Attempts Towards Site-directed Mutation of an Endonuclease Gene and Insertion of a Gene in a Bacteriophage Genome

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

Manas Mahaveer

Indian Institute of Science Education and Research Pune

Dr. Homi Bhabha Road,

Pashan, Pune 411008, INDIA.

Date: December, 2023

Under the guidance of

Supervisor: Prof Saikrishnan Kayarat,

IISER Pune

From May 2022 to Nov 2023

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE

Certificate

This is to certify that this dissertation entitled Attempts Towards Site-directed Mutation of an Endonuclease Gene and Insertion of a Gene in a Bacteriophage Genome towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Manas Mahaveer at Indian Institute of Science Education and Research under the supervision of Prof Saikrishnan Kayarat, Professor, Department of Biology, during the academic year 2022-23.

Prof Saikrishnan Kayarat

Committee:

Dr. Saikrishnan Kayarat

Department of Biology

IISER Pune

Dr. Ramanujam Srinivasan

School of Biological Sciences

NISER

This thesis is dedicated to everyone who faltered, stumbled and fell, but never stopped

Declaration

I hereby declare that the matter embodied in the report entitled Attempts Towards Site-directed Mutation of an Endonuclease Gene and Insertion of a Gene in a Bacteriophage Genome 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 Saikrishnan Kayarat and the same has not been submitted elsewhere for any other degree

Manas Mahaveer

20181145

5th Year BS-MS

IISER Pune

Prof Saikrishnan Kayarat

Professor

Dept. of Biology

IISER Pune

List of Contents

Acknowledgments

I have seen myself change, grow and turn into a different person over the course of this thesis. It was difficult and filled with challenges, but that is what made it worth it. At the end of this thesis year, I am left feeling happy and proud, not because I was perfect throughout, but because I found in myself the ability to bunker down and keep moving forward, even if it was at snail's pace at times. If I were to do it all over again, knowing everything I know now, I would definitely do it differently, but I don't want to exchange the experience I had for anything else.

First and foremost, I would like to thank Professor Saikrishnan Kayarat, my thesis supervisor for believing in me through the course of the entire project, and allowing me to explore my interests while also expanding my horizons. Under his guidance, I was able to experience what it feels like to be part of a research group, yet it never felt like I was thrown into the deep end without anyone looking after me. I learnt to be responsible and accountable for my own progress and experiments, from the conceptualization to the execution of it. I would also like to thank Dr. Srinvasan Ramanujan, my TAC member for his suggestions and insights into my project, which have helped me mould it into its present form.

In my lab, I would like to thank Abhinav, Nikki and Vinayak for being the best set of mentors and guides I could have ever asked for. They allowed me to be inquisitive, offer suggestions and always were ready to listen to my ideas and help refine them. I am grateful for the presence of Ashwin and Rushik as well, the closest thing I had to a bench partner, for keeping me enthused throughout, while also lending their brain and time to my project. Everyone at SK Lab were helpful, patient and inviting, never making me feel out of place and I was able to learn something from each and everyone, not only making me a better researcher, but a better person.

It would be impossible to look back at my time at IISER without mentioning my friends, who became my family away from home. I am grateful for 4G1C and Yuvraj for being a combination of friends and a guide through my time here, even if they did have to leave a year earlier. Milie, Harshit, Gaurav, Shikhara, Arjun, Adithyan, Ram and Abhishek were my adventuring party at IISER, and will always hold a special place in my heart for maintaining my sanity while simultaneously taking part in our insane plans. I am grateful for Mimamsa and the Quiz Club, two student clubs without which my time at IISER could never have been so fulfilling and fun.

I would like to express my gratitude for Dr. Gayathri Pananghat, who has been a guide both academically and otherwise since my first year at IISER Pune, looking out for me always. I thank my parents and grandparents for being a pillar of support and love, ensuring that my time at IISER was comfortable and enjoyable. Lastly, I thank KVPY and IISER-Pune's Biology Department for giving me the resources and opportunity to work on this project.

Contributions

Contributor name	Contributor role
Dr. Saikrishnan Kayarat	Conceptualization Ideas
Manas Mahaveer, Abhinav Masih, Vinayak Tumuluri	Methodology*
-	Software
-	Validation
Manas Mahaveer, Dr. Saikrishnan Kayarat	Formal analysis
Manas Mahaveer, Abhinav Masih, Vinayak Tumuluri	Investigation*
Dr. Saikrishnan Kayarat	Resources
Manas Mahaveer	Data Curation
Manas Mahaveer	Writing - original draft preparation
Manas Mahaveer, Dr. Saikrishnan Kayarat	Writing - review and editing
Manas Mahaveer, Dr. Saikrishnan Kayarat	Visualization
Manas Mahaveer, Dr. Saikrishnan Kayarat, Abhinav Masih (Gibson Assembly)	Supervision
Dr. Saikrishnan Kayarat	Project administration
Dr. Saikrishnan Kayarat	Funding acquisition

*Methodology and Investigation - Manas Mahaveer, Abhinav Masih, Vinayak Tumuluri (Gibson Assembly), Manas Mahaveer (Site-directed Mutagenesis)

Abstract

Anti-microbial resistance is a growing issue for the healthcare sector, with more and more pathogenic strains propping up every day. It isn't enough to just kill the cells, as the DNA released into the environment can still confer the resistance to neighbouring cells through Horizontal Gene Transfer (HGT), and hence there is the need to develop measures that can do a two-fold job: kill the resistant-strain cells and prevent further propagation. SauUSI is a Type-IV Restriction Endonuclease that has been shown to have a "shredding" activity in methylated dsDNA. It creates several cuts between recognition sites, essentially putting the DNA through a biochemical paper-shredder.

In my thesis, I aimed to engineer a lytic phage delivery system for SauUSI. The phage is responsible for cell lysis while the SauUSI is meant to shred all of the cellular DNA, thus making it meaningless in the context of HGT. Using Gibson Assembly as a synthesis method, and subsequent phage rebooting, I tried to create these phage-SauUSI constructs. The second half of the thesis focussed on identifying potential residues responsible for the specific recognition sequence of SauUSI, and generating mutations in the protein to allow us to alter it. This would allow for a widening in the range of DNA that the construct can shred and have potential therapeutic applications.

Introduction

1.1 Anti-Microbial Resistance

Anti-microbial resistance (AMR) is a major healthcare concern around the world, as pathogenic organisms are rapidly developing strains that are resistant to antibiotics. Existing antibiotics lose their effectiveness against these strains, caused by overuse and eventual abuse of antibiotics. This creates pressure on two major fronts: on the healthcare sector to use alternative, less efficient methods to combat diseases, and on the biomedical research sector to keep trying to synthesize and develop novel antibiotics to work on these new resistant-strains. Methicillin-resistant Staphylococcus aureus (MRSA) and Extended-Spectrum beta-lactamase (ESBL) are just an example of many strains that are cropping up in different parts of the world.

Researchers have tried to come up with novel techniques that can be used to counter such strains, and phage therapy is one such development. Bacteriophages, which are bacteria-specific viruses, have been used to target these pathogenic strains and kill such cells. Yosef et al. 2015 for example showed that temperate phages could be used as delivery agents for a CRISPR-Cas system that could revert the resistance of the strain into a state that can then be treated with traditional antibiotics. Lytic phages on the other hand have been used to target and lyse such resistant cells, bursting the cell open and expelling its contents in the surrounding milieu.

1.2 Horizontal Gene Transfer

With using lytic phages as a therapeutic, a new challenge comes forth, namely Horizontal Gene Transfer (HGT). It is the movement of genetic material between organisms by processes other than cell division or reproduction. HGT can occur through three modes: transformation, transduction or conjugation. On cell lysis, the cell contents are taken in by neighbouring, antibiotic-sensitive cells and they “pick up” resistance-conferring genes, thus bringing the system back to square one. In this case, HGT is actually beneficial for the bacteria, but that isn’t always the case. Viruses can

hijack cell machinery and feed off the host, which is obviously not a desirable outcome for the host. Hence, there exist defence mechanisms against such invasions. One such mechanism is the Restriction Modification System (RM-system), where Restriction Endonuclease (REase) identifies foreign DNA and cleaves it, while certain modifications are made to host DNA so that there is no damage done to the host genome.

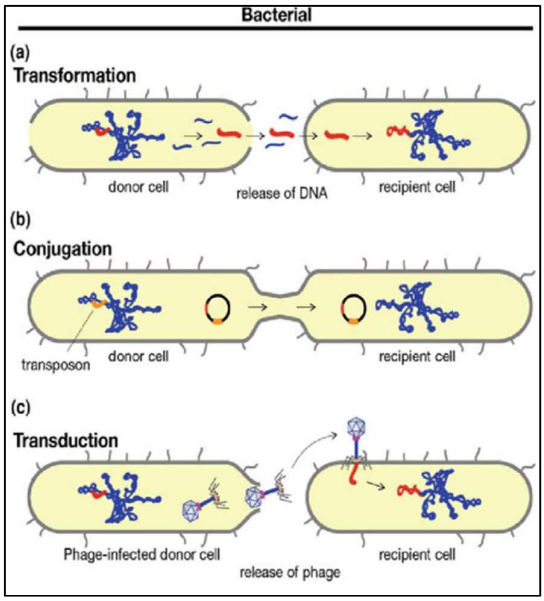


Fig 1.1: Modes of HGT

1.3 Restriction Endonucleases and SauUSI

The REases are categorized into 5 types, based on their cofactors and the target sequences' biochemical properties: Types I, II, III, IV and V. Of these, Type-IV REases target sequences with modifications, like methylation, hydroxymethylation or glucosyl-hydroxymethylation. SauUSI is one such Type-IV REase, and was first discovered in MRSA strains. It was shown to be a defence mechanism against HGT in *S. aureus*. Its cleavage pattern is interesting as in substrates where there exist multiple target sites, the enzyme occupies them all and translocates on ATP hydrolysis along the protein, leading to a "traffic jam" when multiple enzyme molecules converge at the same site. This is quickly followed by the nuclease activity of the enzyme, which makes a cut in the DNA. When this happens along the entire substrate, it is "shredded" into multiple small fragments.

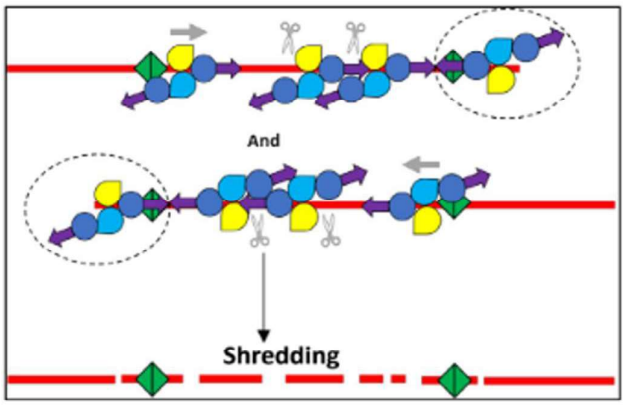


Fig 1.2: Shredding by SauUSI, where green rhombuses are target sites

SauUSI is a 953 amino acid-long protein made of three distinct domains:

1. Nuclease Domain belonging to the Phospholipase D family
2. ATPase Domain belonging to the Helicase Superfamily 2

3. Target Recognition Domain (TRD) that specifically recognizes the sequence 5′-S^{5m}CNGS-3′, where S is either Guanine or Cytosine, and ^{5m}C is 5-methylcytosine or 5-hydroxymethylcytosine.

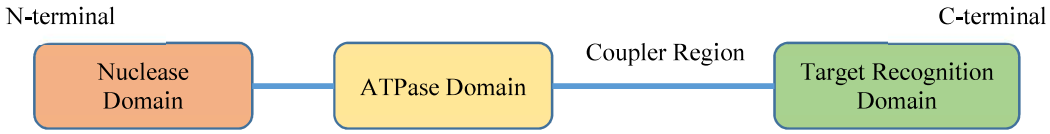


Fig 1.3: Domain Architecture of SauUSI

1.4 Phage Therapy

As a supplement to antibiotics, bacteriophages are used as therapeutics to treat infections where a portion of the pathogen population is antibiotic-resistant. Bacteriophages are suited as a delivery system because they do not target human cells, and can even be specific to only certain types of bacteria. The phages have a capsid, protecting the genome of the virus and a tail-like appendage that is used to attach and invade the host cell through its membrane.

This infection can then progress in one of 2 ways:

a. The viral DNA is incorporated into the host genome, which happens in the case of temperate phages or
b. The viral DNA takes control of the cell replication machinery to replicate its own genome, which is then expressed to form structural proteins that constitute the capsid. Then these viral DNA copies are packaged and released as virions in the surroundings through cell lysis, which happens in lytic phages.

Both types of phages can be engineered to counter AMR, with benefits and drawbacks to both.

1.5 Bacteriophage Selection: T7 Phage

Cell lysis is not a complete method of dealing with AMR, as the cell's resistance-conferring genes, usually stored as plasmids, are exposed to surrounding cells. These cells thus gain resistance and the cycle can perennially continue. Hence, a T7 bacteriophage which is lytic in nature was planned to be the delivery system for SauUSI, a DNA shredder. When combined in this way, it is proposed that the host DNA will first be shredded by the action of the REase, and then once the phage causes

lysis, only tiny fragments of the genome will exist in the milieu, which won't carry any meaningful genetic information.

T7 was selected for construct engineering for a number of reasons. It is a dsDNA phage with a 40kbp genome, which makes it compact to synthesize and manipulate *in vitro*. It codes for essential structural proteins but also has regions in its genome that are deemed non-essential across various studies. These regions can thus be replaced by an insert, while still keeping the phage viable in theory. This viability is however contingent on the size of the insert, even if only non-essential genes are replaced, with variations in the behaviour of the engineered phage as compared to wild type T7 increasing as the insert size increases. It is also compatible with T7 promoters and terminators used for SauUSI expression in plasmids.

1.6 Bacteriophage Engineering through Gibson Assembly

An entire phage genome is much bigger than a plasmid, and hence the approach used to edit the genome must also be efficient in the face of these differences. Traditionally, phages have been edited using the recombination machinery of bacterial or yeast hosts by homologous recombination. However, the phages obtained through such techniques are a mixture of wild-type and edited phages, which necessitates a further screening process.

This is circumvented by *in vitro* assembly, through a process called Gibson Assembly. It is a synthetic biology technique first described by Gibson et al. which allows stitching together a DNA fragment from multiple, overlapping smaller fragments. The reaction has 3 components: a 5' exonuclease, a DNA ligase and a DNA polymerase. The fragments are designed such that they all have around 30-40 bp overlaps on either side with each other. For the reaction to happen, equimolar amounts of DNA fragments are mixed with the components. First, the exonuclease eats at the fragments and creates overhangs on the 3' end, which overlap with other fragments in the mixture. Then, the fragments anneal through these overlapping strands, while the polymerase replaces any extra, missing bases as a result of the exonuclease activity. Finally, the

strand breaks are repaired by the ligase, and a longer, single fragment is formed from the multiple smaller ones.

The bacteriophage genome was similarly cut up into smaller fragments designed with overlaps, and an insert fragment was designed to fit in between two T7 fragments. These fragments were combined using Gibson Assembly and then electroporated into host cells to reboot and generate constructed phages.

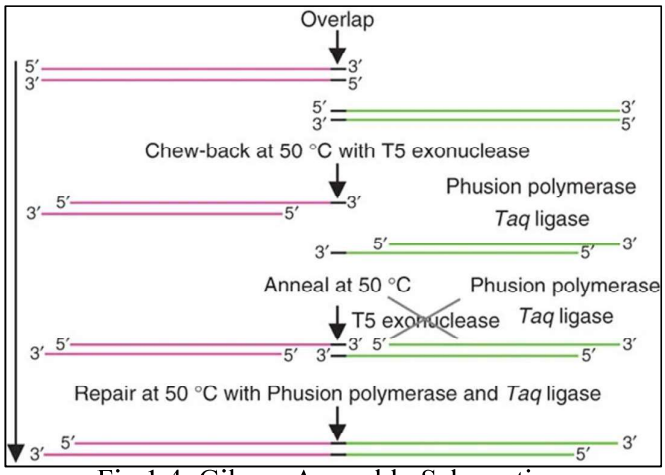


Fig 1.4: Gibson Assembly Schematic

1.7 SauUSI TR Domain

The Target Recognition Domain (TRD) of SauUSI is responsible for identifying the recognition sequence of the REase (5′-S^{5m}CNGS-3′). It is towards the C-terminal of the protein and is connected to the ATPase domain through a coupler region. The TRD is 148 residues long and is a part of the SRA-fold family.

The residue M829 is responsible for stabilizing the cavity formed by flipping of the methylated cytosine while the cytosine in question is flipped into a hydrophobic pocket, interacting with Y831, W868, D858 and I842. This interaction grants specificity to methylated-cytosine, as evidenced by a reduction in affinity to the substrate on replacing the ^{5m}C with cytosine.

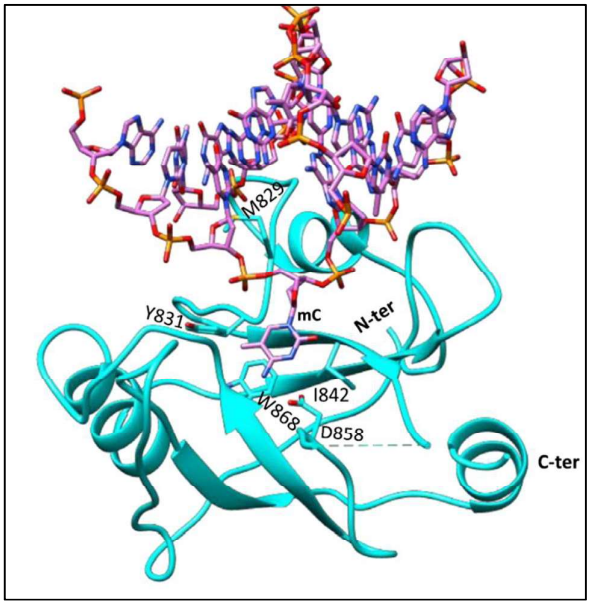


Fig 1.5: TRD of SauUSI (7CLG), with DNA from homolog SUVH6 (6A5N)

The SRA-fold is a highly conserved fold, with variations occurring mostly in the loop regions, which in turn change the sequence a particular TRD recognises. This is very clearly observed in the case of LpnPI and AspBHI, both having their respective TRD similar to that of SauUSI, with some notable

differences. To identify specific residues responsible for sequence recognition in SauUSI, its structure was compared to that of the above-mentioned proteins in Chimera and mutants were made using Restriction-Free (RF) Cloning.

1.8 RF Cloning

RF Cloning is a PCR-based method of inserting a DNA fragment into a circular plasmid completely independent of restriction sites. A linear DNA fragment containing the alteration to the plasmid is first generated through PCR, and this is used as a "megaprimer" for introducing this to the parent plasmid. The ends of the megaprimer are complementary to the plasmid sequence, which anneal and creates a nicked, bubble-like formation with the insert during a PCR reaction.

The template is grown in a host such that it has a different modification than the PCR product, for example, the template is grown in *dcm*$^+$ bacteria, and hence has cytosine methylation, while the PCR reaction mixture has unmethylated dCTP added to it. Then, we digest the entire PCR reaction contents with an enzyme that specifically cleaves cytosine-methylated DNA (such as DpnI). This ensures that the only plasmids that make it through undigested are the ones which contain the insert in them.

Fig 1.6: Schematic for RF Cloning

1.9 Scope of Thesis

SauUSI is an effective REase with well-documented activity, while T7 bacteriophage is also a widely studied system and has been observed to be an effective delivery and lytic system. Combining these two to make a system that can deliver SauUSI into a cell, shred the host DNA and then cause cell lysis and release more such virions into neighbouring cells is a potent attack on resistant pathogenic strains. This acts as both a measure against an AMR cell, and the subsequent HGT that can happen in traditional lytic phage therapy, making it a more complete and effective therapeutic.

The aim of this thesis is in two directions:

a. The first objective is to synthesize a T7 phage construct which has a SauUSI gene insertion. The fragments will be designed and then stitched together using Gibson Assembly, followed by subsequent rebooting. This construct will act as a total counter to AMR populations and prevent any passing on of resistance to antibiotic-sensitive populations. Such a construct can act as a model on which other such constructs can be based on, by changing either the host or the insert depending on the pathogen you want to target, and the most effective method of combating HGT in that system.
b. The second objective will be to modify the recognition sequence of SauUSI, and thus widen the field of hosts on which a construct can act. The insert can be treated as a cassette, and replacing wild-type SauUSI to a mutant in the construct will change the host it can act on. It will also supplement and further what we know about SauUSI's interaction with its substrate, and this can act as a starting platform to studying other proteins with an SRA fold.

Materials and Methods

1.1 Bacteriophage Genome Extraction

T7 bacteriophage genome was extracted from existing high-titre stocks of T7 phage using protocol described elsewhere. A 5ml primary culture of Medium 1 (0.5%peptone, 0.3% yeast extract, 1.5% Agar and 5mM $MgSO_4$) containing NEB® Turbo Competent *E. coli* cells (OD_{600}) was infected with 1ml of T7 bacteriophage stored in SM buffer and incubated at 37°C till lysis was observed. A secondary culture with 200ml of Medium 1 containing NEB® Turbo Competent *E. coli* cells (OD_{600}) was incubated at 37°C post infection with the lysate till lysis was observed again. 200µl chloroform was added to the secondary culture to lyse the entire population and ensure that all the bacteriophages were out of the cells. This was then spun at 4000g for 20 minutes at 25°C to separate the cell debris and the supernatant was passed through a 0.22µm filter. 216µl of L1 Buffer (20 mg/ml RNase A, 6mg/ml DNase I, 0.2 mg/ml BSA, 10mM EDTA, 100mM Tris-HCl, 3M NaCl, pH7.5) was added and then incubated for an hour at 37°C again. After this, we add 30ml of L2 Buffer (30% PEG6000, 3M NaCl) and kept overnight at 4°C. Post overnight incubation, the supernatant solution is then distributed into falcon tubes and centrifuged at 10,000g for 30 minutes at 4°C. The supernatant is discarded and the pellet is resuspended in 9ml L3 Buffer (100mM Tris-Cl, 100mM NaCl, 25mM EDTA, pH 7.5), and 9ml of 4% SDS is added immediately after. This solution is incubated at 70°C for 30 minutes, and then cooled on ice till all of the SDS precipitates. 9ml of L5 Buffer (2.55M CH_3COONa, pH 4.8) is added and then centrifuged for 30 minutes at 10,000g and 4°C again.

A Genomic Tip/100G flow column is taken and equilibrated with 4ml of QBT Buffer (750mM NaCl, 50mM MOPS, pH 7.0, 15% isopropanol, 0.15% Triton X-100). The supernatant is loaded onto it, and then 15ml of QC Buffer (1M NaCl, 50mM MOPS, pH 7.0, 15% isopropanol) is used to wash the column. The DNA in the column is then eluted by using 5ml of QF Buffer (1.25M NaCl, 50mM Tris-Cl, pH 8.5, 15% isopropanol). The DNA in this QF Buffer is then precipitated using 0.7 volumes of isopropanol and pelleted down by centrifuging at 10,000g at 4°C for 15 minutes. This pellet is washed with 2ml of chilled ethanol, and then centrifuged at the same conditions for another 15 minutes. The ethanol is allowed to air-dry, and then the pellet

is resuspended using 100µl of warmed nuclear-free water by shaking at 55°C, 300 rpm for 2 hours.

1.2 T7 DNA Confirmation: Double Layer Overlay Agar

Double Layer Overlay Agar (DLOA) is a technique that allows for host-phage interaction in a petri dish. The bottom layer of the petri dish is prepared with a LB broth with 1.5% agar content. This acts as the source of nutrition for the top layer, which consists of the same nutrient medium, but 0.7% agar i.e. roughly half, mixed in with a population of an overnight culture of DH5α *E. coli* cells. In the top layer, a lawn of the bacterium is formed and when the phage is plated on it, it allows for them to interact and infect the population. The plate, post spreading of phages, is incubated at 37°C for 12 hours, and the lawn clears up as the phage causes cell lysis. Based on the concentration of the phages, the clearance can appear either as plaques i.e. a patch of clearance surrounded by a lawn at low concentrations, while at higher concentration the entire dish becomes fully lysed.

1.3 Fragment Design for T7 Construct

The entire genome of T7 was divided into 8 fragments, with a ninth SauUSI insert fragment. For generating the T7 genome fragments, the template used was the extracted T7 wild-type genome, while for the SauUSI fragment, it used a pHis plasmid with SauUSI cloned into it. The primers were designed using ApE Plasmid Editor.

Primer Name	5'-→ 3' Sequence
Frag 1_Fwd	TCTCACAGTGTACGGACCTAAAGTTC
Frag 1_Rev	CTCTATAGTGAGTCGTATTGATTTGGCG
Frag 2_Fwd	GCGTTCGCGTAACGCCAAATC
Frag 2_Rev	GTTGACTTGAAGTTATGCATAACATTTATCC
Frag 3_Fwd	ACTAAGAGAGGACTTTAAGTATGCATAACTTCAAGTCAAC

Frag 3_Rev	GTTTCAGCTACGTTAGACATGGTGTGTCTCCTTTAG
Frag 4_Fwd	CTCACTAAAGGAGACACACCATGTCTAACGTAGCTGAAACTATCCG
Frag 4_Rev	TGACTGACTCCATACAGTTCTCCTAAGG
Frag 5_Fwd	TAAGCCACCGCTCCTTAGGAG
Frag 5A_Rev	CCCTATAGTGAGTCGTATTATTATTGCTCAGCGGTGGC
Frag 5B_Fwd	ACCGCTGAGCAATAACTAGCCTAGCATAACCCCTTG
Frag 5_Rev	AGGTTGCTCAAACACGGTTATCTTACC
Frag 6_Fwd	ATTGGAGCCTGATGGTAAGATAACC
Frag 6_Rev	CTGAAACTGGTTTAGACTTGTAGAGTG
Frag 7_Fwd	ATTCTGCGTGATACACTCTACAAGTC
Frag 7_Rev	AGGGACACAGAGAGACACTC
SauUSI Int10_Fwd	CTGCCACCGCTGAGCAATAATAATACGACTCACTATAGGGAGAC
SauUSI Int10_Rev	AAGGGGTTATGCTAGGCTAGTTATTGCTCAGCGGTG

Table 2.1: Primer Details for T7 Construct

The fragments were generated using PCR, with specific PCR protocols listed below in Table 2. They were generated using *pfu* polymerase, and purified using QIAGEN PCR Purification Kit.

Fragment Name	Fragment Length (bp)	Annealing Temp (°C) [T_A]	Extension Time (s) [t_E]
Fragment 1	5850	55	360

Fragment 2	2958	55	180
Fragment 3	4875	55	300
Fragment 4	6628	55	410
Fragment 5A	4011	60	240
Fragment SauInt10	3119	55	320
Fragment 5B	2632	60	155
Fragment 6	6628	55	410
Fragment 7	6610	55	410

Table 2.2: Fragment Details and PCR Protocols for T7 Construct

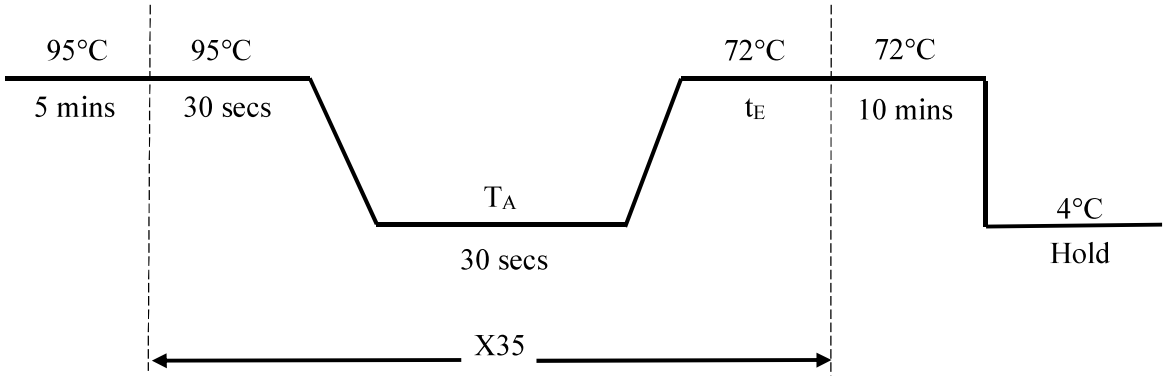


Fig 2.1: Schematic for PCR for Fragment Synthesis

1.4 Fragment Assembly using Gibson Assembly

DNA Fragments were taken in equimolar quantities at 0.4 pmol per fragment in a PCR tube, and mixed with 10µl of NEBuilder HiFi DNA Assembly Master Mix. It was then incubated at 50°C for 1 hour. This process was done in 2 steps, by first creating two larger sub-fragments of the construct, called Fragment A and Fragment B, which were then joined together again by performing the Gibson Assembly reaction again into the full construct genome.

1.5 Phage Rebooting

The reaction product of Gibson Assembly needs to be rebooted to observe whether the construction resulted in a viable phage. 3µl of the reaction product and 1µl of T7 gDNA (as rebooting control) were added to 50µl of NEB® Turbo electrocompetent cells in a 2mm BioRad electroporation cuvette. The cells were then pulsed with parameters of 10µF capacitance, 1.8kV voltage and 600Ω resistance. To the cells, 200µl of SOC Media (LB Media containing 20mM glucose and 10mM magnesium

sulphate) was added for cell recovery, and incubated for 2 hours at 37°C. The cells were then completely lysed using 60µl of chloroform, and centrifuged at 11,000g for 2 minutes to separate the cell debris as a pellet from the free phages in the supernatant. 200µl of this supernatant was then mixed with 500µl of overnight culture of *E.coli* BL21(DE3) cells and 5ml of 0.7% LB agar. This mixture was spread onto a plate with 1.5% LB agar (double layer overlay agar) and incubated at 37°C overnight. Any observed plaques are supposed to be picked using a pipette tip, and then stored in SM buffer at 4°C.

1.6 Structure Analysis of SauUSI TRD

The structure of SauUSI TRD was taken from the complete protein structure (PDB ID:7CLG), and Foldseek Server was used to find protein structure alignments for it with the same fold. From the results of this search, the structures of LpnPI (PDB ID: 4RZL), AspBHI (PDB ID: 4OC8) and UHRF1 (PDB ID: 3FDE) were shortlisted for further structural comparison and analysed using Chimera.

1.7 RF Cloning of SauUSI Mutants

For generating point mutations in the SauUSI TRD, RF Cloning was used. The megaprimers were first synthesized using primers which incorporated the mutation, and then the pHis backbone of the plasmid was generated using PCRs. This was done using *pfu* polymerase, with SauUSI-containing pHis plasmid as the template. The primers were designed using ApE Plasmid Editor.

Primer Name	5'-→ 3' Sequence
S826A_Fwd	TTAGTAAAATATTTAATTGGAATAAAAATGGTTCGGCTGTAATCATG
S826A_Rev	CATGATTACAGCCGAACCATTTTTATTCCAATTAAATATTTTACTAA
S826N_Fwd	TTAGTAAAATATTTAATTGGAATAAAAATGGTTCGAATGTAATCATG
S826N_Rev	CATGATTACATTCGAACCATTTTTATTCCAATTAAATATTTTACTAA

D904N_Fwd	TGTACAGAAAAAAGATGATGATGGTATATATTTTTATTATT TAGGAAC
D904N_Rev	GTTCCTAAATAATAAAAATATATACCATCATCATCTTTTTT CTGTACA
Sau_pHis_Fwd	CTAAATGATTTCAATCAATCTTTA
Sau_pHis_Rev	TAGATAACGATATATATCATCTCTTAC

Table 2.3: Primer Details for TRD Mutant RF Cloning

Product Name	Product Length (bp)	Annealing Temp (°C) [T_A]	Extension Time (s) [t_E]
Megaprimer S826A_Fwd	417	55	75
Megaprimer S826N_Fwd	417	52	75
Megaprimer D904N_Fwd	188	55	45
S826A_pHis	5467	52	480
S826N_pHis	5467	52	480

Table 2.4: Product Details and PCR Protocols for TRD Mutant RF Cloning

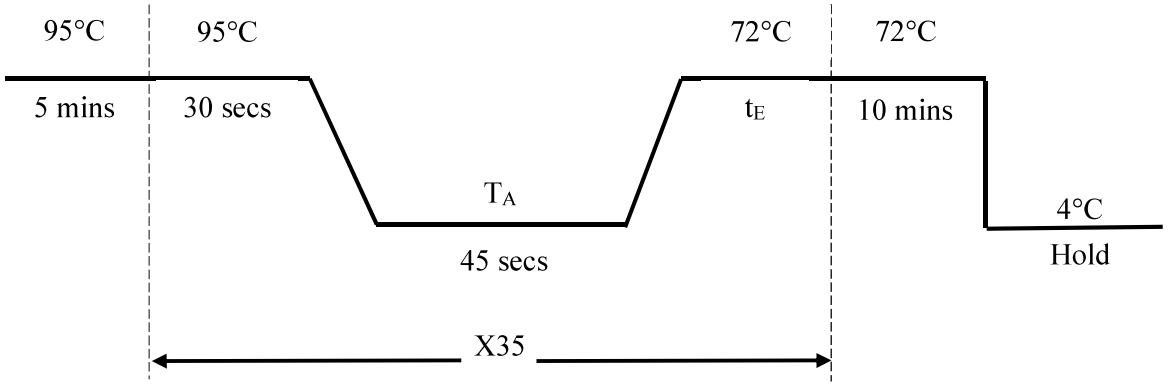


Fig 2.2: Schematic for PCR for Rf Cloning

Results and Discussion

1.1 SauInt10 Design and Synthesis

The SauUSI insert (SauInt10) is a 3,119 bp long sequence. It contains T7 promoter φ10, a Ribosome-binding site (RBS), the SauUSI gene and T7 terminator Tφ. This insert was synthesized into a pHis17 plasmid, which is the template used for the amplification. The PCRs were performed using *pfu* polymerase, and the product was cleaned up and concentrated into sufficient quantities before assembly.

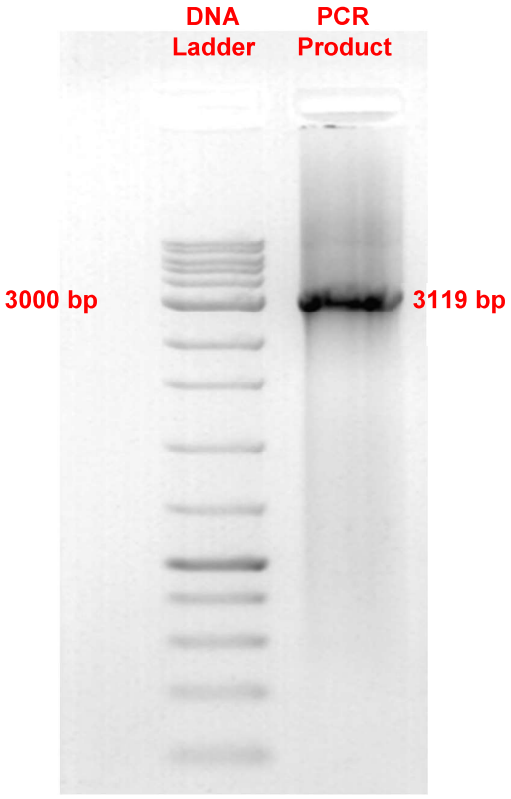


Fig 3.1: 1% Agarose Gel
showing SauInt10 Amplification

1.2 T7 Genome Fragment Synthesis

The T7 genome fragments were of differing lengths, and based on length and primer properties, the extension times and annealing temperatures differed. The PCRs were carried out using *pfu* polymerase, and the products were all cleaned up and concentrated to sufficient quantities before assembly.

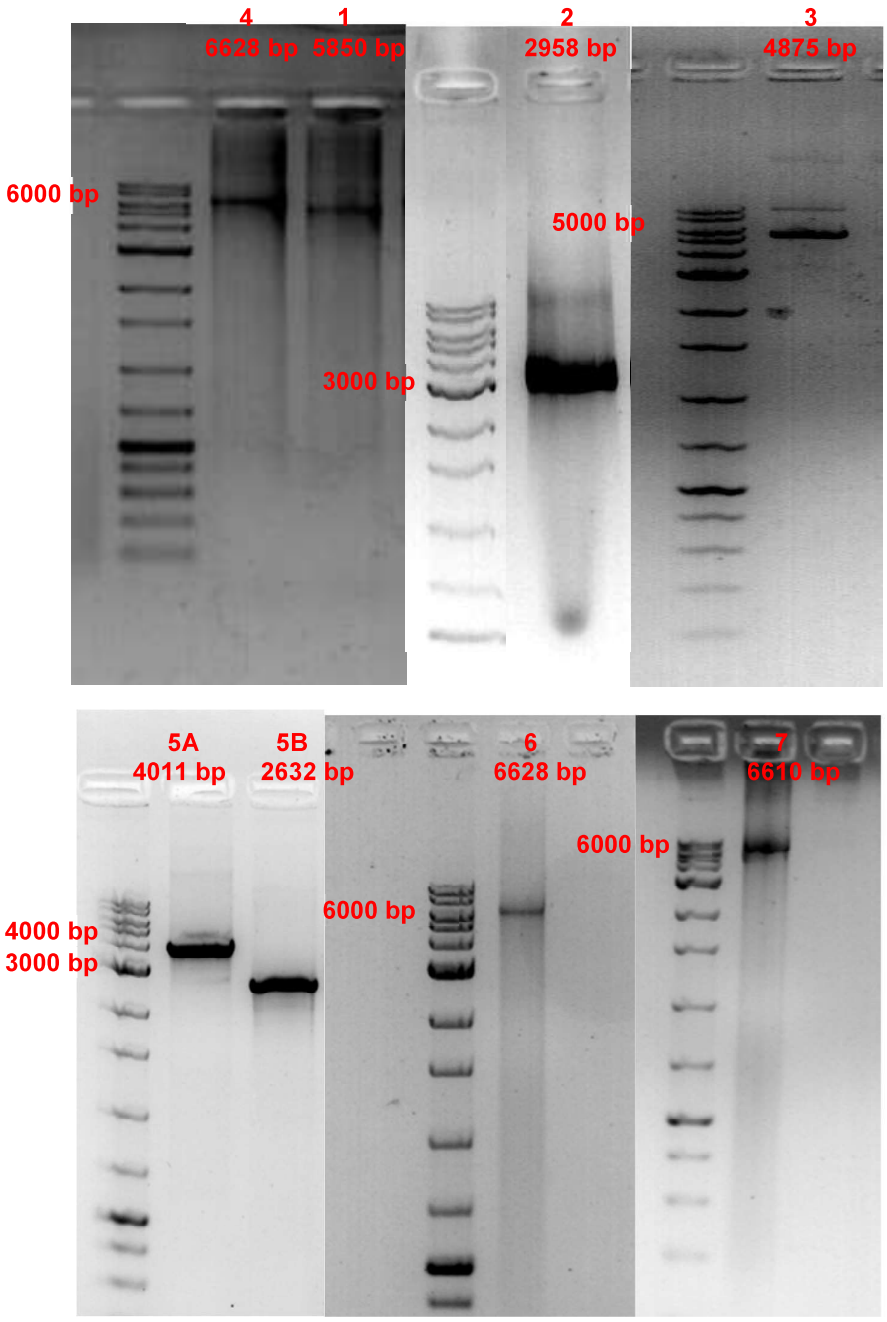


Fig 3.2: 1% Agarose Gels showing T7
Fragment Amplification

1.3 T7-SauInt10 Mutant Bacteriophage

It is necessary to determine a suitable location for insertion, to maximise the chances of designing a viable phage. The larger the size of the insert, the lengthier the packaging time of the virus, and that has an adverse effect on the efficiency of lysis. It has been shown that inserts of 2.5kb can be accommodated into the T7 genome without any major adverse effects on the phage.

There are multiple regions in the T7 genome that have been shown to be non-essential for the phage viability, and hence are suitable sites for replacement/insertions into the genome. The insertion location of choice for this mutant phage was between fragments 5A and 5B, with the exact location being chosen to ensure that no promoters, terminators or RBS for other native genes were affected, so as to prevent any unwanted change in the life cycle of the phage. The location is right after *gp10*, a non-essential region of the genome, and thus perfectly suited for insertion. The SauInt10 design was such that it does not depend on any promoters, RBS or terminators from the T7 genome, and thus can produce SauUSI directly under the influence of the phage's RNA Polymerase.

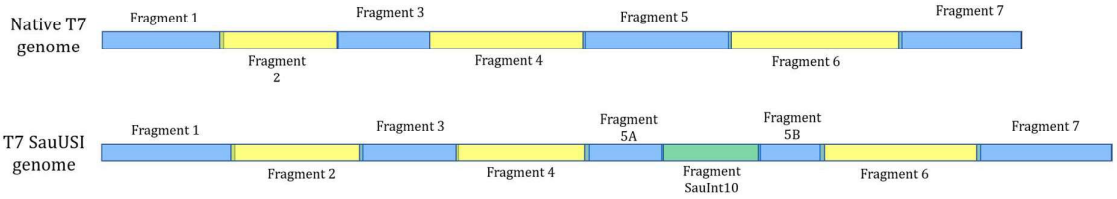


Fig 3.3: Schematic showing T7 native genome fragment design, and SauInt10 location

1.4 Gibson Assembly of T7 Mutant Phage

The fragments were stitched together *in vitro* using NEBuilder® HiFi DNA Assembly. The process has a limitation that it can only stitch together 4-6 fragments in one single reaction, and the entire mutant construct was composed of 9 fragments. To work around this, the entire Gibson Assembly was carried out as a two-step process.

For the first step, fragments 1 to 4 and fragments 5A to 7 were assembled into larger sub-fragments in separate reactions, with the products being labelled Fragment A and Fragment B respectively. Then, a second Gibson reaction was setup with Fragments A and B in the reaction mixture, which were then stitched together to give the entire construct genome. As a Gibson Assembly control, a Positive Control was used which was provided with the kit. The Gibson Assembly Positive Control (GAPC) was composed of 2 fragments of pUC18 plasmid, which contained an ampicillin-resistance gene, split into the 2 fragments. To test whether the phage assembly was successful, the phages needed to be rebooted and tested for their activity.

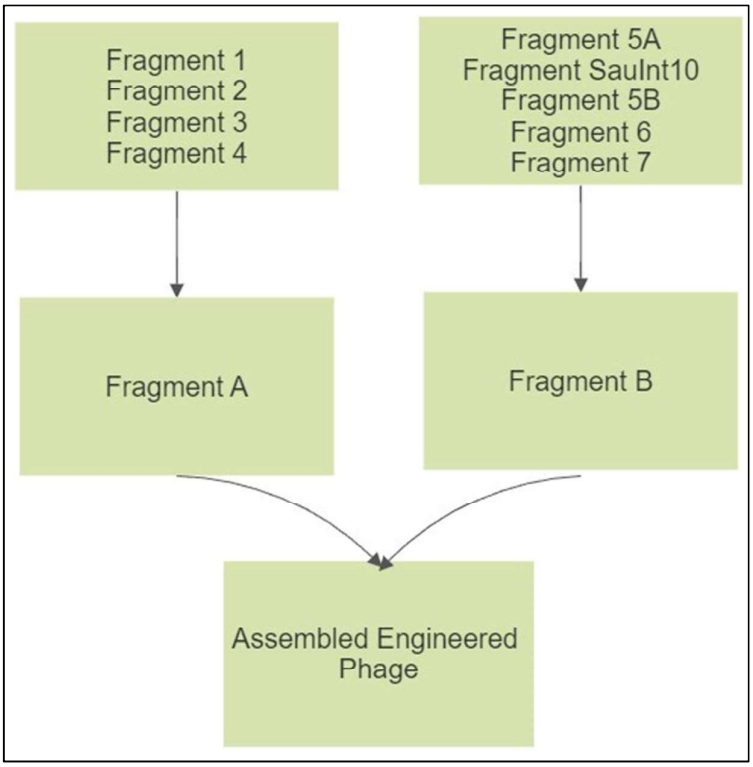


Fig 3.4: Schematic showing Gibson Assembly protocol for Construct Synthesis

1.5 Phage Rebooting and Plating

The Gibson Assembly product needed to be checked to confirm whether the reaction was successful, and if yes, whether the phage was viable or not.

NEB® Turbo Electrocompetent cells were electroporated with the construct assembly product, and as a positive control, the same was done with T7 genomic DNA (EPC). Post transformation and cell recovery, they were plated onto a Double Layer Overlay Agar setup, and compared to a lawn of BL21(DE3) cells that were allowed to grow in such a set-up without any phages introduced to it, which served as a negative control. No lysis was seen in the negative control, thus ruling out the presence of any phage contamination in the cells pre-plating. In the EPC plate, the T7 wild-type phages were rebooted successfully, as seen by a full clearance of the lawn and thus showing that

the rebooting process itself had worked. However, in the plate which contained the construct, a lawn was observed akin to the negative control.(See Fig 3.6)

This lack of plaques in the construct plate can be explained by two possible factors:

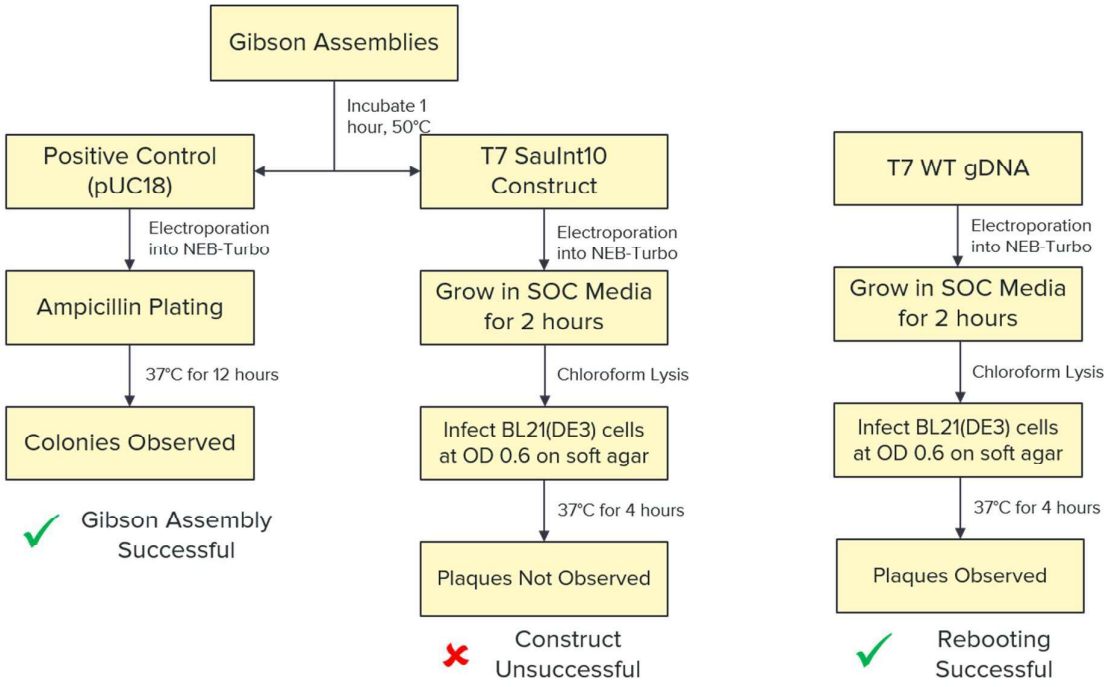


Fig 3.5: Workflow of Phage Rebooting

- **Non-Viable Phage**

The phage that had been formed post-assembly is not viable, either due to the size of the insert or the location of it. If this is the cause, then future experiments can be successful upon finding a more suitable insert location in the T7 genome, or by making suitable deletions in the genome simultaneous to the insertion, so as to preserve the overall size of the phage genome.

Alternatively, if no suitable site can be found in the entire length of the genome, then a phage with a larger genome, such as T4 or T5 can be used, since a 3kb insertion in their genome would be a smaller percentage change in overall genome size.

- **Failed Gibson Assembly**

The two-step Gibson Assembly did not work due to having insufficient concentrations of either of the two sub-fragments after the first Assembly. Since the resultant Gibson Assembly product at the end of the incubation period contains a mixture of full-length product and all possible sub-fragments which are formed after stitching together of

only some of the total fragments in the mixture, the concentration of Fragments A and B cannot be reliably quantified using a Nanodrop machine. Visualization of the product on an agarose gel is also not possible due to low concentrations of the full-length product and presence of intermediates.

To circumvent this, the sub-fragments could be re-amplified in a PCR reaction by using the appropriate primers of only the endpoints of the sub-fragments. The reaction can also be carried out by dividing the genome into 3 sub-fragments instead of 2, to increase the individual yields of full-length products after the first round of assembly.

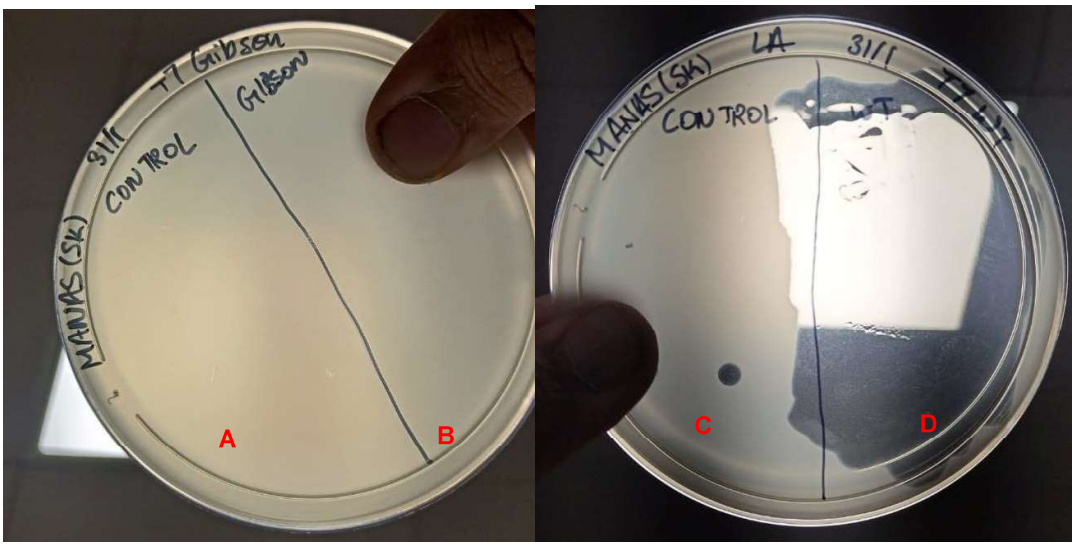


Fig 3.6: (A): Negative Control for Construct Plate, (B): Construct Assembly Product plated, (C): Negative Control for EPC Plate and (D): T7 DNA plated

1.6 Identification of Candidate Residues for TRD Mutations

The structure of SauUSI TRD is very similar to the N-terminal domains of LpnPI and AspBHI, which are both also modification-dependent DNA REases. They are all members of the SRA-fold family of proteins, and this allowed us to superimpose the three on top of each other, and use structural studies done on the two models to infer important residues in SauUSI which play a role in the identification of its target sequence.

Chimera was used to superimpose the three structures, and to generate a sequence alignment based on it. In studies done before, 3 major loop regions were identified in both AspBHI and LpnPI which were responsible for recognising their respective target sequences. Out of these, SauUSI showed a lack of Loop 2B, which has been theorized to aid in recognizing the nucleotide in the +2 position with respect to the ^{m}C.

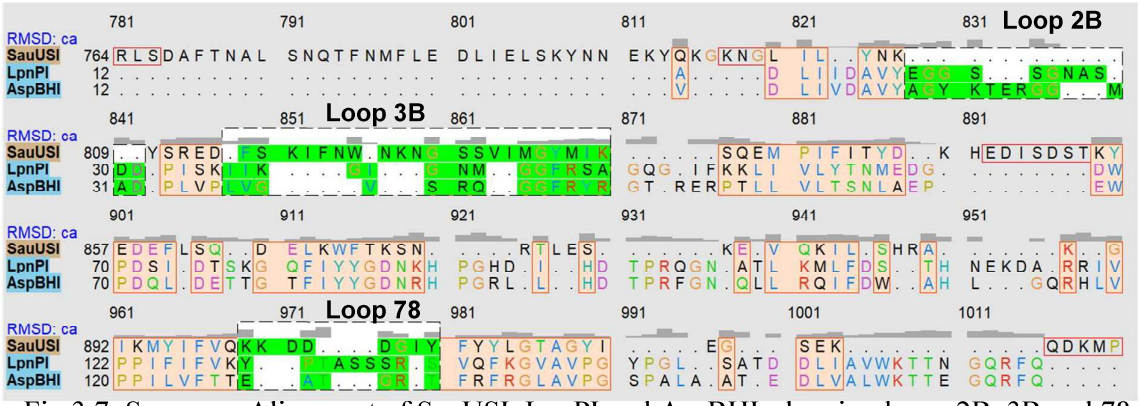


Fig 3.7: Sequence Alignment of SauUSI, LpnPI and AspBHI, showing loops 2B, 3B and 78

However, when it comes to Loop 3B, its length in SauUSI is 21 residues long, as compared to 8 and 9 for LpnPI and AspBHI respectively. This is a difference of roughly 2.5 times, thus we proposed that SauUSI's 3B has a much larger role in sequence recognition than in the other two enzymes. For this, a model of SauUSI was superimposed with a crystal structure of SUVH6, an *Arabidopsis* methyltransferase which had the protein bound to a DNA template (PDB ID: 6A5N). This allowed us to simulate a DNA-bound structure of SauUSI, which itself is not in a DNA-bound state.

On doing so, Ser826 in Loop 3B was spatially close to the -1 nucleotide, and also is biochemically capable of forming hydrogen bonds specifically with guanine or cytosine. Based on these observations, it was identified as a potential candidate for mutation to check whether it contributes to -1 nucleotide recognition in SauUSI. A S826N mutation was also hypothesized for modification of the recognition sequence. Mutating the serine to an asparagine would change the recognition from a G/C to an A, based on literature.

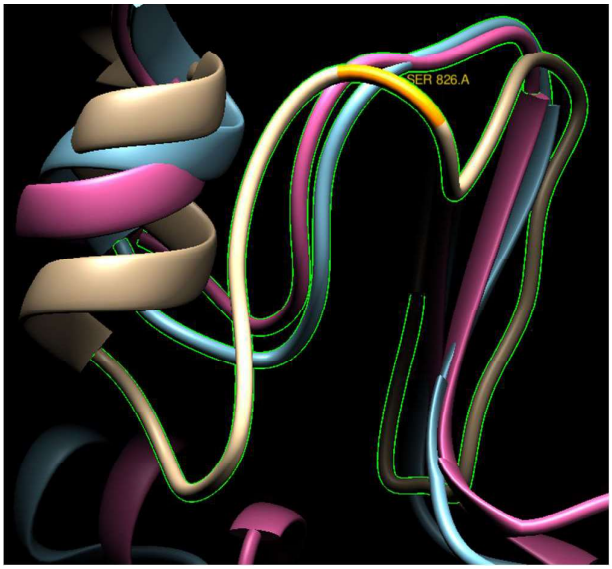


Fig 3.8: Loop 3B of SauUSI (brown), LpnPI (blue) and AspBHI (pink). S826 in orange.

Similarly, when trying to tackle +2 specificity in SauUSI, it was more of a challenge as the DNA template being used as a replacement for SauUSI does not conform to the protein in the same way a native DNA strand would. Moreover, any conformational changes in the loop regions of SauUSI in case of DNA binding would not be present in the structure used.

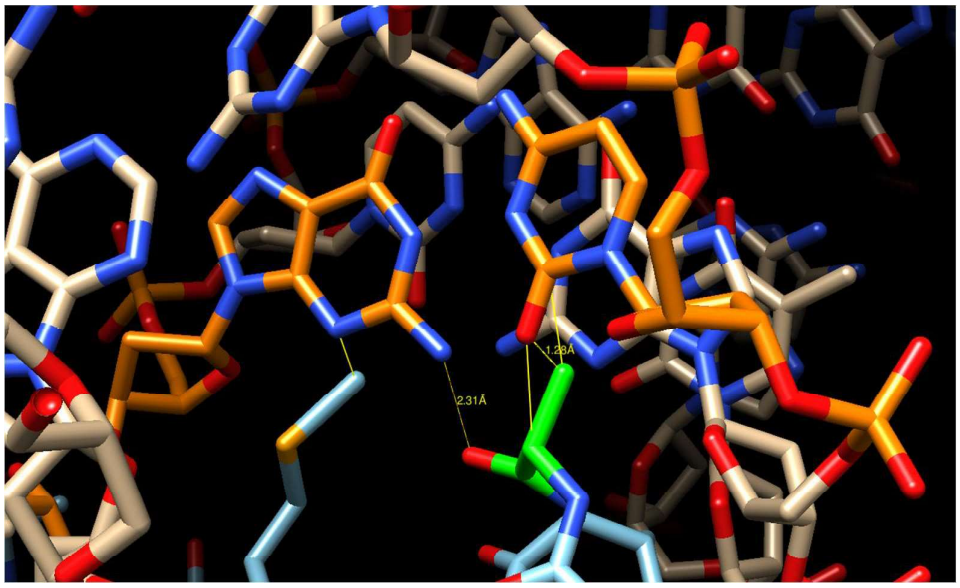


Fig 3.9: Guanine and Cytosine i.e. S at position -1 (orange), interacting with Ser826 (Green). Serine side-chain hydroxyl group was not present in the structure.

To work around these constraints, more emphasis was put on possible biochemical interactions between guanine and residues present in loop 78, which was spatially the closest loop to the DNA template. Based on this, Asp904 was shortlisted as the second candidate for mutation. A D904N mutation was planned to allow for a recognition shift from G to A by SauUSI.

1.7 RF Cloning of SauUSI Mutants

RF Cloning was done for three mutants: S826A, S826N and D904N. The cloning was done with appropriate primers, and using SauUSI-containing pHis plasmid as the template. Post cloning, the mutants S826A and S826N showed positive amplification when visualized in a 1% agarose gel, at 5,467 bp each. The products were transformed into DH5α ultracompetent cells, and plated on 1.5% agar plates containing LB media and ampicillin. Only mutant S826N showed colony growth, which were picked up and a plasmid extraction was done post making cultures from these colonies. The plasmids were sent for sequencing of the mutation, and a clone was confirmed to be positive for the S826N mutation. Since the sequencing results were done following a reverse read,

the underlined codon in *Fig 3.11,* ATT is the Asp826 AAT. Full gene sequencing has not been done.

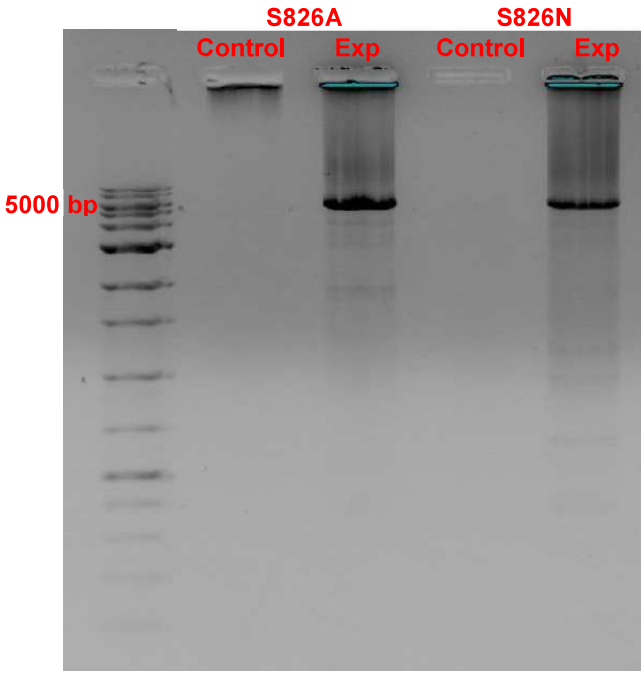


Fig 3.10: 1% Agarose Gels showing
S826A and S826N RF PCRs

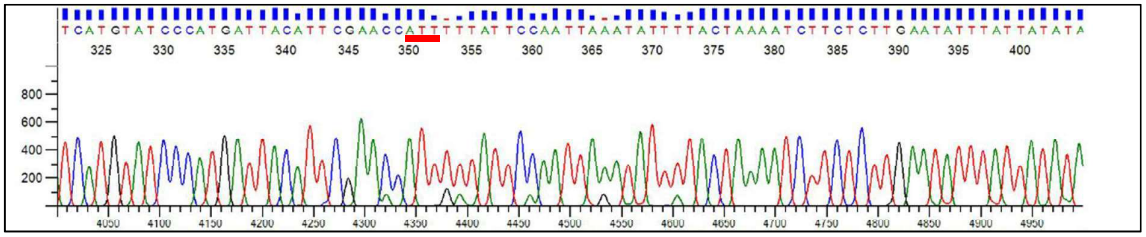
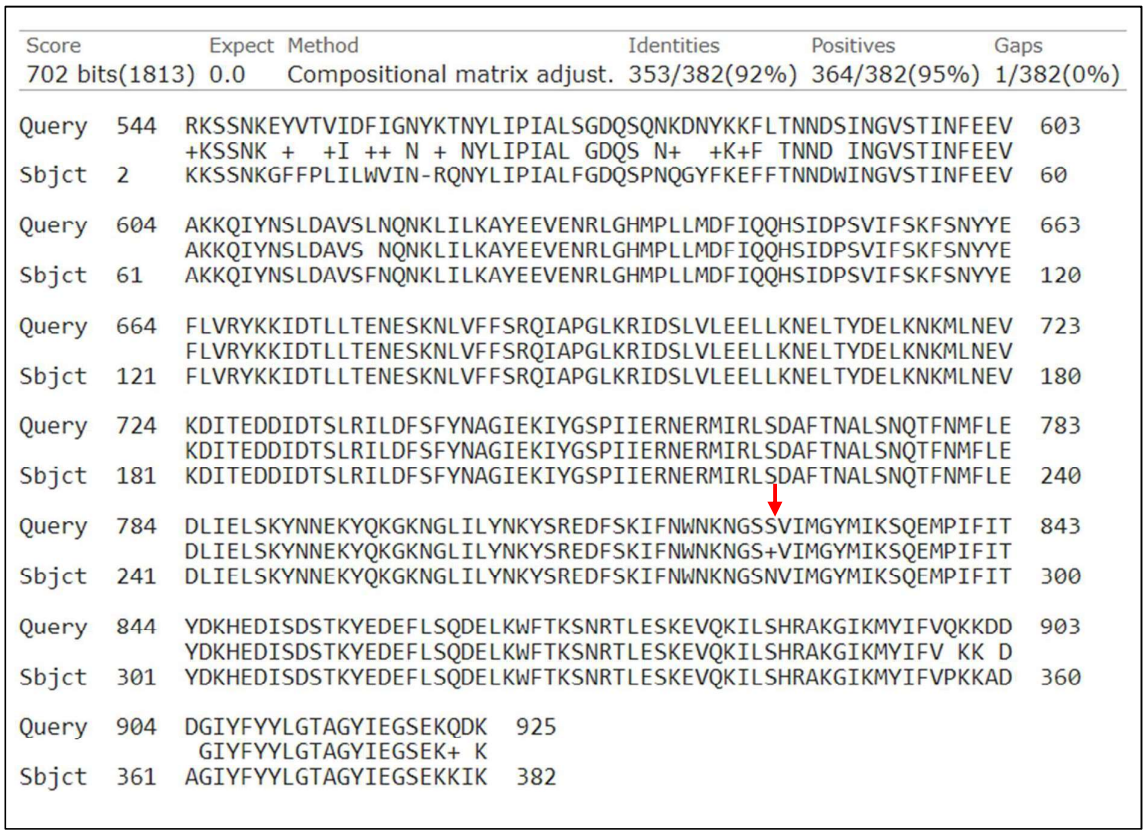


Fig 3.11: Sequencing Result of positive clone (S826N mutation underlined in red).



Score	Expect	Method	Identities	Positives	Gaps
702 bits(1813)	0.0	Compositional matrix adjust.	353/382(92%)	364/382(95%)	1/382(0%)

```
Query  544  RKSSNKEYVTVIDFIGNYKTNYLIPIALSGDQSQNKDNYKKFLTNNDSINGVSTINFEEV  603
            +KSSNK +  +I ++ N + NYLIPIAL GDQS N+  +K+F TNND INGVSTINFEEV
Sbjct  2    KKSSNKGFFPLILWVIN-RQNYLIPIALFGDQSPNQGYFKEFFTNNDWINGVSTINFEEV  60

Query  604  AKKQIYNSLDAVSLNQNKLILKAYEEVENRLGHMPLLMDFIQQHSIDPSVIFSKFSNYYE  663
            AKKQIYNSLDAVS NQNKLILKAYEEVENRLGHMPLLMDFIQQHSIDPSVIFSKFSNYYE
Sbjct  61   AKKQIYNSLDAVSFNQNKLILKAYEEVENRLGHMPLLMDFIQQHSIDPSVIFSKFSNYYE  120

Query  664  FLVRYKKIDTLLTENESKNLVFFSRQIAPGLKRIDSLVLEELLKNELTYDELKNKMLNEV  723
            FLVRYKKIDTLLTENESKNLVFFSRQIAPGLKRIDSLVLEELLKNELTYDELKNKMLNEV
Sbjct  121  FLVRYKKIDTLLTENESKNLVFFSRQIAPGLKRIDSLVLEELLKNELTYDELKNKMLNEV  180

Query  724  KDITEDDIDTSLRILDFSFYNAGIEKIYGSPIIERNERMIRLSDAFTNALSNQTFNMFLE  783
            KDITEDDIDTSLRILDFSFYNAGIEKIYGSPIIERNERMIRLSDAFTNALSNQTFNMFLE
Sbjct  181  KDITEDDIDTSLRILDFSFYNAGIEKIYGSPIIERNERMIRLSDAFTNALSNQTFNMFLE  240

Query  784  DLIELSKYNNEKYQKGKNGLILYNKYSREDFSKIFNWNKNGSSVIMGYMIKSQEMPIFIT  843
            DLIELSKYNNEKYQKGKNGLILYNKYSREDFSKIFNWNKNGS+VIMGYMIKSQEMPIFIT
Sbjct  241  DLIELSKYNNEKYQKGKNGLILYNKYSREDFSKIFNWNKNGSNVIMGYMIKSQEMPIFIT  300

Query  844  YDKHEDISDSTKYEDEFLSQDELKWFTKSNRTLESKEVQKILSHRAKGIKMYIFVQKKDD  903
            YDKHEDISDSTKYEDEFLSQDELKWFTKSNRTLESKEVQKILSHRAKGIKMYIFV KK D
Sbjct  301  YDKHEDISDSTKYEDEFLSQDELKWFTKSNRTLESKEVQKILSHRAKGIKMYIFVPKKAD  360

Query  904  DGIYFYYLGTAGYIEGSEKQDK  925
             GIYFYYLGTAGYIEGSEK+ K
Sbjct  361  AGIYFYYLGTAGYIEGSEKKIK  382
```

Fig 3.12: Alignment of SauUSI Sequence (Query) and Sequencing Result (Sbjct); Arrow pointing at incorporation of S826N mutation

Conclusion

AMR is a growing concern for healthcare industry around the globe, and HGT is a barrier when it comes to treating AMR through phage therapy. To act as a therapeutic, restriction endonucleases can be combined with the phage delivery system to provide a more complete response to antibiotic resistant pathogens.

The thesis has 2 halves, with the first half trying to engineer a modified, mutant T7 bacteriophage with SauUSI gene present in it. This would act as an example of a phage delivery system for a REase into a bacterial system, and cause cell lysis while simultaneously acting as a barrier for HGT to confer any resistance from the host cell to any neighbouring populations. The entire genome was split into fragments, and to incorporate an insert into it, the technique of *in vitro* DNA assembly i.e. Gibson Assembly was used. No phage plaques were formed on agar plates post assembly, which can be attributed to either a lack of viability of such a T7 mutant, or a limitation of the protocol used for synthesizing the viral genome from the smaller fragments.

For carrying on this section of the project, the phage genome needs to be redesigned with suitable deletions and identifying the optimal insertion site for SauUSI. This phage once constructed successfully, then needs to be compared to the wild-type T7 phage on parameters such as lysis time and burst size to illustrate and quantify the effects of the insert on the phage activity. Similarly, the DNA shredding efficiency of the phage will need to be compared to SauUSI introduced directly into a cell population.

The second half of the thesis focussed more on increasing the range of DNA sequences that could be targeted by SauUSI, and shed more light on the mechanism and structural features of the protein that grant it target specificity. The structure of SauUSI was compared to other proteins with similar folds, but on which more extensive studies had already been done. This resulted in showing an important structural difference between SauUSI TRD and two more prominent members of this fold family, LpnPI and AspBHI. We were able to show through structural superimposing and sequence alignment a difference in motifs present in the different enzymes, and how these differences could lead to differences in the manner in which these proteins interact with their DNA substrate.

To further progress on this front, the mutants already synthesized will need to be expressed, and cleavage assays will need to be done with suitable substrates to show a change in the recognition sequence of the mutant, when compared to the wild-type protein. Based on the results of this, more residues can be identified in the protein structure as well as potential candidates for mutational studies. Structures in a DNA-bound form will also help a lot in trying to determine important residues in the DNA-protein interaction of SauUSI.

Exploration in these directions have major applications for global healthcare in the coming decades, with more and more pathogens evolving strains that seek to escape the reach of modern antibiotics. Creating such phage delivery systems, and having a library of insert “cassettes” can lead to a more modular form of therapeutics, in which for a particular pathogen, a phage and cassette combination can be chosen which will have very specific and very efficient response to it.

References

Ando, H, Lemire, S, Pires, DP, and Lu, TK Engineering Modular Viral Scaffolds for Targeted Bacterial Population Editing - PMC. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4785837/. Accessed October 28, 2023.

Davis, MW, and Jorgensen, EM ApE, A Plasmid Editor: A Freely Available DNA Manipulation and Visualization Program - PMC. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9580900/. Accessed October 28, 2023.

Gibson, DG, Young, L, Chuang, R-Y, Venter, JC, Hutchison, CA, and Smith, HO (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6, 343–345.

Horton, JR et al. (2014). Structure and mutagenesis of the DNA modification-dependent restriction endonuclease AspBHI. Sci Rep 4, 4246.

Jensen, JD, Parks, AR, Adhya, S, Rattray, AJ, and Court, DL (2020). λ Recombineering Used to Engineer the Genome of Phage T7. Antibiotics 9, 805.

Kulkarni, M, Nirwan, N, van Aelst, K, Szczelkun, MD, and Saikrishnan, K Structural insights into DNA sequence recognition by Type ISP restriction-modification enzymes - PMC. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4872093/. Accessed October 28, 2023.

Luscombe, NM, Laskowski, RA, and Thornton, JM (2001). Amino acid–base interactions: a three-dimensional analysis of protein–DNA interactions at an atomic level. Nucleic Acids Res 29, 2860–2874.

Mathieu, J, Alvarez, E, and Alvarez, P (2014). Recombination-assisted megaprimer (RAM) cloning. MethodsX 1.

Navarro, F, and Muniesa, M (2017). Phages in the Human Body. Front Microbiol 8, 566.

Pulkkinen, EM, Hinkley, TC, and Nugen, SR (2019). Utilizing in vitro DNA assembly to engineer a synthetic T7 Nanoluc reporter phage for Escherichia coli detection. Integr Biol 11, 63–68.

Sasnauskas, G, Zagorskaitė, E, Kauneckaitė, K, Tamulaitiene, G, and Siksnys, V (2015). Structure-guided sequence specificity engineering of the modification-dependent restriction endonuclease LpnPI. Nucleic Acids Res 43, 6144–6155.

Sukackaite, R, Grazulis, S, Tamulaitis, G, and Siksnys, V (2012). The recognition domain of the methyl-specific endonuclease McrBC flips out 5-methylcytosine. Nucleic Acids Res 40, 7552–7562.

Tumuluri, VS, Rajgor, V, Xu, S-Y, Chouhan, OP, and Saikrishnan, K (2021). Mechanism of DNA cleavage by the endonuclease SauUSI: a major barrier to

horizontal gene transfer and antibiotic resistance in *Staphylococcus aureus*. Nucleic Acids Res 49, 2161–2178.

Vernikos, G, and Medini, D (2014). Horizontal Gene Transfer and the Role of Restriction-Modification Systems in Bacterial Population Dynamics. 169–190.

Xu, S, Corvaglia, AR, Chan, S-H, Zheng, Y, and Linder, P (2011). A type IV modification-dependent restriction enzyme SauUSI from Staphylococcus aureus subsp. aureus USA300. Nucleic Acids Res 39, 5597–5610.

Yosef, I, Manor, M, Kiro, R, and Qimron, U (2015). Temperate and lytic bacteriophages programmed to sensitize and kill antibiotic-resistant bacteria. Proc Natl Acad Sci 112, 7267–7272.