TANDEM REPEAT PROTEINS INSPIRED BY SQUID RING TEETH

A Dissertation in
Engineering Science and Mechanics

by
Abdon Pena-Francesch

© 2017 Abdon Pena-Francesch

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2017
The dissertation of Abdon Pena-Francesch was reviewed and approved* by the following:

Melik C. Demirel  
Professor of Engineering Science and Mechanics  
Director of Center for Research on Advanced Fiber Technologies  
Dissertation Advisor  
Chair of Committee

Joseph P. Cusumano  
Professor of Engineering Science and Mechanics

Reginald F. Hamilton  
Associate Professor of Engineering Science and Mechanics

Ralph H. Colby  
Professor of Materials Science and Engineering and Chemical Engineering

Judith A. Todd  
P.B. Breneman Professor of Engineering Science and Mechanics  
Head of the Department of Engineering Science and Mechanics

*Signatures are on file in the Graduate School
Abstract

Proteins are large biomolecules consisting of long chains of amino acids that hierarchically assemble into complex structures, and provide a variety of building blocks for biological materials. The repetition of structural building blocks is a natural evolutionary strategy for increasing the complexity and stability of protein structures. However, the relationship between amino acid sequence, structure, and material properties of protein systems remains unclear due to the lack of control over the protein sequence and the intricacies of the assembly process.

In order to investigate the repetition of protein building blocks, a recently discovered protein from squids is examined as an ideal protein system. Squid ring teeth are predatory appendages located inside the suction cups that provide a strong grasp of prey, and are solely composed of a group of proteins with tandem repetition of building blocks. The objective of this thesis is the understanding of sequence, structure and property relationship in repetitive protein materials inspired in squid ring teeth for the first time. Specifically, this work focuses on squid-inspired structural proteins with tandem repeat units in their sequence (i.e., repetition of alternating building blocks) that are physically cross-linked via β-sheet structures. The research work presented here tests the hypothesis that, in these systems, increasing the number of building blocks in the polypeptide chain decreases the protein network defects and improves the material properties.

Hence, the sequence, nanostructure, and properties (thermal, mechanical, and conducting) of tandem repeat squid-inspired protein materials are examined. Spectroscopic structural analysis, advanced materials characterization, and entropic elasticity theory are combined to elucidate the structure and material properties of these repetitive proteins. This approach is applied not only to native squid proteins but also to squid-inspired synthetic polypeptides that allow for a fine control of the sequence and network morphology.

The results provided in this work establish a clear dependence between the repetitive building blocks, the network morphology, and the properties of squid-inspired repetitive protein materials. Increasing the number of tandem repeat units in SRT-inspired proteins led to more effective protein networks with superior properties. Through increasing tandem repetition and optimization of network morphology, highly efficient protein materials capable of withstanding deformations up to 400% of their original length, with MPa-GPa modulus, high energy absorption (50 MJ m$^{-3}$),
peak proton conductivity of 3.7 mS cm\(^{-1}\) (at pH 7, highest reported to date for biological materials), and peak thermal conductivity of 1.4 W m\(^{-1}\) K\(^{-1}\) (which exceeds that of most polymer materials) were developed. These findings introduce new design rules in the engineering of proteins based on tandem repetition and morphology control, and provide a novel framework for tailoring and optimizing the properties of protein-based materials.
# Table of Contents

List of Figures ............................................................................................................. viii
List of Tables ................................................................................................................ xv
List of Abbreviations .............................................................................................. xvi
List of Symbols ......................................................................................................... xviii
Acknowledgements .................................................................................................. xx

Chapter 1. General introduction ................................................................................. 1
  1.1 Protein materials ............................................................................................... 2
  1.2 Structural motifs in proteins ............................................................................ 4
  1.3 Cephalopod-derived biological materials ....................................................... 11
  1.4 Scope and outline of the thesis ....................................................................... 14

Chapter 2. Experimental approach ............................................................................. 18
  2.1 Proteomics ........................................................................................................ 18
  2.2 Structural characterization ............................................................................. 22
  2.3 Thermal analysis ............................................................................................. 24
  2.4 Mechanical analysis ......................................................................................... 25
  2.5 Neutron scattering ......................................................................................... 28
  2.6 Conductivity measurements ........................................................................... 32

Chapter 3. Native SRT: sequence and structure ......................................................... 34
  3.1 Introduction ....................................................................................................... 34
  3.2 Results and discussion .................................................................................... 36
    3.2.1 Amino acid sequence analysis ............................................................... 36
    3.2.2 Structural analysis .................................................................................. 42
    3.2.3 Thermal properties ................................................................................ 50
    3.2.4 Mechanical properties .......................................................................... 53
  3.3 Conclusions ..................................................................................................... 55
Chapter 4. Semicrystalline morphology and thermomechanical properties of recombinant SRT ................................................................. 57

4.1 Introduction ........................................................................................................... 57
4.2 Results and discussion ....................................................................................... 62
  4.2.1 Structural analysis of recombinant SRT proteins ........................................... 66
  4.2.2 Thermal analysis of semicrystalline SRT materials ....................................... 74
  4.2.3 Mechanical properties of semicrystalline SRT materials ............................... 77
  4.2.4 Hydration of recombinant SRT protein ...................................................... 79
4.3 Conclusions ......................................................................................................... 93

Chapter 5. SRT-inspired tandem repeat proteins: network structure and mechanics ....... 94

5.1 Introduction ........................................................................................................... 94
5.2 Results and discussion ....................................................................................... 97
  5.2.1 SRT-inspired tandem repeat sequence ....................................................... 97
  5.2.2 Nanostructure of SRT-inspired tandem repeat protein networks .................. 99
  5.2.3 SRT-inspired protein networks: glassy mechanics ...................................... 103
  5.2.4 SRT-inspired protein networks: rubbery mechanics ................................. 107
5.3 Conclusions ......................................................................................................... 117

Chapter 6. SRT-inspired tandem repeat proteins: transport properties ......................... 119

6.1 Introduction ........................................................................................................... 119
6.2 Results and discussion ....................................................................................... 121
  6.2.1 SRT-inspired networks: chain dynamics ................................................... 121
  6.2.2 SRT-inspired protein networks: bulk proton conductivity ......................... 127
  6.2.3 SRT-inspired protein networks: thermal conductivity .............................. 138
6.3 Conclusions ......................................................................................................... 144

Chapter 7. Applications of SRT .................................................................................. 146

7.1 Fabrication and processing of SRT materials .................................................. 146
  7.1.1 Solution processing .................................................................................... 147
  7.1.2 Thermal processing .................................................................................... 149
7.2 SRT pressure sensitive underwater adhesives ............................................... 152
7.3 SRT self-healing materials .............................................................................. 157
7.4 SRT-based photonic devices ........................................................................... 163
Chapter 8. Conclusions ........................................................................................................................................... 168

8.1 Recommendations for future work .................................................................................................................. 170
  8.1.1 Self-assembly of SRT-inspired proteins and nanoconfinement of β-sheets ........................................... 170
  8.1.2 High-throughput screening of protein libraries ....................................................................................... 171
  8.1.3 SRT-based devices .................................................................................................................................... 172
  8.1.4 Controlled assembly of 2D materials ....................................................................................................... 174

Appendix. Representative publications .................................................................................................................. 176

References ................................................................................................................................................................... 178
List of Figures

Figure 1-1. Protein materials can be tailored to match the elasticity of a variety of biological tissues, from fat and kidney tissue (kPa), skin tissue (MPa) to bone tissue (GPa). .................................................. 2

Figure 1-2. Synthetic polymers vs. natural and synthetic proteins. Key advantages of proteins are: precise control over the sequence, precise control over the molecular weight (monodispersity), and the assembly into complex hierarchical structures. Adapted from ref.18 .............................................. 3

Figure 1-3. Description of major natural protein materials (resilin, collagen, silk, elastin, keratin, and SRT), their representative structure and sequence structural motifs. Adapted from ref.16 ....................... 6

Figure 1-4. Anatomy of European common squid (Loligo vulgaris) and its advanced biological materials. .............................................................................................................. 12

Figure 1-5. Squid ring teeth (SRT): a) European common squid (Loligo vulgaris) is caught in the Mediterranean Sea near Tarragona. b) SRT are located in the suction cups along the arms and tentacles. c) Electron microscopy of the suction cups reveals a columnar epithelium (inset). d) Suction cup tissue is characterized by targeting F-actin with phalloidin dyes and nuclei with DAPI staining. .................................. 13

Figure 1-6. Outline of the thesis.......................................................................................................................... 16

Figure 2-1. Digital image correlation (DIC) setup (a) with DMA and digital camera. b) Dog-bone specimen after speckle pain was applied. c) Close-up of the speckle pattern in the DIC area of interest. d) Pixel histogram of the speckled area. e) Particle size analysis of the speckle pattern. ........................................ 26

Figure 2-2. Underwater tensile test setup. a) Instron 5866 load frame equipped with a custom-built liquid chamber. b) Detail of the custom-built liquid chamber. ..................................................................................................... 27

Figure 2-3. Resolution of neutron spectrometers at NCNR: high-flux backscattering spectrometer (HFBS), disk chopper time-of-flight spectrometer (DCS), and neutron spin echo (SE). ......................... 30

Figure 2-4. Quasielastic neutron scattering spectra of hydrated SRT protein. The graph shows the instrumental elastic resolution (solid black line), two Lorentzian functions (red and blue solid lines) used to describe the dynamic processes, the experimental data (yellow dots), and the total fit (dashed line).... 31

Figure 2-5. Nyquist plot of SRT hydrated protein from impedance spectroscopy. Equivalent circuit for the calculation of bulk resistance (inset). ......................................................................................... 32

Figure 3-1. SEM cross-section of SRT from six squid species reveals a highly organized tubular structure. Insets show the varying morphology of SRT across species. ........................................................................ 35

Figure 3-2. SDS-PAGE of SRT from six squid species (Todarodes pacificus, Sepioteuthis lessoniana, Illex argentinus, Loligo vulgaris, Loligo pealei, and Illex illecebrosus). SRT across species are composed of a group of proteins between 15 and 60 kDa. .................................................................................... 36

Figure 3-3. Amino acid composition of Loligo vulgaris native SRT. Major amino acids are glycine (26.2%), tyrosine (12.5%), and histidine (10.9%). ........................................................................................................ 38

Figure 3-4. a) Major amino acids in Loligo vulgaris native SRT and the pKa of their side chain residue. b) ζ-potential measurements of native SRT show an isoelectric point at pH ~ 6.6........................................ 39
Figure 3-5. Amino acid sequence of proteins composing native *Loligo vulgaris* SRT. GLY-rich segments are marked in yellow, AVSTH-rich segments are marked in blue, and proline (P) residues are marked in bold.

Figure 3-6. a) Segment length distribution and b) histidine-proline distances in *Loligo vulgaris* native SRT (over all sequenced proteins).

Figure 3-7. FTIR spectra of *Loligo vulgaris* native SRT. a) SRT spectra exhibits amide I, amide II, amide III, and amide A bands (characteristic of proteins). b) Amide I band is centered at 1630 cm⁻¹ indicating a high fraction of β-sheet structures.

Figure 3-8. Deconvoluted FTIR amide I band of native SRT shows a major β-sheet content. Individual bands are fitted to the spectra, and are assigned to side chains (sc), β-sheet (β), α-helices (α), and turn (t) structures as described in Table 3-2.

Figure 3-9. Schematic representing parallel and antiparallel β-sheet configurations in SRT crystalline segments. Simulation courtesy of Murat Cetinkaya.

Figure 3-10. X-ray diffraction (XRD) of *Loligo vulgaris* native SRT. Major diffraction peaks are observed at 2Θ = 9.3° and 2Θ = 19.2°, which correspond to alternating β-sheet strands.

Figure 3-11. Temperature modulated differential scanning calorimetry (TMDSC) of native SRT. Reversing heat flow (dashed line) shows a step change around 180 °C, indicating a glass transition.

Figure 3-12. Thermal degradation of native SRT measured by thermogravimetric analysis coupled to mass spectrometry (TGA-MS) is observed starting at 210°C.

Figure 3-13. Dynamic mechanical analysis (DMA) of (a) dry (water content ~0.2%) and (b) fully hydrated (water content ~20%) native SRT shows stable GPa moduli in the glassy region, and a glass transition at 190-210 °C and ~40 °C respectively.

Figure 3-14. Dynamic rheology of fully hydrated (water content ~20%) native SRT at 70 °C shows a weak frequency dependence and a rubbery modulus of 3 MPa.

Figure 3-15. Schematic of SRT semicrystalline structure composed of randomly oriented β-sheets (AVSTH-rich segments) in a disordered matrix (GLY-rich segments).

Figure 4-1. Scalable production of structural proteins. a) Economy of scales for polymeric materials. The production cost of recombinant protein-based polymers decreases with the production scale up, and this can constitute a cheaper and/or more environmentally friendly alternative for other material types. a) Bioreactor size as a function of structural protein production.

Figure 4-2. Extraction and expression of SRT proteins. Native SRT protein complexes are extracted directly from squid’s suction cups (upper arrow). Biosynthetic routes (lower arrow) are used to obtain the recombinant and de novo designed proteins with particular molecular weights and sequence. The identified sequences of the protein of interest are produced in the chosen type of expression system, i.e. bacteria or yeast, using genetic engineering toolbox.

Figure 4-3. Native vs. recombinant SRT proteins. a) European common squid (*Loligo vulgaris*) and its native SRT (inset). b) SDS-PAGE of native SRT (group of proteins in the 15 – 60 kDa range) and recombinant Lv18 SRT protein (single protein of 18 kDa). c) amino acid sequence of Lv18 SRT protein. Crystal-forming segments (ATVSH-rich) and disordered segments (GLY-rich) are marked in blue and yellow respectively, separated by proline residues.
Figure 4-4. Bacterial expression of recombinant SRT protein. a) E. coli cells that have recombinant SRT protein (inset shows freeze-dried bacterial cells). b) FTIR full spectra and (c) amide I and II bands of native, recombinant (purified) and overexpressed (inside bacterial cells) SRT proteins.

Figure 4-5. Processing of recombinant SRT materials. Recombinant SRT is extracted from bacterial cells and subjected to solution and thermal processing processing steps that alter the nanostructure and properties of the material.

Figure 4-6. a) FTIR spectra of recombinant SRT materials (cast, wash, annealed, compressed, and methanol-treated. b) Amide I band shows a progressive shift towards the β-sheet region (1620 cm⁻¹) with processing.

Figure 4-7. Deconvoluted amide I FTIR spectra. a) Deconvoluted spectra maximum shifts towards the β-sheet region (1620 cm⁻¹) with processing. b) Individual band fitting of cast Lv18 (lowest β-sheet content) and methanol-treated Lv18 (highest β-sheet content). Bands assigned to side chains (sc), β-sheet (β), α-helices (α), and turn (t) structures as described in Table 4-1.

Figure 4-8. X-ray diffraction (XRD) of recombinant SRT. a) Diffraction pattern of compressed Lv18 shows major diffraction peaks corresponding to alternating hydrogen-bonded β-sheet strands [(100), (200)] and residue alignment along the protein backbone (002). b) Diffraction patterns of processed recombinant SRT proteins show an increasing crystallinity with processing steps (highest for compressed and methanol-treated materials).

Figure 4-9. Temperature modulated differential scanning calorimetry (TMDSC) of processed recombinant SRT materials. a) Total and reversing heat flow of cast and compressed Lv18 show a step change around 160 – 190 °C. b) Specific reversing heat capacity C_p of processed Lv18 exhibits a step change that increase in temperature and decrease in magnitude with processing.

Figure 4-10. a) Schematic for the calculation of the glass transition temperature (T_g) and the increase in heat capacity (ΔC_p). b) ΔC_p linearly decreases with β-sheet content in processed Lv18 materials. Similar results from silk fibroin are plotted for comparison. T_g (inset) increases with β-sheet content.

Figure 4-11. Dynamic mechanical analysis (DMA) of processed Lv18 materials. a) Storage and loss moduli drop at the glass transition temperature (T_g). b) T_g from DMA and DSC measurements is plotted as function of β-sheet content (linear increase).

Figure 4-12. Water absorption measurements of processed Lv18 materials. a) TGA shows water absorption up to 45% (weight loss at 100 °C). b) Coupled mass spectrometry reveals major water loss in the 25 – 200 °C range, followed by release of CO₂ caused by thermal degradation.

Figure 4-13. Elastic fixed window scan. a) Normalized Elastic Incoherent Scattering Intensity of dry and hydrated recombinant Lv18 protein as function of temperature. b) Mean Square Displacement (MSD), <x²(T)>, of dry and hydrated recombinant Lg-SRT18 protein as function of temperature.

Figure 4-14. QENS spectra of (a) dry Lv18 (measured in HFBS) and (b) hydrated Lv18 (measured DCS). The gray solid lines are experimental data, including error bars, and the black solid lines represent the elastic component (resolution). Dry and hydrated spectra were fitted with one and two Lorentzian functions respectively (red and blue).

Figure 4-15. QENS analysis of dry and hydrated Lv18 protein. (a) Full width half maximum Γ(Q) of the (Lorentzian functions) as function of Q². (b) Elastic incoherent structure factor EISF of dry and hydrated Lv18 as function of Q. A confined diffusion model (diffusion of disordered chains) and a 3-site jump diffusion model (methyl group rotation) are used to describe the protein dynamics.
Figure 4-16. Mechanical properties of hydrated processed Lv18 materials. a) DMA analysis shows a decrease in Tg due to the plasticization of the amorphous chains. b) Dynamic rheology shows a frequency-independent rubbery modulus around 1 MPa. .......................................................... 91

Figure 5-1. a) Building block sequence inspired in SRT. b) Rolling circle amplification (RCA) strategy to construct SRT-inspired polypeptides with varying number of building blocks. .......................................................... 96

Figure 5-2. SRT-inspired tandem repeat protein mimics. a) Amino acid building block sequence (repeat unit “n”) based on consensus across species. b) Mass spectrometry and SDS-PAGE of SRT-inspired tandem repeat protein mimics comprising 4, 7, 11, and 25 repeat units (15, 25, 42, and 86 kDa in size respectively). .......................................................... 98

Figure 5-3. Structural characterization of SRT-inspired proteins. a) FTIR and b) XRD data show the same secondary structure and crystalline nanodomains (dominated by β-sheet structures). .......................................................... 100

Figure 5-4. Formation of tie chain connecting strands due to β-sheet nanoconfinement. While short polypeptides can be accommodated into one single crystallite (TR-n4), longer polypeptides must accommodate into several crystallites (leaving a connecting strand in every crystallite jump). ............... 102

Figure 5-5. Tensile testing of SRT-inspired proteins. a) Stress-strain measurements show a similar elastic region, but the ultimate strain is higher with increasing number of repeat units n. b) Fractured samples show brittle fracture for low n, whereas crazing is observed for higher n. .......................................................... 104

Figure 5-6. Digital image correlation (DIC) of SRT-inspired proteins. a) Detail of crazing in TR-n11. b) DIC shows full-field strain measurements for TR-n4, n7, and n11 at the locations marked with letters in c). TR-n4 shows a homogeneous strain along the gauge length, while TR-n7 and n11 show localized strain concentration in the crazing areas. .......................................................... 105

Figure 5-7. a) TGA of hydrated SRT-inspired proteins show the same water absorption for all 4 proteins. b) Dynamic rheology of fully hydrated (water content ~20%) SRT-inspired proteins at 70 °C shows an increase in moduli with higher number of repeat units n. .......................................................... 107

Figure 5-8. Analysis of entanglements. a) FTIR of SRT-inspired protein in 8M urea shows the disruption of β-sheet structures (physical cross-links) as the amide III band intensity decreases. b) Dynamic rheology of SRT-inspired proteins in 8M urea shows the same storage modulus for all samples. .......................................................... 108

Figure 5-9. SRT modulus masterplot. a) The contribution of entanglements and tie chains to the total shear modulus is separated as G'ε ≈ G'ε + G'c. While the entanglement modulus is independent of the number of repeat units n, the tie chain network modulus increases with the repeat units n. b) The elastic effectiveness (fraction of elastically effective strands in the network) is calculated as function of the repeat units n and the crystallite size βc. .......................................................... 110

Figure 5-10. Tensile tests of rubbery SRT-inspired tandem repeat proteins. a) Stress-strain curves show strain-hardening at high elongation. Solid lines are experimental data, dashed lines are fits to freely jointed chain model (Eq. 14). The tensile modulus E shows a linear dependence on reciprocal molecular weight (or number of repeat units). b) While high number of repeats are elastomeric (tie chain-rich), low number of repeats are brittle (defect-rich). .......................................................... 112

Figure 5-11. Cyclic tensile testing of SRT-TR proteins. a) Stress-strain curves for several loading cycles with increasing strain. B) Toughness or energy absorbed per cycle shows three regions with increasing strain: (i) homogeneous deformation of random coils), (ii) unraveling of random coils, and (iii) backbone stretching. Ultimate toughness shows a linear dependence with reciprocal molecular weight. ............... 114
Figure 5-12. Strain-induced β-sheet alignment. a) Amide I band (FTIR) shifts towards the β-sheet region with increasing strain. (inset) β-sheet content increase at strains higher than 50%. b) WAXS. c) SAXS reveals the anisotropy and alignment of the β-sheet nanostructure with increasing strain. .......................... 116

Figure 5-13. Mechanics of SRT-inspired protein networks. a) Elastic effectiveness as function of number if repeat units and crystallite size predicts the density of tie chains and defects. b) Polydispersity in native SRT proteins leads to a more efficient network structure with superior mechanical properties......................... 118

Figure 6-1. Elastic fixed window scans of SRT-TR proteins. a) Normalized elastic intensity as function of temperature of dry, D$_2$O-hydrated, and H$_2$O-hydrated TR-n4 and TR-n11 proteins. b) Averaged mean squared displacements $<x^2>$ of H atoms estimated from Debye-Waller factor ........................................... 121

Figure 6-2. QENS analysis of dry TR-n4 and TR-n11 proteins (measured in HFBS instrument). a) Lorentzian full width half maximum (FWHM) is independent of Q (localized motion). b) Elastic incoherent structure factor (EISF) follows a 3-site jump model for methyl group rotation .................. 123

Figure 6-3. QENS analysis of D$_2$O- and H$_2$O-hydration of TR-n4 and TR-n11 proteins (measured in DCS). a) Lorentzian full width half max (FWHM) fitted to the D$_2$O/protein quasielastic spectra (VD model). b) FWHM fitted to the H$_2$O/protein quasielastic spectra (SS model). c) Elastic incoherent structure factor (EISF) of the D$_2$O/protein quasielastic spectra (confined diffusion model). d) Elastic incoherent structure factor (EISF) of the H$_2$O/protein quasielastic spectra (confined diffusion model).................................. 125

Figure 6-4. Proton conductivity in SRT-inspired tandem repeat proteins. a) Characteristic Nyquist plot proton conductivity. b) Proton conductivity of SRT-TR proteins as function of number of tandem repeats 1/n (pH 7)........................................................................................................................................ 127

Figure 6-5. Kinetic isotope effect in proton conductivity in SRT-TR proteins. a) Nyquist plots and b) conductivity of SRT-TR proteins in H$_2$O and D$_2$O (pH 7). ............................................................................................................................... 128

Figure 6-6. a) Proton conductivity as function of water volume fraction φ (pH 7). b) Reduced proton conductivity (proton conductivity $\sigma_{\text{H}}$ normalized to fully swollen proton conductivity $\sigma_{\text{Hn}}$) as function of water volume fraction φ shows agreement with percolation theory and a power law is fitted to the data with $\phi_c$ as percolation threshold and $\alpha$ as power law coefficient. Inset shows a linear dependence of $\alpha$ on 1/n reciprocal repeat unit number. ......................................................................................... 129

Figure 6-7. Semicrystalline morphology in proton conductivity. a) Amide I FTIR spectra of cast, wash, annealed, compressed, and methanol-treated TR-n11 materials shift towards β-sheet region (1620 cm$^{-1}$) with protein processing. b) Proton conductivity increases with the β-sheet content (pH 7). However, the conductivity is constant in the perpendicular strain direction (anisotropic transport). ............................................................... 132

Figure 6-8. SRT-TR proteins as stretchable proton conductors. a) Nyquist plots of stretched TR-n11 protein films show a decrease in impedance with strain. b) Proton conductivity increases with strain due to the alignment of the protein structure (pH 7). However, the conductivity is constant in the perpendicular strain direction (anisotropic transport). ............................................................................. 134

Figure 6-9. Bulk proton conductivity of biological materials. Proton conductivity benchmark, $\sigma_{\text{Hn}}$, for protein-based materials as a function of inverse molecular weight (M$^{-1}$). SRT-TR proteins have the highest bulk proton conductivity among all reported biological materials to date. .................. 135

Figure 6-10. Thermal conductivity of SRT-TR proteins. a) Thermal conductivity of dry TR proteins is constant, but thermal conductivity of hydrated (water content ~40%) TR proteins scales linearly with 1/n. b) Thermal conductivity switch for TR-n11 shows dynamic thermal conductivity as function of hydration over multiple cycles. ................................................................................................. 138
Figure 6-11. Thermal conductivity switching figure of merit $|Z_{TCS}| = |\Delta k/k_{\text{min}}|$ benchmark of diverse state-of-the-art materials.................................................................................................................. 141

Figure 6-12. Master plot for physical properties of SRT and SRT-TR proteins (mechanical, proton conductivity, and thermal conductivity) as function of reciprocal molecular weight or number of repeat units......................................................................................................................................................... 145

Figure 7-1. a) Native and/or b) synthetic SRT proteins are c) dissolved and processed into a variety of d) complex geometries. SRT materials fabricated by solution-based processing: e) thin films are fabricated by drop casting. f) Sponges by particle-assisted casting and subsequent etching of the particles. g) Nanoparticles are prepared by salting out or by addition of a surfactant. h) Coating of textiles and other substrates is performed by dip coating. i) Microfibers are prepared by electrospinning. j) Complex 3D geometries are fabricated by mold casting................................................... 148

Figure 7-2. a) Native and/or b) synthetic SRT proteins are c) thermally processed into a variety of d) complex geometries. SRT materials fabricated by thermal-based processing: e) fibers and rods are fabricated by extrusion, f) complex 3D geometries are fabricated by injection molding, g) thin films are fabricated by hot pressing, h) fibers are fabricated by drawing, i) nanoscale objects are fabricated by template-assisted nanowetting, and j) patterned surfaces are fabricated by compression molding. ....... 150

Figure 7-3. Pictures of lap joints bonded with SRT protein complex in a) wet and b) dry testing conditions. Glass substrates were holding a 300 g weight for both cases. Tested and separated substrate surfaces show c) adhesive and d) cohesive failure mechanisms for wet and dry conditions............... 153

Figure 7-4. a) Normal tensile adhesion strength of SRT protein complex at room temperature as a function of the preload at 45 °C and 70 °C processing temperature. b) Single-lap shear adhesion of SRT in wet and dry conditions................................................................................................................................. 154

Figure 7-5. Underwater adhesion strength for various surfaces: glass, hydrophilic polymeric surfaces (polylactic acid-PLA and polyvinyl acetate-PVA), and plasma treated hydrophobic polymeric surfaces (polystyrene-PS, polydimethyl siloxane-PDMS) at room temperature. .......................................................... 155

Figure 7-6. Normal tensile adhesion strength as a function of a) ion concentration, and b) pH........... 156

Figure 7-7. Self-healing of SRT specimen........................................................................................................... 158

Figure 7-8. Mechanical testing of healed SRT-TR films. a) Stress-strain curves of healed specimens exhibit higher stress and fail at lower strain. b) Healed specimens break at the pristine region, while the healed area withstands the stress........................................................................................................................................ 159

Figure 7-9. Structural changes in the healing process. a) Amide I FTIR spectra and b) XRD spectra reveal a local increase in β-sheet content caused by the healing process................................................................. 160

Figure 7-10. Proton conductivity of healed SRT-TR specimens. Conductivity is slightly higher after healing due to the local increase in crystallinity (pH 7).............................................................................................. 161

Figure 7-11. Self-healing of SRT-coated textiles: a) healing of cut fabrics, b) single fibers, and c) patching........................................................................................................................................................................ 162

Figure 7-12. SRT whispering gallery mode (WGM) resonators. a) flexible array of SRT resonators. b) Transmission spectrum of a SRT WGM resonance around 672 nm with quality factor $Q_0 = 4.45 \times 10^5$. Inset shows SRT toroidal resonator ................................................................. 164

Figure 7-13. SRT photonic devices. a-b) Add-drop filter. SRT resonator with two coupling tapered-fiber waveguides. Transmission spectrum shows a resonant dip in the transmission port and a resonance peak in
the drop port. c) Light transmission through a SRT microfiber as waveguide. d-e) SRT optical switch. Control OFF: control is off-resonance and signal light is coupled to the resonator. Control ON: control is brought to on-resonance, which increases the temperature of the resonator and the signal is shifted to off-resonant. ......................................................................................................................... 165
# List of Tables

| Table 1-1. Major structural elements in natural protein materials and their building blocks | 8 |
| Table 2-1. FTIR amide I band assignment with secondary structure | 23 |
| Table 3-1. Comparative amino acid composition (%) of SRT, elastin, abductin, and resilin | 38 |
| Table 3-2. Secondary structure content of native SRT determined by FTIR | 46 |
| Table 4-1. Secondary structure analysis of processed recombinant SRT materials | 69 |
| Table 4-2. β-sheet dimensions and crystallinity analysis of processed recombinant SRT materials | 72 |
| Table 4-3. Water absorption (determined by TGA-MS) and β-sheet content of processed Lv18 materials | 81 |
| Table 4-4. Calculation of hydrogen fraction involved in the protein dynamics | 89 |
| Table 5-1. Modulus, ultimate strain, ultimate stress, and toughness of SRT-inspired proteins from tensile testing | 104 |
| Table 5-2. SRT-inspired network modulus and elastic effectiveness | 111 |
| Table 6-1. Fit parameters of elastic incoherent structure factor (EISF) from quasielastic neutron scattering measurements of hydrated TR-n4 and TR-n11 proteins | 126 |
| Table 6-2. Proton conductivity of SRT-TR proteins, activation energy and percolation parameters (pH 7) | 130 |
| Table 6-3. Bulk proton conductivity of biological materials | 137 |
| Table 6-4. Thermal conductivity and $C_p$ of SRT-TR proteins | 139 |
| Table 6-5. Thermal conductivity switching benchmark of diverse state-of-the-art materials | 143 |
| Table 7-1. Dry and underwater adhesion strength of SRT and other biological and synthetic adhesives | 157 |
# List of Abbreviations

### Amino acid codes:
- **A** (Ala): alanine
- **C** (Cys): cysteine
- **D** (Asp): aspartic acid
- **E** (Glu): glutamic acid
- **F** (Phe): phenylanine
- **G** (Gly): glycine
- **H** (His): histidine
- **I** (Ile): isoleucine
- **K** (Lys): lysine
- **L** (Leu): leucine
- **M** (Met): methionine
- **N** (Asn): asparagine
- **P** (Pro): prolin
- **Q** (Gln): glutamine
- **R** (Arg): arginine
- **S** (Ser): serine
- **T** (Tr): threonine
- **U** (Sec): selenocysteine
- **V** (Val): valine
- **W** (Trp): tryptophan
- **Y** (Tyr): tyrosine

### Acronyms:
- **ASTM**: American Society for Testing and Materials
- **ATR**: Attenuated Total Reflection
- **CD**: Circular Dichroism
- **DAPI**: 4’,6-diamidino-2-phenylindole
- **DCS**: Disk Chopper Spectrometer
- **DIC**: Digital Image Correlation
- **DMA**: Dynamic Mechanical Analysis
- **DNA**: deoxyribonucleic acid
- **DOPA**: dihydroxyphenylalanine
- **DSC**: Differential Scanning Calorimetry
- **DWF**: Debye-Waller factor
- **ECM**: Extracellular matrix
- **EINS**: Elastic Incoherent Neutron Scattering
- **EISF**: Elastic Incoherent Structure Factor
- **ELP**: Elastin-like proteins
- **FSD**: Fourier self-deconvolution
- **FTIR**: Fourier transform infrared
- **FWHM**: full width half maximum
- **HFBS**: High-Flux Backscattering Spectrometer
- **HFIP**: 1,1,1,3,3,3-Hexafluoro-2-propanol
- **KIE**: kinetic isotope effect
- **LC-MS/MS**: liquid chromatography – tandem mass spectrometry
- **LCST**: lower critical solution temperature
- **LED**: light-emitting diode
- **LV18**: *Loligo vulgaris* 18 kDa recombinant squid ring teeth protein
- **MALDI-TOF**: matrix-assisted laser desorption/ionization time-of-flight
- **meOH**: methanol
- **MS**: mass spectrometry
- **MSD**: mean squared displacement
NGS: Next Generation Sequencing
NMR: nuclear magnetic resonance
PCR: polymerase chain reaction
PD-RCA: Protected Digestion Rolling Circle Amplification
PLA: polylactic acid
PS: polystyrene
PSA: pressure-sensitive adhesive
PVA: polyvinyl acetate
QENS: Quasielastic Neutron Scattering
RCA: Rolling Circle Amplification
RNA: ribonucleic acid
SAXS: Small Angle X-ray Scattering
SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEM: Scanning Electron Microscopy
SRT: squid ring teeth
SRT-TR: squid ring teeth inspired tandem repeat protein
SS: Singwi-Sjölander model
TDTR: time-domain thermoreflectance
TGA: thermogravimetric analysis
TMDSC: temperature modulated differential scanning calorimetry
TR: tandem repeat
VD: Volino-Dianoux model
WAXS: Wide Angle X-ray Scattering
WGM: Whispering Gallery Mode
XRD: X-ray diffraction
List of Symbols

<x^2>: mean squared displacement

\(a\): Miller index

\(a\): \(\alpha\)-helix structure

\(a\): power law coefficient in percolation

\(A\): cross-section area

\(A_0\): elastic area

\(A_i\): quasielastic area

\(b\): Miller index

\(b\): Kuhn length

\(\beta\): \(\beta\)-sheet structure

\(\beta_c\): \(\beta\)-sheet crystallite capacity

\(c\): Miller index

\(C_p\): specific heat capacity

\(C_p(T)^{\text{glassy}}\): specific heat capacity as function of temperature in the glassy state

\(C_p(T)^{\text{rubbery}}\): specific heat capacity as function of temperature in the rubbery state

\(d\): lattice plane separation

\(\delta\): delta function

\(\delta\): phase angle

\(\Delta C_p\): change in specific capacity

\(\Delta G\): change in Gibbs free energy

\(\Delta H_{\text{Hyd}}\): change in enthalpy from hydrogen bonding interactions

\(\Delta H_{\text{vdW}}\): change in enthalpy from Van der Waals interactions

\(\Delta k\): change in thermal conductivity

\(\Delta S_c\): change in entropy of a flexible chain

\(D\): diffusion coefficient

\(DWF\): Debye-Waller factor

\(E'\): tensile storage modulus

\(E''\): tensile loss modulus

\(E_a\): activation energy

EISF: Elastic Incoherent Structure Factor

\(\varepsilon_{\text{eff}}\): elastic effectiveness parameter

\(\varepsilon_{\text{full-field}}\): full-field strain

\(\varepsilon_{\text{max}}\): maximum strain

\(\varepsilon_{\text{yield}}\): yield strain

\(f\): functionality of network junctions

\(f\): force to stretch a flexible chain

\(G'\): shear storage modulus

\(G''\): shear loss modulus

\(G'_e\): shear storage modulus from entanglements

\(G_{ex}\): extrapolated shear storage entanglement modulus

\(G'_{x}\): total shear storage modulus

\(G''_{x}\): shear storage modulus from network strands

\(G_{ex}\): extrapolated shear storage network modulus

\(I_{el}\): elastic scattering intensity

\(j_0\): zero-order Bessel function

\(j_1\): first-order Bessel function

\(k\): thermal conductivity

\(k\): percolation constant

\(k\): Boltzmann constant

\(k_{\text{min}}\): minimum thermal conductivity in thermal switch

\(L\): electrode separation
\( l \): jump distance in jump diffusion model
\( \lambda \): wavelength
\( \lambda \): draw ratio
\( L_i \): Lorentzian function for quasielastic process \( i \)
\( M_{\text{avg}} \): average molecular weight of SRT amino acid
\( M_c \): average molar mass of an entanglement strand
\( M_n \): average molar mass of a network strand
\( M_w \): molecular weight
\( M_x \): average molar mass of a cross-linked strand
\( n \): number of tandem repeat units
\( N \): number of flexible subunits in a chain
\( v \): number density of strands
\( \omega \): frequency
\( p_{\text{amorphous}} \): fraction of hydrogen atoms in the amorphous phase
\( p_{\text{methyl}} \): fraction of hydrogen atoms involved in methyl group rotations
\( p_{\text{mobile}} \): fraction of mobile hydrogen atoms
\( p_{\text{sphere}} \): fraction of hydrogen atoms involved in confined diffusion
\( Q \): scattering vector
\( Q^* \): crossover Q-point in VD model
\( Q_0 \): quality factor of a resonator
\( R \): gas constant
\( R \): instrumental resolution
\( R \): chain end-to-end distance
\( R_0 \): unstretched chain end-to-end distance
\( R_b \): bulk resistance
\( rc \): random coil
\( r_{\text{conf}} \): confinement radius
\( sc \): side chain
\( S_H \): self-dynamic structure factor
\( \sigma \): stress
\( \sigma_{\text{max}} \): maximum stress
\( \sigma_{\text{yield}} \): yield stress
\( \sigma_{H0} \): saturated proton conductivity
\( S_m \): dynamic structure factor
\( t \): turn structure
\( T \): temperature
\( \Gamma \): full width half maximum of a Lorentzian function
\( \tau_0 \): oscillatory residence time in SS model
\( T_d \): denaturation temperature
\( T_{\text{deg}} \): thermal degradation temperature
\( T_g \): glass transition temperature
\( T_m \): melting temperature
\( \theta \): diffraction angle
\( w \): mass fraction
\( X_{\beta} \): crystallinity (i.e., \( \beta \)-sheet fraction)
\( Z' \): real impedance
\( Z'' \): imaginary impedance
\( Z_{\text{TCS}} \): thermal conductivity switching figure of merit
Acknowledgements

Foremost, I wish to express my deep gratitude for the constant guidance and motivation from my thesis advisor, Dr. Meik C. Demirel. I was fortunate to work closely with him in the adventure of squid biomimetic materials from the very early stage of the project, and I have learned much from his expertise and knowledge. His passion for exploring novel research ideas has been and will be a motivating force in my research career.

I want to thank my thesis committee members Dr. Joseph P. Cusumano, Dr. Reginald F. Hamilton, and Dr. Ralph H. Colby for their invaluable guidance, feedback, and recommendations over the thesis work.

I thank my collaborators Dr. Benjamin D. Allen, Dr. Sergio Florez, Dr. Madhusudan Tyagi, Dr. Patrick E. Hopkins, Dr. Michael Hickner, Dr. Metin Sitti, Dr. Sahin K. Ozdemir, and Dr. Veikko Sariola, whose help is deeply appreciated. I also want to thank my fellow lab partners and graduate students, past and present, (Huihun Jung, Sebastián Ferreyro, Robert Shreiner, Yusuke Kikuchi, Benjamin Barbu, Joan Bars, Dr. Asheesh Lanba, Dr. Mert Vural, Dr. Natalia Domeradzka, Dr. Koray Sekeroglu, etc.) for the good times in between experiments. I especially want to thank Huihun Jung, partner in crime both in and out of the lab, whose friendship I will treasure well beyond graduate school, and whose work has had an immense impact on this thesis. I would like to thank Dr. Salvador Borrós and Dr. Robert Texidó, who initiated me in the exciting world of scientific research. I would also like to thank The Pennsylvania State University, the Office of Naval Research, and the Army Research Office, the Materials Research Institute, and the Engineering Science and Mechanics department for supporting my research work.

Last, I want to thank my friends, family, and especially Chris, for their love, encouragement, and never ending patience and support throughout my graduate studies.

From the bottom of my heart, thank you!

“Never surrender when you’re up against the world”

Never surrender (Saxon)
Nothing is art if it does not come from nature

Antoni Gaudí
Chapter 1. General introduction

Biomaterials have been defined as “substances other than foods or drugs contained in therapeutic or diagnostic systems”, as well as “materials composed of biologically derived components”. ¹,² Biomaterials (including metals, ceramics, and polymers) have played an important role in the diagnosis and treatment of diseases and the improvement of healthcare throughout history.³ However, polymer materials became extensively used in healthcare due to the advances in organic chemistry and polymer synthesis in the last decades. To date, biomaterials have had an enormous repercussion in medicine and a direct positive impact in patients’ lives. Controlled drug delivery systems involving polymers and polymer-coated stents, which have been approved in Europe and United States, are used by tens of millions and thousands of people respectively each year.⁴–⁶

Another area of medicine where biomaterials have had an enormous impact over the last decades is tissue engineering. Every year, millions of patients suffer the loss or failure of an organ or tissue as a result of accidents or disease (cardiovascular disease alone accounted for ~30% of deaths worldwide in 2013).⁷ Depending on the severity of the disease, tissue or organ transplantation is a common therapy to treat these patients, but is extremely limited by organ donor shortage. As of August 2017, 116803 patients (men, women, and children) are on the U.S. national transplant waiting list, while only 33611 transplants were performed in 2016. On average, one person is added to the list every ten minutes, and twenty patients die each day waiting for a transplant.⁸ A solution to the donor shortage is engineering artificial organs and tissue from replacement biomaterials. These organs and tissues are usually engineered from a combination of a patient’s own cells with polymer scaffolds.⁹,¹⁰ The scaffold must be designed to mimic the extracellular matrices found in tissues (composed of proteins and polysaccharides), which control the structure and the function of the new engineered tissue.¹¹–¹³ Therefore, the scaffold must meet two critical criteria: (i) it must be biocompatible (i.e., it must be able to exist within the body without causing damage to the neighboring cells or lead to an inflammatory response), and (ii) its mechanical properties must match those of the original extracellular matrix (Figure 1-1) (the mechanical properties regulate cell adhesion, growth, and gene expression).¹⁴,¹⁵
Natural occurring materials such as proteins (built from amino acids) and polysaccharides usually exhibit adequate biocompatibility for tissue engineering applications, while synthetic polymers may cause negative immune responses from the host body.\textsuperscript{2,10} Furthermore, natural biomaterials can be fabricated with controlled mechanical properties to modulate the cell behavior. For this reason, natural materials, and proteins in particular, have received much attention in the last decades as a promising alternative to polymer biomaterials. Hence, there is a growing interest in identifying specific design parameters in protein-based materials that are critical to performance not only in tissue engineering but also in other areas of biomedical research (biophotonics, bioelectronics, etc.).

### 1.1 Protein materials

Proteins are large biomolecules consisting of long chains of amino acids that fold and hierarchically assemble into complex and well-defined structures.\textsuperscript{16,17} Proteins have several advantages over conventional polymers in terms of control, structure, properties, and synthesis (Figure 1-2).
Figure 1-2. Synthetic polymers vs. natural and synthetic proteins. Key advantages of proteins are: precise control over the sequence, precise control over the molecular weight (monodispersity), and the assembly into complex hierarchical structures. Adapted from ref. 18

The amino acid sequence of proteins can be precisely tuned since a defined sequence is genetically encoded in the DNA. This allows absolute control over stereochemistry, sequence, and chain length. Proteins usually have dispersities of 1.0 (i.e., exact molecular weight) and assemble into complex hierarchical structures (defined by the sequence), whereas polymers mainly form random coil or rigid rod conformations and have statistical distributions of monomer sequences and molecular weights. The precise control of the primary amino acid sequence regulates the assembly into the hierarchical structures, and ultimately governs the resulting physical, chemical, and
biological properties of the material (mechanics, stability, activity, etc.). Posttranslational modification of proteins can provide site-specific side chains to proteins.\textsuperscript{19,20} Additionally, proteins present a biocompatible alternative with superior cell-interactive properties and tailored biodegradability, which makes them a material of interest for biomedical applications.

Naturally occurring proteins can be directly extracted from the native organisms. However, recombinant expression in a variety of hosts allows for protein scalable production from renewable sources. Over the past couple of decades, researchers have explored a wide range of expression systems for the high-yield production of proteins such as bacteria,\textsuperscript{21–23} yeast,\textsuperscript{24,25} plants,\textsuperscript{26} mammalian cell lines,\textsuperscript{27} and transgenic organisms.\textsuperscript{28} Genetically modified \textit{Escherichia coli} (\textit{E. coli}) bacteria is the most established suitable host for industrial-scale production due to the availability of expression vectors and well-understood genetics.\textsuperscript{21,29,30} In addition, recombinant expression of engineered artificial genes allows for the biosynthesis of proteins with specified combinations of the 20 natural amino acids and a variety of unnatural amino acids, expanding the possibilities of protein design.\textsuperscript{31,32}

\subsection*{1.2 Structural motifs in proteins}

Nature has evolved many functional materials across the animal and plant kingdom with hierarchical structures across the mesoscale and nanoscale that are built from protein building blocks (Figure 1-3, Table 1-1). Many of the protein-based biological building blocks converged into a same family of structures despite evolving separately. The major structural elements found in protein polymers, namely coiled-coils, β-sheets, and β-turns/spirals, are briefly reviewed below.

**Helical coiled-coil proteins**

Coiled-coils are bundles of α-helices that are twisted into a superhelix, and are usually found in nature in extracellular matrix proteins.\textsuperscript{33–35} α-helix structures (first predicted by Pauling in 1951)\textsuperscript{36} consist of a helical arrangement of the protein backbone, typically with 3.6 amino acid residues per turn of the helix. Each α-helix is stabilized by hydrogen bonding between the backbone amino and carbonyl groups and those in the next turn of the helix, leaving the amino acid side chains in the outer shell of the helix.\textsuperscript{37} Coiled-coil structures are abundant in naturally occurring proteins.
such as collagen and keratin.

Collagen is the most abundant protein in mammals (compromising up to 30% of whole human body protein content) and is found throughout fibrous tissues such as tendons, cartilage, ligaments, and skin. 38 Its major functions are scaffolding, tissue assembly, and repair. 39,40 Fibrillar collagens (types I, II, and III) have [GPX] repeat, where X is usually occupied by proline and/or hydroxyproline. 41,42 The GPX repeat forms left-handed α-helices that intertwine into right-handed triple helices. 39,40 Due to the fact that natural collagen can cause immunological side effects, recombinant collagen has attracted interest because it can include posttranslational modifications (hydroxyproline). 43 Recombinant hydroxyproline collagen has been synthetized for tissue engineering (corneal substitution, cartilage replacement) as well as biosensing and therapeutics applications. 44–46

Keratin is another helix-structured protein, which forms filaments that can be found in epithelial and epidermal appendages such as hair, nails, horns, hooves, wool, and skin. 47 Due to its high sulfur content used to cross-link the coils, keratin is highly insoluble and mechanically strong, contributing to waterproofing and strengthening of hair and epidermal tissues. 48 α-keratins have a repeating heptapeptide sequence α-[X1X2X3X4X5X6X7] that form right handed α-helices dimers. 48 Within the repeat unit, the first, fourth, fifth, and seventh positions are located at the hydrophobic interface between two α-helices, while the second, third, and sixth positions are exposed to the outside environment. The first and fourth amino acids of the heptapeptide are nonpolar (usually occupied by leucine, hence the name “leucine zippers”). 49 The first and fourth positions are especially important since they form the hydrophobic plane along each helix, and they dominate the inter-helical hydrophobic interactions. 48 The hydrophobic planes align between helices to form dimers, which are further stabilized by hydrogen bonding and crosslinking of cysteine residues via disulfide bonds. 47,48,50 Common heptapeptide units such as EVSALEK, KVSALKE, EIAALEK, KIAALKE, VALEKE, and VAALKEK have been used as supramolecular cross-linkers in keratin-inspired coiled-coil protein-based materials. 48

The hierarchical assembly of coiled-coil domains has been explored in the development of biomedical hydrogels. Since the aggregation of coils is driven by hydrophobic inter-helical interactions, a variety of stimuli can disrupt the association and trigger stimuli-responsive behaviors: temperature, ionic strength, pH, and denaturing buffers. 51,52 In addition, the mechanical
properties and association kinetics can be tailored by adjusting the amino acid composition of the heptapeptides (different side chains protruding from the helix). The control of the association/dissociation of coiled domains has led to shear thinning and self-healing protein-materials, which have found use as injectable biomedical hydrogels.

Figure 1-3. Description of major natural protein materials (resilin, collagen, silk, elastin, keratin, and SRT), their representative structure and sequence structural motifs. Adapted from ref.16

**β-turn/β-spiral elastic proteins**

Most elastic proteins are intrinsically disordered but contain a high fraction of β-turns and polyproline structures. β-turns are small secondary structures involving four amino acids that cause a change in direction of the backbone chain. The first and fourth amino acid of the group form intramolecular hydrogen bonding between their carbonyl and amino groups. As a rule of thumb, glycine and proline amino acids provide flexibility to the backbone chain since they prevent folding. Glycine has no side chain, and therefore the entropic cost of folding is usually too
elevated to partake in ordered structures. On the other hand, proline residues are not flexible due to the pyrrolidine group directly bonded to the backbone, and therefore introduce kinks in the backbone chain and disrupt the association into ordered structures.\textsuperscript{37,62} Elastin, which is found in the extracellular matrix and connective tissue (especially in the skin), is composed of water-soluble monomers that aggregate into non-soluble constructs. It has a common hydrophobic domain VPGVG that exhibits a lower critical solution temperature (LCST). Above this temperature, the hydrophobic domains interact and aggregate into β-turn structures separating from the soluble phase.\textsuperscript{63} Additionally, elastin has lysine residues that, after posttranslational modification into allysine, chemically cross-link the hydrophobic domains yielding non-soluble stretchable elastin.\textsuperscript{64,65} The ability to control and modify specific amino acid residues along the chain provides control over the overall hydrophobicity and aggregation kinetics, yielding thermoresponsive elastic materials. Hence, elastin-like proteins (ELPs) that mostly derive from the VPGVG repeat have been used in drug delivery of pharmaceuticals, tissue engineering, biosensing, and protein purification.\textsuperscript{66–68}

Resilin is another elastic protein with high content of β-turn and β-spiral structures. In nature, it is found in the wing hinge, jumping pads, and vocal cords of some insects.\textsuperscript{69,70} These high frequency functions require very elastic and resilient materials (resilin has up to 95% resilience).\textsuperscript{70} Resilin proteins have three main components that function cooperatively as an energy storage/release mechanism: exon I (water-lubricated elastic domain), exon II (cross-linked to chitosan frame), and exon III (energy-storing component).\textsuperscript{70} Resilin has GGRPSDSYGAPGGGN hydrophilic repeats (glycine and proline provide chain flexibility) that are stabilized via dityrosine cross-linking.\textsuperscript{70,71} Recombinant resilin has been synthetized from exons and consensus repeats, and dityrosine cross-linking has been achieved through enzymatic chemistry and photo-cross-linking.\textsuperscript{72} Synthetic resilin-like proteins have been used in tissue engineering s degradable scaffolds with cell-binding domains.\textsuperscript{73}

Flagelliform silk, which is the silk connecting the lines of a spider web and absorbs the energy of impacting prey, is elastic due to a high content of β-turns and β-spirals.\textsuperscript{74} 90% of flagelliform silk is composed of GPGGX motifs (common β-turn motif) that can be cross-linked via disulfide bonds through incorporating cysteine residues.\textsuperscript{75}
Table 1-1. Major structural elements in natural protein materials and their building blocks

<table>
<thead>
<tr>
<th>Structure</th>
<th>Protein</th>
<th>Origin</th>
<th>Building block</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helical coiled-coils</td>
<td>collagen$^{39,40,76}$</td>
<td>tendon, cartilage, ligaments, ECM</td>
<td>[GPX]</td>
</tr>
<tr>
<td></td>
<td>α-keratin$^{48,50}$</td>
<td>hair, nails, wool</td>
<td>[X1X2X3X4X5X6X7]</td>
</tr>
<tr>
<td>β-sheets</td>
<td>silk fibroin$^{77-79}$</td>
<td>silkworm cocoon</td>
<td>[GAGAGS]</td>
</tr>
<tr>
<td></td>
<td>dragline silk$^{79,81}$</td>
<td>spiderwebs</td>
<td>[AAAAAAAA]</td>
</tr>
<tr>
<td></td>
<td>SRT$^{82-85}$</td>
<td>squid suction cups</td>
<td>[AAASVSTVHHP], [GLY]</td>
</tr>
<tr>
<td></td>
<td>curli$^{86-88}$</td>
<td>biofilms, amyloid fibers</td>
<td>[S(X)5QXGXGNXA(X)3Q]</td>
</tr>
<tr>
<td>β-turns/β-spirals</td>
<td>elastin$^{61,63,89,90}$</td>
<td>skin, connective tissue, ECM</td>
<td>[VPGVG], [AAA]4KAAK</td>
</tr>
<tr>
<td></td>
<td>resilin$^{70,71,91,92}$</td>
<td>insect's wings, jumping pads</td>
<td>[GGRPSDYGAPGGGN]</td>
</tr>
<tr>
<td></td>
<td>flagelliform silk$^{74,93}$</td>
<td>spiderwebs</td>
<td>[GPGGX]</td>
</tr>
</tbody>
</table>

**β-sheet-structured proteins**

β-sheet structures are formed by laterally-connected strands of protein with hydrogen bonding interaction between the backbone carbonyl oxygen and the amino hydrogen atoms, and provide stability and mechanical strength through strong intermolecular interactions. Typically, multiple β strands can be arranged into an extensive hydrogen bonding network with their neighboring strands, forming crystal-like domains in the protein matrix. Since peptide chains have directionality depending on their N-terminus and C-terminus, β-sheet secondary structures have directionality too depending on the hydrogen bonding pattern. In an antiparallel structure, the neighboring β-strands alternate directions (e.g., N-terminus of a strand is adjacent to the C-terminus of the next). Allowing the inter-strand hydrogen bonding between carbonyl and amine groups to be planar, this β-sheet configuration is the most stable. Inversely, a configuration wherein parallel β-sheets are formed by β strands oriented in the same direction (e.g., N-terminus adjacent to another N-terminus) disallows planar inter-strand hydrogen bonding and is, thus, less stable than the antiparallel configuration.$^{37}$

Silk is the most extensively studied β-sheet-structured protein. Spun by a variety of insects, it serves as predatory and protective material, with tensile strength (~700 MPa and ~1 GPa for
Bombyx mori silkworm and Araneus diadematus spider silks respectively) and toughness (~160 MJ m⁻³) surpassing those of high-end synthetic polymers such as Kevlar. Silkworm silk fibroin consists of heavy chain and light chain, which are bound through disulfide bridges and glycoproteins. The heavy chain consists of GAGAGS hydrophobic motifs that associate into stiff pleated β-sheets, while the hydrophilic light chain provides flexibility. Spider silk is composed of several types of silk such as dragline silk (main frame of spider webs) and flagelliform silk (elastic connecting silk, rich in β-turns and spirals as described previously). Dragline silk (spun by the major ampullate gland) contains polyalanine and GA repeats that form pleated β-sheets, and helical and turn domains that provide elasticity. The hydrophobic interactions in the polyalanine domains drive the formation of β-sheets, and govern the semicrystalline morphology and mechanical properties of the material. Silkworm silk is usually obtained directed from the cocoons, which allows for a scalable production, and it is used in commercially available products (textiles, paper, wound dressings) as well as in biomedical applications (drug delivery, tissue repair, biophotonics). Spider silks are usually recombinantly expressed from silk-derived repeats due to the fact that spiders are difficult to farm and direct expression of the full protein is difficult due to its large size, and it is used as a biomaterial.

Curli proteins are β-sheet-rich proteins that are found in amyloid fibers and in E. coli and Salmonella biofilms. Amyloid fibers have recently received significant research efforts in order to understand their aggregation mechanism and their role in neurodegenerative diseases such as Huntington’s, Parkinson’s, and Alzheimer’s diseases. The core of amyloid fibers has S(X)₃QXGXGNXA(X)₃Q repeating motifs that aggregate into cross-β-sheet structures (i.e., β-sheet-turn-β-sheet). Synthetic curli proteins have been fused with designer peptides for the development of functional biofilms with site-specific binding, abiotic, adhesive properties.

**Chemically cross-linked proteins**

In addition to the discussed secondary structural motifs in proteins, chemical cross-linking of proteins is another approach to the formation of mechanically stable protein materials both in naturally occurring proteins and synthetic recombinant proteins. Usually, secondary structure elements (e.g., coiled-coils and β-sheets) act as physical cross-linkers via hydrogen bonding and hydrophobic interactions, offering a reversible and stimuli-responsive aggregation mechanism. On the other hand, chemical cross-linking creates intermolecular covalent bonds which are
generally more stable than physical cross-links. The chemical cross-linking of proteins usually involves side chains that are especially reactive, and are used as cross-linking agents to bind with other specific residues. Hence, the chemical cross-linking of a protein is heavily sequence-dependent, and offer several cross-linking chemistries that apply to specific residues.

**Cross-linking through tyrosine residues**

The phenol functionality of tyrosine residues allows for the cross-linking of proteins via the formation of dityrosine and trityrosine adducts. This cross-linking mechanism is the basis of several elastic biological materials, including resilin (found in the wing's hinge and jumping pads of insects). In nature, resilin is cross-linked *in vivo* through an enzyme-mediated (peroxidases) reaction. This strategy has been mimicked *in vitro* with horseradish peroxidase (single chain β-type hemoprotein), which catalyzes the conjugation of phenol and aniline derivatives in the presence of H₂O₂. Additionally, tyrosine residues can be photo-cross-linked with visible light with ruthenium (II) tris-bipyridyl dication [Ru(bpy)₃]²⁺ and ammoniumpersulfate, yielding resilin-based materials with higher resilience than polybutadiene rubber.

**Cross-linking through lysine residues**

Lysine residues enable the cross-linking through primary and secondary amine groups, which naturally occurs in elastin proteins. In order to replicate this strategy, amine-reactive cross-linkers such as β-[tris(hydroxymethyl) phosphine] propionic acid (THPP) and tetrakis(hydroxymethyl) phosphonium chloride (THPC) have been used in condensation reactions in aqueous solutions and at physiological pH to cross-link amine groups. Due to the fact that this reaction is cytocompatible, it has been successfully used to cross-link biologically compatible elastin- and resilin-based hydrogels for tissue engineering and cell encapsulation.

**Cross-linking through cysteine residues**

The presence of cysteine residues (containing sulfhydryl groups) allows for the formation of intermolecular disulfide bonds. This is observed in naturally occurring proteins such as keratin, where the disulfide bonds stabilize the coiled coils. Additionally, thiol-ene chemistry enables the cross-linking of engineered proteins with periodic cysteine residues in their sequence, as well the cross-linking of protein-polymer hybrids. The use of protein-polymer hybrids opens
many design and structural possibilities for multifunctional hydrogels with programmable properties and controlled aggregation kinetics.53

Cross-linking through other residues

Other cross-linking strategies involve certain peptide tags that specifically bind to specific target domains, such as the SpyCatcher/SpyTag mechanism. The immunoglobulin-like collagen adhesion domain 2 (CnB2) can be split into a peptide tag (SpyTag, with 13 amino acid residues) and a protein target fragment (SpyCatcher, 138 amino acid residues) that reconstitute into a stable isopeptide bond under physiological conditions.120,121 These isopeptides have been used as covalent cross-links between different proteins since SpyCatcher and SpyTag fragment can be included in the amino acid sequence of engineered proteins, yielding recombinant protein-based hydrogels with incorporated functional sequences for cell adhesion.122,123

Many of these building blocks have been adapted to short amino acid sequences ( minimal consensus repeats) that can be incorporated into repeat proteins and provide the function of the native system to engineered protein materials.124 Furthermore, several structural motifs can be merged into a single large gene combining different protein building blocks.125,126 This allows to combine different functions in the same molecule analogous to block copolymers, opening many possibilities for the design of engineered protein materials.127,128

1.3 Cephalopod-derived biological materials

In the search for high-strength materials, mankind has used biological materials from the dawn of civilization (since the use of animal teeth, antlers and skin for hunting and protection) to modern days. More recently, biological materials have attracted the attention of scientists who have looked for inspiration in the animal and plant kingdom for the development of advanced materials (composite materials inspired in mollusk shells, ceramics inspired in bone, or artificial skin).129 Despite the extraordinary progress in the field over the last decades, the world of underwater creatures still remains a mystery. Aquatic animals often hide in the depths of seas and oceans, and the discovery of new species of deep-sea animals continues to this day.130 The study of marine life is often driven by the search for biological materials that provide the fundamental science, the
design rules and the inspiration to develop advanced functional materials and devices. For example, the study of the mussel adhesive plaque led to dopamine based surgical adhesives,\textsuperscript{131} the study of mollusk egg capsules led to the design of high deformation and shock absorbing bioelastomers,\textsuperscript{132} the characterization of the dactyl club of the peacock mantis shrimp provides design strategies for impact resistant composite materials,\textsuperscript{133} and the study of photosensitive systems in echinoderms set the groundwork for the fabrication of microlens arrays through biomineralization approaches.\textsuperscript{134,135}

Figure 1-4. Anatomy of European common squid (\textit{Loligo vulgaris}) and its advanced biological materials.

Cephalopods, and squids in particular, present some of the most extraordinary adaptation and survival strategies found in nature.\textsuperscript{1} Squids have developed several features that allow them to be very efficient predators and have attracted scientists for over a century (Figure 1-4). For example, squids possess a highly sophisticated nervous system with a giant axon (up to 1 mm in diameter) that controls the water jet propulsion system, whose comprehensive study revolutionized the field of neurology.\textsuperscript{136,137} Squids also have a reflective tissue (based on chromatophores and reflectin protein) that is used for camouflage and behavioral displays, allowing squids to become invisible to both predator and prey on command that has inspired the development of invisibility stickers for defense applications.\textsuperscript{138–140} The eyes of squids (and other cephalopods) are capable of distinguishing polarized light due to the orthogonal
organization of the photoreceptors, which give squids an additional advantage in prey detection, navigation and communication among color-shifting cephalopods.\textsuperscript{141,142} The gladius or pen is the major hard component of the squid’s anatomy and it is composed mainly of chitin.\textsuperscript{143} Since squids are invertebrates, the gladius provides a site for muscle attachment and mechanical support to the squid’s mantle and the jet propulsion swimming mechanism. The study of gladii in squids and other chitin-rich materials such as the shells of crustaceans initiated the use of chitin-based materials as a renewable bioplastic for commercial applications.\textsuperscript{144,145} The squid beak is one of the hardest and stiffest organic materials found in nature due to its composite structure of histidine-rich protein matrix and chitin fibers which are mechanically stabilized by L-3,4-dihydroxyphenylalanine (dopa)-histidine crosslinkers.\textsuperscript{146} The comprehensive study of the beak structure provided scientists with molecular assembly strategies to assemble mechanically mismatched materials (i.e. soft and hard) into graded composite materials.\textsuperscript{147}

Figure 1-5. Squid ring teeth (SRT): a) European common squid (\textit{Loligo vulgaris}) is caught in the Mediterranean Sea near Tarragona. b) SRT are located in the suction cups along the arms and tentacles. c) Electron microscopy of the suction cups reveals a columnar epithelium (inset). d) Suction cup tissue is characterized by targeting F-actin with phalloidin dyes and nuclei with DAPI staining.
As another example of squid’s predatory arsenal (and the focus of this thesis), squid have evolved to have teeth inside their suckers which serve to attach the squid to its prey strongly (Figure 1-5). Typically, the muscles on the squid limbs generate large pressure differences at the suckers for holding a wide diversity of objects, and the teeth contribute to a strong grasp. The rings are composed of a broad cylindrical band strengthened by an external ridge that separated two muscles in the sides of the sucker cup that act as a sphincter and regulate the actuation of the sucker. While the bottom part of the ring remains inside of the sucker cup, the free edge of the ring is populated by several rows of conical teeth that increase in height from within outward. The teeth and the ring structure are secreted by columnar epithelial cells in the sides of the sucker cups.

These squid sucker ring teeth (SRT) are solely composed of a group of recently discovered proteins that provide mechanical properties that exceed those of conventional synthetic polymers. SRT proteins vary in molecular weight, but are all composed of repetitive motifs with alternating flexible disordered segments and crystal-forming. This thesis dissertation will investigate the amino acid sequence, the nanostructure, and the properties of SRT proteins for the first time. Specifically, the presented work will focus on structural proteins with tandem repeat units in their sequence (i.e., repetition of alternating building blocks) that are physically cross-linked via β-sheet structures inspired in SRT proteins.

1.4 Scope and outline of the thesis

The objective of this thesis is the understanding of sequence, structure, and property relationship in SRT-inspired repetitive protein networks for the first time. Specifically, this work focuses on structural proteins with tandem repeat units in their sequence (i.e., repetition of alternating building blocks) that are physically cross-linked via β-sheet structures. The research work presented here is based on the hypothesis that, in these systems, increasing the number of building blocks in the polypeptide chain decreases the protein network defects and improves the material properties. Hence, the amino acid sequence, nanostructure, and properties (thermal, mechanical, and conducting) of tandem repeat SRT proteins are examined. This approach is applied not only to native SRT proteins but also to SRT-inspired synthetic polypeptides that allow for a fine control
of the network morphology and properties. The results provided in this work establish a clear dependence between the structural motifs of SRT’s amino acid sequence, the number of repetitive building blocks, the network morphology, and the material properties, providing a refined control over SRT materials and introducing design rules in the optimization of protein-based materials.

This thesis dissertation is outlined below (Figure 1-6):

Chapter 1 serves as a general introduction to protein-based materials, protein structural motifs, cephalopod-derived materials and the discovery of SRT. Chapter 2 provides a summary of the experimental approach and describes the methodology used in the characterization and fabrication of SRT materials, including structural analysis, protein dynamics analysis, and measurement of the thermal, mechanical, and conducting properties.

Chapter 3 investigates the amino acid composition, sequence, and molecular weight of a native SRT protein complex directly extracted from Loligo vulgaris squid species. Structural analysis of the native protein complex by spectroscopic techniques revealed a semicrystalline morphology. The presence of repetitive units in the native sequence driving the self-assembly of crystalline domains is discussed, as well as its impact on thermal stability and mechanical properties. However, the analysis of the structure and properties is limited due to the lack of control over the sequence of natural proteins.

Chapter 4 investigates the structure, dynamics, and properties of a recombinant SRT protein with a defined amino acid primary sequence. The semicrystalline nanostructure of the protein is characterized in detail, and the evolution of the protein morphology with processing conditions is analyzed (yielding protein materials with variable crystallinity). The glass transition of the semicrystalline recombinant SRT protein is studied by thermal and mechanical characterization methods, and the plasticization and hydration dynamics of the protein are studied by quasielastic neutron scattering. As result, a relationship between the varying semicrystalline nanostructure and the thermal properties (i.e., glass transition) is established.
Chapter 5 and Chapter 6 investigate the structure and properties of SRT-inspired synthetic polypeptides with a controlled and repetitive segmented sequence. These SRT mimics share the same basic building blocks but vary in length (i.e., varying total number of building blocks per protein chain). Due to crystal nanoconfinement, SRT mimics form effective or defective protein network structures that can be controlled with the number of total building blocks. In Chapter 5, the network defects are quantified, the glassy and rubbery mechanics of SRT mimics are investigated, and a simple network model for SRT-based materials is proposed. In Chapter 6, the chain dynamics and transport properties (proton and thermal conductivity) of SRT mimics are investigated. The findings presented in these two chapters establish a clear dependence between network defects and overall properties of SRT, and show a possible rationale for optimization of SRT-based materials both in nature and in engineered proteins.

Chapter 7 describes a variety of applications for SRT-based materials. First, the fabrication of SRT materials is discussed, including several solution and thermal-based processing methods. Second, the underwater pressure-sensitive adhesive properties of SRT are reported, and the potential use
of SRT as a bioadhesive is discussed. Third, the self-healing properties of SRT are analyzed, showing promise as self-healing materials and protective coatings in textile applications. Last, SRT-based photonic devices are fabricated, including optical waveguides, filters, and switches based on whispering gallery mode resonators.

Chapter 8 presents the conclusions of this thesis dissertation, and briefly discusses the future research follow-up work including the development of SRT-based devices, SRT-based nanocomposites, and sequence-controlled assembly of SRT proteins.
Chapter 2. Experimental approach

This chapter describes the experimental approach of the thesis, including squid capture, protein analysis, structural characterization (infrared spectroscopy, X-ray diffraction), thermal analysis (calorimetry and gravimetry), mechanical analysis (dynamic rheology, dynamic mechanical analysis, tensile tests, digital image correlation, and adhesion tests), protein dynamics (neutron spectroscopy), and conductivity measurements (impedance and thermoreflectance).

2.1 Proteomics

**Squid and SRT collection:** European common squids (*Loligo vulgaris*) were caught from the coast of Tarragona (Spain). Longfin inshore squids (*Loligo pealei*) were obtained from the Marine Biological Laboratory in Woods Hole (MA, USA). Northern shortfin squids (*Illex illecebrosus*) were obtained from Lund’s Fisheries Inc. in Cape May (NJ, USA). Argentine shortfin squids (*Illex argenties*) were caught in Mar de Plata (Argentina). Bigfin reef squids (*Sepioteuthis lessoniana*) and Japanese flying squids (*Todarodes pacificus*) were caught in Taipei (Taiwan). Suction cups were stored in RNALater solution (Life Technologies, CA) immediately after capture, and stored in −80 °C refrigerator for subsequent RNA Isolation Protocol. The squid ring teeth (SRT) were removed from the tentacles, washed in deionized water and ethanol mixture (70:30 ratio v/v), and vacuum dried in a desiccator.

**Protein size determination by electrophoresis:** 0.2 mg of SRT is dissolved in 1 mL of 5% acetic acid/2 M urea solution, and subjected to sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) for protein separation. The protein gels were stained with Coomassie blue dye.

**Protein size determination by Maldi-TOF:** 0.25 mg/ml SRT protein solution is prepared in dilute HCl solution at pH=3. Sample preparation, data analysis, and processing are performed according to the following details. Protein samples were mixed 1:1 with a saturated solution of sinapinic acid in 50% aqueous acetonitrile containing 0.1% trifluoroacetic acid, 1 microliter of this solution was applied to a stainless steel target plate, and the samples were allowed to crystallize
under ambient conditions. Positive-ion mass spectra were acquired on a Bruker Ultraflextreme MALDI-TOF-TOF instrument operating in linear mode at Penn State Proteomic Center. A factory default method was modified to acquire data over 8,000 – 60,000 m/z range and 15,000 – 70,000 m/z range. Mass spectra were calibrated externally using linear calibration curve and 4 calibrants (Sigma): insulin, cytochrome C ([M+H]+ and [M+2H]+), ubiquitin, and myoglobin ([M+H]+ and [2M+H]+). Spectra were smoothed once using Savitzky-Golay algorithm, the baseline subtraction algorithm was TopHat, and the mass lists were generated based on the centroid peak detection algorithm.

**RNA isolation protocol:** Suction cups were defrosted and RNALater solution was decanted. Any remaining SRT in the suction cups was removed to reduce protein contamination. The tissues samples were homogenized by slicing them into smaller pieces with a clean razor inside a biological hood and resuspended in RNALater solution. The homogenized tissue was disrupted by adding 600 μL of RLT Plus lysis buffer (Qiagen), and kept in room temperature for 2 min (or until the solution color turned yellow) in eppendorf tubes. The solution was centrifuged for 3 min at high speed. For the DNA elimination, the lysate supernatant from last step was transferred to a DNA Eliminator spin column (Qiagen, RNAeasy Mini Kit), and centrifuged for 30 s at 10 000 rpm. 600 μL of 70% ethanol solution was added to the flow through and mixed well by pipetting without centrifugation. For RNA filtering, the solution was transferred to a RNAeasy spin column (Qiagen, RNAeasy Mini Kit) and centrifuged for 15 s at 1000 rpm. Three wash buffer steps were performed according to the Mini Kit user manual. Finally, RNA extraction was completed by adding 50 μL of RNase-free water directly to the spin column membrane, and by collecting the solution via centrifugation for 1 minute at 10 000 rpm. The solution was stored in the fridge for sequencing.

**Sequencing and cloning:** SRT genome sequencing was performed at Penn State Genomics Core Facility and the Bioinformatics Consulting Center. High-throughput sequencing produced paired end reads with read lengths of at least 250 base pairs, which were used to assemble a preliminary transcriptome. The sequence data has been deposited in the National Center for Biotechnology Information BioProject database (PRJNA320263). Peptide sequences from the whole-SRT protein complex were sequenced using liquid chromatography-tandem mass spectrometry (LC-MS/MS) to provide N-terminal biased partial protein sequences that were matched against the putative
transcripts. Details of the iterative bioinformatics approach can be found elsewhere. The open reading frame of the LvSRT protein was cloned into XhoI/BamHI site of the pET-14b vector (Novagen). Digestions of the PCR products and the vector were carried out at 37°C for 3 hours, and purified using Qiagen PCR clean up. Purified PCR fragments were ligated using T4 ligase (New England Biolabs) with a ratio of 3:1 insert to vector. The gene was cloned omitting the predicted signaling peptide and in framed with the his-tag found in the pET-14b vector.

**Construction of SRT tandem repeat (TR) templates by protected digestion rolling circle amplification (PD-RCA):** SRT-TR templates were constructed by Dr. Benjamin D. Allen and Huihun Jung. The process is summarized here, but the details can be found published elsewhere. A 111-bp gene fragment encoding an 18-aa amorphous region and an 11-aa crystalline region was synthesized by Genewiz, cloned into plasmid pCR-Blunt by standard methods, and verified by Sanger sequencing. The insert contains five restriction sites to enable the PDRCA. A circular, nicked version of the insert sequence was prepared as a template for RCA, and the nicking enzyme reaction was heat-inactivated for 20 min at 80°C. 1.5 μL of the heat-inactivated nicking reaction was used as the template in a 10-μL RCA reaction with 1× New England Biolabs (NEB) phi29 polymerase buffer, 1 μg BSA, 1 mM dATP, 1 mM dGTP, 1 mM dTTP, 0.5 mM dCTP, 0.5 mM 5-methyl-dCTP, and 2.5 U NEB phi29 polymerase. The reaction was incubated at 30°C for 24 h and then heat-inactivated for 10 min at 65°C. The heat-inactivated RCA reaction was sequentially digested with ApaI and Acc65I, yielding TRs of various sizes because the random protection of their recognition sites by 5mC (Fig. 2B). TR fragments between 500 and 1,500 bp were isolated from a 1% agarose-TAE gel and purified with an Omega Bio-Tek E.Z.N.A Gel Extraction Kit. The purified fragments were cloned through the Acc65I and ApaI sites into the ORF of an expression vector prepared by site-directed mutagenesis of pET14b. Colony PCR was used to screen for clones with inserts of the desired sizes; diagnostic digestion and Sanger sequencing confirmed the lengths and compositions of the clones after plasmid isolation.

**Protein expression:** SRT recombinant and tandem repeat proteins were expressed by Huihun Jung. A single colony was inoculated and grown overnight in 5 mL LB with ampicillin (100 μg/mL). The overnight culture was scaled up to 2 L (i.e., four by 500 mL LB media) and grown on a shaker at 210 rpm and 37°C for 5 h. When the cultures reached OD600 of 0.7–0.9, isopropyl β-D-1-thiogalactopyranoside was added to the final concentration of 1 mM, and shaking was
continued at 37 °C for 4 h. Then, the cells were pelleted at 21,612 × g for 15 min and stored at −80 °C. After thawing, cell pellets were resuspended in 300 mL lysis buffer (50 mM Tris, pH 7.4, 200 mM NaCl, 1 mM PMSF, and 2 mM EDTA) and lysed using a high-pressure homogenizer. The lysate was pelleted at 29,416 × g for 1 h at 4 °C. The lysed pellet was washed twice with 100 mL urea extraction buffer [100 mM Tris, pH 7.4, 5 mM EDTA, 2 M urea, 2% (vol/vol) Triton X-100] and then washed with 100 mL washing buffer (100 mM Tris, pH 7.4, 5 mM EDTA). Protein collection in the washing step (urea extraction and final wash) was performed by centrifugation at 3,752 × g for 15 min. The resulting recombinant protein pellet was dried with a lyophilizer (FreeZone 6 Plus; Labconco) for 12 h. The final yield of expressed protein was ∼15 mg/1 L bacterial culture.

**Zeta potential measurements:** the zeta potential of SRT films was measured by ZetaSpin. The ZetaSpin method is based on the fact that liquid flow near a solid surface generates surface current that must return as bulk current through the electrolyte in order to complete the circuit. ZetaSpin uses a rotating disk that generates an axisymmetric flow that sweeps positive charge in the diffuse layer outward to the periphery of the disk where it leaves the diffuse layer and flows back through the liquid toward the axis to complete the circuit. The setup consists of a rotating disk (sample is attached to its end) immersed in an electrolyte solution and two reference electrodes (one near the sample on the axis of rotation and the other far away from the sample). When the sample is rotated (controlled externally by a motor), the streaming potential is measured and the zeta potential is calculated based on the theory described elsewhere. Hence, ZetaSpin is useful for surface charge measurements of films and non-soluble bulk materials (in this case, SRT films). Disk-shaped SRT samples were measured in ZetaSpin rotated at 1000 rpm at pH between 3 and 11 (increments of 0.5).
2.2 Structural characterization

**Scanning electron microscopy (SEM):** The surface morphology was characterized by an electron microscope (Philips XL30) at an accelerating voltage of 5 kV (resolution of 3.5 nm at 30 kV). Samples were gold coated (20 Å thick approximately) with a sputter coater (Ladd Research Industries, Model 30800).

**Fourier transform infrared spectroscopy (FTIR):** spectra were collected (Thermo Scientific Nicolet 6700 FT-IR) under attenuated total reflection (diamond crystal) mode using Happ-Genzel apodization with 4 cm⁻¹ resolution from wavenumber 400 to 4000 cm⁻¹. For each spectrum, 256 scans are co-added. Fourier self-deconvolution (FSD) and second derivative of the amide I band (1580-1706 cm⁻¹) was performed by OMNIC software (Thermo Scientific, v7.3). Second derivative was obtained from the original amide I spectra and a nine-point Savitsky-Golay smoothing filter of polynomial degree 5 was applied. FSD was performed with Lorentzian line shape with 25 cm⁻¹ bandwidth and an enhancement factor of 2. The secondary structure content of SRT proteins was calculated from the amide I band (1600-1700 cm⁻¹). Individual bands were fitted to the deconvoluted spectra and were assigned to secondary structural components (Table 2-1). The number and position of the fitted bands were obtained from the second derivative spectra, where the minima in the second derivative spectra indicates the position of a Gaussian band. Curve fitting was performed in OriginPro 8.5 software by fitting of Gaussian functions by a nonlinear least-squares method. First, the initial band positions (taken from the second derivative spectra) were fixed and the width and height were left as free parameter. Then the band positions were allowed to change within a ± 1 cm⁻¹ range using the built-in Levenberg-Marquardt algorithm. The relative areas of the single bands were used in the calculation of the fraction of the secondary structure elements.
Table 2-1. FTIR amide I band assignment with secondary structure

<table>
<thead>
<tr>
<th>Secondary structure</th>
<th>Band position (cm$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>side chains$^{153,155}$</td>
<td>1594-1597</td>
</tr>
<tr>
<td>$\beta$-sheets$^{151-153,156-159}$</td>
<td>1600-1632, 1697-1703</td>
</tr>
<tr>
<td>random coils$^{151,157,158,160,161}$</td>
<td>1640-1655</td>
</tr>
<tr>
<td>$\alpha$-helices$^{157,161}$</td>
<td>1658-1662</td>
</tr>
<tr>
<td>turns$^{151,156,157,160,162}$</td>
<td>1665-1693</td>
</tr>
</tbody>
</table>

**X-ray diffraction (XRD):** wide angle X-ray scattering (WAXS) data were collected in a Rigaku DMAX-Rapid II Microdiffractometer (wavelength $\lambda = 0.154$ nm) using Cu K$\alpha$ source and 30 μm collimator with 10 minute exposure at 50kV and 40mA. The scattering angle $2\Theta$ was collected from $3^\circ$ to $75^\circ$. 2D WAXS diffraction patterns were converted in one-dimensional pattern by integration across all azimuthal angles (avoiding the beam stop). The crystallinity index was calculated as the ratio of the area of crystal peaks to the total area by fitting the Lorentz-corrected WAXS intensity data using Gaussian functions.$^{163}$ The spectra were analyzed with MDI Jade X-ray diffraction software and peak fitting was performed in OriginPro 8.5 software. The major crystalline peaks of interest were at $2\Theta \approx 9.5^\circ$, $19.5^\circ$ and $25^\circ$, which according to Bragg’s Law law ($n\lambda = 2d \sin \Theta$),$^{164}$ correspond to the hydrogen-bond distance between two $\beta$-sheet chains (4.72 Å), the distance between alternating $\beta$-sheet chains (9.44 Å, i.e. unit cell dimension in the hydrogen-bond direction fitting two $\beta$-sheet chains) and the chain length of a single amino acid in an antiparallel $\beta$-sheet structure (3.5 Å, with a two-residue repeat distance of 7.0 Å).$^{165}$ Miller indices were assigned consistent with literature of SRT.$^{84,85,166,167}$ The size of the crystallites were estimated by the analysis of the peak broadening according to Scherrer’s equation, $\Gamma(2\Theta) = k \lambda / (L \cos \Theta)$,$^{168}$ where $\Gamma(2\Theta)$ is the full width at half maximum of the crystalline peak, $k$ is a constant of proportionality (usually 0.9 to 1 for polymers and cubic symmetries), $\lambda$ is the X-ray wavelength, and $L$ is the crystallite size. The small angle X-ray scattering (SAXS) measurements were performed using SAXS/WAXS laboratory beamline (Xeuss 2.0) at the Materials Characterization...
Lab user facility in Penn State. The X-ray source is GeniX3D, which is a compact system of microfocus sealed tube and X-ray optics with wavelength of 1.54 Å. This laboratory beamline is equipped with Pilatus 200K-A detector, and the sample-to-detector distance was 2.5 m, with 20 min of exposure time. The samples were placed into an aluminum holder, with no mica or Kapton windows.

2.3 Thermal analysis

Thermogravimetric analysis – mass spectrometry (TGA-MS): TGA was performed on a TA Instruments Q50 coupled to a Pfeiffer vacuum mass spectrometer in order to measure the protein thermal degradation and water absorption. Samples were heated from 25 °C to 500 °C at a heating rate of 10 °C/min in an argon gas flow (90.0 mL/min). Hydrated samples were immersed in DI water for 2 hours at room temperature before the experiment, and excess surface water was removed before transferring the samples to the instrument. The initial sample weight was in the range between 6.5 mg and 9.0 mg.

Temperature modulated differential scanning calorimetry (TMDSC): TA Instruments Q200 DSC was used for the glass transition measurements of SRT proteins. The instrument was calibrated for specific heat \( C_p \) measurements using Tzero aluminum pans and sapphire standard materials, including cell resistance and capacitance, cell constant and temperature calibration. SRT samples were annealed on a hot plate at 100°C for 30 minutes in order to remove water before the experiment. Samples were heated at 2°C/min from 25°C to 210°C (beginning of thermal degradation) with a modulation period of 60 seconds and a temperature amplitude of 0.318°C. The glass transition was analyzed from the reversing specific heat capacity \( C_p \) data. The increase in heat capacity \( \Delta C_p \) is calculated as the difference between the glassy and rubbery tangent lines:

\[
C_p(T)_{\text{rubbery}} = C_p(T)_{\text{glassy}} + \Delta C_p.
\]

The glass transition temperature is calculated as the step change midpoint:

\[
C_p(T_g) = C_p(T_g)_{\text{glassy}} + \Delta C_p/2.
\]

\( C_p(T)_{\text{glassy}} \) and \( C_p(T)_{\text{rubbery}} \) are the extrapolation of the tangent to the specific reversing \( C_p \) below and above the glass transition respectively.
2.4 Mechanical analysis

**Dynamic mechanical analysis (DMA):** DMA was performed in a TA 800Q DMA instrument, which has a force range of $10^{-4}$ to 18 N, force resolution of $10^{-5}$ N, strain resolution of 1 nm, and a frequency range of 0.01 to 200 Hz. DMA was used for small deformation analysis (oscillatory strain and strain ramp) in order to measure the glassy mechanics and glass transition of SRT proteins. Film-tension and submersion-tension clamps were used for dry and hydrated (saturation) measurements respectively. Sample dimensions were 15 mm × 2.5 mm × 0.2 mm for the dry test and 25 mm × 2.5 mm × 0.2 mm for the hydrated test. Oscillatory temperature sweep experiments were performed at 1 Hz, with amplitude of 2 μm and a rate of 2 °C per minute. Stress–strain experiments were performed with a strain rate of 1% per minute and a preload of 0.01N.

**Digital image correlation (DIC):** In situ full-field deformation measurements of stretched SRT proteins were obtained using DIC together with DMA (Figure 2-1). A speckle pattern was applied to SRT film specimens using an IWATA Micron-CMB airbrush. A thin, uniform white coating of Golden Airbrush Titanium White (#8380) paint was applied on the surface, followed by a speckle pattern of Golden Airbrush Carbon Black (#8040) paint. In situ images of the specimen surface were captured using a Grasshopper GRAS-20S4M/C CCD camera (1600×1200 pixels). Images were captured in parallel to tensile stress measurements at a frequency of 0.5 Hz using Vic-Snap software from Correlated Solutions and were correlated to DMA data capture. The images captured for all the experiments had a resolution of $16 \pm 1$ μm/pixel. The DIC was performed via the software Vic-2D from Correlated Solutions, Inc. The strains in DIC contours are the normal component of strain that coincides with the loading direction, i.e. the longitudinal dimension of the sample. The correlation is performed in an analysis area on the surface referred to as the DIC area. The DIC area spans the entire gage length of the specimens and measures $2 \times 8$ mm$^2$. The DIC area is divided into subsets, which are square regions that are approximately 560 ± 5 μm wide. The correlation algorithm matches pixel intensity maps in a reference image to those in a deformed image in order to determine the displacements and displacements gradients (strains). The correlation parameters step size and filter size can be adjusted to define a virtual strain gage. Thus, the full-field analysis can be imagined as local measurements over many tiny strain gages. For the current analysis, the virtual strain gage size is $400 \pm 5$ μm. The virtual strain gage thus allows the micro-scale full-field strain measurements.
Figure 2-1. Digital image correlation (DIC) setup (a) with DMA and digital camera. b) Dog-bone specimen after speckle pain was applied. c) Close-up of the speckle pattern in the DIC area of interest. d) Pixel histogram of the speckled area. e) Particle size analysis of the speckle pattern.

**Underwater tensile tests:** SRT-TR dog-bone shaped films with 30 μm in thickness, 2 mm in width, and 15 mm in gauge length were cast from SRT/HFIP solution. Specimens were tested in an Instron model 5866 load frame with a 10 N cell inside a 2 gallon custom-built liquid chamber (Figure 2-2). All measurements were performed with the specimens immersed in DI water. Tensile stretching measurements were performed at a rate of 5 mm/min until failure. Cyclic stretching measurements were performed at a rate 5 mm/min to a strain of 0.5, 1, 1.5, 2.5, 5, 7.5, 10, 15, 20, 30, and 40 mm and back to 0 mm in each cycle. Toughness was calculated as the area under the true stress-strain curve for each cycle.
Figure 2-2. Underwater tensile test setup. a) Intron 5866 load frame equipped with a custom-built liquid chamber. b) Detail of the custom-built liquid chamber.

**Rheology:** SRT proteins were compressed in PDMS molds (hydrated, 70 °C) into disk shaped samples of 2 mm in diameter and 1 mm in height. Measurements were performed in a Rheometric Scientific ARES-LS rheometer with 3 mm diameter parallel plate geometry with a 10 mL liquid reservoir in the bottom plate. Samples were adhered to the plates using Click Bond CB200 acrylic adhesive (2 h curing time at room temperature, with modulus in the GPa range). Liquid (ultrapure water) was fed into the reservoir with a peristaltic pump (Ultra Low Flow Mini-pump model 3384, Control Company) with a flow rate of 0.4 mL/min to compensate for evaporation, and the system was equilibrated at 70 °C for 90 minutes. Dynamic frequency scans with 2% strain were performed from 0.001 to 100 rad/s.

In some measurements, 8M urea solution was fed into the liquid cell in order to disrupt hydrogen bonding in the protein network (urea is a known hydrogen bonding disruptor). Hence, H\textsubscript{2}O vs. urea experiments allowed to measure the modulus with and without hydrogen-bonded structural elements (physically cross-linked β-sheet domains vs. random coil chains).

**Adhesion tests:** Tensile and shear tests are performed to estimate the adhesion strength. At least 5 individual tests were performed to ensure statistical significance. Single lap shear adhesion was
measured in a universal testing machine (United Calibration Corp., STM-20) at 1.27 mm/min strain rate (0.05 inch/minute). Adhesive joints were prepared according to ASTM 3163 standard with glass substrates. The adhesive joints were kept underwater initially, and were dried overnight at room temperature for dry adhesion measurement. Single lap shear adhesion was validated using commercially available ethyl cyanoacrylate (“Superglue”) and epoxy (“Loctite 5 minute epoxy”) glues. These samples were cured for 24 hours at room temperature. The adhesion strengths are 3.39 ± 0.60 MPa for ethyl cyanoacrylate and 4.75 ± 1.42 MPa for epoxy. Tensile adhesion strength normal to the substrate surface was determined in a home-built setup with a liquid media reservoir for both underwater and dry adhesion. The glass slides (plain VWR micro slides, 1 mm thick) were pulled apart in the normal direction of the substrate surface (opening mode). Normal tensile adhesion tests at different salt concentration and pH were immersed in the respective solutions. Ultrapure water was titrated with acetic acid and sodium hydroxide to prepare solutions between pH 3 and 12. NaCl and MgCl₂ salts were dissolved in ultrapure water between 0.05 M and 0.5 M (Na⁺ and Mg²⁺) for salt concentration experiments.

### 2.5 Neutron scattering

Neutron scattering consists in the measurement of the neutron intensity scattered into a solid angle after interacting with the sample.¹⁷⁶–¹⁷⁸ Scattered neutron intensity has a coherent and an incoherent contribution. Coherent scattering, used in diffraction experiments where the phase differences between the waves create constructive and destructive interferences at different angles, preserves the relative phases of the scattered waves. The coherent scattering intensity depends on the constructive and destructive interferences amplitudes and is calculated as the sum of the amplitudes. On the other hand, incoherent scattering does not preserve the phase relation between the incident and scattered wave, and the total intensity is calculated by the sum of individual scattered intensities.¹⁷⁹ Thus, coherent scattering contains structural information and the positions of different atoms (collective dynamics), whereas incoherent scattering describes individual dynamics (self-motion of individual atoms). Due to the large incoherent neutron scattering cross section of hydrogen atoms (~ 40 times larger than deuterium and much larger than nitrogen, carbon, and oxygen), incoherent neutron spectroscopy has the ability to selectively probe the self-
dynamics of hydrogen atoms (because the signal from nonexchangeable hydrogen atoms is dominant). Therefore, the use of deuterated water enables the direct observation of the protein dynamics since scattering from solvent is negligible, making QENS an extremely useful tool for the study of hydration dynamics of biological systems (e.g. in myoglobin, lysozyme, RNA, GFP, silk, – casein, etc.).

Depending on the energy exchange between the neutron and the sample, the scattering events can be divided into: (i) elastic scattering, if the exchanged energy is zero or smaller than the energy resolution of the spectrometer; (ii) inelastic scattering, if the energy exchange is not zero and results in the excitation of the sample; (iii) quasielastic neutron scattering (QENS), if the energy exchange is small (slightly higher than the elastic scattering). Hence, QENS is observed as the broadening of the elastic scattering peak, and it arises from slow motions close to zero energy exchange. Quasielastic neutron scattering (QENS) is a spectroscopic technique for studying dynamical modes at molecular level with 1 – 30 Å spatial and pico- to nanosecond temporal resolution. The momentum transfer is defined as \( Q = \frac{4 \pi \sin (\theta / 2)}{\lambda} \), where \( \theta \) is the scattering angle and \( \lambda \) is the wavelength of the incoming neutrons. Due to the small wavelength of neutrons, molecular dynamic processes such as rotations, localized motions, diffusive motions and relaxations can be studied with QENS.

Quasielastic neutron scattering experiments were performed at NIST Center for Neutron Research (NCNR), Gaithersburg MD. Experiments on dry SRT proteins were performed on the high-flux backscattering spectrometer (HFBS), with energy resolution of 1 μeV, dynamic range of ± 15 μeV and Q-range of 0.25 Å\(^{-1}\) – 1.8 Å\(^{-1}\), at 295K. The resolution function was measured at 4K (signal is expected to be completely elastic). Quasielastic experiments on H₂O/D₂O-hydrated SRT-TR polypeptides were performed on the disk chopper time-of-flight spectrometer (DCS), with energy resolution of 64 μeV (wavelength \( \lambda = 6 \) Å), dynamic range of ± 0.5 meV and Q-range of 0.1 Å\(^{-1}\) – 2 Å\(^{-1}\), at 295K. Quasielastic experiments on hydrated Lv18 recombinant SRT protein were performed on DCS, with energy resolution of 22 μeV (wavelength \( \lambda = 9 \) Å), dynamic range of ± 0.35 meV and Q-range of 0.1 Å\(^{-1}\) – 1.3 Å\(^{-1}\), at 295K. Vanadium was used as the resolution function in DCS measurements. Figure 2-3 illustrates the length and timescales available from HFBS and DCS in the selected configurations.
The neutron scattering intensity measured in the quasielastic experiments was converted to the measured dynamic structure factor $S_m(Q, \omega)$, which given an instrumental resolution $R(Q, \omega)$ and the self-dynamic structure factor of hydrogen atoms in the protein $S_H(Q, \omega)$, can be written as:

$$S_m(Q, \omega) = S_H(Q, \omega) \otimes R(Q, \omega)$$  \hspace{1cm} (1)

The self-dynamic structure factor $S_H(Q, \omega)$ can be separated into a purely elastic $A_0(Q)\delta(\omega)$ component and one or more quasi-elastic components centered at $\omega = 0$, $A_i(Q)L_i(\omega)$:

$$S_H(Q, \omega) = DWF \left[ A_0(Q) \cdot \delta(\omega) + \sum_{i=1}^{N} A_i(Q) \cdot L_i(Q, \omega) \right]$$  \hspace{1cm} (2)

$$L_i(Q, \omega) = \frac{1}{\pi} \frac{\Gamma_i(Q)}{\omega^2 + \Gamma_i^2(Q)}$$  \hspace{1cm} (3)

where $\delta(\omega)$ is the delta function (which represents the elastic part of the scattering function), $A_0$ is the elastic component area, $L_i(\omega)$ is a Lorentzian function with full width at half maximum.
(FWHM) $\Gamma_A$, $A_i$ is the quasielastic component area for a particular mode $i$, and DWF is the Debye-Waller factor.\textsuperscript{177,178,193} Therefore, the measured dynamic structure factor can be modeled as a sum of an elastic process and several quasielastic processes (i.e. several types of motion):

$$S_m(Q, \omega) = DWF [A_0(Q) \cdot \delta(\omega) + \sum_{i=1}^n A_i(Q) \cdot L_i(Q, \omega)] \otimes R(Q, \omega)$$  \hspace{0.5cm} (4)

The NIST-developed DAVE software, developed by NCNR,\textsuperscript{194} was used to analyze the data. Lorentzian functions were fit to describe the quasielastic broadening, and the full width at half maximum (FWHM) and the area of the fitted peak were analyzed as function of $Q$ (Figure 2-4).

![Quasielastic neutron scattering spectra of hydrated SRT protein. The graph shows the instrumental elastic resolution (solid black line), two Lorentzian functions (red and blue solid lines) used to describe the dynamic processes, the experimental data (yellow dots), and the total fit (dashed line).](image)

Figure 2-4. Quasielastic neutron scattering spectra of hydrated SRT protein. The graph shows the instrumental elastic resolution (solid black line), two Lorentzian functions (red and blue solid lines) used to describe the dynamic processes, the experimental data (yellow dots), and the total fit (dashed line).

The dependence of the full width half maximum $\Gamma$ and the elastic incoherent structure factor (EISF, which is defined as the ratio of elastic to total scattering) were analyzed as function of $Q$ and $Q^2$ and fit to a series of models (Volino-Dioux, Singwi-Sjölander, three-site jump site potential, and bounded diffusion within confinement models) that describe the protein dynamics and are explained in detail in the results section.\textsuperscript{177,178,195,196}
2.6 Conductivity measurements

**Proton conductivity:** Proton conductivity of TR protein films (approximately 1cm x 3cm x 80 \(\mu\)m, dimensions vary between samples and are examined for each measurement) was determined by AC impedance spectroscopy using a two-probe measurement cell (platinum-coated nickel electrodes) connected to an impedance/gain-phase analyzer (Solatron 1260, analysis with ZView and ZPlot software). Measurements were conducted at an AC amplitude of 100 mV and a frequency interval ranging from 100 Hz to 10 MHz. The real (\(Z'\)) and imaginary (\(Z''\)) components of the impedance were measured simultaneously over the defined frequency range, and were analyzed in a Nyquist plot (Figure 2-5).

![Nyquist plot of SRT hydrated protein from impedance spectroscopy. Equivalent circuit for the calculation of bulk resistance (inset).](image)

The impedance curves were fit to an equivalent circuit (Figure 2-5 inset) consisting of: bulk resistance (\(R_{\text{bulk}}\)) and bulk capacitance (\(C_{\text{bulk}}\)) connected in parallel (accounting for the protein film), connected in series to the film/electrode interface capacitance (\(C_{\text{interface}}\)). This equivalent circuit accounts for the bulk impedance and capacitive effects at the contacts, and has been
extensively used in modelling proton-exchange membranes. In short, the bulk resistance of the films $R_b$ (Ω) was calculated from the Nyquist plot by extrapolating the impedance line to the x-axis (real impedance where the imaginary impedance is zero, $Z'$ at $Z'' = 0$). The proton conductivity of the film can be calculated as $\sigma = L / (R_b A)$, where $\sigma$ is the proton conductivity (S cm⁻¹), $L$ is the length between the electrodes (0.7 cm), $A$ is the cross sectional area of the membrane sample (cm²). Impedance data were collected as a function of temperature (20 °C, 50 °C, and 70 °C) and water content in a controlled humidity chamber. Stretched samples were measured in situ using a home-built stretching stage.

**Time-domain thermoreflectance (TDTR):** thermal conductivity of dry and fully hydrated SRT-TR films was measured by Dr. Patrick E. Hopkins and John Tomko using time-domain thermoreflectance (TDTR). This ultrafast laser pump-probe technique has been a common tool for thermal measurements of thin films, as explained in detail elsewhere. When applying TDTR to measure the thermal conductivity of protein films, the thermoreflectance signal from a transducer beneath the TR film is related to the temporal decay of energy density and transport mechanisms induced from a modulated pump pulse. In fitting the data to a thermal model, the thermo-physical properties of the various layers can be extracted.
Chapter 3. Native SRT: sequence and structure

The major goal of this chapter is to identify the molecular structure of native SRT. In particular, native SRT from European common squid (*Loligo vulgaris*) will be analyzed. Using high-throughput RNA sequencing and proteomics tools, the amino acid composition and sequence of *Loligo vulgaris* SRT is determined. Next, the secondary structure of SRT is analyzed by infrared spectroscopy and X-ray diffraction. Last, native SRT are characterized by thermo-mechanical methods, revealing thermal and mechanical properties that arise from structural elements. Hence, this chapter is divided in amino acid analysis, structural analysis, and thermo-mechanical properties sections, and the relationship between them is discussed.

3.1 Introduction

Squids have evolved teeth inside their suckers that provide a strong predatory grasp of the tentacles and arms. These squid ring teeth (SRT) structures have remarkable mechanical properties, with modulus of 7 GPa and hardness of 0.7 GPa, that exceed those of most engineered polymeric materials.\(^{202}\) They have a structure composed of tightly packed parallel tubular elements with 90 to 300 nm in diameter that are arranged perpendicular to the surface, similar to a honeycomb structure.\(^{202,203}\) Although the teeth shape and size slightly vary across squid species, the organized tubular nanostructure is conserved (Figure 3-1). This structure provides compliance (necessary for bending and actuating the ring structure inside the sucker cup) while maximizes the strength of the material in compression, making SRT very effective predatory appendages.
Figure 3-1. SEM cross-section of SRT from six squid species reveals a highly organized tubular structure. Insets show the varying morphology of SRT across species.

Because of its mechanical properties, it was first believed that SRT were chitin-based materials, as it is very common in hard tissues in insects, marine animals, and even other hard tissues in squid (i.e. gladius and beak). A book published by Williams a century ago in 1909 entitled “The anatomy of the common squid” studied the squid arms and tentacles in detail. At the time, the author claimed that the squid teeth were “composed of a chitin-like substance, which differs from chitin in being soluble in dilute alkalies”. This reported “difference” was the first step into the analysis of SRT structure and composition. In 1977, Nixon and Dilly published a comprehensive study of cephalopod’s sucker surfaces, including the SRT, which concluded that squid’s suckers have teeth that are not chitin-like but are made of a protein. SRT are fully dissolvable in weak acid solutions (5% acetic acid) and in denaturing cocktails used in protein extraction (concentrated urea solutions), and further characterization by Miserez et al. via Ninhydrin-based amino acid analysis confirmed that there is no chitin present in SRT and they are solely composed of proteins. This discovery, together with technological advances in high-throughput RNA sequencing, proteomics, and advanced materials characterization, set the stage for the comprehensive analysis of SRT proteins that is described in this work.
3.2 Results and discussion

3.2.1 Amino acid sequence analysis

The protein distribution of native SRT from six different species (*Todarodes pacificus, Sepioteuthis lessoniana, Illex argentinus, Loligo vulgaris, Loligo pealei, and Illex illecebrosus*) is analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), which revealed that native SRT are protein complexes (Figure 3-2). SRT are composed by a group of 2 to 10 proteins ranging between 15 and 60 kDa. Although the polydispersivity is conserved across species, each squid species shows a distinct molecular weight distribution (with the more abundant proteins in the 40-55 kDa range). This phenomenon is also observed in other species not investigated in this work (*Dosidicus gigas, Euprymna scolopes, and Sepia esculenta*).  

![Figure 3-2. SDS-PAGE of SRT from six squid species (*Todarodes pacificus, Sepioteuthis lessoniana, Illex argentinus, Loligo vulgaris, Loligo pealei, and Illex illecebrosus*). SRT across species are composed of a group of proteins between 15 and 60 kDa.](image)

In general, a monodisperse molecular weight distribution is common in globular and membrane proteins since they have to fold into well-defined structures to perform biological functions. On
the other hand, structural proteins do not have a biological function and they do not require monodispersity as long as they can assemble into their characteristic long-range ordered aggregates. Similarly to synthetic polymers, polydispersity in proteins can be used as a strategy to tune the morphology and properties, and can drive self-assembly into ordered structures. Therefore, two possible reasons for protein polydispersity arise from this observations: a) protein polydispersity is necessary to grow SRT with its characteristic morphology, where proteins different in size drive the assembly of the nanotubular structures. Although it would explain the different distributions across species (causing different morphologies in the rings), a comprehensive study of the protein secretion inside the sucker cups is necessary to evaluate the assembly mechanism (which is extremely challenging due to the need of in vivo analysis in squids), and therefore it remains a speculation. And b) protein polydispersity maximizes the mechanical properties of the material (which will be explored in detail in Chapter 5).

### 3.2.1.1 Amino acid composition

The amino acid composition of native *Loligo vulgaris* SRT (Figure 3-3) shows high composition of glycine, tyrosine, and histidine accounting for approximately 50% of total amino acids. Glycine (26.2%) does not have any side chain and therefore it is the smallest and most flexible amino acid. It is abundant in elastomeric proteins such as collagen, and provides chain flexibility and rubberlike elasticity to the protein. Tyrosine (12.5%) is a hydrophilic amino acid that can form hydrogen bonds between the hydroxyl group of one tyrosine residue and the carboxyl group of another. The high content of tyrosine is likely to contribute to the formation of a hydrogen bonded protein network and provide stability. Histidine (10.9%), which is a positively charged amino acid at physiological pH, is important in load bearing and impact resistance tissues because of the versatility of the imidazole side-chain to stabilize protein networks, for instance through hydrogen-bonding, metal coordination complexation or covalent cross-linking.
Figure 3-3. Amino acid composition of *Loligo vulgaris* native SRT. Major amino acids are glycine (26.2%), tyrosine (12.5%), and histidine (10.9%).

In Table 3-1, the amino acid composition of SRT is also compared with commonly observed noncrystalline bioelastomers\(^2^0^8\) (*e.g.*, elastin\(^9^0\), resilin\(^7^0\), and abductin\(^2^0^9\)). Compared with these bioelastomers, the proteinaceous structure of the SRT does not have any covalent cross-linker and, hence, dissolves in weak alkaline or acidic buffers. High nonpolar to polar ratio of the amino acid side groups of SRT and other bioelastomers listed in Table 3-1 suggests that the chains would be well separated by the presence of solvent molecules (where resilin is an exception for the polarity ratio and the mechanism is described elsewhere\(^7^0\)).

**Table 3-1. Comparative amino acid composition (%) of SRT, elastin, abductin, and resilin.**

<table>
<thead>
<tr>
<th>Side chain</th>
<th>Occurrence in proteins(^2^1^0)</th>
<th>SRT(^a)</th>
<th>SRT(^b)</th>
<th>Elastin(^c)</th>
<th>Abductin(^c)</th>
<th>Resilin(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small(^d)</td>
<td>21.1</td>
<td>39.5</td>
<td>47.4</td>
<td>54.9</td>
<td>70.2</td>
<td>62.0</td>
</tr>
<tr>
<td>Nonpolar</td>
<td>39.9</td>
<td>38.4</td>
<td>37.4</td>
<td>60.9</td>
<td>28.5</td>
<td>22.0</td>
</tr>
<tr>
<td>Polar</td>
<td>22.1</td>
<td>9.1</td>
<td>7.2</td>
<td>7.5</td>
<td>13.1</td>
<td>35.0</td>
</tr>
</tbody>
</table>

\(^a\) *Loligo Vulgaris*\(^2^1^1\)  \(^b\) *Dosidicus Gigas*\(^2^0^2\)  \(^c\) Adapted from Erman *et al.*\(^2^1^2\)  \(^d\) Small amino acids are G, V, A
The major amino acids of SRT are non-charged (Gly, Tyr, Leu, Ala) with the exception of histidine (10.9%). Histidine has an imidazole side chain which is partially protonated and has positive charge at physiological pH. Being the only major charged amino acid, it is expected that histidine dominates the charge of the overall protein. \( \zeta \)-potential measurements of SRT revealed an isoelectric point of \(~6.6\) which is close to the pKa of histidine (6.04) (Figure 3-4). The net charge of the protein is close to zero at physiological buffers, and therefore SRT proteins are neutral in water (squid’s habitat). This has strong implications in protein stability in aqueous environments, since pH around 7 will favor the aggregation of SRT proteins (neutral charge) and lead to stable and stiff materials. Other fibrous proteins that lack histidine have a lower isoelectric point and are not stable in aqueous solutions (silk fibroin has an isoelectric point of 4.5, and non-crystalline silk is totally soluble in water).\(^{213,214}\)

### 3.2.1.2 Amino acid sequence

A combination of high-throughput RNA-sequencing and protein mass spectroscopy was performed to identify several proteins of the *Loligo vulgaris* SRT. mRNA extracted from the suction cups of the epithelium tissues of *Loligo vulgaris* was sequenced to identify the cDNA that aligned with the protein sequences observed in SRT complex. The overall sequencing process is described in the Chapter 2.\(^{82,83}\)
Figure 3-5. Amino acid sequence of proteins composing native *Loligo vulgaris* SRT. GLY-rich segments are marked in yellow, AVSTH-rich segments are marked in blue, and proline (P) residues are marked in bold.

SRT of *Loligo vulgaris* is composed of six major proteins of 18, 22, 26, 30, 45, and 54 kDa in size. The amino acid sequence of each of the proteins is illustrated in Figure 3-5. The analysis of the overall sequence reveals two repetitive motifs across all proteins: AVSTH-rich segments (marked in blue), and GLY-rich segments (marked in yellow). Additionally, the majority of the segments are separated by a proline residue (P). In practice, there are no universal laws regarding the sequence-structure relationship for structural proteins, but certain guidelines are accepted. A common one is the content of certain amino acids, especially of glycine and proline. Both amino acids are commonly seen in disordered peptides, but their contributions are very different. Glycine does not contain any side chain, therefore it is very flexible compared to other amino acids and it brings a higher entropic penalty to the system when it is ordered. On the other hand, proline is a very rigid amino acid due its cyclic structure (the pyrrolidine side chain locks the chain dihedral angle $\phi$), which introduces kinks and disturbs the packing of other amino acids into secondary structures such as $\alpha$-helices and $\beta$-sheets. Hence, proline is usually found as the first or last
residue in α-helices and β-sheets and in turn structures. A notable exception is collagen, where multiple proline residues form polyproline helices and increase the stability of collagen.\textsuperscript{40,216,217}

In the case of SRT, the presence of proline in the amino acid sequence in between two different motifs (AVSTH-rich and GLY-rich segments) suggests the separation between two different secondary structural elements. The length of all segments across the six sequenced proteins of \textit{Loligo vulgaris} SRT is analyzed (Figure 3-6a). AVSTH-rich segments have a conserved length of 10-15 residues approximately, while GLY-rich segments have varied inhomogeneous lengths (up to 82 residues long). Considering the high flexibility due to the glycine content, GLY-rich segments are very likely to be disordered (random coil) and therefore they are not subject to strict length limitations (i.e. size is not a restriction in disordered configurations). On the other hand, the length of AVSTH-rich segments is repetitive within very few amino acids, suggesting that these segments likely align themselves into a defined arrangement limited by size. Considering the hydrophobicity of alanine residues, it is expected that hydrophobic forces drive the aggregation into β-sheet structures and keep them together as observed in other structural proteins.\textsuperscript{218} Moreover, alanine-rich β-sheets in silk fibroin align parallel to each other via alternating methyl side chains of alanine groups and they form a three dimensional stack of a few nanometers in size.\textsuperscript{219} From these observations, it is expected that AVSTH-rich segments are arranged in β-sheet structures 10 to 15 residues long.

![Figure 3-6. a) Segment length distribution and b) histidine-proline distances in \textit{Loligo vulgaris} native SRT (over all sequenced proteins).](image-url)
Histidine residues in the chain segments are very concentrated in the terminal ends (Figure 3-6b). Due to their positive charge, histidine residues are likely to direct the assembly of β-sheet. Their positioning within a 5 residue from the end of the segment (majority is adjacent or 1 residue away) confers anisotropy to the chain. During the folding into β-sheet structures, it is expected that terminal histidine residues from neighboring segments will repel each other due to electrostatic repulsion. For this reason, terminal histidines would favor the formation of antiparallel β-sheet over parallel β-sheet.

### 3.2.2 Structural analysis

The analysis of the amino acid sequence of *Loligo vulgaris* SRT revealed a segmented sequence topology that suggests the presence of β-sheet structures. In this section, *Loligo vulgaris* SRT is structurally characterized by Fourier infrared transform spectroscopy (FTIR) and X-ray diffraction (XRD) in order to quantify the secondary structure content, understand the self-assembly mechanisms, and confirm the previously proposed sequence-structure hypotheses.

#### 3.2.2.1 Secondary structure analysis

The secondary structure content of proteins (i.e. fraction of β-sheets, α-helices, and other structures) has been extensively studied over the years by circular dichroism (CD), nuclear magnetic resonance (NMR), Raman spectroscopy and infrared spectroscopy. However, non-soluble materials present characterization challenges. Attenuated total reflectance (ATR) Fourier transform infrared (FTIR) spectroscopy is especially useful in the study of fibrous proteins and biological materials because it does not require the dissolution of the material and hence the structure of the native material can be analyzed without damage arising from sample preparation. Following this approach, SRT of *Loligo vulgaris* were analyzed by ATR-FTIR spectroscopy.

The FTIR spectra of SRT (Figure 3-7a) gives 4 major bands in the infrared absorption spectrum that are characteristic to proteins, named amide A, I, II, and III (other minor bands are omitted for simplicity). Amide A is found around 3300 cm$^{-1}$ approximately. It is exclusively governed by the N-H stretching vibration and therefore it is insensitive to the polypeptide backbone conformation.
However, it is sensitive to hydrogen bonding since its frequency is dependent on the strength of the hydrogen bond. The amide I band is the most intense band and it can be found between 1600 and 1700 cm\(^{-1}\). The amide I band is originated by the stretching vibration of the C=O (70-85\%) and C-N (10-20\%) groups. The amide I band is highly dependent on the backbone conformation of the protein, and it is insensitive the amino acid side chains. It is for this reason that the amide I band is the most commonly studied for structural determination in proteins and amides. The amide II band is found around 1550 cm\(^{-1}\), and it is mostly related to the in-plane N-H bending vibration (40-60\%) and the stretching vibrations of the C-N (18-40\%) and the C-C (10\%). It is insensitive to side chains and it contains structural information about the secondary structure, although the analysis is more complicated and limited than that of amide I. The amide III band is found around 1200-1400 cm\(^{-1}\), and it is the combination of N-H bending and C-N stretching vibrations. Although it can be used for structure prediction, its analysis is complex due to the overlapping contribution from side chains and several modes of N-H bending in this region.\(^{151,153}\)

![FTIR spectra of *Loligo vulgaris* native SRT](image)

**Figure 3-7.** FTIR spectra of *Loligo vulgaris* native SRT. a) SRT spectra exhibits amide I, amide II, amide III, and amide A bands (characteristic of proteins). b) Amide I band is centered at 1630 cm\(^{-1}\) indicating a high fraction of β-sheet structures.

The amide I band of native *Loligo vulgaris* SRT (Figure 3-7b) has a maximum at 1630 cm\(^{-1}\), which indicates that the material is rich in β-sheet structures (as suggested by the amino acid segmented sequence analysis). There are several rules of thumb for structural prediction from amide I band analysis that have been established from computational and experimental studies over the years. Antiparallel β-sheet structures generate a strong absorption band at 1630 cm\(^{-1}\) and a weaker band
around 1690 cm$^{-1}$. The position of this bands can change depending on the number of strands and
the twisting of the β-sheets. Parallel β-sheets have absorption bands at higher wavenumber than
antiparallel, but the difference can be as little as 4 cm$^{-1}$. For this reason it is experimentally difficult
to differentiate between parallel and antiparallel β-sheet solely from FTIR (especially if both
species might coexist together). Other secondary structures are also included in the main amide I
absorption band. α-helices give rise to an absorption band at 1660 cm$^{-1}$, although the position of
the band depends on the length of the helix. Turns have absorption bands in the 1670-1690 cm$^{-1}$
region, and disordered polypeptide chains (random coil conformation) have absorption bands in
the 1650 cm$^{-1}$ region.$^{151,153,159,227}$ Although a qualitatively analysis of the amide I reveals a structure
rich in β-sheet, it is difficult to separate the contribution of other structural elements and provide
a detailed quantitative structural analysis.$^{228}$ However, Fourier self deconvolution (FSD) of the
amide I band and its corresponding analysis is a more robust and reliable method that can provide
reliable quantitative results.$^{154}$

![Figure 3-8. Deconvoluted FTIR amide I band of native SRT shows a major β-sheet content. Individual bands are fitted to the spectra, and are assigned to side chains (sc), β-sheet (β), α-helices (α), and turn (t) structures as described in Table 3-2.](image)
Figure 3-8 shows the deconvoluted spectra of native *Loligo vulgaris* SRT protein amide I absorption band, and a set of individual secondary structure bands that has been fitted. A total of 11 bands have been fitted to the deconvoluted spectra, giving similar results to FTIR analysis of *Bombyx mori* silk fibroin.\(^{152}\) Each band has been labeled as β-sheets (β), α-helix (α), random coil (rc), turns (t) and side chains (sc).

The band centered around 1594 cm\(^{-1}\) (marked as sc) is assigned to the side chains of the protein, found between the Amide I (1600-1700 cm\(^{-1}\)) and the Amide II region (1510-1580 cm\(^{-1}\)). Although it is not related to the secondary structure, it is still observed in the vicinity of the deconvoluted amide I band. The absorption peak in this region is related to the aromatic ring in the side chains of tyrosine and histidine.\(^{153,155}\) Tyrosine and histidine are contributing strongly to this band since their amino acid fraction is 12.5% (Tyr) and 10.9% (His) for the native SRT protein.

A triplet of bands (marked as β) is fitted to the deconvoluted spectra between 1600 and 1637 cm\(^{-1}\) which are assigned to β-sheets.\(^{151,156-159}\) The band centered around 1612 cm\(^{-1}\) is typically assigned to intermolecular β-sheets formed by molecular aggregation.\(^{157,158,160,229}\) The band centered around 1621 cm\(^{-1}\) is usually assigned to intermolecular β-sheets in crystallized proteins. This band has been assigned to the stacking of antiparallel β-sheets in previous studies on silk fibroin.\(^{152}\) The band centered around 1631 cm\(^{-1}\) is assigned to the formation of intramolecular β-sheets.\(^{152,229}\) Last, another small β-sheet band can be observed at 1697 cm\(^{-1}\), which is due to the splitting of β-sheet bands and is also observed in other β-sheet-rich polypeptides.\(^{151-153,230}\)

A set of bands between the major β-sheet bands and the minor β-sheet band (1635-1700 cm\(^{-1}\) range) are attributed to random coils, α-helices and turn secondary structures. The two bands centered at 1643 cm\(^{-1}\) and 1650 cm\(^{-1}\) (marked as rc) were assigned to random coil conformations.\(^{151,157,158,160,161}\) The band centered at 1661 cm\(^{-1}\) (marked as α) is assigned to α-helix secondary structures.\(^{157,161}\) The three remaining bands centered at 1667, 1679 and 1693 cm\(^{-1}\) are assigned to turns structures.\(^{151,156,157,160,162}\)
Table 3-2. Secondary structure content of native SRT determined by FTIR.

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Secondary structure content (%)</th>
<th>Band position (cm(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>side chains(^{153,155})</td>
<td>-</td>
<td>1594</td>
</tr>
<tr>
<td>β-sheets(^{151–153,156–159})</td>
<td>49.0 ± 1.7</td>
<td>1612,1621,1631,1697</td>
</tr>
<tr>
<td>random</td>
<td></td>
<td></td>
</tr>
<tr>
<td>coils(^{151,157,158,160,161})</td>
<td>21.8 ± 0.7</td>
<td>1643,1650</td>
</tr>
<tr>
<td>α-helices(^{157,161})</td>
<td>9.8 ± 0.7</td>
<td>1661</td>
</tr>
<tr>
<td>turns(^{151,156,157,160,162})</td>
<td>19.4 ± 1.1</td>
<td>1667,1679,1693</td>
</tr>
</tbody>
</table>

The fraction of secondary structure elements was determined by calculating the ratio of the fitted band area to the total deconvoluted amide I band area (excluding the side chains band \(sc\)). The secondary structure composition of *Loligo vulgaris* SRT protein complex is 49.0 ± 1.7 % β-sheets, 21.8 ± 0.7 % random coils, 19.4 ± 1.1 % turns and 9.8 ± 0.7 % α-helix, as listed in Table 3-2. The assignment of the minor β-sheet band at 1697 cm\(^{-1}\) is often controversial because overlaps with the turn bands, and it is difficult to differentiate between the two structures. However, it represents less than the 2% of the total amide I region. The secondary structure composition reported here is similar to that of other β-sheet-rich fibrous proteins such as silk fibroin and amyloids.\(^{152,162,227}\)

The quantitative measurements of the secondary structure of native SRT supports the structure predictions from the segmented sequence analysis. The amino acids of all sequenced *Loligo vulgaris* SRT proteins are distributed between GLY-rich and AVSTH-rich segments in a 0.47 and 0.53 ratio, which agrees with the measured 0.49 fraction of β-sheet content (within error). While AVSTH-rich segments aggregate in β-sheet structures, the long GLY-rich segments are mostly disordered (random coil conformation) amorphous chains that connect the β-sheet together. The high content of turn structures, which help stabilize abrupt directional changes along the protein backbone, would help the formation of antiparallel β-sheets over parallel β-sheets. β-strands (AVSTH-rich segments) in the same chain separated by amorphous segments containing a turn (GLY-rich segments) would be close enough to form antiparallel β-sheets. However, the analysis of the secondary elements and the morphology of SRT solely by FTIR is limited, and further structural characterization is required by diffraction methods.
3.2.2.2 β-sheet analysis

β-sheet structures are formed by laterally-connected strands of protein with hydrogen bonding interaction between the backbone carbonyl oxygen and the amino hydrogen atoms. Typically, multiple β strands can be arranged into an extensive hydrogen bonding network with their neighboring strands, forming crystal-like domains in the protein matrix. Since peptide chains have directionality depending on their N-terminus and C-terminus, β-sheet secondary structures have directionality too depending on the hydrogen bonding pattern. In an antiparallel structure, the neighboring β strands alternate directions (e.g. N-terminus of a strand is adjacent to the C-terminus of the next). Allowing the inter-strand hydrogen bonding between carbonyl and amine groups to be planar, this β-sheet configuration is the most stable. Inversely, a configuration wherein parallel β-sheets are formed by β strands oriented in the same direction (e.g. N-terminus adjacent to another N-terminus) on the other hand disallows planar inter-strand hydrogen bonding and is, thus, less stable than the antiparallel configuration. The β-sheet type (parallel/antiparallel) and the size of the crystallite (number of strands, length along the backbone, and number of stacks) play a major role on the physical properties of fibrous proteins. For this reason, the morphology of native Loligo vulgaris SRT is analyzed here.

Figure 3-9. Schematic representing parallel and antiparallel β-sheet configurations in SRT crystalline segments. Simulation courtesy of Murat Cetinkaya.
Due to the presence of histidine amino acids in the end of each crystalline segment (next to each proline residue that separates the crystalline and amorphous segments), it is suggested that the antiparallel arrangement of β-sheets is more favorable than parallel β-sheets (both inter and intramolecular). Parallel hydrogen bonding would position neighboring His side chains next to each other, resulting in a less stable and less ordered β-sheet stacking due to the high volume of the aromatic ring in His side chains. On the other hand, antiparallel hydrogen bonding alternates the position of the His amino acids in neighboring chains, resulting in a more ordered and stable hydrogen bonding structure (Figure 3-9).

Figure 3-10. X-ray diffraction (XRD) of *Loligo vulgaris* native SRT. Major diffraction peaks are observed at 2Θ = 9.3° and 2Θ = 19.2°, which correspond to alternating β-sheet strands.

Native *Loligo vulgaris* SRT has been characterized by X-ray diffraction (XRD) (Figure 3-10). The 2D diffraction pattern (Figure 3-10 inset) shows approximately equal intensity independent of the orientation, suggesting that the morphology is isotropic. The integration of the 2D pattern across all azimuthal angles result in very broad spectra characteristic of semicrystalline protein structures. Despite the major contribution of the amorphous halo, two crystalline peaks are be observed at 2Θ
= 9.28° and 2Θ = 19.16°. A shoulder towards higher 2Θ is suspected, which is caused by an additional peak at 2Θ ≈ 25°. The d spacings of the crystalline system are calculated according to Bragg’s law $n\lambda = 2d \sin \Theta$,\textsuperscript{164} where $\lambda$ is the X-ray wavelength, giving d spacing of 9.53 Å and 4.63 Å for the major two peaks (2Θ = 9.28° and 2Θ = 19.16° respectively). The measured lattice is characteristic of β-sheet structures, where the measured lattice distance values correspond to the hydrogen-bond distance between two β-sheet chains (4.72 Å) and the distance between alternating β-sheet chains (9.44 Å).\textsuperscript{165} Furthermore, the third peak (2Θ ≈ 25°) corresponds to a d spacing of 3.56 Å, which is the chain length of a single amino acid in a β-sheet structure (3.5 Å, with a two-residue repeat distance of 7.0 Å).\textsuperscript{165} The diffraction pattern and the position of the peaks are consistent with those of other β-sheet-rich fibrous proteins such as silkworm and spider silk.\textsuperscript{234,235}

The size of the β-sheet crystals of native \textit{Loligo vulgaris} SRT was estimated by the analysis of the peak broadening according to Scherrer’s equation, $\Gamma(2\Theta) = k \lambda / (L \cos \Theta)$,\textsuperscript{168} where $\Gamma(2\Theta)$ is the full width at half maximum of the crystalline peak, $k$ is a constant of proportionality (usually 0.9 to 1 for polymers and cubic symmetries), $\lambda$ is the X-ray wavelength, and $L$ is the crystallite size. The wide breadth of the peaks suggests that the β-sheet crystals are small (i.e. a few nanometers in size). Crystal dimensions of 1.5 – 2 nm along the hydrogen bonding direction and 2 – 2.5 nm along the backbone direction are estimated, which correspond to 3 – 4 aggregated β-sheet strands and 6 – 10 amino acids along the chain. An accurate determination of the size from native samples is challenging due to the sequence diversity. AVSTH-rich crystal-forming segments of the native SRT protein complex vary in composition and size, leading to a distribution of crystal dimensions. Variations in the sequence, the small size of crystallites, and the signal contribution from the amorphous background complicate the analysis significantly, and increase the risk of underestimating the size due to artifacts arising from these problems. However, even with these limitations, the measured crystalline structures supports previous results and agrees with the predictions from the sequence analysis (AVSTH-rich crystal-forming segments have 10 – 15 amino acids in length versus 6 – 10 amino acids measured experimentally).
3.2.3 Thermal properties

In this section, the thermal properties of native *Loligo vulgaris* SRT are investigated. Due to its β-sheet-rich structure, it is expected that SRT exhibits properties similar to those of semicrystalline polymers. In this analogy, β-sheet strands (AVSTH-rich segments) aggregate forming the crystalline phase while GLY-rich segments form the amorphous phase. It is then expected to observe two major thermal transitions: a glass transition at $T_g$ and a melting transition at $T_m$.

At very low temperature, the polymer chains are greatly compressed and the modulus is high, but as the temperature is increased, greater mobility results in greater compliance of the molecule and, thus, lower moduli; hence, the polymer expands and the free volume increases. As the temperature continues increasing, the chains in the amorphous domain begin to coordinate large-scale movements in what is known as the glass transition $T_g$. The glass transition is classically described as the transition in which the amorphous domains begin to melt, and it is not observable in pure crystalline materials. If the temperature is further increased, the melting $T_m$ of the crystalline domains occurs, where large-scale crystal slippage occurs and the material starts flowing. However, the transitions above $T_g$ do not occur for cross-linked polymers since the cross-linking prevents the chains from moving past another.\textsuperscript{171} If temperature is further increased, the material thermally degrades at $T_{deg}$ and decomposes into carbon dioxide, water, and short hydrocarbons. $T_g$, $T_m$, and $T_{deg}$, define the operating temperature range over which a polymer can be used since they drastically change the material properties and structure. For instance, while $T_g$ is the upper temperature limit for applications where strength and stiffness of the material are required, it is the lower limit for applications preferring elastomeric behavior. In addition, the material will be irreversibly damaged beyond $T_{deg}$ due to the degradation of the chains. For this reason, the $T_g$ is often the lower temperature limit for polymer processing techniques based on the advantages of plastic deformation methods like extrusion, compression molding, lamination, etc.\textsuperscript{171}

In addition to the aforementioned thermal transitions, which are classic to polymers, proteins also have another transition which affects the overall structure. Protein denaturation ($T_d$) consists in the unfolding of the secondary structure features of the protein such as the β-sheet or α-helix structures into a random coil configuration. Usually, it results in the irreversible aggregation of the protein due to inter-protein hydrophobic/hydrophilic interactions, and entails the loss of function of the protein.\textsuperscript{236} The denaturation process can be originated by temperature, pressure, pH, ion
concentration (salts), or presence of solvents in the protein solution. This transition is commonly observed in globular proteins, and it has been extensively studied since it causes the loss of structure and biological function. However, the implications of protein denaturation in SRT protein complexes will be discussed as well.

Figure 3-11. Temperature modulated differential scanning calorimetry (TMDSC) of native SRT. Reversing heat flow (dashed line) shows a step change around 180 °C, indicating a glass transition.

The thermal transitions of *Loligo vulgaris* SRT were studied by temperature modulated differential scanning calorimetry (TMDSC). Protein samples were previously annealed at 120 °C in order to remove ambient absorbed water (water content after annealing was ~0.2% w/w). Figure 3-11 shows the total heat flow (left axis) and reversing heat flow (right axis) for *Loligo vulgaris* SRT proteins. The reversing heat flow shows a step change around 180 °C, which indicates a glass transition (marked as *T*₁). The observation of the glass transition in a macroscopic measurement is a clear confirmation of the semicrystalline nature of SRT morphology, which supports the sequence analysis hypothesis. The heat capacity increase Δ*C*ₚ during the glass transition process was estimated as 0.23 ± 0.02 J/(g °C), which falls within the expected values for other semicrystalline polymers and proteins. A small endotherm in the total heat flow is observed after the glass transition region that might corresponds to a relaxation peak (probably caused by the previous
annealing of the sample for water removal).\textsuperscript{152,237} Above 210 °C the heat flow signal strongly decreases, which is probably caused by the onset of thermal degradation (T\textsubscript{deg}). Interestingly, no melting endotherms were observed, suggesting that the melting point (T\textsubscript{m}) of the SRT \( \beta \)-sheet crystallites is higher than the degradation temperature. This implies that the physical cross-linking of \( \beta \)-sheets is stable enough to withstand high temperatures up to 200 °C without compromising the physical properties of SRT. This phenomenon is observed in classic elastomers and thermoset polymers, where the cross-linking mechanism is through irreversible covalent bonds (i.e. the polymer thermally degrades before breaking the cross-linking bonds). Silk fibroin exhibits a similar phenomenon, where the material thermally degrades before melting the \( \beta \)-sheet crystals.\textsuperscript{152} However, advanced fast scanning calorimetry techniques (with heating rates as fast as 2000 K/s) enable the analysis at high temperatures without damaging temperature-sensitive materials (extremely short exposure time to high temperatures). Fast scanning calorimetry of silk fibroin revealed the melting of \( \beta \)-sheets at T\textsubscript{m} \sim 260 °C, which is above the thermal degradation temperature.\textsuperscript{238} Due to the similarities in sequence, amino acid content and semicrystalline morphology, it is expected that \( \beta \)-sheets in SRT proteins have a similar melting temperatures above the thermal degradation point (T\textsubscript{m} > T\textsubscript{deg}).

Figure 3-12. Thermal degradation of native SRT measured by thermogravimetric analysis coupled to mass spectrometry (TGA-MS) is observed starting at 210°C.
The thermal degradation of native SRT was further investigated by thermo gravimetric analysis and mass spectrometry (TGA-MS). The weight loss of native *Loligo vulgaris* SRT with temperature is shown in Figure 3-12. A slow weight loss of approximately 5% can be observed in the room temperature to 200 °C region, which is attributed to water absorption in ambient conditions. This is confirmed by mass spectrometry, which showed that the only ions detected in that region were m/z 17 and 18, corresponding to OH⁻ and H₂O respectively. Other than that, SRT is stable in temperatures up to 200 °C. Thermal degradation of the proteins is observed at temperatures above 210°C as a major weight loss. Simultaneous mass spectrometry showed the release of water (m/z 17 and 18), CO₂ (m/z 44), and short hydrocarbons (not shown), confirming the decomposition of the polypeptide.

### 3.2.4 Mechanical properties

Thermo-mechanical characterization was performed on *Loligo vulgaris* SRT protein complex by dynamic mechanical analysis (DMA) (Figure 3-13). SRT in dry conditions (water content ~0.2% w/w) exhibits stable moduli ranging in temperature from 25 to 195°C, with a storage modulus E’ of ~ 2 GPa. At 180 °C the storage and loss moduli exhibit a sharp drop indicating the glass transition in the protein complex (which agrees with calorimetry measurements). The tangent δ is a good indicator of the transition since a great increase towards tan δ = 1 can be observed in the 190-210 °C region. The rubbery plateau modulus above the glass transition cannot be measured in dry conditions due to the imminent thermal degradation of the proteins at temperatures above 220 °C. Therefore, water was used a plasticizer in order to decrease the glass transition temperature and avoid the thermal degradation of the proteins (water content ~20% w/w).²³⁹–²⁴¹ The dynamic mechanical analysis for hydrated SRT protein complex is similar to that in dry conditions, but the transition is occurring at lower temperatures due to the plasticizing effect of water. The plasticizing effect of water can be clearly observed as a decrease in the glass transition temperature (T_g ~ 40 °C). Although the rubbery state is important for the structural and mechanical characterization of the material, it should be noted that the working temperature of SRT in its natural conditions remains 10 – 15 °C (depending on the squid’s habitat), and hence it usually operates in its glassy state (GPa modulus). This is advantageous to the squid since it provides with high modulus teeth for predatory purposes rather.
Figure 3-13. Dynamic mechanical analysis (DMA) of (a) dry (water content ~0.2%) and (b) fully hydrated (water content ~20%) native SRT shows stable GPa moduli in the glassy region, and a glass transition at 190-210 °C and ~40 °C respectively.

Figure 3-14. Dynamic rheology of fully hydrated (water content ~20%) native SRT at 70 °C shows a weak frequency dependence and a rubbery modulus of 3 MPa.

The rubbery modulus of SRT was measured in a rotational rheometer above T_g (T = 70 °C) (Figure 3-14). Both storage and loss moduli show a very weak frequency dependence in the 1 < ω < 100 rad/s region, and a frequency-independent plateau in the 1 rad/s > ω region. This behavior is characteristic of imperfect polymer networks. The weak frequency dependence at high frequencies originates from the relaxation of dangling ends and loops in the network structure. These network defects act as elastically effective strands at high frequencies, but they do not contribute to the
modulus at frequencies lower than their relaxation time (hence, the modulus decreases at lower frequencies). The network reaches its equilibrium modulus at low frequencies (i.e. low frequency plateau). The fact that the modulus does not decay at low frequencies indicates that the \( \beta \)-sheet cross-linked network is very stable, especially when compared to other physically cross-linked protein systems such as coil-coil-based protein systems that behave as liquids in the low frequency limit.

3.3 Conclusions

The amino acid sequence, nanostructure, and thermo-mechanical properties of native SRT from *Loligo vulgaris* squid were analyzed in this chapter. Squid Ring Teeth (SRT) protein chains contain both hard and brittle crystalline (AVSTH-rich) segments and amorphous flexible (GLY-rich) segments, similarly to semi-crystalline polymers and to the structure of silk fibroin. The backbones of the repeating crystalline units align and form a randomly oriented hydrogen bonding network, resulting in physically cross-linked \( \beta \)-sheet crystalline domains responsible for the mechanical properties of the material. On the other hand, disordered domains contain flexible peptide chains that connect the crystallites forming a \( \beta \)-sheet-stabilized protein network (Figure 3-15). The self-assembled nanostructure of SRT gives rise to excellent mechanical properties (GPa modulus) that provide squids with advantageous high strength predatory materials. Hence, a relationship between sequence (segmented repetitive motifs), nanostructure (\( \beta \)-sheet formation), and thermo-mechanical properties (high modulus, glass transition, and network behavior) is proposed.

![Figure 3-15. Schematic of SRT semicrystalline structure composed of randomly oriented \( \beta \)-sheets (AVSTH-rich segments) in a disordered matrix (GLY-rich segments).](image)
Loligo vulgaris native SRT exhibits thermo-mechanical properties similar to those of elastomeric polymers (i.e. glass transition and rubbery behavior) which agrees with the repetitive segmented sequence hypothesis and the structural characterization presented in this chapter. However, comprehensive quantitative analysis of structure and properties of native materials cannot be performed without difficulty for two main reasons: 1) strong limitations in sample preparation of native materials (i.e. directly extracted from the squid) complicate most measurements since most sample preparation protocols (i.e. solubilization) entail structure modification or damage to the material. Therefore, analyses on the protein complex material that are representative of the biological function are limited. 2) There is a significant diversity in the amino acid sequences. Although the main crystalline/amorphous motif is repeated across all sequenced proteins, the length of the segments, the amino acid composition, the amino acid order arrangement, and the overall polypeptide size vary within proteins (even within the same protein chain). With such diversity, it is challenging to model the material without having made reasonable assumptions. In the following chapters, this material modeling and characterization problem is reduced to a more simplified system, where SRT-inspired synthetic polypeptides with controlled sequence are investigated.
Chapter 4. Semicrystalline morphology and thermomechanical properties of recombinant SRT

In this chapter, a synthetic SRT protein Lv18 (18 kDa) with a defined amino acid primary sequence and molecular weight is studied in detail. Lv18 synthetic protein is processed through solution-based and thermal-based methods yielding protein materials with tailored structure and β-sheet content (up to 54 %). The tailored morphology strongly affects the thermo-mechanical properties, causing an increase in modulus and glass transition temperature with increasing crystalline fraction, and a relationship between the tailored semicrystalline morphology and properties is established. Furthermore, the hydration dynamics of Lv18 synthetic SRT protein were analyzed by neutron spectroscopy, revealing an increase in chain mobility (confined within the β-sheet network) and a decrease in glass transition temperature. The control of the protein crystalline structure and the plasticization of the disordered chains provide ample tailoring of the thermo-mechanical properties of SRT.

4.1 Introduction

Biomimetic materials research is an expanding field that has attracted a growing interest over the past years in the search for high-performance materials and efficient design.\textsuperscript{129,243} Recent research on biological materials has expedited the design of biologically-inspired functional materials and devices with applications in adhesion (gecko footpad,\textsuperscript{244,245} mussel-inspired adhesives\textsuperscript{131}), optics (insect-inspired diffraction gratings),\textsuperscript{246} wetting (plant- and insect-inspired superhydrophobic surfaces),\textsuperscript{247,248} and medicine\textsuperscript{2,94} to name a few. While macroscale concepts are the most developed in biomimetics (they commonly involve systems mimicking structural or physical properties of a natural system such as propulsion or gripping mechanisms),\textsuperscript{249,250} nanoscale biomimetic concepts are based on the assembly of molecules or macromolecular complexes.\textsuperscript{68,251} Bio-derived materials (and biopolymers in particular) hold great promise to provide a broad range of solutions to industrial problems.\textsuperscript{252,253} These materials can be shaped into various geometries such as fibers,
colloids, and thin films, and the synthesis or extraction of biopolymers from biological organisms can provide a new generation of recyclable-engineered materials. A multitude of novel materials remain to be discovered through this approach, which in general involves three phases: (i) discovery to extract genomic information or metabolic pathways; (ii) synthesis or expression of the biopolymer for controlling the materials properties; and (iii) biomimicry of the materials properties identified in the discovery and/or expression phases. Understanding the structures and functional characteristics of biopolymers will accelerate the design, fabrication and synthesis of eco-friendly, recyclable, advanced materials, with novel mechanical, optical, and surface (e.g., wetting, friction, anti-icing) properties. Despite the recent technological advances in biology and materials science, the discovery, understanding and development of novel biological and biologically-inspired materials is a challenging task due to three major problems: (i) the consumption of natural resources and experimentation in animals, including specimen collection and capture; (ii) the need for low cost and scalable means of production; and (iii) the extreme diversity within biological organisms.

Figure 4-1. Scalable production of structural proteins. a) Economy of scales for polymeric materials. The production cost of recombinant protein-based polymers decreases with the production scale up, and this can constitute a cheaper and/or more environmentally friendly alternative for other material types. a) Bioreactor size as a function of structural protein production. The numbers are estimated based on recombinant SRT production.
The investigation of squid ring teeth (SRT) proteins and materials faces the same impediments. Even if one sets aside the inherent ethical aspects of squid capture for obtaining the precious material, the direct extraction of SRT from squid tentacles is limited due to price and natural sources. Global capture production in the major fisheries over the last decade is 2.2 annual million ton approximately (including all major squid species for human consumption).\textsuperscript{261} One can make a rough estimation of the overall cost by considering a 0.5 kg average squid (for example, \textit{Loligo vulgaris}) that can produce 100 mg of SRT.\textsuperscript{260} If SRT were extracted from all captured squids, this would yield an approximated 220 ton annually of SRT production. If an efficient and low cost system for extracting SRT without damaging the rest of the animal were to be designed, the production cost could be approximated to a minimum of a $1 per squid ($0.7/squid and $0.3/squid for current collection and handling price, assuming that the whole squid could be sold for human consumption after the process without any additional cost). This would give an estimated minimum production cost of $10 per gram of SRT by means of direct extraction from the animal. Compared to the production cost of a high end polymer ($10 / kg) and to the large production volume in the polymer industry (300 million tons produced per year globally),\textsuperscript{262} the volume and production costs of SRT by direct extraction are several orders of magnitude inferior and more costly (Figure 4-1). Gram quantities of native SRT can be collected from squids for laboratory-scale research with current extraction methods,\textsuperscript{211} but large-scale production is necessary for economically feasible and sustainable protein-based bioplastic production for engineering and medical applications.

Furthermore, the extremely large diversity within squid species significantly complicates the elucidation of the material sequence, structure, and properties. Squids are members of the order \textit{Teuthida}, which is the larger cephalopod order with over 300 species separated in 29 families.\textsuperscript{260} This broad order of cephalopods include species ranging from a few millimeters in size (such as the pygmy squid, \textit{Idiosepius paradoxus}, with 16 mm of length, or the Hawaiian bobtail squid, \textit{Euprymna scolopes}, with 10 to 30 mm in length) to several meters in size (such as the colossal squid, \textit{Mesonychoteuthis hamiltoni}, with 12 to 14 m in length and up to 750 kg, which is the largest known invertebrate), with hundreds of species with diverse anatomy and morphology in between.\textsuperscript{260} Each squid species have characteristic SRT with distinct size, color, and macro- and nano-morphology (number of teeth, teeth geometry and arrangement, size of nanopores, or external support ring) (Figure 3-1).\textsuperscript{85,255,260} Each SRT is composed of a distinct group of proteins...
ranging between 15 and 60 kDa (Figure 3-2), with different molecular weight distribution across species.\textsuperscript{85,166,255} Furthermore, the group of proteins that form one single SRT in a single squid have variations in amino acid composition, and even each of the single proteins have variations in their segmented sequence (variations in amino acid composition, order, and segment length).\textsuperscript{83,85} Hence, the overall diversity in SRT protein sequences significantly complicates the understanding of the structure and properties of SRT, despite having a common repetitive sequence motif (which is a common problem in other biological materials).\textsuperscript{263}

Figure 4-2. Extraction and expression of SRT proteins. Native SRT protein complexes are extracted directly from squid’s suction cups (upper arrow). Biosynthetic routes (lower arrow) are used to obtain the recombinant and de novo designed proteins with particular molecular weights and sequence. The identified sequences of the protein of interest are produced in the chosen type of expression system, i.e. bacteria or yeast, using genetic engineering toolbox.

Next Generation Sequencing (NGS) in conjunction with high throughput proteomics recently introduced an approach to rapidly identify genes that encode for a specific proteinaceous material candidate and enable the scalable and environmentally friendly production of polypeptides by bacterial expression.\textsuperscript{82} Recombinant expression of biopolymers finds solutions to the challenges of biomimetic research by enabling scalable environment-friendly production of biological materials (Figure 4-2): (i) there is no need for animal capture once the protein has been sequenced. (ii) Recombinant expression promotes the large scale, low cost production of biological materials.
through green processing protocols. Over the past couple of decades, researchers have explored a wide range of expression systems for the high-yield production (up to ~ 1 g/L) of biological materials (including but not limited to silk, elastin, and collagen proteins) such as bacteria,\textsuperscript{21-23} yeast,\textsuperscript{24,25} plants,\textsuperscript{26} mammalian cell lines,\textsuperscript{27} and transgenic organisms.\textsuperscript{28} Genetically modified \textit{Escherichia coli} (\textit{E. coli}) bacteria is the most established suitable host for industrial-scale production due to the availability of expression vectors and well-understood genetics.\textsuperscript{21,29,30} (iii) Polypeptides with a precise and well-defined amino acid sequence are produced by recombinant expression, which provides exceptional control over the self-assembly mechanism, nanostructure, and properties of the material. This greatly simplifies the study of biological materials by overcoming the protein diversity challenge, and therefore single isolated proteins can be synthetized and studied.\textsuperscript{82,83,264}

The processing of long-repetitive polypeptides remains one of the major challenges in developing protein-based materials with predictable properties.\textsuperscript{77,265} In most processing methods, proteins are exposed to severe conditions (temperature, pressure, pH, presence of solvents, etc.) that alter the protein nano/micro-structure and ultimately compromise the final properties of the materials. Consequently, many proteins and biological materials with remarkable properties have not yet been transformed into advanced materials and devices due to limitations in stability and bulk processing. Within this context, SRT proteins are superior to most repetitive proteins since they can be bulk processed by an array of methods analogous to those of the polymer industry.\textsuperscript{83}

In this chapter, the structure and properties of a recombinant SRT protein are comprehensively analyzed. Recombinant 18 kDa SRT protein from \textit{Loligo vulgaris} (Lv18) was expressed in \textit{E. coli} bacteria by Dr. Sergio Florez and Huihun Jung.\textsuperscript{83} Lv18 recombinant protein is processed by thermal- and solution-based methods into materials with varying crystallinity with the intention to replicate the natural properties of native SRT (Chapter 3). The evolution of the structure, dynamics, and thermo-mechanical properties of recombinant protein Lv18 are investigated, revealing a clear predictive relationship between protein morphology and material properties.
4.2 Results and discussion

Recombinant 18 kDa SRT protein from *Loligo vulgaris* (Lv18) was chosen among *Loligo vulgaris* native proteins for this study and was expressed in *E. coli* bacteria. The reason for choosing Lv18 was its small size compared to other SRT proteins. Previous collaborative studies demonstrated the successful expression of a 39 kDa recombinant protein from *Dosidicus gigas* (double size approximately), but the protein failed to fully mimic the properties of native SRT. Selecting a smaller protein such as Lv18 for a comprehensive study, on the other hand, has three major advantages over longer protein candidates: (i) due to the small size, the expression and purification processes are simplified and a higher expression yield is possible; (ii) shorter polypeptides have less number of possible conformations, which significantly simplifies the structural analysis of the protein; and (iii) the amino acid sequence of Lv18 is divided into relatively homogeneous segments (Figure 4-3). The SDS-PAGE in Figure 4-3b shows a slightly higher molecular weight for Lv18 due to the incorporation of a histidine tag (“HHHHHHH”, 840.9 Da) and a thrombin cleavage site (“LVPRGS”, 627.8 Da) peptide sequence at the beginning of the polypeptide chain for purification purposes. Due to this modification in the sequence, the initial 23 amino acids (including His-tag and thrombin site) slightly increase the protein molecular weight by 1.5 kDa (as observed in SDS-PAGE). Further expression and analyses of Lv18 protein were performed without his-tag and thrombin cleavage modifications, and the Lv18 sequence is as reported in Figure 4-3c. The crystalline segments of Lv18 (AVSTH-rich segments) have an average length of 10.3 ± 1.5 amino acids, while the amorphous segments (GLY-rich) have an average length of 16.8 ± 3.1 amino acids. The lengths of the two types of segments are relatively constant if compared to other proteins within *Loligo vulgaris* SRT protein complex (with some amorphous segments up to 82 amino acid long). Therefore, Lv18 is the best candidate within *Loligo vulgaris* SRT proteins for a comprehensive structural analysis, with a total of 6 crystal-forming segments per chain.
Figure 4-3. Native vs. recombinant SRT proteins. a) European common squid (*Loligo vulgaris*) and its native SRT (inset). b) SDS-PAGE of native SRT (group of proteins in the 15 – 60 kDa range) and recombinant Lv18 SRT protein (single protein of 18 kDa). c) amino acid sequence of Lv18 SRT protein. Crystal-forming segments (ATVSH-rich) and disordered segments (GLY-rich) are marked in blue and yellow respectively, separated by proline residues.

Generally, the protein expression steps for recombinant SRT protein include: inoculation and growth of *E. coli* culture, induction of protein expression (from this point, the cells use most of their nutrients for the production of the target protein instead of growing), cell collection and washing, lysis (breaking down the cell membrane and releasing the expressed protein), washing and purification. In the case of Lv18, the protein was successfully expressed by Huihun Jung and Dr. Sergio Florez with 90% purity and a yield of 50 mg/L. The protein content was evaluated in the intermediate expression and purification steps (Figure 4-4). FTIR analysis of the overexpressed *E. coli* reveals slightly higher β-sheet content (48%) compared to purified recombinant 18kDa protein (38%). Typical total protein concentration in *E.coli* is estimated as 190 mg/ml, which corresponds to dry weight of 55% or volumetric composition of 40%. Cell membrane proteins are usually rich in β-sheet structures (β-barrels are present in the outer membranes of Gram-negative bacteria, cell wall of Gram-positive bacteria, outer membrane of
mitochondria and chloroplasts), and therefore a high absorption in the β-sheet region and larger β-sheet content in E. coli compared to purified recombinant film is expected. FTIR analysis on the same empty vector strain of E. coli showed 6% lower β-sheet content (42%) compared to overexpressed E. coli. This difference is attributed to overexpressed SRT protein, which should contribute 6-8% increase to β-sheet content given that overexpression yields 15-20% excess recombinant SRT protein in E. coli.

The purified protein (in this case, Lv18) is usually lyophilized (freeze-dried) and is obtained in a final form of fine powder. However, the protein is not usable in this form for most applications and needs to be processed into defined shapes and materials (e.g. films, nanoparticles, rods, fibers, and other 3D complex geometries). Most protein processing methods consist in the solubilization of a protein in solvent and subsequent aggregation by removal of the solvent via ambient or vacuum assisted evaporation. SRT proteins are stabilized by β-sheet structures that act as physical crosslinks, and are not water-soluble. Therefore, the β-sheet elements must be disrupted in order to solubilize the protein. To this purpose, several pH conditions (acidic pH below 3 and basic pH above 10), salts (lithium bromide, calcium chloride, calcium nitrate, guanidinium chloride, etc.), surfactants (sodium dodecylbenzenesulfonate) and organic solvents (dimethylsulfoxide, hexafluoroisopropanol) can be used. Hexafluoroisopropanol (HFIP) is used as the preferred
solvent to solubilize SRT proteins due to the high solubility of SRT (up to 100 mg/mL). However, these conditions can cause proteins to misfold or become kinetically trapped into undesirable assembly states; hence, the physical properties of these samples may vary significantly compared to native assemblies.

Figure 4-5. Processing of recombinant SRT materials. Recombinant SRT is extracted from bacterial cells and subjected to solution and thermal processing steps that alter the nanostructure and properties of the material.

Lv18 protein is subjected to a series of solution-based and thermal-based processing steps that include: 1) solubilization of SRT in HFIP solvent, 2) solvent casting, 3) water washing (H₂O), 4) hot water washing (H₂O, 70 °C), 5) compression molding (H₂O, 70 °C, 1 MPa), and 6) methanol treatment (immersion in methanol for 24 h) (Figure 4-5). In addition, all materials can be re-dissolved in HFIP at any point of the processing chain due to the physical cross-linking nature of β-sheets in SRT proteins, enabling the recycling of SRT proteins for multiple uses. In this chapter, the evolution of the protein nanostructure and thermo-mechanical properties with the processing steps are evaluated in order to replicate the nanostructure and properties observed in native SRT.
4.2.1 Structural analysis of recombinant SRT proteins

4.2.1.1 Secondary structure content

The secondary structure of Lv18 protein materials (as cast, washed, annealed, compressed, and methanol-treated) was analyzed by FTIR spectroscopy. FTIR spectra of Lv18 (Figure 4-6a) exhibited the major absorption bands commonly observed in proteins: amide I (1600-1700 cm\(^{-1}\), C=O stretching vibration), amide II (1550 cm\(^{-1}\), N-H bending vibration), amide III (1200-1400 cm\(^{-1}\), N-H bending and C-N stretching), and amide A (3300 cm\(^{-1}\), N-H stretching).\(^{151,153}\) All Lv18 samples show similar spectra, except cast samples which show additional bands in the 1000-1300 cm\(^{-1}\) region. These sharp and intense bands at 1100 cm\(^{-1}\) and 1192 cm\(^{-1}\) are mainly originated from C-F\(_3\) stretching and O-H bending vibrations, which indicate that HFIP moieties remain trapped in the protein matrix after the casting process.\(^{270}\) After a water wash, the bands do not show up in any of the other samples and the amide III band can be observed, which indicates that HFIP solvent residues are easily washed away.

![FTIR spectra of recombinant SRT materials](image)

![Amide I band](image)

Figure 4-6. a) FTIR spectra of recombinant SRT materials (cast, wash, annealed, compressed, and methanol-treated. b) Amide I band shows a progressive shift towards the β-sheet region (1620 cm\(^{-1}\)) with processing.
Figure 4-6b shows the amide I absorption band in detail. Amide I is highly dependent on the protein backbone conformation (C=O stretching vibration is affected by the hydrogen bonding pattern), and therefore it is a great indicator of conformational changes in the protein matrix. While cast Lv18 shows a symmetric band centered around 1650 cm⁻¹, the center of the other bands progressively shifts towards lower wavenumbers (red shift) together with a decrease in relative intensity in the high wavenumber region (i.e. from symmetric to asymmetric band). The red shifts in the amide I band suggest an increase in β-sheet structures (1610-1632 cm⁻¹ region) and a decrease in disordered chains (1640-1655 cm⁻¹).

![Deconvoluted amide I FTIR spectra](image)

Figure 4-7. Deconvoluted amide I FTIR spectra. a) Deconvoluted spectra maximum shifts towards the β-sheet region (1620 cm⁻¹) with processing. b) Individual band fitting of cast Lv18 (lowest β-sheet content) and methanol-treated Lv18 (highest β-sheet content). Bands assigned to side chains (sc), β-sheet (β), α-helices (α), and turn (t) structures as described in Table 4-1.

In order to perform a robust quantitative analysis of the secondary structure of the Lv18 materials, the Fourier self-deconvoluted amide I is analyzed in Figure 4-7. A clear shift from ~1650 cm⁻¹ in the cast spectrum to ~1615 cm⁻¹ in the methanol (meOH) sample can be observed in the deconvoluted spectra in Figure 4-7a. Furthermore, an increase in intensity at ~1695 cm⁻¹ in the compressed and meOH samples is observed. This clearly suggests an increase in β-sheet content at the expense of disordered random coil chains. Hence, the spectra indicate that the successive processing steps in Lv18 induce the formation of β-sheets from disordered chains.
Individual Gaussian functions were fitted to the deconvoluted spectra (Figure 4-7b) for a quantitative analysis as described previously (Chapter 2). Each band has been labeled as β-sheets (β), α-helix (α), random coil (rc), turns (t) and side chains (sc) for clear comparison, and the secondary structure content is listed in Table 4-1.

The band centered around 1594 cm\(^{-1}\) (marked as sc) comes from the side chains of the protein (histidine and tyrosine), found between the Amide I (1600-1700 cm\(^{-1}\)) and the Amide II region (1510-1580 cm\(^{-1}\)).\(^{153,155}\) This band does not change with the different processing conditions, which is expected since the content of histidine (10.4 %) and tyrosine (15.6 %) amino acids is constant (same amino acid sequence).

A triplet of bands (marked as β) was fitted to the deconvoluted spectra between 1600 and 1637 cm\(^{-1}\) which are assigned to β-sheets.\(^{151,156–159}\) The band centered around 1612 cm\(^{-1}\) is typically assigned to intermolecular β-sheets formed by molecular aggregation.\(^{157,158,160,229}\) The band centered around 1621 cm\(^{-1}\) is a well-known intermolecular β-sheets band in crystallized proteins.\(^{156,271–273}\) These two bands significantly increase with the successive treatments (H\(_2\)O, temperature, pressure, and meOH), indicating the progressive crystallization of the protein via formation of β-sheet structures. The band centered around 1631 cm\(^{-1}\) has been assigned to the formation of intramolecular β-sheets, and its intensity does not vary much between samples.\(^{152,229}\) Although these bands have been identified in inter- and intramolecular β-sheet structures in other crystallized proteins,\(^{160,161}\) it is perilous to use these bands to differentiate between the two by solely infrared spectroscopy, especially since there is no reason to exclude the possibility of either based on the sequence analysis of Lv18. A small β-sheet band around 1697 cm\(^{-1}\) (splitting of β-sheet bands) only appears in the higher crystallinity samples (compressed and meOH). This band intensifies in parallel with the formation of the other β-sheet bands, and it has been reported in other high-crystallinity proteins such as high-crystalline silk fibroin.\(^{151–153,230}\)

The two bands centered at 1645 cm\(^{-1}\) and 1650 cm\(^{-1}\) (marked as rc) are identified as disordered random coil bands.\(^{151,157,158,160,161}\) While these two bands are the most intense in the cast sample, they steadily decrease intensity upon crystallization and increase of β-sheet elements. The band around 1660 cm\(^{-1}\) (marked as α) is identified as α-helix secondary structures, and it decreases in intensity together with the random coil bands.\(^{157,161}\) The remaining bands in the 1665-1993 cm\(^{-1}\) region are assigned to turns structures (marked as t).\(^{151,156,157,160,162}\) The content of turn structures
remains constant in all samples (within 16-18%). Random coil, α-helix, and turn structures represent the disordered amorphous fraction of the protein chains, and the overall disordered fraction decreases as the successive processing steps induce the crystallization of β-sheets.

Table 4-1. Secondary structure analysis of processed recombinant SRT materials.

<table>
<thead>
<tr>
<th>FTIR analysis</th>
<th>Assignment</th>
<th>Band position (cm⁻¹)</th>
<th>Native</th>
<th>Cast</th>
<th>Wash</th>
<th>Annealed</th>
<th>Compressed</th>
<th>meOH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>side chains¹⁵³,¹⁵⁵</td>
<td>1594-1597</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>β-sheets¹⁵¹–¹⁵³,¹⁵⁶–¹⁵⁹</td>
<td>1600-1632,</td>
<td>49</td>
<td>38</td>
<td>43</td>
<td>45</td>
<td>48</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1697-1703</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>random coerce¹⁵¹,¹⁵⁷,¹⁵⁸,¹⁶⁰,¹⁶¹</td>
<td>1640-1655</td>
<td>22</td>
<td>34</td>
<td>27</td>
<td>25</td>
<td>23</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>α-helices¹⁵⁷,¹⁶¹</td>
<td>1658-1662</td>
<td>10</td>
<td>14</td>
<td>13</td>
<td>12</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>turns¹⁵¹,¹⁵⁶,¹⁵⁷,¹⁶⁰,¹⁶²</td>
<td>1665-1693</td>
<td>19</td>
<td>15</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>18</td>
</tr>
</tbody>
</table>

Note: all secondary structure content calculations are within a ± 2% of maximum error.

When comparing the secondary structure evolution of all processed samples (Table 4-1), the effect of the processing parameters is clear. The addition of water, temperature, pressure, and methanol induce the formation of β-sheets and increase the overall crystallinity. Compressed Lv18 has a very similar secondary structure content to native *Loligo vulgaris* SRT protein complex (Chapter 3), demonstrating that the native morphology can be replicated by synthetic recombinant proteins and consecutive processing. The addition of methanol increases the crystallinity further than the native material (this approach is not used in Nature since methanol and most organic solvents are not present in the squid’s natural habitat). A very similar trend has been previously reported in silk fibroin, which is expected due to their sequence and morphology similarities. In previous reports, the crystallization of silk fibroin with temperature and exposure to methanol was described, producing silk fibroin films with a β-sheet from 0% to 56%.¹⁵² Interestingly, a fully amorphous SRT nanostructure (i.e. 0% β-sheet content) has not been achieved (38% was the minimum achieved). This is attributed to the histidine charged residues in the sequence that drive the aggregation of the chains (isoelectric point close to pH 7).²¹¹ The presence of histidine provide an
advantageous protein assembly mechanism for increased crystallinity in an aquatic organism such as squids since, after all, SRT’s function in Nature is favored by a high-strength crystalline morphology.

On the other hand, cast materials have a more disordered structure with only 38% of a β-sheet content. This could be explained by the presence of HFIP solvent moieties trapped in the matrix (as observed in Figure 4-6a), that hinder the formation of a β-sheets. This would leave the protein structure in a kinetically trapped configuration (or quenched). After HFIP residues are washed away by water molecules, β-sheets can form more easily. With increasing temperature and pressure, more energy is provided to the system and neighboring chains are likely brought closer together, which both enhance the formation of β-sheets.\textsuperscript{152, 274} And finally, the exposure to methanol further enhances the formation of a β-sheets via hydrophobic forces.\textsuperscript{152, 275} Although it is extremely useful, FTIR only can provide with the fraction of secondary structure elements. In order to understand the induced crystallization of Lv18, further analysis of the β-sheet structures is performed by X-ray diffraction.

### 4.2.1.2 β–sheet analysis

Lv18 protein films were analyzed by X-ray diffraction to study the evolution of β-sheet crystallites after exposure to humidity, temperature, pressure, and methanol solvent. Figure 4-8 shows the crystallographic analysis of Lv18 compressed materials, which allows for a more accurate analysis than the native protein equivalent in Chapter 3 (higher resolution data can resolve more diffraction peaks). Miller indices are assigned consistently with literature of SRT.\textsuperscript{84, 85, 166, 167} The major crystalline peaks can be observed at $2\Theta = 9.3^\circ$, $19.6^\circ$ and $24.5^\circ$ corresponding to lattice distances $d_{100} = 9.52 \text{ Å}$, $d_{200} = 4.53 \text{ Å}$ and $d_{002} = 3.64 \text{ Å}$ (according to Bragg’s Law $n\lambda = 2d \sin \Theta$).\textsuperscript{164} Additionally, a weak diffraction peak is observed at $2\Theta = 41.1^\circ$ with lattice distance $d_{240} = 2.2 \text{ Å}$, as well as a broad peak corresponding to the amorphous halo (background scattering from amorphous chains). The measured lattice distances correspond to the hydrogen-bond distance between two β-sheet chains (4.72 Å), the distance between alternating β-sheet chains (9.44 Å, i.e. unit cell dimension in the hydrogen-bond direction fitting two β-sheet chains) and the chain length of a single amino acid in an antiparallel β-sheet structure (3.5 Å, with a two-residue repeat distance
of 7.0 Å). Therefore, the intense peak at \(2\Theta = 19.6^\circ\) is attributed to the combination of (120) and (200) reflections and the peak at \(2\Theta = 41.1^\circ\) to the combination of (240) and (023) reflections.\(^{166}\) The \(\beta\)-sheet crystal structure is fitted into an orthorhombic unit cell similarly to other known \(\beta\)-sheet crystals such as spider and silkworm silk.\(^{276,277}\) Although no (0k0) diffraction peaks are observed in the diffraction pattern, the unit cell dimension \(b\) (amino acid side chain direction) can be calculated from the \(d_{120}, d_{240}\) and \(d_{023}\) spacing values. The unit cell dimensions obtained by the diffraction data are \(a = 9.5\) Å (H-bond direction), \(b = 10.2\) Å (amino acid side chain direction) and \(c = 7.3\) Å (chain backbone direction). The resulting crystal structure for SRT has very similar dimensions to the crystal structure of \(Nephila clavipes\) spider silk, which is classified into the Warwicker system group 3b and has an orthorhombic unit cell with dimensions \(a = 10.6\) Å (amino acid side chain direction), \(b = 9.44\) Å (hydrogen bond direction) and \(c = 6.95\) Å (backbone direction).\(^{276,277}\)

---

**Figure 4-8.** X-ray diffraction (XRD) of recombinant SRT. a) Diffraction pattern of compressed Lv18 shows major diffraction peaks corresponding to alternating hydrogen-bonded \(\beta\)-sheet strands [(100), (200)] and residue alignment along the protein backbone (002). b) Diffraction patterns of processed recombinant SRT proteins show an increasing crystallinity with processing steps (highest for compressed and methanol-treated materials).

It should be mentioned that predicting the dimension in the stacking direction (i.e. amino acid side chain direction) is very complex due to the lack of well-defined peaks. The crystalline segments of synthetic polypeptides are rich in alanine, threonine, valine, serine and histidine amino acids, which increase the complexity in the inter-sheet stacking (especially when incorporating large side
groups such as histidine). It is known that different amino acids in the crystalline chains can lead to varying inter-sheet spacing distances (known as non-periodic lattice crystals) due to the effect of the different side groups.\textsuperscript{278,279} For example, silk β-sheet crystals from different species such as \textit{Nephila clavipes} spider or \textit{Bombyx mori} silkworm have crystal-forming sequences with slightly different repeating units (polyalanine \([\text{Ala-Ala-Ala-Ala-Ala}]_n\) segments or alternating glycine-alanine \([\text{Gly-Ala-Gly-Ala-Gly-Ser}]_n\) segments respectively).\textsuperscript{280} Due to the alternating order of glycine and alanine amino acids, one side of the silk chain is populated by methyl groups (alanine residue) while the other side is not (glycine has no side chain residue). This results in an alternating stacking of the β-sheets, where the alanine faces (stacking in the methyl sides) have a greater inter-sheet separation (5.7 Å) than glycine faces (stacking with no side chain residue) (3.5 Å).\textsuperscript{165,278,279} Therefore, given the complex sequence of crystal-forming segments in SRT proteins, the β-sheets might not stack at all due to the presence of more bulky residues (e.g. histidine and threonine) and the electrostatic repulsion from charged amino acids (histidine). At best, one would expect a very inhomogeneous non-periodic inter-sheet stacking or defects in the crystal structure. For these reasons, it is not clear whether β-sheets stack in an ordered fashion or not, despite the prediction inter-sheet spacing distance from orthorhombic cell calculations. However, the absence of crystalline peaks in the stacking direction suggest that the β-sheets are single layered and do not stack. Additionally, no anisotropy was observed in the diffractograms, suggesting that the β-sheets are randomly oriented.

Table 4-2. β-sheet dimensions and crystallinity analysis of processed recombinant SRT materials

<table>
<thead>
<tr>
<th>Lv18</th>
<th>Crystallinity (%)</th>
<th>H-bond direction (Å)</th>
<th>β strands</th>
<th>Backbone direction (Å)</th>
<th>No. residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cast</td>
<td>34.9</td>
<td>14.7 ± 1.7</td>
<td>3.3 ± 0.4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Wash</td>
<td>38.6</td>
<td>18.7 ± 3.7</td>
<td>4.3 ± 0.8</td>
<td>18.8 ± 7.4</td>
<td>5.4 ± 2.1</td>
</tr>
<tr>
<td>Annealed</td>
<td>40.8</td>
<td>15.0 ± 1.3</td>
<td>3.4 ± 0.3</td>
<td>20.3 ± 6.1</td>
<td>6.0 ± 1.7</td>
</tr>
<tr>
<td>Compressed</td>
<td>43.0</td>
<td>15.3 ± 1.1</td>
<td>3.5 ± 0.3</td>
<td>21.1 ± 4.2</td>
<td>6.0 ± 1.2</td>
</tr>
<tr>
<td>meOH</td>
<td>55.9</td>
<td>25.6 ± 4.1</td>
<td>5.7 ± 0.9</td>
<td>26.9 ± 3.1</td>
<td>7.7 ± 0.9</td>
</tr>
</tbody>
</table>
The crystallinity index of Lv18 materials was calculated by fitting the crystalline and amorphous peaks in the Lorentz-corrected wide angle X-ray scattering (WAXS) spectra.\textsuperscript{163} The crystallinity index was calculated as the ratio of the crystalline area to the total area, and listed in Table 4-2. The crystallinity indexes are extremely close to the β-sheet content determined by FTIR, which validates the measurements. The size of the β-sheet crystals (Table 4-2) of Lv18 was estimated by the analysis of the peak broadening according to Scherrer’s equation, $I(2\Theta) = k \lambda / (L \cos \Theta),^{168}$ where $I(2\Theta)$ is the full width at half maximum of the crystalline peak, $k$ is a constant of proportionality (usually 0.9 to 1 for polymers and cubic symmetries), $\lambda$ is the X-ray wavelength, and $L$ is the crystallite size, giving an average crystallite size of ~ 1.5 x 2 nm (hydrogen bonding direction x backbone direction).

Within these dimensions, a β-sheet crystallite can accommodate 3 - 4 β-sheet strands in the hydrogen bonding direction. Considering that Lv18 has 6 crystal-forming segments in its sequence, each chain will fold into 1 – 1.5 β-sheet crystallites, with inter- and intramolecular β-sheets being possible. The crystallite dimension in the hydrogen bonding direction is constant for all samples except for meOH treated Lv18, which is approximately double in size (~ 6 strands, ~ 4.7 Å of separation between strands). This larger crystals are likely formed by the lateral aggregation of two neighboring β-sheet layers driven by hydrophobic forces during the solvent exchange process (similarly reported in silk fibroin crystals).\textsuperscript{275,281} On the other hand, the dimension along the backbone increases from a barely visible peak in the cast sample to 2.7 nm in the methanol treated samples. This suggests the HFIP solvent residues trapped in the protein from the casting process interfere with the β-sheet formation, and only a few amino acids can form hydrogen bonding with the neighboring strand (possibly a few alanine residues due to their hydrophobicity). As water, temperature, and pressure are introduced, the protein chains gain enough energy and mobility to expand the β-sheets by incorporating adjacent amino acids in the crystal-forming segment.\textsuperscript{274,275,281} The maximum measured dimension along the backbone corresponds to ~ 8 amino acids (~ 3.5 Å per amino acid), which falls pretty closely within the segments limits dictated by the sequence analysis (10 amino acid length between proline residues in crystal-forming segments).

The β-sheet analysis agrees with infrared spectroscopy structural analysis of Lv18, showing that the structure of recombinant SRT proteins can be adjusted by several processing conditions.
(humidity, temperature, pressure, and organic solvents) to match that of native SRT and beyond. Next, the impact of the nanostructure in the materials properties will be studied.

4.2.2 Thermal analysis of semicrystalline SRT materials

Thermal properties of Lv18 as function of β-sheet content were analyzed by temperature modulated differential scanning calorimetry (TMDSC).\textsuperscript{282,283} Calorimetric methods have been extensively used in the analysis of thermal properties of biological materials and proteins in particular.\textsuperscript{16,70,152,284}

TMDSC data of cast Lv18 and compressed Lv18 are shown in Figure 4-9a (cast is the most amorphous recombinant sample, and compressed is the closest to native structure; other data have been omitted from the graph for clarity). Protein samples were previously annealed at 120 °C in order to remove ambient absorbed water (water content after annealing was ~0.2% w/w). Total heat flow and reversing heat flow for both Lv18 recombinant proteins is shown, with endotherms as a downward deflection from the heat flow baseline.

![Figure 4-9](image)

**Figure 4-9.** Temperature modulated differential scanning calorimetry (TMDSC) of processed recombinant SRT materials. a) Total and reversing heat flow of cast and compressed Lv18 show a step change around 160 – 190 °C. b) Specific reversing heat capacity $C_p$ of processed Lv18 exhibits a step change that increase in temperature and decrease in magnitude with processing.
The reversing heat flow shows a step change in both samples between 170 and 200 °C, revealing a glass transition (marked as $T_g$). The step change occurs at higher temperatures in the compressed, which is expected since it is more crystalline. A small endotherm in the total heat flow is observed right after the glass transition that possibly corresponds to a relaxation peak (caused by the previous annealing of the sample for water removal). Sup No melting endotherms were observed, indicating that the β-sheet network is extremely stable. However, thermal degradation of the samples occurred around 200 – 210 °C and above (data shown in figure was cut at the thermal degradation onset). This indicates than the β-sheet melting temperature is higher than the degradation temperature. A similar phenomenon is observed in silk fibroin, where thermal degradation occurs around 200 °C but the melting of β-sheets occurs at 260 °C (high temperature calorimetry beyond 200 °C was possible using fast scanning calorimetry techniques as reported by others). Sup Due to the high interaction strength and stability of β-sheet networks, it is not surprising that β-sheet-rich structural proteins have melting temperatures beyond their thermal degradation (similar to thermoset polymers and elastomers).

The specific reversing heat capacity $C_p$ for the complete Lv18 data set (cast, wash, annealed, compressed, and meOH) are shown in Figure 4-9b. While the reversing $C_p$ is the same for all samples below the glass transition temperature ($T < 150$ °C, $C_p^{glassy}$), the step changes of all samples vary in temperature range and intensity depending on the protein morphology (i.e. β-sheet content). β-sheets act as physical cross-links, restricting the mobility of the disordered segments (similar to the role of lamellar crystals in semicrystalline synthetic polymers). The formation of β-sheets have three consequences on the specific reversing $C_p$ measurements: (1) the increase in reversing $C_p$ decreases with β-sheet formation due to the reduction of the mass fraction of disordered mobile segments (i.e. the higher the β-sheet content, the less chains participate in the glass transition event); (2) the glass transition is broadened with increasing β-sheet content due to the physical constrains of disordered chains imposed by the β-sheet crystals (physical cross-links); and (3) the glass transition temperature $T_g$ progressively shifts towards higher temperatures with increasing β-sheet content, due to the reduced mobility of the chains at higher β-sheet content.
Figure 4-10. a) Schematic for the calculation of the glass transition temperature (T_g) and the increase in heat capacity (ΔC_p). b) ΔC_p linearly decreases with β-sheet content in processed Lv18 materials. Similar results from silk fibroin are plotted for comparison. T_g (inset) increases with β-sheet content.

In order to quantitatively characterize the glass transition relaxation in Lv18 semicrystalline materials, reversing C_p data are analyzed following the schematic in Figure 4-10a. The dashed lines represent the extrapolations of the tangents to the specific reversing C_p curves. The tangent below T_g corresponds to the heat capacity in the solid or glassy state (C_p^{glassy}). A linear fit to the 50 °C - 150 °C region gives an estimate of C_p(T)_{glassy} = 1.196 + 3.08 × 10^{-3} T [J/(g °C)] = 0.350 + 3.08 × 10^{-3} T [J/(g K)]. C_p^{rubbery} is the extrapolation of the tangent to the specific reversing C_p beyond the glass transition (until the onset of thermal degradation), with the same slope as C_p^{glassy}. The increase in heat capacity ΔC_p is calculated as the difference between the glassy and rubbery tangent lines: C_p(T)_{rubbery} = C_p(T)_{glassy} + ΔC_p. The glass transition temperature is calculated as the step change midpoint: C_p(T_g) = C_p(T_g)^{glassy} + ΔC_p/2. These calculations are performed for each sample, and ΔC_p and T_g as function of β-sheet content are plotted in Figure 4-10b. The increase in heat capacity ΔC_p during the glass transition shows a clear linear decrease with increasing β-sheet content due to the reduction of disordered mass fraction participating in the relaxation. A two-phase model with crystalline β-sheet and disordered fractions is fitted to the heat capacity data, giving a linear fit of ΔC_p = 0.539 – 0.617 χ_β [J/(g °C)], where χ_β is the fraction of β-
sheets. In a perfect two-phase model, one would expect $\Delta C_p = 0$ in fully crystalline materials ($\chi_\beta = 1$). However, the proposed model slightly deviates from this extrapolation. A possible explanation for this deviation is the need for a three-phase model that includes a rigid amorphous fraction (which are immobilized non-crystalline chains that are heavily constrained and do not contribute to the specific heat increment, similar to semicrystalline synthetic polymers). Since rigid amorphous chains cannot be detected spectroscopically, the model has to be calibrated with fully non-crystalline materials so the heat capacity increase for a completely amorphous material is known ($\chi_\beta = 0$). This is challenging in SRT proteins since fully disordered materials cannot be easily fabricated (the presence of histidine charged amino acids causes an isoelectric point close to physiological values and rapidly initiate the $\beta$-sheet aggregation). However, previous studies on silk fibroin by Hu et al.\(^{152}\) reported a heat capacity model of $\Delta C_p = 0.475 - 0.494 \chi_\beta$ (plotted in dashed lines in Figure 4-10b), which is within error of the Lv18 calorimetry results. The glass transition temperature, $T_g$ (plotted in the inset), also follows a linear trend with the $\beta$-sheet from 167 °C to 193 °C, which is caused by the confinement and immobilization of disordered chains upon $\beta$-sheet crystallization. The thermal analysis results reported in this section are consistent with those of other structural proteins,\(^{152,284,286-289}\) and clearly show that the thermal properties can be adjusted by controlling the protein morphology and $\beta$-sheet crystallization with varying processing conditions (humidity, temperature, pressure, and exposure to organic solvents).

### 4.2.3 Mechanical properties of semicrystalline SRT materials

The mechanical properties of semicrystalline Lv18 proteins were investigated by dynamic mechanical analysis (DMA) (Figure 4-11a). The glassy moduli of all samples are stable up to 140 °C, where they start slowly decreasing for the less ordered samples. The storage modulus $E'$ is in the 1-2 GPa range for all samples below the glass transition temperature $T_g$. This is not surprising since a moduli of ~1 GPa and ~16-36 GPa and beyond are expected for the glassy amorphous phase and $\beta$-sheet crystalline phase respectively, according to computational and experimental studies for similar materials (silk fibroin and amyloids) from the literature.\(^{102,218,231,232,290,291}\) Depending on the semicrystalline morphology, it is expected that the modulus increases for higher ordered and crystalline samples,\(^{218,232,290}\) as it is observed for the compressed and methanol treated
samples (highest β-sheet content). However, the modulus does not increase beyond the GPa range even at high β-sheet content, which suggests that the amorphous phase is dominant. The lack of β-sheet stacking and the random orientation of the β-sheet crystals are a major difference from silk fibroin materials, and very possibly are the cause for weaker crystals and a lower elastic modulus (compared to silk).\textsuperscript{102,218,232} Annealed samples exhibit a sub-GPa glassy modulus, although it is attributed to macro-defects (cracks and deformations) originated from internal stress during the drying process rather than structural defects.

The glass transition relaxation occurs at higher temperature with increasing crystallinity. Cast Lv18 exhibit a decay in moduli at temperatures as low as \(\sim 140 \, ^\circ C\) (possibly due to the presence of residual solvent moieties), while high crystallinity Lv18 (compressed and methanol-treated samples) have stable moduli up to \(\sim 200 \, ^\circ C\). \(T_g\) was calculated from the maximum \(\tan \delta\), where phase angle \(\delta = \arctan (\text{loss modulus } E'' \text{ / storage modulus } E')\), and plotted together with \(T_g\) determined by calorimetry (DSC) in Figure 4-11b. Both glass transition analysis methods are in agreement with each other, showing a linear dependence of \(T_g\) with crystallinity: \(T_g = 212.5 \, \chi_\beta + 88.3 \, (^\circ C)\), where \(\chi_\beta\) is the β-sheet fraction. The increase in \(T_g\) is caused by the physical constraints imposed by the β-sheet network on the disordered amorphous chains (i.e. disordered chain fluctuations are suppressed by the β-sheet crystals and relaxation processes are shifted to higher
temperatures). This glass transition analysis is consistent with those of synthetic semicrystalline polymers and structural proteins, especially silk fibroin (closest analogous material in structure and properties). However, the semicrystalline morphology of Lv18 and SRT proteins in general is limited on both amorphous and crystalline fraction. Histidine charged amino acids drive the β-sheet assembly, impeding the fabrication of fully amorphous materials and resulting in a minimum β-sheet content of ~35% (obtained by self-assembly of current sequences). On the other hand, amorphous segments, voluminous amino acid side chains, and proline amino acids (known β-sheet disruptors) hinder crystallization beyond ~55-60% of β-sheet fraction. Hence, prediction of \( T_g \) based on the presented experiments is only valid in materials with β-sheet content between ~35% and ~60% (and \( T_g \) between 160 °C and 220 °C). Beyond those limits, SRT-based materials cannot be fabricated by means of self-assembly.

The analysis of the rubbery properties of SRT materials is challenging due to the proximity of the glass transition temperature (\( T_g \)) and the thermal degradation temperature (\( T_{\text{deg}} \), ~ 220 °C), which complicates measuring any rubbery property without causing irreversible damage. Plasticizers such as water or glycerol can be used to swell the protein materials, decrease the \( T_g \), and significantly extend the rubbery state to safe working conditions away from degradation. Therefore, the properties of hydrated (plasticized) Lv18 will be studied in the next section.

### 4.2.4 Hydration of recombinant SRT protein

The interaction between proteins and water molecules plays a key role in defining the structure and dynamics of the system, and the physical properties that arise from them. For example, mechanical properties (elastic modulus, elasticity, and \( T_g \)) of structural proteins, such as silk and collagen, are highly dependent on hydration. The hydrogen bonded network is disrupted by the water molecules resulting in the plasticization of the protein chains. Hydration of structural and globular proteins has been analyzed by an array of characterization methods including mechanical, thermal, X-ray scattering and spectroscopic techniques such as dielectric, FTIR, Raman and NMR spectroscopy. Although major conformational changes between dry and hydrated states have been observed by the methods mentioned above, spectroscopic neutron techniques (e.g. QENS) provide molecular...
level resolution to understand hydration effects on protein structure and dynamics, which is necessary to design novel protein-based materials with multifunctional characteristics and improved mechanical performance. In this section, the protein dynamics and mechanical properties upon hydration will be studied.

4.2.4.1 Water absorption

The water absorption of semicrystalline Lv18 protein films was measured by thermo gravimetric analysis and mass spectrometry (TGA-MS, Figure 4-12). Absorbed water in all samples was calculated from the weight loss between room temperature (20 °C) and 230 °C. In this region, the only detected ions in the mass spectrometer were m/z 17 and m/z 18, which correspond to OH\(^-\) and H\(_2\)O (Figure 4-12b). Moreover, a major weight loss was observed at 250 °C, accompanied by the intense release of m/z 17 (OH\(^-\)), m/z 18 (H\(_2\)O), m/z 44 (CO\(_2\)) and larger molecular weight fragments (short hydrocarbons, not shown in figure). This weight loss corresponds to the thermal degradation of the protein, and the decomposition fragments are the major byproducts in the combustion of organic materials.

Figure 4-12. Water absorption measurements of processed Lv18 materials. a) TGA shows water absorption up to 45% (weight loss at 100 °C). b) Coupled mass spectrometry reveals major water loss in the 25 – 200 °C range, followed by release of CO\(_2\) caused by thermal degradation.
Water content of 5% was measured in all samples at ambient conditions (20 °C, 1 atm, 65% relative humidity), and water content below 1% was measured in all samples after a mild dry annealing process (100 °C for 30 minutes). This was consistent with all samples, indicating that ambient humidity only leads to 5% water uptake independently of the protein morphology (crystallinity), which is the reason behind the similar thermo-mechanical properties in ambient conditions regardless of the protein structure. However, water uptake in saturated conditions significantly increases with the amorphous content (Table 4-3).

Table 4-3. Water absorption (determined by TGA-MS) and β-sheet content of processed Lv18 materials

<table>
<thead>
<tr>
<th>Lv18</th>
<th>Water absorption (%) (w/w)</th>
<th>β-sheet content (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>all (dry)</td>
<td>0.2</td>
<td>-</td>
</tr>
<tr>
<td>all (ambient)</td>
<td>5</td>
<td>-</td>
</tr>
<tr>
<td>meOH</td>
<td>20</td>
<td>55</td>
</tr>
<tr>
<td>compressed</td>
<td>24</td>
<td>45</td>
</tr>
<tr>
<td>annealed</td>
<td>41</td>
<td>43</td>
</tr>
<tr>
<td>wash</td>
<td>45</td>
<td>42</td>
</tr>
</tbody>
</table>

Water molecules are hydrogen bond disruptors and interact with the polar groups of hydrogen bonded solids. In proteins, backbone amide groups are responsible for most of the water uptake, but are also responsible for the assembly of the protein secondary structure. Backbone amides interact with neighboring amides forming β-sheets, α-helices, and other diverse structures depending on the backbone arrangement (including disordered domains). When water molecules are introduced, they disrupt the protein structure by replacing the hydrogen bonds (i.e. water molecules form hydrogen bonds with amides instead). However, not all amide groups are available to water molecules since they might be forming highly stable hydrogen bonded structures such as β-sheets (C=O···H-N). With increasing β-sheet content, the fraction of disordered chains with available amide groups decreases, and consequently so does the number of potential binding sites for water molecules. Since β-sheet structures do not have free amide groups, it is expected
that the water absorption increases with the amorphous fraction, which agrees with the TGA-MS measurements.

### 4.2.4.2 Hydration dynamics of recombinant SRT

Quasielastic neutron scattering (QENS)\textsuperscript{177,178,188,189} is a spectroscopic technique for studying dynamical modes at molecular level with 1 – 30 Å spatial and pico- to nanosecond temporal resolution.\textsuperscript{189} Hence, molecular dynamic processes such as rotations, localized motions, diffusive motions and relaxations can be studied with QENS. Due to the large incoherent neutron scattering cross section of hydrogen atoms (~ 40 times larger than deuterium and much larger than nitrogen, carbon, and oxygen), QENS has the ability to selectively probe the self-dynamics of hydrogen atoms (signal from nonexchangeable hydrogen atoms is dominant). Therefore, the use of deuterated water in QENS protein hydration measurements enables the direct observation of the protein dynamics since scattering from solvent is negligible, making QENS an extremely useful tool for the study of hydration dynamics of biological systems (e.g. in myoglobin,\textsuperscript{180} lysozyme,\textsuperscript{181} RNA,\textsuperscript{182} GFP,\textsuperscript{183} silk,\textsuperscript{184–186} casein,\textsuperscript{187} etc.).

### Elastic incoherent neutron scattering (EINS) – protein mobility

Elastic fixed window scans are one of the most common measurements in neutron spectroscopy and are typically used to estimate the protein mobility over a range of temperatures. Fixed window scans measure the incoherent elastic neutron scattering intensity $I_{el}(Q)$ over a fixed elastic resolution window and over an array of detectors with access to different length scales $Q$ (Figure 4-13a). The sum of elastic intensities over the full $Q$-range is a measure of the non-mobile fraction of hydrogen atoms in the sample (including atoms whose dynamics are slower than the instrumental resolution and appear effectively static). Hydrogen motions faster than the instrumental resolution time window ($\approx$ 1 ns) contribute to the non-elastic scattering fraction and are excluded from the elastic scattering intensity (decrease in $I_{el}(Q)$).\textsuperscript{190} At very low temperatures ($T \approx$ 40 K), all hydrogen atoms in Lv18 protein are static (or moving slower than the instrumental resolution) and the normalized elastic intensity is unity. As temperature increases, hydrogen
motions progressively become faster than the instrumental resolution and fall outside the $I_{el}(Q)$ measurement window. As consequence the elastic scattering intensity decreases with temperature. While the elastic scattering intensity of dry SRT protein steadily decreases with temperature, lower intensity is measured for hydrated SRT at $T > 280$ K which indicates higher mobility upon hydration.

Moreover, the Q-dependence of the elastic scattering intensity can provide information about the length scale of the hydrogen motions. Mean square displacements (MSD) of mobile hydrogen atoms can be estimated from the elastic scattering intensity as function of scattering wave vector $Q$ and temperature $T$ through the Debye-Waller Factor (DWF):

$$DWF = \frac{I_{el}(Q, T, \omega = 0)}{I_{el}(Q, T = 40 K, \omega = 0)} = \exp \left[ -\frac{1}{3} Q^2 \langle x^2 \rangle \right]$$

Where $I_{el} (Q, T, \omega = 0)$ and $I_{el} (Q, T = 40 K, \omega = 0)$ are the elastic scattering intensity at temperature $T$ and 40 K respectively. This Gaussian approximation is the standard method for estimating the MSD, but it assumes that all mobile hydrogen atoms undergo the same isotropic motion.\textsuperscript{194,318} Hence, the approximated MSD includes all kind of motions (vibrations, rotations, diffusion, etc.) as a single value over a range of temperature. MSD of hydrogen atoms as function of temperature

Figure 4-13. Elastic fixed window scan. a) Normalized Elastic Incoherent Scattering Intensity of dry and hydrated recombinant Lv18 protein as function of temperature. b) Mean Square Displacement (MSD), $<x^2(T)>$, of dry and hydrated recombinant Lv-SRT18 protein as function of temperature.
in dry and hydrated Lv18 protein were estimated from EINS data and are shown in Figure 4-13b. Two onsets are observed in both dry and hydrated samples: (i) at very low temperatures (40 – 100 K), both dry and hydrated samples show a plateau around zero, revealing a very low mobility and no temperature dependence in the $T < 100$ K temperature range. The first onset is observed at $T \sim 100$ K for both samples, which is common for proteins and is typically attributed to the onset of methyl group rotations. This suggests that dry Lv18 dynamics are dominated by methyl group rotations from 100 K to 460 K (constant slope with $d<x^2>/dT \approx 1.7 \cdot 10^{-3}$ Å$^2$/K). (ii) A second onset is observed for dry Lv18 at 470 K corresponding to the glass transition (also called dynamic transition in neutron scattering literature) of the protein (with $d<x^2>/dT \approx 6.2 \cdot 10^{-3}$ Å$^2$/K), which involves increased mobility due to backbone motion above the $T_g$.

Hydrated Lv18 shows similar behavior to dry SRT in the 40 - 260 K temperature range: extremely low mobility in the 40 – 100 K range followed by an increase of $d<x^2>/dT \approx 1.7 \cdot 10^{-3}$ Å$^2$/K in the 100 – 260 K range due to methyl group rotations (onset at $T \sim 100$ K), revealing that methyl group rotational dynamics are independent of hydration (as previously reported for other proteins). A significant mobility increase is measured at 270 - 280 K caused by the melting of D$_2$O ice. In the 280 – 320 K temperature region (with $d<x^2>/dT \approx 6.4 \cdot 10^{-3}$ Å$^2$/K), hydrated SRT shows increased mobility compared to the dry SRT (methyl rotation dominated), suggesting the presence of an additional dynamic mode that will be investigated below. An additional onset corresponding to the glass transition is observed at $T \sim 325$ K, which is highly dependent on hydration and reveals high mobility ($d<x^2>/dT \approx 20.7 \cdot 10^{-3}$ Å$^2$/K) in the 330 – 350 K temperature range.

Despite the limitations (Gaussian approximation), MSD estimation from EINS data is a useful measurement to compare the mobility of different systems and to observe the temperature onsets of the dynamic modes of proteins, and has been extensively used as such. MSD analysis proves very useful as preliminary measurement to find the temperatures of interest for further analysis by quasielastic scattering. Below, the hydration dynamics of Lv18 at 295K (temperature of interest) are studied.
Protein chain dynamics - QENS analysis

After the elastic fixed window scans, the dynamics modes in the 100-460 K and 280-320 K for dry and hydrated SRT respectively were selected as our regions of interest for studying the protein dynamics by quasielastic neutron scattering analysis. QENS measurements of dry and hydrated Lv18 were performed in backscattering (HFBS) and time-of-flight (DCS) spectrometers respectively. Although HFBS is able to measure the dynamics of dry SRT, additional dynamics upon hydration were not observed in HFBS due to the limited dynamic range of the instrument (±15 μeV).\textsuperscript{190} DCS, which has a wider dynamic range (± 0.35 meV), was able to measure faster dynamic processes and therefore it was used to measure the faster hydrated SRT dynamics.\textsuperscript{191}

![Figure 4-14](image.png)

Figure 4-14. QENS spectra of (a) dry Lv18 (measured in HFBS) and (b) hydrated Lv18 (measured DCS). The gray solid lines are experimental data, including error bars, and the black solid lines represent the elastic component (resolution). Dry and hydrated spectra were fitted with one and two Lorentzian functions respectively (red and blue).

The QENS spectra of dry and hydrated recombinant Lv18 protein is shown in Figure 4-14. Figure 4-14a shows the quasielastic scattering for the dry sample measured on the HFBS. A single Lorentzian was fitted to the dry Lv18 QENS spectra in addition to the elastic component (delta function convoluted with instrumental resolution). A small broadening of the central peaks from the elastic component, which is consistent for all $Q$ values, suggests small mobility of hydrogen
atoms in the dry state. QENS spectra for hydrated Lv18 measured in DCS are shown in Figure 4-14b, exhibiting a larger broadening from the elastic component. This larger quasielastic scattering in the hydrated sample suggests a higher mobility of the hydrogen atoms compared to the dry state. One Lorentzian was insufficient to describe the broadening of hydrated films QENS spectra and therefore two Lorentzian functions were fitted: 1) a narrow Lorentzian (which is $Q$-independent, similar to that observed on HFBS from dry Lv18) and 2) a broader Lorentzian (faster process, which broadens with increasing $Q$ values). The need for two Lorentzians indicated that there are two superimposed dynamic processes occurring in the hydrated sample, which agrees with the MSD data analysis from previous elastic fixed window scan measurements.

![Graph](image)

Figure 4-15. QENS analysis of dry and hydrated Lv18 protein. (a) Full width half maximum $\Gamma(Q)$ of the (Lorentzian functions) as function of $Q^2$. (b) Elastic incoherent structure factor EISF of dry and hydrated Lv18 as function of $Q$. A confined diffusion model (diffusion of disordered chains) and a 3-site jump diffusion model (methyl group rotation) are used to describe the protein dynamics.

Detailed analysis of the quasielastic broadening is provided in Figure 4-15a, where the full width half maximum $\Gamma(Q)$ is plotted as function of $Q^2$ for both dry and hydrated SRT samples (one and two Lorentzians respectively). First, a constant full width half maximum $\Gamma_0$ is observed in the dry sample around $\sim22 \mu eV$ over the measured $Q$ range. This $Q$-independent behavior is
characteristic of localized motions in the time scale resolution window. Considering the localized nature and the temperature onset (MSD analysis), this dynamic process is be attributed to methyl group rotations. Similarly, the narrow Lorentzian of the hydrated sample shows a constant $\Gamma_0$ of $\sim 30$ $\mu$eV over the measured $Q$ range. The measurement of this two $Q$-independent processes in both HFBS and DCS suggest that the same dynamic mode (methyl group rotations) is occurring in both dry and hydrated samples (hydration independent). These results agree with the elastic fixed window scan analysis and are consistent with previous analysis on Green Fluorescence Protein (GFP) where similar $\Gamma(Q)$ were attributed to methyl group rotations in both dry and hydrated samples.\(^{183}\)

On the other hand, the faster process of the hydrated sample (broad Lorentzian) exhibits a linear increase in $\Gamma(Q)$ with $Q^2$ (simple diffusion $Q^2$ – dependence is plotted in figure 5a for comparative purposes). At low $Q$ values ($Q^2 < Q^*^2$), the data suggests the presence of a $Q$-independent plateau, while a linear scaling with $Q^2$ is observed at $Q^2 > Q^*^2$, where $Q^*^2$ is the crossover point. This 2-regime $Q$-dependence is characteristic of diffusion in a confined space and can be described by the Volino and Dianoux (VD) model for bounded diffusion in a potential of spherical symmetry.\(^{177,195}\)

According to the VD model in which a confined particle is diffusing in an impermeable sphere of radius $r_{conf}$, no length-scale dependence can be observed for length scales larger than a confinement radius $r_{conf}$ (hence the $\Gamma_0$ plateau at low $Q$ values), while simple diffusion can be observed for length scales smaller than $r_{conf}$ (higher $Q$ values):

\[
Q^* = \frac{\pi}{r_{conf}} \quad \Gamma_0 (Q < Q^*) = 2D \frac{\pi^2}{r_{conf}} \quad \Gamma (Q > Q^*) = 2DQ^2
\]

(6)

However, not enough data points at low $Q$ values are available in order to establish a clear crossover point $Q^*$, resulting in an inaccurate estimation of $r_{conf}$ from $\Gamma(Q)$ data. For this reason, the elastic incoherent structure factor (EISF) as function of $Q$ is analyzed in Figure 4-15b, which is defined as the ratio of elastic intensity to the total intensity, and provides information on the geometry of the motion and the mobile fraction of hydrogen atoms involved. EISF can be modeled for multiple types of hydrogen motion, and a three-site jump model that describing methyl group rotations is used: \(^{177}\)

\[
EISF_{methyl}(Q) = \frac{1}{3} [1 + 2j_0(Qr\sqrt{3})]
\]

(7)
Where $j_0$ is the zero order Bessel function and $r$ is the rotation radius of a methyl group ($\approx 1$ Å). Therefore, the EISF of dry SRT can be written as:

$$ EISF_{dry}(Q) = 1 - p_{methyl} + p_{methyl}EISF_{methyl}(Q) $$

(8)

Where $p_{methyl}$ is the fraction of mobile hydrogen involved in methyl group rotations, and is weighting the $EISF_{methyl}$ term. The term $1 - p_{methyl}$ represents the elastic intensity from non-methyl atoms (immobile fraction) and $p_{methyl}EISF_{methyl}(Q)$ represents the elastic intensity coming from methyl groups. Despite a slight deviation caused by the small quasielastic broadening and high noise at low Q values in the QENS spectra, this methyl rotation model agrees with the methyl group onset observed in the dry SRT MSD data (elastic fixed window scans).

As suggested by the MSD and QENS analysis above, the hydrated SRT dynamics at 295 K include methyl group rotations as well as the bounded diffusive motion within a sphere. Therefore, a localized spherical diffusion model is considered for the hydrated SRT dynamics:

$$ EISF_{sphere}(Q) = \left[ \frac{3j_1(QR)}{QR} \right]^2 $$

(9)

Where $j_1$ is the first order Bessel function and $R$ is the radius of a sphere in which the diffusion is confined. Assuming that the contribution from methyl group rotation and localized spherical diffusion are additive, the EISF of hydrated SRT $EISF_{hydrated}(Q)$ can be calculated as the combination of methyl group rotation and bounded diffusion within a sphere models. This combination of the two models has been successfully used in previous studies of hydrated green fluorescence protein and lysozyme, and it satisfactorily agrees with the hydrated SRT data. However, it should be noted that SRT is a non-water soluble semicrystalline structural protein rather than a soluble globular protein, and therefore constraints for the confined diffusion are most likely imposed by secondary structure elements (as discussed below). As result of the EISF model fit for the dry and hydrated SRT sample, the fraction of hydrogen atoms involved in the two dynamic modes and the length-scale of the localized diffusion are estimated: $p_{methyl} \approx 0.35$, $p_{sphere} \approx 0.47$ and $R_{sphere} \approx 6$ Å.
Table 4-4. Calculation of hydrogen fraction involved in the protein dynamics

<table>
<thead>
<tr>
<th>AA with methyl groups</th>
<th>no. of AA in Lv18</th>
<th>no. methyl H per AA</th>
<th>no. methyl H in Lv18</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>26</td>
<td>3</td>
<td>78</td>
</tr>
<tr>
<td>L</td>
<td>23</td>
<td>6</td>
<td>138</td>
</tr>
<tr>
<td>T</td>
<td>16</td>
<td>3</td>
<td>48</td>
</tr>
<tr>
<td>V</td>
<td>10</td>
<td>6</td>
<td>60</td>
</tr>
<tr>
<td>I</td>
<td>2</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>M</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Total methyl H</td>
<td></td>
<td></td>
<td>339</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AA in amorphous segments</th>
<th>no. of AA in Lv18</th>
<th>no. of H per AA</th>
<th>no. of H in Lv18</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>46</td>
<td>2</td>
<td>92</td>
</tr>
<tr>
<td>Y</td>
<td>30</td>
<td>8</td>
<td>240</td>
</tr>
<tr>
<td>L</td>
<td>23</td>
<td>10</td>
<td>230</td>
</tr>
<tr>
<td>Total G, Y, L H</td>
<td></td>
<td></td>
<td>562</td>
</tr>
<tr>
<td>Total Lv18 H</td>
<td></td>
<td></td>
<td>1123</td>
</tr>
</tbody>
</table>

Methyl H fraction 0.30  
$p_{\text{methyl}}$ (experimental) 0.35  
Amorphous H fraction 0.50  
$p_{\text{sphere}}$ (experimental) 0.47

The fraction of hydrogen atoms involved in the dynamic modes can be predicted by analyzing the amino acid sequence of the recombinant Lv-SRT18 protein (Table 4-4).\textsuperscript{83,84} The amino acids A, L, T, V, I and M (with an abundance in Lv-SRT18 of 13.3%, 11.8%, 8.2%, 5.1%, 1% and 0.5% respectively) contain one or two methyl groups per amino acid residue. The methyl fraction has been calculated as the ratio of number of hydrogen atoms in methyl groups (weighted by number of groups per residue and individual amino acid abundance) to total number of hydrogen atoms per protein chain, resulting in a predicted methyl fraction $p_{\text{methyl}}' = 0.30$ that agrees with the EISF fitting results. Similarly, the hydrogen fraction involved in the confined diffusive dynamics in the hydrated protein films is predicted from the amino acid sequence analysis. The overall hydrogen atoms in the amorphous segments (from G, Y and L amino acids) is calculated giving a fraction of $p_{\text{amorphous}} \approx 0.50$, which agrees with the measured hydrogen fraction of $p_{\text{sphere}} \approx 0.47$ that exhibit diffusive motion within a confined space upon hydration. The agreement between the predicted and experimental amorphous fraction confirms that only the amorphous segments of the SRT
backbone are responsible for the increased mobility of hydrated SRT, and that the \( \beta \)-sheet crystals are stable in hydrated conditions. However, the mobility of the amorphous chain segments is limited due to the physical constraints imposed by the crystalline \( \beta \)-sheet domains and, as a consequence, they can only diffuse within a confined volume between the \( \beta \)-sheet nanocrystals. Therefore, the effect of the protein semicrystalline morphology has strong effect in the confined chain dynamics of Lv18. These results, which are consistent with those of other structural proteins and polyamides, highlight how the segmented architecture and morphology of SRT proteins is critical in the selective hydration of the system, and will strongly affect the physical properties of the material that arise from the protein secondary structures.

### 4.2.4.3 Mechanical properties upon hydration

The mechanical properties of hydrated Lv18 were analyzed by dynamic mechanical analysis (DMA) and rotational rheometry in completely saturated water environments (immersion fixtures). Both storage and loss moduli are in the MPa – Gpa region at room temperature (Figure 4-16a), which indicates that the \( \beta \)-sheets are stable in water and that the physically cross-linked network is conserved. However, samples with different \( \beta \)-sheet content absorb different amounts of water as discussed previously, and this significantly affects the mechanical properties. The storage modulus at room temperature drops significantly for the washed and annealed samples, which have the lowest crystallinity (42\% and 43\% respectively), highest water absorption (45\% and 41\% absorbed water), and therefore the highest disordered chain mobility. On the other hand, compressed and meOH LV18 samples have storage moduli in close to 1 GPa and have the highest crystallinity (45\% and 55\%), lowest water absorption (24\% and 20\%), and the most constrained disordered domains. As temperature increases, a steady decrease in moduli is observed due to the glass transition event (maximum tan \( \delta \)). Water molecules plasticize the amorphous fraction and increase the disordered chain mobility, and therefore glass transition events occur at much lower temperatures (from \( T_g \text{ dry} \approx 160 \text{ - } 200 \text{ °C} \) to \( T_g \text{ plasticized} \approx 20 \text{ – } 50 \text{ °C} \)).

The observed decrease in glass transition agrees with Couchman-Karasz’s empirical approximation for binary mixes:

\[
T_{g \text{ plasticized}} = \frac{w_{protein}T_{g \text{ protein}} + k_wT_{g \text{ water}}}{w_{protein} + k_w}, \quad k = \frac{\Delta c_{p \text{ water}}}{\Delta c_{p \text{ protein}}} \quad (10)
\]
Where $T_g$ is the glass transition temperature, $w$ is the mass fraction, and $\Delta C_p$ is the increment in specific heat of protein and absorbed water respectively. Couchman-Karasz’s approximation predicts a range of plasticized $T_g$ from -41 °C to 36 °C. However, it does not account for the crystalline fraction which constrains the amorphous chains and results in an increased glass transition temperature for measured Lv18 materials.

Figure 4-16. Mechanical properties of hydrated processed Lv18 materials. a) DMA analysis shows a decrease in $T_g$ due to the plasticization of the amorphous chains. b) Dynamic rheology shows a frequency-independent rubbery modulus around 1 MPa.

The rubbery moduli of Lv18 was measured by rotational rheometry above the $T_g$ (immersed in 70 °C water, Figure 4-16b). In all samples, a very weak frequency dependence was measured above 1 rad/s (due to the relaxation of dangling ends and network defects) and a frequency-independent plateau was measured below 1 rad/s (equilibrium moduli), which indicated that the β-sheet physically cross-linked network is stable even at high temperatures and low frequencies.242 Due to measurement conditions (humidity temperature, and pressure), washed and annealed processing conditions are equivalent to those of compressed Lv18, and consequently the equilibrium modulus is 0.7 MPa. MeOH Lv18, which has higher crystallinity (i.e. higher cross-linking degree), exhibits a slightly higher modulus of 1.0 MPa. The average cross-linking strand can be estimated using entropic elasticity models such as the affine network model:242,323,324
\[ G = \nu kT = \frac{\rho RT}{M_s} \left( \frac{\phi}{\phi_0} \right)^{1/3} \]  

(11)

Where \( \nu \) is the number density of strands (number of network strands per unit volume), \( k \) is Boltzmann constant, \( \rho \) is density, \( R \) is the gas constant and \( M_s \) is the average molar mass of a network strand. \( (\phi/\phi_0)^{1/3} \) is a correction factor accounting for the network swelling, where \( \phi_0 \) and \( \phi \) are the initial and equilibrium swelling protein volume fraction respectively. Swelling in \( \theta \)-solvent is considered in this work for simplicity, although swelling in good solvent might be a suitable alternative for the model (swelling and mechanical properties of SRT in different solvents and plasticizers will be the focus of follow-up work). The affine network model, as in most entropic elasticity models, considers that the contribution to the modulus is \( kT \) per strand. However, the affine model only considers a perfect end-to-end cross-linked network and ignores the effect of network defects and entanglements. In order to estimate the contribution from entanglements, the modulus of Lv18 was measured in 8M urea solution. Urea is a known hydrogen bonding disruptor, and it is extensively used in protein unfolding protocols (urea disrupts secondary structure elements such as \( \beta \)-sheets and \( \alpha \)-helices, and protein chains adopt disordered random coil conformations). The moduli of denatured Lv18 (without \( \beta \)-sheet structures) shows a very similar network behavior, with a lower modulus of 0.5 MPa. This suggests that the contribution from entanglements is significant, and entanglements act as network strands. The affine network model estimates network strands of \( M_s \) Lv18 compressed = 5018 g/mol, \( M_s \) Lv18 meOH = 3513 g/mol, and \( M_{\text{entanglement}} \) Lv18 urea = 5549 g/mol. If assuming an average amino acid molecular weight of \( M_{\text{avg, AA}} = 96.8 \) g/mol (calculated from the sequence), the network strands correspond to \( M_s \) Lv18 compressed = 51.7 AA, \( M_s \) Lv18 meOH = 36.2 AA, and \( M_{\text{entanglement}} \) Lv18 urea = 57.1 AA (where AA denotes amino acids). These calculations are higher than expected if one considers that the average length of amorphous segments is 17 AA. This underestimation of the affine model suggests that Lv18 materials are rich in network defects (such as dangling ends and loops) that do not contribute to the modulus ad are not accounted for in the model. Although the affine network model serves as a first evaluation, it does not accurately describe the protein network structure. Other advanced entropic elasticity models (phantom network, Flory’s dangling end correction, real elastic network theory, etc.) include a number of corrections for structural network defects. However, neither Lv18 nor native SRT proteins are appropriate for a detailed mechanics analysis due to their variation in amino acid sequence and molecular weight. For this reason, a better protein
polymeric system with a well-defined repetitive sequence is necessary in order to understand the self-assembled structure of SRT.

4.3 Conclusions

In this chapter, the structure, chain dynamics, and thermo-mechanical properties of synthetic Lv18 recombinant proteins in dry and hydrated conditions were analyzed. Obtained in a powder form from the bacterial expression process, Lv18 was processed thermally and with organic solvents into materials with varying β-sheet content (from 38 % to 55%). This processing methods highlighted that the controlled exposure to humidity, temperature, pressure, and methanol regulate the crystallization of β-sheets, enabling the fabrication of SRT-based materials with programmable morphology. Protein chain dynamics of Lv18 were investigated by quasielastic neutron scattering, revealing the confined diffusion of Lv18’s amorphous segments within a semicrystalline network. β-sheet crystallites act as physical cross-linkers that constrain the disordered chains, and therefore regulate the chain mobility and the properties that arise from it. The thermo-mechanical properties of Lv18 were modified by adjusting the β-sheet content. Lv18 materials exhibited a glass transition temperature ranging from 160 °C to 210 °C with increasing β-sheet content, and glassy and rubbery moduli of ~ 1 GPa and ~ 1 MPa respectively. Therefore, the morphology and properties of synthetic Lv18 recombinant proteins can be adjusted to match those of native SRT (including *Loligo vulgaris*) by several thermal and solvent methods. The successful replication of native properties and structure eliminates the need for squid capture, and justifies the production of designer SRT proteins by bacterial expression. However, Lv18 protein has a defect-rich network structure (determined by rheological analysis) which causes disparities between the mechanical properties of recombinant and native proteins when subjected to large deformations. These observations suggest that although the semicrystalline morphology is important, it is not the sole defining factor of the physical properties of SRT. Other factors such as molecular weight are likely to play a decisive role in the SRT self-assembly process, structural defects and material properties. The defects responsible for the mechanical disparity cannot be analyzed in Lv18 due to its internal variation in the amino acid sequence, and more homogeneous repetitive sequence is required. In order to solve this problem, a series of repetitive tandem proteins are developed and analyzed in Chapter 5 and 6.
Chapter 5. SRT-inspired tandem repeat proteins: network structure and mechanics

In this chapter, SRT-inspired synthetic tandem repeat proteins with a controlled and repetitive segmented sequence are investigated. SRT-inspired tandem repeat proteins share the same disordered/crystalline building blocks but vary in length (i.e., varying total number of building blocks per protein chain), and assemble into physically cross-linked networks composed of β-sheet crystallites connected to each other via disordered tie chains. However, intramolecular interactions can originate disordered loop structures and introduce defects in the network structure. A network model that quantifies the network structure (effective vs. defective strands) is proposed. The model accurately describes the behavior of SRT materials under both small and large deformations, and is been validated with experimental measurements. Shear modulus, tensile modulus, and toughness increase with the number of repeat units in the protein chain as a result of a higher fraction of elastically effective strands.

5.1 Introduction

Naturally occurring proteins are optimized for diverse physiological functions as result of fine-tuning of properties through years of evolution. In natural occurring structural proteins, convergent evolution has resulted in diverse molecular mechanisms and structures that provide a variety of mechanical functions (elasticity, resiliency, load-bearing, and energy dissipation). Elastin provides elasticity to vertebrate animal tissue such as blood vessels, lung, and skin. Resilin provides efficient energy conservation and resilience in insect wings and jumping appendages. Silk proteins provide though and strong support for spider webs (predatory) and silkworm cocoons (protective). These structural proteins share common motifs in their amino acid sequence, or common “building blocks”. Elastin contains repetitive [VPGVG]ₙ that form hydrophobic disordered and elastic domains, that are cross-linked by lysine-rich hydrophilic domains. Resilin is composed of glycine-rich disordered domains cross-linked via tyrosines. Silk is composed of repetitive disordered glycine-rich domains and hydrophobic alanine-rich domains that assemble into semicrystalline fibers throw a natural drawing process. These tandem repetitions (TR) in amino acid sequence are advantageous because of the intrinsic
promotion of stability through the periodic recurrence of favorable interactions in the protein structure. Tandem-repeat (TR) proteins, found throughout the tree of life and in all eukaryotes, feature a modular design in which a sequence motif encoding 20-40 amino acids is duplicated a handful or scores of times in a single open-reading frame to yield a full-length gene. TR proteins exhibit a wide range of structures and functions, from soluble forms such as ankyrins and HEAT-repeats that serve to bind other biomolecules, to structural fibers such as collagens and silks. Tandem-repeat arrays are thought to expand and contract (typically at rates between 10^{-2} and 10^{-6} per generation) due to polymerase slippage during genome replication, leading to variations in repeat number that are subject to natural selection. For soluble TR proteins, the influence of repeat number on biophysical properties such as thermodynamic stability and binding affinity has been investigated extensively. The relationship between physical properties and tandem repeats in fibrous proteins has been studied as well, although there are no clear established rules due to the high diversity in repeating unit sequences and self-assembly mechanisms. For this reason, repetitive proteins have served as rich inspiration for biomimetic materials design towards the understanding and control of the physical properties of protein-based materials.

As described in Chapter 3, squid sucker ring teeth (SRT) proteins are composed of a group of proteins with varying molecular weight in the 15 – 65 kDa range. These proteins have a segmented design, with alternating disordered GLY-rich amorphous segments and crystal-forming AVSTH-rich segments that assemble into β-sheet-rich networks. Although the general motifs are observed across all studied squid species to date, the molecular weight, the amino acid content, and the length of the building block segments are very diverse even within the same species. This diversity complicates the analysis of the nanostructure and properties of native SRT (Loligo vulgaris in particular, Chapter 3), and for this reason recombinant DNA technology was used to express synthetic SRT proteins (Chapter 4). However, natural sequence diversity within the same recombinant protein hinders the comprehensive analysis of sequence, structure, and properties. Hence, a new design strategy for the expression of SRT-inspired polypeptides with precise control of the sequence, segment length, and molecular weight is required. Recombinant DNA technology has been successfully used in the synthesis of tandem repeats of naturally occurring peptides. However, current methods for DNA polymerization have major limitations: (i) they require multiple sequential steps, (ii) they cannot be run in parallel,
and (iii) they do not offer precise tunable control over a range of molecular weights.\textsuperscript{354} The synthesis of high molecular weight repetitive sequences is complicated due to genetic instability,\textsuperscript{355–357} and researchers often opt for protein cross-linking from a tandem repeat monomer (which does not solve the problem since introduces additional defects in the protein structure such as cyclic chains, and does not necessarily represent the native structure).\textsuperscript{18,109,347}

![Diagram of SRT-TR repeat unit (building block)](image)

Figure 5-1. a) Building block sequence inspired in SRT. b) Rolling circle amplification (RCA) strategy to construct SRT-inspired polypeptides with varying number of building blocks.

Overcoming these limitations, rolling circle amplification (RCA) offers a one-step method to synthetize repetitive proteins from a DNA monomer with precise control over the number of repeats.\textsuperscript{354} Recently, Dr. Benjamin D. Allen and Huihun Jung used protected digestion rolling circle amplification (PD-RCA) to synthetize a library of squid ring teeth – tandem repeat (SRT-TR) proteins with controlled number of repeat units (Figure 5-1), which is summarized here.\textsuperscript{85} A DNA sequence encoding for a SRT-inspired “monomer” was constructed based on consensus sequences derived by inspection of the native SRT proteins of several squid species: \textit{Loligo vulgaris}, \textit{Loligo pealei}, \textit{Todarodes pacificus}, \textit{Euprymna scolopes}, \textit{Dosidicus gigas}, \textit{Sepioteuthis lessoniana}, and \textit{Sepia esculenta}.\textsuperscript{82,85,166} The “monomer” sequence consisted of a crystal-forming segment of PAAASVSTVHHP, and a disordered segment of STGLSYGYGGGLYGYGGGLYG (Figure 5-1). The “monomer” DNA construct was digested and circularized (i). The circularized
“monomer” DNA is used as template as polymerase rolls around it, forming random RCA products (linear oligomers, ii). The RCA products are digested, yielding a library of TR products comprised of an integer number of repeats of the TR “monomer” gene (iii). The TR products are separated by size via electrophoresis, and specific TR DNA oligomers can be selected by direct extraction from the electrophoresis matrix (iv). The selected DNA oligomers are then ligated to an expression vector to create an expression library for TR protein synthesis (v). Hence, this method can generate protein libraries comprising TR polypeptides (with the same building block sequence) with a specified number of repeat units.

The SRT-TR polypeptides synthetized by PD-RCA are an ideal set to study the assembly, structure, and properties of SRT materials because of their precisely controlled structure. Here, the effects of tandem repetition on the structure and mechanical properties in a family of elastomeric proteins inspired from SRT are studied, showing that materials performance depends strongly on repetition number. Furthermore, the results reported here provide insights on the optimization of mechanical properties in proteins through gene duplication, and offer a possible rationale for the polydispersity observed in naturally occurring SRT proteins.

5.2 Results and discussion

5.2.1 SRT-inspired tandem repeat sequence

In this chapter, synthetic SRT-inspired tandem repeat (TR) polypeptides were designed and produced by bacterial expression. Similar to native SRT proteins, these polypeptides are made such that they comprise ordered crystalline and disordered amorphous domains, which contribute to their mechanical properties. However, the synthetic polypeptides differ from their natural counterparts by having a distinct repeating primary amino acid sequence (i.e., while native SRT proteins have significant variations in amino acid segment length, order, and composition, the repeating sequence in TR polypeptide is conserved). In general, the design strategy of repetitive sequences uses three parameters to modulate the properties of synthetic proteins (which in this work are semicrystalline): (i) the amino acid composition of the crystalline/ordered or amorphous regions, usually inspired by sequence motifs in naturally occurring proteins, (ii) the length and
fraction of the amorphous and crystalline regions, and (iii) the number of repeat sequence units $n$ (i.e., number of tandem copies of the amorphous + crystal-forming units). The design strategy was based on the crystal-forming polypeptide sequence of PAAASVSTVHHP, and the amorphous polypeptide sequence of STGTLSYGYGGLYGGLGGLGYG. These motifs were selected from several possible consensus sequences derived by inspection of the native SRT proteins of several squid species: *Loligo vulgaris, Loligo pealei, Todarodes pacificus, Euprymna scolopes, Dosidicus gigas, Sepioteuthis lessoniana,* and *Sepia esculenta.*

![Figure 5-2](image)

Figure 5-2. SRT-inspired tandem repeat protein mimics. a) Amino acid building block sequence (repeat unit “n”) based on consensus across species. b) Mass spectrometry and SDS-PAGE of SRT-inspired tandem repeat protein mimics comprising 4, 7, 11, and 25 repeat units (15, 25, 42, and 86 kDa in size respectively).

Inspired by the native sequence of SRT proteins, four synthetic polypeptides were designed that share the same primary amino acid sequence: $[\text{STGTLSYGYGGLYGGLGGLGYGPAASVSTVHHP}]_n$ (Figure 5-2a). PD-RCA cloning was used by Dr. Benjamin Allen and Huihun Jung to generate a diversity of DNA comprising $n$ copies of the SRT tandem repeat building block. $^{85}$ Selected RCA products with $n = 4, 7, 11,$ and 25 were expressed in *E. coli* to produce a panel of artificial SRT-based proteins that vary only in the repeat
number, but not in the lengths or compositions of their crystalline and amorphous regions. The resulting SRT-inspired polypeptides are:

- TR-n4 (15 kDa): [STGTLSYGYGGLYGGLGYGPLAAASVSTVHHP]_{4}
- TR-n7 (25 kDa): [STGTLSYGYGGLYGGLGYGPLAAASVSTVHHP]_{7}
- TR-n11 (42 kDa): [STGTLSYGYGGLYGGLGYGPLAAASVSTVHHP]_{11}
- TR-n25 (86 kDa): [STGTLSYGYGGLYGGLGYGPLAAASVSTVHHP]_{25}

One of the advantages of the PD-RCA cloning method is that the TR polypeptides have a single molecular weight (molecular weight dispersity ~1.0) since the sequence repetition steps occur in the DNA (and the candidates are selected for consequent expression). This provides a major advantage in the analysis of the self-assembly and properties of SRT-TR proteins due to the full control of the topology of each synthetic SRT-TR polypeptide (including the amino acid composition, length of each segment, and total length of the polypeptide chain by varying the repeat unit number). The molecular weight of the four SRT-TR polypeptides was measured by MALDI-TOF and SDS-PAGE (Figure 5-2b), revealing a single molecular weight of 15, 25, 42 and 86 kDa for each TR polypeptide, which was within the expected length.

5.2.2 Nanostructure of SRT-inspired tandem repeat protein networks

The nanostructure and morphology of SRT-TR synthetic polypeptides was analyzed by FTIR and XRD (Figure 5-3) following analogous methods to previous characterization of native *Loligo vulgaris* SRT protein complex (Chapter 3) and synthetic Lv18 recombinant protein (Chapter 4). FTIR analysis of the amide I band (1600-1700 cm\(^{-1}\)), which rises from carbonyl stretching vibrations and contains information about the backbone conformation, reveals an extremely similar structure of all four TR proteins (Figure 5-3a). It is expected that all TR proteins have the same secondary structure content because they all share the same building block and amino acid primary sequence (i.e., amino acid composition is exactly the same, and only the total number of building blocks is changing). The amide I band has its maximum at 1625 cm\(^{-1}\), which denotes a high content of β-sheet structures (similar to native and recombinant proteins). \(^{151–153,156–159}\)
However, the amide I band of TR proteins is heavily skewed towards low wavenumbers, which suggests a higher β-sheet content than native and recombinant proteins (54%, 48%, and 49% respectively). This 6% increase in β-sheet structures (at the expense of random coil and turn structures) in TR is attributed to a more ordered sequence, similar to the effect of tacticity in polymer crystallization. Crystalline segments with the exact same length and exact amino acid composition will facilitate the assembly of ordered strands and will position alternating proline and histidine residues at the exact same distance in antiparallel configuration. In addition, irregular long amorphous segments are eliminated from the TR sequence (all disordered segments have the same length), decreasing the random coil and turn structures fraction. Hence, ordered repetitive sequences promote the crystallization of β-sheet in SRT-TR synthetic polypeptides.

Figure 5-3. Structural characterization of SRT-inspired proteins. a) FTIR and b) XRD data show the same secondary structure and crystalline nanodomains (dominated by β-sheet structures).

Analysis of the β-sheet crystals by XRD (Figure 5-3b) reveals a very similar spectra to native and recombinant SRT proteins. The major crystalline peaks can be observed at 2θ = 9.1°, 19.8° and 24.4° corresponding to lattice distances d_{100} = 9.68 Å, d_{200} = 4.48 Å and d_{002} = 3.54 Å (according to Bragg’s Law law nλ = 2d sin θ). These lattice distances correspond to the hydrogen-bond separation between two β-sheet strands (4.72 Å), the separation between alternating β-sheet chains (9.44 Å, i.e. two β-sheet strands) and the chain length of a single amino acid in an antiparallel β-
sheet structure (3.5 Å, with a two-residue repeat distance of 7.0 Å). The crystal structure is consistent with previous analysis of SRT proteins and is consistent with the Warticker system group 3b (orthorhombic unit cell). The size of the β-sheet crystals of SRT-TR proteins was estimated by the analysis of the peak broadening according to Scherrer’s equation, $\Gamma(2\Theta) = \frac{k \lambda}{L \cos \Theta}$, where $\Gamma(2\Theta)$ is the full width at half maximum of the crystalline peak, $k$ is a constant of proportionality (usually 0.9 to 1 for polymers and cubic symmetries), $\lambda$ is the X-ray wavelength, and $L$ is the crystallite size. SRT-TR β-sheet crystals have an average crystallite size of ~ 1.8 x 2.4 nm (hydrogen bonding direction x backbone direction), which is slightly larger than recombinant β-sheet crystals (agreeing with FTIR measurements). These dimensions correspond to 4 β-sheet strands in the hydrogen bonding direction and approximately 7 amino acids in the backbone direction, falling within expectations from the repeating sequence design (10 amino acids is the maximum possible backbone length).

Interestingly, the XRD spectra of all SRT-TR polypeptides is extremely similar to each other (overlapping spectra), indicating that the crystal size is conserved independently of the number of repeat units and total molecular weight. A constant crystal size in all TR polypeptides is not necessarily expected despite having the same sequence building block, and this has a strong impact in the network formation and overall physical properties. The formation of β-sheet crystals can be roughly described by the thermodynamics of self-assembly: $\Delta G \approx \Delta H_{HB} + \Delta H_{vdW} - T \Delta S_c$. The free energy change ($\Delta G$) upon incorporation of a crystal-forming strand (AVSTH-rich segment) in a β-sheet crystal can be divided into three terms: (i) hydrogen bonding within the β-sheet ($\Delta H_{HB}$), (ii) van der Waals interactions in the β-sheet ($\Delta H_{vdW}$), and (iii) the free energy of the flexible protein chain ($T \Delta S_c$). The hydrogen bonding (strands attach parallel to the backbone) and van der Waals terms (strands stack on top of the β-sheet) are enthalpic, and drive the self-assembly process. The third term represents the entropic loss due to the folding of the main chain, and opposes the self-assembly process (flexible chains will minimize the energy in a disordered random coil conformation). According to this simplified model, the β-strands will aggregate forming β-sheets until the opposing entropic term overtakes the self-assembly driving forces. At that point of thermodynamic minimum free energy of self-assembly, the β-sheet crystal will stop growing. Although this is a simplified model, it serves as base for understanding the limited growth of β-sheet crystals in SRT proteins and other fibrous proteins such as amyloid fibrils and silk fibroin. The effect of segment length, sequence topology, charged and voluminous amino
acid side chains, and nucleation should be additionally considered in a comprehensive model for β-sheet assembly. Developing and validating such models that can predict protein assembly is a challenging task due to the complexity of natural and synthetic proteins, and it has been the focus of extensive research over the past years.\(^{88,348,354,366-369}\) However, it is clear from the structural characterization of SRT-TR polypeptides that β-sheet crystals are confined to a single-layer sheets (i.e., no multiple stacking) of four strands (hydrogen bonding direction).

![Figure 5-4](image)

Figure 5-4. Formation of tie chain connecting strands due to β-sheet nanoconfinement. While short polypeptides can be accommodated into one single crystallite (TR-n4), longer polypeptides must accommodate into several crystallites (leaving a connecting strand in every crystallite jump).

The presence of proline and histidine amino acids (β-sheet disruptor and charged side chain respectively) and the end of alanine-rich (crystal-forming) segments sets the crystal size limits along the backbone as previously discussed, but the origins of the hydrogen bonding size confinement (number of strands) is not completely clear to date. If one hypothesizes about non-restricted β-sheet crystal growth, SRT-TR polypeptides could possibly assemble into supramolecular β-sheet fibrils similar to amyloid structures.\(^{88}\) SRT-TR chains would fold into inter-chain β-sheet structures and form large crystallites (one big crystal per chain). This would result in weak structures because of: (i) larger β-sheet crystals have inferior mechanical properties (elastic modulus, toughness, and ultimate strength of β-sheet crystals decrease with increasing crystal size),\(^{101,102,231,232}\) and (ii) the connectivity between crystallites would be poor since most of
the amorphous segments would adopt loop conformations (inter-chain hydrogen bonding). On the other hand, nanoconfined crystals such as the ones measured in SRT-TR polypeptides provide superior mechanical properties: (i) nanoconfined β-sheet (especially 3 nm and below) exhibit enhanced mechanical properties due to a more efficient use of hydrogen bonds, more uniform stress distribution, and dissipative molecular stick-slip deformation mechanisms;\textsuperscript{101,102,231,232} (ii) in two materials with the same β-sheet content, smaller crystals will result in an increased number density of β-sheet aggregates compared to a larger crystal network (i.e., smaller in size but larger in number, equivalent to increasing the cross-linking density); and (iii) crystal nanoconfinement forces the polypeptide to form tie chain connections between crystals. Figure 5-4 represents the tightest packing possible conformation (favoring inter-chain interactions) of all SRT-TR proteins given a crystal size of 4 strands. TR-n4 has 4 crystal-forming segments that can occupy a single β-sheet crystallite. On the other hand, TR-n7 has 7 crystal-forming segments. Once the longer chains have filled the capacity for one crystal (β-sheet nanoconfinement), they must continue in the next crystal, leaving an amorphous tie chain connecting the two neighboring β-sheets. Therefore, larger molecular weight (higher number of repeat units \(n\)) will result in increasing tie chain density due to the β-sheet crystal nanoconfinement, yielding a more efficient network structure. The quantification of the protein network morphology and its impact on the mechanical and transport properties of SRT-TR are investigated below.

5.2.3 SRT-inspired protein networks: glassy mechanics

The mechanical properties of cast SRT-TR synthetic polypeptides were analyzed by DMA (Figure 5-5a). TR-n4 is brittle, and shows linear elastic behavior at low strains and then fracture. In contrast TR-n7, TR-n11, and TR-n25 can withstand larger strains compared to TR-n4, and they exhibit irreversible plastic deformation at strains larger than 3%. Young modulus for the polypeptides can be estimated from the linear region of the stress-strain curve, which is \(\sim 0.7-0.8\) GPa (Table 5-1). This is slightly lower than the measured elastic modulus of native SRT (\(\sim 2\) GPa), which could be due to differences in crystallinity (38% - 48%), ambient water in the sample, and trace amount of HFIP from the casting process. Although the elastic modulus, and ultimate strength for all four samples are similar (slightly increasing with molecular weight), their toughness and maximum strain increase with the polypeptide molecular weight. This is due to the yielding process 3%
strains and above for high molecular weight TR proteins. While short molecular weight (TR-n4) proteins exhibit brittle behavior, higher molecular weight proteins (TR-n7, TR-n11, and TR-n25) yield and exhibit crazing before fracture. Crazing lines are formed during the yielding process, and eventually lead to the formation of cracks and ultimately lead to mechanical failure (Figure 5-5b).

![Stress-strain graph](image)

**Figure 5-5.** Tensile testing of SRT-inspired proteins. a) Stress-strain measurements show a similar elastic region, but the ultimate strain is higher with increasing number of repeat units n. b) Fractured samples show brittle fracture for low n, whereas crazing is observed for higher n.

**Table 5-1.** Modulus, ultimate strain, ultimate stress, and toughness of SRT-inspired proteins from tensile testing.

<table>
<thead>
<tr>
<th></th>
<th>E (MPa)</th>
<th>ε&lt;sub&gt;yield&lt;/sub&gt; (%)</th>
<th>σ&lt;sub&gt;yield&lt;/sub&gt; (Mpa)</th>
<th>ε&lt;sub&gt;max&lt;/sub&gt; (%)</th>
<th>σ&lt;sub&gt;max&lt;/sub&gt; (MPa)</th>
<th>Toughness (MJ/m&lt;sup&gt;3&lt;/sup&gt;)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TR-n4</td>
<td>590 ± 207</td>
<td>---</td>
<td>---</td>
<td>3.1 ± 1.2</td>
<td>13.3 ± 3.5</td>
<td>0.24 ± 0.16</td>
</tr>
<tr>
<td>TR-n7</td>
<td>710 ± 86</td>
<td>3.1 ± 0.4</td>
<td>15.1 ± 1.5</td>
<td>5.8 ± 1.8</td>
<td>15.9 ± 1.8</td>
<td>0.66 ± 0.28</td>
</tr>
<tr>
<td>TR-n11</td>
<td>748 ± 147</td>
<td>4.6 ± 1.3</td>
<td>18.2 ± 1.5</td>
<td>15.0 ± 4.4</td>
<td>18.9 ± 1.6</td>
<td>2.22 ± 0.72</td>
</tr>
<tr>
<td>TR-n25</td>
<td>808 ± 93</td>
<td>3.2 ± 0.4</td>
<td>20 ± 1.2</td>
<td>19 ± 6.3</td>
<td>20.8 ± 1.7</td>
<td>2.84 ± 0.63</td>
</tr>
</tbody>
</table>

The yielding process is further investigated by digital image correlation (DIC) and the obtained strain field maps of TR-n4, TR-n7, and TR-n11 are shown in Figure 5-6. TR-n25 was not selected for DIC characterization due to the extremely low quantity of material available for
characterization (yield of high molecular weight proteins is low) and the destructive nature of the DIC experiment. Strain maps of TR-n4 show a homogeneous distribution of strain, as is expected in isotropic elastic materials, and a sudden brittle failure of the protein film (instant failure immediately after the first crack is generated). On the other hand, TR-n7 and TR-n11 strain maps show highly localized strain in the crazing regions. Once a craze is formed, only the localized strain increases upon further stretching while the bulk strain of the sample remains constant (around 3%, blue background in strain field maps). At local strains higher than 25%, the crazes lead to crack formation. With increasing strain beyond this point, the crack progressively propagates and leads to mechanical failure (DIC cannot longer correlate the strain).

Figure 5-6. Digital image correlation (DIC) of SRT-inspired proteins. a) Detail of crazing in TR-n11. b) DIC shows full-field strain measurements for TR-n4, n7, and n11 at the locations marked with letters in c). TR-n4 shows a homogeneous strain along the gauge length, while TR-n7 and n11 show localized strain concentration in the crazing areas.
The β-sheet semicrystalline morphology can explain the tensile deformation behavior of SRT-TR polypeptides, which is in agreement with established polymer fracture models. As described previously, SRT (both native and synthetic) proteins are composed of hard crystalline β-sheet domains and disordered amorphous chains due to the segmented amino acid sequence. In the glassy state, the amorphous regions form a loose network of chains that are tied together through random weak attractive forces (e.g., hydrogen bonds and van der Waals interactions), while β-sheets are stiff (efficient and ordered hydrogen bonding). This weak hydrogen bonding network, which is present in all samples independently of the disordered chain conformation, is responsible for the elastic behavior at low strains. Upon further deformation ($\varepsilon \sim 3\%$, $\sigma \sim 20$ MPa), the weak hydrogen bonding network that keeps the disordered chains together breaks, and the disordered chains are not constrained anymore. At this point, the linear elasticity is terminated and two behaviors are observed: (i) short polypeptides (i.e., TR-n4) exhibit brittle fracture, and (ii) long polypeptides (i.e., TR-n7, TR-n11, and TR-n25) exhibit yielding. If one considers the effect of β-sheet crystal nanoconfinement in the disordered chain conformation (as previously discussed), the disordered fraction in short polypeptides will be rich in loop structures due to abundant inter-chain folding. Once the weak hydrogen bonding network is broken upon deformation, the loop structures cannot withstand stress due to the lack of connectivity between crystals (defective network). Hence, cracks form easily and brittle failure is observed. On the other hand, longer polypeptides are abundant in tie chains that connect the β-sheets due to crystal nanoconfinement, and form a more efficient network that can withstand stress after the initial elastic region. After the breaking of the hydrogen bonds between the disordered chains, a yield process dominated by disordered chain slippage and reorientation of the crystals begins. As the molecular weight of SRT-TR proteins increases, so does the density of connecting tie chains and the efficiency of the network, and therefore it is expected that the toughness and ultimate strain is higher in long polypeptides. Several experimental and computational studies (based on Eyring’s activation rate theory) have accurately described the mechanics of synthetic and natural semicrystalline polymers including silk fibroin, and show analogous mechanics and yielding dependence on molecular weight.
5.2.4 SRT-inspired protein networks: rubbery mechanics

As discussed previously, the kinetics of SRT-TR polypeptides self-assembly and the nanoconfinement of β-sheet crystals lead to protein networks with tie chains and topological defects, such as loops and dangling ends. The presence of such defects has a strong impact on the morphology of the material and complicates the prediction of physical properties.\textsuperscript{328} The estimation of network defects is a challenging task since loops and dangling ends are chemically and spectroscopically identical to tie chains, and advanced isotope labeling spectroscopy is required for detailed quantification\textsuperscript{242,328}. Nonetheless, loop structures have a great impact in the mechanical properties of the material (loops generate topological defects in an elastic network), and therefore the loop and tie-chain content in the SRT-TR polypeptide network can be estimated by mechanical characterization. A quantitative evaluation of the protein network morphology from the glassy mechanics is challenging due to the presence of weak attractive forces and physical constraints in the disordered chains. To understand the correlations between tandem repetition, topological defects and physical properties, the rubbery mechanics of SRT-TR polypeptides are analyzed and modeled as hydrogel networks.

Figure 5-7. a) TGA of hydrated SRT-inspired proteins show the same water absorption for all 4 proteins. b) Dynamic rheology of fully hydrated (water content ~20%) SRT-inspired proteins at 70 °C shows an increase in moduli with higher number of repeat units n.
First, SRT-TR proteins were plasticized with water in order to decrease the glass transition temperature and increase the mobility of the disordered chains (analogous to native and recombinant SRT studies). Water absorption was measured by TGA (Figure 5-7a), revealing a water uptake of approximately 40% in all four polypeptides.

SRT-TR polypeptides were analyzed by oscillatory rheology in order to estimate the molecular network morphology. SRT-TR proteins are dominated by the storage modulus ($G' > G''$) and exhibit a frequency-independent modulus in the low-frequency limit (Figure 5-7b). Such behavior can be explained by a hydrogen-bonding stabilized network where the crystalline β-sheet domains act as physical cross-linker and the disordered amorphous chains are the stretched network strands or tie chains. A weak frequency dependence is observed at $\omega > 0.1$ rad/s, which originates from the gradual relaxations of network defects (such as dangling ends and loop structures, common in imperfect polymer networks). An increase in moduli with molecular weight can be observed, which suggests a higher cross-linking density or a higher network efficiency for high molecular weight SRT-TR. Different from other physically cross-linked protein gels (based on coiled-coil interactions, phase transition aggregation, etc.), SRT-TR polypeptides exhibit a stable MPa modulus over a wide range of frequencies due to the high stability of β-sheet hydrogen bonding interactions.

Figure 5-8. Analysis of entanglements. a) FTIR of SRT-inspired protein in 8M urea shows the disruption of β-sheet structures (physical cross-links) as the amide III band intensity decreases. b) Dynamic rheology of SRT-inspired proteins in 8M urea shows the same storage modulus for all samples.
In order to corroborate the physical cross-linked nature of the self-assembled SRT-TR polypeptides, the viscoelastic response in a concentrated urea solution was measured. Urea is a well-known hydrogen bonding disruptor extensively used in protein unfolding protocols. FTIR analysis of the amide III band (1200 – 1300 cm\(^{-1}\)) shows major reduction of SRT-TR \(\beta\)-sheet structures after immersion in 8M urea solution (Figure 5-8a). The modulus of the SRT-TR gels in urea decreases due to the unfolding of \(\beta\)-sheet domains and the consequent removal of cross-linking points (Figure 5-8b). However, the viscoelastic response still exhibits an overall frequency independent modulus. Without hydrogen-bonded structures present to stabilize the network, these results suggest that chain entanglements are responsible for the modulus of the gels in urea. Therefore, it was concluded that both entanglement and physically cross-linked network strands contribute to the modulus of the assembled SRT-TR polypeptides.

In order to better estimate the chain morphology and to separate the contribution from tie-chains and entanglements, the total modulus is divided into \(G'_{t} \approx G'_{x} + G'_{e}\), where \(G'_{x}\) and \(G'_{e}\) are the modulus of the cross-linked network and entangled network respectively (taken as the equilibrium modulus at the low frequency limit) (Figure 5-9a). \(G'_{e}\) is constant for all SRT-TR mimics and shows no dependence on molecular weight, which is expected for entanglement dominated systems of high molecular weight. On the other hand, \(G'_{x}\) has a linear dependence with \(1/M_w\), which is common for cross-linked networks, indicating that the network strand density increases and the number of defects decreases with molecular weight.\(^{242,323,378}\) Hence, the tie-chain density of the SRT-TR polypeptides as function of repeat unit \(n\) can be estimated by entropic elasticity theory models.\(^{242,323,324}\) Considering that the contribution to the modulus is \(kT\) per strand, the modulus of any network can be estimated as follows according to the affine network model:

\[
G = \nu kT = \frac{\rho RT}{M_s} \tag{12}
\]

Where \(\nu\) is the number density of strands (number of network strands per unit volume), \(k\) is Boltzmann constant, \(\rho\) is density, \(R\) is the gas constant and \(M_s\) is the average molar mass of a network strand. The phantom network model includes a correction factor of \((1 - 2/f)\), where \(f\) is the functionality of the network junctions and accounts for fluctuations of the cross-linking junctions. However, these models considers a perfect end-to-end cross-linked network and ignores the effect of network defects and entanglements. Because of network defects, \(\nu\) is seldom precisely
known and causes discrepancies between predictive and experimental quantification of the modulus. A common modification to the affine network model was introduced by Flory, which corrected for terminal dangling ends: \((1 - 2 \frac{M_d}{M})\), but did not include defective loop strands. More recent models can provide an accurate prediction of the modulus given that the knowledge of the loop content is known beforehand, which is only possible in limited polymer chemistries and isotope labeling. However, none of them accurately describe the SRT protein networks due to the highly defective structures at low molecular weight due to the high content of disordered loop structures (intramolecular interactions).

![Figure 5-9. SRT modulus masterplot.](image)

a) The contribution of entanglements and tie chains to the total shear modulus is separated as \(G' = G'_e + G'_c\). While the entanglement modulus is independent of the number of repeat units \(n\), the tie chain network modulus increases with the repeat units \(n\). b) The elastic effectiveness (fraction of elastically effective strands in the network) is calculated as function of the repeat units \(n\) and the crystallite size \(\beta_c\).

In order to accurately quantify the network defects in SRT polypeptides, a modification of Flory’s dangling ends correction is proposed that accounts for the defects generated by the \(\beta\)-sheet nanoconfinement: while intramolecular \(\beta\)-sheets are not limited in hypothetical infinite size crystals, real \(\beta\)-sheets can only accommodate a limited number of crystalline strands. Once the first crystal is at maximum capacity, the chain will move on to the next neighbor crystal and so on, leaving a connecting tie chain in the process. Hence, an elastic effectiveness parameter can be estimated as \(\varepsilon_{\text{eff}} = (1 - \frac{\beta_c}{n})\), where \(\beta_c\) is the \(\beta\)-sheet crystallite size (\(\approx 4 \beta\) strands for SRT proteins as determined by XRD), and \(n\) is the number of repeats in the polypeptide chain. With this
correction for network defects and additional corrections for the swelling of the network (in water and urea), the modulus of SRT proteins can then be calculated as follows:

\[
G'_{\text{total}} \cong G'_e + G'_x = \rho RT \left[ \frac{1}{M_e} \left( \frac{\phi_e}{\phi_{e0}} \right)^{\frac{1}{3}} + \varepsilon_{\text{eff}} \left( \frac{\phi_x}{\phi_{x0}} \right)^{\frac{1}{3}} \right] = \rho RT \left[ \frac{1}{M_e} \left( \frac{\phi_e}{\phi_{e0}} \right)^{\frac{1}{3}} + \frac{1 - \beta_c}{n} \left( \frac{\phi_x}{\phi_{x0}} \right)^{\frac{1}{3}} \right]
\]

Where \( M_e \) is the average molar mass of an entanglement strand, \( M_x \) is the average molar mass of a network strand (extrapolated to \( n = \infty \)), \( \phi_0 \) and \( \phi \) are the initial and equilibrium swelling protein volume fraction respectively (\( \phi_e \) and \( \phi_x \) denote swelling of entanglement and network strands, measured in 8M urea and in water respectively).

The elastic effectiveness parameter \( \varepsilon_{\text{eff}} \) is plotted against reciprocal number of repeat unit \( n \) for different crystallite strand capacity \( \beta_c \) (Figure 5-9b). Due to \( \beta \)-sheet confinement, the number of tie-chains increases with \( n \) (high \( \varepsilon_{\text{eff}} \) will be achieved with large polypeptide lengths and small \( \beta \)-sheet crystallite dimensions). On the other hand, short polypeptide lengths (\( \beta_c \approx n \)) self-assemble into loop-rich structures with \( \varepsilon_{\text{eff}} \approx 0 \) (high density of network defects). The experimental elastic effectiveness \( \varepsilon_{\text{eff}} \), calculated from shear data is in good agreement with the prediction for a \( \beta \)-sheet crystallite capacity of 4 strands (Table 5-2), and it is anticipated that the proposed model can predict the fraction of elastically effective network strands for any SRT-inspired repetitive proteins with given length and crystallite size.

Table 5-2. SRT-inspired network modulus and elastic effectiveness

<table>
<thead>
<tr>
<th></th>
<th>TR-n4</th>
<th>TR-n7</th>
<th>TR-n11</th>
<th>TR-n25</th>
<th>TR-n( \infty )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n )</td>
<td>4</td>
<td>7</td>
<td>11</td>
<td>25</td>
<td>( \infty )</td>
</tr>
<tr>
<td>Mw (kDa)</td>
<td>15</td>
<td>25</td>
<td>42</td>
<td>86</td>
<td>( \infty )</td>
</tr>
<tr>
<td>( G_t ) (MPa)</td>
<td>1.61 ± 0.23</td>
<td>2.58 ± 0.23</td>
<td>3.41 ± 0.22</td>
<td>3.69 ± 0.18</td>
<td>4.17 ± 0.15</td>
</tr>
<tr>
<td>( G_e ) (MPa)</td>
<td>1.32 ± 0.14</td>
<td>1.45 ± 0.19</td>
<td>1.54 ± 0.23</td>
<td>1.39 ± 0.31</td>
<td>1.42 ± 0.09</td>
</tr>
<tr>
<td>( G_x ) (MPa)</td>
<td>0.29 ± 0.27</td>
<td>1.13 ± 0.30</td>
<td>1.87 ± 0.32</td>
<td>2.30 ± 0.36</td>
<td>2.75 ± 0.15</td>
</tr>
<tr>
<td>( \varepsilon ) (exp.)</td>
<td>0.11 ± 0.08</td>
<td>0.41 ± 0.08</td>
<td>0.69 ± 0.08</td>
<td>0.85 ± 0.07</td>
<td>-</td>
</tr>
<tr>
<td>( \varepsilon ) (model)</td>
<td>0</td>
<td>0.43</td>
<td>0.67</td>
<td>0.84</td>
<td>1</td>
</tr>
</tbody>
</table>

The cross-inking and entanglement moduli of a perfect SRT protein network can be calculated by extrapolation to an infinite number of repeat units \( n \). This hypothetical network will be dominated
by elastically effective strands (tie chains) and will have negligible defects. Therefore, it can be used to estimate the size of the average network strand of SRT-TR polypeptides (regardless of the repeat units). The entanglement modulus (calculated from the urea data set) is \( n \)-independent, and gives an average modulus of \( G_{\infty} = 1.42 \pm 0.09 \) MPa which corresponds to an average entanglement strand of \( M_e = 2268 \pm 146 \) g mol\(^{-1}\). On the other hand, the cross-linking modulus of a perfect SRT network is extrapolated to \( G_{\infty} = 2.75 \pm 0.15 \) MPa which corresponds to an average cross-linked strand of \( M_x = 1277 \pm 174 \) g mol\(^{-1}\). If one analyzes the repeating unit sequence, an average molecular weight per amino acid of 96.8 g mol\(^{-1}\) can be estimated. The average entanglement and cross-linking strands have an estimated length of \( M_e = 23.4 \pm 1.5 \) and \( M_x = 13.2 \pm 1.8 \) amino acids, which fall within the length of the repetitive disordered segment sequence (17 amino acid long). Hence, the proposed model accurately describes the mechanical properties of SRT-TR polypeptides, and it provides design guidelines for optimization of the modulus by controlling the crystal size and number of repetitive building units in protein sequence.

**5.2.4.1 SRT-inspired protein networks under large deformation**

![Figure 5-10. Tensile tests of rubbery SRT-inspired tandem repeat proteins. a) Stress-strain curves show strain-hardening at high elongation. Solid lines are experimental data, dashed lines are fits to freely jointed chain model (Eq. 14). The tensile modulus \( E \) shows a linear dependence on reciprocal molecular weight (or number of repeat units). b) While high number of repeats are elastomeric (tie chain-rich), low number of repeats are brittle (defect-rich).](image)
The mechanical response under large deformation of hydrated SRT-TR polypeptides was investigated, revealing an elastomeric behavior. Figure 5-10a shows a linear stress-strain dependence at low deformation, and increasing stress at higher strains. This strain hardening behavior is explained by the finite extensibility of the protein chains. The Gaussian approximation for a freely jointed chain model accurately describes small deformations (end-to-end distance is much shorter than in fully stretched chains), but underestimates the force required to stretch a chain at high deformations (end-to-end distance close to maximum stretched length)\textsuperscript{242,323,324}. The force required to stretch a freely jointed chains have a Langevin dependence of the normalized end-to-end distance that accounts for the finite length of the chains:\textsuperscript{290,329,379}

\[ f = \frac{kT}{b} L^{-1} \left( \frac{R}{Nb} \right) \]  

(14)

Where \( b \) is the length of the smallest flexible chain subunit (Kuhn length), \( L^{-1} \) is the inverse Langevin function,\textsuperscript{380–382} \( N \) is the total number of flexible subunits in a chain, and \( R \) is the chain end-to-end distance. Due to the finite extensibility of the chains, this dependence of the force \( f \) on the chain elongation \( R \) deviates from the Gaussian approximation and diverges at the maximum end-to-end distance \( R_{max} = bN \). Based on finite chain extensibility, multiple models have been developed and extensively used to describe large scale deformations in polymer networks.\textsuperscript{212,324,329}

This relationship can be rewritten in terms of stress and strain by applying a number of simple conversions and approximations:

- \( E = \frac{\sigma}{\varepsilon} \) (where \( E \) is elastic modulus, \( \sigma \) is stress, and \( \varepsilon \) is strain)
- \( \sigma = \frac{f}{A} \) (\( f \) is stretching force, and \( A \) is chain cross-sectional area)
- \( \nu \approx A R_0 \) (\( \nu \) is the volumetric strand density, \( R_0 \) is the unstretched end-to-end distance)
- \( G = \nu kT \) (\( G \) is shear modulus)
- \( E \approx 3 G \)
- \( R_0 = b \sqrt{N} \)
- \( \lambda = 1 + \varepsilon \) (\( \lambda \) is the draw ratio)

When combined, the stress-strain relationship is written as:

\[ \sigma = \frac{1}{3} E \sqrt{N} L^{-1} \left( \frac{\lambda}{\sqrt{N}} \right) - \sigma_0 \]  

(15)

Where \( \sigma_0 \) is the stress at zero strain. This freely jointed chain model, which has been used in synthetic polymers and biomolecules,\textsuperscript{290,372,379,383} describes accurately the elastomeric behavior of
the SRT polypeptides. The fitting of the stress curves gives $N = 16$-$18$ amino acids per amorphous strand, which agrees with the GLY-rich strands $17$ amino acid sequence. At failure ($350\%$ strain), $R/Nb \approx 1$, indicating that the flexible chains are close to fully stretched. The elastic tensile modulus $E$ (shown in Figure 5-10a inset) increases with molecular weight, suggesting an increase in cross-linking density (or elastically effective strands density) with increasing number of TR repeat units. These repeat unit (and the consequence structural defects) dependence agree with measurements of the rubbery shear modulus by rotational rheology. However, low and high molecular weight polypeptides show different failure modes (Figure 5-10b), similar to the results reported by DIC in glassy SRT-TR films. While TR-n4 is brittle (i.e., breaks at very low stress and deformation), higher molecular weight polypeptides can withstand larger deformations and stress due to a less defective structure.

Figure 5-11. Cyclic tensile testing of SRT-TR proteins. a) Stress-strain curves for several loading cycles with increasing strain. B) Toughness or energy absorbed per cycle shows three regions with increasing strain: (i) homogeneous deformation of random coils), (ii) unraveling of random coils, and (iii) backbone stretching. Ultimate toughness shows a linear dependence with reciprocal molecular weight.

Strain-induced hardening, caused by changes in the secondary structure upon deformation (in the case of SRT, $\beta$-sheet structures), is a commonly observed phenomenon in elastomeric proteins such as silk fibroin$^{231,384}$, keratin$^{385}$, marine snail egg cases$^{132,383,386}$ and hagfish threads$^{387}$ among others. The strain-induced transition of disordered to $\beta$-sheet structures comprehends three major
steps as the material is stretched: (i) homogeneous stretching of the disordered protein chains, (ii) unraveling of the amorphous domains and chain alignment, (iii) stretching of the protein backbone\textsuperscript{231,290,379,384}. The energy dissipation of SRT polypeptides during stretching was measured in cyclic loading tests experiments (Figure 5-11a). The absorbed energy or toughness for every cycle is shown in Figure 5-11b, where the aforementioned three regions (with different toughness slope) can be differentiated: (i) little energy absorbed (below 0.5 MJ/m\textsuperscript{3}) at cycle strains up to 50-60\% (reversible deformation of disordered chains), (ii) moderate increase of the energy absorbed at cycle strains between 50\% and 150\% (unraveling of random coils), (iii) heavy increase of the energy absorbed at cycle strains higher than 150\% (backbone stretching). Note that the toughness vs. cycle strain slopes of the SRT-TR mimics are very similar in region (i) (homogenous stretching), but the slopes are steeper for higher molecular weight proteins in regions (ii) and (iii). Longer SRT-TR polypeptides exhibit higher toughness, and the ultimate toughness (i.e. energy absorbed until failure, Figure 5-11b inset) has a clear linear dependence with reciprocal molecular weight. This \textit{n} dependence of the toughness can be explained by the increased tie chain density of higher molecular weight polypeptides. Longer peptides have more elastically effective networks and therefore a higher number of tie chains will be stretched, causing an increase in absorbed energy.

However, structural changes might be occurring during the stretching process (such as strain-induced crystallization), as suggested by the hysteresis in the cyclic stress-strain curves (Figure 5-11a). Fourier transform IR (FTIR) spectroscopy data of TR-n11 films (Figure 5-12a) provides insight in the structural changes within the protein network during stretching. SRT-TR42 was chosen for this study among the other synthetic polypeptides due to its molecular weight (42kDa), which is most representative of the native SRT protein complex\textsuperscript{211}. The amide I band (1600-1700 cm\textsuperscript{-1}, corresponding to the carbonyl stretching vibration, and containing information about the backbone conformation) shows a progressive shift towards lower wavenumbers with increasing strain (shift towards 1625 cm\textsuperscript{-1}, which is in the $\beta$-sheet region)\textsuperscript{151–153,156–159}. Analysis of the secondary structure elements of the stretched films by Fourier self-deconvolution (FSD) reveals an increase in $\beta$-sheet content from 54\% to 60\% up to 200\% strain. The $\beta$-sheet content does not increase at low deformations (0 – 67 \% strain) because of the reversible deformation of the random coil chains, as measured by cyclic loading. However, $\beta$-sheet content increases sharply upon larger
deformations corresponding to the unraveling of coils, chain alignment and backbone stretching (which would bring neighboring chains closer and favor strain-induced crystallization).

Figure 5-12. Strain-induced β-sheet alignment. a) Amide I band (FTIR) shifts towards the β-sheet region with increasing strain. (inset) β-sheet content increase at strains higher than 50%. b) WAXS. c) SAXS reveals the anisotropy and alignment of the β-sheet nanostructure with increasing strain.

Wide angle X-ray scattering (WAXS, Figure 5-12b) shows very similar diffraction patterns for unstretched and stretched (200% strain) films. The two overlapping diffraction profiles (include diffraction peaks characteristic of β-sheet structures at 2θ ≈ 9.5°, 19.5°, 25°) suggest that the size of β-sheet crystallites is conserved upon stretching (probably due to the presence of proline in the amino acid sequence), and that the small increase in β-sheet content might arise from localized β-sheet formation in the stretched amorphous chains. However, isotropic vs. anisotropic scattering is observed at small diffraction angles. Small angle X-ray scattering (SAXS, Figure 5-12c) provides complementary information on the structure evolution upon deformation, and shows reorientation and anisotropy with stretching of TR-n11 films. While non-deformed films have an isotropic 2D diffractogram, all stretched films have preferential directionality and exhibit
increasing anisotropy with increasing strain. Films stretched to 50%, 100%, and 150% strain show increasingly stretched ellipsoidal diffraction patterns. When stretched to 200%, high anisotropy can be observed due to the splitting in the diffractogram, suggesting that the β-sheet semicrystalline structure of the protein is reorienting and aligning into a nematic phase. Hence, uniaxial deformation causes the alignment of the hierarchical nanostructure over multiple length scales (protein chains, β-sheets, nematic phase). Whether the β-sheet are disrupted and reformed during stretching remains unclear at this point, although computational and experimental work on similar protein systems (silk fibroin) suggest that slippage of β-sheet strands might be possible.231,389 These results suggest that SRT protein networks are isotropic and do not suffer significant structural changes at small deformations (homogeneous stretching of the disordered phase). However, further deformation induces chain alignment, β-sheet crystals reorientation, and possible formation β-sheet long-order fibrils within the protein matrix.166,167,388

5.3 Conclusions

SRT-inspired polypeptides are self-assembled bioelastomers with programmable mechanical properties. Designer SRT-TR proteins with repetitive segmented disordered/crystalline building blocks self-assemble into a semicrystalline protein network rich in β-sheet and disordered structures. Due to the nanoconfined crystallization of β-sheets, the polypeptides form either effective or defective networks depending on the number of repeat units in their sequence. The SRT networks are composed of β-sheet physical cross-linkers connected to each other via disordered tie chains. However, disordered loop structures are originated from intramolecular interactions and introduce defects in the network structure, resulting in inferior mechanical properties. In this chapter, the structure and mechanical properties of SRT-TR polypeptides was analyzed. A model based on entropic elasticity theory is used to predict the network defects as function of the number of repeat units in the polypeptide given a set of nanoconfined β-sheet crystal sizes (Figure 5-13a). The proposed model accurately describes SRT-TR behavior for both small and large deformations, and experimental measurements of shear modulus, tensile modulus, and toughness are in agreement with the predicted properties.
Figure 5-13. Mechanics of SRT-inspired protein networks. a) Elastic effectiveness as function of number of repeat units and crystallite size predicts the density of tie chains and defects. b) Polydispersity in native SRT proteins leads to a more efficient network structure with superior mechanical properties.

Moreover, the results provide design rules for the optimization of the properties of structural proteins and similar biomolecules by increasing repetitions (yielding high-density tie chain networks). This designer optimization approach provides insights into the evolutionary dynamics of gene duplication in naturally occurring repetitive protein materials. Native SRT are formed by a group of homologous proteins with diverse lengths, as described previously. The expression of high molecular weight proteins is energetically expensive for squid’s suction cup tissue, and there is no apparent justification for varying high molecular weight proteins in SRT from a merely structural perspective. However, polydispersity might be a key factor in the effective evolution of material properties in natural systems. The optimization of mechanical properties through gene duplication is a very likely driving force for the expression of high molecular weight repetitive proteins in natural materials (Figure 5-13b).
Chapter 6. SRT-inspired tandem repeat proteins: transport properties

In this chapter, the impact of network defects on the transport properties of SRT-inspired tandem repeat proteins is investigated. Expanding on the network structure analysis from previous chapters, SRT-inspired proteins with varying density of network defects are examined. Increasing the number of repeat units in the protein chain leads to more efficient networks with higher proton and thermal conductivity.

6.1 Introduction

Proton transfer has been extensively studied for over two hundred years as a fundamental process in chemistry, including in acid/base and redox reactions.\textsuperscript{390–395} Research over the last decades has focused on solid-state proton conducting materials (polymers, metal-organic frameworks, ceramic oxides) due to the technological relevance in energy applications and their use in fuel cells, batteries and water electrolysers.\textsuperscript{396–402} Ionomers such as Nafion and its sulfonated derivates represent the state-of-the-art proton conducting materials with some of the highest reported conductivity, and have received much research effort.\textsuperscript{403–407} However, biological proton conducting materials have not been explored to such extent. Proton transport is a ubiquitous phenomenon in nature observed in numerous biomolecules such as bacteriorhodopsin (proton pump) or ATPases (passive conductor).\textsuperscript{394,395,408–411} Some proteins exhibit bulk protonic conduction comparable to analogous conducting materials, and have unique advantages over non-biological materials such as biocompatibility, tunable structure and tunable physico-chemical properties through amino acid sequence control.\textsuperscript{2,78,109,347,412}

In addition, polymer materials have been recently studied as an alternative for conventional thermal management materials (metals and ceramics) despite their poor thermal conductivity (k). Polymers and soft materials offer advantages, including light weight, low cost, easy processability, and flexibility, that surpass conventional thermal materials. These qualities make polymers good candidates for applications where thermal management and flexibility is required, such as flexible
electronics, LED housing, chip encapsulating, and cell phone casing. Driven by these technological incentives, the development of thermally conductive polymers has attracted attention of the scientific community, and multiple approaches to the optimization of thermal properties have been investigated. A common approach is the blending with high-k fillers (metals, ceramics, carbon nanotubes, and graphene most commonly). However, thermally conductive polymer composites require a large filler volume fraction, and they usually entail undesired optical/electrical properties of the composite, and increased cost and processing complexity. As an alternative to polymer composites, recent research has focused on changing the intrinsic thermal conductivity of polymers by adjusting the morphology. While amorphous polymers have low k (in the range of 0.1 W m\(^{-1}\) K\(^{-1}\)), the conductivity of individual polymer chains has been calculated to be as large of 104 W m\(^{-1}\) K\(^{-1}\) (which is higher than most elemental metals). This sets the stage for new approaches in the development of thermally conductive soft materials based on the control of the amorphous morphology and chain alignment.

Because of their programmable nanostructure and control of network defects, SRT present a good platform to investigate the structure-transport properties relationship in protein-based materials. In this chapter, the effects of tandem repetition on the chain dynamics, proton conductivity, and thermal conductivity are analyzed. Tandem repetition, structural morphology, and chain orientation play a key role in the enhancement/inhibition of the transport properties of SRT proteins, and through appropriate control can lead to the highest conductivity values reported to date for biological materials. Because of these unique and programmable set of properties and their controlled nanostructure, SRT-based conducting materials are good candidates for developing bioelectronics devices such as protonic transistors, thermal devices and sensors.
6.2 Results and discussion

6.2.1 SRT-inspired networks: chain dynamics

The SRT-TR protein chain dynamics were measured by neutron scattering spectroscopy. As previously described, QENS can measure molecular dynamic processes such as rotations, relaxations and diffusive motions with 1 – 30 Å and pico- to nanosecond resolution by directly probing the self-diffusion of hydrogen atoms. The lowest (TR-n4) and highest (TR-n11) molecular weight proteins that are in the range of native SRT protein distribution (15 – 65 kDa) were analyzed by elastic and quasielastic scattering in dry, H$_2$O, and D$_2$O conditions.

Fixed window scans (Figure 6-1a) measure the incoherent elastic neutron scattering intensity $I_{el}(Q)$ as function of temperature, which is a measure of the non-mobile hydrogen atoms in the sample. Hydrogen motions faster than the instrumental resolution time window (≈ 1 ns) contribute to the inelastic scattering fraction and are excluded from the elastic scattering intensity. At very low temperatures (T ≈ 4 K) all hydrogen atoms move slower than the instrumental resolution but as temperature increases, hydrogen motions progressively become faster than the instrumental
resolution and fall outside the $I_{el}(Q)$ measurement window. As a consequence, the elastic scattering intensity decreases with temperature. While the elastic scattered intensity of dry TR proteins steadily decreases with temperature, lower intensity is measured for hydrated TR proteins at $T > 280$ K, indicating higher mobility upon hydration.

The measurements were repeated using deuterated water (D$_2$O) to differentiate water and protein dynamics since deuterium is invisible to neutron scattering (negligible scattering cross-section), while it can still plasticize the TR proteins like H$_2$O does. MSD (mean square displacement, Figure 6-1b) of mobile hydrogen atoms were calculated from the elastic scattering intensity as a function of the scattering wave vector $Q$ and temperature $T$ through the Debye-Waller factor (DWF) = $I_{el}(Q, T) / I_{el}(Q, T = 40K) = \exp[-1/3 \, Q^2 \langle x^2 \rangle]$. This Gaussian approximation is the standard method for estimating the MSD and protein mobility, but it approximates MSD of all types of hydrogen motions (vibrations, rotations, diffusion) to a single value over a range of temperatures and $Q$. At very low temperatures (40 K), all samples (dry and hydrated) show a plateau around zero, revealing a very low mobility and no temperature dependence in the $T < 70$ K temperature range. Dry TR polypeptides show an onset at $T \sim 70$ K, which is common for proteins and is typically attributed to the onset of methyl group rotations (constant slope from 70 K to 350 K with $d<x^2>/dT \approx 3.0 \cdot 10^{-3}$ Å$^2$/K). Hydrated (both D$_2$O and H$_2$O) TR proteins show a slope of $d<x^2>/dT \approx 4.4 \cdot 10^{-3}$ Å$^2$/K in the 60 – 230 K region and a slope of $d<x^2>/dT \approx 12.4 \cdot 10^{-3}$ Å$^2$/K in the 260 – 350 K region, with a dynamic transition at 245 K. The higher mobility is attributed to the plasticizing effect of water (i.e. disruption of the hydrogen bonded network). At higher temperatures (above 250 K), TR-n4 (D$_2$O) shows slightly higher mobility than TR-n11 (D$_2$O). In addition, a significant MSD increase in H$_2$O samples is observed at 270-280 K (above the melting point of ice) due to the contribution of hydrogen in bound water. TR-n11 (H$_2$O) exhibits a higher mobility than TR-n4 (H$_2$O) when compared to the D$_2$O baseline. Despite providing information about the protein mobility landscape in a wide range of temperatures, elastic measurements cannot differentiate between the different motions occurring in SRT-TR polypeptides, and further QENS analysis is required.
Figure 6-2. QENS analysis of dry TR-n4 and TR-n11 proteins (measured in HFBS instrument). a) Lorentzian full width half maximum (FWHM) is independent of Q (localized motion). b) Elastic incoherent structure factor (EISF) follows a 3-site jump model for methyl group rotation

QENS spectra was collected at 295 K in high-flux backscattering spectrometer (HFBS, dry proteins) and disk chopper time-of-flight spectrometer (DCS, H₂O and D₂O proteins), which have a dynamic range of ±15 μeV and ±0.5 meV respectively. The instrumental resolution (convoluted with delta function) and a single Lorentzian function (protein motion) were fitted to the experimental data to analyze the quasielastic peak broadening for each sample and each Q value as previously described. The full width half maximum Γ(Q) of dry TR-n4 and TR-n11 proteins is plotted as function of Q² in Figure 6-2a, which reveals a Q-independent behavior at ≈ 10 – 20 μeV characteristic of localized motions. The elastic incoherent structure factor (EISF) of both dry proteins follows a three-site jump model describing methyl group rotations:

\[ EISF_{dry}(Q) = 1 - p_{methyl} + p_{methyl} \cdot EISF_{methyl}(Q), \]

where \( p_{methyl} \) is the fraction of mobile hydrogen involved in methyl group rotations, and is weighting the EISF_{methyl} term.

\[ EISF_{methyl}(Q) = \frac{1 + 2j_0(Q \cdot r \cdot \sqrt{3})}{3}, \]

where \( j_0 \) is the zero order Bessel function and \( r \) is the rotation radius of a methyl group (≈ 1 Å) (Figure 6-2b). This observation of the dry dynamics of dry SRT-TR polypeptides, which are dominated by methyl group rotations, agrees with the measurements of recombinant Lv18 protein dynamics, and those of other protein systems.
Hydration dynamics of TR-n4 and TR-n11 proteins in D$_2$O (protein dynamics only) and H$_2$O (protein + bound water dynamics) are shown in (Figure 6-3). The full width half maximum $\Gamma(Q)$ of the quasielastic peak broadening (Lorentzian function) of D$_2$O- and H$_2$O-hydrated proteins are plotted as function of $Q^2$ in Figure 6-3a and Figure 6-3b respectively. TR protein dynamics (D$_2$O samples, Figure 6-3a) show a $Q$-independent plateau at low $Q$ values, while a linear scaling with $Q^2$ is observed at higher $Q$. This 2-regime $Q$-dependence is characteristic of diffusion in a confined space and can be described by the Volino and Dianoux (VD) model for bounded diffusion in a potential of spherical symmetry.$^{195}$ Fitting of the VD model gives a diffusion coefficient $D = 1.93 \cdot 10^{-4}$ cm$^2$/s (obtained from the slope $\sim DQ^2$ law at higher $Q$) and a confinement volume of radius $r_{conf} = 2.6$ Å.$^{177,195}$ The two proteins show a very similar FWHM over the full $Q^2$ range, indicating that the chain dynamics are the same for both proteins. On the other hand, TR protein in H$_2$O dynamics (Figure 6-3b) show a sublinear scaling with $Q^2$ which is characteristic of jump-diffusion models such as the Singwi-Sjölander (SS) model for combined translational and oscillatory diffusion of H$_2$O molecules.$^{196}$ Fitting of the SS model gives a diffusion coefficient $D = 7.27 \cdot 10^{-4}$ cm$^2$/s (obtained from the slope $\sim DQ^2$ law at low $Q$), an oscillatory residence time $\tau_0 = 1.7 \cdot 10^{-12}$ s, and a jump distance of $l = 2.1$ Å (which is reasonable since the distance between water molecules is usually 3 Å and agrees with previously reports)$^{110,177,196,431}$ However, the evolution of FWHM with $Q$ of TR-n4 and TR-n11 in H$_2$O hydration conditions are very similar, which suggests that the hydrogen dynamics are the same for both hydrated proteins (i.e., same mobility).
Figure 6-3. QENS analysis of D$_2$O- and H$_2$O- hydrated TR-n4 and TR-n11 proteins (measured in DCS). a) Lorentzian full width half max (FWHM) fitted to the D$_2$O/protein quasielastic spectra (VD model). b) FWHM fitted to the H$_2$O/protein quasielastic spectra (SS model). c) Elastic incoherent structure factor (EISF) of the D$_2$O/protein quasielastic spectra (confined diffusion model). d) Elastic incoherent structure factor (EISF) of the H$_2$O/protein quasielastic spectra (confined diffusion model).

The elastic incoherent structure factor (EISF) was analyzed as function of $Q$ (Figure 6-3c and Figure 6-3d), which is defined as the ratio of elastic intensity to the total intensity, and provides information on the geometry of the motions and the mobile fraction of hydrogen atoms involved in the dynamic process. EISF was modeled for confined diffusion of hydrogen atoms within a spherical volume as follows: $EISF_{sph}(Q) = [3j_1(Qr)/Qr]^2$, where $j_1$ is the first order Bessel function and $r$ is the radius of the confined spherical volume. Therefore, the EISF of TR
polypeptides can be written as: $EISF_{TR}(Q) = 1 - p_{mobile} + p_{mobile}EISF_{sph}(Q)$, where $p_{mobile}$ is the fraction of mobile hydrogen atoms, and is weighting the $EISF_{sph}$ term. The term $1 - p_{mobile}$ represents the elastic intensity from immobile fraction. The fraction of hydrogen atoms involved in motion for both proteins in D$_2$O and H$_2$O hydration conditions is calculated and listed in Table 6-1.

Table 6-1. Fit parameters of elastic incoherent structure factor (EISF) from quasielastic neutron scattering measurements of hydrated TR-n4 and TR-n11 proteins.

<table>
<thead>
<tr>
<th></th>
<th>TR-n4</th>
<th>TR-n11</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>Mw</td>
<td>15 kDa</td>
<td>42 kDa</td>
</tr>
<tr>
<td>$p$ (D$_2$O)</td>
<td>0.20</td>
<td>0.11</td>
</tr>
<tr>
<td>$r$ (D$_2$O)</td>
<td>3 Å</td>
<td>4 Å</td>
</tr>
<tr>
<td>$p$ (H$_2$O)</td>
<td>0.26</td>
<td>0.40</td>
</tr>
<tr>
<td>$r$ (H$_2$O)</td>
<td>3 Å</td>
<td>3 Å</td>
</tr>
</tbody>
</table>

While the radius of confined self-diffusion is similar for all samples, the fraction of mobile hydrogen atoms shows different trends in D$_2$O and H$_2$O hydration conditions. TR-n4 (D$_2$O) has higher $p_{mobile}$ than TR-n11 (D$_2$O), indicating that protein chains are more mobile in TR-n4 (deuterium is invisible to neutron scattering, therefore this measurements reflects protein dynamics only). This would agree with the protein network analysis in Chapter 5 that concluded that short SRT proteins (TR-n4) have a high density of network defects (loop structures), while longer polypeptides (TR-n11) are rich in tie chains. It is therefore expected that disordered chains are more physically constrained (lower $p_{mobile}$) in highly cross-linked networks. On the other hand, H$_2$O measurements show the opposite: TR-n11 has higher $p_{mobile}$ than TR-n4. Protein-H$_2$O measurements include protein and bound water dynamics, therefore the increase must arise from the water molecules bound to the protein. This implies that longer SRT polypeptides have a more mobile network of hydrogen bonded water molecules around the protein chains. Since all SRT-TR polypeptides are chemically identical (same building block), the difference in hydration must
come from the network morphology and conformational changes demonstrated in Chapter 5. It is expected that the protein morphology and dynamics in hydrated environments will strongly impact the transport properties of SRT-based materials, and therefore bulk proton and thermal conductivity of SRT-TR polypeptides are investigated next.

6.2.2 SRT-inspired protein networks: bulk proton conductivity

The proton conductivity of hydrated SRT-TR films was measured by electrochemical impedance spectroscopy (EIS) in a two electrode configuration impedance and data were visualized in Nyquist plots (Figure 6-4a). The impedance data shows characteristics of proton conduction, with a semicircle in the high-frequency range (bulk protonic impedance) and a linear region in the low frequency range (film/electrode interface)\textsuperscript{197}. The proton conductivity of the four SRT-TR polypeptides was measured at pH 7 and at different temperatures, and obtained a peak conductivity value of $3.48 \pm 0.30 \text{ mS cm}^{-1}$ for TR-n25 at 70 °C. Activation energies ($E_a$) for conduction were calculated from the linear fit to a temperature-dependent Arrhenius plot, and an average $E_a$ of $0.07 \pm 0.02 \text{ eV}$ was calculated.

Figure 6-4. Proton conductivity in SRT-inspired tandem repeat proteins. a) Characteristic Nyquist plot proton conductivity. b) Proton conductivity of SRT-TR proteins as function of number of tandem repeats $1/n$ (pH 7).
The proton conductivity and low $E_a$ was attributed to the presence of histidine and tyrosine in the amino acid sequence of SRT-TR. Histidine has an undisputed role in proton shuttles catalytic sites in enzymes due to its imidazole side chain (cycling between imidazolium and imidazolate states translocates two protons per event), and has been studied in detail in carbonic anhydrases.\textsuperscript{432–435} It is the major charged group ($\sim 7\%$) in the protein sequence and regulates the net charge of SRT proteins (isoelectric point of 6.7)\textsuperscript{211}. Charged amino acids (arginine, histidine, lysine, aspartic acid and glutamic acid) have shown a critical role in proton conducting proteins due to their ability to donate excess protons, to participate in hydrogen bonding, and to drive the protein assembly into morphologies that are favorable for proton transport\textsuperscript{394,395,424,436}. Tyrosine, which is one of the major amino acids SRT-TR sequence ($\sim 14\%$), is also involved in proton transport in biological systems (e.g., proton-coupled electron transfer in ribonucleotide reductase and photosystem II) due to its phenol side chain,\textsuperscript{437–441} and therefore it is expected to contribute to the bulk proton conductivity of proteins.\textsuperscript{442} Interestingly, proton conductivity of SRT-TR mimics scales almost linearly with repeat unit number $n$ (Figure 6-4b), which is a unique characteristic of SRT-TR proteins and provides additional control over the material properties\textsuperscript{85}.

Figure 6-5. Kinetic isotope effect in proton conductivity in SRT-TR proteins. a) Nyquist plots and b) conductivity of SRT-TR proteins in H$_2$O and D$_2$O (pH 7).

In order to confirm the bulk proton conductivity in SRT proteins, impedance spectroscopy measurements were repeated in deuterium oxide D$_2$O to investigate the nature of the charge carrier (Figure 6-5). The resistance in D$_2$O considerably increased, and the measured conductivity was
reduced 40 – 50 % for all samples, caused by the kinetic isotope effect (KIE). Classical KIE models predict that the rate of a proton hoping in an oxide is proportional to the \( v_{OH} \) vibration frequency.\(^{392,443,444}\) The OH/OD vibrational frequencies follow an approximate ratio of \( v_{OH} / v_{OD} \approx \sqrt{2} \).\(^{445}\) Hence, the mobility of a proton is expected to be \( \sim 1.4 \) times higher than that of a deuteron, and so should the conductivity. Although the predicted KIE often deviates from experimental measurements in diverse materials, measuring a decrease in conductivity of \( \sim 30 - 50 \% \) after deuterium substitution is a good indicator of proton hoping through the Grotthuss mechanism (proton hoping between neighboring molecules in a hydrogen bond network). The decay in conductivity caused by the KIE is consistent with proton conductivity observation in multiple other materials, and provides evidence for proton conduction.\(^{392,397,424,426,442-444,446,447}\)

**Figure 6-6.** a) Proton conductivity as function of water volume fraction \( \varphi \) (pH 7). b) Reduced proton conductivity (proton conductivity \( \sigma_{H} \) normalized to fully swollen proton conductivity \( \sigma_{H0} \)) as function of water volume fraction \( \varphi \) shows agreement with percolation theory and a power law is fitted to the data with \( \varphi_c \) as percolation threshold and \( \alpha \) as power law coefficient. Inset shows a linear dependence of \( \alpha \) on \( 1/n \) reciprocal repeat unit number.

As additional evidence for proton conduction, the proton conductivity in all four SRT-TR polypeptides was measured as a function of water content (Figure 6-6). Fully hydrated SRT-TR films have enough water molecules to form a hydrogen-bonded network (including water molecules bound to protein residues and backbone) that allows for Grotthus-type proton transfer (i.e., protons migrate along a chain of hydrogen-bonded water molecules).\(^{297,390-392,394}\) Lower conductivity was measured for decreasing water content, and no conductivity was observed below
a humidity threshold, which agrees with percolation theory of protein protonic transport (i.e. a minimum number of conducting water molecules is required for any transport to occur). The reduced conductivity (proton conductivity $\sigma_H$ normalized to fully swollen proton conductivity $\sigma_{H0}$) as function of water volume fraction follows a power law above the percolation threshold $\sigma_H/\sigma_{H0} = k (\phi - \phi_c)^\alpha$, where $\sigma_H$ is conductivity, $\sigma_{H0}$ is conductivity at saturation water content, $\phi$ is the water volume fraction, $\phi_c$ is the water volume fraction percolation threshold, $\alpha$ is the power law coefficient and $k$ is a constant (Table 6-2). The power law coefficient $\alpha$ which denotes the dimensionality of the system increases linearly with $1/n$, suggesting that the bound water conductive pathways are more efficient at higher tandem repetition (Figure 6-6 inset).

Table 6-2. Proton conductivity of SRT-TR proteins, activation energy and percolation parameters (pH 7).

<table>
<thead>
<tr>
<th>n</th>
<th>TR-n4</th>
<th>TR-n7</th>
<th>TR-n11</th>
<th>TR-n25</th>
<th>TR-n(\infty)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mw (kDa)</td>
<td>15</td>
<td>25</td>
<td>42</td>
<td>86</td>
<td>$\infty$</td>
</tr>
<tr>
<td>$\sigma_H$, 20 °C, $\phi=0.3$ (mS/cm)</td>
<td>0.47 ± 0.13</td>
<td>0.72 ± 0.13</td>
<td>0.93 ± 0.11</td>
<td>1.75 ± 0.21</td>
<td>3.14</td>
</tr>
<tr>
<td>$\sigma_H$, 50 °C, $\phi=0.3$ (mS/cm)</td>
<td>0.93 ± 0.04</td>
<td>1.19 ± 0.08</td>
<td>1.85 ± 0.10</td>
<td>2.61 ± 0.30</td>
<td>4.11</td>
</tr>
<tr>
<td>$\sigma_H$, 70 °C, $\phi=0.3$ (mS/cm)</td>
<td>1.38 ± 0.12</td>
<td>1.54 ± 0.4</td>
<td>2.70 ± 0.08</td>
<td>3.48 ± 0.30</td>
<td>4.58</td>
</tr>
<tr>
<td>$E_A$, $\phi=0.3$ (eV)</td>
<td>0.08 ± 0.02</td>
<td>0.06 ± 0.03</td>
<td>0.08 ± 0.02</td>
<td>0.05 ± 0.03</td>
<td>0.07</td>
</tr>
<tr>
<td>$\phi_c$ (percolation)</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>N/A</td>
</tr>
<tr>
<td>$\alpha$ (percolation)</td>
<td>1.3</td>
<td>1.2</td>
<td>1.1</td>
<td>1</td>
<td>N/A</td>
</tr>
<tr>
<td>k (percolation)</td>
<td>7.5</td>
<td>7.5</td>
<td>7.5</td>
<td>7.5</td>
<td>N/A</td>
</tr>
</tbody>
</table>

All four SRT-TR polypeptides are chemically identical since they are based on the same repetitive building block (i.e. same amino acid composition and order within the crystal/amorphous
segments). Therefore, the different proton conductivity of SRT-TR polypeptides with different repeat unit number $n$ must arise from structural and morphological differences. Extensive experimental and computational research on conjugated conducting polymers has shown a clear relationship between conductivity and morphology where amorphous tie-chains connecting the crystallites favor the charge transport (long molecular weight chains in semicrystalline polymers provide better connectivity between ordered regions).\textsuperscript{451–455} The findings presented here suggest that a similar relationship might be valid for proton conduction in protein materials and other biomolecules.\textsuperscript{424,426,427,442,456,457} In general, proton conductivity scales linearly with the mobility and the concentration of protons in protein films.\textsuperscript{396} As demonstrated by QENS measurements, the mobility of protons (dynamics of bound water revealed the same diffusion coefficient, with identical FWHM in QENS spectra) is the same independently of the protein molecular weight. However, the concentration of mobile hydrogen (fraction of mobile hydrogen in bound water) is significantly higher in higher molecular weight proteins (TR-n11) despite having the same water absorption. Therefore, the proton conductivity is proportional to the fraction of hydrogen involved in the process and to the effectiveness of the network (tie chain density): $\sigma_H \sim p_{\text{free}} \sim \varepsilon_{\text{eff}}$. Due to crystalline $\beta$-sheet nanoconfinement, long polypeptides can connect the $\beta$-sheet ordered domains through tie chain strands, which facilitates the proton transport and allows a higher number of protons to participate in the process.

### 6.2.2.1 Semicrystalline morphology in proton conducting SRT-inspired protein networks

The interplay between microstructure and transport phenomena in synthetic polymer systems has been extensively studied in order to understand the conductive properties of organic materials.\textsuperscript{451–455} Most conjugated polymers have a semicrystalline morphology, where semi-ordered domains and amorphous disordered chains coexist. In such microstructure, the ordered phase is responsible for charge transport due to an energy barrier to move from ordered to amorphous regions.\textsuperscript{451,454} Hence, control of the semicrystalline morphology is key to optimizing the conducting properties. Following a similar approach, the effect of the semicrystalline morphology of SRT-TR polypeptides on the proton conductivity is investigated here.
First, the conductivity of TR-n11 protein materials with varying crystallinity (i.e., β-sheet content) was measured. TR-n11 materials with varying crystallinity were prepared by solution-based and thermal-based processing methods that include: 1) casting from HFIP solutions, 2) water washing (H₂O), 3) hot water washing (H₂O, 70 °C), 4) compression molding (H₂O, 70 °C, 1 MPa), and 5) methanol treatment (immersion in methanol for 24 h). These materials were exposed to water washing, temperature, pressure, and organic solvents in processing parameters analogous to those of Lv18 protein materials (Chapter 4).

Figure 6-7a shows the amide I FTIR spectra of the TR-n11 materials with varying crystallinity. Shifts in the amide I band (sensitive to C=O stretching vibrations) denote changes in the backbone conformation due to the rearrangements in the hydrogen bonding pattern. A shift towards lower wavenumber can be observed from cast samples (centered around 1650 cm⁻¹, random coil region) to meOH treated samples (centered around 1620 cm⁻¹, β-sheet region). These shifts indicate a progressive increase in β-sheet structures (1610-1632 cm⁻¹ region) and a decrease in disordered chains (1640-1655 cm⁻¹), which is consistent with previous structural analysis of Lv18 recombinant SRT protein. FSD analysis of the amide I band gives an increasing β-sheet content of 41 %, 54 %, 55 %, 59 %, and 63 % for cast, wash, annealed, compressed, and meOH-treated TR-n11 samples respectively (Figure 6-7b).

![Figure 6-7. Semicrystalline morphology in proton conductivity. a) Amide I FTIR spectra of cast, wash, annealed, compressed, and methanol-treated TR-n11 materials shift towards β-sheet region (1620 cm⁻¹) with protein processing. b) Proton conductivity increases with the β-sheet content (pH 7).](image)
Proton conductivity measurements by impedance spectroscopy revealed an increasing proton conductivity with β-sheet content (Figure 6-7b), up to a ~50 % increase at the maximum crystallinity (methanol treatment). Interestingly, water uptake is significantly lower for more crystalline samples (~45 % to ~20 % of water absorption for wash and meOH samples respectively), as reported for Lv18 proteins. This implies that higher crystallinity proteins are more efficient conductors even with significantly lower absorbed water. In β-sheet structures, the amide groups are involved in very stable hydrogen bonding interactions with neighboring strands and they are not available to water molecules, while disordered chains are responsible for the major water absorption (water molecules interact with amides in amorphous chains). However, it is expected that water molecules form a hydrogen bonding shell around the β-sheet layers (due to the contribution of polar amino acids in the crystal segments such as serine, threonine, and histidine). The presence of these water molecule clusters around the β-sheets would enhance the overall hydrogen bonding network, and yield more efficient conductive pathways for protons. Hence, the β-sheet content and protein network morphology prove important structural parameters in controlling the conductive properties of structural proteins and SRT in particular.424,442,457

Furthermore, the effect of backbone orientation on proton conductivity is analyzed. The effect of single chain conformation and orientation in conjugated polymers has been studied in recent experimental and computational work.451,453,454 In conjugated polymers the energy barrier for charge to hop to a neighboring chain is significantly higher than the barrier for hopping along the same chain.453 However, kinks and turns in randomly oriented chains that are not aligned with the external field hinder the charge transport along the chain, and inter-chain hoping becomes possible at the expense of the overall charge conductivity. Hence, the orientation of the disordered chains can be controlled in order to optimize the transport properties by alignment with the field. This approach is tested on SRT proteins. The structural evolution upon stretching was analyzed in Chapter 5, which revealed the strain hardening of TR-n11 polypeptides when deformed up to 300 % strain. FTIR spectroscopy and SAXS diffraction analysis showed that the disordered chains align with the load direction resulting in an anisotropic structure, and for this reason it is expected that proton transport will be sensitive to stretching of the protein.
Figure 6-8. SRT-TR proteins as stretchable proton conductors. a) Nyquist plots of stretched TR-n11 protein films show a decrease in impedance with strain. b) Proton conductivity increases with strain due to the alignment of the protein structure (pH 7). However, the conductivity is constant in the perpendicular strain direction (anisotropic transport).

TR-n11 films were progressively stretched to 270 % strain and the proton conductivity was measured by impedance spectroscopy. The Nyquist plots of the stretched films show a clear decrease in the impedance semicircle size with increasing strain (Figure 6-8a). This indicates a decrease in resistance (intercept at $Z'' = 0$) and increase in conductivity. The increase in conductivity with strain is shown in Figure 6-8b. It can be observed that the conductivity sharply increases in the 0 % - 75 % region approximately, and then increases at a lower rate up to 275 % strain (measurements at further strain resulted in damage in the samples). The regions correspond to the homogeneous stretching of the randomly oriented disordered chains, followed by the unraveling and orientation of the disordered strands (which agree with the network analysis in Chapter 5). Furthermore, the conductivity was measured in the parallel and perpendicular directions with respect to the stretching in order to evaluate the anisotropy (Figure 6-8b inset). While the conductivity along the stretching direction increases with strain, it remains approximately constant in the perpendicular direction. This indicates that the alignment of the tie chains along a specific direction generates more efficient conductive pathways for proton transport. However, proton transport is not improved in the misaligned direction (most likely due to the necessity for charge hoping between non-connected chains, analogous to conjugated
polymers).\textsuperscript{451,453} A similar effect has been reported in oriented sulfonated polymers (Nafion), where the preferential alignment of the water channels leads to anisotropic conductivity.\textsuperscript{458} Although proton transport along the backbone has been observed in some proteins such as gramicidin (helical proton channel),\textsuperscript{395,459,460} no clear relationship between protein morphology and bulk conductive properties has been proposed to date.\textsuperscript{424,442,457} The measurement of anisotropic conductivity (with an increase of $\sim x2$) agrees with the morphology analysis of stretched TR-n11, and provides further evidence on the importance of chain orientation and morphology in the transport properties of protein materials.

![Figure 6-9](image)

Figure 6-9. Bulk proton conductivity of biological materials. Proton conductivity benchmark, $\sigma_H$, for protein-based materials as a function of inverse molecular weight ($M_w$). SRT-TR proteins have the highest bulk proton conductivity among all reported biological materials to date.

The proton conductivity of SRT-TR reported here compares favorably to those of other proton conducting materials, and are among the best reported for any bulk biological material up to date such as silk fibroin, maleic chitosan, keratin, collagen, reflectin, ampullae of Lorenzini jelly, melanin, lysozyme, maleic chitosan, bovine serum albumin (Figure 6-9, Table 6-3). Besides the
high proton conductivity values, SRT proteins provide a platform for engineering protein-based materials with tailored properties by controlling the sequence and repeating building block units. Repeating units allow for the formation of stable bioelastomeric networks capable of withstanding large deformations, which is a current limitation in state-of-the-art proton conducting materials. Hence, the work presented here introduces design rules in the engineering of protein-based materials with programmable properties.
Table 6-3. Bulk proton conductivity of biological materials

<table>
<thead>
<tr>
<th>Material</th>
<th>Mw [kDa]</th>
<th>$\sigma$ [mS/cm]</th>
<th>$E_A$ [eV]</th>
</tr>
</thead>
<tbody>
<tr>
<td>TR-n4 *</td>
<td>15</td>
<td>$1.38 \pm 0.12$</td>
<td>0.08 ± 0.02</td>
</tr>
<tr>
<td>TR-n7 *</td>
<td>25</td>
<td>$1.94 \pm 0.2$</td>
<td>0.06 ± 0.02</td>
</tr>
<tr>
<td>TR-n11 *</td>
<td>42</td>
<td>$2.70 \pm 0.22$</td>
<td>0.08 ± 0.02</td>
</tr>
<tr>
<td>TR-n25 *</td>
<td>86</td>
<td>$3.48 \pm 0.3$</td>
<td>0.05 ± 0.02</td>
</tr>
<tr>
<td>TR-n∞ *</td>
<td>$\infty$</td>
<td>3.66</td>
<td>–</td>
</tr>
<tr>
<td>SRT * * Loligo vulgaris</td>
<td>36 (avg)</td>
<td>$2.36 \pm 0.42$</td>
<td>–</td>
</tr>
<tr>
<td>SRT * * Loligo pealei</td>
<td>40 (avg)</td>
<td>$2.20 \pm 0.24$</td>
<td>–</td>
</tr>
<tr>
<td>SRT * Ilex illecebrosus</td>
<td>35 (avg)</td>
<td>$2.33 \pm 0.47$</td>
<td>–</td>
</tr>
<tr>
<td>TR-n7 polyA (no histidines) *</td>
<td>25</td>
<td>$0.26 \pm 0.22$</td>
<td>–</td>
</tr>
<tr>
<td>Silk (cocoon) *</td>
<td>~ 390</td>
<td>$0.18 \pm 0.22$</td>
<td>–</td>
</tr>
<tr>
<td>Silk (reconstituted) *</td>
<td>~ 150</td>
<td>$0.28 \pm 0.21$</td>
<td>–</td>
</tr>
<tr>
<td>Maleic chitosan (gladius) *</td>
<td>~ 50</td>
<td>$0.46 \pm 0.21$</td>
<td>–</td>
</tr>
<tr>
<td>Keratin *</td>
<td>40-60</td>
<td>0.001</td>
<td>–</td>
</tr>
<tr>
<td>Collagen *</td>
<td>115-235</td>
<td>0.02</td>
<td>1.32 ± 0.03</td>
</tr>
<tr>
<td>Reflectin *</td>
<td>40</td>
<td>$1.2 \pm 1$</td>
<td>0.22 ± 0.05</td>
</tr>
<tr>
<td>AoL Jelly *</td>
<td>20-200</td>
<td>$1.8 \pm 0.9$</td>
<td>–</td>
</tr>
<tr>
<td>Melanin *</td>
<td>0.3</td>
<td>$0.71 \pm 0.48$</td>
<td>–</td>
</tr>
<tr>
<td>Lysozyme *</td>
<td>14.3</td>
<td>$10^{-5}$</td>
<td>–</td>
</tr>
<tr>
<td>Maleic chitosan *</td>
<td>50</td>
<td>0.7</td>
<td>–</td>
</tr>
<tr>
<td>Bovine Serum Albumin (monolayer)</td>
<td>66.5</td>
<td>$4 \times 10^{-3}$</td>
<td>–</td>
</tr>
<tr>
<td>Bovine Serum Albumin (mat) *</td>
<td>66.5</td>
<td>0.5</td>
<td>–</td>
</tr>
<tr>
<td>Poly aspartic acid p(Asp)$_{100}$</td>
<td>133.1</td>
<td>0.5</td>
<td>–</td>
</tr>
<tr>
<td>YYACAYY peptide *</td>
<td>0.92</td>
<td>0.1</td>
<td>–</td>
</tr>
</tbody>
</table>

* This work
6.2.3 SRT-inspired protein networks: thermal conductivity

The thermal conductivity of SRT-TR polypeptides was measured using time-domain thermoreflectance (TDTR) by Dr. Patrick E. Hopkins and John Tomko. Figure 6-10 shows the thermal conductivity of both dry and hydrated SRT-TR polypeptides. The conductivity of dry SRT proteins is constant and independent of the number of repeat units \( n \) (\( k \approx 0.33 \text{ W m}^{-1} \text{ K}^{-1} \)). For atomically disordered systems, the thermal conductivity is driven by the heat capacity and diffusivity, which is related to both the mean square displacement (MSD) and the force constant among neighboring atoms. As demonstrated by neutron spectroscopy, dry SRT-TR proteins have very low mobility (dynamics dominated by localized methyl group rotations) in their glassy state since the disordered strand are constrained by a dense network of inter- and intra-chain hydrogen bonding. A change in conductivity with increasing \( n \) (in the glassy) state is not expected since chain dynamics are equally hindered in all proteins. The reported \( k \) falls within those reported for amorphous polymers, usually regarded as thermal insulators (0.1 – 0.6 W m\(^{-1}\) K\(^{-1}\)).

![Figure 6-10. Thermal conductivity of SRT-TR proteins. a) Thermal conductivity of dry TR proteins is constant, but thermal conductivity of hydrated (water content ~40%) TR proteins scales linearly with 1/n. b) Thermal conductivity switch for TR-n11 shows dynamic thermal conductivity as function of hydration over multiple cycles.](image-url)
Interestingly, hydrated SRT-TR polypeptides exhibit higher thermal conductivities than their dry analogous. This is unexpected since the bulk modulus of SRT protein films decrease upon hydration. Even if one considers the thermal conductivity of water ($k_{\text{water}} \approx 0.6 \text{ W m}^{-1} \text{ K}^{-1}$), the measured conductivities are significantly higher (up to 1.3 W m$^{-1}$ K$^{-1}$ for TR-n25). However, this increase can be explained by the higher MSD measured by neutron spectroscopy, which translates into an increase in thermal diffusivity as the weak nonbonding interchain interactions (which are abundant in the dry state and are known to inhibit thermal conductivity) are removed by water molecules.  

Table 6-4. Thermal conductivity and $C_p$ of SRT-TR proteins

<table>
<thead>
<tr>
<th>Sample</th>
<th>TR-n4</th>
<th>TR-n7</th>
<th>TR-n11</th>
<th>TR-n25</th>
<th>TR-n$\infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>4</td>
<td>7</td>
<td>11</td>
<td>25</td>
<td>$\infty$</td>
</tr>
<tr>
<td>Mw (kDa)</td>
<td>15</td>
<td>25</td>
<td>42</td>
<td>86</td>
<td>$\infty$</td>
</tr>
<tr>
<td>$k_{\text{dry}}$ (W/mK)</td>
<td>0.35 ± 0.05</td>
<td>0.34 ± 0.06</td>
<td>0.27 ± 0.05</td>
<td>0.33 ± 0.05</td>
<td>0.33</td>
</tr>
<tr>
<td>$C_p_{\text{dry}}$ (J/g °C)</td>
<td>1.30 ± 0.02</td>
<td>1.27 ± 0.02</td>
<td>1.28 ± 0.02</td>
<td>1.27 ± 0.04</td>
<td>-</td>
</tr>
<tr>
<td>$k_{\text{hydrated}}$ (W/mK)</td>
<td>0.58 ± 0.03</td>
<td>0.93 ± 0.05</td>
<td>1.12 ± 0.13</td>
<td>1.30 ± 0.07</td>
<td>1.43</td>
</tr>
<tr>
<td>$C_p_{\text{hydrated(protein)}}$ (J/g °C)</td>
<td>2.78 ± 0.08</td>
<td>2.80 ± 0.02</td>
<td>2.80 ± 0.01</td>
<td>2.80 ± 0.03</td>
<td>-</td>
</tr>
<tr>
<td>$C_p_{\text{hydrated(protein)}}$ (J/g °C)</td>
<td>1.83 ± 0.08</td>
<td>1.88 ± 0.04</td>
<td>1.87 ± 0.03</td>
<td>1.88 ± 0.04</td>
<td>-</td>
</tr>
<tr>
<td>$Z_{\text{TCS}}$</td>
<td>0.79 ± 0.22</td>
<td>1.87 ± 0.34</td>
<td>2.46 ± 0.54</td>
<td>2.99 ± 0.36</td>
<td>3.40</td>
</tr>
</tbody>
</table>
Furthermore, the thermal conductivity of hydrated SRT-TR films scales linearly with $1/n$ to a maximum of $k_{TR-n\infty} = 1.43 \text{ W m}^{-1} \text{ K}^{-1}$ (Table 6-4). This $n$ dependence agrees with the network defects calculations (elastic effectiveness $\varepsilon_{\text{eff}}$ parameter) from Chapter 5, which predict high density of network defects (loop strands and dangling ends) in low molecular weight proteins (TR-$n4$) and high density of effective strands (tie chain strands) in high molecular weight proteins (TR-$n25$). Hence, the results suggest that tie chains are the rate-limiting step for thermal transport (analogous to electrical conductivity in conjugated conducting polymers).\textsuperscript{420,421,454}

The thermal conductivity can be estimated using the minimum thermal conductivity model (MTCM), which describes thermal transport in highly disordered amorphous materials. In this model, the thermal conductivity $k$ scales with the velocity of sound (which further depends on elastic modulus $E^{1/2}$). As described in Chapter 5, the modulus of SRT-TR proteins scales with the elastic effectiveness parameter $\varepsilon_{\text{eff}}$, and therefore it is reasonable that thermal conductivity $k$ scales with $\varepsilon_{\text{eff}}^{1/2}$. In general, phonons are affected by a combination of four scattering processes:\textsuperscript{420,476} (i) Intramolecular phonon scattering depends on the covalent bonds along the protein chain (thermal conductivity along single polymer chains can be as high as $10^4$ W m$^{-1}$ K$^{-1}$), which is constant for all SRT-TR since the sequence building block is the same.\textsuperscript{418,422} (ii) Intermolecular phonon scattering is caused by weak interchain non-bonding interactions (i.e., van der Waals and electrostatic), which are considered a bottleneck in polymer heat transfer.\textsuperscript{421,422} Weak non-bonding interactions are expected to be low in hydrated SRT proteins due to the plasticizing effect of water (water molecules disrupt weak interactions in the protein matrix), and therefore no significant change in conductivity is expected from interchain scattering. (iii) Boundary phonon scattering is caused by scattering at the chain ends. Low molecular weight polymers have a higher fraction of dangling end structures, while the fraction in high molecular weight polymers will progressively decrease. Therefore, it is expected that the conductivity is inhibited by boundary scattering at low molecular weight.\textsuperscript{414,420,474} (iv) Structural phonon scattering arises from morphology and chain orientation.\textsuperscript{414,420,474,476} In a polymer crystal, the chains are densely packed and aligned in a specific direction (i.e., lamellae structures). The oriented crystal-forming chains can serve as efficient conductive pathways along the crystal direction due to the stiff covalent bonding (anisotropic thermal conductivity has been measured in crystalline polymers, with differences between axial and perpendicular conductivities of at least an order of magnitude).\textsuperscript{414,420} The measured SRT-TR films all have the same isotropic $\beta$-sheet
fraction, therefore the change in conductivity must not originate here but in the disordered amorphous phase. The random orientation and curvature of amorphous chains cannot propagate vibrational waves efficiently, resulting in a low mean-free path of the phonons.\textsuperscript{420} For this reason, low thermal conductivity is observed in most amorphous polymers (loosely packed coiled and entangled interchain structures inhibit phonon propagation). However, the tie chain strands that connect the crystalline domains in high molecular weight proteins would enhance the heat transport, which agrees with the measured increase in thermal conductivity ($1/n$ dependence, equivalent to tie chain fraction $\varepsilon_{\text{eff}}$). Therefore, the control of thermal conductivity through increasing number of tandem repeat units $n$ can lead to conductivities up to 1.43 W m$^{-1}$ K$^{-1}$, which exceeds those of conventional polymers. It is expected that by increasing the crystalline fraction and aligning the disordered chains through mechanical stretching, the thermal conductivity can be increased further, as has been demonstrated for amorphous polymers in recent research.\textsuperscript{419,421–423,477}

![Image](image.png)

Figure 6-11. Thermal conductivity switching figure of merit $|Z_{\text{TCS}}| = |\Delta k/k_{\text{min}}|$ benchmark of diverse state-of-the-art materials.
The thermal conductivity change between dry and hydrated states is completely reversible for at least twenty cycles (Figure 6-10b). This cycling behavior enables the potential use of SRT-TR polypeptides as thermal conductivity switches with applications in temperature regulation, thermal storage, and sensing. Most thermal switches are based on liquid-solid and solid-solid reversible phase transitions (which include metal-semiconductor-insulator transitions). The thermal conductivity switching behavior can be evaluated by a dimensionless figure of merit \( |Z_{TCS}| = \frac{|\Delta k/k_{min}|}{\text{change in thermal conductivity divided by smaller thermal conductivity}} \). Most materials that exhibit reversible thermal switching do so at high temperature (700 – 1000 K range for high \( Z_{TCS} \)), and most require liquid-solid transitions in order to achieve high \( Z_{TCS} \). SRT-TR polypeptides can overcome these limitations since they exhibit thermal switching upon hydration, and can operate at room temperature without any loss of properties. \( Z_{TCS} \) of SRT-TR polypeptides is higher than those of other reported materials in the 0 °C to 100 °C range (Figure 6-11, Table 6-5), which makes SRT proteins an attractive material for thermal devices.
Table 6-5. Thermal conductivity switching benchmark of diverse state-of-the-art materials

| Material            | Switching Temperature [K] | Thermal Conductivity Switch Figure of Merit $|Z_{TCS}| = |\Delta k/k_{\text{min}}|$ |
|---------------------|----------------------------|-----------------------------------------------|
| TR-n4 *             | 295                        | 0.79 ± 0.22                                  |
| TR-n7 *             | 295                        | 1.87 ± 0.34                                  |
| TR-n11 *            | 295                        | 2.45 ± 0.54                                  |
| TR-n25 *            | 295                        | 2.99 ± 0.36                                  |
| TR-n∞ *             | 295                        | 3.40                                          |
| H₂O                 | 273                        | 2.62                                          |
| h-C/C₁₈H₃₄           | 291                        | 2.27                                          |
| LiₓCoO₂              | 295                        | 0.459                                         |
| C₇₀                  | 298                        | 0.67                                          |
| PZT                 | 298                        | 0.13                                          |
| RbCl                 | 298                        | 0.85                                          |
| B nanoribbon         | 300                        | 0.516                                         |
| SWCNT/n-C₁₈H₃₈      | 303                        | 2.01                                          |
| Ga                   | 303                        | 0.46                                          |
| PEG4000             | 313.15                     | 0.28                                          |
| K                    | 337                        | 0.80                                          |
| VO₂                  | 340                        | 0.50                                          |
| RM257                | 353                        | 0.71                                          |

* This work
6.3 Conclusions

In summary, the effects of tandem repetition and network morphology on the chain dynamics and transport properties (proton and thermal conductivity) of SRT-TR polypeptides were analyzed in this chapter. Tandem repetition clearly shows both proton and thermal conductivity enhancement through control of the protein network morphology and defects. Low molecular weight repetitive proteins (TR-n4) are rich in deficient disordered structures (loops and dangling ends), which introduce network defects that inhibit the transport phenomena. As the molecular weight and number of tandem repetitions increase, the defective disordered structures are replaced by effective network strands (tie chains) that covalently connect the β-sheet crystallites and enhance the conductive pathways. By reducing the network defects and maximizing the effective strands (high density of tie chains), SRT proteins exhibit a peak proton conductivity of 3.66 mS/cm at pH 7 (highest conductivity reported to date for a biological material) and a peak thermal conductivity of 1.43 W m⁻¹ K⁻¹ (which exceeds that of most polymer materials). The reported conductivity values fall short when compared to non-biological materials (sulfonated polymers such as Nafion exhibit proton conductivities of 70 mS/cm, and metals such as copper have thermal conductivities of 400 W m⁻¹ K⁻¹) used in industrial applications such as fuel cell membranes and heat exchangers. However, the proton and thermal conductivity of SRT-based materials, together with their programmable mechanical properties (flexible and stretchable), make them an attractive alternative for biomedical applications (including bioelectronics and biosensing).
Figure 6-12. Master plot for physical properties of SRT and SRT-TR proteins (mechanical, proton conductivity, and thermal conductivity) as function of reciprocal molecular weight or number of repeat units.

Moreover, the results offer a window into the evolutionary dynamics of genetic duplication in naturally occurring tandem-repeat protein materials. Each squid ring-tooth is formed from multiple homologous sequences with diverse polypeptide lengths; the distribution of lengths varies greatly between different squid species. While the functional properties of individual sequences vary greatly with polypeptide length, the different mixtures of sequences present in different squid species lead to the same average molecular weight ($\bar{\varepsilon}_{\text{avg}}$) and overall functional properties (Figure 6-12). This suggests that polydispersity may be a key factor in the effective evolution and optimization of material properties in natural systems. Collectively, the results suggest a new rationale via tandem repetitions for the discovery and engineering of biomaterials with extreme tunable and dynamic functional properties, and introduce design rules for the optimization of transport properties in protein-based materials.
Chapter 7. Applications of SRT

SRT proteins have physical and chemical properties that make that can be tailored by regulation of the primary amino acid sequence and the protein nanostructure (as described in previous chapters). In addition to biocompatibility and biodegradability, SRT protein materials have tunable mechanical, thermal, optical, and conducting properties that make them attractive programmable materials for a variety of applications. For example, the semicrystalline morphology can be tailored to adjust the modulus, stretchability, water absorption, thermo-optic coefficient and glass transition of SRT-based materials, which are critical in the performance of free-standing materials, durable, flexible, and efficient protein devices. In this chapter, the processing of SRT and fabrication of SRT-based materials (from nanoscale to macroscale) through solution-based and thermal-based methods are discussed. Then, this chapter explores potential applications of SRT proteins in the biomedical field, and showcases SRT-based proof-of-concept materials and devices. First, the pressure sensitive adhesive properties of SRT are investigated, showing an underwater adhesion strength higher than other biological adhesives. Second, the self-healing properties of SRT are studied, showing that the material properties are mostly conserved after the healing process and exhibiting potential use of SRT proteins as self-healing coatings in textiles. Last, flexible SRT-based photonic devices (optical waveguides, filters, and switches based on whispering gallery mode resonators) are fabricated, exhibiting high quality factors and high power efficiency.

7.1 Fabrication and processing of SRT materials

Despite the recent advances in protein engineering, processing of long-repetitive polypeptides remains one of the major challenges in developing protein-based materials. In most processing methods, proteins are exposed to severe conditions (temperature, pressure, pH, presence of solvents, etc.) that alter the protein nano/micro-structure and ultimately compromise the final properties of the materials. Consequently, many proteins and biological materials with remarkable properties have not yet been transformed into advanced materials and devices due to limitations in stability and bulk processing. Within this context, SRT proteins are superior to most repetitive proteins since they can be bulk processed by an array of methods analogous to those of
the polymer industry without significant loss in structure and properties. These processing and fabrication possibilities of SRT proteins are highlighted in this section.

### 7.1.1 Solution processing

Solution-based processing is the most common method used in most fibrous or structural proteins. Roughly, it involves the solubilization of a protein in solvent and subsequent aggregation by removal of the solvent via ambient or vacuum assisted evaporation. Water-soluble proteins can be easily dissolved in aqueous buffers, but the solubilization of structural proteins requires the disruption of aggregated protein structure. This step can involve acidic or basic conditions and the addition of salts or organic solvents. However, disruption of crosslinks is only possible for weakly bonded assemblies (e.g., hydrogen-bonded β-sheet proteins such as silk and SRT or hydrophobic interactions in helical polypeptides). For these proteins, aggregation or precipitation can be re-initiated by several methods such as salting out, isoelectric precipitation, the addition of miscible solvents, or evaporation of the solvent. However, these conditions can cause proteins to misfold or become kinetically trapped into undesirable assembly states; hence, the physical properties of these samples may vary significantly compared to native assemblies. Solution-based processing also involves the generation of solvent waste (often hazardous), increased processing time due to drying and purification steps, and the possibility of irreversible aggregation. For example, a myriad of solution-based processing methods for the fabrication of silk-based materials have been developed but all require prior degumming and solubilization of silk fibroin, which involves additional dialysis steps and the use of harsh chemicals. Similarly, many solution-based methods have been proposed for the fabrication of aligned collagen but the resulting mechanical properties of the reconstituted materials are significantly lower than natural collagen due to the lack of hierarchical structure.

Solution-based processing is applicable to SRT proteins as it is to many other biological materials. SRT proteins are stabilized by β-sheet structures that act as physical crosslinks, and are not water-soluble. Therefore, the β-sheet elements must be disrupted in order to solubilize the protein. To this purpose, several pH conditions (acidic pH below 3 and basic pH above 10), salts (lithium bromide, calcium chloride, calcium nitrate, guanidinium chloride), surfactants (sodium dodecyldiphenylsulfonate) and organic solvents (dimethylsulfoxide, hexafluoroisopropanol) can be used. SRT protein solutions can be prepared with high protein concentrations (above 100
mg/mL) depending on the solubility limit of the specific solvent cocktail (i.e., higher protein concentration results in high viscosity that complicates the processing).

Figure 7-1. a) Native and/or b) synthetic SRT proteins are c) dissolved and processed into a variety of d) complex geometries. SRT materials fabricated by solution-based processing: e) thin films are fabricated by drop casting. f) Sponges by particle-assisted casting and subsequent etching of the particles. g) Nanoparticles are prepared by salting out or by addition of a surfactant. h) Coating of textiles and other substrates is performed by dip coating. i) Microfibers are prepared by electrospinning. j) Complex 3D geometries are fabricated by mold casting.

SRT-based materials have been recently fabricated using solution-based processing (Figure 7-1). Transparent and flexible freestanding films (Young’s modulus of 0.7-0.8 GPa) with thickness ranging from a few micrometers to several hundred micrometers were be easily fabricated by drop casting. SRT-based scaffold sponges were prepared by a particle templating casting process, which consists of casting a SRT-micro/nanoparticle composite and subsequent etching of the particles. This process allows for a broad range of particle materials because SRT is not soluble in water or most organic solvents, which facilitates particle etching. The pore size is designed by
selecting the appropriate particle size, and ultimately controls the mechanical properties of the final sponge.\textsuperscript{265,492} SRT nanoparticles were prepared by multiple aggregation methods based on salting out, solvent exchange or isoelectric aggregation.\textsuperscript{83} SRT proteins are used for biomolecule encapsulation and controlled release, such as the encapsulation of enzymes, DNA, and dyes.\textsuperscript{265,493,494} Coating of complex substrates such as knitted or woven fabrics were performed by dip coating to fabricate advanced textiles that share SRT’s surface properties.\textsuperscript{495} Microfibers were spun by several methods including miscible solvent exchange and electrospinning.\textsuperscript{83} Complex objects such as toroidal whisper-gallery-mode microresonators were made by micro/nano-mold casting, enabling the fabrication of protein-based photonic devices.\textsuperscript{496} These showcased materials are varied examples of the capabilities of SRT-protein solution-based processing, and find applications in the food, pharmaceutical, textile and biomedical industry.

7.1.2 Thermal processing

Thermoplastic processing consists of softening the processed material with heat, forming it into a particular shape while soft, and hardening by cooling. These processing methods offer a series of advantages over solution processing that primarily derive from the elimination of solvents in the process. Eliminating solubilization and drying steps reduce the total processing time. Structure and properties of the materials are usually conserved since defects arising from residual solvent are reduced. Due to the reversible nature of their thermal transitions, thermoplastic materials can be processed multiple times over their life cycle and therefore can be recycled. These aspects, together with the versatility and low cost of the production systems, make thermoplastic processing methods the most extensively used in the polymer industry.

Most structural proteins are not thermoplastic since they are chemically cross-linked (crosslinking via disulfide bonds in keratin, tyrosin in Resilin, and lysine in elastin).\textsuperscript{47,72,337} The chemical cross-linking prevents the protein chains from moving past another and therefore material does not flow. For this reason, the processing of structural proteins presents many limitations,\textsuperscript{497} and solution casting is most commonly used.\textsuperscript{492} Non-chemically cross-linked proteins that exhibit thermoplasticity (including but not limited to casein, whey, soy, wheat gluten, gelatin, and SRT) can be processed by thermal-based methods.\textsuperscript{239,498–502} Thermoplastic proteins are typically processed above their glass transition temperature $T_g$, above which the polypeptide chains relax and the mechanical properties dramatically drop. However, the processing temperature should not
exceed the denaturation temperature $T_d$, at which the protein unfolds and irreversibly aggregates (for globular proteins), or the thermal degradation temperature $T_{\text{deg}}$. The temperature range in which the protein can be processed (between $T_g$ and $T_d$ or $T_{\text{deg}}$) is typically very narrow and often limits the processing options. Hence, plasticizers such as water, or certain oils or alcohols (e.g., glycerol, butanediol) are mixed with protein to increase the mobility of the protein chains and therefore decrease the $T_g$. \cite{83,167,211,239,388,499,503} This results in a broader processing temperature range and expands the processing capabilities of the plasticized material system (i.e., the material is softer at lower temperatures and therefore it is easier to shape).

Figure 7-2. a) Native and/or b) synthetic SRT proteins are c) thermally processed into a variety of d) complex geometries. SRT materials fabricated by thermal-based processing: e) fibers and rods are fabricated by extrusion, f) complex 3D geometries are fabricated by injection molding, g) thin films are fabricated by hot pressing, h) fibers are fabricated by drawing, i) nanoscale objects are fabricated by template-assisted nanowetting, and j) patterned surfaces are fabricated by compression molding.
SRT proteins (both native and engineered) have demonstrated a great versatility in their processing capabilities including thermoplastic processing methods. SRT proteins are stable up to temperatures of 200 °C and have a glass transition temperature in dry conditions of $T_g \sim 185$ °C. Therefore, the β-sheet crystals do not fully melt (estimated melting temperature of 260 °C) and SRT is processed in its rubbery state. Aided by plasticizers such as water, the $T_g$ is decreased to room temperature or even below 0 °C, and the processing temperature range is significantly broadened (as described previously). This enables the thermoplastic processing of SRT proteins in mild conditions of hydration (5% to 45% of water content), temperature ($T > 40$ °C) and pressure (1 MPa). These conditions facilitate easy processing, extend the manufacturing possibilities and working conditions, and make SRT proteins a promising source for developing protein-based functional materials. Conventional thermoplastic processing methods commonly used in polymer production systems have been successfully tested with SRT proteins (Figure 7-2). Paste extrusion, injection molding, hot pressing, and fiber drawing were easily adapted to SRT by addition of plasticizers, and a diversity of SRT-based micro/macro-materials are successfully fabricated by these versatile methods (fibers, films, 3D geometries, etc.). Likewise, the plasticizing of SRT allows for 3D protein printing, exploring new manufacturing possibilities for bioprinting and tissue engineering. In addition, the thermoplastic properties of SRT can also be exploited for the fabrication of nanomaterials such as thin films, nanotubes, nanoparticles, and complex nanoscale geometries and patterns for multiple applications in biotechnology (drug delivery, biosensing, surface wettability, etc.).

In addition to solution-based and thermal-based processing, SRT proteins can be exposed to external stimuli (pressure, temperature, and solvents) to control the semicrystalline morphology in several post-processing steps (as detailed previously in Chapter 4). This allows for further design and fabrication possibilities (since the material properties arise from the protein nanostructure), including materials with gradient properties.

The versatility and easy processing of SRT proteins allow for materials fabrication without loss of mechanical properties (the nanostructure can be recovered if desired). This represents a major advantage over other structural proteins, and opens up new research directions for the use of SRT in multiple applications (sensing, adhesives, drug delivery, etc.). In particular, the use of SRT in
self-healing coatings, bioadhesives, photonic devices and nanocomposites were explored, and are described next.

7.2 SRT pressure sensitive underwater adhesives

Pressure-sensitive adhesives (PSA) can adhere to a variety of surfaces with pressure contact, and can be classified according to their physical state (e.g., aqueous, solvent and hot melt). The chemistry of the adhesive formulation is based on the polymer type, molecular weight, level of cross-linking, and additives (e.g., tackifiers, fillers, stabilizers, and plasticizers). Commercial PSA’s are generally obtained from petroleum based chemicals, such as acrylics, acetates, nitriles, and special (e.g., styrene based) block copolymers. However, synthetic PSAs present cytotoxicity problems in medical uses, and for this reason natural adhesives have recently gained attention. Although natural materials are used as adhesives (e.g., natural rubber, soy bean protein), the number of reports related to bio-based PSA’s is limited.

Other natural elastomers made from protein extracts have received significant interest as eco-friendly functional materials for underwater adhesion. Natural adhesives, such as mussel, sandcastle worm glue and gecko feet, provide adhesion in wet or dry conditions by different adhesion mechanisms. Mussel adhesion is based on a mixture of dihydroxyphenylalanine (DOPA)-containing proteins, which are able to provide various chemical-based surface interactions under wet conditions, including metal coordination, hydrogen bonding, and hydrophobic interactions. Gecko feet rely on Van der Waals interactions by exploiting the principles of friction and wetting at nanoscale for dry and wet adhesion respectively.
In this section, the pressure-sensitive adhesion of SRT proteins (native *Loligo vulgaris* SRT protein complex) is studied.\(^{211}\) SRT proteins are not adhesive in the glassy state, but are hard and stiff instead. If one considers the natural function of SRT in squid species in their natural habitats (deep, cold water), high modulus and hardness are required for effective predatory materials (it is used as teeth but not as adhesive in nature). However, SRT proteins are adhesive above the glass transition (Figure 7-3), which is not exploited in nature. As previously discussed in this work, SRT can be plasticized with water reducing the glass transition temperature to \(T_g \approx 30 \, ^\circ\text{C}\). This enables the use of SRT as an adhesive in wet environments (including completely immersed underwater) in mild conditions close to body temperature. Hence, SRT does not require any drying process or toxic solvents, and have good adhesion strength to a range of substrates at low temperatures, which is interesting for biomedical applications (as sealant and surgical adhesive). Furthermore, the adhesion remains after bringing the protein back to the glassy state after adhesion (drying after processing underwater). Pictures of single-lap joints bonded with SRT protein are shown in Figure 7-3a and Figure 7-3b for wet and dry testing condition respectively. Analysis of the unbounded surfaces show that the failure is adhesive (i.e., failure at the substrate-protein interface) in underwater conditions (Figure 7-3c), but cohesive (bulk protein failure) in dry conditions (Figure
7-3d).

![Figure 7-4. a) Normal tensile adhesion strength of SRT protein complex at room temperature as a function of the preload at 45 °C and 70 °C processing temperature. b) Single-lap shear adhesion of SRT in wet and dry conditions.](image)

SRT exhibits an adhesive behavior similar to that of a hot melt PSA when compressed above the glass transition temperature and cooled down. The adhesion strength increases abruptly (i.e., two orders of magnitude) above a critical preload pressure (Figure 7-4a). The temperature for secure bonding at 45 °C or 70 °C provide similar adhesion strength but higher preload is required for the former. However, a constant maximum adhesion is achieved above preloads of 1 MPa. The adhesive is stable underwater at least for six months. The normal tensile adhesion strength of the protein film is approximately 1.50 ± 0.23 MPa for underwater and 0.24 ± 0.08 MPa for dry conditions. The adhesion strength of SRT protein complex was also quantitatively measured by shear tests according to the ASTM standard 3163. Single-lap underwater shear adhesion strength (i.e., 2.51 ± 0.55 MPa for underwater and 0.17 ± 0.05 MPa for dry conditions) is larger than that in tension (Figure 7-4b), which is common for soft materials.\textsuperscript{511}

In contrast to other biological adhesives, the adhesion mechanism of SRT cannot possibly arise from covalent bonding with the substrate due to the lack of any chemical cross-linker (including cysteine residues) or any post-translational modification. The mechanism of adhesion could be due to several effects combining surface chemistry, surface roughness, residual stresses at the interface,
and/or mechanical instability. The adhesion phenomenon could be divided into two categories (*i.e.*, above and below glass transition temperature, $T_g$). First, for the case below $T_g$ (glassy state), the protein film is not adhesive because it is very stiff (*i.e.*, modulus of $\sim$1 GPa),$^{83}$ and the deformation is limited under compressive loading. Second, for the case above $T_g$ (rubbery state), the material is soft (*i.e.*, $\sim$ 1 MPa) and deforms easily, which increases the total area of contact with the surface. Therefore, the adhesion increases via nonspecific Van der Waals and hydrogen bonding interactions. When cooled below $T_g$, the material modulus increases by three orders of magnitude (*i.e.*, $\sim$ 1 GPa), and the adhesion strength increases drastically because of lowered roughness and shielded charge in water.$^{512,513}$

![Figure 7-5. Underwater adhesion strength for various surfaces: glass, hydrophilic polymeric surfaces (polylactic acid-PLA and polyvinyl acetate-PVA), and plasma treated hydrophobic polymeric surfaces (polystyrene-PS, polydimethyl siloxane-PDMS) at room temperature.](image)

SRT protein films adhere only to hydrophilic surfaces, but don’t adhere to hydrophobic surfaces underwater. Therefore, the surface of these substrates were oxidized with corona plasma surface treatment. Figure 7-5 shows the adhesion data for various surfaces including glass, hydrophilic polymeric (polylactide, PLA, and polyvinyl acetate, PVA), and plasma-treated (*i.e.*, to create hydroxyl sites on the polymer surface for improved adhesion) hydrophobic surfaces:
polymethylmethacrylate (PMMA), polystyrene (PS), and polydimethylsiloxane (PDMS). This strong affinity for polar substrates suggests that SRT adhesion is possibly caused by hydrogen bonding (very likely from the hydroxyl groups in tyrosine residues, which are concentrated in the plasticized amorphous mobile chains). In order to support this concept, the effect of salt additions (Figure 7-6a) and pH (Figure 7-6b) on the adhesion strength was tested. The presence of ions in the solution during the adhesion tests substantially reduces the adhesion strength (ions disrupt the hydrogen bonding interactions between protein and substrate). pH between 5 and 10 has no apparent effect on the adhesion strength (mild conditions). However, strong acid and basic conditions decrease the adhesion strength, and ultimately dissolve the protein.

![Figure 7-6](image.png)

Figure 7-6. Normal tensile adhesion strength as a function of a) ion concentration, and b) pH.

Although the detailed chemistry of SRT adhesion is not fully understood yet, these findings set the groundwork for developing SRT-based PSA adhesives with tailored properties (controlled by amino acid sequence and morphology). SRT proteins exhibit an underwater adhesion stronger than those of other biological and bioinspired adhesives (Table 7-1), and could be used as adhesive in an array of fields in the future, including dental resins, bandages for wound healing, and surgical sutures in the body, all of which require wet adhesion.
Table 7-1. Dry and underwater adhesion strength of SRT and other biological and synthetic adhesives

<table>
<thead>
<tr>
<th></th>
<th>Dry adhesion (MPa)</th>
<th>Underwater adhesion (MPa)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Biological adhesives</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRT protein</td>
<td>0.24 ± 0.08&amp;, 0.17 ± 0.05#</td>
<td>1.50 ± 0.23&amp;, 2.51 ± 0.55#</td>
</tr>
<tr>
<td>Mussel adhesive plaques</td>
<td>~0#</td>
<td>0.09-0.28 ± 0.08#</td>
</tr>
<tr>
<td>Gecko footpad(510)</td>
<td>0.08 ± 0.02#</td>
<td>0.02 ± 0.01#</td>
</tr>
<tr>
<td><strong>Bioinspired adhesives</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mussel (DOPA-based)</td>
<td>0.12 ± 0.01&amp;</td>
<td>0.086 ± 0.01&amp;</td>
</tr>
<tr>
<td>mimic(514)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sandcastle worm glue</td>
<td>~0#</td>
<td>0.10 ± 0.1#</td>
</tr>
<tr>
<td>mimic(515)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gecko mimic(516)</td>
<td>0.03 ± 0.1&amp;</td>
<td>~0&amp;</td>
</tr>
<tr>
<td><strong>Synthetic adhesives</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poly(vinyl acetate)(517)</td>
<td>4 ± 1#</td>
<td>~0#</td>
</tr>
<tr>
<td>Ethyl cyanoacrylate(517)</td>
<td>7 ± 1#</td>
<td>~0#</td>
</tr>
<tr>
<td>Epoxy(517)</td>
<td>11 ± 2#</td>
<td>0.07 ± 0.03#</td>
</tr>
</tbody>
</table>

\& tensile and \# shear adhesion tests

7.3 SRT self-healing materials

Self-healing materials are able to partially or completely heal damage inflicted on them, in particular by repairing cracks.\(^{518-520}\) Multiple design strategies have been explored in the development of self-healing materials to date (mostly polymer materials), including the encapsulation of healing agents (curing agents and catalysts),\(^{521-523}\) stimuli-activated polymer chain diffusion,\(^{524-526}\) dynamic and reversible cross-linking,\(^{527-529}\) metal complexation,\(^{530,531}\) ionic interactions,\(^{532}\) \(\pi-\pi\) interactions,\(^{533}\) multiphase separation,\(^{534-536}\) and hydrogen bonding interactions.\(^{537-540}\) Biocompatible materials with tunable properties that self-heal in wet environment would greatly benefit biomedical applications, in particular by extending the lifetime of implants, but most self-healing polymeric chemistries are not suitable for aqueous environments and entail health hazards. Underwater self-healing materials can be found in nature,\(^{131,507,541,542}\)
and significant research efforts have focused on mimicking the interfacial chemistry of marine barnacles and mussels for biomedical applications.  

Figure 7-7. Self-healing of SRT specimen.

In this section, the self-healing properties of SRT proteins are explored. SRT proteins exhibit self-healing above the glass transition temperature (Figure 7-7). In most self-healing materials, stiffness comes at the cost of dynamic healing: weak intermolecular interactions result in a less stiff material, but more dynamic healing, and vice versa. This tradeoff leads to a general self-healing strategy based on a multiphase material composed of a self-healing phase and a hard reinforcing phase.  

As previously described in this work, SRT proteins have a segmented amino acid sequence with soft and hard segments. Soft GLY-rich segments (disordered, amorphous) are strongly plasticized by water, lowering the glass transition temperature close to body temperature and softening the material. On the other hand, hard AVSTH-rich segments self-assemble into nanoconfined crystalline β-sheet domains to reinforce the material (analogous to physical cross-links). In the glassy state, the protein is stiff due to a dense hydrogen bonding network in the amorphous phase and the deformation is very limited (∼ 1 GPa modulus). Therefore, no autohesion is observed in the glassy state. However, the amorphous chains can move past each other in the rubbery state, softening the material (∼ 1 MPa modulus) and activating the self-healing behavior. The chains diffuse within the protein matrix in the plasticized rubbery state when pressure is applied (similar to the previously discussed pressure-sensitive adhesion), repairing the damaged area.
Figure 7-8. Mechanical testing of healed SRT-TR films. a) Stress-strain curves of healed specimens exhibit higher stress and fail at lower strain. b) Healed specimens break at the pristine region, while the healed area withstands the stress.

When SRT healed materials are cooled/dried to the glassy state after the healing process, no change is observed in the mechanical properties (hydrogen bonding network is dominant in the amorphous phase, resulting in a stiff and brittle material). Furthermore, the mechanical properties in the glassy (i.e., dry) state are stable after an undefined number of healing cycles. However, the rubbery mechanics of healed materials are different due to local changes in the morphology (Figure 7-8). The ultimate stress and extensibility of healed SRT-TR proteins are reduced to half approximately, which is usual for repaired elastomers (Figure 7-8a). However, the modulus and overall stress upon stretching are increased after healing. Furthermore, mechanical failure upon stretching occurs in the pristine area rather than the healed area (Figure 7-8b). This is surprising since it indicates that the healed area must be stronger than the original material, possibly due to local changes in the micro/nanostructure.
Figure 7-9. Structural changes in the healing process. a) Amide I FTIR spectra and b) XRD spectra reveal a local increase in β-sheet content caused by the healing process.

Structural analysis of the healed samples revealed an increased crystallinity in the healed area. This is observed in FTIR (Figure 7-9a) as a shift in the amide I band towards ~1625 cm⁻¹, and in XRD (Figure 7-9b) as a narrowing of the diffraction peaks in comparison to the pristine material. This indicates a localized increase in the β-sheet crystalline structure content. This observation agrees with the semicrystalline analyses of SRT upon external stimuli previously discussed in this work (Chapter 4). Temperature, pressure, and water, all of which are required in the self-healing process (causing the plasticization and diffusion of the protein chains), induced an increase in β-sheet content. These localized changes in the protein morphology explain the hardening of the healed areas and the mechanical failure in the pristine areas. The healed material becomes a composite material with stiff healed regions and soft stretchable regions that deform non-homogeneously (resulting in a non-homogeneous stress distribution). Interestingly, the proton conductivity of SRT healed materials do not drop after healing but slightly increase (Figure 7-10). This can be explained again by the increase in crystallinity, which was reported to aid in proton transport in Chapter 6.
Figure 7-10. Proton conductivity of healed SRT-TR films. Conductivity is slightly higher after healing due to the local increase in crystallinity (pH 7).

SRT proteins can be coated into a variety of materials such as fibers and fabrics to create self-healing advanced textiles. As proof of concept, knitted wool and cotton fabrics were coated with SRT proteins and successful cut repair, single fiber repair, and patching were demonstrated (Figure 7-11). Furthermore, SRT coatings can incorporate biomolecules (such as enzymes) for more complex protective functionalities for advanced functional fabrics that are mechanically stable and enzymatically active after cycles of healing. Such complex films may enable a coating for multifunctional applications in defense programs such as Second Skin, where advanced fabrics protect wearers from chemical and biological warfare agents by restricting the pores of the clothing to block out harmful agents. A mechanically stable, self-healing film containing enzymes such as organophosphorous hydrolase, which neutralize chemical agents, would prove critical to such an application to actively neutralize these threats as well as ensure no micro or macro scale defects permit leaks.
In summary, SRT proteins hold great promise to provide a broad range of solutions for textile applications due to their self-healing properties. Furthermore, utilization of fiber coating technologies to prepare such composite fabrics offers the advantages of low cost and versatile processing compatible with modern advanced textile-manufacturing. The programmable properties of SRT offer an additional customization advantage in the design of smart multifunctional textiles. Self-healing high strength structural proteins may provide novel mechanical properties for clinical applications such as orthopedic devices for repair, and treatment strategies as novel biodegradable matrices to deliver and stimulate bioactive molecules for wound healing.
7.4 SRT-based photonic devices

Protein-based devices provide an alternative to synthetic polymers and glasses granting comparable or better performances while presenting diverse functionalities and clean processing capabilities. Recent bio-inspired photonic devices made from proteinaceous materials lay the groundwork for many functional device applications such as electroluminescence in peptide nucleic acids, multiphoton absorption in amyloid fibers, sensing with protein single fibers, stimulus responsive protein whispering-gallery-mode (WGM) lasers, WGM emitting protein microcavity assemblies along with silk waveguides, optical fibers, and inverse opals. The control and manipulation of light flow and light–matter interactions using whispering-gallery-mode (WGM) resonators has created significant interest in various fields of science, including but not limited to biosensing and detection, cavity-QED, optomechanics, and parity-time symmetric photonics. WGM resonators are currently manufactured using standard lithography techniques with conventional materials such as silica, silicon, and silicon nitride. However, recent developments in optical technologies have revealed the strong need for developing soft, biocompatible, and biodegradable photonic devices and photonic structures with novel functionalities that cannot be attained with current optical materials. Towards this aim, all-polymer and polymer-coated silica WGM resonators, as well as silica WGM resonators encapsulated in low-index polymers have been fabricated using conventional and commercially available polymers, such as polydimethylsiloxane (PDMS) and polystyrene (PS) to address the need for flexible structures.
Figure 7-12. SRT whispering gallery mode (WGM) resonators. a) flexible array of SRT resonators.
b) Transmission spectrum of a SRT WGM resonance around 672 nm with quality factor $Q_0 = 4.45 \times 10^5$. Inset shows SRT toroidal resonator.

Structural proteins provide an excellent platform for the development of functional, flexible, and biodegradable photonic devices with programmable structure and properties. In this section, WGM resonators were fabricated from SRT proteins and their use as add-drop filters and optical switches was explored (in collaboration with Dr. Huzeyfe Yilmaz and Dr. Sahin Özdemir, who performed the optical measurements). Flexible chips containing arrays of WGM resonators were fabricated from SRT protein solution, yielding toroidal resonators of 54 μm in diameter (Figure 7-12a). Although, proteinaceous WGM resonators have been demonstrated before, the SRT resonators are superior in performance and perform significantly better than conventional materials in specific applications (quality factors $Q_0$ ranging from $10^4$ to $10^5$, Figure 7-12b).
Figure 7-13. SRT photonic devices. a-b) Add-drop filter. SRT resonator with two coupling tapered-fiber waveguides. Transmission spectrum shows a resonant dip in the transmission port and a resonance peak in the drop port. c) Light transmission through a SRT microfiber as waveguide. d-e) SRT optical switch. Control OFF: control is off-resonance and signal light is coupled to the resonator. Control ON: control is brought to on-resonance, which increases the temperature of the resonator and the signal is shifted to off-resonant.

Flexible protein-based photonic devices (add-drop filters, waveguides, and optical switches) were fabricated from SRT proteins (Figure 7-13). Add-drop filters are frequently integrated in optical communication architectures (such as optical filters, multiplexers and routers), as well as used as...
schemes for high performance optical sensing devices. SRT-based add-drop filters were fabricated by coupling SRT WGM resonators to two separate fiber-taper waveguides. The add-drop filter (Figure 7-13a) routes the light from the input waveguide to the drop port in the second waveguide when the wavelength of the light coincides with the resonance wavelength of the SRT resonator. Non-resonant light passed through the input waveguide to the transmission port. The add-drop filter performs with an efficiency of approximately 51 %, as shown by the dip in the transmission port and the peak in the drop port (Figure 7-13b). This efficiency shows the successful performance of a proof-of-concept device, but it could be optimized by decreasing scattering losses and improving the waveguide-resonator coupling. Wholly SRT waveguides were fabricated by electro-spinnning and engaged to a silica fiber taper via van der Waals attractive forces (Figure 7-13c). Evanescent coupling of light from the silica fiber taper to the protein fiber was achieved, demonstrating the successful performance of SRT waveguides. “On/off” optical switches using SRT-based WGM structures was fabricated (Figure 7-13d). The transmission of a signal field through a coupled waveguide was switched between “on” and “off” states (unit and zero transmission respectively) by a control field via the thermal response of SRT protein. “On”: when the control field is on-resonance, the field built up inside the cavity is absorbed by the protein material and transformed into heat, which in turn shifts the resonance modes through the strong thermo-optic and thermal expansion responses of the SRT material. As a result, the signal field becomes off-resonant and its transmission increases to a value close to unity. “Off”: when the control field is adjusted to be off-resonant, the cavity cools down and the resonances return back to their initial positions. As a result, signal field becomes on-resonant again and its transmission drops to near-zero value (Figure 7-13e).

In summary, SRT protein WGM microcavities provide a platform to manipulate photons within microscale volumes of structural proteins at high power efficiency. Waveguides, add-drop filters, and optical switches were fabricated and demonstrate a wide range of potential applications for SRT-based photonic devices. SRT resonators perform switching with a pump power three orders of magnitude less than that is required by silica resonators. This striking performance is both due to the strong thermo-optic coefficients and negative coefficient of thermal expansion of SRT proteins (described in detail elsewhere). These results show that protein-based photonic
devices can be used for low power consumption applications. Moreover, the crystallinity and morphology of SRT proteins can be engineered at the molecular level, which will provide tuning of the thermo-optic response of the device. This approach will expedite the design, fabrication and synthesis of eco-friendly, recyclable, flexible optical materials and devices.
Chapter 8. Conclusions

This thesis dissertation analyzes the structure-property relationship of squid sucker ring teeth (SRT) proteins for the first time. Native SRT proteins and SRT-inspired synthetic repetitive proteins have been examined by spectroscopic structural analysis and advanced materials characterization, to elucidate the protein morphology and mechanical, thermal, and conducting properties. The initial hypothesis that states, increasing the number of amino acid building blocks in the protein chain leads to less defective networks and improved material properties, has been validated. In order to reach this conclusion, the following milestones have been achieved:

The amino acid composition and sequence of native SRT from *Loligo vulgaris* squids were unraveled using Next Generation Sequencing (NGS), revealing repetitive motifs in the sequence with alternating flexible disordered segments (rich in glycine) and crystal-forming segments (rich in alanine and histidine). This alternating repetitive sequence drives the self-assembly into a semicrystalline protein morphology, where the crystal-forming segments form nanoscale β-sheet structures are connected by disordered glycine-rich segments. The β-sheet domains act as physical cross-links, stabilizing the protein network and providing remarkable thermal (stability up to 200 °C) and mechanical properties (~ 1 GPa in modulus).

The morphology of a synthetic SRT protein with a defined amino acid primary sequence and molecular weight, was tailored to match that of its native equivalent. The synthetic protein was processed from powder to a variety of nano-, micro-, and macroscale materials through a series of solution-based and thermal-based methods, yielding SRT materials with tailored β-sheet content and structure. The tailored semicrystalline morphology strongly affected the material properties, causing an increase in modulus, glass transition temperature, and proton conductivity with increasing crystalline fraction due to the increased ordering of the protein chains.

Next, SRT-inspired synthetic tandem repeat polypeptides with a controlled and repetitive segmented sequence were analyzed. SRT-TR polypeptides share the same disordered/crystalline building blocks but vary in length (*i.e.*, varying total number of building blocks per protein chain), and assemble into networks composed of β-sheet physical cross-linkers connected to each other via disordered tie chains. However, disordered loop structures originate from intramolecular...
interactions and introduce defects in the network structure, resulting in inferior properties. A simple network model based on entropic elasticity theory has been proposed, describing the protein network structure (density of effective vs. defective strands) and mechanical properties as function of the number of repeat units in the polypeptide chain given a set of confined β-sheet crystallite dimensions. According to the model, the network defects increase linearly with the reciprocal number of repeat units (i.e., high number of repeat units lead to effective networks), which has been experimentally validated with measurements of shear modulus, tensile modulus, and toughness.

Furthermore, the effects of tandem repetition and network morphology on the chain dynamics and transport properties (proton and thermal conductivity) of SRT-TR polypeptides were analyzed. Tandem repetition enhances both proton and thermal conductivity through control of the protein network morphology and defect density. Low molecular weight repetitive proteins are rich in defective disordered structures (loops and dangling ends), which introduce network defects that inhibit the transport phenomena. As the molecular weight and number of tandem repetitions increase, the defective disordered structures are replaced by effective network strands (tie chains) that covalently connect the β-sheet crystallites and enhance the conductivity due to decreased phonon scattering and increased concentration of mobile hydrogen.

Therefore, the control of the protein nanostructure through tandem repetition and β-sheet crystallization allows for the tuning of SRT materials at the molecular level. This introduces new design rules in the engineering of synthetic protein materials with optimized programmable properties, and provides insight into the evolutionary dynamics of gene duplication in naturally occurring repetitive protein materials. Such level of control has led to tailored SRT materials that are stretchable up to 400% of their original length, exhibit a peak proton conductivity of 3.66 mS cm⁻¹ (highest than any biological material reported to date), and a peak thermal conductivity of 1.43 W m⁻¹ K⁻¹ (which exceeds that of most polymer materials). The high proton and thermal conductivity of SRT-based materials, together with their programmable mechanical properties (flexible and stretchable), make them an attractive material for bioelectronics and biosensing applications.

Last, a variety of applications for SRT-based materials have been explored, including SRT-based adhesives (with an underwater adhesion strength at least ten times stronger than other biological
adhesives), SRT-based self-healing materials and coatings, and SRT-based photonic devices (optical waveguides, filters, and switches based on whispering gallery mode resonators). These tunable materials and devices showcase the great potential of SRT proteins, which will have a strong impact in the biomedical field.

8.1 Recommendations for future work

As a follow-up to the work presented in this thesis dissertation, the following research directions are proposed:

8.1.1 Self-assembly of SRT-inspired proteins and nanoconfinement of β-sheets

This dissertation studies the effect of tandem repetition on the nanostructure and properties of proteins with a primary amino acid sequence inspired in SRT. However, variations on the SRT sequence (including amino acid composition, location along the chain, and length of the amorphous and crystal-forming segments) have not been explored in this work. Although the role of individual amino acids (such as histidine and proline) have been discussed, a clear and complete map of amino acid interactions within SRT-like sequences is still missing. Variations in the amino acid sequence and segment length will drive the assembly into β-sheet domains of different dimensions that will define the structure and bulk properties of the material.

The formation of β-sheet crystals can be roughly described by the thermodynamics of self-assembly: \[ \Delta G \approx \Delta H_{\text{HB}} + \Delta H_{\text{vdW}} - T\Delta S_c. \] The free energy change (\( \Delta G \)) upon incorporation of a crystal-forming strand (AVSTH-rich segment) in a β-sheet crystal can be divided into three terms: (i) hydrogen bonding within the β-sheet (\( \Delta H_{\text{HB}} \)), (ii) van der Waals interactions in the β-sheet (\( \Delta H_{\text{vdW}} \)), and (iii) the free energy of the flexible protein chain (\( T\Delta S_c \)). The hydrogen bonding (strands attach parallel to the backbone) and van der Waals terms (strands stack on top of the β-sheet) are enthalpic, and drive the self-assembly process. The third term represents the entropic loss due to the folding of the main chain, and opposes the self-assembly process (flexible chains will minimize the energy in a disordered random coil conformation). According to this simplified model, the β-strands aggregate forming β-sheets until the opposing entropic term overtakes the
self-assembly driving forces. At that point of thermodynamic minimum free energy of self-assembly, the β-sheet crystal will stop growing. Therefore, calculating these terms for an extensive library of SRT-like sequences is necessary to accurately predict the self-assembly mechanism and resulting nanostructure. The effect of segment length, sequence topology, charged and voluminous amino acid side chains, and nucleation should be additionally considered in a comprehensive model for β-sheet assembly. Although this is a simplified model, it serves as base for understanding the limited growth of β-sheet crystals in SRT proteins and other fibrous proteins such as amyloid fibrils and silk fibroin. Developing and validating such models that can predict protein assembly is a challenging task due to the complexity of natural and synthetic proteins, but hold the key to the understanding of the self-assembly process and optimization of protein structure/properties. Both computational studies (molecular dynamics simulations on the stability of protein structures with selected amino acid permutations) and the complementary synthesis and characterization of protein libraries (random amino acid content within the crystal-forming segments) will be the follow-up step towards a comprehensive understanding of the self-assembly of SRT-like protein materials.

### 8.1.2 High-throughput screening of protein libraries

Despite the recent advances in proteomics and materials science, structural characterization and prediction of repetitive protein systems from their primary amino acid sequence remains a challenge. Most techniques used to predict protein structure rely on databases that are rarely applicable to structural-protein assemblies. Experimental characterization for this class of proteins often employs spectroscopic techniques such as Fourier transform infrared spectroscopy (FTIR), X-ray crystallography, Raman spectroscopy, or circular dichroism. However, these well extended structural characterization methods have limitations in high-throughput screening of protein libraries due to the large number of sequence permutations in protein libraries.

Recently, our group has shown that it is also possible to use ultrafast laser-probing spectroscopy to characterize protein structure–property relationships. Transient thermal reflectivity (TTR) experiments using purified SRT and silk proteins as well as overexpressed SRT recombinant protein in *E. coli* showed “proof of concept” quantification of protein crystallinity in real time for the first time. Our novel TTR technique theoretically enables screening of $10^8$-$10^9$ different
structural polypeptide sequences for protein assembly in hours, a feat that would be impossible to achieve with existing screening tools such as fluorescence, immunostaining, or functional assays. The integration of transient thermal reflectivity in protein library screening is a very promising platform for the mapping of sequence-assembly, and will help to identify unnatural structural motifs with exquisite resolution through library design.479–581

8.1.3 SRT-based devices

8.1.3.1 Biophotonic devices

The possibilities of SRT-based devices have been described in this dissertation and published elsewhere.496,574 In particular, optical waveguides, add-drop filters, and switches were fabricated solely from SRT protein and showed excellent performance (high quality factor, flexibility, and high power efficiency). However, protein-based optical devices have properties that have not been explored in this work and would provide tunable devices for a wide range of biophotonics applications. As described in this work, the semicrystalline morphology of SRT proteins can be modified by adjusting the amino acid sequence, but it is also affected by external stimuli such as temperature, pressure, humidity, and exposure to certain organic solvents. As described elsewhere, SRT-based resonators with high and low crystallinity exhibited different thermo-optic coefficients and therefore different resonance wavelength shifts with temperature.496 This structure-dependent shifts can be used to design WGM-based flexible biosensors. In addition, ultrasensitive biomolecule detection has been previously demonstrated in optical cavities,556–558,562 and a similar approach can be applied to protein-based resonators. This would require further functionalization of SRT resonators so the biomolecule of interest can bind to the resonator surface. Although further research is necessary to develop fully responsive sensors, SRT has shown biomolecule (enzymes) encapsulation capabilities through surface engineering that will prove useful in the design and functionalization of SRT-based biosensors.495 These controllable factors will provide tuning and specificity of the thermo-optic response of the SRT flexible devices.

8.1.3.2 Protonic devices
In this thesis work, the proton conductivity of SRT proteins has been investigated. However, the use of SRT in functional protonic devices has not been explored in this work. Protonic devices (proton transistors) from biological materials (such as reflectin and chitosan) have been recently reported for the use in bioelectronics.\textsuperscript{424,425,427} SRT proteins have several advantages over current biological materials used in proton transport, and would outperform current state-of-the-art bioprotonic devices. First, SRT proteins have the highest proton conductivity reported to date for a biological material as described in this thesis, which will allow for higher operating current. Second, SRT proteins are stable extreme conditions of temperature (up to 200 °C), humidity (non-soluble, therefore it can be completely immersed in aqueous solutions), and pH (stable between pH 4 and 10). Third, SRT materials are mechanically strong but flexible, allowing for the deformation of the materials and the potential integration for in-body applications. But more importantly, SRT transport properties can be controlled by adjusting the nanostructure. This can be done by exposure to external stimuli: pressure, strain, humidity, temperature, exposure to organic solvents, etc. Such tuning of the nanostructure and the resulting proton transport properties enable the use of SRT proteins in sensors and memory devices that use mechanical (pressure, strain) and chemical cues as external input, which will be of great interest for in-body sensing in biomedical applications.

\textbf{8.1.3.3 Thermal devices}

The thermal conductivity of SRT proteins has been investigated in this thesis work, outperforming most conventional amorphous polymers. Specifically, the effect of network defects in the transport properties was studied, revealing a strong conductivity dependence on the tie chain density and overall crystallite connectivity. However, additional research is necessary in order to further optimize the thermal transport properties and apply them in the design and fabrication of thermal devices. As described in this thesis, the protein network morphology can be finely adjusted by controlling the number of repeat units in the sequence (tie chain density) and exposing the protein material to external stimuli such as temperature, pressure, humidity, strain, and organic solvents. It is expected that a protein structure with high density of tie chains and crystalline $\beta$-sheet domains that are aligned in a preferential direction would facilitate the thermal transport as reported for other polymer systems.\textsuperscript{419,421,423,477} Such highly ordered morphology would enhance the anisotropic thermal transport along the alignment direction, and further increase the thermal
conductivity. Furthermore, the plasticizing effect of water in SRT and the consequent changes in chain mobility and thermal conductivity will be enhanced by a finely controlled ordering in the protein structure. A material with this characteristics would offer thermal switching capabilities with a record-high reversible switching ratio (dry/hydrated state) at room temperature. Together with the self-healing properties, biocompatibility, and flexibility, SRT thermal-devices will find thermal management applications in the biomedical field.

8.1.4 Controlled assembly of 2D materials

Due to their controllable sequence, programmable properties, and easy processability, SRT proteins have shown exceptional potential in the development of electrical, thermal, and photonic devices. 2D-layered materials establish the foundation of next-generation, programmable, flexible, optically superior, energy efficient and mechanically strong materials and devices. Moreover, exploitation of unique material properties at the nanoscale opens up new doors for fundamental research. Recent advances in the nanotechnology of 2D materials combined with parallel improvements in biotechnology and synthetic biology have demonstrated that more complex composite materials with properties engineered precisely to optimize performance can be achieved. SRT proteins are an excellent platform to control the alignment of 2D materials in protein-based composites. Our group reported the synthesis and fabrication of 2D molecular composites, comprising layers of atomically thin crystalline inorganic materials and layers of semi-crystalline self-assembling proteins. The amino-acid sequence of the proteins, which dictates the degree of crystallinity and alignment of the protein layers, can also be used to control the interactions at the 2D material/protein interface, ultimately dictating the functional physical properties (e.g., electrical resistivity and thermal conductivity) of the devices. SRT synthetic polypeptides have the ability to use protein interfaces in contact with 2D materials to control interfacial chemistry, electrical contact resistance, and thermal boundary resistance, which are nanoscale characteristics that are important to the operation of flexible 2D devices made from these materials. Our preliminary work has shown the successful assembly of SRT-based nanocomposites from 2D materials including transition metal carbides (titanium carbide MXene) and graphene derivatives (graphene oxide). The layered 2D composite systems, in which interlayer distances can be precisely and finely tuned by the molecular weight of the protein, offer a very interesting platform to study how optical, electrical, thermal and mechanical properties can
be controlled in a molecular composite. Furthermore, the addition of SRT to the protein-based composites provides responsiveness to external stimuli (humidity, pH, temperature). Successful development of programmable 2D composites will have a significant impact on multiple applications in various fields (e.g., synthetic biology, autonomy, nanotechnology, and energy) and open new avenues of 2D materials research.
Appendix. Representative publications

# 14 “Structural protein-based flexible whispering gallery mode resonators”
H Yilmaz, A Pena-Francesch, R Shreiner, H Jung, Z Belay, MC Demirel, S Ozdemir, L Yang
*ACS Photonics* DOI: 10.1021/acsphotonics.7b00310 (2017)

#13 “Programmable molecular composites of tandem proteins with graphene oxide for efficient bimorph actuators”
M Vural, Y Lei, A Pena-Francesch, H Jung, B Allen, M Terrones, MC Demirel
*Carbon*, Volume 118, July 2017, Pages 404–412 (cover page)

#12 “Ultrafast laser-probing spectroscopy for studying molecular structure of protein aggregates”
*Analyst*, 2017,142, 1434-1441 (cover page)

#11 “Molecular tandem repeat strategy for elucidating mechanical properties of high-strength proteins”
*Proceedings of the National Academy of Sciences* 113 (23), 6478–6483 (2016)

#10 “Self-healing textile: enzyme encapsulated layer-by-layer structural proteins”
D Gaddes, H Jung, A Pena-Francesch, G Dion, S Tadigadapa, WJ Dressick, MC Demirel
*ACS Applied Materials & Interfaces* 8 (31), 20371-20378

#9 “Protein-based flexible whispering gallery mode resonators”
H Yilmaz*, A Pena-Francesch*, L Xu, R Shreiner, H Jung, SH Huang, SK Ozdemir, MC Demirel, L Yang (*equal contribution)
*SPIE OPTO*, 97450I-97450I-6 (2016)

#8 “Segmented molecular design of self-healing proteinaceous materials”
V Sariola*, A Pena-Francesch*, H Jung, M Çetinkaya, C Pacheco, M Sitti, MC Demirel (*equal contribution)
*Scientific reports* 5, 13482 (2015)

#7 “Biomimicry of the *manduca sexta* forewing using SRT protein complex for FWMAV development”
SC Michaels, KC Moses, RJ Bachmann, R Hamilton, A Pena-Francesch, A Lanba, MC Demirel, RD Quinn
*Biomimetic and Biohybrid Systems*, 86-91 (2015)

#6 “Recent advances in nanoscale bioinspired materials”
MC Demirel, M Cetinkaya, A Pena-Francesch, H Jung

#5 “Materials fabrication from native and recombinant thermoplastic squid proteins”
*Advanced Functional Materials* 24 (47), 7401-7409 (2014) (cover page)

#4 “A Renewable Bioplastic Made From Squid Proteins”
MC Demirel, A Pena-Francesch, W Mills

#3 “Pressure sensitive adhesion of an elastomeric protein complex extracted from squid ring teeth”
A Pena-Francesch, B Akgun, A Miserez, W Zhu, H Gao, MC Demirel
*Advanced Functional Materials* 24 (39), 6227-6233 (2014)

#2 “Tailoring the LCST of Thermosensitive Hydrogel Thin Films Deposited by iCVD”
A Pena-Francesch, L Montero, S Borrós

#1 “Accelerating the design of biomimetic materials by integrating RNA-seq with proteomics and materials science”
PA Guerette, S Hoon, Y Seow, M Raida, A Masic, FT Wong, VHB Ho, KW Kong, MC Demirel, A Pena-Francesch, S Amini, GZ Tay, D Ding, A Miserez
*Nature biotechnology* 31 (10), 908-915 (2013) (cover page)
References


(74) Hayashi, C. Y.; Lewis, R. V. Evidence from Flagelliform Silk cDNA for the Structural Basis of Elasticity


31, 261–272.


(105) Spiess, K.; Lammel, A.; Scheibel, T. Recombinant Spider Silk Proteins for Applications in Biomaterials.


(128) Qiu, W.; Teng, W.; Cappello, J.; Wu, X. Wet-Spinning of Recombinant Silk-Elastin-like Protein Polymer Fibers with High Tensile Strength and High Deformability. *Biomacromolecules* 2009, 10, 602–608.


Copley, J. R. D.; Cook, J. C. The Disk Chopper Spectrometer at NIST: A New Instrument for Quasielastic


(205) Pace, C. N.; Horn, G.; Hebert, E. J.; Bechert, J.; Shaw, K.; Urbanikova, L.; Scholtz, J. M.; Sevcik, J.


(323) Flory, P. J. *Principles of Polymer Chemistry*; Cornell University Press, 1953.


(355) Fuchs, S. Genetic Instability within Regions Encoding Repetitive Proteins as a Driver of Adaptation. *Protein"
Sci. 2015, 24, 121.


(386) Wasko, S. S.; Tay, G. Z.; Schwaighofer, A.; Nowak, C.; Waite, J. H.; Miserez, A. Structural Proteins from


(445) Glerup, M.; Poulsen, F. W.; Berg, R. W. Vibrational Spectroscopy on Protons and Deuterons in Proton


(476) Chen, G. *Nanoscale Energy Transport and Conversion: A Parallel Treatment of Electrons, Molecules,*


2010, 22, 5424–5430.


215


Abdon was born in the beautiful city of Tarragona, Spain, on May 2\textsuperscript{nd} 1988. He graduated in Mechanical Engineering from Institut Químic de Sarriá (Barcelona) in 2011. He obtained a M. Sc. in Chemical Engineering from Institut Químic de Sarriá in the Materials Engineering Group under the guidance of Prof. Salvador Borros in 2013, where he developed tailored thermo-responsive thin film hydrogels by initiated chemical vapor deposition (iCDV) for anti-biofouling applications. He later pursued his graduate studies at The Pennsylvania State University in the Engineering Science and Mechanics department, where he joined Prof. Melik C. Demirel’s research group. During his graduate studies, Abdon investigated advanced protein materials inspired in squid proteins, and extensively worked on the sequence-structure-property of repetitive structural protein biopolymers, the self-assembly, and fabrication/processing of protein-based materials. His research interests are biomimetics, soft functional materials, and nanotechnology.

For more details visit: https://sites.psu.edu/abdonpf/