Probing the structure and metal binding of the magnetite biomineralisation protein MmsFcc via NMR spectroscopy.

Fraser Beaumont

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

The University of Sheffield
Faculty of Science
School of Biosciences

23/09/2024

Abstract

Magnetite (Fe_3O_4) nanoparticles are an established technology, whose further adoption is limited by the scope of current synthesis methods. Magnetite nanoparticles are also found in magnetotactic bacteria, which mediate the biomineralisation of a variety of nanoparticle morphologies through the magnetosome organelle. Magnetosome membrane (Mms) proteins have been demonstrated to greatly improve the quality of nanoparticles produced by chemical synthesis methods. Of these proteins, MmsF has been identified as a key mediator of nanoparticle synthesis *in vivo* and *in vitro*, where it interacts directly with the growing nanoparticle through a 13-residue loop sequence (DRDDEFVYFHAKQ). This study utilises an artificial antiparallel coiled-coil display system, called MmsFcc, for the functional and structural study of the MmsF loop via NMR spectroscopy. This thesis first describes the optimisation of the purification strategy for MmsFcc, alongside the reduced function mutant MmsFcc D49S. We carry out backbone and sidechain assignment of MmsFcc, then utilise this to experimentally determine the loop structure of MmsF. In this structure, some sidechains have well-defined positions while others seem much more variable, which has interesting implications for how MmsF functions to control magnetite morphology. Next, we perform a series of metal titrations to assess the metal binding properties of MmsFcc. We report that MmsFcc interacts transiently with both Fe^{2+} and Zn^{2+}, but not with Fe^{3+} and Ca^{2+}, and that this interaction induces a conformational change. Furthermore, this interaction is dependent on the deprotonation of histidine 58 in the MmsF loop. We propose that the histidine dependant interaction with metal induces a conformational change towards a lower energy bound structure, which positions the cation in an optimal conformation for the correct crystallisation into magnetite. This complex is then released, due to the transient nature of the interaction.

Abbreviations

- MNP: Magnetite Nanoparticle
- NMR: Nuclear magnetic resonance
- MRI: Magnetic resonance imaging
- MTB: Magnetotactic bacteria
- Mms: Magnetosome membrane specific
- Mam: Magnetosome associated membrane
- OATZ: Oxic-Anoxic transition zone
- MAI: Magnetosome island
- AMB-1: *Magnetospirillum magneticum* AMB-1
- MSR-1: *Magnetospirillum gryphiswaldense* MSR-1
- cc: coiled-coil
- MmsFcc: Magnetosome membrane specific F coiled-coil
- SAXS: Small angle X-Ray scattering
- CD: Circular dichroism
- RDC: Residual dipolar coupling
- PRE: Paramagnetic relaxation enhancement
- NOE: Nuclear Overhauser effect
- TEM: Transmission electron microscopy
- HSQC: Heteronuclear single quantum coherence
- NOESY: Nuclear Overhauser effect spectroscopy
- LB: Lysogeny broth
- SDS-PAGE: Sodium dodecyl sulphate polyacrylamide gel electrophoresis
- TEMED: Tetramethylethylenediamine
- EDTA: Ethylenediaminetetraacetic acid
- DTT: Dithiothreitol
- *E.* coli: *Escherichia coli*
- TEV: Tobacco etch virus
- GuHCl: Guanidine hydrochloride
- MWCO: Molecular weight cutoff
- TSP: Trimethylsilylpropanoic acid
- Ni-NTA: Nickel-nitrilotriacetic acid
- SEC: Size exclusion chromatography
- IPTG: Isopropyl ß-D-1-thiogalactopyranoside
- TFE: Trifluoroethanol
- LCMS: Liquid chromatography–mass spectrometry
- NUS: Non-uniform sampling
- TOCSY: Total correlation spectroscopy
- RMSD: Root mean square deviation

Author declaration

I declare that all work presented in this thesis is my own, and that any republished work and ideas are properly accredited.

Acknowledgements

Firstly, I would like to thank the Engineering and Physical sciences research council for funding this project. Thank you to Dr Sarah Staniland for giving me a change and getting me started on this project. Special thanks to Professor Mike Williamson for becoming my primary supervisor, and for being extremely generous with your wisdom and time. It has been a privilege to work with you. Special thanks to Dr Barbara Ciani for welcoming me into her research group once my supervisor left, it is very much appreciated. Thank you to Dr Michelle Rowe for lending her expertise in running the NMR experiments, and for being a great companion during magfill and experiment time. Special thanks to Jess Hurcum for being my day one lab mate, best of luck with the rest of your PhD. I would like to thank my two master's students, Isobel Walker and Edward Carlile, for their hard work and patience as I learned to become a better teacher. Thank you to Dr David Finger and Dr Reuben Ouanounou for generously donating the His-TEV protease used in this chapter, and for lending their expertise using AKTA chromatography systems. I would like to say thank you to the wider PhD community across Biosciences and Chemistry, who have always been willing to lend an ear. Thank you to all the support staff and porters who carry out crucial work which allows research to take place.

I thank everybody in the Reson8 NMR community who have been instrumental in improving my understanding of the principles underlying NMR. This extends to the CCPN development team, for developing the freely available software essential for completing this project.

I would also like to thank my parents for being incredibly supportive of my journey from a child interested in understanding the world to the person I am today. I am eternally grateful for the support of my partner, Isobel, without you I wouldn't have been able to do this. I would also like to thank my cat Milkshake, for being the best writing companion I could ask for.

Table of contents

Table of figures

Table of tables

Table of appendix

Introduction

1.1 Magnetite nanoparticles as an established technology, whose function relies on precise control of their structure

Magnetite nanoparticles possess many qualities that make them attractive for a variety of applications. Like other nanoparticles, they can pass through membrane barriers, and possess a large surface area to volume ratio, making them excellent candidates for coating with functional molecules such as drugs (Ghazanfari et al., 2016; Luchini & Vitiello, 2019; Price et al., 2018). Nanoparticles can also be coated with compounds to adjust properties such as: biocompatibility, targeting, and aggregation (Gao et al., 2020; León Félix et al., 2019; Luchini & Vitiello, 2019). What separates magnetite nanoparticles from other nanoparticles is their intrinsic magnetism, which greatly increases the potential applications of MNPs. Of all the composite materials that MNPs can consist of, magnetite is an attractive option due to its biocompatibility. An example of this application is in targeted drug delivery: nanoparticles coated with drugs can be directed towards the therapeutic target region via a magnetic probe, greatly reducing off-target effects. The magnetic properties of magnetite nanoparticles are dictated by the size, morphology, and crystallinity of the nanoparticles. Furthermore, small changes to these factors can alter the sample's magnetic properties, meaning that maintaining sample homogeneity is of critical importance. One property that is of particular importance to researchers is the phenomenon of superparamagnetism. Sufficiently small magnetite nanoparticles are superparamagnetic, meaning that they possess no magnetic memory. The ability to switch magnetism on and off is an attractive property for some cases, such as the treatment of cancer cells through magnetic hyperthermia (Liu et al., 2020). This is one reason why it is important to be able to synthesize nanoparticles of specific sizes.

To maintain a sufficient degree of control over in vitro synthesis methods, high reaction temperatures, and the use of toxic chemicals are often required. These parameters raise concern over the sustainability of MNP production, and the use of toxic chemicals can render samples unsuitable for biomedical applications (Khan et al., 2019; Ruby et al., 2006; Rui et al., 2010). One of the most well-established biomedical applications for MNPs is as a contrast agent in for magnetic resonance imaging (MRI) diagnostic techniques. Contrast

agents work by speeding up the rate of relaxation, thus increasing the signal outputted during scans. Studies have established that longitudinal (T_1) magnetism of cubic nanoparticles relaxes at a faster rate than that spherical nanoparticles, thus providing a greater contrast and subsequent image quality (Zhen et al., 2010). This example highlights the importance of controlling nanoparticle morphology, and the need for the ability to synthesise different shapes. Another application of superparamagnetic nanoparticles is the previously mentioned magnetic hyperthermia treatment of cancer cells. In this therapy, MNPs are localised to the tumour site, then subjected to an alternating magnetic field in order to raise the temperature of surrounding cancerous tissues, promoting cell death (Wahajuddin and Arora, 2012; Liu et al., 2020). This method has been suggested as a theranostic approach to cancer therapy, merging both diagnosis and therapy, due to MNPs' previously discussed role as MRI contrast agents (Pouliquen et al., 1991; Wei et al., 2017). Clearly, MNPs have many demonstratable biomedical applications. It has been suggested that the further uptake of this technology is limited by the cost of synthesis methods, and through the limited control conferred by synthesis methods in the absence of toxic chemicals (Rawlings et al., 2019). A potential solution to this problem exists in the proteome of magnetotactic bacteria, which are introduced in Section 1.2.

1.2. Magnetotactic bacteria and the magnetosome

Magnetotactic bacteria (MTB) are a polyphyletic group of bacteria found ubiquitously in both fresh and saltwater habitats, characterized by their ability to biomineralize iron oxide or iron sulphide magnetic nanoparticles; this synthesis takes place in an organelle called the Magnetosome (Uebe & Schüler, 2016). Through the function of a series of Magnetosome Membrane Specific (Mms) and Magnetosome Associated Membrane (Mam) proteins, these bacteria can precisely synthesize nanoparticles with a variety of morphologies and magnetic properties (Figure 1.1).

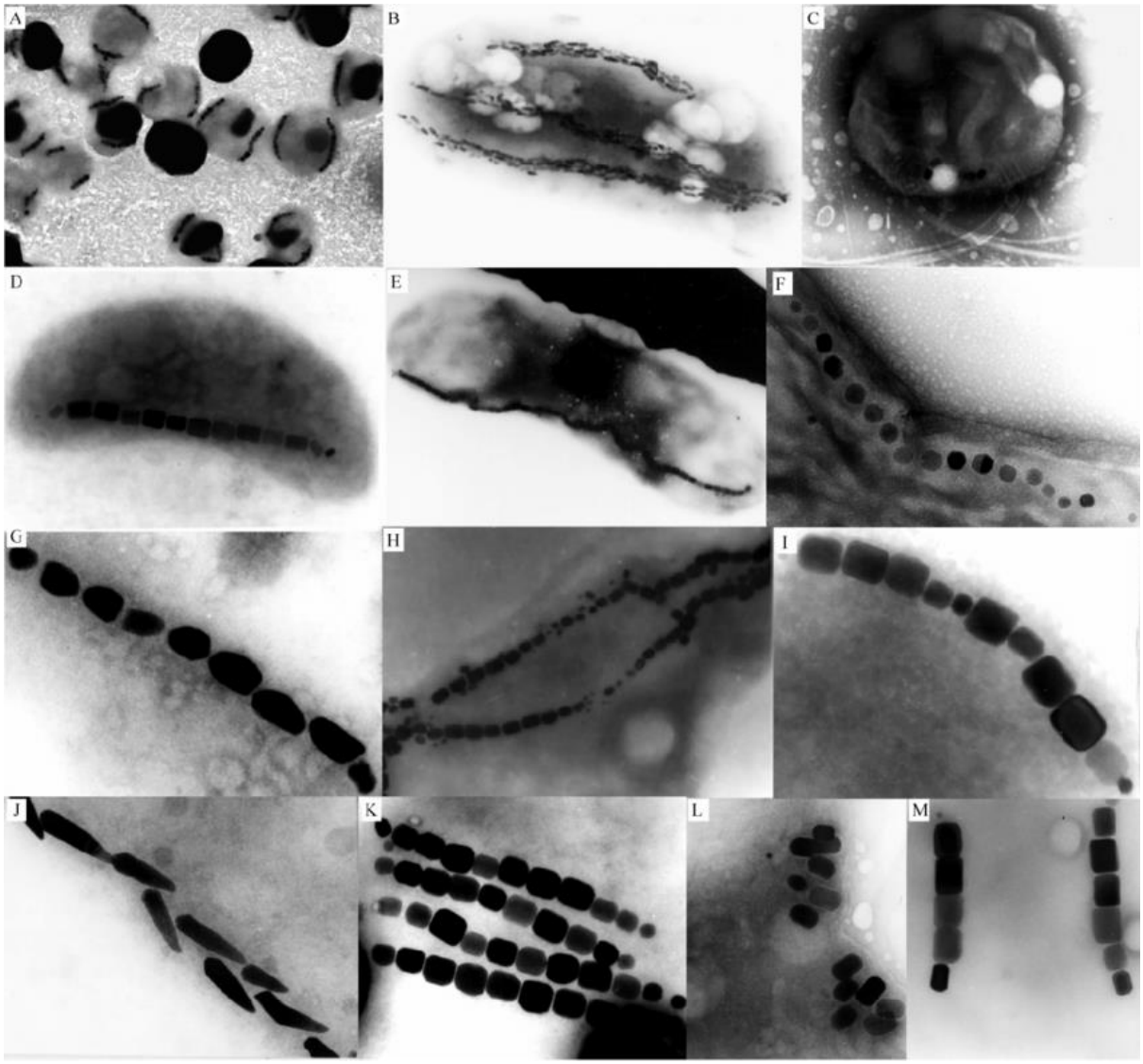


Figure 1.1. Diversity of magnetotactic bacteria and magnetosomes. Panels A–F are transmission electron micrographs (TEM) of a variety of bacterial morphologies seen in MTB. A) coccoid. B) Elongated bacterium with multiple magnetosome chains. C) coccoidal bacterium (D) *Vibrio* bacterium with magnetosome chain. E) Rod-shaped bacterium with one magnetosome chain. (F) *Spirilla* bacterium. Pannels G–M are electron micrographs of a variety of nanoparticle morphologies, bacterial morphologies, and patterns of magnetosome organisations seen in MTB. With the exception of (F), all examples are from uncultivated MTB, which were magnetically collected from environmental samples. Figure taken from (Scheffel et al., 2005).

Remarkably, this synthesis is reproducible amongst individuals of a species and does not require the use of high temperatures and pressures. This is in stark contrast to most commercially viable methods used today to synthesize magnetic nanoparticles. For this reason, the proteins responsible for this synthesis have become a key research focus for the development of more sustainable synthesis methods (Rawlings et al., 2019; Yew et al.,

2016). The application of MTB for the improved synthesis of MNPs can be categorised into two main fields. The first is the use of the bacteria as tools of biotechnology to grow nanoparticles, then either harvest the magnetosomes or to utilise the bacterial culture. Secondly, and the focus of this thesis, is the addition of key magnetosome proteins to commercial synthesis of MNPs to modulate nanoparticle properties. The end goal of this line of research is the ability to precisely synthesise nanoparticles to customer specifications of size and morphology.

Magnetosome mediated biomineralisation is a complex process that involves a variety of different proteins. It is a taxing process for the bacterium, with them sequestering and biomineralizing enough that roughly 2-3% of dry cellular weight is iron (Lin *et al.*, 2013). Clearly, this and the diversity observed demonstrate this is an important process. Despite this, there is still some debate as to why the magnetosome has evolved. The most commonly accepted theory for why these bacteria synthesise magnetosomes is that the bacterium aligns itself along the Earth's geomagnetic field lines using the intracellular magnetics, and this alignment is utilised to direct swimming toward the organism's preferred environment. MTB are either microaerophiles or anaerobes, meaning that they prefer an environment containing little to no oxygen. MTB are primarily found in the Oxic-Anoxic Transition Zone (OATZ), where oxygenated water or sediment mixes with deoxygenated water or sediment (Figure 1.2). The directed movement of these bacteria is termed Magnetotaxis, although this is perhaps a misnomer since taxis implies the directed movement toward or away from an external stimulus. A more accurate description would be chemotaxis (in this case movement towards a region of optimal oxygen concentration), assisted by the internal magnetic compass provided by the magnetosomes.

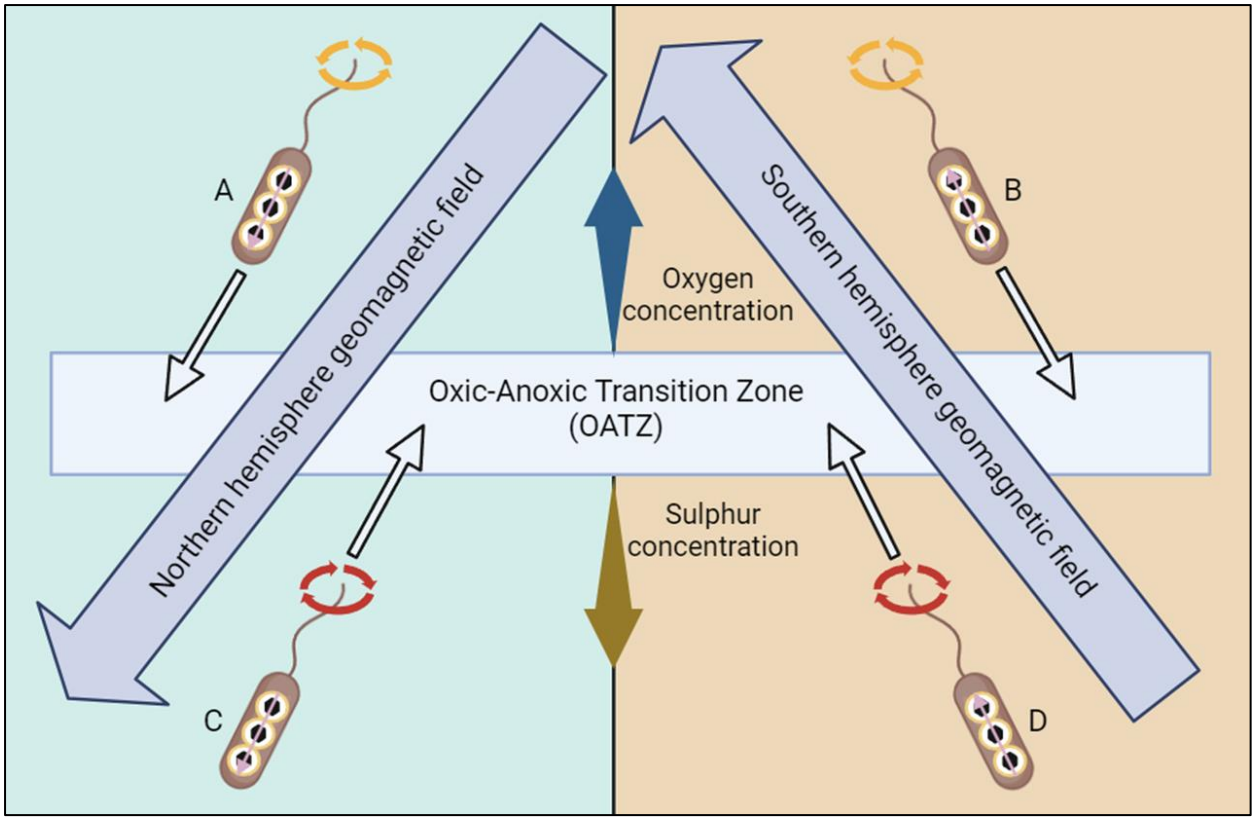


Figure 1.2. Diagram representing the proposed function of magnetosomes, to provide a one-dimensional axis of direction along the earth's geomagnetic field lines. The magnetic moments of the compass needle-like magnetosomes align with the earth's geomagnetic field. In the northern and southern hemisphere, whether bacteria swim against or with the magnetic alignment is flipped, as demonstrated by the pairs A+B, and C+D. Figure adapted from https://www.nature.com/scitable/knowledge/library/bacteria-that-synthesize-nano-sized-compasses-to-15669190/.

To biomineralize nanoparticles, the magnetosome relies on the function of non-specific proteins, which are involved in other cellular processes (an example being iron uptake from the environment). There also exists a class of genes specific to the biomineralization process, which are all located in a region of the genome called the Magnetosome island (MAI). The MAI constitutes roughly 2% of a typical *Magnetospirillum* genome and contains roughly 30 mam and mms genes. A list of these genes, and the predicted function of these proteins is provided in Table 1.1.

Protein Name	MTB strain(s)	Predicted Function	References
AOR	*M. magneticum* AMB - 1	Iron absorption and transport	(Wahyudi et al., 2003)
MagA	*M. magneticum* AMB - 1	Iron absorption and transport	(Lohße et al., 2011)

MamA	*M. magneticum* AMB - 1	Iron absorption and transport	(Grünberg et al., 2004; Lohße et al., 2014)
MamV	*M. magneticum* AMB - 1	Iron absorption and transport	(Grünberg et al., 2004)
MamB	*M. gryphiswaldense* MSR - 1	Iron absorption and transport	(Grünberg et al., 2004)
MamH	*M. gryphiswaldense* MSR - 1	Iron absorption and transport	(Uebe et al., 2011)
MamZ	*M. gryphiswaldense* MSR - 1	Iron absorption and transport	(Uebe et al., 2011)
MamM	*M. gryphiswaldense* MSR - 1	Iron absorption and transport	(Grünberg *et al*., 2004)
MamT	*M. gryphiswaldense* MSR - 1	Magnetosome formation	(C. et al., 2022)
MamL	*M. gryphiswaldense* MSR - 1	Membrane Invagination	(Lohße et al., 2014)
MamI	*M. magneticum* AMB - 1	Membrane Invagination	(Murat et al., 2010)
MamQ	*M. gryphiswaldense MSR - 1*	Membrane Invagination	(Murat *et al*., 2010)
MamY	*M. gryphiswaldense MSR - 1*	Membrane Invagination	(Tanaka et al., 2010)
MpsA	*M. magneticum AMB - 1*	Membrane Invagination	(Tanaka et al., 2010)
Mms16	*M. gryphiswaldense MSR - 1*	Membrane Invagination	(Okamura et al., 2001)
MamU	*M. magneticum AMB - 1*	Membrane Invagination	(Nudelman et al., 2016)
MamX	*M. magneticum AMB - 1*	Iron absorption and transport	(Yang et al., 2013)
MamG	*M. gryphiswaldense MSR - 1*	Magnetite crystal size and morphology control	(Scheffel et al., 2008)
MmsF	*M. magneticum* AMB - 1	Magnetite crystal size and morphology control	(Rawlings et al., 2014)
MamC / Mms13/ Mam12	*M. gryphiswaldense MSR - 1*	Magnetite crystal size and morphology control	(Arakaki et al., 2003)
MamD / Mms7	*M. magneticum AMB - 1*	Magnetite crystal size and morphology control	(Nudelman et al., 2016)

MamO	*M. magneticum AMB - 1*	Magnetite crystal size and morphology control	(Hershey et al., 2016)
MamN	*M. magneticum AMB - 1*	Magnetite crystal size and morphology control	(Yan et al., 2017)
MamS	*M. magneticum AMB - 1*	Magnetite crystal size and morphology control	(Murat et al., 2010)
MamR	*M. gryphiswaldense MSR - 1*	Magnetite crystal size and morphology control	(Murat et al., 2010)
FtsZ - like	*M. gryphiswaldense MSR - 1*	Magnetite crystal size and morphology control	(Arakaki et al., 2014)
Mms5	*M. magneticum AMB - 1*	Magnetite crystal size and morphology control	(Arakaki et al., 2014)
Mms6	*M. gryphiswaldense MSR - 1*	Magnetite crystal size and morphology control	(Nudelman et al., 2016)

Table 1.1. A comprehensive summary of known magnetosome specific proteins with references to key papers. Table adapted from (Ying et al., 2022).

As shown in Figure 1.1, the variety of MNPs synthesised by MTB is vast, highlighting the potential for proteins involved in mediating magnetite crystal size and morphology to be developed into a toolbox for customisable nanoparticle synthesis. The majority of MTB species synthesise Magnetite (Fe_3O_4) crystals, while some synthesise Greigite (Fe_2S_4) crystals (Lefevre & Bazylinski, 2013). Studies have primarily focussed on Magnetite synthesis since it is more biocompatible than Greigite. *Magnetospirillum magneticum* AMB-1 (AMB-1) and *Magnetospirillum gryphiswaldense* MSR-1 (MSR-1), have each been studied extensively due to their early adoption as laboratory strains (as evidenced by Table 1.x). Through these studies, it has been possible to identify and characterise many Mms and Mam proteins key to nanoparticle synthesis. Magnetosome genes can be broadly categorised into two groups based on their evolutionary conservation. The first, called 'core' magnetosome genes, are named so since they are conserved across all MTB species. This core group consists of 9 genes (mamA, mamB, mamE, mamI, mamK, mamM, mamO, mamP, and mamQ) (Taoka et al., 2023). Non-universally conserved proteins can give rise to the variation seen across MTB species. Magnetosome biogenesis is a complex process, which can be broken down into 4 stages. These stages, and the proteins associated with them are listed in Table 1.2. Evolutionarily conserved core magnetosome proteins are highlighted in bold.

1) Membrane formation	2) Alignment and positioning of magnetosomes	3) Maturation and lumen condition management	4) Magnetite nucleation and biomineralisation
MamB, **MamI**, MamL, **MamQ**, **MamM**	**MamA**, MamY, McaB, MamJ, **MamK**	**MamE**, **MamO**, **MamA**, **MamB**, MamH, **MamM**, MamN, MamZ, MamX, MamT, **MamP**	Mms13, Mms7, Mms6, Mms5, MmsF
1) Membrane formation			
	2) Alignment of magnetosome		
		3) Maturation of magnetosome and lumen	
			4) Biomineralization

Table 1.2. The four stages of magnetosome development, with known key proteins listed. A Gannt chart is included to visualise the overlapping of these stages (times are an approximation). A clear box denotes that the process is not occurring, while green denotes when the process is occurring. Figure adapted from (Taoka et al., 2023)

From Table 1.2, it can be seen that magnetosome development is a complex process which can be categorised into four overlapping stages. Firstly, invagination of cell membrane in an 'empty' bacterium. The budding vesicle remains attached in the early stages as membrane proteins are incorporated. Secondly, alignment of the magnetosomes through interaction between the membrane protein MamJ and the actin-like filamentous protein MamK occurs throughout the development timeline of the Magnetosome. Thirdly, the magnetosome is 'matured' through the processing of membrane proteins by the protease MamE. Simultaneously, transport and redox proteins modulate the conditions of the lumen for subsequent nanoparticle biomineralization. Finally, nanoparticle growth and nucleation is then mediated by proteins known to interact with the crystal. Many proteins, including MamA, MamB and MamM, are active at different stages of magnetosome development. Interestingly, Magnetite nucleation and biomineralization is mediated by the action of only non-conserved proteins. These magnetosome membrane specific proteins are often seen bound to extracted magnetite nanoparticles, through their direct interaction with the crystal (Arakaki et al., 2003; Grünberg et al., 2004). These proteins have been shown to mediate nanoparticle morphology, and are therefore an attractive candidate for improving methods of nanoparticle synthesis (Rawlings et al., 2014, 2016). The standard approach used to

examine the activity of these proteins *in vitro* is through the synthesis of nanoparticles via chemical co-precipitation, which is discussed in Section 1.3

1.3 Chemical co-precipitation of Magnetite as an analytical tool for observed magnetosome protein function

The chemical co-precipitation of magnetite is a simple chemical reaction used to synthesise crude nanoparticles from a mix of Fe^{2+}/Fe^{3+} by the stepwise addition of a strong base such as sodium hydroxide. A simplified reaction scheme is displayed below:

$$Fe^{2+} + 2Fe^{3+} + 8OH^- \rightarrow 2FeO_2H + Fe^{2+}.2H_2O + 2OH^- \rightarrow Fe_3O_4 + 4H_2O$$

If performed under 'ecological conditions' (i.e. atmospheric pressure, room temperature), then the control over important nanoparticle batch qualities such as morphology, crystallinity, and size distribution is severely limited. The creation of other iron oxide species is also prolific in this case. Remarkably, with only the addition of certain Mms proteins (e.g. MmsF, Mms6) it is possible to produce MNPs with desirable properties from this benchtop co-precipitation (Amemiya et al., 2007; Lohße et al., 2011; Rawlings et al., 2014; Valverde-Tercedor et al., 2015). To achieve this, only a small amount of protein is required (10 µg per 1 mL reaction solution). This is displayed in Figure 1.3, where the addition of self-assembled MmsF proteinosomes improves the quality of nanoparticles produced by room temperature co-precipitation. MmxF and MamF are non-functional homologues of MmsF, which also exist in the in AMB-1 genome; they are discussed further in Section 1.4.

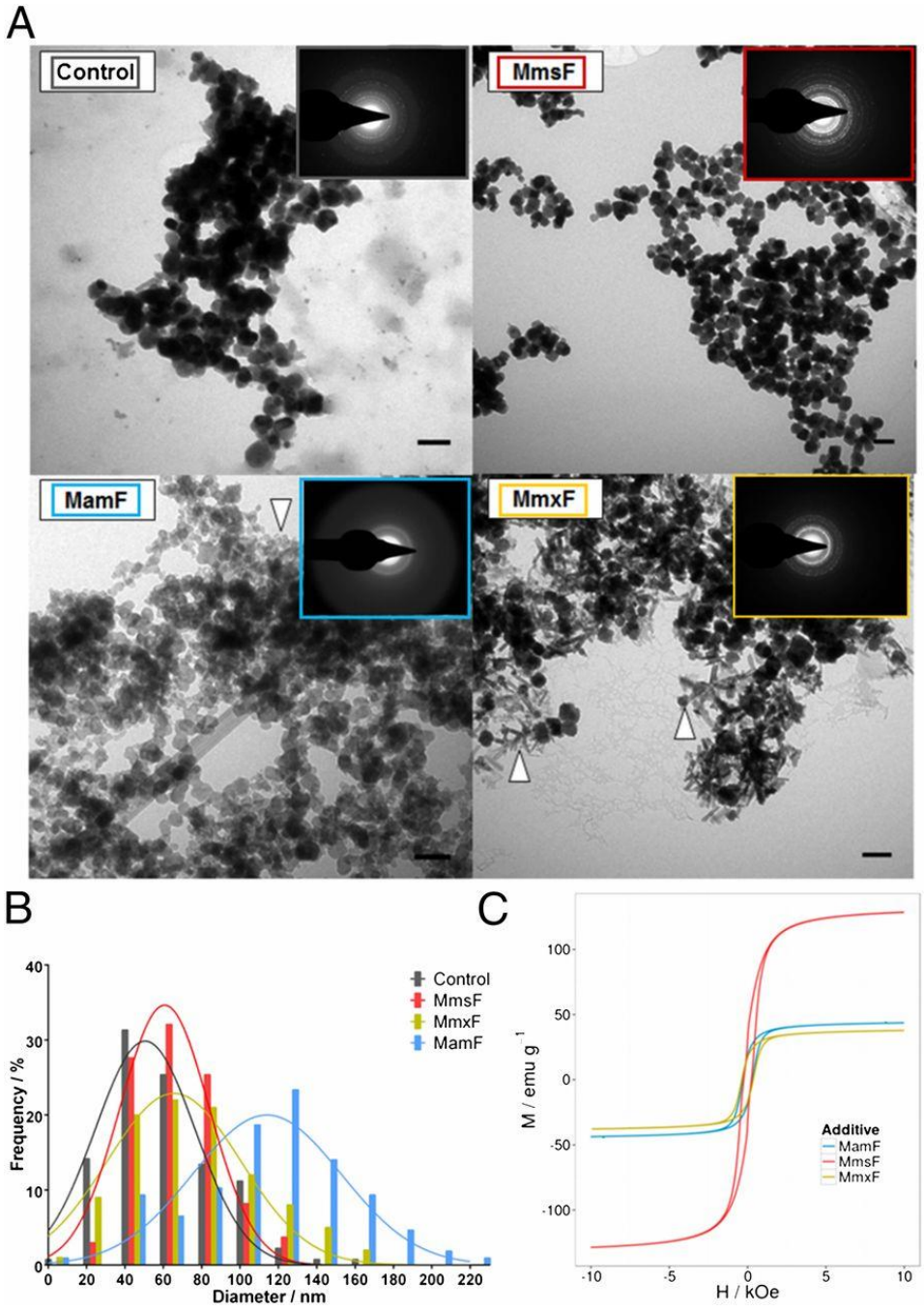


Figure 1.3. (A) Transmission Electron Microscopy (TEM) images of MNPs produced in the absence of protein (control), and in the presence of MmsF, MamF, and MmxF. 100 nm scale bars are included, along with a selected area electron diffraction image (coloured boxes). (B) Histogram of nanoparticle diameter distributions. (C) Vibrating sample magnetometry performed at 298 K. Figure taken from Rawlings *et al,* 2014 with permission.

In this figure, nanoparticle quality is analysed by observing 3 different properties. The first is by measuring nanoparticle diameters, which is displayed in the histogram in panel B. MmsF produces a more homogenous sample of nanoparticles than the negative control, and both non-functioning mutants. Secondly, the selected area electron diffraction images are

included as a measure of the quality of crystal structure constituting the nanoparticles. MmsF nanoparticles gave the most intense and uniform diffraction pattern, indicating that the nanoparticles are ordered and crystalline. Finally, vibrating sample magnetometry measurements (panel C) show MmsF nanoparticles to have the highest magnetisation saturation when compared to homologues. These data support the theory that these proteins are involved in mediating nanoparticle biomineralisation *in vivo*, and that they represent an exciting avenue towards the end goal of customisable, ecologically considerate, batch production of MNPs. Most of these discoveries have been made through genetic manipulation, usually via observing the phenotype of deletion mutants (Grünberg et al., 2004; Katzmann et al., 2010; Murat et al., 2012). Therefore, little is known about how they function mechanistically. However, initial attempts to solve this have been slowed by difficulties encountered when attempting to purify the proteins. Since they are membrane proteins, they experience many of the difficulties classically faced when purifying membrane proteins, such as low yields and poor solubility. Nonetheless, progress has been made. The next section looks in depth at MmsF, which is the focus of this thesis.

1.4 The Magnetosome Membrane Specific protein F (MmsF) is a key regulator of magnetite biomineralisation

The biomineralization protein MmsF was first characterised by Murat *et al* in their 2012 paper (Murat et al., 2012). In this study, they demonstrate that MmsF is a major regulator of magnetite biomineralization in the AMB strain when attempting to identify the genetic cause of the ΔR3 mutant phenotype. This phenotype, initially characterised by the deletion of the mamCDF and mms6 gene clusters (Murat et al., 2010), causes the AMB strain to produce nanoparticles roughly 40% the diameter of wildtype crystals. To investigate this effect, they created a series of mutants which possessed different combinations of the genes in these 2 clusters. The ΔmmsF mutant produced a phenotype almost identical to the initial ΔR3 strain. Interestingly, a mutant strain containing deletions of every gene in the mamCDF and mms6 clusters except for mmsF resulted in nanoparticles that were wildtype in appearance and nearly wildtype in magnetic properties. This study strongly suggests a key role for MmsF in the proper biomineralisation of nanoparticles. Furthermore, by demonstrating the localisation of MmsF to the nanoparticle, they show that it performs this function through

direct interaction with the growing MNP. For these reasons, MmsF has been the focus of further research.

The Staniland group demonstrated that purified MmsF had a positive effect when added to magnetite co-precipitation reactions (Figure 1.3) (Rawlings et al., 2014). Interestingly, MmsF was found to self-assemble into water-soluble structures termed proteinosomes, likely through the semi-ordered burying of the hydrophobic faces of the transmembrane helices. Addition of these proteinosomes to a room-temperature co-precipitation of magnetite improved overall size and sample homogeneity, while giving the nanoparticles a more defined octahedral morphology. This strongly suggests that MmsF directly interacts with the growing nanoparticle to control the biomineralisation process. Furthermore, the assembly and subsequent functionality of these MmsF proteinosomes suggest that MmsF may form higher order structures in the magnetosome, something that is commonly observed in membrane proteins through structures such as membrane rafts. How MmsF would do all these things is unknown, however the sequence of MmsF provides a clue. Figure 1.4 shows the sequence of MmsF, in which a proposed functional loop sequence sits between the predicted transmembrane helices one and two.

Figure 1.4. Sequence of MmsF with predicted transmembrane spanning helices (TMS) annotated. Proposed functional loop of MmsF (DRDDEFVYFHAKQ) is coloured red. Figure adapted from (Rawlings et al., 2014).

The N-terminus of MmsF is located in the cytoplasm, as demonstrated through the fusion of GFP to the N-terminus of MmsF (Murat et al., 2012b). This means it can be determined that the proposed functional loop of MmsF sits in the magnetosome lumen. This is strong evidence that it is responsible for mediated protein-nanoparticle interactions. Furthermore, it consists of many residues typically associated with coordinating metal cations, such

aspartates, glutamates, and histidine. Interestingly, in AMB there exists 2 homologues of MmsF, termed MamF and MmxF. Their loop sequences are shown in Figure 1.5.

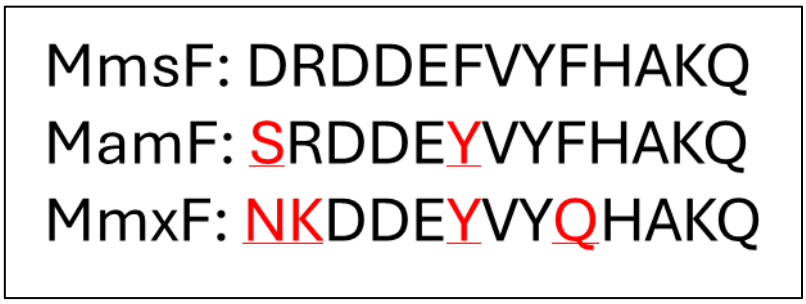


Figure 1.5. Predicted functional loop sequences of MmsF and its non-functional homologues, MamF and MmxF. Residues which differ in the non-functional homologues are highlighted in red. Adapted from (Rawlings et al., 2014).

These homologues were predicted to be non-functional, since they do not rescue the mutant phenotype observed in ΔmmsF mutants (assuming they are expressed). When used in co-precpitation reactions, nanoparticles produced do not match the quality and dimensions of those produced by MmsF (Figure 1.3). This is surprising, since they each have near identical loop sequences when compared to MmsF. Interestingly, the C-terminal HAKQ sequence remains unchanged between the three mutants. This, along with the presence of histidine (a classic example of a metal binding amino-acid), suggests that this sequence is of particular importance. Rawlings *et al* demonstrated in their 2014 paper that purified MamF and MmxF assembled into similar proteinosome structures in solution, yet their addition to co-precipitation reactions yielded poorer quality nanoparticles in comparison to MmsF (Rawlings et al., 2014). This strongly suggests that the residues substituted are key to proper biomineralisation, and that MmsF function is highly sequence specific. To further understand the residue-specific function of this sequence, an artificial coiled-coil peptide display system was developed (Section 1.5) (Rawlings et al., 2019).

1.5 MmsF coiled-coil (MmsFcc) is an artificially designed scaffold protein which displays the functional loop of MmsF

Further functional investigation into MmsF was likely to be hindered by low yields, while the self-assembly into proteosomes would make structural studies challenging. To remedy this, the group looked to develop an artificially designed protein for displaying the loop. Of the potential scaffold systems, coiled-coils proved to be an attractive option. Proteins containing

coiled-coil motifs are ubiquitous in living organisms, where the coiled-coil mostly functions as a key structural component. This structural motif was first proposed by Francis Crick in 1952, as a solution to the X-ray diffraction pattern of α-keratin, which closely matched the diffraction pattern of a typical α-helix, but with a noticeable difference. This difference was explained by the supercoiling of two alpha-helices around each other. Since this, extensive research has shown the widespread adoption of coiled-coil motifs in protein structures. Canonical coiled-coil proteins contain a heptad repeat sequence (ABCDEFG), where residues A and D are conserved hydrophobic residues (Truebestein & Leonard, 2016). Coiled-coils were an attractive candidate for displaying the loop of MmsF, since they are relatively inert for proteins, and could be designed to mimic the predicted distance between MmsF transmembrane helices (9.5 Å). For MmsFcc, the coiled-coil backbone was derived from an artificial sequence designed by Gurnon, Whitaker and Oakley, 2003. In this study, they were able to successful design and express a single α-helix which reliably dimerised to form stable antiparallel coiled-coils in solution. The antiparallel conformation is heavily favoured by ensuring that e−e‘, g−g‘, a−d‘, and d−a‘ interactions are sterically favourable in the antiparallel conformation only. In MmsFcc, the antiparallel conformation is further favoured by the inclusion of the interconnecting loop. A leucine zipper motif is also present in this scaffold, with residues spaced in the following pattern: L-X-X-X-L-X-X-L (with L denoting leucine or isoleucine). This keeps the coiled-coil in the correct register by sterically inhibiting sliding of either alpha helix. Rawlings *et al* termed this new protein MmsF coiled-coil (MmsFcc) (Figure 1.6).

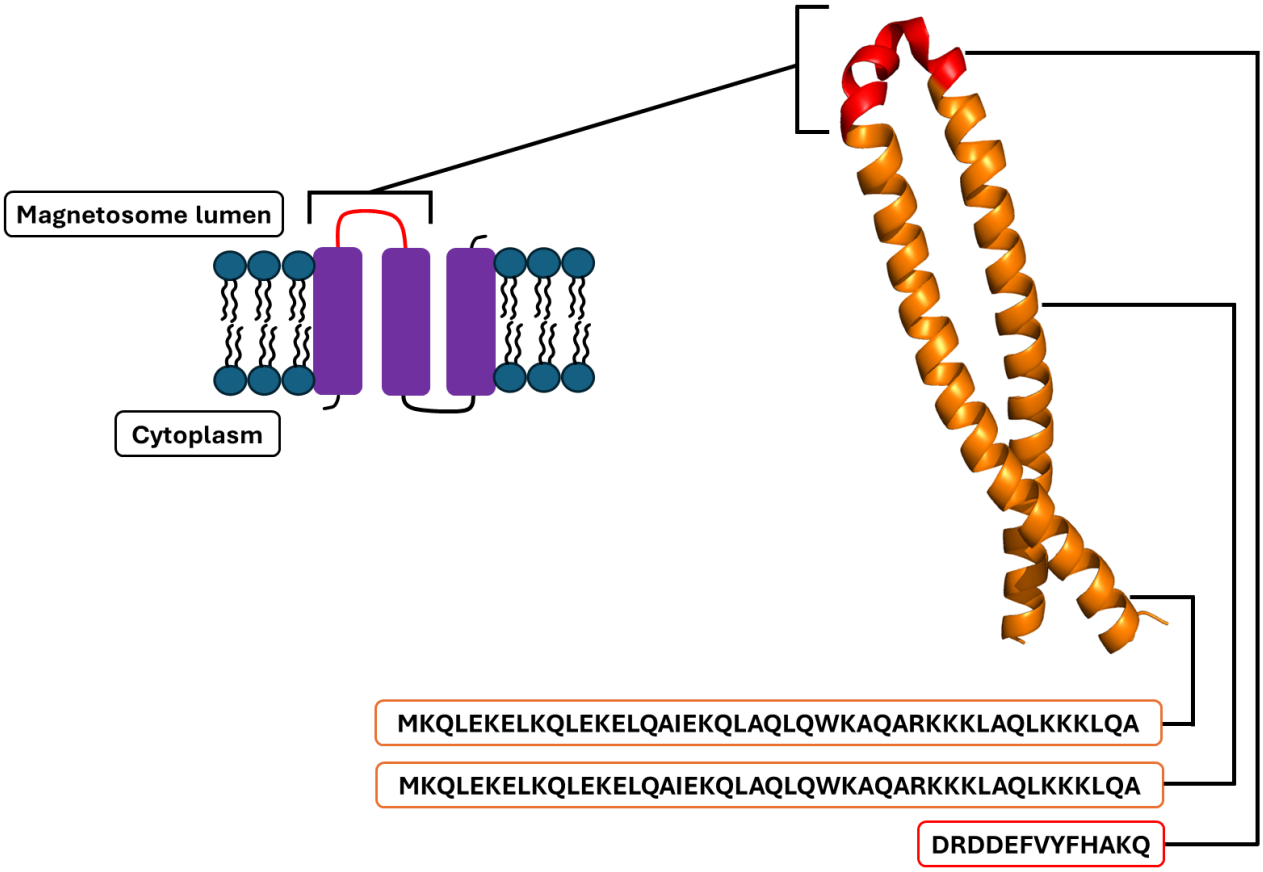


Figure 1.6. Artificially designed coiled-coil scaffold system for displaying the functional loop of MmsF. Here, the functional loop (red) is grafted onto a coiled-coil scaffold (orange), consisting of two identical α-helix sequences. This new protein is called MmsFcc.

The ExPASy ProtParam tool was used to determine protein parameters based on the protein sequence (Gasteiger et al., 2003). MmsFcc is a 107-residue protein with an approximate molecular weight of 12.7 kDa, and a pI of 10.00. An eight-residue histidine tag was attached to the N-terminus of MmsFcc, spaced with a TEV protease recognition sequence. These two features were added to allow for the purification of MmsFcc via HisTag affinity to the Ni^{2+} cation, and the subsequent removal of the tag via cleavage with TEV protease. The full sequence of His-MmsFcc and MmsFcc, and the plasmid map of pPR-1BA1-His-MmsFcc, is provided in the Appendix sections 7.1-7.3. In this system, the loop of MmsF was placed on a coiled-coil scaffold designed to mimic the predicted distance between TMS1 and TMS2 in full-length MmsF (Rawlings et al., 2019).

With protein sequences not created through evolution, there is always some concern that the protein will not behave as intended. Despite this, properly folded MmsFcc was expected to be more water-soluble than MmsF and thus yield greater amount of protein. Rawlings *et*

al were successful in expressing and purifying MmsFcc. They were also able to show, though small angle x-ray scattering and circular dichroism experiments, that a solution of MmsFcc was monomeric and folded into a coiled coil conformation. Furthermore, Rawlings *et al* demonstrated that when MmsFcc was added to co-precipitations, it functioned almost identically to MmsF proteinosomes (Figure 1.7.) (Rawlings et al., 2019).

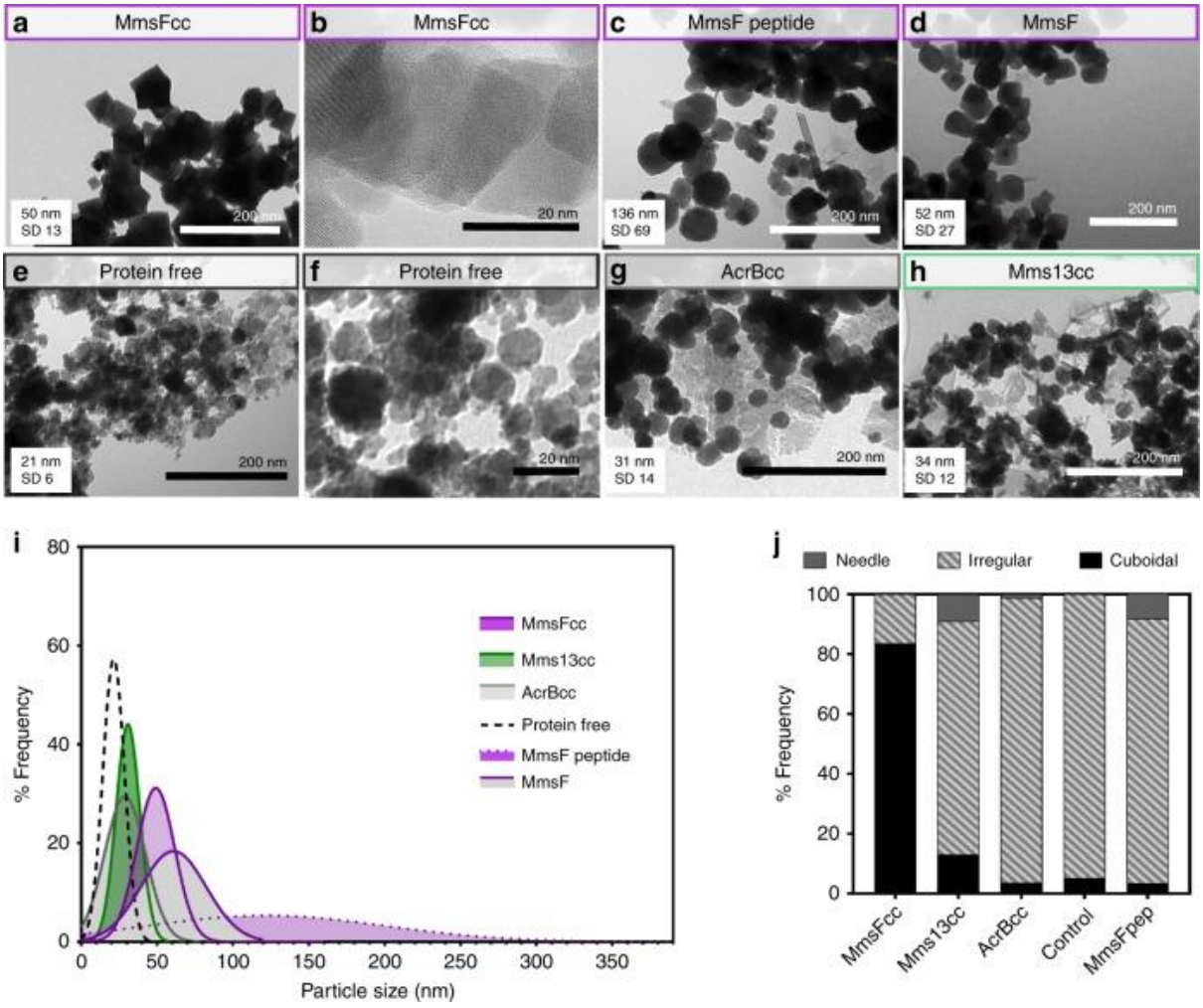


Figure 1.7. Analysis of magnetite nanoparticles produced from the addition of MmsFcc. Panels A-H display TEM images of iron oxide nanoparticles produced by room temperature co-precipitation with the addition of different proteins. Panel I displays the size distribution of nanoparticles in each case. Panel J quantifies the morphology of nanoparticles. MmsF peptide is the 13-resiude loop sequence (DRDDEFVYFHAKQ). AcrBcc is a negative control consisting of a loop derived from *E. coli* acriflavine efflux protein AcrB. Mms13cc is derived from Mms113, which is another biomineralization protein found to co-purify with AMB-1 crystals and increase nanoparticle size in co-precipitations (Arakaki et al., 2003; Valverde-Tercedor et al., 2015). Figure taken from Rawlings *et al,* 2019 with permission.

Some interesting observations can be made from this figure. Firstly, it proves that the loop was responsible for the control over magnetite biomineralisation exhibited by MmsF.

Furthermore, this function is dependent on the restriction of this loop sequence by two alpha-helices, as evidenced by the nanoparticles produced by the free MmsF peptide. This implies that MmsF function is dependent on a structured loop. The fact that the displayed loop mimics MmsF functions suggests that, if the native protein's function requires the formation of higher-order assemblies, then this assembly may be mediated by the loop itself. Crucial to this thesis is the fact that, since nanoparticles produced by MmsFcc were nearly identical to those produced by MmsF, the coiled-coil display system is an appropriate system for studying the loop of MmsF in detail. Of the technologies available to do this, Nuclear Magnetic Resonance (NMR) spectroscopy represented an attractive avenue for doing so and is discussed further in Section 1.6.

1.6 NMR spectroscopy is an attractive tool for structural and functional analysis of MmsFcc

NMR spectroscopy is a widely adopted technology in the structural and functional analysis of proteins (Furukawa et al., 2016; Hu et al., 2021; Kleckner & Foster, 2011; Pellecchia et al., 2002; Puthenveetil & Vinogradova, 2019). The technique utilises the nuclear property of spin to acquire atom-specific information on molecules. A proton or neutron with an unpaired spin has a magnetic moment and can be thought of as acting like a small magnet. NMR spectroscopy works by first aligning these unpaired spins with a strong external magnetic field, termed B_0. A series of electromagnetic pulses are then applied to the sample, which shift the direction of some spins away from the B_0 field. As the magnetic moments relax to B_0, a detectable voltage is produced, which constitutes the signal observed by NMR. One advantage of NMR, over other techniques such as X-ray crystallography and Cryo-electron microscopy, is the ability to observe soluble proteins in their natural state.

While powerful, NMR spectroscopy has some limitations. Firstly, it is limited in the size of molecules that it can observe. This is so because the sensitivity of experiments decreases due to increased tumbling, causing a faster T_2 relaxation rate. This can be somewhat mitigated with higher field-strength magnets, and the use of increased sensitivity cryo-probes. Another limitation of NMR is the requirement for isotopic labelling of proteins. It is not possible to observe nuclei with paired spins, since they have no nuclear magnetic moment. Proteins primarily consist of proton, carbon, and nitrogen, with the most common

naturally occurring isotopes of each being ^{1}H, ^{12}C, and ^{14}N. Both ^{12}C and ^{14}N contain no unpaired spins, and the next most abundant 'NMR active' isotopes (^{13}C and ^{15}N) do not constitute enough of the proteins' make-up to produce a viable NMR signal. This is mitigated by the isotopic labelling of proteins; this is often achieved by growing cells expressing recombinant protein in a medium containing only isotopically labelled sources of carbon and/or nitrogen (examples include ^{13}C glucose/glycerol, ^{15}N ammonium sulphate/chloride) (Wüthrich, 1986). When compared to unlabelled growths in rich media such as LB medium or autoinduction medium, the cost is greater, and the growth potential is lower. This can make the study of low-yielding proteins cost-limited. Nonetheless, NMR spectroscopy has greatly contributed to the field of protein studies. In this thesis, we aim to solve the structure of MmsFcc and analyse its interaction with metals through titration experiments.

A standard approach towards determining the structure of proteins equivalent in size to MmsFcc in solution is listed below:

1. Assign NMR spectra to determine which peaks represent which residues in the protein sequence.
2. Use this assignment to experimentally determine angle and distance restraints.
3. Input distance restraints into molecular dynamic calculations, such as simulated annealing, to determine best approximation of protein structure in solution.

Experimental restraints are experimentally derived distances or angles between certain nuclei, which are used to define how a protein folds in simulations. They can be determined by a variety of phenomena, with the main three being: observation of the Nuclear Overhauser effect (distance) (NOE), Residual dipolar coupling (angle) (RDC), and Paramagnetic Relaxation Enhancement (distance) (PRE) (K. Chen & Tjandra, 2012; Clore & Iwahara, 2009; Hu et al., 2021). Of these three, NOEs are the most commonly used since they provide accurate distance information and are the most abundant of the three while requiring the least experimental setup (Puthenveetil & Vinogradova, 2019). The Overhauser effect was first demonstrated in electrons in 1953 by the American physicist Albert Overhauser, then identified in between nuclei by French physicist Ionel Solomon in 1955 (Overhauser, 1953; Solomon, 1955) and involves the transfer of nuclear spin polarization from one population of spin-active nuclei to another via a process called cross-relaxation. NOEs can be measured to accurately determine the distance between specific nuclei in a

molecule (Hu et al., 2021). For this thesis, experimentally determined NOEs were used to solve the structure of MmsFcc (Chapter 4).

AlphaFold is a rapidly growing tool in the field of protein structural determination. AlphaFold refers to a series of machine-learning based artificial intelligence programmes that have demonstrated an ability to accurately predict protein folding, requiring just the protein's amino acid sequence (Jumper et al., 2021). Large sections of the academic community have been quick to adopt this new technology, for understandable reasons. AlphaFold is free and easy to use and can produce structural predictions that can otherwise take years of expensive experimental work. With regards to MmsFcc, it is still appropriate to solve the structure experimentally for two main reasons. The first, is that it is an artificially designed protein. Since AlphaFold is trained on a library of natural proteins and works by comparing the given sequence to a database of sequences, it is therefore likely that it might struggle to predict artificial sequences accurately (Jumper et al., 2021). Also, this thesis intended to analyse the effect of point mutants on MmsF loop structure. AlphaFold, at least in its current iteration, failed to predict the misfolding effect of missense mutants, implying that it struggles to predict the significance of certain mutations (Buel & Walters, 2022).

NMR spectroscopy is also an excellent tool for assaying the interaction of proteins and ligands in solution. It is able to detect weak affinity interactions, such as those with millimolar K_D values. The most prevalent example of this is in the field of drug discovery, where NMR is initially used as a high-throughput method for screening libraries of ligands for therapeutically targeted proteins; these methods often involve the observation of the ligand one-dimensional proton spectra (Emwas et al., 2020; Pellecchia et al., 2002). It is also possible to perform more detailed studies, through the performance of protein-ligand titrations to determine residue-specific interaction information (Furukawa et al., 2016; Williamson, 2013). This process involves adding an increasing concentration of ligand or protein to a solution containing a fixed concentration of the opposite molecule, then observing ligand or protein spectra to derive information about the mechanism of binding. If these spectra are assigned (meaning it is known which residues different peaks in the spectrum represent, see Chapter 4), it is possible to determine residue-specific information from these experiments. Since MmsFcc is a proposed metal-binding protein, its interaction

with metals was a focus of this thesis. The metals assayed are discussed further in Section 1.8, and the experimental evidence is shown in Chapter 5.

The study of Magnetosome proteins with NMR spectroscopy is a novel field, with only a couple of papers published on Mms6 (Ma et al., 2017; Rawlings et al., 2016). In 2016, Rawlings *et al* assessed Mms6 function via NMR spectroscopy by utilising a truncated construct containing the C-terminal sequence of Mms6. This sequence (KSRDIESAQSDEEVELRDALA) was selected since it resides in the magnetosome lumen and contains acidic residues which are likely to bind to positively charged metal ions. This sequence was shown to have a positive effect on nanoparticles produced by co-precipitation and was shown to possess specificity for ferrous iron (Fe^{2+}) through interaction with the DEEVE motif. Ma *et al* expanded on this finding, by demonstrating that the DEEVE motif interacts with magnetite crystals (Ma et al., 2017).

1.7 Mutants of MmsF derived from non-functional homologues provide clues for the roles of specific residues

One major advantage of the MmsFcc scaffold system when compared to MmsF was that it would be easier to design and express point mutants targeting the functional loop of MmsF, to investigate the importance of specific residues. This is relevant when considering the homologues MamF and MmxF, which are non-functional because of only two or four substitutions respectively in the loop. In unpublished work, Dr Somner developed a series of mutants using the coiled-coil scaffold (Figure 1.8).

MmsFcc (D49S):	SRDDEFVYFHAKQ
MmsFcc (F54Y):	DRDDEYVYFHAKQ
MmsFcc (F57Q):	DRDDEFVYQHAKQ
MmsFcc (F54Y, F57Q):	DRDDEYVYQHAKQ
MmsFcc (R50K, F54Y, F57Q):	DKDDEYVYQHAKQ
MmsFcc:	DRDDEFVYFHAKQ
MamFcc:	SRDDEYVYFHAKQ
MmxFcc:	NKDDEYVYQHAKQ

Figure 1.8. Loop sequences of point mutants derived from the non-functional homologues of MmsFcc. Substituted residues are highlighted in red. Adapted from unpublished data with the authors' permission.

These mutants were designed to bridge the differences between the three sequences, allowing for the importance of each substituted residue to be assayed via chemical co-precipitation reactions (Figure 1.9).

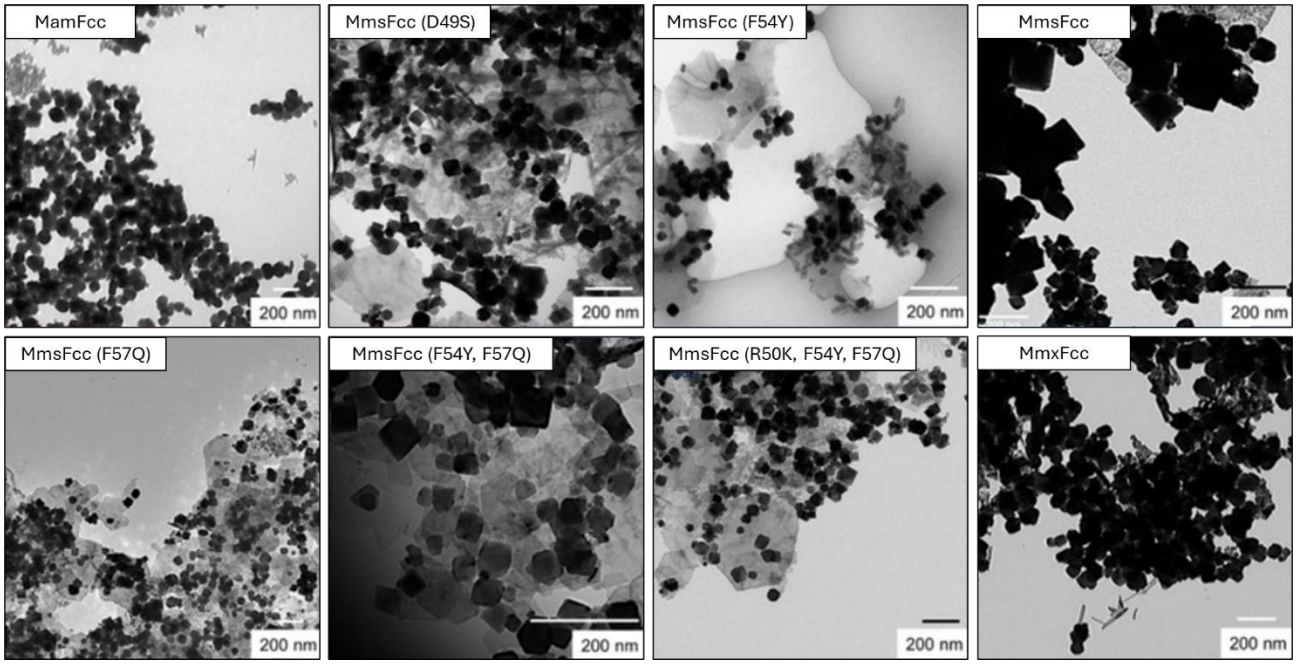


Figure 1.9. TEM images of nanoparticles produced by co-precipitation of magnetite with the addition of MmsFcc and mutant derivatives. Images are at different zoom levels, as visualised by the 200 nm scale bars. Data reproduced from unpublished data with the authors' permission.

Through observation of randomly selected regions from these TEM scans, nanoparticle morphology was quantified to determine mutant viability (Table 1.3).

Protein	Sequence	Size distribution (nm)	Morphology
MamFcc	SRDDEYVYFHAKQ	80.52±44.70	Misshapen
MmsFcc D49S	SRDDEFVYFHAKQ	22.42±18.76	64% cubo-octahedral
MmsFcc F54Y	DRDDEYVYFHAKQ	46.91±18.36	76% cubo-octahedral
MmsFcc	DRDDEFVYFHAKQ	49.22±18.01	86% cubo-octahedral
MmsFcc F57Q	DRDDEFVYQHAKQ	50.72±13.53	Misshapen
MmsFcc F54Y, F57Q	DRDDEYVYQHAKQ	60.45±19.07	77% cubo-octahedral
MmsFcc R50K, F54Y, F57Q	DKDDEYVYQHAKQ	28.73±21.40	58% cubo-octahedral
MmxFcc	NKDDEYVQFHAKQ	70.55±40.69	Misshapen

Table 1.3. Analysis of Figure 1.9 nanoparticle morphology. Morphology of nanoparticles were judged and tallied by original author. Data reproduced from unpublished data with authors' permission.

Since MmsFcc produces nanoparticles of cubo-octahedral morphology, the functionality of these mutants can be classified based on the ratio of cubo-octahedral : misshapen nanoparticles that they produce. Using this metric, F54Y and (F54Y, F57Q) can both be defined as 'functional'. D49S and (R50K, F54Y, F57Q) are both classified as having reduced function. This is so as while they are able to produce nanoparticles with the correct morphology, the fraction of cubo-octahedral nanoparticles is smaller and the nanoparticles themselves have a smaller size distribution than wildtype (~50 nm average for wildtype compared to ~25 nm). Finally, MamFcc, MmxFcc, and F57Q each produce misshapen nanoparticles, defining them as non-functional. From these results, two mutants were highlighted as being of particular interest. The first, F57Q, was a non-functional mutant. The second, (F54Y, F57Q), rescues this loss-of-function through the substitution of phenylalanine to tyrosine. Since these residues do not typically bind metals through their sidechains, it is instead likely that a conformational change is responsible for the reduced function observed. Another observation is that, in all cases, magnetite crystals were created with greater size than those produced in absence of protein. This implies that, despite losing control of nanoparticle morphology, the presence of these proteins still promoted nanoparticle growth. In the case of MamFcc and MmxFcc, the average diameter of nanoparticles produced is greater than wildtype, suggesting that the wildtype sequence functions to

restrict the excessive growth of the magnetite crystal. This is relevant when considering that the desirable property of superparamagnetism, especially in relation to human application of synthesised nanoparticles, is reliant on the nanoparticle being a sufficiently small size (Wallyn et al., 2019; Wei et al., 2017).

1.8 The interaction of proteins and metals is diverse, and an essential component of life

Many proteins have evolved to interact with metals. Approximately one third of reported structures contain a metal cation, which is essential for their correct function (Lu et al., 2012). For some of these proteins, such as zinc fingers, this incorporation serves to stabilise the structure of the protein (Cassandri et al., 2017a). In the case of metal co-factors in metalloenzymes, the cation is incorporated into the active site where it catalyses enzymatic activity (although it could be argued in some cases that these metals serve both a catalytic and structural function) (A. Y. Chen et al., 2019). The interaction between protein and metal can be transient, with allosteric control of protein function being widely reported (Baksh & Zamble, 2020; Reyes-Caballero et al., 2011). However, because metal cations react readily, excessively large populations of free metals in cells can have deleterious effects. For this reason, control of metal populations by organisms is an essential process and a tricky balancing act. To mediate this, proteins have evolved to interact, transport, and sequester metal ions. Clearly then, the affinity between protein and ligand (typically measured by calculating K_D), can vary greatly. For protein-metal interactions, a K_D in the millimolar range is considered to represent 'weak binding' (Wang et al., 2009). In the case of MmsF, we hypothesised that its interaction with iron would be short lived, given its predicted activity inside the magnetosome. Furthermore, MmsF function would likely not be inhibited by a perceived weaker binding, due to the high concentration of iron atoms present inside the magnetosome. We predict that MmsF functions by presenting iron (either Fe^{2+}, Fe^{3+}, or both), in the correct position to be properly incorporated during magnetite biomineralization. For this reason, both Fe^{2+} and Fe^{3+} were tested for their interaction with MmsFcc.

Two other metal ions, Zn^{2+} and Ca^{2+}, were tested for their interaction with MmsFcc. Calcium and zinc have been included to assess the specificity of MmsF, whilst also providing

alternative divalent cations in the case that the paramagnetism of Fe^{2+} was problematic for NMR experiments. Paramagnetic components in solution NMR can cause the loss of signal due to line broadening as a result of paramagnetic relaxation enhancement (a phenomenon which can actually be desirable when determining the structures of certain proteins) (Clore & Iwahara, 2009). Two major factors were considered when determining which metals to select for screening. The first was the co-ordinate complexes that said metal commonly forms during protein-metal interactions. Secondly, the preference of binding partner was considered. This can be conceptualised by the concept of 'hard' and 'soft' ions. This system places ions on a scale of 'hardness', with harder ions having a larger charge, smaller ionic radius, and low potential for polarisation (Pearson, 1968). Because of this, hard ions preferentially bind other hard ions in an event similar to an ionic bond. Conversely, soft ions have a smaller charge, bigger ionic radius, and strong potential for polarisation. Therefore, they prefer to bind other soft ions in an event similar to covalent bonding. Common soft binding partners in proteins are nitrogen in amide groups and sulphurs in thiols, while common hard binding partners are oxygens in carboxylate. The preferred co-ordination complexes and binding partners of the four mentioned cations are discussed next.

Iron is the second most abundant metal in the earth's crust and exists in two main redox states in solution (Hans Wedepohl, 1995). Its ability to readily convert between these two redox states, Fe^{2+} and Fe^{3+}, makes it useful to organisms when assisting acid-base and electron transfer reactions. Whist essential for life, iron must be carefully regulated by cells since both redox states have toxic properties (Sánchez et al., 2017). Fe^{2+} is a ready electron donor and is therefore toxic, in particular through catalysing the Fenton reaction and the Haber-Weiss reaction resulting in increased oxidative stress (A. M. Koorts and Viljoen, 2007). Fe^{3+} on the other hand readily forms iron hydroxide precipitates in solution, which can also be toxic to cells. For these reasons, organisms have evolved complex systems of iron uptake, sequestering, transport, and application. Both Fe^{2+} and Fe^{3+} commonly form six-coordinate octahedral complexes but have also been shown to form four-coordinate tetrahedral complexes (Eaton et al., 1971; Johnson et al., 2005). The coordination chemistry displayed in these complexes, however, is remarkable varied and changeable. Both redox states can interchange between having bound and unbound coordination sites and can accept different ligands. Different ligands alter the redox potential and radius of the iron atom, allowing for

precise tailoring of these factors to fit the job required (Sánchez et al., 2017). Generally speaking, the two main redox states do have preferred binding partners. Fe^{3+}, due to its hardness, prefers oxygen coordinators such as carboxylate (aspartate/glutamate) and phenolate (tyrosine). Fe^{2+} on the other hand prefers nitrogen donors such as histidine, and sulphur ligands such as methionine and cysteine (Pearson, 1963).

-

Perhaps the most well-known example of iron in biology is in the oxygen-transport protein, haemoglobin. Human haemoglobin is a tetrameric molecule consisting of two α-globin and two β-globin monomers. These units assemble to produce 4 oxygen binding sites, consisting of heme groups (Figure 1.10).

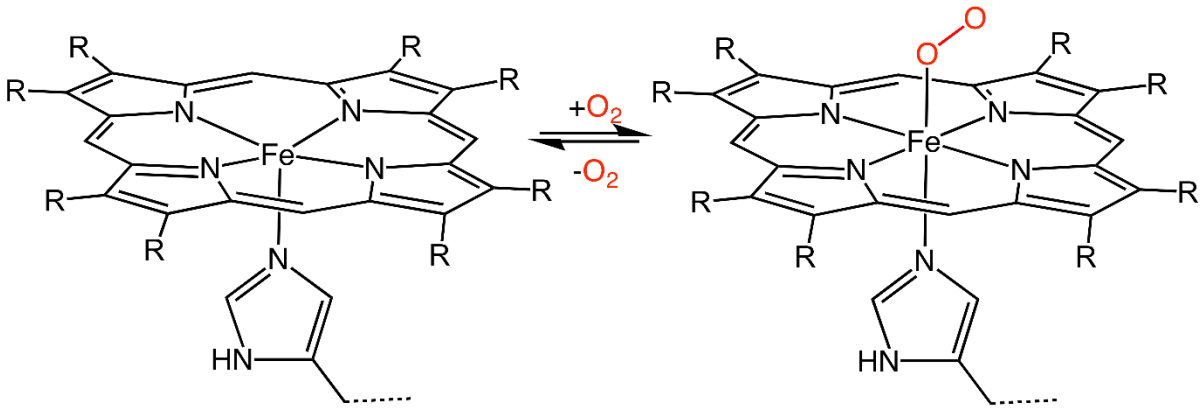


Figure 1.10. Schematic of haem group structure, demonstrating O_2 coordination to the free axial site of iron. By Smokefoot - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=63789703.

In haem, iron has formed an octahedral co-ordination complex with four pyrrole rings positioned in a plane, with two axial binding sites present. In humans, the first is occupied by a histidine sidechain and the second is where oxygen binds. It is also a great example of how the variable redox potential and size of the iron cofactor allows it to have a great effect on protein structure and function. In a haemoglobin molecule with no oxygen bound, the larger radius of the Fe^{2+} causes distortion of the coordinating tetrapyrrole structure. Upon binding of the initial oxygen, the oxygenation to Fe^{3+} reduces the radius and stops warping the ring. This change shifts the axial histidine ring 0.4 Å closer to the complex, resulting in allosteric changes to the conformation of the other haem groups in the molecule, which result in an increased affinity for O_2 at the other haem groups (Collman et al., 2004; Shikama, 2006).

Another classic example of iron binding is the iron storage protein, ferritin. Ferritin is an ubiquitous and highly conserved protein found in almost all living organisms, where it functions as an iron storage protein. In humans, ferritin stores large amounts of iron (up to 4500 atoms per ferritin 24-mer) in an enclosed cavity as ferrihydrite crystals (Koorts & Viljoen, 2007b). In this system, ferritin effectively buffers free iron concentration via controlled sequestering in response to intracellular iron levels. Ferritin structure resembles a large shell formed by 24 symmetrically arranged protein subunits, the H-subunit and the L-subunit (L-subunit not found in bacteria/archaea), with 2 types of channels facilitating the transport of iron, chelators, and oxidants/reductants (Ebrahimi et al., 2015). Inside the 80 Å diameter cavity, iron sequestration begins via the controlled oxidation of Fe^{2+} to Fe^{3+} in the presence of oxygen by the ferroxidase centre of the H-subunit This produces insoluble ferrihydrite, which can be released from storage (although the ability to do so, and mechanisms driving this release vary between organisms) (Sudarev et al., 2023). Eukaryotic cells primarily release iron from ferritin through the action of lysosomes (Kidane et al., 2006)

-

Zinc is another biologically relevant metal. Unlike iron, it only exists in the Zn^{2+} state and is completely redox inert. Furthermore, Zn^{2+} has a similar charge and near identical ionic radius as Fe^{2+}, making it a suitable candidate for analysis. Historically, zinc was thought to only bind proteins in a permanent fashion, as structural components of zinc fingers or in the active site of metalloenzymes (Maret & Li, 2009). However, further research into zinc proteins has shown regulatory roles for zinc, requiring a transient binding event (Patzer & Hantke, 2000; VanZile et al., 2002). Zinc primarily forms tetrahedral complexes, like those seen in Zinc finger proteins (Cassandri et al., 2017; Sénèque & Latour, 2010). For catalytic purposes, zinc is used primarily for its properties as a strong Lewis acid (Maret, 2013). As mentioned earlier, zinc finger proteins are an extensively studied family of proteins, whose structure is reliant on the proper co-ordination of one or two zinc cations. The co-ordination complex formed is always tetrahedral, and involves either four cysteines, three cysteines and a histidine, or two cysteines and two histidines (Figure 1.11).

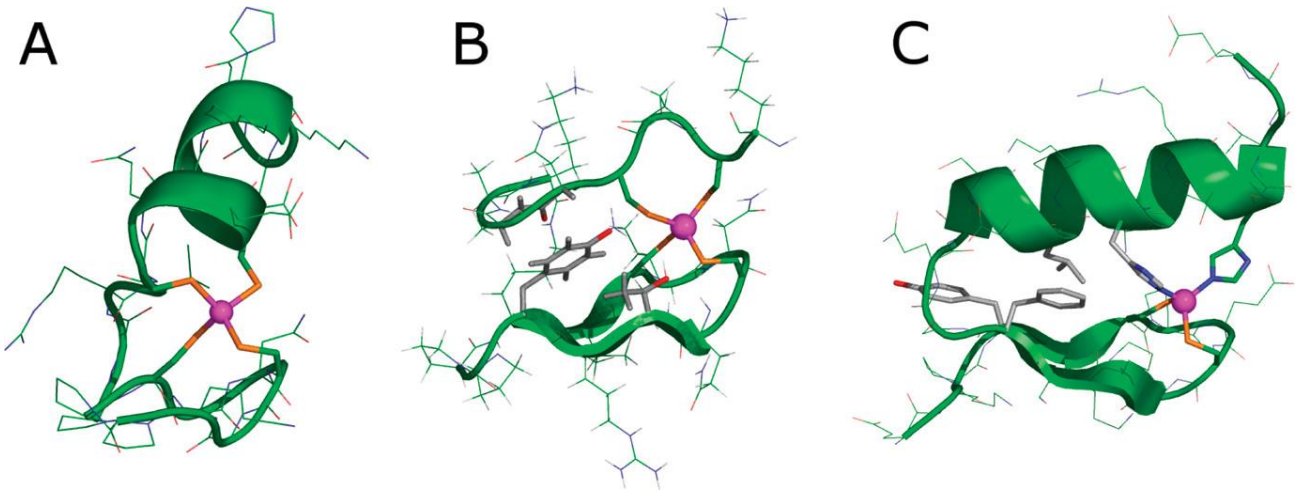


Figure 1.11. Tetrahedral co-ordination of zinc (purple) by three zinc finger proteins. A and B demonstrate co-ordination by cysteines solely. C demonstrates co-ordination by two cysteines and two histidines. Figure taken from (Sénèque & Latour, 2010).

-

Calcium is also considered essential for life. It is of particular interest to this study since it is involved in the most well-known examples of biomineralization, the formation of bones. The mineral component of bones is called Hydroxyapatite, with the chemical formula $Ca_{10}(PO_4)_6(OH)_2$. The intracellular crystallisation of Ca^{2+} and inorganic phosphate (Pi) is a spontaneous process, which is actually inhibited in most cells. Osteoblasts overcome this to mediate the nucleation and growth of hydroxyapatite, although rather surprisingly a lot is unknown about this. Relevant to this thesis is the formation of membrane-enclosed Ca-P particles called matrix vesicles. These, like the magnetosome, provide a locally concentrated microenvironment for hydroxyapatite mineralisation. However, unlike magnetosomes, the biomineralisation process appears be quite different. Embedded proteins and phospholipids control biomineralization primarily by mediating nucleation and the sudden influx of phosphorus. The crystal is thought to grow until it pierces the vesicle, where it is deposited into extracellular bone matrix (Ansari et al., 2021; Beck, 2019). It is unclear whether proteins, like MmsF, influence biomineralisation through direct interaction with the growing crystal. It should be stated that this is just one of many components that forms complex bone organs, which is much more complex than nanoparticle biomineralization.

Calcium biomineralization also occurs during the formation of calcium carbonate-based shells, commonly seen in invertebrate species such as molluscs. In the case of molluscs, crystalline and amorphous calcite is deposited into the growing shell, the process of which is

mediated by a series of enzymes and shell matrix proteins. Similarly to the system observed in magnetosomes, these proteins possess a large degree of control over the crystallisation process by mediating nucleation and crystal growth. The matrix itself serves as a platform for the displaying of relevant proteins, similar in function perhaps to the magnetosome membrane (Marin et al., 2007). Differences in expression of these proteins largely contributes to the range of morphologies observed in molluscs (Yarra et al., 2021).

In a more general sense, calcium is found in solution in the Ca^{2+} oxidation state. It possesses a varied co-ordination chemistry, being able to form complexes with co-ordination numbers ranging from two to eight, although it is mostly found in six to eight co-ordinate complexes. Ca^{2+} is a harder ion than Zn^{2+} and Fe^{2+}, which prefers to be co-ordinated by oxygens such as those found in carboxylate groups (Katz et al., 1996a).

1.9. A conceptual model for how MmsFcc could mediate nanoparticle formation

In the absence of MmsF, the co-precipitation of iron oxides from a ferrous/ferric solution produces an array of disordered iron oxides and iron hydroxides. The simple addition of MmsF (or MmsFcc) greatly increases the ratio of crystalline magnetite (Fe_3O_4) formed. From the currently available body of research, it is not clear how MmsF could control the co-precipitation to this degree. It is known that MmsF is found tightly bound to purified magnetite crystals, strongly suggesting that its function is carried out via direct interaction with the growing crystal (Murat et al., 2012). It is also known that this function is largely the result of the thirteen-residue loop which presides in the magnetosome lumen, as demonstrated by the research carried out with MmsFcc (Rawlings et al., 2019).

Of the two iron ions, Fe^{3+} is more insoluble. This means Fe^{3+} will readily incorporate into insoluble products of co-precipitation faster than Fe^{2+}. This favours the formation of ferric products and skews the ratios of Fe^{2+}/Fe^{3+} required for magnetite formation. One of the ways that MmsFcc may control magnetite formation is to improve the incorporation of Fe^{2+} into the growing crystal. This may be achieved by directly binding the ion, creating a local concentration greater than the whole magnetosome where Fe^{2+} incorporation is required.

MmsF also improve the homogeneity of nanoparticle sizes produced by co-precipitation. This may be achieved by the direct interaction of MmsF and the nanoparticle, which again is evidenced by tightly bound protein in purified crystals (Murat et al., 2012). MmsF, either through a charge or shape interaction, could bind to properly formed faces of the crystal. This would halt excessive growth at the site of interaction, by blocking the incorporation of further ions.

1.10. Project aims

This project aimed to utilise the MmsFcc construct to perform an in depth functional and structural analysis of MmsF and its non-functional homologues. The first step to achieving this was to establish an efficient expression strategy, using cultures of *E. coli* grown in isotopically labelled minimal medium. Next was to optimise the purification of MmsFcc, utilising nickel column affinity to the N-terminal HisTag of MmsFcc. A backbone assignment of the MmsFcc ^{15}N HSQC spectrum would then be performed, which would allow for the assessment of MmsFcc-metal titrations through the observed changes to amide chemical shift values. A backbone assignment would also allow for the completion of the sidechain assignment of the MmsFcc ^{13}C HSQC spectrum, which we predicted would be necessary for elucidating the structure of MmsFcc. To do this, NOEs would be determined from various HSQC-NOESY spectra, then used to provide distance restrains in simulated annealing calculations. If possible, a bound structure of MmsFcc would also be elucidated to more accurately determine MmsFcc-metal interaction. This process would then be repeated on the mutants of interest (F57Q and F54Y/F57Q) primarily, then if possible other mutants would be examined. Unfortunately, due to unforeseen difficulties expressing and purifying the protein, not all of these aims were achieved. Chapter three describes the process of optimising the express and purification of the protein. Chapter four describes the backbone assignments of MmsFcc and MmsFcc D49S, and the sidechain assignment of the MmsFcc functional loop. Chapter four then details the NOE-based structural calculation of the MmsFcc loop, with additional restraints derived from AlphaFold predictions used to define the structure of the coiled-coil (alongside experimentally derived TALOS dihedral restraints). Chapter five then shows the results obtained from the titrations of Fe^{2+}, Fe^{3+}, Zn^{2+}, and Ca^{2+} against MmsFcc and MmsFcc D49S. Chapter six then brings these experimental results together into a more general discussion.

Chapter 2. Materials and Methods

2.1. Recipes

Below is a list of commonly used recipes for this project.

Ampicillin stock

For all MmsFcc plasmids, the antibiotic Ampicillin was at a working concentration of 100 µg/mL. A 1000x stock of ampicillin was prepared by mixing 0.1 g of ampicillin sodium salt in 8 mL of Milli-Q H_2O. Once fully dissolved, the volume was topped up to 10 mL then filter sterilised with a 0.2 µm syringe filter. This was then aliquoted into 1 mL aliquots and stored at -20°C. If thawed, excess ampicillin was discarded.

LB medium

LB medium was prepared following the recipe in Table 2.1, then sterilised by autoclave. If required, ampicillin was added once the medium had cooled to room temperature.

Reagent	**Amount**
Tryptone	10g
NaCl	10g
Yeast extract	5g
Medium was made to 1 L final volume with Milli-Q grade H_2O	

Table 2.1. LB medium recipe.

Agar plates

Agar plates were prepared using Thermo Scientific™ Nutrient agar; 28 g of powder was added per 1 L distilled water. The solution was then sterilised by autoclave then cooled to 50°C in a water bath. Once cooled, antibiotics were added then poured into petri dishes.

M9 minimal medium

M9 salts were prepared to the following recipe (Table 2.2).

Reagent	**Amount**
Na_2HPO_4	6g

KH_2PO_4	3g
NaCl	0.5g
Medium was made up to a volume of 950 mL, then adjusted to pH 7.4. After pH adjustment, volume was then topped up to 1 L then sterilised by autoclaving.	

Table 2.2. M9 salts recipe.

Before use, M9 minimal medium was prepared by adding the M9 components (Table 2.3) to 1 L of M9 salts (Table 2.2). Ammonium chloride/sulphate and/or glucose was substituted with the same isotopically labelled compound when required.

Reagent	**Amount**
Trace elements (Table 2.4)	650 µL
Glucose 0.1 g/mL (filter sterilised)	2 g
Ammonium chloride or ammonium sulphate 0.1 g/mL (filter sterilised)	1 g
Thiamine 100 mg/ml (filter sterilised)	1 mL
$MgSO_4$ 1 M (autoclave sterilised)	1 mL
$CaCl_2$ 0.1 M (autoclave sterilised) (was added last, and the solution mixed instantly to avoid precipitation)	1 mL

Table 2.3. Additives for 1 L of M9 minimal media.

Reagents	**Amount**
$CaCl_2.2H_2O$	550mg
$MnSO_4.H_2O$	140mg
$CuSO_4.5H_2O$	40mg
$ZnSO_4.7H_2O$	220mg
$CoCl_2.6H_2O$	45mg
$Na_2MoO4.2H_2O$	26mg
H_3Bo_4	40mg
KI	26mg
Solution was made to 70 mL with distilled water, then adjusted to pH 8.0. After adjusting, 500 mg of EDTA was added then the pH adjusted to 8.0. 375 mg of $FeSO_4$.7H2O was then added, before making the solution to a final volume of 100 mL. Solution was then sterilised by autoclave.	

Table 2.4. Recipe for 100 mL of trace elements.

18% SDS-PAGE gels

SDS-PAGE gels were prepared following the recipe below. First, stock solutions of upper and lower buffers were made (Tables 2.5-2.6).

Reagents	**Final concentration**
Tris	1.5 M
SDS	0.4% (w/v)
Buffer was made up to 4/5^{ths} of final volume, then adjusted to pH 8.0. Buffer was then topped up to full volume then filter sterilised.	

Table 2.5. Upper buffer recipe.

Reagents	**Final concentration**
Tris	0.5 M
SDS	0.4% (w/v)
Buffer was made up to 4/5^{ths} of final volume, then adjusted to pH 6.8. Buffer was then topped up to full volume then filter sterilised.	

Table 2.6. Lower buffer recipe.

Stacking and resolving buffers were then made up to the following recipes (Tables 2.7-2.8). Solutions were mixed gently to avoid incorrect formation of acrylamide matrix.

Reagents	**Volume**
Upper buffer (table 2.5)	2.5 mL
40% bis/acrylamide	1.1 mL
dH_2O	6.4 mL
10% Ammonium persulfate	110 µL

Table 2.7. Stacking buffer recipe.

Reagents	**Volume**
Lower buffer (table 2.6)	2.5 mL
40% bis/acrylamide	4.5 mL

dH_2O	3 mL
10% Ammonium persulfate	100 µL

Table 2.8. Resolving buffer recipe.

Gels were set using the bio-RAD Mini-PROTEAN® gel making kit. Once the cassette was set up, 10 µL of TEMED was added to the resolving buffer to induce polymerisation. The resolving buffer was then poured into the cassette, filling up 4/5 of the cassette. A small layer of isopropanol was added on top, then left for 30 minutes to set. Once set, the isopropanol was poured off then 10 µL of TEMED was added to the stacking buffer. This was then poured into the cassette until full, then a 15-well comb inserted, and the gel allowed to set for 30 minutes. Gels were stored for up to 2 weeks at 4°C by wrapping in wet paper roll and tinfoil.

2.2 Preparation of chemically competent *E. coli*

Cells were prepared for transformation by treatment with rubidium chloride. The following buffers were used (Tables 2.9-2.10).

Reagent	**Final concentration**
RbCl	100mM
$MnCl_2$-4.H_2O	50 mM
Potassium acetate	30mM
$CaCl_2$-2.H_2O	10mM
Glycerol	15% (w/v)
Components were dissolved in distilled water, then filtered sterilised. The buffer was stored at 4°C.	

Table 2.9. TFB1 buffer recipe.

Reagent	**Final concentration**
0.2M MOPS pH 6.5	10mM
RbCl	10mM
$CaCl_2$-2.H_2O	75mM
Glycerol	15% (w/v)
Components were dissolved in distilled water, then adjusted to pH 6.5 before being filter sterilised. This buffer was stored at 4°C.	

Table 2.10. TFB2 buffer recipe.

Cells from a glycerol stock were streaked on an agar plate, which was incubated at 37°C overnight, to get a single colony. A 10 mL LB starter culture was prepared and inoculated with a single colony; this starter culture was incubated overnight at 37°C while rotating. This culture was used to inoculate 500 mL of LB, which was then incubated at 37°C while rotating at 200 rpm, until an OD 600 of 0.5 was reached. The culture was then cooled on ice for fifteen minutes, then harvested by centrifugation at 3000 rpm, 4°C for 10 minutes. The supernatant was discarded, and the pellet resuspended in 30 mL of TFB1 buffer. The cells were then incubated on ice for 90 minutes then spun down at 4000 RPM, 4°C for five minutes. The supernatant was discarded, and the pellet resuspended in 6 mL of TFB2 buffer on ice. 50-200 µl aliquots were flash frozen in liquid nitrogen. These competent cells were stored for up to one year at -80°C.

2.3 Transformation of chemically competent *E. coli*

MmsFcc and MmsFcc mutant plasmids were transformed into chemically competent *E. coli* cells using the same protocol. For protein expression, BL21 DE3 cells were used. For plasmid miniprepping, XL-10 gold cells were used. Both strains have no antibiotic resistance. All MmsFcc plasmids contained an ampicillin resistance gene, so 100 µg/mL ampicillin was used for selection. The protocol used is described below.

Competent cells and plasmid DNA were thawed on ice, then 1 µL of plasmid DNA was added to 50 µL of cells and gently mixed. After mixing, the cells were chilled on ice for ten minutes. Cells were then heat shocked at 42°C for two minutes. After heat shocking, cells were immediately transferred to ice and left for two minutes. 200 µL of LB medium was added, then the mix incubated at 37°C for one hour while rotating. 50 µL of cells were spread onto a 100 µg/ml ampicillin agar plate for selection, the plates were incubated overnight at 37°C.

2.4 Miniprepping of plasmid DNA and sanger sequencing

Plasmid DNA was miniprepped using the Monarch plasmid miniprep kit. A full protocol can be found on the NEB website (https://www.neb.com/en-sg/-/media/nebus/files/manuals/manualt1010.pdf?rev=e79c527b32dd4c7f823db7a05ce6ad62&hash=43C57F3E6F44121123B7C44D529EBB7B).

All sequencing was performed by GENEWIZ, inc. Plasmid DNA concentration was calculated using a NanoDrop™ One UV-Vis spectrophotometer. The T7 promoter forward primer was used (TAATACGACTCACTATAGGG). The T7 reverse primer (GCTAGTTATTGCTCAGCGG) was initially trialled but did not amplify when used. For this reason, the T7 forward promoter was deemed sufficient for sequencing.

2.5 Small scale colonies for preparation of analytical SDS-PAGE samples

Colonies of 5-50 mL were used for trialling growth conditions. For these colonies, a 1:100 volume ratio of antibiotic-containing starter culture was used. For example, with a 10 mL culture, a 10 µL starter culture was used. Antibiotic-containing cultures were then grown at 37°C while rotating until OD600 of 0.6 was reached. The cells were then induced with IPTG and grown in the test conditions. Cells were harvested by centrifugation at 8000 xg for 10 minutes. Pellets were resuspended in sonication buffer at a 5% (w/v) ratio (Table 2.11).

Reagents	**Final concentration**
Tris	25 mM
NaCl	100 mM
NaN_3	3 mM
Buffer components were dissolved in 4/5 final volume of distilled water. The pH was then adjusted to 7, then topped up to final volume. The buffer was filter sterilised.	

Table 2.11. Sonication buffer recipe.

These samples were then sonicated in a sonication bath for 30 minutes at full power, mixing the sample regularly. If needed, soluble and insoluble fractions were separated by centrifugation at 15,000 xg for 10 minutes. The insoluble pellet was resuspended in the same volume of sonication buffer used previously. These samples were stored for up to 1 month at -20°C.

2.6 SDS-PAGE

4 µL of Bio-Rad Precision plus dual colour protein standards were loaded into SDS-PAGE gels to provide a protein ladder. All SDS-PAGE experiments were performed in Bio-Rad vertical electrophoresis cells following manufacturer guidelines. The following recipes were used for SDS running buffer and SDS loading buffer respectively (Table 2.12-2.13).

Reagents	**Amount**
1 M Tris pH 6.5	1.5 mL
Sodium dodecyl sulphate	0.6 g
Bromophenol blue	0.03 g
Glycerol	2.4 mL
Buffer components were mixed with 5 mL of Milli-Q H_2O, then adjusted pH to 6.5. Volume was then topped up to 7.5 mL then stored at -20°C in 0.5 mL aliquots.	

Table 2.12. 4x SDS loading buffer recipe for 7.5 mL.

Reagents	**Concentration**
Tris	250 mM
Glycine	1.92 M
SDS	10 % (w/v)
Buffer components were mixed with four-fifths the final volume of Milli-Q H_2O, then adjusted pH to 8.3. Afterwards, the solution was topped up to final volume.	

Table 2.13. 10x SDS-PAGE running buffer recipe.

10x SDS running buffer was diluted to 1x with distilled water, then poured into the electrophoresis cell. For sample preparation one part of 4x SDS loading buffer was added to three parts of sample, then the mixture was denatured by heating at 95°C for 5 minutes. For high quality gels, the electrophoresis was run at 100V, until the dye ran off the bottom of the gel. Early on in the project, gels were run at 180V for up to 1 hour, although this could overheat the sample and affect the quality of gel produced.

To visualise protein bands, gels were stained for up to 1 hour in READYBLUE™ protein stain, then washed overnight in distilled water.

2.7. Optimal expression and purification protocol

The following is the finalised protocol for purifying MmsFcc and its mutants from a 500 mL culture.

Day 1

For all growth steps, a working concentration of 100 µg/mL ampicillin was used. Chemically competent BL21 DE3 cells were transformed (Protocol 2.3), then incubated overnight at 37°C.

Day 2

10 mL of LB and 20 mL of M9 starter culture was prepared per 500 mL of culture required (Protocol 2.1). These starter cultures were inoculated with cells from the same colony, to maintain genetic homogeneity. These starter cultures were incubated overnight at 37°C while rotating at 200 rpm. 2L baffled flasks containing 500 mL of M9 minimal media were also prepared, then incubated overnight at 37°C while rotating at 200 rpm. This was done to heat the growth medium to 37°C, allowing for the instant addition of starter culture at the start of day 3.

Day 3

Cultures were inoculated with both the M9 and LB starter cultures. The OD600 was checked two hours post inoculation, then every subsequent hour until an OD600 of 0.6 was reached. Once reached, ITPG was added to a final concentration of 1 µM. The cultures were then incubated at 30°C overnight while rotating at 200 rpm.

Day 4

The next day, cells were pelleted through centrifugation, by centrifuging at 8000 rpm at 4°C, for fifteen minutes using a F500 rotor. If pausing, the pellet was stored for up to 2 weeks at -80°C. It was found to be essential to use GuHCl with a purity of at least 99.5%. The following buffers were used for the purification (Tables 2.14-2.18).

Reagents	**Final concentration**
Tris	25 mM
NaCl	100 mM
NaN_3	3 mM
Buffer components were dissolved in 4/5 final volume of distilled water. The pH was then adjusted to 7, then topped up to final volume. The buffer was filter sterilised. On	

the day of purification, a SIGMAFAST™ Protease Inhibitor Tablets was dissolved in a 50 mL aliquot.

Table 2.14. Sonication buffer recipe.

Reagents	**Final concentration**
Tris	25 mM
GuHCl	8 M
GuHCl was first weighed out into a measuring beaker, then dissolved in water by slowly adding distilled water while stirring on a magnetic stir plate at 100°C. During this process, Tris was added from a 1M pH 7 stock. The solution was then sterilised by filter sterilisation.	

Table 2.15. Extraction buffer recipe.

Reagents	**Final concentration**
Tris	25 mM
GuHCl	6 M
GuHCl was first weighed out into a measuring beaker, then dissolved in water by slowly adding distilled water while stirring on a magnetic stir plate at 100°C. During this process, Tris was added from a 1M pH 7 stock. The solution was then sterilised by filter sterilisation.	

Table 2.16. Wash buffer recipe.

Reagents	**Final concentration**
Tris	25 mM
Imidazole	500 mM
GuHCl	8 M
On the day of buffer preparation, a 2M stock of imidazole at pH 7 was prepared. GuHCl was first weighed out into a measuring beaker, then dissolved in water by slowly adding distilled water while stirring on a magnetic stir plate at 100°C. During this process, Tris	

and imidazole were both added from pH 7 stocks. The solution was then sterilised by filter sterilisation.

Table 2.17. Elution buffer recipe.

Reagents	**Final concentration**
Tris	250 mM
NaCl	1 M
NaN_3	30 mM
Buffer components were dissolved in 4/5 final volume of distilled water. The pH was then adjusted to 7, then topped up to final volume. The buffer was filter sterilised. On the day of dialysis, the dialysis buffer was diluted to 1x with Milli-Q water.	

Table 2.18. 10x Tris dialysis buffer recipe.

The pellet of a 500 mL culture was resuspended in 7.5 mL of Sonication buffer (Table 2.14) using a Pasteur pipette. The suspension was then sonicated on ice at 40% amplitude 6x 30s on, 30s off cycle. Sonicate was promptly submerged in ice to avoid heat related degradation. The lysate was then spun down for 15 minutes at 24,000 rpm at 4°C using a JA 25.50 rotor. The supernatant was discarded, then the insoluble pellet resuspended in 12.5 mL of extraction buffer (Table 2.15). Large clumps of insoluble pellet were broken apart by using a small plastic Pasteur pipette. The suspension was then sonicated on ice at 50% amplitude 6x 30s on, 30s off cycle, then mixed vigorously at room temperature for 30 minutes using a magnetic stir plate.

Before purification, the sample was first filter sterilised using a 0.2 µm syringe filter, then diluted to 6 M GuHCl with sonication buffer (Table 2.14). The sample was then ready for column application.

All purifications were performed using 1 mL Cytiva HisTrap FF columns at a 1 mL/min flow rate on an ÄKTA Start. The column was primed with five column volumes of wash buffer (Table 2.16). The sample was then applied, then the column again washed with five column volumes of wash buffer. After washing, protein was eluted with five column volumes of elution buffer (Table 2.17).

Elution fractions containing protein were collected, then diluted to 25 mL using wash buffer (Table 2.16). The sample was dialysed overnight against 4 L of tris dialysis buffer, using 3.5 kDa MWCO dialysis tubing (Protocol 2.11). The dialysis buffer was stirred, and the dialysis performed at 4°C overnight.

Day 6

The sample bag was swapped into a fresh 4 L of dialysis buffer, then left to dialyse for a further 2-4 hours. At this point, precipitated protein was removed by centrifuging the sample at 9000 xg for fifteen minutes. Precipitated protein could be recovered by resuspension in 25 mL wash buffer, then by repeating the same dialysis steps (Protocol 2.12).

Protein concentration was approximated by measuring A280 absorbance using a NanoDrop™ One UV-Vis spectrophotometer in triplicate. TEV protease cleavage was set up by adding EDTA and DTT to a final concentration of 1 and 5 mM respectively. TEV protease was then added in a 1:10 TEV:MmsFcc concentration ratio. The cleavage reaction was incubated overnight at 34°C (Protocol 2.09).

Day 7. Cleanup of cleavage reaction through nickel column

The sample was concentrated using an Amicon 50 mL ultrafiltration unit with a 1 kDa regenerated cellulose membrane. Nitrogen gas was passed over the sample, with a maximum system pressure of 50 psi applied. The sample was concentrated to a volume of 6.25 mL, then diluted to 25 mL with extraction buffer. The denatured sample was then passed down a 1mL nickel column. The flow through was collected, then run down the column a further two times. After this, the column was washed with five column volumes of wash buffer, before eluting impurities with five column volumes of elution buffer.

Flow through fractions containing cleaved MmsFcc were collected, pooled then diluted to 25 mL with wash buffer. The sample was dialysed overnight against 4 L of BisTris dialysis buffer, using 3.5 kDa MWCO dialysis tubing (Table 2.19) (Protocol 2.11). The dialysis buffer was stirred, and the dialysis performed at 4°C overnight. For protein destined for metal titrations, sodium azide was omitted.

Reagents	**Final concentration**

BisTris	250 mM
NaCl	1 M
NaN_3 (optional)	30 mM
Buffer components were dissolved in 4/5 final volume of distilled water. The pH was then adjusted to 6.5, then topped up to final volume. The buffer was filter sterilised. On the day of dialysis, the dialysis buffer was diluted to 1x with Milli-Q water.	

Table 2.19. 10x BisTris dialysis buffer recipe.

Day 8:

The sample bag was swapped into a fresh 4 L of dialysis buffer, then left to dialyse for a further two to four hours. Any precipitated protein was removed by centrifuging the sample at 9000 rpm for fifteen minutes. Sample concentration was measured by A280 absorbance. For concentration of volumes down to 2 mL, ultrafiltration was performed. This was usually sufficient to concentrate samples to 100 µM, which was the highest working concentration for all 2-dimensional experiments. For 3-dimensional experiments, further concentration was required. This was done using a 2 mL Amicon 3 kDa centrifugal concentrator. Samples were spun at 3000 xg in ten-minute intervals. In between intervals, the membrane was washed by gently pipetting the solution over it.

MmsFcc could be stored for up to two weeks at 4°C, at a concentration no greater than 100 µM. Longer term storage at -20°C and -80°C was also possible, but some protein was lost to the freeze thaw cycle.

2.8 General preparation of NMR samples

The majority of NMR experiments were performed using a Bruker 800 MHz Neo spectrometer, equipped with a high sensitivity TCI cryoprobe. Some experiments were performed on a Bruker 600 MHz Neo spectrometer, equipped with a high sensitivity TCI cryoprobe. Unless specified, it can be assumed that the 800 MHz spectrometer was used. NMR samples were prepared in NE-UP5 standard tubes, due to the high field strengths used.

Samples were prepared to a final volume of 550 µL, consisting of 10% D_2O and 5mM TSP. Samples were pipetted into the NMR tube using a Gilson pipette, then the tube was capped and centrifuged by hand centrifuge.

2.9 Optimal TEV cleavage and denatured Ni-NTA reaction cleanup

At the stage when optimal purification protocol requires cleavage by TEV protease (Day six of Section 2.7), the MmsFcc sample contains the following buffer components (Table 2.20).

Tris pH 7	25 mM
NaCl	100 mM
Protein	X µM

Table 2.20. The composition of MmsFcc sample before cleavage by TEV protease. The concentration of protein is a variable value, denoted by X.

Efficient cleavage by TEV protease required the addition of DTT and EDTA, and a concentration ratio of 10x MmsFcc to 1x TEV protease. EDTA was added from a 0.1 M stock solution, which was stored at room temperature. With regards to DTT, it proved essential that a solution of DTT had to be prepared on the day of cleavage (Section 3.6). The following reaction conditions were used for TEV protease cleavages (Table 2.21).

Tris pH 7	25 mM
NaCl	100 mM
Protein	X µM
TEV protease*	X/10 µM
EDTA*	1 mM

DTT*	5 mM

Table 2.21. Reaction conditions for TEV protease cleavage of MmsFcc. *denotes components that are added on the day of cleavage.

The reaction was incubated overnight at 34°C. Cleanup of the reaction via affinity between a nickel column and the cleaved His-Tags and His-TEV required the denaturing of the sample with 6 M GuHCl. This was required cause MmsFcc displayed affinity for the nickel matrix via its structured metal binding loop. For larger volume samples, the sample volume was first reduced to at least 12.5 mL with an Amicon ultrafiltration unit (Protocol 2.7), this was done to not waste GuHCl. The sample was then mixed with 8 M GuHCl extraction buffer (Table 2.15) to a final concentration of 6 M GuHCl. The denatured sample was then passed over a 1 mL HisTrap FF column using an AKTA start chromatography system, at a flow rate of 1 mL/min. The flow through was collected, which contained mostly pure cleaved protein.

2.10 Size exclusion chromatography

All size exclusion chromatography was performed using an AKTA start chromatography system. For pre-poured columns, the sample was applied directly on top of the bead bed, then passed through the column using the AKTA pump. For prepacked columns, sample was applied using a 2 mL superloop.

2.11 Dialysis

Due to the size of MmsFcc (12.7 kDa), all dialysis was performed using Spectra/Por™ 3.5 kDa molecular weight cutoff dialysis membrane. The desired exchange buffer was prepared from a 10x stock by dilution with Milli-Q H_2O before preparing the dialysis bag. Dialysis membrane was first washed with Milli-Q water to hydrate the membrane. One end of the membrane was then folded over twice and sealed with a plastic clip, then the sample was placed inside the bag using a Pasteur pipette. The other end was then folded over and sealed with a plastic clip. The sample bag was then placed in the exchange buffer, with a foam float attached to the top. The exchange buffer was then placed in a 4°C fridge and stirred gently with a magnetic stir bar overnight. In many cases (due to high concentrations of GuHCl), it was essential to perform a second dialysis stage. This was done with a fresh amount of

exchange buffer, and the sample bag was transferred to the fresh buffer and allowed to mix for two to four hours.

2.12 Recovery of precipitated protein

It was possible to recover protein from the insoluble precipitate that formed while MmsFcc was refolding during the dialysis process. Precipitate was pelleted by centrifuging the protein sample for 15 minutes at 9000 xg. After centrifugation, the supernatant was transferred to a fresh tube. The remaining pelleted precipitate was then resolved by resuspending with 6 M GuHCl wash buffer (Table 2.16). For every 2 mL of sample used in the previous dialysis step, 1 mL of GuHCl was buffer was used to resuspend the precipitate. After resuspension, protein can be recovered by dialysis against a non-denaturing buffer (Protocol 2.11). With this method, it was possible increase protein yield by 50 % for each growth.

2.13 AlphaFold2 structural predictions

All AlphaFold2 structural predictions were performed using the AlphaFold2 collaborative server called ColabFold, hosted at the following web address : https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb

This service aims to improve the accessibility of AlphaFold for simple use cases. It combines MMseqs2, a programme which searches for protein sequence homology, with AlphaFold2 to reduce the computational load and produce predicted structures quickly. Sequences were simply pasted into the query sequence input bar, then the prediction ran following the webpages instructions. No settings were changed, ColabFold predicts five structures, which are ranked by pLDDT score (for monomers). This is a per-residue measure of local confidence in the structure, meaning that there can be variance in confidence across a structure. The five structures produces were aligned then compared in PyMol, to check the similarity of the structures. The pLDDT plots produced by CollabFold were also compared. Once it was determined that all structures were near-equal, the highest ranked structure was then used for analysis or figure making.

2.14 Preparation of Fe^{2+}, Fe^{3+}, Ca^{2+}, and Zn^{2+} titration samples

The following metal salts were used when preparing ligand stocks:

- $FeCl_2$ (Fe^{2+})
- $FeCl_3$ (Fe^{3+})
- $CaCl_2$ (Ca^{2+})
- $ZnSO_4$ (Zn^{2+})

$ZnCl_2$ was trialled but formed insoluble Zinc chloride hydroxide when dissolved in water. Titration samples were prepared from a mix of protein stock and ligand stock. The protein stock was first prepared to the following conditions (Table 2.22).

BisTris pH 6.5	25 mM
NaCl	100 mM
MmsFcc	100 µM

Table 2.22. Composition of protein stock for metal titrations.

Titration samples were prepared to a final volume of 550 µL, by mixing the following components:

- 400 µL protein stock (Table 2.22)
- 100 µL Ligand stock (Table 2.23 or 2.24) + NMR buffer (Table 2.25)
- 45 µL D_2O
- 5 µL of 100 mM TSP in D_2O

This resulted in a final protein concentration of 72 µM. The ligand stock was prepared to a concentration where adding 100 µL to the titration sample resulted in a final concentration of 2.304 mM ligand, which is 32x the concentration of MmsFcc. The ionic strength of the titration sample was balanced by reducing the concentration of NaCl in the ligand stock (Table 2.23-2.24).

Component	**Concentration**
BisTris pH 6.5	25 mM
NaCl	74.66 mM
$ZnSO_4$/$CaCl_2$/$FeCl_2$	12.67 mM

Buffer was prepared to four-fifths the final volume with Milli-Q H_2O, then pH adjusted using 0.1 M NaOH and 0.1 M HCl. After pH adjustment, the sample was topped up to final volume with Milli-Q H_2O, then filter sterilised.

Table 2.23. Ligand stock recipe for Zn^{2+}, Fe^{2+} or Ca^{2+} titrations.

Component	Concentration
BisTris pH 6.5	25 mM
NaCl	62 mM
$FeCl_3$	12.67 mM
Buffer was prepared to four-fifths the final volume with Milli-Q H_2O, then pH adjusted using 0.1 M NaOH and 0.1 M HCl. After pH adjustment, the sample was topped up to final volume with Milli-Q H_2O, then filter sterilised.	

Table 2.24. Ligand stock recipe for Fe^{3+} titrations.

An NMR buffer containing no ligand or protein was used to top up titration samples that needed less than 100 µL of ligand buffer (Table 2.25).

Component	Concentration
BisTris pH 6.5	25 mM
NaCl	62 mM
Buffer was prepared from stock solutions of BisTris pH 6.5 and NaCl, after mixing the solution was made to the final volume with Milli-Q H_2O	

Table 2.25. NMR Buffer recipe.

The following schematic was used to produce all titration samples, by mixing the components listed (Table 2.26)

	Protein stock (µL)	D_2O (µL)	100 mM TSP in D_2O (µL)	Ligand stock (µL)	NMR buffer (µL)
0x	400	45	5	0	100
2x	400	45	5	6.25	93.75
4x	400	45	5	12.5	87.5
8x	400	45	5	25	75
16x	400	45	5	50	50

32x	400	45	5	100	0

Table 2.26. Schematic for the preparation of 0x, 2x, 4x, 8x, 16x, and 32x titration samples.

After mixing, the pH of the 0x sample was recorded using a Orion Combi pH Micro electrode. Since pH affects chemical shifts, it was important to maintain a consistent pH across all titration samples. 2x-32x samples were carefully adjusted to match the pH of the 0x sample using 0.1M solutions of NaOH and HCl, with a variation from the target pH of ±0.02 allowed. Once adjusted, samples were then transferred to the NMR tubes following Protocol 2.8. All titration spectra were recorded overnight using the same experimental parameters.

2.15 Preparation of Fe^{2+} titration samples

Due to the propensity of Fe^{2+} to oxidise to Fe^{3+} in solution, extra caution was taken to reduce the availability of dissolved oxygen. Fe^{2+} was prepared following the same protocol described in Section 2.15, yet the following nitrogen flushing steps were included.

- Ligand buffer was flushed with nitrogen for five minutes before addition of $FeCl_2$.
- After pH adjustment of ligand stock, the solution was flushed for one additional minute.
- Protein stock was flushed with nitrogen for one minute before addition of ligand stock.
- NMR buffer was also flushed for five minutes before being added to the titration sample.
- After the pH adjustment of titration sample, the sample was quickly transferred to an NMR tube. Here, it was gently flushed for one minute (this was achieved by attaching a glass Pasteur pipette to the nitrogen outlet with parafilm). After flushing, the NMR tube lid was attached and then sealed with parafilm.

2.16 Preparation of 300x titration samples

300x titration samples were prepared mostly following the same procedure as described in Protocol 2.15. The following ligand stock recipe was instead used (Table 2.27).

Component	Concentration
BisTris pH 6.5	25 mM

$ZnSO_4/FeCl_2$	108 mM
Buffer was prepared to four-fifths the final volume with Milli-Q H_2O, then pH adjusted using 0.1 M NaOH and 0.1 M HCl. After pH adjustment, the sample was topped up to final volume with Milli-Q H_2O, then filter sterilised.	

Table 2.27. Ligand stock recipe for 300x titrations of Fe^{2+}/Zn^{2+}. For Fe^{2+}, nitrogen flushing steps described in Protocol 2.16 were applied.

NaCl was excluded, since the metal salt was already introducing an excess of anions (108 mM SO_4^2, 216 mM Cl^-). To mitigate this, protein was dialysed into the following buffer (Tables 2.28).

Component	Concentration
BisTris pH 6.5	25 mM
NaCl	82.5 mM
Buffer was prepared from stock solutions of BisTris pH 6.5 and NaCl, after mixing the solution was made to the final volume with Milli-Q H_2O	

Table 2.28. Dialysis buffer for protein used in 300x titrations.

With this, the concentration of negative charges is roughly equal to that contributed by a 100 mM concentration of NaCl.

2.17 Preparation of MmsFcc pH titration samples

To assess the effect of pH on the MmsFcc HSQC spectrum, a range of pH 5.5-7.5 was trialled. To achieve this, the following samples were prepared:

- pH 5.5
- pH 6.0
- pH 6.2
- pH 6.4
- pH 6.6
- pH 6.8
- pH 7.0
- pH 7.5

pH 5.5 and 7.5 samples were each prepared by dialysing the protein into buffer of the target concentration (Table 2.29). For all samples, a final concentration of 72 µM MmsFcc was used.

Component	Concentration
BisTris pH 5.5/7.5	25 mM
NaCl	100 mM
BisTris power was weighed out, then dissolved in Milli-Q H_2O to a final volume of four-fifths. NaCl was added from a stock solution, then the solution was mixed, and pH adjusted. Afterwards mixing the solution was made to the final volume with Milli-Q H_2O, then filter sterilised.	

Table 2.29. Dialysis buffer for pH 5.5/7.5 titration samples.

The pH 6-7 samples were prepared from a 100 µM stock of protein at pH 6.5 (Table 2.x). 400 µL of stock protein was aliquoted into Eppendorf tubes, then 45 µL of D_2O and 5 µL of 100 mM TSP in D_2O was added. The pH was adjusted with the gradual addition of 0.1 M solutions of NaOH or HCl. Once the target pH was reached, the sample was topped up to 550 µL with an NMR buffer pH adjusted to the target pH.

2.18 NMR spectroscopy experimental parameters

Table 2.30. lists all pulse sequences used for all nitrogen HSQC experiments used during this project, with the number of scans also listed. All double and triple resonance spectra were processed using a gaussian window function. All titration HSQC spectra (chapter 5) used the standard Bruker pulse sequence (hsqcetfpf3gpsi2) with 16 scans.

Experiment name	Figure number (spectrum colour)	No. of scans	Pulse sequence	Reference
Proton 1D	3.12-3.13	8	zgesgp	T.-L. Hwang & A.J. Shaka, J. Magn. Reson., ; Series A 112 275-279 (1995)
^{15}N-HSQC	3.14	32	hsqcetfpf3gpsi2	A.G. Palmer III, J. Cavanagh, P.E. Wright & M. Rance, J. Magn. ; Reson. 93, 151-170 (1991)
^{15}N-HSQC	3.15 (all spectra)	16	hsqcetfpf3gpsi2	A.G. Palmer III, J. Cavanagh, P.E. Wright & M. Rance, J. Magn. ; Reson. 93, 151-170 (1991)
^{15}N-HSQC	3.16 (all spectra)	16	hsqcetfpf3gpsi2	A.G. Palmer III, J. Cavanagh, P.E. Wright & M. Rance, J. Magn. ; Reson. 93, 151-170 (1991)

^{15}N-HSQC	3.17 (all spectra)	16	hsqcetfpf3gpsi2	A.G. Palmer III, J. Cavanagh, P.E. Wright & M. Rance, J. Magn. ; Reson. 93, 151-170 (1991)
^{15}N-HSQC	3.19	32	hsqcetfpf3gpsi2	A.G. Palmer III, J. Cavanagh, P.E. Wright & M. Rance, J. Magn. ; Reson. 93, 151-170 (1991)
^{15}N-HSQC	3.29 (black)	32	hsqcetfpf3gpsi2	A.G. Palmer III, J. Cavanagh, P.E. Wright & M. Rance, J. Magn. ; Reson. 93, 151-170 (1991)
^{15}N-HSQC	3.29 (blue)	8	hsqcetfpf3gpsi2	A.G. Palmer III, J. Cavanagh, P.E. Wright & M. Rance, J. Magn. ; Reson. 93, 151-170 (1991)
^{15}N-HSQC	3.31 (all spectra)	16	hsqcetfpf3gpsi2	A.G. Palmer III, J. Cavanagh, P.E. Wright & M. Rance, J. Magn. ; Reson. 93, 151-170 (1991)
^{15}N-HSQC	4.7	4	hsqcetfpf3gpsi2	A.G. Palmer III, J. Cavanagh, P.E. Wright & M. Rance, J. Magn. ; Reson. 93, 151-170 (1991)
^{15}N-HSQC	4.8	4	hsqcetfpf3gpsi2	A.G. Palmer III, J. Cavanagh, P.E. Wright & M. Rance, J. Magn. ; Reson. 93, 151-170 (1991)
^{15}N-HSQC	4.11	4	hsqcetfpf3gpsi2	A.G. Palmer III, J. Cavanagh, P.E. Wright & M. Rance, J. Magn. ; Reson. 93, 151-170 (1991)
^{15}N-HSQC	4.12	4	hsqcetfpf3gpsi2	A.G. Palmer III, J. Cavanagh, P.E. Wright & M. Rance, J. Magn. ; Reson. 93, 151-170 (1991)
^{15}N-HSQC	4.16	32	hsqcetfpf3gpsi2	A.G. Palmer III, J. Cavanagh, P.E. Wright & M. Rance, J. Magn. ; Reson. 93, 151-170 (1991)
^{15}N-HSQC	4.17	4	hsqcetfpf3gpsi2	A.G. Palmer III, J. Cavanagh, P.E. Wright & M. Rance, J. Magn. ; Reson. 93, 151-170 (1991)
^{15}N-HSQC	4.19 (Black)	32	hsqcetfpf3gpsi2	A.G. Palmer III, J. Cavanagh, P.E. Wright & M. Rance, J. Magn. ; Reson. 93, 151-170 (1991)
^{15}N-HSQC	4.19 (Orange)	4	hsqcetfpf3gpsi2	A.G. Palmer III, J. Cavanagh, P.E. Wright & M. Rance, J. Magn. ; Reson. 93, 151-170 (1991)

Table 2.30. A list of scans, pulse sequences and references for proton 1D and ^{15}N-HSQC experiments.

Table 2.31 lists all experiments used for the assignment and structural calculation of MmsFcc and MmsFcc D49S. For both the wildtype and the mutant, the same pulse sequences and number of scans were used.

Experiment name	No. of scans	Pulse sequence	Reference
HNCA	16	hncagpwg3d	P. Schanda, H. v. Melckebeke & B. Brutscher, ; J. Am. Chem. Soc. 128, 9042-9043 (2006)
HN(CO)CA	16	hncocagpwg3d	P. Schanda, H. v. Melckebeke & B. Brutscher, ; J. Am. Chem. Soc. 128, 9042-9043 (2006)
HNCACB	16	cbcanhgpwg3d	S. Grzesiek & A. Bax, J. Magn. Reson. 99, 201-207 (1992)
HN(CO)CACB	16	cbcaconhgpwg3d	S. Grzesiek & A. Bax, J. Biomol. NMR 3, 185-204 (1993)
HNCO	16	hncogpwg3d	S. Grzesiek & A. Bax, J. Magn. Reson. 96, 432 - 440 (1992)
HN(CA)CO	16	hncacogpwg3d	R.T. Clubb, V. Thanabal & G. Wagner, J. Magn. Reson. 97, 213-217 (1992)
Aliphatic ^{13}C HSQC		hsqcctetgpsisp	G.W. Vuister & A. Bax, J. Magn. Reson. 98, 428-435 (1992)
Aliphatic ^{13}C HSQC-NOESY	16	noesyhsqcetgpsi3d	A.G. Palmer III, J. Cavanagh, P.E. Wright & M. Rance, J. Magn. ; Reson. 93, 151-170 (1991)
Aromatic ^{13}C HSQC		trosyargpphwg	K. Pervushin, R. Riek, G. Wider & K.Wuethrich, J. Am. Chem. Soc. 120, ; 6394-6400 (1998)
Aromatic ^{13}C HSQC-NOESY	8	noesyhsqcetgpsi3d	A.G. Palmer III, J. Cavanagh, P.E. Wright & M. Rance, J. Magn. ; Reson. 93, 151-170 (1991)
^{15}N HSQC-TOCSY	32	dipsihsqcf3gpsi3d	A.G. Palmer III, J. Cavanagh, P.E. Wright & M. Rance, J. Magn. ; Reson. 93, 151-170 (1991)
^{15}N HSQC-NOESY	24	noesyhsqcf3gpsi3d	A.G. Palmer III, J. Cavanagh, P.E. Wright & M. Rance, J. Magn. ; Reson. 93, 151-170 (1991)
^{15}N-HSQC	8	hsqcetfpf3gpsi2	A.G. Palmer III, J. Cavanagh, P.E. Wright & M. Rance, J. Magn. ; Reson. 93, 151-170 (1991)
HCCH	16	hcchdigp3d	(L.E. Kay, G.Y. Xu, A.U. Singer, D.R. Muhandiram & J. D. Forman-Kay ; J. Magn. Reson. B 101, 333 - 337 (1993))
HCC	16	hcchdigp3d2	(L.E. Kay, G.Y. Xu, A.U. Singer, D.R. Muhandiram & J. D. Forman-Kay ; J. Magn. Reson. B 101, 333 - 337 (1993))

Table 2.31. A list of scans, pulse sequences and references for assignment and structural calculation NMR experiments.

Chapter 3: Optimisation of MmsFcc expression and purification

This chapter describes the challenges encountered when attempting to purify sufficient amounts of MmsFcc and the mutant MmsFcc D49S, and the solutions found to overcome them. Since studying proteins the size of MmsFcc (15.2 kDa with the HisTag, 12.7 kDa without) with NMR spectroscopy requires the uniform isotopic labelling of protein with 'NMR active' isotopes of carbon and nitrogen (^{13}C, ^{15}N), it was essential to adapt the purification protocol inherited from previous researchers (Hu et al., 2021a). This required a switch from 'rich media', such as LB or autoinduction media, to minimal medium. For this project, the M9 minimal media recipe was used, and this recipe is given in Section 2.1. Switching to minimal media increased cost due to the price of isotopic source compounds, reduced yield due to lower growth potential of minimal media and required the establishment of new optimal expression conditions. The fully optimised protocol is described in section 2.7. Due to time and financial constraints, the only 2 constructs that were extensively studied were MmsFcc and MmsFcc D49S. As discussed in Section 1.7, the loss of function (F57Q) and the rescue of function mutant (F54Y, F57Q) were both of interest to this study. However, they were never successfully purified during this project. This was because the plasmid stocks were lost during the handover of resources, while replacement plasmids ordered from GenScript did not express protein when trialled.

Further information on the pulse programmes used, and the number of scans collected can be found in Section 2.18.

3.1 Transformation of *E. coli* with MmsFcc plasmids and sequencing

Plasmids for MmsFcc and mutants were handed over from a previous researcher, so their integrity was confirmed by plasmid miniprepping and sequencing. For plasmid purification, XL-10 gold cells were made chemically competent through treatment with ruthenium chloride (Section 2.2). 50 µL of competent cells were transformed with 1 µL of plasmid DNA. All miniprepping of plasmid DNA was performed using the Monarch® Plasmid miniprep kit

(Section 2.3). All sequencing for this project was outsourced to GENEWIZ, inc. through their Sanger sequencing service; the T7 forward and reverse primers were selected. Only the forward primer worked, and the trace alignments for MmsFcc and MmsFcc D49S sequencing with the forward primer are shown in Figures 3.1-3.2.

Figure 3.1. Trace alignment of pPR-IBA1_MmsFcc Sanger sequencing, using the T7 forward primer.

Figure 3.2. Trace alignment of pPR-IBA1_MmsFcc-D49S Sanger sequencing, using the T7 forward primer.

From these trace alignments, it was clear that the plasmids inherited from previous researchers were still viable. During this project, sequencing was performed after every miniprep to check that the sequence remained valid. Now that gene integrity had been confirmed, the next step was to establish optimal growth conditions.

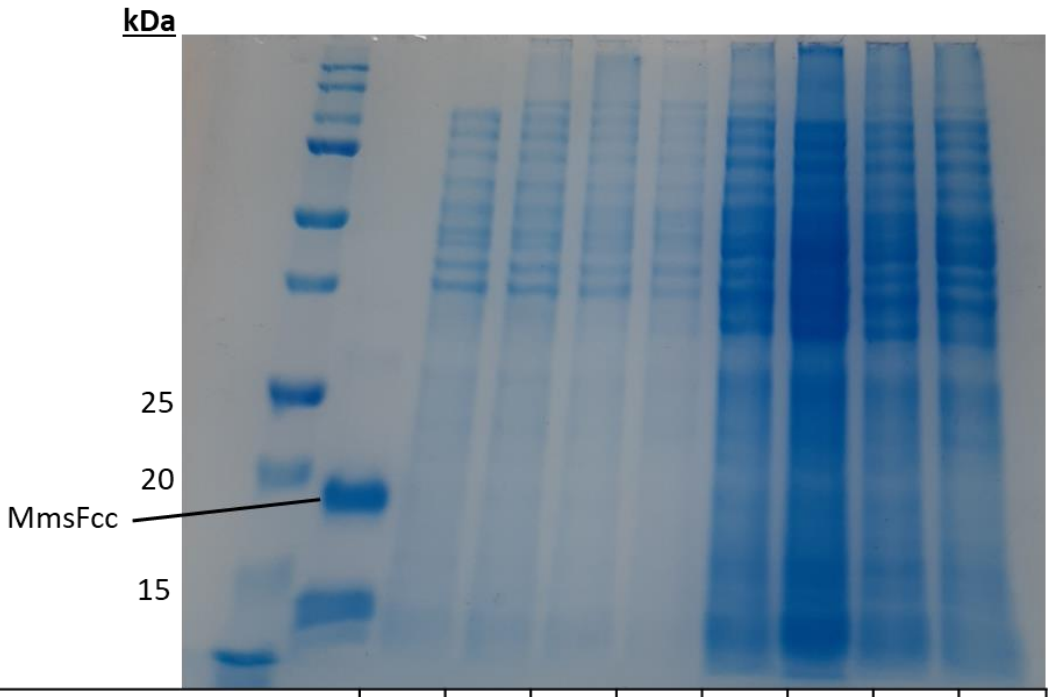
3.2 MmsFcc expression trials in rich and minimal media

Prior to this project, MmsFcc and its mutants were purified with a denaturing protocol. This was performed because the protein tended to aggregate into insoluble intracellular inclusion bodies. *In vitro* refolding of MmsFcc is possible due to its simple tertiary structure, which has been extensively proven by (Rawlings et al., 2019a). Prior to this project, researchers working on MmsFcc opted to grow cultures in auto-induction media. The option of auto-induction was not explored for this project, since it is challenging to source autoinduction medium capable of isotopically labelling proteins with ^{13}C and/or ^{15}N . Commercially available labelled mixes are available but expensive, and preparation of materials from scratch is complex and expensive (Crowley & Rafferty, 2019).

Before settling on a denaturing purification strategy, the possibility of expressing significant amounts of protein in the soluble fraction was investigated. To assess this possibility, a range of growth conditions were trialled. Since inclusion bodies form as a cellular response to excessive protein expression, common strategies for mitigating their formation centre on slowing the rate of expression. Another approach is to introduce the cell to stress conditions, to promote the co-expression of stress-induced protein folding chaperones (Francis & Page, 2010). To slow protein expression, a range of incubation temperatures (18°C, 25°C, and 30°C) and IPTG concentrations (0.1 mM, 1 mM) were trialled. Cells grown in minimal medium experience stress conditions due to the lack of available growth nutrients.

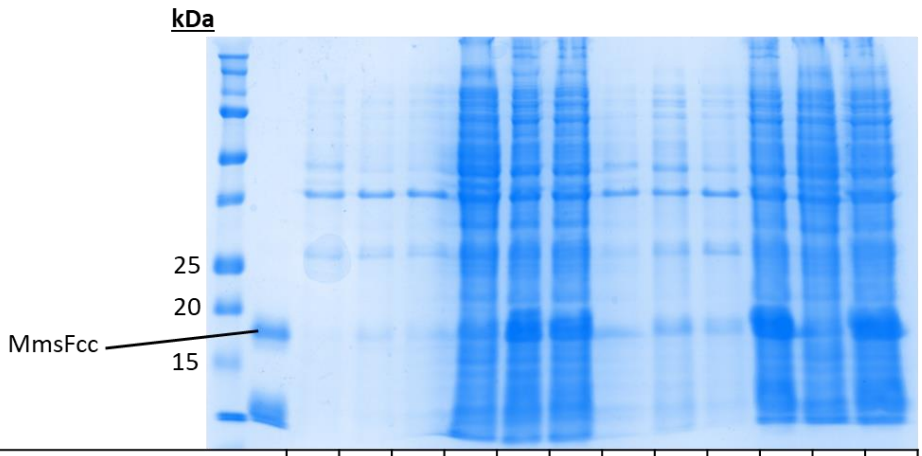
Small scale BL21-DE3 *E. coli* expression trial cultures were prepared for analysis by sodium-dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE); this preparation is described in Sections 2.5-2.6. During initial rounds of testing, cultures were inoculated with cells from a 30% glycerol stock of transformed BL21 DE3 cells stored at -80°C. It was

eventually discovered that the use of a freezer stock resulted in undetectable levels of protein expression, as evidenced by Figures 3.3-3.4.



Soluble (S) or insoluble (IS) fraction	IS	IS	IS	IS	S	S	S	S
IPTG concentration [mM]	0	0.1	1	1	0	0.1	1	1
Presence of 3% ethanol	×	×	×	✓	×	×	×	✓

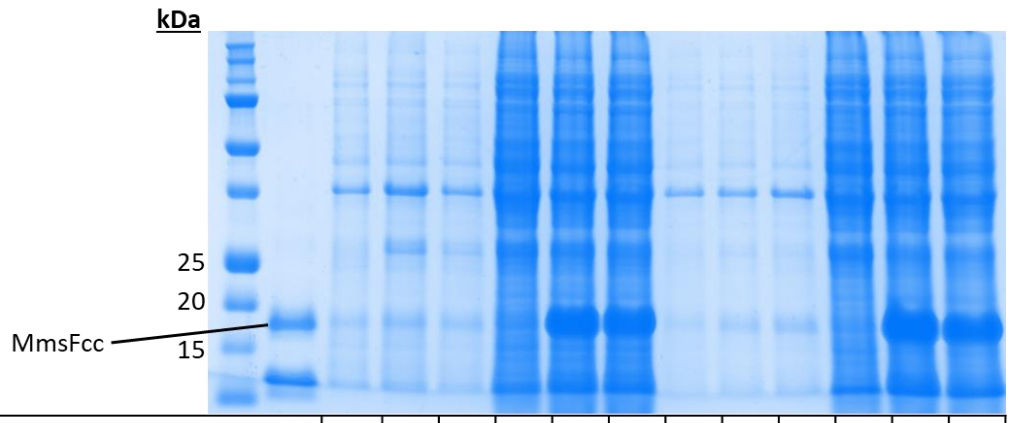
Figure 3.3. Expression trial of MmsFcc in BL21 DE3 LB cultures, incubated overnight at 30°C, then analysed via SDS-PAGE. MmsFcc markers are from samples inherited from a previous student. Cultures were inoculated with a frozen stock of cells.



Soluble (S) or insoluble (IS) fraction	S	S	S	IS	IS	IS	S	S	S	IS	IS	IS
Temperature (°C)	25	25	25	25	25	25	30	30	30	30	30	30
IPTG concentration [mM]	0	0.1	1	0	0.1	1	0	0.1	1	0.1	0	1

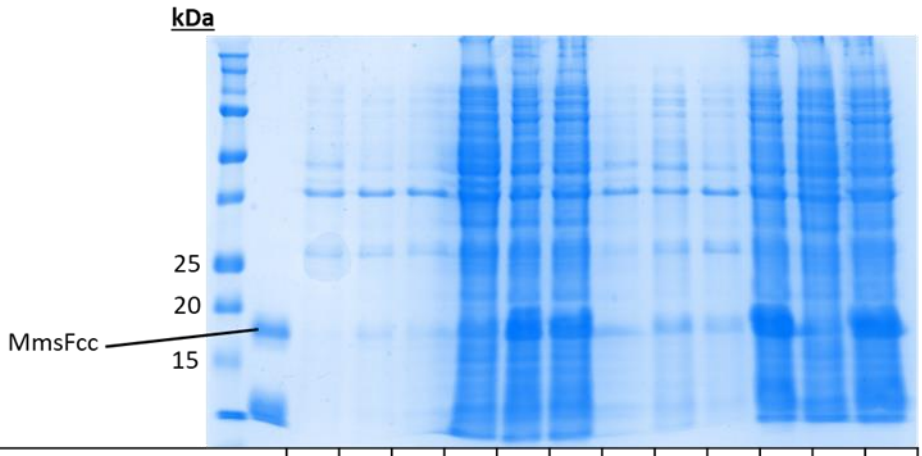
Figure 3.4. Expression trial of MmsFcc in BL21 DE3 LB cultures, incubated overnight at 25 or 30°C, then analysed via SDS-PAGE. Cultures were inoculated with cells transformed and plated the previous day.

Markers of MmsFcc were included in these gels, using frozen samples of protein inherited from previous researchers. In these lanes, there is a second band around 10 kDa. These are likely a degradation product and are fully explored in Section 3.8. From these Figures, there is significantly more protein expressed in the same growth conditions when a fresh transformation is performed. As a result, many expression trials were not truly representative due to the use of frozen cell stocks and have therefore not been included in this thesis. In the interest of time, the more unusual conditions were not repeated. IPTG concentrations of 0.1 and 1 mM were re-trialled along with incubation temperatures of 25 and 30 degrees. These expression conditions were then analysed by SDS-PAGE in Figures 3.5-3.6.



Soluble (S) or insoluble (IS) fraction	S	S	S	IS	IS	IS	S	S	S	IS	IS	IS
Temperature (°C)	25	25	25	25	25	25	30	30	30	30	30	30
IPTG concentration [mM]	0	0.1	1	0	0.1	1	0	0.1	1	0	0.1	1

Figure 3.5. Expression trial of MmsFcc in BL21 DE3 minimal media cultures, incubated overnight at 25°C or 30°C, then analysed via SDS-PAGE.

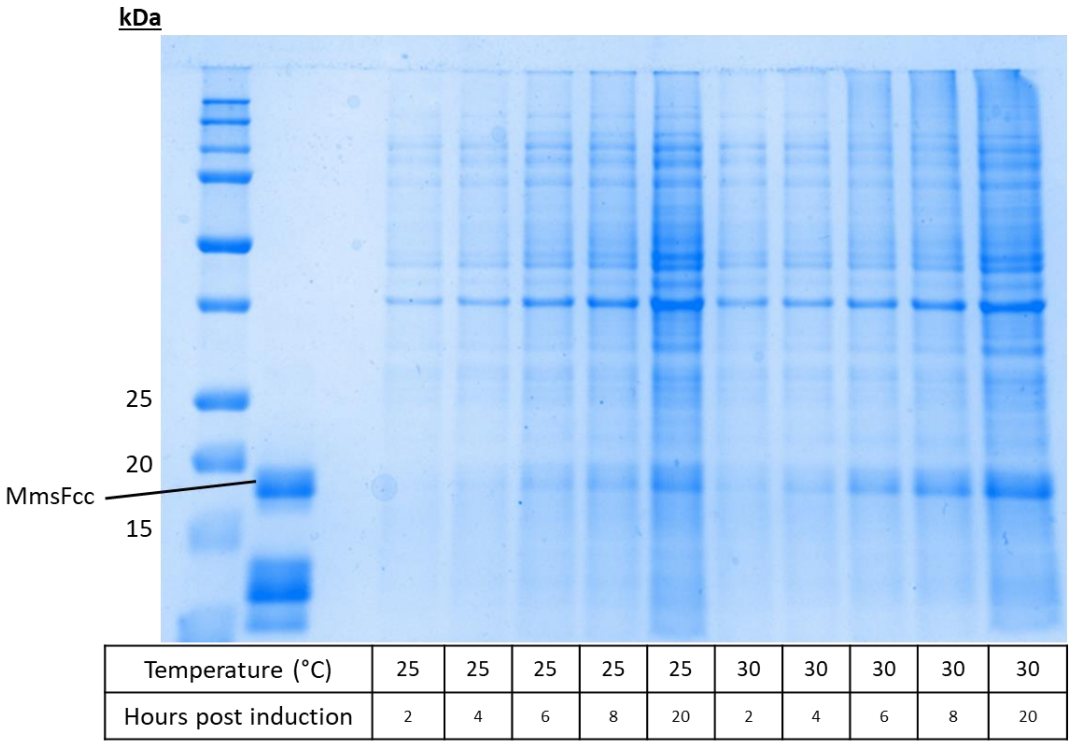


Soluble (S) or insoluble (IS) fraction	S	S	S	IS	IS	IS	S	S	S	IS	IS	IS
Temperature (°C)	25	25	25	25	25	25	30	30	30	30	30	30
IPTG concentration [mM]	0	0.1	1	0	0.1	1	0	0.1	1	0.1	0	1

Figure 3.6. Expression trial of MmsFcc in BL21 DE3 LB cultures, incubated overnight at 25°C or 30°C, then analysed via SDS-PAGE.

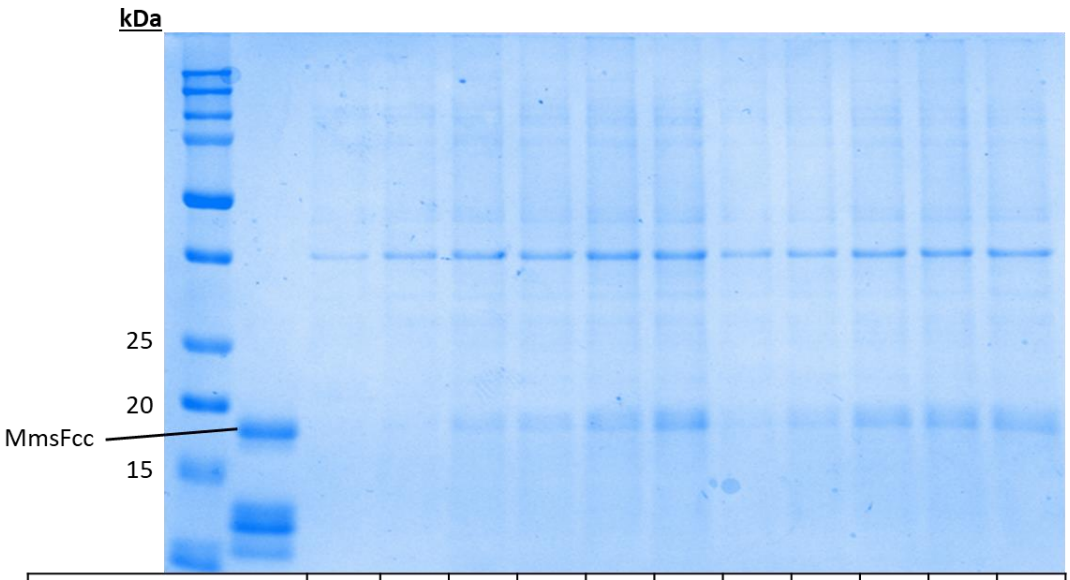
From these results, it is apparent that the amount of inclusion bodies are unaffected by changes to temperature, IPTG concentration, or the use of minimal media. For these reasons, the purification strategy was altered towards extraction with a strong denaturant. Guanidine hydrochloride (GuHCl) was selected since it had been used to purify MmsFcc in previous work on this project. As a result, the aim of growth conditions switched from tempering overexpression to producing as much protein as possible. An incubation temperature of 30°C, and a final IPTG concentration of 1 mM were established.

Before moving forward, the expression timeline of MmsFcc was examined via time trials. This was done to determine whether expression levels remained consistent post induction, and to see if protein degradation occurred in a time dependant manner. Samples were taken post induction at two-hour intervals, and the resulting cell lysate assessed via SDS-PAGE (Figures 3.7-3.8).



Temperature (°C)	25	25	25	25	25	30	30	30	30	30
Hours post induction	2	4	6	8	20	2	4	6	8	20

Figure 3.7. SDS-PAGE of MmsFcc whole cell lysate expression timeline in LB media, using a final IPTG concentration of 1 mM.



Temperature (°C)	25	25	25	25	25	25	30	30	30	30	30
Hours post induction	0	2	4	6	8	20	2	4	6	8	20

Figure 3.8. SDS-PAGE of MmsFcc whole cell lysate expression timeline in minimal media, using a final IPTG concentration of 1 mM.

These results show that, over a 20-hour period, the rate of protein expression does not slow significantly over time with no degradation occurring. It was therefore likely that expression of MmsFcc would not be improved by adding IPTG incrementally. Therefore, it was decided to add the full amount of IPTG at once, then the culture would be allowed to incubate overnight.

Now that growth conditions were established, the next step was to scale up cultures to volumes of 0.5-4 litres. With this scale-up, it became apparent that BL21 DE3 cultures transformed with MmsFcc then grown in minimal media were often slow growing. These cultures would often take 8-15 hours to reach an OD600 of 0.6. To combat this, a range of different starter cultures and growth conditions were trialled. Using two overnight starter cultures of 20 mL LB and 40 mL M9 minimal media per 1 L culture was found to be most effective. Alongside this, using a 2 L baffled flask per 500 mL culture helped to increase growth speed through increased aeration. In combination, these methods reduced the time to reach OD600 of 0.6 to 4-7 hours consistently.

3.3 Optimising the purification protocol

This section goes through the purification strategy stepwise, highlighting at each stage how the apparent properties of MmsFcc proved challenging, and explaining the modifications made to mitigate these issues. A fully optimised protocol is described in section 2.7.

The harvesting, sonication, and extraction of MmsFcc from inclusion bodies (with 8M GuHCl) remained largely unchanged. The first major obstacle was met when attempting to elute MmsFcc from the nickel column. Initial purification attempts involved loading the protein onto the column in a 6 M GuHCl protein extraction buffer, followed by an extensive wash with 0 M GuHCl wash buffer. This allowed MmsFcc to refold on column, meaning that when eluted, the eluent contained folded protein. Using this method removed the need to exchange GuHCl from the sample post elution, although some form of buffer exchange would still be required due to the high concentration of imidazole present in the elution buffer (0.5 M). When attempting this refolding protocol, it was quickly identified that MmsFcc had a propensity to stick to the nickel column. Attempts to elute folded protein with 0.5 M imidazole would result in a broad elution peak, as shown in Figures 3.9 and 3.10.

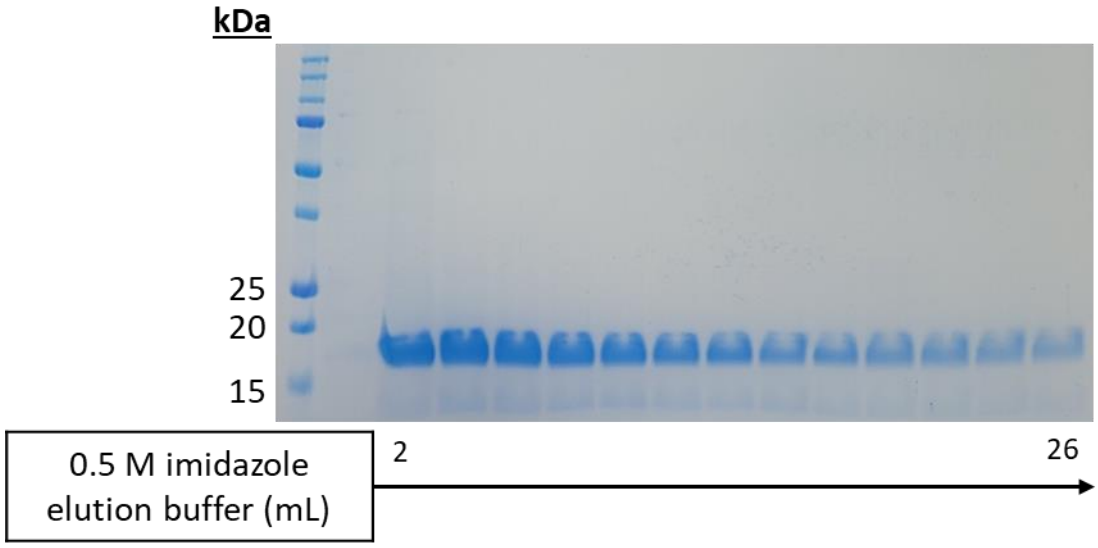


Figure 3.9. SDS-PAGE of native elution of MmsFcc from a 1 mL Ni-NTA column, using 500 mM Imidazole elution buffer.

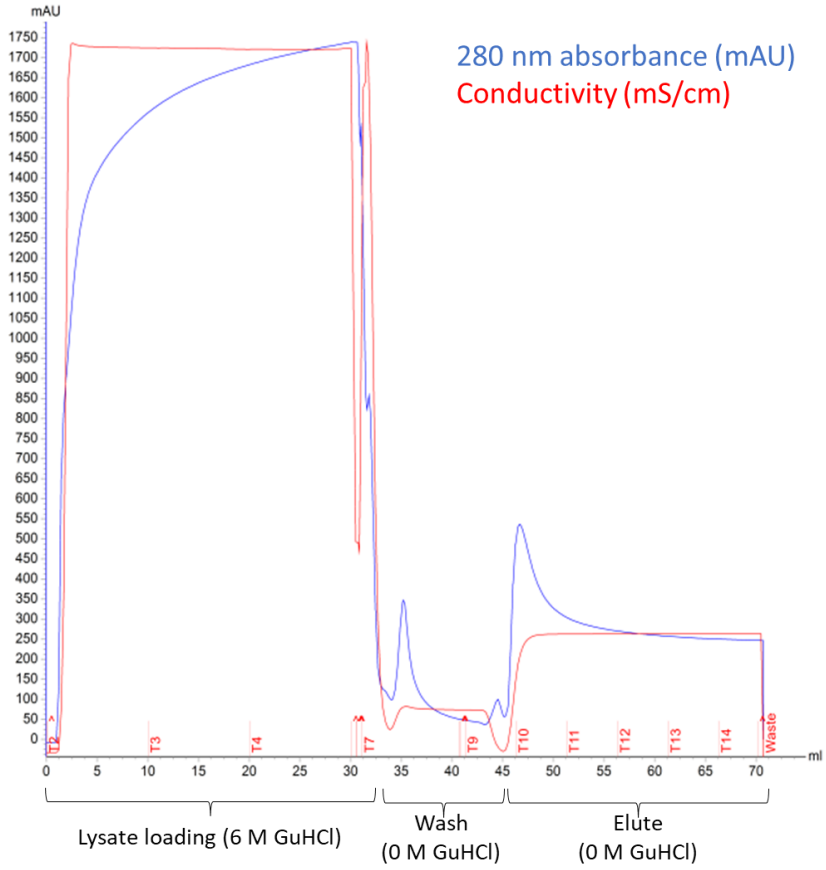


Figure 3.10. AKTA start chromatograph of refolding purification of MmsFcc. Figure 3.9 shows the presence of protein in fractions after the elution peak, representing a broad elution profile.

This was reducing yield significantly since protein remained stuck on the column. It was hypothesized that the stickiness of MmsFcc was due to the metal binding loop, and that this property was reliant on the loop being structured. Therefore, it was decided to conduct the entire purification under denaturing conditions. The resulting chromatograph in Figure 3.11 shows a much sharper elution peak when compared with Figure 3.10. High concentrations of GuHCl (>100 mM) are incompatible with SDS-PAGE, so a gel of the denatured purification has not been included.

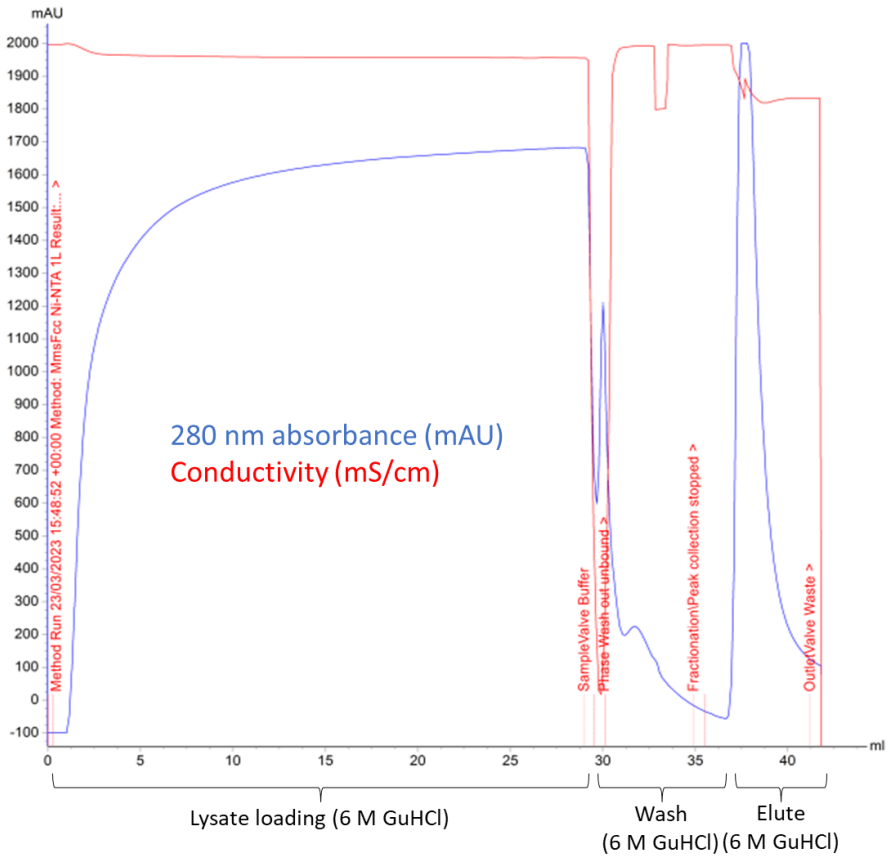


Figure 3.11. AKTA start chromatograph of a fully unfolded purification of MmsFcc.

-

Attention was then turned to improving the subsequent protein refolding and buffer exchange. Of the many options for buffer exchange, dialysis was selected for two reasons. The first was that MmsFcc had been shown to stick to nickel columns, so column methods such as desalting were potentially problematic. The use of protein concentrators for buffer exchange was avoided since the precipitation of MmsFcc is concentration dependant. It was predicted that the addition of non-denaturing buffer to a concentrated protein retentate would cause significant loss due to precipitation in the centrifugal concentrator. Of the remaining options, dialysis was selected (Section 2.11).

After elution, fractions were pooled then diluted to 50 mL with 6M GuHCL buffer. This was transferred to a 3.5 kDa MWCO dialysis bag and dialysed overnight in 4 L of the desired buffer at 4°C whilst stirring gently. Diluting the sample was found to reduce the occurrence of protein precipitation, ultimately improving yield likely due to the reduction of concentration dependant aggregation. In the morning, the buffer was replaced with fresh

buffer and the dialysis allowed to continue for two to four hours (Section 2.11). Recovery of precipitated protein was possible, and the method is described in Section 2.12.

Because of these dilutions, it was required regularly to concentrate 50-60 mL samples to volumes of 0.5-5 mL. This was first attempted with Vivaspin® 20 mL 3.5 kDa molecular weight cutoff (MWCO) centrifugal concentrators, which resulted in significant loss of protein (Section 2.12). Instead, the use of Amicon® 50 mL ultrafiltration chambers with a Millipore® 1 kDa regenerated cellulose membrane was explored (Section 2.8). These two methods were compared by concentrating two equivalent 25mL samples of His-MmsFcc to 2 mL. The centrifugal method saw a loss of 75%, in comparison the ultrafiltration method had a loss of just 30%. This is likely due to the removal of a concentration gradient in the ultrafiltration system, although the choice of membrane material could have also played a role. For this reason, ultrafiltration was performed for all large volume concentration. For preparation of 0.5 mL NMR samples, Amicon® 3 kDa MWCO 2 mL centrifugal concentrators with regenerated cellulose membranes were selected.

It was now possible to produce samples of MmsFcc with sufficient purity and concentration for NMR. The next step was to optimise the NMR buffer conditions.

3.4 Optimisation of NMR buffer conditions

A proton 1D experiment was performed on a purified sample of His-MmsFcc. The general preparation of NMR samples is described in section 2.8. The resulting spectrum is displayed in Figures 3.12 and 3.13. Details on NMR experiments performed are found in Section 2.18.

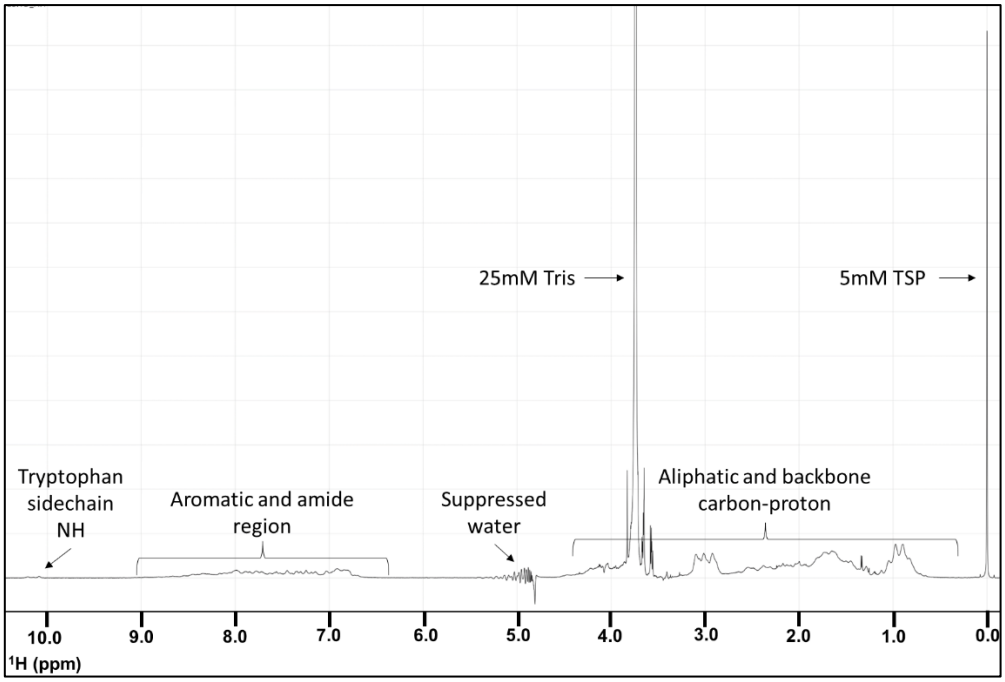


Figure 3.12. ^{1}H spectrum of 400 µM ^{15}N-MmsFcc in 25 mM Tris pH 7, 100 mM NaCl, 5 mM TSP, 3 mM NaN_3.

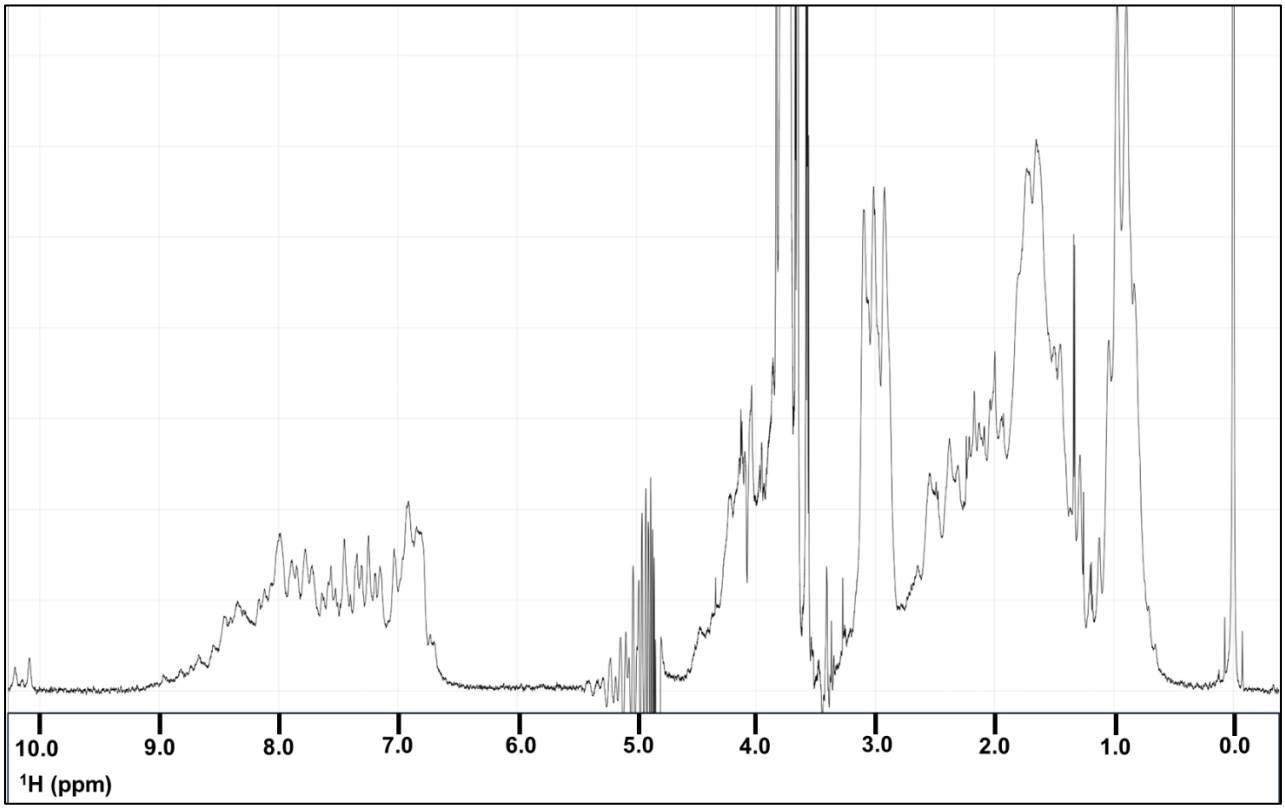


Figure 3.13. His-MmsFcc ^{1}H spectrum, scaled in relation to the intensity of protein peaks.

This spectrum provides information on protons in the sample. From this spectrum it is not possible to determine much about the sample. It can also be determined that the sample does not contain any small molecule contaminants at relevant concentrations.

A ^{1}H-^{15}N heteronuclear single-quantum coherence (HSQC) experiment was performed next on the same sample, producing a spectrum representing the NH groups in the sample (Figure 3.14).

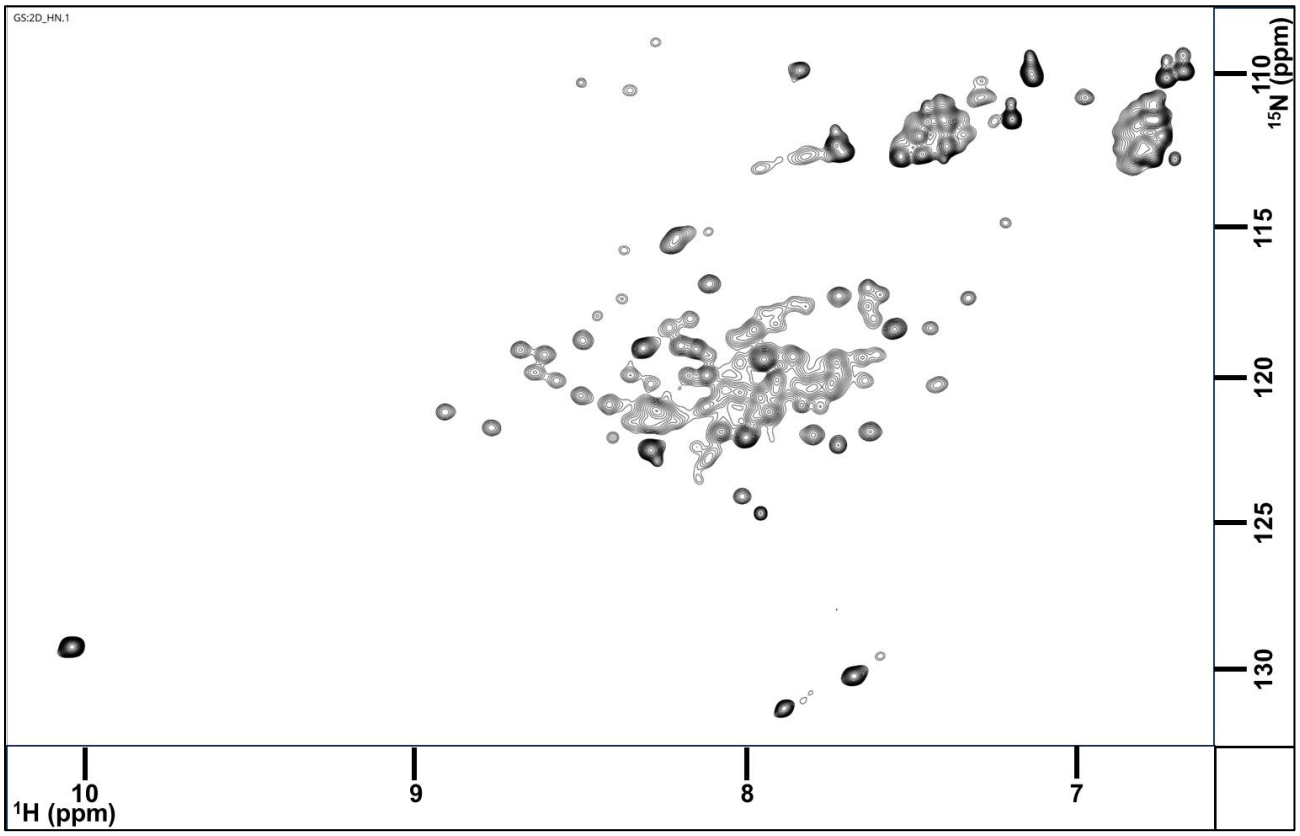


Figure 3.14. ^{1}H-^{15}N HSQC spectrum of 400 µM MmsFcc in 25 mM Tris pH 7, 100 mM NaCl, 5 mM TSP, 3 mM NaN_3.

This HSQC spectrum provides a useful snapshot of protein stability. For a sample of this concentration, many peaks are weaker than expected. This suggests that the protein is experiencing lots of dynamics, which results in lost signal. This is not surprising, since His-MmsFcc is an artificially designed protein. Systems intended to keep the protein folded and stable, such as the burying of the helical hydrophobic faces and the leucine zipper, may not be as effective as those created through evolution. Another aspect of the protein's design that could be causing increased dynamics is the N-terminal His-Tag. Having a metal binding tag on a metal binding protein, especially with the two being situated at opposite ends of the protein, creates the potential for multiple protein-protein and protein-metal interactions. Removal of the tag was an option, due to the presence of a TEV protease cleavage site in the protein sequence. Removing the tag was initially avoided, since it would

result in protein loss, due to the extra processing steps required. Tag removal is described in section 3.6.

This spectrum also has lots of broad, overlapping peaks. While this can suggest that the protein structure is not very stable, this HSQC profile was to be expected since a coiled-coil consisting of 2 identical helical sequences would result in many nuclei having the same or similar chemical environments. Nevertheless, different buffer conditions were trialled to improve spectrum quality.

-

The following samples were prepared from the same purification of His-MmsFcc, using dialysis to alter buffer conditions. BisTris, Tris and Potassium phosphate buffers were all trialled at pH 7 (Figure 3.15).

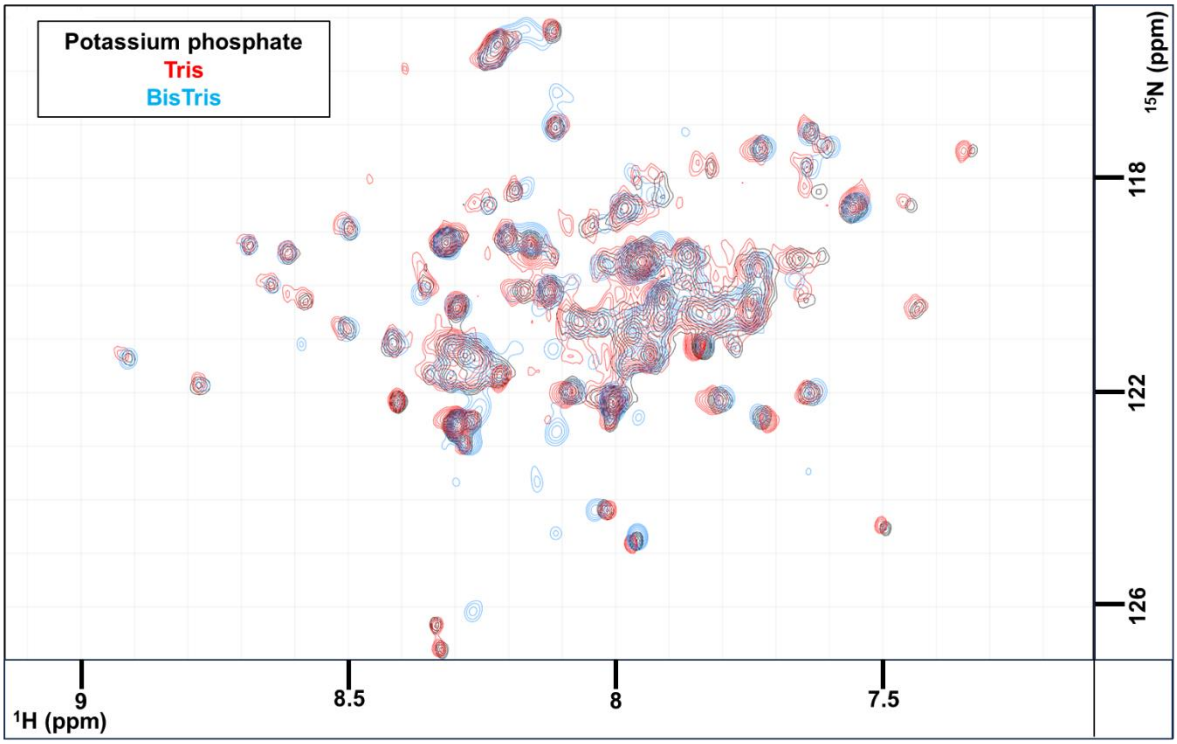


Figure 3.15. ^{1}H-^{15}N HSQC spectra of MmsFcc in 25 mM potassium phosphate pH (black), Tris (red), and BisTris (blue). All samples were at pH 6.5 and contained 100 mM NaCl.

These spectra suggest that BisTris is the best buffering component, since the peaks in its spectrum have the most uniform shape. There does not appear to be any loss of signal with its use, so BisTris was selected as the buffering component. Next, a higher concentration of NaCl was trialled:

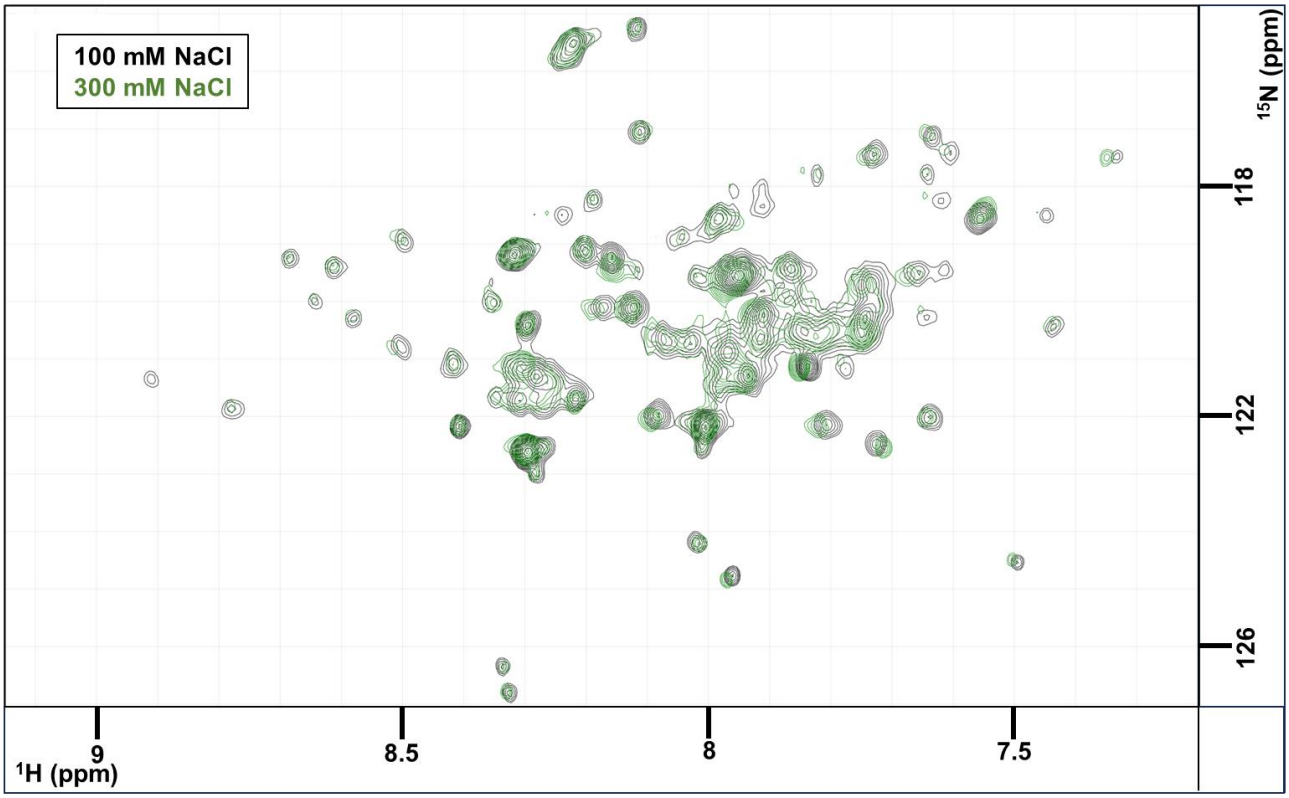


Figure 3.16. The effect of increased salt concentration on MmsFcc spectra. Both samples contain 25 mM potassium phosphate at pH 6.5. NaCl concentrations of 100 mM (black) and 300 mM (green) were trialled.

The addition of 300 mM salt resulted in significant loss of signal, while not improving the sharpness of peaks. Therefore, a lower concentration of salt (100 mM) was selected. Next, the addition of EDTA was trialled. EDTA is a metal chelator, therefore it was hypothesised that its addition would remove metal-protein interactions and improve sample quality. Addition of EDTA did not affect spectral quality, so EDTA was not added to the standard NMR buffer. Finally, the addition of trifluoroethanol (TFE) was trialled (Figure 3.17). TFE is a compound commonly used to stabilize α-helices (Corrêa & Farah, 2007; Shiraki et al., 1995)

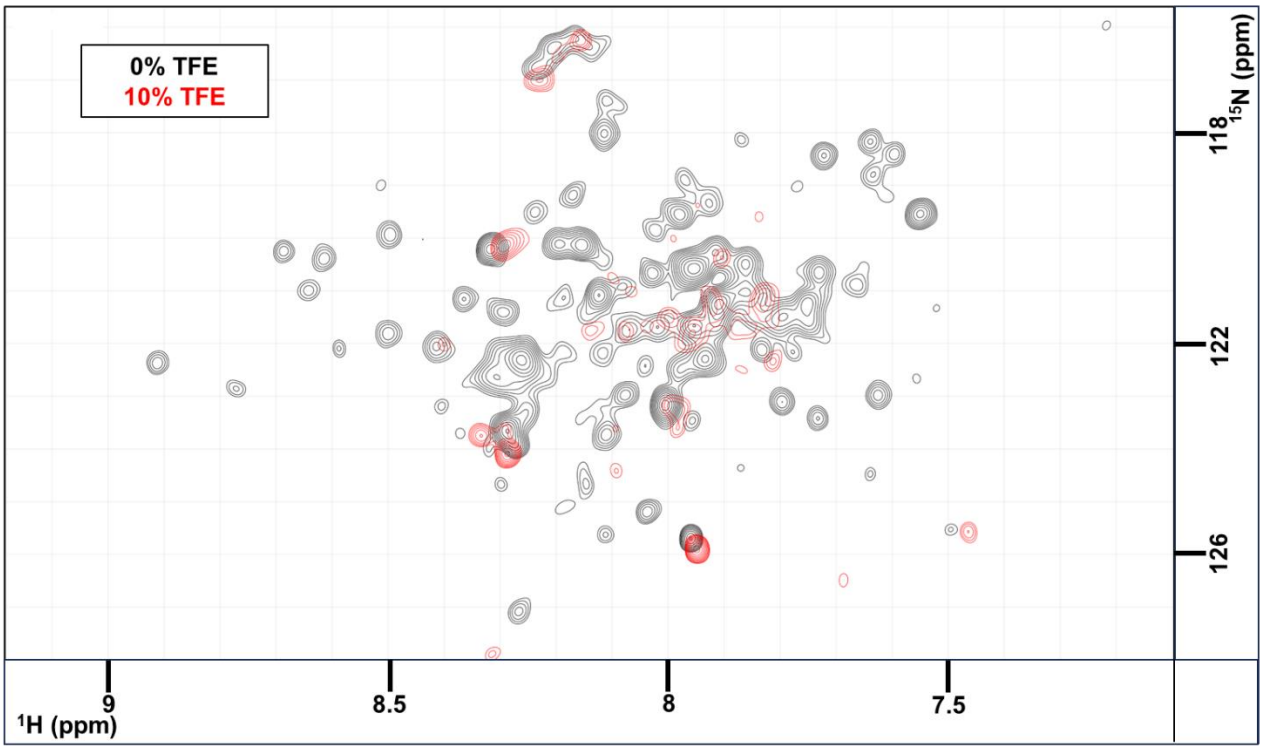


Figure 3.17. MmsFcc HSQC spectra in 25 mM BisTris pH 6.5, 100 mM NaCl, 0% TFE (black) and 10% TFE (red).

The addition of 10% TFE greatly altered the spectrum of MmsFcc, which suggests a significant change in structure. Since this did not result in an improvement to spectral quality, TFE was not selected for the final NMR buffer. Since it was planned to add sodium azide to samples, the effect of 3 mM azide was checked by HSQC. This addition of azide left the spectrum unchanged.

From these trials, the following NMR buffer conditions were established:

- 25 mM BisTris pH 6.5
- 100 mM NaCl
- 3 mM NaN_3

At this point, most realistic changes to buffer composition had been trialled. Another option for improving spectrum quality, the removal of the His-Tag, was not explored at this point. It was predicted that the small tag was not worth removing, since protein NMR often works fine with a small tag attached. The next step was to perform a backbone assignment of the His-MmsFcc backbone chemical shifts.

3.5 Triple resonance experiments show the need to remove His-Tag

A backbone assignment of His-MmsFcc was attempted. In this, a series of 3D (^{1}H-^{13}C-^{15}N) experiments were performed with the aim of assigning the HSQC spectra of MmsFcc and MmsFcc D49S. A 400 µM sample of ^{13}C ^{15}N His-MmsFcc was prepared, then a series of backbone assignment experiments were performed. The assignment was unsuccessful, owing to the lack of signal in the carbon dimensions. 3 planes from the CBCANH experiment are displayed in Figure 3.18.

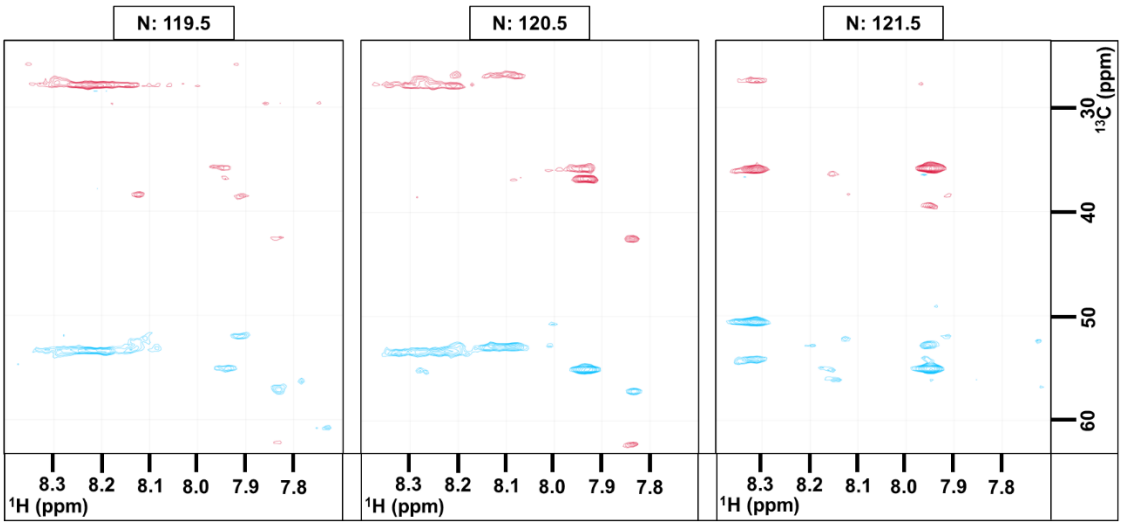


Figure 3.18. Planes from CBCANH experiment on His-MmsFcc. Sample contains 400 µM His-MmsFcc, 25 mM BisTris pH 6.5, 100 mM NaCl, 3 mM NaN_3. Red peaks represent Cβ's, while blue peaks represent Cα's.

These 3 planes represent the most densely populated region of the spectrum. The amount of signal is significantly less than expected, and the peaks present are not very sharp. It was possible to assign the N-terminal tag and this assignment is shown in Figure 3.19. How assignments are performed is discussed in detail in Sections 4.1-4.3.

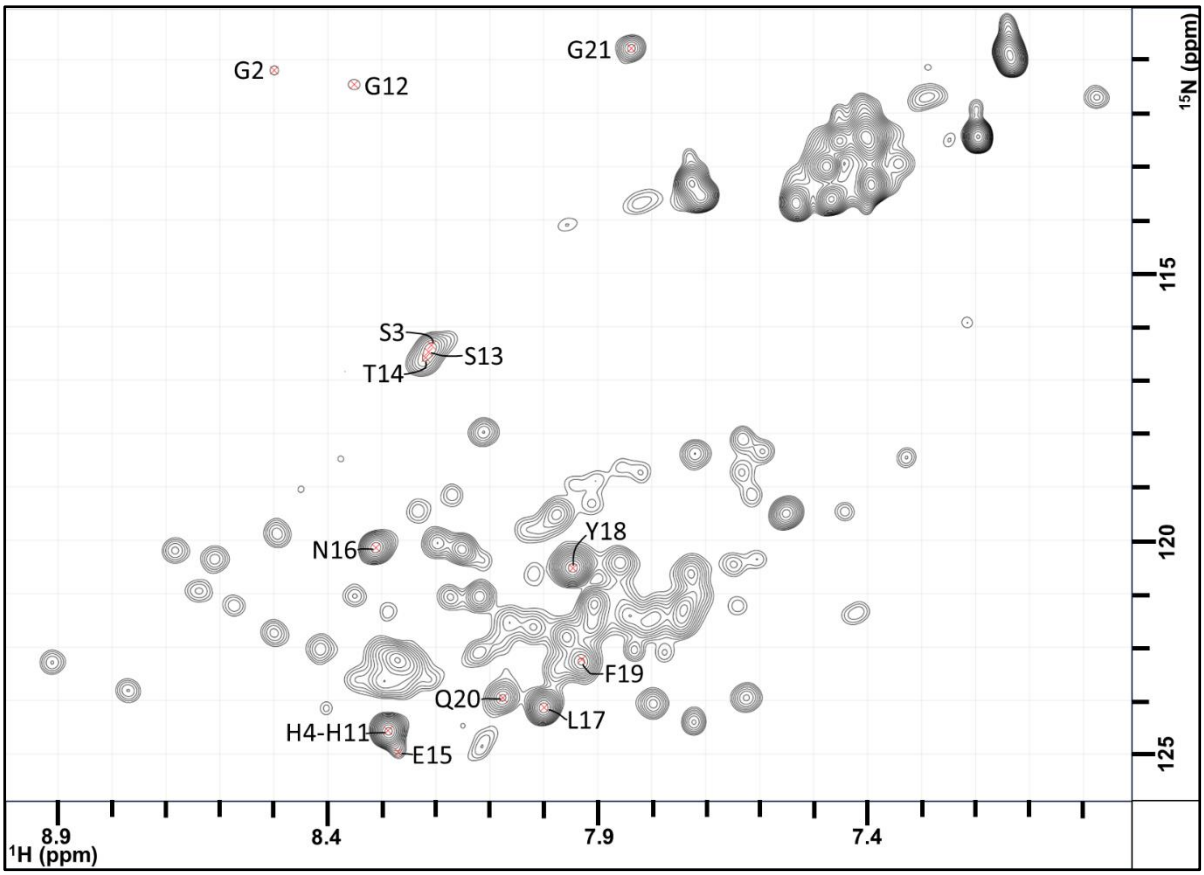


Figure 3.19. Partial assignment of His-MmsFcc 400 µM HSQC spectrum. N-terminal HisTag peaks are marked with a red cross and annotated.

It was hypothesized that the N-terminal His-tag was responsible for this by mediating intermolecular or intramolecular interactions. The nature of this interference could have been investigated by trialling the effect of concentration on the broadness of peaks. This was not considered at the time, and has not been trialled for this thesis as a result. Nonetheless, it was decided to remove the tag. The next section describes the optimization of His-tag removal by TEV protease.

3.6 TEV protease cleavage of the N-terminal His-tag

Tobacco etch virus (TEV) protease is a highly specific cysteine protease which recognises the peptide sequence TENLYQ/G, cleaving between Q6 and G7. This specificity has seen the widespread adoption of TEV as a protease for the removal of affinity tags (Nam et al., 2020; Raran-Kurussi et al., 2017a). For MmsFcc, a His-TEV construct was selected to allow for a simple clean up procedure. By passing the solution down a nickel column, the cleaved tags, His-MmsFcc and His-TEV are immobilised to the bead matrix. The resulting flow through is a

pure sample of cleaved protein. The buffer conditions for the initial cleavage attempt are listed in Table 3.1, developed from the standard TEV protease buffer conditions:

BisTris pH 6.5	25 mM
NaCl	100 mM
DTT (from freezer stock)	1 mM
EDTA	0.5 mM
MmsFcc	40 µM
TEV protease	4 µM

Table 3.1. Initial TEV cleavage sample components.

The cleavage mixture was left overnight at 4°C since TEV retains 1/10th of its activity at 4°C (when compared to its optimum temperature of 34°C), allowing for preservation of MmsFcc. After cleavage, an aliquot was taken then the rest of the mixture run down a 1 mL HisTrap column (Section 1.x). Aliquots of each step were taken, then analysed via SDS-PAGE in Figure 3.20.

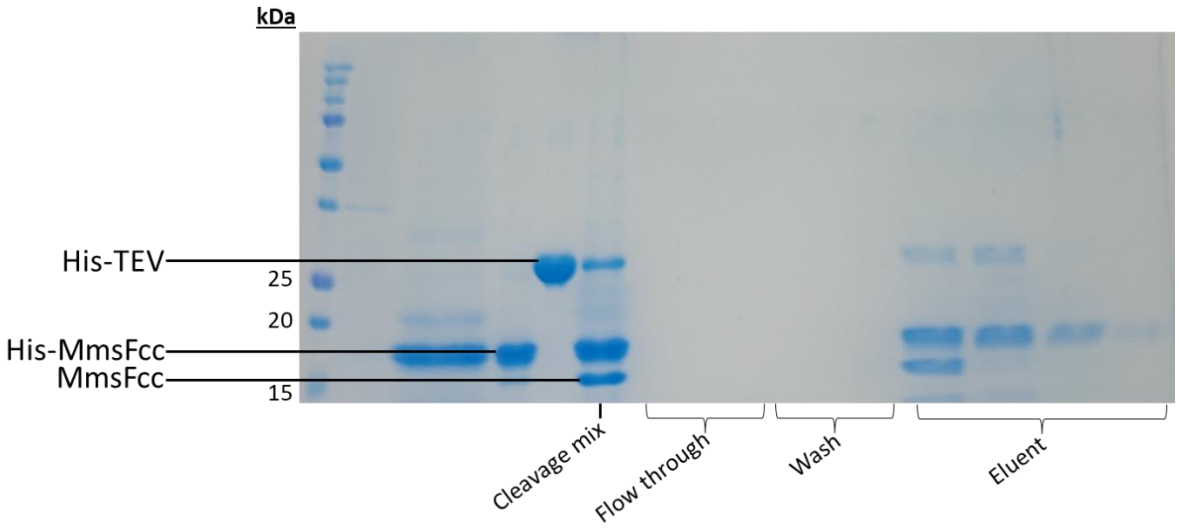
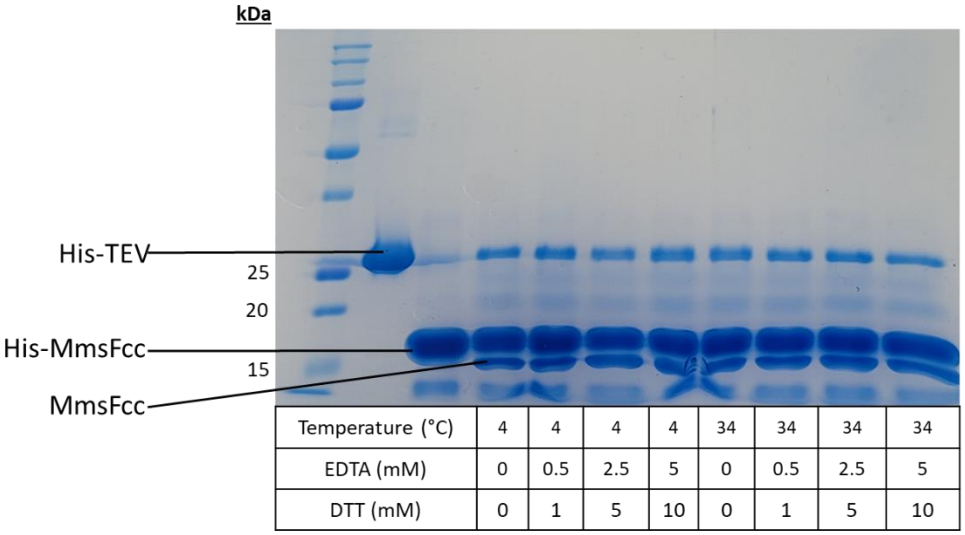


Figure 3.20. SDS-PAGE of 4°C overnight cleavage of MmsFcc by TEV, and subsequent cleanup via nickel affinity. The first lane was used for checking another sample's degradation status and is not relevant to this experiment.

From this gel, two issues can be identified. Firstly, the ratio of intact to cleaved protein is not sufficient, evidenced by the large proportion of His-MmsFcc in the cleavage mixture lane. Secondly, the cleaved MmsFcc has retained a tight affinity for the Ni-NTA matrix, only releasing when outcompeted with 0.5 M imidazole. This is strong evidence that the structured loop of MmsFcc has either a strong affinity for the Ni^{2+} cation, or that there is some dimerization keeping the cleaved protein in the matrix.

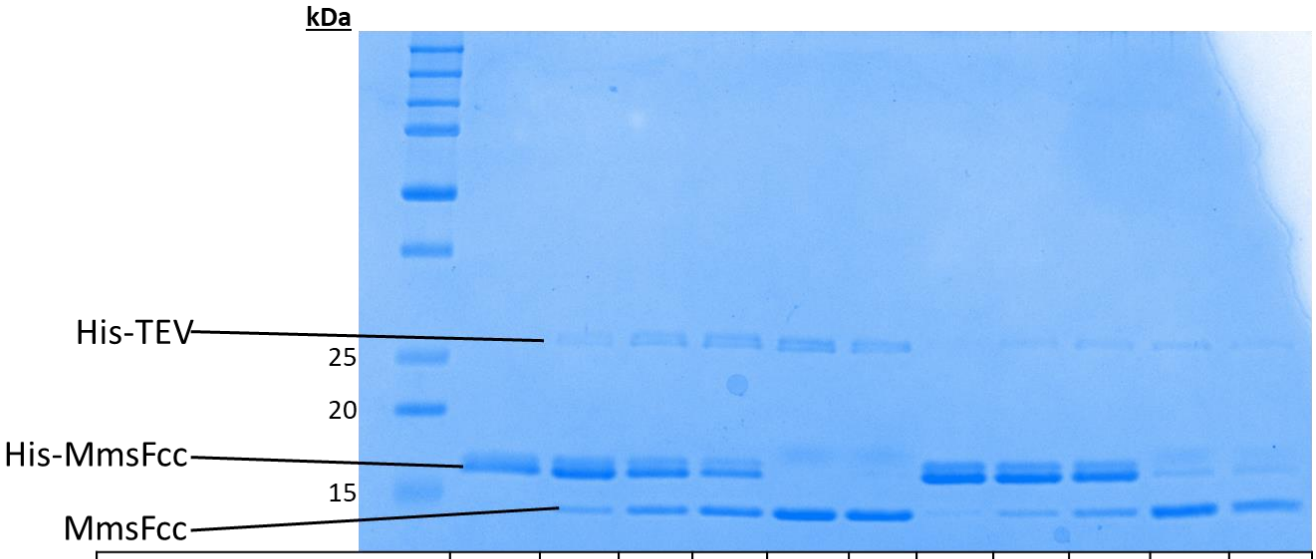
To improve enzymatic activity, a range of different conditions were trialled. The reaction temperature was also raised to 34°C, which proved to be not detrimental to protein integrity. During these initial trials, it was discovered that the cleavage reaction would not function optimally unless a freshly prepared stock of DTT was used. Figure 3.21 below shows a lack of cleavage with a freezer stock of DTT.



Temperature (°C)	4	4	4	4	34	34	34	34
EDTA (mM)	0	0.5	2.5	5	0	0.5	2.5	5
DTT (mM)	0	1	5	10	0	1	5	10

Figure 3.21. SDS-PAGE analysis of TEV activity in 25 mM potassium phosphate (pH 7.0), with varying DTT, EDTA concentrations and incubation temperatures. TEV concentration was kept at a 10:1 ratio of MmsFcc to TEV. DTT is from frozen aliquots.

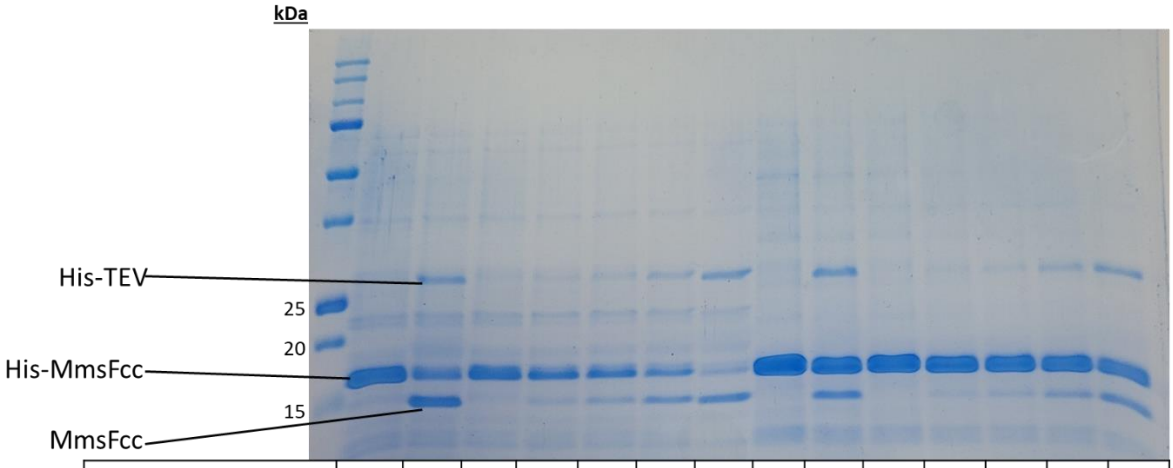
When trialling TEV protease efficiency over time, it was first discovered that fresh DTT was essential. DTT is a reducing agent that is commonly used in molecular biology to maintain cysteine sidechains in a reduced state. Its activity is not stable in solution, since it is highly reactive, and its oxidation is irreversible under normal conditions. This is so because the oxidation of DTT results in a hexagonal ring molecule pinned together by a disulphide bridge. To ensure activity, DTT solutions were prepared on the day of the cleavage. In this experiment, near total cleavage of the product is achieved in conditions comparable to those used in Figure 3.22.



MmsFcc : TEV ratio	0	5:1	5:1	5:1	5:1	5:1	10:1	10:1	10:1	10:1	10:1
Hours post TEV addition	0	2	4	6	8	20	2	4	6	8	20

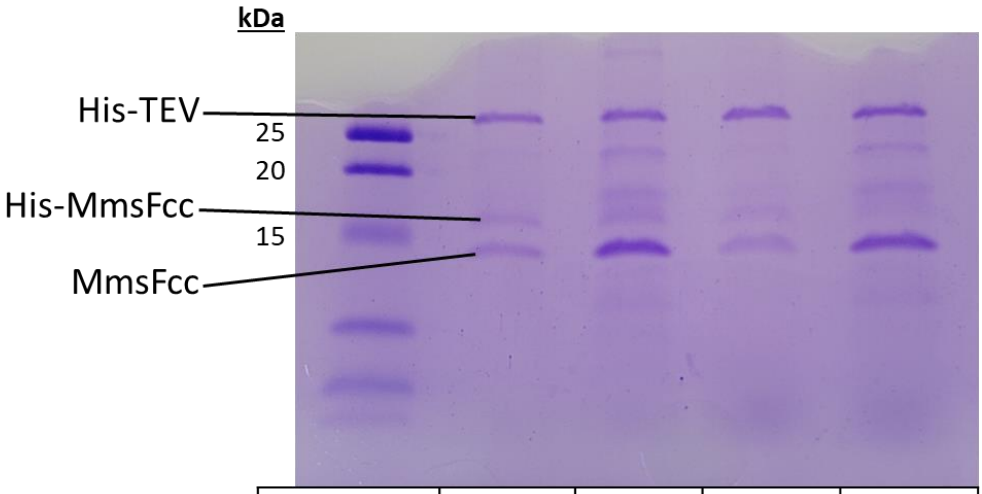
Figure 3.22. SDS-PAGE analysis of TEV activity in 25mM potassium phosphate (pH 7.0), 1 mM EDTA, 5 mM DTT at different timepoints post addition of TEV protease. All cleavages performed at 34°C, with freshly made DTT.

When the 34°C lanes from Figure 3.21 are compared to the 10:1 lanes from Figure 3.22, it can be seen that cleavage is greatly improved with the switch to freshly prepared DTT. From Figure 3.22, it was observed that TEV protease's activity did not deteriorate over time. It did however take at least 8 hours for sufficient cleavage. Therefore, it was decided to perform cleavages overnight. Now that this was established, TEV's activity in different buffers were assayed. Specifically, potassium phosphate, BisTris and Tris all at pH 7. The effect of GuHCl concentration on cleavage efficiency was also examined. An aliquot of protein was taken after only one round of dialysis, where the GuHCl concentration was roughly 130 mM. The rest of the cleavage conditions were trialled with protein subject to two rounds of dialysis, with an estimated 0.16 mM GuHCl concentration. These results are displayed in Figures 3.23-3.24.



MmsFcc : TEV ratio	0	10:1	0	100:1	50:1	25:1	10:1	0	10:1	0	100:1	50:1	25:1	10:1
Rounds of dialysis	1	1	2	2	2	2	2	1	1	2	2	2	2	2
25 mM kPi	✓	✓	✓	✓	✓	✓	✓	×	×	×	×	×	×	×
25 mM BisTris	×	×	×	×	×	×	×	✓	✓	✓	✓	✓	✓	✓

Figure 3.23. Comparison of TEV activity in potassium phosphate and BisTris buffers. All reactions performed at 34°C overnight.



DTT (mM)	10	5	10	5
25 mM kPi	✓	✓	×	×
25 mM Tris	×	×	✓	✓

Figure 3.24. Comparison of TEV activity in potassium phosphate and Tris buffers. All reactions performed at 34°C overnight.

These results show that TEV activity was mostly the same in Tris or potassium phosphate, but that cleavage was reduced when buffered by BisTris. Figure 3.23 also shows that a GuHCl concentration of ~130 mM reduced activity significantly, meaning that the 2-step dialysis process was necessary for maximum cleavage. From these trials, Tris was selected as the buffer compound since it was already used in the purification stage.

-

Attention was next turned to the clean-up of the cleavage reaction. As seen in Figure 3.20, cleaved protein remained stuck to the nickel column under native conditions.

One option was to denature the solution by addition of 8 M GuHCl, then to pass down the nickel column. This was not the ideal solution, as loss of protein was predicted to be high due to the extra transfers and steps this protocol would require. As a result, other solutions were explored first. One method attempted was size exclusion chromatography (SEC). SEC was an attractive option, since it allowed for simultaneous clean up and buffer exchange into BisTris. Superdex 75 resin was selected, due to its separation range of 3,000-72,000 Daltons. two different column dimensions were also trialled. The first was a prepacked 10/300 mm column. The second was a self-poured 10/500 mm column. A cleavage mixture of MmsFcc was first concentrated to a volume of 2 mL. This was then split into two, and each half ran down the two columns.

For the 10/300mm column, a flow rate of 1 ml/min was used. 0.4 mL fractions were collected. The attempt was analysed by SDS-PAGE in Figure 3.25.

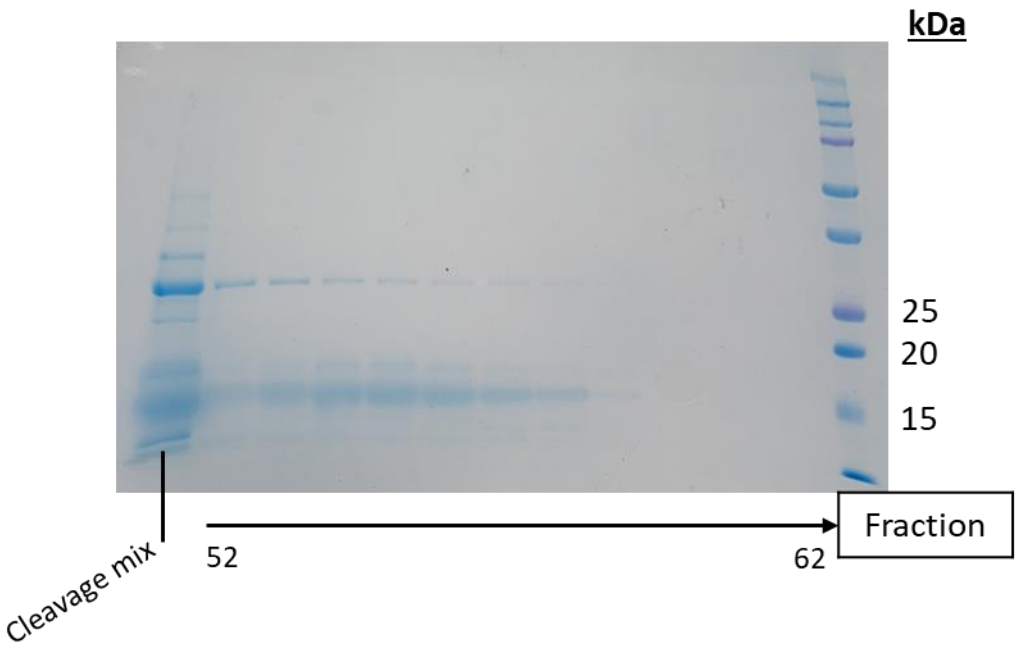


Figure 3.25. SDS-PAGE analysis of attempted MmsFcc TEV cleavage cleanup, using a 10/300 mm Superdex 75 prepacked SEC column. Flow rate = 1 mL/min, fraction volume = 0.4 mL.

With this attempt, separation was not even sufficient to remove TEV from the sample. For the 10/500 mm column, a flow rate of 1 mL/min was used. A fraction volume of 1 mL was selected, due to the larger column volume. The results were analysed by SDS-PAGE in Figure 3.26.

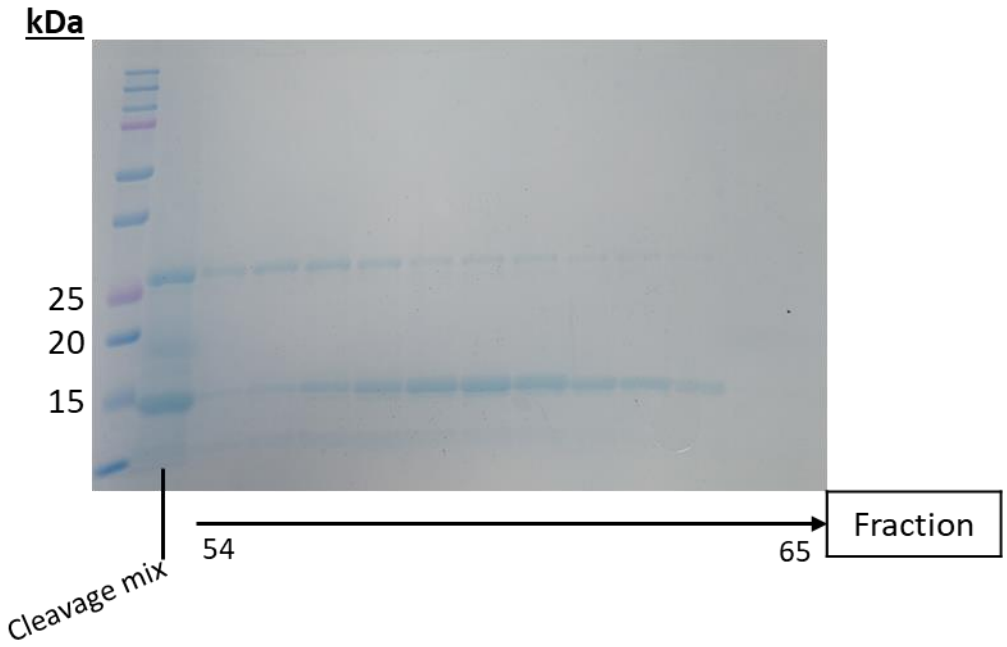


Figure 3.26. SDS-PAGE analysis of attempted MmsFcc TEV cleavage cleanup, using a x/x Superdex 75 selfpoured SEC column. Flow rate = 1 mL/min, fraction volume = 1 mL.

The change in column dimensions did somewhat improve separation when compared to Figure 3.25. However, it was still not sufficient. Furthermore, around 75% of protein was lost in this trial. Any attempt to improve separation by increasing the column length would likely result in more loss, so SEC was abandoned.

-

A denatured cleanup was performed on a cleavage reaction containing 8.045 mg of MmsFcc, this protocol is described in Section 2.7. Afterwards, 3.83 mg of cleaved protein was left in solution. This roughly 50% loss was favourable when compared with the 75% loss observed when using size exclusion. The resulting chromatograph is displayed in Figure 3.27.

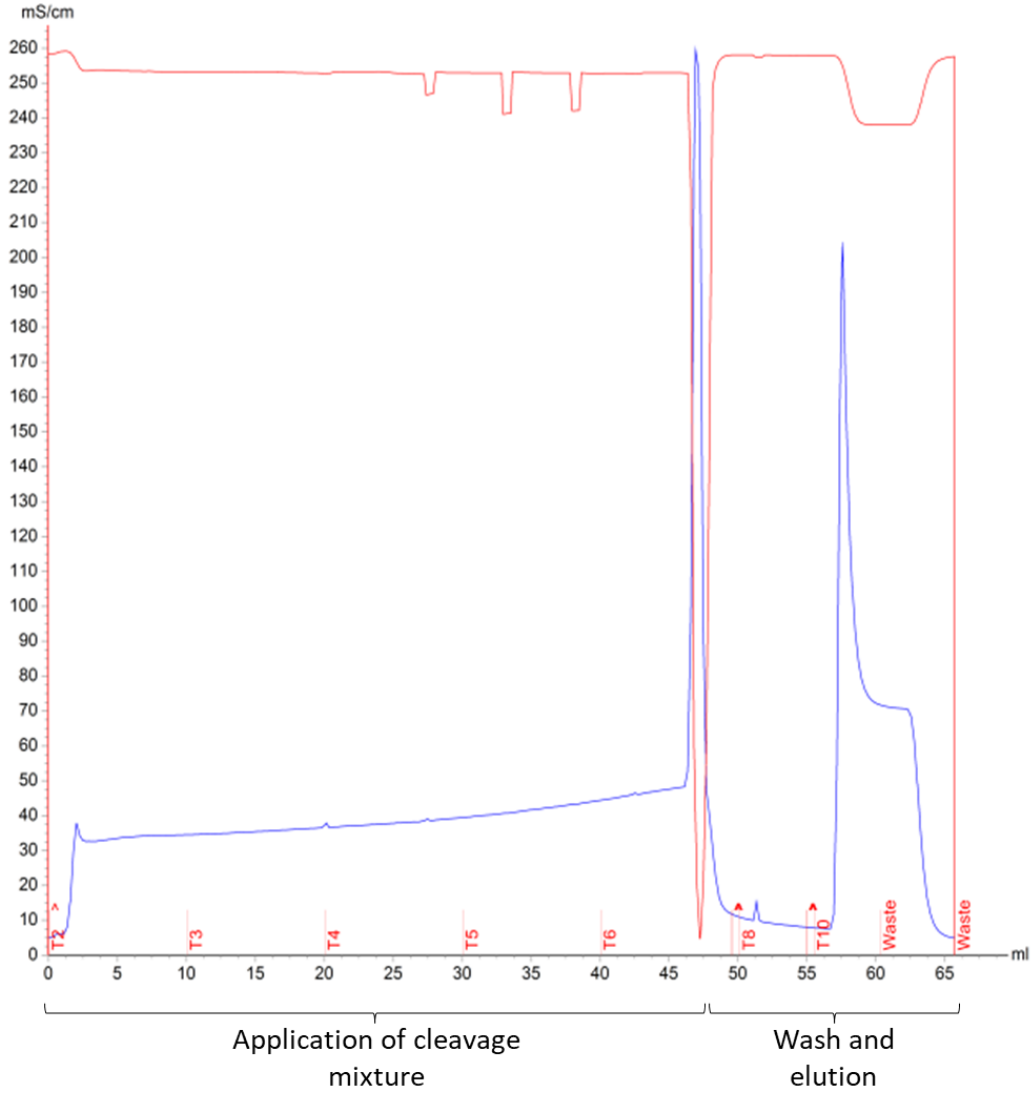


Figure 3.27. AKTA start chromatograph of denatured cleanup of TEV cleavage reaction. Large spike in sample application is result of air getting onto column, from allowing the sample valve to be open for too long.

A native clean-up of a fully cleaved sample was not trialled. In hindsight, this was an error since the tendency of MmsFcc to stick to the nickel column could have been a result of dimerization with His-MmsFcc. Nonetheless, it was now possible to produce decently pure and concentrated samples of cleaved MmsFcc and MmsFcc D49S. Below is a summary of the fully optimised protocol:

1. Freshly transform BL21 DE3 and select on ampicillin plates.
2. Prepare minimal media and grow starter cultures overnight.
3. Inoculate then grow until an OD600 of 0.6 is reached. Induce expression with IPTG, then incubate cells overnight at 30°C.

4. Pellet cells, then lyse by sonication. Centrifuge sonicate to separate insoluble fraction.
5. Resuspend insoluble fraction in 8M GuHCl, then mix at room at temperature.
6. Centrifuge to remove lipids and other insoluble, then filter supernatant.
7. Perform Ni-NTA purification under denaturing conditions.
8. Dilute then dialyse sample twice to fully remove GuHCl.
9. Cleave with TEV protease.
10. Concentrate sample, then add GuHCl to a 6M concentration.
11. Perform Ni-NTA cleanup under denaturing conditions.
12. Dialyse cleaved fraction twice to fully remove GuHCl.
13. Concentrate dialysed protein.

The evidence showing that the removal of the HisTag improved spectral quality is shown in Section 3.7.

3.7 Removal of His-Tag improved the spectral quality of three-dimensional experiments

The next step was to prepare a concentrated, double-labelled sample of cleaved protein for backbone assignment experiments (Chapter 4). This was initially done with MmsFcc D49S, since it expressed better than the wildtype. However, to allow for comparison with the partially assigned spectrum of His-MmsFcc spectrum, data from a 400 µM $^{13}C^{15}N$ MmsFcc sample has been used for this Section. This sample was prepared following the fully optimised protocol described in Section 2.7. SDS-PAGE was performed on the 400 µM sample of MmsFcc to assess purity (Figure 3.28).

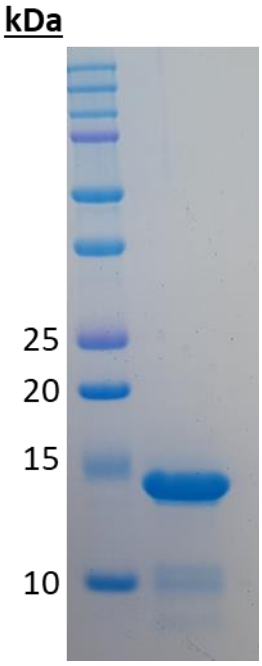


Figure 3.28. SDS-PAGE analysis of 400 µM NMR sample of MmsFcc.

At this concentration, the contaminant band first present in figures 3.3-3.4 is visible. This is investigated in section 3.8 and is determined to likely be a degradation product of MmsFcc. Besides this band, the sample is concentrated and pure.

A comparison of ^{1}H-^{15}N HSQC spectra for His-MmsFcc and MmsFcc validates SDS-PAGE data confirming the cleavage (Figure 3.29).

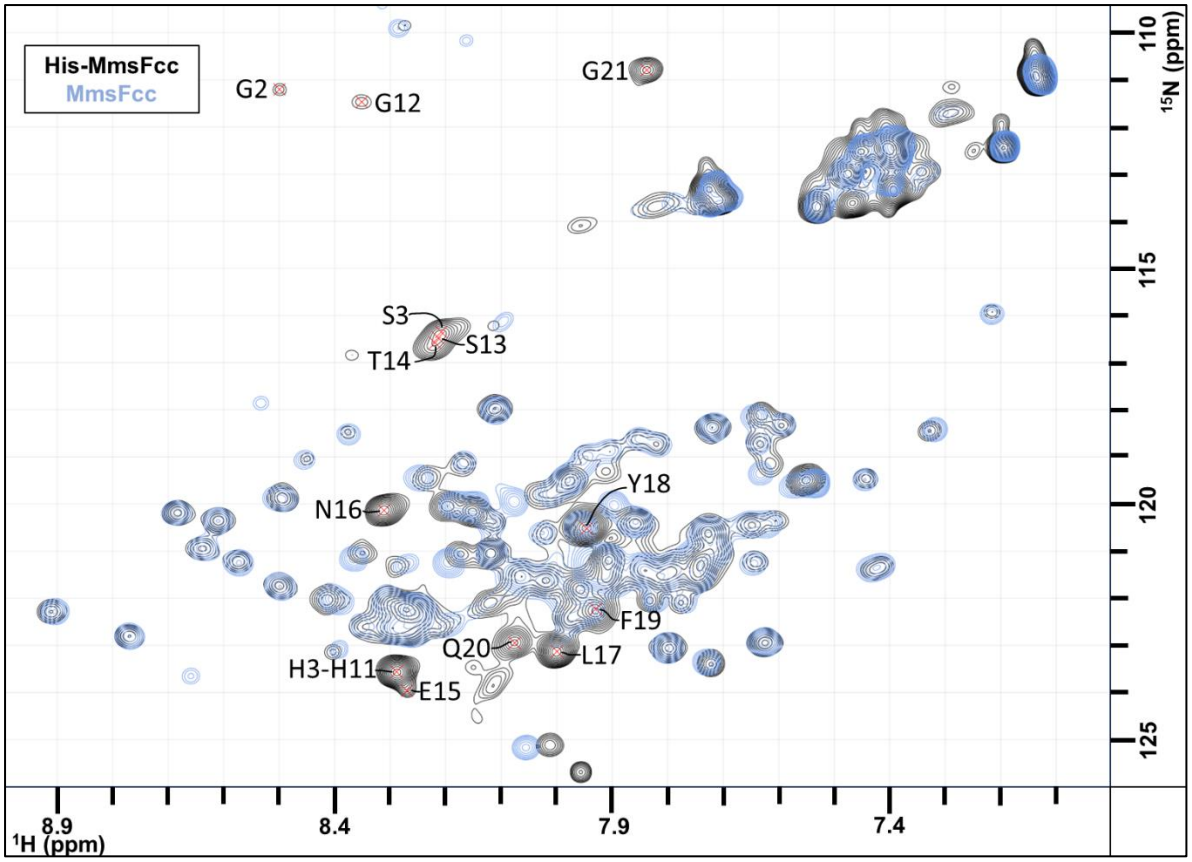


Figure 3.29. HSQC spectra of His-MmsFcc (black) and MmsFcc (blue) overlaid. N-terminal tag peaks are assigned.

In Figure 3.29, all peaks representing the N-terminal tag from the His-MmsFcc spectrum are not present in the MmsFcc spectrum, showing the cleavage to be successful.

The removal of the tag greatly improved the quality of spectrum produced. This was most evident in the improvement to the uniformity of peak intensities. This is visualised by the box and whisker plot shown in Figure 3.30. For these calculations, every backbone amide peak height was recorded. Then a mean and standard deviation was calculated from the full dataset. The two datasets were then normalised to have the same mean intensity. This was necessary to be able to compare the 2 datasets, since MmsFcc had significantly more intense peaks.

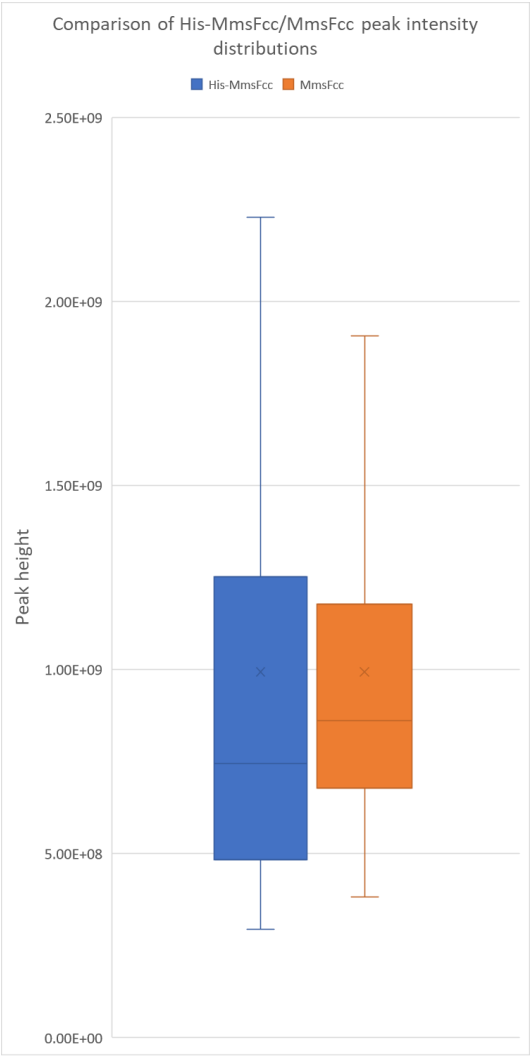


Figure 3.30. Box and whisker plot comparing the distributions of peak intensities between His-MmsFcc and MmsFcc (including His-Tag peaks). Both datasets were collected at the same concentration (~400 µM) on the same magnet with the same number of scans (4 scans). Both datasets were processed identically, and both datasets normalised to have the same mean intensity.

In Figure 3.30, the distribution of peak intensities is smaller when the His-Tag is removed. This improvement in spectral quality is likely the result of reduced interactions involving the HisTag. As previously mentioned, there is also an improvement to signal strength when the tag is cleaved. By overlaying these two spectra, then matching the contour levels, this improvement becomes visually clear. This is represented in Figure 3.31. It should be noted that, despite using similar concentrations and collection times etc, different acquisition parameters (such as receiver gain) can greatly affect signal strength. Nevertheless, the difference is so apparent that any perceived improvement is likely to be a result of a better-behaved protein.

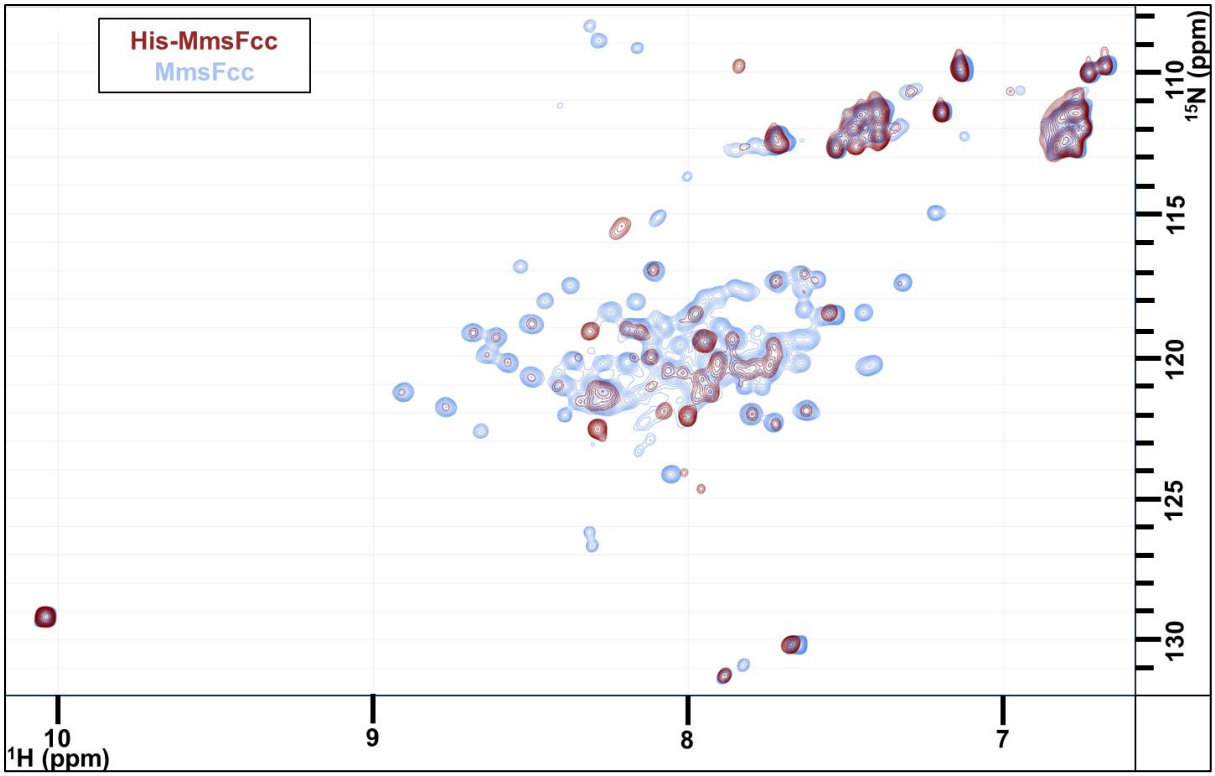


Figure 3.31. His-MmsFcc and MmsFcc HSQC spectra overlaid with contour levels matched.

Figures 3.30 and 3.31 show significant improvement of HSQC spectrum quality. The next step was to assign the backbone shifts using data collected from cleaved MmsFcc. All assignments are described in detail in Chapter 4. The next section goes into detail about the contaminant band observed in SDS-PAGE gels of MmsFcc and MmsFcc D49S.

3.8 Analysis of contaminant band shows it is a degradation product of MmsFcc

It was noticed early on that there was a protein contaminant in samples of purified MmsFcc. This was first seen in SDS-PAGE gels, as a small band around the 10 kDa marker. It was hypothesized that this was some degradation product occurring during the expression or purification of MmsFcc. It was present in both wildtype and D49S samples, as evidenced in Figures 3.32 and 3.3. Once TEV protease cleavage had been performed, there was stronger evidence that this contaminant came from MmsFcc degradation (Figure 3.32).

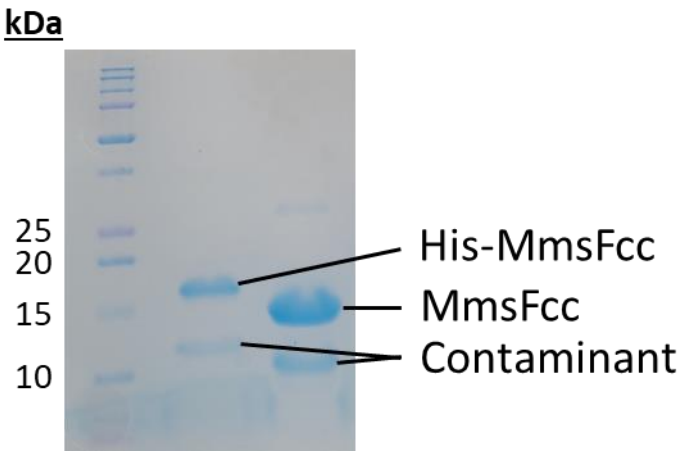


Figure 3.32. SDS-PAGE of MmsFcc D49S sample after cleavage, cleanup, and refolding.

In this gel, it appears that cleavage by TEV protease has caused both the MmsFcc and contaminant bands to experience the same reduction in size. This suggests that the tag is being removed from both protein species, suggesting that the contaminant band is a degradation product of MmsFcc containing the N-terminal His-Tag. These results were corroborated by liquid chromatography-mass spectroscopy (LCMS). The spectrum is displayed in Figure 3.33. In this, a ^{15}N sample of MmsFcc D49S was submitted.

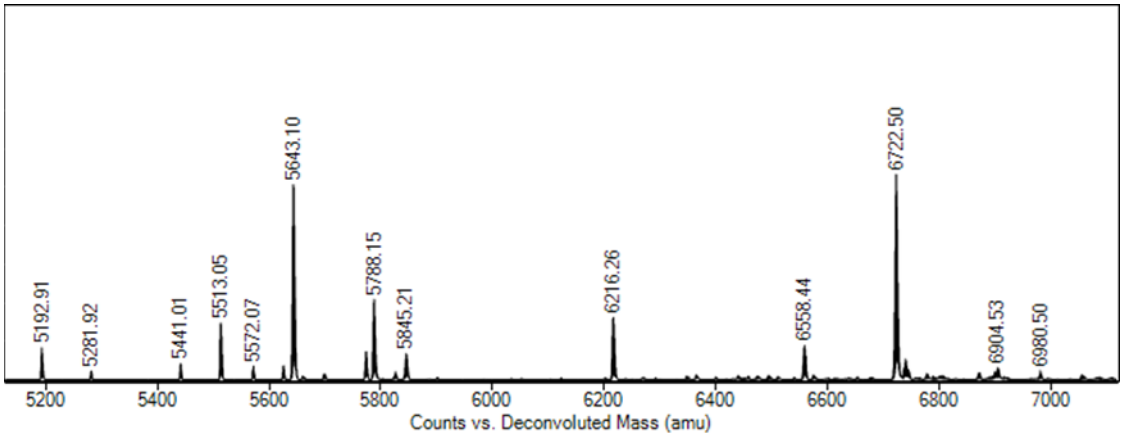


Figure 3.33. LCMS Spectrum of D49S, displaying the region containing contaminant band. Y-axis represents counts, a measure of the amount of a molecule was detected. It has been excluded since data provided by the Mass spec facility did not include this measure.

Figure 3.34 shows the predicted mass of potential degradation products of MmsFcc D49S match the deconvoluted mass of peaks observed in Figure 3.33. Sine this degradation product is observed from the initial purification of MmsFcc, it is highly likely that the degradation products observed originate from the N-terminal end of the protein. This is so as this fragment would contain the N-terminal HisTag, which would allow for the fragment to

be carried through the purification. Mass calculations of D49S fragments were done assuming a 99% population of ^{15}N.

6558 Dalton fragment: GPSMKQLEKELKQLEKELQAIEKQLAQLQWKAQARKKKLAQLKKKLQASRDDEFVY
5643 Dalton fragment: GPSMKQLEKELKQLEKELQAIEKQLAQLQWKAQARKKKLAQLKKKLQ
5192 Dalton fragment: PSMKQLEKELKQLEKELQAIEKQLAQLQWKAQARKKKLAQLKK
5788 Dalton fragment: PSMKQLEKELKQLEKELQAIEKQLAQLQWKAQARKKKLAQLKKKLQAS
5845 Dalton fragment: MKQLEKELKQLEKELQAIEKQLAQLQWKAQARKKKLAQLKKKLQASRD

Figure 3.34. Degradation products of ^{15}N D49S that match the mass of peaks in figure 3.33.

It was decided to not pursue removing the contaminant, since there was not a viable route for doing so given previous issues with size exclusion chromatography. It was hypothesized that the contaminant shouldn't interfere with results too much, due to its significantly weaker concentration and lack of a structured loop.

3.9 Chapter discussion

The optimization of MmsFcc and MmsFcc-D49S purification was a challenge. Originally predicted to be a roughly one-year endeavour, it wasn't until till the start of the 3rd year that a cleaved, double labelled sample of MmsFcc D49S sufficiently concentrated for assignment was produced. A major contributor toward slowing down this research was the slow growth rate of cells transformed with MmsFcc and its mutants. As mentioned above, 1 litre cultures would often take up to 12-15 hours to reach a sufficient OD600 for IPTG induction. We hypothesize that this is due to MmsFcc likely being toxic to the cell, due to its metal binding activity. The basal expression repressing strain BL21(DE3) pLysS was trialled, but ultimately had no effect.

Poor expression also contributed to this delay. For the first year, only wildtype MmsFcc was worked on. From the plasmids inherited from previous researchers of this project, it was only possible to sequence the mutants D49S and F54Y. Although these weren't the most interesting mutants, the initial purification of D49S yielded three times more protein (~10 mg of His-D49S) than the best attempts of wildtype at the time. It was hypothesized that this was due to D49S being less toxic than wildtype, evidenced by its reduced capacity to mediate nanoparticle synthesis *in vitro*. As a result, a lot of the optimization work was done on D49S, and the first successful assignment was of D49S (Section 4.2). Eventually, once

purification was fully optimized, 500 mL of wildtype culture yielded 3.82 mg of cleaved protein while 500 mL of D49S culture yielded 4.45 mg on average.

As discussed in this chapter, the choice of a His-Tag was potentially a mistake in hindsight. Specific issues such as the metal binding loop sticking to the nickel column have been highlighted in this section. Anecdotally, MmsFcc behaved much better with the tag removed, in reference to the protein's propensity to stick to surfaces or precipitate. It is however understandable why this tag was selected. It provides a simple and cheap purification system, with columns being easily recharged and reused. Alternative options for purification strategies are as follows:

- Cation exchange chromatography of MmsFcc. MmsFcc has a pKa of 10, meaning that it carries an overall positive charge at neutral pH. This means it could potentially be purified using cation exchange chromatography. This method was not explored, since
- SEC. This was trialled in the project and proved to be not viable with MmsFcc.
- Untagged purification by nickel affinity. MmsFcc shows an affinity for Ni^{2+}, evidenced by Figure 3.23. Therefore, untagged MmsFcc could have been refolded after extraction, then run down a nickel column under native conditions. The sample would then have to be centrifuged and filtered to remove all insoluble components, which would remove a significant fraction of precipitated protein.
- Epitope tags such as FLAG and c-myc provide high specificity and would likely not mediate interactions in the same way as His. This approach was not viable for this project, due to the cost of the antibody columns required.
- Small protein affinity tags, such as STREP are viable options. Their columns are cheaper than epitope tags but have less reusability and protein binding capacity than the nickel columns.
- Large protein affinity tags, such as maltose binding protein have cheaper and more robust substrate-based affinity columns. These larger tags also help to express misfolding proteins into the soluble fraction, potentially removing the need for MmsFcc extraction by GuHCl. It could be argued that, after cleavage, there would be no need to remove the free tag protein from the NMR sample. This is so as the protein would not be visible due to its large size (for example, maltose binding

protein is 42.5 kDa). This was not considered during the completion of this project, and would have likely been an attractive alternative to Ni-NTA purification.

In the end, the decision was made to stick to the system inherited, since any change would have resulted in a large set up cost and the re-establishing of expression conditions. The next chapter described the successful backbone assignments of MmsFcc and MmsFcc D49S, and the successful sidechain assignment and subsequent elucidation of MmsFcc structure.

Chapter 4: Backbone and side-chain assignments of MmsFcc were used to determine the structure of the functional loop

This chapter describes the successful backbone and sidechain assignments of MmsFcc and MmsFcc D49S (Sections 4.1-4.3). These assignments were then used to determine the structure of MmsFcc, using experimentally determined NOE distance restraints in molecular dynamics calculations (Sections 4.5-4.6). The process of assigning spectra involves determining the specific residue and atom that peaks represent, which is done with a series of three-dimensional (3D) experiment pairs. An assigned HSQC spectrum is required for many analytical protein NMR techniques, such as titration assays (Chapter 5) and structural determination.

AlphaFold2 predictions of MmsFcc and MmsFcc D49S are used in Chapter 4 and Chapter 5, to assist with the creation of figures, and for the fixing of the coiled-coiled display scaffold in structural calculations (Section 4.5). The protocol for AlphaFold2 predictions is described in Section 4.12.

The sequence of MmsFcc was challenging for assignment, due to the presence of 2 nearly identical α-helix sequences (Figure 4.1).

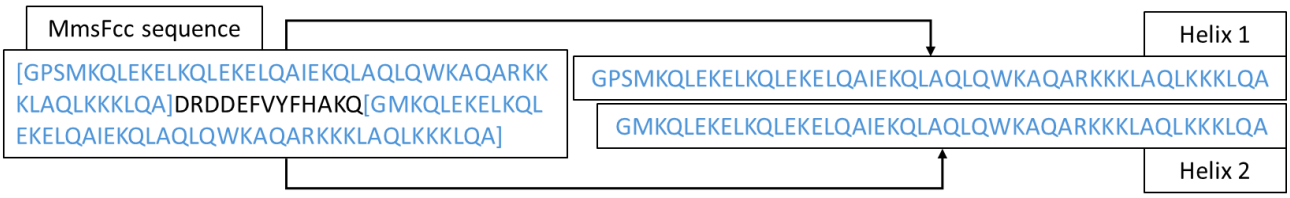


Figure 4.1. The two near identical coil sequences in MmsFcc.

It was predicted that equivalent residues (e.g. W30/W89) in the two identical α-helices would likely have the same chemical shift values, due to the residues having near identical chemical environments (Figure 4.2). This meant that some HSQC peaks would represent two helical residues.

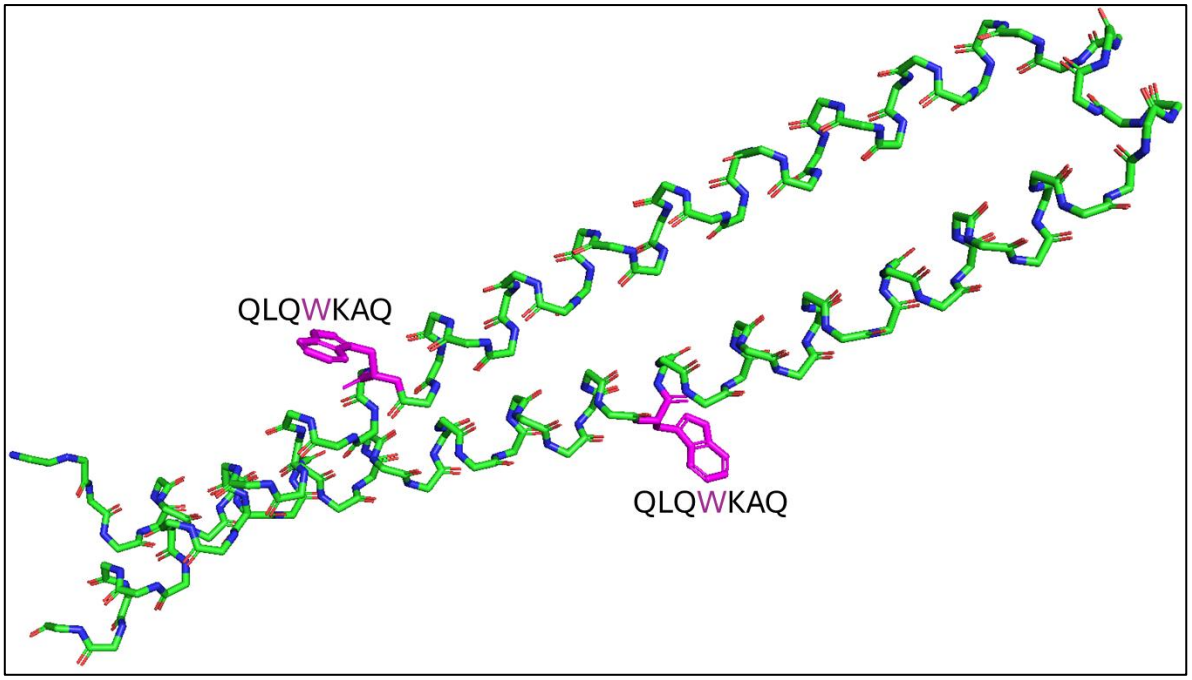


Figure 4.2. AlphaFold2 predicted structure of MmsFcc. W30 and W89 (purple) are predicted in a near identical chemical environment, resulting in identical chemical shift values.

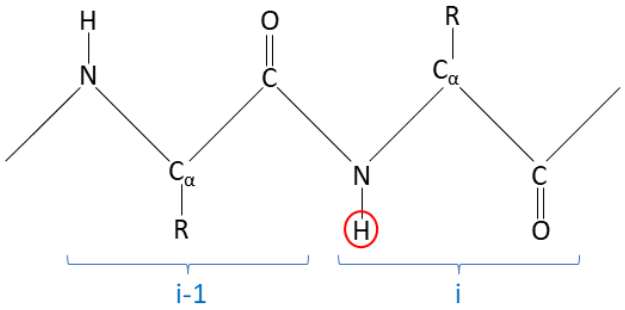
Whilst being an unfortunate side-effect of the protein's design, it was ultimately not detrimental to the goals of this thesis. While it was not possible to observe the behaviour of the coiled-coil section of the protein, it was still likely that the functional metal binding loop could be studied in detail. The process of doing this started with the assignment of the ^{15}N HSQC spectrum, the background of which is explained in Section 4.1. A backbone assignment was first attempted on MmsFcc D49S because the mutant yields more protein than the Wildtype.

Further information on the pulse programmes used, and the number of scans collected can be found in Section 2.18.

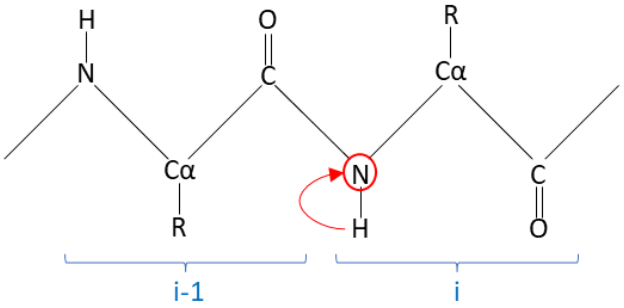
4.1 Introduction to backbone assignments

The standard approach to assigning proteins the size of MmsFcc requires labelling with ^{13}C and ^{15}N isotopes (Hu et al., 2021a). This labelling allows for the observation of nitrogen and carbon nuclei. To assign MmsFcc backbone chemical shifts, a series of triple resonance (^{1}H-^{13}C-^{15}N) NMR experiments were performed. A simplified visualisation of the HNCA experiment is displayed in Figure 4.3

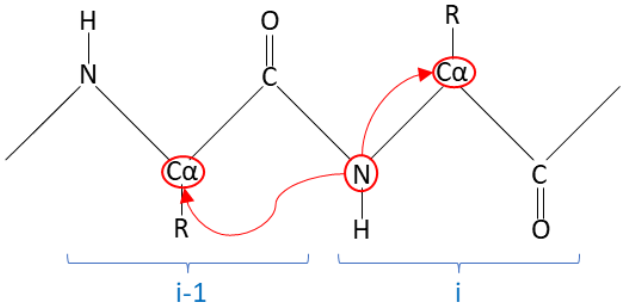
1) Magnetization is excited on the proton



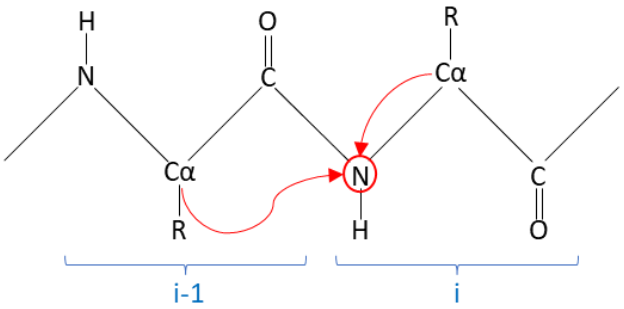
2) Magnetization is then transferred to the amide nitrogen. A delay (t_1) is included to allow the nitrogen to precess



3) Magnetization is then transferred to both the i and i-1 α-carbons. Another delay (t_2) is included to allow carbons to precess



4) Magnetization is then returned to both the amide nitrogen



5) Finally, magnetization is returned to the proton for direct detection

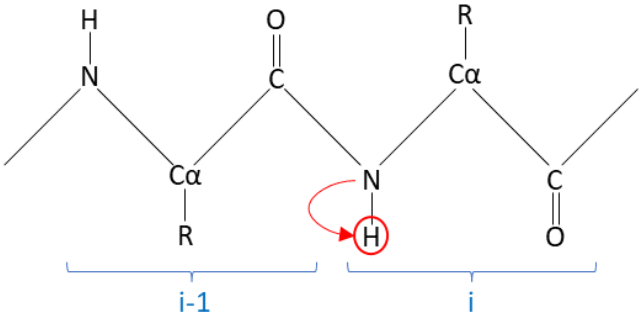


Figure 4.3. Simplified diagram of HNCA magnetisation transfer steps.

To reduce experiment run time, Non-uniform sampling (NUS) can be used. In an experiment with NUS, a fraction of the full dataset is collected. NUS algorithms then use this data to predict the uncollected data (Gołowicz et al., 2020). NUS is typically performed on larger proteins, or on proteins which are unstable at room temperature. Since MmsFcc was neither of these things, and spectrometer availability wasn't limited, NUS was not used in any MmsFcc experiments.

Data from 3D experiments is displayed in strips, which is shown in Figure 4.4 with an HNCA spectrum.

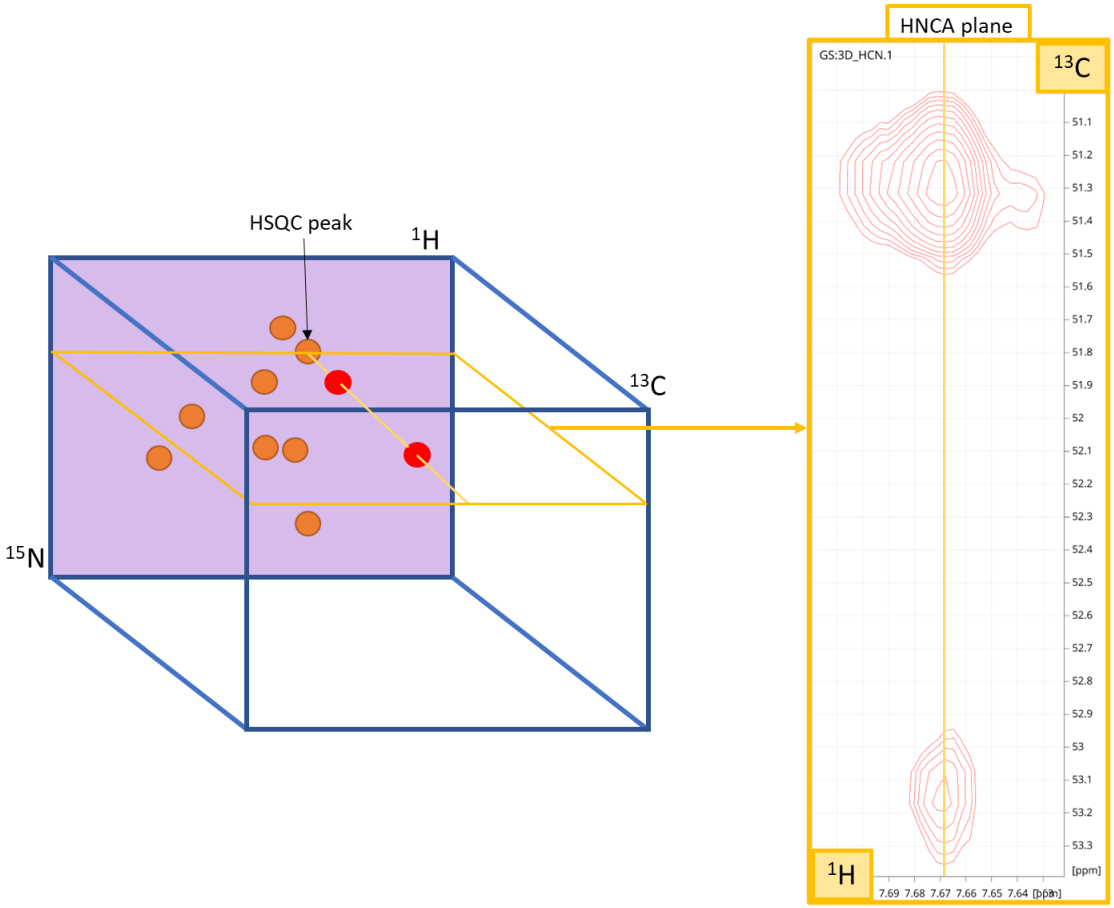


Figure 4.4. How strips are used to interpret a three-dimensional NMR spectrum. The purple plane represents a nitrogen HSQC, with orange circles representing HSQC peaks. The yellow section shows a ^{15}N plane with red HCα peaks. Data is from a HNCA experiment run on 400 µM MmsFcc in 25 mM BisTris pH 6.5, 100 mM NaCl.

In this strip, the two carbon frequencies and the amide proton frequency are defined as being in the same spin system. If the NH HSQC peak is from residue $_i$, then the carbon peaks in the same spin system represent the Cα_i and Cα_{i-1} nuclei. For assignments, triple resonance experiments are paired to allow for easier determination of C$_i$ and C$_{i-1}$. In this case, the HNCA

is paired with an HN(CO)CA spectrum. The HN(CO)CA only collects $C\alpha_{i-1}$ carbon shifts, so by pairing the HNCA and HN(CO)CA experiments, it is possible to determine which peak represents each carbon. This can be seen in Figure 4.5, in the purple plane at the carbon frequency of 54.6 ppm. Knowing this, ppm values can then be matched to begin linking HSQC peaks in sequence.

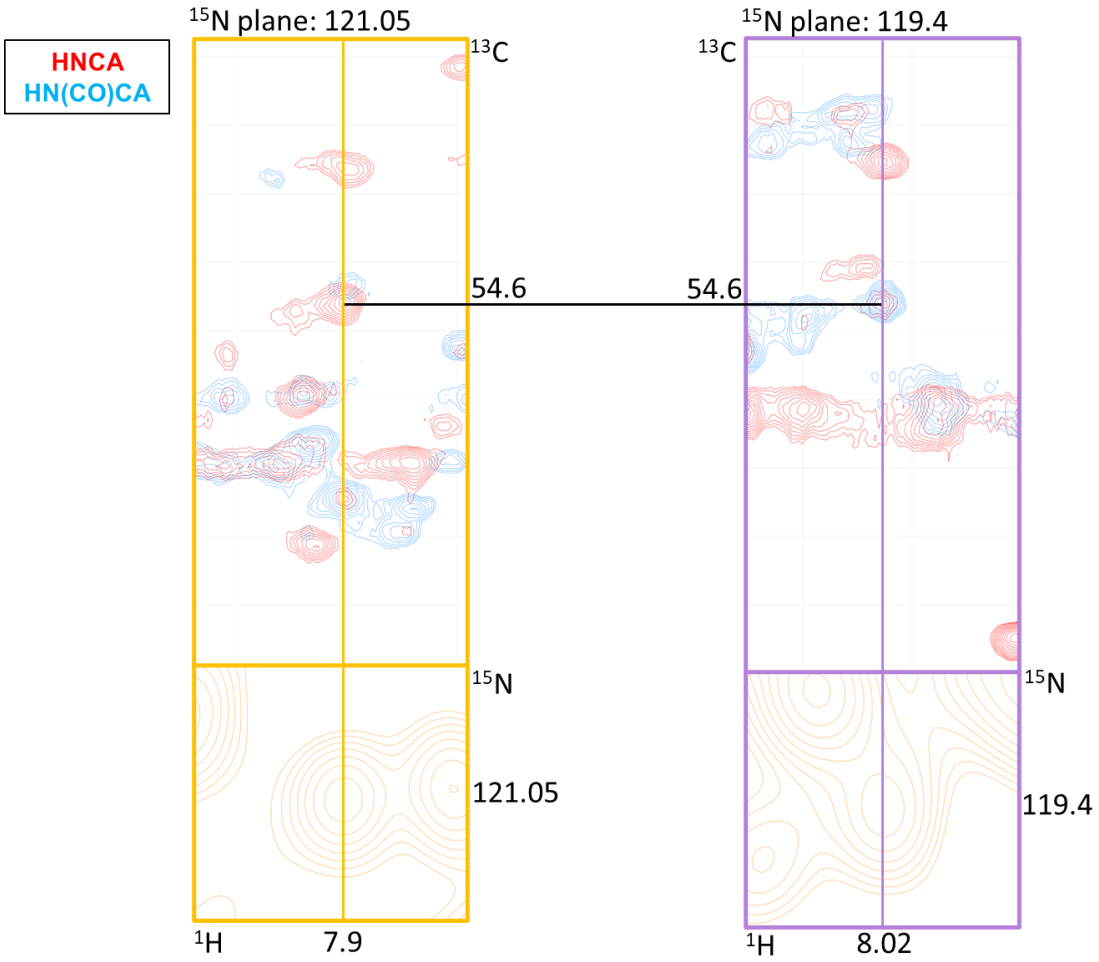


Figure 4.5. Visualisation of HSQC peak pairing using carbon data from HNCA (red) and HN(CO)CA (blue) experiments. Matched Carbon frequencies of 54.6 ppm in the orange and purple planes

In this example, the orange and purple spin systems have a carbon frequency of 54.6 ppm. Since the purple is an *i-1* frequency, this connection indicates that the purple residue is the next one after the orange. This connection is then validated by checking whether the other carbon atoms (Cβ, and carbonyl (Co)) also match. This process is then repeated to link chains of spin systems together. Some amino acids can be assigned based on their carbon chemical shift values. For example, alanine is the only amino acid to have a Cβ value up field from 20

ppm. Another example is glycine, which is the only amino acid to have no Cβ. To assign the chains of linked residues, the characteristic peaks are matched to the protein sequence. For example, in MmsFcc it was possible to assign the sequence X-A-X-X-G-X as H-A-K-Q-G-M, since this was the only part of the sequence that had a glycine three residues C-terminal to alanine.

For the backbone assignment a 400 µM sample of ^{13}C, ^{15}N-labelled His-MmsFcc D49S was prepared following Section 2.7. The successful backbone assignment of MmsFcc D49S is described in Section 4.2.

4.2 Backbone assignment of MmsFcc D49S

The experiments performed in the backbone assignment of D49S, and the nuclei they observe, are listed in Table 4.1.

HNCA	$C\alpha_i$, $C\alpha_{i-1}$	HN(CO)CA	$C\alpha_{i-1}$
HNCACB	$C\alpha_i$, $C\beta_i$, $C\alpha_{i-1}$, $C\beta_{i-1}$	HN(CO)CACB	$C\alpha_{i-1}$, $C\beta_{i-1}$
HNCO	Co_i, Co_{i-1}	HN(CA)CO	Co_{i-1}

Table 4.1. List of backbone assignment experiment pairs used, with the observed nuclei listed.

The first step was to determine whether there was sufficient signal in the carbon region to perform the assignment. When comparing the 121.5 ppm nitrogen plane of cleaved D49S to His-MmsFcc, both signal intensity and peak sharpness are greatly improved (Figure 4.6). A

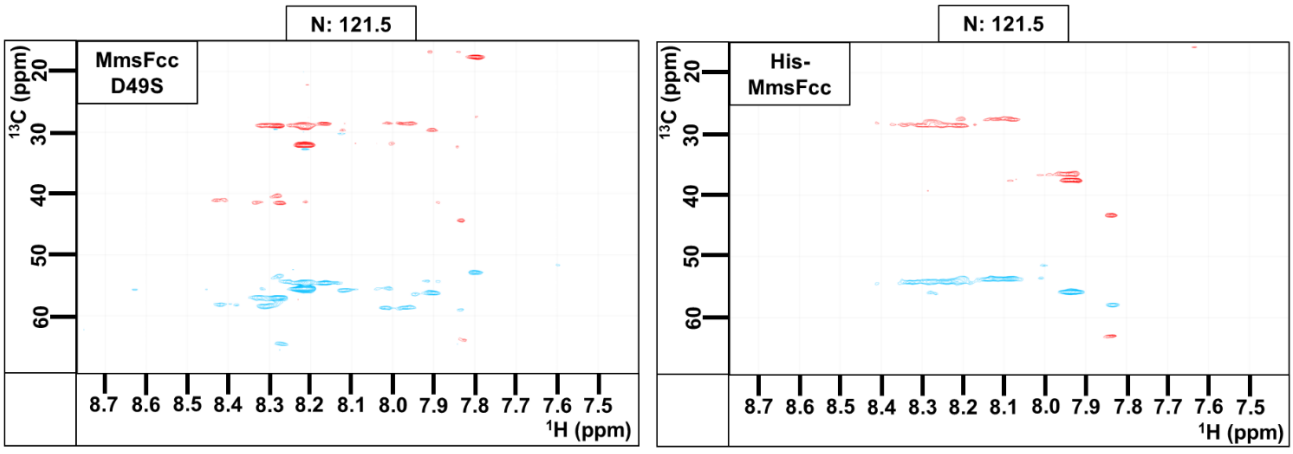


Figure 4.6. Comparison of constant-time HNCACB nitrogen planes from D49S and His-MmsFcc samples. Both samples were roughly 400 µM concentration, and both datasets were processed identically.

It was therefore possible to partially assign the MmsFcc D49S backbone shifts. The assigned HSQC spectrum is shown in Figure 4.7, with densely populated regions of the spectrum expanded in Figure 4.8.

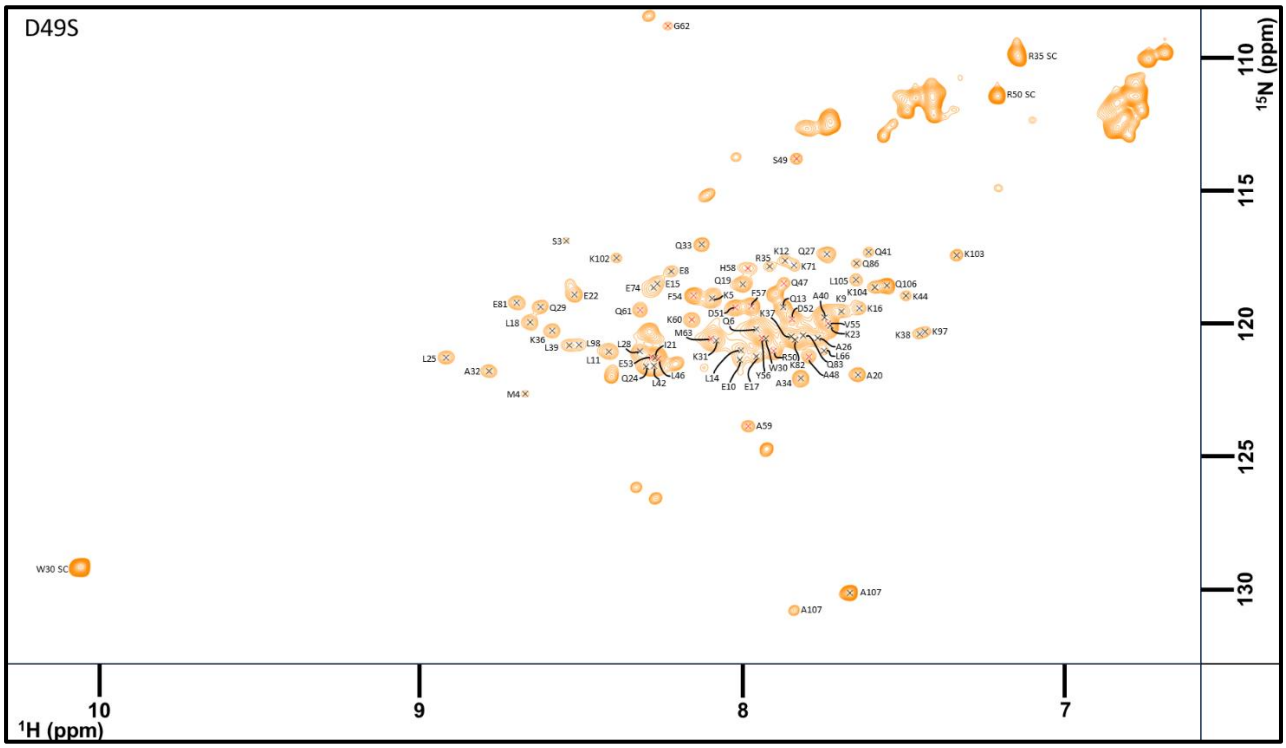
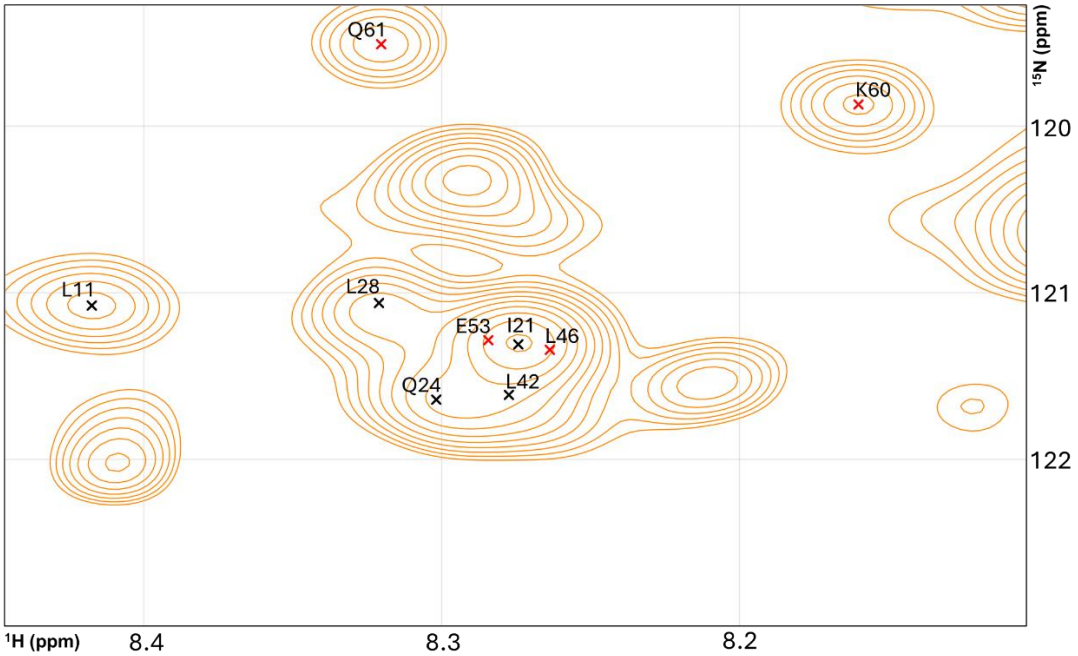


Figure 4.7. Partially assigned D49S HSQC spectrum. Sidechain peaks are denoted with SC. Residues surrounding and in the loop are marked with a red cross.



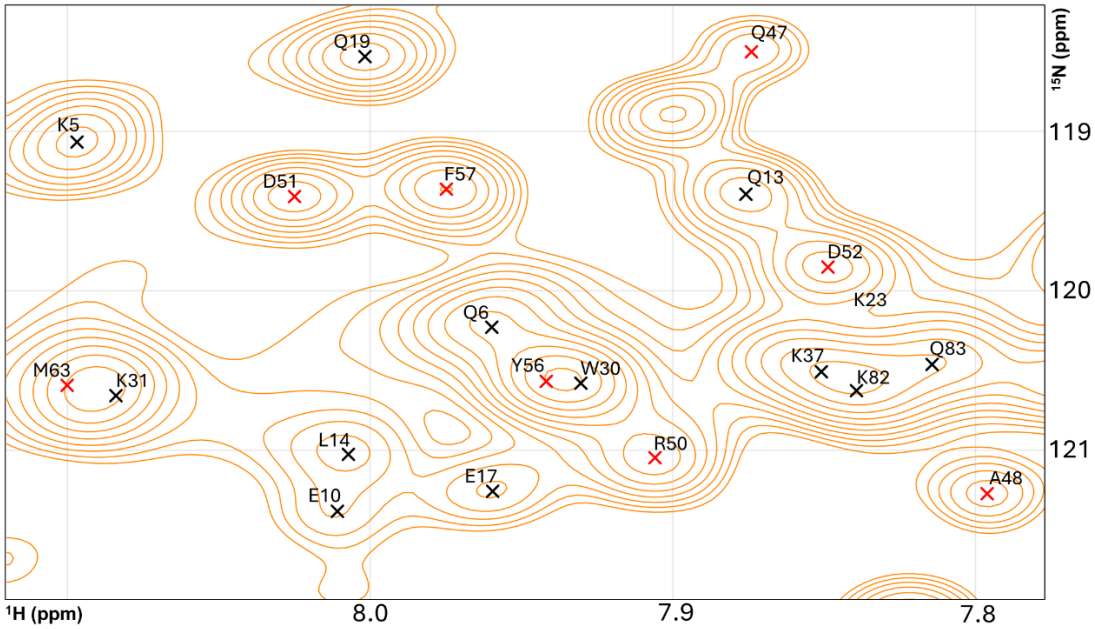


Figure 4.8. Zoom of densely populated regions of D49S HSQC spectrum. Residues surrounding and in the loop are marked with a red cross.

From these Figures, it can be seen that not all HSQC peaks are assigned. These peaks are unassignable to the MmsFcc sequence and are discussed in detail in Section 4.4. The assignment of MmsFcc D49S ^{1}H -^{15}N HSQC was ultimately successful; of the assignable residues, only L7, K45, K64, and Q65 were unable to be assigned. This was due to insufficient signal strength. A diagram representing the assignment of some of the metal binding loop residues is presented in Figure 4.9.

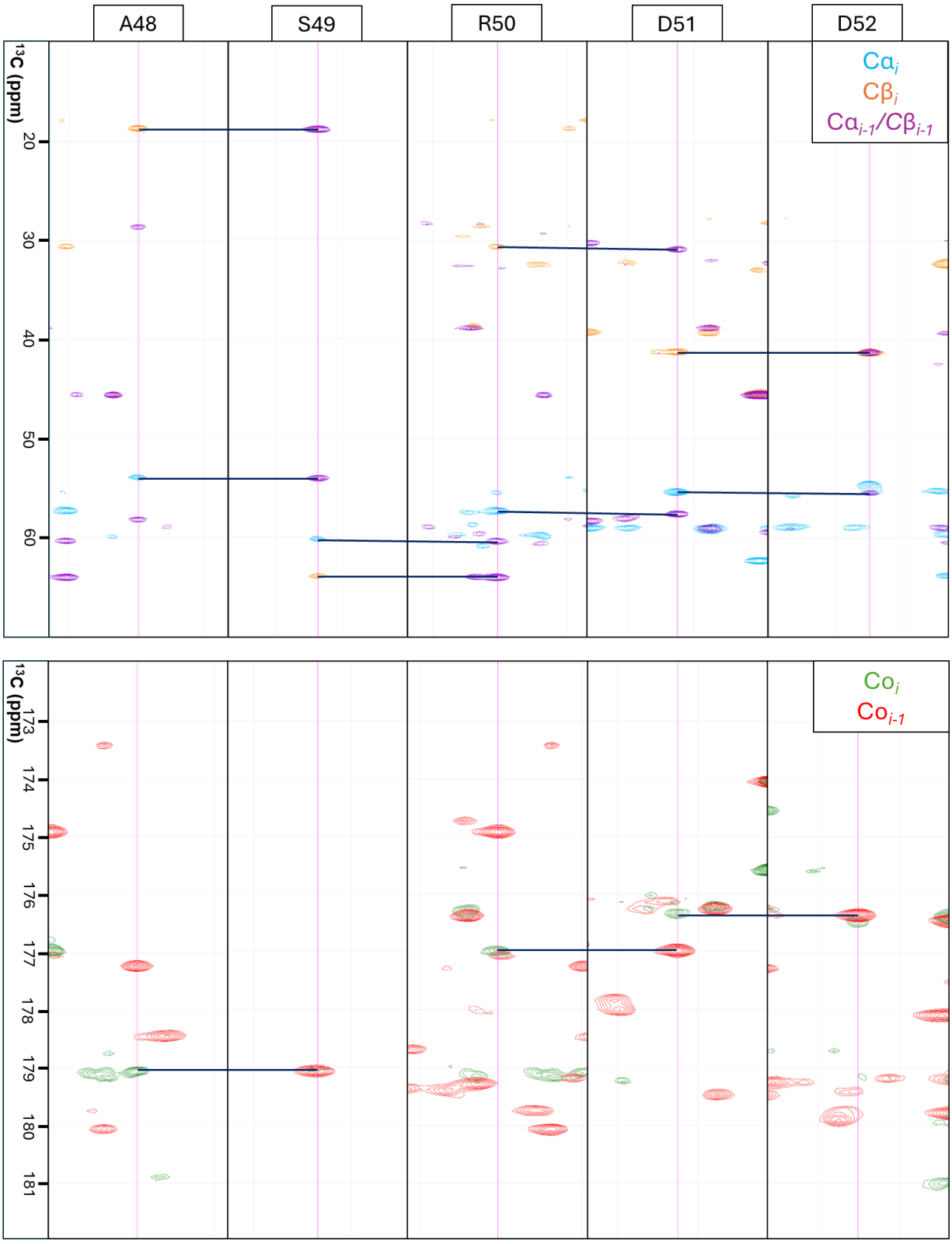


Figure 4.9. Assignment of A48-D52 in MmsFcc D49S, using pairs of triple resonance experiments. In the first panel, the CBCANH spectrum (blue & orange) is overlaid on the CBCA(CO)NH spectrum (purple). In the second panel, the HNCO spectrum (green) is overlaid on the HN(CA)CO spectrum (red).

The hypothesis that many equivalent coiled-coil residues would be represented by the same peak was shown to be true. An assessment of the difference in peak position for equivalent coiled-coil residues is performed in Figure 4.10. For this figure, the following calculation was used to determine the shift in peak position:

$$\Delta\delta=\sqrt{[(Coil1peak_H- Coil2peak_H)^2+ 0.14*(Coil1peak_N-Coil2peak_N)^2]}$$

Nitrogen values were normalised to proton by multiplying them by 0.14. This rough approach is a commonly used method of normalization, by weighting the H and N shift changes proportionally to their shift range (M. P. Williamson, 2013a).

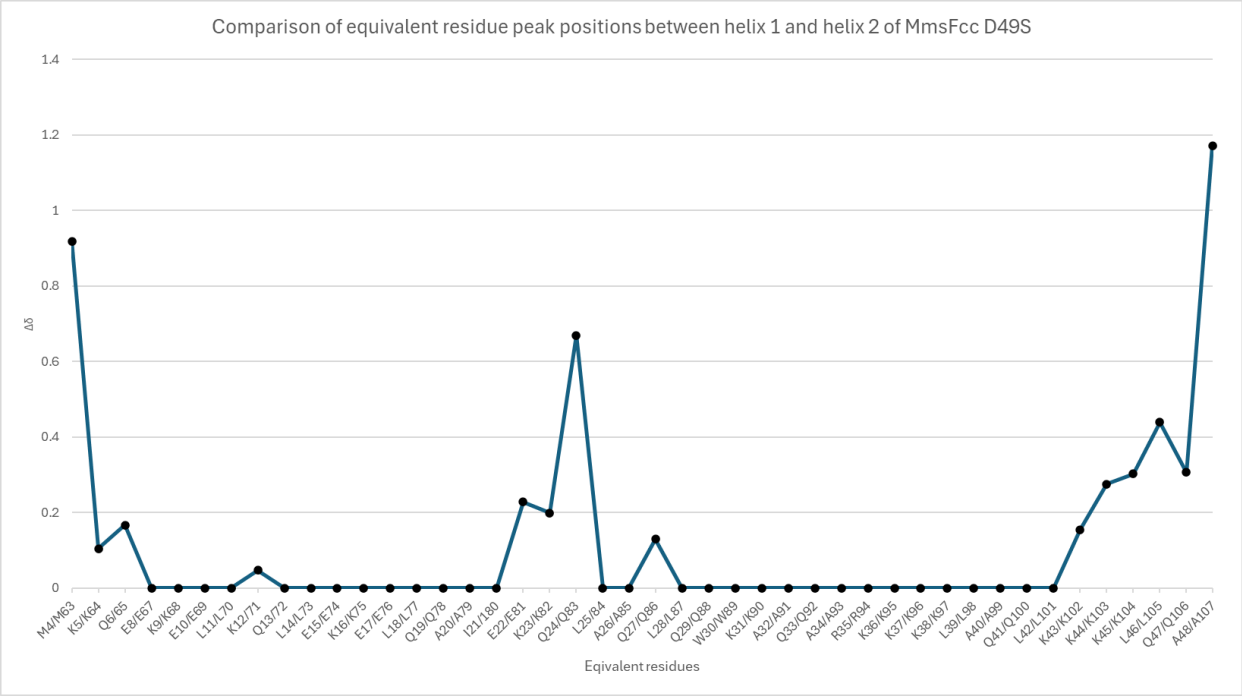


Figure 4.10. Comparison of equivalent residue peak positions between helix 1 and 2 of MmsFcc D49S.

There is a predictable divergence of peak positions towards the ends, caused by the presence of the unique loop sequence. The peak divergence observed in the centre of the coiled-coil is unexpected. It is not possible to conclude what is causing this, but it could likely be due to protein dynamics. It is possible that the protein is not behaving like a proper coiled-coil, resulting in less opportunity for conformational averaging. A table containing all chemical shifts determined from these experiments is found in Appendix x. Chemical shift values for residues in the second α-helix were taken from the equivalent 1st α-helix residue if a unique HSQC peak could not be found. For some peaks that were unassignable in the

HSQC, it was possible to determine carbon chemical shift values using i-1 frequencies in the carbon spectra (e.g. L7, K45). With the assignment complete, it was then time to focus on the more biologically relevant wildtype protein. The sucessful backbone assignment of MmsFcc is described in Section 4.3.

4.3 Backbone assignment of MmsFcc and comparison of MmsFcc and D49S spectra

MmsFcc was assigned using the same experiments and strategies described in Section 4.2. A 400 µM sample of ^{13}C, ^{15}N labelled MmsFcc was used for the backbone assignment. The assigned MmsFcc D49S spectrum was used as a guide for this assignment, which was mostly useful for coiled-coil peaks. The fully assigned HSQC spectrum is presented in Figure 4.11, with densely populated regions of the spectrum focused on in Figure 4.12.

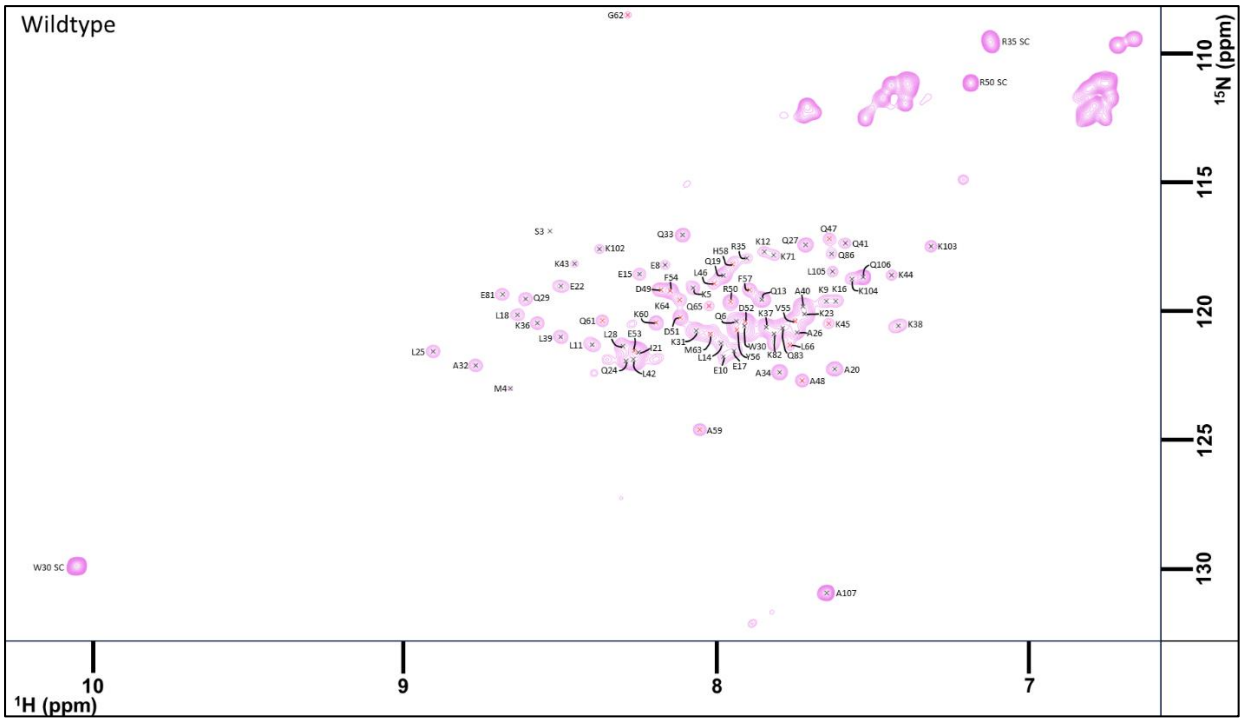


Figure 4.11. Fully assigned MmsFcc HSQC spectrum. Sidechain peaks are denoted with SC. Residues surrounding and in the loop are marked with a red cross.

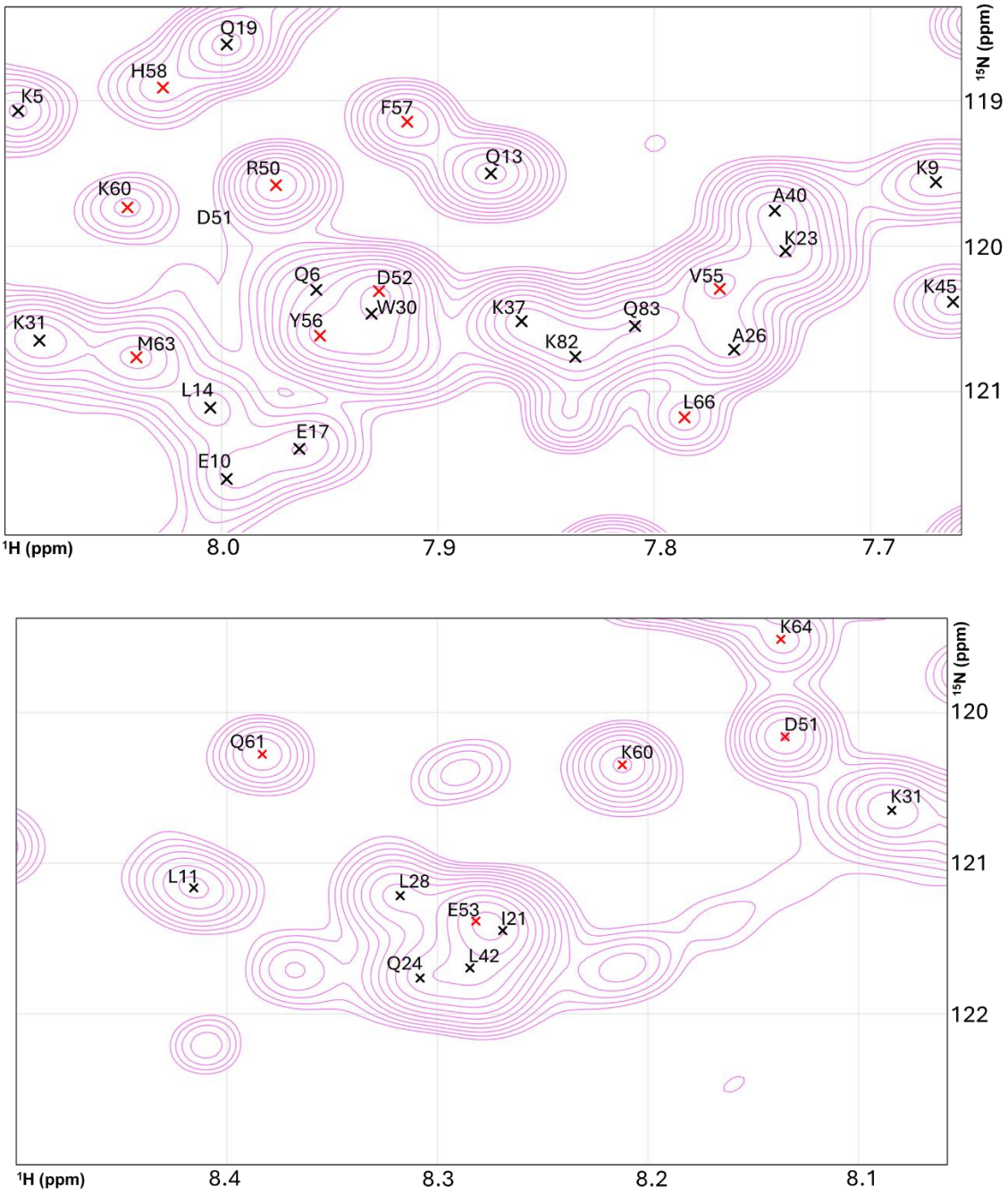


Figure 4.12. Expanded view of densely populated regions of MmsFcc HSQC spectrum. Residues surrounding and in the loop are marked with a red cross.

As with D49S, there are some unassignable peaks which are discussed in Section 4.4. The chemical shift values derived from the backbone assignment are listed in Appendix x. The assignment of MmsFcc was successful, with L7 being the only unassignable residue. This residue was unassignable due to a lack of signal in the carbon spectra. The assignment of some of the metal-binding loop residues is presented in Figure 4.13.

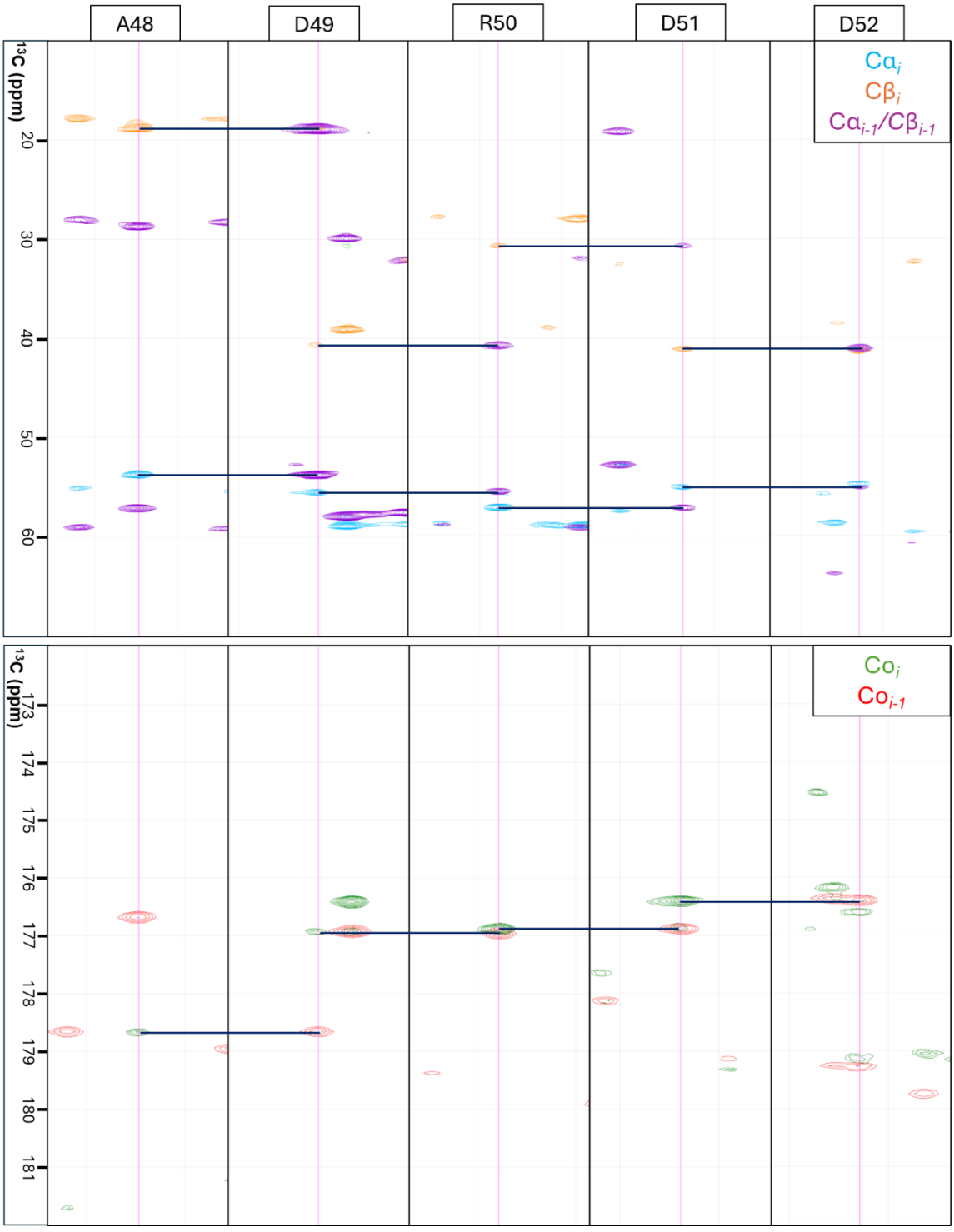


Figure 4.13. Assignment of A48-D52 in MmsFcc, using pairs of triple resonance experiments. C_α (blue), C_β (orange), and C_o (green) peaks are matched to pC_α & pC_β (purple), and pC_o (red) to connect HSQC residues.

An assessment of the difference in peak position for equivalent coiled-coil residues is performed in Figure 4.14. The following equation was used to calculate peak shifts:

$$\Delta\delta=\sqrt{[(Coil1peak_H- Coil2peak_H)^2+ 0.14*(Coil1peak_N-Coil2peak_N)^2]}$$

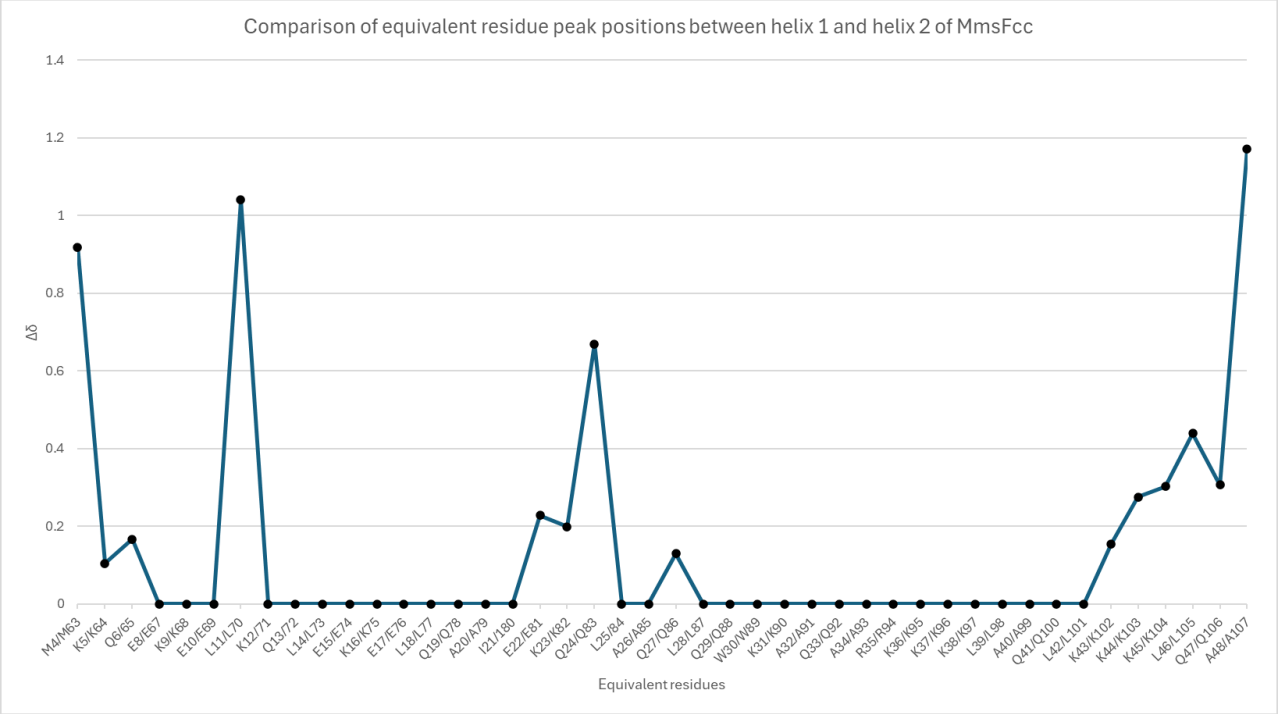


Figure 4.14. Comparison of equivalent residue peak positions between helix 1 and 2 of MmsFcc.

Figure 4.14 shows similar results to Figure 4.10, suggesting that coiled-coil backbone sequence is responsible for the distribution observed. A comparison of peak positions in wildtype and D49S HSQC spectra in performed in Figure 4.15. The following equation was used to calculate the difference:

$$\Delta\delta=\sqrt{[(WT_H-D49S_H)^2+0.14*(WT_N-D49S_N)^2]}$$

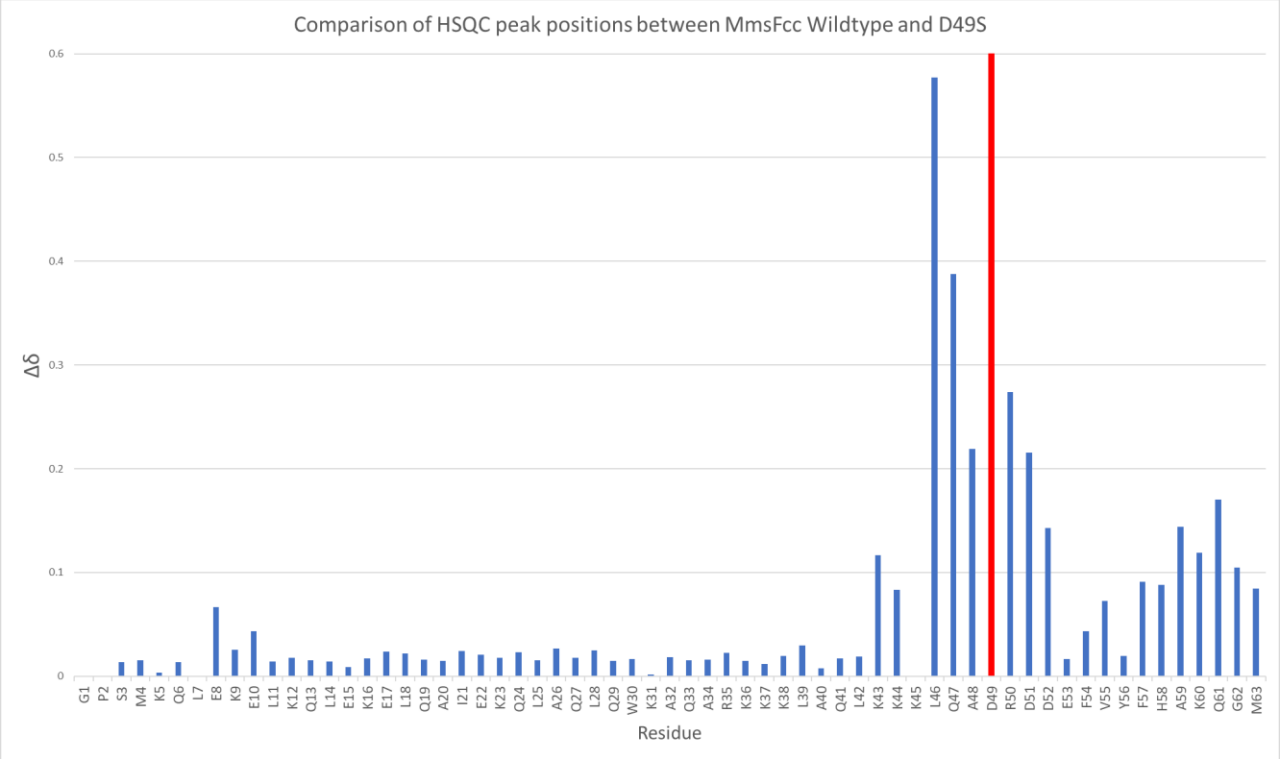


Figure 4.15. Difference in peak position for Wildtype and D49S ^{15}N-HSQC spectra. Residues 62-107 have been excluded for legibility. The position of the mutation is highlighted in red.

Predictably, most of the difference is seen around the site of mutation (residue 49). Typically, a point mutation which does not affect structure usually affects HSQC peaks from residues 3 C-terminal and 3 N-terminal to the site of mutation. It can be seen in Figure 4.15 that residues further than 3 C-terminal to residue 49 are significantly altered, suggesting some form of conformational change as a result of the mutation.

-

In both assignments, there was a group of peaks which did not fit the sequence of either wildtype or D49S. These peaks are discussed in Section 4.4.

4.4 The origin of unassignable peaks could not be determined

As mentioned previously, there was a class of peaks which did not fit the sequence of MmsFcc or MmsFcc D49S. These peaks were weak in the nitrogen HSQC, intense in the carbon planes of triple resonance experiments, present in both MmsFcc and D49S spectra, and were able to be linked by their carbon chemical shifts. Ultimately, it was not possible to determine the origin of these peaks. The HSQC peaks can be seen in Figures 4.16 and 4.17. They are labelled @number, which is how CCPN version three labels unassigned residues.

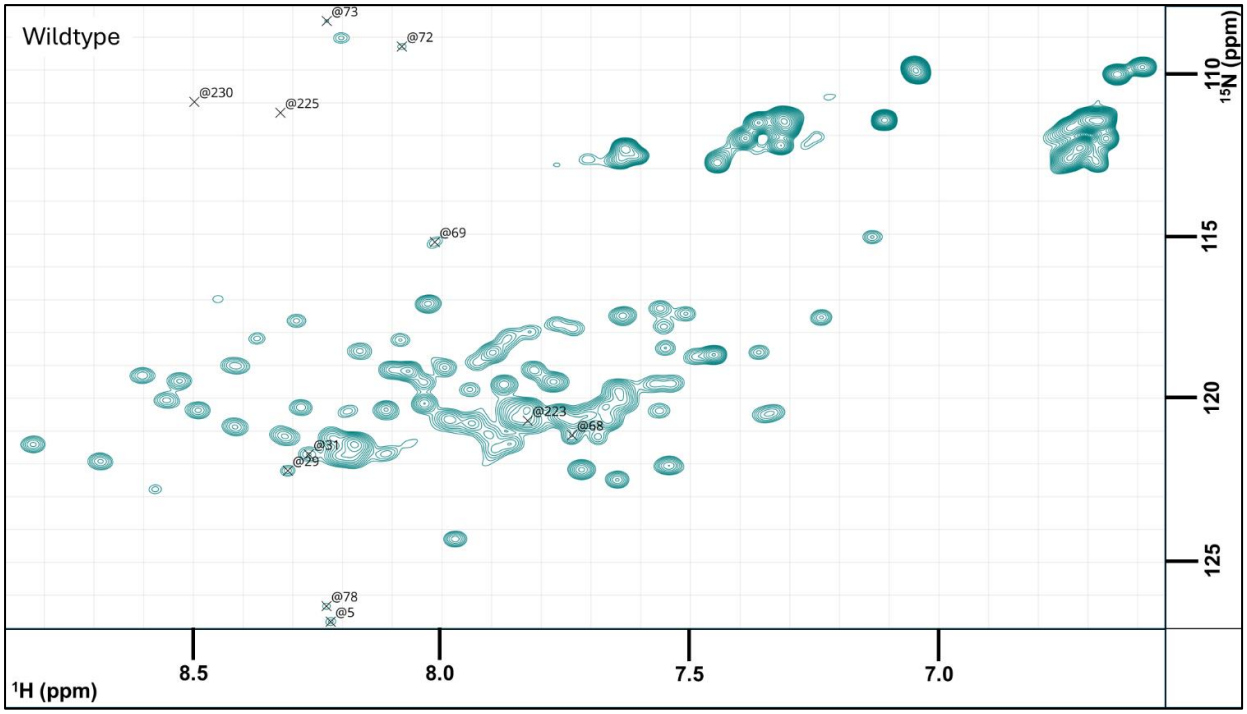


Figure 4.16. Wildtype HSQC with additional unassigned peaks labelled.

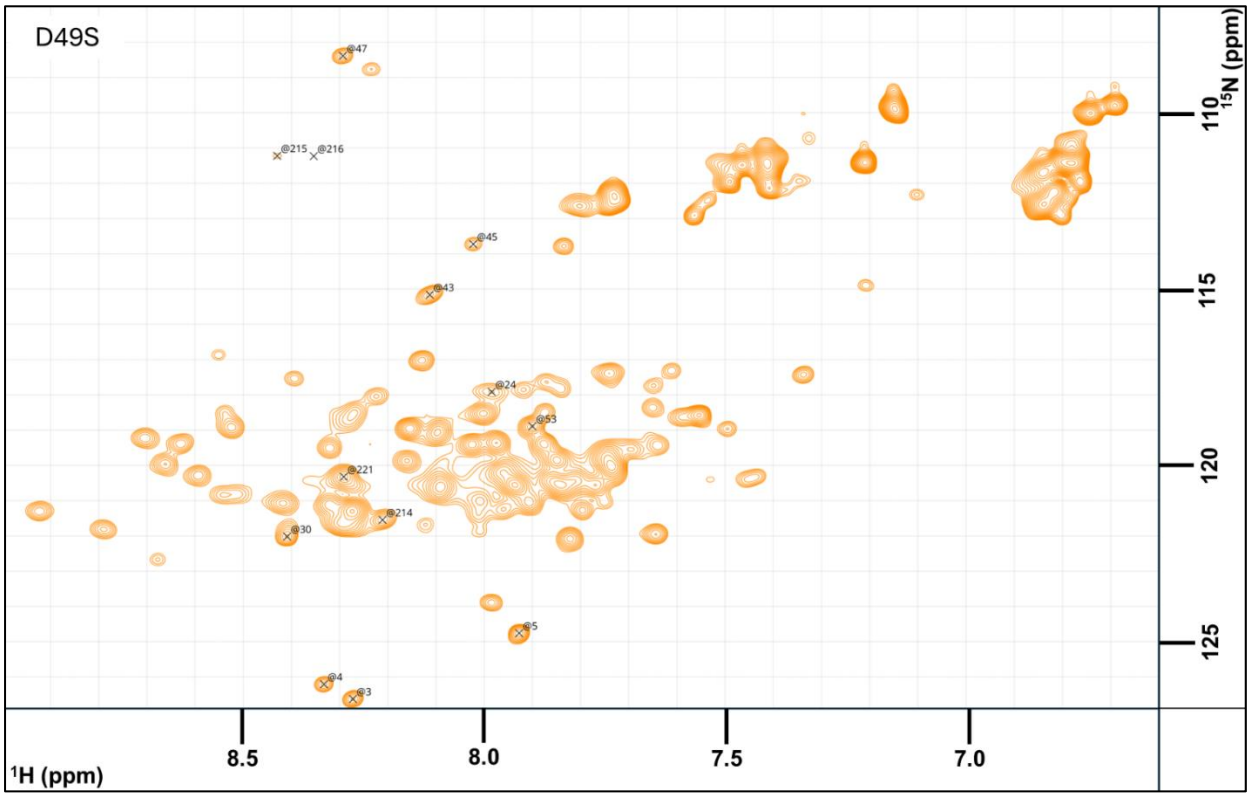


Figure 4.17. D49S HSQC with additional unassigned peaks labelled.

From these two HSQCs it can be seen that some peaks exist in the same location in both spectra (such as @78 in wildtype and @4 in D49S), while others are unique to each protein. These residues are potentially in a random coil conformation since the proton shifts are

mostly in-between 8-8.5 ppm. A list of the determined *i* and *i-1* residues for these spin systems is provided in Table 4.2.

Wildtype			**D49S**		
Spin system	***i-1 residue***	***i residue***	**Spin system**	***i-1 residue***	***i residue***
73	A	G	47	A	G
72	G	G	215	T	G
236	S	G	216	S	G
225	T	G	45	G	T
235	G	T	24	?	?
69	?	T	53	G	?
50	A	?	221	A	?
46	?	?	214	?	?
29	?	?	30	?	?
78	T	A	5	S	?
5	T	A	4	A	T
			3	A	T

Table 4.2. Determination of *i* and *i-1* residues for wildtype and D49S unusual peaks.

One unusual feature of these spin systems was the relatively high intensity of the carbon peaks, when compared to their root HSQC peak. This is shown in Figure 4.x.

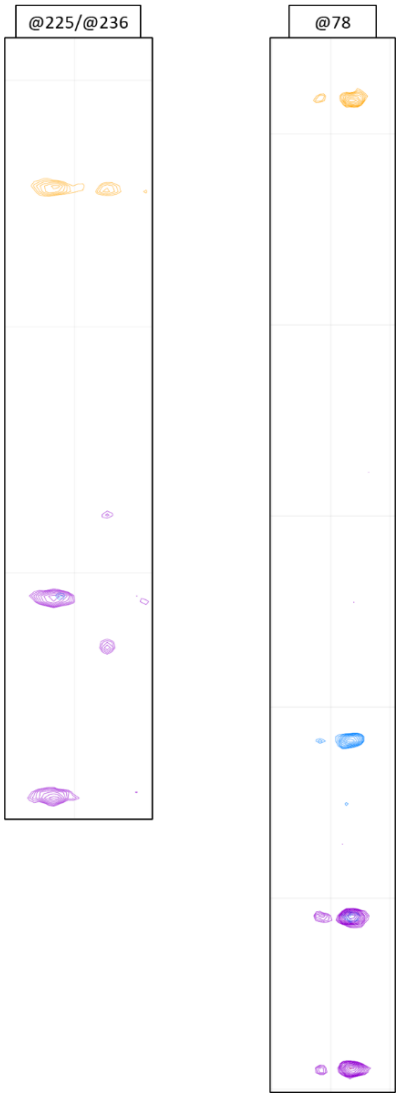


Figure 4.18. @225/@236 and @78 carbon strips from MmsFcc. HNCACB (blue and orange) and HN(CO)CACB (purple) spectra included.

This difference in intensity is likely due to slower carbon relaxation rates when compared to the main MmsFcc peaks. If these peaks are from a random coil peptide, then amide peaks will be more affected by the loss of signal due to increased mobility.

It was possible to link some of these spin systems sequentially, suggesting that they are from the same peptide. The result of this for wildtype and D49S is shown in Tables 4.3 and 4.4.

Wildtype sequential spin systems					
PEAK	@46	@29	@69	@5	@73
RESIDUE	?	?	T	A	G

Table 4.3. Partial linking of unassigned wildtype peaks.

D49S sequential spin systems

PEAK	@215	@45	@4	@221	@214	@30	@43	@3	@47	@53	@5
RESIDUE	G	T	A	?	?	?	T	A	G	?	S

Table 4.4. Partial linking of unassigned D49S peaks.

There is no sequence in the protein that matches these predicted sequences. To try and understand the origin of the addition peaks the spectrum of His-MmsFcc and MmsFcc were overlaid in Figure 4.19, with unusual and HisTag peaks labelled.

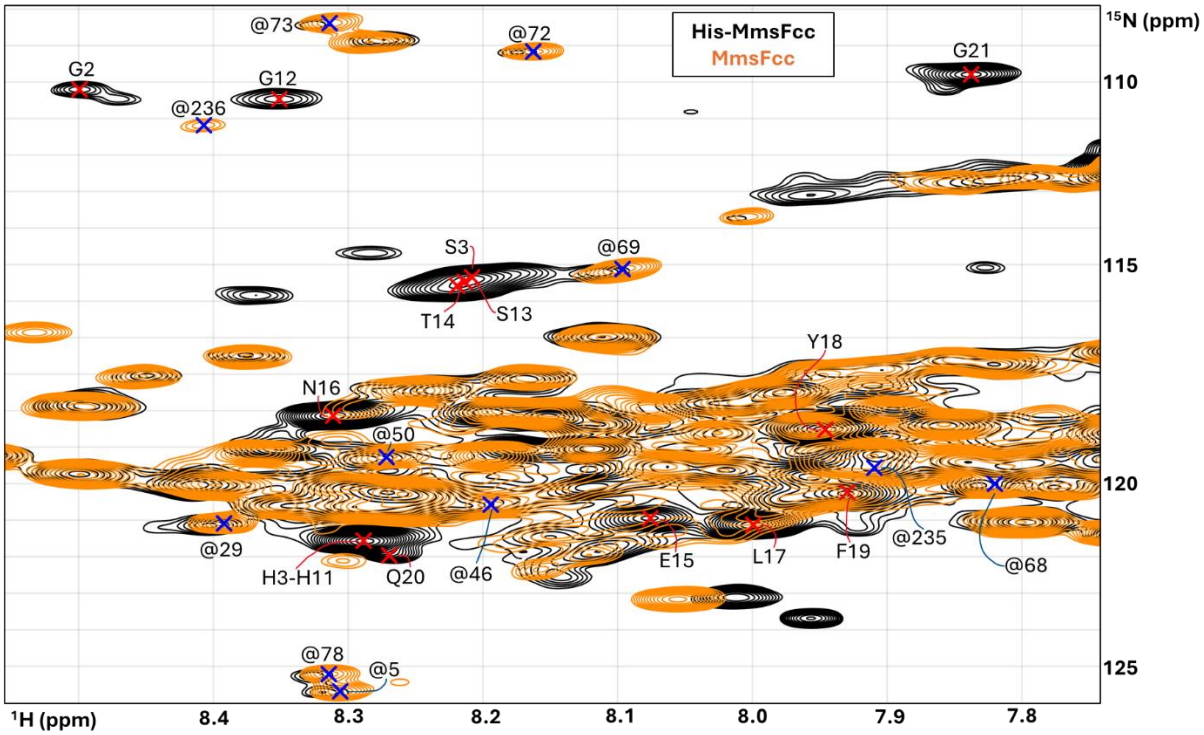


Figure 4.19. Comparison of His-MmsFcc (black) and MmsFcc (orange) HSQC spectra. Tag peaks from the His-MmsFcc spectrum are labelled in red, while additional unassigned peaks are labelled in blue.

Every additional peak, except for @236, is also present in the His-MmsFcc spectrum. This suggests that the presence of these peaks is independent of the HisTag cleavage. While it was not possible to determine the origin of these peaks, they did not populate the areas of the spectrum which were of interest to us. Critically, they were also unaffected by addition of metal cations (Chapter 5). For these reasons, the unusual peaks were ignored. More could have been learned about them by investigating the carbon HSQC spectra, and NOEs from these peaks. This was not done in the interest of time.

4.5 Partial sidechain assignment of MmsFcc

The sidechain carbons of MmsFcc were assigned using a series of NOESY and TOCSY experiments (Table 4.5). Due to the difference in the range of carbon chemical shift values for aliphatic and aromatic sidechains, aliphatic and aromatic C-HSQC spectra are collected separately and assigned using different strategies. A list of the experiments performed, and the samples used is provided in Table 4.5.

Experiment	Sample used	Note
Aliphatic ^{13}C HSQC	450 µM $^{13}C^{15}$N-MmsFcc	8-78 ppm carbon range
Aliphatic ^{13}C HSQC-NOESY	450 µM $^{13}C^{15}$N-MmsFcc	8-78 ppm carbon range. Detects NOEs from aliphatic protons.
HCCH-TOCSY	450 µM $^{13}C^{15}$N-MmsFcc	Detects protons in aliphatic residues
HCC-TOCSY	450 µM $^{13}C^{15}$N-MmsFcc	Detects carbons in aliphatic residues
Aromatic ^{13}C HSQC	450 µM $^{13}C^{15}$N-MmsFcc	103-143 ppm carbon range
Aromatic ^{13}C HSQC-NOESY	450 µM $^{13}C^{15}$N-MmsFcc	103-143 ppm carbon range. Detects NOEs from aromatic protons.
^{15}N HSQC	200 µM ^{15}N-MmsFcc	-
^{15}N HSQC-TOCSY	200 µM ^{15}N-MmsFcc	Detects α, β protons in residue. Sometimes detects extra protons in residue
^{15}N HSQC-NOESY	200 µM ^{15}N-MmsFcc	Detects NOEs from amide protons.

Table 4.5. Experiments used to perform sidechain assignments of MmsFcc and MmsFcc D49S.

All the experiments in Table 4.5 were also performed on a 650 µM sample of $^{13}C^{15}$N MmsFcc D49S, or on a 250 µM sample of ^{15}N MmsFcc D49S.

To begin assigning both aliphatic and aromatic sidechains, the hydrogen chemical shift values for Hα and Hβ atoms are needed. These can be derived from a ^{1}H-^{15}N HSQC-TOCSY. TOCSY data from a 200 µM sample of ^{15}N labelled MmsFcc are presented in Figure 4.20.

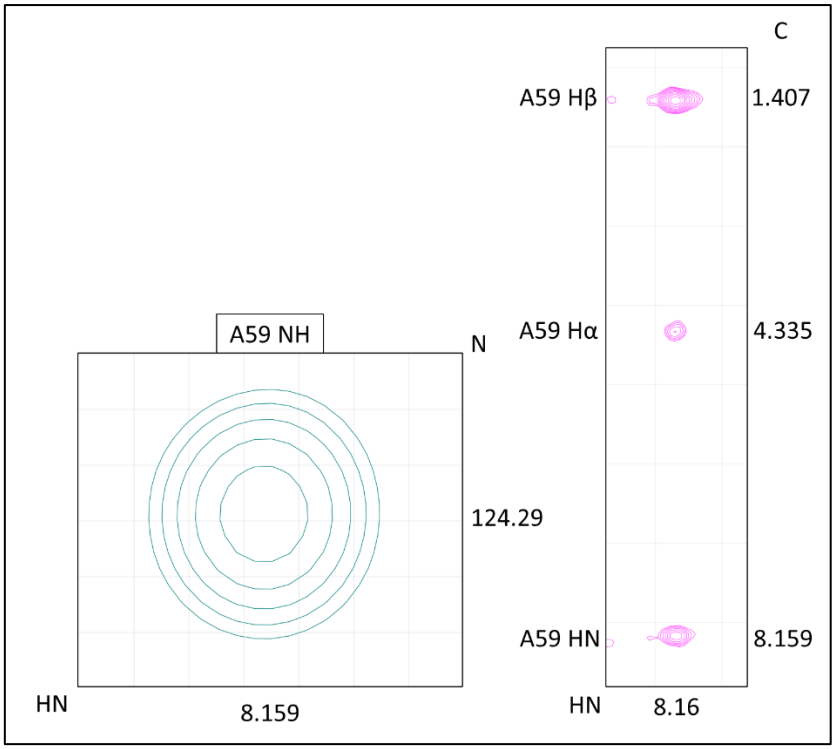


Figure 4.20. HSQC (teal)-TOCSY (blue) spectrum from 200 µM MmsFcc. A59 HN HSQC peak is correlated to A59 Hα and Hβ TOCSY peaks.

Carbon chemical shifts from the backbone assignment and proton shifts from the HSQC-TOCSY were used to assign α and β peaks in the MmsFcc ^{1}H-^{13}C HSQC spectrum, as shown in Figure 4.21. In this spectrum, orange peaks represent carbons bound to zero or two CH_x. If it is bound to one or three CH_x, then the peak is blue. This is the result of a constant time (CT) experiment being used.

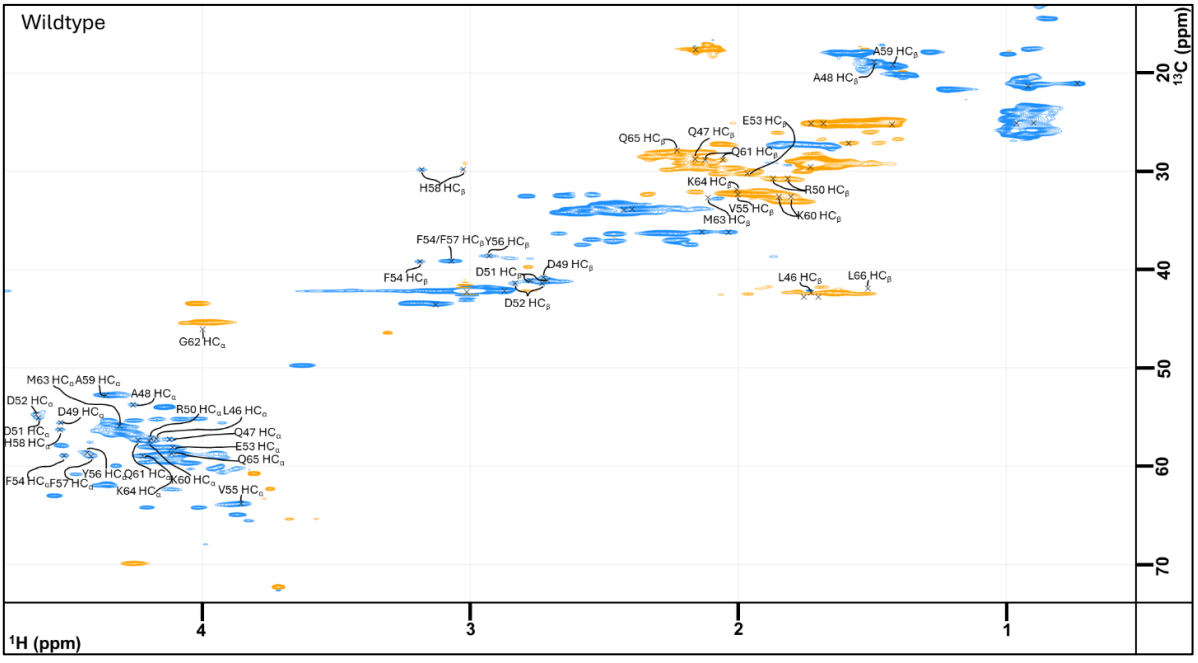


Figure 4.21. ^{1}H-^{13}C HSQC spectrum of MmsFcc, partially assigned with α and β from the metal-binding loop.

To assign carbon peaks further down the sidechain a new combination of carbon observing experiments were used. HCCH-TOCSY gives chemical shifts for all CH protons in a residue, while HCC-TOCSY gives chemical shifts for all CH carbons in a residue. Both spectra are displayed in Figure 4.22.

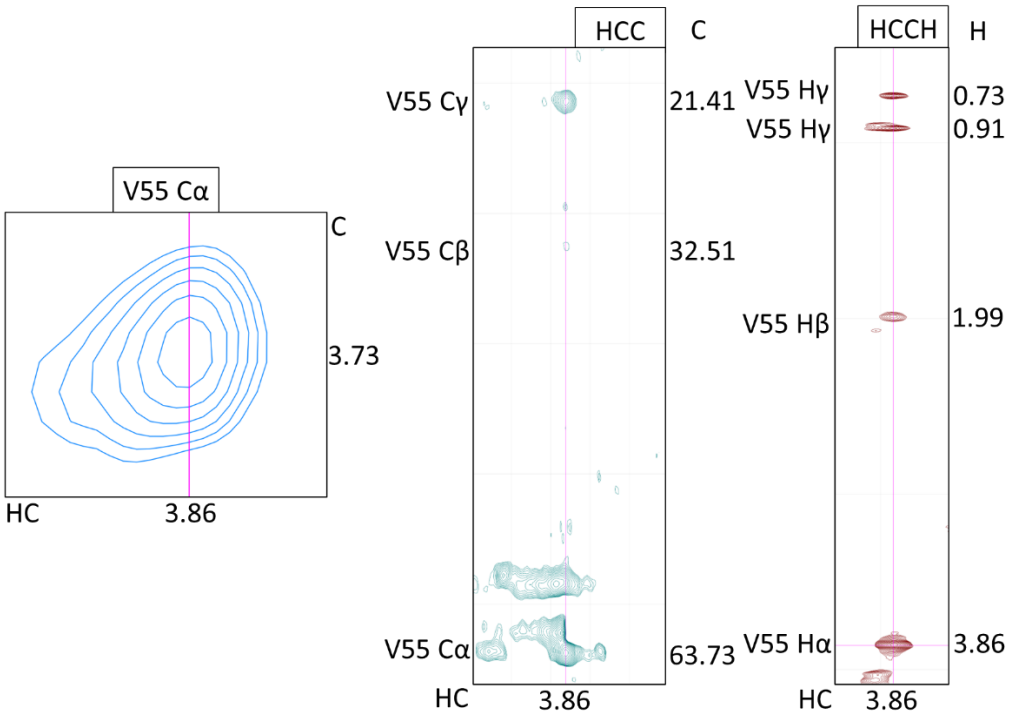


Figure 4.22. HCC and HCCH strips of V55, showing CH chemical shift values. Data from HCCH-TOCSY and HCC-TOCSY, observed on the V55 Cα carbon plane.

In Figure 4.22, sidechain chemical shifts are being observed from the V55 Cα plane. The chemical shift values are used to assign peaks in the carbon HSQC. These assignments were validated by observing other planes. In Figure 4.23, the V55 Cγ carbon plane was observed in the HCCH-TOCSY and HCC-TOCSY spectra.

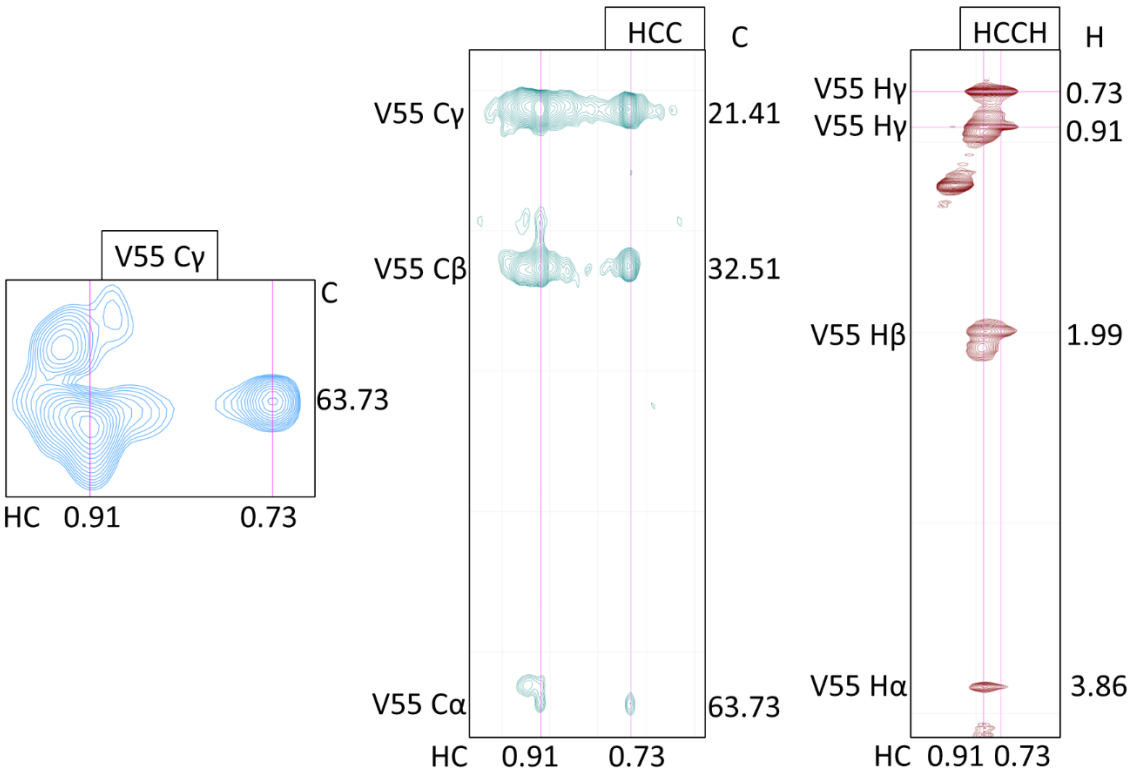


Figure 4.23. HCC and HCCH strips of V55, showing CH chemical shift values. Data from HCCH-TOCSY and HCC-TOCSY, observing the V55 Cγ carbon plane.

In the HCCH and HCC spectra, the intensities of peaks representing the same nuclei in different carbon planes provide further validation for the sidechain assignment. In general, peaks closer to the root peak will be more intense. This can be seen in Figure 4.23, where the Cγ-Cβ-Cα descend in intensity. Groups which produce strong signals, such as methyl groups, can skew this pattern. This is seen in Figure 4.22, where the V55 Cγ methyl peak is much more intense than the closer Cβ peak. This process was repeated to assign the rest of the aliphatic sidechain carbons in the MmsFcc metal-binding loop. This assignment is shown

in Figure 4.24.

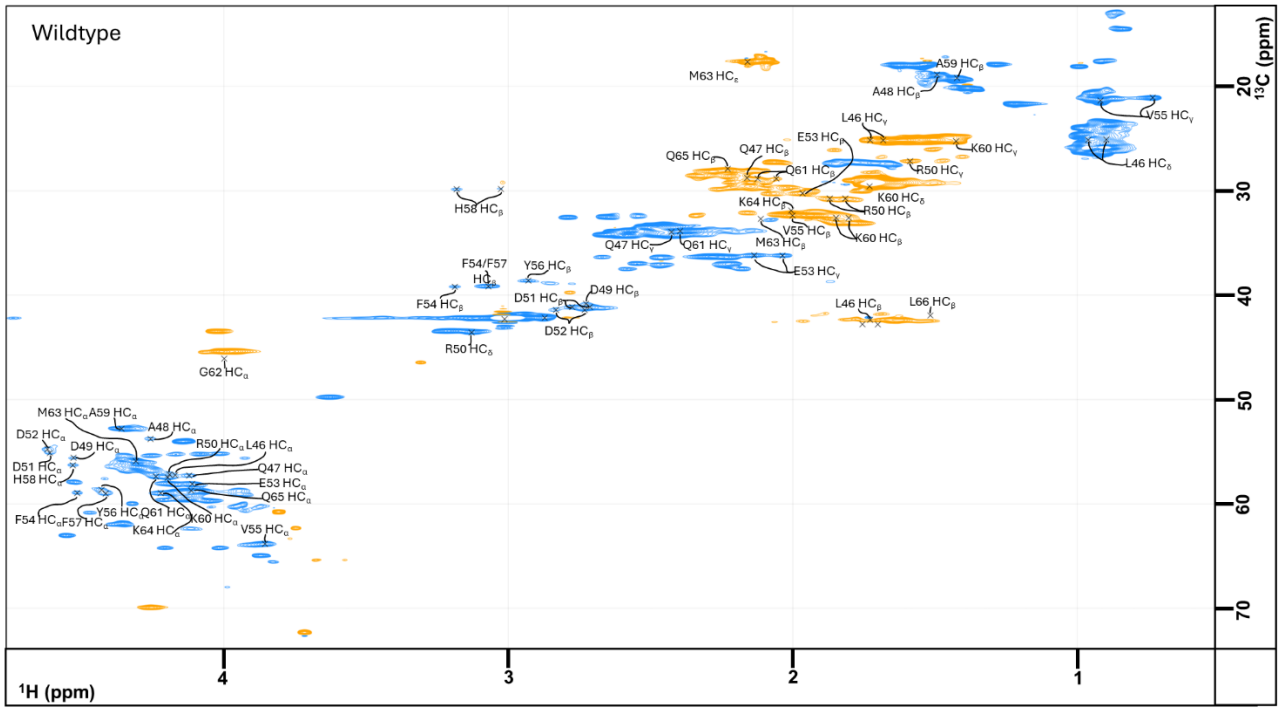


Figure 4.24. Aliphatic C-HSQC with full sidechain assignments for the metal binding loop.

Next, the aromatic C-HSQC spectrum was assigned. The unassigned spectrum of MmsFcc is shown in Figure 4.25. As with the aliphatic HSQC, the colour of the peak is related to the number of covalently bonded carbons. Red peaks are bound to one or three carbons, while blue peaks are bound to zero or two.

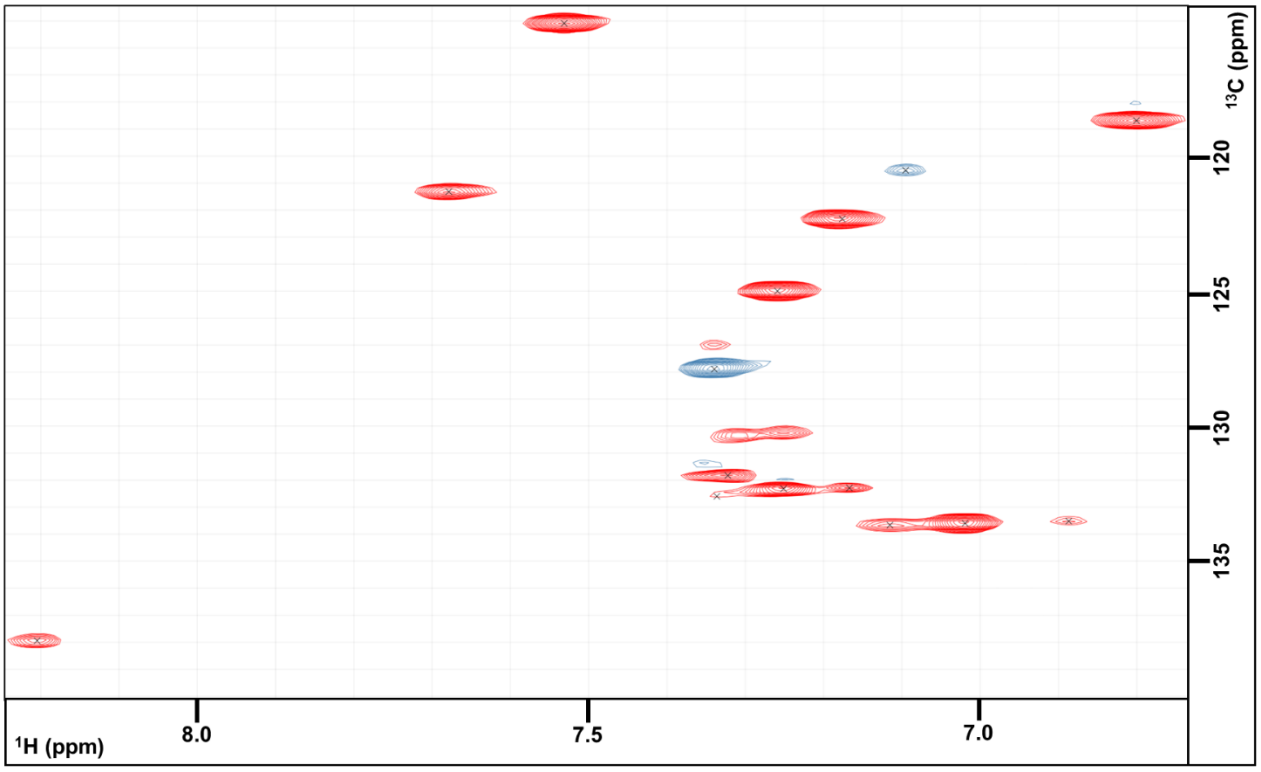


Figure 4.25. MmsFcc Aromatic ^{13}C-HSQC.

Due to the relatively small number of aromatic sidechain nuclei, it was possible to assign this spectrum using reference chemical shift values and NOEs. Tryptophan was assigned first, since it is possible to determine intramolecular NOEs to ring carbons from the NHε1 peak in the nitrogen HSQC spectrum. The IUPAC-recommended labelling of Tryptophan sidechain residues is shown in Figure 4.26.

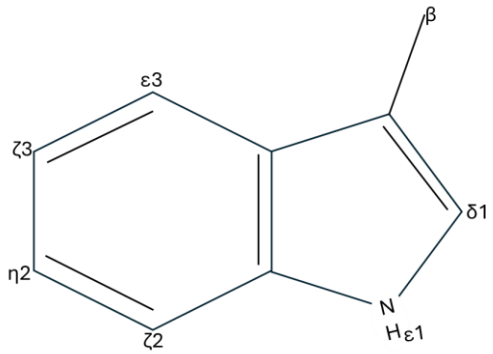


Figure 4.26. Labelled sidechain of Tryptophan. For simplicity, only nuclei bound to protons have been labelled.

An NOE between W30's Hε1 and Hδ1 was identified using the ^{15}N HSQC-NOESY and ^{13}C aromatic HSQC-NOESY spectra in Figure 4.27.

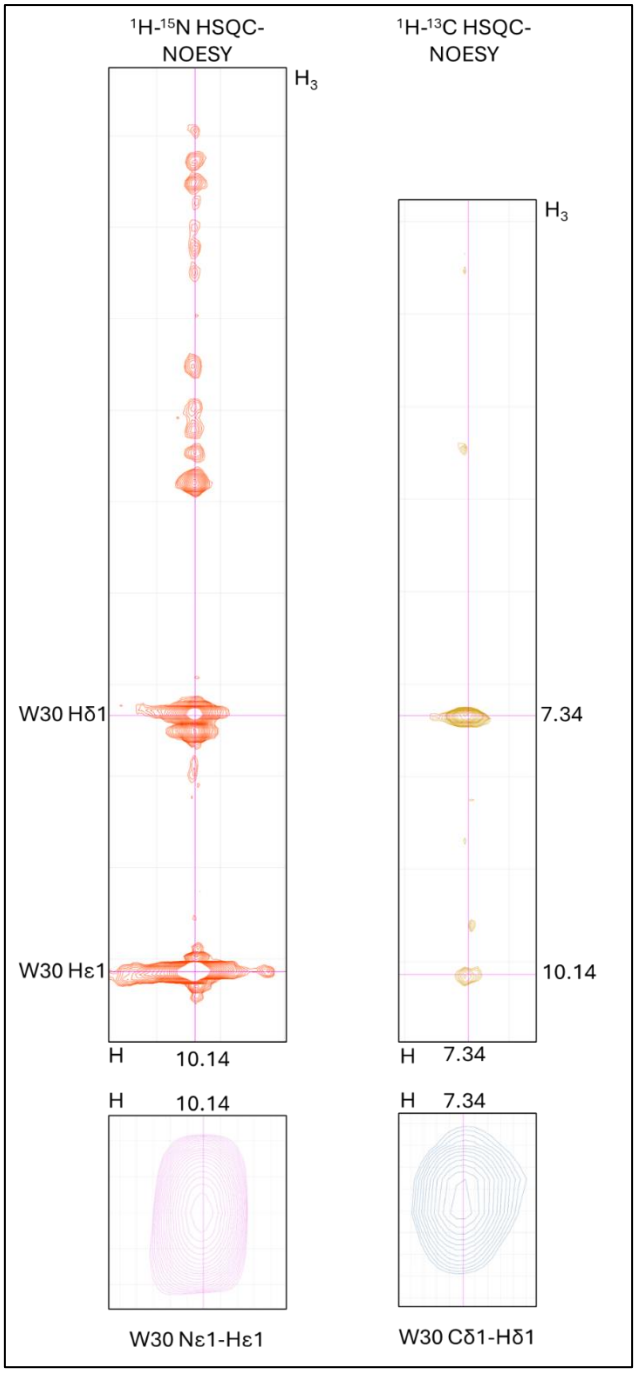


Figure 4.27. Strong intraresidue NOE between W30 Hε1 and Hδ1, identified from ^{1}H-^{15}N HSQC-NOESY and validated with ^{1}H-^{13}C HSQC-NOESY.

An NOE from the NHε1 to CHζ2 was then detected using the same method. From this, it was possible to detect NOEs around the second ring of tryptophan using the aromatic NOESY spectrum and fully assign all possible tryptophan sidechain signals. Next, the tyrosine 56 residue was assigned. This was done by detecting the intramolecular NOE between Y56 Hβ-Hδ using the aliphatic-NOESY spectrum. This was then validated using the aromatic NOESY.

The aromatic NOESY was then used to identify the NOE between Y56Hδ-Hε and to assign the relevant peak in the aromatic HSQC.

After this, only phenylalanine and histidine sidechains were left to be assigned. The lone peak at 8.21, 138 ppm was assigned as H58 CHε1 based on the typical chemical shift values for these nuclei (as per the BMRB database). Phenylalanine CHδ peaks were assigned by observing NOEs from F54 and F57 Hβ to their respective Hδ. The NOE observed between F54/F57 CHδ and F54/F57 CHε peaks were used to assign the CHε peaks. It was not possible to determine an NOE between phenylalanine Hε-Hζ, so reference chemical shift values were used instead. The full assignment is shown in Figure 4.28.

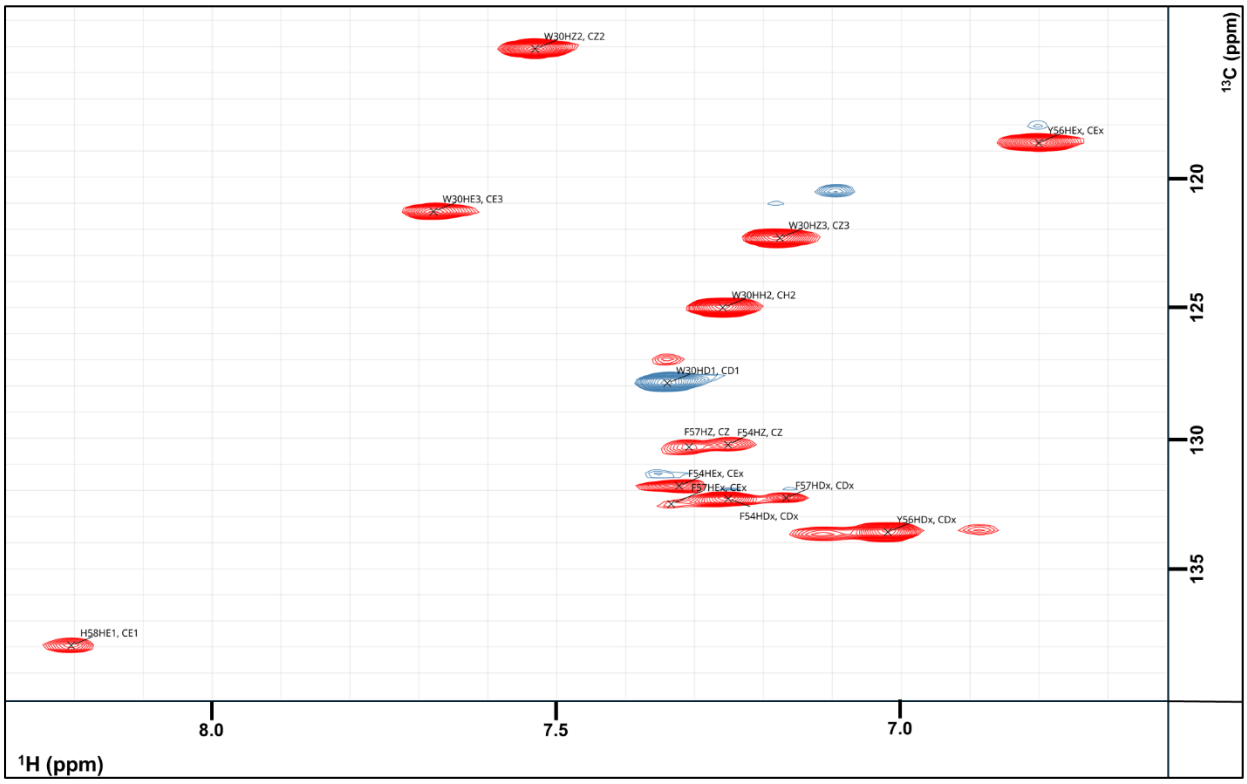


Figure 4.28. Aromatic HSQC of MmsFcc with all aromatic sidechains assigned.

-

Due to time constraints, the structure of MmsFcc D49S was not investigated. As a result, the carbon HSQC were not fully assigned and have therefore been excluded from this section.

With these assignments complete, the next step was to determine the structure of MmsFcc. This process is described in Section 4.6.

4.6 Determining the structure of MmsFcc

The structure of MmsFcc was determined by restrained molecular dynamics with a restrained molecular dynamics simulation, using the programme Crystallography and NMR system (CNS) (Brünger et al., 1998). This programme simulates the gradual cooling of the protein, where it attempts to fold the protein into a structure with the lowest total energy. Restraints are given to the programme which define energies for the simulation to satisfy. For MmsFcc, the following types of restraints were applied:

- Hydrogen bond restraints
- Backbone dihedral angle restraints
- NOE distance restraints

Hydrogen bonds and backbone dihedral angle restraints were both derived from experimental data by performing a TALOS-N prediction (Shen & Bax, 2013, 2015). TALOS-N is a neural network-based programme which has been trained to predict backbone angles from backbone chemical shift data. TALOS-N produces a list of ϕ, ψ, and χ angles for each residue, which can be used to determine whether that residue is in an α-helix or β-sheet conformation. The TALOS output for MmsFcc is presented in Appendix 6.12. A value has been excluded when the software was unable to calculate a value. All input files used to determine the structure of MmsFcc will be uploaded to the Protein database and are available upon request.

-

The strategy for solving the structure of MmsFcc started with ensuring the correct folding of the coiled-coil backbone (i.e. if the AlphaFold predicted coiled coil structure is accurate), then determining the structure of the loop by using experimental NOE's. Since the loop of MmsFcc was the primary focus of this study, it was plausible to consider whether the scaffold of the protein required structural determination. It would have been possible to encode CNS to reproduce the coiled-coil conformation predicted by AlphaFold. Another solution would have been to exclude these residues entirely, and to instead encode the MmsFcc loop into a closed circle. Ultimately it was decided to include the coiled-coil component of MmsFcc to ensure that a rigid structure for this region did not impede the correct folding of the loop. To fix the coiled-coil backbone, the AlphaFold2 prediction of MmsFcc structure was used to provide additional restraints to supplement experimental

data. This was necessary since it was not possible to determine many experimental NOEs from the coiled-coils.

Firstly, the dihedral restraints were formatted into the following CNS-compatible format for angular restraints:

assign (resid and name) (resid and name)

(resid and name) (resid and name) A B C D

The following values were then inputted:

A. Energy constant ($kcalmol^{-1}rad^{-2}$)
B. Angle of restraint (°)
C. Range around restrained angle (°)
D. Exponent

The energy constant and the exponent remained unchanged for all dihedral restraints. Examples of ϕ, ψ, and χ restraints inputs are shown below:

!phi angle 3

assign (resid 2 and name C) (resid 3 and name N)

(resid 3 and name CA) (resid 3 and name C) 1.0 -66.703 17.358 2

!psi angle 3

assign (resid 3 and name N) (resid 3 and name CA)

(resid 3 and name C) (resid 4 and name N) 1.0 -151.442 17.358 2

!chi1 angle 3

assign (resid 3 and name N) (resid 3 and name CA)

(resid 3 and name CB) (resid 3 and name OG) 1.0 60.0 30.0 2

For dihedral angles obtained from TALOS, the range was set as two times the standard deviation from TALOS, as typical 95% confidence limits.

-

The correct formation of the two α-helices was next. Since α-helices have a predictable hydrogen bonding pattern, a sequence can be forced into an α-helix by adding hydrogen bond distance restraints. Distance restraints have the following format:

assign (resid and name) (resid and name) A B C

The following values were then inputted:

A. Target distance (Å)
B. Lower bound difference (Å)
C. Upper bound difference (Å)

In this format, B and C define the range of possible allowed values around the given distance. To make a sequence folding to an α-helix, the following input was used:

assign (resid i and name O) (resid i+4 and name N) 2.9 0.2 0.2

This is a tight distance restraint which forces the i+4 hydrogen bond characteristic of α-helical secondary structure. In MmsFcc, residues 3-46 and 64-106 were confined to an α-helix conformation, based on the TALOS predicted secondary structure shown in Appendix 6.12. While TALOS has predicted many of the loop residues to be in helical conformation, these have not been restrained with hydrogen bonds to allow the peptide chain to turn 180°.

-

MmsFcc was designed to have a leucine zipper (Gurnon et al., 2003; Landschulz et al., 1988; Vinson et al., 2002). Leucine zippers are common motifs in coiled-coils, which function to keep the two coils in register by interlocking along the hydrophobic faces of the helices. The AlphaFold structure is shown in Figure 4.29, with leucine zipper residues highlighted.

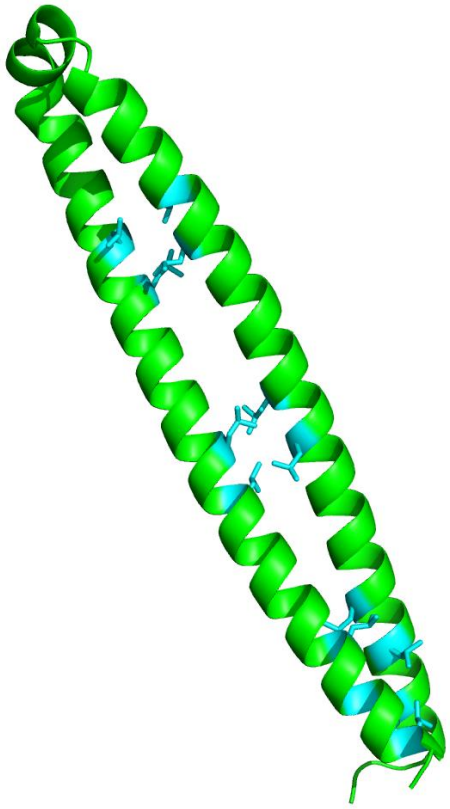


Figure 4.29. AlphaFold2 prediction of MmsFcc structure, with leucine residues highlighted in blue.

Distances were measured between leucine Cα-Cα, Cβ-Cβ, and Cδ-Cδ for the following zipper pairs from the AlphaFold2 prediction:

- L7-L105
- L11-L101
- L14-L98
- L25-L87
- L28-L84
- L39-L73
- L42-L70

These distances were inputted into the simulations as distance restraints, the coiled-coil was twisting excessively. To fix this, the distance between the ends of each α-helix was measured from the AlphaFold structure then formatted into a distance restraint, essentially designed to keep the structure linear and stop it getting over-twisted. The restraints used are provided below:

assign (resid 1 and name CA) (resid 62 and name CA) 69.1 4 0.0

assign (resid 48 and name CA) (resid 107 and name CA) 66.8 4 0.0

With these restraints, it was possible to reliably simulate reproducible coiled-coil structures with low energy penalties. The next step was to use experimental NOEs between loop residues to define distance restraints and determine the metal-binding loop structure. The Nuclear Overhauser effect is a phenomenon that occurs between spin-active nuclei which are in close proximity (Solomon, 1955). NOEs are detectable by NMR and provide information on the distance between nuclei. Therefore, they can used to calculate distance restraints in molecular dynamics calculations.

-

Experimental NOEs were collected from the ^{15}N HSQC-NOESY, ^{13}C Aromatic NOESY, and ^{13}C Aliphatic NOESY spectra. For the ^{15}N HSQC-NOESY spectrum, one typically sees inter-residue NOEs between NH_i and NH_{i+1} for α-helical sequences. Therefore, these NOEs were included if observed in the loop sequence, since they strongly suggested a helical structure. To calibrate the distances from observed NOEs, the standard approach involves measuring the intensities of NOEs for protons separated by a known fixed distance. For example, the distance between *i* and *i+1* backbone amide protons in an α-helix conformation is 3.2 Å. For MmsFcc, peak heights were selected as a measure of intensity. For the ^{15}N HSQC-NOESY spectrum, NH_i and NH_{i+1} NOEs were recorded from α-helical residues and their intensities averaged. These are displayed in Table 4.6.

I	*i+1*	peak height
e8	k9	2026000000
q19	a20	4526000000
a20	i21	4251000000
q33	a34	4059000000
k103	k104	2328000000

Table 4.6. NOEs of known distance from 15N HSQC-NOESY spectrum, used to calculate distance for NH-NH NOEs.

Some NOEs appear to be twice as intense as others. This is because 2 identical NOEs between residues in each of the α-helices are being observed (e.g. q19-a20 in Table 4.7). Since two NOEs were contributing to this intensity, the intensity value was halved before being included in calculations. An average distance for this NOE was measured from the

AlphaFold structure, then both values were used in the following equation to define NOE distance restraints:

$$Distance_{obs} = Distance_{ref}^{-4} * (Intensity_{obs}/Intensity_{ref})^{-4}$$

The exponent of -4 (instead of the expected -6) was used because conformational flexibility tends to reduce the r-6 dependence to something closer to r^{-4} (M. Williamson, 2011). Distance restraints were given a minimum bound of 1.8 Å, which is the minimum distance between protons before steric clashes occur. This process was repeated for each type of NOE detected when solving the structure of MmsFcc, using reference NOEs involving the same nuclei as the observed. All observed loop NOEs used to determine the structure of MmsFcc, and the calculated distances, are listed in Table 4.7.

Proton 1	Proton 2	Upper bound (Å)	Lower bound (Å)
F57 Hδ	F57 Hα	3.8	1.8
D51 Hβ	Q61 Hβ	5.0	1.8
F57 Hα	A59 Hβ	5.5	1.8
F54 Hε	D51 Hα	3.8	1.8
F54 Hε	Q61 Hβ	3.8	1.8
Y56 Hδ	A59 Hβ	5.3	1.8
Y56 Hε	E53 Hβ	4.7	1.8
F57 HD	K60 Hγ	4.5	1.8
K44 HN	K45 HN	3.0	1.8
K45 HN	L46 HN	3.1	1.8
L46 HN	Q47 HN	3.2	1.8
A48 HN	D49 HN	2.6	1.8
F54 HN	V55 HN	2.6	1.8
D51 HN	Q47 Hβ	3.1	1.8
F54 HN	F54 Hδ	3.1	1.8
Y56 HN	Y56 Hδ	3.1	1.8
F57 HN	F57 Hδ	3.4	1.8
K60 HN	A59 Hα	3.6	1.8

Table 4.7. Summary of NOE distance restraints for MmsFcc metal-binding loop.

No NOEs were picked for coiled-coil residues since these spectral regions were too crowded to accurately determine NOE pairs. It was not possible to confidently pick more NOEs than presented in Table 4.7, which represent a lower number of NOEs than usual for structural determination. This is to be expected of a 13-residue loop sequence that is not buried into a protein structure.

-

After the final set of restraint and parameter files were prepared, 120 calculations of MmsFcc structure were performed. The 55 simulations with the lowest energy violations were then ranked and selected for water refinement. A water refinement improves the quality of the structural calculation by including water molecules into the simulation, resulting in more realistic forcefields and thus better energies (Chen et al., 2004). The 20 water refined structures with the most negative energy were selected as the best structures of MmsFcc. The statistics of these structures are shown in Table 4.8.

Experimental NOE distance restraints (residues 44-61)	
Intraresidue	4
Sequential	5
Short range	7
Long range	2
Number of restraints per loop residue	1.38
Average total energy violations	8.44
Average energy violation per NOE restraint	0.46
AlphaFold derived distance restraints (coiled-coil residues only)	
Intraresidue	0
Sequential	0
Short range	0
Long range	18
Total	18
TALOS-N derived restraints	
Dihedral angle restraints (all residues)	241
Hydrogen bond restraints (coiled-coil residues)	84
RMSD (Å)	
Average RMSD heavy atoms (all residues)~	1.95
Ramachandran plot summary (Molprobity)$^{+}$	
Most favoured regions (% of residues)	98.2 %
Allowed regions (% of residues)	1.8 %
Ideal geometry deviations and close contacts^{+}	
Average number of close contacts per structure*	3.4
RMS deviation for bond angles	0.6 °
RMS deviation for bond lengths	0.004 Å

Table 4.8. Statistical analysis of the twenty best structural predictions of MmsFcc. ˜Pairwise alignment of structures performed using mTM-align software (Dong et al., 2018). $^{+}$Values calculated by Protein Structure Validation Software (PSVS) version 2.0: https://montelionelab.chem.rpi.edu/PSVS/psvs2/. *Close contacts defined as within 1.6 Å for H atoms, and within 2.2 Å for heavy atoms. Residues 1-48 and 62-107 are coiled-coil residues.

The number of restraints per residue represents a lower than typically expected value for a trusted structure. For an acceptable structure, typically 10-20 restraints are observed on average (Fowler et al., 2021). This is significantly more than observed in the loop of MmsFcc. One reason for this is that NOEs which provide no useful information, such as many intraresidue NOEs, were excluded for the sake of brevity. Typically, loops also contain less NOEs. These twenty structures were aligned using the mTM-align software, with the resulting alignment shown in Figure 4.30.

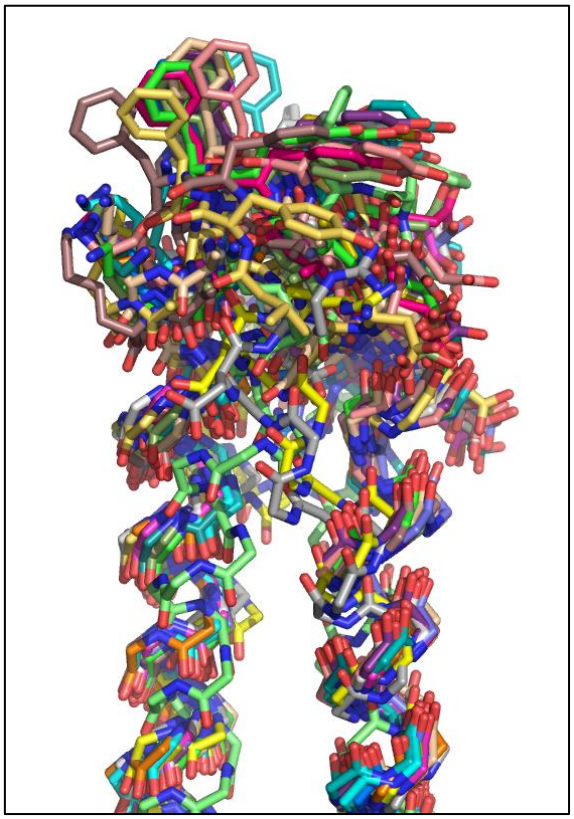


Figure 4.30. mTM-align alignment of twenty best MmsFcc structures in PyMol, aligning all residues. The sidechains of all residues are excluded in this visualisation, except for residues 49-61.

From this alignment, it is hard to determine whether there is any structural similarity between the loop sidechains of MmsFcc. This is so since the alignment attempts to align all residues in the structure which has resulted in the misalignment of the loop residues for the better alignment of the coiled coil residues. The loop residues of these twenty structures (residues 49-61) were aligned using the PyMol align command, in order to gauge the uniformity of the loop structure. This is represented in Figure 4.30

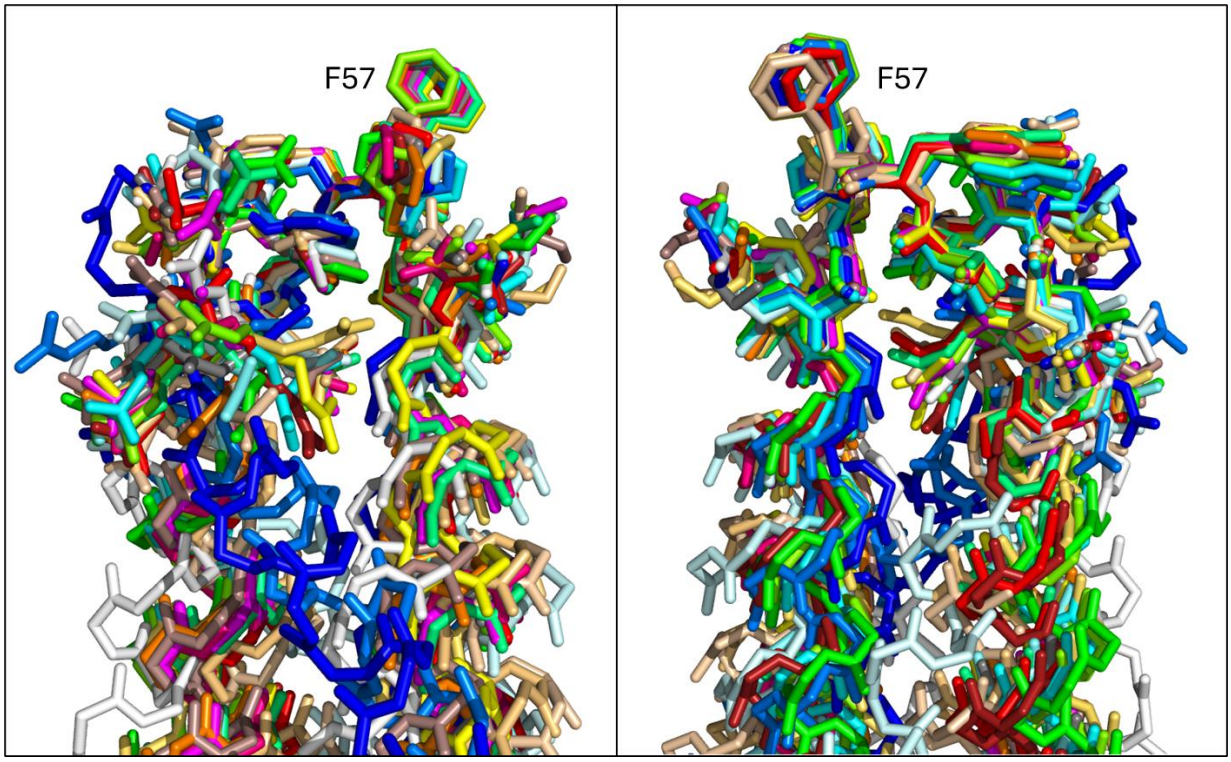


Figure 4.31. PyMol alignment of twenty best MmsFcc structures, aligning only residues 49-61.

From this alignment, we can see that there is some uniformity in the loop structures of these twenty refinements. This uniformity appears to be most apparent in the centre of the loop. One residue of interest is F57, which appears to be consistently exposed on the surface of the loop, which is an untypical arrangement for the hydrophobic residue. The structure of MmsFcc is discussed further in the next section.

4.7 Chapter discussion

As discussed in Chapter 1, histidine is a common key residue in metal binding proteins (especially for iron ligands), making it likely to be a key component of MmsFcc metal binding.

One interesting aspect of this structure is the contrast of positions between F54 and F57. Phenylalanine has a very hydrophobic sidechain and is typically buried within protein structures. F54 is behaving like this, while the F57 sidechain is presented to the solvent. This is an energetically unfavourable position, which suggests that it is functionally important. There is strong evidence that the structure of MmsFcc changes upon binding of metals, discussed in detail in Chapter 5. Co-precipitation data from the F54/F57 double mutant presented in Section 1.x strongly suggests that these residues play a key structural role in

proper MmsFcc function. One of the potential explanations is that some form of stacking interaction was present, which was replaced by a different (hydrogen bond-like?) interaction in the F54Y/F57Q double mutant. This structure does not support this hypothesis, at least in the unbound state. It is important to remember that these co-precipitation reactions show how effective the mutant is at producing full sized nanoparticles, so any potential explanation for this observation may not be elucidated from looking at the free structure. Another explanation for the position of the F57 sidechain is that it is mediating the formation of higher order assemblies of MmsF. Rawlings *et al* demonstrated in their 2014 paper that MmsF forms soluble proteinosomes in solution, suggesting that native protein function may involve some higher order assemblies. It should be said that Rawlings *et al* did show in their 2019 paper that MmsFcc was monomeric in solution, however it is possible that this formation of assemblies is reliant on a sufficient concentration of iron.

A closer look at the structures suggests a potential binding site for metals (Figure 4.32).

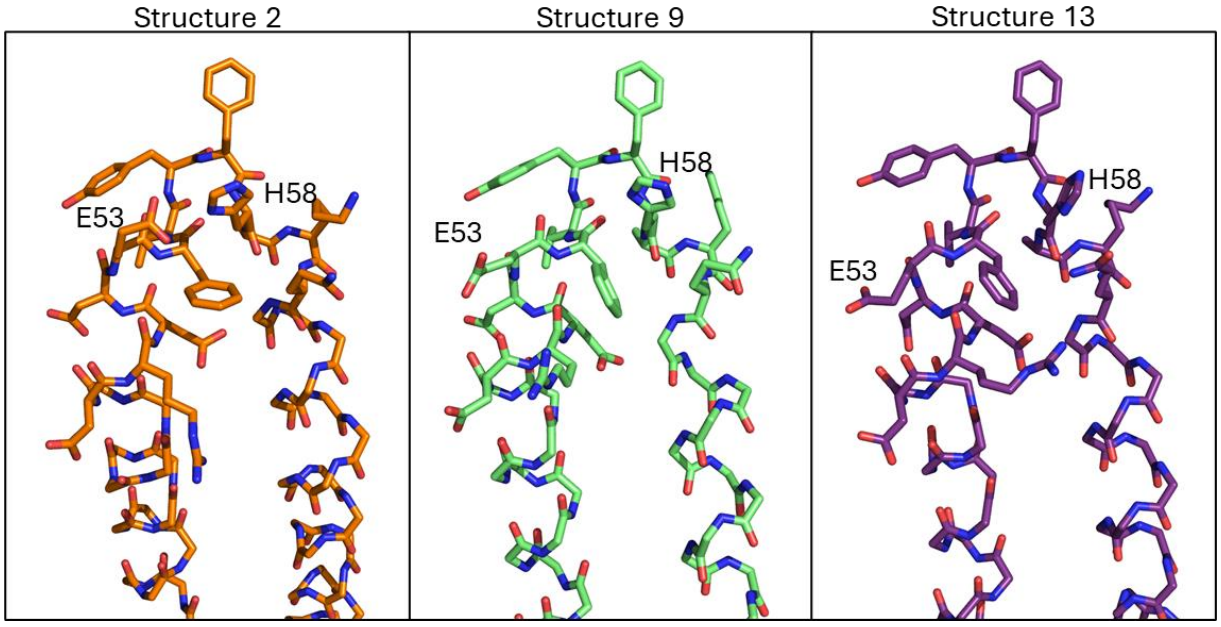


Figure 4.32. Comparison of 3 MmsFcc structures, displaying the variety of positions that the glutamate sidechain can occupy.

Figure 4. shows how the sidechain of glutamate 53 can adopt a variety of positions. Critically, in structure two the carboxylate group comes into proximity with the imidazole ring of histidine, creating a potential metal binding site. It should be noted that, of the twenty structures selected, this degree of closeness was only observed in structure 2. By considering the structure of the loop in the context of its energy landscape, a possible

explanation for this observation is that free MmsFcc exists in a range of higher energy conformations. Initial interaction with metal may then induce a conformational change to a single lower energy structure, which positions the metal optimally for proper chelation. Due to the relative weak affinity, this complex is then likely released quicky. A strong candidate for this proposed initial interaction would be the three aspartate residues (D49, D51, D52). It was thought that these sidechains would play an important role in metal binding, based on their prevalence in the literature (Section 1.8). However, a closer look at them in the structure shows there is little structural evidence for their role in metal binding (Figure 4.32).

This lack of structure may instead be explained by a role in changing the conformation of the loop, upon transient interaction with metal. It should be noted that it was difficult to collect NOEs from aspartate residues, since they regularly occupied crowded regions of spectra. Given that the structural determination programme used includes electrostatic terms which accounts for the repulsion of similarly charged sidechains, it is also likely that a candidate metal binding site would not be obvious from these free structures. This is so as the presence of the metal cation may be required to facilitate the closer interaction of similarly charged residues. The interaction of MmsFcc and MmsFcc D49S with metals was analysed via titrations, which is presented in Chapter 5.

Chapter 5: MmsFcc and MmsFcc D49S selectively bind Fe^{2+} and Zn^{2+}, with a mechanism dependant on histidine deprotonation

A well-established method of assaying protein-ligand interaction via NMR is through ligand titrations (Becker et al., 2018a; Furukawa et al., 2016b; M. P. Williamson, 2013a). This is usually done by adding increasing concentrations of a ligand to fixed concentration of protein, then recording spectra to observe the effect of ligand concentration on chemical shifts and peak line shapes. In this case of MmsFcc, a series of metals were titrated against MmsFcc and MmsFcc D49S, with ^{15}N HSQC spectra being recorded at each interval. The metals selected were Fe^{2+}, Fe^{3+}, Ca^{2+}, and Zn^{2+}. Fe^{2+} and Fe^{3+} were selected since they are the biologically relevant redox states of iron, which is the primary component of the magnetite nanoparticles formed by magnetotactic bacteria. Since Fe^{2+} is paramagnetic, it was predicted that adding large concentrations of the cation would have deleterious effects on spectral quality, such as paramagnetic line broadening. For this reason, Zn^{2+} and Ca^{2+} were also selected as alternative divalent cations.

The addition of a ligand can alter a range of different chemical shift influencing factors, either by causing a conformational change, or by through-space or through-bond effects at the site of binding. Despite this range of factors that contribute to the effect observed when adding ligand, it is usually true that the residues affected are either directly involved in binding or exist near the binding site. This has established NMR titrations as a powerful tool for assaying protein-ligand interactions in solution (M. P. Williamson, 2013a). The rate of exchange between bound and free conformations will also affect the appearance of peaks. If the rate of exchange is at least ten times faster than the chemical shift difference (in Hz) between bound and free, a single peak is observed. This is called fast exchange, and the observed peak is an average of the ratio of bound/free conformations. If the rate of exchange is at least ten times slower than Δδ, the protein is in slow exchange. In this scenario, two distinct peaks representing bound and free exist simultaneously. If Δδ is roughly equivalent to the exchange rate, then the effect of intermediate exchange is observed. In this scenario, the two peaks are broadened to the point that they become undetectable (Kleckner & Foster, 2011b). A bound signal can be observed if the protein is saturated with ligand. This is visualised in Figure 5.1.

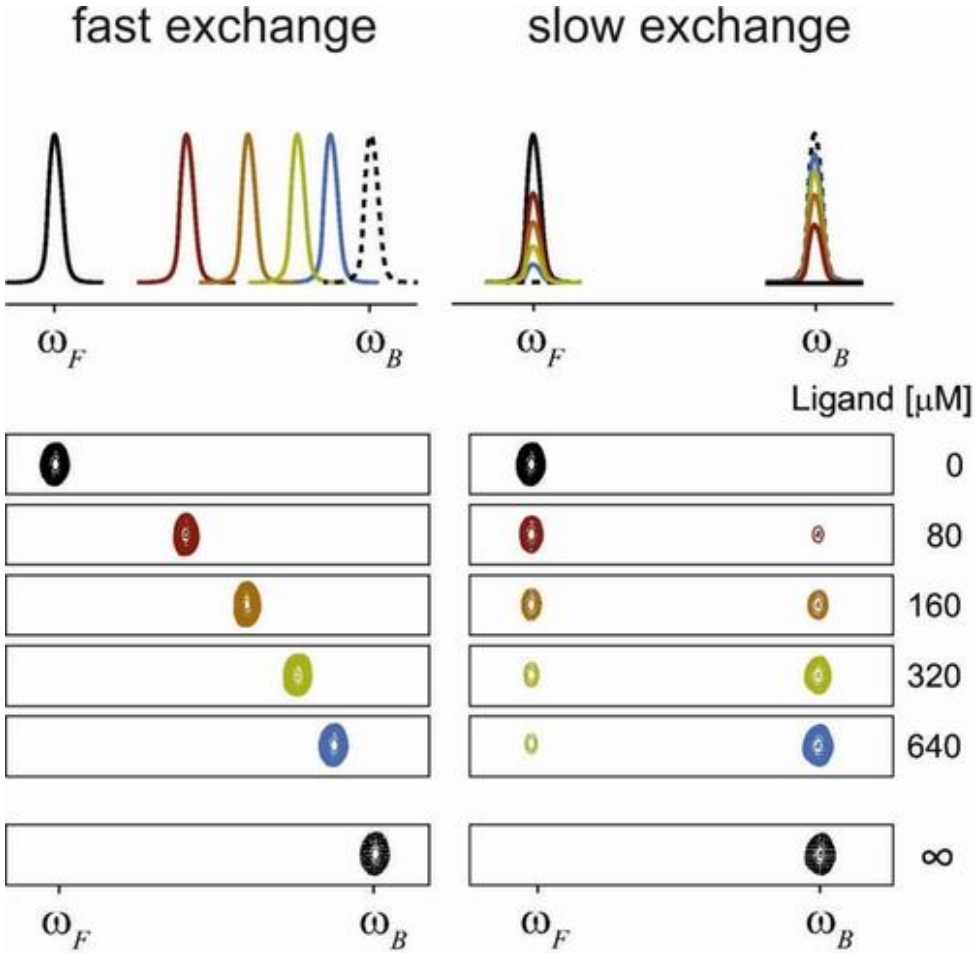


Figure 5.1. One and two-dimensional representation of the effect of exchange rate on titrating peaks. ω_B and ω_F denote 100% bound and 100% free respectively. Figure taken from (Becker et al., 2018).

It is important to keep buffer conditions as consistent as possible when performing titrations, since small differences in ionic strength or pH can alter shift values. The full methods describing the preparation of titration samples is found in Section 2.14. For each ligand concentration, a separate NMR sample was prepared. This requires more protein than the alternative method of adding ligand to the same sample but allowed for more precise control of pH and concentrations and thus more accurate results. A working concentration of 72 μM protein was used for all titrations. For the rest of this chapter, concentrations of ligand are given in relation to the protein concentration (e.g. 32x titration means 32 times the concentration of protein). In general, the metal salt was dissolved in buffer conditions identical to the protein, then the pH adjusted before addition. In the ligand buffer, the concentration of NaCl was adjusted to keep the ionic strength of the sample as consistent as possible. It is possible for ligand binding to alter pH, for this reason sample pH was adjusted

after addition of ligand to match the pH of the free sample. In the case of Fe^{2+}, efforts were made to keep the sample under nitrogen. This was necessary since Fe^{2+} is readily oxidised to Fe^{3+} by oxygen dissolved in water. The preparation of Fe^{2+} titration samples is described in Section 2.15.

Trial titrations showed the chemical shift changes to be quite small, for this reason titration concentrations of 2x, 4x, 8x, 16x and 32x were selected as standard. Initial observations are made in Section 5.1. Further information on the pulse programmes used, and the number of scans collected can be found in Section 2.18.

5.1 Initial observations show that addition of Zn^{2+} and Fe^{2+} causes line broadening; Ca^{2+} and Fe^{3+} do not bind

In a typical protein-ligand titration, the direction in which titrating peaks move is consistent. Some titration datapoints were not included for analysis since the titrating peaks had shifted in a direction not consistent with the rest of the dataset. Since it was still possible to overlay the majority of peaks on top of each other, it was presumed that the error was a result of poor sample preparation. Datapoints excluded are listed in Table 5.1.

MmsFcc						
	2x	4x	8x	16x	32x	300x
Fe^{2+}	×	✓	✓	✓	×	✓
Fe^{3+}	✓	✓	✓	✓	✓	×
Ca^{2+}	✓	✓	✓	✓	✓	×
Zn^{2+}	✓	✓	✓	✓	✓	✓
MmsFcc D49S						
	2x	4x	8x	16x	32x	300x
Fe^{2+}	✓	✓	✓	✓	✓	×
Fe^{3+}	✓	✓	✓	✓	×	×
Ca^{2+}	×	✓	✓	✓	✓	×
Zn^{2+}	✓	×	✓	✓	✓	×

Table 5.1. Table showing unviable datasets from MmsFcc and MmsFcc D49S metal titrations. Datasets excluded from analysis are denoted with a ×.

300x samples were trialled in an attempt to saturate the protein with ligand, creating a near fully bound spectrum. This was only done for wildtype with Zn^{2+} and Fe^{2+}, since these 2 combinations showed the most binding. For the lower concentration titrations, such as D49S + Zn^{2+} 4x, this is likely due to error in sample preparation resulting in slight changes in conditions that affect chemical shift such as pH and ionic strength. These errors are

amplified in titration datasets such as these, where the overall shift differences are relatively small when compared to other examples of protein-ligand titrations. Since MmsFcc's interaction with Fe^{2+} was of great interest to this study, multiple attempts were made to produce a viable 32x sample. Unfortunately, none of these attempts worked. It is possible that this is due to paramagnetic broadening effected by the Fe^{2+} cations, although the 32x Fe^{2+} for D49S was viable and matched the trend. Because the effect of paramagnetic broadening is related exponentially to distance, it is possible that tighter binding in the wildtype results in more severe paramagnetic broadening.

When these titrations were analysed, it was immediately noticed that peak intensity appeared to reduce, and this reduction correlated with the concentration of ligand added. Loss of peak intensity as a result of line broadening is a commonly observed phenomenon in protein-ligand titrations (M. P. Williamson, 2013a). This effect was observed only in Fe^{2+} and Zn^{2+} titrations, for both wildtype and D49S spectra (although the reduction in intensity was lower in D49S titrations). Spectra from Zn^{2+} and Fe^{2+} titrations against MmsFcc are displayed in Figures 5.2 and 5.3.

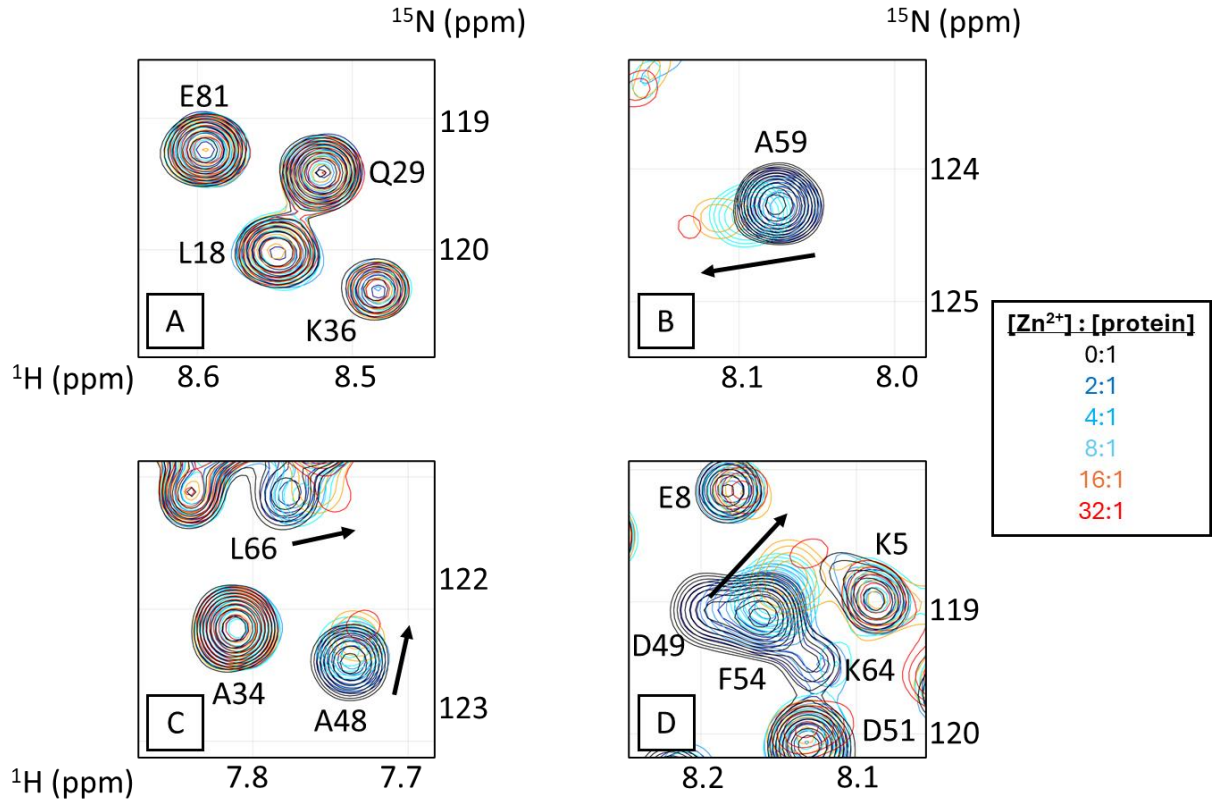


Figure 5.2. HSQC spectra of Zn^{2+} titrations against MmsFcc, demonstrating the additive line broadening effect induced by protein-ligand interaction. Panel A shows the consistent peak

intensities of peaks unaffected by addition of metal. Panels B-D highlight peaks that reduce in intensity because of metal addition.

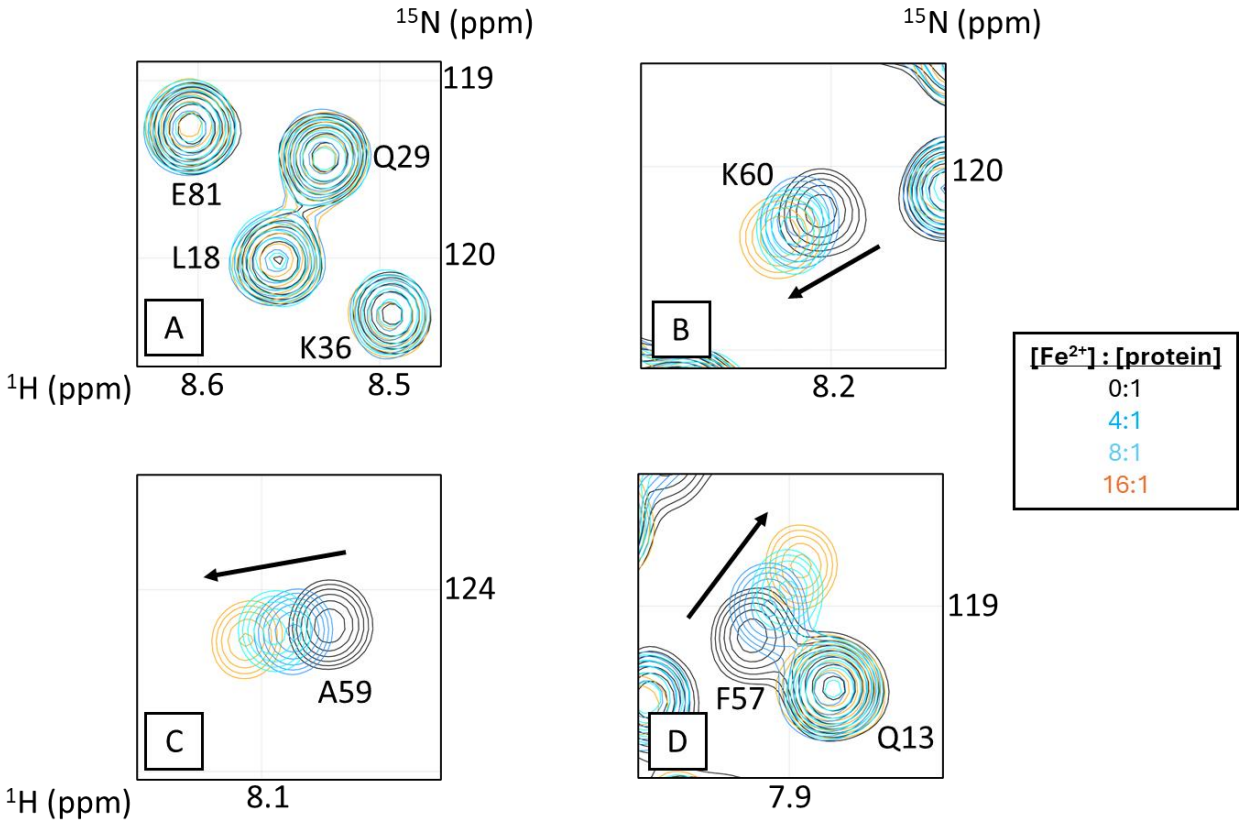


Figure 5.3. HSQC spectra of Fe^{2+} titrations against MmsFcc, demonstrating the additive line broadening effect induced by protein-ligand interaction. Panel A shows the consistent peak intensities of peaks unaffected by addition of metal. Panels B-D highlight peaks that reduce in intensity because of metal addition.

From these spectra, it is apparent that the reduction in intensity is much greater in Zn^{2+} titrations than with Fe^{2+}. Peak heights represent a measure of signal intensity. Therefore, by dividing the peak heights derived from the highest concentration spectrum by the free peak heights, the degree of change in peak intensity could be derived.

Where Peak Height A refers to the peak height in the free spectra, and Peak Height B prefers to the peak height in the spectra with the highest ligand concentration. The results of this analysis are shown in Tables 5.2 and 5.3. Residues 46-66 were included since they were the only residues to exhibit this effect. Some residues (such as D49, E53) were not included since they inhabited populated regions of the spectra, making it impossible to accurately determine peak heights.

	0x Peak height	32x Peak height	0x/32x
L46	6.81E+08	2.49E+08	2.74
A48	1.13E+09	3.36E+08	3.37
R50	1.19E+09	2.17E+08	5.47
D51	1.04E+09	5.17E+08	2.02
F54	1.60E+09	3.20E+08	4.99
F57	1.08E+09	9.61E+07	11.20
H58	7.26E+08	4.46E+08	1.63
A59	1.14E+09	1.83E+08	6.22
K60	8.20E+08	1.75E+08	4.68
Q61	1.15E+09	2.79E+08	4.11
G62	5.60E+08	1.98E+08	2.83
K64	5.32E+08	1.73E+08	3.08
Q65	7.65E+08	5.66E+08	1.35
L66	7.03E+08	3.23E+08	2.17

Table 5.2. Difference in MmsFcc HSQC peak height after addition of 32x Zn^{2+}.

	0x Peak height	16x Peak height	0x/16x
L46	6.07E+08	3.89E+08	1.56
Q47	7.67E+08	5.81E+08	1.32
A48	1.06E+09	6.75E+08	1.57
R50	1.1E+09	4.43E+08	2.48
D51	1.12E+09	4.37E+08	2.57
F54	1.33E+09	9.4E+08	1.41
F57	8.84E+08	6.82E+08	1.30
H58	7.57E+08	4.46E+08	1.70
A59	9.02E+08	6.64E+08	1.36
K60	7.34E+08	6.48E+08	1.13
Q61	7.31E+08	4.81E+08	1.52
G62	4.64E+08	7.55E+08	0.61
K64	6.13E+08	4.92E+08	1.25
Q65	6.4E+08	5.55E+08	1.15
L66	5.92E+08	4.29E+08	1.38

Table 5.3. Difference in MmsFcc HSQC peak height after addition of 16x Fe^{2+}.

Both the magnitude of intensity change, and the residues affected are different for the two metals. This supports the evidence in Section 5.2 suggesting that the mode of binding for the two metals is different. Another possibility is that the rate of exchange is different in both cases. As mentioned previously, a few potential explanations exist for this line broadening effect, which are explored in detail in Section 5.2.

-

300x concentration spectra were recorded for Fe^{2+} and Zn^{2+} with wildtype. Both saturated spectra were collected overnight with 128 scans, in an attempt to reveal peaks whose intensity was severely reduced by line broadening (Protocol 2.16) (Figure 5.4-5.5)

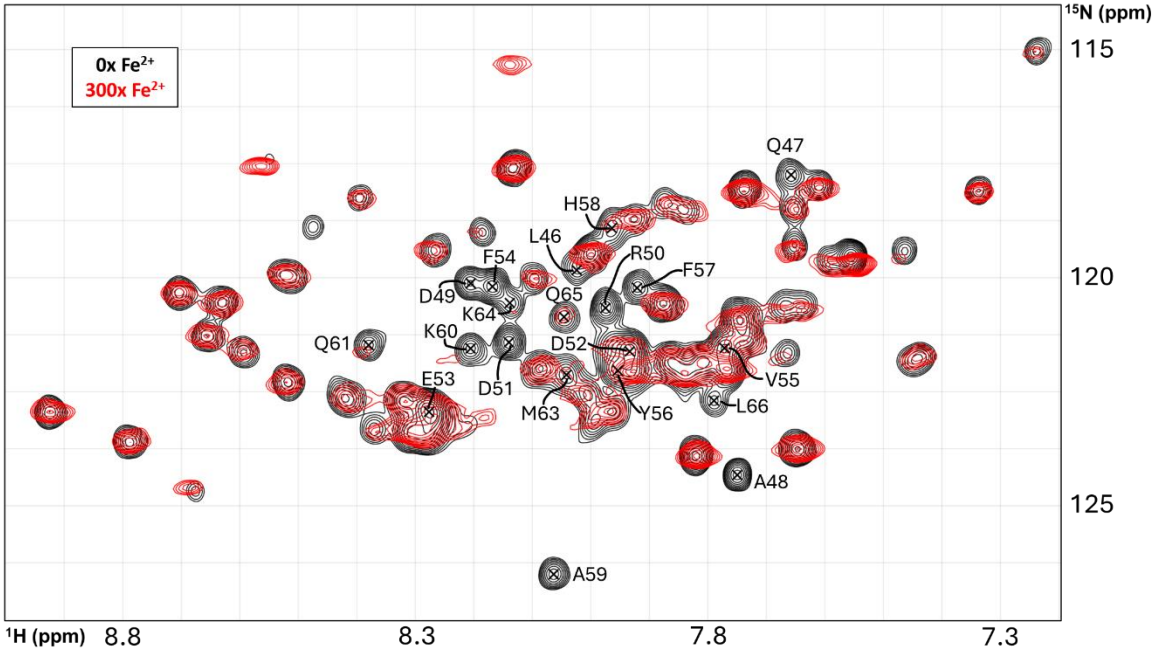


Figure 5.4. HSQC spectra of free MmsFcc and MmsFcc + 300x Fe^{2+} overlaid. Loop residues 46-66 are annotated. 300x spectrum was collected on a 600 MHz spectrometer with a cryoprobe.

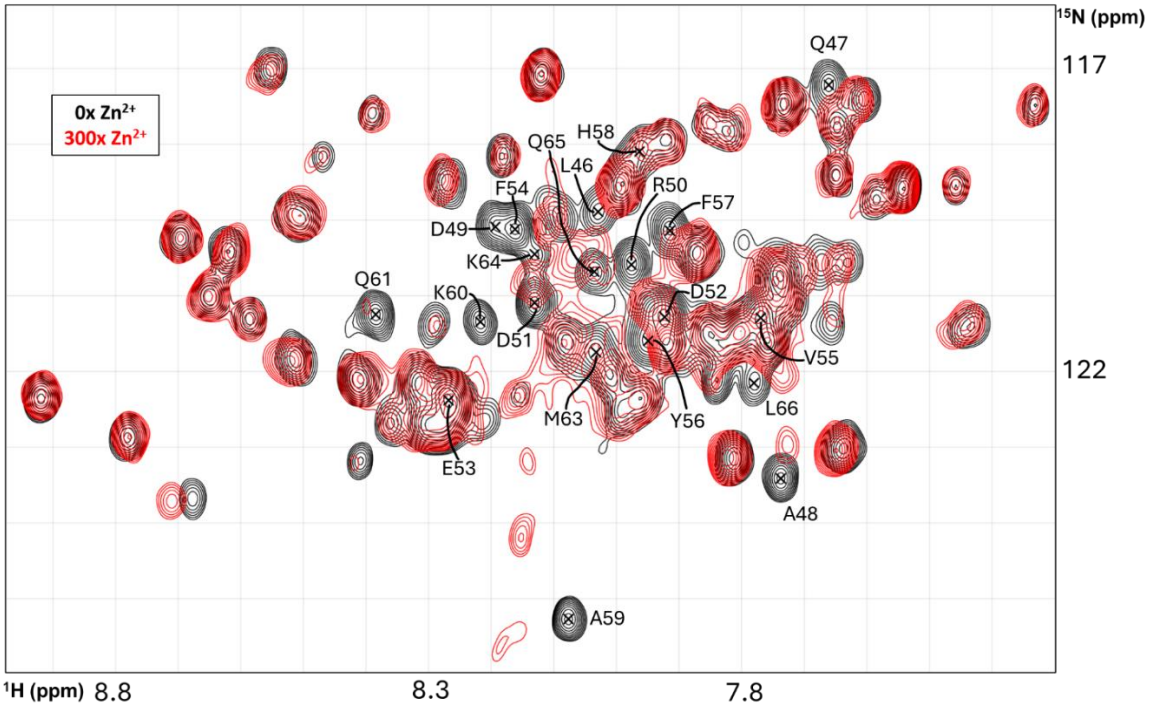


Figure 5.5. HSQC spectra of free MmsFcc and MmsFcc + 300x Zn^{2+} overlaid. Loop residues 46-66 are annotated.

Saturation of MmsFcc with Fe^{2+} causes the majority of peaks between residues 46-66 to completely disappear due to line broadening, with exception of a couple of very broadened peaks at K60, Q61, Q65. If this is not the result of paramagnetic line broadening, then something related to the binding event is causing this broadening.

For saturation with Zn^{2+}, the same loss of peaks is observed except for a few (A48, D51, L66). In this case, there is no possibility of paramagnetic line broadening, so the effect observed is due to exchange. To properly interpret this data, an estimation of the percentage of bound protein is needed. This requires the calculation of the dissociation constant, K_D, the process of which is described in Section 5.3. The next section takes a closer look at the specific residues shifting in each titration scenario.

5.2 MmsFcc and MmsFcc D49S each interact differently with different metals

While a variety of factors are known to influence chemical shift to different degrees, it is generally true that the largest chemical shift changes represent residues undergoing the biggest conformational change. For this reason, the total difference in peak position for each residue was calculated using the following equation:

$$\Delta\delta=\sqrt{[\Delta\delta H^2+ (0.14*\Delta\delta N)^2]}$$

For each scenario, the free spectrum and the highest ligand concentration spectrum was compared. As discussed previously, nitrogen shifts were multiplied by 0.14 to avoid over favouring them in the analysis.

The titration of Ca^{2+}, Fe^{2+}, Fe^{3+}, and Zn^{2+} against both MmsFcc and MmsFcc D49S were analysed in this fashion, with MmsFcc titrations presented first in Figures 5.6-5.9. For these analyses, it was not possible to track peak position for the following peaks due to overcrowding in the spectrum: Q6, L7, I21, K23, Q24, A26, L28, W30, K37, L42, D49, I80, K82, Q83, A85, W89, K96, and L101. For loop peaks in crowded spectral regions (e.g. E53, V55), changes to the shift of the most prominent peak in that region was attributed to the loop residue. This was done since the data shows that mainly loop and loop-adjacent peaks shift

upon titration with metals. However, this is inherently less accurate than tracking a single peak, so these datapoints should be treated with more caution.

For MmsFcc + Fe^{2+}, the 16x titration point was used for comparison.

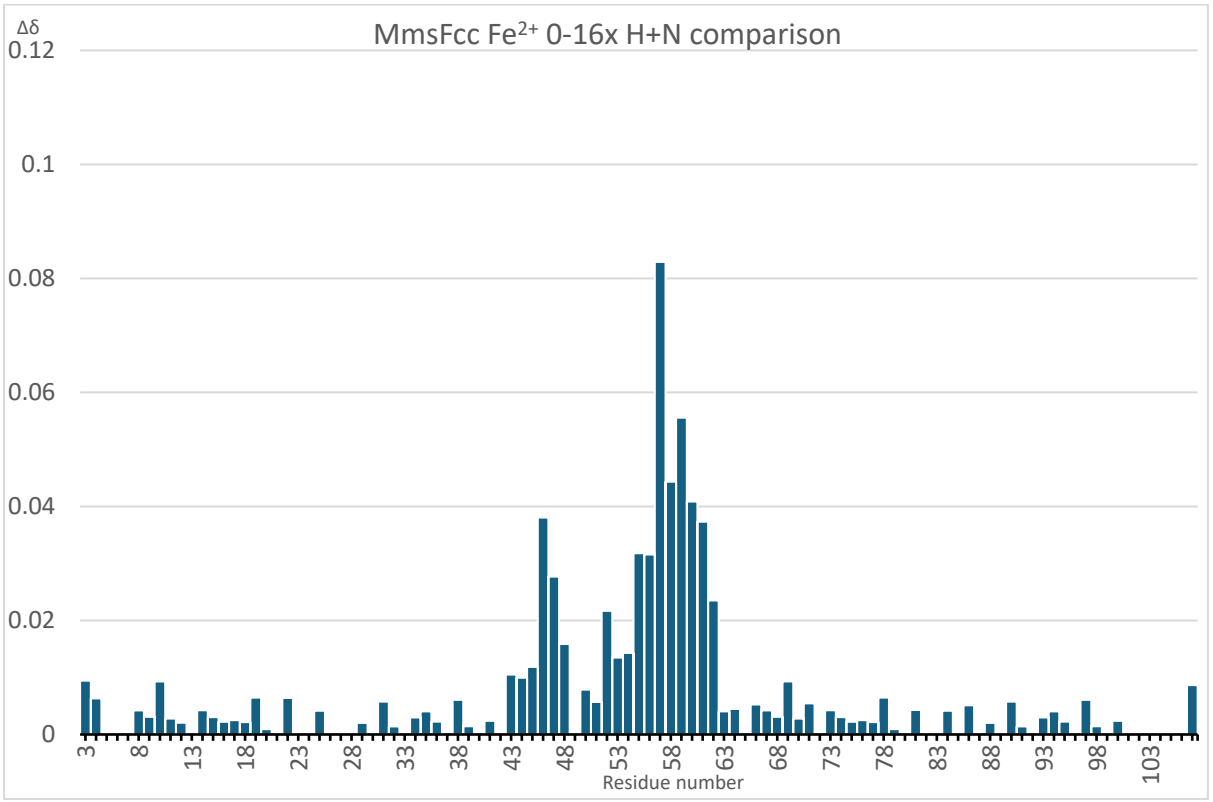


Figure 5.6. Difference in MmsFcc HSQC peak position for 0x and 16x Fe^{2+} titration points. Proton and nitrogen Δδ were calculated separately.

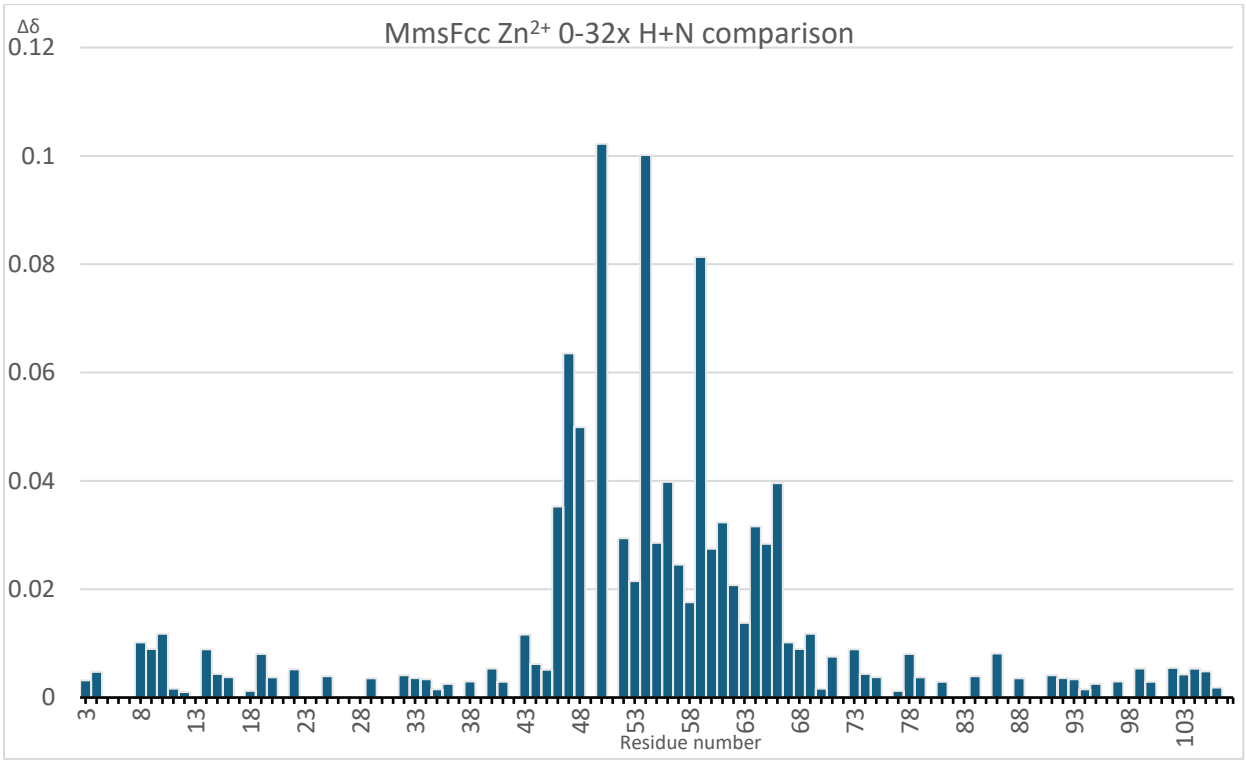


Figure 5.7 Difference in MmsFcc HSQC peak position for 0x and 32x Zn^{2+} titration points. D51 residue excluded since it was not possible to track peak migration. Proton and nitrogen Δδ were calculated separately.

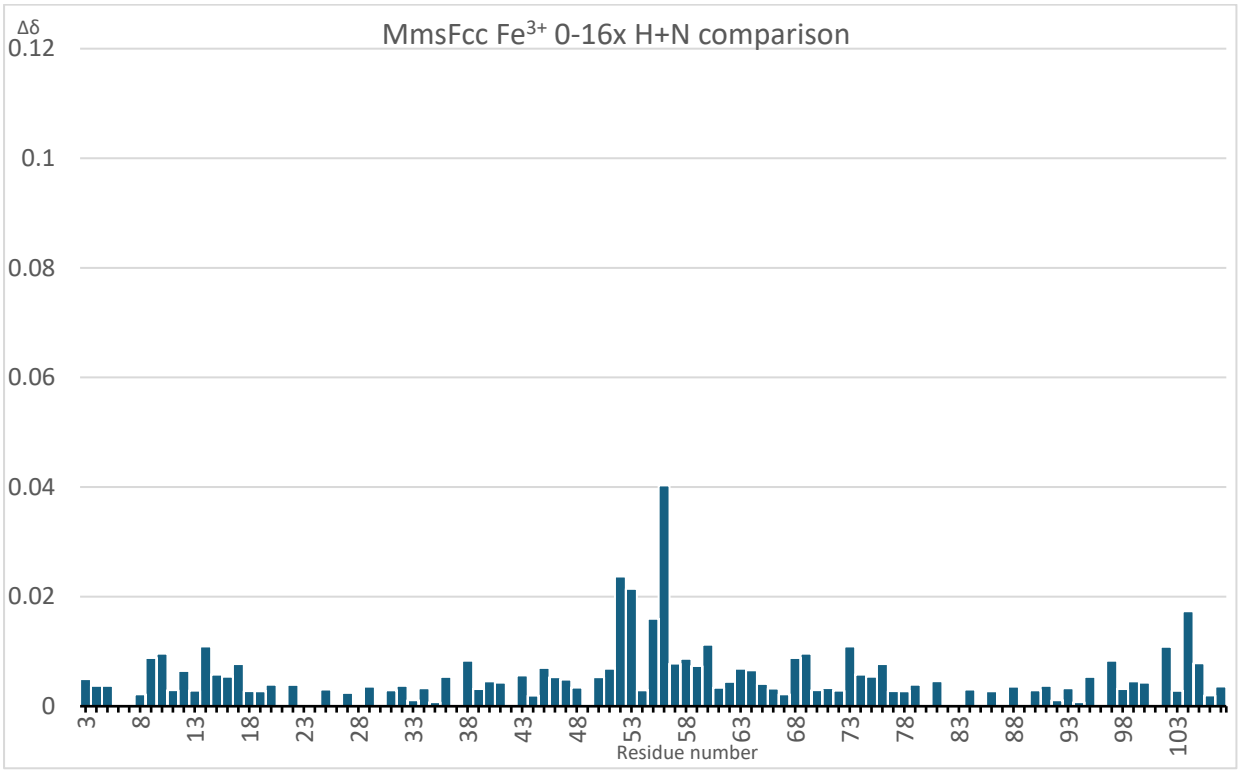


Figure 5.8 Difference in MmsFcc HSQC peak position for 0x and 32x Fe^{3+} titration points. Proton and nitrogen Δδ were calculated separately.

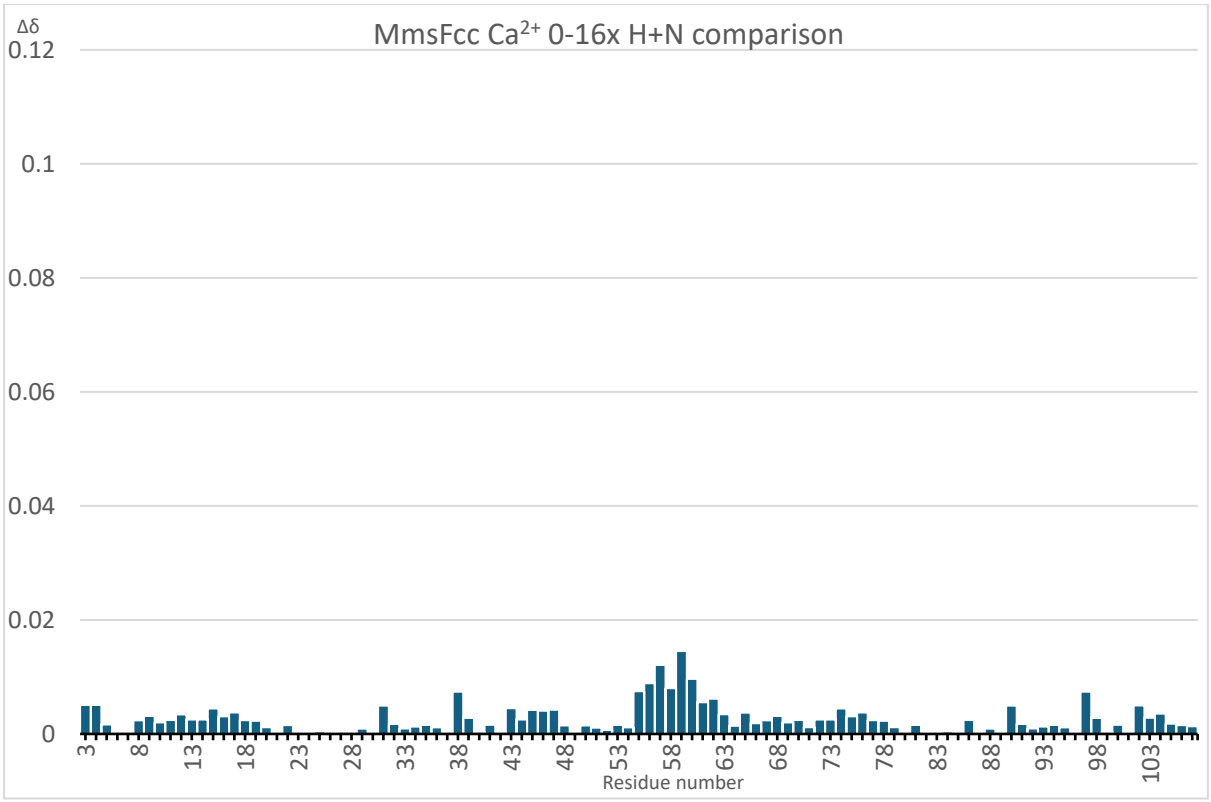


Figure 5.9. Difference in MmsFcc HSQC peak position for 0x and 32x Ca^{2+} titration points. Proton and nitrogen Δδ were calculated separately.

As expected, the largest shift changes are confined to the metal binding loop (residues 49-61). While this is the case, it was not uncommon to see significant shift changes in coiled-coil residues surrounding the loop (specifically L46, Q47, A48, G62, M63, K64, Q65, L66). These residues are not present in the native protein sequence, where their equivalent residues would likely be buried as part of the transmembrane helix. They are also not typically metal binding residues, and for these reasons are likely to not be directly involved with metal binding. It is instead strong evidence that there is a conformational change related to the binding event, which is affecting the chemical environment of residues bordering the loop.

Interestingly, addition of Fe^{2+} has a much greater effect on chemical shifts than addition of Fe^{3+}. This fits the hypothesis suggested in section 1.9, where MmsFcc may increase the formation of Magnetite crystals by improving the incorporation of more soluble Fe^{2+} ions. While the Fe^{3+} cation is more positively charged, and therefore more likely to have a stronger affinity for negative charges, in solution this means it is more shielded by water. This likely explains the difference observed. Because of this affinity for water, Fe^{3+} has a propensity to form insoluble hydroxide precipitate in solution. However, this is concentration dependant,

and no visible precipitate was observed in the samples analysed. It is plausible that MmsF may possess affinity for Fe^{3+} in the magnetosome lumen, although it is hard to speculate because the microenvironment maintained by the magnetosome has not been elucidated.

Ca^{2+} appears to not bind to either protein, which could be due to calcium's preferred coordinate complexes. Both Zn^{2+} and Fe^{2+} tend to form complexes with coordinate numbers of 4-6 (co-ordinate number 6) complexes, whereas Ca^{2+} tends to form complexes with a co-ordinate number of 6-8 (Katz et al., 1996b; Laitaoja et al., 2013; Sánchez et al., 2017a). It is therefore possible that the loop of MmsFcc has specificity for forming lower coordinate number complexes.

Focusing on the 2 binding cations, the binding mechanism of Zn^{2+} to MmsFcc differs from that of Fe^{2+}, evidenced by the different profiles in Figures 5.x and 5.x. With Zn^{2+}, residues at the N-terminal end of the loop shift more than those at the C-terminal end. This profile is reversed in the Fe^{2+} titrations. In both cases, there is a localisation of activity in the residues surrounding histidine 58. This may indicate a role for histidine, even if the histidine NH peak itself does not change much. The importance of histidine is further explored in Section 5.5.

These same analyses was performed on MmsFcc D49S mutants (Figure 5.10-5.13). For MmsFcc D49S + Fe^{3+}, the 16x titration point was used for comparison. For these analyses, it was not possible to track peak position for the following peaks due to overcrowding in the spectrum: Q6, L7, I21, K23, Q24, A26, L28, W30, K31, K37, L42, K43, K45, L46, K64, Q65, I80, K82, Q83, A85, W89, K90, K96, A99, and L101.

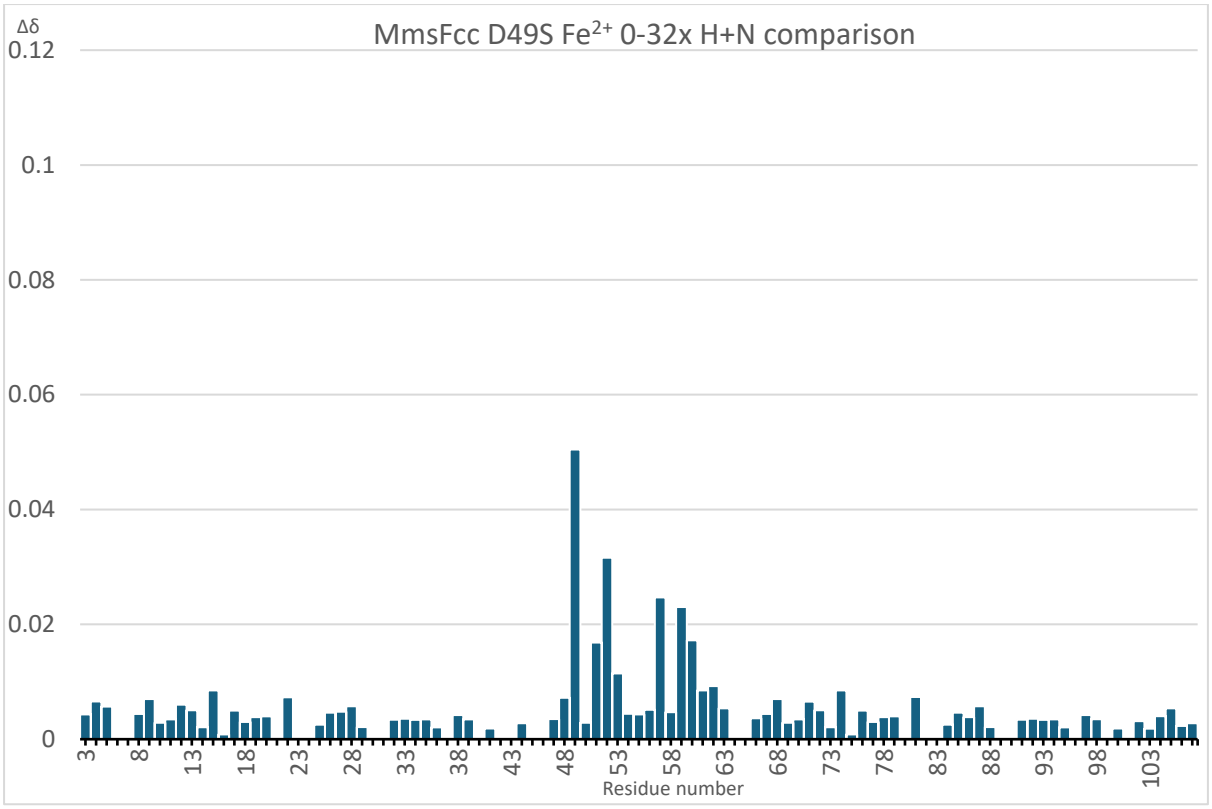


Figure 5.10. Difference in MmsFcc D49S HSQC peak position for 0x and 32x Fe^{2+} titration points. Proton and nitrogen Δδ were calculated separately. Nitrogen Δδ was then weighted by 0.14 before being added to Proton Δδ.

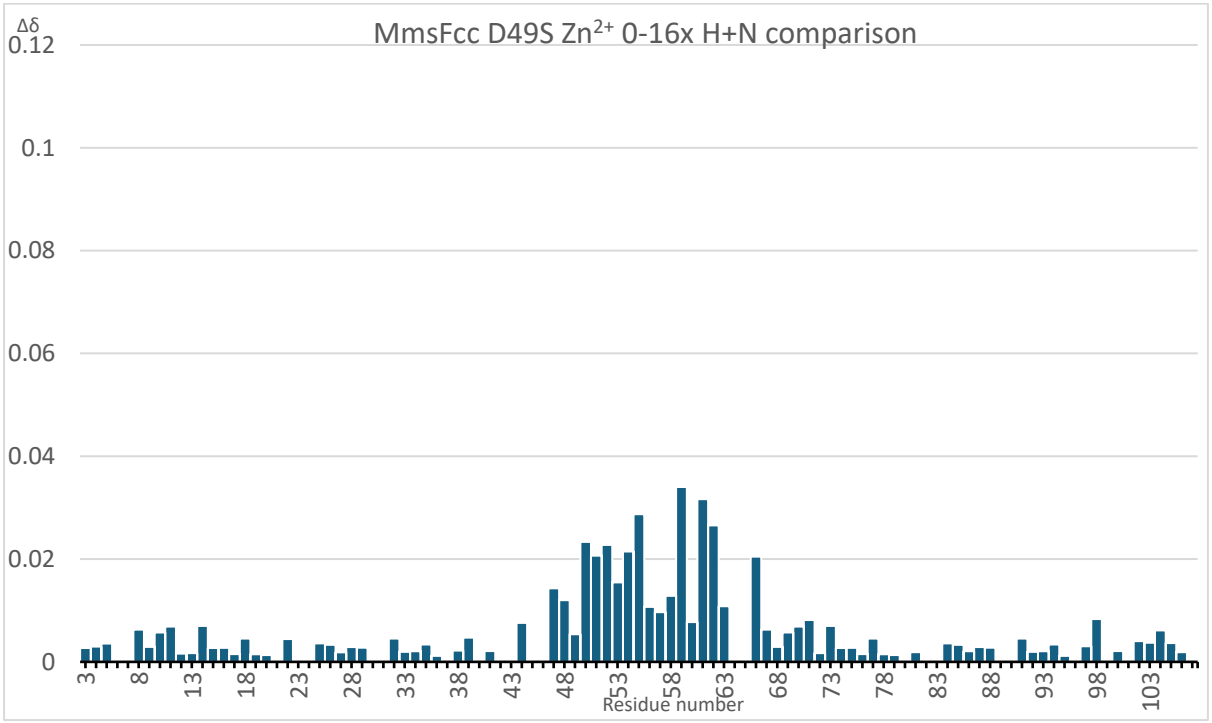


Figure 5.11. Difference in MmsFcc D49S HSQC peak position for 0x and 32x Zn^{2+} titration points. Proton and nitrogen Δδ were calculated separately. Nitrogen Δδ was then weighted by 0.14 before being added to Proton Δδ.

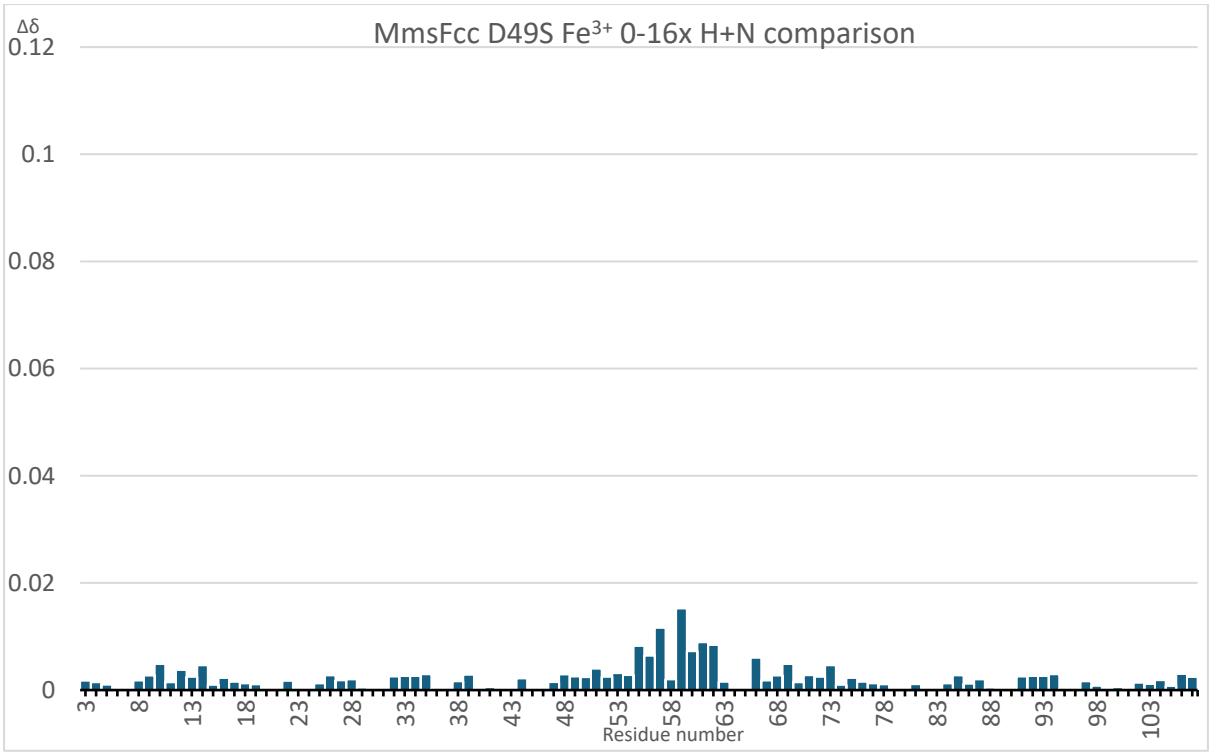


Figure 5.12. Difference in MmsFcc D49S HSQC peak position for 0x and 16x Fe^{3+} titration points. Proton and nitrogen Δδ were calculated separately. Nitrogen Δδ was then weighted by 0.14 before being added to Proton Δδ.

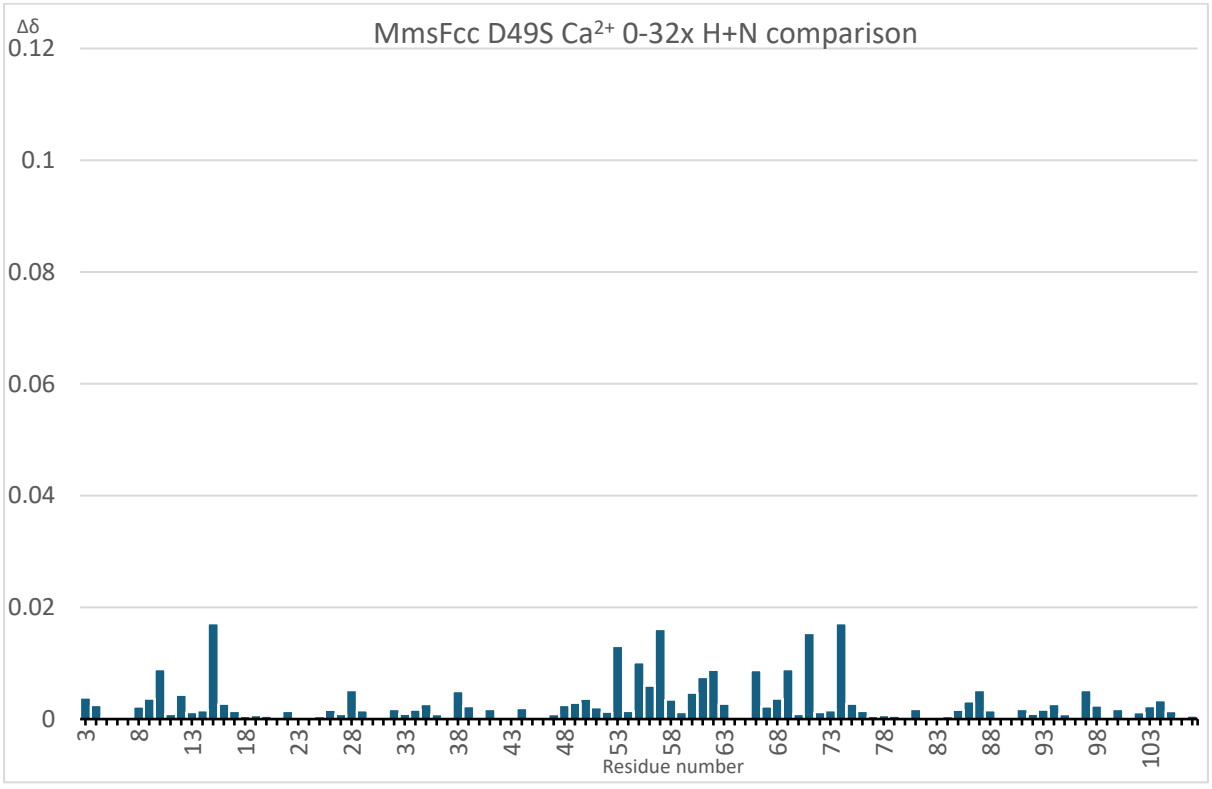


Figure 5.13. Difference in MmsFcc D49S HSQC peak position for 0x and 32x Ca^{2+} titration points. Proton and nitrogen Δδ were calculated separately. Nitrogen Δδ was then weighted by 0.14 before being added to Proton Δδ.

From these graphs, a few obvious observations can be made. Generally, the magnitude of shift change is much greater in the Wildtype than it is in D49S titrations. This is likely due to a stronger protein-metal interaction with Wildtype and supports the data discussed in Section 1.7. suggesting that MmsFcc D49S is less functional in *in vitro* chemical co-precipitations of magnetite.

In Figures 5.10 and 5.11, the N-terminal loop peaks shift more on average for Zn^{2+}, while the C-terminal peaks shift more for Fe^{2+}. Interestingly, this observation is flipped in the D49S titrations. This suggests that D49 is important for the correct binding of metal to MmsFcc, given that its substitution for a non-co-ordinating sidechain (serine) significantly alters the chemical shift pattern for the two ligands.

The difference in proton and nitrogen shifts were calculated separately, and displayed in Figures 5.14-5.17, to see if there is a trend.

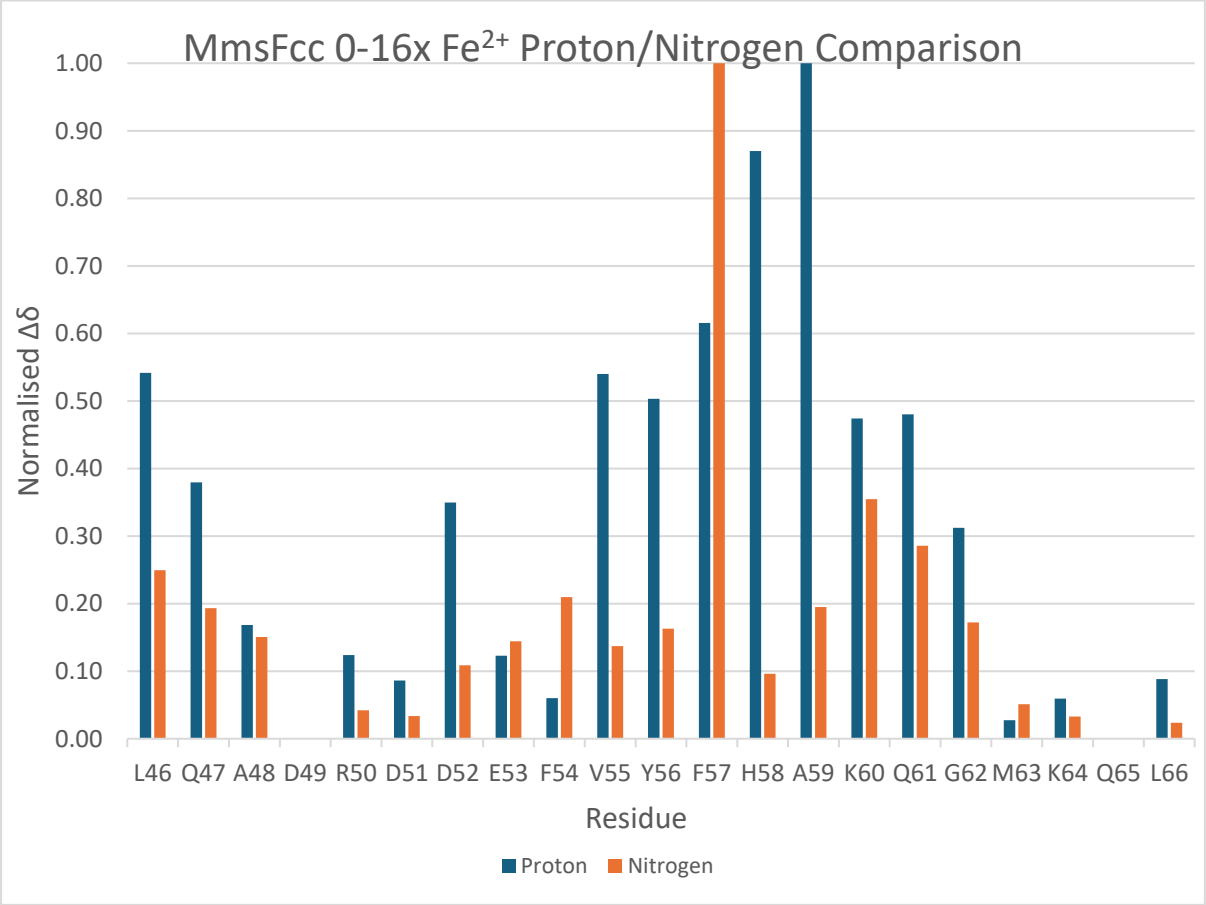


Figure 5.14. Difference in MmsFcc HSQC chemical shifts for 0x and 16x Fe^{2+} titration points. Proton and nitrogen Δδ were kept separate, then normalised to 1 by diving the dataset by largest shift difference.

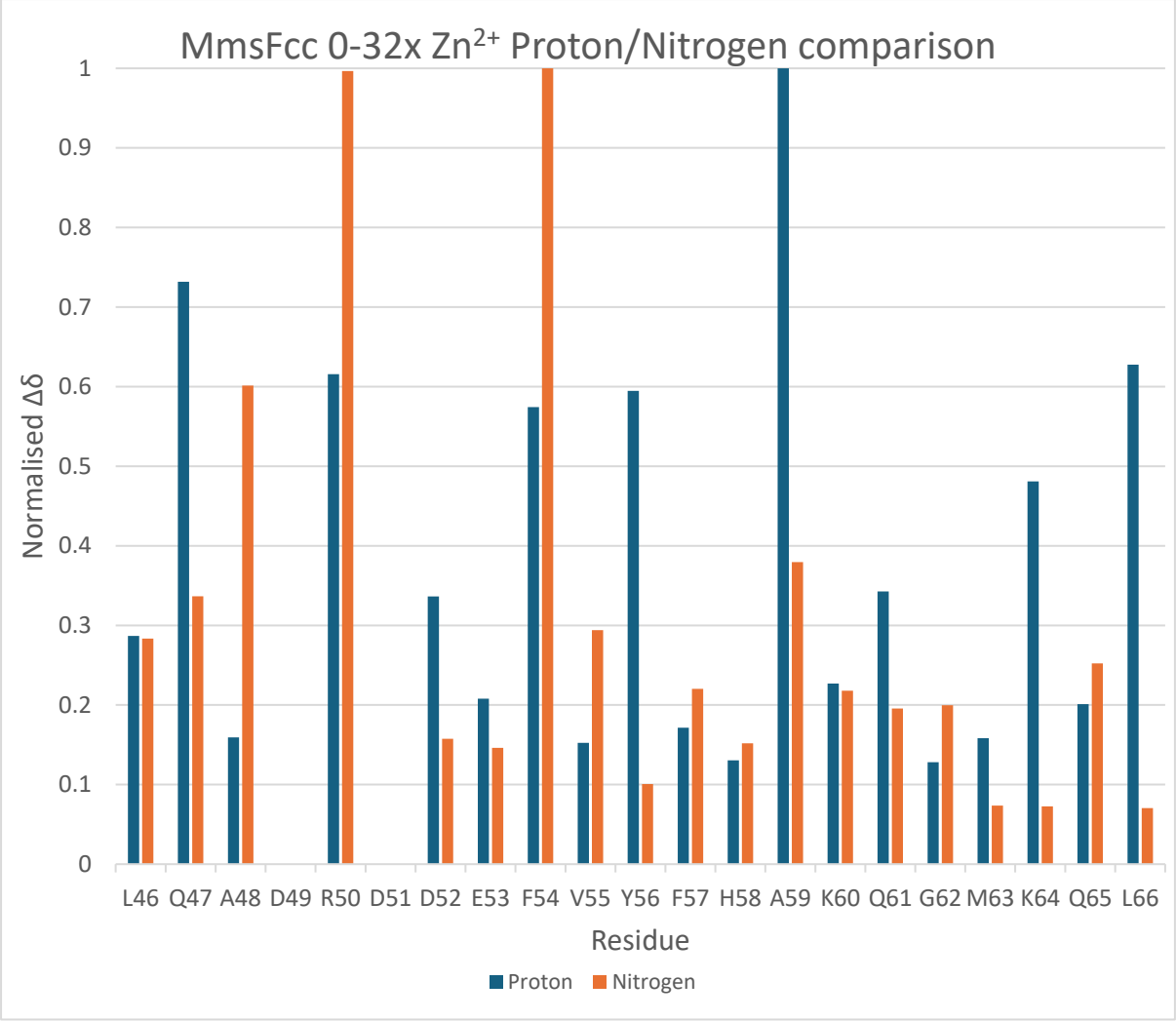


Figure 5.15. Difference in MmsFcc HSQC chemical shifts for 0x and 32x Zn^{2+} titration points. Proton and nitrogen Δδ were kept separate, then normalised to 1 by diving the dataset by largest shift difference.

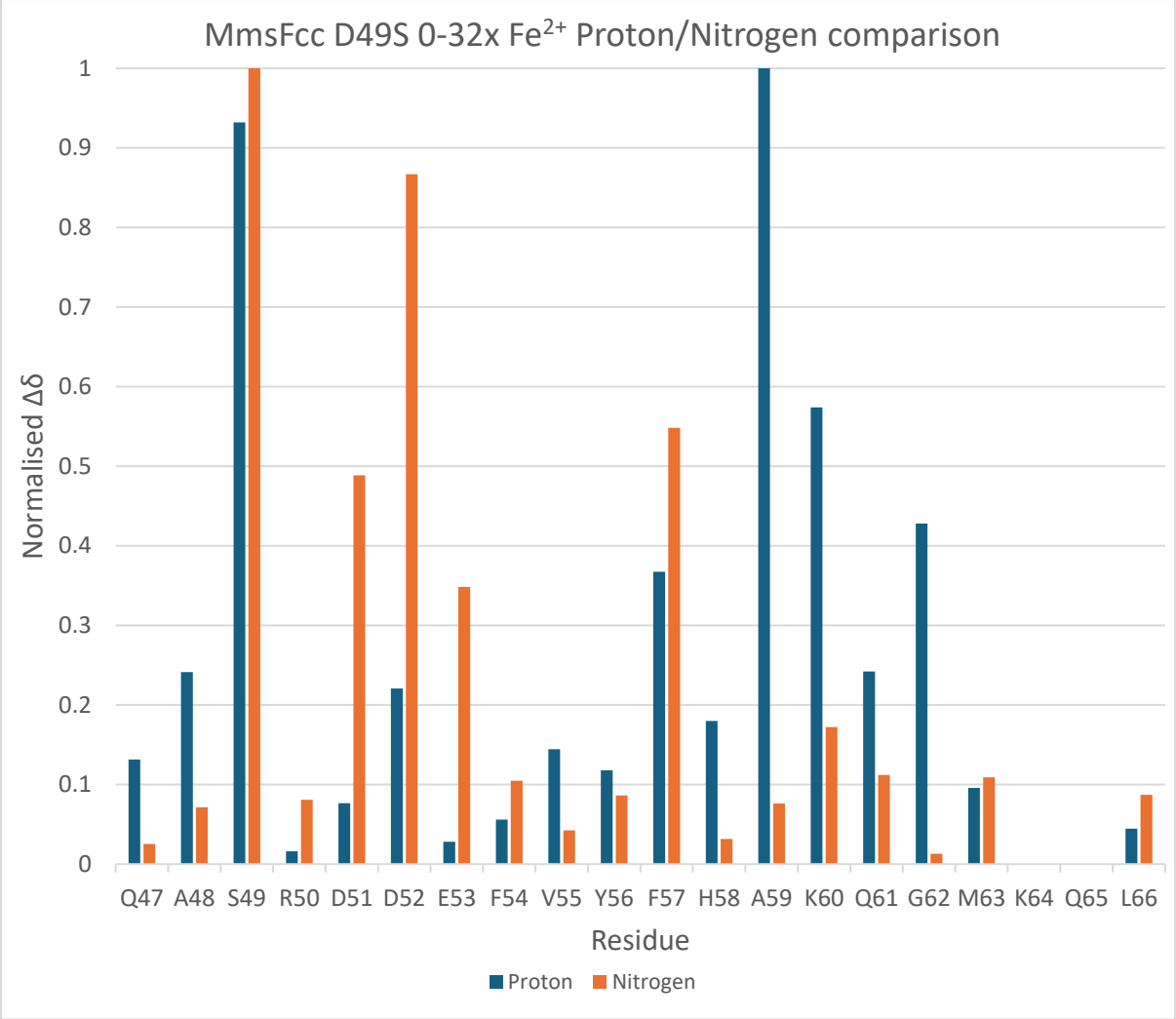


Figure 5.16. Difference in MmsFcc D49S HSQC chemical shifts for 0x and 32x Fe^{2+} titration points. Proton and nitrogen Δδ were kept separate, then normalised to 1 by diving the dataset by largest shift difference.

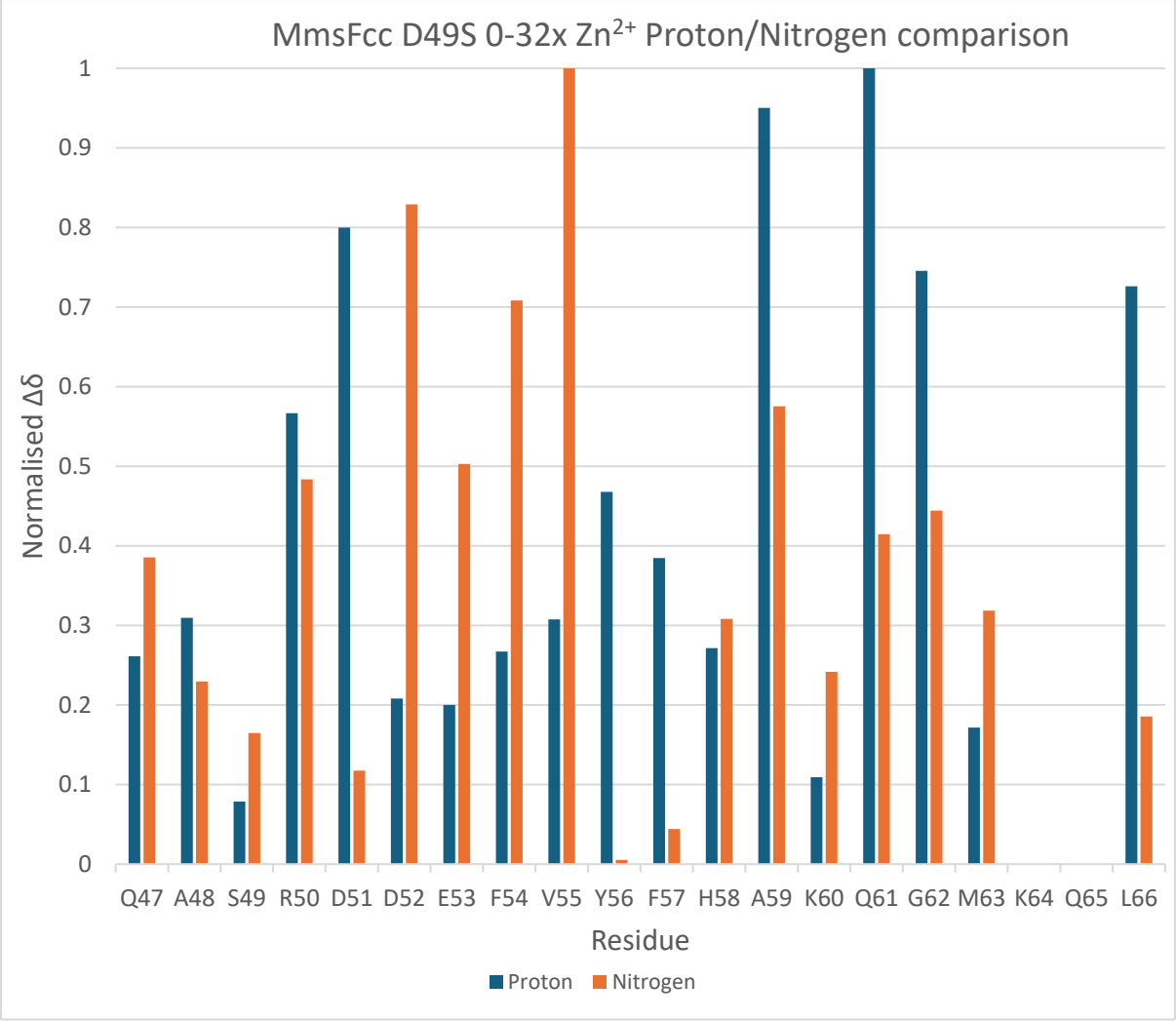


Figure 5.17. Difference in MmsFcc D49S HSQC chemical shifts for 0x and 32x Zn^{2+} titration points. Proton and nitrogen Δδ were kept separate, then normalised to 1 by diving the dataset by largest shift difference.

It is problematic to draw conclusions from this analysis since it is not possible to determine the root cause of the changes in pattern from HSQC spectra alone. This is especially true with Nitrogen shifts, which are influenced by a range of factors such as the sidechain conformation of the preceding residue and hydrogen bonding. With proton shifts, most observed change is due to changes in hydrogen bonding (M. P. Williamson, 2013a).

5.3 Determination of K_d values using least squares fitting, analysis of 300x ligand spectra

It is possible to use the chemical shift changes induced by ligand binding to calculate the dissociation constant, K_D, of the interaction. The interaction between MmsFcc and metal cations was hypothesised to be a one-site reversible binding event. This is so as the thirteen-

residue functional loop is likely too small to be able to co-ordinate more than one binding partner. In the same vein, it is likely not structured enough to trap a binding partner permanently. This binding model fits our hypothesis for MmsFcc function, where Fe^{2+} cations would need to bind transiently before subsequent incorporation into the growing crystal. For these reason, the following equation was used to fit to the observed data (Trevitt et al., 2024):

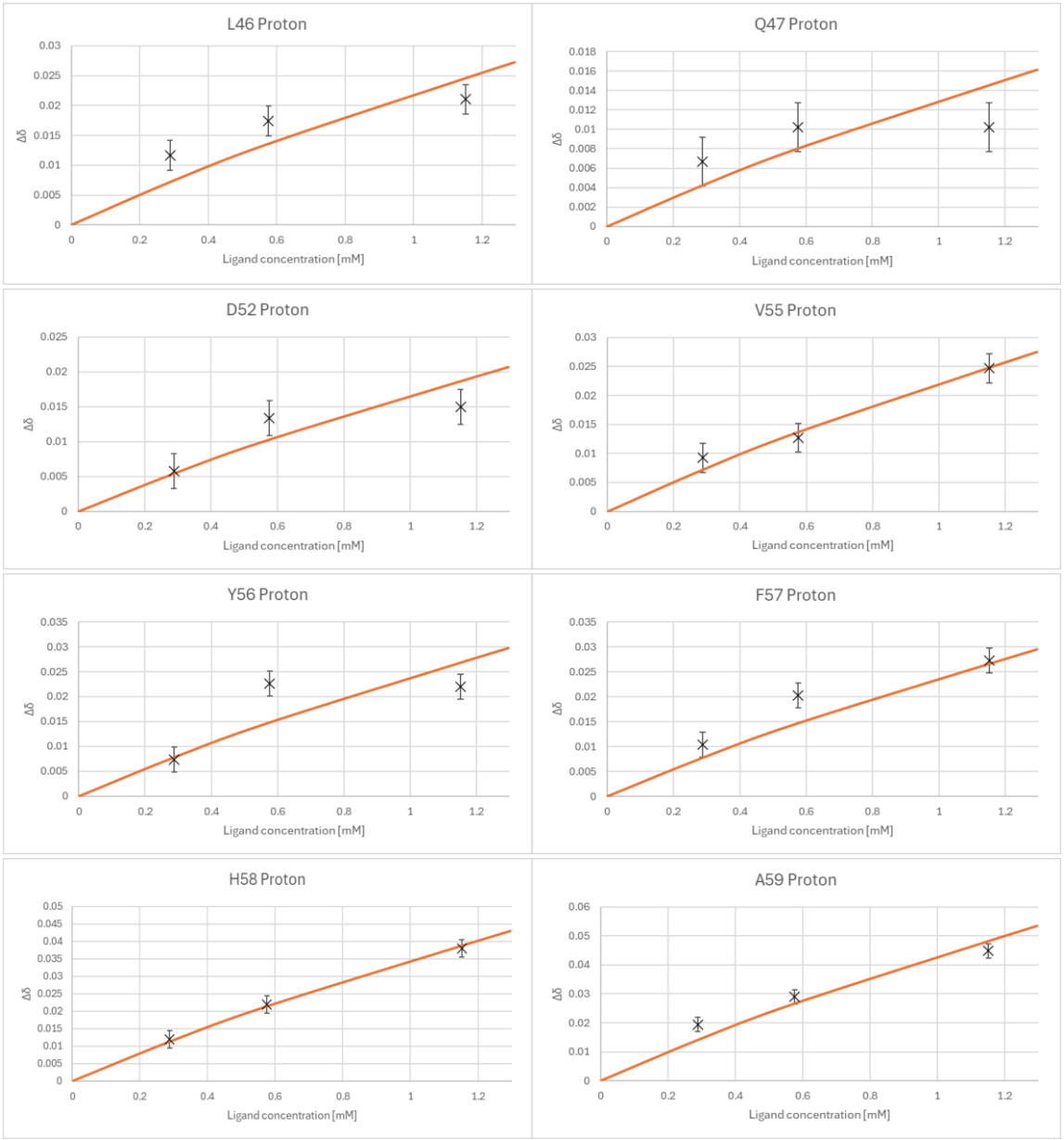
$$\Delta\delta_{obs} = \Delta\delta_{max}\{([P]+[L]+K_d)-[([P]+[L]+K_d)^2-4[P][L]]^{1/2}\}/2[P]$$

Proton and nitrogen shifts were not combined for this analysis. Determining which shifts were included in K_d calculations was achieved by calculating $\Delta\delta_{32x}$ (or δ_{16x} if there was no 32x dataset), then selecting only the shift changes greater than the mean plus one standard deviation.

Figure 5.7 shows that there are some titrating peaks present in the 300x Zn^{2+} spectrum, which could be included in K_D calculations. However, while these peaks did move in the same direction as the lower concentration titration spectra, they proved to be difficult to fit. Furthermore, there is a significant proportion of peaks in the spectrum which do not overlay properly with the free spectrum, implying that the perceived shift difference might not be simply due to the binding event. For these reasons, the 300x spectrum has not been included in K_D calculations.

Nitrogen chemical shift values were reduced by multiplication with a scale factor of 0.14, in order to equalise the magnitude of proton and nitrogen chemical shift values. Titration ligand concentrations were inputted into the one-site binding equation, along with initial guess values $\Delta\delta_{max}$ and K_D values of 1 ppm and 1 mM to produce poorly fitted predicted shift changes. For these predictions, a unique $\Delta\delta_{max}$ value was used for each set of $\Delta\delta$ (e.g. A59 H), while the same K_D was applied to every equation in a specific protein + metal combination. The squared difference between these predicted values and the observed shift changes was calculated. To fit the predicted values to the observed, the excel solver add-on was used to minimise the sum of the squared differences by changing $\Delta\delta_{max}$ and K_D values. These constants can be used to predict a binding curve, which can be plotted against the observed datapoints. An estimation of error was made by considering the inherent error in measuring chemical shifts. This was defined as 0.005 ppm for proton and 0.036 ppm for

nitrogen. The nitrogen error was derived from dividing 0.005 ppm by 0.14, to match the larger magnitude of the nitrogen chemical shift scale. These errors were then normalised to 1 alongside the data, then added to the plots. These plots for MmsFcc + Fe^{2+} are shown in Figures 5.18-5.19.



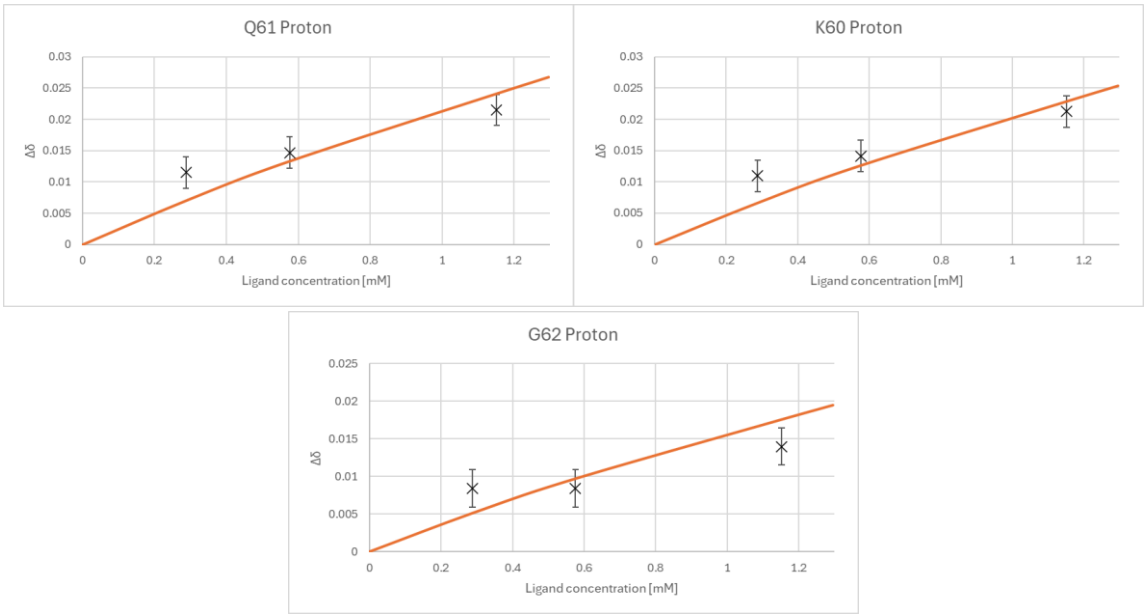


Figure 5.18. Binding curve fitted to proton chemical shift changes (using least square fit method) for MmsFcc + Fe^{2+} titrations. Observed data represented by crosses. Orange line represented fitted binding curve.

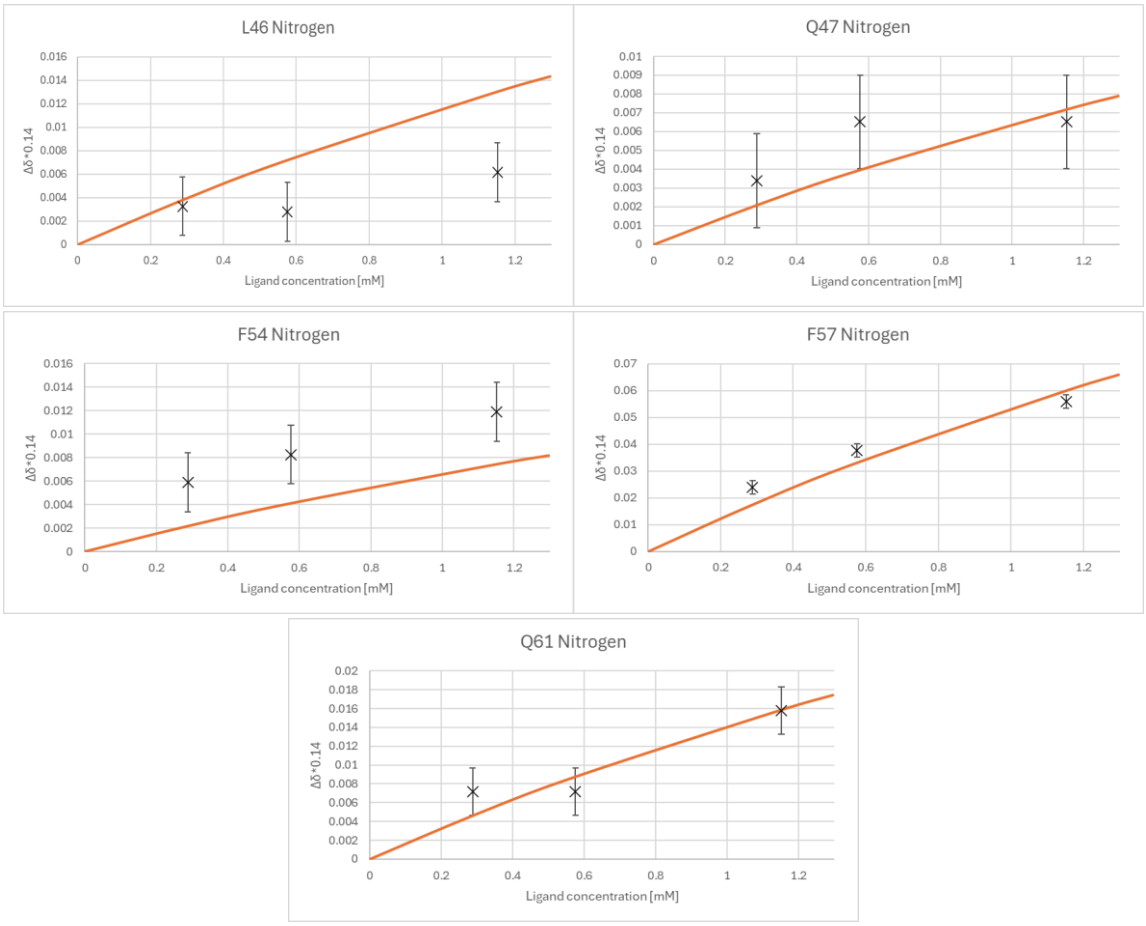


Figure 5.19. Binding curve fitted to nitrogen chemical shift changes (using least square fit method) for MmsFcc + Fe^{2+} titrations. Observed data represented by crosses. Orange line represented fitted binding curve.

The shape of the fitted curves indicate weak binding, which is corroborated by the K_d values seen in Table 5.4. There is also a degree of uncertainty, which cannot only be attributed to measurement errors. This is typical of weak binding, meaning that some caution should be applied when interpreting these calculated K_d values. Binding curve plots for MmsFcc + Zn^{2+}, MmsFcc D49S and Zn^{2+} and MmsFcc D49S and Fe^{2+} can be found in the appendix.

A table of the calculated K_D values is provided below, error estimates were calculated by inputting the values defined by the error bars into the one-site binding equation:

Cation	MmsFcc K_D [mM]	MmsFcc D49S K_D [mM]
Fe^{2+}	4.8± 2.6	5.6±2.5
Zn^{2+}	19.6±19	56.3±28.5
Fe^{3+}	Could not calculate	Could not calculate
Ca^{2+}	Could not calculate	Could not calculate

Table 5.4. K_d values and error estimates for MmsFcc and MmsFcc D49S interactions with the four trialled metals.

Larger K_D values for MmsFcc D49S indicate less affinity for metals, again further supporting the hypothesis that its reduced function in chemical co-precipitation reactions is due to weaker affinity for iron. From these values, it is apparent that MmsFcc and MmsFcc D49S both have a stronger affinity for Fe^{2+} than Zn^{2+} However, due to the high degree of uncertainty, and the weakness of these interactions, this difference is likely negligible. For Fe^{3+} and Ca^{2+}, any interaction that did occur produced chemical shift changes too small to possibly determine a K_d value. All these K_D values are in the millimolar range, which represents an affinity on the lower end of range typically seen in protein-metal interactions (Wang et al., 2009). The biological context of the native protein may explain this observation. The magnetosome employs many proteins to uptake and sequester iron from the surrounding environment, creating a concentrated pool of iron inside the magnetosome lumen (Uebe & Schüler, 2016). In this context, a millimolar affinity for the ligand would likely be sufficient due to saturation of the ligand. Furthermore, since K_D is largely defined by the K_{off} rate, a more transient binding event allowed by a weaker affinity may be key to the proper crystallization of the growing nanoparticle.

It should be mentioned that, while the affinity between protein and ligand is contributing to the K_D values observed, this value is only an approximation of all events occurring in relation to the binding. The possibility of further events such as conformational changes occurring are discussed in detail in Section 5.6.

5.4 Fe^{2+} binding is greatly reduced when the histidine sidechain is deprotonated

It was hypothesized that the histidine sidechain ring would play an important role in metal binding, evidenced by the widely observed metal binding properties observed in the literature. The calculated structure of MmsFcc also suggests a role for histidine, since its sidechain position is strongly conserved across the 20 structures, and it also sits in a position where it could potentially form a metal-binding site (Chapter 4.6). Titration data supports this hypothesis, with the amide groups of residues surrounding the histidine experiencing large chemical shift changes upon addition of ligand. Since histidine primarily interacts with metals through its side-chain ring, the H58 HSQC peak itself is not expected to experience much change in chemical shift, due to the through-bond effect of binding to the sidechain being mostly attenuated before it reaches the backbone amide.

To investigate further, it was first important to understand the protonation state of the histidine sidechain in our experiments. In the context of metal binding, the deprotonated form of histidine is considered 'active' and able to co-ordinate metals. Monomeric histidine has a sidechain pK of ~6.03, meaning that a population of histidine would have ~50% of the sidechains protonated at pH 6.03. This is close to our standard pH of 6.5, meaning that titrations were being performed in the pH range where the histidine sidechain will experience the largest fluctuations in protonation state. To better understand the protonation state of this residue, a series of pH titrations were performed. Discontinuous samples of MmsFcc were adjusted to the following pH values: 5.5, 6.0, 6.2, 6.4, 6.6, 6.8, 7.0, 7.5 (Protocol 2.17). ^{1}H-^{15}N HSQC spectra were recorded for each pH, and then overlaid in Figure 5.20.

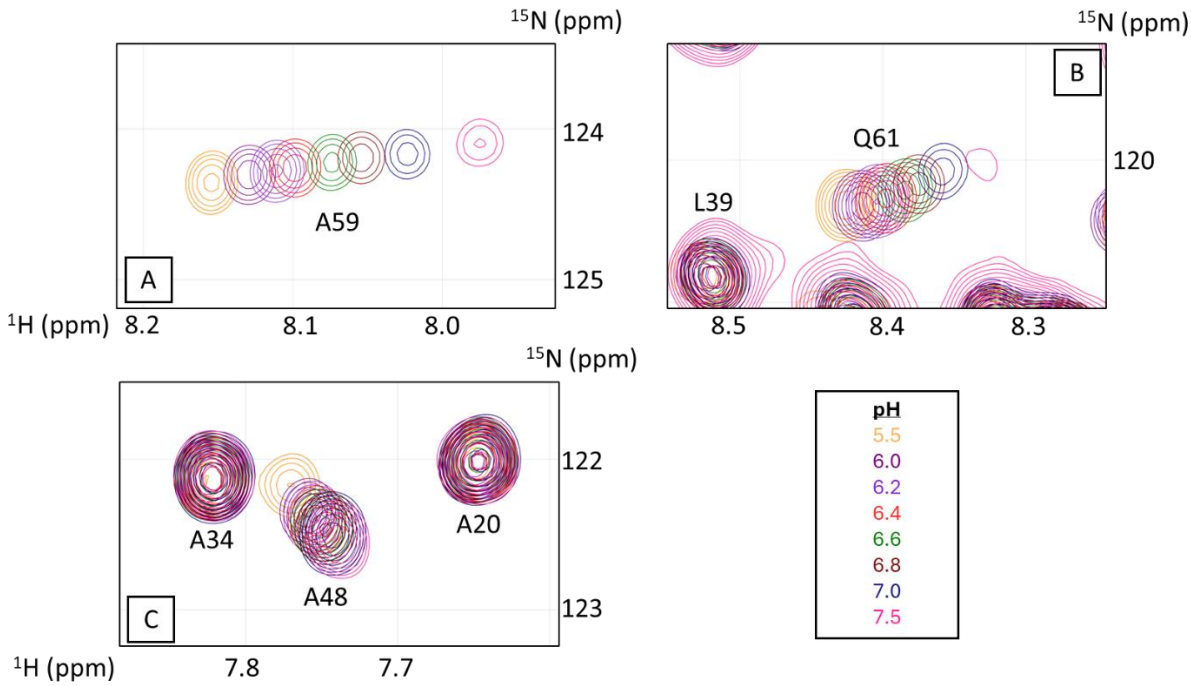


Figure 5.20. HSQC spectra of MmsFcc at different pH, ranging from pH 5.5-7.5. Different pH spectra are colour coded.

In Figure 5.20, it is noticeable that residues either in or directly adjacent to the metal-binding loop experience chemical shift changes, while coiled-coil resides remain stationary. Since the range of pH used only affects histidine sidechains, all observed shift changes are the result of changes to histidine protonation. Therefore, these changes can be used to calculate a pK. Proton δ values were recorded for the residues F57, H58, A59, Q61. For each residue, the following equation was used to fit proton δ values:

$$\delta H = (a + b*10^{(pH-pK)})/(1+10^{(pH-pK)})$$

Where a is the δ at low pH, and b is the δ at high pH. The square of the difference between each pH titration and its equivalent prediction was then calculated. Like in the K_D calculations, the sum of these square differences was calculated, then minimised using the excel solver addon by allowing the programme to alter values a, b, and pK. This produced pK predictions for each residue, which are presented in Table 5.5

Residue	Predicted pK
F57	6.71
H58	6.81
A59	6.79
Q61	6.83

Table 5.5. Predicted pK values of H58 using proton chemical shift values from different residues.

These values were averaged to a pK of 6.79. The pK can then be used to calculate the approximate percentage of histidines which are protonated at different pH conditions, which is done using the following equation:

$$pH=pK+\log([A^-]/[HA])$$

Which can be rearranged to give the ratio of deprotonated/protonated as such:

$$10^{(pH-pK)}=[A^-]/[HA]$$

Using this equation, at pH 6.5 approximately 49.7% of histidines are deprotonated. Meaning that if our hypothesis is true, then we are observing approximately 50% activity at pH 6.5. To test the metal binding capacity of MmsFcc at lower pH, a series of Fe^{2+} titrations were performed at pH 5.5. At this pH, approximately 5.1% of histidines are deprotonated. A comparison of Δδ for 0x and 16x Fe^{2+} titration points for MmsFcc at pH 6.5 and 5.5 is shown in Figure 5.21. For this comparison, the following equation was used:

$$\Delta\delta=\sqrt{[(\text{Titration1H}\delta-\text{Titration2H}\delta)^2+0.14*(\text{Titration1N}\delta-\text{Titration2N}\delta)^2]}$$

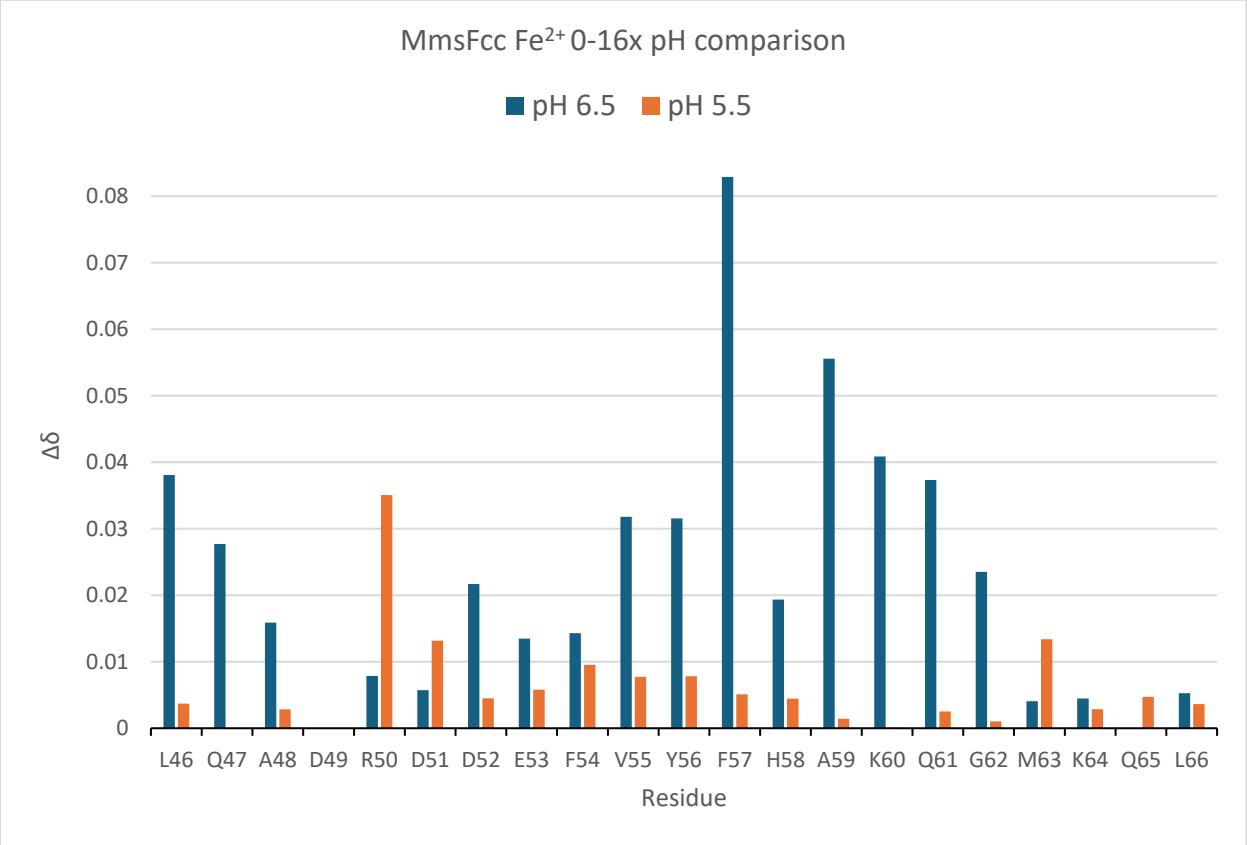


Figure 5.21. Comparison of MmsFcc-Fe^{2+} interaction at pH 5.5 and 6.5. Blue bars represent the difference in peak position between 16x and 0x spectra at pH 6.5. Orange bars represent the difference in peak position between 16x and 0x spectra at pH 5.5.

From this figure, it can be seen that MmsFcc + Fe^{2+} interaction is drastically reduced at pH 5.5, which strongly suggests that a deprotonated histidine is essential for binding to Fe^{2+}. The perceived increase in shift for residues R50, D51, and M63 is probably attributable to noise. In the next section, some of the concepts explored in this chapter are brought together and discussed.

5.5 Chapter discussion

The main question that arises from this chapter is the cause of the line broadening observed in Figures 5.2-5.3. One potential cause was the effect of intermediate exchange; however, this can be disproven with some calculations. Since intermediate exchange is defined as when $K_{off} \approx \Delta\delta$, $\Delta\delta$ can tell us what rate of exchange would have to be for this to happen. The largest observed shift in the wildtype + Fe^{2+} titrations for proton is 0.0448 ppm for residue A59. On an 800 MHz spectrometer, this equates to 35.84 Hz. This would mean that, for the effects of (East et al., 2020)intermediate exchange to be seen, then MmsFcc would be releasing ligand at a rate of ~35 times per second. This would imply much tighter binding than what both the K_D calculations and the proposed *in vivo* function of MmsF each suggest (discussed in Section 5.3). It is also important to consider the proportion of bound protein when considering this line broadening. This is especially relevant to the 300x concentration spectra for wildtype + Fe^{2+}/Zn^{2+} (Figures 5.4,-5.5), both of which showed the near total loss of all loop signals. Using the K_D calculations, it is possible to determine the estimated percentage of bound protein at this ligand concentration. For Fe^{2+}, MmsFcc is estimated to be 96% bound. For Zn^{2+}, MmsFcc is estimated to be 88% bound. In the case of intermediate exchange, we would expect to see a peak representing the bound population at this degree of ligand saturation, meaning that we can confidently discredit this proposition (East et al., 2020).

Line broadening is commonly observed in titrations and is often not the result of off/on exchange (M. P. Williamson, 2013b). Often, conformational changes related to ligand binding, such as the structural rearrangement of the protein after binding or ligand induced dimerisation, occur at rates which result in broadening. Assuming a single binding site, a

conformational change would explain why almost every loop residue experiences chemical shift change. In this case, the chemical shift difference would stem from changes to the protein structure encompassing each residue, not through direct interaction with the ligand.

It is difficult to determine exactly what conformation change is occurring from this titration from this analysis alone. More information about the system could be found from performing line-shape analysis on the broadening peaks. It is possible to derive kinetic and interaction information from line shape, though analysis with programmes such as TITAN (Waudby et al., 2016). This analysis was not performed due to time constraints.

It is difficult to predict the binding event model based on the literature and titration data, since there is enough variation in co-ordinate complex shapes and co-ordinate partner preferences to explain multiple potential models. Instead, conclusive evidence would be derived from a bound structure. This was the intention of this study, however the line broadening observed upon saturation with ligand prompted us to not pursue this avenue, given that any bound spectrum would likely experience line broadening in the carbon dimensions. It is possible that this could have been remedied with clever experimental design, but this was unfortunately out of the timescale of this project. Nonetheless, many useful and novel observations can be made from analysing the structural and binding data of MmsFcc, which is discussed further in Chapter 6.

Chapter six: Conclusions and future work

6.1 Conclusions from Chapter 3

As discussed in detail in Chapter 3, the scope of this project was limited by difficulties when attempting to purify MmsFcc. The very properties that make MmsFcc such an interesting protein to research (artificially designed coiled-coil scaffold, metal binding activity), were likely to have directly contributed to this. Although anecdotal, some observations of MmsFcc based on its behaviour during the purification process can be made. Firstly, the formation of inclusion bodies can be the result of overexpression but is more commonly observed in proteins with toxic properties. It can also be due to the protein aggregation due to misfolding (Bhatwa et al., 2021). MmsFcc is likely toxic to *E. coli*, through its demonstrated interaction with free iron and zinc (Chapter 5). Evidence for this exists with the higher yields of D49S mutants, which has been demonstrated experimentally to have a weaker affinity for free metals (Section 5.2). We also know that MmsFcc tends to aggregate, demonstrated by the large amount of protein recoverable from precipitate (Protocol 2.12). In any case, both the presence of misfolding protein and metal binding would explain the slow growths experienced during this project.

Another issue encountered during this project was the affinity for Nickel columns that MmsFcc demonstrated (Figure 3.9). This was clearly mediated by the structured loop, since removal of loop structure via denaturing removed this affinity (Figure 3.11). Considering the importance of deprotonated histidine 58 for the binding of iron (Section 5.4), it is possible that MmsFcc affinity for the nickel matrix could have been greatly reduced through the lowering of pH. pH titrations had already demonstrated that MmsFcc was stable at pH 5.5, and the effect of lower pH's (and thus less 'active' histidine) could easily be assessed with HSQC experiments. While we consider other options for tags, each ultimately has its drawbacks. This affinity for positive charges is likely what made the protein 'sticky' (i.e. experiencing significant loss on surfaces such as column matrices, tubes, and membranes).

Difficulties in cleaving the HisTag were not foreseen, since in theory the recognition site should have been readily available for TEV binding. Typically, poor TEV cleavages are the result of buried cleavage sites (Raran-Kurussi et al., 2017b). Since TEV has been used extensively to cleave HisTags, it is unlikely that metal binding by the HisTag is causing this

effect. Instead, the presence of the metal binding loop of MmsFcc must be driving this loss of efficiency, likely through tag-loop interactions mediated by metal ions.

6.2 Conclusions from Chapter 4

Solving the structure of MmsFcc in its free state provided some insight into how it may function. Firstly, the sidechains of many residues (D49-E53, K60-Q61) adopt a range of conformations, although this is to be expected for soluble residues in a loop that is surrounded by water. The low RMS deviation observed for backbone atoms (0.43 Å for residues 49-61) suggests that the backbone of the loop is structured. Importantly, histidine 58 is amongst the residues where the sidechain position was more conserved (F54-A59). The prevalence of histidine metal binding in literature, and the evidence presented in Section 5.4, strongly suggest a key role in MmsFcc mediated biomineralisation for histidine. The conservation of histidine 58 sidechain orientation suggests that its correct positioning in the loop structure is important for MmsFcc function.

Probably the largest unanswered question from the structures obtained is role of the three aspartate residues (D49, D51, and D52). The D49S substitution has some effect on MmsFcc function, evidenced by the larger K_d values observed in D49S titrations when compared to MmsFcc wildtype. The reduction in size and uniformity of D49S mediated nanoparticles (Section 1.4) also supports this statement. As mentioned previously, the molecular dynamics programme used to determine MmsFcc structures includes values for electrostatic charges across the molecule. This would likely explain why there is no obvious structural role for these three aspartate residues, given that their close incorporation would be energetically unfavourable. Since MmsFcc mediated biomineralisation is likely a complex multi-stage process, it is entirely possible that a change in protein structure at a stage in the process (perhaps allowed by the inclusion of a charge balancing cation) results in the incorporation of these residues into a more conventional metal binding site. This is, however, quite challenging to investigate given the likely transient nature of these stages. Further structural investigation of MmsFcc could be possible through the use of Cryo-EM. By fusing MmsFcc to a large soluble tag such as GB1 or MBP, it would likely be possible to determine MmsFcc loop structure through Cryo-EM images of the fused protein construct. Although likely tricky, this construct could potentially be mixed with nanoparticles from different stages of the

biomineralisation process, in order to determine how the structure of MmsFcc may change over the course of Nanoparticle synthesis. This approach was not attempted in NMR spectroscopy, due to the predicted challenges of incorporating large magnetic bodies into the NMR sample.

Some confusing observations did arise from the ligand free structure of the MmsFcc loop. Primarily, the odd position of phenylalanine 57, whose hydrophobic aromatic ring was presented to the solvent in all 20 structures, is not typical of phenylalanine. This residue is clearly important for the correct biomineralisation of magnetite nanoparticles, as evidenced by the loss of function mutant F57Q (Figure 1.7) and the non-functional homologue MmxF. Unfortunately, due to difficulties getting a plasmid that would express, we were not able to perform any experiments on either of these mutants (F57Q, F54Y/F57Q). It goes without saying that by solving these structures (even in the free state) and assessing their metal binding capacity, much more could have been elucidated about the roles of these apparently critical residues.

In the introduction, the idea that AlphaFold may struggle with artificially designed proteins was discussed (Section 1.6). This is counterintuitive when considering that the AlphaFold structure was used to derive distance restraints for the coiled-coil backbone (Section 4.6). The coiled-coil predicted by AlphaFold looked very reasonable, which is probably due to the presence of common features seen in natural proteins (i.e. the heptad sequence repeat typical of coiled-coils, the leucine zipper). The AlphaFold structure does vary from our experimentally derived structure in the loop: in the AlphaFold prediction the loop is helical with 2 turns to bridge the ends of the coiled-coils (Figure 6.1).

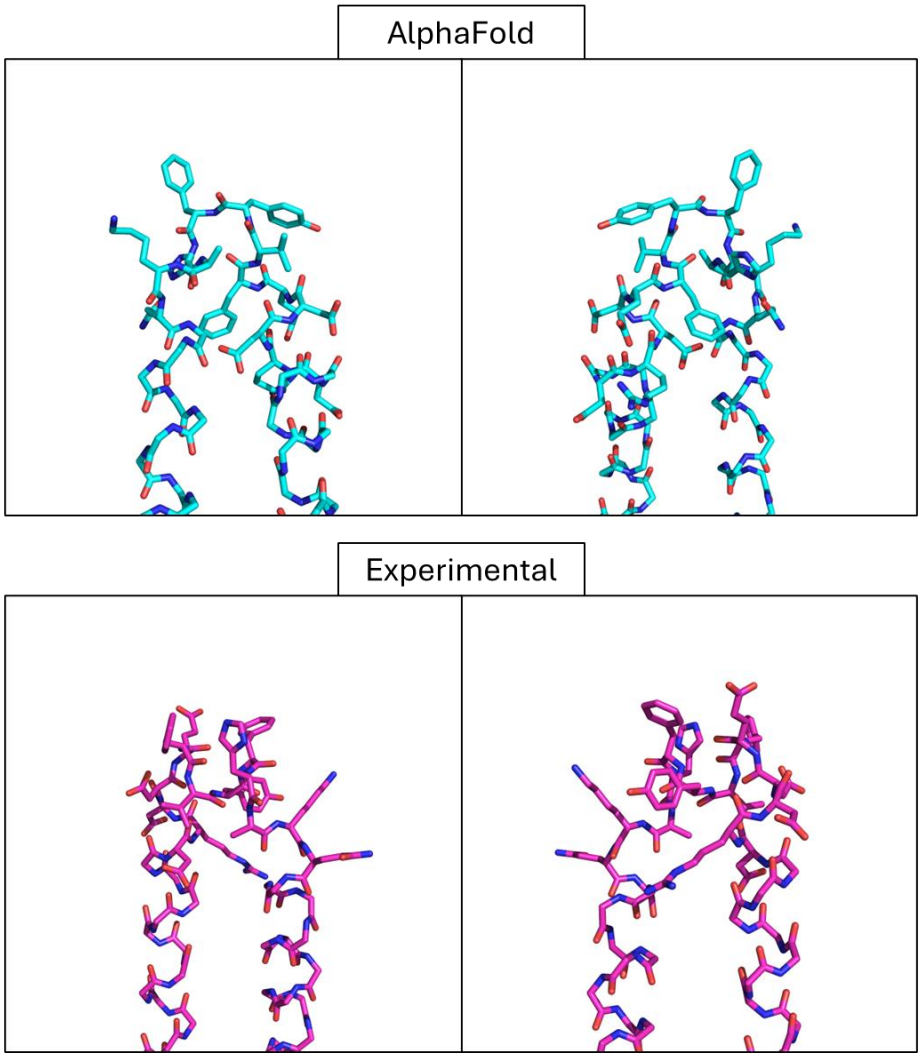


Figure 6.1. Difference in loop structures for experimental structure (magenta) and AlphaFold (teal).

From this figure, it is visually clear that there is little structural similarity. This is confirmed by a quick alignment of the loop residues in PyMol, resulting in a RMSD of 5.54 angstroms from the two structures displayed in Figure 6.1. This result is backed up by the lack of experimental NOEs present in the AlphaFold structure. The following experiment NOEs were investigated and found to not be present in the AlphaFold structure:

Proton 1	Proton 2	Upper bound (Å)	Lower bound (Å)
D51 Hβ	Q61 Hβ	5.0	1.8
F54 Hε	D51 Hα	3.8	1.8
F54 Hε	Q61 Hβ	3.8	1.8
Y56 Hδ	A59 Hβ	5.3	1.8
D51 HN	Q47 Hβ	3.1	1.8

Table 6.1. Five experimental NOEs which were not identifiable in Figure 6.1 AlphaFold structure.

-

It is likely that much more could have been learnt about MmsFcc if a bound structure was solved. However, this was not pursued during this project due to the line broadening observed in Section 5.1.This could be achieved through total saturation of the protein with ligand, resulting in a fully bound spectrum. However, the concentration of iron required to achieve this would likely be problematic due to the paramagnetic broadening. I t could certainly be possible with Zinc, but this would be harder to draw comparisons to the *in vivo* function of MmsF.

6.3 Conclusions from Chapter 5

It is apparent that some form of conformational change event is the cause of line broadening observed upon interaction with the ligands Fe^{2+} and Zn^{2+}, since the possibility of protein-ligand exchange broadening was discounted in Section 5.6. This leaves a rather open-ended question for this thesis, being how exactly the loop of MmsFcc changes structure. We suggested in chapter 5 that this conformational rearrangement could occur through an initial ligand interaction to mediate ligand binding, or it could happen after ligand binding as part of the mechanism of MmsFcc function. Another possibility that could be affecting the spectrum is dimerisation with other MmsFcc molecules; the organization of MmsF into proteinosomes has been discussed as an indicator for a role in higher order structures in the correct function of MmsF (Section 1.4). NMR can detect dimerisation events which have millimolar affinity, through measurement of T1 and T2 relaxation it is possible to calculate the rotational correlation time, which is inversely related to protein size (Aramini et al., 2011). Other potential experimental methods include SAXS and CD, although these are less sensitive for weaker binding events.

The significant reduction in binding between MmsFcc and Fe^{3+} when compared to Fe^{2+} strongly suggests that MmsFcc function is related to its preferential binding of ferrous iron, which supports the hypothesis that part of MmsFcc function is mediated by this interaction. As discussed in Section 1.9, this could be actioned by MmsFcc improving the incorporation of more soluble Fe^{2+} into the growing magnetite crystal, thus maintaining the correct ratio of ferrous/ferric iron required for proper magnetite formation. If Magnetite formation requires the presence of Fe^{2+} at the site of nucleation, then MmsFcc could also improve outcomes of

co-precipitation by incorporating Fe^{2+} into the nucleation process. It should be said that pH would likely greatly affect the nature of these interactions, already evidenced by the significant reduction of ferrous iron binding by MmsFcc when pH is reduced from 6.5 to 5.5 (Section 5.4). A more thorough investigation into the effect of pH on MmsFcc interactions would likely provide valuable insight into how MmsFcc functions.

Other methods were discussed for calculating K_D, especially Isothermal calorimetry (ITC). ITC was an attractive option since it is possible to fit two and three-site binding models to the data produced, which would allow us to determine whether a singular or multiple binding event was occurring. ITC could have also potentially identified whether MmsFcc was forming higher order structures (Brautigam, 2015). However, ITC was not pursued since our NMR data suggested weak binding, which was out of the scope of ITC detection at the protein concentrations we were able to produce. It is possible that the actual binding affinity is tighter, and that the K_D we observed is from the slower conformational change. So, for this reason, preliminary trials of ITC would still be a worthwhile endeavour.

Analysis of MmsFcc data is still ongoing, in preparation for submission to journals. For this, we have collaborated with Dr David Cooke of The University of Huddersfield, to produce computer simulations of MmsFcc binding to iron from molecular dynamics. This work is still in progress. We also intend to analyse the line shapes of broadening peaks via TITAN software, with this it should be possible to determine key dynamic parameters such as the rate of conformational change (Waudby et al., 2016). Again, this analysis is ongoing and not ready in time for this thesis.

As demonstrated in Section 1.7, one advantage of the coiled-coil system is how adaptable it is to new sequences. Dr Somner was successful in producing 7 expressible mutant MmsF (Gernert et al., 1995)sequences, while Rawlings et al demonstrated that other loop sequences could be attached to the coiled-coil backbone (Rawlings et al., 2019b). The mutants that Dr Somner designed were derived from the non-functional homologues of MmsF, since this was the only information available suggesting which residues were key. From this thesis, it is possible to identify new key residues that could be mutated. Our structure (Section 4.6) suggests that the glutamate 53 sidechain is key for binding, through its proximity to the already established important histidine 58. Therefore, both make for promising candidates for substitution with inert residues (such as alanine).

6.4 Concluding remarks

Although this project did not achieve many of the goals it set out to achieve, it has still been a promising demonstration of the potential for the coiled-coil system in studying the functional loops of proteins. In this thesis, we presented the experimentally determined structure of a loop using an artificially designed protein, on which there exists very little literature. One aspect of this project that was given less attention was the behaviour of the designed antiparallel coiled-coil scaffold. Both of the research fields involved (the design of artificial protein sequences, coiled-coil motifs) are extensive in their own right, and there would be many interesting things to discern from further experimentation in the coiled-coil scaffold. This is especially the case with the antiparallel orientation of this protein, which is much less commonly observed than the parallel conformation (Gernert et al., 1995).

It should be mentioned that, as is typical in PhD projects, the majority of data reported in this thesis was collected and analysed in the final year of experimental work. With expression and purification parameters established, it is likely that continuation of this work would yield more fruitful results and help us to better understand how Mms proteins work to mediate nanoparticle biomineralisation. The work on MmsFcc D49S suggests that MmsFcc mutants express and purify similarly to wildtype (or in the case of D49S, even better). This suggests that further work on existing mutants (F57Q, F54Y/F57Q) or on proposed mutants (E53A, H58A), would not require extensive re-optimisation of the purification protocol.

Bibliography

Amemiya, Y., Arakaki, A., Staniland, S. S., Tanaka, T., & Matsunaga, T. (2007). Controlled formation of magnetite crystal by partial oxidation of ferrous hydroxide in the presence of recombinant magnetotactic bacterial protein Mms6. *Biomaterials*, *28*(35), 5381–5389. https://doi.org/10.1016/j.biomaterials.2007.07.051

Ansari, S., de Wildt, B. W. M., Vis, M. A. M., de Korte, C. E., Ito, K., Hofmann, S., & Yuana, Y. (2021). Matrix Vesicles: Role in Bone Mineralization and Potential Use as Therapeutics. *Pharmaceuticals*, *14*(4). https://doi.org/10.3390/ph14040289

Arakaki, A., Webb, J., & Matsunaga, T. (2003). A novel protein tightly bound to bacterial magnetic particles in Magnetospirillum magneticum strain AMB-1. *Journal of Biological Chemistry*, *278*(10), 8745–8750. https://doi.org/10.1074/jbc.M211729200

Arakaki, A., Yamagishi, A., Fukuyo, A., Tanaka, M., & Matsunaga, T. (2014). Co-ordinated functions of Mms proteins define the surface structure of cubo-octahedral magnetite crystals in magnetotactic bacteria. *Molecular Microbiology*, *93*(3), 554–567. https://doi.org/10.1111/mmi.12683

Aramini, J. M., Ma, L.-C., Zhou, L., Schauder, C. M., Hamilton, K., Amer, B. R., Mack, T. R., Lee, H.-W., Ciccosanti, C. T., Zhao, L., Xiao, R., Krug, R. M., & Montelione, G. T. (2011). Dimer Interface of the Effector Domain of Non-structural Protein 1 from Influenza A Virus: AN INTERFACE WITH MULTIPLE FUNCTIONS *. *Journal of Biological Chemistry*, *286*(29), 26050–26060. https://doi.org/10.1074/jbc.M111.248765

Baksh, K. A., & Zamble, D. B. (2020). Allosteric control of metal-responsive transcriptional regulators in bacteria. *Journal of Biological Chemistry*, *295*(6), 1673–1684. https://doi.org/https://doi.org/10.1074/jbc.REV119.011444

Beck, L. (2019). Expression and function of Slc34 sodium–phosphate co-transporters in skeleton and teeth. *Pflügers Archiv - European Journal of Physiology*, *471*(1), 175–184. https://doi.org/10.1007/s00424-018-2240-y

Becker, W., Bhattiprolu, K. C., Gubensäk, N., & Zangger, K. (2018a). Investigating Protein-Ligand Interactions by Solution Nuclear Magnetic Resonance Spectroscopy. *Chemphyschem : A European Journal of Chemical Physics and Physical Chemistry*, *19*(8), 895–906. https://doi.org/10.1002/cphc.201701253

Becker, W., Bhattiprolu, K. C., Gubensäk, N., & Zangger, K. (2018b). Investigating Protein-Ligand Interactions by Solution Nuclear Magnetic Resonance Spectroscopy. *Chemphyschem : A European Journal of Chemical Physics and Physical Chemistry*, *19*(8), 895–906. https://doi.org/10.1002/cphc.201701253

Bhatwa, A., Wang, W., Hassan, Y. I., Abraham, N., Li, X.-Z., & Zhou, T. (2021). Challenges Associated With the Formation of Recombinant Protein Inclusion Bodies in Escherichia coli and Strategies to Address Them for Industrial Applications. *Frontiers in Bioengineering and Biotechnology*, *9*. https://doi.org/10.3389/fbioe.2021.630551

Brautigam, C. A. (2015). Fitting two- and three-site binding models to isothermal titration calorimetric data. *Methods*, *76*, 124–136. https://doi.org/https://doi.org/10.1016/j.ymeth.2014.11.018

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., & Warren, G. L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallographica. Section D, Biological Crystallography*, *54*(Pt 5), 905–921. https://doi.org/10.1107/s0907444998003254

Buel, G. R., & Walters, K. J. (2022). Can AlphaFold2 predict the impact of missense mutations on structure? *Nature Structural & Molecular Biology*, *29*(1), 1–2. https://doi.org/10.1038/s41594-021-00714-2

C., M. H., M., W. K., P., A. A., & Arash, K. (2022). Global Analysis of Biomineralization Genes in Magnetospirillum magneticum AMB-1. *MSystems*, *7*(1), e01037-21. https://doi.org/10.1128/msystems.01037-21

Cassandri, M., Smirnov, A., Novelli, F., Pitolli, C., Agostini, M., Malewicz, M., Melino, G., & Raschellà, G. (2017). Zinc-finger proteins in health and disease. *Cell Death Discovery*, *3*(1), 17071. https://doi.org/10.1038/cddiscovery.2017.71

Chen, A. Y., Adamek, R. N., Dick, B. L., Credille, C. V, Morrison, C. N., & Cohen, S. M. (2019). Targeting Metalloenzymes for Therapeutic Intervention. *Chemical Reviews*, *119*(2), 1323–1455. https://doi.org/10.1021/acs.chemrev.8b00201

Chen, J., Im, W., & Brooks, C. L. (2004). Refinement of NMR Structures Using Implicit Solvent and Advanced Sampling Techniques. *Journal of the American Chemical Society*, *126*(49), 16038–16047. https://doi.org/10.1021/ja047624f

Chen, K., & Tjandra, N. (2012). The use of residual dipolar coupling in studying proteins by NMR. *Topics in Current Chemistry*, *326*, 47–67. https://doi.org/10.1007/128_2011_215

Clore, G. M., & Iwahara, J. (2009). Theory, Practice, and Applications of Paramagnetic Relaxation Enhancement for the Characterization of Transient Low-Population States of Biological Macromolecules and Their Complexes. *Chemical Reviews*, *109*(9), 4108–4139. https://doi.org/10.1021/cr900033p

Collman, J. P., Boulatov, R., Sunderland, C. J., & Fu, L. (2004). Functional Analogues of Cytochrome c Oxidase, Myoglobin, and Hemoglobin. *Chemical Reviews*, *104*(2), 561–588. https://doi.org/10.1021/cr0206059

Corrêa, F., & Farah, C. S. (2007). Different effects of trifluoroethanol and glycerol on the stability of tropomyosin helices and the head-to-tail complex. *Biophysical Journal*, *92*(7), 2463–2475. https://doi.org/10.1529/biophysj.106.098541

Crowley, E. L., & Rafferty, S. P. (2019). Review of lactose-driven auto-induction expression of isotope-labelled proteins. *Protein Expression and Purification*, *157*, 70–85. https://doi.org/https://doi.org/10.1016/j.pep.2019.01.007

Dong, R., Peng, Z., Zhang, Y., & Yang, J. (2018). mTM-align: an algorithm for fast and accurate multiple protein structure alignment. *Bioinformatics*, *34*(10), 1719–1725. https://doi.org/10.1093/bioinformatics/btx828

East, K. W., Skeens, E., Cui, J. Y., Belato, H. B., Mitchell, B., Hsu, R., Batista, V. S., Palermo, G., & Lisi, G. P. (2020). NMR and computational methods for molecular resolution of allosteric pathways in enzyme complexes. *Biophysical Reviews*, *12*(1), 155–174. https://doi.org/10.1007/s12551-019-00609-z

Eaton, W. A., Palmer, G., Fee, J. A., Kimura, T., & Lovenberg, W. (1971). Tetrahedral Iron in the Active Center of Plant Ferredoxins and Beef Adrenodoxin*. *Proceedings of the National Academy of Sciences*, *68*(12), 3015–3020. https://doi.org/10.1073/pnas.68.12.3015

Emwas, A.-H., Szczepski, K., Poulson, B. G., Chandra, K., McKay, R. T., Dhahri, M., Alahmari, F., Jaremko, L., Lachowicz, J. I., & Jaremko, M. (2020). NMR as a "Gold Standard" Method in Drug Design and Discovery. *Molecules (Basel, Switzerland)*, *25*(20). https://doi.org/10.3390/molecules25204597

Fowler, N. J., Sljoka, A., & Williamson, M. P. (2021). The accuracy of NMR protein structures in the Protein Data Bank. *Structure*, *29*(12), 1430-1439.e2. https://doi.org/https://doi.org/10.1016/j.str.2021.07.001

Francis, D. M., & Page, R. (2010). Strategies to optimize protein expression in E. coli. *Current Protocols in Protein Science*, *Chapter 5*(1), 5.24.1-5.24.29. https://doi.org/10.1002/0471140864.ps0524s61

Furukawa, A., Konuma, T., Yanaka, S., & Sugase, K. (2016). Quantitative analysis of protein-ligand interactions by NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *96*, 47–57. https://doi.org/10.1016/j.pnmrs.2016.02.002

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., & Bairoch, A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, *31*(13), 3784–3788. https://doi.org/10.1093/nar/gkg563

Gernert, K. M., Surles, M. C., Labean, T. H., Richardson, J. S., & Richardson, D. C. (1995). The Alacoil: A very tight, antiparallel coiled-coil of helices. *Protein Science*, *4*(11), 2252–2260. https://doi.org/https://doi.org/10.1002/pro.5560041102

Gołowicz, D., Kasprzak, P., Orekhov, V., & Kazimierczuk, K. (2020). Fast time-resolved NMR with non-uniform sampling. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *116*, 40–55. https://doi.org/https://doi.org/10.1016/j.pnmrs.2019.09.003

Grünberg, K., Müller, E. C., Otto, A., Reszka, R., Linder, D., Kube, M., Reinhardt, R., & Schüler, D. (2004). Biochemical and Proteomic Analysis of the Magnetosome Membrane in Magnetospirillum gryphiswaldense. *Applied and Environmental Microbiology*, *70*(2), 1040–1050. https://doi.org/10.1128/AEM.70.2.1040-1050.2004

Gurnon, D. G., Whitaker, J. A., & Oakley, M. G. (2003). Design and Characterization of a Homodimeric Antiparallel Coiled Coil. *Journal of the American Chemical Society*, *125*(25), 7518–7519. https://doi.org/10.1021/ja0357590

Hans Wedepohl, K. (1995). The composition of the continental crust. *Geochimica et Cosmochimica Acta*, *59*(7), 1217–1232. https://doi.org/https://doi.org/10.1016/0016-7037(95)00038-2

Hershey, D. M., Ren, X., Melnyk, R. A., Browne, P. J., Ozyamak, E., Jones, S. R., Chang, M. C. Y., Hurley, J. H., & Komeili, A. (2016). MamO Is a Repurposed Serine Protease that Promotes Magnetite Biomineralization through Direct Transition Metal Binding in Magnetotactic Bacteria. *PLoS Biology*, *14*(3), e1002402. https://doi.org/10.1371/journal.pbio.1002402

Honarmand Ebrahimi, K., Hagedoorn, P.-L., & Hagen, W. R. (2015). Unity in the Biochemistry of the Iron-Storage Proteins Ferritin and Bacterioferritin. *Chemical Reviews*, *115*(1), 295–326. https://doi.org/10.1021/cr5004908

Hu, Y., Cheng, K., He, L., Zhang, X., Jiang, B., Jiang, L., Li, C., Wang, G., Yang, Y., & Liu, M. (2021a). NMR-Based Methods for Protein Analysis. *Analytical Chemistry*, *93*(4), 1866–1879. https://doi.org/10.1021/acs.analchem.0c03830

Hu, Y., Cheng, K., He, L., Zhang, X., Jiang, B., Jiang, L., Li, C., Wang, G., Yang, Y., & Liu, M. (2021b). NMR-Based Methods for Protein Analysis. *Analytical Chemistry*, *93*(4), 1866–1879. https://doi.org/10.1021/acs.analchem.0c03830

Johnson, E., Cascio, D., Sawaya, M. R., Gingery, M., & Schröder, I. (2005). Crystal Structures of a Tetrahedral Open Pore Ferritin from the Hyperthermophilic Archaeon Archaeoglobus fulgidus. *Structure*, *13*(4), 637–648. https://doi.org/https://doi.org/10.1016/j.str.2005.01.019

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021a). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021b). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Katz, A. K., Glusker, J. P., Beebe, S. A., & Bock, C. W. (1996a). Calcium Ion Coordination: A Comparison with That of Beryllium, Magnesium, and Zinc. *Journal of the American Chemical Society*, *118*(24), 5752–5763. https://doi.org/10.1021/ja953943i

Katz, A. K., Glusker, J. P., Beebe, S. A., & Bock, C. W. (1996b). Calcium Ion Coordination: A Comparison with That of Beryllium, Magnesium, and Zinc. *Journal of the American Chemical Society*, *118*(24), 5752–5763. https://doi.org/10.1021/ja953943i

Kidane, T. Z., Sauble, E., & Linder, M. C. (2006). Release of iron from ferritin requires lysosomal activity. *American Journal of Physiology-Cell Physiology*, *291*(3), C445–C455. https://doi.org/10.1152/ajpcell.00505.2005

Kleckner, I. R., & Foster, M. P. (2011). An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica et Biophysica Acta*, *1814*(8), 942–968. https://doi.org/10.1016/j.bbapap.2010.10.012

Koorts, A. M., & Viljoen, M. (2007). Ferritin and ferritin isoforms I: Structure-function relationships, synthesis, degradation and secretion. *Archives of Physiology and Biochemistry*, *113*(1), 30–54. https://doi.org/10.1080/13813450701318583

Laitaoja, M., Valjakka, J., & Jänis, J. (2013). Zinc Coordination Spheres in Protein Structures. *Inorganic Chemistry*, *52*(19), 10983–10991. https://doi.org/10.1021/ic401072d

Landschulz, W. H., Johnson, P. F., & McKnight, S. L. (1988). The Leucine Zipper: A Hypothetical Structure Common to a New Class of DNA Binding Proteins. *Science*, *240*(4860), 1759–1764. https://doi.org/10.1126/science.3289117

Lefevre, C. T., & Bazylinski, D. A. (2013). Ecology, Diversity, and Evolution of Magnetotactic Bacteria. *Microbiology and Molecular Biology Reviews*, *77*(3), 497–526. https://doi.org/10.1128/mmbr.00021-13

Lohße, A., Borg, S., Raschdorf, O., Kolinko, I., Tompa, E., Pósfai, M., Faivre, D., Baumgartner, J., & Schüler, D. (2014). Genetic dissection of the mamAB and mms6 operons reveals a gene set essential for magnetosome biogenesis in Magnetospirillum gryphiswaldense. *Journal of Bacteriology*, *196*(14), 2658–2669. https://doi.org/10.1128/JB.01716-14

Lohße, A., Ullrich, S., Katzmann, E., Borg, S., Wanner, G., Richter, M., Voigt, B., Schweder, T., & Schüler, D. (2011). Functional analysis of the magnetosome Island in magnetospirillum gryphiswaldense: The mamAB operon is sufficient for magnetite biomineralization. *PLoS ONE*, *6*(10), 25561. https://doi.org/10.1371/journal.pone.0025561

Lu, C.-H., Lin, Y.-F., Lin, J.-J., & Yu, C.-S. (2012). Prediction of metal ion-binding sites in proteins using the fragment transformation method. *PloS One*, *7*(6), e39252. https://doi.org/10.1371/journal.pone.0039252

Ma, K., Zhao, H., Zheng, X., Sun, H., Hu, L., Zhu, L., Shen, Y., Luo, T., Dai, H., & Wang, J. (2017). NMR studies of the interactions between AMB-1 Mms6 protein and magnetosome Fe3O4 nanoparticles. *J. Mater. Chem. B*, *5*(16), 2888–2895. https://doi.org/10.1039/C7TB00570A

Maret, W. (2013). Zinc Biochemistry: From a Single Zinc Enzyme to a Key Element of Life. *Advances in Nutrition*, *4*(1), 82–91. https://doi.org/https://doi.org/10.3945/an.112.003038

Maret, W., & Li, Y. (2009). Coordination Dynamics of Zinc in Proteins. *Chemical Reviews*, *109*(10), 4682–4707. https://doi.org/10.1021/cr800556u

Marin, F., Luquet, G., Marie, B., & Medakovic, D. (2007). Molluscan Shell Proteins: Primary Structure, Origin, and Evolution. In *Current Topics in Developmental Biology* (Vol. 80, pp. 209–276). Academic Press. https://doi.org/https://doi.org/10.1016/S0070-2153(07)80006-8

Murat, D., Falahati, V., Bertinetti, L., Csencsits, R., Körnig, A., Downing, K., Faivre, D., & Komeili, A. (2012). The magnetosome membrane protein, MmsF, is a major regulator of magnetite biomineralization in Magnetospirillum magneticum AMB-1. *Molecular Microbiology*, *85*(4), 684–699. https://doi.org/10.1111/j.1365-2958.2012.08132.x

Murat, D., Quinlan, A., Vali, H., & Komeili, A. (2010). Comprehensive genetic dissection of the magnetosome gene island reveals the step-wise assembly of a prokaryotic organelle. *Proceedings of the National Academy of Sciences*, *107*(12), 5593–5598. https://doi.org/10.1073/pnas.0914439107

Nam, H., Hwang, B. J., Choi, D. Y., Shin, S., & Choi, M. (2020). Tobacco etch virus (TEV) protease with multiple mutations to improve solubility and reduce self-cleavage exhibits enhanced enzymatic activity. *FEBS Open Bio*, *10*(4), 619–626. https://doi.org/10.1002/2211-5463.12828

Nudelman, H., Valverde-Tercedor, C., Kolusheva, S., Perez Gonzalez, T., Widdrat, M., Grimberg, N., Levi, H., Nelkenbaum, O., Davidov, G., Faivre, D., Jimenez-Lopez, C., & Zarivach, R. (2016). Structure-function studies of the magnetite-biomineralizing magnetosome-associated protein MamC. *Journal of Structural Biology*, *194*(3), 244–252. https://doi.org/10.1016/j.jsb.2016.03.001

Okamura, Y., Takeyama, H., & Matsunaga, T. (2001). A magnetosome-specific GTPase from the magnetic bacterium Magnetospirillum magneticum AMB-1. *The Journal of Biological Chemistry*, *276*(51), 48183–48188. https://doi.org/10.1074/jbc.M106408200

Overhauser, A. W. (1953). Polarization of Nuclei in Metals. *Phys. Rev.*, *92*(2), 411–415. https://doi.org/10.1103/PhysRev.92.411

Patzer, S. I., & Hantke, K. (2000). The Zinc-responsive Regulator Zur and Its Control of theznu Gene Cluster Encoding the ZnuABC Zinc Uptake System in Escherichia coli *. *Journal of Biological Chemistry*, *275*(32), 24321–24332. https://doi.org/https://doi.org/10.1074/jbc.M001775200

Pearson, R. G. (1963). Hard and Soft Acids and Bases. *Journal of the American Chemical Society*, *85*(22), 3533–3539. https://doi.org/10.1021/ja00905a001

Pearson, R. G. (1968). Hard and soft acids and bases, HSAB, part 1: Fundamental principles. *Journal of Chemical Education*, *45*(9), 581. https://doi.org/10.1021/ed045p581

Pellecchia, M., Sem, D. S., & Wüthrich, K. (2002). Nmr in drug discovery. *Nature Reviews Drug Discovery*, *1*(3), 211–219. https://doi.org/10.1038/nrd748

Puthenveetil, R., & Vinogradova, O. (2019). Solution NMR: A powerful tool for structural and functional studies of membrane proteins in reconstituted environments. *Journal of Biological Chemistry*, *294*(44), 15914–15931. https://doi.org/10.1074/jbc.REV119.009178

Raran-Kurussi, S., Cherry, S., Zhang, D., & Waugh, D. S. (2017a). Removal of Affinity Tags with TEV Protease. *Methods in Molecular Biology (Clifton, N.J.)*, *1586*, 221. https://doi.org/10.1007/978-1-4939-6887-9_14

Raran-Kurussi, S., Cherry, S., Zhang, D., & Waugh, D. S. (2017b). Removal of Affinity Tags with TEV Protease. *Methods in Molecular Biology (Clifton, N.J.)*, *1586*, 221. https://doi.org/10.1007/978-1-4939-6887-9_14

Rawlings, A. E., Bramble, J. P., Hounslow, A. M., Williamson, M. P., Monnington, A. E., Cooke, D. J., & Staniland, S. S. (2016). Ferrous Iron Binding Key to Mms6 Magnetite Biomineralisation: A Mechanistic Study to Understand Magnetite Formation Using pH Titration and NMR Spectroscopy. *Chemistry - A European Journal*, *22*(23), 7885–7894. https://doi.org/10.1002/chem.201600322

Rawlings, A. E., Bramble, J. P., Walker, R., Bain, J., Galloway, J. M., & Staniland, S. S. (2014). Self-assembled MmsF proteinosomes control magnetite nanoparticle formation in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(45), 19094–19099. https://doi.org/10.1073/pnas.1409256111

Rawlings, A. E., Somner, L. A., Fitzpatrick-Milton, M., Roebuck, T. P., Gwyn, C., Liravi, P., Seville, V., Neal, T. J., Mykhaylyk, O. O., Baldwin, S. A., & Staniland, S. S. (2019). Artificial coiled coil biomineralisation protein for the synthesis of magnetic nanoparticles. *Nature Communications*, *10*(1), 1–9. https://doi.org/10.1038/s41467-019-10578-2

Reyes-Caballero, H., Campanello, G. C., & Giedroc, D. P. (2011). Metalloregulatory proteins: metal selectivity and allosteric switching. *Biophysical Chemistry*, *156*(2–3), 103–114. https://doi.org/10.1016/j.bpc.2011.03.010

Sánchez, M., Sabio, L., Gálvez, N., Capdevila, M., & Dominguez-Vera, J. M. (2017). Iron chemistry at the service of life. *IUBMB Life*, *69*(6), 382–388. https://doi.org/10.1002/iub.1602

Scheffel, A., Gärdes, A., Grünberg, K., Wanner, G., & Schüler, D. (2008). The major magnetosome proteins MamGFDC are not essential for magnetite biomineralization in Magnetospirillum gryphiswaldense but regulate the size of magnetosome crystals. *Journal of Bacteriology*, *190*(1), 377–386. https://doi.org/10.1128/JB.01371-07

Scheffel, A., Gruska, M., Faivre, D., Linaroudis, A., Plitzko, J. M., & Schüler, D. (2005). An acidic protein aligns magnetosomes along a filamentous structure in magnetotactic bacteria. *Nature 2005 440:7080*, *440*(7080), 110–114. https://doi.org/10.1038/nature04382

Sénèque, O., & Latour, J.-M. (2010). Coordination Properties of Zinc Finger Peptides Revisited: Ligand Competition Studies Reveal Higher Affinities for Zinc and Cobalt. *Journal of the American Chemical Society*, *132*(50), 17760–17774. https://doi.org/10.1021/ja104992h

Shen, Y., & Bax, A. (2013). Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *Journal of Biomolecular NMR*, *56*(3), 227–241. https://doi.org/10.1007/s10858-013-9741-y

Shen, Y., & Bax, A. (2015). Protein Structural Information Derived from NMR Chemical Shift with the Neural Network Program TALOS-N. In H. Cartwright (Ed.), *Artificial Neural Networks* (pp. 17–32). Springer New York. https://doi.org/10.1007/978-1-4939-2239-0_2

Shikama, K. (2006). Nature of the FeO2 bonding in myoglobin and hemoglobin: A new molecular paradigm. *Progress in Biophysics and Molecular Biology*, *91*(1), 83–162. https://doi.org/https://doi.org/10.1016/j.pbiomolbio.2005.04.001

Shiraki, K., Nishikawa, K., & Goto, Y. (1995). Trifluoroethanol-induced Stabilization of the α-Helical Structure of β-Lactoglobulin: Implication for Non-hierarchical Protein Folding. *Journal of Molecular Biology*, *245*(2), 180–194. https://doi.org/10.1006/JMBI.1994.0015

Solomon, I. (1955). Relaxation Processes in a System of Two Spins. *Physical Review*, *99*(2), 559–565. https://doi.org/10.1103/PhysRev.99.559

Sudarev, V. V, Dolotova, S. M., Bukhalovich, S. M., Bazhenov, S. V, Ryzhykau, Y. L., Uversky, V. N., Bondarev, N. A., Osipov, S. D., Mikhailov, A. E., Kuklina, D. D., Murugova, T. N., Manukhov, I. V, Rogachev, A. V, Gordeliy, V. I., Gushchin, I. Yu., Kuklin, A. I., & Vlasov, A. V. (2023). Ferritin self-assembly, structure, function, and biotechnological applications. *International Journal of Biological Macromolecules*, *224*, 319–343. https://doi.org/https://doi.org/10.1016/j.ijbiomac.2022.10.126

Tanaka, M., Arakaki, A., & Matsunaga, T. (2010). Identification and functional characterization of liposome tubulation protein from magnetotactic bacteria. *Molecular Microbiology*, *76*(2), 480–488. https://doi.org/10.1111/j.1365-2958.2010.07117.x

Taoka, A., Eguchi, Y., Shimoshige, R., & Fukumori, Y. (2023). Recent advances in studies on magnetosome-associated proteins composing the bacterial geomagnetic sensor organelle. *Microbiology and Immunology*, *67*(5), 228–238. https://doi.org/10.1111/1348-0421.13062

Trevitt, C. R., Yashwanth Kumar, D. R., Fowler, N. J., & Williamson, M. P. (2024). Interactions between the protein barnase and co-solutes studied by NMR. *Communications Chemistry*, *7*(1), 44. https://doi.org/10.1038/s42004-024-01127-0

Truebestein, L., & Leonard, T. A. (2016). Coiled-coils: The long and short of it. *BioEssays*, *38*(9), 903–916. https://doi.org/https://doi.org/10.1002/bies.201600062

Uebe, R., Junge, K., Henn, V., Poxleitner, G., Katzmann, E., Plitzko, J. M., Zarivach, R., Kasama, T., Wanner, G., Pósfai, M., Böttger, L., Matzanke, B., & Schüler, D. (2011). The cation diffusion facilitator proteins MamB and MamM of Magnetospirillum gryphiswaldense have distinct and complex functions, and are involved in magnetite biomineralization and magnetosome membrane assembly. *Molecular Microbiology*, *82*(4), 818–835. https://doi.org/10.1111/j.1365-2958.2011.07863.x

Uebe, R., & Schüler, D. (2016). Magnetosome biogenesis in magnetotactic bacteria. In *Nature Reviews Microbiology* (Vol. 14, Issue 10, pp. 621–637). Nature Publishing Group. https://doi.org/10.1038/nrmicro.2016.99

Valverde-Tercedor, C., Montalbán-López, M., Perez-Gonzalez, T., Sanchez-Quesada, M. S., Prozorov, T., Pineda-Molina, E., Fernandez-Vivas, M. A., Rodriguez-Navarro, A. B., Trubitsyn, D., Bazylinski, D. A., & Jimenez-Lopez, C. (2015). Size control of in vitro synthesized magnetite crystals by the MamC protein of Magnetococcus marinus strain MC-1. *Applied Microbiology and Biotechnology*, *99*(12), 5109–5121. https://doi.org/10.1007/s00253-014-6326-y

VanZile, M. L., Chen, X., & Giedroc, D. P. (2002). Structural Characterization of Distinct α3N and α5 Metal Sites in the Cyanobacterial Zinc Sensor SmtB. *Biochemistry*, *41*(31), 9765–9775. https://doi.org/10.1021/bi0201771

Vinson, C., Myakishev, M., Acharya, A., Mir, A. A., Moll, J. R., & Bonovich, M. (2002). Classification of Human B-ZIP Proteins Based on Dimerization Properties. *Molecular and Cellular Biology*, *22*(18), 6321–6335. https://doi.org/10.1128/MCB.22.18.6321-6335.2002

Wahyudi, A. T., Takeyama, H., Okamura, Y., Fukuda, Y., & Matsunaga, T. (2003). Characterization of aldehyde ferredoxin oxidoreductase gene defective mutant in Magnetospirillum magneticum AMB-1. *Biochemical and Biophysical Research Communications*, *303*(1), 223–229. https://doi.org/https://doi.org/10.1016/S0006-291X(03)00303-6

Wallyn, J., Anton, N., & Vandamme, T. F. (2019). Synthesis, principles, and properties of magnetite nanoparticles for in vivo imaging applications—A review. In *Pharmaceutics* (Vol. 11, Issue 11). MDPI AG. https://doi.org/10.3390/pharmaceutics11110601

Wang, X., Kirberger, M., Qiu, F., Chen, G., & Yang, J. J. (2009). Towards predicting Ca2+-binding sites with different coordination numbers in proteins with atomic resolution. *Proteins: Structure, Function, and Bioinformatics*, *75*(4), 787–798. https://doi.org/https://doi.org/10.1002/prot.22285

Waudby, C. A., Ramos, A., Cabrita, L. D., & Christodoulou, J. (2016). Two-Dimensional NMR Lineshape Analysis. *Scientific Reports*, *6*(1), 24826. https://doi.org/10.1038/srep24826

Wei, H., Bruns, O. T., Kaul, M. G., Hansen, E. C., Barch, M., Wiśniowska, A., Chen, O., Chen, Y., Li, N., Okada, S., Cordero, J. M., Heine, M., Farrar, C. T., Montana, D. M., Adam, G., Ittrich, H., Jasanoff, A., Nielsen, P., & Bawendi, M. G. (2017). Exceedingly small iron oxide nanoparticles as positive MRI contrast agents. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(9), 2325–2330. https://doi.org/10.1073/pnas.1620145114

Williamson, M. (2011). *How Proteins Work* (First).

Williamson, M. P. (2013). Using chemical shift perturbation to characterise ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *73*, 1–16. https://doi.org/https://doi.org/10.1016/j.pnmrs.2013.02.001

Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids* (Vol. 1). Wiley Interscience.

Yan, L., Da, H., Zhang, S., López, V. M., & Wang, W. (2017). Bacterial magnetosome and its potential application. *Microbiological Research*, *203*, 19–28. https://doi.org/https://doi.org/10.1016/j.micres.2017.06.005

Yang, J., Li, S., Huang, X., Li, J., Li, L., Pan, Y., & Li, Y. (2013). MamX encoded by the mamXY operon is involved in control of magnetosome maturation in Magnetospirillum gryphiswaldense MSR-1. *BMC Microbiology*, *13*(1), 203. https://doi.org/10.1186/1471-2180-13-203

Yarra, T., Blaxer, M., & Clark, M. S. (2021). A Bivalve Biomineralization Toolbox. *Molecular Biology and Evolution*, *38*(9), 4043–4045.

Yew, Y. P., Shameli, K., Miyake, M., Kuwano, N., Bt Ahmad Khairudin, N. B., Bt Mohamad, S. E., & Lee, K. X. (2016). Green Synthesis of Magnetite (Fe3O4) Nanoparticles Using Seaweed (Kappaphycus alvarezii) Extract. *Nanoscale Research Letters*, *11*(1), 276. https://doi.org/10.1186/s11671-016-1498-2

Ying, G., Zhang, G., Yang, J., Hao, Z., Xing, W., Lu, D., Zhang, S., & Yan, L. (2022). Biomineralization and biotechnological applications of bacterial magnetosomes. *Colloids and Surfaces B: Biointerfaces*, *216*, 112556. https://doi.org/https://doi.org/10.1016/j.colsurfb.2022.112556

Appendix

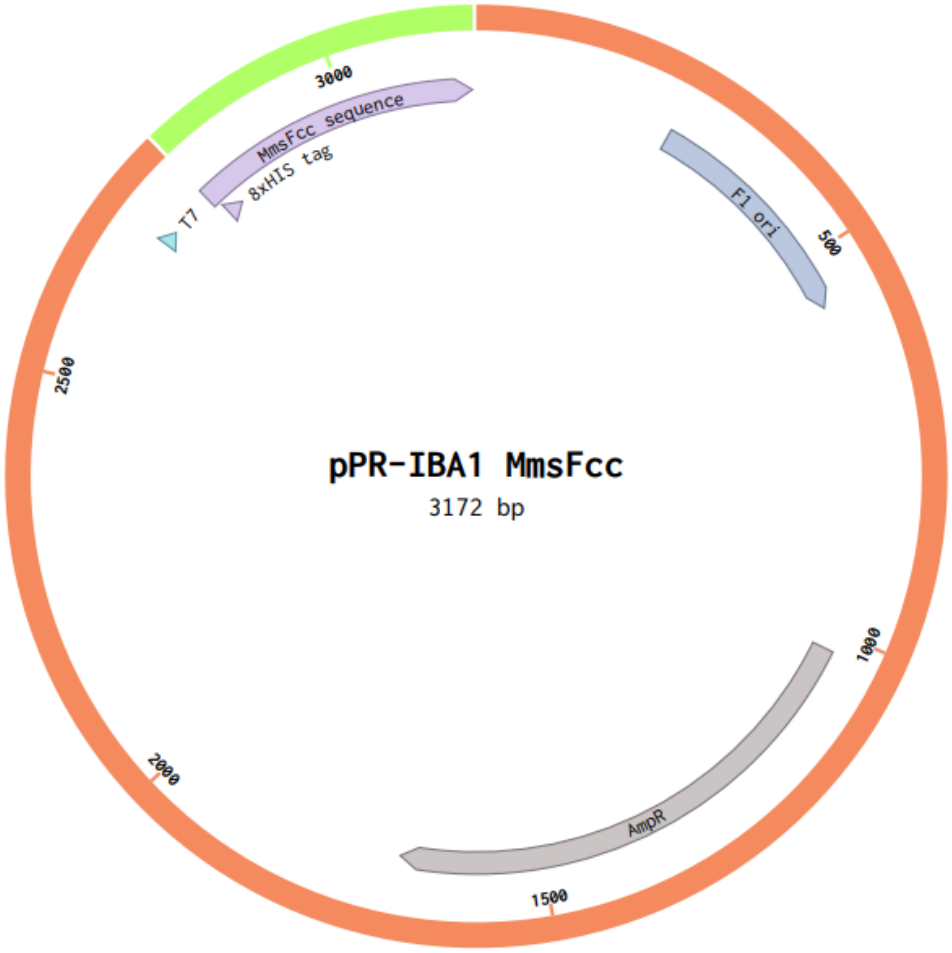
Appendix 1. Numbered amino acid sequence of MmsFcc/MmsFcc D49S.

G1	**Q33**	**Q65**	**K97**
P2	**A34**	**L66**	**L98**
S3	**R35**	**E67**	**A99**
M4	**K36**	**K68**	**Q100**
K5	**K37**	**E69**	**L101**
Q6	**K38**	**L70**	**K102**
L7	**L39**	**K71**	**K103**
E8	**A40**	**Q72**	**K104**
K9	**Q41**	**L73**	**L105**
E10	**L42**	**E74**	**Q106**
L11	**K43**	**K75**	**A107**
K12	**K44**	**E76**	
Q13	**K45**	**L77**	
L14	**L46**	**Q78**	
E15	**Q47**	**A79**	
K16	**A48**	**I80**	
E17	**D/S49**	**E81**	
L18	**R50**	**K82**	
Q19	**D51**	**Q83**	
A20	**D52**	**L84**	
I21	**E53**	**A85**	
E22	**F54**	**Q86**	
K23	**V55**	**L87**	
Q24	**Y56**	**Q88**	
L25	**F57**	**W89**	
A26	**H58**	**K90**	
Q27	**A59**	**A91**	
L28	**K60**	**Q92**	
Q29	**Q61**	**A93**	
W30	**G62**	**R94**	
K31	**M63**	**K95**	
A32	**K64**	**K96**	

Appendix 2. Amino acid sequence of His-MmsFcc

MGSHHHHHHHHGSTENLYFQGPSMKQLEKELKQLEKELQAIEKQLAQLQWKAQARKKKLAQLKKKLQA
DRDDEFVYFHAKQGMKQLEKELKQLEKELQAIEKQLAQLQWKAQARKKKLAQLKKKLQA

Appendix 3. pPR-IBA-1 MmsFcc plasmid map.



Appendix 4. MmsFcc D49S chemical shifts derived from backbone assignment.

	nH	N	C_α	C_β	C_o
P2	x	x	62.31	31.38	177.38
S3	8.55	116.89	56.87	63.25	177.39
M4	8.68	122.67	56.75	31.03	178.50
K5	8.10	119.07	57.96	31.03	179.75
Q6	7.96	120.23	54.75	27.22	175.05
L7	x	x	57.17	40.90	179.25
E8	8.22	118.04	59.10	28.68	179.59
K9	7.70	119.56	58.76	31.29	179.82
E10	8.01	121.38	55.55	28.48	177.12
L11	8.42	121.08	58.20	40.96	178.92
K12	7.87	117.64	58.08	31.05	180.43
Q13	7.88	119.39	57.96	27.09	179.21
L14	8.01	121.03	57.28	41.00	179.19
E15	8.27	118.51	59.00	28.77	179.42

K16	7.64	119.44	58.81	31.40	179.89
E17	7.96	121.26	58.63	28.58	179.48
L18	8.66	119.96	57.38	40.72	178.93
Q19	8.00	118.53	58.25	27.27	179.46
A20	7.64	121.95	54.33	16.87	181.73
I21	8.27	121.31	64.60	40.38	179.40
E22	8.52	118.92	59.38	28.78	180.30
K23	7.73	120.09	58.46	31.32	176.89
Q24	8.30	121.64	57.22	28.90	180.33
L25	8.92	121.30	57.64	41.54	179.65
A26	7.77	120.56	54.30	16.80	181.42
Q27	7.74	117.39	58.08	27.59	179.78
L28	8.32	121.06	56.92	41.47	177.14
Q29	8.63	119.39	58.37	27.11	179.75
W30	7.93	120.58	59.83	27.56	179.63
K31	8.08	120.66	59.09	31.80	179.83
A32	8.79	121.81	54.74	16.78	179.90
Q33	8.13	117.02	58.13	27.08	179.16
A34	7.82	122.08	54.28	16.87	182.20
R35	7.92	117.85	57.60	28.78	178.28
K36	8.59	120.28	59.76	31.38	180.25
K37	7.85	120.51	58.62	31.30	179.55
K38	7.45	120.41	58.08	31.27	179.18
L39	8.54	120.84	57.74	41.32	178.55
A40	7.75	119.79	54.26	16.85	180.98
Q41	7.61	117.31	57.78	27.73	179.67
L42	8.28	121.61	57.09	41.57	177.37
K43	8.54	118.53	59.40	31.34	179.86
K44	7.50	118.96	58.52	31.26	179.68
K45	x	x	58.09	31.50	178.70
L46	8.26	121.34	56.22	40.67	178.65
Q47	7.87	118.50	56.91	27.39	177.74
A48	7.80	121.27	52.93	17.71	179.56
S49	7.83	113.78	59.23	62.94	176.36
R50	7.91	121.05	56.15	29.71	177.36
D51	8.02	119.41	54.30	40.11	176.86
D52	7.85	119.85	53.60	40.29	176.97
E53	8.28	121.28	57.07	28.96	177.38
F54	8.15	118.96	57.96	38.13	176.97
V55	7.74	120.01	62.76	31.34	176.36
Y56	7.94	120.57	57.71	37.62	176.79
F57	7.97	119.36	58.03	38.13	176.71
H58	7.99	117.93	55.35	28.81	175.38
A59	7.98	123.89	51.94	18.14	178.68
K60	8.16	119.87	56.92	31.46	178.24
Q61	8.32	119.51	56.41	27.74	177.95

G62	8.23	108.76	45.20	x	176.15
M63	8.10	120.59	56.71	27.43	175.52
K64	x	x	x	x	x
Q65	x	x	57.72	26.91	178.46
L66	7.75	121.04	57.10	41.48	179.27
E67	8.22	118.04	59.10	28.68	179.59
K68	7.70	119.56	58.76	31.29	179.82
E69	8.01	121.38	55.55	28.48	177.12
L70	8.42	121.08	58.20	40.96	178.92
K71	7.84	117.81	58.11	31.07	180.37
Q72	7.88	119.39	57.96	27.09	179.21
L73	8.01	121.03	57.28	41.00	179.19
E74	8.28	118.66	59.07	28.81	179.50
K75	7.64	119.44	58.81	31.40	179.89
E76	7.96	121.26	58.63	28.58	179.48
L77	8.66	119.96	57.38	40.72	178.93
Q78	8.00	118.53	58.25	27.27	179.46
A79	7.64	121.95	54.33	16.87	181.73
I80	8.27	121.31	64.60	40.38	179.40
E81	8.70	119.23	59.40	28.18	180.59
K82	7.84	120.63	58.54	31.38	179.66
Q83	7.81	120.46	58.13	27.94	179.66
L84	8.92	121.30	57.64	41.54	179.65
A85	7.77	120.56	54.30	16.80	181.42
Q86	7.65	117.74	57.84	27.60	179.73
L87	8.32	121.06	56.92	41.47	177.14
Q88	8.63	119.39	58.37	27.11	179.75
W89	7.93	120.58	59.83	27.56	179.63
K90	8.08	120.66	59.09	31.80	179.83
A91	8.79	121.81	54.74	16.78	179.90
Q92	8.13	117.02	58.13	27.08	179.16
A93	7.82	122.08	54.28	16.87	182.20
R94	7.92	117.85	57.60	28.78	178.28
K95	8.59	120.28	59.76	31.38	180.25
K96	7.85	120.51	58.62	31.30	179.55
K97	7.44	120.32	57.95	31.30	179.26
L98	8.51	120.82	57.77	41.33	178.58
A99	7.75	119.79	54.26	16.85	180.98
Q100	7.61	117.31	57.78	27.73	179.67
L101	8.28	121.61	57.09	41.57	177.37
K102	8.39	117.54	59.27	31.29	179.67
K103	7.34	117.43	57.71	31.29	179.43
K104	7.59	118.65	57.24	31.71	178.43
L105	7.65	118.36	55.03	41.77	177.67
Q106	7.55	118.58	54.90	28.06	175.16
A107	7.67	130.19	53.08	19.06	183.09

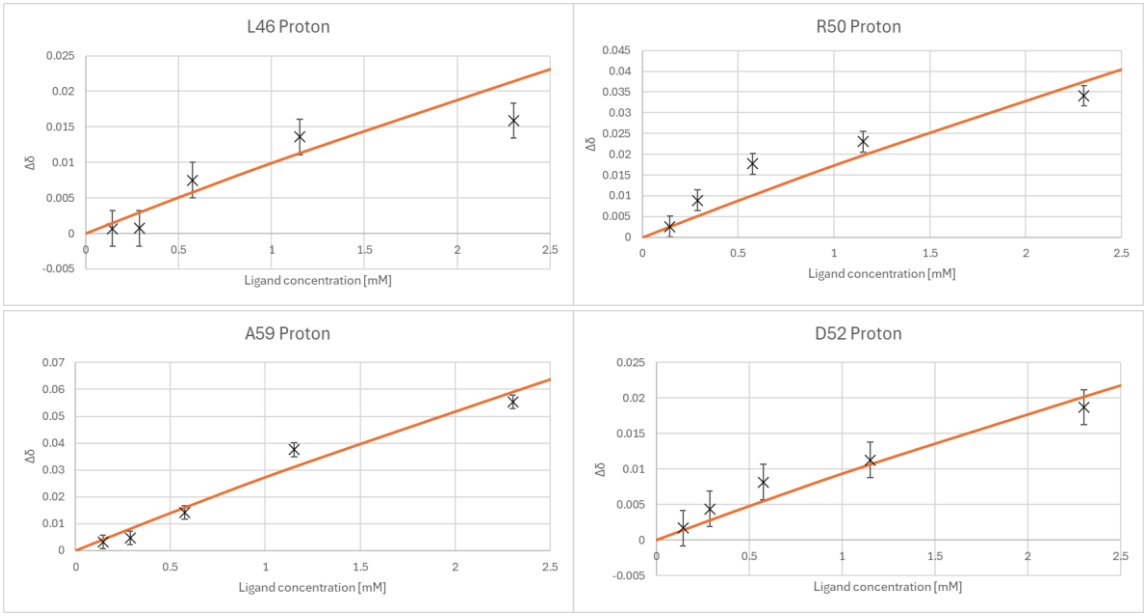
Appendix 5. MmsFcc chemical shifts derived from backbone assignment.

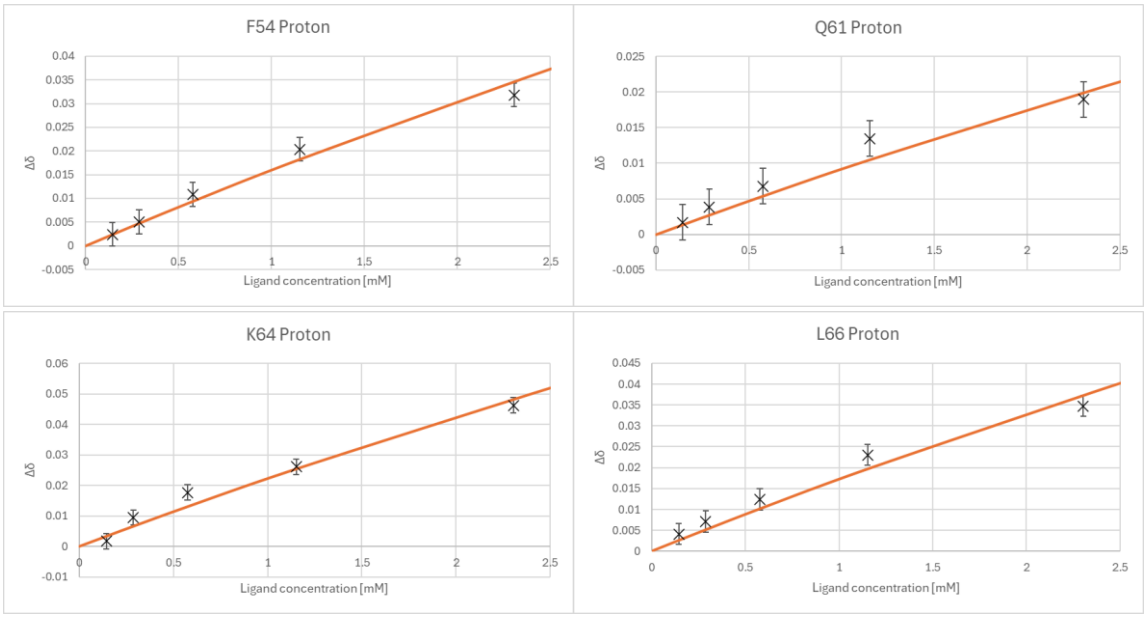
	nH	N	Cα	Cβ	Co
P2	x	x	63.00	32.28	176.89
S3	8.63	116.98	57.86	64.09	175.33
M4	8.76	122.77	57.66	32.10	178.01
K5	8.18	119.07	58.81	32.08	179.27
Q6	8.04	120.30	58.67	28.05	178.69
L7	x	x	58.15	41.93	178.76
E8	8.27	118.23	59.99	29.76	179.07
K9	7.75	119.56	59.82	32.20	179.36
E10	8.08	121.60	59.60	29.50	179.07
L11	8.50	121.16	59.13	42.00	179.47
K12	7.95	117.74	58.96	31.93	180.00
Q13	7.96	119.50	58.90	27.99	178.71
L14	8.09	121.11	58.05	41.71	179.05
E15	8.35	118.56	59.99	29.77	178.98
K16	7.72	119.56	59.68	32.26	179.41
E17	8.05	121.39	59.73	29.49	179.00
L18	8.74	120.06	58.31	41.61	178.48
Q19	8.08	118.61	59.22	28.21	178.94
A20	7.73	122.05	55.38	17.90	181.22
I21	8.35	121.45	58.19	42.49	179.42
E22	8.60	119.00	60.38	29.81	179.80
K23	7.82	120.03	59.60	32.15	179.96
Q24	8.39	121.76	59.36	29.87	179.82
L25	9.01	121.41	58.76	42.47	179.43
A26	7.85	120.71	55.30	17.82	180.91
Q27	7.82	117.49	58.98	28.60	179.31
L28	8.40	121.21	57.98	42.33	179.80
Q29	8.71	119.48	59.35	28.16	179.27
W30	8.01	120.46	60.89	28.56	179.12
K31	8.17	120.65	60.14	32.89	179.32
A32	8.87	121.92	55.52	17.86	179.45
Q33	8.21	117.12	58.93	27.96	178.66
A34	7.90	122.18	55.19	17.84	181.70
R35	8.01	117.98	58.54	29.73	178.74
K36	8.67	120.37	60.66	33.08	179.75
K37	7.94	120.52	59.62	32.36	179.05
K38	7.52	120.47	58.91	32.15	178.70
L39	8.60	120.87	58.68	42.25	178.04
A40	7.83	119.76	55.25	17.94	181.01
Q41	7.69	117.42	58.75	28.52	179.16
L42	8.37	121.69	58.03	42.49	179.07
K43	8.56	118.18	60.43	32.34	179.37

K44	7.55	118.60	59.12	32.16	179.09
K45	7.75	120.38	59.11	25.80	178.17
L46	8.11	118.91	57.24	20.10	169.90
Q47	7.74	117.26	57.01	28.63	176.68
A48	7.83	122.48	53.77	18.80	178.67
D49	8.28	119.14	55.59	40.73	176.92
R50	8.06	119.58	57.15	30.74	176.87
D51	8.22	120.16	55.11	41.08	176.40
D52	8.01	120.31	54.78	41.32	176.61
E53	8.37	121.38	58.04	29.92	176.94
F54	8.25	119.17	58.90	39.13	176.41
V55	7.85	120.29	63.80	32.34	176.37
Y56	8.04	120.62	58.66	38.52	176.18
F57	8.00	119.14	58.89	38.96	176.08
H58	8.05	118.20	56.35	29.81	174.89
A59	8.16	124.29	52.80	19.18	178.11
K60	8.30	120.35	57.47	32.53	177.64
Q61	8.47	120.28	57.33	28.69	177.46
G62	8.39	109.02	46.08	x	179.82
M63	8.12	120.76	56.06	32.50	177.78
K64	8.22	119.51	58.88	31.91	179.35
Q65	8.13	119.73	58.71	27.77	178.45
L66	7.87	121.18	58.20	41.96	x
E67	8.27	118.23	59.99	29.76	179.07
K68	7.75	119.56	59.82	32.20	179.36
E69	8.08	121.60	59.60	29.50	179.07
L70	8.50	121.16	59.13	42.00	179.47
K71	7.92	117.86	59.07	31.93	173.30
Q72	7.96	119.50	58.90	27.99	178.71
L73	8.09	121.11	58.05	41.71	179.05
E74	8.35	118.56	59.99	29.77	178.98
K75	7.72	119.56	59.68	32.26	179.41
E76	8.05	121.39	59.73	29.49	179.00
L77	8.74	120.06	58.31	41.61	178.48
Q78	8.08	118.61	59.22	28.21	178.94
A79	7.73	122.05	55.38	17.90	181.22
I80	8.35	121.45	58.19	42.49	179.42
E81	8.78	119.31	60.30	29.09	180.08
K82	7.92	120.76	59.50	32.37	179.16
Q83	7.89	120.55	59.22	28.89	179.15
L84	9.01	121.41	58.76	42.47	179.43
A85	7.85	120.71	55.30	17.82	180.91
Q86	7.74	117.82	57.23	28.61	176.65
L87	8.40	121.21	57.98	42.33	179.80
Q88	8.71	119.48	59.35	28.16	179.27
W89	8.01	120.46	60.89	28.56	179.12

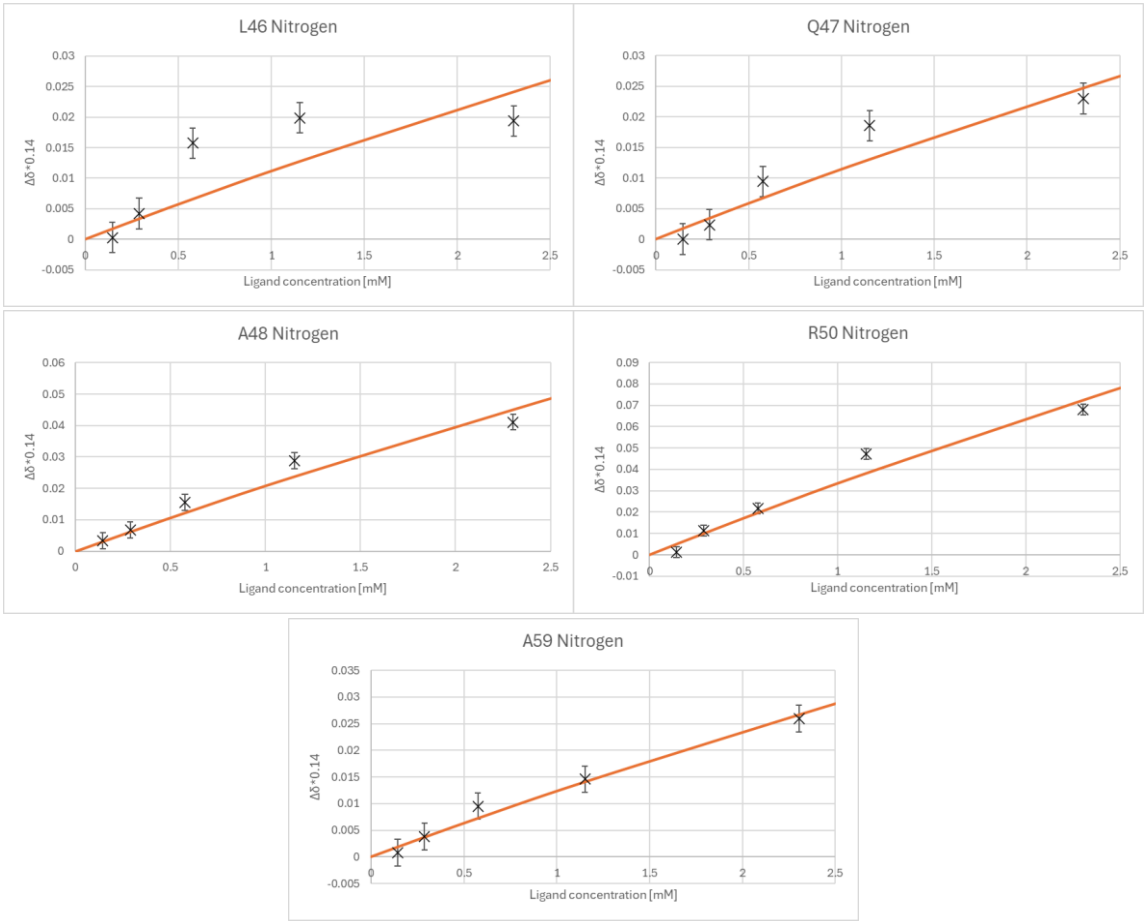
K90	8.17	120.65	60.14	32.89	179.32
A91	8.87	121.92	55.52	17.86	179.45
Q92	8.21	117.12	58.93	27.96	178.66
A93	7.90	122.18	55.19	17.84	181.70
R94	8.01	117.98	58.54	29.73	178.74
K95	8.67	120.37	60.66	33.08	179.75
K96	7.94	120.52	59.62	32.36	179.05
K97	7.52	120.47	58.91	32.15	178.70
L98	8.60	120.87	58.68	42.25	178.04
A99	7.83	119.76	55.25	17.94	181.01
Q100	7.69	117.42	58.75	28.52	179.16
L101	8.37	121.69	58.03	42.49	179.07
K102	8.48	117.64	60.28	32.38	179.14
K103	7.42	117.54	58.76	32.16	178.94
K104	7.67	118.75	58.28	32.68	177.95
L105	7.73	118.47	56.05	42.73	177.17
Q106	7.64	118.67	55.85	29.11	174.65
A107	7.75	130.30	53.97	20.14	182.60

Appendix 6. Binding curve fitted to proton chemical shift changes (using least square fit method) for MmsFcc + Zn^{2+} titrations. Observed data represented by crosses. Orange line represented fitted binding curve.

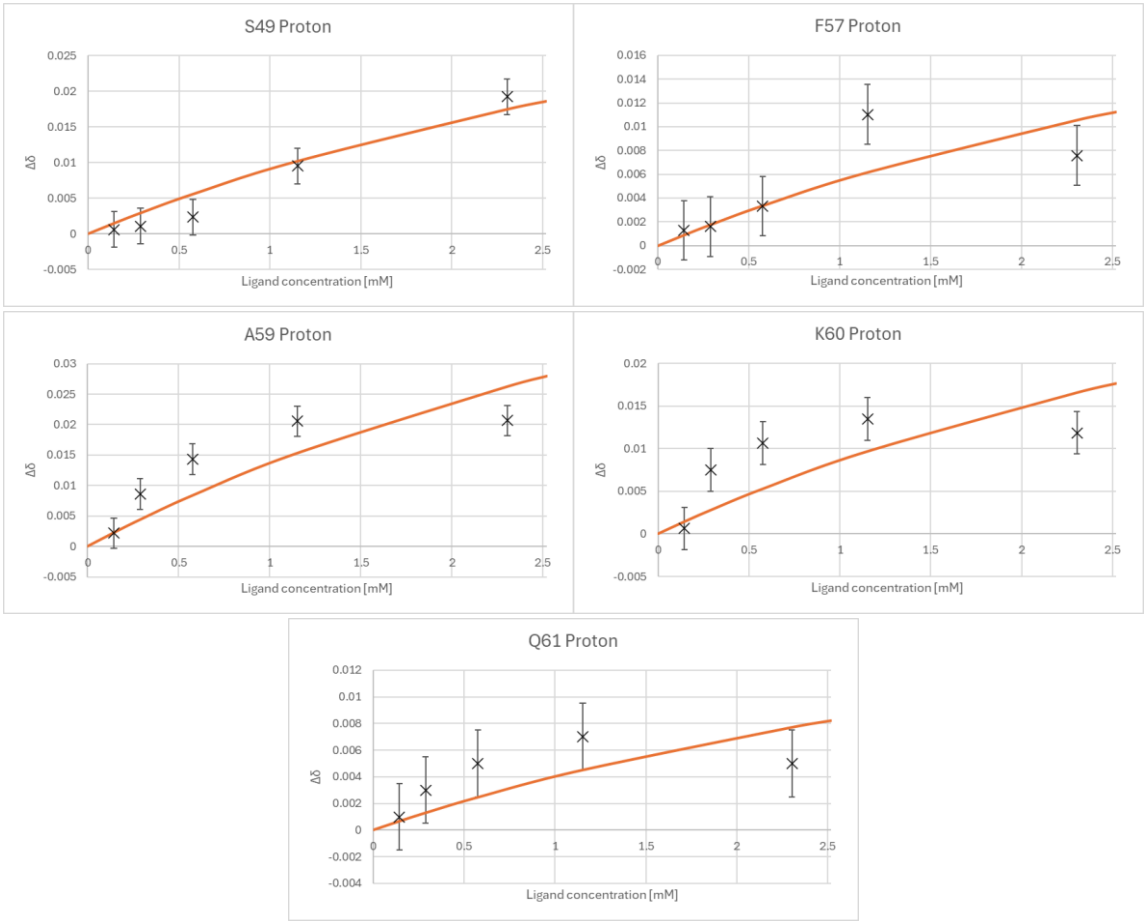




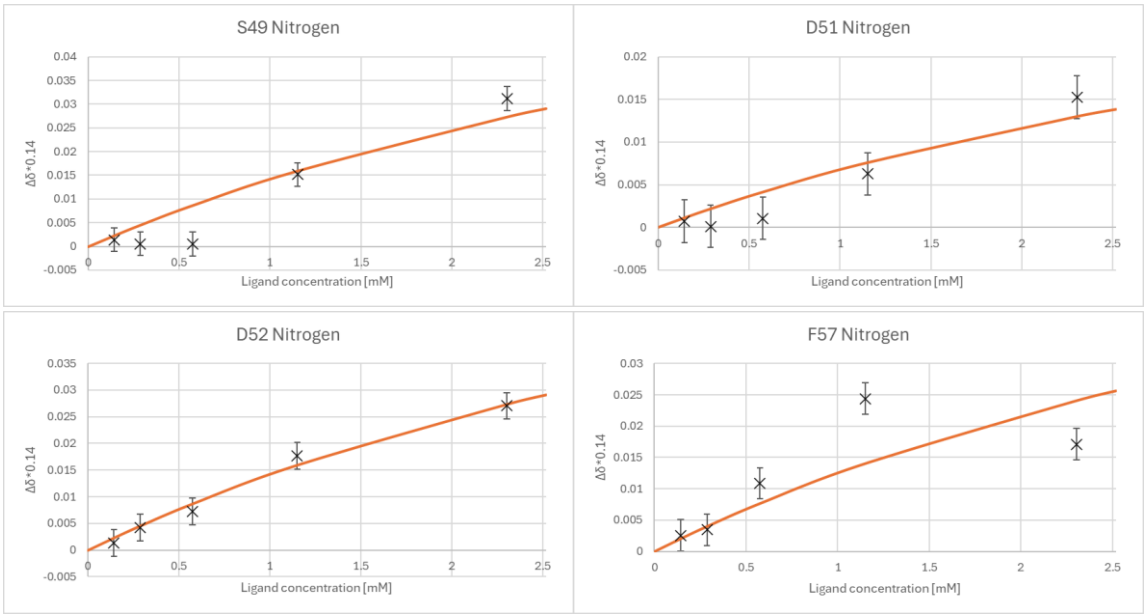
Appendix 7. Binding curve fitted to nitrogen chemical shift changes (using least square fit method) for MmsFcc + Zn^{2+} titrations. Observed data represented by crosses. Orange line represented fitted binding curve.



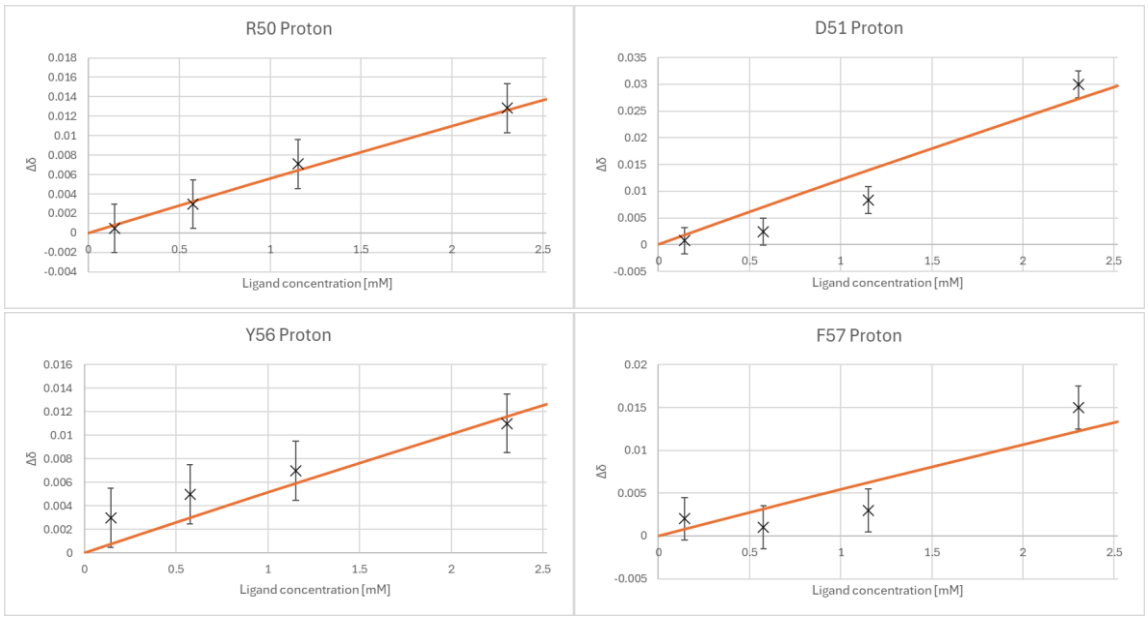
Appendix 8. Binding curve fitted to proton chemical shift changes (using least square fit method) for MmsFcc D49S + Fe^{2+} titrations. Observed data represented by crosses. Orange line represented fitted binding curve.

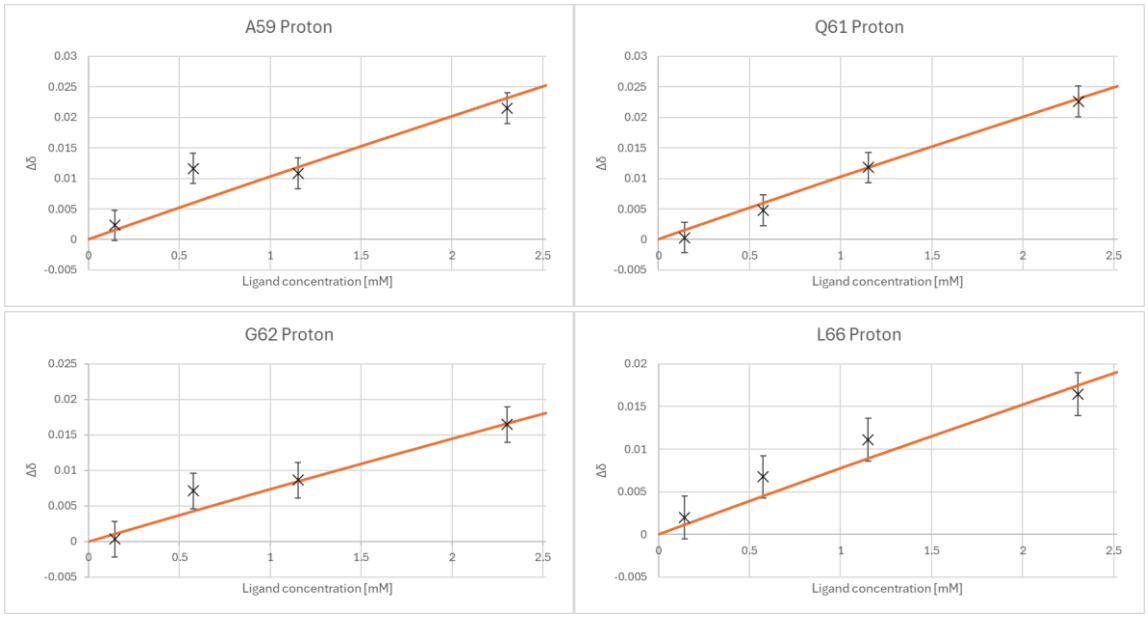


Appendix 9. Binding curve fitted to nitrogen chemical shift changes (using least square fit method) for MmsFcc D49S + Fe^{2+} titrations. Observed data represented by crosses. Orange line represented fitted binding curve.

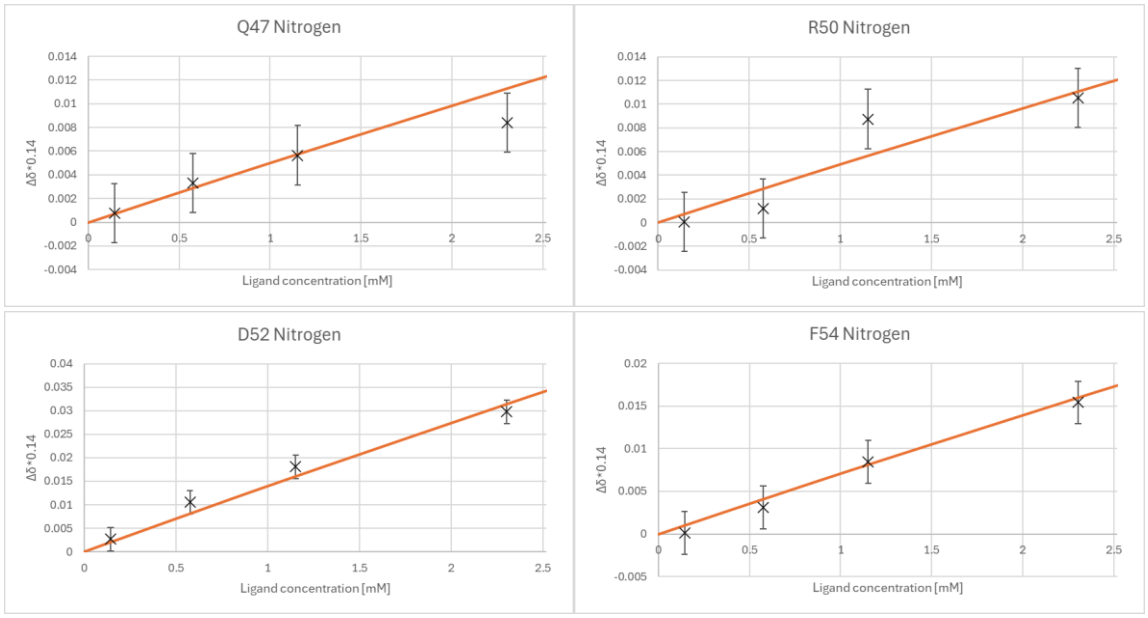


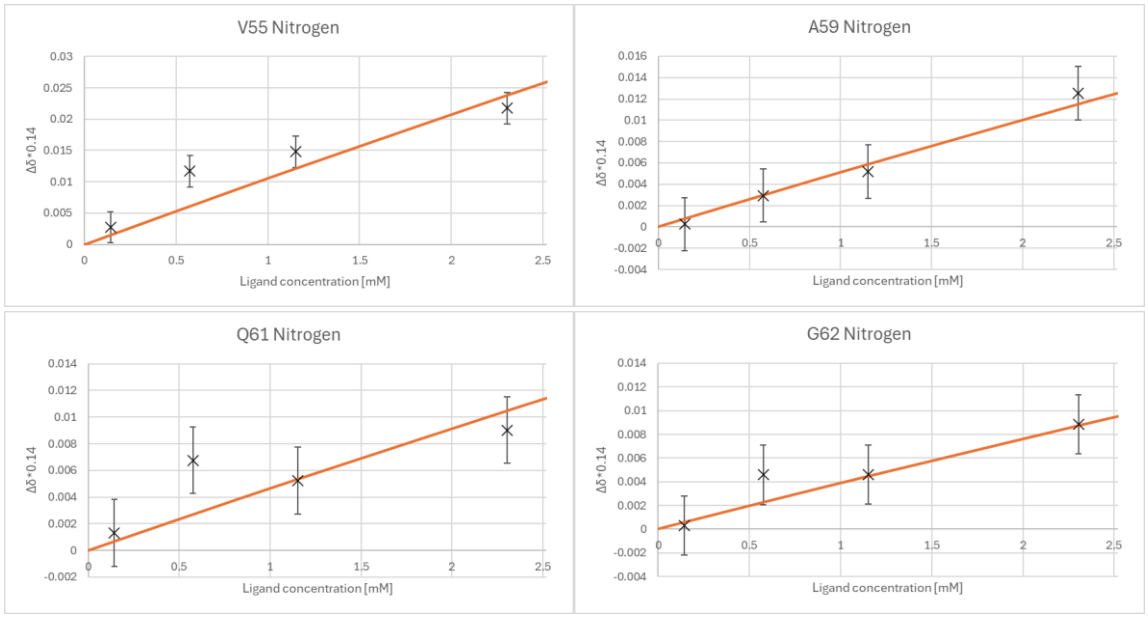
Appendix 10. Binding curve fitted to proton chemical shift changes (using least square fit method) for MmsFcc D49S + Zn^{2+} titrations. Observed data represented by crosses. Orange line represented fitted binding curve.





Appendix 11. Binding curve fitted to nitrogen chemical shift changes (using least square fit method) for MmsFcc D49S + Zn^{2+} titrations. Observed data represented by crosses. Orange line represented fitted binding curve.





Appendix 12. TALOS-N predictions of MmsFcc backbone dihedral angles and secondary structure conformation. The S^2 value is an estimate of local flexibility, based on the library of chemical shift values that TALOS predicts from. A lower value predicts more flexibility, which explains the pattern of values seen in the loop compared to the coiled-coil residues. The loop is predicted to be more flexible, especially at the C-terminal end.

Residue	Phi	Psi	Chi	S^2	Conformation	Confidence
P2	-	-	-	-	Loop	None
S3	-66.703	151.442	68.8	0.610	Helix	Strong
M4	-62.150	-32.455	-67.2	0.741	Helix	Strong
K5	-67.685	-39.336	-	0.836	Helix	Strong
Q6	-68.697	-38.569	-68.8	0.871	Helix	Strong
L7	-64.545	-44.628	-	0.886	Helix	Strong
E8	-62.015	-40.696	-	0.900	Helix	Strong
K9	-64.835	-40.997	-	0.905	Helix	Strong
E10	-64.704	-41.445	-	0.907	Helix	Strong
L11	-62.144	-41.920	-	0.904	Helix	Strong
K12	-64.141	-38.968	-	0.897	Helix	Strong
Q13	-64.944	-41.427	-	0.894	Helix	Strong
L14	-64.751	-42.810	-	0.893	Helix	Strong
E15	-61.797	-42.056	-	0.902	Helix	Strong
K16	-64.618	-42.430	-	0.901	Helix	Strong
E17	-65.476	-40.563	-	0.900	Helix	Strong
L18	-65.280	-39.025	-69.7	0.893	Helix	Strong
Q19	-64.687	-39.420	-65.7	0.895	Helix	Strong
A20	-65.522	-38.207	-	0.903	Helix	Strong
I21	-72.830	-32.823	-67.8	0.913	Helix	Strong
E22	-60.770	-43.505	-69.7	0.921	Helix	Strong

K23	-66.789	-42.477	-	0.915	Helix	Strong
Q24	-65.376	-42.060	-	0.909	Helix	Strong
L25	-63.382	-42.574	-69.6	0.902	Helix	Strong
A26	-64.054	-39.237	-	0.896	Helix	Strong
Q27	-64.835	-42.067	-67.7	0.891	Helix	Strong
L28	-63.965	-42.466	-69.5	0.890	Helix	Strong
Q29	-63.238	-41.60	-	0.897	Helix	Strong
W30	-66.362	-40.132	-	0.902	Helix	Strong
K31	-64.468	-43.355	-	0.904	Helix	Strong
A32	-63.282	-41.055	-	0.904	Helix	Strong
Q33	-63.184	-42.870	-	0.903	Helix	Strong
A34	-65.489	-41.402	-	0.905	Helix	Strong
R35	-66.516	-39.353	-	0.903	Helix	Strong
K36	-62.811	-41.734	-	0.901	Helix	Strong
K37	-64.905	-40.725	-	0.890	Helix	Strong
K38	-65.432	-41.727	-	0.886	Helix	Strong
L39	-63.526	-42.899	-	0.887	Helix	Strong
A40	-64.695	-39.420	-	0.889	Helix	Strong
Q41	-65.467	-41.999	-65.3	0.890	Helix	Strong
L42	-65.490	-43.150	-	0.887	Helix	Strong
K43	-61.604	-41.705	-	0.892	Helix	Strong
K44	-66.533	-39.716	-	0.882	Helix	Strong
K45	-67.177	-39.218	-	0.858	Helix	Strong
L46	-66.952	-37.654	-70.4	0.818	Helix	Strong
Q47	-67.144	-37.760	-65.0	0.799	Helix	Strong
A48	-67.621	-36.119	-	0.783	Helix	Strong
S49	-67.894	-35.893	-76.6	0.789	Helix	Strong
R50	-70.359	-36.425	-67.3	0.780	Helix	Strong
D51	-69.299	-32.615	-74.4	0.794	Loop	Strong
D52	-67.489	-37.148	-75.6	0.787	Loop	Strong
E53	-67.217	-38.334	-	0.774	Loop	Strong
F54	-71.358	-33.625	-75.4	0.726	Helix	Strong
V55	-68.002	-34.669	178.8	0.692	Helix	Strong
Y56	-68.739	-33.295	-76.3	0.673	Loop	Strong
F57	-69.369	-31.348	-76.7	0.657	Helix	Strong
H58	-79.210	-24.492	-72.3	0.630	Loop	Strong
A59	-69.299	-38.225	-	0.603	Loop	Generous
K60	-66.340	-35.682	-	0.619	Loop	Strong
Q61	-66.478	-39.852	-63.6	0.652	Helix	Strong
G62	-65.131	-41.082	0.000	0.718	Helix	Strong
M63	-70.843	-38.110	-66.2	0.771	Helix	Strong
K64	-66.963	-38.993	-	0.832	Helix	Strong
Q65	-67.701	-37.847	-	0.864	Helix	Strong
L66	-64.468	-45.046	-	0.879	Helix	Strong
E67	-61.569	-41.483	-	0.895	Helix	Strong
K68	-64.558	-42.695	-	0.902	Helix	Strong

E69	-65.399	-41.362	-	0.907	Helix	Strong
L70	-64.257	-42.137	-	0.903	Helix	Strong
K71	-64.456	-37.080	-	0.896	Helix	Strong
Q72	-67.052	-42.540	-	0.894	Helix	Strong
L73	-64.477	-43.074	-	0.893	Helix	Strong
E74	-61.126	-42.065	-	0.902	Helix	Strong
K75	-64.318	-42.743	-	0.901	Helix	Strong
E76	-65.289	-40.523	-	0.900	Helix	Strong
L77	-66.781	-37.462	-67.9	0.876	Helix	Strong
Q78	-65.657	-35.897	-63.6	0.869	Helix	Strong
A79	-67.842	-36.484	-	0.870	Helix	Strong
I80	66.198	-41.808	-65.9	0.896	Helix	Strong
E81	-63.144	-38.599	-68.6	0.910	Helix	Strong
K82	-64.761	-42.317	-	0.907	Helix	Strong
Q83	-65.818	-41.072	-	0.901	Helix	Strong
L84	-62.989	-41.998	-71.3	0.888	Helix	Strong
A85	-64.459	-37.955	-	0.874	Helix	Strong
Q86	-68.031	-41.203	-65.6	0.859	Helix	Strong
L87	-65.232	-41.314	-	0.867	Helix	Strong
Q88	-62.183	-42.361	-	0.885	Helix	Strong
W89	-66.376	-40.137	-	0.902	Helix	Strong
K90	-64.275	-43.730	-	0.904	Helix	Strong
A91	-63.432	-40.913	-	0.904	Helix	Strong
Q92	-62.465	-43.200	-	0.903	Helix	Strong
A93	-65.518	-41.183	-	0.905	Helix	Strong
R94	-66.516	-39.353	-	0.903	Helix	Strong
K95	-62.811	-41.734	-	0.901	Helix	Strong
K96	-64.921	-40.711	-	0.890	Helix	Strong
K97	-65.432	-41.727	-	0.886	Helix	Strong
L98	-62.517	-43.273	-	0.887	Helix	Strong
A99	-64.626	-40.051	-	0.889	Helix	Strong
Q100	-65.367	-41.104	-64.9	0.889	Helix	Strong
L101	-63.754	-41.306	-	0.874	Helix	Strong
K102	-62.458	-37.227	-	0.871	Helix	Strong
K103	-68.137	-35.132	-	0.791	Helix	Strong
K104	-67.180	-32.377	-	0.726	Helix	Strong
L105	-73.960	-23.111	-69.8	0.707	Helix	Strong
Q106	-94.207	-4.867	-66.9	0.678	Helix	Warn
A107	-	-	-	-	Loop	None