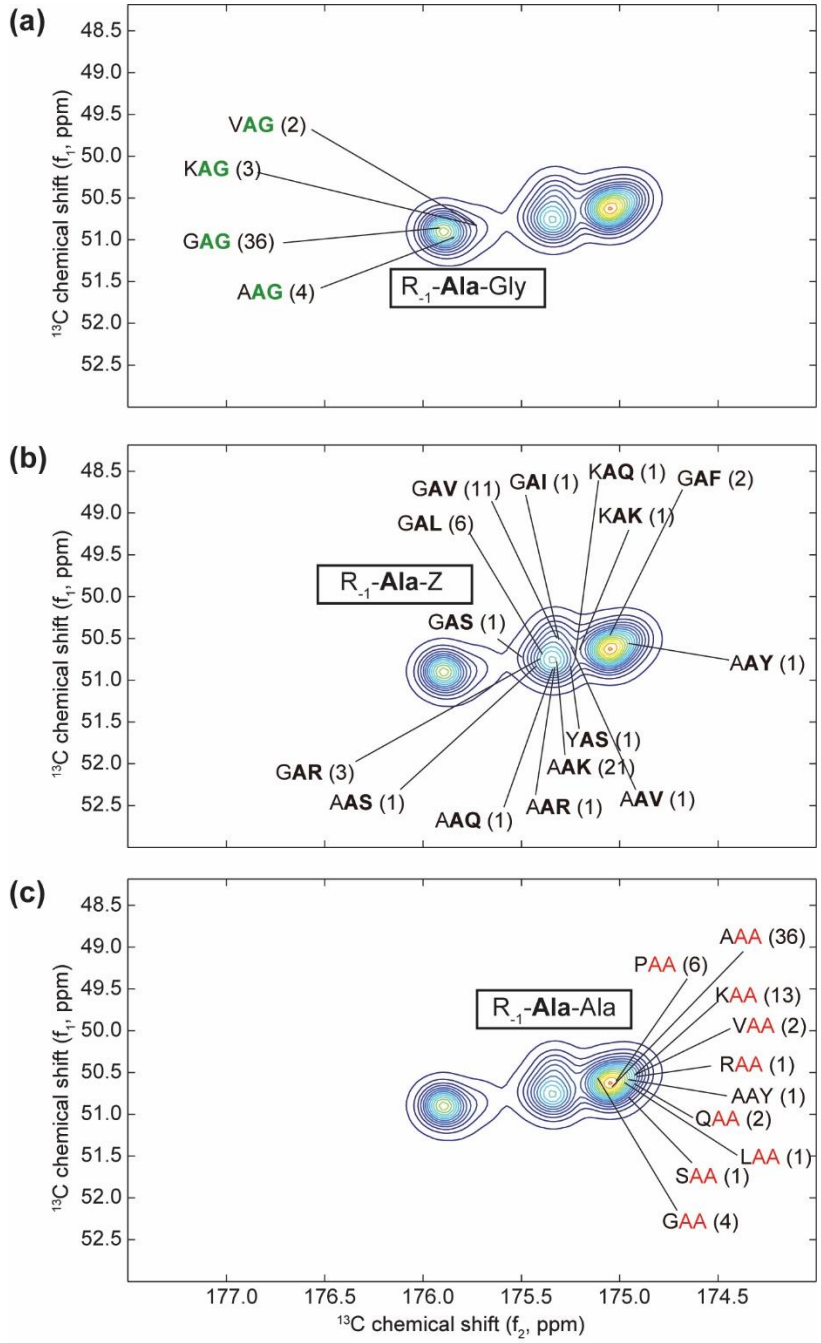


Figure A.6 Predicted ^{13}CO-^{13}Cα correlations of random coil Ala populations for (a) R$_{-1}$-Ala-Gly, (b) R$_{-1}$-Ala-Z (Z≠Ala, Gly), and (c) R$_{-1}$-Ala-Ala motifs. Each 2D spectrum reflects the expected ^{13}CO-^{13}Cα correlations for all R$_{-1}$-Ala-R$_{+1}$ motifs in rat tropoelastin (Table 3.1, Chapter 3.A.1). The primary structure is obtained from ElastoDB [1].

[1] He, D., et al., *Comparative genomics of elastin: Sequence analysis of a highly repetitive protein.* Matrix Biology, 2007. 26(7): p. 524-40.

APPENDIX 8. COMPARISON OF ^{13}CO-ALA LINESHAPES IN HYDRATED AND DENATURED ELASTIN

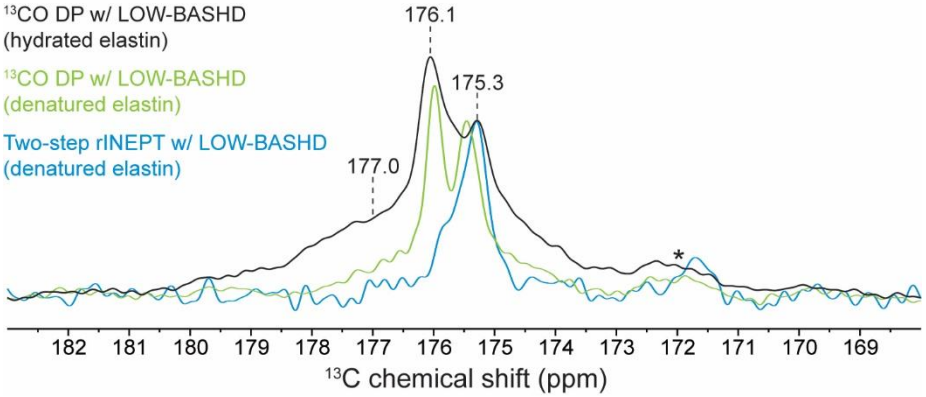


Figure A.7 Superposition of ^{13}CO-Ala lineshapes for hydrated and denatured elastin at 20° C. Black and **green** represent *J*-decoupled ^{13}CO-Ala DP lineshapes of hydrated and denatured elastin, respectively. **Blue** represents *J*-decoupled ^{13}CO-Ala DP lineshape of denatured elastin. Each lineshape is scaled to the height of the 175.3 ppm peak. Vertical dashed lines indicate features in the ^{13}CO-Ala DP spectrum of hydrated elastin, α-helical (177 ppm) and random coil populations (176.1 and 175.3 ppm). Asterisk indicates natural-abundance ^{13}CO-Gly signal.

APPENDIX 9. COMPARISON OF ^{13}C-ALA LINESHAPES ACQUIRED BY CP, RINEPT, AND DP EXPERIMENTS

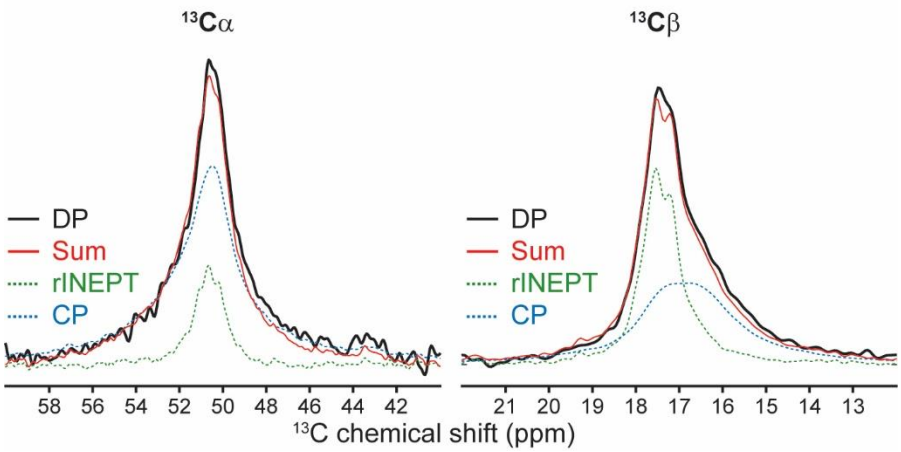


Figure A.8 Superposition of ^{13}C-Ala lineshapes obtained by DP and those obtained by selective detection experiments. ***Black*** represents DP lineshape. The sum of CP (***blue***) and rINEPT (***green***) lineshapes are shown in ***red***. The ^{13}C DP, rINEPT and CP spectra were acquired with 256, 1024 and 6144 transients, respectively. Each signal was processed with 10 Hz of line broadening, and the height of each peak is scaled to the maximum intensity of the DP spectrum.

APPENDIX 10. ALDL AND DLNL PARAMETERS FOR MD SIMULATIONS

A patch entry for ALDL crosslink (in the topology file)

```
PRES ALDL       0.00   ! patch for Allysine Aldol between 2 Lys
                       !
                       !  1HB1  1HG1       2HE1   2HD1   2HG1
                       !      |     |            |      |     |
                       !  --1CB--1CG--1CD==2CE--2CD--2CG--
                       !      |     |     |            |      |
                       !  1HB2 1HG2  1CE==1OE  2HD2 2HG2
                       !                  |
                       !                1HE1
                       !
GROUP
ATOM 1CD  CG2DC1  0.12
ATOM 1CE  CD         0.16
ATOM 1OE  O          -0.33
ATOM 1HE1 HR1        0.05
GROUP
ATOM 2CE  CG2DC1  -0.15
ATOM 2HE1 HGA4       0.15
GROUP
ATOM 1CG  CT2    -0.18
ATOM 1HG1 HA      0.09
ATOM 1HG2 HA      0.09
GROUP
ATOM 2CD  CT2    -0.18
ATOM 2HD1 HA      0.09
ATOM 2HD2 HA      0.09
DELETE ATOM 1HD1
DELETE ATOM 1HD2
DELETE ATOM 1HE2
DELETE ATOM 2HE2
DELETE ATOM 1NZ
DELETE ATOM 1HZ1
DELETE ATOM 1HZ2
DELETE ATOM 1HZ3
DELETE ATOM 2NZ
DELETE ATOM 2HZ1
DELETE ATOM 2HZ2
DELETE ATOM 2HZ3
BOND 1CD 2CE
BOND 1CE 1OE
BOND 1HE1 HR1
```

A patch entry for DLNL crosslink (in the topology file)

```
PRES DLNL       0.00   ! patch for dehydro-Lysinonorleucine crosslink
                       ! between 2 Lys
                     !
                     !   1HD1  1HE1  2HE1  2HD1
                     !       |         |       |    |
                     !   --1CD---1CE---1NZ==2CE---2CD--
                         !       |         |    ..       |
                         !    1HD2   1HE2          2HD2
                     !
                     !
                     !
GROUP
ATOM 1CE   CT2    0.04
ATOM 1HE1  HA     0.09
ATOM 1HE2  HA     0.09
GROUP
ATOM 1NZ   NS1     -0.60
ATOM 2CE   CC1B   0.23
ATOM 2HE1  HE1    0.15
DELETE ATOM 2NZ
DELETE ATOM 2HZ1
DELETE ATOM 2HZ2
DELETE ATOM 2HZ3
DELETE ATOM 2HE2
DELETE ATOM 1HZ1
DELETE ATOM 1HZ2
DELETE ATOM 1HZ3
BOND 1NZ 2CE
```

Parameters for ALDL and DLNL crosslinks (in the parameter file)

```
CG2DC1 CD O     140.000   125.0000 ! TAKEN From propionaldehyde
CD CG2DC1 CG2DC1  60.00    120.00 ! TAKEN from RETINOL RTAL unmodified
CC1B  CT2   383.000     1.504 ! from CC1B CT3
NS1  CC1B CT2    52.00     123.00 !sum=242.5
CC1B CT2  HA      45.00     111.50     ! 1-butene; from propene, yin,adm jr., 12/95
CT2  CT2  CC1B    32.00    112.20  20   ! 1-butene; from propene, yin,adm jr., 12/95
CT2  CC1B HE1     40.00    116.00    00   ! 1-butene; from propene, yin,adm jr., 12/95
CT2  NS1  CC1B    67.00    111.00
CT2  NS1  CC1B CT2    12.0000  2   180.00
HE1  CC1B CT2  CT2     0.1200  3     0.00    ! butene, yin,adm jr., 12/95
HA  CT2   CC1B HE1      0.0000  3     0.00  ! butene, adm jr., 2/00 update
NS1 CC1B CT2 CT2       0.5000  1   180.00  00 ! 1-butene, adm jr., 2/00 update
NS1 CC1B CT2 CT2      1.3000  3   180.00
NS1 CC1B CT2  HA       1.0000  2   180.00
```

Notes:

1. Entries for topology and parameter files were obtained from a variety of parameters from small molecules stored on CHARMM GUI library archive <http://www.charmm-gui.org/?doc=archive>
2. Some parameters were also gathered from charmm general force field v.2b7 for small molecule drug design <http://mackerell.umaryland.edu/~kenno/cgenff>

APPENDIX 11. EXAMPLE MD CONFIGURATION FILE AND ANALYSIS SCRIPTS

MDrun.conf (an example of restart Job)

```
#####################################################################
## JOB DESCRIPTION:        restart Job                              ##
#####################################################################
structure          ./E25ALDLwbion-rev.psf
coordinates        ./E25ALDLwbion.pdb
set outputname     equil-4
set temperature    300
firsttimestep      20102000
bincoordinates     ./equil-3.restart.coor
binvelocities      ./equil-3.restart.vel
extendedSystem     ./equil-3.restart.xsc
#####################################################################
## SIMULATION PARAMETERS                                           ##
#####################################################################
# Input
paraTypeCharmm             on
parameters          /Users/Jhonsen/Documents/ALLYSINEALDOL/par_all22_DES_ALDL-rev.inp
#temperature         $temperature   :#Commented for restarting jobs
# Force-Field Parameters
exclude             scaled1-4
1-4scaling          1.0
cutoff              10.0
switching           on
switchdist          8.0
pairlistdist        12.5
# Integrator Parameters
timestep            2.0  ;# 2fs/step
rigidBonds          all  ;# needed for 2fs steps
nonbondedFreq       1
fullElectFrequency  2
stepspercycle       10
# Constant Temperature Control
langevin            on    ;# do langevin dynamics
langevinDamping     5     ;# damping coefficient (gamma) of 1/ps
langevinTemp        $temperature
#langevinCol         O
langevinHydrogen    no    ;# don't couple langevin bath to hydrogens
# Periodic Boundary Conditions
#cellBasisVector1     34.5   0.0   0.0 :#Commented for restarting jobs
#cellBasisVector2     0.0    44.9  0.0 :#Commented for restarting jobs
#cellBasisVector3     0.0    0.0  41.7  :#Commented for restarting jobs
#cellOrigin         -14.5  6.3  20.9        :#Commented for restarting jobs
margin 5
wrapAll             on
# PME (for full-system periodic electrostatics)
PME                 yes
```

```
PMEGridSizeX        40
PMEGridSizeY        48
PMEGridSizeZ        45
# Constant Pressure Control (variable volume)
useGroupPressure      yes ;# needed for rigidBonds
useFlexibleCell       no
useConstantArea       no
langevinPiston        on
langevinPistonTarget  1.01325 ;#  in bar -> 1 atm
langevinPistonPeriod  200.0
langevinPistonDecay   100.0
langevinPistonTemp    $temperature
# Output
outputName          $outputname
binaryoutput        no
restartfreq         10000     ;# 2 fs x 10,000steps = every 20ps
dcdfreq             1000
xstFreq             1000
outputEnergies      100
outputPressure      100
outputTiming        1000
#############################################################
## EXECUTION SCRIPT                                        ##
#############################################################
#minimize   2000     :#Commented for restarting jobs
reinitvels    $temperature
run 10000000 ;# 20 ns with 2fs/step
```

STRUCTcount.tcl

```
set out [open ResiStruct-ALDL7.txt w]
set sel [atomselect top alpha]
set helcounts [list]
set turncounts [list]
set extcounts [list]
set bridgecounts [list]
set coilcounts [list]
set picounts [list]
set 310counts [list]
foreach ind [$sel get index] {
   lappend helcounts 0
   lappend turncounts 0
   lappend extcounts 0
   lappend bridgecounts 0
   lappend coilcounts 0
   lappend picounts 0
   lappend 310counts 0
}
# set i with the frame you wanna start analyze
for {set i 2000} {$i < 4003} {incr i} {
   animate goto $i
   display update ui
```

```
    mol reanalyze top
    set structs [$sel get structure]
puts -nonewline $out frame:$i
puts  $out content:$structs
for {set j 0} {$j < [llength $structs]} {incr j} {
        if {[lindex $structs $j] == "H"} then {
        set helcounts [lreplace $helcounts $j $j [expr [lindex $helcounts $j] + 1] ]}
        if {[lindex $structs $j] == "T"} then {
        set turncounts [lreplace $turncounts $j $j [expr [lindex $turncounts $j] + 1] ]}
        if {[lindex $structs $j] == "E"} then {
        set extcounts [lreplace $extcounts $j $j [expr [lindex $extcounts $j] + 1] ]}
        if {[lindex $structs $j] == "B"} then {
            set bridgecounts [lreplace $bridgecounts $j $j [expr [lindex $bridgecounts $j] + 1] ]}
        if {[lindex $structs $j] == "C"} then {
        set coilcounts [lreplace $coilcounts $j $j [expr [lindex $coilcounts $j] + 1] ]}
        if {[lindex $structs $j] == "G"} then {
        set 310counts [lreplace $310counts $j $j [expr [lindex $310counts $j] + 1] ]}
        if {[lindex $structs $j] == "I"} then {
        set picounts [lreplace $picounts $j $j [expr [lindex $picounts $j] + 1] ]}
}
}
puts $out frame:ALLhelix:$helcounts
puts $out frame:ALLTurn:$turncounts
puts $out frame:ALLExt:$extcounts
puts $out frame:ALLBridge:$bridgecounts
puts $out frame:ALL310:$310counts
puts $out frame:ALLPi:$picounts
puts $out frame:ALLCoil:$coilcounts
#puts $out frame:ALLhelix:$helcounts
close $out
```

Percentahelix.tcl

```
set out [open pctHELIX-ALDL3-400ns.txt w]
set sel [atomselect top alpha]
set helcounts [list]
foreach ind [$sel get index] {
   lappend helcounts 0
}
for {set i 0} {$i < [molinfo top get numframes]} {incr i} {
            animate goto $i
            display update ui
            mol reanalyze top
            set structs [$sel get structure]
            puts -nonewline $out frame:$i,
            puts -nonewline $out content:$structs,
            #puts -nonewline index
            set vecG [lsearch -all $structs G]
            set vecH [lsearch -all $structs H]
            set vecT [lsearch -all $structs T]
```

```
        set vecE [lsearch -all $structs E]
        set vecB [lsearch -all $structs B]
        set vecI [lsearch -all $structs I]
        set vecC [lsearch -all $structs C]
        #puts $vex
        # Identified 2ndary structure divided by the number of residues for E25
        set pctG [expr [llength $vecG] * 100 / 18]
        set pctH [expr [llength $vecH] * 100 / 18]
        set pctT [expr [llength $vecT] * 100 / 18]
        set pctE [expr [llength $vecE] * 100 / 18]
        set pctB [expr [llength $vecB] * 100 / 18]
        set pctI [expr [llength $vecI] * 100 / 18]
        set pctC [expr [llength $vecC] * 100 / 18]
        puts $out pct-H,G,T,E,B,I,C:$pctH,$pctG,$pctT,$pctE,$pctB,$pctI,$pctC
        }
close $out
```

Countwatnum.tcl

```
set mol [molinfo top]
for {set r 1} {$r < 19} {incr r} {
        set shel [atomselect $mol "same residue as (water within 3.6 of backbone and resid $r)"]
        set nf [molinfo $mol get numframes]
        set fil [open water-bbres$r.txt w]
                for {set i 0} {$i<$nf} {incr i} {
                $shel frame $i
                $shel update
                puts $fil "res:$r frame:$i num-wat:[$shel num]"
                }
close $fil
                }
```

CountOwater.tcl

```
set mol [molinfo top]
for {set r 1} {$r < 19} {incr r} {
        set shel [atomselect $mol "name OH2 and water within 3.6 of (backbone and name O and resid
$r )" ]
        set nf [molinfo $mol get numframes]
        set fil [open numwat-OxyOxy-$r.txt w]
                for {set i 0} {$i<$nf} {incr i} {
                $shel frame $i
                $shel update
                puts $fil "res:$r frame:$i num-wat:[$shel num]"
                }
close $fil
                }
```

Counthb.tcl

```
set mol [molinfo top]
set prot [atomselect $mol "protein and nitrogen"]
set prot2 [atomselect $mol "protein and oxygen"]
set nf [molinfo $mol get numframes]
set fil [open allhydrogenbond.txt w]
for {set i 0} {$i<$nf} {incr i} {
$prot frame $i
$prot2 frame $i
set nhb [llength [lindex [measure hbonds 3.5 45 $prot $prot2] 0]]
puts $fil "$i: $nhb"
}
close $fil
```

RMSF.tcl

```
#Prior to running this calculation align the molecule first
# ERROR MESSAGE (if it comes up) due to " changed to Ò by textedit
set outfile [open Rmsf-DLNL574K-alignALL_ref.txt w]
set sel [atomselect top "protein and name CA"]
set rms [measure rmsf $sel first 0 last 2499 step 1]
for {set i 0} {$i < [$sel num]} {incr i} {
puts $outfile "[expr {$i + 1}] [lindex $rms $i]"
}
close $outfile
```

HBdefineindex.rtf (HBond definition input for VMD's Timeline plugin)

(index 52 15 53) or (index 63 25 64) or (index 73 35 74) or (index 88 51 89) or (index 100 62 101) or (index 116 72 117) or (index 126 82 127) or (index 136 99 137) or (index 158 115 159) or (index 168 125 169) or (index 193 135 194)

PickHB.csh (sorting HBonds by each interaction)

```
#!/bin/csh
grep -A 2004 'freeSelLabel 52 15 53' TL-BBHB-ALDL5 > BB-HB-1.txt
grep -A 2004 'freeSelLabel 63 25 64' TL-BBHB-ALDL5> BB-HB-2.txt
grep -A 2004 'freeSelLabel 73 35 74' TL-BBHB-ALDL5> BB-HB-3.txt
grep -A 2004 'freeSelLabel 88 51 89' TL-BBHB-ALDL5> BB-HB-4.txt
grep -A 2004 'freeSelLabel 100 62 101' TL-BBHB-ALDL5> BB-HB-5.txt
grep -A 2004 'freeSelLabel 116 72 117' TL-BBHB-ALDL5> BB-HB-6.txt
grep -A 2004 'freeSelLabel 126 82 127' TL-BBHB-ALDL5> BB-HB-7.txt
grep -A 2004 'freeSelLabel 136 99 137' TL-BBHB-ALDL5> BB-HB-8.txt
grep -A 2004 'freeSelLabel 158 115 159' TL-BBHB-ALDL5> BB-HB-9.txt
grep -A 2004 'freeSelLabel 168 125 169' TL-BBHB-ALDL5> BB-HB-10.txt
grep -A 2004 'freeSelLabel 193 135 194' TL-BBHB-ALDL5> BB-HB-11.txt
```

Phipsitotxt.csh (splitting torsion-angles by residue)

```
#!/bin/csh
grep '1 ALA' phi-psi.txt >1ALA_phipsi.txt
grep '2 ALA' phi-psi.txt >2ALA_phipsi.txt
grep '3 ALA' phi-psi.txt >3ALA_phipsi.txt
```

```
grep '4 LYS' phi-psi.txt >4LYS_phipsi.txt
grep '5 SER' phi-psi.txt >5SER_phipsi.txt
grep '6 ALA' phi-psi.txt >6ALA_phipsi.txt
grep '7 ALA' phi-psi.txt >7ALA_phipsi.txt
grep '8 LYS' phi-psi.txt >8LYS_phipsi.txt
grep '9 VAL' phi-psi.txt >9VAL_phipsi.txt
grep '10 ALA' phi-psi.txt >10ALA_phipsi.txt
grep '11 ALA' phi-psi.txt >11ALA_phipsi.txt
grep '12 LYS' phi-psi.txt >12LYS_phipsi.txt
grep '13 ALA' phi-psi.txt >13ALA_phipsi.txt
grep '14 GLN' phi-psi.txt >14GLN_phipsi.txt
grep '15 LEU' phi-psi.txt >15LEU_phipsi.txt
```

StripTermCatPdbGmx.csh (Transitioning Namd to Gromacs)

```
grep -v ' TIP' ../../E25-ALDL-3/E25ALDLwbion.pdb >tmp
grep -v ' ION' tmp > tmp2
grep -v ' CY' tmp2 > tmp
grep -v ' CAY' tmp > tmp2
grep -v ' CY' tmp2 > tmp
grep -v ' OY' tmp > tmp2
grep -v ' OE' tmp2 > tmp
grep -v ' NT' tmp > tmp2
grep -v ' CAT' tmp2 > tmp
grep -v ' OT1' tmp > tmp2
grep -v ' OT2' tmp2 > tmp
mv tmp first.pdb
awk '/ATOM/{print$2-1}' first.pdb > first.idx
catdcd -o alldcdCombined.trr -otype trr -i first.idx -stride 50 -dcd ../../E25-ALDL-3/equil-2.dcd ../../E25-
ALDL-3/equil-3.dcd ../../E25-ALDL-3/equil-4.dcd ../../E25-ALDL-3/equil-5.dcd ../../E25-ALDL-3/equil-6.dcd
../../E25-ALDL-3/equil-7.dcd ../../E25-ALDL-3/equil-8.dcd ../../E25-ALDL-3/equil-9.dcd ../../E25-ALDL-
3/equil-10.dcd ../../E25-ALDL-3/equil-11.dcd ../../E25-ALDL-4/equil-2.dcd ../../E25-ALDL-4/equil-3.dcd
../../E25-ALDL-4/equil-4.dcd ../../E25-ALDL-4/equil-5.dcd ../../E25-ALDL-4/equil-6.dcd ../../E25-ALDL-
4/equil-7.dcd ../../E25-ALDL-4/equil-8.dcd ../../E25-ALDL-4/equil-9.dcd ../../E25-ALDL-4/equil-10.dcd
../../E25-ALDL-4/equil-11.dcd ../../E25-ALDL-5/equil-2.dcd ../../E25-ALDL-5/equil-3.dcd ../../E25-ALDL-
5/equil-4.dcd ../../E25-ALDL-5/equil-5.dcd ../../E25-ALDL-5/equil-6.dcd ../../E25-ALDL-5/equil-7.dcd
../../E25-ALDL-5/equil-8.dcd ../../E25-ALDL-5/equil-9.dcd ../../E25-ALDL-5/equil-10.dcd ../../E25-ALDL-
5/equil-11.dcd ../../E25-ALDL-6/equil-2.dcd ../../E25-ALDL-6/equil-3.dcd ../../E25-ALDL-6/equil-4.dcd
../../E25-ALDL-6/equil-5.dcd ../../E25-ALDL-6/equil-6.dcd ../../E25-ALDL-6/equil-7.dcd ../../E25-ALDL-
6/equil-8.dcd ../../E25-ALDL-6/equil-9.dcd ../../E25-ALDL-6/equil-10.dcd ../../E25-ALDL-6/equil-11.dcd
../../E25-ALDL-7/equil-2.dcd ../../E25-ALDL-7/equil-3.dcd ../../E25-ALDL-7/equil-4.dcd ../../E25-ALDL-
7/equil-5.dcd ../../E25-ALDL-7/equil-6.dcd ../../E25-ALDL-7/equil-7.dcd ../../E25-ALDL-7/equil-8.dcd
../../E25-ALDL-7/equil-9.dcd ../../E25-ALDL-7/equil-10.dcd ../../E25-ALDL-7/equil-11.dcd
pdb2gmx -f first.pdb -o cluster.gro -ignh -missing
```

CreateFEL-1.csh (preparing free-energy landscape)

```
#!/bin/csh
cd Analysis-263K/FEL-RMSDvsRg
perl ~/Documents/TCLscripts/sham.pl -i1 ../../CaRMSD-resid4to17-allE19.txt  -i2 ./Rgyr.txt -data1 1 -data2
1 -o gshaminput3.xvg
g_sham -f gshaminput3.xvg -ls FEL-RMSDvsRgyr-263K.xpm
xpm2ps -f FEL-RMSDvsRgyr-263K.xpm -o FEL-RMSDvsRgyr-263K.eps
```

CreateFEL-2.csh (converting output into text format)

```
#!/bin/csh
cd Analysis-263K/FEL-RMSDvsBBHBond
python ~/Documents/TCLscripts/xpm2txt.py -f FEL-RMSDvsBBHBond-263K.xpm -o FEL-
RMSDvsBBHBond-263K.txt
```

APPENDIX 12. TIMELINE PLOTS FOR EX25-ALDL AND EX19-DLNL TRAJECTORIES

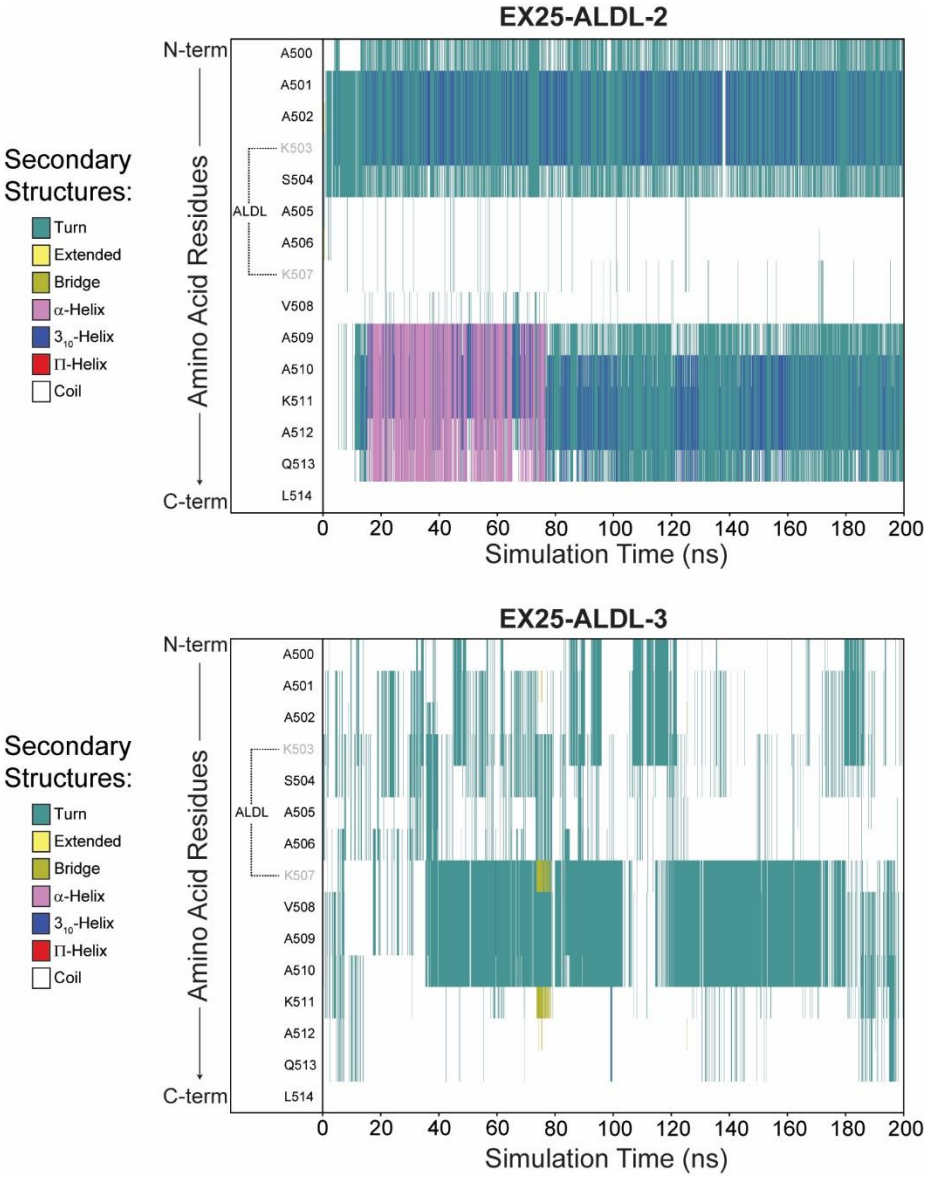


Figure A.9 Timeline plots of secondary structures for EX25-ALDL-2 and -3 trajectories at 300 K. Horizontal axes indicate the length of a single MD trajectory (200 ns). Colors indicate secondary structures based on STRIDE.

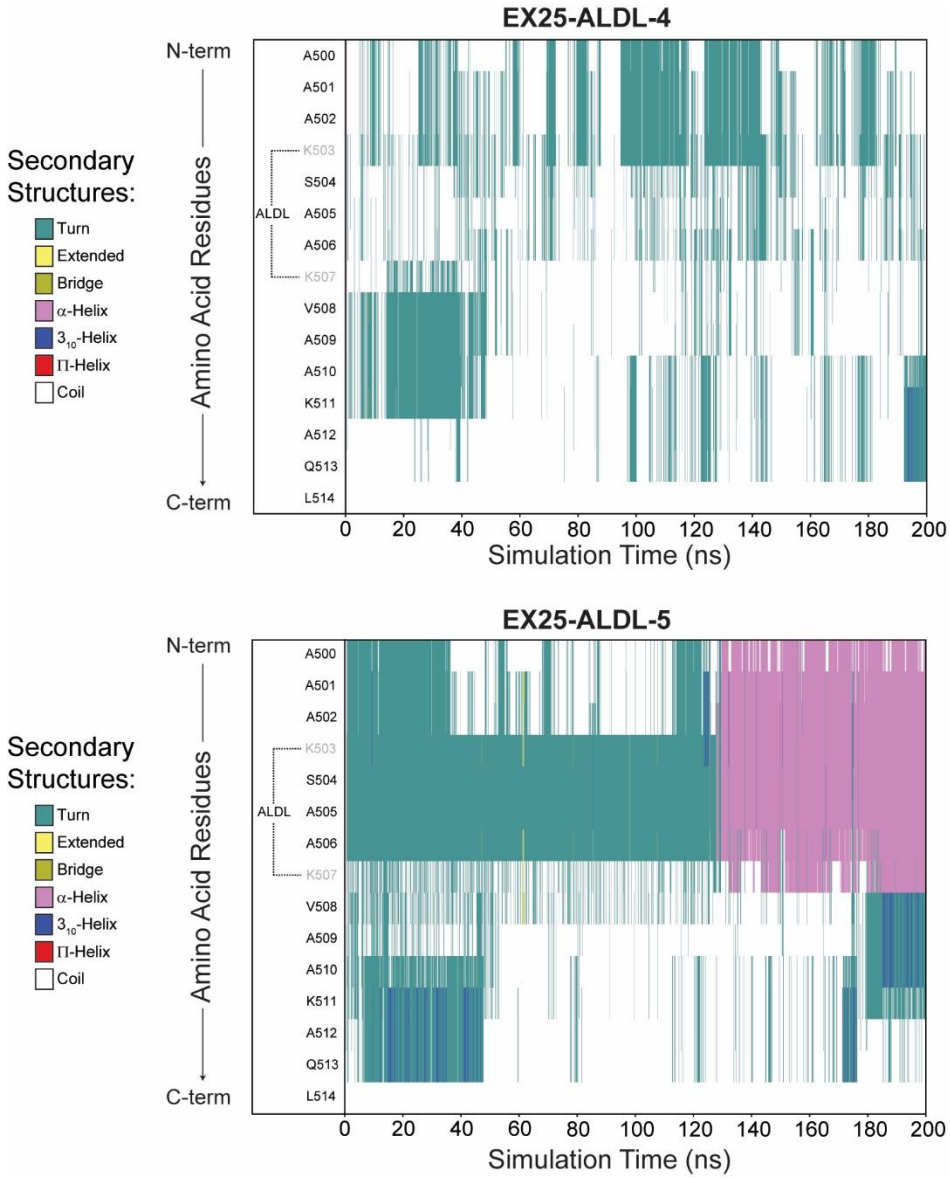


Figure A.10 Timeline plots of secondary structures for EX25-ALDL-4 and -5 trajectories at 300 K. Horizontal axes indicate the length of a single MD trajectory (200 ns). Colors indicate secondary structures based on STRIDE.

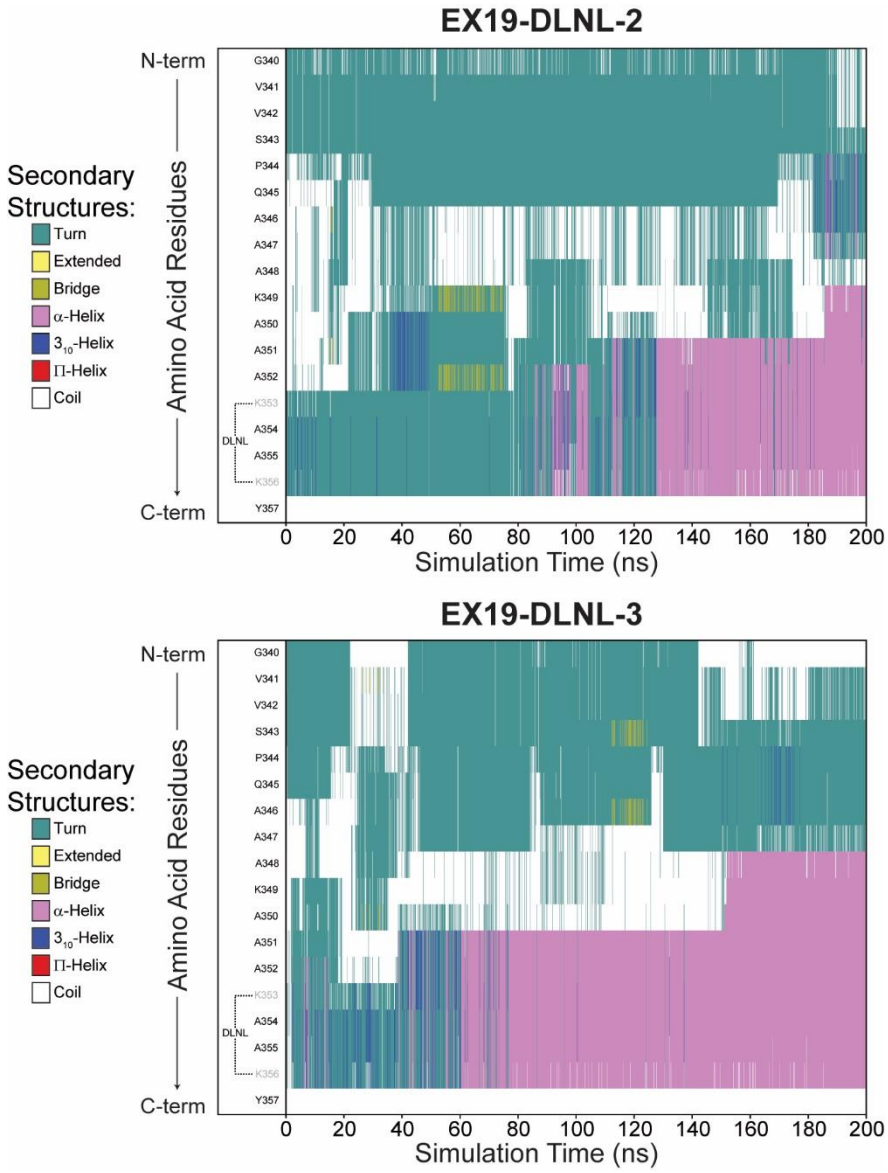


Figure A.11 Timeline plots of secondary structures for EX19-DLNL-2 and -3 trajectories at 300 K. Horizontal axes indicate the length of a single MD trajectory (200 ns). Colors indicate secondary structures based on STRIDE.

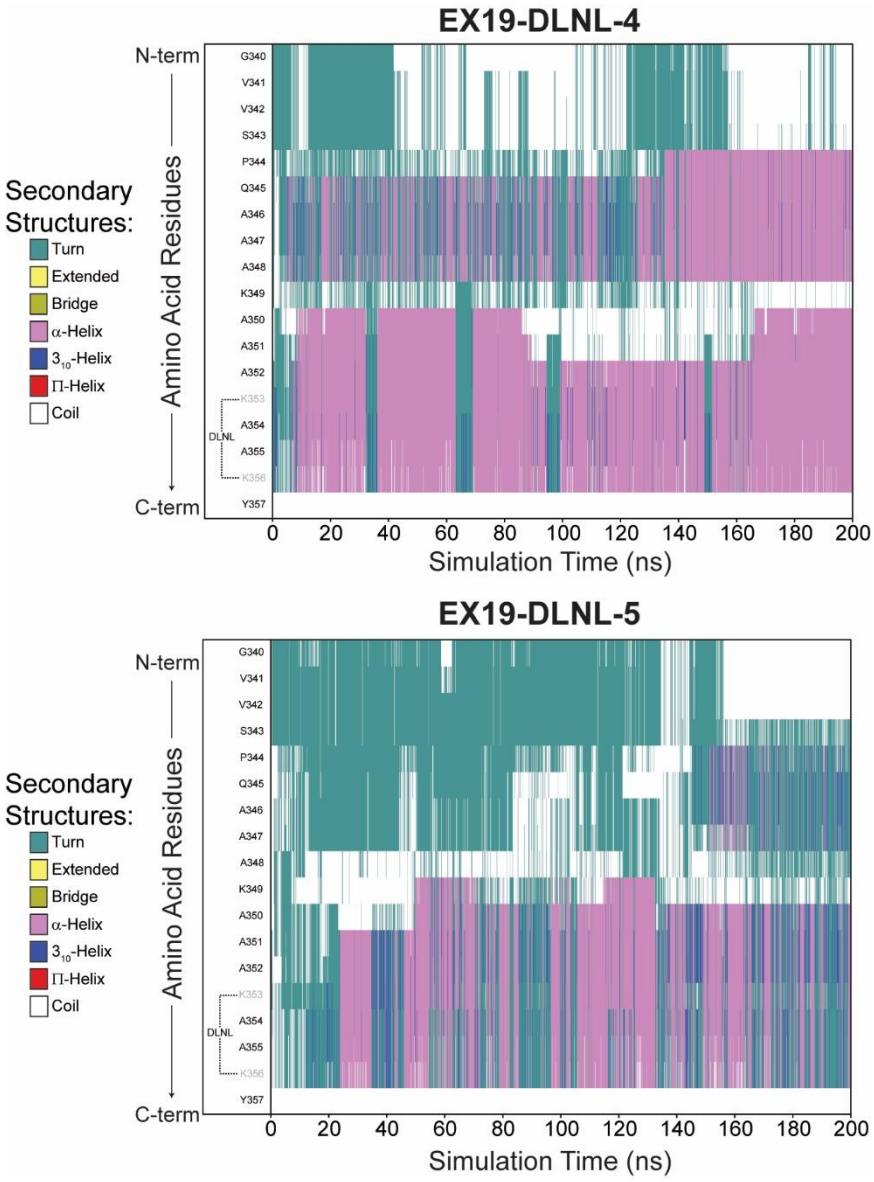


Figure A.12 Timeline plots of secondary structures for EX19-DLNL-4 and -5 trajectories at 300 K. Horizontal axes indicate the length of a single MD trajectory (200 ns). Colors indicate secondary structures based on STRIDE.

APPENDIX 13. Cα-RMSF PLOTS FOR EX25-ALDL AND EX19-DLNL TRAJECTORIES

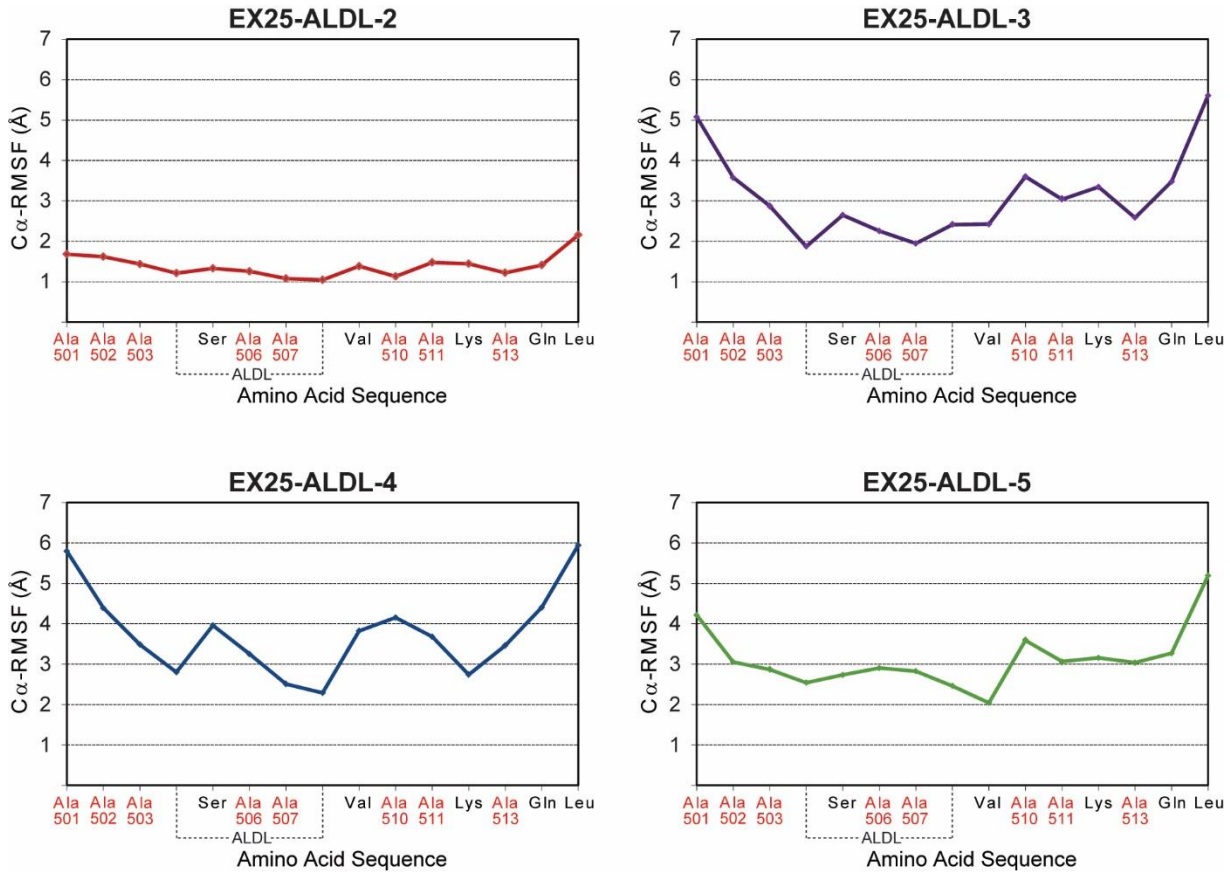


Figure A.13 Cα-RMSF plots for EX25-ALDL trajectories at 300 K. Each plot represents a single 200 ns trajectory; RMSF's are calculated with respect to average Cα coordinates within the last 100 ns, during which α-helices are more likely to be adopted. Amino acid sequences show the relative position of each alanine (A) in the chain.

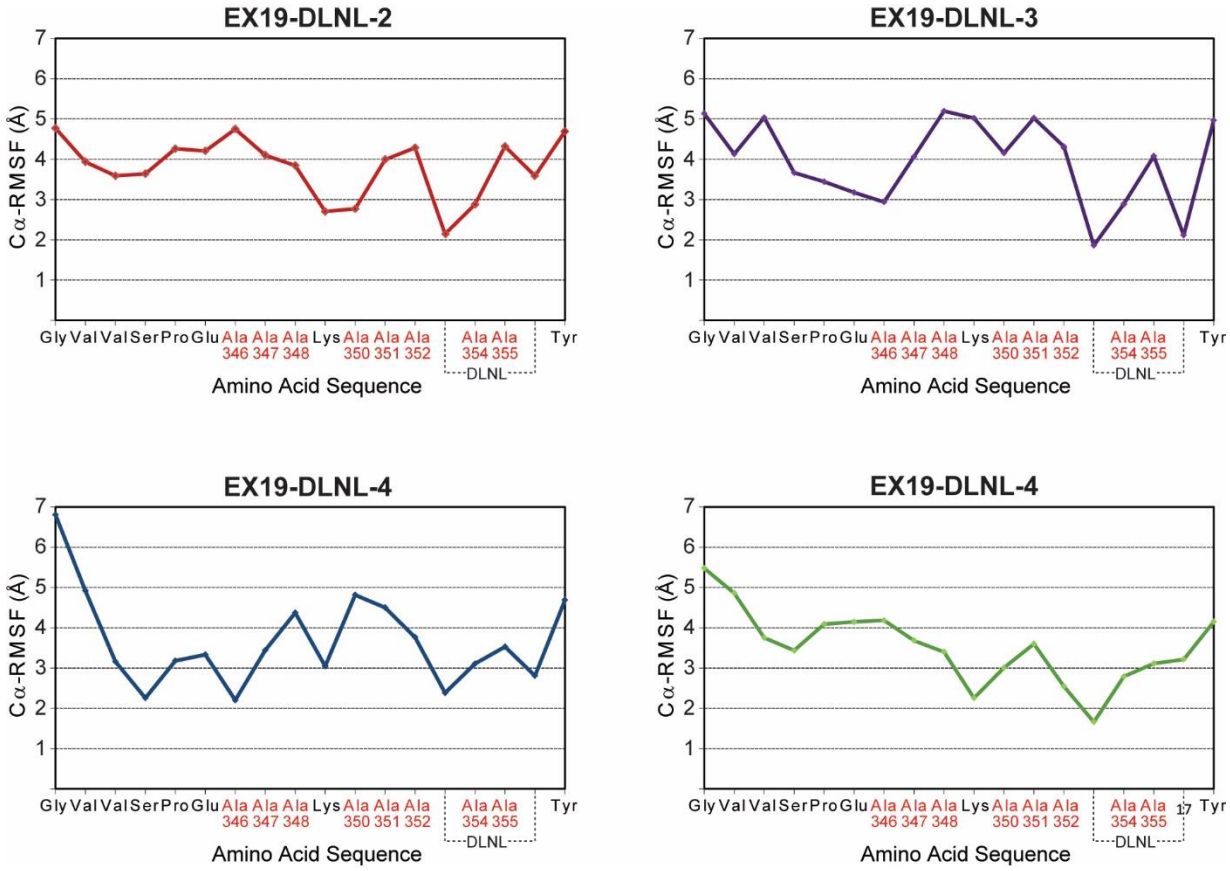


Figure A.14 Cα-RMSF plots for EX19-DLNL trajectories at 300 K. Each plot represents a single 200 ns trajectory; RMSF's are calculated with respect to average Cα coordinates within the last 100 ns, during which α-helices are more likely to be adopted. Amino acid sequences show the relative position of each alanine (A) in the chain.

APPENDIX 14. AVERAGE COORDINATION NUMBER OF WATER MOLECULES IN EX25-ALDL AND EX19-DLNL SYSTEMS

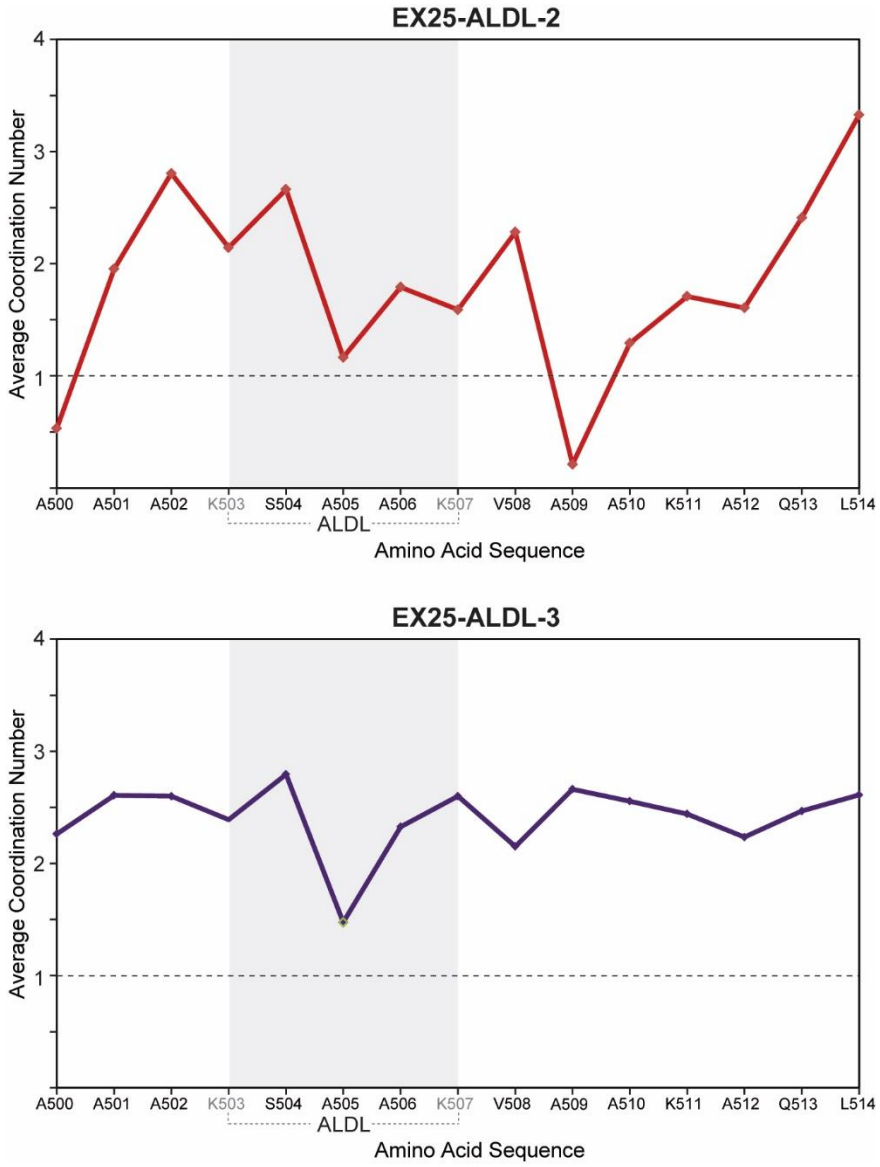


Figure A.15 Average coordination number of water for each amino acid in EX25-ALDL-2 and -3. The first hydration shell is defined as the region where water molecules are found within 3.6 Å of the peptide's backbone. The average coordination number of water in is calculated over the last 100 ns of the MD trajectory, during which a substantial α-helical segment is adopted. Gray shade indicates amino acids that are tethered by ALDL crosslink (i.e., modified K503 and K507). Horizontal dashed line indicates the average coordination number of 1, typical for carbonyl oxygens found in the α-helical structure.

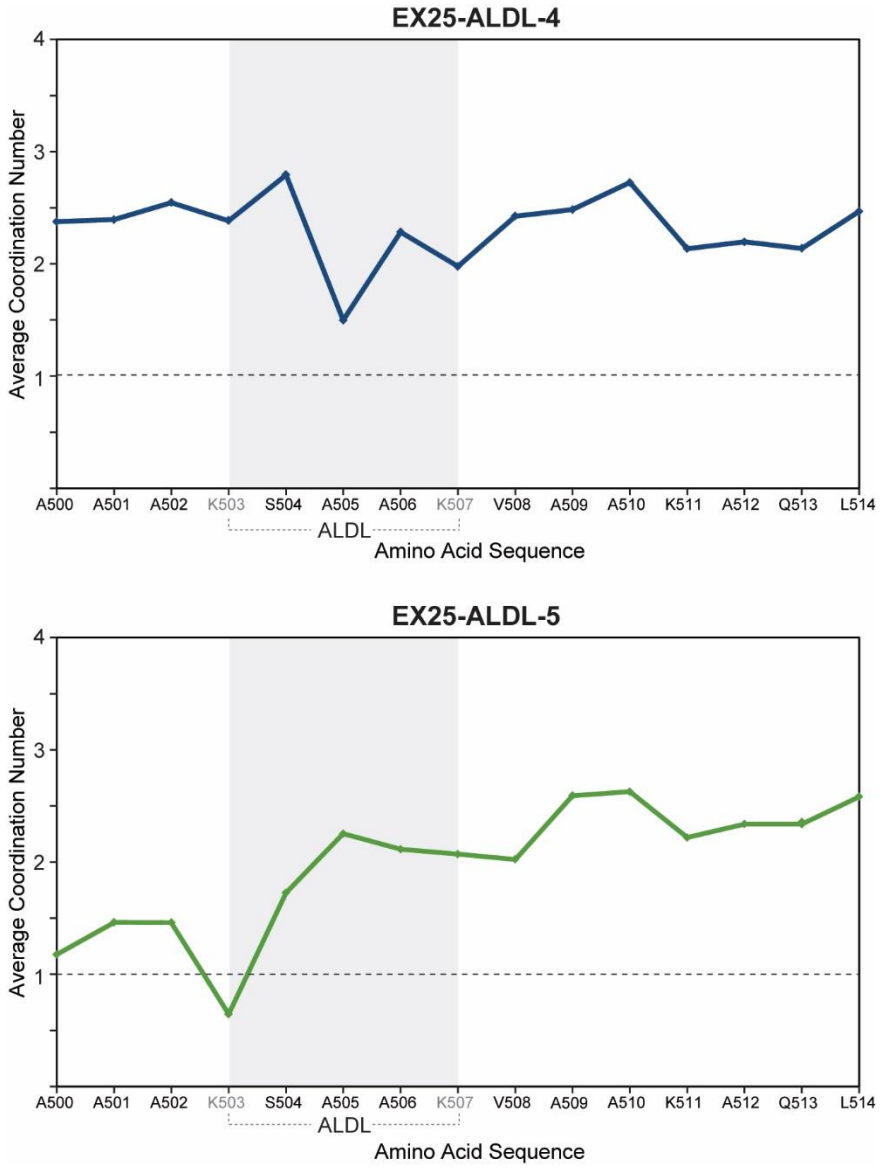


Figure A.16 Average coordination number of water for each amino acid in EX25-ALDL-4 and -5. The first hydration shell is defined as the region where water molecules are found within 3.6 Å of the peptide's backbone. The average coordination number of water in is calculated over the last 100 ns of the MD trajectory, during which a substantial α-helical segment is adopted. Gray shade indicates amino acids that are tethered by ALDL crosslink (i.e., modified K503 and K507). Horizontal dashed line indicates the average coordination number of 1, typical for carbonyl oxygens found in the α-helical structure.

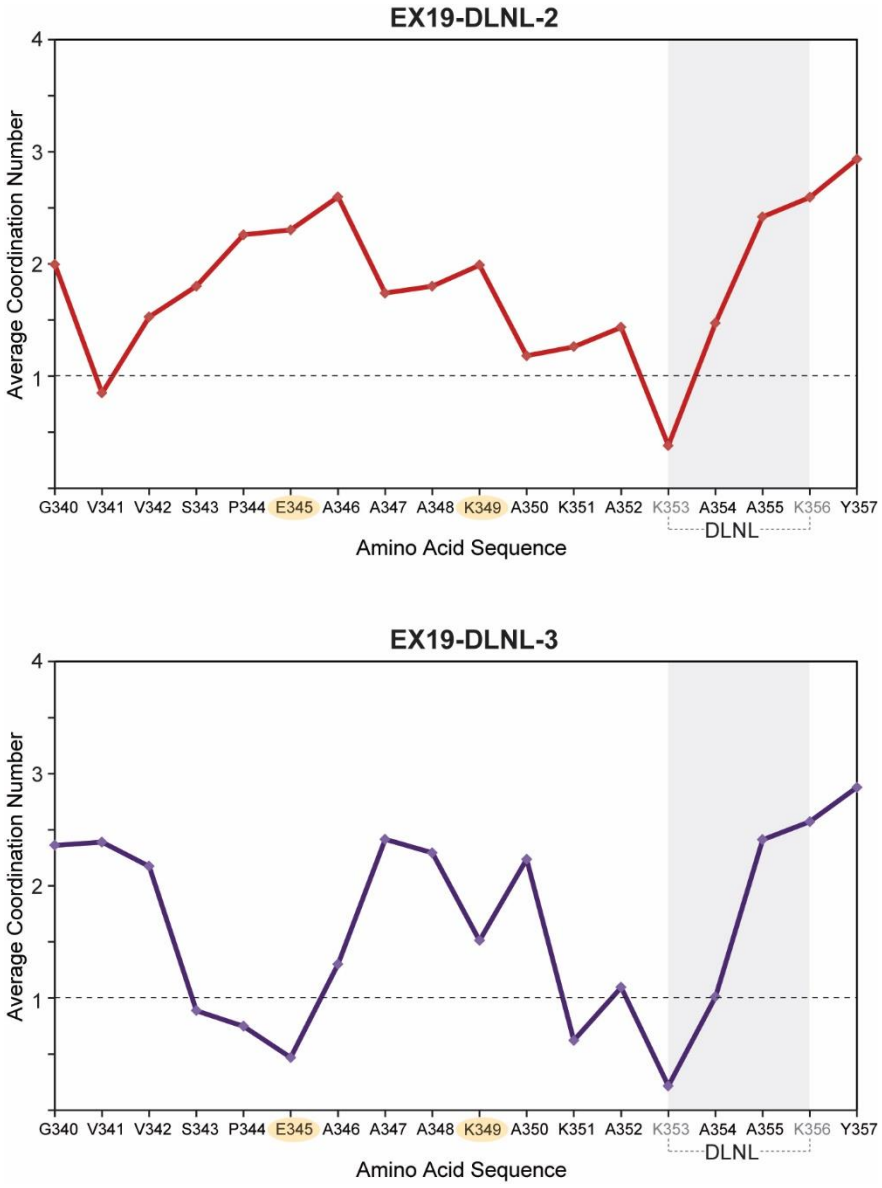


Figure A.17 Average coordination number of water for each amino acid in EX19-DLNL-2 and -3. The first hydration shell is defined as the region where water molecules are found within 3.6 Å of the peptide's backbone. The average coordination number of water in is calculated over the last 100 ns of the MD trajectory, during which a substantial α-helical segment is adopted. Gray shade indicates amino acids that are tethered by DLNL crosslink (i.e., modified K353 and K356). Horizontal dashed line indicates the average coordination number of 1, typical for carbonyl oxygens found in the α-helical structure. Positive and negative charges in K349 and E345, respectively, (yellow shaded) indicate the salt-bridge interaction in EX19-DLNL.

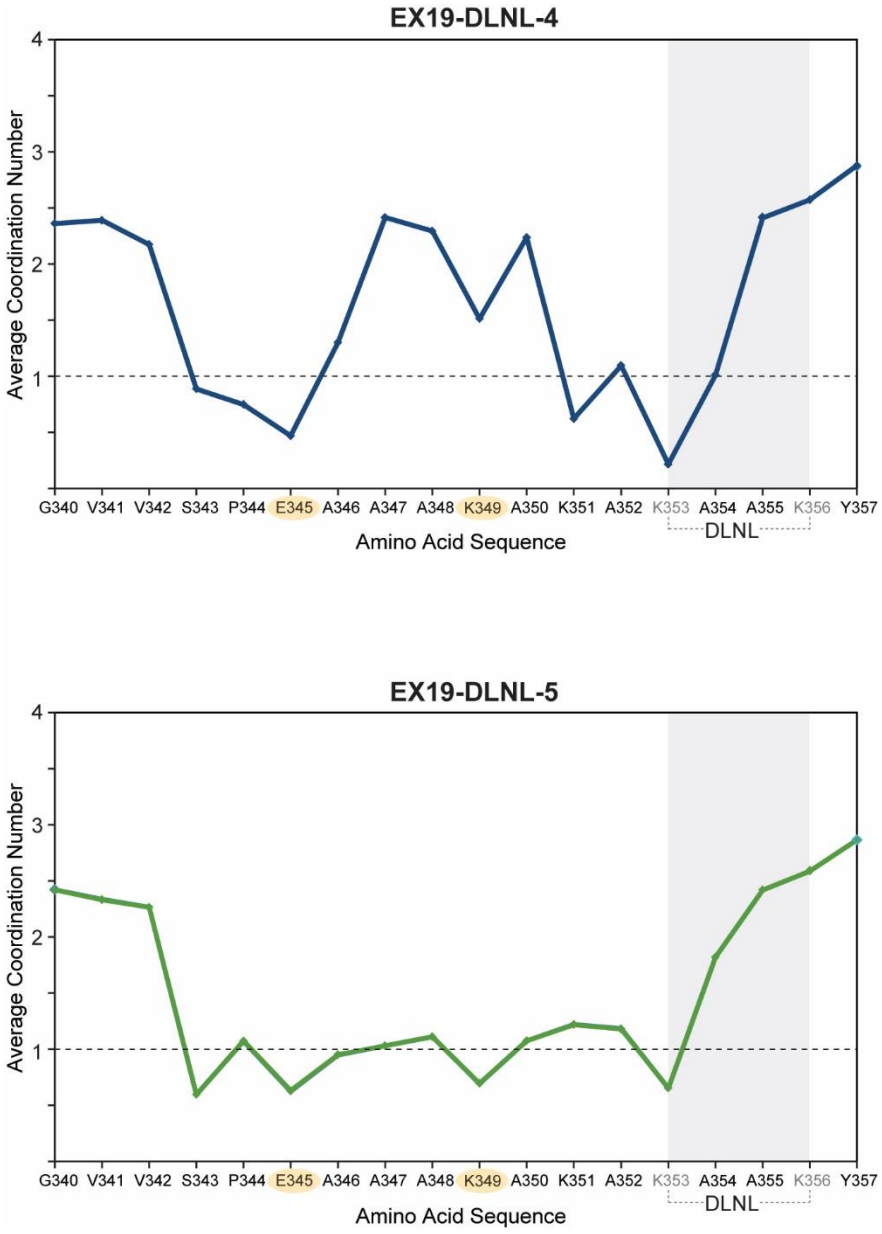


Figure A.18 Average coordination number of water for each amino acid in EX19-DLNL-4 and -5. The first hydration shell is defined as the region where water molecules are found within 3.6 Å of the peptide's backbone. The average coordination number of water in is calculated over the last 100 ns of the MD trajectory, during which a substantial α-helical segment is adopted. Gray shade indicates amino acids that are tethered by ALDL crosslink (i.e., modified K353 and K356). Horizontal dashed line indicates the average coordination number of 1, typical for carbonyl oxygens found in the α-helical structure. Positive and negative charges in K349 and E345, respectively, (yellow shaded) indicate the salt-bridge interaction in EX19-DLNL.

APPENDIX 15. BACKGROUND INFORMATION ON THE HELIX-COIL TRANSITION THEORY

Introduction to the Helix-Coil Transition Theory

The general problem that the helix-coil transition theory aims to solve deals with the probability of occurrence of a particular conformation in a given molecule with N linked units, in which each unit exists in several states. In a protein-based polymer, each unit corresponds to each amino acid in the sequence that can either exist in the helix (h) or the coil (c) state. A residue is in an h state if its dihedral angles are found within the region defined as the α-helix. Otherwise this residue is in a c state. Hence, each polymer has 2^N states, which can be generated by $(h+c)^N$. For example, a polymer with N=12 may have a particular conformation that contains alternating h's and c's, such as

$$hhhcchhcchc \qquad \text{(eqn. A.5.1)}$$

The partition function for a homopolymer system containing N residues encompass all possible conformations, calculated by

$$Z(N) = \sum_{\{h,c\}} exp[-\frac{G\{h,c\}}{RT}] \qquad \text{(eqn. A.5.2)}$$

where G{h,c} is the free energy for each conformation, and {h,c} indicates a given sequence of h's and c's. The probability of occurrence of a given conformation is therefore,

$$P\{h,c\} = {exp[-\frac{G\{h,c\}}{RT}]}\Big/{Z(N)} \qquad \text{(eqn. A.5.3)}$$

Furthermore, the partition function for a particular part of the molecule is referred as the "statistical weight", which is a general notation for exp(-G/RT). For instance, the statistical weight of a polymer sequence containing only c states is,

$$u_i = exp[-\frac{G_i(c)}{RT}] \qquad \text{(eqn. A.5.4)}$$

where i is the length of the sequence. Correspondingly, the statistical weight of a sequence with j units containing only h states is,

$$v_j = exp[-\frac{G_j(h)}{RT}] \qquad \text{(eqn. A.5.5)}$$

Therefore, the conformation of a polymer with N=12 has the component statistical weights of helix and coil sequences that include,

$$\begin{matrix} h\,h\,h\,c\,c\,c\,h\,h\,c\,c\,c\,c \\ (\ v_3\ )(u_3)\,(v_2)(\ u_4\ ) \\ c\,c\,c\,c\,c\,h\,h\,c\,c\,c\,c \end{matrix}$$

$$\begin{matrix}(\ \ u_3\ \ )(\ v_3\ )(\ u_4\ ) \\ etc.\end{matrix} \qquad \text{(eqn. A.5.6)}$$

, and the partition function for this polymer is

$$Z(12) = v_3u_3v_2u_4 + u_3v_3u_4 + etc. \qquad \text{(eqn. A.5.7)}$$

Matrix Treatment of the Partition Function

The partition function for a polymer represents the sum over all states of the system, i.e., all combinations of h's and c's. As described earlier, this function is obtained by the explicit summation of statistical weights, w and u, by

$$Z(N) = (w + u)^N \qquad \text{(eqn. A.5.8)}$$

, in which the assignment for each residue is independent from other (adjacent) residues. However, this summation is not applicable to a complex system, in which the statistical weight assigned to each residue requires the knowledge of states of other residues in the sequence. A such system requires matrix multiplication to construct the partition function, i.e.,

$$Z(N) = (Matrix)^N \qquad \text{(eqn. A.5.9)}$$

The matrix treatment of the partition function was used by Lifson and Roig to assign the statistical weight to a given residue, *i*, based on the states of the flanking residues, *i-1* and *i+1*. Specifically, a 3×3 matrix, *W*, assigns the *ith* residue as *w*, *v*, or *u*, according to the h or c state of its (flanking) neighbors, by

$$W = \begin{array}{cc|ccc} & i+1 & h & c & c\cup h \\ i-1 & i & h & h & c \\ \hline h & h & w & v & 0 \\ h & c & 0 & 0 & u \\ c & c\cup h & v & v & u \end{array} \qquad \text{(eqn. A.5.10)}$$

, where the notation c ∪ h means c or h. The rows of this matrix correspond to the statistical weight values for residues *i-1* and *i*, whereas the columns correspond to those for residues *i* and *i+1*. For instance, if a polymer with N=3 has a conformation described by $\underline{\mathrm{h}}\mathbf{h}\underline{\mathrm{c}}$, then the statistical weight assigned to the *ith* residue (in the **h** state) is *v*, because the *(i-1)th* and *(i+1)th* residues are found in the $\underline{\mathrm{h}}$ and $\underline{\mathrm{c}}$ states, respectively.

With the matrix *W* and the rules of matrix multiplication, the partition function can be represented by the product

$$Z(N) = \boldsymbol{u}W^{\boldsymbol{n-2}}\boldsymbol{u}^{+} \qquad \text{(eqn. A.5.11)}$$

, where **u** and $\mathbf{u}^+$ are the row and column vectors, respectively, which represents the two terminal residues of the chain by

$$\boldsymbol{u} = (v\ v\ u) \qquad \boldsymbol{u}^{+} = \begin{pmatrix} v \\ u \\ v+u \end{pmatrix} \qquad \text{(eqn. A.5.12)}$$

The vector $\mathbf{u}$ is obtained from the last row of the matrix W, the vector $\mathbf{u}^{+}$ is obtained from the sum of the last two columns in W. Apart from the special cases that consider very small chains, these two terms can be simplified into the row and column vectors

$$\boldsymbol{e} = (0\ 0\ 1) \qquad \boldsymbol{e}^{+} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \qquad \text{(eqn. A.5.13)}$$

such that the partition function can now be written as

$$Z(N) = \boldsymbol{e}W^{\boldsymbol{n}}\boldsymbol{e}^{+} \qquad \text{(eqn. A.5.14)}$$

Matrix Multiplication in Lifson-Roig's Model

The matrix treatment of the partition function, $\mathrm{Z(N)}$, is illustrated (below) for polymers with N=3 and N=4. A polymer containing 3 residues has 8 (i.e., 2^3) conformations that include

$$hhh\ hhc\ hch\ chh\ hcc\ chc\ cch\ ccc \qquad \text{(eqn. A.5.15)}$$

The statistical weights for each residue in these states are obtained from the product of eqn. A.5.11, which is expressed by

$$Z(3) = (v \quad v \quad u)\begin{pmatrix} w & v & 0 \\ 0 & 0 & u \\ v & v & u \end{pmatrix}\begin{pmatrix} v \\ u \\ u+v \end{pmatrix} \qquad \text{(eqn. A.5.16)}$$

, resulting in

$$Z(3) = vwv + vvu + vuv + uvv + vuu + uvu + uuv + uuu \qquad \text{(eqn. A.5.17)}$$

A polymer containing 4 residues has 16 (i.e., 2^4) conformations that include

$$\begin{matrix} hhhh & hhhc & hhch & hchh \\ hhcc & hchc & chhc & chch \\ cchh & hcch & hccc & chhh \\ chcc & cchc & ccch & cccc \end{matrix} \qquad \text{(eqn. A.5.18)}$$

The statistical weights for each residue in these states are obtained from the product of eqn. A.5.11, which is expressed by

$$Z(4) = (v \quad v \quad u)\begin{pmatrix} w & v & 0 \\ 0 & 0 & u \\ v & v & u \end{pmatrix}\begin{pmatrix} w & v & 0 \\ 0 & 0 & u \\ v & v & u \end{pmatrix}\begin{pmatrix} v \\ u \\ u+v \end{pmatrix} \qquad \text{(eqn. A.5.19)}$$

, resulting in

$$Z(4) = vwwv + vwvu + vvuv + vuvv +$$
$$uvwv + vvuu + vuvu + uvvu +$$
$$uvuv + uuvv + vuuv + vuuu +$$
$$uvuu + uuvu + uuuv + uuuu \quad \text{(eqn. A.5.20)}$$

APPENDIX 16. MATLAB SCRIPT TO SIMULATE ^{13}CO NMR LINESHAPES BASED ON BIXON-SCHERAGA-LIFSON'S HELIX-COIL THEORY AND ULLMAN'S APPROACH

```
% parameters
w=1.0; %h-bondedness 1.0
v=0.03;
s=1.7; %sigma
A=0.1; %line broadness
u=1.0;
a=1+sqrt(v)*(s-1);% alpha

dh=180; %chemical shift for helix, in eqn. 5.10
dc=175; %chemical shift for coil, in eqn. 5.10

x=[173:0.1:180]; %Range of x-coordinates
% Eqn. 5.7
W=[w*s  v  0;
       0   0  v;
       a*u u  u];
% Eqn. 5.9
H=[w*s  v  0;
     0   0  v;
     0   0  0];

% To produce Figures 5.8-5.9,
%n_vec=[7 8 10 11 12 14 15 ];
%Gn=[1 1 1 1 2 3 1];
n_vec=[2 3 4 6 7 8 9];
Gn=[10 8 3 1 1 1 1];

Ind_mat=[];

 for jj=1:length(n_vec)
    n=n_vec(jj);
     % Eqn. 5.4
    Zn=[1 1 1]*W^n*[0;0;1];

    Pnk_vec=[];
    Ink_mat=[];

    for ii=1:n % Summation should be from 1 to n, but not to n-2
        i=ii;
        % Eqn. 5.8
        Fni=[1 1 1]*W^(i-1)*H*W^(n-i)*[0;0;1];
        % Eqn. 5.3
        Pni=Fni/Zn;
        % Store Pni into Pni_vec
        Pni_vec=cat(1,Pni_vec,Pni);
        % Eqn. 5.10
```

```
        dni=dh*Pni+dc*(1-Pni);
        % Eqn. 5.11
        Ink=1/A*exp(-(x-dnk).^2/(2*A^2));
        % Store Ink into Ink_mat
        Ink_mat=cat(1,Ink_mat,Ink);
    end

    % Eqn. 5.12
    Ind=sum(Ink_mat);

    % Store Ind into Ind_mat
    Ind_mat=cat(1,Ind_mat,Ind);
    hold on
    plot(x,Gn(jj)*Ind)
    hold off
    set(gca,'xdir','reverse')

end

Gn_mat=repmat(Gn',1,length(x));
hold on
plot(x,sum(Gn_mat.*Ind_mat),'r');
hold off

 %Plot for Helicity, to produce Figure 5.7
 %Remove comments to run
 %figure
 %hold on
 %for ii=1:7 ;%length of residues
 %    plot(Pni_vec(ii)*[1 1],[0 1])
 %end
 %plot(0*[1 1],[0 1],'k')
 %plot(1*[1 1],[0 1],'k')
 %hold off
 %set(gca,'xdir','reverse')
 %xlim([0 1])
```

APPENDIX 17. ^{13}CO-ALA LINESHAPES AT 37, 20, 10, 0, AND -4 °C

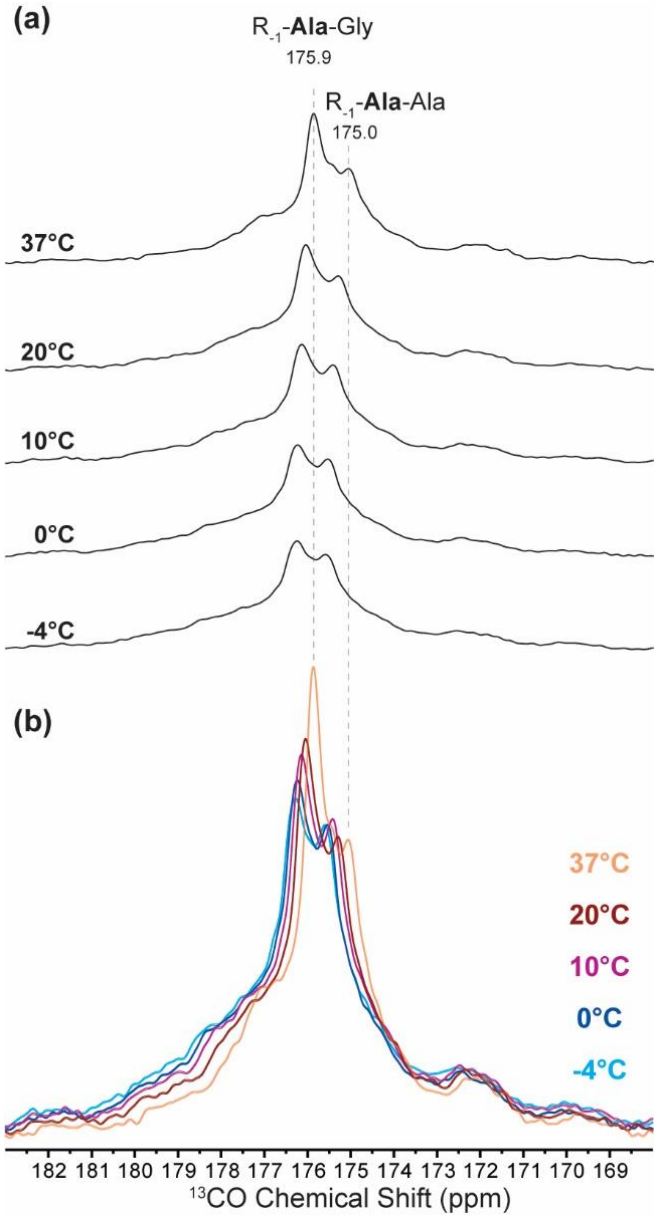


Figure A.19 (a) VT spectra of ^{13}CO DP with LOW-BASHD homonuclear decoupling, as a function of temperature and (b) superposition of the spectra. LOW-BASHD decoupling was implemented to remove the line broadening effect from *J*-couplings and to simplify peak assignments. Dotted lines (⋮) on the VT spectra indicate the ^{13}C chemical shifts assigned for alanines in the R$_{-1}$-**Ala**-Gly, R$_{-1}$-**Ala**-Z (Z≠Ala,Gly) and R$_{-1}$-**Ala**-Ala motifs.

APPENDIX 18. MISCELLANEOUS SCRIPTS FOR REMD SIMULATIONS

EX19DLNL_BASE.namd

```
# shared simulation options for replica exchange
structure          EX19DLNL.psf
coordinates        EX19DLNL.pdb
# Forcefield parameters
paraTypeCharmm          on
parameters      ../../par_all22_DES_ALDL-rev2.inp
exclude            scaled1-4
1-4scaling         1.0
switching          on
switchdist         15.0
cutoff             16.0
pairlistdist       18.0
#GBIS parameters
GBIS             on
ionConcentration  0.15
alphaCutoff       14
solventDielectric 80
#hydrophobic Energy
sasa             on
surfaceTension    0.006
# Integrator parameters
timestep              1
nonbondedFreq      2
fullElectFrequency    4
#rigidBonds all
stepspercycle    20
# Some Temp Control in replica.namd script
# ...langevin, langevinTemp,
# ...
# ... langevinDamping (originally 10.0)
langevin on
langevinDamping 5.0
margin             10.0
# output stuff, but some included in the replica.namd
# ...outputname, dcdFreq, outputEnergies
# ...
outputTiming      1000   ;# every 10ps
#outputPressure    100    ;# similar to outputEnergies [steps_per_run/10]
```

Fold_EX19DLNL.conf

```
# configuration for replica exchange scripts
# run simulation:
#   mkdir output
#   (cd output; mkdir 0 1 2 3 4 5 6 7)
#   mpirun -np 8 -hostfile hostfile $bindir/namd2 +replicas 8 job0.conf +stdout output/%d/job0.%d.log
# the number of MPI ranks (-np) must be a multiple of the number of replicas (+replicas)
# to continue:
```

```
#   mpirun -np 8 -hostfile hostfile $bindir/namd2 +replicas 8 job0.conf +stdout output/%d/job1.%d.log
# increase num_runs below if job completed, or use latest restartXXX.tcl file available
# be sure to increment jobX for +stdout option on command line
# view in VMD:  source job0.conf; source ../show_replicas.vmd
# add continued:   source job1.conf; source ../show_replicas.vmd
# show both:  vmd -e load_all.vmd
# sort into single-temperature trajectories:
#   $bindir/sortreplicas output/%s/fold_alanin.job0 8 10
#   $bindir/sortreplicas output/%s/fold_alanin.job1 8 10
set num_replicas 16
set min_temp 263
set max_temp 574
set steps_per_run 1000
#steps_per_run is exchange frequency 1000(1fs/step) = 1ps
set num_runs 5000
# total simulations = steps-per-run * num-runs = 1ps * 5000 = 5 ns
# num_runs should be divisible by runs_per_frame * frames_per_restart
set runs_per_frame 10
set frames_per_restart 10
set namd_config_file "EX19DLNL_base.namd"
#set output_root "output/%s/fold_alanin" ; # directories must exist
set output_root "output/%s/fold_EX19DLNL" ; # directories must exist
# the following used only by show_replicas.vmd
set psf_file "EX19DLNL.psf"
set initial_pdb_file "EX19DLNL.pdb"
set fit_pdb_file "EX19DLNL.pdb"
```

Job0.conf

```
source fold_EX19DLNL.conf
# prevent VMD from reading replica.namd by trying command only NAMD has
if { ! [catch numPes] } { source ../../replica.namd }
```

Job1.conf

```
source fold_EX19DLNL.conf
source [format $output_root.job0.restart5000.tcl ""]
set num_runs 10000
# prevent VMD from reading replica.namd by trying command only NAMD has
if { ! [catch numPes] } { source ../../replica.namd }
```

REMDScript

```
#!/bin/bash
#----------------------------------------------------
# Example SLURM job script to run MPI applications on
# # TACC's Stampede system.
#SBATCH -J Jdremd  # Job name
#SBATCH -e err.%j.txt # error output
#SBATCH -p development # Queue name
#SBATCH -N 1 # Total number of nodes requested (16 cores/node)
#SBATCH -n 16 # Total number of mpi tasks requested
#SBATCH -t 02:00:00 # Run time (hh:mm:ss) - 2 hrs
```

```
#SBATCH -A TG-MCB140225 # <-- Allocation name to charge job against
#SBATCH --mail-user=jhonsen@hawaii.edu
#SBATCH --mail-type=ALL
#SBATCH --dependency=afterok:138201 #equivalent of -hold_jid
# Launch the MPI executable named "job0.conf"
module load namd/2.9
export MKL_MIC_ENABLE=1
export I_MPI_EAGER_THRESHOLD=262144
cd /work/01309/jhonsen/RepTEST/example
mkdir output
(cd output; mkdir 0 1 2 3 4 5 6 7)
ibrun -o 0 -n 8 namd2 +replicas 8 job0.conf >& /output/job0.%d.log
ibrun -o 8 -n 8 namd2 +replicas 8 job1.conf >& /output/job1.%d.log
wait
```

APPENDIX 19. GOR IV AND AGADIR PREDICTIONS OF HELICAL CONTENT IN RAT TROPOELASTIN

Table A.5. Predicted α-Helical Content in Tropoelastin's Crosslinking Domains using AGADIR [1-9] and GOR IV [10] algorithms

EXON	Amino Acid Sequences	%-Helical Content	
		AGADIR[$]	GOR IV
4	ALGPGGKPPKP	0.3	0.0
6	GGVLVPGGGAGAAAAYKAAAKA	4.9	36.4
8	GAVVPQLGAGVGAGGKPGKVP	0.0	0.0
10	GARFPGVGVLPGVPTGTGVKAKVP	0.1	0.0
12*	GVGPFGGQQPGVPLGYPIKAPKLP	0.0	0.0
15*	GVGSQAAVAAAKAAKY	7.5	43.8
17	GAGTPAAAAAAKAAAKAAKY	19.4	70.0
19	GAVSPAAAAKAAAKAAKY	15.4	66.7
21	GASQAAAAAAAAKAAKY	26.6	70.6
23	GTPAAAAAAAAAKAAAKAGQY	26.9	61.9
25	GAGTPAAAKSAAKAAAKAQY	4.5	55.0
27	VPGSLAASKAAKY	2.1	0.0
29	GAPAAAAAAKAAAKAAQY	12.2	72.2
31	GVSPAAAAKAAKY	9.4	30.8
35	GKPPKPYGGALGALGYQ	0.2	0.0

Notes:

[#] Exons 12 and 15 are subject to alternative splicing [11]

[$] Parameters for AGADIR predictions:

- Temperature = 300 K
- N- and C-termini are acetylated and amidated
- Ionic strength = 0.15 M
- Solution pH = 7.0

[1] Serrano, L., J.-L. Neira, J. Sancho, and A.R. Fersht, *Effect of alanine versus glycine in [alpha]-helices on protein stability.* Nature, 1992. **356**(6368): p. 453-455.

[2] Muñoz, V., and L. Serrano, *Elucidating the folding problem of helical peptides using empirical parameters.* Nature Structural & Molecular Biology, 1994. **1**(6): p. 399-409.

[3] Muñoz, V., and L. Serrano, *Elucidating the folding problem of helical peptides using empirical parameters. II. Helix macrodipole effects and rational modification of the helical content of natural peptides.* Journal of Molecular Biology, 1995. **245**(3): p. 275-296.
[4] Muñoz, V., and L. Serrano, *Elucidating the folding problem of helical peptides using empirical parameters. III. Temperature and pH dependence.* Journal of Molecular Biology, 1995. **245**(3): p. 297-308.
[5] Villegas, V., A.R. Viguera, F.X. Avilés, and L. Serrano, *Stabilization of proteins by rational design of α-helix stability using helix/coil transition theory.* Folding and Design, 1996. **1**(1): p. 29-34.
[6] Muñoz, V., and L. Serrano, *Local versus nonlocal interactions in protein folding and stability – an experimentalist's point of view.* Folding and Design, 1996. **1**(4): p. R71-R77.
[7] Muñoz, V., and L. Serrano, *Development of the multiple sequence approximation within the AGADIR model of α-helix formation: Comparison with Zimm-Bragg and Lifson-Roig formalisms.* Biopolymers, 1997. **41**(5): p. 495-509.
[8] Viguera, A.R., V. Villegas, F.X. Avilés, and L. Serrano, *Favourable native-like helical local interactions can accelerate protein folding.* Folding and Design, 1997. **2**(1): p. 23-33.
[9] López-Hernández, E., P. Cronet, L. Serrano, and V. Muñoz, *Folding kinetics of Che Y mutants with enhanced native α-helix propensities.* Journal of Molecular Biology, 1997. **266**(3): p. 610-620.
[10] Garnier, J., J.-F. Gibrat, and B. Robson, *GOR method for predicting protein secondary structure from amino acid sequence*, in *Methods in Enzymology*. 1996, Academic Press. p. 540-553.
[11] Pierce, R.A., A. Alatawi, S.B. Deak, and C.D. Boyd, *Elements of the rat tropoelastin gene associated with alternative splicing.* Genomics, 1992. **12**(4): p. 651-658.