UNIVERSITY OF SHEFFIELD

DOCTORAL THESIS

The University Of Sheffield.

Towards Patterned Protein Based Nanoparticle Arrays

Author:

Lukas Jasaitis

Supervisor:

Dr Sarah S. Staniland

Prof Jennifer R. Potts

Prof Steven D. Johnson

A thesis submitted in fulfillment of the requirements

for the degree of Doctor of Philosophy

in the

Department or Chemistry

January 30th, 2021

"There is no need to build a labyrinth when the entire universe is one."

— Jorge Luis Borges, Labyrinths

Abstract

To meet the exponential growth in the demand for data storage, capacity innovations in hard drive disk (HDD) technologies must be sought. A bit-pattern media (BPM) approach is the frontier of HDD storage technology. It consists of metallic nano-scale patterns of less than 10 nm with 20 nm periodicity and has the potential to significantly increase the data storage density, where each patterned island acts as a single bit. The conventional nano-scale manufacturing methods have so far failed to produce such patterns at scale. There are proteins in nature that can self-assemble into nano-patterns at the desired feature sizes without high energy input under ambient conditions. It is proposed here that a bio-inspired approach could provide useful input to the BPM field by constructing a nano-scale pattern using protein self-assembly exclusively. The project looks to the field of magnetic nanoparticle biomineralisation with attempts to functionalise the protein constructed patterns.

The work here characterises a known, unusually rigid rod-like protein domain G52-E-G53 (originating from the SasG protein found in *Staphyloccocus aureus*) as a building block for protein pattern construction by creating extended linear assemblies through genetic fusion with two orthogonal pairs of coiled-coil forming α-helices. The assembly is validated through a robust series of biophysical and microscopy investigations. The G52-E-G53 coiled-coil constructs are also granted the ability to specifically interact with magnetite nanoparticles through a fusion with a known magnetite binding peptide. Initial work is also performed to create two-dimensional patterns using G52-E-G53 as a rigid linker. The work also shows strong evidence that a magnetite binding peptide and iron nucleation membrane protein Mms6 can be used as genetic fusions to impart improved biomineralisation/particle binding capacity to a well characterised S-layer protein SgsE. Modularity of a different magnetite binding peptide is also demonstrated with respect to different loop displaying scaffold proteins. Lastly, phage display is used to discover three new seven amino acid peptides for CoPt $L1_0$ nanoparticle binding and biomineralisation with one of the peptides showing evidence of improving magnetic properties of CoPt chemical synthesis products.

Together this multi-pronged approach demonstrates substantial leads that could be the basis of a field of magnetic mineral templating on protein patterns using specific magnetic nanoparticle binding moieties.

Acknowledgements

First of all I would like to thank Sarah Staniland for offering me the position those five long years ago. I appreciate that greatly and I am very glad that I accepted. I have learned a great deal of science over these years and knowing how to make nanomagnets is more than a biologist like me could ever have asked for. I learned more than just how to be a scientist though, I also received great insights into how things work in academia and life in general. Sarah taught me how to talk to seek out collaborations and create 'win-win' relationships in the world of science! Thanks for all the encouragement to undertake extracurricular activities and develop myself during this time.

I don't think this section would be complete without mentioning all the training I received from Andrea Rawlings. I started this PhD having never successfully purified a protein and being the strange person who enjoyed the 'cloning-bit' more. A couple of months in I was purifying proteins and optimising my own protocols. A similar progression could be described for any other lab skill and I am grateful to Andrea for that.

Thanks to Jen and Steve for being great co-supervisors! And thanks to Steve's group for being so welcoming during my visits to York for QCM-D.

It is incredible, how many people you meet in four years, I could write a book each for most of these relationships, but I have to keep it short. Thanks, Sam, for not being afraid to dream big in your scientific pursuits. Thank you, Laura, for all of the conversations about PhD anxieties and looking on the bright side! Thank you, Chris, for financial advice. Thank you, Lori, for your great taste in music. Thanks, Rosie, for exemplifying great work ethic! Thanks, Zainab, for showing me how you culture magnetic bacteria. Thanks Reuben for a like mind and a great taste in music as well. Thanks to Amelia for always listening to my phage display questions. Thanks to Ben for python and the mug.

Thanks to all of the post doctoral staff whose job was not to teach me, but who none the less showed me expert tips on how to perform science well, especially Dave Finger and Chatchawal Phansopa. Thanks to Ciani group for all the great biochemistry and CD training.

Thanks to all my students Molly, Tinglu and Simran for listening to me talk all day over-explaining everything and for being a great help.

Joe Gaunt, thank you for the business mentorship!

Thank you Dimitrios for making my dream to do iGEM come true and for staying an amazing friend who challenges my opinions to this day! Thank you, Mat, for staying my friend even after moving!

Thank you Laia for being an inspiring scientist and a great AFM teacher and for your help with the paper on the protein fibres!

Everyone at Singer instruments. And especially Jose, who stayed a dear friend.

Thanks to Adam and Tim, and all the random encounters at the climbing wall.

Thanks to Rytis who really came through with proof-reading, patiently parsing through my esoteric text. Veronica and Jose, thanks for pointing out last minute mistakes and for the nice words about my writing!

Thanks to Erikas and Goda for all the great chats, trips to london and life advice (even if I don't take any of it, I will always appreciate hearing it - never change!). Thanks to Julija for keeping me healthy through conversations and dietary advice!

I really appreciate the patience of my housemate Kristle, thank you for all the great conversations and thank you for sharing with us your great baking. And thank you Keelan for the science-philosophy chats and the room when I needed it! Many

thanks to Matthew for being curious. And thanks to Igors for returning my excitement for Science through his sheer existence.

Thank you to everyone else I missed. I am sure I did!

I probably would not have been able to finish this PhD and the thesis without the support of my dear Iris. You really held my hand through all of this! When I was at my lowest you showed me colour and built me back up. And I love you for that! Looking forward to what is next for us.

Great thanks to all of my family for support. Laura, Rolandai, Laima, Rimai, Martynai, Rasa, Mariau, Irena, Močiute, Tėti, you have all been great. Thanks to my brother Laurynas and his wife Egle for always checking in and making sure I am not overworking! Thank you Jonas, Urte, Jokubas, Augustas, Meda and Eva for being great sources of joy.

My life's worth of gratitude to my mother Daiva. She brought me up with enough curiosity and provided care and support through my formative years to create the conditions for me to learn and grow. In more ways than one, I would not be here without her.

Declaration of Authorship

I, Lukas Jasaitis, declare that this thesis titled: "Towards Patterned Protein Based Nanoparticle Arrays" and the work presented in it are my own. I confirm that except for where specific references have been made to other sources the work in this thesis is the original work of the author and has not been submitted wholly or in part for another degree.

Chapter 3

Initial SasG G52-E-G53 constructs with coiled-coil pairs AN4/BN4 and P5/P6 for linear were designed by Dr Dan Peters from University of York who also performed preliminary circular dichroism and non-denaturing polyacrylamide electrophoresis experiments (results not included in this thesis). ESI-TOF analysis was performed by Sharon Spey and Simon Thorpe (Faculty of Science Mass Spectrometry Centre). QCM-D experiments were performed with the aid of Callum Silver at the University of York. Atomic force microscopy results were obtained in consultation (and in a few cases in collaboration) with Laia Pasquina-Lemonche (The University Of Sheffield, Department of Physics). Some of the 30-50 nm magnetite samples were synthesised by Dr Andrea Rawlings.

Chapter 4

The HR-add, HR-anchor, Ank4-add and Ank4-anchor amino acid sequences were designed and computationally assessed by Fabio Parmeggiani (University of Bristol) who also proposed the strategy for the on-surface assembly. ESI-TOF analysis was performed by Sharon Spey and Simon Thorpe (Faculty of Science Mass Spectrometry Centre). The first LSPR experiment was performed with the help of Dr Mahmoud Abdelhamid (Tim Craggs group, The University of Sheffield). Transmission electron microscopy was performed at the Electron Microscopy Facilities (The University of Sheffield, Bio-molecular Sciences/Molecular Biology Biotechnology departments). Grid carbon coating and uranyl formate solution for staining were prepared by Svetomir Tzokov and Chris Hill. Gold sputtered surfaces were prepared by Dr Rosie Jarrald.

Chapter 5

ESI-TOF analysis was performed by Sharon Spey and Simon Thorpe (Faculty of Science Mass Spectrometry Centre). Transmission electron microscopy was performed at the Electron Microscopy Facilities (The University of Sheffield, Bio-molecular Sciences/Molecular Biology Biotechnology departments). Grid carbon coating and uranyl formate solution for staining were prepared by Svetomir Tzokov and Chris Hill. The cloning of SgsE-A3 was performed with an MChem student Simran Basi who also carried out the SgsE-EGFP urea experiment and the study on SgsE-STOP and SgsE-Mms6 as magnetite synthesis additives. Simran also performed the particle sizing for the experiment. Some of the binding experiments used 10 nm sized magnetite synthesised by Chris Legge (The University of Sheffield, Chemistry Department). Some of the 30-50 nm magnetite samples were synthesised by Dr Andrea Rawlings.

Chapter 6

The monobody project was conceptualised by Dr Andrea Rawlings (The University of Sheffield, Department of Chemistry) who also designed and ordered the MB-E8 constructs. MB-E8 and MB-CTRL protein purification and CD was performed together with a summer project student Molly C John (The University of Sheffield, Molecular Biology Biotechnology) who also carried out the ELISA optimisation experiments. Some of the E8cc proteins used were purified by Dr Lori Somner. CoPt $L1_0$ particles used as the target materials was obtained from Dr Rosie Jarrald who also provided the XRD spectra. The phage display protocol was written by Dr Andrea Rawlings. The phage display process was carried out with great help from an MChem student Tinglu Xua. Transmission electron microscopy was performed at the Electron Microscopy Facilities (The University of Sheffield, Bio-molecular Sciences/Molecular Biology Biotechnology departments). Grid carbon coating was prepared by Svetomir Tzokov and Chris Hills. The Excel spreadsheet for Scherrer calculation was prepared by Dr Andrea Rawlings. ESI-TOF analysis was performed by Sharon Spey and Simon Thorpe (Faculty of Science Mass Spectrometry Centre). Attempts at nanoparticle annealing were performed with the help of Neil Hind who operated the high temperature furnace (Department of Materials Science, The University of Sheffield). Some of the 30-50 nm magnetite samples were synthesised by

Dr Andrea Rawlings.

Contents

List of Figures

List of Tables

List of Abbreviations

2D	two dimensional
3D	three dimensional
a b c d e f g	amino acid locations on a coiled-coil
A3	magnetite binding peptide
aa	amino acid
AFM	atomic force microscopy
AIM	auto induction medium
Ank4	variant of an ankyrin repeat protein
APTS	3-Aminopropyltriethoxysilane
BPM	bitt-patterned media
BSA	bovine serum albumin
°C	degrees Celsius
CBS	citrate buffer saline
CC	Coiled-coil
CD	circular dichroism
CoPt A1	cobalt platinum with randomly distributed cobalt and platinum atoms
CoPt $L1_0$	cobalt platinum with annealed crystal planes
CPU	central processing unit
Da	Daltons - g/mol
DLS	dynamic light scattering
DMF	Dimethylformamide
DNA	deoxyribonucleic acid
DTT	Dithiothreitol
E. coli	*Escherichia coli*
ECL	Enhanced Chemiluminescence

EDTA	Ethylenediaminetetraacetic acid
EGFP	enhanced green fluorescent protein
ELISA	enzyme linked immunosorbent assay
ESI-TOF	electrospray ionisation time of flight
FFT	Fast Fourier Transform protocol
FN3	fibronectin domain 3
FPLC	Fast protein liquid chromatography
FWHM	full width at half maximum
g	gram
G. stearothermophilus	*Geobacillus stearothermophilus*
GRAVY	grand average of hydropathy
GTP	guanosine triphosphate
GuHCl	guanidine hydrochloride
h	hydrophobic
HDD	hard drive disk
HPLC	high performance liquid chromatography
HR	protein HR00C3-2
HRP	horse radish per-oxidase
HSQC	heteronuclear single quantum coherence
in	inch
IPTG	Isopropyl-b-D-thiogalactoside
k	kilo
KIH	'knobs-in-holes'
l	litre
LB	Luria Bertani
LSPR	Localised Surface Plasmon Resonance
MAI	magnetosome island
mAU	milli Absorbance Units
MB	monobody
min	minute
ml	milli litre

MOPS	(3-(N-morpholino)propanesulfonic acid)
MRE	mean residue ellipticity
MW	molecular weight
ng	nanogram
Ni-NTA	Nitrilotriacetic acid
nm	nanometre
NMR	Nuclear magnetic resonance
OD595	optical density at 595 nanometre wavelength
Oe	Oersted
OTS	octadecylTrichlorosilane
p	polar
PBS	phosphate buffer saline
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
pI	isoelectric point
pIII	protein displayed on M13 phage surface
PLL	poly-l-lysine
PLO	poly-l-ornithine
pN	pico newton
QCM-D	Quartz crystal monitoring with Dissipation
rpm	revolutions per minute
S-layer	Surface layer
S. aureus	*Staphylococcus aureus*
S1	one of the variants of subunit S1
S2	one of the variants of subunit S2
SDS-PAGE	sodium dodecyl polyacrylamide electrophoresis
SSD	solid state drive
Strep	Strep-tag II
SUMO	Small Ubiquitin-like Modifier
Tb	terabyte

TBS	tris buffer saline
TBS-T	tris buffer saline tween
TEM	Transmission Electron Microscopy
Tm	melting temperature
μ**m**	micro metre
μ**M**	micro molar
UV	ultraviolet
XRD	x-ray diffraction

Dedicated to my Grandmothers - Regina and Agnutė…

Chapter 1

Introduction

1.1 Introduction

The modern world currently stores more than 33 zettabytes of data [1]. This is a thousand times more than it was at the turn of the century (5 exabytes in 2003 [2]) and is expected to rise to 175 zettabytes by 2025 [1]. To meet such growing demand, data storage technologists have sought radical new possibilities to increase world's data storage capacity. This work builds on efforts to use biological materials and know-how extracted from prior studies on magnetic bacteria with the aim of constructing bit patterned magnetic data storage media for improved hard drive disks. The work is highly multidisciplinary and touches on fields of magnetic nanoparticle production, bio-mineralisation, protein self-assembly and protein engineering among others. Chapter 1 introduces the reader to the world of hard drive disk data storage, the justification for the desirability of bit-patterned media, nanotechnology using proteins and previous attempts at biologically inspired nano-patterning of nanoparticles. Some of the results chapters also contain separate literature reviews on subjects relevant to the specific chapters as that this will allow the reader to better navigate this multidisciplinary work.

Despite its broad scope, this work does have a single underlying thesis question: can a bit patterned media hard-drive disk of desired dimensions be constructed by utilising protein self-assembly and specific protein-inorganic material interactions?

1.2 Data-Storage: a Historical Overview

It is not controversial to say that the ability to dissect patterns about the natural world and pass them on to other generations has played a key role in humans developing as a species. Very early on, the capacity for the human brain to store all the acquired information was exhausted and methods of imprinting data onto external objects were developed. Arguably, the earliest discovered instance of data storage external to the human brain was discovered on the Ishango bone (found in what is now The Democratic Republic of The Congo), dated 6,500-9,000 B.C. [3] which possessed notches proving the existence of mathematical knowledge as well as a log of lunar phases. Another bone showing clearly defined notches was also found in Border Cave in Lemombo Mountains (South African Republic and Eswatini border) and was dated to 35,000 B.C. [3]. Jumping ahead, written word has long been the medium for propagation of human knowledge [4] until it was replaced by printing technologies [5, 6, 7]. A cornerstone development in data registering was triggered by the 1665 bubonic plague, where officials concerned about public health education started chelating and publishing granular detail on incidents of mortality [8]. It was not until the invention of the magnetic tape recording, however, that the data storage started to involve complicated contraptions used today (the first tapes were invented by Fritz Pfleumer and demonstrated in 1935, where magnetic iron particles were glued to narrow strips of paper [9]). By 1950s, magnetic hard drive disks (HDDs) started to become commercialised [10, 11] which had the advantage of offering random access of data as compared to magnetic tapes [12]. HDD technology has improved steadily since its invention (a 10^8 improvement in terms of data storage density) by optimising and minimising hardware components [12]. Today, solid-state drives (SSD) - transistor based data storage, are steadily replacing HDDs in consumer electronics like desktops and laptops [13]. While SDDs are overtaking the personal local data storage market, the big data storage is moving to 'cloud' storage [14]. The term itself has been criticised for being misleading as it hides the immense amounts physical infrastructure needed to support today's data storage needs [15]. Indeed, the 'cloud' server warehouses use HDDs as their main storage for providing real-time access to customer's data and even magnetic tapes for archival purposes

[16]. Future server-side data storage technologies may rely on innovations in deoxyribonucleic (DNA) data storage, however so far they have higher latency, higher energy expenditure and the costs are currently at around 800 million dollars per terabyte [17]. For these reasons, HDD and magnetic tape technology appear to be of crucial importance for the world's data storage needs today and in the future (expected to increase 2 fold by the year 2025 [1]). The HDD industry, however, has met the limits of data storage densities afforded by component minimisation and looks to technologies like heat assisted magnetic recording and bit-patterned media to further increase data storage densities in HDDs [18].

1.3 Hard-drive Discs and their Future

1.3.1 Magnetism

In order to understand HDD technology, an introduction to magnetism on the atomic scale is needed. Magnetism on an atomic level is most commonly caused by electron spin [12, 19]. While all atoms may be magnetically responsive to an applied external magnetic field, only few maintain this magnetisation once the application of the field is discontinued [20]. Paramagnetic atoms contain at least a single unpaired electron, thus such atoms have a magnetic moment, which can be directed by the applied magnetic field. Due to thermal fluctuations the direction of magnetic moment is lost once the magnetic field is discontinued. Materials built out of atoms with unpaired electrons will have different magnetic behaviours depending on how the atoms are arranged in a crystal lattice of the material. These arrangements are labelled as ferromagnetic, antiferromagnetic and ferrimagnetic (Figure 1.1).

In a ferromagnetic arrangement, the electron spin of each atom is aligned in the same direction. This is due to the electrons being exchange coupled (a quantum mechanical phenomenon) in a parallel direction. Because of the parallel alignment of the individual electron moments, the atoms collectively have a net magnetic moment. Antiferromagnetic arrangement has no net magnetic moments due the electrons being exchange coupled in an anti-parallel direction. Ferrimagnetic materials also have the electrons exchange coupled in an anti-parallel direction, however, here the neighbouring individual atomic magnetic moments have unequal magnitudes,

thus there is a net magnetic moment. Different atoms will be compatible with different arrangements, for example, ferromagnetic materials typically are made up of Fe, Co or Ni or their alloys or oxides.

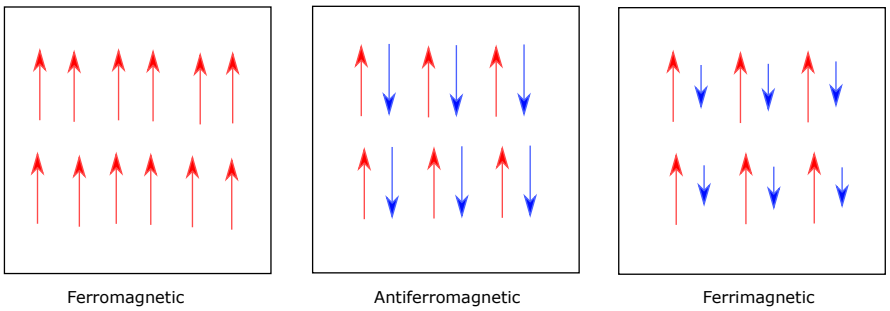


FIGURE 1.1: Ferromagnetic, antiferromagnetic and ferrimagnetic electron spin arrangements. Arrows indicate the direction and magnitude of magnetic moments.

In nature, however, atoms are generally not found in extensive continuous parallel or anti-parallel arrangements. In order to achieve magnetostatic energy minimisations, materials form localised domains of coupled atoms with net magnetic moments (Figure 1.2).

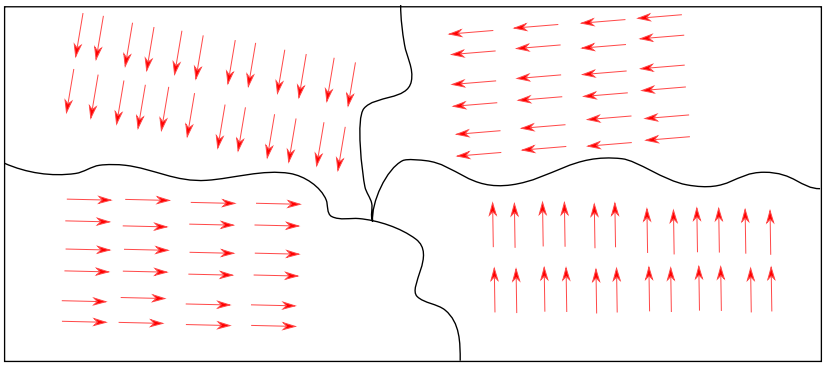


FIGURE 1.2: A diagram of a bulk material arranged in magnetic domains. Red arrows indicate locally coupled atomic magnetic moments.

Each of the domains points the magnetic moment to a compatible end of another domain. In bulk, this results in a material without a net spontaneous magnetic moment. The magnetic domain formation is balanced between the energy reduced from minimising magnetostatic energy and increased energy caused by the existence of

domain-to-domain walls. Atomic composition, the shape of the material and environmental conditions all contribute to defining the maximum size of a single domain before it begins to split into sub-domains.

Magnetic materials possess a property termed magnetic anisotropy [19], meaning that a material will have a preferred axis for magnetisation, where it is easier to magnetise the said material (requiring a lower applied magnetic field). This axis is called the 'easy' axis (as opposed to the 'hard' axis). Magnetic anisotropy can arise from the crystal structure, shape or environmental conditions of the material. The anisotropy that arises from the crystal structure called magnetocrystalline anisotropy is intrinsic to the material itself rather than being induced externally. An example of magnetocrystalline anisotropy can be seen from a body centred cubic structure of iron atoms (Figure 1.3).

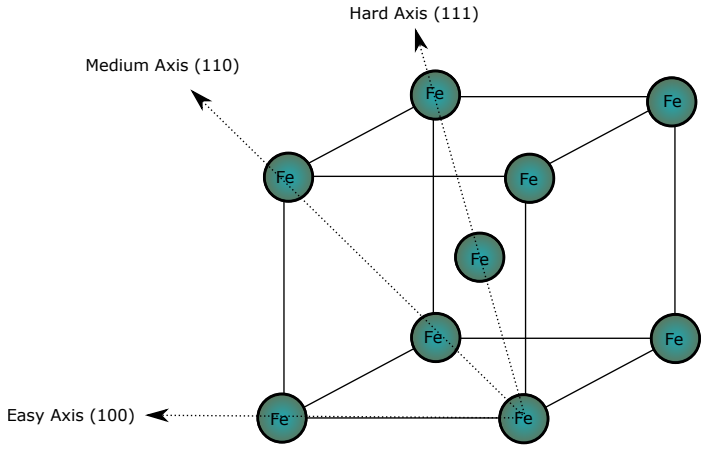


FIGURE 1.3: A diagram of iron body centred cubic (bcc) crystal structure. Axes for magnetisation are listed. Crystal planes are listed as (100), (110), and (111).

Here, iron is shown to have three axes: 'easy', 'medium' and 'hard' corresponding to different crystal planes (100), (110), and (111), respectively. These, again, indicate the axis where the least amount of magnetic field needs to be applied to reach full magnetisation of the material. The higher the magnetocrystalline anisotropy is, the harder it is to change the magnetic orientation of that material. This applies to

both spontaneous demagnetisation due to thermal fluctuations and intended magnetic orientation change through an applied magnetic field. For single domain magnets, magnetocrystalline anisotropy is the major contributor to a magnetic property known as coercivity. The coercivity of magnetic materials can be quantified by obtaining and analysing magnetic hysteresis loops [19]. The property is measured in Oersted units (Oe). Values below 100 Oe mean that the material is magnetically 'soft', whereas materials with values over 1000 Oe are considered to be magnetically 'hard' [20]. Values in-between signify 'semi-hard' magnets. Magnetic data storage applications generally use 'hard' magnets as this property allows for long-term retention of information.

1.3.2 Magnetic Nanoparticles

Nanoparticles also play an important role in magnetic data storage present and future. Magnetic nanoparticles are pure metal, metal oxide or metal alloy particles that are up to 100 nm in size and are magnetic [21]. Understanding magnetic materials at nano-scale (bellow 100 nm) requires specific consideration. As mentioned above, depending on the material, nanoparticles will have differing critical volumes at which they will always have a single magnetic domain. Increasingly small particles are more susceptible to thermal fluctuation effects that can 'flip' their magnetic moment spontaneously. This relationship between particle volume (V), thermal energy (T) and maintenance of magnetic orientation (magnetic anisotropy - k_μ) is defined by the following equation:

$$k_\mu V = k_B T \tag{1.1}$$

Here, k_B is the Boltzmann constant. The right side of this equation defines the temperature required for the magnetic moment orientation of a given particle to change. This means that with decreasing particle volume, the magnetic orientation becomes less stable at lower temperatures. On the other hand, an increase in anisotropy would increase the temperature at which the magnetic moment orientation can be maintained.

Magnetic nanoparticles have been attractive to various fields, including data storage [181], environmental [22] and medical [22, 23] applications. Magnetocrystalline anisotropy defines the magnetic properties of a given nanoparticle [21] and various nanoparticle crystal lattices may come in an array of geometries [19]. This project focuses on two well-known magnetic nanoparticles: magnetite and cobalt platinum alloy.

1.3.3 Magnetite

Magnetite (Fe_3O_4) is one the three most common iron oxides that exist in nature, the other two being maghemite (γ-Fe_2O_3) and hematite (α-Fe_2O_3) [24]. Magnetite is ferrimagnetic at room temperature, but its magnetic properties may vary depending on the synthesis method (which also defines individual particle shape and size) [24] (Figure 1.4).

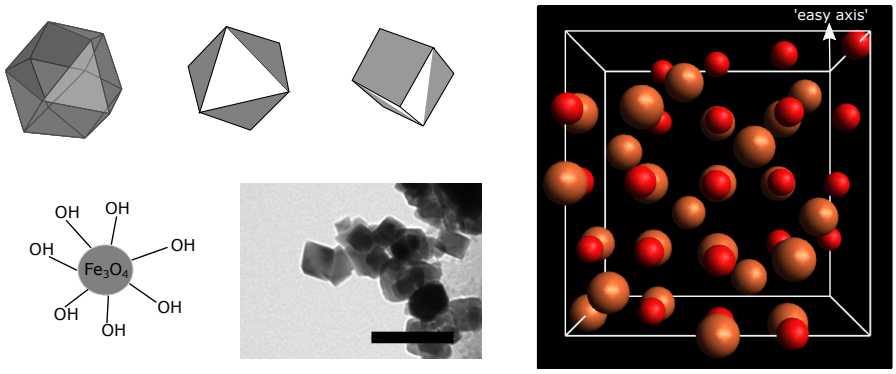


FIGURE 1.4: A summary of the magnetite nanoparticle. Top left: some of the geometries that magnetite can assume. Bottom left: a diagrammatic representation of magnetite surface. Bottom middle: a transmission electron microscopy image of magnetite nanoparticles (scale bar - 50 nm). Right: magnetite crystal structure.

Magnetite can assume the shapes of tetrahedrons, cubes, dodecahedrons and others. Its surface in solution is covered with hydroxyl groups. Magnetite can be produced through chemical synthesis methods like room temperature co-precipitation from aqueous solutions and through top-down physical methods of grinding down naturally occurring bulk materials [25]. A typical magnetite co-precipitation reaction can be described with the following equation.

$$Fe^{2+} + 2Fe^{3+} + 8OH^{-} \longrightarrow Fe_3O_4 + 4H_2O \tag{1.2}$$

The reaction follows a pathway of initial nucleation succeeded by crystal growth. Magnetite synthesis chemistry is well understood and simple changes to the reaction conditions can yield particles of desired properties (size and shape) for example, changing the sequence of added components can produce particles that are 10 nm in diameter instead of 50 nm [26]. Reactions can be performed at room temperature without usage of harsh solvents [27]. All together, magnetite is a well-studied magnetic nanoparticle with attractive features for nanomedicine [28]. It can also act as a model nanoparticle for studying inorganic rigid material interactions with biomolecules. However, as magnetic materials, iron oxides are too soft for use in hard drive disc data storage [12] as they have reported values of 23 Oe around room temperature [29] (although Oe of 190 has been reported for magnetite particles at critical size of 76 nm at room temperature [30]).

1.3.4 CoPt

Various cobalt and platinum alloys have been used as films for continuous media data storage [31]. When cobalt platinum (CoPt) alloy nanoparticles are formed through synthesis methods they usually exhibit a crystal arrangement known as A1 [32] (Figure 1.5**A**).

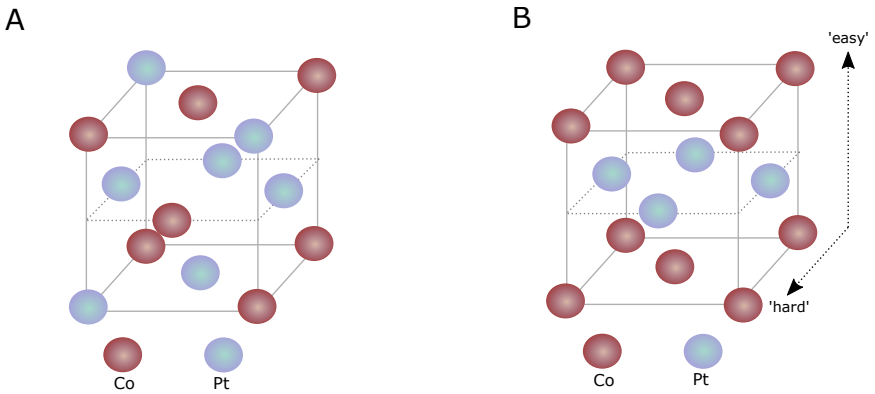


FIGURE 1.5: **A** The A1 arrangement of a CoPt alloy crystal. **B** CoPt $L1_0$ crystal arrangement. 'easy' and 'hard' axes indicate the directionality of magnetic anisotropy in CoPt $L1_0$ crystals.

Here, cobalt and platinum atoms are arranged randomly in the crystal lattice. These materials are magnetic, but have low coercivity values (0-50 Oe). On the other hand, the arrangement seen in Figure 1.5**B** is called $L1_0$ [33, 34]. $L1_0$ crystals have each atom type lined up on separate planes, this allows for more cobalt atoms to be placed side by side and thus have stronger exchange coupling between them, ultimately enhancing the magnetic properties of the material [35]. Values on the scale of 1-10 kOe have been reported [31]. These magnetic properties make $L1_0$ CoPt a suitable candidate for data storage; however processing after synthesis has been identified as a major bottleneck [181]. Various production methods have been shown to arrive at CoPt $L1_0$ crystals with varying degrees of success [36, 37, 38, 39, 40], but have all relied on energy intensive annealing protocols (heating to up to 800 °C) making these routes environmentally unfriendly and potentially commercially inviable. Not only that, but high temperature annealing may cause unwanted particle size growth through agglomeration and crystal lattice deformation [32].

Sections below illustrate the importance of retaining magnetisation at small particle sizes for future data storage applications. CoPt $L1_0$ nanoparticles have the advantage of retaining magnetisation at single domain sizes of 6 nm, whereas magnetite becomes super-paramagnetic below 20 nm [41].

1.3.5 How do Hard-drive Disks Work?

The purpose of a given HDD device is to store data reliably and to maintain it in absence of power for 10 years (span of time until an HDD reaches obsolescence) [12]. The component parts of a typical modern HDD can be seen in Figure 1.6.

It typically consists of a stack of discs each covered with several layers of magnetic and non-magnetic material, a read-write head consisting of a sensor and an inductive writer, a spindle motor that rotates the disc, an actuator that moves the head along the radial direction and a lithographically printed circuit board that controls the HDD and interfaces with the central processing unit (CPU) [12]. Until 1980s, the magnetic films were manufactured by spin coating, which was subsequently replaced by sputtering. The manufacturing process produces a continuous film of granulated magnetic metallic particles (in the 1990s it was CoCrPtTa, but has been

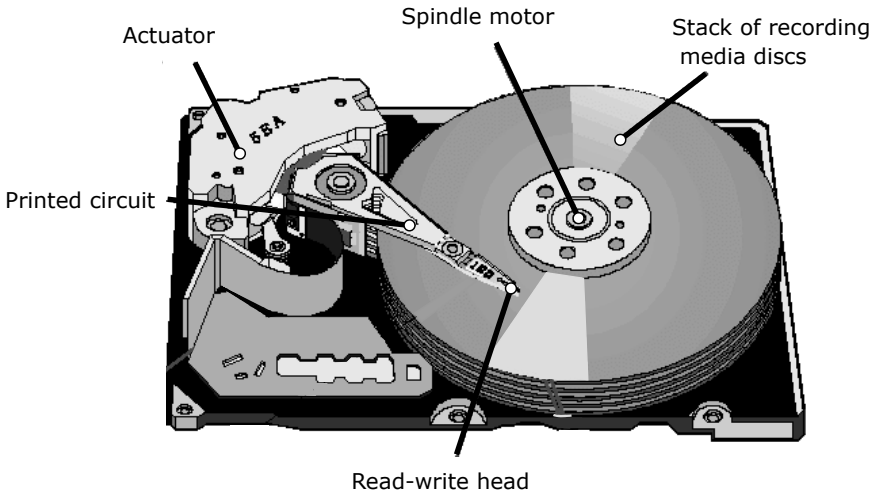


FIGURE 1.6: A diagram of component parts found in a HDD drive. Adapted from [182].

CoCrPtB since the 2000s) [12]. The read-write head defines sections of these particle films as bits (diagrammed in Figure 1.7).

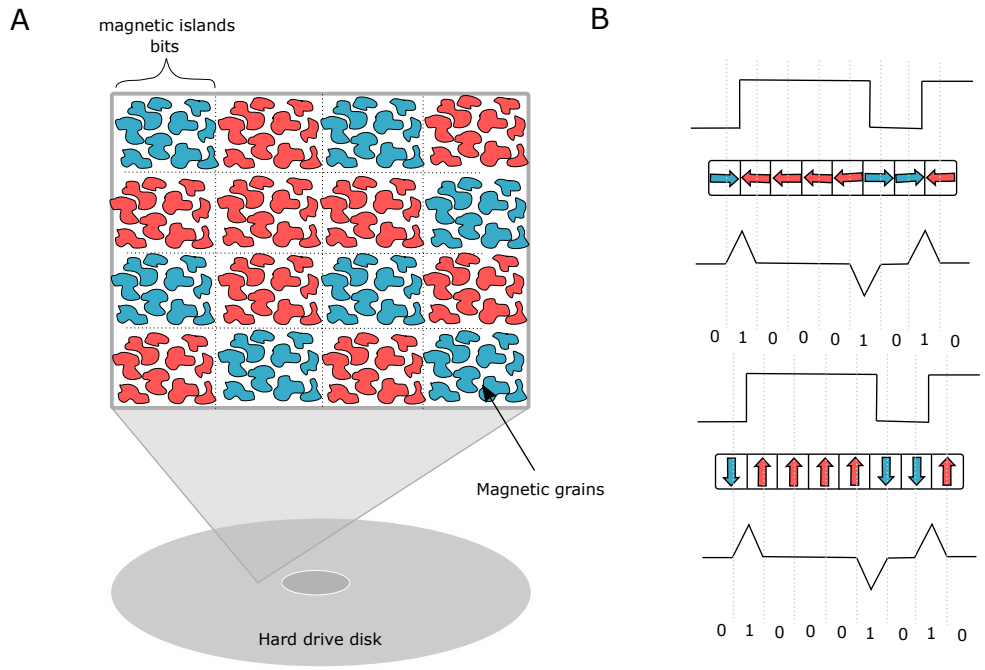


FIGURE 1.7: Data recording on granular magnetic disk media. **A** Diagram showing fields of magnetic grains defined as bits. **B** Top: longitudinal magnetic recording. Bottom: perpendicular magnetic recording. Red and blue indicate opposing magnetic moments.

As shown in Figure 1.7**A** arbitrary regions of magnetic grains are magnetically polarised in either direction. These regions are termed magnetic islands. Transitions between the defined magnetic islands are registered either as bits of '1' or '0' for binary data storage. For example a '1' bit can be defined when two adjacent magnetic islands have opposing magnetic moments (Figure 1.7**B**), where '0' would then be defined by two magnetic islands with the same magnetic moment orientation. Historically, magnetic island magnetic polarity was defined along the plane of the magnetic disc - longitudinally, however, in recent years, perpendicularly defined magnetic island magnetic moments have taken over due to increased areal data storage densities afforded by the method [13] (Figure 1.7**B**).

Since magnetic anisotropy is material dependent, reduction of the grain and thus the island size can be continued if a higher anisotropy material is selected. Values of k_μV that are 60 higher than k_BT are required for reliable data storage for up to 10 years [12]. While selecting higher anisotropy materials would solve the issue of

thermal fluctuations, these materials would not be compatible with magnetic continuous film storage as extensive strong exchange coupling between grains would ensue, making bit to bit boundaries wide, resulting in bits that are too large [12]. This issue would be overcome by having discrete physically separated islands of high coercivity magnetically hard materials acting as single bits of information. This goal was termed bit patterned media (BPM) [12].

1.3.6 Bit-Patterned Media

Conventional, continuous film magnetic recording media are reaching their limits at around 1 Tb/in^2[13], bit pattern media promises storage densities of 10 Tb/in^2[18] and beyond. A comparison between the two types of media can be seen in Figure 1.8.

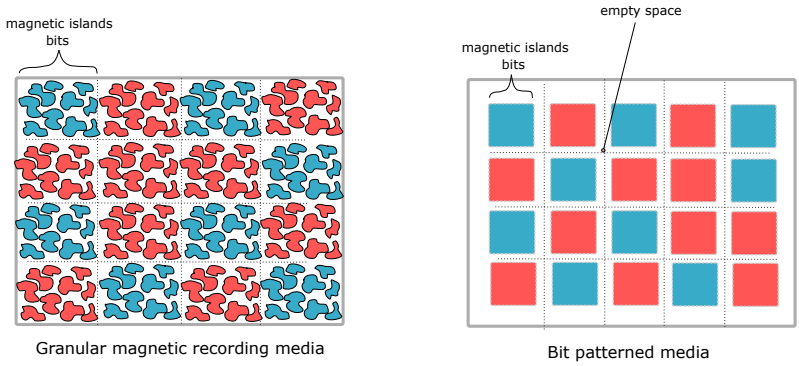


FIGURE 1.8: A comparison between granular recording media and bit-pattern media.

Here, bits are no longer defined as an arbitrary collection of magnetic grains on a continuous film, but rather bits are individual physically separated magnetic islands. The physical separation allows for magnetic material used to be strongly exchange coupled and thus resist thermal fluctuations at smaller grain sizes. Such strong exchange coupling would not be feasible in a continuous film. Because of this, bit patterned media can reduce the overall size of magnetic islands while at the same time making single grains slightly larger thus increasing the signal to noise ratio. One key difference between continuous film and bit pattern media is that the location of bits is no longer defined by the read-write head, but is predetermined in advance.

The biggest obstacle to adopting bit patterned media drives is the fabrication of the media itself [42]. Processes conventionally used for nano-scale patterning (such as electron beam lithography, nano-imprinting and pattern etching) in semiconductor industry have been projected to be too slow and expensive for viable product deployment [18]. Not only that, but the cost targets are much lower and the desired area to be patterned is much larger than for semiconductor needs [42]. For example, to construct a bit pattern medium of an areal density of 1 Tb/in^2 one would need feature sizes of 12.5 nm [12]. Other sources indicate the need for feature sizes of less than 10 nm (width of the magnetic island) with a periodicity below 20 nm [43]. To date, bit pattern media drives have not yet been commercialised [44], though various top-down patterning approaches have been tried to produce 30 nm $L1_0$ FePtCu dots with 60 nm density [45] or sub 10 nm features [43], but none of the methods were viable at scale for commercial bit pattern media production.

1.3.7 Bit-Patterned Media Criteria

Considering the potential reward of achieving BPM, this PhD investigates radical methods that may lead to construction of bit-pattern media without the energy intensive procedures listed above. An overview of the ultimate goal for the project can be seen in Figure 1.9.

Requirements for bit pattern media can be listed as follows [12]:

- method for creating an extensive nano-pattern with feature sizes below 10 nm for high density storage.
- ability to create nano-islands of high hard magnets for long term storage resistant to thermal fluctuations.
- there must be uniformity of shape and magnetic properties between the islands for compatibility with other standardised parts within the storage device.

Here, bit-pattern media is suggested to be achievable by, first, constructing a pattern on the surface and then attaching single domain magnetic nanoparticles to that surface at defined locations. Each nanoparticle would then act as a single bit of information. The method also needs to be environmentally friendly thus the pattern must

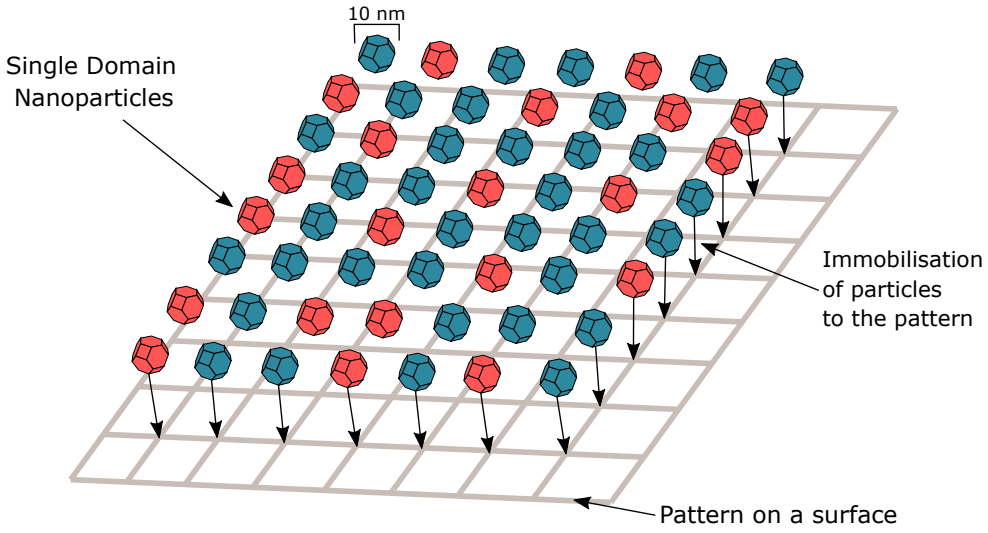


FIGURE 1.9: A diagram of the ultimate-goal for the project.

not be constructed using conventional 'top-down' lithography methods, but rather be spontaneously assembled from biologically derived molecules in a - 'bottom-up' fashion.

1.4 Nature and Nanotechnology

The proposition that biological molecules may help deliver innovations in hard drive disc technology will seem outlandish to most. However, it is often missed how deeply biologically based technologies already penetrate all levels of human society. The following sections aim to introduce the reader to the rationale behind conceptualising biologically inspired production of bit-patterned media.

A groundbreaking innovation took place in 1970s that allowed for the first instance of human-controlled precise splicing of strings of DNA [46], thus ushering the recombinant DNA technology age. This technology combined with innovations in DNA synthesis technologies opens up a plethora of technological possibilities mediated by biological molecules. For example, enzymes are commonly used to improve industrial scale food ingredient production (amylase for sugar production [47]). Personal care products such as face creams or hair protective agents including engineered polymerising proteins have also recently started being produced through

recombinant methods [48]. Various biological solutions have been applied to tissue engineering applications [49]. Biological engineering is also known to produce biodegradable, but robust animal-free leather, screws or packaging materials [50]. Materials like spider silk have long been shown to be useful as elastic, but strong materials [51] and have been recently shown to be producible at scale [52]. Biological components have also been shown to be usable as coatings for gold in electronics applications [53]. The above applications still use biology somewhat conventionally in the sense that, DNA is used to either change the behaviour of the cell or to create a useful protein to perform a function, however, DNA molecules themselves have attracted the attention of technologists. The four DNA bases 'A', 'C', 'G' and 'T' can be see in Figure 1.10**A**.

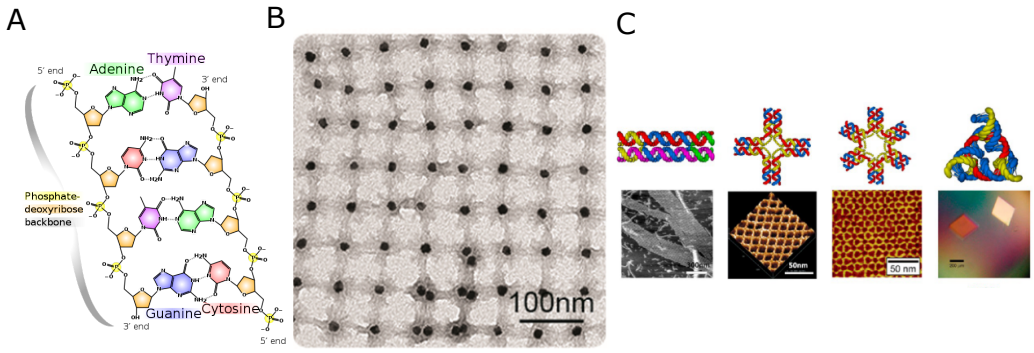


FIGURE 1.10: **A** The chemistry of DNA component parts. **B** Patterns created using DNA origami connecting nanoparticles with a coating - adapted from [183]. **C** 2D patterns constructed using DNA origami - adapted from [184].

DNA as a molecule is well understood and the predictability of the DNA base interactions 'A' : 'T' and 'C' : 'G' allow researchers to encode folding behaviours to create nano-scale shapes, patterns and molecular motors [54, 55]. From this understanding, the field of DNA origami has emerged, where instead of paper, the DNA backbone is folded [56]. The construction of dynamic and static DNA based nanostructures can be potentially used in diagnostics, light harvesting, single molecule reactions, drug delivery, molecular computing and lithography (patterning) [57]. The knowledge of predictable DNA base-pairing has also enabled demonstrations of DNA based-computing [58, 59]. However, the most relevant for this PhD project is the application of DNA origami in nano-scale patterning. Examples of patterning can be seen on Figure 1.10**B**. Here, gold nanoparticles were functionalised with

a flower shaped DNA coating, which then connected them in a pattern [183]. Figure 1.10**C** shows particle independent two-dimensional patterns created by DNA origami [184]. Patterning has also been achieved at sub 10 nm scale, however, extended long-ranging arrays have not been produced [60]. This illustrates the drawbacks of DNA based nano-patterning, inability to create extended long-ranging arrays, which is a requirement for BPM [61].

1.5 Proteins as Nano-manufacturing Building Blocks

Thus, attention is now directed to the other large molecule class for biologically inspired nanotechnology - proteins. From the perspective of nano-scale pattern creation, proteins have the advantage over DNA that instead of four building blocks, ('A', 'C', 'G' and 'T') proteins are composed of 20 different amino acids, each with diverse chemical features (Figure 1.11).

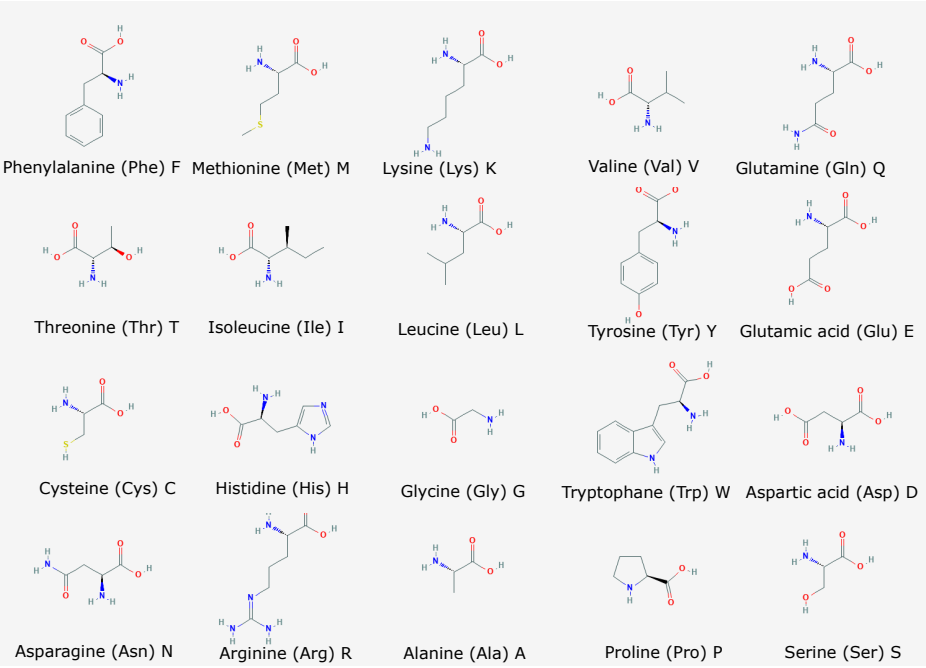


FIGURE 1.11: The set of biologically occurring amino acids with three letter and one letter nomenclature.

This compositional diversity grants proteins the ability to assemble into more shapes and to have more functions. For example, it has been shown that materials made of proteins collagen or tropoelastin have more than 50 different molecular interactions ready for technological exploitation [62].

1.5.1 Diversity of Protein Structures

Before exploring in more detail the specific protein system chosen for this project, a brief introduction into protein structure is needed.

In proteins, sequences of amino acids are connected in chains through peptide bonds [63] (Figure 1.12**A**). Such sequences are then called peptides or polypeptides.

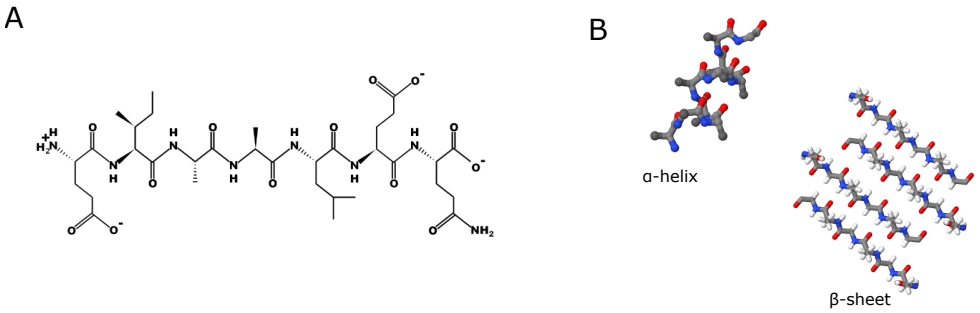


FIGURE 1.12: **A** A short sequence of peptide bond connected amino acids. **B** Short polypeptide chain arrangements in an α-helical or β-sheet conformations.

A peptide is a continuous chain starting from a nitrogen connected to a hydrogen (NH) followed by a carbon connected to an amino acid side chain (α-carbon), followed by a carbon forming a double bond with an oxygen. That carbon is then connected to a nitrogen (the peptide bond) and the order repeats. The chain can be described as (-NH-CαHR-CO-). Polypeptides are free to rotate at certain angles between NH-CαHR and CαHR-CO, but not between CO-NH, as the peptide bond has partial double bond character. As polypeptide chains get longer they will start to assume different conformations (within allowed angles of the peptide chain bond rotations) to form secondary structures (Figure 1.12**B**). The assumed secondary structure is defined by the amino acid sequence. Same amino acids are hydrophobic, thus need to be hidden from the solution, while others are hydrophilic, which tend to be exposed to the surface of a given protein. One such secondary structure is the α-helix, where the peptide chain arranges itself in a spiral (3.6 residues make up a single complete helical turn) [64]. Most α-helices twist in a right-handed fashion (due to inherent chirality of naturally occurring amino acids). Amino acids such as alanine,

leucine, arginine, methionine and lysine have been identified as having the highest propensity for α-helix formation, whereas proline and glycine are the least expected to form one [65]. Another abundant example of secondary structure is the β-sheet (Figure 1.12**B**). Here, stretches of peptide chains (in this context named β-strands) are aligned in parallel or anti-parallel directions to form stabilising hydrogen bond interactions between each protein backbone [63]. β-sheets have a tendency to stack side-by-side. The diversity of protein structures for large part originates from different combinations of secondary structure motifs (Figure 1.13).

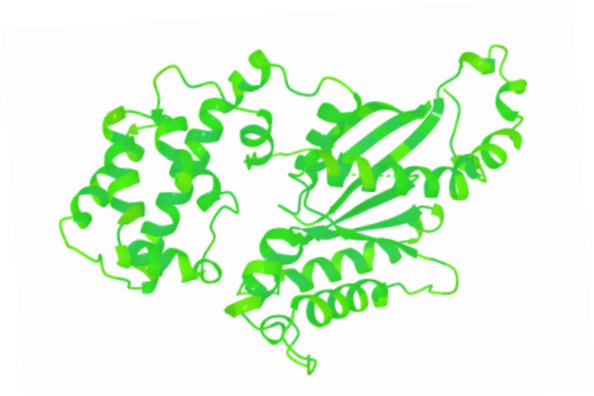


FIGURE 1.13: Cartoon representation of the three-dimensional structure of G-protein alpha subunit 1, showing a complicated mixture of α-helices and β-sheets. Structure obtained using the PDB-viewer:1GDD.

The figure shows a G-protein alpha subunit 1 involved in guanosine triphosphate (GTP) hydrolysis. It consists of multiple α-helices and β-sheets interacting in a complex pattern. This structure was accessed from the publicly available Protein Data Bank [66] that annually releases more than ten thousand protein structures and currently stores 173,000 protein structures. There are also publicly accessible servers with precise standardised classifications of protein structures [67]. It can be calculated that for a protein composed of a 100 amino acid long polypeptide chain there can be 20^{100} possible combinations [68]. Such diversity in primary protein structure could translate to a large diversity of available protein structures. Depending on the classification method there may currently be between 1,000 - 10,000 naturally occurring protein folds discovered [68] which could act as nanotechnology building blocks.

Thus, proteins can offer a diverse library of nano-scale structures, the folding space for which can be incredibly diverse. The selection of 20 amino acids can offer increased chemical diversity for functionalisation of the achieved assemblies when compared to DNA origami assemblies. Any protein structure combination can be relatively easily tested due to cheap gene synthesis technologies. The following sections will cover some of the examples of naturally existing and human engineered proteins that can form patterns on the scale required for bit-pattern media construction.

1.5.2 S-layer Proteins

Surface layer (S-layer) proteins are found on surfaces of most gram positive and gram negative bacterial and on almost all archaea [69]. They have attracted the interest of nanotechnologists because S-layers form precise two-dimensional crystal assemblies with unit cell dimensions between 3 to 30 nm and have a tendency for spontaneous self-assembly *in vitro* [70]. Depending on the microbial species they are derived from, S-layers may form oblique, square or hexagonal porous (2 to 8 nm) lattices [70] shown in Figure 1.14.

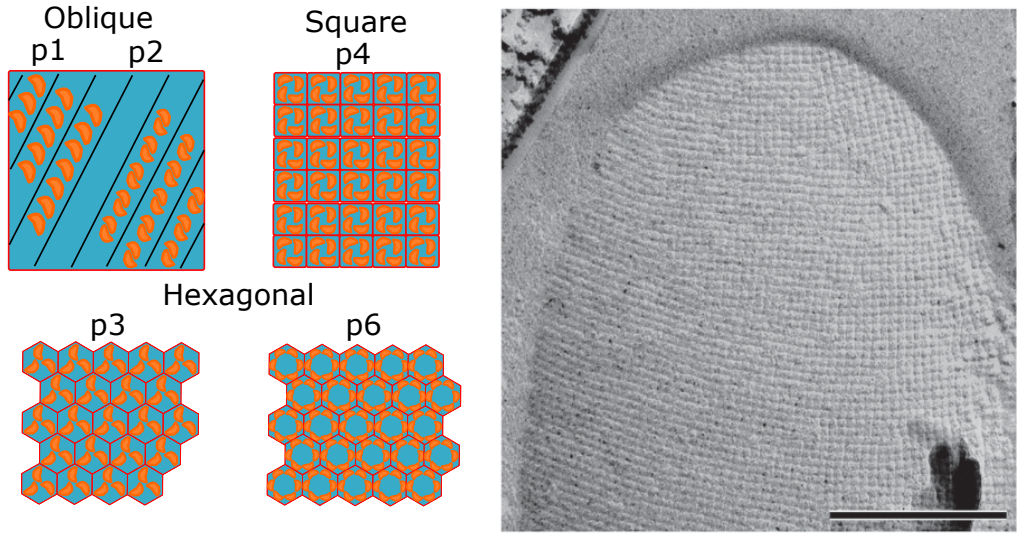


FIGURE 1.14: **A** Possible S-layer lattices - adapted from [185]. **B** A transmission electron microscopy image of a feeeze-etched negatively stained S-layer surface on *Desulfotomaculum nigrificans* strain NCIB 8706 adapted from Pum et. al., 2014 [186]. Scale bar - 200 nm.

Depending on the organism, S-layers can serve a shape determination and maintenance, a protection (against biotic and abiotic stresses), a cell adhesion, a molecular sieving/filtration, an antigenic variation (for host avoidance in case of pathogens), a bioremediation through mineralisation and even an enzymatic function (or a combination of any of the above) [69]. While there is generally very little conservation in terms of primary sequences of S-layer proteins between different species, there is a high degree of structural identity [69]. In general, S-layers have two main domains: N-terminal, which is responsible for cell wall anchoring (which generally exhibits more primary sequence conservation) and a C-terminal, which is responsible for self-assembly [69]. Most S-layer protein sub-units are between 440-1645 amino acids long (40-200 kDa), some strains assemble the S-layer with one sub-unit, other have a multiple component assembly [69]. Biogenesis studies revealed that S-layers are continuously synthesised *in vivo* during cell growth, but the expression is tightly controlled. [186]. It is also important to note that S-layers do not require an underlying pattern to create their patterns and that crystal growth is controlled by sub-unit to sub-unit interactions [71]. S-layer sub-unit tendency to form extended 2D lattices *in vitro* has made it difficult to attain 3D crystals for X-ray crystallographic structure determination, however, a few exceptions do exist [72, 73, 74]. When extracted from the host (either directly or through heterologous recombinant expression), S-layer

sub-units readily reassemble in solution into tubes, ribbons, vesicles, flat-sheets or on surfaces such as round lipid capsules, flat air-water interfaces, solid supports, self-assembled mono-layers and lipid films [70, 186] (Figure 1.15).

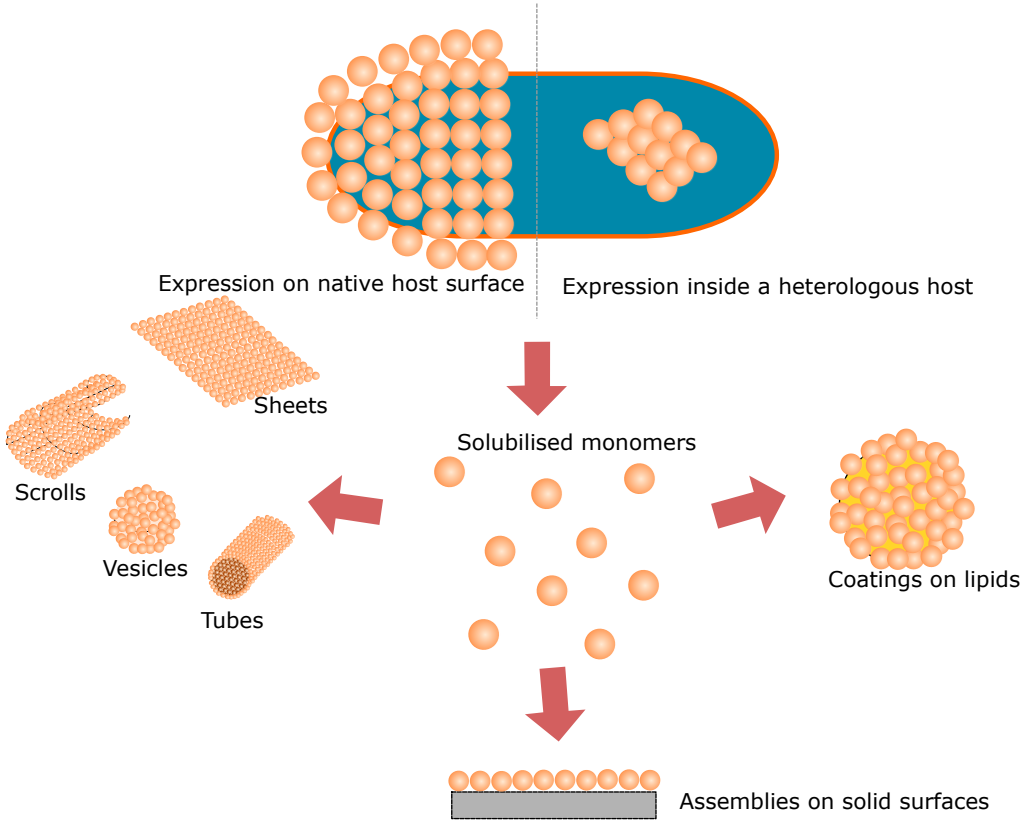


FIGURE 1.15: Different possible S-layer protein assembly fates. Redrawn from Pum et. al., 2014 [186].

The assembly is normally triggered by removal of denaturants like urea or guanidine hydrochloride (present in high molar excess during extraction) in presence of Ca^{2+} (or other divalent ions in some cases) [186]. S-layer assembly is still a subject of intense research with most studies confirming a multi-step process triggered by calcium ions [75, 76, 77, 78]. A two-step pathway can be described as follows: an initial nucleation of monomers of a unit cell takes place (amorphous cluster), followed by slow maturation into fully folded crystalline sub-units. Once a unit cell is mature, the rest of the S-layer grows from the edges of the unit cell, more quickly than the initial cell (for this reason the self-assembly process was described as autocatalytic [79]). It has been suggested that the process is entropically driven, as there is a net loss of hydrophobic surface area upon assembly [80]. Similar general self-assembly

pathway describes the assembly in solution and on surface [186]. Ca^{2+} concentration has been shown to control the number of S-layer rods obtained [81], exemplifying a known feature of S-layer - the fact that their assembly can be tuned by adjusting environmental conditions [186]. When assembling on surfaces, different adsorption kinetics can be observed, with surfaces like 3-Aminopropyltriethoxysilane (APTS) or octadecylTrichlorosilane (OTS) showing faster adsorption rates than bare SiO_2 [82]. This causes SiO_2 surfaces to be covered in larger crystalline domains, because while protein to surface adsorption is affected by the surface, the protein to protein assembly kinetics may not be. S-layer mosaicism when assembling on solid surfaces has long been known [70] and is caused by there being multiple nucleation points on the surface from which the S-layer lattices grow outwards until they meet.

What makes S-layers so interesting from the nanotechnology point of view is their ability to retain the assembly capability when fused to other functional proteins. S-layers have been assembled when fused to enzymes [83, 84, 85], to antigens [86, 87], streptavidin [88, 89] and others.

1.5.3 Artificially Designed 2D Protein Assemblies

Artificial design of ordered two-dimensional protein assemblies is an active and exciting field for construction of functionalised nano-scale arrays. It could be considered as the next step in biological surface nano-patterning, following the successes of DNA origami with the added advantage of potential for more functionality afforded by the 20 amino acids that make up the protein components. The field of artificial protein macro assembly design is, by its nature, multidisciplinary and it involves protein structure and interaction simulations, understanding of biophysics and requires broad knowledge of naturally occurring structural protein modules [90]. The initial conceptualisation of artificial self-assembling protein arrays was elucidated in 2001 [91] (Figure 1.16).

Here, a known protein dimer is genetically fused to another known protein trimer. Depending on the angle at which these two are fused, the protein may self-assemble into three-dimensional ball structures or two-dimensional arrays. The initial study

only demonstrated assembly of tetrahedral three-dimensional assemblies and protein filaments [91]. All of the proteins in the study were derived from known naturally occurring protein components with determined three-dimensional structure (e.g. the tetrahedral assemblies were made up of a bromoperoxidase gene fused to an influenza virus matrix protein M1 through a linker found in ribosomal protein L9 [91]). Another approach, is to take a naturally occurring protein oligomeric assembly and use computational methods to reconfigure the surface exposed amino acid residues to enable extended two-dimensional assembly, like in the example with hexameric protein STM4215 from *S. typhimurium* [187, 92]. Some assemblies designed successfully using this method can be seen in Figure 1.17.

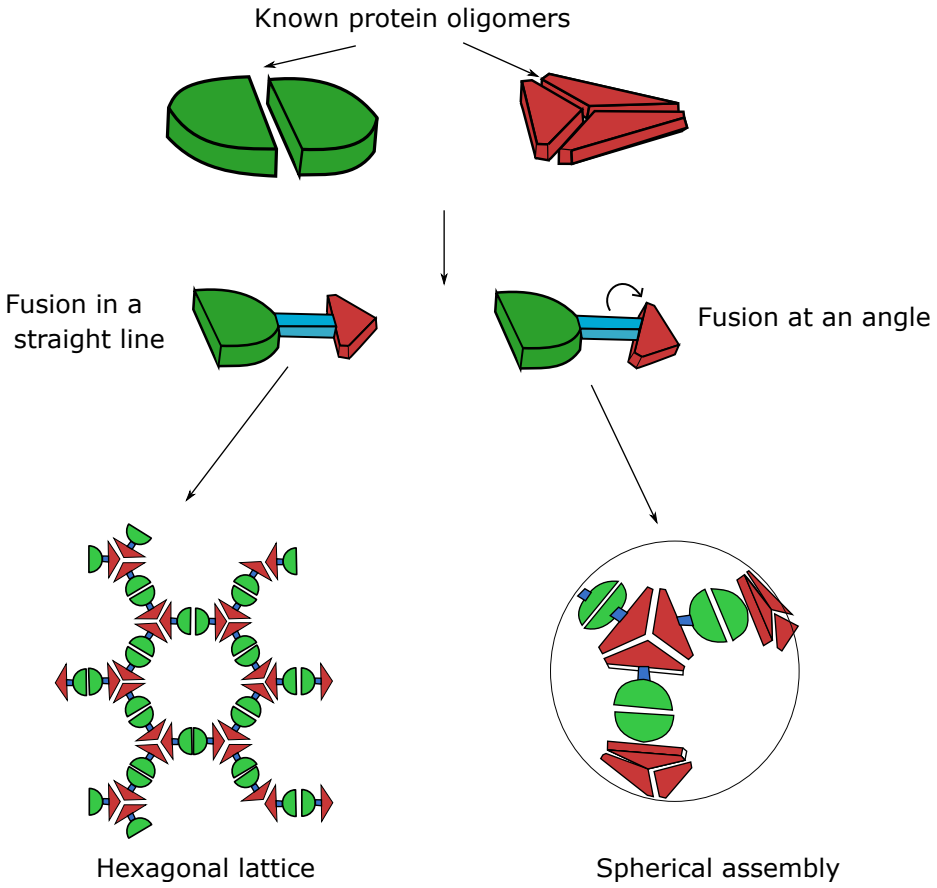


FIGURE 1.16: An early conceptualisation of engineered two- and three-dimensional protein assemblies as described by Padilla et.al.,2001[188].

However, even the most recent successful two-dimensional protein array assembly study used naturally occurring components (wheat germ agglutinin) as the starting building block for the design [93] (the study does, however, include a non-protein ligand in the designed assemblies). Lastly, a recent approach (at the moment

of writing only demonstrated *in silico*) uses a spruce budworm antifreeze protein (strong amyloid-like protein) fused to a multimeric archaeal extremophile derived protein [94]. The two proteins are described as resistant to extreme acidic pH and high denaturant concentrations and, as the authors argue, such component qualities may be retained in the context of extended assemblies, making them more useful for technological applications. These studies demonstrate that the field is still growing, that it relies heavily on discovery of new protein folds with interesting structural properties and that more experimental demonstrations of computationally predicted assemblies are needed.

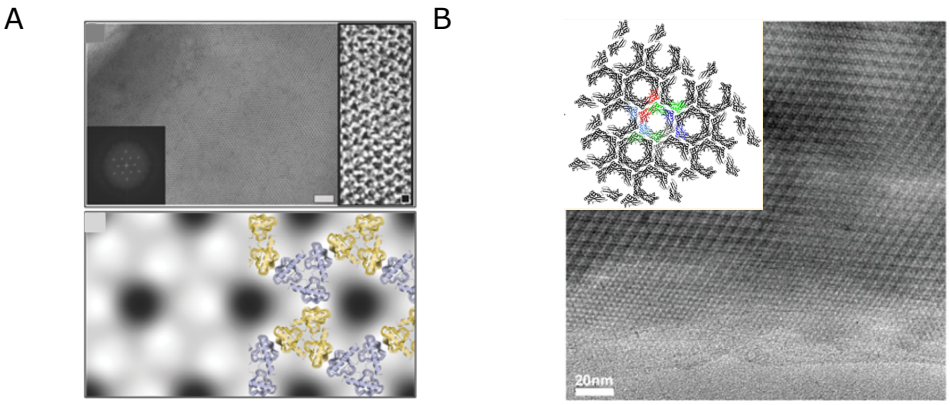


FIGURE 1.17: Transmission electron microscopy images of designed two-dimensional assemblies from Gonen et.al.2015 [189] and Matthaei et.al.2015 [187].

One of the goals of this project is to construct a new protein self-assembly system based on proteins with known desired physical properties. The following section introduces the reader to a potential building block for protein-based nano-patterning.

1.5.4 SasG is a *Staphylococcus aureus* Surface Protein with Incredible Physical Properties

SasG is a surface protein found in the gram positive bacterium *Staphylococcus aureus* (*S. aureus*) and was identified alongside other pathology implicated adhesin proteins [95]. *S. aureus* is a well known pathogen, especially common in hospital acquired infections due to its ability to colonise epithelial and medical equipment surfaces and to form a bio-film [96]. *S. aureus* bio-film formation is a complex multi-step process that involves teichoic acids, proteins, polysaccharide intercellular adhesin - poly-N-acetylglucosamine (PNAG) and other molecules [97]. SasG plays a complex role in

bio-film formation. Strains that do not possess SasG are still able to form bio-films [98], yet SasG containing bio-films tend to be more antibiotic resistant [99] and SasG knockouts display lowered pathogenicity in silkworm larva infection models [100]. Other studies show that SasG plays a more bio-film promoting role [101], the degree of which seems to scale with levels of SasG gene expression [102]. First structural studies on the SasG gene identified a multi domain arrangement that included an A domain, a B repeat domain (each made up of G5 and E domains constituted of 128 residues) that varies in length within any given population and a C-terminal sorting signal peptide made up of a cell wall anchoring motif followed by a stretch of hydrophobic residues [95, 103]. Early electron microscopy evidence shows *S. aureus* cells expressing SasG to have extended (53 nm) thin fibrils protruding from their surface [104]. The same study showed that a minimum of five B repeats is needed for SasG to contribute to bio-film production. This hinted that the length of the extended fibrils (function of the number of E-G5 repeats) determined whether or not bio-film promoting properties were exhibited, thus multiple structural studies followed in order to understand how E-G5 domains can form such long fibres.

G5 and E Domains

First, a G5-E domain repeat unit was studied in isolation and a 39 % β-sheet composition was determined using CD spectroscopy and secondary structure fitting algorithms [105]. Later, X-ray crystal structures were solved for domain combinations: E-G5 and G51-E-G52 [106] (Figure 1.18).

FIGURE 1.18: Cartoon representation G51-E-G52 crystal structure (PDB: 3TIQ).

The data revealed that both E and G5 domains form extended flat β-sheet structures and, while they did not have a compact hydrophobic core, their thermodynamic stability was only slightly lower than those of globular proteins of similar size.

Both constructs were revealed to be extended rod-like proteins (G51-E-G52 measured to be 17 nm long and 2 nm wide). Heteronuclear Single Quantum Coherence Nuclear Magnetic Resonance (HSQC-NMR) evidence revealed that folding of E and G5 domains is highly context dependent. The technique correlates ^{1}H and ^{15}N NMR signals, thus mapping the protein backbone through detection of amide groups. This information allows for determination whether or not a protein is folded. The E domain seemed to only fold when followed by a G5 domain (for example in constructs E-G5 or G51-E-G52), but not when it is the last C-terminal domain (in a G5-E construct, the E domain would not be folded). The study concluded that the two domains fold (and unfold) in a cooperative manner by interlocking via domain interfaces. The apparent lack of a stabilising hydrophobic core is compensated by a 'strategic' distribution of non polar side chains where they are buried within the hydrophobic 'stems' of longer charged side chains such as glutamate or lysine. Further stability is achieved by tyrosine, phenylalanine, isoleucine and leucine at inter domain interfaces forming pseudohydrophobic cores. Later on, the importance of the instability of the E domain for SasG ability to form extended fibres was illustrated [107]. They study that SasG extends because of obligate folding cooperativity of the E domains and stability of the interdomain interfaces. In a G5-E-G5 construct, both G5 domains would start folding first, once folded the C-terminal G5 domain would present an interface to the preceding E domain which would cause the E domain to fold. The folding of the E domain, thus presents a new interface for the N-terminal G5 domain allowing it to achieve its final fold. In such a way then, the E domain effectively couples the folding of the two separated G5 domains thermodynamically. The same study also investigated the mechanical properties of the G5-E-G5 construct using AFM pulling experiments, which revealed unfolding strengths of 250 pN and 420 pN for E and G5 domains respectively (in longer domain repeats such as G51-E-G52-E-G53-E-G54-E-G55-E-G56-E-G57 all E domains unfold first and considering that they can readily reform upon relaxation of mechanical stress, E domains can be seen acting as force buffers). This was later corroborated by a study showing SasG withstanding forces of 500 pN in the context of the cell surface [108]. The mechanical strength exhibited by these domains was close to Ig domains found in the strength paradigm protein titin. Molecular dynamics simulations revealed that this strength

arises from tandemly arrayed clamp motifs within the folded domains constituted of β-sheets and their side-chain packing. Interestingly enough, simulations show that pulling G5-E-G5 apart unfolds the E domain first without disrupting the interfaces, thus further implicating the E domain as a force buffer. Finally, the influence of E domain on a downstream G5 domain was demonstrated [109]. The study showed that E-G5 domain is more stable than G5 in isolation and, furthermore, G5 domains with misfolding causing mutations still fold when preceded by the E domain. They propose two alternative folding pathways. The lowest energy pathway, wherein G5 folds at a C-terminal nucleation point, allowing it to then form the interface with the E domain. Once the E-domain is folded, the interface imparts further stability onto the G5 domain. A less entropically favourable pathway is for the E-G5 interface to fold first and subsequently stabilise both domains, this has been shown to be possible experimentally. The unfavourability arises from the fact that for the interface to fold the intrinsically disordered E domain needs to assume a partial fold first. Taken together, the structural studies provide a robust characterisation of a small, elongated, mechanically and thermally stable rod-like β-sheet protein than can act as a building block for bio-nanotechnology. However, in order to make use of the unique properties of SasG rod-like proteins in constructing nanowire networks, a way to connect them is needed.

1.5.5 Coiled-coil Super-secondary Structural Protein Motifs

One of the most well studied and understood protein folding motifs arguably is the coiled-coil [110, 111] where two α-helices form a twisted self-stabilising structure (Figure 1.19). First proposed by Francis Crick in 1952 [112] (and then later expanded upon in 1953 [113, 114]) as a potential structure for α-keratin to account for twists in its constituent α-helices, coiled-coils have become one of the most prominent motifs for rational protein design [115, 116]. Crick proposed a slight twist to the α-helices so that the side-chains would pack tightly in a 'knobs-in-holes' (KIH) pattern (Figure 1.19**C**). The systematic KIH pattern for side-chain packing outweighs the free energy cost of having two slightly twisted α-helices and grants coiled-coils their mechanical strength.

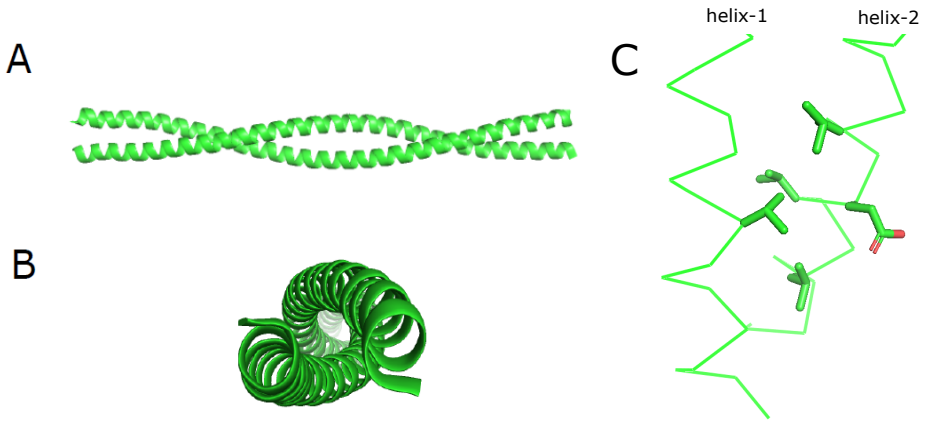


FIGURE 1.19: Representative cartoon coiled-coil structure (Coiled-coil dimerisation domain from coretexillin I, PDB:1D7M) **A** Side-view showing the twist in alpha helical components. **B** Top-down view of the structure. **B** Representation of knobs-in-holes packing in GCN4 Leucine Zipper coiled-coil protein (PDB:2ZTA [190]) where side chain of residue *a* on helix-1 is packed withing side chains of residues *d*, *g*, *a*, and *d*-1 of helix-2. The structures were generated using PyMol.

Over time, interest started building in these protein structures and in 1975 one of the more useful ways to conceptualise them was proposed in a form of a helical-wheel diagram [117] (Figure 1.20) alongside a summary of insights generated on coiled-coil amino acid arrangement patterns discovered until that point with specific references to a known coiled-coil protein tropomyosin.

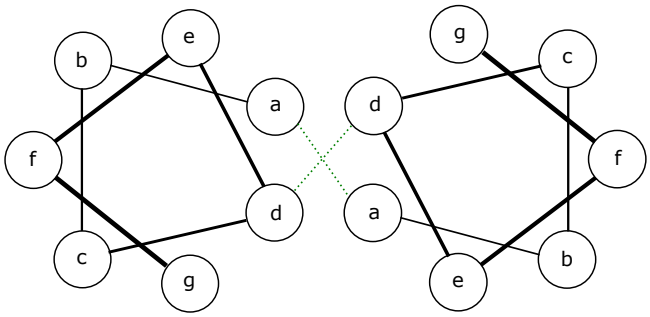


FIGURE 1.20: Coiled-coil helical wheel diagram.

Coiled-coils have been observed to be made up of two or more α-helices wrapped around each other. The α-helices can be arranged either in a parallel or anti-parallel manner (determined by amino acid composition and solution conditions) [110, 118]. Coiled-coil forming alpha helices generally exhibit a hpphppp amino acid pattern

repeat (h-hydrophobic, p-polar). Each of the seven amino acid positions in the repeat are by convention labeled as ***a b c d e f g***. In general, ***a d*** positions are occupied by small hydrophobic residues and comprise the tightly packed hydrophobic core, ***b c f*** positions are solvent exposed and are mostly occupied by uncharged polar amino acids such as alanine or glutamine, lastly ***e g*** positions are occupied by charged side-chain residues most commonly arginine, lysine and glutamate [117]. One of the best studied examples of coiled-coil domains is the leucine zipper GCN4 dimerisation motif [119]. It is an obligate heterodimer, meaning that the final coiled-coil is comprised of two unique chains that would otherwise be unfolded in separation. Coiled-coil design principles learned from the GCN4 motif were the basis for the first study to demonstrate functional ***de novo*** designed coiled-coils [120]. Therein two peptide sequences denoted ACID and BASE were designed. Each of them was comprised of four heptad repeats. The ACID peptide contained glutamate residues in the ***g*** and ***e*** positions, whereas BASE peptide populated ***g*** and ***e*** positions with lysine. This created a preference for the peptides to form heterodimers as homodimers would result in like charges being brought in close proximity. The study put an asparagine at the ***a*** position in the second heptad of each peptide. Asparagine in that position on both peptides would result in an unfavourable juxtaposition of the polar asparagine and a hydrophobic core residue if the peptides were to be arranged in anti-parallel direction, thus asparagine would promote a parallel arrangement of the ACID and BASE peptides where the two asparagines get buried in the hydrophobic core by forming a hydrogen bond (molecular dynamics studies have since complicated this model slightly, suggesting that asparagines at the hydrophobic core are more dynamic than expected [121]). The theoretical predictions in that study were validated experimentally and thus a precedent was set for *de novo* designed coiled-coil technology.

Coiled-coil Formation

While the general amino acid patterns resulting in coiled-coils were known arguably since the 1970s, multiple studies have been conducted since to understand the exact sequence of events that takes place during coiled-coil formation. Specific knowledge on how coiled-coils form would inform the design of self-assembly systems

that use coiled-coils as interfaces. Initial studies showed that coiled-coils form an unfolded collision complex before finally assembling into a coiled-coil and that there is an ionic strength dependence for the assembly rate [122]. Observations have been made that some peptide chains that fit the generic coiled-coil amino acid pattern do not necessarily form a coiled-coil [123]. A homology study compared a terminal sequence in GCN4 to multiple other known coiled-coil sequences and identified what they coined as the trigger sequence [124]. Their identified trigger fit the pattern of Leu-Gln-X-C-h-X-C-X-C-C-X (where X- any amino acid, C- a charged amino acid, h-hydrophobic amino acid) and would form autonomous partial α-helical units prior to coiled-coil dimerisation. They proposed that upon dimerisation of the two helices the rest of the coiled-coil would 'zip-up' along the polypeptide chain. Removing the trigger sequences abolished coiled-coil formation [125]. Trigger sequences have also been shown to be highly determinant of the coiled-coil oligomerisation state [126]. Later studies showed trigger sequences that do not fit a specific pattern, but rely on inter- and intrahelical salt-bridge patterns [123, 127]. A network of intrachain side-chain hydrogen bonding as well as salt-bridge interactions were shown to cause the trigger α-helix to form [127]. The study showed that destabilising the α-helix slows down the coiled-coil formation. They also proposed that trigger sequences can explain how highly repetitive long chain coiled-coils fold 'in-register', in other words, they do not form staggered assemblies. Increasing helical propensity for any given part of the coiled-coil will allow it to act as a trigger sequence, further emphasising that trigger sequences do not need to fit specific amino acid patterns. A later study corroborated this by introducing only two mutations at ***c*** and ***b*** positions that imparted trigger sequence qualities to a heptad and increased the thermal stability of the coiled-coil by 30 °C [128]. While this is by no means an exhaustive explanation of the processes involved in coiled-coil assembly, it is a good primer that is useful when thinking about coiled-coils as protein engineering building blocks to be used as inter protein anchors.

Coiled-coils as Inter-protein Linkages

To assess coiled-coil applicability for linking G5-E-G5 domains one can look at native biological systems where coiled-coil facilitated linkages are ubiquitous (for example, in mediating membrane fusion SNARE proteins [129] and Golgins [130], yeast two-hybrid systems [131] or receptor binding as is the case for hemagglutinin [132]). They have also been successfully employed in synthetic biology studies to link DNA origami structures into long chains [133], create protein-only nanowires [134], join protein domains in cellular environments [135], protein detection [136], cell labelling [137], drug delivery [138] and countless others. The idea of designed coiled-coils as a protein polymerisation toolbox was specifically tested in a study that took a number of previously developed coiled-coils known to form homo-dimers, homo-trimers, homo-tetramers and homo-pentamers [139]. The study showed that different designed coiled-coils have varying degrees of success when it comes to achieving expected polymerisation when fused to the green fluorescent protein (GFP). By default, a 6 glycine linker was chosen to separate the coiled-coil monomers from GFP. The dimer and one of the trimers tested formed expected assemblies as determined by size exclusion chromatography, analytical ultracentrifugation and native mass spectrometry. Other oligomers either failed to form or formed aggregates. These issues were overcome by extending the coiled-coil heptad length (in the case of a trimer) and by extending the linker to 9 glycine residues and introducing asparagine instead of tryptophan in the ***d*** position to increase solubility (in case of the tetramer). These studies show that coiled-coils are viable protein engineering components for creating a diverse variety of protein-protein assemblies. Some care must be taken, however, during the design process as tailored modifications might be needed depending on the coiled-coil (and possibly the fusion partner) chosen.

1.6 Bit-Patterned Media Criteria Revisited

The sections above show that it may be feasible to construct nano-scale patterns using proteins, however, a need to connect the proteins to magnetic components would still exist.

Requirements for bit pattern media revisited:

- method for creating an extensive nano-pattern with feature sizes below 10 nm for high density storage **by using protein self-assembly**.
- ability to create nano-islands of high coercivity and anisotropy magnetic material for long term storage resistant to thermal fluctuations.
- there must be uniformity of shape and magnetic properties between the islands for compatibility with other standardised parts within the storage device.

The following sections show that protein-magnetic nanoparticle interactions are not only possible, but are indeed feasible and that understanding of how this works is growing. One these interactions can be integrated into a pattern producing protein self-assembly system for creation of two-dimensional magnetically functionalised arrays.

1.7 Discovery of Biomineralisation Proteins and Peptides

1.7.1 Nature Interacting with Minerals

While it may seem that the worlds of minerals and biological molecules are separate, they are actually inextricably linked through the process called biomineralisation. Changes in the geosphere during the Ediacaran and Cambrian transition, for example, triggered origination of what we know as animal skeletons - bones [140]. Countless of species produce diverse molluscan shells made of calcium carbonate with incredible precision [141]. More exotic examples of biomineralisation include the ice-binding proteins, that protect various organisms (like fish and algae) from ice-crystals forming within the organism at sub-zero temperatures [142, 143] or chitons - molluscs with sharp biomineralised magnetite teeth [144]. It has also been proposed that many of the found natural metal ore deposits around earth are a product of biological metabolism [145]. Human technologists have sought to adapt these natural biomineralisation processes for creating high precision ultra-hard devices for cutting (in case of the chiton teeth [144]) or to enable low energy manufacturing of future materials in construction (in case of calcium carbonates [146, 147]).

In the examples above, proteins play essential roles in producing the biominerals,

thus this thesis draws attention to the body of work on magnetotactic, where proteins for controlled nanoparticle formation and immobilisation may be found.

1.7.2 Magnetotactic Bacteria

Magnetotactic bacteria have been identified to be able to orient themselves according to the Earth's magnetic poles using chains of intracytoplasmic membrane engulfed nanomagnets termed magnetosomes [148]. Multiple species of magnetotactic bacteria have been since identified, including *Magnetospirillum gryphiswaldense* MSR-1, *Magnetospirillum magneticum* AMB-1 and *Magnetococcus* MC-1 [149]. A conserved gene region between the known model species has been identified and was termed as the magnetosome island (MAI). Understanding magnetosome biosynthesis and the control of engulfed nanoparticle formation is still an area of active research [150], but general representation of the magnetosome can be seen in Figure 1.21.

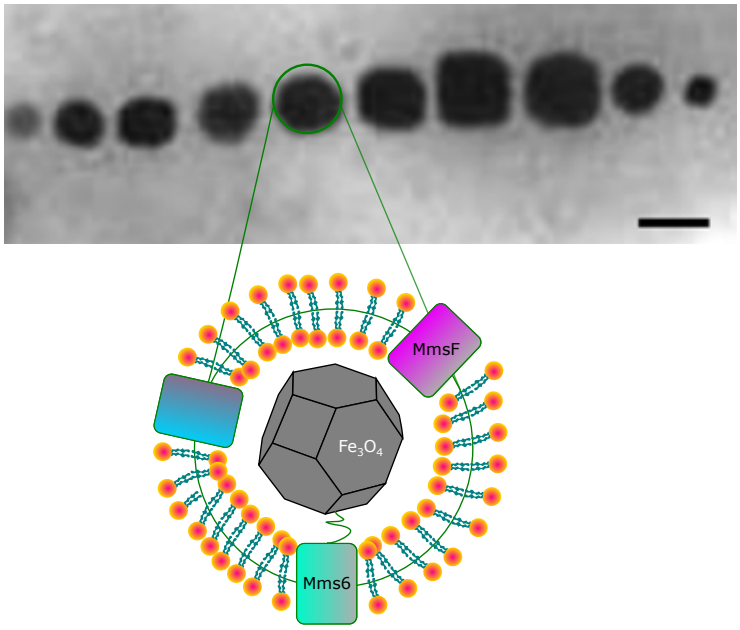


FIGURE 1.21: An adapted image of *Magnetospirillum magneticum* AMB-1 strain [191] with a diagram of a magnetosome. Scale bar - 50 nm.

Magnetosomes are composed of a core inorganic magnetite nanoparticle engulfed in a lipid bilayer membrane. The membrane is populated with transmembrane and membrane associated proteins (Mms6, MamG, MamD, MmsF, MamF, MamC and others, depending on the species [151]). These proteins act as iron ion transporters,

iron ion nucleation sites and shape control facilitators [152]. Crystal size and shape is thought to be controlled through magnetosome proteins binding to specific crystal phases of the growing magnetite particle [150]. Magnetosome properties have been shown to be subject to modification through doping with metals other than iron. *Magnetospirillum sp.* were grown in presence of cobalt in the growth medium, resulting in magnetosomes with 0.2-1.4 % of cobalt content and increased coercivity by 36-45 % [153, 154]. Structured particles containing Mn [155] and Ni [156] have also been achieved through other doping studies. These studies suggest that metals other than iron may be templated by the magnetic bacteria proteins.

The precise control of magnetic nanoparticle synthesis may be ideal for the bit-pattern media requirement for uniformly sized magnetic particles.

1.7.3 Magnetite Interacting Proteins *in vitro*

While there is a complex arrangement of cellular infrastructure within magnetic bacteria to control the shape and size of the internal magnetic nanoparticles, proteins seem to be playing a major role in the process. Thus, understanding their behaviour outside the cell is necessary if one is to replicate the nanoparticle synthesis successes outside of the bacterial host. Two crystal size and shape control proteins, Mms6 and MmsF [157] originating from *Magnetospirillum sp.* have garnered attention for *In vitro* nanoparticle synthesis studies. MmsF plays a major role in crystal size and shape definition and deletion studies showed that complementation with MmsF restores magnetite synthesis phenotype in *Magnetospirillum sp.* [158]. *In vitro* heterologous expression of MmsF produces water stable aggregates that, when used to supplement a chemical magnetite synthesis reaction, produce larger and more defined nanoparticles with enhanced magnetic properties and a narrower size distribution [159]. Mms6 has been identified by investigating proteins bound to *Magnetospirillum sp.* magnetite nanoparticles [160]. It was shown to bind iron ions and was thought to be responsible for inhibiting magnetite crystal growth, thus controlling the morphology. The same initial study also demonstrated particle size and shape control in chemical synthesis reactions outside the cell. A mutational study identified residues

E44, E50 and R55 as responsible for ferric iron binding [192]. Cubo-octahedral morphology particles with a narrow size distribution (instead of octahedral) were produced using Mms6 as a synthesis additive [161]. The proteins were found to remain tightly associated with the particles even after synthesis. What is more, even a truncated version of the Mms6 protein containing only the suspected binding peptide DIESAQSDEEVE was shown to promote growth of more defined shape and size magnetite nanoparticles [162, 163]. Lastly, Mms6 was used as an additive to produce $CoFe_2O_4$ nanoparticles with more control than otherwise would have been possible [164]. Cobalt doped magnetite synthesis enhanced with recombinant Mms6 was also shown to produce high coercivity nanoparticles with narrow size distribution [165].

1.7.4 Phage Display

Phage display is a selection technique that allows for identification of high affinity peptide and whole variable protein binders to biotic or abiotic surfaces [166] (Figure 1.22).

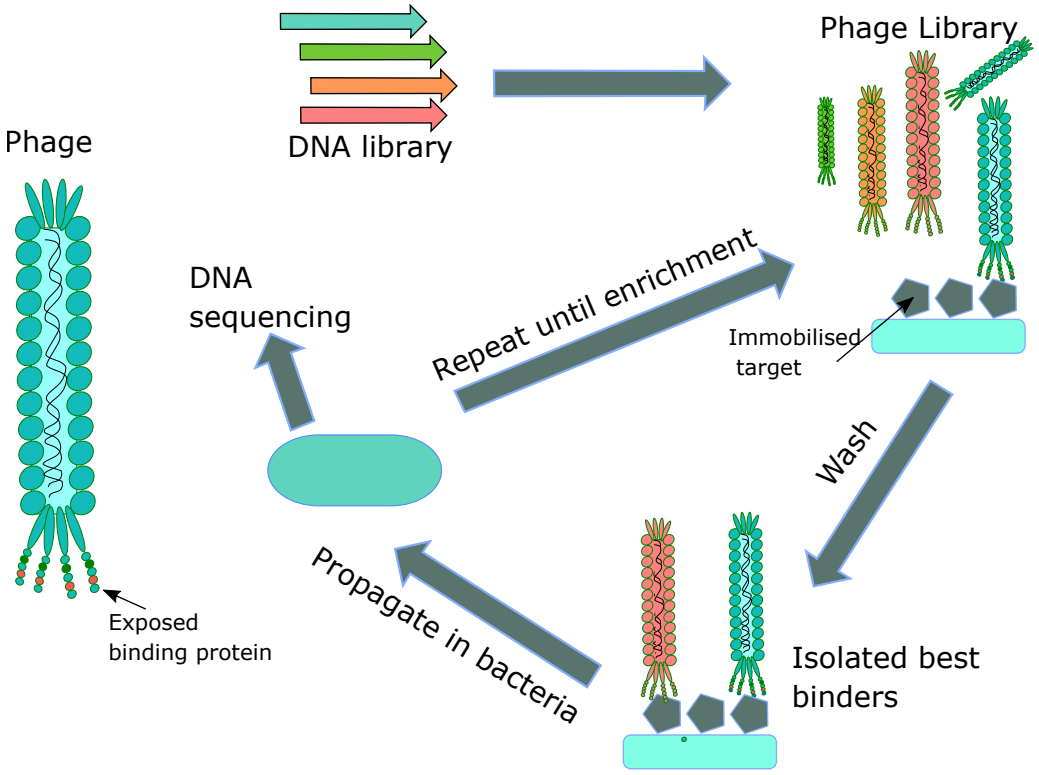


FIGURE 1.22: A demonstration of a phage display cycle.

The technique relies on libraries of modified bacteriophage particles containing genetic material that encodes an exposed binding moiety on the phage particle surface. The libraries can either be exhaustive (contain all possible variations of a given peptide sequence) or can only vary specific sites within a peptide of interest. Libraries incubated with the target material are screened through various washing and elution steps multiple times, enriching the constructs with the highest affinity. These sequences for the constructs are then identified by determining the viral genetic material [167]. This technique can be used to identify peptide sequences of variable lengths that have affinity for inorganic materials with high specificity [168]. Linear peptides have been discovered for both magnetite [169] and CoPt nanoparticles [170, 181]. The study focused on magnetite [169] identified two high affinity sequences HYIDFRW and TVNFKLY that were then shown to effectively immobilise magnetite nanoparticles when fused to other peptides, proving that the binding is independent of the phage display system. Phage display libraries have been crafted to contain protein scaffolds that expose peptide loops (that are then varied through phage display) [171]. A study using such a library was carried out to discover magnetite binding peptides [172]. A noteworthy peptide sequence HNHKSKKHK termed A3 (after the plate location during experimentation, listed in supplementary materials of the discovery paper [172]) was identified to be a good binder both when in the context of a protein scaffold and when tested as a linear peptide (immobilised on a nitrocellulose membrane, unpublished Staniland group results). It was also shown not to interact with DNA, which is a common issue for positively charged peptides (unpublished Staniland group results). These properties would make this A3 peptide a universal immobilisation moiety to attach nanoparticles to other proteins or their assemblies.

1.7.5 Understanding Peptide Interactions with Magnetic Nanominerals

The exciting results obtained from studying magnetite biomineralisation proteins have led to multiple studies trying to understand the basis of metal ion and nanoparticle interactions with proteins and peptides. A study was published wherein nitrocellulose membranes spotted at discreet locations with hexamer homo peptides

were probed for interaction with 13.5 nm magnetite nanoparticles in various conditions [173]. The rationale for the study was to identify which amino acids were most likely to interact with the nanoparticles so that the gained insights could be used for rational design of magnetite binding peptide sequences. They found that charged amino acids like D, E, R, K and H were the best binders at pH 7.4 and pH 8 tris buffered saline (TBS) buffers. They observed that in citrate buffered saline (CBS) pH 6 amino acids D and E did not bind. They have also identified that neither of the uncharged amino acid peptides showed any change in binding in different buffers. This led them to devise a model where buffer molecule ions may be competing with peptides for nanoparticle binding. They then proposed that certain amino acid peptides (D and E) could be used as reversible affinity tags - to bind the particles in TBS and be released in CBS. A follow-up study then showed that the peptide binding function is retained in solution by expressing green fluorescent protein with an E octamer at N-terminus [174]. An analogous study using peptide spotted nitrocellulose membrane arrays was also carried out with CoPt nanoparticles [193]. The design of the peptide arrays was different in that heteropeptides with known affinity have been used as leads for alanine scanning, amino acid substitution and truncation study (the peptides: KTHEIHSPLLHK, HNKHLPSTQPLA, KSLSRHDHIHHH, SVSVGMKPSPRP, VISNHRESSRPL - were discovered in previous studies through phage display). The aim of the study was not only to find which peptide sequences (enabled by key amino acids and motifs) bind CoPt the best, but also which would act as best biotemplates for CoPt $L1_0$ synthesis reactions. For this reason, three assays were carried out in the presence of the membranes, which were: pre-formed particle binding, precursor Co and Pt ion binding and CoPt synthesis reactions. They concluded that presence of lysine residues at terminal sites, presence of methionine, presence of basic residues and absence of acidic residues were important for binding (in all three assays). They also observed that shorter sequences show higher affinity and that a KSLS motif was recurring. A set of peptides was then designed using the insights and they were used as additives for CoPt synthesis additives individually. Reactions supplemented with peptide KSLSRGMK showed improved coercivity and evidence of CoPt $L1_0$ presence.

1.8 Bio-inspired Particle Patterning

A body of research exists in the field of nano-patterning concerned with combining peptide-nanoparticle binding and top down lithography techniques with aims of creating patterns for bit pattern media and other nano-electronics. An early study showed a stamp assisted patterning of silver binding peptides (discovered through phage display) achieve tracks of patterned silver nanoparticles with sub 5 μm periodicity [175]. Mms6 protein was demonstrated to be specifically immobilised (through a specific interaction with a self-assembled mono-layer) on patterned squares obtained via a technique called micro contact printing [176]. The pattern was achieved by stamping a protein resistant self-assembling mono-layer and then filling the gaps in with a self-assembling mono-layer prone to binding proteins. Performing a partial oxidation magnetite synthesis reaction over a few hours at 80 °C produced magnetic patterns of magnetic material only at squares that had Mms6 immobilised (the magnetic islands were 20 μm^2). Micro contact printing was also used to make gold track patterns to which Mms6, mutagenised to contain cysteine, was specifically attached [177]. Following a partial oxidation reaction, the tracks were covered with magnetite at the same locations where Mms6 was immobilised. A similar study showed a successful decrease of track periodicity to 357 nm by employing a patterning technique called interference lithography, where gold tracks were revealed by attacking an X-ray sensitive oligo ethylene glycol (OEG)-thiolate layer at an angle [178]. Magnetite synthesis reactions were then again used to successfully produce magnetic material tracks. CoPt surface immobilisation was also demonstrated on multiple occasions. Initially, a dual affinity peptide HPPMNASHPHMH-GSG-KTHEIHSPLLHK, made up of an N-terminal sequence with affinity for silica connected a C-terminal sequence with affinity for CoPt $L1_0$ (discovered through phage display) via GSG flexible linker was used to successfully mineralise CoPt from metal ion precursors with evidence of $L1_0$ crystal structure being present as determined by X-ray diffraction [179]. The same peptide with affinity for CoPt $L1_0$ was also engineered to contain a C-terminal cysteine (CGSGKTHEIHSPLLHK) and was applied to interference

lithography and micro contact printing patterned gold nano-tracks [180]. Discrete magnetic tracks with sub 200 nm periodicity were obtained through a room temperature synthesis reaction.

1.9 Conclusion and Thesis Layout

This thesis aims to move a step towards biologically derived single particle bit patterned media. The premise is that magnetic nanoparticle interacting proteins and peptides can be patterned by fusing them to periodical self-assembling protein structures, which in turn would act as scaffolds for magnetic nanoparticles. The ultimate aim is to construct nano-scale nanoparticle arrays with-dimensions meeting the requirements for bit pattern media without utilising any 'top-down' lithography methods for patterning or high energy synthesis reactions for the production of the magnetic material.

Chapter 3 describes the work to construct linear, periodical protein structures based on SasG G52-E-G53 domain based fibril proteins connected in chains through a coiled-coil component, whereas Chapter 4 catalogues the efforts of creating two-dimensional assemblies based on SasG G52-E-G53 domains with coiled-coils and a computationally *de novo* designed component. Chapter 5 covers the work performed on a series of modified S-layer SgsE proteins. Chapters 3, 4 and 5 attempt to functionalise the protein assemblies using a magnetite binding peptide A3 discovered through phage display [172]. While the ultimate bit pattern media would have to contain materials with harder magnetic properties, magnetite here acts as a model nanoparticle that is cheap and easy to produce and is well understood. Chapter 6 follows characterisation of a previously discovered magnetite binding peptide sequence in different peptide displaying scaffolds. The chapter also introduces the discovery and characterisation of three new peptide sequences for $L1_0$ CoPt binding and biotemplated synthesis. A summary of the projects juxtaposed to the current bit-pattern media manufacturing state of the art can be seen in Figure 1.23.

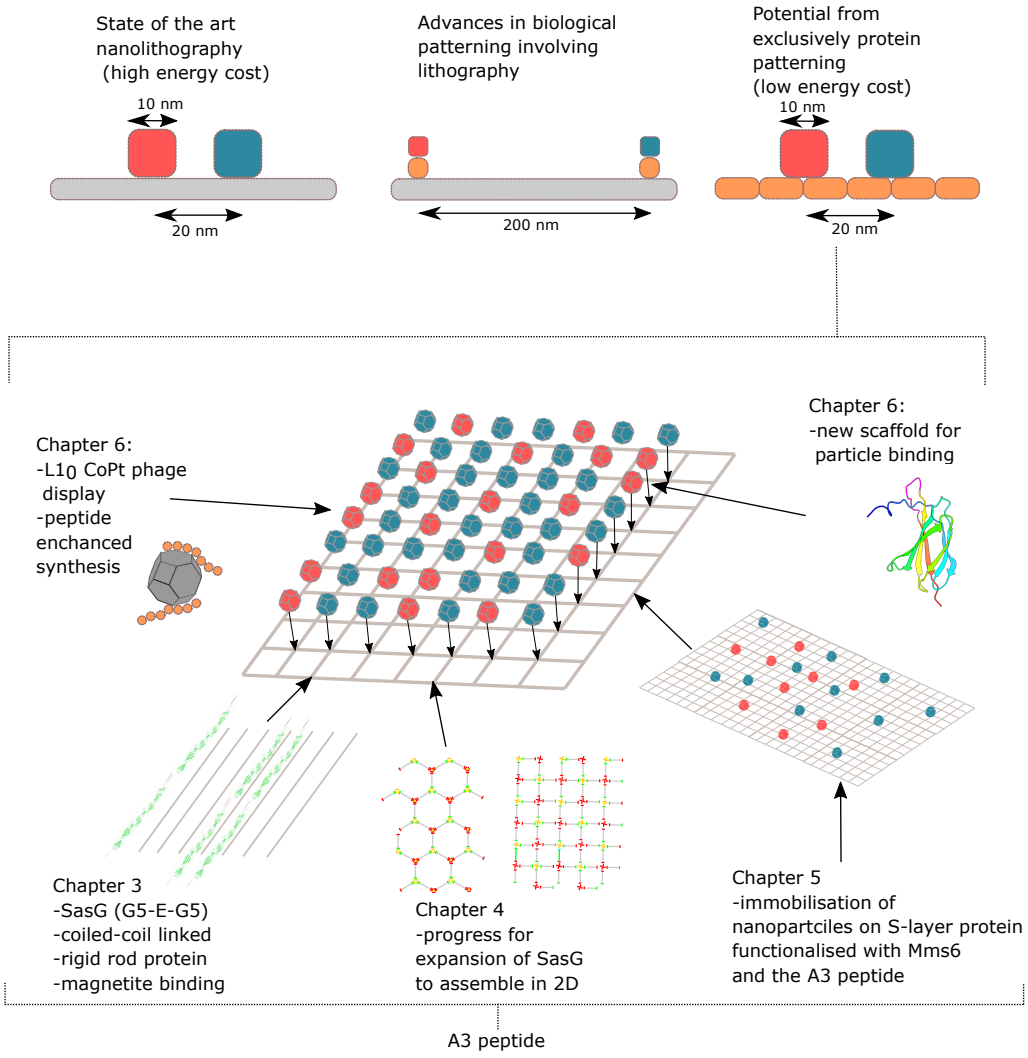


FIGURE 1.23: Top: a comparison of scale and input energy in different approaches to achieving bit patterned media. Bottom: the summary of results chapters in relation to the ultimate goal of constructing bit patterned media.

Chapter 2

Methodology

2.1 Reagent Sterilisation and Sterile Conditions

All molecular biology reagents and equipment requiring sterilisation were treated in an autoclave at 121 °C for 15 minutes. Reagents, incompatible with high temperature treatment, were sterilised using 0.2 μm filters. Sterile conditions were maintained by routinely sterilising the work area with 70 % ethanol and performing all the work near a Bunsen burner. The 70 % ethanol sterilisation was only performed after the Bunsen burner had been turned off. Any biological waste was treated with Rely+On Virkon (Lanxess, Germany) before being disposed of in the main waste streams. Ultra pure water was obtained using Milli-Q EQ 7000 (Merck, US).

2.2 Centrifugation

All small scale centrifugation (1.5 ml) was performed in a GenFuge bench-top micro centrifuge (Progen Scientific, London, UK). Any larger volume centrifugation was performed using Heraeus Megafuge 40R (Thermo Fisher Scientific, US) which has a temperature control module. Centrifugation up to 4700 rpm was performed using TX-750 Swinging Bucket Rotor (Thermo Fisher Scientific, US), whereas speeds up to 14,500 rpm were reached using a F15x-6x100y rotor (Thermo Fisher Scientific, US).

2.3 Ultrasonic Processing

All ultrasonic processing was performed using a Cole-Parmer 130-Watt Ultrasonic Processor (Cole-Parmer, UK). The probe was contained within a soundproof cabinet.

2.4 *E. coli* Strains

Throughout this work three strains of *E. coli* K12 were used: XL10 gold (Stratagene, US) for DNA amplification, BL21 CodonPlus DE3 RIPL (Agilent, US) for protein production and ER2738 (NEB, US) for phage amplification. XL10 gold cells have a modified genome for increased transformation efficiency of large plasmid molecules and are restriction enzyme deficient increasing the success of DNA extraction procedures. BL21 CodonPlus DE3 RIPL cells are engineered to contain more copies of genes for tRNA synthesis to promote a higher abundance of protein expression. The cells contain additional copies of rare codons for certain amino acids. The strain is compatible with T7 RNA polymerase promoter expression constructs and is low in proteases (Lon and OmpT). ER2738 strain is rapidly growing and contains the F-factor plasmid making them prone to phage infection.

2.4.1 Culturing Media

Luria-Bertani (LB) medium was used for most routine *E. coli* culturing tasks. The medium was composed of tryptone (10 g/l), yeast extract (5 g/l) and NaCl (10 g/l) dissolved in deionised water and sterilised by autoclave. Phage display was performed using a lower NaCl concentration (5 g/l). Solid media plates were prepared by making a 1.5 % agar solution in LB, sterilising in autoclave and pouring the liquid once it has reached 56 °C. Protein overexpression was carried out in autoinduction terrific broth medium (from here on referred to as AIM) including trace elements (purchased from Formedium, UK) which contained tryptone (12 g/l), yeast extract (24 g/l), $(NH4)_2SO_4$ (3.3 g/l), KH_2PO_4 (6.8 g/l), Na_2HPO_4 (7.1 g/l), glucose (0.5 g/l), lactose (2.0 g/l), $MgSO_4$ (0.15 g/l) and trace elements (0.03 g/l) suspended in distilled water and sterilised by autoclave. As *E. coli* cultures grow, if present, glucose is consumed first [194]. The glucose levels are optimised in AIM so that it is depleted when cells reach density of OD_{595} 0.5-0.6 - the optimal density to start heterologous protein over-production. After glucose depletion, *E. coli* cells switch their metabolism to start consuming lactose. Metabolic products of lactose (allolactose) activate the expression of T7 polymerase controlled by the *lac* promoter in BL21 strains. Upon activated expression of T7 RNA polymerase, any plasmids containing

protein open reading frames preceded by a T7 promoter will start to be expressed. Thus, the use of AIM provides an automated method of inducing heterologous protein expression in *E. coli* strains containing the T7 RNA polymerase.

2.4.2 Antibiotic Selection

In order to select for correct plasmid *E. coli* transformants and to maintain plasmids in transformants in liquid culture, antibiotic selection is needed. Throughout the project kanamycin, carbenicillin and tetracycline were used at 40 μg/ml , 100 μg/ml and 20 μg/ml final concentrations, respectively. The antibiotics were prepared at a 500 times working concentration in water (and 1:1 mixture of water:ethanol for tetracycline), filter sterilised and stored in aliquots at -20 °C to be thawed immediately before use. When used together with LB agar, molten LB agar was cooled to 56 °C to prevent thermal antibiotic inactivation.

2.4.3 Overnight Cultures

Standard overnight cultures were obtained by mixing 5 ml of LB medium with an appropriate amount of needed antibiotic in 50 ml centrifuge tubes. Either a single colony or a scraping was then taken with a sterile loop and mixed into the medium with antibiotics. The tubes were then placed into a 37 °C incubator shaking at 225 rpm for up to 16 hours.

2.4.4 Transformation of *E. coli* with Plasmid DNA

Chemically competent *E. coli* of various strains were transformed with plasmid DNA using a heat shock protocol. First, competent *E. coli* aliquots were taken out of -80 °C storage and were allowed to thaw for 10-20 minutes under sterile conditions, plasmid DNA was then added and was allowed to adhere to positively charged competent *E. coli* outer membranes for 20 minutes. Plasmid DNA volume varied from 0.5 μl for miniprep purified plasmid DNA to 2.5 μl for plasmid DNA obtained by ligation or polymerase chain reaction (PCR) extension (see the section on site directed mutagenesis). The cell tubes were then quickly moved to a 42 °C heat block for 45 seconds to induce a heat shock causing the DNA to enter the *E. coli* cells.

The cells were then returned to ice for 3 minutes, supplemented with 150 μl of LB and were allowed to re-grow for 60 minutes at 37 °C. A suspension of 100 μl was then applied to moderately dry LB agar plates (containing the appropriate antibiotic) and spread under sterile conditions. The spread cells were allowed to dry for 5-10 minutes and were placed into a 37 °C incubator overnight. The next day, successful transformants produced individual colonies. The plates were stored in 4 °C for up to two weeks.

2.4.5 Long-term Storage of Strains in Glycerol

Placing *E. coli* strains and their transformants in glycerol allows for long-term keeping and easy access when needed. Glycerol stocks were prepared, first, by setting up overnight cultures of the desired strain and mixing the cultures in a 3:1 ratio with 80 % glycerol (autoclave sterilised) in 2 ml cryogenic tubes producing a final glycerol concentration of 20 %. The tubes were then placed into a -80 °C freezer. The strain recovery was performed by transferring the frozen tubes to ice, dipping an inoculation loop into the tube under sterile conditions and using it to spread cells on an LB agar plate with the appropriate antibiotic. The plates were then dried and placed in a 37 °C incubator overnight.

2.4.6 Producing Competent Cells

In order for *E. coli* cells to be receptive to foreign plasmid DNA they need to be made chemically competent. Strains were bought in already competent initially. To produce more stock of the desired strains the cells were subjected to a rubidium chloride treatment. First, 10 ml of overnight cultures were grown in LB which was used to inoculate a large 500 ml flask of LB which was then grown at 37 °C until it reached OD_{595} of 0.5. The culture was then cooled on ice for 15 minutes and transferred to sterile centrifuge bottles which were centrifuged for 10 minutes at 4500 rpm. The pellets were recovered and gently resuspended in 30 ml of a filter sterilised solution of RbCl (100 mM), $MnCl_2$-$4H_2O$ (50 mM), glycerol (15 %). Afterwards, the resuspension was cooled on ice for 15 minutes and spun down for 5 minutes at 4000 rpm. The pellets were then resuspended in 6 ml of filter sterilised MOPS (10 mM), RbCl

(10 mM), $CaCl_2$-$2H_2O$ (75 mM), glycerol (15 %). Small volumes of cells were then distributed into 1.5 ml tubes, flash frozen in liquid nitrogen and placed in a -80 °C freezer.

2.5 DNA Design and Manipulation

Various DNA manipulation and characterisation techniques were employed during the PhD project with the goal of producing various DNA constructs for protein production. Different plasmids were employed depending on application. All protein expression plasmids contained a T7 promoter. Expression plasmids containing the desired insert were constructed by using polymerase chain reaction (PCR) to amplify target inserts with extended adaptor sequences for restriction enzyme treatment and subsequent ligation into like cut plasmid DNA [195]. The template DNA for PCR was either in the form of commercially synthesised linear DNA or as plasmids in Staniland group's storage or shared by collaborators. Small modifications to plasmid sequences were created using site directed mutagenesis. Some constructs used for expression were synthesised and sub-cloned commercially (this is noted where relevant).

2.5.1 Plasmid Component Breakdown

pET28a plasmid was used for protein expression in Chapters 3, 4 and 6. The plasmid system provides a T7 promoter, kanamycin resistance cassette, multiple cloning site containing multiple common restriction enzyme sites and a thrombin cleavage site followed by a poly-histidine (6) tag for protein detection and purification. Chapter 5 concerned with the production of S-layer proteins utilised pPR-IBA1, which is compatible with single enzyme *Bsa*I cloning, contains an ampicillin/carbenicillin resistance cassette and is smaller making it useful for cloning large (>2,000 base pair genes) S-layer genes. pET24a was utilised in expression of small peptide sequences (individual α-helices) as it contains a SUMO-1 tag (and a carbenicillin resistance cassette). SUMO sequences provided a soluble tag that also increased the intermediate (before cleavage) molecular weight/size to make the expressed constructs more compatible with standard purification techniques [196].

2.5.2 Ordering Commercially Synthesised DNA

The project required synthesis of various length *in vitro* synthesised DNA to be ordered from commercial sources. The gene for the S-layer protein SgsE was ordered as GeneArt String (Thermo Fisher Scientific, US). The linear double stranded DNA sequence codon optimised for protein expression in *E. coli* arrived in powder form and was re-suspended in sterile ultra pure water to a concentration of 50 ng/μl and was stored at -20 °C. Short single stranded oligonucleotide sequences (to act as primers) were ordered as desalted, dry powders (Thermo Fisher Scientific, US) that were resuspended in sterile ultra pure water at a concentration of 100 μM and were stored at -20 °C. Sequences used in Chapter 4 were ordered as GeneArt Strings (Thermo Fisher Scientific, US) sub-cloned in pUC57, were resuspended at 50 ng/μl and stored at -20 °C. Monobody - E8 was ordered as a subcloned gene in pET28a from Genscript (US).

2.5.3 Agarose Gel Electrophoresis

Agarose gel electrophoresis was used to separate DNA by size. Typically, 0.8 % agarose gels were prepared using a tris-acetate buffer with ethylenediaminetetraacetic acid (EDTA) at pH 8. The gel mixtures were supplemented with SYBR Safe DNA Gel Stain (Thermo Fisher Scientific, US) for DNA visualisation under UV light. The gels were visualised using the ChemiDocGel Imaging System (Bio-Rad, US).

2.5.4 Polymerase Chain Reaction - Restriction Enzyme Cloning

PCR is an efficient method to produce numerous copies of template DNA [197]. The copies can be modified by using primers that posses slightly different DNA sequences than the template. Restriction enzyme cloning required a high fidelity polymerase that has a low rate of copying error. This project used the KOD Hot Start polymerase (Merck, US). A typical reaction mixture can be seen in Table 2.1.

The primers were designed to be specific for template DNA and to carry adaptor sequences for restriction enzyme cloning. The general design rules included: optimised melting temperature (Tm) to be 55-65 °C and be as close for forward and

Component	Volume	Final Concentration
10X KOD Polymerase Buffer	5 μl	1X
25 mM MgSO4	3 μl	1.5 mM
dNTP mixture (2 mM each)	5 μl	0.2 mM (each)
PCR Grade Water	32.5 μl	
Forward Primer (10 μM)	1.5 μl	0.3 μM
Reverse Primer (10 μM)	1.5 μl	0.3 μM
DNA template	0.5 μl	
KOD polymerase	1 μl	0.02 units/μl
Total	50 μl	

TABLE 2.1: A typical reaction mixture for high fidelity DNA amplification by KOD PCR.

reverse primers as possible, having a G or C nucleotide at 3′ ends and A or T nucleotides at 5′ prime ends, having 17-21 nucleotide (nt) overlap for each primer with the template DNA and not being 40 nt in total length. PCR relies on repeated thermal cycling for DNA amplification which was performed in T100 Thermal Cycler (Bio-Rad, US). A typical PCR reaction is shown in Table 2.2.

Step	Temperature	Time	Repeats
Polymerase activation	95 °C	2 min	1
DNA melting	95 °C	20 s	20-40 cycles
Primer Annealing	Lowest primer Tm °C	10 s	20-40 cycles
DNA extension	70 °C	10 s/kb	20-40 cycles
Final DNA extension	70 °C	2 min	1
Hold	4 °C	infinite; until retrieved	1

TABLE 2.2: A thermocycler protocol for high fidelity DNA amplification by KOD PCR.

The degree of success for a given PCR reaction was assessed via agarose gel electrophoresis and confirmed by presence of strong singe bands at expected molecular weights. If there were no bands or if there were multiple unexpected bands, annealing step temperature was adjusted. Products of successful PCR reactions were purified using spin column based GeneJET PCR Clean-Up kits (Thermo Fisher Scientific, US). The products were then used for restriction enzyme digestion. Throughout the course of this project, two rounds of PCR were used to graft longer (up to 30 base pairs) stretches of DNA onto PCR products. In these cases the PCR products of the first reactions were purified and used as templates for the second.

2.5.5 Restriction Enzyme cloning Into Expression Plasmids

Purified PCR products containing flanking restriction enzyme sites were digested with appropriate combinations of enzymes (obtained from NEB, US) for up to 3 hours at 37 °C. The reaction products were purified the same way as PCR products. Target plasmids vectors were also digested for up to 3 hours at 37 °C. The linearised digested plasmids were then separated from undigested circular plasmids by gel electrophoresis and gel extraction using gel extraction kits (GeneJET Gel Extraction Kit, Thermo Fisher Scientific, US). Both, digested inserts and plasmid vectors were quantified using UV-vis spectroscopy. Various molar ratios of insert and vector were prepared and mixed with Instant Sticky Ligase Master Mix (NEB, US) for 5 minutes on ice before using the mixture to transform *E. coli* XL10 gold cells using the standard transformation method described above.

2.5.6 Colony PCR

Colony PCR was used for preliminary assessment of restriction enzyme cloning procedures. The technique follows the same principle as PCR, but uses whole cells (containing target DNA inside) as template and does not require a high fidelity polymerase. The lack of high fidelity requirement comes from the fact that the aim is to see that the reaction detects the target DNA without concern for accurate copying of the said DNA. DreamTaq Green PCR Master Mix (NEB, US) that utilises the Taq polymerase was used for colony PCR in this project. Colonies obtained from restriction enzyme procedures were used to inoculate a clean tube with water and the master mix supplemented with a forward primer specific for insert DNA and a reverse primer specific for the vector (universal T7 forward or reverse primers were used throughout the project) at 0.3 μM each. Reaction mixtures were the subjected to a standard Taq polymerase PCR cycle and the products were analysed on agarose gel electrophoresis. Tubes with colony inoculated water that corresponded to lanes containing the correct size PCR amplification products were used to prepare overnight cultures for plasmid DNA extraction using GeneJET Plasmid Miniprep Kits (Thermo Fisher Scientific, US).

2.5.7 Site Directed Mutagenesis

Small DNA alterations (up to 30 nucleotide long) were created in existing plasmids using a Site directed mutagenesis protocol. Sets of two overlapping primers were designed to have Tm °C of around 70 °C, G or C at 3' ends, T and A at 5' prime ends and to contain either omitted (for deletions) or additional nucleotides (for insertions). Reaction mixtures were prepared the same way as for PCR with KOD polymerase. The extension times were set to accommodate the length of the whole plasmid being mutated, annealing temperature was set to 70 °C. The reaction products were then digested with *Dpn*I at 37 °C for 3 hours to remove any of the original template plasmid and the resulting mixture was used to transform competent *E. coli* XL10 cells. The resulting colonies were used to prepare overnight cultures for plasmid DNA extraction using GeneJET Plasmid Miniprep Kits (Thermo Fisher Scientific, US).

2.5.8 DNA Sequencing

Plasmid DNA received from collaborators or obtained by commercial cloning, in-house restriction enzyme cloning, site-directed mutagenesis or phage display was validated by Sanger sequencing using services from GeneWiz (UK). The plasmid solutions were diluted in ultra pure water to 50-100 $\mu g/\mu l$ and were mailed in bubble wrap protected envelopes to GeneWiz. Sequencing was performed using universal T7 primers (forward or reverse) for expression plasmids. S-layer fusion genes were sequenced with a total of three sequencing reactions: with T7 forward and reverse primers and a custom primer specific for a stretch of sequence near the middle of the S-layer gene. Successful results were confirmed by translating the obtained DNA sequences into protein sequences and using ClustalW align algorithms to detect mismatches with expected protein sequences [198]. If mismatches were detected, Sanger sequencing chromatograms were inspected to see if the mismatches had been caused by a sequencing error or a true unexpected mutation.

2.5.9 Phage Display

An M13 phage based Phage Display procedure was followed to screen for 7 amino acid peptide sequences interacting with $L1_0$ CoPt nanoparticles. The procedure was adapted from the NEB protocol for Ph. D 7 aa phage display library by Dr Andrea Rawlings. Each screening round was performed three times to enrich the strongest binding 7 amino acid sequences. All of the reagents used were sterilised. Each phage display 'round' was carried out over the span of four days. On day one *E. coli* ER2738 cells were revived from glycerol storage on tetracycline supplemented LB agar plates and placed in a 37 °C incubator overnight. On day two, plate grown *E. coli* ER2738 cells were placed into liquid LB medium supplemented with tetracycline in the morning (picking a single colony) to be used for phage titration later. Panning rounds were carried out while *E. coli* cultures were developing. First, 0.1 mg of annealed $L1_0$ CoPt nanoparticles were weighed using a high sensitivity weight balance. The nanoparticles were resuspended in 500 μl PBS pH 7.4 and ultrasonically treated three times for 1 minute each at 40 % amplitude. The nanoparticles were then collected magnetically (using a magnetic 1.5 ml tube rack - Thermo Fisher Scientific, US), the PBS pH 7.4 was removed and nanoparticles were resuspended in 300 μl of 1x caseine blocking solution (Merck, US) for 1 hour to 'block' the nanoparticles. The particles were then collected and washed three times with 300 μl. The nanoparticles were then mixed with 1 ml of 1x caseine blocking solution mixed with 10 μl of either the naive Ph. D 7 aa phage display library (for screening round 1) or the eluted phage particles from a previous round (for screening rounds 2 and 3). The solution was incubated at room temperature for 1 hour on a roller. The nanoparticles were, again, collected magnetically and the supernatant was discarded into a solution of Virkon, thus, removing any unbound phage particles. The nanoparticles were washed three times with PBS pH 7.4. The nanoparticles were then resuspended in PBS pH 7.4 and moved to a new tube. This step aims to avoid collecting any phage adhered to the plastic wall of the first tube where binding was allowed to take place. The PBS pH 7.4 was removed again and the nanoparticles were incubated with 300 μl of 200 mM glycine, 0.2 % BSA pH 2.2 for 10 minutes (low pH elution of the phage). The supernatant was then collected into a fresh and

was neutralised with 500 μl of 1 M tris pH 7.4. Subsequently, the nanoparticles were washed with PBS pH 7.4 and incubated with 300 μl of 100 mM triethylamine pH 10 for 10 minutes (high pH elution of the phage). The supernatant was collected and mixed with the low pH elution. This process effectively washed off and collected any phage specifically bound to the magnetic nanoparticles. To confirm presence of and quantify the M13 phage a titration process combined with blue/white screening was carried out. Sterile conditions were used to prepare IPTG/Xgal (Isopropyl-β-D-thiogalactoside/ 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) containing LB agar plates. Pre-grown liquid medium *E. coli* ER2738 cells were then mixed with the eluted phage particles (from the previous screening step)(200 μl cells with 10 μl of phage) and were left to stand for 3 minutes at room temperature. During this step, the phage particles would infect the bacterial cells. The infected cells were then mixed with 50 °C top agar (0.7 % percent agar in LB) and applied evenly to IPTG/Xgal plates. The plates were incubated at 37 °C overnight. An overnight culture was also set up for procedures the next day. On day three, the IPTG/Xgal plates with phage infected *E. coli* ER2738 from day two were inspected for blue coloured plaques. These correspond to patches of *E. coli* cells that were infected with the one of the phage particles of the peptide displaying M13 library. The plasmid encoding the gIII gene for peptide display also encodes a β-galactosidase gene (downstream of an IPTG inducible promoter) the protein product of which catalyses Xgal to produce a visible blue product. No naturally occurring wild phage particles were detected as indicated by lack of any discoloured plaques observed during the process. Titration plates containing 10-20 plaques were used to count the plaques and calculate plaque forming unit concentration of the eluted phage mixtures (pfu/ml). In order to produce enough phage, particles for a subsequent screening round, a phage amplification process was carried out by mixing the whole of eluted phage from day two with 25 ml of LB in a 250 ml conical flask with 250 μl of an overnight culture of *E. coli* ER2738 cells and incubating them at 37 °C for 5 hours. During this process the phage would infect the *E. coli* ER2738 cells, replicate within them and escape back into the growth medium without destroying the host. After incubation, the phage particles were separated from the cells by centrifugation at 12,000 rpm for 10 minutes, twice. The supernatant containing the phage particles was then moved into a

new tube and was supplemented with a precipitation buffer (15 ml of phage particle suspension mixed with 3 ml of 2.5 M NaCl with 20 % w/v PEG8000) overnight at 4 °C. The highly hydrophobic and high salt conditions create an insoluble environment causing the phage particles to precipitate and drop out of solution. On day four, the phage particles in the precipitation buffer from day three were collected via centrifugation at 12,000 rpm for 15 minutes. A white precipitate was obtained (precipitated phage) and, after the removal of the supernatant, was resuspended in 1 ml PBS and centrifuged in a new tube for 12,000 rpm for 5 minutes. The supernatant was then transferred to another new tube and was supplemented with 200 μl of precipitation buffer and was incubated for 1 hour on ice for a second precipitation after which the solution was centrifuged again for 10 minutes at 12,000 rpm. The phage precipitate was resuspended in 200 μl of PBS pH 7.4 after the removal of the supernatant. The phage was then mixed with glycerol (for a 40 % final concentration) and stored in -20 °C. This was then used for another titration to quantify pfu/ml and as a starting material for the following titration rounds. After the third round, the phage particles were amplified and the DNA was extracted for Sanger sequencing.

Determining the Amino Acid Sequence of the Highest Binders

The phage display results were obtained by, first, preparing a sterilised deep well 24 well plate with 4 ml of *E. coli* ER2738 cells (prepared by mixing 500 μl of an overnight culture with 50 ml of LB medium with tetracycline) in alternating wells (to avoid cross-contamination). Individual blue plaques (12 in total) from a phage titration plate were picked and were used to inoculate the wells. The cells were then allowed to grow for 5 hours at 37 °C to allow the phage to multiply. The cell-phage suspensions from each well were mixed (1:1) with 80 % glycerol and labelled based on the well location (A1, A2, A3, B2, B4, B6, C1, C2, C3, D2, D4, D6) and stored at -80 °C as a stock. The rest of the suspension (10 μl) was then used to inoculate a set of wells in another 24 well plate prepared with 4 ml of LB medium. This was incubated overnight at 37 °C with 225 rpm agitation. Each of the wells was the processed using GeneJET Plasmid Miniprep Kits (Thermo Fisher Scientific, US) and the resulting plasmids were submitted for DNA sequencing with a sequencing primer -96gIII (5′ CCCTCATAGTTAGCGTAACTG 3′) which binds downstream of

the gIII gene in reverse. The 7 amino acid sequences were identified by translating the reverse complement of the sequencing results into the corresponding amino acid sequence and identifying motifs FYSHS-XXXXXXX-GGGS, where XXXXXXX is the variable region.

2.6 Protein Production and Analysis

2.6.1 Protein Sequence Analysis

Theoretical protein molecular weights, isoelectric points and extinction coefficients were calculated using ExPASy [199] online tools and the BioPython package for Python [200].

2.6.2 Phosphate Buffer

The solution referred to as 'phosphate buffer' throughout the project was prepared by dissolving Na_2HPO_4-$7H_2O$ and $NaH_2PO_4H_2O$ to make two separate 1 M solutions. To prepare a solution at pH 7.4 the 1 M solutions of Na_2HPO_4-$7H_2O$ and $NaH_2PO_4H_2O$ were mixed in a ratio of 80:20. The solution was then diluted to the desired molarity (20 mM most frequently). The pH was then measured and adjusted to pH 7.4 using either a solution of 20 mM Na_2HPO_4 or 20 mM NaH_2PO_4.

2.6.3 Protein Overexpression in *E. coli* BL21 DE3

A general protein over-expression protocol was followed for all of the proteins purified in the project. Large 2 litre flasks containing 400 ml of AIM supplemented with the appropriate antibiotic were inoculated with 1 ml of mature overnight cultures of *E. coli* BL21 DE3 cells transformed with different expression plasmids. The cultures were then incubated at 37 °C for 8 hours with shaking at 225 rpm to reach OD_{595} 0.5 and consume the present glucose and for induction of the T7 polymerase expression through lactose metabolism to occur. The temperature was then switched to 25 °C and incubated for 48 hours to allow for protein production to proceed. The cells were harvested by centrifugation at 3600 rpm for 20 minutes in 4 °C. The obtained pellets

were on average 8 g per 400 ml flask and were either stored at -80 °C or processed directly for subsequent purification steps.

2.6.4 Extracting Soluble Protein from Cell Pellets

A standard ultrasonic treatment was used for all lysis procedures during purified protein preparation. Pellets that were centrifuged immediately after the overexpression procedure or thawed from -80 storage (on ice for 0.5 to 1 hour) were processed in the same manner. The pellets were resuspended in the lysis buffer (specific lysis buffers are described below) at 20 % (w/v) with 100 μl EDTA-free protease inhibitor cocktail solution and 1 μl of benzonase. EDTA-free protease inhibitor cocktail solution was prepared by dissolving a single tablet in 1 ml of lysis buffer. The lysis was carried out with sample tubes being placed in ice-cold water. Ultrasonic treatment was carried out in intermittent bursts (20 seconds on:30 seconds off) for a total of 6 minutes 30 seconds. A centrifugation process was then carried out at 14,500 rpm for 45 minutes at 4 °C. The obtained soluble supernatant was then separated from the insoluble cell debris by pouring.

2.6.5 Ni^{2+} Affinity Purification

All affinity purification was performed using Ni-NTA (Ni^{2+} immobilised on nitrilotriacetic acid) agarose loose resins or Fast protein liquid chromatography (FPLC) columns (Amintra Ni-NTA Affinity Resin (Expedeon, UK) or HisTrap HP (GE Healthcare Life Sciences, USA). Each protein construct was designed to be produced containing N- or C-terminal poly-histidine tags (either 8, 6 or 5 histidine residues). Histidine residues are known to have affinity for metal ions such as Ni^{2+} or Co^{2+} (only Ni^{2+} was used in this project) [201]. The interaction between, histidine and the metal ions can be out-competed by introducing a high concentration of imidazole that also has affinity for the metal ions. In cases where loose resin was used, centrifugation clarified lysates were separated from the insoluble fractions (by pouring into a separate tube). The Ni-NTA resin was washed with ultra pure water to remove the storage solution (20 % ethanol) and was washed with binding buffer (imidazole (10-30 mM), NaCl (150 mM), tris (50 mM) pH 7.4) and was subsequently mixed with

the lysate and incubated between 1 hour to overnight at 4 °C. imidazole in the binding buffer prevents low affinity non-specific binding by native *E. coli* proteins. The bound resin was then collected in a bench-top gravity column and washed (wash buffer: imidazole (50 mM), NaCl (300 mM), tris (50) pH 7.4) to remove weakly associated proteins. The elution was performed with a buffer containing imidazole (400 mM), NaCl (150 mM), tris (25 mM) pH 7.4). At such a high concentration imidazole starts occupying Ni^{2+} sites instead of the poly-histidine tagged proteins, thus causing them to be eluted. Elutions were performed in 0.5 ml fractions which were analysed using Sodium Dodecyl Sulphate–Polyacrylamide Gel Electrophoresis (SDS-PAGE) for presence of the protein of the correct size and then pooled where appropriate. Where FPLC was used, centrifugation clarified lysates were filtered (0.4 μm) before application to a 150 ml Superloop (Cytiva, USA) on an ÄKTA pure (Cytiva, USA) system. The samples were eluted in imidazole (300 mM), NaCl (150 mM), tris (25 mM) pH 7.4, collecting fractions automatically in a 96 deep well plate. Protein presence was detected via 280 nm absorbance detection module in milli-absorbance units (mAU). The elution peaks were pooled. The elution buffer is incompatible with most downstream applications, thus it was exchanged into a working buffer by using a desalting column, size exclusion chromatography or dialysis.

2.6.6 Desalting Columns

Quick buffer exchange was performed using HiTrap Desalting columns connected to an ÄKTA pure (Cytiva, USA) system. The columns are prepacked with Sephadex G-25 resin that traps small molecules causing large molecules, such as proteins, to be eluted first. Since the columns are equilibrated in the desired buffer, the eluted protein is also eluted in the said buffer.

2.6.7 Size Exclusion Chromatography

Where maximal protein purity was required, Ni^{2+} affinity purification was followed by size exclusion chromatography (SEC). The technique operates on the same principle as desalting columns, but allows higher resolution fractionation of different

protein species afforded by the higher total volume of the packed resin. Either a Superdex 200 (SP200) Increase 10/300 GL column (Cytiva, USA) or a HiLoad 16/600 Superdex 75 pg column (Cytiva, USA) were used for volumes up to 0.5 ml or 5 ml, respectively. The columns were washed with ultra pure water to remove the storage solution (20 % ethanol) and were pre-equilibrated in the target buffer prior to sample application. The samples were run at the flow rate of 0.4 ml/min. The presence of the protein was tracked using a 280 nm absorbance detection module and elutions were fractionated at a volume of 0.5 ml in 96 deep well plates. The fractions were then analysed on SDS-PAGE and were pooled where appropriate. All procedures were performed in a 4 °C cold room. All buffers were pre-incubated in a 4 °C to avoid air bubbles from occurring.

The Superdex 200 (SP200) Increase 10/300 GL column was also used for determining the molecular weight of proteins. This can be performed by calibrating the column with globular proteins of known molecular weight and calculating the expected elution volumes for a given molecular eight. The calibration curve obtained can be seen in Figure 2.1.

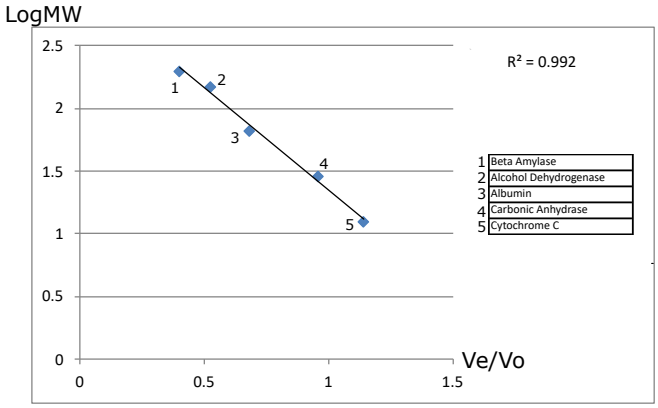


FIGURE 2.1: A calibration curve derived from known molecular weight standards. Y axis - log10(molecular weight of the standard), X axis elution volume divided by void column volume.

2.6.8 Increasing Protein Concentration

Purified protein concentration was increased where needed using Amicon Ultra-15 Centrifugal Filters (Merck, US). Columns with filters with 10,000 kDa cut-off were

used. The technique uses centrifugal force to promote buffer and any components below 10,000 kDa to pass a filter, the protein is thus retained in a smaller final volume resulting in a suspension containing a higher concentration of protein. Samples were applied to the top section of the column and were spun at 4700 rpm until the desired volume was reached (20 - 60 minutes).

2.6.9 Dialysis

Dialysis offers a reliable way of removing undesirable buffer components and replacing them with others. A sample with a mixed buffer composition is placed in a membrane with a known molecular weight cut-off (3,500 or 10,000 in this project) which is then sealed and placed into a reservoir with the desired final buffer. The sample to reservoir volume ratio was 1:1000 and dialysis was repeated 3 times, at least 4 hours each, preparing a fresh buffer each time. The procedure was performed in 4 °C.

2.6.10 Sodium Dodecyl Sulphate–Polyacrylamide Gel Electrophoresis

SDS-PAGE was used for routine investigations of sample purity and for preliminary estimation of protein molecular weight. The technique can be used to separate proteins based on their molecular weight alone by utilising a porous polyacrylamide matrix and an applied electrical current. Influence on mobility caused by protein secondary structure or net charge (originating from a unique amino acid composition) is eliminated by the presence of a molar excess of SDS that unfolds the protein (by adhering to the polypeptide chain) and grants it a negative charge. The resulting net charge of the SDS denatured protein molecule is proportional to the length of the polypeptide chain. Bis-Tris Precast Gels (Expedeon, UK) were used for medium to large proteins(10 - 200 kDa proteins), whereas Mini-PROTEAN Tris/Tricine Precast Gels (Bio-Rad, US) were used for small proteins (<10 kDa). All samples were prepared by mixing 1:3 with NuPAGE LDS Sample Buffer (4X) (Thermo Fisher Scientific, US) and heating them to 90 degrees for 5-10 minutes for initial denaturation. The sample buffer contains LDS (equivalent to SDS), Coomassie G250 and Phenol

Red for tracking electrophoresis progress and glycerol to cause the sample to be deposited in the gel wells. The sample buffer also adjusts the sample pH to 8.5. The gels were run by applying a constant 200 V current for 40-55 minutes after which they were stained with InstantBlue Coomassie Protein Stain (Expedeon, UK) for 15 minutes or longer. The gels were the washed and visualised using ChemiDocGel Imaging System (Bio-Rad, US). PageRuler™ Prestained Protein Ladder, 10 to 180 kDa (Thermo Fisher Scientific, US) was run along side any samples for molecular weight estimations.

2.6.11 Expression-tag Cleavage

pET28a encodes an amino acid sequence (LVPRG) downstream of the N-terminal poly-histidine tag that can be recognised by thrombin protease for specific polypeptide chain cleavage. Thrombin CleanCleave Kit (Sigma-Aldrich, USA) was used where thrombin cleavage was necessary. The kit includes a thrombin protease immobilised on agarose beads for easy separation of the enzyme from the processed proteins. In short, the reactions were allowed to proceed in 50 mM Tris-HCl pH 8 with 10 mM $CaCl_2$ for 16 hours at room temperature. The resulting reaction products were separated from the thrombin agarose and were applied to equilibrated Ni-NTA resin in order to capture the cleaved poly-histidine tags and any un-cleaved proteins. The cleaved proteins were collected in the flow through and were buffer exchanged using an appropriate method.

2.6.12 Protein Quantification

Protein concentration was determined by measuring the absorbance at 280 nm of purified protein samples and calculating the concentration using a derivation from Beer's law:

$$c = A/\epsilon * L \tag{2.1}$$

Where A is absorbance, ϵ - extinction coefficient and L - path length of the measurement cell. The extinction coefficients were determined using ProtParam (ExPASy). The absorbance readings at 280 nm were taken using a VersaWave spectrophotometer (Expedeon, UK). Alternatively, protein concentration was measured by performing the Bicinchoninic acid assay (BCA). The assay offers the advantage of giving signal proportional to the amount of polypeptide chains (rather than presence of highly 280 nm absorbing residues tryptophan or tyrosine). The presence of peptide bonds causes Cu^{2+} to be reduced to Cu^{+} proportionately. The Cu^{+} ions then form a purple-coloured complex that absorbs light at 562 nm. A typical BCA assay was performed by preparing 100 μl of bicinchoninic acid and Cu^{2+} mixture and inoculating it with a serial dilution of known concentration bovine serum albumin (BSA) and the unknown concentration samples at 10 μl in 96 well plate. Each inoculation was performed in duplicate. The plates were then incubated at 37 °C over-night. Readings at 562 nm were then taken. Standard concentration curves were then obtained based on the BSA readings and concentrations of the samples were calculated.

2.6.13 Protein-Blots

Western blots and dot blots were used to detect specific presence of proteins containing a poly-histidine tag. Nitrocellulose membranes for the Western blot were obtained by performing SDS-PAGE analysis and transferring the eletrophoretically separated proteins using the Trans-Blot Turbo Transfer System (Bio-Rad, USA). Dot blot membranes were obtained by spotting 5μl of either pure protein or cell lysate containing the protein onto a nitrocellulose membrane. In both cases the membrane was then placed into a casein blocking buffer (Sigma-Aldrich, USA) diluted in tris buffer saline pH 7.4 supplemented with 0.5 % (v/v final) tween 20 (TBS-T) for overnight at room temperature with gentle agitation. The blocking solution was then removed and fresh blocking solution mixed with anti Penta-His Tag Antibody conjugated with horse radish per-oxidase (HRP) (QIAGEN, Germany) was added and incubated for 1 hour at room temperature. The antibody solution was removed and

the membranes were washed with (TBS-T) three times 5 minutes each. The antibodies immobilised on the membrane bound proteins were then visualised by incubating the membranes with ECL Western Blotting Substrates (Bio-Rad, USA) containing luminol and hydrogen peroxide. The two reagents applied to the membranes containing HRP bound to the antibody cause the luminol solution to be catalysed and emit light at 428 nm. The emission was detected in a ChemiDocGel Imaging System (Bio-Rad, US).

2.6.14 Native-PAGE

Native-PAGE (or non-denaturing PAGE) is a technique that allows for separation of protein without denaturation. The technique parallels the steps followed in SDS-PAGE, however SDS is not included and thus the proteins are not unfolded or charge neutralised. The gels were poured using Mini-PROTEAN Tetra Handcast System (Bio-Rad, USA). The glass plates used for making Native-PAGE gels were dedicated for this procedure (and not used for making SDS-PAGE gels to avoid any introduction of a denaturant). The gels were made up of two layers: a stacking gel (low acrylamide) to line up all of the applied protein so that they enter the resolving gel (high acrylamide) at the same time, where proteins are then separated. The stacking gel was prepared in 0.5 M tris HCl pH 6.8 solution and contained 4 % acrylamide. The resolving gel was prepared using 1.5 M tris HCl pH 8.8 with variable amounts (7.5 - 12 %) of acrylamide, depending on the experiment. Stacking and resolving gel formation was catalysed by a solution of 10 % (w/v) amoniumpersulfate (APS) and tetramethylethylenediamine (TEMED), at a final concentration (v/v) of 1 and 0.1 %, respectively. To avoid thermal denaturation of the proteins caused by the heat from the running current, the gels were run at 4 °C at a low voltage (80 V). The gels were visualised following an InstantBlue Coomassie Protein Stain (Expedeon, UK) treatment.

2.6.15 Cysteine-dye Labelling

Protein containing cysteine residues can be chemically linked with dye molecules through maleimide groups that react to sulfhydryl groups. This project uses DyLight 650 Maleimide dyes (Thermo Fisher Scientific, US) that fluoresce at 672 nm wavelength. The purified cysteine containing proteins (500 mg/ml) were exchanged (desalting columns) into a 100 mM sodium phosphate pH 6.0 solution containing 10 mM DTT and 2.5 mM EDTA and incubated for 90 minutes at 37 °C to reduce any disulphide bonds. The proteins were then exchanged into PBS with 1 mM EDTA. DyLight 650 Maleimide dye resuspended in DMF (1 mg in 100 μl) was then added to the protein suspension (20 μl for 1 ml of protein) and incubated at room temperature overnight. The excess dye was removed from the solution via dialysis (10,000 Da cut-off). Comparison of absorbance at 280 nm (total protein) to 652 nm (dye) was used to determine the degree of labelling.

2.6.16 Electro Spray Ionisation - Time of Flight Mass Spectrometry

Electro spray ionisation - Time of Flight mass spectrometry (ESI-TOF) provides an accurate method to determine the molecular weight of proteins with sub-Dalton precision [202]. Electron spray ionisation converts proteins into gaseous species. The gaseous proteins are then separated in the mass spectrometer based on the mass to charge ratio (m/z) with smaller ions with the same charge arriving at the detector earlier. The ionisation method used retains covalently interacting atoms intact. After deconvolution the mass spectrum of the sample can be obtained. Liquid chromatography (LC) was carried out using Agilent 1260 Infinity instrumentation using a Phenomenex Aeris Widepore XB-C18 2.1mm x 50mm, 3.6 micron column. 0.1 % formic acid was used as solvent A and acetonitrile solution with 0.1 % formic acid was used as solvent B. The gradient procedure was performed over 15 minutes (5% solvent B to 95% solvent B) at a flow rate of 0.4 ml/ml. The injection volume was 1-10 μl (concentration of up to 0.5 mg/ml). Agilent Technologies 6530 Accurate Mass LC-MS QToF was employed for ESI-TOF. Positive ion mode was performed with the m/z range of 100-3200 used. The drying gas temperature was 350 °C. All LC and MS steps were Operated by University of Sheffield, Department of Chemistry, Mass

Spectrometry facilities. The obtained masses from deconvoluted results were compared to calculated molecular weight with and without an N-terminal methionine residue. Methionine is occasionally removed in *E. coli* during protein production and is determined by the residues in the sequence immediately downstream of the initiating methionine [203]. Methionine is removed if the second residue is either proline, serine, glycine, valine, alanine or threonine.

2.6.17 Circular Dichroism

Circular dichroism (CD) is a technique that can be used to quickly estimate protein secondary structure [204]. It operates by detecting differences between the sample absorbance of circularly polarised light in the right and left directions. Different polypeptide backbone geometries result in different levels of absorption of the left and right circularly polarised light waves. This forms the basis of detecting unique traces of circular dichroism along near-UV range of wavelengths. Over years of experimental analysis, certain circular dichroism spectra have been observed to correspond to specific protein secondary structure motifs (Figure 2.2).

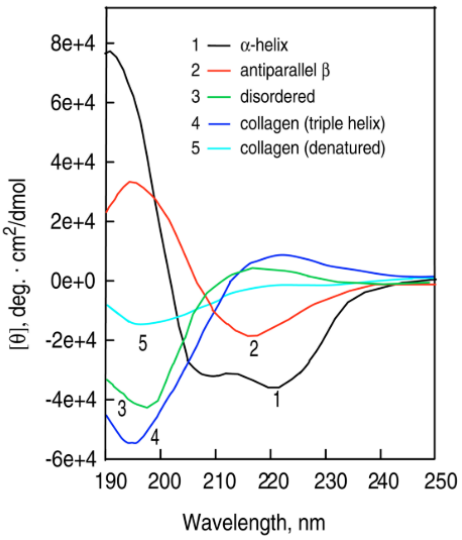


FIGURE 2.2: Commonly observed CD spectra for different secondary structure motifs. Adapted from Greenfield, 2007 [231].

Circular dichroism was carried out using J-810 Spectropolarimeter (Jasco, Japan) equipped with a Peltier thermoelectric type temperature control unit. Pure protein

samples were prepared in a capped 1 mm path length cuvette at 100 μg/ml concentration (although experiment specific adjustments have been made). Buffers high in Cl^- were avoided to prevent interference with the CD signal. Measurements of CD spectra were taken in triplicate within the wavelength range of 280 to 190 nm, with wavelength pitch of 1 nm. Buffer spectra were also obtained and were subtracted from the sample spectra. Spectra acquisition was performed either at a single 25 °C temperature point or over a temperature range between 5 and 90 °C with a temperature pitch of 5 °C. Thermal denaturation curves were obtained by obtaining CD signals at discrete wavelength positions (e.g. at 222 nm) with gradually increasing temperature (pitch 1 °C). The data were converted to units of mean residue ellipticity for comparison using the formula:

$$\theta_{\mathrm{MRW}} = \frac{MRW \cdot \theta_{\mathrm{obs}}}{10 \cdot d \cdot c} \tag{2.2}$$

Where θ_{MRW} is mean residue ellipticity, MRW - mean residue weight (calculated by dividing the molecular weight of the polypeptide by number of amino acids in the chain), θ_{obs} - CD signal, d - path length of the cuvette and c - concentration of the protein in g/ml. The units obtained are degrees cm^2 $dmol^{-1}$ $residue^{-1}$.

Circular dichroism spectra can be deconvoluted using a selection of empirical algorithms that can be accessed through the Dichroweb server [205]. The server provides access to various machine learning algorithms trained on standardised data sets of CD spectra annotated with known secondary structure quantities (determined through other methods such as X-ray crystallography). Spectra obtained in this study were analysed using the CDSSTR algorithm [206] with four different data-sets (Set4 [206], Set7[206], SP175 [207] and SP180) and secondary structure values were averaged and plotted with standard deviations (population).

2.6.18 Standard Soluble Protein Purification

All protein (unless otherwise noted) in Chapter 3, Chapter 4 and Chapter 6 were purified using a standard protocol. *E. coli* BL21 cells containing the appropriate expression plasmids were grown using the standard protocol in AIM (above). The cells were either used immediately after or frozen at -80 °C and thawed in room

temperature before use. For purification at scale, 3-4 cell pellets were lysed in NaCl (150 mM), imidazole (20 mM), tris HCl (50 mM) pH 7.4 with protease inhibitors and benzonase (and 1 mM DTT if the proteins contained cysteine residues). The soluble fraction was separated via centrifugation at 12,500 rpm at 4 °C for 45 minutes. The soluble fractions were pooled and filtered with 0.4 μm syringe filters and subjected to HisTrap HP columns. The proteins were eluted in NaCl (150 mM), imidazole (300 mM), tris HCl (50 mM) pH 7.4 and the elution fractions were pooled and applied to a HiLoad 16/600 Superdex 75 pg column (Cytiva, USA) equilibrated in either PBS pH 7.4 or Na_2HPO_4-$7H_2O$/$NaH_2PO_4H_2O$ phosphate buffer (20 mM) (with 1mM DTT if cysteine residues were present). The fractions corresponding to absorbance peaks were analysed by SDS-PAGE and fractions containing the protein of the expected size were pooled. Protein concentrations were measured using the BCA assay or UV-vis spectroscopy (if they contained DTT as it causes a colour change in the BCA reaction). Proteins were either stored at 4 °C for up to 4 days or at - 4 °C.

2.6.19 Coiled-coil-SUMO Fusion Protein Purification

E. coli BL21 cells containing the appropriate pET24a plasmids with His-SUMO-cc constructs were grown using the standard protocol in AIM. The cells were lysed in NaCl (150 mM), imidazole (20 mM), tris HCl (50 mM) pH 7.4 with protease inhibitors and benzonase, centrifuged at 12,500 rpm and the soluble fraction was filtered using 0.4 μm syringe filters subjected to HisTrap HP columns. The eluted fractions corresponding to the peaks were then pooled and subjected to a desalting column and exchanged into tris (20 mM) pH7.4, NaCl (150 mM), DTT (2mM) for cleavage with Ulp1 protease. The cleavage reaction was allowed to take place for 2 hours at room temperature (the degree of success was determined by performing SDS-PAGE with Mini-PROTEAN tris/Tricine Precast Gels). The cleaved and uncleaved His-SUMO proteins were removed by subjecting the reaction mixtures to Ni-NTA benchtop columns. The collected flow-through containing the cleaved α-helices was exchanged into PBS pH 7.4 using a desalting column. The peptide concentration was determined using UV-vis spectroscopy.

2.6.20 S-layer Protein Denaturing Purification

The S-layer proteins containing pellets were produced using the standard protein expression protocol in AIM (see above). The purification and protein re-assembly protocol was adapted from literature[208]. The cells were lysed via ultrasonic treatment in NaCl (150 mM), imidazole (20 mM), tris HCl (50 mM) pH 7.4 containing 8 M urea to disrupt any S-layer sheets formed within the cell during growth. The soluble fraction (containing urea solubilised S-layers) was separated using centrifugation at 12,500 rpm at 4 °C for 45 minutes. The supernatant was then transferred to fresh tubes. 0.5 ml of Ni-NTA loose agarose beads were added to each 40 ml of solubilised cell lysate. The binding was carried out overnight at 4 °C. The resin was then washed with NaCl (150 mM), imidazole (20 mM), tris HCl (50 mM) pH 7.4 and the S-layer proteins were eluted in NaCl (150 mM), imidazole (300 mM), tris HCl (50 mM) pH 7.4. The eluted protein suspensions were then placed in dialysis tubing with 10,000 Da cut off. The tubing was placed into 10 mM $CaCl_{2+}$ solutions at 4 °C with gentle agitation for at least 4 (and no longer than 16) hours. The solution was changed to ultra pure water and was incubated for another 4 hours. The last step repeated additional two times. A white (or green in case of SgsE-EGFP) precipitate appeared after the second incubation with ultra pure water. The precipitated solution (corresponding to assembled S-layer proteins) was then removed from dialysis tubing and stored in sterile tubes until needed at 4 °C. The concentration was measured using the BCA assay or UV-vis. UV-vis spectroscopy was carried out by solubilising the purified S-layers with 6 M urea (1:1 dilution, 3M final urea) and then taking the measurement with 3 M urea as the blank solution. The S-layer re-assembly procedure is depicted in Figure 2.3.

2.6.21 Coiled-coil Scaffold Protein Purification

Chapter 6 demonstrates purification and assay of a loop exposing coiled-coil protein (E8cc) which was shown to be prone to forming inclusion bodies in the cell in a previous study, where a method for denaturing purification and refolding was proposed [209]. The cells containing the E8cc sequence in pPR-IBA1 were grown using the standard over-expression protocol in AIM (see above). The obtained cell pellets

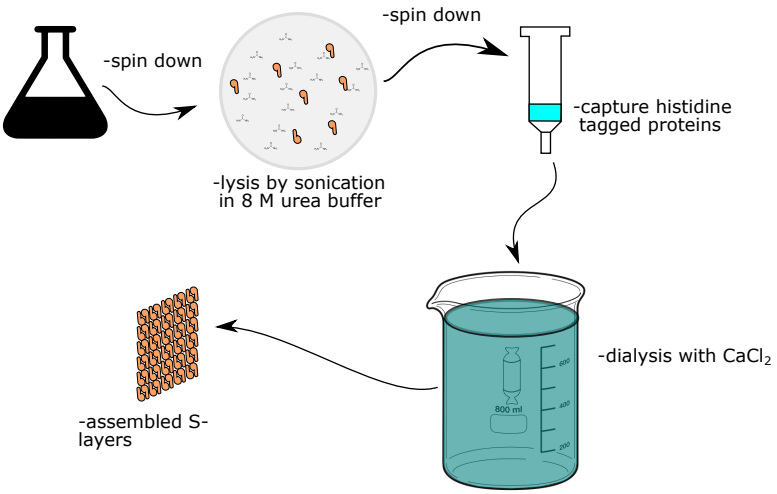


FIGURE 2.3: A diagrammatic representation of SgsE purification and re-assembly strategy. The S-layer proteins expressed in *E. coli* cells are collected by centrifugation. After the removal of growth medium, the cells are lysed through sonication in a buffer containing 8 M urea which solubilises the S-layer protein monomers. The lysates are then clarified by removing insoluble debris through centrifugation. Poly-histidine tag containing S-layer monomers are then captured, washed from undesired proteins and eluted using Ni-NTA gravity columns. The eluted proteins are subjected to dialysis with $CaCl_2$ followed by dialysis in ultra-pure water to remove residual urea and induce S-layer protein self-assembly into S-layer sheets.

were lysed in NaCl (150 mM), tris HCl (25 mM) pH 7.4 with 8 M GuHCl using ultrasonic treatment. The lysate was then centrifuged for 30 minutes at 12,000 rpm. The supernatant was then incubated with Ni-NTA loose agarose beads (equilibrated in the same buffer) for 30 minutes. The resin was then placed into a bench top column. Unbound solution was allowed to flow through. The column with the resin was then washed in NaCl (150 mM), tris HCl (25 mM) pH 7.4 buffer with decreasing concentrations of GuHCl (4 M, 2 M, 1 M, 0.5 M). This was followed by a wash in NaCl (150 mM), imidazole (40 mM), tris HCl (25 mM) pH 7.4 and an elution step in NaCl (150 mM), imidazole (400 mM), tris HCl (25 mM) pH 7.4. The eluted protein was then subjected to SEC in Superdex 200 (SP200) Increase 10/300 GL columns with PBS pH 7.4 as the elution buffer. The concentration was determined via UV-vis spectroscopy. The proteins were stored at -20 °C.

2.6.22 Commercially Synthesised Peptides

Peptides A1 - YHPLRNH, A5 - KSPIHVP and B6 - GHSQQST discovered via Phage Display in Chapter 6 were synthesised by Genscript (US) through 'Fast Peptide Synthesis' service. The peptides were received at 85 % purity and were re-suspended in ultra-pure water to a final concentration of 8 mg/ml. The suspensions were then stored at -20 °C in small aliquots.

2.6.23 Dynamic Light Scattering

Dynamic light scattering (DLS) is a technique that allows for indirect measurement of the hydrodynamic radius by calculating it from the obtained information on the movement of particles. Small particles move more quickly than large particles [210]. The basic experimental set-up uses an incident laser directed at the cuvette and a detector that registers the patterns in scattering. The relationships within the system can be described by the equation:

$$D = \frac{k_B T}{6\pi\eta R_H} \tag{2.3}$$

Where, D is the diffusion coefficient (stands in for particle speed), k_B is the Boltzmann constant, T is the temperature, η is viscosity and R_H is the hydrodynamic radius.

DLS was performed using ZetaPals Zeta potential Analyzer (Brookhaven Instruments, US) using the particle size analysis function. Protein samples were prepared at 100 μg/ml in 20 mM phosphate buffer pH 7.4. The proteins were centrifuged at 12,500 rpm (bench-top centrifuge) for 10 min to remove any particulates and only the supernatants were used. The samples were prepared in clean UV compatible cuvettes. The cuvettes were capped and general sterile technique was followed to prevent bacterial contamination as the experiment duration spanned multiple days with proteins being stored at room temperature. The obtained hydrodynamic radius readings were only considered as a means to track appearance of large objects (fibrils) rather than determining sizes of discrete protein sub-unit lengths. The DLS

chart presented was generated by plotting time against the species most predominant within a given sample at the time point.

2.6.24 Quartz Crystal Monitoring with Dissipation

Quartz crystal monitoring with Dissipation (QCM-D) is a technique that relies on acoustic vibrations obtained by oscillating a piezoelectric crystal [211]. The oscillations are achieved via application of alternating current to the piezoelectric crystal. The QCM-D apparatus then records the frequency of the oscillation. The current is then short circuited and the dissipation of the oscillation is also recorded. Mass adsorbed onto QCM-D crystals (that can also be coated) affects the frequency and dissipation values. Frequency is reduced as more mass is accumulated, whereas dissipation depends on the elasticity of the material (more elastic materials show higher increases in dissipation). The changes in mass can be calculated using the following equation (Sauerbrey equation):

$$\Delta m = \frac{C}{n}\Delta f \tag{2.4}$$

Where, Δm is the change in mass, C is –17.7 Hz ng/cm^2 (constant value for crystals with resonant frequency of 5 MHz which depends on the crystal thickness and density), n is the harmonic number (frequency and dissipation can be recorded at different multiples of the resonant frequency - 3, 5, 7,...) and Δ f is the change in frequency.

The experiments were performed using the Q-Sense E4 QCM-D (Q-Sense AB, SE) system. Gold-coated QCM-D sensors (with resonant frequency of 5 MHz) were cleaned using a UV-ozone treatment for 10 minutes. The sensors were then ultrasonically treated in a 2 % Hellmanex solution, and then two times in ultra-pure water followed by drying with nitrogen and a final UV-ozone treatment for 30 minutes. Finally, the sensors were treated with 20 % ethanol. The sensors were mounted and the system was equilibrated in 20 mM phosphate buffer pH 7.4. In experiments with 20 mM phosphate buffer pH 4.5 washing, these buffers were run prior to application of the protein to observe the magnitude of the effect on the readings. The proteins

were prepared at 150 μg/ml in 20 mM phosphate buffer pH 7.4 and were applied to the systems at a flow rate of μ/min at 20 °C. The results were graphed by plotting time against the 3rd overtone reading of dissipation and frequency changes. The data plotting mass on surface was calculated using the Sauerbrey equation.

2.6.25 Localised Surface Plasmon Resonance

Localised Surface Plasmon Resonance (LSPR) is a technique that allows for detection of changes in mass on surface by tracking shifts in resonance peak position caused by a change in the surface refractive index [212]. Xnano LSPR instrument (Insplorion) was employed in Chapter 4 to detect protein-protein interactions. The apparatus uses sensor chips containing a gold layer coated with glass and an outer gold layer (for immobilisation). The experiments were run by confirming the presence of the plasmon peak corresponding to the gold sensor (520 nm) in air and then supplying the system with the buffer to be used in the experiment. The solution was supplied to the system using a peristaltic pump. The temperature of the experiments was observed (did not exceed 26 °C). Proteins were applied at 150 μg/ml in TBS pH 7.4 until saturation. Centroid position (a mathematical estimation of the resonance peak position) was used to track the material on the surface. After saturation the material was washed with the same buffer and the binding partner was introduced. Surface immobilisation was achieved through non-specific interactions with the gold surface. The estimation of the number of interactions on the surface was performed as follows. First an assumption was made that the protein applied first had fully saturated the surface. Then the fraction of total mass for each protein applied to the surface was estimated (by comparing the LSPR signal at the point of washing after the addition of each protein layer) and assuming that the highest point of signal after washing corresponds to 100 % mass on surface. From the mass ratio a ratio of moles was determined by taking into account the molecular mass of the protein molecules.

2.7 Protein Structure Visualisation and Modelling

2.7.1 Protein 3D structure visualisation

Protein structure visualisations were obtained using PyMOL [213] or ChimeraX [214]. The structural information was obtained from PDB (rcsb.org [215]). Where exact structural information was not available (in the case of SasG coiled-coil construct visualisation or visualisation of mutated binding loop regions) trRosetta modelling was used [216] (accessed through robetta.bakerlab.org [217]). trRosetta builds the protein structure based on energy minimisation. Restraints (derived from multiple sequence alignments) are included to make the process more efficient. The algorithm also refers to homologous templates of known structure to inform predictions.

2.8 Nanoparticle Synthesis

2.8.1 CoPt

CoPt A1

Cobalt platinum A1 crystal lattice nanoparticle synthesis was performed under anaerobic conditions in ultra pure water. First, ultra pure water was degassed to remove any oxygen. Constituent metal salts $CoSO_4 7H_2O$ (30 mM) and Na_2PtCl_4 (10 mM) were then resuspended in the degassed water. In case of the peptide additive study, the metal salts were supplemented at this step for a final peptide concentration of 0.1 or 0.01 mg/ml. The metal salt solutions were mixed in a 1:1 ratio in a reaction vessel with nitrogen sparging for 5 minutes (performed in the fume cupboard). Once the salts had fully dissolved a freshly made solution of $NaBH_4$ (25 mM) was added to the salt mixture (ratio 1:1 with the salts combined). The clear solutions produced a dark precipitate immediately and the reactions were left to run to completion for 45 minutes. The products were magnetically collected and washed three times with degassed ultra pure water. The samples were then dried in a vacuum oven overnight at 60 °C.

Annealed CoPt $L1_o$

The annealing procedure was performed by placing dry CoPt A1 nanoparticles in a quartz cuvette cleaned with a mixture of nitric and hydrochloric acids (molar ratio 1:3) for 5 minutes (care must be taken when performing this step). Single zone horizontal tube furnace (TSH18/75/300, Elite Thermal Systems) was used for the heating protocol. The samples were inserted into the furnace at room temperature and heated at the rate of 1 °C /minute until it reached 800 °C and was then kept for 1 hour. The samples were then slowly cooled down to room temperature. The heating was performed in an Argon gas atmosphere.

2.8.2 Magnetite

Magnetite 30-50 nm

Room temperature co-precipitation reactions were performed to produce 30-50 nm sized magnetite nanoparticles. Ultra pure water was prepared by nitrogen sparging for 30 minutes prior to the experiment to prevent oxidation of the metal salts. $Fe_2(SO_4)_3$ and $FeSO_4$ were prepared at 1:1 ratio to make up a solution containing a total of 50 mM Fe dissolved in ultra pure water. The salts were then placed in a reaction vessel with stirring and nitrogen sparging. If protein additives were used, they were injected at this step. The formation of the nanoparticles was triggered by a controlled 20 μl/min addition of NaOH (500 mM) (a 10 ml iron suspension required 4 ml of NaOH (50 mM)). The reactions were allowed to proceed for 3 hours. The products were then magnetically collected and washed three times with degassed ultra pure water. The samples were then dried in a vacuum oven overnight at 60 °C.

Magnetite 10 nm

Reverse addition room temperature co-precipitation reactions were performed to produce 10 nm sized magnetite nanoparticles. Ultra pure water was prepared by nitrogen sparging for 30 minutes prior to synthesis to prevent oxidation of the metal salts. A reaction vessel was prepared with 12 ml of ultra pure water mixed with 8 ml of NaOH (500 mM) and continuous nitrogen sparging. Iron metal salts $Fe_2(SO_4)_3$ and $FeSO_4$ were weighed for 0.1489 g and 0.139 g, respectively, and were dissolved

in 8 ml of ultra pure water. The iron suspension was then drip-fed into the reaction vessel at the continuous rate of 50 μl/min. The reactions were allowed to proceed for 160 minutes, after which the reaction products were magnetically collected and washed with ultra pure water three times before drying in a vacuum oven overnight at 60 °C.

2.9 Nanoparticle Characterisation

2.9.1 Powder X-ray Diffraction

X-ray diffraction allows for identification of crystalline material [218]. An X-ray diffraction apparatus consists of an electron source, a sample plate and an X-ray detector. X-rays can be scattered by atoms of a crystal surface and the diffraction can be detected at an angle (2θ). Different diffraction patterns correspond to different crystalline materials. XRD was performed in a Bruker D8 Powder diffractometer. Dried particle samples were ground down to a fine powder and were placed on Si low background sample holders. The resulting diffraction spectra were background subtracted before analysis. The scans were performed between 20-90 °in 0.02 °steps each lasting 1.5 seconds. 40 keV X-rays were generated at with a copper source (wavelength = 1.54049).

2.9.2 Estimating Particles Sizes Using the Scherrer Equation

Nano-scale crystallite sizes can be calculated from XRD spectra using the Scherrer equation [219]. It has been observed that small crystalline domains cause XRD peaks to broaden. This relationship is described as.

$$d = \frac{k \cdot \lambda}{FWHM \cdot cos(2\theta)} \tag{2.5}$$

Where d is the mean crystallite diameter in the sample, k is the shape factor (0.9 can be used for spherical crystals [220]), FWHM - full width at half maximum (width of the chosen peak at half of the total intensity measured in degrees), 2θ - the angle of the peak chosen and λ is the wavelength of the X-rays used. In this project

2θ = 40 °was chosen as it was the dominant peak, allowing for the most accurate measurement. Peak width measurements performed taken by hand.

2.9.3 Magnetic Susceptibility Measurements

The magnetic properties of synthesised nanoparticles were assessed by measuring magnetic susceptibility (χ) which measures how much of a given material may become magnetised in an applied magnetic field. The technique can be summarised using the following formula.

$$\chi = \frac{M}{H} \tag{2.6}$$

Where χ is magnetic susceptibility, M is the magnetic moment and H is the intensity of the applied field. Bartington MS2G susceptibility measurement system was used. Blank readings were first obtained by placing empty 0.6 ml centrifuge tubes into the sample holder. Materials of known mass were then placed and readings were taken in triplicate. Plotted values were normalised for mass.

2.10 Probing for Protein-Nanoparticle/Metal Interactions

2.10.1 Magnetite ELISA

An adaptation to enzyme linked immunosorbent assay (ELISA) was used to probe interactions between magnetite nanoparticles and various proteins [172]. The assay involves 'blocking' well dispersed nanoparticles in a non-specific protein rich solution, incubating protein of interest and then probing for the presence of the said protein with protein of interest specific antibodies.

All steps throughout the assay were performed at room temperature. The incubation steps were carried out on a rotisserie mixer. Here, particle solutions at 30 mg/ml were ultrasonically treated for 1 minute at 40 % amplitude. The particles were then distributed between 1.5 ml centrifuge tubes, 30 μl in each tube. A solution of 2x caseine blocking solution (Merck, US) diluted in PBS-Tween (Tween final concentration 0.5 % v/v) was then added to the particles for a 1 hour incubation. Afterwards,

10 μl of each protein to be tested was added (typical concentrations ranged from 0.5 - 1 mg/ml) directly to the blocking solutions for another 1 hour incubation. The particles were then magnetically collected, supernatants were removed and the particles were washed three times with PBS-Tween. An anti-poly-histidine antibody (rabbit) (Cell Signalling Technology, US) diluted in the blocking solution (1:2000) was then added for a 1 hour incubation. The solution was then removed and particles were washed three times again. An anti-rabbit alkaline phosphatase solution (Thermo Fisher Scientific, US) diluted in the blocking solution (1:10000) was then added and incubated for 1 hour. The solution was then removed and the particles were washed with PBS-Tween. The particles were then re-suspended in 500 μl of PBS-Tween and transferred to fresh 1.5 ml centrifuge tubes. This step was included to separate the particles from any antibody non-specifically bound to the tube plastic. The particles were then washed two more times with PBS-Tween which was then removed. The magnetically separated particles were then exposed to 300 μl of BluePhos Substrate (VWR, US) containing 5-bromo-4-chloro-3-indoyl phosphate each. The solution was then incubated with the particles for 15 min to 10 hours (depending on the protein concentration used). During the incubation the alkaline phosphatase would catalyse 5-bromo-4-chloro-3-indoyl phosphate resulting in a product with absorbance at 600 nm. Absorbance readings then were taken using FLUOstar Omega Microplate Reader (BMG Labtech, UK).

Each protein to be tested was run in triplicate and plate readings were performed in duplicate. Each experiment was internally controlled (suspected binding proteins were run together with known non-binding ones). Each experiment also included a control with particles without any protein to be tested to determine the levels of background catalysis of the substrate.

2.10.2 Phage ELISA

Phage ELISA was performed to investigate the binding of the phage to $L1_0$ and A1 CoPt nanoparticles. The principle of the assay is similar to magnetite ELISA. Particle solutions are 'blocked' in protein solution and are incubated with phage particles. A

specific phage binding antibody is then used to detect if any phage was retained after washing.

All steps throughout the assay were performed at room temperature. The incubation steps were carried out on a rotisserie mixer. Here, 2.5 mg of CoPt particles were weighed and resuspended in 1.5 ml of PBS. The solution were then ultrasonically treated for 2 minutes at 40 % amplitude and distributed equally into 1.5 ml tubes (100 μl each). The 2x caseine blocking solution (Merck, US) diluted in PBS-Tween was then used to 'block' the particles for 1 hour. The enriched phage particles were were then added (1 μl of each single plasmid phage diluted to $1x10^{21}$ plaque forming units), an equal amount of pooled phage from panning round 3 was also added, each to separate tubes. The phage particles were incubated for 1 hour and then washed three times with PBS-Tween. A primary anti-M13 antibody (rabbit) (Merck, US) was prepared in the 2x caseine blocking solution (1:10000 dilution). The rest of the procedure was carried out exactly like Magnetite ELISA (see above).

2.10.3 Monobody - SDS-PAGE Based Nanoparticle Binding Assays

Due to potential inaccessibility of the poly-histidine tag for ELISA based probing in the monobody proteins (MB-E8, MB-CC) an SDS-PAGE based assay was developed (Chapter 6). 5 nanomoles of each protein to test was incubated with 70 μl of 30 mg/ml magnetite nanoparticle solutions (30-50 nm) that were blocked in 3 % skimmed milk solution made up in PBS-Tween for one hour in advance. The proteins were incubated at room temperature overnight. The particles were then washed three times with PBS-Tween. The particle solutions were then resuspended in 250 μl of PBS-Tween and were mixed 1:1 with 2X SDS-PAGE loading buffer and boiled for 10 minutes to release any bound protein. The particles were then allowed to settle and the supernatants were analysed on SDS-PAGE (tris glycine gels). Lane band intensity profiles were plotted to detect unique bands not found in samples that were only exposed to the 3 % milk blocking solutions.

An alternative assay, without protein rich 'blocking' was also developed. Proteins being studied were prepared in equal molar concentrations and were serially diluted

in PBS pH 7.4. 35 µl of the magnetite solution was then incubated with 250 µl of each protein dilution at room temperature for 5 hours and 30 minutes on a rotisserie mixer at room temperature. The protein bound particles were then magnetically collected and washed with PBS-T once. The washed particles were incubated with a 100 mM glycine solution at pH 2 for 5 minutes to elute the bound protein. The solution was then neutralised with 1 M tris pH 7.4. The particles were then again collected magnetically and the solution containing the protein was removed. The protein solution was mixed with and SDS-PAGE loading buffer and incubated at 95 degrees for 10 minutes. 9 µl of each sample was then applied to individual wells on 12 % tris-Glycine gels and run until full separation. The gels were then stained for an hour. The extent of particle binding for each protein was assessed by quantifying the intensities (with respect to the background) of the bands corresponding to the correct protein size. The quantification was performed using the volume tools option found in the Image Lab gel analysis software package.

2.10.4 Luminol Iron Binding Assay

A luminol based iron binding assay was adapted from Rawlings et.al., 2020 [192]. It relies on the fact that Fe^{2+} has been observed to catalyse the luminol reactions to produce chemiluminescence [221]. Here, 100 μl of assembled S-layer proteins at 500 μg/ml in separate tubes were mixed with 2 μl of a 1:1 mixture of Fe^{2+}:Fe^{3+} (50 mM total iron concentration) producing a protein suspension with a final 1 mM iron concentration. This was incubated for 1 hour in the dark at room temperature. The tubes were supplied with 500 μl of ultra pure water and were then centrifuged for 10 minutes at 12,500 rpm (bench-top centrifuge). This collected the insoluble S-layer sheets at the bottom of the tubes. The supernatant was removed and 300 μl of ultra pure water was used to re-suspend the S-layers followed by another centrifugation for 10 minutes at 12,500 rpm. This was repeated one more time (the resuspension and centrifugation acted as washing steps). Finally, the solution was removed and S-layer pellets were resuspended in 100 μl of 6 M urea, this was incubated for 30 minutes. The incubation with urea is meant to denature the proteins and release the bound iron into the solution. 40 μl of ECL Western Blotting Substrates (containing luminol) were distributed to individual wells in a 96 well plate (three wells for each

protein tube). The urea dissolved proteins were then added to the luminol containing wells, 20 μl. Readings were taken within 30 minutes (or until a signal could be detected). Plates were inserted in to the ChemiDocGel Imaging System and emission of light at 428 nm was detected and images were taken. The signal was quantified using volume tools option found in the Image Lab gel software.

2.10.5 Nitrocellulose Membrane Experiments

Nitrocellulose membrane binding experiments rely on the premise that protein adsorbed to the membrane can retain its function and, thus ability to bind nanoparticles or metal ions. Magnetite specific methodology can be found in early studies on Mms6 proteins [160]. Here, three types of assays were performed: iron ion binding, magnetite synthesis biotemplation and pre-formed magnetite binding assays (demonstrated for CoPt work with SPOT peptide arrays [193]). First, nitrocellulose membranes (0.2 μm) were spotted with 5 μl drops of 500 μg/ml protein suspensions and were allowed to dry for 1-2 minutes. The membranes were then blocked in 6 % skimmed milk (iron binding), 3 % skimmed milk (synthesis reaction) or 3 % BSA (pre-formed particle binding) made up in PBS-Tween with gentle agitation. All experiments were performed at room temperature. Three distinct versions of binding reactions were then performed (listed below). The membranes were visualised after washing while still damp. The results were analysed on the ChemiDocGel Imaging System, by measuring the spot intensity (corresponding to material deposition co-localised with the immobilised proteins). The spot intensity was normalised to proximate background intensity (averaged three spots around the target location). In the case of iron binding assay, the intensity of spots was not normalised due the background staining being higher than that of the spotted proteins. There, absolute values were used for analysis.

Binding Iron Ions

After a 60 minute incubation with a 6 % milk solution, the membrane was washed with ultra pure water and incubated with a solution of EDTA (0.1 M) for 60 minutes followed by a wash in ultra pure water. This step removes any metals present in initial blocking solution. The membranes were then placed in a 1:1 mixture of Fe^{2+}:Fe^{3+}

(50 mM total iron concentration) and incubated at room temperature in the dark for 16 hours. The membranes were then rinsed with ultra pure water and visualised.

Synthesis of magnetite on a membrane

An incubation of the spotted protein membrane with 70 ml of 1:1 mixture of Fe^{2+}:Fe^{3+} (50 mM total iron concentration) was set-up as was in the ion binding experiments (with the exception of blocking taking place in 3 % skimmed milk). After a 45 minute incubation with iron salts, 14 ml of NaOH (1M) was added. A black precipitate immediately formed. The reactions were left to proceed for 1 hour after which the membranes were transferred to a new container and were washed three times with ultra pure followed by visualisation.

Binding-Preformed Particles

Peptide spotted membranes were incubated with 3 % BSA for 1 hour. The membranes were then rinsed with ultra pure water and were immersed in 24 mg/ml of either a 10 nm or a 30-50 nm particle solution. The solution was left to incubate for 16 hours without agitation (swirling was found to cause the particles to coalesce around the middle of the container, whereas with the containers stationary the particles were found to be evenly distributed throughout the area of the whole membrane). The membranes were then gently rinsed with no direct stream of liquid on the membrane and visualised.

2.11 Microscopy

2.11.1 Fluorescence Microscopy

Fluorescent microscopy images were taken on a Nikon inverted Ti (dual camera) system using an oil immersion lens Plan Apo VC 100x oil. Excitation laser of 640 nm and emission filters of 660 nm were used to detect DyLight 650 (ThermoFisher Scientific, US) labelled protein assemblies. Samples were incubated at room temperature at 200 µg/ml concentration. 5 µl of sample was mixed with a 5 µl drop of Thermo Scientific Shandon Immu-Mount directly on a standard microscope slide onto which

a cover slip was then mounted. The cover slips were then fixed using nail varnish. Image analysis and processing was performed using Fiji [222].

2.11.2 Atomic Force Microscopy

Atomic Force Microscopy (AFM) is a versatile technique that allows for visualisation of topologies of samples in air and liquid environments at sub 10 nm scale [223]. The general layout of an AFM apparatus can be seen in Figure 2.4.

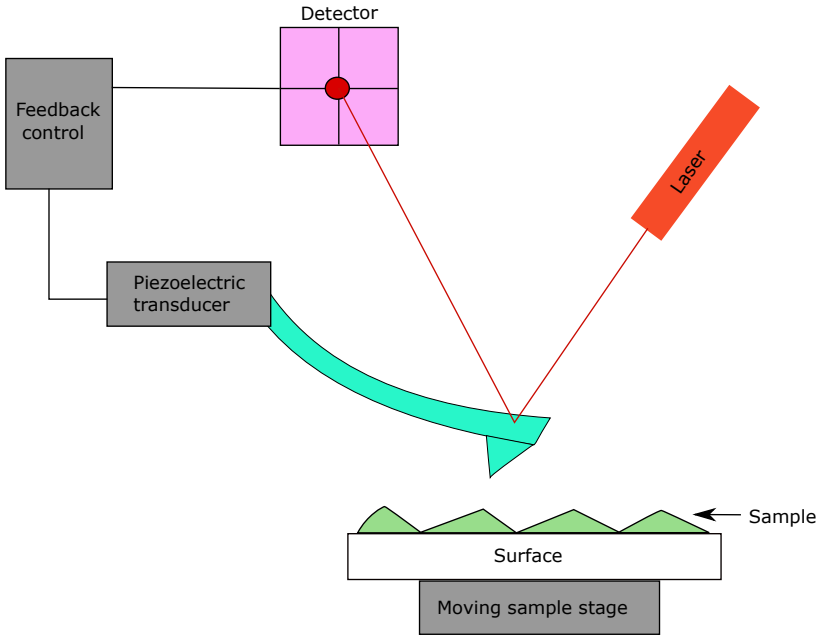


FIGURE 2.4: Component parts of an AFM apparatus.

The system directs a laser to the top of a flexible cantilever the reflection from which is detected by a diode sensor. Small cantilever bends can be detected. The technique relies on a sharp Silicon nitride probe placed on the cantilever. The probe can either interact with the surface continuously (contact mode) or intermittently (tapping mode). Throughout this project, tapping mode was used. In tapping mode the cantilever oscillates at its resonant frequency near the surface. The image is scanned by moving the cantilever in x and y directions along the sample by contracting and expanding the piezo electric component attached to the cantilever. If surface topology (z axis) changes during the movement along the x and y axes, the probe interacts differently with the surface (tip to surface interaction in tapping mode is usually attractive Van der Waals [223]) changing its amplitude, this is detected via a

change in the reflected angle of the laser. Since the mode 'aims' to keep the amplitude of the oscillation constant, the piezoelectric component adjusts the z axis. The recorded changes in the z axis of the piezoelectric component are used to generate the topology images.

In this project AFM was carried out using a Veeco Dimension 3000 with TESPA V2 tips (Bruker, DE) for air tapping mode and MLCT (Bruker, DE) cantilevers with 0.6 N/m spring constant for imaging in solution. For liquid AFM the cantilever holders were washed with dish soap, isopropanol and water (two times) followed by blow drying with nitrogen prior to any experiments. The images were analysed using Gwyddion open source software [224] and NanoScope Analysis tools (Bruker, DE). Image processing generally included levelling of the data by mean plane subtraction before analysis.

Sample preparations were performed on bare or coated mica surfaces. The bare mica was cleaved using sticky tape producing an atomically flat surface. Mica was either coated in poly-L-lysine [225] (PLL) (0.01 % low molecular weight solution, Merck, US) or poly-L-ornithine (PLO) (0.01 % low molecular weight solution, Merck, US)[226] that have been shown to impart positive charge on the surface. The positive charge was postulated to be instrumental in immobilising the SasG protein assemblies (S1 and S2 combination was expected to have a net negative charge) that were visualised in liquid or the sample preparations that were washed prior to drying. The exact coating protocol varied between samples, but the general procedure followed a fresh cleavage of the mica surface, application of 30 μl of PLO or PLL for a 5 minutes incubation at room temperature, three washes with 100 μl of HPLC grade water (VWR, US) and drying with nitrogen (the tap was equipped with a 0.2 μ filter). Differently prepared surface controls can be seen in Figure 2.5.

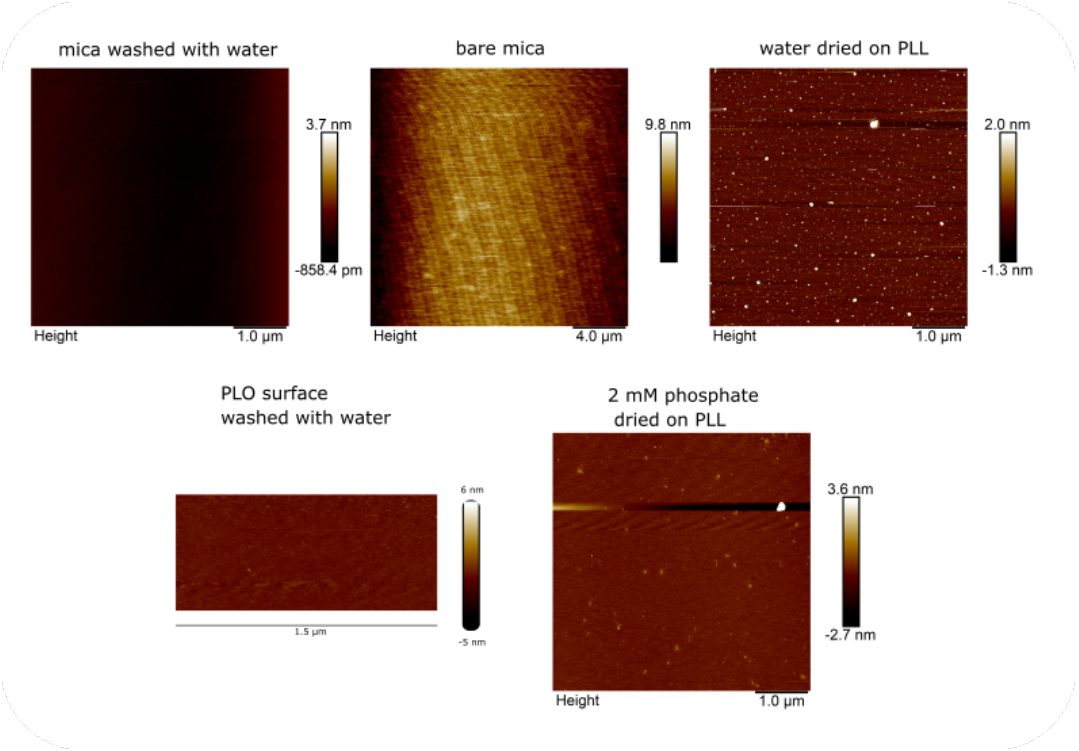


FIGURE 2.5: AFM controls for surfaces used throughout the projects visualised without exposure to proteins. Bare and water-washed mica surfaces as well as water-washed poly-l-ornithine (PLO) surfaces (coated mica) showed no contaminant particular matter. HPLC grate water or 2 mM phosphate buffer dried on poly-l-lysine (coated mica) showed a distribution of contaminant particles. None-of the negative controls showed formations such as were obtained with various protein incubations during the projects in Chapter3, Chapter4, Chapter5

SasG Fibril Preparations for AFM in Liquid

The images containing S1A3 and S2A3 fibrils seen in Chapter 3 were prepared as follows. The proteins were mixed in 1:1 ratio at the final protein concentration of 100 μg/ml and incubated at room temperature for 34 hours. The preparation was performed under sterile conditions. The surface was coated with PLO and the surfaces were incubated with 1 μg/ml of the protein assemblies for 3 hours (the samples were kept in closed containers to prevent evaporation). The sample preparation was carried out the same way for S1A3, S2A3 and G52-E-G53 controls. The images were taken in 20 mM phosphate buffer pH 7.4.

SasG Fibril Preparations for AFM in Air

For visualisation in air, the S1A3 and S2A3 assemblies were diluted to 100 ng/ml in 20 mM phosphate buffer pH 7.4 and applied to a mica surface coated with PLO. The incubation was carried out for 2 hours followed by three 100 μl washes with HLPC grade water and drying under nitrogen before visualisation.

SasG Amyloid-like Preparations for AFM in Air

The amyloid-like assemblies of S1x and S2 proteins were achieved by mixing the proteins in 100 ng/ml total concentration in 1:1 ratio and subjecting the samples to a thermal treatment. The thermal control was achieved using a thermocycler and followed a regimen of 95 °C for 2 minutes, 70 °C for 1 minute, 55 °C for 1 minute, 30 °C for 1 minute, 20 °C for 1 minute, 4 °C for 10 minutes in sequence. Serial dilutions of the proteins were then prepared using HPLC grade water. This meant that each dilution had decreasing amounts of buffer components. The samples at different concentrations were then applied to bare freshly cleaved mica surfaces as 50 μl drops. The drops were allowed to evaporate at room temperature before the samples were visualised.

AFM Sample Preparation for Chapter 4

The 'ring' structures were achieved by preparing PLO coated freshly cleaved mica surfaces. The surfaces were incubated with a fresh mixture of total protein concentration of 100 μg/ml HR-anchor with S2A3 (1:1 ratio) for two hours. A mixture of HR-add with S2A3 at the same ratio and concentration was then added at an equal volume (removing the previously applied suspension without washing) for a 22 hour incubation. This was washed three times with 100 μl of HPLC grade water and dried with nitrogen before visualisation. The control of S2A3 with HR-add and the Ank4 attempts were visualised after a 2 hour incubation followed by washing and drying.

S-layer AFM Preparations

The images with SgsE-EGFP in dry tapping mode were obtained by drying a 50 μl drop of 100 μg/ml assembled S-layer on a freshly cleaved bare mica surface. The images in liquid were taken by immobilising SgsE-EGFP assemblies on a PLL coated mica surface (incubation of 2 hours). The immobilisation was followed by three washes with 100 μl of HPLC grade water. The visualisation was performed in HPLC grade water.

2.11.3 Transmission Electron Microscopy

Transmission Electron Microscopy is a technique that allows for visualisation of matter at with 0.1 nm resolution [232]. A schematic of an electron microscope apparatus can be seen in Figure 2.6.

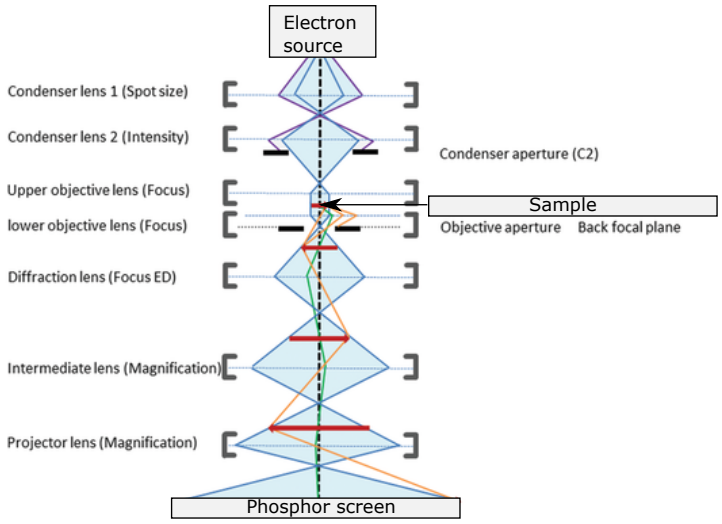


FIGURE 2.6: Component parts of an TEM apparatus adapted from Franken et.al.,2020 [232].

An electron microscope generally includes an electron source (either a field-emission gun or a heated tungsten or lanthanum hexaboride) to generate electrons, condenser lenses to increase the speed of the electrons and direct then down a vacuum tube, specimen holder, objective lenses and projector lenses (to collect the scattered electrons) and a phosphor screen with a camera to form an image and translate it into a digital format. The lenses are electromagnetic coils with positive charge. Vacuum is needed to prevent electron scattering in air. In this set-up electrons pass

through the sample and depending on electron density of the sample an image is reconstructed. Because electrons need to be able to pass through the sample, the samples have to be thin. Most biological samples lack the needed electron density, thus negative staining protocols have been developed [227]. Negative staining protocols use heavy metal materials such as uranyl formate to create an electron dense background that blocks the transmission of the passing electrons. First, biological material is applied to a carbon coated copper grid, then the negative stain is applied. After imaging, areas where biological material was not present remain dark.

In this project TEM was used for metallic nanoparticle and protein assembly visualisations. The particles were diluted to 0.01 mg/ml concentration in water and were ultrasonically treated. Then, 5 μl of a particle samples were applied to carbon coated copper grids for 1 minute and were then blotted with filter paper. For biological samples, carbon coated copper grids were first plasma treated to render the carbon surface hydrophilic. Samples were prepared within 1 hour after plasma treatment. Proteins at 100 μg/ml concentration were applied to the surfaces as 5 μl drops for 1 minute, the protein was then blotted and washed with 50 μl drops of distilled water twice (blotting in between) and then soaked in 1 % (w/v) of uranyl formate solution (20 μl), followed by blotting to dry and then another soaking in uranyl formate for 20 seconds and blotting to dry again. All grids were further dried using a vacuum pump. TEM images were obtained using a FEI Tecnai G2 Spirit electron microscope together with Gatan DigitalMicrograph software.

The particle size analysis was performed using Fiji software by measuring individual particles along the long axis. S-layer nanosheet Fourier Transform analysis was performed by identifying ordered regions and running a standard Fast Fourier Transform protocol (FFT). Threshold adjustment tools were then used to remove high frequency image information. An inverse FFT function was the performed and images with more clearly visible S-layer periodicity were obtained from which the lattice angle measurements were derived (measured manually).

2.12 Other Data Analysis and Data Visualisation

2.12.1 Statistical Tests

All statistical tests were performed using Prism GraphPad software (versions 7-9). Standard deviations were calculated using the standard deviation of the population formula. Significance value signs ns, *, **, *** represented 'not significant', $P \leq 0.05$, $P \leq 0.01$, $P \leq 0.001$, $P \leq 0.0001$, respectively. P values were obtained by performing t-tests (unpaired, nonparametric with Welch's correction [228]). Smoothing of circular dichroism spectra and thermal denaturation curves was accomplished using the Savitzky-Golay method [229].

2.12.2 Data Visualisation

Graphical plotting was performed using Prism GraphPad software (versions 7-9) and the Matplotlib Python plotting package [230].

2.12.3 Amino Acid colour schemes

The ClustalX colour scheme used in Chapter 4 can be decoded using Table 2.3.

Hydrophobic	BLUE
Positive charge	RED
Negative charge	MAGENTA
Polar	GREEN
Cysteines	PINK
Glycines	ORANGE
Prolines	YELLOW
Aromatic	CYAN
Unconserved	WHITE

TABLE 2.3: ClustalX colour scheme table.

Chapter 3

Results: SasG nanofibrils

3.1 Chapter Summary

Chapter 3 focuses on a fusion protein-based approach for creating magnetite functionalised nanowires. In abstract, the idea was to create modular self-assembling fibres capable of binding magnetic nanoparticles which could later be immobilised on solid surfaces (Figure 3.1). The modular components were intended to act as spacers for templating magnetite placement on surfaces. The approach followed choosing a mechanically strong small rod-like protein SasG to act as a spacer, two sets of orthogonal coiled-coils to join the rod-like protein together and a magnetite binding motif to provide functionality for the resulting assemblies. This results in an obligate two sub-unit assembly. Choosing a two component system introduces flexibility in tuning the spacing for magnetite placement for future patterning applications.

The chapter starts with a description of the design of coiled-coil and SasG protein fusions. The results cover the biophysical and structural characterisation of the SasG native G5-E-G5 domain and coiled-coil adapted constructs. This is followed by a multi-disciplinary characterisation of interactions between the two sub-units and direct visualisation of their assembly. The suitability of this system for inorganic magnetic nanoparticle patterning is also investigated through magnetite nanoparticle binding assays. Unexpected behaviours of the system are also evaluated.

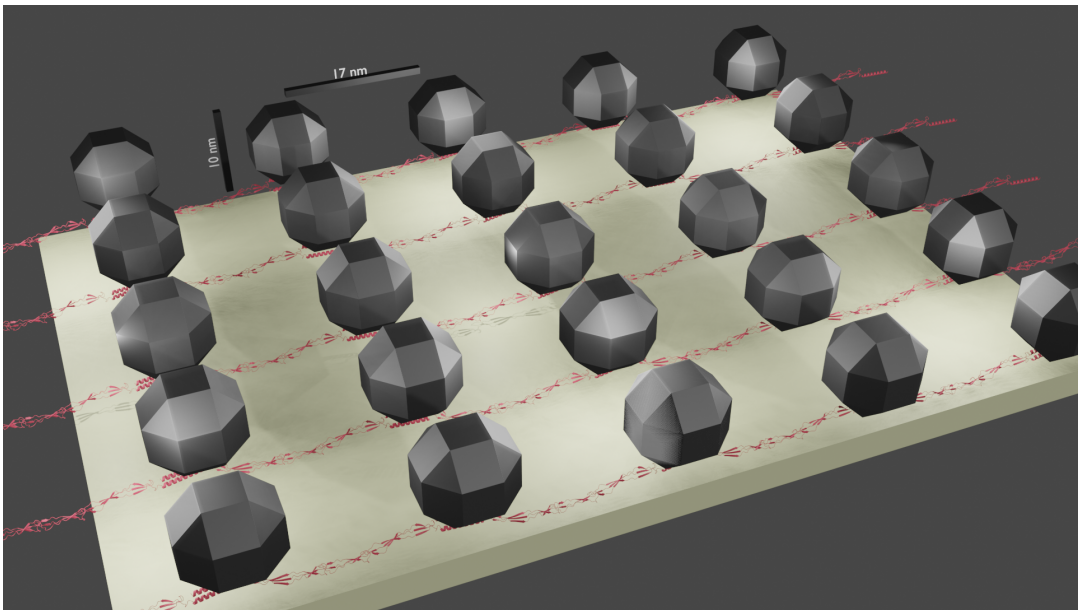


FIGURE 3.1: A three-dimensional rendering of the proposed SasG G5-E-G5 domain assembly. The dark metallic shapes represent 10 nm sized magnetite nanoparticles. The red structures underneath - predicted assembly behaviour of the G5-E-G5 domain with coiled-coil linkages. The objects are to scale (within 1-2 nm). Scale bars: horizontal - 17 nm, vertical - 10 nm.

3.2 Introduction: Self-assembly Components

3.2.1 Two sets of *de novo* Designed Orthogonal Coiled-coils

Two pairs of coiled-coil heterodimer forming α-helices, termed P5/P6 [248] and AN4/BN4 [249], were selected as suitable non-covalent linkage for constructing SasG G5-E-G5 nanowires due to their demonstrated high stability and specificity (each coil only forms with its designated α-helical partner). Both sequence pairs are shown in Figure 3.2.

A

```
                gabcdef  gabcdef   gabcdef  gabcdef
P5 SPED ENAALEE KIAQLKQ KNAALKE EIQALEY G
P6 SPED KNAALKE EIQALEE ENQALEE KIAQLKY G
```

B

```
                gabcdef gabcdef  gabcdef   gabcdef
CC-AN4 G EIAALEQ EIAALEK ENAALEW  EIAALEQ GG
CC-BN4 G KIAALKQ KIAALKY KNAALKK KIAALKQ GG
```

FIGURE 3.2: **A** α-helix forming amino acid sequences termed P5/P6 designed and demonstrated to form a coiled-coil [248]. **B** α-helix forming amino acid sequences termed AN4/BN4 designed and demonstrated to form a coiled-coil [249].

Both pairs are comprised of four heptad repeats and rely on electrostatic interactions between opposing ***e*** and ***g*** positions to form parallel (with respect to the peptide chain orientation) specific coiled-coils (for an explanation of coiled-coil terminology refer to section 1.5.5 in Chapter 1). The P5/P6 pair exhibits a more complicated hydrophobic core residue pattern of alternating asparagine and isoleucine residues. Because coinciding asparagine and isoleucine residues are highly energetically unfavourable this further ensures dimer specificity and correct parallel orientation (only if both peptides exhibit this pattern are they be able to form a coiled-coil). The study also positioned glutamine and alanine residues in the ***b***, ***c*** and ***f*** positions for their contribution to helical propensity. The C-terminal ***f*** position contains a tyrosine residue for better UV absorbance. Both peptides were capped with serine-proline-glutamate-aspartate sequence at C-terminus acting as a helix breaker (useful as linkers for fusion proteins) and a glycine residue at N-terminus. AN4 and BN4 heterodimeric coiled-coil pair in contrast has leucine and isoleucine residues in all ***a*** and ***b*** positions, respectively, except for asparagine at ***a*** of the third heptad which confers parallel orientation preference onto the coiled-coil. The AN4 and BN4 have either a tryptophan or a tyrosine in one of the ***f*** positions for solubility and UV tracking during purification and experimentation. Both termini were capped with glycine residues. Both pairs of sequences were biophysically characterised. The P5/P6 pair has a 40 °C melting temperature, whereas AN4/BN4 transitioned at 81 °C as determined by CD thermal denaturation experiments. The differences in melting temperatures for the two pairs, introduce a potential differential control trigger - one could determined which coiled-coil associations are allowed by changing the temperature. All peptides were completely disordered in isolation at room temperature (determined using CD spectra) apart from BN4. BN4 was seen to dimerise with itself (analytical ultra centrifugation experiments), but the homodimer was much less stable (transition temperature of 25 °C compared 40 ° of AN4/BN4 heterodimer). AN4/BN4 pair was also compared to AN3/BN3 (three heptad coiled-coil) and the former showed a dissociation constant 600 times lower than the latter, meaning that vast stability gains can be made by extending the coiled-coil by just one heptad repeat. Experiments for both coiled-coils were performed at physiological pH (7.0-7.4).

Both coiled-coil pairs have seen follow-up studies investigating their biophysical properties and applicability in 'bottom-up' synthetic biology. P5/P6 pair has been used to construct self-assembling polyhedral structures (pyramids) [233]. The self-assembling system involved 6 parallel and anti-parallel coiled-coil pairs demonstrating the specificity of the interactions in a crowded molecular environment. AN4/BN4 coiled-coil pair has been modified each with a histidine at ***b*** and ***f*** positions at terminal heptads providing Ni^{2+} coordination sites [234]. This resulted in a 10 pN increase in mechanical strength in presence of Ni^{2+} compared to no Ni^{2+} in AFM pulling experiments. A molecular dynamics and AFM pulling study showed that there is a mechanical strength increase when comparing AN4/BN4 to AN3/BN3 [235]. This corroborates the previous thermal denaturation result demonstrating that melting temperature improvements can stand in for mechanical strength. This is useful in protein engineering as determination of thermal stability is much less labour intensive experimentally than AFM pulling (also corroborated here [236]). Lastly, strand displacement showed that any homodimers formed by BN4 get quickly out-competed when exposed to the binding partner AN4 [237].

3.2.2 SasG Coiled-coil Fusion Protein Construction

SasG G52-E-G53 structure was selected for this project as it is the most similar to the rest of repeats (SasG G5 and E domains have subtle amino acid sequence differences depending on the repeat) and has a fully solved X-ray crystal structure (PDB accession code: 4WVE [107]). The general picture of a two component self assembly system can be seen in Figure 3.3.

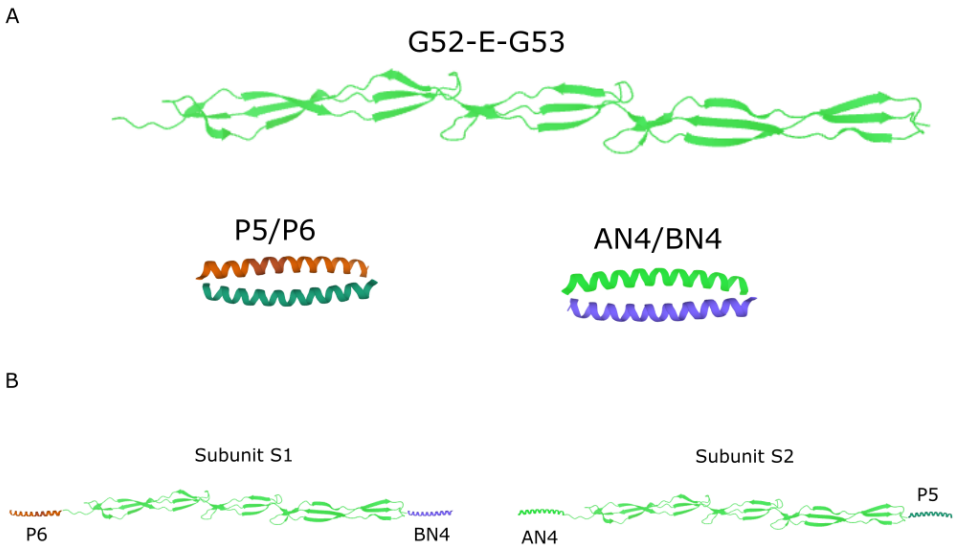


FIGURE 3.3: **A** A cartoon representation of component parts used in Chapter 3 including the G52-E-G53 (PDB:4WVE) and two coiled-coil pairs P5/P6 and AN4/BN4. **B** A proposed two component system for extended nanowire assembly.

By creating two DNA constructs encoding for P5-G52-E-G53-BN4 and AN4-G52-E-G53-P5 one would expect to produce two distinct proteins (denoted S1 and S2) that would assemble in a pattern of [-S1-S2-]n in chains of indefinite length. Sub-units can also be functionalised with binding moieties at their termini. An example displaying sequences of one version of the sub-unit pair used in this chapter is shown in Figure 3.4.

S1A3 (P6-GEG-BN4)
MGSSHHHHHHSSGLVPRGSHMASSPEDKNAALKEEIQALEEENQALEEKIAQLKYGGKYGPVKGDSIVEKEEIPFEKERKFNPD
LAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETITPGHRDEFDPKLPTGEKEEVPGKPGIKNPET
GDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKEEITKDPINE
LTEYGPETGGKIAALKQKIAALKYKNAALKKKIAALKQGGGSGS-HNHKSKKHK
S2A3 (AN4-GEG-P5)
MGSSHHHHHHSSGLVPRGSHMASGGEIAALEQEIAALEKENAALEWEIAALEQGGKYGPVKGDSIVEKEEIPFEKERKFNPDLAP
GTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETITPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGD
VVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKEEITKDPINELT
EYGPETGGSPEDENAALEEKIAQLKQKNAALKEEIQALEYGGSGS-HNHKSKKHK

FIGURE 3.4: Top: Amino acid sequence for the S1A3 construct. Bottom: Amino acid sequence for the S2A3 construct.

The figure shows two amino acid sequences with an N-terminal poly-histidine purification tag followed by a thrombin cleavage site (LVPRG). Both sequences contain the G52-E-G53 as the 'backbone' of the structure (tinted grey). The coiled-coil

forming peptides P6, BN4, AN4 and P5 are tinted yellow, red, green and blue, respectively. two glycine residue linkers are used at each intersection with the G52-E-G53 protein. At C-terminus is the magnetite nanoparticle binding sequence A3 separated by a mixed glycine/serine linker (tinted brown). Thus, the two constructs are termed S1A3 and S2A3. Throughout the chapter different alterations of this sequence will be used and will be introduced where appropriate. All of the amino acid sequences can be found in Table A.1.

3.2.3 Conclusion to the Introduction

In summary, this two sub-unit system should assemble in a specific step-wise manner creating extended long nanofibres. Placing a binding peptide at C-terminus would hopefully allow for discreet length functionalisation of the fibres. Once functionalised, they may facilitate achieving the ultimate goal of constructing self-assembling nanoparticle patterns on surfaces for bit pattern media (Figure 3.5).

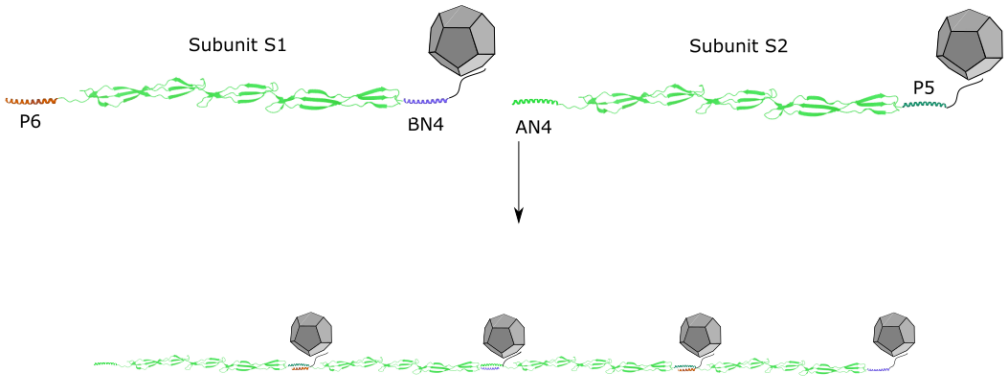


FIGURE 3.5: Expected assemblies from the two component system. Shapes in grey represent magnetite nanoparticles. Particles and protein not represented to scale.

3.3 Production and Characterisation of the Constructs

3.3.1 DNA Constructs

All SasG constructs (G52-E-G53, S1A3, S2A3, S1x, S2x, S2, cysS1A3 and cysS2A3) were obtained by using restriction enzymes to clone the sequences into pET28a at *Nhe*I and *BamH*I sites to obtain final expression constructs containing N-terminal poly-histidine (six repeat) tags. DNA sequences for expression of AN4, BN4, P5

and P6 separately from the G52-E-G53 backbone were subcloned into pET24a downstream of a SUMO cleavage sequence (preceded by the SUMO protein itself). The translated protein open reading frames used throughout chapter are summarised in Table 3.1.

Construct	Exp. Mass (no Met)	pI	Abs 0.1%	Description
G52-E-G53	25722.76	5.51	0.232	SasG repeat unit 
S1A3	34181.51	6.86	0.261	sub-unit S1 with A3 binding peptide at C-terminus 
S2A3	34211.10	5.29	0.379	sub-unit S2 with A3 binding peptide at C-terminus 
S1x	32653.85	6.02	0.274	sub-unit S1 with no magnetite binding peptide 
S2x	32813.60	4.96	0.395	sub-unit S2 with no magnetite binding peptide 
S2	33924.62	5.10	0.382	sub-unit S2 with a magnetite binding peptide QRAQSVSKK 

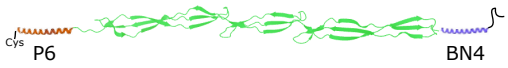
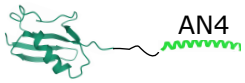
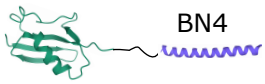
cysS1A3	34197.58	6.93	0.261	Sub-unit S1 with A3 binding peptide at C-terminus and a cysteine residue near the N-terminus 
cysS2A3	34227.16	5.29	0.378	Sub-unit S2 with A3 binding peptide at C-terminus and a cysteine residue near the N-terminus 
SUMO-AN4	16185.89	4.96	0.432	SUMO protein fusion to AN4 α-helix 
SUMO-BN4	16098.27	9.05	0.185	SUMO protein fusion to BN4 α-helix 
SUMO-P5	16503.27	5.24	0.181	SUMO protein fusion to P5 α-helix 
SUMO-P6	16561.31	5.15	0.180	SUMO protein fusion to P6 α-helix 

TABLE 3.1: Constructs used in Chapter 3. Characteristics were calculated by translating the DNA sequences into amino acid sequences and performing protein parameter analysis (ProtParam - ExPASy).

3.3.2 G52-E-G53

Before any assessment of the coiled-coil two sub-unit system could be performed, the G52-E-G53 'backbone' needed to be characterised for its biophysical properties.

A plasmid construct was created to produce G52-E-G53 with an N-terminal six histidine purification tag cleavable by thrombin protease (Table A.1 for sequence information). The protein was produced using the auto-induction growth medium protocol (see methods) and purified on a Ni-NTA followed by a gel filtration step and eluted in 20 mM phosphate buffer pH 7.4. Typical purification results visualised using SDS-PAGE are shown in Figure 3.6A. Based on SDS-PAGE, the molecular weight appears to be 39 kDa. Furthermore, size exclusion chromatography shows a major peak corresponding to a 72 kDa protein (SEC calibrated for globular proteins). This might appear puzzling as the theoretical size estimations of the translated DNA sequence indicate a 25853.84 Da protein. The same protein preparation analysed using ESI-TOF mass spectrometry returned a single peak at 25723 Da (Figure 3.6C). This is an almost perfect match for the protein with N-terminal methionine removed (25722.76 Da). This confirms that the correct construct was successfully produced. Aberrant SDS-PAGE mobility of SasG constructs has been shown before, where purified full-length SasG protein ran as larger bands up to 40 kDa than expected [238], this might be due to the E domain that is enriched in disorder promoting residues. Domains with such amino acid composition have been shown before to change SDS-PAGE mobility of the whole protein [239]. The perceived larger size on size exclusion chromatography is most likely arising from the fact that the column calibration was performed with globular protein standards, as G52-E-G53 is an extended rod-like protein it might not fit into the small pores that other compact globules of similar molecular weight would. Because of this, G52-E-G53 would elute earlier than expected (the relationship between folded protein shape and size exclusion chromatography mobility is discussed in Chakrabarti et. al., 2018 [240]).

The quality of the purified G52-E-G53 protein was assessed using circular dichroism to check for correctly assumed secondary structure (Figure 3.7A).

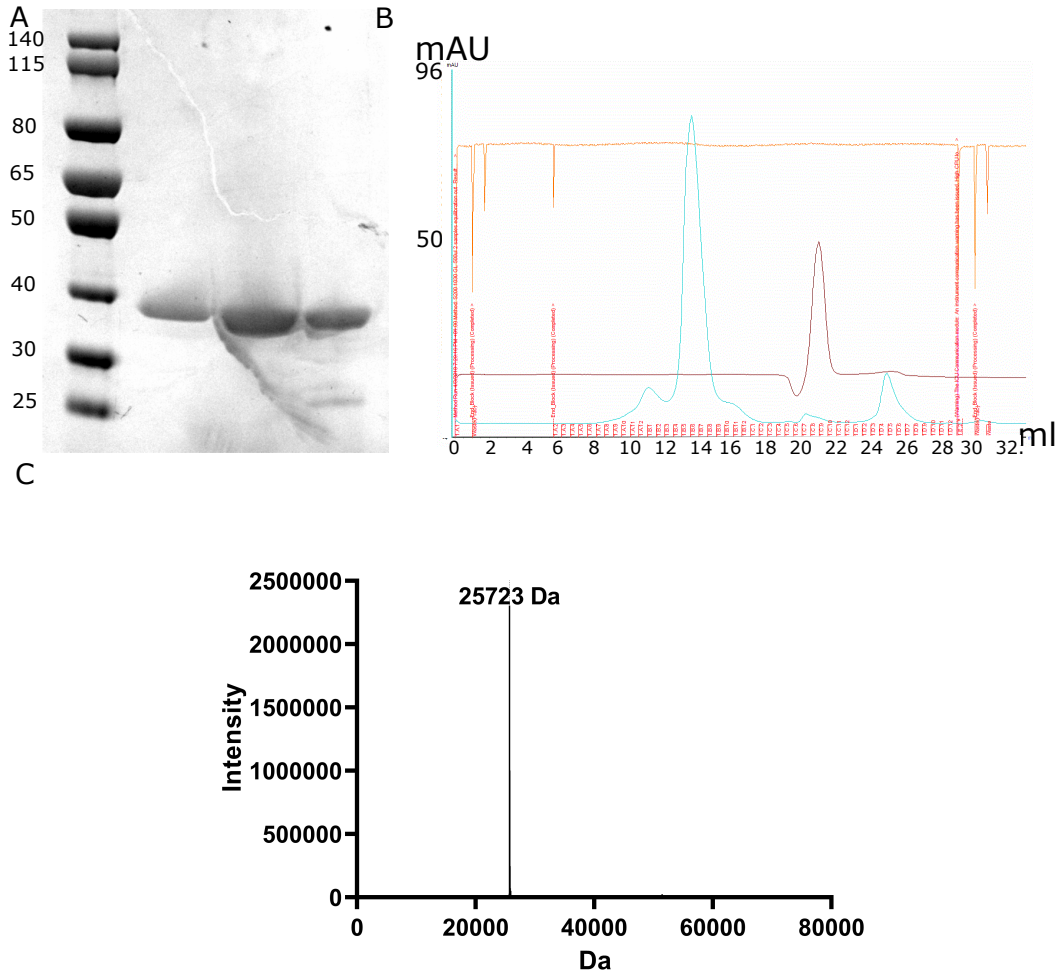


FIGURE 3.6: **A** SDS-PAGE of G52-E-G53 after size exclusion chromatography purification. G52-E-G53 can be seen running at around 40 kDa based on the molecular weight marker (PageRuler in MOPS). **B** Size exclusion chromatography (SP200 column) UV trace (mAU) of G52-E-G53 eluted in 20 mM phosphate buffer pH 7.4. **C** ESI-TOF mass spectrum of G52-E-G53 purified via size exclusion chromatography.

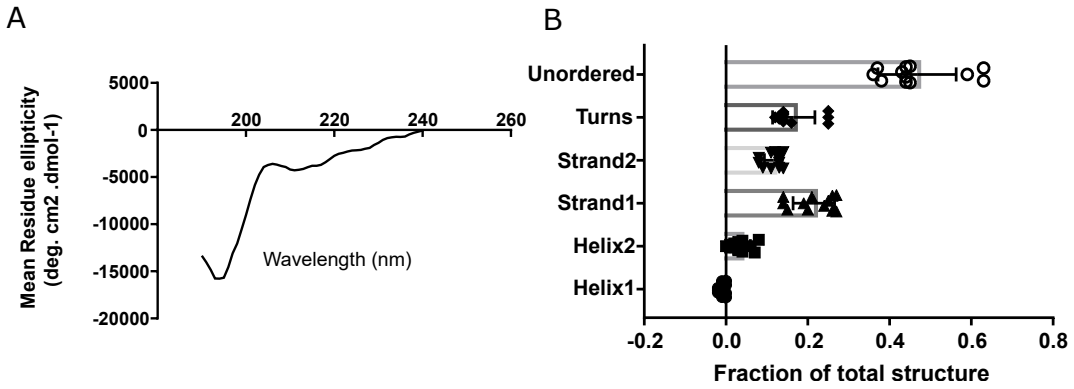


FIGURE 3.7: **A** G52-E-G53 CD spectrum at room temperature in 20 mM phosphate buffer pH 7.4. Plotted mean values from three replicates. Samples analysed at 2.5 μM. **B** Dichroweb analysis using CDSSRT algorithm. Individual points represent pooled three repeats and replicate analyses using data sets 4, 7, SP175 and SP180 (n=12). Error bars represent standard deviation (population).

The spectrum shows a negative peak at 218 nm corresponding to β-sheet content present and a 195 nm negative peak that has been shown to correspond collagen-like triple helix structure [241] that is known to be present in G5 and E domains [107]. The obtained spectrum was reproducible throughout the project and it matched closely with the published spectrum of an E-G5 construct [105] which should have approximately the same structural composition. The literature values indicate 39 % β-sheet, 32 % random-coil, 24 % β-turn and 4 % helix which is close to the values obtained here: 34 % β-sheet, 47 % random-coil, 17 % β-turn and 4 % helix. A 5 % lower total β-sheet content would be expected from this experiment due the fact that the G52-E-G53 also contained a purification tag, which is most likely randomly coiled (effectively reducing the β-sheet content percentage). Additionally, use of circular dichroism instrumentation capable of producing spectra beyond 190 nm could provide more information for the algorithms to determine secondary structure composition more accurately [205]. Nevertheless, the analysis here provides sufficient evidence that G52-E-G53 assumes the expected secondary structure. G52-E-G53 with the poly-histidine tag cleaved and un-cleaved was compared in Figure 3.8A.

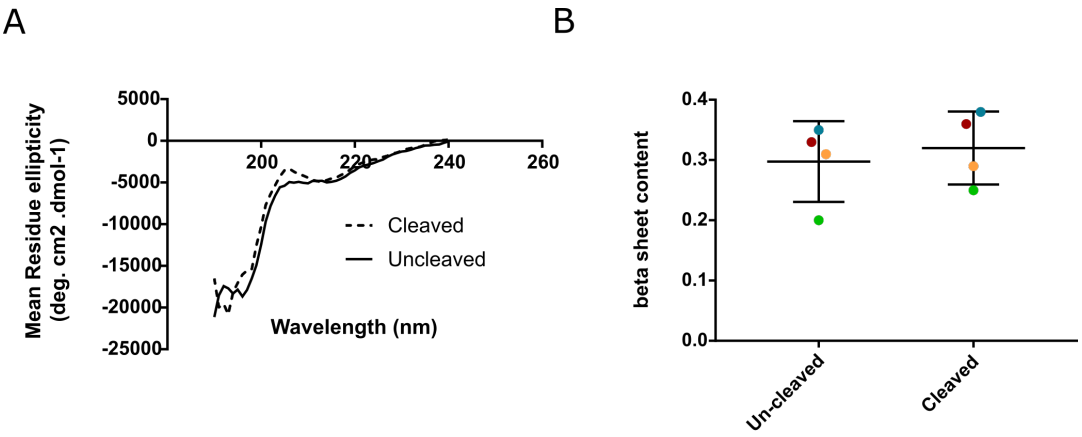


FIGURE 3.8: **A** G52-E-G53 CD spectra of poly histidine tag cleaved (dashed line) and un-cleaved protein preparations (solid line). Both spectra were obtained in 20 mM phosphate buffer pH 7.4 using 2.5 μM of G52-E-G53. **B** Dichroweb CDSSTR analysis of the spectra in **A**. Different colours represent β-sheet content values obtained using different reference sets: yellow - reference set 4, green - reference set 7, blue - reference set SP175 and red - reference set SP180.n=4

Both spectra showed the 218 nm and 195 nm negative peaks. Dichroweb fitting analysis in Figure 3.8B showed close to the same estimated β-sheet composition with a slight bias towards higher values for the cleaved preparations (differences are not statistically significant). This shows that the protein assumes the expected secondary structure regardless whether or not the tag is cleaved. It is also apparent that CDSSTR fitting sometimes gives lower values when using certain reference sets. Reference set 7 seems to have a bias for lower β-sheet content in this experiment. Since this result showed no significant difference between cleaved and un-cleaved preparations of G52-E-G53, all subsequent characterisation is performed on G52-E-G53 containing an N-terminal poly-histidine tag.

3.3.3 G52-E-G53 Thermal and pH Stability

The purified G52-E-G53 protein was tested for its folding state at different pH conditions (Figure 3.9A).

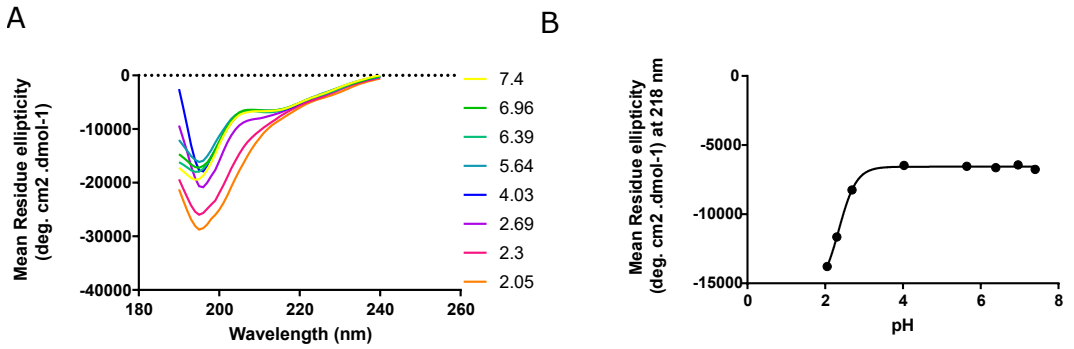


FIGURE 3.9: **A** G52-E-G53 CD spectra at different pH environments. **B** Tracked changes at 218 nm band over a range of different pH conditions at protein concentration of 2.5 μM. The pH environments were obtained by adding concentrated HCl into 20 mM phosphate buffer (starting pH 7.4) and re-tracing the pH values using a calibration curve.

In the range between pH 7.4 and 2.69 (not inclusive) the G52-E-G53 secondary structure seems to be maintained. However, at the range of pH 2.69 and 2.05 the structure starts to unravel. A strong change towards 195 nm indicates appearance of random coil structure (195 nm band is an indicator for collagen triple helix and random coils, thus the mere presence of a negative peak is not sufficient and one must look at the relative intensity compared between difference conditions). Tracking the changes at 218 nm (Figure 3.9**B**) caused by the widening of the negative 195 nm peak one can infer a transition at pH 2.5. This is a rather broad range of tolerance for pH fluctuations and it might be useful if G52-E-G53 is to be used as a static (unchanging) component in pH based molecular switching systems. The thermal stability of the protein was also assessed and CD melting curves tracked at 222 nm and 218 nm can be seen in Figure 3.10.

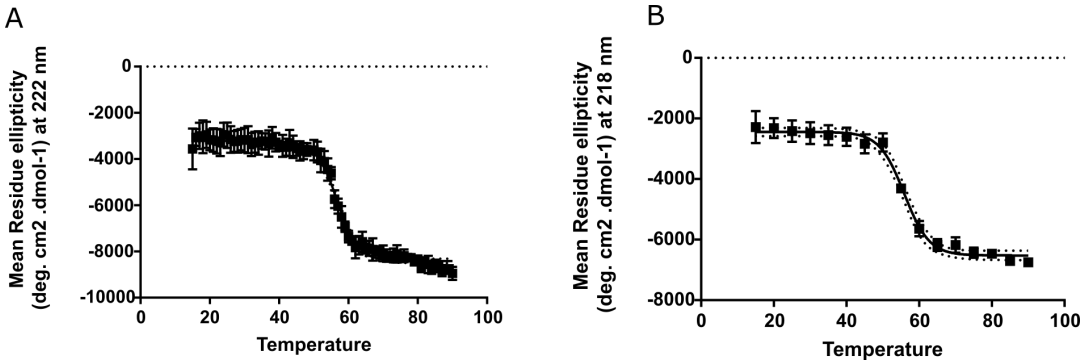


FIGURE 3.10: **A** G52-E-G53 CD melting curve tracked at the 222 nm band. **B** G52-E-G53 CD melting curve tracked at the 218 nm band. The experiment was performed in 20 mM phosphate buffer pH 7.4 in triplicate. Proteins analysed at 2.5 μM. Error bars are standard deviations (population). Solid lines indicate a non-linear fit (included for easier visualisation of the trends).

A transition of 56.6 °C (+/-0.6) can be extracted from either of the graphs. Tracking at 222 nm was performed as coiled-coil containing constructs are investigated later in Figure 3.19, where familiarity with this melt profile is necessary for interpretation of the data. The melt profiles are corroborated by Dichroweb (CDSSTR) fitting analysis on CD spectra taken at a range of temperatures (Figure 3.11).

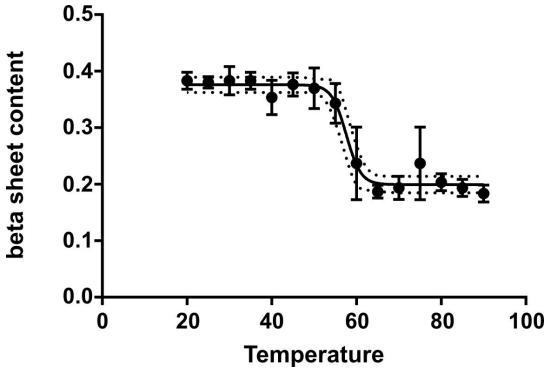


FIGURE 3.11: G52-E-G53 spectra in 20 mM phosphate buffer pH 7.4 at different temperatures analysed using Dichroweb CDSSTR algorithm with reference set SP175. Experiments were performed in triplicate. Proteins analysed at 2.5 μM. Error bars represent standard deviations (population). Solid line indicates a non-linear fit.

The thermal denaturation curves in Figures 3.10 and 3.11 show that G52-E-G53 is in a folded state with the transition to an unfolded state with a midpoint at 56 °C. If the expressed protein was not folded such transition would not be observed. Denaturation curves in Figure 3.10 tracking circular dichroism signal at 222 nm and 218 nm seem to indicate that the protein is gaining secondary structure with increasing temperature (based on what would be expected for fully α-helical or β-sheet protein [242]), however as Figure 3.7 indicates, G52-E-G53 resembles collagen-like triple helix structure for which more positive circular dichroism values are expected at the band range between 210-222 nm when the protein is at a folded state (i.e. tracking CD signal during thermal denaturation at this range would see a decrease).

Reading a circular dichroism spectrum at room temperature after running the thermal denaturation experiments returns the G52-E-G53 protein to the same structural composition (Figure A.1). Fitting the data using Dichroweb (CDSSTR) reveals unchanged levels of β-sheet content displaying the reversibility of denaturation.

Taken together, with the previous section, these data show G52-E-G53 to be a stable β-sheet containing protein at a broad range of pH values that can also spontaneously refold after thermal denaturation (transition temperature 56.6 °C (+/-0.6)).

3.3.4 Assessment of AN4/BN4 and P5/P6 Orthogonality

The two coiled-coil pairs AN4/BN4 and P5/P6 have not been tested for their orthogonality in the literature before. One can try to predict the outcome of mixing the two pairs based on the first principles of coiled-coil assembly discussed in Chapter 1. Figure 3.12 juxtaposes peptides from the two sets in all possible combinations (other than the designed expected ones).

A parallel P5/AN4 coiled-coil is unlikely to form as two C- and N-terminal heptads place glutamate residues at corresponding ***e*** and ***a*** positions. This would destabilise the dimer as glutamate negative charges would be clashing. Furthermore, there would be mismatched asparagine/isoleucine pair at ***a*** position of the first heptad

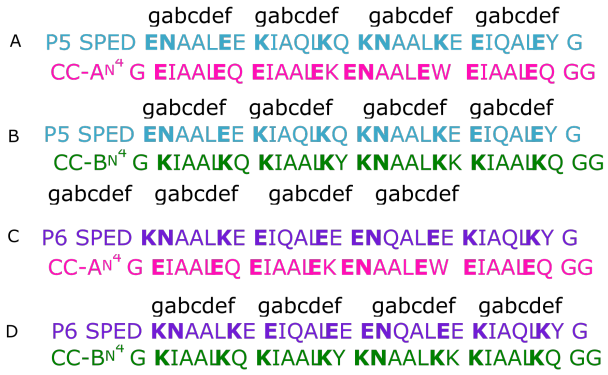


FIGURE 3.12: Juxtaposition of possible α-helix combinations: **A** P5/AN4. **B** P5/BN4. **C** P6/AN4. **D** P6/BN4.

pair. An anti-parallel arrangement (not shown) would result in same charge electrostatic clashes and even more mismatched asparagine/isoleucine residue pairs. Similar problems would arise with a P5/BN4 pair, this time with lysine residues clashing in the middle two heptads. P6/AN4 and P6/BN4 pairs, again, would have electrostatic clashes at either central (glutamate) or terminal heptads (lysine), respectively, and asparagine/isoleucine mismatches. Considering, multiple examples in literature, where such peptide chain incompatibilities result in disordered monomers as opposed to coiled-coil heterodimers [120, 248], it might be safe to conclude that these two coiled-coil pairs should be orthogonal to each other and not form unexpected dimers when mixed. This would only be expected to be the case at physiological pH where glutamate and lysine residues are both charged. pH has been shown to change the behaviour of coiled-coil assemblies depending on amino acid composition [243, 244, 245, 246]. In order to test orthogonality experimentally, a strategy for expressing individual coiled-coil heptad repeats was devised and is illustrated in Figure 3.13.

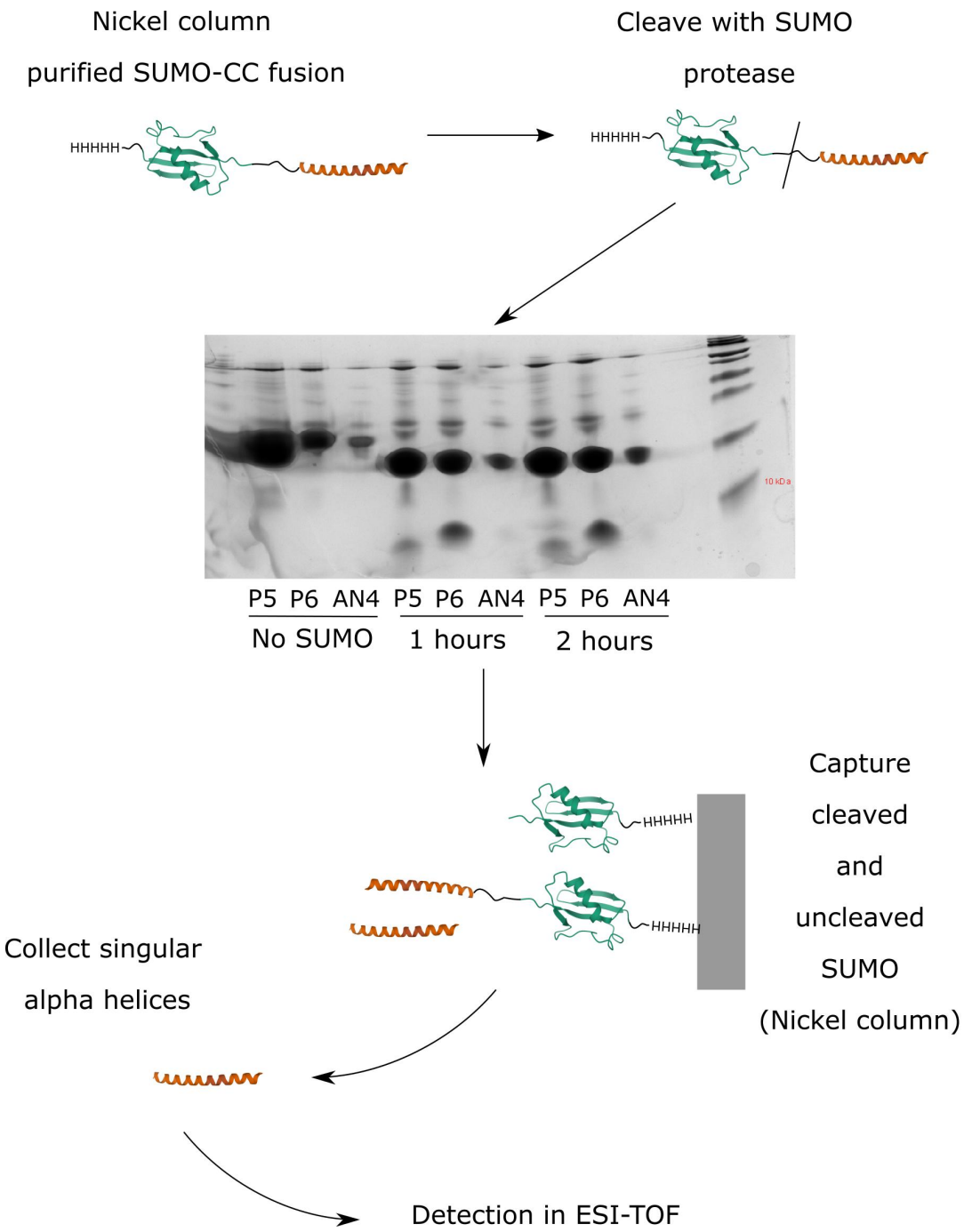


FIGURE 3.13: A diagram displaying individual alpha-helix expression/purification strategy. SUMO-CC stands for SUMO protein with either of the α-helices (P5, P6, An4 or BN4). Middle: SDS-PAGE Tris-Tricine gel of SUMO-cc fraction digested for different amounts of time. Structural models are not to scale.

The heptad repeat encoding DNA (for P5, P6, An4 and BN4) was cloned into a plasmid containing a six histidine followed by Smt3 gene encoding a soluble expression tag SUMO-1. The heptad repeats were placed downstream of the sequence separated by SUMO protease recognition site (glycine-glycine). The rationale was

to attach a relatively large, soluble, UV-absorbent protein for trace-ability during expression and purification stages of the individual α-helices. After a first round of affinity purification, the six histidine SUMO-1 were cleaved with a SUMO protease. Tris-Tricine SDS-PAGE gels were run to assess purification outcomes (Tris tricine gels allow for better separation at low molecular weight ranges <15 kDa). The mixed protein populations (His-SUMO-protease, His-SUMO-1, His-SUMO-1-CC, and CC) were then clarified by running them through a nickel affinity column. The flow through was collected, and was analysed on ESI-TOF. Only the heptad-repeats were expected to be present in the flow-through as all other components would have been immobilised via the six-histidine tags. The obtained ESI-TOF results can be seen in Table 3.2.

Construct	Expected Mass	ESI-TOF mass
AN4	3297.57	MS-failed
BN4	3354.09	3354.09
P5	3759.10	3758.9195 (with impurities)
P6	3817.14	2803.4363

TABLE 3.2: Cleaved heptad repeat ESI-TOF mass spectrometry results (Mass in Da).

BN4 and P5 mass spectra showed detectable peaks for exact expected size for a cleaved heptad repeat. Unfortunately, due to the fact that the purifications were performed a week apart P5 and P6 heptad repeats were frozen and thawed before ESI-TOF was performed. P6 spectrum does not show the correct 3817.14 Da molecular weight, but rather a 2803.4363 fragment that corresponds to a truncated sequence of KEEIQALEEENQALEEKIAQLKYG. While P5 spectrum showed the correct peak it

was only 28 % of the total population. The other fragments of 1601.74 Da and 2597.34 Da seem to fit sizes of truncated sequences SSPEDENAALEEKIA and SSPEDENA ALEEKIAQLKQKNAAL, respectively. It seems that the freeze thaw cycle might have caused the degradation of the peptides into smaller fragments. AN4 showed low expression (undetected on the gel) and was not detected on ESI-TOF. Due to time constraints (causing inability to repeat the purifications) CD experiments were nevertheless attempted (Figure 3.14).

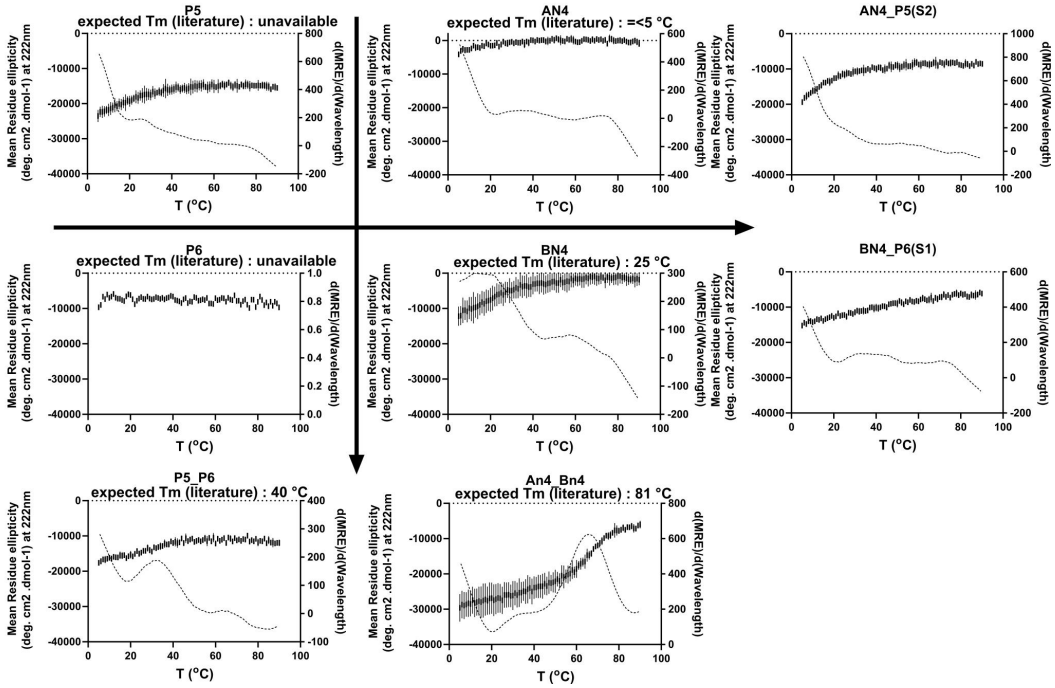


FIGURE 3.14: CD thermal denaturation curves tracking CD signal at 222 nm to check for orthogonality between the two sets of coiled-coil pairs. Melts performed in triplicate. Error bars represent standard deviation (population). Dashed line show first order derivatives for transition temperature estimations.

To assess if the poor quality purification outcomes impair the ability of the heptad repeats to form heterodimeric coiled-coils P5/P6 and AN4/BN4 were mixed and mean residue ellipticity was tracked at 222 nm. P5/P6 pair produced a novel melting profile (compared to individual P5 and P6 profiles) with a positively shifted transition temperature at 36 °C. Same was observed for the AN4 and BN4 coiled-coil heterodimer. A novel profile with a 68 °C transition was produced. Both 36 °C and 68 °C values fall slightly short of the expect literature 40 °C and 81 °C values that are expected for these coiled-coils. This is most likely due to the fact that the peptide mixtures are either at very low concentration (AN4) or are fragmented (P5/P6).

Nevertheless, this still shows that the peptides are present and functional (only to a lesser degree). Peptide mixtures of AN4/P5 and BN4/P6 were thus tested to see if they can form undesired heterodimers. The particular mixtures were chosen as they occur in S1 and S2 respectively, and would cause creation of homo-polymer assemblies of the sub-units should they prove to interact. The AN4/P5 mixture did not seem to produce a novel transition temperature and looks like an average of the two melting profiles suggesting a lack of interaction. BN4/P6 is more difficult to interpret, but it does not seem to produce a profile with a clear transition temperature. These results imply that, AN4/BN4 and P5/P6 may be sufficiently orthogonal for usage in specific assembly designs.

3.3.5 S1 and S2 Construct Comparison

As discussed above, in order to join G52-E-G53 protein rods into extended structures, the proteins were modified with coiled-coils. Multiple variants have been constructed for different purposes, their sequences and descriptions can be accessed at Table A.1. The constructs are also graphically depicted in Figure 3.15. The initial assumption was that the modifications would not change the expected behaviour of the two sub-unit system, namely that sub-unit S1 will interact with sub-unit S2 via coiled-coil pair BN4/AN4 at one terminus, and that they will interact via P5/P6 coiled-coil pair at the other. As the project evolved, different amino acid sequence modifications have been applied to the sub-units to fit changing application needs. This resulted in not all characterisation techniques being applied to every construct. There is, however, enough overlap in terms of characterisation between the different constructs to extrapolate the findings to cover the whole set of sub-unit S1 and S2 pairs.

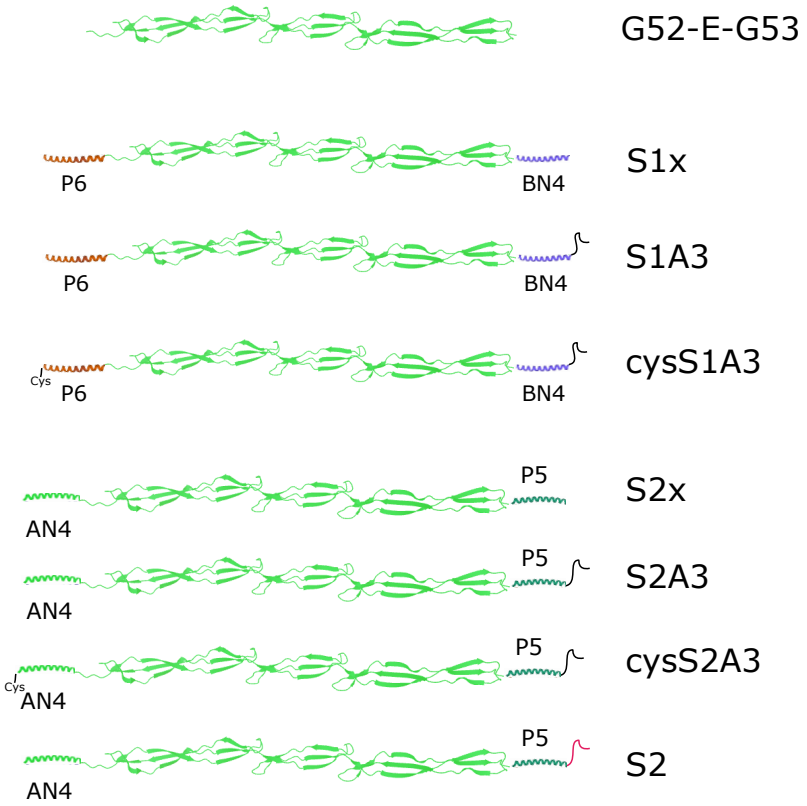


FIGURE 3.15: Diagrammatic summary of S1 and S2 protein variants analysed in Chapter 3.

3.3.6 S1 and S2 Production and Validation

Just as G52-E-G53, all of the sub-unit constructs were cloned into the pET28a plasmid to produce N-terminally poly-histidine (6) tagged protein products. The proteins were over-expressed using the autoinduction medium expression protocol (see Chapter 2). Proteins were purified using Ni-NTA affinity purification tandem with a subsequent size exclusion chromatography step. The final elution was performed in 20 mM phosphate buffer pH 7.4. This particular buffer was chosen, because the two coiled-coil pairs have been characterised and shown to be active at physiological pH ranges [248, 249]. The choice of including phosphates only, as opposed to PBS, was because NaCl would interfere with CD and QCM-D measurements. A low molarity of the buffer component was also assumed to interfere with AFM experiments to a lesser degree. Appendix Figure A.4 shows CD spectra comparison between different

molarity buffers for S1x that indicate little effect of buffer molarity on protein structure. A full evaluation of successful construct purifications can be seen in Figure 3.16.

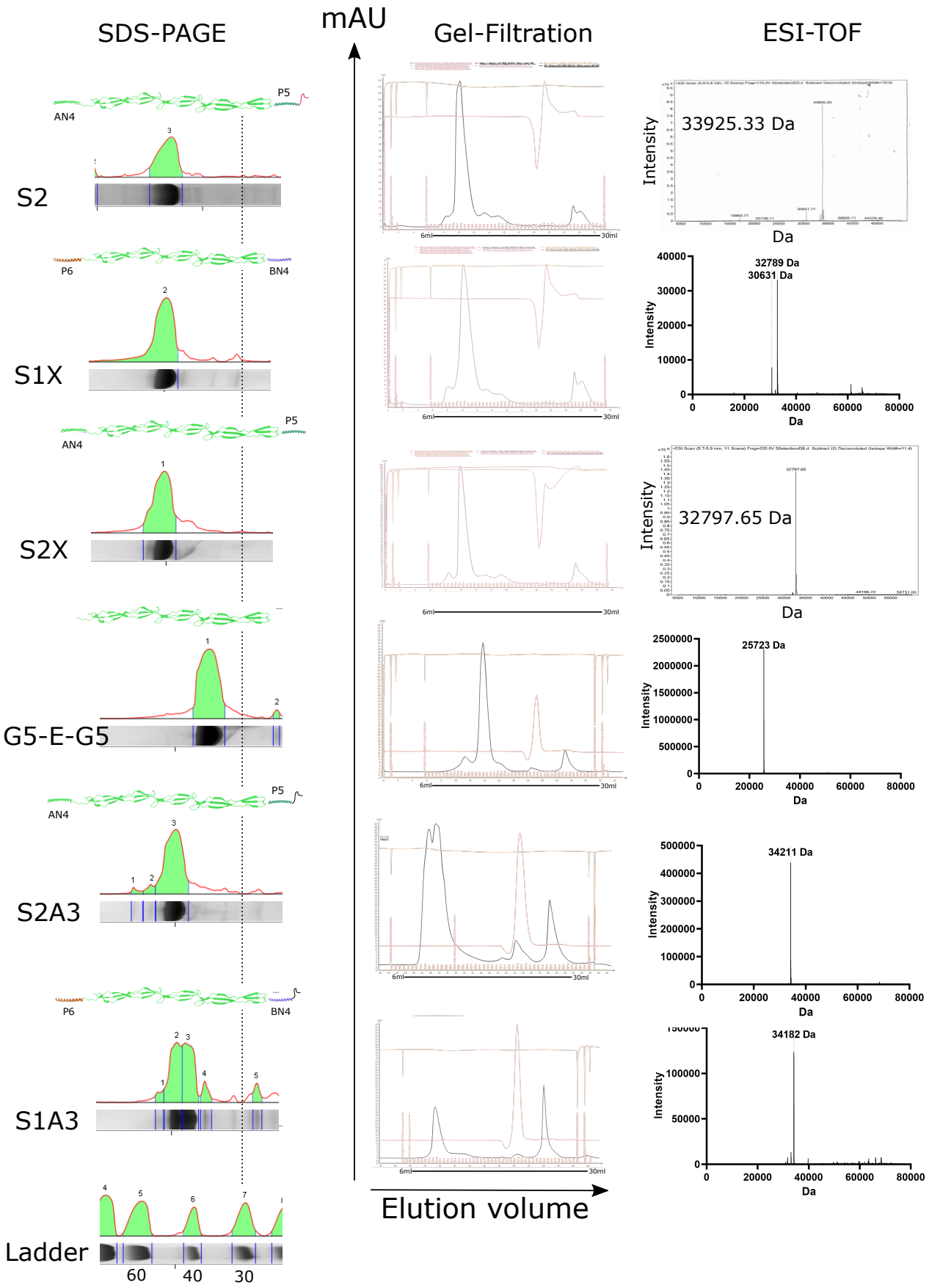


FIGURE 3.16: Purification of S1 and S2 variants. Column one: Tris Glycine SDS-PAGE gels. Column two: size exclusion chromatography traces - elution in 20 mM phosphate buffer pH 7.4. Column three: ESI-TOF results. The size exclusion chromatography traces are adjusted for contrast, but otherwise not altered. Higher resolution representative gel filtration plots can be found in Figure A.2

Depending on the construct, each purified protein (apart from G52-E-G53) should move as an approximately 34 kDa band on SDS-PAGE, but the sizes based on the molecular weight markers came out to be 42-45 kDa (similar offset to the observed aberrant mobility of G52-E-G53). Furthermore, the gel-filtration results show the proteins eluting at the volumes that would correspond to 400 kDa proteins. To assess, whether or not the proteins were interacting with the separation medium, size exclusion chromatography was also performed with 20 mM phosphate pH 7.4 buffer with 150 mM NaCl and no change was observed (see Figure A.3). The mobility shifts can be explained by the presence of the E domain, which causes a mobility shift in the G52-E-G53 protein. The purification of the correct construct was, however, confirmed by analysing the size exclusion chromatography elution fractions corresponding to the prominent peaks using ESI-TOF. The results are shown in Table 3.3.

Construct	Expected Mass (no methionine)	ESI-TOF mass
S1A3	34181.35	34181.67
S1x	32784.86	32789
S2A3	34210.92	34211
S2	33924.62	33925.33
S2x	32797.35	32797.65
G5-E-G5	25722.64	25723

TABLE 3.3: SasG construct ESI-TOF mass table (Mass in Da).

All of the sub-units appear to be have been correctly expressed and of high purity. The small mass discrepancies were within the accuracy limits of ESI-TOF (0.01

%). Having confirmed that the correct polypeptide chains were obtained, the folding states of sub-units were assessed using CD (Figure 3.17).

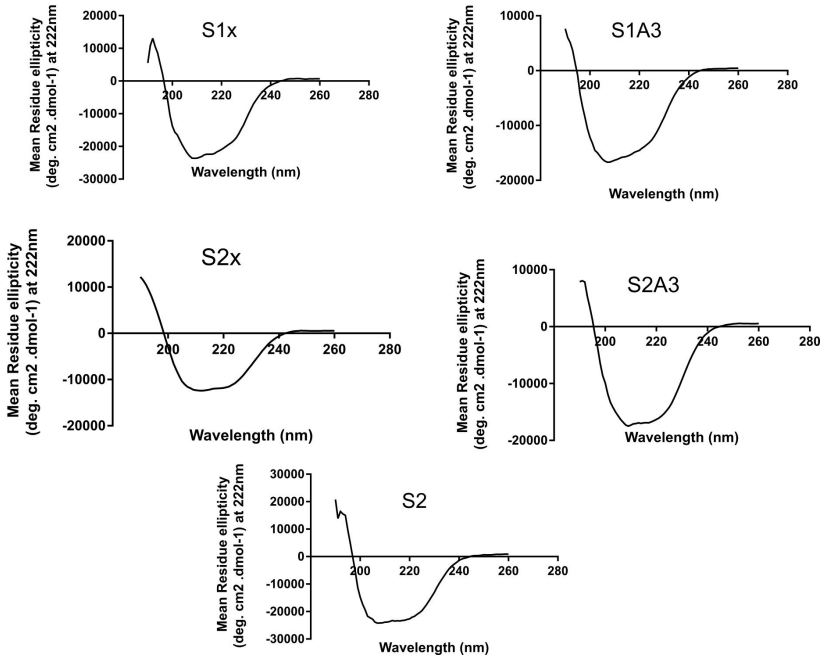


FIGURE 3.17: CD spectra of S1x, S2A3, S2x, S2A3, S2 in 20 mM phosphate buffer pH 7.4 at 20 °C. Proteins analysed at 2.5 μM.

Initially, the shapes of the CD spectra were puzzling as they did not match any expected spectra. Looking at individual features of the spectra, they seem to indicate deformed α-helical spectra with negative peaks at 222 and 208 nm and a positive peak at 190 nm. The 208 nm peak is, however, deeper relative to the 222 nm and does not fit the expectations for a conventional α-helical spectrum, furthermore the height of the 190 nm positive peak seems to be too low. These spectra seem to indicate a mixed structure of α-helices and G52-E-G53. To test this, a theoretical spectrum was created by arithmetically combining spectra of G52-E-G53 and a known coiled-coil [209] (Figure 3.18).

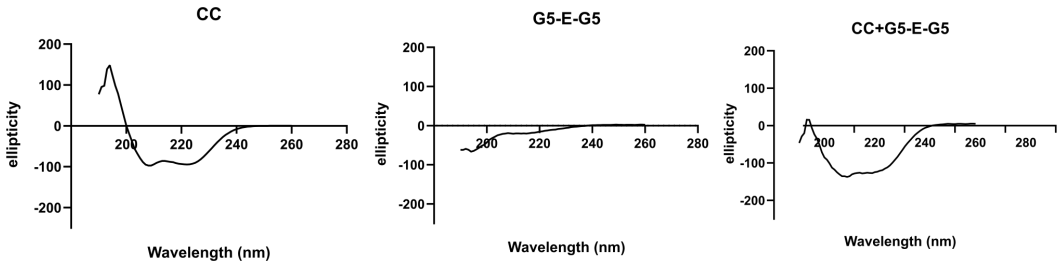


FIGURE 3.18: Theoretical spectrum generated by arithmetically averaging spectra of G52-E-G53 and a known *de novo* designed coiled-coil ccE8 [192].

The resulting spectrum looks like the experimentally derived spectra for the different sub-unit versions. From this, a conclusion can be made that purified S1 and S2 sample contain β-sheet and α-helical structure. By design [249, 248], when sub-units are not mixed, the coiled-coil forming peptides should be disordered and not assume any α-helical structure (with the exception of BN4 present in sub-unit S1 constructs[249]). Here, they all show α-helical presence. It could be the case that being in fusion with G52-E-G53 imparts the peptides with enough stability to fold into α-helices even without their coiled-coil partners. Attempts were made at estimating the sub-unit thermal stability by tracking mean residue ellipticity signal at the 222 nm wavelength band in Figure 3.19**AB**.

Both sub-units exhibit complex denaturation profiles. Considering that each of the sub-units is made up of three components, G52-E-G53 backbone and two unique coiled-coil forming heptad repeats, there should be three unique transitions emerging. Both, S1 and S2, show a negative transition at 53 ° C (+/-1.5 ° C) and 57 ° C respectively. This must correspond to the G52-E-G53 transition (56.6) (Figure 3.19C). With multiple repeats, regardless of exact sub-unit construct, S1 always displays the G52-E-G53 transition at a lower temperature, suggesting a possible destabilisation of the G52-E-G53 structure arising from the fusion with the α-helices. The only other easily discernible transition can be seen in S2 variants at 30 ° C (+/- 1.5 ° C). Due to the fact that the melting profiles are complex, it may be that extraction of exact transition temperatures is not feasible, however it is clear that regardless of the specific sub-unit variant, each sub-unit displays a unique denaturation profile that is reproducible. Overlapping profiles obtained with or without the binding peptide A3 indicate that the presence of a highly charged nanoparticle binding peptide (in

the case of A3 sub-unit versions) does not change the stability of the protein.

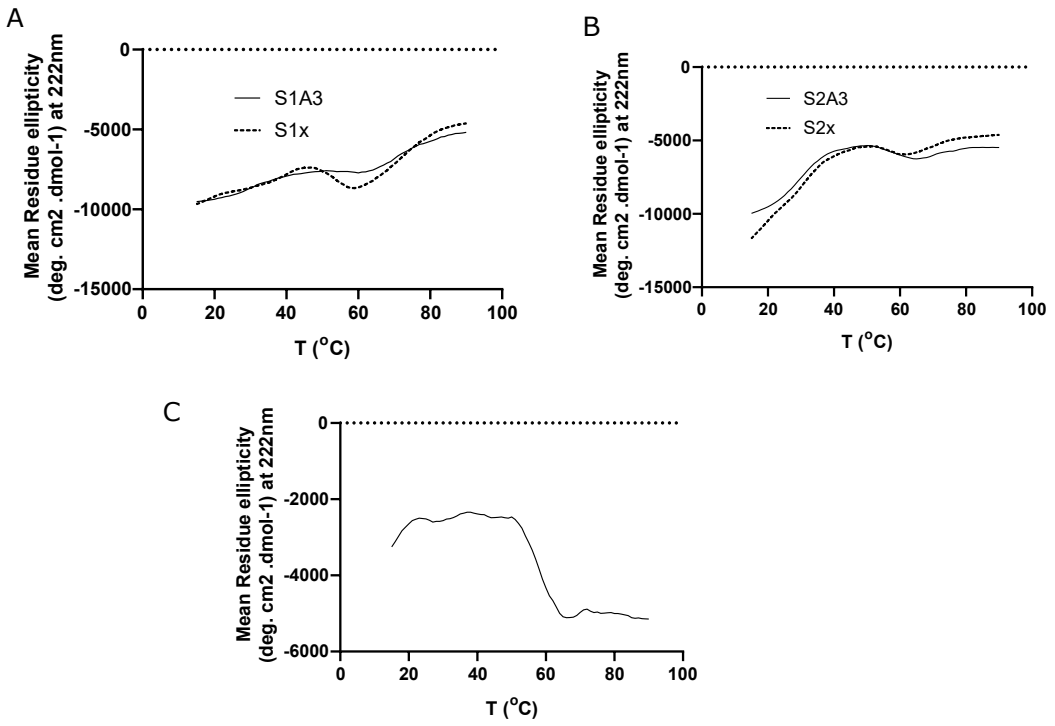


FIGURE 3.19: Thermal denaturation curves tracking mean residue ellipticity at 222 nm band in 20 mM phosphate buffer pH 7.4 **A** Sub-unit S1 variants. **B** Sub-unit S2 variants. **C** G52-E-G53. Proteins analysed at 2.5 μM.

Taken together, the section above characterises the individual components in the two sub-unit system. The G52-E-G53 and sub-unit S1/S2 variants are expressed as folded proteins with reproducible thermal stability values. This lays down the background for investigating the interactions between the sub-unit pair.

3.4 Probing for Sub-unit Interaction

This section focuses on demonstrating that sub-units S1 and S2 interact (regardless of the particular variant). Chronologically, Non-denaturing 'Native' PAGE showed the first hints of interaction between the sub-units. These results were corroborated by reproducible structural changes observed for mixed sub-unit pairs when analysed in CD. QCM-D then shows a specific interaction in a dynamic/flow system. Lastly, direct visual evidence is obtained using AFM in liquid and dry environments.

3.4.1 Native PAGE Interaction Studies

Native PAGE is a technique that allows for detection of protein complexes. Since no denaturant is added, protein complexes held by non-covalent interactions should run as higher molecular weight bands. The exact movement speed of the proteins is dependent on the folding state, thus it is difficult to assess absolute sizes. The Native PAGE experiments were run using a sub-unit pair S1x/S2. First, a concentration trial was performed (Figure 3.20).

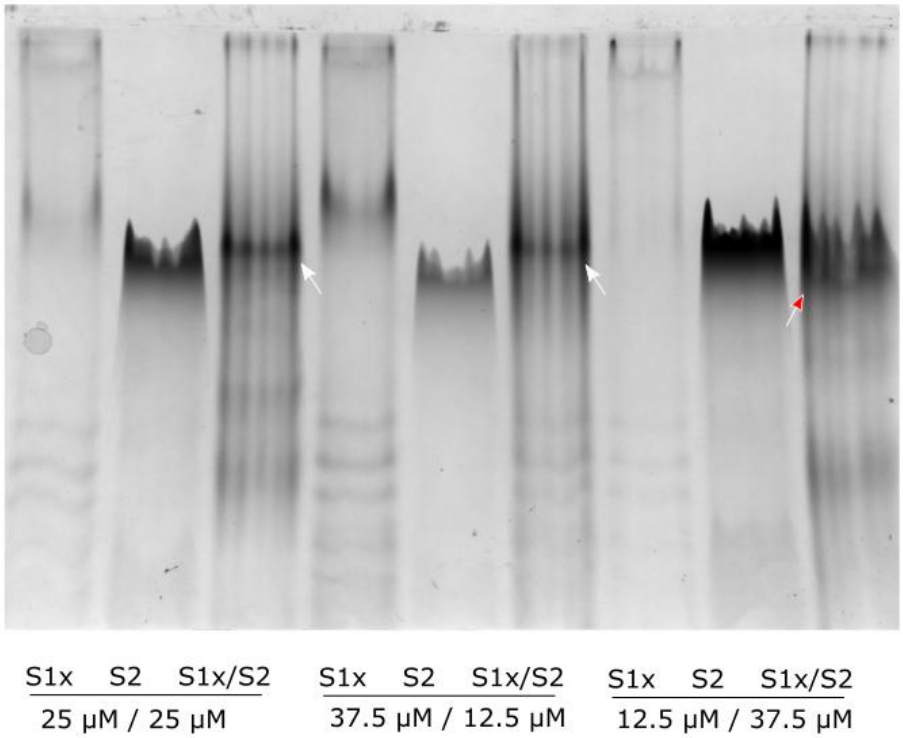


FIGURE 3.20: Native PAGE concentration study. 8 % polyacrylamide gel. Lanes 1, 4 and 7 contain protein S1x at 25 μM, 37.5 μM and 12.5 μM, respectively. Lanes 2, 5 and 8 contain protein S2 at 25 μM, 12.5 μM and 37.5 μM, respectively. Lanes 3, 6 and 9 contain protein S1x/S2 mixtures at 25 μM/25 μM, 37.5 μM/12.5 μM and 12.5 μM/37.5 μM, respectively. Gels run for 3 hours at 80V (constant).

The first three lanes show an equimolar mixture of the two sub-units. There seems to be a novel band appearing in lane with the two sub-unit mixture. The same band appears at 3:1 mixture of S1x:S2. At inverse concentration ratio, the band does not seem to appear. It might be that S1x concentration measurements were overestimated (looking at the S1x it seems that there is very little protein material). The novel band is only slightly different from that of S2, for better separation another gel was also run for six hours (Figure 3.21).

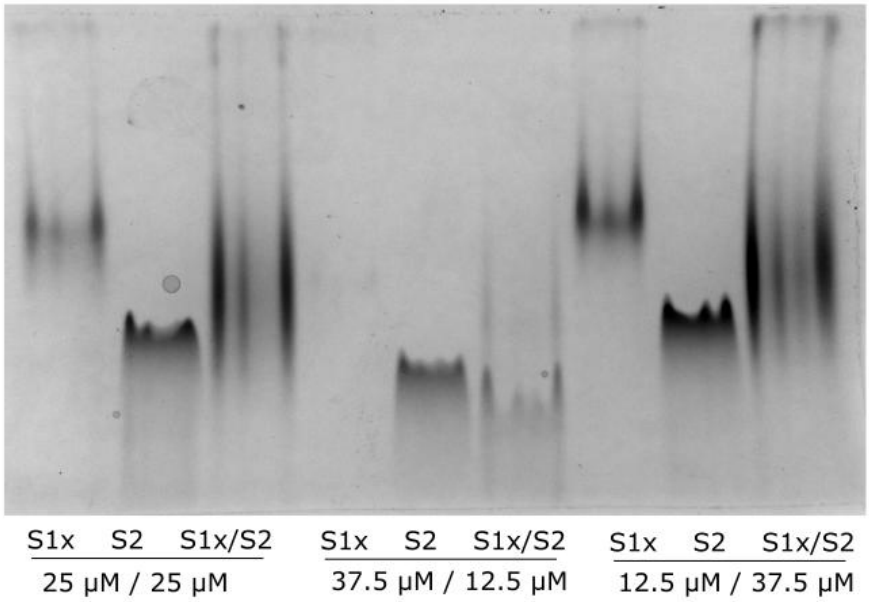


FIGURE 3.21: Native PAGE concentration study repeat. 8 % polyacrylamide gel. Gel run until full band separation (six hours).

With a longer run time, the novel intermediate band shows smearing. Generally, such stretched bands mean that there is a population of different objects with incrementally different properties. Native PAGE separates objects based on hydrodynamic radius and based on charge. It may be that the intermediate band appears because S1x/S2 mixture is forming assemblies of variable lengths. Each of the assemblies would have a different isoelectric point depending on how many of each sub-unit it contains and these would run at slightly different speeds. At pH 8.8 (pH of the resolving gel) S1x would be expected to run higher than S2. Another Native PAGE experiment was also conducted to assess whether or not pre-heating the sample has any effect on sub-unit assembly (Figure 3.22).

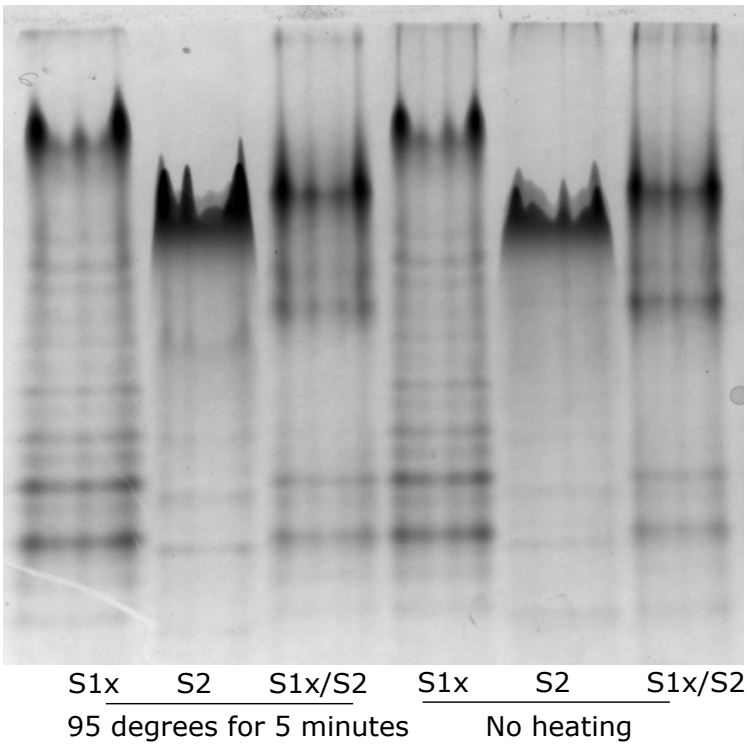


FIGURE 3.22: Native PAGE heating experiment. 8 % polyacrylamide gel. Gels run for 4 hours at 80V (constant). Proteins analysed at 25 μM concentration.

The hypothesis was that heating would disrupt any potential homodimers that either S1x or S2 would have formed during storage prior to the experiment. The experiment was performed with a less pure set of samples, however the same novel band appears in S1x/S2 mixtures (equimolar concentration) regardless of the heating status. The individual sub-unit lanes also run exactly the same as in the previous experiments, potentially suggesting that heating has no effect on the homo-oligomerisation status. Together, Native PAGE results reproducibly show an interaction between S1x/S2 mixtures, however it is difficult to infer the exact nature of the interaction.

3.4.2 Protein-protein interaction studies using CD

In order to study the structural changes taking place during inter sub-unit polymerisation, CD was employed. First, thermal denaturation curves were obtained by tracking mean residue ellipticity at 222 nm of sub-unit pairs S1x/S2x and S1A3/S2A3 (Figure A.5).

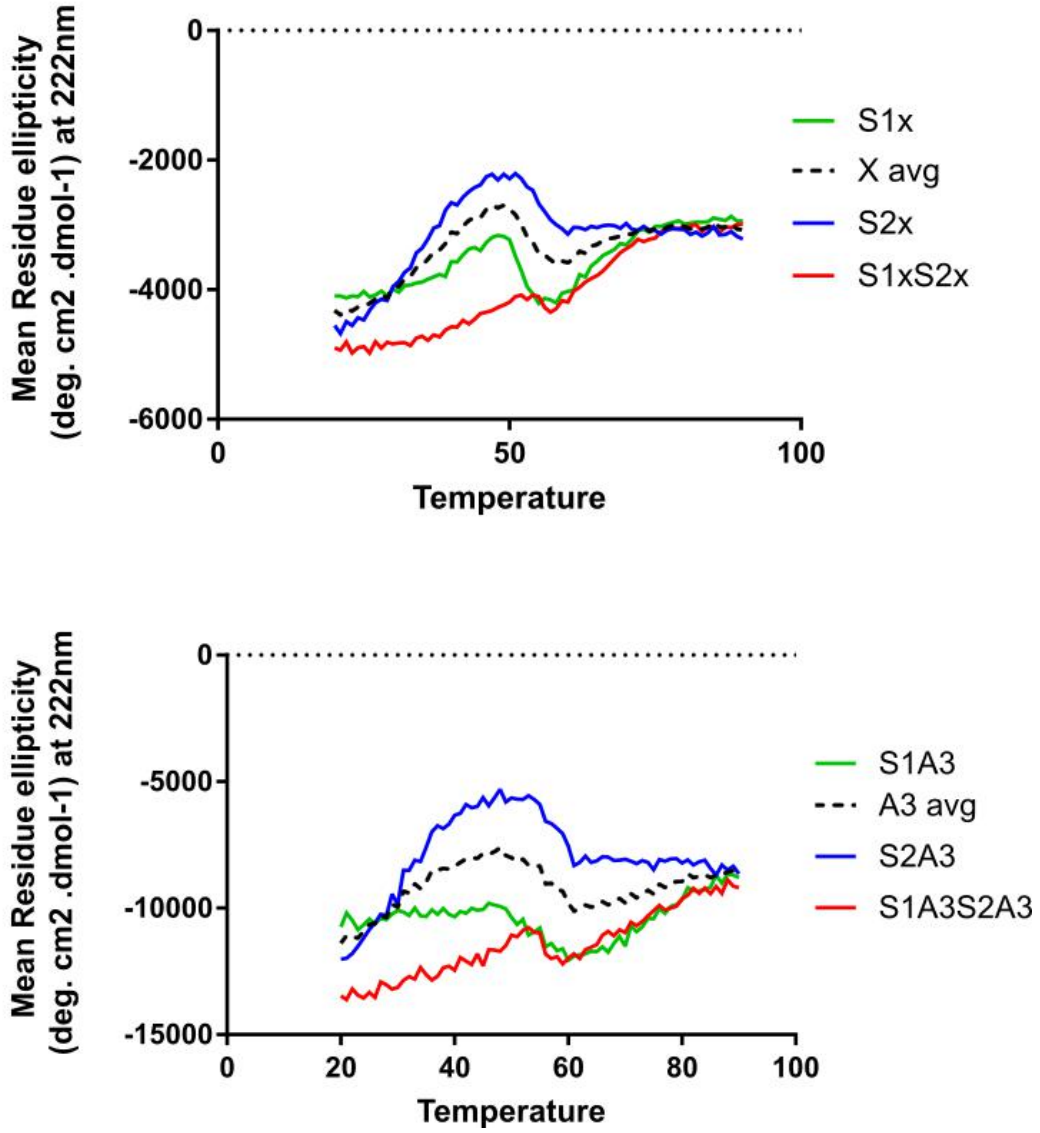


FIGURE 3.23: Thermal denaturation curves tracking mean residue ellipticity at 222 nm in 20 mM phosphate buffer pH 7.4. The puncture lines signify an arithmetically derived average profile of the respective pairs of S1x/S2x and S1A3/S2A3. Proteins analysed at 2.5 μM.

Both pairs were run in parallel to rule out any effect of presence of the binding peptide on structural changes. Thermal denaturation profiles produced by S1x/S2x and S1A3/S2A3 mixtures emerge to be novel, thus suggesting an interaction between the sub-units. If the sub-units were not interacting, one would expect to obtain a profile that is the average of the two individual profiles. This is not the case and is much more clearly seen in Figure 3.23, the average mixtures do not overlap with the experimentally derived melting profiles of the mixed sub-units. Due to the complexity of the spectra (G52-E-G53 crowds the signal), it is difficult to extract melting temperatures. A figure from pooled data sets (appendix Figure A.8) does

show three transitions at 43 °C, 55.25 °C , 71.4 °C that most likely correspond to P5/P6, G52-E-G53 and AN4/BN4 transitions respectively. This strongly evidences a specific coiled-coil interaction between the two sub-units. A more detailed view of this is provided by spectra taken at different temperature points in Figure A.6.

The spectra presented side by side show that S1A3 does not see any significant structural changes until it reaches 50 °C, where a negative peak corresponding to random coils starts appearing indicating the unravelling of α-helical content. Sub-unit S2A3, on the other hand, starts exhibiting structural re-configurations at 30 °C and reaches random-coiled conformation at 35 °C. Lastly, S1A3/S2A3 not only exceeds the stability of S2A3, but is also structurally stable beyond 50 °C, where S1A3 starts unravelling its structure. Only at 55 °C does the S1A3/S2A3 mixture start losing α-helical structure. This re-affirms the suggestion that the sub-units are interacting, and shows that they interact via stabilised α-helices within coiled-coil structures. The same experiment was performed for sub-unit mixture S1x/S2x, showing that this effect is present regardless of the presence of the A3 binding peptide (see appendix Figure A.7).

3.4.3 QCM-D Interaction Studies

Having established that the sub-units interact, S1A3 and S2A3 have been modified to replace Ser19 residue with cysteine to produce constructs labelled cysS1A3 and cysS2A3 (see Chapter 2) for QCM-D experiments. QCM-D is a technique that can allow assessment of changes in mass on surfaces in real time (see Chapter 2). Gold surfaces were chosen and it was thus assumed that having cysteine residues at N-termini would allow for immobilisation of sub-units in specific orientations (N-terminus immobilised on surface and C-terminus exposed to the solution) as the cysteine side-chains should form gold-thiol bonds. A specificity experiment was run to assess if an immobilised layer of one sub-unit will preferentially interact with its binding partner (Figure 3.24).

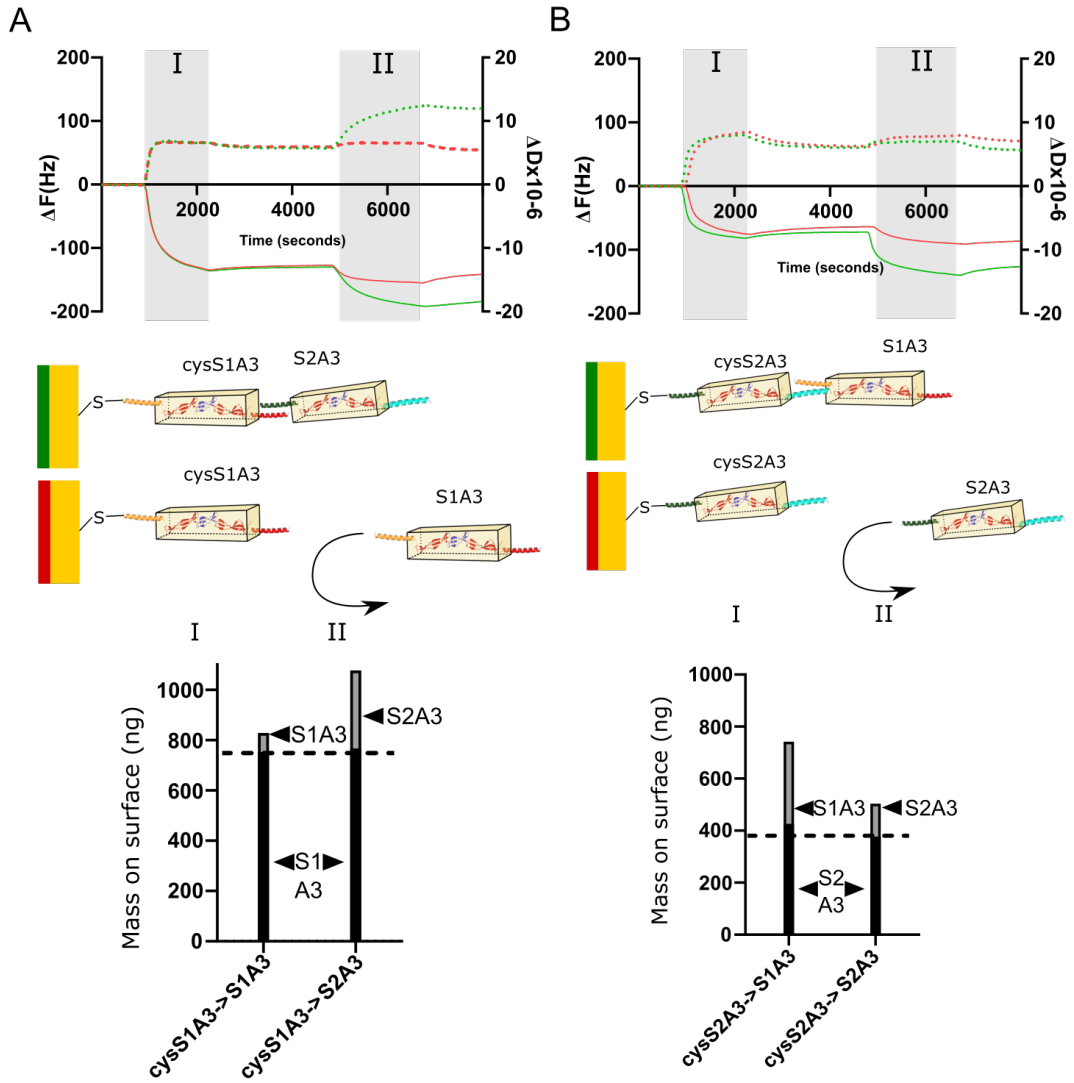


FIGURE 3.24: QCM-D assembly specificity studies. **A** Sub-unit S1 specificity experiment. **I** shows an increase in mass (as indicated by a negative frequency change) by applying cysS1A3 (solid lines – frequency, dashed lines - dissipation). In stage **II** the cysS1A3 layer was exposed to S2A3 (green) and showed an increase in mass that persisted after washing, whereas cysS1A3 layer exposed to S1A3 showed a much smaller increase in mass (red). **B** An inverse experiment applying cysS2A3 first. A layer of cysS2A3 exposed to S1A3 (green) shows an increase in mass higher than cysS2A3 exposed to S2A3 (red). Proteins were applied at 150 μg/mL at a 20 μL/min flow rate. Curly arrows represent sub-units not interacting.

The experiment shows a larger increase in mass on surface when cysS1A3 is exposed to S2A3 and cysS2A3 is exposed to S1A3 than when cysS1A3 is exposed to S1A3 or when cysS2 is exposed to S2A3. This most likely arises from the fact that the immobilised layers of cysS1A3 and cysS2A3 expose the coiled-coil forming heptad repeats BN4 and P5 ,respectively, to the solution. Only S2A3 is then able to attach itself to the cysS1A3 coated surface via the BN4/AN4 coiled-coil interaction. The same happens in channels where cysS2A3 is immobilised, only S1A3 containing a

P6 heptad repeat can interact with the exposed P5 forming a coiled coil. This effect was observed during multiple experiments mass changes of which were pooled and graphed in Figure 3.25.

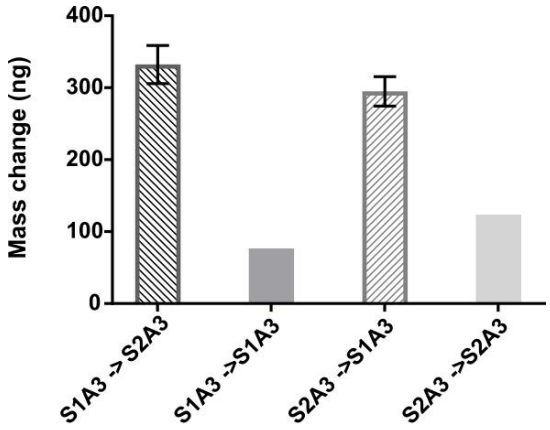


FIGURE 3.25: Pooled mass change data from different QCM-D experiments showing reproducible binding specificity between S1A3 and S2A3. Error bars are standard deviations (population). S1A3->S2A3 and S2A3->S1A3 results were pooled from three separate experiments each.

The fact that the mass increase seen above is due to coiled-coil interactions is further confirmed by an experiment where a layer of cysS2A3 is exposed to either S1A3 or G52-E-G53 (Figure 3.26).

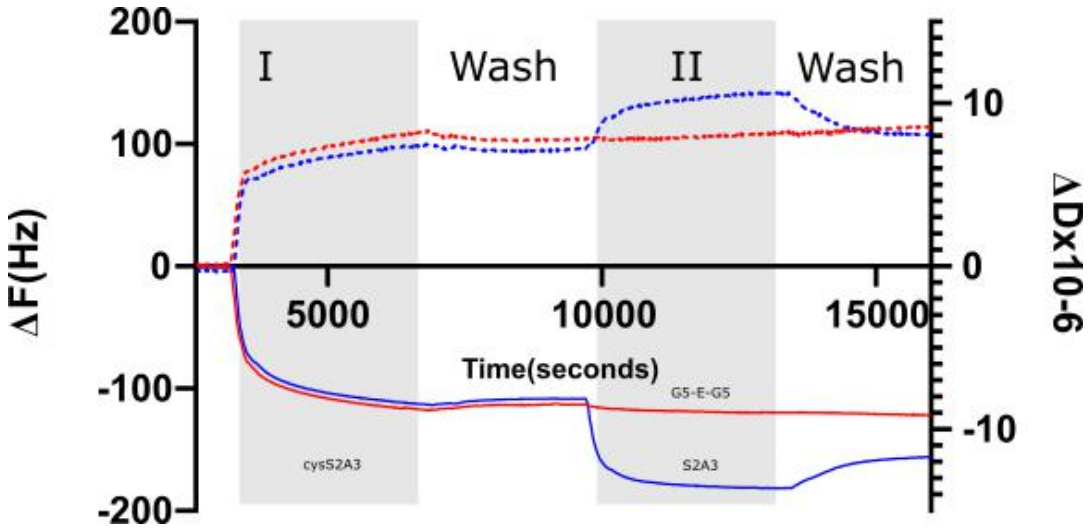


FIGURE 3.26: QMC-D experiment showing the need for coiled-coils for interaction. Blue solid trace shows frequency changes where cysS2A3 is exposed to S1A3. Red - shows cysS2A3 exposed to G52-E-G53. Dissipation traces are plotted as dashed lines.

There appears to be no interaction between the cysS2A3 layer (solution exposed P5 heptad repeats) and G52-E-G53 containing no coiled-coil forming moieties. As

discussed above, the coiled-coil pairs have been designed to interact at physiological pH conditions. It has been shown that certain electrostatically bound coiled-coil pairs unravel at extreme pH conditions [243, 120]. Glutamate side-chain assumes a neutral charge at pH of around 4, with glutamate side-chains neutral, positively charged lysine residues would start repelling each other, thus washing with a solution at pH in that range is expected to disrupt the coiled-coil interaction. This was tested for assemblies of cysS2A3->S1A3 and cysS1A3->S2A3 in Figure 3.27.

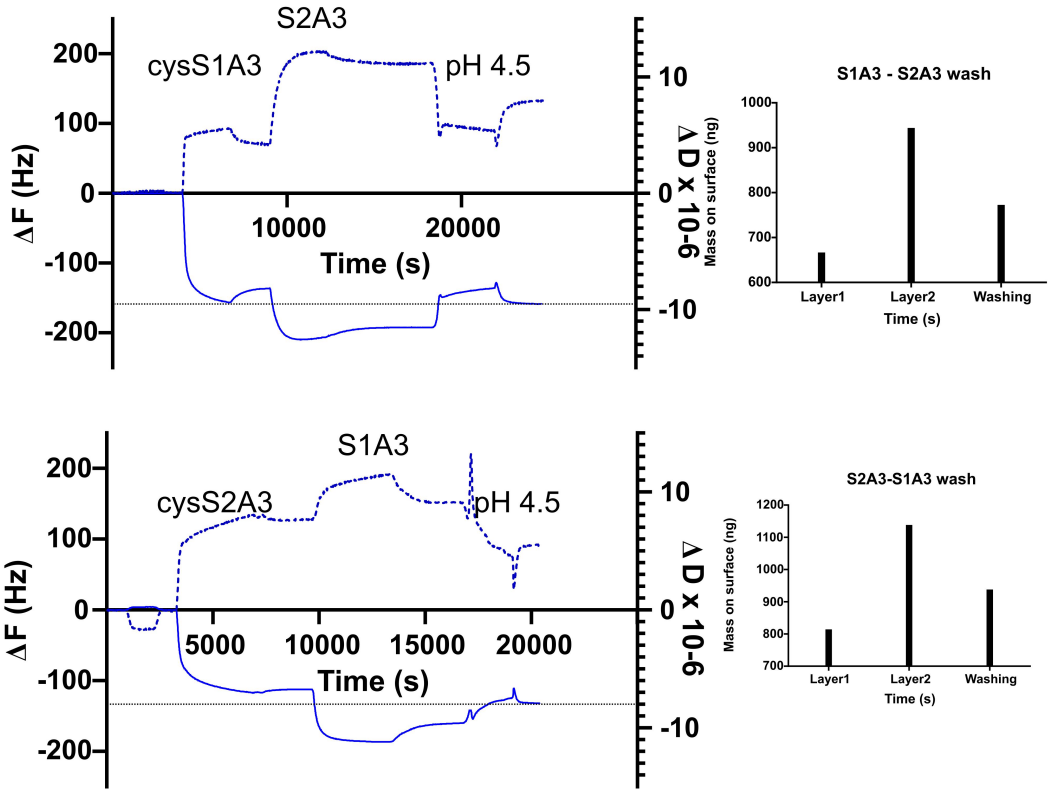


FIGURE 3.27: Figure QCM-D and pH dependence. The horizontal dashed lines represent the final frequency change at the end of the experiment following the final buffer pH 7.4 rinse. Frequency - solid lines, dissipation -dashed lines. The effect of the low pH buffer solution are tested prior to the application of the proteins.

In both cases the second layer constituted of either S2A3 or S1A3 is removed by a wash with phosphate solution at pH 4.5 (the pH was adjusted using HCl). Upon initial application of the low pH solution a number of transient fluctuations can be seen. These are stabilised by returning to phosphate buffer at pH 7.4. The calculated mass on surface at time-points corresponding to Layer1, Layer2 and the state after washing is also plotted (right of the QCM-D curves in Figure 3.27). The

mass on surface after washing is much smaller than that of the Layer2. Both curves cysS1A3->S2A3 and cysS2A3->S1A3 demonstrate a pH sensitive assembly of sub-units via coiled-coils AN4/BN4 and P5/P6, respectively. The fact that after washing the calculated mass was slightly higher, can be explained by the drift inherent to the experimental system. Growing longer sub-unit assemblies on the surface was also attempted (Figure 3.28).

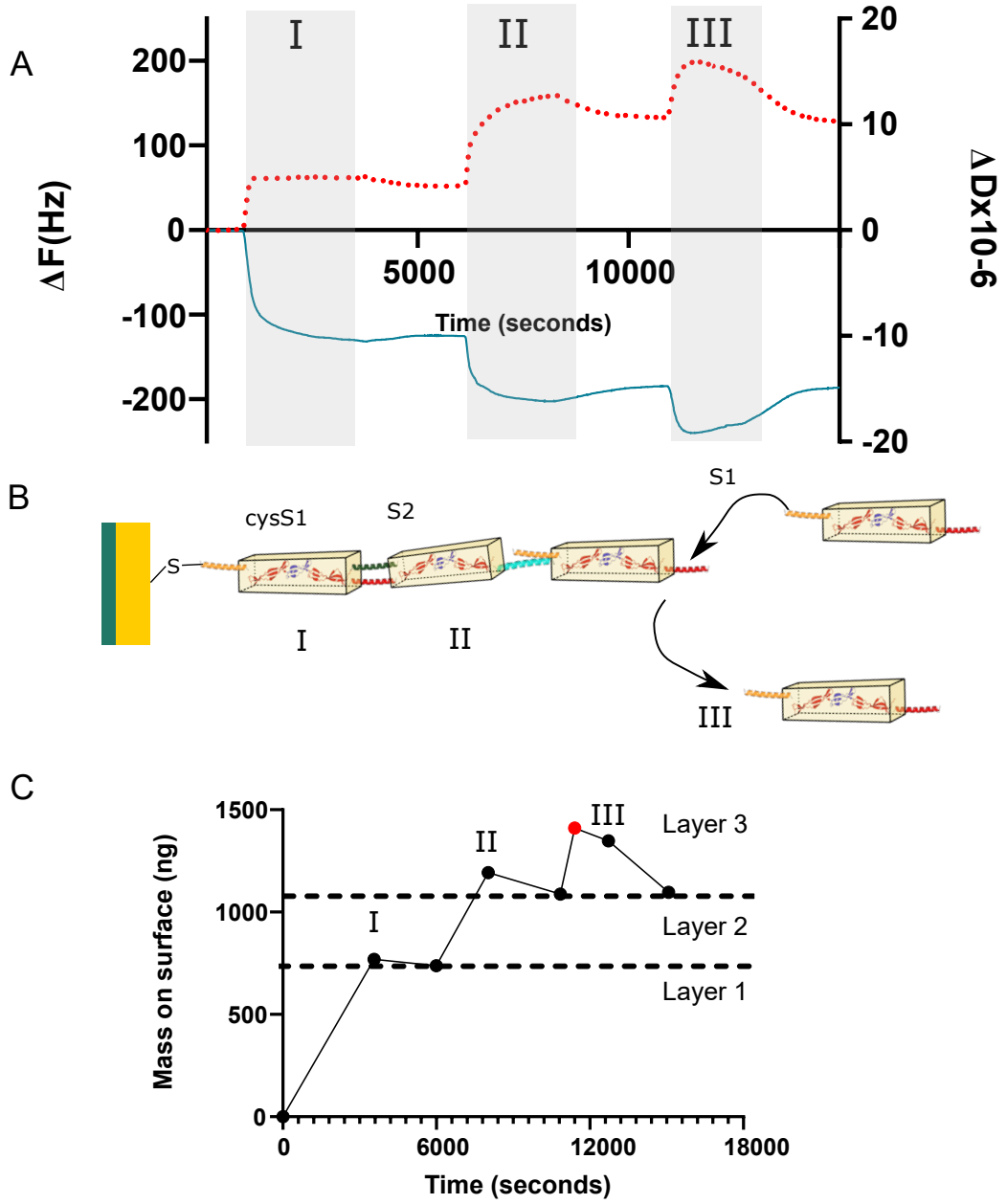


FIGURE 3.28: QCM-D experiment showing the maximum length that can be grown on a surface. **A** Maximal continuous rod assembly on a gold surface. **B** A theoretical illustration of higher order assembly in **A**, where an applied third layer S1A3 transiently interacts with the S2A3 layer, but is then lost. **C** Tracks relative mass changes corresponding to different protein layers observed in **A**. The red data point signifies the peak mass on the surface corresponding to initial adsorption of S1A3.

Figure 3.28 is divided into three phases (**I**, **II** and **III**). **I** shows formation of a saturated layer of cysS1A3 followed by a buffer wash. **II** shows a further increase in mass after adding S2A3 until saturation. Finally, at stage **III**, applying S1A3 to the surface shows an initial increase in mass associated with surface binding followed

by a slight decrease prior to washing, indicating desorption from the surface and a complete return to frequency corresponding to two layers after washing. The reduction in mass before a washing step might arise due to desorption of S1A3 caused by sheer forces or an instability of a surface bound three layer assembly (molecular crowding). The transient mass increase only seems to take place when a sub-unit layer is exposed to an expected binding partner as shown in Figure 3.29.

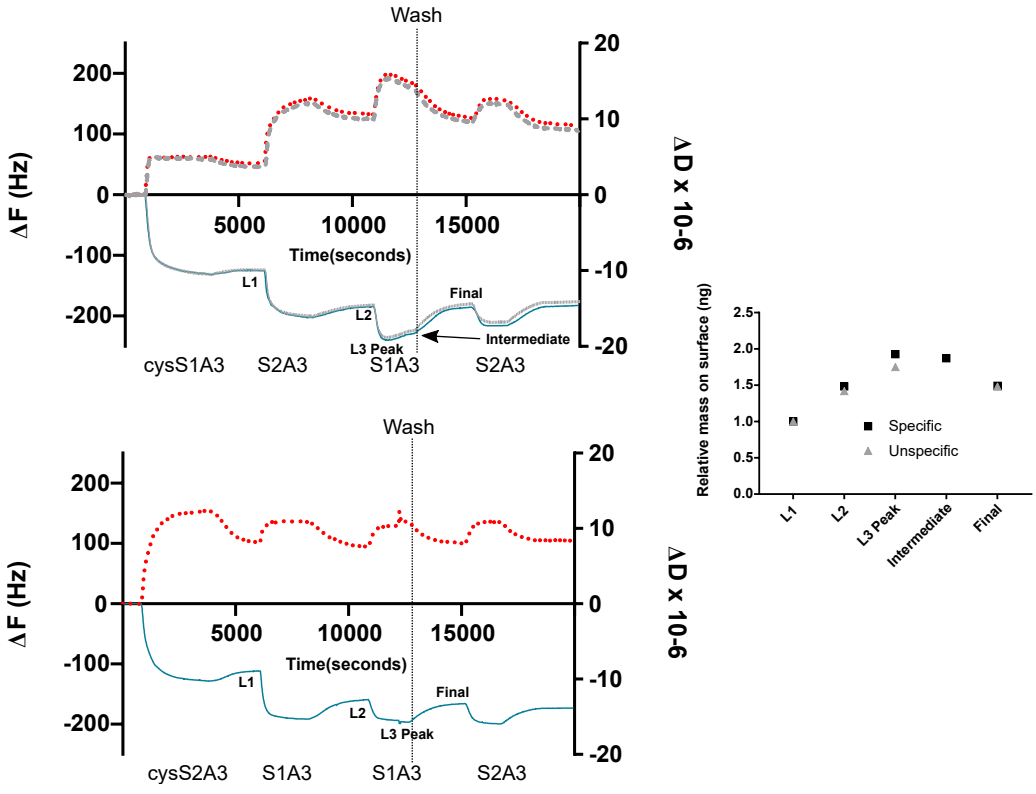


FIGURE 3.29: QCM-D third layer behaviour specificity. The relevant pH 7.4 wash is indicated by a perpendicular dotted line. On the left: detected mass on surface at labelled time points corresponding to molecular events. Values plotted are relative mass changes on surface.

The top QCM-D curve shows two identical repeats of an attempted three layer assembly of cysS1A3->S2A3->S1A3. Upon sample application, both curves show the transient increase and then a reduction in mass, even prior to washing. The curve below shows an analogous third layer growth attempt with cysS2A3->S1A3->S1A3. In this case, the applied S1A3 is not expected to interact with the S1A3 on the surface and the observed curve confirms this. The curve resembles frequency

changes expected for a non interacting protein. The figure shows that, while in both scenarios the final outcome is the same - washing returns frequency levels to only two stable layers, in the case of cysS1A3->S2A3-S1A3, sub-unit S1A3 'attempts' to form a third layer, but is destabilised by physical factors, most likely inherent to the experimental system and not to the protein assembly itself. Immobilisation of protein assemblies on a surface increases the local protein concentration (relative to the overall solution), such molecular crowding may cause the assemblies to be disrupted. Effects of molecular crowding on surface assemblies have been assessed before using on-surface CD [247], where molecular crowding effects were shown to change conformation of coiled-coil binding partners.

3.4.4 Visualisation of Sub-unit Assemblies

As a first means of direct visualisation, cysteine containing cysS1A3 was labelled with a fluorescent dye (Dylight650). The resulting labelled proteins was assigned a name of S1cys650. A time-course assembly experiment was carried out and visualised under a fluorescence microscope (Figure 3.30).

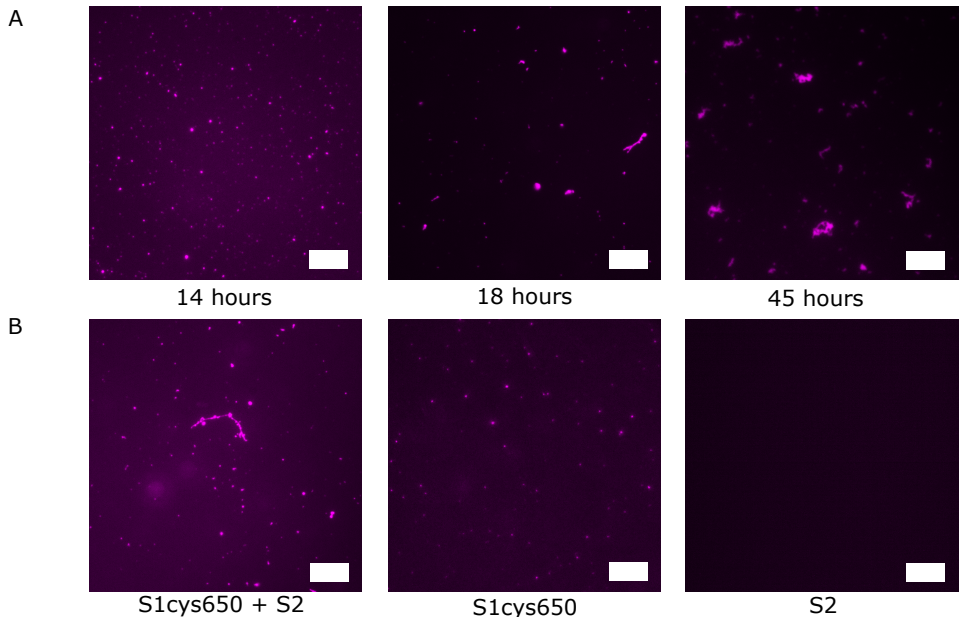


FIGURE 3.30: **A** A time course experiment visualising S1A3cys650/S2A3 mixtures in PBS pH 7.4. **B** Comparison of S1cys650+S2A3, S1cys650 and S2A3 at the 18 hour time-point. Proteins were prepared at concentrations of 100 μg/ml. Scale bars are 10 micrometers. Images presented with artificially enhanced contrast.

In can be seen that, for the first 14 hours mixtures of labelled S1cys650 and unlabelled S2A3, do not form elongated structures. At the 18 hour time-point various elongated aggregates start appearing. The structures appear even more aggregated at the 45 hour time point. No assembly/aggregation at the 18 hour time point can be seen in samples containing only S1cys650 or unlabelled S2A3. This shows low resolution evidence of specific sub-unit assembly. The time dependent assembly of large structures is also corroborated by a DLS experiment (Figure 3.31).

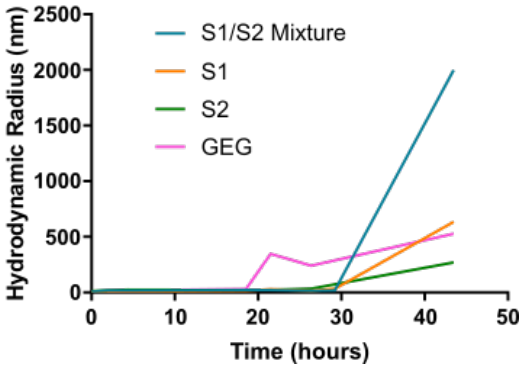


FIGURE 3.31: DLS time-course assembly experiment in 20 mM phosphate buffer pH 7.4. Proteins were prepared at concentrations of 100 μg/ml.

Here, S1A3/S2A3 mixtures are shown to assemble/aggregate starting at the 30 hour time point and continue to assemble until the 40 hour mark (where the experiment was concluded). Individual sub-units do start to assemble/aggregate at the same time point, but the kinetics seem to be significantly different. G52-E-G53 starts to assemble/aggregate earlier on, but does not reach the levels of S1A3/S2A3 mixtures. The fluorescence microscopy and DLS data do not give discrete information about the hetero-polymers formed by S1A3/S2A3 mixtures, but they do inform as to which are the time points wherein large assemblies can be expected. DLS data also shows that having coiled-coil forming heptad repeats on either end the G52-E-G53 protein must impart it with additional solubility as the singular G52-E-G53 aggregated at an earlier point. The discrepancy between optimal times for assembly determined by DLS and fluorescent microscopy can be accounted for by the fact that DLS was performed in 20 mM phosphate pH 7.4 and did not contain NaCl, which might have altered the assembly kinetics. These data were used to inform the design of a liquid AFM experiment. A 33 hour time point was chosen as, based on DLS, it is when S1A3/S2A3 should show structures of approximately 1000 nm in radius and

individual sub-units should only produce relatively small assemblies/aggregates. The results can be seen in Figure 3.32.

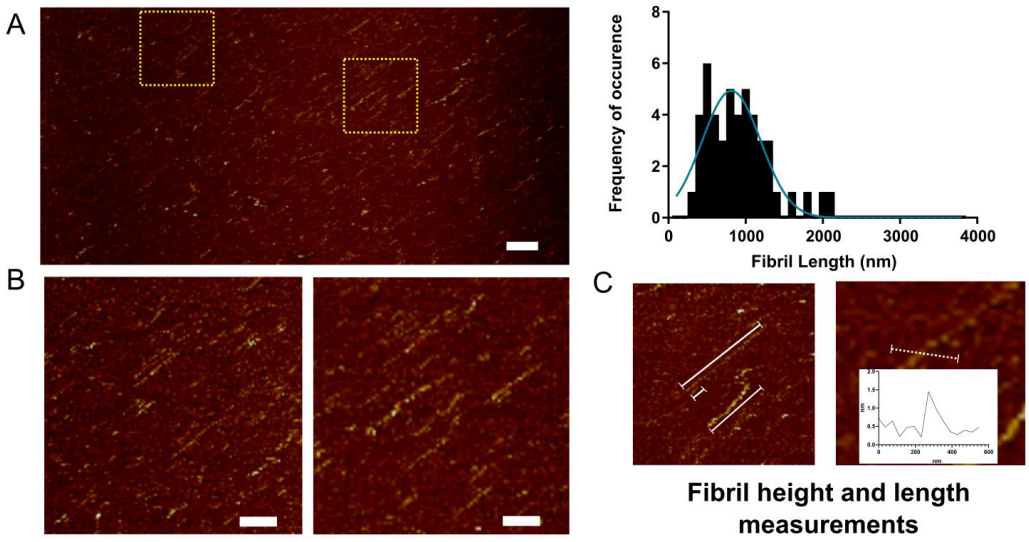


FIGURE 3.32: Liquid AFM experiment performed with S1A3/S2A3 on a PLO surface in HPLC grade water. **A** Large view of the scan showing multiple long objects. Regions to be magnified highlighted with yellow dashed lines. Distribution of object lengths is presented it the histogram on the right (n=98). Scale bar - 1 μm. **B** Magnified regions showing assembled fibres. **C** Demonstration of how heights and length measurements were taken. Pixel-pixel distance 39 nm. S1A3 and S2A3 assemblies are expected to be slightly negatively charged in solution and thus attracted to positively charged PLO surfaces.

The figure shows a distribution of long extended fibres. When measured for height, the fibres average at 1 nm, which matches the expected G52-E-G53 rod thickness. Measurements of the fibre lengths reveal a size distribution of around 1000 nm. This may mean that the similarly sized objects observed in DLS are indeed elongated fibres (rather than generic protein aggregates). A closer inspection of the fibres shows that they are composed of small individual units. The pixel-pixel distance on the x-y plane for the images is 39 nm making exact measurements of 17 nm length sub-units unattainable, however the individual segments making-up the fibres have dimensions of elongated structures. A separate experiment was performed with the same conditions (buffer, temperature, time) to assess if sub-units S1A3, S2A3 and G52-E-G53 form comparable structures (Figure 3.33).

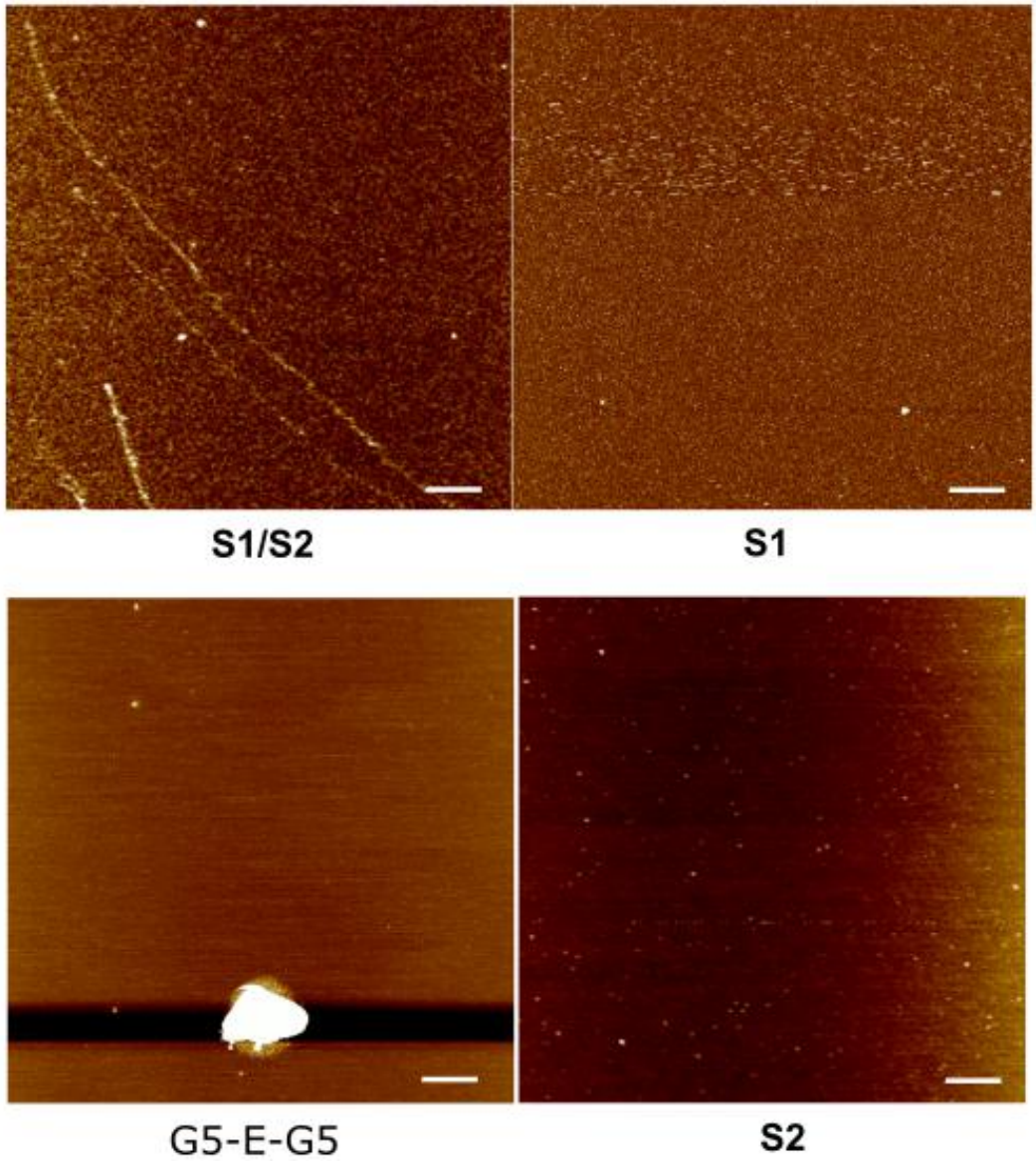


FIGURE 3.33: Liquid AFM experiment performed with S1A3/S2A3, S1A3, S2A3 and G52-E-G53 on a PLO surface in HPLC grade water. Samples incubated for 33 hours at room temperature at 100 μg/ml concentration and applied to the surface at 5 μg/ml. Scale bars - 500 nm.

The results of the repeats show that S1A3/S2A3 mixtures assemble into long chains, S1A3 and S2A3 both show no large structures immobilised on the surface and G52-E-G53 contained large aggregates (50-150 nm in height). This shows that the S1A3/S2A3 fibre assembly is reproducible and only takes place when both sub-units are present in the mixture. The observed G52-E-G53 aggregates seem to correspond to the size of the objects observed in DLS. While the liquid AFM results show long micro scale assemblies, the resolution of the images did not allow the investigation of the composition of the fibres. In order to obtain higher resolution

information, dry - air tapping mode images of the assembled fibrils were also taken (Figure 3.34).

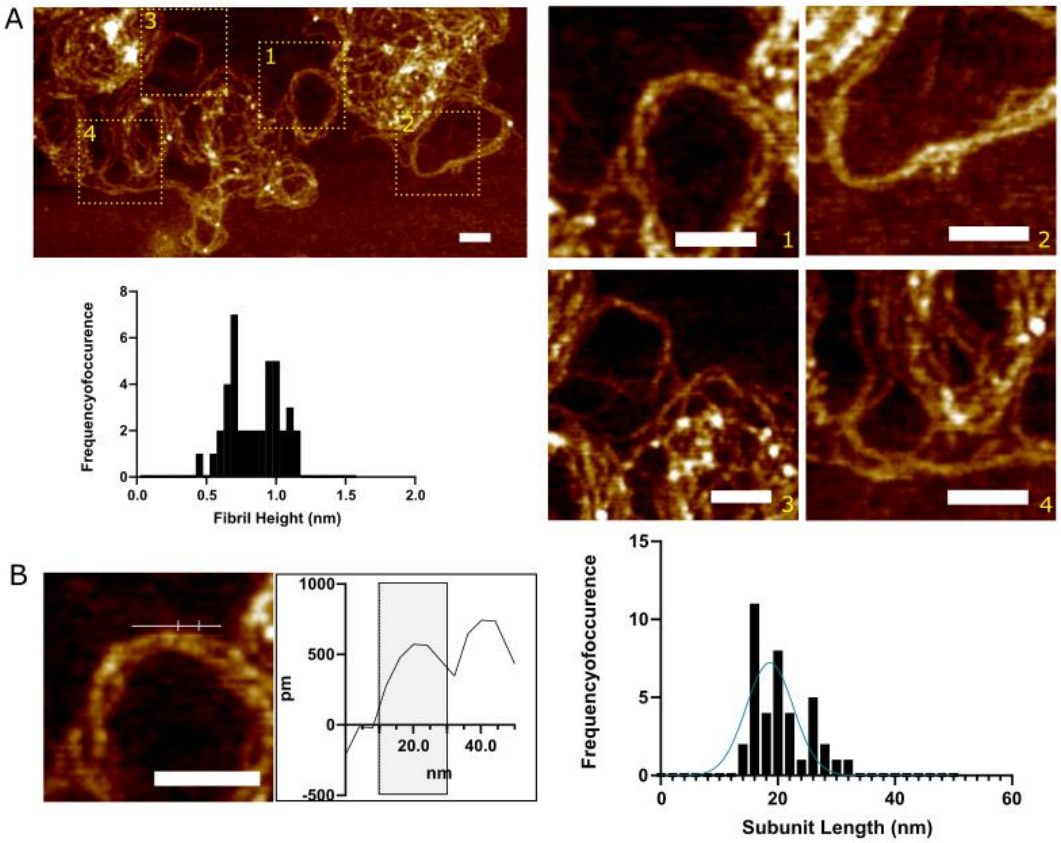


FIGURE 3.34: **A** Dry tapping mode AFM experiment on S1A3/S2A3 mixtures with large protein assemblies being detected. Proteins were prepared by mixing the two sub-units at 100 ng/mL on a poly-l-ornithine coated surface for 2 h then washed with water and quickly dried. Magnifying parts of the large image (yellow dotted outlines) reveals rope-like assemblies made up of small segments. Quantifying multiple height sections showed that the rope-like fibrils were around 1 nm, consistent with width of a single chain of the protein (n=26). **B** Shows the method for measuring sub-unit length (n=50). All Scale bars represent 100 nm. The pixelpixel distance was 4 nm. Samples prepared and images taken in collaboration with Laia-Pasquina Lemonche

Visualising the self-assembly of mixtures of S1A3 and S2A3 in air revealed curled fibrous structures that were also close to 1 nm in height. Since the sample was prepared by incubating the proteins on the surface and then drying them, it could be that the fibril aggregation in this way is an artifact of the drying process, nevertheless it further confirms the ability of S1A3 and S2A3 to form long fibres. The curled fibres seem to be forming intertwined ropes. A close examination of the segments reveals that the fibres are made up sub-units that are on average 20 nm in length. The sub-units are assembled end to end longitudinally. These two observations show

that the fibres are assembled of S1A3 and S2A3 along the long axis (and not stacking side-by-side) hinting that the assembly is mediated via terminal coiled-coils.

In summation, the microscopy results combined with DLS define the parameters for S1A3 and S2A3 assembly and show that in various contexts the sub-units form long (micron scale) fibres in a mutually dependent manner.

3.5 Fibril Macro State

At the outset of the project an assumption was made that S1 and S2 variants may form some non-specific interactions with themselves when in isolation, thus a heating protocol was used to try and dis-associate any interactions with self. The sub-unit variants S1x and S2 were mixed and placed in a thermocycler that was programmed to change the temperature to 95 °C for 2 minutes, 70 °C for 1 minute, 55 °C for 1 minute, 30 °C for 1 minute, 20 °C for 1 minute and 4 °C for 10 minutes in sequence. The hypothesis was that a short burst of thermal energy would disrupt any pre-formed interactions by the monomers and a steady decrease in temperature would facilitate the self assembly of the two component system. The resulting assemblies were visualised via dry air tapping mode AFM and TEM.

3.5.1 Heating/Cooling Protocol AFM Assemblies

Thermally treated S1x/S2 mixtures were dried on bare freshly cleaved uncoated mica surfaces at various concentrations (achieved through a serial dilution) (Figure 3.35).

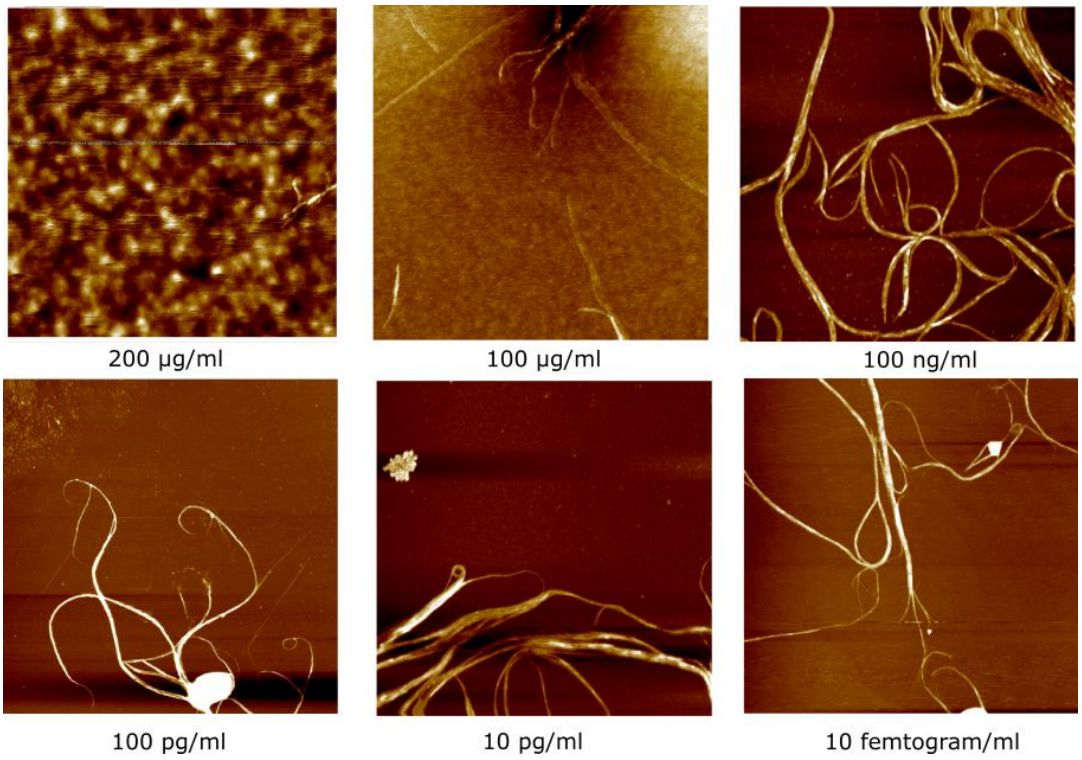


FIGURE 3.35: Dry air tapping mode AFM images of different concentrations of S1x/S2 assembled after thermal treatment. The listed sample concentration refers to the concentration of the protein applied to the mica surface before drying. Drying accomplished through passive evaporation at room temperature. The image area is 5 μm^2.

The concentration range reveals that there is too much material when protein is dried at 100-200 μg/ml. The sample preparation at 100 μg/ml does, however, show long and thick fibres. The fibres are more visible at concentrations below 100 ng/ml. These fibres look different from un-heated fibres in Figure 3.34. Multiple single fibres seem to be aggregating into thicker fibres when heated as opposed to forming ropes (when not heated). A close-up image can be seen in Figure 3.36.

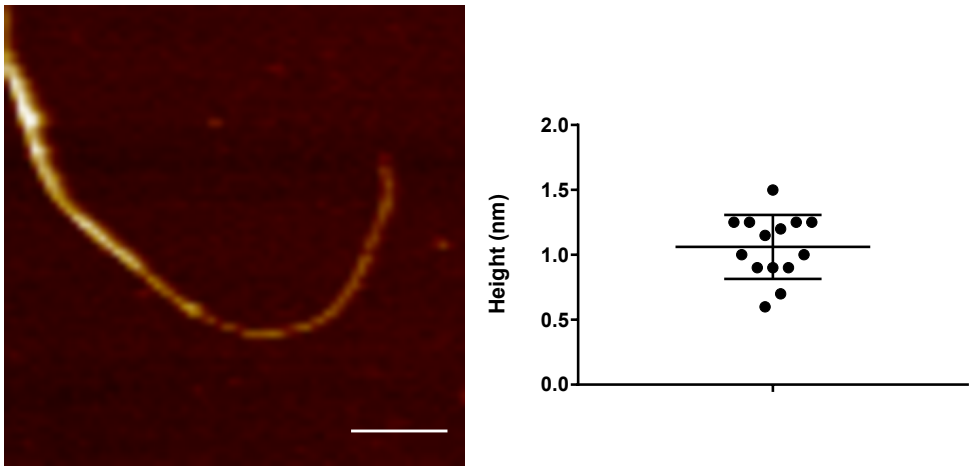


FIGURE 3.36: Dry air tapping mode AFM images of S1x/S2 assembled after thermal treatment at 100 ng/ml. Drying accomplished through passive evaporation at room temperature. Scale bar in 100 nm. Right: a swarm plot of height measurements along the fibre, n=14.

The close-up image shows that the ends of the large fibres are 1 nm in height, corresponding to an expected height for the G52-E-G53 backbone. It is however unclear if the fibre is build up of sub-units interacting end-to-end (like in unheated assemblies) or if there is side-by-side stacking of the sub-unit components. This question becomes more pertinent when assessing G52-E-G53 without terminal helices under the same preparation conditions (Figure 3.37).

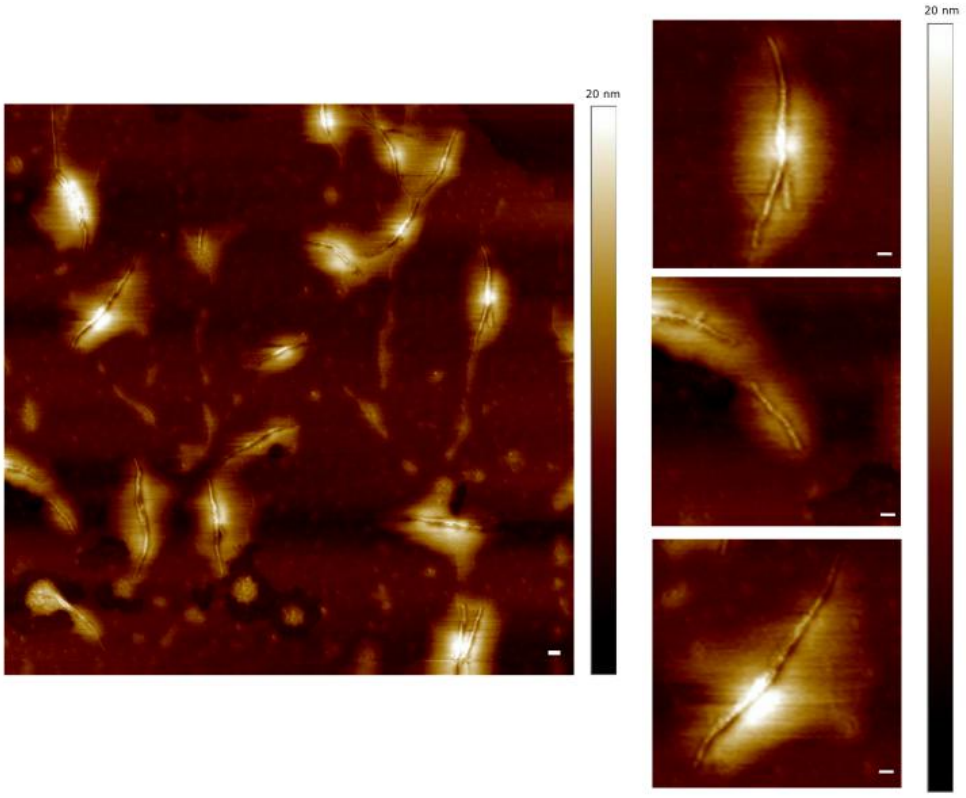


FIGURE 3.37: Dry air tapping mode AFM images of G52-E-G53 assembled after thermal treatment at 100 ng/ml. Drying accomplished through passive evaporation at room temperature. Scale bar is 100 nm. Right: close-up panels of the fibres.

The images reveal that G52-E-G53 also assembles into fibres when thermally treated. The assemblies exhibit a different overall morphology showing shorter fibres that do not adhere to each other. This behaviour is unexpected as G52-E-G53 does not possess terminal α-helices for end-to-end assembly. A closer look at one of the fibres is shown in Figure 3.38.

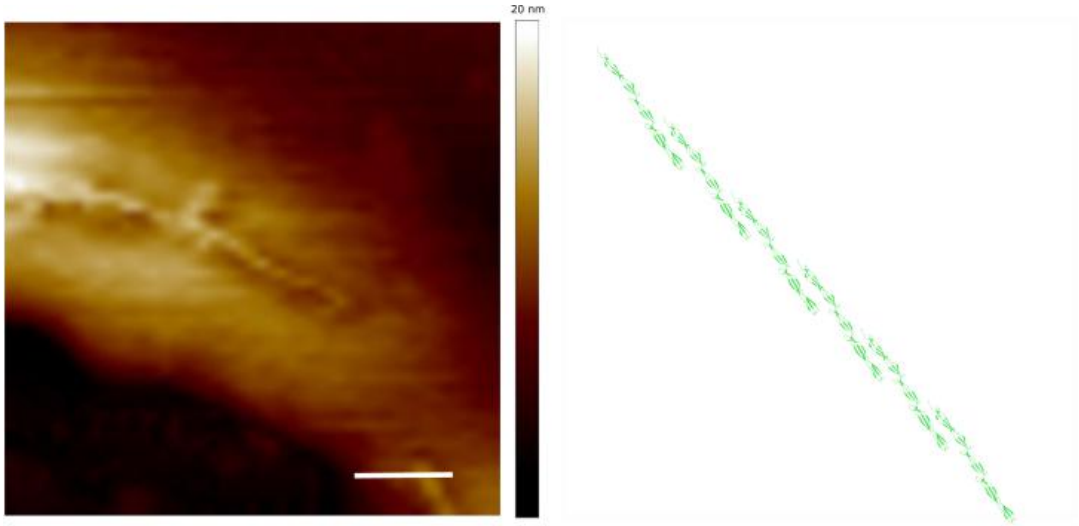


FIGURE 3.38: A model for the mode of G52-E-G53 assembly after thermal treatment. Scale bar - 100 nm.

The G52-E-G53 formed fibre appears to be made up of small uniform individual segments. The segment to segment interaction, however is not end-to-end, but appears to be stacking sideways (as shown on the right of Figure 3.38). This thus led to a conceptualisation of a model for what causes the aggregated fibres of S1x/S2 mixtures in Figure 3.35. The schematic of what may be occurring is drawn in Figure 3.39.

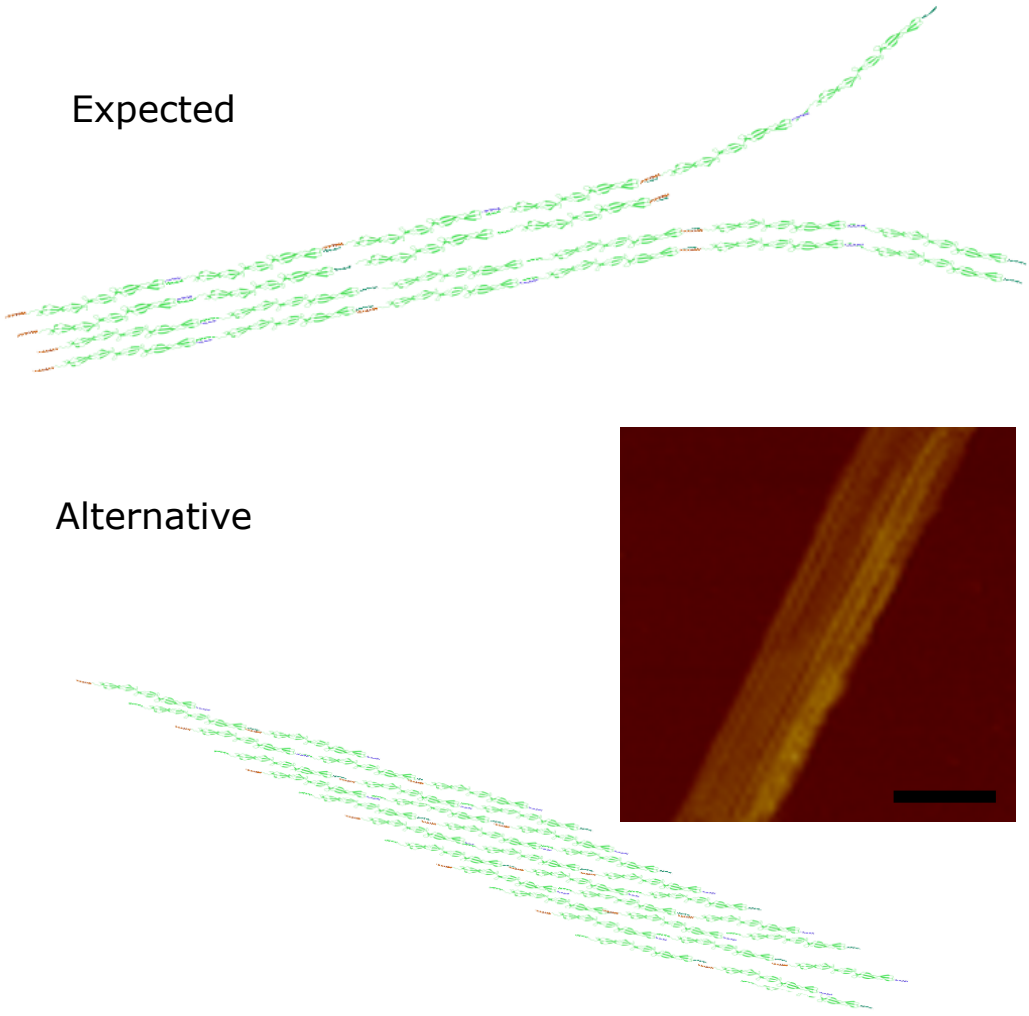


FIGURE 3.39: A conceptualisation of how S1x/S2 mixtures may be forming fibres after a heating protocol. Top: 'expected' assembly where fibres are assembling longitudinally and the aggregating side-by-side. Bottom: alternative assemblies where long chains are formed by G52-E-G53 stacking stabilised by coiled-coil interaction side-by-side. Scale bar - 100 nm.

The figure shows two scenarios for how thick fibres in Figure 3.35 could be assembling. The 'expected' mode of assembly shows the S1x/S2 forming long chains through coiled-coil interactions. The thickness of the fibres is then achieved through singular assemblies adhering side-to-side. Based on the observations from data on heated G52-E-G53 assemblies an alternative model for assembly is proposed. Here, S1x/S2 sub-units stack along the length of the β-sheet G52-E-G53 backbone to indefinite length. The stacked G52-E-G53 fibres are then stabilised by the presence of the

coiled-coil interactions between them. Coiled-coil stabilisation explains why G52-E-G53 backbone composed fibres are shorter and do not aggregate side-by-side. The heated S1x/S2 fibres exhibit a tightly nit internal molecular structure. Same assembly can be observed for mixtures of S1x/S2, S1x/S2x and S1A3/S2A3 (Figure A.10).

This section shows that the S1/S2 fibres exhibit an interesting behaviour, when thermally treated and that coiled-coils may be participating in formation of thick protein fibres. While it may not be relevant for nano-patterning, this may be explored further as a basis for nano-material manufacturing.

3.5.2 S1A3 and S2A3 Interaction with Magnetite

If S1/S2 fibres are to be used as a patterning platform, the ability of the sub-units S1A3 and S2A3 to interact with magnetic nanoparticles needs to be investigated. To assess the capacity of magnetite binding peptide A3 (HNHKSKKHK) to interact with bare iron oxide (magnetite) nanoparticles, an adapted ELISA assay was performed. The assay follows an incubation of the protein of interest with a nanoparticle solution (blocked in e.g. casein or 3 % milk). Any unbound protein is washed off and the presence of bound protein is detected using a specific antibody. The antibody is then detected. The ELISA results can be seen in Figure 3.40.

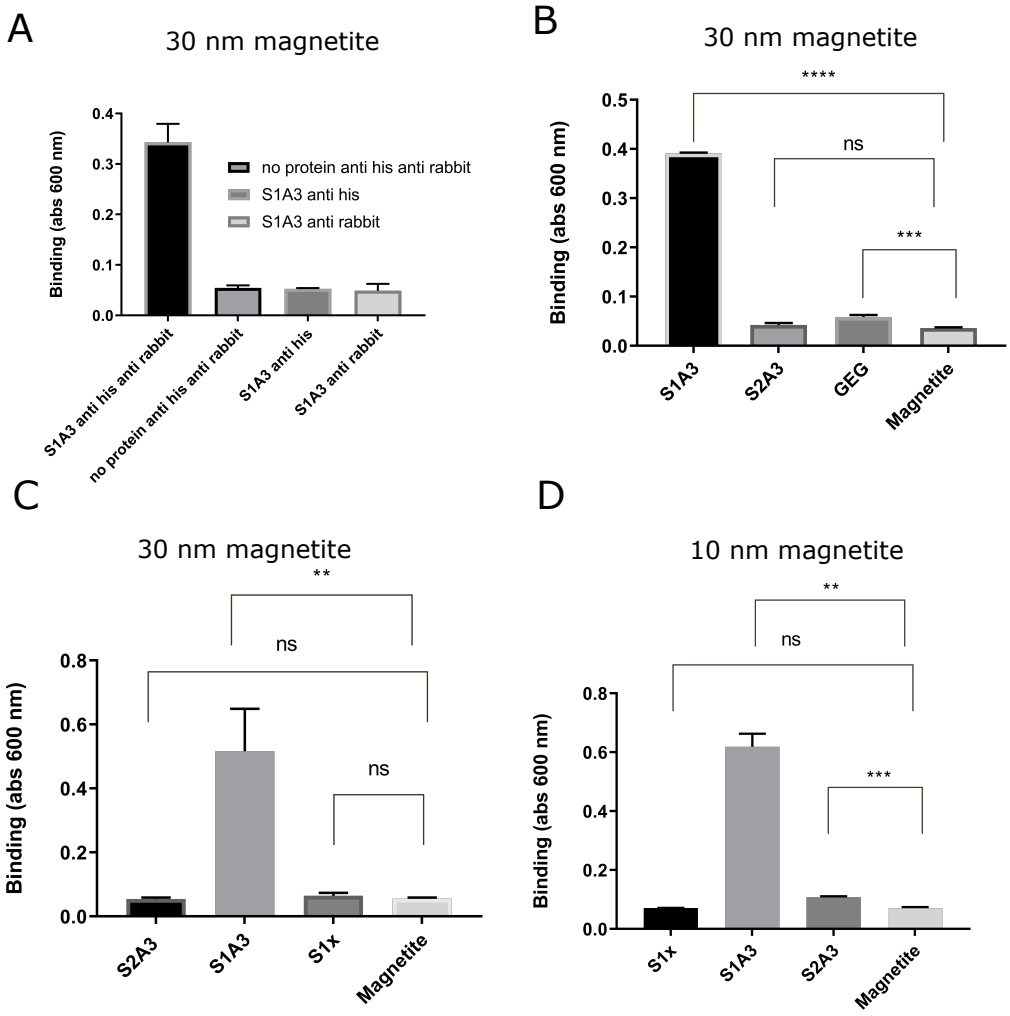


FIGURE 3.40: Magnetite binding ELISA **A** ELISA optimisation experiment with 30 nm magnetite nanoparticles and S1A3. **B** An experiment showing that S2A3 does not bind the particles. **C** An experiment showing that S1x does not bind 30 nm magnetite particles. **D** An experiment showing that S1A3 binds 10 nm particles. Error bars are standard deviations (population), n=3. Blocking conditions performed in casein.

Figure 3.40**A** shows an initial experiment to determine the viability of the assay.

Here, tubes containing S1A3 show a signal (absorbance at 600 nm) after incubation with α-poly-histidine (rabbit) and α-rabbit conjugated to alkaline phosphatase antibodies. Incubating this mixture with a substrate (5-bromo-4-chloro-3-indoyl phosphate) causes a colorimetric reaction. No colour change was detected when the substrate was incubated with tubes that contained only the two antibodies, S1A3 with only α-poly-histidine antibody or S1A3 with only α-rabbit conjugated to alkaline phosphatase antibody. This demonstrates that the experimental design is valid and that neither of the antibodies binds to the particles, meaning that any binding is facilitated by the specific interaction between the particles and S1A3. The same experimental procedure was then used to investigate the binding of S1A3, S2A3, and G52-E-G53 to 30 nm magnetite nanoparticles (Figure 3.40B). The experiment shows that, surprisingly, S2A3 does not bind the particles. What is more, G52-E-G53 has a small yet, statistically significant increase in binding (compared to particles incubated without any protein). The importance of carrying the A3 peptide for particle binding is illustrated Figure 3.40C. Here, S1x shows no binding signal. The same effect of S1A3 binding and S1x and S2A3 no binding, was observed when the experiment was carried out with 10 nm magnetite nanoparticles showing that the binding capacity granted by the A3 peptide is particle size independent.

Taken together these binding results show that addition of a small peptide at a terminal end can impart a nanoparticle binding function to a rod-like protein. The results for S2A3 reveal that the gain of function is not guaranteed though and that a specific proximal amino acid context may be important. Figure 3.41 compares the terminal ends of S1x, S1A3 and S2A3.

```
S1-A3/1-62  EITKDPINELTEYGPETGGKIAALKQKIAALKYKNAALKKKIAALKQGGGSGSHNHKSKKHK
S1x/1-47    EITKDPINELTEYGPETGGKIAALKQKIAALKYKNAALKKKIAALKQ---------------
S2A3/3-63   DPINELTEYGPETGGSPEDENAALEEKIAQLKQKNAALKEEIQALEYGGSGSHNHKSKKHK-
```

FIGURE 3.41: Comparison of amino acid sequence for S1A3, S1x and S2A3 at C-termini.

The closer inspection of the amino acid sequence shows that S1A3 has an additional glycine residue at the linker region (GGSGS), which is longer than S2A3 (GSGS). The discrepancy came about because the terminal glycine was present as

part of the design of the BN4 (found at the C-terminus of S1A3) and was not included for P6 (found at the C-terminus of S2A3). This difference may grant the additional flexibility S1A3 terminus compared to S2A3 causing the difference in binding ability. The P6 α-helix N-terminus also carries significant local negative charge in the glutamate residues, whereas BN4 α-helix N-terminus is positively charged due to lysine residues being present. It may be that the highly positively charged peptide A3 (HNHKSKKHK) is being attracted to the negative charges in S2A3 preventing the peptide from binding the nanoparticles or that the positive charges at S1A3 N-terminus are repelling the A3 peptide making it extended and available for binding. In either case, care must be taken when designing nanoparticle binding proteins in the future.

3.6 Chapter 3: Conclusion

This chapter characterises SasG repeat G52-E-G53 fused with coiled-coils as an addition to the protein engineering tool-kit. The coiled-coil interaction is shown to be specific making the system useful in targeted assembly applications and the sub-units are shown to assemble into long fibres. The assembly process is tolerant of terminal modifications with short (<10 amino acid) peptides which can also be shown to be effective in binding magnetite nanoparticles. For the system to be used as means to achieving controlled nanoparticle patterning on surfaces a lot of work remains ahead. While there does not seem to be any issue in obtaining long chains of S1 and S2 sub-units, there is a need to prevent the fibres from curling or aggregating. Perhaps a surface immobilisation assisted strategy may help to control the protein assembly. Furthermore, particle binding determinants must be investigated more deeply.

Chapter 4

Results: Computationally designed components for designed 2D SasG arrays

4.1 Chapter Introduction

Chapter 4 builds on Chapter 3 with attempts to construct SasG G52-E-G53 based two-dimensional protein arrays. Two sets of hub proteins are introduced to direct coiled-coil based SasG assembly to assume a two-dimensional pattern. The chapter starts with an introduction to collaborator's work (Fabio Parmeggiani) on computational design of cyclic oligomeric proteins to act as hubs. The design of two new sets of cyclic proteins compatible with the SasG coiled-coil system is then described (Design work by Fabio Parmeggiani). The production of these proteins is demonstrated together with validation of their interaction with S2 (one of the SasG-coiled-coil system components).

4.1.1 A Method for Computational Design of Oligomeric Protein Assemblies

Fallas et. al., 2017 [254] describe a computational method to create cyclical oligomeric protein assemblies with symmetries of C2, C3, C4. The method uses naturally occurring proteins and varies the amino acid residues on the surface to achieve a geometry of interaction of the desired oligomerisation state (dimer ,trimer, tetramer) and symmetry. The self-assembly of the designed oligomers is driven by chemical and

shape complementary. The designs were created by, first, docking low resolution rigid body structures followed by a full atom side chain optimisation. The study demonstrated successful assembly of 21 out 96 computationally designed proteins.

4.1.2 Modified Oligomeric Assembly Designs

A collaborator, Fabio Parmeggiani (part of the initial study) used the Fallas et al., 2017 [254] findings to design two sets of components compatible with the SasG coiled-coil system (Chapter 3). Two oligomer backbones were used: Ank4 which was modelled after ankyrin repeat proteins and was shown to form a tetrameric assembly and HR00C3-2 (from here on referred to as 'HR') which is a completely *de novo* designed repeat protein (structure not confirmed experimentally) shown to assemble into trimers (oligomerisation states validated by size exclusion chromatography [254]). The goal of the strategy is to design interfaces between SasG G52-E-G53 domains for assembly into two-dimensional patterns via protein hubs capable of magnetic nanoparticle binding. In this approach, the coiled-coil pairs P5/P6 and AN4/BN4 characterised in Chapter 3 are used. Sub-unit S2 (AN4-G52-E-G53-P5) is used as originally designed, however the coiled-coil α-helices from sub-unit S1 (P6 and BN4) are separated and attached to the structures of two sets of HR or Ank4 oligomeric assemblies. The expected behaviours for each of the sets of hub proteins is demonstrated in Figure 4.1.

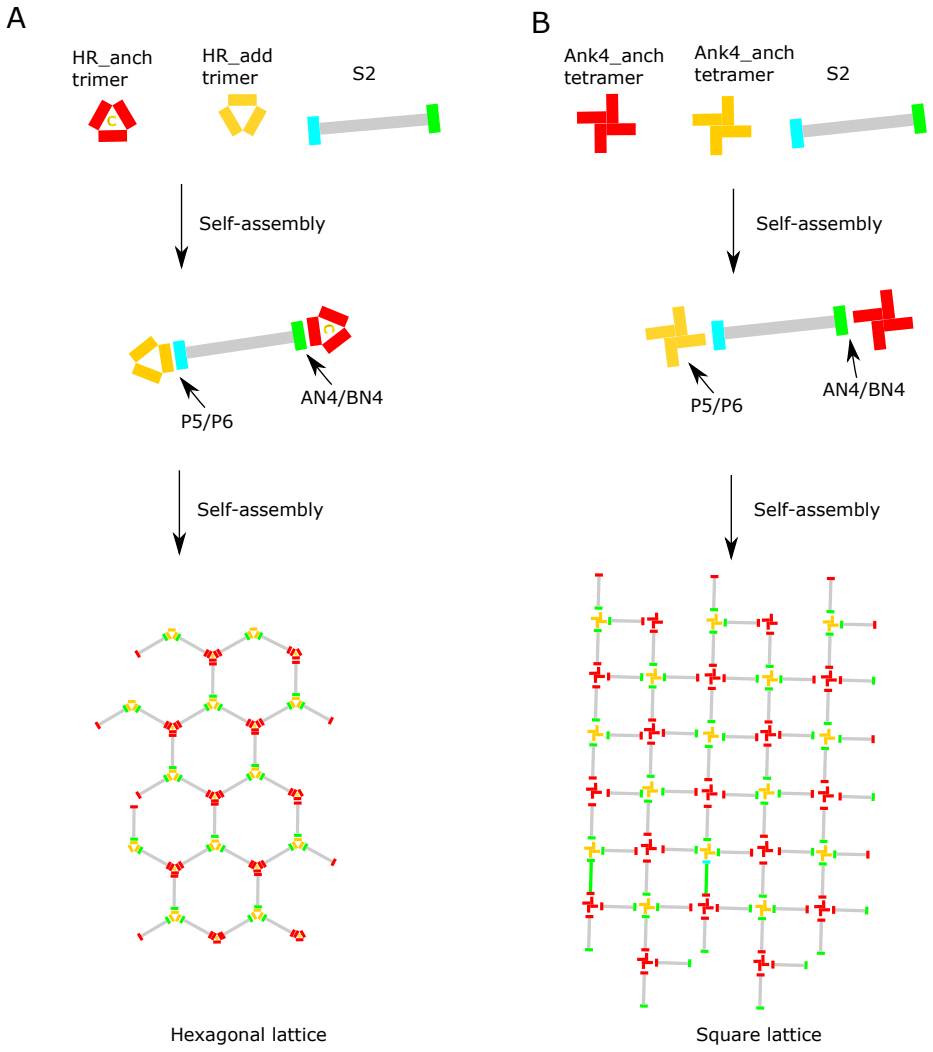


FIGURE 4.1: Diagrammatic representations of possible lattices constructed using a combination of hub proteins from [254] and the SasG coiled-coil sub-unit S2. **A** The expected behaviour and resulting lattice upon mixture of S2 with HR-add and HR-anchor - formation of hexagonal lattices. **B** The expected behaviour and resulting lattice upon mixture of S2 with Ank4-add and Ank4-anchor - formation of hexagonal lattices. (Differences between '-add' and '-anchor' variants are explored below).

The expectation is that each of the modified hub protein sets will retain their ability to form symmetrical trimers and tetramers when using HR and Ank4, respectively. Figure 4.1**A** shows how S2 interacting with HR-anchor and HR-add would form hexagonal lattices. S2 would form coiled-coil interactions with both hub assemblies. Each of the hub assemblies would have three contact points with S2. The cumulative outcome of these interactions can produce hexagonal lattices. Such lattice assembly would rely on the flexible glycine-glycine linker to accommodate the necessary angles. Figure 4.1**B** shows the expected interactions between S2 and Ank4 hub proteins. Again, S2 would form coiled-coil interactions with both hub assemblies. Each Ank4-add and Ank4-anchor tetramer, however, would form four interactions with S2. Eventually, such interaction pattern would result in a square lattice.

The design layout of Ank4 based assemblies can be seen in Figure 4.2. Figure 4.2**A** shows the overall expected tetrameric assembly of the protein. Coloured locations within the structures represent where terminally fused A3 peptide is expected to be presented (red) and the coiled-coil forming α-helix fusion site (green). Yellow signifies where a cysteine residue would be presented in the final fold. Ank4-add is the variant of the protein containing a C-terminal P6 α-helix, whereas Ank4-anchor contains a C-terminal BN4 α-helix, the sequence comparison can be seen in Figure 4.2**B**. Both constructs contain a C-terminal poly-histidine (6) tag for purification. Ank4-anchor also contains an amino acid replacement E166C. The cysteine can be seen to be surface accessible in the trRosetta models presented in Figure 4.2**C**. The rationale behind the cysteine substitution is to have one hub act as a surface anchoring hub.

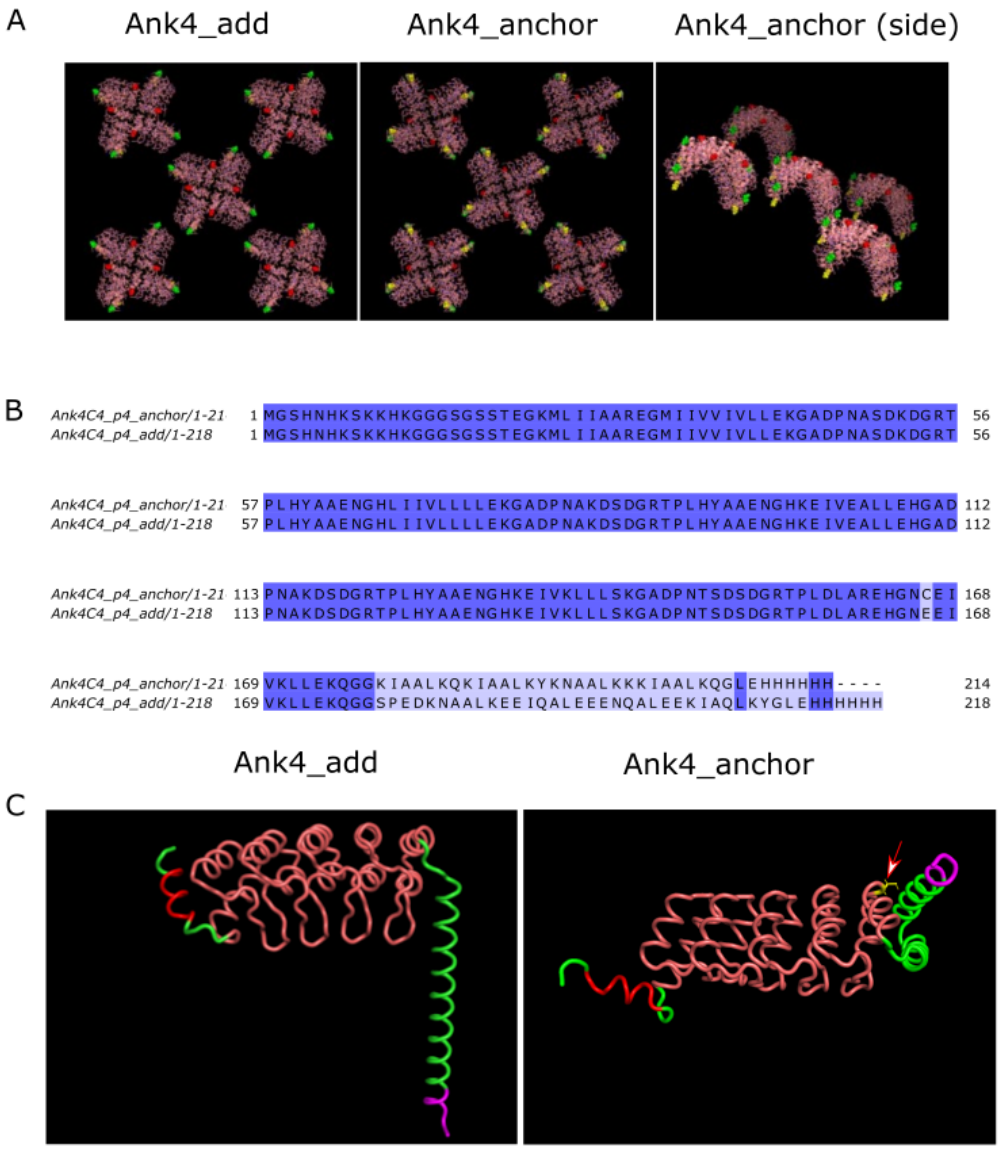


FIGURE 4.2: **A** Low resolution structures of Ank4-add and Ank4-anchor top down (left to middle) and sideways (right). **B** Amino acid sequence comparison of Ank4-add and Ank4-anchor. **C** trRosetta predicted structures of Ank4-add and Ank4-anchor. Colour scheme: red - magnetite binding peptide A3, salmon - Ank4 protein, green - P6 and BN4 α-helix for Ank4-add and Ank4-anchor, respectively, pink - poly-histidine tag.

An equivalent design approach was applied to the trimeric HR self assembling protein (Figure 4.3). The design, again, fuses P6 and BN4 coiled-coil forming α-helices to the HR backbone to create two distinct hub structures. Figure 4.3**A** shows low resolution structures of expected trimeric assemblies of the HR variants. Figure 4.3**A** shows amino acid sequences of the two variants. HR-add contains an N-terminal magnetite binding peptide A3, the P6 α-helix and a C-terminal poly-histidine (6) tag. HR-anchor contains the same components, except for the BN4 α-helix instead of P6. HR-anchor also contains a cysteine residue for surface immobilisation. It must be noted that with this design HR-add or Ank4-add would be exposing the P6 α-helix at the C-terminus. S2 exposes the P5 (binding partner of P6) at the C-terminus as well. This may present an issue for the formation of the coiled-coil as the pair is designed to form a parallel dimer. There is, however, sufficient flexibility within both proteins (granted by the glycine linkers) to accommodate an arrangement for the formation of a parallel P5/P6 coiled-coil.

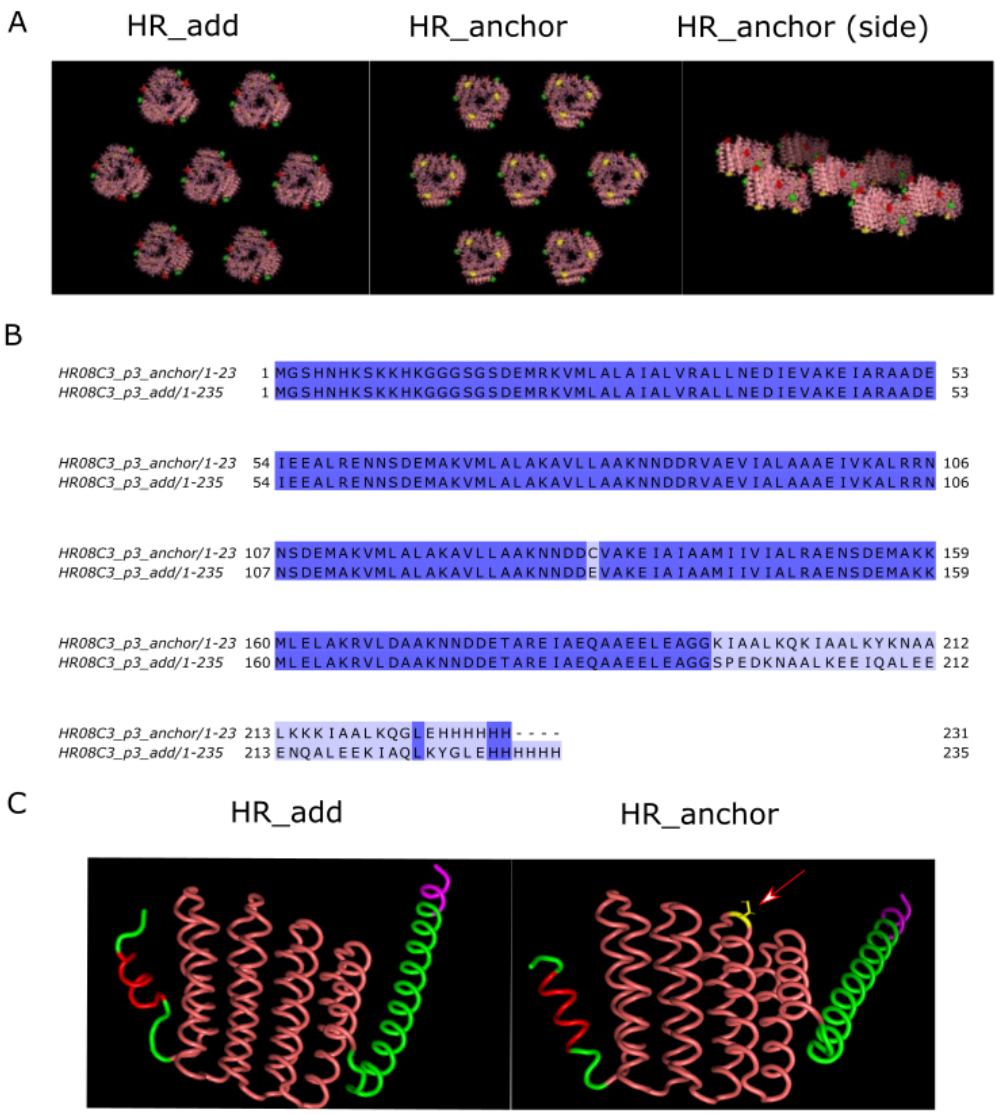


FIGURE 4.3: **A** Low resolution structures of HR-add and HR-anchor top-down (left to middle) and sideways (right). **B** Amino acid sequence comparison of HR-add and HR-anchor. **C** trRosetta predicted structures of HR-add and HR-anchor. Color scheme: red - magnetite binding peptide A3, salmon - HR protein backbone, green - P6 and BN4 α-helix for HR-add and HR-anchor, respectively, pink - poly-histidine tag.

During the design process both HR and Ank4 components were assessed for their ability to retain their self-assembly capabilities (using the same computational tools as demonstrated in Fallas et al., 2017 [254]*) while accommodating for the modifications: A3 peptide, either P6 or BN4 and the terminal poly-histidine tag. Both sets of sequences contain flexible glycine-glycine linker between the self-assembling part of the protein and the fused coiled-coil forming α-helical region.

Dynamic modelling performed by Fabio Parmeggiani

To avoid the aggregation of the potential protein assemblies in solution, a surface assembly strategy was envisaged (Figure 4.4).

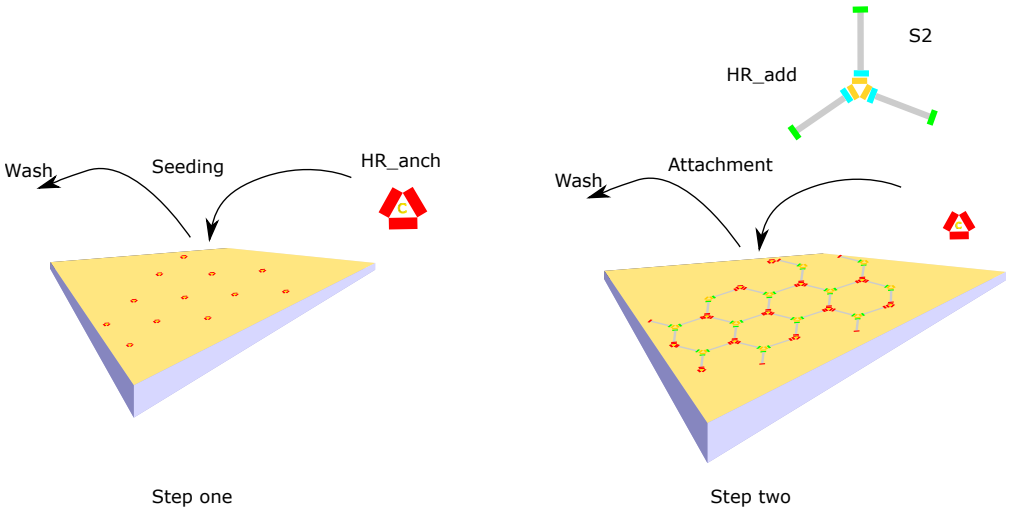


FIGURE 4.4: Strategy to sequentially assemble HR-anch lattices on surfaces. The area in yellow is meant to represent a gold surface.

First, a low concentration solution of HR-anchor would be incubated with a gold surface to seed the surface with points for attachment. Then, HR-add pre-assembled with S2 would be introduced together with additional HR-anchor. This assembly procedure would be repeated until a full lattice is assembled. Same strategy would be followed for Ank4 variants. The lattice would be expected to be 1-3 nm in height with local height peaks corresponding to the locations of the hubs (as can be predicted from comparing the protein structures of the component parts (Figure 4.5).

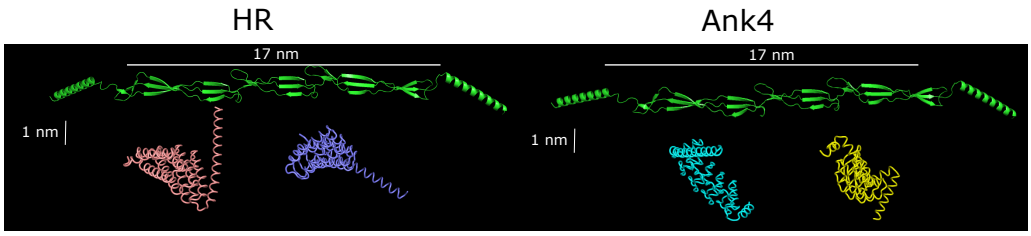


FIGURE 4.5: Size comparisons between SasG G52-E-G53 domain sub-unit S2 and the hub assembling proteins (HR-add, HR-anchor, Ank4-add, Ank4-anchor).

Finally, the assembled lattices of either of the sets of proteins are expected to specifically immobilise magnetite (the design is expected to be modular, however,

allowing for exchanging the particle binding peptide for binding other materials) at the hub locations. Should this strategy succeed, nanoparticle patterns with 17-30 nm periodicity could be achieved (measured based on the length of the G52-E-G53 domain). This chapter catalogues the progress made towards that goal during this PhD.

4.2 Results: Ank4

4.2.1 Purification

Ank4-add and Ank4-anchor DNA sequences were purchased as synthetic genes and were sub-cloned into pET28a plasmids for over-expression in *E. coli*. The expected parameters of the produced proteins are listed in Table 4.1.

Construct	Expected Mass (no M)	Expected Mass	pI
Ank4-add	23467.93 Da	23599.12 Da	5.76
Ank4-anchor	22965.96 Da	22834.77 Da	8.78

TABLE 4.1: Ank4-add and Ank4-anchor expected protein characteristics. M - methionine. Generated using ProtParam (Expasy).

The differences in the isoelectric points are caused by the presence of a glutamate residue rich P6 α-helix in Ank4-add and a lysine rich BN4 in Ank4-anchor. The standard protein over-expression protocol was carried out and protein purification was attempted (Figure 4.6).

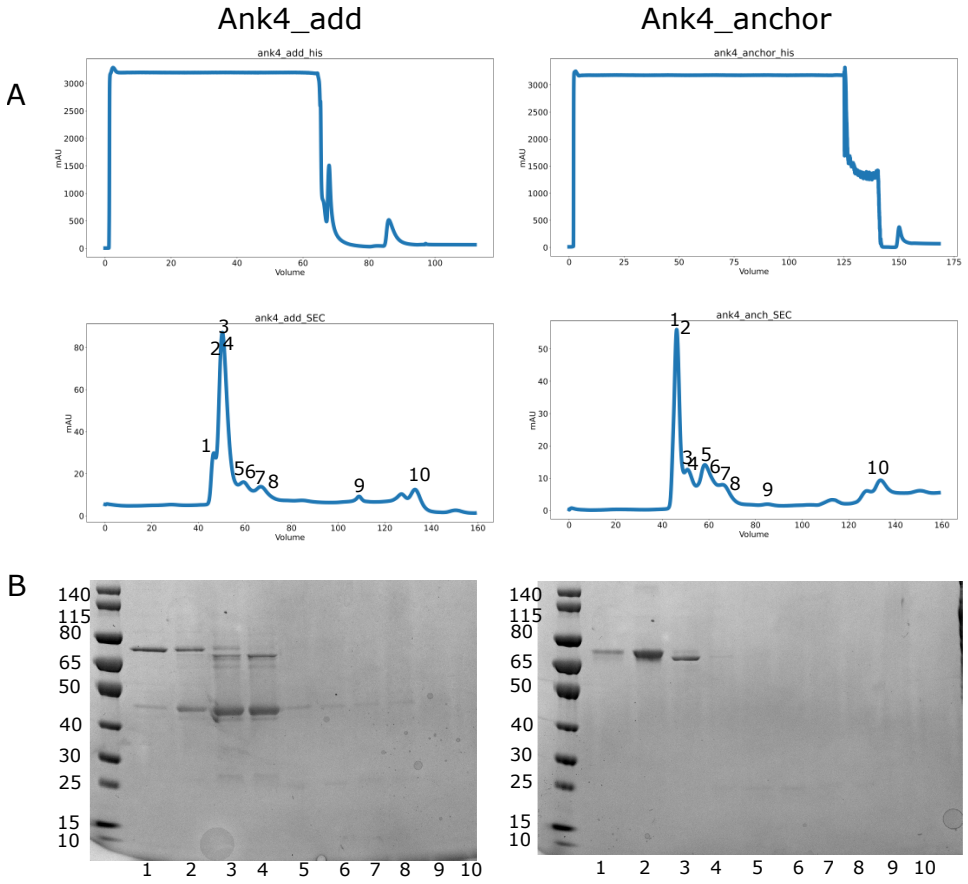


FIGURE 4.6: Purification of Ank4 variants **A** Top: His-Trap column purification performed on the Äkta pure system for Ank4-add and Ank4-anchor. High concentration (300 mM) imidazole was introduced at elution volume of 85 and 145ml, respectively. Bottom: size exclusion chromatography performed on a HiLoad 16/600 Superdex 75 pg column for Ank4-add and Ank4-anchor. Size exclusion chromatography was performed in PBS pH 7.4. 1 mM DTT was included for Ank-anchor samples. X axes - elution volume, Y axes - mAU. **B** SDS-PAGE analysis of elution fractions corresponding to the labelled peaks in **A**.

The Ni-NTA pre-packed column purification showed discrete elution peaks after application of high molarity imidazole solution implying that the constructs containing the poly-histidine tag were successfully expressed. Size exclusion chromatography was chosen as a purification refinement step. In both cases chromatograms with multiple peaks were produced and were analysed on SDS-PAGE individually (Figure 4.6**B**). For Ank4-add, peaks labelled '1' and '2' had SDS-PAGE bands with mobility of around 47 and 72 kDa, which are a close match for theoretical molecular weights of dimers (51 kDa) and trimers (76 kDa), '3' and '4' display the same bands with faint bands running at monomer size (around 25 kDa) and some contaminants, peaks labelled '5','6','7','8' all show faint bands at 47 kDa and at 25 kDa. Ank4-anchor analysis shows bands at 72 kDa for peaks '1','2','3' and faint monomer bands for peaks '5','6','7','8' (peak '4' seems to be a faint mixture of both sizes). The fractions producing similar SDS-PAGE banding patterns were pooled and analysed on ESI-TOF to confirm the exact molecular mass of the purified protein fractions (Figure A.12).

The information about the most prominent ESI-TOF peaks is tabulated in Table A.2. All Ank4-add samples revealed a distribution of sizes corresponding to exact size of monomers - 23469 Da, dimers - 46937 Da and trimers - 70404 Da (sizes correspond to expected sizes without the N-terminal methionine). The intensity of the signal progressively decreases for larger species. There was also a 23715 Da species at comparable intensities to those of the the monomer (at least 50 %) in all Ank4-add samples. Ank4-anchor fractions show a more diverse distribution of mass spectra peaks. None of the peaks matches the exact expected size of monomers (22965 Da), dimers (45931 Da), trimers (68897 Da) or tetramers (91863 Da). However, fraction 5-6 spectrum shows a 23552 Da species (587 Da larger than expected for Ank4-anchor monomer) for which there are also peaks corresponding to dimers (47102 Da) and trimers (70655 Da) on the same spectrum.

The fractions were analysed using a Western blot to confirm the identity of the proteins by probing for a poly-histidine tag (Figure 4.7).

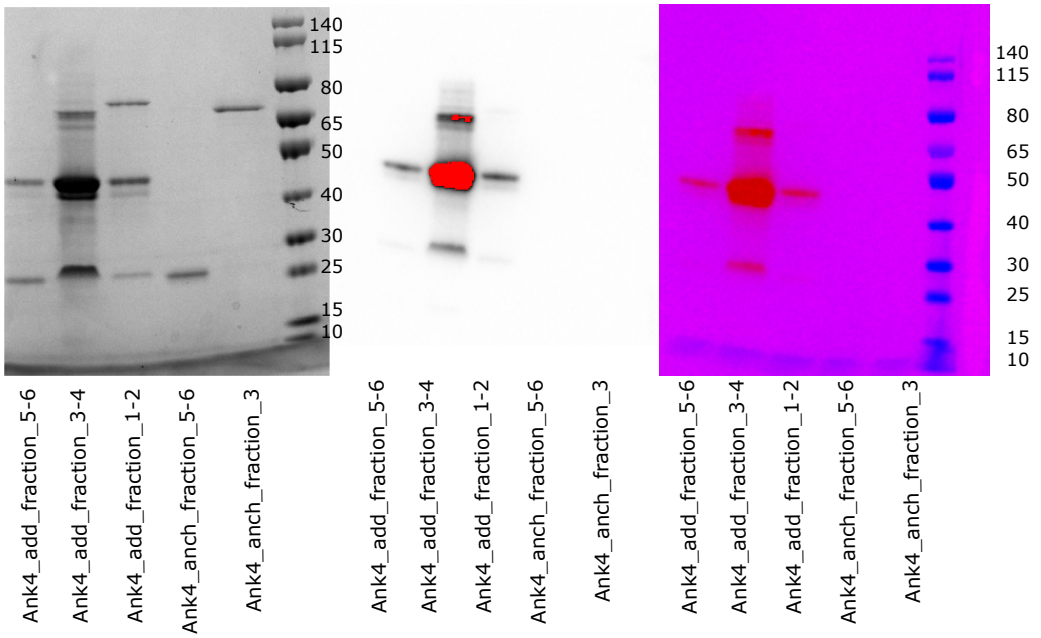


FIGURE 4.7: Western blot analysis of Ank4-add and Ank4-anchor fractions obtained from pooling fractions from Figure 4.6. Left: SDS-PAGE gel with separated fractions. Middle: a Western blot run with the same samples as the SDS-page gel probed for a poly-histidine tag. Right: overlay image of the Western blot with a protein marker.

The Western blot analysis shows Ank4-add being detected by the poly-histidine specific antibody at various apparent oligomerisation states. The Western blot shows positive signal for bands corresponding to 23 kDa, 46, kDa and 70 kDa corresponding to Ank4-add monomers, dimers and trimers. Surprisingly neither of Ank4-anchor fractions showed any signal. All of the purified fractions were also analysed using CD (Figure 4.8).

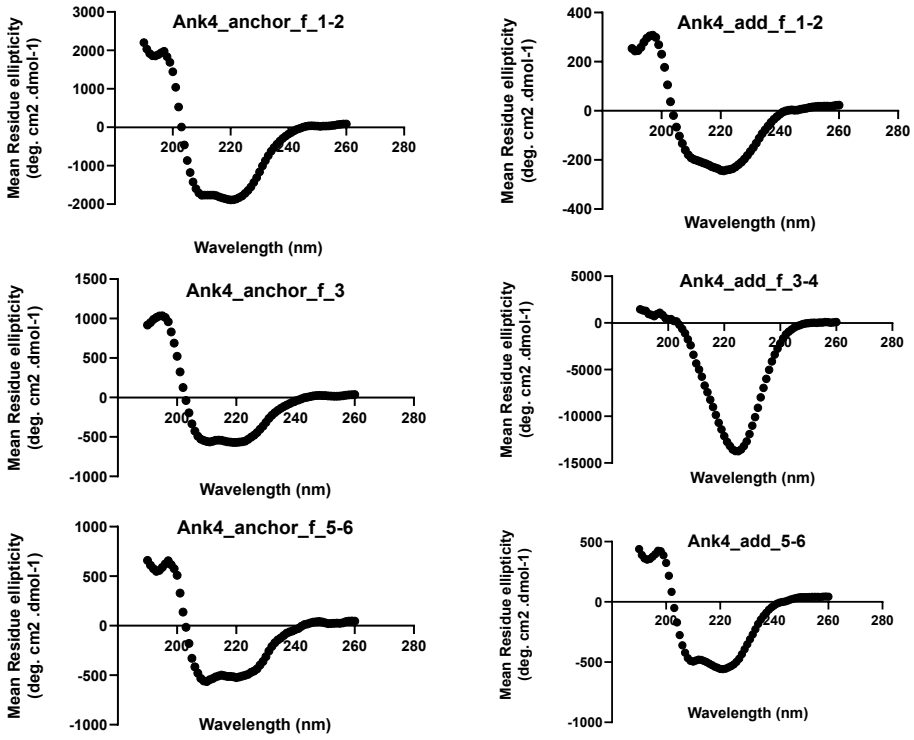


FIGURE 4.8: CD spectra of Ank4-add and Ank4-anchor pooled fractions from Figure 4.6. CD spectra performed in PBS pH 7.4. Proteins analysed at 2.5 μM concentrations.

The spectra show all of the fractions (except for Ank4-add fractions 3-4) to be folded into α-helical structures with peak minima at 220 and 205 nm bands and a peak maximum at 189 nm. These values have also been reported for ankyrin proteins before [250]. Ank4-add fractions 3-4 show a singular peak minimum at 222 nm. Similar spectrum has been shown to correspond to protein aggregates [251].

Together, these data show that Ank4-add is produced at the exact expected size and folds into an α-helical protein that appears to assemble into oligomers. The self-assembly behaviour can be inferred from incomplete denaturation of the proteins in SDS-PAGE analysis (the protein to protein interactions are stronger than denaturant to protein interactions) and from remnants of dimeric and trimeric species in ESI-TOF experiments (ESI-TOF is known for allowing detection of non-covalently interacting protein assemblies [252]). Interestingly, no tetramer sized bands or mass spectrometry peaks were detected. Dimeric peaks in ESI-TOF were at 5 % of the monomeric peak and trimeric peaks were only slightly more intense than background noise, implying that large assemblies are progressively less likely to be observed using this technique, thus meaning that tetramers are too bulky or unstable to reach the detector. Fractions 5-6 were chosen to be used during subsequent purifications for any downstream experiments as they showed good folding (CD) and appeared to be the purest. Ank4-anchor results were less clear. The lack Western of blot signal implies that there may be some structural changes that occur in this version of the hub protein, not only that, but ESI-TOF shows an off-set in molecular weight. Nevertheless, the detection of species two and three times larger than the monomer (though of the wrong size) implies that there may be similar oligomerisation taking place as for Ank4-anchor. Subsequent purifications for downstream experiments used fractions equivalent to 5-6 for the Ank4-anchor protein.

4.2.2 CD investigation: Ank4-add thermal stability and interactivity with S2

Ank4-add was analysed for its ability to re-fold after thermal treatment (Figure 4.9).

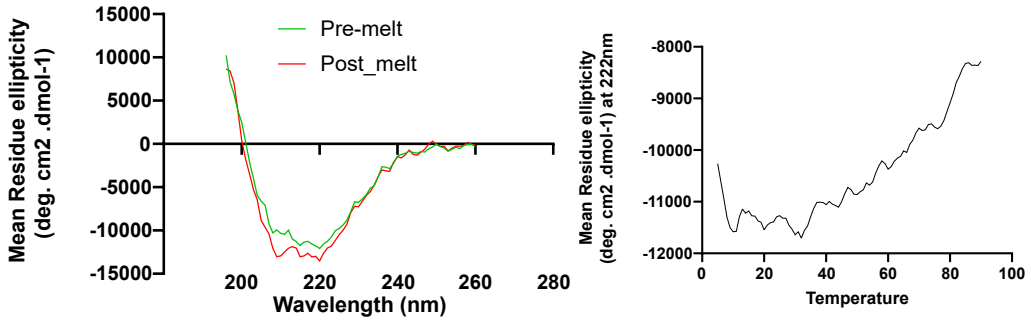


FIGURE 4.9: Left: CD spectra of Ank4-add before and after a thermal denaturation experiment (heating to 90 °C). Right: thermal denaturation curve of Ank4-add. Experiments performed in TBS pH 7.4. Proteins analysed at 2.5 μM concentrations.

The experiment shows that Ank4-add re-folds to its original state after thermal treatment. Unfortunately, a melting temperature could not be reliably determined from this experiment. With the aim of assessing whether or not there was an interaction, Ank4-add was also mixed with S1x and S2x and subjected to thermal denaturation experiments using CD (Figure 4.10).

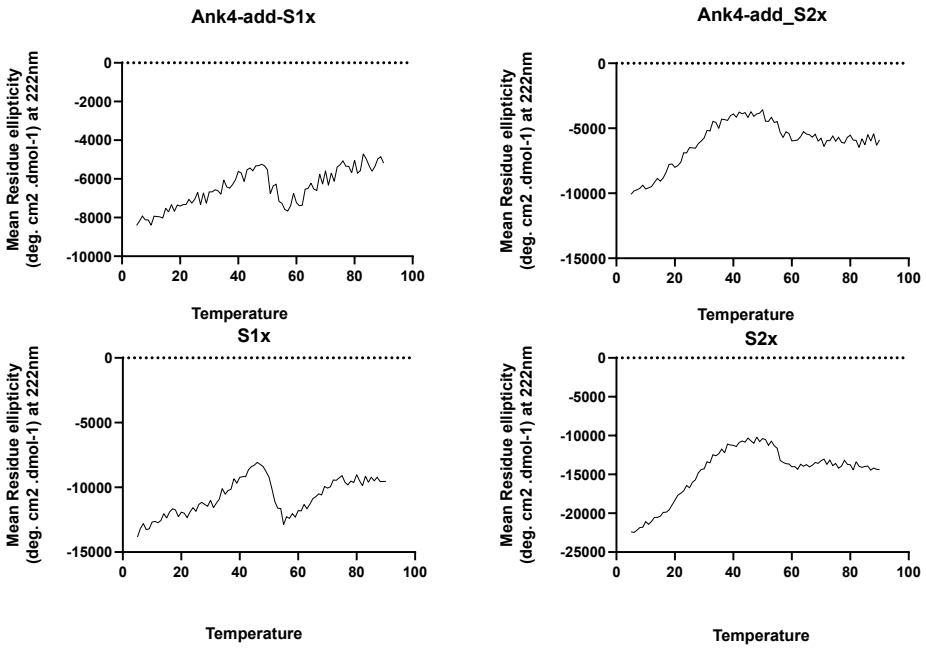


FIGURE 4.10: Top: Thermal denaturation curves of Ank4-add mixed with either S1x or S2X. Bottom: thermal denaturation curves of S1x and S2x. Experiments performed in TBS pH 7.4. Proteins analysed at 2.5 μM concentrations.

No change in shape of the thermal denaturation curves (and thus stability) could

be inferred. The premise of the experiment was to see if the presence of Ank4-add would increase the thermal stability of S2x preferentially by forming the designed P5/P6 coiled-coil. No interaction is expected with S1x, thus it would not change the thermal stability. This negative result may not necessarily mean that there is no interaction between S2x and Ank4-add, but rather that the experimental set-up needs optimisation. This type of an experiment was used in Chapter 3, with a two component system (S1 and S2) where two components were of the same molecular weight and almost identical structural composition making the experimental design less complex.

4.2.3 LSPR Specificity Experiments

Localised surface plasmon resonance experiments were carried out to investigate the ability of Ank4-add and Ank4-anchor constructs to interact with S2x. The technique allows for detection of accumulated biological material on the surface (probed by a change in optical properties). The goal was to coat a gold coated LSPR surface with either Ank4-add and Ank4-anchor and after a buffer wash to introduce S2x into the flow system. The experiment with immobilised Ank4-anchor can be seen in Figure 4.11.

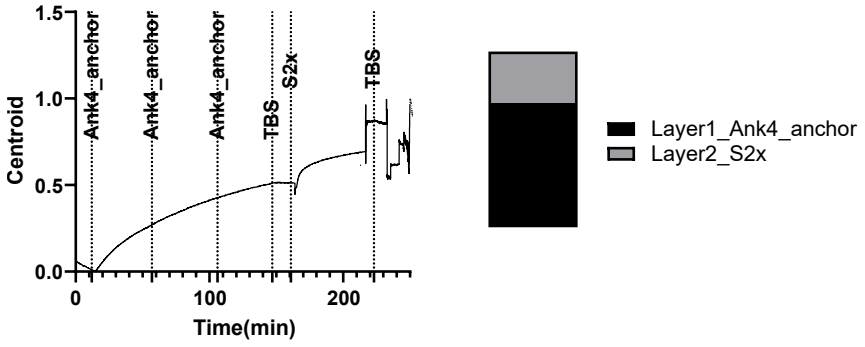


FIGURE 4.11: LSPR experiment with Ank4-anchor and S2x. Left: plotted Xnano LSPR experiment (Y: algorithmically adjusted changes in the extinction peak - centroid, X: time). Centroid data were normalised to the highest value. Experiments were carried out at 50 μl/min at 26 °C. Specific time-points where a new solution is introduced are designated with vertical dotted lines. Each protein was used at 2.5 μM in TBS pH 7.4. The initial equilibration was performed in TBS pH 7.4. Right: comparison of the amount of material estimated for each layer of the added proteins after washing (measured at the lowest point after washing for the fist layer and the highest point for S2x as the information after washing was not available).

The immobilisation of Ank4-anchor on the gold surface was expected to be facilitated by a combination of the presence of cysteine residues forming gold-thiol bonds and non-specific protein adsorption. The experiments were started with equilibration in TBS buffer pH 7.4 and an attempt to saturate the surface by introducing Ank4-anchor (2.5 μM). The three distinct points designating addition of Ank4-anchor correspond to 'topping-up' the source sample tube with additional protein (the time to saturation was longer than initially expected). Once near-saturation was achieved, the layer was washed with TBS pH 7.4. A flow of S2x (2.5 μM) was then introduced and saw a sharp increase in signal and eventual saturation. The system became unstable before the final wash. At which point a final wash attempted and the experiment was stopped. The instability may have been caused by presence of bubbles, ambient dust or the aggregation of the protein sample in the source tube. Measuring the amount of material at the highest saturation point at around 210 minute timepoint there seems to be a 72 % : 28 % division between the Ank4-anchor base layer and S2x layer. An equivalent set of experiments was performed with Ank4-add (Figure 4.12).

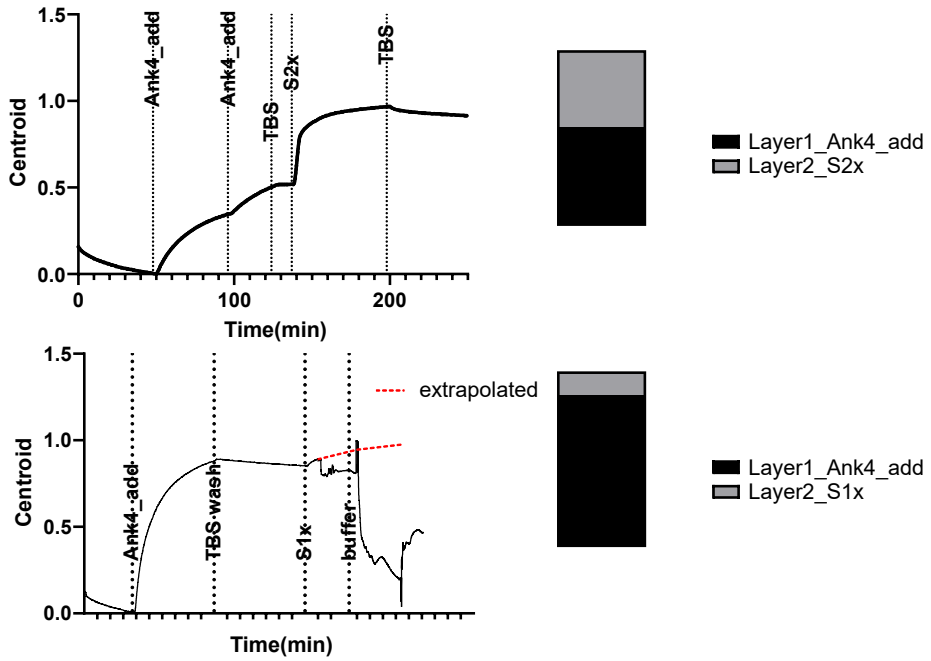


FIGURE 4.12: LSPR experiment with either S2x (top) or S1x (bottom) and Ank4-add. Left: plotted Xnano LSPR experiments (Y: algorithmically adjusted changes in the extinction peak - centroid, X - time). Centroid data were normalised to the highest value. Experiments were carried out at 50 μl/min at 26 °C. Specific time-points where a new solution is introduced are designated with vertical dotted lines. Each protein was used at 2.5 μM in TBS pH 7.4. The initial equilibration was performed in TBS pH 7.4. The puncture lines in red (bottom plot) represent an extrapolation of the expected curve. Right: comparison of the amount of material estimated for each layer of the added proteins after washing. S1x layer was estimated from the highest point of the extrapolated part of the plot.

Here, the immobilisation of Ank4-add relies solely on the non-specific interaction between the protein and the gold surface. The first experiment shows an initial attempt to saturate the surface (and an additional 'top-up' with Ank4-add midway, the apparent difference in kinetics may be explained by the fact the concentration of the second Ank4-add solution was underestimated). After a TBS pH 7.4 wash, S2x was introduced into the system (2.5 μM). There was a sharp increase in material on surface reaching complete saturation. The accumulated layers were then washed with TBS pH 7.4 and most of the accumulated material was retained. The protein coverage (based on the signal values after washing) was 56 % Ank4-add and 44 % S2x. Assuming the amount of material of each protein is directly proportional to the accumulated signal under these experimental conditions, it can be calculated (the

calculation takes into account the molecular weight of each protein to calculate molar ratios based on the percentage of respective protein material on surface) that every tetrameric Ank4-add hub would interact with 2.4 molecules of S2x. The incomplete coverage may be a molecular crowding/concentration based limitation or may be due to the fact that not all interaction sites (P6 α-helices) are accessible. That may arise due to the proteins within the saturated layer sterically blocking each other. It is known that poly-histidine tags have affinity for gold surfaces [253]. If Ank4-add proteins are applied to the system in an oligomeric state, it may then mean that one side of each oligomer is preferentially oriented to face the surface (blocking access to P6 α-helices) while the other side is exposed to the solution. This results in a lower amount of immobilisation sites for S2x than theoretically possible. An analogous experiment with S1x instead of S2x was also performed as a control (Figure 4.12 (bottom)). The surfaces were saturated with Ank4-add and washed with TBS pH 7.4 before application of S1x (2.5 μM). The change to the protein solution did not cause a rapid increase in mass as observed in the experiment with S2x. Unfortunately, the system became unstable before full saturation could be reached. An extrapolation of the curve was plotted to nevertheless measure and compare the amount of Ank4-add and S1x material on the surface (which was measured from the lowest point after washing with buffer for Ank4 and the highest point of the extrapolation for S1x). The protein coverage was 87 % for Ank4-add and 13 % for S1x. Applying the same calculation as above it appears that there are 0.4 S1x molecules for every Ank4-add tetramer. This lower binding strongly implies that Ank4-add surface immobilised protein layers have a specific preference for S2x. This is most likely enabled by the specific P5/P6 coiled-coiled interactions between Ank4-add and S2x. It is interesting that Ank4-anchor surface showed a lower amount of binding (calculated 1.14 molecules of S2x for every Ank4-anchor tetramer) than Ank4-add. The difference in values may be explained by how the cysteine containing Ank4-anchor hubs are deposited on the gold surface. The coiled-coil forming α-helix BN4 may also behave differently to P6 in this context. Still, there seem to be more interactions than the negative control experiment, suggesting that there is specificity to the Ank4-anchor and S2x interaction. Taken together these LSPR results reveals that both Ank4-add and Ank4-anchor show tendency to specifically interact with their designed binding

partner S2x.

4.2.4 Ank4 Microscopy

Attempts were made to obtain visual proof of Ank4-add and Ank4-anchor assemblies with S2. Uranyl formate stained TEM was chosen for the initial investigations. A total protein mixture (1:1:2 for Ank4-add:Ank4-anchor:S2) concentration of 500 μg/ml was prepared in PBS pH 7.4 and incubated at room temperature for 72 hours after which the samples were applied to carbon coated copper grids and stained with uranyl formate (Figure 4.13).

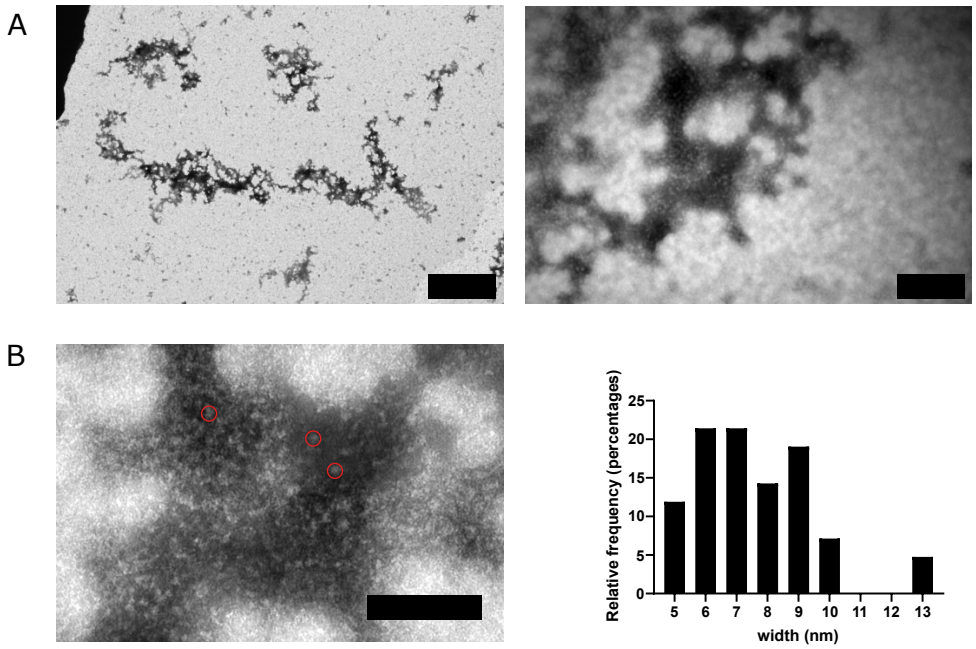


FIGURE 4.13: TEM of Ank4-add, Ank4-anchor and S2 attempted assemblies. **A** Overview images of the objects observed. Scale bars - 2000 nm and 200 nm (left to right). **B** Left: a close-up view showing objects suspected to proteins. Scale bar - 100 nm. Right: frequency distribution of measured width of the objects observed (n=72).

No protein assemblies could be seen, instead the samples showed excessive amounts of high contrast precipitate (Figure 4.13**A**). Magnification of the dark precipitated areas reveals bright patches that may correspond to the protein (Figure 4.13**B**). The quantified widths of the bright patches reveal an average size of 7-8 nm that may correspond to the Ank4-add or Ank4-achor hubs, however clearer images would need to be taken to validate this conclusion. The presence of large aggregates is most likely caused by the presence of phosphates (from PBS) which are known to cause precipitation of uranyl solutions. Interestingly, however, an equivalent sample preparation mixed with 10 nm magnetite nanoparticles (500 μl of 500 μg/ml

protein solution with 10 μl of 30 mg/ml magnetite) produced magnetite nanoparticle networks in Figure 4.14.

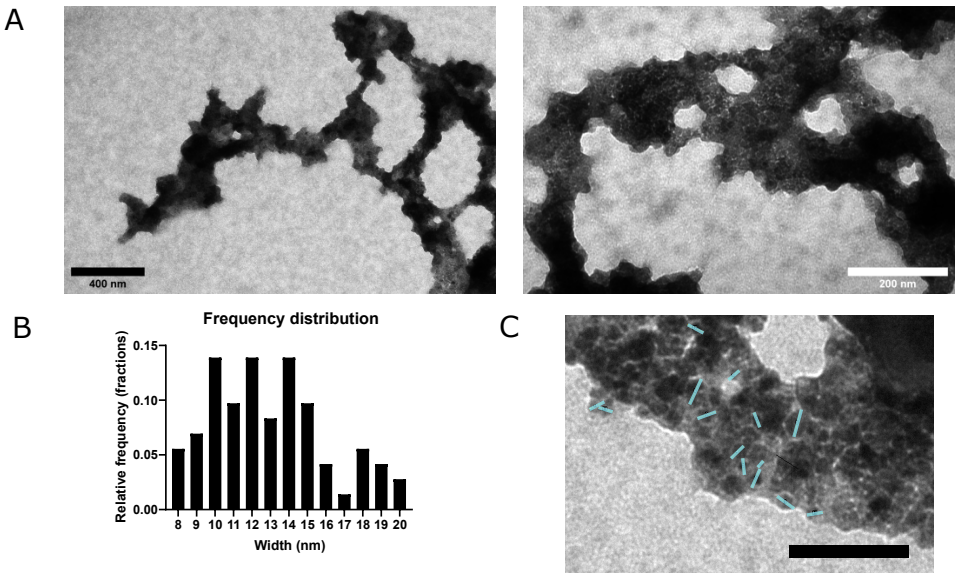


FIGURE 4.14: TEM of Ank4-add, Ank4-anchor and S2 attempted assemblies with 10 nm magnetite. **A** Overview images of what appears to be networks of nanoparticles. Scale bars - 400 nm and 200 nm (left to right). **B** A frequency distribution of nanoparticle sizing measurements, n=72. **C** Method for particle size measurement. Areas where discrete particle are visible were selected and particles were measured across the long axis. Scale bar - 100 nm.

The resulting images showed extended networks of what appeared to be templated 10-13 nm particles. Such behaviour is uncharacteristic to TEM sample preparations of magnetite nanoparticles in isolation. This arrangement of particles may be caused by magnetite being templated to the protein present in the samples. The protein architectures observed are not of the expected dimensions, however (expected 17 nm x 17 nm lattice parameter compared to 100-200 nm^2 observed). This may imply that the proteins are assembling preferentially in one dimension and only occasional branch points emerge. Unfortunately, constructs not containing the magnetite peptide were not available to see if the same templation would be observed.

4.3 Results: HR

4.3.1 Purification

HR-add and HR-anchor DNA sequences were ordered as synthetic genes and were sub-cloned into pET28a plasmids for over-expression *E. coli* in the same manner as Ank4 variants. The expected parameters of the produced proteins are listed in Table 4.1.

Construct	Expected Mass (no M)	Expected Mass	pI
HR-add	25451.67 Da	25582.87 Da	5.16
HR-anchor	24818.51 Da	24949.71 Da	7.98

TABLE 4.2: HR-add, HR-anchor protein characteristics calculated using ProtParam (ExPASy).

The differences in the isoelectric points, again like in Ank4 constructs, are caused by the presence of a glutamate residue rich P6 α-helix in HR-add and a lysine rich BN4 in HR-anchor. The same over-expression and purification protocol (as Ank4 variants) was applied (Figure 4.15).

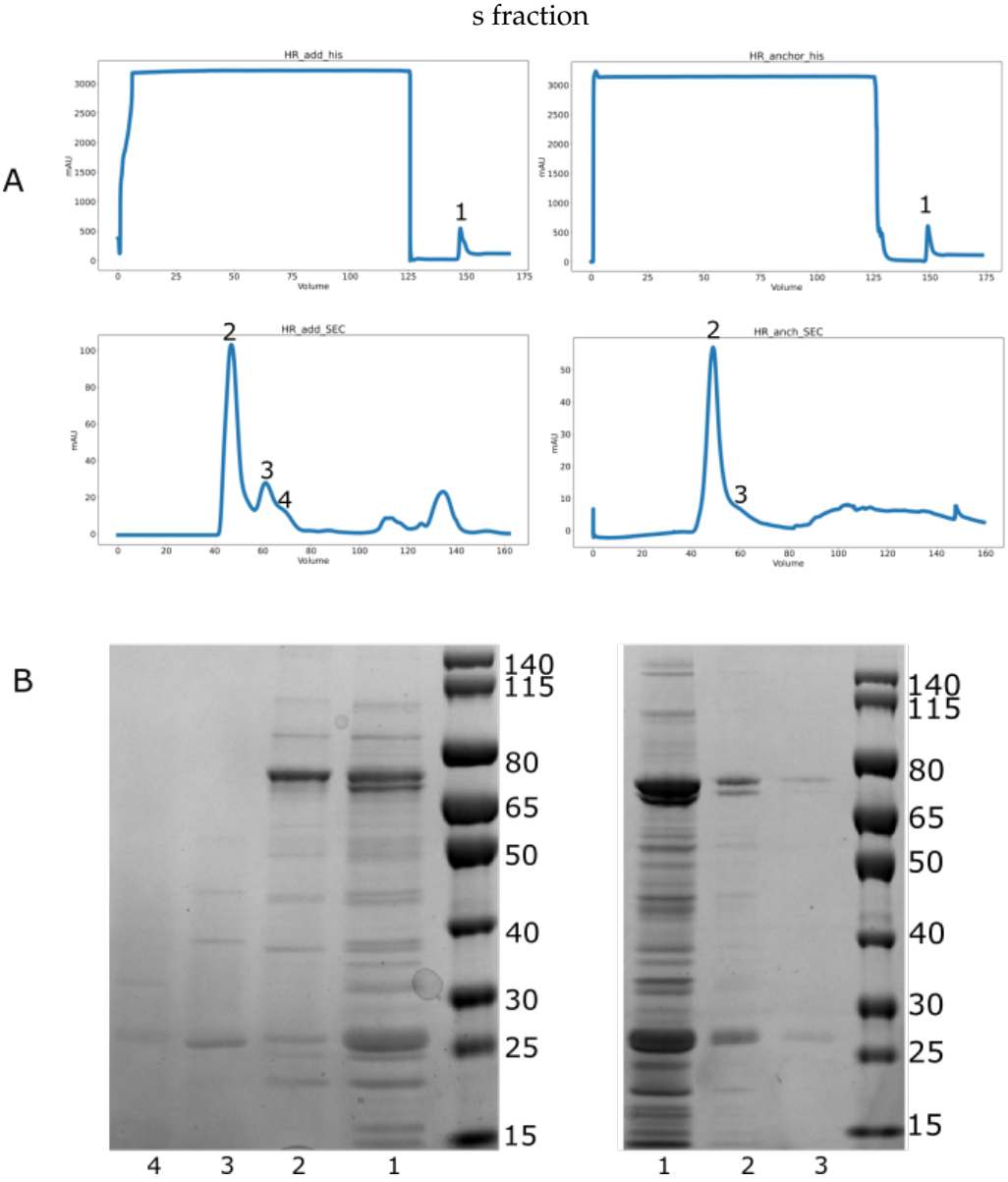


FIGURE 4.15: Purification of HR variants **A** Top: His-Trap column purification performed on the Äkta pure system for HR-add and HR-anchor. High concentration (300 mM) imidazole was introduced at elution volume of 150. Bottom: size exclusion chromatography performed on a HiLoad 16/600 Superdex 75 pg column for HR-add and HR-anchor. Size exclusion chromatography was performed in PBS pH 7.4. 1mM DTT was included for HR-anchor samples. X axes - elution volume, Y axes - mAU. **B** SDS-PAGE analysis of elution fractions corresponding to the labelled peaks in **A**.

The His-Trap column purifications showed discrete elution (300 mM imidazole) peaks implying that the proteins are interacting with Ni^{2+} ions via the C-terminal poly-histidine tags. The proteins eluted from His-Trap column were directly applied to size exclusion chromatography columns. The peaks produced were analysed using SDS-PAGE. A sample of the His-Trap elutions was also analysed (labelled '1'). The fractions produced two prominent bands around 25 and 75 kDa together

with what appears to be contaminants. The sizes closely correspond to the expected monomer, trimer sizes 25451.67, 76355.01 and 24818.5, 74455.5 Da for HR-add and HR-anchor, respectively. For HR-add the fraction corresponding to the gel filtration peak '2' shows an enrichment of the trimer size, whereas fraction '3' (and possibly '4') is enriched for monomers. Due to there being slightly fewer contaminants visible, fraction '3' was chosen for further analysis and experimentation. Fraction '2' was chosen for Ank4-anchor due to there being relatively few contaminants, but significantly more of the target protein compared to fraction '3'.CD spectra were obtained for the two chosen HR-add and HR-anchor fractions (Figure 4.16).

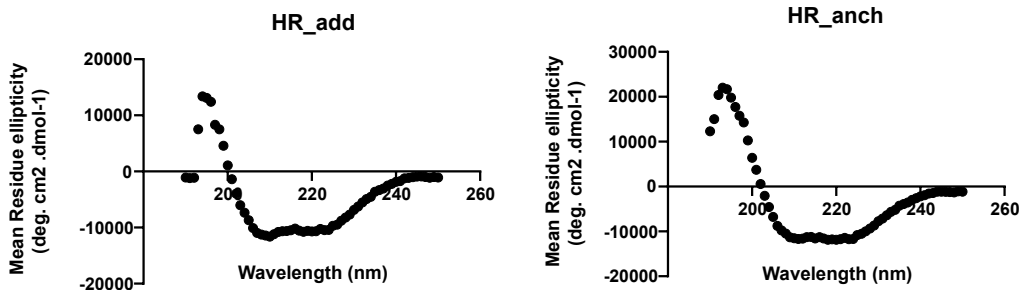


FIGURE 4.16: CD spectra of HR-add and HR-anchor fractions from Figure 4.15. CD spectra performed in PBS pH 7.4 at room temperature. Proteins analysed at 2.5 μM concentration.

Both samples produced spectra resembling those of α-helical proteins with peak minima at 222 and 208 nm and peak maximum near 190 nm. While there is no experimental structural information available for this protein, this does correspond well with the trRosetta models showing α-helical bundles in the protein fold. Unfortunately, ESI-TOF results could not be produced to confirm the exact molecular weights of the purified proteins, but the SDS-PAGE results showing these proteins to be 25 and 75 kDa and sensible CD determined structural composition give confidence that correct proteins have been produced.

4.3.2 Attempts to visualise HR-add, HR-anchor, S2 assemblies using AFM

Dry tapping mode AFM was chosen for attempts at construction of the designed HR-add, HR-anchor, S2 arrays (AFM experiments were performed with the S2A3 variant of the protein). Initially, an attempt was made to obtain flat surfaces of gold through gold sputtering, however the surface roughness was identified to be too great to observe the desired objects (A.11), thus the preliminary work was performed on PLL or PLO coated mica surfaces (for immobilisation of potential assemblies). In-solution mixtures of all three components (HR-add, HR-anchor and S2A3) were visualised first (Figure 4.17).

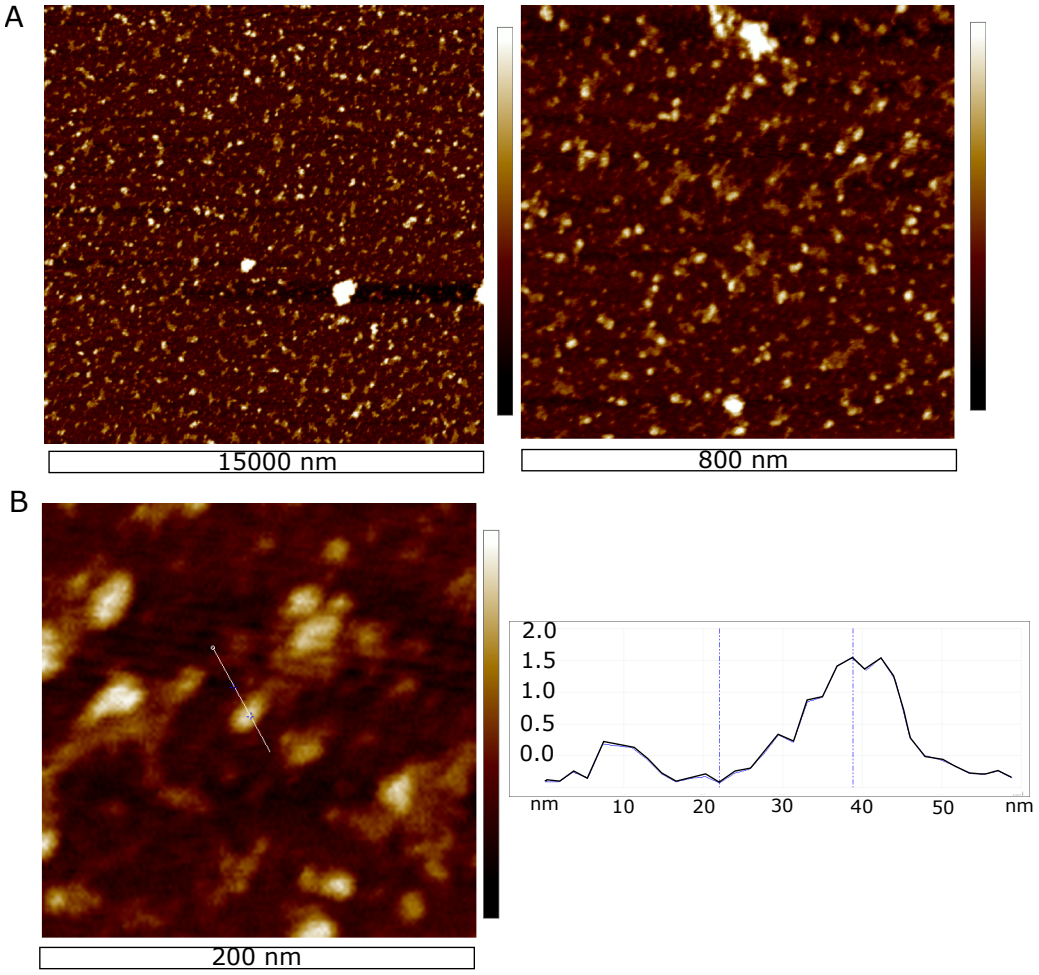


FIGURE 4.17: Dry tapping mode AFM experiment with in-solution mixtures of HR-add, HR-anchor and S2A3 on a PLL surface. **A** Shows overview images showing abundant presence of potentially biological material. **B** Height profile of one of the objects. The width of the square images is written in white boxes below. Height scales on all images were 10 nm.

The preliminary sample preparation followed a procedure of a 20 minute coating

with PLL that was washed with 1 ml of HPLC grade water and dried with nitrogen. The protein mixtures (1:1:1 ratio for each component) were diluted to 300 ng/ml in HPLC water (progressively reducing the buffer concentration) and applied to the coated surface as a 30 μl drop, which was allowed to dry by evaporation (over 2-3 hours at room temperature). The samples were then visualised. The rationale behind the design of sample preparation was to have the simplest possible method investigate if any expected protein networks could be observed. Unfortunately, no assemblies were found. The surfaces showed a distribution of objects that were measured to be 1-3 nm in height. The homogeneous distribution of spaces between observed objects indicates that the surface coating successfully immobilised the bio molecules on surface. It may be that the lack of assemblies was caused by the components being mixed in solution in advance, where they may have assembled and aggregated. An alternative more step-wise preparation was attempted (Figure 4.18).

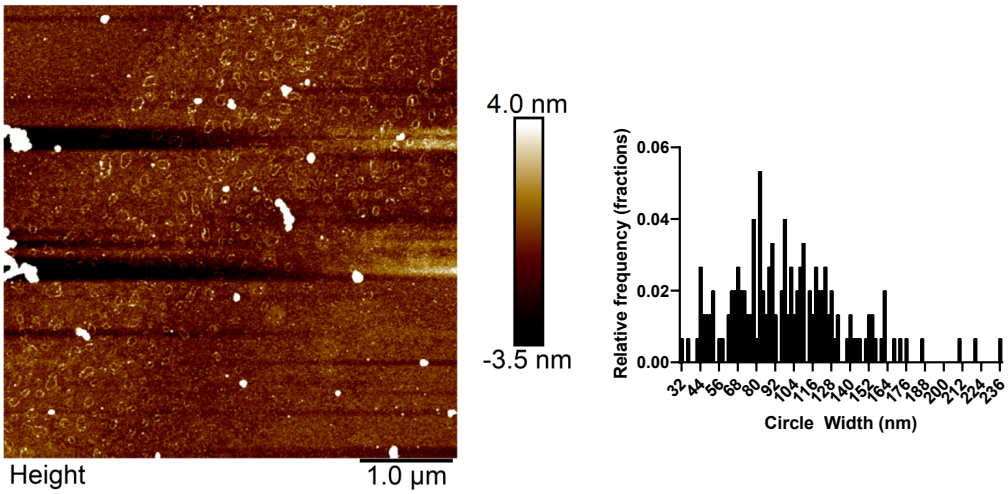


FIGURE 4.18: Dry tapping mode AFM experiment with step-wise assembly of HR-anchor, HR-add and S2A3 on a PLO surface. Left: result of step-wise assembly with multiple visible circular objects. Right: frequency distribution histogram of measured circle widths (n=150). Oval-shaped objects were measured along their length. The vertical scale bar represents the height range of the image.

The sample was prepared by coating a mica surface with PLO for five minutes and then washing (four gentle washes with 100 μl of HPLC grade water). A premixed solution ($\leq$30 minutes in advance) of HR-anchor and S2A3 was then applied to the surface at a concentration of 100 μg/ml for 2 hours. The liquid was then removed (but not washed) and equal volumes of S2A3 and HR-add (at equimolar

concentrations) were applied and incubated on the surface for 18 hours at room temperature (the sample surface was kept in a small tightly closed container to prevent drying). The protein solution was then removed and the surface was washed with HPLC grade water (3 times with 100 μl each), the samples were then dried with nitrogen and visualised. The rationale behind the sample preparation was to to first seed the surface with HR-anchor/S2A3 complexes and then supply HR-add with additional S2A3 for networks to form around the seeded HR-anchor proteins. Omitting the wash step between the two protein incubations was to retain a small amount of HR-anchor in solution for extended network assembly. The resulting images showed unexpected arrays of circular objects. The objects were within a height range of 1-3 nm. The circular assembly diameters were also quantified and ranged between 32-236 nm and were 80 nm on average. There were also large aggregates that were difficult to identify. Higher magnification images were also taken and separate images for individual circular objects were extracted (Figure 4.19).

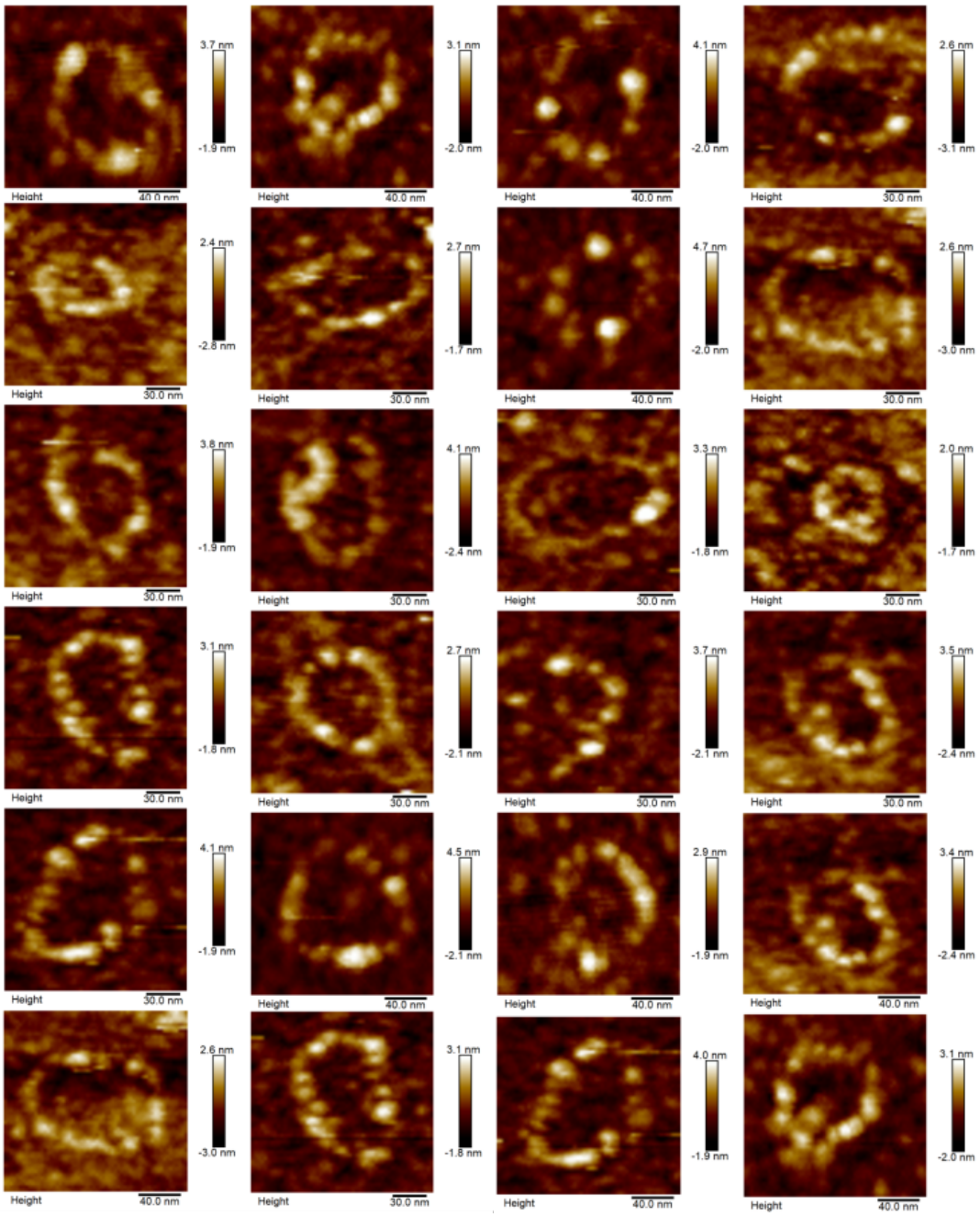


FIGURE 4.19: Extracted magnifications of the circles observed in Figure 4.18. The vertical scale bar represents the height range of the image. The scale bars are either 40 nm or 50 nm (labels under individual images).

The circles observed appear to be composed of individual segments. The segments themselves seem to be variable in height. If these were composed of HR-add and HR-anchor linked by S2A3 bridges, one would indeed expect a series of 3 nm hubs interspersed by 1 nm bridges. Width measurements of the structural model of S2A3 show the expected height of 1-1.5 nm (depending on where along the protein

it is measured), whereas measurements of either of the HR-add or HR-anchor structural models show a 2.7-3.3 nm width. An attempt to quantify the hub and linker heights can be seen in (Figure 4.20).

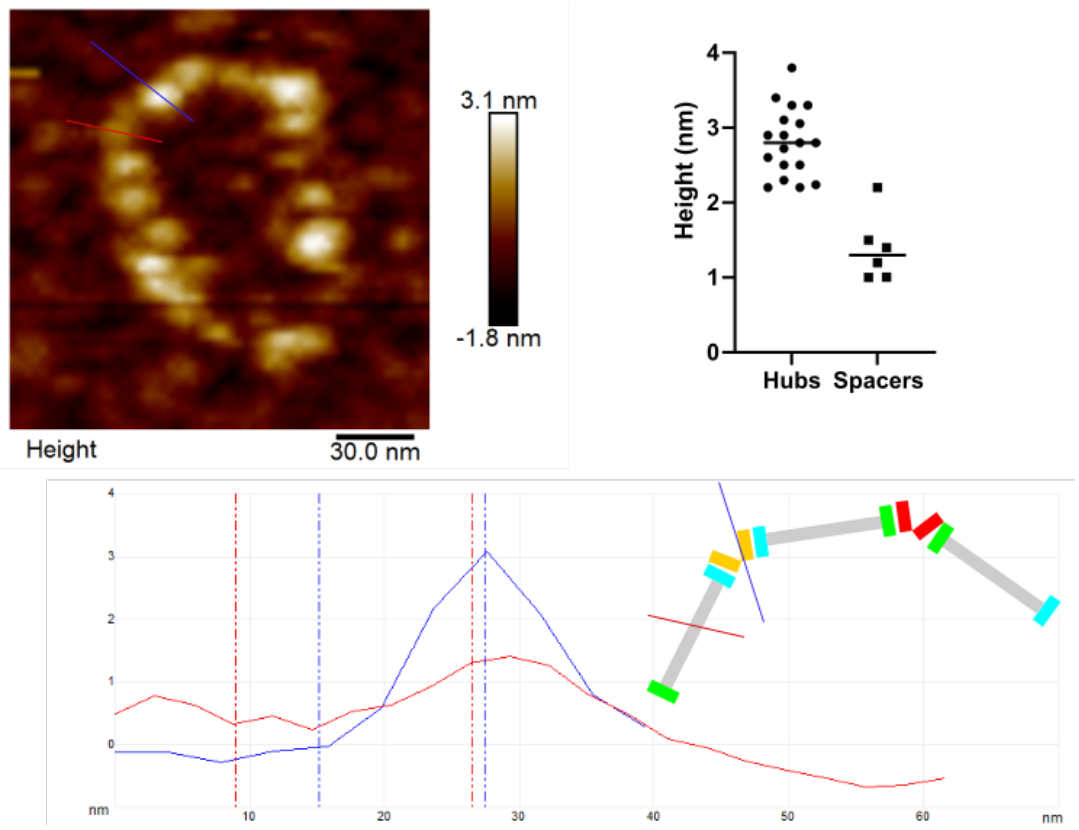


FIGURE 4.20: Quantifying the heights of segments composing circles displayed in Figure 4.19. Top right: swarm plots of heights of hubs (suspected HR-add or HR-anchor, n=17) and spacers (suspected S2A3, n=6). Bottom: height profile measurement example of a hub (blue) and a spacer (red) together with a graphical representation of suspected objects being measured.

The measurements do reveal that the higher hubs seem to correspond to $\sim$ 2.9 nm and the linker to $\sim$ 1.3 nm. The measurement method was to identify regions where clearly visible 'dimers' of highest objects (hubs) were connected by a lower region (higher than background) to a another pair of larger objects. While there are some outliers within both sets of measurements that do not fit the expected size ranges, they can be accounted for by slight variations of the underlying coating height or different sample material depositing on top of each other. Nevertheless, the measurements correspond well with the observed circle structures being assemblies of HR-anchor, HR-add and S2A3. The fact that stretches of hub-hub-spacer-hub-hub-spacer-hub-hub can be seen indicates that S2A3 is able to form coiled-coil

interactions with HR-add and HR-anchor simultaneously. A graphical representation of how such circular assemblies could form is shown in Figure 4.21

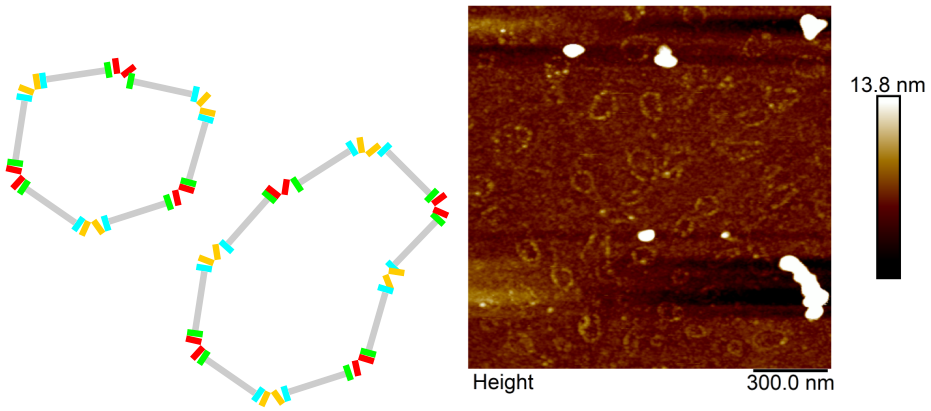


FIGURE 4.21: A diagrammatic representation of a potential explanation for HR-add, HR-anchor and S2A3 assembling into circular structures.

While the original design of the three component self-assembly system expects the formation of hexagonal lattices, circular assemblies are also theoretically possible, if each of the hubs only formed two interactions with S2A3 instead of the expected three. If the two S2A3 sub-units joined to a hub have an angle with respect to each other (which is expected as the hubs are C3 symmetry assemblies) an eventual formation of a closed loop - a circle is possible. It is unclear whether or not the hubs observed in the AFM images are composed of HR protein dimers. It may well be that the hubs are trimeric, but not clearly visible. Circle measurements in Figure 4.18 show that the lower end of the size distribution is 32 nm which corresponds to the smallest possible diameter of a circle assembled in this manner (assuming that 90 °C angles are not allowed for the assemblies to form 17 nm^2 squares). Two scenarios can be proposed about how such circles may assemble. First, the hubs may be trimers in solution, but can become dimeric through a conformational change induced by two coiled-coils forming in presence of S2A3, alternatively the hub does not become dimeric, but the conformational changes cause the third coiled-coil forming moiety to become inaccessible. Second, the hubs actually each form three interactions with S2A3 and form hexagonal lattices on the surface, which are disrupted by the drying process during sample preparation. This may explain why the circles were observed

so close to each other. Future experiments in liquid may reveal if the second hypothesis is true.

4.4 Chapter 4 Conclusion

Taken together these results demonstrate the initial characterisation of two sets of scaffolds previously designed to assemble into trimers or tetramers [254] modified with a set of coiled-coils and a magnetite binding peptide to create two-dimensional assemblies of protein for immobilisation of magnetic nanoparticles. Both sets of newly designed proteins have been produced successfully and seem to possess expected α-helical structure and both sets seem to retain their ability to oligomerise as seen by high molecular weight bands persisting even in denaturing SDS-rich conditions. The method to assess coiled-coil based interaction used in Chapter 3 was not successful in this context, optimisation may be possible, however the high α-helical content in either Ank4 or HR proteins would make it difficult to detect the changes caused by the coiled-coil forming α-helices unravelling. Two divergent strategies are shown for characterising the interaction of the newly designed hubs with S2 components. For Ank4, an LSPR method is employed showing that both variants are able to immobilise their binding partner S2x and that Ank4-add has a strong specificity for S2x rather than for S1x. The AFM results for the HR variants mixed with S2A3 reveal structures that potentially are either precursors to the expected hexagonal lattices or are the results of the lattices being disrupted due to experimental conditions. The AFM visualised circular objects made up of HR-add, HR-anchor and S2 are substantially smaller and more consistent in arrangement that S1 and S2 assemblies in Chapter3. This shows that introducing a hub component has the potential to achieve more controlled two dimensional protein assemblies. In conclusion, Chapter 4 shows that the SasG G52-E-G53 domain coiled-coil fusion proteins have potential to be integrated with existing designed protein frameworks.

Chapter 5

Results: S-layer nano-sheets

5.1 Chapter 5 Summary

Chapter 4 concludes the work in characterisation of novel nano-patterning proteins, whereas Chapter 5 skips-ahead to attempt a demonstration of the ability to immobilise nanoparticles onto two-dimensional protein patterns via genetic fusion of an S-layer protein with magnetite binding peptide A3 and the magnetite biomineralisation protein Mms6.

The chapter introduction covers literature on previous attempts to use S-layer proteins for metallic nanoparticle patterning. The introduction is concluded with a focus on a specific S-layer protein SgsE and a summary of the patterning strategy for this thesis.

5.2 Introduction to S-layer Protein Work

5.2.1 S-layer Protein-Based Nano-patterning

Due to their intrinsic ability to form ordered patterns, S-layer proteins have been previously investigated as metal and metal nanoparticle patterning platforms for data storage and nano-electronics. S-layer proteins from *Geobacillus stearothermophilus* [255, 256, 257], *Lactobacillus kefiri* [258], *Lysinibacillus sphaericus* [259, 260, 261], *Bacillus subtilis* [262], and *Sporosarcina ureae* [263, 264] have all been used for patterning, but the most commonly used S-layer proteins were the hexagonal lattice HPI (originating from *Deinococcus radiodurans*) [265, 266, 267, 268, 269] and tetragonal lattice

SbpA (originating from *Bacillus sphaericus*) [255, 270, 271, 272, 273, 274, 275, 276, 277, 278]. A summary of the literature can be found in Table 5.1.

Species	S-layer	Particle Type	Particle Application	Particle size range
Deinococcus radiodurans	HPI	Co, FeCo, Fe, CoNi, NiFe[265], Au(0) [266], citrate-capped Au(0) [279, 267, 268, 269], Pt-DENS [268]	sputter deposition [265], pre-formed incubation [266, 279, 280] [267, 268, 269]	1.8 nm-10 nm
Bacillus sphaericus	SbpA	Au(0) [270], FePt [276, 277, 278] Pd(0) [273, 274] CdSe [272], CdS [255], Pt(0) [271], citrate-capped Au(0) [272]	pre-formed incubation [272], chemical reduction [255, 270] [271, 273, 274], sputter deposition [276, 277, 278]	0.8 nm-6 nm
Geobacillus stearo-thermophilus	-	CdS [255]	chemical reduction [255]	8 nm
Lactobacillus kefiri	-	Pt [258]	chemical reduction [258]	-
Lysinibacillus sphaericus	SbpA, Slp1	Au(0) [259, 260] prussian blue [281]	pre-formed incubation [259, 260, 281]	5 nm-6 nm
Bacillus subtilis	-	Au(0) [262]	chemical reduction [262]	4 nm-6 nm
Sporosarcina ureae	-	Pt(0) [263], Pd(0) [264]	chemical reduction [263, 264],	1.2 nm-2 nm

TABLE 5.1: Summary of S-layer protein studies attempting to achieve discrete inorganic metal nano-patterning.

Three distinct approaches to patterning could be inferred from the available literature: incubation of pre-formed nanoparticles in liquid media in presence of an

S-layer surface, chemical reduction of metal salts on S-layer surfaces and lastly sputter deposition of gaseous metals onto S-layer surfaces. Most studies using the aqueous pre-formed particle approach relied on metal nanoparticles with coatings, such as citrate capping for Au [282, 272, 269, 259] or dendrimer coated Pt nanoparticles [268] making the particles more soluble and in the case of dendrimer coatings, providing a way to tune particle size. Other studies used bare colloidal gold [266, 272, 260] and Prussian blue [281] particles. The process generally starts by depositing already assembled S-layer sheets on copper grids (tailored for TEM) and incubating particle solutions at room temperature in the time range between 20 minutes to 2 hours (some exceptional protocols describe binding within a few seconds [266] or 24 hours [281]). The particles are then washed off and the layers are visualised under TEM. The chemical reduction method relies on incubating surface deposited assembled S-layers with various metal salt solutions (10 mM CdCl2 [255], 10 mM $HAuCl_4.3H_2O$ [270, 262, 283, 284], 3 mM K_4PdCl_4 [264, 263, 284], 2 mM Na_2PdCl_4 [273, 274], 1 mM silver nitrate [262]. The salt solutions are incubated anywhere between 5 minutes and 4 days at room temperature for the metal ions to interact with the S-layer surfaces. The patterning is then achieved by addition of reducing agents such as H_2S, 3 mM NaN_3, 6 mM hydrazine solution or 22 mM dimethylaminoboran hydride. The reducing agents trigger biomineralisation of the metal salts within the exposed S-layer protein cavities. A few studies observed the metal pattern to become more discrete with longer exposures under the TEM, suggesting that electron attack may act to etch unpatterned metals away in these samples [270, 271]. Lastly, the gaseous sputter deposition method uses gas phase nanoparticles and applies them to the S-layers in vacuum, for example, a study with a *Bacillus sphaericus* S-layer, used FePt particles from gas phase to sputter a pattern on a surface, the resulting nanoparticle patterns matched the underlying S-layer lattice parameters [276]. The FePt nanoparticles were also annealed to assume $L1_0$ crystal structure. A follow-up study combined the sputtering method with an applied magnetic field during nanoparticle deposition on the S-layer surface [277]. The applied magnetic field reduced particle agglomeration and increased regularity in particle density on the surface. An alternative vapor deposition strategy was demonstrated with a *Deinococcus radiodurans* S-layer [265] where Co, FeCo, Fe, CoNi and NiFe were deposited on top

of S-layers as 2.5 nm thick layers that were later etched with argon ions to reveal the patterns. All studies achieved discrete patterning to variable degrees. The clearest patterns achieved were in studies focusing on coated gold nanoparticles. Magnetic particles provided their challenge in patterning due to their tendency to agglomerate together. The mode of particle to S-layer interaction was postulated to be electrostatic [272] and hydrogen bond based [268], but in general it was assumed that the three-dimensional porous structure of the S-layer surface was acting as a trap for the nanoparticles. Only a few studies used S-layers that are not native [270, 259, 260]. There, S-layers were modified with cysteine to immobilise gold nanoparticles through thiol bonds.

5.2.2 New Strategy for S-layer based Nano-patterning

This project aims to build on the S-layer nanoparticle literature by, this time, fusing an S-layer to proteins and peptides known to have specific affinity for magnetic nanoparticles rather than relying on an the intrinsic ability of S-layers to non-specifically immobilise nanoparticles through their three-dimensional architecture. Magnetite was chosen again as the model magnetic nanoparticle. Peptide A3 and a whole Mms6 have been chosen as magnetite binding moieties to be fused to the S-layer protein of choice. Mms6 was chosen with the expectation that it may facilitate templated biomineralisation of magnetite on the surface of the S-layer.

The S-layer chosen for this project is SgsE from *Geobacillus stearothermophilus* NRS 2004/3a, first described in some depth by Schaffer et.al., 2002 [285] where the primary amino acid sequence was reported. Native SgsE S-layer sheets present lattice parameters of a = 11.6 nm, b = 9.4 nm and $gamma$ = 78°. Full length SgsE was shown to be amenable to heterologous expression in *Lactococcus lactis* NZ9000 where SgsE was shown to retain self-assembly capabilities *in vitro* displaying assembly of 700 nm - 3 μm long cylinders (90-300 nm in diameter) and 4 μm sheets. Assembly triggered in presence of *G. stearothermophilus* sacculi (cell wall fractions of dead cells), hydrophobic and hydrophilic surfaces showed the versatility of the assembly capabilities. The SgsE protein was shown to be a viable fusion protein platform for enzyme immobilisation [208]. There, two truncated versions $SgsE_{131\text{-}903}$ and $SgsE_{331\text{-}903}$

were genetically fused through C-terminus to RmlA (a 33.2 kDa enzyme that converts thymidylmonophosphate nucleotide to glucose-1-phosphate). Immunogold labelling showed that C-terminus was more accessible thus C-terminus was chosen for the fusions. All constructs were successfully expressed in *E. coli* and purified via N-terminal poly-histidine tags. All of the constructs retained their ability to form S-layers in solution, on spherical silicon nanoparticles and silicon wafers. The enzyme fused S-layers retained close to 100 % activity (when compared to soluble RmlA) and showed an improved shelf-life up to three-months of storage in 4 °C (to compare soluble RmlA activity dropped to 20 % by week 4). The same $SgsE_{131\text{-}903}$ truncations have also been shown to retain assembly capabilities when genetically fused (C-terminal) to three different GFP mutants [286]. The GFP proteins fused to assembled SgsE retained their native spectral qualities. A follow-up study then showed that SgsE S-layers can assemble on solid surfaces with two fluorescent proteins fused to the SgsE monomer unit on both N- and C- termini at the same time [287]. Lastly, a stability study showed that EGFP fused to N- or C- terminal ends of SgsE have increased stability [288]. Thermal stability of the SgsE protein in the context of a fusion was not reduced (compared to native SgsE).

The literature listed above shows that SgsE may be a suitable S-layer protein to immobilise magnetic nanoparticle binding proteins/peptides. First, it was proven to accommodate fusion proteins and it was shown that genetically fusing them to the C-terminus exposes them to the surface of the formed layers. Second, the ability of the S-layer to assemble the discreet periodical patterns on various surfaces and in solution is also not affected in a context of a fusion. Lastly, the fusion S-layer appears to stay as stable as its native counterparts and it appears to impart additional stability to its fusion partner, which may be because this S-layer protein originates from *G. stearothermophilus* NRS 2004/3a which is a known thermo-extremophilic organism.

5.3 Production and Characterisation of SgsE Constructs

5.3.1 Design and Sequences of SgsE Fusion Proteins

The truncated $SgsE_{131\text{-}903}$ was chosen as the S-layer platform for this project. The protein sequence for SgsE was accessed from the Uniprot data base (accession number: Q8VTF1). It was then reverse translated to obtain a corresponding DNA sequence (using arbitrarily assigned codons). The DNA sequence was then submitted to the GeneOptimizer (Thermofisher Scientific, US) online tool to obtain a codon optimised DNA sequence for protein expression in *E. coli*. The sequence was then ordered (GeneArt synthesis, Thermofisher Scientific, US) as a double stranded linear sequence with an added N-terminal poly-histidine (6) tag and a C-terminal GGS flexible linker for fusion protein construction (Figure A.13). The sequence was then sub-cloned into pPR-IBA1 (downstream of a T7 promoter) to produce an array of different fusion constructs (Figure A.14). Graphical representations of the constructs can be seen in Figure 5.1 .

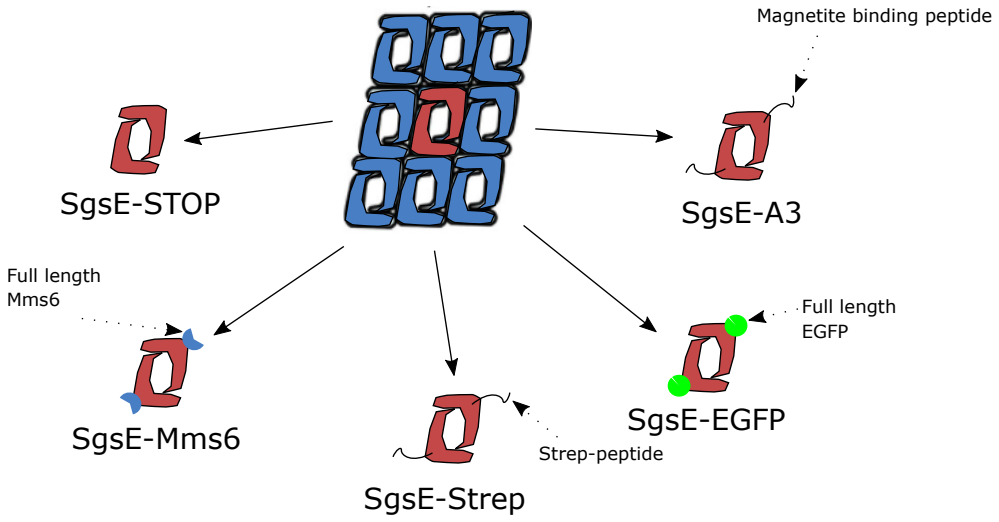


FIGURE 5.1: A diagram of all constructs featured in Chapter 5: SgsE-STOP, SgsE-A3, SgsE-EGFP, SgsE-Strep and SgsE-Mms6.

Throughout the project $SgsE_{131\text{-}903}$ was referred to only as SgsE for brevity. SgsE-STOP was the construct that contained a 'Stop' codon right after the flexible linker and had an estimated molecular weight of 83920.03 Da. This construct was to act as a

negative control for particle/metal binding experiments. SgsE-A3 (expected molecular weight of 85045.34 Da) contained the magnetite biding peptide A3 (HNHKSKKHK) at the C-terminus. SgsE-Strep (expected molecular weight of 85118.33 Da) contained a streptavidin binding peptide and it acted as a control for S-layer assembly with a generic peptide not selected for magnetite binding. SgsE-EGFP (expected molecular weight of 110843.37 Da) was the largest of the constructs used in the project and was designed with the intention of tracking S-layer assembly status and validating that the protocols employed during this project facilitated maintenance of functionality in SgsE fusions (EGFP stands for enhanced green fluorescent protein). Finally, SgsE-Mms6 is a fusion protein between SgsE and the magnetite interacting/iron nucleation membrane protein Mms6 (expected molecular weight of 98691.96 Da).

5.3.2 Expression SgsE Constructs

Ni-NTA purified proteins were analysed on SDS-PAGE before the final dialysis reassembly step (Figure 5.2).

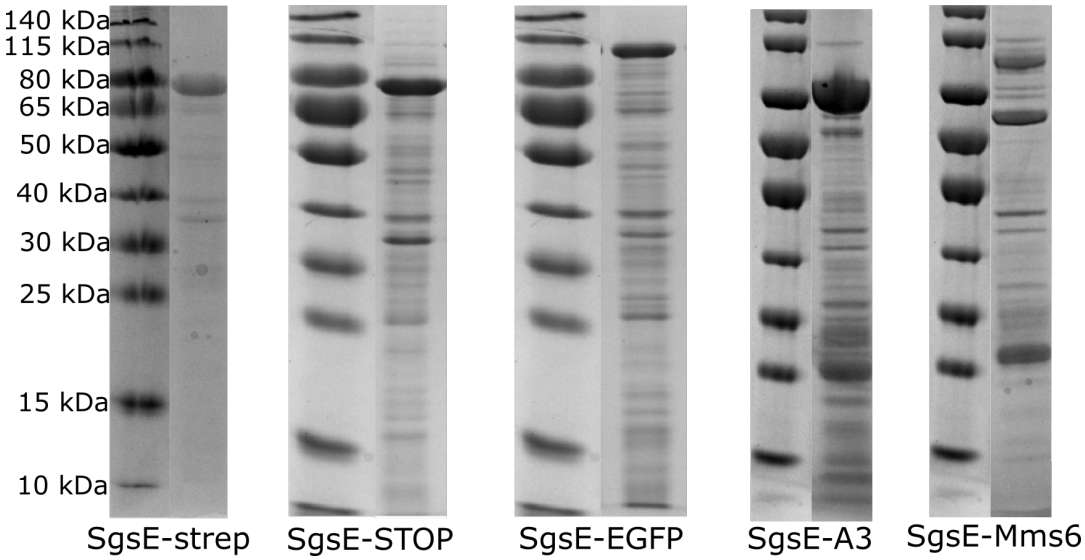


FIGURE 5.2: SDS-PAGE of Ni-NTA purified SgsE fusion proteins analysed in on tris-glycine stacking gels. Elution fractions in 400 mM imidazole buffers have been run on separate gels. The gel images were cropped to show the elution fractions only.

The elution fractions showed enriched bands at expected corresponding molecular weight sizes. The contaminant bands represent *E. coli* proteins binding to the

Ni-NTA resin non-specifically. SgsE-Mms6 consistently (between multiple purifications and growth regimes) produced two high molecular weight bands, one - at the expected molecular weight, other - slightly lower than SgsE without a fusion. The aberrant gel movement of the protein may be due to Mms6 being membrane protein with intrinsically disordered tendencies (intrinsically disordered proteins have been shown to show SDS-PAGE movement at unexpected sizes and to exhibit laddering [239]). Ni-NTA purified and assembled SgsE-STOP and SgsE-A3 were analysed using ESI-TOF to precisely confirm the molecular weight of the obtained proteins (Figure 5.3).

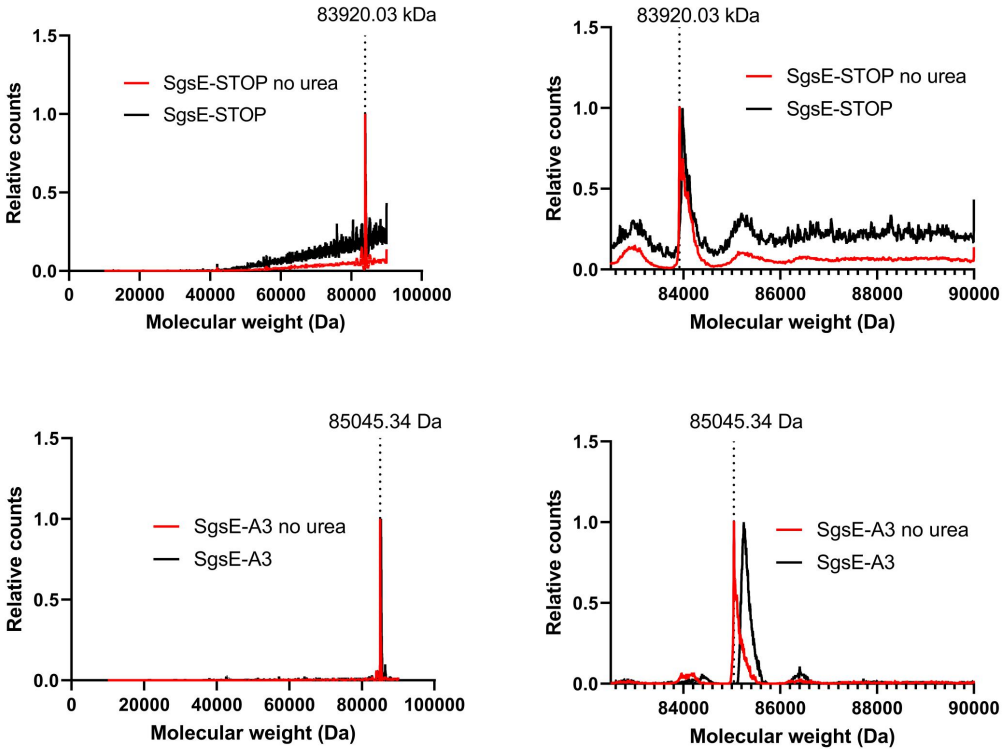


FIGURE 5.3: ESI-TOF on SgsE-STOP and SgsE-A3. Left: full mass spectra, on the right: spectra focused on the approximate size of the target protein. Both SgsE-STOP and SgsE-A3 show their respective correct expected sizes in samples run without 3M Urea.

Initially, the samples were diluted to a final concentration of 3 M urea before being subjected to ESI-TOF in order to dissolve the S-layer sheets. Both SgsE-STOP and SgsE-A3 showed sizes larger than theoretically calculated (58 Da excess for SgsE-STOP and 200 Da - for SgsE-A3). Performing the same analysis on samples not dissolved in urea produced results showing the expected sizes for both constructs (83920.03 and 85045.34 for SgsE-STOP and SgsE-A3, respectively). These results show that the methods employed in this project produce correctly sized SgsE monomers and that urea can potentially associate with the protein in a way that is not disrupted during the electrospray ionisation process or that it can induce the appearance of unexpected post translational modifications. To confirm that the purified S-layer proteins were functional, the dialysed assemblies were analysed using TEM with uranyl formate negative staining (Figure 5.4).

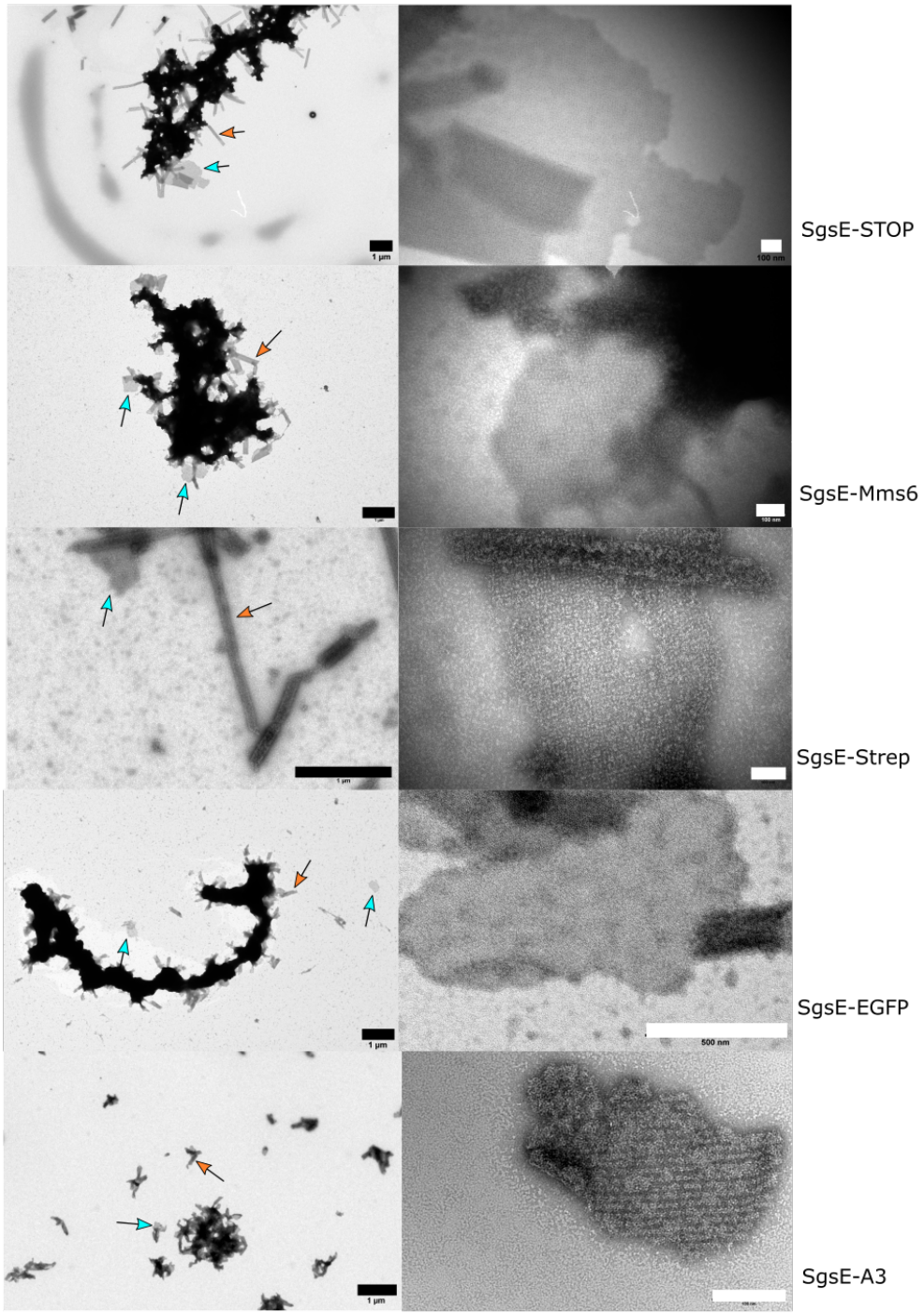


FIGURE 5.4: TEM images of different SgsE fusion proteins re-assembled with Ca^{2+} and dialysed in ultra pure water. Each of the proteins assembles into a mixture of flat sheets and aggregated rods. Each sample was stained with uranyl formate for enhanced contrast. Scale bars on the left - 1000 nm, right - 100 nm. (SgsE-EGFP scale bar - 500 nm). Arrows denote assembled SgsE sheets (cyan) and rods (orange).

Each of the SgsE dialysis products showed typical forms for S-layer proteins assembled *in vitro* which have also been previously described for recombinant SgsE

[289]. This included a mixture of rods and sheets (rods have been suspected to be linear sheets that have reached a critical length causing them to 'scroll-up'[71]). All of the fusion protein microscopy preparations showed the assemblies (rods and sheets) further aggregate into large agglomerates (Figure 5.4 on the left). Sheets and rods from all samples were seen to have periodical patterns (Figure 5.4 - right). Close-up images of the sheets were taken to extract periodicity information using Fourier transform image analysis (Figure 5.5).

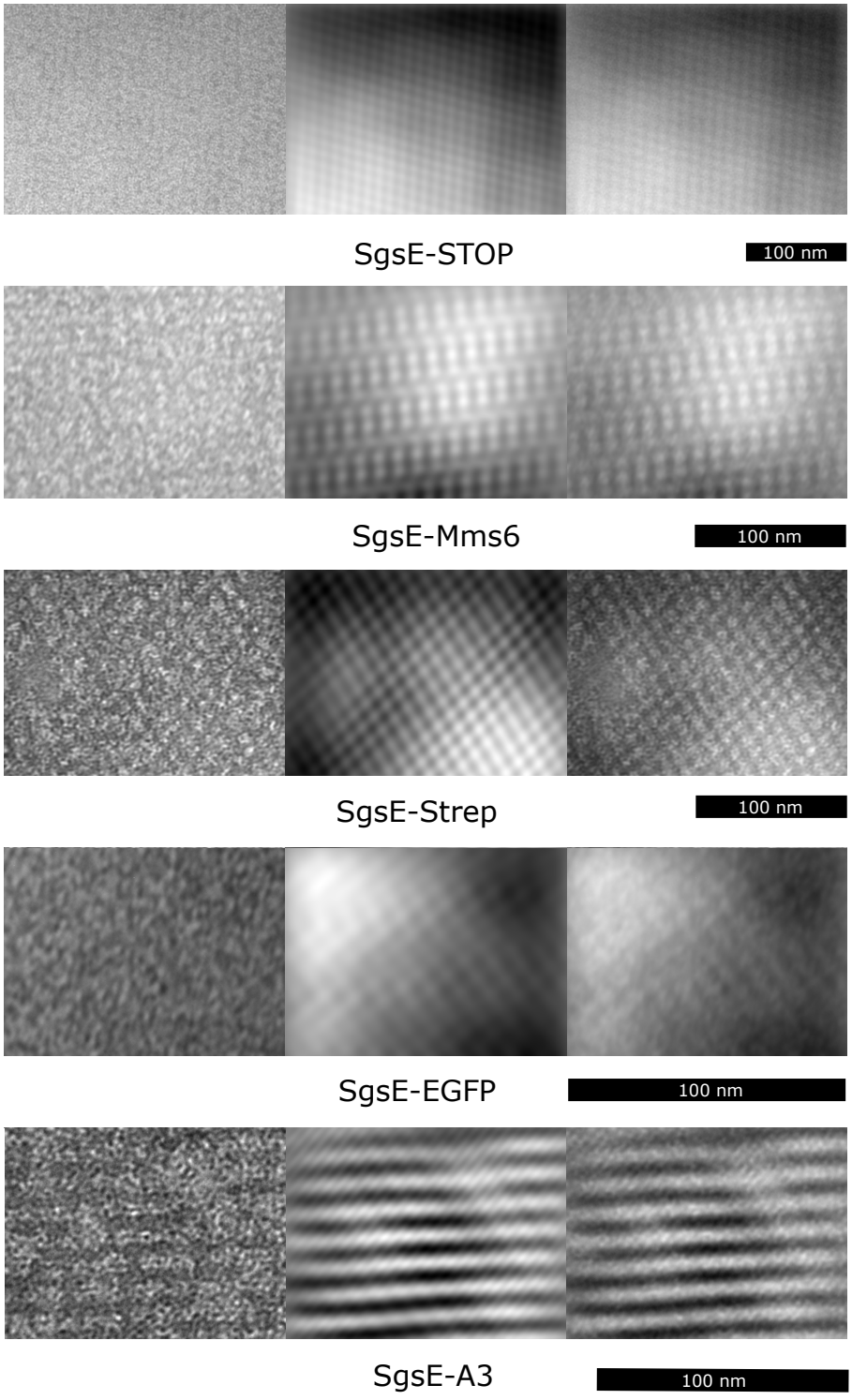


FIGURE 5.5: TEM close-up images of SgsE fusion proteins. First column shows a close-up of obtained flat S-layer sheets. Second - a Fourier transform processed image uncovering the periodicity present in the first image. Third - an overlay of the first and second columns. Scale bars for columns one and two are the same as seen under the overlay images.

The Fourier transform analysis allows for image abstraction that eliminates noise and enriches any repeating signals present in the images. All fusion constructs

showed periodical structures emerging. SgsE-STOP and SgsE-Strep showed two-dimensional periodical features that fit the expected lattice angle (summarised in Table 5.2) for SgsE S-layer sheets [285] specifically the oblique angle of the unit cell $\gamma = 78°$. SgsE-Mms6 and SgsE-EGFP samples exhibited patterns with imperfections similar to each other where it can be seen that the separate S-layer pattern lines overlap with each other. These imperfections may be caused by the fact that in these cases SgsE is fused to bigger partner proteins than in the case of SgsE-Strep or SgsE-A3. SgsE-A3 shows a clear periodicity in one dimension with a hint of there being periodicity in the other. The lack of a clear view of a two-dimensional periodicity may have been caused by the starting images (before performing the Fourier transform) not being of high enough resolution or the particular S-layer sheets imaged having too much unspecific debris obstructing the attempts to extract periodicity information.

S-layer	γ
SgsE-STOP	80°
SgsE-Strep	81°
SgsE-EGFP	81°
SgsE-A3	83.6°
SgsE-Mms6	86.6°

TABLE 5.2: SgsE lattice angles measured from images in Figure 5.5.

SgsE-EGFP assemblies were also visualised in liquid (PLL coated surface) and dry AFM and under a light microscope (Figure 5.6). The same preparation can be seen to form rods and sheets under different microscopy conditions. The low resolution light microscopy images show green sheets which implies the sheets observed in TEM and AFM indeed array the EGFP fusion at nanoscale. This acts as a validation that the conditions used to purify the SgsE fusion proteins allow for correct assembly of SgsE-S layers while maintaining the function of the fused moieties.

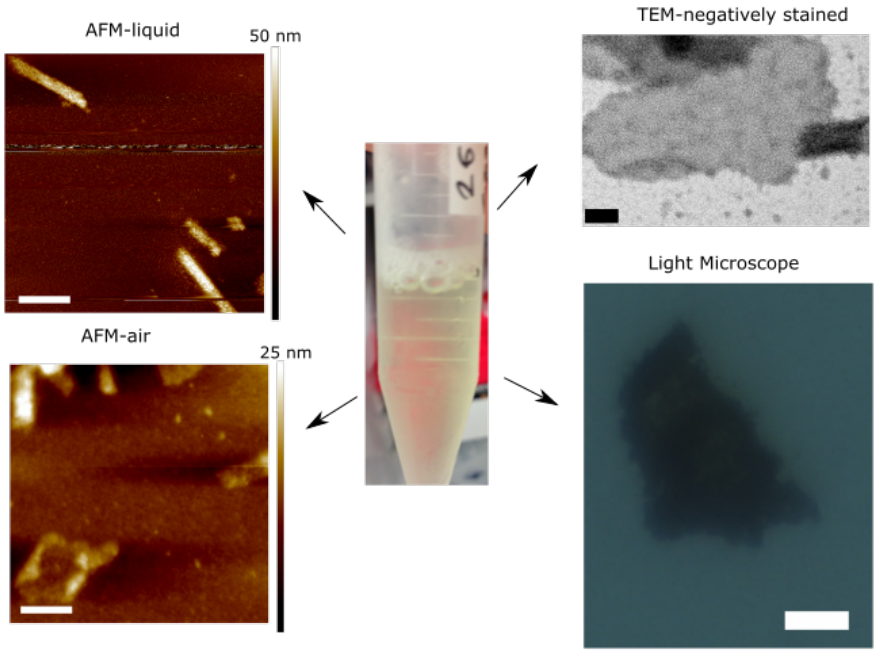


FIGURE 5.6: SgsE-EGFP assembly visualised on AFM-liquid, AFM-tapping air, uranyl formate stained TEM and under a light microscope. Centre: eluted SgsE-EGFP showing fluorescence. AFM scale bars - 1 μm, TEM - 500 nm, light microscopy 1 μm.

5.3.3 SgsE S-layer Fusion Protein Stability

SgsE-STOP, SgsE-A3 and SgsE-Mms6 assembled S-layer sheets were investigated for their thermal stability by tracking the CD signal at 222 nm (Figure 5.7 and summarised in Table 5.3). SgsE-A3 and SgsE-STOP showed a melting temperature of 60 ° C, whereas SgsE-Mms6 was at 62 ° C. This may be an indication that Mms6 fusion is providing the assemblies additional stability (perhaps through Mms6 moieties associating amongst themselves through their exposed hydrophobic side-chains), however the results are not conclusive and further investigation may be required. A previous study has identified $SgsE_{133\text{-}903}$ to have a melting temperature of 71 ° C using differential scanning calorimetry. It is unlikely that two additional amino acids (as the study used $SgsE_{133\text{-}903}$ instead of $SgsE_{131\text{-}903}$) at the N-terminus would impart an increase of 11 °C stability over the SgsE-STOP construct, thus this must arise from the different experimental set ups and sample preparation. Nonetheless, these stability data show that A3 and Mms6 fusions do not dramatically change the thermal stability of SgsE S-layers.

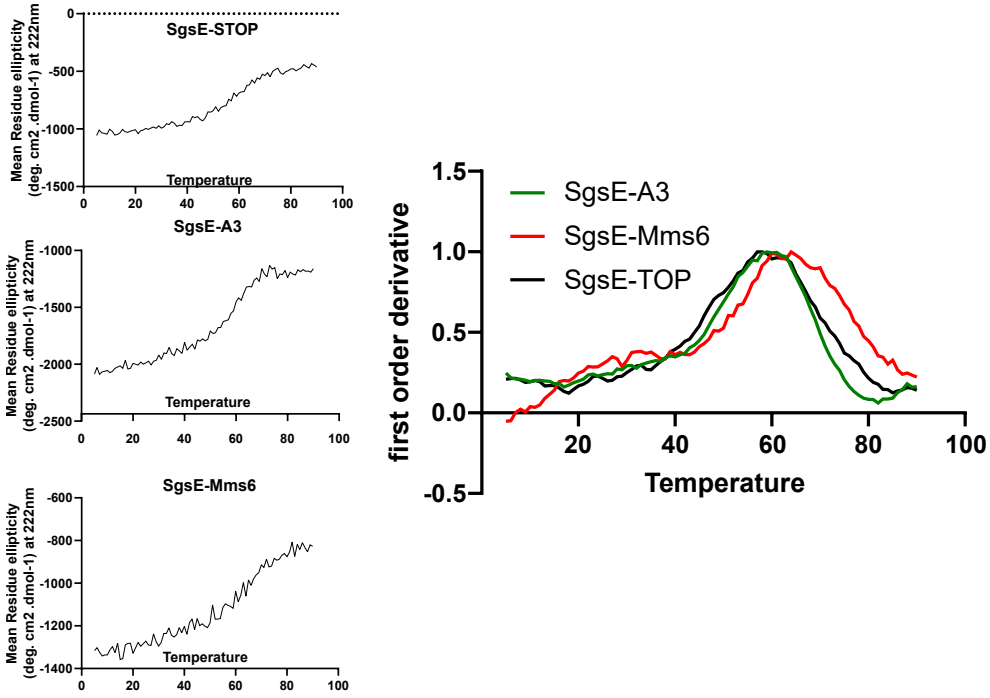


FIGURE 5.7: Temperature dependant folding curves comparing SgsE-A3, SgsE-Mms6 and SgsE-STOP. Bottom: first order derivatives revealing transition temperatures at 60 °C for SgsE-A3 and SgsE-STOP. SgsE-Mms6 appears to exhibit a slightly higher 62 °C transition temperature. The data was obtained by running thermal denaturation experiments and tracking circular dichroism at 222 nm. Proteins (2μM) analysed in ultra-pure water.

S-layer	Tm (°C)
SgsE-STOP	60
SgsE-A3	60
SgsE-Mms6	62

TABLE 5.3: Assembled SgsE S-layer thermal stability information extracted from Figure 5.7.

SgsE-EGFP was used to study the stability of the assembled S-layers in urea (Figure 5.8). An assay was designed to identify the concentration of urea needed to dissolve assembled SgsE fusion S-layer sheets. Briefly, equal amounts of SgsE-EGFP S-layers were resuspended in different urea concentrations to measure the amount of SgsE-EGFP solubilised and obtain a saturation curve. Around 500 mM urea was shown to be required to release half of the SgsE-EGFP monomers from the S-layer

sheets. This information was used throughout the project whenever purified and then dissolved S-layers were needed.

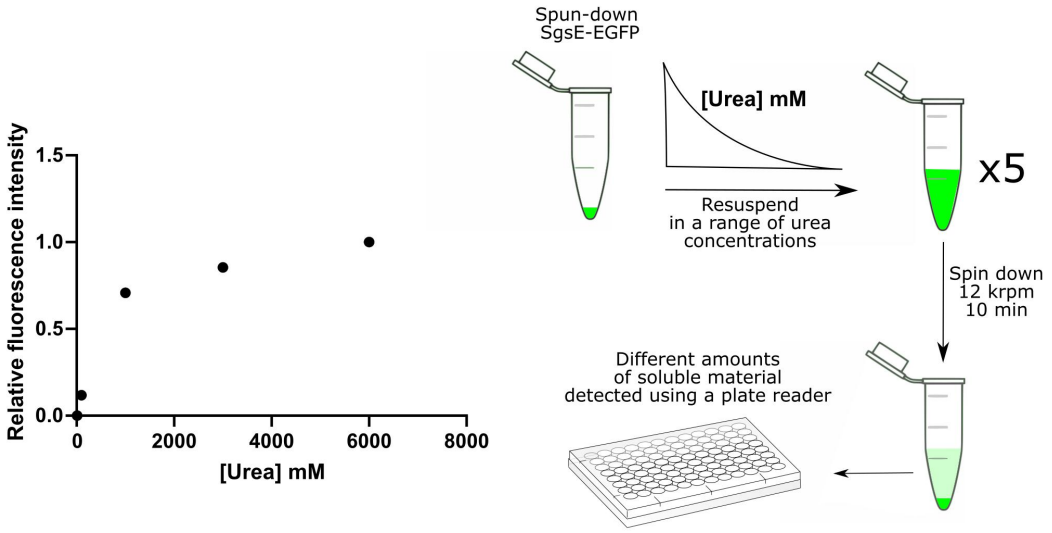


FIGURE 5.8: S-layer urea stability experiment. Left: a plotted fluorescence curve based on variable urea concentrations. Right: graphical explanation of the experimental design. A tube with pre-assembled SgsE-EGFP S-layer was spun-down at 10 krpm for 10 minutes in a separate tube for each urea concentration to be tested. The settled S-layer pellets were resuspended in variable urea concentrations and supernatants containing different amounts of solubilised SgsE-EGFP monomers were measured in a plate reader (excitation 488 nm, emission 509 nm).

5.4 Attempts at Visualisation of Interaction Between SgsE-A3 and Nanoparticles

Since the ultimate goal of the project is to obtain visual proof of protein-based nanoparticle patterning, direct visualisation of SgsE-STOP and SgsE-A3 interactions with magnetite nanoparticles was attempted. Magnetite with mean particle size of 10 nm was chosen as this size would be the most compatible with SgsE S-layer periodicity. The first attempts to visualise binding are shown in Figure 5.9.

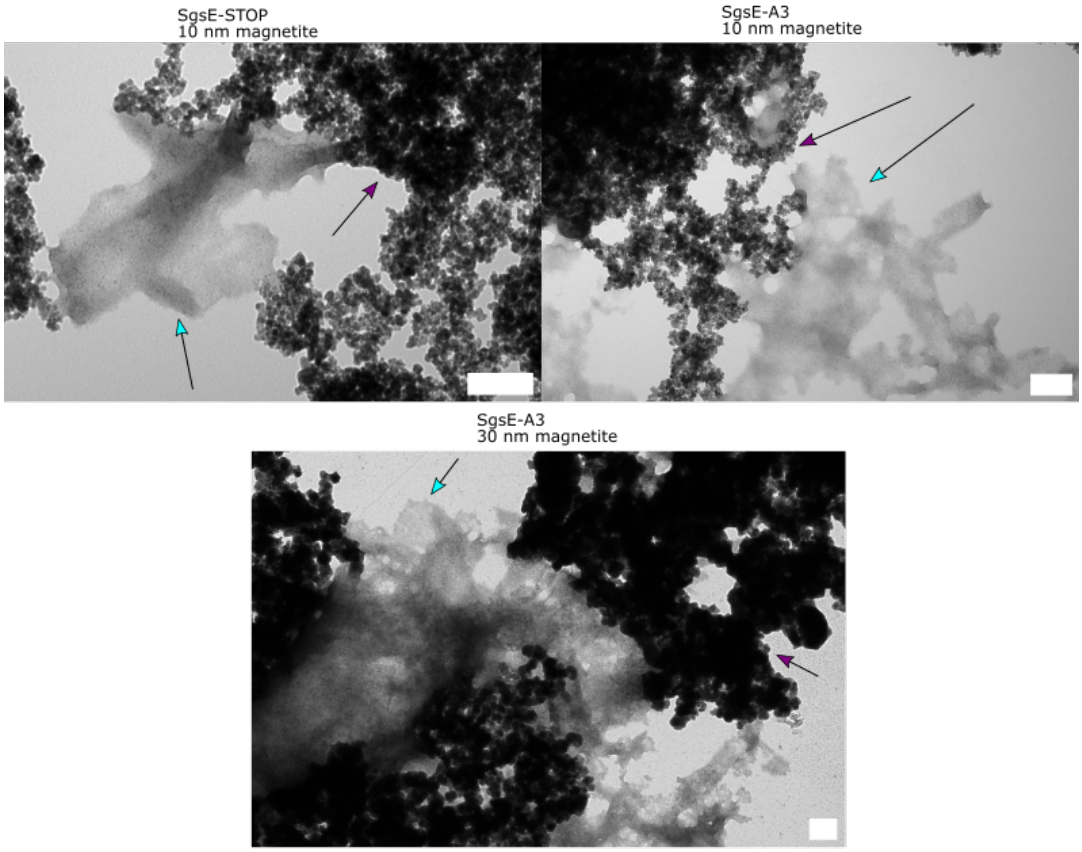


FIGURE 5.9: TEM images of attempts to bind 10 nm and 30 nm magnetite to SgsE-STOP and SgsE-A3. Particle solutions were ultrasonically treated for 1 minute at 40 % amplitude and mixed with 500 μ g/ml of each protein an hour in advance of application to freshly hydrophilised carbon coated copper grids. Scale bars - 100 nm. Arrows indicate what is expected to be SgsE S-layer assemblies (cyan) and magnetite nanoparticles (purple).

The images show large aggregates of 10 nm sized magnetite nanoparticles bound by large S-layer protein aggregates. Particles 30 nm in size were also attempted to the same effect. The S-layers appear faint (low in contrast) as the images are not stained. Staining was omitted as the contrast obtained from the uranyl formate stain may obscure presence of the magnetite nanoparticles on the S-layers. It is most likely that any interaction between the two types of aggregates is not-specific and is not enabled by the presence of the A3 peptide in the case of SgsE-A3. Clearly, any specific interaction between an SgsE-A3 binding site and an individual particle would be out-competed by particle-particle aggregation caused by magnetic attraction. To overcome this, a set of sample preparation conditions was tested for binding of 10 nm magnetite particles (Figure 5.10).

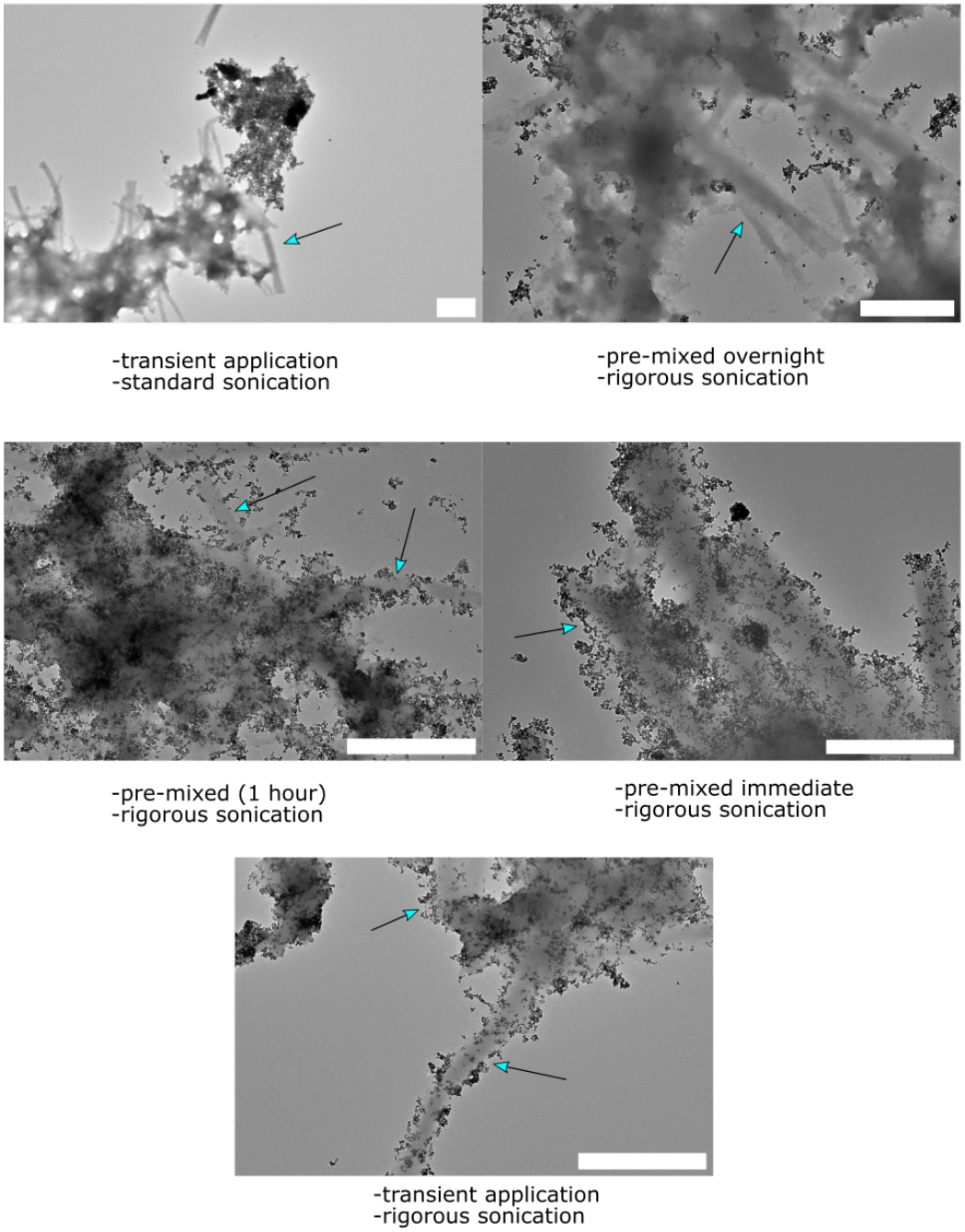


FIGURE 5.10: TEM images of attempts to bind 10 nm magnetite to SgsE-A3. Transient application refers to the S-layer protein being applied to a freshly hydrophilised carbon coated copper grids for 1 minute then washed and incubated with a particle solution for 1 minute before washing and drying. Standard sonication refers to 1 minute ultrasonic treatment at 40 % amplitude. Rigorous sonication refers to particle solutions being ultrasonically treated for 3 minutes at 70 % amplitude, the resuspended supernatant being moved to a fresh tube and the sonication protocol being repeated. Pre-mixing refers to combining various particle solutions with the S-layer protein solutions prior to application to freshly hydrophilised carbon coated copper grids. All samples were unstained. Arrows (cyan) indicate where SgsE assembled rods are. Scale bars represent 1 μm.

Two variables were identified to be important to achieving single particle visualisation under the electron microscope: extent of ultrasonic treatment and time from ultrasonic treatment to sample being dried on the hydrophilised carbon coated copper grids. First, an attempt was made to use the same ultrasonic treatment as before (1 minutes at 40 % amplitude), but apply the particles directly to the protein covered grids (as opposed to pre-mixing them in solution before grid application), this preparation was termed 'transient'. The preparation produced large protein aggregates bound to large nanoparticle aggregates. A rigorous ultrasonic treatment was thus developed where particles were ultrasonically treated for 3 minutes at 70 % amplitude. The supernatant was then separated from any undissolved particles and the ultrasonic treatment was repeated two more times, this was termed 'rigorous' sonication. 'Rigorously' treated particles pre-mixed with S-layers in solution overnight showed improved results, with particle aggregates being smaller and attached to various locations around the S-layer aggregate. The ability to obtain individual particles attached to the S-layers was progressively improved with decreasing the time of incubation of particles with S-layers. Finally 'transient' application of 'rigorously' ultrasonically treated magnetite nanoparticles was shown to produce the least aggregated particle distributions around the S-layers. The identified conditions were used to compare magnetite particle binding to SgsE-STOP and SgsE-A3 (Figure 5.11).

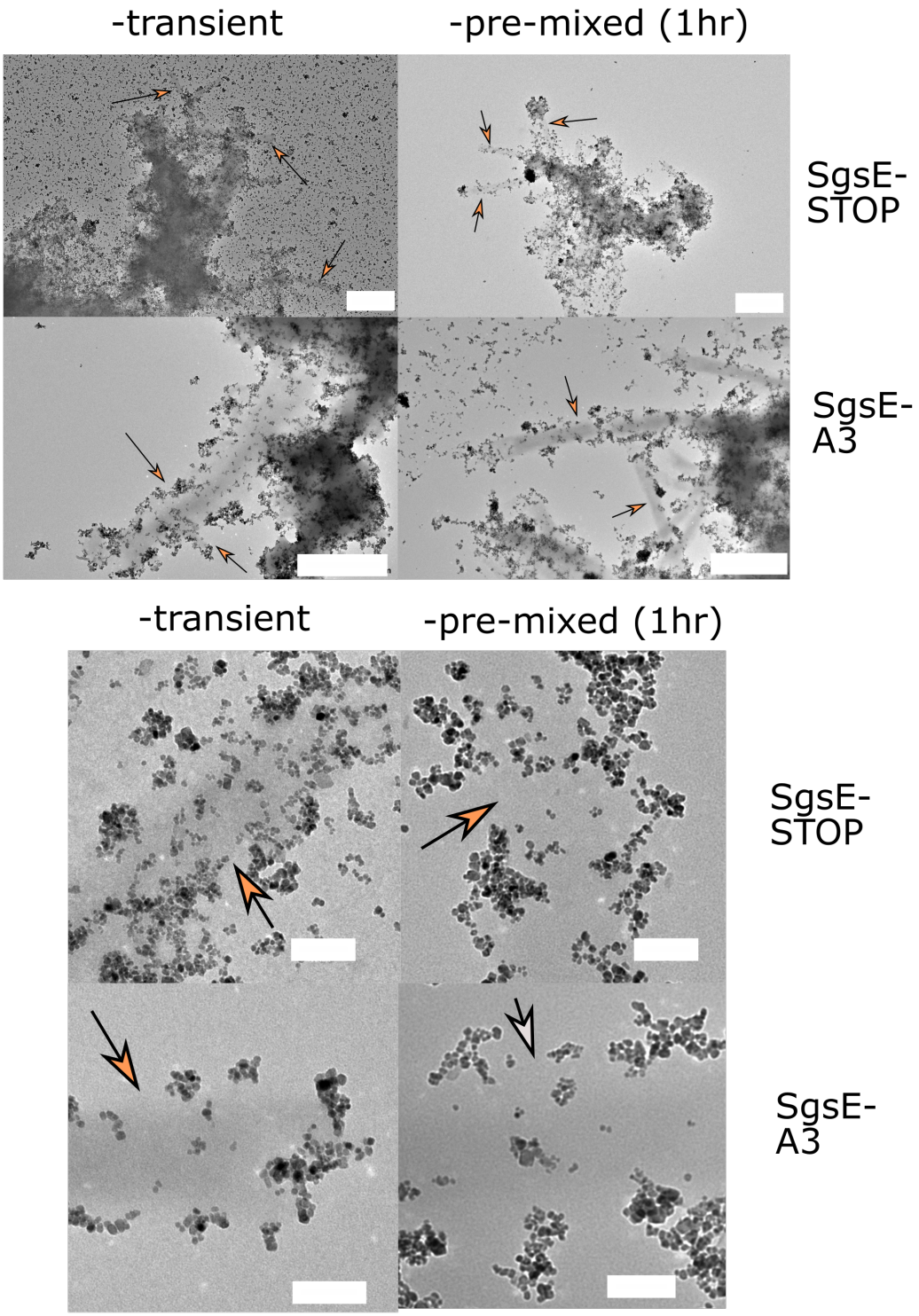


FIGURE 5.11: TEM images of comparing SgsE-STOP with SgsE-A3 and their interaction with 10 nm magnetite nanoparticles. Transient application refers to the S-layer protein being applied to freshly hydrophilised carbon coated copper grids for 1 minute, then washed and incubated with a particle solution for 1 minute before washing and drying. The particles used followed a 'rigorous' sonication where particle solutions were ultrasonically treated for 3 minutes at 70 % amplitude, the resuspended supernatant was moved to a fresh tube and the sonication protocol was repeated. Pre-mixing refers to combining the particle solutions with the S-layer protein solutions prior to application to freshly hydrophilised carbon coated copper grids. All samples were unstained. Scale bars - 1000 nm, except bottom-left where it is 100 nm. Orange arrows indicate where assembled S-layer rods are located. Bottom: higher magnification areas from images above. Top panel scale bars 1000 nm, bottom - 100 nm.

Surprisingly, both S-layer types show particles attached along the lengths of aggregates and rods. This was unexpected as SgsE-STOP does not contain the A3 binding peptide and thus is not expected to interact with the nanoparticles. An attempt to differentiate types of binding between the two S-layers can be seen in Figure 5.12.

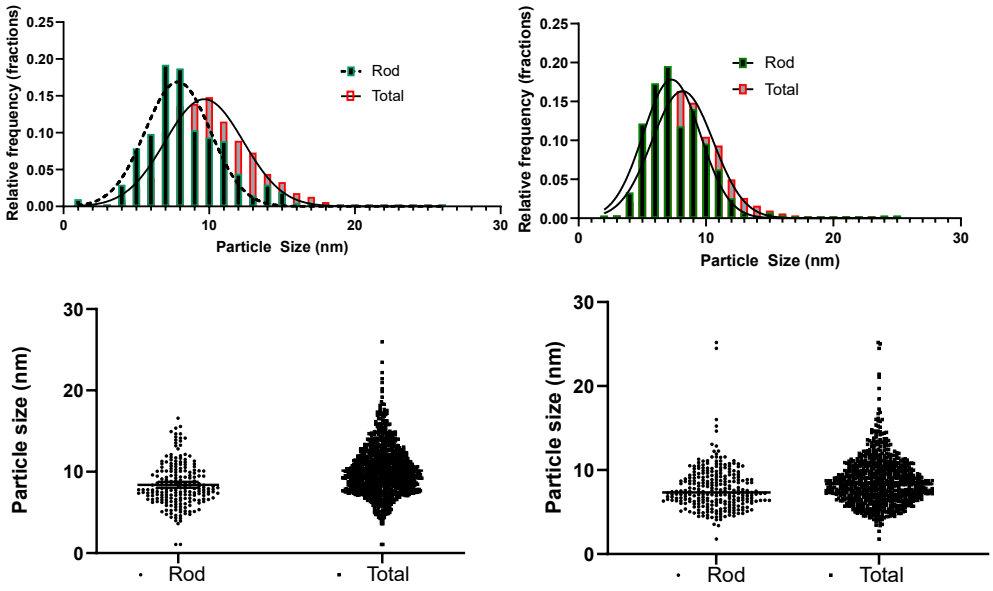


FIGURE 5.12: Attempted quantification of SgsE-STOP and SgsE-A3 particle binding as evidenced by microscopy. Comparison of numbers and sizes of particles found on S-layer rods and particles found non-specifically bound to the hydophilised carbon coated copper grids. The analysis followed identification of S-layer rods on each image and counting particles and their sizes inside the designated areas (n=200). This was then followed by counting and sizing of particles outside of the bounds of the rods (n=1000).

The data show that in both SgsE-STOP and SgsE-A3 samples there is a slight bias towards smaller particles to be found on the areas where S-layer rods can be seen as compared to the general size distribution in a given image. There is also a slightly smaller size distribution on rods in both cases. What this may mean is that the small sub 10 nm particles are being selectively templated on the S-layer rods, even though the periodicity cannot be clearly seen. This may possibly be due to sample preparation conditions not being optimal yet. The particle coverage on the rods is not full, thus periodicity might not be visible. Furthermore, the analysis is made more difficult by the fact that, so far, only rods have been visualised interacting with the particles. The protein purification method used seems to be strongly biased toward the assembly of rods instead of sheets. Any particle pattern periodicity present may be distorted by the fact that the rod patterns will be curved in difficult to predict

trajectories. A close-up image of one of the nanoparticle coated rods can be seen in Figure 5.13.

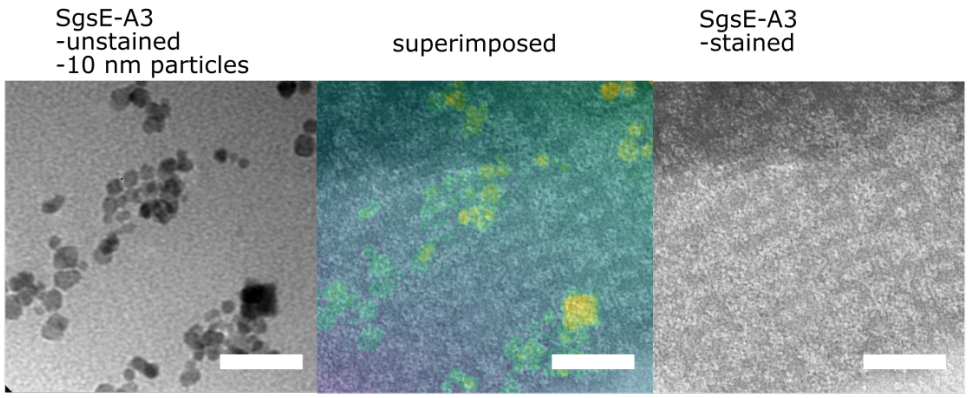


FIGURE 5.13: Close-ups of SgsE-A3 incubated with 10 nm magnetite nanoparticles juxtaposed with an SgsE-A3 negative stain preparation. Scale bars represent 100 nm.

Here, a close-up image of a SgsE-A3 rod shows some particles fall into what looks like orderly arrangements. For comparison, an image from a negatively stained sample is superimposed. While this is by no means proof that the particles in the first image are indeed being specifically patterned, it nevertheless shows that the dimensions of the observed particle patterns and the actual expected underlying S-layer patterns are in close alignment.

5.5 SgsE Construct Biomineralisation and Particle Binding

5.5.1 SgsE-Mms6 in Biomineralisation

One of the main premises of Chapter 5 was that SgsE-Mms6 S-layers would retain the function of Mms6 with respect to controlling magnetite formation by nucleating iron ions. By doing so, patterning of magnetic material may be achieved by specific biotemplation on Mms6 moieties of SgsE-Mms6 S-layers. This hypothesis was first tested by carrying out magnetite room temperature co-precipitation synthesis reactions with added SgsE-Mms6 S-layers (Figure 5.14).

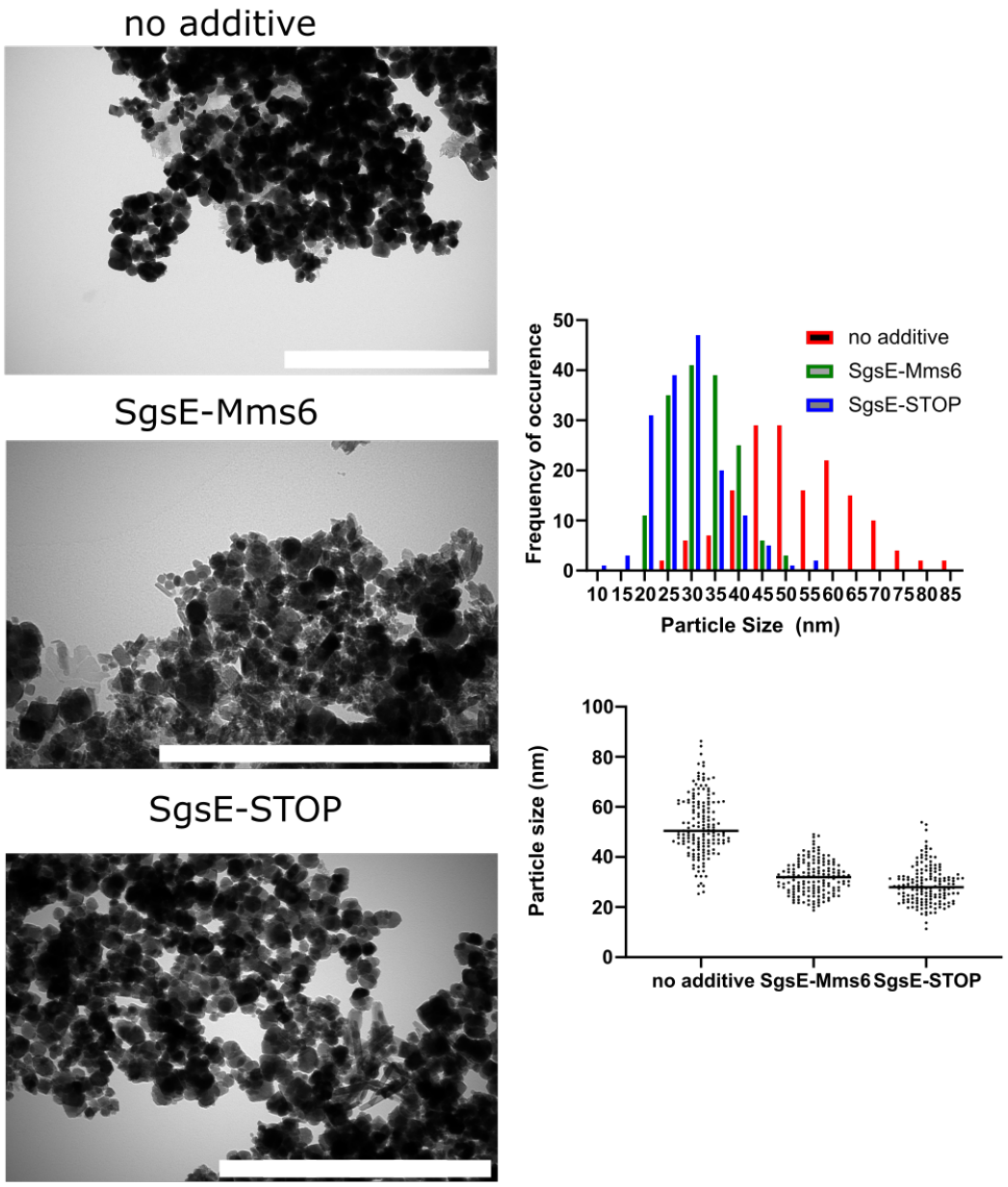


FIGURE 5.14: Room-temperature co-precipitation reactions for magnetite synthesis with SgsE-STOP and SgsE-Mms6 as additives. Top: unstained transmission electron microscopy images. Scale bars: 500 nm. Bottom left: particle size estimation from particle diameter measurements, n=160. Bottom right: same data as a swarm plot, n=160.

The resulting magnetite nanoparticles analysed on TEM reveal that adding SgsE-Mms6 and SgsE-STOP (as a control) decreases average particle size in both cases (50 nm with no additive to 30 nm with S-layer additives). This may imply that there is no Mms6 specific influence on particle synthesis and that the decrease in nanoparticle size is caused by the increase in nucleation sites corresponding to the porous

S-layer sheets. Synthesis products with SgsE-Mms6 S-layers as additives do exhibit a slightly narrower size distribution than when SgsE-STOP was used suggesting potential specific size control.

Batch magnetite synthesis reactions were also performed in the presence of assembled S-layers of SgsE-Mms6, SgsE-STOP and SgsE-A3 spotted on nitrocellulose membranes (Figure 5.15).

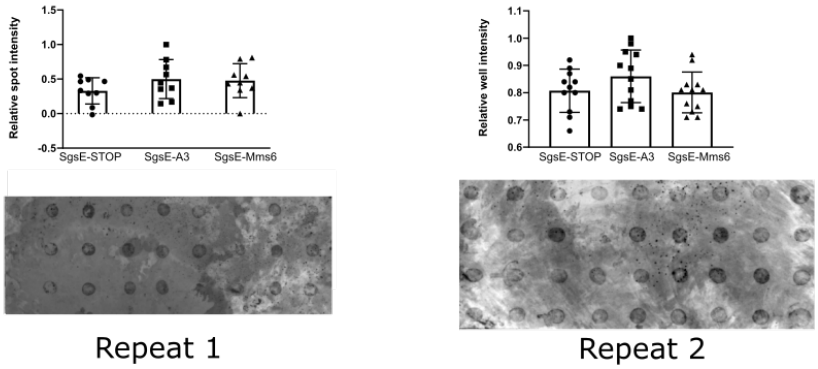


FIGURE 5.15: Batch magnetite synthesis reactions on SgsE-STOP, SgsE-A3, SgsE-Mms6 spotted nitrocellulose membranes. A 100 ml 50 mM Iron3+/Iron-total=0.5 solution was incubated with membranes spotted with 5 µl of each protein at 500 µg/ml for 40 minutes. A 20 ml solution of 1M NaOH was the added and membranes were visualised within 15 minutes. Quantification was performed by measuring spot intensities using ImageLab volume measurement tools and subtracting an average of three same diameter spots found locally around the measured spot. Error bars are standard deviations. Bottom: two images of the membranes obtained. Repeat 1 - n=9, repeat 2 - n=12.

The membrane results showed that SgsE-A3 S-layers retain more material than SgsE-Mms6 or SgsE-STOP, however the experimental noise was too high to make any conclusions. The initial expectation was that SgsE-Mms6 would show a stronger signal under these experimental conditions because of its ability to nucleate iron. It was hypothesised that this would anchor the synthesis products co-localised with SgsE-Mms6 spotting. It appears that a more important factor for material deposition in this experimental setting was the ability to capture the particles that formed in solution resulting in SgsE-A3 immobilising slightly more material.

5.5.2 SgsE Construct Differential Binding to pre-formed Nanoparticles

Assembled SgsE-STOP, SgsE-A3 and SgsE-Mms6 S-layers were investigated in a series of experiments for their ability to interact with pre-formed magnetite nanoparticles. The experimental series was intended to validate that either SgsE-A3 or SgsE-Mms6 S-layers can specifically immobilise magnetic material more so than SgsE-STOP. First, a membrane binding experiment was carried out with pre-formed 10 nm magnetite nanoparticles (Figure 5.16).

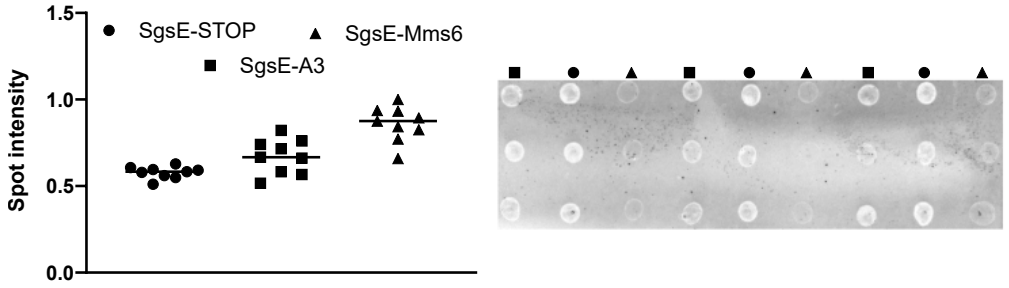


FIGURE 5.16: Pre-formed 10 nm magnetite particle spotted nitrocellulose membrane for SgsE-STOP, SgsE-A3 and SgsE-Mms6. The membrane was blocked with 3 % BSA. Well sonicated particles were incubated for 16 hours in room temperature. Right: obtained membrane image. The proteins were spotted in repeated columns - the shapes indicate which protein represents the column. Left: a swarm plot of quantified spots, n=9.

The results identified SgsE-Mms6 to be the best binder followed by SgsE-A3 and then by SgsE-STOP. The significance tests show statistical significance between SgsE-STOP and SgsE-A3 signals (*) and between SgsE-A3 and SgsE-Mms6 (**). Same experimental conditions were then also used to perform a membrane binding experiment using pre-formed 30-50 nm magnetite nanoparticles (Figure 5.17).

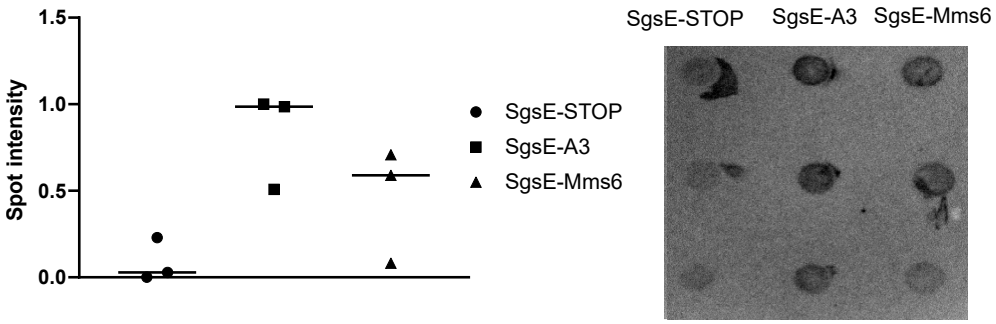


FIGURE 5.17: Pre-formed 30-50 nm magnetite particle spotted nitrocellulose membrane for SgsE-STOP, SgsE-A3 and SgsE-Mms6. The membrane was blocked with 3 % BSA. Well sonicated particles were incubated for 16 hours in room temperature. Left: swarm plot of the quantified spots, n=3. Right: image of the obtained membrane.

In both cases when incubating with 10 nm and 30-50 nm sized particles, SgsE-STOP shows the lowest amount of binding as expected, since it does not possess a specific binding interface. SgsE-Mms6 seems to bind to more 10 nm magnetite nanoparticles than SgsE-A3, whereas this relationship is inverse for binding of 30-50 nm sized magnetite nanoparticles. This may mean that Mms6 presented on an assembled

S-layer has high affinity for small magnetite nanoparticles and that larger nanoparticles do not fit a formed binding pocket on the Mms6 protein. In such a scenario SgsE-A3 would have the advantage by only interfacing with the particles through a structurally dynamic charged linear peptide. A series of magnetite ELISA assays was also performed to corroborate these results (Figure 5.18).

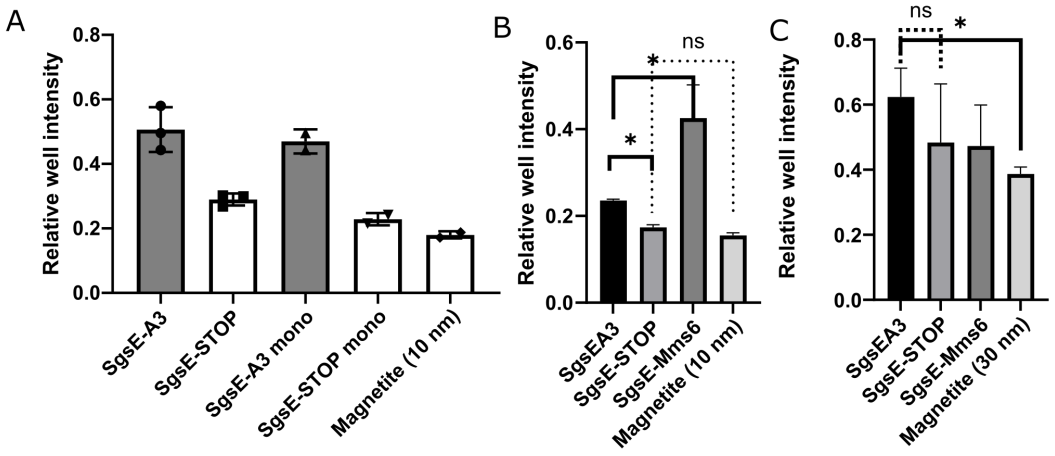


FIGURE 5.18: **A** Magnetite ELISA assessing binding of SgsE-STOP and SgsE-A3 to 10 nm magnetite nanoparticles. SgsE-STOP and SgsE-A3 (as labelled on the graph) are assembled S-layer samples incubated with the particles. '-mono' stands for incubated monomers (obtained by centrifugation of assemblies at 12.5 krpm for 10 minutes and using the supernatant). **B** A repeat of the assay with assembled S-layers and included SgsE-Mms6 with 10 nm magnetite nanoparticles. **C** A repeat of the assay with assembled S-layers and included SgsE-Mms6 with 30-50 nm magnetite nanoparticles. Error bars - standard deviation (population), n=3.

In Figure 5.18**A** SgsE-STOP and SgsE-A3 are compared for their ability to bind 10 nm magnetite particles. Fully assembled and monomeric versions (obtained by centrifugation of assemblies and extracting the supernatant, respectively) of the proteins were tested. SgsE-A3 showed higher binding as monomer and as an assembly. Same assay was performed with assemblies only (Figure 5.18**B**), but with SgsE-Mms6 also included. Just like in Figure 5.16 SgsE-Mms6 binds the 10 nm particles more than SgsE-A3 and SgsE-STOP. An ELISA assay with 30-50 nm particles and fully assembled S-layers was also performed Figure 5.18C and the same trend could be seen as was in Figure 5.17, however the results had too much noise to fully assess the differences between SgsE-A3 and SgsE-STOP binding.

5.5.3 SgsE Construct Iron Ion Binding

The three S-layer constructs SgsE-STOP, SgsE-A3 and SgsE-Mms6 were also compared for their ability to bind iron ions. An iron binding-luminol assay is shown in Figure 5.19.

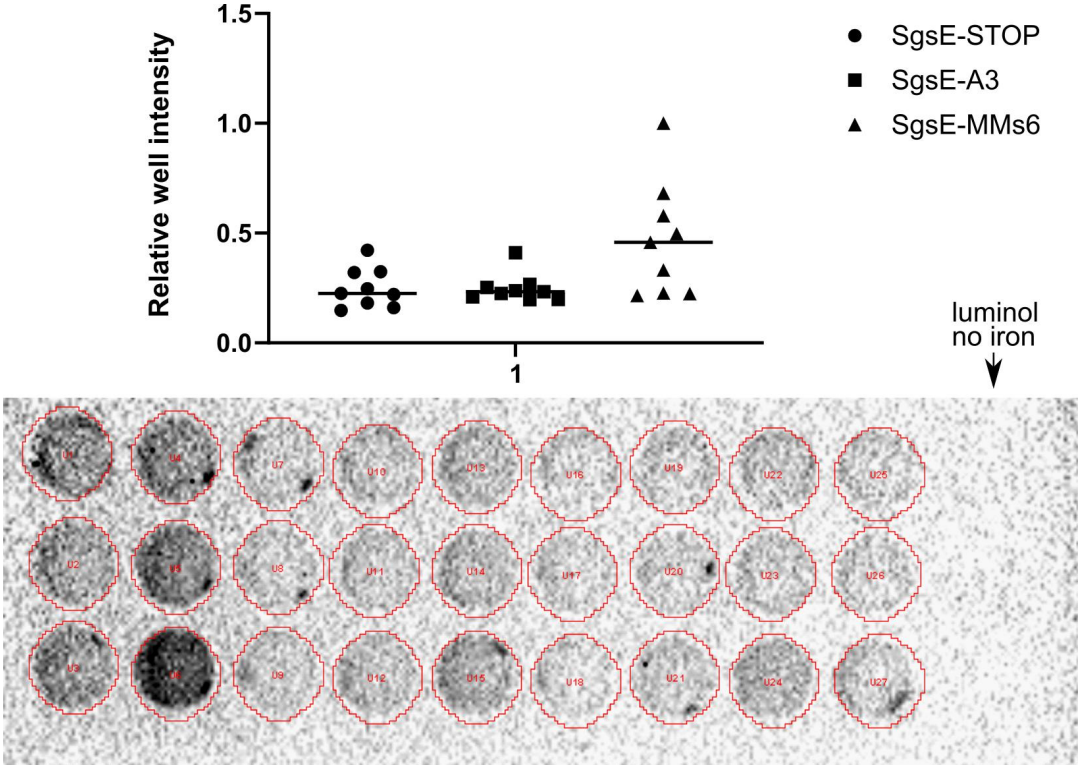


FIGURE 5.19: Luminol assay to detect iron binding to SgsE-STOP, SgsE-A3 and SgsE-Mms6. 100 μl of SgsE (A3; STOP; Mms6) at 500 µg/ml was distributed into three tubes each, spun-down and mixed with 1 mM iron ($Iron^{3+}$/Iron-total=0.5) for 1 hour. Each incubation was washed three times, denatured with 3M urea and mixed with a luminol solution and visualised immediately. The plate was inserted into the ChemiDoc visualiser and signal in each well was saved in a form on an image which was then quantified using ImageLab. Top: swarm plot of the obtained values, n=9.

The assay showed that SgsE-Mms6 absorbed the most material and that SgsE-STOP and SgsE-A3 showed approximately equal amounts (most likely non-specific) binding. A membrane assay to assess iron binding was also performed (Figure A.15) which replicated the SgsE-Mms6 binding result, but interestingly, also showed SgsE-A3 binding slightly stronger than SgsE-STOP. The implication of the results is that in the context of an assembled S-layer protein SgsE-Mms6 seems to retain one of the suspected functions of Mms6 which is to bind iron ions. This result may form the basis of studying ion nucleation proteins on S-layer sheets.

5.5.4 Conclusions

The work outlined here demonstrates another example of S-layer protein SgsE being used as a fusion protein for attempts at arriving at a nanotechnological application. The protein is successfully modified with a Strep-tag, A3 magnetite binding peptide, Mms6 and EGFP and retains its ability to assemble onto extensive sheets and rods.

SgsE-Mms6 is shown to be able to bind iron ions and small (10 nm) nanoparticles. This may form a basis for SgsE (or perhaps other S-layers) to be used as a platform for studying biomineralisation. While the expression of Mms6 was the most difficult of all of the fusion proteins used, the production of this membrane protein fusion was achieved without altering the expression and purification protocols used for other fusions.

SgsE-A3 is demonstrated to interact with nanoparticles specifically in binding assays when compared to SgsE-STOP. TEM results however complicate the picture, revealing that both versions of the protein produce rods that are coated with nanoparticles. This discrepancy suggests that the nanoparticle coating observed on TEM is most likely originating from nonspecific interactions between the particles and S-layer periodical cavities. It may well be that there is specific SgsE-A3 facilitated templating taking place, but is not visible via protocols attempted here. Sufficiently challenging conditions for particle assembly on to S-layer need to be created to observe only specific A3 peptide mediated binding (akin to the intensive blocking conditions in the binding assays). Initial direct visualisation attempts focused on SgsE-A3 rather than SgsE-Mms6 due to concerns about Mms6 fusion effects on the lattice structure and a lower overall purity of the protein preparations. The membrane binding and magnetite ELISA results, however, reveal that SgsE-Mms6 looks like a much stronger binder for the small 10 nm magnetite particles.

These results show promising leads for S-layer based nano-patterning through specific particle binding moieties. To visualise bio-templated patterning based on these fusion proteins one would need to optimise S-layer assembly protocols to obtain

more planar surfaces, either in solution or by on-surface assembly. This may make the elucidation of any patterns forming more clear. Considering that SgsE-Mms6 was the best binder for 10 nm magnetite, on would need to optimise the expression and purification protocols for this protein to remove impurities. The TEM results showed a distorted two-dimensional lattice of assembled SgsE-Mms6 S-layers removal of impurities may restore the correct lattice. However, it is also possible that Mms6 hydrophobic surface intrinsically cannot be arrayed in such a manner. It that case, perhaps, the improved binding function (over the linear A3 peptide) may be transferred to a more workable scaffold that is more soluble and would not disrupt S-layer lattice assembly. Chapter 6 starts with an introduction to such protein scaffolds.

Chapter 6

Results: Prospecting for new Binding sequences

6.1 Chapter 6 Summary

Chapter six is a contribution to discovering and characterising nanoparticle binding peptide sequences beyond the A3 peptide discussed in Chapters 3 and 5 (and included in the design in Chapter 4). The results describe work on a magnetite binding peptide E8 and evaluation of its binding to the nanoparticles when it is engineered into different peptide displaying scaffolds in Part 1. This is followed by a phage display study to identify new binding peptide sequences specific to cobalt platinum nanoparticles in Part 2. The explorations in alternative magnetic nanoparticle binding moieties could in future be used together with the self-assembling systems described in previous chapters for bit-pattern media assemblies.

6.2 Part 1 Introduction: to Peptide Loop Exposing Protein Scaffolds

6.2.1 Adhiron Binding Protein

Previous studies in the Staniland group have discovered a magnetite binding peptide, termed E8 [172]. The E8 peptide exhibits specificity for magnetite nanoparticles and was the only peptide out of a pool of screened magnetite binding peptides that did not bind DNA molecules making it a useful component in engineering magnetic molecular devices in biological settings (unpublished Staniland group results). The peptide sequence was determined to be: AHMYTKAQT. The peptide was discovered using a phage display library containing a scaffold protein - adhiron (Figure 6.1).

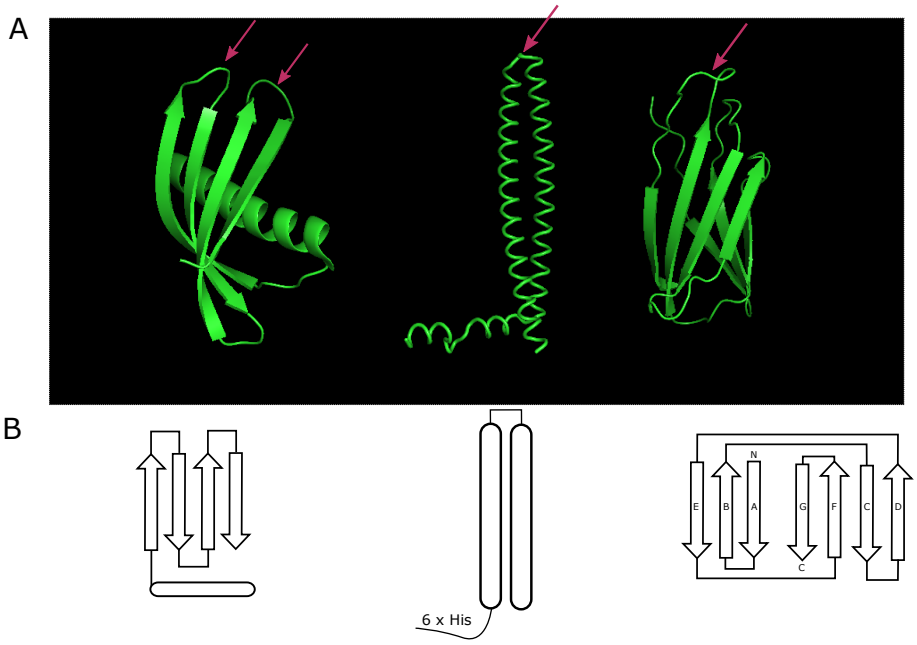


FIGURE 6.1: **A** 3D structural models for adhiron (4N6T), coiled-coil (trRosetta model) and FN3 (1TTF). Arrows indicate variable loops for binding. **B** structure schematics of the models above. Arrows represent β-strands, rounded squares - α-helices. Structure visualisations generated using PyMol.

Adhirons (more lately known as 'affymers') are loop displaying scaffold proteins derived from a consensus sequence of distantly related phytocystatin proteins found in plants [171]. They are small, monomeric, highly soluble and stable, furthermore they lack glycosylation sites and disulphide bonds. This makes them to be attractive binding loop displaying scaffolds. The initial study created a library of

1.3 x 10^{10} clones containing two randomised loops that were then screened against a SUMO protein. The emerging enriched sequences within the two loops showed high sequence identity to known SUMO protein binding motifs thus indicating that the phage library was capable of producing loop exposing scaffolds that bind to the desired material. The same library was later used to search for sequences that can interact with cubic magnetite nanoparticles [172] (the study focused on sequences for usage as biomineralisation additives, the E8 sequence is only listed in the supplementary information). The study discovered a series of magnetite interacting peptides that were then shown to influence room temperature co-precipitation reactions for production of more homogeneous (in size and shape) cubic magnetite particles. The same study also showed that when using the same peptides outside of the context of an adhiron scaffold the same size control effect was not exhibited [172]. This suggested that having the peptides constrained in a scaffold is important for their function. Since the E8 peptide containing adhiron possessed interesting binding qualities there was interest to test it in different contexts. To test that the peptide can be used as a modular part for protein engineering, binding capacity had to be assessed outside of the original adhiron scaffold: as a free peptide and as a peptide constrained in different scaffold proteins. As a free peptide, E8 was tested in an ELISA assay (results from Dr Lori Somner's thesis) which showed specific binding compared to a random amino acid sequence (albeit at lower intensity than when in an adhiron scaffold). Testing the peptide in a different scaffold would show that increased binding capacity is not emerging from a synergy between the qualities of the peptide and the scaffold itself (one can rule out the scaffold binding to the particles by itself as that has been tested [172]). If a peptide is to be an agent considered to solely confer binding capacity to a scaffold protein (by replacing the loop region with the said peptide) it needs to be shown to function in at least three different scaffold proteins independently (the adhiron counts as the first demonstration of that).

6.2.2 Coiled-coil Peptide Display Scaffold

The second scaffold was chosen to be a coiled-coil (Dr Andrea Ralwings' and Dr Lori Somner's design). As discussed in Chapter 1, coiled-coil primary sequence to structure relationship is so well understood that it is possible to design novel

loop exposing scaffolds with relative ease. Early examples of such designs were used to expose binding loops for integrin αIIβ3 [290] and immuno-competent cell surface antigens [291]. Both studies created heterodimeric coiled-coils design to assemble in an anti-parallel fashion. Both coiled-coil components were located on the same amino acid chain and were separated by a binding sequence. As the coiled-coil forms, the binding sequence (beginning and ending with helix breaking amino acids G and P, respectively) get 'looped out'. Both studies managed to achieve specific interaction with intended target molecules. Staniland group have designed a homodimeric intrachain anti-parallel coiled-coil scaffold by adapting design principles elucidated in Gurnon et. al., 2003[292]. The structure for which can be seen in Figure 6.1. The scaffold was shown to be effective at displaying peptide loops found in the lumen of magnetotactic bacteria magnetosomes (magnetite associated proteins Mms13 and MmsF) [209]. These loops are normally presented as solution exposed linker regions for transmembrane domains. A major bottle neck in studying and utilising proteins like Mms13 or MmsF is their transmembrane nature. The coiled-coil scaffolds MmsFcc and Mms13cc displaying their respective peptide loops formed soluble, easily expressible 7 nm rod proteins. Both constructs showed specific interaction with magnetite nanoparticles (compared to a coiled-coil construct with a transmembrane connecting loop region from acriflavin efflux protein AcrB from *E. coli*, termed AcrBcc, not expected to have affinity for magnetite) and allowed production of more magnetic magnetite nanoparticles when included as additives in room temperature co-precipitation reactions of iron. These results gave precedence for this scaffold to be useful for displaying the novel magnetite interacting peptide E8. The peptide was cloned into the coiled-coil scaffold (construct named E8cc - Figure 6.1) and was evaluated using a magnetite adapted ELISA in Lori Somner's thesis showing specific interaction with magnetite when compared to AcrBcc negative control.

6.2.3 Monobody Peptide Display Scaffold

Building on the work done prior, this thesis introduces a new E8 peptide loop displaying construct that utilises a fibronectin domain 3 (FN3) (later coined as monobody [293] - a small antibody mimic). FN3 is a small autonomously folding domain

that resembles immunoglobulin domains (found in antibodies) (Figure 6.1). It has a β-sandwich structure composed of seven β-strands. The strands are connected via linker loops which were randomised to create a phage display library that was successfully screened against ubiquitin [294]. The initial study established FN3 as a plausible platform for binding loop display which exhibited useful properties like small size, high stability, lack of disulphide bonds or post-translational modifications and the ability to retain correct folding upon modification of the loop region. A series of studies that followed identified the loop tolerance for different length insertions [293], showed viability for fusion with oligomerisation domains to increase avidity [295], achieved 1.1 pM binding to other bio-molecules [296] and was utilised in FRET studies where the monobody was designed to bind organic dye molecules. A summary of monobody applications can be found in Chandler et. al., 2020 [297].

The work to fully characterise the E8 peptide in different scaffolds before the writing of this thesis can be seen in Table 6.1. Thus, part 1 of this chapter attempts to show that E8 peptide displayed on a monobody scaffold matches the binding capacity of E8cc and out-competes the monobody displaying the control loop.

Scaffold	Loop	Description	Reference
Adhiron	AHMYTKAQT	E8 peptide, magnetite binding	[172]
Adhiron	DWWEAGVFM	CTRL peptide, non-binding	[172]
Coiled-coil	AHMYTKAQT	E8 peptide, magnetite binding	Lori Somner's thesis

Coiled-coil	ENVERVMAEE GLPPKEATRKS MGQL	control peptide, non-binding	Lori Somner's thesis
Monobody	AHMYTKAQT	E8 peptide, magnetite binding	This work
Monobody	DWWEAGVFM	CTRL peptide, non-binding	This work

TABLE 6.1: Summarised work for displaying the magnetite binding peptide E8 on different scaffolds.

6.3 Part 1 Results: Magnetite Binding Peptide E8

6.3.1 Sequence Comparison of Coiled-coil, Adhiron and Monobody Constructs

Figure 6.2A shows full sequences of E8cc, Adhiron-E8 and MB-E8 (and their respective controls, denoted 'CTRL').

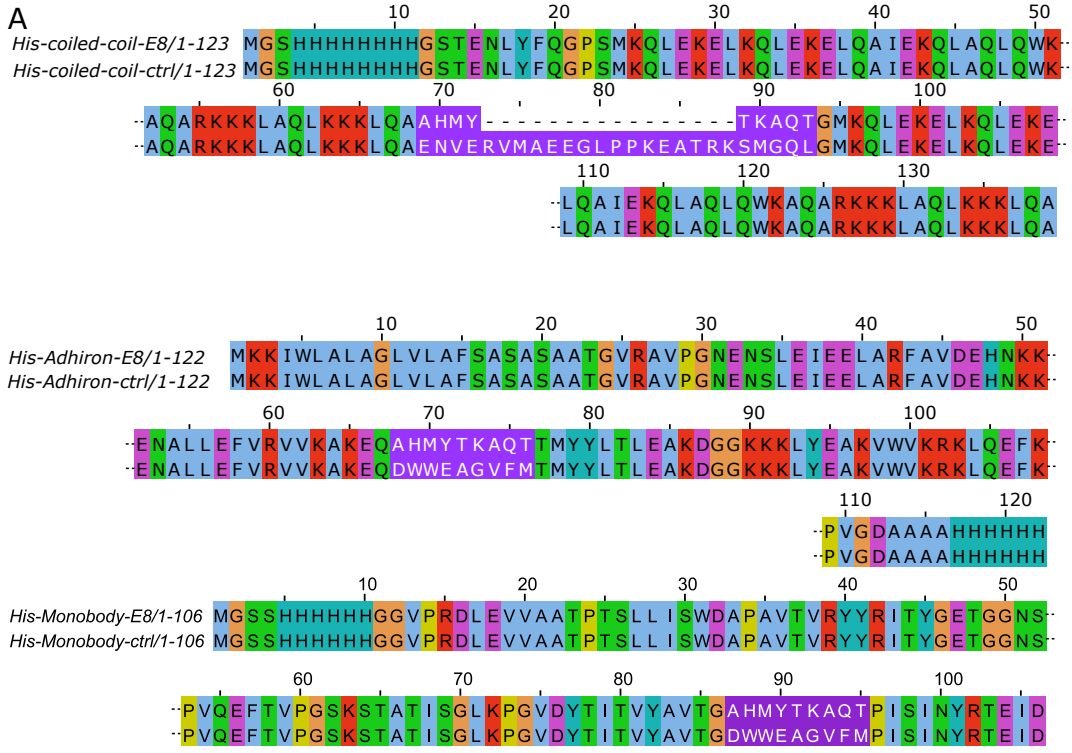


FIGURE 6.2: **A** Full sequences of different scaffolds containing the E8 magnetite binding peptide or the control non-binding sequence (highlighted in purple). Other amino acids highlighted based on the Clustal scheme - see appendix for the legend. **B** E8 and 'CTRL' sequences aligned for comparison.

All of the constructs posses either a 6 poly-histidine tag (Adhiron-E8, MB-E8) or an 8 poly-histidine tag (E8cc) for purification and ELISA purposes. For the design of the new monobody sequence, either the magnetite binding loop E8 (AHMYTKAQT) or the non-binding control sequence ('CTRL':DWWEAGVFM) were inserted between glycine and proline residues to replace the native sequence RGDSPASSK [293]. The 'CTRL' sequence has been shown not to bind magnetite in an adhiron scaffold [172]. All scaffolds exhibit relatively simple structures composed of short β-strands and/or α-helices (Figure 6.1). The comparisons between intended binding and non-binding loops are shown in Figure 6.2B. The 'CTRL' sequence has a higher percentage of aromatic and negatively charged residues, whereas the E8 peptide possesses histidine, lysine and methionine residues (previously implicated in nanoparticle binding [193]. Figure 6.3 shows predicted trRosseta models for MB-E8 and MB-CTRL.

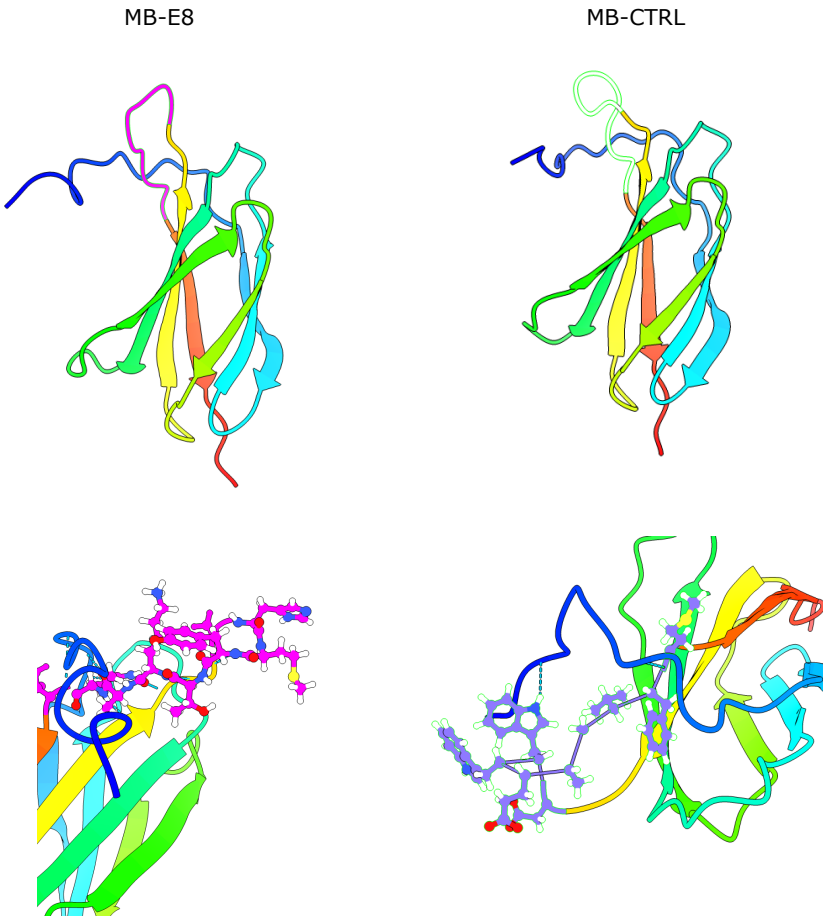


FIGURE 6.3: trRosseta models for MB-E8 and MB-CTRL. Loop regions highlighted in pink and white for E8 and CTRL, respectively.

Based on the homology based models the novel peptide loop replacements do not seem to disrupt the fold of the scaffold. trRosseta modelling accuracy metrics (not shown) indicate less certainty for the output structure for the loop (and the terminal 6 x His tag) regions to match the actual structures. This is not surprising as such loops in those positions are novel and thus would not be homologous with any previously determined structures.

6.3.2 Purification and Quality Control of the Monobody Protein.

The modified monobody sequence MB-E8 was codon optimised for expression in *E. coli* and cloned into a pET28a vector downstream of a T7 promoter (the sub-cloning was ordered as commercially available service from Genscript). A pET28a vector was obtained by using site-directed mutagenesis to replace the E8 loop with the CTRL loop. The plasmids were used to transform *E. coli* BL21 (DE3). Over-expression of the plasmid encoded proteins was performed as described in Chapter 2. The produced cell pellets were lysed by sonication , subjected to 1 ml Äkta His Trap Nickel columns with a subsequent gel filtration purification with elution in 20 mM phosphate buffer pH 7.4 using the standard methods as described in Chapter 2. Gel filtration fractions corresponding to the dominant peaks were analysed on SDS-PAGE (Figure 6.4A - only MB-E8 is displayed here, but same was repeated for MB-CTRL) and showed pure protein between 10 kDa and 15 kDa indicating the expression of the correct protein (the expected theoretical sizes for MB-E8 and MB-CTRL were 11440.7 Da and 11530.81 Da, respectively).

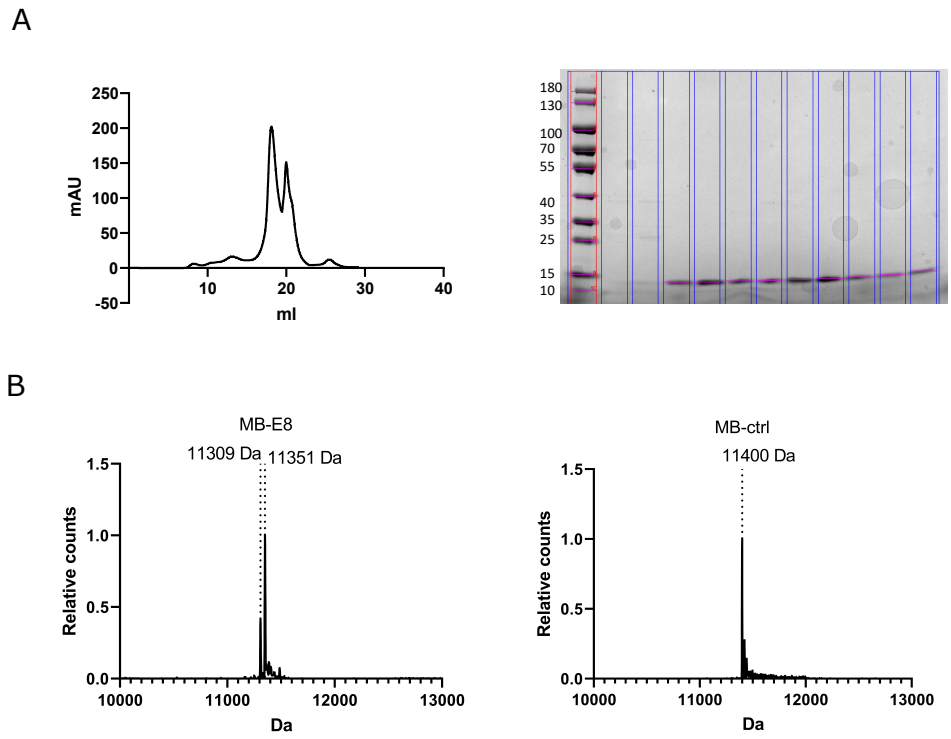


FIGURE 6.4: **A** Left: gel filtration chromatogram of MB-E8 on an SP200 column eluted in 20 mM phopshate buffer pH 7.4. Right: SDS-PAGE image of different fractions of the dominant peak corresponding to the chromatogram on the left. **B** ESI-TOF spectra of MB-E8 (left) and MB-CTRL (right). Relative counts - values of ESI-TOF that have been normalised to the highest value in each spectrum. Dotted lines indicate expected theoretical values.

MB-E8 gel filtration elution volume for the dominant peak corresponded to that of 11.8 kDa (when compared to a soluble globular protein calibration curve Chapter 2). The value closely matches that of the theoretical size. This indicates that the protein is mostly in a monomeric state. The gel filtration fractions of the dominant peaks were pooled and analysed on ESI-TOF (Figure 6.4B). Both purified MB-E8 and MB-CTRL showed presence of peaks matching expected theoretical sizes of their respective primary sequence with a removed methionine (MB-E8:11309.54 Da, MB-CTRL:11399.62 Da). MB-E8 mass spectrum shows a species (more than twice as prominent) 42 Da larger than expected the expected value. This may indicate that the E8 loop is binding to materials present during the purification process. The scaffolds used in this study were assessed for correct folding in 20 mM phosphate pH 7.4 buffer on CD (Figure 6.5).

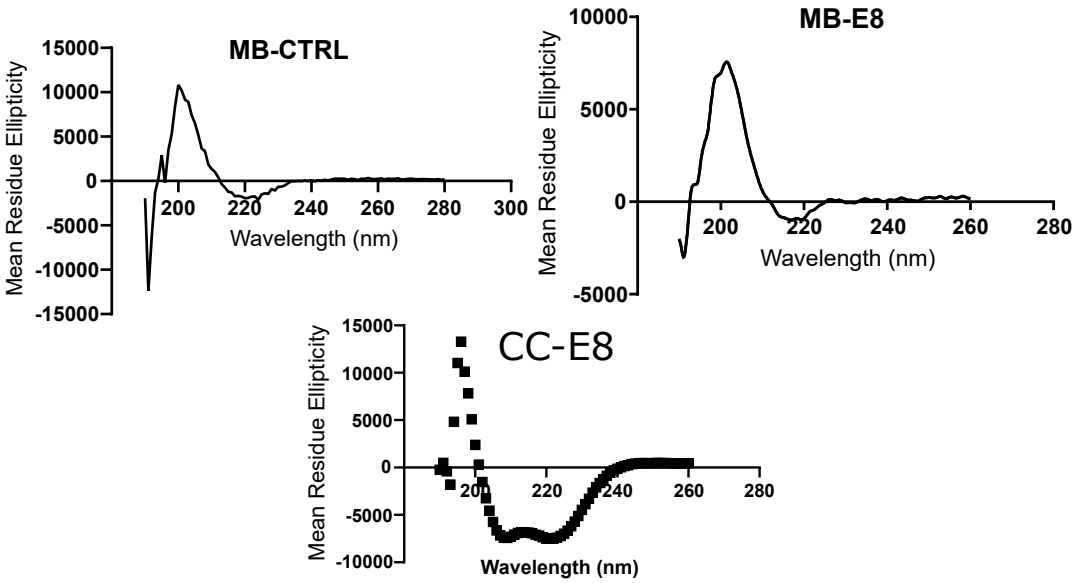


FIGURE 6.5: CD spectra of scaffolds used in the chapter. Spectra for MB-CTRL and MB-E8 were performed in 20 mM phosphate buffer pH 7.4. CC-E8 analysis was performed in ultra-pure water. Protein were analysed at 5 μM concentrations.

MB-E8 and MB-CTRL show similar spectra that resemble those found in literature (see supplementary materials in Hackel et. al., 2008[296]). Specifically, the spectra show negative peaks at 218 nm indicating presence of β-sheet content. These results give confidence that both control and magnetite binding scaffolds are folded correctly, thus most likely displaying the binding loops for binding as intended. E8cc was also purified (as described in Chapter 2) to act as a positive control for subsequent experiments. The CD spectra for E8cc show two negative peaks at 222 nm and 208 nm indicating a strong presence of α-helices. This suggests that the coiled-coil scaffold was also produced in a correctly folded form. Together these results suggest that correct scaffolds are produced in a folded form and that any results of binding (or lack thereof) would be caused by the displayed peptide loops.

6.3.3 Magnetite ELISA

To investigate the ability of MB-E8 and MB-CTRL to bind magnetite the proteins were tested in an initial magnetite ELISA experiment (Figure 6.6A) with a sample of magnetite synthesised using standard conditions (size 30-50 nm).

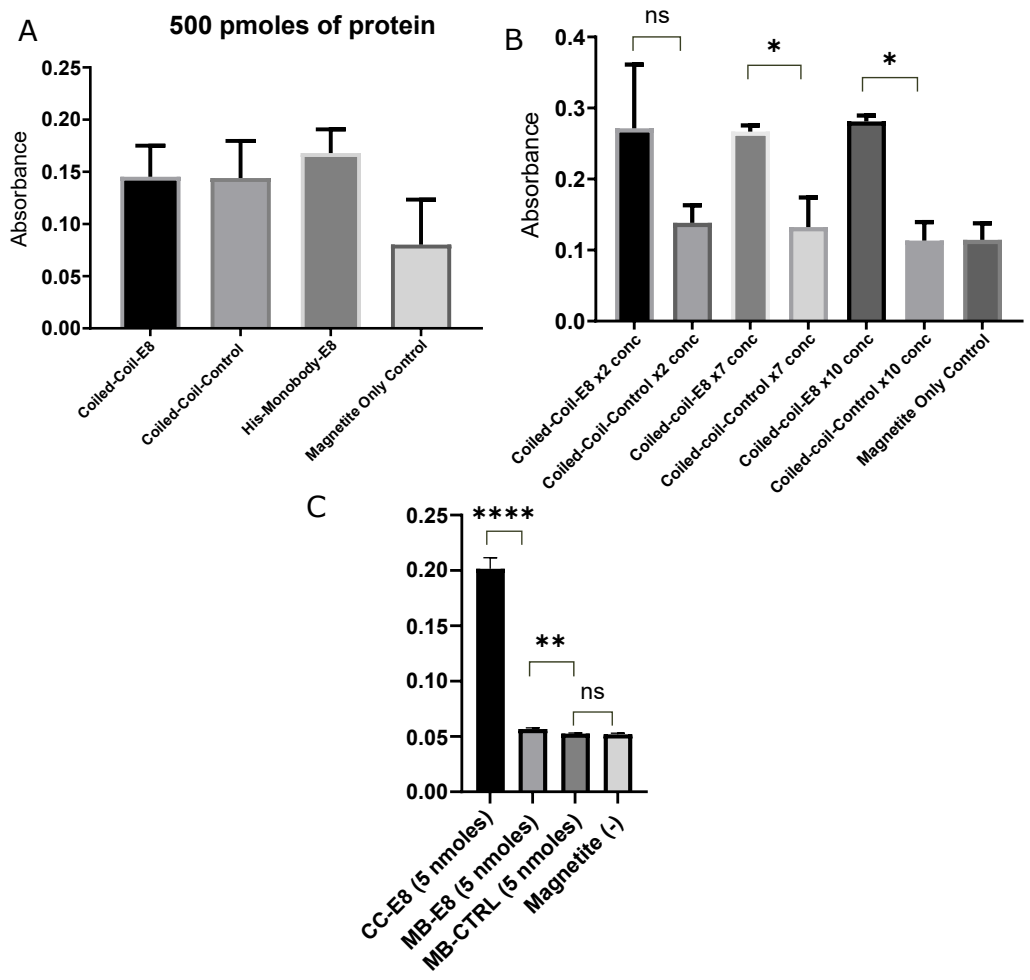


FIGURE 6.6: **A** The Fist magnetite ELISA with low (500 pmole in 250 µl) concentrations. **B** E8cc concentration study to find optimal ELISA concentrations. **C** Magnetite ELISA repeat with optimised conditions including CC-E8 (5 nmoles), MB-E8 (5 nmoles), MB-CTRL (5 nmoles). Mean values plotted, n=3. All error bars are standard deviations (population). Particles used: 30-50 nm magnetite.

The experiment was performed by adding a total of 500 pmoles of each protein to 250 µl of casein blocked magnetite nanoparticles at particle concentration of 3 mg/ml. No significant differences between E8cc, MB-E8 or control-CC could be seen. A concentration optimisation experiment was then performed with E8cc (acting as a positive control) shown in Figure 6.6**B**. Starting at the 1 nmole a clearer effect of binding was started to be observed, but was not yet sufficient to obtain a statistically significant result. Increasing the amount of E8cc to 3.5 nmoles showed the same amount of average binding, but with less uncertainty. Lastly, a 5 nmole sample showed the same amount of binding as one with 3.5 nm. An experiment with MB-E8 and MB-CTRL was then repeated with 5 nmoles of protein in the same volume and concentration of casein blocked magnetite nanoparticles (Figure 6.6**C**). The

results showed E8cc with a four times larger signal than MB-E8, MB-CTRL and magnetite only negative control. There was a small, but statistically significant increase in signal for MB-E8 over MB-CTRL and the magnetite control. The ELISA signal is strongly dependant on the ability of the primary antibody to access its poly-histidine tag target, thus a negative (or in this case minuscule) binding effect may not necessarily represent the true extent of magnetite binding capacity for the MB-E8 protein. The ELISA result suggested a need for an investigation into poly-histidine tag accessibility on the MB-E8 protein.

6.3.4 Assessing MB-E8's Antibody Binding Accessibility

To understand more about the differences of poly-histidine tag accessibility to antibody, MB-E8 and CC-E8 were spotted on a nitro-cellulose membrane at known concentrations and were probed with a poly-histidine specific antibody (Figure 6.7).

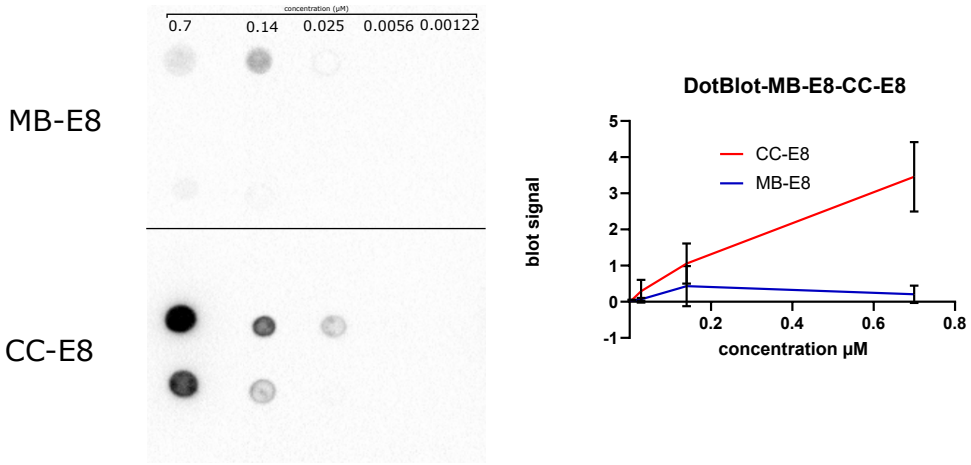


FIGURE 6.7: A dot blot experiment to determine poly-histidine tag accessibility/antibody binding propensity. MB-E8 and E8cc were spotted in duplicate, blocked and probed with anti-poly-histidine antibody conjugated with HRP and visualised with the ECL substrate. Right: the values were quantified and plotted, n=2. Error bars are sample standard deviations (population).

The blots show that there is significantly less signal for the MB-E8 throughout the tested range of concentrations as compared to E8cc. A closer examination of 3D structures of these two proteins can be seen in Figure 6.8**A**. The figure shows that E8cc structurally separates the E8 binding loop from the poly-histidine tag

region, whereas MB-E8 positions both moieties at the same location in the three-dimensional plane. (Figure 6.8B) attempts to depict how such an juxtaposition of the poly-histidine tag and a binding loop may prevent the detection of the protein in ELISA studies. Assuming that the MB-E8 binds the nanoparticles, the binding event itself would cause the poly-histidine tag to be physically obscured. The probing antibody molecules would then be too large to navigate to the binding sites. The last factor determining the lower ELISA signal for MB-E8 is the fact that the poly-histidine tag is only 6 residue long compared to the 8 residue long tag on CC-E8 (Figure 6.2).

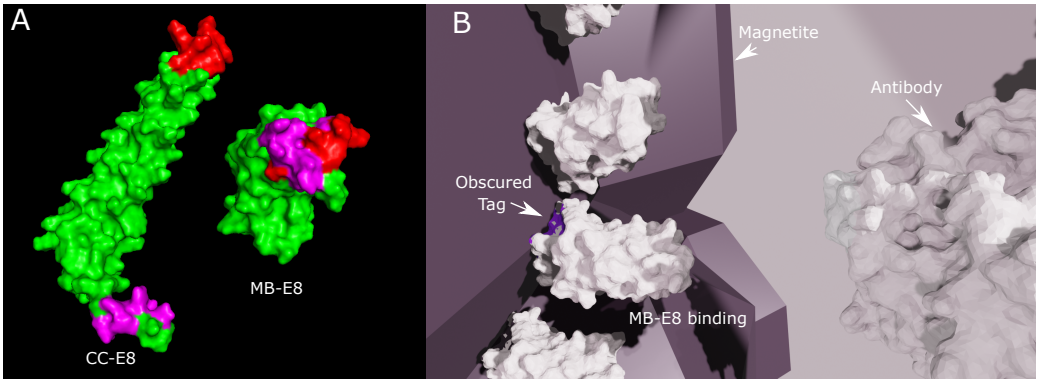


FIGURE 6.8: **A** A comparison of poly-histidine tag placement relative to the binding loop in E8cc and MB-E8. Poly-histidine tag is highlighted in magenta, the E8 binding loop is highlighted in red. Structures demonstrated as topology models. **B** A three-dimensional render to demonstrate how poly-histidine tag placement on MB-E8 may affect its accessibility to a probing antibody.

These two factors combined would result in a cumulative reduction of ELISA signal for MB-E8. First, from the antibody having less affinity for the poly-histidine tag and second, from the antibody having less access to poly-histidine tag in the context of the target protein being bound to a nanoparticle.

6.3.5 Alternative Binding Assays

Considering the poly-histidine tag accessibility issue concerning the MB-E8 discussed above, alternative methods to detect protein-nanoparticle binding were developed. First, 5 nmoles of each of the protein scaffolds was incubated with 500 μM of 3 mg/ml nanoparticle solution in 3 % skimmed milk in PBS-T for 16 hours after which the samples were washed and heated to 95 degrees for 10 minutes and analysed on SDS-PAGE (Figure 6.9).

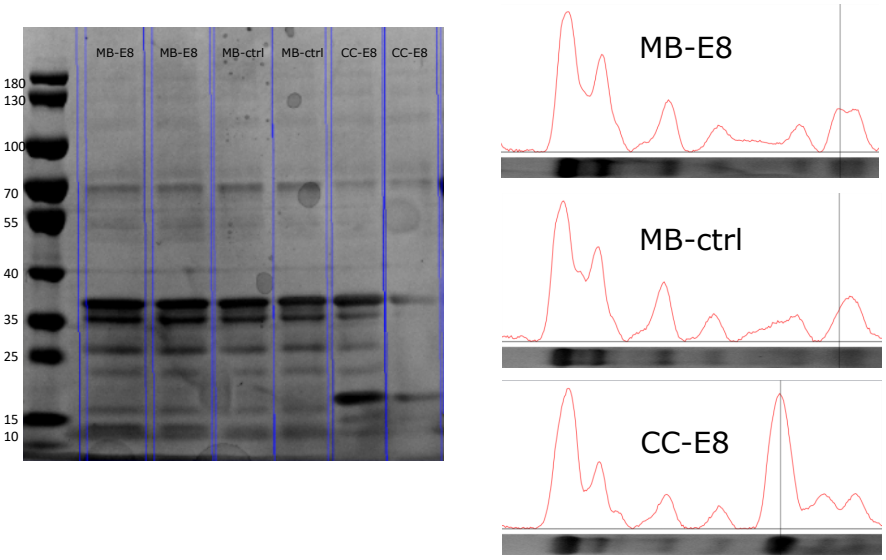


FIGURE 6.9: Left: SDS-PAGE 30-50 nm magnetite nanoparticle binding assay in 3 % skimmed milk in PBS-T. Each of the three scaffold proteins, MB-E8, MB-CTRL and E8cc was tested in duplicate. Right: gel lane profiles plotted. Perpendicular dotted lines show expected protein sizes for each sample.

A more detailed description of the assay can be found in methods (Chapter 2). The resulting gel image showed a laddering of protein bands corresponding to milk proteins. MB-E8 and E8cc lanes contained unique bands at their respective expected sizes that were not present in MB-CTRL suggesting that the E8 peptide enables binding to the nanoparticles. A range of scaffold protein concentrations was also tested in absence of 3 % milk (Figure 6.10).

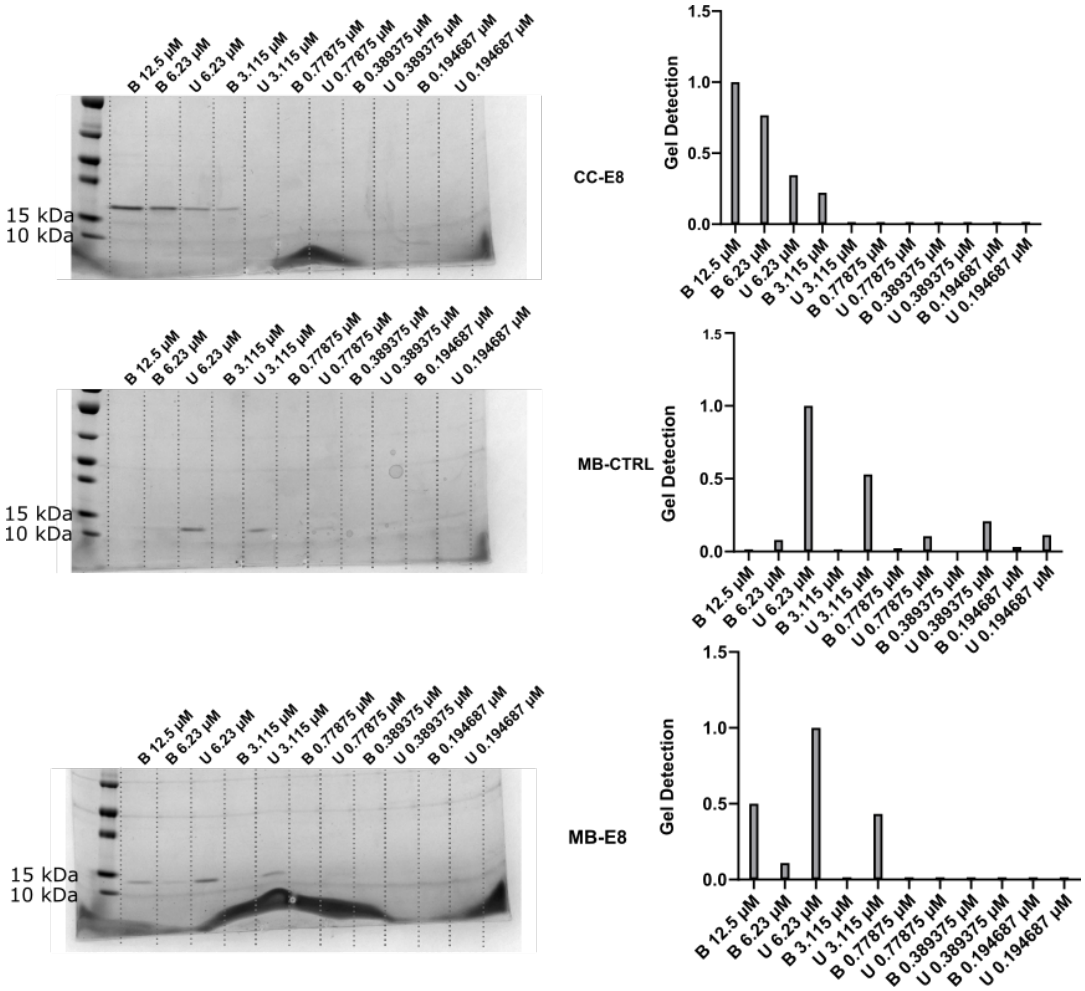


FIGURE 6.10: SDS-PAGE 30-50 nm magnetite binding assay for MB-E8, MB-CTRL and E8cc in absence of blocking reagents. Left: SDS-PAGE gels for each of the scaffold proteins. Different concentrations of scaffold proteins incubated with particles analysed on separate lanes. B - bound fraction, U - unbound fraction. Right: quantified intensities of bands corresponding to different concentrations of bound and unbound fractions for each of the scaffolds. Intensities have been normalised to the highest value.

The experiment compares three separate SDS-PAGE gels that analyse pH based elution of proteins after incubation with magnetite nanoparticles - 'bound' fractions. The E8cc shows intense bands in 'bound' fractions down to the concentration of 3.115 μM. There appears to be some 'unbound' protein, but it is less compared to amount of 'bound' fraction (which can be seen in the 6.23 μM comparison). MB-CTRL shows almost no signal in the bands corresponding to the 'bound' fractions with most protein being collected from the 'unbound' fraction. Lastly, MB-E8 shows signal in the 'bound' fraction when protein were incubated at 12.5 μM and relatively little binding at incubation concentration of 6.23 μM. While the assay most likely has a very high detection limit (limited by the lowest amount of protein detectable

by the gel staining method chosen), it shows that even without blocking conditions MB-CTRL does not bind to magnetite nanoparticles as well (if at all) as MB-E8 or E8cc. The assay shows that there is still a two fold stronger signal for E8cc binding compared to MB-E8. This may be an indication of true higher affinity or a function of the primary amino acid sequence being more prone to binding the SDS-PAGE gel staining reagents (different amino acid composition has been show to have different binding to Coomassie blue staining [298]).

Nevertheless, the alternative binding assays above show that simply transferring a magnetite binding peptide into a loop region of a known scaffold protein confers it the ability to bind nanoparticles without additional sequence optimisation.

6.4 Part 2 Introduction: to 7 amino acid Phage Display for CoPt L10

Chapter 1 discusses the need for production of small, magnetically hard nanoparticles and identifies thermally annealed CoPt $L1_o$ as good candidates to be used as single particle bits when constructing bit patterned media. While production methods already exist, they require intense energy input and thus are not scale-able. It has been postulated before that it may be possible to template the production of CoPt $L1_o$ particles using specific protein/peptide molecules. This bio-templating would thus make it unnecessary to have a thermal annealing step in the production of these particles. A general experimental scheme (discussed in Jarrald et.al., 2020 [193]) has been defined where one can produce the crystal of the desired material using an energy intensive method, then using a bio-molecule display technique to identify molecules that bind specifically to the produced crystal and then using the identified biological material as a templating agent for synthesising the original crystal material without the high energy input. A phage display approach has already been demonstrated to discover 12 amino acid long peptides with biotemplating potential [181] for CoPt $L1_o$. More recently, however, a rational design study revealed that shorter peptides 7-8 amino acid peptides might have better overall performance as CoPt $L1_o$ synthesis additives and pre-formed CoPt $L1_o$ particle binders. Rational design lacks the ability, however, to explore the full possible amino acid sequence space. Commercially available 7 amino acid peptide phage display systems offer the full coverage of all possible 7 amino acid combinations and a phage display screening for CoPt $L1_o$ binding using 7 amino acid libraries has not been demonstrated so far in literature. Thus, part 2 of Chapter 4 aims to perform phage display with such a library using $L1_o$ CoPt magnetic nanoparticles as the target.

6.5 Part 2 Results: 7 Amino Acid Phage Display for CoPt L10

6.5.1 CoPt Nanoparticles: Production of the Starting Material for Phage Display

The phage display project was started with an attempt to produce CoPt $L1_o$ using the standard high energy synthesis method. CoPt A1 particles were first synthesised by using a method of mixing water dissolved $CoSO_4$ salts with Na_2PtCl_4 under inert conditions with a subsequent reduction with $NaBH_4$. The resulting particles were magnetic and showed characteristic CoPt A1 X-ray diffraction peaks (Figure 6.11).

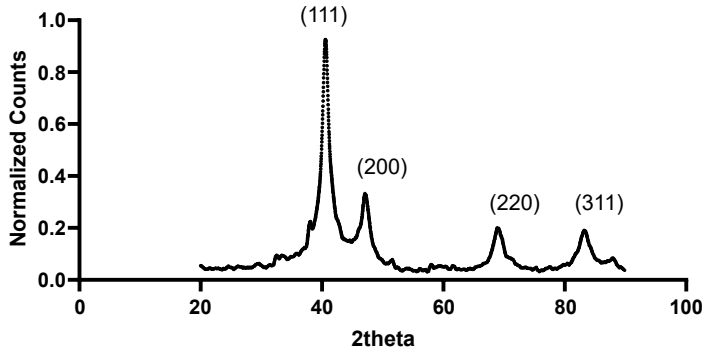


FIGURE 6.11: CoPt A1 X-ray diffraction spectrum. In brackets - labelled characteristic CoPt A1 peaks.

The spectrum shows the expected peaks at 40.1, 46.9, 68.4 and 82.1 corresponding to CoPt A1 crystal phases (111), (200), (220) and (311), respectively, signifying a successful synthesis. There also appear to be low intensity peaks present at 33 and 55 degrees that could not be assigned to A1 CoPt and most likely signify a small degree of impurity in the obtained product. A peak at 33 also correlates with (110) phase of CoPt $L1_o$, small amount of which may have been produced during the synthesis reaction. The same sample was subjected to a heat treatment in attempt to anneal the crystal planes of the particles to obtain CoPt $L1_o$. The heat treatment was carried out by placing the particle sample in a clean quartz cuvette capable of withstanding heat treatments beyond 1000 °C. The particles were inserted into the furnace at room temperature and heated slowly (1 °C/min) to 800 °C under an argon gas atmosphere. The samples were then cooled passively at room temperature. Annealed CoPt $L1_o$ can be distinguished from CoPt A1 either by measuring magnetic coercivity or by interpreting X-ray diffraction peaks. Surprisingly, the particles collected

from the furnace were no longer magnetic and showed an amorphous XRD signal (Figure 6.13).

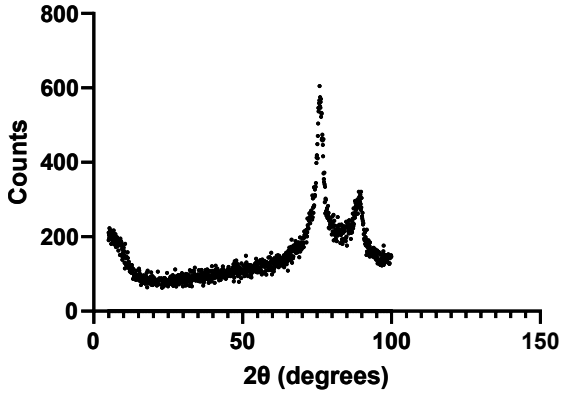


FIGURE 6.12: X-ray diffraction spectrum of attempted sample annealing.

The spectrum appears to be lacking any recognisable peaks corresponding to either CoPt $L1_o$ or CoPt A1. The same procedure was repeated two more times to same results. Dr Rosie Jarrald had previously managed to produce CoPt $L1_o$ using a thermal annealing process. The XRD spectra obtained can be seen in Figure 6.13.

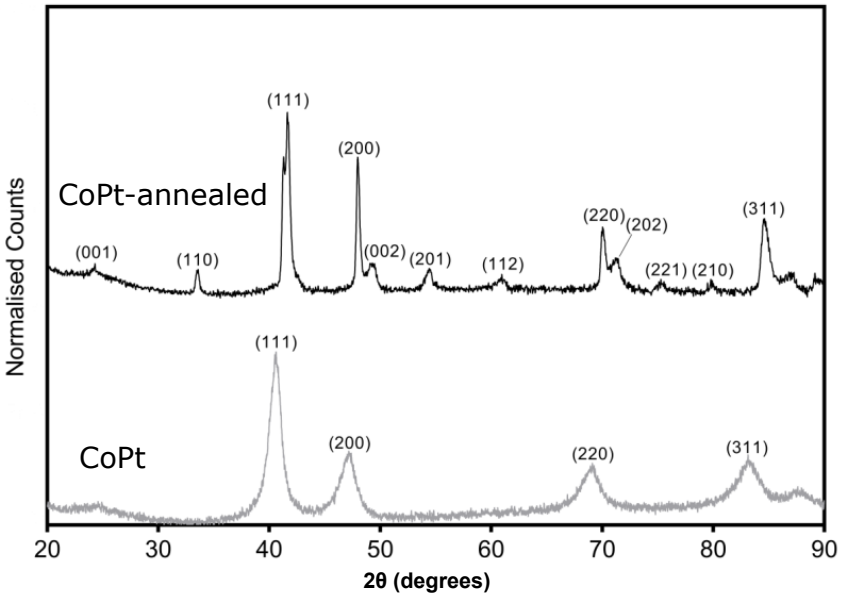


FIGURE 6.13: X-ray diffraction spectra of CoPt as-synthesised and CoPt after a thermal annealing process.

The resulting CoPt $L1_o$ spectra* show additional peaks appearing after the thermal annealing procedure. The peaks corresponding to (001), (110), (002), (202), (221) and (210) crystal phases expected from CoPt $L1_o$ XRD spectrum were all visible. It remains puzzling why multiple attempts to replicate this procedure had failed. It

Spectra adapted from Dr Rosie Jarrald's PhD thesis

may be that the 25 °C lower temperature used here was too low to allow for the necessary crystal lattice re-arrangements (Dr Rosie Jarrald's protocol followed heating to 825). The material that produced the CoPt $L1_o$ resembling XRD spectrum (Figure 6.13) was used for the phage display in later sections.

6.5.2 Phage Display on Annealed CoPt Nanoparticles

A naive commercially available Ph.D 7 M13 phage display library (NEB, UK) was used to identify seven amino acid sequences with high binding affinity for annealed $L1_o$ CoPt. The phage display procedure is described in detail in (Chapter2) and is summarised in Figure 6.14. At no point during the process was there any appearance of uncoloured plaques indicating that no wild phage contamination occurred. All twelve sequencing analyses returned correct and unmutated sequences matching pIII M13 protein (found on the phage surface) with variable 7 amino acid regions.

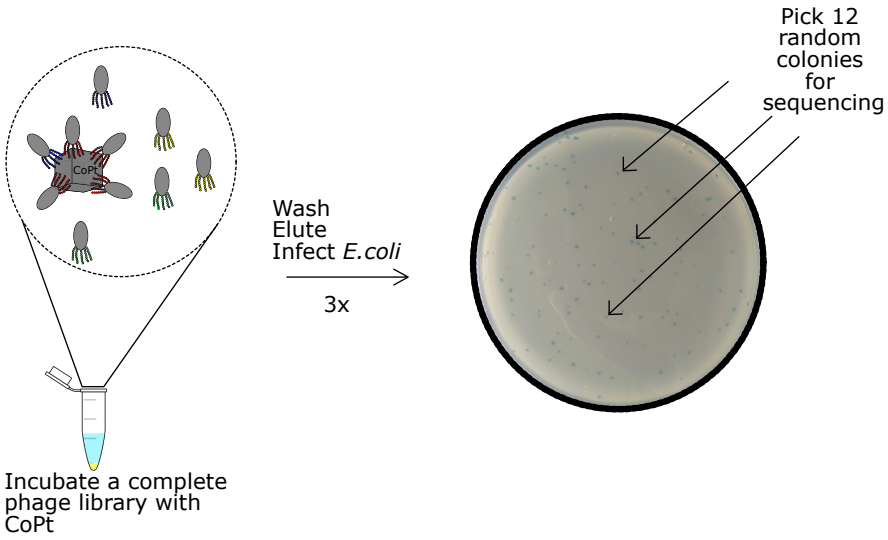


FIGURE 6.14: A graphical description of the phage display process for $L1_o$ CoPt selection.

The obtained sequences are summarised in Table 6.2. The results showed a total of three unique sequences: A1 - YHPLRNH, A5 - KSPIHVP and B6 - GHSQQST (peptide IDs were assigned based on the plasmid purification wells showing the first instance of the sequence). A1 and B6 sequences occurred twice each, while A5 occurred 8 times out the total of twelve counted. The higher occurrence of one peptide sequence over the others may imply higher initial counts of the phage carrying that

sequence in the final eluted pool indicating that this sequence had been enriched due to having highest affinity for the target material.

Peptide ID	Peptide Sequence	Number of Hits
A1	YHPLRNH	2
A5	KSPIHVP	8
B6	GHSQQST	2

TABLE 6.2: CoPt $L1_o$ 7 amino acid Phage Display results summarising numbers of discovered sequences.

6.5.3 Discovered CoPt $L1_o$ 7 Amino Acid Peptide Sequence Analysis

In order to interpret the phage display results, a series of bioinformatics analyses were performed. A breakdown of obtained sequences can be seen in Figure 6.15.

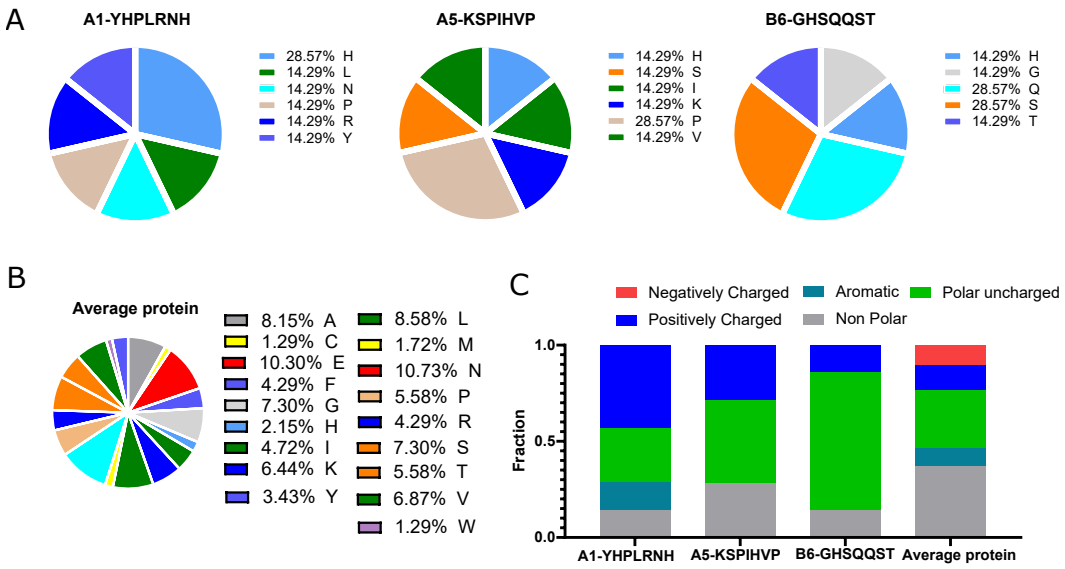


FIGURE 6.15: **A** Pie charts representing the amino acid composition of each phage display discovered peptide (A1, A5, B6). **B** A composition diagram of an average protein [300]. **C** Bar charts showing fractions of negatively charged, aromatic, polar uncharged, positively charged and non-polar within the phage display discovered peptides A1, A5, B6.

The amino acid composition analysis reveals that A1 is relatively high in positively charged residues (42 %), followed by A5 (28 %) with B6 having the fewest amount (14 %). All sequences, have at least a single histidine residue (with A1 having two). A1 and A5 both possess proline residues and A1 is the only one of the three to have a single aromatic residue (tyrosine). Neither of the enriched residues (occurring more than once per peptide) within the three discovered peptides can be seen to be more commonly occurring in the 'average protein' data set [300]. In fact, histidine, proline and glutamine are seen to occur less often in a composition of an 'average protein'. The amino acid composition shows promise that the peptides would be good binders of either platinum ions or pre-formed particles as histidine has been shown to occur frequently in CoPt $L1_o$ binding peptides [193, 181], and is also known as the amino acid residue to bind Pt atoms in cisplatin drugs [299]. The peptides were analysed for their theoretical biochemical properties (Figure 6.16).

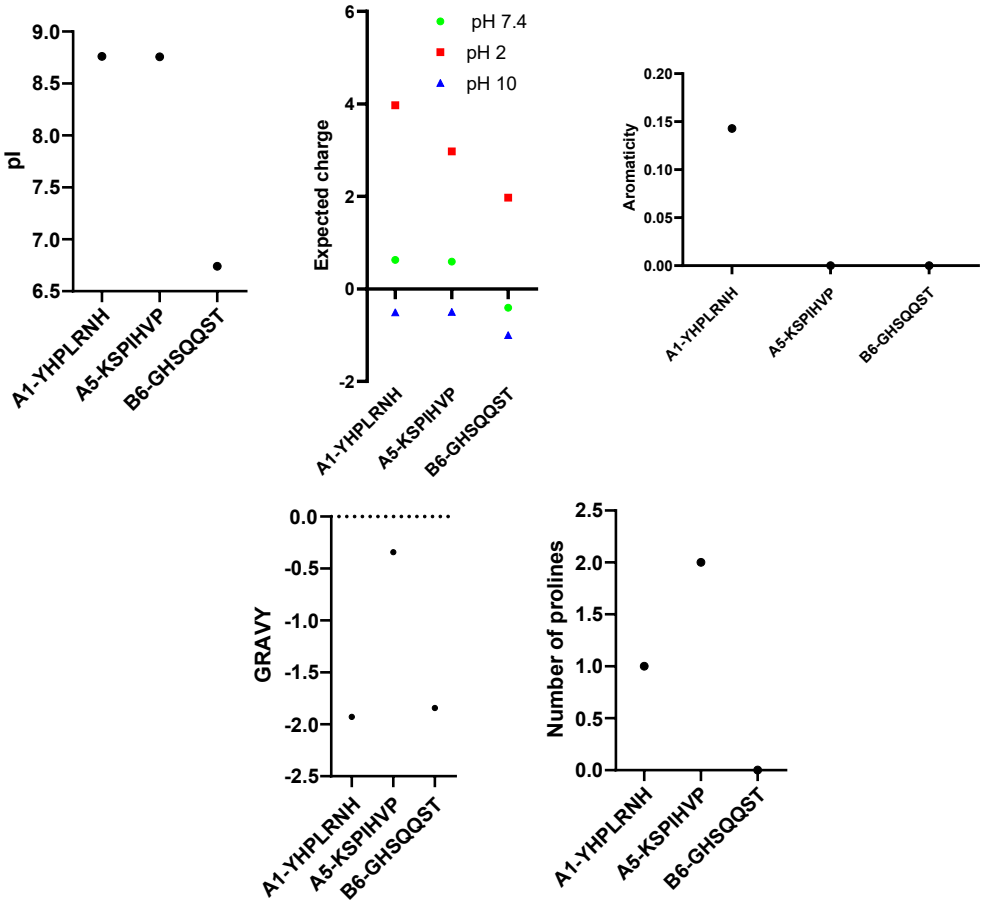


FIGURE 6.16: Analysis of A1, A5, B6 showing differences in pI, expected charges at different pH, aromaticity, grand average of hydropathy - GRAVY and number of proline residues.

The calculated biochemical metrics show that A1 and A5 carry the same pI of 8.76, with B6 being lower at 6.75. Under physiological pH 7.4 conditions at which the binding experiments have been carried out, A1 and A5 would have had a net positive charge, while B6 may have had a slight net negative charge. A1 and B6 were calculated to be hydrophilic (low GRAVY score), while A5 had a GRAVY score of near zero originating from a presence of a valine and isoleucine residues. A5 also had the largest number of proline residues (suggesting a structural component to the nature of particle binding).

For a more robust interpretation of the phage display results the peptides were also compared to known CoPt $L1_o$ binding/biotemplating sequences from literature. The sequences investigated are listed in Table 6.3.

Peptide ID	Peptide Sequence	Reference
CoPt-Reiss2005-1	KTHEIHSPLLHK	[181]
CoPt-Jarrald2020-1	KSLSRHMHIHHH	[193]
CoPt-Jarrald2020-2	KSLSRMDHIHHH	[193]
CoPt-Jarrald2020-3	KSLSRHD	[193]
CoPt-Jarrald2020-4	KSLSRMDK	[193]
CoPt-Jarrald2020-5	KSLSMK	[193]
CoPt-Jarrald2020-6	KSLSRGMK	[193]

TABLE 6.3: CoPt $L1_o$ binding/biotemplating sequences discovered in literature.

The peptides in the set above were demonstrated to facilitate controlled CoPt $L1_o$ in their respective studies. The sequences compared for their amino acid composition (Figure 6.17)

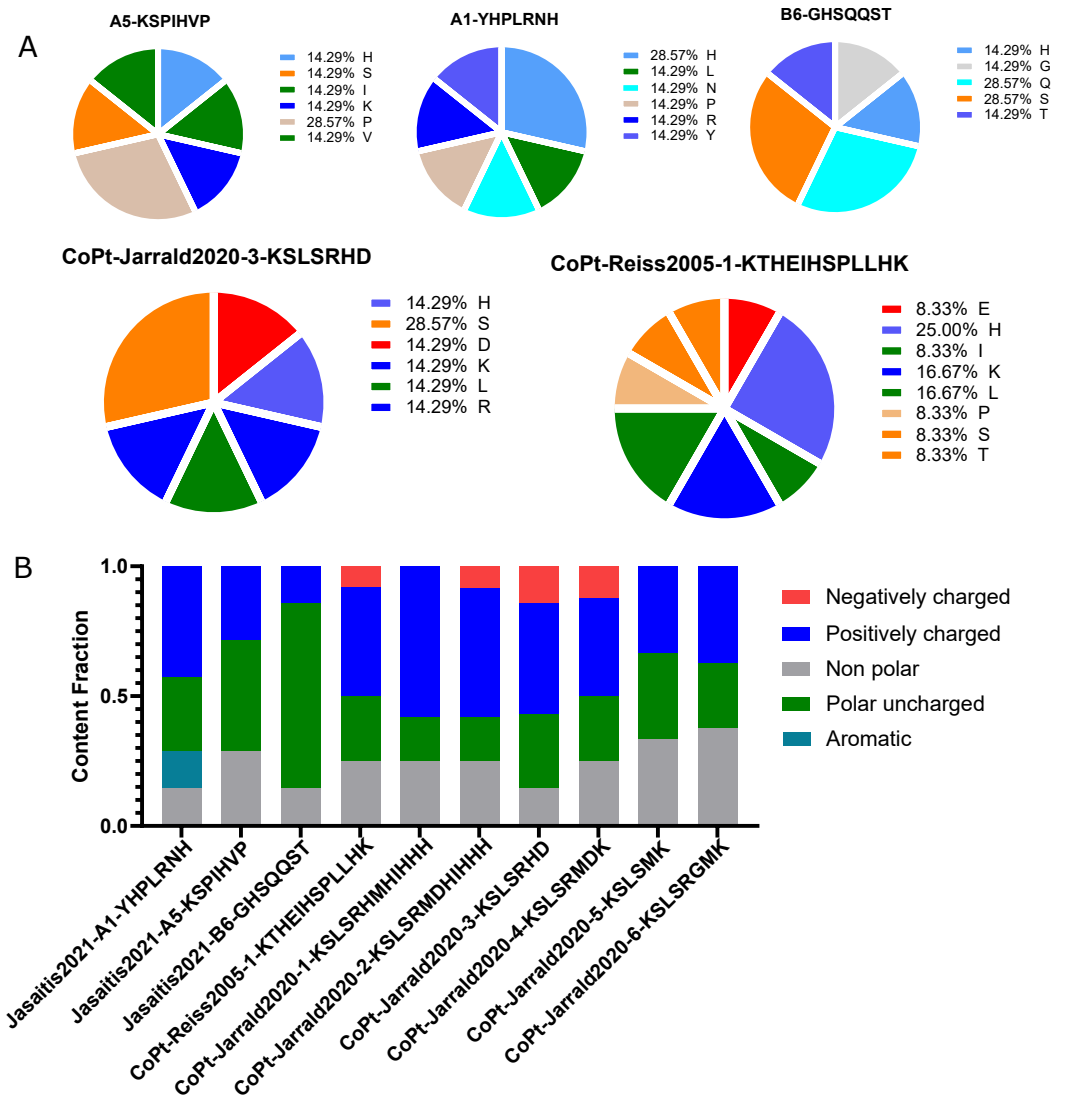


FIGURE 6.17: **A** Pie charts representing the amino acid composition of each phage display discovered peptide (A1, A5, B6) compared to other peptides targeted at CoPt $L1_o$. A 7 amino peptide obtained through rational design (Jarrald2020-3) [193] and a phage display derived 12 amino acid peptide (Reiss2005)[181]. **B** Bar charts showing fractions of negatively charged, aromatic, polar uncharged, positively charged and non-polar residues within peptides described in **B** and additional peptides found in Jarrald et.al., 2020[193].

The amino acid composition analysis shows that the reference sequences possess high amounts of histidine and have at least one positively charged lysine residue. These profiles are similar to those of A5 and A1, whereas B6 appears to be an outlier. Interestingly, both CoPt-Reiss2005-1-KTHEIHSPLLHK and CoPt-Jarrald2020-3-KSLSRHD possess a single negatively charged residue, whereas neither of the new set of peptides did. From the amino acid type breakdown in Figure 6.17 A5 and B6 peptides seem to be on the lower end in terms of positively charged amino acid content, whereas A1 is close to average. B6 seems to have an uncharacteristically (for

CoPt $L1_o$ binding peptide) high percentage of polar uncharged residues. A series of calculations for biochemical properties were also carried out to compare the new set of peptides to the previously discovered ones (Figure 6.18).

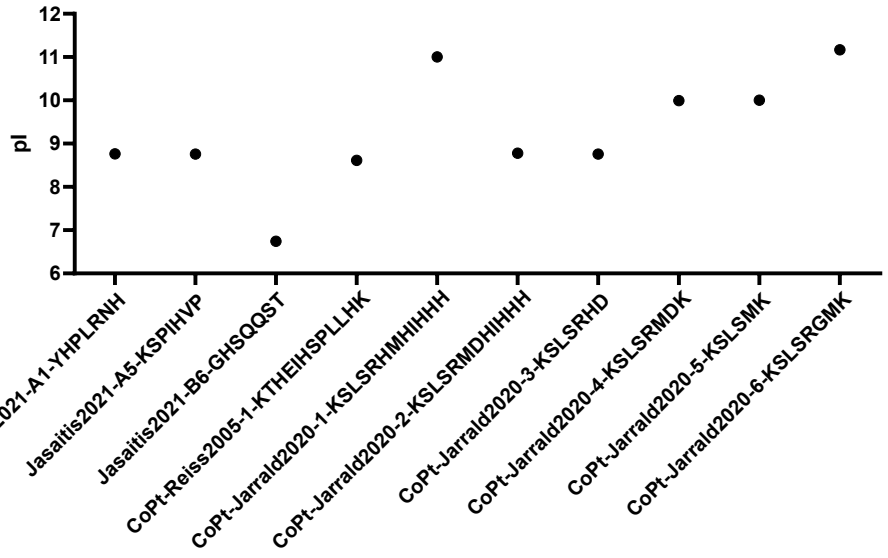


FIGURE 6.18: Isoelectric point (pI) comparison between known CoPt $L1_o$ binding sequences and peptides A1, A5, B6.

In the general CoPt $L1_o$ selected population the isoelectric points tend to fall between 8.7-11. Peptide B6 seems to be an outlier due to being composed mostly of glycine, glutamine, serine and threonine residues. Only some members of the rationally designed data set have pI values below 10. The information regarding proline residue content was also plotted (Figure 6.19).

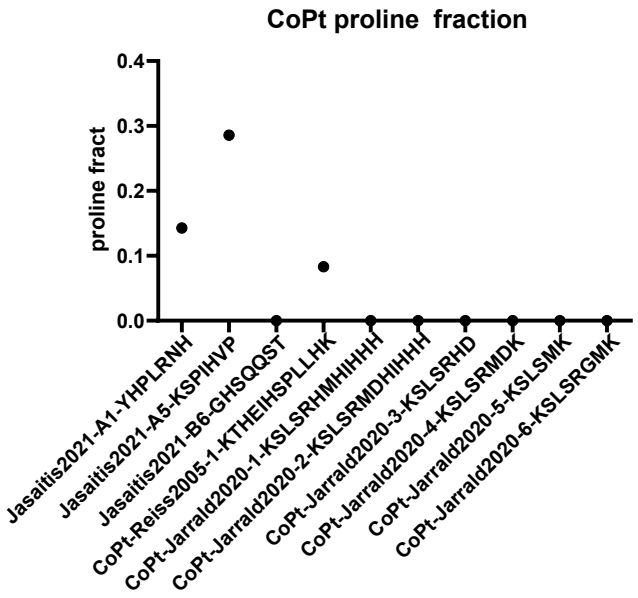


FIGURE 6.19: Proline content comparison between known CoPt $L1_o$ binding sequences and peptides A1, A5, B6. Proline fract - fraction of proline within each sequence.

Other than A1 and A5 discovered here, only the other phage display selected peptide CoPt-Reiss2005-1-KTHEIHSPLLHK possesses a proline residue. There seems a negative bias against proline residues in the rational design study. Peptides A1 and A5 seem to have an excess of fractional proline content compared to any of the previously discovered peptides with A5 having near 30 % of the peptide content being proline residues. Lastly, the peptides were compared for their charged amino acid content (Figure 6.20).

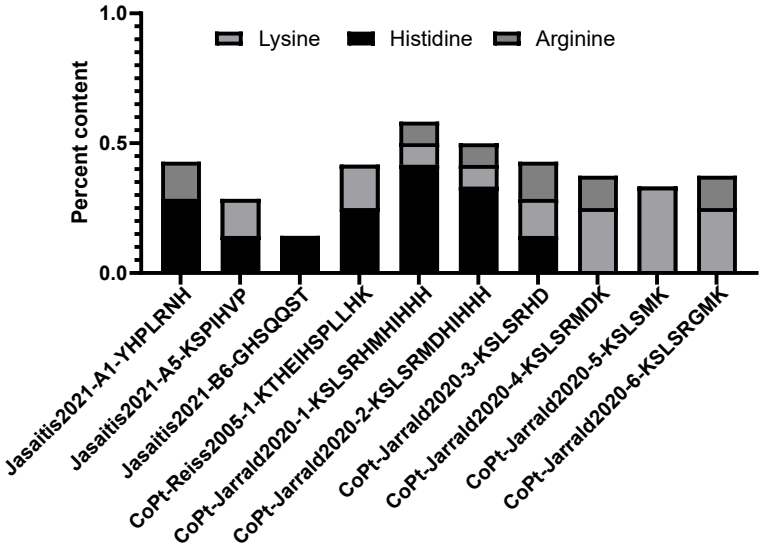


FIGURE 6.20: Lysine, histidine and arginine content comparison between known CoPt $L1_o$ binding sequences and peptides A1, A5, B6.

The fractional charged amino acid content values reveal that peptide A1 matches the overall average (among the compared peptides) charged content of 40 %. A1 has the same amount of charged residues as the previously phage display discovered 12 amino acid peptide CoPt-Reiss2005-1-KTHEIHSPLLHK and the same length (7 amino acid) peptide from the rationally designed data set CoPt-Jarrald2020-3-KSLSRHD (the peptides are also identical in their pI values Figure 6.18). The exact make-up of charge residues however differs with A1 being composed of a mix of histidine and arginine (2:1 ratio), CoPt-Reiss2005-1-KTHEIHSPLLH - composed of histidine and lysine (2:1 ratio) and CoPt-Jarrald2020-3-KSLSRHD - composed of histidine, lysine and arginine (1:1:1 ratio). Peptide B6 has the lowest charged amino acid fraction of all the compared peptides (14 % accounted for by a single histidine residue). Most of the sequences include at least one histidine (with the exception of three outliers).

Overall, there seem to be significant sequence composition similarities between the discovered A1 and A5 peptides and the known sequences. In particular, the phage display discovered CoPt-Reiss2005-1-KTHEIHSPLLHK sequences shared similar amounts of positively charged residues, equal isoelectric points and presence of proline residues. CoPt-Jarrald2020-3-KSLSRHD is also similar for the mentioned properties with the

exception of proline content. Peptide B6 appears to lack similarities to the overall CoPt $L1_o$ data set, implying that the binding mechanism causing the enrichment of this sequence in the phage pool, must be different.

6.5.4 Discovered CoPt $L1_o$ 7 Amino Acid Particle Binding

To compare the relative CoPt $L1_o$ binding intensities of each of the discovered peptides, the isolated phage particles displaying A1, A5 and B6 sequences were subjected to a phage adapted ELISA assay demonstrated in Figure 6.21**A**.

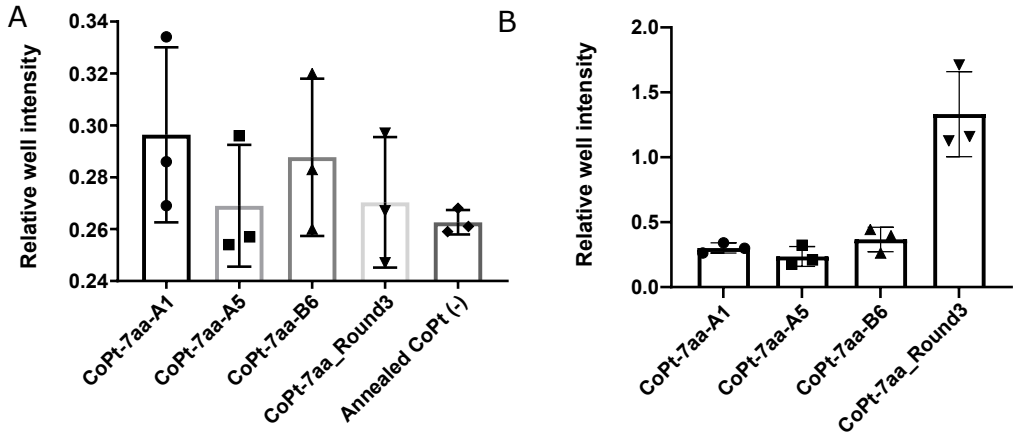


FIGURE 6.21: Phage ELISA experiments. **A** Peptide A5, A1 and B6 displaying isolated phage particles tested against CoPt $L1_o$. **B** Peptide A5, A1 and B6 displaying isolated phage particles tested against freshly synthesized CoPt A1. Error bars - standard deviation (population), n=3.

The phage pool of third screening round (Round3) was used as a control (this pool would contain the three identified peptide sequence displaying phages and potentially other undetected sequences - only 12/70 of obtained plaques have been sequenced randomly, thus there may have been other sequences in the pool). The assay failed to produce robust results as the variability in the ELISA signal was too high causing large and overlapping standard deviations (population) among all tested phage samples. The standard deviation is low for the annealed CoPt (CoPt $L1_o$) control. This means that the variable signal in the sample tubes is originating from the differential binding of the probing antibody to the phage particles and that it is not the nanoparticles themselves triggering catalysis of the colorimetric reagent. This indicates that binding events are taking place. The low resolution results are sensible in light of the observed difficulty of CoPt $L1_o$ sample homogenisation prior to the experiment. This issue prevented complete mixing and equal distribution between

the sample tubes causing some to contain less target material than others. Since the ELISA signal is dependent of the number of binding sites available to the probing antibody (phage particles) which in turn are dependent on the number of binding sites available to the phage particles (CoPt $L1_o$), an unequal distribution of particles would cause high variability in the signal. An attempt to overcome this, was with an analogous ELISA experiment with CoPt A1 as the target (Figure 6.21**B**). The results showed that Round3 phage pool has 4-5 times more binding to the nanoparticles (unfortunately the baseline reading for particles without a phage sample was not obtained). What this implies is that if A1, A5 and B6 are capable of binding CoPt $L1_o$), this binding is specialised and it precludes these peptides from binding CoPt A1 as efficiently as the phage pool from Round3. Round3 composition should be made up of 82 % of other, unknown peptide sequences that may have higher affinity for CoPt A1.

6.5.5 Potency of the Discovered CoPt $L1_o$ as CoPt Synthesis Additives

The ability of the three phage display discovered peptides to bio-template CoPt crystal assembly was also investigated. The three peptides A1, A5 and B6 were obtained through commercial *in vitro* synthesis (see methods) as linear 7 amino acid peptides. To understand the peptide effectiveness as nanoparticle bio-templates, standard CoPt Synthesis reactions were set up as indicated in the methods section (Chapter 2). Each of the peptides was used to supplement separate reactions at two different final peptide concentrations (0.1 mg/ml and 0.01 mg/ml). The processed (washed and dried) reaction products were in dark magnetic powder form. The products of reactions with 0.01 mg/ml of peptide additives were analysed using TEM (Figure 6.22).

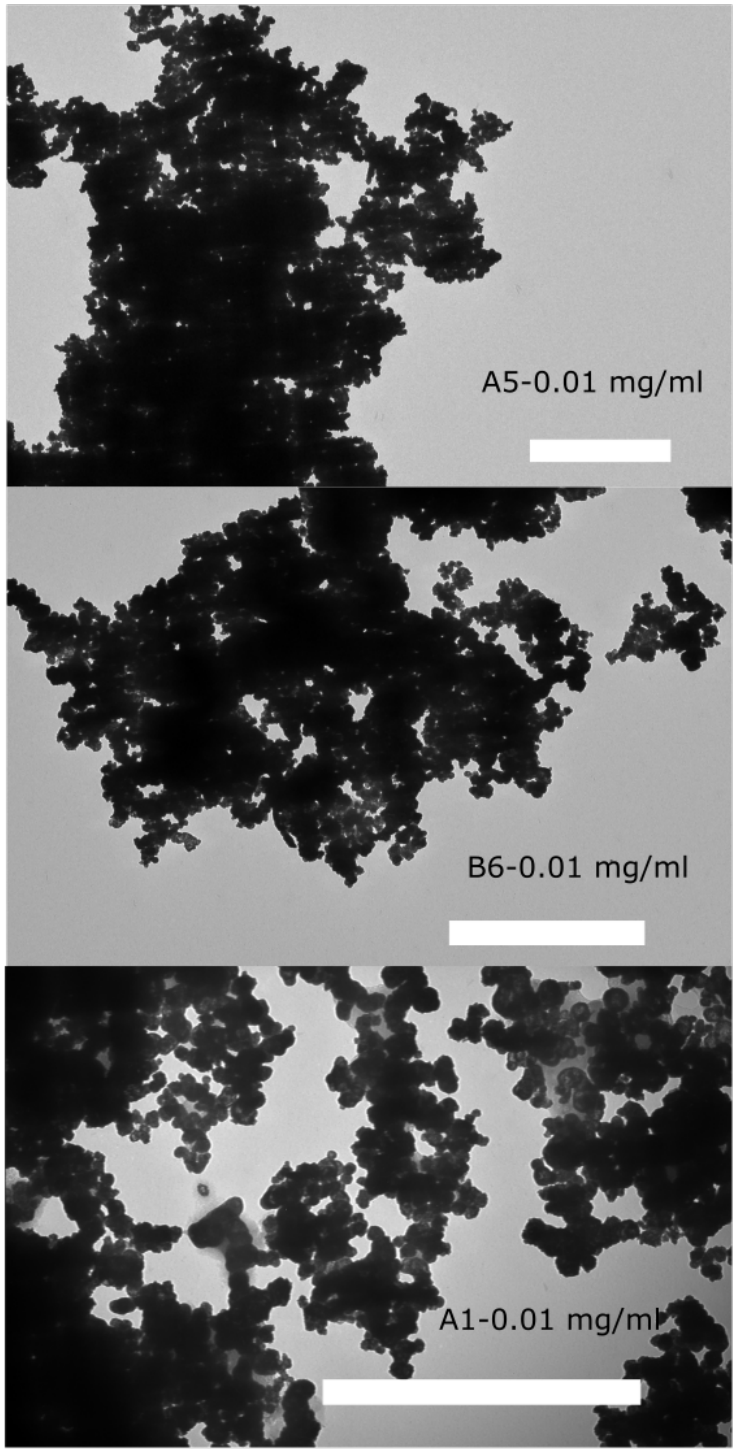


FIGURE 6.22: TEM images of CoPt synthesis results with peptides A1, A5 and B6 as additives (0.01 mg/ml). All scale bars - 1 μm.

Each sample showed clusters of high contrast material. Multiple scans were obtained, however, no regions of highly dispersed particles could be found for discrete particle size analysis. Individual particles were only visible on edges of clusters, thus further optimisation of sample preparation to overcome particle-particle aggregation is needed (quantification of particles on the outer edges of visualised particle cluster may produce biased results). As an alternative particle size analysis method, X-ray diffraction was employed. (Figure 6.23).

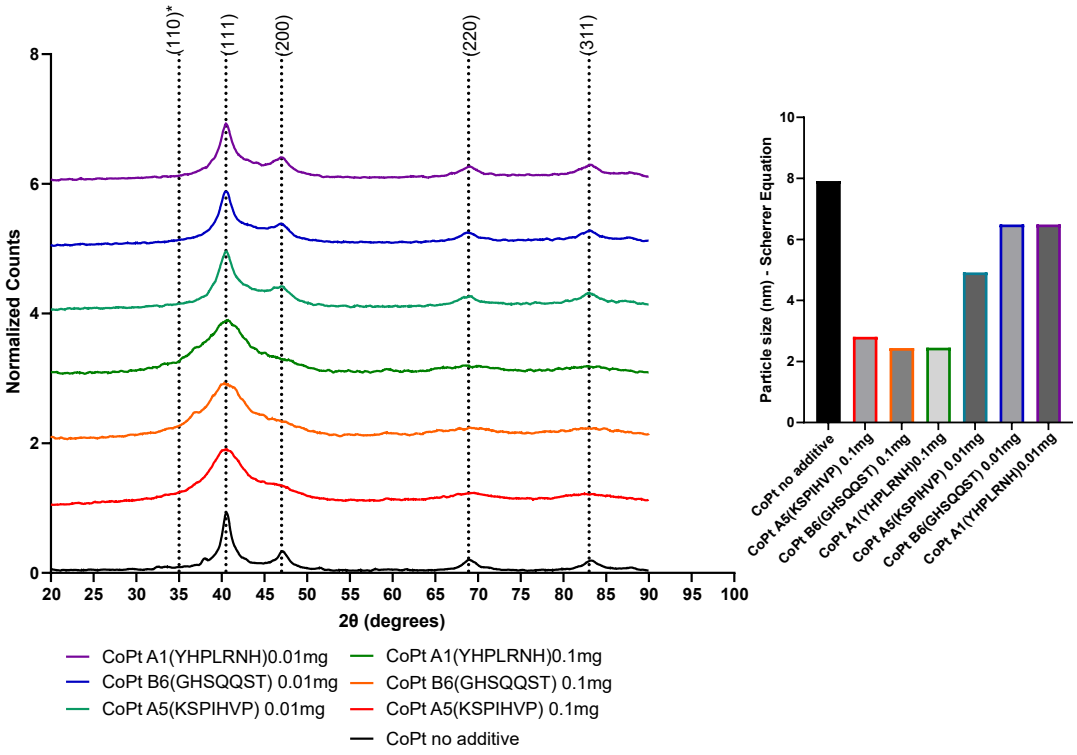


FIGURE 6.23: Right: XRD spectra of CoPt particles synthesised with A1, A5, B6 peptides as additives at 0.01 mg/ml and 0.1 mg/ml concentrations. Crystal phases corresponding to peaks 40.1, 46.9, 68.4 and 82.1 are denoted as (111), (200), (220) and (311). Left: plotted particle size values obtained from applying the Scherrer (see methods) analysis on the corresponding XRD spectra.

All spectra showed characteristic CoPt peaks at 40.1, 46.9, 68.4 and 82.1 corresponding to CoPt A1 crystal phases (111), (200), (220) and (311). Neither of the synthesis reactions shows presence of the (110) peak (around 35 degrees) that would signify the presence of CoPt $L1_o$. The peaks obtained from synthesis reactions with peptide additives show a broadening of the expected peaks which is known to signify smaller crystalline domains. In order to estimate the crystalline domain sizes

of the samples the Scherrer equation was applied to the XRD spectral information (Table 6.4).

Peptide ID	Peptide Concentration (mg/ml)	Calculated size
No additive	n/a	7.91
A1(YHPLRNH)	0.01	6.49
A5(KSPIHVP)	0.01	4.92
B6(GHSQQST)	0.01	6.49
A1(YHPLRNH)	0.1	2.45
A5(KSPIHVP)	0.1	2.81
B6(GHSQQST)	0.1	2.44

TABLE 6.4: Summary of particle sizes obtained from Scherrer analysis of XRD spectra shown in Figure 6.23.

The results determined the presence of crystalline domains between 2.44 nm and 7.91 nm (depending on the sample), at this size range CoPt crystalline domain size would correspond to the individual particle size (in other words - particles are expected to be single domain). The calculations first confirm that without additives the synthesis reaction produced nanoparticle sizes that have been reported before - 7.91 nm [193]. The synthesis reactions with peptide concentration of 0.01 mg/ml produced particles that are smaller by 1.42 nm when A1 and B6 peptides were added and by 2.99 nm with the A5 peptide. The synthesis reactions with higher additive concentrations showed a further reduction in obtained particles sizes. A1 and B6 were smaller than the control particles by 5.47 nm, whereas A5 was smaller by 5.1 nm. The same samples were also analysed for their magnetic susceptibility (Figure 6.24).

The data shows that all synthesis reactions with 0.01 mg/ml of either of peptide additives produced material with magnetic susceptibility that is two fold higher than the control reaction. The reaction peptide A5 appears to have produced particles with slightly higher susceptibility than A1 or B6. Reactions with 0.1 mg/ml of either peptide A1 or B6 produced material with magnetic susceptibility values of

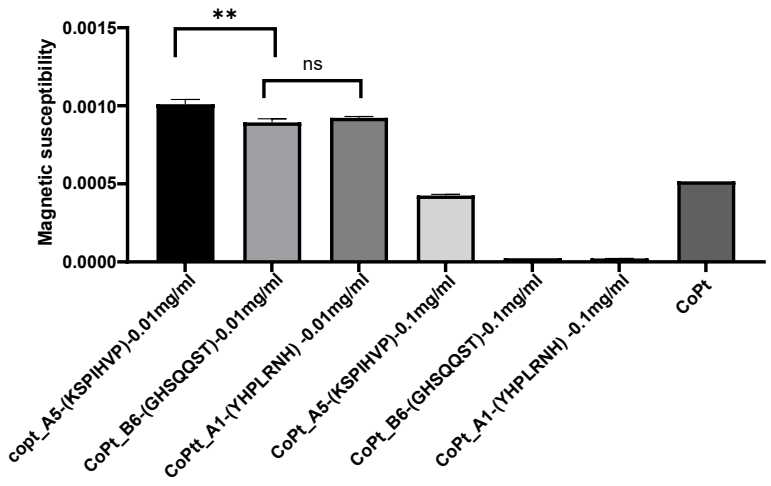


FIGURE 6.24: Magnetic susceptibility analysis of synthesis results with peptide additives. Measured in cm^3/g. Mean values plotted, n=3. Error bars are standard deviations (population).

near-zero, however when used as an additive at 0.1 mg/ml peptide A5 produced particles with magnetic susceptibility values just slightly lower than the control.

This section reveals that peptide additives in CoPt synthesis reactions have an effect on the properties of the product. At first glance it appears that using additives at 0.01 mg/ml produces particles that are smaller yet more magnetically susceptible, whereas additives at 0.1 mg/ml appear to make small particles with diminished or lost magnetic susceptibility. It is possible that such effects would be caused purely by the presence of peptide chain backbone irrespective of the specific amino acid sequence. However, peptide A5 seems to stand out as a unique case among the three discovered peptides as it produces differently sized particles in either of the concentrations, furthermore the small 2.81 nm particles produced in presence of A5 retain their magnetic susceptibility and the value is comparable to the control. This implies that there must be some biotemplation taking place ensuring that small crystals maintain some order. The effect may be arising due to the presence of two proline residues resulting in a curved peptide.

6.6 Conclusion to Chapter 6

This chapter shows broad specific scaffold independent binding capacity for a magnetite binding peptide E8 when it is displayed as a component of a small protein.

These results are the first instance of showing FN3-monobody scaffold being used to specifically bind to magnetic nanoparticles via the variable loop regions (it has been terminally conjugated before).

While it is not fully clear if the CoPt material produced with the new phage display discovered peptides is indeed CoPt $L1_o$. There is strong evidence that the new set of peptides may be nano-technologically useful for CoPt immobilisation and aided synthesis.

Together these results add to the tool-kit for constructing protein-based nanoparticle arrays by introducing new ways to attach nanoparticles to proteins and by introducing promising leads to producing CoPt $L1_o$ the ultimate material for bit-patterned-media.

Chapter 7

Discussion and Concluding remarks

7.1 Thesis Summary

The aim of this thesis was to progress towards establishing biological routes to creating magnetic nanoparticle patterns with features at near 10 nm scale. Within the individual chapters there already exist small localised discussions concerned with the technical aspects of the work. This final chapter revisits the results obtained during the course of this PhD, establishes the relationship with the broader literature for future directions and finally attempts to draw connections between the chapters with respect to their relevance to BPM. Even though this project was focused on BPM, the extent of use cases for the findings goes beyond data storage applications.

7.2 Chapters 3-4 Outcomes

The chapter concerned with SasG domain G52-E-G53 constructs aims to characterise this domain as a nano-pattern building block. The proven mechanical properties [107] combined with its unusual extended β-sheet rod-like structure make this protein an ideal candidate for self-assembly systems where precise spacing between features is desirable (as in BPM). Here, the production of the intact G52-E-G53 protein and pH dependent structural integrity is assessed. The protein seems to maintain its structural features at pH values as low as 3. The assessment of orthogonality between the two coiled-coil pairs used to join G52-E-G53 domains into chains gives a

strong indication that sub-units S1 and S2 should not form assemblies in absence of their respective binding partner. Combined with the observed coiled-coil pH sensitivity demonstrated by the QCM-D experiments (and the demonstrated G52-E-G53 pH tolerance), this presents a controllable self-assembly system. The assembly process is triggered by the mixture of the two sub-unit solutions and can be deactivated at will by adjusting the pH without disrupting the integrity of the G52-E-G53 backbone. Such control has the advantage over a homo-polymer assembly, where the singular sub-unit would have a tendency to aggregate during the production process. As designed, S1 and S2 assemble via the coiled-coils. This was inferred from the fact that as the sub-units are mixed, α-helical CD signature persists at higher temperatures, meaning that the addition of the partner sub-unit stabilises the α-helices. Since the only additional molecular entity upon sub-unit mixture is their complementary α-helix it can be concluded that the stability originates from coiled-coil formation. This, however does not prove that both terminal α-helices form coiled-coils. For this, QCM-D provides good evidence. Assuming that the gold immobilisation strategy orientated the sub-unit surface deposition as intended, application of either cysS1A3 or cysS2A3 to the surface would have only produced BN4 and P5 binding sites, respectively. Since the subsequent application of the respective binding partners showed specific immobilisation it can be concluded that both terminal α-helices are able to form coiled-coils on either of the sub-units. Liquid AFM showed that the two sub-unit system was able to assemble with the same specificity. The analysed lengths of the assemblies after a 32 hour incubation in liquid were on average 1000 nm.

In absence of their binding partner, sub-units S1 and S2 both showed α-helical stability at temperatures beyond the literature values [248, 249]. It cannot be ruled out that the α-helices, originally designed to be disordered in isolation, would be stabilised in the context of a fusion protein by non-covalent interactions with the neighboring domain [301] (which would correspond to G52-E-G53 in this context). An estimated folding behaviour for such a scenario can be seen in Figure 7.1.

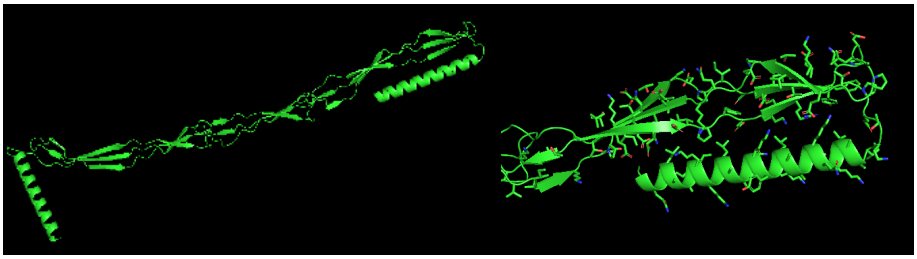


FIGURE 7.1: Artificially generated demonstrative model on how a terminal α-helix could align itself to the G52-E-G53. S1x structure depicted.

The CD experiments also showed that regardless of whether or not the sub-units carry a C-terminal magnetite binding peptide A3, their interaction behaviour does not change. Both sub-units carrying the peptide modification were tested for magnetite particle binding and the results showed that only S1A3 was granted the ability to bind the particles. This indicates that including a short linear peptide at a terminal end of potentially flexible α-helix may not be a reliable strategy. The short peptide length makes it prone to local charge based interactions, perhaps obstructing the ability of the peptide to bind particles. Additionally, the source literature on the terminal α-helices used [248, 249] show that the C-terminus of S1A3 should be more structurally stable than S2A3, which may also be a cause for different binding capacities. Options for the linear peptide placement optimisation are considered in Chapter 3, but here, an alternative is proposed. Part 1 of Chapter 6 discusses a series of scaffold proteins for presenting magnetite binding loops. One of the discussed scaffolds may be genetically fused to the C-terminal regions of both G52-E-G53 coiled-coil sub-units with hopes that it delivers a more reproducible gain of function.

As the dual function to self-assemble and to interact with magnetite nanoparticles (in case of S1A3) was demonstrated, the question remains whether or not both events can take place simultaneously. In other words, can the assembled two sub-unit fibres be patterned with magnetite nanoparticles. A few attempts at showing this were unsuccessful (not presented in the thesis). The technique of choice was AFM, where only particle aggregates were observed. It has been shown before that protein fibres with specific magnetite binding moieties can be visualised [302], the

study engineered a population of *Salmonella typhimurium* cells to produce flagella exposing a magnetite binding loop WWWSVTEFLRG and used TEM to show binding to pre-formed magnetite nanoparticles. The protocol followed an ultrasonic treatment of particles at 80 % amplitude for 2 minutes. A similar ultra-sonic treatment to that was used in Chapter 5 could be used here to obtain discrete particle binding to S1A3/S2A3 fibres, however, extensive optimisation to balance the two functions (particle binding and protein self-assembly) may be needed before nanoparticle templated fibres can be visualised in the future.

The assembled fibres obtained in liquid were 1000 nm in length, assuming a 17 nm sub-unit length would mean that there are 60 sub-units in a fibre. Assuming that nanoparticle binding was achieved for every sub-unit each fibre would carry 60 bits or 7.5 bytes. In order to store information on any feasible useful scale, much longer fibre assemblies would be needed. There may be space for optimisation of the assembly protocol already used, however, as far as practical utility of the linear fibres with respect to patterning bits is concerned, there may be a need for a compromise to include a method of top-down lithography [303] to create rows of low resolution tracks to anchor S1A3 and S2A3 assemblies (Figure 7.2).

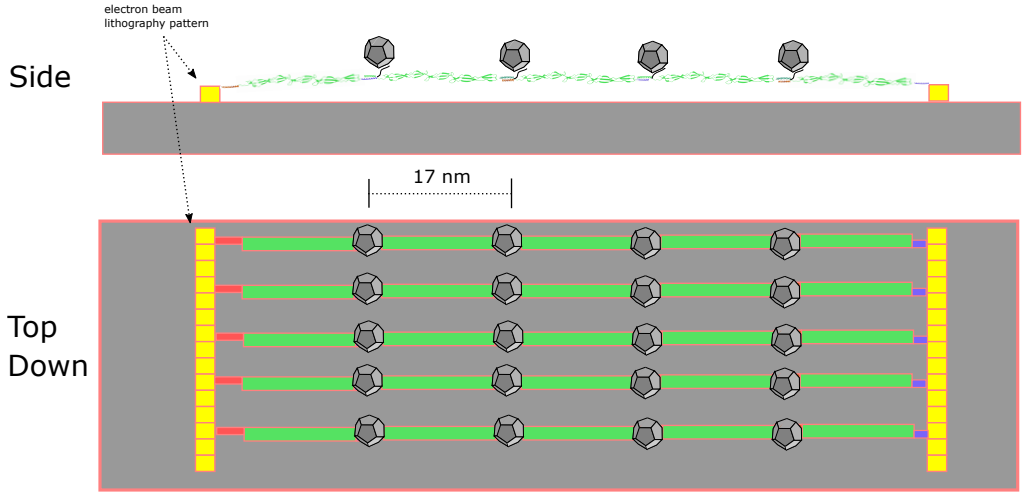


FIGURE 7.2: A compromise strategy for S1A3 and S2A3 based patterning. Lithographically patterned anchoring points (seen in gold). First anchoring sub-units would be applied, then additional soluble sub-units would be added until a formation of a fixed length fibre is achieved. Any other fibre lengths would be washed off.

While interference lithography is a slow process, this strategy would only require patterning of low resolution tracks (e.g. gold) with 1000 nm periodicity. The intervening space would then be patterned by the immobilisation and assembly of G52-E-G53 coiled-coil rod assemblies that would then anchor the particles.

There may yet not be a need for a compromise with 'top-down' lithography methods if the strategy discussed in Chapter 4 proves to be successful. Here, two sets of hub assemblies were shown to have promise in forming extensive structures with G52-E-G53 containing sub-unit S2. Ank4 constructs were validated for their capacity to form a specific interaction with S2, while HR constructs were successfully visualised in AFM. The resulting circular assemblies were a surprising outcome, while multiple reasons may have caused this, it may be that the method of connecting the hub assemblies with S2 through glycine linked coiled-coils is too flexible. SasG G51-E-G52-E-G53 domains were recently shown to be successfully joined in post-translational covalent linkage through use of a set of orthogonal intein linkers [304]. Intein sequences can join adjacent residues into a peptide bond. The study showed assembly of what would correspond to chains of covalently linked G51-E-G52-E-G53 chains spanning 162 nm validated by SDS-PAGE. Microscopy evidence would need to be obtained to prove that such linkage maintains the extended rod structure, however adapting this strategy and placing two pairs of orthogonal inteins (in the same manner as the coiled-coil design) onto the component sets in Chapter 4 may deliver a more constrained assembly system fit for BPM.

Lastly, G52-E-G53 coiled-coil proteins exhibited an unexpected behaviour that was described in Chapter 3 as resembling that of amyloids. Amyloid - is a broad term to describe a fibrillar protein aggregate that is composed of stacked β-sheets [305] (although α-helical versions of this phenomenon have also been discovered [306]). Their formation usually follows a pathway of initial aggregation of a few monomers in an aggregation nucleus (where the β-sheet structure is distorted or changed) that then catalyses the aggregation of other structurally similar proteins in the solution to form long fibres [305]. Amyloid formation is major component of Alzheimer's and Parkinson's diseases [305]. β-sheet amyloid fibre formation has been shown to

be inducible by thermal treatment. It has also been suggested that amyloid fibres may be a useful platform for molecular electronics and tissue engineering [307, 308]. Circular dichroism has been used to study amyloid formation induced structural changes within monomers before[309]. Experiments in this thesis reveal that G52-E-G53 shows no structural changes before and after a 90 °C heating protocol, but showed the amyloid like assemblies on AFM. This may mean that there are either no structural changes during the formation of the fibres, that the structural changes are too subtle to detect on CD or that the fibre formation only takes place during drying on surface in preparation for AFM. Indeed, the study using CD to investigate amyloids, describes a protocol of filtering objects smaller than 100 kDa to eliminate the presence of monomers that may obscure the signal [309]. Further investigations using techniques like thioflavin T aggregation tracking or the observed amyloid-specific intrinsic fluorescence [305] may be of use as a new model for amyloid formation in a disease context, but perhaps more interestingly, the observed G52-E-G53 coiled-coil construct fibres resemble those of spider dragline silk (Figure 7.3).

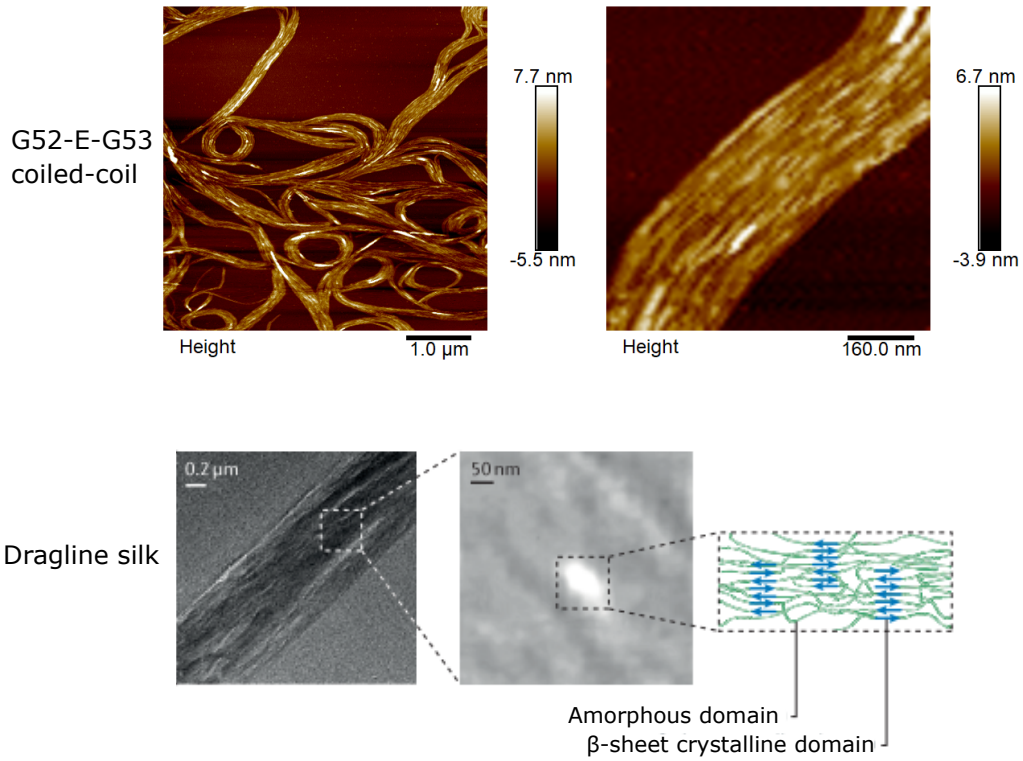


FIGURE 7.3: Top: S1xS2 assembly on AFM after heating to 90 °C. Bottom: spider dragline silk image obtained by bright field transmission electron microscopy. Image adapted from Yarger et.al., 2018 [319].

Spider dragline silk is composed of nanocrystalline β-sheet proteins interspersed with protein regions that are amorphous [319]. These materials exhibit exceptional strength and elasticity. These properties make spider silk a prime target for protein-based materials in textile, optics and electronics, synthetic muscles and biomedicine [310]. The thick long G52-E-G53 protein fibres were obtained through a heating protocol in this project, it may be possible that the spider spinning process exhibiting high sheer forces to partly denature the spider silk proteins [319] would induce a similar change in the G52-E-G53 coiled-coil constructs. Reverse-engineering of these spider silk proteins is an ongoing process [319], thus new alternative protein components like G52-E-G53 coiled-coil constructs may provide a significant addition to the field.

7.3 Chapter 5 Outcomes

Chapter 5 catalogued prior attempts to use S-layers as a basis for metal nano-patterning where approaches of patterning pre-formed nanoparticles and reduction of metal salts on S-layer were tried. This PhD project attempts to use genetic fusion of known biomineralisation and particle binding moieties to grant an S-layer better ability to bind or mineralise nanoparticles. Five different versions of the SgsE S-layer with variable C-termini were produced. While the SgsE-EGFP is a novel modification the non-enhanced GFP protein has been fused to SgsE before [286]. SgsE-STOP and SgsE-Strep showed the production of two-dimensional lattices with expected parameters. It was difficult to obtain a high resolution image to determine if SgsE-A3 also forms a two-dimensional lattice that is not distorted, however given the sequence similarity between A3 and Strep-tag II, one would expect SgsE-A3 to assemble as successfully as SgsE-Strep (especially considering the relative small size of the peptides compared to the whole SgsE protein). Nevertheless, it was shown that SgsE-A3 is ordered along one axis, at least. SgsE-EGFP and SgsE-Mms6 showed distorted lattice parameters. It may that the bulky fusion proteins exposed on the surface of the S-layer are swaying in solution and are thus obscuring the underlying S-layer pattern. In the case of SgsE-Mms6 it is not clear if the lattice distortion is caused by a high presence of a persistent contaminant protein (SDS-PAGE: 69 kDa).

The distorted SgsE-Mms6 lattice was one of the reasons why it was not used in the initial TEM experiments to test binding to 10 nm magnetite nanoparticles. The TEM experiments with SgsE-STOP and SgsE-A3 required robust optimisation of particle preparation and showed particle adherence to the S-layer rods of both proteins. It was not possible to investigate the patterning behavior on S-layer sheets, as in preparations without staining single layer sheets do not provide enough contrast. Despite extensive optimisation, particle agglomeration is still a persistent issue contributing to difficulty in quantifying the difference in particle binding between SgsE-A3 and SgsE-STOP. However, some regions on 10 nm nanoparticle coated SgsE-A3 rods showed localised particle arrangements that may fit the expected S-layer lattice parameters.

The reason for the difficulty in quantification of differential particle binding between SgsE-A3 and SgsE-STOP S-layer sheets was made more clear by the biochemical binding assays, where the difference in binding was shown to be small compared to SgsE-Mms6. SgsE-A3 turned out to be a better binder of mixed size 30-50 nm magnetite nanoparticles. These results were shown by the developed membrane binding experiments and then corroborated by the magnetite ELISA. While SgsE-STOP consistently shows the least amount of binding, TEM indicates that some binding capacity does exist, which is possibly enabled by S-layer porous cavities. What is more, SgsE-STOP showed an almost identical effect on a magnetite synthesis reaction when compared to SgsE-Mms6. The results of binding studies suggest that assay sensitivity may need to be improved. Recently, a study testing iron binding and particle precipitation employed the Xnano LSPR instrument to successfully compare native Mms6 to an Mms6 mutant [192]. Differential particle binding was also tested by using QCM-D [311]. This thesis project was focused on validating the particle binding visually, however with more insight gained it became clear that future ventures to achieve patterning may require robust characterisation of the exact molecular events taking place, thus QCM-D and LSPR techniques may help elucidate and better quantify the extent interaction between metals and their particles with native and modified S-layers.

Since 10 nm particles were chosen for their size compatibility with BPM, finding visual TEM proof of SgsE-Mms6 interacting with the particles would be the next logical step. For patterning however, the incorrect lattice parameters observed in stained SgsE-Mms6 images may be an issue. If the lattice distortion arises from the observed impurity, a different purification protocol may prove useful. Membrane proteins are notorious for difficulties incurred during their expression and purification [312, 313]. Membrane proteins have a hydrophobic exterior that causes their aggregation and accumulation in inclusion bodies. The SgsE-Mms6 fusion protein was produced in the exact same manner as the other constructs, the assumption was that the large precipitous S-layer sheet (that nevertheless precipitates in an ordered fashion unlike inclusion bodies) would align the hydrophobic surfaces of the Mms6 fusion thus hiding them away from the solution. An S-layer fusion to a membrane protein has not been produced before [314]. Perhaps a project to optimise the SgsE-Mms6 purification with detergents used in standard membrane protein production protocols would prove to produce higher purity protein that assembled the correct lattice. Nevertheless, despite the impurities the produced SgsE-Mms6 fusion protein maintained some of the Mms6 functionality as shown by the iron binding experiments. This shows, that in principle, the method of purifying a membrane protein as an S-layer fusion is valid to produce functional proteins. Membrane proteins make up 25-35 % of the human genome, but are over-represented as drug targets [315]. This signifies the importance of finding new and better ways to study them. Further investigations in an S-layer fusion approach may open up new avenues in streamlined membrane protein production.

There still remains the issue of particle to particle agglomeration if an S-layer is to be used as a patterning platform for BPM. Even if the S-layers functionalised with a binding protein/peptide are able to immobilise particles, an insulating layer may be needed to prevent the particles from adhering to each other. In order to maintain the protein exclusive approach a new fusion strategy with a *Methanococcus jannaschii* small Heat Shock protein [320] (MjHSP) is proposed (Figure 7.4).

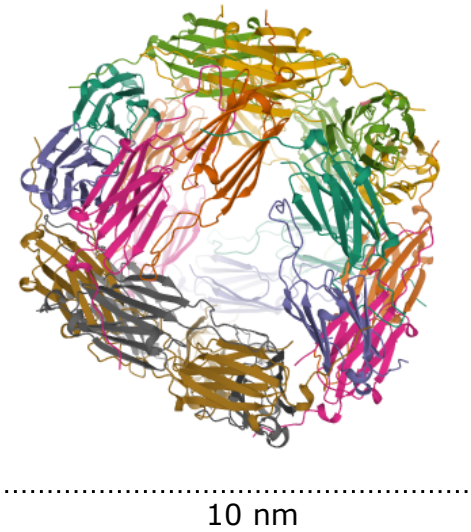


FIGURE 7.4: *Methanococcus jannaschii* small Heat Shock protein cartoon structure. PDB: 1SHS [320].

This MjHSP protein has been shown previously to facilitate mineralisation of CoPt $L1_0$ nanoparticles inside with controlled shape and size [170] by exposing a specific phage display discovered peptide on the interior of the protein-based cage. The cage is an assembly of multiple protein monomers, thus a strategy could be followed, where the sequence for the protein monomer is genetically fused to an S-layer protein. Addition of separately purified soluble MjHSP monomers could then potentially form the protein cages immobilised on the S-layer protein. The MjHSP monomers would also be encoded to contain a biomineralisation peptide. A solution of Co and Pt precursor salts could then be introduced for mineralisation of CoPt particles resulting in an array of physically separated metallic nanoparticles.

7.4 Chapter 6 Outcomes

This Chapter focuses on arguably the most important component of the all-protein BPM assembly strategy - the particle binding interfaces. The work on the protein scaffolds and the E8 peptide reveals that a known binding peptide can maintain its binding capacity regardless of the scaffold that it is in. This is a significant result as it proves that nanoparticle binding peptides can be used as modular components. Peptide scaffold independence has implications for protein based two-dimensional surface patterning. For example, Chapter 5 discusses a fusion protein SgsE-Mms6

that exhibited problems during expression and purification. It may be that one of the soluble scaffolds discussed in Chapter 6 would be more compatible with the correct S-layer lattice assembly than the others scaffolds. Choosing the most compatible scaffold first, knowing that the loop will retain functionality regardless, will make the design of such self-assembly systems easier and more modular. A future study could be designed to test if peptides other than E8 with affinities for magnetite or other metallic nanoparticles would exhibit the same scaffold independence. For such studies different tag-independent particle binding screening methods could be employed for more robust and high-throughput testing [316, 317].

The phage display study revealed three potential binding peptides to CoPt $L1_0$. The most frequently occurring peptide KSPIHVP had an identical N-terminus to KSLSRHD (same length peptide from the rational designed peptide study [193]) and overall similar biochemical properties. The unique effects on the particle size and magnetism indicate that this peptide may be an important addition the set of known CoPt synthesis biotemplation peptides [193, 170]. The excessive presence of proline residues, however, suggests that perhaps the peptide exhibits a different mode of particle synthesis control. A positional scanning survey could be carried out by creating peptide arrays derived from KSPIHVP to understand which residues grant the control over CoPt synthesis using the methods of SPOT peptide array synthesis [318].

The 12 amino acid long peptides previously discovered through phage display have been used to produce CoPt $L1_0$ in an MjHSP protein cage [170]. A future investigation and comparison could be carried out to see if the new KSPIHVP peptide improves this process. A new study could be designed to replace the A3 peptide used in Chapters 3-5 with KSPIHVP to assess if the ability to immobilise CoPt $L1_0$ is maintained in different protein contexts as was shown for S1A3 and SgsEA3 with magnetite nanoparticles. Exchanging all of the A3 peptides for the discovered CoPt $L1_0$ peptide and validating binding would take the protein self-assembly systems characterised here one significant step closed to being usable as biological BPM production building blocks.

7.5 Final Outlook

This work presents the many challenges involved in attempts to use nature's nano-scale control abilities for small magnet patterning. The challenges, however, should not overshadow the exciting new leads discovered here promising easier and more environmentally friendly methods to construct data storage devices and perhaps other types of nano-materials. Protein self-assembly is a source of infinite possibilities if the conditions are right. The way is there, one just has to find it.

Appendix A

Additional information

A.1 Amino Acid Sequences for Chapter 3

Construct	Sequence	Description
S1-A3	MGSSHHHHHHSSGLVPRGSHMASSPE DKNAALKEEIQALEEENQALEEKIAQL KYGGKYGPVKGDSIVEKEEIPFEKERKF NPDLAPGTEKVTREGQKGEKTITTPTL KNPLTGEIISKGESKEEITKDPINELTEY GPETITPGHRDEFDPKLPTGEKEEVPGK PGIKNPETGDVVRPPVDSVTKYGPVKG DSIVEKEEIPFEKERKFNPDLAPGTEKV TREGQKGEKTITTPTLKNPLTGVIISKGE PKEEITKDPINELTEYGPETGGKIAALK QKIAALKYKNAALKKKIAALKQGGGS GSHNHKSKKHK	SasG-G52-E-G53 -domainwith-P6 -and-BN4-α-heli ces-and-A3-mag netite-binding-p eptide

cysS1-A3	MGSSHHHHHHSSGLVPRGCHMASSP EDKNAALKEEIQALEEENQALEEKIAQ LKYGGKYGPVKGDSIVEKEEIPFEKERK FNPDLAPGTEKVTREGQKGEKTITTPT LKNPLTGEIISKGESKEEITKDPINELTE YGPETITPGHRDEFDPKLPTGEKEEVPG KPGIKNPETGDVVRPPVDSVTKYGPVK GDSIVEKEEIPFEKERKFNPDLAPGTEK VTREGQKGEKTITTPTLKNPLTGVIISK GEPKEEITKDPINELTEYGPETGGKIAA LKQKIAALKYKNAALKKKIAALKQGG GSGSHNHKSKKHK	SasG-G52-E-G53 -domainwith-P6 -and-BN4-α-heli ces-and-A3-mag netite-binding-p eptide-N-termin al-cysteine
S1x	MGSSHHHHHHSSGLVPRGSHMASSPE DKNAALKEEIQALEEENQALEEKIAQL KYGGKYGPVKGDSIVEKEEIPFEKERKF NPDLAPGTEKVTREGQKGEKTITTPTL KNPLTGEIISKGESKEEITKDPINELTEY GPETITPGHRDEFDPKLPTGEKEEVPGK PGIKNPETGDVVRPPVDSVTKYGPVKG DSIVEKEEIPFEKERKFNPDLAPGTEKV TREGQKGEKTITTPTLKNPLTGVIISKGE PKEEITKDPINELTEYGPETGGKIAALK QKIAALKYKNAALKKKIAALKQ	SasG-G52-E-G53 -domainwith-P6 -and-BN4-α-heli ces

cysS2-A3	MGSSHHHHHHSSGLVPRGCHMASGG EIAALEQEIAALEKENAALEWEIAALE QGGKYGPVKGDSIVEKEEIPFEKERKFN PDLAPGTEKVTREGQKGEKTITTPTLK NPLTGEIISKGESKEEITKDPINELTEYG PETITPGHRDEFDPKLPTGEKEEVPGKP GIKNPETGDVVRPPVDSVTKYGPVKGD SIVEKEEIPFEKERKFNPDLAPGTEKVT REGQKGEKTITTPTLKNPLTGVIISKGEP KEEITKDPINELTEYGPETGGSPEDENA ALEEKIAQLKQKNAALKEEIQALEYGG SGSHNHKSKKHK	SasG-G52-E-G53 -domain-with-A N4-and-P5-α-he lices-and-A3-ma gnetite-binding- peptide-N-termi nal-cysteine
S2	MGSSHHHHHHSSGLVPRGSHMASGG EIAALEQEIAALEKENAALEWEIAALE QGGKYGPVKGDSIVEKEEIPFEKERKFN PDLAPGTEKVTREGQKGEKTITTPTLK NPLTGEIISKGESKEEITKDPINELTEYG PETITPGHRDEFDPKLPTGEKEEVPGKP GIKNPETGDVVRPPVDSVTKYGPVKGD SIVEKEEIPFEKERKFNPDLAPGTEKVT REGQKGEKTITTPTLKNPLTGVIISKGEP KEEITKDPINELTEYGPETGGSPEDENA ALEEKIAQLKQKNAALKEEIQALEYGG GQRAQSVSKK	SasG-G52-E-G53 -domainwith-A N4-and-P5-α-he lices-and-QRAQ SVSK-magnetite -binding-peptid e

S2A3	MGSSHHHHHHSSGLVPRGSHMASGG EIAALEQEIAALEKENAALEWEIAALE QGGKYGPVKGDSIVEKEEIPFEKERKFN PDLAPGTEKVTREGQKGEKTITTPTLK NPLTGEIISKGESKEEITKDPINELTEYG PETITPGHRDEFDPKLPTGEKEEVPGKP GIKNPETGDVVRPPVDSVTKYGPVKGD SIVEKEEIPFEKERKFNPDLAPGTEKVT REGQKGEKTITTPTLKNPLTGVIISKGEP KEEITKDPINELTEYGPETGGSPEDENA ALEEKIAQLKQKNAALKEEIQALEYGG SGSHNHKSKKHK	SasG-G52-E-G53 -domain-with-A N4-and-P5-α-he lices-and-A3-ma gnetite-binding- peptide
S2x	MGSSHHHHHHSSGLVPRGCHMASGG EIAALEQEIAALEKENAALEWEIAALE QGGKYGPVKGDSIVEKEEIPFEKERKFN PDLAPGTEKVTREGQKGEKTITTPTLK NPLTGEIISKGESKEEITKDPINELTEYG PETITPGHRDEFDPKLPTGEKEEVPGKP GIKNPETGDVVRPPVDSVTKYGPVKGD SIVEKEEIPFEKERKFNPDLAPGTEKVT REGQKGEKTITTPTLKNPLTGVIISKGEP KEEITKDPINELTEYGPETGGSPEDENA ALEEKIAQLKQKNAALKEEIQALEYG	SasG-G52-E-G53 -domain-with-A N4-and-P5-α-hel ices

G5-E-G5	MGSSHHHHHHSSGLVPRGSHMASKY GPVKGDSIVEKEEIPFEKERKFNPDLAP GTEKVTREGQKGEKTITTPTLKNPLTG EIISKGESKEEITKDPINELTEYGPETITP GHRDEFDPKLPTGEKEEVPGKPGIKNP ETGDVVRPPVDSVTKYGPVKGDSIVEK EEIPFEKERKFNPDLAPGTEKVTREGQK GEKTITTPTLKNPLTGVIISKGEPKEEIT KDPINELTEYGPET	SasG-G52-E-G53 -domain
SUMO-AN4	MGSSHHHHHHGGSTSGSDSEVNQEA KPEVKPEVKPETHINLKVSDGSSEIFFK IKKTTPLRRLMEAFAKRQGKEMDSLRF LYDGIRIQADQTPEDLDMEDNDIIEAH REQIGGSGGEIAALEQEIAALEKENAA LEWEIAALEQGS	SUMO-1-follow ed-by-GG-for-cl eavege-and-AN 4-α-helix
SUMO-BN4	MGSSHHHHHHGGSTSGSDSEVNQEA KPEVKPEVKPETHINLKVSDGSSEIFFK IKKTTPLRRLMEAFAKRQGKEMDSLRF LYDGIRIQADQTPEDLDMEDNDIIEAH REQIGGSGKIAALKQKIAALKYKNAAL KKKIAALKQGS	SUMO-1-follow ed-by-GG-for-cl eavege-and-BN4 -α-helix
SUMO-P5	MGSSHHHHHHGGSTSGSDSEVNQEA KPEVKPEVKPETHINLKVSDGSSEIFFK IKKTTPLRRLMEAFAKRQGKEMDSLRF LYDGIRIQADQTPEDLDMEDNDIIEAH REQIGGSSPEDENAALEEKIAQLKQKN AALKEEIQALEYG	SUMO-1-follow ed-by-GG-for-cl eavege-and-P5- α-helix

SUMO-P6	MGSSHHHHHHGGSTSGSDSEVNQEA KPEVKPEVKPETHINLKVSDGSSEIFFK IKKTTPLRRLMEAFAKRQGKEMDSLRF LYDGIRIQADQTPEDLDMEDNDIIEAH REQIGGSSPEDKNAALKEEIQALEEEN QALEEKIAQLKYG	SUMO-1-follow ed-by-GG-for-cl eavege-and-P6-α-helix

TABLE A.1: SasG construct table

A.2 Chapter3: Supplementary figures

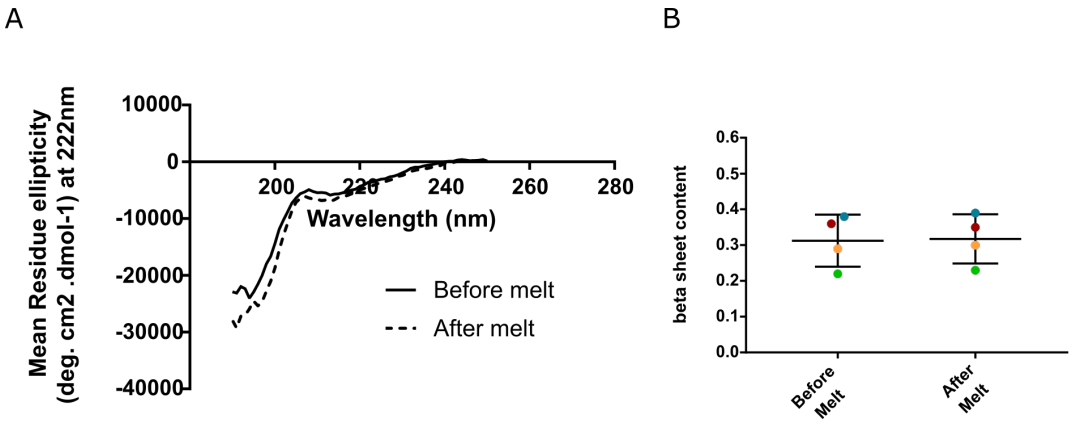


FIGURE A.1: **A** G52-E-G53 spectra in 20 mM phosphate buffer pH 7.4 before (solid line) and after the thermal denaturation experiment (puncture line). **B** Dichroweb CDSSRT fitted β-sheet composition data from spectra in A. Data points represent values obtained using different reference sets. Error bars are standard deviations (population).

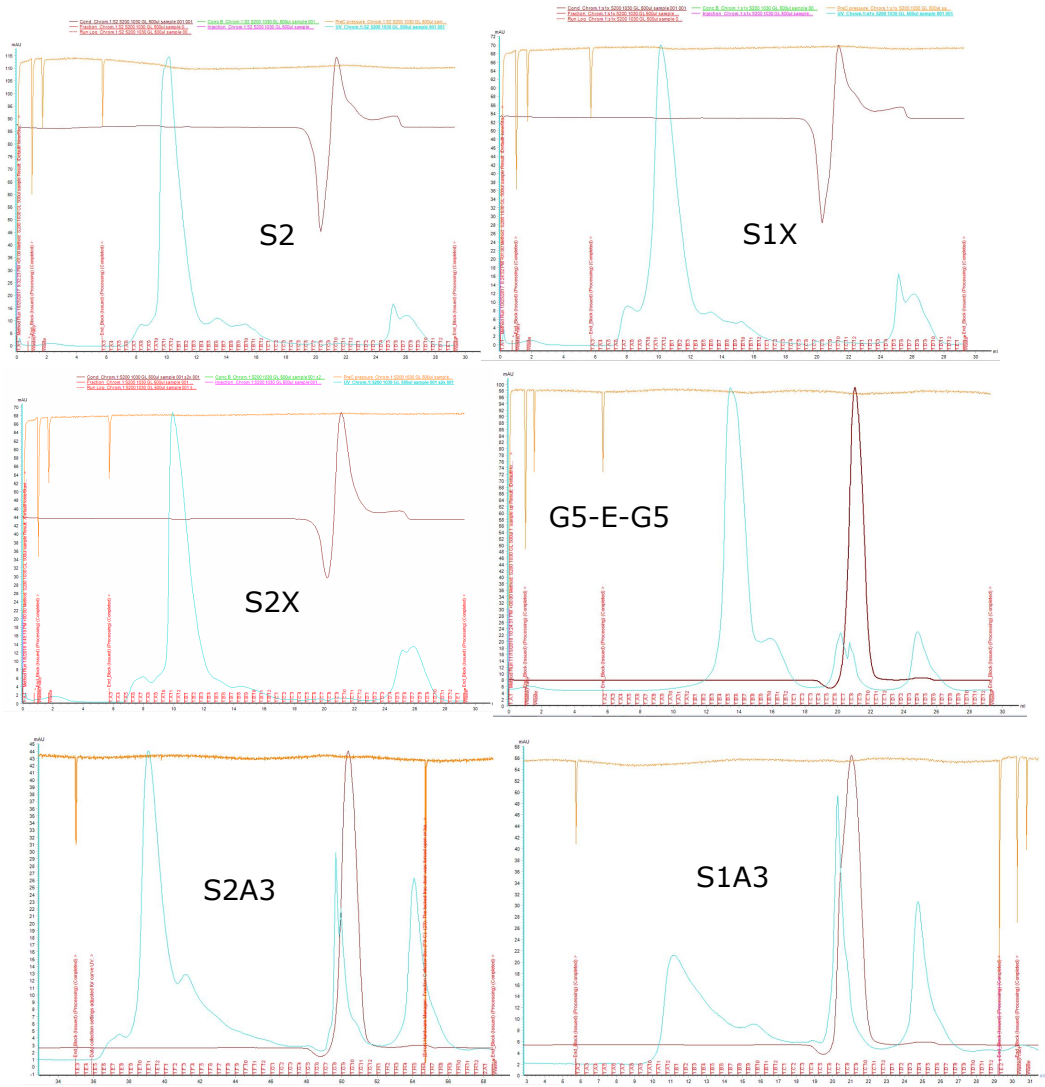


FIGURE A.2: Gel filtration chromatograms of S1x, S2x, S1A3, S2A3, S2 and G52-E-G53 in 20 mM phosphate buffer pH 7.4.

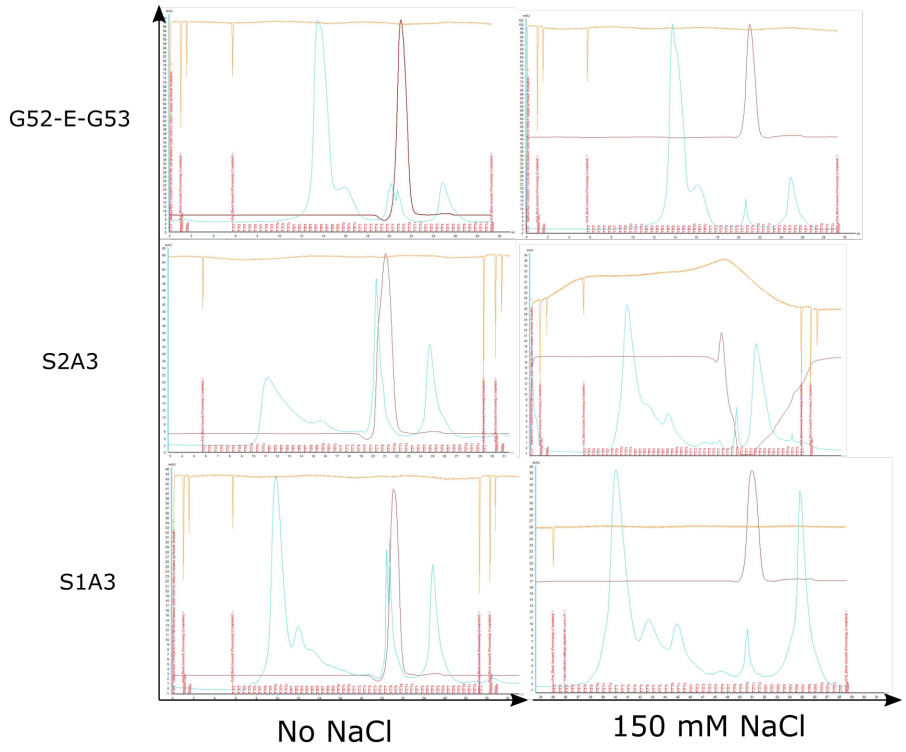


FIGURE A.3: Gel filtration chromatograms with and without 150 mM NaCl in 20 mM phosphate buffer pH 7.4.

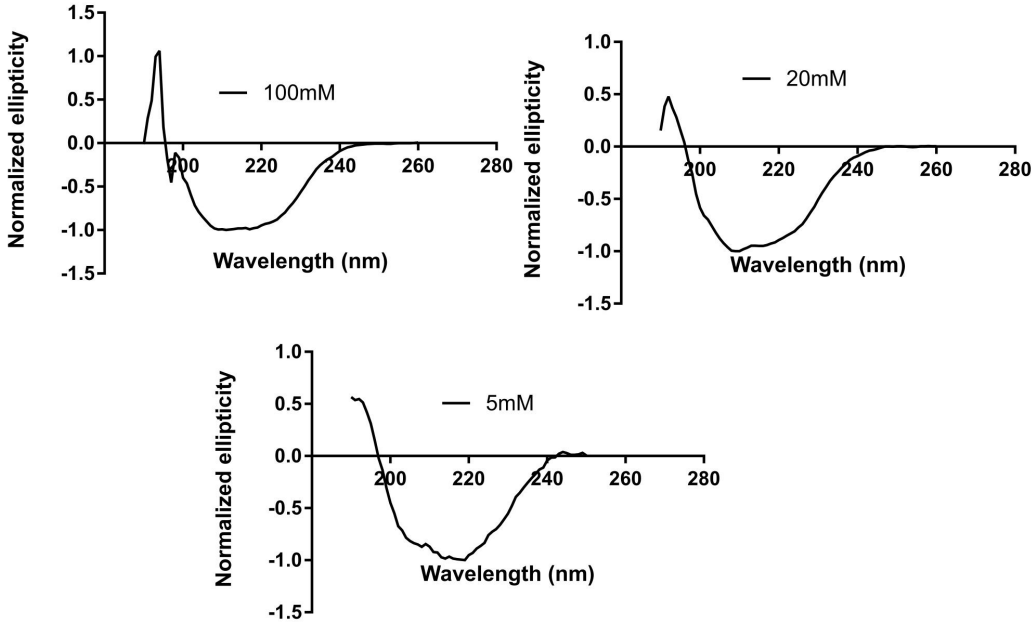


FIGURE A.4: S1x CD spectra over a buffer range 5-100 mM phosphate pH 7.4.

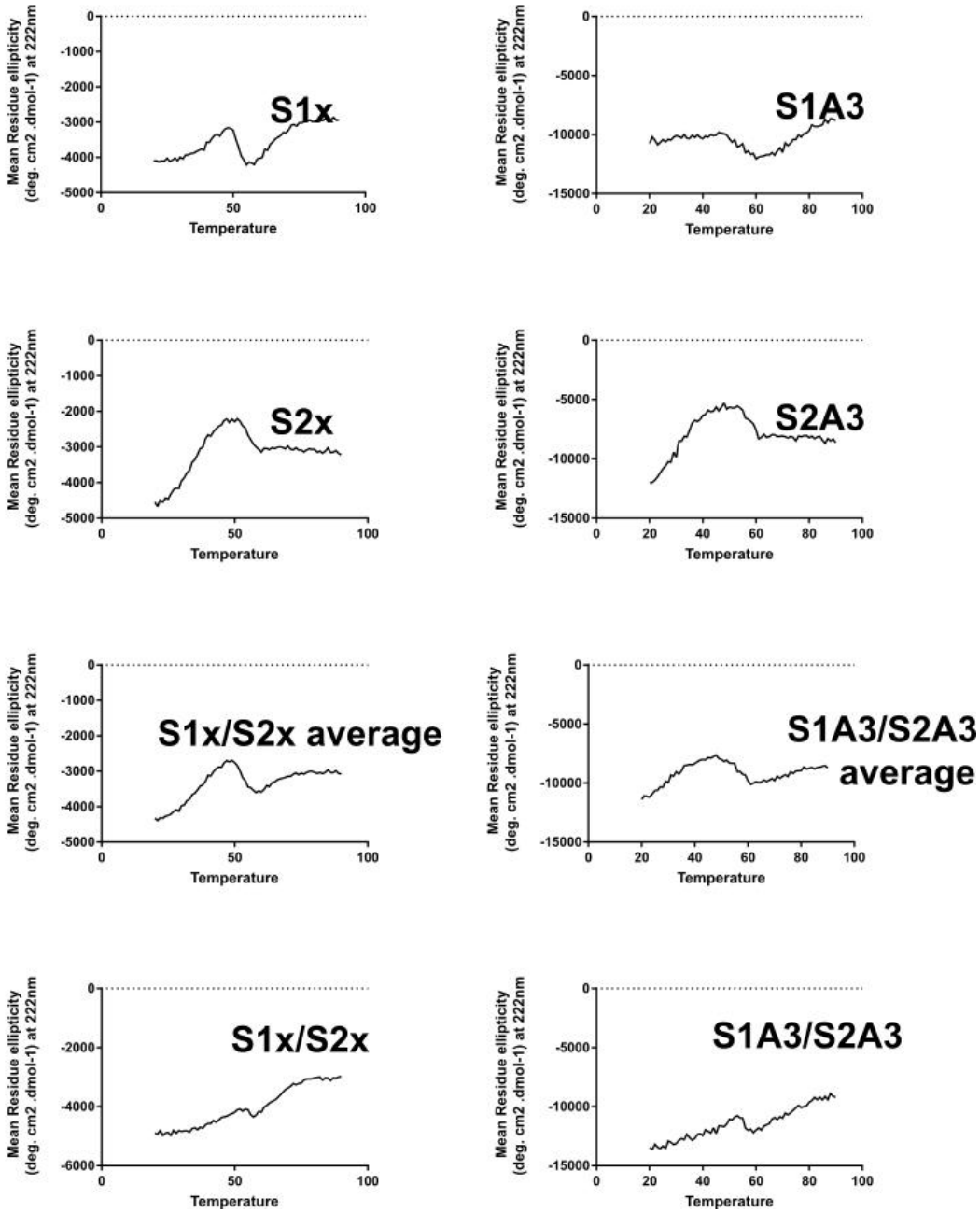


FIGURE A.5: A comparison between thermal denaturation curves for SasG G52-E-G53 constructs with and without the A3 peptide tracking mean residue ellipticity at 222 nm in 20 mM phosphate buffer pH 7.4.

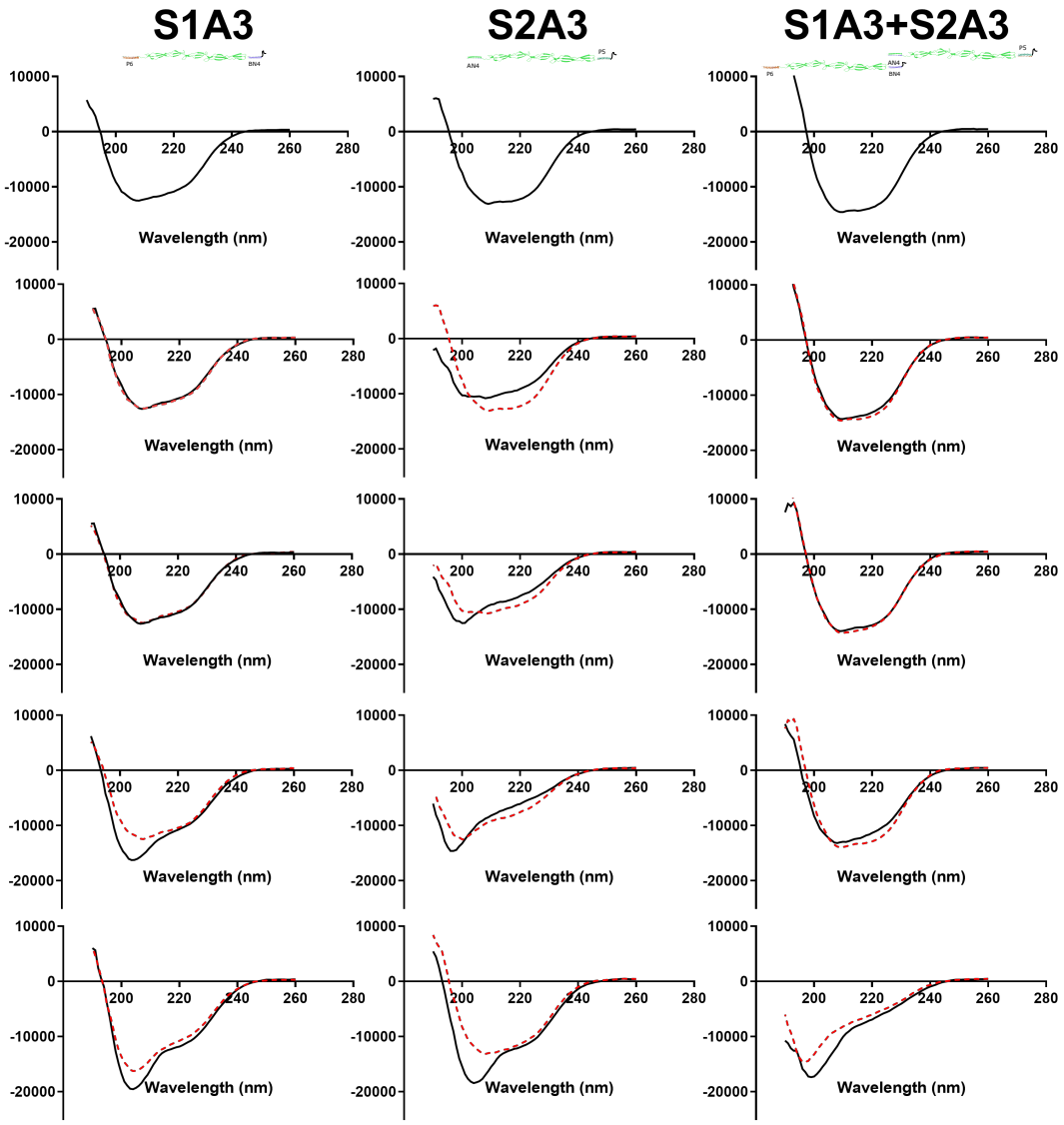


FIGURE A.6: CD spectra of S1A3, S2A3 and S1A3/S2A3 in 20 mM phosphate buffer pH 7.4. Each row represents spectra taken at different temperatures. Traces in red represent the trace for the same protein at a previous temperature step. Proteins analysed at 2.5 μM concentration.

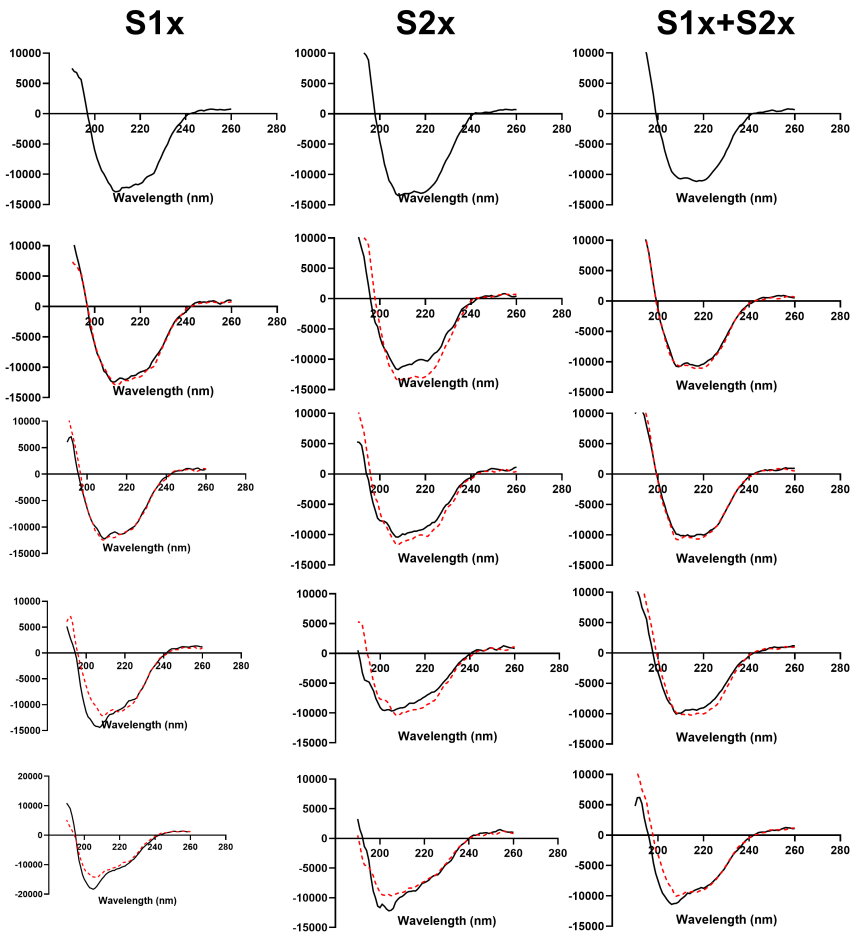


FIGURE A.7: CD spectra of S1x, S2x and S1x/S2x in 20 mM phosphate buffer pH 7.4. Each row represents spectra taken at different temperatures. Traces in red represent the trace for the same protein at a previous temperature step. Proteins analysed at 2.5 μM concentration.

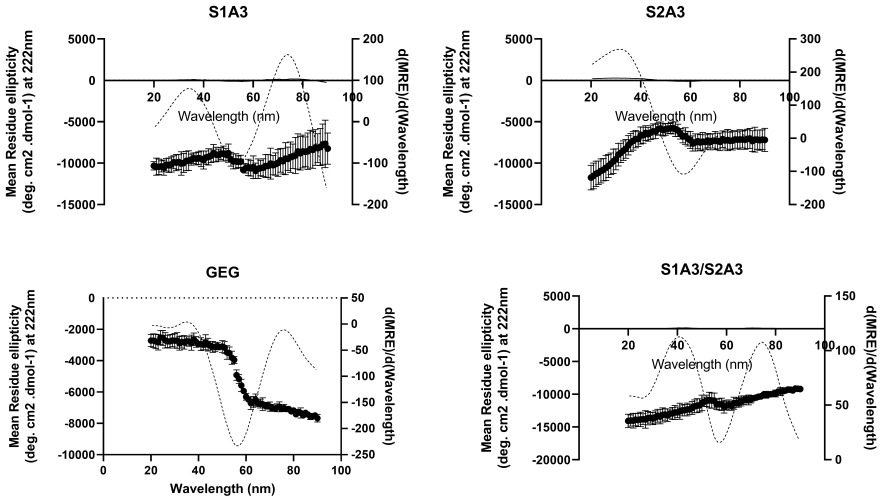


FIGURE A.8: Thermal denaturation curves for S1A3, S2A3, G52-E-G53 and S1A3/S2A3 constructs with and without the A3 peptide tracking mean residue ellipticity at 222 nm in 20 mM phosphate buffer pH 7.4. Data from different experiments was pooled to extract error bards (standard deviation -population). Dotted lines represent first order derivatives of average curves.

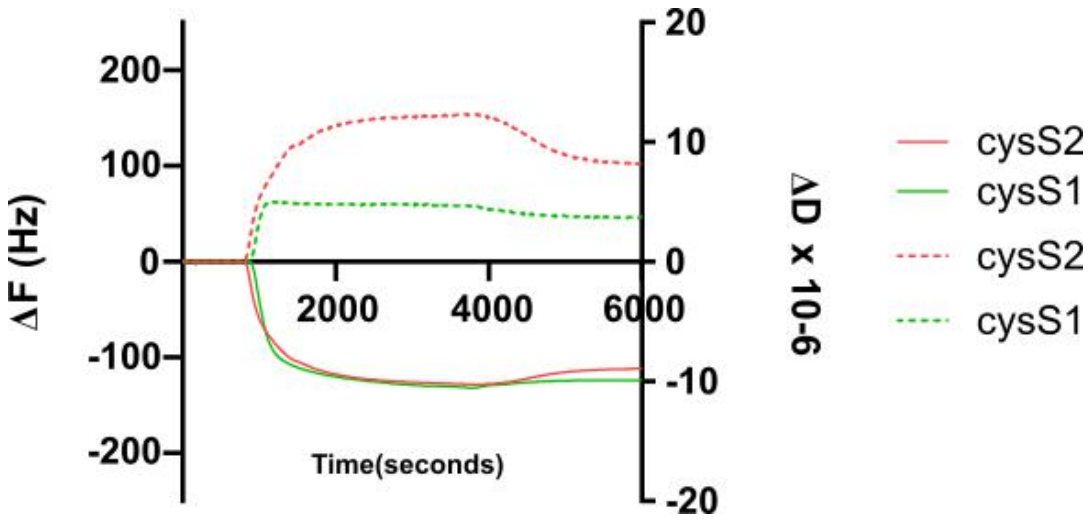


FIGURE A.9: A comparison of dissipation profiles for saturated layers of either cysS1A3 or cysS2A3. For the same mass on surface (determined by the frequency change) cysS1A3 shows a much lower value of dissipation than cysS2A3.

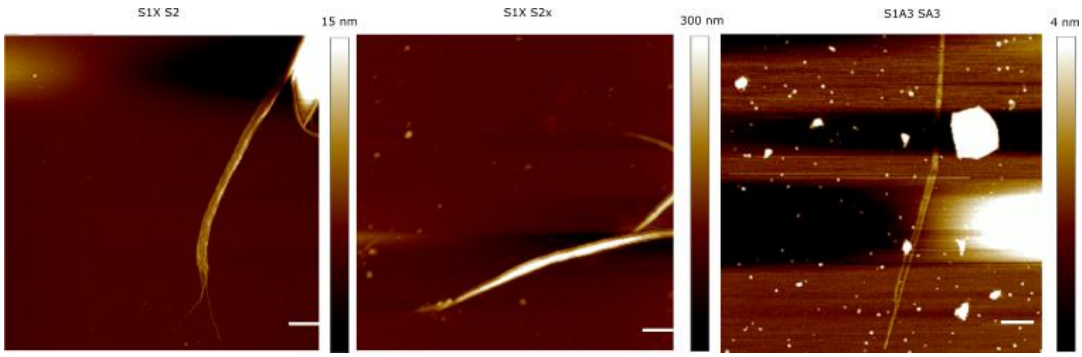


FIGURE A.10: Dry air tapping mode AFM images of S1x/S2, S1x/S2x and S1A3/S2A3 assembled after thermal treatment at 100 ng/ml. Drying accomplished through passive evaporation at room temperature. Scale bar in 500 nm.

A.3 Chapter4: Supplementary figures

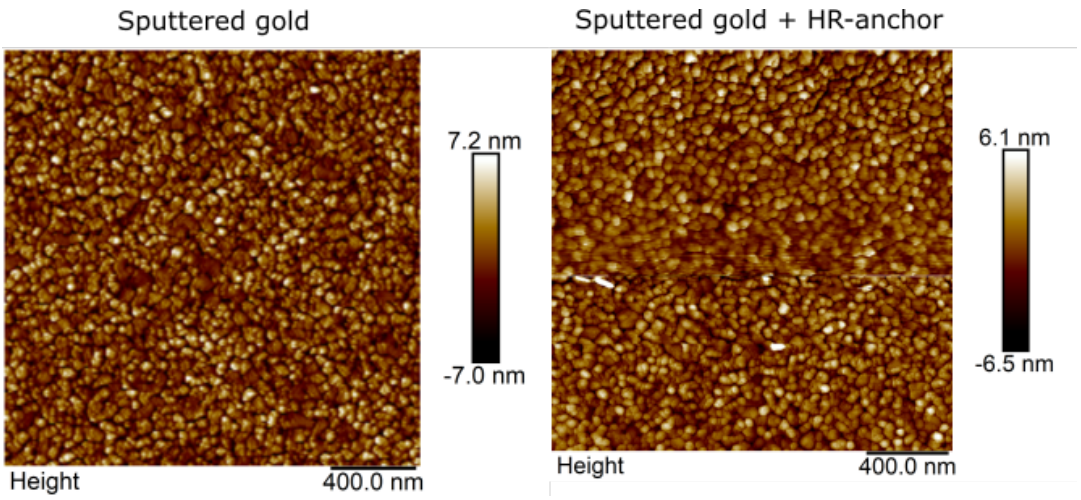


FIGURE A.11: Left: gold surface obtained by sputter deposition [321]. Right: gold surface obtained by sputter deposition incubated with HR-anchor protein for 20 minutes and washed with HPLC grade water 3 times before drying.

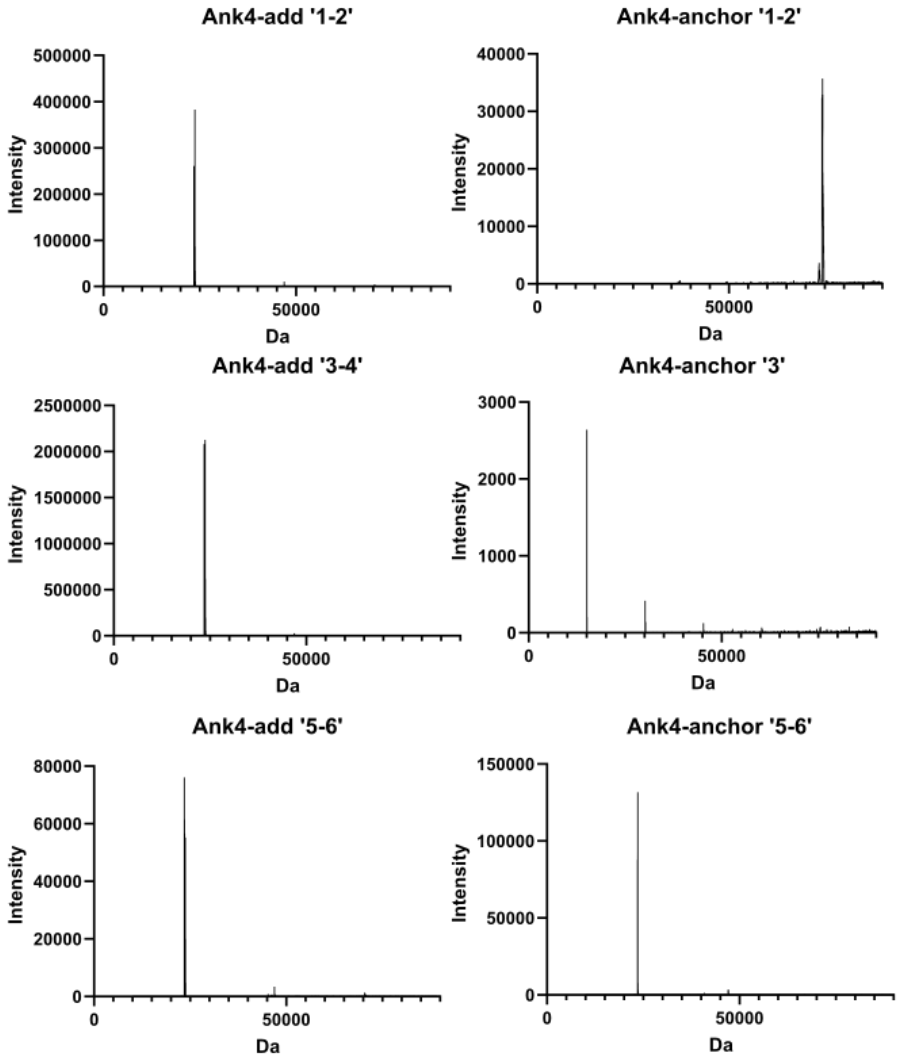


FIGURE A.12: ESI-TOF mass spectra of Ank4 size exclusion chromatography fractions. Left: Ank4-add ESI-TOF spectra. Right: Ank4-anchor ESI-TOF spectra.

Protein	Fraction	Species I	Species II	Species III	Species IV
Ank4-add	fractions 1-2	23469	23715	46937	70404
Ank4-add	fractions 3-4	23469	23715	46936	70404
Ank4-add	fractions 5-6	23715	23469	46937	70403

Ank4-anchor	fractions 1-2	74406	74289	73523	-
Ank4-anchor	fraction 3	15094	30187	45283	-
Ank4-anchor	fractions 5-6	23552	47102	70655	-

TABLE A.2: Tabulated ESI-TOF results from Figure A.12

A.4 Chapter5: Supplementary figures

HHHHHHATKLDKMRQELKAAVDAKDLKKAEELYHKISYELKTRTVILDRVYGQSTR
ELLRSTFKADAQALRDSLIYDITVAMKAREAQDAVKAGNLDKAKAALDQVNQYVSKVTDA
FKAELQKAAQDAKAAYEAALPPKVESVTAVNAKTLEIKFNKAVDAATVIDNKGTSDTSDD
VVKATAITLKAIDDQVPVSTVKASLSDDKKTLKLVVDGAQFFTKRYVVDIKNVKTLDGKD
VPAYTTTIDTTDSVRPSVLSFSYADNGLTLKVKFSEPLASVGTVKLYDGTTEISVSPKFT
AGDDEMTINLASSSVPVNKDLTLKIFGAVDYNGNVINPNPAELTVKKTTVDITKPVVQSI
EAVNTKTVKVTFSEKLLSAPTIKIGGQTASVSVDSTGLVYTATLASALSKGVYAVEVSDY
KDLAGNSGDAYTKVVQLKADNTAPKFVSSQVVKINGVEHLVLTFDEEVTTGSNITVVQSS
DKYIDENNVLKAVGADLKTTSDNFKLYLPTDGKSKSVALNISSLPKGTYTVTLPNGLVSD
LADNPYAERKQITFVRGSDSLTTKPALDKDYDSNGVKADNNNELVFAFTQNLDASALNLS
NFNINGLTVTKAVFDGDTKHIRVTLAPGANTWTGTHVITISNIKNTSGLVMDTVTVNEYM
KENVAPTFTATLTSADVIRVDFSEPVANATISRALSANNFIVKVDGNVVTVSNVYEDNNA
TNLVQGSKGYKTVYLKLQSPVTDLSKPITLSATDIVDVNQTGAITDNNVVGNNVSATVVN
VAKGGS

FIGURE A.13: SgsE sequence as ordered (GeneArt synthesis, Thermofisher Scientific, US).

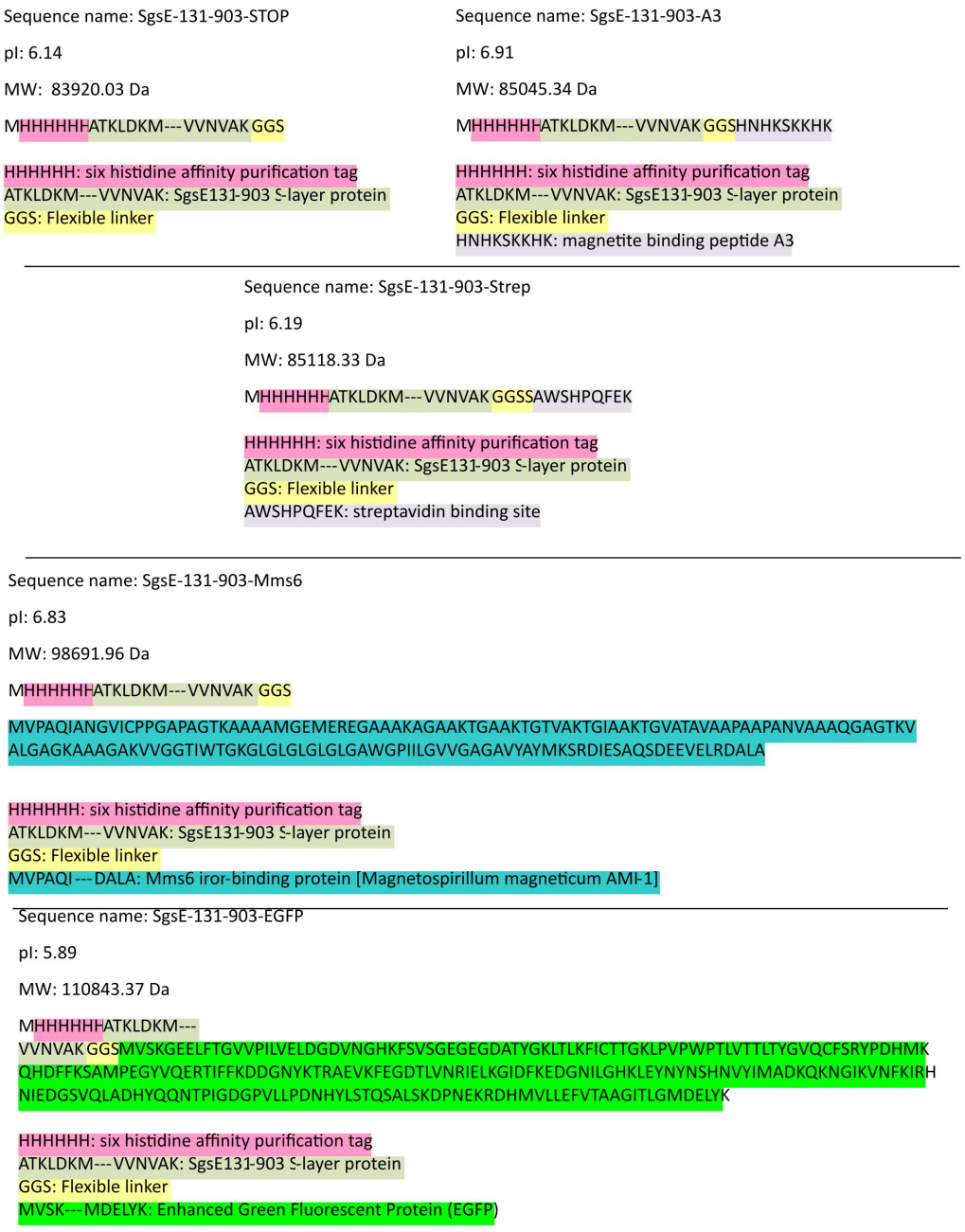


FIGURE A.14: A diagram of all constructs featured in Chapter 5. pI - calculated isoelectric point of the protein, MW - molecular weight (both values calculated using ProtParam (ExPASy).

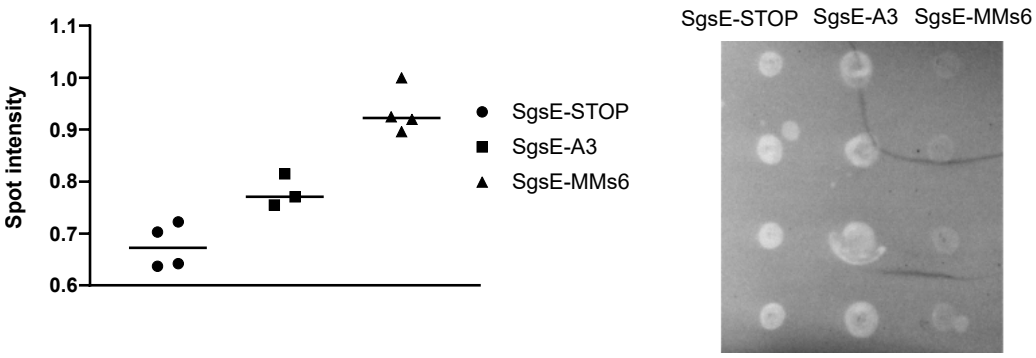


FIGURE A.15: Iron ion binding to SgsE-STOP, SgsE-A3, SgsE-Mms6 spotted nitrocellulose membranes. A 100 ml 50 mM $Iron^{3+}$/Iron-total=0.5 solution was incubated with membranes spotted with 5 µl of each protein at 500 µg/ml for 16 hours. Quantification was performed by measuring spot intensities using ImageLab volume measurement tools and subtracting an average of three same diameter spots found locally around the measured spot.

Appendix B

Rational Design and Self-Assembly of Coiled-Coil Linked SasG Protein Fibrils

The work presented in Chapter 3 was published as:

Jasaitis, Lukas, Callum D. Silver, Andrea E. Rawlings, Daniel T. Peters, Fiona Whelan, Lynne Regan, Laia Pasquina-Lemonche, Jennifer R. Potts, Steven D. Johnson, and Sarah S. Staniland. "Rational Design and Self-Assembly of Coiled-Coil Linked SasG Protein Fibrils." ACS Synthetic Biology 9, no. 7 (2020): 1599-1607.

ACS SyntheticBiology

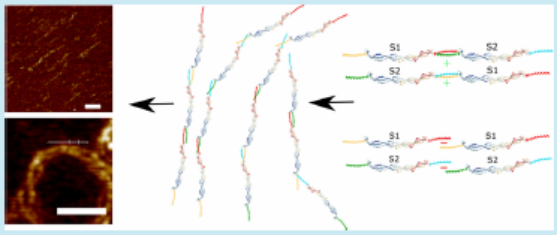
pubs.acs.org/synthbio Letter

Rational Design and Self-Assembly of Coiled-Coil Linked SasG Protein Fibrils

Lukas Jasaitis, Callum D. Silver, Andrea E. Rawlings, Daniel T. Peters, Fiona Whelan, Lynne Regan, Laia Pasquina-Lemonche, Jennifer R. Potts, Steven D. Johnson, and Sarah S. Staniland*

Cite This: ACS Synth. Biol. 2020, 9, 1599−1607 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: Protein engineering is an attractive approach for the self-assembly of nanometer-scale architectures for a range of potential nanotechnologies. Using the versatile chemistry provided by protein folding and assembly, coupled with amino acid side-chain functionality, allows for the construction of precise molecular "protein origami" hierarchical patterned structures for a range of nanoapplications such as stand-alone enzymatic pathways and molecular machines. The *Staphylococcus aureus* surface protein SasG is a rigid, rod-like structure shown to have high mechanical strength due to "clamp-like" intradomain features and a stabilizing interface between the G5 and E domains, making it an excellent building block for molecular self-assembly. Here we characterize a new two subunit system composed of the SasG rod protein genetically conjugated with *de novo* designed coiled-coils, resulting in the self-assembly of fibrils. Circular dichroism (CD) and quartz-crystal microbalance with dissipation (QCM-D) are used to show the specific, alternating binding between the two subunits. Furthermore, we use atomic force microscopy (AFM) to study the extent of subunit polymerization in a liquid environment, demonstrating self-assembly culminating in the formation of linear macromolecular fibrils.

KEYWORDS: *nanorods, fibrils, protein engineering, SasG, self-assembly*

Self-assembling protein fibrils are attractive materials for use in nanotechnology due to a growing knowledge of peptide folding, enabling designed assemblies and versatile selection of chemical functionality. In recent years, numerous studies have focused on creating macromolecular complexes and structures using protein building blocks.[1−3] The geometries of such structures have ranged from relatively simple 1D nanowires[3] to more complex hierarchical architectures.[4] One advantage of using protein building blocks and self-assembly in this way is the ability to design and assemble nanomaterials with precision, reproducibility, and well-defined physical and chemical properties. While there have been many excellent attempts to create *de novo* protein assemblies,[5−7] many still rely on engineering of naturally occurring molecular components.

An example of one such building block with great potential is SasG: a rod-like protein found on the surface of *Staphylococcus aureus* that is implicated in biofilm formation.[8−10] The repetitive region of the protein is composed of repeating G5 and E domains that consist of triple stranded beta-sheets and a collagen-like triple helix without additional covalent interactions (*e.g.*, disulfide bonds).[11] Tandemly arrayed pairs of G5 and E domains form long rod-like structures of specific length depending on the number of alternating 7 nm (G5) and 5 nm (E) domains (Figure 1A).

The structural resilience for both domains is comparable to Ig domains in titin, a polypeptide known as the "strength-paradigm" of protein rigidity and strength.[15] Due to these

Received: March 23, 2020
Published: June 17, 2020

https://dx.doi.org/10.1021/acssynbio.0c00156
ACS Synth. Biol. 2020, 9, 1599−1607

Bibliography

[1] Reinsel David, Gantz John, and Rydning John. *The Digitization of the World*. Tech. rep. 2018.

[2] V. Shobana and N. Kumar. "Big data - A review". In: *International Journal of Applied Engineering Research* 10.55 (2015), pp. 1294–1298. ISSN: 09739769. DOI: 10.26634/jit.6.1.13507.

[3] Paulus Gerdes. "On Mathematics in the History of Sub-Saharan Africa". In: *Historia Mathematica* 21.3 (1994), pp. 345–376. ISSN: 1090249X. DOI: 10.1006/hmat.1994.1029.

[4] Bernhard Bischoff. *Manuscripts and Libraries in the Age of Charlemagne*. Vol. 1. Cambridge University Press, 2007.

[5] Jixing Pan. "On the origin of printing in the light of new archaeological discoveries". In: *Chinese Science Bulletin* 42.12 (1997), pp. 976–981.

[6] Pow Key Sohn. "Printing since the 8th Century in Korea". In: *Koreana* 7 (1993), pp. 4–9.

[7] John Man. *Gutenberg: How one man remade the world with words*. Wiley, 2002.

[8] Stephen J. Greenberg. "The 'Dreadful Visitation': Public health and public awareness in seventeenth-century London". In: *Bulletin of the Medical Library Association* 85.4 (1997), pp. 391–401. ISSN: 00257338.

[9] Peter McMurray. "Once Upon Time: A Superficial History of Early Tape". In: *Twentieth-Century Music* 14.1 (2017), pp. 25–48. ISSN: 14785730. DOI: 10.1017/S1478572217000044.

[10] Rajeshwar Dass. "Review of Various Data Storage Techniques". In: *International Journal of Science and Research* 2.2 (2013), pp. 2319–7064. URL: www.ijsr.net.

[11] Bharat Bhushan. "Historical evolution of magnetic data storage devices and related conferences". In: *Microsystem Technologies* 24.11 (2018), pp. 4423–4436. ISSN: 09467076. DOI: 10.1007/s00542-018-4133-6. URL: https://doi.org/10.1007/s00542-018-4133-6.

[12] S.N. Piramanayagam and Tow C. Chong. *Developments in Data Storage - Materials Perspective*. 2012. ISBN: 9780470501009. URL: http://library1.nida.ac.th/termpaper6/sd/2554/19755.pdf.

[13] Kunpot Mopoung and Sakuntam Sanorpim. "Material development of perpendicular recording for high areal density hard disk drives". In: *Journal of Magnetics* 24.3 (2019), pp. 437–447. ISSN: 12261750. DOI: 10.4283/JMAG.2019.24.3.437.

[14] Martin Strohbach, J€org Daubert, Herman Ravkin, and Mario Lischka. "Big Data Storage". In: *New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe*. 2016, pp. 1–303. ISBN: 9783319215693. DOI: 10.1007/978-3-319-21569-3.

[15] Nathan Ensmenger. "The Environmental History of Computing". In: *Technology and Culture* 59.4 (2018), S7–S33. ISSN: 14699494. DOI: 10.1080/14623528.2020.1715533.

[16] R H Dee. "Magnetic Tape for Data Storage: An Enduring Technology". In: *Proceedings of the IEEE* 96.11 (2008), pp. 1775–1785. ISSN: 1558-2256. DOI: 10.1109/JPROC.2008.2004311.

[17] Luis Ceze, Jeff Nivala, and Karin Strauss. "Molecular digital data storage using DNA". In: *Nature Reviews Genetics* 20.8 (2019), pp. 456–466. ISSN: 14710064. DOI: 10.1038/s41576-019-0125-3. URL: http://dx.doi.org/10.1038/s41576-019-0125-3.

[18] Y. Shiroishi, K. Fukuda, I. Tagawa, H. Iwasaki, S. Takenoiri, H. Tanaka, H. Mutoh, and N. Yoshikawa. "Future options for HDD storage". In: *IEEE Transactions on Magnetics* 45.10 (2009), pp. 3816–3822. ISSN: 00189464. DOI: 10.1109/TMAG.2009.2024879.

[19] Nicola A. Spaldin. *Magnetic Materials Fundamentals and Applications*. Cambridge University Press, 2012. ISBN: 9780511781599. DOI: https://doi.org/10.1017/CBO9780511781599.

[20] B D (Bernard Dennis) Cullity. *Introduction to magnetic materials*. 2nd ed. Hoboken, N.J.: Wiley for IEEE, 2005. ISBN: 0471477419.

[21] Srikanth Singamaneni, Valery N. Bliznyuk, Christian Binek, and Evgeny Y. Tsymbal. "Magnetic nanoparticles: Recent advances in synthesis, self-assembly and applications". In: *Journal of Materials Chemistry* 21.42 (2011), pp. 16819–16845. ISSN: 09599428. DOI: 10.1039/c1jm11845e.

[22] Leena Mohammed, Hassan G. Gomaa, Doaa Ragab, and Jesse Zhu. "Magnetic nanoparticles for environmental and biomedical applications: A review". In: *Particuology* 30 (2017), pp. 1–14. ISSN: 16742001. DOI: 10.1016/j.partic.2016.06.001. URL: http://dx.doi.org/10.1016/j.partic.2016.06.001.

[23] Saima Gul, Sher Bahadar Khan, Inayat Ur Rehman, Murad Ali Khan, and M. I. Khan. "A Comprehensive Review of Magnetic Nanomaterials Modern Day Theranostics". In: *Frontiers in Materials* 6.July (2019), pp. 1–15. ISSN: 22968016. DOI: 10.3389/fmats.2019.00179.

[24] Amyn S. Teja and Pei-Yoong Koh. "Synthesis, properties, and applications of magnetic iron oxide nanoparticles". In: *Progress in Crystal Growth and Characterization of Materials* 55.1 (2009), pp. 22 –45. ISSN: 0960-8974. DOI: https://doi.org/10.1016/j.pcrysgrow.2008.08.003. URL: http://www.sciencedirect.com/science/article/pii/S0960897408000168.

[25] Siddharth V Patwardhan and Sarah S Staniland. *Green Nanomaterials*. 2053-2563. IOP Publishing, 2019. ISBN: 978-0-7503-1221-9. DOI: 10.1088/978-0-7503-1221-9. URL: http://dx.doi.org/10.1088/978-0-7503-1221-9.

[26] Yoshiteru Mizukoshi, Tatsuya Shuto, Naoya Masahashi, and Shuji Tanabe. "Preparation of superparamagnetic magnetite nanoparticles by reverse precipitation method: Contribution of sonochemically generated oxidants". In: *Ultrasonics Sonochemistry* 16.4 (2009), pp. 525–531. ISSN: 13504177. DOI: 10.1016/j.ultsonch.2008.12.017. URL: http://dx.doi.org/10.1016/j.ultsonch.2008.12.017.

[27] Sophie Laurent, Delphine Forge, Marc Port, Alain Roch, Caroline Robic, Luce Vander Elst, and Robert N Muller. "Magnetic iron oxide nanoparticles: synthesis, stabilization, vectorization, physicochemical characterizations, and biological applications". In: *Chemical reviews* 108.6 (2008), pp. 2064–2110.

[28] Attarad Ali, Hira Zafar, Muhammad Zia, Ihsan ul Haq, Abdul Rehman Phull, Joham Sarfraz Ali, and Altaf Hussain. "Synthesis, characterization, applications, and challenges of iron oxide nanoparticles". In: *Nanotechnology, Science and Applications* 9 (2016), pp. 49–67. ISSN: 11778903. DOI: 10.2147/NSA.S99986.

[29] Robert L. Rebodos and Peter J. Vikesland. "Effects of oxidation on the magnetization of nanoparticulate magnetite". In: *Langmuir* 26.22 (2010), pp. 16745–16753. ISSN: 07437463. DOI: 10.1021/la102461z.

[30] Qing Li, Christina W. Kartikowati, Shinji Horie, Takashi Ogi, Toru Iwaki, and Kikuo Okuyama. "Correlation between particle size/domain structure and magnetic properties of highly crystalline Fe3O4 nanoparticles". In: *Scientific Reports* 7.1 (2017), pp. 1–4. ISSN: 20452322. DOI: 10.1038/s41598-017-09897-5. URL: http://dx.doi.org/10.1038/s41598-017-09897-5.

[31] Tsann Lin. "Magnetic, recording and structural characteristics of sputtered co- cr-pt films for longitudinal recording". In: *Journal of Magnetism and Magnetic Materials* 86 (1989), pp. 159–168.

[32] Yaxin Wang, Xiaolong Zhang, Yang Liu, Yuhong Jiang, Yongjun Zhang, and Jinghai Yang. "Microstructure and magnetic properties of L1 0 CoPt nanoparticles by Ag addition". In: *Journal of sol-gel science and technology* 70.3 (2014), pp. 528–533.

[33] Cheikh Ndao. "Direct Method for Qualitative Assessment of the Fraction of Ordered and Disordered Phases in Thick FePt Films". In: *Journal of Magnetism and Magnetic Materials* 403 (2015). DOI: 10.1016/j.jmmm.2015.11.077.

[34] Jai Lin Tsai, Hsin Te Tzeng, and Bing Fong Liu. "Magnetic properties and microstructure of graded Fe/FePt films". In: *Journal of Applied Physics* 107.11 (2010). ISSN: 00218979. DOI: 10.1063/1.3446198.

[35] Jeotikanta Mohapatra, Meiying Xing, Jacob Elkins, and J. Ping Liu. "Hard and semi-hard magnetic materials based on cobalt and cobalt alloys". In: *Journal of Alloys and Compounds* 824 (2020), p. 153874. ISSN: 09258388. DOI: 10.1016/j.jallcom.2020.153874. URL: https://doi.org/10.1016/j.jallcom.2020.153874.

[36] J. A. ABOAF, S. R. HERD and E. KLOKHOLM. "Magnetic Properties and Structure of Cobalt- Platinum Thin Films". In: *IEEE Transactions on Magnetics* 19 (1983), p. 1514. ISSN: 00218979. DOI: 10.1063/1.1682811.

[37] C. Petit, S. Rusponi, and H. Brune. "Magnetic properties of cobalt and cobalt-platinum nanocrystals investigated by magneto-optical Kerr effect". In: *Journal of Applied Physics* 95.8 (2004), pp. 4251–4260. ISSN: 00218979. DOI: 10.1063/1.1686906.

[38] Y. J. Zhang, Y. T. Yang, Y. Liu, Y. X. Wang, L. L. Yang, M. B. Wei, H. G. Fan, H. J. Zhai, X. Y. Liu, Y. Q. Liu, N. N. Yang, Y. H. Wu, and J. H. Yang. "A novel approach to the synthesis of CoPt magnetic nanoparticles". In: *Journal of Physics D: Applied Physics* 44.29 (2011). ISSN: 00223727. DOI: 10.1088/0022-3727/44/29/295003.

[39] Yang Liu, Yanting Yang, Yongjun Zhang, Yaxin Wang, Xiaolong Zhang, Yuhong Jiang, Maobin Wei, Yanqing Liu, Xiaoyan Liu, and Jinghai Yang. "A facile route to synthesis of CoPt magnetic nanoparticles". In: *Materials Research Bulletin* 48.2 (2012), pp. 721–724. ISSN: 00255408. DOI: 10.1016/j.materresbull.2012.11.019. URL: http://dx.doi.org/10.1016/j.materresbull.2012.11.019.

[40] Prakash Karipoth and Raphael Justin Joseyphus. "Evolution of High Coercivity in CoPt Nanoparticles Through Nitrogen Assisted Annealing". In: *Journal of Superconductivity and Novel Magnetism* 27.9 (2014), pp. 2123–2130. ISSN: 15571947. DOI: 10.1007/s10948-014-2564-6.

[41] Scott J Kemp, R Matthew Ferguson, Amit P Khandhar, and Kannan M Krishnan. "Monodisperse magnetite nanoparticles with nearly ideal saturation magnetization". In: *RSC advances* 6.81 (2016), pp. 77452–77464.

[42] Thomas R. Albrecht, Hitesh Arora, Vipin Ayanoor-Vitikkate, Jean Marc Beaujour, Daniel Bedau, David Berman, Alexei L. Bogdanov, Yves Andre Chapuis, Julia Cushen, Elizabeth E. Dobisz, Gregory Doerk, He Gao, Michael Grobis, Bruce Gurney, Weldon Hanson, Olav Hellwig, Toshiki Hirano, Pierre Olivier Jubert, Dan Kercher, Jeffrey Lille, Zuwei Liu, C. Mathew Mate, Yuri Obukhov, Kanaiyalal C. Patel, Kurt Rubin, Ricardo Ruiz, Manfred Schabes, Lei Wan, Dieter Weller, Tsai Wei Wu, and En Yang. "Bit-patterned magnetic recording: Theory, media fabrication, and recording performance". In: *IEEE Transactions on Magnetics* 51.5 (2015). ISSN: 00189464. DOI: 10.1109/TMAG.2015.2397880. arXiv: 1503.06664.

[43] Thomas R. Albrecht, Hitesh Arora, Vipin Ayanoor-Vitikkate, Jean Marc Beaujour, Daniel Bedau, David Berman, Alexei L. Bogdanov, Yves Andre Chapuis, Julia Cushen, Elizabeth E. Dobisz, Gregory Doerk, He Gao, Michael Grobis, Bruce Gurney, Weldon Hanson, Olav Hellwig, Toshiki Hirano, Pierre Olivier Jubert, Dan Kercher, Jeffrey Lille, Zuwei Liu, C. Mathew Mate, Yuri Obukhov, Kanaiyalal C. Patel, Kurt Rubin, Ricardo Ruiz, Manfred Schabes, Lei Wan, Dieter Weller, Tsai Wei Wu, and En Yang. "Bit-patterned magnetic recording: Theory, media fabrication, and recording performance". In: *IEEE Transactions on Magnetics* 51.5 (2015). ISSN: 00189464. DOI: 10.1109/TMAG.2015.2397880. arXiv: 1503.06664.

[44] Cherie R. Kagan, Taeghwan Hyeon, Dae Hyeong Kim, Ricardo Ruiz, Maryann C. Tung, and H. S.Philip Wong. "Self-Assembly for electronics". In: *MRS Bulletin* 45.10 (2020), pp. 807–814. ISSN: 08837694. DOI: 10.1557/mrs.2020.248.

[45] C. Brombacher, M. Grobis, J. Lee, J. Fidler, T. Eriksson, T. Werner, O. Hellwig, and M. Albrecht. "L1 0 FePtCu bit patterned media". In: *Nanotechnology* 23.2 (2012), pp. 0–4. ISSN: 09574484. DOI: 10.1088/0957-4484/23/2/025301.

[46] Stanley N Cohen, Annie C Y Chang, Herbert W Boyer, and Robert B Helling. "Construction of biologically functional bacterial plasmids in vitro". In: *Proceedings of the National Academy of Sciences* 70.11 (1973), pp. 3240–3244.

[47] Wangsa Tirta Ismaya. "The Role of Protein Biochemistry in Biotechnology". In: *Inovasi Online* 19.23 (2014), pp. 11–18.

[48] David N. Breslauer. "Recombinant Protein Polymers: A Coming Wave of Personal Care Ingredients". In: *ACS Biomaterials Science and Engineering* 6.11 (2020), pp. 5980–5986. ISSN: 23739878. DOI: 10.1021/acsbiomaterials.0c01038.

[49] Jerome A. Werkmeister and John A.M. Ramshaw. "Recombinant protein scaffolds for tissue engineering". In: *Biomedical Materials* 7.1 (2012). ISSN: 1748605X. DOI: 10.1088/1748-6041/7/1/012002.

[50] Nadia C Abascal and Lynne Regan. "The past , present and future of protein-based materials". In: (2018).

[51] Aleksandra P. Kiseleva, Pavel V. Krivoshapkin, and Elena F. Krivoshapkina. "Recent Advances in Development of Functional Spider Silk-Based Hybrid Materials". In: *Frontiers in Chemistry* 8.June (2020), pp. 1–20. ISSN: 22962646. DOI: 10.3389/fchem.2020.00554.

[52] Jun Xu, Qinglin Dong, Ye Yu, Baolong Niu, Dongfeng Ji, Muwang Li, Yongping Huang, Xin Chen, and Anjiang Tan. "Mass spider silk production through targeted gene replacement in Bombyx mori". In: *Proceedings of the National Academy of Sciences of the United States of America* 115.35 (2018), pp. 8757–8762. ISSN: 10916490. DOI: 10.1073/pnas.1806805115.

[53] Yingfeng Li, Ke Li, Xinyu Wang, Mengkui Cui, Peng Ge, Junhu Zhang, Feng Qiu, Chao Zhong, and Chao Zhong. "Conformable self-assembling amyloid protein coatings with genetically programmable functionality". In: *Science Advances* 6.21 (2020), pp. 1–12. ISSN: 23752548. DOI: 10.1126/sciadv.aba1425.

[54] Huajie Liu and Dongsheng Liu. "DNA nanomachines and their functional evolution". In: *Chemical Communications* 19 (2009), pp. 2625–2636. ISSN: 13597345. DOI: 10.1039/b822719e.

[55] K. E. Dunn, M. C. Leake, A. J.M. Wollman, M. A. Trefzer, S. Johnson, and A. M. Tyrrell. "An experimental study of the putative mechanism of a synthetic autoomous rotary DNA nanomotor". In: *Royal Society Open Science* 4.3 (2017). ISSN: 20545703. DOI: 10.1098/rsos.160767.

[56] Pengfei Wang, Travis A. Meyer, Victor Pan, Palash K. Dutta, and Yonggang Ke. "The Beauty and Utility of DNA Origami". In: *Chem* 2.3 (2017), pp. 359–

382. ISSN: 24519294. DOI: 10.1016/j.chempr.2017.02.009. URL: http://dx.doi.org/10.1016/j.chempr.2017.02.009.

[57] Kosti Tapio and Ilko Bald. "The potential of DNA origami to build multifunctional materials". In: *Multifunctional Materials* 3.3 (2020). ISSN: 23997532. DOI: 10.1088/2399-7532/ab80d5.

[58] Katherine E. Dunn, Martin A. Trefzer, Steven Johnson, and Andy M. Tyrrell. "Towards a bioelectronic computer: A theoretical study of a multi-layer biomolecular computing system that can process electronic inputs". In: *International Journal of Molecular Sciences* 19.9 (2018). ISSN: 14220067. DOI: 10.3390/ijms19092620.

[59] Russell Deaton, Max Garzon, John Rose, D. R. Franceschetti, and S. E. Stevens. "DNA Computing: A Review". In: *Fundamenta Informaticae* 35.1-4 (1998), pp. 231–245. ISSN: 01692968. DOI: 10.3233/fi-1998-35123413.

[60] Isaac Gállego, Brendan Manning, Joan Daniel Prades, Mònica Mir, Josep Samitier, and Ramon Eritja. "DNA-Origami-Aided Lithography for Sub-10 Nanometer Pattern Printing". In: *Proceedings* 1.10 (2017), p. 325. ISSN: 2504-3900. DOI: 10.3390/proceedings1040325.

[61] Jeanette Nangreave, Dongran Han, Yan Liu, and Hao Yan. "DNA origami: A history and current perspective". In: *Current Opinion in Chemical Biology* 14.5 (2010), pp. 608–615. ISSN: 13675931. DOI: 10.1016/j.cbpa.2010.06.182. URL: http://dx.doi.org/10.1016/j.cbpa.2010.06.182.

[62] Tara D. Sutherland, Mickey G. Huson, and Trevor D. Rapson. "Rational design of new materials using recombinant structural proteins: Current state and future challenges". In: *Journal of Structural Biology* 201.1 (2018), pp. 76–83. ISSN: 10958657. DOI: 10.1016/j.jsb.2017.10.012. URL: https://doi.org/10.1016/j.jsb.2017.10.012.

[63] Salam Al Karadaghi. "Basics of Protein Structure". In: (2014). URL: http://www.proteinstructures.com/Structure/Structure/amino-acids.html.

[64] C. Chothia. "Principles that determine the structure of proteins." In: *Annual review of biochemistry* 53 (1984), pp. 537–572. ISSN: 00664154. DOI: 10.1146/annurev.bi.53.070184.002541.

[65] C. Nick Pace and J. Martin Scholtz. "A helix propensity scale based on experimental studies of peptides and proteins". In: *Biophysical Journal* 75.1 (1998), pp. 422–427. ISSN: 00063495. DOI: 10.1016/s0006-3495(98)77529-0. URL: http://dx.doi.org/10.1016/S0006-3495(98)77529-0.

[66] Helen Berman, Kim Henrick, and Haruki Nakamura. "Announcing the worldwide Protein Data Bank". In: *Nature Structural Biology* 10.12 (2003), p. 980. ISSN: 10728368. DOI: 10.1038/nsb1203-980.

[67] Ian Sillitoe, Natalie Dawson, Tony E. Lewis, Sayoni Das, Jonathan G. Lees, Paul Ashford, Adeyelu Tolulope, Harry M. Scholes, Ilya Senatorov, Andra Bujan, Fatima Ceballos Rodriguez-Conde, Benjamin Dowling, Janet Thornton, and Christine A. Orengo. "CATH: Expanding the horizons of structure-based functional annotations for genome sequences". In: *Nucleic Acids Research* 47.D1 (2019), pp. D280–D284. ISSN: 13624962. DOI: 10.1093/nar/gky1097.

[68] Rachel Kolodny, Leonid Pereyaslavets, Abraham O. Samson, and Michael Levitt. "On the universe of protein folds". In: *Annual Review of Biophysics* 42.1 (2013), pp. 559–582. ISSN: 1936122X. DOI: 10.1146/annurev-biophys-083012-130432.

[69] Chaohua Zhu, Gang Guo, Qiqi Ma, Fengjuan Zhang, Funing Ma, Jianping Liu, Dao Xiao, Xiaolin Yang, and Ming Sun. "Diversity in S-layers". In: *Progress in Biophysics and Molecular Biology* 123 (2017), pp. 1–15. ISSN: 00796107. DOI: 10.1016/j.pbiomolbio.2016.08.002. URL: http://dx.doi.org/10.1016/j.pbiomolbio.2016.08.002.

[70] Dietmar Pum, Angela Neubauer, Erika Györvary, Margit Sára, and Uwe B. Sleytr. "S-layer proteins as basic building blocks in a biomolecular construction kit". In: *Nanotechnology* 11.2 (2000), pp. 100–107. ISSN: 09574484. DOI: 10.1088/0957-4484/11/2/310.

[71] Dietmar Pum and Uwe B. Sleytr. "Reassembly of S-layer proteins". In: *Nanotechnology* 25.31 (2014). ISSN: 13616528. DOI: 10.1088/0957-4484/25/31/312001.

[72] Ekaterina Baranova, Rémi Fronzes, Abel Garcia-Pino, Nani Van Gerven, David Papapostolou, Gérard Péhau-Arnaudet, Els Pardon, Jan Steyaert, Stefan Howorka, and Han Remaut. "SbsB structure and lattice reconstruction unveil Ca2+ triggered S-layer assembly". In: *Nature* 487.7405 (2012), pp. 119–122. ISSN: 00280836. DOI: 10.1038/nature11155.

[73] Ch Evrard, J. P. Declercq, T. Debaerdemaeker, and H. König. *The first successful crystallization of a prokaryotic extremely thermophilic outer surface layer glycoprotein*. Vol. 214. 8. 1999, pp. 427–429. DOI: 10.1524/zkri.1999.214.8.427.

[74] Michael D. Jones, Anson C.K. Chan, John F. Nomellini, Michael E.P. Murphy, and John Smit. "Surface-layer protein from Caulobacter crescentus: Expression, purification and X-ray crystallographic analysis". In: *Acta Crystallographica Section:F Structural Biology Communications* 72 (2016), pp. 677–680. ISSN: 2053230X. DOI: 10.1107/S2053230X16011638.

[75] Jonathan Herrmann, Po Nan Li, Fatemeh Jabbarpour, Anson C.K. Chan, Ivan Rajkovic, Tsutomu Matsui, Lucy Shapiro, John Smit, Thomas M. Weiss, Michael E.P. Murphy, and Soichi Wakatsuki. "A bacterial surface layer protein exploits multistep crystallization for rapid self-assembly". In: *Proceedings of the National Academy of Sciences of the United States of America* 117.1 (2020), pp. 388–394. ISSN: 10916490. DOI: 10.1073/pnas.1909798116.

[76] Andreas Breitwieser, Jagoba Iturri, Jose-Luis Toca-Herrera, Uwe Sleytr, and Dietmar Pum. "In Vitro Characterization of the Two-Stage Non-Classical Reassembly Pathway of S-Layers". In: *International Journal of Molecular Sciences* 18.2 (2017), p. 400. ISSN: 1422-0067. DOI: 10.3390/ijms18020400. URL: http://www.mdpi.com/1422-0067/18/2/400.

[77] Behzad Rad, Thomas K. Haxton, Albert Shon, Seong Ho Shin, Stephen Whitelam, and Caroline M. Ajo-Franklin. "Ion-specific control of the self-assembly dynamics of a nanostructured protein lattice". In: *ACS Nano* 9.1 (2015), pp. 180–190. ISSN: 1936086X. DOI: 10.1021/nn502992x.

[78] Seong-Ho Shin, Sungwook Chung, Babak Sanii, Luis R Comolli, Carolyn R Bertozzi, and James J De Yoreo. "Direct observation of kinetic traps associated with structural transformations leading to multiple pathways of S-layer

assembly." In: *Proceedings of the National Academy of Sciences of the United States of America* 109.32 (2012), pp. 12968–73. ISSN: 1091-6490. DOI: 10.1073/pnas.1201504109. URL: http://www.pnas.org/cgi/content/long/109/32/12968.

[79] Sungwook Chung, Seong Ho Shin, Carolyn R. Bertozzi, and James J. De Yoreo. "Self-catalyzed growth of S layers via an amorphous-to-crystalline transition limited by folding kinetics". In: *Proceedings of the National Academy of Sciences of the United States of America* 107.38 (2010), pp. 16536–16541. ISSN: 10916490. DOI: 10.1073/pnas.1008280107.

[80] Leonardo Maestri Teixeira, Aaron Strickland, Sonny S. Mark, Magnus Bergkvist, Yajaira Sierra-Sastre, and Carl A. Batt. "Entropically driven self-assembly of Lysinibacillus sphaericus S-layer proteins analyzed under various environmental conditions". In: *Macromolecular Bioscience* 10.2 (2010), pp. 147–155. ISSN: 16165187. DOI: 10.1002/mabi.200900175.

[81] Manfred Bobeth, Andreas Blecha, Anja Blüher, Michael Mertig, Nuriye Korkmaz, Kai Ostermann, Gerhard Rödel, and Wolfgang Pompe. "Formation of tubes during self-assembly of bacterial surface layers". In: *Langmuir* 27.24 (2011), pp. 15102–15111. ISSN: 07437463. DOI: 10.1021/la203430q.

[82] Aitziber Eleta Lopez, Susana Moreno-Flores, Dietmar Pum, Uwe B. Sleytr, and José L. Toca-Herrera. "Surface dependence of protein nanocrystal formation". In: *Small* 6.3 (2010), pp. 396–403. ISSN: 16136810. DOI: 10.1002/smll.200901169.

[83] Judith Ferner-Ortner-Bleckmann, Nicola Gelbmann, Manfred Tesarz, Eva M. Egelseer, and Uwe B. Sleytr. "Surface-layer lattices as patterning element for multimeric extremozymes". In: *Small* 9.22 (2013), pp. 3887–3894. ISSN: 16136810. DOI: 10.1002/smll.201201014.

[84] Judith Ferner-Ortner-Bleckmann, Angelika Schrems, Nicola Ilk, Eva M. Egelseer, Uwe B. Sleytr, and Bernhard Schuster. " Multitechnique study on a recombinantly produced Bacillus halodurans laccase and an S-layer/laccase fusion protein ". In: *Biointerphases* 6.2 (2011), pp. 63–72. ISSN: 1934-8630. DOI: 10.1116/1.3589284.

[85] Fusion With. "Evaluation of a New System for Developing Particulate Enzymes Based on the Surface (S) -Layer Protein (RsaA) of Caulobacter crescentus". In: 127 (2005).

[86] a Breitwieser, E M Egelseer, D Moll, N Ilk, C Hotzy, B Bohle, C Ebner, U B Sleytr, and M Sara. "A recombinant bacterial cell surface (S-layer)-major birch pollen allergen-fusion protein (rSbsC/Bet v1) maintains the ability to self-assemble into regularly structured monomolecular lattices and the functionality of the allergen". In: *Protein Eng* 15.3 (2002), pp. 243–249. ISSN: 0269-2139. URL: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve{\&}db=PubMed{\&}dopt=Citation{\&}list{_}uids=11932495.

[87] A. Breitwieser, S. Küpcü, S. Howorka, S. Weigert, C. Langer, K. Hoffmann-Sommergruber, O. Scheiner, U. B. Sleytr, and M. Sára. "2-D protein crystals as an immobilization matrix for producing reaction zones in dipstick-style immunoassays". In: *BioTechniques* 21.5 (1996), pp. 918–925. ISSN: 07366205. DOI: 10.2144/96215rr05.

[88] Carina Huber, Eva M. Egelseer, Nicola Ilk, Uwe B. Sleytr, and Margit Sára. "S-layer-streptavidin fusion proteins and S-layer-specific heteropolysaccharides as part of a biomolecular construction kit for application in nanobiotechnology". In: *Microelectronic Engineering* 83.4-9 SPEC. ISS. (2006), pp. 1589–1593. ISSN: 01679317. DOI: 10.1016/j.mee.2006.01.109.

[89] Carina Huber, Jing Liu, Eva M. Egelseer, Dieter Moll, Wolfgang Knoll, Uwe B. Sleytr, and Margit Sára. "Heterotetramers formed by an S-layer-streptavidin fusion protein and core-streptavidin as a nanoarrayed template for biochip development". In: *Small* 2.1 (2006), pp. 142–150. ISSN: 16136810. DOI: 10.1002/smll.200500147.

[90] Yen Ting Lai, Neil P. King, and Todd O. Yeates. "Principles for designing ordered protein assemblies". In: *Trends in Cell Biology* 22.12 (2012), pp. 653–661. ISSN: 09628924. DOI: 10.1016/j.tcb.2012.08.004. URL: http://dx.doi.org/10.1016/j.tcb.2012.08.004.

[91] Jennifer E. Padilla, Christos Colovos, and Todd O. Yeates. "Nanohedra: Using symmetry to design self assembling protein cages, layers, crystals, and filaments". In: *Proceedings of the National Academy of Sciences of the United States of America* 98.5 (2001), pp. 2217–2221. ISSN: 00278424. DOI: 10.1073/pnas.041614998.

[92] Alexander Thomas, James F. Matthaei, and François Baneyx. "A Self-Assembling Two-Dimensional Protein Array is a Versatile Platform for the Assembly of Multicomponent Nanostructures". In: *Biotechnology Journal* 13.12 (2018), pp. 1–7. ISSN: 18607314. DOI: 10.1002/biot.201800141.

[93] Rongying Liu, Zdravko Kochovski, Long Li, Yue wen Yin, Jing Yang, Guang Yang, Guoqing Tao, Anqiu Xu, Ensong Zhang, Hong ming Ding, Yan Lu, Guosong Chen, and Ming Jiang. "Fabrication of Pascal-triangle Lattice of Proteins by Inducing Ligand Strategy". In: *Angewandte Chemie - International Edition* 59.24 (2020), pp. 9617–9623. ISSN: 15213773. DOI: 10.1002/anie.202000771.

[94] Rachel A. Baarda, Tegan L. Marianchuk, Michael D. Toney, and Daniel L. Cox. "In silico stress-strain measurements on self-assembled protein lattices". In: *Soft Matter* 14.40 (2018), pp. 8095–8104. ISSN: 17446848. DOI: 10.1039/C8SM00412A.

[95] Fiona M. Roche, Mary Meehan, and Timothy J. Foster. "The Staphylococcus aureus surface protein SasG and its homologues promote bacterial adherence to human desquamated nasal epithelial cells". In: *Microbiology* 149.10 (2003), pp. 2759–2767. ISSN: 13500872. DOI: 10.1099/mic.0.26412-0.

[96] James E. Cassat, Paul M. Dunman, Fionnuala McAleese, Ellen Murphy, Steven J. Projan, and Mark S. Smeltzer. "Comparative genomics of Staphylococcus aureus musculoskeletal isolates". In: *Journal of Bacteriology* 187.2 (2005), pp. 576–592. ISSN: 00219193. DOI: 10.1128/JB.187.2.576-592.2005.

[97] Michael Otto. "Staphylococcal biofilms. In: Romeo, T. (ed.) Bacterial biofilms (current topics in microbiology and immunology)". In: *Springer-Verlag* (2008), pp. 207–228.

[98] Junni Tang, Juan Chen, Ji Liu, Rong Zhang, Rongsheng Yang, and Lianghong Chen. "Effects of different cultivation conditions on staphylococcus aureus biofilm formation and diversity of adhesin genes". In: *Journal of Food Safety* 32.2 (2012), pp. 210–218. ISSN: 01496085. DOI: 10.1111/j.1745-4565.2012.00370.x.

[99] Keli Cristine Reiter, Bárbara Villa, Thiago Galvão da Silva Paim, Gustavo Enck Sambrano, Caio Fernando de Oliveira, and Pedro Alves D'Azevedo. "Enhancement of antistaphylococcal activities of six antimicrobials against sasG-negative methicillin-susceptible Staphylococcus aureus: An in vitro biofilm model". In: *Diagnostic Microbiology and Infectious Disease* 74.2 (2012), pp. 101–105. ISSN: 07328893. DOI: 10.1016/j.diagmicrobio.2012.05.034. URL: http://dx.doi.org/10.1016/j.diagmicrobio.2012.05.034.

[100] Keigo Yonemoto, Akio Chiba, Shinya Sugimoto, Chikara Sato, Mitsuru Saito, Yuki Kinjo, Keishi Marumo, and Yoshimitsu Mizunoe. "Redundant and Distinct Roles of Secreted Protein Eap and Cell Wall-Anchored Protein SasG in Biofilm Formation and Pathogenicity of Staphylococcus aureus". In: *Infection and Immunity* 87.4 (2019). Ed. by Nancy E. Freitag, pp. 1–15. ISSN: 0019-9567. DOI: 10.1128/IAI.00894-18. URL: http://iai.asm.org/lookup/doi/10.1128/IAI.00894-18.

[101] Bernhard Krismer and Andreas Peschel. "Does Staphylococcus aureus nasal colonization involve biofilm formation?" In: *Future Microbiology* 6.5 (2011), pp. 489–493. ISSN: 1746-0913. DOI: 10.2217/fmb.11.37. URL: https://www.futuremedicine.com/doi/10.2217/fmb.11.37.

[102] Rachid Achek, Helmut Hotzel, Ibrahim Nabi, Souad Kechida, Djamila Mami, Nassima Didouh, Herbert Tomaso, Heinrich Neubauer, Ralf Ehricht, Stefan Monecke, and Hosny El-adawy. "Phenotypic and molecular detection of biofilm formation in staphylococcus aureus isolated from different sources in Algeria". In: *Pathogens* 9.2 (2020). ISSN: 20760817. DOI: 10.3390/pathogens9020153.

[103] Fiona M. Roche, Ruth Massey, Sharon J. Peacock, Nicholas P.J. Day, Livia Visai, Pietro Speziale, Alex Lam, Mark Pallen, and Timothy J. Foster. "Characterization of novel LPXTG-containing proteins of Staphylococcus aureus

identified from genome sequences". In: *Microbiology* 149.3 (2003), pp. 643–654. ISSN: 13500872. DOI: 10.1099/mic.0.25996-0.

[104] Rebecca M. Corrigan, David Rigby, Pauline Handley, and Timothy J. Foster. "The role of Staphylococcus aureus surface protein SasG in adherence and biofilm formation". In: *Microbiology* 153.8 (2007), pp. 2435–2446. ISSN: 13500872. DOI: 10.1099/mic.0.2007/006676-0.

[105] Deborah G. Conrady, Cristin C. Brescia, Katsunori Horii, Alison A. Weiss, Daniel J. Hassett, and Andrew B. Herr. "A zinc-dependent adhesion module is responsible for intercellular adhesion in staphylococcal biofilms". In: *Proceedings of the National Academy of Sciences of the United States of America* 105.49 (2008), pp. 19456–19461. ISSN: 00278424. DOI: 10.1073/pnas.0807717105.

[106] Dominika T Gruszka, Justyna A Wojdyla, Richard J Bingham, Johan P Turkenburg, Iain W Manfield, Annette Steward, Andrew P Leech, Joan A Geoghegan, Timothy J Foster, Jane Clarke, et al. "Staphylococcal biofilm-forming protein has a contiguous rod-like structure". In: *Proceedings of the National Academy of Sciences* 109.17 (2012), E1011–E1018.

[107] Dominika T. Gruszka, Fiona Whelan, Oliver E. Farrance, Herman K.H. Fung, Emanuele Paci, Cy M. Jeffries, Dmitri I. Svergun, Clair Baldock, Christoph G. Baumann, David J. Brockwell, Jennifer R. Potts, and Jane Clarke. "Cooperative folding of intrinsically disordered domains drives assembly of a strong elongated protein". In: *Nature Communications* 6 (2015), pp. 1–9. ISSN: 20411723. DOI: 10.1038/ncomms8271.

[108] Cécile Formosa-Dague, Pietro Speziale, Timothy J. Foster, Joan A. Geoghegan, and Yves F. Dufrêne. "Zinc-dependent mechanical properties of Staphylococcus aureus biofilm-forming surface protein SasG". In: *Proceedings of the National Academy of Sciences of the United States of America* 113.2 (2016), pp. 410–415. ISSN: 10916490. DOI: 10.1073/pnas.1519265113.

[109] Dominika T. Gruszka, Carolina A.T.F. Mendonça, Emanuele Paci, Fiona Whelan, Judith Hawkhead, Jennifer R. Potts, and Jane Clarke. "Disorder drives cooperative folding in a multidomain protein". In: *Proceedings of the National*

Academy of Sciences of the United States of America 113.42 (2016), pp. 11841–11846. ISSN: 10916490. DOI: 10.1073/pnas.1608762113.

[110] Derek N Woolfson. "Coiled-Coil Design: Updated and Upgraded". In: *Fibrous Proteins: Structures and Mechanisms, Subcellular Biochemistry*. Vol. 82. 2017, pp. 35–61. ISBN: 978-3-319-49672-6. DOI: 10.1007/978-3-319-49674-0. URL: http://link.springer.com/10.1007/978-3-319-49674-0.

[111] Efrosini Moutevelis and Derek N. Woolfson. "A Periodic Table of Coiled-Coil Protein Structures". In: *Journal of Molecular Biology* 385.3 (2009), pp. 726–732. ISSN: 00222836. DOI: 10.1016/j.jmb.2008.11.028. URL: http://dx.doi.org/10.1016/j.jmb.2008.11.028.

[112] FHS Crick. "Is a-Keratin a Coiled Coil?" In: *Nature* 4334 (1952), pp. 882–883.

[113] FHS Crick. "Packing of alpha-helices: Simple Coiled-Coils". In: *Acta Crystallographica* 6 (1953), pp. 986 –997.

[114] FHS Crick. "The Fourier Transform of a Coiled-Coil". In: *Acta Crystallographica* 2.6 (1953), p. 685. DOI: 10.1107/S0365110X53001952.

[115] Ivan V Korendovych and William F. DeGreado. "De novo protein design , a retrospective". In: *Quarterly Reviews of Biophysics* 53.e3 (2019), pp. 1–33.

[116] Elizabeth H.C. Bromley, Richard B. Sessions, Andrew R. Thomson, and Derek N. Woolfson. "Designed α-helical tectons for constructing multicomponent synthetic biological systems". In: *Journal of the American Chemical Society* 131.3 (2009), pp. 928–930. ISSN: 00027863. DOI: 10.1021/ja804231a.

[117] A. D. McLachlan and Murray Stewart. "Tropomyosin coiled-coil interactions: Evidence for an unstaggered structure". In: *Journal of Molecular Biology* 98.2 (1975), pp. 293–304. ISSN: 00222836. DOI: 10.1016/S0022-2836(75)80119-7.

[118] Linda Truebestein and Thomas A Leonard. "Coiled-coils: The long and short of it". In: *BioEssays* 38 (2016), pp. 903–916. DOI: 10.1002/bies.201600062.

[119] Erin K. O'Shea, Juli D. Klemm, Peter S. Kim, and Tom Alber. "X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil". In: *Science* 254.5031 (1991), pp. 539–544. ISSN: 00368075. DOI: 10.1126/science.1948029.

[120] K O Erin, Kevin J Lumbl, and Peter S Kimt. "Peptide ' Velcro *' design of a heterodimeric coiled coil". In: *Current Biology* 3.10 (1993), pp. 658–667.

[121] Franziska Thomas, Ai Niitsu, Alain Oregioni, Gail J Bartlett, and Derek N Woolfson. "Conformational Dynamics of Asparagine at Coiled-Coil Interfaces". In: *Biochemistry* 56 (2017), pp. 6544–6554. DOI: 10.1021/acs.biochem.7b00848.

[122] Hans Wendt, Lukas Leder, Harri Ha¨rma, Ilian Jelesarov, Antonio Baici, and Hans Rudolf Bosshard. "Very Rapid , Ionic Strength-Dependent Association and Folding of a Heterodimeric". In: *Biochemistry* 2960.96 (1997), pp. 204–213. DOI: 10.1021/bi961672y.

[123] Peter Burkhard, Jörg Stetefeld, Sergei V Strelkov, and Peter Burkhard. "Coiled coils : a highly versatile protein folding motif". In: *Trends in Cell Biology* 11.2 (2001), pp. 82–88.

[124] RICHARD A. Kammerer, THERESE Schulthess, RUTH Landwehr, ARIEL Lustig, JU¨RGEN Enger, UELI Aebi, and MICHEL O. Steinmetz. "An autonomous folding unit mediates the assembly of two-stranded coiled coils". In: *Proceedings of the National Academy of Sciences* 95 (1998), pp. 13419–13424.

[125] Richard A Kammerer, Victor A Jaravine, Sabine Frank, Therese Schulthess, Ruth Landwehr, Ariel Lustig, Carlos Garcı, and Michel O Steinmetz. "An Intrahelical Salt Bridge within the Trigger Site Stabilizes the GCN4 Leucine Zipper". In: *The Journal of biological chemistry* 276.17 (2001), pp. 13685–13688. DOI: 10.1074/jbc.M010492200.

[126] Barbara Ciani, Saša Bjelić, Srinivas Honnappa, Hatim Jawhari, Rolf Jaussi, Aishwarya Payapilly, Thomas Jowitt, Michel O. Steinmetz, and Richard A. Kammerer. "Molecular basis of coiled-coil oligomerization-state specificity". In: *Proceedings of the National Academy of Sciences of the United States of America* 107.46 (2010), pp. 19850–19855. ISSN: 10916490. DOI: 10.1073/pnas.1008502107.

[127] Michel O Steinmetz, Ilian Jelesarov, William M Matousek, Srinivas Honnappa, Wolfgang Jahnke, John H Missimer, Sabine Frank, Andrei T Alexandrescu, and Richard A Kammerer. "Molecular basis of coiled-coil formation". In: *PNAS* 104.17 (2007), pp. 7062–7067.

[128] Igor Drobnak, Helena Gradis, Ajasja Ljubetic, Estera Merljak, and Roman Jerala. "Modulation of Coiled-Coil Dimer Stability through Surface Residues while Preserving Pairing Specificity". In: *Journal of the American Chemical Society* 139 (2017), pp. 8229–8236. DOI: 10.1021/jacs.7b01690.

[129] RB Sutton, D Fasshauer, R Jahn, and AT Brunger. "Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 A resolution". In: *Nature* 395.September (1998), pp. 347–53.

[130] Alison K Gillingham and Sean Munro. "Finding the Golgi : Golgin Coiled-Coil Proteins Show the Way". In: *Trends in Cell Biology* 26.6 (2016), pp. 399–408. ISSN: 0962-8924. DOI: 10.1016/j.tcb.2016.02.005. URL: http://dx.doi.org/10.1016/j.tcb.2016.02.005.

[131] J. C. Hu. "A guided tour in protein interaction space: Coiled coils from the yeast proteome". In: *Proceedings of the National Academy of Sciences of the United States of America* 97.24 (2000), pp. 12935–12936. ISSN: 00278424. DOI: 10.1073/pnas.97.24.12935.

[132] John J . Shehel and Don C. Wiley. "RECEPTOR BINDING AND MEMBRANE FUSION IN VIRUS ENTRY: The Influenza Hemagglutinin". In: *Annual Reviews in Biochemistry* (2000), pp. 777–810. ISSN: 0022-538X. DOI: 10.1128/JVI.06147-11. URL: http://arjournals.annualreviews.org/doi/abs/10.1146/annurev.biochem.68.1.863{\%}5Cnpapers3://publication/uuid/37777FFE-E284-4B0C-9648-9DF205012ADD.

[133] Alex Buchberger, Chad R Simmons, Nour Eddine Fahmi, Ronit Freeman, and Nicholas Stephanopoulos. "Hierarchical Assembly of Nucleic Acid/Coiled-Coil Peptide Nanostructures". In: *Journal of the American Chemical Society* 142 (2020), pp. 1406–1416. DOI: 10.1021/jacs.9b11158.

[134] Dominic J Glover, Samuel Lim, Dawei Xu, Nancy B Sloan, Ye Zhang, and Douglas S Clark. "Assembly of Multicomponent Protein Filaments Using Engineered Subunit Interfaces". In: *ACS Synthetic Biology* 7 (2018), pp. 2447–2456. DOI: 10.1021/acssynbio.8b00241. URL: https://pubs.acs.org/sharingguidelines.

[135] Caitlin L Edgell, Abigail J Smith, Joseph L Beesley, Nigel J Savery, and Derek N Woolfson. "De Novo Designed Protein-Interaction Modules for In-Cell Applications". In: *ACS Synthetic Biology* 9 (2020), pp. 427–436. DOI: 10.1021/acssynbio.9b00453.

[136] Carolin Mueller and Tom N Grossmann. "Coiled-Coil Peptide Beacon : A Tunable Conformational Switch for Protein Detection Communications Angewandte". In: *Angewandte Chemie - International Edition* 57 (2018), pp. 17079–17083. DOI: 10.1002/anie.201811515.

[137] Tina Lebar, Duško Lainšček, Estera Merljak, Jana Aupič, and Roman Jerala. "A tunable orthogonal coiled-coil interaction toolbox for engineering mammalian cells". In: *Nature Chemical Biology* (2020). DOI: 10.1038/s41589-019-0443-y.

[138] Jian Yang, Yasuhito Shimada, Rene C L Olsthoorn, B Ewa Snaar-jagalska, and Herman P Spaink. "Application of Coiled Coil Peptides in Anticancer Drug Delivery Using a Zebrafish Xenograft Model Liposomal". In: *ACS Nano* 10 (2016), 74287435 Article. DOI: 10.1021/acsnano.6b01410.

[139] Ajitha S. Cristie-David, Aaron Sciore, Somayesadat Badieyan, Joseph D. Eschweiler, Philipp Koldewey, James C. A. Bardwell, Brandon T. Ruotoloa, and E. Neil G. Marsh. "Evaluation of de novo-designed coiled coils as off-the-shelf components for protein assembly". In: *Molecular Systems Design & Engineering* 2 (2017), pp. 140–148. DOI: 10.1039/c7me00012j.

[140] Duncan J.E. Murdock. "The 'biomineralization toolkit' and the origin of animal skeletons". In: *Biological Reviews* 95.5 (2020), pp. 1372–1392. ISSN: 1469185X. DOI: 10.1111/brv.12614.

[141] Xiaorui Song, Zhaoqun Liu, Lingling Wang, and Linsheng Song. "Recent advances of shell matrix proteins and cellular orchestration in marine molluscan shell biomineralization". In: *Frontiers in Marine Science* 6.FEB (2019). ISSN: 22967745. DOI: 10.3389/fmars.2019.00041.

[142] Peter L. Davies. "Ice-binding proteins: A remarkable diversity of structures for stopping and starting ice growth". In: *Trends in Biochemical Sciences* 39.11

(2014), pp. 548–555. ISSN: 13624326. DOI: 10.1016/j.tibs.2014.09.005. URL: http://dx.doi.org/10.1016/j.tibs.2014.09.005.

[143] Maya Bar Dolev, Ido Braslavsky, and Peter L. Davies. "Ice-Binding Proteins and Their Function". In: *Annual Review of Biochemistry* 85 (2016), pp. 515–542. ISSN: 15454509. DOI: 10.1146/annurev-biochem-060815-014546.

[144] James C. Weaver, Qianqian Wang, Ali Miserez, Anthony Tantuccio, Ryan Stromberg, Krassimir N. Bozhilov, Peter Maxwell, Richard Nay, Shinobu T. Heier, Elaine DiMasi, and David Kisailus. "Analysis of an ultra hard magnetic biomineral in chiton radular teeth". In: *Materials Today* 13.1-2 (2010), pp. 42–52. ISSN: 13697021. DOI: 10.1016/S1369-7021(10)70016-X. URL: http://dx.doi.org/10.1016/S1369-7021(10)70016-X.

[145] Benjamin Carlos Bostick. "Massive ore deposits from microscopic organisms". In: *Geology* 47.2 (2019), pp. 191–192. ISSN: 19432682. DOI: 10.1130/focus022019.1.

[146] Navdeep K. Dhami, M. Sudhakara Reddy, and M. Sudhakara Mukherjee. "Biomineralization of calcium carbonates and their engineered applications: A review". In: *Frontiers in Microbiology* 4.OCT (2013), pp. 1–13. ISSN: 1664302X. DOI: 10.3389/fmicb.2013.00314.

[147] Varenyam Achal, Abhijit Mukherjee, Deepika Kumari, and Qiuzhuo Zhang. "Biomineralization for sustainable construction - A review of processes and applications". In: *Earth-Science Reviews* 148 (2015), pp. 1–17. ISSN: 00128252. DOI: 10.1016/j.earscirev.2015.05.008. URL: http://dx.doi.org/10.1016/j.earscirev.2015.05.008.

[148] RICHARD Blakemore. "Magnetotactic Bacteria". In: *Science* 190.October (1975), pp. 1–4.

[149] Lei Yan, Shuang Zhang, Peng Chen, Hetao Liu, Huanhuan Yin, and Hongyu Li. "Magnetotactic bacteria, magnetosomes and their application". In: *Microbiological Research* 167.9 (2012), pp. 507–519. ISSN: 09445013. DOI: 10.1016/j.micres.2012.04.002. URL: http://dx.doi.org/10.1016/j.micres.2012.04.002.

[150] René Uebe and Dirk Schüler. "Magnetosome biogenesis in magnetotactic bacteria". In: *Nature Reviews Microbiology* 14.10 (2016), pp. 621–637. ISSN: 17401534. DOI: 10.1038/nrmicro.2016.99.

[151] Shiran Barber-Zucker and Raz Zarivach. "A look into the biochemistry of Magnetosome biosynthesis in magnetotactic bacteria". In: *ACS Chemical Biology* 12.1 (2017), pp. 13–22. ISSN: 15548937. DOI: 10.1021/acschembio.6b01000.

[152] Hila Nudelman and Raz Zarivach. "Structure prediction of magnetosome-associated proteins". In: *Frontiers in Microbiology* 5.JAN (2014), pp. 1–17. ISSN: 1664302X. DOI: 10.3389/fmicb.2014.00009.

[153] Sarah Staniland, Wyn Williams, Neil Telling, Gerrit Van Der Laan, Andrew Harrison, and Bruce Ward. "Controlled cobalt doping of magnetosomes in vivo". In: *Nature Nanotechnology* 3.3 (2008), pp. 158–162. ISSN: 17483395. DOI: 10.1038/nnano.2008.35.

[154] Lourdes Marcano, David Munoz, Rosa Martín-Rodríguez, Inaki Orue, Javier Alonso, Ana García-Prieto, Aida Serrano, Sergio Valencia, Radu Abrudan, Luis Fernández Barquín, Alfredo García-Arribas, Alicia Muela, and M. Luisa Fdez-Gubieda. "Magnetic Study of Co-Doped Magnetosome Chains". In: *Journal of Physical Chemistry C* 122.13 (2018), pp. 7541–7550. ISSN: 19327455. DOI: 10.1021/acs.jpcc.8b01187.

[155] Lourdes Marcano, Iñaki Orue, Ana Garcĺa-Prieto, Radu Abrudan, Javier Alonso, Luis Fernández Barquĺn, Sergio Valencia, Alicia Muela, and M. Luisa Fdez-Gubieda. "Controlled Magnetic Anisotropy in Single Domain Mn-doped Biosynthesized Nanoparticles". In: *Journal of Physical Chemistry C* 124.41 (2020), pp. 22827–22838. ISSN: 19327455. DOI: 10.1021/acs.jpcc.0c07018.

[156] Safia Anjum, Rabia Tufail, Hafsa Saleem, Rehana Zia, and Sira Riaz. "Investigation of Stability and Magnetic Properties of Ni- and Co-Doped Iron Oxide Nano-particles". In: *Journal of Superconductivity and Novel Magnetism* 30.8 (2017), pp. 2291–2301. ISSN: 15571947. DOI: 10.1007/s10948-017-4044-2.

[157] Shiran Barber-Zucker, Noa Keren-Khadmy, and Raz Zarivach. "From invagination to navigation: The story of magnetosome-associated proteins in magnetotactic bacteria". In: *Protein Science* 25 (2016), pp. 338–351. ISSN: 1469896X. DOI: 10.1002/pro.2827.

[158] Dorothée Murat, Veesta Falahati, Luca Bertinetti, Roseann Csencsits, André Körnig, Kenneth Downing, Damien Faivre, and Arash Komeili. "The magnetosome membrane protein, MmsF, is a major regulator of magnetite biomineralization in Magnetospirillum magneticum AMB-1". In: *Molecular Microbiology* 85.4 (2012), pp. 684–699. ISSN: 0950382X. DOI: 10.1111/j.1365-2958.2012.08132.x.

[159] Andrea E. Rawlings, Jonathan P. Bramble, Robyn Walker, Jennifer Bain, Johanna M. Galloway, and Sarah S. Staniland. "Self-assembled MmsF proteinosomes control magnetite nanoparticle formation in vitro". In: *Proceedings of the National Academy of Sciences of the United States of America* 111.45 (2014), pp. 19094–19099. ISSN: 10916490. DOI: 10.1073/pnas.1409256111.

[160] A Arakaki, J Webb, and T Matsunaga. "A novel protein tightly bound to bacterial magnetic particles in Magnetospirillum magneticum strain AMB-1". In: *JOURNAL OF BIOLOGICAL CHEMISTRY* 278.10 (2003), pp. 8745–8750. ISSN: 0021-9258. DOI: 10.1074/jbc.M211729200.

[161] Yosuke Amemiya, Atsushi Arakaki, Sarah S Staniland, Tsuyoshi Tanaka, and Tadashi Matsunaga. "Controlled formation of magnetite crystal by partial oxidation of ferrous hydroxide in the presence of recombinant magnetotactic bacterial protein Mms6". In: *BIOMATERIALS* 28.35 (2007), pp. 5381–5389. ISSN: 0142-9612. DOI: 10.1016/j.biomaterials.2007.07.051.

[162] Atsushi Arakaki, Fukashi Masuda, Yosuke Amemiya, Tsuyoshi Tanaka, and Tadashi Matsunaga. "Control of the morphology and size of magnetite particles with peptides mimicking the Mms6 protein from magnetotactic bacteria". In: *Journal of Colloid and Interface Science* 343.1 (2010), pp. 65–70. ISSN: 00219797. DOI: 10.1016/j.jcis.2009.11.043. URL: http://dx.doi.org/10.1016/j.jcis.2009.11.043.

[163] Jos J.M. Lenders, Cem L. Altan, Paul H.H. Bomans, Atsushi Arakaki, Seyda Bucak, Gijsbertus De With, and Nico A.J.M. Sommerdijk. "A bioinspired co-precipitation method for the controlled synthesis of magnetite nanoparticles". In: *Crystal Growth and Design* 14.11 (2014), pp. 5561–5568. ISSN: 15287505. DOI: 10.1021/cg500816z.

[164] Tanya Prozorov, Pierre Palo, Lijun Wang, Marit Nilsen-Hamilton, DeAnna Jones, Daniel Orr, Surya K Mallapragada, Balaji Narasimhan, Paul C Canfield, and Ruslan Prozorov. "Cobalt ferrite nanocrystals: Out-performing magnetotactic bacteria". In: *ACS NANO* 1.3 (2007), pp. 228–233. ISSN: 1936-0851. DOI: 10.1021/nn700194h.

[165] Johanna M. Galloway, Atsushi Arakaki, Fukashi Masuda, Tsuyoshi Tanaka, Tadashi Matsunaga, and Sarah S. Staniland. "Magnetic bacterial protein Mms6 controls morphology, crystallinity and magnetism of cobalt-doped magnetite nanoparticles in vitro". In: *Journal of Materials Chemistry* 21.39 (2011), pp. 15244–15254. ISSN: 09599428. DOI: 10.1039/c1jm12003d.

[166] Tomaž Bratkovič. *Progress in phage display: Evolution of the technique and its applications*. 2010. DOI: 10.1007/s00018-009-0192-2.

[167] Alaa A.A. Aljabali, Sachin N. Shah, Richard Evans-Gowing, George P. Lomonossoff, and David J. Evans. "Chemically-coupled-peptide-promoted virus nanoparticle templated mineralization". In: *Integrative Biology* 3.2 (2011), pp. 119–125. ISSN: 17579694. DOI: 10.1039/c0ib00056f.

[168] Aw Wei Liang Alvin, Masayoshi Tanaka, and Mina Okochi. "Characterization of particulate matter binding peptides screened from phage display". In: *Journal of Bioscience and Bioengineering* 123.5 (2017), pp. 621–624. ISSN: 13474421. DOI: 10.1016/j.jbiosc.2016.12.014. URL: http://dx.doi.org/10.1016/j.jbiosc.2016.12.014.

[169] Fei You, Guangfu Yin, Ximing Pu, Yucan Li, Yang Hu, Zhongbin Huang, Xiaoming Liao, Yadong Yao, and Xianchun Chen. "Biopanning and characterization of peptides with Fe3O4 nanoparticles-binding capability via phage

display random peptide library technique". In: *Colloids and Surfaces B: Biointerfaces* 141 (2016), pp. 537–545. ISSN: 18734367. DOI: 10.1016/j.colsurfb.2016.01.062. URL: http://dx.doi.org/10.1016/j.colsurfb.2016.01.062.

[170] Michael T. Klem, Debbie Willits, Daniel J. Solis, Angela M. Belcher, Mark Young, and Trevor Douglas. "Bio-inspired synthesis of protein-encapsulated CoPt nanoparticles". In: *Advanced Functional Materials* 15.9 (2005), pp. 1489–1494. ISSN: 1616301X. DOI: 10.1002/adfm.200400453.

[171] Christian Tiede, Anna A.S. Tang, Sarah E. Deacon, Upasana Mandal, Joanne E. Nettleship, Robin L. Owen, Suja E. George, David J. Harrison, Raymond J. Owens, Darren C. Tomlinson, and Michael J. McPherson. "Adhiron: A stable and versatile peptide display scaffold for molecular recognition applications". In: *Protein Engineering, Design and Selection* 27.5 (2014), pp. 145–155. ISSN: 17410134. DOI: 10.1093/protein/gzu007.

[172] Andrea E. Rawlings, Jonathan P. Bramble, Anna A.S. Tang, Lori A. Somner, Amy E. Monnington, David J. Cooke, Michael J. McPherson, Darren C. Tomlinson, and Sarah S. Staniland. "Phage display selected magnetite interacting Adhirons for shape controlled nanoparticle synthesis". In: *Chemical Science* 6.10 (2015), pp. 5586–5594. ISSN: 20416539. DOI: 10.1039/c5sc01472g.

[173] Silvia A. Blank-Shim, Sebastian P. Schwaminger, Monika Borkowska-Panek, Priya Anand, Peyman Yamin, Paula Fraga-García, Karin Fink, Wolfgang Wenzel, and Sonja Berensmeier. "Binding patterns of homo-peptides on bare magnetic nanoparticles: Insights into environmental dependence". In: *Scientific Reports* 7.1 (2017), pp. 1–11. ISSN: 20452322. DOI: 10.1038/s41598-017-13928-6.

[174] S. P. Schwaminger, S. A. Blank-Shim, I. Scheifele, P. Fraga-García, and S. Berensmeier. "Peptide binding to metal oxide nanoparticles". In: *Faraday Discussions* 204 (2017), pp. 233–250. ISSN: 13645498. DOI: 10.1039/c7fd00105c.

[175] Rajesh R. Naik, Sarah J. Stringer, Gunjan Agarwal, Sharon E. Jones, and Morley O. Stone. "Biomimetic synthesis and patterning of silver nanoparticles". In: *Nature Materials* 1.3 (2002), pp. 169–172. ISSN: 14761122. DOI: 10.1038/nmat758.

[176] Johanna M. Galloway, Jonathan P. Bramble, Andrea E. Rawlings, Gavin Burnell, Stephen D. Evans, and Sarah S. Staniland. "Nanomagnetic arrays formed with the biomineralization protein Mms6". In: *Journal of Nano Research* 17.February (2012), pp. 127–146. ISSN: 16625250. DOI: 10.4028/www.scientific.net/JNanoR.17.127.

[177] Scott M. Bird, Andrea E. Rawlings, Johanna M. Galloway, and Sarah S. Staniland. "Using a biomimetic membrane surface experiment to investigate the activity of the magnetite biomineralisation protein Mms6". In: *RSC Advances* 6.9 (2016), pp. 7356–7363. ISSN: 20462069. DOI: 10.1039/c5ra16469a.

[178] S. M. Bird, O. El-Zubir, A. E. Rawlings, G. J. Leggett, and S. S. Staniland. "A novel design strategy for nanoparticles on nanopatterns: Interferometric lithographic patterning of Mms6 biotemplated magnetic nanoparticles". In: *Journal of Materials Chemistry C* 4.18 (2016), pp. 3948–3955. ISSN: 20507526. DOI: 10.1039/c5tc03895b.

[179] Johanna M. Galloway, Jennifer E. Talbot, Kevin Critchley, Jim J. Miles, and Jonathan P. Bramble. "Developing Biotemplated Data Storage: Room Temperature Biomineralization of L10 CoPt Magnetic Nanoparticles". In: *Advanced Functional Materials* 25.29 (2015), pp. 4590–4600. ISSN: 16163028. DOI: 10.1002/adfm.201501090.

[180] J. M. Galloway, S. M. Bird, J. E. Talbot, P. M. Shepley, R. C. Bradley, O. El-Zubir, D. A. Allwood, G. J. Leggett, J. J. Miles, S. S. Staniland, and K. Critchley. "Nano- and micro-patterning biotemplated magnetic CoPt arrays". In: *Nanoscale* 8.22 (2016), pp. 11738–11747. ISSN: 20403372. DOI: 10.1039/c6nr03330j.

[181] Brian D. Reiss, Chuanbin Mao, Daniel J. Solis, Katherine S. Ryan, Thomas Thomson, and Angela M. Belcher. "Biological routes to metal alloy ferromagnetic nanostructures". In: *Nano Letters* 4.6 (2004), pp. 1127–1132. ISSN: 15306984. DOI: 10.1021/nl049825n.

[182] T Yan and R Lin. "Finite element modeling and modal testing of vibration characteristics of disk drives". In: *Proceedings of SPIE - The International Society for Optical Engineering* 4753 (2002).

[183] Robert Schreiber, Ibon Santiago, Arzhang Ardavan, and Andrew J. Turberfield. "Ordering Gold Nanoparticles with DNA Origami Nanoflowers". In: *ACS Nano* 10.8 (2016), pp. 7303–7306. ISSN: 1936086X. DOI: 10.1021/acsnano.6b03076.

[184] Fei Zhang, Jeanette Nangreave, Yan Liu, and Hao Yan. "Structural DNA nanotechnology: State of the art and future perspective". In: *Journal of the American Chemical Society* 136.32 (2014), pp. 11198–11211. ISSN: 15205126. DOI: 10.1021/ja505101a.

[185] Tobias J. Günther, Matthias Suhr, Johannes Raff, and Katrin Pollmann. *Immobilization of microorganisms for AFM studies in liquids*. Vol. 4. 93. 2014, pp. 51156–51164. ISBN: 9783319391960. DOI: 10.1039/c4ra03874f.

[186] Dietmar Pum and Uwe B. Sleytr. "Reassembly of S-layer proteins". In: *Nanotechnology* 25.31 (2014). ISSN: 13616528. DOI: 10.1088/0957-4484/25/31/312001.

[187] James F. Matthaei, Frank DiMaio, Jeffrey J. Richards, Lilo D. Pozzo, David Baker, and François Baneyx. "Designing Two-Dimensional Protein Arrays through Fusion of Multimers and Interface Mutations". In: *Nano Letters* 15.8 (2015), pp. 5235–5239. ISSN: 15306992. DOI: 10.1021/acs.nanolett.5b01499.

[188] Jennifer E. Padilla, Christos Colovos, and Todd O. Yeates. "Nanohedra: Using symmetry to design self assembling protein cages, layers, crystals, and filaments". In: *Proceedings of the National Academy of Sciences of the United States of America* 98.5 (2001), pp. 2217–2221. ISSN: 00278424. DOI: 10.1073/pnas.041614998.

[189] Shane Gonen, Frank DiMaio, Tamir Gonen, and David Baker. "Design of ordered two-dimensional arrays mediated by noncovalent protein-protein interfaces". In: *Science* 348.6241 (2015), pp. 1365–1368.

[190] Erin K O'Shea, Juli D Klemm, Peter S Kim, and Tom Alber. "X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil". In: *Science* 254.5031 (1991), pp. 539–544.

[191] Atsushi Arakaki, Ayana Yamagishi, Ayumi Fukuyo, Masayoshi Tanaka, and Tadashi Matsunaga. "Co-ordinated functions of Mms proteins define the surface structure of cubo-octahedral magnetite crystals in magnetotactic bacteria". In: *MOLECULAR MICROBIOLOGY* 93.3 (2014), pp. 554–567. ISSN: 0950-382X. DOI: 10.1111/mmi.12683.

[192] Andrea E Rawlings, Panah Liravi, Sybilla Corbett, Alex S Holehouse, and Sarah S Staniland. "Investigating the ferric ion binding site of magnetite biomineralisation protein Mms6". In: *PLOS ONE* 15.2 (2020). ISSN: 1932-6203. DOI: 10.1371/journal.pone.0228708.

[193] Rosie M. Jarrald, Aw W. Liang Alvin, Andrea E. Rawlings, Masayoshi Tanaka, Mina Okochi, and Sarah S. Staniland. "Systematic Screening and Deep Analysis of CoPt Binding Peptides Leads to Enhanced CoPt Nanoparticles Using Designed Peptides". In: *Bioconjugate Chemistry* 31.8 (2020), pp. 1981–1994. ISSN: 15204812. DOI: 10.1021/acs.bioconjchem.0c00348.

[194] F William Studier. "Protein production by auto-induction in high-density shaking cultures". In: *Protein expression and purification* 41.1 (2005), pp. 207–234.

[195] Ralph Rapley. "Basic Techniques in Molecular Biology". In: Oct. 2007, pp. 1–12. DOI: 10.1385/1-59259-870-6:001.

[196] Raymond J Peroutka III, Steven J Orcutt, James E Strickler, and Tauseef R Butt. "SUMO fusion technology for enhanced protein expression and purification in prokaryotes and eukaryotes". In: *Heterologous Gene Expression in E. coli* (2011), pp. 15–30.

[197] Randall K Saiki, Stephen Scharf, Fred Faloona, Kary B Mullis, Glenn T Horn, Henry A Erlich, and Norman Arnheim. "Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia". In: *Science* 230.4732 (1985), pp. 1350–1354.

[198] Julie D Thompson, Desmond G Higgins, and Toby J Gibson. "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice". In: *Nucleic acids research* 22.22 (1994), pp. 4673–4680.

[199] Elisabeth Gasteiger, Alexandre Gattiker, Christine Hoogland, Ivan Ivanyi, Ron D Appel, and Amos Bairoch. "ExPASy: the proteomics server for in-depth protein knowledge and analysis". In: *Nucleic acids research* 31.13 (2003), pp. 3784–3788.

[200] Peter J A Cock, Tiago Antao, Jeffrey T Chang, Brad A Chapman, Cymon J Cox, Andrew Dalke, Iddo Friedberg, Thomas Hamelryck, Frank Kauff, Bartek Wilczynski, and Others. "Biopython: freely available Python tools for computational molecular biology and bioinformatics". In: *Bioinformatics* 25.11 (2009), pp. 1422–1423.

[201] Joshua A Bornhorst and Joseph J Falke. "[16] Purification of proteins using polyhistidine affinity tags". In: *Methods in enzymology* 326 (2000), pp. 245–254.

[202] Kerstin Strupat. "Molecular weight determination of peptides and proteins by ESI and MALDI". In: *Methods in enzymology* 405 (2005), pp. 1–36.

[203] Ph-Herve Hirel, M J Schmitter, Philippe Dessen, Guy Fayat, and Sylvain Blanquet. "Extent of N-terminal methionine excision from Escherichia coli proteins is governed by the side-chain length of the penultimate amino acid". In: *Proceedings of the National Academy of Sciences* 86.21 (1989), pp. 8247–8251.

[204] Sharon M Kelly, Thomas J Jess, and Nicholas C Price. "How to study proteins by circular dichroism". In: *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* 1751.2 (2005), pp. 119–139.

[205] Lee Whitmore and B. A. Wallace. "Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases". In: *Biopolymers* 89.5 (2008), pp. 392–400. ISSN: 00063525. DOI: 10.1002/bip.20853.

[206] N Sreeranna and RW Woody. "Estimation of protein secondary structure from CD spectra: Comparison of CONTIN, SELCON and CDSSTR niethods with an expanded reference set". In: *Anal. Biochem* 287 (2000), pp. 252–26.

[207] Jonathan G Lees, Andrew J Miles, Frank Wien, and BA Wallace. "A reference database for circular dichroism spectroscopy covering fold and secondary structure space". In: *Bioinformatics* 22.16 (2006), pp. 1955–1962.

[208] Christina Schäffer, René Novotny, Seta Küpcü, Sonja Zayni, Andrea Scheberl, Jacqueline Friedmann, Uwe B. Sleytr, and Paul Messner. "Novel biocatalysts based on S-layer self-assembly of geobacillus stearothermophilus NRS 2004/33: A nanobiotechnological approach". In: *Small* 3.9 (2007), pp. 1549–1559. ISSN: 16136810. DOI: 10.1002/smll.200700200.

[209] Andrea E. Rawlings, Lori A. Somner, Michaela Fitzpatrick-Milton, Thomas P. Roebuck, Christopher Gwyn, Panah Liravi, Victoria Seville, Thomas J. Neal, Oleksandr O. Mykhaylyk, Stephen A. Baldwin, and Sarah S. Staniland. "Artificial coiled coil biomineralisation protein for the synthesis of magnetic nanoparticles". In: *Nature Communications* 10.1 (2019). ISSN: 20411723. DOI: 10.1038/s41467-019-10578-2. URL: http://dx.doi.org/10.1038/s41467-019-10578-2.

[210] Jörg Stetefeld, Sean A. McKenna, and Trushar R. Patel. "Dynamic light scattering: a practical guide and applications in biomedical sciences". In: *Biophysical Reviews* 8.4 (2016), pp. 409–427. ISSN: 18672469. DOI: 10.1007/s12551-016-0218-6. URL: http://dx.doi.org/10.1007/s12551-016-0218-6.

[211] Matthew C. Dixon. "Quartz crystal microbalance with dissipation monitoring: Enabling real-time characterization of biological materials and their interactions". In: *Journal of Biomolecular Techniques* 19.3 (2008), pp. 151–158. ISSN: 15240215.

[212] Jianhua Zhou, Yangyang Wang, Li Zhang, and Xuemeng Li. "Plasmonic biosensing based on non-noble-metal materials". In: *Chinese Chemical Letters* 29.1 (2018), pp. 54–60.

[213] Schrödinger, LLC. "The {PyMOL} Molecular Graphics System, Version˜1.8". 2015.

[214] Thomas D Goddard, Conrad C Huang, Elaine C Meng, Eric F Pettersen, Gregory S Couch, John H Morris, and Thomas E Ferrin. "UCSF ChimeraX: Meeting modern challenges in visualization and analysis". In: *Protein Science* 27.1 (2018), pp. 14–25. DOI: https://doi.org/10.1002/pro.3235. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/pro.3235.

[215] Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, Talapady N Bhat, Helge Weissig, Ilya N Shindyalov, and Philip E Bourne. "The protein data bank". In: *Nucleic acids research* 28.1 (2000), pp. 235–242.

[216] Jianyi Yang, Ivan Anishchenko, Hahnbeom Park, Zhenling Peng, Sergey Ovchinnikov, and David Baker. "Improved protein structure prediction using predicted interresidue orientations". In: *Proceedings of the National Academy of Sciences of the United States of America* 117.3 (2020), pp. 1496–1503. ISSN: 10916490. DOI: 10.1073/pnas.1914677117.

[217] David E. Kim, Dylan Chivian, and David Baker. "Protein structure prediction and analysis using the Robetta server". In: *Nucleic Acids Research* 32.WEB SERVER ISS. (2004), pp. 526–531. ISSN: 03051048. DOI: 10.1093/nar/gkh468.

[218] H Stanjek and W Häusler. "Basics of X-ray Diffraction". In: *Hyperfine Interactions* 154.1-4 (2004), pp. 107–119.

[219] A. L. Patterson. "The scherrer formula for X-ray particle size determination". In: *Physical Review* 56.10 (1939), pp. 978–982. ISSN: 0031899X. DOI: 10.1103/PhysRev.56.978.

[220] J Il Langford and A J C Wilson. "Scherrer after sixty years: a survey and some new results in the determination of crystallite size". In: *Journal of applied crystallography* 11.2 (1978), pp. 102–113.

[221] Andrew L Rose and T David Waite. "Chemiluminescence of luminol in the presence of iron (II) and oxygen: oxidation mechanism and implications for its analytical use". In: *Analytical chemistry* 73.24 (2001), pp. 5909–5920.

[222] Johannes Schindelin, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch, Curtis Rueden, Stephan Saalfeld, Benjamin Schmid, and Others. "Fiji: an open-source platform for biological-image analysis". In: *Nature methods* 9.7 (2012), pp. 676–682.

[223] Peter Eaton and Paul West. *Atomic force microscopy*. Oxford university press, 2010.

[224] David Nečas and Petr Klapetek. "Gwyddion: An open-source software for SPM data analysis". In: *Central European Journal of Physics* 10.1 (2012), pp. 181–188. ISSN: 18951082. DOI: 10.2478/s11534-011-0096-2.

[225] Malte Bussiek, Norbert Mücke, and Jörg Langowski. "Polylysine-coated mica can be used to observe systematic changes in the supercoiled DNA conformation by scanning force microscopy in solution." In: *Nucleic acids research* 31.22 (2003). ISSN: 13624962. DOI: 10.1093/nar/gng137.

[226] Alessandro Podestà, Marco Indrieri, Doriano Brogioli, Gerald S. Manning, Paolo Milani, Rosalinda Guerra, Laura Finzi, and David Dunlap. "Positively charged surfaces increase the flexibility of DNA". In: *Biophysical Journal* 89.4 (2005), pp. 2558–2563. ISSN: 00063495. DOI: 10.1529/biophysj.105.064667.

[227] Melanie Ohi, Ying Li, Yifan Cheng, and Thomas Walz. "Negative staining and image classification—powerful tools in modern electron microscopy". In: *Biological procedures online* 6.1 (2004), pp. 23–34.

[228] Graeme D Ruxton. "The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test". In: *Behavioral Ecology* 17.4 (2006), pp. 688–690.

[229] Cynthia K Larive and Jonathan V Sweedler. *Celebrating the 75th anniversary of the acs division of analytical chemistry: A special collection of the most highly cited analytical chemistry papers published between 1938 and 2012*. 2013.

[230] J D Hunter. "Matplotlib: A 2D graphics environment". In: *Computing in Science & Engineering* 9.3 (2007), pp. 90–95. DOI: 10.1109/MCSE.2007.55.

[231] Norma J. Greenfield. "Using circular dichroism spectra to estimate protein secondary structure". In: *Nature Protocols* 1.6 (2007), pp. 2876–2890. ISSN: 17542189. DOI: 10.1038/nprot.2006.202.

[232] Linda E Franken, Kay Grünewald, Egbert J Boekema, and Marc C A Stuart. "A Technical Introduction to Transmission Electron Microscopy for Soft-Matter: Imaging, Possibilities, Choices, and Technical Developments". In: *Small* 16.14 (2020), p. 1906198.

[233] Helena Gradišar, Sabina Božič, Tibor Doles, Damjan Vengust, and Iva Hafner-bratkovič. "Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments". In: *Nature chemical biology* 9.6 (2013), pp. 362–366. DOI: 10.1038/nchembio.1248.Design.

[234] Isabell Tunn, Alberto S. de Léon, Kerstin G. Blank, and Matthew J. Harrington. "Tuning coiled coil stability with histidine-metal coordination". In: *Nanoscale* 10 (2018), pp. 22725–22729. DOI: 10.1039/c8nr07259k.

[235] Melis Goktas, Chuanfu Luo, Ruby May A. Sullan, Ana E. Bergues-Pupo, Reinhard Lipowsky, Ana Vila Verde, and Kerstin G. Blank. "Molecular mechanics of coiled coils loaded in the shear geometry". In: *Chemical Science* 9 (2018), pp. 4610–4621. DOI: 10.1039/c8sc01037d.

[236] Patricia Lopez-Garcia, Melis Goktas, Ana E Bergues-pupo, Beate Koksch, Daniel Varon Silva, and Kerstin G Blank. "Structural determinants of coiled coil mechanics". In: *PCCP* 21 (2019), pp. 9145–9149. DOI: 10.1039/c9cp00665f.

[237] Mike C Groth, W Mathis Rink, Nils F Meyer, and Franziska Thomas. "Kinetic studies on strand displacement in de novo designed parallel heterodimeric coiled coils". In: *Chemical Science* (2018), pp. 4308–4316. DOI: 10.1039/c7sc05342h.

[238] Moo-Jin Suh, David J. Clark, Prashanth P. Parmer, Robert D. Fleischmann, Scott N. Peterson, and Pieper Rembert. "Using chemical derivatization and mass spectrometric analysis to characterize the post-translationally modified Staphylococcus aureus surface protein G". In: *Biochim Biophys Acta* 23.1 (2010), pp. 1–7. ISSN: 15378276. DOI: 10.1038/jid.2014.371. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2688697/pdf/nihms-110443.pdf.

[239] Lilia M. Iakoucheva, Amy L. Kimzey, Christophe D. Masselon, Richard D. Smith, A. Keith Dunker, and Eric J. Ackerman. "Aberrant mobility phenomena of the DNA repair protein XPA". In: *Protein Science* 10.7 (2009), pp. 1353–1362. DOI: 10.1110/ps.ps.40101.

[240] Atis Chakrabarti. "Separation of monoclonal antibodies by analytical size exclusion chromatography". In: *Antibody Engineering*. IntechOpen, 2018.

[241] Norma J. Greenfield. "Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions". In: *Nature Protocols* 1.6 (2007), pp. 2527–2535. ISSN: 17542189. DOI: 10.1038/nprot.2006.204.

[242] Norma J Greenfield. "Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions". In: *Nature protocols* 1.6 (2006), pp. 2527–2535.

[243] Bojana Apostolovic and Harm Anton Klok. "pH-sensitivity of the E3/K3 heterodimeric coiled coil". In: *Biomacromolecules* 9.11 (2008), pp. 3173–3180. ISSN: 15257797. DOI: 10.1021/bm800746e.

[244] Rahi M. Reja, Mohsina Khan, Sumeet K. Singh, Rajkumar Misra, Anjali Shiras, and Hosahudya N. Gopi. "PH sensitive coiled coils: A strategy for enhanced liposomal drug delivery". In: *Nanoscale* 8.9 (2016), pp. 5139–5145. ISSN: 20403372. DOI: 10.1039/c5nr07734f.

[245] Robert Lizatovic, Oskar Aurelius, Mikael Akke, Derek T Logan, Ingemar Andre, Oskar Aurelius, Olof Stenstro, and Ingemar Andre. "A De Novo Designed Coiled-Coil Peptide with a Reversible pH-Induced Oligomerization Switch". In: *Structure* 24 (2016), pp. 946–955. DOI: 10.1016/j.str.2016.03.027.

[246] Naoki Nishikawa, Yoshiaki Tanizawa, Shoichi Tanaka, Yasunobu Horiguchi, Hitomi Matsuno, and Asakura Tetsuo. "pH Dependence of the Coiled-Coil Structure of Keratin Intermediate Filament in Human Hair by 13 C NMR Spectroscopy and the Mechanism of Its Disruption". In: *Polymer Journal* 30.2 (1998), pp. 125–132.

[247] Simon J. White, Steven D. Johnson, Mark A. Sellick, Agnieszka Bronowska, Peter G. Stockley, and Christoph Wälti. "The influence of two-dimensional organization on peptide conformation". In: *Angewandte Chemie - International Edition* 54.3 (2015), pp. 974–978. ISSN: 15213773. DOI: 10.1002/anie.201408971.

[248] Helena Gradišar and Roman Jerala. "De novo design of orthogonal peptide pairs forming parallel coiled-coil heterodimers". In: *Journal of Peptide Science* 17.2 (2011), pp. 100–106. ISSN: 10752617. DOI: 10.1002/psc.1331.

[249] Franziska Thomas, Aimee L. Boyle, Antony J. Burton, and Derek N. Woolfson. "A set of de novo designed parallel heterodimeric coiled coils with quantified dissociation constants in the micromolar to sub-nanomolar regime". In: *Journal of the American Chemical Society* 135.13 (2013), pp. 5161–5166. ISSN: 00027863. DOI: 10.1021/ja312310g.

[250] Mark E Zweifel and Doug Barrick. "Studies of the ankyrin repeats of the Drosophila melanogaster Notch receptor. 1. Solution conformational and hydrodynamic properties". In: *Biochemistry* 40.48 (2001), pp. 14344–14356.

[251] Matthew J Gage, Brian G Lefebvre, and Anne S Robinson. "Determinants of protein folding and aggregation in P22 tailspike protein". In: *Misbehaving Proteins*. Springer, 2006, pp. 247–264.

[252] Shibdas Banerjee and Shyamalava Mazumdar. "Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte". In: *International journal of analytical chemistry* 2012 (2012).

[253] M Ozboyaci, D B Kokh, and R C Wade. "Three steps to gold: mechanism of protein adsorption revealed by Brownian and molecular dynamics simulations". In: *Physical Chemistry Chemical Physics* 18.15 (2016), pp. 10191–10200.

[254] Jorge A. Fallas, George Ueda, William Sheffler, Vanessa Nguyen, Dan E. McNamara, Banumathi Sankaran, Jose Henrique Pereira, Fabio Parmeggiani, T. J. Brunette, Duilio Cascio, Todd R. Yeates, Peter Zwart, and David Baker. "Computational design of self-assembling cyclic protein homo-oligomers". In: *Nature Chemistry* 9.4 (2017), pp. 353–360. ISSN: 17554349. DOI: 10.1038/nchem.2673.

[255] Wayne Shenton, Dietmar Pum, Uwe B. Sleytr, and Stephen Mann. "Synthesis of cadmium sulphide superlattices using self-assembled bacterial S-layers". In: *Nature* 389.6651 (1997), pp. 585–587. ISSN: 00280836. DOI: 10.1038/39287.

[256] Dieter Moll, Carina Huber, Birgit Schlegel, Dietmar Pum, Uwe B. Sleytr, and Margit Sára. "S-layer-streptavidin fusion proteins as template for nanopatterned molecular arrays". In: *Proceedings of the National Academy of Sciences of*

the United States of America 99.23 (2002), pp. 14646–14651. ISSN: 00278424. DOI: 10.1073/pnas.232299399.

[257] Tae Jung Park, Seok Jae Lee, Jong Pil Park, Min Ho Yang, Jong Hyun Choi, and Sang Yup Lee. "Characterization of a bacterial self-assembly surface layer protein and its application as an electrical nanobiosensor". In: *Journal of Nanoscience and Nanotechnology* 11.1 (2011), pp. 402–407. ISSN: 15334880. DOI: 10.1166/jnn.2011.3264.

[258] Patricia A. Bolla, Agustina Sanz, Sofía Huggias, José F. Ruggera, María A. Serradell, and Mónica L. Casella. "Regular arrangement of Pt nanoparticles on S-layer proteins isolated from Lactobacillus kefiri: synthesis and catalytic application". In: *Molecular Catalysis* 481.June (2020), pp. 1–9. ISSN: 24688231. DOI: 10.1016/j.mcat.2018.12.011. URL: https://doi.org/10.1016/j.mcat.2018.12.011.

[259] Jilin Tang, Helga Badelt-Lichtblau, Andreas Ebner, Johannes Preiner, Bernhard Kraxberger, Hermann J. Gruber, Uwe B. Sleytr, Nicola Ilk, and Peter Hinterdorfer. "Fabrication of highly ordered gold nanoparticle arrays templated by crystalline lattices of bacterial S-layer protein". In: *ChemPhysChem* 9.16 (2008), pp. 2317–2320. ISSN: 14397641. DOI: 10.1002/cphc.200800507.

[260] Helga Badelt-Lichtblau, Birgit Kainz, Christine Völlenkle, Eva Maria Egelseer, Uwe B. Sleytr, Dietmar Pum, and Nicola Ilk. "Genetic engineering of the S-layer protein SbpA of Lysinibacillus sphaericus CCM 2177 for the generation of functionalized nanoarrays". In: *Bioconjugate Chemistry* 20.5 (2009), pp. 895–903. ISSN: 10431802. DOI: 10.1021/bc800445r.

[261] Ulrike Jankowski, Mohamed L. Merroun, Sonja Selenska-Pobell, and Karim Fahmy. "S-Layer protein from Lysinibacillus sphaericus JG-A12 as matrix for Au III sorption and Au-nanoparticle formation". In: *Spectroscopy* 24.1-2 (2010), pp. 177–181. ISSN: 07124813. DOI: 10.3233/SPE-2010-0408.

[262] Sarang S. Puranik, Hrushikesh M. Joshi, S. B. Ogale, and K. M. Paknikar. "Hydrazine based facile synthesis and ordered assembly of metal nanoparticles (Au, Ag) on a bacterial surface layer protein template". In: *Journal of*

Nanoscience and Nanotechnology 8.7 (2008), pp. 3565–3569. ISSN: 15334880. DOI: 10.1166/jnn.2008.135.

[263] M Mertig, R Kirsch, W Pompe, and H Engelhardt. "Fabrication of highly oriented nanocluster arrays by biomolecular templating". In: *The European Physical Journal D* 9.1-4 (1999), pp. 45–48. ISSN: 14346060. DOI: 10.1007/s100530050397.

[264] Reiner Wahl, Harald Engelhardt, Wolfgang Pompe, and Michael Mertig. "Multivariate statistical analysis of two-dimensional metal cluster arrays grown in vitro on a bacterial surface layer". In: *Chemistry of Materials* 17.7 (2005), pp. 1887–1894. ISSN: 08974756. DOI: 10.1021/cm048497p.

[265] M. Panhorst, H. Brückl, B. Kiefer, G. Reiss, U. Santarius, and R. Guckenberger. "Formation of metallic surface structures by ion etching using a S-layer template". In: *Journal of Vacuum Science and Technology B: Microelectronics and Nanometer Structures* 19.3 (2001), pp. 722–724. ISSN: 10711023. DOI: 10.1116/1.1364699.

[266] Simon R. Hall, Wayne Shenton, Harald Engelhardt, and Stephen Mann. "Site-specific organization of gold nanoparticles by biomolecular templating". In: *ChemPhysChem* 2.3 (2001), pp. 184–186. ISSN: 14394235. DOI: 10.1002/1439-7641(20010316)2:3<184::AID-CPHC184>3.0.CO;2-J.

[267] Sonny S. Mark, Magnus Bergkvist, Xin Yang, Leonardo M. Teixeira, Parijat Bhatnagar, Esther R. Angert, and Carl A. Batt. "Bionanofabrication of metallic and semiconductor nanoparticle arrays using S-layer protein lattices with different lateral spacings and geometries". In: *Langmuir* 22.8 (2006), pp. 3763–3774. ISSN: 07437463. DOI: 10.1021/la053115v.

[268] Sonny S. Mark, Magnus Bergkvist, Xin Yang, Esther R. Angert, and Carl A. Batt. "Self-assembly of dendrimer-encapsulated nanoparticle arrays using 2-D microbial S-layer protein biotemplates". In: *Biomacromolecules* 7.6 (2006), pp. 1884–1897. ISSN: 15257797. DOI: 10.1021/bm0603185.

[269] Sonny S. Mark, Magnus Bergkvist, Parijat Bhatnagar, Colin Welch, Andrew L. Goodyear, Xin Yang, Esther R. Angert, and Carl A. Batt. "Thin film processing using S-layer proteins: Biotemplated assembly of colloidal gold etch

masks for fabrication of silicon nanopillar arrays". In: *Colloids and Surfaces B: Biointerfaces* 57.2 (2007), pp. 161–173. ISSN: 09277765. DOI: 10.1016/j.colsurfb.2007.01.015.

[270] S Dieluweit, D Pum, and U B Sleytr. "Formation of a gold superlattice on an S-layer with square lattice symmetry". In: *Supramolecular Science* 5.97 (1998), pp. 15–19. URL: http://linkinghub.elsevier.com/retrieve/pii/S0968567797000734{\%}5Cnpapers3://publication/uuid/F3641383-5F59-47D2-B6AD-A72B62E0F87B.

[271] M. Mertig, R. Wahl, M. Lehmann, P. Simon, and W. Pompe. "Formation and manipulation of regular metallic nanoparticle arrays on bacterial surface layers: An advanced TEM study". In: *European Physical Journal D* 8.3 (2001), pp. 317–320. ISSN: 14346060. DOI: 10.1007/s100530170119.

[272] Erika Györvary, Andrea Schroedter, Dimitri V. Talapin, Horst Weller, Dietmar Pum, and Uwe B. Sleytr. "Formation of nanoparticle arrays on S-layer protein lattices". In: *Journal of Nanoscience and Nanotechnology* 4.1-2 (2004), pp. 115–120. ISSN: 15334880. DOI: 10.1166/jnn.2004.229.

[273] K. Pollmann, M. Merroun, J. Raff, C. Hennig, and S. Selenska-Pobell. "Manufacturing and characterization of Pd nanoparticles formed on immobilized bacterial cells". In: *Letters in Applied Microbiology* 43.1 (2006), pp. 39–45. ISSN: 02668254. DOI: 10.1111/j.1472-765X.2006.01919.x.

[274] Karim Fahmy, Mohamed Merroun, Katrin Pollmann, Johannes Raff, Olesya Savchuk, Christoph Hennig, and Sonja Selenska-Pobell. "Secondary structure and Pd(II) coordination in S-layer proteins from Bacillus sphaericus studied by infrared and X-ray absorption spectroscopy". In: *Biophysical Journal* 91.3 (2006), pp. 996–1007. ISSN: 00063495. DOI: 10.1529/biophysj.105.079137.

[275] Katrin Pollmann and Sabine Matys. "Construction of an S-layer protein exhibiting modified self-assembling properties and enhanced metal binding capacities". In: *Applied Microbiology and Biotechnology* 75.5 (2007), pp. 1079–1085. ISSN: 01757598. DOI: 10.1007/s00253-007-0937-5.

[276] Ute Queitsch, Elias Mohn, Franziska Schäffel, Ludwig Schultz, Bernd Rellinghaus, Anja Blüher, and Michael Mertig. "Regular arrangement of nanoparticles from the gas phase on bacterial surface-protein layers". In: *Applied Physics Letters* 90.11 (2007), pp. 1–4. ISSN: 00036951. DOI: 10.1063/1.2713163.

[277] Ute Queitsch, Elias Mohn, Anja Blüher, Beate Katzschner, Michael Mertig, Ludwig Schultz, and Bernd Rellinghaus. "Regular arrangement of gas phase prepared in-flight annealed FePt nanoparticles on s layers". In: *IEEE Transactions on Magnetics* 44.11 PART 2 (2008), pp. 2756–2759. ISSN: 00189464. DOI: 10.1109/TMAG.2008.2002244.

[278] Ute Queitsch, Christine Hamann, Franziska Schäffel, Bernd Rellinghaus, Ludwig Schultz, Anja Blüher, and Michael Mertig. "Toward dense biotemplated magnetic nanoparticle arrays: Probing the particle-template interaction". In: *Journal of Physical Chemistry C* 113.24 (2009), pp. 10471–10476. ISSN: 19327447. DOI: 10.1021/jp9020992.

[279] Magnus Bergkvist, Sonny S. Mark, Xin Yang, Esther R. Angert, and Carl A. Batt. "Bionanofabrication of ordered nanoparticle arrays: Effect of particle properties and adsorption conditions". In: *Journal of Physical Chemistry B* 108.24 (2004), pp. 8241–8248. ISSN: 15206106. DOI: 10.1021/jp0492801.

[280] Rachel M. Abaskharon and Feng Gai. "Meandering Down the Energy Landscape of Protein Folding: Are We There Yet?" In: *Biophysical Journal* 110.9 (2016), pp. 1924–1932. ISSN: 15420086. DOI: 10.1016/j.bpj.2016.03.030. URL: http://dx.doi.org/10.1016/j.bpj.2016.03.030.

[281] Elsa Valero, Miguel Martín, Natividad Gálvez, Purificación Sánchez, Johannes Raff, Mohamed L. Merroun, and Jose M. Dominguez-Vera. "Nanopatterning of Magnetic CrNi Prussian Blue Nanoparticles Using a Bacterial S-Layer as a Biotemplate". In: *Inorganic Chemistry* 54.14 (2015), pp. 6758–6762. ISSN: 1520510X. DOI: 10.1021/acs.inorgchem.5b00555.

[282] Sonny S. Mark, Magnus Bergkvist, Xin Yang, Leonardo M. Teixeira, Parijat Bhatnagar, Esther R. Angert, and Carl A. Batt. "Bionanofabrication of metallic and semiconductor nanoparticle arrays using S-layer protein lattices with

different lateral spacings and geometries". In: *Langmuir* 22.8 (2006), pp. 3763–3774. ISSN: 07437463. DOI: 10.1021/la053115v.

[283] Sonja Selenska-Pobell, Thomas Reitz, Rico Schönemann, Thomas Herrmansdörfer, Mohamed Merroun, Andrea Geißler, Juan Bartolomé, Fernando Bartolomé, Luis Miguel García, Fabrice Wilhelm, and Andrei Rogalev. "Magnetic au nanoparticles on archaeal s-layer ghosts as templates". In: *Nanomaterials and Nanotechnology* 1.2 (2011), pp. 8–16. ISSN: 18479804. DOI: 10.5772/50955.

[284] Matthias Suhr, Nancy Unger, Karen E. Viacava, Tobias J. Günther, Johannes Raff, and Katrin Pollmann. "Investigation of metal sorption behavior of Slp1 from Lysinibacillus sphaericus JG-B53: a combined study using QCM-D, ICP-MS and AFM". In: *BioMetals* 27.6 (2014), pp. 1337–1349. ISSN: 15728773. DOI: 10.1007/s10534-014-9794-8.

[285] Christina Schäffer, Thomas Wugeditsch, Hanspeter Kählig, Andrea Scheberl, Sonja Zayni, and Paul Messner. "The surface layer (S-layer) glycoprotein of Geobacillus stearothermophilus NRS 2004/3a. Analysis of its glycosylation". In: *Journal of Biological Chemistry* 277.8 (2002), pp. 6230–6239. ISSN: 00219258. DOI: 10.1074/jbc.M108873200.

[286] Birgit Kainz, Kerstin Steiner, Marco Möller, Dietmar Pum, Christina Schäffer, Uwe B. Sleytr, and José L. Toca-Herrera. "Absorption, steady-state fluorescence, fluorescence lifetime, and 2D self-assembly properties of engineered fluorescent S-layer fusion proteins of Geobacillus stearothermophilus NRS 2004/3a". In: *Biomacromolecules* 11.1 (2010), pp. 207–214. ISSN: 15257797. DOI: 10.1021/bm901071b.

[287] Birgit Kainz, Kerstin Steiner, Uwe B. Sleytr, Dietmar Pum, and José L. Toca-Herrera. "Fluorescence energy transfer in the bi-fluorescent S-layer tandem fusion protein ECFP-SgsE-YFP". In: *Journal of Structural Biology* 172.3 (2010), pp. 276–283. ISSN: 10478477. DOI: 10.1016/j.jsb.2010.07.002.

[288] Birgit Kainz, Gerhard Stadlmayr, Heinrich Schuster, Dietmar Pum, Paul Messner, Christian Obinger, and Christina Schäffer. "A Fusion Tag to Fold on : The S-Layer Protein SgsE Confers Improved Folding Kinetics to Translationally Fused Enhanced Green Fluorescent Protein". In: 22.9 (2012), pp. 1271–1278.

[289] René Novotny, Andrea Scheberl, Marc Giry-Laterriere, Paul Messner, and Christina Schäffer. "Gene cloning, functional expression and secretion of the S-layer protein SgsE from Geobacillus stearothermophilus NRS 2004/3a in Lactococcus lactis". In: *FEMS Microbiology Letters* 242.1 (2005), pp. 27–35. ISSN: 03781097. DOI: 10.1016/j.femsle.2004.10.036.

[290] David G. Myszka and Irwin M. Chaiken. "Design and Characterization of an Intramolecular Antiparallel Coiled Coil Peptide". In: *Biochemistry* 33.9 (1994), pp. 2363–2372. ISSN: 15204995. DOI: 10.1021/bi00175a003.

[291] Aijun Tang, Chun Wang, Russell J. Stewart, and Jindřich Kopeček. "The coiled coils in the design of protein-based constructs: Hybrid hydrogels and epitope displays". In: *Journal of Controlled Release* 72.1-3 (2001), pp. 57–70. ISSN: 01683659. DOI: 10.1016/S0168-3659(01)00262-0.

[292] Daniel G. Gurnon, Jennifer A. Whitaker, and Martha G. Oakley. "Design and characterization of a homodimeric antiparallel coiled coil". In: *Journal of the American Chemical Society* 125.25 (2003), pp. 7518–7519. ISSN: 00027863. DOI: 10.1021/ja0357590.

[293] Vincent Batori, Akiko Koide, and Shohei Koide. "Exploring the potential of the monobody scaffold: Effects of loop elongation on the stability of a fibronectin type III domain". In: *Protein Engineering* 15.12 (2002), pp. 1015–1020. ISSN: 02692139. DOI: 10.1093/protein/15.12.1015.

[294] Akiko Koide, Charles W. Bailey, Xiaolin Huang, and Shohei Koide. "The fibronectin type III domain as a scaffold for novel binding proteins". In: *Journal of Molecular Biology* 284.4 (1998), pp. 1141–1151. ISSN: 00222836. DOI: 10.1006/jmbi.1998.2238.

[295] Jinzhu Duan, Jinsong Wu, C. Alexander Valencia, and Rihe Liu. "Fibronectin type III domain based monobody with high avidity". In: *Biochemistry* 46.44 (2007), pp. 12656–12664. ISSN: 00062960. DOI: 10.1021/bi701215e.

[296] Benjamin J. Hackel, Atul Kapila, and K. Dane Wittrup. "Picomolar Affinity Fibronectin Domains Engineered Utilizing Loop Length Diversity, Recursive Mutagenesis, and Loop Shuffling". In: *Journal of Molecular Biology* 381.5 (2008), pp. 1238–1252. ISSN: 00222836. DOI: 10.1016/j.jmb.2008.06.051.

[297] Peter G. Chandler and Ashley M. Buckle. "Development and Differentiation in Monobodies Based on the Fibronectin Type 3 Domain". In: *Cells* 9.3 (2020), p. 610. DOI: 10.3390/cells9030610.

[298] Michael Fountoulakis, Jean Francois Juranville, and Michael Manneberg. "Comparison of the Coomassie brilliant blue, bicinchoninic acid and Lowry quantitation assays, using non-glycosylated and glycosylated proteins". In: *Journal of Biochemical and Biophysical Methods* 24.3-4 (1992), pp. 265–274. ISSN: 0165022X. DOI: 10.1016/0165-022X(94)90078-7.

[299] Simon W M Tanley and John R Helliwell. "Structural dynamics of cisplatin binding to histidine in a protein". In: *Structural Dynamics* 1.3 (2014), p. 34701.

[300] Malcolm H Smith. "The amino acid composition of proteins". In: *Journal of Theoretical Biology* 13 (1966), pp. 261–282. ISSN: 0022-5193. DOI: https://doi.org/10.1016/0022-5193(66)90021-X. URL: http://www.sciencedirect.com/science/article/pii/002251936690021X.

[301] Laura K Henchey, Andrea L Jochim, and Paramjit S Arora. "Contemporary strategies for the stabilization of peptides in the α − *helicalconformation*". In: *Current opinion in chemical biology* 12.6 (2008), pp. 692–697.

[302] Eva Bereczk-Tompa, Mihaly Posfai, Balazs Toth, and Ferenc Vonderviszt. "Magnetite-Binding Flagellar Filaments Displaying the MamI Loop Motif". In: *CHEMBIOCHEM* 17.21 (2016), pp. 2075–2082. ISSN: 1439-4227. DOI: 10.1002/cbic.201600377.

[303] Andreas J Wolf, Hubert Hauser, Volker Kübler, Christian Walk, Oliver Höhn, and Benedikt Bläsi. "Origination of nano-and microstructures on large areas by interference lithography". In: *Microelectronic Engineering* 98 (2012), pp. 293–296.

[304] Filipe Pinto, Ella Lucille Thornton, and Baojun Wang. "An expanded library of orthogonal split inteins enables modular multi-peptide assemblies". In: *Nature Communications* 11.1 (2020), pp. 1–15. ISSN: 20411723. DOI: 10.1038/s41467-020-15272-2. URL: http://dx.doi.org/10.1038/s41467-020-15272-2.

[305] Clemens F. Kaminski and Gabriele S. Kaminski Schierle. "Probing amyloid protein aggregation with optical superresolution methods: from the test tube to models of disease". In: *Neurophotonics* 3.4 (2016), p. 041807. ISSN: 2329-423X. DOI: 10.1117/1.nph.3.4.041807.

[306] Einav Tayeb-Fligelman, Orly Tabachnikov, Asher Moshe, Orit Goldshmidt-Tran, Michael R. Sawaya, Nicolas Coquelle, Jacques Philippe Colletier, and Meytal Landau. "The cytotoxic Staphylococcus aureus PSMα3 reveals a cross-α amyloid-like fibril". In: *Science* 355.6327 (2017), pp. 831–833. ISSN: 10959203. DOI: 10.1126/science.aaf4901.

[307] Elizabeth H.C. Bromley, Kevin Channon, Efrosini Moutevelis, and Derek N. Woolfson. "Peptide and protein building blocks for synthetic biology: From programming biomolecules to self-organized biomolecular systems". In: *ACS Chemical Biology* 3.1 (2008), pp. 38–50. ISSN: 15548929. DOI: 10.1021/cb700249v.

[308] Simona Poviloniene, Vida Časaite, Virginijus Bukauskas, Arunas Šetkus, Juozas Staniulis, and Rolandas Meškys. "Functionalization of α-synuclein fibrils". In: *Beilstein Journal of Nanotechnology* 6.1 (2015), pp. 124–133. ISSN: 21904286. DOI: 10.3762/bjnano.6.12.

[309] Steven J. Roeters, Aditya Iyer, Galja Pletikapiä, Vladimir Kogan, Vinod Subramaniam, and Sander Woutersen. "Evidence for Intramolecular Antiparallel Beta-Sheet Structure in Alpha-Synuclein Fibrils from a Combination of Two-Dimensional Infrared Spectroscopy and Atomic Force Microscopy". In: *Scientific Reports* 7.January (2017), pp. 1–11. ISSN: 20452322. DOI: 10.1038/srep41051. URL: http://dx.doi.org/10.1038/srep41051.

[310] Thierry Lefèvre and Michèle Auger. "Spider silk as a blueprint for greener materials: A review". In: *International Materials Reviews* 61.2 (2016), pp. 127–153. ISSN: 17432804. DOI: 10.1080/09506608.2016.1148894.

[311] Scott M. Bird, Andrea E. Rawlings, Johanna M. Galloway, and Sarah S. Staniland. "Using a biomimetic membrane surface experiment to investigate the activity of the magnetite biomineralisation protein Mms6". In: *RSC Advances* 6.9 (2016), pp. 7356–7363. ISSN: 20462069. DOI: 10.1039/c5ra16469a.

[312] Sinéad M. Smith. *Strategies for the purification of membrane proteins*. Vol. 1485. 2017, pp. 389–400. ISBN: 9781607619123. DOI: 10.1007/978-1-4939-6412-3_21.

[313] Khadija Mathieu, Waqas Javed, Sylvain Vallet, Christian Lesterlin, Marie Pierre Candusso, Feng Ding, Xiaohong Nancy Xu, Christine Ebel, Jean Michel Jault, and Cédric Orelle. "Functionality of membrane proteins overexpressed and purified from E. coli is highly dependent upon the strain". In: *Scientific Reports* 9.1 (2019), pp. 1–15. ISSN: 20452322. DOI: 10.1038/s41598-019-39382-0.

[314] Nicola Ilk, Eva M Egelseer, and Uwe B Sleytr. "S-layer fusion proteins—construction principles and applications". In: *Current opinion in biotechnology* 22.6 (2011), pp. 824–831.

[315] Philip D. Laible, Heather N. Scott, Lynda Henry, and Deborah K. Hanson. "Towards higher-throughput membrane protein production for structural genomics initiatives". In: *Journal of Structural and Functional Genomics* 5.1-2 (2004), pp. 167–172. ISSN: 1345711X. DOI: 10.1023/B:JSFG.0000029201.33710.46.

[316] Irem Nasir, Warda Fatih, Anja Svensson, Dennis Radu, Sara Linse, Celia Cabaleiro Lago, and Martin Lundqvist. "High throughput screening method to explore protein interactions with nanoparticles". In: *PloS one* 10.8 (2015), e0136687.

[317] Carlos Tassa, Jay L Duffner, Timothy A Lewis, Ralph Weissleder, Stuart L Schreiber, Angela N Koehler, and Stanley Y Shaw. "Binding affinity and kinetic analysis of targeted small molecule-modified nanoparticles". In: *Bioconjugate chemistry* 21.1 (2010), pp. 14–19.

[318] Ronald Frank. "Spot-synthesis: an easy technique for the positionally addressable, parallel chemical synthesis on a membrane support". In: *Tetrahedron* 48.42 (1992), pp. 9217–9232.

[319] Jeffery L. Yarger, Brian R. Cherry, and Arjan Van Der Vaart. "Uncovering the structure-function relationship in spider silk". In: *Nature Reviews Materials* 3.3 (2018). ISSN: 20588437. DOI: 10.1038/natrevmats2018.8.

[320] Kyeong Kyu Kim, Rosalind Kim, and Sung-Hou Kim. "Crystal structure of a small heat-shock protein". In: *Nature* 394.6693 (1998), pp. 595–599.

[321] Kiyotaka Wasa, Isaku Kanno, and Hidetoshi Kotera. *Handbook of sputter deposition technology: fundamentals and applications for functional thin films, nanomaterials and MEMS*. William Andrew, 2012.