

**Biochemical and biophysical characterisation of  
the non-methylated state of FrzCD,  
the cytoplasmic receptor of the Frz pathway**

**A thesis submitted towards partial fulfillment of the requirements  
of the BS-MS Dual Degree Program**



**By**

**Yaikhomba Mutum**

**20121029**

With

**Dr. Gayathri Pananghat**

Department of Biology

Indian Institute of Science Education and Research Pune

Thesis Advisor

**Dr. M S Madhusudhan, IISER Pune**

## Certificate

This is to certify that this dissertation entitled “Biochemical and biophysical characterisation of the non-methylated state of FrzCD, the cytoplasmic receptor of the Frz pathway” towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research (IISER), Pune represents the research carried out by Yaikhomba Mutum at Indian Institute of Science Education and Research, Pune under the supervision of Dr. Gayathri Pananghat, Assistant Professor, Department of Biology, IISER Pune during the academic year 2016 - 2017.

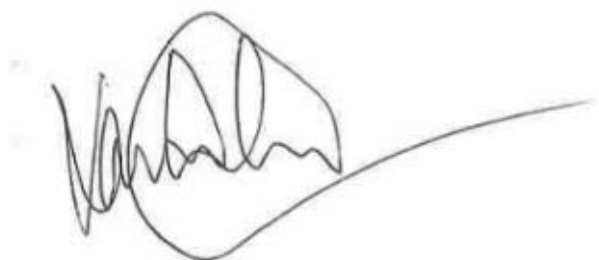


Dr. Gayathri Pananghat  
Assistant Professor  
Department of Biology  
IISER Pune

Date: 20th March 2017

## Declaration

I hereby declare that the matter embodied in the report entitled “Biochemical and biophysical characterisation of the non-methylated state of FrzCD, the cytoplasmic receptor of the Frz pathway” are the results of the investigations carried out by me at the Department of Biology, Indian Institute of Science Education and Research, Pune, under the supervision of Dr. Gayathri Pananghat and the same has not been submitted elsewhere for any other degree.

A handwritten signature in black ink, appearing to read 'Yaikhomba Mutum', with a long horizontal flourish extending to the right.

Yaikhomba Mutum

20121029

Integrated MS of The 2012 Batch

IISER Pune

Date: 20 March 2017

*This thesis is dedicated to* 

*without whose will for my freedom of exploring the other side of the world,*

*I will not be anywhere near today.*

*Also for Gai, Sai and Yai.*

## Abstract

The complex lifecycle of *Myxococcus xanthus* is facilitated by two types of motility, in addition to several other pathways that regulate its lifecycle. The *frz* pathway modulates the cellular reversal frequency and plays an important role in regulating both the adventurous and social motilities of this bacteria. The foremost enzyme in this *frz* pathway – FrzCD is a cytoplasmic chemoreceptor. Different post-translational modifications of FrzCD correspond to different states of the receptor and can activate or repress the signalling pathway resulting in modulation of the cellular reversal frequency. In this study, we sought out to characterise one of the states – the unmethylated form of the receptor, using a combination of biophysical approaches to decipher the oligomeric state of the protein. Biochemical means to dissect the protein into its constituent N and C-terminal domains to assign its structure and function, and subsequent biophysical characterisation to assess the quality of the protein sample were carried out. During this process, we discovered that the N-terminal domain of FrzCD binds to DNA, the first instance of a methyl-accepting chemosensory protein (MCP) binding to DNA.

## **Table of Contents**

Chapter 1: Introduction	Page no.
1.1 Chemotaxis in <i>E. coli</i>	5
1.2 Motility in <i>Myxococcus xanthus</i>	6
1.3 The Frizzy (Frz) pathway	7
1.4 Objectives	9
Chapter 2: Materials and Methods	
2.1 Design of FrzCD N and C-terminal domain constructs	10
2.2 Cloning of FrzCD constructs	10
2.3 Overexpression and solubility profile of FrzCD constructs	12
2.4 Protein Purification	13
2.5 Gel filtration and SEC-MALS	14
2.6 Crystallisation	16
2.7 DNA binding assays using EMSA	16
Chapter 3: Experimental results	
3.1 Cloning, purification and crystallisation trials with FrzCD-h6	18
3.2 Design and purification of FrzCD constructs	19
3.3 Oligomeric state of the protein constructs	24
3.4 Crystallisation trials of FrzCD1-86-h6	27
3.5 DNA binding experiments	28
Chapter 4: Discussion	32
Chapter 5: Conclusions and the future	36
Chapter 6: References	37

## **Abbreviations**

CH<sub>3</sub> – methyl group.

DNA – Deoxyribonucleic acid.

EDTA – Ethylenediamine tetra-acetic acid.

EMSA – Electrophoretic mobility shift assay .

Frz – Frizzy.

FrzCD-h6 – the full length protein FrzCD with a hexa-histidine tag at the C-terminus.

HAMP domain – A coiled-coil dimerisation domain commonly found in Hisdine kinase, Adenylate cyclase, Methyl accepting protein and Phosphatases.

LB - Luria Bertani.

MCP – Methyl accepting chemotaxis protein..

MPD – 2-Methyl-2,4-pentanediol.

N106 – the N-terminal domain construct FrzCD1-106-h6.

N86 – the N-terminal domain construct FrzCD1-86-h6.

Ni-NTA – Nickel-Nitralotriacetic acid.

OD<sub>600</sub> – Optical density measured at 600 nm.

PAGE – Polyacrylamide Gel Electrophoresis.

PCR – Polymerase Chain Reaction.

PEG – Polyethylene Glycol.

RF cloning – Restriction (enzyme) Free cloning.

SDS – Sodium Dodecyl Sulphate.



## List of Figures

Figure number	Figure title	Page number
1.1	The Che pathway in <i>E. coli</i> .	5
1.2	The life cycle of <i>Myxococcus xanthus</i> .	6
1.3	The Frz pathway.	7
1.4	The Phenotype of Frz mutants in development phase and FrzCD localisation in cells.	8
1.5	Effect of methylation on reversal frequency.	9
2.1	The procedure for RF cloning.	10
2.2	The PCR cycle.	12
2.3	Calibration plot for size exclusion chromatography.	15
2.4	The sitting drop method for protein crystallisation.	16
3.1	Cloning of FrzCD-h6.	18
3.2	Purification of FrzCD-h6.	18
3.3	Crystallisation of FrzCD-h6.	19
3.4	SUPERFAMILY alignment of FrzCD.	20
3.5	PROMALS3D alignment of FrzCD.	20
3.6	The truncated constructs of FrzCD.	20
3.7	Cloning of FrzCD constructs.	21
3.8	Solubility profile of C-terminal domain constructs.	22
3.9	Solubility profile of the N-terminal domain constructs.	22
3.10	Mass spectrum of the soluble constructs.	23
3.11	SEC-MALS profile of soluble constructs.	26
3.12	SEC-MALS elution profile of FrzCD1-106-h6.	26
3.13	Non-homogeneous oligomeric state of FrzCD1-106-h6.	27
3.14	Crystallisation of FrzCD1-86-h6.	28
3.15	Protein absorbance before and after ion exchange chromatography.	29
3.16	The soluble FrzCD constructs bind DNA.	29
3.17	Binding experiments of FrzCD-h6 with plasmid.	30
3.18	Binding experiments of FrzCD-h6 with decreasing DNA length.	30-31
4.1	Disordered region prediction in FrzCD using PrDOS.	32
4.2	Alignment of FrzCD with Aer2 using Promals3D.	33
4.3	Sequence alignment of HAMP1 against other HAMPs.	35

## **List of Tables**

Table number	Description	Page number
Table 2.1	Primers used for PCR amplification.	11
Table 2.2	DNA sequences for DNA binding assays.	17
Table 3.1	Summary of purified proteins.	24
Table 3.2	Oligomeric state of proteins from Size exclusion chromatography.	25
Table 3.3	Oligomeric state of proteins from SEC-MALS.	25

## Acknowledgements

Be it Manipur or in Science, one needs freedom to survive – imagine and do things constructively. Not only did I get this very generously during the last two years for whichever experiment I wanted to do, but there was always an inner conscience – Ma'am to reflect back and help me streamline my thoughts. This generosity combined with her sharp and contrasting reasoning – to see things differently when I was hell-bent on the interpretations of the result was quite insightful (and humbling!). The same is with Sai Sir and his lab, for which anything related to DNA – discussions, oligos, buffers were all just a step away. I very much recall the discussions this year about our DNA binding experiments in PAGE gels. To prove it inconclusive, when I was so insistent upon the result is another tribute to her reasoning and analysis (and her wit too!). Thanks Ma'am & Sir for everything!

This aspect of giving ample freedom so that one may learn from their own mistakes, even at the cost of running a lab when resources are meager, is something I will never forget and will carry on – both in my life as well as in science. I seek your blessings Ma'am.

Alongside conducting my experiments, I also had the opportunity to meet many people and learn of them too. I thank Dr. Radha Chauhan, Pravin Devangan and others of Prof. Shekhar Mande's lab in NCCS who were there throughout the SEC-MALS run.

There are many things I feel nostalgic here too – about learning things from scratch from people who came for my transformation experiments at 11 PM, when I joined the lab two and half years ago. This was even when despite their gut feeling that it may not work, they came for my insatiable and uncontrollable curiosity. Thanks Da Birjeet, Rajnandani, Jyoti Di, Priyanka Didi, Mahesh and Ishtiyah Bhaiya for all that you have taught me and bringing up the entire lab. I also thank my friends in IISER Pune, other members of our and Sai Sir's lab for all the help in every possible way – small or big, personal or lab related, and the very interesting discussions we had in our lab meetings. I also thank Mohan Reddy for the Mass spectrometry experiments and Siddharth Adithyan for the beautiful figures.

Thanks IISER Pune for whom I had the great fortune of meeting many people – their science and their life. Thanks Dr. Srivatsan, Prof Hotha, Jayakannan, K N Ganesh, the chemistry department for all the discussions we had and their warm welcome every time.

Towards the end, I thank my family and my well wishers who has been throughout this long journey away from home and the incredible hope and blessings you shower me always. Most importantly, my mother who has always been listening – even when I'm pitifully wrong. I feel so nostalgic now. I will stop here.

# Chapter 1.

## Introduction.

To survive, all living beings must sense the changes in the environment around it and respond appropriately. This process of chemotaxis involves sensing of nutrients, toxins, pH, oxygen level or light in the environment around it and a motile response apt for the stimulus sensed (Zhulin, I.B., 2001). This fundamental process has been well studied in *Escherichia coli* (Wadhams, G.H. et al., 2004). Here, a modified two-component system senses changes in nutrients and modulates the direction of flagellar rotation to bias the movement in accordance with the signal. Two-component systems consist of a histidine kinase signal sensor and a response regulator that can be phosphorylated by the sensor (Zhulin, I.B., 2001). The phosphorylated response regulator can then activate downstream effectors and modulate motility or bring about a change in the gene expression to alter the physiology of the cell apt for the local environment (Kirby, J.R., 2009).

### 1.1 Chemotaxis in *E. coli*.

In *E. coli*, a transmembrane homodimeric coiled coil MCP (methyl-accepting chemotaxis protein) senses the signal and conveys the signal through its conserved C-terminal signalling domain (Ferris, H.U. et al., 2014). It signals the cytoplasmic histidine kinase, CheA through the adaptor protein, CheW to catalyse a trans-phosphorylation of a conserved histidine residue in CheA (Figure 1.1). The phosphate is then relayed to the response regulator CheY. The phosphorylated CheY then binds to the FliM, the switch of the flagellar motor. This alters the rotation of some of the flagella which disassemble the flagellar bundle and causes the cell to tumble. The cell can then redirect its movement and take up a new direction where the nutrients are available.

To ensure a sustained directed movement towards increasing nutrient concentration, the sensor MCP has

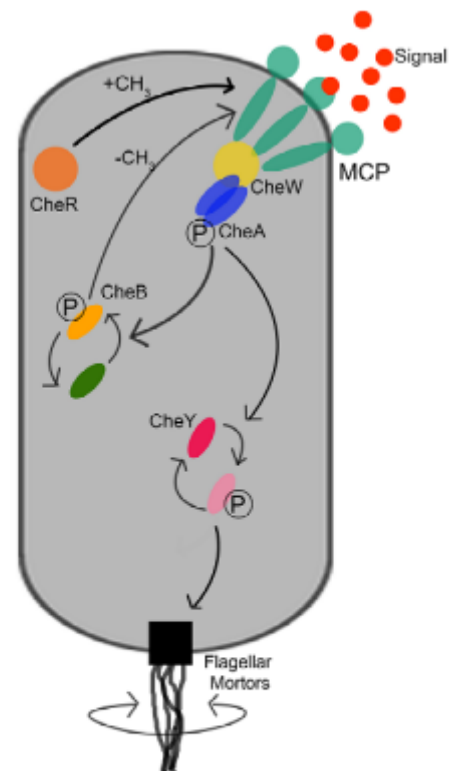


Figure 1.1: The Che pathway that directs movement in *E. coli*. Adapted from [www.chemotaxis.biology.utah.edu](http://www.chemotaxis.biology.utah.edu) and [www.2016.igem.org](http://www.2016.igem.org)

to be modulated to make sure that the flagellar bundle doesn't disassemble. As a corollary, the bacteria must tumble at a higher rate when nutrient concentrations are decreasing and choose a direction that is high in nutrients. This is achieved by bringing the current stimulus response back to the pre-stimulus level, a process called adaptation. The sensing of a gradient is carried out through adaptation, and is brought about by post-translational methylation, in response to the external signal, of certain glutamate residues of the sensor MCP in the C-terminal signalling domain. CheR is a constitutively active methyltransferase. A methylated MCP receptor is more sensitive to decreasing nutrient concentration and activates the signalling pathway. During this process, the histidine kinase also phosphorylates CheB, a methylesterase that can remove methyl groups from the MCP. The demethylated MCP is now 'desensitised' and not active in the existing amount of signal. A nutrient concentration lower than the current one is now required to activate the receptor and bring about tumbling. For increasing concentrations of the nutrient, the signalling pathway is not activated, hence the bacteria swims in the same direction for a longer time period (Wadhams, G.H. et al., 2004).

Though, the chemotaxis mechanism in *E. coli* has been well studied and understood, the variations in motility regulation in other bacteria have been not completely understood, such as the *Myxococcus sp.*, which has a complex life cycle (Figure 1.2), different types of motility and can move rhythmically in groups (Kaiser, D., 2003).

## 1.2 Motility in *Myxococcus xanthus*.

*Myxococcus xanthus* is a predatory gram-negative social bacteria and has two types of movements - the adventurous and the social motility (Zusman, D.R. et al., 2007). While social motility, responsible for coordinated group movements in the bacteria, is established by synchronous extensions and retractions of the type IV pilus from the leading pole of the bacteria, gliding or A-motility is powered by molecular motors that lie along the helical cytoskeletal track of the rod-shaped bacterium (Nan, B. et al, 2010, 2011). A periodic reversal in its cellular polarity occurring every 6 – 13 minutes is a feature of both these movements

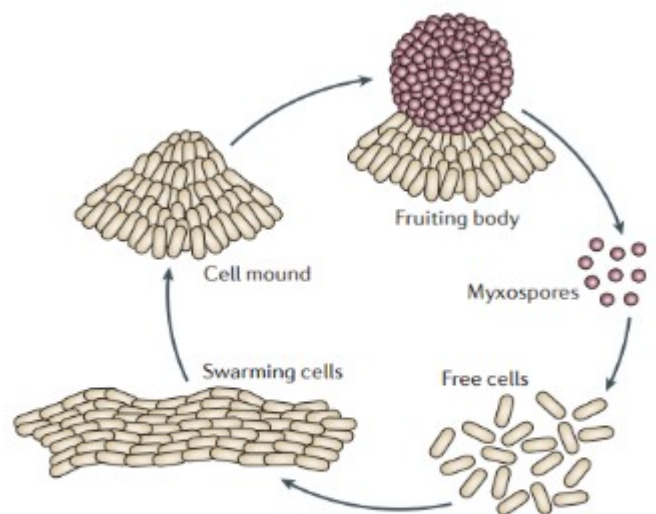


Figure 1.2: The lifecycle of myxococcus xanthus. Adapted from Westra, E.R. et al., 2014.

(McBride, M.J. et al., 1989, Bustamante, V.H. et al. 2004). When starved on a nutrient poor medium, it forms spores that are heat and chemically resistant and the development phase of the bacteria ensues. Once nutrients are available, the spores can then hatch out and the vegetative cycle of the bacterium continues. The reversal frequency in *Myxococcus xanthus* is modulated by a chemosensory pathway very similar to the *E. coli* chemosensory pathway described above (Zusman, D.R. et al., 2007). The pathway responsible for this consists of the Frz pathway proteins, comprising of FrzCD, FrzA, FrzE, and FrzZ, that are equivalents of MCP, CheW, CheA-Y and CheY-Y respectively of the *E.coli* Che pathway (Figure 1.1 and 1.3, Zusman, D.R., et al., 2007).

### 1.3. The Frizzy (Frz) pathway.

The components of the Frz pathway in *Myxococcus xanthus* closely resembles this established paradigm of the Che pathway (Figure 1.1, Kaiser, D., 2003). In addition to regulating motility, the Frz pathway also has established roles in vegetative cycle of the bacteria – it controls coordinated group movements like swarming, regulates the reversal frequency in its movement and in the development cycle of the bacteria (Bustamante, V.H. et al, 2004). The deletion of components in the Frz pathway increases the cellular reversal frequency, disrupts its group behaviour like swarming in the vegetative cycle and can no longer form fruiting bodies (Weinberg, R.A. et al., 1989, McBride, M.J. et al., 1992, Bustamante, V.H. et al., 2004). Studying how movements are regulated and co-ordinated in this bacterium in different environmental contexts is core to understanding the features the are responsible for motility in general. The Frz pathway of this bacteria is thus a model system to understand different motilities, its regulation and co-ordination in groups and in environment specific contexts.

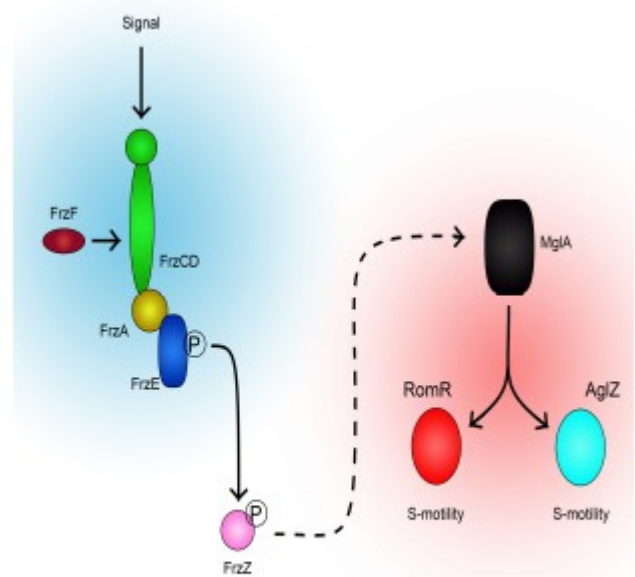


Figure 1.3: The Frz pathway.  
Adapted from Zusman, D.R. et al., 2007

The foremost of the enzymes in the Frz pathway, FrzCD is the homolog resembling the MCP of the Che pathway (McBride, M.J., et al., 1992). Though the signal it senses is not known, the C-terminal domain of FrzCD from residues 130-417 shares about 30%

sequence identity with the Tsr, serine receptor of *E. coli* and 40% identity with the aspartate receptor Tar, of *Salmonella typhimurium* (Mc Bride, M.J. et al., 1989). Though its N-terminal domain is of an unknown fold, sequence analysis predicts that the C-terminal domain is a coiled coil.

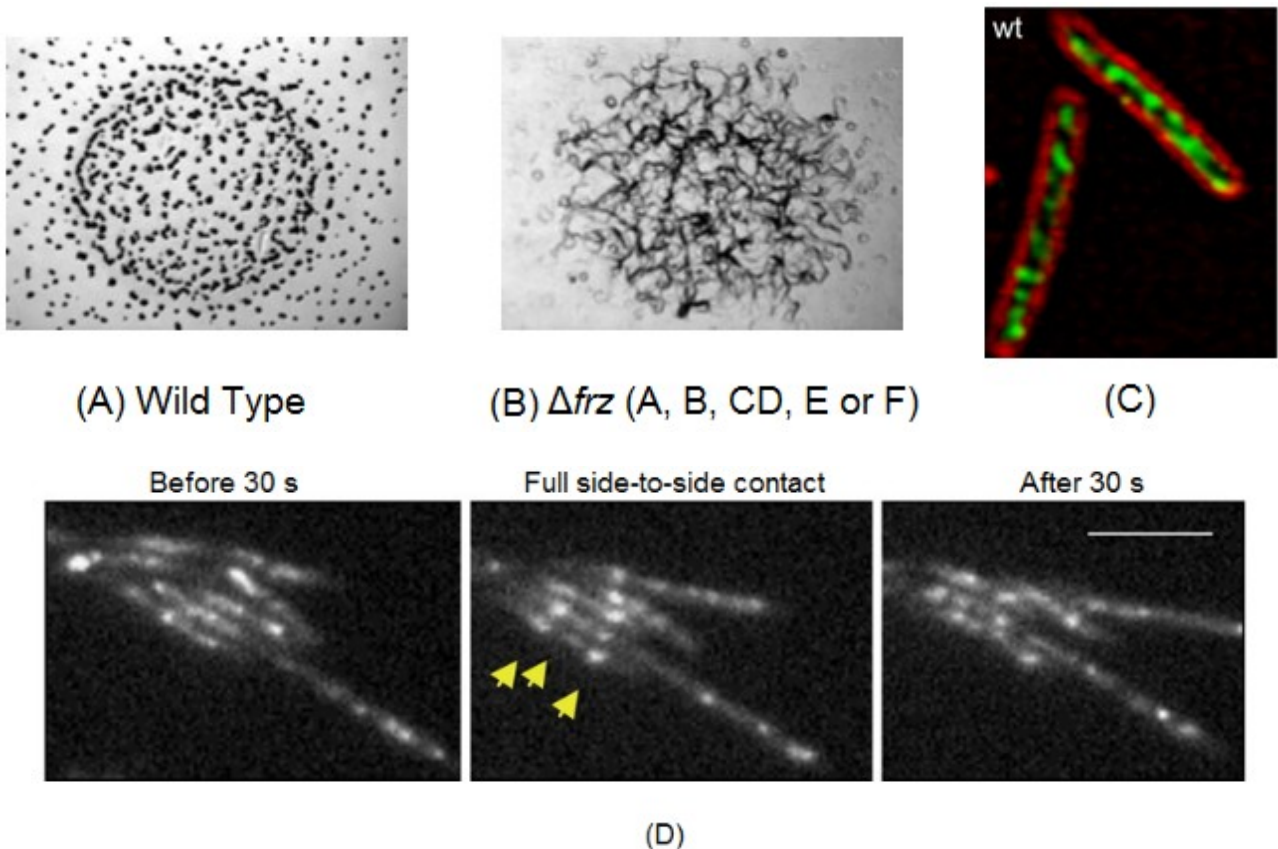


Figure 1.4: Role of Frz in *Myxococcus xanthus* (A) Morphology in wild type Frz (B) Frz deletion mutants fail to form fruiting bodies. (Adapted from Bustamante, V.H. et al., 2004)

(C) In vivo fluorescence microscopy of Frz clusters which align when cells are close to each other (D) (Adapted from Mauriello, E.M.F. et al., 2009).

FrzCD is a cytoplasmic MCP and is observed to be localised in a helical pattern in the cytoplasm (Figure 1.4C). These dynamic helical clusters of FrzCD align with each other when the cells are in direct side-side contact with each other (Figure 1.4D, Mauriello, E.M.F. et al., 2009).

Earlier studies have also shown that the protein is methylated in vivo by the methyltransferase FrzF and the protein exists in a variety of methylated states (Figure 1.5, McBride, M.J. et al., 1989 and McCleary et al., 1990). However, the most interesting feature of this protein is that methylation of different sites on this receptor can modulate



the cellular reversal frequency by increasing or decreasing it (Figure 1.5B, Astling, D.P. et al., 2006, Scott, A.E. et al., 2008).

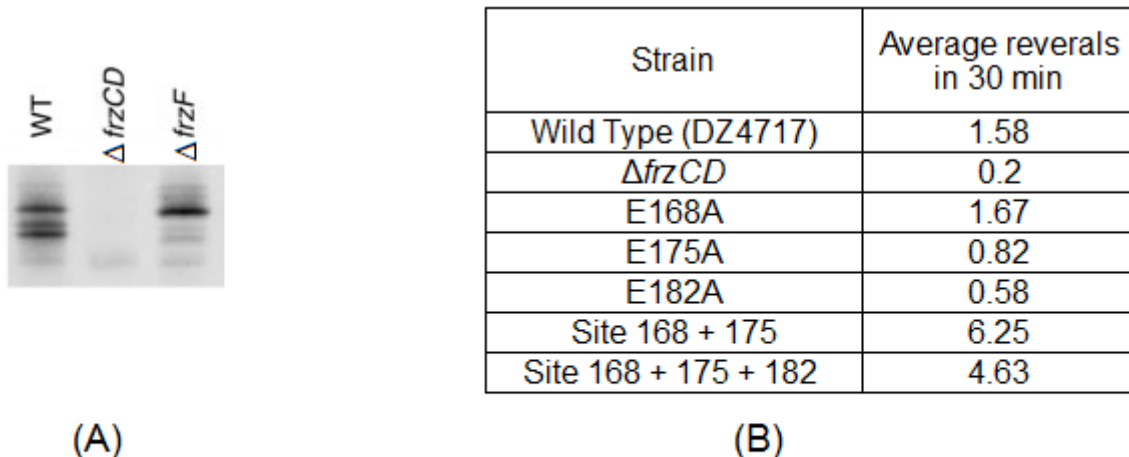


Figure 1.5: (A) FrzCD is methylated and exists in a combination of methylated states in vivo (Bustamante, V.H. et al., 2004). (B) Effect of alanine mutations of FrzCD methylation sites on cellular reversal frequency (Adapted from Scott, A.E. et al., 2008).

## 1.4 Objectives.

How methylation at different sites of FrzCD changes the state of the receptor is unknown, though it has an obvious phenotypic effect, reflected as changes in the reversal frequency. The main goal of the project is to understand the FrzCD signalling pathway by using a combination of biochemical and biophysical techniques. For instance, does methylation of a specific glutamate residue change the conformation of the receptor or the oligomeric state of the FrzCD and change the biochemical affinity with its downstream partners? Finding the structure of the receptor in its non-methylated form is the first step in this project. This will provide us a standard against which the different conformations of the protein in different methylated contexts can be compared with.

The main objectives of this project are:

- biophysical and biochemical characterisation of the cytoplasmic receptor in its native, non-methylated form.
- obtain a crystallisable construct, in case the full length protein is not amenable for crystallization.
- find the structure and function of the unknown N-terminal domain of FrzCD.



## 2. Materials and Methods

### 2.1 Design of FrzCD N and C-terminal domain constructs.

The FrzCD sequence was analysed using online servers – Superfamily (<https://www.supfam.org/SUPERFAMILY>, Wilson, D. et al., 2009), Promals3D ([www.prodata.swmed.edu/promals3d/promals3d.php](http://www.prodata.swmed.edu/promals3d/promals3d.php), Pei, J. et al., 2007) and pcoils (<https://toolkit.tuebingen.mpg.de/pcoils>, Alva, V. et al., 2016) to get an idea of the domains of FrzCD against available structures of its homologs. Subsequently, based on the alignment, the constructs were designed.

### 2.2 Cloning of FrzCD constructs.

Cloning involves amplification of the gene of interest from the genomic DNA, its insertion into an appropriate locus onto the vector and then extracting plasmids from the bacteria. The required gene was amplified from the template (genomic DNA of *Myxococcus xanthus* for the full length *frzCD* or a pHis17 vector containing the *frzCD* gene insert for amplifying the truncated constructs). The gene was then introduced into the pHis17 vector using restriction-free cloning (Figure 2.1). In this procedure, the amplified PCR product, containing flanking vector sequence on both sides, can act as a mega-primer and extend over the vector during a PCR reaction. This amplification process generated a plasmid containing a nick in both the strands.

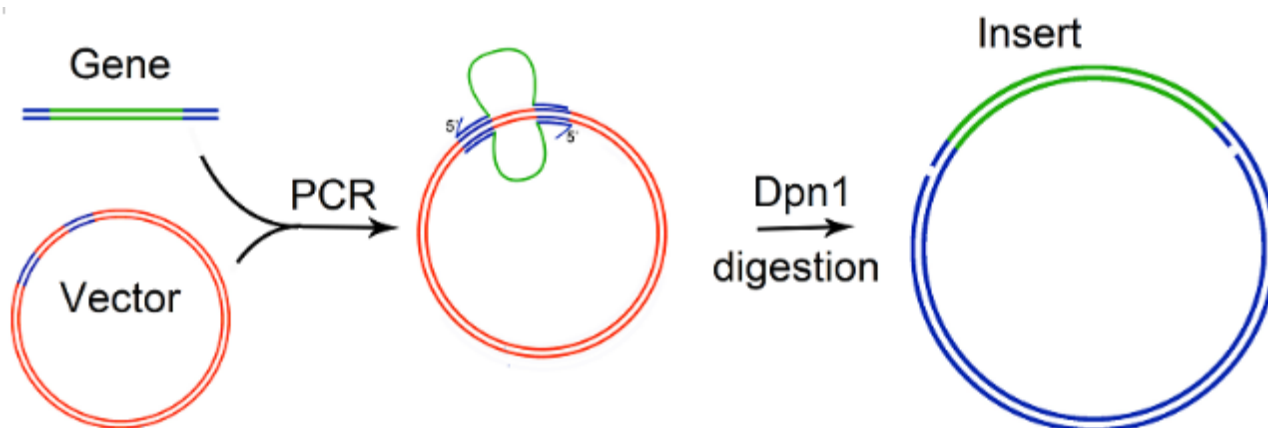


Figure 2.1: The restriction enzyme free method of gene insertion into a vector. Adapted from [www.rf-cloning.org](http://www.rf-cloning.org)

The template vector is then digested by Dpn1, which specifically cleaves only methylated DNA. After overnight Dpn1 digestion, NEB electro-competent cells were electroporated and plated onto LB (Luria Bertani) agar plates containing 0.1 mg/ml Ampicillin. Single colonies were then picked up and grown in LB broth containing 0.1 mg/ml of Ampicillin for 18 hours. The plasmids were then extracted by the alkaline lysis method. Subsequently, a

restriction digestion of the plasmids was carried out by incubation for 3 hours to confirm the insert size. The gene contains restriction sites of Nde1 and BamH1 on either side. The positive clones were then finally confirmed by sequencing.

A typical 50 µl PCR (Polymerase Chain Reaction) for amplifying the gene is given below while table 2.1 lists the primers used for generating the relevant constructs.

- Template DNA: ~200 ng.
- Forward and reverse primers: 1 µl of 20 µM stock each.
- AccuPrime™ Pfx: 0.5 µl.
- 10X AccuPrime™ Pfx Buffer: 5 µl.
- AccuPrime™ Pfu DNA Polymerase: 0.5 µl.
- De-ionised water/ Milli-Q®: make up the volume to 50 µl.

Table 2.1: Primers used for PCR amplification of the constructs.

<b>Primer</b>	<b>Primer Sequence (5' -----&gt; 3')</b>	<b>Length</b>
<i>frzCD-f</i>	GTTTAACTTTAAGAAGGAGATATACATATGTC CCTGGACACCCCAACG	49
<i>frzCD.h6-r</i>	GCGGTTCAAGGCCGACGGATCCCATCATCAT CATCATCATAAAAGC	47
<i>frzCD1-86-r</i>	GCTTTTAATGATGATGATGATGATGGGATCCA TGCTTGCGGTGCTCGGAG	50
<i>frzCD1-106-r</i>	GCTTTTAATGATGATGATGATGATGGGATCC GCCCTCGCGCACCAGGCCG	50
<i>frzCD84-415-f</i>	GTTTAACTTTAAGAAGGAGATATACATATGCG CAAGCATGTGGCGGCG	48
<i>frzCD107-415-f</i>	GTTTAACTTTAAGAAGGAGATATACATATGGA CCTGTCCCGGTGGAACAC	50
<i>t7-promoter-f</i>	TAATACGACTCACTATAGGG	20
<i>t7-terminator-r</i>	GCTAGTTATTGCTCAGCGG	19

For RF (restriction free) cloning, the reaction mixture contains

- Template DNA: ~500 ng.
- Gene insert: ~1000 ng.
- AccuPrime™ Pfx: 0.5 µl.
- 10X AccuPrime™ Pfx Buffer: 5 µl.
- AccuPrime™ Pfu DNA Polymerase: 0.5 µl.
- De-ionised water/ Milli-Q®: make up the volume to 50 µl.

A typical PCR cycle involves the following heating and cooling cycle to activate the enzyme, amplify and anneal the amplified gene (Figure 2.2).

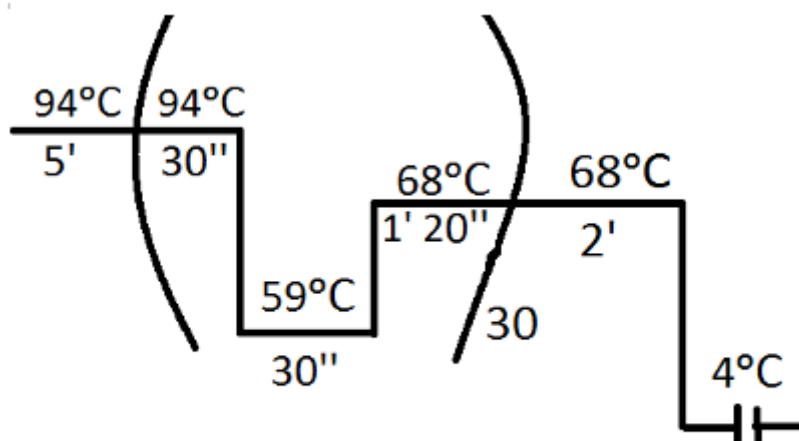


Figure 2.2: The PCR cycle employed here.  
Note: 20 cycles for RF cloning.

Components for Dpn1 digestion:

- 9.5  $\mu\text{l}$  of RF-PCR product
- 0.5  $\mu\text{l}$  of Dpn1 enzyme.

After electroporation with the Dpn1 digested PCR product, cells were immediately revived using 200  $\mu\text{l}$  of 2X LB and incubated for an hour at 37°C. To confirm the correct gene insert into the pHis17 vector, restriction digestion using the enzymes Nde1 and BamH1 was carried out, which will result in a release of the insert corresponding to the target gene size.

Restriction digestion reaction mixture:

- plasmid: ~100 ng.
- Nde1: 0.3  $\mu\text{l}$
- BamH1: 0.3  $\mu\text{l}$
- 10X NEB Cutsmart Buffer: 1  $\mu\text{l}$
- Milli-Q™: make up the volume to 10  $\mu\text{l}$ .

This reaction mixture after 3 hours at 37°C was loaded onto a 0.8% agarose gel and run for 40 minutes at 110 V.

### 2.3 Overexpression and solubility profile of FrzCD constructs.

After confirming the recombinant plasmid by DNA sequencing, BL21AI *E. coli* cells were transformed by heat-shock treatment at 42 °C using ~100 ng of the recombinant plasmid. Soon, the cells were revived using 200  $\mu\text{l}$  of 1XLB and incubated at 37° C for an hour.

Later, they were plated onto LB Agar containing 100 µg/ul of ampicillin and grown at 37° C for 16 hours.

To check the solubility of the protein, 20 - 25 BL21AI single colonies were picked up and grown in a test tube containing 13 ml of LB broth with 100 µg/ul of ampicillin at 37°C till the growth reached an optical density of 0.8 – 1, monitored at 600 nm. 6 ml of the cells were transferred into a fresh sterile test tube, induced with 0.2% L-arabinose and then grown at 30° C for 5 hours with constant shaking of 180 rpm. The remaining 6 ml of the cells are left as a control for the experiment and not induced.

After pelleting down the cells at 4,000 g, cells were resuspended in 600 µl of a buffer of pH 8 containing 50 mM Tris and 200 mM NaCl in a 1.5 ml eppendorf. To check the solubility in other buffers, cells were resuspended in the lysis buffer of requisite pH and salt. While salt used here is NaCl whose concentration could also be varied, a buffer of pH 6.5 was maintained using sodium cacodylate. Cells were lysed at 4° C by sonication at 60% amplitude (81 W, 20 kHz) for one minute in a 1" on and a 3" off cycle. The lysate was then spun at 21,000 g for 10 minutes at 4° C to pellet all insoluble parts of the lysate. A 10 µl aliquot of the fractions are taken just after sonication and after spinning the lysate down. Later, induced and uninduced fractions were compared by running the heated sample in an SDS-PAGE.

## **2.4 Protein Purification.**

All procedures for protein purification were carried out at 4°C unless mentioned otherwise. 500 ml of culture was resuspended in 35 ml of Buffer A (200 mM NaCl, 50 mM Tris pH 8) and lysed by sonication for 5 minutes (60 % amplitude (81 W, 20 kHz) in a 1" on and 3" off cycle). The cell lysate was immediately spun at 39,191 g for an hour to remove membrane fractions and other insoluble particles. The supernatant containing the soluble fraction of the lysate was then loaded onto a 5 ml Ni-NTA column at 5 ml/min allowing only the hexahistidine tagged protein to bind to the column. The column was then given an extensive wash of 8 column volumes (40 ml) with 2% buffer B (containing 10 mM imidazole; buffer B contains 500 mM imidazole in addition to all components of buffer A) and 5% buffer B (containing 25 mM imidazole) to remove other non-specific impurities that might have bound the column. The protein was then eluted out of the column by increasing the imidazole concentration. The imidazole concentration in the column was serially increased from 10 %B, 20 %B, 50 %B to 100 %B and 30 ml of each were collected as 5 ml fractions. The fractions containing the protein were then pooled together and dialysed for 3 hours in a buffer containing 25 mM NaCl, 50 mM Tris pH 8 and 1 mM EDTA to lower the high salt

and imidazole content of the protein fractions. The presence of low salt is essential for the protein to bind the ion exchange column. Ion exchange chromatography also removes additional protein impurities and bound DNA. The protein after dialysis was spun at 39,191 g for 30 minutes and loaded onto an ion exchange column using 50 ml Superloop™ and eluted using increasing amount of salt (up to 1 M NaCl). The fractions containing the protein were then concentrated using a concentrator of 3 kDa cutoff. The proteins were then flash frozen using liquid nitrogen in thin walled PCR tubes in aliquots of 20 - 30 µl and stored at -81°C.

The protein concentration was estimated by Bradford method. Comparing the absorbance of the purified protein at 595 nm with a standard (Bovine serum albumin curve) yields us an estimate of the protein concentration. Later, the quality of the protein was confirmed through both mass spectrometry (to check proteolytic degradation) and size exclusion chromatography (to check that the protein is a homogeneous sample and are not aggregated).

### **Note:**

1) The type of ion exchange columns and the pH of the buffer used depends on the pI of the protein. For FrzCD-h6 (pI of 5.4) and FrzCD:1-106-h6 (pI of 7), the anion exchange column (MonoQ 4.6, GE healthcare) was used and for FrzCD:1-86-h6 (pI of 9.6) a cation exchange column (MonoS 4.6, GE healthcare) was used with a binding running buffer at pH 8.

## **2.5 Gel filtration and SEC-MALS.**

Globular proteins elute on the basis of size in size exclusion chromatography. Bigger proteins elute from the column earlier than smaller ones. These columns also have an exclusion limit - beyond which the molecular weights cannot be resolved. Based on this a standard graph of  $\text{Log}M_w$  can be plotted against the partition coefficient -

$$K_{av} = \frac{V_e - V_0}{V_c - V_0}$$

Here  $V_c$  is the column volume,  $V_0$  is the void volume of the column and  $V_e$  is the elution volume of the protein.

The molecular weight standards used for this runs were Blue Dextran ( $M_w = 2000,000$ ), Thyroglobulin ( $M_w = 670,000$ ), Apoferritin ( $M_w = 440,000$ ), Beta-amylase ( $M_w = 200,000$ ),

Alcohol dehydrogenase ( $M_w = 150,000$ ), Bovine serum albumin ( $M_w = 67,000$ ) Carbonic anhydrase ( $M_w = 29,000$ ) and Cytochrome C ( $M_w = 13,600$ ). From this, a calibration plot was calculated. The molecular weight of the protein constructs can then be estimated if the elution volume is known.

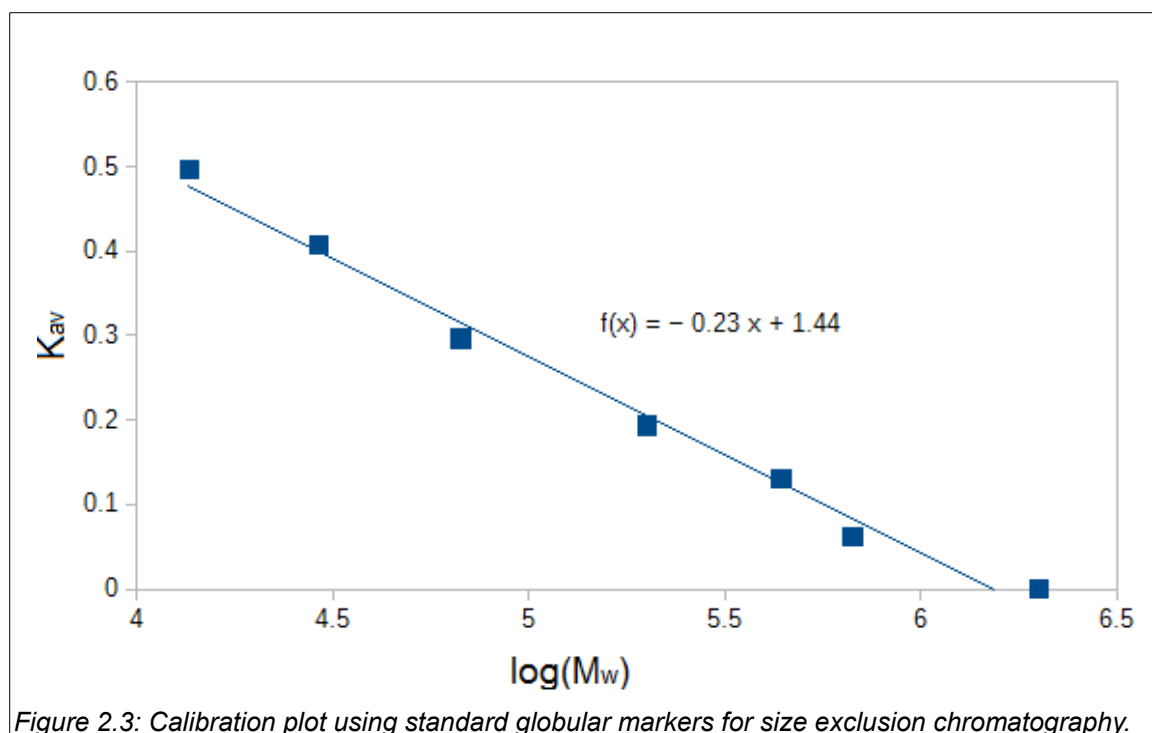


Figure 2.3: Calibration plot using standard globular markers for size exclusion chromatography.

Before loading onto the size exclusion column, the column was equilibrated with buffer of 50 mM Tris pH 8 containing 50 mM NaCl at 4°C. The protein was diluted to 100  $\mu$ l with the same running buffer to 10 mg/ml and spun at 21,000 g for 10 minutes. Later the supernatant was pipetted out and injected into the column. The run was carried out at 0.5 ml/min and monitored by UV at 230 nm, 280 nm and 260 nm.

The procedure is the same for a SEC-MALS run (Size exclusion chromatography with Multi angle light scattering as a detector), except that the detectors used to detect protein elution are different. In this technique, a size exclusion column separates a mixture of protein sample into its constituent oligomers. The homogeneous protein molecules are then passed through a polarised laser light at 658 nm and the intensity of the light scattered is measured at different angles. By measuring the intensity of the scattered light and the change in refractive index of the sample, the solution molecular weight of the sample can be calculated. The advantage of this detection technique is that it assumes nothing about the shape of the sample for estimating molecular mass unlike size exclusion chromatography which is a technique applicable only for globular proteins.

## 2.6 Crystallisation.

Crystallisation trials were carried out using the sitting drop method of vapour diffusion. In this method, a saturated protein solution (10 mg/ml) is mixed with a buffered solution containing precipitants and is allowed to equilibrate with a reservoir containing the buffered solution. Because the reservoir has a lower vapour pressure, the drop slowly evaporates (Figure 2.4). During this process, if conditions are just right, then the protein can form crystals.

For initial screening, the sitting drop method of crystallisation was carried out in 96 (8 X 12) well plates. Here, the reservoir containing 50  $\mu$ l of the condition is equilibrated with 200 nl of 1:1 v/v protein : condition and kept at 18°C in a vibration-free incubator. The plates were observed everyday for the first week and then once every week from the second week on.

To improve upon the condition hits, optimisation conditions were set up for sitting drop in 48 well plates (6 X 8). The reservoir volume was 100  $\mu$ l and a 1  $\mu$ l drop containing 1:1 v/v (protein : condition) is equilibrated at 18°C in a vibration-

free incubator. The plates were observed everyday for the first week and then once every week from the second week onwards.

All the stock solutions used for crystallisation were filtered with 0.2  $\mu$ m filter into a fresh sterilized falcon. The protein to be used for crystallisation was diluted accordingly with a buffer of pH 8 containing 50 mM Tris and 50 mM NaCl and spun at 21,000 g for 20 minutes at 4°C and the supernatant transferred onto a fresh tube. This supernatant was used for crystallisation.

## 2.7 DNA binding assays using EMSA.

EMSA (electrophoretic mobility shift assay) is a technique to visualise the amount of protein bound to the DNA in a reaction mixture. It relies on the principle that the mobility of the protein-DNA complex during gel electrophoresis is retarded compared to the free DNA

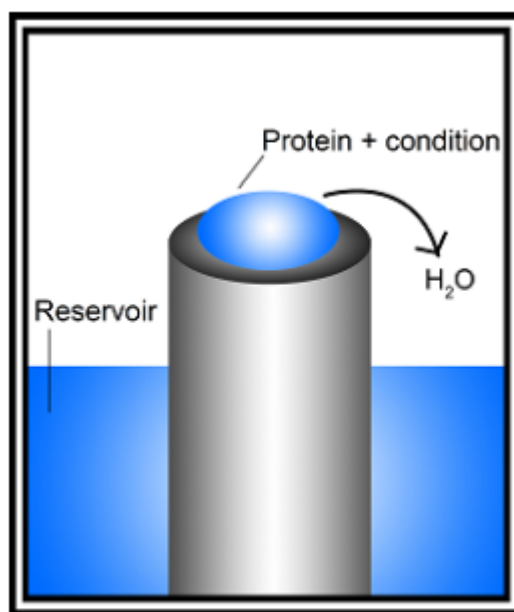


Figure 2.4: The sitting drop method of protein crystallisation. Adapted from [http://hamptonresearch.com/documents/growth\\_101/4.pdf](http://hamptonresearch.com/documents/growth_101/4.pdf)

and this can be used to assess the relative binding affinity of the protein-DNA complex (Hellman, L.M. et al., 2007). DNA binding assays were carried out in a Mg<sup>2+</sup> free buffer. It is a buffer of pH 7.4 containing 50 mM Tris, 50 mM NaCl, 1 mM dithiothreitol and 10% glycerol. Here, after adding the DNA and the protein in the required amounts, the reaction mixture was incubated at 25°C at for 20 minutes and loaded onto an (0.8 - 2%) agarose gel with 6X BPB (a loading dye) and run for 60 minutes at 110 V to see if there is a shift of the free DNA band. Shorter DNA oligos (below 147 bp) had to be run at higher gel percentages to resolve them. *Myxococcus xanthus* genome is GC rich. Thus, wherever possible, we've tried to use a GC rich sequence for binding assays. The sequences of the DNA oligo used are tabulated (Table 2.2).

Table 2.2: The different DNA sequences used for binding assays with FrzCD-h6.

ds DNA used	Length	Sequence (5' ---> 3')	GC content (%)
Oligo8	8	GCGATCGC	75
Oligo35	35	GTCACCTGCTCTAGCTAATAGACTGAGCCGAGGTG	54
Oligo70	69	CTTGCAGTAGAGCTGACCATGATTACGCCATCAGCAGCTCCAGGTCGTA CCTCCAGCTACCAATCCCCG	57
Oligo147	147	CTGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCT AGCACCGCTTAAACGCACGTACGCGCTGTCCCCCGCGTTTTAACCGCCA AGGGGATTACTCCCTAGTCTCCAGGCACGTGTGAGATATATACATCCTGT	56
Oligo235	235	GGTGATGTACGAAGAGGAGTTCACCATAATCAACGCCGTTTGCGACCG GCTTACCAAGGACGCGAACGCGAAGGTGGTCTTCCTCGTCGACAAGAA CGGGCAGCTCATCTCCTCCGCGGGTCAGACGCAGAACATCGACACCAC GTCACTGGCCTCGCTGACGGCCGGTAACGTGGCCGCGATGGGTGGCCT GGCCAAGCTGATTAGGGAGAACGAGTTCCTCAACCAGTTCAC	60
Oligo316	316	GTTTAACTTTAAGAAGGAGATATACATATGCGCAAGCATGTGGCGGCGC AGGCGGTGGAAGGCATCCTCGCCGGCGTCCAGGAGACGAGCGACGCG GCCCGCGTCATCAACCTCGCCACGCAGCAGCAGCGCACGGCCACCGA GCAGGTGGTGGCCTCCATGGCGGAGATCGAGGACGTGACGCGCCAGA CGACGCAGGCCTCGAAGCAGGCCACGGGCGCGGCCGCGGAGCTGAC GCAGTTGGCCGGACGGCTGGCCGAGCTCATCAAGCGGTTCAAGGCCG ACGGATCCCATCATCATCATCATTAATAAAGC	64
Oligo1.3 kbp	1270	frzCD gene	68
Linearised plasmid	3851	Nde1 digested pHis17 vector containing the FrzCD gene	56
Circular Plasmid	3851	pHis17 vector containing the FrzCD gene	56



### 3. Experimental results

#### 3.1 Cloning, purification and crystallisation trials with FrzCD-h6.

After successfully cloning the FrzCD-h6 gene into the pHis17 vector (Figure 3.1), the protein was overexpressed in BL21AI cells and purified using Ni-NTA and passed through the anion exchange column to remove nucleotide and other additional impurities (Figure 3.2). Size exclusion chromatography confirmed that the protein is not in the aggregated form. However, the initial crystallisation screen of 1440 conditions using the sitting drop method of crystallisation was not successful. The protein precipitated in most conditions or clear drops were observed (Figure 3.3).

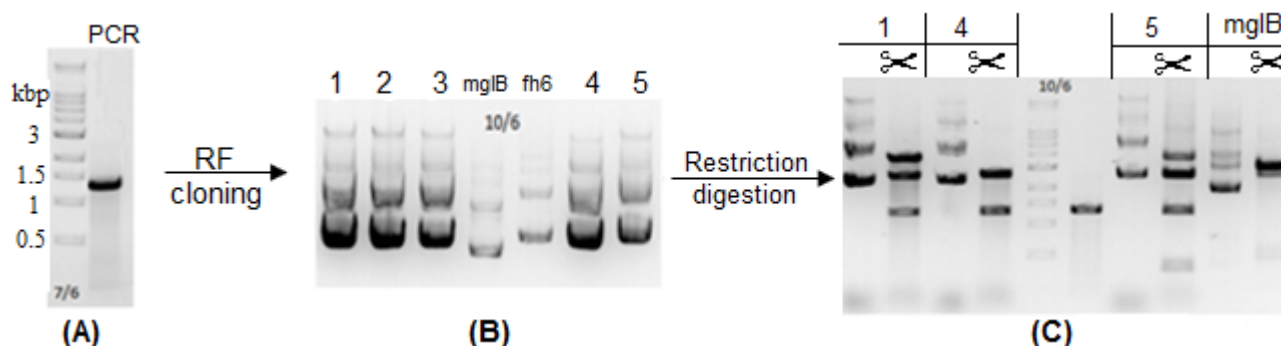


Figure 3.1: Cloning of *frzCD* gene into the *pHis17* vector using restriction-free cloning first by amplification from the genome (a), inserting the gene into the *pHis17* vector (b) and eventually confirmed by restriction digestion (c). Lane 5 of subfigure (c) is the Supermix DNA ladder and lane 6 is the amplified *frzCD* gene used for reference. *mgIB* is the plasmid used as a template here, a gene insert size of 500 bp.

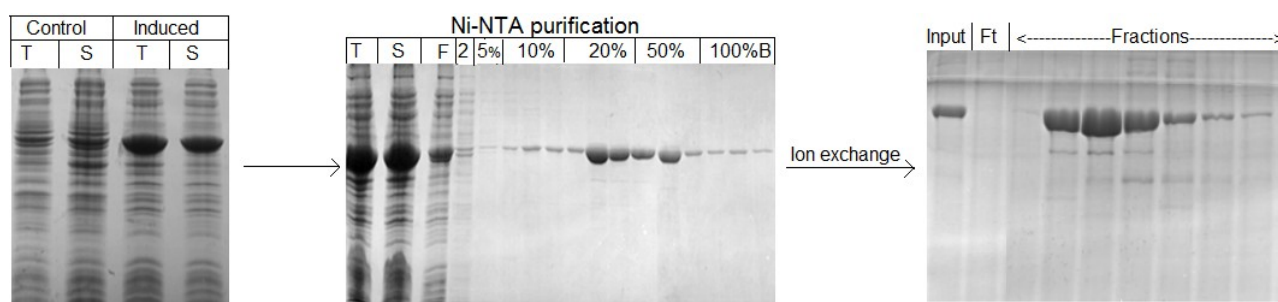


Figure 3.2 The purification profile of FrzCD. Here the total bacterial lysate (T) is spun and the supernatant (S) is loaded onto the Ni-NTA column. Flow through (F) is collected and reloaded again to the column. For ion exchange, the dialysed protein containing low salt is the input and the flow through of the column (Ft) is also analysed to see if the protein has been bound to the column.

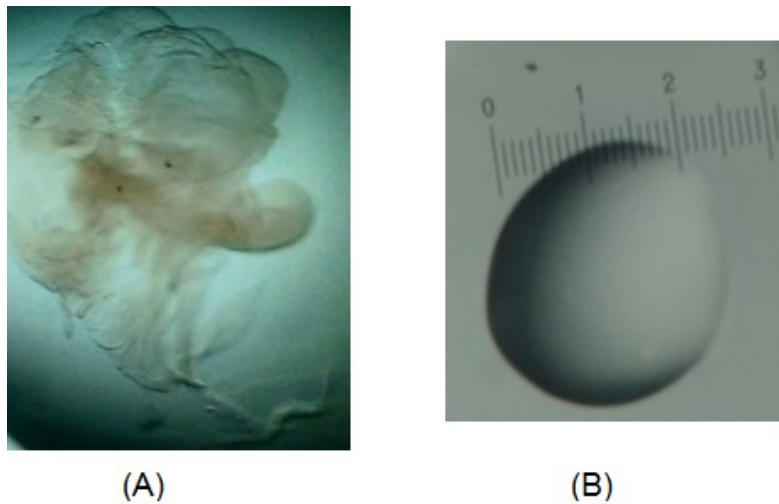


Figure 3.3: Crystallisation of FrzCD was not successful – (a) precipitate and (b) a clear drop during the protein crystallisation.

### 3.2 Design and purification of FrzCD constructs.

A superfamily search of the FrzCD sequence assigned the stretch of 106-417 sequence of FrzCD to the MCP signalling domain with pdb id: 2ch7 (cytoplasmic domain of an MCP from *Thermotoga maritima*) as its closest structural homolog (Figure 3.4). To get a better idea of domains in FrzCD, available structures of MCPs (obtained by PSI-BLAST in the pdb database) were aligned with FrzCD using PROMALS3D, a structure based sequence alignment software (Figure 3.5). It predicted a stretch of 20 more residues in the N-terminal domain (86-106) that aligned well with another MCP pdb id: 3zx6 (C-terminal domain of the serine receptor in *E. coli* fused with a HAMP domain). The protein was thus truncated at 2 points - before and after the helix, as shown in the figure. Together, these generated 4 constructs - 2 N-terminal domain constructs and two C-terminal domain constructs (Figure 3.6).

- N terminal mutants - FrzCD:1-86-h6 and FrzCD:1-106-h6 (denoted as N86 and N106 respectively in the figure, h6 denotes the Histidine tag at the C terminus).
- C-terminal mutants - FrzCD:84-417-h6 and FrzCD:107-417-h6 (denoted as C84 and C107 respectively in the figure, h6 denotes the Histidine tag at the C terminus).

```

FrzCD: 107 .....DLSRWNTTTEDPQLGPLLEGFGKVIETLRTFVREIneaalrlessanqvlaasTQHETSS
2ch7: 1      kdvqtetfsv-----AESIEEISKANEEITNQLLGISKEM.....DNISTRI

FrzCD: 167 TEQAAAIHETTATMEELKHASAQIAENAGSVARVAEETLGAARAGRGAIGEFIQAMQQIRSDGVAVADSI
2ch7: 43    ESISASVQETTAGEEISSATKNIAADSAQQAASFADQSTQLAKEAGDALKKVIEWTRMISNSAKDVERVV

FrzCD: 237 AKLSKRVERIGTVVEVIDEIADRSDLLALNAALEGSRAGEAGKGFSSIVAAEMRRLAENVLDSTKEIKNLI
2ch7: 113   ESFQKGAEIITSFVETINAIQAEQTNLLALNAAIEAARAGEAGRGFAVVADEIRKLAEESSQQAENVRRVV

FrzCD: 306 TEIREATAAAAGAAEASKSATESGEKLGAVAAQAVEGILAGVQETSDAARVINLATQQORTATEQVVASM
2ch7: 183   NEIRSIADAGKVSSEITARVEEGTKLADEADEKLNISIVGAVERINEMLQNIAAAIEEQNAAVDEITAM

FrzCD: 378 AEIEDVTRQTTQASKQATGAAAEITQLAGRLAELIKRFK-----
2ch7: 253   TENAKNAEEITNSVKEVNRARLQEISASTEETSRVQTIQRENVQMLKEIVARYK

```

Figure 3.4: The result of SUPERFAMILY search for FrzCD sequence. ([www.supfam.org/SUPERFAMILY](http://www.supfam.org/SUPERFAMILY))

```

2ch7 -----
FrzCD 1  MSLDTPNEKPAGKARARKAPASKAGATNAASTSSSTKAITDTLLTVLSGNLQARVPKELVGE SGVELAHL 70
3zx6 -----

                ↓                ↓

2ch7 1  -----GSHMKDVQTETFSVAESIEE 20
FrzCD 71 LNQVLDQFAASEHRKHVAAQEIIDQALDALIGLVREGDLSRWNTTTEDEHQLGPLLEGFGKVIETLRTFVRE 140
3zx6 1  -----TITRPIIEL-SNTVDKIAEGNLEAEVPHQN-----RADEIGILAKS 40
consensus AA .....h.pl...hpc
consensus ss      hhhhhhhh hhhhhhhh eee hhhhhhhhhhhhhhhhhhhhh

```

Figure 3.5: PROMALS3D alignment of FrzCD structures against available structures of the C-terminal domain of MCP ([www.prodata.swmed.edu/promals3d/promals3d.php](http://www.prodata.swmed.edu/promals3d/promals3d.php)).

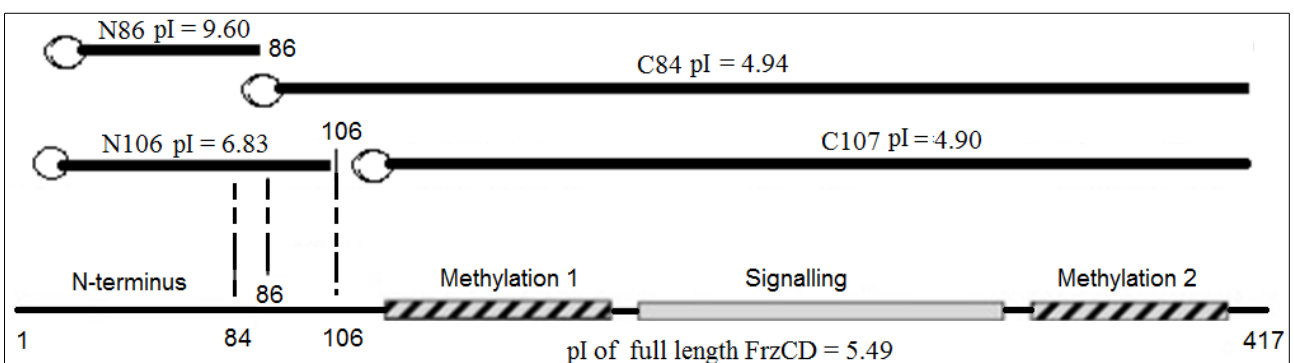


Figure 3.6: The N and C terminal domain constructs of FrzCD. Figure adapted from Bustamante, V.H. et al., 2004.

It was also conspicuous that the pI of the protein shifted from basic (9.6) to acidic (4.9) as we move from the N to the C-terminal domain. This later, along with other experimental data motivated us to carry out DNA binding assays as would be presented later sections.

After amplifying the required constructs from the FrzCD gene, these genes were introduced into the pHis17 vector by restriction-free cloning (Figure 3.7).

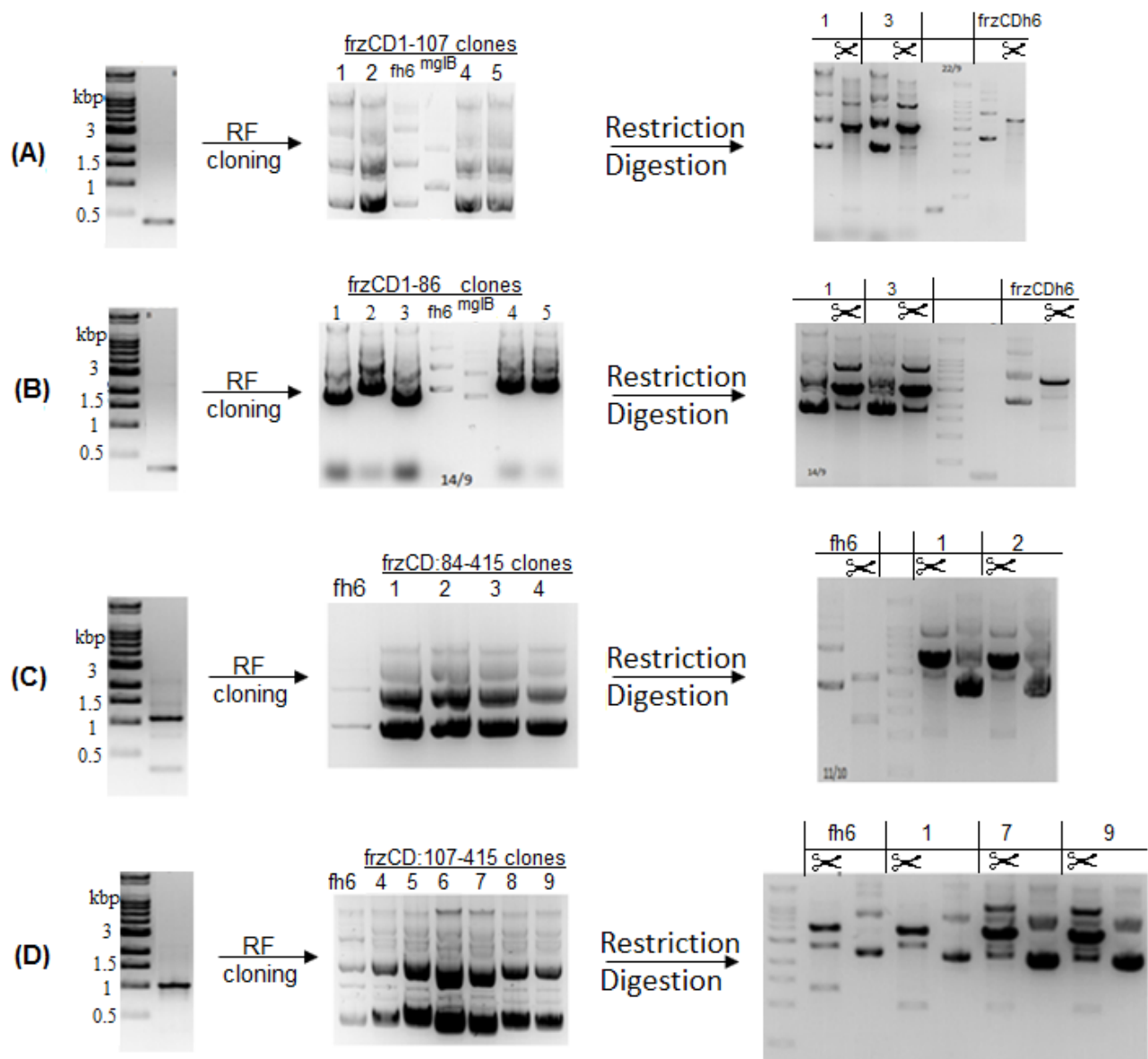
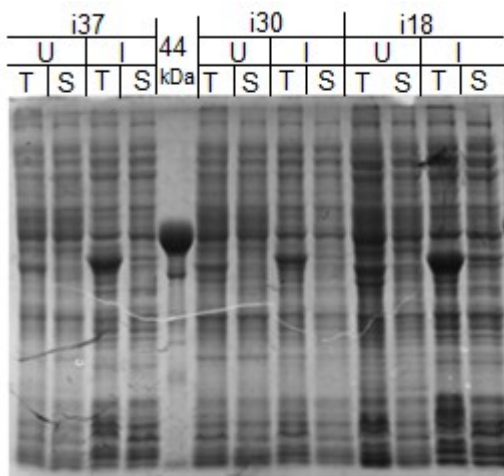
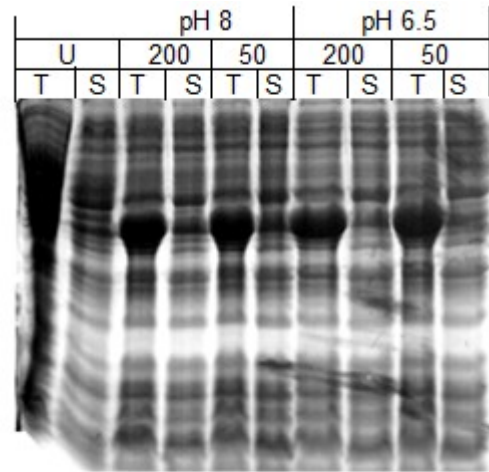


Figure 3.7: Cloning of *frzCD* constructs into pHis17 vector a) *frzCD*1-106-h6 b) *frzCD*.1-86-h6 c) *frzCD*84-417-h6 and d) *frzCD*107-417-h6.

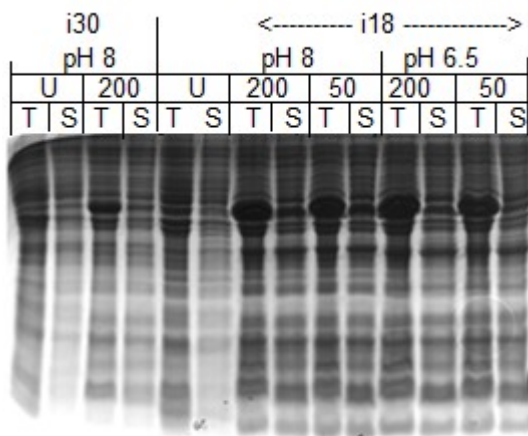
The proteins were then overexpressed in BL21AI *E.coli* cells. The solubility profile of each of these constructs were then analysed in different buffers (Figure 3.8 and 3.9). Both C-terminal constructs FrzCD:84-417-h6 and FrzCD:107-417-h6 were not soluble in buffers of different pH - 6.5 and 8 or salt - 50 mM and 200 mM NaCl (Figure 3.8). The addition of 10% glycerol didn't improve the solubility either. Inducing them at a lower temperature i.e. 18°C instead of 30°C had no effect on the solubility profile of these constructs.



(A)

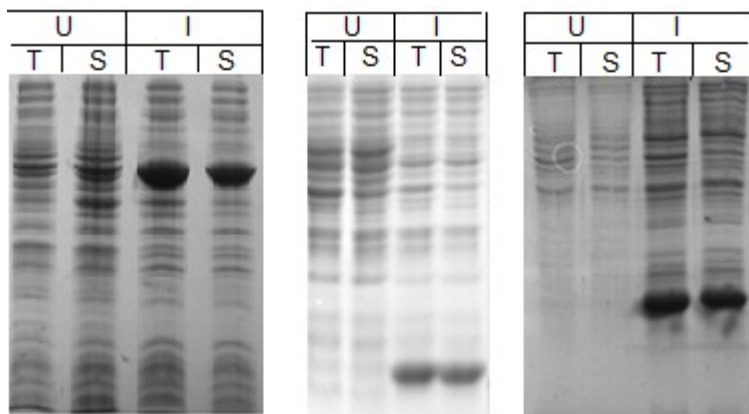


(B)



(C)

Figure 3.8: Solubility profile of the C terminal domain constructs. FrzCD106-417-h6 was insoluble despite inducing the BL21AI cultures at different temperatures (a) and using different lysis buffers of varying salt and pH (b). The same was with FrzCD:84-417-h6 (c). i37, i30 and i18 represent the induction temperatures at 37°C, 30°C and 18°C after growing the bacterial cultures till an  $OD_{600}$  of 0.8. NaCl concentration in mM is indicated as 200 and 50. T is the total lysate while S is the supernatant of it, obtained after spinning. U-uninduced or control and I-Induced fractions.



(A)

(B)

(C)

Figure 3.9: The soluble constructs – (a) the full length FrzCD-h6, (b) FrzCD1-106-h6 and (c) FrzCD1-86-h6.

It is worth mentioning that the expression profile remains the same when induced with arabinose from 0.04% to 0.2% arabinose for FrzCD-h6 and the other soluble constructs. Once the constructs were soluble in the various buffers tried, large volumes (500 ml) of



bacterial culture were grown for protein purification. There was a considerable increase in overexpression of FrzCD when BL21AI cells (transformed with FrzCD-pHis17 plasmid) were grown as 500 ml culture in 2 l flasks compared to 500 ml culture in 1 l flask, presumably due to increased aeration. Subsequently, all bacterial cultures were grown as 500 ml in 2 l flasks. It yielded about 10 times better overexpression of FrzCD than 1 l culture in 2 l flasks.

The two N-terminal domain his-tagged constructs were soluble in a buffer of pH 8 containing 200 mM NaCl and 50 mM Tris (Figure 3.9). Subsequently, these two soluble constructs were purified - first, using affinity based Ni-NTA chromatography to get a single band, removing all other proteins that do not contain the his-tag. The protein was then injected into ion exchange column to remove bound DNA and yielded concentrated protein fractions. However, proteolytic degradation of 20 residues from the N terminus of the protein in all the 3 constructs - the full length, FrzCD:1-86-h6 and FrzCD:1-106-h6 was observed, prior to concentrating it. The addition of 1 mM EDTA after Ni-NTA purification stopped the degradation as confirmed by mass-spectrometry (Figure 3.10). Table 3.1 summarises the various purification trials and the observations.

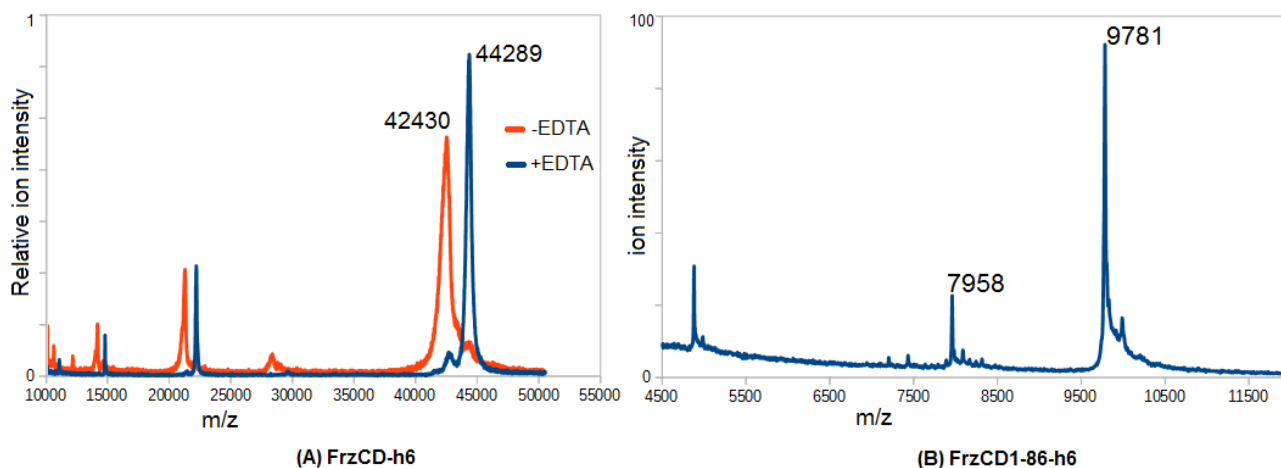


Figure 3.10: Mass spectrometry of the full length protein and FrzCD1-86-h6 shows that the protein purified without the addition of 1 mM EDTA is proteolytically degraded.

Note:

1. FrzCD:1-106-h6 bound very weakly to the anion exchange (MonoQ) column. However, when the same protein dialysed at a pH of 7 was passed through MonoS column (10/100) with a running buffer of pH 7, the protein eluted at a very broad conductivity range (from 14 to 40 mS/cm), indicating that this truncated protein construct is quite inhomogeneous in its oligomeric state.
2. FrzCD:1-86-h6 precipitates at low temperatures i.e. 4°C. Hence, concentration of the protein had to be carried out at 25°C. Moreover, we couldn't determine the concentration

by Bradford method as the absorbance at 595 nm did not decrease with successive dilutions. Denaturing it by SDS and adding Bradford reagent didn't help either. Thus, we used a qualitative method of measuring concentration by running the concentrated protein onto an SDS-PAGE gel and comparing it with a BSA standard. Moreover, monitoring absorbance at 280 nm to determine the concentration is not an option as the protein construct doesn't have any tryptophan.

Table 3.1 : A summary of each purification for all different constructs.

<b>Protein</b>	<b>Purification steps</b>	<b>Culture volume</b>	<b>Yield</b>	<b>Remarks</b>
FrzCD.h6	Ni-NTA, Superdex75, MonoQ	0.9 l	26 mg/ml 150 $\mu$ l	Nuclease contaminated, degraded protein.
	Ni-NTA, MonoQ	1 l	60 mg/ml 170 $\mu$ l	Nuclease contaminated, degraded protein.
	Ni-NTA, MonoQ	1 l	40 mg/ml 135 $\mu$ l	Nuclease contaminated, degraded protein.
	Ni-NTA, MonoQ	1 l	50 mg/ml 200 $\mu$ l	Assay quality, degraded protein.
	<b>Ni-NTA, MonoQ</b>	<b>1 l</b>	<b>10 mg/ml 600 <math>\mu</math>l</b>	<b>Assay quality, intact protein.</b>
FrzCD1-86.h6	Ni-NTA, Superdex75, MonoS	1 l	10 mg/ml 300 $\mu$ l	Nuclease contaminated, degraded protein.
	Ni-NTA, Ammonium sulfate precipitation, Superdex75	0.5 l	20 mg/ml 100 $\mu$ l	Nuclease contaminated, degraded protein.
	Ni-NTA, MonoS	1 l	15 mg/ml 630 $\mu$ l	Nuclease contaminated, degraded protein.
	Ni-NTA, MonoS	1 l	40 mg/ml 560 $\mu$ l	N/A
	<b>Ni-NTA, MonoS</b>	<b>1 l</b>	<b>30 mg/ml 600 <math>\mu</math>l</b>	<b>Assay quality, intact protein</b>
FrzCD1-106.h6	Ni-NTA, MonoQ, MonoS	1.5 l	50 mg/ml 400 $\mu$ l	Nuclease contaminated, degraded protein.
	<b>Ni-NTA, MonoQ</b>	<b>1 l</b>	<b>30 mg/ml 260 <math>\mu</math>l</b>	<b>Assay quality, intact protein.</b>

### 3.3 Oligomeric state of the protein constructs.

It is essential that the protein purified is a pure and homogeneous sample for biochemical and crystallisation experiments. Difference in protein activity can occur in different oligomeric states and a non-homogeneous sample can disrupt crystal growth. Size exclusion chromatography separates molecules based on their size. By comparing the

elution volume of the sample with the standard globular proteins, this method can be used to identify the oligomeric state of the protein.

The full length protein FrzCD-h6 ( $M_w = 43.6$  kDa) and the N terminal constructs – FrzCD:1-106-h6 ( $M_w = 12$  kDa) and FrzCD1-86-h6 ( $M_w = 10$  kDa) exhibited two distinct oligomers in size exclusion chromatography. The protein also elutes differently in different size exclusion columns of different ranges and buffers of varying salt. Thus, the estimated molecular weight of the protein varied - depending on the buffer and the column used. This reflects on the protein being a non-homogenous mixture. This is summarised in Table 3.2.

Table 3.2: The oligomeric state of protein as inferred from the different size exclusion chromatography columns.

Protein	$M_w$ (Da)	Oligomeric state determined from different columns		
		Enrich650	Superdex200	Superdex75
FrzCD1-86.h6	9929	23.90	-	2.98
FrzCD1-106.h6	11993	23.98	-	6.09
		11.11		
FrzCD.h6	44581	11.02	3.07	1.66
		4.49	4.95	

To get a reliable estimate of the oligomeric state of the protein, the proteins were analysed by SEC-MALS (size exclusion chromatography in conjunction with Multi-angle light scattering detector). MALS estimates the molecular weight of the protein using the intensity of scattered light. Moreover, the intensity of the scattered light depends on the molecular weight of the protein and not the shape. At 2 mg/ml, the molecular weight of the proteins were estimated by comparing the scattered intensities with that of the Bovine serum albumin standard. The result of the oligomeric state is summarised in Figure 3.11 and Table 3.3.

Table 3.3: The oligomeric state of protein as inferred from the SEC-MALS.

Protein	$M_w$ (Da)	Experimental $M_w$ (Average)	Oligomeric state
FrzCD1-106.h6	11993	4000 – 40 kDa	400,000 – 4
FrzCD1-86.h6	9929	100 -10 kDa	10 – 1
FrzCD.h6	44581	98.2 kDa	2



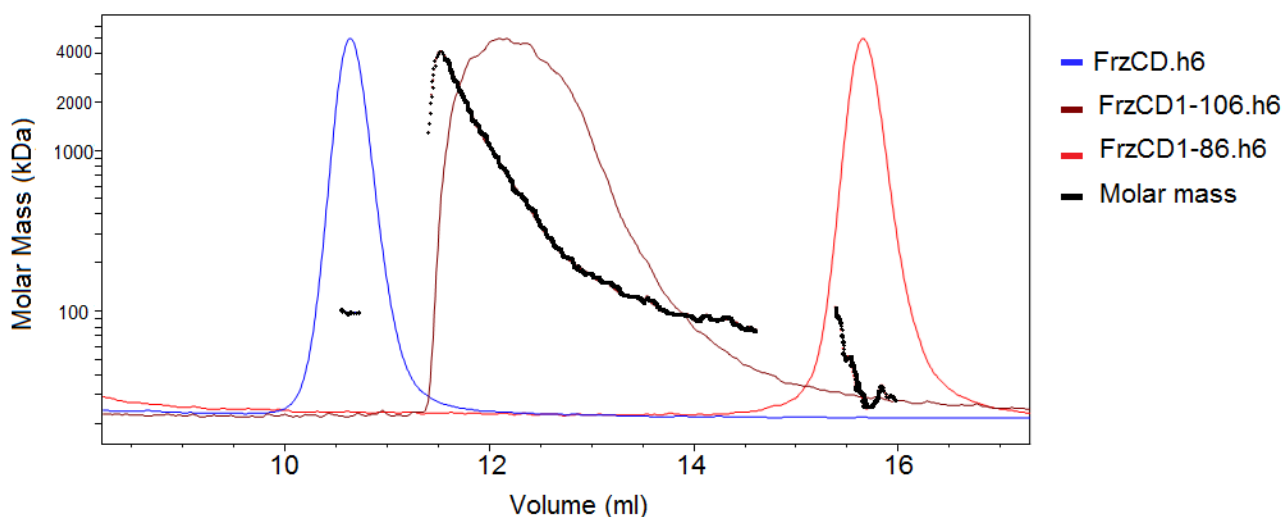


Figure 3.11: SEC-MALS profile of the different constructs of FrzCD. The solution molecular weight is indicated by the black line for each of the constructs.

FrzCD:1-106-h6 is a very heterogeneous mixture whose oligomeric state varies from 4-400,000. The scattering intensity peak during elution has an exponential distribution indicating aggregation and inhomogeneity of the sample. The scattering intensity peak precedes the differential refractive index peak when the protein elutes (Figure 3.12). The sudden increase in the scattering intensity profile as soon as the protein elutes indicates that a very large complex elutes first followed by aggregates of lower size. Moreover, the fact that a protein of 4 MDa molecular weight elutes much away from the void volume in a size exclusion column whose exclusion limit is 650 kDa indicated the non-specific interaction of the protein with the column beads. Moreover, the running buffer that was used during the course of elution contained low salt - 50 mM NaCl and 50 mM Tris of pH 8. When the salt content of the running buffer was increased to 150 mM NaCl, the protein eluted everywhere beyond the void volume, as monitored in UV 280 nm (Figure 3.13).

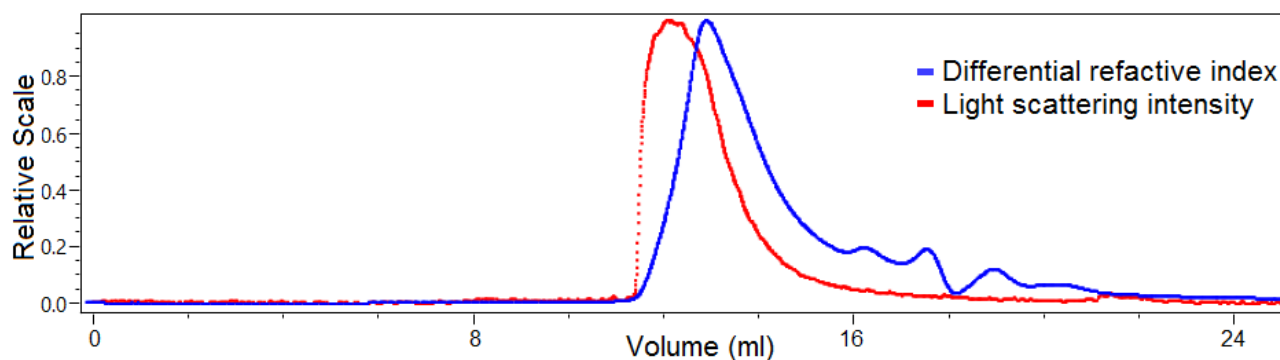


Figure 3.12: The SEC-MALS elution and scattering profile of FrzCD1-106-h6. The running buffer here is of pH 8 and contains low salt – 50 mM NaCl.

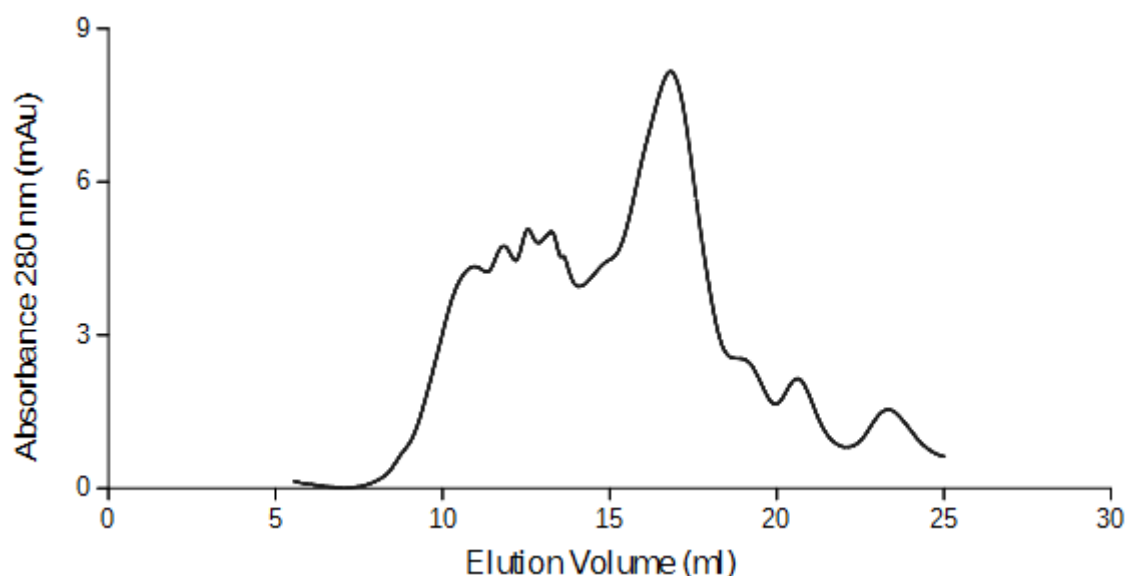


Figure 3.13: SEC profile of FrzCD1-106-h6 elutes in a very broad range of volume, reflecting its inhomogeneous oligomeric state as the salt concentration in the running buffer is increased to 150 mM NaCl.

FrzCD1-86-h6 also has a variable molecular weight with an average molecular mass of 38 kDa as revealed by SEC-MALS. The SEC-MALS profile of the full length protein showed that it is homogeneous with a solution molecular weight of 100 kDa at 2 mg/ml (50  $\mu$ M). The scattering profile was constant throughout the elution of the protein indicating that the good quality of the protein purified and homogeneity of the sample (Figure 3.11).

### 3.4 Crystallisation trials of FrzCD-1-86-h6

Though the SEC-MALS profile of FrzCD1-86-h6 was not encouraging i.e. it had a variable oligomeric state, we went ahead with crystallisation hoping that some of the solutions in the crystallisation conditions would homogenise the oligomeric state of FrzCD-1-86-h6. The first crystallisation screen of FrzCD:1-86-h6 was carried out in 288 conditions at 15 mg/ml and 7 mg/ml using the sitting drop method of vapour diffusion. FrzCD:1-86-h6 crystallised as spherulites in the following conditions (Figure 3.14 a).

1. 0.2 M Ammonium acetate, 0.1 M Tris pH 8.5, 25% PEG3350.
2. 0.2 M Ammonium sulphate, 0.1 M Tris pH 8.5, 25% PEG3350.
3. 0.2 M Lithium sulphate, 0.1 M Tris pH 8.5, 25% PEG3350.
4. 0.1 M Carboxylic acid mix, 0.1 M Tris pH 8.5, 30% w/v PEG550MME\_PEG20k
5. 0.1 M Carboxylic acid mix, 0.1 M Tris pH 8.5, 37.5% w/v MPD\_P1k\_PEG3350

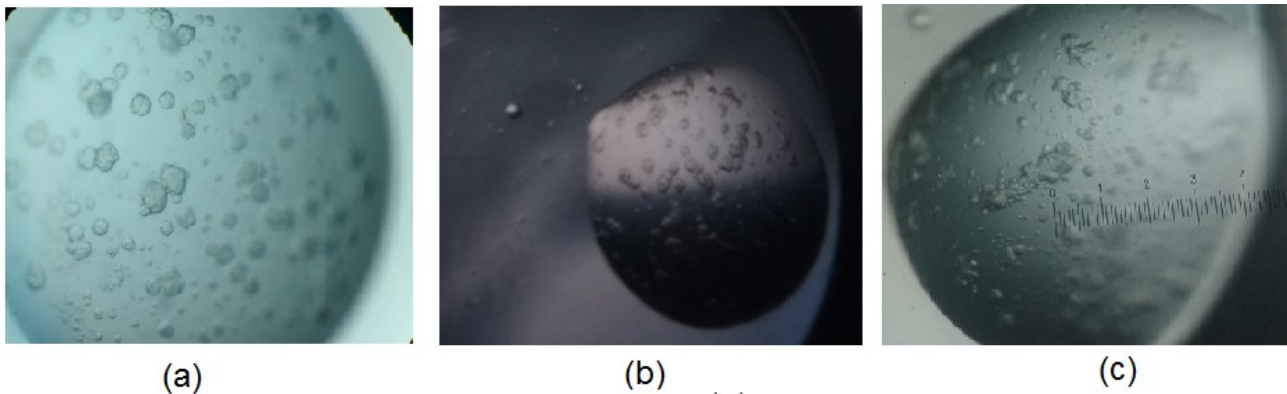


Figure 3.14 a) Crystallisation hits of N86. They were further optimised to get (b) and (c).

Subsequently, to grow them into single crystals, we set up an optimisation screen of the conditions 1, 2 and 3 at 10 mg/ml and 7 mg/ml. In the optimisation screen, the salt was varied from 0.1 M to 0.3 M and PEG3350 concentrations were varied from 15% to 30%, while the pH was maintained the same. However, there was only a difference in the amount of nucleation across this screen (Figure 3.14 b and c). Though the shape of the clusters didn't vary, the sizes and the number of the clusters varied and would not form single crystals.

Subsequently, we tried an additive screen (Hampton Research) for two of the conditions 2 and 3 at 15 mg/ml and 10mg/ml. These 96 conditions consist of various small molecules at high concentration - amino acids, salts, dissociating agents, detergents etc. that can affect the solubility of the protein molecules. There was hardly any variation in the shape of the crystals - they were still clusters, and this approach did not help in obtaining single crystals.

### 3.5 DNA binding experiments

During purification, the protein constructs (FrzCD-h6, FrzCD1-86-h6) had an unusually high 260/280 nm absorbance ratio of the protein after Ni-NTA chromatography which later decreased after passing through the ion exchange column. This together with the very basic pI of the N-terminal domain led to the hypothesis that it may be a DNA binding protein.

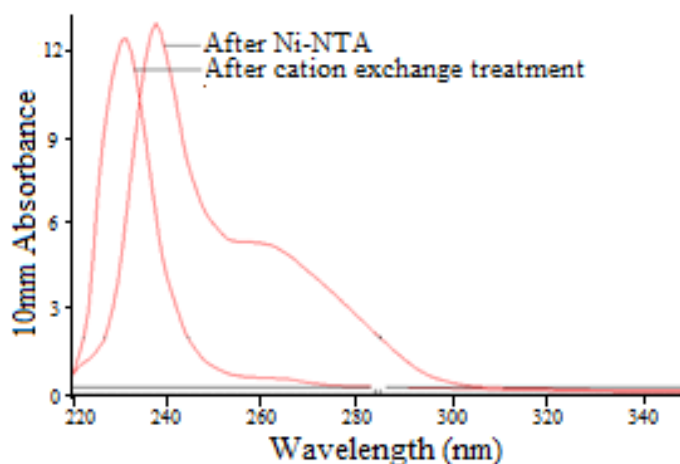


Figure 3.15: The peak at 260 nm disappears upon passing the protein through ion exchange column.

Moreover, *in vivo* fluorescence experiments that show that FrzCD colocalises with the nucleoid lent credibility to this DNA binding hypothesis (personal communication, Tam Mignot). Because the N-terminal domain is very basic in nature and the earlier observation of bound DNA to the affinity purified protein (Figure 3.15) we proceeded with the binding assays of the protein with DNA. The binding was visualised by EMSA (electrophoretic mobility shift assay). We observed that all the three constructs - the full length protein and the two N terminal domain constructs bind the plasmid DNA (Figure 3.16).

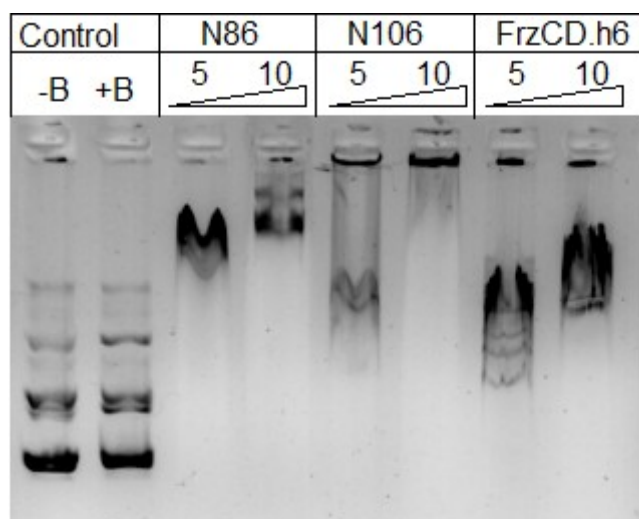


Figure 3.16: All the 3 constructs - the full length and the N terminal constructs bind to DNA. In this qualitative assay, ~300 ng of plasmid was used and equilibrated with 5  $\mu$ M and 10  $\mu$ M of the different proteins. The control lanes contain no protein and are equilibrated without buffer (-B) and with the DNA binding buffer (+B). The protein constructs used here: N86 (FrzCD1-86-h6), N106 (FrzCD1-106-h6) and FrzCD-h6 (the full length protein). The numbers indicated are the protein concentrations used in  $\mu$ M.

However, crystallisation of the protein bound to the DNA is only feasible if the DNA length is small which otherwise would increase the degree of freedom for free DNA, hindering crystallisation. To aid crystallisation of the full length protein, DNA binding experiments with the full length FrzCD-h6 were carried out with dsDNA of various lengths in decreasing order - 1.2 kbp, 316 bp, 147 bp, 70 bp, 30 bp and 8 bp (Figure 3.18).

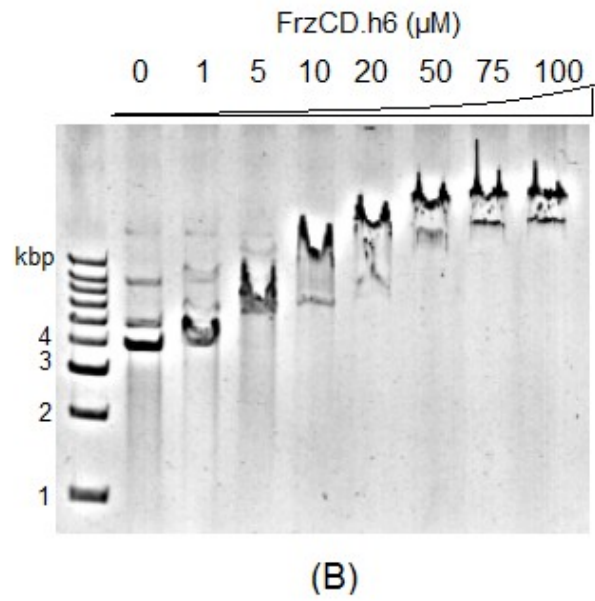
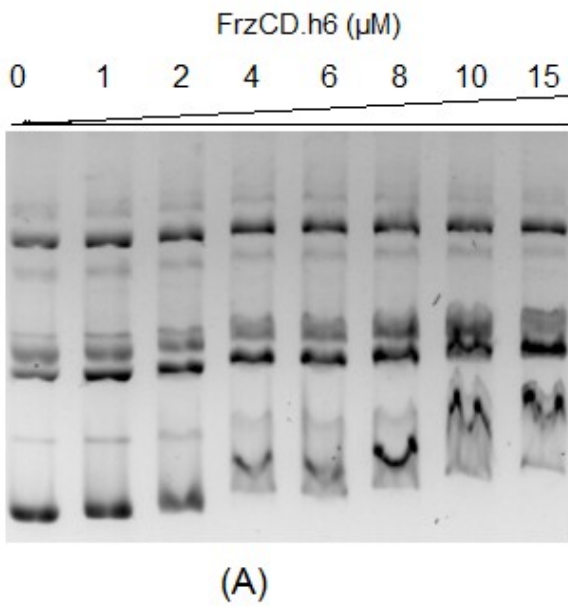
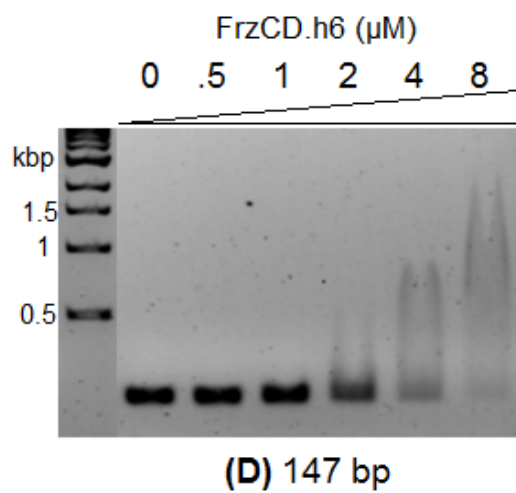
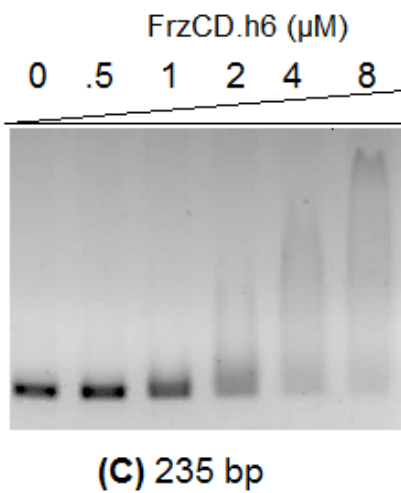
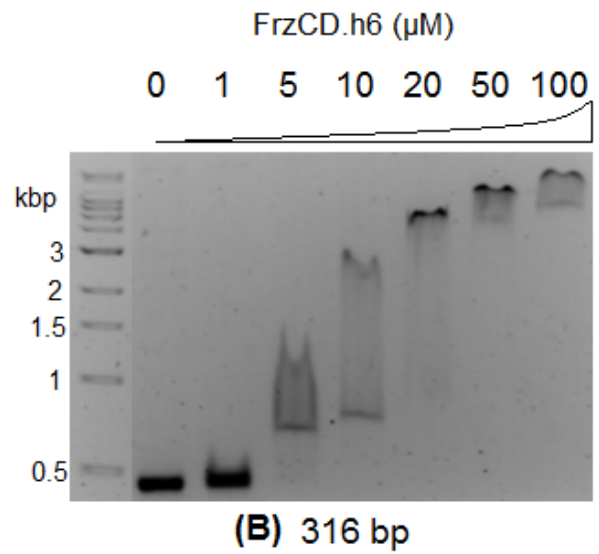
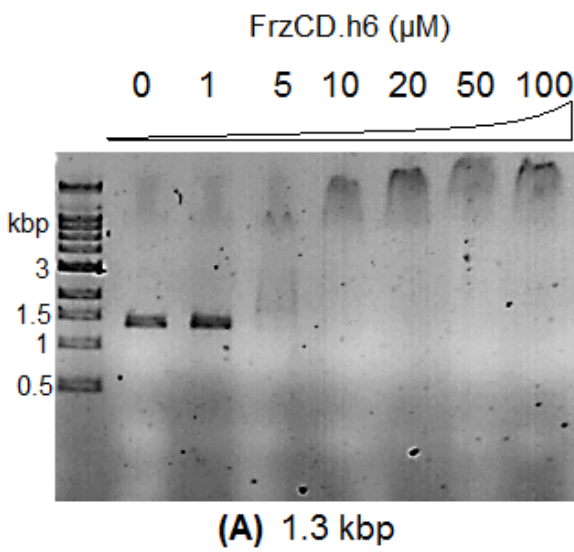


Figure 3.17: Concentration dependent binding experiment of the full length protein FrzCD-h6 with (a) 12 nM of plasmid and (b) with 10 nM of linearised plasmid.



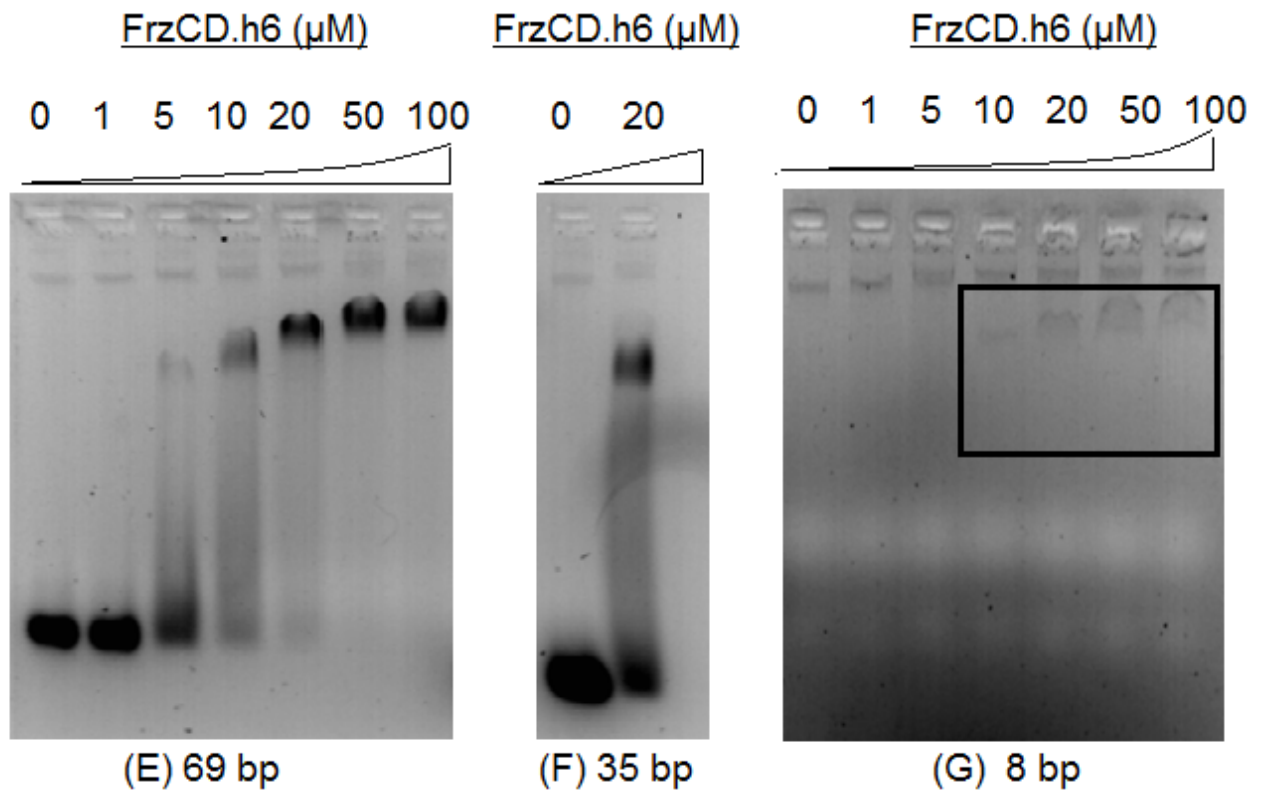


Figure 3.18: Concentration dependent binding experiment of the full length protein with DNA of various lengths. The DNA concentrations used were (a) 4 nM - 1.3 kbp (b) 30 nM - 316 bp (c) 63 nM - 235 bp, (d) - 147 bp, (e) 500 nM - 69 bp, (f) 500 nM - 35 bp and (g) 500 nM - 8 bp.

## 4. Discussion.

FrzCD has two distinct domains - a C terminal coiled coil signalling domain and a N-terminal domain of unknown function and fold. Earlier experiments have shown that there is no phenotypic effect on A and S motility upon the deletion of the N-terminal domain till residue 130 of FrzCD (Bustamante, V.H. et al., 2004). Also, a chimera construct, where the N-terminal domain of FrzCD is replaced by the NarX sensor domain, shows that signals can be transduced directly from the N-terminal domain to the C-terminal domain through the HAMP domain (Xu, Q. et al., 2007). The HAMP domain, a coiled coil oligomerisation domain that is a feature of all MCPs, that lies between the N and the C terminal domain is crucial for this signal transduction mechanism (Wang S., 2012).

Dissecting the protein into the N and the C terminal domains is one of the first steps geared towards understanding the structure and function of the N-terminal domain. However, it is clear from the SEC-MALS experiments that the domains were not dissected appropriately in our constructs as the molecular weight of both the N-terminal constructs – FrzCD1-106-h6 and FrzCD1-86-h6 were not homogeneous. Even if the protein was dissected into its constituent domains, the N and the C-terminal domains are unstable to be without each other. If this latter statement is true, this assumption combined results of the SEC-MALS experiments also indicate that the N-terminal domain is a disordered segment without the appropriate binding partner or signal. This is supplemented by the prediction that the first 35 residues of FrzCD is a disordered region – Figure 4.1 using PrDOS, a protein disorder prediction server (Ishida,T., et al., 2007). The heterogeneity of the protein sample could also be a reason why crystallisation experiments of FrzCD1-86-h6 were not successful.

1	<b>MSLDTPNEKP</b>	<b>AGKARARKAP</b>	<b>ASKAGATNAA</b>	<b>STSSSTKAIT</b>	DTLLTVLSSN	50
51	LQARVPKELV	GESGVELAHL	LNQVLDQFAA	SEHRKHVAAQ	EIDQALDALI	100
101	GLVREGDLSR	WNTTTEDPQL	GPLLEGFGKV	<b>IE</b>		134

Figure 4.1: The N terminal domain of FrzCD is disordered. Predicted disordered residues (coloured in red) and ordered residues (black) using the PrDOS server.

To get another estimate of the domain boundaries of N-terminal domain, we are now looking closely the HAMP domain. The HAMP domain will indicate the domain boundaries between the N-terminal and C-terminal domains. Structural alignment of FrzCD against







protein FrzCD binds preferentially to the supercoiled form of the plasmid, in the presence of other forms of plasmid like the linear and circular forms. Further foolproof competition experiments are required to validate this hypothesis. This would give further credibility to our in vitro experiments if this is true, again supporting the physiological relevance of the full length protein binding to the nucleoid (supercoiled DNA).

Though the data here requires further optimisation in terms of the relative concentrations of the DNA used in the binding experiments, the DNA shifts observed in EMSA experiments hint that the binding events of protein (FrzCD) to the DNA are not independent. If binding were independent, multiple DNA shifts would have been seen.

We are also hopeful that these DNA binding experiments will help in our crystallisation experiments. A coiled coiled protein would make fewer crystal contacts compared to a globular protein as it would have make poor contacts in all but one orientation – where the coiled coil domains (rods) align next to each other. Crystallisation of a highly coiled coil protein will be favored if crystal contacts between the protein residues are maximised. This (low crystal contacts) could be reason for failure of our crystallisation experiments with the full length protein, FrzCD-h6. A way to increase side-side contacts between protein is by increasing the lateral contacts of FrzCD bound to the DNA. Because multiple protein molecules can bind to a reasonable stretch of DNA through the N-terminal domain, the chances of the coiled coil C-terminal domain making lateral contacts seem promising. If crystallisation conditions are just right, how cooperative effect between the receptors are achieved could also be elucidated.

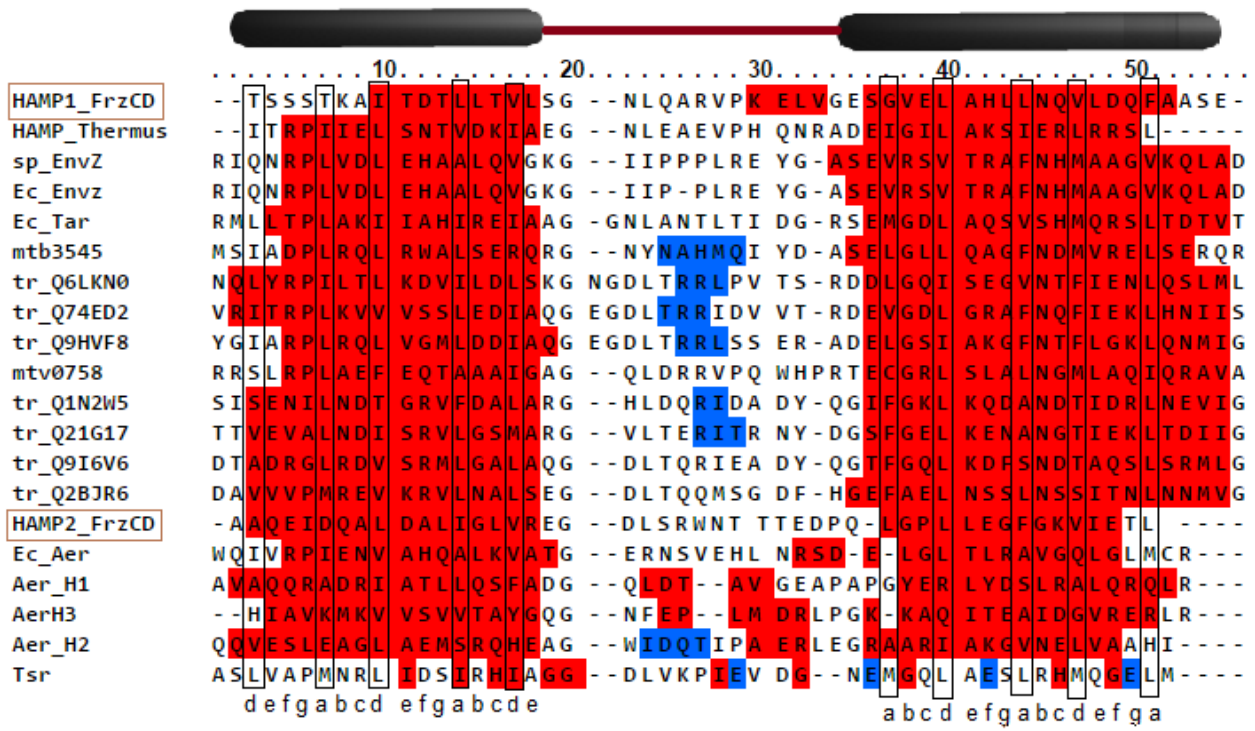


Figure 4.3: The residues 24-80 labelled as HAMP1\_FrzCD aligns well against other annotated HAMP domains with a characteristic heptad repeat, as shown below the sequences in the figure. The boxed residues are the hydrophobic residues located at 'a' and 'd' positions of the heptad repeat in the HAMP helix. The residues highlighted in red are predicted to be a helix and blue –  $\beta$ -sheet. Created using PRALINE server (<http://zeus.few.vu.nl>).

## 5. Conclusions and the future.

The dissection of this cytoplasmic chemoreceptor into its constituent domains is the first step towards understanding the individual functions of the domains of FrzCD. Though the dissection of the protein into its constituent domains yielded either soluble (N-terminal domain constructs) or insoluble aggregates (the C-terminal domain constructs), we are hopeful that we can design better constructs with a stronger picture of the domain boundaries of FrzCD from the better bioinformatic analysis to be carried out. We could also assign a new function to the protein from a pure biochemistry angle - the DNA binding aspect. Our work is further supported by the independent in-vivo and in-vitro experiments done in the Tam Mignot group. Though it is qualitative and speculative at this point of time, the co-operative mode of DNA binding is a relevant feature to the members of the MCP family. This is the first instance of an MCP binding to DNA, and the relevance of this feature with respect to signal sensing and downstream phosphorylation/methylation is currently unknown.

We have identified a shortest segment of DNA that binds to FrzCD, which is an ideal ligand to initiate crystallization experiments of the full length protein (the most physiologically relevant construct). Competition experiments to determine if the protein binds preferentially to the supercoiled form of the plasmid are also in progress. The co-operative mode of DNA binding will be further validated by other quantitative experiments. Once we are successful in this part of the project, we will design methylation mimics of the FrzCD to check if there is a change in oligomeric state by SEC-MALS or a difference in binding affinity to DNA. Crystallisation experiments of the methylation mimics will reveal if there is a conformational change in the protein. If none of these changes are prominent, biochemical experiments with its downstream partners - FrzF, the methyl-transferase, FrzA & FrzB – the two adaptor proteins and FrzE, the histidine kinase will be conducted to assess biochemical affinity or a difference in catalytic activity. If none of the changes are still prominent, we will have to think of evaluating how the change in receptor state is conveyed to modulate reversal frequency. Perhaps, it is imagination that will define the course of this project.

To cut the tale of this long journey short, this journey of thousand miles began with a single step - and this step has given us new interesting and hopeful directions, though we fell in the well for quite sometime.

## 6. References.

- Alva, V., Nam, S.-Z., Söding, J., and Lupas, A.N. (2016). The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic Acids Res.* **44**, 348, W410-415.
- Astling, D.P., Lee, J.Y., and Zusman, D.R. (2006). Differential effects of chemoreceptor methylation-domain mutations on swarming and development in the social bacterium *Myxococcus xanthus*. *Mol. Microbiol.* **59**, 45–55.
- Bustamante, V.H., Martínez-Flores, I., Vlamakis, H.C., and Zusman, D.R. (2004). Analysis of the Frz signal transduction system of *Myxococcus xanthus* shows the importance of the conserved C-terminal region of the cytoplasmic chemoreceptor FrzCD in sensing signals. *Mol. Microbiol.* **53**, 1501–1513.
- Chayen, N.E., and Saridakis, E. (2008). Protein crystallization: from purified protein to diffraction-quality crystal. *Nat. Methods* **5**, 147–153.
- Ferris, H.U., Zeth, K., Hulko, M., Dunin-Horkawicz, S., and Lupas, A.N. (2014). Axial helix rotation as a mechanism for signal regulation inferred from the crystallographic analysis of the *E. coli* serine chemoreceptor. *J. Struct. Biol.* **186**, 349–356.
- GE Healthcare Life Sciences, Gel Filtration Calibration kits – [https://www.gelifesciences.com/gehcls\\_images/GELS/Related%20Content/Files/1326706518989/litdoc28951560\\_20161014092124.pdf](https://www.gelifesciences.com/gehcls_images/GELS/Related%20Content/Files/1326706518989/litdoc28951560_20161014092124.pdf)
- Hampton Research, Sitting drop vapor diffusion crystallisation, [http://hamptonresearch.com/documents/growth\\_101/4.pdf](http://hamptonresearch.com/documents/growth_101/4.pdf)
- Hellman, L.M., and Fried, M.G. (2007). Electrophoretic Mobility Shift Assay (EMSA) for Detecting Protein- Nucleic Acid Interactions. *Nat. Protoc.* **2**, 1849–1861.
- Ishida, T., and Kinoshita, K. (2007). PrDOS: Prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* **35**, 460–464.
- Kaiser, D. (2003). Coupling cell movement to multicellular development in myxobacteria. *Nat. Rev. Microbiol.* **1**, 45–54.
- Kirby, J.R. (2009). Chemotaxis-like regulatory systems: unique roles in diverse bacteria. *Annu. Rev. Microbiol.* **63**, 45–59.

- Mauriello, E.M.F., Astling, D.P., Sliusarenko, O., and Zusman, D.R. (2009). Localization of a bacterial cytoplasmic receptor is dynamic and changes with cell-cell contacts. *Proc. Natl. Acad. Sci. U. S. A.* *106*, 4852–4857.
- McBride, M.J., Kohler, T., and Zusman, D.R. (1992). Methylation of FrzCD, a methyl-accepting taxis protein of *Myxococcus xanthus*, is correlated with factors affecting cell behavior. *J. Bacteriol.* *174*, 4246–4257.
- McBride, M.J., Weinberg, R. a, and Zusman, D.R. (1989). “Fizzy” aggregation genes of the gliding bacterium *Myxococcus xanthus* show sequence similarities to the chemotaxis genes of enteric bacteria. *Proc. Natl. Acad. Sci. U. S. A.* *86*, 424–428.
- McCleary, W.R., McBride, M.J., and Zusman, D.R. (1990). Developmental sensory transduction in *Myxococcus xanthus* involves methylation and demethylation of FrzCD. *J. Bacteriol.* *172*, 4877–4887.
- Nan, B., and Zusman, D.R. (2011). Uncovering the Mystery of Gliding Motility in the Myxobacteria. *Annu. Rev. Genet.* *45*, 21–39.
- Nan, B., Mauriello, E.M.F., Sun, I.H., Wong, A., and Zusman, D.R. (2010). A multi-protein complex from *Myxococcus xanthus* required for bacterial gliding motility. *Mol. Microbiol.* *76*, 1539–1554.
- Pei, J., and Grishin, N. V. (2007). PROMALS: Towards accurate multiple sequence alignments of distantly related proteins. *Bioinformatics* *23*, 802–808.
- Scott, A.E., Simon, E., Park, S.K., Andrews, P., and Zusman, D.R. (2008). Site-specific receptor methylation of FrzCD in *Myxococcus xanthus* is controlled by a tetra-trico peptide repeat (TPR) containing regulatory domain of the FrzF methyltransferase. *Mol. Microbiol.* *69*, 724–735.
- Shuishu Wang (2012). Bacterial Two-Component Systems: Structures and Signaling Mechanisms, Protein Phosphorylation in Human Health, Dr. Cai Huang (Ed.), InTech, DOI: 10.5772/48277. Available from: <http://www.intechopen.com/books/protein-phosphorylation-in-human-health/bacterial-two-component-systems-structures-and-signaling-mechanisms>.
- Wadhams, G.H., and Armitage, J.P. (2004). Making sense of it all: bacterial chemotaxis. *Nat.Rev.Mol.Cell Biol.* *5*, 1024–1037.

- Weinberg, R.A., and Zusman, D.R. (1989). Evidence that the *Myxococcus xanthus* frz genes are developmentally regulated. *J. Bacteriol.* *171*, 6174–6186.
- Westra, E.R., Buckling, A., and Fineran, P.C. (2014). CRISPR-Cas systems: beyond adaptive immunity. *Nat. Rev. Microbiol.* *12*, 317–326.
- Wilson, D., Pethica, R., Zhou, Y., Talbot, C., Vogel, C., Madera, M., Chothia, C., and Gough, J. (2009). SUPERFAMILY - Sophisticated comparative genomics, data mining, visualization and phylogeny. *Nucleic Acids Res.* *37*, 380–386.
- Xu, Q., Black, W.P., Mauriello, E.M.F., Zusman, D.R., and Yang, Z. (2007). Chemotaxis mediated by NarX-FrzCD chimeras and nonadapting repellent responses in *Myxococcus xanthus*. *Mol. Microbiol.* *66*, 1370–1381.
- Zhulin, I.B. (2001). The superfamily of chemotaxis transducers: from physiology to genomics and back. *Adv. Microb. Physiol.* *45*, 157–198.
- Zusman, D.R., Scott, A.E., Yang, Z., and Kirby, J.R. (2007). Chemosensory pathways, motility and development in *Myxococcus xanthus*. *Nat. Rev. Microbiol.* *5*, 862–872.