Structural Analysis of Silk Proteins using X–ray and Neutron Scattering

A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy

submitted by
Imke Greving
(Nee: Diddens)

Oxford Silk Group
Department of Zoology
University of Oxford
Jesus College

Michaelmas 2012

Contents

List of Figures

List of Tables

List of Abbreviations

AA	Amino acid
AFM	Atomic Force Microscopy
b	Scattering length
c_s	Spring constant
CD	Circular Dichroism
$\Delta\Omega$	Solid angle element of the detector
$\Delta\rho$	Scattering density difference
η	Detector efficiency
d	Distance
D_{max}	Maximum dimension of a particle
$\frac{d\sigma}{d\Omega}$	Differential scattering cross–section
ϵ	Depth of well in Lennard–Jones potential
Φ	Flux
FFT	Fourier transformation
$F(\vec{q})$	Form factor
$\gamma(\vec{r})$	Auto correlation function
I	Intensity
$\vec{k}$	Wave vector
λ	Wavelength
L	Counter length
l_k	Kuhn length
l_p	Persistence length
M_w	Molecular weight

μ	Attenuation factor
ν	Partial molar volume
N	Number of particles/ atoms
N_k	Number of Kuhn length segments
$p(r)$	Pair distribution function
$p_c(r)$	Pair distribution function of the cross–section
$\vec{q}$	Scattering vector
ρ	Density
r_c	Radius of the cross–section
ρ_e	Electron density
R_g	Radius of gyration
σ	Scattering cross–section of single atom
σ_{LJ}	Finite distance in Lennard–Jones potential
$S(\vec{q})$	Structure factor
SAS	Small angle scattering
V	Volume
z_t	Deflection of cantilever in z direction

In dedication to Björn and Jannik.

I, Imke Greving, declare that this thesis titled 'Structural Analysis of Silk Proteins using X–ray and Neutron Scattering' and the work presented in it are my own.

I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

Acknowledgments

Writing this doctoral thesis would have been impossible without the help and support of all the kind people around me, from the academic community as well as family and friends.

First of all, I would like to thank Prof. Fritz Vollrath for his academic and personal support throughout my time in Oxford. Thanks to his wealth of ideas, his unfailing motivation and enthusiasm his group is a unique working place. This fascinating scientific project would not have been possible without the man, the myth, the legend: Cedric Dicko. The invaluable discussions and his unsurpassed knowledge were very valuable for me and of high importance for the success of this project. Although he left for Sweden 3 month after I started in Oxford he was always available for gathering ideas in countless Skype discussions and providing huge support in writing beamtime proposals and during all our experiments at the large–scale facilities. A very special thank you is reserved for Ann Terry, not only my second supervisor, but now also a very good friend. She was always there when I was overcome by the 'last minute panic'. Her peaceful nature and inexhaustible patience combined with her unrestricted personal support were particularly invaluable for me in the last days and hours before the submission of my thesis. I am also most grateful to my thesis examiners, Dr. Barbara Gabrys and Dr. Manfred Burghammer, who provided encouraging and constructive feedback during my viva.

Above all, I am deeply grateful to my husband Björn for his self–sacrificing

support and never ending patience even in the most difficult times. A very big thank you to my little son Jannik for teaching me that there are more important things in life than just work. I am also grateful to my parents for their financial support throughout my (not always straight) career path and for the priceless support in the first three month after Jannik was born. My mother–in–law also earns a very big thank you for her great support and the countless times she traveled to Oxford to help us with Jannik.

I have enjoyed my time in the Oxford Silk Group very much, and it will be difficult for me to find such a good and friendly working environment in the future. A very big thank you to all the members of the Oxford silk Group, in particular to Chris Holland for his countless corrections of my English writing and, together with Maxime Boulet-Audet, a special thanks for all the help during beamline experiments. Additionally I am grateful to David Porter for the enlightening and helpful discussions and to my friend Maxie Roessler for all the help with the Chemistry. A special thanks to Tom for passing me on the black cloud. Anna, Fuji and Steph also deserve big thanks for always having an 'open ear' for me.

I would like to acknowledge the academic, financial and technical support of the University of Oxford and the Department of Zoology and their staff, in particular Toni Wheeler for her invaluable support and help with grants and funding issues. The EPSRC I gratefully acknowledge for the 'Next Generation User' stipend, which provided the financial basis of this research. I would also like to thank my collaborators in America, Prof. Hannes Schniepp and Minzhen Cai, for their support during my visit and their help with paper writing.

Last but by no means least, I would like to thank all my friends and supporters in Great Britain, France, America and elsewhere for their encouragement and support during my dissertation.

I General Introduction

I.1 Silk structure and function

Silk proteins (spidroin in spiders and fibroin in Lepidoptera insects) are constructed from highly repetitive amino acid sequences (the primary structure) containing mainly alanine, glycine and serine with varying small side chain residues depending on the species [11, 13]. In the case of the commercialized *Bombyx mori* cocoon silk the amino acid sequence has been fully determined [134] (see table I.1). For *Nephila edulis* silk the complete sequence remains unknown [22] although structures for the N and C terminal region have been determined from recombinant spider silks [4, 40]. The proteins involved in the fibre formation of *Bombyx mori* are the heavy chain fibroin (Fhc, 350kDa), the light chain (35kDa) and a protein chain called P25 [12].

Despite these differences in the side chain composition between *Bombyx mori* silks and spider dragline silks [12], the rheology [49] as well as the overall

	Fibroin (*Bombyx mori*)	Spidroin (MA)
Crystal content fibre	40 – 50%	35%
Amino acids	Glycine (45%)	Glycine (37,1%)
	Alanine (29%)	Alanine (21,1%)
	Serine (12%)	Serine (4,5%)
Amino acid sequence	GAGAGS	GPGGX
Protein	Heavy, light chain fibroin and P25	MASp 1 MASp 2

Table I.1: Structural differences in silks. [42, 19]

structure of the final fibre are surprisingly similar. The semi–crystalline silk fibre contains crystal like structures as well as less crystalline blocks up to totally amorphous parts [12, 47, 52]. The highly ordered crystalline structures are β – sheet structures (silk II) composed of short side–chained amino acids allowing a closer packing of the chains [46]. The less well–ordered blocks (silk I) contain mainly helical structures (e.g. alpha–helix) while the amorphous or disordered part consists of elongated unfolded polymer chains lightly connected by H–bonding [12]. An overview of secondary structures is given in figure I.1.

Interestingly, the exceptional mechanical properties of silk fibers seem to be due to the overall molecular structure, at the nanoscale level, instead of the complex hierarchical structure [100, 127, 12, 36]. Molecular modelling of silk fibers mechanical properties shows that there are mainly two controlling factors: on the one hand the ratio between ordered and disordered fractions and on the other hand the state of hydration of the protein chains [100]. These two parameters have to be controlled by the animal during the spinning process in a precise way to prevent accidental aggregation and mis–folding [126, 104]. The animal accomplishes that during the complex

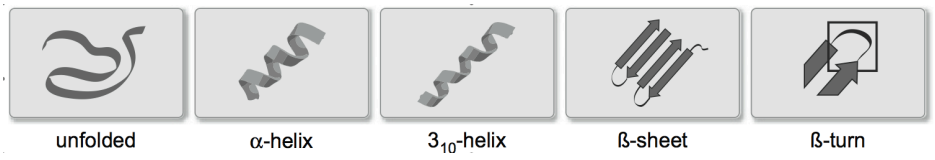


Figure I.1: Overview of secondary structures (adapted from [46]).

spinning process to form the desired proportions of ordered and amorphous structures forming a well–defined nano–fibrillar architecture [14, 106, 125]. It

is known that this process is controlled by changes in the chemical environment (pH, ions) of the silk proteins in the gland/duct and through physical stretching, aligning and shearing of the silk proteins during the extrusion process [23, 49, 8]. There seems to be substantial differences in the change of these parameters in the glands of spiders and silk worms. Findings from Knight et al. [71] indicated that the pH in the gland of a *Nephila* spider changes gradually from 6.9 to 6.3 as the silk proteins pass down the duct to form the thread. The ion concentration of Na^+ and Cl^- seems to decrease while an increase of K^+, P and S has been observed. Magoshi et al [85] stated that the pH change in the gland of the *Bombyx mori* silkworm is much more dramatic and goes down to pH 4.8 near the spinneret (silk spinning organ). In both cases, it has been suggested that metal cations are involved in some way to refold the silk protein molecules while the secretion of H^+ may assist in this process by changing the ionization of charged groups on the protein. More recent findings on the artificial spinning of regenerated silk find evidence [82] for the stated pH change during the spinning process. Studies on regenerated silk proteins indicate an increase of β –sheet structures when the pH is low (pH 5.2) while a more alkaline pH and increasing Ca^{2+} – ion concentration maintain in a random coil conformation [135, 82]. Additionally, studies of recombinant spider silk proteins suggests that the terminal regions play a key role in the storage and assembly process of major ampullate silk proteins acting as a switch for pH induced dimerization [4, 40, 79, 41].

Despite these insights, the details of the conversion process of silk proteins during the spinning process remains unclear. The starting parameters, the initial shape and structure of the silk protein in solution, have not been

resolved yet. Neither NMR, FTIR studies [2, 48], nor x–ray [3] seem to show any evidence of β – sheet structures in the precursor (un–spun) silk protein gel. It is widely believed that the proteins are mainly in a 'random coil' state but there are indications for helical structures present in silk glands shown by NMR, FTIR and circular dichroism (CD) [48, 20, 21] as well as in regenerated fibroin silk solutions [82] .

Previous studies to determine the shape and size of fibroin proteins have either examined dehydrated molecules [67, 61] or have been proposed indirectly from polarized optical microscopy [124]. Viney [124] stated from birefringence experiments that silk spidroin showed an evidence of supramolecular rod–like structures in the liquid precursor state. Later a rod–like shape (60 nm x 15 nm) has been observed for fibroin [61] by studying the dehydrated protein with AFM. An anisotropic type of shape has also been suggested by Knight et al as being one of the conformational states in liquid crystalline spinning of silks [72]. Jin et al. in contrast proposed a micellar like structure of the proteins in the silkworm gland by SEM, although the gland was treated with methanol which is known to induce conversions of silk proteins (sizes of micelles 100 – 200 nm, figure I.2) [67]. However, Foo et al. [30] supports the proposed micellar formation based on a change of charges in the system, which allows the heavy chain to collapse into a micellar structure of 200–300 nm in size.

Additionally most studies of silk proteins in solution have, due to the difficulties in handling the native silk proteins without accidental aggregation, focussed on regenerated silk or reconstituted silk proteins of various preparation types [87, 135, 82, 40]. This might have lead to the discrepancy over

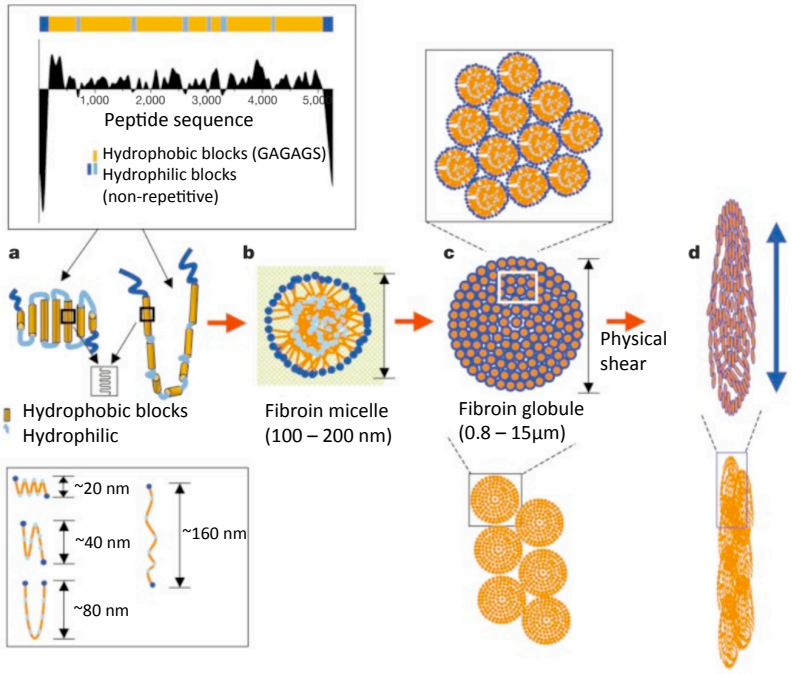


Figure I.2: Proposed micellar model of chain folding and shear processing of silk proteins (adapted from [67]).

whether fibroin forms globular [67, 30] or rod like [61] structures and could explain the variation is sizes given for the proteins. Therefore, in this thesis, the aim is to study fibroin and spidroin in a physiologically relevant environment in order to characterize the protein structure, shape and potential assembly process of native silk proteins.

An introduction to the methods used in this thesis is given in the second chapter. The first section (II.1) gives a basic introduction to scattering techniques with a focus on the small angle scattering both from neutron and x–ray. The basics of data reduction and analysis are presented and discussed in light of large and complex macromolecules such as silks. The following section (II.2) describes the atomic force microscopy (AFM) characterization methodology to study silk macromolecules, including the data reduction, modelling and interpretation. In the third section (II.3) the sample preparation of the solutions is explained and the effect of D_2O on the solutions is discussed.

Chapter III demonstrates the suitability of small angle scattering (SAS) to investigate the morphology and structure of native silk fibroin at near in vivo conditions. A comparison of native (SF) and reconstituted (RSF) silk proteins as a function of concentration is performed and different preparation methods for RSF are explored.

Single molecule and shear induced self–assembly of native silk proteins is studied in chapter IV using AFM. The shape and potential shear induced assembly pathways of silk and its dependence on protein concentration are investigated.

The structural characterization of silks is expanded to include spider silk

proteins in chapter V. The four principal spider silks major ampullate (MA), minor ampullate (MI), flagelliform (Flag) and cylindriform (Cyl) are studied using SAXS and CD. The effect of protein concentration is investigated followed by a comparison of the four silks at similar protein concentration. A link between the structural shape and function is considered and modelling of the SAXS data provides an insight into the flexibility of the spider silks.

In chapter VI, the findings from the thesis are summarized. Similarities between the different silks and their relevance for the hierarchical ordering of proteins are discussed.

Two papers have been published from the work in this thesis, thus far, based on the work in chapter III [38] and chapter IV [37] (see Appendix A).

II Methods

II.1 Small angle scattering

Small angle scattering (SAS) is a technique which provides information about large scale sizes, shapes and orientations of components inside the sample. The probe can be neutrons (SANS), x–rays (SAXS) or light (SALS or just LS) which makes this technique accessible to a large variety of samples. The basic theory and laws (like e.g. Guinier or Porod) are common for all the probes although they interact with the samples in different ways. A further difference between neutrons, x–rays on one hand and the LS on the other are the length scales that can be accessed. While the wavelength range of neutrons and x–rays lie in the range between 0.01 up to 3 nm the range accessed by LS is an order of magnitude larger (visible light range 400 – 600 nm) [84]. The importance of the SAS method for biology and polymer science becomes obvious when one considers the length scales typically measured. This range goes from bonding length dimensions of 0.1 nm over average R_g of polymers of 1 – 10 nm up to micelles or larger scale macro molecules with maximum sizes of 1000 nm. Neutrons offer the following advantages over x–rays: Neutrons are non destructive which is especially important when it comes to biological samples. Labelling of the sample becomes much easier with neutrons due to the fact that they are very sensitive to isotopes. In particular the difference between H_2O and D_2O in neutron scattering be-

comes very valuable for complex systems. By selectively labelling parts of the molecule with D_2O these parts can become invisible for the neutrons and substructures of the molecule can be studied. One of the main disadvantages in neutron scattering are the relatively long acquisition times compared to x–ray experiments at a synchrotron source like for example, the ESRF. This becomes problematic in the case of unstable systems or if fast transitions are of interest. Light scattering is also widely applied to soft condensed matter but requires low concentration and low opacity samples. Keeping these limitations in mind often a combination of the techniques becomes the ideal solution to study colloidal, protein and polymer systems to gain information about sizes, folding states and potential shapes in solution (see table of comparison II.1).

II.1.1 Basic theory of small angle scattering

The basic theory of small angle scattering (SAS) is similar for all three techniques, but there are substantial differences when it comes to the way the probe interacts with the sample. While x–rays interact with the electrons of an atom, neutrons interact with the core of the atom. Using the x–ray technique one can then study inhomogenities of the electron density while probing with neutrons detects the so–called scattering length density (see table II.1.1). This explains also the aforementioned difference between H_2O and D_2O in neutron scattering. While in x–rays the scattering intensity scales with Z^2, in neutron scattering the intensity varies significantly between isotopes. Despite these fundamental differences the basic mathematical formal-

Method	Samples	Advantages	Limitations
Crystallography	Crystals	Very high resolution (up to 0.1 nm) revealing fine detail of atomic structure (e.g. of the active centres)	Crystals required. Flexible portions are not seen. Structure may be influenced by crystal packing forces
NMR	Dilute solutions ($\sim$5–10 mg ml^{-1})	High-resolution (0.2–0.3 nm) in solution	Hardly applicable for MM exceeding $\sim$50 kDa
Cryo-EM	Frozen very dilute solutions ($<$1 mg ml^{-1})	Low amount of material. Direct visualization of particle shape and symmetry	Low ($\sim$1 nm) resolution. Hardly applicable for MM less than $\sim$200 kDa
SAXS/SANS	Dilute and semi-dilute solutions (from $\sim$1 to $\sim$100 mg ml^{-1})	Analysis of structure, kinetics and interactions in nearly native conditions. Study of mixtures and non-equilibrium systems. Wide MM range (few kDa to hundreds MDa)	Low ($\sim$1–2 nm) resolution. Requires additional information to resolve ambiguity in model building
Static and dynamic light scattering, ultracentrifugation	Very dilute solutions ($<$1 mg ml^{-1})	Non-destructive. Low amount of material. Simplicity of the experiments	Yield overall parameters only

Figure II.1: Advantages and limitations of important methods for structure analysis. [119]

ism is the same for neutron and x–ray scattering. This chapter is focussed on small angle neutron scattering and will point out differences to x–ray scattering. Introduction to scattering techniques in general can be found in the literature [118, 84, 129, 29, 43].

Table II.1: Source of contrast in scattering techniques (adapted from [84])

	Neutrons	X–rays	Light
Contrast	Difference of scattering length Δb^2	Difference of electron density ${\Delta\rho_e}^2$	Difference of refractive index $[\delta n/\delta_c]^2$

II.1.1.1 SAS intensity and cross–section

The basic parts of a scattering experiment are the source of radiation/ particles, the sample to be studied and a detector to monitor the energy of the probe scattered by the sample (figure II.2). In an ideal case the radiation (being x–rays or neutrons) can be described as a planar monochromatic wave. Each atom i at position $\vec{r}$ in the sample volume interacting with the radiation gives rise to a spherical symmetric wave. The elastic (no energy exchange) scattering process itself can be separated into two parts, coherent and incoherent scattering. Coherent scattering is the interference of the spherical waves from each atom and holds therefore information about the spatial distribution of the atoms in the sample. Incoherent scattering depends on the correlation between the positions of the same nucleus at different times (e.g. t=0 and t=1). Considering that the particle's motion is on average isotropic within the interaction time and length scale (as is a valid assump-

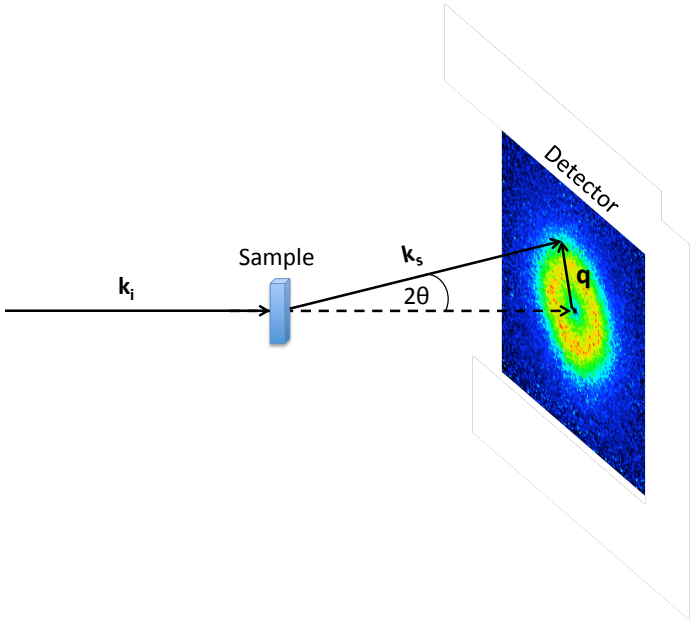


Figure II.2: Basic setup for a scattering diffraction experiments. The beam produced by the source is scattered by the sample and monitored by the detector. The scattering vector q (momentum transfer) is the difference between the incident wave vector k_i and the scattered wave vector k_s (adapted from [39]).

tion for diluted protein solutions), we can approximate the incoherent part of the scattering to an angle independent background which will be neglected for now (see section II.1.4.1). If b_i is the fraction of radiation scattered by atom/particle i the total scattering amplitude is the sum of the N wave functions A_i as [84]

$$A(\vec{q}) \propto \sum_{i=1}^{N} b_i e^{-i\vec{q}\vec{r}_i} \tag{II.1}$$

with $\vec{q}$ being the scattering vector (or momentum transfer) defined as (compare figure II.2)

$$q = |\vec{q}| = |\vec{k_s} - \vec{k_i}| = \frac{4\pi}{\lambda} sin(\Theta) \tag{II.2}$$

where $|\vec{k}| = k = \frac{2\pi}{\lambda}$ stands for the wave vector and $|\vec{k_i}| = |\vec{k_s}|$ in the case of elastic scattering (no energy transfer). It is convenient to use a continuous function

$$A(\vec{q}) \propto \int_{V_p} \rho(\vec{r}) e^{-i\vec{q}\cdot\vec{r}} d^3\vec{r} \tag{II.3}$$

introducing the 'scattering density distribution' $\rho(\vec{r_i}) = \sum b_i/V$ which is proportional to the scattering length density in neutron scattering. For x–rays $\rho(\vec{r_i})$ is proportional to the electron density distribution in the sample [84] since the photons are scattered from the spatially extended electron density distribution instead of a point source like nucleus (neutron scattering). However, taking into account that these atomic electron densities are roughly centered around the atoms each scattering centre can be identified with an atomic position or atomic side so the introduced formalism (equation II.3) is also valid for x–rays. The scattering length for an atom in x–rays is given

by the integral over all electrons

$$\begin{aligned} b_i^x(\vec{q}) &= b_e f(\vec{q}) \\ &= b_e \int_{Atom} \rho(\vec{r}) \exp(-i\vec{q}.\vec{r}) d^3r \end{aligned} \tag{II.4}$$

where $\rho(r)$ is the charge distribution in the atom and b_e=2.8 x 10^{-15}m the scattering length for a single electron. The $f(\vec{q})$ is the atomic form factor for x–ray scattering and depends on the scattering angle Θ. In most small angle scattering experiments where $2\Theta < 5^\circ$ the integral approaches the number of electrons in the system and $b_i = zb_e$ with z being the number of electrons [73]. Therefore the same formalism applied to neutron scattering can in this case also be used for x–ray scattering. The scattering amplitude however is not directly accessible by the experiment but only the complex conjugate A^*A which is proportional to the intensity measured. Since it is common to normalize the scattered intensity by the incident flux and solid angle element of the detector $\Delta\Omega$ the differential scattering cross–section is now introduced by

$$\frac{d\sigma}{d\Omega} = \frac{\text{no. of neutr./x–rays scattered per second into a detector element } \Delta\Omega}{\Phi\Delta\Omega} \tag{II.5}$$

where the flux Φ is the number of incident neutrons/ x–rays per second per unit cross–sectional area of the incident beam (typically in units of $\text{cm}^{-2}\text{s}^{-1}$). $\frac{d\sigma}{d\Omega}$ is usually in units of barn/str where 1 barn = 10^{-24}cm^2. The total cross–section σ is found by integrating the differential cross–section over all possible angles. For neutron scattering the total scattering cross–section of a single

atom is $\sigma_n = 4\pi b_1^2$. The intensity in counts per second measured in a detector solid angle element $\Delta\Omega$ is then given by

$$I(\vec{q}) = C'\Phi\frac{d\sigma}{d\Omega}(\vec{q})\Delta\Omega \tag{II.6}$$

with C' taking into account the experimental parameters and the instrumental setup. Considering now that the irradiated volume is large enough to contain all possible configurations of these identical particles and that these particles are diffusing in time (Brownian dynamics) the time average is the measured quantity. The scattering intensity becomes

$$I(\vec{q}) \propto \frac{1}{V}\langle\left\{\int_V \rho(\vec{r_i})e^{-i\vec{q}\cdot\vec{r_i}}d^3\vec{r_i}\right\}\left\{\int_V \rho(\vec{r_j})e^{i\vec{q}\cdot\vec{r_j}}d^3\vec{r_j}\right\}\rangle \tag{II.7}$$

where the statistical average $\langle\rangle$ is taken over all orientations and available positions of the particles. Introducing the centre position of a particle $\vec{r_i}$ with $\vec{r} = \vec{r_i} + \vec{u}$ the scattering intensity can be written as

$$I(\vec{q}) \propto \frac{1}{V}\langle\left\{\sum_{i=1}^{N} e^{-i\vec{q}\cdot\vec{r_i}}\int_{V_p}\rho(\vec{u})e^{-i\vec{q}\cdot\vec{u}}d^3\vec{u}\right\}\left\{\sum_{j=1}^{N} e^{i\vec{q}\cdot\vec{r_j}}\int_{V_p}\rho(\vec{v})e^{i\vec{q}\cdot\vec{v}}d^3\vec{v}\right\}\rangle \tag{II.8}$$

with V_p being the volume of a single particle. Re–arranging equation II.8 gives

$$I(\vec{q}) \propto \frac{N}{V}\langle\left\{\int_{V_p}\int_{V_p}\rho(\vec{u})\rho(\vec{v})e^{-i\vec{q}\cdot(\vec{u}-\vec{v})}d^3\vec{u}\right\}\left\{\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}e^{i\vec{q}\cdot(\vec{r}_j-\vec{r}_i)}\right\}\rangle \tag{II.9}$$

II.1.1.2 Form factor

In a very diluted system the particle positions are not correlated, so $r_i = r_j$ and the right term in equation II.9 becomes 1. The remaining part of equation II.9 can be written as

$$I(\vec{q}) \propto \frac{1}{V} \int_{V_p} \int_{V_p} \rho(\vec{r}_1)\rho(\vec{r}_2) e^{-i\vec{q}\cdot(\vec{r}_1-\vec{r}_2)} d^3\vec{r}_1 d^3\vec{r}_2 \tag{II.10}$$

which expresses the correlation between densities measured at two points r_1 and r_2 separated by $\vec{q}$ of the same scattering particle. Substituting $\vec{r} = \vec{r}_1 - \vec{r}_2$ so that $\vec{r}_2 = \vec{r}_1 - \vec{r}$ we obtain

$$I(\vec{q}) \propto \frac{1}{V} \int_{V_p} \int_{V_p} \rho(\vec{r}_1)\rho(\vec{r}_1 - \vec{r}) e^{-i\vec{q}\cdot\vec{r}} d^3\vec{r}_1 d^3\vec{r} = \int_V \gamma(\vec{r}) e^{-i\vec{q}\cdot\vec{r}} d^3\vec{r} \tag{II.11}$$

where $\gamma(\vec{r})$ is the auto correlation function of the scattering density difference inside the particle.

$$\gamma(\vec{r}) = \int_V \rho(\vec{r}_1)\rho(\vec{r}_1 - \vec{r}) d^3\vec{r}_1 \tag{II.12}$$

Equation II.11 is the single particle form factor $F(\vec{q})$ holding information about the particle shape

$$F(\vec{q}) = \frac{1}{V_p} \int_{V_p} \gamma(\vec{r}) e^{-i\vec{q}\cdot\vec{r}} d^3\vec{r} \tag{II.13}$$

Studying dilute protein solutions where no orientation or alignment is expected (isotropic) we can simplify equation II.13 by averaging over all orientations ($\Omega = 4\pi$) using the expression $< e^{i\vec{q}\vec{r}} >_\Omega = \frac{sin(qr)}{qr}$ where q and r are

the scalar magnitudes of their respective vector quantities

$$\begin{aligned} F(q) &= \langle F(\vec{q}) \rangle_{\Omega} \\ &= \frac{4\pi}{V_p} \int_0^{D_{max}} r^2 \gamma_p(r) \frac{\sin(qr)}{qr} dr \end{aligned} \tag{II.14}$$

with D_{max} being the maximum point–point distance in the particle. The distance correlation function $\gamma_p(r)$ gives the probability to be inside the particle after moving a distance r from a random point inside the particle. For a homogenous sphere of radius R $\gamma_p(r)$ can be given analytically. The characteristic function $\gamma_0(r)$ for a sphere is given by normalization of $\gamma_p(r)$ for the volume of the sphere V_s

$$\gamma_0(r) = \frac{\gamma(r)}{V_s} = 1 - \frac{3}{4}\frac{r}{R} + \frac{1}{16}(\frac{r}{R})^3 \tag{II.15}$$

$\gamma_0(r)$ goes to zero at $2R$ since the largest possible distance between 2 points in the sphere is its diameter. Here the 'pair distribution function' $p(r) = r^2\gamma(r)$ is introduced which can be obtained by Fourier transformation (FFT) of the single particle form factor

$$p(r) = \frac{r}{2\pi^2} \int_0^{\infty} qF(q)sin(qr)dq \tag{II.16}$$

and is a valuable tool for the data interpretation. Figure II.3 gives an example of pair distribution functions for typically geometric shapes together with the relevant scattering curves. Keeping in mind the limitations (mono–disperse diluted system, no inter–particle interactions) this method often allows a rough estimate of the overall shape of the (homogenous) particle in solution.

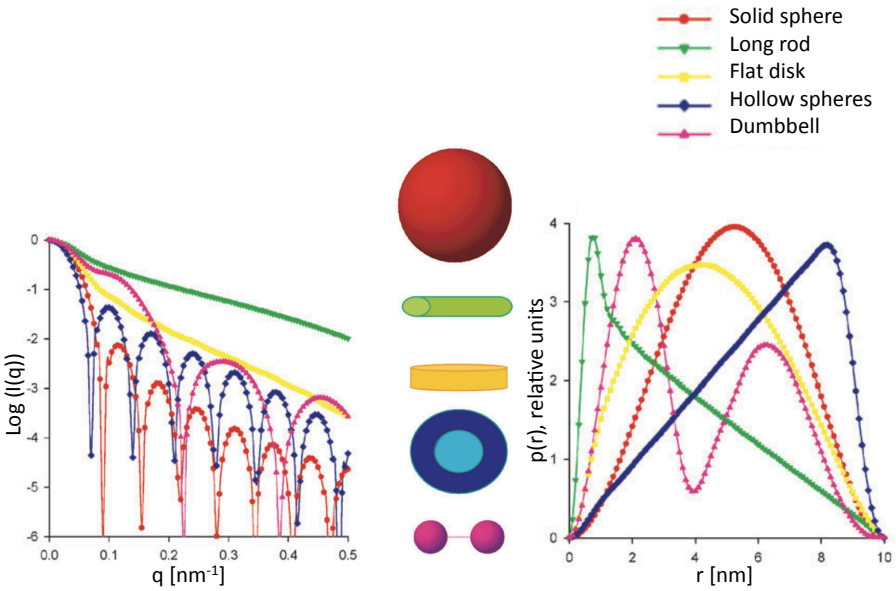


Figure II.3: Theoretical scattering curves of basic geometric shapes and their pair distribution functions obtained from Fourier transformation of the intensity (adapted from [119]).

However, since the averaging is over all orientations ($\Omega = 4\pi$) the phase information is lost and therefore a back transformation is far from trivial [84, 119]. Additionally the data are measured over a limited q–range, and are affected by instrumental smearing, which makes the correct computation of the p(r) difficult.

A method called Inverse Fourier Transformation (IFT) first proposed by O. Glatter [34] takes these problems into account when transferring the data into real–space. The data are approximated (smoothed) and corrected for instrumental smearing before the actual transformation. This method is also

implemented in the program GNOM by Svergun et al [121]. While a direct transformation would transfer the data directly from data space ($\vec{q}$) into real space (r) this method is creating an optimized function system which 'allows the calculation of the desired solution on the 'way back' [34]'. A maximum particle distance is introduced 'D_{max}' which is the upper limit of the largest particle of the dimension 'D' in the sample ($D_{max} \geq D$) [84]. It follows that $p(r) = 0$ for anywhere outside the particle $r \geq D_{max}$ in the case of diluted systems (with no inter–particle interactions so that the structure factor becomes unity). The maximum dimension of the particle D_{max} does not have to be known but can be obtained iteratively [119]. As an estimate for the maximum particle dimension possible in the measured q–range is given by the sampling theorem (Shannon channels) [119]. This implies that a scattering curve sampled with increments of $\Delta q \leq q_{min}$, with q_{min} being the lowest q measured holds information for all particles with a maximum dimension

$$D_{max} = \frac{\pi}{q_{min}} \tag{II.17}$$

Given the mentioned smearing effects and the loss of information due to counting statistics one would generally stay below this limit. However equation II.17 is a good first approximation for the D_{max} of the studied particle if the dimensions are unknown.

II.1.1.3 Structure factor

The second term in equation II.9 accounts for the correlation between particles i and j. In the case of N spherical particles with identical 'interactions'

this term is the structure factor $S(\vec{q})$

$$S(\vec{q}) = 1 + \frac{1}{N}\langle\sum_{i=1}^{N}\sum_{j\neq i}^{N} e^{i\vec{q}\cdot(\vec{r}_j-\vec{r}_i)}\rangle \tag{II.18}$$

when going to a continuous limit

$$S(\vec{q}) = 1 + \frac{N-1}{V}\int_V g(\vec{r})e^{-i\vec{q}\cdot\vec{r}}d^3\vec{r} \tag{II.19}$$

$g(\vec{r})$ being the radial pair distribution function giving the probability to find a second particle in a distance r from the first particle. Equation II.18 reveals, that for $q = 0$ $S(q = 0) = N$ which means that all the particles are scattering in phase, independently of knowing whether they are correlated or not. Therefore we introduce the measured structure factor $S_m(\vec{q})$

$$S_m(\vec{q}) = S(\vec{q}) - \frac{N-1}{V}\delta(\vec{q}) \tag{II.20}$$

Going to the continuous limit and performing the averaging over all possible orientations assuming an isotropic sample $\langle\rangle_\Omega$ it follows

$$S(q) = \langle S_m(\vec{q})\rangle = 1 + \frac{4\pi N}{V}\int_0^\infty r^2(g(r)-1)\frac{sin(qr)}{qr}dr \tag{II.21}$$

where $N >> 1$ for the particle solution scattering case and therefore $N-1$ is replaced by N. The information is contained in the radial pair distribution function $g(r)$ which can be obtained by Fourier transformation of the structure factor. Furthermore the $Ln(g(r))$ is directly related to the potential energy function describing the interactions between the scattering particles.

Using the definition $G(r)$

$$G(r) = \frac{4\pi N r^2}{V} g(r) \tag{II.22}$$

gives a damped, oscillating density distribution function and its maximum is related to the distance of each nearest neighbor. A separation of the form factor $F(q)$ and the structure factor $S(q)$ is not as easy as it looks in the first place, especially for more complex systems where interactions cannot be neglected. However a separation of the two terms for a semi–dilute solution is possible by using measurements at different concentrations or/and in different solvent conditions (pH, ionic strength etc.) [119, 84].

II.1.1.4 Particle in solution and contrast

So far the scattering intensity contribution from the solvent has not been taken into account. Making the approximation that the particles are surrounded by a homogenous (in the studied length scale regime) solvent with constant density $\overline{\rho}$, equation II.3 can be integrated over the whole illuminated volume V so the total scattering amplitude can be written as

$$A(\vec{q}) \propto \int_V \rho(\vec{r}) e^{-i\vec{q}\cdot\vec{r}} d^3\vec{r} \tag{II.23}$$

Expressing the density $\rho(\vec{r})$ by its mean $\overline{\rho}$ and the fluctuations $\rho(\vec{r}) = \overline{\rho} + \Delta\rho(\vec{r})$ equation II.23, becomes

$$A(\vec{q}) \propto \int_V \Delta\rho(\vec{r}) e^{-i\vec{q}\cdot\vec{r}} d^3\vec{r} + \overline{\rho} \int_V e^{-i\vec{q}\cdot\vec{r}} d^3\vec{r} \tag{II.24}$$

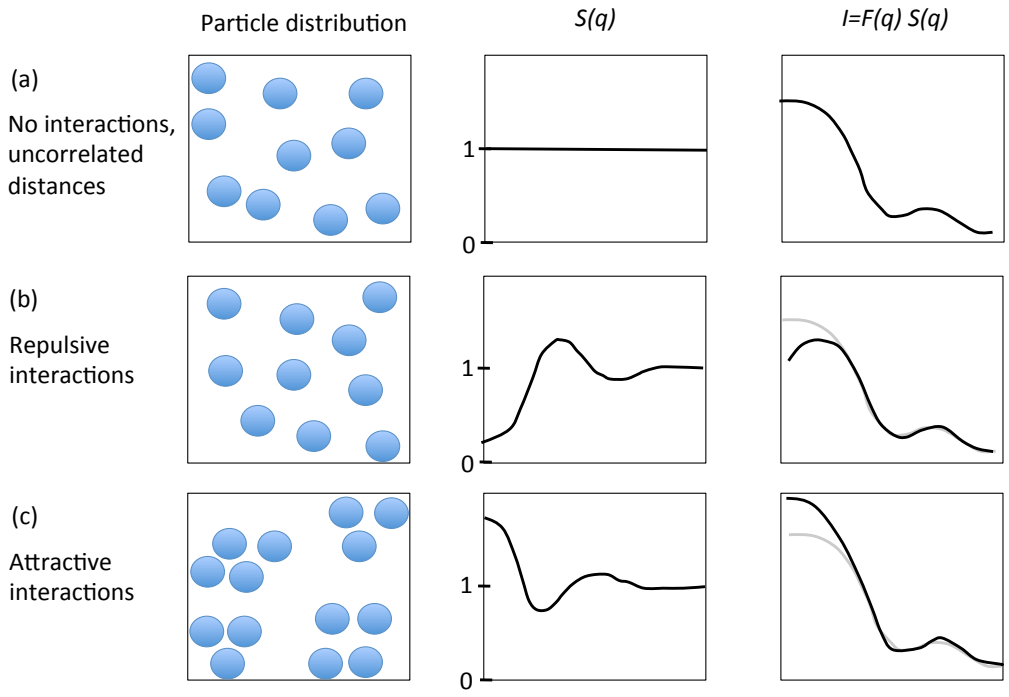


Figure II.4: In a dilute system (a), with random particle distribution the structure factor $S(q)$ is 1 and the scattering curve represents the experimental form factor of the particle. If the concentration is increased (b) and the particles have a repulsive interaction a decrease of intensity of the scattering curve at low q can be observed ((black curve in (b) compared to light grey). In contrast, if the particles have attractive interactions (they tend to aggregate) the intensity at low q is increased (black in (c)) compared to the system with no interactions (light grey, redrawn from [77]).

Any cross–terms coming from the complex conjugation of $A(\vec{q})$ are accounted to be zero since typically the sample volume is much larger than the volume of the sample. The second term is the scattering contribution of the whole solvent. Taking the macroscopic sample volume into account this term becomes a delta function for the complex conjugate of $A(\vec{q})$, giving a contribution only at $A(\vec{q} = 0)$. Experimentally this region is not accessible, since only contributions for $q > 0$ are detected. Comparing $A(\vec{q})$ of a protein in vacuum (equation II.3) with the first term in equation II.24 reveals that the only difference is the $\Delta\rho$ in equation II.24 which accounts for the fluctuations in scattering length density of the protein compared to the average scattering length of the solvent $\overline{\rho}$, the contrast.

$$\Delta\rho(\vec{r}) = \rho(\vec{r}) - \overline{\rho} \tag{II.25}$$

Therefore we can replace the ρ in the derivation of the intensity above with $\Delta\rho$. It is important to note that the contrast $\Delta\rho$ term can also become negative (e.g. if the density of the particle is lower than that of the solvent). Inhomogeneities in the solvent are usually accounted as very small (in the length scales studied by SAS) and their contribution rarely depends on the scattering angle. Experimentally any contribution from the solvent is taken into account by subtraction of the scattering of a pure solvent from the sample scattering (see section II.1.4). In a two phase system the neutron scattering length density difference is

$$\Delta\rho_n = [(N_A \sum_i \frac{b_s}{\nu_s}) - (N_A \sum_i \frac{b_p}{\nu_p})] \tag{II.26}$$

where b_i are the scattering amplitudes and ν_i are the partial molar volumes (cm^3 mol^{-1}). $\Delta\rho_n$ gives the contrast in every neutron small angle scattering experiment and has the unit of $length^{-2}$. If the scattering length densities of the solvent and the sample match there is no contrast and therefore no intensity. The so called 'match point' is determined by measuring the sample in different D_2O / H_2O ratios (see also section II.3). This can be very useful if it comes to more complex systems with several components hence one of the components can then be 'matched out'.

Note that in this derivation multiple scattering processes (e.g. a neutron scattered twice during its way through the sample) have not been taken into account. Particularly in thick samples which scatter well the contribution of multiple scattering can become significant. In case of x–ray scattering of diluted protein solutions multiple scattering can be neglected. Measuring proteins in water with neutrons however can lead to multiple scattering if the sample thickness is too large [84]. For a sample thickness of 1 mm and below multiple scattering in diluted protein solutions is known to have no significant contribution of multiple scattering to the total intensity [84]. Also incoherent scattering (see section II.1.1.1) contributes to the total scattering amplitude and has been neglected in this description as mentioned at the beginning. It gives rise to a flat 'background' which has to be added to the total scattering amplitude. Incoherent scattering (Compton scattering) is often neglected in the case of SAXS on protein solutions, as only small angles and light elements are considered. In neutron scattering, however, the incoherent background can become significant in particular if the particles of interest are measured in water. The incoherent background has therefore

to be subtracted from the data before further analysis or interpretation (see section II.1.4).

II.1.2 Non perfect solutions and polydispersity

So far the studied system had the assumption that the particles in the sample are of the same size and shape (monodisperse). Keeping the limitation of a dilute solution (no inter–particle interactions) now the case of polydispersity is discussed. For polydisperse systems or systems with a mixture of particles it is in principle not possible to determine the shape and size distributions of a system simultaneously from the scattering data [84]. In particular in the case of anisotropic particles the identification of polydispersity by SAXS becomes impossible [5]. The polydispersity of a system is ideally determined by dynamic light scattering (DLS) prior to the experiments. DLS however requires a large sample volume (0.5 mL) which is in some cases difficult to obtain. If the amount of polydispersity is known for a system this can be taken into account for the determination of the differential cross–section. The intensity in the idealized case of a monodisperse system of spherically symmetric particles the scattering cross–section is

$$\frac{d\sigma}{d\Omega}(q) = N_p V_p^2 (\Delta\rho_n)^2 F(q) S(q) \tag{II.27}$$

with N_p being the number of particles, $\Delta\rho_n$ the difference in scattering length density of the particle and the solvent, V_p the volume of the particles and $F(q)$, $S(q)$ are the form and structure factor as introduced in the sections above (II.1.1.2, II.1.1.3). Including size polydispersity in the scattering cross–

section for a dilute system ($S(q) = 1$) equation II.27 becomes [84]

$$\frac{d\sigma}{d\Omega}(q) = (\Delta\rho_n)^2 \int_0^\infty D(R)V(R)^2F(q,R)dR \qquad \text{(II.28)}$$

with $D(R)$ being the number size distribution, $V(R)$ the volume of the particle with characteristic length R (for spheres the radius) and form factor $F(q, R)$. In the case of polydisperse spheres R is a radius. Protein and nucleic acid solutions rarely display this kind of size polydispersity described by equation II.28 [73] but this equation is often applicable to micelles, micro–emulsions and nano–particles. In these cases the particle form factor is often considered as spherical and equation II.28 is used to obtain the size distribution function $D(R)$ by indirect Fourier transformation [73]. It is clear that all structural parameters obtained from a polydisperse system represent the ensemble average. In many form factors polydispersity can be included analytically by using the normalized Schultz–Zimm distribution [84]. In the more complex case of size and shape polydispersity the data can be fitted using a combination of simple bodies (like spheres or cylinders) or a singular value decomposition (SVD) can be applied to the data [73].

II.1.3 SAS instruments

The instruments used to collect the data presented in this thesis are outlined in the following two sections for neutron and x–ray instruments.

II.1.3.1 SANS instruments D22 and D11

Most of the presented data were collected at the small angle diffractometer D22 at the ILL (Institut Laue–Langevin, Grenoble, France). In a steady–state reactor source, such as the ILL the neutrons are continuously produced by fission processes. A schematic layout of the SANS steady–state pinhole instrument D22 can be found in figure II.1.3.1.

The wavelength of the incoming neutrons is selected through a mechanical velocity selector consisting of a drum with helically curved absorbing slits at its surface. This allows to vary the selected wavelength between 4.6 and 40 Å depending on the rotation speed of the velocity selector. The narrow Bragg–peaks of polycrystalline powder Silver behenate are used as a standard to calibrate the wavelength [39]. Directly after the selector unit a set of vertical and horizontal slits are mounted in order to define the size of the beam. If a higher instrumental resolution is required for an experiment the slit sizes can be reduced. Afterwards the neutrons pass through a monitor, which is a low efficiency detector used for data normalization during the measurement.

Since neutrons, unlike x–rays, cannot be easily focussed a series of waveguides is necessary to collimate the neutron beam. For D22 the collimation composed of eight guides with a cross–section of 55 x 40 mm^2 and a length yielding free flight paths of 1.4 up to 17.6 m. The length of each segment is calculated so that by removing or adding a section the flux decreases or increases by a factor of 2 [39]. Between two guide sections an anti–parasitic diaphragm is placed to minimize parasitic scattering from the wave guides. The size of the beam in front of the sample is determined by an aperture made

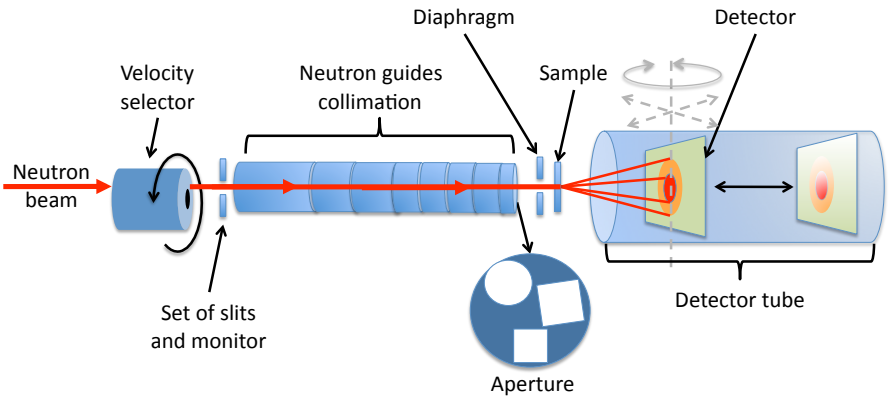


Figure II.5: A schematic representation of the SANS instrument D22 at the ILL (adapted from [39]). The wavelength range is 4.6 – 40 Å and the maximum sample–detector distance is 17.6 m [39].

of B_4C covered by cadmium. There are different shapes and sizes available depending on the sample geometry. The area detector (^{3}He) is placed in a evacuated detector tube of 20 m length and 2 m in diameter and can be moved from 1.1 to 18 m from the sample position [39]. A beam stop made of B_4C and cadmium, is placed in front of the detector preventing the direct beam from damaging the detector during the scattering experiment. For the transmissions the beam stop is replaced by an attenuator. In this way a total q–range from 8 x 10^{-4} to 0.8 Å^{-1} is achievable.

The D22 beamline is located close to the high brilliance horizontal cold source of the reactor which leads together with the large cross–section of the neutron guide, the short rotor and high transmissions of its velocity selector to a flux at the sample position of up to 10^8 neutrons/s/cm^2. The ILL has another operational small angle diffractometer D11. While comparable in wavelength range the main difference between D22 and D11 is the longer sample–detector distance available at D11. The detector of D11 can be placed at any distance between 1.2 and 39 m allowing to access a q range of 3 x 10^{-4} to 1 Å in real space. Additionally the flux at the sample position is higher at D22 compared to D11. The solution scattering sample environment is similar for both beamlines. A sample changer unit can be placed in position taking 22 cuvette cells (rectangular standard cells of 1 or 2 mm thickness) which can be thermostatically controlled by a water bath.

II.1.3.2 SAXS instrument ID14–3

The ID14–3 instrument at the ESRF (Grenoble, France) is dedicated to small angle x–ray scattering (SAXS) experiments of biological macromolecules in

solution. The incident beam size on the Ge(220) monochromator is defined by slits before the beam is focussed by a toroidal mirror. Further slits remove parasitic scattering and tune the resolution of the instrument [96]. The beam stop is mounted to block the direct beam and prevent damage of the 1M Pilatus detector, both being situated in the evacuated flight tube behind the sample unit. For normalization of the x–ray intensity a diode is mounted on the beam stop. The length of the flight tube itself can be varied from 0.5 to 2.5 m. ID14–3 is operated at a fixed wavelength of 0.931 Å (13.32 keV) with an accessible q–range of 0.05–5.8 nm^{-1} at a sample–detector distance of 2.5 m (standard operation mode) [96].

Using an automated sample loader, the liquid sample is transferred from the 96 well microplates into a 2 mm diameter quartz Lindemann capillary, mounted in the evacuated sample unit. The system automatically cleans the capillary and flow lines between each run. A camera is mounted above the capillary to allow the user to visually monitor sample loading and cleaning. This is in particular useful to see e.g. air bubbles in the system. It is also possible to flow the sample during the experiment minimizing radiation damage effects. The storage unit (samples in the well plates) as well as the capillary unit can be temperature controlled (2–60 °C) independently of each other.

II.1.4 Data reduction

In any scattering experiment the information of interest is the scattering intensity of the sample (e.g. the protein) and the angle at which this intensity

intersects the plane of the detector. The measured intensity however contains also the scattering from the solvent, the measuring container and is affected by instrumental and experimental parameters. Furthermore the data are measured typically by a two dimensional detector and radial averaging is performed to obtain the final scattering curve. Generally the following steps are done to reduce the data in a small angle experiment (isotropic sample) [39].

- Normalization of the 2D data with respect to the incident scattering intensity
- Correction for sample transmission and external background (e.g. electronic noise, masking beamstop)
- Radial averaging of the two dimensional scattering pattern giving the intensity as a function of q
- Background subtraction (subtraction of solvent and sample container scattering)
- Intensity calibration on absolute scale (in cm^{-1}) (if used)

II.1.4.1 SANS

The data from D22 and D11 experiments were reduced using the Matlab based GRASP software package [16]. In the following a quick outline is given of the data reduction steps for the SANS data presented in this thesis.

In a neutron experiment the collimated (ideally total parallel) radiation beam is partially transmitted, scattered and absorbed by the sample. The il-

luminated sample volume V is given by the beam size (cross–section) and the sample thickness d. The direct transmitted beam is blocked by the beam stop and the scattered intensity is monitored by the detector at a certain angle and detector distance L_d (solid angle detector element $D\Omega = dxdy/(L_d)^2$). The scattering intensity can therefore expressed as

$$I(\lambda, \Theta) = I_i(\lambda)\Delta\Omega\eta(\lambda)TV\frac{d\sigma}{d\Omega}(q) \tag{II.29}$$

I_i is the incident flux, η is the detector efficiency, T is the transmission, and $d\sigma/d\Omega(q)$ is the differential scattering cross–section introduced above (equ. II.5). The transmission accounts for the attenuation of the intensity on the way through the sample of thickness d and is therefore expressed as $T = e^{-\mu d}$ with μ being the linear radiation specific attenuation factor [84]. The transmission is measured by exchanging the beam stop with an attenuator. The transmission value of the sample is defined as the ratio of the outgoing to the incoming intensity at scattering angle $2\Theta = 0$. Since the sample is measured within a container the sample transmission is defined as the ratio of the transmission through the sample and container T_{s+c} and the transmission of the empty container T_c. Similarly the empty beam transmission can be defined as ratio between the empty container T_c and the incident beam intensity or 'empty beam', T_{eb}.

$$T_s = \frac{T_{s+c}}{T_c}; T_e = \frac{T_c}{T_{eb}} \tag{II.30}$$

The intensity of the sample (solute in solvent) corrected for empty con-

tainer scattering, background and electronic noise (I_{cd}) becomes

$$I_{corr} = \frac{1}{T_s T_e}(I_s - I_{cd}) - \frac{1}{T_e}(I_c - I_{cd}) \tag{II.31}$$

with I_c being the scattering intensity of the empty container.

As we can see from equation II.29 the data have to be normalized with respect to the incident intensity I_i. This value can be measured directly with a calibrated monitor installed at the sample position or be measured by the detector using a calibrated attenuator. The standard setting in GRASP normalizes the intensities for 'standard monitor'. The raw data counts n and errors $\sqrt{n}$ are subsequently divided by the number of actual monitor counts for the duration of the measurement [16]. Each detector needs a characteristic time to be able to separate (and therefore count) two events. Therefore the intensity has to be corrected for the 'dead–time' τ of the detector. In GRASP this is done by relating the real number of incident neutrons n to the apparent number actually measured by the detector n_m

$$n_m = \frac{n}{1 + n\tau} \tag{II.32}$$

The detector efficiency has also to be taken into account for the total scattering intensity of the sample. The detector efficiency η was calculated for one detector distance by GRASP software package using a water sample (1mm) [16]. Water or vanadium are often used, since they provide a uniform incoherent background and so fluctuations in the efficiency of individual detector elements show up. At the same time water is often used for absolute

scaling of the intensity (in units of cm^{-1}). However, it has to be taken into account that in the case of water the scattering is not totally isotropic due to inelastic and multiple scattering depending on the wavelength distribution, the geometry and the detector [39, 84]. Therefore it is extremely important to measure the calibrations at the same settings and conditions as the sample measurement. Due to long acquisition times at long sample to detector distances however it is desirable if the calibration data measured in one configuration can be also used for correction and calibration in another configuration. Ideally a calibration measurement would be done in one optimal instrumental configuration while the sample data maybe measured in one or more configurations. The data reduction program GRASP allows for this by using a data base of measurements of the neutron flux as a function of collimation and wavelength [16]. Taking relative values (differences in flux between different collimations and wavelength) this method allows the scaling of the non–calibrated detector distances with respect to the detector distance on an absolute scale. Also a cross–calibration with a second protein standard is used for evaluating the M_w of the sample

$$(\frac{d\Sigma}{d\Omega})_{sample} = \frac{I_{sample}}{I_{standard}}(\frac{d\Sigma}{d\Omega})_{standard} \tag{II.33}$$

This approach is mainly used in x–rays where the intensities are not on an absolute scale.

Another important correction is the solid angle correction taking into account that the spherical waves are measured by a flat detector. The geometric distortion has to be corrected for and is most significant for short

sample to detector distances [39].

As mentioned above the incoherent scattering contribution to the total cross–section cannot be neglected in neutron scattering. Incoherent scattering arises from the correlation between the positions of the same nucleus at different times and holds therefore no structural information about the sample (see also section II.1.1.1). The differential cross–section can be decomposed so that

$$\frac{d\Sigma}{d\Omega} = (\frac{d\Sigma}{d\Omega})_{coherent} + (\frac{d\Sigma}{d\Omega})_{incoherent} \tag{II.34}$$

The first term in equation II.34 is the coherent scattering cross–section as discussed above. The second term is the incoherent differential scattering cross–section and is considered to be independent of q in small angle solution scattering and therefore a (flat) background. In the case of a dilute mixture of (deuterated) proteins in hydrogenous solvent the incoherent contribution from the protein/polymer can be considered to be small and it is often sufficient to subtract the solvent scattering. If the protein concentration becomes high or a protonated polymer or protein is measured in a deuterated solvent one of the most common approaches listed below have to be performed [84]

- The incoherent background of the sample can be estimated with a measurement at the highest possible q since at high angles the coherent scattering becomes negligible. In this case the scattering curve levels off to a constant asymptotic value and this can be taken as a good approximation for the internal incoherent background of the sample.

- When the exact sample composition is known, a reference sample with

no structure containing the same amount of H and D molecules (e.g. H_2O/ D_2O mixture) can be measured.

- In the case that the q–dependence of the differential cross–section is known

$$\frac{d\Sigma}{d\Omega} = Aq^{-d} + B \tag{II.35}$$

where B represents the sample internal background. Plotting $\frac{d\Sigma}{d\Omega}q^d$ versus q^d yields

$$\frac{d\Sigma}{d\Omega}q^d = A + Bq^d \tag{II.36}$$

so the background B can be determined from the slope of the plot.

- In principle the incoherent background can also be calculated with the tabulated values of the incoherent scattering length. It has to be taken into account that the tabulated values are for bound atoms and that the spin incoherence is not taken into account.

II.1.4.2 SAXS

The SAXS data acquired at ID14–3 (ESRF, Grenoble) were reduced using the automated data reduction pipeline provided at the beamline. In general similar corrections as for the SANS data are applied to the x–ray scattering data. The incident intensity for normalization I_i is measured by an ion–chamber mounted in front of the sample unit. Instead of measuring with B_4C at the sample position, a dark–field measurement is performed with the shutter closed (beam off) to measure the electronic background noise. A diode mounted on the beamstop measures the transmission of the sample.

The detector efficiency for the Pilatus detector is also measured prior to the experiment and implemented in the automated data reduction pipeline [96]. The scattering from the solvent (or buffer) I_b is subtracted from the 1D data of the sample (protein in solvent I_{s+p}) using [74]

$$I_s = I_{s+b} - \alpha I_b \tag{II.37}$$

where the factor α accounts for the volume fraction with $\alpha = 1 - c\nu$, c being the protein concentration (g/mL) and ν the average partial molar volume of the protein 0.73 cm^3/g [74].

II.1.5 Data analysis

The following types of small angle data analysis have been derived for model systems, i.e. for no polydispersity (all particles have the same shape and size) and in some cases even an idealised two phase system. These techniques have been successfully used for polymer solutions and systems, which do not entirely fulfill all these requirements [43], [29], but one should keep in mind that assumptions have been made and that the observed values can only be seen as an average (see also section II.1.2). An overview about the limitations of the techniques can be found in table II.2.

II.1.5.1 Scattering invariant

A useful tool for SAS and especially polymers in solution is the scattering invariant Q_I. It considers the (coherent) scattering of the whole sample and is independent of the topology or geometry of the sample, assuming

Table II.2: Limitations of standard SAS analysis, [29, 43].

Analysis	Result	Assumptions	Requirements/ Limitations
Scattering Invariant	Volume fractions or contrast, data consistency	Two media approximation, constant scatt. length density	Independent of topology, no detailed structures, large $q-$range required
Guinier	General or specific dimensions	Dilute system, no interactions: $S(q)$=1	Guinier law $qR_g << 1.3$
Porod	Scattering from interfaces, specific surface area	Only very dilute system, infinitive sharp interfaces, randomly distributed	Good quality data at high angle needed, exact background subtraction
I(q=0)	Particle volume, molecular weight	Dilute system, no interactions	Guinier region at low q

a sample of two incompressible phases, a particle embedded in a matrix, with constant scattering length densities and volume fractions [129]. The total amount of scattering depends, as mentioned earlier, on the contrast (difference in scattering length density) and the volume fraction of each phase in the sample. Q_I for an incompressible two phase system is therefore

$$Q_I = \frac{1}{V}\int q^2 I(q) dq \tag{II.38}$$

This equation relates quantities from the sample directly to the scattered intensity.

II.1.5.2 Guinier approximation and molecular weight

Guinier analysis is only valid for dilute, monodisperse systems in which the uniform particles are distributed randomly (no correlations). Dilute means that there is no interaction between the scattering particles in the solvent. The intensity from a dispersion of N_p identical particles of volume V_p in a continuum volume V (sample volume) is

$$I(q) = \frac{N_p}{V}\Delta\rho^2 V_p^2 F(q) S(q) \tag{II.39}$$

In the case of a very diluted system and assuming no interactions $S(q) = 1$ so the angular dependence of the scattering intensity from the single particles can be calculated $I \sim F(q)$. As noted above the form factor $F(q)$ is strongly dependent on the shape and size of the scattering particle. However, Guinier pointed out that this is only valid at very small angles, corresponding to

inverse scattering vectors much larger than the size of the particle $qR_g << 1$ (Guinier Approximation). The intensity becomes therefore

$$I(q) \sim I(0)e^{-(\frac{q^2R_g^2}{3})} \tag{II.40}$$

where R_g is the radius of gyration (for homogenous particles only). Plotting $ln(I(q))$ vs q^2 should be linear and R_g can be obtained from the slope. Depending on the shape of the particle R_g can then be related to a size parameter R (examples in table II.3). Strictly speaking the Guinier approximation is only valid in the range of $qR_g << 1$ although for solution systems $qR_g < 1.3$ has been accepted in the literature [43].

Table II.3: Radius of gyration for different shapes [43].

Spheres	$R_g^2 = \frac{3}{5}R^2$
Cylinder	$R_g^2 = \frac{L^2}{12} + \frac{d^2}{8}$
Gaussian Coil	$R_g^2 = \frac{1}{6} < L^2 >$

In the case of polydisperse systems, the Guinier plot is curved and the obtained R_g at the linear part gives only the average value weighted by the size of the larger particles due to the fact that the larger particles scatter at lower angles [129]. For some systems it is difficult to have no interactions in the system which is required for the Guinier analysis. For systems which cannot be diluted but the interactions are known to be weak an extrapolation of the experimental R_g to zero concentration is an accepted tool [29].

If the intensity is known on an absolute scale, $I(q = 0)$ is related to the relative molecular weight M_w by extrapolation of the intensity to zero angle.

This implies that the data at low angle are of good quality and that the assumptions made for the Guinier analysis are kept in mind. From equation II.39 the intensity at zero angle $I(q=0)$ becomes

$$I(q=0) = n(\Delta\rho)^2 {V_p}^2 \tag{II.41}$$

with n being the number density of particles (N_p/V). Taking the substitutions [65]

$$n = \frac{cN_A}{M_w} \quad and \quad V_p = \frac{\nu M_w}{N_A} \tag{II.42}$$

the molecular weight can be determined using

$$M_w = \frac{I(0)N_A}{c(\Delta\rho\nu)^2} \tag{II.43}$$

where N_A is the Avogadro constant, c the concentration of the protein (in g cm^{-3}), ν the partial molar volume and $\Delta\rho = (\rho_{sam} - \rho_{sol})$ is the difference in the scattering length density of the sample and the solvent. The partial molar volume v is the inverse of the density of the particles in solution, and in the case of sufficiently large proteins (>60kDa) is known to be 0.74 cm^3 g^{-1} [28, 108]. Theoretically calculating ν for proteins from the AA sequence can be performed using the equation II.42. Taking the theoretical volume of the protein V_p by summation of each of the AA volumes (given by Jacrot [65]) and dividing this by the theoretical M_w for fibroin heavy chain protein 391.55 kDa gives a $\nu = 0.733$.

For x–ray scattering where a direct determination of the scattering intensity on an absolute scale is not possible secondary standards have to be

used for the molecular weight calculation. Often a water sample is used as a calibration standard or, more commonly, a protein standard with known M_w is measured [92].

II.1.5.3 Porod limit

Correlations much smaller than the typical length scales in the sample volume scatter in the Porod region, at high q, which provides information about the local structure of the scattering objects themselves. Again it is assumed that the structure factor $S(q) = 1$, so the system has to be very dilute with no inter–particle interactions. This assumption has to be strict due to the sensitivity to inter–particle correlations at high angle. In this case the form factor $F(q)$ represents the single particle scattering.

Porod showed that for an infinitively sharp interface the surface area per unit volume $A = S/V$ is related to the scattering intensity by

$$A \sim q^n I(q) \tag{II.44}$$

which is known as the Porod limit. This implies, that fitting a horizontal line to $log(I(q))$ vs $log(q)$, the Porod plot, yields information about the fractal dimension of the scattering objects. In the case of polymer coils, the Porod slope n is related to the excluded volume parameter $v_e x$ as its inverse $n = 1/v_{ex}$ [43]. The data at high angle (large q) has to be of good quality and the background subtraction has to be very precise or the tail of the scattering curve cannot be interpreted in terms of the surface area. Porod's law is valid at all distances for which the interface is seen as sharp. For samples where

the scattering length density gradually changes from one phase to the other, the intensity has been calculated by convoluting the ideally sharp profile with various smoothing functions [129].

II.1.5.4 Kratky plot

The Kratky plot ($q^2I(q)$ vs q) is used as a characterization method for the folding state of a protein [24, 90]. Originally introduced by Kratky [35] for obtaining the persistence length of a Gaussian coil molecule the interpretation of this useful tool has been established to characterize folded and unfolded chains in solution scattering [24]. In the case of a molecule in solution showing characteristic Gaussian coil configuration, the Kratky plot reaches a flat plateau region [35]. If the chain of the molecule is totally unfolded the Kratky plot lacks a characteristic plateau and the curve keeps rising linearly with q [35, 24]. In contrast a fully folded structure would show a bell shaped peak at low q [24, 90]. States in between the fully folded and the extended plateau region are counted as partially folded structures. Flexible multi–domain proteins can also potentially be identified by using the Kratky plot. These proteins show a combination of the characteristics of a partially unfolded state and an unfolded state [90].

II.2 Atomic force microscopy (AFM)

In this section the basic principle of Atomic Force Microscopy (AFM) is outlined. Later dynamic amplitude modulation technique used in chapter IV is described. A brief introduction into image processing focussing on the techniques used in this thesis is provided followed by examples of the image analysis.

II.2.1 Basic principles of AFM

AFM, originally invented by Binnig et al. [7], is a high resolution scanning probe microscope and has become a powerful tool to obtain images from proteins, polymers and biological samples on a nanometre scale [33]. In contrast to scattering techniques such as SAXS, AFM allows imaging in real space making a direct visualization of the molecule possible. The general principle of an AFM is shown in figure II.6.

The probe is a cantilever which is typically made out of silicon or silicon nitride. The tip of the cantilever scanning the sample surface has a curvature in the order of nanometre. The spring–like cantilever is now, due to the interacting forces between the tip and the sample surface, more or less deflected (z_t) depending on the type and strength of the forces (F) as well as its spring constant (c_s) according to Hooks law:

$$F = c_s z_t \tag{II.45}$$

The attractive forces are the van der Waals forces and the repulsive forces

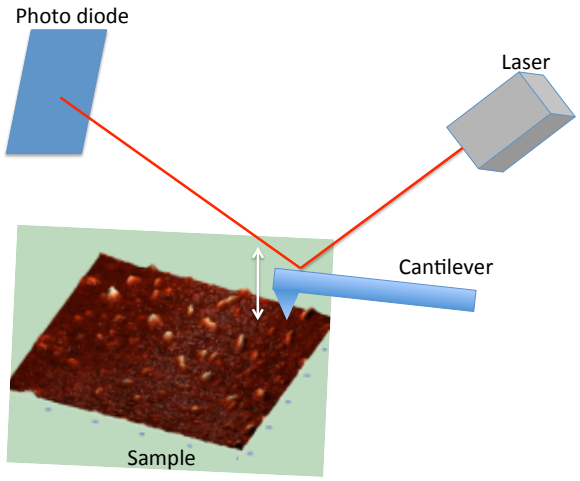


Figure II.6: The cantilever tip scans the surface of the sample. Its deflection, depending on the spring constant and the tip–sample interactions, is monitored by a laser beam deflected off the cantilever and detected by a photodetector.

are due to the overlapping electron orbitals (Pauli and ionic forces). Typically the involved forces are in the range from 10^{-11} to 10^{-6}N (for comparison two covalent bonded atoms at an $\approx$ Å distance are of the order of 10^{-9}N) [91]. The interplay of these forces can be described in a Lennard–Jones potential

$$V(d) = 4\epsilon[(\frac{\sigma_{LJ}}{d})^{12} - (\frac{\sigma_{LJ}}{d})^{6}] \tag{II.46}$$

where ϵ is the depth of the potential well, σ is the finite distance at which the inter–particle potential is zero and d is in this case the distance between the tip and the sample surface. The first term in brackets describes the repulsive interactions (short range forces) and the second term the attractive forces (long range) as indicated by dashed lines in figure II.7.

The deflection of the cantilever is detected using a laser beam pointing at the cantilever (figure II.6). The change in the reflected laser beam is then observed via a photo detector. The digitalized signal of the monitored deflection (z-movement) in combination with the lateral position of the tip (x–y) gives a topography image of the scanned surface.

Two force regimes are distinguished in AFM: contact and non–contact mode depending on the chosen force regime (see figure II.7). For proteins, polymers and biological samples, the non–contact mode is often chosen, since the tip is not touching the sample avoiding deformation or displacement of the molecules. There are two modes of operation, either the static (dc) or the dynamic (ac) mode (see figure II.8). In the static mode the deflection (= force) is kept constant by regulating the height of the sample relative to the cantilever tip (*equiforce* mode) or the height is kept constant and the

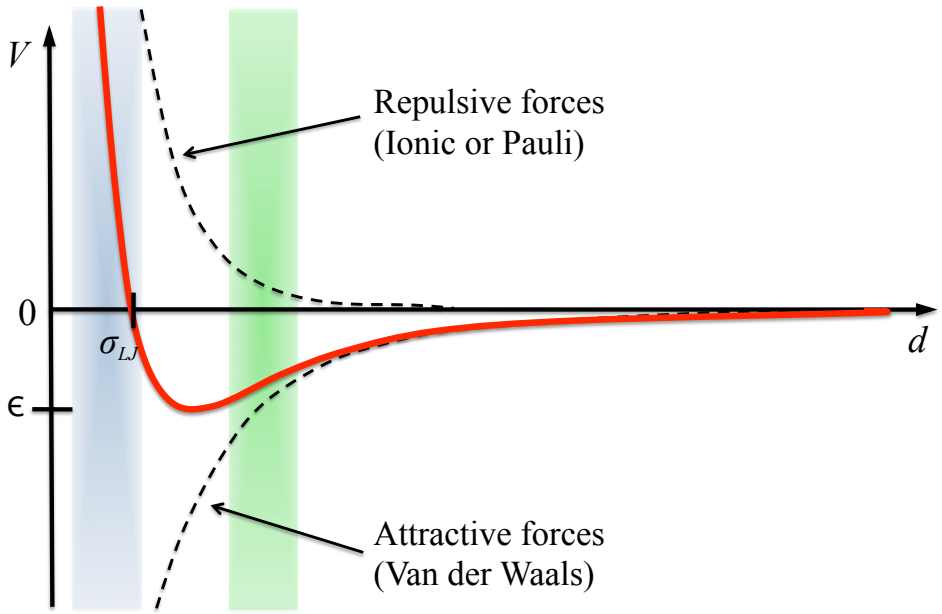


Figure II.7: Schematic drawing of the Lennard–Jones potential and the interplay of repulsive and attractive forces as a function of the sample–tip distance d. The coloured bars represent the force regions for contact–mode (blue) and non–contact mode (green).

deflection is regulated (*variable deflection* mode) [91].

II.2.2 Dynamic AFM

In this thesis the dynamic mode is used for the experiments in chapter IV. A comprehensive review of dynamic AFM has been published by Garcia et al [33]. In contrast to the static mode the cantilever is oscillating at a frequency close to its resonance frequency in the dynamic mode. In this case the amplitude, the resonance frequency and the phase shift of the oscillation are linked to the tip–surface interactions. Any of these parameters could be used as a feedback parameter (kept constant) for acquiring the surface topography but mainly two types of modulations are used: the amplitude modulation (dynamic AM–mode) and the frequency modulation (dynamic FM–mode). Due to the fact that the majority of the FM–mode experiments are conducted in UHV, while AM–mode experiments are performed in air or liquid the latter is more often used for soft matter and a description of this will be the focus of this chapter. For the dynamic AM–mode the cantilever is excited (using e.g. a piezoelectric actuator) at a fixed frequency (normally close to the cantilever's free resonance frequency). By using the oscillation amplitude as a feedback parameter the sample topography is imaged. In the AM–mode the two main experimental quantities of interest are the amplitude and the phase shift. Usually the amplitude modulation involves a range of 1–100 nm, where the analysis of large oscillation amplitudes is more complicated. This is easily explained taking into account that during one large oscillation the force gradient varies significantly leading to non–linear

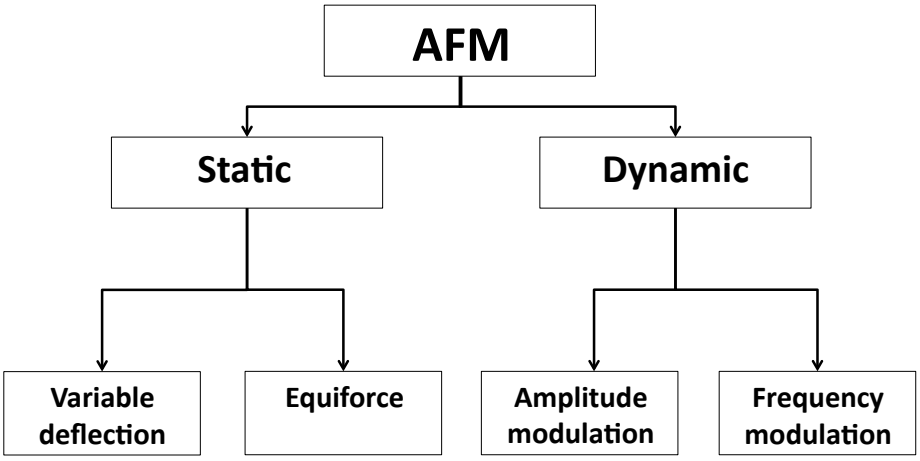


Figure II.8: Possible modes of operation for AFM. For soft condensed matter, the dynamic amplitude modulation (AM–AFM) mode is often used since in this mode the risk of deformation and displacement of the macromolecules is minimized.

features in the dynamics of the tip motion.

Given the presence of repulsive and attractive forces and their non–linear character as indicated by the Lennard–Jones potential, a coexistence of two stable oscillations is found: a low oscillation and a high oscillation amplitude branch. For '*true*' non–contact mode the first one is chosen implying no mechanical tip–sample contact. This mode has been used in chapter IV. A basic model of a point mass spring provides some insights into the motion of the cantilever–tip ensemble:

$$m\ddot{z} + c_s z + \frac{m\omega_0}{Q}\dot{z} = F_{ts} + F_0 cos(\omega t) \tag{II.47}$$

where F_0 is the amplitude and ω the angular frequency of the driving force, while ω_0 is the angular resonance frequency. Q is the 'Quality factor' accounting for parameters from the cantilever (e.g. length, thickness, ...) and c_s is the force constant of the cantilever. F_{ts} accounts for the tip–surface interaction forces and is dependent on parameters like the tip–sample distance d as well as the tip radius. If tip–sample interactions are absent, equation II.47 describes the motion of a forced harmonic oscillator with damping. Although the harmonic models used to describe dynamic AM–AFM give useful insights into the basic principle of the technique, they do not allow quantitative or even semi–quantitative agreements with the experiments and have even lead to misconceptions [33]. However for the purpose of this basic introduction into the AM–AFM technique this model gives an idea of the basic principle of this method of imaging.

Simultaneously to the topography image the lag in phase signal of the tip oscillation relative to the excitation signal can be monitored while the feedback keeps the amplitude constant. The phase signal is sensitive to surface forces (e.g. adhesion) but also to bulk material properties (e.g. elasticity or viscoelasticity). This combination often allows an enhancement of contrast and is particularly useful if the resulting topography image is difficult to interpret or analyze due to height differences. Extracting material properties from phase imaging however is not straight forward and requires often the use of a model. Since the phase images presented in this work are not used

for the analysis of material properties, no further details will be presented here.

II.2.3 AFM experimental setup

The experimental setup for the AFM experiments presented in chapter IV is shown in figure II.9. A light microscope with a resolution of 1 μm is mounted precisely above the scanning unit to enable easy navigation of the sample as well as taking snapshots with an installed microscope camera (left figure in II.9). The scanning unit itself is shown in a close–up, with the main components indicated (right figure of II.9). A quadrant photodiode detects the deflected laser light from the AFM probe. The AFM must be isolated from any external vibrations.

II.2.4 Imaging artifacts

When analyzing the sizes and shapes of single molecular structures absorbed on a surface it is important to keep in mind the asymmetry between lateral (along the surface) and vertical (normal to the sample surface) scanning direction. The studied polymer or protein structures can appear smaller than expected in the vertical direction (height). This can e.g. be caused by protein deformation, for example, resulting from the imaging itself or attractive forces between substrate and sample [97, 109]. A smaller appearance in vertical direction can also result from protein dehydration [33, 60], while a quick drying of the structures after deposition is thought to minimize this effect by 'freezing' the shape. Additionally modulations of the tip–sample gap

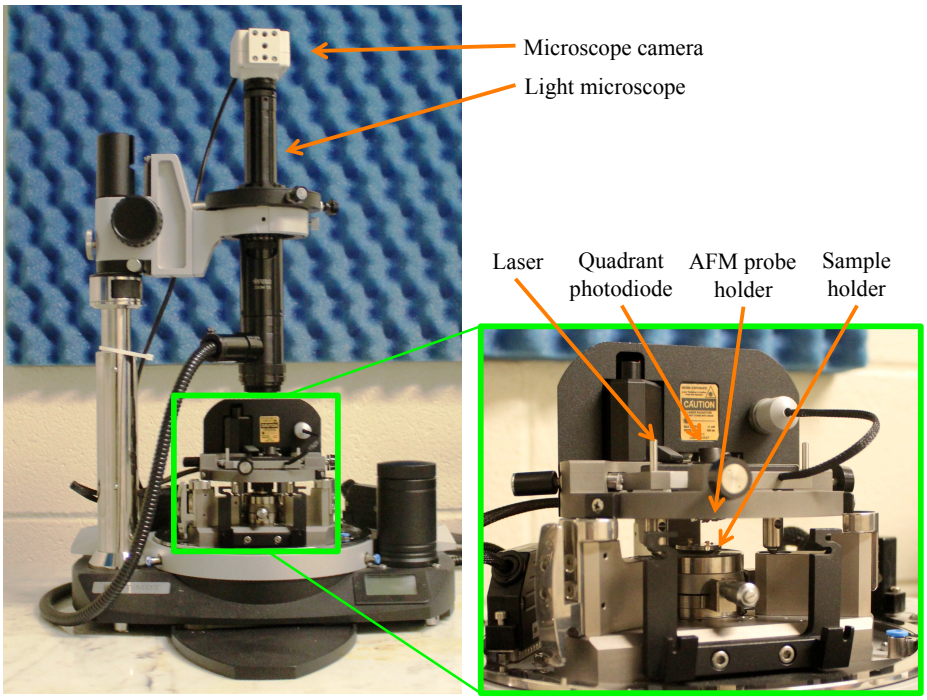


Figure II.9: The figure on the left shows the NTEGRA Prima Scanning Microscope. On top of the AFM setup a light microscope allows instant visualization of the sample surface for easier navigation. The right image shows a close–up of the scanning unit under the light microscope unit. The arrows indicate the positions of the photodiode, laser, AFM probe and sample holder.

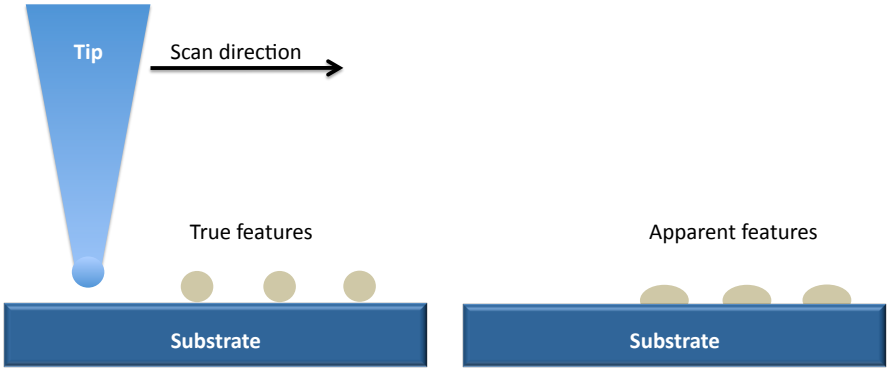


Figure II.10: This schematic drawing visualizes the broadening effect in single molecule studies in the lateral direction caused by the tip–sample convolution. This effect becomes significant when the size of the AFM probe is similar to the size of the observed features.

caused by variation in the spatial tip–sample interactions [33, 93] as well the averaging of those interactions over a larger area than the size of the imaged features, leads to an apparent decrease in the height of the imaged structures [109].

While the height appears decreased the width of the molecules from the images is often overestimated, in particular if the tip size is in the size range of the studied molecules. This leads to a broadening of the features caused by the convolution of the tip and the sample (see figure II.10). Additionally, potentially attractive interactions between the molecule and the substrate

can lead to a flattening of the structures so that they appear larger. In order to improve the lateral accuracy of the AFM measurements, it would be necessary to determine the size and shape of each of the AFM tips using an external size standard [114]. This procedure, however, appears to be extremely difficult to perform in the case of the 'super sharp' tips used in this work. It was found that the tips broke quite easily making an individual tip calibration prior to each sample measurement impossible. Reducing the significant uncertainty in the protein molecule's height determined via dynamic–mode AFM requires a rigorous understanding of the effective tip–sample potential, which is still an area of active research [109]. In particular when the size of the AFM probe approaches that of the molecule imaged these uncertainties in overall dimensions become substantial and have to be kept in mind in particular if a quantitative analysis of the sizes of the imaged structures is performed.

II.2.5 AFM image processing

Often it is necessary to optimize the image quality before analyzing the AFM image. In this section the standard methods in AFM image processing are briefly reviewed focussing on the techniques used in chapter IV. A detailed and recent review on image processing in AFM can be found in 'Atomic Force Microscopy', chapter 5, by P. Eaton et al [27]. The images presented in this thesis (chapter IV) have been processed using the instrument software NOVA AFM [26].

A very common distortion found in the raw AFM images is caused by

either tilted samples or non–linear scanner behaviour. This can cause the surface features to not be visible or to appear deformed in the raw images. Therefore the first step in image processing is often to 'level' the sample. A widespread method for this AFM image leveling is the so called 'line–by–line levelling' or 'polynomial fitting'. Each line in the image is fitted with a polynomial equation which is afterwards subtracted from the scan line. Then the average height of each line is set equal to the previous corrected line [27]. For most images a first order polynomial fit will be sufficient; sometimes, in particular in large scans, a second order is applied. In the case of single particle imaging this method can cause additional artefacts especially if higher order polynomials are used for fitting (3 or more). The isolated features might then be included in the polynomial fitting causing streak marks in the image. Therefore the isolated features might need to be excluded from the line–by–line leveling. An example is given in figure II.11. Generally the horizontal scanning axis is used for the leveling method since this axis is usually the fast scanning axis, thus any discontinuities introduced by changes in the imaging conditions over the time of the scan will best accounted for. Other methods of accounting for height distortions are 'two–dimensional plane fitting' and 'three–point fitting'. In the first method an automatically fitted flat plane is subtracted from the image, which works well if the background is completely flat. If large features are present, or the background has a terrace like feature this reduces the accuracy of the fit. Therefore often a more manual approach is needed, which is the three–point fitting. This method allows the user to identify 3 points on the image as a plane which is then subtracted from the image. In this manner the user can

distinguish between the features from the substrate and the sample allowing a more accurate subtraction. However, both methods cannot account for any curvature in the images, since a flat plane is subtracted therefore the polynomial method explained above is more common and has been used for leveling backgrounds in chapter IV.

Filtering is applied to the images if the noise level is too high to obtain a reasonable result. Irregularities in surface roughness, humidity, sample impurities as well as mechanical or electronic instabilities can cause noise in an AFM image. Either matrix filtering or Fourier filtering can be used to reduce the noise level. In matrix filtering adjacent points in the image are averaged by replacing a centre pixel in a square matrix of an image so that certain frequency components are removed. Two types of matrix filters are common, the 'low–pass' and 'high–pass' filters. In the low–pass filters, low–frequency components can pass, while the high frequency part is reduced, thus having a smoothing effect on the data (contrast reduced). If the high–frequency part is reduced too much a blurring is observed. High–pass filters have the opposite effect: the low–frequency components are reduced leading to a sharpening and often allowing a better edge detection.

In Fourier filtering the frequency components of the image are calculated (FFT images) and the unwanted frequency components can be identified and removed before back–transformation is performed. Since most unwanted sources of noise have a characteristic frequency this method makes it possible to identify these sources and remove or suppress them. If the features of interest are on a noisy background it is possible to select the features and remove the background noise around them before back–transformation into

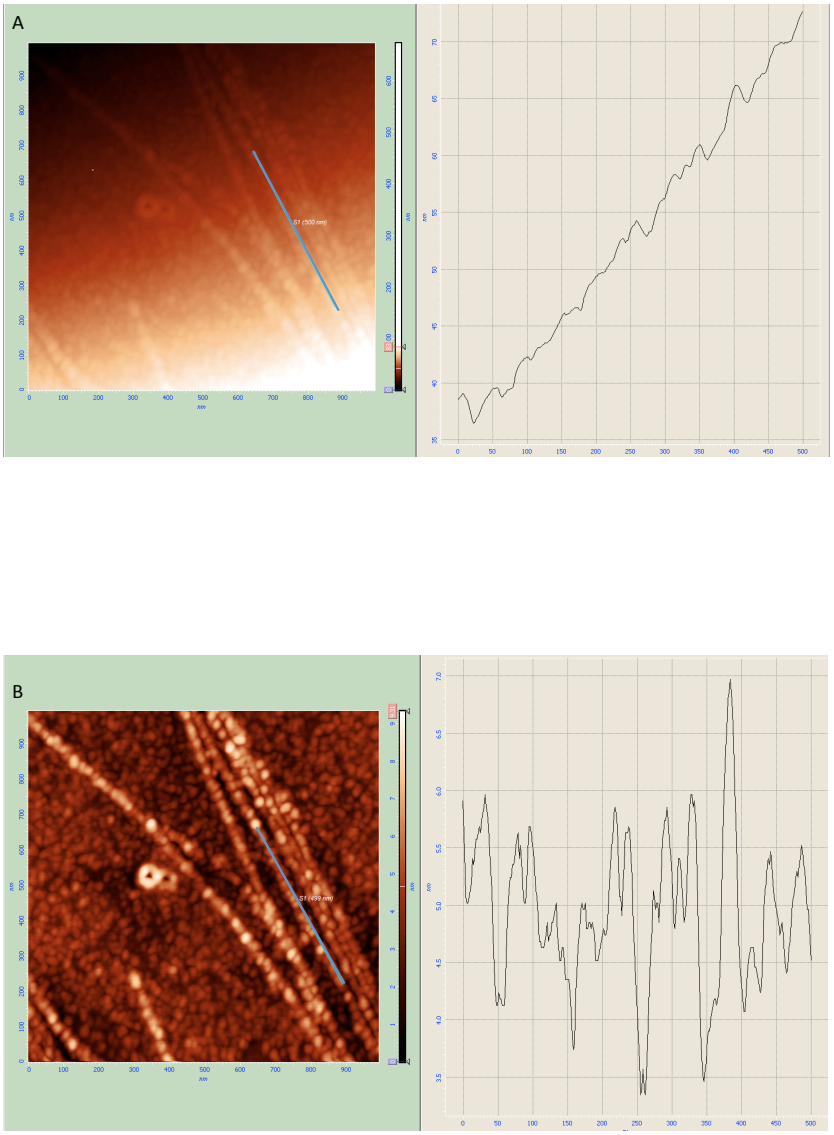


Figure II.11: These images give an example of 'line–by–line' levelling in AFM image processing using the NOVA–AFM analysis software package. The raw image is shown in the top image (A) with its cross–section on the right (taken along the line indicated in blue in the topography image) revealing the slope of the background. The processed image is shown in (B) where the cross–section taken over the same range reveals the corrected data.

real space. However, care must be taken, since this procedure might also create artefacts. In this thesis the FFT filter has not been used for data treatment, however the FFT is a useful tool to discover periodicity in the image and analyze spacings (see example below).

II.2.6 AFM image analysis

One major advantage of AFM is that most values of interest can be extracted directly from the AFM images. The determination of the cross–sections for the single molecules and the width of the fibrils of silk were determined using the cross–sectional tool in NOVA AFM [26]. Having corrected the background by levelling, the FWHM of the cross–section of the feature can be determined as indicated in figure II.12. The values obtained for the particle width are likely to be overestimated, while the height is likely to be underestimated for reasons stated in the section II.2.4 above. For example the diameter of the single molecule analysis was an average from 12 cross–sections of single and cluster units. The statistical analysis of the volumes of these objects was done by Minzhen Cai using 91 single and 85 clusters (consisting of 2-4 subunits). The background was 'subtracted' from the molecules using the program written by Prof. Schniepp (ScanHaSee). Afterwards Matlab was used to extract the apparent volume from each molecule or cluster.

For aligned molecules at higher concentrations ('sheared' samples) the apparent fibrillar width as well as the 'bead' size along the fibrils can be determined using the the cross–sectional tool over sections across a bundle or along the fibril respectively. Due to the potential artefacts stated in section

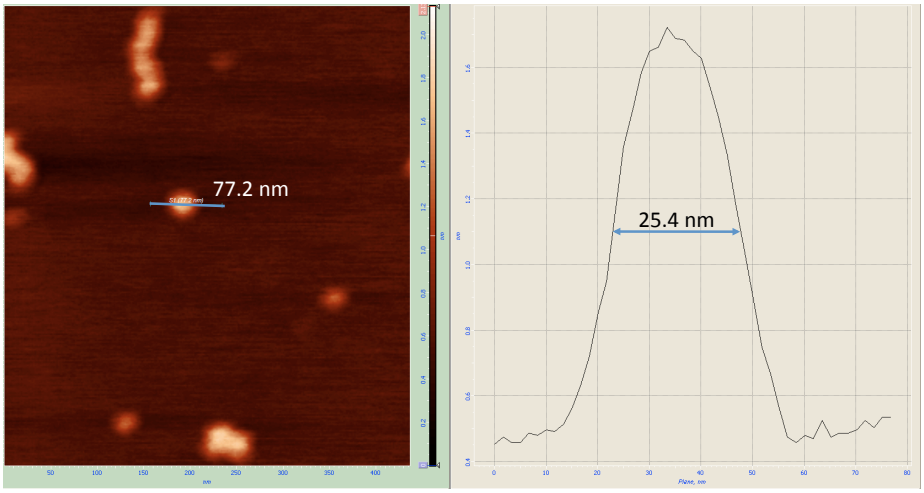


Figure II.12: Image from the AFM NOVA software [26] measuring a cross–section of a single feature at FWHM. The line taken for the cross–section is indicated as a blue bar in the topography image (left). The height of the feature is $\sim$ 1.2 nm.

II.2.4 and the roughness in these images as mentioned in chapter IV the errors are significant, however averaging over a large number of beads reduces the errors. Therefore 5 cross–sections of fiber bundles were taken from the topography image of figure IV.5 (chapter IV) to estimate the average width of a fibril. Assuming close packing the length of the cross–section was therefore divided by the number of beads and an average fibrillar width obtained (error determined from standard deviation). The average bead diameter along the fibrils has been calculated using 6 large cross–sections. In this case the error has been calculated taking the counting error of the number of beads per cross–section into account (the standard deviation error would be smaller).

As mentioned above in section II.2.5, FFT can also be used to determine periodic structures in AFM imaging and therefore be applied to the cross–sections to determine the periodicity along the fibril. The highest frequency peak represents the structural feature with the highest spatial frequency. Making the assumption that the objects are touching each other this frequency corresponds to the centre–to–centre distance of the objects (peak–to–peak in the topography images) which is, in this case, the diameter of the beads. Therefore, the same cross–sections used for the diameter analysis along the fibrils have been analyzed using FFT analysis to reveal the periodicity of the beads and determine the average length of the protein molecules along the fibrils. Using NOVA AFM software the average periodicity was calculated and was found to be in good agreement with the cross–sectional analysis from above. A larger error arises when compared to the cross–sectional analysis due to the fact that the highest frequency peak could in some cases not be identified clearly. In the example given in figure

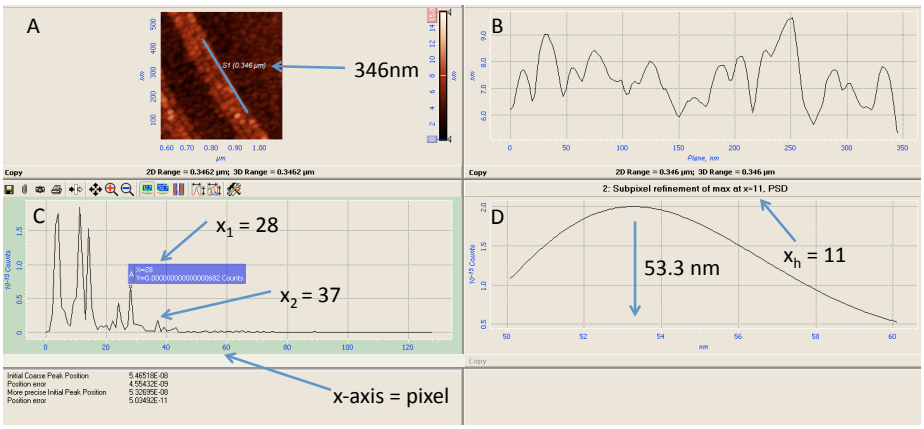


Figure II.13: Example of FFT analysis obtained from NOVA AFM. The cross–section data (B) obtained from the topographic image (A) is Fourier transformed to obtain the highest frequency peak which corresponds to the smallest periodicity of the structures (C). In this example the highest frequency peak is not clearly identified (x_1 and x_2 in C). The conversion of the x–axis in (C) from pixel into a length is performed for the highest peak at $x = 11$ (D).

II.13 the highest frequency peak (x_2) is of very low y–value and might well be counted as noise. Using just the spectra with clearly identifiable highest frequency peaks (3 out of 6) yields an average value for the bead diameters in close agreement with the cross–sections, which shows the validity of this approach.

The periodicity along the thin fibrils (chapter IV, figure IV.6) is more difficult to obtain compared to the larger fibrils due to the sizes of the fibril

approaching the tip–size used for the high resolution images (5 nm). However the periodicity is visible in parts of the fibril and can be roughly estimated form the topography image ((A) in figure II.14) using the cross–sectional tool (figure II.14 (B)).

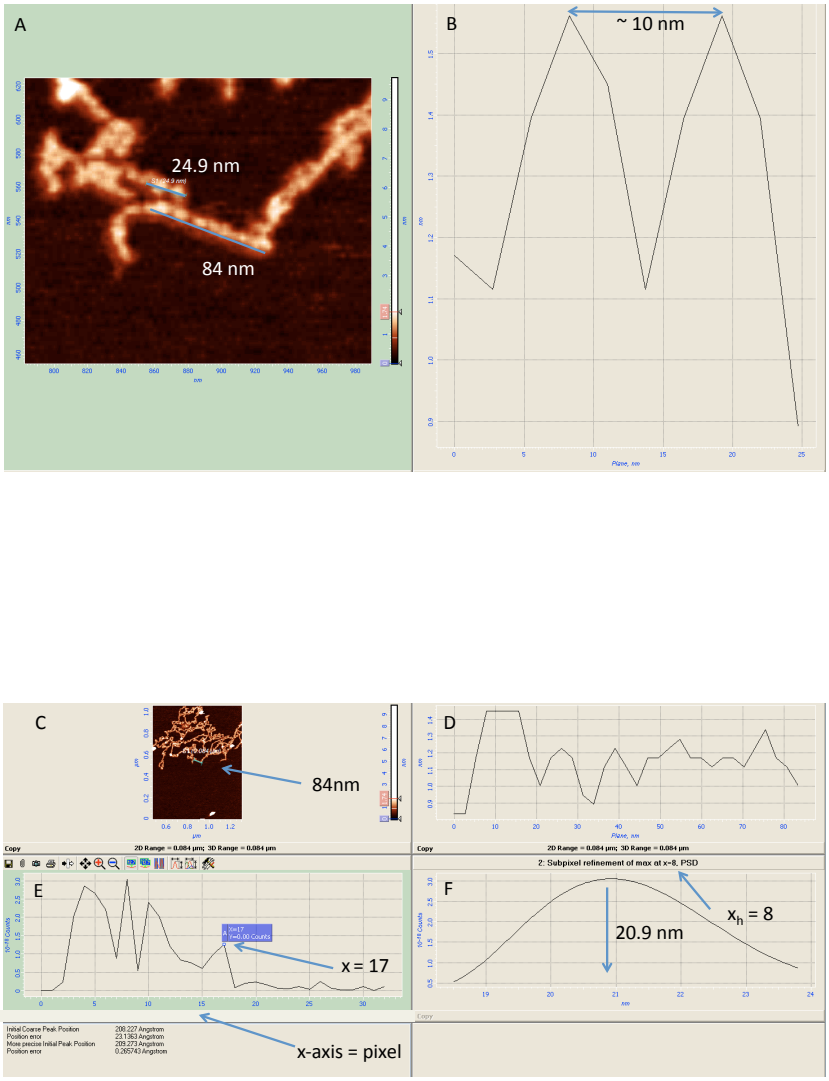


Figure II.14: The periodicity along the thin fibrils is clearly visible in the topography image (A). The cross–section of a short high contrast region reveals a peak–to–peak distance of about 10 nm. Using the FFT analysis of the larger cross–section (indicated by second line in (A)) gives a periodicity of 9.8 nm which is in good agreement with the cross–section in (A).

II.3 Sample preparation

A principal reason why precursor native silk has not been studied more widely is the accessibility and difficulty of handling such materials. Research therefore has been focused either on the already converted native silk fibre (being mostly *Bombyx mori* due to availability again) or on so called regenerated silk proteins. The regenerated silk is made by dissolving *Bombyx mori* silk fibres in LiBr and there are various different types of preparation methods. There are mainly 3 types of silk samples from either native or regenerated sources: aqueous silk gel, silk fibres and silk films. The silk films can be cast from a stock solution of native or regenerated silk proteins. By gently drying the proteins stay in their unconverted structure similar to the solution and the resulting film is an ideal sample to study the dynamics and hydration effects of the proteins.

II.3.1 Preparation of native *Bombyx mori* fibroin solutions

Due to the quantity of sample needed, the neutron experiments focussed on silkworm silk rather than spider silk. It is important to note that due to the short stability of the solutions, the dissection and solutions have to be prepared within 24h of starting the experiment. A working protocol was established to study silk proteins in a wide and physiologically relevant concentration range without aggregation.

The precursor silk, protein gel, has been extracted from the glands (see

figure II.15) of the *Bombyx mori* silkworms, which were in the 5th instar. The glands were dissected into petri dishes filled with milliq water. The epithelium was punctured at the turns between the posterior, middle and anterior sections using fine tweezers directly after dissection and the gel like fibroin was allowed to ooze out of the glands. After approximately 10 minutes the epithelium was gently removed with tweezers under a microscope and care was taken not to shear the fibroin 'dope'. During the whole time the gland was totally immersed in milliq water to prevent the dope from drying out. The glands were washed in distilled water (to remove the sericin coating) every 15 minutes 3 times. In between washings the petri dish was kept on ice as this helped prevent aggregation. Finally the fibroin dope was carefully retrieved from the petri dish using curved tweezers, gently blotted and transferred into 2 mL eppendorf tubes. The eppendorf tubes were filled completely with milliq water so that little or no air bubbles were left. Samples with larger amounts of air in the eppendorf tubes appeared less stable over time, which means that they showed floculation compared to completely filled eppendorfs. The fibroin dope was left to dissolve in the horizontally stored eppendorf tubes in the fridge at 6 °C for 24h to obtain a stock solution. The obtained stock solutions were then transferred into new eppendorf tubes so that non–dissolved parts remained in the old tube. The concentrations have been estimated from the dry weight of the protein and vary between 20 – 40 mg/mL depending on the gland size. The samples were either diluted first, or directly loaded in 1 mm (or if dissolved in D_2O in 2 mm) quartz cells for SANS experiment. For the SAXS experiments the samples were loaded into the well–plates. All samples were prepared and measured

at room temperature (controlled 20 °C in SAXS and SANS setups).

Figure II.15: Dissected glands of *Bombyx Mori*.

A visual inspection and in some cases CD measurements of the samples were made prior to loading into the sample cells. Any gelled or opaque samples were rejected from further measurements. Even samples which were handled exactly the same way may show gelation when others did not. It is well known that these samples are particularly shear sensitive. Therefore, great care had to be taken when loading the samples into the cells. Data sets which showed steep increase at low q (aggregation) or showed visual signs of aggregation after the experiment were not used in the data analysis.

II.3.1.1 Effect of D_2O

Dissolving the native silk proteins in D_2O appeared to enhance the aggregation of the solution especially at higher concentrations. The glands seem to dissolve much easier in H_2O than in D_2O. The fibroin in D_2O phase separated into a gelled part and a low viscosity fluid (both initially transparent). Little is known about this effect although D_2O is used for studying silk proteins in very dilute solutions [70]. The apparent 'pH' of D_2O is slightly higher than

that of milliQ water. Due to the sensitivity of silk to pH changes this might cause immediate aggregation of the protein, although aggregation occurred even in buffered protein solutions at a pH of 6. Despite being chemically similar to H_2O, D_2O has an additional proton, which might cause a change in the local environment of the protein and therefore cause chaotropic aggregation. Despite the difficulties outlined above, a match point plot was obtained from a series of SF proteins in D_2O and H_2O. The series was prepared by mixing a H_2O stock solution and a D_2O stock solution of the protein by weight to obtain the right ratio. The 'true' amount of D_2O was estimated using the dryweight concentration of the solutions since the protein is always hydrated and has $\approx$ 20 wt.% (H_2O) associated to the protein in the gland. The matchpoint of native silk was extrapolated (see figure II.16) to $\approx 27 \pm 7$ % which is low given the stated matchpoint region for proteins being between 35–45% D_2O [64]. Interestingly the RSF solutions appeared less affected by D_2O.

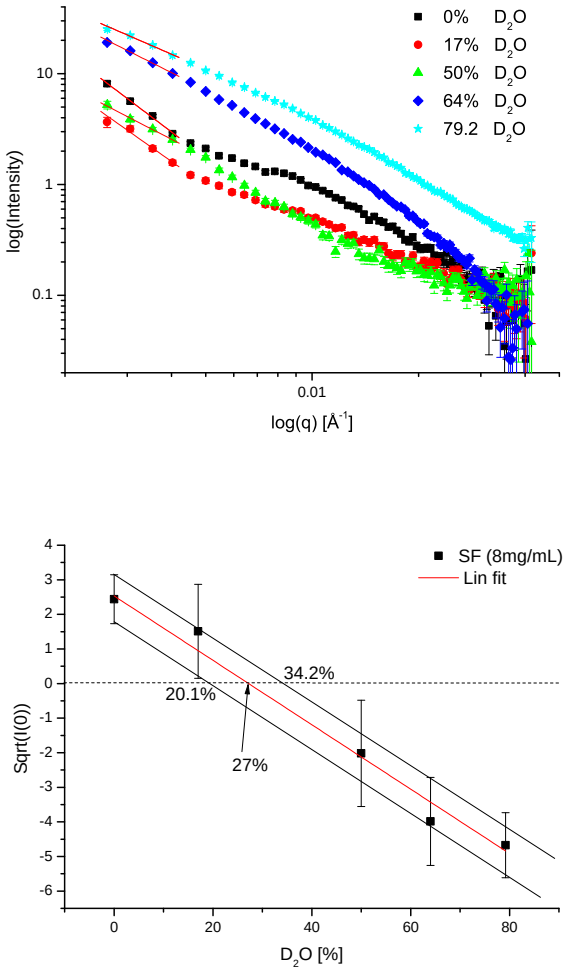


Figure II.16: Plot showing the match point of silk fibroin. The samples showed visual signs of aggregation in the D_2O shortly after preparation and the H_2O sample had flocculation after the experiment. The scattering curves of the series indicate aggregation due to the observed increase at low q (upper image). Extrapolating to zero angles by a linear fit, the I_0 indicates the D_2O sample with the lowest contrast. This so called match point is obtained by plotting the I_0 intensities as a function of D_2O concentration. The prefix of the three highest intensities have been switched to obtain the match point at zero contrast by linear fitting (black lines indicate error band). Standard deviations of linear fits have been weight due to lower scattering intensity close to the match point.

III Native and Reconstituted Silkworm Silk

III.1 Introduction

As mentioned in chapter I.1 substantial progress has been made into understanding silk's fibre morphology [36, 107, 131] and linking this to the mechanical properties [127, 100, 75, 112]. Surprisingly, comparatively less in–depth research has focused on the physical chemistry of the precursor material i.e. the raw un–spun silk, as it is stored in the gland. Indeed, due the difficulties in handling these high molecular weight proteins without aggregation, most of the experiments on silk protein folding and assembly have been performed on so–called reconstituted or regenerated silk fibroin (RSF) [67, 137, 87]. Fundamental differences however between the flow behaviour of native silk fibroin (SF) solutions extracted from the silkworm gland and typical RSF solutions have been revealed [50]. Links to structures, shape, clustering and gel network were hypothesized but yet remain to be demonstrated. Here the limitations and pitfalls associated with the study of reconstituted silk proteins are addressed and fundamental differences between native and reconstituted silks are revealed. A novel sample preparation protocol for the native silks was established (chapter II.3) and validated using small angle scattering (SANS/SAXS) to compare the shape and structure relationships

of native and reconstituted silks and the effect of protein concentration.

Typically the preparation of RSF starts with the degumming of silkworm silk fibres in a soap solution at 70 °C to remove the sericin coating. This is followed by the reconstitution or regeneration process during which the clean fibres are dissolved in the presence of a chaotropic agent, such as lithium bromide (LiBr) at temperatures ranging from 60–100°C. It is remarkable that in spite of a widespread use of the above recipe the effect of LiBr on the silk molecules is not fully understood. Various studies found, for example, a reduction in molecular weight [139] but no measurable changes of behaviour in solution [67]. Additionally, the preparation methods for reconstituted silk vary between research groups [9, 86, 138] and often the exact method of preparation is not clearly mentioned in the published studies [57, 9, 88, 138]. This has led to the publication of a wide range of shapes and sizes to describe silk proteins in solution and consequently, a highly divergent interpretation of how silk proteins fold and assemble into fibres [87, 124, 67, 72, 30, 61]. Here the focus is on the LiBr route as it is the most popular method, but one should note that other approaches exist with emphasis on milder reconstitution processes [130].

Small angle scattering has proven ideal to characterize the overall structure, shape and folding state of proteins and biopolymers in solution [119, 73, 102]. Within this context a combination of CD and SANS/SAXS is presented in order to evaluate the role and interplay of structures and morphologies of silk proteins behaviour in solution. This work demonstrates the suitability of small angle scattering as an important tool in the unravelling of silk shape and structure relationships.

III.2 Experimental methods

Here a brief description of the sample preparation methods used for this chapter is given. A detailed discussion is found in chapter II section II.3.

Native silk fibroin solutions (SF): SF protein solutions were prepared from silk proteins (fibroin) extracted from the glands of 5th instar *Bombyx mori* silkworms as described in chapter II.3. The silk fibroin solution, which was gel–like, was allowed to gently ooze out of the glands, washed in distilled water (to remove the sericin coating) and gently blotted before being diluted in deionized water (milliQ water) or D_2O at 6°C for 24 h. The obtained stock solutions had protein concentrations varying between 20 – 40 mg/ mL as calculated from the mass of dried protein within an aliquot of stock solution dried under vacuum at 80°C for one hour. A concentration series of native SF was prepared by further dilution of the stock solution in either H_2O or D_2O. D_2O however induced gelation of the native silk solutions but appeared to have no visual effect on RSF (for discussion see chapter II.3). Therefore it was decided to study silk proteins upon dilution with H_2O only. The dry weight concentrations for the native silks mentioned in [38] were the concentrations obtained from the retrieved samples after the experiment. Given the large amount of sample needed for the SANS experiments it was not practical to keep an aliquot of each dilution to determine a more precise dry weight. The calculation of the M_w however appeared reasonable, therefore the dry weight concentration was not critically examined. The M_w calculation in Greving et al [38] was based on an underestimated contrast value of 1.77 x 10^{10}cm^{-2} instead of 2.33 x 10^{10}cm^{-2}. A comparison of the

M_w obtained from SAXS and SANS at equivalent concentrations showed a difference of about 100 kDa. Careful appraisal of the raw data, reduction protocols and reduced data suggested no obvious origin for the discrepancy in M_w between SANS and SAXS. This left the sample concentration and its determination as the main source of error. Pooling the sample concentration stock solutions (dry weight measurement) and comparing them to the concentration of sample in cuvettes; it was found that concentration of the sample in cuvettes were systematically 30% higher than in the stock solutions. The cuvettes used are open neck design without PTFE stoppers. Although the cuvette cells were sealed with parafilm the discrepancy might be explained through evaporation, since the samples were often left in the cells for up to 24h after preparation before they were retrieved from the cells for dry weight calculation. Indeed some sample loss was observed. Additionally gelation of the high concentration samples occured when the samples were left for more than 24h which made a complete recovery often difficult. The low angle SAS intensity was however always measured within 2–3 hours after the preparation. Reducing the obtained dry weight concentrations from the retrieved samples by 30% and recalculating the M_w using 2.33 x 10^{10}cm^{-2} leads to M_w values close to the obtained values from SAXS (see M_w comparison inset in figure III.5) .

Reconstituted silk solutions (RSF): For the standard reconstituted silk fibroin (RSF) *Bombyx mori* silk fibres were degummed for two hours in a 0.5 w% sodium carbonate solution (Na_2CO_3) at 70 °C. Then the dried degummed silk fibres were dissolved in 9M LiBr solution at 70 °C. This protocol is widely used in silk research [50, 139]. To obtain a better under-

standing of this method, 6 different fibre to LiBr ratios were prepared by dissolving 3.5 g, 3 g, 2.5 g, 2 g, 1 g and 0.3 g, respectively, of degummed *Bombyx mori* in 10 mL 9M LiBr until they were totally dissolved (3 minutes plus additional 3 minutes to allow the solution to cool down). Each of the obtained RSF solutions were dialysed against milliQ water at room temperature to remove the LiBr (2 l per dialysis tube, 4 changes of water, 3 times 2 hours and 10 h overnight). From the obtained stock solutions the respective concentration series were prepared by dilution in milliQ water and the dry weights were estimated from an aliquot of each prepared dilution. The protein concentration was estimated from the dry weight method and ranged from 16 mg/ mL for the 1 g RSF stock solution up to 130 mg/ mL for the 3.5 g RSF stock solution.

Small angle neutron scattering (SANS): All the samples were carefully loaded into quartz cells (Hellma, QS–type) of 1 mm path length and measured at 20°C the day after the preparation of the stock solutions. A minimum equilibration time of one hour was left before loading and measuring the prepared solutions. The SANS experiments were conducted at the small angle neutron diffractometer D22 at the ILL in Grenoble (France) at a wavelength of 6 Å. For each native concentration two sample–detector distances of 17 m and 2 m were used and combined to cover a wide q–range. In the case of the lowest concentration (SF 2.5 mg/ mL) a sample–detector distance of 10 m instead of 17 m was chosen with a shortened collimation length, in order to increase the incident flux without significantly reducing the angular resolution. The RSF data were additionally measured at 4 m to strengthen the data overlap in the q–range of interest. A buffer sample of H_2O was

measured under identical conditions. All the data were reduced to absolute intensities using the data reduction software GRASP [16] (see also chapter II.1.4). The background subtraction was achieved by adjusting its scattering intensity relative to the difference in transmission observed with the sample. This prevented any over– or under–subtraction of the scattering intensity curves.

Small angle x–ray scattering (SAXS): The samples were measured at the bio–SAXS beamline ID14–3 at the ESRF (Grenoble) using the automated sample changer system based on an HPLC auto sampler [96]. The samples were pipetted into a 96 well plate before sealing with crystal clear tape. Each sample was loaded into the quartz capillary (1 mm diameter) at the lowest flow rate ($0.5\mu m$/s) and a background sample (e.g. H_2O) was measured prior to and after the sample. Between samples, the capillary, syringe and connecting pipe work were flushed with 12M urea and rinsed with water. Data were collected at a fixed sample detector distance of 2.43 m, wavelength 0.93 Å . Exposure times of 5 seconds maximum were chosen in the flowing system and up to 50 frames were taken per sample of 30 μL. Clear evidence of radiation damage was found after exposure times exceeding 5 seconds for a static position. Under normal measurement conditions, this was avoided by continuing to flow the sample through the capillary during the collection of x–ray data. Data were reduced using the data reduction pipeline from ID14–3 [96] and background subtraction was performed using PRIMUS software package [74] (see chapter II.1.4 for details).

Data analysis: Details of the data analysis for SAS is found in chapter II.1.5. The radius of gyration R_g was calculated using the Guinier approach

assuming a spherical geometry and an isotropic sample. From the obtained linear curve the slope was fitted within the Guinier region for solution scattering ($q \cdot R_g < 1.3$) [102]. The analysis was performed using the PRIMUS software package [74]. Further discussion on the validity of the calculation is provided in the results and discussion sections. The SAXS data were normalized for protein concentration and for the SANS–SAXS comparison (insets in figure III.8) scaled to overlap the SANS data (figure III.2).

Electrophoresis gel: An SDS–PAGE was run for the samples previously measured by SANS: the native silk sample SF (at 22 mg/ mL) and an aliquot of each of the different RSF samples at a concentration of 16 mg/ mL. 10 μL aliquot were mixed with 10 μL of llaemli (SDS/ mercaptoethanol/ stain/ molecular weight ladder, Precision Plus Protein Standards, All Blue from Biorad, catalog 161–0373) and loaded in the gel. Note that the solutions were not temperature denatured to prevent any further dissociation.

Circular dichroism (CD): Selected solutions from the SANS concentration series of SF and RSF were measured using a Jasco 810 circular dichroism spectrophotometer, the day after the SANS experiment. To achieve suitable spectra collection the samples were measured in quartz cells (Hellma) of either 0.01 mm or 0.1 mm pathlength depending on the protein concentration. Typically, three consecutive spectra were recorded and averaged. A resolution of 1 nm step size and a scan rate of 200 nm/ min were choosen. The requirement for spectroscopic transparency did not allow the full concentration range to be measured, instead only key concentrations were possible (see section III.3). The buffer (H_2O) was also collected under identical conditions. The data were reduced and pre–processed using the CDTool software pack-

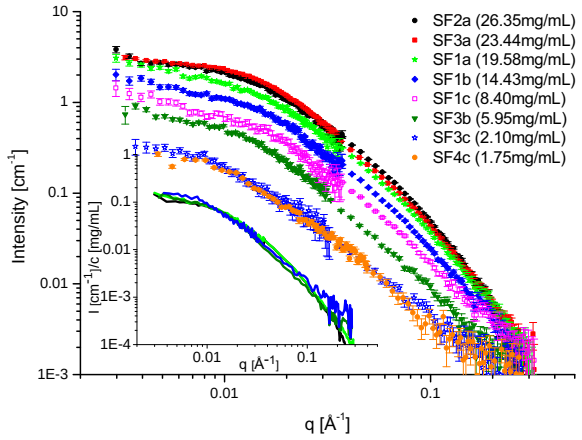
age [81]. Complementary measurements on similar samples were performed at the synchrotron radiation circular dichroism (SRCD) CD1 beamline at Aarhus University, Denmark [20]. The data collection was identical to the afore mentioned conditions. Note that the units are given in mdeg simply because of the concentration determination method.

III.3 Results and discussion

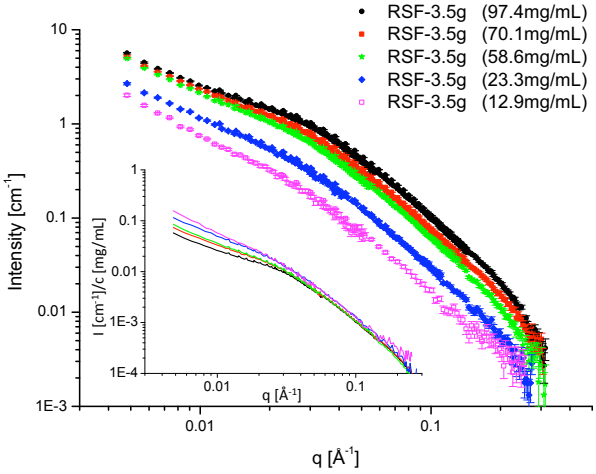
Figure III.1 compares the small angle neutron scattering data of native (SF, panel (a)) versus reconstituted (RSF, panel (b)) silk fibroin. The inset in each panel shows the data after normalization for the fibroin (silk protein) concentration. The compared data sets in figure III.1 reveal distinctive differences in the shape of the scattering curves. Native silk (SF) shows an extended flat region at low q whereas reconstituted silk (RSF) shows a significant increase in intensity in the same region. While the native silk fibroin (SF) starts to monotonically tail off at $q > 0.015$ Å^{-1}, the RSF shows instead a marked broad peak between 0.03 and 0.04 Å^{-1}.

III.3.1 Concentration effects

In order to distinguish between the form factor $P(q)$ and the structure factor $S(q)$ (see chapter II.1) a concentration series was measured. Decreasing the solution concentration will minimize the inter–particle interactions (such that $S(q)$ becomes unity) until only the form factor $P(q)$ remains. Good consistency was achieved in the overall shape and intensity of the scattering curves of native silk fibroin (SF) over a number of samples prepared in the same



(a)



(b)

Figure III.1: The SANS scattering curves of native (SF, panel (a), numbers represent different stock solutions) and reconstituted silk fibroin (RSF, panel (b) for 3.5 g of RSF dissolved in LiBr). The silk protein concentration is given in brackets. In contrast to the SF, which shows a flat region at low q, the RSF has a steep slope at low q, which may be attributed to large aggregates. In the insets the SANS curves are scaled for silk protein concentration. SF (a) shows good overlap for the different concentrations, whereas RSF (b) shows a strong departure at lower q with changing concentrations.

manner but from different silkworms and for similar fibroin concentrations (figure III.1). Within the concentration range from 26 mg/ mL to 4 mg/ mL, the shape of the SF scattering curve remains constant (figure III.1 (a) inset), implying that the $S(q)$ is indeed unity for these native silk fibroin solutions and that the form factor $P(q)$ does not change. For low concentrations ($<$ 5 mg/ mL) however the SANS signal–to–noise ratio is very low, even after long data collection times ($>$ 3 h per sample). Therefore comparable SAXS measurement have been conducted at ID14–3 (ESRF). Although only 2 dilution series of native SF could be measured and one of them obtained from early instar worms a direct comparison of the SAXS data and the SANS data reveals a difference in the overall shape of the curves between high and low concentrations (figure III.2). Normalizing the intensity for protein concentrations suggests a change in shape of the native SF curve (see figure III.2) at very low protein concentrations ($<$ 5 mg/mL).

Examining the concentration series of RSF (figure III.1 (b)) prepared using 3.5 g of silk fibre dissolved in LiBr (see experimental methods III.2) reveals a rather different behaviour in comparison to the SF. It is immediately apparent that the scattering intensity increases at low q, which would arise from large–scale aggregates present in the solution. Moreover when the data is normalized for protein concentration the curves do not overlay and there is a shoulder present in the scattering curve with increasing protein concentration (figure III.1 (b) inset). In contrast to SF a large variation in the overall shape of the SANS data curves obtained from different RSF stock solutions is found. The only difference in the preparation of these RSF stock solutions is the amount of fibre dissolved in 10 mL 9M LiBr (fibre to LiBr

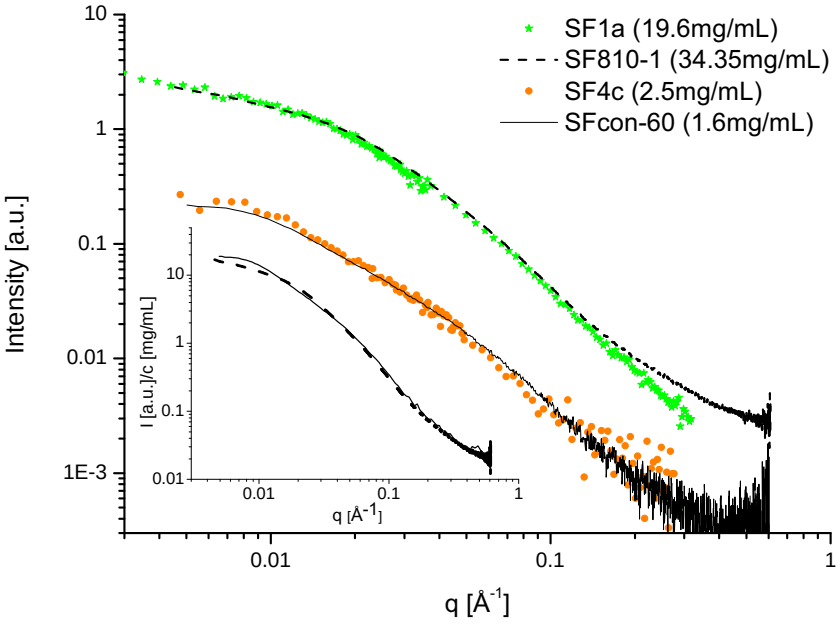
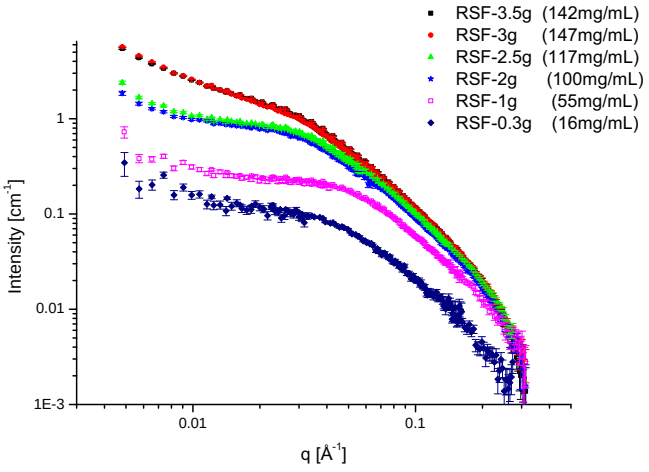


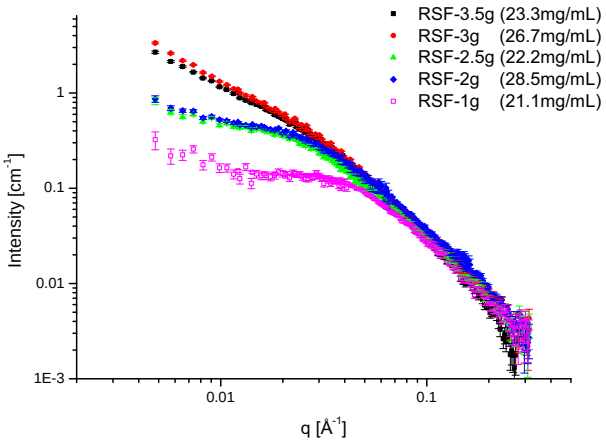
Figure III.2: SANS (symbols) and SAXS (lines) curves of native silk fibroin (SF) obtained from different silkworms (SAXS curves scaled to match SANS intensities). The overall shape is very similar, and the low concentration SAXS suggests a deviation in shape at low angles (inset). The deviation at high angles between the SANS (green stars) and SAXS (dashed line) is resulting from over subtraction in the SANS data set.

dissolution ratio) (figure III.3 (a)). High fibre to LiBr ratios (3.5 g and 3 g fibre per 10 mL 9M LiBr) show an increase in scattering intensity at low angle, this intensity being reduced for lower fibre to LiBr ratios (2.5 g and 2 g) becoming almost flat at low q for the lowest dissolution ratio used (1 g and 0.3 g) (figure III.3 (a)). This effect is the same even if the solutions are further diluted to a similar protein concentration of 20 mg/ mL (figure III.3 (b)) suggesting that the chaotropic LiBr is unable to dissolve as effectively the protein fibres if the fibre to LiBr ratio is too high and large sized aggregates remain in the RSF solution.

A complete dilution series was measured for each of the different RSF stock solutions. Since the differences between 3.5 g and 3 g as well as the 2.5 g and 2 g are small as already indicated from the stock solutions (figure III.3) the dilution series of 2.5 g and 1 g are presented in figure III.4. Comparing the dilution series of these lower fibre/ LiBr ratios ($<$ 2.5 g) (figure III.4) reveals that the presence of a structure factor at higher fibroin concentrations is also not negligible, even in the samples prepared from lower fibre/LiBr ratios ($<$ 2.5 g) where larger scale aggregates appear absent (figure III.4). However, at concentrations lower than 30 mg/ mL in these lower fibre/LiBr ratio samples the shape appears to be less influenced by the structure factor and these data sets can be analysed using the Guinier approach (see chapter II.1.5).

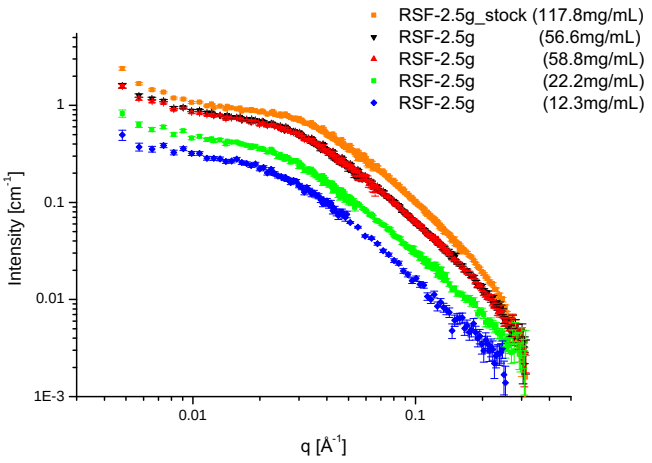


(a)

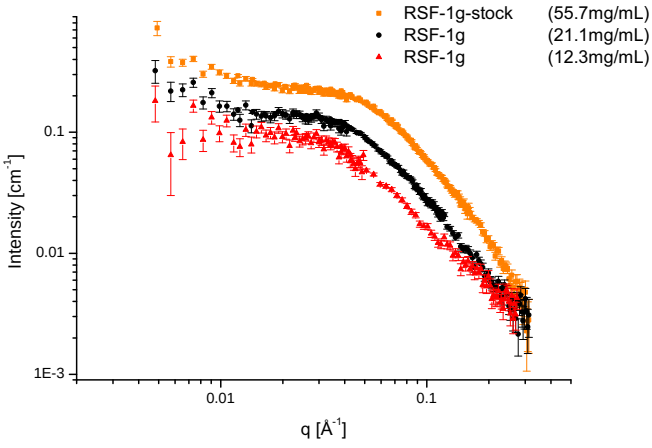


(b)

Figure III.3: Panel (a), SANS data of the RSF stock solutions prepared from different fibre/ LiBr ratios. An increase in intensity at low q, indicative of large sized aggregates, is seen for the higher fibre to LiBr ratio solutions of 3.5 g and 3 g fibre per 10 mL 9M LiBr, which actually overlay each other. Panel (b), dilution of these stock solutions to a similar protein concentration still shows an increase in intensity at low q for the higher fibre to LiBr ratio whilst the apparent correlation peak seems to move to lower q.



(a)



(b)

Figure III.4: At 2.5 g (a) and 2 g (data not shown), that is for low fibre to LiBr ratios, the increase in intensity at low q is much less pronounced than in the high fibre samples (3 g and 3.5 g of fibres per 10 mL LiBr (see figure III.3)). Additionally, the increase in intensity at low q is most pronounced for the high protein concentration samples (100 mg/ mL – 50 mg/ mL) (a) and seems to disappear at low protein concentrations (b). It is therefore attributed to a contribution from the structure factor.

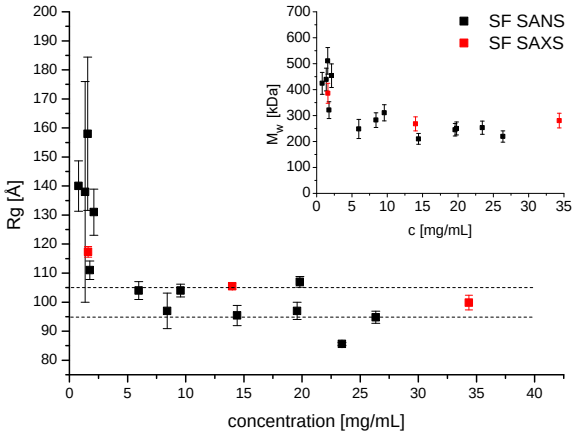
III.3.2 Guinier analysis

Figure III.5 shows the calculated R_g for native silk (SF, (a)) and reconstituted silk fibroin (RSF, (b)) as a function of protein concentration from the SANS data. For SF the obtained R_g values from complementary SAXS data are shown in red (figure III.5 (a)). At SF concentrations above 10 mg/ mL, the R_g is found in a range of 95 to 105 Å. The variation might be explained by the extended shape of the protein, allowing only a limited amount of points to be fitted within the Guinier region, which is not taken into account in the error calculation. The SF3a highest concentration sample has however a significantly lower R_g compared to the other SF dilutions which might be explained by the slight deviation in shape at around 0.2 $Å^{-1}$. For concentrations below 5 mg/ mL the trend suggests an increase of the R_g in good agreement with the SAXS data (red squares in figure III.5 (a)). The inset of figure III.5 (a) shows the molecular weight M_w calculated from the intensity at zero angles $I(0)$ obtained from Guinier analysis (see chapter II.1.5.2) for the SANS and SAXS data. Great care must be taken when determining the $I(0)/c$ for the SANS measurements due to the mentioned uncertainty in protein concentration (see III.2) and the low signal to noise ratio in particular for the low concentration samples. Taking together the obtained results from SANS and SAXS suggests that the M_w does not change significantly down to a concentration of 5 mg/mL (insert in figure III.5, (a)). An average value of 267.8 ± 52.4 kDa is obtained for the SANS M_w down to a concentration of 5 mg/mL (error from standard deviation). This value appears not unreasonable given stated literature value for the heavy chain fibroin of 350 kDa

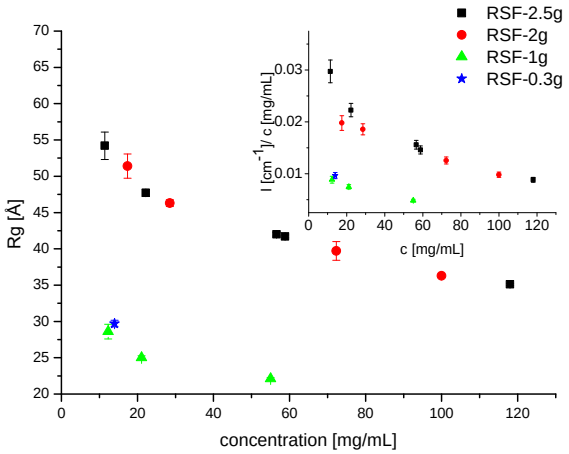
[134, 111] and is consistent with the results obtained from SAXS (inset in figure III.5 (a)). The apparent increase of the M_w calculated from the SANS data at concentrations below 5 mg/ mL might suggest protein–protein associations. This trend is again confirmed by the SAXS data as indicated by the red squares in figure III.5, (a).

In the case of the reconstituted silk fibroin (RSF) the Guinier approximation can only be applied to the low fibre to LiBr ratio samples (2.5 g and lower), as mentioned above. The R_g was calculated for the concentration series of RSF at 2.5 g, 2 g, 1 g and 0.3 g fibre/ LiBr ratio even though the validity for concentrations higher than 30 mg/ mL must be questioned due to a likely contribution of the structure factor.

The results show that the R_g for RSF is in general two to three times smaller when compared to the native silk fibroin. While the native silk shows an R_g between 95 and 110 Å the RSF values increase in a similar concentration range from 45 to 55 Å for the higher fibre/ LiBr ratios (2 g and 2.5 g). Importantly, a distinct behaviour depending on the fibre/ LiBr ratio was found. For the preparations using 1 g and 0.3 g fibre/ 10 mL LiBr, a much smaller R_g is calculated, which is also reflected in the $I(0)$ normalized by concentration which is proportional to the (M_w) (figure III.5 (b)). For the RSF prepared with more than 2 g fibre per 10ml LiBr larger R_g and higher values for $I(0)/c$ suggesting a stronger reduction in molecular weight (M_w) of the silk fibroin once the proportion of silk fibre to LiBr exceeds a certain ratio. Given the chemical strength of the chaotropic agent required to dissolve the strongly bonded silk proteins, and that a reduction of the molecular weight of fibroin has been shown in the literature [139] this would appear



(a)



(b)

Figure III.5: The R_g for native silk fibroin (SF) (a) (SANS black, SAXS red) and reconstituted silk fibroin (RSF) (b). Insets: (a) SF M_w's obtained from SANS (black) and SAXS (red) reveal a similar trend. (b) The intensities at zero angle $I(0)$ for RSF normalized by protein concentration (proportional to molecular weight, see equation II.43). Interestingly the R_g of the SF samples is in general significantly larger than that of the RSF. Additionally the SF appears to have a constant R_g over a wider range of concentrations while the RSF shows a clear increase in R_g over the range of concentrations investigated.

a reasonable assumption. An attempt to verify this was made by running gel electrophoresis of the same samples used in the SANS measurements directly after the experiment, gel electrophoresis being an established method to determine the molecular weight (M_w) distribution of proteins in solution.

III.3.3 Gel electrophoresis

An SDS–PAGE gel of aliquots of the samples measured by SANS is shown in figure III.6 revealing a smeared region from 250 to 50 kDa characteristic of the high polydispersity for the RSF samples. Consequently, the R_g calculated by small angle scattering for the RSF can only be interpreted as a size average weighted to the larger sizes. However, the bands for RSF 1 g and 0.3 g samples are sharper suggesting a lower polydispersity. The RSF 1 g fibre per 10 mL LiBr shows a weak band at very low M_w namely 15 kDa, in contrast to the higher fibre to LiBr ratio RSF samples, (figure III.6). Thus, when a greater amount of LiBr is used in the preparation, the M_w of the fibroin protein is greatly reduced. The bands at 25 kDa observed in the high fibre to LiBr ratio RSF and the native SF sample are most likely the light chain fibroin and the so called P25 which have published molecular weights of 25 kDa and 30 kDa, respectively [134, 111]. The reduced M_w for the RSF indicated by the low R_g and shown by the gel is reflected in the estimation of the molecular weight. The calculated M_w for the RSF 1 g yields 17.9 kDa, which is close to the 15 kDa shown by the gel (figure III.6). A larger M_w value calculated from the scattering compared to the gel is expected, since the higher scattering contribution of the larger structures to the obtained

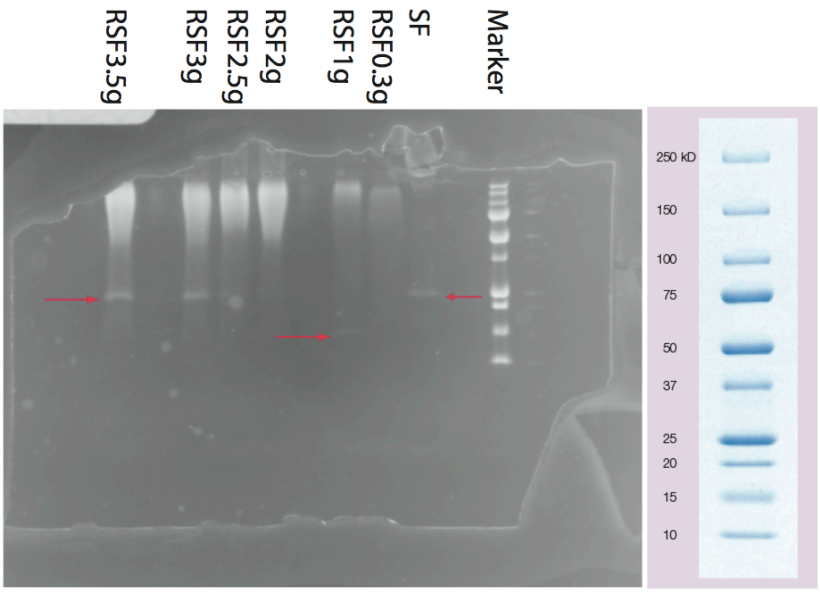


Figure III.6: After the SANS measurements the 16 mg/ mL RSF and the 22 mg/ mL SF samples were loaded into a SDS–PAGE gel. Ladder to the right gives the explicit molecular weight for the marker line in the gel. All the RSF samples penetrated into the gel and show a high polydispersity in M_w ranging from 250 kDa down to 50 kDa (smeared bands). The majority of the native silk sample (SF) did not penetrate into the gel which is expected for a M_w of the order of 350 kDa as cited in the literature [134, 111]. The band at 25 kDa shown clearly in the 3.5 and 3 g RSF and the SF samples is identified as the P25 and light–chain proteins which have molecular weights of 30 and 25 kDa [134, 111], respectively.

average $I(0)$. The 2.5 g and 2 g samples have molecular weights between 40 and 60 kDa, which is in reasonable agreement with the literature [130, 63]. In clear contrast to the RSF the native silk sample does not penetrate into the gel apart from the band at 25 kDa, which is expected since the published molecular weight of the heavy chain fibroin is 350 kDa; previous workers have shown a gel electrophoresis of native silk having two clear narrow bands between 350 kDa and 400 kDa, and a band at 25 kDa [130].

This work has clearly shown a difference in the small angle neutron scattering for native and reconstituted silk solutions. It is interesting to now determine whether these differences in the size of the scattering entities can be coupled to a difference in the structure of the protein in solution. Circular dichroism (CD) spectroscopy measures the difference in absorption of left–handed polarized light versus right–handed polarized light that arise from chirality. For this reason, CD is a highly sensitive tool for determining secondary and tertiary structures of proteins.

III.3.4 Circular dichroism of SF and RSF

Complementary CD spectra obtained from aliquots of the samples after the SANS measurements are shown in III.7 (a) and (b), using a conventional CD machine. Optical limitations from conventional instruments prevent access to the full spectral range especially for lower wavelengths (below 190 nm) and thus limits the range of measurable concentrations. In particular as the protein concentration increases, strong absorption makes it difficult to measure below 190–200 nm (figure III.7 (a), (b) $>$ 10 mg/ mL) therefore the

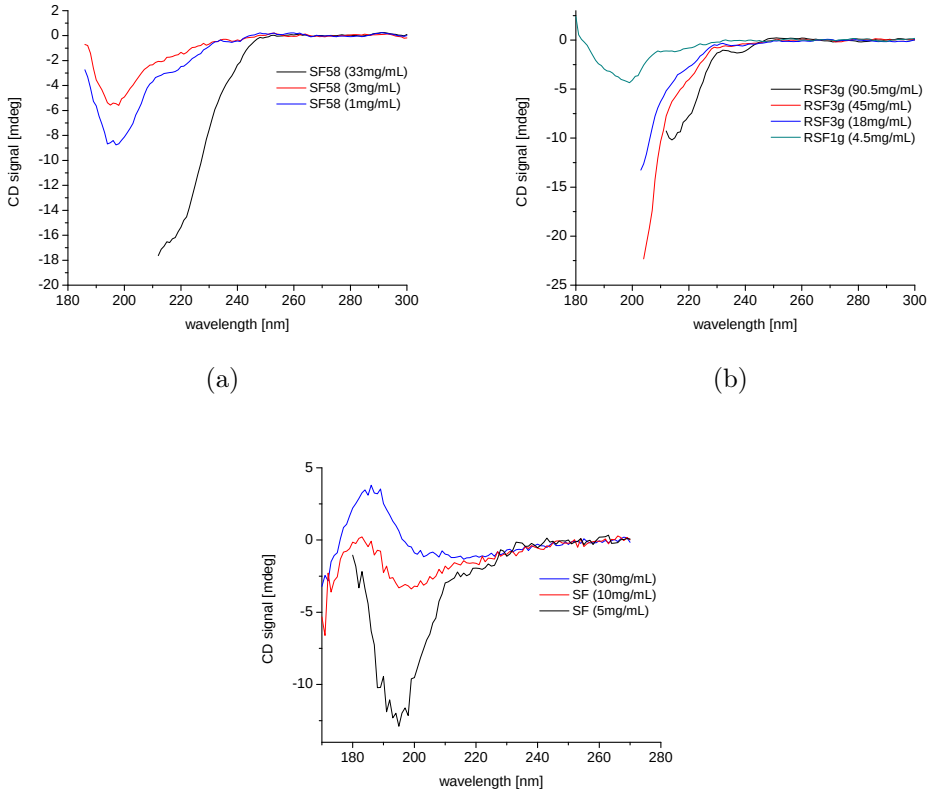


Figure III.7: CD spectra of SF (a), (c) and RSF (b). Panel (a), conventional CD instrument spectra for SF (sample no. 58) solutions of concentrations 33, 3 and 1 mg/ mL. Panel (b), RSF 3 g and 1 g spectra obtained using a conventional CD instrument, with concentrations of 90, 45 and 18 mg/ mL for RSF 3 g and 4.5 mg/ mL for RSF 1 g. Panel (c), SF spectra measured on synchrotron radiation instrument, for concentrations of 30, 10 and 5 mg/ mL. It appears that both native (SF) and reconstituted (RSF) proteins unfold with decreasing concentration (see text).

high concentration samples curves in III.7 (a) and (b) are not shown at lower wavelengths. Synchrotron radiation CD (SRCD) allows to access a broader spectral region of SF proteins as shown in figure III.7 panel (c).

Typically a helical refolded silk shows two minima at 220 and 204 nm of comparable intensity [20, 22]. When the concentration is decreased, below a critical concentration of 10 mg/ mL, the minimum at 204 nm is progressively shifted to 199 nm with a higher (more negative) intensity than the minimum at 220 nm. This structural rearrangement has been attributed to an increase of structural disorder in the silk proteins, [21, 22]. A structural unfolding associated with both sample groups is observed as the protein concentration of the solution is decreased, although this effect is less pronounced in RSF (Figure III.7 (b)).

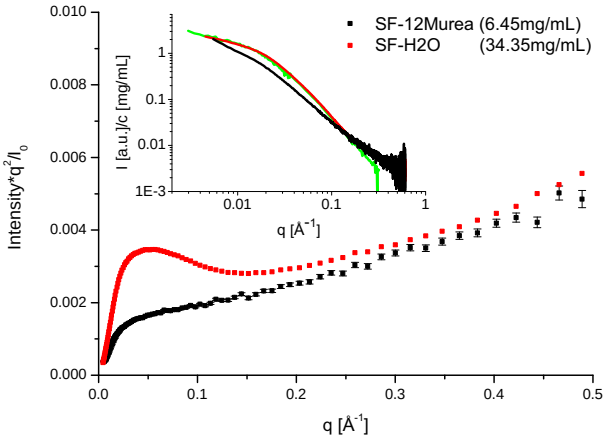
Taking the structural unfolding and the apparent increase in R_g together, the results suggest that the two events are linked and that the increase of R_g corresponds to a swelling of the silk structure (both for SF and RSF) with an unfolding of the silk protein chains.

III.3.5 Kratky analysis

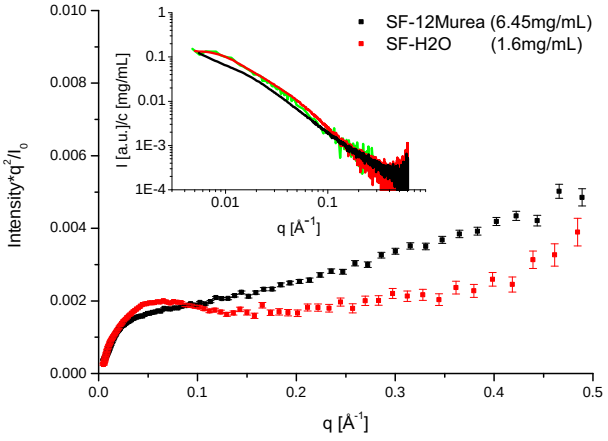
To further investigate the folding state of the native SF protein a comparison with the totally unfolded chain is needed. Urea is known to destroy hydrogen bonds and is therefore often used to study the unfolding of proteins [89, 31, 115]. Further to the observation of the potential shape change at low concentrations, urea was used to study the unfolded SF protein. The folding state of a protein can be studied using a Kratky plot (see chapter

II.1.5). For SANS of proteins in H_2O this plot is seldom used due to the high incoherent background arising from the solvent and the difficulties of background subtraction. Therefore the Kratky analysis focusses on the complementary SAXS data obtained at the ESRF. Figure III.8 shows the SAXS data of native SF in H_2O in comparison with a 12M urea sample. The insets in the graphs reveal a good overlap of the SANS and SAXS data of native SF in H_2O at similar protein concentration. At protein concentrations above 5 mg/mL the 12M urea sample shows a clear deviation from the shape of the scattering of the SF in H_2O at high and low angle (data normalized for protein concentration) III.8 inset (a). The Kratky plot reveals that the SF protein in 12M urea is indeed fully unfolded, while the SF in H_2O indicates partial folding (figure III.8). Reducing the protein concentration in the H_2O sample below a critical concentration of 5 mg/mL reveals that the shape of the scattering curve does look similar to the SF in 12M urea III.8 inset (b). The Kratky plot of the low concentration SF does not show a clear maximum, however a plateau region is evident (figure III.8 (b)) in contrast to the SF in 12M urea. The differences in the folding state of the higher concentration SF in H_2O compared to the low concentration sample might suggest a transition from a partially folded to a random coil state of the protein chain [90]. This loss of structure is also indicated by CD data and therefore supports the above hypothesis.

A comparison of SF and RSF at relatively high protein concentration in a Kratky plot suggests a major difference between the two silk types (figure III.9). The RSF sample lacks the maximum in the Kratky plot and suggests a more randomly oriented structure compared to the SF sample. However



(a)



(b)

Figure III.8: Kratky plots of SAXS data from native SF in H_2O (red line) and 12M urea (black line) at concentrations above (a) and below (b) 5 mg/mL. The inset figures reveal a good overlap of the SANS (green; (a) SF3b 5.95 mg/mL and (b) SF4c 1.75 mg/mL from figure III.1) and SAXS (red) data of SF in H_2O. At higher SF concentration (a) the deviation between the SAXS curve of SF in H_2O and SF in urea (inset (a)) becomes evident. The Kratky plot (a) reveals that the H_2O sample at 8mg/ml is partially folded in comparison to the totally unfolded protein in 12M urea. Below the critical concentration of 5 mg/mL the scattering curves of the H_2O and the urea sample show a more similar shape (inset in (b)) but the Kratky plot of the SF in H_2O reveals the presence of a plateau region.

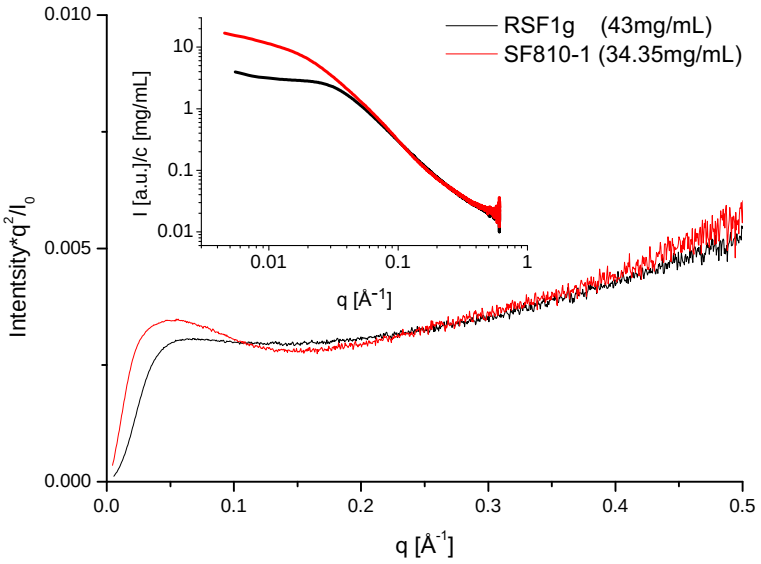


Figure III.9: The Kratky plot of an SF SAXS sample (red curve) in comparison with an RSF sample (black curve) indicating a difference in the folding states of the two silks. While the SF sample suggests a partially folded state, the RSF lacks a maximum and shows a plateau region.

care must be taken since the absence of the peak in the RSF sample might be caused by the high polydispersity revealed from the gel (figre III.6).

III.3.6 Native silks equilibrium conformation

In this study native silk proteins with a protein concentration range from 1.5 mg/mL to 35 mg/mL were investigated. No clear structure factor was observed at the higher protein concentrations measured. Indeed the concentration normalized curves above 5 mg/ml overlay (figure III.1). Typically for globular proteins an effect of inter–particle interactions is observed at concentrations as low as 1 to 10 mg/mL [102]. It is therefore interesting

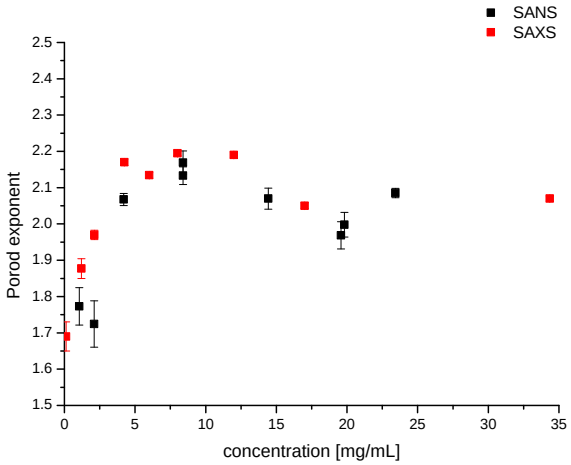
to hypothesize why these silk proteins appear to show no structure factor. It should be remembered that the silk fibroin and accompanying chaperone and glyco proteins, sugars etc, [111] which make up silk dope within a gland, are expressed and stored for significant periods of time without spontaneous aggregation. Rheology studies [50, 53], as well as AFM studies (chapter IV) have shown evidence of molecular alignment prior to fibre formation. Together this gives a strong indication of a non–interacting system at higher protein concentrations. This would imply that there are repulsive interactions present in the system preventing an inter–chain interaction and therefore an uncontrolled aggregation. Based on the charge distribution along the chain a charged micellar model has been suggested by Foo et al [30]. In a charged system with repulsive interactions however one would expect to see a clear structure factor superimposed on the form factor. This would lead to a decrease in $I(q)$ at low q at higher concentration which is not observed. Additionally the intensity normalized for protein concentration ($I(0)/c$) appears constant down to 5 mg/mL and does not increase progressively as one would expect upon suppression of a form factor. This suggests that the calculated M_w of the scattering entity in the small angle data at these concentrations is comprised principally of an individual silk chain. In this case inherent stability of the proteins maybe imparted if the individual proteins do not have strong interactions with neighboring proteins, such as would lead to a structure factor.

Another explanation for the absence of a structure factor however could be that the native silk protein is not in an equilibrium state upon dilution. This would also explain the observed variation in R_g in the concentration

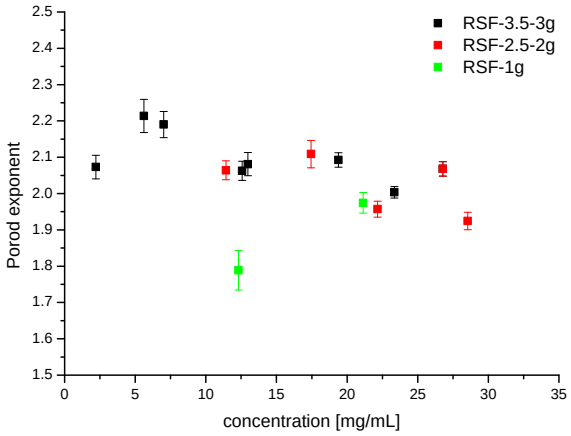
range of 35 down to 5 mg/mL. The present structure factor might in this case be overlaid by the different form factors of the SF proteins.

At concentrations below 5 mg/mL however, the apparent R_g and M_w increase and the Porod exponent reduces approaching values similar to those of swollen coils (1.7) (figure III.10 (a)). The R_g increase, the decrease in Porod exponent and the Kratky analysis suggest that this change is due to a loss of structure, which is supported by the CD data observations. The apparent increase in the M_w observed in these dilutions might indicate protein–protein interactions. An increased aggregation propensity upon dilution in H_2O was also reported for very dilute native SF samples by [122]. In the here presented work the dilution series has been prepared using H_2O rather than a charge balanced buffer since evidence from our laboratory showed that dissection and dilution into a 'buffer' solution induced visual aggregation. A potential explanation for these observations might be that at very low concentrations the silk structure might no longer be stabilized by charge, which could lead to swelling and loss of structure of the protein chain. This loss of structure might enable the chains to interact and leads to the observed increase in molecular weight.

In RSF however, where the reconstitution process has broken up the protein chain by molecular weight a different mechanism would be expected. Indeed a progressive increase in R_g and M_w in a comparable concentration range to the SF was observed. Additionally a comparison of the Porod exponent obtained from the RSF dilutions suggests no clear change in a similar concentration range compared to SF (figure III.10 (b)). Although the CD data indicate a loss of structure upon dilution this effect is less evident com-



(a)



(b)

Figure III.10: Comparison of Porod exponent as a function of protein concentration in SF (a) and RSF (b). A linear fit over the region from 0.03 – 0.1 shows a decrease in the slope at concentrations below 5 mg/mL. This also supports the hypothesis of a loss in structure upon dilution. Fitting the high q part for the different RSF dilutions reveals no reduction in the porod exponent with concentration (b).

pared to native silk (figure III.7). It has to be kept in mind however, that even in concentrations below 30 mg/mL the structure factor is not negligible in RSF, which might also affect the slope in the chosen fitting region for the Porod exponent. Additionally the change in the Porod exponent might be masked by the observed high polydispersity of the RSF solutions. Anyhow given the preparation method used to obtain the RSF solutions it seems not unlikely that the potential ability of the chains to prevent or enable inter–chain interaction might be lost. This hypothesis is also supported by rheological studies [50, 53], which showed a clear difference in the flow properties between native silks and the standard reconstituted silk.

III.3.7 Summary

The results show the validity of using small angle scattering to determine the sizes of aggregates within silk fibroin solutions, of both native and reconstituted silks. Neutron scattering has proven particularly suited to this, since this technique is known to be non–destructive in contrast to x–ray scattering. With small angle x–ray scattering, radiation induced changes would be observed in these solutions at room temperature for longer than 5 seconds exposure (see Experimental methods III.2). Additionally since the SAXS beamline ID14-3 has a flowing sample changer system to reduce the radiation damage and allow longer data collections [96] particular care had to be taken due to the shear sensitivity of silk proteins [50]. Since one express aim of this study was to examine the native fibroin solutions at near *in vivo* conditions, SANS seemed more suited for these solutions, in particular to

check for radiation and flow induced changes. One disadvantage in neutron scattering however are the relatively long collection times needed to achieve reasonable signal to noise ratios. Also the determination of the background at high angles becomes difficult due to the incoherent contribution of the solvent if measured in H_2O. With SANS, one would typically use the difference in the scattering length density between hydrogenated and deuterated solvents, e.g. H_2O or D_2O, to give further information about the nature of the scattering entities in the solution. In solution SANS it is therefore a wide spread approach to use D_2O as a solvent for the proteins [119, 84]. Interestingly the silk protein appeared visually stable in H_2O for up to 3 days, but showed increased aggregation propensity when dissolved in D_2O (see chapter II.3.1.1 for more information). Given the nature of silk proteins, being large proteins in a hydrogenated gel state it is next to impossible to know the exact amount of hydrogen or deuterium in the system. Therefore in that case no additional insight would be derived from using D_2O, however, the possibility of using selective deuteration of, for example the protein is being investigated. It will also be of interest whether the shape and nature of the aggregates observed in fibroin solutions also varies with the different preparation methods, this forming the basis for on–going studies.

III.4 Conclusions

The here presented, novel small angle neutron scattering data reveal significant differences between the molecular weights and sizes of (i) native silk (SF) fibroin extracted from *Bombyx mori* silkworm glands and (ii) reconsti-

tuted silk (RSF) fibroin prepared using the LiBr dissolution method. RSF solutions contain large–scale aggregates unless the ratio of fibre to LiBr used in the preparation is chosen low. However, when the ratio of fibre to LiBr is lower than 2 g/ 10 mL 9M LiBr, a marked reduction in the molecular weight of the fibroin protein was observed as measured by electrophoresis gels, albeit with a very high polydispersity. Accompanying the change in M_w is a change in the radius of gyration from 56 Å to 26 Å. In contrast, the SF samples do not show the presence of any large sized aggregates and display a nominal radius of gyration of 100 $\pm$5 Å across the measured concentration range.

Complementary circular dichroism measurements suggest that a progressive unfolding of silk proteins towards a more disordered structures is associated with the change in R_g as the concentration of silk protein in solution is reduced. Although similar structural unfolding is observed in SF and RSF, the latter is more prone to form large aggregates. The Kratky analysis on native SF in H_2O and 12M urea reveals that the silk protein is not totally unfolded at low protein concentrations in H_2O but shows a typical Kratky plot of a fully unfolded chain when measured in 12M urea. The difference in the Kratky plot between a higher and a lower protein concentration of SF in H_2O suggests the transition from a partial folded state to a random coil state which is supported by the observations from CD as well as a reduction in the Porod exponent. A comparison of RSF and SF in a Kratky plot indicates that the RSF tends more towards a Gaussian coil characteristic while a comparable SF sample shows features of a partially folded protein. This seems also be supported by the Porod analysis of RSF, where no clear change in the Porod exponent with concentration is observed.

In summary, the work presented in this chapter demonstrates that there are significant differences between the sizes as well as the folding state, and thus probably the shapes, of reconstituted and native silk fibroin proteins in solution. This must call into question the validity of using reconstituted silk fibroin to interpret the behavior of native silk proteins. Furthermore, this conclusion might help explain why most of the results published on reconstituted silk vary significantly from one another concerning absolute numbers, sizes and shapes of the protein molecules [50].

IV Single Molecule and Shear–induced Self–assembly of Native Silk Proteins

The work presented in this chapter was a collaborative study with the group of Prof. Schniepp at The College of William and Mary (Virginia, US). The AFM instrument was operated by Ms. Minzhen Cai and she used a Matlab based program for the statistical volume analysis of the single molecule images (see also chapter II.2.6). This work has been published in Biomacromolecules [37].

IV.1 Introduction

Native silk's protein shape and structure prior to spinning are key to understanding silk spinning. In the literature a variety of different shapes, sizes and proposed steps of protein assembly are suggested as indicated in chapter I.1. The main reason for this might be the wide spread approach of using silk substitutes instead of the native protein. Based on reconstituted silk (RSF) and recombinant silk (genetically expressed proteins) mainly two different types of shapes have been suggested for silk proteins: rod–like [58] and micellar/globular [98, 94, 67]. The results from different research groups diverge significantly as expected since the solution scattering and gel–electrophoresis

revealed a clear size and M_w dependence on the preparation method used for RSF (chapter III). The substantial spread in the reported protein sizes, ranging from a few to several hundred nanometers [76, 67, 58] might well be explained by the mentioned variations in the preparation methods of these proteins [58, 67, 38, 94]. This uncertainty and spread associated with artificial silk in terms of protein mass and shape highlights again the importance of studying native silk protein.

Concerning the shape of native precursor SF proteins on a macromolecular scale there is hardly any data in the literature. The most relevant work in this respect is a single–molecule AFM study of native *Bombyx mori* silk fibroin (SF) by Inoue et al. who suggested a rod–like shape for SF [61]. A rodlike structure of SF was also suggested by Viney et al. from birefringence measurements [124]. However, the model that Inoue et al. [61] developed is based on contact–mode AFM images that were acquired in air, where tip–sample interactions lead to significant lateral forces that are strong enough to frequently displace, deform and disrupt biomolecules. This suggests that the results reported by Inoue et al. [61] are based on significantly distorted images. For this reason I decided to repeat Inoue's experiment to clarify the shape of native silk proteins under the AFM. Additionally AFM in contrast to scattering techniques, allows a direct observation of the molecules. However care has to be taken in interpreting the AFM results for the single protein shape since the proteins are imaged in air and are absorbed on a surface in contrast to the scattering experiments in chapter III.2. To reduce this the AFM data obtained in this study are collected with a much gentler imaging technique than in the case of the AFM study by Inoue et al. [61].

The findings presented here suggest that their model needs to be corrected.

As outlined in the introduction I.1 the hierarchical structure of the silk fiber has been studied extensively e.g. by NMR or x–ray scattering [99, 75, 123]. In addition, AFM and electron microscopy provides evidence for larger structural building blocks of the fiber, so called nano–fibrils with diameters in the range 20 – 170 nm [101, 103, 25]. Some authors have suggested that such nano–fibrils are actually bundles of even smaller nano–fibrils with diameters of 5 nm [42], or with diameters of 10 –15 nm [132]. However, there is no direct experimental evidence supporting the existence of these 5 – 15 nm diameter nano–fibrils. Little is understood about the process of fibre formation itself. In particular any details about the protein assembly from the silk dope are still unknown. Starting at the level of individual, aqueously dissolved protein molecules, this multi–scale process is extremely challenging to study experimentally. Known contributing factors in this process are pH changes along with a shift in ion concentration [19, 18], which is accompanied by shear and flow–elongation (pre–aligning the protein molecules prior to fiber formation [72, 8]).

The experimental approach chosen here is based on high–resolution AFM imaging of dry samples made from an aqueous solution of native SF. Non–destructive non–contact mode was used, where the AFM probe is guided across the surface sensing attractive van der Waals forces [33], so that the tip rarely touches the sample and interactions between tip and sample are minimized (chapter II.2). In some samples, the aqueous protein solution has been sheared during the drying process using a spin coater. These shear–dried samples thus represent a snapshot of the structural response of silk protein

solution to shear, which is then revealed via AFM imaging with nanometer resolutions. While spin coating has been used with reconstituted or recombinant silk before [68, 66] this work is the first report of application of such a procedure with native silk proteins. Overall, the presented experiments provide novel insights into the protein shape of native SF molecules, as well as a first direct evidence for their molecular–scale self–assembly behavior under shear conditions.

IV.2 Experimental methods

IV.2.1 Sample preparation

The native SF stock solutions were prepared according to chapter II.3. Additionally RSF1g and RSF3g reconstituted silk solutions were prepared according to the preparation protocol in chapter III.2. Due to time constraints no good quality single molecule images of RSF were obtained, therefore no data of RSF are shown in this chapter. For AFM sample preparation, six of the obtained SF stock solutions with a concentration of 10 $\pm$ 0.5 mg/mL were diluted by a factor of 1:10, 1:100 and 1:1000 using picopure water (Synergy Millipore, Billerica, MA; resistivity 18 MΩ cm at 25° C). For high–resolution AFM imaging, all samples were prepared using freshly cleaved, atomically smooth mica sheets as substrates. The protein was deposited on the substrates from the diluted aqueous solutions in two distinct ways. In the first way ('sheared samples'), a shear flow was created in the SF of 3 different stock solutions by spin–coating a 200 μm droplet on the substrate at 2000

rpm using a WS–650SZ Spin Processor (Laurell Technologies Corporation, North Wales, PA). After 2 minutes the samples were dry and ready for AFM examination and 4 different samples have been imaged. In the second way of preparing the samples great care was taken to avoid any shearing of the solution ('non–sheared samples'). A droplet was therefore placed on the mica surface and kept under humid atmosphere for 30 minutes. Afterwards the mica was gently flushed with deionized water and immediately dried with nitrogen gas (Nitrogen 5.0, GTS Welco, Allentown, Pennsylvania). In total 6 different stock solutions were used to prepare the 'non sheared' samples and 9 samples of different dilutions were imaged. In all of the images a bead–like structures with comparable sizes was observed.

IV.2.2 Atomic force microscopy (AFM)

AFM measurements were conducted at room temperature using a NTEGRA Prima Scanning Probe Laboratory (NT–MDT, Zelenograd, Russia see also II.2.3). A general introduction into the basic principles of the AFM method as well as the image analysis is given in chapter II.2. In order to minimize tip–sample interactions and deformations of the soft protein molecules, scanning was carried out in true non–contact mode [33]. The AFM probes were ACTA silicon cantilevers (APPNANO, Santa Clara, California) with a typical resonance frequency of f = 300 kHz, a radius of curvature of $r < 10$ nm, and a spring constant of k = 40 N/m. In order to get AFM images with highest resolution, ultra–sharp tips with a radius of curvature $r < 5$ nm were used (APPNANO, ACTA–SS silicon cantilevers, f = 300 kHz, k = 40 N/m).

The high–resolution images were obtained using tip vibration amplitudes of less than 50 nm.

IV.3 Results

IV.3.1 SF non–sheared at low concentration

Non–contact mode AFM images of a non–sheared sample prepared from a 1:1000 diluted silk solution are shown in figures IV.1 a and b. The images show small, isolated objects of different size. The smallest objects are most likely individual silk fibroin molecules and have a round shape of 20–25 nm diameter, determined as the full width at half maximum (FWHM) of topographical cross–sections; the corresponding heights above background are ca. 1.2 nm (see chapter II.2.6). The larger objects feature more irregular shapes with the higher–magnification scan (figure IV.1 b) showing the larger objects to be clusters of several of the smaller, round objects. A statistical analysis of the volumes of these objects is performed, taking into account 91 single objects, as well as 85 clusters consisting of 2–4 subunits. Figure IV.1 c shows a histogram with the volumes of each of these objects. As expected the measured volumes increase with the number of subunits in each cluster. For figure IV.1 d the volumes of clusters are divided by the number of apparent subunits in each cluster. Objects consisting of only one subunit ('Singles', data in black), feature an asymmetric histogram dominated by small particles with volumes of less than 250 nm^3. Clusters consisting of 2–4 subunits feature roughly symmetric volume distributions centered around 750 nm^3.

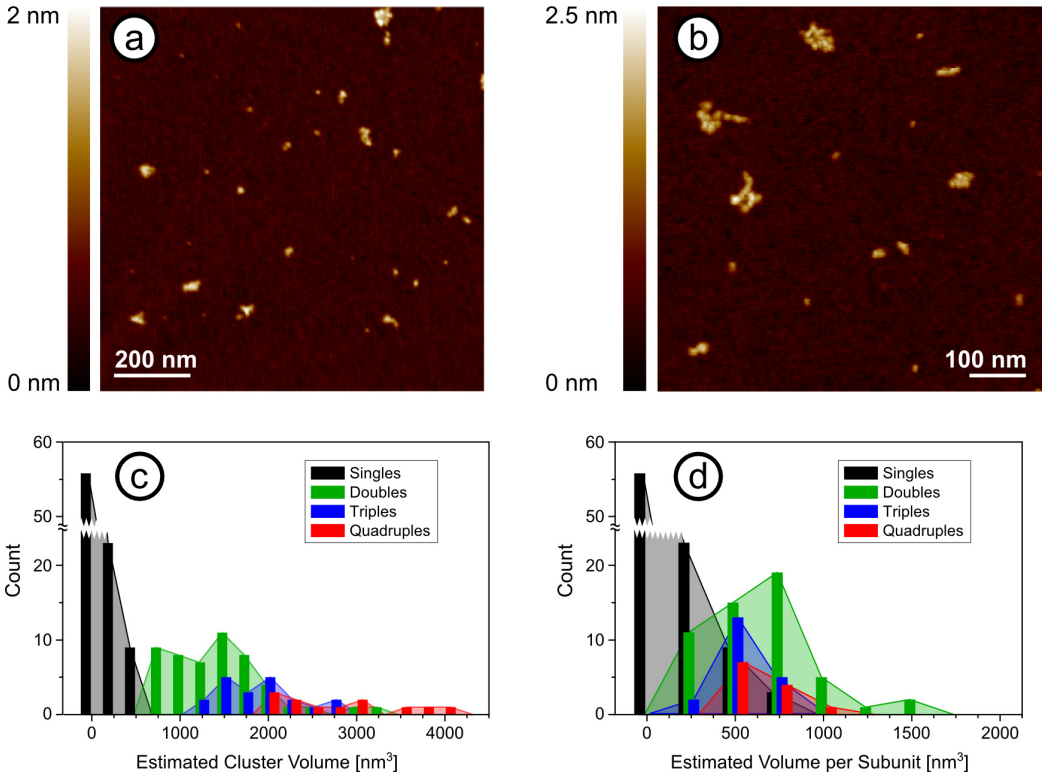


Figure IV.1: Non–contact AFM topography images from native silk molecules (SF) deposited on a cleaved mica surface from a 1:1000 diluted solution without shear. The smaller features are round, have a diameter of 20–25 nm, a height of about 1.2 nm, and are most likely individual silk proteins. The larger objects are clusters of several proteins. (c): Histograms showing the volumes of proteins and protein clusters determined from AFM topography data. (d): Same as (c), except that volumes are divided by the number of apparent subunit within each cluster.

In order to compare the obtained AFM results with volume estimates of single proteins based on their molecular weight (M_w) the formula $V = \nu M_w/N_A$ is used, where N_A is Avogadro's number and ν=0.725 cm^3/g is the partial molar volume for fibroin [133, 28]. For the known molecular weight of the heavy chain silk fibroin, 350 kDa, this leads to a volume of $V = 415$ nm^3; for a heavy chain linked by a disulfide bond to a light chain, 375 kDa, this leads to a volume of V = 451 nm^3. Due to the inaccuracies in the AFM measurements of protein volumes (see chapter II.2.4) a 1:1 correspondence between the measured values (315–750 nm^3) and predicted numbers (415–450 nm^3) cannot be established with certainty. Nevertheless, the results suggest the following interpretation of the AFM data: The larger ones of the single–unit, round objects are either individual heavy chains or heavy chains with a light chain attached. Since heavy and light chain proteins are linked via a disulfide bond the latter option seems more likely [62]. The smaller single–unit objects are likely to be P25 fibroin molecules, being only interlinked to the other silk protein by weak non–covalent hydrophobic interactions [62]. In the clusters, each of the apparent subunits corresponds to one heavy chain or heavy–light chain complex. Within the cluster potential individual light chains and P25 fibroins might be present, although not recognizable as distinct topography features.

IV.3.2 SF non–sheared at high concentration

When the protein concentration is increased to the 1:10 level a completely covered surface with homogeneous and isotropic coverage of small objects is

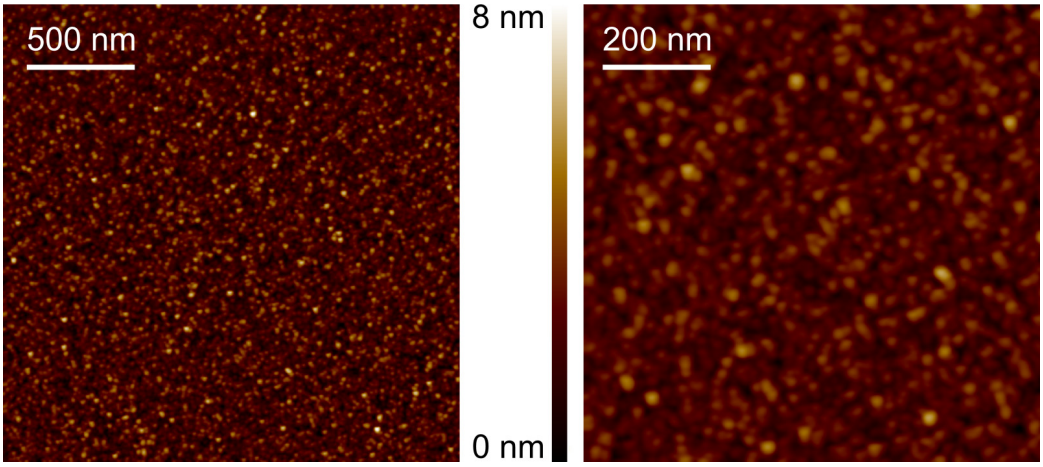


Figure IV.2: Non–contact mode AFM topography images of silk protein prepared without shearing on a mica substrate from a 1:10 diluted aqueous protein solution.

observed as revealed by the corresponding AFM images in figure IV.2 at two different magnification levels. Although more difficult to determine due to the surface roughness, many of these objects feature diameters in the 20–25 nm range and are thus likely to correspond to individual heavy chain proteins in analogy to the interpretation of figure IV.2. There is no manifest symmetry, anisotropy, positional or directional correlation or organized structure of any kind. The protein positions and directions are completely random, uncorrelated and not organized. The same is true for the organization of the individual proteins in the reported clusters observed at the lower 1:1000 concentration; no apparent organization into structure above the single–molecule level, other than the 'condensation' into clusters itself.

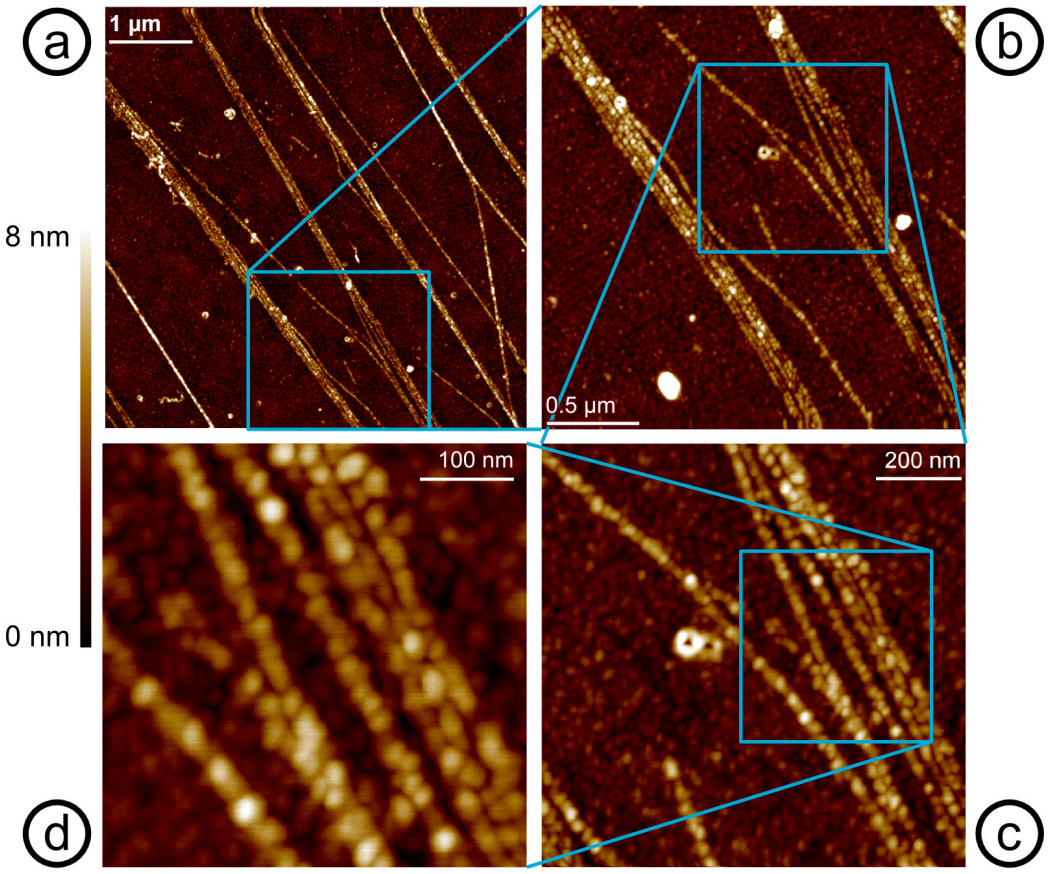


Figure IV.3: Non–contact mode AFM topography images showing a sample which has been prepared by spin–coating a 1:10 silk solution. The images show self–assembly of silk protein into linear, thread–like structures.

IV.3.3 SF under shear

In contrast to the observations on the un–sheared solutions, a significant and intriguing molecular self–organization of the silk proteins is observed when the silk solutions are exposed to shear during the deposition onto the mica substrates. Proteins that are deposited under shear from the 1:10 solution are featured in figure IV.3, which shows a series of non–contact mode AFM images at different magnifications. Figure IV.3a features bright, diagonal lines on a dark background. In terms of size, shape, distribution, and density of the proteins, the structure in the dark areas of figure IV.3 looks very similar to figure IV.2 i.e. the image of a sample made from the same silk solution without shear. This suggests that the dark areas of figure IV.3 represent homogeneously distributed, randomly positioned protein molecules (or small agglomerates thereof). In contrast, the bright, diagonal lines, represent elevated structures on the surface protruding from the 'amorphous' background of randomly placed proteins by about 2–4 nm. These lines have varying widths, ranging from 25–250 nm, and they are branching and forking in several areas. They are predominantly orientated in radial direction, normal to the spinning axis of the sample as shown in figure IV.4.

The higher–magnification image in figure IV.3b shows that each of the lines consists of a discrete number of constituent nano–fibrils. Figure IV.3c reveals that each of these nano–fibrils consists of a discrete series of elements, exhibiting a 'beads on a string' type of morphology. Based on their diameters, 20–25 nm, each of these 'beads' is probably a single silk protein. The morphology and organization of each of the beads is not highly uniform; as

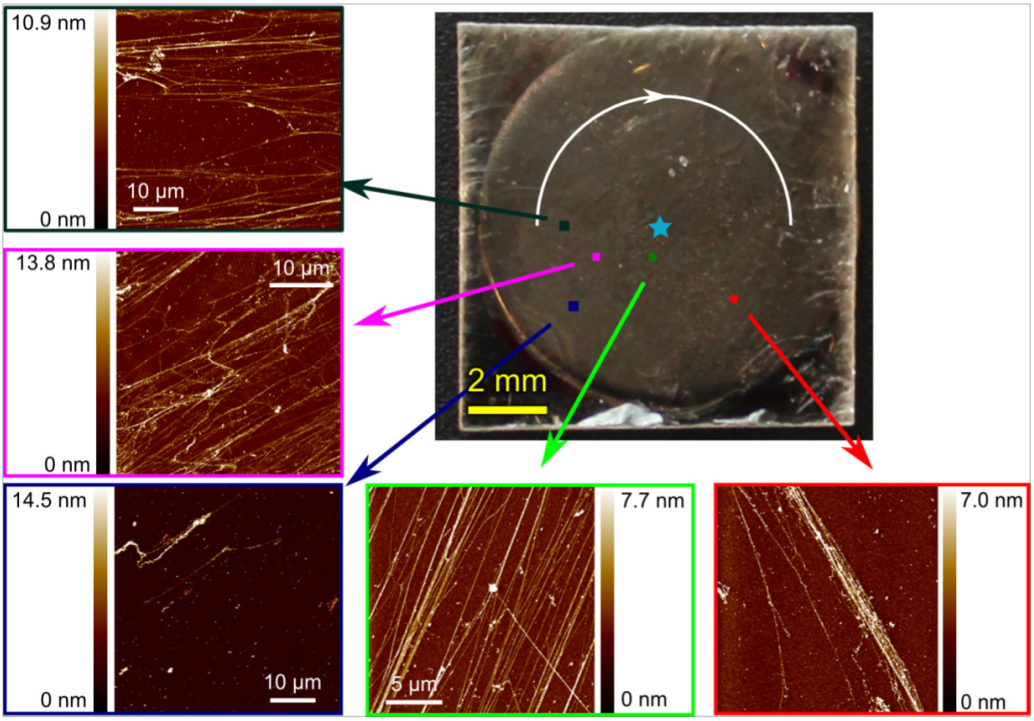


Figure IV.4: The large image in the top right corner shows a photograph of the mica substrate from above. A square mica slide (about 12 x 12 mm) is mounted on a metallic puck (about 11 mm in diameter). The rotation axis of the spin coating is indicated by the light blue star. The five pictures grouped around the photograph are AFM topography scans of the sample taken at the indicated positions. All the fibrils visible in the AFM scans directly point away from the spinning axis, in radial direction.

can be seen in figure IV.3c–d they vary in height (indicated by differences in color) and in diameter. The beads are not lined up perfectly within each of the nano–fibrils; their position is laterally offset by up to 10 nm within the nano–fibril (see figure IV.3d). A phase image corresponding to the non–contact mode topography images in figure IV.3 is shown in figure IV.5. The diameter of an individual nano–fibril is determined from this phase image to be about 25 nm. The width of a bundle consisting of 9 nano–fibrils shown in figure IV.5 allows an estimation of the fibrillar width. Assuming close packing of the fibrils in the bundle the diameter of a single fibril can be estimated with higher precision, since the error due to the finite diameter of the probe (5–10 nm) is distributed over the entire bundle (213 nm) and thus less significant. The obtained value 213 nm / 9 $\approx$ 23.7 nm is close to the 25 nm obtained for a single fibril, which suggests that size overestimation due to probe diameter can be neglected even for single–bundle measurements in this case.

Further the length of the protein molecules along the direction of the nano–fibrils has been investigated. Four topography cross–sections in the center of the nano–fibrils, following the fibril direction, are generated from AFM imaging data. Fourier transforms of these cross–sections are carried out and the frequency peak corresponding to the protein periodicity is identified (see also chapter II.2.6). The corresponding periodicities are 17.52 nm, 16.63 nm, 17.20 nm, and 15.32 nm, leading to an average of 16.7 $\pm$ 1.0 nm. The length of the proteins along the fibers, $\approx$ 17 nm, is thus significantly smaller than the observed width of the proteins perpendicular to the fibril axis, $\approx$ 24 nm. In some areas, highlighted by blue boxes in figure IV.5, this

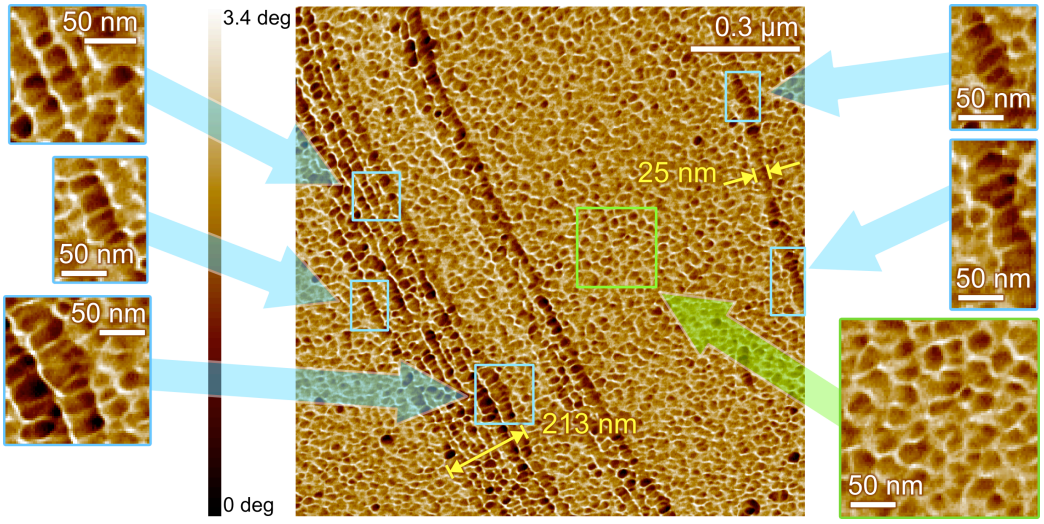


Figure IV.5: Non–contact AFM phase image of the same sample shown in figure IV.3. Areas in colored boxes are featured with higher magnification on the left and right sides of the main image. The majority of the sample looks like the area highlighted in the green box–unorganized, round protein molecules. The blue boxes highlight areas in the nano–fibrils featuring anisotropic conformation of the proteins. Sections across a single nano–fibril and a bundle of 9 nano–fibrils are indicated in yellow.

shape anisotropy of the beads in the fibers is obvious. In contrast, all the proteins between the fibrils (a representative area is highlighted in the green box) are randomly distributed and round (isotropic) with similar diameters, comparable to the film–like structures observed without shear.

When proteins from a 1:1000 silk solution are deposited on mica under shear, another mode of protein self–organization on the substrate is observed (see figure IV.6). Prepared under these conditions the great majority of the protein on the surface is organized into sparsely distributed, large assemblies. Figure IV.6a shows a non–contact mode AFM topography image featuring several of these assemblies of different size; in total, 50 of these structures are imaged. Figures IV.6b and IV.6c show a topography image at higher magnification and its corresponding phase image, respectively. Figure IV.6d shows a phase image at a yet higher magnification. The structures appear like random coils formed out of one or several threads ('nano–fibrils'), similar to observations from other 1–dimensional macromolecules, such as DNA [45] or synthetic polymers [69]. The string–like, 1–dimensional morphology of these assemblies suggests that they are not formed on the surface but in solution upon application of shear. The high–resolution AFM image in figure IV.6 (right panel) reveals that these nano–fibrils have a diameter on the order of 8 nm and an apparent height of about 0.6 nm. Like other AFM data of proteins the diameter is most likely overestimated, whereas the height is an underestimate (see also chapter II.2.4). Also a periodic structure along the direction of the nano–fibrils is noticeable, hinting at the next smaller level of structural organization. Eight short topography cross–sections, 2–5 periods long, are taken, and the average period is determined to be 8.6 $\pm$ 1.2 nm.

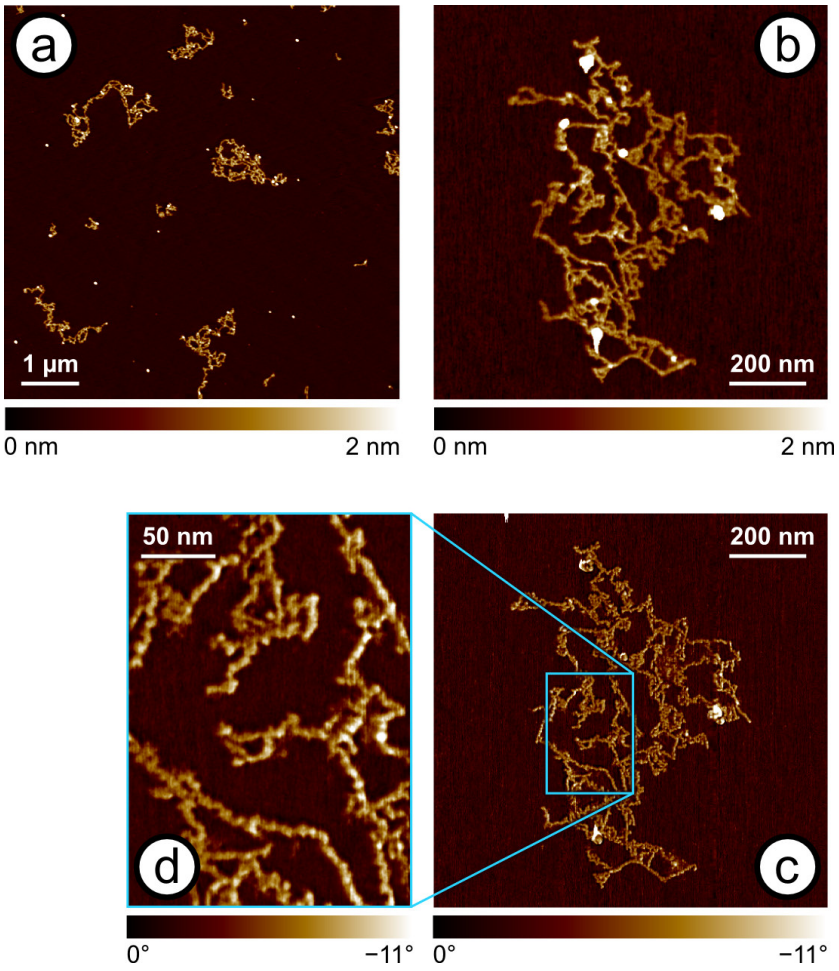


Figure IV.6: Non–contact AFM images of silk deposited from an aqueous 1:1000 solution by spin–coating, which induced shear in the solution. Panels (a) and (b): topography image; panels (c) and (d): phase images. The proteins self–assemble into nano–fibrils. The diameter is less than one of the globular single proteins shown in figure IV.1, which are deposited from the same solution. This indicates that the protein undergoes conformational changes during the spin–coating process.

IV.4 Discussion

A comparison between the results for sheared and un–sheared samples reveals several fundamental differences. In the un–sheared samples the protein is randomly distributed on the substrate. At the lower, 1:1000 concentration, individual protein molecules of globular shape are observed on the surface, along with clusters that most likely form on the substrate upon adsorption. However the results from chapter III where an increase in M_w is observed at very low protein concentrations might also suggest that the observed clusters have already formed in solution. The lateral dimensions of the protein is determined to be 20–25 nm, and its apparent height on the substrate, 1.2 nm. These findings are in clear contrast to the only other AFM work on native silk proteins in the literature by Inoue et al. [61] who found rod–like structures of 60 nm length under similar conditions. The discrepancy can be easily explained by the fact that Inoue et al. used contact–mode AFM, which is known to disrupt, displace and distort biomolecules during imaging (see chapter II.2). The data presented in this work have been taken in non–contact mode thus represents protein structure without comparable artifacts. Furthermore the resolution of the images in the work of Inoue et al. [61] is low compared to the AFM images presented here.

At the higher 1:10 concentration a completely covered substrate is observed, where the proteins are packed more densely. This close–packed morphology does not allow measuring the protein diameter with high precision or the protein height. However, given the similar shape and comparable diameter, one might suspect that the protein molecules feature the same con-

formation in both cases, just at different packing densities. Importantly no spontaneous fibrillation or elongation of the globular shapes upon increasing the protein concentration is observed in this work as reported in Inoue's work [61] as well as in studies on recombinant spider silks [94] and reconstituted silks [113, 67]. Exposing silk solutions to shear however, can trigger such transformations quite easily as shown in this work. Therefore it is possible that these authors have accidentally sheared their samples to a small degree during normal sample handling, for instance while pipetting the solutions.

When the protein solutions are exposed to shear, two interesting modes of protein organization depending on the concentration are found. At the higher 1:10 concentration, part of the protein on the surface is organized into nano–fibrils (and branching bundles thereof) that run on the substrate surface almost straight (figure IV.4). The size of the smallest visible building blocks in these nano–fibrils is similar to the dimensions of the single protein molecules and film–like structures observed without shear. Hence, these nano–fibrils observed here are likely to be shear aligned single protein molecules. It would be very interesting to know whether these 25 nm wide nano–fibrils are related to the recently observed 20 nm wide fibrils in cross–sections of *Bombyx mori* fibers [25] or to the 90–170 nm wide nano–fibrils reported elsewhere in the literature [101, 103]. In a very recent confocal microscopy and *in situ* rheology study the authors report the formation and visualization of fibrils in native SF under shear [51]. The 'smallest' observable structures in their confocal microscopy images are in the order of 300 nm [51]. Interestingly my findings seem to suggest that single fibrils often group together into 'bundles' as indicated in figure IV.3 and IV.5, which appear to

have very similar diameter to the observed structures by Holland et al. [51].

At the lower 1:1000 concentration sheared samples the protein on the surface has the shape of thin strings with a diameter of about 8 nm that are randomly coiled. Their diameter is thus significantly smaller than any of the other structures obtained in this work. All the other structures featured in figures IV.1 – figure IV.5 have building blocks with lateral dimensions in the range of 17–25 nm. This includes the single molecules (and clusters thereof) featured in figure IV.1, the homogeneously distributed, close–packed proteins in figures IV.2 – IV.5, and the proteins forming the nano–fibrils in figure IV.3 and figure IV.5. Therefore, the nano–fibrils observed in figure IV.6 cannot simply be a chain of the proteins observed in the other figures; instead the protein must have undergone a transformation into a thinner and longer conformation upon assembly into the nano–fibrillar morphology. The diameter of these nano–fibrils, $\approx$ 8 nm, is close to what Hakimi et al. (5 nm) [42] or Zhao et al. (10–15 nm) [132] have proposed. However, both groups did not present experimental evidence for the existence of such small nano–fibrils. The periodicity observed along these thin nano–fibrils, 8.6 $\pm$ 1.2 nm, is much shorter than the dimensions of individual proteins and is thus likely to represent an intra–molecular structure of the proteins in the elongated conformation. In the absence of any other experimental evidence on the nano–fibrils in the literature, however, further conclusions cannot be drawn at this point. It is worth noting that structural elements of similar size have been detected for spider silk by Grubb and Jelinski, who have found 7 nm as a lower limit for the length of β–sheet crystals along the fiber axis and other authors report an average fibril periodicity of 7 nm from SAXS

[110]. Additionally studies on the molecular structure of *Bombyx mori* silk suggested a bead–of–string model similar to our observations, however with a smaller periodicity of about 4 nm along the fibril [127].

The mechanism of assembly for both forms is currently unknown. The straighter, thicker nano–fibrils from the 1:10 concentration may, for instance, be formed through attractive interactions between single proteins. In that case, attractive forces between proteins may deform the protein molecules and thus account for the observed anisotropy. For the lower 1:1000 concentration we hypothesize that the conformational change might have been possible due to unfolding of the silk protein (loss in structure) at low concentrations, also observed in the work on solution scattering on native silk (chapter III, [37]), which allows an aggregation into thin fibrils. Further investigation will also be needed to determine the influence of the substrate on the formation of the reported structures. All experiments here are carried out using mica, which is a strongly hydrophilic, high–energy surface which has a significant negative surface charge under these conditions. Given the similarity between the observed features on the film–like structures and the globular shapes observed in direct contact with the mica surface the results suggest that the hydrophilic mica surface is not the main driving force for the observed structures. However it would be interesting to see if the same structures are observed on a low–energy, hydrophobic surface such as highly oriented pyrolytic graphite (hopg).

IV.5 Conclusion

Native silk proteins, obtained from silkworm glands, are diluted with pico–pure water to different concentrations and deposited on mica substrates with and without shearing the solution. The samples are investigated by non–contact mode AFM for structural analysis of the obtained protein structures. Deposition of low protein concentrations onto a mica substrate yielded samples featuring single protein molecules and small agglomerates thereof, with random organization. The single molecules featured appear to be of globular shape with diameters in the range 20–25 nm and an apparent height of $\approx$ 1.2 nm, which is in clear contrast to the observed rod–like structures by Inoue et al. [61]. At an increased protein concentration the substrate is completely covered with objects of similar size and shape in random and uncorrelated positions and orientations. No spontaneous fibrillation at higher protein concentrations is observed as reported previously in recombinant spider silks and reconstituted silkworm silk [94, 113]. When shear is applied during deposition by spin–coating the silk protein solution onto the substrates, fibrillar structures of two different kinds are obtained, depending on the protein concentration. At high (1:10) protein concentrations single molecules that are shear aligned into threadlike structures and bundles thereof are observed. The average sizes of the aligned molecules are comparable to the single molecule diameters, though the molecules appear slightly compressed in the direction along the fibrils. This suggest a pre–fibrillar state with the average fibrillar width of $\sim$ 25 nm close to the 20 nm observed for nano fibrils in AFM in *Bombyx mori* fibres [25]. At low (1:1000) silk protein con-

centration, randomly coiled nano–fibrils of much thinner diameter ($\approx$ 8 nm) are found, which might correspond to similar sizes observed in the literature [110]. To organize into this structure, the protein has to undergo a significant conformational transformation, as the protein exhibits a significantly larger diameter in single–molecule conformation. Therefore I hypothesize that the observed conformational change is possible due to an unfolding of the protein at low concentrations [37] allowing a subsequent aggregation into thin fibrils.

V SAS on Native Spider Silk Proteins

V.1 Introduction

Spiders are able to spin silks with different mechanical properties according to their function in the web. The secret of spider silk's unique mechanical performance appears to lie in the ratio of disordered and ordered fractions inside the final fibre [99, 128]. This makes spider silk an ideal model to study the structure function relationship in silks (figure V.1). There has been considerable debate about whether the protein sequence inherently leads to the mechanical properties or whether it is the spinning process, which confers the degree of disorder/order in the fibre [47, 99]. However, what is clear is that there is a link between the mechanical properties and the molecular order, e.g. high toughness and crystallinity down to high extensibility and disorder. Therefore the folding and assembly of the silk proteins itself has attracted a lot of attention in various areas of science [78], [59].

Despite important findings for the shape and potential assembly processes of recombinant spider silks [4, 41, 32] the structure of the native silk protein before it is spun into a fibre remains under debate. NMR suggests a random coil type of structure in the precursor solution (i.e. dope) of *Bombyx mori* fibroin [2, 132] as well as major ampullate silks [1, 55]. In contrast circular

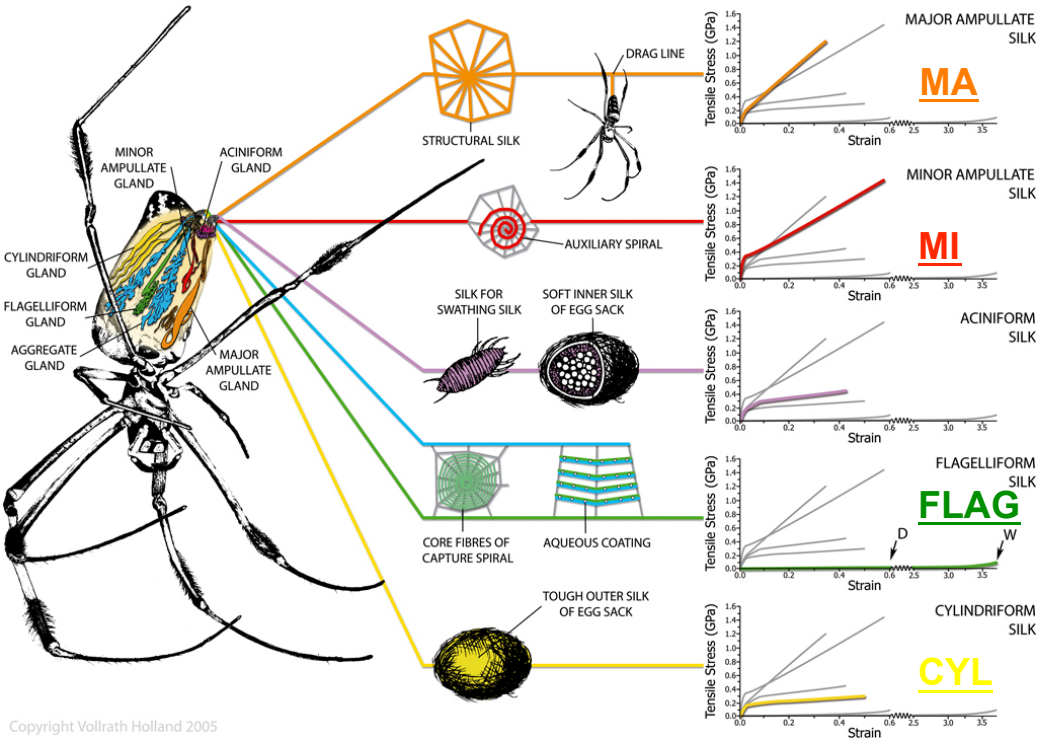


Figure V.1: Spider silk proteins of the golden orbweb spider genus *Nephila*. The spider is able to spin different types of silk depending on their function in the web. Highlighted are the 4 different silks analyzed in this study. Image reproduced from [127].

dichroism (CD) and other spectroscopy techniques (e.g. FTIR) indicate the presence of secondary/ tertiary structures in the native proteins which are specific to each silk type [21]. To clarify the link between conformation and function of silk proteins in solution prior to spinning, in this chapter four different native silk proteins from *Nephila edulis* spiders are investigated using a combination of CD (secondary structure) and SAXS (shape/ size). Typically, small angle scattering provides information on shape, size and interactions of proteins in solution [119, 73] while CD indicates the secondary and tertiary structures. Combining CD and SAXS our results reveal that each of the different silk proteins exhibits a specific shape and structure in solution linking local molecular order to macroscopic functionality. Further modeling using a semi–flexible polymer approach suggests a relationship between structural disorder and molecular chain flexibility.

V.2 Methods

Nephila edulis (Tetragnathidae), golden orb spiders, were raised in a greenhouse under controlled humidity and temperature conditions. The major (MA), minor (MI), flagelliform (Flag) and cylindriform (Cyl) glands were retrieved by dissection of mature female spiders. The glands were gently peeled to remove the epithelium and the proteins were dissolved in small eppendorf tubes in milliq water over night at 6 degree to equilibrate. Apart from MI all silk proteins showed high solubility and clear uniform stock solutions were obtained. MI in contrast did not dissolve completely and the obtained stock solutions showed a tendency to flocculate. The protein concentrations of the

stock solutions were obtained from dry weight calculations following drying of an aliquot in a vacuum oven for 2 hours. The error in protein concentration of the stock solution was estimated to be ±0.1 mg/mL. The dilution series were all prepared by weight instead of volume for more accuracy. All dilutions were prepared from the initial stock solution of the protein and diluted according to the stated dilution ratio by weight (sample–1 = stock solution; sample–2 = dilution 1:1; sample–4 = dilution 1:4; sample–8 = 1:8 and sample–16 = 1:16). The major ampullate stock solutions had concentrations of 18.1 and 20 mg/mL respectively, the minor ampullate samples 1.4 and 2.7 mg/mL, the cylindriform stock solutions had concentrations of 3 and 1.5 mg/mL. The protein concentration for the flagelliform dilution series was not obtained. However a concentration of 3 mg/mL is estimated for the Flag stock solution due to scaling for the high q region with another flagelliform stock solution (0.3 mg/mL) and the MA, MI and Cyl solutions. Note that the pHs of all silk solutions were not controlled resulting in a pH averaging around 6.5.

The protein solutions were measured at the bio–SAXS beamline ID14–3 at the ESRF (Grenoble) using the automated sample changer system which is a modified HPLC pump system [96]. The samples were pipetted into a 96 well plate before sealing with crystal tape. Each sample was loaded into the quartz 1 mm diameter capillary at the lowest flow rate (0.5 μm/s) and a buffer sample was measured prior and after the sample. Between samples, the capillary, syringe and connecting pipe work were flushed with 12 M urea solution. Data were collected at a fixed sample detector distance of 2.43 m, wavelength 9.3 Å. Exposure times of maximum of 1 second were chosen in

the flowing system and up to 50 frames were taken per sample of 30 μL. Radiation damage was checked for carefully and evidence was found after exposure times exceeding 5 seconds at a single position as determined by the SAXS signal. Under normal measurement conditions, radiation damage was avoided by continuing to flow the sample through the capillary during the collection of x–ray data. Data were reduced using the data reduction pipeline from ID14–3 [96]. The obtained three column ascii files for each frame were averaged using PRIMUS [74], where care was taken to exclude empty frames or frames affected by the meniscus of the solution. Guinier analysis was conducted using PRIMUS [74] fulfilling the Guinier approximation for solution scattering ($q \cdot R_g$ <1.3) [73]. The obtained M_w values were calculated by using BSA as a calibration standard and the errors are estimated to be 10% [140]. In GNOM standard settings were chosen for the $p(r)$ and the D_{max} was estimated visually from the fit [120] using a q_max of 0.4 nm^{-1}. For the Porod regime the region from 0.3 – 1 nm^{-1} was fitted using Origin, the errors are calculated from linear regression. For MI a fitting region of 0.5 – 2 nm^{-1} was chosen.

The same stock solutions were also measured at the circular dichroism (CD) spectrophotometer instrument Jasco 810 at the EMBL (Grenoble) prior to the SAS experiment. The samples were measured in quartz cells (Hellma) of either 0.01 mm or 0.1 mm path length (depending on protein concentration). Three consecutive spectra were recorded and averaged using a 1 nm resolution step and 200 nm/min scan rate. The data were reduced and analyzed using the CDTool software [81].

V.3 Results

V.3.1 Circular Dichroism (CD) of native spider silk proteins

Prior to our SAXS measurements, each of the stock solutions were measured by Circular Dichroism (CD) (figure III.7). The spectra are consistent with known CD spectra [21] showing distinct structural features between the 4 types of silk proteins. While major ampullate (MA) shows a strong negative peak at 199 nm indicating random coil structure it also shows a plateau at 217 nm suggesting the presence of beta–like structures (sheets and turns). Overall the spectrum suggests a polyproline–II confirmation for MA [21]. The spectrum of minor ampullate silk (MI) suggests a 3_1–helix structure, showing a strong negative peak at around 208 nm. The spectra of flagelliform (Flag) silk indicates beta–spiral confirmation as predicted by Zhou et al. [136] while cylindriform (Cyl) shows an alpha helical dominated spectrum [21].

From the CD spectra the so–called folding index can be extracted for each silk [22]. The folding index is defined as

$$\gamma = \frac{\Theta_1}{\Theta_2} \tag{V.1}$$

with Θ_1 and Θ_2 being the measured ellipticities (mdeg) of the two minima found at around 220 and 200 nm in silks CD spectra (see figure V.2). A folding index of around 1 is typical for a folded helical structure, while a value below 0.5 denotes a partially to fully unfolded structure [22].

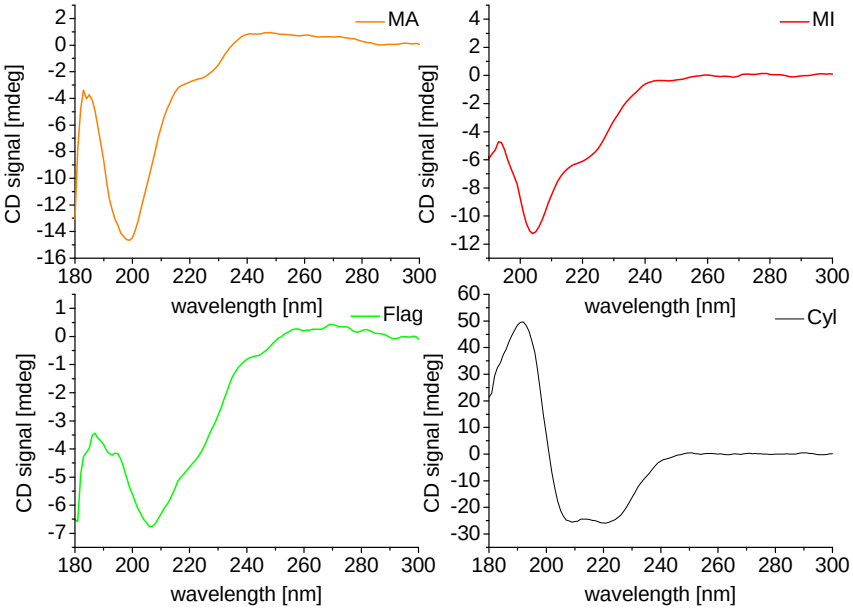


Figure V.2: All stock solutions were measured by Cicular Dichroism (CD) prior to the SAXS experiments to check the quality of the protein solutions. The CD spectra of MA, MI, Flag and Cyl silk proteins in H_2O suggest clear structural differences between the 4 silks as reported in the literature [21].

Table V.1: Folding index of principal spider silks

Silk	Folding index	Error ±
MA-1	0.249	0.003
MI2	0.543	0.032
Flag-2	0.645	0.032
Cyl-1	0.991	0.011

The folding index for the four silk proteins is increasing from MA to Cyl, indicating an increase in the amount of folded structures in the silk proteins (table V.1). To gain further understanding of the silk structure in solution we investigated the scattering intensity measured by SAXS for each of the 4 types of native spider silk protein. For clarity the SAXS results for the silks have been separated and for each silk a presentation of the concentration effects and overall scattering feature is presented.

V.3.2 Dilution series of native spider silks

V.3.2.1 Major ampullate

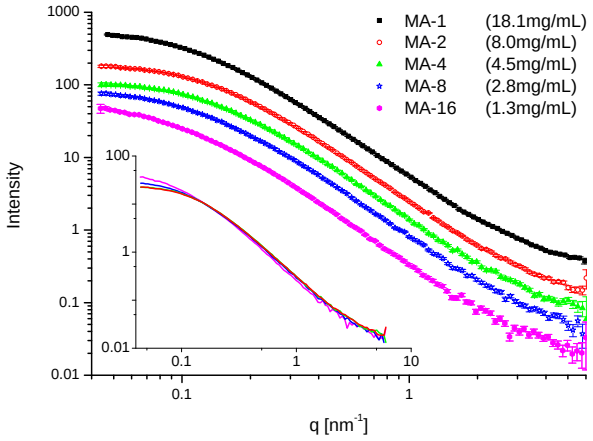
Figure V.3 shows typical SAXS spectra of major ampullate silk at concentration varying between 18.1 to 1.3 mg/mL. A normalization for protein concentration (inset in figure V.3) indicates that the overall shape of the scattering curve does not vary significantly since the curve of the stock solution MA–1 (18.1mg/mL) is overlapping the MA-8 dilution (2.8 mg/mL) (figure V.3, (b)). This suggests the absence of inter–particle interactions at these concentrations. The lowest concentration sample (1.3 mg/mL) however is showing an increase in intensity at low q compared to the higher concentration samples. For the dilutions 1 to 8 an R_g of 10–12 nm is obtained by Guinier analysis. From the intensity at zero angle the molecular weight for MA is estimated at an average of 530 $\pm$ 71 kDa for the measured stock solutions and the dilutions down to 2.8 mg/mL. These values are consistent with a dimer and values reported in the literature [116]. The slope in the high q region of the MA scattering curve, so–called Porod regime, is indicative

of the protein interface. Extraction of the Porod exponent yields a value of 2.04 ±0.004 and 2.09 ±0.007 for dilution 1 and 8 whereas dilution 16 on the other hand has an exponent of 2.17 ±0.01.

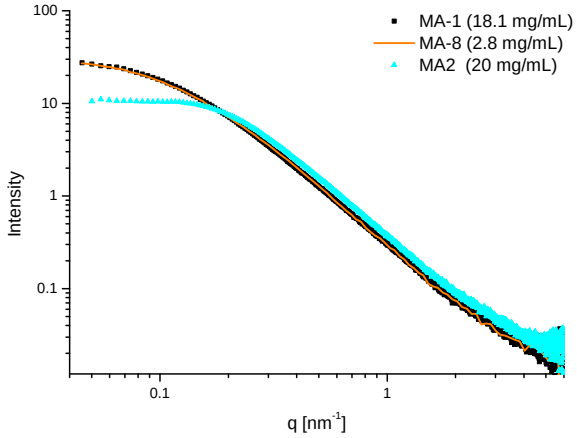
In one instance, a measured stock solution of MA showed a clear deviation in shape as compared to the other stock solutions (figure V.3). The R_g for this sample is approximately half the value for the other MA samples (6.2 ±0.8 nm) with a molecular weight of 223 kDa. This molecular weight value is in keeping with the 250 kDa reported from gel electrophoresis for monomeric state [116].

V.3.2.2 Minor ampullate

Figure V.4 shows a typical scattering curve for MI silk proteins. The smaller protein availability of MI as compared to MA precluded a detailed concentration analysis. However, two distinct states were observed: MI1 representing the sets of data with increasing intensity at low q and MI2 with an extended Guinier region indicating the absence of larger scale aggregates. The Guinier analysis of MI2 gives an R_g value of 3.34 ± 0.132 nm with a molecular weight estimate (from zero angle intensity) of 36 kDa. This value appears unrealistically low perhaps due to an overestimation of the protein concentration given the stated M_w for MI spidroin 1 and 2 (351 kDa/ 255 kDa) [10]. The Porod exponent for the MI2 sample is found to be 1.89 ± 0.01, while the MI1 in the linear region gives an exponent of 3.56 ± 0.04.



(a)



(b)

Figure V.3: (a) SAXS curves of protein concentration series of native MA. The numbers 1 to 16 state the dilution ratio (1= stock concentration) and the values in brackets state the measured dry weight concentration. Inset shows curves normalized for protein concentration suggesting the absence of structure factor. The lowest concentration shows a slightly higher intensity (MA–16). (b) Interestingly one of the measured MA samples showed a totally different shape to the other stock solutions (MA2, cyan curve). The M_w calculated from the I(0) gives 223 kDa, the calculated R_g and D_{max} values are approximately half the value obtained for the other MA samples suggesting a monomer state for this sample. Note also the complete overlap from dilution MA–8 (orange line) and the stock solution MA–1 (black rectangles).

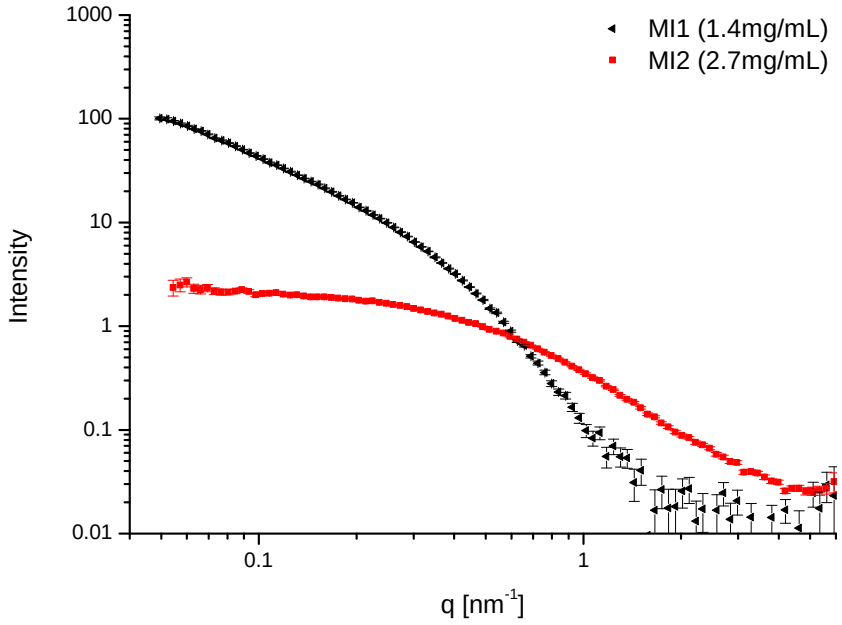


Figure V.4: (a) SAXS scattering curves of two stock solutions of minor ampullate silk at a protein concentration of 2.7 mg/mL (red curve) and 1.4 mg/mL (black curve). The black curve (MI1) shows an increase in intensity at low q compared to MI2, typically an indication of aggregation.

V.3.2.3 Flagelliform

Figure V.5 shows a typical concentration series for flagelliform silk. For all concentrations a feature at 0.3 nm^{-1} corresponding to 21 nm in real space (figure V.5) is observed. This knee like feature does not disappear with lowering the protein concentration, which suggests that the feature is related to the shape and not to an interaction or structure factor effect. The absence of a measureable Guinier region precludes any accurate measurement of size and molecular weight for the Flag dilutions. A rough estimate by extrapolation to zero angle intensity suggests a molecular weight of 1200 kDa. For a monomeric state the reported value of the M_w of Flag is 360 kDa [40] estimated from partial DNA sequences suggesting an oligomeric state for the Flag protein in solution. Under the assumption that flagelliform is an elongated particle, the cross–sectional radius of the protein can be estimated. Using Primus [74] and BayesApp [44] analysis tools we found a cross–sectional radius in the range of 5.4 – 5.7 nm for the dilutions 2–16. Determination of the Porod exponent for flagelliform yield an average value of 2.0 $\pm$ 0.007 independent of concentration.

V.3.2.4 Cylindriform

Figure V.6 shows a typical concentration series of the egg case silk, cylindriform. Normalization of the intensity for dilution ratio reveals that the overall shape of the curve is not changing significantly (inset in figure V.6). The R_g for cylindriform is estimated at a value of 12 – 13 nm for the Cyl–1 to Cyl–4 samples, with an estimated molecular weight of 300–321 kDa. Despite the

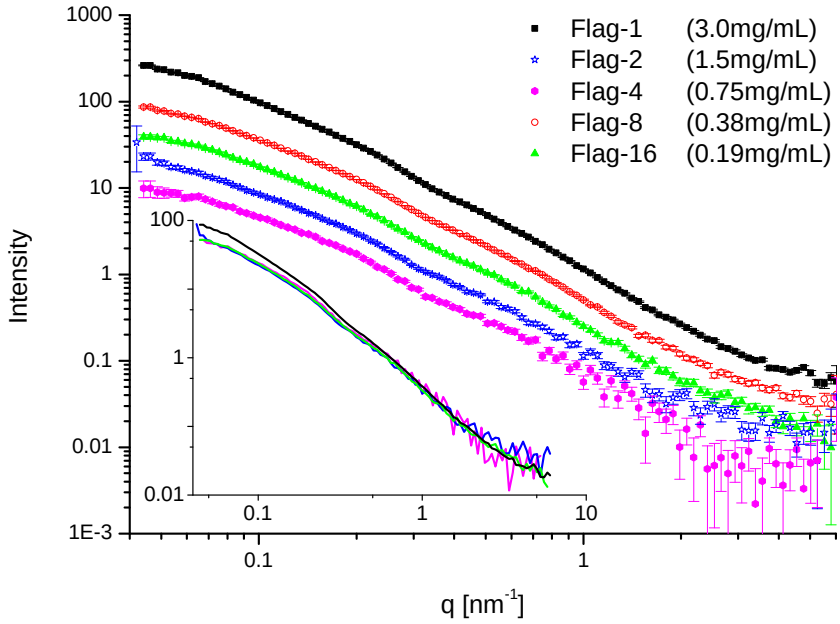


Figure V.5: SAXS concentration series of native flagelliform silk protein. Note the concentration is only an estimate from the re–scaling of the SAXS data as no protein concentration was measured. flagelliform has a distinct feature at 0.3 nm^{-1} corresponding to 21 nm in real space, which appears to be an intramolecular shape feature. Inset: Curves normalized for dilution ratio. Typically a deviation of the normalized intensity at highest protein concentration suggests the presence of a structure factor (interparticle interactions).

uncertainty in protein concentration the values found are in keeping with the values found by other researchers, i.e. 366 kDa for cylindriform silks [83]. The Porod exponent for all cylindriform silks lies between 1.77 and 1.84 and is therefore significantly lower than the exponent of the other 3 silks.

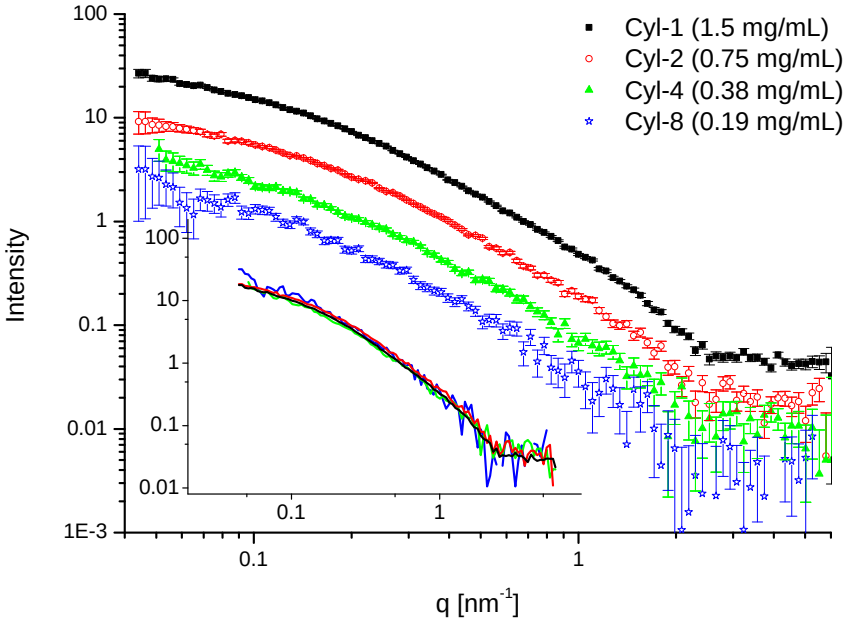


Figure V.6: Solution SAXS of cylindriform silk at different protein concentrations. Number in legend states dilution ratio (1=stock solution; 2= stock diluted bei half). Scaling with MA and MI suggests a protein concentration of 1.5 mg/mL for the Cyl–1. Inset shows same curves normalized for dilution ratio.

V.3.3 Comparison of 4 spider silks

The analysis of the concentration series of each of the spider silk proteins suggests that a structure factor (intermolecular interactions) can be neglected

at protein concentrations below 3 mg/mL for all silks. This allows a further analysis and direct comparison of the four silks at similar protein concentrations. For the major ampulate the dilution 8 is used for further analysis. Figure V.7 shows an overlay of the scattering profiles of the four silk proteins at a comparable concentration revealing clear difference in overall shape. Table V.2 shows a numerical comparison of the R_g's and Porod exponents for all silks. MI shows a significantly lower R_g in comparison with the MA and Cyl. Flag lacks a Guinier region and is showing a shape feature at 0.3 nm^{-1} in contrast to the other three silks. The Porod region also indicates differences in the silks at similar concentration (see table V.2). While the exponents for MA and Flag are relatively close to the Gaussian coil value of 2, the values for MI and Cyl are below 1.9. Cyl shows the lowest Porod exponent of 1.77 indicating a difference in local or surface structure compared to the other silks (see table V.2).

Table V.2: Comparison of R_g and Porod exponent

Spider silk	Protein conc. [mg/mL]	R_g [nm]	Err(R_g) [nm]	Porod	Err(Porod) [nm]
MA1–8	2.8	12.95	0.11	2.09	0.01
MI2	2.7	3.34	0.13	1.87	0.01
Flag–2	1.5	5.74	0.04	1.96	0.01
Cyl–1	1.5	13.70	0.34	1.77	0.01

V.3.4 Distance distribution function $p(r)$

The radius of gyration can also be obtained from the distance distribution function $p(r)$, which is known to be more accurate than the Guinier analy-

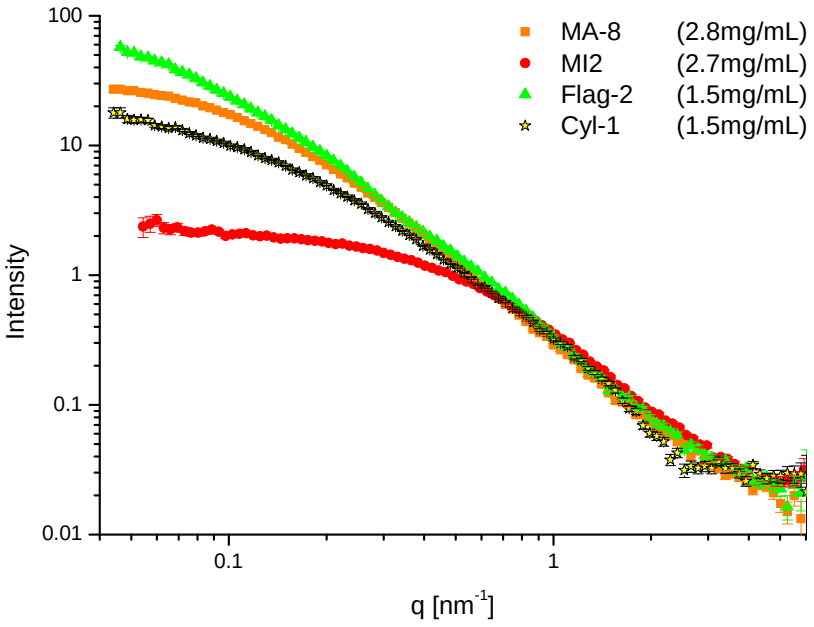


Figure V.7: A direct comparison of the four spider silk proteins at comparable protein concentration reveals clear differences in the overall shape of the protein (intensity normalized for protein concentration).

sis, taking this time the whole scattering curve into account (chapter II.1.5). Additionally the protein maximum dimension (D_{max}) is extracted from the $p(r)$ analysis. The $p(r)$ function has been obtained by indirect Fourier transformation method (using GNOM [121], see figure V.8) as well as Bayesian statistics (Bayes Fit [44]). The advantage of the Bayesian statistics over GNOM is that the D_{max} is estimated from the program and no value has to be suggested by the user in contrast to GNOM.

The R_g values obtained from the $p(r)$ function (see chapter II.1) are slightly larger than the R_g from Guinier analysis for MA, MI and Cyl, which is typical for partially disordered proteins [105]. For flagelliform a value of 20 nm was found from the GNOM analysis. The overall shape of the $p(r)$ curves is showing a similar anisotropic shape with a D_{max}/R_g ratio above 3 for all 4 silks. This anisotropy of the $p(r)$ is typical for elongated particle shapes (e.g. rods, compare chapter II.1) but also often observed in flexible systems [6]. Interestingly Flag (green in figure V.8) shows a very distinct $p(r)$ function, arising from the feature at $0.3\ \mathrm{nm}^{-1}$ in the scattering curve. This double peak feature in the $p(r)$ indicates a dumbbell shape of the protein in solution typical for multi–domain proteins [90]. The D_{max} estimate is found to be over 50 nm for MA, Flag and Cyl, with the exception of MI with a D_{max} of 16 nm.

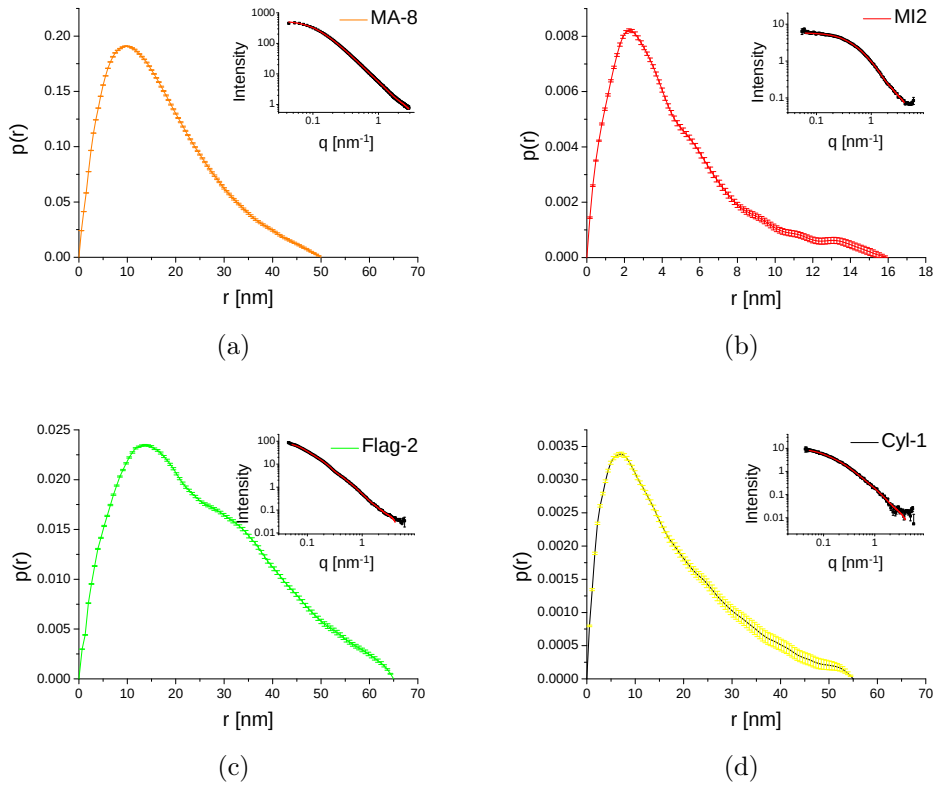


Figure V.8: The $p(r)$ curves for the 4 types of spider silk protein obtained from GNOM [121]. The insets show the fits (red line) to the data (black squares). Assuming mono–dispersity for each of the systems the D_{max} was evaluated from the best fit, especially at low q values. Overall good to reasonable fits were obtained (insets). The $p(r)$ of all silks suggests an anisotropic/ elongated structure. Interestingly the $p(r)$ for the flagelliform silk shows a multi–domain or dumbbell shape for the protein in solution [102].

V.3.5 Semi–flexible analysis of silk proteins in solution

V.3.5.1 Kratky and Holtzer analysis

The folding state of the four spider silk proteins is investigated using Kratky plots [102, 90]. The scattering intensity (normalized for protein concentration) is therefore multiplied with q^2 and plotted as a function of q, yielding a characteristic shape depending on the folding state in solution (see chapter II.1). Completely folded proteins exhibit a bell shaped curve, an opening of these bell shape indicates partial folding while a flat plateau region with no maximum is indicating a random coil confirmation [102, 90]. Figure V.9 compares the Kratky plots of all 4 spider silk proteins.

The MA dimer state (figure V.9, MA, orange curve) exhibits a maximum in the Kratky plot at about 0.5 nm^{-1} indicating partial folding. The presence of a maximum is less pronounced in the putative monomeric state (figure V.9, MA, black curve). MI2 (figure V.9, MI, red curve) shows a partially folded structure. The low data quality at moderate q obscure the presence of a broad peak indicative of further folding. The increase, however, in intensity at higher q values is evident in both MA and MI2 samples. Interestingly the MI1 sample which showed an increasing intensity at low q in the scattering curve compared to MI2 (MI1 black curve in figure V.4) has a distinct bell shaped curve in the Kratky plot indicating a fully folded structure (figure V.9, MI1, black). The Flag indicates partial folding, which is more pronounced in the Kratky plot for Cyl. The increase in intensity at high q values seen in all 4 silks is typically an indication for a flexible systems and elongated 'rod–like' shapes. A measure for the flexibility of a polymer (or extended

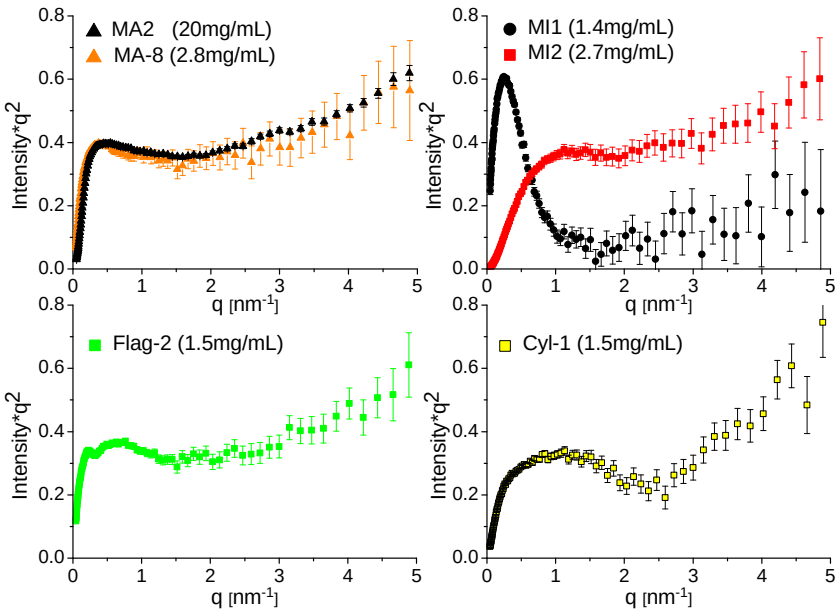
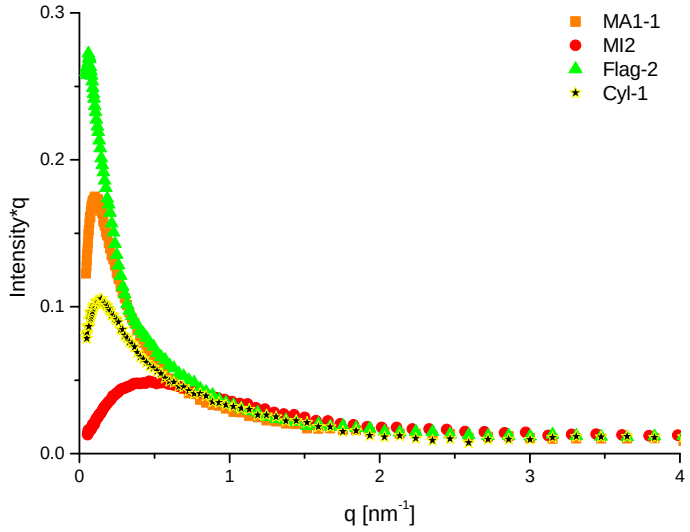


Figure V.9: Kratky plots of the 4 principal spider silks. The MA dimer state (orange, MA1-2) clearly indicates partial folding while this state seems to be less obvious in the potential monomer state (MA2, black). For the MI (MI2, red) no clear maximum is observed however a deviation from the plateau at around 2 nm^{-1} is evident. Interestingly the MI1 (black) sample showing an increasing intensity at low q in the scattering curve (figure V.4) shows clear indication of a fully folded structure. The Flag indicates partial folding, which is more obvious in the Cyl Kratky plot. All 4 silk proteins show some degree of flexibility indicated by the continuous increase in the curves towards higher q values.

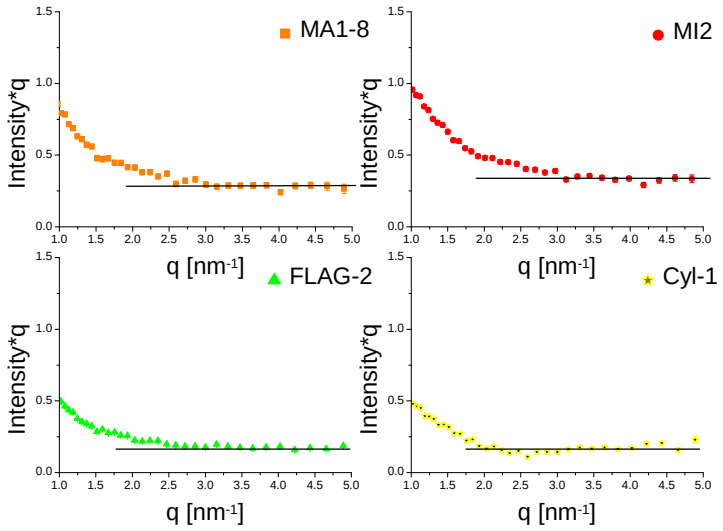
protein) chain is the persistence length. As mentioned in the methods section the persistence length (l_p) can be extracted from the Kratky plot in case of Gaussian coil behavior [35] as follows:

$$l_p = \frac{6}{\pi q^*} \tag{V.2}$$

In the particular cases of flexible systems and elongated 'rod–like' shapes, the Holtzer plot is preferred to Kratky plot [15, 54] for the characterization of flexible systems. The so–called Holtzer plot or bending rod plot often allows a more precise determination of the persistence length compared to the Kratky analysis [15]. The Holtzer plot is obtained by plotting the intensity (normalised for protein concentration) multiplied by q as a function of q. A characteristic Holtzer plot for flexible and semi–flexible polymer chain reaches a maximum at low q values and shows a q^{-1} dependence at high angles resulting in a flat plateau. Stiff chains in contrast do not show a peak in the low q [15]. The departure point from the plateau line at high q, is marked by a q^* which is a measure of the persistence length l_p (half the statistical Kuhn length l_k). Different ways have been proposed to calculate the persistence length from the estimated q^* of the Holtzer analysis [80, 56] and it was found that for various chain length $q^* = 3.5/l_p$ is a good approximation [80]. Figure V.10 shows a comparative Holtzer plot of the four silks. All four silks exhibits a clear peak at low q and a flat plateau at high q characteristic of flexible systems. The departure point from the plateau region at high q (q^*) is obtained by linear regression in this region (see figure V.10). The values obtained for q^* and l_p are summarized in table V.3.



(a)



(b)

Figure V.10: (a) Holtzer plots of the four spider silk proteins normalized for protein concentrations. The flat plateau and the peak at low q characteristic for a flexible system are present in all 4 silk proteins. (b) Estimation of q^* from the Holtzer plot (deviation from plateau region). The coloured data are the binned data and the line overlays the fit obtained by linear regression.

Table V.3: Results obtained from Holtzer analysis

Silk	q^* [nm^{-1}]	l_p [Å]	flexibility $1/l_p$
MA1–8	2.75 ± 0.2	12.73 ± 0.86	0.079 ± 0.005
MI2	3.0 ± 0.25	11.67 ± 0.90	0.086 ± 0.006
FLAG–2	2.5 ± 0.15	14.00 ± 0.79	0.071 ± 0.004
Cyl–1	2 ± 0.25	17.5 ± 1.94	0.057 ± 0.006

A comparison of the values for the length parameters derived from q^* of the four spider silk protein reveals an increase in the persistence length l_p going from MA/ MI to Flag and Cyl. The parameter $1/l_p$ is often used as an indicator for flexibility of the protein chain and suggests that Cyl is the most rigid of the four silks. Interestingly this is in line with our findings from the CD analysis and the folding index (section V.3.1). An increase in the folding index, in other words secondary structure, is correlated to a drop in the flexibility index determined from the scattering data.

V.3.5.2 Worm–like chain modeling of silk proteins

After confirming that silks are flexible systems using the Kratky and Holtzer analysis, a worm–like chain (WLC) model is fitted to the SAXS data, which is a standard analytical model for flexible systems in polymer physics. The WLC model allows for a more accurate determination of l_p [95] compared to the asymptotic methods since the whole data range is used for fitting, taking the cross–section of the chain as well as excluded volume effects into account

(method 3 in reference [95]). The form factor for the WLC is given by [84]

$$S(q) \approx [(1 - \chi(q, L, l_b)S_{coil}(q, L, l_b))+ \\ \chi(q, L, l_b)S_{rod}(q, L)]\Gamma(q, L, l_b)S_{cs}(qR_{cs}) \tag{V.3}$$

with S_{coil} representing the scattering function for a flexible polymer with excluded volume effects, S_{rod} describes the scattering function of a rigid rod, the cross–sectional function is given by $S_{cs}(qR_{cs})$, χ being the function describing the crossover region between the random walk regime (q^{-2}) and the rigid rod regime (q^{-1}). Γ is an additional correction term needed for the crossover region [84, 95].

Using the flexible cylinder model (WLC model, [95]) from SansView we obtain reasonable fits for the 4 silks (figure V.11), the fit parameters are summarized in table V.4. A perfect fit of the model is not expected due to the indication of partial folding, however a qualitative comparison of local structural differences in between the silks is possible.

Table V.4: Results obtained from WLC fits

Silk	l_k [Å]	L [Å]	r_c	N_k	l_p [Å]	flexibility $1/l_p$
MA1–8	34.43 ± 5.60	2327.40 ± 71.79	7.38 ± 3.45	67.59 ± 16.08	17.22 ± 2.80	0.058 ± 0.009
MI2	29.80 ± 1.86	210.00 ± 7.45	8.83 ± 0.15	7.05 ± 0.70	14.90 ± 0.93	0.067 ± 0.004
FLAG–2	38.10 ± 3.32	3759.10 ± 239.99	7.53 ± 3.41	98.66 ± 14.90	19.05 ± 1.66	0.052 ± 0.005
Cyl–1	55.50 ± 1.01	905.20 ± 9.09	13.55 ± 0.12	16.31 ± 0.46	27.75 ± 0.50	0.036 ± 0.001

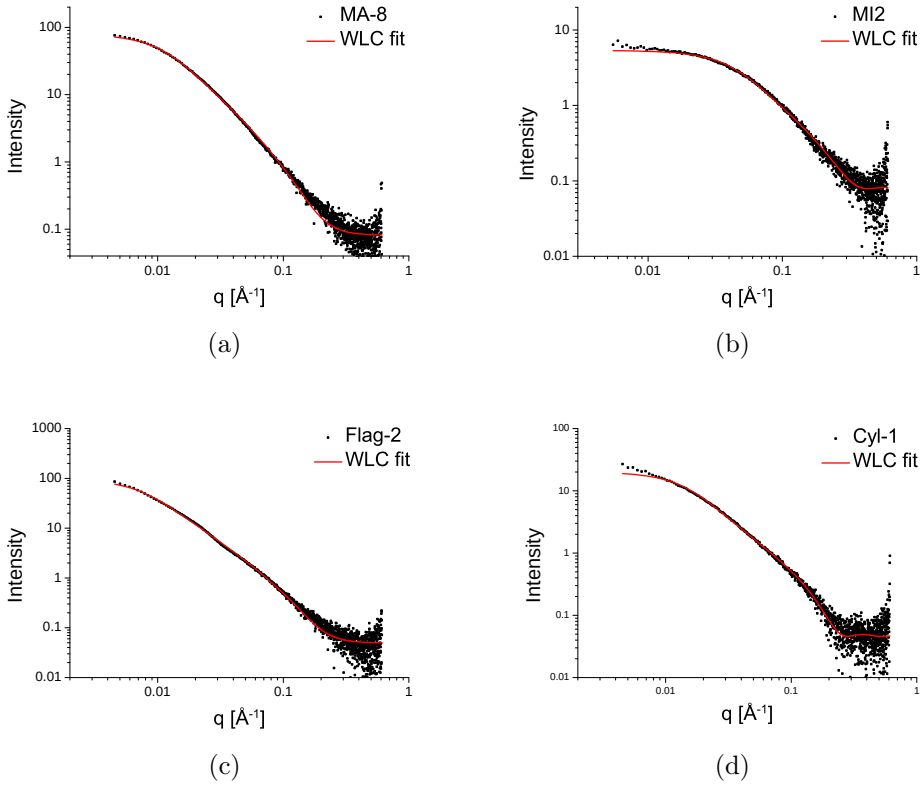


Figure V.11: WLC fit of the four silk proteins at similar protein concentration. While MA (a) and Flag (c) are in good agreement with the model at small q values, the high q region was difficult to fit and a radius polydispersity had to be applied. Cyl (d) and MI (b) in contrast show good overlap with the model at high q while a deviation at low q is observed.

Figure V.12 shows the flexibility ($1/l_p$) obtained from the WLC and Holtzer analysis as a function of the folding index obtained from the CD data (see section V.2). This comparison reveals that the WLC analysis confirms clearly the trend indicated by the Holtzer analysis suggesting a drop in the flexibility with increasing folding index.

From the WLC fit parameters other characteristic features of the silk protein can be calculated, namely for example the radius of gyration.

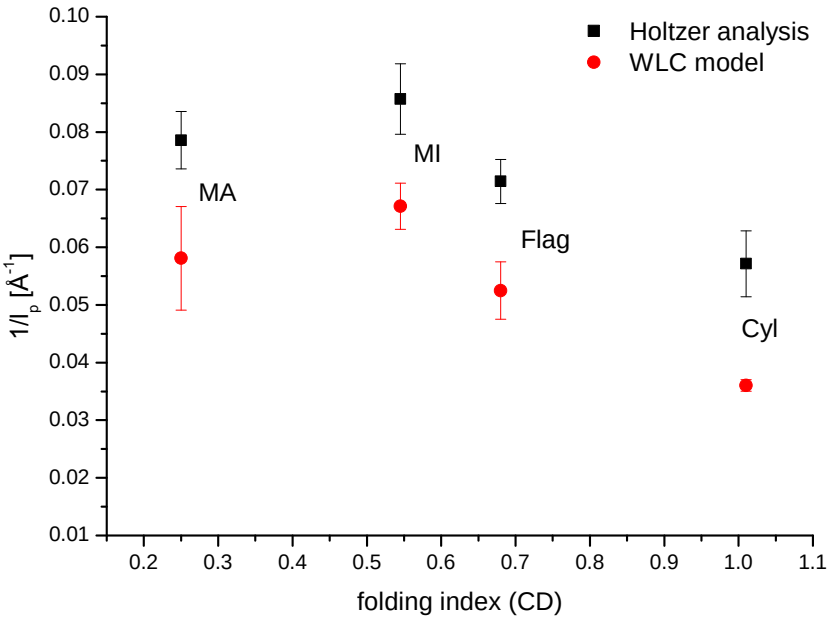


Figure V.12: Flexibility ($1/l_p$) from SAXS analysis as a function of folding index obtained from CD data shows a positive correlation. The trend observed from Holtzer analysis (black squares) is confirmed by the WLC model (red circles). With increasing folding index (CD) the flexibility of the proteins (SAXS) is decreasing. This is expected since a folding index of 1 obtained for Cyl from CD data denotes mainly folded structures (helical structures) and suggests therefore a lower flexibility index (SAXS).

The radius of gyration is estimated as follows [105]

$$R_g = l_k \sqrt{\frac{Y}{6} - \frac{1}{4} + \frac{1}{4Y} - \frac{1}{8Y^2}} \tag{V.4}$$

with $Y = \frac{L}{l_p}$ which is valid for L/l_k >10. The radii of gyration are in good agreement with the values obtained from Guinier/ $p(r)$ analysis (table V.5). The errors appear unreasonably large, resulting from error propagation. For MI this 'simple' R_g calculation is strictly not valid since $L/l_k = 7.14$<10, however the obtained R_g is comparable within errors to the R_g value from the Guinier analysis (compare table V.5 and V.2).

Table V.5: R_g values calculated from equation V.4.

Silk	R_g [nm]	Err(Rg) [nm]
MA1–8	14.95	10.58
MI2	4.31	0.57
FLAG–2	21.76	10.89
Cyl–1	12.65	7.25

V.3.5.3 Cross–sectional analysis of flexible rods

Although obtain from the WLC fit, the radius of the flexible rods does not inform on its anisotropy. To access the shape of the flexible rod cross–section, we can compute the cross–sectional $p(r)$ from a truncated scattering curve at q^*. The truncation at q^* (in real space related to the persistence length) therefore fulfil the assumption of a 1 dimensional rod, thus allowing the application of the Hankel transform to obtain the cross–sectional $p(r)$, henceforth the $p_c(r)$.

Figure V.13 shows the respective cross–sectional $p_c(r)$ for the 4 silks and the fits to the truncated scattering intensities. Remarkably all 4 $p_c(r)$ show a non–circular cross–section. The cross–section profiles are reminiscent of elliptical cross–sections. Notably, the calculated radius of the cross–section from the $p_c(r)$ are in keeping with the radius obtained from the WLC (table V.5). Overall the cross–sectional $p_c(r)$ reveals a non–trivial shape for the cross–section of each of the silks. Interestingly the cross–section of Cyl is nearly twice the size of the other 3 silks.

V.4 Summary and Discussion

Clear differences are observed in the small angle scattering of the four principal spider silk proteins, MA, MI, Flag and Cyl in solution. Table V.6 summarizes the key properties and findings for each silk from the combination of SAXS and CD.

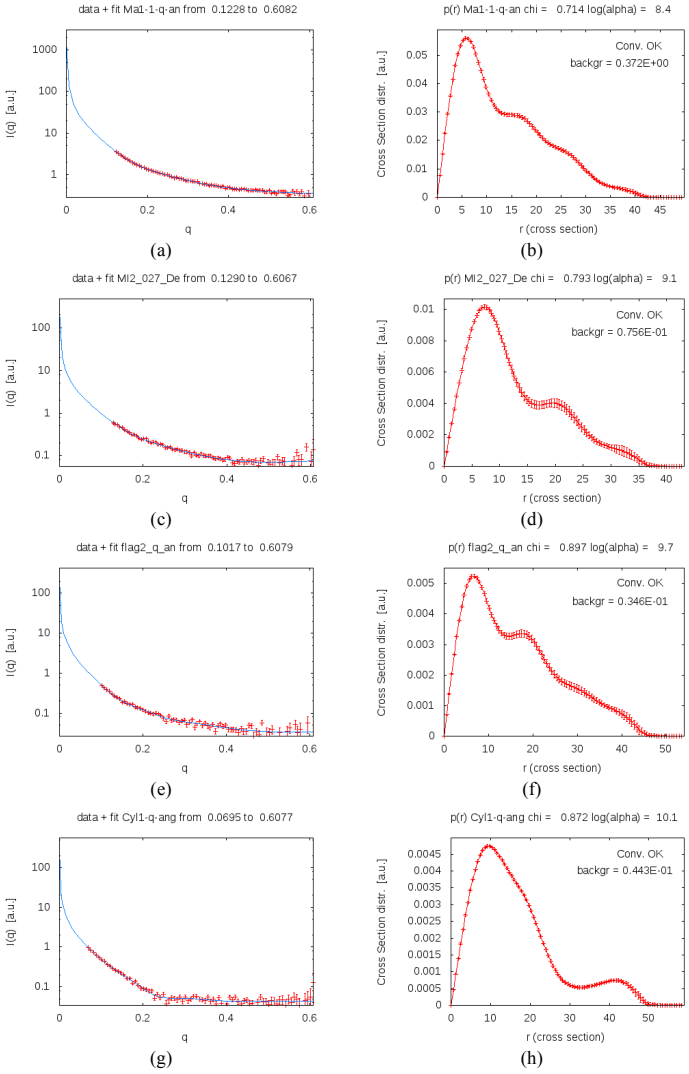


Figure V.13: Cross–sectional $p_c(r)$ fits obtained from BAYES.APP [44]. The first maximum representing the smallest distance in the cylinder cross–section is in good agreement with the radii of the cylinder obtained from WLC.

Table V.6: Summary of data for the four principal spider silks studied

Silk protein	Function	AA–sequence	Structure in solution (CD)	M_w [kDa] (literature)	M_w [kDa] (SAXS)	R_g [nm] from p(r)	D_{max} [nm]	Porod exponent
Major ampullate (MA)	Dragline, structural web	$(GA)_n//(A)_n$ GPGGX/ GPGQQ	Proline II	250 monomer, 550-650 [117] dimer	223 monomer, 530 ±71 dimer	12.95 ± 0.11	50	2.09 ±0.1
Minor ampullate (MI)	Auxiliary spiral	$(GA)_n//(A)_n$	GGX 3_1helix	351 - 255 [10]	36	3.86 ± 0.06	16	1.87 ±0.1
Flagelliform (Flag)	Core fibres of capture spiral	GPGGX// GGX	β–spiral, 3_1helix	360 [40] 157 [protein databank]	≈1200 extrapol. of I(0)	20.05 ± 1.08	> 65	1.96 ±0.1
Cylindriform (Cyl)	Egg sac	$(A)_n//(S)_n//$ $(SA)_n$	GX helix–like	366 [83]	300-321	14.4 ± 0.5	55	1.77 ±0.1

MA consists of two proteins spidroin I and spidroin II with similar molecular weights of 250 kDa [117]. The molecular weight measured by SAXS of 530 ±71 kDa suggests that the protein exhibits a dimer state in H_2O in agreement with the literature [117] (see table V.6). In one sample a lower M_w value of 223 ±22 kDa was found indicating a putative monomeric state of the MA protein. Studies on recombinant silks suggest that at neutral pH (7.2) the monomer state is predominant while at lower pH the proteins are in a dimer state [32, 41]. Given the pH of milliq water used for the dilution series (pH 6.5) a dimeric state of the protein is expected. Since the pH of the solution was not buffered, it is however possible that some solutions had a higher pH and thus the MA protein was also measured in a monomeric state in solution. A monomeric state in solution is also suggested from the SAXS of the Cyl in agreement with the published values of Lin et al. [83]. Similarly to MA, the MI also shows two states of the protein in solution, accompanied with a total change in shape. However one appears to be aggregated since the Kratky plot (figure V.9) shows a fully folded structure supported by the higher folding index of CD. This propensity to aggregate is seen by the difficulties of preparing a non aggregated sample (see also section V.2). Interestingly the obtained M_w of Flag also shows an oligomeric state. The reported M_w of Flag is 360 kDa [40] whereas the M_w estimated from the SAXS data is 1.2 MDa. In addition Flag in contrast to the other silks shows a more complex SAXS curve (figure V.5) and the corresponding $p(r)$ suggests multi–domain or dumbbell shape whereas the other 3 silks just show an anisotropic/ elongated structure (figure V.8). Molecular modeling of silk proteins from their sequence also indicates an anisotropic shape, the

so–called bead–of–string model described by Vollrath and Porter [127]. For Flag in particular the large side chains would favour an extended structure of the protein in solution [21]. The anisotropic shape of the four spider silk proteins is also reflected in the ratio of D_{max}/R_g which is above 3 for all the 4 silks. Such anisotropy is observed in elongated but might also indicate flexibility of the molecules [105].

A further characterization of the folding state of the 4 silk proteins in solution by Kratky analysis shows a dual behaviour in the Kratky plot indicated by a maximum in the plot and an increase in intensity at high angles suggesting an inherent flexibility for these proteins (V.9). The Holtzer analysis reveals that all 4 silk proteins follow the characteristics of a semi–flexible polymer chain as confirmed by fitting to a WLC model. The extracted flexibility index from the WLC fitting shows a negative correlation with the folding index obtained from CD (figure V.12) indicating a decrease in the flexibility of the proteins with increasing folding index. Interestingly the l_p as obtained from the WLC analysis are similar (within errors) for MA, MI and Flag yielding values around 30 Å, while Cyl exhibits a significantly larger l_p of 55 Å. Typically a polypeptide chain adopting random confirmations yield an l_p of 9.35 Å($\approx$ 3 amino acids) [105]. Comparing this value to the obtained l_p's of the four spider silks suggests that silks do not adopt a random coil confirmation upon dilution in H_2O. Taking together the findings from Kratky, Holtzer analysis and WLC fitting silks might be classified as intrinsically disordered proteins (IDP's). The structural contribution from the folded termini regions [4, 40] give rise to the specific CD signal while the rest of the protein chain remains partially or fully disordered. The high

flexibility might also explain the observed random coil like signal in the NMR analysis.

The semi–flexible nature of silk proteins is characterised by specific persistence lengths and numbers of Kuhn length segments. The nature and dimension of the Kuhn segment are generally strongly dependent on the chemical structure and local interaction of the protein chains. Here the length and shape of the cross–section $p_c(r)$ for all 4 silks are remarkably similar (figure V.13). What seems to vary is the dimension of the Kuhn length segments' and the overall shape of the proteins. These in turn depend on the differences in secondary structure content. It is important to remind ourselves that with Kuhn length segments the ratio between the Kuhn length and the cross–sectional radius defines the anisotropy. It is this local anisotropy of the polymeric chain that can result in the appearance of supramolecular structures that will eventually influence the macroscopic properties of the polymer system. Therefore altering the Kuhn length segment properties can translate into the ability to direct supramolecular assembly, and eventually spinning.

VI General Discussion and Conclusions

In this thesis the structure of different silk proteins is investigated using neutron and x–ray small angle scattering (SANS and SAXS) in combination with circular dichroism (CD) and atomic force microscopy (AFM). In chapter III a comparative SANS study of native silk fibroin (SF) and reconstituted silk fibroin (RSF) in solution is presented. A sample preparation protocol had to be established in order to handle the high molecular weight (M_w) proteins without accidental aggregation (chapter II.3). It is found that SF has a higher aggregation propensity in D_2O compared to H_2O and therefore the protein is studied in H_2O. Major differences are observed in the overall shape as well as the R_g's and M_w's between SF and RSF. While SF shows in general no overall change in the shape of the scattering curve, the RSF samples show a large variation when measured by SAS. It is found that the observed differences in the shape of the RSF scattering curves is caused by the variation in the amount of silk fibre dissolved per unit lithium bromide (fibre/LiBr ratio) used for the preparation. A gel electrophoresis reveals a high polydispersity of the RSF proteins in solution (figure III.6) and the presence of large scale aggregates for fibre/LiBr ratios > 2.5 g is revealed by the SAS results. The R_g as well as the apparent M_w are found to be clearly reduced compared to the values obtained for native SF. While the R_g and M_w values for RSF increase as a function of concentration, the SF

proteins show a different behaviour. Over the studied range from 35 down to 5mg/mL the R_g and M_w appear not to change significantly. At protein concentrations of SF below 5 mg/mL however an increase in R_g and M_w is observed accompanied by a decrease in the Porod exponent from about 2.1 down to 1.7 (figure III.10) while RSF appears to show no change in the Porod exponent. CD measurements reveal a loss of structure at low concentrations. Together this might suggest that the SF protein structure at low concentrations is no longer stabilized by charge and this might allow inter–particle interaction reflected in an increase in M_w. This indicates that concentration might play a major role in the storage and assembly of the silk proteins as suggested from CD studies on native spider silk proteins [21].

In order to investigate the effect of concentration on the shape and assembly of silk proteins an atomic force microscopy study was conducted in collaboration with Prof. Schniepps group. The study on native silk proteins clarifies the contradiction about the overall shape of the native silk protein being more globular than rod like (see chapter IV). Globular shapes of 20–25 nm in diameter are observed. Comparing this to the average R_g value for SF found from SAXS a spherical object (chapter II.1.5.2) would yield a diameter of 32 nm which appears not unreasonable given that the AFM samples are dried in air. Interestingly a clear dependence on the concentration is observed in the shear induced assembly of the silk proteins (chapter IV). While at high concentrations a shear induced alignment of the globular particles is found, an aggregation into thin fibrils occurs upon lowering the protein concentrations. This observation might also be explained by an unfolding at low concentrations allowing the proteins to form inter–particle interactions as

suggested from the SAS experiment. Interestingly a periodicity along these thin fibrils is revealed to be in the order of 8 nm with a width of 7 nm. These values are likely overestimated (see chapter II.2 for details), however, interestingly a bead–of–string model for silk with bead sizes in the order of 4 nm along the fibre has been suggested for silks by molecular modeling (see chapter II.2).

Since spider silks provide an ideal model to examine the structure–property relations in native proteins the SAXS and CD studies were extended to include the Major (MA), minor (MI), flagelliform (Flag) and cylindriform (Cyl) silks of *Nephila edulis* (chapter V). The extended $p(r)$, the Kratky and Holtzer analyses reveal that the silks are similar to semi–flexible polymers, as confirmed by fitting to a worm–like chain model. The values obtained from the fitting reveal similarities in the local structure of MA, MI and FLAG silks despite their overall difference in shape and this and the similarities to silk fibroin are explored in more detail below.

Using the WLC model [95] to fit the native SF sample suggests a similar Kuhn length l_k and cross–sectional R_g compared to the native spider silk samples. A reasonable WLC fit of RSF, however, could not be obtained due to the high polydispersity revealed by gel electrophoresis (chapter III). Although the Holtzer analysis cannot yield precise values due to the cross–sectional radius and excluded volume effects, it is shown in chapter V that this type of analysis can be insightful and can be used for qualitative comparison of the q^* and therefore the obtained persistence length values (l_p). Assuming that the contributions from the structure factor are small at high angles a qualitative comparison of the MA with SF and RSF using the Holtzer analysis

becomes possible (figure VI.1). Interestingly MA and SF exhibit a similar q^* of ≈3.5 nm, the two RSF samples in comparison yield a q^* closer to 4 (figure VI.1).

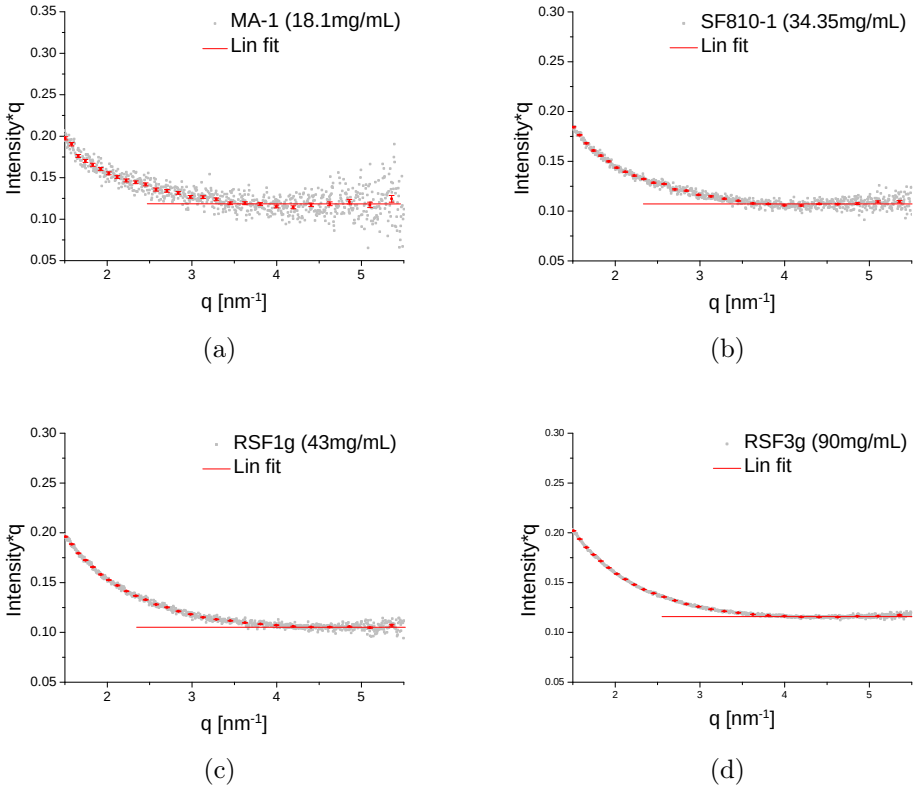


Figure VI.1: Holtzer plots of SF and MA at stock concentrations reveal a similar q^* of about 3.5 nm^{-1}. The two RSF samples in contrast show a higher q^* of 4 nm^{-1}. The grey data represent the un–binned data, the binned data are the red squares. The linear fit for obtaining the plateau (region 4 – 5.5 nm^{-1}, slope set 0) is overlaid by a red line to demonstrate the deviation point.

Although the effect of the high polydispersity of the RSF samples on the overall scattering curve cannot be neglected, the contributions at high angles are likely to be small. Given the methods used for the sample preparation of

standard RSF a difference in the local structure of native SF and RSF would be expected. Spectroscopic techniques like CD or FTIR however show only minor differences in the spectra of the two silk types [17]. The influence of potential different structures (as indicated by the Kratky analysis in chapter III), however, will give an average value for any local length scale parameter like e.g. the Kuhn length l_k. This qualitative comparison of the q^* values of MA, SF and RSF might therefore indicate a difference in local interactions between the protein and its solvent for RSF and native silks. Indeed one of the key findings from the comparison of the Kratky plots for RSF and SF is that RSF appears too solubilized following the processing with LiBr. The lack of folding which means that any pre–ordained structures are absent in the RSF solution may help to explain why it is so difficult to spin a fibre from RSF. This work serves to highlight the importance of studying native, functional proteins at close to *in vivo* conditions in order to understand the way in which Nature processes these large biomacromolecules. Indeed the pre–ordained structures imparted by control of the local environment of the protein chains, as signified by such parameters as the persistence and the Kuhn length, serve to underline how complex hierarchical structures can be achieved.

VI.1 Future Work

One major challenge of this work was to handle the native silk samples without causing any spontaneous aggregation. In order to achieve this, a careful protocol was developed which has proved very successful at generating re-

producible, large quantities of native proteins in solution. The coupling of small angle scattering with spectroscopic techniques provides a valuable insight into the structure of such proteins and it could be valuable to make such measurements simultaneously. This experimentally causes some difficulties especially with respect to the various pathlengths of each technique.

Since reproducible solutions of the silk solutions can now be made, it would be very enlightening to control the ionic strength and pH of the silk solutions. It is already known that certain metal ions cause β–sheet structures to develop within the silk dope and such interactions could be studied in greater detail, again through the application of small angle scattering and fluorescence studies, e.g. EXAFS. In this thesis, dimerisation of the spider silk was observed but the pH of the solutions was not controlled. Clearly this study should be extended to consider the effects of pH on silk.

Much of the work presented in this thesis has used neutron scattering, however, very little deuteration of the proteins or the solvent was achieved without aggregation. Since the association of water to the protein in these hydrogen bonded materials has a major effect on the mechanical properties, it would be very interesting to explore further the selective labelling of either the protein or the solvent in order to gain more details from our small angle scattering experiment. This would also prove beneficial to inelastic neutron scattering measurements.

This thesis compares native and reconstituted silks. In this way, spiders and silkworms are a ready source of large volumes of native, functional proteins. Typically, more research has focussed on readily available proteins such as lysosyme which has lost its functionality through lyophilisation. A major

effort within the Oxford Silk Group has been to explore the biomimetics of silk and silk spinning. I would suggest that even more insights can be gained by examining other naturally occurring biopolymers to see what else we can borrow from Nature.

A Appendix

The content of chapter III and chapter IV is based on the publications Greving et al. 2010 and 2012 [38, 37]. Greving et al. 2010 has been published in Soft Matter and Greving et al. 2012 has been published in Biomacromolecules. Figures A.1 and A.2 show the respective front matters of the publications.

PAPER www.rsc.org/softmatter | Soft Matter

Small angle neutron scattering of native and reconstituted silk fibroin

Imke Greving,†[a] Cedric Dicko,*[a] Ann Terry,[ab] Phil Callow[c] and Fritz Vollrath[a]

Received 15th March 2010, Accepted 24th June 2010
DOI: 10.1039/c0sm00108b

Silks spun by spiders and insects are independently evolved proteinaceous biomaterials with fascinating physical properties attracting scientists from a wide range of disciplines using a wide range of analytical tools. Here we demonstrate the suitability of neutron small angle scattering (SANS) to investigate the morphology and structure of native silk fibroin at near in vivo conditions. Comparing native and reconstituted silk we observed significant differences in sizes, molecular weights, refolding and interactions. These observations question the validity of a presently widespread approach in silk analysis *i.e.* studying reconstituted silk with the goal to gain important insights into the mechanisms involved in the formation (storage and spinning) of native silks.

1. Introduction

The complex self-assembly of biological macromolecules into functional 3-dimensional structures is a highly researched subject area with potential applications ranging from biomedical to high performance structural materials. The complexity, however, of the processes involved can be daunting. A starting point is to consider the balance of forces involved in natural biopolymer self-assembly. Typically the forces responsible for the control and assembly of silk molecules involve van der Waals, electrostatic, steric, entropic and solvation forces, as well as hydrogen-bonding and hydrophobic interactions. The final molecular interplay is, however, not fully understood. Within this context natural silks provide an ideal platform to assess the role of intermolecular, interparticle and surface forces prior to and during fibre formation.

Natural silks are protein-based materials that are extrusion spun at ambient temperatures and pressures from an aqueous solution. The results are fibres that can display exceptional mechanical properties.[1,2] Linking the processing of the material to its mechanical properties has proven to be rather elusive. Yet understanding them will shed some light on important protein-folding prerequisites to achieve macroscopic function.[3,4]

In the last 30 years many research groups worldwide have made important inroads into understanding fibre morphology,[5–7] linking this to the mechanical properties[8–11] and artificial spinning of reconstituted silk.[12] Surprisingly, comparatively less analytical in-depth research has focused on the physical chemistry of the precursor material *i.e.* the raw un-spun silk, as it is stored in the gland. Indeed, due to the limited quantities available and the difficulties in handling these high molecular weight proteins without aggregation, most of the experiments on silk protein folding and assembly have been performed on so-called reconstituted or regenerated silk fibroin (RSF).[12–14] Here we aim to address the limitations and pitfalls associated with the study of reconstituted silk proteins. To achieve this we established and validated a novel protocol using small angle neutron scattering (SANS) to compare the shape and structure relationships of native and reconstituted silks.

Typically the preparation of RSF starts with the degumming of silkworm silk fibres in a soap solution at 70 °C to remove the various sericin coatings. This is followed by the reconstitution or regeneration process during which the clean fibres are dissolved in the presence of a chaotropic agent, such as lithium bromide (LiBr) at temperatures ranging from 60–100 °C. It is remarkable that in spite of a widespread use of the above recipe the effect of LiBr on the silk molecules is not fully understood. Various studies found, for example, a reduction in molecular weight[15] but no measurable changes of behaviour in solution.[13] Additionally, the preparation methods for reconstituted silk vary between research groups[16–18] and often the exact method of preparation is not clearly mentioned in the published studies.[16,18–20] This has led to the publication of a wide range of shapes and sizes to describe silk proteins in solution and consequently, a highly divergent interpretation of how silk proteins fold and assemble into fibres.[12,13,21–24] We focus here on the LiBr route as it is the most popular method, but one should note that other approaches exist with emphasis on milder reconstitution processes.[25]

The effect of preparation on the solution behaviour of silks was clarified in a recent rheological study that demonstrated fundamental differences between the flow behaviour of native silk fibroin (SF) solutions freshly extracted from the silkworm gland and typical RSF solutions.[26] Links to structures, shape, clustering and gel network were hypothesized but yet remain to be demonstrated. Within this context we turned our attention to SANS to evaluate the role and interplay of structures and morphologies in silk proteins behaviour in solution.

We demonstrate that neutron small angle scattering could become an important tool in the unravelling of silk shape and structure relationships. Small angle scattering is commonly used to study the shape, structure and interactions of proteins and polymers in solution using light, X-rays or neutrons as the probe. While light scattering examines very large size structures, both

[a]*University of Oxford, Department of Zoology, Tinbergen building, Oxford, OX1 3PS, UK. E-mail: cedric.dicko@zoo.ox.ac.uk; Tel: +44(0)18650271216*
[b]*ISIS Facility, Science and Technology Facilities Council, Rutherford Appleton Laboratory, Harwell Science and Innovation Campus, Didcot, OX11 0QX, UK*
[c]*Institut Laue Langevin, BP 156, 6 rue Jules Horowitz, 38042 Grenoble, Cedex 9, France*
† Nee: Diddens

Figure A.1: Greving et al. 2010 (front matter) [38], Soft Matter 2010

Article
pubs.acs.org/Biomac

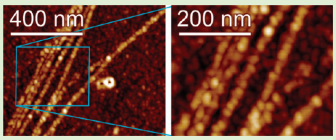
Shear-Induced Self-Assembly of Native Silk Proteins into Fibrils Studied by Atomic Force Microscopy

Imke Greving,[†] Minzhen Cai,[‡] Fritz Vollrath,[†] and Hannes C. Schniepp*[,‡]

[†]Department of Zoology, University of Oxford, The Tinbergen Building, Oxford, OX1 3PS, United Kingdom
[‡]Applied Science Department, The College of William and Mary, Williamsburg, Virginia 23187, United States

Supporting Information

ABSTRACT: Noncontact mode atomic force microscopy was used to investigate native silk proteins prepared in different ways. Low protein concentrations revealed that single protein molecules exhibit a simple, round shape with apparent diameters of 20−25 nm. Shearing the native protein solutions after extraction from the gland and prior to drying led to a beads-on-a-string assembly at the nanometer scale. Protein concentration had a significant effect on the morphology of the protein assemblies. At higher protein concentrations, shear-induced alignment into nanofibrils was observed, while lower concentrations lead to the formation of much thinner fibrils with a width of about 8 nm.



INTRODUCTION

Silk is a natural high performance fiber featuring a rare combination of high strength and high breaking strain. Spun from a protein gel by insects and spiders at ambient temperatures and pressures, this biopolymer outperforms most manmade fibers.[1,2] It is generally believed that the outstanding toughness of silk is due to its sophisticated hierarchical structure involving amorphous and crystalline protein phases and their interplay.[3−6] The ability to artificially make silk with similarly good properties is a desirable goal given the many interesting engineering and biomedical applications of a wide range of natural silks.[7,8] Before synthetic high-performance silk fibers can be developed, however, several major challenges need to be overcome. For example, silk molecular and supramolecular structure needs to be better understood and so do the processes giving rise to the hierarchical assembly of the silk proteins into the final fiber.

Native high-molecular weight silk proteins are relatively difficult to handle. Accordingly, previous studies of silk proteins on a molecular level have mainly focused on regenerated or reconstituted silk (as well as recombinantly produced spider silk peptides) as these materials are more easily handled and prepared.[6−14] However, all of these artificially made silks have substantially different protein molecular weights[15,16] and, moreover, reconstituted and native silk dopes differ significantly in their rheological behavior.[24] Since artificial silks differ in both molecular-scale and macroscopic properties from natural silks it seems imperative to develop a rigorous understanding of the structure of native silk as well as its self-assembly behavior. In this work, we visualize native silk fibroin (SF) from the silk worm *Bombyx mori* at molecular resolutions using atomic force microscopy (AFM). Our results surpass previous studies in image resolution and provide new structural information about silk fibroin. Moreover, we use a sample preparation technique that allows us to visualize the onset of shear-induced, molecular-scale self-assembly of native SF for the first time.

Silkworm silk fibroin consists of three proteins with different molecular weights (M_w): heavy chain fibroin (350 kDa), light chain fibroin (25 kDa), and so-called P25 (30 kDa); these proteins occur in the mass ratios 6:6:1.[17] Concerning the spatial conformation of native SF molecules there is hardly any data in the literature. The most relevant work we found in this respect is a single-molecule AFM study of native *Bombyx mori* SF by Inoue et al. who suggest a rod-like shape for SF,[18] which was also suggested by Viney et al. from birefringence.[19] However, the model that Inoue et al. developed is based on contact-mode AFM images that were acquired in air, where tip−sample interactions lead to significant lateral forces[20] that are strong enough to frequently displace, deform, and disrupt biomolecules.[21] We believe that the data reported by Inoue et al. is significantly distorted through limitations of their imaging approach. Our data, collected with a much gentler imaging technique, suggests that their model needs to be corrected.

For artificial silks, there is structural protein data available; however, the results from different research groups diverge significantly. Two different types of shapes have been suggested for silk proteins: rod-like[22] and micellar/globular.[12,23] Moreover, there is a substantial spread in the reported protein sizes, ranging from a few to several hundred nanometers. This is in line with the observed differences in molecular weight,[14−16] which are due to variations in the preparation methods for these artificial silks between research groups, using different chaotropic agents and so on.[15,16,23,24] This uncertainty and spread associated with artificial silk in terms of protein mass

Received: October 26, 2011
Revised: December 24, 2011
Published: February 21, 2012

Figure A.2: Greving et al. 2012 (front matter) [37], Biomacromolecules 2012

Bibliography

[1] T Asakura. Structure of *Bombyx mori* silk fibroin in aqueous solution. *Die Makromolekulare Chemie*, 7:755–759, 1986.

[2] T Asakura, H Suzuki, and Y Watanabe. Conformational characterization of silk fibroin in intact *Bombyx mori* and pilosamia cynthia ricini silkworms by carbon-13 nmr spectroscopy. *Macromolecules*, 16:1024–1026, 1983.

[3] T Asakura, T Yamane, Y Nakazawa, T Kameda, and K Ando. Structure of *Bombyx mori* silk fibroin before spinning in solid state studied with wide angle x-ray. *Biopolymers*, 58(5):521–525, 2001.

[4] G Askarieh, M Hedhammar, K Nordling, A Saenz, C Casals, A Rising, J Johansson, and S D Knight. Self-assembly of spider silk proteins is controlled by a ph-sensitive relay. *Nature*, 465(7295):236–U125, 2010.

[5] G Beaucage, HK Kammler, and SE Pratsinis. Particle size distributions from small-angle scattering using global scattering functions. *Journal of Applied Crystallography*, 37:523–535, 2004.

[6] P Bernado and D I Svergun. Structural analysis of intrinsically disordered proteins by small-angle x-ray scattering. *Mol Biosyst*, 8(1):151–167, 2012.

[7] G Binnig, H Rohrer, and C Gerber. Surface studies by scanning tunneling microscopy. *Physical review letters*, 1982.

[8] M Boulet-Audet, F Vollrath, and C Holland. Rheo-attenuated total reflectance infrared spectroscopy: a new tool to study biopolymers. *Physical Chemistry Chemical Physics*, 13(9):3979–3984, 2011.

[9] X Chen, Z Shao, D Knight, and F Vollrath. Conformation transition kinetics of *Bombyx mori* silk protein. *Proteins: Structure*, 68:223–231, 2007.

[10] M Colgin. Spider minor ampullate silk proteins contain new repetitive sequences and highly conserved non spacer regions. *Protein Sci*, 7:667–672, 1998.

[11] C Craig. Evolution of arthropod silks. *Annual review of entomology*, 42:231–267, 1997.

[12] C Craig. *Tracing Evolution from Molecules to Genes to Phenotypes*. Oxford University Press, 2003.

[13] C Craig, M Hsu, D Kaplan, and N Pierce. A comparison of the composition of silk proteins produced by spiders and insects. *International journal of biological macromolecules*, 24:109–118, 1999.

[14] C Craig and C Riekel. Comparative architecture of silks, fibrous proteins and their encoding genes in insects and spiders. *Comparative Biochemistry and Physiology*, 133:493–507, 2002.

[15] P Denkinger. Determination of chain stiffness and polydispersity from static light scattering. *Journal of Polymer Science Part B:*, 29:589–600, 1991.

[16] C Dewhurst. Grasp software, August 2009. http://www.ill.eu/instruments-support/instruments-groups/groups/lss/grasp/.

[17] C Dicko, 2012. Personal communication.

[18] C Dicko, JM Kenney, D Knight, and F Vollrath. Transition to a beta-sheet-rich structure in spidroin in vitro: The effects of ph and cations. *Biochemistry-Us*, 43(44):14080–14087, 2004.

[19] C Dicko, JM Kenney, and F Vollrath. beta-silks: Enhancing and controlling aggregation. *Adv Protein Chem*, 73:17–+, 2006.

[20] C Dicko, D Knight, J Kenney, and F Vollrath. Structural conformation of spidroin in solution: a synchrotron radiation circular dichroism study. *Biomacromolecules*, 5:758–767, 2004.

[21] C Dicko, D Knight, JM Kenney, and F Vollrath. Secondary structures and conformational changes in flagelliform, cylindrical, major, and minor ampullate silk proteins. temperature and concentration effects. *Biomacromolecules*, 5(6):2105–2115, 2004.

[22] C Dicko, D Porter, J Bond, JM Kenney, and F Vollratht. Structural disorder in silk proteins reveals the emergence of elastomericity. *Biomacromolecules*, 9(1):216–221, 2008.

[23] C Dicko, F Vollrath, and JM Kenney. Spider silk protein refolding is controlled by changing ph. *Biomacromolecules*, 5(3):704–10, 2004.

[24] S Doniach. Changes in biomolecular conformation seen by small angle x-ray scattering. *Chem Rev*, 101(6):1763–1778, 2001.

[25] N Du, Z Yang, XY Liu, Y Li, and HY Xu. Structural origin of the strain-hardening of spider silk. *Advanced Functional Materials*, 21(4):772–778, 2011.

[26] P Eaton and P West. *NY–MDT Image Analysis Reference Manual*. NY–MDT Co, 2006. http://www.ntmdt.com.

[27] P Eaton and P West. *Atomic Force Microscopy*. Oxford University Press, 2010.

[28] H Fischer, I Polikarpov, and AF Craievich. Average protein density is a molecular-weight-dependent function. *Protein Sci*, 13(10):2825–2828, 2004.

[29] J Fitter, T Gutberlet, and J Katsaras. *Neutron Scattering in Biology*. Springer, 2006.

[30] C Foo, E Bini, J Hensman, D Knight, and R Lewis. Role of ph and charge on silk protein assembly in insects and spiders. *Applied Physics A: Materials Science & Processing*, 82:223–233, 2006.

[31] F Gabel, M Jensen, and G Zaccaï. Quantitative modelfree analysis of urea binding to unfolded ubiquitin using a combination of small angle x-ray and neutron scattering. *Journal of the American*, 131:8769–8771, 2009.

[32] WA Gaines, MG Sehorn, and WR Marcotte. Spidroin n-terminal domain promotes a ph-dependent association of silk proteins during self-assembly. *Journal of Biological Chemistry*, 285(52):–, 2010.

[33] R Garcia and R Perez. Dynamic atomic force microscopy methods. *Surf Sci Rep*, 47(6-8):197–301, 2002.

[34] O Glatter. A new method for the evaluation of small-angle scattering data. *Journal of Applied Crystallography*, 10:415–421, 1977.

[35] O Glatter and O Kratky. *Small Angle X-ray Scattering*. Academic Press, 1982.

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

[37] I Greving, M Cai, F Vollrath, and HC Schniepp. Shear-induced self-assembly of native silk proteins into fibrils studied by atomic force microscopy. *Biomacromolecules*, 13:676–682, Feb 2012.

[38] I Greving, C Dicko, AE Terry, P Callow, and F Vollrath. Small angle neutron scattering of native and reconstituted silk fibroin. *Soft Matter*, 6(18):4389–4395, 2010.

[39] I Grillo. *Small-Angle Neutron Scattering and Applications in Soft Condensed Matter*. Springer Netherlands, 2008.

[40] F Hagn, L Eisoldt, JG Hardy, C Vendrely, M Coles, T Scheibel, and H Kessler. A conserved spider silk domain acts as a molecular switch that controls fibre assembly. *Nature*, 465(7295):239–U131, 2010.

[41] F Hagn, C Thamm, T Scheibel, and H Kessler. ph-dependent dimerization and salt-dependent stabilization of the n-terminal domain of spider dragline silk-implications for fiber formation. *Angew Chem Int Edit*, 50(1):310–313, 2011.

[42] O Hakimi, DP Knight, F Vollrath, and P Vadgama. Spider and mulberry silkworm silks as compatible biomaterials. *Compos Part B-Eng*, 38(3):324–337, 2007.

[43] B Hammouda. *Probing nanoscale structures - The SANS Toolbox*. Online available, 2008.

[44] S Hansen. Bayesian estimation of hyperparameters for indirect fourier transformation in small-angle scattering. *Journal of Applied Crystallography*, 33:1415–1421, 2000.

[45] HG Hansma, J Vesenka, C Sirgerist, G Kelderman, H Mmorrett, RL Sinsheimer, V Elings, C Bustamante, and PK Hansma. Reproducible imaging and dissection of plasmid dna under liquid with the atomic force microscope. *Science*, 256(5060):1180–1184, 1992.

[46] JG Hardy, LM Roemer, and TR Scheibel. Polymeric materials based on silk proteins. *Polymer*, 49(20):4309–4327, 2008.

[47] C Hayashi, N Shipley, and R Lewis. Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins. *International journal of biological macromolecules*, 24:271–275, 1999.

[48] D Hijirida, K Do, C Michal, S Wong, and D Zax. 13c nmr of *Nephila clavipes* major ampullate silk gland. *Biophysical Journal*, 71:3442–3447, 1996.

[49] C Holland, AE Terry, D Porter, and F Vollrath. Comparing the rheology of native spider and silkworm spinning dope. *Nature Materials*, 5(11):870–874, 2006.

[50] C Holland, AE Terry, D Porter, and F Vollrath. Natural and unnatural silks. *Polymer*, 48(12):3388–3392, 2007.

[51] C Holland and J Urbach. Direct visualization of shear dependent silk fibrillogenesis. *Soft Matter*, 8:2590, 2012.

[52] C Holland and F Vollrath. *Biomimetic principles of spider silk for high performance fibres. Chapter 7.* A.G. Woodhead Publishing, 2008.

[53] C Holland, F Vollrath, AJ Ryan, and OO Mykhaylyk. Silk and synthetic polymers: Reconciling 100 degrees of separation. *Advanced materials*, Nov 2011.

[54] A Holtzer. Interpretation of the angular distribution of the light scattered by a polydisperse system of rods. *Journal of Polymer Science*, 17(85):432–434, 1955.

[55] M Hronska, JD van Beek, PTF Williamson, F Vollrath, and BH Meier. Nmr characterization of native liquid spider dragline silk from nephila edulis. *Biomacromolecules*, 5(3):834–839, 2004.

[56] H Hsu and W Paul. Standard definitions of persistence length do not describe the local "intrinsic" stiffness of real polymer chains. *Macromolecules*, 2010.

[57] X Hu, D Kaplan, and P Cebe. Determining beta-sheet crystallinity in fibrous proteins by thermal analysis and infrared spectroscopy. *Macromolecules*, 39(18):6161–6170, 2006.

[58] T Huang, P Ren, and B Huo. Atomic force microscopy of the topography of regenerated observations of the silk fibroin aggregations. *J Appl Polym Sci*, 106(6):4054–4059, 2007.

[59] W Huang, R Begum, T Barber, V Ibba, NCH Tee, M Hussain, M Arastoo, Q Yang, LG Robson, S Lesage, T Gheysens, NJV Skaer, DP Knight, and JV Priestley. Regenerative potential of silk conduits in repair of peripheral nerve injury in adult rats. *Biomaterials*, 33(1):59–71, 2012.

[60] N Van Hulst, M Garcia-Parajo, and M Moers. Near-field fluorescence imaging of genetic material: toward the molecular limit. *Journal of structural*, 1997.

[61] S Inoue, J Magoshi, T Tanaka, Y Magoshi, and M Becker. Atomic force microscopy: Bombyx mori silk fibroin molecules and their higher

order structure. *Journal of Polymer Science Part B: Polymer Physics*, 38(11):1436–1439, 2000.

[62] S Inoue, K Tanaka, F Arisaka, S Kimura, and K Ohtomo. Silk fibroin of *Bombyx mori* is secreted, assembling a high molecular mass elementary unit consisting. *Journal of Biological Chemistry*, 275(51):40517–40528, 2000.

[63] Y Iridag and M Kazanci. Preparation and characterization of *Bombyx mori* silk fibroin and wool keratin. *J Appl Polym Sci*, 100(5):4260–4264, 2006.

[64] B Jacrot. The study of biological structures by neutron scattering from solution. *Reports on Progress in Physics*, 39:911–953, 1976.

[65] B Jacrot and G Zaccai. Determination of molecular-weight by neutron-scattering. *Biopolymers*, 20(11):2413–2426, 1981.

[66] C Jiang, X Wang, R Gunawidjaja, Y-H Lin, MK Gupta, DL Kaplan, RR Naik, and VV Tsukruk. Mechanical properties of robust ultrathin silk fibroin films. *Advanced Functional Materials*, 17(13):2229–2237, 2007.

[67] HJ Jin and DL Kaplan. Mechanism of silk processing in insects and spiders. *Nature*, 424(6952):1057–1061, 2003.

[68] F Junghans, M Morawietz, U Conrad, T Scheibel, A Heilmann, and U Spohn. Preparation and mechanical properties of layers made of

recombinant spider silk proteins and silk from silk worm. *Appl Phys A-Mater*, 82(2):253–260, 2006.

[69] A Kiriy, G Gorodyska, S Minko, W Jaeger, P Stepanek, and M Stamm. Cascade of coil-globule conformational transitions of single flexible polyelectrolyte molecules in poor solvent. *Journal of the American Chemical Society*, 124(45):13454–13462, 2002.

[70] D Knight, 2009. Personal communication.

[71] DP Knight and F Vollrath. Changes in element composition along the spinning duct in a *Nephila* spider. *Naturwissenschaften*, 88(4):179–182, Jun 2001.

[72] DP Knight and F Vollrath. Biological liquid crystal elastomers. *Philos T Roy Soc B*, 357(1418):155–163, 2002.

[73] MHJ Koch, P Vachette, and DI Svergun. Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Q Rev Biophys*, 36(2):147–227, 2003.

[74] PV Konarev, VV Volkov, AV Sokolova, MHJ Koch, and DI Svergun. Primus: a windows pc-based system for small-angle scattering data analysis. *Journal of Applied Crystallography*, 36:1277–1282, 2003.

[75] I Krasnov, I Diddens, N Hauptmann, G Helms, M Ogurreck, T Seydel, SS Funari, and M Müller. Mechanical properties of silk: Interplay of deformation on macroscopic and molecular length scales. *Physical review letters*, 100(4):048104, 2008.

[76] T Krishnaji, W Huang, O Rabotyagova, E Kharlampieva, I Choi, VV Tsukruk, R Naik, P Cebe, and DL Kaplan. Thin film assembly of spider silk-like block copolymers. *Langmuir*, 27(3):1000–1008, 2011.

[77] C Krywka. Dissertation, 2008. http://hdl.handle.net/2003/25184/.

[78] AS Lammel, X Hu, S-H Park, DL Kaplan, and TR Scheibel. Controlling silk fibroin particle features for drug delivery. *Biomaterials*, 31(16):4583–91, Jun 2010.

[79] M Landreh, G Askarieh, K Nordling, M Hedhammar, A Rising, C Casals, J Astorga-Wells, G Alvelius, SD Knight, J Johansson, H Joernvall, and T Bergman. A ph-dependent dimer lock in spider silk protein. *Journal of molecular biology*, 404(2):328–336, 2010.

[80] S Lecommandoux, F Checot, R Borsali, M Schappacher, A Deffieux, A Brulet, and JP Cotton. Effect of dense grafting on the backbone conformation of bottlebrush polymers: Determination of the persistence length in solution. *Macromolecules*, 35(23):8878–8881, 2002.

[81] J Lees, B Smith, F Wien, A Miles, and B Wallace. Cdtool—an integrated software package for circular dichroism spectroscopic data. *Analytical biochemistry*, 332:285–289, 2004.

[82] X Li, L Wu, M Huang, H Shao, and X Hu. Conformational transition and liquid crystalline state of regenerated silk fibroin in water. *Biopolymers*, 89(6):497–505, 2008.

[83] Z Lin, W Huang, J Zhang, J-S Fan, and D Yang. Solution structure of eggcase silk protein and its implications for silk fiber formation. *P Natl Acad Sci Usa*, 106(22):8906–8911, 2009.

[84] P Lindner and T Zemp. *Neutrons, X-rays and Light: Scattering Methods Applied to Soft Condensed Matter*. Elsevier, 2002.

[85] J Magoshi, Y Magoshi, and S Nakamura. Mechanism of fiber formation of silkworm, 1994.

[86] A Martel, M Burghammer, R Davies, and E DiCola. A microfluidic cell for studying the formation of regenerated silk by synchrotron radiation small-. *Biomicrofluidics*, 2:024104–1, 2008.

[87] A Martel, M Burghammer, RJ Davies, E Di Cola, C Vendrely, and C Riekel. Silk fiber assembly studied by synchrotron radiation saxs/waxs and raman spectroscopy. *Journal of the American Chemical Society*, 130(50):17070–17074, 2008.

[88] A Matsumoto, J Chen, AL Collette, U-J Kim, GH Altman, P Cebe, and DL Kaplan. Mechanisms of silk fibroin sol-gel transitions. *J Phys Chem B*, 110(43):21630–21638, 2006.

[89] ER McCarney, JE Kohn, and KW Plaxco. Is there or isn't there? the case for (and against) residual structure in chemically denatured proteins. *Crit Rev Biochem Mol Biol*, 40(4):181–9, 2005.

[90] HDT Mertens and DI Svergun. Structural characterization of proteins and complexes using small-angle x-ray solution scattering. *J Struct Biol*, 172(1):128–141, 2010.

[91] E Meyer. Atomic force microscopy. *Progress in surface science*, 41:3–49, 1992.

[92] E Mylonas and DI Svergun. Accuracy of molecular mass determination of proteins in solution by small-angle x-ray scattering. *Journal of Applied Crystallography*, 40:S245–S249, 2007.

[93] S Van Noort, K Van der Werf, and B De Grooth. Height anomalies in tapping mode atomic force microscopy in air caused by adhesion. *Ultramicroscopy*, 1997.

[94] E Oroudjev, J Soares, S Arcdiacono, JB Thompson, SA Fossey, and HG Hansma. Segmented nanofibers of spider dragline silk: Atomic force microscopy and single-molecule force spectroscopy. *P Natl Acad Sci Usa*, 99:6460–6465, 2002.

[95] JS Pedersen and P Schurtenberger. Scattering functions of semiflexible polymers with and without excluded volume effects. *Macromolecules*, 29(23):7602–7612, 1996.

[96] P Pernot, P Theveneau, and T Giraud. New beamline dedicated to solution scattering from biological macromolecules at the esrf. *Journal of Physics:*, 247:012009, 2010.

[97] L Pietrasanta, A.Schaper, and T Jovin. Probing specific molecular conformations with the scanning force microscope. complexes of plasmid dna and anti-z-dna antibodies. *Nucleic Acids Research*, 22(16):3288–3292, 1994.

[98] T Pirzer, M Geisler, T Scheibel, and T Hugel. Single molecule force measurements delineate salt, ph and surface effects on biopolymer adhesion. *Phys Biol*, 6(2):025004, 2009.

[99] D Porter and F Vollrath. Silk as a biomimetic ideal for structural polymers. *Adv Mater*, 21(4):487–492, 2009.

[100] D Porter, F Vollrath, and Z Shao. Predicting the mechanical properties of spider silk as a model nanostructured polymer. *Eur Phys J E*, 16(2):199–206, 2005.

[101] P Poza, J Perez-Rigueiro, M Elices, and J Llorca. Fractographic analysis of silkworm and spider silk. *Eng Fract Mech*, 69(9):1035–1048, 2002.

[102] CD Putnam, M Hammel, GL Hura, and JA Tainer. X-ray solution scattering (saxs) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys*, 40(3):191–285, Aug 2007.

[103] S Putthanarat, N Stribeck, S Fossey, and R Eby. Investigation of the nanofibrils of silk fibers. *Polymer*, 41:7735–7747, 2000.

[104] S Rauscher, S Baud, M Miao, FW Keeley, and R Pomes. Proline and glycine control protein self-organization into elastomeric or amyloid fibrils. *Structure*, 14(11):1667–1676, 2006.

[105] V Receveur-Brechot and D Durand. How random are intrinsically disordered proteins? a small angle scattering perspective. . . . *Protein and Peptide Science*, 2012.

[106] C Riekel, M Muller, and F Vollrath. In situ x-ray diffraction during forced silking of spider silk. *Macromolecules*, 32:4464–4466, 1999.

[107] C Riekel and F Vollrath. Spider silk fibre extrusion: combined wide-and small-angle x-ray microdiffraction experiments. *International journal of biological macromolecules*, 29:203–210, 2001.

[108] KA Rubinson, C Stanley, and S Krueger. Small-angle neutron scattering and the errors in protein structures that arise from uncorrected background and intermolecular interactions. *Journal of Applied Crystallography*, 41:456–465, 2008.

[109] S Santos, V Barcons, HK Christenson, J Font, and NH Thomson. The intrinsic resolution limit in the atomic force microscope. *Plos One*, 6(8):e23821, 2011.

[110] D Sapede, T Seydel, VT Forsyth, MA Koza, R Schweins, F Vollrath, and C Riekel. Nanofibrillar structure and molecular mobility in spider dragline silk. *Macromolecules*, 38(20):8447–8453, 2005.

[111] F Sehnal and M Zurovec. Construction of silk fiber core in lepidoptera. *Biomacromolecules*, 5(3):666–74, 2004.

[112] T Seydel, K Koelln, I Krasnov, I Diddens, N Hauptmann, G Helms, M Ogurreck, S-G Kang, MM Koza, and M Müller. Silkworm silk under tensile strain investigated by synchrotron x-ray diffraction and neutron spectroscopy. *Macromolecules*, 40(4):1035–1042, 2007.

[113] H Shulha, CWP Foo, DL Kaplan, and VV Tsukruk. Unfolding the multi-length scale domain structure of silk fibroin protein. *Polymer*, 47(16):5821–5830, 2006.

[114] S Singamaneni and V Tsukruk. Probing soft matter with the atomic force microscopies: Imaging and force spectroscopy. *Polym Rev*, 50:235–286, 2010.

[115] R Sinibaldi, M.G Ortore, F Spinozzi, S de Souza Funari, J Teixeira, and P Mariani. Sans/saxs study of the bsa solvation properties in aqueous urea solutions via a global fit approach. *Eur Biophys J*, 37(5):673–681, 2008.

[116] A Sponner, B Schlott, F Vollrath, E Unger, F Grosse, and K Weisshart. Characterization of the protein components of *Nephila clavipes* dragline silk. *Biochemistry-Us*, 44(12):4727–4736, 2005.

[117] A Sponner, W Vater, W Rommerskirch, F Vollrath, E Unger, F Grosse, and K Weisshart. The conserved c-termini contribute to the properties of spider silk fibroins. *Biochem Bioph Res Co*, 338(2):897–902, 2005.

[118] GL Squires. *Introduction to the Theofy of Thermal Neutron Scattering*. John Wiley, 1996.

[119] D Svergun and M Koch. Small-angle scattering studies of biological macromolecules in solution. *Reports on Progress in Physics*, 66:1735–1782, 2003.

[120] D Svergun and A Semenyuk. Small-angle-scattering-data treatment by the regularization method. *Crystallographica Section A*, A44:244–250, 1988.

[121] DI Svergun. Determination of the regularization parameter in indirect - transform methods using perceptual criteria. *Journal of Applied Crystallography*, 25:495–503, 1992.

[122] Y Tashiro and E Otsuki. Studies on posterior silk gland of silkworm bombix-mori .4. ultracentrifugal analyses of native silk proteins, especially fibroin extracted from middle silk gland of mature silkworm. *Journal of Cell Biology*, 46(1):1–&, 1970.

[123] Y Termonia. Molecular modeling of spider silk elasticity. *Macromolecules*, 27(25):7378–7381, 1994.

[124] C Viney. Natural silks: archetypal supramolecular assembly of polymer fibres. *Supramolecular Science*, 4(1-2):75–81, 1997.

[125] F Vollrath, T Holtet, HC Thogersen, and S Frische. Structural organization of spider silk. *P Roy Soc Lond B Bio*, 263(1367):147–151, 1996.

[126] F Vollrath and DP Knight. Liquid crystalline spinning of spider silk. *Nature*, 410(6828):541–548, 2001.

[127] F Vollrath and D Porter. Spider silk as archetypal protein elastomer. *Soft Matter*, 2(5):377–385, 2006.

[128] F Vollrath, D Porter, and C Holland. There are many more lessons still to be learned from spider silks. *Soft Matter*, 7(20):9595–9600, 2011.

[129] CE Williams, RP May, and A Guinier. *Small Angle Scattering of X-rays and Neutrons*. John Wiley, 1993.

[130] H Yamada, H Nakao, Y Takasu, and K Tsubouchi. Preparation of undegraded native molecular fibroin solution from silkworm cocoons. *Mat Sci Eng C-Bio S*, 14(1-2):41–46, 2001.

[131] Z Yang, DT Grubb, and LW Jelinski. Small-angle x-ray scattering of spider dragline silk. *Macromolecules*, 30(26):8254–8261, 1997.

[132] C Zhao and T Asakura. Structure of silk studied with nmr. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 39:301–352, 2001.

[133] HP Zhao, XQ Feng, and HJ Shi. Variability in mechanical properties of *Bombyx mori* silk. *Materials Science and Engineering: C*, 27(4):675–683, 2007.

[134] C Zhou, F Confalonieri, N Medina, and Y Zivanovic. Fine organization of *Bombyx mori* fibroin heavy chain gene. *Nucleic Acids Research*, 12:2413–2419, 2000.

[135] P Zhou, X Xie, DP Knight, X-H Zong, F Deng, and W-H Yao. Effects of ph and calcium ions on the conformational transitions in silk fibroin using 2d raman correlation spectroscopy and 13c solid-state nmr. *Biochemistry-Us*, 43(35):11302–11, Sep 2004.

[136] YT Zhou, SX Wu, and VP Conticello. Genetically directed synthesis and spectroscopic analysis of a protein polymer derived from a flagelliform silk sequence. *Biomacromolecules*, 2(1):111–125, 2001.

[137] J Zhu, Y Zhang, H Shao, and X Hu. Electrospinning and rheology of regenerated *Bombyx mori* silk fibroin aqueous solutions: The effects of ph and concentration. *Polymer*, 49(12):2880–2885, 2008.

[138] ZH Zhu, K Ohgo, and T Asakura. Preparation and characterization of regenerated *Bombyx mori* silk fibroin fiber with high strength. *Express Polym Lett*, 2(12):885–889, 2008.

[139] B Zuo, L Dai, and Z Wu. Analysis of structure and properties of biodegradable regenerated silk fibroin fibers. *J Mater Sci*, 41(11):3357–3361, 2006.

[140] E Mylonas, D Svergun. Accuracy of molecular mass determination of proteins in solution by small–angle X–ray scattering. *Journal of Applied Crystallography*, 40:S245–S249, 2007.