UNIVERSITY OF TRENTO

Doctoral School in Civil, Environmental and Mechanical Engineering

Curriculum 3. Modelling and Simulation

Vincenzo Fazio

Multiscale models based on statistical mechanics and physically-based machine learning for the thermo-hygro-mechanical behavior of spider-silk-like hierarchical materials

2024 - Doctoral thesis

UNIVERSITY OF TRENTO - Italy
Department of Civil, Environmental and Mechanical Engineering

Doctoral School in Civil, Environmental and Mechanical Engineering
Topic 3. Modelling and Simulation - XXXVI cycle 2020/2023

Doctoral Thesis - April 2024

Vincenzo Fazio

Multiscale models based on statistical mechanics and physically-based machine learning for the thermo-hygro-mechanical behavior of spider-silk-like hierarchical materials

Supervisors
Nicola Maria Pugno, University of Trento
Giuseppe Puglisi, Polytechnic University of Bari
Luca Deseri, University of Trento

Credits of the cover image: Federica Fazio

University of Trento
Doctoral School in Civil, Environmental and Mechanical Engineering
http://web.unitn.it/en/dricam
Via Mesiano 77, I-38123 Trento
Tel. +39 0461 282670 / 2611 - *dicamphd@unitn.it*

Acknowledgements

This thesis was possible thanks to several fundamental contributions. First of all, I would like to thank my advisors.

Prof. Giuseppe Puglisi, my mentor from the very beginning, with his competence and knowledge, has contributed significantly to the effectiveness of my work. I am particularly grateful for his availability and patient guidance.

Prof. Nicola Maria Pugno, the invaluable supervisor of this work, with whom I discussed each key issue and the directions to be followed. His expertise, availability and kindness were truly precious. Moreover, this work and many related experiences have been made possible by the availability of funding from my supervisor.

I am grateful to Prof. Luca Deseri for his kind encouragement and cooperation in this work.

I would also like to thank Prof. Giuseppe Florio, Salvatore Di Stefano, Luca Bellino and Claudia Binetti for their great support and many discussions, which have contributed greatly to my growth as a researcher and as a person.

Finally, I would like to express my deep gratitude to my parents and friends for their tireless encouragement and constant presence throughout the years of my work for this thesis, and beyond.

Vincenzo Fazio

Abstract

Scientists are continuously fascinated by the high degree of sophistication found in natural materials, arising from evolutionary optimisation. In living organisms, nature provides a wide variety of materials, architectures, systems and functions, often based on weak constituents at the lower scales.

One of the most extensively studied natural materials is spider silk, renowned for its outstanding mechanical properties, which include exceptional strength and toughness. Owing to its wide range of properties, which vary depending on factors such as the type of silk (up to seven) that each spider can produce, and the species of spider, it can be considered a class of semi-crystalline polymeric material. Indeed, spider silk cleverly combines, depending on the application required, the great deformability of an amorphous phase with the stiffness and strength conferred by pseudo-crystals consisting of specific secondary structures of some of the proteins constituting the material. Based on the countless studies conducted on spider silk, it is now also clear that its remarkable performance are the result of a sophisticated optimisation of the material's hierarchical structure. Nevertheless, many of the multiscale mechanisms that give rise to the striking macroscopic properties are still unclear. Many open problems are also related to the relevant effects of environmental conditions and in particular on temperature and humidity, strongly conditioning the mechanical performances.

In this thesis, aimed at unveiling some of these open problems, we introduce a multiscale model for the thermo-hygro-mechanical response, starting with the influence of water molecules modifying the microstructure, up to the effects at the macroscopic scale, including softening, increase in elongation at break and supercontraction, i.e. the shortening (up to half the initial length) of the spider threads in wet environments.

Thereafter, we describe how the supercontraction effect can be adopted to obtain humidity-driven actuators, and in particular, we determine the maximum actuation force depending on the silk properties at the molecular scale and on the constraining system representing other silk threads or the actuated device. The spider silk actuation properties turned out to be extraordinary, making spider silk potentially the best performing humidity-driven actuator known to date in terms of work density (2.19 KJ/m^3).

As observed in many natural materials, spider silks are characterized by a strong variability in both chemical and structural organization, as for example

described in the recently published experimental database of properties at different scales of about a thousand different spider silks, where evident correlations among quantities are scarce. This large variability makes the theoretical understanding of the observed material behavior, in relation of the complex hierarchical structure, particularly intriguing.

To address this novel amount of experimental data without losing sight of theoretical analytical modelling, we propose a new data modelling methodology to obtain simple and interpretable relationships linking quantities at different scales. In particular, we employ a symbolic regression technique, known as 'Evolutionary Polynomial Regression', which integrates regression capabilities with the Genetic Programming paradigm, enabling the derivation of explicit analytical formulas, finally delivering a deeper comprehension of the analysed physical phenomenon. Eventually, we provide insights to improve our multiscale theoretical model accounting for the humidity effects on spider silks. This approach may represent a proof of concept for modelling in fields governed by multiscale, hierarchical differential equations.

We believe that the analytical description of the macroscopic behaviour from microscale properties is of great value both for the full understanding of biological materials, as well as in the perspective of bioinspired materials and structures.

Contents

Chapter 1

Introduction

1.1 Hierarchical biological and bioinspired materials

Scientists are impressed by the sophisticated and miniature structures found in natural materials. Over the past 500 million years, proven materials have emerged through natural selection processes. In all living organisms, be they basic or complex, nature offers an array of materials, architectures, systems, and functions. Typically, a precise hierarchical organization of these natural materials across diverse length scales gives rise to their unique performance.

Multiscale, hierarchically patterned materials, such as lotus leaves, butterfly wings and gecko adhesive pads, are abundant in nature, where microstructures are usually used to enhance mechanical stability, while nanostructures provide the main functionality, i.e. wettability, structural colour or dry adhesion [Bae et al., 2014].

Nature is also a prolific producer of composite materials. Examples of natural composites include crustacean shells, mollusc shells, vertebrate bones and teeth, and spider silk. These materials exhibit a high degree of sophistication, with their various components assembled following a clearly defined pattern. The highly elaborated performances characterizing biological materials result also from time-dependent processes [Sanchez et al., 2005]. Even the ultra-soft membranes that enclose cells also exhibit extraordinary properties that can be their structure on multiple length scales. Studies of biological materials often reveal surprising refinements that have been developed through slow evolutionary engineering. These materials are hierarchically structured and display unique properties that are influenced by the structure and the generative processes at all levels of the biological structural hierarchy.

Interestingly, nature operates within a limited range of materials. In biological tissues, the components are primarily proteinaceous, while rigid composites often consist of calcium carbonates, calcium phosphates, and silica [Council et al., 1994]. On the other hand, regarding the use of synthetic hierarchical concepts, the range of materials is potentially great. Additionally, many structural variables can be more easily altered in synthetic materials than in natural materials. In the control of composition and fabrication processes, numerous variables can be altered, although not always independently. These variables include atomic structure, molecular structure, nanostructures, microstructures, and interfaces at all levels, as well as the size, distribution, and morphology of constituents of cells and other substructures. It is also possible to enhance performance through the engineering of dislocation and other defect structures. The interest in these analyses is not only related to the theoretical aspects, but also thinking of bioinspired synthetic hierarchical structures, having great potential and continuously being realized due to advancements in processing technology, which provide precise control over all levels of structural arrangement.

Although natural materials possess exceptional combinations of properties, they also have important limitations. For example, the constituent components of material composites maintain these properties over a temperature range that often is too narrow for most engineering designs. Another example is the mechanical performance of the spider silk, the main focus of this thesis, which undergoes significant changes in properties when exposed to wet environmental conditions. Thus, models about architecture, composite elements and adhesion in mechanical collaboration are useful for learning from nature and applying to other material components to produce analogous synthetic structures that overcome the performance limitations of the inspiring natural counterpart [Chen and Pugno, 2013].

It is crucial to acknowledge the abundance of literature addressing microscopic descriptions and various phenomenological models elucidating macroscopic bulk properties. On the other hand, there has been insufficient progress in theoretical, multiscale investigations that systematically connect mechanical properties at different length scales. To achieve this goal, it is essential to fill the gap by using appropriate multiscale techniques to highlight insights at each hierarchical level [Bechtle et al., 2010; Meyers et al., 2008].

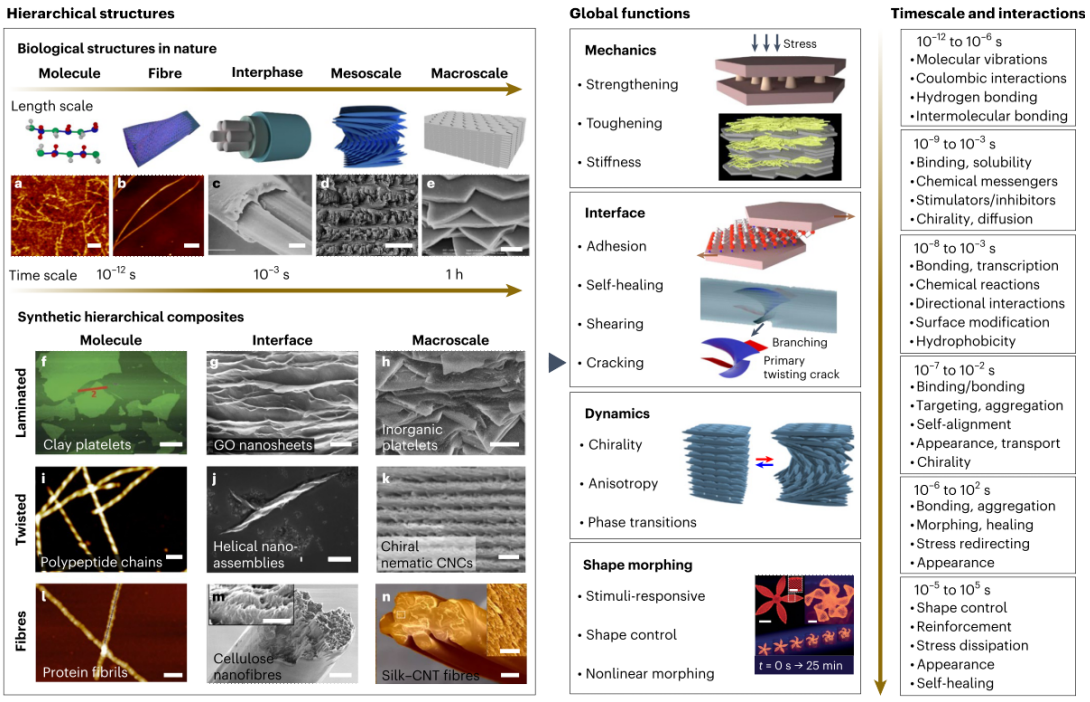

Figure 1.1: Hierarchical bio-inspired composite designs in space and time and significant contributions to mechanical functionality. Top left: representative hierarchical biological structures from nature, showing schematics (top) and actual morphologies (bottom). AFM images of the β-sheet secondary structure of silk fibroin (a) and silk nanofibrils (b). SEM images of single-filament silk fiber (c) hierarchical structure of the dactyl club of the stomatopod (d), natural Cristaria plicata nacre (e). Scale bars: 50 nm (a); 0.5 μm (b); 5 μm (c); 20 μm (d); 1 μm (e). Bottom left: synthetic and hybrid composite material morphologies. AFM image of PVA-coated core–shell clay nanoplatelets (f), SEM images of the layered nanostructure of graphene oxide nanosheets combined with silk fibroin (g) and artificial hybrid nacre materials (h), AFM image of twisted amyloid fibrillar bundles (i), SEM image of self-assembled helices from Cadmium telluride nanoparticles (h), Optical microscopy image of a hierarchically organized polysaccharide composite (k). AFM image of silica-deposited protein core–shell nanofilaments (l), SEM images of nanostructured artificial cellulose nanofibrils with fractured areas shown at higher magnification in the inset and regenerated silk–CNT fibres, shown at higher magnification in the inset. Scale bars: 1 μm (f); 600 nm (g); 1 μm (h); 100 nm (i); 100 nm (j); 10 μm (k); 50 nm (l); 2 μm (m); 400 nm (inset of m); 20 μm (n); 2 μm (inset of n). The hierarchical structures translate into global functions (middle) with a range of timescales and characteristic interactions for each order of magnitude (right). Scale bars (shape morphing panel): 5 mm (main); 2.5 mm (inset). Reproduced from [Nepal et al., 2023].

1.1.1 Typical features

The presence of hierarchies results in a 'smart' response, involving structures at different scales. The behaviour of energy dissipation and fracture is regulated by the properties and functions resulting from the specific arrangement of the

multiscale structure. This arrangement depends on both the function of the tissue or cells and the surrounding environment [Roberts et al., 2002].

The hierarchy formation is mediated by the process of assimilation, which determines how protein structures attach together and attain a specific conformation. During this evolution, other phenomena driven by specific functions of the components arise, such as growth, fiber assembly, or the generation of a network, as seen in the cases of lamins and spider silks [Bini et al., 2004; Herrmann and Aebi, 2004; Rammensee et al., 2008].

Knowledge of these arrangements is essential for deriving mathematical models to describe these materials and designing bioinspired materials.

For example, toughness is typically provided by controlled interfacial features such as hydrogen bonds, friction, or chain assembly straightening and stretching. Thus, knowledge of behaviour at the smallest scale, coupled with the ability to control processes such as DNA recombination or amino acid chain assembly, are the key ingredients for engineering structures at this level. The scale hierarchy in time and length, along with the corresponding experimental techniques, are illustrated in Figure...

Hierarchical materials and systems in biology are characterised by the recurrent use of molecular constituents, leading to diverse properties that can vary significantly even among seemingly similar elementary units. These properties adapt to meet specific performance requirements; for instance, structural elements can be oriented to achieve distinct properties, displaying controlled and often complex shapes. Recurring characteristics of the material include fatigue resistance, resiliency, and self-repair capacity. Another noteworthy characteristic is its sensitivity to, and critical dependence on, the presence of water, which will be analysed in depth in this thesis.

In many biosystems, such a high level of structure-function integration is associated with other aspects such as miniaturisation, which aims to accommodate a maximum of elementary functions in a small volume, and hybridisation between inorganic and organic components, which optimises complementary possibilities [Sanchez et al., 2005].

The complexity and diversity of protein materials generate an enormous number of different structures. Therefore, it is often necessary to categorise them by introducing distinguishing features [Alon, 2007]. A crucial concept in this perspective is the distinction between diffuse and peculiar properties. For example, it is widely acknowledged that protein secondary structures or synthesis and growth phenomena at lower scales exhibit diffuse, often universal properties that are present in almost all biological materials. These properties include primary,

secondary, and higher-order structures, sacrificial joints and weak links, and organization in hierarchical structures. In contrast, higher-order assembled textures such as the titin protein or bone shells display specialized properties that are highly dependent on their function. Thus, it is suggested that diversity increases with scales, indicating that specific functions or specialized protein structures are associated with structural diversity.

In the following, we delineate the key common features among many natural materials, delving into the case of proteinaceous materials, starting from the observation that the properties of complex multiscale structured materials are largely controlled by weak interactions at the nano-scale, particularly hydrogen bonds, acting as sacrificial bonds.

Sacrificial bonds

Sacrificial bonds and hidden lengths in structural molecules and compounds were found to significantly enhance their properties. This is achieved by providing a reversible, molecular-scale energy-dissipation mechanism. Sacrificial bonds are defined as bonds that rupture before strong bonds (often covalent bonds in the backbone) fail under deformation. The hidden length refers to the portion of the molecule that was prevented from stretching by the sacrificial bond.

Due to the nanoscopic nature of this mechanism, single molecule force spectroscopy using an Atomic Force Microscope (AFM) has been a useful tool to investigate this mechanism. Indeed, this instrument, which was initially developed to produce high-resolution images of surfaces, is capable of generating force-extension curves for individual protein molecules.

The understanding of this sacrificial bonding mechanism remained unclear until two decades ago. In 1997, Rief et al. [Rief et al., 1997] conducted a study on the mechanical properties of single titin chains using single-molecule force spectroscopy by AFM. The force–extension curves exhibited a characteristic saw-tooth pattern (see Figure 1.2), indicating the successive unraveling of individual domains within a single titin molecule. While the study did not delve into the mechanism of unraveling, it demonstrated that individual titin domains unravel one at a time. This observation suggested that AFM offered a potential avenue for elucidating sacrificial bonding mechanisms in biological materials [Smith et al., 1999].

In several biological cases, sacrificial bonds have been found to be reversible, providing a "self-healing" property to the material. When a biological material is stretched or loaded, a large amount of energy is dissipated in the biopolymer through the rupture of sacrificial bonds and the release of hidden lengths, thereby

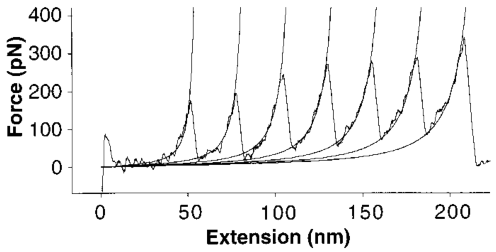

Figure 1.2: A force-extension curve resulting from the stretching of a single Ig8 titin fragment. The curve exhibits a distinctive sawtooth pattern with seven peaks. This pattern can be explained by the stepwise increase in the contour length of a polymer, whose elastic properties are well described by the wormlike chain model (WLC) [Bustamante et al., 1994; Marko and Siggia, 1995]. Reproduced from [Rief et al., 1997].

ensuring the high toughness of the biological material. These sacrificial bonds, such as ionic bonds, hydrogen bonds, and metal–ligand coordination bonds, are generally weaker than covalent bonds, and reversible. They can disrupt prior to covalent bonds and self-repair to maintain the integrity of the biological materials. Examples of sacrificial bonds are the Hydrogen bonds in Titin [Rief et al., 1997] and spider silks [De Tommasi et al., 2010] and Van der Waals forces in Gecko feets [Gao and Yao, 2004]

Recently, great progress in synthesis of polymeric materials containing of sacrificial bonds has been achieved [Zhou et al., 2017]. Scientists have fabricated biomimetic materials using various inorganic and organic components. The concept of sacrificial bonds has recently been applied to other synthetic materials like hydrogels and elastomers, resulting in impressive performance, particularly in terms of toughness. By introducing a variable proportion of isotropically prestretched chains that can break and dissipate energy before the material fails, brittle unfilled elastomers were significantly reinforced in stiffness and toughness using sacrificial bonds [Ducrot et al., 2014].

From weak constituents to high performances

Nature has mastered the art of crafting robust and efficient materials and structures from inherently weak constituents. This ingenious feat is made possible through the implementation of hierarchical mechanisms, a strategy that spans multiple length and time scales to impart strength, resilience, and adaptability. This fascinating phenomenon is evident in various biological materials, showcasing the triumph of hierarchical organization over the limitations of individual components. Many biological materials start with constituents that, in isolation, may be considered weak or exhibit limited mechanical properties. These constituents can include proteins, polymers, minerals, and other elements with

inherently modest strength. For instance, the beta sheets nano crystals give rise to the exceptional resistance of spider silk in a counter-intuitive way compared to the weak hydrogen bonds they consist of [Keten et al., 2010]. A further example in this direction are the collagen fibers, found in connective tissues, which are individually flexible and relatively weak, yet they play a pivotal role in the formation of remarkably strong tissues and structures, such as bones. Thus, hierarchical organization involves arranging constituents from the nanoscale to the macroscale, creating a cascade of reinforcing interactions.

In the bioinspired perspective, a novel design paradigm exploiting simple components and merging structure and materials in the design of hierarchical structure, resulting in enhanced mechanical properties, is emerging. As an example case, Naraghi et al. [Naraghi et al., 2010] made innovative carbon nanotubes bundles where the presence of hierarchy and abundance of relatively weak interaction (entropic elasticity of polymers and van der Waals interactions) resemble the structure of many biological materials. They obtained enhanced ductility without decreasing the strength, so coupling typically competing properties.

Protein structures

A comprehensive understanding of protein structure serves as a fundamental basis for investigating the relationship between structure and function and for accurately describing biological materials from a modelling perspective. These materials have a hierarchical constitution and are often complex, with a wide range of functional applications. However, they share common features in their basic structural organization, which are present in almost all protein materials.

At the most basic level, the 21 amino acids found in humans are arranged in a sequential chain of precise length, measured in terms of the number of amino acids. This chain is responsible for cell activities such as repair, survival, regeneration, and growth [Hughes and Dougan, 2016; Goriely, 2017]. This is known as the primary structure, which is illustrated in Figure 1.3. The folding of a polypeptide sequence is influenced by the position of the amino acids along the chain and their interactions with the surrounding environment, as described by Roberts et al. [Roberts et al., 2002]. This is a long-standing theoretical problem that is currently being approached from various perspectives, including machine learning [Butler et al., 2018], statistical mechanics [Bellino et al., 2019; Manca et al., 2013; Makarov, 2009], and molecular dynamics [Marin-Gonzalez et al., 2017; Karplus and McCammon, 2002; Hsu et al., 2020]. Regarding the deduction of protein folding stability, this issue is particularly significant due to the potential for protein mutations that can result in malfunctioning and pathology.

Conversely, a mismatch in the protein sequence arrangement may cause various issues. Hereditary metabolic disorders, caused by the degradation of specific amino acids, are a well-known medical issue. Aliu et al. [Aliu et al., 2018] have studied the effects of these disorders. One example is sickle cell anaemia, a blood disorder characterized by an anomaly in the transport of red blood cells due to the rigid, sickle-like shape of hemoglobin [Rees et al., 2010].

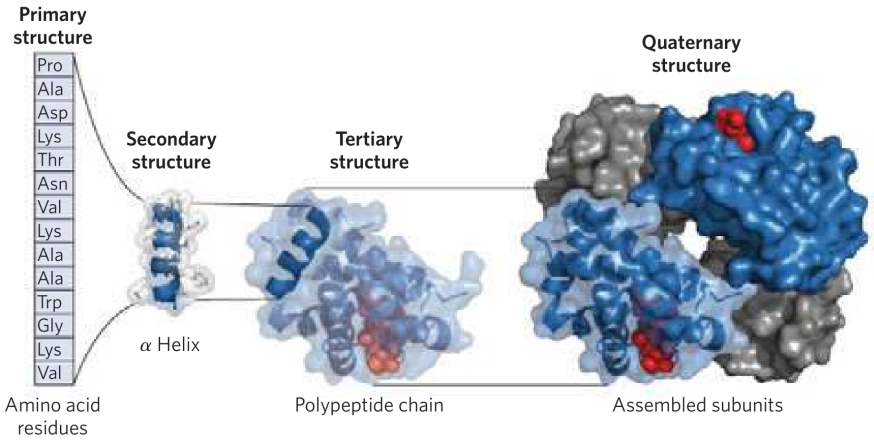

Figure 1.3: The structure of a protein is organized into four levels: primary, secondary, tertiary, and quaternary. The primary structure is the linear sequence of amino acids joined by peptide bonds. The secondary structure refers to local folded structures within the chain, such as the α-helix, stabilized by hydrogen bonds. The tertiary structure is the overall three-dimensional shape of a single polypeptide chain, resulting from interactions between amino acid side chains. Proteins with multiple polypeptide chains exhibit a quaternary structure, which involves the arrangement of these subunits in a functional protein complex. Reproduced from [Nelson et al., 2008]

The secondary structure is formed when the polypeptide chain folds, combining the strong local covalent bonds between amino acids with weaker non-local hydrogen bonds. Two of the most common secondary configurations adopted by proteins in nature are the β-sheet and α-helix, as shown in Figure 1.3, respectively. Depending on the direction of the amino acid planes, two conformational β-sheet configurations can be achieved, anti-parallel or parallel strands. Furthermore, many types of helices can be found, depending on the pitch of the helix, as predicted theoretically by Linus Pauling in 1951, for which he received the Nobel Prize [Pauling et al., 1951]. Pauling assumed that the peptide bonds are planar, that all the amino acid residues are equivalent with respect to the backbone conformation, and that the bonds among amide protein to an oxygen atom of another residue are of the hydrogen type [Edison, 2001].

Tertiary structures are formed when secondary structures, which are typically planar, fold into various geometric shapes in three dimensions. The arrangement

of these structures not only determines their shape but also their function. The three-dimensional shapes are primarily regulated by hydrophobic or hydrophilic interactions, where non-polar solutes are linked to minimize their interaction with water [Baldwin, 2013]. For example, tropoelastin, the most elastic protein in the human body, has a specific conformational arrangement of primary, secondary, and tertiary structures. This arrangement facilitates hydrophobic and hydrophilic interactions through hydrogen bonds, allowing it to obtain a shape that can link with other tropoelastin proteins and undergo biological processes [Baldock et al., 2011; Yeo et al., 2016].

Lastly, quaternary structure refers to the coupling and interaction of multiple protein chains and subunits into a tightly packed arrangement, each composed of its own primary, secondary and tertiary structure. Non-covalent bonds, such as Van der Waals forces and hydrogen bonds, typically regulate it. Sir John Kendrew and Max Perutz first discovered it in the 1950s using X-ray diffraction techniques. [Kendrew et al., 1958].

1.1.2 Examples of hierarchical biological materials

In this Section, we describe some examples of hierarchical biological materials, chosen for their particular relevance in materials science, due to their structural properties and interesting properties that make them similar to spider silk, the main subject of this thesis. Bone has a very fascinating and complex hierarchical structure that allows it to serve very different functions: in facts bones protect the various other organs of the body, produce red and white blood cells, store minerals, provide structure and support for the body, and enable mobility.

On the other hand, hair and cellulose were chosen for some similarity with the spider silk. Indeed, hair is, like silk, a protein material and contains some secondary protein structures that are also very common in spider silk. Cellulose, on the other hand, has the very interesting characteristic that it can contract with moisture. We will therefore describe how this contraction takes place in the case of cellulose, whereas, the contraction of spider silk, is the subject of a very in-depth study within the Chapter 2.

Bone

Bone is a complex biological material that exhibits a diverse array of structures across a range of length scales that work together to fulfill mechanical, biological and chemical functions. Its structures work together to provide structural support, protection, storage for healing cells, and maintenance of mineral ion

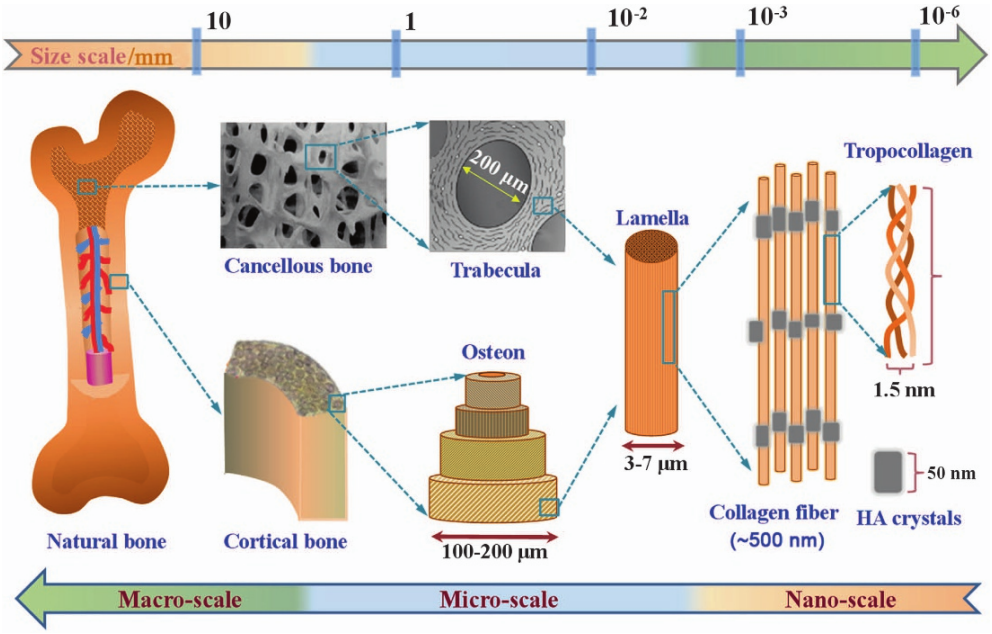

Figure 1.4: The multiscale structure of natural bone. Reproduced from [Gao et al., 2017].

homeostasis. and maintenance of mineral ion homeostasis. The hierarchical and complex nature of bone architecture is pivotal in comprehending its final properties.

The fundamental constituent of bone material is the mineralized collagen fibril, consisting of the fibrous protein collagen, a component also present in skin, tendon, and a variety of other soft tissues. The second component is the mineral dahllite, also known as carbonated apatite. Finally, water forms the third major component, creating with the other two components an ordered structure, the mineralized collagen fibril. To delve into the ultimate properties of bone, understanding the properties of its component phases and their structural relationships across hierarchical levels is crucial [Rho et al., 1998; Gao et al., 2017; Reznikov et al., 2014].

In terms of hierarchical organization, the bone structure, as proposed by Gao et al. [Gao et al., 2017], includes: (i) the macrostructure: cancellous and cortical bone; (ii) the microstructure (from 10 to 500 mm): Haversian systems, osteons, single trabeculae; (iii) the sub-microstructure (1–10 mm): lamellae; (iv) the nanostructure (from a few hundred nanometers to 1 mm): fibrillar collagen and embedded mineral; and (v) the subnanostructure (below a few hundred nanometers): molecular structure of constituent elements, such as mineral, collagen, and non-collagenous organic proteins.

At the macrostructural level, two primary structures are identified: cortical

(compact) and cancellous (trabecular) bone [Weiner and Wagner, 1998; Elias et al., 2013]. Cortical bone boasts a dense outer shell, while cancellous bone features a porous, trabecular interior [Nguyen et al., 2012]. Microstructural differences between them involve histological evaluation, highlighting cortical and cancellous bone as either variations of a single material or distinct entities. Cancellous bone, being metabolically active, undergoes frequent remodeling, influencing its structural and material properties [McKittrick et al., 2010].

At the microstructural level, mineralized collagen fibers organize into lamellae [Weiner and Wagner, 1998]. These lamellae form osteons or Haversian systems in cortical bone, while cancellous bone comprises trabeculae with diverse cellular structures. Mechanical properties vary between cortical and cancellous bone, presenting challenges in expressing bone properties in single values due to regional heterogeneity. The sub-microstructure involves lamellae, presenting challenges in understanding collagen fiber arrangements [Weiner and Wagner, 1998]. Experimental methods such as nanoindentation aid in measuring mechanical properties at the sub-micron level, offering insights for theoretical models in bone mechanics.

Collagen fibers at the nanostructure level are surrounded and infiltrated by mineral [Robinson, 1952; Weiner and Price, 1986; Fratzl et al., 1992]. Specific attachment sites of macromolecules onto the collagen framework are not precisely known, but studies indicate periodic labeling along collagen molecules and fibers. Crystal orientation within collagen fibrils, along with nanocrystalline bone apatite composition, is crucial.

Examining the subnanostructure reveals crystals and collagen fibrils at the tens of nanometers scale. Mature plate-shaped apatite crystals within collagen fibrils limit primary growth, exhibiting specific orientation. Type I collagen, the primary organic component [Hodge, 1963] self-assembles into fibrils with a specific periodicity. Non-collagenous organic proteins may regulate mineral deposits, necessitating further studies for conclusive understanding.

The hierarchical organization of bone, spanning macro to subnanostructures, contributes to its heterogeneity and anisotropy, ultimately shaping its mechanical, biological, and chemical functions. The complex structure of bone and its relationship to function has been studied for more than 320 years and is still far from fully understood [Reznikov et al., 2014].

Hair

Hair is a filamentous appendage of the skin in vertebrates that serves to protect the body mainly against cold and wet conditions. It is a proteinaceous material, and much of the research in hair science is fundamental to protein science. Hair

is renowned for its remarkable mechanical properties and very long degradation time [Matsunaga et al., 2013]. Healthy human hair has a Young's modulus of 2.0-3.7 GPa and a failure stress and strain of approximately 117 MPa and 35%, respectively [Chou et al., 2015].

The hierarchical structure of hair is shown in Figure 1.5. A typical hair fiber has a diameter ranging from 50 to 100 μm and is covered by an outermost layer known as the cuticle, composed of dead, overlapping cells forming a protective layer around the hair. The inner part of the hair is called cortex and it is composed of cortical cells that are approximately 100 μm long and 1-6 μm thick. These cortical cells consist of macrofibrils, which have a diameter of 0.1-0.4 μm [Harland et al., 2014]. At the nanometer scale, macrofibrils are composed of intermediate filaments (IF) embedded in a high-sulfide content matrix. Each IF has a diameter of approximately 7.5 nm and is formed by eight protofilaments. An intermediate filament (IF) is composed of four right-handed α-helix chains, known as protofilaments. Therefore, a total of thirty-two chains form an IF [Voet and Voet, 2010]. The mechanical strength and durability of hair are determined by one of the main keratin proteins: trichocyte keratin. This protein constitutes the intermediate filaments and interacts with the matrix and crosslinks with disulfide bonds [Matsunaga et al., 2013; Chou et al., 2015].

The intricate arrangement of the cortex and cuticle contributes to the unique mechanical and physical properties of human hair, making it an intriguing example of a hierarchical biological material.

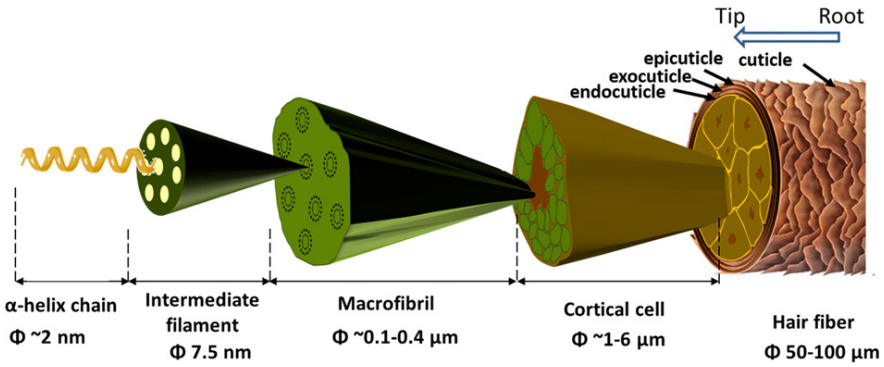

Figure 1.5: Scheme of the hierarchical structure in human hair. Reproduced from [Yu et al., 2017].

As is common in biological materials, the tensile properties are dependent on various factors: high relative humidity or high temperature can decrease the Young's modulus and increase the extensibility [Robbins and Robbins, 2012; Rebenfeld et al., 1966]; Additionally, the hair's yield stress decreases as the strain rate decreases, relative humidity increases, and temperature increases [Yu et al.,

2017]. These environmental factors can lead to changes in the appearance and texture of the hair, highlighting the sensitivity of hair to its surrounding conditions.

Interestingly, the hair swells when exposed to humid environments [Stam et al., 1952]. The secondary and tertiary structure of the IFs and the matrix are thought to control the swelling behaviour of keratins [Robbins and Robbins, 2012]. Even if the maximum swelling occurs in the diametral dimension of the hair and only a secondary swelling occurs in the longitudinal dimension that is along the axis of the fibres, the longitudinal swelling is the mechanism fundamental for the hair hygrometer [Trowbridge, 1896; Whipple, 1921; Mehrabian and Hashemian, 2003].

Understanding the hierarchical structure and structure-function relationships in human hair is important not only for its biological significance but also for its relevance in various fields, including materials science and mechanical actuation besides cosmetic research.

Cellulose

Cellulose is a natural linear biopolymer, the most abundant natural polysaccharide, with a complex hierarchical structure, which is essential to its mechanical properties and functionality. Specifically, it is a high molecular weight polysaccharide composed of repeating cellobiose units that form a linear chain in which both intra-chain and inter-chain molecular hydrogen bonds that bind the chains together [Di Donato et al., 2015], which in turn produce microfibrils, matrices, and multilayered cell walls (Fig. 1.6).

The mechanical response of natural plants to humidity, such as pinecones [Andres et al., 2014] and ice plants [Harrington et al., 2011], with a sophisticated hydration-dependent unfolding of their seed capsules, has inspired researchers to investigate the deformation of structures in response to humidity changes. The humidity-induced response is due to the asymmetric swelling behavior of cellulose with different orientations in plant tissue at high humidity [Dawson et al., 1997].

Interestingly, a cellulose fibre in a humid environment typically swells, as do hair and also hemp and wood, being predominantly composed of cellulose. On the other hand, a rope composed of fibers (which swell) arranged in a helical fashion, contracts, because the swelling is greater in the direction of the diameter than of the length of the fibers (ratio estimated at 20:1) and this leads to an increase in diameter and a reduction in the pitch of the helix, the length of the fibers having to remain almost constant (compared to the increase in diameter) [Ridge, 1953]. However, there is a possibility of fiber contraction with moisture.

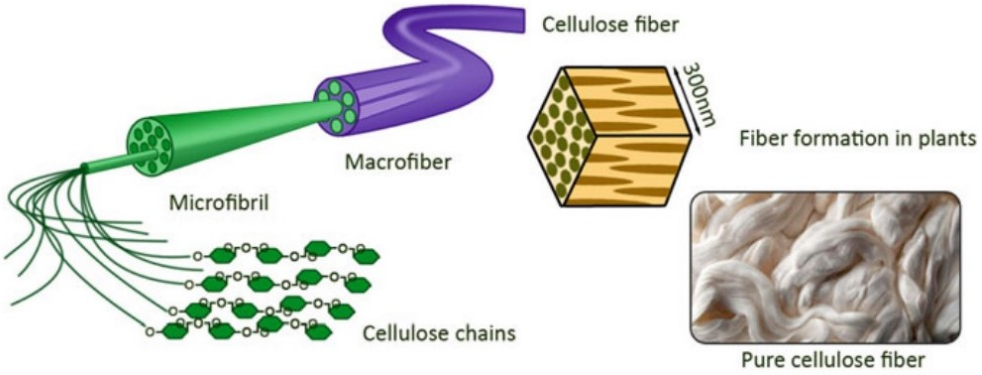

Figure 1.6: Cellulose fiber hierarchical composition starting from cellulose chains. Adapted from [Ramawat and Ahuja, 2016].

This occurs when macromolecules are frozen in a non-equilibrium condition in an elongated configuration [Bowling et al., 2001]. In this way, as occurs in spider silk and extensively addressed within this thesis, water molecules of hydration can break the bonds that freeze macromolecules in elongated configuration, allowing them to return to the equilibrium condition by contracting from the elongated frozen condition. Thus the cellulose fiber contracts due to the strong network of hydrogen bonds among all chains [Krässig, 1993] returning to the conformational equilibrium of individual chains [Le Moigne et al., 2010].

Contraction upon swelling occurs in both native and regenerated fibers due to the extended conformational state of cellulose chains resulting from the spinning process for regenerated cellulose fibers or the bio-deposition process for native cellulose fibers. As anticipated, the contraction is caused by the change in the mean conformation of cellulose chains from an extended state to a more condensed state.

For regenerated fibers, the high-viscosity cellulose solution passes through a spinneret where shear forces orient the cellulose chains in the spinning direction. Before being coagulated in a water bath, the resulting cellulose fibers are stretched in an air gap. The cellulose chains are therefore frozen in an extended conformation, out of equilibrium. This result in the well-known highly oriented morphology of regenerated cellulose, with both the crystalline and non-crystalline phases having cellulose chains that are strongly oriented in the direction of the fiber [Lenz et al., 1993].

According to Bowling et al. [Bowling et al., 2001], cellulose chains in cotton hairs or wood fibers are in an extended conformational state due to the bio-deposition mechanism. The Cellulose Synthase Systems extrude cellulose chains that condense into crystalline microfibrils before they can fully relax, thus main-

taining the chains in an extended conformational state. During partial dissolution or swelling, molecular mobility enables cellulose chains from the non-crystalline phase (also known as the amorphous phase) to return to their equilibrium conformation, resulting in a significantly reduced end-to-end distance. The value of the equilibrium end-to-end distance depends on the swelling or dissolving agent.

Eventually, thanks to its humidity responsiveness resulting from its rich hydrophilic groups [Li et al., 2021], cellulose is an excellent material for developing green humidity sensors. A common mechanism for producing moisture sensors from cellulose is to employ the bending of two or more layers of cellulose with different orientations [Tian et al., 2018]. Cellulose can be used in various types of sensors, including resistance, capacitance, colorimetry, shape deformation, and capacitive micromachined ultrasonic transducer (CUMT) sensors [Li et al., 2021].

Spider silk

Finally, we present the case of the spider silk, where the amino acids aggregate in particular proteins that in turn are assembled in structures that give rise to an exceptional macroscopic behavior, which leads spider silk to be one of the most resistant materials in nature and with resistances comparable to those of high strength steels, as well as a toughness that is surpassed only by a few manmade high-performance materials.

Its hierarchical structure will be described in detail in Section 2.1, yet here we introduce such material, starting with a brief discussion about the animals that produce it *i. e.* the spiders, to provide insight into the origin of its high performance and variable properties, and ending with the various applications of both natural and bio-inspired silk.

Spiders are arthropods characterized by eight limbs and spinnerets that enable silk production. Found on every continent except Antarctica, they have demonstrated remarkable adaptability to a wide range of land habitats. Spiders have evolved to overcome various environmental challenges, including large fluctuations in temperature and humidity. Their ability to thrive in diverse environments is evidence of the incredible strategies and structures they have developed to navigate and flourish in different ecological settings [Canals et al., 2015]. The number of known spider species is estimated to be around 50,000 in 136 families recorded by taxonomists [Catalog, 2024].

Spiders are renowned for their remarkable ability to produce silk, a protein fiber with diverse functions. Spider silk is used to create webs for prey capture, nests, and cocoons, and for various other purposes such as suspending themselves, floating, or gliding away from predators. The elaborate structure of spider silk

contributes to its exceptional properties, making it one of the toughest materials known [Blackledge, 2012]. Research on spider silk has attracted significant attention due to its potential applications in various fields, including the development of high-performance materials and biocompatible hybrids [Kiseleva et al., 2020].

The hierarchical organization and unique properties of spider silk have sparked interest in understanding its natural structuring and synthesis. This natural fiber has also garnered attention for its potential in engineering novel polymeric materials inspired by nature, highlighting the significance of unraveling the structure-function relationship of spider silk proteins [Römer and Scheibel, 2008].

Given that each spider can produce between two and seven different types of silk, the number of distinct spider silks exceeds 100,000 [Vollrath, 1992; Gosline et al., 1994]. These numbers demonstrate the remarkable biological variability of these materials, resulting in significant physical and chemical diversity.

Spider silk combines outstanding strength and extensibility, and a wide range of mechanical properties can be achieved with only slight changes in chemical structure [Vollrath and Porter, 2006]. For this reason, along with its intriguing humidity-driven actuation properties, spider silk is one of the most extensively researched natural materials. To give an idea, there are approximately 5000 new results per year on Google Scholar in the last three years that mention spider silk.

The complexity of spider silk requires the combined efforts of several disciplines, including biology, mechanics, materials science, chemistry, biochemistry, genetics and nanotechnology, to fully understand the unique properties of spider silk and explore its many applications.

The mechanical properties of spider silk, along with other physical properties such as thermal conductivity [Huang et al., 2012] and piezoelectricity [Karan et al., 2018], make it a compelling candidate for a wide range of technological applications. In addition to these properties, the biological characteristics of spider silk, including its ability to support cell adhesion and proliferation, as well as its resistance to bacterial proliferation, have positioned it as a top contender among potential biological materials for biomedical technologies [Kundu et al., 2014; Aigner et al., 2018; Holland et al., 2019]

In recent decades, scientific exploration has extended beyond the use of spider silk fibers, with a focus on biomedical applications utilizing various forms of spider silk. These applications typically originate from a spider silk solution, which can be processed through chemical treatments to produce nanospheres or nanoparticles. These nanostructures have demonstrated potential in drug delivery [Lammel et al., 2011], as well as in the development of films [Hümmerich et al.,

2006], hydrogels [Schacht and Scheibel, 2011], and foams [Schacht et al., 2016], all of which hold promise for diverse biomedical technologies [Kluge et al., 2008].

1.2 Multiscale models and soft materials

In the following we provide an introduction to multiscale modelling, analysing the advantages and disadvantages of different methodologies and providing a classification of the most common techniques. We also introduce the main concepts regarding modelling of soft materials, being the foundation from which the field of modelling of biological and hierarchical materials has been developing over the last decades.

1.2.1 Importance of multiscale models for scientific and biomimetic interests

The concept of multiscale modelling has been gaining ground in recent decades in a variety of fields, from complex systems such as aircraft to the field of materials, where it is enabling major advances in the design of new materials and metamaterials and in the full understanding of the behaviour of complex natural systems. For example, the gecko adhesion, despite being studied for several years, has been effectively described only within the framework of the multiscale modelling [Chen and Pugno, 2013].

Multiscale theoretical models provide a framework to comprehend the intricate relationships between the different length scales within hierarchical materials. Multiscale modelling aims to describe procedures that simulate continuum-scale behaviour using information from computational models of finer scales in the system, rather than relying on empirical constitutive models. This approach allows scientists to bridge the gap between the macroscopic properties of a material and its underlying micro and nanostructures.

By integrating information from various scales, multiscale models enable the prediction of material properties with improved accuracy [Jancar et al., 2010]. This is crucial for designing materials with tailored characteristics for specific applications. The work by Jancar et al. [Jancar et al., 2010] exemplifies the application of multiscale models in understanding of the fundamental physical relationships between nanoscale structural variables and the macroscale properties of polymer nanocomposites.

Understanding the dynamic behavior of hierarchical materials requires consideration of interactions across different scales. Multiscale modeling, facilitates the

exploration of dynamic phenomena, such as deformation and failure mechanisms, providing insights that may not be apparent at individual scales [Weinan, 2011].

The importance of multiscale modelling is emerging also with recent studies on biological materials, where complex relationships between structures and functions make it possible to obtain extraordinary macroscopic compositions in the face of microscopic components that in themselves have apparently common or even weak properties, as described above.

How nature develops the mechanical strength of natural materials is often still unknown, but an interesting insight pursued by scientists is that the form and microstructure of natural materials are intimately linked by their common origin, the growth of the organ [Elices, 2000]. Growth therefore means that form and microstructure are generated as part of the same process [Fratzl and Weinkamer, 2007].

The path to artificially obtain inspired materials with such interesting properties is still long owing to the intrinsic difference in how nature and men produce materials. In fact, while men design and construct a material according to some previously determined specifications, in a static process, nature follows a dynamic process in that biological structures are able to reshape to adapt to changes in external conditions [Fratzl and Weinkamer, 2007]. For instance, it is enough to think of a tree that can adapt its growth throughout its whole life in response to environmental changes such as the wind direction [Gardiner et al., 2016].

A further intrinsic advantage of biological structure is to generally include several materials. Indeed, it was elucidated that combining structural hierarchy and mixing of various materials can result in superior resistance, which cannot be achieved only by the hierarchy in a homogeneous phase [Bosia et al., 2012].

Several of these characteristics of natural materials have been effectively described within the framework of multiscale modelling. References such as [Mielke, 2006] and [Banasiak and Miekisz, 2008] showcase the efficacy of multiscale modeling in achieving a comprehensive understanding of materials behavior. The significance of developing multiscale theoretical models for understanding hierarchical material is therefore of paramount importance, contributing both to scientific inquiry and inspiring innovative biomimetic design principles.

1.2.2 Classification of multiscale modelling methods

Multiscale modeling involves deriving equations, parameters, or simulation algorithms to depict behavior at a specific length scale based on the physics at a finer scale. This finer scale encompasses electrons, atoms, molecules, and their assem-

blies, as well as mesoscale structures like phases or grains. The different scales may be defined with several terms, for example micro, atomistic or discrete scale for the lowest scale and macro, coarse, continuum scale for the highest. Between these two extreme scales may be defined several mesoscales.

A large number of such methods have been developed, taking a range of approaches to bridging across multiple length and time scales. Here we introduce some key concepts of multiscale modelling and review a selection of methods from several categories of models, including techniques developed in recent years that integrate new fields such as machine learning and materials design. In previous reviews of multiscale approaches [Fish et al., 2021], two methods are distinguished, the resolved-scale and the upscaling.

In the resolved-scale methods, fine-scale and coarse-scale responses, that may contain diverse physics, are computed simultaneously and the flow of information among scales is located at the interface. This method can be outlined as global-local approach. In particular, separate physics models and discretization schemes are typically used in different spatial domains, so that multiple scales can be simulated concurrently. This approach is particularly useful when the fine scale is important in small subregion of the problem.

A popular method in this direction is the Quasicontinuum (QC) method, first conceptualized by Tadmor, Ortiz, and Phillips in 1996 [Tadmor et al., 1996]. Aimed at simulating an atomistic system without explicitly addressing each individual atom, QC establishes a framework in which certain degrees of freedom are selectively eliminated, expediting force/energy calculations. This approach is complemented by adaptive model refinement to ensure the preservation of full atomistic detail in specific regions where necessary, while employing continuum assumptions to reduce computational demands elsewhere.

The upscaling is usually employed to manage hierarchies and it consists of finding relationships that characterize the behavior at a specific length scale, based on the physics at a smaller size, on condition that at a smaller scale, physics and structure are more clearly understood. The upscaling methods may include approaches where the same physics is employed at different scales, usually referred as homogeneization methods, as well as approaches where different physics are employed at different scales. In recent years, data-driven upscaling methods have emerged, where the macro-scale response is inferred through machine learning or automated model-order reduction. In these methods, the macroscale is typically constructed from a large number of data simulations at lower scales using data modelling approaches. In this case, a closed-form description of interscale dependence is typically not required. This thesis addresses the issue of the use of

data-driven techniques for the improvement of 'classical' physics-based models, and outlines a new methodology for this purpose in section 4.

More in detail, indeed, although machine learning has been effectively used to automate data processing and ensure high accuracy and repeatability of outcomes, it ignores basic physical principles, which may result in incorrectly presented problems or non-physical solutions. On the other hand, multiscale modeling has proven to be a successful method for combining multiscale, multiphysics data and revealing mechanisms that underpin the origin of function. Multiscale modeling, on the other hand, typically falls short of effectively combining large datasets from numerous sources and resolution levels. In the section 4, we explore potential synergies between machine learning and multiscale modeling to produce robust prediction models that take into account the underlying physics to handle challenging problems within the framework of materials science.

1.3 Modelling of soft biological materials

In the field of solid mechanics, the theory of elasticity is generally used to refer to the study of materials whose behaviour is described by Hooke's law, the simplest constitutive equation, describing the linear stress-strain curve of materials that undergo small deformations. It fits well with ordinary 'hard' materials such as metals or crystalline solids, where elasticity is given by binding forces that hold atoms in equilibrium positions. Equilibrium is achieved by balancing the attractive and repulsive forces between the particles, minimising the internal energy of the system [Euler, 2008].

Two major classes of materials behavior are excluded from such theory, namely, inelastic and elastic subject to large deformations. The latter includes the case of elastomers, which, according to the IUPAC definition, are polymers that display rubber-like elasticity [Alemán et al., 2007], for which the theory of non-linear elasticity is needed. Many biological soft materials, among which the spider silks, exhibit rubber-like elasticity, and due to the fact they are composed of macromolecules, they are called biopolymers.

For such materials, the elastic force does not originate from binding energy, but their behavior originates from an entropy-driven elasticity. Roughly speaking, rubber consists of flexible macromolecules that are in continuous Brownian motion at normal temperatures due to thermal agitation. The origin of rubber elasticity is rooted in the fact that when we apply a force to a piece of rubber, the long molecules in the network are stretched and if this force is released they spring back to random shapes. In other words, the external work done during

deformation is required to force the polymer molecules from a more probable into a less probable state. The restoring force arises from the tendency to return to conformations with higher entropy. During this process, molecular bonds are not stretched, and as a result, the internal energy of the molecules remains unchanged. Since there are fewer ways to realise a stretched molecular conformation, the polymer chain reduces entropy during mechanical stretching.

The interplay between the two described types of elasticity gives rise to the complex behaviour of technologically and naturally relevant 'soft' materials like elastomers and biological matter.

It is noteworthy that nonlinear elasticity is generally required to model large deformations of rubbery materials even in the elastic range. For even higher stresses, materials exhibit plastic behavior, that is, they deform irreversibly. In order to study rubber, it is important to be aware of both non-linear elasticity theory and the inelastic behaviour of bodies. Plasticity theory is particularly useful in distinguishing between plastic and elastic deformations, which rubber-like materials often experience simultaneously.

Biological materials are often modelled based on the assumptions of incompressibility, isotropy and hyperelasticity, as we will detail below. The hyperelastic model is a special case of nonlinear elasticity that allows for certain assumptions and simplifications that have been validated by a large body of experimental evidence.

1.3.1 Rubber-like materials

Modelling of soft biological materials is rooted in fundamental work that has been done over the last century to explain the behaviour of rubber and rubberlike materials.

A classical definition of rubberlike material, given by Treolar [Treloar, 1975], includes the presence of long-chain molecules, with freely rotating links, weak secondary forces between the molecules, and an interlocking of the molecules at a few places along their length to form a three-dimensional network.

In its virgin state, natural rubber is a highly viscous liquid. To give it a permanent shape, the raw material is crosslinked by adding a filler, commonly black carbon, to interlink molecular sequences or strands in the polymeric network. Crosslinking makes the rubber stronger and more rigid, but still very elastic.

Mechanical behaviour: key features

Experiments show that the elasticity, at least for small strains, is of the entropic type and is regulated by unfolding and refolding effects in the macromolecules [Treloar, 1975]. This is evidenced by the observation that the tension in stretched rubber at a fixed length linearly increases with temperature [Meyer and Ferri, 1935].

At increasing stretch, two major effects must be taken into account. On one hand, the behavior of the chains becomes non-Gaussian, meaning that the simple approximation of a Gaussian distribution for the probability of the end-to-end vector of a single chain is no longer valid. On the other hand, complex inelastic phenomena are observed.

Furthermore, at the macroscale, the elastic modulus of crosslinked rubber decreases as the material is stretched, and this effect increases with the maximum stretch. This damage effect is commonly known as the Mullins effect [Mullins, 1948] and has been extensively analysed in the literature [Diani et al., 2009].

A mesoscopic interpretation of the softening behaviour observed in cross-linked rubber-like materials was first proposed by Bueche in 1960 [Bueche, 1960]. According to Bueche's interpretation, the macroscopic softening results from the alteration of the polymeric network at the mesoscale, including the breaking of weak bonds between rubber molecules and filler particles. Therefore, the elasticity properties of a rubber specimen depend on the degree of previously attained maximum strain.

Bueche's perspective implies that a rubber specimen under strain becomes an anisotropic material, with the elastic modulus depending on the direction of deformation, as chains oriented in different directions undergo different strain histories. Also Mullins observed this phenomenon in his pioneering work [Mullins, 1948]. The breaking and reforming of bonds at the microscopic results in a continuous variation of the natural (unstressed) configuration of the network, which is discussed in detail in [Rajagopal and Srinivasa, 2004] and [Rajagopal and Wineman, 1992].

Complex models and theoretical discussions have been developed from such experimental evidences. The presence of residual strains associated with damage anisotropy, as shown in [Göktepe and Miehe, 2005; Diani et al., 2006], has been demonstrated in numerical analyses of anisotropic damage models for rubber materials.

Many biological soft materials exhibit similar behaviour, achieving their remarkable mechanical properties through clever microscale organization, where the material is often described as composed of hard pseudo-crystals and soft macro-

molecular network [Gosline et al., 1984; Ji and Gao, 2004] In particular, for the case of protein materials, of interest for this thesis, typical hard pseudo-crystals are achieved through specific protein secondary structures such as α-helices and β-sheets. These structures have two significant effects: first, an initial stiffening of the material, similar to the effect of additives in rubber, and second, a hidden length effect due to the phenomenon that as the material stretches, they may unfold [Fantner et al., 2006; Puglisi et al., 2017].

To meet the increasing demand for models that associate the microscopic properties with the macroscopic behaviour of the material, the theoretical framework of multiscale rubber elasticity has been pivotal, in increasingly deep collaboration between classical and new aspects of continuum mechanics, taking into account relevant non-linear and history-dependent effects [Puglisi and Saccomandi, 2016].

Main model classes

A first fundamental step was accomplished around 1930 [Guth and Mark, 1934; Kuhn, 1934] when molecular models were introduced to model the mechanical behaviour of natural rubber and its previously introduced properties. However, these models failed to describe the complex non-linear effects and history dependence of rubber-like materials, and to treat the large deformation regime in a proper and general framework.

The Mooney paper in 1940 [Mooney, 1940] and subsequent work by Rivlin [Rivlin, 1947], which introduced the possibility of describing non-linear elasticity by introducing the strain energy function, marked a major development in the field.

A further significant advance was then made with the advent of non-Gaussian molecular theory. These theories provide a valuable tool for gaining a deeper theoretical understanding of how low-scale properties impact macroscopic behavior. On the other hand, such theories may not capture important effects or lead to very expensive computations. A classical example is the absence in isotropic incompressible models of the dependence on the so-called C_2 energy term, which we will introduce in the following (see [Puglisi and Saccomandi, 2016] for a detailed discussion on this topic). Therefore, a synergy between phenomenological and statistical multi-scale models is needed. In this respect, there have been several important advances in the description of macroscopic behaviour with complex phenomena, such as rate and damage effects, residual stretching, growth and remodelling, which are still under investigation.

Statistical mechanics provided a further framework for describing the entropic character of network elasticity extending the above-described ones to encompass

non-Gaussian networks, enthalpic and non-equilibrium effects, and alterations in natural configurations. Thus, statistical mechanics led to a new class of multi-scale models that derive how macroscopic constitutive laws depend on microscale parameters.

These models address a limitation of typical phenomenological models, which often lack a direct relation to the underlying molecular level. Indeed, the form of such models, including energy function of the deformation invariant (often in polynomial form), as we will see in detail in the following, make them useless for a direct design of the material [Trentadue et al., 2021], and have often lead to fitting numerical problems and experimental problems to find the corresponding parameters [Destrade et al., 2017].

1.3.2 Basic equations and strain energy functions

In this section, we recall the basic equations and some essential aspects of nonlinear elasticity, of interest in the following analytical treatment for the modelling of rubber-like materials, in order to provide a rigorous and quantitative description of some of the previous remarks.

We start by considering the deformation of a body $\Omega, \Omega \ni X \mapsto x = x(X)$, its gradient $\boldsymbol{F} = \mathrm{Grad}(x)$ and the left Cauchy-Green deformation tensor $\boldsymbol{B} = \boldsymbol{F}\boldsymbol{F}^{\mathrm{T}}$. The principal invariants of $\boldsymbol{B}$ are

$$I_1 = \mathrm{tr}(\boldsymbol{B}), \quad I_2 = \frac{\left(I_1^2 - \mathrm{tr}\left(\boldsymbol{B}^2\right)\right)}{2} \quad \text{and} \quad I_3 = \det \boldsymbol{B}. \tag{1.1}$$

For a hyperelastic and isotropic material [Ogden, 1997], we may introduce a strain-energy density function $W = W\left(I_1, I_2, I_3\right)$.

It should be noted that the typical moduli of linear elasticity, such as the tensile modulus or the bulk modulus, can be derived in the limit of infinitesimal deformation. As the bulk compression modulus of rubber is quite high compared to its tensile modulus, it is common to treat rubber as incompressible. As in this case, only isochoric deformations are admissible, it results $I_3 = 1$, and $W = W\left(I_1, I_2\right)$.

In this setting, the Cauchy stress tensor $\boldsymbol{T}$ is given by:

$$\boldsymbol{T} = -p\boldsymbol{I} + 2\frac{\partial W}{\partial I_1}\boldsymbol{B} - 2\frac{\partial W}{\partial I_2}\boldsymbol{B}^{-1}, \tag{1.2}$$

where p is the Lagrange multiplier taking into account the constraint of incom-

pressibility. Moreover, the equilibrium equation in absence of body forces is

$$\operatorname{div} \boldsymbol{T} = \mathbf{0}, \tag{1.3}$$

with div the divergence operator with respect to x. The fundamental problem in this framework is to determine the functional form of the strain energy density W. In the following, we provide an overview of the most significant strain energy functions.

Neo-Hookean

The neo-Hookean model introduce one of the simplest strain energy functions. It involves a single parameter and provides a mathematically simple constitutive model for the non-linear deformation behavior of isotropic rubber-like materials. Its strain energy function is

$$W = \frac{\mu}{2}(I_1 - 3) \tag{1.4}$$

where $\mu > 0$ is the shear modulus for infinitesimal deformations.

This model is satisfactory in describing the macroscopic mechanical behaviour of rubber elasticity only for small and moderate shear deformations [Ogden, 1997]. Nevertheless, it was of fundamental importance in understanding many theoretical aspects of nonlinear elasticity theory because of its analytical simplicity.

Mooney–Rivlin

Rivilin improved the fitting to data by introducing a dependence of the strain energy function on both the first and second invariants. This resulted in a more general model than the neo-Hookean model, known as the Mooney-Rivlin model. The strain energy function is assumed to be linear in both the first and second invariants of the Cauchy-Green strain tensor. Mooney developed this model purely from a phenomenological perspective [Mooney, 1940]. For an incompressible material, the strain energy density function can be expressed as:

$$W = C_1(I_1 - 3) + C_2(I_2 - 3), \tag{1.5}$$

with C_1 and C_2 two material constants to calibrate in order to fit experimental data.

According to Mooney [Mooney, 1940], the form 1.5 is the most general for large deformations of an incompressible hyperelastic, undeformed isotropic material.

The Mooney-Rivilin model has been extensively studied, despite not accurately describing any rubber-like material within experimental error. It is com-

monly used as the first example for general results of isotropic incompressible materials, for which several analytical solutions have been found.

Generalized neo-Hookean models and Gent Model

Although Rivlin proposed introducing the dependence of W on the second invariant I_2, there are several models of strain energy functions that depend on the first invariant I_1 only. It is worth noting that in molecular theory, I_1 is related to the mean squared end-to-end distance of the chains, but in general the chains cannot have a completely arbitrary shape and length. Overcoming this restriction requires the second invariant I_2, which is instead related to the surface area of the material. However, introducing this invariant may lead to complications in the calculations, leading to the widespread use of strain energy functions which depend only on the first invariant in a non-linear way. These functions are called generalized neo-Hookean models and have the following general form:

$$W = W\left(I_1\right). \tag{1.6}$$

To consider the finite extensibility of the polymeric chains constituting the elastomer network, some models of the form 1.6 introduce a distribution function for the end-to-end distance of the polymeric chain that is not Gaussian. These models are usually referred to as non-Gaussian models.

From a phenomenological perspective, these models can be divided in two categories: power-law models and models with limiting chain extensibility.

An example of the first class, widely used in biomechanics, was proposed by Fung [Fung, 1967] as follows

$$W = \frac{\mu}{2b}\exp\left[b\left(I_1 - 3\right) - 1\right] \tag{1.7}$$

where $b > 0$ is a dimensionless stiffining costant, and $\mu > 0$ is the shear modulus for infinitesimal deformations.

An example of the second class, proposed by Gent [Gent, 1996] and of particular relevance for the modelling approach adopted in this thesis, is based on the concept of limiting chain extensibility. In this case, the strain energy density function is structured in such a way that it exhibits a singularity when the first invariant of the left Cauchy-Green deformation tensor reaches a limit value I_m. Such strain energy density function is given by

$$W = -\frac{\mu J_m}{2}\ln\left(1 - \frac{I_1 - 3}{J_m}\right) \tag{1.8}$$

where μ is the shear modulus and $J_m = I_m - 3$, with $I_m > 0$ a limiting constant value for I_1, accounting for the limiting chain extensibility.

Notice that in the limit where $I_m \to \infty$, the Gent model reduces to the Neo-Hookean model.

Roughly speaking, for the multiscale approach that we detail in Chapter 2, a Gent-type constitutive law will be employed, with a history-dependent limit threshold and a variable natural configuration, based on the classical multiplicative decomposition of the deformation gradient in a growth and in an elastic component.

Note that the model 1.8 may not be suitable for accurately predicting the mechanical properties of rubber-like materials across the full range of deformations. Instead, it is designed to reflect material behaviour for large strains [Pucci and Saccomandi, 2002].

Constitutive equations of the form 1.6 often lead to closed-form analytical solutions for many interesting problems (see [Horgan and Saccomandi, 1999b; Horgan and Saccomandi, 1999a; Horgan and Saccomandi, 2001c; Horgan and Saccomandi, 2001a; Horgan and Saccomandi, 2001b; Horgan and Saccomandi, 2002]).Such solutions are useful for understanding the mechanical properties of materials and as benchmarks for more complex numerical calculations.

Other models

For the sake of completeness, we also report on other successful models and approaches in the modelling of rubber-like materials. In 1951, Rivlin and Saunders [Rivlin and Saunders, 1951] pointed out that neither the neo-Hookean nor the Money-Rivlin models were adequate to precisely describe the experimental properties of typical rubber materials. They found that $\partial W/\partial I_1$ is independent of both I_1 and I_2, and that $\partial W/\partial I_2$ is independent of I_1 and decreases with increasing I_2. They thus derived the following form for the strain energy function:

$$W = C\,(I_1 - 3) + f\,(I_2 - 3)\,, \tag{1.9}$$

where C is a constant while f is a function having a decreasing slope as I_2 increases. In the work of Obata [Obata et al., 1970], it is found that neither $\partial W/\partial I_1$ nor $\partial W/\partial I_2$ can be considered constant, and therefore that each should depend on both I_1 and I_2 [De Pascalis, 2010].

Moreover, a generalization of the Gent model (Eqn 1.8) is the so-called Gent-Gent model proposed by Pucci and Saccomandi [Pucci and Saccomandi, 2002]. They improve the Gent model by introducing the I_2 invariant so that the strain

energy function becomes

$$W = -C_1 J_m \ln\left(1 - \frac{I_1 - 3}{J_m}\right) + C_2 \ln\left(\frac{I_2}{3}\right) \tag{1.10}$$

with C_1 and C_2 two constants. Pucci and Saccomandi proposed this model using a purely phenomenological methodology, obtaining successful predictions for the behaviour of rubber-like materials [Ogden et al., 2004].

Several authors have pointed out that Rivlin's principal invariants approach is not very practical for fitting experimental data because of the possible propagation of experimental errors (see for example [Valanis and Landel, 1967]). Therefore, it may be interesting to consider the possibility of expressing the strain energy directly in terms of the principal stretches. For this reason, Valanis and Landel [Valanis and Landel, 1967] postulated the strain energy function as a sum of functions each depending on a single stretch:

$$W = w(\lambda_1) + w(\lambda_2) + w(\lambda_3), \tag{1.11}$$

in which the function $w(\lambda)$ is, by symmetry, the same for each of the extension ratios, in the case of isotropic material.

Accordingly, Ogden [Ogden, 1972] proposed the following form in terms of powers of the principal stretches $\lambda_j, j = 1, 2, 3$, as:

$$W = \sum_{p=1}^{N} \frac{\mu_p}{\alpha_p} \left(\lambda_1^{\alpha_p} + \lambda_2^{\alpha_p} + \lambda_3^{\alpha_p} - 3\right) \tag{1.12}$$

with μ_p and α_p two material constants. Assuming incompressibility, we can rewrite as

$$W(\lambda_1, \lambda_2) = \sum_{p=1}^{N} \frac{\mu_p}{\alpha_p} \left(\lambda_1^{\alpha_p} + \lambda_2^{\alpha_p} + \lambda_1^{-\alpha_p}\lambda_2^{-\alpha_p} - 3\right) \tag{1.13}$$

In general, the shear modulus is obtained from

$$2\mu = \sum_{p=1}^{N} \mu_p \alpha_p. \tag{1.14}$$

The material behaviour of rubbers can be described very accurately by the Ogden model with $N = 3$ and by calibrating the material parameters. Interestingly, for specific values of material constants, the Ogden model reduces to either the Neo-Hookean ($\alpha = 2$) or the Mooney-Rivlin model ($N = 2, \alpha_1 = 2, \alpha_2 = -2$, with the constraint condition $\lambda_1\lambda_2\lambda_3 = 1$).

Jones and Treloar [Jones and Treloar, 1975] and Ogden [Saccomandi and Ogden, 2004] show the consistency of biaxial strain experiments with the Valanis-Landel 1.11 and Ogden models.

A different approach is the so-called Rivlin-Signorini method. First Murnaghan [Murnaghan, 1937] and then Rivlin [Rivlin, 1960] and Signorini [Signorini, 1943] approximated the material response functions by polynomials in the appropriate invariants. Thus constant coefficients of the polynomial rather than the functions characterise a given material. Applications of the Rivlin-Signorini method can be found in [Martin and Carlson, 1977; Singh, 1967].

Eventually, a further two parameter limiting chain extensibility model with a dependence on I_2 has been proposed by Horgan and Saccomandi [Horgan and Saccomandi, 2004] where

$$W = -\frac{\mu}{2} J \ln\left(\frac{1}{(J-1)^3}\left(J^3 - J^2 I_1 + J I_2 - 1\right)\right), \tag{1.15}$$

or in terms of the principal stretches of the deformation

$$W = -\frac{\mu}{2} J \ln\left(\frac{\left(1-\frac{\lambda_1^2}{J}\right)\left(1-\frac{\lambda_2^2}{J}\right)\left(1-\frac{\lambda_3^2}{J}\right)}{\left(1-\frac{1}{J}\right)^3}\right). \tag{1.16}$$

where μ is the shear modulus for infinitesimal deformations and the parameter $J > 1$ is a new limiting chain parameter for which

$$\max\left(\lambda_1^2, \lambda_2^2, \lambda_3^2\right) < J \tag{1.17}$$

We note that such constraint is in terms of the principal stretches rather than the first invariant as in 1.8. The model 1.16 is more complex in form but has two advantages over the basic Gent model. First, as pointed out in [Horgan and Saccomandi, 2002; Horgan and Saccomandi, 2006], the limiting chain condition expressed in terms of the principal invariant is less physically accessible than 1.17. Furthermore, the lack of dependence on the second invariant in the basic Gent model implies some physical limitations (see, e.g., [Horgan and Smayda, 2012; Puglisi and Saccomandi, 2016] for a discussion of the important role of the second invariant).

1.4 Objectives of the thesis

The main objective of this thesis, given the particular importance of the multiscale description of biological materials described above, is to describe in terms of multiscale models some of the mechanisms that regulate the behaviour of biological materials, with a focus on spider silk, one of the most studied hierarchical materials, also in relation to different conditions of humidity and temperature.

The investigation begins by examining the primary proteins that constitute the material. Subsequently, the structures that these proteins typically form and their arrangement within the material at the macroscopic scale are explored. In this case, the material is a thread of spider silk.

Once the structures associated with certain mechanical properties have been identified, this study aims to investigate how temperature and humidity affect these structures. As is common with many natural materials, including spider silk, the literature shows that temperature and humidity have a significant impact on their structure.

The mechanical behaviour of the material, i.e. the stress-strain behaviour, will be investigated, as well as the effect of loading history, a key issue for biological polymers, as well as for polymers in general and elastomers in particular, as discussed above. The impact of temperature and humidity-induced structural changes on the mechanical behaviour of the material will also be investigated.

As a second goal, we want to investigate the possibility of exploiting the extraordinary property of spider silk to contract significantly in a humid environment in order to construct humidity-driven actuators. A multiscale perspective will be employed once again, starting from the interaction of water molecules with the proteins that compose the material and leading to a description of the macroscopic behaviour.

Therefore, the different proteins will be described, especially concerning their secondary structures, which are considered to be the most responsible for regulating the behaviour of silk when it undergoes a transition from a dry to a wet environment. To ensure the validity of the proposed model, it will be continuously compared with experiments on the interaction of spider silk with a wet environment available in the literature.

A multiscale description of the mechanisms leading to the macroscopic behaviour of the moisture-driven spider silk actuator will then allow exploring the influence of the different model parameters, all of which have a clear physical interpretation, on the actuator properties. In other words, we want to investigate how, by modifying the structure at the lower scales, for instance by engineering

chains of amino acids of different lengths or bonds between molecules of different strengths, we can vary the response of the actuator to meet different requirements that can be defined at the actuator design stage.

It is interesting to contextualize the actuation properties of the humidity-induced contraction of spider silk with other actuators both humidity driven and with other operating principles. In comparison to other actuators, the supercontraction of spider silk will exhibit surprising properties.

We will also explore the possibility of different types of actuated devices, including the simpler case of a linear spring and the more complex case of a bistable spring to maximise the amount of obtained work.

In recent years, there have been significant advancements in two areas, not only in the field of materials. These areas include experimentation at increasingly smaller scales and exponential growth in computing power. The production of data has increased significantly in terms of both quantity and quality. The processing of such data appears to be seamless.

In fact, the processing of data made possible by newly available computing power has significant drawbacks and unresolved problems. The main issue is the interpretability of the results obtained. The most widespread data modelling technique, artificial neural networks, allows for great performance in processing data and finding possible correlations between them. However, the results are difficult to be interpreted due to the inherent nature, commonly referred to as the black box, of the ANN approach, which will be discussed in more detail later.

In this context, our aim is to acknowledge the availability of vast amounts of data and computing power, but to keep the scientist, who is able to produce physically interpretable models, at the center of the knowledge production process. In other words, our purpose is to develop a methodology that enables the processing of the large amounts of data available in order to deliver results that can be interpreted by the scientist who is the expert in the problem under investigation, i.e. in such a way that effective progress can be made in the theoretical knowledge of the problem under investigation.

In order to describe the proposed methodology in a detailed and operational way, the spider silk will again be used as an example of biological material. The hierarchical nature of this material will be considered to illustrate the potential advantages of using data modelling techniques that provide easily interpretable results, including the fundamental advantage of incorporating knowledge gained from modelling large amounts of data into models previously developed for the multiscale physical description of the material of interest.

This approach aims to explore the potential of harnessing newly available

computational power to advance our knowledge of not only the multiscale mechanisms that give rise to the extraordinary behaviour of spider silk, but also, more generally, of phenomena governed by multiscale hierarchical differential equations that can be treated in a similar way.

Chapter 2

Spider Silks Mechanics: Predicting Humidity and Temperature Effects

This chapter is based on the paper published in the Journal of the Mechanics and Physics of Solids (volume 164, 104857, 2022, DOI 10.1016/j.jmps.2022.104857) [Fazio et al., 2022] and deals with the mechanics of spider silk in different environmental conditions and the contraction of the spider silk in wet environments.

Abstract

We deduce a microstructure inspired model for humidity and temperature effects on the mechanical response of spider silks, modeled as a composite material with a hard crystalline and a soft amorphous region. Water molecules decrease the percentage of crosslinks in the softer region inducing a variation of natural length of the macromolecules. The resulting kinematic incompatibility between the regions crucially influences the final mechanical response. We demonstrate the predictivity of the model by quantitatively reproducing the experimentally observed behavior.

2.1 Introduction

Due to their extraordinary properties, spider silks represent one of the most intensively studied materials, also in the spirit of biomimetics [Zhao et al., 2014]. The availability of more and more sophisticated experimental techniques let in the last decades a deeper understanding –both from a chemical and structural point of view– of the complex multiscale, hierarchical material structure at the base of their notable mechanical behavior. Nevertheless, due to the complexity of its

behavior, many important phenomena regulating its loading history dependence, rate, temperature and humidity effects, remain unclear [Pérez-Rigueiro et al., 2021], especially when multiscale effects are taken in consideration. In this paper we propose a model for describing and predicting humidity and temperature effects and with a special focus on a crucial phenomenon known as *supercontraction* effect. Firstly addressed in [Work, 1977], it consists in a shrinkage of the fiber up to 50% of its initial length, when immersed in water or in high humidity environment strongly modifying the mechanical performances. As we show, the comprehension of such striking behavior let us deduce a predictive model for the experimental thermohygromechanical behavior of spider silks, starting from a description at the molecular scale.

At the molecular scale, spider silks are composed by an amorphous matrix of oligopeptide chains and by pseudo-crystalline regions made up principally of polyalanine β-sheets [Elices et al., 2011; Sponner et al., 2007] with dimensions between 1 and 10 nm [Keten and Buehler, 2010], mostly oriented in the fiber direction [Jenkins et al., 2013]. The cross section of the fiber is highly organized in the radial direction [Li et al., 1994; Eisoldt et al., 2011; Sponner et al., 2007]. Moreover, the chemical and structural composition varies according with the different silks produced by the different glands [Cranford et al., 2014] and of course the different species. Here, to fix the ideas, we focus on the most performant case of dragline silk.

More in detail, the thread is covered by a skin, with a chemical and physical protection function, that does not play a role in supercontraction and mechanical response [Yazawa et al., 2019]. We thus neglect it in the model. Next, the core can be schematically decomposed as in Fig. 2.1. The major constituent of the external part [Li et al., 1994; Brown et al., 2011] are proteins (Major ampullate Spidroin 1, MaSp1) organized into β-pleated sheets. We refer to this fraction as *hard region*. The internal part, here referred as *soft region*, is instead mainly constituted by proteins with a proline content preventing the formation of β-sheet structures [Sponner et al., 2007] (Major ampullate Spidroin 2, MaSp2). This fraction has a significantly lower crystallinity and macromolecules with weaker crystal domains, typically in the form of α-helix and β-turns [Sponner et al., 2007; Nova et al., 2010]. The different crystallinity is also due to the shear stress at the spinning duct wall inducing the formation of harder crystal domains in form of β-sheets mainly in the outer region [Giesa et al., 2016; Brown et al., 2011].

Based on the previous description and referred literature, we model the silk fiber as a composite material with a hard external fraction of crystalline chains and a soft internal fraction of amorphous chains. Moreover, by following the

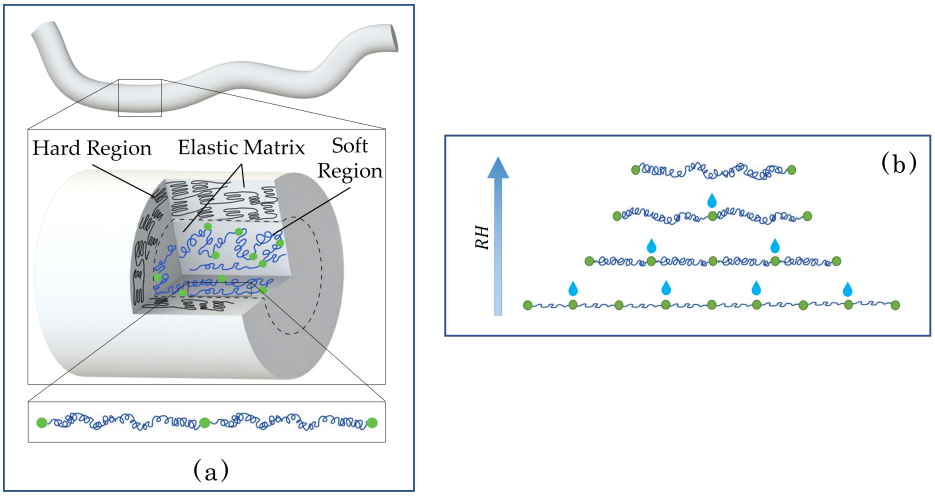

Figure 2.1: Cartoon of the silk fiber and its microstructure (a). The outer hard region characterized by crystals of β-sheets is represented by black chains, the soft region molecules are drafted in blue, whereas the embedding elastic matrix is grey. In (b) we schematize the incremental effects of water molecules (added as RH increases) disrupting crystal domains (green dots) and inducing entropic recoiling of chains.

classical approach for polymeric and biopolymeric materials [Flory and Erman, 1982], we suppose that the hard and soft fractions of macromolecules aligned with the fiber axis are embedded in a tridimensional elastic matrix, describing the complex macromolecular network with inner and intrachains connections.

Due to the different crystalline composition, the humidity affects differently the hard and soft fraction. Water hardly breaks the H-bonds of the compact β-sheet domains in the hard fraction [Yazawa et al., 2019]. On the other hand, here, we may observe a misalignment of the crystals with respect of the fiber direction that increases as the relative humidity RH grows [Eles and Michal, 2004]. Since the material stiffness grows with the alignment of the crystals [Du et al., 2006], water induces a progressive damage in this fraction that we introduce in our model. On the contrary, water content strongly influences the crystal percentage in the soft internal core [Du et al., 2006; Elices et al., 2005; Elices et al., 2011], because α-helices and β-turns are much more easily broken by water molecules. In particular, the experiments exhibit a non uniform variation of the silk properties with a localized transition at a specific value of RH, hereon indicated as RH_c, known as *supercontraction* threshold [Fu et al., 2009]. Finally, an important effect in the evolution of the natural chain length is induced from the stretch history (see [Puglisi et al., 2017] and reference therein for a theoretical detailed discussion of such phenomenon). Indeed, as the end-to-end distance of the molecules changes, β-sheets undergo unravelling with a corresponding increase of the number of available monomers, here considered in the hard fraction.

2.2 Model

According with classical Statistical Mechanics results [Rubinstein and Colby, 2003], the expectation value of the end-to-end distance for ideal chains is

$$L_n =< r^2 >^{1/2}= b\, n^{1/2}, \tag{2.1}$$

where n is the number of Kuhn segments with length b. We refer to the end-to-end distance as natural length for the reason that it is the length that the chain naturally assumes when no force is applied. As we will detail later n depends on humidity in the soft fraction, whereas it depends on the maximum attained stretch in the hard fraction. As a result the natural length of these fractions, assumed identical after spinning, vary depending on the external force and humidity fields, inducing a kinematic incompatibility as detailed in the following. We remark that in the recent paper [Cohen et al., 2021], based on this observation, the authors proposed a model with a material undergoing a hard→soft transition reproducing a mutation from glassy to rubbery state induced by RH variations, efficiently describing the variations of stiffness and supercontraction length in correspondence of the supercontraction threshold.

2.2.1 Soft Region

To consider the disruption process of H-bonds induced by hydration [Du et al., 2006], let us introduce the function $m = \hat{m}(\mathrm{RH})$ assigning the number of links in the generic humidity state, with $m(0) = m_o$ and $m(100) = m_f$ (initial and permanent number of H-bonds) [Vollrath and Porter, 2006]. To the knowledge of the authors, no direct measurement of $\hat{m}$ is available, so that we consider a Gaussian probability density of rupture events

$$d(\mathrm{RH}) = \frac{\hat{m}(\mathrm{RH}) - m_o}{m_f - m_o} = \int_0^{\mathrm{RH}} \frac{1}{\sqrt{2\pi\bar{s}^2}}\, e^{-\frac{(\mathrm{RH}-\mathrm{RH}_c)^2}{2\bar{s}^2}}. \tag{2.2}$$

Here $d \in (0,1)$ is a 'damage' type parameter, measuring the percentage of broken links and we assume that the Gaussian is centered in the critical value RH_c (see Fig. 2.2).

To obtain the corresponding variation of the natural length based on (2.1), assume that n_o^s is the (mean) number of chain free monomers when the silk is spun. Here and in the following we indicate by the apexes s, h, m, and t the soft, hard, matrix and homogenized (total) quantities. If we identify the number of H-bonds with the number of domains in which the chain is divided

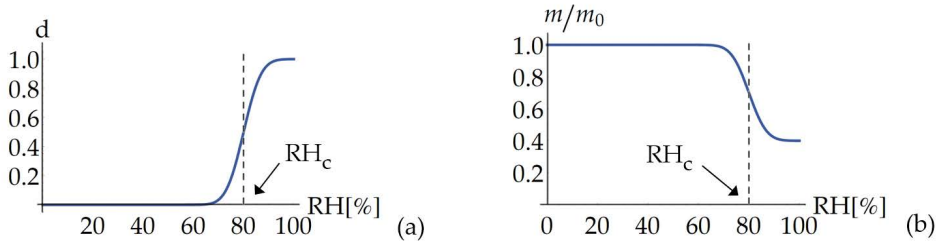

Figure 2.2: (a) Damage parameter d representing the percentage of broken links as a function of humidity. (b) Influence of the relative humidity on the number of H-bonds with respect to the initial number of H-bond. Here $\mathrm{RH}_c = 80\%$, $\bar{s} = 5.5$ and $m_f/m_o = 0.4$.

(see the scheme Fig. 2.1(b)), the mean number of free monomers in each domain is $n^s = \hat{n}^s(\mathrm{RH}) = n_o^s/\hat{m}(\mathrm{RH})$ corresponding to a *natural length*

$$L_n^s = \hat{m}(\mathrm{RH})\sqrt{n_o^s/\hat{m}(\mathrm{RH})}\, b^s = \sqrt{n_o^s\, \hat{m}(\mathrm{RH})}\, b^s. \tag{2.3}$$

We obtain in this way an analytic measure of the shrinkage chain effect induced by humidity. Observe that instead the *contour length* is fixed: $L_c^s = n^s\, b^s$.

2.2.2 Hard Region

As anticipated, following [Du et al., 2006] we assume that the elastic modulus of the crystalline region decreases with RH by considering a (phenomenological) damage function in the macroscopic model deduced in section 2.2.3. On the other hand, while β-sheet crystals are affected only in their orientation by humidity, large strain can induce important unravelling effects as fully described in [Puglisi et al., 2017] with conformational transitions inside the secondary structure [Cranford et al., 2014; Giesa et al., 2016], from a coiled configuration [Yarger et al., 2018] to an unfolded state [Cranford et al., 2014; Puglisi et al., 2017] (see the scheme in Fig. 2.3). Thus the mean number n^h of available free monomers depends on the maximum attained value of the end-to-end distance so that we assume the existence of a function $\hat{n}$ such that $n^h = \hat{n}^h(L_{max}^h)$. Observe that for simplicity we assume that the unfolding is irreversible with n^h monotonically increasing with L_{max}. More general hypotheses could be introduced [De Tommasi et al., 2010] and a discussion of this function will be given later.

As a result, for the hard fraction the natural and contour lengths are assigned

$$L_n^h = \sqrt{\hat{n}^h(L_{max})}\; b^h, \qquad L_c^h = \hat{n}^h(L_{max})\, b^h. \tag{2.4}$$

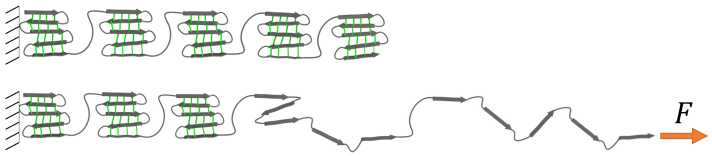

Figure 2.3: Cartoon of a chain of the hard region undergoing a folded → unfolded transition when subjected to a force.

2.2.3 From single chain to macro laws

Both in the amorphous and crystalline region we adopt a Worm Like Chain (WLC) energy density (per unit chain contour length L_c) in the form proposed in [De Tommasi et al., 2013] $\varphi_e = \varphi_e(L, L_c) = \kappa \frac{L^2}{L_c - L}$ where $\kappa = \frac{k_B T}{4 l_p}$, T is the temperature, k_B the Boltzmann constant and l_p the persistence length measuring the flexibility of the chain (see *e.g.* [Rubinstein and Colby, 2003] for details). This energy respects the limit extensibility condition, $\lim_{L \to L_c} \varphi_e(L, L_c) = +\infty$, and allows for explicit calculations. Moreover, following [Trentadue et al., 2021], we extend this function to consider that, as described above, the end-to-end distance L can be decomposed in a variable (zero force) natural length measured by (2.1) and the remaining length measuring the elastic elongation $L_e = L - L_n$. Thus we assume an energy and a force-elongation law for a single chain

$$\begin{aligned} \varphi_e &= \kappa \frac{L_e^2}{L_c - L} \\ f &= \frac{\partial \varphi_e}{\partial L} = \kappa \left[\left(\frac{L_c - L_n}{L_c - L} \right)^2 - 1 \right], \end{aligned} \tag{2.5}$$

with the force decreasing to zero as the length attains its (history dependent) natural length ($L = L_n$ or $L_e = 0$).

Notice that the total energy considering all the chains of the network can be determined as $\Phi = N_v \varphi_e L_c = N_a N_l \varphi_e L_c = N_a \varphi_e$ where for the sake of simplicity we consider $N_l L_c = 1$, with N_v, N_a, N_l the number of chains per unit volume, area, length respectively.

We remark that the proposed model can be inscribed in the theory of Thermodynamics with internal variables [Coleman and Gurtin, 1967] in the simple case when there is a single external variable L and a single internal variable L_{max}. In our simple setting of isothermal processes, to verify the thermodynamic consistency of the model we consider the Clausius-Duhem inequality, requiring the positivity of the dissipation rate $\Gamma = f\dot{L} - \dot{\varphi}_e(L, L_{max}) \geq 0$. Since at given RH the only material fraction involved in the dissipation is the hard one, undergoing unfolding effects and variations of the natural length regulated by L_{max} according with (2.4), the internal energy dissipation rate reduces to $Q'(L_{max}) = -\partial_{L_{max}} \varphi_e(L, L_{max})$.

Thus, in view of (2.5), we obtain $\Gamma = Q'(L_{max})\dot{L}_{max} \geq 0$ that is satisfied under our assumption that $\hat{n}^h$ is increasing.

Eventually, to obtain the macroscopic behavior of the thread we consider, following [Grubb and Ji, 1999], the classical *affinity hypothesis* [Rubinstein and Colby, 2003] that identifies the macroscopic stretches with the macromolecular ones. We then first evaluate the following stretch measures of the different fractions

$$\begin{aligned} &\lambda^i = \tfrac{L}{L^i_o} && \text{total stretch,} \\ &\lambda^i_e = \tfrac{L^i_e}{L^i_o} && \text{elastic stretch,} \\ &\lambda^i_n = \tfrac{L^i_n}{L^i_o} && \text{permanent stretch,} \\ &\lambda^i_c = \tfrac{L^i_c}{L^i_o} && \text{contour stretch,} \end{aligned} \qquad i = h, s, m, t, \tag{2.6}$$

with $L^i_o = b^i \sqrt{n^i_0}$ denoting the initial natural length of the chain. in particular, the permanent stretch measures the variation of the natural length, having the same role of plastic streth in classical non linear plasticity theories (see [De Tommasi et al., 2015] for a detailed theoretical discussion).

For the soft region, it can be deduced using Eqns. (2.3) and (2.6),

$$\lambda^s_n = \frac{L^s_n}{L^s_o} = \frac{\sqrt{n^s_o\ \hat{m}(\mathrm{RH})}\ b^s}{\sqrt{n^s_o\ m_o}\ b^s} = \sqrt{\frac{\hat{m}(\mathrm{RH})}{m_o}}. \tag{2.7}$$

If we consider a Gaussian distribution for the breaking rate of the bonds as in Eqn. (2.2) we get the following evolution of the permanent stretch as RH varies

$$\lambda^s_n = \sqrt{1 + d(\mathrm{RH})\left(\frac{m_f}{m_o} - 1\right)}. \tag{2.8}$$

A typical variation of the damage function and number of domains under our Gaussian probability choice is represented in Fig. 2.2, whereas the variation of the permanent stretch of the soft fraction is reported in Fig. 2.4. We remark that the parameters needed to compute the variation of the permanent stretch as a function of humidity are m_f/m_o, together with RH_c and $\bar{s}$ assigning the Gaussian distribution function. On the other hand, the corresponding expression for the contour length is $L^s_c = m\frac{n^s_o}{m}\ b^s = n^s_o b^s$, so that the contour stretch of the amorphous part is constant

$$\lambda^s_c = \frac{L^s_c}{L^s_o} = \frac{n^s_o\ b^s}{\sqrt{n^s_o\ m_o}\ b^s} = \sqrt{\frac{n^s_o}{m_o}}. \tag{2.9}$$

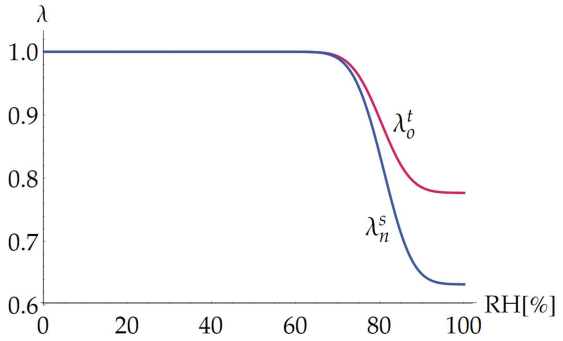

Figure 2.4: Dependence of the permanent stretch of the soft region λ_n^s from the relative humidity, in blue (assumed parameters $\bar{s} = 5.5$ and $m_f/m_o = 0.4$). Observe the abrupt decrease around the supercontraction threshold $\mathrm{RH}_c = 80\%$. For comparison, the unstressed stretch of the overall fiber λ_o^t is represented in pink (assumed parameters $E^s = 1.35$ MPa, $\mu = 2.5$ MPa, $\lambda_c^s = 2.62$).

Under an additive assumption and given the number of chains per unitary reference area N_a^s, the (Piola, engineering) stress using Eqns. $(2.5)_2$ and (2.6) is given by

$$\sigma^s = E^s \left[\left(\frac{\lambda_c^s - \hat{\lambda}_n^s(\mathrm{RH})}{\lambda_c^s - \lambda^s} \right)^2 - 1 \right], \tag{2.10}$$

where the permanent and contour stretches are given by Eqns. (2.8) and (2.9), whereas $E^s = N_a^s \kappa^s$ is the elastic modulus of the soft fraction.

For the hard region, the contour and permanent stretches are calculated by using Eqns. (2.4) and (2.6). Let n_o^h be the initial number of available free monomers, the initial natural length $L_o^h = \sqrt{n_o^h}\, b^h$ can be used to calculate the contour stretch as

$$\lambda_c^h = \frac{L_c^h}{L_o^h} = \frac{n^h b^h}{\sqrt{n_o^h}\, b^h} = \frac{n^h}{\sqrt{n_o^h}} = \frac{\hat{n}^h(\lambda_{max})}{\sqrt{n_o^h}} \tag{2.11}$$

and the permanent stretch as

$$\lambda_n^h = \frac{L_n^h}{L_o^h} = \frac{\sqrt{n^h}\, b^h}{\sqrt{n_o^h}\, b^h} = \sqrt{\frac{n^h}{n_o^h}} = \sqrt{\frac{\hat{n}^h(\lambda_{max})}{n_o^h}}. \tag{2.12}$$

Since the effective stretch-induced unfolding depends on the unknown size and strength crystals distribution, following [Trentadue et al., 2021] we assume a simple power law

$$\lambda_c^h = c_1 (\lambda_{max}^h)^{c_2}. \tag{2.13}$$

On the other hand, since by Eqn. (2.4) the permanent and contour stretches are related by $\lambda_n^h = \left(\lambda_c^h / \lambda_{co}^h\right)^{\frac{1}{2}}$, where $\lambda_{co}^h \equiv c_1$ is the initial contour stretch, by

Eqn. (2.13) the permanent stretch is given by

$$\lambda_n^h = (\lambda_{max}^h)^{\frac{c_2}{2}}. \tag{2.14}$$

It is important to stress that thanks to previously described microstructure based analysis of the permanent and contour stretches, they are analytically related, so that in this fraction the permanent stretches are directly related to damage, both measured by the unknown function $\hat{n}^h$, here assumed depending on the two only parameters c_1 and c_2.

If then, as anticipated, we consider a damage taking care of the described humidity induced crystal misalignment, the stress-stretch relation for the hard part by using Eqns. (2.5) and (2.6) is

$$\sigma^h = \hat{E}^h(\mathrm{RH}) \left[\left(\frac{\hat{\lambda}_c^h(\lambda_{max}^h) - \hat{\lambda}_n^h(\lambda_{max}^h)}{\hat{\lambda}_c^h(\lambda_{max}^h) - \lambda^h} \right)^2 - 1 \right] \tag{2.15}$$

with

$$\hat{E}^h(\mathrm{RH}) = (1 - \alpha\ \mathrm{RH}) E_o^h, \tag{2.16}$$

where $E_o^h = N_a^h \kappa^h$ is the modulus in the dry condition and α measures the humidity induced damage rate.

This macroscopic constitutive damage assumption is based on the experimental observation, recalled above, of a crystal misalignment induced by humidity. Indeed, the Wide-Angle X-ray Scattering (WAXS) measurements reported in Fig. 2.5(a) and reproduced from [Yazawa et al., 2020], indicate a linear dependence of the Full Width at Half Maximum (the statistical measure of the orientation of crystalline β-sheets adopted in [Yazawa et al., 2020]) from RH, indicating a lower orientation of the nano crystals at higher RH values. Such misalignment affects the elastic modulus described in [Du et al., 2006]. The resulting modulus variation assumed in (2.16) is represented in Fig. 2.5(b).

Eventually, by following [Flory and Erman, 1982], we consider a matrix embedding hard and soft fractions describing the complex macromolecular network composing the spider thread, with inner and intrachains connections. The total free energy is then calculated as the sum of the energy of ideally isolated hard and soft chains described above and elongated along the fiber direction plus an energy term taking care of chains interactions. Under a simple Neo-Hookean assumption (corresponding at the molecular scale to harmonic network chains [Rubinstein and Colby, 2003]) for incompressible material, the matrix stress for a uniaxial

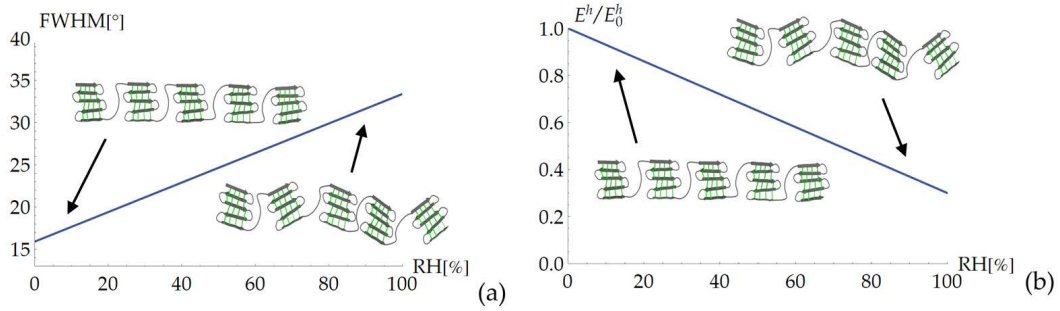

Figure 2.5: (a) WAXS measurements of dragline silk fibers at different RHs are used to calculate the orientation of crystalline β-sheets with respect to the fiber axis [Yazawa et al., 2020] with the result that FWHM linearly increases with RH. This proves a reduced orientation of the nano crystals under higher values of RH as schematized by the cartoons. (b) Corresponding damage function in (2.16): the reduction of the elastic modulus of the hard part is associated to the lack of orientation of the crystal as the humidity increases. Here we assume $\alpha = 0.007$.

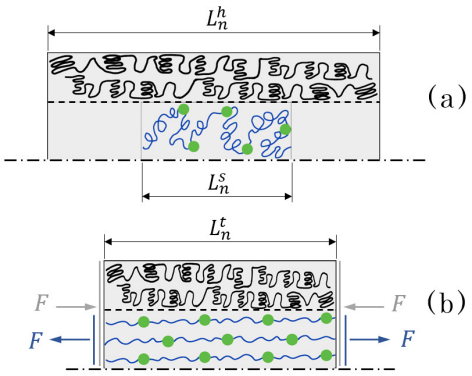

Figure 2.6: Outer and inner regions have different natural lengths (a) leading to a prestretch of the internal amorphous part (b).

extension is

$$\sigma^m = \mu \left(\lambda^h - \frac{1}{(\lambda^h)^2} \right), \tag{2.17}$$

with μ the material shear modulus. Observe that we assume that the matrix natural length coincides with the hard fraction initial one, so that $\lambda^m = \lambda^h$.

We are now in the position of deducing, based on all the microstructure based assumptions and the affinity hypothesis, the macroscopic behavior of the spider thread. Our deduction is based on the main assumption that the spun initial length of the fractions are the same: $n^h b^h = n^s b^s$ (kinematic compatibility). After exposition to humidity the soft region chains reduce their natural length (see Fig. 2.6). The kinematic compatibility then imposes that the different regions undergo the same stretch λ_o^t that corresponds to zero overall initial stress. On the other hand, when the fiber is subjected to a force $F > 0$, it undergoes

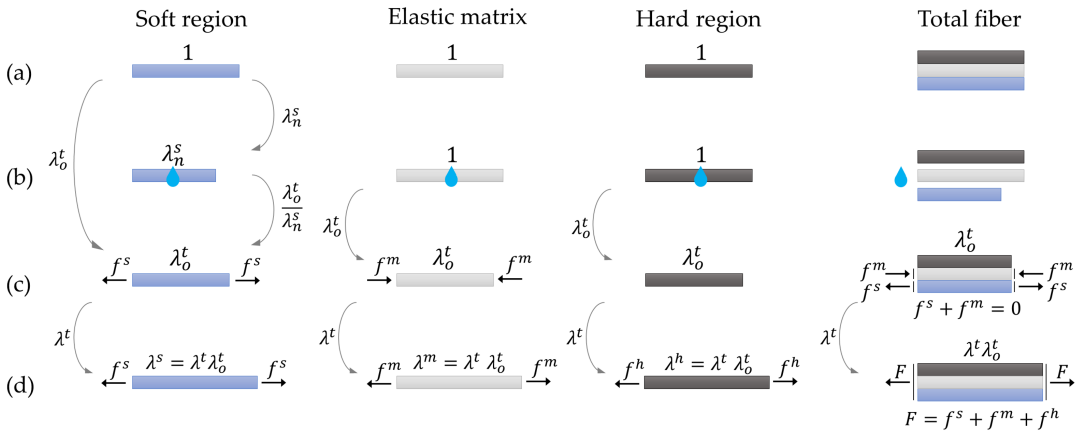

Figure 2.7: Scheme of the evolution of the configurations of the different composing phases and of the whole fiber: (a) dry natural configuration, (b) humid condition without external forces where selfstresses (reported in (c)) are induced by kinematic compatibility, (d) non zero external force configurations.

a stretch λ^t, starting from the new natural configuration λ_o^t. As a final result, the stretches for the soft region (that varies its permanent stretch according to humidity), hard region and matrix starting from the spun initial length (see the scheme in Fig. 2.7) are given by $\lambda^s = \lambda^h = \lambda^m = \lambda^t \lambda_o^t$, where λ^t represents the experimentally measured stretch.

The overall stress-stretch relation is

$$\sigma^t(\lambda^t) = \Theta(\lambda_n^h - \lambda^t \lambda_o^t)(1 - \alpha \text{ RH}) E_o^h \left[\left(\frac{\lambda_c^h - \lambda_n^h}{\lambda_c^h - \lambda^t \lambda_o^t} \right)^2 - 1 \right]$$

$$+ E^s \left[\left(\frac{\lambda_c^s - \lambda_n^s}{\lambda_c^s - \lambda^t \lambda_o^t} \right)^2 - 1 \right] + \mu \left(\lambda^t \lambda_o^t - \frac{1}{(\lambda^t \lambda_o^t)^2} \right) \tag{2.18}$$

where Θ is the step function considering that the hard fraction chains are not able to sustain any compressive force ($\sigma^h = 0$ if $\lambda^h < \lambda_n^h$). Observe that from this equation at $\lambda^t = 1$ and $\sigma^t = 0$ we determine λ_o^t.

In Fig. 2.8 we describe the behavior of the model here proposed, representing the stress-stretch curves for different humidity conditions (RH $= 0\%, 70\%, 85\%, 90\%, 100\%$). In the dry condition (RH $= 0\%$) the natural length of the hard and soft fraction coincide ($\lambda_n^h = \lambda_n^s$) and the hard fraction participates to the mechanical response from the beginning. Consequently the force is carried mainly by the much stiffer hard fraction for all the elongation set. The behavior is similar for RH $<$ RH$_c$ where a lower stiffness is due to the hard domain disorder inducing a damage according with Fig. 2.5 and Eqn. 12 (see the curve RH $= 70\%$). Once RH $>$ RH$_c$, the role played by the amorphous fraction becomes much

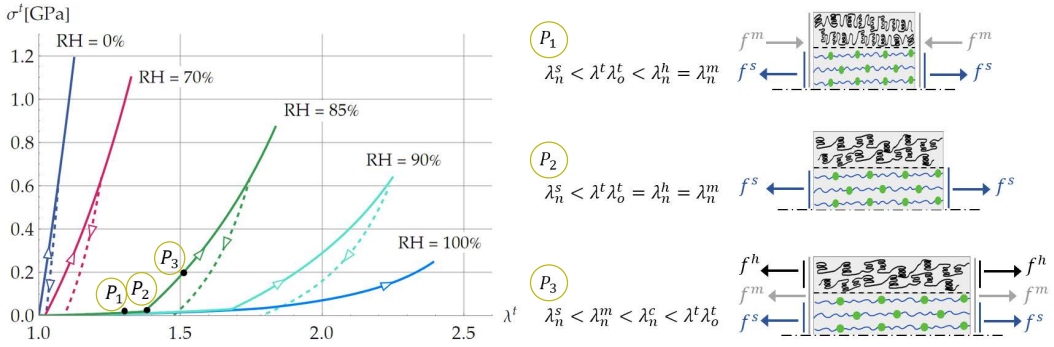

Figure 2.8: Theoretical stress-stretch curves at different humidity conditions (continuous lines correspond to loading and dashed to unloading curves). Three points of the curve RH = 85% are marked to illustrate different regimes schematically illustrated on the right: P_1 hard phase is unloaded, P_2 hard phase in its natural configuration, P_3 hard phase in traction. Here $E_o^h = 2.16$ GPa, $E^s = 18$ MPa, $\mu = 0.14$ MPa, $c_1 = 1.33$, $c_2 = 0.75$, $\alpha = 0.0099$, $m_f/m_o = 0.3$, $\bar{s} = 6.5$, $\lambda_c^s = 1.65, \mathrm{RH}_c = 80\%, \lambda_{lim} = 1.34$.

more relevant because its natural length λ_n^s undergoes a significant decrease (see Fig. 2.4). This leads to a consequent substantial decrease of the initial (zero force) length of the total fiber λ_o^t. Thus, as long as the fibrils of the hard region do not reach again their natural length, the mechanical response is given only by the amorphous phase and the matrix (see *e.g.* point P_1 in Fig. 2.8). Then, as soon as the hard region starts to be stretched (point P_2) the force starts again to be mainly sustained by the stiffer crystalline phase (see *e.g.* point P_3). In this way we give a theoretical interpretation, in full agreement with the experiments (see Fig 2.9 and Fig. 2.9), of the 'localized' material hardening observed in spider threads . The behavior is similar also for higher values of the humidity (RH = 90%), with different values of λ_t distinguishing the two regimes. At extreme humidity conditions (last curve at RH = 100%) the mechanical behavior may be given by the only amorphous phase and matrix if the ultimate stretch of the fiber is lower than the transition threshold.

Even though in this paper we are focussed on humidity effects on monotonic stress-stretch curves, in Fig. 2.8 we also show through dashed lines the system behavior when subjected to unloading. This figure let us show that, based on the microstructure interpretation, the proposed model is able to describe not only the fundamental macroscopic damage effect, but also the experimentally observed presence of residual stretches [Vehoff et al., 2007]. Interestingly, permanent deformations are not deduced as usually independently from damage, *e.g.* through the introduction of new variables, whereas both damage and residual stretches descend from the unfolding of the hard domains.

A comment about a second important aspect of the variable mechanical behavior is now in order: *temperature effects*. Indeed, when the temperature at fixed RH is increased, the silk undergoes an effect of link scissions as described for humidity [Plaza et al., 2006]. Moreover, also temperature growth is accompanied by fiber contraction [Glišović and Salditt, 2007] and again the experiments show the existence of a critical value where such effects of link scission and length variation are strongly localized. In analogy with polymer mechanics this value is indicated as glass transition temperature T_g. In particular, in [Fu et al., 2009] the authors obtained an experimental linear relation between T_g and RH. Of course such a relation would ask a theoretical description that by itself appears to be very interesting, but it is out of the aims of this paper. Instead, to show that our model can reproduce also the experimental temperature effects, we *phenomenologically* assume a Gaussian dependence of the number of links from temperature in Eqn. (2.2) (where RH is substituted by T) and then we modify correspondingly the constitutive equation Eqn. (2.18). Accordingly RH_c is substituted by T_g. The efficacy of these assumptions are well supported by the experimental comparison considered in the following section.

The final aspect of the model regards the humidity and temperature dependence of the limit stretch: to this hand we need a *fracture criterion*. Based on the considerations in [Yazawa et al., 2020] we here assume that the fracture is regulated by the hard fraction and in particular that the fracture condition is $\lambda^h = \lambda_{lim}$, where λ_{lim} is a given constitutive parameter. As we show in the following section, this criterion is successful with the exception of the fully dry condition where the breakage is known to be induced by localized damage defects [Yazawa et al., 2020].

2.3 Experimental validation

In this final section we verify the effectiveness of the proposed model in predicting the mechanical behavior of spider silks by quantitatively comparing the main experimental effects induced by humidity and temperature variations on different silks with the theoretical behavior.

Consider first the tensile response under variable RH for a highly stretchable silk (Argiope trifasciata fibers, reproduced by [Elices et al., 2005]). As shown in Fig. 2.9 this silk exhibits a remarkable dependence of the mechanical response on humidity. We may observe two different regimes in accordance with the silk experimental response: for $\mathrm{RH}<\mathrm{RH}_c$ the behavior is almost linear and this is due

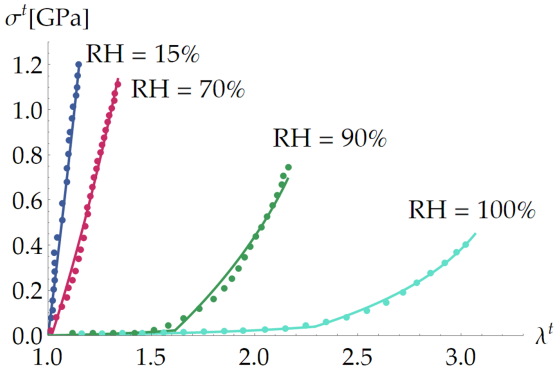

Figure 2.9: Theoretical (continuous lines) *vs* experimental (dots, reproduced from [Elices et al., 2005]) stress-stretch curves for *Argiope trifasciata* spider fibers at different RH and $T = 20\,$℃. Here $E_o^h = 2.2$ GPa, $E^s = 13.5$ MPa, $\mu = 0.14$ MPa, $c_1 = 1.3$, $c_2 = 0.87$, $\alpha = 0.0094$, $\bar{s} = 8.5$, $m_f/m_o = 0.12$, $\lambda_c^s = 1.62$, $\lambda_{lim} = 1.34$, $\mathrm{RH}_c = 85\%$.

in our model to the dominance of the hard fraction; for RH>RH_c we have two regimes. Initially the silk is highly stretchable, with high deformations at very low forces. In this regime the numerical simulations show that the hard region is shorter than its natural length, so it does not contribute to the fiber stiffness. When this length is attained, the fiber exhibits a sudden hardening. In Fig. 2.9 it is possible also to verify the efficacy of the introduced fracture criterion. Indeed, we calibrated the hard fraction limit stretch to reproduce the experimental limit at RH = 70% and then we predicted the RH = 90% and 100% cases with errors of only 0.17% and 1.49%, respectively. As anticipated the prediction is less accurate for very low humidities. Observe the possibility of predicting damage, hardening localization and fracture strain at very different values of the humidity with fixed parameters.

To further test the efficiency of the model, we show the possibility of predicting the influence of humidity and temperature on other important material parameters, such as elastic modulus, supercontraction stretch of unrestrained fibers and limit stretches (Fig. 2.10). It is important to remark, regarding the predictivity properties of the model, that in this prediction of the experiments *we fixed the material parameters* and changed only RH_c at different temperatures *using the experimental values in* [Plaza et al., 2006]. In Fig. 2.10 we also reproduce the limit stretch for the available testing temperature (55 ℃).

Furthermore, we consider the effects induced by variable temperature at fixed RH in [Plaza et al., 2006]. We evaluated T_g at given RH = 50% using the relation reported in the same paper. The results exhibited in Fig. 2.11 show again an accurate reproduction of the experiments. We remark that a comparison of the values of E at RH = 50% and different temperatures lead to different values of the elastic moduli estimated by Fig. 2.10 and Fig. 2.11, thus showing that the

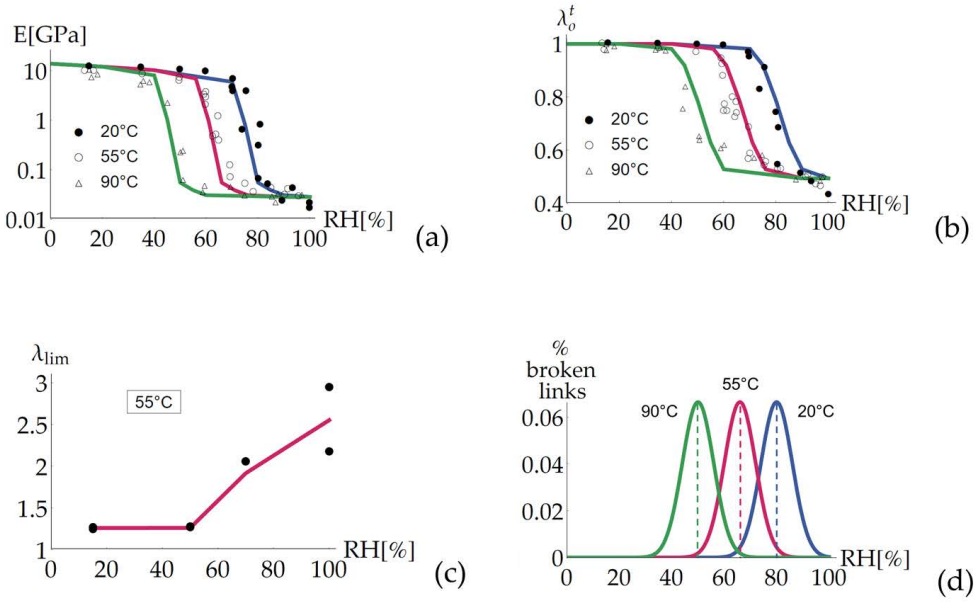

Figure 2.10: Theoretical (continuous lines) *vs* experimental (dots) curves for the initial elastic modulus (a), initial (zero force) stretch (b), limit stretch (c) and assumed Gaussian distribution of broken links (d) as a function of RH for *Argiope trifasciata* fibers [Plaza et al., 2006]. Here $E_o^h = 4.05$ GPa, $E^s = 39.6$ MPa, $\mu = 0.2$ MPa, $c_1 = 1.4$, $c_2 = 0.75$, $\alpha = 0.0065$, $\bar{s} = 0.6$, $m_f/m_o = 0.23$, $\lambda_c^s = 2.05$, $\lambda_{lim} = 1.26$, and $\text{RH}_c = 80, 66, 50\%$ for $T = 20, 55, 90\,^{\circ}\text{C}$, respectively.

data refers to different silks so that different parameters have been used in the two figures.

In addition to the *Argiope trifasciata* spider fibers (Fig. 2.9), we consider tensile tests performed on a *Nephila clavata* spider fiber under various RHs $(0\%, 75\%, 97\%)$ reproduced from [Yazawa et al., 2020]. In Fig. 2.12 we report the comparison between the experimental results and the theoretical model. Despite this silk shows a remarkably different response to the humidity variations, the proposed model is once again significantly successful in quantitatively predicting the observed experimental behavior. In Fig. 2.12 we also test the above described rupture hypothesis for the *Nephila clavata* fibers by using the value corresponding to the experimental break at RH = 75% to predict the breaking strain at RH = 97% with an error of 0.22%. This remarkable small error confirms the plausibility of the proposed rupture criterion. Observe anyway that this hypothesis in this silk cannot be applied to the fully dry case where the breakage is typically induced by localized defects [Yazawa et al., 2020].

As a last comparison with experimental results, in Fig. 2.13 we show the possibility of predicting with remarkable accuracy the complex mechanical response of a dragline silk in dry condition (reproduced from [Gosline et al., 1999]). No-

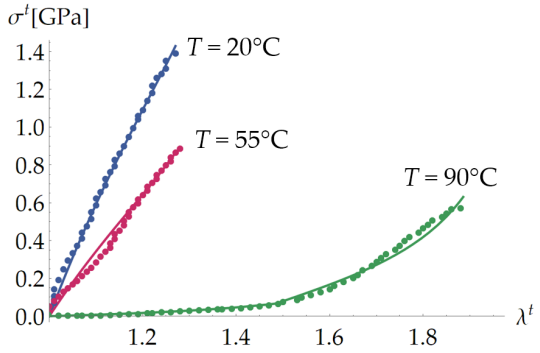

Figure 2.11: Theoretical (continuous lines) *vs* experimental (dots) stress-stretch curves for *Argiope trifasciata* spider fibers at different temperature at fixed RH = 50% (reproduced from [Plaza et al., 2006]). Here $T_g = 84\,°C$, $E_o^h = 3.83$ GPa, $E^s = 32.7$ MPa, $\mu = 2$ MPa, $c_1 = 1.36$, $c_2 = 1.25$, $\alpha = 0.00995$, $\bar{s} = 4.5$, $m_f/m_o = 0.345$, $\lambda_c^s = 1.49$, $\lambda_{lim} = 1.27$.

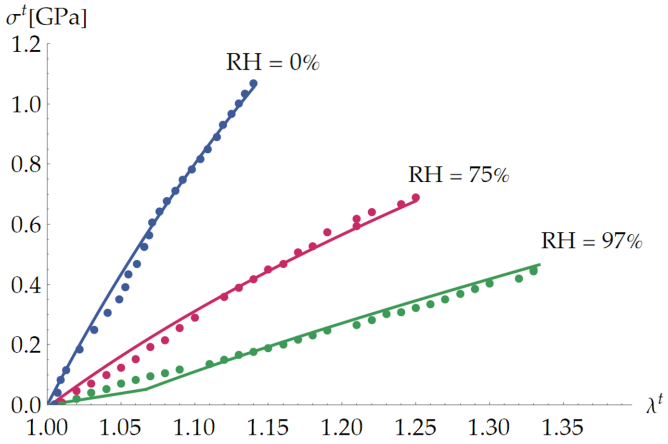

Figure 2.12: Theoretical (continuous lines) *vs* experimental (dots) stress-stretch curves for *Nephila clavata* fibers at different humidity and $T = 25\,°C$ (reproduced from the experiments at strain rate of $3.3\times10^{-3}\ s^{-1}$ in [Yazawa et al., 2020]). Here $E_o^h = 5.6$ GPa, $E^s = 1.47$ GPa, $\mu = 1.1$ MPa, $c_1 = 1.35$, $c_2 = 1.46$, $\alpha = 0.009$, $m_f/m_o = 0.878$, $\bar{s} = 3$, $\lambda_c^s = 4.6, \lambda_{lim} = 1.25$, $\mathrm{RH}_c = 80\%$.

tice that here, the thread is produced by a third species of spider, the *Araneus diadematus*.

2.4 Discussion

A comment is in order. As anticipated previously, in the recent paper [Cohen et al., 2021] the authors proposed a different model well reproducing with 8 parameters the experiments in Fig.2.10a),b). Our model, being based on a more detailed description of the microstructure, considering the two different fractions and stretch induced unfolding effects, ends up with 10 parameters. On the other

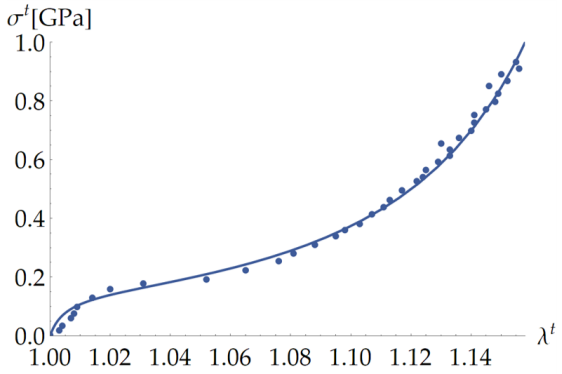

Figure 2.13: Theoretical (continuous lines) *vs* experimental (dots) stress-stretch curves for *Araneus diadematus* spider fibers (reproduced from [Gosline et al., 1999]). Here $E_o^h = 6.75$ GPa, $E^s = 13.6$ MPa, $\mu = 1$ MPa, $c_1 = 1.005$, $c_2 = 1.979$, $\alpha = 0$, $m_f/m_o = 0.13$, $\bar{s} = 4.5$, $\lambda_c^s = 1.25$, $\mathrm{RH}_c = 84\%$.

hand, the proposed model is able to reproduce the whole stress-strain curves as function of both temperature and humidity, with unloading, increasing damage, fracture and residual stretches. It's important to observe that the most diffuse models efficiently describing damage and residual stretches in soft polymeric materials (without any temperature and humidity effects) adopt 7-8 parameters (see [De Tommasi et al., 2019] and references therein).

2.5 Conclusions

The described ability of the proposed model of predicting the experimental behavior of different mechanical properties make us confident that it well reproduces the humidity and temperature effects at the molecular scales. This is supported even more by predicting the behavior of different silks and environmental conditions. We strongly believe that this is a consequence of our microstructure deduction of the material response function. The physical meaning of all the adopted (microscopic) parameters opens up the possibility of applying the proposed model not only to other protein materials with similar structures [Puglisi et al., 2017], but also in the design of bioinspired materials employing chosen specific proteins [Greco et al., 2021; Liu et al., 2019].

Chapter 3

"Water to the ropes": a predictive model for the supercontraction stress of spider silks

This chapter is based on the paper published in Extreme Mechanics Letters (volume 61, 102010, 2023, 10.1016/j.eml.2023.102010) [Fazio et al., 2023] and deals with the forces developed by the spider silk when restrained and exposed to wet environments. The possibility of using such system as a humidity driven actuator is also discussed.

Abstract

When humidified at different moisture conditions, restrained spider silk fibers can exhibit a very high supercontraction phenomenon. The hydration water molecules induce a Hydrogen-bonds disruption process that, due to entropic effects, decreases the natural -zero force- end-to-end chains length. By considering a bundle of macromolecules, we describe supercontraction as a possible actuation system and determine the maximum actuation force depending on the silk properties at the molecular scale and on the constraining system representing other silk threads or the actuated device. The comparison with experimental results of *Argiope trifasciata* silk fibers show the effectiveness of the proposed model in quantitatively predicting the experimental actuation properties. The considered historical case study of obelisk rescue in Saint Peter's Square (Rome) through ropes hydration is discussed evidencing the optimal performances of this natural material adopted as moisture powered actuator: we obtain a work density of 2.19 kJ/m^3 making spider silk the most performant hydration driven active material.

Moreover we obtain a power density of the order of 730 W/kg about three times the most performant carbon nanotube actuators making such material very competitive as compared with all types of actuator. The analytic description of the macroscopic actuation parameters from microscale properties shows the possibility of adopting our approach also in the field of bioinspired artificial silks design, possibly considering also important non-linear effects in the actuated system.

3.1 Introduction

Spider silks have been increasingly the focus in very wide research and technological fields due to their extreme mechanical properties [Gosline et al., 1986] such as extraordinary strength and toughness, self-healing, and environmental adaptability. Such properties are often unattained by artificial materials (see [Pérez-Rigueiro et al., 2021] and references therein). This outstanding material response, resulting by a complex hierarchical structure organization, attracted the attention in the important field of biomimetics [Zhao et al., 2014; Greco et al., 2021; Arndt et al., 2022]. The deduction of the fiber material response, starting from silk structure at the molecular scale, represents a demanding theoretical problem, not completely clear, especially regarding the adaptability to different environmental (humidity and temperature) and loading conditions.

Here, by extending the recent results in [Fazio et al., 2022], we focus on the important role of hydration on the actuation material properties of spider silks. Indeed, a striking effect observed in spider silks is the so called *supercontraction effect*, addressed, to the knowledge of the author, for the first time in 1977 [Work, 1977], that occurs when a spider silk thread is exposed to humidity. Depending on the silk composition, the experiments show the existence of a Relative Humidity (RH) threshold beyond which the fiber contracts up to a half of its initial (dry) length. The experimentally observed contraction depends on several factors, including spider species [Boutry and Blackledge, 2010], type of silk (among the up seven different ones that some spiders can produce [Vollrath, 1992; Gosline et al., 1994]), environmental conditions [Plaza et al., 2006] and hydration rate [Agnarsson et al., 2009a].

Interestingly, thinking both to the humidity effects on spider webs and to the possibility of adopting such behavior as a natural actuation device, constrained humidified silks generates a stress that can be measured trough a load cell [Guinea et al., 2003; Ene et al., 2011]. In the case of fully locked end-points (fixed end-to-end length) the higher magnitude of the stress is attained, typically of the order of tens of MPa as experimentally measured by [Agnarsson et al., 2009a; Guinea

et al., 2003; Blackledge et al., 2009]. Clarifying and predicting the mechanisms that originate this phenomenon, both in natural and artificial silks, is considered of great interest also in the perspective of adopting this behavior at the base of mechanical actuation [Dong et al., 2021] or humidity sensing devices [Zhang et al., 2022]. This aspect is the focus of the following analysis.

Two types of experiments are typically performed to describe the supercontraction effect. In the first type, the dry silk sample is free to shorten and the length variation under increasing humidity is measured. In the second case the fiber is fixed at its dry length and no contractions is allowed and the force applied to fix the length at increasing humidity is measured. The maximum attained value of the stress is then measured, *supercontraction stress* [Bell et al., 2002; Savage et al., 2004] and it represents the previously recalled actuation force. Intermediate boundary conditions (see [Florio and Puglisi, 2019] for a theoretical discussion of the influence of different boundary conditions on the material response of constrained systems) represent real phenomena with the silk constrained by deformable devices such as other silks fibers in the web or external attaching systems.

At the molecular scale, the supercontraction effect is due to hydration-induced hydrogen bonds breaking. More in detail, at the molecular scale spider silk is composed by an amorphous matrix of oligopeptide chains and pseudo-crystalline regions, made up principally of polyalanine β-sheets [Elices et al., 2011; Sponner et al., 2007] with dimensions between 1 and 10 nm [Keten and Buehler, 2010] oriented in the direction of the fiber [Jenkins et al., 2013]. The chains are highly hydrogen bonded, with a medium-low density of H-bonds in the amorphous part whereas the nanocrystallites are characterized by a high density of H-bonds [Yarger et al., 2018]. Supercontraction is the result of the entropic recoiling of the macromolecules when the hydration water molecules break the H-bonds naturally present in the virgin silk, that previously fixed the macromolecules in a natural elongated conformation induced during the spinning of the fiber [Du et al., 2006; Elices et al., 2005; Elices et al., 2011].

In the following, we propose an approach to determine the supercontraction stress arising when a dry silk fiber is humidified with monotonically increasing RH and the length is fixed. Specifically, we reframe the multiscale approach for the humidity and thermomechanical response of spider silks effects recently proposed in [Fazio et al., 2022], where the silk thread is modeled as a composite material made by hard crystalline fraction and a soft amorphous fraction embedded in an elastic tridimensional matrix reproducing the network effect [Flory and Erman, 1982]. Regarding the hydration phenomenon, it is important to remark that,

due to their different conformations and chemical composition, hard crystalline domains are hydrophobic whereas the soft amorphous fractions are hydrophilic. This results in two different responses to the hydration for the two different fractions. Indeed, the water molecules cannot penetrate the hydrophobic crystalline domains and the only humidity effect in this fraction is an increased misalignment of the crystallites with respect to the fiber axis.

On the other hand, the hydration water molecules decrease the percentage of crosslinks in the softer region, inducing a modification of the natural (zero-force) length of the chains composing this region. We then assume that during hydration the natural lengths of the hard region and the matrix do not vary, so that the supercontraction stress is a direct effect of the variation of the H-bonds percentage and resulting variation of the natural length of the silk molecules belonging to the soft region (as detailed in [Fazio et al., 2022]). As a consequence, in the specific experiment of fixed total length we may deduce that the mechanical response is due to the only configurational changes of the soft fraction (see the scheme in Fig. 3.1(a)). Instead, as anticipated above, when restrained to its natural (zero force) length, if humidified the fiber cannot contract and supercontraction forces develop (see Fig. 3.1(b)).

In this work we provide a model to predict the supercontraction stress when the environmental humidity is monotonically increased. Further, we enhance the previous model in [Fazio et al., 2022] of the amorphous chain by reducing the number of fit parameters based on classical results of reaction kinetic theory. The effectiveness of the modifications of the model is demonstrated by quantitatively predicting the experimental supercontraction stress for restrained spider Argiope trifasciata silk fibers.

We then show the efficiency of the actuation properties of spider silks fibers under external variable humidity conditions by considering a historical example when in 1586 Pope Sisto V asked for the erection of the 350 tons and 25 meters high obelisk in Saint Peter's Square (Rome, Italy). In that case the abrupt fall of the obelisk during erection was avoided by their activation shortening through hydration of the ropes adopted for the erection. To explicitly describe through a virtual example the incredible actuation properties of this material we design a virtual system based on the actuation of spider silks to attain the same results of the ropes.

We then study the fundamental effect of the elasticity of the actuated device, typically neglected in previous studies, and show how actuation properties such as the actuation work and power depend not only on the spider silks, but also on the stiffness of the interacting actuated device. We then consider the optimization

problem for a linear actuated systems to determine the values of the relative (actuation and actuated) stiffness leading to the maximum actuation work. As a result, we are able to explicitly determine the actuation properties of silks and we show that a work density of 2.19 kJ/m^3 can be attained so that spider silk, to the knowledge of the authors, can be considered the most performant hydration driven active material. Moreover we determine a power density of 730 W/kg that is about three times the most performant carbon nanotube actuators. We then deduce that also in a comparison with general actuation systems spider silks result as very competitive actuators. Eventually, as a simple prototypical amplified actuation device, we also show the possibility of strongly increasing the actuation by considering multistable responses of the actuated system.

We believe that our study can open up the understanding and the application of spider silks and artificial silks as actuation devices. Based on our analytic results we are also confident that the proposed model will be important in the design of very efficient actuation materials and devices.

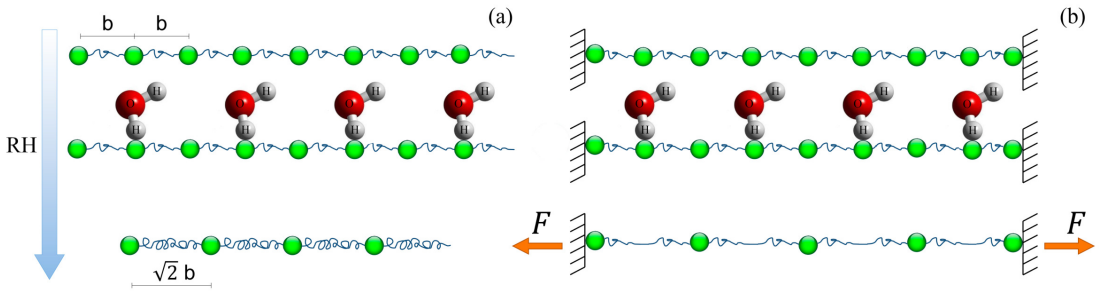

Figure 3.1: Cartoon of the supercontraction effect. Water molecules reduces the number of H bonds and, due to entropic effects, the chain is subjected to a reduction of its natural end-to-end length. In (a) we show the case of humidification for an unconstrained chain and the resulting contraction: *e.g.* two Kuhn segments of total length $2b$ coalesce in a segment of length $\sqrt{2}b$. In (b) we show the case of (perfectly) constrained molecule and the resulting force applied to the constraint of particular interest in this paper where actuation properties are considered.

3.2 Model

Here for the spider thread we adopt the theoretical model recently proposed in [Fazio et al., 2022] where the authors describe the spider thread as composed by a multiphase material. Thus the stress is additively decomposed as the sum of the the soft fraction contribution σ_{soft}, depending on RH, the hard fraction contribution σ_{hard}, active only for lengths larger than the natural one (see [Fazio et al., 2022] for details), and the elastic matrix contribution, representing an

external network, here modeled as Neo-Hookean with modulus μ. The total constitutive response of the silk is then

$$\sigma(\lambda) = \sigma_{soft}(\lambda) + \sigma_{hard}(\lambda) + \mu\left(\lambda_{\rm eq} - \frac{1}{\lambda_{\rm eq}^2}\right). \tag{3.1}$$

Notice that in this paper, the hard region chains are slack (they do not sustain any compressive load), so that here we may neglect their contribution. The extension to the contemporary application of external humidity and mechanical loading can be obtained by simply considering the contribution $\sigma_{\rm hard} \neq 0$.

On the other hand here we consider the case of fixed length and determine the resulting stress

$$\sigma_{act} = \sigma(1) \equiv \sigma_{soft}(1).$$

Let us then consider the only soft fraction. The silk thread (hygro-thermo-mechanical) behavior results as an average response of the (parallel) silks molecules spider silk [Plaza et al., 2006]. Specifically, we suppose that the soft fraction is composed of identical molecules composed by a number n of Kuhn segments each of length b. The expectation value of the unloaded end-to-end length of an ideal chain can then be evaluated by using a classical result of the Statistical Mechanics [Rubinstein and Colby, 2003] as

$$L_n =< r^2 >^{1/2} = n^{1/2}\, b, \tag{3.2}$$

whereas its contour length is $L_c = n\, b$.

We indicate by m_b and m the numbers of broken and unbroken H-bonds, respectively, with

$$m_b + m = m_o, \tag{3.3}$$

where m_o is the initial (dry) number of bonds. We then describe the disruption process induced by the hydration water molecules n_w [Du et al., 2006] based on the classical Michaelis-Menten kinetics adopted to describe the enzymatic reaction as regulated by the concentration of a substrate. The considered reaction (see Fig. 3.2) is

$$m + n_w \underset{k^-}{\overset{k^+}{\rightleftharpoons}} m_b.$$

so that the rate of bonds breaking is

$$\frac{\mathrm{d}\, m_b}{\mathrm{d}t} = k^+ n_w(\mathrm{RH})\, m - k^- m_b. \tag{3.4}$$

When the equilibrium of the transition is attained, we get $\frac{m_b}{n_w(\mathrm{RH})\, m} = \frac{1}{k_d}$ with

$k_d = \frac{k^-}{k^+}$. By substituting in Eqn. (3.3) we obtain the Michaelis-Menten equation [Johnson and Goody, 2011]

$$\frac{m(\mathrm{RH})}{m_o} = \frac{1}{1 + \dfrac{n_w(\mathrm{RH})}{k_d}}. \tag{3.5}$$

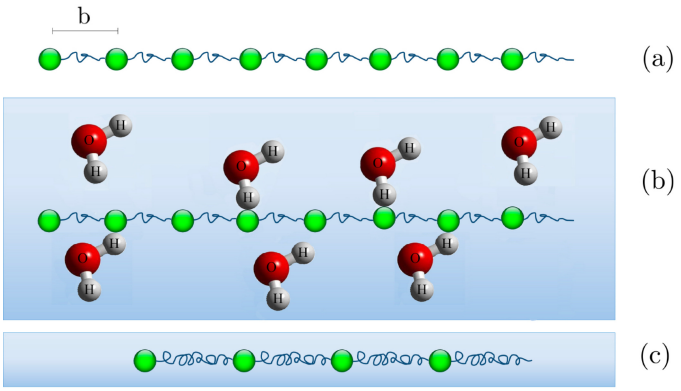

Figure 3.2: Scheme of the hydration reaction. An example of a chain composed by $m_0 = 8$ (blue) Kuhn segments of length b in the initial dry condition is schematized in (a). In the example of wet environment shown in (b) we suppose to have $n_w = 7$ water molecules and $m_b = 4$ water molecules that bound and break links, so that only $m = 4$ links are kept in the wet condition (c).

To take into account the experimental observation of the existence of a super-contraction threshold RH_c, such that as the relative humidity grows the spider silk fiber sharply contracts with increasing humidity as the threshold is approached from above RH [Fu et al., 2009], we consider the simple law

$$n_w(\mathrm{RH}) = \begin{cases} 0 & \text{if } \mathrm{RH} < \mathrm{RH}_c, \\ k_p(\mathrm{RH} - \mathrm{RH}_c) & \text{if } \mathrm{RH} \geq \mathrm{RH}_c. \end{cases} \tag{3.6}$$

Thus we get the final relation

$$\frac{m(\mathrm{RH})}{m_o} = \begin{cases} 1 & \text{if } \mathrm{RH} < \mathrm{RH}_c, \\ \frac{1}{1+k(\mathrm{RH}-\mathrm{RH}_c)} & \text{if } \mathrm{RH} \geq \mathrm{RH}_c, \end{cases} \tag{3.7}$$

with

$$k = \frac{k_p\, k^+}{k^-} \tag{3.8}$$

representing the main parameter regulating the hydration induced debonding effect of H-bonds. The resulting trend is schematized in Fig. 3.3.

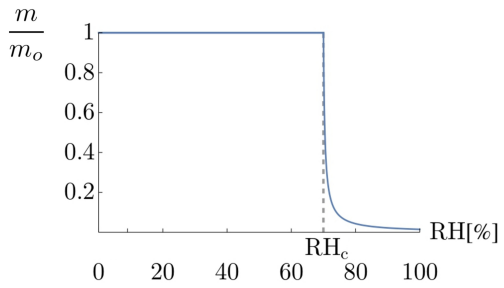

Figure 3.3: Influence of the humidity on the relative (as compared with the initial, dry condition) number of H-bonds. Assumed parameters $\mathrm{RH}_c = 70\%, k = 2.22$.

Following [Fazio et al., 2022], to obtain the variation of the natural length of the 'representative' chain with the humidity, we assign the initial (mean) number of chain monomers n_o when the silk is spun (virgin) and we identify the number of links m with the number of domains in which the chain is divided (see the scheme in Fig. 3.2). In the generic humidity state, the (mean) number of monomers in each domain is therefore $n(\mathrm{RH}) = n_o/m(\mathrm{RH})$. Based on Eqn. (3.2) we determine the natural length of the whole chain:

$$L_n(\mathrm{RH}) = m\sqrt{n_o/m(\mathrm{RH})}\; b = \sqrt{m(\mathrm{RH})\; n_o}\; b. \tag{3.9}$$

Following [De Tommasi et al., 2013], we adopt a Worm Like Chain (WLC) type energy density per unit chain contour length L_c, $\varphi_e = \varphi_e(L, L_c) = \kappa \frac{L^2}{L_c - L}$ where $\kappa = \frac{k_B T}{4 l_p}$, T is the temperature, k_B the Boltzmann constant, and l_p the persistence length. This energy respects the limit extensibility condition, $\lim_{L \to L_c} \varphi_e(L, L_c) = +\infty$, and allows for explicit calculations. Moreover, as described above, we consider that the end-to-end distance L can be decomposed into a variable (zero-force) natural length measured by (3.2) and the remaining elastic component $L_e = L - L_n$, as firstly proposed in [Trentadue et al., 2021]. Thus we assume an energy and a force-elongation law for a single chain

$$\begin{aligned} \varphi_e &= \kappa \frac{L_e^2}{L_c - L} = \kappa \frac{(L - L_n)^2}{L_c - L}, \\ f &= \frac{\partial \varphi_e}{\partial L} = \kappa \left[\left(\frac{L_c - L_n}{L_c - L} \right)^2 - 1 \right], \end{aligned} \tag{3.10}$$

with the force decreasing to zero as the length attains its natural length ($L = L_n$ or $L_e = 0$).

In order to deduce the macroscopic behavior of the thread, we consider the classical *affinity hypothesis* [Rubinstein and Colby, 2003] that identifies the macroscopic stretches with the macromolecular ones. We can then introduce the fol-

lowing stretch measures

$$\begin{aligned} \lambda = \tfrac{L}{L_o} \quad &\text{total stretch,} & \lambda_e = \tfrac{L_e}{L_o} \quad &\text{elastic stretch,} \\ \lambda_n = \tfrac{L_n}{L_o} \quad &\text{permanent stretch,} & \lambda_c = \tfrac{L_c}{L_o} \quad &\text{contour stretch,} \end{aligned} \tag{3.11}$$

with $L_o = b\sqrt{n_o}$ denoting the initial natural length.

The natural stretch for the soft region can be deduced using Eqns. (3.9) and (3.11):

$$\lambda_n(\mathrm{RH}) = \frac{L_n(\mathrm{RH})}{L_o} = \frac{\sqrt{n_o\, m(\mathrm{RH})}\, b}{\sqrt{n_o\, m_o}\, b} = \sqrt{\frac{m(\mathrm{RH})}{m_o}} \tag{3.12}$$

with $m(\mathrm{RH})/m_o$ given by Eqn. (3.7) so that the variation of the natural length with humidity explicitly depends from physically based parameters. On the other hand, the corresponding expression for the contour length is $L_c = m\frac{n_o}{m}\, b = n_o b$, so that the contour stretch of the amorphous part is

$$\lambda_c = \frac{L_c}{L_o} = \frac{n_o\, b}{\sqrt{n_o\, m_o}\, b} = \sqrt{\frac{n_o}{m_o}}. \tag{3.13}$$

Under an additive assumption, the engineering stress (Piola Kirchhoff, force divided by initial cross-sectional area) is determined using Eqns. $(3.10)_2$ and (3.11)

$$\sigma_{soft} = E\left[\left(\frac{\lambda_c - \hat{\lambda}_n(\mathrm{RH})}{\lambda_c - \lambda}\right)^2 - 1\right], \tag{3.14}$$

where the permanent stretch depends by RH and it is given by Eqn. (3.12) and Eqn. (3.7) whereas the contour stretch is constant (see Eqn. (3.13)). Here $E = N_a\kappa$ is the elastic modulus of the soft fraction with N_a the number of chains per unitary reference area. Indeed the total energy considering all the chains of the network can be determined as $\Phi = N_v\varphi_e L_c = N_a N_l \varphi_e L_c = N_a\varphi_e$ where for the sake of simplicity we consider $N_l L_c = 1$, with N_v, N_a, N_l the number of chains per unit volume, area, and length respectively.

Observe that the supercontraction stress $\sigma_{act} = \sigma_{soft}(1)$ depends on three only material parameters all having a clear microstructure interpretation. In Fig. 3.4 we represent the possibility of designing the material response by changing the material parameters, namely the module of the soft fraction E, the contour stretch of the chains of the soft fraction λ_c and the constant k regulating on the H-Bonds disruption kinetics. The influence of possible deformable constraints are considered in the following. In the simulations reported in Fig. 3.4 typical values of spider silks have been used for material parameters, as it will be evident from

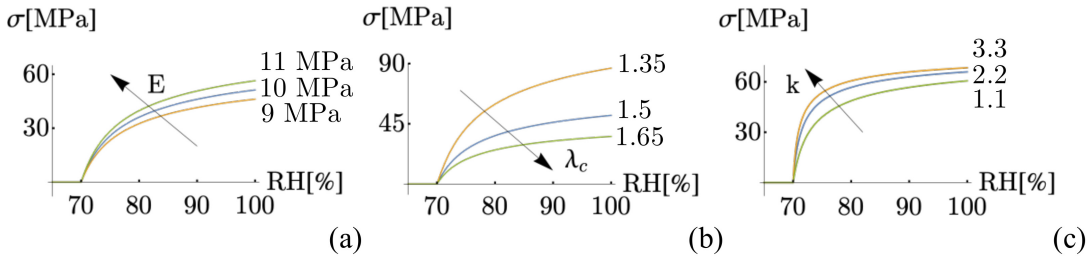

Figure 3.4: Influence of the material parameters on the macroscopic stress-stretch plot. (a) Elastic modulus soft region $E = 9, 10, 11$ MPa, $\lambda_c = 1.5, k = 0.45$, (b) Contour stretch soft region $\lambda_c = 1.35, 1.5, 1.65, E = 10$ MPa, $k = 0.45$, (c) constant depending on the H-bonds disruption kinetics $k = 1.1, 2.2, 3.3, \lambda_c = 1.5, E = 10$ MPa. $\mathrm{RH}_c = 70\%$.

the experimental validation in the following section. For the elastic modulus and the contour stretch, variations of $\pm 10\%$ were considered, while for the constant k the arbitrary variation of $\pm 50\%$ was considered to obtain an appreciable variation in the stress-stretch plot.

3.3 Experimental validation

To test the effectiveness of the proposed model in quantitatively describing the experimental behavior, we consider in Fig. 3.5 a test on spider dragline silk (*Argiope trifasciata*) reproduced from [Guinea et al., 2003], in an experiment when the fiber is clamped at maximum length with zero force and exposed to monotonically increasing humidity content while measuring the force. In the figure the value of the engineering stress is reported.

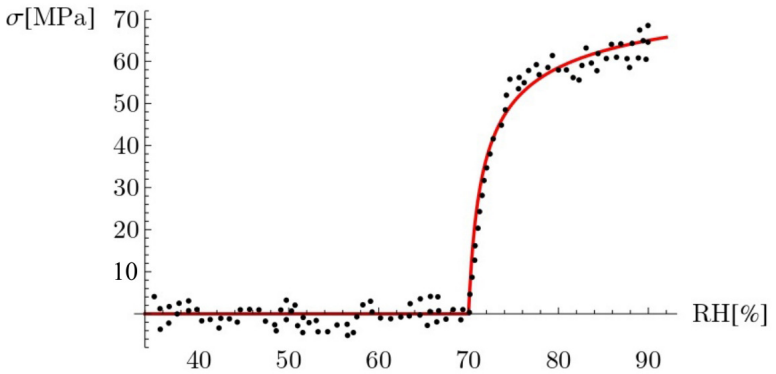

Figure 3.5: Stress-humidity curve for an *Argiope trifasciata* spider: dots are the experiments reproduced from [Guinea et al., 2003], whereas the continuous line represents the theoretical prediction with parameters $E = 10.3$ MPa, $\lambda_c = 1.5, k = 2.22, \mathrm{RH}_c = 70\%$.

As expected for restrained fibers tested in controlled environment with monotonically increasing RH a sudden development of supercontraction stress occurs only once the RH overcomes the threshold of $\mathrm{RH}_c \cong 70\%$ with mainly no stress variation below this critical threshold. A steep increasing of stress is observed in the narrow range between RH$\sim$ 70% and RH$\sim$ 75%. For higher RHs the stress keeps growing, but with a much lower slope until RH=100% when the maximum actuation stress is attained.

Notice that the optimal fit parameters are fully compatible with the ones employed in [Fazio et al., 2022] to reproduce the stress-stretch behavior of fibers of the same spider species (*Argiope trifasciata*) where they were deduced based on their micromechanical interpretation. Indeed here we assume $E = 10.3$ MPa, $\lambda_c = 1.5$ for the elastic modulus and the contour stretch, respectively, in place of $E = 13.5$ MPa, $\lambda_c = 1.62$ employed in [Fazio et al., 2022] for the soft region. Likewise, the assumed supercontraction threshold $\mathrm{RH}_c = 70\%$ match the one experimentally found during the experimental test of the fiber [Guinea et al., 2003]. These differences are then justified by the observation that the experiment refers to a different tested silk of the same spider species, given that, even within the same specie, a variability in the silk behavior can be observed [Madsen et al., 1999]. The only "arbitrary fitting" parameter is therefore the constant k regulating the decreasing of the number of H-bonds when the humidity overcomes the critical threshold. To the knowledge of the authors, no experimental values are available so that we assume $k \cong 2$ as a best fit parameter. The model turns out to be accurate in quantitatively reproducing the experimental behavior, based on the obtained final relation (3.14).

3.4 Elastic interaction

We previously introduced two special cases of boundary conditions, namely the case of a restrained thread with fixed end-to-end distance studied in the previous section, and the case of unrestrained supercontraction discussed in detail in [Fazio et al., 2022]. In this section we explore the actuation properties of the spider silk thread during supercontraction in the more realistic case when elastic interactions at the boundary are considered, as schematized in Fig. 3.6. This scheme mimics for example the case of spider webs where the different threads, *i.e.* the radii and the capture spiral, respond (and thus contract) in different ways to changes in humidity environment. In particular, the capture spiral is composed of elastic flagelliform silk that is coated by hygroscopic glue droplets. The glue droplets contain water that maintains flagelliform silk in a continuously supercontracted

state [Guinea et al., 2010] so that no further contraction occurs in wet external environment. On the other hand, the Major Ampullate (MA) silk forming the radii contracts when the threads are immersed in wet environment. In the simplified scheme for studying the elastic interaction among the threads, the MA silk is represented by the radius silk thread that when contracts may elongate the spring (capture spiral thread). In the following, for simplicity, we consider the capture spiral thread elasticity by constraining the radial thread with a linear spring.

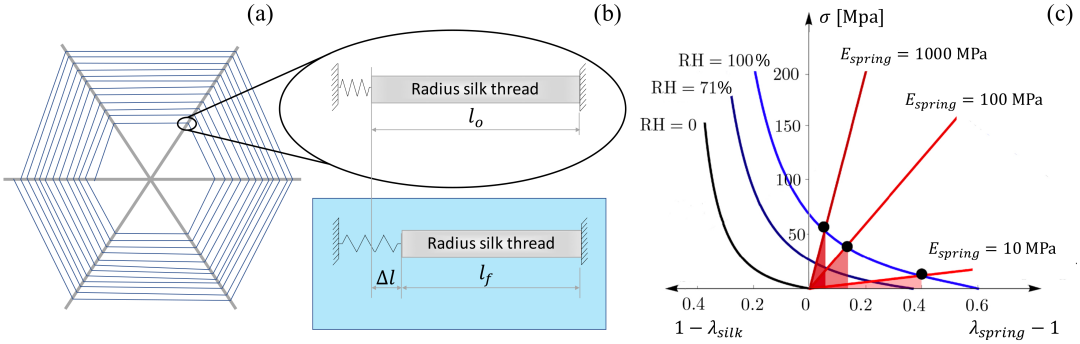

Figure 3.6: (a) scheme of a spider web and the radius silk thread-spring interaction. The spider web is composed by radii (grey) and capture spiral (blue). Due to different composition, the radii may undergo contraction in humid conditions, whereas the capture spiral threads do not contract. (b) simple scheme of the elastic interaction between the radius silk contracting and elongating spiral silks modeled as elastic springs. (c) Stress-strain behaviors of the silks at different humidity and different stiffnesses of the springs represented in the same diagram using the compatibility condition (3.16). Different equilibrium configuration corresponding to humidity saturation and different actuated device stiffness are represented by black dots. Assumed parameters: $E = 10$ MPa, $\lambda_c = 1.5$, $k = 2.22, \mathrm{RH}_c = 70\%, \mathrm{RH} = 100\%, \mu = 1$ MPa, $E_{spring} = 30, 300, 1000$ MPa

Since we are interested to work and power density, hereon we can assume unit reference area and length for both the actuated device and spider thread. We have then to respect the simple equilibrium condition (suppose for simplicity of notation that the two devices have unitary area)

$$\sigma_{\text{spring}} = \sigma_{\text{silk}}. \tag{3.15}$$

On the other hand the compatibility condition of the strain, imposing that the shortening of the silk thread from the dry condition, measured by $1-\lambda_{\text{silk}}$, equals the elongation of the actuated device $\lambda_{\text{spring}} - 1$, gives

$$\lambda_{\text{spring}} - 1 = 1 - \lambda_{\text{silk}} \tag{3.16}$$

with

$$\sigma_{\text{spring}} = E_{\text{spring}}(1 - \lambda_{\text{silk}}). \tag{3.17}$$

Then the equilibrium stretch λ_{eq} is uniquely defined using Eqn. (3.15) by solving

$$E\left[\left(\frac{\lambda_c - \hat{\lambda}_n(\text{RH})}{\lambda_c - \lambda_{\text{eq}}}\right)^2 - 1\right] + \mu\left(\lambda_{\text{eq}} - \frac{1}{\lambda_{\text{eq}}^2}\right) = E_{\text{spring}}(1 - \lambda_{\text{eq}}). \tag{3.18}$$

By increasing humidity, as supercontraction occurs, the equilibrium stretch λ_{eq} grows as shown in Fig. 3.6 (c). In the figure, using Eq. (3.16), we represent in the same diagram the stress of the actuated device versus its elongation $(\lambda_{spring} - 1)$ and the stress of the silk thread as a function of the silk elongation $(1 - \lambda_{silk})$, so that the equilibrium configurations correspond to the intersections of the two stress-strain curves. The corresponding actuation work is represented by shaded areas. The dependence of the stretch and non-dimensional stress $\bar{\sigma} := \frac{\sigma_{\text{act}}}{\sigma_{\text{act,max}}}$ on the stiffness ratio $\xi = \frac{E_{\text{spring}}}{E}$ is represented in Fig. 3.7. Notice that for high values of ξ the stretch reaches the saturation value corresponding to the restrained supercontraction ($\lambda = 1$) and the silk approaches the restrained condition actuation stress ($\sigma_{\text{act}} = \sigma_{\text{act,max}}$). On the other hand, the supercontraction stress is zero for low values of ξ (corresponding to unrestrained silk thread), where the contraction stretch attains its minimum value $\lambda_{\text{eq}} = \lambda_{\text{sc}}$ known as supercontraction stretch.

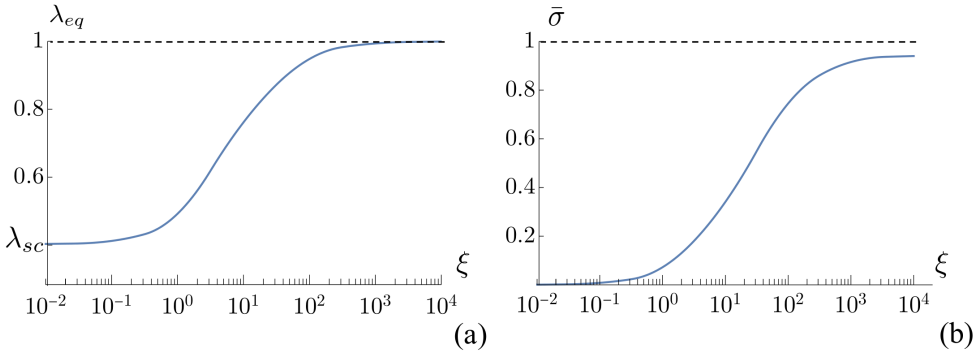

Figure 3.7: Actuation dependence on spring modulus. Assumed parameters: $E = 10$ MPa, $\lambda_c = 1.5$, $k = 2.22, \text{RH}_c = 70\%, \text{RH} = 100\%, \mu = 1$ MPa

3.5 Water to the ropes!

In 1586 Pope Sisto V, wanting to embellish Saint Peter's Square (Rome, Italy), ordered that the large obelisk -that is still admired there- be erected. The work,

which was entrusted to the architect Domenico Fontana, presented serious difficulties. The obelisk weighed 350 tons and was 25 meters high, so that Fontana had to do calculations and engage scaffolding, winches, pulleys, hemp ropes, 800 men, and 140 horses. The obelisk was almost in place when people saw the ropes overheat dangerously, with the risk that they caught fire. The monolith would have fallen to the ground. Then in the silence there was a scream: "Daghe l'aiga ae corde!" (expression of the Ligurian language meaning "Water to the ropes"). The advice was immediately followed with excellent results. To thwart the danger had been Captain Benedetto Bresca, Ligurian sailor of Sanremo, who knew well that the hemp ropes are heated by the clutch of the winches and also shorten when they are wet. Thus the water-induced contraction of hemp ropes was employed to erect the obelisk thus using it as a humidity driven actuator.

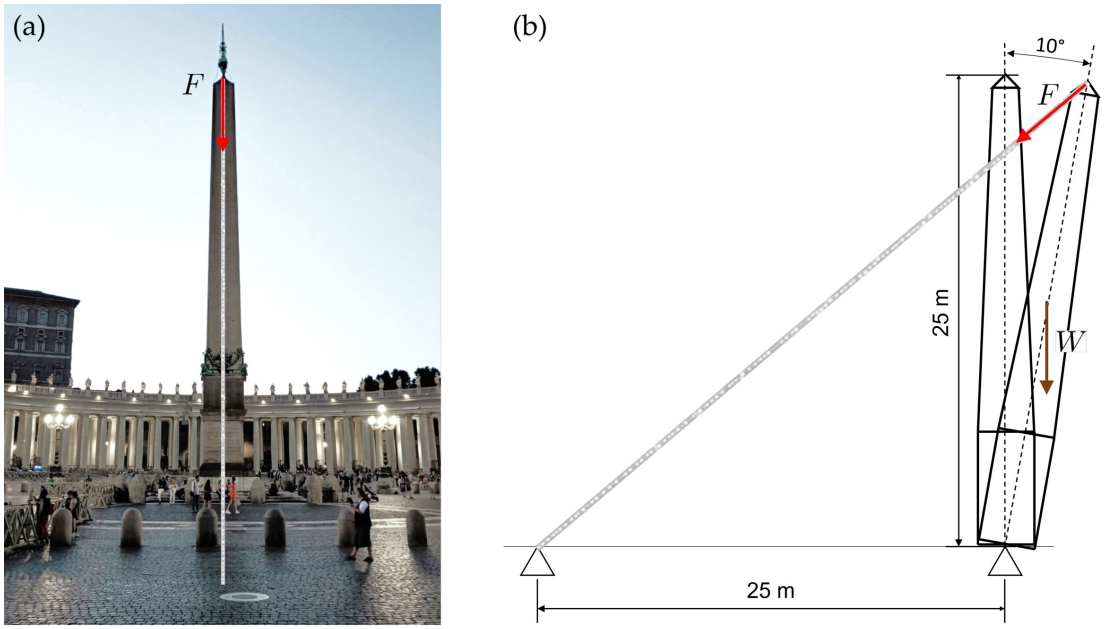

Figure 3.8: Scheme of the equilibrium recovery for the obelisk, considering a weight of 3.5 MN applied in the center of mass in a configuration of 10° from the equilibrium vertical configuration, obtained based on supercontraction spider silk actuation. Considering the parameters obtained for the *Argiope trifasciata* silk fiber (Fig. 3.5), we calculated the diameter of the spider silk thread needed that is only 10 cm.

Here, to exemplify through a 'real' example the strong actuation properties of spider silks when subjected to humidity, we consider the possibility of erecting the obelisk by employing the supercontraction properties of the spider silk. More explicitly, we determine the diameter of the silk thread needed to rotate the obelisk in the vertical direction, starting from a misalignment angle of 10° and assuming that the distance of the constraint is equal to the height (25 m) of the obelisk (see Fig. 3.8). For simplicity we consider the center of gravity at the half

of the height where the 350 tons weight is applied.

First observe that according with the description of previous section, also in this case the length of the thread is not fixed, because it decreases as the obelisk rotates. In particular, by simple geometric argument it is easy to determine that to recover the vertical configuration the thread should undergo a contraction $\lambda_{eq} = 0.92$. Based on the previously considered physical parameters (see Fig. 3.7(a)) such a contraction corresponds to consider an elastic spring with a stiffness ratio $\xi = 60$. We may then determine (see Fig. 3.7(b)) the related actuation stress $\sigma_{act} = 47$ MPa. The diameter of the spider silk thread is then calculated by a simple rotational equilibrium equation (see Fig. 3.8(b)) and it results of about 10 cm.

3.6 Actuation properties

To exemplify the incredible actuation properties of the considered spider silk, we may evaluate the work density of previous example as $W_d = \frac{1}{2}\sigma_{act}(1 - \lambda_{\text{silk}}) = 1814\,\frac{\text{kJ}}{\text{m}^3}$. The work density dependence on the boundary conditions is described in Fig. 3.9, where it is represented as a function of the stiffness ratio, supercontraction stretch, and stress. The maximum attainable work density $W_d = 2846\frac{\text{kJ}}{\text{m}^3}$, considering a spider silk density of $\rho = 1.3$ g/cm^3 [Stauffer et al., 1994], corresponds to a work capacity of $W_c = 2.19$ kJ/kg. To the knowledge of the authors, this value is higher than the maximum ever recorded for a moisture powered actuator, namely the hybrid poly(diallyldimethylammonium chloride) Carbon nanotube yarns (PDDA/CNT) actuator providing a work capacity up to $W_c = 2.17$ kJ/kg [Kim et al., 2016]. These values are well above (over 50 times) the mean work capacity of human muscle of $W_c = 0.039$ kJ/kg [Madden et al., 2004].

Eventually, by considering a realistic contraction time of $t = 3$ s [Agnarsson et al., 2009b] we obtain an actuation power density

$$P = \frac{W_d}{\rho\, t} = 730\,\frac{\text{W}}{\text{kg}}. \tag{3.19}$$

Also this value is of absolute relevance compared with human muscle ($P = 50 \div 280$ W/kg) and Carbon Nanotube actuators ($P = 10 \div 270$ W/kg) [Madden et al., 2004].

Further, we represent the operating points in the stress-stretch plot as intersections among the stress-stretch curve of the silk threads and some possible curves of the actuated device, in order to point out some interesting considerations on the actuation properties of the spider silk. In particular, in Fig. 3.6(c)

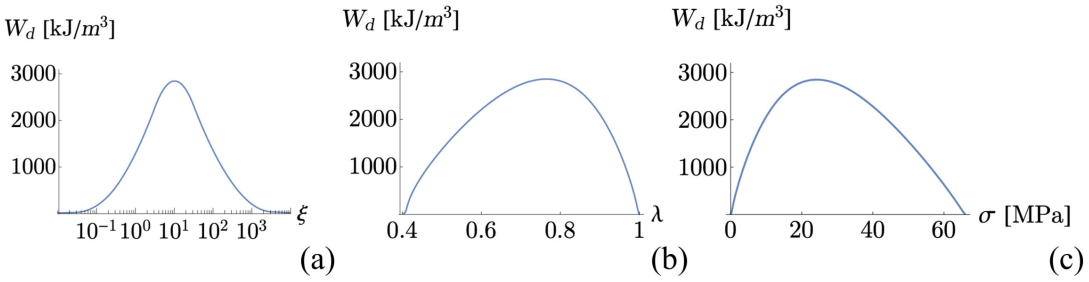

Figure 3.9: Work density generated by spider dragline silk as a function of the stiffness ratio (a), supercontraction stretch (b) and stress (c). Assumed parameters: $E = 10$ MPa, $\lambda_c = 1.5$, $k = 2.22, \mathrm{RH}_c = 70\%, \mathrm{RH} = 100\%, \mu = 1$ MPa

we show the case of linear actuated device, for different RH and for different spring stiffness. In the case of the highest RH (blue curve), we indicate the different operating points (as black circles) and the corresponding work (as red areas below the spring curves). Note that the maximum area is obtained at an 'intermediate' value of the spring modulus, whereas for the highest and lowest spring modulus, the work decreases and it is null with spring modulus ideally going to zero and infinity, corresponding to the special cases of unrestrained and full restrained supercontraction, respectively, typically considered in the experiments. These observations are consistent with those on the work density in Fig. 3.9(a) and remark the crucial role of actuated device stiffness in defining the optimal actuation properties of the system.

3.6.1 Bi-stable actuated device

Eventually we show the possibility of obtaining an even significantly higher work by considering non linear effects. Harnessing bistability and multistability in various structures and soft materials, indeed, has recently sparked increased interest in high-performance soft actuators and soft robots [Chi et al., 2022]. In particular we investigate the interesting case of bistable actuated devices with a region of material instability where the energy is concave and in particular we consider the simple case of a cubic stress-strain relation $\sigma_{spring}(\lambda)$ with a negatively sloped strain domain (see Fig. 3.10(b)). Thus if we write the non-convex energy of the nonlinear actuated device (see Fig. 3.10(a))

$$\Phi_{\text{spring}}(\lambda_{\text{spring}}) = \int_1^{\lambda_{\text{spring}}} \sigma_{\text{spring}}(\lambda)\, d\lambda \tag{3.20}$$

and the silk energy density

$$\Phi_{\text{silk}}(\lambda_{\text{silk}}) = \int_1^{\lambda_{\text{silk}}} \sigma_{\text{silk}}(\lambda)\, d\lambda \tag{3.21}$$

with total energy (assume for simplicity of notation unitary length)

$$\Phi_{\text{tot}}(\lambda_{\text{silk}}) = \Phi_{\text{spring}}(\lambda_{\text{spring}}) + \Phi_{\text{silk}}(\lambda_{\text{silk}}) = \Phi_{\text{spring}}(\lambda_{\text{spring}}) + \Phi_{\text{silk}}(2 - \lambda_{\text{spring}}), \tag{3.22}$$

the stationarity equation

$$\frac{d\Phi_{\text{tot}}(\lambda)}{d\lambda} = 0 \tag{3.23}$$

delivers again the equilibrium condition (3.15). This equation (see Fig. 3.10(d)) has one single solution for low values of RH belonging to the right stable equilibrium branch. When RH is increased the system has two distinct equilibrium solutions. For example in the figure for RH = 75% we denote the solutions by b (first stable branch) and B (second stable branch) with

$$\begin{aligned} \Phi_{\text{tot}}(\text{B}) - \Phi_{\text{tot}}(\text{b}) &= \int_{\lambda_b}^{\lambda_B} \Phi'_{\text{tot}}(\lambda)\, d\lambda \\ &= \int_{\lambda_b}^{\lambda_B} \left(\sigma_{\text{spring}}(\lambda_{\text{spring}}) - \sigma_{\text{silk}}(2 - \lambda_{\text{spring}})\right) d\lambda_{\text{spring}} \end{aligned} \quad . \tag{3.24}$$

We are then lead to the analogue of the Maxwell construction for bistable devices (see *e.g.* [Puglisi and Truskinovsky, 2000]) and the system has a strain discontinuity when the area between the actuated and activation stress-strain systems equals zero. In Fig. 3.10(d) we optimized such condition by choosing the material response of the non-linear spring in such a way that the Maxwell condition is attained exactly in correspondence with the fully wet condition RH = 100% (see the shaded green area in Fig. 3.10(d) and the corresponding energy in Fig. 3.10(c)). Thus if we suppose to increase the humidity starting from the dry condition (point o in the figure) the system has one single equilibrium configuration up to RH $\sim$ 70% (see point a in the figure). For larger values of RH two distinct solutions are possible (the intersections with the negatively sloped part of the sigmoidal curve are unstable equilibrium solutions [Puglisi and Truskinovsky, 2000]), but only when the equal area condition is attained the solution in the larger strain stable equilibrium branch represents the lowest energy. The system then follows the path $o - c - C$ when it 'jumps' to the second branch and the Maxwell condition is attained. As a result the system exhibits a large increase of actuation work (shaded purple area in the figure) that for the same

spider silk properties considered above leads to a work density $W_d = 10900 \frac{\text{kJ}}{\text{m}^3}$, approximately four time larger than the maximum work density calculated for the simplest case of the actuated device composed by the linear spring. The work density obtained by the moisture powered spider silk is higher with respect to many others actuators, included Liquid Crystalline Polymers ($W_d = 56 \frac{\text{kJ}}{\text{m}^3}$), Conductive Polymers ($W_d = 100 \frac{\text{kJ}}{\text{m}^3}$) and Ferroelectric Polymers ($W_d = 1000 \frac{\text{kJ}}{\text{m}^3}$) [Madden et al., 2004]."

3.7 Conclusions

We addressed the issue of quantifying the supercontraction stress arising when a restrained spider silk fiber is hydrated at different humidity conditions and the corresponding actuation properties. By extending the recent approach that we proposed in [Fazio et al., 2022], we deduced a microstructure inspired model taking into account of the H-bond disruption process induced by external humidity in the hydrophilic macromolecular chains composing the spider silk. The num-

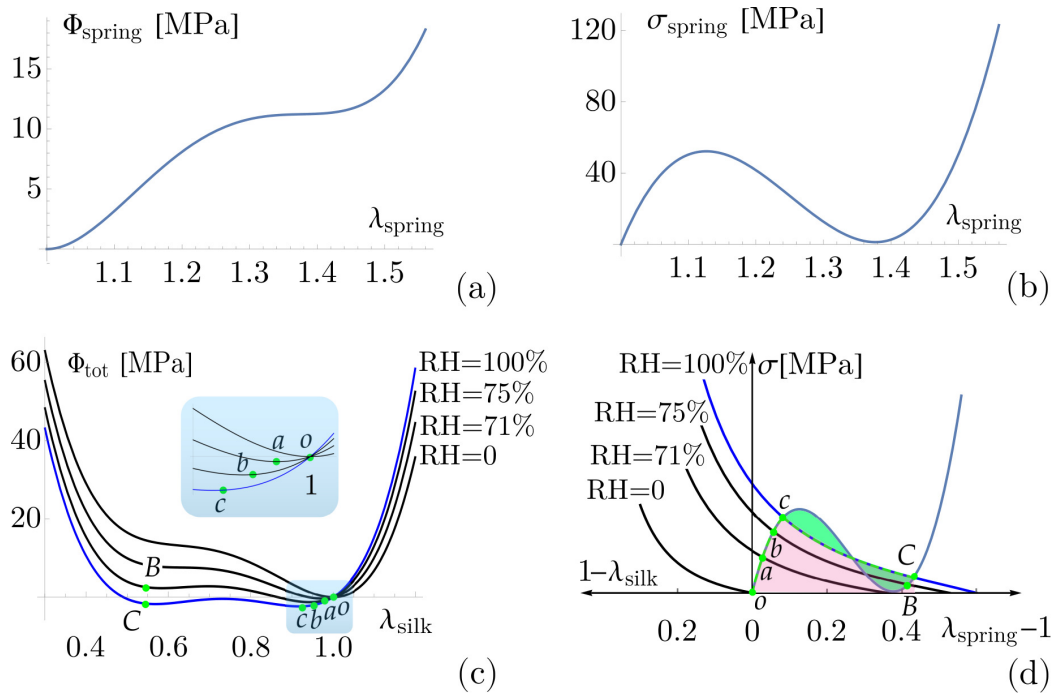

Figure 3.10: Non-convex energy-strain a) and non-monotonic nonlinear relation b) of the actuated device. Dots represents stable equilibrium solutions for different values of RH in the $\Phi_{\text{tot}} - \lambda_{\text{silk}}$ space c) and stress-strain space d). Stress-strain behaviors of the silks at different humidity and different stiffnesses of the springs represented in the same diagram using the compatibility condition (3.16). Assumed values: nonlinear spring $\sigma_{\text{spring}}(\lambda_{\text{spring}}) = 6.5 \times 10^9(\lambda_{\text{spring}}(4.653 + (\lambda_{\text{spring}} - 3.755)\lambda_{\text{spring}}) - 1.898)$; silk, $\lambda_c = 1.5$, $k = 2.22, \text{RH}_c = 70\%, \text{RH} = 0, 71, 75, 100\%, \mu = 1$ MPa.

ber of H-bonds determines then the natural (zero-force) length of a chain that can be quantified by means of classical statistical mechanics results. In particular, the force-elongation relationship is obtained by considering a WLC type energy with a humidity-dependent natural configuration. We then deduce the macroscopic behavior based on the classical affinity hypothesis identifying the macroscopic stretches with the macromolecular ones. The model has been tested against experiments of restrained silk fibers exposed to increasing moisture content while measuring the fiber stress. By considering material parameters of the thread coherent with previously determined ones [Fazio et al., 2022], we obtain that the theoretical prediction well reproduces the experimental behavior. We then exploit the important influence of boundary conditions in determining the maximum supercontraction stress and stretch, neglected by previous literature.

We then apply the proposed model, by considering known experimental values of the material parameters, to predict the supercontraction stress arising in spider threads at different humidities and boundary conditions reproducing effective real biological devices [Greco et al., 2021]. In particular we stress the fundamental role of the stiffness of the constraint device, representing the other spider threads in the web or the possible device considered in the actuation process. This would be clearly of interests in the field of the mechanical actuation by using the humidity as driving load [Dong et al., 2021].

Moreover we determine, based on the introduced elastic interaction hypothesis, the important power and work densities parameters, showing the superior properties of humidity based actuators constituted by spider silks as compared with the most performing artificial materials. Eventually we show the possibility of obtaining even higher work by considering multistable responses of the actuated system.

Chapter 4

Physically based machine learning for hierarchical materials

This chapter is based on the paper published in Cell Reports Physical Science (volume 5, 101790, 2024, 10.1016/j.xcrp.2024.101790) [Fazio et al., 2024] and deals with the possibility of employing data driven techniques to improve the comprehensions of multiscale phenomena and their modelling.

Abstract

We present a novel approach for hierarchical phenomena, which relies on a data modeling technique. Our methodology let us deduce *analytical* relationships, preserves the physical reliability and allows the integration of multiscale theoretical methods.

In multiscale phenomena, complex structure-function relationships emerge across different scales, making predictive modeling challenging. The recent scientific literature is exploring the possibility of leveraging machine learning, with predominant focus on Neural Networks, excelling in data fitting but often lacking insight into essential physical information. We propose the adoption of a symbolic data modelling technique, the "Evolutionary Polynomial Regression", that integrates regression capabilities with the Genetic Programming paradigm, enabling the derivation of explicit analytical formulas, finally delivering a deeper comprehension of the analyzed physical phenomenon. To demonstrate the key advantages of our multiscale numerical approach, we consider the spider silk case. Based on a recent multiscale experimental dataset, we deduce the dependence of the macroscopic behavior from lower scales parameters, also offering insights for improving a recent theoretical model by some of the authors. Our approach

may represent a *proof of concept* for modelling in fields governed by multiscale, hierarchical differential equations.

4.1 Introduction

Multiscale models play a crucial role in different fields of theoretical and applied science, especially due to the increasing possibility of experimental analyses and technologies working down to the micro and nano scales such as Atomic Force Microscopy (AFM), optical tweezers, magnetic tweezers [Bustamante et al., 2000], etc. As a matter of fact, in different fields a huge experimental literature delivering big data libraries on hierarchical systems, starting from the nano and micro scales, up to the macro scale, is now available. These experimental observations represent a potential new fundamental tool for a theoretical advancing in several fields. Such advance requires the deduction of new numerical/theoretical tools delivering correct physical interpretation of impact in engineering [Chen and Pugno, 2013], medicine [Ashley, 2016], physiology [Zhang et al., 2012], biology [McLennan et al., 2012] and physics [Ji et al., 2022].

Within this context, there is a growing debate concerning the need for effective methodologies capable of facilitating the interaction between theoretical insights and empirical data. In this perspective, machine learning approaches, in a broad sense, appear to be the most promising tools. However, it is important to point out that machine learning *per se*, does not inherently possess the capability to automatically incorporate scientific knowledge, which is crucial for avoiding unphysical results. Indeed, in the digital age, the possibility of new instruments such as unprecedented power of calculation and machine learning techniques, has opened up exciting possibilities for analyzing the vast amount of experimental data now accessible. However, such analysis, can lead to a corresponding increase in the theoretical understanding and modelling of the resulting physical system, only if adequate numerical instruments of data modelling are available. On the other hand, as in every transition, the digital transition brings significant risks and drawbacks if not deeply analyzed in its possible effects. Thus, machine learning can lead to a scientific knowledge growth or obscuration, rationalization or unclearness, access to deeper theoretical models or reliance on purely data mining approaches.

Among data-driven techniques, many of which have been developed in recent years, Artificial Neural Networks (ANN) is the most adopted to model complex, non-linear processes including multiscale hierarchical phenomena [Alber et al., 2019]. However, by design, this method inherently leads to black-box represen-

tations of the studied system, often resulting in limited or no interpretability. In Section 4.2, we provide a brief review of this method, highlighting its advantages and disadvantages, along with a selection of examples showcasing its applications to multiscale problems.

In this work, we trace a rational way in the direction of deducing different tools for the modelling of multiscale phenomena based on machine learning techniques, with the potential to significantly advance scientific knowledge. Our approach is distinctive in that it relies on the establishment of fundamental *analytical approximation relations*, crucial to achieve a fully effective and fruitful synergy between data and theoretical modeling. Thus, we recognize that the multiscale character typically corresponds to a hierarchical organization, involving a natural selection of dependent and independent variables. We adopt a Genetic Algorithm-based approach, which deduces analytical relations through a Pareto front type optimization, *i.e.* the Evolutionary Polynomial Regression (EPR) method. A comprehensive description on the history, concepts and motivation is provided in the following dedicated section.

To analyze the efficiency of the proposed approach in treating complex multiscale hierarchical phenomena, we here consider the field of constitutive modelling of complex material behaviors. Specifically, we focus on the paradigmatic example of spider silk, one of the most studied and complex natural materials due to its extreme mechanical properties, particularly its strength and toughness. We rely our analysis on the availability on recent experimental observations on a large number of silks from different spider species from all over the world, where several material properties at different involved scales have been cataloged for the fist time in a comprehensive database [Arakawa et al., 2022]. In this respect, it is worth noting that previous data modeling results are founded only on statistical properties of available data (statistical results based on correlation analysis) [Arakawa et al., 2022], and this allows for a very partial attainment of the potential impact of such experimental results.

This work aims to be general within the framework of a multiscale description of physical phenomena and the deduction of larger scales properties from the structures at lower scales. Indeed in the formulation of the specific case study here analyzed, we have considered three scales starting from the *micro* (protein) scale, to the *macro* scale passing through the *meso* scale. We explicitly impose in our approach that these three scales interact with each other in a hierarchical way. In particular, we consider the three possibilities of deduction of the meso from the micro properties, a successive macro from meso and eventually an interesting direct micro to macro deduction.

Our results show the effectiveness of the proposed method to deduce new physical knowledge on the studied phenomenon. In particular, regarding the considered example of spider silk, among other results, we deduce a functional relation between the thermal degradation temperature and the parameters describing the micro scale proteins structure, a very simple relationship between the diameter of the silk thread and the meso scale properties, and finally, a straightforward and effective relationship that describes how to deduce the macro scale supercontraction property as a direct function of micro scale parameters. Additionally, we identify a meso scale variable that do not depend on the considered micro variables, suggesting the importance of other micro variables in shaping the mesoscale structure of silk material. Thus, even the hierarchical structure of the involved variables results from the proposed approach, suggesting different micro-meso, meso-macro, and micro-macro relationships depending on the considered variables together with the determination of the effective dependent-independent variables.

We show that the proposed methodology also allows to enhance existing physical approaches by increasing the understanding of the underlying physical processes. In this respect, we have also identified possible directions for further investigating some relationships that have already been theorized. To this end, we interpret the obtained results in relation with a recent physically-based model introduced by the authors [Fazio et al., 2022; Fazio et al., 2023].

We argue that this is a first step toward a more effective adoption of the new availability of data and data modelling techniques that can be of fundamental help in several fields of multiscale phenomena as compared with the diffuse ANN physically based approaches.

4.2 Artificial Neural Networks and multiscale phenomena

Among data-driven techniques, many of which have been developed in recent years, Artificial Neural Networks (ANN) is the most adopted to model complex, non-linear processes including multiscale hierarchical phenomena. Loosely speaking, ANN uses models consisting of multiple processing elements (neurons) connected by links of variable weights (parameters) to deduce typically “black-box" representations of the analyzed systems. Learning in ANN involves adjusting the parameters (weights) of interconnections in a highly parametrized system.
In a few words, the main widely recognized disadvantages of ANN model construc-

tion are the curse of dimensionality, overfitting issues and parameter estimation [Haykin, 1999; Giustolisi and Laucelli, 2005]. The well-known curse of dimensionality refers to the exponential increase in the need of parameters when the model input space grows. This means that the number of connections exponentially raises and in a such widened space the training set of input becomes more sparse or the amount of data needed to preserve a constant level of accuracy increases exponentially. On the other hand, in such a way, ANN acquires greater flexibility in mapping events with complex structure. However, this leads often to overfitting problems, that is ANN tends to fit training data too precisely due to the large number of parameters resulting in the propensity to generate poor predictions for events not close to the training data set. A further disadvantage of using ANN is the difficulty of incorporating knowledge derived from known physical laws into the learning process due to the inherent complexity of its framework.
Despite these drawbacks, several significant results in this field have been reported. We may recall that Gu et al. [Gu et al., 2018] have employed finite elements analysis together with convolutional Neural Network algorithms to predict and optimize the toughness of hierarchical composite systems and validated their results through additive manufacturing and testing. Good predictive performances were obtained also by Neural Network methods in linking the elastic properties of composite materials to their meso scale structure, in particular, the three-dimensional microstructure to its effective (homogenized) properties [Cecen et al., 2018]. Recently, in [Linka and Kuhl, 2023] the authors adopted ANN to choose, among an a priori specified class of specific constitutive models depending on the right Cauchy Green deformation tensor invariants, the model that best reproduces stress-strain behaviors under different classes of deformation. While the approach is interesting, it is highly oriented by the specific knowledge of the problem and restricted to the special case when the class of constitutive laws is already known: *i.e.*, the stress dependence on the deformation invariants. With the aim of reducing the computational burden associated with the numerical solution of describing active force in the cardiac muscle tissue, ANN were employed to build a reduced order model starting from high fidelity mathematical models [Regazzoni et al., 2020]. The implications are thus fundamental and let us obtain relevant information for problems that have long been theoretically unresolved, such as the recalled long-lasting problem of predicting the protein structures from amino acid sequences [Baek et al., 2021]. On the other hand, the main drawback in the perspective of extending the knowledge for the theoretical modelling of such phenomena is that ANN leads to “black-box" approaches. There is then a strong limitation on the ‘operational’ advantages due to the lack of interpretabil-

ity of the artificial intelligence results. Some very recent works address this issue [Murdoch et al., 2019; Du et al., 2019], but this is still an open problem [Doshi-Velez and Kim, 2017; Molnar, 2020] due to the intrinsic nature of the approach, summarized above.

4.3 Symbolic Machine-Learning using EPR: History, Concepts and Motivation

Digital transition is defined as the review of processes, using products based on digital products, technologies (hardware) and strategies (software), to increase efficiency. *The simpler, more accessible, and representative collection and evaluation of data relating to processes is the knowledge base to provide useful information for efficiency. The process to be made efficient is here the scientific knowledge.*

The technological event, then, does not explain and is not alone the digital transition. In fact, at the basis of today's digitalization there are always humans who developed the theories, paradigms and concepts that generated the strategies, methods, and algorithms. The latter have not only allowed the development of digital technologies together with supporting the evolution of electronics, but strategies of the digital transition are also the basis to make efficient the processes themselves.

The scientific studies that have generated the possibility of today's *digital transition*, impacting definitively on the development of both *digital strategies*, pertain to the fields of mathematical logic and mathematics. They have developed throughout the last few centuries.

We report some fundamental stages, which, in the past have given rise to the scientific and conditions of the digital transition and the specific symbolic machine-learning strategy here used, named Evolutionary Polynomial Regression [Giustolisi and Savic, 2006].

Alan Turing already in the thirties of the '900 introduces the concepts of algorithms and calculating machines that will later lead to the development of computers. In fact, he is considered the father of information technology and of the concept of *machine learning* which has nowadays entered everyday life with the idea of *artificial intelligence*. It is useful to clarify, without dwelling excessively on the subject, that *machine learning* or *data-modelling* or *data-driven* are more appropriate terms than the generic one of *artificial intelligence*. The latter is, in our opinion, an abused 'slogan', rooted often in motivations distant from scientific

reality, which should not be used for at least two reasons:

- it is misleading with respect to the possibility of digital machines to make all the processes more efficient using an unsupervised strategy, *i.e.*, without the intervention of the human reasoning. This is especially true when digitalization is used to support and make more efficient the progress of the scientific knowledge.

- It is not yet known what human intelligence is, both at the mechanistic and biological level of brain functioning; therefore, it is impossible to build machines that simulate something that is unknown in the intrinsic mechanisms that, in addition, generate human consciousness, for example, a concept itself that is difficult to understand.

The idea of *artificial intelligence* was born with McCulloch and Pitts in 1943 [McCulloch and Pitts, 1943] when they published a work showing a simple system of artificial neurons able to perform basic logical functions. At least in theory, this system could learn in the same way that humans learn by using experience through the trial and error that strengthens or weakens the connections between neurons. *Artificial neural networks* are *machine learning* techniques based on this idea from McCulloch and Pitts. They brought to the today well-known *artificial neural networks* that were already programmed in the first personal computers when Rumelhart, Hinton and Williams developed the Error Back-Propagation method [Rumelhart et al., 1986], in 1986, to train them or rather to calibrate the weights of the "synapses" that connect the neurons simulating, in a "very simplified way", the functioning of the human brain. Note that *artificial neural networks* can be seen today as a category *machine learning* strategies which are based on the original paradigm with developments of the mathematical structure and learning strategies.

As said, we propose the *symbolic machine-learning* strategy, in particular EPR, searching for models using a multi-objective strategy based on evolutionary optimization by genetic algorithms. Thus, as a summary for the reader, we report a brief history of the origin of evolutionary optimization to better understand the motivations of adopting EPR together with a Multi-Objective Genetic Algorithm (EPR-MOGA) [Giustolisi and Savic, 2009].

The idea of *evolutionary optimization* was born in the last century and is nowadays a relevant component of process efficiency strategies, that in our context can be identified with the scientific knowledge. In 1973, Ingo Rechenberg was the pioneer of evolutionary calculation and artificial evolution, [Rechenberg, 1971], whose theories were taken up again in 1975 by John Holland who develops the

theory of *genetic algorithms* reported in the book Adaptation in Natural and Artificial Systems [Holland, 1992]. In 1989, David Goldberg, a student of John Holland and hydraulic engineer, wrote a book [Goldberg, 1989] which became the milestone for the use of *genetic algorithms*.

The tools allowing for the optimization with *evolutionary calculation strategies* such as *genetic algorithms* are essential to attain system efficiency. Indeed, they allow cost-benefit (efficiency) problems to be solved by considering more than a single objective, contrary to most of the classical techniques. Moreover, these strategies allow for the adoption of the so-called *Pareto front of optimal or efficient solutions* [Pareto, 1906] from the cost-benefit point of view or more solutions with different trade-offs based on the objective functions. These may become a *decision support* for the efficiency of any process, that in the case of this paper regards scientific knowledge.

In this respect, we remark that EPR-MOGA uses *genetic algorithm* to search for *symbolic models* of data because the strategy is to search for the best trade-off models complexity vs data fitting. We point out that *Symbolic modelling* of data is a specific strategy internal to the paradigm of *genetic programming*.

In 1992, John Koza developed the paradigm of *genetic programming*, showing [Koza, 1992] the possibility of creating machines that program themselves to solve problems postulated by humans. *Genetic programming* integrates *machine learning*, in a wider sense with respect to the original studies, with *evolutionary optimization* in an original way. Much of what is proposed today as *artificial intelligence* refers to the paradigm of *genetic programming*. *Symbolic modelling* is a specific application of Koza's paradigm to obtain models by means of the integration of *machine learning* and *genetic algorithms* in form of *symbolic formulas* that can be evaluated as such by the expert. This is a paradigm alternative to that of *artificial neural networks* (see Section 4.2), which are general mathematical structures characterized by the "universal" ability of interpolating data, but, for this reason, not suitable for the interpretation of the results with respect to the physical knowledge of the expert and the required cause-effect relation.

Roughly speaking, the key idea of EPR-MOGA is that the domain of formulae interpolating data is very wide also because of the unavoidable data errors. In other words, we can argue that many formulas might exist of different structures which interpolate the same data with a similar accuracy. However, a clear scientific interpretation is crucially favoured by simple polynomial structures because the parameters can be simply evaluated being a problem of linear optimization, *i.e.*, a single set of constants exist given a standard error function oppose to artificial network whose training depends on the initial "guess" of the weights

(parameters).

In summary, EPR is a two-stage strategy: (i) a polynomial structure is selected and (ii) the constants are calculated. Each monomial, is composed of one constant and the product of independent variables. If we assign to each of those independent variables an exponent or they are argument of logarithm or exponential functions we obtain a very wide family of possible, non-linear, formulae models with the same characteristics of being linear in parameters. Thus, EPR models coding is through exponents and functions for independent variables and the maximum number of monomials (parameters). They are prior assumptions of the expert human which are the only candidate for model building.

The model building is based on the evolution of polynomial structures, which are solutions of a genetic algorithm; each solution is assigned as a set of exponents (where the null exponent reduces the number of independent variables and of monomials) that determines the model structure and parameters.

Thus, as explicitly described in the following, a genetic algorithm determines evolving analytical solutions with the single objective of best fitting to data, possibly with constraints such as the statistical relevance of a monomial. Then the algorithm searches for the optimal values of the constant polynomial parameters, based on a sequence of linear optimizations.

The further development of EPR [Giustolisi and Savic, 2009] was to use the MOGA strategy. In this way the optimization searches for the best trade-off of models complexity versus fitting to data. The complexity is measured with two functions to be minimized, the number of monomials and the number of independent variables used. In this way, a Pareto set of symbolic models is obtained with two competing terms: the model parsimony and the fitting to data.

This is a very effective innovation with respect to standard machine-learning, in addition useful in scientific knowledge support. In fact, the expert human assumes the exponents, functions, and independent variables and the strategy returns a front of models which is a decision support for the expert at the end of the model search. The symbolic structure of the Pareto front of models, together with possible recursive functional terms and independent involved variables, allows for the selection of the adopted analytical model by the expert in a more robust way with respect to a pure statistical assessment.

In a few words, EPR-MOGA returns the model formulas with the best trade-offs of complexity (parsimony) versus fitting to data. The expert chooses the best model looking at the whole Pareto front and its symbolic structure, also considering the increase of complexity versus the effective improvement in terms

of fitting to data. In the following section, we give, for the help of the reader, a brief introduction to the mathematical treatment of numerical optimization problems based on EPR algorithms. We refer to other works [Giustolisi and Savic, 2006; Giustolisi and Savic, 2009] for a detailed description of the method.

4.4 EPR algorithm

In the simple setting considered in this paper EPR generates explicit mathematical expressions to fit a set of data points starting from the symbolic equation

$$Y = a_0 + \sum_{j=1}^{m} a_j \mathbf{X}_1^{\mathbf{ES}_{j1}} \mathbf{X}_2^{\mathbf{ES}_{j2}} \ ... \ \mathbf{X}_k^{\mathbf{ES}_{jk}} \tag{4.1}$$

where Y is the considered output dependent variable, $\mathbf{X}$ is the vector of input variables and a_o is a bias term. Thus, we assume that Y can be expressed as a polynomial functions composed of m terms, here represented by products of powers of the $\mathbf{X}_i$ generated by the algorithm. In the more general case, other simple functions can be considered instead of powers [Giustolisi and Savic, 2009]. Observe that, as previously anticipated, each of the m terms is linearly dependent on the unknown parameters a_j. The power exponents $\mathbf{ES}_{ji}$ are selected from a predetermined set of values.

Synthetically, EPR is performed in two steps: a) structure identification and b) parameter estimation. The first stage entails simultaneously determining the best 'arrangement' of the independent variables and the related exponents. A multi-objective genetic algorithm termed OPTIMOGA, which stands for OPTImized Multi-Objective Genetic Algorithm, is used to finalize this optimization. This algorithm is based on the MOGA strategy, introduced in the previous section and extensively described elsewhere [Giustolisi et al., 2004; Giustolisi and Savic, 2009]. We refer the readers to those papers for more detailed information.

We remark that, since the user defines a priori the set of candidate exponents, the possible negligible input variables are obtained by including zero among them. This represents a fundamental option for the important aspect, recalled in the introduction, of determining the effective independent variables. The values of the parameters a_j are determined in a second stage using the linear Least Squares (LS) approach, which minimizes the Sum of Squared Errors (SSE). In addition to the usual LS search, the LS is typically performed by searching for only positive values (constraints $a_j > 0$). This choice can be removed by the EPR algorithm and it can be a fortiori justified in our physical model of spider silks by referring to

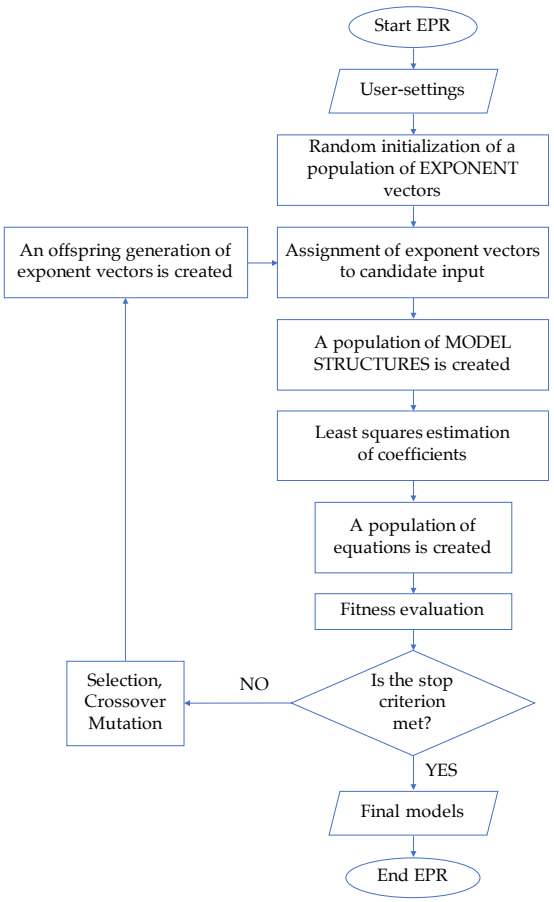

Figure 4.1: Flow-chart of EPR working phases

only positive values of the input and output variables. However, this choice helps in avoiding overfitting, by excluding sequences of terms with negative/positive a_j values that may result from the modelling of the data noise [Giustolisi et al., 2007].

Moreover, the uncertainty of the coefficients (a_j) is evaluated during the search and the distribution of estimated pseudo-polynomial coefficients is used to eliminate those parameters whose value is not sufficiently larger than zero [Giustolisi and Savic, 2004; Giustolisi and Savic, 2006]. Indeed, it may be argued that a low coefficient value with respect to the variance of estimates relates to terms that describe noise rather than the underlying function of the phenomenon being studied.

The algorithm is depicted in the flowchart shown in Figure 4.1. As a starting point, the candidate independent variables, the general polynomial structure, the functions composing the monomials, the candidate exponents, and the maximum number of terms are assigned, possibly based on the initial knowledge of the physical phenomenon. The exponents can reflect the types of relationships between the inputs and output. For example, if the vector of candidate exponents is chosen to be $\mathbf{ES} = [-1, -0.5, 0, 0.5, 1]$, if the maximum number of terms is $m = 4$

and if the candidate independent input variables are $k = 3$, the polynomial regression problem is to find a matrix of exponent $\mathbf{ES}_{4\times3}$. In a first stage, an initial population of matrix of exponents is generated. An example of such a matrix is

$$\mathbf{ES}_{4\times3} = \begin{bmatrix} 1 & 0.5 & 0 \\ 0 & 0 & 1 \\ 0 & -0.5 & 1 \\ -1 & 0 & 0.5 \end{bmatrix} \tag{4.2}$$

so that the Equation (4.1) is:

$$Y = a_o + a_1\,\mathbf{X}_1\,\mathbf{X}_2^{0.5} + a_2\,\mathbf{X}_3 + a_3\,\mathbf{X}_2^{-0.5}\,\mathbf{X}_3 + a_4\,\mathbf{X}_1^{-1}\,\mathbf{X}_3^{0.5} \tag{4.3}$$

The adjustable parameters a_j are then computed by minimizing the SSE as a cost function. It follows the evaluation of the fitness function: if the termination criterion is satisfied, the output results are shown, otherwise a new matrix of exponents is generated through Genetic Algorithm (GA) including crossover, mutation and ranking selection [Giustolisi and Savic, 2006]. Then, again, the adjustable parameters are calculated and the fitness function evaluated until the termination criterion is satisfied.

The equally performing models are those composing the Pareto dominance front [Pareto, 1896; Giustolisi and Savic, 2009] and since EPR returns the whole set of formulae of the Pareto front, the final choice of the model among different possible relations can be based on physical considerations [Giustolisi, 2004]. In this respect, we observe that genetic algorithms generates formulae/models for f, coded in tree structures of variable size, performing a global search of the expression for f as symbolic relationships among $\mathbf{X}$, while the parameters a_j play a role only in the optimization process. On the other hand, ANN goal is to map f, without focusing on the level of knowledge of the functional relationships among $\mathbf{X}$. This is why we argue that EPR represents a better tool for data-driven knowledge discovery.

4.5 Case study: spider silk

Spider silk is one of the most studied natural materials due to its extreme mechanical properties, particularly its strength and toughness, which overcome many high-performance man-made materials. Furthermore, spider silks are regarded as a fundamental material for a new class of high-performance fibers in the context of biomimetics [Greco et al., 2021; Arndt et al., 2022]. The availability of increas-

ingly sophisticated experimental techniques has allowed for a deeper understanding – both chemically and structurally – of the complex multiscale, hierarchical material underlying their notable mechanical behavior. Despite this, many relevant phenomena governing their loading history dependence, rate, temperature, and humidity effects remain unknown [Pérez-Rigueiro et al., 2021].

At the molecular level, spider silks are made up of an amorphous matrix of oligopeptide chains and pseudo-crystalline regions composed primarily of polyalanine β-sheets [Elices et al., 2011; Sponner et al., 2007] with dimensions ranging from 1 to 10 nm [Keten and Buehler, 2010], mostly oriented in the fiber direction [Jenkins et al., 2013]. The radial cross section of the fiber is highly organized [Li et al., 1994; Eisoldt et al., 2011; Sponner et al., 2007]. Furthermore, the chemical and structural composition varies according to the different silks produced by the different glands and, of course, the different species. Here, we focus on the most performing and extensively studied type of silk known as *dragline*.

Many biological examples of evolutionary material optimization represent the possibility of obtaining unreached material performances at the macro scale, based on a clever, hierarchical organization of weak composing materials at the lower scales [Keten et al., 2010]. A further enrichment in biological structure is to possibly include different composing materials [Bosia et al., 2012]. The analytical description of how the macroscopic performances result from these complex low scale material organizations is far from being reached and represents a coveted benchmark not only for their theoretical interest, but also in the crucial field of bioinspired material design [Arndt et al., 2022; Liu et al., 2016].

4.5.1 Micro scale

Spider dragline silk fibers (also known as Major Ampullate silk) are constituted by structural proteins called Spidroins, which are divided into two major subtypes, MaSp1 and MaSp2. The overall sequence architectures of the two subtypes are similar, with a highly repetitive core region flanked by small N-terminal and C-terminal domains (NTD and CTD, respectively). The repetitive regions, which account for 90% of the primary structure, are composed of alternating runs of polyalanine and multiple glycine-rich motifs arrayed in tandem. Moreover, very recent studies, prompted primarily by advances in proteomics and sequencing technologies, paint a more complex picture of dragline silk composition than a simple MaSp1/MaSp2 dichotomy [Arakawa et al., 2022]. Despite this complexity, here we only consider the protein MaSp1 and MaSp2 which are widely recognized as the two main composing the spider silk. From the secondary structure point

of view, the MaSp1 is mainly organized into pseudo-crystalline polyalanine β-pleated sheets [Li et al., 1994; Brown et al., 2011]. On the other hand, the MaSp2 is mainly constituted by proteins with a proline content preventing the formation of β-sheet crystals [Sponner et al., 2007], resulting in a structure with significantly lower crystallinity and macromolecules with weaker crystal domains, typically in the form of α-helix and β-turns [Sponner et al., 2007; Nova et al., 2010].

We remark that, as recognized in polymer mechanics [Flory and Erman, 1982] and described also for the spider silk case in [Fazio et al., 2022], the number of monomers of the macromolecule (*i.e.*, protein for the silk case) is fundamental for the mechanical behavior of the material. Based on the fact that (i) the mechanical behavior of the spider silk material is to be ascribed to the repetitive region features more than the terminal region of the protein [Hayashi et al., 1999], and (ii) the pseudo-crystalline β-sheets, mainly present in the MaSp1, are recognized to be the most impactful feature in determining the exceptional strength of the spider silk [Yarger et al., 2018], here we consider the following three quantities describing the protein scale of the silk material.

- Length of the repetitive region of the protein MaSp1 in terms of number of amino acids
- Length of the repetitive region of the protein MaSp2 in terms of number of amino acids
- Length of the polyalanine β-sheet in the protein MaSp1 in terms of number of alanine amino acid

4.5.2 Meso scale

At the meso scale we consider the proteins' secondary structure, how macromolecules are arranged in the fiber and properties regarding the chemical and structural stability of the polymer. In particular, we analyze the following material properties:

- Birefringence. It reflects the degree of molecular orientation of silk protein chains. The birefringence of the dragline silk fiber was calculated from the retardation value and silk fiber diameter [Arakawa et al., 2022].
- Degree of crystallinity. It was calculated based on wide-angle x-ray scattering (WAXS) analysis [Arakawa et al., 2022]. In particular, it was obtained

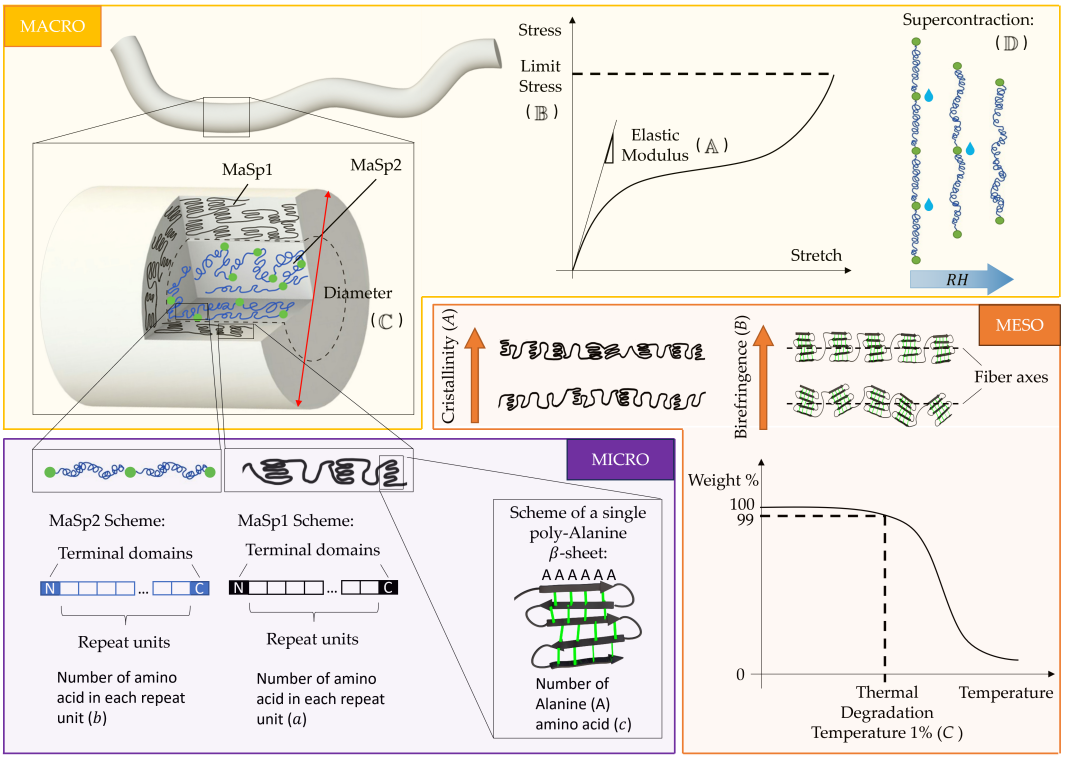

Figure 4.2: Scheme of the hierarchical structure of spider silks and of the considered variables at the different scales in the data modelling analysis.

as the ratio of the total area of the separated crystalline scattering components to that of the crystalline and amorphous scattering components as resulting from the 1D profile obtained by the two-dimensional (2D) diffraction.

- Degradation temperature. This quantity gives a measure of the chemical and structural stability of the silk. In [Arakawa et al., 2022] the thermal degradation temperature has been defined as the temperature that yielded 1% weight loss in the silk samples. Indeed, heating leads to changes of the molecular weight that in turn decreases the mass due to the production of gaseous byproducts of the chemical reactions.

4.5.3 Macro scale

Spider silk is a very interesting material from the point of view of its mechanical performance at the macroscopic scale. In particular, here we focus on the material stiffness and strength. The Young's modulus, of the order of tens of GPa, is above man-made polymers and at the top among other natural materials. The strength is even more interesting, being comparable with high strength steels (order of

magnitude of 1 GPa) and with the most performing man-made composites like the carbon and kevlar reinforced composites [Ashby and Johnson, 2013]. The reason for these outstanding properties with respect to standard materials, is not yet clear, with a relevant role also of the extremely small diameter of dragline spider silk [Porter et al., 2013]. For this reason, we also consider the diameter in the properties at the macro scale. Finally, we address the very significant role of hydration in the material behavior of spider silks. Indeed, a striking effect observed in spider silks is the so called *supercontraction*, addressed, to the knowledge of the author, for the first time in 1977 [Work, 1977], that occurs when a spider silk thread is exposed to humidity. Depending on the silk composition, the experiments show the existence of a Relative Humidity (RH) threshold beyond which the fiber contracts up to half of its initial (dry) length. This also results in the possibility of exploiting the supercontraction in the actuation field [Fazio et al., 2023]. The experimentally observed contraction depends on several factors, including spider species [Boutry and Blackledge, 2010], type of silk (among the up seven different ones that some spiders can produce [Vollrath, 1992; Gosline et al., 1994]), environmental conditions [Plaza et al., 2006] and hydration rate [Agnarsson et al., 2009a]. The quantities we consider at the macro scale are therefore the following:

- Young's modulus, obtained from the stress-strain curves determined through tensile tests of single dragline silk fibers conducted at 25°C and RH$\approx$ 50% [Arakawa et al., 2022].
- Tensile strength, calculated as the breaking force determined by tensile test divided by the undeformed cross-sectional areas of the fiber samples determined by SEM observations [Arakawa et al., 2022].
- Diameter, determined by SEM observations [Arakawa et al., 2022].
- Maximum supercontraction, calculated as $(L_0 - L_f)/L_0$, where L_0 is the length in dry condition and L_f in fully wet conditions (RH=100%) [Arakawa et al., 2022].

4.6 Modelling strategy

For all the EPR run the maximum number of terms has been set to 3 and the chosen set of candidate exponents has been $[-1, -0.5, 0, 0.5, 1]$ to keep the expression as simple as possible thus letting their physical interpretability. Moreover, the ex-

pressions were optimized with a bias term a_o since this element may compensate for the possible lack of relevant inputs in the model.

The choice of the maximum number of terms is justified by comparing the expressions provided in the appendix and their corresponding performances in Figure 4.4. Indeed, we achieve nearly maximal performance with just one or two terms, and adding a third term to the expression does not result in a significant improvement in fitting performance. In any case, this represents an optimization parameter that can be easily varied. Moreover, the choice of the candidate exponents can be considered as the simplest choice to consider the important possibility of determining the non relevance of a candidate input (0), of a linear direct or inverse dependence (1 and -1) and just a simple non linear direct or inverse dependence (0.5 and -0.5). Other more rich choices, depending on the considered problem, could be considered. As a matter of fact, the proposed model, differently from the widely used ANN approaches, requires a systematic connection between the scientist and the machine learning results. All these choices are therefore guided by the specific physical problem at hand. Thus, they are part of the modeling and of the the scientist's physically guided data preprocessing.

In this regard, we also remark that we considered the possibility of more complex elementary functions, and verified the optimality of our choice of power functions. It's worth noting that such comparisons are computationally inexpensive compared to similar possibilities in ANN, which is another notable advantage of the approach here proposed.

Once again, we refer the readers interested to the numerical performances of EPR to e.g. [Giustolisi and Savic, 2006; Giustolisi and Savic, 2009] and References therein. Here we aim to focus on the applicability of such already tested numerical efficient approach to hierarchical problems in material science, a field of wide interest in the recent literature on physically based data modelling techniques.

As anticipated above, to exemplify the proposed approach, we focus on the challenging case of spider silks. In particular, we refer to the recent experimental campaign proposed in [Arakawa et al., 2022], where the authors analyzed the properties, at different scales, of approximately 1000 different silks. From our perspective, this presents a significant opportunity for scientists interested in unraveling the 'secrets' behind the remarkable mechanical properties of this material in relation to its hierarchical structure.

In the original paper, the authors already proposed a statistical correlation analysis, based on the classical study of the Pearson correlation coefficient. Here we show how, our new data modeling approach, combined with our theoretical understanding of the model, especially as presented in [Fazio et al., 2022; Fazio

et al., 2023], allows us to gain deeper physical insights in the considered experimental data.

Among the material properties analyzed in the paper, we chose the ones reported in Table 4.1 with the corresponding adopted symbols[1] (the type of font distinguishes the scales). The role of these variables in the material hierarchical structure and response of spider silk is schematized in the Figure 4.2.

Micro scale	a	Length of the repetitive region of MaSp1
	b	Length of the repetitive region of MaSp2
	c	Length of the polyalanine β-sheet in the MaSp1
Meso scale	A	Crystallinity
	B	Birefringence
	C	Thermal degradation temperature (1% loss)
Macro scale	$\mathbb{A}$	Young's modulus
	$\mathbb{B}$	Tensile strength
	$\mathbb{C}$	Diameter
	$\mathbb{D}$	Supercontraction

Table 4.1: Material properties considered for the data modelling case study divided by scales.

As a main parameter of accuracy, we report for the different numerical results the Coefficient of Determination[2] R^2. The physical valence of the expressions found is discussed, by following [Arakawa et al., 2022], also through the comparison with the correlation matrix represented in Fig. 4.3 obtained by calculating the Pearson correlation coefficient[3] between the different considered variables for the analyzed silks. Observe that, from the database [Arakawa et al., 2022] we considered only the data where the searched output and the considered input are reported simultaneously. Thus, since there are some experimental properties missing for some silks of the database, the number of silks composing the training

[1] Each quantity is considered with the unit of measurement reported in the original database, namely GPa for Young's modulus and limit stress, μm for the diameter, °C for the thermal degradation temperature. All the micro scale properties are expressed in terms of number of amino acids, whereas the supercontraction and the crystallinity are two nondimensionional quantities ranging in (0,1). As a result, the parameters a_j (see Eqn. (4.1)) estimated by means the minimization of the SSE, can be dimensional quantities.

[2] Here we consider the classical definition $R^2 = 1 - \sum_{i=1}^{N} \frac{(x_i^{num} - x_i^{exp})^2}{(x_i^{exp} - \bar{x}^{exp})^2}$, where the x_i^{num} are the output variables of the numerical test and x_i^{exp} are the corresponding experimental values, with $i = 1, ...N$, where N is the number of experimental observations considered as dependent variables. Observe that EPR also considers other, not explicitly reported here, indicators of performance, *e.g.* the sum of squared errors (SSE). As a result R^2 does not necessarily increase as the complexity of the expressions grows.

[3] It is a measure of linear correlation between two sets of data $\{x_i, i = 1, ..., n\}$ and $\{y_i, i = 1, ..., n\}$ with n the number of data, defined as $\rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$, where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ are the mean values.

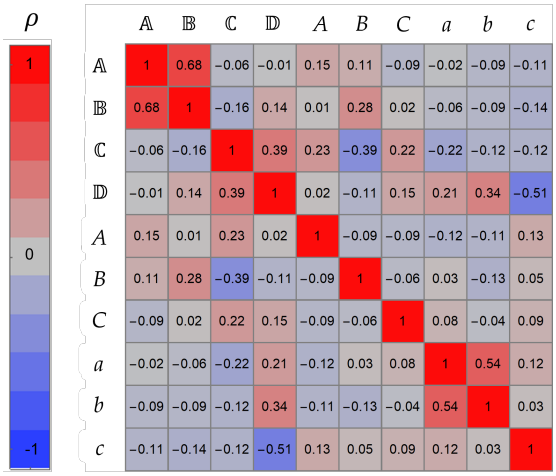

Figure 4.3: Experimental Pearson correlations among the material properties at the three scales considering all the silk reported in [Arakawa et al., 2022].

set is different for each considered output. In so doing, for each target output, we consider the maximum possible available information, in terms of number of silks.

The EPR technique has returned a series of polynomial expressions for each dependent variable. In Fig. 4.4(d,e,f) we report the variation of the accuracy of the analytical expressions in reproducing the experimental data. Thus, in a Pareto front approach, the model let us choose the best formulae considering parsimony (simple expression) and accuracy. Observe that, the analysis of the whole expression set, reported in the Appendix A, allows for a rational choice of the most suitable material relations (reported in Table 4.2) that can be selected considering not only parsimony and performance, but also analyzing the physical interpretation of the experimental matrix correlations in Fig. 4.3.

4.7 Results and Discussion

In this section, we consider different possible data analysis. Specifically, we examine three potential deduction scenarios: deriving meso from micro properties, then macro from meso properties, and eventually a direct deduction from micro to macro. On one hand, this facilitates the identification of relationships between variables and the potential analytical forms of these relationships. On the other hand, it highlights the significant role of scales in the considered model and the analysis of the best-performing multiscale functional relationships.

In particular, we point out that, based on the possibility of analyzing the functional relations at different scales with immediate control over functionality

and accuracy, we do not assume in advance that a sequential micro-meso-macro variable dependence is the most reasonable as typical of multiscale approaches. We instead suppose, that also a direct micro-macro variables relations can be observed.

In what follows, we show that this is actually the case and we find a direct effective relation between a micro and a macro variable. Moreover, we obtain that one of the meso variables does not depend from the considered micro variables, thus suggesting the possibility that other micro variables could be important for the meso scale structure of the silk material. These results exhibit the efficiency of the model in selecting the correct functional dependence, the possibility of missing dependence among measured variables and the role of the complex interactions among the different involved scales.

4.7.1 Meso from micro

Firstly, the meso scale properties have been calculated using all the micro scale quantities as independent variables (see Fig. 4.4(a)) according to Eq. (4.1). The results of the accuracy are reported in Fig. 4.4(d) and the resulting functional dependencies are reported in Table 4.2.

As a common property of the considered numerical tests (see in particular the variables A and C in Fig. 4.4(d)), we may typically distinguish two regimes of the performance curves. In the first regime, the performance increases rapidly with the number of expressions and thus with the model's complexity. In the second regime, the performance curves stabilize in a saturation band. This indicates an easy way of selecting an optimal model complexity and to avoid overfitting due to possible noise of the considered data.

Regarding the selected functional dependence, first, we observe that the crystallinity A decreases with b, in accordance with the general correlation matrix (Fig. 4.3). The presence of the bias term is coherent with the value of $R^2 = 11\%$, since, as recalled before, the bias may compensate for the lack of relevant inputs in the model.

The birefringence B shows a very low accuracy $R^2 < 5\%$, coherent with the experimental results that show a very low Pearson correlation between B and a, b, c (see Fig. 4.3). We remark that, in this case, the EPR method avoided data overfitting that could have resulted in more performing, but physically misleading expressions deduced by other numerical approaches. We therefore conclude, in this case, that the considered meso scale quantity, the birefringence B, cannot be predicted starting from the considered micro scale properties and we consider

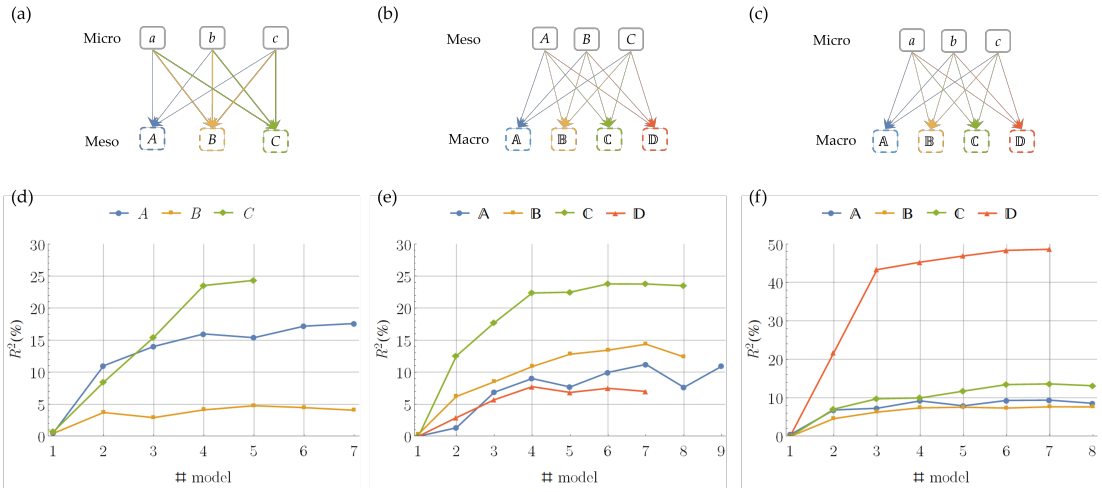

Figure 4.4: Prediction of material properties using two scales at a time: (a,d) meso from micro experimental properties, (b,e) macro from meso experimental properties, (c,f) macro from micro experimental properties. (a,b,c) Scheme of the strategy to obtain each quantity: solid (Dashed) box indicates experimental (obtained from EPR) quantities. (d,e,f) EPR Model performance in terms of R^2 plotted against the number of the found model.

instead the B as an independent variable to compute the macro scale quantities in the following.

On the other hand, in the case of the Thermal Degradation Temperature C, the EPR found expressions with higher R^2. In this case, the selected expression provides a quantitative estimate of the target quantity with a trend increasing with a and c, in accordance with the experimental correlation matrix in Fig. 4.3.

Scale	Expression	R^2(%)	# model
Meso from micro	$A = 3.5562\frac{1}{b} + 0.10262$	11	2
	$C = 3186.7046\frac{1}{a} + 0.86787ac^{0.5} + 45.2672$	23.54	4
Macro from meso	$\mathbb{A} = 0.091301\frac{B^{0.5}}{A} + 29.2668A$	9.01	4
	$\mathbb{B} = 0.013837\frac{B^{0.5}}{A} + 0.014276AC$	12.81	5
	$\mathbb{C} = 0.81928\frac{A^{0.5}C}{B}$	22.37	4
Macro from Micro	$\mathbb{D} = +0.061926\frac{b}{c} + 0.0047393$	43.35	3

Table 4.2: Prediction across two scales: selected explicit expressions

4.7.2 Macro-meso

As a second data modelling analysis, we consider the possible functional dependence of the macro properties from the meso scale quantities, now considered as independent variables (see Fig. 4.4(b). The results of the accuracy are reported in Fig. 4.4(e) and the resulting functional dependencies are reported in Table 4.2.

Regarding the Young's modulus $\mathbb{A}$, the chosen expression correctly reports the monotonic growth with crystallinity A, as can be immediately deduced by comparing the derivative of the expression for $A > 0$ with the experimental correlation matrix.

Regarding the limit stress $\mathbb{B}$, the selected expression correctly reports the highest experimental correlation, namely the positive one with the Birefringence B.

The expression chosen for the diameter $\mathbb{C}$ has the highest accuracy among the macro-meso case ($R^2 = 22.37\%$) with a very simple expression composed of only a single term that includes all three variables at the meso scale. The correlation is positive for A and C and negative for B in accordance with the experiments.

For this last case, in Table 4.3, we report the complete Pareto front of formulas obtained as the output of the EPR method, along with the corresponding accuracy (R^2) for each expression. It is worth noting that the trend of accuracy concerning the model's complexity is illustrated in Figure 4.4(e). By comparing the set of expressions with the accuracy of the 8 models found by EPR, we observe a rapid increase in accuracy ($R^2 = 0 \rightarrow R^2 \approx 12.5$) when the inverse dependence on variable B is introduced in the model 2. A further significant increase in accuracy ($R^2 \approx 12.5 \rightarrow R^2 \approx 18$) is achieved by considering the dependency on $A^{0.5}$ in the model 3. The last substantial accuracy improvement ($R^2 \approx 18 \rightarrow R^2 \approx 22.5$) is obtained by including the linear dependence on variable C in the model 5. This expression is considered the most suitable for describing the relationship between the diameter and the meso scale variables. It combines relatively high accuracy with a simple and interpretable structure. The models from 6 to 8, while more complex, do not significantly enhance predictive accuracy ($R^2 \approx 22.5 \rightarrow R^2 \approx 23.5$). This implies that the additional terms in these expressions, compared to the previous one, describe noise in the data rather than playing a real physical role in the considered phenomenon.

Finally, we note that the dependence on $1/B$ is maintained from the model 2 to the model 4, indicating the robustness of this relationship. It is considered reliable as it was preserved even when EPR attempted to reduce expression complexity. On the other hand, the analysis of the complete Pareto front permits the identification of terms that appear in only one model (see the models 5 to

# model	Expression	$R^2(\%)$
1	$\mathbb{C} = 2$	0
2	$\mathbb{C} = 81.9474\frac{1}{B}$	12.53
3	$\mathbb{C} = 177.4203\frac{A^{0.5}}{B} + 0.051737$	17.73
4	$\mathbb{C} = 0.81928\frac{A^{0.5}C}{B}$	22.37
5	$\mathbb{C} = 0.00021001C + 0.80165\frac{A^{0.5}C}{B}$	22.49
6	$\mathbb{C} = 0.0037544\frac{C}{B^{0.5}} + 0.76892\frac{A^{0.5}C}{B}$	23.79
7	$\mathbb{C} = 0.011782\frac{C}{B} + 0.0025068\frac{C}{B^{0.5}} + 0.76003\frac{A^{0.5}C}{B}$	23.77
8	$\mathbb{C} = 0.028813\frac{C}{B} + 0.69464\frac{A^{0.5}C}{B} + 0.010027\frac{A^{0.5}C}{B^{0.5}}$	23.50

Table 4.3: Pareto front of models for predicting the Diameter ($\mathbb{C}$) from the meso scale variables

8); such terms are likely to be weakly related to the physical phenomenon, but rather specific to the error present in the data. Similar considerations apply when examining the complete Pareto fronts for each considered output variable, as reported in Appendix A. The Pareto front has been extensively discussed in this case, which is particularly suitable for explanatory purposes due to the shape of the Pareto front (Figure 4.4(e)), combining a rapid increase for the initial models and a clear performance saturation band for the more complex models.

We remark that the final choice of the appropriate equation requires an evaluation of the resulting physical consequences. This may necessitate further experimental and theoretical investigations, as is common in the analysis of any scientific open problem. In our opinion, it is only a continuous efficient interaction between data modelling, with analytical formulas, and scientific interpretation of them, that can ensure the desired advancement of the understanding of the physical phenomena.

Eventually, we consider the selected expression for the supercontraction $\mathbb{D}$. In this case, we are not able to produce a good estimate of the target output from the meso variables ($R^2 < 8\%$).

4.7.3 Macro-micro

As anticipated, we now consider the possibility of direct dependence between macro and micro variables. Thus, the macro properties have been calculated also using all the micro scale quantities as independent variables (see Fig. 4.4(c)). The results of the accuracy are reported in Fig. 4.4(f) and the resulting functional

dependencies are reported in Table 4.2.

In this case, regarding the Young's modulus ($\mathbb{A}$), the limit stress ($\mathbb{B}$) and the diameter ($\mathbb{C}$), the values of R^2 are generally low. On the other hand, the supercontraction $\mathbb{D}$ is predicted with a relatively high accuracy ($R^2 > 40\%$), and the selected expression ($R^2 = 43.35\%$) provides a reasonably precise quantitative estimate of the supercontraction, higher than the ones deduced from the meso variables. This suggests the intriguing possibility of a direct influence of micro variables on the macroscopic supercontraction variable representing still a debated effect of spider silks behavior [Fazio et al., 2022; Cohen et al., 2021]. Moreover, we remark the simplicity of the obtained analytical expression including the two most relevant experimental correlations between the supercontraction and the micro scale properties, *i.e.* the positive one with b and the negative one with c. This result demands by itself a theoretical investigation, which will be the focus of our future studies.

In summary, we observe that by employing this direct macro-micro deduction, from one side we obtain a relatively precise estimation of the supercontraction property that was missing from the macro-meso analysis, but from a modelling point of view, we deduce the possibility of modelling the supercontrction as a macro variable with a direct functional dependence from the micro ones. Moreover, the low accuracy in predicting the other macro variables ($\mathbb{A}$, $\mathbb{B}$, $\mathbb{C}$) directly from the micro ones, enlighten the importance of the meso scale structures in generally determining the macro properties of the material, as expected from the classical hierarchical dependence. For the particular case of the spider silk, this reflects established results in literature pointing out the dependence of the silk thread macroscopic behavior from the secondary structures of the proteins [Yazawa et al., 2020; Hayashi et al., 1999], here described by the meso scale variables.

4.7.4 Discussion on the accuracy of the EPR formulae for the spider silk case

A general comment is in order on the accuracy of the relationship found by EPR. The coefficient of determination of expressions found by the EPR method is generally low if compared with other frameworks where EPR was applied [Berardi et al., 2008; Creaco et al., 2016; Balf et al., 2018], but this was expected for the study case of spider silks, as in biological materials a high intrinsic variability for experimental observations is known [Cook et al., 2014]. Also, for this particular material, a meaningful variability of the mechanical property of silks taken from

the same individual under similar conditions is well recognized (see e.g. [Madsen et al., 1999]). Further, the characteristics of the spider silks have high sensitivity to a large number of parameters, among which starvation, reeling speed [Madsen et al., 1999] other than the more expected spider species [Boutry and Blackledge, 2010], type of silk (among the up seven different ones that some spiders can produce [Vollrath, 1992; Gosline et al., 1994]), environmental conditions [Plaza et al., 2006] and hydration conditions [Agnarsson et al., 2009a]. In a very recent work [Greco et al., 2022], the variability of spider silk properties has been directly compared to that of carbon fibers, and significantly higher variability in spider silk in all properties considered has been reported. For these reasons, even if the R^2 of the expressions found by means EPR is generally not as high as other frameworks, the performance of the data modeling strategies are considered satisfactory. Furthermore, we believe that the results we have described strongly demonstrate the feasibility of our proposed approach when compared to the more commonly used approaches, typically based on ANNs. This approach allows us to deduce both analytical results and important physical properties related to the problem at hand, thereby establishing a new way of investigation in the considered field.

4.8 Theoretical vs experimental correlations

While the objective of this paper is general and mainly related to the exhibited possibilities of obtaining information on the considered physical properties, in this section, we show operatively this possibility by comparing experimental and theoretical results.

Going to the considered case of spider silks, we are now in the position of deducing the theoretical meso and macro response based on the only micro properties to compare with the experimental meso and macro response. Specifically, we have one set of data for which all properties are experimentally known, and on the other hand, we construct a set of theoretical data based solely on the experimentally determined properties at the micro scale. The properties at higher scales (*i.e.*, meso and macro) of the theoretical dataset are then calculated using the explicit relationships selected in Table 2. The purpose is to demonstrate the applicability of EPR-derived relationships to predict macro scale properties based on micro-scale knowledge. We have chosen to utilize pairwise correlations between variables as a means of comparison between experimental data and the relationships learned through Evolutionary Polynomial Regression (EPR).

Coherently with the hierarchical assumption of our model, we first deduce the meso scale variables by the Micro ones and then, based on previous analytical

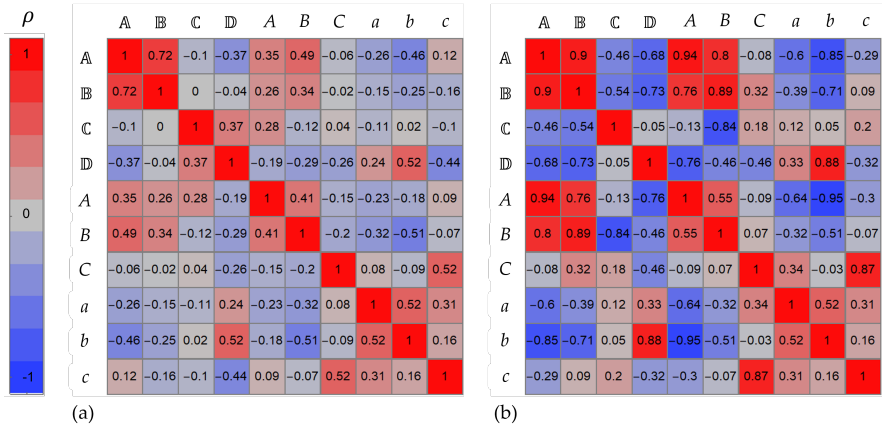

Figure 4.5: Pearson correlations among the material properties at the three scales. (a) Experimental correlations obtained considering a subset of silks for which all the analyzed properties are reported simultaneously (b) Correlations among the material properties obtained from the data modelling EPR approach (macro and meso) starting from the known micro experimental properties.

results, we deduce the macro variables. Accordingly with previously described numerical data analysis, also the meso variable B (birefringence) is considered here as an independent variable. On the other hand, in the special case of the supercontraction $\mathbb{D}$, we assume that it directly depends on micro variables. Notice that all these relevant physical information has been deduced by previous data modelling.

Regarding the experimental data, we consider a subset of the silks analyzed by [Arakawa et al., 2022] and in particular only those for which all the 10 considered properties (see Table 4.1) are known simultaneously (the so obtained subset consists of 35 silks). As a possible comparison between the theoretical and experimental data sets, we consider the Pearson correlation coefficients for each pair of properties. The comparisons of the correlations tables for theoretical and experimental results are reported in Fig. 4.5. They show a satisfactory correspondence almost extensible to all the data and a satisfying result in terms of the values of the correlation coefficients. To get a global comparison we also adopt a positive definite relative error

$$e_r = \frac{|e_a|}{e_m} \tag{4.4}$$

where $e_a = \rho_t - \rho_e$ is the absolute error, ρ_t and ρ_e are the theoretical and experimental Pearson coefficients, respectively. Here e_m is the mean error that since ρ_m and ρ_e range in the interval $(-1, 1)$, we assume as $e_m = 1$. The average value of the relative error by considering all the possible pairs of the correlation matrix $\overline{e_r} = 0.33$, with $0 < \overline{e_r} < 2$, indicates that the functional dependence found

by the EPR method reproduces in a reasonably accurate way the experimental correlations among the properties of spider silks.

Eventually, as evidenced in [Linka and Kuhl, 2023], an important extension of the proposed approach would be to consider a Bayesian framework for the uncertainty quantification in order to compute each output in terms of statistical distribution with a mean and a confidence interval by also taking into account the input data variability.

4.8.1 Towards integrating data driven knowledge and physical modelling

The possibility of advancing, based on the proposed approach, the understanding of the underlying physical relationships, can be exhibited by considering possible progresses in existing theoretical settings. To this end, we here explicitly refer to the recent works previously proposed by some of the authors [Fazio et al., 2022; Fazio et al., 2023], where the dependence of supercontraction on the MaSp2 protein has been addressed. In that context, the MaSp2 protein was treated using the classical approach of multiscale analysis of soft macromolecular materials, based on classical statistical mechanics approach to quantify the expected length of the protein's macromolecules. On the other hand, based on the EPR method, an explicit relationship between the length of the repetitive unit of the MaSp2 protein and supercontraction can be inferred. As an extension of this work, this relationship will be employed to enhance the microstructure-based model considered in the previous works [Fazio et al., 2022; Fazio et al., 2023]. Presumably, from the EPR findings, it will be possible to establish a quantitative relationship between micro scale variables describing the primary structure of the MaSp2 protein and supercontraction. In other words, owing to the interpretable relationships obtained through EPR, it is feasible to extend the prediction of macroscopic supercontraction behavior towards the precise primary structure of the involved proteins. Indeed, in the previously proposed theoretical multiscale approach, this prediction was based on more general properties of macromolecules behavior without specifying the detailed primary structure properties obtained in previous analysis. This example is just one illustration of our approach, but it serves as a representative instance of how our work intends to improve the theoretical understanding in material science. With each of the relationships considered in Table 2, and more broadly, as we explore the Pareto front of expressions, additional relationships discovered among variables are subjects of ongoing study by the authors. These investigations aim to contribute significantly to the expansion of our

knowledge concerning the multiscale mechanisms responsible for the remarkable characteristics of spider silk. This extension is the subject of the forthcoming research of the authors.

4.9 Conclusions

We showed the possibility of adopting, based on a Genetic Programming approach, data modelling techniques, innovative in thew field of material science, particularly suitable for the deduction of analytical models for multiscale problems. Our approach is based on the Evolutionary Polynomial Regression (EPR) method which, as we showed, lets us deduce models that are both accurate and simple, able to describe the dependence of macro scale variables from the ones at lower scales, in their hierarchical order itself deduced by a careful analysis of the data. The best performing models are those located on the Pareto dominance front, which takes into account both accuracy and parsimony and are returned by the EPR algorithm. The final choice of the model can be then based on physical considerations.

To explicitly show the possibility of acquiring physical insight in a complex multiscale problem, and to evidence the key advantages of our multiscale approach compared to classical, non-physically based techniques, we referred to the materials science field and in particular to the complex case of spider silk: a biological material with exceptional properties hugely analyzed also in the spirit of bioinspiration. The choice of this specific case is due to the observation that such remarkable properties are strictly based on an evolutionary hierarchical optimization and the macroscopic spider silk behavior is the result of noteworthy mesostructures emerging from the aggregation of amino acids at the molecular scale. For this intriguing and very complex material many phenomena underlying the multiscale structure and the complex energetic exchanges among the scales ensuring their notable properties are still strongly unclear. We then used this paradigmatic example to show how the presented data modelling approach can be useful in several directions: determine dependent and independent variables, indicate their hierarchical organization, deduce explicit relations among different groups of variables. In this direction we also want to remark that a possible important role can be attained by a following dimensional analysis (e.g. Buckingham theorem) that should be related to the possible absence of variables at the different scales. This is another important aspect that is beyond the scope of this paper and will be the subject of our future investigation.

Furthermore, we showed that the proposed approach let us overcome the

overfitting problem typically observed in the analysis of big data within the ANN framework diffusely adopted in this field. Based on this, new physical knowledge is acquired, that can be used as a starting point for determining new analytical models, suggesting new experiments, and applying more focused data modelling analysis. We also strive to enrich existing physical approaches by enhancing our comprehension of the underlying physical processes. In this context, we investigate the potential for enhancing some authors' previously introduced physically-based theoretical analyses, leveraging the insights obtained from our current approach.

In this sense, we assert that Machine Learning or Artificial Intelligence can have a significant impact on scientific knowledge only if the data modelling approaches are in continuous synergy with the scientific interpretation of the results. We argue thus that a new mixed genetic programming - theoretical approach can be a new fruitful approach in material science, but also in fields as diverse as biology and medicine.

Chapter 5

Conclusions and future perspectives

Throughout this thesis, we explored multiscale models and physically based data-driven techniques for hierarchical materials, with a specific emphasis on biological and bioinspired materials and in particular spider silks. Here we summarise the key steps and achievements. In Chapter 1 we outlined the motivation and context of the subject of this thesis. We then introduced hierarchical biological and bioinspired materials, looking in detail at some typical characteristics in biomaterials of engineering interest. In particular, we described the crucial role of sacrificial bonds and analyzed some material structural properties letting to achieve high performance from weak constituents in some of the most studied natural materials. Furthermore, we explored protein structures and properties, as in the following part of the thesis, we will focus on a protein material with extraordinary properties, namely spider silk.

Before focusing on the latter, we also reviewed other hierarchical biological materials of great scientific interest, such as bone, hair and cellulose. We then presented the case of spider silk, to give a preliminary idea of what spider silk is, how it is produced and why, and the reasons why there is great scientific interest in this material. We then also introduced the complexity of the material, also given the variability of spider silks that exist in the world, which has nevertheless led to the design of bioinspired materials with diverse properties and therefore diverse applications, a brief list of which we have reported.

Afterward, we introduced the importance of multiscale modelling for natural soft materials. We then reported on the fundamental steps of the scientific community in modelling the class of rubber-like materials, to which many biological soft materials, including spider silk, belong. Before getting into the heart of our modelling proposal, we set the objectives of the thesis.

The first significant result is the model proposed in Chapter 2 where we deduced a microstructure inspired model for humidity and temperature effects on

the mechanical response of spider silks, modeled as a composite material with a hard crystalline and a soft amorphous region.

We have considered that when humidified at different moisture conditions, unrestrained spider silk fibers can exhibit a very high supercontraction phenomenon due to hydration water molecules inducing a Hydrogen-bonds disruption process.

We then modelled the water molecule's effect in decreasing the percentage of crosslinks in the softer region and therefore inducing a variation of the natural length of the macromolecules due to entropic effects. The resulting kinematic incompatibility between the hard and soft regions crucially influences the final mechanical response, giving rise to the experimentally observed behaviour.

The described ability of the proposed model to predict the experimental results in terms of different mechanical properties make us confident that it well reproduces the humidity and temperature effects at the molecular scales. This is supported even more by predicting the behavior of different silks and environmental conditions. We strongly believe that this is a consequence of our microstructure deduction of the material response function.

The physical meaning of all the adopted microscopic parameters opens up the possibility of applying the proposed model not only to other protein materials with similar structures but also in the design of bioinspired materials employing chosen specific proteins.

Afterward, in Chapter 3, we have addressed the case of restrained spider silk fibers humidified at different moisture conditions. They can generate significant so-called supercontraction forces. By considering a bundle of macromolecules, we described supercontraction as a possible actuation system and determined the maximum actuation force depending on the silk properties at the molecular scale and on the constraining system representing other silk threads or the actuated device.

Again, the comparison with experimental results, in particular of *Argiope trifasciata* silk fibers, showed the effectiveness of the proposed model in quantitatively predicting the experimental actuation properties.

We considered the historical case study of obelisk rescue in Saint Peter's Square (Rome) through ropes hydration is discussed evidencing the optimal performances of this natural material adopted as humidity-powered actuator.

For such actuator, we obtained a work density of 2.19 kJ/m^3 making spider silk the most performant hydration-driven active material and a power density of the order of 730 W/kg, about three times the most performant carbon nanotube actuators making such material very competitive as compared with all types of actuator.

Also in this case, the analytic description of the macroscopic actuation parameters from microscale properties shows the possibility of adopting our approach also in the field of bioinspired artificial silks design, possibly considering also important non-linear effects in the actuated system.

Eventually, in Chapter 4, we have outlined a methodology for extracting information from a large amount of data, which may possibly be employed to enhance the models previously presented. We propose this approach based on the fact that today there is a huge experimental literature providing large data libraries on hierarchical systems, starting from the nano- and microscale up to the macroscale. The recent scientific literature is exploring the possibility of leveraging machine learning, with a predominant focus on neural networks, excelling in data fitting, but often lacking insight into essential physical information.

We showed the possibility of adopting, based on a Genetic Programming approach, data modelling techniques, innovative in thew field of material science, particularly suitable for the deduction of analytical models for multiscale problems. Our approach is based on the EPR method which, as we showed, lets us deduce models that are both accurate and simple, able to describe the dependence of macro scale variables from the ones at lower scales, in their hierarchical order itself deduced by a careful analysis of the data. The best-performing models are those located on the Pareto dominance front, which takes into account both accuracy and parsimony and is returned by the EPR algorithm. The final choice of the model can be then based on physical considerations.

To explicitly show the possibility of acquiring physical insight in a complex multiscale problem, and to evidence the key advantages of our multiscale approach compared to classical, non-physically based techniques, we referred to the case of spider silk, where its remarkable properties are strictly based on an evolutionary hierarchical optimization and the macroscopic spider silk behavior is the result of noteworthy mesostructures emerging from the aggregation of amino acids at the molecular scale.

By employing this paradigmatic example, we showed how the presented data modelling approach can be useful in several directions: determine dependent and independent variables, indicate their hierarchical organization, and deduce explicit relations among different groups of variables. In this direction, we also want to remark that a possible important role can be attained by a following dimensional analysis (e.g. Buckingham theorem) that should be related to the possible absence of variables at the different scales. This is another important aspect that is beyond the scope of this thesis and may be the subject of future investigation.

Furthermore, we showed that the proposed approach let us overcome the overfitting problem typically observed in the analysis of big data within the ANN framework diffusely adopted in this field.

Based on this, new physical knowledge is acquired, that can be used as a starting point for determining new analytical models, suggesting new experiments, and applying more focused data modelling analysis. We also strive to enrich existing physical approaches by enhancing our comprehension of the underlying physical processes. In this context, we investigate the potential for enhancing the previously introduced physically-based theoretical analyses, leveraging the insights obtained from the data modelling approach.

In this sense, we assert that Machine Learning or Artificial Intelligence can have a significant impact on scientific knowledge only if the data modelling approaches are in continuous synergy with the scientific interpretation of the results. We argue thus that a new mixed genetic programming-theoretical approach can be a fruitful approach in material science, but also in fields as diverse as biology and medicine.

Bibliography

Agnarsson, I., Boutry, C., Wong, S.-C., Baji, A., Dhinojwala, A., Sensenig, A. T., and Blackledge, T. A. (2009a). Supercontraction forces in spider dragline silk depend on hydration rate. *Zoology*, 112(5):325–331.

Agnarsson, I., Dhinojwala, A., Sahni, V., and Blackledge, T. A. (2009b). Spider silk as a novel high performance biomimetic muscle driven by humidity. *Journal of Experimental Biology*, 212(13):1990–1994.

Aigner, T. B., DeSimone, E., and Scheibel, T. (2018). Biomedical applications of recombinant silk-based materials. *Advanced materials*, 30(19):1704636.

Alber, M., Buganza Tepole, A., Cannon, W. R., De, S., Dura-Bernal, S., Garikipati, K., Karniadakis, G., Lytton, W. W., Perdikaris, P., Petzold, L., et al. (2019). Integrating machine learning and multiscale modeling—perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. *NPJ digital medicine*, 2(1):115.

Alemán, J., Chadwick, A. V., He, J., Hess, M., Horie, K., Jones, R. G., Kratochvíl, P., Meisel, I., Mita, I., Moad, G., et al. (2007). Definitions of terms relating to the structure and processing of sols, gels, networks, and inorganic-organic hybrid materials (iupac recommendations 2007). *Pure and Applied Chemistry*, 79(10):1801–1829.

Aliu, E., Kanungo, S., and Arnold, G. L. (2018). Amino acid disorders. *Annals of translational medicine*, 6(24).

Alon, U. (2007). Simplicity in biology. *Nature*, 446(7135):497–497.

Andres, C. M., Zhu, J., Shyu, T., Flynn, C., and Kotov, N. A. (2014). Shape-morphing nanocomposite origami. *Langmuir*, 30(19):5378–5385.

Arakawa, K., Kono, N., Malay, A. D., Tateishi, A., Ifuku, N., Masunaga, H., Sato, R., Tsuchiya, K., Ohtoshi, R., Pedrazzoli, D., et al. (2022). 1000 spider

silkomes: Linking sequences to silk physical properties. *Science Advances*, 8(41):eabo6043.

Arndt, T., Greco, G., Schmuck, B., Bunz, J., Shilkova, O., Francis, J., Pugno, N. M., Jaudzems, K., Barth, A., Johansson, J., et al. (2022). Engineered spider silk proteins for biomimetic spinning of fibers with toughness equal to dragline silks. *Advanced Functional Materials*, page 2200986.

Ashby, M. F. and Johnson, K. (2013). *Materials and design: the art and science of material selection in product design*. Butterworth-Heinemann.

Ashley, E. A. (2016). Towards precision medicine. *Nature Reviews Genetics*, 17(9):507–522.

Bae, W.-G., Kim, H. N., Kim, D., Park, S.-H., Jeong, H. E., and Suh, K.-Y. (2014). Scalable multiscale patterned structures inspired by nature: the role of hierarchy. *Advanced Materials*, 26(5):675–700.

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557):871–876.

Baldock, C., Oberhauser, A. F., Ma, L., Lammie, D., Siegler, V., Mithieux, S. M., Tu, Y., Chow, J. Y. H., Suleman, F., Malfois, M., et al. (2011). Shape of tropoelastin, the highly extensible protein that controls human tissue elasticity. *Proceedings of the National Academy of Sciences*, 108(11):4322–4327.

Baldwin, R. L. (2013). The new view of hydrophobic free energy. *FEBS letters*, 587(8):1062–1066.

Balf, M. R., Noori, R., Berndtsson, R., Ghaemi, A., and Ghiasi, B. (2018). Evolutionary polynomial regression approach to predict longitudinal dispersion coefficient in rivers. *Journal of Water Supply: Research and Technology—AQUA*, 67(5):447–457.

Banasiak, J. and Miekisz, J. (2008). *Multiscale problems in the life sciences: from microscopic to macroscopic*. Springer Science & Business Media.

Bechtle, S., Ang, S. F., and Schneider, G. A. (2010). On the mechanical properties of hierarchically structured biological materials. *Biomaterials*, 31(25):6378–6385.

Bell, F. I., McEwen, I. J., and Viney, C. (2002). Supercontraction stress in wet spider dragline. *Nature*, 416(6876):37–37.

Bellino, L., Florio, G., and Puglisi, G. (2019). The influence of device handles in single molecule experiments. *Soft Matter*, 15(43):8680–8690.

Berardi, L., Giustolisi, O., Kapelan, Z., and Savic, D. (2008). Development of pipe deterioration models for water distribution systems using epr. *Journal of Hydroinformatics*, 10(2):113–126.

Bini, E., Knight, D. P., and Kaplan, D. L. (2004). Mapping domain structures in silks from insects and spiders related to protein assembly. *Journal of molecular biology*, 335(1):27–40.

Blackledge, T. A. (2012). Spider silk: a brief review and prospectus on research linking biomechanics and ecology in draglines and orb webs. *The Journal of Arachnology*, 40(1):1–12.

Blackledge, T. A., Boutry, C., Wong, S.-C., Baji, A., Dhinojwala, A., Sahni, V., and Agnarsson, I. (2009). How super is supercontraction? persistent versus cyclic responses to humidity in spider dragline silk. *Journal of Experimental Biology*, 212(13):1981–1989.

Bosia, F., Abdalrahman, T., and Pugno, N. M. (2012). Investigating the role of hierarchy on the strength of composite materials: evidence of a crucial synergy between hierarchy and material mixing. *Nanoscale*, 4(4):1200–1207.

Boutry, C. and Blackledge, T. A. (2010). Evolution of supercontraction in spider silk: structure–function relationship from tarantulas to orb-weavers. *Journal of Experimental Biology*, 213(20):3505–3514.

Bowling, A. J., Amano, Y., Lindstrom, R., and Brown, R. M. (2001). Rotation of cellulose ribbons during degradation with fungal cellulase. *Cellulose*, 8(1):91–97.

Brown, C. P., MacLeod, J., Amenitsch, H., Cacho-Nerin, F., Gill, H. S., Price, A. J., Traversa, E., Licoccia, S., and Rosei, F. (2011). The critical role of water in spider silk and its consequence for protein mechanics. *Nanoscale*, 3(9):3805.

Bueche, F. (1960). Molecular basis for the mullins effect. *Journal of Applied Polymer Science*, 4(10):107–114.

Bustamante, C., Macosko, J. C., and Wuite, G. J. (2000). Grabbing the cat by the tail: manipulating molecules one by one. *Nature Reviews Molecular Cell Biology*, 1(2):130–136.

Bustamante, C., Marko, J. F., Siggia, E. D., and Smith, S. (1994). Entropic elasticity of λ-phage dna. *Science*, 265(5178):1599–1600.

Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O., and Walsh, A. (2018). Machine learning for molecular and materials science. *Nature*, 559(7715):547–555.

Canals, M., Veloso, C., and Solís, R. (2015). Adaptation of the spiders to the environment: the case of some chilean species. *Frontiers in Physiology*, 6:220.

Catalog, W. S. (2024). World spider catalog. version 24.5.

Cecen, A., Dai, H., Yabansu, Y. C., Kalidindi, S. R., and Song, L. (2018). Material structure-property linkages using three-dimensional convolutional neural networks. *Acta Materialia*, 146:76–84.

Chen, Q. and Pugno, N. M. (2013). Bio-mimetic mechanisms of natural hierarchical materials: A review. *Journal of the Mechanical Behavior of Biomedical Materials*, 19:3–33.

Chi, Y., Li, Y., Zhao, Y., Hong, Y., Tang, Y., and Yin, J. (2022). Bistable and multistable actuators for soft robots: Structures, materials, and functionalities. *Advanced Materials*, 34(19):2110384.

Chou, C.-C., Lepore, E., Antonaci, P., Pugno, N., and Buehler, M. J. (2015). Mechanics of trichocyte alpha-keratin fibers: Experiment, theory, and simulation. *Journal of Materials Research*, 30(1):26–35.

Cohen, N., Levin, M., and Eisenbach, C. D. (2021). On the origin of supercontraction in spider silk. *Biomacromolecules*, 22(2):993–1000.

Coleman, B. D. and Gurtin, M. (1967). Thermodynamics with internal state variables. *Journal of Chemical Physics*, 47:597–613.

Cook, D., Julias, M., and Nauman, E. (2014). Biological variability in biomechanical engineering research: Significance and meta-analysis of current modeling practices. *Journal of biomechanics*, 47(6):1241–1250.

Council, N. R. et al. (1994). *Hierarchical structures in biology as a guide for new materials technology*, volume 464. National Academies Press.

Cranford, S., Pugno, N., and Buehler, M. (2014). *Silk and Web Synergy: The Merging of Material and Structural Performance*, volume 5, chapter 12, pages 219–268.

Creaco, E., Berardi, L., Sun, S., Giustolisi, O., and Savic, D. (2016). Selection of relevant input variables in storm water quality modeling by multiobjective evolutionary polynomial regression paradigm. *Water Resources Research*, 52(4):2403–2419.

Dawson, C., Vincent, J. F., and Rocca, A.-M. (1997). How pine cones open. *Nature*, 390(6661):668–668.

De Pascalis, R. (2010). The semi-inverse method in solid mechanics: Theoretical underpinnings and novel applications.

De Tommasi, D., Millardi, N., Puglisi, G., and Saccomandi, G. (2013). An energetic model for macromolecules unfolding in stretching experiments. *Journal of the Royal Society, Interface / the Royal Society*, 10:20130651.

De Tommasi, D., Puglisi, G., and Saccomandi, G. (2010). Damage, self-healing, and hysteresis in spider silks. *Biophysical journal*, 98(9):1941–1948.

De Tommasi, D., Puglisi, G., and Saccomandi, G. (2015). Multiscale mechanics of macromolecular materials with unfolding domains. *Journal of the Mechanics and Physics of Solids*, 78:154–172.

De Tommasi, D., Puglisi, G., Toma, E., and Trentadue, F. (2019). A predictive model for the hysteretic and damage behavior of rubberlike materials. *Journal of Rheology*, 63(1):1–10.

Destrade, M., Saccomandi, G., and Sgura, I. (2017). Methodical fitting for mathematical models of rubber-like materials. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2198):20160811.

Di Donato, P., Poli, A., Taurisano, V., Nicolaus, B., et al. (2015). Polysaccharides from bioagro-waste for new biomolecules. In *Polysaccharides: bioactivity and biotechnology*, pages 603–637. Springer International Publishing.

Diani, J., Brieu, M., and Vacherand, J. (2006). A damage directional constitutive model for mullins effect with permanent set and induced anisotropy. *European Journal of Mechanics-A/Solids*, 25(3):483–496.

Diani, J., Fayolle, B., and Gilormini, P. (2009). A review on the mullins effect. *European Polymer Journal*, 45(3):601–612.

Dong, L., Qiao, J., Wu, Y., Ren, M., Wang, Y., Shen, X., Wei, X., Wang, X., Di, J., and Li, Q. (2021). Programmable contractile actuations of twisted spider dragline silk yarns. *ACS Biomaterials Science & Engineering*, 7(2):482–490.

Doshi-Velez, F. and Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.

Du, M., Liu, N., and Hu, X. (2019). Techniques for interpretable machine learning. *Communications of the ACM*, 63(1):68–77.

Du, N., Liu, X. Y., Narayanan, J., Li, L., Lim, M. L. M., and Li, D. (2006). Design of superior spider silk: from nanostructure to mechanical properties. *Biophysical journal*, 91(12):4528–4535.

Ducrot, E., Chen, Y., Bulters, M., Sijbesma, R. P., and Creton, C. (2014). Toughening elastomers with sacrificial bonds and watching them break. *Science*, 344(6180):186–189.

Edison, A. S. (2001). Linus pauling and the planar peptide bond. *Nature structural biology*, 8(3):201–202.

Eisoldt, L., Smith, A., and Scheibel, T. (2011). Decoding the secrets of spider silk. *Materials Today*, 14(3):80–86.

Eles, P. T. and Michal, C. A. (2004). Strain Dependent Local Phase Transitions Observed during Controlled Supercontraction Reveal Mechanisms in Spider Silk. *Macromolecules*, 37(4):1342–1345.

Elias, C. N., Meyers, M. A., Valiev, R. Z., and Monteiro, S. N. (2013). Ultrafine grained titanium for biomedical applications: An overview of performance. *Journal of Materials Research and Technology*, 2(4):340–350.

Elices, M. (2000). *Structural biological materials: design and structure-property relationships*. Elsevier.

Elices, M., Guinea, G. V., Pérez-Rigueiro, J., and Plaza, G. R. (2005). Finding inspiration in argiope trifasciata spider silk fibers. *Jom*, 57(2):60–66.

Elices, M., Plaza, G. R., Pérez-Rigueiro, J., and Guinea, G. V. (2011). The hidden link between supercontraction and mechanical behavior of spider silks. *Journal of the mechanical behavior of biomedical materials*, 4(5):658–669.

Ene, R., Papadopoulos, P., and Kremer, F. (2011). Supercontraction in nephila spider dragline silk–relaxation into equilibrium state. *Polymer*, 52(26):6056–6060.

Euler, M. (2008). Hooke's law and material science projects: exploring energy and entropy springs. *Physics Education*, 43(1):57.

Fantner, G. E., Oroudjev, E., Schitter, G., Golde, L. S., Thurner, P., Finch, M. M., Turner, P., Gutsmann, T., Morse, D. E., Hansma, H., et al. (2006). Sacrificial bonds and hidden length: unraveling molecular mesostructures in tough materials. *Biophysical journal*, 90(4):1411–1418.

Fazio, V., De Tommasi, D., Pugno, N. M., and Puglisi, G. (2022). Spider silks mechanics: Predicting humidity and temperature effects. *Journal of the Mechanics and Physics of Solids*, 164:104857.

Fazio, V., Pugno, N. M., Giustolisi, O., and Puglisi, G. (2024). Physically based machine learning for hierarchical materials. *Cell Reports Physical Science*, page 101790.

Fazio, V., Pugno, N. M., and Puglisi, G. (2023). "water to the ropes": A predictive model for the supercontraction stress of spider silks. *Extreme Mechanics Letters*, 61:102010.

Fish, J., Wagner, G. J., and Keten, S. (2021). Mesoscopic and multiscale modelling in materials. *Nature materials*, 20(6):774–786.

Florio, G. and Puglisi, G. (2019). Unveiling the influence of device stiffness in single macromolecule unfolding. *Scientific reports*, 9(1):4997.

Flory, P. J. and Erman, B. (1982). Theory of elasticity of polymer networks. 3. *Macromolecules*, 15(3):800–806.

Fratzl, P., Groschner, M., Vogl, G., Plenk Jr, H., Eschberger, J., Fratzl-Zelman, N., Koller, K., and Klaushofer, K. (1992). Mineral crystals in calcified tissues: a comparative study by saxs. *Journal of Bone and Mineral Research*, 7(3):329–334.

Fratzl, P. and Weinkamer, R. (2007). Nature's hierarchical materials. *Progress in materials Science*, 52(8):1263–1334.

Fu, C., Porter, D., and Shao, Z. (2009). Moisture effects on antheraea pernyi silk's mechanical property. *Macromolecules*, 42(20):7877–7880.

Fung, Y. (1967). Elasticity of soft tissues in simple elongation. *American Journal of Physiology-Legacy Content*, 213(6):1532–1544.

Gao, C., Peng, S., Feng, P., and Shuai, C. (2017). Bone biomaterials and interactions with stem cells. *Bone research*, 5(1):1–33.

Gao, H. and Yao, H. (2004). Shape insensitive optimal adhesion of nanoscale fibrillar structures. *Proceedings of the National Academy of Sciences*, 101(21):7851–7856.

Gardiner, B., Berry, P., and Moulia, B. (2016). Wind impacts on plant growth, mechanics and damage. *Plant science*, 245:94–118.

Gent, A. N. (1996). A new constitutive relation for rubber. *Rubber chemistry and technology*, 69(1):59–61.

Giesa, T., Perry, C. C., and Buehler, M. J. (2016). Secondary structure transition and critical stress for a model of spider silk assembly. *Biomacromolecules*, 17(2):427–436.

Giustolisi, O. (2004). Using genetic programming to determine chezy resistance coefficient in corrugated channels. *Journal of Hydroinformatics*, 6(3):157–173.

Giustolisi, O., Doglioni, A., Savic, D., and Laucelli, D. (2004). A proposal for an effective multi-objective non-dominated genetic algorithm: The optimised multi-objective genetic algorithm—optimoga. *OPTIMOGA, Report*, 7.

Giustolisi, O., Doglioni, A., Savic, D. A., and Webb, B. (2007). A multi-model approach to analysis of environmental phenomena. *Environmental Modelling & Software*, 22(5):674–682.

Giustolisi, O. and Laucelli, D. (2005). Improving generalization of artificial neural networks in rainfall-runoff modelling. *Hydrological Sciences Journal*, 50(3).

Giustolisi, O. and Savic, D. (2004). A novel genetic programming strategy: evolutionary polynomial regression. In *Hydroinformatics*, pages 787–794. World Scientific.

Giustolisi, O. and Savic, D. (2009). Advances in data-driven analyses and modelling using epr-moga. *Journal of Hydroinformatics*, 11(3-4):225–236.

Giustolisi, O. and Savic, D. A. (2006). A symbolic data-driven technique based on evolutionary polynomial regression. *Journal of Hydroinformatics*, 8(3):207–222.

Glišović, A. and Salditt, T. (2007). Temperature dependent structure of spider silk by X-ray diffraction. *Applied Physics A*, 87(1):63–69.

Göktepe, S. and Miehe, C. (2005). A micro–macro approach to rubber-like materials. part iii: The micro-sphere model of anisotropic mullins-type damage. *Journal of the Mechanics and Physics of Solids*, 53(10):2259–2283.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, Machine Learning*. Addison-Wesley.

Goriely, A. (2017). *The mathematics and mechanics of biological growth*. Springer.

Gosline, J., Guerette, P., Ortlepp, C., and Savage, K. (1999). The mechanical design of spider silks: from fibroin sequence to mechanical function. *Journal of Experimental Biology*, 202(23):3295–3303.

Gosline, J., Pollak, C., Guerette, P., Cheng, A., DeMont, M., and Denny, M. (1994). Elastomeric network models for the frame and viscid silks from the orb web of the spider araneus diadematus. In *ACS Symposium Series*, volume 544, pages 328–341. Washington, DC: American Chemical Society,[1974]-.

Gosline, J. M., DeMont, M. E., and Denny, M. W. (1986). The structure and properties of spider silk. *Endeavour*, 10(1):37–43.

Gosline, J. M., Denny, M. W., and DeMont, M. E. (1984). Spider silk as rubber. *Nature*, 309(5968):551–552.

Greco, G., Arndt, T., Schmuck, B., Francis, J., Bäcklund, F. G., Shilkova, O., Barth, A., Gonska, N., Seisenbaeva, G., Kessler, V., et al. (2021). Tyrosine residues mediate supercontraction in biomimetic spider silk. *Communications Materials*, 2(1):43.

Greco, G., Mirbaha, H., Schmuck, B., Rising, A., and Pugno, N. M. (2022). Artificial and natural silk materials have high mechanical property variability regardless of sample size. *Scientific Reports*, 12(1):3507.

Grubb, D. and Ji, G. (1999). Molecular chain orientation in supercontracted and re-extended spider silk. *International journal of biological macromolecules*, 24(2-3):203–210.

Gu, G. X., Chen, C.-T., Richmond, D. J., and Buehler, M. J. (2018). Bioinspired hierarchical composite design using machine learning: simulation, additive manufacturing, and experiment. *Materials Horizons*, 5(5):939–945.

Guinea, G. V., Cerdeira, M., Plaza, G. R., Elices, M., and Pérez-Rigueiro, J. (2010). Recovery in viscid line fibers. *Biomacromolecules*, 11(5):1174–1179.

Guinea, G. V., Elices, M., Pérez-Rigueiro, J., and Plaza, G. (2003). Self-tightening of spider silk fibers induced by moisture. *Polymer*, 44(19):5785–5788.

Guth, E. and Mark, H. (1934). Zur innermolekularen, statistik, insbesondere bei kettenmolekiilen i. *Monatshefte für Chemie und verwandte Teile anderer Wissenschaften*, 65:93–121.

Harland, D. P., Walls, R. J., Vernon, J. A., Dyer, J. M., Woods, J. L., and Bell, F. (2014). Three-dimensional architecture of macrofibrils in the human scalp hair cortex. *Journal of structural biology*, 185(3):397–404.

Harrington, M. J., Razghandi, K., Ditsch, F., Guiducci, L., Rueggeberg, M., Dunlop, J. W., Fratzl, P., Neinhuis, C., and Burgert, I. (2011). Origami-like unfolding of hydro-actuated ice plant seed capsules. *Nature communications*, 2(1):1–7.

Hayashi, C. Y., Shipley, N. H., and Lewis, R. V. (1999). Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins. *International journal of biological macromolecules*, 24(2-3):271–275.

Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*. International edition. Prentice Hall.

Herrmann, H. and Aebi, U. (2004). Intermediate filaments: molecular structure, assembly mechanism, and integration into functionally distinct intracellular scaffolds. *Annual review of biochemistry*, 73(1):749–789.

Hodge, A. J. (1963). Recent studies with the electron microscope on ordered aggregates of the tropocollagen molecules. *Aspects of protein structure*, pages 289–300.

Holland, C., Numata, K., Rnjak-Kovacina, J., and Seib, F. P. (2019). The biomedical use of silk: past, present, future. *Advanced healthcare materials*, 8(1):1800465.

Holland, J. H. (1992). *Adaptation in natural and artificial systems*. MIT press.

Horgan, C. and Saccomandi, G. (2001a). Large deformations of a rotating solid cylinder for non-gaussian isotropic, incompressible hyperelastic materials. *J. Appl. Mech.*, 68(1):115–117.

Horgan, C. and Saccomandi, G. (2004). Constitutive models for atactic elastomers. In *Waves And Stability In Continuous Media*, pages 281–294. World Scientific.

Horgan, C. O. and Saccomandi, G. (1999a). Pure axial shear of isotropic, incompressible nonlinearly elastic materials with limiting chain extensibility. *Journal of elasticity*, 57:307–319.

Horgan, C. O. and Saccomandi, G. (1999b). Simple torsion of isotropic, hyperelastic, incompressible materials with limiting chain extensibility. *Journal of elasticity*, 56:159–170.

Horgan, C. O. and Saccomandi, G. (2001b). Anti-plane shear deformations for non-gaussian isotropic, incompressible hyperelastic materials. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 457(2012):1999–2017.

Horgan, C. O. and Saccomandi, G. (2001c). Pure azimuthal shear of isotropic, incompressible hyperelastic materials with limiting chain extensibility. *International journal of non-linear mechanics*, 36(3):465–475.

Horgan, C. O. and Saccomandi, G. (2002). Constitutive modelling of rubber-like and biological materials with limiting chain extensibility. *Mathematics and mechanics of solids*, 7(4):353–371.

Horgan, C. O. and Saccomandi, G. (2006). Phenomenological hyperelastic strain-stiffening constitutive models for rubber. *Rubber chemistry and technology*, 79(1):152–169.

Horgan, C. O. and Smayda, M. G. (2012). The importance of the second strain invariant in the constitutive modeling of elastomers and soft biomaterials. *Mechanics of Materials*, 51:43–52.

Hsu, C. C., Buehler, M. J., and Tarakanova, A. (2020). The order-disorder continuum: linking predictions of protein structure and disorder through molecular simulation. *Scientific reports*, 10(1):1–14.

Huang, X., Liu, G., and Wang, X. (2012). New secrets of spider silk: exceptionally high thermal conductivity and its abnormal change under stretching. *Advanced Materials*, 24(11):1482–1486.

Hughes, M. L. and Dougan, L. (2016). The physics of pulling polyproteins: a review of single molecule force spectroscopy using the afm to study protein unfolding. *Reports on Progress in Physics*, 79:–076601.

Hümmerich, D., Slotta, U., and Scheibel, T. (2006). Processing and modification of films made from recombinant spider silk proteins. *Applied Physics A*, 82:219–222.

Jancar, J., Douglas, J., Starr, F. W., Kumar, S., Cassagnau, P., Lesser, A., Sternstein, S. S., and Buehler, M. (2010). Current issues in research on structure–property relationships in polymer nanocomposites. *Polymer*, 51(15):3321–3343.

Jenkins, J. E., Sampath, S., Butler, E., Kim, J., Henning, R. W., Holland, G. P., and Yarger, J. L. (2013). Characterizing the secondary protein structure of black widow dragline silk using solid-state NMR and X-ray diffraction. *Biomacromolecules*, 14(10):3472–3483.

Ji, B. and Gao, H. (2004). Mechanical properties of nanostructure of biological materials. *Journal of the Mechanics and Physics of Solids*, 52(9):1963–1990.

Ji, H., Daughton, W., Jara-Almonte, J., Le, A., Stanier, A., and Yoo, J. (2022). Magnetic reconnection in the era of exascale computing and multiscale experiments. *Nature Reviews Physics*, 4(4):263–282.

Johnson, K. A. and Goody, R. S. (2011). The original michaelis constant: translation of the 1913 michaelis–menten paper. *Biochemistry*, 50(39):8264–8269

Jones, D. and Treloar, L. (1975). The properties of rubber in pure homogeneous strain. *Journal of Physics D: Applied Physics*, 8(11):1285.

Karan, S. K., Maiti, S., Kwon, O., Paria, S., Maitra, A., Si, S. K., Kim, Y., Kim, J. K., and Khatua, B. B. (2018). Nature driven spider silk as high energy conversion efficient bio-piezoelectric nanogenerator. *Nano Energy*, 49:655–666.

Karplus, M. and McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nature structural biology*, 9(9):646–652.

Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R., Wyckoff, H., and Phillips, D. C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181(4610):662–666.

Keten, S. and Buehler, M. J. (2010). Nanostructure and molecular mechanics of spider dragline silk protein assemblies. *Journal of The Royal Society Interface*, 7(53):1709–1721.

Keten, S., Xu, Z., Ihle, B., and Buehler, M. J. (2010). Nanoconfinement controls stiffness, strength and mechanical toughness of β-sheet crystals in silk. *Nature materials*, 9(4):359–367.

Kim, S. H., Kwon, C. H., Park, K., Mun, T. J., Lepró, X., Baughman, R. H., Spinks, G. M., and Kim, S. J. (2016). Bio-inspired, moisture-powered hybrid carbon nanotube yarn muscles. *Scientific Reports*, 6. Cited by: 52; All Open Access, Gold Open Access, Green Open Access.

Kiseleva, A. P., Krivoshapkin, P. V., and Krivoshapkina, E. F. (2020). Recent advances in development of functional spider silk-based hybrid materials. *Frontiers in Chemistry*, 8:554.

Kluge, J. A., Rabotyagova, O., Leisk, G. G., and Kaplan, D. L. (2008). Spider silks and their applications. *Trends in biotechnology*, 26(5):244–251.

Koza, J. R. (1992). Genetic programming, on the programming of computers by means of natural selection. a bradford book. *MIT Press*.

Krässig, H. A. (1993). Cellulose : structure, accessibility, and reactivity. Gordon and Breach Science Publishers, Amsterdam.

Kuhn, W. (1934). Über die gestalt fadenförmiger moleküle in lösungen. *Kolloid-Zeitschrift*, 68:2–15.

Kundu, B., Kurland, N. E., Bano, S., Patra, C., Engel, F. B., Yadavalli, V. K., and Kundu, S. C. (2014). Silk proteins for biomedical applications: Bioengineering perspectives. *Progress in polymer science*, 39(2):251–267.

Lammel, A., Schwab, M., Hofer, M., Winter, G., and Scheibel, T. (2011). Recombinant spider silk particles as drug delivery vehicles. *Biomaterials*, 32(8):2233–2240.

Le Moigne, N., Bikard, J., and Navard, P. (2010). Rotation and contraction of native and regenerated cellulose fibers upon swelling and dissolution: the role of morphological and stress unbalances. *Cellulose*, 17(3):507–519.

Lenz, J., Schurz, J., and Wrentschur, E. (1993). Properties and structure of solvent-spun and viscose-type fibres in the swollen state. *Colloid and Polymer Science*, 271:460–468.

Li, S., McGhie, A., and Tang, S. (1994). New internal structure of spider dragline silk revealed by atomic force microscopy. *Biophysical journal*, 66(4):1209–1212.

Li, Z., Wang, J., Xu, Y., Shen, M., Duan, C., Dai, L., and Ni, Y. (2021). Green and sustainable cellulose-derived humidity sensors: A review. *Carbohydrate Polymers*, 270:118385.

Linka, K. and Kuhl, E. (2023). A new family of constitutive artificial neural networks towards automated model discovery. *Computer Methods in Applied Mechanics and Engineering*, 403:115731.

Liu, D., Tarakanova, A., Hsu, C. C., Yu, M., Zheng, S., Yu, L., Liu, J., He, Y., Dunstan, D. J., and Buehler, M. J. (2019). Spider dragline silk as torsional actuator driven by humidity. *Science Advances*, 5(3):eaau9183.

Liu, Y., Luo, D., and Wang, T. (2016). Hierarchical structures of bone and bioinspired bone tissue engineering. *Small*, 12(34):4611–4632.

Madden, J. D., Vandesteeg, N. A., Anquetil, P. A., Madden, P. G., Takshi, A., Pytel, R. Z., Lafontaine, S. R., Wieringa, P. A., and Hunter, I. W. (2004). Artificial muscle technology: physical principles and naval prospects. *IEEE Journal of oceanic engineering*, 29(3):706–728.

Madsen, B., Shao, Z. Z., and Vollrath, F. (1999). Variability in the mechanical properties of spider silks on three levels: interspecific, intraspecific and intraindividual. *International journal of biological macromolecules*, 24(2-3):301–306.

Makarov, D. E. (2009). A theoretical model for the mechanical unfolding of repeat proteins. *Biophysical journal*, 96(6):2160–2167.

Manca, F., Giordano, S., Palla, P. L., Cleri, F., and Colombo, L. (2013). Two-state theory of single-molecule stretching experiments. *Physical Review E*, 88(3):032705.

Marin-Gonzalez, A., Vilhena, J. G., Perez, R., and Moreno-Herrero, F. (2017). Understanding the mechanical response of double-stranded dna and rna un-

der constant stretching forces using all-atom molecular dynamics. *Proceeding of the National Academy of Science (PNAS)*, 114:7049–7054.

Marko, J. F. and Siggia, E. D. (1995). Stretching dna. *Macromolecules*, 28(26):8759–8770.

Martin, S. E. and Carlson, D. E. (1977). The behavior of elastic heat conductors with second-order response functions. *Zeitschrift für angewandte Mathematik und Physik ZAMP*, 28:311–329.

Matsunaga, R., Abe, R., Ishii, D., Watanabe, S.-i., Kiyoshi, M., Nöcker, B., Tsuchiya, M., and Tsumoto, K. (2013). Bidirectional binding property of high glycine–tyrosine keratin-associated protein contributes to the mechanical strength and shape of hair. *Journal of structural biology*, 183(3):484–494.

McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5:115–133.

McKittrick, J., Chen, P.-Y., Tombolato, L., Novitskaya, E., Trim, M., Hirata, G., Olevsky, E., Horstemeyer, M., and Meyers, M. (2010). Energy absorbent natural materials and bioinspired design strategies: a review. *Materials Science and Engineering: C*, 30(3):331–342.

McLennan, R., Dyson, L., Prather, K. W., Morrison, J. A., Baker, R. E., Maini, P. K., and Kulesa, P. M. (2012). Multiscale mechanisms of cell migration during development: theory and experiment. *Development*, 139(16):2935–2944.

Mehrabian, M. A. and Hashemian, M. (2003). Measuring the relative humidity of humid air using a hair hygrometer. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, 217(5):379–385.

Meyer, K. H. and Ferri, C. (1935). Sur l'élasticité du caoutchouc. *Helvetica Chimica Acta*, 18(1):570–589.

Meyers, M. A., Chen, P.-Y., Lin, A. Y.-M., and Seki, Y. (2008). Biological materials: structure and mechanical properties. *Progress in materials science*, 53(1):1–206.

Mielke, A. (2006). *Analysis, Modeling and Simulation of Multiscale Problems*.

Molnar, C. (2020). *Interpretable machine learning*. Lulu. com.

Mooney, M. (1940). A theory of large elastic deformation. *Journal of applied physics*, 11(9):582–592.

Mullins, L. (1948). Effect of stretching on the properties of rubber. *Rubber chemistry and technology*, 21(2):281–300.

Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., and Yu, B. (2019). Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences*, 116(44):22071–22080.

Murnaghan, F. D. (1937). Finite deformations of an elastic solid. *American Journal of Mathematics*, 59(2):235–260.

Naraghi, M., Filleter, T., Moravsky, A., Locascio, M., Loutfy, R. O., and Espinosa, H. D. (2010). A multiscale study of high performance double-walled nanotube- polymer fibers. *ACS nano*, 4(11):6463–6476.

Nelson, D. L., Lehninger, A. L., and Cox, M. M. (2008). *Lehninger principles of biochemistry*. Macmillan.

Nepal, D., Kang, S., Adstedt, K. M., Kanhaiya, K., Bockstaller, M. R., Brinson, L. C., Buehler, M. J., Coveney, P. V., Dayal, K., El-Awady, J. A., et al. (2023). Hierarchically structured bioinspired nanocomposites. *Nature materials*, 22(1):18–35.

Nguyen, L. H., Annabi, N., Nikkhah, M., Bae, H., Binan, L., Park, S., Kang, Y., Yang, Y., and Khademhosseini, A. (2012). Vascularized bone tissue engineering: approaches for potential improvement. *Tissue Engineering Part B: Reviews*, 18(5):363–382.

Nova, A., Keten, S., Pugno, N. M., Redaelli, A., and Buehler, M. J. (2010). Molecular and Nanostructural Mechanisms of Deformation, Strength and Toughness of Spider Silk Fibrils. *Nano Letters*, 10(7):2626–2634.

Obata, Y., Kawabata, S., and Kawai, H. (1970). Mechanical properties of natural rubber vulcanizates in finite deformation. *Journal of Polymer Science Part A-2: Polymer Physics*, 8(6):903–919.

Ogden, R. (1997). *Non-linear Elastic Deformations*. Dover Civil and Mechanical Engineering. Dover Publications.

Ogden, R. W. (1972). Large deformation isotropic elasticity–on the correlation of theory and experiment for incompressible rubberlike solids. *Proceedings*

of the Royal Society of London. A. Mathematical and Physical Sciences, 326(1567):565–584.

Ogden, R. W., Saccomandi, G., and Sgura, I. (2004). Fitting hyperelastic models to experimental data. *Computational Mechanics*, 34:484–502.

Pareto, V. (1896). *Cours d'Economie politique*, volume 1. F. Rouge.

Pareto, V. (1906). *Manual of Political Economy*. Oxford University Press.

Pauling, L., Corey, R. B., and Branson, H. T. (1951). The structure of proteins: two hydrogen-bonded helical configuration of the polypeptide chain. *Proceedings of the National Academy of Science PNAS*, 37(4):205–211.

Pérez-Rigueiro, J., Elices, M., Plaza, G. R., and Guinea, G. V. (2021). Basic principles in the design of spider silk fibers. *Molecules*, 26(6):1794.

Plaza, G. R., Guinea, G. V., Pérez-Rigueiro, J., and Elices, M. (2006). Thermo-hygro-mechanical behavior of spider dragline silk: Glassy and rubbery states. *Journal of Polymer Science Part B: Polymer Physics*, 44(6):994–999.

Porter, D., Guan, J., and Vollrath, F. (2013). Spider silk: super material or thin fibre? *Advanced Materials*, 25(9):1275–1279.

Pucci, E. and Saccomandi, G. (2002). A note on the gent model for rubber-like materials. *Rubber chemistry and technology*, 75(5):839–852.

Puglisi, G., De Tommasi, D., Pantano, M. F., Pugno, N. M., and Saccomandi, G. (2017). Micromechanical model for protein materials: From macromolecules to macroscopic fibers. *Physical Review E*, 96(4):042407.

Puglisi, G. and Saccomandi, G. (2016). Multi-scale modelling of rubber-like materials and soft tissues: an appraisal. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 472(2187):20160060.

Puglisi, G. and Truskinovsky, L. (2000). Mechanics of a discrete chain with bi-stable elements. *Journal of the Mechanics and Physics of Solids*, 48(1):1–27.

Pérez-Rigueiro, J., Elices, M., Plaza, G. R., and Guinea, G. V. (2021). Basic principles in the design of spider silk fibers. *Molecules*, 26(6):1794.

Rajagopal, K. R. and Srinivasa, A. R. (2004). On the thermomechanics of materials that have multiple natural configurations part i: Viscoelasticity and classical plasticity. *Zeitschrift für angewandte Mathematik und Physik ZAMP*, 55:861–893.

Rajagopal, K. R. and Wineman, A. S. (1992). A constitutive equation for nonlinear solids which undergo deformation induced microstructural changes. *International Journal of Plasticity*, 8(4):385–395.

Ramawat, K. and Ahuja, M. (2016). Fiber plants: an overview. *Fiber Plants*, pages 3–15.

Rammensee, S., Slotta, U., Scheibel, T., and Bausch, A. R. (2008). Assembly mechanism of recombinant spider silk proteins. *Proceedings of the National Academy of Sciences*, 105(18):6590–6595.

Rebenfeld, L., Weigmann, H. D., and Dansizer, C. (1966). Temperature dependence of the mechanical properties of human hair in relation to structure. *J. Soc. Cosmet. Chem*, 17:525–538.

Rechenberg, I. (1971). Evolutionsstrategie. *Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*, (PhD thesis). Reprinted by Fromman-Holzboog (1973).

Rees, D. C., Williams, T. N., and Gladwin, M. T. (2010). Sickle-cell disease. *The Lancet*, 376(9757):2018–2031.

Regazzoni, F., Dedè, L., and Quarteroni, A. (2020). Machine learning of multiscale active force generation models for the efficient simulation of cardiac electromechanics. *Computer Methods in Applied Mechanics and Engineering*, 370:113268.

Reznikov, N., Shahar, R., and Weiner, S. (2014). Bone hierarchical structure in three dimensions. *Acta biomaterialia*, 10(9):3815–3826.

Rho, J.-Y., Kuhn-Spearing, L., and Zioupos, P. (1998). Mechanical properties and the hierarchical structure of bone. *Medical engineering & physics*, 20(2):92–102.

Ridge, B. (1953). Synthetic polymer fibres. *Journal of the Textile Institute Proceedings*, 44(2):P48–P65.

Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M., and Gaub, H. E. (1997). Reversible unfolding of individual titin immunoglobulin domains by afm. *science*, 276(5315):1109–1112.

Rivlin, R. S. (1947). Torsion of a rubber cylinder. *Journal of Applied Physics*, 18(5):444–449.

Rivlin, R. S. (1960). Some topics in finite elasticity. *Proceedings of the First Naval Symposyum on Structural Mechanics*.

Rivlin, R. S. and Saunders, D. (1951). Large elastic deformations of isotropic materials vii. experiments on the deformation of rubber. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 243(865):251–288.

Robbins, C. R. and Robbins, C. R. (2012). *Chemical and physical behavior of human hair*, volume 4. Springer.

Roberts, K., Alberts, B., Johnson, A., Walter, P., and Hunt, T. (2002). Molecular biology of the cell. *New York: Garland Science*, 32(2).

Robinson, R. A. (1952). An electron-microscopic study of the crystalline inorganic component of bone and its relationship to the organic matrix. *JBJS*, 34(2):389–476.

Römer, L. and Scheibel, T. (2008). The elaborate structure of spider silk: structure and function of a natural high performance fiber. *Prion*, 2(4):154–161.

Rubinstein, M. and Colby, R. (2003). *Polymer Physics*. Oxford University Press.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088):533–536.

Saccomandi, G. and Ogden, R. W. (2004). *Mechanics and thermomechanics of rubberlike solids*, volume 452. Springer.

Sanchez, C., Arribart, H., and Giraud Guille, M. M. (2005). Biomimetism and bioinspiration as tools for the design of innovative materials and systems. *Nature materials*, 4(4):277–288.

Savage, K. N., Guerette, P. A., and Gosline, J. M. (2004). Supercontraction stress in spider webs. *Biomacromolecules*, 5(3):675–679.

Schacht, K. and Scheibel, T. (2011). Controlled hydrogel formation of a recombinant spider silk protein. *Biomacromolecules*, 12(7):2488–2495.

Schacht, K., Vogt, J., and Scheibel, T. (2016). Foams made of engineered recombinant spider silk proteins as 3d scaffolds for cell growth. *ACS biomaterials science & engineering*, 2(4):517–525.

Signorini, A. (1943). Trasformazioni termoelastiche finite. *Annali di Matematica pura ed applicata*, 22:33–143.

Singh, M. (1967). Small finite deformations of elastic dielectrics. *Quarterly of Applied Mathematics*, 25(3):275–284.

Smith, B. L., Schäffer, T. E., Viani, M., Thompson, J. B., Frederick, N. A., Kindt, J., Belcher, A., Stucky, G. D., Morse, D. E., and Hansma, P. K. (1999). Molecular mechanistic origin of the toughness of natural adhesives, fibres and composites. *Nature*, 399(6738):761–763.

Sponner, A., Vater, W., Monajembashi, S., Unger, E., Grosse, F., and Weisshart, K. (2007). Composition and Hierarchical Organisation of a Spider Silk. *PLoS ONE*, 2(10):e998.

Stam, P. B., Kratz, R. F., and White JR, H. J. (1952). The swelling of human hair in water and water vapor. *Textile research journal*, 22(7):448–465.

Stauffer, S. L., Coguill, S. L., and Lewis, R. V. (1994). Comparison of physical properties of three silks from nephila clavipes and araneus gemmoides. *The Journal of Arachnology*, 22(1):5–11.

Tadmor, E. B., Ortiz, M., and Phillips, R. (1996). Quasicontinuum analysis of defects in solids. *Philosophical magazine A*, 73(6):1529–1563.

Tian, P., Gao, X., Wen, G., Zhong, L., Wang, Z., and Guo, Z. (2018). Novel fabrication of polymer/carbon nanotube composite coated janus paper for humidity stress sensor. *Journal of colloid and interface science*, 532:517–526.

Treloar, L. G. (1975). The physics of rubber elasticity.

Trentadue, F., De Tommasi, D., and Puglisi, G. (2021). A predictive micromechanically-based model for damage and permanent deformations in copolymer sutures. *Journal of the Mechanical Behavior of Biomedical Materials*, 115:104277.

Trowbridge, C. C. (1896). The use of the hair hygrometer. *Science*, 4(81):62–65.

Valanis, K. and Landel, R. F. (1967). The strain-energy function of a hyperelastic material in terms of the extension ratios. *Journal of Applied Physics*, 38(7):2997–3002.

Vehoff, T., Glišović, A., Schollmeyer, H., Zippelius, A., and Salditt, T. (2007). Mechanical Properties of Spider Dragline Silk: Humidity, Hysteresis, and Relaxation. *Biophysical Journal*, 93(12):4425–4432.

Voet, D. and Voet, J. G. (2010). *Biochemistry*. John Wiley & Sons.

Vollrath, F. (1992). Spider webs and silks. *Scientific American*, 266(3):70–77.

Vollrath, F. and Porter, D. (2006). Spider silk as a model biomaterial. *Applied Physics A*, 82(2):205–212.

Weinan, E. (2011). *Principles of multiscale modeling*. Cambridge University Press.

Weiner, S. and Price, P. A. (1986). Disaggregation of bone into crystals. *Calcified tissue international*, 39:365–375.

Weiner, S. and Wagner, H. D. (1998). The material bone: structure-mechanical function relations. *Annual review of materials science*, 28(1):271–298.

Whipple, F. J. W. (1921). The theory of the hair hygrometer. *Proceedings of the Physical Society of London*, 34(1):i.

Work, R. W. (1977). Dimensions, birefringences, and force-elongation behavior of major and minor ampullate silk fibers from orb-web-spinning spiders—the effects of wetting on these properties. *Textile Research Journal*, 47(10):650–662.

Yarger, J. L., Cherry, B. R., and Van Der Vaart, A. (2018). Uncovering the structure–function relationship in spider silk. *Nature Reviews Materials*, 3(3):1–11.

Yazawa, K., Malay, A. D., Masunaga, H., Norma-Rashid, Y., and Numata, K. (2020). Simultaneous effect of strain rate and humidity on the structure and mechanical behavior of spider silk. *Communications Materials*, 1(1):10.

Yazawa, K., Malay, A. D., Masunaga, H., and Numata, K. (2019). Role of Skin Layers on Mechanical Properties and Supercontraction of Spider Dragline Silk Fiber. *Macromolecular Bioscience*, 19(3):1800220.

Yeo, G. C., Tarakanova, A., Baldock, C., Wise, S. G., Buehler, M. J., and Weiss, A. S. (2016). Subtle balance of tropoelastin molecular shape and flexibility regulates dynamics and hierarchical assembly. *Science advances*, 2(2):e1501145.

Yu, Y., Yang, W., Wang, B., and Meyers, M. A. (2017). Structure and mechanical behavior of human hair. *Materials Science and Engineering: C*, 73:152–163.

Zhang, M., Liu, Z., Zhang, Y., Zhang, Y., Yang, X., Zhang, J., Yang, J., and Yuan, L. (2022). Spider dragline silk-based fp humidity sensor with ultra-high sensitivity. *Sensors and Actuators B: Chemical*, 350:130895.

Zhang, P., Su, J., and Mende, U. (2012). Cross talk between cardiac myocytes and fibroblasts: from multiscale investigative approaches to mechanisms and functional consequences. *American Journal of Physiology-Heart and Circulatory Physiology*, 303(12):H1385–H1396.

Zhao, N., Wang, Z., Cai, C., Shen, H., Liang, F., Wang, D., Wang, C., Zhu, T., Guo, J., Wang, Y., et al. (2014). Bioinspired materials: from low to high dimensional structure. *Advanced Materials*, 26(41):6994–7017.

Zhou, X., Guo, B., Zhang, L., and Hu, G.-H. (2017). Progress in bio-inspired sacrificial bonds in artificial polymeric materials. *Chemical Society Reviews*, 46(20):6301–6329.

Appendix A

EPR Expressions

A.1 Cristallinity from micro properties

$$A = 0.19253 \tag{S1.1}$$

$$A = 3.5562\frac{1}{b} + 0.10262 \tag{S1.2}$$

$$A = 3.9403\frac{1}{b} + 0.0097339c + 0.013178 \tag{S1.3}$$

$$A = 0.0010868b + 2.0441\frac{c^{0.5}}{b} \tag{S1.4}$$

$$A = 0.007611\frac{b}{a^{0.5}} + 1.961\frac{c^{0.5}}{b} \tag{S1.5}$$

$$A = 0.0077062\frac{b}{a^{0.5}} + 2.4415\frac{1}{b} + 0.38037\frac{c}{b} \tag{S1.6}$$

$$A = 0.0079125\frac{b}{a^{0.5}} + 10.8129\frac{1}{cb} + 0.50353\frac{c}{b} \tag{S1.7}$$

A.2 Birefringence from micro properties

$$B = 45.0698 \tag{S2.1}$$

$$B = 88.7638\frac{1}{b^{0.5}} + 31.2338 \tag{S2.2}$$

$$B = 188.8929\frac{1}{a} + 54.9397\frac{1}{b^{0.5}} + 31.4271 \tag{S2.3}$$

$$B = 571.6574\frac{1}{a} + 1.7581\frac{a}{b^{0.5}} + 19.1995 \tag{S2.4}$$

$$B = 120.3924\frac{b^{0.5}}{a} + 18.2244\frac{a}{b} + 7.3676 \tag{S2.5}$$

$$B = 135.0812\frac{b^{0.5}}{a} + 19.3865\frac{a}{b} + 0.036473a + 2.329 \tag{S2.6}$$

$$B = 141.865\frac{b^{0.5}}{a} + 17.6973\frac{a}{b} + 0.69564\frac{a}{b^{0.5}} \tag{S2.7}$$

A.3 Thermal degradation temperature from micro properties

$$C = 226.8169 \tag{S3.1}$$

$$C = 3.1984c + 201.1538 \tag{S3.2}$$

$$C = 785.1137\frac{1}{c} + 13.9926c + 13.4973 \tag{S3.3}$$

$$C = 3186.7046\frac{1}{a} + 0.86787ac^{0.5} + 45.2672 \tag{S3.4}$$

$$C = 2482.7659\frac{1}{a} + 383.6809\frac{1}{c} + 0.27232ac + 25.2964 \tag{S3.5}$$

A.4 Young's Modulus from meso properties

$$\mathbb{A} = 9.4674 \tag{S4.1}$$

$$\mathbb{A} = 0.86198B^{0.5} + 3.7949 \tag{S4.2}$$

$$\mathbb{A} = 0.60871\frac{1}{A} + 29.105A \tag{S4.3}$$

$$\mathbb{A} = 0.091301\frac{B^{0.5}}{A} + 29.2668A \tag{S4.4}$$

$$\mathbb{A} = 1.3608\frac{B^{0.5}}{AC^{0.5}} + 29.2732A \tag{S4.5}$$

$$\mathbb{A} = 20.1089\frac{B^{0.5}}{AC} + 1.9687AC^{0.5} \tag{S4.6}$$

$$\mathbb{A} = 2.3655\frac{B}{AC} + 0.098189\frac{1}{A} + 2.0067AC^{0.5} \tag{S4.7}$$

$$\mathbb{A} = 27.51\frac{1}{AC} + 2.2479\frac{B}{AC} + 2.0041AC^{0.5} \tag{S4.8}$$

$$\mathbb{A} = 10.3522\frac{B^{0.5}}{AC} + 1.3585\frac{B}{AC} + 1.9953AC^{0.5} \tag{S4.9}$$

A.5 Limit Stress from meso properties

$$\mathbb{B} = 1 \tag{S5.1}$$

$$\mathbb{B} = 0.14344B^{0.5} + 0.2349 \tag{S5.2}$$

$$\mathbb{B} = 0.09374\frac{1}{A} + 3.1305A \tag{S5.3}$$

$$\mathbb{B} = 0.013904\frac{B^{0.5}}{A} + 3.1765A \tag{S5.4}$$

$$\mathbb{B} = 0.013837\frac{B^{0.5}}{A} + 0.014276AC \tag{S5.5}$$

$$\mathbb{B} = 0.011517\frac{B^{0.5}}{A} + 0.0017796ACB^{0.5} + 0.20612 \tag{S5.6}$$

$$\mathbb{B} = 0.021735\frac{1}{A} + 0.0090613\frac{B^{0.5}}{A} + 0.001894ACB^{0.5} + 0.14047 \tag{S5.7}$$

$$\mathbb{B} = 0.012666\frac{B^{0.5}}{A} + 0.0048689AC + 0.001261ACB^{0.5} + 0.09517 \tag{S5.8}$$

A.6 Diameter from meso properties

$$\mathbb{C} = 2 \tag{S6.1}$$

$$\mathbb{C} = 81.9474\frac{1}{B} \tag{S6.2}$$

$$\mathbb{C} = 177.4203\frac{A^{0.5}}{B} + 0.051737 \tag{S6.3}$$

$$\mathbb{C} = 0.81928\frac{A^{0.5}C}{B} \tag{S6.4}$$

$$\mathbb{C} = 0.00021001C + 0.80165\frac{A^{0.5}C}{B} \tag{S6.5}$$

$$\mathbb{C} = 0.0037544\frac{C}{B^{0.5}} + 0.76892\frac{A^{0.5}C}{B} \tag{S6.6}$$

$$\mathbb{C} = 0.011782\frac{C}{B} + 0.0025068\frac{C}{B^{0.5}} + 0.76003\frac{A^{0.5}C}{B} \tag{S6.7}$$

$$\mathbb{C} = 0.028813\frac{C}{B} + 0.69464\frac{A^{0.5}C}{B} + 0.010027\frac{A^{0.5}C}{B^{0.5}} \tag{S6.8}$$

A.7 Supercontraction from meso properties

$$\mathbb{D} = 0.31695 \tag{S7.1}$$

$$\mathbb{D} = 1.4429\frac{1}{B^{0.5}} + 0.093513 \tag{S7.2}$$

$$\mathbb{D} = 11.8488\frac{A^{0.5}}{B} + 0.18064 \tag{S7.3}$$

$$\mathbb{D} = 0.00085048C + 10.8458\frac{A^{0.5}}{B} \tag{S7.4}$$

$$\mathbb{D} = 0.00086571C + 158.9386\frac{A^{0.5}}{C^{0.5}B} \tag{S7.5}$$

$$\mathbb{D} = 0.00050622C + 3.8693e-05CB^{0.5} + 12.6449\frac{A^{0.5}}{B} \tag{S7.6}$$

$$\mathbb{D} = 0.00067818C + 2.1548e-05CB^{0.5} + 172.7934\frac{A^{0.5}}{C^{0.5}B} \tag{S7.7}$$

A.8 Young's Modulus from micro properties

$$\mathbb{A} = 8.5718 \tag{S8.1}$$

$$\mathbb{A} = +150.4503\frac{1}{b} + 4.8595 \tag{S8.2}$$

$$\mathbb{A} = +103.8759\frac{1}{a} + 93.8966\frac{1}{b} + 3.4673 \tag{S8.3}$$

$$\mathbb{A} = +18.4325\frac{c}{a} + 36.55\frac{1}{c} \tag{S8.4}$$

$$\mathbb{A} = +111.3405\frac{c}{ab^{0.5}} + 37.8293\frac{1}{c} \tag{S8.5}$$

$$\mathbb{A} = +712.237\frac{c}{ab} + 5.6688\frac{b^{0.5}}{c} \tag{S8.6}$$

$$\mathbb{A} = +120.4316\frac{c}{ab^{0.5}} + 21.7225\frac{1}{c} + 2.2107\frac{a^{0.5}}{c} \tag{S8.7}$$

$$\mathbb{A} = +121.6927\frac{c}{ab^{0.5}} + 0.40972\frac{b}{c} + 114.29\frac{a^{0.5}}{cb} \tag{S8.8}$$

A.9 Limit Stress from micro properties

$$\mathbb{B} = 1.2533 \tag{S9.1}$$

$$\mathbb{B} = +17.8802\frac{1}{a} + 0.77347 \tag{S9.2}$$

$$\mathbb{B} = +5.1279\frac{1}{a^{0.5}} + 3.14\frac{1}{c} + 0.025086 \tag{S9.3}$$

$$\mathbb{B} = +146.7895\frac{1}{a^{0.5}b} + 0.014999b + 0 \tag{S9.4}$$

$$\mathbb{B} = +0.26238\frac{b}{a^{0.5}c^{0.5}} + 24.7876\frac{1}{b} + 0 \tag{S9.5}$$

$$\mathbb{B} = +0.29999\frac{b}{a} + 0.056394\frac{b}{c} + 24.7595\frac{1}{b} + 0 \tag{S9.6}$$

$$\mathbb{B} = +132.185\frac{1}{a^{0.5}b} + 0.052774\frac{b}{a^{0.5}} + 0.056459\frac{b}{c} + 0.037522 \tag{S9.7}$$

$$\mathbb{B} = +0.44217\frac{b}{a^{0.5}c} + 49.6034\frac{c^{0.5}}{a^{0.5}b} + 0.054095\frac{b}{c} \tag{S9.8}$$

A.10 Diameter from micro properties

$$\mathbb{C} = 1.5415 \tag{S10.1}$$

$$\mathbb{C} = +44.5973\frac{1}{b} + 0.43669 \tag{S10.2}$$

$$\mathbb{C} = +1030.4476\frac{1}{ab} + 0.82562 \tag{S10.3}$$

$$\mathbb{C} = +1367.643\frac{1}{ab} + 0.013763b \tag{S10.4}$$

$$\mathbb{C} = +8702.7135\frac{1}{acb} + 0.01363b + 0.17322 \tag{S10.5}$$

$$\mathbb{C} = +9507.752\frac{1}{acb} + 0.0019587cb \tag{S10.6}$$

$$\mathbb{C} = +9491.6515\frac{1}{acb} + 0.001894cb + 7.7482e-05ac \tag{S10.7}$$

$$\mathbb{C} = +9609.6043\frac{1}{acb} + 0.0014633cb + 7.4705e-05a^{0.5}cb \tag{S10.8}$$

A.11 Supercontraction from micro properties

$$\mathbb{D} = +0.32479 \tag{S11.1}$$

$$\mathbb{D} = +0.0055013b + 0.097111 \tag{S11.2}$$

$$\mathbb{D} = +0.061926\frac{b}{c} + 0.0047393 \tag{S11.3}$$

$$\mathbb{D} = +0.53648\frac{1}{c} + 0.050578\frac{b}{c} \tag{S11.4}$$

$$\mathbb{D} = +0.8009\frac{1}{c} + 0.0072816\frac{a^{0.5}b}{c} \tag{S11.5}$$

$$\mathbb{D} = +15.9524\frac{1}{cb} + 0.008755\frac{a^{0.5}b}{c} \tag{S11.6}$$

$$\mathbb{D} = +430.4092\frac{1}{acb} + 0.0090635\frac{a^{0.5}b}{c} \tag{S11.7}$$

Scientists are continuously fascinated by the high degree of sophistication found in natural materials, arising from evolutionary optimisation. One of the most studied is the spider silk, a semi-crystalline polymer renowned for its outstanding mechanical properties, which are the result of an ingenious hierarchical structure. Nevertheless, many of the multiscale mechanisms that give rise to the macroscopic properties are still unclear, especially in relation to temperature and humidity, which strongly condition the mechanical performance. In this thesis, aimed at unveiling some of these open problems, we introduce a multiscale model for the thermo-hygro-mechanical response, starting from describing the microstructure, up to the effects at the macroscopic scale, including softening and supercontraction in response to hydration. Thereafter, we describe how spider silks can be adopted to obtain humidity-driven actuators. We determine the maximum actuation force depending on the silk properties at the molecular scale, and on the constraining system representing other silk threads or the actuated device. To address a novel database of recently published multiscale experimental data of about a thousand different spider silks, we propose a data modelling methodology based on symbolic regression techniques to obtain simple and interpretable relationships linking quantities at different scales. This approach may provide a proof of concept for modelling in fields governed by multiscale, hierarchical differential equations. We believe that the analytical description of the macroscopic behaviour from microscale properties is of great value both for a full understanding of biological materials and for the perspective of designing bioinspired materials and structures.

Vincenzo Fazio obtained his bachelor in Mechanical Engineering at Politecnico di Bari, with an experimental thesis on viscoelasticity in polymers. He attended the MSc Mechanics of Materials for Engineering and Integrity of Structures at the Ecole Normale Supérieure Paris-Saclay, with an internship at the ESPCI University, sponsored by Hutchinson, on unsteady-state friction, which led to a publication in an international journal. For his master's thesis in Mechanical Engineering at Politecnico di Bari, he explored the fatigue behaviour of materials. During his PhD, he has authored three articles published in international journals. He has presented his research at several conferences, also as an invited speaker. He has been awarded a GNFM Young Researchers Project. His research activity delves into theoretical exploration and machine learning techniques for the multiscale modelling of the behaviour of rubber-like materials, with focus on spider silks.