In-Silico ANALYSIS FOR HOST RANGE IDENTIFICATION FOR BROAD SPECTRUM BACTERIOPHAGE (Phi 6) TO CONTROL Pseudomonas syringae AND ITS PATHOVARS

M.Sc. Thesis

2024

Submitted to

CENTRAL DEPARTMENT OF BIOTECHNOLOGY

Tribhuvan University

Kirtipur, Kathmandu, Nepal

Sujata Pokhrel

In-Silico ANALYSIS FOR HOST RANGE IDENTIFICATION FOR BROAD SPECTRUM BACTERIOPHAGE (Phi 6) TO CONTROL *Pseudomonas syringae* AND ITS PATHOVARS

M.Sc. Thesis

2024

Submitted to

CENTRAL DEPARTMENT OF BIOTECHNOLOGY

Tribhuvan University

Kirtipur, Kathmandu, Nepal

By

Sujata Pokhrel

Supervisors

Prof. Dr. Rajani Malla

Senior Scientist Dr. Pramod Aryal

Asst. Prof. Pragati Pradhan

Registration No: 5-2-33-325-2013

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my supervisor **Prof**. **Dr. Rajani Malla** for her support and guidance for the accomplishment of my thesis throughout the final semester of the graduate degree.

I am indebted towards my mentor and supervisor **Dr. Pramod Aryal** for his guidance through this entire process of thesis work since the very beginning. He has been for supportive to me from the acceptance of my thesis proposal to the final submission of the work. His consistent motivation and encouragement motivated me whenever I was challenged during my thesis period.

Similarly, I would like to express my gratitude towards my co-supervisor **Asst. Prof. Pragati Pradhan** who has been supporting me through the wet laboratory works and guiding me through.

First and foremost, I would like to express my sincere gratitude to **Alpha Agro Private Limited**, Ex-students of **Budhanilkanta School** (400A Batch) and **NMB Bank Limited** for providing the computer and Software. And my sincere thanks to our seniors of 9th batch, Central Department of Biotechnology and the department itself for helping us in establishing the computer lab where I could work on my thesis.

I would also like to extend my gratitude to "**KCRE Excellent Student Thesis grant 070122022"** for granting me the thesis award for completion of my thesis work.

I am also thankful to **Prof Dr. Krishna Das Manandhar**, HOD of Central Department of Biotechnology, Tribhuvan University for allowing me to complete my thesis.

A huge thanks to my senior **Mr. Samiran Subedi** and my classmate **Mr. Aashish Pokharel** who guided me with the problems in operating the software and technology used in this research work.

I would also like acknowledge my friends **Ms. Smita Shrestha, Mr. Binod Khadka, Ms. Pooja Shrestha** and my other classmates for assisting me during the time of need.

LIST OF ABBREVIATIONS

AA	Amino acids
AIRs	Ambiguous Restraints
BE	Binding Energy
BSA	Buried Surface Area
BLAST	Basic Local Alignment Search Tool
FASTA	Fast Adaptive Shrinkage Threshold Algorithm
FCC	Fraction of Common Contacts
HADDOCK	High Ambiguity Driven protein-protein DOCKing
PAE	Predicted Aligned Error
PDB	Protein Data Bank
pLDDT	predicted local distance difference test
PRODIGY	PROtein binDIng energy
Psa	*Pseudomonas syringae* pv. Actinidiae
Pph	*Pseudomonas syringae* pv. phaseolicola
Psav	*Pseudomonas syringae* pv. savastanoi
Pss	*Pseudomonas syringae* pv. Syringae
pTM	predicted template modelling
ip TM	interface predicted template modelling
pv.	Pathovar
RBP	Receptor Binding Protein

LIST OF TABLES

LIST OF FIGURES

Table of Contents

ABSTRACT

Phage spike protein P3 of Phi6 is involved in the attachment of the phage with its host *Pseudomonas syringae* via the membrane protein Type IV pilin. This study aims to carry out binding affinity prediction and its analysis, and molecular docking to identify Phi 6 as broad spectrum bacteriophage to control infection of plants by *Pseudomonas syringae* pathovars. Binding affinity prediction of P3 and Pilin complexes using the HADDOCK prodigy software inferred the efficacy of Phi6 against *Pseudomonas syringae* pathovars like *Pseudomonas syringae* pv. phaseolicola, *Pseudomonas syringae* pv. syringae, *Pseudomonas syringae* pv. actinidiae and savastanoi suggesting Phi6 as a broad spectrum bacteriophage as potential candidate to control *Pseudomonas syringae* and related pathogens. This inference was further validated by docking individual proteins using HADDOCK 2.4 and its results were in coherence to the finding that Phi6 is indeed a broad spectrum phage. The study provides an insight into the interaction dynamics between bacteriophage attachment proteins and bacterial pilins, underscoring the potential of exploiting these interactions to develop novel biocontrol measures. By detailing the intermolecular forces and binding affinities that govern these critical interactions, our study provides a foundation for the rational design of therapeutics aimed at mitigating the impact of pathogenic Pseudomonas on agriculture. This research represents a significant step towards the innovative treatment of plant diseases and the protection of global crop yields.

Keywords: Binding affinity, Broad spectrum, HADDOCK 2.4, P3, Phi 6, *Pseudomonas syringae*, Type IV pilin.

1. INTRODUCTION

1.1. Background

Pseudomonas syringae is a very common and widespread pathogen that infects woody plants and herbs causing a huge loss of agricultural crops. This broad ranged plant pathogen has more than sixty pathovars causing all kinds of diseases in citrus fruits and most of the crop species that are economically important (Xin et al., 2018). The impact of *Pseudomonas syringae* on agriculture is substantial due to its ability to cause diseases such as bacterial speck, blight, and canker across various crops, including tomatoes, kiwifruits, and beans. It is capable of surviving in different environments including water supplies, which makes control measures more difficult and encourages the spread of the infection. It infects plants by entering through wounds or stomata, or other natural openings, and causes symptoms including leaf spots, blights, and cankers. The bacterium is skilled at creating a protein known as ice nucleation active (INA) protein, which increases the harm that frost does to plants and intensifies its effects, particularly in temperate climates.

Since different strains of *Pseudomonas syringae* are widely spread in agricultural crops, we need a cheap and sustainable solution to prevent as well as treat pathogenic infections caused by this bacteria. The prevalent method for eradicating the effects of pathogenic strains of *Pseudomonas syringae* and related pathogen involves copper rich compounds. However, the capacity and efficacy of these compounds seems to be diminishing. Furthermore, these chemical compounds are found to leave trace amounts of toxin harmful to the consumer. This certainly calls for a new approach for treating these plant pathogens. In current times, phage therapy has emerged as one such potential method for treating such pathogenic plant species in phage therapy since phages are specific to a particular bacterium, have the capacity to replicate themselves and are non-toxic in nature because of lower dispersal of detrimental compounds in their habitat (Brüssow, 2012a). Besides, a single phage can infect multiple hosts despite of higher variability of their surface receptors which is definitely advantageous over chemical treatments such as use of antibiotics in plants or animals.

A therapeutic approach using bacteriophage has the potential to manage diseases, however. one needs a large library of pertinent phages along with detail knowledge of

interaction between phages and their hosts, i.e., different strains of *Pseudomonas syringae* (Svircev et al., 2018). Traditional culturing and performing wet lab assays are preferred and a reliable method for determining the host specificity and range, however culturing bacteriophage in laboratory environment can be challenging and result in ambiguities. Similarly, the genomes and proteomes of bacteria are shaped by the coevolution of bacteria and their phages. This implies that these data from *In Silico* analysis can be useful substitutes for researching and figuring out phage-host specificity. (Edwards et al., 2015).

Strong antibacterial activity can be found in proteins produced by bacteriophages. Therefore, by identifying and describing novel therapeutic targets, the exploration of phage-host interactions might guide the search for small molecules. Phage proteins have naturally developed strategies to disrupt microorganisms. Moreover, tiny groups of amino acids create hotspots in multiprotein complexes involved in functions like cell regulation, where they provide most of the free energy generated during interactions (Bogan & Thorn, 1998). By examining how proteins of phages interfere with the protein–protein interfaces, we can recognize potent spots and target them while developing new drugs. Since these medications may be modified to target widely conserved bacterial systems, they may be able to circumvent the need to diagnose a particular harmful bacterium, doing away with the original phage's restriction on specificity.

1.2. Current Studies

Control methods for *Pseudomonas syringae* include cultural practices, chemical treatments, and biological control strategies. However, these methods have their limitations, such as the development of resistance to chemical pesticides and the environmental impact of their use. Using hostile bacteria is one of the more promising biological control techniques. Certain strains of Pseudomonas possess the capacity to inhibit soil borne diseases and stimulate plant growth, suggesting their potential as biocontrol agents (Mavrodi et al., 2012). In addition, the U.S. Environmental Protection Agency has registered integrated biological control approaches that use Pseudomonas strains for disease suppression, indicating the usefulness of these tactics in controlling *Pseudomonas syringae*-caused plant diseases (Stockwell & Stack, 2007).

The use of bacteriophages is a very promising method of treating bacterial plant diseases (Jones et al., 2007). Bacteriophages, sometimes known as phages, constitute viruses that attack bacteria, the most prevalent organisms in nature. The International Committee on Taxonomy of Viruses (ICTV) reports that there are 165 phage species that infect different species of Pseudomonas (Suttle, 2005). Research on bacteriophages targeting *Pseudomonas aeruginosa*, a relative of *Pseudomonas syringae*, demonstrates the ability of bacteriophages to control bacterial infections effectively. Studying the molecular mechanisms underlying these interactions between Pseudomonas species and their phages can help develop phage-based treatments targeted at *Pseudomonas syringae* (De Smet et al., 2017b). The Pseudomonas phage Pf-10 is a part of the biopesticide "Multiphage," which is used to stop bacterial infections in agricultural crops brought on by *Pseudomonas syringae* (Kazantseva et al., 2021). Pf-10's ability to lyse Pseudomonas bacteria and its stability over a broad pH and temperature range make it a good choice for biocontrol applications.

Similarly, identifying bacteriophages with appropriate specificity from environmental isolates is a labor- and money-intensive procedure that needs to be done (Nilsson, 2014; Schmidt, 2019). These isolates are frequently then mixed to create "phage cocktails" that have the intended host range (Altamirano & Barr, 2019). However, challenges such as phage resistance and environmental stability need to be addressed to maximize the efficacy of phage therapy in agricultural settings (Chegini et al., 2020).

In addition to eliminating plant diseases, phage application in agriculture also fosters the growth and health of plants. The interaction between biocontrol agents and the plant environment, including the microbiome and physical conditions, plays a significant role in the success of bacteriophage therapy (Holtappels et al., 2021).

1.3. Hypothesis

1.3.1. Null Hypothesis

H_0: There is no host range specificity for broad spectrum bacteriophage and Pseudomonas syringae pathovars.

1.3.2. Alternate Hypothesis

H_A: There is host range specificity for broad spectrum bacteriophage and Pseudomonas syringae pathovars.

1.4. Objective

1.4.1. General Objective

- To identify broad spectrum bacteriophage against *Pseudomonas syringae* and its pathovars.

1.4.2. Specific Objective

- To identify phage surface proteins and their potential receptors in *Pseudomonas syringae*.
- To retrieve amino-acid sequences of phage surface proteins and prepare it.
- To build the 3D structures of the proteins using the sequences.
- To perform protein-protein docking and simulations.
- To identify the host range for suggested phage to imply as therapy against wide range of hosts relating to *Pseudomonas syringae*.

1.5. Rationale

Plant pathogens like *Pseudomonas syringae* and its pathovars have affected a wide range of agricultural crops causing significant economic losses worldwide. Conventional methods to control this pathogen, including chemical pesticides and crop rotation, have limitations due to environmental concerns, pesticide resistance, and the pathogen's diverse host range. This research focusses on identifying the host range for potent bacteriophages to control different phyto-pathogens of the Pseudomonas family through the study of the binding affinity of the pathovars with their phages.

1.6. Scope of the Study

The research study focuses on developing bacteriophage therapy against the plant pathogen *Pseudomonas syringae* and its related pathogen by utilizing the in-silico techniques for host detection. The outcomes may pave the way for using the phages against *Pseudomonas syringae* and its pathovars as a biocontrol method of managing diseases in agriculture, which would lessen the need for chemical interventions and promote ecosystem health.

2. LITERATURE REVIEW

2.1. Review of literature related to *Pseudomonas syringae*

2.1.1. Plant Pathogenic Pseudomonas species

The Pseudomonas genus is one of the bacterial genera with the greatest ecological significance that consists of plant commensals (such as *Pseudomonas stutzeri* and *Pseudomonas fluorescens*), pathogenic species like *Pseudomonas syringae*, can also be used for bioremediation (*Pseudomonas putida*) as well as pathogens of insects (*Pseudomonas entomophila*), humans as well as animals (*Pseudomonas aeruginosa*) (Silby et al., 2011). Due to the enormous metabolic variety of this genus and the exceptional ability of different strains to respond to stressors in the environment, it is widely distributed ecologically, even in freshwater and soil. Despite being recognized as a pathogenic species, many isolates of this species are non-pathogenic and commensal in nature.

2.1.2. Pathovars of *Pseudomonas syringae*

The pathogenic strain of Pseudomonas, also known as *Pseudomonas syringae* van Hall 1902, was initially isolated from a diseased plant *Syringa vulgaris* (lilac). Collectively known as *Pseudomonas syringae* complex, this plant pathogen is further classified into nine genomospecies based on DNA–DNA hybridization, over 60 pathovars based on pathogenic features, and 13 phylogenetic groups (phylogroups) based on multilocus sequence analysis (Berge et al., 2014). In the Pseudomonas genus, the *Pseudomonas fluorescens*-like major branch contains a monophyletic group made up of the Pseudomonas syringae species complex (Garrido-Sanz et al., 2016). These phylogroups further divide into 2 groups of the seven canonical lineages with late branches and the six non-canonical lineages that branch out early (Adamczyk et al., 2012). The canonical lineages represent the strains with phenotypic characters associated to the *Pseudomonas syringae*. The conserved effector locus (CEL) and exchangeable effector locus (EEL) 21, which encode a type III secretion system (T3SS), flanking these strains' typical tripartite pathogenicity islands. Whereas the CEL encodes the three highly conserved syntenic effector genes hopAA11, hopM1, and avrE, the effectors produced by the EEL vary between pathovars and strains. The T3SS is essential for pathogenesis in a variety of

interactions between animals and plants because it translocate b effector proteins of bacteria inside the cells of host (Clarke et al., 2010).

The ability to induce immune-associated programmed host cell death, or the hypersensitive response, in resistant plants, ice nucleation activity, and the iaaL gene, which is involved in the inactivation of the plant hormone auxin, are among the characteristics shared by the *Pseudomonas syringae* lineages. The plant-specific canonical PGs contain the iaaL gene (Glickmann et al., 1998). *Pseudomonas viridiflava* and *Pseudomonas cichorii*, two plant diseases with a wide host range and resemblance to *Pseudomonas syringae*, are among the six early branching lineages.

Many agricultural and non-agricultural *Pseudomonas syringae* isolates have been collected and sequenced, revolutionizing our knowledge of *Pseudomonas syringae* diversity and evolution. Despite the species' original classification as a pathogen, it has now been discovered that a large number of strains that phylogenetically belong to the species are not pathogenic and cohabit with plants as commensals.

2.1.3. Growth and Infection Progression in *Pseudomonas syringae*

There are two interconnected phases of growth in *Pseudomonas syringae* species, one is the epiphytic phase, and the other is the endophytic phase. During this phase, the pathogenic bacteria lives on plant surfaces and superficial tissues, specifically on the parts of plants above the ground. But the bacteria actually invade and enters the plant tissue in the endophytic phase. In this phase, the bacteria colonize the intercellular apoplast space and multiply to initiate infection (Xin et al., 2018b). In fact, for some strains of *Pseudomonas syringae,* the epiphytic populations on the surface of the plant can serve as excellent indicators of future endophytic populations and disease outbreaks whenever conditions become favorable. (Rouse, 1985).

2.1.4. Determinants of Pathogenicity of *Pseudomonas syringae*

The pathovars of *Pseudomonas syringae* employ a variety of virulence factors, such as the Type III secretion system, effector proteins, and various phytotoxins, to induce disease in a specific host species. Using these systems, the pathogen enters plant cells and translocate factors that promote virulence, influencing transcription. It also has evolved

virulence factors like, the phytotoxins coronatine and T3Es that impair stomatal defense of its host plant and hence establish infection.

2.1.4.1. Type III Secretion System (T3SS)

Pseudomonas syringae pathogen uses the type III secretion system (T3SS) to influence transcription and transfer virulence components into plant cells (Block & Alfano, 2011). The T3SS family of genes are found in the hypersensitive response and pathogenicity (Hrp) island area of the bacterial chromosome, which is a complicated network of effector proteins, chaperone, and secretion machinery (Collmer et al., 2000). The T3SS effector proteins (TTEs), encoded by the avirulence (avr) genes, are secreted through the needlelike export system into the host cell cytosol and therein elicit the plant hypersensitivity response (HR) (Morel & Dangl, 1997). The host's attempt to contain and isolate the invasive infection is seen in this reaction. Once impacted, the cells undergo apoptosis, which causes soft rot and browning of the plant; these necrotic lesions hinder the pathogen's ability to acquire resources and prevent it from spreading to adjacent plant tissues.

2.1.4.2. Type IV Pili

The Type IV Pili (T4P) are flexible, proteinaceous filaments that can reach lengths of several micrometers and have a diameter of 5–8 nm. Typically found at either or both poles of a cell, they facilitate twitching motion, a very effective and adaptable type of flagellar-independent microbial membrane mobility (Jarrell & McBride, 2008). Numerous other bacterial functions and traits, such as surface adherence and colonization, biofilm formation, genetic material uptake, and pathogenicity, have also been linked to T4P (Nudleman & Kaiser, 2004). These structures also assist the bacteria to withstand climatic shocks like rain splatters, air flow etc. that could disrupt bacterial invasions of plant tissues.

The adherence ability of the Type IV pili is thought to be a conditional virulence factor that is directly related to the colonization of epiphytes by *Pseudomonas syringae* pathovars. However, these structures aren’t only useful in adherence of the pathogen to its plant host, but they also act as initial site of cellular attachment for bacteriophages like phi6 and other members of ds RNA virus family. This attachment is facilitated by the Type

IV pilin protein, which is a chromosomally encoded bacterial protein (Burdman et al., 2011).

The pili is composed of the Type IV Pilin, also called the pilin protein which is a chromosomally encoded protein of *Pseudomonas syringae* groups that allows cells of bacteria to adhere with external parts of the plant like leaf surfaces. These structural proteins are tiny (7 to 20 kDa) and have an alpha-helical, hydrophobic, conserved N terminus with a transmembrane (TM) domain and an interaction between protein and protein motif (Melville & Craig, 2013).

2.1.4.3. Phytotoxins

The pathogenic *Pseudomonas syringae* species are also capable of producing plant toxins like phaseolotoxin, syringopeptin, tabtoxin, syringolin, syringomycin, and coronatine etc., that play equal role in its virulence strategy (Bender et al., 1999).

One such phytotoxins, coronatine mimics the plant hormone jasmonate is active in the form of jasmonoyl isoleucine (JA-Ile) and directly binds with its receptor to activate it. This phytotoxin mediated signaling cause stomatal opening and hence pathogenicity. Coronatine takes advantage of endogenous antagonistic association between salicylic acid and jasmonate signaling that occurs following the PAMP activation (Zheng et al., 2012).

2.2. Review of literature related to Bacteriophage

2.2.1. Bacteriophage and its History

Bacteriophages, also known as phages, are viruses that kill Bacteria and archaea. Frederick Twort, a British scientist, and Felix D'Herelle, a French-Canadian microbiologist, independently discovered these in 1915 and 1917, respectively. "Bacteriophage," which means "bacteria eater," was first used by D'Herelle. Bacteriophages are obligatory parasites of bacteria, and they need a host bacterium to replicate itself. This ubiquitous organism plays an important part in preserving the planet's microbiological balance. There are an estimated 1032 bacterial phages on the earth. Bacteriophages, often known as phages, are found wherever their bacterial host is present. (Wittebole et al., 2013). Structurally, bacteriophage is made of an outer protein coat, Nucleic acid (might be DNA

or RNA, but never both), and tail with structures like collar, sheath, baseplate, tail fiber and pin.

Nucleic acid, also known as phage genome, is present in capsid coated by coat protein. Based on the type of bacteriophage, double- or single-stranded DNA or RNA can make up the phage genome, and genome can be circular or linear. In most phages, the capsid is joined to a tail structure composed of proteins encoded by phages. The size of length of the tail varies according to their family and some classes of phage may not have tail. The core and sheath are attached to a hexagonal base plate and one or more tail fiber are attached to the tail end. Due to specificity nature to the host bacteriophage are named after host. The phages that infect Pseudomonas species are called Pseudomonas phages and these belong to the Cystoviridae family (Naureen et al., 2020).

2.2.2. Life Cycle of Bacteriophage

One can better understand the host range of bacteriophages in relation to the life cycle of the pathogen. Like any other viruses, phages also remain as an inert particle with their genetic material enclosed safely inside the protective protein covering. However, the virus can direct biological processes to generate additional viral particles inside its host cell. Generally, phages multiply in the cytoplasm but in the case of eukaryotic cells, most DNA viruses replicate inside the nucleus. The lifecycle of a bacteriophage begins with the initial adsorption of the phage on to the host which later progresses into both temperate and lytic life cycles. After infection, the lytic phages proliferate quickly and create progeny then disintegrate its host whereas, before becoming lytic, the temperate phages first integrate their genome into the host genome, form a lysogen, and remain in that state until stimuli cause them to change. Bacteriophages life cycles also involve persistent infections as well as pseudolysogeny which are comparatively less mentioned, and their widespread presence is also unknown (Siringan et al., 2014).

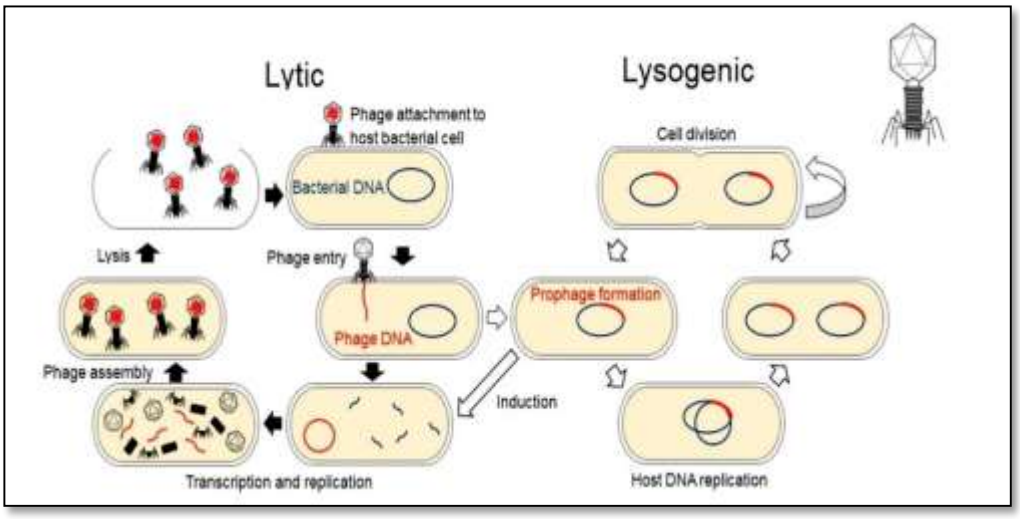


Figure 1: Life Cycle of Bacteriophage (Batinovic et al., 2019)

During the lytic phase of a virulent phage, there are four sequential stages of infection that finally lead to the lysis of the host. In the first step, the phage encounters a suitable host and initiate attachment by interacting with a suitable receptor. During this attachment, the phage enters the host cell and inserts its genetic material. The second step is surpassing the defense mechanism inside the cell to infect it sustainably. The process takes place when the tail sheath contracts, injecting viral DNA into the cell wall and membrane like a hypodermic needle. Then, by the third step, the bacteriophage phage hijacks the host resources to produce its progeny. It biosynthesizes all necessary viral enzymes, base plates, tail fibers, capsomeres, and sheaths needed to assemble new viruses. Finally, the viral progeny matures out then escapes from the host cell into the environment and restarts the life cycle. During this release, lysin and other phage proteins break down the cell wall; this process is referred to as lysis. The studies dedicated to exploring molecular mechanisms involved in broad spectrum bacteriophages focus on the initial attachment stage of this life cycle.

Contrastingly, in lysogenic life cycle, after the injection of phage genome in the bacterial host, rather than destroying the host, the genome of a phage merges with the chromosomes of bacteria to become a part of it. Prophages are phage genomes that have been incorporated. A lysogen is a prophage-carrying bacterial host. Usually, this lysogenic conversion occurs when the host encounter stress. The phage genome multiplies along with the host genome as the bacteria duplicates the phage's DNA and transfers it to new daughter cells along with its chromosome during reproduction. The presence of phage

genome in bacterial cells may change the phenotype of bacterium and this extra gene may lead the bacterial cell to more virulence (Ofir & Sorek, 2018).

2.2.3. Phage Therapy

Phage formally known as bacteriophages, are virus that kill the targeted bacteria during its replication. The use of such viruses as therapeutic agents for the treatment of infectious pathogens is known as phage therapy (PT). Bacteriophages are the most common biological entities in nature and predators of bacteria. Phage therapy started since the discovery of phage by d' Herelle. Bacteriophage therapy is a sustainable and inexpensive way to deal with bacterial infections. When it comes to a plant pathogenic bacteria like *Pseudomonas syringae*, bacteriophage therapy can prove to be a reliable approach since it is specific to target bacteria and has the ability to self-replicate. Moreover, the therapy is non-toxic in nature thereby it releases only a minimal amount of detrimental chemical in nature (Brüssow, 2012a).

Phages are viruses that only infect bacteria; they do not directly harm plants or mammals. When a virulent phage infects a bacterium, it usually replicates quickly, lyses the bacterium, and releases a large number of progeny phages. After then, these phages can spread to nearby bacteria and infect them. As a result, when target pathogens are present, the phage population grows, thereby amplifying the therapeutic response to the bacterial infection. This is a clear benefit over alternative therapies like antibiotics.

2.2.3.1. History of Phage Therapy

The history of phage therapy starts with the discovery of bacteriophage which has been a matter of extensive debates and controversies over decades for priority claims. It was proposed that bacteriophages may be used to prevent and treat bacterial illnesses before the discovery and widespread use of antibiotics. The Indian rivers Ganga and Yamuna were found to have antibacterial activity in 1896, according to a report by British bacteriologist Ernest Hanbury Hankin, who worked as the government's chemical examiner and bacteriologist for the United Provinces and Central Provinces of India. This activity destroyed cultures of bacteria that cause cholera (Sulakvelidze et al., 2001). And he also suggested that these small heat liable substances pass through small-pored porcelain filter and were actually limiting spread of cholera epidemics. After two years

later, Nikolay Fyodorovich Gamaleya observed a similar phenomenon while working with *Bacillus subtilis* (Hermoso et al., 2007).

The main factor influencing the decision to halt commercial phage production at the time was the development of broad spectrum antibiotics. Phage therapy was initially developed and researched extensively in Russia, Georgia, and Poland, but it was eventually offered by the Eliava Institute and other institutions as well, including the Hirszfeld Institute of Immunology and Experimental Therapy in Wroclaw, Poland (Deresinski, 2009.

2.2.3.2. Phage Therapy in *Pseudomonas syringae*

Bacteriophages against the Pseudomonas genus are also known as Cystoviruses (phages f6-f14). These phages have genome made up of dsRNA viruses enclosed inside a lipid-membrane (Mindich et al., 1999). The first isolate and hallmark member of the pseudomonas phages is f6 that infects *Pseudomonas syringae* pv. phaseolica and has a 13,379 base pair long genome that is segmented into large, medium, and small dsRNA encoding total 14 gene products (Jäälinoja et al., 2007). The nucleocapsid shell of this bacteriophage is surrounded by a lipid membrane and a spike protein P3 protrudes from the membrane being anchored to the membrane protein P6 (Laurinavičius et al., 2004).

Despite being a promising approach to treat diseases, we need a better knowledge of the interaction between phages and hosts as well as a large collection of pertinent phages to use in therapy against *Pseudomonas syringae* species (Svircev et al., 2018).

2.3. Review of Literature Related to Host Range

2.3.1. Host Range of Bacteriophages

Host range of bacteriophage simply means the type of bacterial species that the phage is capable of infecting and transducing its DNA successfully. The main benefit of phages is their capacity to bind target bacteria with species and sub-species specificity; nevertheless, this capacity also has a major drawback, as the host ranges of isolated phages are frequently too limited or do not meet the needs of a particular application.

Based on their host range, phages can be classified into three fundamental types and the process involved in achieving it, that is, phages with three different host ranges: host-switching, broad, and narrow. Narrow-host-range phages are the most studied type as

these can infect only one host and is specific towards it. Contrastingly, broad-host-range species of phages can infect multiple bacteria. This capability is actually present in all individual phages but all species do not necessarily practice it. And the last is the host-switching groups that forms a grey area between the two previous types. The quasispecies here hold the ability to infect multiple bacterial host, yet the individual one can only infect the single host.

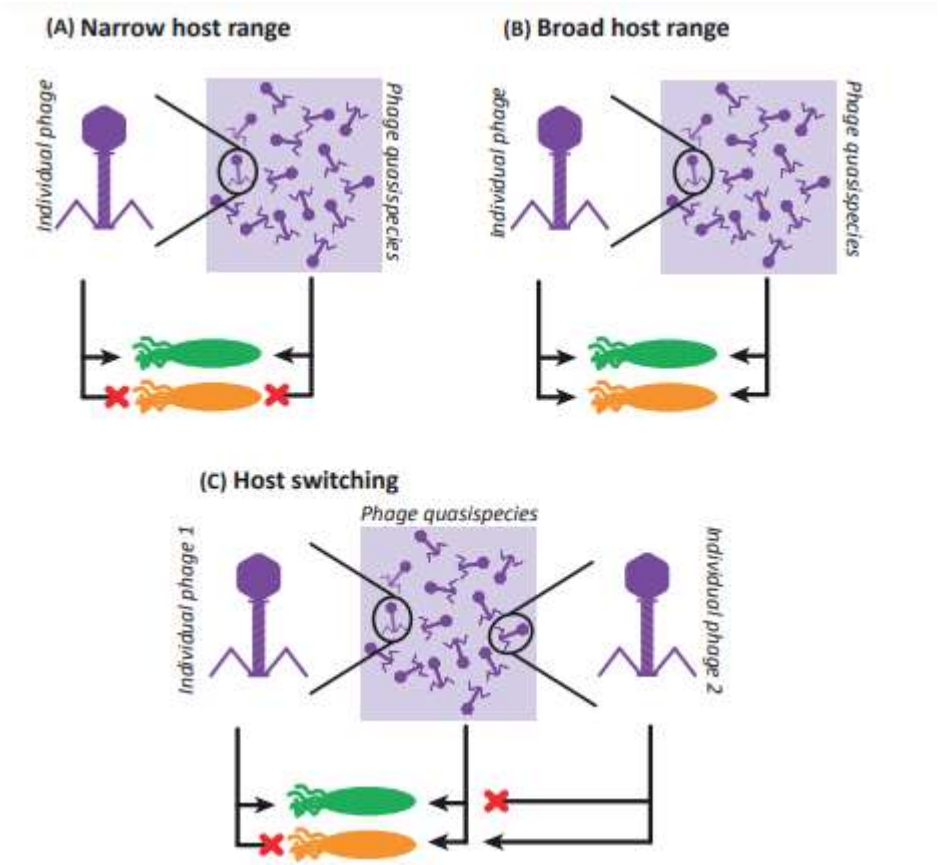


Figure 2: Subdivision of Bacteriophages Based on Host Range (Tu et al., 2017b)

The phenotypic processes comprise single particles with a large host range that hold the power to infect several host ranges. Whereas in genotypical mechanisms of host-switching phages, the whole quasispecies of a phage might infect numerous hosts when the individuals within can infect only one. Hence, the change in individual phage genotype results to broad host range (Tu et al., 2017b).

Bacteriophages having a broad host range may finish their life cycle in a variety of hosts that are classified differently, and these do not have distinct differentiation within varied levels of the wide host range. Both the bacteriophages which can invade multiple strains

of same species of bacteria and those that can infect various strains of the bacterium are considered constituting broad host range.

2.3.2. Host Infection Phenomena

Phages are viruses which attack and infect bacteria. Like other viruses, phages are obligatory parasites, that is, rely on their host to produce and discharge fresh virus particles. It is essential to comprehend the structural and molecular mechanisms behind phage-host interactions in order to effectively utilize phages in biotechnology and medicine. Furthermore, a fuller mechanistic understanding of this mechanism will shed light on the ecology and evolution of phages as well as their significance in the development of the human microbiome and microbial ecology at large. Both a phage-encoded receptor binding protein (RBP) and a host-encoded receptor (or receptors) are necessary for phage-host interactions. Every tailed phage uses its tail and related RBPs to communicate with the host and open a passageway for DNA to enter the cell. Despite this common tactic, the specific aspects of the protein machinery required for host adsorption vary throughout phage species. These characteristics may affect the phage host range. The receptor-binding proteins (RBPs) at the tip of the bacterium's tail are used by the phage to adhere to the bacterial surface, initiating the infection process of Pseudomonas syringae. These RBPs frequently develop into tail spike proteins (TSPs) or long or short tail fibers, which enable the phage to attach to particular receptors on the host bacterium's surface.

2.3.3. Determinants of Host Range

Phage receptor-binding molecules and host receptor compounds interact to cause bacteriophage attachment to the host bacterial surface, a crucial step in the phage infection process. Furthermore, phage receptor-binding proteins are essential for phage infection because the adherence of the receptor-binding protein to bacterial surface proteins initiates infection by tailed phages. (Węgrzyn, 2006).

2.3.3.1. Receptor Binding Proteins (RBPs)

Bacteriophages utilize the receptor binding proteins in their tail to attach to a particular receptor on the host bacterium's surface. These particular proteins establish a connection between the bacteriophage and its host by binding with bacterial receptors on the cell surface. But this isn't a one-on-one connection. Multiple phages may be able to infect the

same bacterium by binding to different bacterial receptors for bacteriophage adhesion. (Chaturongakul & Ounjai, 2014). These RBPs of tailed phages can localize in different substructures of the tail, namely fibers, spikes or baseplates. Although the architectures of tail spikes are essentially the same, their receptor-binding regions and the host-specific specificities they bestow are different. Moreover, enzymes that break through the cell membrane and promote DNA ejection are frequently carried by tail spikes. Tail fibers often don't have any enzymatic activity, in opposition to tail spikes, however certain tail fiber genes have sections encoding depolymerases. Similar to tail spikes, tail fibers are engaged in reversible and/or irreversible adsorption and feature extremely varied receptor-binding domains. The phage capsid's DNA is expelled by conformational changes that are initiated by tail fibers. There are other membrane-penetrating proteins, which differ from tail fibers and tail spikes in both structure and function. Their primary purpose is to cross the cell barrier, although very little is known about penetration of the inner membrane. Also, the phages harbor many different types of enzymes that break down peptidoglycan locally, such as lysozymes, endopeptidases, N-acetylmuramyl-l-alanine amidases, N-acetylglycosaminidase, and lytic trans-glycosylases. Enzymes targeting other structures in bacteria can also be detected in phages (Nobrega et al., 2018).

As phages are specific in their infection process, most express a single RBP whereas some exhibit multiple RBPs capable of recognising multiple different receptors. An in-depth molecular understanding of RBP-host interactions is required to understand how phages bind to and infect the bacteria of their interest (Ando et al., 2015); (De Jonge et al., 2019). Certain phage proteins contribute to the explanation of phage specificity. The initial stage of phage propagation involves these proteins recognizing and binding to specific receptors on the cell wall of bacteria (Samson et al., 2013). Their tail fiber proteins' C-terminal domain is the primary factor that determines host specificity. Simply altering this C-terminal region will change the host specificity of both phages.

2.3.3.2. Phage Receptors

Phage receptors, which consist of proteins, lipopolysaccharides, teichoic acids, and capsules, are molecules that are exposed to the cell surface and are encoded by bacteria (Heller, 1992). Similar to phage RBPs, the host receptors' type and structure play a critical role in defining the degree of specificity in phage-host interactions. Specificity is also

influenced by the location, quantity, and density of receptors on the cell surface. The lipid and lipoprotein composition, thickness, homogeneity, and receptors available for phage adsorption of the cell wall of Gram-positive and Gram-negative bacteria are significantly different from one another. The receptor that is displayed on the surface of cells could be a teichoic acid, protein, or carbohydrate moiety, or a mix of these (Casey et al., 2018).

2.4. Review of Literature Related to Phage Host Range Identification

Phage host range identification is a critical aspect of bacteriophage research, essential for understanding phage-bacteria interactions and applying phages in various fields, such as phage therapy, biocontrol, and microbial ecology. The host range of a bacteriophage determines which bacterial species or strains it can infect, a crucial factor for its application in targeting specific bacterial pathogens without affecting beneficial microbiota. The process of identifying a phage's host range typically involves isolating the phage and testing its infectivity against a panel of bacterial strains. Advances in molecular biology and genomics have further refined host range determination, allowing researchers to identify phage receptor-binding proteins and predict host range based on phage genomic data.

2.4.1. Prevalent Techniques for Phage Host Range Identification

The prevalent methods of phage host range identification involve a combination of experimental assays and, more recently, bioinformatics approaches. These methods are crucial for understanding the interactions between bacteriophages and their bacterial hosts, which has implications for phage therapy, biocontrol, and microbial ecology studies. Following are some of the currently used methods for such analysis:

2.4.1.1. Spot Tests

This simple method of phage host identification involves spotting a phage suspension onto a bacterial lawn cultured on an agar plate. A small volume of phage suspension is spotted onto a bacterial lawn that has been spread on an agar plate and incubate for a time period. After incubation, the presence of clear zones or plaques indicates where phages have lysed the bacteria. These are quick and easy to perform, allowing for the screening of multiple phage-bacterium combinations to determine which bacteria are

susceptible to which phages. But it is limited to qualitative analysis rather than a quantitative one.

2.4.1.2. Efficiency of Plating (EOP)

EOP is a more quantitative approach than spot tests, providing a ratio of the phage's ability to form plaques on different hosts compared to a reference host. It involves calculating the ratio of plaque-forming units (PFUs) on a test bacterial strain to the PFUs on a standard or reference strain. This method provides a more nuanced understanding of the phage's host range.

2.4.1.3. Host Range Mutants

This approach involves selecting or inducing mutations in phages to observe changes in their host range. By identifying which mutations expand or alter the host range, researchers can pinpoint specific genes or proteins involved in host recognition and attachment. This method provides insights into the molecular determinants of phage specificity and can be useful for engineering phages with desired host ranges.

2.4.1.4. Receptor Binding Analysis

Since phage infection starts with the binding to specific receptors on the bacterial surface, understanding this interaction is key to determining host range. Techniques like affinity chromatography can isolate and identify bacterial proteins that bind to phage tail fibers or other attachment structures. Protein foot printing can identify the interaction sites between phage proteins and bacterial receptors, giving a molecular-level understanding of the host range.

2.4.1.5. Cross-Infectivity Assays

These involve systematically testing a phage against a diverse panel of bacterial strains or species to map out its host range. Such assays provide comprehensive data on which bacterial strains are susceptible or resistant to the phage, offering a landscape view of its infectivity. This method is particularly useful in applications where a broad or specific host range is crucial, such as in phage therapy or biocontrol.

2.4.2. In-silico Techniques for Host Range Identification

In-silico techniques for phage host range identification are rapidly evolving, driven by advances in bioinformatics and the growing availability of genomic data. These methods offer a complementary approach to traditional experimental techniques, providing insights into the intricate relationships between phages and their bacterial hosts. Studies have demonstrated that genomic fingerprints, including codon use patterns, dinucleotide frequencies, and GC content, are frequently shared among phages and their hosts. Phage-host couples may be accurately predicted by finding their genetic similarities using tetra nucleotide frequency correlations (Ahlgren et al., 2016). Similarly, machine learning techniques have been applied to predict phage-bacterial infection networks, showcasing the potential of these models in host range prediction. In order to infer virus–host relationships, ML approaches utilize 'features', i.e., measurable properties of the object being analyzed such as the nucleotide and amino acid content of the viral genome, amino acid properties, and protein domains (Young et al., 2020).

Similarly, the CRISPR-Cas system in bacteria provides a historical record of phage infections. In-silico analysis of CRISPR spacers can reveal past interactions between phages and hosts, providing insights into host range. By comparing phage genomes to CRISPR spacer sequences in bacterial genomes, the potential host-phage relationships can be studied.

2.4.3. Protein-Protein Analysis for Host Range Prediction

Interactions between bacteriophage proteins and bacterial proteins are important for efficient infection of the host cell so these interactions may contribute to understanding the infectious interactions between bacteria and phages. The protein-protein interaction analysis focuses on identifying and characterizing the interactions between phage proteins, particularly those involved in host recognition and attachment, and bacterial surface receptors. Receptor Binding Proteins (RBPs) located on the phage's tail or capsid are key players in phage-host interactions as these are responsible for recognizing and binding to specific receptors on the bacterial cell surface, dictating the host specificity of the phage. Thus, analyzing the structure and biding properties of these RBPs, one can predict the host range that phage can affect. A study delved into the structural analysis of RBPs from phage T4, providing insights into the molecular basis of host recognition.

Besides, structural bioinformatics can predict the interaction between phage RBPs and bacterial receptors, even in the absence of empirical interaction data (Spinelli et al., 2005). This predictive approach was illustrated in research which applied structural modeling to predict protein-protein interactions, offering a framework that can be adapted for phage-host interaction studies (Xu & Dunbrack, 2019).

2.5. Protein-Protein Docking

Proteins bond in an extremely particular way after recognizing one another, usually in a crowded environment. In this process, molecules diffuse across a highly populated environment of various proteins and other biomolecular structures before precisely and structurally attaching (docking) to their intended protein partner. This process is quite astonishing, especially considering the vast size of these macromolecules, the great diversity of their structures, and the high density of the biomolecular milieu. Since its inception, protein docking—the process of inferring the structure of a protein-protein complex from the structures of the individual proteins—has undergone substantial development. This is due to the addition of more suitable energy functions, the use of potent methods for sampling the energy landscapes, and the utilization of the rapidly expanding corpus of information regarding protein interactions and structures. The principles of protein interaction that we currently understand significantly better than in the past, which aids in the development of improved docking techniques. Undoubtedly, the remarkable advancements in computer hardware have also been instrumental in facilitating novel approaches to protein interaction modeling and frequently permitting the application of previously impracticable concepts. However, because of their inherent nature, some fundamental docking principles have remained unexpectedly unaltered. As it was in the early days of the docking industry, steric and physicochemical complementarity remains the cornerstone of most docking techniques (Vakser, 2014b).

2.6. HADDOCK

HADDOCK is an information-driven flexible docking approach for the modelling of biomolecular complexes. Using this software, docking occurs in different degrees of flexibility and in different type of chemical environments through three different stages. The first stage involves randomization and rigid minimization in which the interacting bodies are treated as rigid bodies. Thus, these are separated in space and are randomly

rotated about their centers of mass. This is followed by rigid body minimization step, where molecules are allowed to rotate and translate to optimize the interaction. Moving to the second stage where semi-flexible simulated annealing occurs in torsion angle space. This particular process introduces flexibility to the interacting partners through 3 stepped molecular dynamics based refinement. This refinement occurs so that the complex can achieve interface packing. Here, first the molecules are kept rigid and inly their orientations are optimized then flexibility is introduced in the interface while residues move their side chains. Finally, the backbone and side-chains of the flexible interface are freed. Now, at the last stage, the docked molecules are refined in Cartesian space using explicit solvent. In fact, this refinement occurs only if the user allows it, otherwise, only a short minimization occurs and results are displayed. If allowed, the complex immerges in a solvent shell, usually water, to improve the energetics of the interaction.

HADDOCK was created to incorporate experimental data into the docking process and provide highly accurate simulation of protein-protein, protein-nucleic acid, and protein-ligand interactions. Understanding intricate biological processes and the field of drug design, where knowledge of molecular interactions can guide the creation of novel treatments, are two areas in which this is very helpful. To assist with the docking process, this integrative tool incorporates data from multiple experimental sources, including bioinformatics, mutagenesis, and NMR. By incorporating experimental data, conventional docking methods—which frequently struggle with the dynamic nature of protein interactions can overcome some of their shortcomings (Van Zundert et al., 2016; Honorato et al., 2021).

2.7. HADDOCK Prodigy

HADDOCK Prodigy (PROtein binDIng enerGY prediction tool) is a tool inside the HADDOCK server and has been designed to study the binding affinity between the proteins in a complex. Using this tool, we can use the structural data gained from protein docking using HADDOCK and predict the binding efficiency between any two interacting proteins. Prodigy assesses different parameters of interaction like as intermolecular contact, hydrogen bond the molecule, and its hydrophobic effects. This tool provides a detailed analysis of thermodynamics involved to predict binding free energy of interacting

complex. It aids in the identification of important factors influencing binding affinity, directs the engineering of proteins with improved interaction characteristics, and makes the process of looking for possible therapeutic candidates easier. Prodigy, an expansion of the HADDOCK docking program, is notable for combining energetic predictions and structural modeling, providing a thorough method for researching protein interactions (Vangone & Bonvin, 2015).

3. MATERIALS AND METHODOLOGY

3.1. Selection of bacterial protein targets and development of leads

3.1.1. Selection of bacteria's host attachment protein and obtaining their genomic sequences

Among the several membrane proteins involved in interaction of *Pseudomonas syringae* with varied range of hosts, Type IV pilin (Pilin) protein that allows bacterial cells to adhere to external structures of the host plant was taken as the protein target for further experimentation. The whole genome sequence of this protein of *Pseudomonas syringae*, its pathovars and *Pseudomonas aeruginosa* were taken for carrying out alignment and structural analysis.

3.1.2. Alignment studies for Pilin protein of *Pseudomonas syringae* pilin with other pathovars

Different alignment studies based on sequence comparison were performed for the sequence of *Pseudomonas syringae* Pilin protein deposited in Uniprot database to explore the close association and relation to *Pseudomonas syringae* pv. syringae (Pss), *Pseudomonas syringae* pv. actinidiae (Psa), *Pseudomonas syringae* pv. *savastanoi* (Psav), *Pseudomonas syringae* pv. phaseolicola (Pph) and *Pseudomonas aeruginosa*.

3.1.2.1. Alignment search using Blast P for *Pseudomonas syringae* Type IV pilin

The sequence for *Pseudomonas syringae* Type IV Pilin was searched in Uniprot database and BLAST tool in the same database was employed for alignment studies with the intention to find the matches in sequences to Pss, Psa, Psav, Pph and *Pseudomonas aeruginosa*.

3.1.2.2. Multiple Sequence Alignment search CLUSTAL OMEGA for *Pseudomonas syringae* Type IV Pilin

The protein sequences of *Pseudomonas syringae* pilin(A0A1C7YW38), Pss pilin (A0A0Q0BCD2), Psa pilin (A0A7Z6U3M0), Psav pilin (A0A3M5E045), Pph pilin (A0A7Z6UU48) and *Pseudomonas aeruginosa* pilin (A0A3M5E045) were downloaded in FASTA sequences, and the Clustal Omega tool was used for performing the multiple

sequence analysis and construction of phylogenetic tree amongst the three strains in consideration.

3.1.3. Selection of Phage Receptor Binding Protein

The phage receptor binding protein P3 that binds with type IV pilus of its bacterial host to initiate infection was searched in Uniprot database. The FASTA sequence of this protein sequence (P11129) was retrieved for further experimental analysis.

3.2. Prediction of structures of the protein and protein complexes

Since the experimental 3D structure of our proteins of interest were not available in Protein Data Bank (PDB), we predicted the structures from the protein sequences using the AlphaFold Colab (Jumper et al., 2021) notebook and used AlphaFold Multimer for predicting the structures for protein complexes.

3.2.1. Load Sequences in AlphaFold Colab

The protein sequences for P3 protein of Phi6 retrieved earlier was copied and pasted in the query sequence blank for AlphaFold Colab notebook and hit run all after connecting to a server. The same process was repeated for Pilin protein of *Pseudomonas syringae*, its pathovars Pss, Psa, Psav, and Pph and *Pseudomonas aeruginosa*. Since the structure of *Pseudomonas syringae* pv. phaseolicola (Pph) was taken as a reference value for the study, its structured was also built using AlphaFold.

Likewise, for predicting structure of the P3 and Pilin protein complex, the sequences were inserted together separate by a colon sign ":" and allowed to run.

3.2.2. Prediction and selection of structures of individual proteins and protein-protein complexes

After a successful runtime, the predicted structures were analysed to select the best fitted one with higher sequence coverage and optimal parameters.

3.2.3. Structural Alignment Analysis

For validation of the predicted structure, the protein structure was built using I-TASSER and Swiss Model to find the best possible structure for docking. I-TASSER also uses TM-

align structural alignment program to match the first I-TASSER model to all structures in the PDB library.

3.3. Predict binding affinity of Protein-Protein Complexes using HADDOCK Prodigy

The Haddock Prodigy web server was used to predict the binding affinity of the protein-protein complexes built earlier. This tool predicts the binding affinity based on intermolecular contact between the protein's chains in the complex.

3.3.1. Load PBD and specify the chain

The best suited protein complex structure was chosen and uploaded in HADDOCK Prodigy Web portal. Then, chain IDs were specified for both the interactor proteins in the complex and temperature was set to default setting of 25°C. A personalized job ID was given to identify the run and similar set of steps was repeated for all five complexes of Spike Protein P3 and Type IV Pilin protein of different species of Pseudomonas and their pathovars.

3.4. Perform protein-protein docking using HADDOCK

To perform a comparative study, the proteins were separately docked with each other using the HADDOCK software. For this experimental study, HADDOCK web server 2.4 was used.

3.4.1. Select the PDB files and edit them

The PDB files of structures of P3 and different pilin proteins were visualized in PyMOL for presence of extra atoms, ions, charges, and water molecules. Then the PDB files were uploaded in PBD Tools webserver available for the software. All necessary editing in the structure was made following the given commands and the output was downloaded.

3.4.2. Load PBD files

The HADDOCK webserver was opened in the browser and logged in with the credentials. An option of Submit a new job was clicked to open the HADDOCK submission. A definitive Job name was given for the docking process and the edited PDB structure of P3 protein was input as Molecule 1 whereas the edited Pilin Protein of *Pseudomonas syringae* was input as Molecule 2. The procedure was repeated after each cycle by replacing Molecule

2 with Pilin protein of other species and pathovars. For each job, a different job name was specified such as:

P3-Pilin 0: for P3 protein and pilin protein of *Pseudomonas syringae*

P3-Pilin 1: for P3 protein and pilin protein of *Pseudomonas syringae* pv. syringae

P3-Pilin 2: for P3 protein and pilin protein of *Pseudomonas syringae* pv. savastanoi

P3-Pilin 3: for P3 protein and pilin protein of *Pseudomonas syringae* pv. actinidiae

P3-Pilin 4: for P3 protein and pilin protein of *Pseudomonas aeruginosa*

P3-Pilin 5: for P3 protein and pilin protein of *Pseudomonas syringae* pv. phaseolicola

3.4.3. Check the Input Parameters

Once loaded, the Input parameters were set for both the protein molecules. The active and passive residues for interaction during docking were given. These residues were selected after analyzing active residue pairs suggested during binding affinity experiment done earlier with protein-protein complex using HADDOCK Prodigy.

3.4.4. Check the docking parameters and hit run

For docking parameters the pre-set default value was used as we don't yet have specific requirement mentioned for docking these proteins in previous research. The Submit button was clicked to initiate the docking process.

3.4.5. Docking Analysis and structure selection

After successful docking run, the obtained structure were visualized, the values were analyzed and a suitable docked structure was selected.

4. RESULT AND DISCUSSION

4.1. Search for bacterial protein targets

The wide spread invasion of crop and cash crops around the world by plant pathogenic species of *Pseudomonas* and development of resistance to copper derivatives has led to the need of novel approach for treating the infections. Amongst the number of receptors involved in establishing pathogenic infection with different plant species, the Type IV Pilin protein of *Pseudomonas syringae*, which is responsible for facilitating the attachment of this plant pathogen on the leaf surface of host plant was chosen as it is also involved in interaction for attachment of bacteriophage. This chromosomally encoded pilin protein acts as the initial site of cellular attachment for bacteriophages like Phi 6 and other members of dsRNA virus from the Cystoviridae family and this attachment is mediated by the viral attachment protein also known as Phage Spike protein P3 (Mindich, 2006; Gottlieb et al. 1988). Once attached, the retraction in pilus moves Phi 6 particles in a closer proximity to the bacterial cell membrane and allows the fusion of membrane leading to entry of phage into the bacterial cell through the activity of viral lytic enzyme (Mindich, 2006). Following the fusion, the phage Phi 6 makes use of its host cell metabolism to cause a lytic infection and release mature virions to infect host cells expressing the type IV pilus (Gottlieb et al. 1988). Considering the indispensable nature of Type IV pilin protein in the infection of Pseudomonas strains by bacteriophage phi 6, it was considered as a target protein and its receptor P3 as a receptor protein involved in this protein-protein interaction.

4.1.1. Selection of bacterial protein and obtaining genomic sequences

The Type IV Pilin protein was chosen as a protein target and its protein sequences was obtained from Uniprot database. These protein sequences provide the basic about the preferred codons, and it was seen that Type IV Pilin of *Pseudomonas syringae* consisted of 142 amino acid sequences on it whereas its pathovars syringae, savastanoi and actinidiae had 144 amino acid sequences. Similarly, the protein sequence for *Pseudomonas syringae* pathovar phaseolicola was 139 amino acids and P aeruginosa was 158 amino acids

4.1.2. Alignment studies for Type IV Pilin protein

Different alignment studies based on sequence comparison using different algorithms and parameters were performed for the sequence of *Pseudomonas syringae* Type IV pilin protein deposited in Uniprot database to explore the close association and relation to other pathovars like Psa, Pss, Psav, and Pph and distantly related *Pseudomonas aeruginosa* pilin proteins.

4.1.2.1. Alignment search using Blast P for *Pseudomonas syringae* Type IV pilin genome

The primary sequences of *Pseudomonas syringae* was compared with that its pathovars and *Pseudomonas aeruginosa* and it was found that *Pseudomonas syringae* shares identity percentage of 81.69% with *Pseudomonas syringae* pv. syringae, 96.74% with *Pseudomonas syringae* pv. actinidiae and 76.76% with *Pseudomonas syringae* pv. savastanoi. However, the information regarding identity with *Pseudomonas syringae* pv. phaseolicola and *Pseudomonas aeruginosa* wasn't obtained.

Table 1: p-BLAST results for Pilin protein of Pseudomonas syringae pathovars with reference of Pseudomonas syringae pilin

Subject sequence	Max score	Total score	Query coverage	E-value	Percentage identity
Pss	228	228	100%	5e-74	81.69%
Psa	273	273	100%	8e-92	96.45%
Psav	205	205	100%	3e-65	76.76%

4.1.2.2. Multiple Sequence Alignment Studies

The multiple sequence alignment studies for *Pseudomonas syringae*, its four pathovars, and *Pseudomonas aeruginosa* was performed using the Clustal Omega tool. The FASTA sequences for the strains *Pseudomonas syringae* (Uniprot Accession ID: A0A1C7YW38), Pss (Uniprot Accession ID: A0A0Q0BCD2), Psa (Uniprot Accession ID: A0A7Z6U3M0), Psav (Uniprot Accession ID: A0A3M5BU11), Pph (Uniprot Accession ID: A0A3M5E045) were used for the study. From the sequence alignment studies, the results show the conservation of residues in one domain, the transmembrane domain (residues 12-30) of the *Pseudomonas syringae* transmembrane domain.

Figure 3: Sequence alignments for Pseudomonas syringae, its pathovars, and Pseudomonas aeruginosa

The results from the multiple sequence alignments were then interpreted to study the phylogenetic relationship between these three closely related strains of the bacteria. It appears as if the *Pseudomonas syringae* is more closely related to *Pseudomonas syringae* pv. phaseolicola and *Pseudomonas aeruginosa* in comparison to *Pseudomonas syringae* pv. syringae, *Pseudomonas syringae* pv. actinidiae, and *Pseudomonas syringae* pv. savastanoi.

Figure 4: Phylogenetic tree generation for Pseudomonas syringae, its pathovars, and Pseudomonas aeruginosa.

4.1.3. Selection of Phage Receptor Binding Protein (RBP)

The Spike Protein P3 of bacteriophage Phi6, also called as phage receptor binding was known to bind with Type IV pilin of its host bacteria, *Pseudomonas syringae* and its pathovars thus, it was selected as a protein to dock with our target protein. The FASTA sequence of the Spike protein P3 (Uniprot Accession Id: P11129 P3_BPPH6) was retrieved

from the Uniprot database and it was subjected to blast with intention to find matches with a protein with predicted 3D structure. However, the results showed that protein has lesser than 30% identity with other proteins and none with a verified experimental structure.

4.2. Prediction of structures of the protein and protein complexes

The 3D structures of our protein of interest the phage receptor, Pilin and its receptor binding protein, P3 as well as their complexes were predicted using AlphaFold and AlphaFold Multimer respectively.

4.2.1. Selection of Protein Structure

The software suggested five different structures with different sequence coverage, predicted Local Distance Difference Test (pLDDT), and predicted Template Modelling (pTM) for individual structures and only the one with highest values for these parameters were selected. Similarly, for the complexes five different models with different sequence coverage, pLDDT, pTM, and interface predicted Template Modelling (ip TM).

Table 2: pLDDT and pTM score of P3 protein of Phi6 and Pilin protein of Pseudomonas syringae, its pathovars and Pseudomonas aeruginosa.

Organism	**Protein**	**pLDDT value**	**pTM value**
Pseudomonas syringae	Pilin	89.5	0.773
Pseudomonas syringae pv. Syringae	Pilin	94.2	0.821
Pseudomonas syringae pv. Savastanoi	Pili	94.2	0.817
Pseudomonas syringae pv. Phaseolicola	Pilin	90.6	0.787
Pseudomonas aeruginosa	Pilin	94.4	0.828
Phi 6	Spike Protein P3	27.4	0.179

In case of pilin protein of *Pseudomonas syringae*, the structure with highest pLDDT value of 89.5 and pTM value of 0.773 was selected. Similarly, the pLDDT value and pTM value

for four pathovars and *Pseudomonas aeruginosa* as displayed above in the table were above 50 indicating the high confidence of the structure. Whereas, for Spike protein P3, the pLDDT value was below 50 and pTM value below 0.2 suggesting the lower confidence of the structure. To find a more reliable structure, the structure was predicted using Swiss model but the sequence coverage tended to be much lower than the coverage obtained from AlphaFold. Hence the structure predicted by AlphaFold was used.

Similarly, for building the protein-protein complexes, AlphaFold Multimer was used and the structure with highest pLDDT, pTM and iP TM value was selected for further analysis. Ideally the model was supposed to have pLDDT value greater than 50 for a higher confidence, however, our values were close to 40 and hence accepted. Likewise, the pTM value were over 0.2 indicating probability of correct interface modeling and ip TM values over 0.15.

Table 3: pLDDT, pTM, and ip TM scores for P3-Pilin complexes of Phi6 and different species and pathovars of Pseudomonas.

Complex	**Organisms**	**pLDDT value**	**pTM value**	**ip TM value**
P3-Pilin 0	Phi 6 & *Pseudomonas syringae*	35.9	0.274	0.152
P3-Pilin 1	Phi 6 & *Pseudomonas syringae* pv. syringae	36.9	0.273	0.155
P3-Pilin 2	Phi 6 & *Pseudomonas syringae* pv. savastanoi	38	0.28	0.151
P3-Pilin 3	Phi 6 & *Pseudomonas syringae* pv. actinidiae	36.9	0.273	0.155
P3-Pilin 4	Phi 6 & *Pseudomonas syringae* pv. phaseolicola	35.7	0.283	0.166
P3-Pilin 5	Phi 6 & *Pseudomonas aeruginosa*	38.4	0.285	0.145

4.2.2. Ramachandran plot analysis

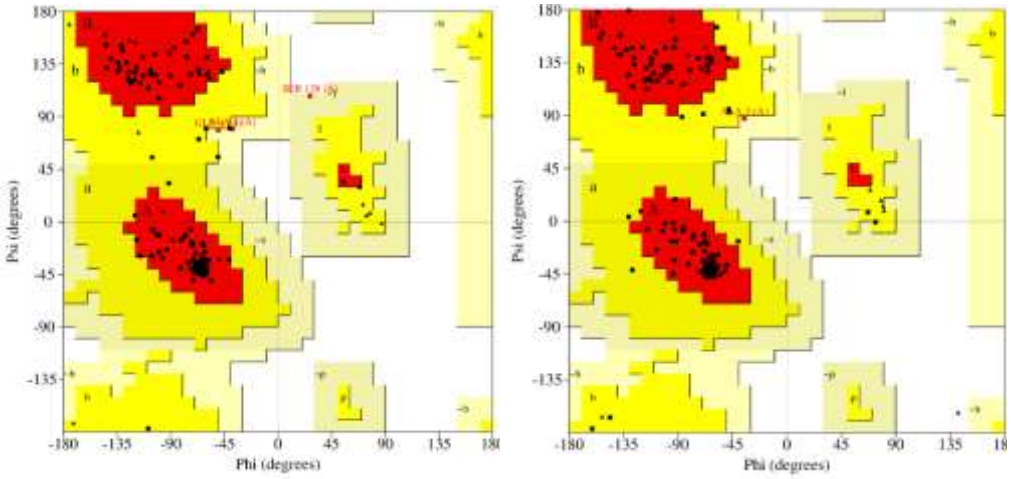
The main basic requirement for a protein model is correct stereochemistry. Anomalies, such as phi/psi angle combinations that are placed in disallowed regions, steric collisions, and unfavorable bond lengths and angles are analyzed by programs such as PROCHECK

(Laskowski, 1993) and WHATCHECK (Hooft et. al., 1996). These programs also analyze these stereochemical features of the residues in the model and give an evaluation of the overall quality of a model or structure (Bhattacharya et. al., 2007). Highlighting unrealistic conformations within the model by analysis of bond geometry by looking at Ramachandran plots is important. Certain conformations of phi and psi angles are forbidden in protein structures because they result in steric hindrance, or clashes between atoms. A good model will generally have 90% of its residues in the allowable regions of a Ramachandran plot (Laskowski, 1993).

For our predicted protein structures in consideration, the majority of residues were in allowable regions indicating that there were only few phi/psi angle combinations in the disallowed region as depicted in the table and figures below. Only the pilin protein of *Pseudomonas syringae* pv. phaseolicola (0.8%) and P3 protein (3.3%) had residues in disallowed regions.

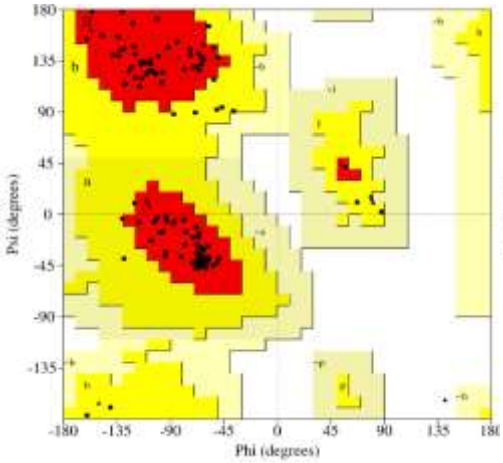
Table 4: Ramachandran Plot Analysis for predicted structures

Protein	Organism	Residues in allowed region	Residues in additional allowed region
P3	Phi 6	62.9%	26.6%
Pilin	*Pseudomonas syringae*	89.8%	9.4%
Pilin	*Pseudomonas syringae* pv. syringae	89.9%	10.1%
Pilin	*Pseudomonas syringae* pv. savastanoi	91.5%	8.5%
Pilin	*Pseudomonas syringae* pv. actinidiae	91.5%	8.5%
Pilin	*Pseudomonas aeruginosa*	88.2%	8.8%
Pilin	*Pseudomonas syringae* pv. phaseolicola	88.3%	8.3%

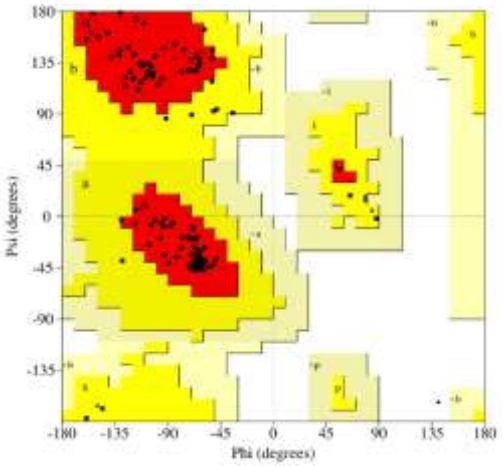


a. Pseudomonas syringae

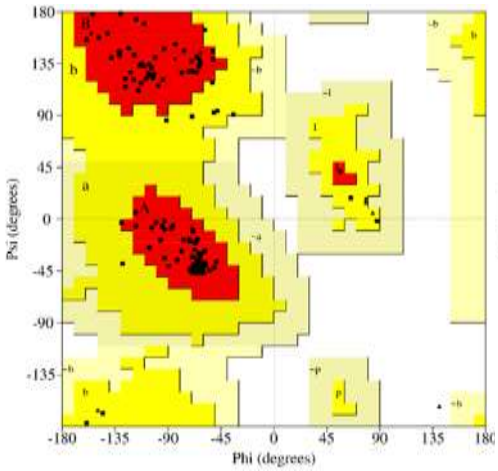
b. Pseudomonas syringae pv. syringae



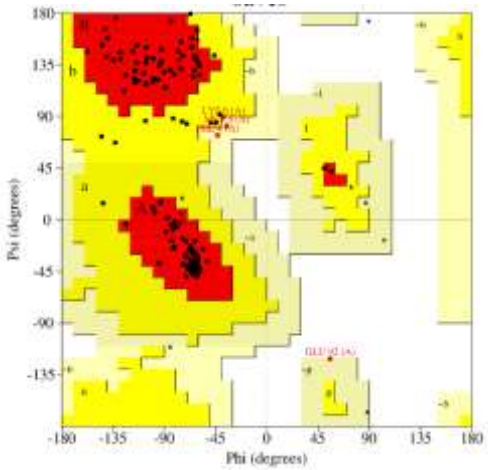
c. Pseudomonas syringae pv. savastanoi



d. Pseudomonas syringae pv



e. Pseudomonas aeruginosa



f. Pseudomonas syringae pv. Phaseolicola

Figure 5: Ramachandran plot analysis using SAVES V 5.0 for amino residues of predicted structure of Pilin protein for a. Pseudomonas syringae, b. Pseudomonas syringae pv. syringae, c. Pseudomonas syringae pv. savastanoi, d. Pseudomonas syringae pv. actinidiae, e. Pseudomonas aeruginosa and f. Pseudomonas syringae pv. phaseolicola respectively.

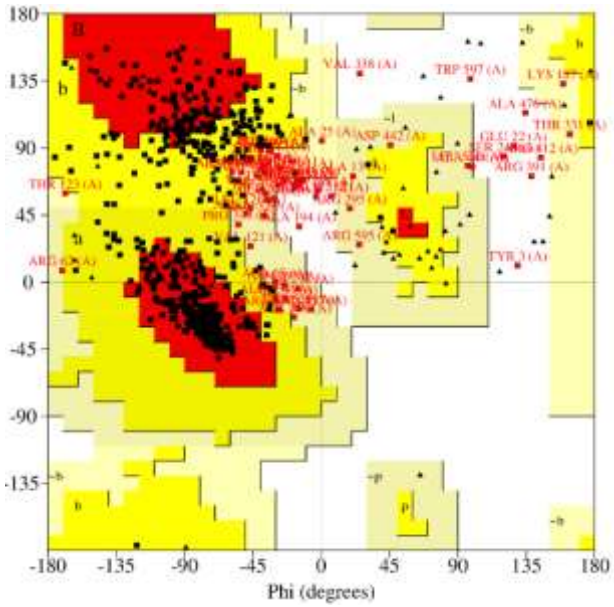


Figure 6: Ramachandran plot analysis using SAVES V 5.0 for amino residues of predicted structure of P3 protein

4.2.3. Structure Validation Analysis

For validation of the structure predicted by AlphaFold, the protein sequences were taken and the 3D model of the protein was built using I-TASSER. The top model built with the software had a C-score of -1.83, an estimated TM-score of 0.49 +/- 0.15 and estimated RMSD of 12.3 +/- 4.3 A°. The C-score of the structure is within the range of (-5, 2) indicating the average confidence.

Figure 7: P3 protein 3D structure using I-TASSER

After the structure assembly simulation, the software used the TM-align structural alignment program to match the top structure model with all structures in the PDB library and presented top 10 identified structural analogs. A TM score above the cutoff value of

0.5 indicates a correct topology of the structure. Similarity for top model was 2.01 which is an acceptable value.

Top 10 Identified stuctural analogs in PDB

Click to view	Rank	PDB Hit	TM-score	RMSD[a]	IDEN[a]	Cov	Alignment
○	1	3ai7A	0.923	2.01	0.153	0.958	Download
○	2	3ahgA	0.923	2.02	0.150	0.958	Download
○	3	8io8A	0.899	2.46	0.139	0.954	Download
○	4	2g25B	0.730	4.50	0.069	0.872	Download
○	5	3rimA	0.704	4.59	0.092	0.841	Download
○	6	1itzC	0.694	4.42	0.086	0.821	Download
○	7	5nd5A	0.693	4.51	0.097	0.823	Download
○	8	1ay0B	0.690	4.43	0.083	0.812	Download
○	9	8wa8A	0.689	3.91	0.090	0.790	Download
○	10	4c7vA	0.688	4.33	0.079	0.812	Download

Figure 8: Score of top 10 PDB hits from structural alignment

Figure 9: Structural Alignment of Predicted Protein with top hit PBD ID: 3ai7A

4.3. Predict binding affinity of Protein-Protein Complexes using HADDOCK Prodigy

The Haddock Prodigy web server was used to predict the binding affinity of the protein-protein complexes of Spike Protein P3 with Type IV Pilin protein of different pathovars of *Pseudomonas syringae* and *Pseudomonas aeruginosa*. It was observed that the P3-Pilin complex of Phi 6 with *Pseudomonas syringae* pv. phaseolicola had the highest binding affinity of -17.3kcal mol-1 whereas the P3-Pilin complex of Phi 6 and *Pseudomonas*

aeruginosa had the lowest binding affinity of -6.6. Likewise, the binding affinity of P3-Pilin for *Pseudomonas syringae* pv. actinidiae (-12.7) and *Pseudomonas syringae* pv. syringae (-12.2) was closer to that of highest value.

Table 5: Binding affinity and Kd prediction for P3-Pilin Complexes

Protein-Protein Complex	**Organisms**	**Number of intermolecular contacts**	**Binding Affinity (kcal mol-1)**	**Dissociation Constant (Kd at 25°C)**
P3-Pilin 0	Phi 6 & *Pseudomonas syringae*	62	-8.9	3.2e-07
P3-Pilin 1	Phi 6 & *Pseudomonas syringae* pv. syringae	112	-12.2	1.2e-09
P3-Pilin 2	Phi 6 & *Pseudomonas syringae* pv. savastanoi	83	-8.1	1.1e-06
P3-Pilin 3	Phi 6 & *Pseudomonas syringae* pv. actinidiae	118	-12.7	4.7e-10
P3-Pilin 4	Phi 6 & *Pseudomonas aeruginosa*	36	-6.6	1.4e-05
P3-Pilin 5	Phi 6 & *Pseudomonas syringae* pv. phaseolicola	181	-17.3	2.1e-13

The total number of intermolecular contacts for P3-Pilin complex 5 was 181 which is obvious since the phage Phi6 was discovered against the pathovar phaseolicola. Here as well, the number of intermolecular contacts of complex with pilin protein of pathovars syringae and actinidiae were closer to that of the pv phaseolicola. The lowest number of intermolecular contacts were seen with *Pseudomonas aeruginosa*.

Table 6: Active residues for different P3-Pilin complexes

Complex	**Active Residues on P3**	**Active Residues on Pilin**
P3-Pilin 0	Val99, Gly100, Tyr101, Ala102, Cys103, Pro104, Gly105, Leu106, Asp107, Leu108, Arg110, Ala111, Ala268, Ser269, Lys270, Trp271, Gly520, Arg522, Arg617, Gln621, Leu625	Tyr30, Tyr33, Glu41, Ala45, Ser48, Ala49, Thr51, Cys52, Val53, Thr54, Glu55, Thr56, Val57, Ala58, Ser59, Ala68, Cys69,Ala70, Thr71, Gly72, Phe73, Val74, Pro75, Thr76, Thr77, Tyr78, Met82, Thr118, Trp120, Thr121
P3-Pilin 1	Gln130, Ser132, Asp189, Trp190, Glu191, Gly192, Ser193, Ala194, Val195, Ala196, Asp197, Trp214, Gln215, Ser216, Val217, Gly218, Gly219, Ala220, Tyr222, Ser224, Asp281, Arg522, Asp533, Ala534,	Leu22, Ala23, Ala26, Ile27, Ser29, Tyr30, Asn31, Tyr33, Thr34, Leu35, Lys36, Ala37, Lys38, Ser40, Glu41, Ser46, Ala47, Pro48, Ala49, Gln50, Gln51, Ala52, Leu53, Ala54, Leu55, Ala56, Asn58, Gln59, Thr80, Lys82,

	Asp535, Glu536, Gln537, Asp538, Tyr615, Arg619, Lys 620, His623, Ser624, Ala627, Ser628	Tyr83, Val123, SER125, Ala126, Gln127, Trp130, 143, Thr144
P3-Pilin 2	Ser118, Ala119, Ser120, Val121, Thr122, Thr123, Gly124, Ser125, Gln127, Trp190, Glu191, Thr192, Ser193, Ala194, Val195, Gln219, Gln266,Lys 270, Gln352	Leu62, Ser63, Ala64, Ala65, Thr66, Thr67, Ser70, Ala75, Ser86, Thr88, Val90, Gly91, Thr92, Ser93, Ala94, Thr95, Gln96, Gly97, Thr100, VAl101, Val102, Ala109, Asp110, Val111, Asp112, Ala113, Lys114, Thr115, Thr135
P3-Pilin 3	Ser120, Thr122, Thr123, Gln127, Gln130, Asp189, Trp190, Glu191, Gly192, Ser193, Ala194, Val195, VAl217, Gly218, Gly218, Gly219, Ala220, LGy221, Tyr222, Trp271, Met273, Pro275, Leu276, Leu277, Met278, Ile279, Val280, Asp281, Arg522, Val531, Val532, Asp533, Ala534, Asp535, Glu536, Gln537, Asp618, Arg619, Lys620, Gln621	Leu22, Ala23, Ala24, Val25, Ala26, Ile27, Pro28, Ser29, Tyr30, Tyr33, Thr34, Lys36, Ala37, Lys38, Se40, Glu41, Ala42, Ala47, Pro48, Ala49, Gln50, Gln51, Ala52, Leu53, Ala54, Leu55, Ala56, Phe57, Asn58, Asp59, Gly60, Leu71, Gly72, Leu73, Ser81, Lys82, Tyr83, Ala105, Thr106, Met120, Cys122, Ser125, Ala126, Gln127, Cys128,Thr 129, Trp130, Lys143, Thr144
P3-Pilin 4	Val121, Thr122, Thr123, Gly124, Ser125, Leu126, Gln127, Glu191, Gly192, Ser193, Ala194, Val195, Thr245, Tyr251, Met273, Pro275	Glu66, Val68, Asn72, Thr74, Gln77, Lys89, Ser91, Gly94, Ile96, Thr97, Val98, Lys121, Gly122, Ala123, Val124, Ser140, Ser141, Thr142, Pro143, Ala144
P3-Pilin 5	Leu85, Asp86, Trp88, Gly89, Gly90, Thr91, Gly92, Asp93, Ala94, Trp95, Gln96, Gln97, 122, Thr123, Ser125, Gln127, Gln130, Ser193, Ala194, Ala196, Asp197, Tyr199, Trp214, Gln215, Ser216, Val217, Gly218, Gly219, Ala220, Thr245, Ser246, Gln352, Gly353, Gln355, Tyr357, Gly520, Arg522, Val531, Asp533, Asp535, Glu536	Val53, Ala54, Ile57, Gln60, Val62, Ala63, Asn64, Gly65, Leu68, Asn69, Thr70, Asp75, Thr76, Ser84, Val85, Ala86, Val87, Asn88, Asn89, Gly90, Val91, Ile92, Thr93, Val94, Val95, Ala98, Asn101, Val102, Ala103, Ser104, Thr105, Leu106, Val107, Leu108, Thr109, Pro110, Val111, Leu112, Gly113, Ile117, Thr18, Trp119, Asn120, Lys121, Ala122, Gly123, Ser124, Gly125, Cys126, Leu127, Val128, Ala129, Gln130, Ala131, Ala133, Ile135, Leu139

Apart from binding affinity and interface contacts, the information about active residue pairs in contact at a distance of 5.5 A° and the interacting pdb file were also achieved from the prodigy result.

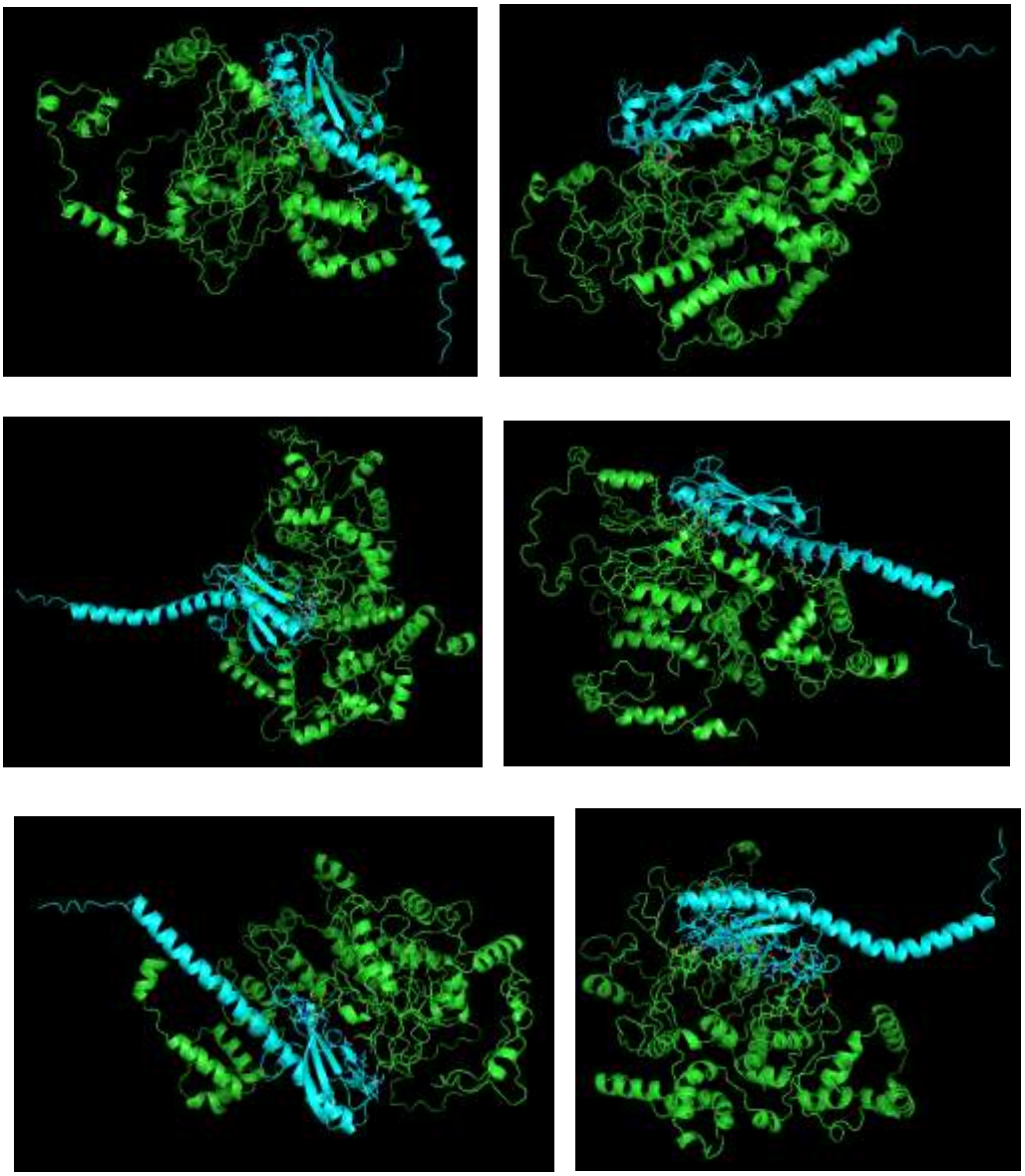


Figure 10: Protein interfaces in P3-Pilin Complex 0, P3-Pilin Complex 1, P3-Pilin Complex 2, P3-Pilin Complex 3, P3-Pilin Complex 4, and P3-Pilin Complex 5 respectively.

4.4. Molecular docking Simulation using HADDOCK

4.4.1. Protein Preparation

The Phi 6 spike protein P3 and Type IV Pilin protein were selected as receptor protein binding protein and phage receptor protein to understand the host-range of bacteriophage against plant pathogenic *Pseudomonas syringae*. The predicted structure for these proteins were obtained from AlphaFold. Before docking simulations in HADDOCK web server, the conversion of PDB into pdbqt format and ending the files with an end statement is a must. The obtained structure were first visualized in PyMOL for presence of water molecules, ions, and any crystallization agent. As none of them were present, we proceeded to edit the PDBs using the pdb tools available in the server.

The pdb file had single chain hence, it wasn't necessary to add a TER statement. After giving the command for a unique name, validation of structure and tidy, the desired output was achieved. The procedure was repeated for the P3 protein and all Type IV Pilin protein structures and the files were saved.

4.4.2. Identification of Active Binding Site

It is necessary to define the active binding sites between the interacting molecules before initiating the docking process. The transmembrane domain was also stated to be a protein-protein interaction site for the Type IV pilin but the active binding site for P3 protein wasn't defined. Thus, we took reference from the active residues pairs defined in the HADDOCK Prodigy results. We analysed the active residues for each pair of P3-Pilin complexes and enlisted them to use for defining input parameters.

4.4.3. Upload Dockable Protein Structure

The edited pdb files were uploaded in the HADDOCK webserver and P3 protein structure was loaded in the place for Molecule 1 whereas Pilin protein was loaded in the Molecule 2. The process was repeated for Pilin protein of all Pseudomonas species and their pathovars.

4.4.4. Define Input Parameters

The active residues as displayed in the table (6) was chosen for each combination of P3 and Pilin Protein for respective organisms and for passive residues, actively define passive residues around the active residues was chosen. By default setting, the minimum percentage of relative solvent accessibility to automatically define surface neighbors of active residues as passive was set to 40.0. The histidine protonation state wasn't specified for the run and the flexible segments defined were set to automatic.

4.4.5. Define Docking Parameters

Since the active residues for this docking were known, the ambiguous distance restraints was already defined and due to this none other distance restraints were altered from the preset settings. Likewise, on default setting, the random exclusion for ambiguous restraints (AIRs). The sampling parameters was set as following:

No. of structures for rigid body docking: 1000

No. of trials for rigid body minimization: 5

No. of structures for semi-flexible refinement: 200

Solvent used: Water

No. of structures for final refinement: 200

No. of structures to analyze: 200

Cutoff Value for neighboring residues: 5 A°

And for clustering parameters, Fraction of Common Contacts (FCC) was used with a cutoff value of 0.66 and minimum cluster size of 4. And the scoring parameters for docking were set as Evdw at 0.01 and Eelec at 1.0 and rest. The rest of the parameters were untouched and allowed to be at the pre-defined default setting.

4.4.6. Docking Summary

The protein-protein docking using HADDOCK was successful for all combination of P3 and Pilin proteins and best ten clusters were displayed with their respective HADDOCK score, z-value and other values. For P3 and Pilin protein of *Pseudomonas syringae*, the software clustered 113 structures in 9 clusters, representing the 56% of the water-refined models. And for P3 and Pilin protein of *Pseudomonas syringae* pv. syringae, 184 structures were clustered in 4 clusters which represented 92% of water-refined models HADDOCK generated. Likewise the for P3 protein and Pilin protein *Pseudomonas syringae* pv. savastanoi, HADDOCK clustered 85 structures in 13 clusters and these represented 42% of the water-refined model HADDOCK generated. Whereas for P3 protein docking with Pilin protein of *Pseudomonas syringae* pv. actinidiae, the software clustered 94 structures in 10 clusters, which represented 47% of the water-refined models HADDOCK generated. Similarly for docking with pilin protein of *Pseudomonas syringae* pv. phaseolicola, HADDCOK clustered 120 structures in 10 clusters, which represents 60% of the water-refined models HADDOCK generated. Lastly, for P3 docking with pilin protein of *Pseudomonas aeruginosa*, the software clustered 171 structures in 10 clusters, which represents 85% of the water-refined models.

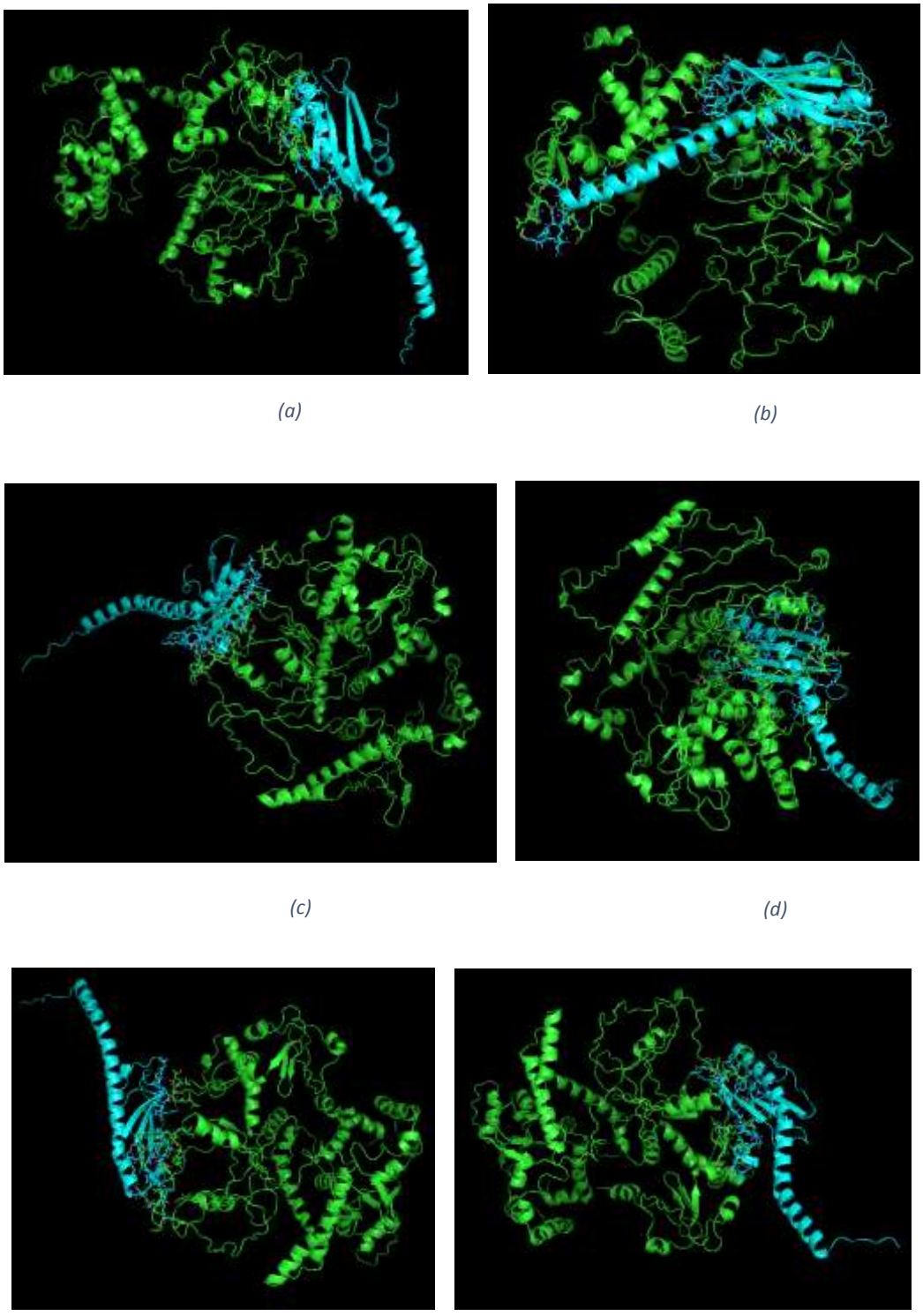
4.4.7. Ranking and Selection

The HADDOCK software ranked the clusters and its structure based on the top value of each parameter such as HADDOCK score, cluster size, z-score, buried surface area, RMSD from the overall lowest-energy structure and other energy score. Since, our intent was to study the binding of two protein molecules, we took z-score, HADDOCK score, and energy values into consideration to select our structure for comparative study. These value for selected clusters and the selected docked protein structures are presented below.

Table 7: HADDOCK docking results for P3 and Pilin

Job name → / **Parameters** ↓	P3-Pilin 0	P3-Pilin 1	P3-Pilin 2	P3-Pilin 3	P3-Pilin 4	P3-Pilin 5
HADDOCK score	-77.3+/-4.1	-60.9 +/- 13.4	-41-1 +/- 19.4	-9.5+/- 13.2	-3.7 +/- 10.7	-112.3 +/-4.8
Z-score	-2.5	-1.1	-2.6	-1.7	-1.6	-2.2
RMSD from overall lowest structure	0.8 +/- 0.5	1.1+/-0.6	1.5+/- 0.9	6.0 +/- 0.4	15.9 +/- 0.1	0.9 +/- 0.6
Van der Waals energy	-67.1 +/- 5.1	60.4+/- 4.6	-71.4 +/- 8.0	-75.8 +/- 8.8	-70.0 +/- 7.4	-62.2 +/- 13.0
Electrostatic energy	-163.6 +/- 31.4	-199.7 +/- 31.8	-393.7+- 93.5	-315.7 +/-28.5	-69.6 +/- 21.7	-362.3 +/-42.1
Desolvation Energy	-5.9 +/-1.6	-0.3+/- 2.4	11.7+/- 2.6	2.9 +/- 1.0	7.9 +/- 2.8	21.9 +/- 4.7
Buried Surface Area	1889.4+/- 42.9	2301.4+/- 104.5	3425.0 +/-327.6	3244.9 +/-64.4	2512.8 +/- 105.3	2379.0 +/- 73.9
Restraints Violation Energy	284.0 +/- 63.4	397.2 +/- 83.3	973.0 +/ 129.8	1265.4 +/- 108.2	1022.0 +/- 88.8	145.5 +/- 58.7

And out of the suggested four structure for each top cluster of each protein-protein, the best structure was chosen for each docking protein structures and visualized on PyMOL.



(a) *(b)*

(c) *(d)*

(e) *(f)*

Figure 11: Protein-protein docked models for Pseudomonas syringae, Pseudomonas syringae pv. syringae, Pseudomonas syringae pv. savastanoi, Pseudomonas syringae pv. actinidiae, Pseudomonas aeruginosa and Pseudomonas syringae pv. phaseolicola respectively using HADDOCK.

4.4.8. Analysis of docking results

The obtained docking results were analysed in reference to the docking results for P3 protein of Phi6 and pilin protein of *Pseudomonas syringae* pv. phaseolicola. The docking

scores for other sets of protein-protein docking were interpreted in relevance to the values obtained for the reference. The HADDOCK scores suggest that P3-Pilin 5 has the most favorable interaction, with a significantly lower docking score of -112.3 and a corresponding Z-score of -2.2, indicating a strong binding affinity compared to the other complexes which is obvious as this is our reference structure. The P3-Pilin 0 and P3-Pilin 1 had higher HADDOCK score closer to reference model while P3-Pilin 0 and P3-Pilin 2 had a z-score above the reference model suggesting a stronger binding affinity as well. Whereas, the HADDOCK score for P3-Pilin 4 model was extremely low even though it had a considerable z-score subjecting it questionable status about its binding affinity.

Similarly, The van der Waals and electrostatic energy components point to significant noncovalent interactions, with P3-Pilin 5 (the reference model) again showing a strong favorable profile. These energies contribute substantially to the binding affinity and specificity of the interactions, key factors in the development of small molecule inhibitors that can disrupt these interactions for therapeutic purposes (Meller & Porollo, 2012). In this study, the Van der Waals energy values for P3-Pilin models 0, 2, 3, and 4 were higher than the reference value, suggesting these interactions to be even more favorable. Coming to the desolvation energy, the negative values across the models indicate that the formation of these complexes is energetically favored upon the removal of water molecules from the binding interface. The highest desolvation energy of 11.7 +/- 2.6 for P3-Pilin 3 model even higher than reference model 5suggest, it is the most favored interactions whereas the interaction of P3-Pilin 4 is least favored with a value of -21.9 +/- 4.7.

Additionally, the low RMSD values for P3-Pilin 0 and P3-Pilin 4 suggest a consistent and precise docking result that closely matches the reference structure. These results imply a reliable model for these complexes that may be relevant for further in silico experiments and validation (M. van Dijk et al., 2013). Furthermore, the higher Buried Surface Area (BSA) for all the docked models suggest that the interactions between the proteins were stronger as these are indicative of the surface area that becomes inaccessible to solvent after docking. The BSA values are indicative of the extent of the interface between P3 and Pilin, with P3-Pilin 2 and P3-Pilin 3 showing the most extensive buried surface areas. This could correlate with a more significant interface formed, potentially leading to stronger

and more stable interactions, which can be an attractive feature for the development of therapeutic agents that target these interfaces (M. van Dijk et al., 2013).

5. SUMMARY

The in-silico studies to analyse host-pathogen interaction is today's necessity, even more so in case of bacteriophages which are difficult to culture and obtain a pure strain. When it comes to treating pathogenic infection in plants, such studies could be even more useful in generating target specific solutions without having to deal with the consequences of and resistance to use of chemical compounds like copper oxides and other copper-rich compounds. The biocontrol method using phages are both target specific and infer none to extremely lower harm to the soil quality as well. In quest to suggest to an effective phage with wide host range to treat infection of crops by *Pseudomonas syringae* and related pathogens, an in-silico molecular analysis was performed in which interaction of the established phage Phi6 with different pathovars of *Pseudomonas syringae*, *Pseudomonas syringae* itself and *Pseudomonas aeruginosa* was analysed. The study was conducted by taking the proteins involved in establishing successful phage-host attachment which were phage membrane protein, Spike protein P3 as the receptor binding protein and the bacterial membrane protein Type IV Pilin as the phage receptor. The 3D structure of these protein-protein complexes were built using AlphaFold Multimer and their binding affinity was predicted using HADOCK Prodigy and all the achieved results were compared with reference to binding affinity of P3-Pilin complex of Phi 6 and *Pseudomonas syringae* pv. phaseolicola. Furthermore, the 3D structures of these proteins were predicted using the AlphaFold Colab tool and the molecular docking of the target protein P3 was done against pilin proteins of different Pseudomonas species and their pathovars for further analysis.

6. CONCLUSION

We have derived the following conclusions from this study:

- The role of P3 protein and pilin protein in successful attachment of pseudomonas phages like Phi 6 to their host *Pseudomonas syringae* depicts them as probable target protein and protein-ligand.
- Phi 6 isolated against the plant pathogen *Pseudomonas syringae* pv. phaseolicola could be effective in dealing other plant pathogenic strains of *Pseudomonas syringae* related pathogens.
- The results obtained for this study suggest that, the Phi 6 could be useful in treating infectious plant diseases caused by its pathovars such as *Pseudomonas syringae* pv. actinidiae, *Pseudomonas syringae* pv. syringae, *Pseudomonas syringae* pv. savastanoi and *Pseudomonas syringae.*
- However, its efficacy against the related species *Pseudomonas aeruginosa* was questionable.

7. RECOMMENDATIONS

Here is a list of recommendations for this study:

- Although, the computational studies gave a preliminary idea on Phi 6 being effective against *Pseudomonas syringae* pathovars, experimental validation of this result and field trial is necessary before actually implementing this biocontrol measure
- The interactions between P3 protein of phi 6 and Pilin protein of different pathovars of *Pseudomonas syringae* to be verified by culture of the organisms, protein isolation and their analysis using NGS sequencing.
- Chromatographic analysis and spectroscopy of the isolated protein and comparison with molecular simulation results could be the way forward to strengthen the candidacy of phi 6 and similar other phages as potential suitor for biocontrol of plant pathogenic Pseudomonas species.
- Mutagenesis studies of the predicted 3D structure to predict possible prevention of phage entry.

REFERENCES

Adamczyk, L., Agakishiev, G., Aggarwal, M. M., Ahammed, Z., Alakhverdyants, A. V., Alekseev, I., Alford, J., Anderson, B. D., Anson, C. D., Arkhipkin, D., Averichev, G. S., Balewski, J., Banerjee, A., Barnovska, Z., Beavis, D. R., Bellwied, R., Betancourt, M. J., Betts, R. R., Bhasin, A., . . . Zoulkarneeva, Y. (2012). Directed flow of identified particles INAU+AUCollisions aTSNN=200 GEV. *Physical Review Letters*, *108*(20). https://doi.org/10.1103/physrevlett.108.202301

Ahlgren, N. A., Ren, J., Lu, Y. Y., Fuhrman, J. A., & Sun, F. (2016). Alignment-free d_2^* oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. *Nucleic Acids Research*, *45*(1), 39–53. https://doi.org/10.1093/nar/gkw1002

Altamirano, F. L. G., & Barr, J. J. (2019). Phage therapy in the postantibiotic era. *Clinical Microbiology Reviews*, *32*(2). https://doi.org/10.1128/cmr.00066-18

Ando, H., Lemire, S., Pires, D. P., & Lu, T. K. (2015). Engineering modular viral scaffolds for targeted bacterial population editing. *Cell Systems*, *1*(3), 187–196. https://doi.org/10.1016/j.cels.2015.08.013

Banerjee, S., Walder, F., Büchi, L., Meyer, M., Held, A. Y., Gattinger, A., Keller, T., Charles, R., & Van Der Heijden, M. G. A. (2019). Agricultural intensification reduces microbial network complexity and the abundance of keystone taxa in roots. *the ISME Journal*, *13*(7), 1722–1736. https://doi.org/10.1038/s41396-019-0383-2

Batinovic, S., Wassef, F., Knowler, S. A., Rice, D. T. F., Stanton, C. R., Rose, J., Tucci, J., Nittami, T., Vinh, A., Drummond, G. R., Sobey, C. G., Chan, H. T., Seviour, R. J., Petrovski, S., & Franks, A. E. (2019). Bacteriophages in natural and artificial environments. *Pathogens*, *8*(3), 100. https://doi.org/10.3390/pathogens8030100

Bender, C. L., Alarcón-Chaidez, F., & Gross, D. C. (1999). Pseudomonas syringaePhytotoxins: Mode of action, regulation, and biosynthesis by peptide and polyketide synthetases. *Microbiology and Molecular Biology Reviews*, *63*(2), 266–292. https://doi.org/10.1128/mmbr.63.2.266-292.1999

Berge, O., Monteil, C. L., Bartoli, C., Chandeysson, C., Guilbaud, C., Sands, D. C., & Morris, C. E. (2014). A User's Guide to a Data Base of the Diversity of Pseudomonas syringae and Its Application to Classifying Strains in This Phylogenetic Complex. *PloS One*, *9*(9), e105547. https://doi.org/10.1371/journal.pone.0105547

Block, A., & Alfano, J. R. (2011). Plant targets for Pseudomonas syringae type III effectors: virulence targets or guarded decoys? *Current Opinion in Microbiology*, *14*(1), 39–46. https://doi.org/10.1016/j.mib.2010.12.011

Bogan, A. A., & Thorn, K. S. (1998). Anatomy of hot spots in protein interfaces. *Journal of Molecular Biology/Journal of Molecular Biology*, *280*(1), 1–9. https://doi.org/10.1006/jmbi.1998.1843

Brüssow, H. (2012a). What is needed for phage therapy to become a reality in Western medicine? *Virology*, *434*(2), 138–142. https://doi.org/10.1016/j.virol.2012.09.015

Brüssow, H. (2012b). What is needed for phage therapy to become a reality in Western medicine? *Virology*, *434*(2), 138–142. https://doi.org/10.1016/j.virol.2012.09.015

Burdman, S., Bahar, O., Parker, J. K., & De La Fuente, L. (2011). Involvement of Type IV pili in pathogenicity of plant pathogenic bacteria. *Genes*, *2*(4), 706–735. https://doi.org/10.3390/genes2040706

Casey, E., Van Sinderen, D., & Mahony, J. (2018). In vitro characteristics of phages to guide 'Real Life' phage therapy suitability. *Viruses*, *10*(4), 163. https://doi.org/10.3390/v10040163

Chaturongakul, S., & Ounjai, P. (2014). Phageâ€“host interplay: examples from tailed phages and Gram-negative bacterial pathogens. *Frontiers in Microbiology*, *5*. https://doi.org/10.3389/fmicb.2014.00442

Chegini, Z., Khoshbayan, A., Moghadam, M. T., Farahani, I., Jazireian, P., & Shariati, A. (2020). Bacteriophage therapy against Pseudomonas aeruginosa biofilms: a review. *Annals of Clinical Microbiology and Antimicrobials*, *19*(1). https://doi.org/10.1186/s12941-020-00389-5

Clarke, C. R., Cai, R., Studholme, D. J., Guttman, D. S., & Vinatzer, B. A. (2010). Pseudomonas syringae Strains Naturally Lacking the Classical P. syringae hrp/hrc Locus Are Common Leaf Colonizers Equipped with an Atypical Type III Secretion System. *Molecular Plant-microbe Interactions*, *23*(2), 198–210. https://doi.org/10.1094/mpmi-23-2-0198

Collmer, A., Badel, J. L., Charkowski, A. O., Deng, W., Fouts, D. E., Ramos, A. R., Rehm, A. H., Anderson, D. M., Schneewind, O., Van Dijk, K., & Alfano, J. R. (2000). Pseudomonas syringae Hrp type III secretion system and effector proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(16), 8770–8777. https://doi.org/10.1073/pnas.97.16.8770

De Jonge, P. A., Nobrega, F. L., Brouns, S. J., & Dutilh, B. E. (2019). Molecular and evolutionary determinants of bacteriophage host range. *Trends in Microbiology*, *27*(1), 51–63. https://doi.org/10.1016/j.tim.2018.08.006

De Smet, J., Hendrix, H., Blasdel, B. G., Danis-Wlodarczyk, K., & Lavigne, R. (2017a). Pseudomonas predators: understanding and exploiting phage–host interactions. *Nature Reviews. Microbiology*, *15*(9), 517–530. https://doi.org/10.1038/nrmicro.2017.61

De Smet, J., Hendrix, H., Blasdel, B. G., Danis-Wlodarczyk, K., & Lavigne, R. (2017b). Pseudomonas predators: understanding and exploiting phage–host interactions. *Nature Reviews. Microbiology*, *15*(9), 517–530. https://doi.org/10.1038/nrmicro.2017.61

Deresinski, S. (2009). Bacteriophage therapy: exploiting smaller fleas. *Clinical Infectious Diseases/Clinical Infectious Diseases (Online. University of Chicago. Press)*, *48*(8), 1096–1101. https://doi.org/10.1086/597405

Edwards, R. A., McNair, K., Faust, K., Raes, J., & Dutilh, B. E. (2015). Computational approaches to predict bacteriophage–host relationships. *FEMS Microbiology Reviews*, *40*(2), 258–272. https://doi.org/10.1093/femsre/fuv048

Frampton, R. A., Taylor, C., Moreno, A. V. H., Visnovsky, S. B., Petty, N. K., Pitman, A. R., & Fineran, P. C. (2014). Identification of Bacteriophages for Biocontrol of the Kiwifruit Canker Phytopathogen Pseudomonas syringae pv. actinidiae. *Applied and Environmental Microbiology*, *80*(7), 2216–2228. https://doi.org/10.1128/aem.00062-14

GaláN, J. E., & Collmer, A. (1999). Type III Secretion Machines: Bacterial Devices for Protein Delivery into Host Cells. *Science*, *284*(5418), 1322–1328. https://doi.org/10.1126/science.284.5418.1322

Garrido-Sanz, D., Meier-Kolthoff, J. P., Göker, M., Martín, M., Rivilla, R., & Redondo-Nieto, M. (2016). Genomic and Genetic Diversity within the Pseudomonas fluorescens Complex. *PloS One*, *11*(2), e0150183. https://doi.org/10.1371/journal.pone.0150183

Glickmann, E., Gardan, L., Jacquet, S., Hussain, S., Elasri, M., Petit, A., & Dessaux, Y. (1998). Auxin Production Is a Common Feature of Most Pathovars of Pseudomonas

syringae. *Molecular Plant-microbe Interactions*, *11*(2), 156–162. https://doi.org/10.1094/mpmi.1998.11.2.156

Gottlieb, P., Metzger, S., Romantschuk, M., Carton, J., Strassman, J., Bamford, D. H., Kalkkinen, N., & Mindich, L. (1988). Nucleotide sequence of the middle dsRNA segment of bacteriophage ϕ6: Placement of the genes of membrane-associated proteins. *Virology*, *163*(1), 183–190. https://doi.org/10.1016/0042-6822(88)90245-0

Heller, K. J. (1992). Molecular interaction between bacteriophage and the gram-negative cell envelope. *Archives of Microbiology*, *158*(4), 235–248. https://doi.org/10.1007/bf00245239

Hermoso, J. A., García, J. L., & García, P. (2007). Taking aim on bacterial pathogens: from phage therapy to enzybiotics. *Current Opinion in Microbiology*, *10*(5), 461–472. https://doi.org/10.1016/j.mib.2007.08.002

Holtappels, D., Fortuna, K., Lavigne, R., & Wagemans, J. (2021). The future of phage biocontrol in integrated plant protection for sustainable crop production. *Current Opinion in Biotechnology*, *68*, 60–71. https://doi.org/10.1016/j.copbio.2020.08.016

Honorato, R. V., Koukos, P. I., Jiménez-García, B., Tsaregorodtsev, A., Verlato, M., Giachetti, A., Rosato, A., & Bonvin, A. M. (2021). Structural biology in the clouds: the WENMR-EOSC ecosystem. *Frontiers in Molecular Biosciences*, *8*. https://doi.org/10.3389/fmolb.2021.729513

Jäälinoja, H. T., Huiskonen, J. T., & Butcher, S. J. (2007). Electron Cryomicroscopy Comparison of the Architectures of the Enveloped Bacteriophages φ6 and φ8. *Structure*, *15*(2), 157–167. https://doi.org/10.1016/j.str.2006.12.004

Jarrell, K. F., & McBride, M. J. (2008). The surprisingly diverse ways that prokaryotes move. *Nature Reviews. Microbiology*, *6*(6), 466–476. https://doi.org/10.1038/nrmicro1900

Jones, J., Jackson, L., Balogh, B., Obradovic, A., Iriarte, F., & Momol, M. (2007). Bacteriophages for plant disease control. *Annual Review of Phytopathology*, *45*(1), 245–262. https://doi.org/10.1146/annurev.phyto.45.062806.094411

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

Kazantseva, O. A., Buzikov, R. M., Pilipchuk, T. A., Valentovich, L. N., Kazantsev, A. N., Kalamiyets, E. I., & Shadrin, A. M. (2021). The Bacteriophage Pf-10—A Component of the Biopesticide "Multiphage" Used to Control Agricultural Crop Diseases Caused by Pseudomonas syringae. *Viruses*, *14*(1), 42. https://doi.org/10.3390/v14010042

Laurinavičius, S., Käkelä, R., Bamford, D. H., & Somerharju, P. (2004). The origin of phospholipids of the enveloped bacteriophage phi6. *Virology*, *326*(1), 182–190. https://doi.org/10.1016/j.virol.2004.05.021

Mavrodi, O. V., Walter, N., Elateek, S., Taylor, C. G., & Okubara, P. A. (2012). Suppression of Rhizoctonia and Pythium root rot of wheat by new strains of Pseudomonas. *Biological Control*, *62*(2), 93–102. https://doi.org/10.1016/j.biocontrol.2012.03.013

Meller, J., & Porollo, A. (2012). Computational methods for prediction of Protein-Protein interaction Sites. In *IntechOpen eBooks*. https://doi.org/10.5772/36716

Melville, S., & Craig, L. (2013). Type IV pili in Gram-Positive bacteria. *Microbiology and Molecular Biology Reviews*, *77*(3), 323–341. https://doi.org/10.1128/mmbr.00063-12

Mindich, L., Qiao, X., Qiao, J., Onodera, S., Romantschuk, M., & Hoogstraten, D. (1999). Isolation of Additional Bacteriophages with Genomes of Segmented Double-Stranded RNA. *Journal of Bacteriology*, *181*(15), 4505–4508. https://doi.org/10.1128/jb.181.15.4505-4508.1999

Morel, J., & Dangl, J. L. (1997). The hypersensitive response and the induction of cell death in plants. *Cell Death and Differentiation*, *4*(8), 671–683. https://doi.org/10.1038/sj.cdd.4400309

Naureen, Z., Dautaj, A., Anpilogov, K., Camilleri, G., Dhuli, K., Tanzi, B., Maltese, P. E., Cristofoli, F., De Antoni, L., Beccari, T., Dundar, M., & Bertelli, M. (2020). Bacteriophages presence in nature and their role in the natural selection of bacterial populations. *PubMed*, *91*(13-S), e2020024. https://doi.org/10.23750/abm.v91i13-s.10819

Nilsson, A. S. (2014). Phage therapy—constraints and possibilities. *Upsala Journal of Medical Sciences*, *119*(2), 192–198. https://doi.org/10.3109/03009734.2014.902878

Nobrega, F. L., Vlot, M., De Jonge, P. A., Dreesens, L. L., Beaumont, H. J. E., Lavigne, R., Dutilh, B. E., & Brouns, S. J. J. (2018). Targeting mechanisms of tailed bacteriophages. *Nature Reviews. Microbiology*, *16*(12), 760–773. https://doi.org/10.1038/s41579-018-0070-8

Nudleman, E., & Kaiser, D. (2004). Pulling Together with Type IV Pili. *Microbial Physiology*, *7*(1–2), 52–62. https://doi.org/10.1159/000077869

Ofir, G., & Sorek, R. (2018). Contemporary phage Biology: From classic models to new insights. *Cell*, *172*(6), 1260–1270. https://doi.org/10.1016/j.cell.2017.10.045

Rombouts, S., Volckaert, A., Venneman, S., Declercq, B., Vandenheuvel, D., Allonsius, C. N., Van Malderghem, C., Jang, H. B., Briers, Y., Noben, J. P., Klumpp, J., Van Vaerenbergh, J., Maes, M., & Lavigne, R. (2016). Characterization of Novel Bacteriophages for Biocontrol of Bacterial Blight in Leek Caused by Pseudomonas syringae pv. porri. *Frontiers in Microbiology*, *7*. https://doi.org/10.3389/fmicb.2016.00279

Rouse, D. I. (1985). A model relating the probability of foliar disease incidence to the population frequencies of bacterial plant pathogens. *Phytopathology*, *75*(5), 505. https://doi.org/10.1094/phyto-75-505

Samson, J. E., Magadán, A. H., Sabri, M., & Moineau, S. (2013). Revenge of the phages: defeating bacterial defences. *Nature Reviews. Microbiology*, *11*(10), 675–687. https://doi.org/10.1038/nrmicro3096

Schmidt, C. (2019). Phage therapy's latest makeover. *Nature Biotechnology*, *37*(6), 581–586. https://doi.org/10.1038/s41587-019-0133-z

Silby, M. W., Winstanley, C., Godfrey, S. A., Levy, S. B., & Jackson, R. W. (2011). Pseudomonasgenomes: diverse and adaptable. *FEMS Microbiology Reviews*, *35*(4), 652–680. https://doi.org/10.1111/j.1574-6976.2011.00269.x

Siringan, P., Connerton, P. L., Cummings, N. J., & Connerton, I. F. (2014). Alternative bacteriophage life cycles: the carrier state ofCampylobacter jejuni. *Open Biology*, *4*(3), 130200. https://doi.org/10.1098/rsob.130200

Spinelli, S., Desmyter, A., Verrips, C. T., De Haard, H. J. W., Moineau, S., & Cambillau, C. (2005a). Lactococcal bacteriophage p2 receptor-binding protein structure suggests a common ancestor gene with bacterial and mammalian viruses. *Nature Structural & Molecular Biology*, *13*(1), 85–89. https://doi.org/10.1038/nsmb1029

Spinelli, S., Desmyter, A., Verrips, C. T., De Haard, H. J. W., Moineau, S., & Cambillau, C. (2005b). Lactococcal bacteriophage p2 receptor-binding protein structure suggests a common ancestor gene with bacterial and mammalian viruses. *Nature Structural & Molecular Biology*, *13*(1), 85–89. https://doi.org/10.1038/nsmb1029

Stockwell, V. O., & Stack, J. P. (2007). Using Pseudomonas spp. for Integrated Biological Control. *Phytopathology*, *97*(2), 244–249. https://doi.org/10.1094/phyto-97-2-0244

Sulakvelidze, A., Alavidze, Z., & Morris, J. G. (2001). Bacteriophage therapy. *Antimicrobial Agents and Chemotherapy*, *45*(3), 649–659. https://doi.org/10.1128/aac.45.3.649-659.2001

Suttle, C. (2005). The viriosphere: the greatest biological diversity on Earth and driver of global processes. *Environmental Microbiology*, *7*(4), 481–482. https://doi.org/10.1111/j.1462-2920.2005.803_11.x

Svircev, A., Roach, D., & Castle, A. (2018). Framing the Future with Bacteriophages in Agriculture. *Viruses*, *10*(5), 218. https://doi.org/10.3390/v10050218

Tu, J., Park, T., Morado, D. R., Hughes, K. T., Molineux, I. J., & Liu, J. (2017a). Dual host specificity of phage SP6 is facilitated by tailspike rotation. *Virology*, *507*, 206–215. https://doi.org/10.1016/j.virol.2017.04.017

Tu, J., Park, T., Morado, D. R., Hughes, K. T., Molineux, I. J., & Liu, J. (2017b). Dual host specificity of phage SP6 is facilitated by tailspike rotation. *Virology*, *507*, 206–215. https://doi.org/10.1016/j.virol.2017.04.017

Vakser, I. A. (2014a). Protein-Protein docking: From interaction to interactome. *Biophysical Journal*, *107*(8), 1785–1793. https://doi.org/10.1016/j.bpj.2014.08.033

Vakser, I. A. (2014b). Protein-Protein docking: From interaction to interactome. *Biophysical Journal*, *107*(8), 1785–1793. https://doi.org/10.1016/j.bpj.2014.08.033

Van Dijk, M., Visscher, K. M., Kastritis, P. L., & Bonvin, A. M. J. J. (2013). Solvated protein–DNA docking using HADDOCK. *Journal of Biomolecular NMR*, *56*(1), 51–63. https://doi.org/10.1007/s10858-013-9734-x

Van Zundert, G., Rodrigues, J., Trellet, M., Schmitz, C., Kastritis, P., Karaca, E., Melquiond, A., Van Dijk, M., De Vries, S., & Bonvin, A. (2016). The HADDOCK2.2 Web Server: User-Friendly Integrative modeling of biomolecular complexes. *Journal of Molecular Biology/Journal of Molecular Biology*, *428*(4), 720–725. https://doi.org/10.1016/j.jmb.2015.09.014

Vangone, A., & Bonvin, A. M. (2015). Contacts-based prediction of binding affinity in protein–protein complexes. *eLife*, *4*. https://doi.org/10.7554/elife.07454

Węgrzyn, G. (2006). *Modern Bacteriophage Biology and Biotechnology 2006*.

Wittebole, X., De Roock, S., & Opal, S. M. (2013). A historical overview of bacteriophage therapy as an alternative to antibiotics for the treatment of bacterial pathogens. *Virulence*, *5*(1), 226–235. https://doi.org/10.4161/viru.25991

Xin, X., Kvitko, B., & He, S. Y. (2018a). Pseudomonas syringae: what it takes to be a pathogen. *Nature Reviews. Microbiology*, *16*(5), 316–328. https://doi.org/10.1038/nrmicro.2018.17

Xin, X., Kvitko, B., & He, S. Y. (2018b). Pseudomonas syringae: what it takes to be a pathogen. *Nature Reviews. Microbiology*, *16*(5), 316–328. https://doi.org/10.1038/nrmicro.2018.17

Xu, Q., & Dunbrack, R. L. (2019). Principles and characteristics of biological assemblies in experimentally determined protein structures. *Current Opinion in Structural Biology*, *55*, 34–49. https://doi.org/10.1016/j.sbi.2019.03.006

Young, F., Rogers, S., & Robertson, D. L. (2020). Predicting host taxonomic information from viral genomes: A comparison of feature representations. *PLOS Computational Biology/PLoS Computational Biology*, *16*(5), e1007894. https://doi.org/10.1371/journal.pcbi.1007894

Zheng, X., Spivey, N. W., Zeng, W., Liu, P., Fu, Z. Q., Klessig, D. F., He, S. Y., & Dong, X. (2012). Coronatine Promotes Pseudomonas syringae Virulence in Plants by Activating a Signaling Cascade that Inhibits Salicylic Acid Accumulation. *Cell Host & Microbe*, *11*(6), 587–596. https://doi.org/10.1016/j.chom.2012.04.014

Appendices

7.1. DNA Sequence and genome annotation databases

EMBL	**http://www.ebi.ac.uk/embl/**	**General nucleotide sequence database**
GenBank	http://www.ncbi.nlm.nih.gov/Genbank/	General nucleotide sequence database

7.2. Protein databases

UniProt	**http://www.ebi.ac.uk/uniprot/**	**Enzyme nomenclature database providing extensive information on all enzymes with an associated EC number**
PSORTdb	http://db.psort.org/	Repository of experimentally determined

7.3. HADDOCK Softwares

HADDOCK Prodigy	**https://rascar.science.uu.nl/prodigy/**	**A webserver/tool to predict binding affinity based on interfacial contact between the protein complexes**
HADDOCK	https://rascar.science.uu.nl/haddock2.4/	A software to perform protein-protein docking

7.4. Experimental Results

7.4.1. Colony characteristics of isolated bacteria on Nutrient Agar

Sample	Media	Color	Shape	Margin	Configuration	Elevation	Opacity
R1C2	Nutrient agar	Creamy white	circular	Entire	Smooth	Convex	Opaque
	Sucrose peptone agar	White mucoid	circular domed shaped	Entire	Smooth	Convex	Opaque
R2C3	Nutrient agar	Creamy white	Circular	Entire	Smooth	Convex	Opaque
	Sucrose peptone agar	White mucoid	Circular domed shaped	Entire	Smooth	Convex	Opaque

7.4.2. Gram staining of isolated bacteria

Sample	Staining method	Stain used	Color of organism	Shape of organism	Result
R1C2	Grams staining	Crystal violet Gram's iodine Decolorizer Safranin	Red	Rod-shaped	Gram negative
R2C3					

7.4.3. Biochemical test of isolated bacteria

Sample	Test	Reagent used	Observation	Result
R1C2	Catalase test	3% H2O2	Effervescence of gas	Catalase positive
	Oxidase test	Tetramethyl-p-phenylenediamine	No development of purple color	Oxidase negative
R2C3	Catalase test	3% H2O2	Effervescence of gas	Catalase positive
	Oxidase test	Tetramethyl-p-phenylenediamine	No development of purple color	Oxidase negative

7.4.4. Levan production test of isolated bacteria

Sample	**Test**	**Media used**	**Colony morphology**	**Result**
R1C2	Levan production test	Sucrose nutrient agar media	White, mucoid and highly domed colony	Levan producing organism
R2C3	Levan production test	Sucrose nutrient agar media	White mucoid and highly domed colony	Levan producing organism

7.4.5 Images of Biochemical Tests

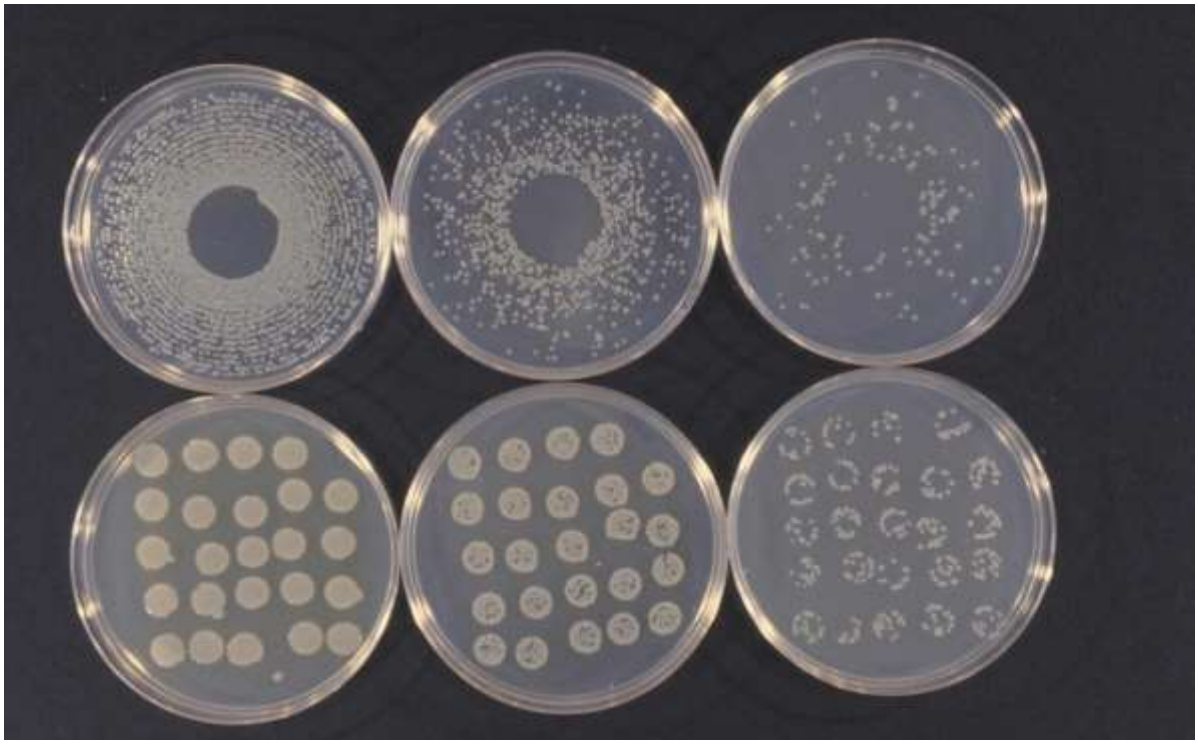


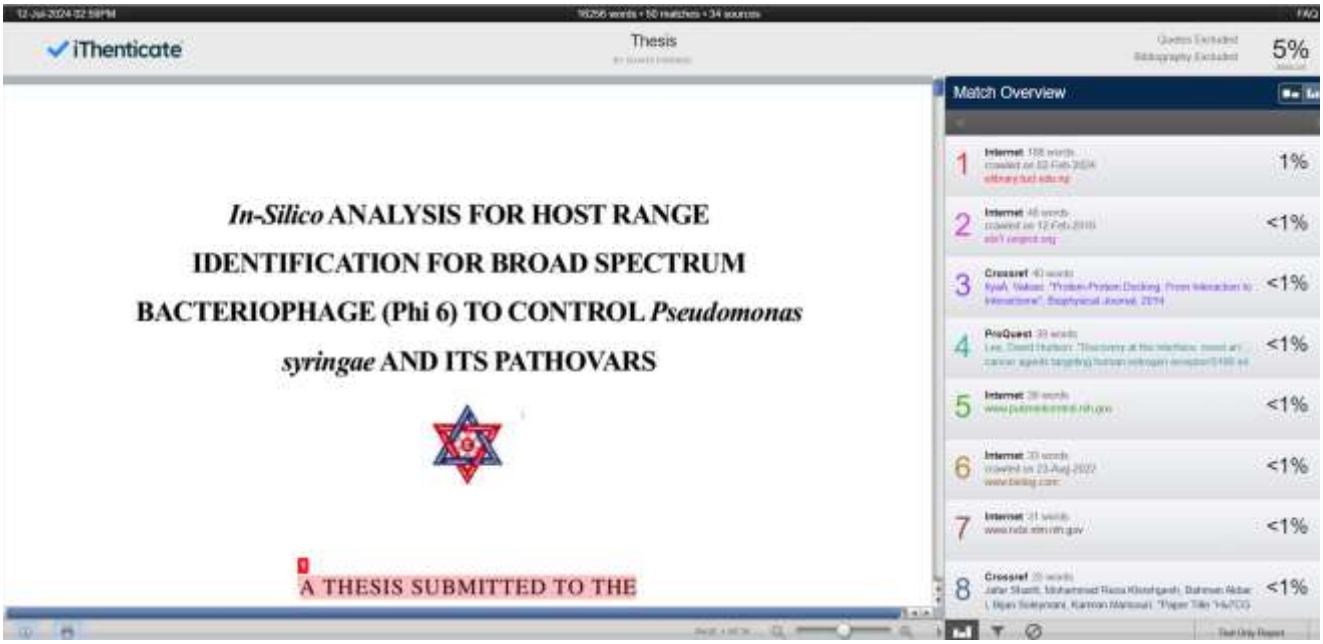
Figure 12: Screeing of isolates on Nutrient Agar

Screening of isolates on Sodium acetate Nitrate medium.

Gram-staining results

Oxidase Test

Catalase Test



Thesis

ORIGINALITY REPORT

5%
SIMILARITY INDEX

PRIMARY SOURCES

1	elibrary.tucl.edu.np Internet	188 words — 1%
2	ebi1.uniprot.org Internet	48 words — < 1%
3	IlyaA. Vakser. "Protein-Protein Docking: From Interaction to Interactome", Biophysical Journal, 2014 Crossref	40 words — < 1%
4	Lee, David Hudson. "Discovery at the interface: novel anti-cancer agents targeting human estrogen receptor/S100 interactions.", Proquest, 2015. ProQuest	38 words — < 1%
5	www.pubmedcentral.nih.gov Internet	38 words — < 1%
6	www.biolog.com Internet	33 words — < 1%
7	www.ncbi.nlm.nih.gov Internet	31 words — < 1%
8	Jafar Sharifi, Mohammad Reza Khirehgesh, Bahman Akbari, Bijan Soleymani, Kamran Mansouri. "Paper Title "Hu7CG2: A Novel Humanized Anti-	25 words — < 1%