Use of helical assembly scaffolds as tools for alleviating preferred orientation problem in Cryo-Electron Microscopy

A Thesis

submitted to

Indian Institute of Science Education and Research Pune in partial fulfilment of the requirements for the BS-MS Dual Degree Programme

by

Aishwarya Shahare

Indian Institute of Science Education and Research Pune

Dr. Homi Bhabha Road,

Pashan, Pune 411008, INDIA.

April, 2023

Under the guidance of

Supervisor: Gayathri Pananghat,

Associate Professor, Department of Biology, IISER Pune

From May 2022 to Mar 2023

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE

Certificate

This is to certify that this dissertation entitled **"Use of helical assembly scaffolds as tools for alleviating preferred orientation problem in Cryo-Electron Microscopy"** towards the partial fulfilment of the BS-MS dual degree program at the Indian Institute of Science Education and Research, Pune represents work carried out **by Aishwarya Shahare** at the Indian Institute of Science Education and Research Pune under the supervision of **Dr. Gayathri Pananghat, Associate Professor, Department of Biology,** during the academic year 2022-2023.

Dr. Gayathri Pananghat

Thesis Advisory Committee:

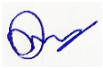
Dr. Gayathri Pananghat, IISER Pune

Dr. Radha Chauhan, NCCS, Pune

This thesis is dedicated to all my well-wishers.

Declaration

I hereby declare that the matter embodied in the report entitled **"Use of helical assembly scaffolds as tools for alleviating preferred orientation problem in Cryo-Electron Microscopy"** are the results of the work 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.



Aishwarya Shahare

Date: 10/ 04/ 2023

Table of Contents

List of Tables

List of Figures

Abstract

The preferred orientation issue has currently emerged as a major challenge for the structural analysis of proteins in cryo-electron microscopy (cryo-EM). Smaller proteins (<100 kDa) are particularly challenging to identify and align in noisy low-dose pictures of cryo-EM. Many smaller proteins and their liganded complexes play a significant role in a biological process but their structure is unknown. Therefore, a tool that assists in structure determination of smaller proteins will be useful for structural biology applications. Megabodies have been recently used to solve the preferred orientation issue of smaller proteins, which consists of a nanobody attached to a large scaffold protein connected by a linker. It increases attached protein size and makes it easy to align and observe in 2D images. To overcome the problem of orientational preference, we have developed a megabody, consisting of a helical filament scaffold that can provide views from all orientations of the protein due to its helical symmetry. We have used ParM, a bacterial actin that polymerizes in the presence of ATP, as the choice for helical scaffold. As a proof-of-principle, we have attached a nanobody that binds to ALFA tag, a helical peptide sequence, to ParM. This enables any protein with the ALFA tag to decorate the helical megabody of nanobody-tagged ParM filaments. Towards this goal, we have explored the use of linker lengths of three different lengths between nanobody and ParM. Polymerization assay through pelleting has been used for monitoring efficiency of filament formation and co-pelleting of the ALFA-tagged small protein (SofG) was carried out for validating the tool design. Moreover, our results show that ParM helical megabody can also be used as a solubility purification tag for nanobodies.

Acknowledgments

I would like to thank everyone for your assistance as I near the completion of my one-year thesis project. Without you, this project would not have been feasible. First of all, I would want to extend my sincere appreciation to my guide Dr. Gayathri Pananghat for her invaluable advice and support during this research. I feel incredibly fortunate to be a member of her lab and to work with her. Second, I want to convey my deep gratitude to my mentor Soumyajit Dutta for his unwavering support and important advice during every stage of my project.

I also like to express my gratitude to Dr. Radha Chauhan for her thoughtful remarks and recommendations on this endeavour. I also like to thank Puja Kumari for her assistance in getting the protein purified. Moreover, I would like to thank Maria Sabu, Shefali Sonarkar, and Prajakta Umbarker for their prior work on this project during semester projects.

In addition, I want to thank Joyeeta, Rinku, Apurba, Mrinmayee, Sukanya, and Vani for always being accessible and helpful. I want to express my sincere appreciation to every one of the G3 and SK lab staff for maintaining a welcoming environment that allows for productive work. Akhilesh, Ranjana, and the rest of my batchmates made this year special for me by being kind and nice, and I cannot even begin to say how grateful I am to them. I want to end by saying how much I appreciate my family's love and support.

Contributions

Contributor name	Contributor role
Dr. Gayathri Pananghat, Soumyajit Dutta, Aishwarya Shahare	Conceptualization Ideas
Dr. Gayathri Pananghat, Soumyajit Dutta, Aishwarya Shahare	Methodology
Dr. Gayathri Pananghat	Validation
Aishwarya Shahare	Formal analysis
Dr. Gayathri Pananghat, Soumyajit Dutta, Aishwarya Shahare	Investigation
Dr. Gayathri Pananghat	Resources
Aishwarya Shahare	Data Curation
Aishwarya Shahare	Writing - original draft preparation
Aishwarya Shahare, Dr. Gayathri Pananghat	Writing - review and editing
---	Visualization
Dr. Gayathri Pananghat, Soumyajit Dutta	Supervision
Dr. Gayathri Pananghat	Project administration
Dr. Gayathri Pananghat	Funding acquisition

Chapter 1. Introduction

This chapter introduces about preferred orientation problem during structure determination using cryo-EM and the strategy that we have proposed for overcoming it in detail.

1.1 Cryo-EM

Cryogenic electron microscopy (cryo-EM) is a high-resolution electron microscopy technique used to image biomolecules in cryogenic conditions (Murata et al., 2018).
To visualize these biomolecules using cryo-EM, they are suspended in water (usually a buffer) and flash frozen within a thin layer of vitreous ice. Thus, here the specimens are observed in ice. Therefore, it is known as “Cryo-EM”.
The particles are frozen in random orientations on the sample grid.
Once the particles get embedded in ice film, they are bombarded with electrons to obtain different 2D views for an individual protein particle. Different 2D views are obtained due to their random orientations on the sample grid. These 2D views are aligned and then views of the same orientation averaged to obtained a class average. Each class average contributes to reconstruct the 3-D structure of a protein (Nakane et al., 2020). However, averaging is not a straightforward process. It also involves image-processing methods including accurate alignment methods and dedicated software tools to build a 3D model for the protein (Nwanochie et al., 2019).

1.1.2. Preferred orientation problem

Cryo-EM can provide the 3D structure for a limited number of proteins because of the factors like the non-random orientation of ice particles and their size, which can have an impact on the resolution of 3D reconstructions. It is difficult to study small particles (<100 kDa) or strongly oriented molecules by cryo-EM. These proteins rarely orient themselves with perfect randomness when vitrified on a cryo-EM grid which is necessary to obtain different views. Some proteins because of their morphology orient themselves in a particular orientation with their long axis parallel to the grid.
Some proteins because of their surface properties, get stuck to the air-water interface (AWI), which makes them get adsorbed and diffuse in the air-water interface (Li et al., 2021). This leads to showing them called “preferred orientation”. Because of the

preferred orientation, some important views can be missed in reconstructing their 3D structure.

1.1.3 Different ways to overcome preferred orientation problem

The preferred orientation problem can be reduced by improving the particle distribution and randomness in ice (Naydenova et al., 2017) or by increasing the size of particles. To improve particle distribution in ice and randomness, different strategies have been employed which may reduce the challenges of preferred orientation. These include:

(a) By collecting more sets of data to get additional missing views (Kühlbrandt, 2014).

(b) By tilting the grids to increase views

(c) By minimizing the time for protein to stick at the air-water interface (Liu et al., 2022)

(d) Alternative grid support (Noble et al., 2018)

(e) Modulation of air-water interface by adding surfactants which can prevent stickiness (Chen et al., 2019, Liu et al., 2022).

Apart from the particle distribution, particle size prevents the highest possible resolution for cryo-EM reconstructions. Smaller proteins have a high chance to go into preferential orientation as they are easily adsorbed (Uchański et al., 2021). Also, it is difficult to recognise frozen-hydrated samples in noisy low-dose pictures and for alignment and averaging. Structural analysis of larger molecules is relatively easier when compared to smaller particles. Larger particles have sufficient features for alignment and averaging to make it easier to determine their position and orientation accurately.

Next, to enlarge the size of the protein, smaller target proteins were genetically fused to multimeric scaffolds (Cabral et al., 2022) or recognition domains but were limited in practice because of the flexibility of linker regions (Uchański et al., 2021).

However, to date, such methods have not been able to be broadly applied to resolve the cryo-EM particle orientation issue.

1.2 Megabodies

To overcome the problem of preferred orientation, recently, megabodies have been developed (Uchański et al., 2021), Mega means big, developed from nanobody).

1.2.1 Design principles of megabodies

Megabody is a chimeric protein, which consists of a nanobody (single domain antibody) and a large scaffold protein connected by a linker (Uchański et al., 2021). Nanobody is a single-domain antibody that is used for binding to a protein (Uchański et al., 2020, Wu et al., 2021). The C-terminus of the nanobody is attached to the N-terminus the of scaffold protein by a linker (Fig. 1.2). Linker can vary according to the size of the scaffold protein to which the nanobody is attached.

1.2.2 Rationale behind the megabody concept

Nanobody is attached to the large scaffold protein. Large scaffold protein helps in increasing the size of the targeted protein which may increase randomness and particle distribution in the ice leading to a reduction in a preferred orientation. It can also prevent it from getting absorbed and denature in the air-water interface. This chimeric protein can then be easily managed and tilted while imaging to get different orientations. It will help in identifying and aligning the smaller proteins in noisy images of Cryo-EM. This makes the 3D construction of them easy.

1.2.3 Design of megabodies on a helical scaffold

We have made megabody construct with a helical scaffold with the motive of getting all views of the attached protein since just binding with the larger scaffold does not ensure views from all sizes but increases the size which can make it easy to observe and tilt and align.

The distribution of views all along the helical axis occurs automatically in a helical assembly because the arrangement of molecules along the filament axis provides rotated views of the molecules due to the helical symmetry. The design of such a tool will consist of the following components: a) a helical filament acting as a scaffold b) a protein with a strong binding affinity to a peptide tag c) a small protein tagged with the peptide (Fig. 1.1).

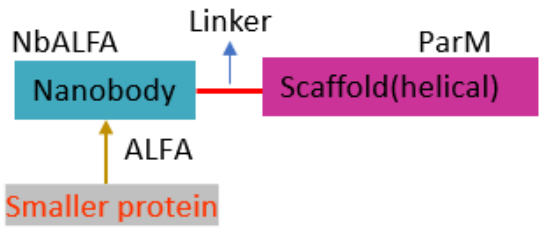


Fig.1.1 Primary structure of megabody attached to a smaller protein via ALFA tag.

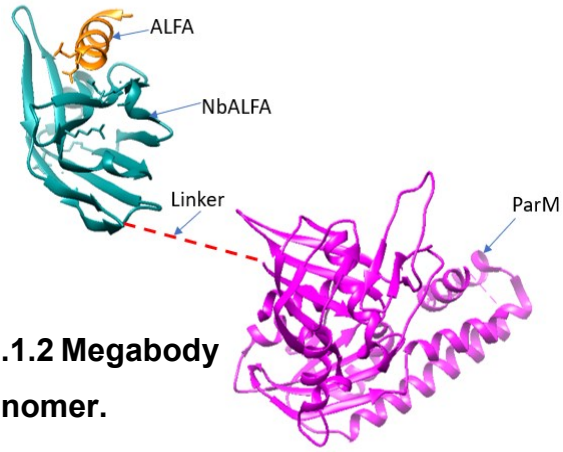


Fig.1.2 Megabody monomer.

We have used nanobody as NbALFA which has a high affinity towards the ALFA tag. So, a target protein is tagged with ALFA to get attached to megabody. We used particularly ALFA tag because of its smaller size (Götzke et al., 2019). Second, we used ParM as a scaffold protein as it forms the helical filament on polymerization which can help in getting all-orientated views of smaller proteins (Fig.1.3).

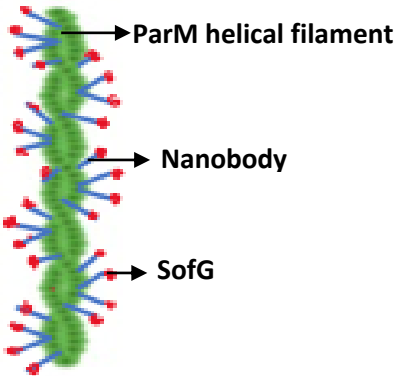


Fig. 1.3 Schematic representation of megabody bound to SofG-ALFA

In this construct, linker length plays a significant role. It should not be very small or very large because of issues with clashes or flexibility respectively.

1.2.3.1 Helical filament scaffold ParM

ParM is an actin homologue. Like actin, it forms a double-helical filament but with a left helical twist (Gayathri et al., 2013, Orlova et al., 2007, Popp et al., 2008). Actin filaments show a right helical twist. ParM shows dynamic instability. We need filament conformation of ParM for getting all orientations of protein due to its helical

arrangement of monomers within the filament. Filament conformation of ParM is maintained as long as the ATP cap is maintained. Once the ATP cap depletes, ParM polymerization will stop and will go back from filament state to monomeric state. ParM monomers hydrolyze ATP in its filament state and thus can lead to depletion in ATP cap. Hence ParM will not be polymerized after some time. To maintain the polymerized state of ParM, we have mutated the residues of ParM which react with gamma phosphate and lead to the hydrolysis of ATP. Thus, on mutation, the ATP will not be hydrolyzed and ParM will be maintained at polymerized state. We can also take a non-hydrolyzable analog of ATP i.e., AMP-PNP. We used AMP-PNP in the experiments for getting the polymerized ParM.

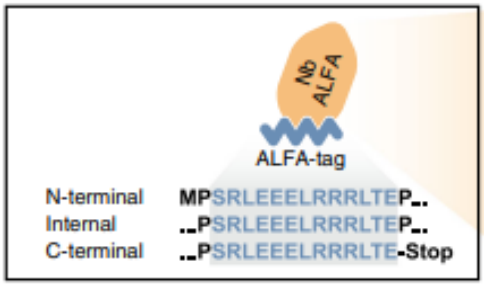
1.2.3.2 Role of linker length

In the megabody construct, nanobody NbALFA is covalently attached to the scaffold protein ParM by a linker. Here, linker length (number of residues for the linker) plays a very important role. A very small linker might affect the folding of a protein and thus can decrease the expression. Very large linkers lead to flexibility between the attached proteins preventing the attached protein from following the helical symmetry. Therefore, the linker length should be as such which can make the linker rigid (Yao et al., 2019) and not lead to clashes or excess flexibility. It should not be very small or very large.

To get the optimum length of the linker, we have made the megabody with different linker lengths, namely- 3, 5 and, 10 to see at which length of linker, megabody expresses well and leads to a better expression of the attached protein on binding.

1.2.3.3 ALFA system for nanobody – protein interaction

ALFA system consists of nanobody NbALFA and ALFA tag. NbALFA is a single domain antibody (Götzke et al., 2019). It consists of a single variable antibody domain. Since these antibodies are much smaller than common antibodies, they are known as “nanobodies”. They can bind selectively to the antigen same as the whole antibody. The ALFA tag is a very adaptable tag that forms a stable alpha helix in solution. This alpha helix makes the tag more compact and smaller and thus can easily be compatible with the proteins. The sequence of NbALFA and ALFA is shown in Fig.1.4.



ALFA sequence

NbALFA sequence

Figure 1.4 ALFA and NbALFA sequence

(Adapted from: Götzke *et al.*,2019)

In order to bind to the ALFA tag, nanobody NbALFA has a paratope on its N-terminus, consisting of a five stranded beta-sheet. Residues of five-stranded beta-sheet form a hydrophobic cavity which makes the centre axis of NbALFA parallel to the ALFA peptide. Also, the ALFA tag establishes many polar and hydrophobic interactions with NbALFA for binding tightly to it (Götzke et al., 2019).

NbALFA shows a higher affinity towards the ALFA tag as compared to other systems. So, a target protein is tagged with ALFA in order to get attached to the nanobody. Both nanobody NbALFA and ALFA tag are smaller in size which makes their binding stronger.

We have chosen the ALFA system particularly, for nanobody-protein interaction, because of its strong binding compared to others and the versatility of the ALFA tag.

1.2 Objectives

In order to overcome the preferred orientation problem in cryo-EM, we aim to develop megabody with a helical scaffold as a helical scaffold protein can provide all orientations because of its helical symmetry. We have chosen ParM protein as the megabody scaffold (as it forms helical filament on polymerization) and nanobody NbALFA. The specific objectives of this thesis include the designing, cloning, purification of megabody constructs (of different linker lengths), and SofG-ALFA construct. In order to get different orientations of SofG protein, it needs to interact with megabody. This thesis reports the attempts to co-express SofG-ALFA with megabody at various levels. The polymerization of megabody constructs and interaction of SofG-

ALFA with the megabodies is confirmed by AMPPNP-based pelleting assays. Polymerization is essential for ParM in megabody so that NbALFA is exposed in a helical manner and bind to target small protein for getting all views. This strategy can also be used for the purification of nanobodies in larger amounts by the polymerization of ParM. Similarly, if we are successful in binding the smaller proteins properly to ParM, then this megabody construct can be used as a solubility tag to produce more amount of smaller proteins through a facile purification step through ATP-based pelleting.

The detailed list of objectives of this thesis is listed below.

- To clone megabody constructs with different linker lengths.
  (a) NbALFA-GSP-ParM
  (b) NbALFA-GSPGS-ParM
  (c) NbALFA-ENLYFQG-GSP-ParM
  where GSP, GSPGS denotes the sequence of the linker residues connecting NbALFA to ParM, and ENLYFQG denotes the TEV protease cleavage site.
- To express above megabody constructs in BL21AI and SHuffle strain and their polymerization assay.
- To clone the smaller protein SofG-ALFA construct and its expression in BL21AI strain and SHuffle strain.
- Co-expression of megabody constructs and SofG-ALFA construct in BL21AI and SHuffle strains and polymerization assay.
- To purify, Megabody and SofG-ALFA proteins.
- To purify nanobody by using ParM as a tag by inserting TEV protease cleavage sequence in between NbALFA and GSP of megabody.

Chapter 2. Materials and Methods

We have made megabody with three different lengths~ (a) GSP (3 aa.), (b) GSPGS (5 aa.), (c) ENLYFQGGSP (10 aa.). We chose *Myxococcus xanthus* SofG as the protein target, a small protein of approximately 27 kDa.

2.1 Primer Design

The vector-specific forward primer (T7 forward) was used for the cloning of all the following constructs. Reverse primers were designed for cloning all the following constructs.

(a) Megabody

The primer design of the NbALFA-GSP-ParM (Nb3) construct was carried out by a previous project student (Shefali Sonarkar and Prajakta Umbarkar) in the lab. For the other two constructs, NbALFA-GSPGS-ParM (Nb5) and NbALFA-TEV-GSP-ParM (Nb10) which have linker lengths 5 and 10 respectively, reverse primers were designed for extension of linker GSP. The strategy for extension of the linker with the help of reverse primer is given in detail in section 2.2. A minimum of 18 nucleotides are maintained at the two ends of the insertion sequence for proper annealing. The primer was also checked for non-specific binding to any other sites on the plasmid by sequence alignment. Primers used for the cloning of all megabody constructs are tabulated below in table 2.1.

(b) SofG-ALFA

For cloning of SofG-ALFA, two reverse primers, SofG-ALFA and ALFA-Stop-His are used. SofG-ALFA reverse primer has the first half of the ALFA sequence and ALFA-Stop-His has the second half of the sequence. Both reverse primers were designed in such a way that they have an 18-nucleotide overlapping sequence and a minimum of 18 nucleotides were maintained at either end for proper annealing to the template. The strategy for insertion of the whole ALFA sequence with the help of reverse primer is given in detail in section 2.2. Primers used for the cloning of the SofG-ALFA construct are tabulated below in table 2.1.

Primers were also designed for inserting His tag in between SofG and ALFA for purification of SofG-ALFA. Two reverse primers which were used are His-ALFA

rev. primer and ALFA-vector. They have 18-nucleotide overlapping sequence and 18 nucleotides on either end for annealing to template. The sequences for these primers are shown in table 2.1.

(c) Non-hydrolysing mutant of ParM

ParM protein hydrolyses ATP during polymerisation. ParM filaments shows dynamic instability (Garner et al., 2004). Filament conformation of ParM will be maintained as long as ATP remain in excess. Once ATP finishes or hydrolyse completely, ParM filament will disassembles and change to monomeric state. We need ParM filament for helical symmetry. In order to keep the ParM filament stable, ATP hydrolysis has to be stopped. ATP hydrolysis can be stopped in two ways: (a) by mutating the residues of ParM which hydrolyses ATP, (b) by using a non-hydrolysable analog of ATP (AMP-PNP). Glu 148 (E148) and Asp 161 (D161) residues of ParM are considered to be reacted with gamma phosphate of ATP and thus responsible for hydrolysis. In order to make non-hydrolysing mutant of ParM the following residues were mutated to alanine (Orlova et al., 2007). Forward primer with E148A mutation was designed for preventing the hydrolysis as tabulated below in table 2.1.

Table 2.1 List of primers

No.	Primer name	Sequence (5'-3')	Length
1.	NbALFA-Lnk5-ParM Rev	GAATACCAACATGGAACCCGGAGAGCCTGAAGACACCGT	39
2.	NbALFA-TEV-GSP-ParM Rev	TACCAACATCGGAGAGCCACCTTGGAAGTACAGGTTCTCTGAAGACACCGTCACTTG	57
3.	ParM E148A FP	GAGTTAGATTCTTTATTAGCGATAGATCTCGGGGGCACC	39
4.	T7 forward	TAATACGACTCACTATAGGG	20
5.	T7 reverse	GCTAGTTATTGCTCAGCGG	19
6.	SofG-ALFA Rev 1	GCGGCGGCGCAGTTCTTCTTCCAGGCGGCTCGGTCGCCCTTCTCCGCTGCG	51
7.	ALFA-STOP-His Rev 2	ATGATGATGATGGGATCCTTATTCGGTCAGGCGGCGGCGCAGTTCTTC	48
8.	His-ALFA reverse	GCGGCGGCGCAGTTCTTCTTCCAGGCGGCTCGGATGATGATGATGATGATG	52
9.	ALFA-vector	GTGGTGGTGGTGAAGCTTTTATTCGGTCAGGCGGCGGCGCAGTTCTTC	49

2.2 Cloning

(1) Cloning of Megabody constructs:

NbALFA-GSP-ParM (Nb3) megabody construct was already cloned in pHis17 ampicillin resistant vector.

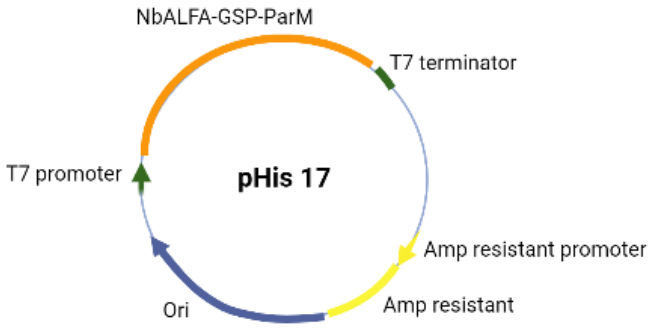


Fig.2.1. Representation of cloned NbALFA-GSP-ParM protein in pHis17 Amp^R plasmid.

2.2.1. Restriction-free cloning (RF cloning) for megabody constructs

All the megabody constructs were cloned by restriction free (RF) cloning (Ent et al., 2006). Both the megabodies Nb5 and Nb10 were cloned by using Nb3 as a template.

In the first step, Nb5 or Nb10 gene was amplified using vector-specific forward primer (T7 forward) and specific reverse primer. Reverse primers were designed in such a way that they could add the extra sequences after the linker in the megabody Nb3. It consists of the last sequence of NbALFA, bases corresponding to codons for GSP residues, the sequence to be added, and the initial part of ParM. So, the PCR product in the case of Nb5 is T7-NbALFA-GSPGS-$ParM_{initial}$. PCR product in the case of Nb10 is T7-NbALFA-ENLYFQG-GSP-initial sequence of ParM. The PCR product was then verified on an agarose gel. If the PCR product showed band of expected size, then it was purified with a Qiagen PCR purification kit. This purified product was used as a megaprimer for the second PCR to amplify the whole plasmid.

Then in RF cloning, this PCR product was used as a megaprimer to insert NbALFA-GSPGS/ENLYFQGGSP sequence in between the T7 promoter and ParM sequence in $ParM^{WT}$ template. The concentration of template and megaprimer used were different for both constructs. Concentrations are mentioned at the tables 2.3 and 2.4

for each construct and also the extension time. The products were checked in agarose gel in order to ensure cloning worked. The product was then DpnI digested for 3-5 hrs and transformed into NEB Turbo electro-competent cells and incubated at 37˚C for 12 hrs. Then colonies were screened by colony PCR or restriction digestion by NdeI and BamHI enzymes for knowing the correct insert band. All clones are verified by sequencing.

Fig.2.2 shows the cloning for all of the megabody constructs.

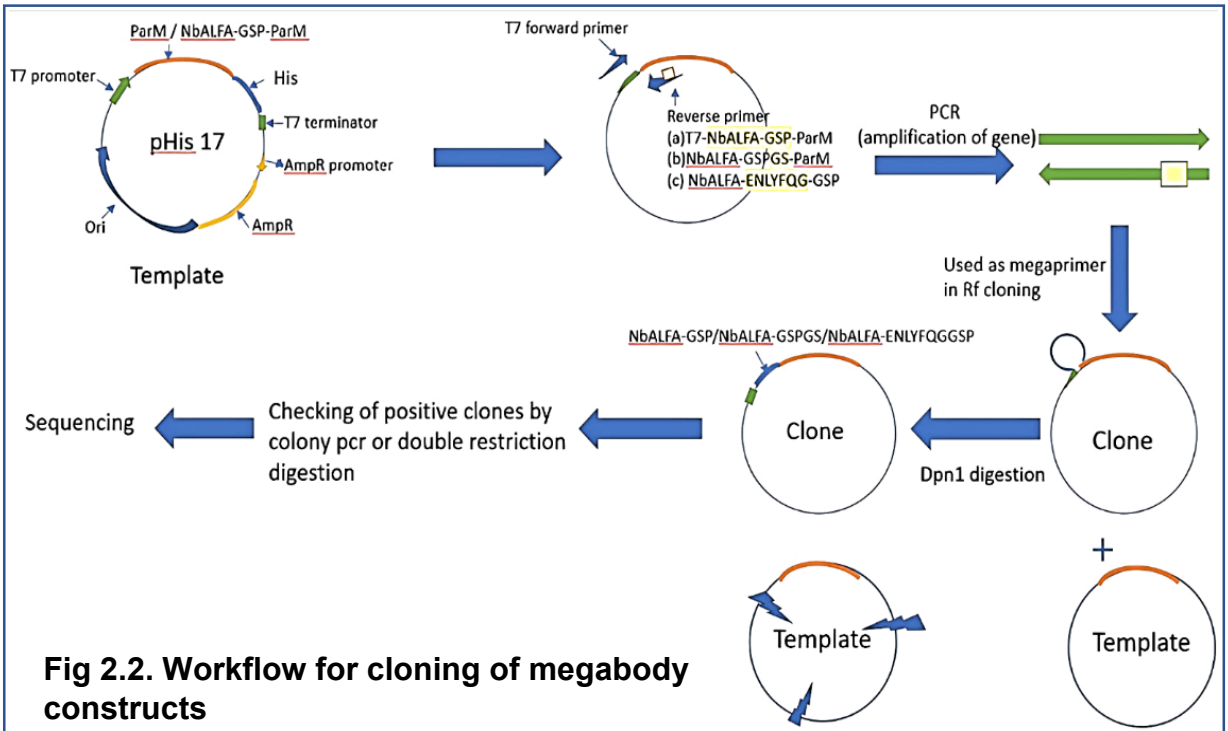


Fig 2.2. Workflow for cloning of megabody constructs

Table 2.2 PCR components and conditions

Components	Volume(µl)
Template(NbALFA-GSP-ParM=116.9ng/µl)	1
dNTPs	2
Pfu buffer	5
T7 forward primer	0.75
NbALFA GSPGS/TEV-GSP rev primer	0.75
Pfu polymerase	1
MiliQ	39.5

PCR conditions

Initial denaturation	95◦C (5 min)	
Denaturation	95◦C (30 sec)	×35
Annealing	56◦C (45sec)	
Extension	72◦C	
Final extension	72◦C (5 min)	
Hold	4◦C	

The reverse primer used is specific for the each megabody construct (section 2.1). Both the constructs are purified with same pcr conditions except extension time.

Table 2.3 (a) RF components of Nb5

Components	Volume(µl)	
	Test	Control
Template(Wt. ParM plasmid=119ng/µl)	2.1	2.1
Megaprimer(PCR2 product=T7-NbALFA-GSPGS-ParM,197ng/µl)	5.076	-
dNTPs (2.5mM)	3	3
Pfu buffer(10x)	5	5
Pfy polymerase	1.2	1.2
MiliQ water	33.624	38.7

(b) RF PCR conditions for Nb5

Initial denaturation	95◦C (5 min)	
Denaturation	95◦C (30 sec)	
Annealing	56◦C (1 min)	×35
Extension	72◦C (4min)	
Final extension	72◦C (10 min)	
Hold	4◦C	

Table 2.4 (a) RF components of Nb10

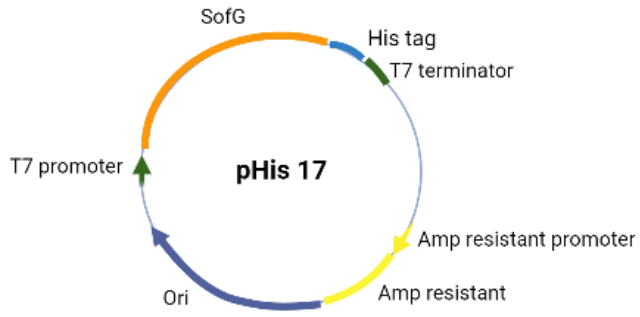
Components	Volume(µl)	
	Control	Test
Template(Wt. ParM plasmid=430ng/µl)	0.6	0.6
Megaprimer(pcr product=T7-NbALFA-TEV-GSP-ParM,168ng/µl)	-	6
dNTPs (2.5mM)	3	3
Pfu buffer(10x)	5	5
Pfu polymerase	1.2	1.2
MiliQ	40.2	34.2

(b) RF PCR conditions for Nb10

Initial denaturation	95◦C (5 min)	
Denaturation	95◦C (30 sec)	
Annealing	56◦C (1 min)	×35
Extension	72◦C (4 min)	
Final extension	72◦C (10 min)	
Hold	4◦C	

(2) Cloning of SofG-ALFA construct:

The SofG protein was already cloned into the pHis17 vector with a hexa-histidine tag at the C-terminal of the protein (SofG-6xHis) (Figure 2.3).



Template

Fig.2.3 Representation of cloned SofG-ALFA protein in pHis17 plasmid.

2.2.1. Restriction-free cloning (RF cloning) of SofG construct

SofG-ALFA was also cloned by RF cloning. Firstly, the SofG-ALFA gene was amplified using the SofG-6xHis as the template with the vector-specific forward primer (T7 forward) and the two reverse primers, SofG-ALFA and ALFA-stop-His. In PCR1, the first half of ALFA sequence was inserted after SofG by using the SofG-ALFA reverse

primer and T7 forward primer. The PCR1 product is T7-SofG-first half of ALFA. In PCR2, the second half of ALFA sequence was inserted by ALFA-stop-His primer. Both reverse primers are designed in such a way that they have an overlapping sequence. So in PCR2, PCR1 product will be extended. Now at the end of PCR2, the whole ALFA sequence will be inserted after the SofG gene. The PCR2 product was then verified by an agarose gel and purified with a Qiagen PCR purification kit.

This purified PCR2 product was then used as a megaprimer in RF cloning to amplify the whole plasmid. In RF cloning, the Fib protein gene in pHis17 Kan^R plasmid is used as a template where SofG-ALFA will replace the Fib gene in order to make the plasmid kanamycin resistant. So, the RF product would be the SofG-ALFA gene in the pHis17 Kan^R vector. The further steps will be the same as for megabody.

Fig.2.4 shows cloning for SofG-ALFA.

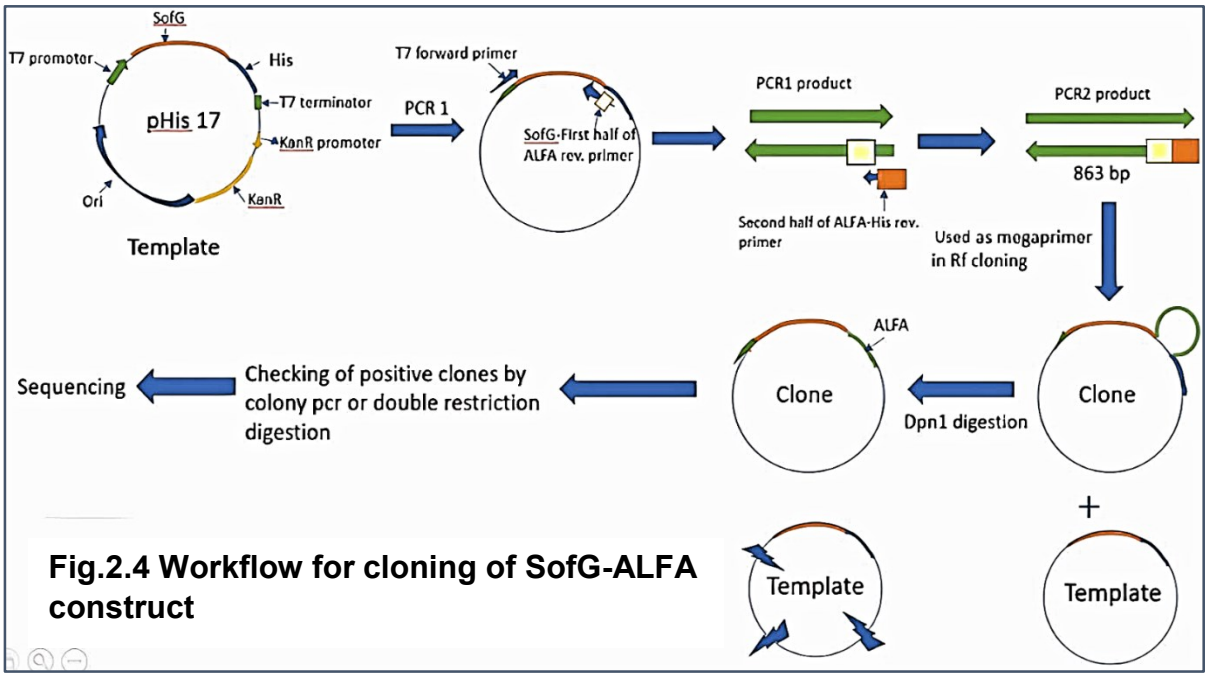


Fig.2.4 Workflow for cloning of SofG-ALFA construct

SofG-ALFA cloning

Table 2.5 (a) PCR1 components

Components	Volume(µl)
Template (SofG-pHis plasmid,86.7ng/µl)	1.5
dNTPs (2.5mM)	2
Pfu buffer (10x)	5
T7 for primer (20µM)	0.75
SofG-alpha rev. primer (20µM)	0.75
MiliQ	39
Pfu polymerase	1

(b) PCR1 conditions

Initial denaturation	95◦C (5 min)	
Denaturation	95◦C (30 sec)	
Annealing	58/60◦C (30 sec)	×35
Extension	72◦C (50 sec)	
Final extension	72◦C (5 min)	
Hold	4◦C	

Table 2.6 (a) PCR2 components

Components	Volume(µl)
Template (SofG-pHis plasmid,86.7ng/µl)	1.5
dNTPs (2.5mM)	2
Pfu buffer (10x)	5
T7 for primer (20µM)	0.75
SofG-alpha rev. primer (20µM)	0.75
MiliQ	39
Pfu polymerase	1

(b) PCR2 conditions

Initial denaturation	95◦C (5 min)	
Denaturation	95◦C (30 sec)	
Annealing	58/60◦C (30 sec)	×35
Extension	72◦C (52 sec)	
Final extension	72◦C (5 min)	
Hold	4◦C	

Table 2.7 (a) RF components

Components	Volume(µl)	
	Test	Control
Template(Fib plasmid, 270ng/µl)	3	3
Megaprimer(PCR2 product,70.7ng/µl)	14.5	-
dNTPs (2.5mM)	3	3
Pfu buffer(10x)	5	5
Pfu polymerase	1.2	1.2
MiliQ water	23.3	37.8

(b) RF PCR conditions

Initial denaturation	95◦C (5 min)	
Denaturation	95◦C (30 sec)	
Annealing	56◦C (1 min)	×30
Extension	72◦C (3 min 25sec)	
Final extension	72◦C (10 min)	
Hold	4◦C	

However, the cloned SofG-ALFA construct has a stop codon before His tag. Further cloning was done in order to insert His tag in between the SofG protein gene and the ALFA sequence by RF cloning. His tag is needed for the purification of SofG-ALFA by affinity chromatography which will be discussed in detail in chapter 3.

SofG-His-ALFA gene was amplified using vector-specific forward primer (T7 forward) and the two reverse primers, His-ALFA and ALFA-vector. SofG-His was used as a template. The PCR product is T7-SofG-His-ALFA. The PCR product was then verified by an agarose gel and purified with a Qiagen PCR purification kit. This purified PCR product was then used as a megaprimer in RF cloning to amplify the whole plasmid. RF cloning didn't work in this case.

2.2.2 Restriction digestion method

The Restriction digestion method was used to clone the His-SofG-ALFA construct since RF cloning was not successful for this construct.

Cloned SofG-ALFA plasmid and template N-His FrzB plasmid digested by Nde1 and BamH1 enzymes (since the restriction sites of these two enzymes are at the two ends of the gene). The whole digestion products were then loaded onto an agarose gel and ran till all bands are resolved. The bands of interest SofG-ALFA gene and N-His vector were cut out from the gel and the DNA was purified using a Qiagen gel extraction kit.

The digested plasmid was then gone through a TSAP treatment for 40-45 minutes at 37°C to avoid any self-ligation of the plasmid. The TSAP enzyme was later denatured by heating the sample at 75°C for 10 min. Both the digested products were then incubated together (100 ng of vector + 500 ng of insert) with T4 DNA ligase for 10 hrs at 16°C. The ligation product was then directly transformed into NEB-turbo electro-competent cells by electroporation. Clones selected by restriction digestion were also verified by sequencing.

2.3 Protein overexpression and solubility check

(a) Megabodies

For the overexpression of proteins, all the megabody constructs were first transformed in *E-coli* BL21AI strain by heat shock and plated on ampicillin containing plate. The cells are incubated at 37˚C for 12 hrs. Only the cells where megabody plasmids are inserted will be grown since it has ampicillin resistance. Then few colonies are inoculated in ampicillin-containing lb broth and incubated at 37˚C till the OD reached 0.6. Once it reached 0.6, they were induced with 0.2% arabinose and kept at 18˚C for overexpression. After 12-16 hrs, the cells were pelleted down. The 5 ml pellets were re-suspended in 500 µl lysis buffer and were sonicated to release the contents of the cell. The total fraction and the soluble fraction were taken for loading on SDS gel for electrophoresis. The proteins were visualized on the SDS gel.

Similarly, we did an expression check in SHuffle strain as it is a strain that correctly folds the proteins which have disulphide bonds (Lobstein et al., 2012, Ke et al., 2014, Ren et al., 2016). NbALFA has disulphide bonds in its structure (Götzke et al., 2019). These cells have different optimal temperatures for expression. Cells after transformation were incubated at 30˚C for 12 hrs. Pre-induction and post-induction temperatures were 30˚C and 16˚C respectively. These cells were induced with 0.4 mM IPTG.

(b) SofG-ALFA

Expression of SofG-ALFA construct was done in a similar way as megabody construct except here since SofG-ALFA construct was cloned in kanamycin-resistant plasmid, kanamycin was used in place of ampicillin as antibiotic.

(d) Co-purification:

All the megabodies are co-purified with the SofG-ALFA construct in the following two ways:

(i) By co-transformation of both plasmids with different antibiotic resistance.
(ii) Co-purification by mixing the lysates before sonication/after sonication.

(i) By co-transformation of both plasmid

Firstly, we tried co-expression by inserting two plasmids with different antibiotic resistance within the same cell. We transformed plasmids, megabody gene pHis17 Amp^R and SofG-ALFA pHis17 Kan^R in BL21AI strain *E. coli* cells and incubated at 37˚C for 12 hrs. Only the cells where both the plasmids are inserted will grow. Then few colonies were inoculated in LB broth (where both kanamycin and ampicillin antibiotics were added) and incubated at 37˚C till the OD reached 0.6. Further steps are the same as for constructs alone.

(ii) By mixing the lysates before sonication/after sonication.

The proteins can also be copurified after mixing the lysates. We have checked the co-expression by mixing the lysates both before and after sonication. Here, the proteins were grown individually and pelleted down. They were resuspended in 500 µl lysis buffer. 250 µl of each was used mixed with each other (volume ratio = 1: 1) and the remaining 250 µl was used as a control. Co-expression was also checked with 1: 2 (megabody: SofG-ALFA) volume ratio to increase the expression of SofG-ALFA. Their interaction was checked by pelleting assay.

2.4 Polymerization/ Pelleting assay

The plasmids with megabody insert or SofG-ALFA insert were transformed in E. coli BL21-AI cells or SHuffle cells. The cells were then grown in 10 ml LB broth and induced with 0.2% arabinose in BL21AI or 0.4 mM IPTG in SHuffle strain. 10 ml pellet was then resuspended in 500 µl lysis buffer, sonicated and centrifuged to get the soluble lysate. This soluble cell lysate was used for polymerization assays. 97 µl of this soluble cell lysate was then added with final concentration of 5 mM AMPPNP (non-hydrolysable

analogue of ATP) and 5 mM of $MgCl_2$ and centrifuged at 100,000 × *g*, 4˚C in a Beckman TLA-120.2 rotor for 25 min. After ultracentrifugation, pellet, wash and supernatant fractions were collected and loaded onto SDS–PAGE gel, to check for the presence of polymerized construct in the pellet fraction. The experiment was repeated with and without nucleotide. The polymerization assay was performed for the interaction of megabody and SofG-ALFA constructs and their expression of them separately as a control.

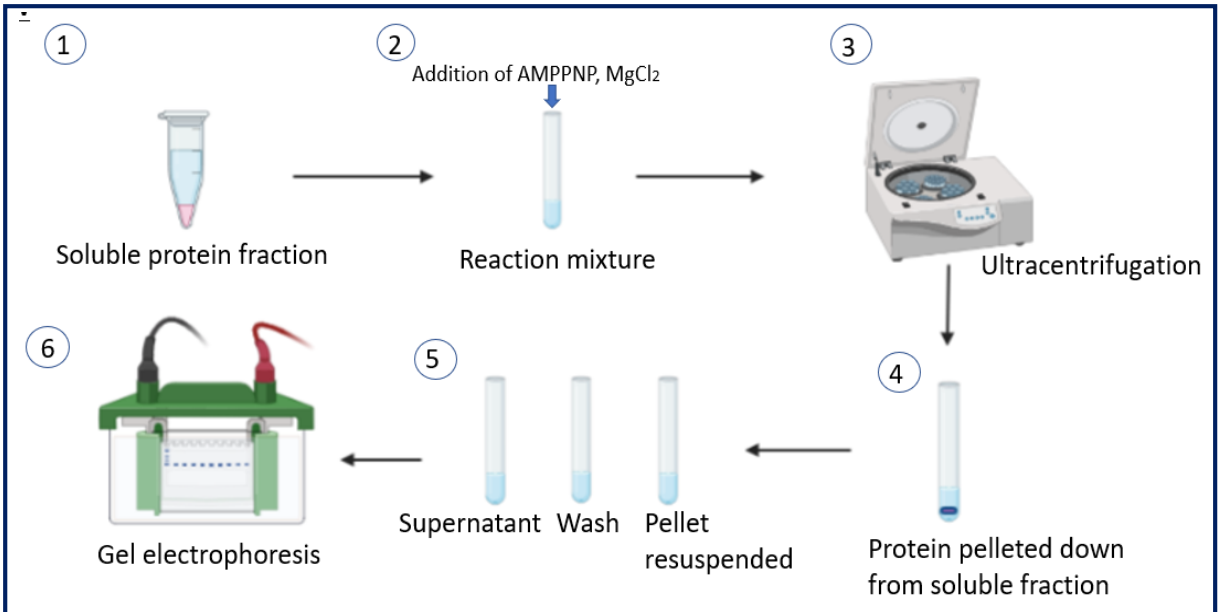


Fig. 2.5 Workflow for pelleting assay

2.5 Protein purification

The methods of purification of all the constructs are detailed in this section.

2.5.1 Affinity chromatography (for 6xHis-SofG-ALFA)

SofG-ALFA construct has been cloned with a Hexa-Histidine tag at the N-terminal of the protein to purify it with the help of Ni-NTA affinity chromatography (followed the protocol from paper: Kanade et al., 2021). For purification, *E. coli* BL21-AI cells transformed with plasmids were grown in 5 L of LB media with the same standardized condition for the overexpression of protein (as mentioned in section 2.3). Cultures were pelleted down by centrifugation at 5500 rpm for 25 min. The cell pellets were resuspended in 150 ml of lysis buffer and sonicated for 3 min (pulse: 1 sec on 3 sec off) with a medium probe at 60% amplitude twice at an interval of 5 min. Lysed cells

were then centrifuged at 18000 x *g*, 4°C for 40 min. 5 ml Ni-NTA column (HisTrap, GE Healthcare) was washed with water and equilibrated with buffer A (200 mM KCl + 50 mM Tris pH 8 + 25 mM Imidazole + 5 mM $MgCl_2$) for binding . The supernatant was then loaded into the column (HisTrap, GE Healthcare) pre-equilibrated with buffer A. Flow-through was collected simultaneously. After passing the entire supernatant through the column, the column was then passed with 30 ml of buffer A to remove any unbound protein. 30 ml each of 2%, 5% of buffer B (200 mM KCl + 50 mM Tris pH 8 + 500 mM imidazole + 5 mM $MgCl_2$) was passed through the column to remove non-specific bounded proteins. The bound proteins were then eluted with an increasing percentage of Buffer B from 10, 20, 50, to 100%. The imidazole in buffer B has more affinity towards the Ni-NTA and competes with the His-tag of the protein. Then 5 ml fraction of 30 ml each of 10%, 20%, 50%, and 100% of buffer B was collected. 15 µl of each eluted fraction was then mixed with 15 µl of 2x SDS dye and ran on a 12% SDS gel. The fractions that are the purest were then pooled and dialyzed in buffer (Tris 50 mM pH = 8, KCl 50 mM). The 35 ml of dialyzed protein was concentrated to 0.27 ml with a 10 kDa centricon (Sartorius, Vivaspin turbo 15) by centrifugation at 4000 rpm at 4°C. The protein after concentrating was then aliquoted into thin-walled PCR tubes and flash frozen, and stored at -80°C.

2.5.2 Ion exchange chromatography (for megabodies)

Since the megabody constructs do not have hexa-Histidine tag in the plasmid for purifying with affinity chromatography, they were purified by ion exchange chromatography, following the published protocol for ParM (Gayathri et al., 2012). Firstly, for the purification of megabody Nb5, the pI of protein was checked. The pI of Nb5 protein came out to be 5.32. We selected anion exchange chromatography for purification. Buffers with a pH higher than 5.32 (pH 8) were made in order to maintain a negative charge on the proteins. SHuffle cells transformed with Nb5 gene plasmid were grown in 2 L of LB media with the same standardized condition for the overexpression of protein (as mentioned in section 2.3). Cultures were then centrifuged at 5500 rpm for 25 min to collect the pellet. The cell pellets were resuspended in 60 ml of lysis buffer (each 1 litre in 30 ml lysis buffer) and sonicated for 3 min (pulse: 1 sec on 3 sec off) with a medium probe at 60% amplitude. Lysed cells were then centrifuged at 18000 x *g* for 50 min at 4 °C. The supernatant was then collected into a separate conical flask. All the tubing of the FPLC system (AKTA Prime)

were washed with water and then with the corresponding buffers at 30ml/min in waste mode. The supernatant was then loaded into the QHP column and pre-equilibrated with buffer A (50 mM KCl + 50 mM Tris pH 8). Flow-through was collected simultaneously. After passing the entire supernatant through the column, the column was washed with 30 ml of buffer A to remove any unbound protein. The bound proteins were then eluted with an increasing percentage of Buffer B (1000 mM KCl, 50 mM Tris pH 8). The percentage of buffer B will automatically increase since the gradient (0% to 50% B over 20 column volumes) is already set. 20 µl of each eluted fraction was then mixed with 20 µl of 2x SDS dye and ran on a 12% SDS gel. The fractions where the band was observed were collected and used for pelleting down ParM protein by polymerizing by ATP addition.

For pelleting ParM polymers, the final concentration of 5 mM ATP and 5 mM $MgCl_2$ was added to 25 ml of the soluble cell lysate and centrifuged at 100,000 × *g* in a Beckman TLA-120.2 rotor for 25 min. Pellet and supernatant fractions after ultracentrifugation were loaded onto SDS–PAGE gel, to check for the presence of polymerized construct in the pellet fraction. However, there was no pellet observed with naked eye. The supernatant was then used for concentrating the protein. The protein was concentrated with a 10 kDa centricon (Sartorius, Vivaspin turbo 15) by centrifugation at 4000 rpm at 4°C.

Concentrated protein was used for dialysis in buffer A. Dialysis helps in purifying the protein by removing the small, unwanted compounds in the solution by passing it through a semi-permeable membrane.

The dialyzed protein was then loaded into a MonoQ column. The column was pre-equilibrated with buffer C (Tris 50 mM pH 7.5 KCl 25 mM). 100 fractions were collected (one ml volume in each fraction). 10 fractions which show a peak at 280 nm were then run on a 12% SDS gel (20 µl of sample + 20 µl of 2 x SDS dye) to check the purity of the protein. 10 ml fraction was then again pooled and dialyzed in buffer C. The dialyzed protein was again concentrated to 1 ml with a Sartorius 10 kDa centricon. The protein after concentrating was then aliquoted into PCR tubes and flash frozen, and stored at -80°C.

2.6 Concentration estimation of protein

Concentration of purified Nb5 protein was checked by Bradford assay. For this BSA protein was taken as standard. Different dilutions of BSA protein (1 mg/ml) were made from 0, 0.1 to 1 in MiliQ. Similarly, Nb5 protein was diluted to 1/15, 1/20 in buffer (Tris 50 mM pH = 7.5, KCl 25 mM). After making dilutions, 5 µl of BSA and Nb5 protein dilutions were loaded onto the 96 well plate. 5 µl of only buffer was also taken as a control. 250 µl of Bradford reagent was added to all of these wells. The absorbance of each of the dilutions was then checked with the help of plate reader at 595 nm. The absorbance values of BSA standards were then used for making the concentration v/s absorbance curve (Fig. 2.6).

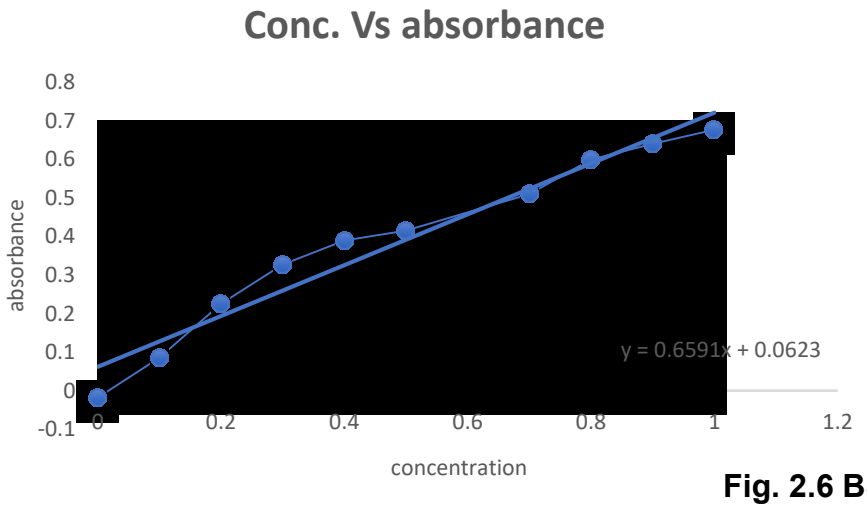


Fig. 2.6 BSA standard curve

The slope of the graph was used for calculating the concentration of purified Nb5 protein with the help of absorbance values of protein dilutions. The concentration of purified Nb5 protein came out to be 11.9 mg/ml.

Chapter 3. Results

The results of experiments performed for the interaction of SofG-ALFA with the helical filament scaffold are described in this chapter.

3.1. Cloning

The results of various steps of cloning all constructs are detailed in this section.

3.1.1. Cloning of megabodies

NbALFA-GSP-ParM (Nb3) construct was already cloned by a previous student in the lab. RF Cloning was done for the other two megabody constructs Nb5 and Nb10. Nanobody ALFA gene (Nb5, Nb10) was amplified using Nb3 as the template with T7 forward primer and gene-specific reverse primer. The reverse primer consists of the last sequence of NbALFA, GSP, the sequence to be added, and the initial part of ParM (Section 2.2.1). So, the PCR product in the case of Nb5 is the T7-NbALFA-GSPGS-initial sequence of ParM (485 bp). PCR product in the case of Nb10 is T7-NbALFA-TEV-GSP-initial sequence of ParM (491 bp). The PCR amplification of the gene with GS insert was performed at two different annealing temperatures (56˚C and 58˚C). A clear band below 500 bp (expected size of = 485 bp) was observed at both the annealing temperatures with equal intensity (Figure 3.1a). The PCR product is then used for RF cloning where wild-type ParM in pHis 17 vector was used as a template. The PCR product will anneal through T7 and ParM sequence to the template. Sample before RF cloning was taken in order to ensure on agarose gel if RF worked. A band around 4 kb at test sample (expected size = 4.9 kb) was observed in agarose gel (Figure 3.1b). Also, the band at T_0 (before PCR) around 500 bp was not present in T showing that megaprimer has been probably inserted into the template. Also, no band was observed in the control sample because of the absence of megaprimer. The RF product was then treated with Dpn1 enzyme for 3-5 hrs and transformed into NEB Turbo electro-competent cells and incubated at 37˚C for 12 hrs. Then randomly selected eight colonies were screened by colony PCR for verification of gene insertion. In colony PCR, wild ParM and Nb3 were used as templates in negative and positive control respectively with T7 forward and reverse primer. Out of the eight selected

colonies, four colonies showed the band (expected size = 1427 bp) the same as the positive control (Figure 3.1). These four were then verified by sequencing. Out of the four, one of them gave a positive clone.

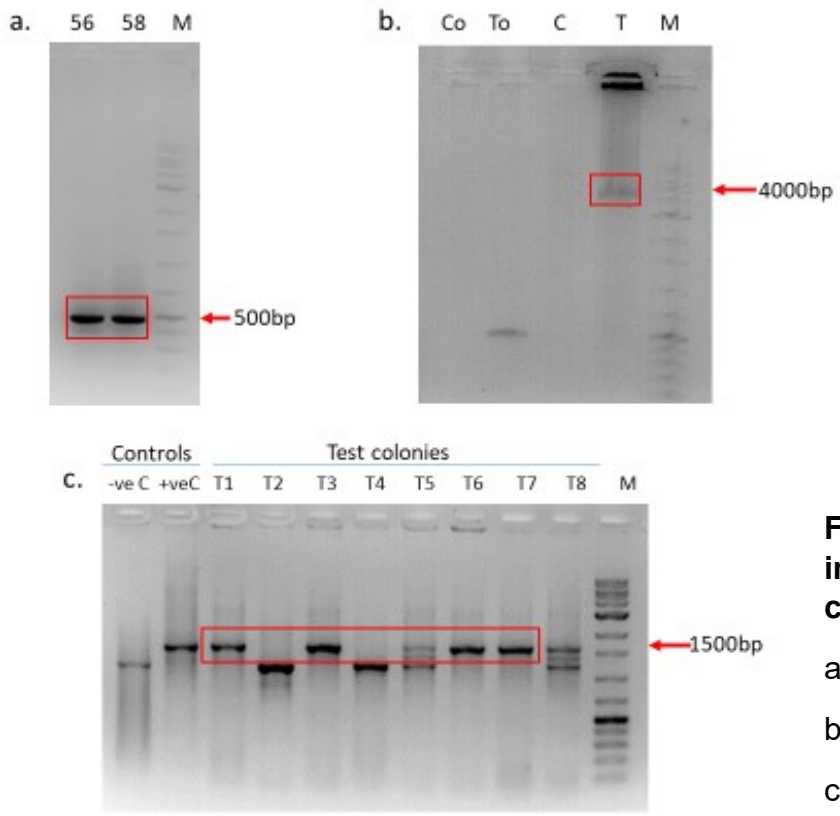


Figure 3.1 Agarose gel images of steps of RF cloning of Nb5.

a) PCR amplification.

b) RF PCR.

c) Colony PCR.

In a similar way, for Nb10, the PCR amplification with TEV insert was performed at two different annealing temperatures (56˚C, 58˚C). A clear band below 500 bp (expected size = 491 bp) was observed at both the annealing temperatures with equal intensity (Figure 3.2a). The PCR product was then used for RF cloning. In RF cloning PCR, a band around 4 kb at the test sample (expected size = 3.9 kb) was observed in agarose gel (Figure 3.1b). Also, band at T around 500 bp was absent when compared to T_0, suggesting that the insert has been incorporated, and no band is observed in the control sample.

The RF cloning PCR product was treated with DpnI and transformed into electro-competent cells. Five colonies were randomly selected, and verified by colony PCR. All five colonies showed a band around 1500 bp (expected size = 1592 bp) same as the positive control (Figure 3.2). Four clones were verified by sequencing. Out of the four, one of them gave a positive clone.

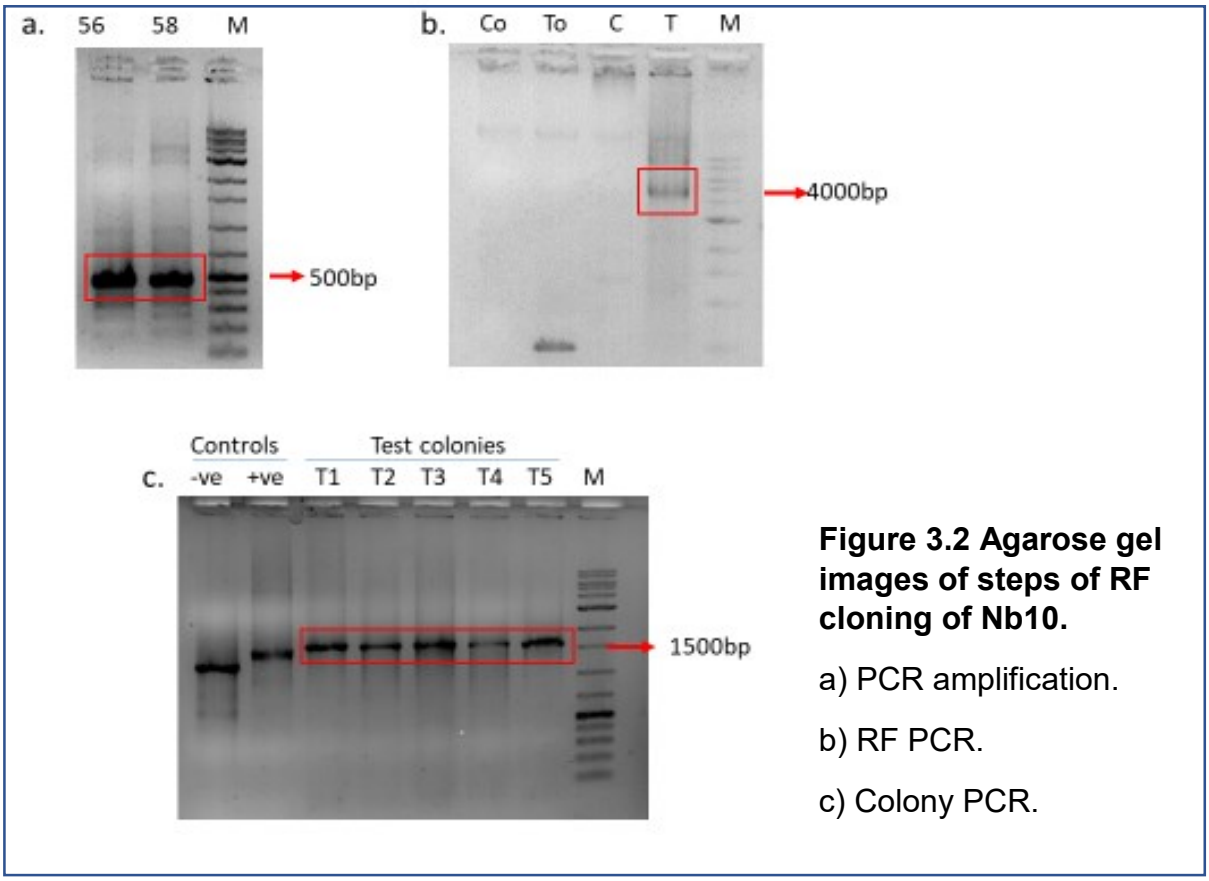


Figure 3.2 Agarose gel images of steps of RF cloning of Nb10.

a) PCR amplification.

b) RF PCR.

c) Colony PCR.

3.1.2 Cloning of SofG-ALFA

The SofG-ALFA construct was cloned by RF cloning. Firstly, the SofG-ALFA gene was amplified using the SofG-His6 as the template with the vector-specific forward primer (T7 forward) and the two reverse primers, SofG-ALFA and ALFA-stop-His respectively for two PCR reactions (Section 2.2.1). In PCR1, the first half of the ALFA sequence was inserted after SofG by using the SofG-ALFA reverse primer and T7 forward primer. The PCR amplification was performed at two different annealing temperatures 58˚C and 60˚C. The PCR1 product is T7-SofG-first half of ALFA (833 bp). A clear band around 800 bp was observed at both annealing temperatures (Figure 3.3a). PCR products obtained from both the annealing temperatures were purified with a Qiagen PCR purification kit and used as a template for the next PCR amplification for insertion of the second half of ALFA. Second PCR amplification was performed with T7 forward primer and ALFA-stop-pHis reverse primer. The PCR2 product was T7-SofG60-whole ALFA sequence-pHis (863 bp). This PCR amplification was performed at 58˚C, 60˚C out of which a bright band around 800 bp was observed at 60˚C (Figure 3.3b). Also, the band below 500 bp was observed, showing non-specific binding. 60˚C PCR

product with a prominent band around 800 bp was purified and used as a megaprimer in RF cloning. In RF cloning, a band above 3 kb for the test sample (expected size = 3.5 kb) was observed in agarose gel (Figure 3.3c). The band at T around 800 bp is negligible as compared to T_0 suggesting that the RF PCR worked. No band was observed in the control sample. Also, a light band was observed at C_0 (for plasmid) and T_0 (for megaprimer) before the PCR reaction.

The RF PCR product was treated with DpnI enzyme and transformed into electro-competent cells. Five colonies were randomly selected, and verified by colony PCR. In colony PCR, pHis17 Kan^R plasmid backbone with Fib insert and PCR1 product were used as templates in negative and positive control respectively with T7 forward and ALFA-pHis reverse primer. Out of the five colonies, colonies T2, T3 and T5 showed a band above 800 bp (expected size = 863 bp) slightly above positive control (Figure 3.3d). These three colonies were further checked by double digestion check. pHis17 Kan^R fib plasmid was used as a control. After digestion, bands around 2.5 kb (vector part) and 850 bp (insert) were observed for all digested test colony plasmids (Figure 3.3e). The control plasmid also showed a 2.6 kb band (vector part) and a 1.5 kb (fib insert) band. A double band was observed for the vector in the T5 colony. Therefore, only T2 and T3 test colonies were sent for sequencing. Out of these two, one of them gave a positive clone.

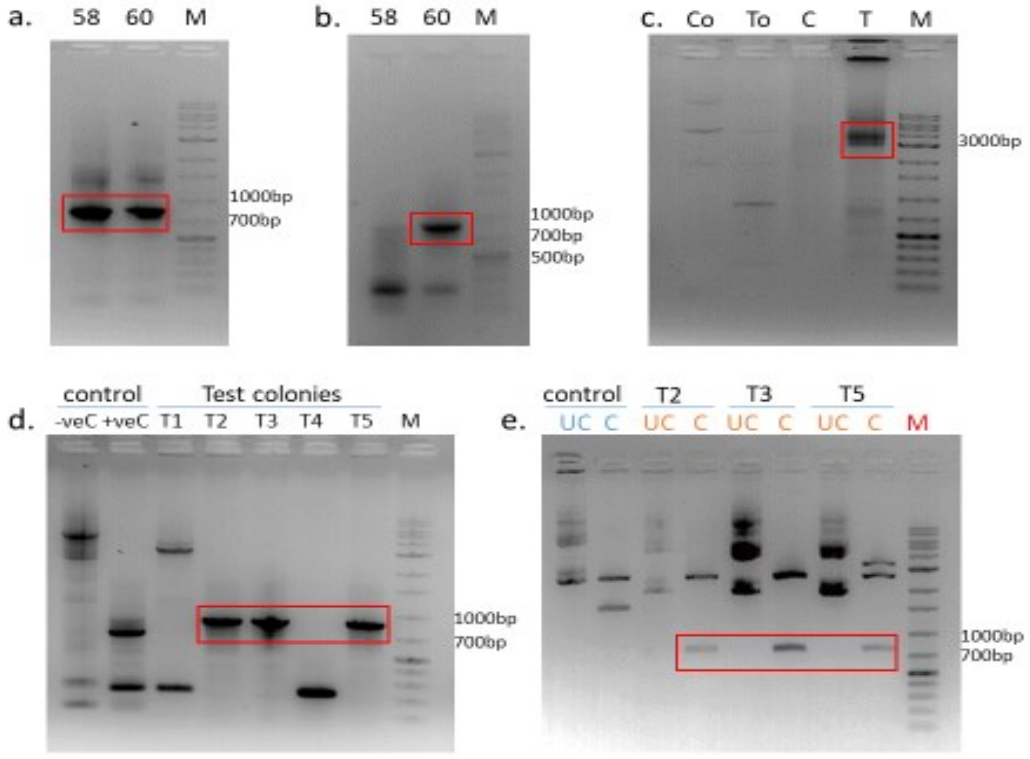


Figure 3.3 Agarose gel images of steps of RF cloning of SofG-ALFA.

a) PCR 1, b) PCR 2, c) RF PCR, d) Colony PCR, e) Double digestion with NdeI and BamH1.

For purification of SofG-ALFA, His tag was inserted in between SofG and ALFA by RF cloning. The RF cloning was done by using SofG-His as a template. His-ALFA and ALFA-vector were used as reverse primers and T7 as the forward primer (Section 2.2.1). In the first PCR, the first half of ALFA sequence was inserted with His-ALFA reverse primer. The PCR1 product is the T7-SofG-His-ALFA part of the primer (851bp). This PCR was performed at two annealing temperatures 58°C, 60°C and in both clear band near 850 bp was observed (Figure 3.4a). This product was then purified and used as a template in the second PCR to insert the second half of the ALFA sequence which was done with the help of ALFA-vector reverse primer. The PCR2 product is T7-SofG-His-ALFA (881 bp). Here also, a clear bright band of the expected size was observed at both the annealing temperatures (Figure 3.4b). PCR2 product was purified and used for RF cloning. Upon observing the RF PCR product on an agarose gel, a smear was observed (Figure 3.4c). This construct was then digested with DpnI enzyme and transformed into electrocompetent cells. Here 9 colonies are selected and verified by colony PCR. Two colonies showed the expected band. These two colonies were then sent for sequencing. However, both of them came out to be negative. Next, we tried cloning the construct with the restriction digestion method, from which we got the positive clone.

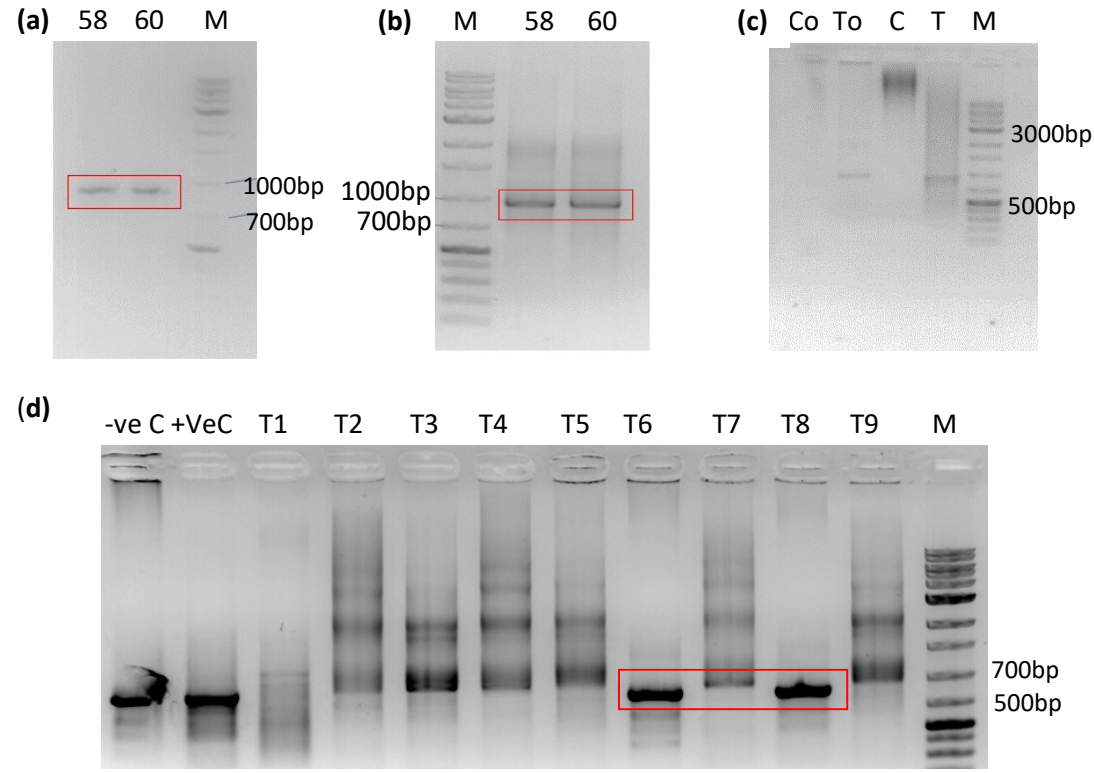


Figure 3.4 Agarose gel images of steps of RF cloning of SofG-H6-ALFA.

a) PCR 1, b) PCR 2, c) RF PCR, d) Colony PCR.

3.2 Protein over-expression and solubility check

All the constructs were transformed into the *E. coli* expression strain BL21-AI or SHuffle cells, and protein was over-expressed with the standardized conditions as mentioned in section 2.3.

3.2.1 Megabodies

(a) NbALFA-GSP-ParM (Nb3)

Nb3 construct was grown in both the BL21AI strain and the SHuffle strain. For the BL21AI strain, 10 ml culture of the cells was grown at 37˚C till OD reach 0.6. After reaching OD 0.6, half of the culture (5 ml) was induced with 0.2% arabinose (arabinose helps in the overexpression of protein), and the remaining half was kept uninduced. After induction, the culture was kept at 18˚C for 16 hrs. After 16 hrs, the culture was pelleted down.

For the SHuffle strain, 10 ml culture was grown till OD_{600} reach 0.6. At $0D_{600}$ reached 0.6, half of the culture was induced with 0.4 mM IPTG. After induction, the culture was grown at 16˚C for 18 hrs. The culture was also grown at 30˚C post-induction.

Along with it, wild type ParM was also grown with the same conditions as for megabody as control. The cells were processed for sonication, and the lysed cells were centrifuged to obtain the soluble fraction in the supernatant. Then for checking the expression total and soluble fraction of both induced (I) and uninduced (UI) were run through the SDS gel. Figure 3.5 shows the expression of megabody Nb3.

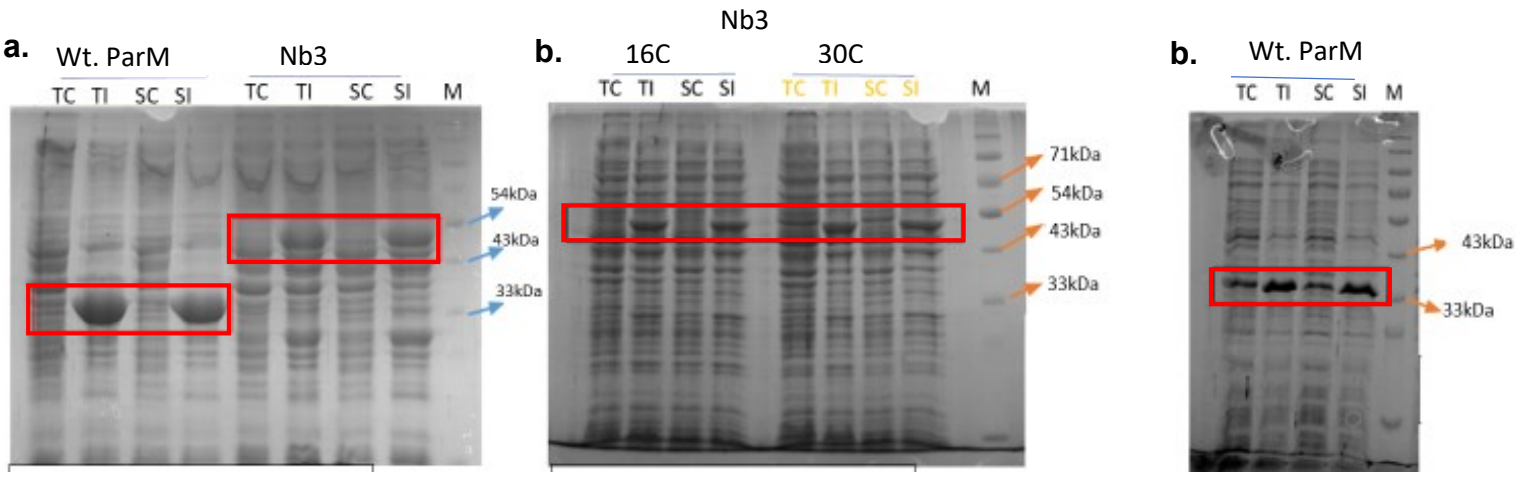


Figure 3.5 SDS gel image of expression check of Nb3. a) Expression in BL21AI strain, b) Expression in SHuffle strain.

Nb3 construct showed a bit less expression as compared to wild type ParM in BL21AI strain, however, it is the same in the SHuffle strain. Nb3 showed significant expression in both strains.

(b) NbALFA-GSPGS-ParM (Nb5)

Nb5 construct was grown with the same conditions as Nb3. Figure 3.6 shows the expression of megabody Nb5. Here no control was kept.

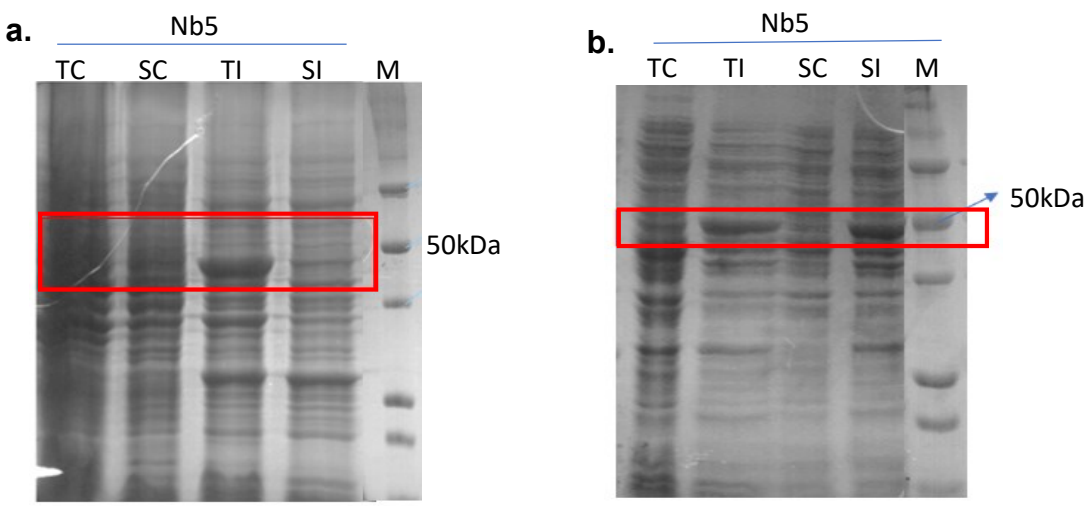


Figure 3.6 SDS gel image of expression check of Nb5. a) Expression in BL21AI strain, b) Expression in SHuffle strain.

Nb5 showed more solubility in the SHuffle strain as compared to the BL21AI strain (Figure 3.6). It showed very less solubility in the BL21AI strain.

(c) NbALFA-TEV-GSP-ParM (Nb10)

Nb10 was also grown with the same conditions as other megabody constructs. Since the expression of Nb3 and Nb5 was good in SHuffle strain, Nb10 expression was only checked in SHuffle strain. The expression check of Nb10 in BL21AI strain is yet to be checked.

Figure 3.7 below shows the expression of Nb10 construct.

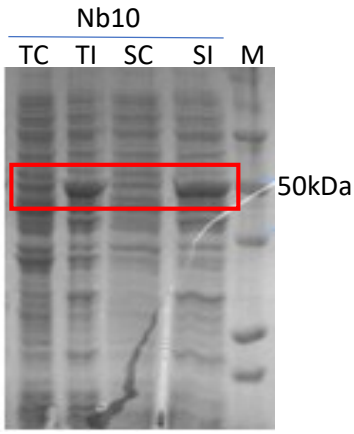


Figure 3.7 SDS gel image of expression check of Nb10.

Nb10 shows good solubility in SHuffle strain.

3.2.2 SofG-ALFA

His tagged SofG-ALFA (28.4 kDa) construct was grown in the BL21AI, BL21DE3, Rosetta and SHuffle strains.

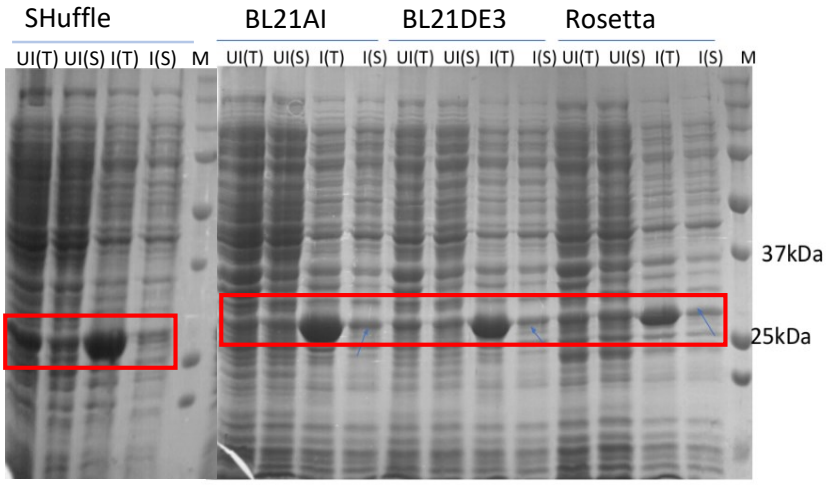


Figure 3.8 SDS gel image of expression check of 6xHis-SofG-ALFA in different strains.

His_6-SofG-ALFA construct shows almost the same solubility in all strains. Since the solubility of SofG-ALFA shows a bit more expression in the BL21AI strain as compared to other strains, the following construct was grown and purified in the BL21AI strain. However, the purification of the SofG-ALFA construct in the BL21AI strain was not successful.

3.2.3 Co-expression

The plasmid with megabody and SofG-ALFA insert were transformed into the BL21AI or SHuffle strain for co-expression. Only the cells transformed with both the plasmids will grow. The cells were then grown in LB broth with both antibiotics (Amp and Kan).

The co-expression of megabodies construct with SofG-ALFA construct are as shown below (Figure 3.9).

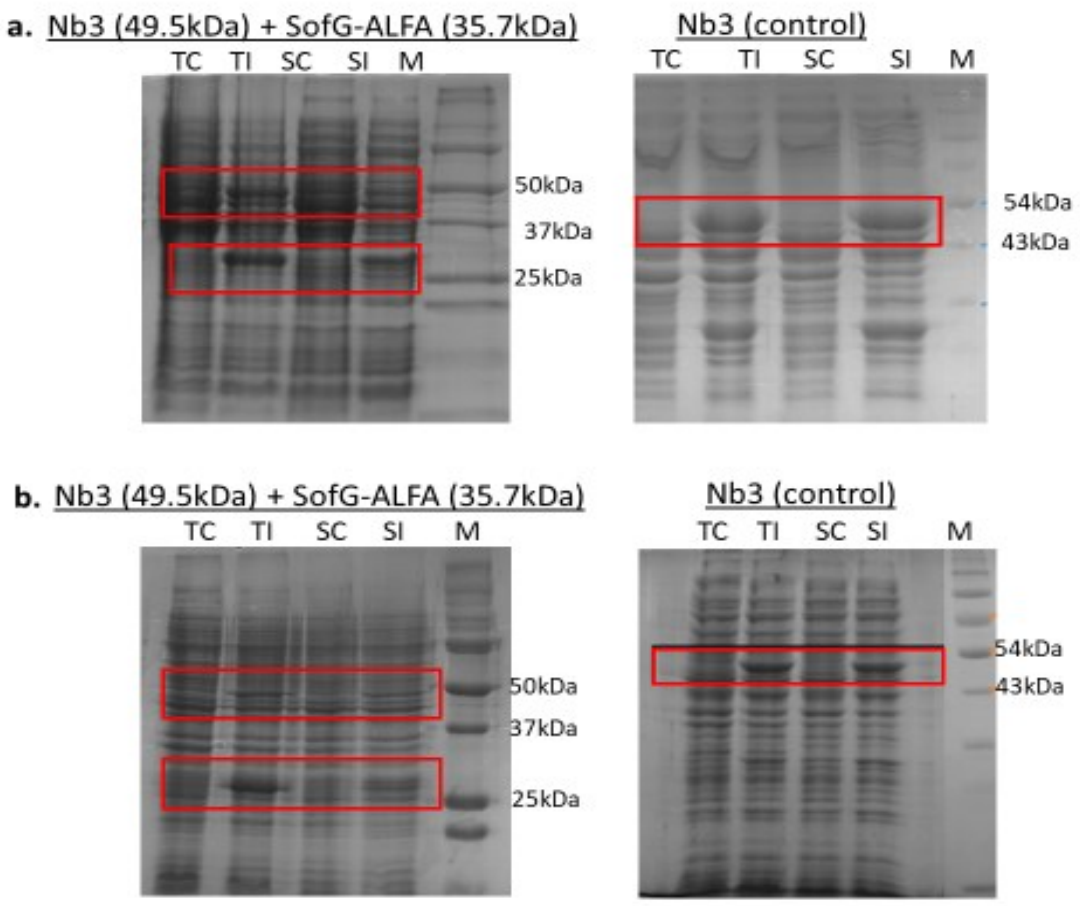


Figure 3.9 Co-expression of SofG-ALFA and NbALFA-GSP-ParM construct.

(a) In BL21AI strain.

(b) In SHuffle strain.

The expression of megabody Nb3 is getting less on co-expression with SofG-ALFA when it is compared with the control in both strains. This can be due to improper interaction between them or dilution of the construct. In order to see whether interaction causes the less expression of megabody or not, we have also checked the coexpression of (a) Nb3 and SofG (without ALFA tag to attach to Nb3), (b) ParM and SofG-ALFA (No Nb to attach to ALFA tag), where there will be no interaction (Figure 3.10).

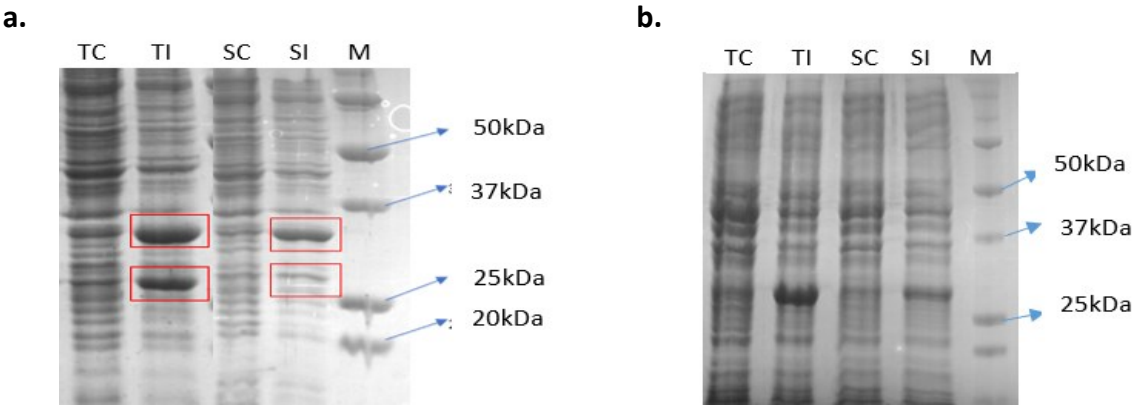


Figure 3.10

(a) **Expression of ParM with SofG-ALFA.**

(b) **Expression of Nb3 with SofG.**

In the co-expression of ParM (35.7 kDa) with SofG-ALFA (28.4 kDa), both proteins express as well as their individual expressions as they are not interacting (Figure 3.10a). This shows that it's the interaction that causes the decrease in expression of megabody expression on co-expressing with SofG-ALFA.

On the other hand, in the co-expression of Nb3 (49.5 kDa) with SofG (27.6 kDa), Nb3 expression gets negligible unfortunately (Figure 3.10b). This can be due to the cell incompatibility as here we are expecting good expression of both as they are not interacting. Next, we tried co-expression with Nb5. The figure below shows the co-expression of Nb5 with SofG-ALFA.

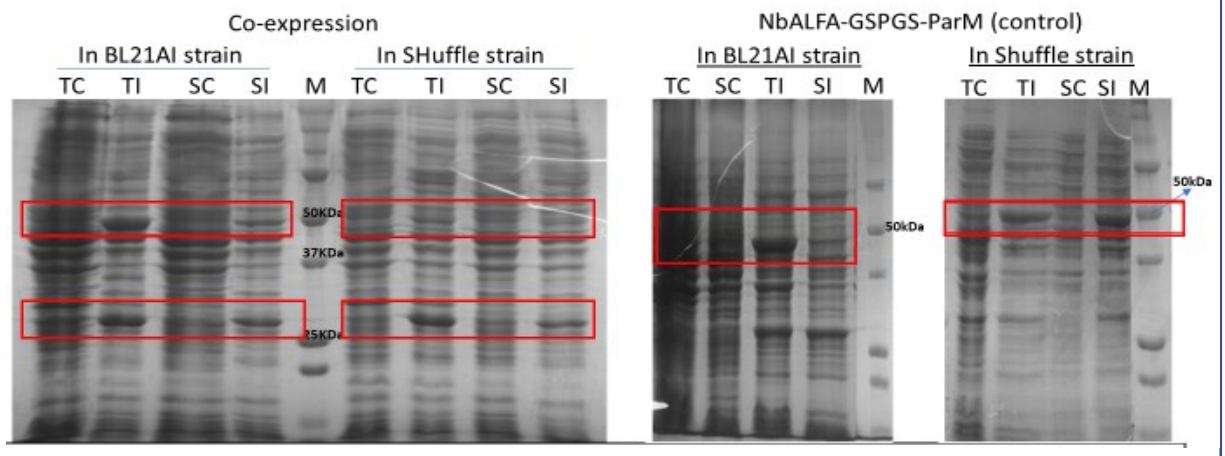


Figure 3.11 Expression of Nb5 with SofG-ALFA in BL21AI and SHuffle strain

Nb5 expression decreases on co-expressing with SofG-ALFA construct in only the SHuffle strain, not in the BL21AI strain (Figure 3.11). Nb5 expression is already less in the BL21AI strain.

Since the expression of megabodies was getting less on co-expression, the proteins were then attempted to co-purify by mixing the soluble lysates.

3.3 Pelleting assay

The polymerization assay was performed for all megabodies, SofG-ALFA, and their interaction in both BL21AI and SHuffle strains was monitored. 10 ml culture was grown for megabody and SofG-ALFA separately. Pelleting assay for all constructs was performed by the addition of 5 mM AMPPNP and 5 mM $MgCl_2$. The experiment was repeated with and without nucleotide. After ultracentrifugation, pellet, wash, and supernatant fractions were loaded onto SDS–PAGE gel, to check for the presence of polymerized construct in the pellet fraction (Section 2.4).

3.3.1 Megabodies

Megabodies are monomeric. As ParM polymerises in presence of AMP-PNP, megabody should also polymerize. The figure below shows pelleting assay gel image for Nb3 megabody construct.

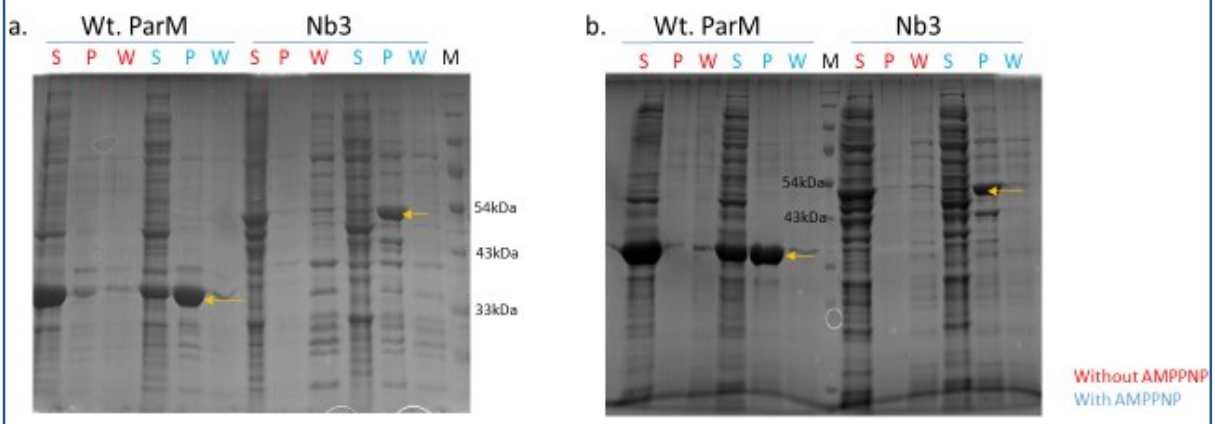


Figure 3.12 Pelleting assay gel images of Nb3:

(a) In BL21AI strain, (b) In SHuffle strain.

Nb3 is showing significant pellet in the presence of AMPPNP and not without AMPPNP. However, the expression of Nb3 in pellet fraction is less as compared to ParM. Thus, Nb3 could polymerize in both strains significantly in the presence of AMPPNP (and not without AMPPNP) but less as compared to wild type ParM showing nanobody insertion is affecting the polymerization of ParM, in the megabody (Figure 3.12).

Pelleting assay for Nb5 and Nb10 was done as a control in co-expression (section 3.3.3).

3.3.2 SofG-ALFA

SofG-ALFA protein is known to be monomeric. It does not polymerize. If SofG interacted with the megabody, it will be present in the pellet on binding with megabody. Thus, its pelleting assay was not done separately but as a control in co-expression (section 3.3.3).

3.3.3 Megabody + SofG-ALFA

(a) By co-transformation of plasmids

Both the plasmids were co-transformed within the cells. The figure below (Figure 3.13) shows the co-expression of megabodies with SofG-ALFA by co-transformation.

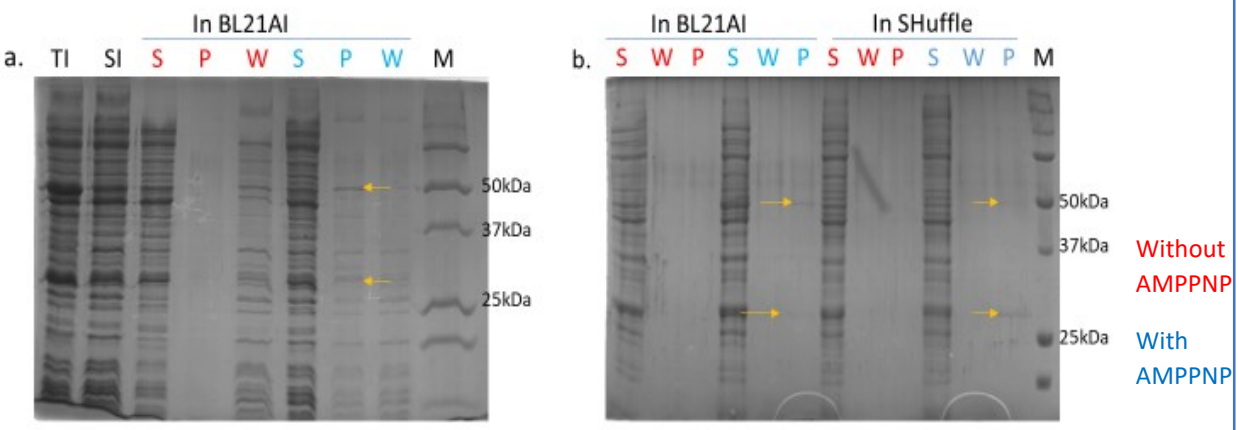


Figure 3.13 Pelleting assay gel images of megabody with SofG-ALFA: (a) NbALFA-GSP-ParM + SofG-ALFA, (b) NbALFA-GSPGS-ParM + SofG-ALFA

From figure 3.13, in Nb3 with SofG-ALFA and Nb5 with SofG-ALFA, both SofG-ALFA and megabody were observed in the pellet with AMP-PNP and not without AMP-PNP. Since SofG-ALFA is bound to Nb3, it should also show pellet in co-expression. However, their expression was very less and hence the polymerization assay results inconclusive. Therefore, we tried a different strategy to improve their expression i.e. mixing the lysates.

We also tried pelleting assay for ParM + SofG-ALFA to confirm whether its interaction is responsible for the decreased expression. Figure 3.14 shows the co-expression of ParM with SofG-ALFA.

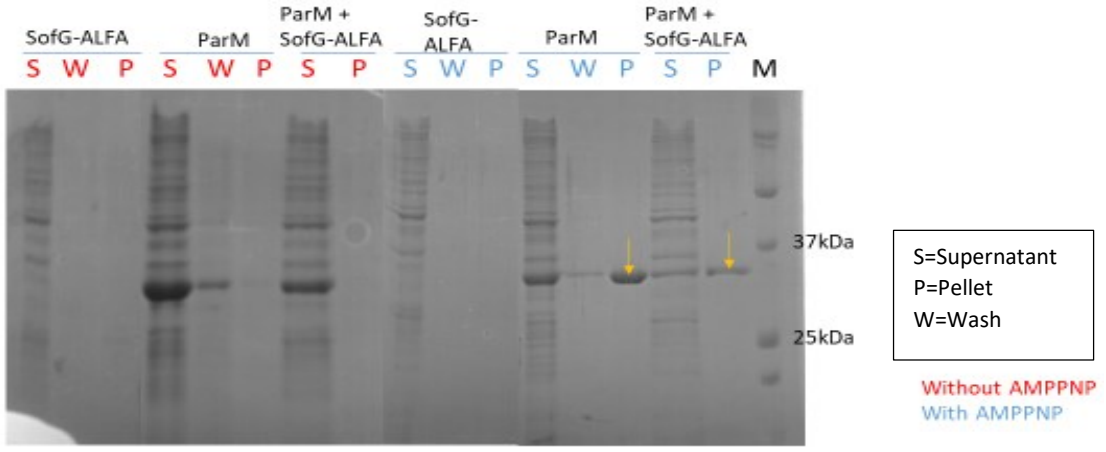


Figure 3.14 Pelleting assay gel image for co-expression of wild type ParM and SofG-ALFA

other hand, in the co-expression of Nb3 with SofG, only Nb3 should show the pellet and not SofG. As SofG does not have an ALFA tag so it will not bind.

(b) By mixing the lysates

Here, cultures were grown separately and pelleted down. These cultures were then resuspended in 500 µl lysis buffer and then mixed (250 µl SofG-ALFA + 250 µl Nb3). This mixed soluble lysate (250 µl SofG-ALFA + 250 µl Nb3) was then used for proceeding with pelleting assay to monitor their interaction. If SofG-ALFA interacted with Nb3, it will also be present in the pellet fraction. The figure below shows the co-expression of megabodies with SofG-ALFA by mixing the lysates (Figure 3.15).

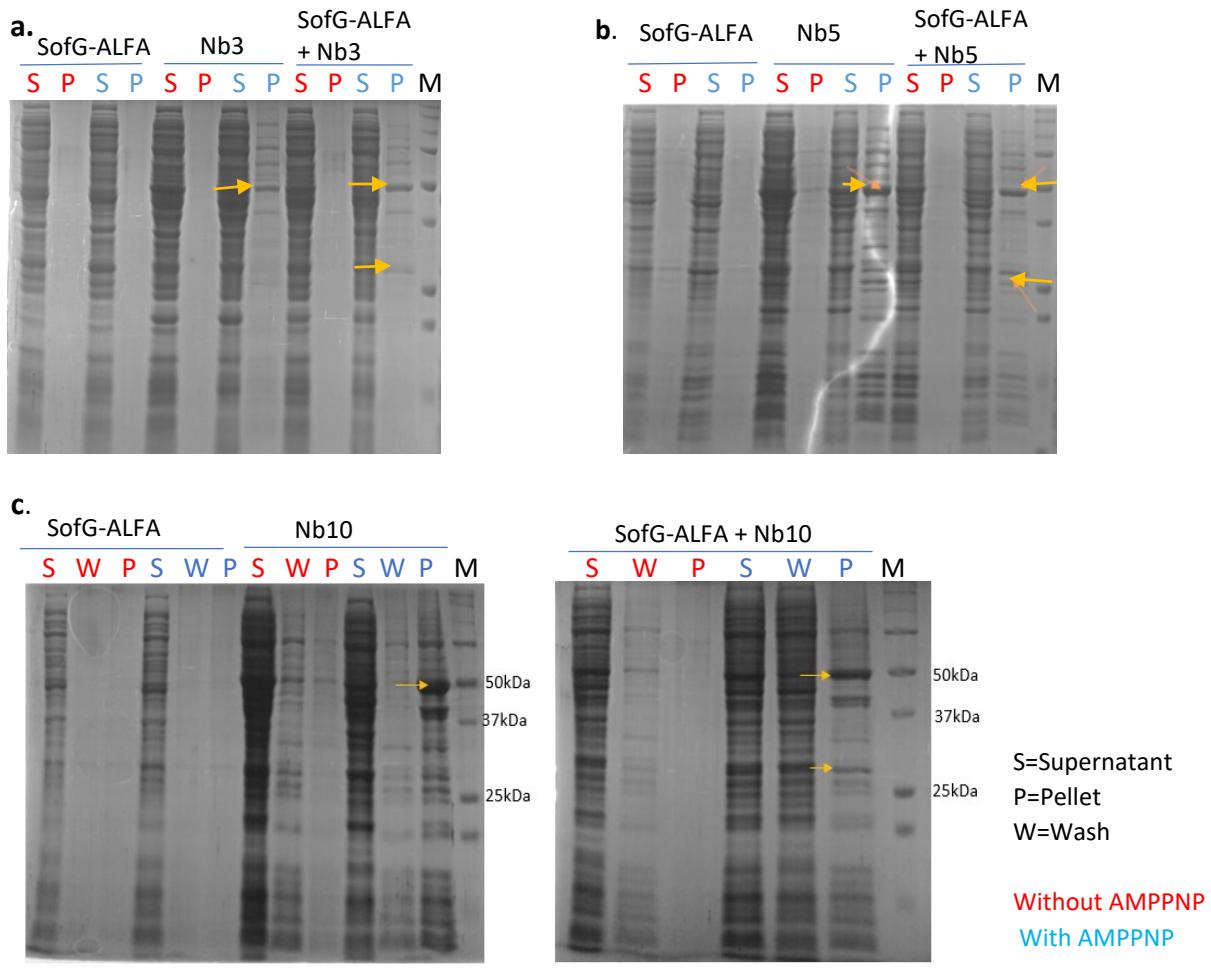


Figure 3.15 Pelleting assay gel image of megabodies with SofG-ALFA: (a) Nb3 + SofG-ALFA, (b) Nb5 + SofG-ALFA, (c) Nb10 + SofG-ALFA

On mixing the lysates, constructs showed more expression as compared to expression in co-expression by co-transformation. Expression of megabodies was not significantly

decreased on mixing with SofG-ALFA when compared to megabody alone (Figure 4.4). SofG-ALFA was present in the pellet fraction of the mixture and not in SofG-ALFA alone. This shows that SofG-ALFA is binding to megabody constructs.
However, there were still a lot of impurities observed in the pellet fraction. The pellet fraction should be clear only with the protein of interest. However, here some light bands are also observed which correspond to non-specific binding.
Therefore, we proceeded with the purification of proteins. Purification of proteins can help get rid of other proteins and pelleting assay with purified proteins can confirm their interaction.

3.4 Protein purification

To confirm the interaction of megabody with SofG-ALFA, both were purified separately. All details for purification are described below.

(a) Megabodies

Since megabodies did not have His tag, they were purified by anion exchange chromatography. For purification of Nb5, 2 litres of cultures were grown under the same conditions as for 10 ml culture (Section 2.3). Before proceeding with the purification, the expression of megabody Nb5 was checked to confirm that the construct is expressing. 2 L culture was then processed to 60 ml soluble lysate after centrifugation. The protein was been purified from this 60 ml of lysate. In the first round of purification, binding of protein to the Q HP column, washing, and elution of protein was carried out with increasing gradient of buffer B (Section 2.5). The 5 ml fractions of increasing percentage of buffer B were collected. 20 µl of each eluted fraction was then mixed with 20 µl of 2x SDS dye and run on a 12% SDS gel. Figure 3.16 shows the expression of some fractions of different percentages.

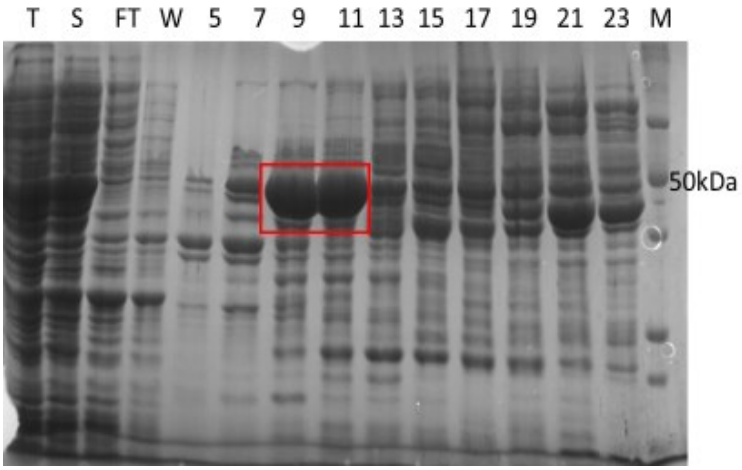


Figure 3.16 Purification profile of eluent protein (alternate fractions) from QHP column.

The molecular weight of Nb5 protein is 49.6 kDa which was observed in fractions 9, 11 at the positions corresponding to the overexpressed band in the total, and supernatant lanes of the cell lysate. No protein was observed in flow through which showed that the protein was successfully bound to the column. Therefore, column tubes 8, 9, 10, 11, 12 (25 ml) were selected for polymerizing the enriched fractions. In the pelleting assay, 25 ml of protein was added with 5 mM ATP and 5 mM $MgCl_2$. However, no pellet was observed after the ultracentrifugation. So, the supernatant was then concentrated and dialyzed. This dialyzed protein was passed through the MonoQ column where again the second round of purification takes place. The protein was bound to the column, the column is washed and the elution of protein with increasing concentration gradient of buffer B.

1 ml fractions were collected. 20 µl of each eluted fraction was then mixed with 20 µl of 2x SDS dye and ran on a 12% SDS gel (Section 2.5). Figure 3.17 shows the gel image of these fractions.

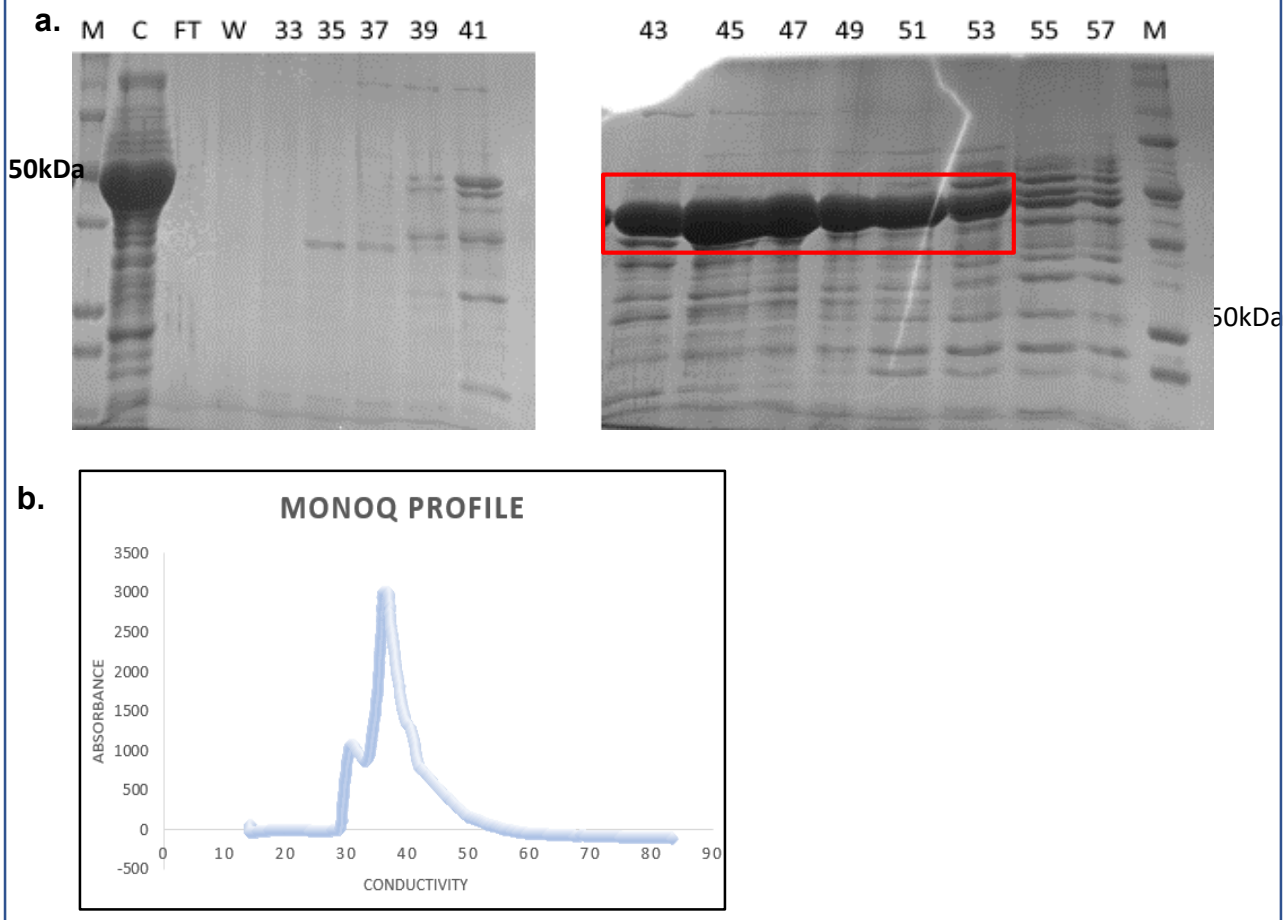


Figure 3.17 (a) Purification profile (alternate fractions) for elution from MonoQ column, (b) MonoQ profile

The protein was observed in fractions 43-53. No protein was observed in flow through Therefore, fractions 42-51 (10 ml) were selected for further purification. The selected fractions were then dialyzed in gel filtration buffer (Tris 50 mM pH 7.4, KCl 25 mM). After dialysis, we concentrated the protein into 1 ml. 1 ml concentrated protein was spun for 10-15 minutes. The supernatant was collected and aliquoted 20µl in each Eppendorf. The figure below shows the purity of the concentrated purified protein.

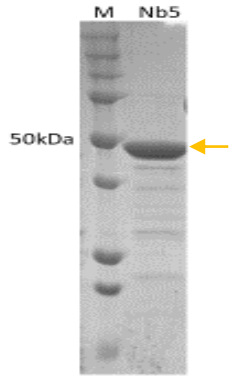


Figure 3.18 Expression of purified Nb5 protein.

In the figure 3.18, there were some light bands observed along with Nb5 protein band. However, the Nb5 protein was purified to reasonable purity, sufficient for the purpose of our assays.

(b) SofG-ALFA

SofG-ALFA was purified by affinity chromatography since it has His tag. For purification 5 L of culture was grown in BL21AI strain cells with the same conditions as for 10 ml culture (Section 2.3). The expression of SofG-ALFA was checked before purification. 5 L culture was then processed to 150 ml soluble lysate after centrifugation. The protein was been purified from this 150 ml of lysate. In purification, 150 ml of lysate was passed through the nickel-NTA column for binding. This column was then washed with buffer A followed by elution of protein with an increasing amount of buffer B. Then 5 ml fractions of 30 ml each of 10%, 20%, 50%, and 100% of buffer B were collected. 15 µl of each eluted fraction was then mixed with 15 µl of 2x SDS dye and ran on a 12% SDS gel (Section 2.5.1). Figure 3.19 shows the gel image of these eluted fractions.

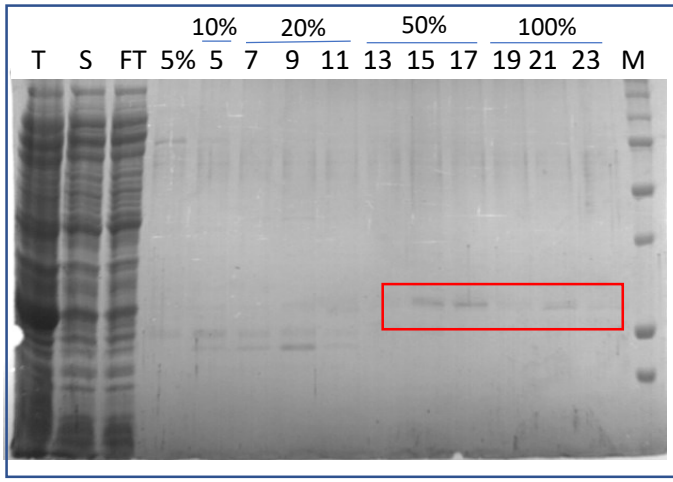


Figure 3.19 Expression of alternate fractions of eluent protein from Ni-NTA column.

The band for H6-SofG-ALFA was observed in the fractions 15, 17, 21, 23. Therefore, we selected 14, 15, 16, 17, 18, 21, 22 (35 ml) fractions for dialysis. Since we suspected loss of protein in flow-through, the flow was checked for the presence of protein with Bradford. There was no colour change which shows the absence of protein in flow through. Then, 35 ml of protein sample from the fractions where protein is supposed to be present was dialyzed and concentrated to 270 µl. The concentrated protein was then spun. The supernatant was collected and aliquoted 20 µl in each Eppendorf. The figure below the purity of the concentrated protein.

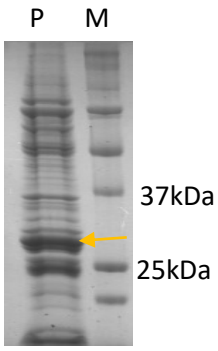


Figure 3.20 Expression of purified SofG-ALFA protein.

From the figure above, we observed the bands for other proteins along with SofG-ALFA. Hence the SofG-ALFA has not been successfully purified.

Chapter4. Discussion and Conclusions

For alleviating the preferred orientation problem of smaller proteins, we developed megabodies which can target the proteins by ALFA tag. For these SofG and megabody proteins need to interact successfully. The three megabodies with different linker lengths have been successfully cloned. We had chosen the smaller protein like SofG as it is not characterized yet. For the binding of megabody with SofG, SofG-ALFA has been successfully cloned.

If SofG-ALFA interacts with megabody constructs successfully, then its structure can be easily analysed by Cryo-EM as the megabody construct will give all views of the protein. For this, the megabody needs to be in polymerized state. Megabodies are successfully polymerizing in the presence of AMP-PNP with almost the same efficiency as that of ParM which is confirmed by pelleting assay. SofG-ALFA was also checked for polymerization. From the repeated pelleting assays, we confirmed that SofG-ALFA does not polymerize or pellet down from solution.

The interaction of megabody with SofG-ALFA was checked by pelleting assay. Megabody can be co-expressed with SofG-ALFA by co-transformation but here the expression of proteins was very low. The expression of constructs in co-expression was improved upon mixing the lysates. However, there were a lot of impurities observed in the pellet fraction which does not confirm their interaction. For this, the constructs need to be purified and mixed for interaction.

The purification of megabody Nb5 is done. The purification for other megabody constructs is remaining to be done. Purification of SofG-ALFA was not done successfully and hence needs to be taken up as a future objective.

Chapter 5. Future prospectives

1. Since megabody expression was getting less in co-expression by co-transformation, we are trying to find different ways to get good expression of both proteins on interaction which are as follows:
   a. The proteins can be purified individually and mixed in the column for interaction. Currently, our main aim is to successfully purify the megabodies and SofG-ALFA construct.
   b. By inserting both genes in the pETDuet vector by cloning and then co-express.
2. To compare the co-expression of purified SofG-ALFA with different purified megabody constructs. Megabody which is more soluble and shows better binding might be a suitable megabody for imaging the smaller proteins. The successful formation of SofG decorated on the helical scaffold can then be visualized by TEM or Cryo-EM in order to see whether it is showing different views of SofG protein.
3. The most soluble and stable megabody construct in co-expression can then be checked with different smaller proteins just to ensure this megabody can successfully target any protein without affecting its function. For this, the proteins need to be cloned with an ALFA tag. Then this megabody can be used for the purification of different proteins.
4. Different antibody-tag can be used in place of NbALFA for interaction.
   We can compare which system works well for this interaction.
5. The use of megabody construct as a purification strategy for nanobodies and smaller proteins is remaining and requires to be explored further.

Chapter 6. References

1. Cabral, A, Cabral, JE, and McNulty, R (2022). Cryo-EM for Small Molecules. Curr Protoc 2.

2. Chen, J, Noble, AJ, Kang, JY, and Darst, SA (2019). Eliminating effects of particle adsorption to the air/water interface in single-particle cryo-electron microscopy: Bacterial RNA polymerase and CHAPSO. J Struct Biol X 1.

3. Van Den Ent, F, and Löwe, J (2006). RF cloning: A restriction-free method for inserting target genes into plasmids. J Biochem Biophys Methods 67, 67–74.

4. Garner, EC, Campbell, CS, and Dyche Mullins, R Dynamic Instability in a DNA-Segregating Prokaryotic Actin Homolog.

5. Gayathri, P, Fujii, T, Møller-Jensen, J, Van Den Ent, F, Namba, K, and Löwe, J (2012). A bipolar spindle of antiparallel ParM filaments drives bacterial plasmid segregation. Science (1979) 338, 1334–1337.

6. Gayathri, P, Fujii, T, Namba, K, and Löwe, J (2013). Structure of the ParM filament at 8.5Å resolution. J Struct Biol 184, 33–42.

7. Götzke, H et al. (2019). The ALFA-tag is a highly versatile tool for nanobody-based bioscience applications. Nat Commun 10.

8. Kanade, M, Singh, NB, Lagad, S, Baranwal, J, and Gayathri, P (2021). Dual specificity of a prokaryotic GTPase-activating protein (GAP) to two small Ras-like GTPases in Myxococcus xanthus. FEBS Journal 288, 1565–1585.

9. Ke, N, and Berkmen, M (2014). Production of disulfide-bonded proteins in Escherichia coli. Curr Protoc Mol Biol 2014, 16.1B.21-16.1B.21.

10. Kühlbrandt, W (2014). The resolution revolution. Science (1979) 343, 1443–1444.

11. Li, B, Zhu, D, Shi, H, and Zhang, X (2021). Effect of charge on protein preferred orientation at the air–water interface in cryo-electron microscopy. J Struct Biol 213.

12. Liu, N, and Wang, HW (2022). Better Cryo-EM Specimen Preparation: How to Deal with the Air–Water Interface? J Mol Biol.

13. Lobstein, J, Emrich, CA, Jeans, C, Faulkner, M, Riggs, P, and Berkmen, M (2012). SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. Microb Cell Fact 11.

14. Murata, K, and Wolf, M (2018). Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. Biochim Biophys Acta Gen Subj 1862, 324–334.

15. Nakane, T et al. (2020). Single-particle cryo-EM at atomic resolution. Nature 587, 152–156.

16. Naydenova, K, and Russo, CJ (2017). Measuring the effects of particle orientation to improve the efficiency of electron cryomicroscopy. Nat Commun 8.

17. Noble, AJ, Wei, H, Dandey, VP, Zhang, Z, Tan, YZ, Potter, CS, and Carragher, B (2018). Reducing effects of particle adsorption to the air–water interface in cryo-EM. Nat Methods 15, 793–795.

18. Nwanochie, E, and Uversky, VN (2019). Structure determination by single-particle cryo-electron microscopy: Only the sky (and intrinsic disorder) is the limit. Int J Mol Sci 20.

19. Orlova, A, Garner, EC, Galkin, VE, Heuser, J, Mullins, RD, and Egelman, EH (2007). The structure of bacterial ParM filaments. Nat Struct Mol Biol 14, 921–926.

20. Popp, D, Narita, A, Oda, T, Fujisawa, T, Matsuo, H, Nitanai, Y, Iwasa, M, Maeda, K, Onishi, H, and Maéda, Y (2008). Molecular structure of the ParM polymer and the mechanism leading to its nucleotide-driven dynamic instability. EMBO Journal 27, 570–579.

21. Ren, G, Ke, N, and Berkmen, M (2016). Use of the SHuffle strains in production of proteins. Curr Protoc Protein Sci 2016, 5.26.1-5.26.21.

22. Uchański, T et al. (2021). Megabodies expand the nanobody toolkit for protein structure determination by single-particle cryo-EM. Nat Methods 18, 60–68.

23. Uchański, T, Pardon, E, and Steyaert, J (2020). Nanobodies to study protein conformational states. Curr Opin Struct Biol 60, 117–123.

24. Wu, X, and Rapoport, TA Cryo-EM structure determination of small proteins by nanobody-binding scaffolds (Legobodies).

25. Yao, Q, Weaver, SJ, Mock, JY, and Jensen, GJ (2019). Fusion of DARPin to Aldolase Enables Visualization of Small Protein by Cryo-EM. Structure 27, 1148-1155.e3.