A Multimodal Lens on Protein Assembly

From Silk to Seeds

Francis, Juanita

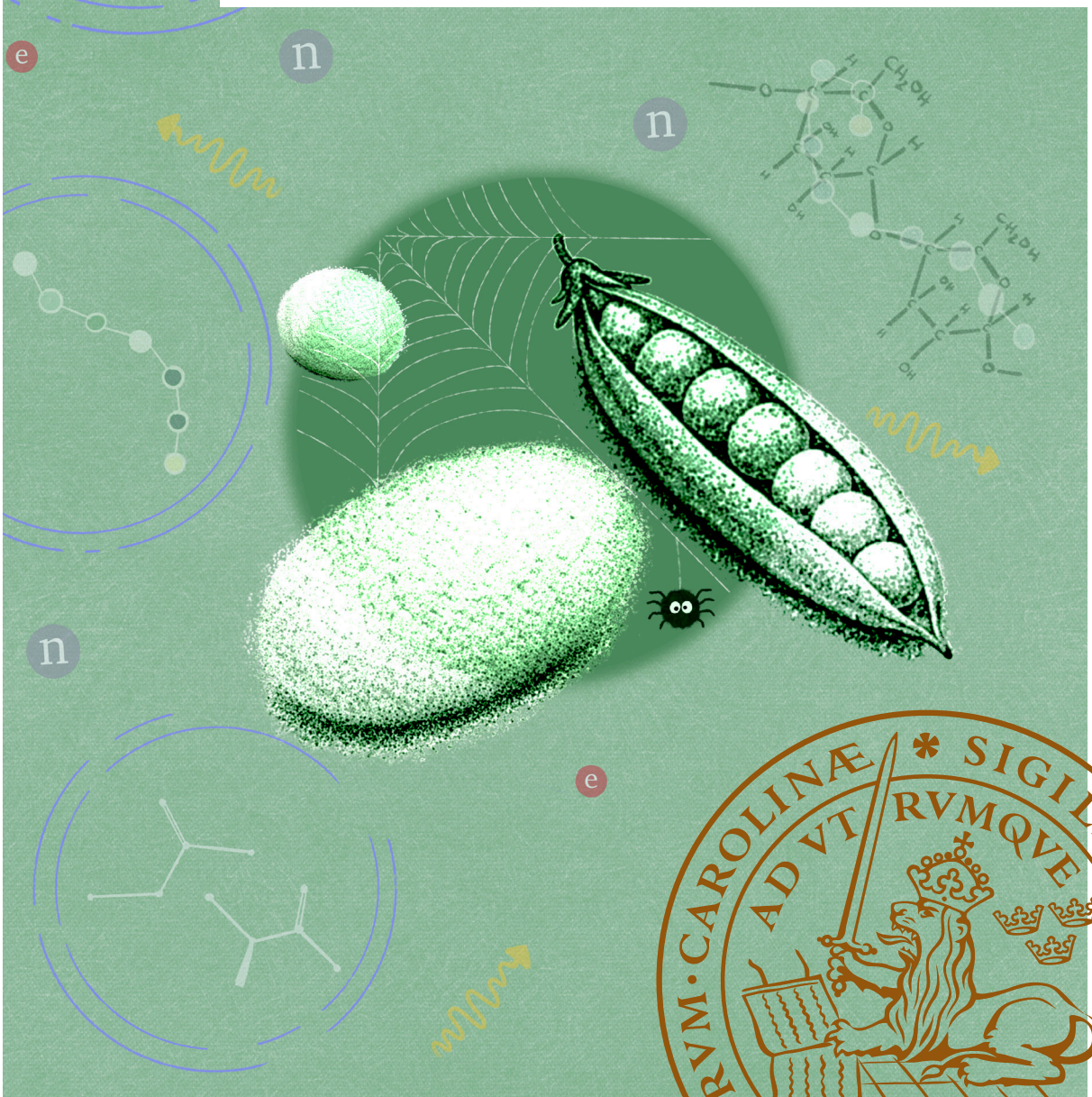
2026

Document Version:
Publisher's PDF, also known as Version of record

A Multimodal Lens on Protein Assembly

From Silk to Seeds

JUANITA FRANCIS
DEPARTMENT OF CHEMISTRY | FACULTY OF ENGINEERING | LUND UNIVERSITY



Nature is an excellent teacher when we take the time to look carefully. I hope this work reflects that and adds something useful to our current understanding.

— Juan

Printed by Media-Tryck, Lund 2026 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

Division of Pure and Applied Biochemistry
Department of Chemistry

ISBN 978-91-8104-874-2

A Multimodal Lens on Protein Assembly

From Silk to Seeds

A Multimodal Lens on Protein Assembly

From Silk to Seeds

Juanita Francis

LUND
UNIVERSITY

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Faculty of Engineering at Lund University to be publicly defended on Friday the 27^{th} of March at 9.00 in lecture hall B, Kemicentrum, Sölvegatan 39, Lund

Faculty opponent
Professor Alberto Saiani

Division of Pharmacy & Optometry
The University of Manchester, United Kingdom

Organisation: LUND UNIVERSITY

Division of Pure and Applied Biochemistry

Department of Chemistry

Document name: DOCTORAL DISSERTATION

Date of issue: 2026-03-27

Author(s): Juanita Francis

Sponsoring organisation: The Swedish Foundation for Strategic Research (SSF) SwedNESS, and Agenda2030 graduate school

Title and subtitle: A Multimodal Lens on Protein Assembly – From Silk to Seeds

Abstract:

Protein-based materials rely on controlled assembly across multiple length scales. Key organisational features often emerge before macroscopic fibres or gels form. Yet the pathways linking molecular organisation to material formation remain poorly resolved. This is partly because structural evolution across length scales is rarely measured concurrently within the same sample. This thesis examines protein assembly as a multiscale, pathway-dependent process. Silk proteins serve as the primary model system, as they form highly organised fibres under mild aqueous conditions. Yellow pea proteins provide a complementary multicomponent case in which processing history and co-existing components influence assembly. The work focuses on the solution and pre-gelation states. In reconstituted silk fibroin from *Bombyx mori*, time-resolved multimodal measurements are used. Neutron scattering is combined with ultraviolet and fluorescence spectroscopy to follow structural evolution from molecular to mesoscopic scales within the same samples. These measurements reveal structured intermediate states prior to extensive β-sheet formation. The accessibility and persistence of these intermediates depend on the assembly environment, showing that processing conditions select assembly pathways rather than simply triggering final states. Native and recombinant cylindriform spidroins are examined as a simpler spider silk system. Small-angle X-ray scattering and ensemble modelling reveal elongated, modular solution architectures, linking solution-state organisation to its multi-domain architecture. Recombinant constructs further demonstrate the sensitivity of assembly to junction stability and solution context. In pea protein–starch mixtures, contrast-variation neutron scattering shows that proteins and residual starch are largely independent in solution. Thermal treatment induces co-assembly in a processing-history-dependent manner. Together, these studies show that protein-based material formation is shaped upstream of macroscopic assembly. Resolving organisation across length scales highlights the roles of pre-organisation, pathway-dependent intermediates, and solvent participation. The multimodal and scattering-based approaches established here provide a general strategy for investigating hierarchical protein assembly in both silk-like and food-relevant systems.

Key words: protein assembly**,** silk**,** pea protein, starch, small-angle scattering, pH-triggered gelation, Multi-domain proteins

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language English, Swedish

Number of pages: 190

ISBN (e-version): 978-91-8104-875-9

ISBN (print): 978-91-8104-874-2

Signature

Date 2026-02-12

A Multimodal Lens on Protein Assembly

From Silk to Seeds

Juanita Francis

LUND
UNIVERSITY

Cover photo (front) by Juanita Francis

Cover photo (back) taken by Ewa Krupinska

Lund University, Faculty of Engineering

Division of Pure and Applied Biochemistry, Department of Chemistry

ISBN 978-91-8104-874-2 (print)

ISBN 978-91-8104-875-9 (electronic)

Printed in Sweden by Media-Tryck, Lund University

Lund 2026

Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

MADE IN SWEDEN

Your research isn't the final word, it's a question mark — with a bibliography
- from someone who recently finished their PhD

Table of Contents

Abstract

Protein-based materials rely on controlled assembly across multiple length scales. Key organisational features often emerge before macroscopic fibres or gels form. Yet the pathways linking molecular organisation to material formation remain poorly resolved. This is partly because structural evolution across length scales is rarely measured concurrently within the same sample.

This thesis examines protein assembly as a multiscale, pathway-dependent process. Silk proteins serve as the primary model system, as they form highly organised fibres under mild aqueous conditions. Yellow pea proteins provide a complementary multicomponent case in which processing history and co-existing components influence assembly.

The work focuses on the solution and pre-gelation states. In reconstituted silk fibroin from *Bombyx mori*, time-resolved multimodal measurements are used. Neutron scattering is combined with ultraviolet and fluorescence spectroscopy to follow structural evolution from molecular to mesoscopic scales within the same samples. These measurements reveal structured intermediate states prior to extensive β-sheet formation. The accessibility and persistence of these intermediates depend on the assembly environment, showing that processing conditions select assembly pathways rather than simply triggering final states.

Native and recombinant cylindriform spidroins are examined as a simpler spider silk system. Small-angle X-ray scattering and ensemble modelling reveal elongated, modular solution architectures, linking solution-state organisation to its multi-domain architecture. Recombinant constructs further demonstrate the sensitivity of assembly to junction stability and solution context.

In pea protein–starch mixtures, contrast-variation neutron scattering shows that proteins and residual starch are largely independent in solution. Thermal treatment induces co-assembly in a processing-history-dependent manner.

Together, these studies show that protein-based material formation is shaped upstream of macroscopic assembly. Resolving organisation across length scales highlights the roles of pre-organisation, pathway-dependent intermediates, and solvent participation. The multimodal and scattering-based approaches established here provide a general strategy for investigating hierarchical protein assembly in both silk-like and food-relevant systems.

Popular science summary

Many of the materials we rely on every day, such as plastics, fibres, coatings, and food gels, are made using energy-intensive processes that rely on high temperatures, strong chemicals, or extensive mechanical processing. In contrast, nature routinely produces high-performance materials under mild conditions. Silk is a striking example. Spiders and silkworms spin fibres that are lightweight, strong, and durable using water-based protein solutions at ambient temperature. Understanding how these proteins organise themselves before fibres and gels are formed is a key step toward developing more sustainable technologies.

This thesis explores how proteins gradually organise into larger structures, from individual molecules to extended networks. Silk proteins serve as a central model system, while pea proteins provide a complementary multi-component example relevant to sustainable food production. Although these systems differ in purpose and biological origin, both rely on controlled protein organisation to achieve their final properties.

A central focus of the work is what happens to proteins while they are still in a solution state. Rather than forming solid materials in a single step, silk proteins pass through transient arrangements that help maintain stability and guide assembly. By combining several experimental approaches that probe structure at different length scales, this thesis examines how silk proteins respond as their environment slowly changes. In regenerated silkworm silk, gradual changes in acidity reveal key stages that govern the transition from soluble proteins to organized networks.

The thesis also examines a lesser-known spider silk used to protect spider eggs. Unlike many other spider silks, this protein is already partly organised in solution, which may allow faster and more reliable fibre formation. Understanding this pre-organised state provides insight into how protein design influences material formation.

In parallel, the work explores pea protein systems used in plant-based foods. Here, interactions between proteins and naturally occurring carbohydrates such as starch strongly influence how gels form and how their internal structure changes.

By studying silk and plant proteins side by side, this thesis highlights shared principles of protein organisation across biological systems. These insights help explain how nature controls structure so efficiently and may guide the development of greener materials and food processes inspired by biology.

Populärvetenskaplig sammanfattning

Många av de material vi förlitar oss på i vardagen, såsom plaster, fibrer, beläggningar och livsmedelsgeler, framställs med energikrävande processer som bygger på höga temperaturer, starka kemikalier eller omfattande mekanisk bearbetning. Naturen producerar däremot rutinmässigt högpresterande material under milda förhållanden. Silke är ett tydligt exempel. Spindlar och silkesmaskar spinner fibrer som är lätta, starka och hållbara med hjälp av vattenbaserade proteinlösningar vid rumstemperatur. Att förstå hur dessa proteiner organiserar sig innan fibrer och geler bildas är ett viktigt steg mot mer hållbara teknologier.

Denna avhandling undersöker hur proteiner gradvis organiserar sig till större strukturer, från enskilda molekyler till utsträckta nätverk. Silkesproteiner fungerar som ett centralt modellsystem, medan ärtproteiner utgör ett kompletterande flerkomponentsexempel med relevans för hållbar livsmedelsproduktion. Trots att systemen skiljer sig åt i syfte och biologiskt ursprung bygger båda på kontrollerad proteinorganisation för att uppnå sina slutliga egenskaper.

Ett huvudfokus i arbetet är vad som händer med proteiner när de fortfarande befinner sig i lösning. I stället för att bilda fasta material i ett enda steg passerar silkesproteiner genom övergående arrangemang som bidrar till stabilitet och styr sammansättningen. Genom att kombinera flera experimentella metoder som undersöker struktur på olika längdskalor analyserar avhandlingen hur silkesproteiner reagerar när deras omgivning förändras långsamt. I regenererat silke från silkesmask visar gradvisa förändringar i surhetsgrad viktiga steg som styr övergången från lösliga proteiner till organiserade nätverk.

Avhandlingen behandlar också ett mindre känt spindelsilke som används för att skydda spindelägg. Till skillnad från många andra spindelsilken är detta protein redan delvis organiserat i lösning, vilket kan möjliggöra snabbare och mer tillförlitlig fiberbildning. Förståelsen av detta förorganiserade tillstånd ger insikt i hur proteindesign påverkar materialbildning.

Parallellt undersöks ärtproteinsystem som används i växtbaserade livsmedel. Här har interaktioner mellan proteiner och naturligt förekommande kolhydrater, såsom stärkelse, stor betydelse för hur geler bildas och hur deras inre struktur förändras.

Genom att studera silkes- och växtproteiner sida vid sida belyser denna avhandling gemensamma principer för proteinorganisation i biologiska system.

Dessa insikter bidrar till att förklara hur naturen kontrollerar struktur så effektivt och kan vägleda utvecklingen av grönare material och livsmedelsprocesser inspirerade av biologin.

Acknowledgements

As the saying goes, it takes a village…

I would first like to thank the Swedish Foundation for Strategic Research (SSF), Agenda2030, and SwedNess for the financial support, which made this work possible.

Cedric, Felix, Andrew and Wolfgang made up my supervisory team. Yes, that is a lot of men in one room, but honestly, they are some of the most loving and nurturing people I know.

Cedric, I was immediately sold on the research topics, but one of the main reasons I took on this challenge was because of you. Your passion for research is obvious, but your love for teaching is even more profound. I have been one of many lucky students to have you as a teacher, and one of the few to have you as a supervisor and mentor. Thank you for supporting me through whatever academic or personal challenges occurred along the way. You helped me stand my ground, think more critically, and do so with an open mind.

Felix: You can say a lot in a very short amount of time. I thought my friend Dana (also German), talked fast, but you definitely beat her to that. At first, I was overwhelmed by the depth and speed of what you taught, but over time, repetition and your patience made all the difference. Thank you for all you have taught me and our great chats.

Andrew: Although you were often very busy throughout my PhD, the times we did meet were always greatly appreciated. Your guidance and support, especially during beam time at ANSTO, and the ease of having honest, real conversations meant a lot to me.

Wolfgang: Thank you for adopting me into the LP3 family, my professional growth here has been obvious. You were always encouraging, gave sound advice, and never doubted me. P.S. Sorry for breaking all those columns and apparently bringing bad luck to the expensive equipment, but at least we got new equipment!

I am so grateful to the ANSTO team, especially **Elliot and Jitendra**, for their guidance, support, and welcoming me into their Australian life.

Life at the TBK did not start out easy; COVID made sure of that. Although my time there was brief before moving to LP3, I watched the department regain its sense of identity, warmth, and kindness (often accompanied by excellent cakes). To my adopted LP3 family, **Ewa, Heather, Anna, Celeste, Victoriia, Zoe, and Maria**, you were loud (mainly Ewa), but you made office life truly fun. Thank you for the laughs, love, and for proving that noise-cancelling headphones

actually work. Being part of LP3 and the biology department allowed me to meet incredible PhD students and postdocs who are now friends, and I am grateful for the spontaneous moments and shared memories that made this journey so special.

To the **SwedNess** fam, what began as 20 strangers quickly became many great friendships. Once a year, we were brought together, and over time, we confided in one another through shared struggles and beautiful moments. Thank you! **Gigi**, our friendship has not always been easy, but it has always been deeply valued. Thank you for keeping me grounded and for giving me the space to be myself. Through **Agenda2030**, I had the pleasure of meeting colleagues from many disciplines. Thank you for the thoughtful discussions that broadened my research perspectives.

Moving to Sweden has been the best decision I've ever made, not only for the life I get to live, but for the friends who became family along the way. **Jojo, Ling, Shruti, and Ingrid**, I've watched you graduate, earn your degrees, and be merry, now watch me!! **Jade**, you somehow made this group commit to weekly horoscopes, but I love you anyway. You ladies, Shuai and Anna-Marie, have made this journey something truly special, and I'm so grateful for all the beautiful memories we've shared and those still to come. To my Movie Hoes (can I say that?), **Gillis, Selim, Frida, Mathilda, Gabby**, occasionally **Lisa, Herman, Daniel, and Kokos** (the family cat), I love you all so much. You probably have no idea what I actually did during my PhD, but you always remind me that I'm doing something cool. **Isa, Dana, Love** if we could live in the same commune, I'm sure we would. We've grown up together and been through so much, and I look forward to the future knowing I'll have you by my side. Can't wait for that Italian trip… To the **Danes** (**Sigrid and co**), you brought joy and beers throughout my PhD; another round of Roskilde please!! To my Canadian friends incl. **Nicole and Chels**, it's rare we get to meet, but when we do, it's like we're kids and teenagers again.

Sadly, much of my PhD has been spent away from my family back in Canada. Even when we were usually 6 hours apart, you were always with me in the moment. Thank you **Mom, Dad, and JP** for the constant love and support. To the fam across the Bridge, **Wumi, Auntie Doreen, Maya, Penelope, and Quentin**, you give me a headache sometimes, but that's exactly what family is for, right? And to the family 1 hour behind **Julie, Steve, Ella and Joe**, thanks for loving me like one of your own. Also, thanks for sharing what he likes to believe is your favourite family member…

SAAAAMMMMM!! 8 years and counting, that is mad. I could say so much, but to keep within the 2 pages, I'll simply say thank you for the endless support. You have always remained patient, grounded and steady, somehow managing to bring me back to myself every time. My love for you is endless. We still have so much life to live together, so let's get on with it <3.

List of papers

Paper I
Francis J., Houston J., Jackson A., Dalgliesh R., Martel A., Porcar L., Roosen-Runge F., Dicko C. (2026). pH-Triggered Clustering Regulates β-sheet Activation in Silk Assembly. *Commun Chem.* https://doi.org/10.1038/s42004-025-01875-7

Paper II
Francis J., Haas S., Jackson A., Roosen-Runge F., Yoshioka T., Dicko C. (2026). Acid vapour-induced transition in silk films favouring silk I polymorphs: implications for the control of silk II emergence.
Manuscript in preparation for submission

Paper III
Francis J.*, Dessì F.*, Jackson A., Gilbert E.P., Mata J.P., Schirone D., Roosen-Runge F. (2026). Truncated Protein Extraction Modulates Pea Protein–Carbohydrate Thermal Co-gelation.
*These authors contributed equally to this work.
Accepted for publication in *Food Hydrocolloids*.

Paper IV
Francis J., Lund M., Dicko C. (2026). SAXS-guided insights into the unusual extended modular architecture of native cylindriform silk protein.
Manuscript in preparation for submission

Paper V
Francis J. (2025). Challenges in Recombinant Production of a Minimal Cylindriform Spidroin.
Short communication (not submitted for publication)

Author's contribution to the papers

Paper I
The author designed the study together with the co-authors. The author performed the experimental investigation, curated the data, conducted the formal analysis and validation, and led the visualisation with input from the co-authors. The author drafted the original manuscript and coordinated the editing and revision of the manuscript for publication together with a co-author.

Paper II
The author contributed to the data interpretation, validation, and visualisation. The author co-drafted the original manuscript and, together with the co-authors, edited and revised it.

Paper III
The author contributed to the conceptualisation and methodology of the study and carried out experimental investigation, data curation, formal analysis, interpretation, validation, and visualisation. The author co-drafted the manuscript with two co-authors and contributed to review and revision, including during peer review.

Paper IV
The author contributed to the data interpretation, validation, and visualisation. The author drafted, edited, and revised the manuscript together with a co-author.

Paper V
The author conceptualised the study and methodology, performed the experimental investigation, curated and analysed the data, conducted validation and visualisation, and drafted the original manuscript. Regarding the yeast research, the author's contributions were limited to the data analysis and interpretation. The manuscript was edited and revised by the author for inclusion in this thesis.

Abbreviations

AE	Alkaline extraction.
ASG	Anterior silk gland.
ATR-FTIR	Attenuated total reflectance Fourier transform infrared spectroscopy.
ATSAS	Analysis tools for small-angle scattering.
CALVADOS3	Coarse-graining approach to liquid–liquid phase separation via an automated data-driven optimisation scheme.
CORAL	COmplexes with RAndom loops.
CT	C-terminal or C-terminus domain(s).
CV	Contrast variation.
D$_{max}$	Maximum particle dimension.
DSC	Differential scanning calorimetry.
EOM	Ensemble optimisation method.
FibH	Fibroin heavy chain.
FibL	Fibroin light chain.
Gal8N	N-terminal domain of galectin-8.
GdL	Glucono-δ-lactone.
His$_{6/8}$	Polyhistidine affinity tag (six or eight residues).
IMAC	Immobilized metal affinity chromatography.
kDa	Kilodalton.

MCR-ALS	Multivariate curve resolution–alternating least squares.
mEGFP	Monomeric enhanced green fluorescent protein.
MSG	Middle silk gland.
NT	N-terminal or N-terminus domain(s).
NUrF	Small-angle neutron scattering, ultraviolet–visible spectroscopy, Raman spectroscopy, and fluorescence spectroscopy.
PPIs	Protein–protein interactions.
PSG	Posterior silk gland.
R_g	Radius of gyration.
RP	Repetitive protein domain (e.g. RP1, RP2).
RSF	Reconstituted silk fibroin.
SANS	Small-angle neutron scattering.
SAXS	Small-angle X-ray scattering.
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis.
SE	Salt extraction.
SEC	Size-exclusion chromatography.
SLD	Scattering length density.
SUMO	Small ubiquitin-like modifier tag.
TEV	Tobacco etch virus.
ThT	Thioflavin T.
TuSp1	Tubuliform silk protein 1.
USANS	Ultra-small-angle neutron scattering.
UV-Vis	Ultraviolet–visible.
WAXS	Wide-angle X-ray scattering.
WB	Western blotting.

Introduction

Motivation and Overview

Protein-based materials such as structural fibres and food gels display functional properties that emerge from organisation across multiple length scales.

Many protein systems can be induced to coagulate or solidify in response to physicochemical conditions. However, these transitions do not necessarily imply ordered assembly and may instead yield amorphous or kinetically trapped states (Bishop et al., 2009). Silkworm and spider silks represent a distinct assembly paradigm in this respect. Few protein systems combine comparable hierarchical organisation, material performance, and assembly efficiency under mild aqueous conditions (Craig, 2003; Vollrath & Knight, 2001).

Despite their evolutionary distance, the spinning processes of silkworms (insects) and spiders (arachnids) show strong functional convergence. This convergence supports integrating insights from spider and silkworm systems toward a more general understanding of silk spinning (Craig, 2003).

Plant protein systems, such as pea proteins, provide a complementary contrast. In these systems, functional networks are typically obtained through multistep processing, which increases material and energy demands in industrial settings. Understanding how hierarchical organisation can emerge under mild or reduced processing conditions is therefore of broad interest. Such insights may inform strategies for designing protein-based materials with reduced processing requirements and improved control over material properties.

Historically, experimental studies of silk have focused on the properties of the final fibre rather than on the assembly process itself. Recent experimental and computational advances have improved access to earlier stages of silk assembly. This has generated diverse models of intermediates based on phase behaviour, compartmentalisation, or discrete stages (Breslauer, 2025; Lay et al., 2025). Differences in emphasis and level of description make it unclear which features are essential for hierarchical assembly and which reflect system-specific contingencies.

To rationalise these diverse observations, hierarchical protein assembly is often framed as a sequence of structural transitions across length scales. In this view, molecular-level organisation constrains mesoscale structure, which in turn governs macroscopic material formation (Hayes et al., 2021; Lay et al., 2025).

In silks, this organisation is thought to arise from an interplay between intrinsic sequence features, environmental conditions such as pH, ionic strength, and solvent composition (Askarieh et al., 2010; Lay et al., 2025). Disentangling the relative contributions of these factors is essential, as differences in *how* assembly proceeds, not simply *whether* it occurs, determine the structure and properties of the resulting material.

For example, silkworm and spider silks are composed of distinct protein sequences and exhibit different mechanical properties under native spinning conditions. Nevertheless, silkworm silk can approach the properties of a spider silk when spun under spider-like conditions (Shao & Vollrath, 2002). This convergence does not imply equivalence of the proteins or their native properties. Rather, it suggests that spinning acts on a protein solution that is already structured in a way that enables a specific response to processing.

Maintaining assembly competence at extremely high protein concentrations in silk glands under conditions of thermodynamic metastability is a non-trivial physicochemical challenge. Despite their intrinsic aggregation propensity, silk proteins remain soluble and mobile until spinning is initiated (Andersson et al., 2016; Lefèvre et al., 2011; Vollrath & Knight, 2001). Several contributing factors to this stability have been proposed, including sequence amphiphilicity, terminal domains, glandular gradients in pH and hydration, and suppression of irreversible β-sheet locking (Askarieh et al., 2010; Boulet-Audet et al., 2014; Jin & Kaplan, 2003; Terry et al., 2004; Vollrath & Knight, 2001).

These stabilising features are not merely protective. They influence the structural organisation of the concentrated solution dope and therefore likely constrain subsequent hierarchical assembly. Many descriptions of silk formation implicitly assume that molecular, mesoscopic, and macroscopic events are coordinated as assembly progresses. However, in practice, this coordination is reconstructed from separate structural, spectroscopic, and scattering measurements performed at different length and time scales. The absence of direct multiscale correlation limits our ability to mechanistically link proposed intermediates to final material outcomes.

In response to this limitation, the central question addressed here is: *"Which structural and molecular features of the silk protein solution dope are already established before spinning, and which constrain hierarchical assembly during subsequent processing?"* This thesis tests the hypothesis that hierarchical assembly outcomes are selected upstream of macroscopic formation. These outcomes depend on pathway-dependent solution-state organisation, including transient intermediates and solvent-mediated structure.

To address the lack of correlated multiscale measurements, a multimodal experimental strategy is established, centred on a coordinated neutron–UV–fluorescence (NUrF) platform (Dicko et al., 2020). Scattering and spectroscopic

measurements probe the same evolving assembly process on complementary timescales. Together with integrative structural modelling, this approach enables correlated observation of molecular and structural changes during assembly. The strategy is applied to model systems in which sequence, processing history, and architectural flexibility can be systematically varied. Accordingly, two silk systems are examined: regenerated *Bombyx mori* silk fibroin as a native-derived system, and a mini-engineered spidroin derived from spider cylindriform silk as an architecturally tunable model.

Pea proteins represent a multicomponent assembly landscape that contrasts with the highly specialised silk systems described above. Whereas silk assembly is encoded within a defined protein architecture, the pea protein systems explored here are multicomponent. They involve interactions with retained native constituents such as starch, which influence protein organisation and gel network formation. Contrast-variation small-angle neutron scattering enables protein and carbohydrate contributions to be separated. This allows the effects of reduced extraction and processing history on protein self-assembly, phase behaviour, and gelation to be assessed.

Aims and scope

This thesis examines how protein-based materials organise hierarchically in response to pre-assembly solution-state organisation, molecular architecture, and environmental conditions. Two contrasting protein systems are considered: silk proteins undergoing triggered self-assembly, and multicomponent pea protein systems in which interactions with co-existing biomolecules and processing history modulate network formation. Together, these systems provide a basis for identifying principles linking solution-state organisation to material outcomes, with relevance for protein processing and food-relevant gel systems.

The specific objectives are to:

1. Characterise how environmental conditions govern both the phase behaviour and assembly pathways of silk proteins. Protein concentration, pH, and solvent composition are examined using multimodal, time-resolved approaches to resolve sol–gel transitions, transient intermediates, and higher-order structure formation.

2. Explore co-assembly behaviour in pea protein–starch systems. Emphasis is placed on how extraction and processing conditions shape gel-like network formation.

3. Examine the solution-state architecture of a folded multi-domain cylindriform silk protein. Experimentally guided modelling is used to establish a solution-state structural reference at the level of overall architecture and domain organisation.

4. Develop and study recombinant cylindriform spidroin systems. These models are used to test the robustness and organisational consequences of native domain architectures outside their biological context.

The thesis is based on the following original research papers:

- **Paper I**: Investigates the gelation of regenerated silk fibroin under gradual acidification. A multimodal NUrF platform is used to identify intermediate states and associated structural, molecular, and physicochemical transitions.

- **Paper II**: Investigates acid vapour–induced gelation of regenerated silk fibroin thin films, using SAXS, WAXS, and IR spectroscopy to track conformational changes within silk polymorphism.

- **Paper III**: Explores co-gelation of pea proteins and carbohydrates, showing how truncated extraction methods influence starch retention and network formation.

- **Paper IV**: Characterises the native fold of the cylindriform silk protein using SAXS-guided modelling, and discusses challenges in capturing native-like conformations.

- **Paper V**: Reports the design and characterisation of a minimal recombinant cylindriform spidroin, providing insights into domain interactions and assembly behaviour.

Chapter 1 – Natural Systems of Hierarchical Protein Assembly

Molecular self-assembly is a ubiquitous phenomenon in biology. It enables the formation of ordered structures across molecular, mesoscopic, and macroscopic length scales (J. Li, 2017). Such assemblies arise from diverse biological building blocks, including lipids, carbohydrates, nucleic acids, and proteins, and support functions ranging from molecular recognition to mechanical reinforcement (Halperin-Sternfeld et al., 2017).

Among these, proteins stand out as particularly versatile. Their sequence-encoded architectures allow access to a wide range of conformations and interactions. As a result, multiple competing assembly pathways can arise, enabling complex hierarchical assemblies with tailored functional properties. This flexibility has motivated extensive efforts to translate biological assembly strategies into engineered biomaterials (Nanda et al., 2011).

A central challenge is that hierarchy remains difficult to predict from the primary sequence alone. Protein assembly rarely follows a single, two-state route (Dobson, 2003). This reflects not only sequence degeneracy but also the strong influence of pathway history and environmental context on assembly outcomes.

This chapter surveys natural protein systems in which hierarchical self-assembly is regulated by defined environmental triggers. It focuses on mechanisms where external cues, including pH, solvent exposure, or heat, drive transitions from soluble precursors to ordered networks. Silk proteins (fibroin and spidroin) serve as the primary model systems for evolutionarily optimised assembly, while pea protein gels provide a supplementary platform for examining multicomponent, food-relevant networks. Together, these systems provide complementary perspectives on how molecular design and environmental conditions shape protein self-assembly across time and length scales.

Protein assembly by design

Natural proteins self-assemble through pathways shaped by evolution, providing reference points for understanding hierarchical assembly. Designed protein systems instead use biomimetic strategies to mimic or re-engineer these pathways under controlled biochemical and physical conditions (J. Li, 2017; Zhu et al.,

2021). Biomimetic approaches range from imitating natural motifs to creating entirely new architectures. These strategies offer opportunities to engineer materials with specific, tunable properties (J. Li, 2017). Central to this endeavour is understanding the design principles that govern both natural and synthetic protein assembly, as well as the limitations that arise when pathway control is incomplete.

Protein assemblies are commonly classified as either finite or extended structures. Finite assemblies, such as haemoglobin, often carry out specific biochemical functions (Jensen et al., 1998). This thesis focuses on extended assemblies. These are generally polymeric and exhibit varying degrees of crystalline or semi-crystalline order, as seen in silk proteins and plant protein gel-like networks (Hardy et al., 2008).

Protein assembly is governed by a set of interrelated design principles that influence how proteins fold, interact, and organise into higher-order structures. These principles, summarised in Table 1, span intrinsic features such as sequence patterns and domain arrangement, as well as extrinsic factors like pH and temperature. Together, these principles describe how proteins transition from individual chains to functional, multi-scale assemblies with distinct material properties. Importantly, they do not uniquely determine the resulting assembly pathways or outcomes (Zhu et al., 2021).

Table 1: Key design principles in protein assembly and their roles (contributing factors rather than deterministic rules)

Design Principle	Role in Protein Assembly
Sequence patterns	Define folding pathways, interaction specificity, and self-assembly potential through recurring motifs and repeats that dictate responsiveness to environmental triggers (Zhu et al., 2021)
Domain architecture	Modular domains (e.g., β-sheets, coiled-coils) direct higher-order motifs and enable hierarchical organisation (Del Sol & Carbonell, 2007).
Environmental cues	External stimuli such as pH, ionic strength, temperature, and shear forces modulate assembly kinetics and structural transitions (Zhu et al., 2021).
Dynamic flexibility	Intrinsic disorder and conformational plasticity permit reversible assembly, adaptability, and regulation of function (Bhattarai & Emerson, 2020).
Hierarchical Propagation	Local interactions propagate to mesoscale and macroscale structures, linking molecular design to emergent material properties (Sun et al., 2020).

The underlying driving forces and molecular interactions are summarised in

Figure 1. These include hydrophobic effects, hydrogen bonding, electrostatic interactions, and mechanical influences that contribute to protein self-assembly.

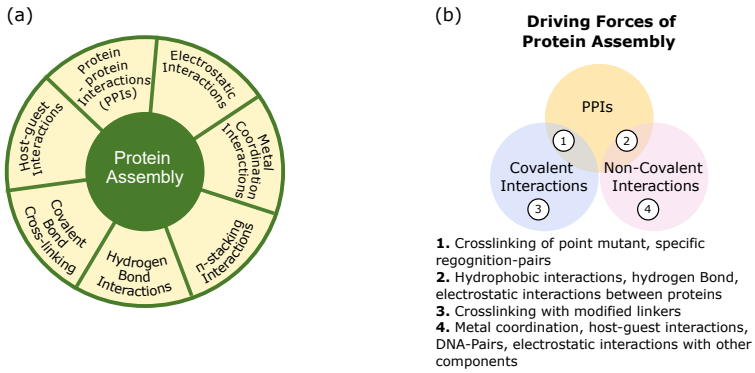


Figure 1: Driving forces and interactions involved in protein assembly. (a) Classification of molecular interactions contributing to protein assembly. (b) Conceptual classification showing how protein–protein interactions (PPIs) arise from covalent and non-covalent interaction types during assembly. Adapted from Y. Li et al. (2023), published under a Creative Commons Attribution (CC BY) license.

These principles provide a point of reference, but their application to biological systems must account for the inherently non-linear and context-dependent nature of assembly pathways (Zhu et al., 2021). For silk proteins, the process is referred to as *silk fibrillogenesis*, describing the series of events by which soluble precursors are converted into ordered fibrous structures. Multiple models have been proposed to describe this transformation (Brookstein et al., 2024; Eliaz et al., 2022; Jin & Kaplan, 2003; Landreh et al., 2024; Mohammadi et al., 2018; Parent et al., 2018; Vollrath & Knight, 2001). Each provides valuable mechanistic insight, yet none fully captures the system's complexity. With the current understanding, it is clear that a dynamic interplay between sequence, structure, and processing conditions exists. Analysing how these factors interact during assembly remains difficult to resolve experimentally.

Silk Proteins: Fibroins and Spidroins

Silk proteins provide a well-studied example of hierarchical protein self-assembly. Among them, silkworm fibroins and spider spidroins are especially notable for forming highly ordered fibrillar structures with outstanding mechanical properties (Breslauer, 2025; Holland et al., 2012; Vepari & Kaplan, 2007; Vollrath & Knight, 2001).

The biological functions of silk differ substantially between silkworms and spiders. Silkworms spin a single silk type to construct protective cocoons, yielding a relatively uniform material in substantial quantity (Andersson et al.,

2016). Spiders, by contrast, produce multiple silks for prey capture, locomotion, shelter, and egg protection, among other functions. These spidroins differ in their underlying sequences, reflecting the behavioural and ecological diversity of the species. Yet, each type is made in comparatively small amounts (Zheng & Ling, 2019). This functional range is mirrored at the molecular level by the regulatory and sequence diversity of spidroins. These differences, in turn, shape their assembly pathways.

Although their functions differ, both fibroins and spidroins share core design principles: repetitive sequence motifs, high protein storage within the gland, environmentally responsive assembly, and a propensity to form β-rich regions (Andersson et al., 2016). These shared features make silk a useful model for studying how molecular architecture, structural organisation, and environmental cues influence hierarchical assembly.

Fibroins (native and regenerated)

Native Fibroin

Molecular architecture and composition: Silk from the domesticated silkworm (*Bombyx mori*) is the most extensively studied silk system, largely because it is readily accessible and can be produced in substantial quantities. This silk is spun exclusively for cocoon formation, where it protects the pupa against environmental stress (F. Chen et al., 2012; Lujerdean et al., 2022). At the molecular level, silkworm silk is a composite material composed of *fibroin*, which forms the structural core, and sericin, a glue-like coating that binds the fibres together. Fibroin itself is a heterodimer consisting of a heavy chain (FibH, $\sim$350 kDa) and a light chain (FibL, $\sim$25 kDa), covalently linked by a single disulphide bond at their C-termini (CT) (Moreno-Tortolero et al., 2024; Tanaka et al., 1999). Its amino acid composition is dominated by glycine (43 %), alanine (30 %), and serine (12 %), which give rise to the repetitive architecture of the FibH chain. Gly–Ala–rich motifs (e.g. GAAS, GAGAGY) within the repetitive domains drive antiparallel β-sheet formation. This sequence-encoded molecular architecture is intrinsic to the nature of silk fibres and underpins their high strength. The non-repetitive FibL chain contributes to hydrophobicity and elasticity (Qi et al., 2017; Zhou et al., 2000).

Glandular processing and spinning environment: In vivo, assembly occurs within a specialised silk gland, where fibroin is produced in the posterior silk gland (PSG) and sericin in the middle silk gland (MSG). Following synthesis, fibroin enters the MSG lumen and is stored as a concentrated aqueous solution, or dope (Dong et al., 2016). During spinning, this solution is transported through the

anterior silk gland (ASG), where gradual changes in pH, ionic environment, and flow conditions trigger conformational rearrangements. Elongational flow, shear, and dehydration progressively align the protein chains. These processes promote the transformation from soluble Silk I structures, often described as hydrated and metastable packing states. During this transition, the protein secondary structure reorganises into β-sheet-rich domains, yielding the crystalline Silk II state characteristic of the final fibre (Brookstein et al., 2024; Terry et al., 2004; Wan et al., 2021).

Assembly pathways and intermediates: Although the overall transformation from soluble to fibrous states is well established, the molecular steps linking Silk I and Silk II remain unresolved. Different models emphasise distinct aspects of the process, which may reflect differences in scale and experimental access rather than mutually exclusive mechanisms. These include: sequence motifs and terminal domains acting as molecular switches (Eliaz et al., 2022; Parent et al., 2018); the combined effects of pH, ions and dehydration in the glandular duct (Brookstein et al., 2024; Holland et al., 2012; Terry et al., 2004); and hierarchical pathways in which fibroin passes through micellar or clustered intermediates before consolidating into fibrillar networks (Eliaz et al., 2022; Landreh et al., 2024; Mohammadi et al., 2018; Vollrath & Knight, 2001).

What unites these perspectives is the recognition that silk assembly proceeds through a sequence of structural checkpoints marked by transient intermediates. These checkpoints regulate whether fibroin remains soluble, aggregates prematurely, or progresses toward fibril and fibre formation. Yet the precise identity, ordering, and kinetics of these intermediates remain debated (Lay et al., 2025).

This uncertainty motivates the use of regenerated silk fibroin (RSF) as an experimentally accessible model system to explore these transitions. Although native silk solutions were considered, their extraction and handling outside the gland proved challenging, making systematic measurements impractical.

Reconstituted silk fibroin (RSF)

When silk fibres from *B. mori* cocoons are extracted and regenerated in the laboratory, they can be re-dissolved to form RSF solutions. The process typically involves degumming to remove sericin. Fibroin is then dissolved in the chaotropic salt lithium bromide (LiBr) and dialysed to obtain an aqueous protein solution (Figure 2) (X. Chen et al., 2007; Zuo et al., 2006). The extracted RSF retains the core molecular design of fibroin but is stripped of the glandular environment, terminal regulation, and processing gradients that normally govern storage and silk spinning *in vivo*.

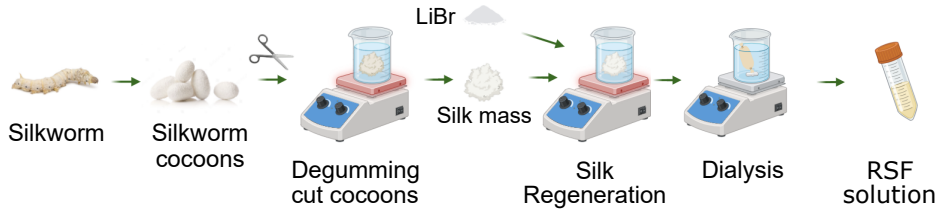


Figure 2: Workflow for RSF preparation. Silkworm cocoons are cut and degummed to remove sericin, leaving a fibroin-rich mass. The fibroin is dissolved in 9.3 M LiBr and subsequently dialysed against water to remove the salt, yielding an aqueous RSF protein solution.

While RSF is widely used, several studies have raised concerns about its suitability as a representative model for native silk protein solutions (Boulet-Audet et al., 2014; Greving et al., 2010; Moreno-Tortolero et al., 2024). Reported shortcomings include changes in molecular weight distribution, loss or destabilisation of critical components (such as the N-terminal pH-switch domain), and disruption of native folding and packing. As such, RSF lacks the monodispersity and pre-organised structure characteristic of its natural counterpart. This makes it more susceptible to less controlled phase transitions during gelation (Huang et al., 2023).

Even with these limitations, RSF retains the core molecular interactions involved in silk assembly, including hydrogen bonding, hydrophobic effects, and β-sheet formation, and responds to pH and ionic strength in a broadly similar manner to native fibroin (Breslauer, 2025). While its assembly pathway likely diverges from that of native silk, RSF remains the predominant system for investigating silk assembly *in vitro*. In this context, RSF does not replicate the precise execution of native spinning. Instead, it probes the robustness of the underlying assembly logic. By demonstrating that key interactions between sequence motifs and external triggers persist in a perturbed system, RSF provides a useful baseline for identifying the minimum requirements for hierarchical protein organisation. For this reason, it remains highly relevant.

Building on this perspective, **Paper I** (Chapter 3) examines RSF gelation under gradual acidification, resolving assembly intermediates and key structural checkpoints using the **NUrF** platform, which integrates **N**eutron small-angle scattering, **U**V–Vis spectroscopy, and **F**luorescence measurements. **Paper II** (Chapter 3) extends this analysis to acid-vapour induction, placing these features within the broader context of silk polymorphism, Silk I and Silk II structures. Together, these studies clarify the transition from soluble fibroin to insoluble β-sheet-rich networks and provide new mechanistic insight into silk assembly.

Spidroins (native and recombinant)

While fibroin from *B. mori* has provided the most accessible and widely studied model of silk assembly; spider silks illustrate the broader diversity of silk protein systems. Among specialised fibre types, dragline (major ampullate) silk has been the most intensively studied, renowned for its superior mechanical strength and toughness (Rising & Johansson, 2015). Dragline spidroins have therefore become the standard model for understanding spider silk mechanics and molecular assembly. However, they present several limitations as a system for investigating the fundamentals of bio-inspired polymer processing. Dragline spidroins often lack well-defined solution structures, are prone to insolubility, and require considerable downstream processing to prevent premature aggregation (Dicko et al., 2008).

Cylindriform (tubuliform) silk offers a biologically meaningful alternative system for studying protein assembly. Produced exclusively by female spiders, it forms the protective egg casings that provide a robust barrier against environmental fluctuations and biological threats (Lin et al., 2009). Like silkworm silk, cylindriform silk is spun largely as a one-time event, in contrast to the continuous or on-demand spinning strategies characteristic of dragline silk. This distinct biological context is thought to be reflected in differences in the underlying assembly pathway.

Spidroins that make up dragline silk are generally described as highly disordered in solution. Cylindriform spidroins, on the other hand, have been reported to adopt a more ordered, predominantly α-helical solution state (Clark et al., 2019; Wen et al., 2017). This difference has led to the proposal of a comparatively direct α-to-β conformational transition during fibre formation. As illustrated in Figure 3, lower glycine content and reduced chain flexibility favour a predominantly α-helical organisation prior to spinning (Dicko et al., 2008). During fibre formation, these helical regions are converted into densely packed, β-sheet-rich structures that confer the mechanical robustness required for egg sac protection (Tian & Lewis, 2006).

The shorter glandular duct and a reported reduced dependence on strong pH-driven fibrillation (Wang et al., 2021) are consistent with a model in which cylindriform silk assembly requires fewer conditioning steps than dragline silk. However, the structural basis for this remains incompletely understood.

From an evolutionary perspective, such an arrangement may minimise the energetic and temporal costs associated with preparing the spinning dope. A more structured starting state could reduce the need for strong physicochemical gradients and high shear, limit susceptibility to misfolding, and enable rapid fibre formation, a strategy well suited to a single, high-volume deployment (Craig, 1997; Dicko et al., 2004).

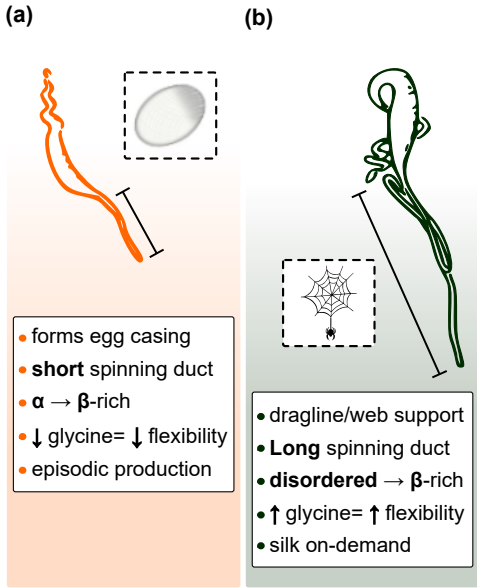


Figure 3: Comparison of cylindriform and major ampullate silk systems. Schematic illustration highlighting key differences between cylindriform (a) and major ampullate (b) silk glands and their respective fibre functions. Cylindriform silk shows lower glycine content and reduced chain flexibility with a transition from α-helical to β-rich structure, whereas major ampullate silk shows higher glycine content, greater flexibility, and a transition from disordered to β-rich structure. Conceptual illustration informed by Dicko *et al.* (2008).

Taken together, these features position cylindriform silk as a useful model system for investigating hierarchical protein assembly. They suggest that its conformational transition is comparatively simple, making it well suited to direct structural characterisation by small-angle scattering requiring minimal sample conditioning.

Native cylindriform spidroins are produced exclusively in the cylindriform glands of female spiders during the reproductive season and rapidly lose their native behaviour outside physiological conditions, making their extraction experimentally challenging (Lin et al., 2009; Wen et al., 2017). These constraints motivate the development of a bio-engineered recombinant cylindriform spidroin. The recombinant constructs retain essential domain architecture and support controlled production and systematic perturbation. In contrast to RSF, they offer greater control over the starting material. Sequence definition and compositional homogeneity are preserved by avoiding solubilisation of pre-assembled fibres into heterogeneous fragments. Both the native cylindriform spidroin from *Nephila edulis* and an engineered recombinant mini-cylindriform spidroin are explored in **Papers IV and V** (Chapters 5 and 6). **Paper IV** focuses on the native fold using modelling guided by small-angle X-ray scattering (SAXS), while **Paper V** examines the designed mini-construct.

Pea proteins and carbohydrates

Pea proteins (*Pisum sativum*) provide a complementary system in which hierarchical assembly occurs during gelation within a mixed biopolymer environment. This enables examination of how co-existing components influence protein interactions and network formation.

Beyond this role, they are central to the development of plant-based foods, where they are highly valued for their favourable nutritional profile, versatility, and reduced environmental impact (Kumar et al., 2022; Shanthakumar et al., 2022).

From a processing perspective, conventional extraction methods for pea protein isolates and concentrates are often inefficient. These approaches typically involve multi-step purification with high water and energy demands. As a result, they generate significant waste, which limits scalability and economic viability. This is particularly problematic given the increasing demand for sustainable and resource-efficient plant protein production (Bagasettihalli Kariyappa et al., 2025; Mondor & Hernández-álvarez, 2022).

Pea proteins are primarily composed of globulins (legumin and vicilin) and albumins. During wet extraction, residual carbohydrates may remain associated with the protein fraction. Recent studies have shown that limited fractionation during purification can significantly alter gelation behaviour compared with highly purified isolates. In particular, retention of native carbohydrates such as starch influences network formation during thermal gelation (Kornet et al., 2021). Starch can modify unfolding and aggregation pathways during heat-induced gelation, while non-starch polysaccharides, including pectins and hemicelluloses, may further influence network organisation (Adamczyk et al., 2022; Kornet et al., 2021).

Despite this, spatial and structural interactions between proteins and residual carbohydrates in less refined fractions remain comparatively underexplored. Addressing this gap provides phase-specific insight into biopolymer organisation in solution and gel networks. It directly links molecular interactions to mesoscale structure under external stimuli.

This work further demonstrates how reduced fractionation shapes solution-state organisation and final network architecture. These findings can help guide low-refinement extraction strategies that reduce energy input while maintaining structural robustness in minimally processed plant-based foods.

Paper III (Chapter 4) characterises minimally processed pea protein extracts before and after heat-induced gelation, using differential scanning calorimetry (DSC), rheology, and contrast-variation small- and ultra-small-angle neutron scattering (CV-SANS/USANS). However, for the focus of this thesis, scattering results are emphasised.

Across the silk and pea systems discussed in this chapter, assembly is experimentally expressed as a sol–gel transition. In both cases, soluble precursors reorganise into viscoelastic, cross-linked networks. This progression involves molecular and structural changes across multiple length scales. Capturing these changes requires multimodal experimental and computational approaches that are sensitive to both molecular structure and mesoscale organisation. Chapter 2 outlines the analytical toolkit used in this thesis.

Chapter 2 – Multiscale Techniques for Characterising Protein Sol–Gel Systems

As outlined in Chapter 1, fibroin, spidroin, and pea proteins assemble hierarchically, with molecular interactions propagating into mesoscale networks and, ultimately, macroscopic materials. No single experimental method can track this process across all relevant length and time scales.

This chapter introduces the experimental approaches employed in this thesis. Scattering techniques, including SAXS, contrast-variation (CV)-SANS, USANS, and WAXS, provide structural information from molecular to mesoscale organisation. Spectroscopic probes such as fluorescence, UV absorbance, and IR spectroscopy report on conformational and chemical transitions. Two integrative approaches are employed: the multimodal time-resolved NUrF platform and SAXS-guided simulations. NUrF integrates complementary probes to resolve assembly correlations, whereas SAXS-guided simulations integrate modelling with scattering data to refine structural organisation. Together, these methods are used to dissect the structural and dynamic pathways of protein sol–gel assembly. The following sections outline how scattering and spectroscopy address these challenges, and where their limitations remain.

Scattering Techniques from Å to µm

Structural methods such as X-ray diffraction, nuclear magnetic resonance, and electron microscopy provide atomic-level detail. However, they often require crystallisation or non-native sample preparation, making them poorly suited for dynamic, hydrated systems (Richards, 2018). Light scattering techniques, such as dynamic light scattering, can probe proteins in solution. However, they are highly sensitive to aggregation and polydispersity. For heterogeneous or evolving systems, this often leads to rapid signal deterioration (Stetefeld et al., 2016). Alternatively, small-angle scattering with X-rays (SAXS/WAXS) or neutrons (SANS/USANS) offers a complementary approach. These techniques probe proteins and their networks in solution and resolve organisation from Ångström motifs to micrometre-scale structures (Jeffries et al., 2021). Neutron scattering

further enables contrast variation through isotopic substitution, allowing selective emphasis of proteins, carbohydrates, or solvent.

Theory in brief: Small-angle scattering probes elastic scattering of X-rays or neutrons at low angles. In elastic scattering, the radiation retains its energy and encodes structural information through changes in direction rather than wavelength. It yields an intensity profile $I(q)$ as a function of the scattering vector $q = \frac{4\pi \sin(\theta)}{\lambda}$, where 2θ is the scattering angle and λ the wavelength.

The signal reflects coherent interference from all atoms in the sample. It is commonly expressed as $I(q) = P(q)S(q)$. Here, $P(q)$ is the **form factor** that describes the size and shape of the scatterers, while $S(q)$ is the **structure factor** that captures inter-particle correlations.

In dilute systems, where inter-particle interactions are negligible, $S(q) \approx 1$ gives direct access to $P(q)$. Features in $I(q)$ relate to real-space dimensions via $d \approx \frac{2\pi}{q}$, where d is a characteristic distance in real space.

Low-q intensities reflect overall size (Guinier region), in which peaks correspond to regular, repeating spacings. High-q reports on finer structural features. In this way, a single experiment can capture structural information across molecular and mesoscale length scales (Jackson, 2010; Jeffries et al., 2021; National Institute of Standards and Technology, 2023).

The principles of SAXS and SANS are conceptually similar, but their interactions differ. In both cases, scattering contrast is governed by the scattering length density (SLD), a measure of a material's scattering strength per unit volume. X-rays, as electromagnetic radiation, scatter from electron clouds and are therefore sensitive to electron density (Stribeck, 2007). Neutrons, being neutral particles, interact with atomic nuclei, and their scattering lengths vary irregularly across elements and isotopes, Figure. 4a. Both methods use transmission geometry, where a collimated beam passes through the sample. The scattered intensity is recorded at small angles on a two-dimensional detector, Figure. 4b.

Small- and Wide- Angle X-ray Scattering (SAXS/WAXS)

SAXS can provide information on solution structures in the nanometre range, while WAXS extends coverage to Å-scale order, such as crystalline β-sheet reflections (Liu et al., 2013).

In this work, SAXS was used to probe protein organisation and assembly in solution. WAXS, in turn, was applied selectively to assess short-range molecular ordering and packing in RSF thin films (Narayanan, 2008). X-ray measurements were carried out at synchrotron facilities P62 PETRA III (DESY, Hamburg) and CoSAXS (MAX IV, Lund).

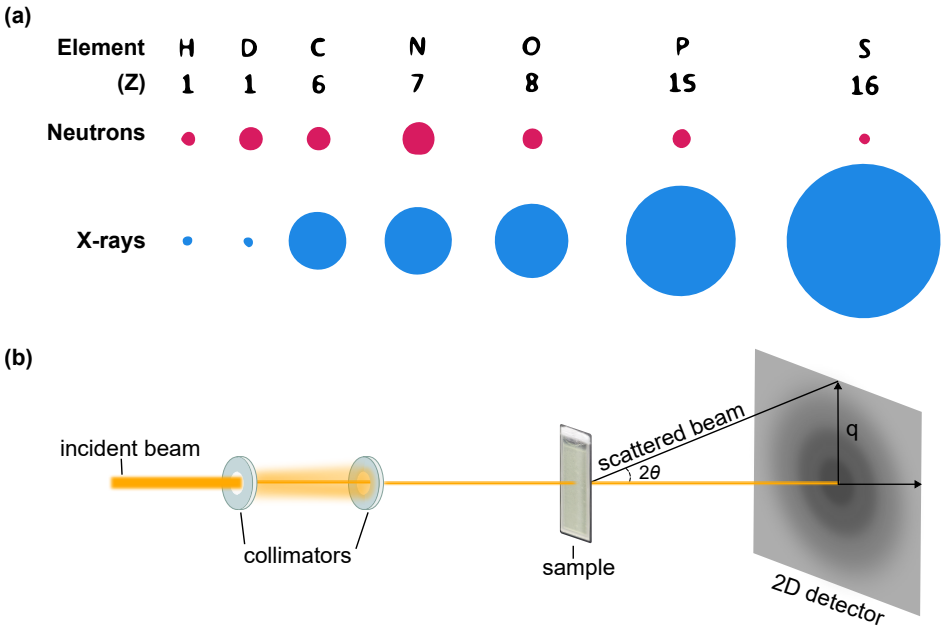


Figure 4: Neutron/X-ray scattering contrast and SANS geometry. (a) Neutrons interact with atomic nuclei and show irregular, isotope-dependent scattering lengths, whereas X-rays scatter from electron clouds. The coloured blobs indicate the relative sign and magnitude of coherent scattering lengths. (b) Schematic of a SANS geometric setup showing the incident beam, sample, scattering angle 2θ, scattering vector q, and detection on a 2D detector.

X-ray scattering can be limited in soft-matter systems due to weak electron-density contrast in aqueous environments. High photon fluxes can also lead to radiation damage at large absorbed doses, inducing protein aggregation or denaturation. Common mitigation strategies include reducing flux or using flowing sample cells. While effective in protecting samples, these approaches reduce signal-to-noise and/or spatial resolution.

An alternative mitigation strategy was adopted during SAXS experiments at DESY. Measurements were performed at high photon energy (18 keV) to reduce absorption and limit radiation damage and beam-induced aggregation while maintaining high photon flux. This preserved sample integrity and signal quality. The shorter X-ray wavelength increased the minimum accessible q, limiting sensitivity to the largest length scales.

Neutron scattering therefore offers a complementary and, in some cases, essential approach for hydrated soft-matter systems. Through contrast variation, specific components can be selectively enhanced or suppressed, providing structural insight that cannot be achieved with X-rays alone (Jeffries et al., 2021).

Ultra- and Small-Angle Neutron Scattering (USANS/SANS) and Contrast Variation

SANS extends scattering measurements into the mesoscopic regime and provides the unique possibility of contrast variation. Neutron scattering lengths vary irregularly across elements and isotopes, allowing hydrogen–deuterium substitution to tune the SLD of the solvent (Figure 5) (Gilbert, 2019). By adjusting H_2O/D_2O ratios, specific components can be selectively highlighted or contrast-matched. This enables different phases within multicomponent systems to be emphasised or suppressed (National Institute of Standards and Technology, 2023).

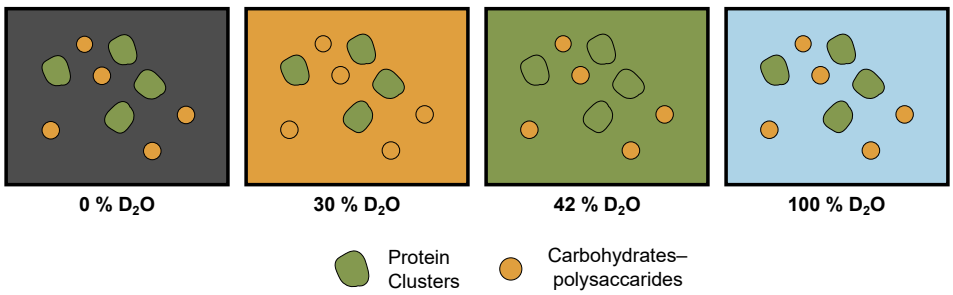


Figure 5: Illustration of contrast variation in neutron scattering using H_2O/D_2O mixtures. Adjusting solvent contrast enables selective enhancement or suppression of protein and carbohydrate contributions. This enables phase-specific analysis of multicomponent protein assemblies, including protein–starch and protein–solvent interactions (Gilbert, 2019). Adapted from Houston et al., *Food Hydrocolloids* (2026). © 2025 Elsevier Ltd. Reproduced with permission.

CV-SANS was applied to disentangle protein and carbohydrate contributions in pea protein–starch systems and to highlight protein–protein correlations in regenerated silk fibroin assemblies. Measurements were carried out at large-scale neutron sources (ILL, ISIS, ANSTO), which provide the flux and extended q-range required for soft-matter studies. USANS, using Bonse–Hart crystal optics, extended access to lower-q, enabling detection of structures beyond the conventional SANS window (Rehm et al., 2013, 2018).

SAS Limitations: Despite their versatility, scattering methods have inherent limitations. The data are indirect and require modelling or assumptions for interpretation, which can introduce ambiguity in complex or polydisperse systems (Jeffries et al., 2021; Stribeck, 2007).

This challenge is particularly evident in biopolymer systems such as silk fibroin and pea proteins. Their assemblies span broad size distributions, ranging from nanometre-scale clusters to extended fibrillar networks. This polydispersity arises from intrinsic compositional heterogeneity as well as from structural evolution during assembly.

Dilute protein solutions further complicate measurements by producing weak

signals. These conditions demand high-flux sources and careful background subtraction, often at the expense of time resolution.

Together, these constraints highlight the need to combine scattering with complementary probes that directly report on chemical and conformational changes during assembly.

Molecular Probes for Protein Assembly

Spectroscopic probes complement scattering by providing chemical and conformational sensitivity that cannot be inferred directly from $I(q)$. They enable tracking of secondary-structure formation, side-chain environment, and aggregation kinetics when measured over time.

Fluorescence Spectroscopy

Fluorescence spectroscopy is a highly sensitive method for monitoring protein structure, capable of detecting changes at nanomolar concentrations. It relies on the re-emission of absorbed light at longer wavelengths (Stokes shift), with emission intensity and wavelength strongly dependent on the local environment (dos Santos Rodrigues et al., 2023). Proteins contain intrinsic fluorophores such as tryptophan, tyrosine, and phenylalanine. Changes in their emission spectra report on solvent exposure, unfolding, or conformational rearrangements (dos Santos Rodrigues et al., 2023; Royer, 2006).

Extrinsic fluorescence was used to monitor Thioflavin T (ThT) binding, a sensitive marker of β-sheet–rich structures. ThT fluorescence provided kinetic traces of fibril formation and network development during gelation, complementing scattering data on hierarchical assembly.

ThT signals can be influenced by quenching and photobleaching. They may also depend on solution viscosity and interactions with ordered or confined surfaces that do not necessarily indicate β-sheet formation. Accordingly, ThT was interpreted cautiously and always in conjunction with structural techniques (Arad et al., 2020; Xue et al., 2017).

UV-Visible Spectroscopy

Ultraviolet-visible (UV-Vis) spectroscopy probes electronic transitions in peptide bonds and aromatic residues, including tryptophan, tyrosine, and phenylalanine. It provides information on protein concentration and changes in local environments (Antosiewicz & Shugar, 2016).

Derivative analysis of UV absorbance spectra was intended to monitor conformational changes involving aromatic side chains. However, experimental

limitations prevented its application in this work.

More generally, UV absorbance offers limited structural specificity and is further complicated by turbidity in samples that aggregate or gel.

Infrared (IR) Spectroscopy

Attenuated Total Reflectance Fourier Transform Infrared (ATR-FTIR) probes molecular vibrations that change dipole moment, yielding characteristic bands for proteins and polysaccharides.

For proteins, the amide I band, $\sim$1600-1700 cm^{-1}, (C=O stretch, $\sim$1650 cm^{-1}) is particularly sensitive to secondary structure, enabling detection of β-sheet formation during gelation (Barth, 2007).

In starch-containing systems, IR absorption in the 1000–1150 cm^{-1} region (C–O and C–O–C stretching) provided a fingerprint for polysaccharide content.

Although performed offline, ATR-FTIR provided complementary confirmation of the ordering processes in RSF and starch-protein gels. These results reinforced trends observed with scattering and fluorescence (Warren et al., 2016).

Integrative Approaches: Resolution, Sensitivity, and Timescale

While scattering and spectroscopy provide complementary insights into protein assembly, each method alone leaves gaps in resolution, sensitivity, or timescale. Integrative approaches address these gaps by combining multiple probes within a coordinated experimental workflow. This enables dynamic processes to be tracked more completely, from molecular changes to mesoscale structure.

A similar limitation applies to experimental and computational approaches. Each provides a partial perspective, and their integration, for example, through SAXS-guided simulations, enables more constrained and informative structural interpretation.

NUrF: A Multimodal, Time-Resolved Platform

The NUrF platform couples SANS with quasi-simultaneous optical spectroscopy (UV–Vis, Raman, and fluorescence) in a single, beamline-compatible setup, Figure. 6 (Dicko et al., 2020). By acquiring scattering and spectroscopic signals under identical conditions, NUrF enables direct correlation of hierarchical structural evolution with conformational and aggregation kinetics in real time.

This integrated design avoids inconsistencies associated with sequential offline measurements and yields synchronised kinetic traces across multiple length scales.

In this thesis, NUrF was used to monitor RSF gelation under gradual acidification using SANS, UV–Vis, and fluorescence, linking molecular ordering to mesoscale network formation as assembly progressed. Raman spectroscopy was not configured for these experiments.

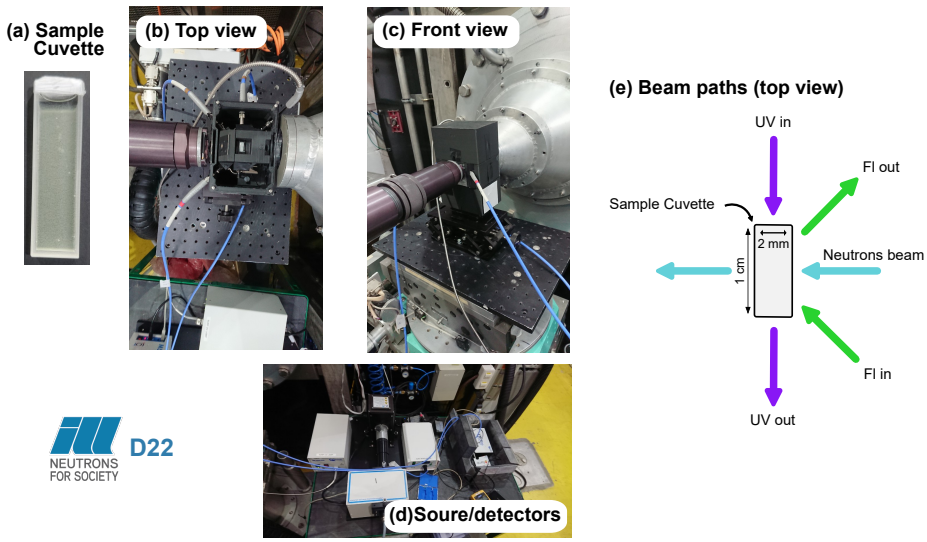


Figure 6: NUrF optical-neutron and instrument layout and optical beam paths. a) Sample cuvette (2 mm optical path, 1 cm height) used for simultaneous neutron scattering and optical spectroscopy. (b) Top-view photograph of the cuvette holder and optical assembly mounted on ILL D22. (c) Arrangement of the light-source modules and detector units for UV–Vis absorption and fluorescence readout. (d) Front-view alignment of the optical head relative to the neutron beam. (e) Schematic of beam paths showing UV excitation (UV in/out, purple arrows), fluorescence collection (Fl in/out, green arrows), and the neutron beam intersecting the sample volume (blue arrows) (adapted from *Paper I*).

Simulations: SAXS-Guided Structural Insights

Scattering experiments yield intensity profiles in reciprocal space. These data do not uniquely define structure and therefore require modelling to infer real-space organisation. This inherent degeneracy means that multiple conformational states can reproduce similar scattering curves.

SAXS-guided simulations address this inverse problem by generating conformational ensembles constrained both by molecular physics and agreement with experimental data. Rather than fitting a single static structure, the approach identifies physically plausible ensembles that reproduce the measured scattering profile.

This strategy was applied here to native cylindriform silk. SAXS-guided simulations provided insight into the distribution of folded states, domain organisation, and conformational flexibility in solution. By bridging reciprocal-space observables and real-space molecular models, the simulations supplied

molecular context to the scattering data and constrained structural interpretation.

Chapter 3 – RSF Gelation under Gradual Acidification

pH-Driven Gelation via Gradual Acidification

Acidification is known to play a central role in silk fibrillation and assembly (Terry et al., 2004). In **Paper I**, this concept was explored experimentally through gradual, *in-situ* acidification of RSF solutions. The **NUrF** platform, combining time-resolved SANS and optical spectroscopy, was used to capture the multiscale structural evolution.

The goal was to reconstruct, under controlled laboratory conditions, the coupled chemical and structural changes that prepare silk proteins for spinning. Gradual hydrolysis of glucono-δ-lactone (GdL) produced a slow, homogeneous pH decrease. This mimics the protonation gradient in the silk gland while excluding flow or shear effects (Del Giudice et al., 2017; Ferreira et al., 2020). This allowed acidification to be studied as a single driving factor, with protein and inducer concentration systematically varied.

Three protein concentrations (5, 10, and 40 mg mL^{-1}) were investigated to capture concentration-dependent differences in assembly behaviour. These concentrations were selected for optical transparency and experimental reproducibility. While they remain well below *in vivo* levels, they preserve the intrinsic pH- and ion-responsiveness of silk proteins, as reflected in reproducible structural transitions across concentrations (Koeppel & Holland, 2017).

The study's strength lay in both the experimental control and the analytical approach. Model selection, fitting strategy, and decomposition of overlapping scattering features transformed complex time-resolved data into interpretable structural and kinetic insights.

Multimodal Monitoring with the NUrF Platform

During pH-driven RSF assembly, separating temporal and structural analysis would obscure transient intermediates evolving across both time and length scale. The NUrF platform addresses this limitation by enabling its coordinated tracking. This revealed transient intermediates that support a pathway involving kinetically distinct structural checkpoints, separated by reproducible transitions. Ordered

domains appeared only after thresholds in compaction, correlation, and solvation were reached. These features represent mesoscale (10–100 nm) precursors to the β-sheet ordering of Silk II.

The NUrF experiments generated an extensive time-resolved dataset integrating several kinetic dimensions. SANS profiles were collected every five minutes, fluorescence every 2.5 minutes, and pH was monitored offline every six seconds, over runs lasting up to 780 minutes. Although the article primarily reports the resulting structural and molecular insights, a substantial part of the scientific challenge lay in designing an analytical workflow capable of handling these evolving, heterogeneous datasets. To address this, each probe was first analysed independently before aligning the data on a common time axis.

Analytical Workflow

Time-resolved SANS fitting is inherently difficult because the observed sample's structure changes continuously, and no known single scattering model applies across all time points. The presence of structurally distinct stages rendered global fitting unstable. This complexity was addressed using a hierarchical fitting strategy implemented in Python.

Model selection was guided by dominant features in each q-region, combining multiple scattering functions in a two-step workflow. Simplified models were first used to stabilise parameter space. These estimates then seeded full-range fits to avoid divergence or trapping in local minima. Parameters were linked sequentially across time, allowing each curve to provide initial estimates for the next and ensuring smooth, physically consistent trends. Only correlation features, $S(q)$, could be reliably extracted, as form-factor information ($P(q)$) became masked by intrinsic polydispersity and network heterogeneity (Stribeck, 2007).

The UV data presented further complications: collected UV absorbance traces saturated rapidly, and a file-saving error rendered them unusable. This led to repurposing the 450 nm fluorescence intensity profiles as a turbidity proxy. At this wavelength, the detected signal is dominated by scattered excitation light and therefore provides a reliable measure of sample opacity during aggregation.

Model-free SANS traces, fitted parameters, pH monitoring, and fluorescence intensity signals corresponding to turbidity and ThT emission (proportional to β-sheet development) were integrated under a common time axis. This alignment enabled direct comparison of structural and spectroscopic evolution, as illustrated in Figure. 7.

The evolution of β-sheet was further validated by offline FTIR measurements, where kinetic traces at 1620 and 1695 cm^{-1} provided independent markers of emerging ordered structure. These FTIR measurements confirmed the onset and progression of β-sheet formation detected in the NuRF analysis and ensured

consistency across the scattering and spectroscopic datasets.

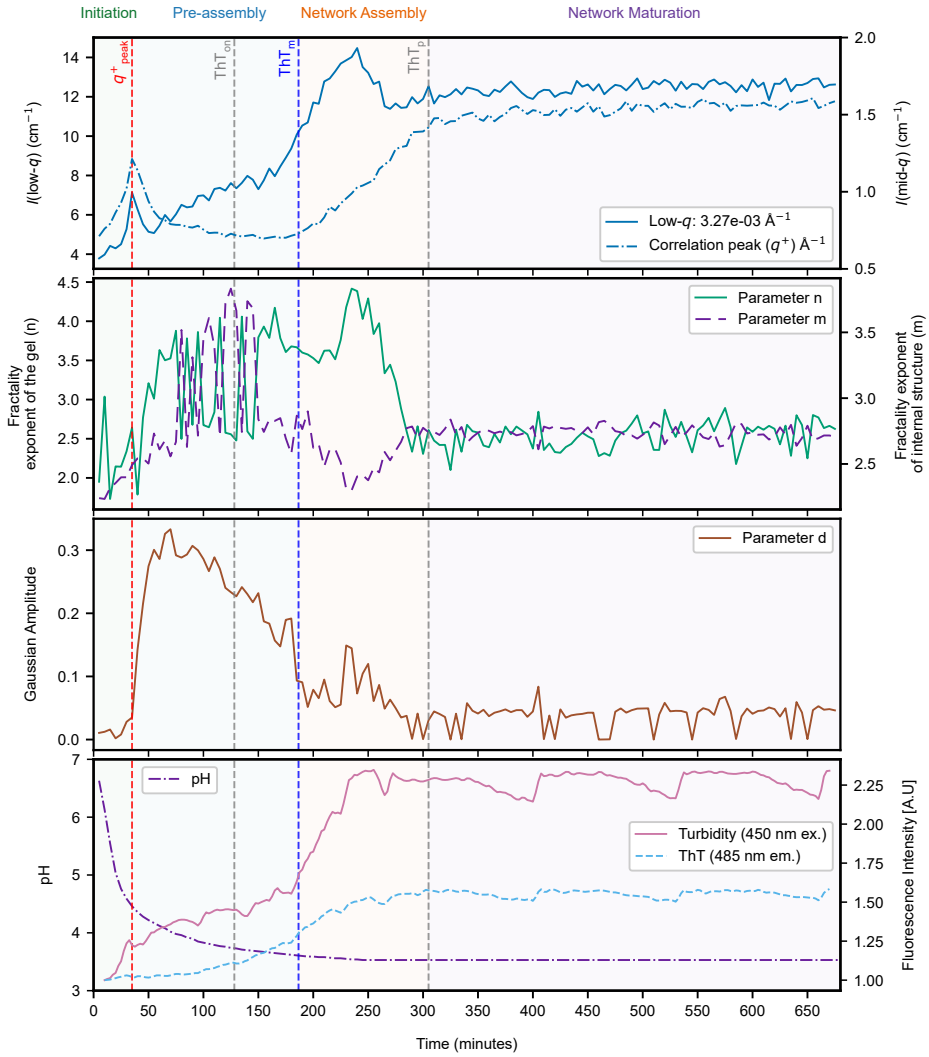


Figure 7: Time-resolved NUrF analysis of RSF gelation at 10 mg mL^{-1} with 1% (w/v) GdL. Top: model-free SANS metrics showing low-q intensity (solid blue) and correlation-peak intensity $I(q^*)$ (dashed dot blue). Middle panels: structural parameters from hierarchical model fits (network fractal exponent n, internal-structure exponent m, and Gaussian amplitude d). Bottom: pH, turbidity, and ThT fluorescence. Vertical dashed lines mark key kinetic points, and shaded regions indicate the four gelation stages. Adapted from Figure 3 in *Paper I*.

Deconvolving Structural Components via SANS MCR-ALS

The power of the NUrF platform is further realised in the subsequent component analysis. Multivariate curve resolution–alternating least squares (MCR-ALS)

was applied to the time-resolved SANS and fluorescence datasets to disentangle overlapping scattering signatures arising during gelation (Jaumot et al., 2015; Tauler et al., 2009). This approach resolved the data into discrete, physically interpretable components, reflecting mesoscale correlations and concurrent molecular ordering during network formation.

A recognised challenge of MCR-ALS is rotational ambiguity, where multiple decompositions may satisfy the data unless sufficient constraints are available. In this system, the simultaneous evolution of SANS and fluorescence kinetics provided such constraints, as both probes follow the same timeline while reporting on different features of the assembly process.

Non-negativity and block-coupling further guided the decomposition toward physically meaningful solutions. These constraints enforce realistic intensities and link SANS and fluorescence contributions to the same underlying species. Unimodality was particularly important, as it imposed kinetically plausible behaviour. This constraint allowed each component to emerge and decay only once during gelation, consistent with an ordered assembly process.

The resulting decomposition revealed a reproducible sequence of processes: the decay of soluble coils, the transient emergence of correlated intermediates, and the accumulation of a β-sheet-rich network consistent with gel maturation (Figure 8). Together, these components provided a coherent, kinetic view of RSF assembly that complemented the model-based analysis and clarified how mesoscale correlations and molecular ordering unfold during acid-triggered gelation.

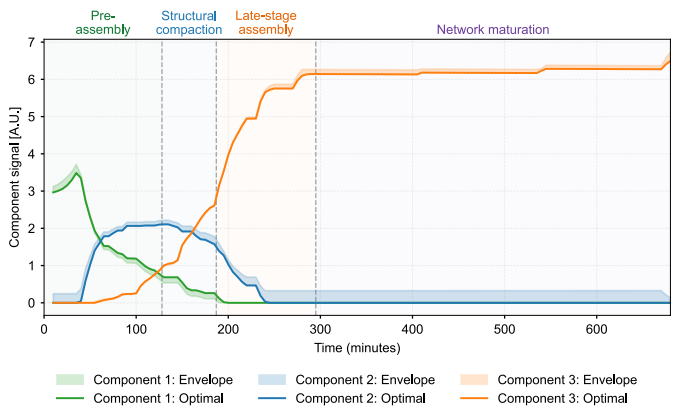


Figure 8: Multivariate component profiles obtained from MCR-ALS. The temporal evolution of three kinetic species during RSF gelation is shown. Shaded regions indicate the main stages of assembly: Pre-assembly, Structural compaction, Late-stage assembly, and Network maturation. Adapted from Figure 5 **Paper I.**

Comparison to MeOH-Triggered Assembly

To contextualise the biological relevance of pH-driven assembly, an alternative assay using methanol (MeOH) was performed to induce RSF fibrillation. Unlike the gradual acidification pathway, MeOH acts primarily as a dehydrating agent. Rapid removal of bound water collapses protein chains and triggers almost immediate β-sheet aggregation. This effectively bypasses intermediate correlation states observed under slow acidification and produces a visibly denser, kinetically arrested network.

Representative 1D scattering profiles (Figure. 9) illustrate these differences. MeOH-induced gels display a strong low-q upturn and lack the mid-q correlation peak that characterises pH-driven assembly, underscoring that dehydration and protonation follow fundamentally distinct structural routes to gelation.

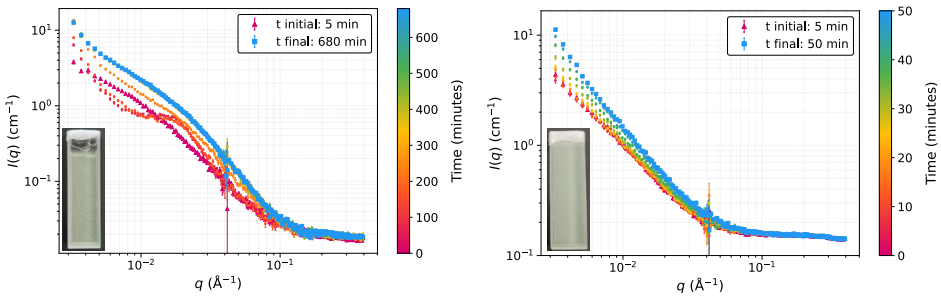


Figure 9: SANS intensity profiles comparing RSF gelation under gradual acidification (left) and methanol treatment (right). Time-resolved scattering curves illustrate the differing kinetics and structural evolution, with image insets showing the corresponding macroscopic samples after each experimental run. The colour scale indicates elapsed time after initiating gelation. Adapted from **Paper I**.

Solvent Reorganisation and Contrast Effects

Using the principles of CV-SANS, we conducted a contrast study external to **Paper I**. Under the same NUrF measurement conditions, we examined how the scattering profile of RSF responded to changes in solvent contrast.

Figure. 10 shows that at the reported solvent to protein contrast-match out condition (42 % D_2O), the initially featureless scattering (red triangles) confirmed homogeneous hydration of fibroin (Banc et al., 2016; Gilbert, 2019). As gelation progressed, a low-q upturn emerged, reflecting large-scale density fluctuations (blue squares). This contrast is consistent with the release and redistribution of H_2O originally confined within protein cavities. Such water populations are likely not readily exchangeable or are weakly bound to protein side chains. Their redistribution is expected to alter the local H/D balance and the effective scattering-length density of the protein phase. Together, these observations indicate that solvent reorganisation accompanies silk network formation and demonstrate the

value of SANS for detecting contrast-sensitive characteristics.

Building on this concept, the next chapter applies CV-SANS to exploit natural contrast differences in a mixed pea protein–carbohydrate system, providing complementary insight into solvent–matrix interactions in a heterogeneous system.

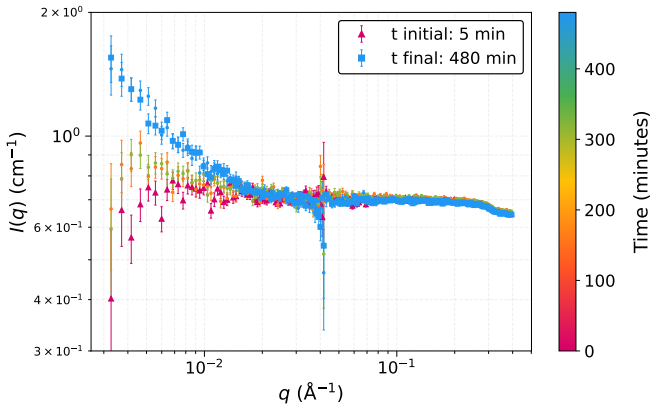


Figure 10: Protein contrast-matched SANS of RSF during GDL-induced gelation. Time-resolved scattering curves recorded at the protein match point (42 % D_2O) show the evolution of a 10 mg mL^{-1} RSF solution during acidification by 1 GDL in the NUrF environment. The colour scale indicates elapsed time after initiating gelation.

Structural Pathways under Acid Vapour Exposure

Acid vapour exposure provides a means to probe silk fibroin reorganisation within a confined and highly concentrated protein matrix. In contrast to the dispersed chains examined in **Paper I**, cast RSF films represent an amorphous yet densely packed starting state with restricted molecular mobility. This configuration more closely approximates the spatial constraints encountered in the silk gland and allows reorganisation to be examined under conditions where diffusion and large-scale rearrangements are limited.

The experiment begins by equilibrating these films under controlled humidity. This introduces hydration without inducing premature structural rearrangement. Once the acid vapour is applied, the pH change modulates local interactions within this pre-condensed network. This sequence enables tracking of hydration, mesoscale organisation, and crystalline polymorphism as they evolve during treatment.

SAXS, WAXS, and Infrared Spectroscopy

The integration of SAXS, WAXS, and infrared spectroscopy provides a multiscale view of the assembly process. SAXS data revealed a two-stage evolution during treatment. During the initial swelling phase, the relative invariant, Q', decreased as the solvent filled the matrix pores. Upon acidification, Q' rose significantly. This increase suggests the emergence of well-defined internal domains as the protein chains reorganise.

Infrared measurements using deuterated water vapour (D_2O) were used to track hydration kinetics. Deuteration increased sensitivity to hydrogen-bond exchange by reducing overlap with H_2O vibrational bands.

Unfortunately, the thickness of the films led to saturation in the amide I–III regions. As a result, the IR data primarily reported changes in hydration and molecular accessibility, providing a qualitative assessment of matrix plasticisation rather than a quantitative measure of secondary structure.

WAXS provided the clearest insight into how these rearrangements translate into crystalline order. Contrary to the common expectation that acid exposure promotes rapid β-sheet formation, the results presented in **Paper II** show that acid vapour stabilises Silk I motifs over extended timescales, without immediate conversion to a Silk II-dominated state (Figure 11).

Rather than driving rapid β-sheet ordering, vapour-acid treatment stabilises intermediate crystalline packing and allows metastable states to persist. This behaviour is mechanistically significant, as favouring the Silk I polymorph delays irreversible Silk II formation and postpones premature aggregation. These findings suggest that specific hydration and concentration conditions may help preserve metastability prior to spinning.

The mechanism by which the system ultimately progresses toward Silk II remains unresolved. Additional mechanical inputs during downstream processing, such as shear and elongational flow, are likely required to complete this transition and would be an interesting direction for future investigation.

Comparison to Progressive pH-Induced Gelation

Assembly in the film matrix diverges clearly from the solution-phase behaviour observed in **Paper I**. In solution, acidification drives the system toward β-sheet-rich networks. In the film matrix, however, the process is redirected toward the metastable Silk I state. This comparison highlights how the physical environment, specifically concentration effects, dictates the assembly route.

This divergence also underscores the importance of probe sensitivity. ThT fluorescence, used in the solution studies, monitors the general emergence of β-sheet environments. It cannot, however, distinguish between amorphous and

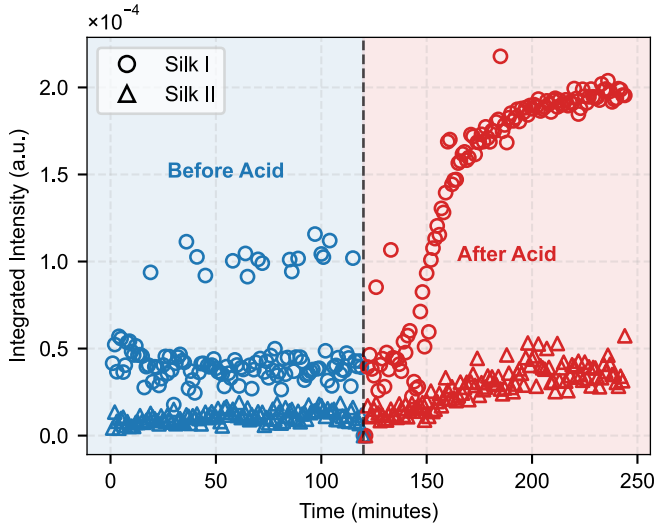


Figure 11: Time-resolved evolution of Silk I and Silk II signatures during acidification, quantified from WAXS data. Integrated scattering intensities corresponding to Silk I (circles) and Silk II (triangles) are shown as a function of time. The dashed line marks the onset of acid exposure, separating the pre-acid (blue) and post-acid (red) regimes. Acidification induces a pronounced increase in the Silk I signal, while Silk II develops more gradually, indicating distinct structural responses during network evolution. Adapted from **Paper II.**

fully crystalline states. WAXS provides the resolution required to identify specific crystalline polymorphs.

These studies relied not only on experimental control but on rigorous data exploitation. Time-resolved, multimodal datasets required deliberate model selection and hierarchical fitting strategies to extract physically consistent trends. Beyond protein structure alone, the analyses revealed that changes in solvation accompany and influence assembly, with contrast-sensitive SANS providing direct access to these effects.

This focus on solvent–matrix interactions motivates the transition to the next chapter. Rather than a purified silk system, the work now turns to a naturally heterogeneous pea protein–carbohydrate matrix, where intrinsic contrast between components can be exploited directly. Applying the contrast principles established here enables interrogation of how hydration and matrix organisation are coupled in complex, minimally processed protein-based materials.

Chapter 4 – Co-Gelation of Pea Proteins and Carbohydrates

From Single-Source Matrices to Multi-component Coupling

The time-resolved studies of RSF established that SANS can capture structural features that are not readily accessible to other techniques. These experiments focused on a single protein system. Measurements at different solvent contrasts showed that changing the solvent SLD alters which structural contributions are visible during assembly.

This methodological insight provides the bridge to the current chapter and **Paper III**. While the RSF system is dominated by the assembly of a single protein matrix, pea protein extracts are inherently heterogeneous and provide intrinsic contrast between protein and carbohydrate components. This contrast enables direct observation of how these components organise before and after gelation in minimally processed extracts, without requiring reconstitution or isotopic labelling. Contrast-variation SANS thus offers a uniquely suited structural probe for resolving protein–carbohydrate organisation in systems where the underlying interactions are not known.

Extraction Conditions and Starch Retention

Paper III examines pea protein extracts obtained through salt (SE) and alkali (AE) extraction, as shown in Figure. 12. These truncated workflows omit the protein precipitation step and therefore retain small but significant amounts of co-extracted carbohydrates, primarily starch, which are typically removed in commercial isolates (Kornet et al., 2021).

The SE method employs milder conditions to preserve the near-native state of the protein subunits, whereas the AE route involves a higher pH that may promote structural rearrangement (Asen et al., 2023). Compositional analysis in the study confirmed higher protein content in SE extracts and greater starch retention in AE extracts. In the context of this chapter, these differences establish well-defined starting points for examining how initially decoupled protein- and carbohydrate-

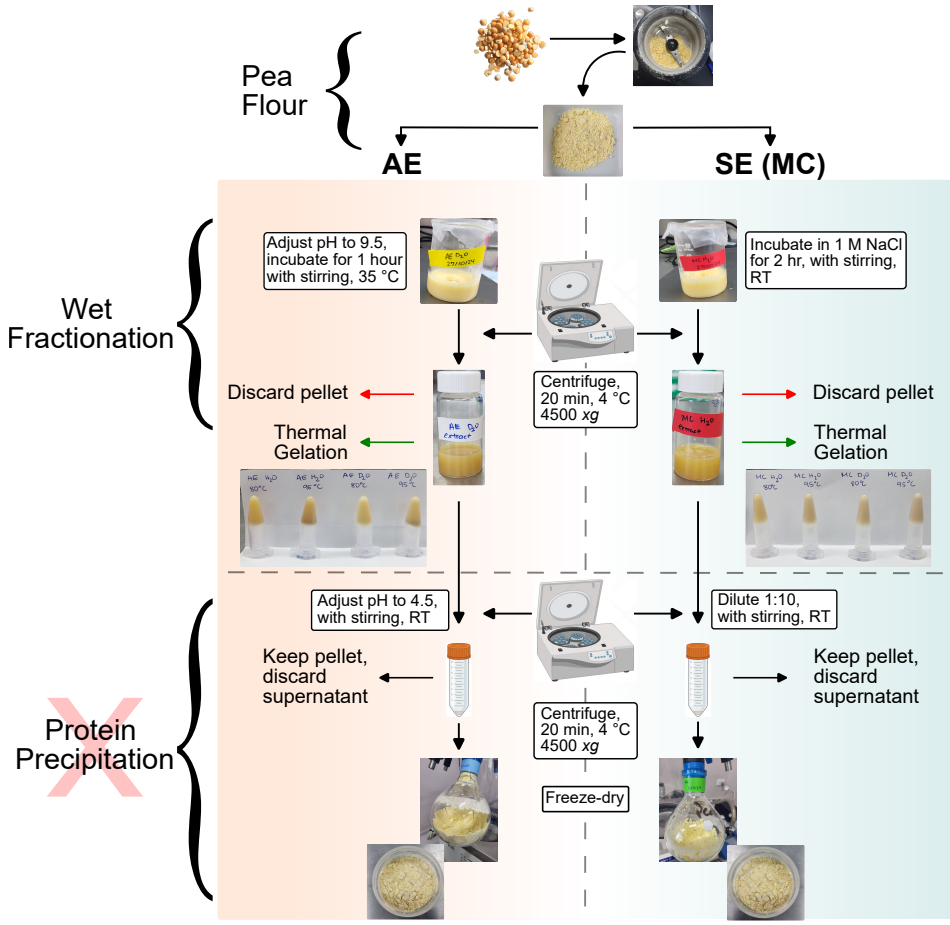


Figure 12: Schematic overview of the pea protein extraction workflows. Both alkaline extraction (AE) and salt extraction (SE/MC) pathways are shown through the wet-fractionation and thermal-gelation steps. Full extraction would typically proceed with protein precipitation, followed by optional drying.

rich populations reorganise into either a coupled or structurally partitioned gel network.

While rheological and calorimetric data from the broader study describe the kinetic and thermal signatures of gelation, they do not resolve how protein and carbohydrate components reorganise within the network. This section, therefore, focuses on the structural insight provided by CV-SANS and USANS. The compositional differences between AE and SE extracts establish distinct starting conditions for testing how protein–carbohydrate coupling and mesoscale connectivity emerge in the gel state.

SANS with Contrast Variation for Network Analysis

CV-SANS provided a direct means to assess how the protein and carbohydrate phases redistribute before and after thermal gelation. By tuning the solvent SLD across 0–100 % D_2O, scattering contributions from each phase could be selectively emphasised or suppressed, enabling a more structure-informed interpretation of reorganisation in the mixed system. Literature values place the approximate contrast-match points of the protein and carbohydrate (polysaccharide) fractions at 42 % and 30 % D_2O, respectively (Gilbert, 2019). These values guided the selection of solvent contrasts and provided a reference for interpreting the low- and mid-q minima observed in the scattering profiles.

The broad spatial range accessed by these SANS measurements, extending from $q \sim 6 \times 10^{-4}$ to ~ 0.75 Å^{-1}, favoured a strategy of robust parametrisation over detailed structural modelling. As in the RSF studies, the emphasis was on characterising network correlations rather than assigning unique particle shapes. In the pea system, contrast-dependent scattering profiles were analysed for the solution and gel end states using quadratic fits across the contrast series to extract apparent match points associated with protein- and carbohydrate-dominated contributions. Comparison of these match points between solution and gel states provides a structural measure of whether the two components remain decoupled or form a coupled network upon gelation.

In the solution state, distinct minima were observed in the low- and mid-q regions, indicating that carbohydrate-rich and protein-rich structures contribute independently to the scattering signal, respectively. Upon gelation, these minima converged, most clearly in the SE samples (Table 2), consistent with the formation of a structurally integrated network in which both components contribute to the same mesoscale correlations. In contrast, AE gels retained separated minima, indicating incomplete coupling and the persistence of structurally distinct protein- and carbohydrate-rich domains. This scattering-based distinction directly reflects differences in network organisation that are not accessible from bulk measurements alone.

Table 2: Calculated contrast match points (in % D_2O) obtained from quadratic fits of low-q (1 – 2 × 10^{-3} Å^{-1}) and mid-q (3 – 9 × 10^{-2} Å^{-1}) weighted mean intensities for both SE and AE solution and gel CV-SANS analysis. Values represent mean ± standard error from the fits.

Sample	Low-q (%)	Mid-q (%)
SE Solutions	13.6 ± 0.3	40.9 ± 0.1
SE Gels	33.4 ± 0.2	34.4 ± 0.1
AE Solutions	12.9 ± 0.2	39.2 ± 0.1
AE Gels	22.1 ± 0.2	34.2 ± 0.1

Extension into the ultra-small-angle range ($q \sim 3.5 \times 10^{-4}$ Å^{-1}) provided

additional insight into the hierarchical organisation of the final gel networks. USANS revealed that SE gels exhibit persistent density correlations extending to length scales of up to 2 μm, consistent with a well-integrated and percolating network. In contrast, AE gels displayed a more truncated structural signature, with scattering dominated by finite, aggregate-like structures and no detectable long-range correlations. These differences indicate that, while both systems form gels, only the SE extraction pathway leads to mesoscale connectivity across multiple length scales.

These USANS results complement the contrast-variation findings by linking component-level mixing to the final network architecture. While both protein and carbohydrate fractions contribute to the gel matrix, their spatial distribution and degree of integration depend strongly on the extraction pathway. The analysis shows that structural hierarchy in these systems is not an intrinsic property of the protein alone. It emerges from the extent of phase coupling permitted by the processing history.

A natural extension of this work would be to perform time-resolved CV-SANS studies on comparable systems. This would directly map the evolution of phase-specific correlations onto the development of mechanical properties. Such an approach extends the kinetic analysis established in the RSF studies to systems with intrinsic component contrast.

Chapter 5 – Solution-State Architecture of Native Cylindriform Spidroin (TuSp1)

The challenge with multi-domain proteins

Chapter 1 introduced hierarchical assembly as an emergent outcome of protein architecture, processing conditions, and biological function. This chapter narrows the focus to the solution state architecture of Tubuliform silk protein 1 (TuSp1), the dominant protein component of cylindriform silk. It explores which structural features are intrinsically encoded at the single-molecule level.

Multi-domain proteins such as TuSp1 pose a fundamental challenge. Linking many domains into a single polypeptide greatly increases the number of possible intramolecular contacts. Many of these contacts are non-native, often leading to frustrated folding landscapes where productive folding competes with misfolding and aggregation. In the cell, folding is supported through co-translational synthesis, chaperone activity, and post-translational regulation (Rajasekaran & Kaiser, 2024). The relative contribution of these mechanisms varies across systems. The structural conformation of large multi-domain proteins is therefore difficult to predict and stabilise. Outside the cellular environment, these supports are absent, increasing susceptibility to misfolding and aggregation.

Spider silk proteins offer a particularly informative system for examining these conformational challenges. As large, repetitive multi-domain chains, they must suppress premature intermolecular association and misfolding to remain soluble at high concentrations prior to the onset of rapid fibre assembly.

Structural background from previous work Comparative solution-state studies have shown that native spider proteins adopt non-globular, elongated conformations, while differing in the extent of local structural order (Greving et al., 2020). Dragline and flagelliform silks are dominated by intrinsically disordered conformations. By contrast, TuSp1 retains pronounced local secondary structure, most notably substantial α-helical content, while remaining globally extended. In this context, local folding tunes chain stiffness and effective persistence length. It does so without driving the protein into a compact, globular tertiary structure.

Multi-domain cylindriform spidroin can thus be regarded as assemblies of folded domains linked by long, flexible segments. These generate conformational heterogeneity (Bernadó et al., 2007; Lin et al., 2009). Most experimental observables are ensemble-averaged over many possible domain arrangements. Structural characterisation must therefore (i) constrain the geometry of the individual folded domains and (ii) explicitly account for global conformational variability (Bernadó et al., 2007).

TuSp1 is particularly well-suited to SAXS-guided modelling because its architecture comprises defined folded terminal domains connected by repetitive regions. The domains can be treated as rigid structural units, while global architecture arises from flexible inter-domain connectivity (Lin et al., 2009). This hierarchical organisation provides structural anchors that reduce modelling degeneracy compared to fully disordered systems.

The availability of extended tubuliform spidroin sequences in recent years further clarified domain–repeat organisation. Earlier sequence data were fragmentary due to the highly repetitive nature of silk genes. More complete genomic sequences published in Wen et al., 2017 and Wen et al., 2020 enabled more accurate construct design and ensemble modelling of defined domain–repeat architectures.

The SAXS data analysed in this chapter were originally acquired as part of the study by Greving et al., 2020. While their work established the global size and shape of native cylindriform spidroin, it did not resolve how the individual folded domains are arranged along the protein chain or how they sample different relative positions in solution.

This analytical approach is extended in **Paper IV** and forms the basis of the present chapter. Here, the same SAXS data were re-analysed using SAXS-guided structural modelling to resolve how locally folded units are organised within the extended solution architecture of TuSp1. The protein was described as a modular chain consisting of 18 folded repeat domains flanked by N- and C-terminal domains and connected by flexible linkers (NT-$(RP1)_{17}$-$RP2$-CT). This reflects the number of structurally distinguishable units accessible within the SAXS length scales used here, rather than the biochemical repeat count inferred from the sequence. However, sequence-based analyses have estimated approximately 21 repeats in native TuSp1 (Lin et al., 2009).

SAXS Measurements and Fitting Strategy

SAXS was used to obtain a reduced description of the size, shape, and conformational heterogeneity of native TuSp1 in solution. As an ensemble-averaging technique, SAXS reports global properties that are sensitive to

conformational flexibility and overall molecular organisation (Jeffries et al., 2021). By fixing the geometry of each folded domain with homology models, the modelling effort was restricted to the flexible linker regions and the relative positioning of the domains (Waterhouse et al., 2018). The resulting SAXS fits demonstrate that the experimental data are consistently captured within this computational approach (Figure 13a). The data were then analysed using two complementary workflows implemented in the ATSAS (Advanced Tools for Small-Angle Scattering) software suite:

1. **COmplexes with RAndom Loops CORAL rigid-body modelling:** This tests whether a single representative arrangement of the domains can capture the dominant features of the scattering curves.

2. **Ensemble Optimisation Method (EOM):** This determines whether a distribution of conformations drawn from a broad pool is required to reproduce the experimental data (Manalastas-Cantos et al., 2021).

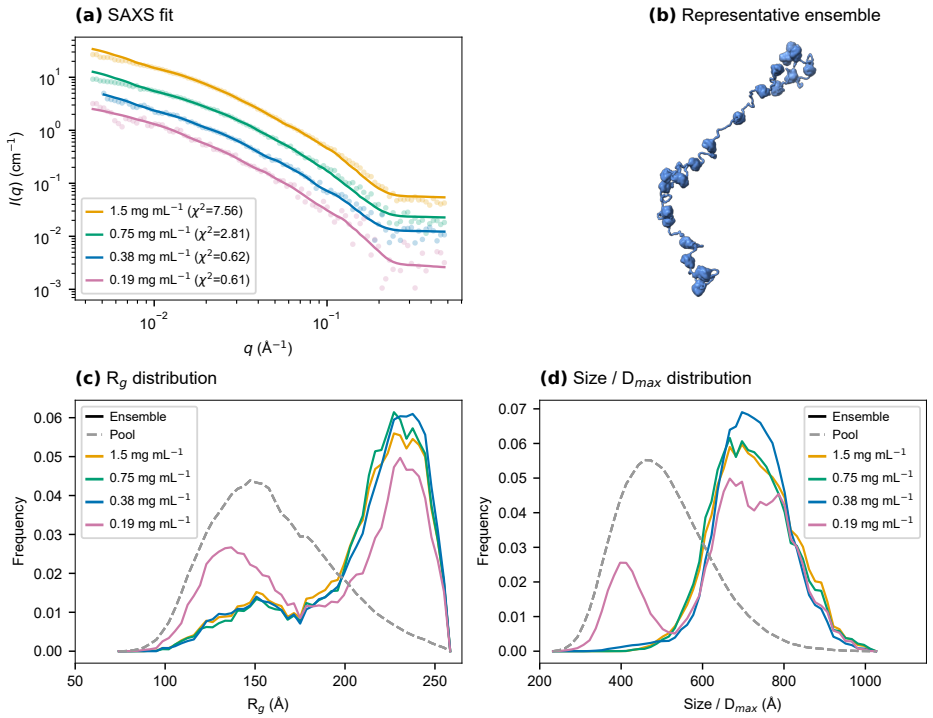


Figure 13: SAXS-guided ensemble modelling of native cylindriform spidroin TuSp1. (a) SAXS fits obtained using ensemble optimisation (EOM) across a concentration series. (b) Representative conformations from the selected ensembles illustrate an elongated, modular architecture. (c) Radius of gyration (R_g) distributions for the selected ensembles compared with the initial conformational pool. (d) Corresponding size (D_{max}) distributions. Reproduced from **Paper IV**.

Structural features that are consistent across both analyses are taken as robust characteristics of the native protein. Across the concentration series, the SAXS data consistently support a non-globular, modular architecture. Representative ensemble conformations show elongated chains composed of discretely folded domains connected by flexible linkers (Figure 13b).

The EOM-selected ensembles are enriched in larger radius of gyration (R_g) and larger maximum particle dimension (D_{max}) values compared to the initial random pool (Figure 13c and d). This shift indicates that individual TuSp1 molecules predominantly sample extended conformations, with their global single-molecule architecture governed by flexible inter-domain connectivity. Consequently, this integrated modelling workflow provides the structural foundation for the subsequent coarse-grained simulations and functional interpretations.

Simulated Models and Folded Conformations

While SAXS analysis constrains the global architecture of TuSp1, coarse-grained simulations provide a physical rationale for how that architecture emerges from the underlying sequence. These simulations are not meant to generate a unique native ensemble. Instead, they test whether the experimentally inferred organisation arises naturally under realistic solution conditions (150 mM NaCl at pH 5, 7, and 9) based on the sequence.

Simulations were performed using the coarse-graining approach to liquid–liquid phase separation via an automated data-driven optimisation scheme (CALVADOS3) on the full 18 repeat domain construct of TuSp1. In this model, each amino acid is represented by a single bead whose interactions depend on residue identity and the specified solution conditions. No pre-defined rigid domains are imposed (von Bülow et al., 2025).

The simulated size metrics mirror the SAXS-derived results (Figure 14):

1. **Radius of gyration,** R_g: Distributions are broad at all pH values, indicating heterogeneous ensembles of extended conformations. The pH 7 ensemble is the most compact and narrowly distributed. At pH 5, the distribution broadens while remaining largely overlapping with pH 7. At pH 9, the distribution shifts toward larger values, indicating expansion of the ensemble under basic conditions.

2. **End-to-end distance,** R_{ee}: The pH 7 ensemble shows the smallest end-to-end distances, while pH 5 exhibits increased heterogeneity and pH 9 shifts toward larger values.

3. **Orientational correlation functions**: correlations decay rapidly with increasing lag time, indicating limited long-range orientational coupling between domains.

4. **Contact probability**: intramolecular contacts are dominated by short sequence separations, with a rapid decay in contact probability at longer separations, indicating that long-range contacts remain rare.

Rapid decay of orientational correlations and the dominance of intramolecular contacts between residues close in sequence indicate weak long-range coupling between domains. This supports an extended global architecture across all pH values.

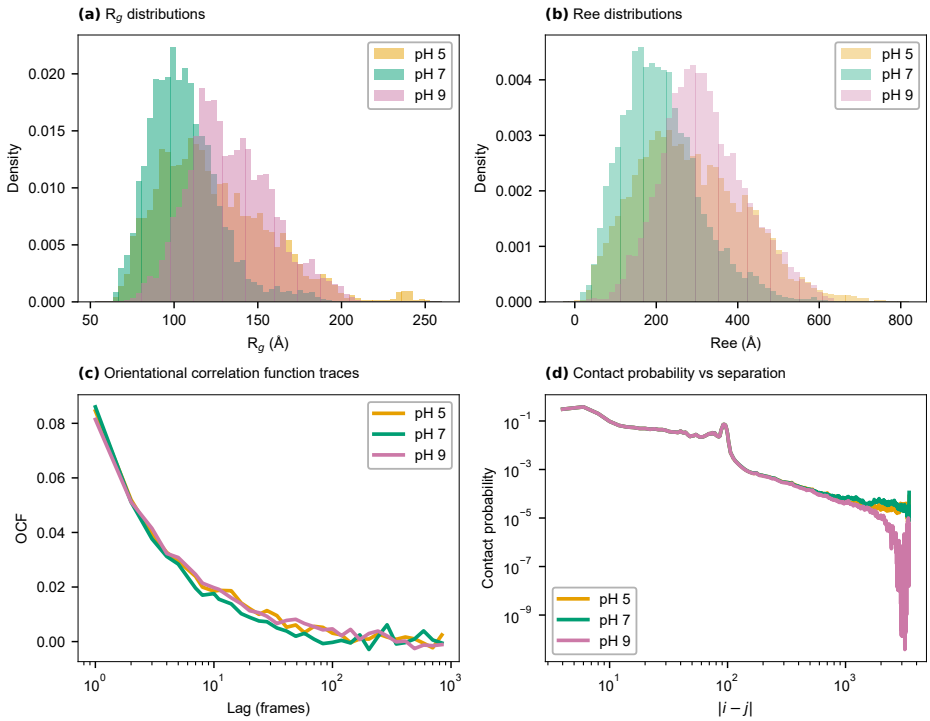


Figure 14: Sequence-based coarse-grained simulations of native cylindriform spidroin TuSp1. Summary of CALVADOS 3 simulation results at pH 5, 7, and 9 (150 mM NaCl), illustrating (top) simulated scattering profiles and global size distributions and (bottom) orientational correlations and residue–residue contact probabilities as a function of sequence separation. The simulations provide a physical interpretation of the extended, weakly compact solution architecture inferred from SAXS-based modelling. Adapted from **Paper IV**.

Although extensive disorder in linker regions can promote global compaction in other aggregation-prone spider silk proteins (Askarieh et al., 2010), such behaviour is not observed here for TuSp1. Instead, the present results show that

TuSp1 adopts a predominantly extended architecture in solution. This architecture arises from modular domain organisation and limited long-range intramolecular coupling. Extending CALVADOS3 simulations to assemblies containing multiple copies of the 18-repeat construct would further elucidate how such architectures pack and interact in a crowded solution.

The approaches used in this study constrain the global architecture and overall flexibility of native cylindriform spidroin (TuSp1), but they do not resolve how stability is distributed across individual domain–linker junctions. Furthermore, it remains unclear how robust this architecture is in its native-like environment. These open questions motivate the transition to Chapter 6, where a reduced recombinant spidroin is used to probe domain connectivity and junction stability directly.

Chapter 6 – Recombinant Cylindriform Spidroin

Understanding multi-domain proteins using a mini-spidroin

Multi-domain proteins must balance the stability of folded regions with the flexibility of the linkers that connect them. In spidroins, this balance is further shaped by the physicochemical constraints of storage and spinning.

As suggested in Chapter 1, the short and narrow spinning duct of cylindriform silk may enable a more direct assembly route. This route is further supported by the predominantly α-helical solution state of TuSp1. Together, these features may reduce the need for the extensive conditioning steps observed in dragline silk and RSF. Consequently, TuSp1 serves as a contrasting case study. What remains unclear is how much of this behaviour is encoded in the TuSp1 architecture itself, and how much depends on the stabilising environment of the gland.

Paper V addresses this question by evaluating a 62.5 kilodalton (kDa) recombinant mini-spidroin derived from TuSp1. The construct was originally intended as a platform for generating phase and stability diagrams. These diagrams were designed to map environmental triggers and structural features leading to phase transitions. However, persistent instability during expression and purification prevented systematic exploration across conditions. The following sections, therefore, treat the mini-spidroin as a diagnostic probe of architectural robustness rather than as a complete model system. This framing allows us to assess which architectural features remain robust outside the gland and which may require explicit engineering for stable recombinant production.

Miniaturising TuSp1: construct design in brief

TuSp1 is a massive ($\approx$ 434 kDa), multi-domain spidroin. Its size and repetitive sequence render the full architecture difficult to produce and manipulate (Tian & Lewis, 2006; Wen et al., 2020). To address these constraints, a 62.5 kDa mini-spidroin was designed. It contains a single copy of the N-terminal domain (NT), repetitive protein domains 1 and 2 (RP1 and RP2), and the CT domain, joined by their native linker regions (Figure 15). This architecture reduces the repetitive core

to a scale compatible with recombinant production while preserving the canonical domain arrangement of cylindriform silk.

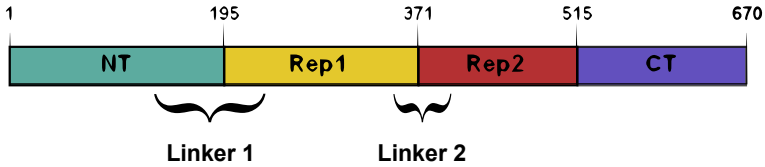


Figure 15: Schematic representation of the minimal 62.5 kDa TuSp1. The mini-spidroin comprises a single copy of each unique domain (NT, RP1, RP2, and CT) connected by their native linker regions. This design preserves the canonical architecture of cylindriform spidroins while reducing the repetitive core to a length compatible with recombinant expression.

A range of fusion tags ($His_{6/8}$, SUMO, mEGFP, and Gal8N) were placed at either terminus to aid expression, solubility, and purification. In parallel, truncated variants (e.g., NT-RP1, RP2-CT) were generated to isolate individual domain-linker regions. Together, these constructs enabled a systematic mapping of which architectural elements remain stable in heterologous hosts and which junctions are most vulnerable to degradation.

Expression and Purification Challenges

Expression screening

The expression strategy aimed to identify which heterologous systems could tolerate the combined domain architecture of the mini-spidroin. Screening across bacterial (*Escherichia coli*), yeast (*Pichia pastoris*), and insect cells provided a comparative view of TuSp1 construct performance.

In practice, meaningful expression was only achieved in bacterial hosts. While the more stringent environments of yeast and insect systems yielded low or undetectable expression, screening within different *E. coli* strains revealed a clear trade-off between total yield and structural integrity. Both *E. coli* Tuner (DE3) and Rosetta (DE3) showed high protein production. This was accompanied by substantial proteolytic degradation despite the use of protease inhibitors, as well as co-purification of the DnaK chaperone. This indicates that the host's standard folding machinery is insufficient to maintain the integrity of the multi-domain construct during synthesis.

However, in *E. coli* ArcticExpress, the mini-spidroin was produced in a more stable and recoverable form. This system relies on low-temperature induction (10–12 °C) and co-expression of the cold-active chaperones Cpn10 and Cpn60. The requirement for reduced thermal energy and specialised chaperones suggests that domain junctions within the mini-spidroin are particularly sensitive during folding.

Under these conditions, proteolytic cleavage and internal fragmentation observed in Tuner and Rosetta were largely avoided.

As an exploratory step toward downstream SANS studies, protein production was carried out in *E. coli* BL21 in a bioreactor (1 L working volume). This enabled controlled, high-density growth for protein per-deuteration using D_2O-based media. Even under the regulated conditions of the bioreactor, the multi-domain construct was partitioned predominantly into the insoluble fraction. As a result, it was unsuitable for biophysical analysis. Full optimisation of bioreactor parameters was beyond the scope of this study. Further refinement of the expression protocol could therefore result in improved protein recovery.

The recurring instability observed across expression hosts reflects intrinsic properties of the construct, while also highlighting limitations of the individual host systems. The same failure modes recurred across hosts with distinct folding and quality-control constraints. Detailed results are presented in **Paper V**. Here, the emphasis is on how these recurring patterns guided the transition from broad screening to a viable production and purification strategy.

Purification

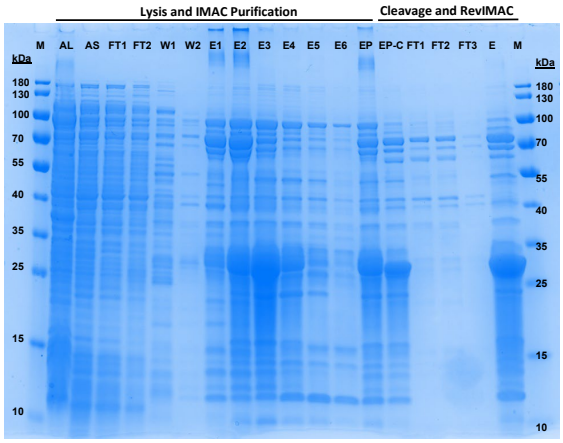
Purification was built directly on the insights gained from the expression trials. Rather than serving as a routine route to homogeneous material, purification outcomes further exposed construct instability.

Purification progress was routinely monitored by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) and Western immunoblotting (WB). This followed immobilised metal affinity chromatography (IMAC), Tobacco Etch Virus (TEV) protease cleavage, and reverse IMAC.

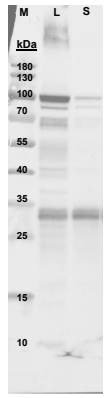
Representative purification analysis (SDS-PAGE and WB) of the fusion protein, His–mEGFP–NT–RP1–RP2–CT, from Tuner (DE3) lysates revealed extensive heterogeneity. The fusion protein co-eluted with multiple lower-molecular-weight species, indicating substantial proteolytic fragmentation (Figure 16). This demonstrates that instability is present from the start and persists throughout purification.

To assess whether this instability could be mitigated under more supportive expression conditions, purification was tested using protein expressed in ArcticExpress *E. coli* (Figure 17).

SDS–PAGE analysis was performed on the 62.5 kDa NT–RP1–RP2–CT construct following size-exclusion chromatography (SEC). A dominant band migrates slightly above the 55 kDa marker. Peak recovery is observed in fractions C2–C8, consistent with the expected molecular weight of the target construct. Higher-molecular-weight contaminants, including DnaK (≈70 kDa), elute predominantly in earlier fractions (B4–B12). This indicates effective



(a) IMAC purification workflow.



(b) Western blot of expression samples.

Figure 16: Purification behaviour of His–mEGFP–NT–RP1–RP2–CT expressed in *E. coli* Tuner (DE3). (a) SDS–PAGE of IMAC, TEV cleavage, and reverse IMAC samples. M, molecular weight marker (PageRuler^{TM} Prestained Protein Ladder (Thermo Scientific)); L, lysate; S, soluble fraction; FT, IMAC flow-through; W, wash; E, imidazole elutions; EP, pooled elutions; Clea, sample after TEV cleavage; reIMAC FT/E, reverse IMAC flow-through and elution. The target protein (≈ 92 kDa) co-elutes with lower-molecular-weight species indicative of proteolytic fragmentation. **(b)** Western immunoblot of a representative expression samples. Reproduced from **Paper V**.

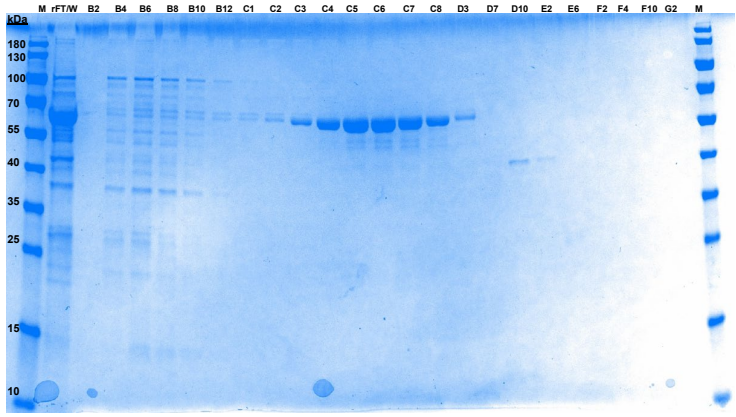


Figure 17: SDS–PAGE of SEC fractions for His_6–mEGFP–NT–RP1–RP2–CT expressed in *E. coli* ArcticExpress. SEC fractions collected following IMAC purification are shown. Lane M indicates the molecular weight marker; rFT/W denotes reverse IMAC flow-through and wash fractions. Adapted from **Paper V**.

polishing by SEC following initial capture by IMAC.

In contrast to the persistent heterogeneity observed for protein expressed in Tuner (DE3) and Rosetta (DE3), expression in ArcticExpress E. coli yielded material that tolerated downstream purification. Additional chaperone-

displacement steps were required to remove co-purifying DnaK, but these conditions enabled recovery of intact mini-spidroin suitable for further analysis.

This behaviour contrasts with the stability of full-length cylindriform spidroins under native storage conditions. In the gland, the protein is stored at very high concentrations within a narrowly regulated environment (Lefèvre et al., 2011; Lin et al., 2009). Such conditions are not reproduced during recombinant expression, where dilution and host-mediated folding dominate. While intact mini-spidroin can be produced and recovered, the absence of the full-length context or native stabilising environment renders its junctions particularly vulnerable. Expression in ArcticExpress E. coli provides additional folding support that partially compensates for these instabilities, enabling reliable recovery of intact protein.

Implications for multi-domain spidroin design

Taken together, the expression and purification stages define the practical limits of the current mini-spidroin design. These results identify where cylindriform architectures are compatible with recombinant production and where additional optimisation is required. Three broader conclusions emerge from this analysis:

1. Localised stability. Apparent structural stability is associated with folded regions of the construct. Where instability arises predominantly at tag–domain and domain–domain junctions, which impose the dominant burden on recombinant production.

2. Host buffering: Host systems differ in how much of this burden they can buffer, but none remove it entirely. Consequently, successful expression depends as much on construct design as on host choice.

3. Engineering Requirements: A functional recombinant TuSp1 analogue will likely require engineered linkers or fusion partners that relieve strain at key junctions, rather than further truncation of the native architecture.

When considered alongside earlier chapters, these findings place clear constraints on attempts to recreate cylindriform architectures outside their native context. The results in this chapter highlight that recreating such a pre-organised solution state is non-trivial in recombinant systems. By systematically probing the architectural limits of the mini-spidroin across constructs and host systems, this diagnostic approach defined clear design constraints to guide future cylindriform analogues.

Conclusions

This thesis demonstrates that protein-based materials cannot be understood solely as the outcome of a final triggering event such as acidification, dehydration, or heating. Across silk and pea protein systems, key aspects of material structure are established upstream of macroscopic gelation or fibre formation, arising from solution-state organisation, transient intermediates, and solvent participation.

In reconstituted silk fibroin, time-resolved multimodal measurements reveal reproducible intermediate states with emerging spatial organisation prior to extensive β-sheet network formation. These are pathway-dependent and differ with how environmental conditions are imposed. Gradual acidification and methanol-induced assembly access distinct structural routes, showing that environmental history biases accessible pathways rather than accelerating a common endpoint. The unexpected persistence of Silk I under acid-vapour exposure suggests that controlling intermediate states may delay irreversible structural transitions. This points to a plausible route for maintaining metastable organisation. Across the silk systems studied, the observed behaviour is most consistent with kinetically biased assembly under metastable conditions, rather than equilibration toward a unique thermodynamic minimum.

Contrast-variation neutron scattering provides direct evidence that solvent organisation is coupled to silk gelation. The emergence of mesoscale structure under protein contrast-matching conditions indicates heterogeneous solvent redistribution within the assembling phase. These findings indicate that water is structurally involved in collapse and network formation, rather than acting solely as a passive plasticiser.

The solution-state analysis of native cylindriform spidroin, supported by SAXS-guided simulations, addresses the role of pre-organisation at the single-molecule level. TuSp1 consistently adopts an unusually extended, modular architecture in solution, composed of folded domains connected by flexible linkers. When removed from the native biological context, as in recombinant constructs, junctional connectivity within this architecture is only marginally stable. Environment-dependent stabilisation therefore localises to specific multi-domain arrangements. This instability underscores the importance of biological environments in preserving pre-organised solution states prior to fibre formation.

The pea protein–starch system extends the concept of upstream control to a multicomponent context. Thermal triggering alone does not guarantee

a structurally integrated network, as SE extracts formed a unified protein–carbohydrate gel matrix, whereas AE gels retained more segregated components. Processing history thus acts as a structural variable that constrains which assembly pathways are accessible and which network morphologies can form.

Taken together, this work reframes protein assembly as a pathway-dependent process governed by pre-organisation, transient intermediates, solvent behaviour, and molecular context. Rather than resolving silk or pea protein gelation into a single unified model, the thesis defines experimentally grounded constraints and shows where simplified interpretations break down. In doing so, it establishes transferable concepts that provide a framework for understanding and steering hierarchical assembly in both biomimetic fibres and sustainable food-relevant systems. It also highlights where further mechanistic insight is still required.

Future Studies

The results of this thesis point to several directions for future work aimed at further clarifying how solution-state organisation, solvent behaviour, and molecular context bias hierarchical protein assembly in both fibrous and gel-forming systems. Silk and pea protein networks provide complementary reference systems for extending these approaches and exploring how pathway-dependent assembly emerges under different molecular and processing constraints.

- **Time-resolved contrast-variation SANS across protein systems.** Extending CV-SANS to time-resolved measurements in reconstituted silk fibroin and pea protein–carbohydrate mixtures would allow the evolution of individual phases to be followed directly during assembly. This would enable phase-specific kinetic pathways to be linked to the emergence of mesoscale network structure.
- **Pathway selection under controlled physicochemical perturbations.** The NUrF platform could be extended to probe silk and pea protein assembly under controlled changes in solvent composition and ionic strength, as well as under external mechanical inputs such as shear and extensional flow. Such experiments would enable direct testing of how specific perturbations redirect assembly pathways under both processing-relevant and biologically relevant conditions.
- **Stability and folding in modular silk architectures.** Further exploration of chaperone-supported expression systems for cylindriform spidroins would clarify how folded domains and domain–linker junctions are stabilised outside the native glandular environment. This has the potential to distinguish intrinsic architectural constraints from host-dependent folding effects.
- **Mapping solution-state phase behaviour.** Systematic phase diagrams for spidroins and pea protein systems as a function of concentration, pH, ionic strength, and solvent composition would establish the boundaries between soluble, pre-assembled, gelled, and phase-separated states, providing a structural context for interpreting pathway-dependent assembly outcomes.
- **Integrating scattering and modelling across systems.** Combining scattering experiments with SAXS-guided modelling and coarse-grained

simulations would enable the minimal architectural and physicochemical requirements underlying network and fibre formation to be defined in both silk and plant protein materials.

- **Multi-molecule coarse-grained simulations of silk assembly.** Extending coarse-grained simulations such as CALVADOS3 from single-chain representations to multi-molecule systems would enable collective effects and intermolecular contacts to be examined explicitly. This would provide insight into how modelled solution-state molecular architecture could translate into higher-order network formation and fibre development.
- **Statistical and Monte Carlo approaches to assembly landscapes.** Monte Carlo and related statistical simulation strategies could be used to sample assembly landscapes across a broad range of concentrations, solvent compositions, and interaction strengths. These approaches would complement molecular simulations by mapping dominant pathways and metastable states. They would also help identify transition regimes that are difficult to access using single-trajectory or time-limited simulations.

References

Adamczyk, G., Kaszuba, J., Kapusta, I., Tarahi, M., Hedayati, S., & Shahidi, F. (2022). Effects of Mung Bean (Vigna radiata) Protein Isolate on Rheological, Textural, and Structural Properties of Native Corn Starch. *Polymers*, *14*(15), 3012. https://doi.org/10.3390/POLYM14153012

Andersson, M., Johansson, J., Rising, A., Andersson, M., Johansson, J., & Rising, A. (2016). Silk Spinning in Silkworms and Spiders. *International Journal of Molecular Sciences*, *17*(8). https://doi.org/10.3390/IJMS17081290

Antosiewicz, J. M., & Shugar, D. (2016). UV–Vis spectroscopy of tyrosine side-groups in studies of protein structure. Part 2: selected applications. *Biophysical Reviews*, *8*(2), 163. https://doi.org/10.1007/S12551-016-0197-7

Arad, E., Green, H., Jelinek, R., & Rapaport, H. (2020). Revisiting thioflavin T (ThT) fluorescence as a marker of protein fibrillation – The prominent role of electrostatic interactions. *Journal of Colloid and Interface Science*, *573*, 87–95. https://doi.org/10.1016/J.JCIS.2020.03.075

Asen, N. D., Aluko, R. E., Martynenko, A., Utioh, A., & Bhowmik, P. (2023). Yellow Field Pea Protein (Pisum sativum L.): Extraction Technologies, Functionalities, and Applications. *Foods*, *12*(21), 1–34. https://doi.org/10.3390/foods12213978

Askarieh, G., Hedhammar, M., Nordling, K., Saenz, A., Casals, C., Rising, A., Johansson, J., & Knight, S. D. (2010). Self-assembly of spider silk proteins is controlled by a pH-sensitive relay. *Nature*, *465*(7295), 236–238. https://doi.org/10.1038/nature08962

Bagasettihalli Kariyappa, P., Sethi, S., & Arora, B. (2025). Extraction and characterization of plant-based proteins from agro-processing by-products—A review. *Food Biomacromolecules*. https://doi.org/10.1002/FOB2.70025

Banc, A., Charbonneau, C., Dahesh, M., Appavou, M. S., Fu, Z., Morel, M. H., & Ramos, L. (2016). Small angle neutron scattering contrast variation reveals heterogeneities of interactions in protein gels. *Soft Matter*, *12*(24), 5340–5352. https://doi.org/10.1039/C6SM00710D

Barth, A. (2007). Infrared spectroscopy of proteins. *Biochimica et Biophysica Acta - Bioenergetics*, *1767*(9), 1073–1101. https://doi.org/10.1016/j.bbabio.2007.06.004

Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., & Svergun, D. I. (2007). Structural Characterization of Flexible Proteins Using Small-Angle X-ray Scattering. *Journal of the American Chemical Society*, *129*(17), 5656–5664. https://doi.org/10.1021/JA069124N

Bhattarai, A., & Emerson, I. A. (2020). Dynamic conformational flexibility and molecular interactions of intrinsically disordered proteins. *Journal of Biosciences*, *45*(1), 1–17. https://doi.org/10.1007/s12038-020-0010-4

Bishop, K. J., Wilmer, C. E., Soh, S., & Grzybowski, B. A. (2009). Nanoscale forces and their uses in self-assembly. *Small*, *5*(14), 1600–1630. https://doi.org/10.1002/smll.200900358

Boulet-Audet, M., Terry, A. E., Vollrath, F., & Holland, C. (2014). Silk protein aggregation kinetics revealed by Rheo-IR. *Acta Biomaterialia*, *10*(2), 776–784. https://doi.org/10.1016/j.actbio.2013.10.032

Breslauer, D. N. (2025). Silk Proteins. In *Reference module in life sciences*. Elsevier. https://doi.org/10.1016/B978-0-443-24738-5.00029-X

Brookstein, O., Shimoni, E., Eliaz, D., Dezorella, N., Biran, I., Rechav, K., Sivan, E., Kozell, A., & Shimanovich, U. (2024). The natural material evolution and stage-wise assembly of silk along the silk gland [Preprint, review complete]. *bioRxiv*. https://doi.org/10.1101/2024.04.16.589504

Chen, F., Porter, D., & Vollrath, F. (2012). Structure and physical properties of silkworm cocoons. *Journal of the Royal Society Interface*, *9*(74), 2299–2308. https://doi.org/10.1098/rsif.2011.0887

Chen, X., Shao, Z., Knight, D. P., & Vollrath, F. (2007). Conformation transition kinetics of Bombyx mori silk protein. *Proteins: Structure, Function, and Bioinformatics*, *68*(1), 223–231. https://doi.org/10.1002/PROT.21414

Clark, D. P., Pazdernik, N. J., & McGehee, M. R. (2019). Protein structure and function. In D. P. Clark, N. J. Pazdernik, & M. R. McGehee (Eds.), *Molecular biology* (pp. 445–483). Academic Press. https://doi.org/10.1016/B978-0-12-813288-3.00014-8

Craig, C. L. (1997). Evolution of arthropod silks. *Annual Review of Entomology*, *42*, 231–267. https://doi.org/10.1146/annurev.ento.42.1.231

Craig, C. L. (2003). *Spiderwebs and silk: Tracing evolution from molecules to genes to phenotypes*. Oxford University Press.

Del Giudice, A., Dicko, C., Galantini, L., & Pavel, N. V. (2017). Time-Dependent pH Scanning of the Acid-Induced Unfolding of Human Serum Albumin Reveals Stabilization of the Native Form by Palmitic Acid Binding. *Journal of Physical Chemistry B*, *121*(17), 4388–4399. https://doi.org/10.1021/acs.jpcb.7b01342

Del Sol, A., & Carbonell, P. (2007). The Modular Organization of Domain Structures: Insights into Protein–Protein Binding. *PLOS Computational Biology*, *3*(12), e239. https://doi.org/10.1371/JOURNAL.PCBI.0030239

Dicko, C., Engberg, A., Houston, J. E., Jackson, A. J., Pettersson, A., Dalgliesh, R. M., Akeroyd, F. A., Venero, D. A., Rogers, S. E., Martel, A., Porcar, L., & Rennie, A. R. (2020). Nurf—optimization of in situ uv–vis and fluorescence and autonomous characterization techniques with small-angle neutron scattering instrumentation. *Rev. Sci. Instrum.*, *91*(7), 075111. https://doi.org/https://doi.org/10.1063/5.0011325

Dicko, C., Knight, D., Kenney, J. M., & Vollrath, F. (2004). Secondary structures and conformational changes in flagelliform, cylindrical, major, and minor ampullate silk proteins. temperature and concentration effects. *Biomacromolecules*, *5*(3), 758–767. https://doi.org/10.1021/bm034486y

Dicko, C., Porter, D., Bond, J., Kenney, J. M., & Vollrath, F. (2008). Structural disorder in silk proteins reveals the emergence of elastomericity. *Biomacromolecules*, *9*(1), 216–221. https://doi.org/10.1021/bm701069y

Dobson, C. M. (2003). Protein folding and misfolding. *Nature*, *426*(6968), 884–890. https://doi.org/10.1038/NATURE02261

Dong, Z., Zhao, P., Zhang, Y., Song, Q., Zhang, X., Guo, P., Wang, D., & Xia, Q. (2016). Analysis of proteome dynamics inside the silk gland lumen of *Bombyx mori*. *Scientific Reports*, *6*, 21158. https://doi.org/10.1038/srep21158

dos Santos Rodrigues, F. H., Delgado, G. G., Santana da Costa, T., & Tasic, L. (2023). Applications of fluorescence spectroscopy in protein conformational changes and intermolecular contacts. *BBA Advances*, *3*, 100091. https://doi.org/10.1016/J.BBADVA.2023.100091

Eliaz, D., Paul, S., Benyamin, D., Cernescu, A., Cohen, S. R., Rosenhek-Goldian, I., Brookstein, O., Miali, M. E., Solomonov, A., Greenblatt, M., Levy, Y., Raviv, U., Barth, A., & Shimanovich, U. (2022). Micro and nano-scale compartments guide the structural transition of silk protein monomers into silk fibers. *Nature Communications*, *13*, 3532. https://doi.org/10.1038/s41467-022-35505-w

Ferreira, B. M., Andersson, N., Atterling, E., Engqvist, J., Hall, S., & Dicko, C. (2020). 3D Structure and Mechanics of Silk Sponge Scaffolds Is Governed by Larger Pore Sizes. *Frontiers in Materials*, *7*, 506339. https://doi.org/10.3389/fmats.2020.00211

Gilbert, E. P. (2019). Small-angle X-Ray and neutron scattering in food colloids. *Current Opinion in Colloid & Interface Science*, *42*, 55–72. https://doi.org/10.1016/J.COCIS.2019.03.005

Greving, I., Dicko, C., Terry, A., Callow, P., & Vollrath, F. (2010). Small angle neutron scattering of native and reconstituted silk fibroin. *Soft Matter*, *6*(2), 438–446. https://doi.org/10.1039/c0sm00108b

Greving, I., Terry, A. E., Holland, C., Boulet-Audet, M., Grillo, I., Vollrath, F., Dicko, C., Greving, I., Terry, A. E., & Dicko, C. (2020). Structural Diversity of Native Major Ampullate, Minor Ampullate, Cylindriform, and Flagelliform Silk Proteins in Solution. *Biomacromolecules*, *21*(8), 3387–3393. https://doi.org/10.1021/acs.biomac.0c00819

Halperin-Sternfeld, M., Ghosh, M., & Adler-Abramovich, L. (2017). Advantages of self-assembled supramolecular polymers toward biological applications. In *Supramolecular chemistry of biomimetic systems* (pp. 9–35). Springer Singapore. https://doi.org/10.1007/978-981-10-6059-5_2

Hardy, J. G., Römer, L. M., & Scheibel, T. R. (2008). Polymeric materials based on silk proteins. *Polymer*, *49*(20), 4309–4327. https://doi.org/10.1016/J.POLYMER.2008.08.006

Hayes, O. G., Partridge, B. E., & Mirkin, C. A. (2021). Encoding hierarchical assembly pathways of proteins with DNA. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(40), e2106808118. https://doi.org/10.1073/PNAS.2106808118;WGROUP:STRING:PUBLICATION

Holland, C., Vollrath, F., Ryan, A. J., & Mykhaylyk, O. O. (2012). Silk and synthetic polymers: Reconciling 100 degrees of separation. *Advanced Materials*, *24*(1), 105–109. https://doi.org/doi.org/10.1002/adma.201103664

Houston, J. E., Schweins, R., Cowieson, N. P., Smith, G. N., & Scotti, A. (2026). A small-angle scattering structural characterization of regular versus gluten-free spaghetti. *Food Hydrocolloids*, *172*, 111855. https://doi.org/10.1016/J.FOODHYD.2025.111855

Huang, L., Shi, J., Zhou, W., & Zhang, Q. (2023). Advances in Preparation and Properties of Regenerated Silk Fibroin. *International Journal of Molecular Sciences*. https://doi.org/10.3390/ijms241713153

Jackson, A. J. (2010). *Introduction to small-angle neutron scattering and neutron reflectometry*. NIST Center for Neutron Research. https://www.ncnr.nist.gov/summerschool/ss10/pdf/SANS_NR_Intro.pdf

Jaumot, J., de Juan, A., & Tauler, R. (2015). MCR-ALS GUI 2.0: New features and applications. *Chemometrics and Intelligent Laboratory Systems*, *140*, 1–12. https://doi.org/10.1016/J.CHEMOLAB.2014.10.003

Jeffries, C. M., Ilavsky, J., Martel, A., Hinrichs, S., Meyer, A., Pedersen, J. S., Sokolova, A. V., & Svergun, D. I. (2021). Small-angle X-ray and neutron scattering. *Nature Reviews Methods Primers*, *1*(1), 1–39. https://doi.org/10.1038/s43586-021-00064-9

Jensen, F. B., Fago, A., & Weber, R. E. (1998). Hemoglobin structure and function. In S. F. Perry & B. L. Tufts (Eds.), *Fish physiology* (pp. 1–40, Vol. 17). Academic Press. https://doi.org/10.1016/S1546-5098(08)60257-5

Jin, H. J., & Kaplan, D. L. (2003). Mechanism of silk processing in insects and spiders. *Nature*, *424*(6952), 1057–1061. https : / / doi . org / 10 . 1038 / nature01809

Koeppel, A., & Holland, C. (2017). Progress and Trends in Artificial Silk Spinning: A Systematic Review. *ACS Biomaterials Science and Engineering*, *3*(3), 226–237. https://doi.org/10.1021/acsbiomaterials.6b00669

Kornet, R., Veenemans, J., Venema, P., van der Goot, A. J., Meinders, M., Sagis, L., & van der Linden, E. (2021). Less is more: Limited fractionation yields stronger gels for pea proteins. *Food Hydrocolloids*, *112*. https://doi.org/10.1016/J.FOODHYD.2020.106285

Kumar, M., Tomar, M., Potkule, J., Reetu, Punia, S., Dhakane-Lad, J., Singh, S., Dhumal, S., Pradhan, P. C., Bhushan, B., Anitha, T., Alajil, O., Alhariri, A., Amarowicz, R., & Kennedy, J. F. (2022). Functional characterization of plant-based protein to determine its quality for food applications. *Food Hydrocolloids*, *123*, 106986. https://doi.org/10.1016/j.foodhyd.2021.106986

Landreh, M., Osterholz, H., Chen, G., Knight, S. D., Rising, A., & Leppert, A. (2024). Liquid-liquid crystalline phase separation of spider silk proteins. *Communications Chemistry*, *7*(1), 1–8. https://doi.org/10.1038/s42004-024-01357-2

Lay, M. G., Oktaviani, N. A., Malay, A. D., & Numata, K. (2025). Exploring the self-assembly of silk proteins through liquid-liquid phase separation. *Polymer Journal 2025 57:8*, *57*(8), 799–814. https://doi.org/10.1038/s41428-025-01040-w

Lefèvre, T., Boudreault, S., Cloutier, C., & Pézolet, M. (2011). Diversity of Molecular Transformations Involved in the Formation of Spider Silks. *Journal of Molecular Biology*, *405*(1), 238–253. https://doi.org/10.1016/J.JMB.2010.10.052

Li, J. (2017). Molecular biomimetics and molecular assembly. In *Supramolecular chemistry of biomimetic systems* (pp. 3–7). Springer Singapore. https://doi.org/10.1007/978-981-10-6059-5_1

Li, Y., Tian, R., Shi, H., Xu, J., Wang, T., & Liu, J. (2023). Protein assembly: Controllable design strategies and applications in biology. *Aggregate*, *4*(3). https://doi.org/10.1002/agt2.317

Lin, Z., Huang, W., Zhang, J., Fan, J. S., & Yang, D. (2009). Solution structure of eggcase silk protein and its implications for silk fiber formation. *Proceedings of the National Academy of Sciences of the United States*

of America, *106*(22), 8906–8911. https : / / doi . org / 10 . 1073 / PNAS . 0813255106

Liu, L., Boldon, L., Urquhart, M., & Wang, X. (2013). Small- and wide-angle x-ray scattering studies of biological macromolecules in solution. *Journal of Visualized Experiments*, (71), 4160. https://doi.org/10.3791/4160

Lujerdean, C., Baci, G. M., Cucu, A. A., & Dezmirean, D. S. (2022). The Contribution of Silk Fibroin in Biomedical Engineering. *Insects*, *13*(3), 286. https://doi.org/10.3390/INSECTS13030286

Manalastas-Cantos, K., Konarev, P. V., Hajizadeh, N. R., Kikhney, A. G., Petoukhov, M. V., Molodenskiy, D. S., Panjkovich, A., Mertens, H. D. T., Gruzinov, A., Borges, C., Jeffries, C. M., Svergun, D. I., & Franke, D. (2021). ATSAS 3.0: Expanded functionality and new tools for small-angle scattering data analysis. *Journal of Applied Crystallography*, *54*, 343–355. https://doi.org/10.1107/S1600576720013412

Mohammadi, P., Aranko, A. S., Lemetti, L., Cenev, Z., Zhou, Q., Virtanen, S., Landowski, C. P., Penttilä, M., Fischer, W. J., Wagermaier, W., & Linder, M. B. (2018). Phase transitions as intermediate steps in the formation of molecularly engineered protein fibers. *Communications Biology 2018 1:1*, *1*(1), 1–12. https://doi.org/10.1038/S42003-018-0090-Y

Mondor, M., & Hernández-álvarez, A. J. (2022). Processing Technologies to Produce Plant Protein Concentrates and Isolates. *Plant Protein Foods*, 61–108. https://doi.org/https://doi.org/10.1007/978-3-030-91206-2_3

Moreno-Tortolero, R. O., Luo, Y., Parmeggiani, F., Skaer, N. J. V., Walker, R., Serpell, L. C., Holland, C., & Davis, S. A. (2024). Molecular organization of fibroin heavy chain and mechanism of fibre formation in *bombyx mori*. *Communications Biology*, *7*(1), 786. https://doi.org/10.1038/s42003-024-06474-1

Nanda, V., Zahid, S., Xu, F., & Levine, D. (2011). Computational design of intermolecular stability and specificity in protein self-assembly. *Methods in Enzymology*, *487*(100), 575–593. https://doi.org/10.1016/B978-0-12-381270-4.00020-2

Narayanan, T. (2008). Synchrotron small-angle x-ray scattering. In *Soft matter characterization* (pp. 899–952). Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-4465-6_17

National Institute of Standards and Technology. (2023). *The sans toolbox* [Accessed: 2025-09-26]. https://www.nist.gov/system/files/documents/2023/04/14/the_sans_toolbox.pdf

Parent, L. R., Onofrei, D., Xu, D., Stengel, D., Roehling, J. D., Bennett Addison, J., Forman, C., Amin, S. A., Cherry, B. R., Yarger, J. L., Gianneschi, N. C., & Holland, G. P. (2018). Hierarchical spidroin micellar nanoparticles as the fundamental precursors of spider silks. *Proceedings of the National*

Academy of Sciences of the United States of America, *115*(45), 11507–11512. https://doi.org/10.1073/pnas.1810203115

Qi, Y., Wang, H., Wei, K., Yang, Y., Zheng, R. Y., Kim, I. S., & Zhang, K. Q. (2017). A Review of Structure Construction of Silk Fibroin Biomaterials from Single Structures to Multi-Level Structures. *International Journal of Molecular Sciences 2017, Vol. 18, Page 237*, *18*(3), 237. https://doi.org/10.3390/IJMS18030237

Rajasekaran, N., & Kaiser, C. M. (2024). Navigating the complexities of multi-domain protein folding. *Current Opinion in Structural Biology*, *86*, 102790. https://doi.org/10.1016/J.SBI.2024.102790

Rehm, C., Brûlé, A., Freund, A. K., & Kennedy, S. J. (2013). Kookaburra: The ultra-small-angle neutron scattering instrument at OPAL. *Journal of Applied Crystallography*, *46*(6), 1699–1704. https://doi.org/10.1107/S0021889813025788

Rehm, C., De Campo, L., Brûlé, A., Darmann, F., Bartsch, F., & Berry, A. (2018). Design and performance of the variable-wavelength Bonse-Hart ultra-small-angle neutron scattering diffractometer KOOKABURRA at ANSTO: *Journal of Applied Crystallography*, *51*(1), 1–8. https://doi.org/10.1107/S1600576717016879

Richards, J. J. (2018). *Contrast variation small-angle neutron scattering: Identifying the unique fingerprints for the structure of proteins* [Summer school module]. NIST Center for Neutron Research. https://www.nist.gov/system/files/documents/2018/06/06/ng7sansc.pdf

Rising, A., & Johansson, J. (2015). Toward spinning artificial spider silk. *Nature Chemical Biology*, *11*, 309–315. https://doi.org/10.1038/nchembio.1789

Royer, C. A. (2006). Probing protein folding and conformational transitions with fluorescence. *Chemical Reviews*, *106*(5), 1769–1784. https://doi.org/10.1021/cr0404390

Shanthakumar, P., Klepacka, J., Bains, A., Chawla, P., Dhull, S. B., & Najda, A. (2022). The current situation of pea protein and its application in the food industry. *Molecules*, *27*(16), 5354. https://doi.org/10.3390/molecules27165354

Shao, Z., & Vollrath, F. (2002). Surprising strength of silkworm silk. *Nature*, *418*(6899), 741. https://doi.org/10.1038/418741a

Stetefeld, J., McKenna, S. A., & Patel, T. R. (2016). Dynamic light scattering: a practical guide and applications in biomedical sciences. *Biophysical Reviews*, *8*(4), 409. https://doi.org/10.1007/S12551-016-0218-6

Stribeck, N. (2007). Polydispersity and heterogeneity. In *X-ray scattering of soft matter* (1st). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-69856-2_1

Sun, H., Li, Y., Yu, S., & Liu, J. (2020). Hierarchical Self-Assembly of Proteins Through Rationally Designed Supramolecular Interfaces. *Frontiers in Bioengineering and Biotechnology*, *8*, 529688. https://doi.org/10.3389/FBIOE.2020.00295

Tanaka, K., Inoue, S., & Mizuno, S. (1999). Hydrophobic interaction of P25, containing Asn-linked oligosaccharide chains, with the H-L complex of silk fibroin produced by Bombyx mori. *Insect Biochemistry and Molecular Biology*, *29*(3), 269–276. https://doi.org/10.1016/S0965-1748(98)00135-0

Tauler, R., Maeder, M., & de Juan, A. (2009). Multiset Data Analysis: Extended Multivariate Curve Resolution. *Comprehensive Chemometrics*, *2*, 473–505. https://doi.org/10.1016/B978-044452701-1.00055-7

Terry, A. E., Knight, D. P., Porter, D., & Vollrath, F. (2004). pH induced changes in the rheology of silk fibroin solution from the middle division of Bombyx mori silkworm. *Biomacromolecules*, *5*(3), 768–772. https://doi.org/10.1021/BM034381V

Tian, M., & Lewis, R. V. (2006). Tubuliform silk protein: A protein with unique molecular characteristics and mechanical properties in the spider silk fibroin family. *Applied Physics A*, *82*(2), 265–273. https://doi.org/10.1007/s00339-005-3433-8

Vepari, C., & Kaplan, D. L. (2007). Silk as a biomaterial. *Progress in Polymer Science*, *32*(8-9), 991–1007. https://doi.org/https://doi.org/10.1016/j.progpolymsci.2007.05.013

Vollrath, F., & Knight, D. P. (2001). Liquid crystalline spinning of spider silk. *Nature*, *410*(6828), 541–548. https://doi.org/10.1038/35069000

von Bülow, S., Yasuda, I., Cao, F., Schulze, T. K., Trolle, A. I., Rauh, A. S., Crehuet, R., Lindorff-Larsen, K., & Tesei, G. (2025). Software package for simulations using the coarse-grained calvados model. *arXiv*. https://doi.org/10.48550/arXiv.2504.10408

Wan, Q., Yang, M., Hu, J., Lei, F., Shuai, Y., Wang, J., Holland, C., Rodenburg, C., & Yang, M. (2021). Mesoscale structure development reveals when a silkworm silk is spun. *Nature Communications 2021 12:1*, *12*(1), 1–8. https://doi.org/10.1038/s41467-021-23960-w

Wang, J., Yuan, W., Qin, R., Fan, T., song Fan, J., Huang, W., Yang, D., & Lin, Z. (2021). Self-assembly of tubuliform spidroins driven by hydrophobic interactions among terminal domains. *International Journal of Biological Macromolecules*, *166*, 1141–1148. https://doi.org/10.1016/J.IJBIOMAC.2020.10.269

Warren, F. J., Gidley, M. J., & Flanagan, B. M. (2016). Infrared spectroscopy as a tool to characterise starch ordered structure—a joint FTIR–ATR, NMR,

XRD and DSC study. *Carbohydrate Polymers*, *139*, 35–42. https://doi.org/10.1016/J.CARBPOL.2015.11.066

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., De Beer, T. A., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research*, *46*(W1), W296–W303. https://doi.org/10.1093/NAR/GKY427

Wen, R., Liu, X., & Meng, Q. (2017). Characterization of full-length tubuliform spidroin gene from Araneus ventricosus. *International Journal of Biological Macromolecules*, *105*, 702–710. https://doi.org/10.1016/j.ijbiomac.2017.07.086

Wen, R., Wang, K., & Meng, Q. (2020). Two novel tubuliform silk gene sequences from Araneus ventricosus provide evidence for multiple loci in genome. *International Journal of Biological Macromolecules*, *160*, 806–813. https://doi.org/10.1016/j.ijbiomac.2020.05.141

Xue, C., Lin, T. Y., Chang, D., & Guo, Z. (2017). Thioflavin T as an amyloid dye: Fibril quantification, optimal concentration and effect on aggregation. *Royal Society Open Science*, *4*(1). https://doi.org/10.1098/rsos.160696

Zheng, K., & Ling, S. (2019). De Novo Design of Recombinant Spider Silk Proteins for Material Applications. *Biotechnology Journal*, *14*(1). https://doi.org/10.1002/biot.201700753

Zhou, C. Z., Confalonieri, F., Medina, N., Zivanovic, Y., Esnault, C., Yang, T., Jacquet, M., Janin, J., Duguet, M., Perasso, R., & Li, Z. G. (2000). Fine organization of Bombyx mori fibroin heavy chain gene. *Nucleic Acids Research*, *28*(12), 2413–2419. https://doi.org/10.1093/NAR/28.12.2413

Zhu, J., Avakyan, N., Kakkis, A., Hoffnagle, A. M., Han, K., Li, Y., Zhang, Z., Choi, T. S., Na, Y., Yu, C. J., & Tezcan, F. A. (2021). Protein assembly by design. *Chemical Reviews*, *121*(22), 13701–13796. https://doi.org/10.1021/acs.chemrev.1c00308

Zuo, B., Dai, L., & Wu, Z. (2006). Analysis of structure and properties of biodegradable regenerated silk fibroin fibers. *Journal of Materials Science*, *41*(11), 3357–3361. https://doi.org/10.1007/s10853-005-5384-z