The Dissertation Committee for Daniel Joseph Garry Certifies that this is the approved version of the following dissertation:

The Impact of Engineered Design Constraints upon Bacteriophage T7 Evolution

Committee:

Andrew Ellington, Supervisor

James Bull

Jeffrey Barrick

Bryan Davies

Ian Molineux

The Impact of Engineered Design Constraints upon Bacteriophage T7 Evolution

by

Daniel Joseph Garry

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2017

Dedication

To my mother and father

Acknowledgements

This work would not have happened without many people who have driven, educated, challenged, and befriended me along the way. I first and foremost need to thank my parents, whose focus on education was only secondary to a focus on trying to live a good life. I hope that I may continue the example they have set forth not just for me but for Tom and Kate as well. I want to acknowledge Tom and Kate for being pretty cool siblings whom I am proud of. I am grateful to have such a family as mine.

I have had the wonderful opportunity to learn from a variety great educators over the years and so I want to acknowledge a small fraction of them here. I want to thank Drs. William Hardie and Tim LeCras, both of Cincinnati Children's Hospital, for allowing me the opportunity to be exposed to real research and for giving me more education thorough that experience than I gave back in lab work. I'd like to also thank Dr. Trung Phuoc Huynh (Joe!) for your humor, candor, and encouragement. I want to thank the educators on my committee, for their support, comments, and encouragement, in particular Jim Bull for being a de facto advisor. I want to acknowledge my real advisor, Andy, for his support and for creating this unique lab in which I was able to learn to work and think independently as well as learn so much from you and everyone else in the lab. We are truly scientific wildcatters in the land of misfit toys.

Finally I'd like to acknowledge some of those members of the Ellington lab, from whom I have befriended and learned so much. Adam Meyer, in particular, though not a wizard, nonetheless served as a mentor for me, but I am also deeply indebted to Johnny, Jared, Andre, Jimmy, Ross, Shaunak and all the other members of the Ellington lab in addition to a myriad of other buddies within the UT Austin community. Thanks folks!

The Impact of Engineered Design Constraints upon Bacteriophage T7 Evolution

Daniel Joseph Garry, Ph.D.

The University of Texas at Austin, 2017

Supervisor: Andrew Ellington

Since the establishment of molecular biology until the present, a shift has taken place by which engineering of biology has increased in its scope, power, and influence. Synthetic biology is the latest term that advances the overall goal of engineering control of biological systems. Enabled by the recent progress in the ability to read and write nucleic acids, synthetic biology stands out in its desire for standardization of parts, processes, assays, and even organisms. This standardization has led to multiple successful outcomes which justify its utility. However, this standardization via synthetic biology is imbued with an implicit hubris: it may be that certain aspects of biological phenomena are too complex to lend themselves to systemization. Engineering biology, unlike other engineering disciplines, is not obviously governed by a simple series of equations but, rather, united by one universal theme: evolution happens.

Because of this, much work has been done in order to study not just how evolution works and but also into controlling, accelerating, and biasing it in a laboratory setting. The former work can be generally classified as “evolutionary biology” and the latter called “directed evolution”. In this dissertation, work is presented which blurs the

line between evolutionary biology and directed evolution yielding unanticipated outcomes and new methods of biological control.

Three projects are presented in which evolution of bacteriophage T7 is studied in response to rationally engineered design constraints. In the first, the predictability of bacteriophage T7 promoter evolution in response to a novel RNA polymerase is determined. In the second, the evolution of the entirety of the transcriptional apparatus of bacteriophage T7 is studied in the context of an evolutionarily adapted bacteriophage population. Finally, the construction, exploration, and optimization of a designed bacteriophage T7 life-cycling system reliant on proteases are described. Together, this work provides evidence that although the ability to rationally design biology is important, various ways to rationally control and direct evolution offer a complementary strategy to the systematization of synthetic biology.

Table of Contents

List of Tables

List of Figures

INTRODUCTION

Discovery, Classification, and Diversity of Bacteriophages

Every good story has to begin somewhere. For bacteriophages, broadly defined as viruses that infect bacteria, their origin is hypothesized to occur soon after the advent of the first prokaryotic cells on Earth over three billion years ago (Hendrix, et al. 1999; Pedulla, et al. 2003). The similarities between structural and replication proteins among viruses that infect the three different domains of life suggest that viruses may be older than the Last Universal Common Ancestor (LUCA) (Forterre 2006). Beyond just existing (and likely infecting) the LUCA, it has been suggested that viruses drove the differentiation among the three domains of life (one hypothesized mechanism suggests DNA viruses fused with RNA cells to create DNA cells from a RNA-celled LUCA) (Forterre 2006). Of the modern-day viruses that infect the three domains of life, bacteriophages are by far the most abundant group. It is estimated that there are over 1031 bacteriophages on Earth, with 1025 infections estimated to occur every second, making bacteriophages the most abundant biological entities on Earth (Pedulla, et al. 2003).

Bacteriophages were first discovered in 1915, when British pathologist Fredrick Twort described a filter-passing virus which could turn a micrococci streak "glassy" and was also able to infect *Staphylococcus aureus* without action on *E. coli* cells (Twort 2011). Twort did not follow up with his discovery, informing readers at the end of his Lancet article with the line "I regret that my financial considerations have prevented my carrying these researches to a definite conclusion, but I have indicated the lines along which others more fortunately situated can proceed" (Twort 2011).

Two years later and independently of Twort, French-Canadian microbiologist Felix d'Herrelle isolated invisible microbes which lysed Shigella cultures from the stools (and in one case, urine) of patients recovering from dysentery (D'Herelle 2007). Interestingly, d'Harrelle was able to isolate these invisible microbes from only patients in the midst of recovery from dysentery rather than healthy individuals or those with dysentery who had not recovered (D'Herelle 2007). This led d'Harrelle to conclude that "an antagonistic microorganism" was acting against the Shigella infection (D'Herelle 2007). These "antagonistic micoorganisms" he named "bacteriophages", which literally means "bacteria eater". Unlike Twort, d'Herrelle was financially able to continue work studying these new and mysterious microbes. D'Herelle, even in that first 1915 paper, realized that bacteriophages could be used therapeutically. He proceeded to carry out a rabbit study in which these bacteriophages were successfully protective against Shigella and later used bacteriophages to treat humans, a treatment which in general has become known as bacteriophage therapy (Salmond and Fineran 2015).

This bacteriophage therapy was similar in effect but mechanistically different than the anti-bacterial therapeutics first pioneered by German immunologist Paul Ehrlich (Bosch and Rosich 2008). Erchlich's creation, Salvarsan, was not just the first antimicrobial chemotherapeutic agent but also was so successful and widely acclaimed that Warner Brothers produced a 1940 movie titled Dr. Ehrlich's Magic Bullet in exaltation of his work (Bosch and Rosich 2008). D'Herelle's ultimate success with bacteriophage therapy was much more limited than that of Ehrlich's creation of antimicrobial chemotherapeutic drugs. After d'Herelle's early successes, such as treating patients with phage against plague, various pharmaceutical companies began selling phage preparations during the 1920s and 1930s to target a number of bacterial infections (Salmond and Fineran 2015).

Bacteriophage therapy spread with mixed results and was subsequently called into question, eventually falling under investigation by the American Medical Association in 1934 (Summers 2001). The AMA concluded that the success of phage therapy was ambiguous and expressed concerns regarding the lack of criteria for purity of the phage mixture, the lack of standardization of phage preparations, and the lack of general understanding of how bacteriophage therapy worked (Summers 2001). Indeed, critics of D'Herelle questioned whether his invisible microbes were viruses in the first place; rather, suggesting that they were enzymes (Keen 2012; Summers 2001). These doubts, combined with the outbreak of World War II (which D'Herelle spent under house-arrest in Vichy, France) and the widespread adoption of antibiotics, doomed the further adoption and spread of bacteriophage therapy in the United States (Summers 2001). Bacteriophage therapy would, however, continue to be used by the Soviet Union and has been continuously studied in Eastern Europe up to the present day (Housby and Mann 2009; Summers 2001).

It was only later, in 1940, when Helmut Ruska, together with his brother Ernst Ruska, imaged a bacteriophage using their recently created technique of electron microscopy, that the debate concerning the viral nature of the bacteriophage phenomenon was settled (Kruger, et al. 2000; Salmond and Fineran 2015). In 1967, the International Committee for Taxonomy of Viruses (the ICTV) was founded in order to use information such as phage shape gained from techniques such as electron microscopy to classify bacteriophages (Adams, et al. 2017). In addition to morphology (enveloped/non-enveloped, tail-shape) the ICTV has used nucleic acid type (ssDNA, dsDNA, ssRNA, or dsRNA) as well as genome organization (circular, linear, or segmented) to classify bacteriophages into 10 families within one Order (Caudovirales) (Ackermann 2003; Krupovic, et al. 2011). Bacteriophage genomes range in size from the 3.5kB ssRNA

genome of the levivirus GA to the 497.5kB dsDNA genome of myovirus G (Krupovic, et al. 2011). Over 5,000 individual phages have been isolated yet only 750 have been sequenced as of 2011 (Hatfull and Hendrix 2011). 70% of these sequenced genomes infect only 12 different bacterial hosts (Hatfull and Hendrix 2011). This dearth of information is especially striking in consideration of the estimate that there are over 100 million phage species worldwide encoding for an estimated 2.5 billion ORFs (Rohwer 2003).

How Bacteriophages helped define Molecular Biology

Little did the classical biologists, biochemists, immunologists, and physicists of the 1930s know that advances in bacteriophage research in the next two decades would help spawn the creation of a new paradigm of inquiry now recognized as molecular biology. In 1935, while the burgeoning prospect of bacteriophage therapy began fading in the West, the utility of bacteriophages was just beginning. In that year, German theoretical physicist Max Delbrück, together with Soviet biologist Timofeev-Ressovsky and German biophysicist K.G. Zimmer, would publish a paper that was a result of their working group on how physics can be applied to biology in order to understand the nature of the gene (Summers 1993). Their conclusion, based in part on the work by the University of Texas professor H. J. Muller that ionizing radiation caused mutations in flies, was that genes had to have "a kind of stability similar to that of molecules" (Delbruck 1970). While this notion of the physical nature to the structure and function of the gene seems trivial by modern standards, at the time it and most other elements of molecular biology were thoroughly unexplained.

Based on this work Delbrück was invited to Caltech in 1937 on a fellowship (Delbruck 1970). Soon after arriving there he met post-doctoral research fellow Emory

Ellis, who introduced Delbrück to bacteriophages (Summers 1993). Ellis, funded through the generosity of wealthy mining industrialist Seeley Wintersmith Mudd to study viral carcingogenesis, decided that a more fundamental understanding of viral infection was first necessary (Summers 1993). The relative ease of manipulation and speed of life cycle attracted Ellis to bacteriophages, even though the viral nature of bacteriophage was an open debate and, at the time, was an idea held by a minority of researchers (Summers 1993). In Ellis and bacteriophages, Delbrück, who was educated as a theoretical physicist under Niels Bohr and had developed a deep interest in solving "the riddle of life", saw an opportunity to develop a "deeper understanding of this process through a quantitative experimental approach" (Delbruck 1970). To Delbrück, phages were the easiest system by which he would be able to interrogate the nature of the gene and its function in life. Ellis and Delbrück published their first and only paper together in 1939 describing one-step growth experiments using an *E. coli* infecting bacteriophage (Ellis and Delbruck 1939). In that paper, they developed key ideas that would later become standardized throughout phage biology such as the plaque assay, absorption, latent period, and viral burst size (Ellis and Delbruck 1939; Salmond and Fineran 2015).

While funding constraints moved Ellis back to cancer research, Delbrück continued on studying bacteriophages (Summers 1993). Delbrück would publish two papers independently on the role of various growth conditions on phage absorption and on the nature of cell lysis (Delbruck 1940a, b). In 1940, at the end of his fellowship at Caltech, Delbrück would move to Vanderbilt University where he would teach in the physics department until 1947 (Cairns 2007). It was there that Delbrück and Italian microbiologist Salvador Luria would perform the experiments studying the nature of bacteriophage resistance that they would later be awarded a Nobel Prize (Luria and Delbruck 1943; Salmond and Fineran 2015).

Delbrück and Luria had noticed that there would be secondary growth hours after initial phage lysis of a liquid bacterial culture. These bacterial cells that made up the secondary growth step were resistant specifically to the phage used to cause the initial lysis. Delbrück and Luria realized that these phage-resistant bacterial cells could have become resistant in two different and mutually exclusive ways: that the cells were already resistant to the phage before exposure (the mutation hypothesis) or that the cells acquired immunity to the phage after exposure (the acquired immunity hypothesis). They developed a theoretical and mathematical framework from which they were able to develop experimental tests using T1 bacteriophage and *E. coli* B cells to determine which of the two hypotheses was correct (Luria and Delbruck 1943). Their experiments showed that when replicate bacterial cultures, all started from the same original culture, are independently exposed to phage, the resultant counts of resistant cells had a high variance. This confirmed the mutational hypothesis because the acquired hypothesis presupposes a less varied number of resistant cell counts. Their experiment was the first to show that mutations arise independently of selection pressure and, as Luria later put it "attracted attention to the remarkable possibilities of bacterial genetics" (Cairns 2007).

The pioneering work by Delbrück and Luria had indeed drawn the attention of many others. In 1945, Delbrück and Luria, together with Alfred Hershey, set up the first of twenty-six annual summer Phage courses at the Cold Springs Harbor Laboratory (Cairns 2007). This class was the impetus for the creation of the so-called "Phage Group" which consisted of those interested in bacterial and phage genetics. In addition to providing the environment to facilitate the exchange of results and ideas the Phage courses also allowed Delbrück by edict to codify which phages, host strains, and techniques were supposed to be used by those in the group (Cairns 2007). This standardization led to the "T" nomenclature ("T" for "type") used now for T1 through T7

phages based on the first electron micrographs of *E. coli* infecting phages published in 1943 (Cairns 2007; Luria, et al. 1943).

The 1946 Symposium opened up to much fanfare regarding Edward Tatum and George Beadle's one-gene-one-enzyme hypothesis. In their seminal 1941 paper, they noticed that X-ray irradiated Neurospora crassa (common bread mold) would commonly lose the ability to grow in the absence of a single metabolite. They were able to isolate different strains, each of which would not grow in the absence of their respective metabolite while still being able to grow on media lacking another metabolite (Beadle and Tatum 1941). This data led them to conclude that one gene produced one enzyme responsible for a single step in a metabolic pathway. This and other auxotrophic data presented in support of the one-gene-one-enzyme hypothesis was challenged by Delbrück, who argued that no current series of experiments at the time could be devised to disprove the hypothesis (Cairns 2007). As a result, Norman Horowitz realized he could use temperature-sensitive mutations to lend support to the one-gene-one-protein hypothesis and thus the meeting spurred the development the powerful innovation of the use of temperature-sensitive mutants in bacterial genetics (Cairns 2007). For their discovery, Beadle and Tatum would be awarded half of the 1958 Nobel Prize in Physiology or Medicine (Norrby 2008).

It was at the 1946 Cold Springs Harbor Phage Symposium that Joshua Lederberg and Edward Tatum announced their discovery of genetic recombination via sexual reproduction in *E. coli*, which later would be shown to be because of bacterial conjugation (Lederberg and Tatum 1946; Tatum and Lederberg 1947). In their experiment, they noticed that two different triple auxotrophic *E. coli* cell lines (the first a threonine, leucine, and thiamine auxotroph and the second a biotin, phenylalanine, and cysteine auxotroph) which would not grow if plated individually would indeed grow after

being in physical contact for a time with one another and then subsequently grown in a culture without all six nutrients (thus becoming prototrophic) (Tatum and Lederberg 1947). This work confirmed bacterial recombination in the form of conjugation and would be the basis for Lederberg and Tatum sharing half of the 1958 Nobel Prize (Norrby 2008). Recombination in bacteriophages had been shown earlier that year when Albert Hershey had used T2 phages with two independent mutations (one for clear/cloudy plaques and the other host-range dependence) to show that multiple infections in the same cell could occur with the result that phages also could experience recombination (Hershey 1946). Not to be outdone, Delbrück in the same year had independently discovered phage recombination (Cairns 2007; Delbrück and Bailey 1946).

Incredibly, this work was done before it became widely accepted that DNA was the genetic material of living things. In 1944, biochemists (and decidedly non-Phage Group members) Oswald Avery, Colin MacLeod, and Maclyn McCarty published their data continuing their work on a system where a live but non-pathogenic R strain of pneumococcus, when injected into mice with a large amount of heat-killed pathogenic S strain of pneumococcus, would lead to lethality to the mice in which blood was generated live and pathogenic R strain of pneumococcus (presumably through transformation of the R strain with a factor that made them pathogenic from the S strain) (Avery, et al. 1944). With this exciting model of cellular transformation, in the 1944 paper, the Rockefeller Institute trio would take these results one step further and show that only when the live but non-pathogenic strain of R strain pneumococcus was transformed *in-vitro* with the DNA component of the heat-killed pathogenic S strain pneumococcus (versus the RNA or protein component) did the subsequent cells become virulent and lethal to mice (Avery, et al. 1944).

While this paper "strongly suggest[ed] that nucleic acids, at least those of the desoxyribose [sic] type, possess different specificities as evidenced by the selective action of the transforming principle" the conclusive evidence that genes were made of DNA was not found until 1952. In that year Albert Hershey and his lab assistant Martha Chase would publish their results in which they separately used S35 and P32 isotopes to radiolabel the protein and the nucleic acid, respectively, of T2 phage (Hershey and Chase 1952). These either protein or nucleic acid-tagged T2 phages were then used to infect unlabeled *E. coli* cells. After initial phage absorption, they blended off the T2 phage capsids from the now infected *E. coli* cells which were collected via centrifugation. These *E. coli* cells were then lysed which led to detection of P32 but not S35 in the slurry. These data showed that the T2 DNA was initially protected by a protein capsid which then entered into the *E. coli* cells. Hershey and Chase correctly concluded that this means that genes were made out of DNA, not protein. Hershey (but not his lab assistant Martha Chase) would share the 1969 Nobel Prize with Luria and Delbrück for "their discoveries concerning the replication mechanism and the genetic structure of viruses".

In April 1953 James Watson and Francis Crick would publish the structure of DNA (Watson and Crick 1953). For their work, heavily influenced by the crystal structures generated by Rosalind Franklin and Maurice Wilkins at the Cavendish Laboratory, they would later win the 1962 Nobel Prize, notably with Wilkins and without Franklin who had passed away (Maddox 2003). While Watson is known by many for the discovery of the structure of DNA he is less known for his initial work within the Phage Group. Watson went to Indiana University for his graduate work to study under Salvador Luria (Cairns 2007). While there, he worked with Renato Dulbecco on the properties of T2 bacteriophage inactivated with X-rays (Watson 1950). It was at the urging of Delbrück and Luria that Watson would go to work on the structure of DNA at the

Cavendish Laboratory at University of Cambridge (Cairns 2007). While Watson's subsequent work was decidedly non-phage related it is worth noting that in 2004 as the Director of Cold Springs Harbor Laboratory he co-authored a Nature Biotechnology correspondence in opposition of a News and Views piece in which he defended the future prospects of phage therapy (Schoolnik, et al. 2004).

Phage Group member Gunter Stent would later observe that "unlike the discovery of the alpha-helix [by Linus Pauling et al. in 1951], the discovery of the DNA double helix opened up enormous vistas to the imagination. It provided highroad to understanding how the genetic material functions." (Cairns 2007; Pauling, et al. 1951). Even in their seminal Nature paper Watson and Crick speculated that "[i]t has no escaped our notice that specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material" (Watson and Crick 1953). In 1958 Matthew Meselson and Franklin Stahl would publish their results on DNA replication in *E. coli* using N14 and N15 containing media to show the semi-conservative nature of DNA (Meselson and Stahl 1958). In a five year span from 1961-1966 the work of many researchers, initially using cell-free protein synthesis using randomly ordered RNA and later deciphering the RNA codons via determining ribosomal-bound aminoacyl-tRNAs, quickly established the nature of the genetic code, transcription, and translation (Nirenberg 2004).

The rapid development of molecular biology, only a fraction of which is actually cited here, led from the discovery of the structure of the double helix in 1953 to Crick's 1970 Nature paper clarifying and codifying the known mechanisms of information flow through the various biological components now known as the "Central Dogma of Molecular Biology" (Crick 1970).

Interestingly, these later discoveries that fleshed out the structural basis of the Central Dogma were admittedly not made by those in the Phage Group. Rather, they were made by those that would call themselves physicists or biochemists. Those who have written about the Phage Group note that Delbrück had a dislike for biochemists, stated in a 1949 lecture that, a physicist:

> Listening to the story of modern biochemistry he might become persuaded that the cell is a sack full of enzymes acting on the substrates converting them through various intermediate stages either into cell substance or into waste products…The enzymes must be situated in their proper strategic positions to perform their duties in a well-regulated fashion. They in turn must be synthesized and must be brought into position by maneuvers which are yet understood, but which, at first sight at least, do not necessarily seem to differ in nature from the rest of biochemistry…And yet this program of explaining the simple through the complex smacks suspiciously of the program of explaining atoms in terms of complex mechanical models. It looks sane until paradoxes crop up and come into sharper focus, and this will not happen until the behavior of living cells has been carried into far greater detail. This analysis should be done on the cell's own terms and theories should be formulated with fear of contradicting molecular physics. I believe that it is in this direction that physicists will show the greatest zeal and will create a new intellectual approach to biology which would lend meaning to the ill-used term biophysics. (Cairns 2007)

Delbrück certainly helped pioneer this "new intellectual approach to biology" which we now call molecular biology, and certainly used bacteriophages as his chariot. Ultimately though, the aforementioned transition occurred between those fundamental discoveries from the Phage Group to those by the (Delbrück-maligned) biochemists after the discovery of the structure of DNA. This can be explained in almost philosophical terms as a differing of mindset between the information-driven Phage Group and the structurally-driven biochemists (Cairns 2007). Delbrück's initial interest in biology stems from some hope articulated by Bohr that the complexity of life could be born out through the discovery of some sort of rules, or "complementary principles", of physics that were yet theretofore unrecognized (Cairns 2007).

Of this so-called "Copenhagen Spirit" which developed out of Bohr's thoughts about physics and motivated Delbrück and others, Stent writes, "it was the Copanhagen Spirit that provided the philosophical infrastructure for navigating latter-day biological thought between the Scylla of crude biochemical reductionism, inspired by 19th century physics, and the Charybdis of obscurantist vitalism, inspired by 19th century romanticism" (Cairns 2007). In many ways this paradox vis-à-vis reductionism and holism which was so motivating during that time period is reflected in the present as the creative dichotomy between pure physical reductionism and the study of biological consciousness (or even the most recent effort to interpret emergent biological phenomena using systems biology).

Discovery and Uses of Bacteriophage T7

In 1970, with the establishment of the Central Dogma, the field of molecular biology had been firmed defined and began to be used as a basis of inquiry into the manifold phenomena of life. This inquiry included continued work into learning about bacteriophages. Though far from being the central driver into advancing the field as it had before, work on bacteriophages was far from over with even as those in the Phage Group hung up their mouth pipettes. The benefits of studying bacteriophage, namely, their relative speed, ease, and low-cost, still held true. The early standardization of which phages to study among the Phage Group by Delbrück resulted in its own frozen accident by which the T-phages became the preferred and even dominant bacteriophages as subjects of research. Because the subject matter of this thesis is the bacteriophage T7, it is worthwhile to describe a bit about its origins and utility.

There are a few pieces of evidence that relate to the discovery of bacteriophage T7. According to molecular biologist and historian W.C. Summers (via personal

communication with Max Delbrück), T7 was discovered when Delbrück isolated only one kind of phage from a commercial laboratory that he was visiting that was advertising a polyvalent phage mixture for bacteriophage therapy (Summers 2001). T7 first occurs in the literature in a 1944 paper by Demerec and Fano in which they say that T7 was "isolated from the standard anti-coli-phage mixture prepared by Dr. W. J. MacNeal" who they would later in their acknowledgements mention was "of the New York Post-Graduate Medical School and Hospital" (Demerec and Fano 1945). Whether this anti-coli-phage mixture was actually created by Dr. MacNeal or originated from a commercial mixture used by him created by a company (maybe even the same one Delbrück visited) has been lost to time.

Regardless of its origin, bacteriophage T7 became a useful tool for studying DNA replication and RNA transcription (Calendar 2006). Incredibly, eight of the 55 proteins coded for in the T7 genome have been well-characterized enough and have enough interesting functionality to currently be available commercially: its DNA polymerase, RNA polymerase, lysozyme, exonuclease, endonuclease I, Ocr protein, DNA ligase, and helicase. The T7 DNA polymerase, together with the *E. coli* thioredoxin protein, has DNA polymerase activity with high fidelity at a rapid rate and with 3' to 5' exonuclease activity and is sold as purified protein by New England Biolabs for second strand synthesis in site-directed mutagenesis protocols (Biolabs 2017c). The T7 RNA polymerase is a DNA-dependent RNA polymerase with high *in-vitro* activity and specificity for its own T7 promoter and is sold as purified protein by New England Biolabs for RNA synthesis (Biolabs 2017f). The T7 lysozyme has the ability to inhibit the RNA polymerase and is sold on an inducible plasmid by New England Biolabs as part of *E. coli* protein expression systems (Biolabs 2017a). The T7 exonuclease, working in the 5' to 3' direction, removes mononucleotides from both phosphyrolated and

unphosphorylated duplex DNA and is sold as a purified protein by New England Biolabs for molecular cloning purposes (Biolabs 2017e). The T7 endonuclease I recognizes non-perfectly matched DNA and Holliday junctions and is sold as a purified protein by New England Biolabs for DNA repair and genome editing detection (Biolabs 2017d). The T7 Ocr protein mimics the structure of B-DNA such that it binds to and inhibits type I restriction-modification systems and is sold as a purified protein by Lucigen as an additive to increase transposon efficiency in bacterial cells (Lucigen 2017). The T7 DNA ligase catalyzes the formation of 5' to 3' phosphodiester bonds without acting on blunt end ends and is sold as a purified protein for molecular cloning purposes (Biolabs 2017b). Finally, the T7 helicase has highly processive 5' to 3' helicase activity together with primase activity and is sold commercially as a purified protein by Biohelix for isothermal DNA amplification techniques (Biohelix 2017).

Far from being plundered to model DNA replication and RNA polymerization (or other pieces for later commercialization and utility), bacteriophage T7 continued to be the subject of investigation into its life-cycle and evolution. The genetic mapping of T7 was first published in part in 1969 and in full in 1972 (the complete sequence was first published in 1983) (Calendar 2006; Dunn and Studier 1983; Studier 1972, 1969). With a genetic map (and, later complete genome) in hand at such a relatively early time in the progression of molecular biology, bacteriophage T7 stood out as a model system for experimental evolution. While not attempting to be exhaustive, it is worth mentioning a few of the bacteriophage T7-based experimental evolution systems to illustrate its versatility.

It was noticed bacteriophage T7 could tolerate an accelerated mutation rate through the use of growing and lysing *E. coli* cells grown in the highly mutagenic chemical N-methyl-N'-nitro-N-nitrosoguianidine (NNNG). Indeed, later, in an effort to

use NNNG to reach lethal mutagenesis in bacteriophage T7 populations, the phage tolerated an astounding 245 mutations per genome over 200 generations yet failed to become extinct (Springman, et al. 2010). Earlier, however, an experiment was designed in which NNNG would induce mutations subsequently detected with various restriction enzyme digests (Bull 1993). This increased mutation rate enabled the creation of the first known experimentally-derived phylogeny from which the previous a prori-based phylogenetic algorithms could be checked, verified, and compared to one another, with one method in particular found to be greater than 98% accurate (Hillis, et al. 1992). This work, for the first time, directly supported the phylogenetic methods which were and still are being used to estimate branching points, branch lengths, and ancestral genotypes (Hillis, et al. 1992). The same protocol was extended to show the effect of parallel molecular evolution of deletions and nonsense mutations with regards to phylogenetic analyses (Cunningham, et al. 1997).

Through various complementation experiments between bacteriophages and plasmids it was discovered that that bacteriophage T7 could recombine with plasmids, and in so doing provided an important tool by which to study gene variants and gene placement in the context of the bacteriophage (Campbell, et al. 1978). This technique allowed a bacteriophage T7 mutant to be created without its own RNA polymerase which was subsequently passaged on the related bacteriophage T3 RNA polymerase in order to study the regulatory evolution and effect on overall fitness of T7 (Bull, et al. 2007). In this study, the 9 of 16 promoters picked up the single mutation which conferred high activity on the T3 RNAP. Regulatory evolution using the T7 RNA polymerase was interrogated in a different way by recombining it back in at various locations throughout the bacteriophage T7 genome in order to investigate gene order (Springman, et al. 2005). Two of these lines picked up enabling mutations in the T7 terminator while one resulted

in a recombination of the T7 RNA polymerase back closer it's wild-type location early in the genome. Other bacteriophage T7 adaption experiments have been done on deletions of a key DNA metabolism gene (its DNA ligase) as well as in response to a T7 promoter-driven, plasmid-based, suicide circuit (Bull and Molineux 2008).

Other bacteriophage T7 experimental evolution experiments have continued on in this tradition by attempting to rationally design or expand the genetic code. One group, inspired by rational design and engineering principles, refactored the leftmost 11,515 base pairs of the 40kB genome in order to remove overlapping genes, remove unwanted restriction sites, and add desirous restriction sites (Chan, et al. 2005). Subsequent passaging of the refactored bacteriophage T7 kept 60-70% of the added design elements (Springman, et al. 2012). Another example of rational design, in this case, for the explicit purpose of viral attenuation, was the codon-modification of the gene 10, which codes for the major capsid protein (Bull, et al. 2012). Four versions of the gene 10, all coding for the same amino acid content of the wild-type major capsid protein, were constructed with decreasing fraction of preferred codons (wild-type: 68% preferred codons, variants: 50%, 30%, 20%, and 10% preferred codons). These were then recombined into T7Δ10 bacteriophage (phage lacking the gene 10) and passaged. Overall fitness initially decreased linearly with decreased preferred codon content. Interestingly, the 10% preferred codon variant did regain fitness, however, at a very slow rate, suggesting that codon de-optimization, at least in this proof-of-concept bacteriophage T7 model, was a viable strategy for short-term viral attenuation.

Finally, one of the most recent examples of bacteriophage T7 experimental evolution is the discovery that it will use an expanded genetic code in order to proceed toward evolutionarily higher fitness (Hammerling, et al. 2014). Bacteriophage T7 was passed in an *E. coli* strain which had an orthogonal tRNA and aminoacyl-tRNA-

synthetase pair to provide 3-iodotyrosine at amber codons. After 50 transfers, one of the phage populations was sequenced and it was discovered that 55% of that mutator strain population (which used an error-prone T7 DNA polymerase to stimulate mutations) had an amber mutation in the T7 holin gene. They found that the amber-containing phages were able to out-compete spontaneous rescue mutants at the same position in co-culture experiments in the same experimental conditions. This is one of a multitude of examples of how bacteriophage T7, far from being fully understood and described, has aided and will continue to aid in investigating intriguing evolutionary questions.

It is also worth mentioning here that bacteriophage T7 has been used for phage display. Phage display is a selection technique by which phage phenotype, the display of a protein or peptide on the phage capsid surface, is linked to its genotype, which is encapsulated in the phage. The phages, containing the DNA coding for the peptide or protein variants on their surface, are then subjected to *in-vitro* binding or "panning" whereas a particular target of choice (a small-molecule, nucleic acid, peptide, or protein) is tethered to a surface on which the phages are added and then washed away. Those variants that successfully bind are isolated and enriched while those that do not bind are washed away and lost. Those high-affinity variants are then assayed for desired activity. While there display systems based on many phages, including, filamentous/M13 phages, T4 phage, and lambda phage, the advantages of the T7 phage display system are: 1) the systems high tolerance for an extreme chemical environment, 2) the capsid is not involved with phage absorption thus the peptide/protein of interest should not alter infection and 3) it is not necessary to secrete the displayed peptides/proteins through the periplasm and cell membrane (which ultimately does limit the inclusion of post-translational modifications) (Bazan, et al. 2012). Together, phage display is a highly versatile and useful technology to discovering as well as increasing the affinity of

peptide/protein-small-molecule, peptide/protein-peptide/protein, and peptide/protein-nucleic-acid interactions of which bacteriophage T7 is one tool in the toolbox.

Transition from Understanding to Engineering of Biology

From Crick's clear articulation of the Central Dogma in 1970 until the current day the pendulum of research has swung from efforts to purely understand biology (what could be called "basic research") to those efforts to engineer biology (what could be called "applied research"). The panoply of nucleic acids, proteins, and biologically relevant chemistries among the diversity of life on Earth has provided a rich starting point for the development of applied tools and techniques. These tools and techniques developed from the applied research standpoint can be and often are used cyclically to further enable the exploration and discovery of additional basic research. Of course, bacteriophages, directly or indirectly, have had a role in the shaping of these tools and techniques. A few of the key discoveries and bacteriophage tie-ins are noted here.

Sequencing

The ability to actually discern the exact order of nucleotides in DNA or RNA as well as amino acids in proteins, collectively known as sequencing, was one of the fundamental advancements in applied research which helped bring about the era of engineering biology. Though this sequencing, or "reading" of the DNA, RNA, and proteins, was initially promoted by the establishment of the genetic code the ability, to do so readily (or, more so, inexpensively) took time and technological development. The first nucleic acid ever sequenced was the *S. cerevisiae* alanine tRNA in 1965 (Heather and Chain 2016). Closely following this development was the sequence of the first protein-coding sequence and genome, that of the bacteriophage MS2 coat protein and later its full 3,959 nucleotide ssRNA genome, in 1972 and 1976, respectively (Fiers, et al.

1976; Heather and Chain 2016; Salmond and Fineran 2015). While these early successes in RNA sequencing were based upon using radiolabeled partial-digestion and 2D fractionation, it was not until Fred Sanger developed chain-termination sequencing that the first DNA genomes were sequenced (Heather and Chain 2016). This chain-termination sequencing used radiolabeled ddNTPs incorporated into the 3' ends of DNA pieces by DNA polymerases. Four of these incorporation reactions (one for each ddNTP) would then be resolved in each of four lanes using polyacrylamide electrophoresis. Using this technique, Sanger was able to sequence the first DNA genome: that of the 5,375 base pair dsDNA bacteriophage ΦX174 (Sanger, et al. 1977).

Sanger's early methods would be further developed and modified (most importantly by moving to flourometric detection and scaling/multiplexing) to accommodate more complex genomes. This culminated in the completion of the Human Genome Project in 2001, in which bacteriophage-derived M13 vectors and T4 DNA ligase played crucial roles (Lander, et al. 2001; Salmond and Fineran 2015). It was around this time that the second-generation of sequencing technology (collectively known as next generation sequencing) was initially developed by 454 Life Sciences but, conceptually, extends to IonTorrent and Illumina sequencing technology (Heather and Chain 2016). In all of these methods, DNA becomes isolated in some way from each other (either by emulsion PCR or ligation to a surface), subsequently amplified using DNA polymerase, and then detected in a highly parallelized fashion. It is worth mentioning that a third generation of sequencing technology (still considered next generation sequencing) known as single-molecule sequencing. One example of this technology, initially developed by the now-defunct Helicos Biosciences and now used by Pacific Biosciences, detects, in real-time, fluorescent nucleotides as they are incorporated into the nascent polymerization product by DNA polymerase (Heather and Chain 2016).

This uniquely enables PacBio to create incredibly long reads. Another example of a single-molecule sequencing technology is the use of nanopores such that the newly polymerized DNA product gets threated through a pore in a synthetic membrane which is able to detect specific nucleotides by the application and detection of voltage (Heather and Chain 2016). These next-generation sequencing technologies, broadly enabling for DNA sequencing, are also used for RNA sequencing: all that is necessary has been to take an extra step by which RNA transcripts of interest are converted into cDNA by a reverse-transcriptase step (Ozsolak and Milos 2011). These technologies have led to widespread ease of sequencing genomes: currently, there are over 7000 viral, 100,000 prokaryotic, and 4500 eukaryotic genomes in the NCBI genome database (NCBI 2017).

Protein sequencing, in relation to RNA or DNA sequencing, has been historically and is presently more difficult on account of more variation of substituents (four standard nucleotides versus twenty standard amino acids), lack of selective amplification (there is no protein equivalent to PCR), and heterogeneity among various proteins made inside of a cell (which, though an issue, can be overcome by selective tagging and purification techniques). Initial work into protein sequencing focused on a technique developed by Peer Edman where chemical cleavage of the amino terminus of a protein is used to identify each amino acid in a step-wise manner (Steen and Mann 2004). This burdensome technique wound up being largely displaced by protein mass spectrometry, in which proteins are cleaved, ionized, and have their masses are measured relative to their displacement in a vacuum (Steen and Mann 2004). Though protein mass spectrometry has become the current state-of-the-art, with various techniques able to pick up on single amino acid substitutions, isotope shifts, and protein modifications, there is some hope that future developments will lead to single-molecule labeling or nanopore-esque-based protein sequencing.

Synthesis

Paralleling the increasing capability for sequencing or "reading" of biological material has been efforts towards synthesis, or, by analogy, "writing" these DNA, RNA, and proteins. The first synthetic gene ever built was the 77 base-pair tRNA in 1940 (Hughes, et al. 2011). De novo DNA synthesis using phosphoramidite chemistry progressed into the 1950s and can be noted for being error-prone and limited in length (Kosuri and Church 2014). While the same basic phosphoramidite chemistry has persisted as the state-of-the-art, major advancements have occurred: multiplexing, ligation-supplemented synthesizing, and error-correcting protocols have enabled the reliable construction of larger and larger assemblies up to the gene level (Hughes, et al. 2011; Kosuri and Church 2014). These advancements have facilitated the complete chemical syntheses of multiple genomes, including the first fully synthetic genome, that of bacteriophage ΦX174 in 2003 (which, fittingly, and noted previously, was the first whole genome sequenced) (Hughes, et al. 2011; Salmond and Fineran 2015). While costs and time requirements have decreased (a gBlock from IDT currently cost run from $89 USD for 125-250 base pairs up to $549 USD for 2751-3000 base pairs in a few weeks while the original 77 nucleotide tRNA, though cost unknown, required five years of work) they have not decreased at the same rate as sequencing costs (Hughes, et al. 2011; Integrated DNA Technologies 2017).

RNA synthesis is almost always done *in-vitro* or *in-vivo* (rather than by chemical synthesis) using DNA as its template and enzymes such as the T7 RNA polymerase. While chemical protein synthesis of peptides is routine, the creation of larger polypeptides on the protein-scale are almost always also created *in-vivo* or *in-vitro*.

Copy, Cut/Pasting, and Editing

As the development of sequencing and synthesis progressed, others were discovering or developing the technology for being able to copy, cut/paste, and edit DNA (which, by proxy, also includes RNA and proteins). The ability to directly copy DNA via polymerase chain reaction, though routine and instrumental after, only matured in the mid-1980s as a the product of work by many over the previous 20 years (Bartlett and Stirling 2003). Until then, it took another set of tools to allow the creation of transgenic constructs and novel gene circuits. The existence of bacteriophages (or, more specifically, prokaryotic defense mechanisms in which to defend against them) led to the development of this suite of tools which have been critical in the development of this cut/paste (and, later editing) technology.

The discovery of restriction-modification systems, which are hypothesized to have evolved in prokaryotes for the purpose of protecting the cell against foreign DNA a la bacteriophages, is one example of a discovery in basic research which ultimately drives technology in the use of restriction enzymes (Salmond and Fineran 2015). Until the more novel forms of DNA cloning in the last decade or so, restriction enzymes and DNA ligases (also bacteriophage derived) were the only tools by which various pieces of discordant DNA could be stitched together to make novel transgenic constructs or genetic circuits. These key discoveries, together with PCR, enabled the burgeoning field of genetic engineering.

Satisfyingly similar to the discovery and technological exploitation of restriction-modification systems has been the discovery and technological exploitation of CRISPR/Cas systems. CRISPR systems, the arrays of which were first noticed in early prokaryotic sequencing as runs of small identical sequences interspersed by wildly different intervening sequences of roughly the same size, were eventually postulated and shown to provide sequence-specific immunity against bacteriophages in 2007

(Barrangou, et al. 2007; Salmond and Fineran 2015). Within five years, CRISPR/Cas systems, which enable sequence-specific, RNA guided nuclease activity towards DNA in almost all parts of the biological tree of life, would become widely adopted for their gene editing ability (Salmond and Fineran 2015). Rapid genome engineering by CRISPR/Cas systems, following on from zinc-finger nuclease and TALEs, has enabled development in basic research (animal models and genetic variation), biotechnology (fuel, food, and materials), and medicine (gene therapy and drug development) (Hsu, et al. 2014).

One of the earliest and most widely adopted of the CRISPR/Cas9 systems was a type II CRISPR/Cas system which was chosen for its simplicity for adaption to biotechnological purposes: one protein, Cas9, interacts with one RNA, the sgRNA , for full nuclease activity (Salmond and Fineran 2015). A nicking Cas9 has been created by inactivating one of the two nuclease domains while a nuclease deficient Cas9 has been created by inactivating both the nuclease domains (Qi, et al. 2013). The nuclease-deficient Cas9 (which is often called dCas9) still retains DNA binding function and can act as a stand-alone repressor or an activator when fused to prokaryotic RNA polymerase subunits (Bikard, et al. 2013). Cas9 as well as dCas9 have been further used within the context of conjugative plasmids and phagemids with the explicit purpose of allowing selective gene targeting or gene modulation inside of prokaryotes (Bikard, et al. 2014; Citorik, et al. 2014). This work gives rise to the hope that such systems can be exploited to root out specific, antibiotic-resistant bacterial strains or modulate the medically relevant prokaryotic communities of the human microbiome. If so, this newest iteration of technology would truly represent the wheel turning full circle back to phage-based therapy.

Molecular Methods to Engineer Bacteriophage

The ability to read and write nucleic acids (and therefore proteins) has further enabled modern methods to engineer and evolve bacteriophage genomes. The genomes of bacteriophages can be altered inside of cells through means of homologous recombination (as in the case of bacteriophage T7), *in-vivo* protein-directed recombination (as in the case of use of the lambda phage recombination machinery and electroporated homologous ssDNA or dsDNA), and even the use of the CRISPR/Cas9 system (as has been shown for multiple phages, including the use of a type-I-E CRISPR-Cas system to engineer bacteriophage T7) (Kiro, et al. 2014; Pires and Cleto 2016). Bacteriophage genomes can also be altered and engineered *in-vitro*. The previously mentioned creation of the refactored bacteriophage T7 utilizing particular engineering design strategies was done by the division of the T7 genome into three pieces on three plasmids which, when transformed into the same cell, led to successful phage production (Chan, et al. 2005). Another technique to engineer and evolve bacteriophages has used a cell-free transcription-translation system derived from *E. coli* cells in which bacteriophages T7 and ΦX174 have been shown to replicate and assemble into infective virons (Shin, et al. 2012). Finally, bacteriophage genomes can get around host toxicity problems associated with bacteriophage genome assembly by using assembly in a foreign and distantly related species, in this case, of bacteriophage T7 onto a plasmid in *S. cerevisiae*, which then can be prepped out of the yeast and used to transform *E. coli* in which the bacteriophage will begin life cycling (Ando, et al. 2015).

These mostly modern methods of engineering bacteriophages have renewed interest in their therapeutic and biotechnological utility. In addition to previously mentioned use of bacteriophage as delivery-systems for effector Cas9/dCas9 systems, there are numerous examples of engineered bacteriophages that have been modified to deliver and overexpress a wide variety of antimicrobial proteins (Pires and Cleto 2016).

As but one example, *lexA3*, a protein which represses the SOS DNA repair system in *E. coli*, was overexpressed and delivered by an engineered M13 phage, thus enhancing the antibacterial activity of an exogenous antibiotic (Lu and Collins 2009). Additionally, these modern methods have been used to enable host range expansion of lytic bacteriophages via swapping of tail fibers (Ando, et al. 2015). Engineering efforts such as the work into decreasing the impact of engineered phages on mammalian systems (thereby decreasing immunogenicity) and the use of engineered non-lytic phages from wild-type lytic phages (thereby mitigating worry over LPS-based toxic shock from lysed bacterial cells) has sparked a renewed interest in the use of bacteriophage therapy (Barbu, et al. 2016; Pires and Cleto 2016). Beyond their therapeutic potential, engineered bacteriophages have been created for bacterial detection and diagnostic purposes (Pires and Cleto 2016). Finally, ability to conjugate or display various chemicals and proteins, such as antibiotics, peptides, and antibodies, on the surface of bacteriophages has led to their use as chassis for delivery of these materials *in-vivo* (Pires and Cleto 2016). Researchers have used bacteriophages in the field of materials science in unexpected ways, going as far as to fabricate cobalt-oxide nanowires from engineered M13 bacteriophages for the purposes of increasing storage capacity in lithium ion batteries (Nam, et al. 2006; Pires and Cleto 2016). These increasingly complex and exotic uses of bacteriophages surely were unanticipated by those founding members of the Phage Group and represent bacteriophage's contribution to the modern manifestation of engineering and systematizing biology called synthetic biology.

Synthetic Biology: the Latest Term for a Powerful Concept

At the time of his death physicist Richard Feynman, whose foray into phage work during a sabbatical summer led to his inclusion on a paper in 1962, had on his blackboard

the statement “what I cannot create, I do not understand” (Edgar, et al. 1962). This statement expresses the ethos behind the relatively recent creation of so-called “synthetic biology”. The number of answers to the question what exactly is synthetic biology are likely as numerous as people asked. The history and coverage of synthetic biology can be found elsewhere in very excellent reviews and perspectives (Cameron, et al. 2014; Gardner and Hawkins 2013). The following is a brief attempt to extricate meaning from the term in order to understand the field as a whole and the motivation behind this thesis work.

Synthetic biology is the newest title in a long line of titles such as genetic engineering, biotechnology, bioengineering, biodesign, and biological engineering which focus on one central concept: control. Synthetic biology is guided by innate human desire (and perhaps hubris) that believes matter can be and should be shaped in our desired way such that we can design the world around us to fit whatever specifications we want. This almost godlike ability to shape matter, in this case biological matter, is a guiding ideal far away from the reality of the situation. The reality is that we simply do not know enough about the various scales of biological phenomenon and how they interact with one another to reach this totality of control. We do, however, have to tools and technologies in place (including the few listed in the previous section) to increase our ability to read and write DNA, RNA, and proteins. Together with the inherited burden of human knowledge, this technology is being used for creation in order to test and increase our understanding of biological phenomenon. Synthetic biology is the attempt to do this.

Synthetic biology uniquely stands out from the others in its stated desire for standardization. In a 1999 DARPA white paper, Adam Arkin and Drew Endy write that:

> Without standardization, the qualitative design methods used in other engineering fields are simply inapplicable. [In practice, rational design of biological systems] is usually realized through an expensive step-wise trial and error approach or through mutation and selection. Furthermore, these otherwise practical approaches are limited in terms of problems they can solve. To address this deficiency, we propose herein a program to produce a set of well-characterized and systematized biological components that can be generically assembled to create custom biological circuitry. (Gardner and Hawkins 2013)

This standardization of parts and of processes has led to the creation of the Registry of Standard Biological Parts and it was indeed the driving force behind notable successes such as the production of artemisinin in yeast or the creation of a completely chemically synthesized bacterial genome (Cameron, et al. 2014).

Cognizant of its successes, however, it may be tempting to ignore some of the inadequacies of synthetic biology. Efforts to standardize can (and arguably have, in the case of synthetic biology) concentrate power, money, and influence to a small geographic subset of labs and tastemakers. Not everyone has the ability as J. Craig Venter does to have a whole institute devoted to building a fully synthetic bacterial chromosome (Gibson, et al. 2010). In that 1.08 million base pair Mycoplasma mycoides genome, JCVI inserted various watermarks, including misquoting Feynman's "What I cannot create, I do not understand" as "I cannot understand what I cannot create" (Gibson, et al. 2010). Though trivial and amusing, this error points to an uncomfortable truth: even the best of rational design can still miss the mark because of human error. As such, sometimes complete control is unnecessary and there is a roll for the uncontrolled: enter directed evolution.

Directed Evolution Schemes

Directed evolution schemes have been around since humans first began domesticating animals, farming crops, and brewing beer in 6000 BC (Buchholz and

Collins 2013). Though more was accomplished through ignorance of the system than knowledge during these early efforts (the details behind fermentation were not worked out until the 1850s) their success is undeniable: for instance, a quick look at the size difference between modern sweet corn and its ancestor teosinte is ample evidence of successful breeding and selection (Buchholz and Collins 2013). As biotechnological tools have been developed, the veil of mystery previously cloaking biological phenomenon has been lifted. These tools enabled better understanding of the system and, ultimately, the building of new systems in experimenter-defined evolutionary contexts and the birth of modern directed evolution of biomolecules.

Modern directed evolution methods share certain commonalities. All usually begin with a library of variants (rationally designed, caused by site-directed mutagenesis, or allowed to mutate *in-vivo*) which are then subjected to some sort of selection scheme in which each variants genotype is linked to some sort of manifesting phenotype. Depending on the method of selection, this manifested phenotype will enable (or disable) the further enrichment of that variant, together with other selected-for variants, from the previous pool of variants. After this enrichment, the new pool may get further mutagenized, experience further rounds of selection, or both. It is important here to discern between a screen and a selection within the context of directed evolution. A screen "requires the inspection of individual phenotypes" whereas a selection "bypasses the need to individually inspect each library member and instead links and activity of interest to physical separation of the encoding DNA or survival of the organism producing active library members" (Packer and Liu 2015).

A concrete example of a directed evolution technique mentioned previously is phage display technology. A M13 bacteriophage population, each displaying a single-chain variable fragment (scFv) antibody variant on its surface, can be panned to select for

antibodies which bind to a protein of interest (let us use T7 RNA polymerase in this example). Bound T7 RNA polymerase on a surface of a well or on beads are washed with the phage library. Those scFv variants with partial binding activity to the T7 RNA polymerase will bind while those that do not bind will be washed away. Then, the phages which contain the nucleic acid code (i.e. genotype) for those high-affinity scFVs are then re-amplified via PCR or even through infection and replication in *E. coli* cells. Phage display, together with mRNA, ribosome, and cell surface display, are all subtypes of a selection type known as affinity selections (Packer and Liu 2015). Affinity selections are powerful technologies which has proven effective as developing therapeutic antibodies as well as other binding proteins (Packer and Liu 2015).

Although affinity selections were some of the earliest directed evolution techniques they are far from the only ones. Compartmentalized Self-Replication (CSR) used water-in-oil emulsions to isolate individual *E. coli* cells from each other in which is expressed a DNA polymerase variant which would amplify a copy of itself via emulsion PCR if that variant was active in the desired selection environment (Ghadessy, et al. 2001). Compartmentalized Partnered Replication (CPR) extended the CSR technology to allow the directed evolution of any gene of interest as long as it could be linked to the production of DNA polymerase inside of *E. coli* cells (Ellefson, et al. 2014). CPR is not the only extensible development of a new selection system from a previous methodology: a yeast display system, named YESS, has been exploited for the evolution of bond-cleaving enzymes such as proteases which, when active, cleave and retain an epitope tag used to enrich successful variants via FACS sorting (Yi, et al. 2013).

Yet another example is the exploitation of the production of an antibiotic resistance marker or metabolically essential gene (i.e. complemented auxotrophy) through some means with the overall survival of a bacterial cell as a whole. A final

example of a directed evolution system is phage assisted continuous evolution (PACE) in which activity of the gene of interest, which is expressed in an *E. coli* cell from inside a modified M13 bacteriophage genome, is linked to the production of the M13 pIII gene which is a coat protein reported to be essential for its replication (Packer and Liu 2015). In PACE, the production of the pIII coat protein thus allows active variants to bud out of that infected *E. coli* cell and spread to other, uninfected *E. coli* cells which are provided in a constant in-flow media to replace spent and infected media (Packer and Liu 2015).

These last two selections, antibiotic selection/complemented auxotrophy and PACE, are distinct from those previous mentioned because they are usually ran as continuous selections. Continuous selections (unlike affinity selections, CSR, CPR, and YESS systems, which are discrete selections) are set up in such a way to allow for all of the aforementioned steps of selection (creation of variation, pressure of selection to select for favorable variants, and enrichment of those variants) to occur at the same time. In theory, continuous evolution is advantageous because it minimizes the amount of time and manual work necessary to turn rounds relative to discrete selection methodologies. In practice, continuous evolution strategies can be particularly finicky to set up and, by their continuous nature, suffer the disadvantage of allowing so-called "cheaters", variants which have somehow short-circuited the selection, to quickly come to dominate the variant library population.

What is the roll of directed evolution in the era of synthetic biology? Efforts towards biological standardization via synthetic biology are imbued with an implicit hubris: it may be that certain aspects of biological phenomena are too complex to lend themselves to systemization. Engineering biology, unlike other engineering disciplines, is not obviously governed by a simple series of equations but, rather, united by one universal theme: evolution happens. Evolution is the great unifier among all biological

phenomena. Because of this, much work has been done in order to study not just how evolution works and but also into controlling, accelerating, and biasing it in a laboratory setting. The former work can be generally classified as “evolutionary biology” and the latter called “directed evolution”. In this dissertation, work will be presented which blurs the lines between evolutionary biology and directed evolution yielding unanticipated outcomes and new methods of biological control. This work provides evidence that although the ability to rationally design biology is important, various ways to rationally control and direct evolution offer a complementary strategy to the systematization of synthetic biology.

CHAPTERS

Chapter 1: Predicting Evolution of the Transcriptional Regulatory Network in a Bacteriophage

Abstract

Prediction of evolutionary paths has been a desirable but elusive goal, and requires a deep knowledge of both underlying mechanisms that relate genotype to phenotype, and an understanding of how phenotype impacts organismal fitness. We have relied on foreknowledge of the sequence specificity of a T7 RNA polymerase variant (G78-KIRV) that recognizes a unique promoter that is three nucleotides away from the wild-type to predict and guide the evolution of the entire bacteriophage T7 regulatory network. A mutant of T7 phage lacking its RNA polymerase gene was passaged on a bacterial strain providing G78-KIRV in *trans*, resulting in higher overall fitness in the population. As the phage adapted to G78-KIRV, we anticipated the evolutionary paths to higher fitness could be predicted based on *in-vitro* transcription rates of various promoters. The predictions failed to capture some of the promoter evolution pathways, but these in turn revealed new insights into the promoter specificities of the G78-KIRV polymerase, further demonstrating the close predictive relationship between sequence specificity and phage fitness. Overall, this study points toward the feasibility of predicting evolution in well characterized, simple systems.

Introduction

The historical dichotomy between studying living systems through reductionist molecular biology or holistic evolutionary biology has given way to a unification termed

the functional synthesis (Dean and Thornton 2007). This synergistic approach is increasingly used to empirically validate evolutionary hypotheses and has been applied to such disparate subjects as insecticide resistance, mouse coat color, and antibiotic resistance (Dean and Thornton 2007; Hoekstra, et al. 2006; Newcomb, et al. 1997; Weinreich, et al. 2006). A long-term goal of the functional synthesis that may now be within reach is evolutionary prediction of both phenotypes and the genetic pathways leading to those phenotypes. Predicting organismal moves on a fitness landscape requires systems in which the genotype-phenotype relationship is well-characterized, and the fitness consequences of the different phenotypes are also understood. At this time, these properties are most feasibly extracted from systems in which there are relatively few alternative molecular states that nonetheless have major effects on fitness.

Among the simplest, yet still complex, systems that are particularly amenable to evolutionary prediction is the transcriptional regulatory network of bacteriophage T7. Bacteriophage T7 is a lytic, 40kB dsDNA podovirus that infects *E. coli* (Molineux 2006). The genome encodes nearly 60 genes and is unusual among phages in that most gene expression requires a phage-encoded RNA polymerase. Gene expression is controlled by 17 phage promoters in the genome, and there are a number of overlapping transcripts. However, underlying this complex regulatory network is a single phage RNA polymerase (T7 RNAP) that interacts with a highly defined and well-characterized promoter. We and others have previously used protein engineering and directed evolution to craft T7 RNAP variants that recognize alternative promoter sequences (Ellefson, et al. 2014; Meyer, et al. 2015).

In particular, we have generated a T7 RNAP variant, G78-KIRV (Ellefson, et al. 2014) that can utilize a promoter variant that is three nucleotides away (CGG) from the wild-type T7 promoter (GAC) at positions -11 to -9 (Rong, et al. 1998), and that is

orthogonal to the wild-type, in that the polymerases cannot readily cross-utilize one another's promoters. The synthetically-derived G78-KIRVpolymerase enables us to perturb the regulatory network of T7 phage, and to observe the adaptation of the network to a major shift in specificity of the transcriptional apparatus. Most importantly, the observed evolutionary pathways for promoters can be compared with initial predictions based on the known and fixed promoter specificities of the mutant G78-KIRV.

Results

Initial prediction of a single mutation path to a mutant promoter triplet

The G78-KIRV RNA polymerase was a product of directed evolution; presumably it had never before been experienced in the evolutionary history of the wild-type T7 phage. Thus, it was unclear to what extent the native bacteriophage T7 was able to evolve in response to this polymerase. This was especially true given that the polymerase had been evolved to transcribe from a promoter that was three mutations from the consensus T7 promoter. By forcing T7 the use this new RNA polymerase in place if its own, the phage was confronted with a potentially deep fitness valley that required the acquisition of several mutations to regain robust replication during passaging.

To determine the most likely single-step evolutionary paths for the bacteriophage promoter network utilizing G78-KIRV RNAP, we carried out *in-vitro* transcription assays using purified G78-KIRV RNAP on the wild-type T7 promoter (GAC), on the CGG G78-KIRV promoter, as well as on the three single- and three double-mutant promoters intermediate between the wild-type and the G78-KIRV promoter (Figure 1.1). These transcription reactions confirmed the previously observed orthogonality (lack of activity of heterologous promoter-RNAP combinations) but, more importantly revealed

the transcriptional activity of the six intermediate mutant promoters (Figure 1.1). While G78-KIRV RNAP had been evolved to transcribe only the mutant promoter triplet CGG, it nonetheless (and not surprisingly) had a higher activity on the six intermediate mutant promoters than on the T7 wild-type triplet (GAC). Of the three promoters one mutation from wild-type, the GAG promoter displayed the highest activity. Of the three double mutants, both the CAG and GGG promoters displayed higher transcriptional activity than CGC (Figure 1.1). These results thus provided a foundation for predicting the evolution of T7 RNAP promoters in the phage challenged with G78-KIRV polymerase.

T7 evolves to use the alternative polymerase

In initiating the evolution of wild-type T7 in response to the new RNA polymerase (G78-KIRV), we first determined whether a phage variant whose genome lacked the wild-type RNAP gene (T7Δ1) could grow with the G78-KIRV polymerase expressed in *trans* from a plasmid. T7Δ1 was complemented by the plasmid, but its plaques were noticeably smaller than plaques of wild-type T7 (Figure 1.2), suggesting a fitness deficiency. The fact that plaques could be generated at all was surprising given the 3 mutation distance between the optimal triplet for G78-KIRV and the promoters found in T7 wild-type. It seems likely that the overexpression of the G78-KIRV polymerase before phage infection is at least partially responsible for the complementation observed. The reduced initial growth meant that the phage could be expected to evolve higher fitness after serial passaging, as seen in the previous study complementing T7Δ1 with bacteriophage T3 RNAP (Bull, et al. 2007).

Four replicate lines of T7Δ1 were subjected to repeated serial transfer on the complementing host (Fig. 1.3). By providing the RNAP in *trans*, all regulatory evolution was forced to occur in the promoters. BL21 *E. coli* cells harboring the G78-KIRV RNAP expression plasmid were grown for one hour with 100uM IPTG induction, at which point

T7Δ1 phage were added at a multiplicity of infection (MOI) of 0.01-0.1 and the culture grown until lysis. This protocol allowed expression of the complementing RNAP in all cells prior to phage infection. The parental T7Δ1 lysed the culture within ~90 minutes, whereas either wild-type T7 phage or T7Δ1 with the wild-type polymerase provided in *trans* fully lysed the culture in ~30 minutes. Over the course of the first 50 passages the lysis time decreased, ultimately going from 90 minutes to 30 minutes in the presence of IPTG induction (Figure 1.4). At this point, serial passaging was continued without IPTG induction, relying only on the leakiness of the LacUV5 promoter to drive expression of the G78-KIRV RNAP. This resulted in an initial increase in culture lysis time to ~60 minutes, which then again decreased back to ~30 minutes during passages 51-100 (Figure 1.5).

After 100 passages the culture lysis time dramatically decreased in all four replicates – measured on un-induced G78-KIRV RNAP cells (Figure 1.6A). Because time to lyse a culture is an indirect fitness measure, fitness assays of growth on plates were performed in which phage were allowed to grow for a fixed time on un-induced G78-KIRV RNAP cells. Fitness assays were carried out for the wild-type T7 phage, the ancestor T7Δ1 phage, and the four evolved T7Δ1 phage lines at passage 100. These fitness assays showed that the four evolved strains reached nearly the same doublings per hour as wild-type T7 phage on un-induced G78-KIRV RNAP cells (Figure 1.6B) and we therefore set out to determine the mutations for increased fitness–presumably to the regulatory network of the T7 promoters.

Limited promoter changes and promoters that changed in the evolved lines

All four adapted lines were sequenced at two time points, after transfer 50 and after transfer 100 (utilizing MiSeq 2x250 reads with at least 3.6x105 reads per sample). Across the four lines, substitutions occurred in only five of the 16 T7 promoters: ϕOL,

ϕ1.5, ϕ2.5, ϕ6.5, and ϕOR (Figure 1.7). In the initial 50 passages all four lines exhibited an anticipated GAG single change toward the promoter specificity of G78-KIRV in one to three of the promoters (Table 1.1). But unanticipated CCC and GCC changes were also observed in all four lines in one to three of the promoters (Table 1.1). By passage 100, promoters were found with two substitutions toward the G78-KIRV triplet (Figure 1.7). But yet another unanticipated change (AAG) was found in one to two promoters in two lines (Table 1.1, Figure 1.8).

Individual promoters followed different paths. Most promoters failed to evolve at all, but even those that did evolve often failed to carry the same substitutions as others that evolved. The ϕ2.5 promoter evolved in only one of the four lines (in line 3 to GAG), and the mutation did not fix in the population, occurring in 46.6% of the phage by passage 50 but only 25.6% by passage 100 (Figure 1.8, Table 1.2). In contrast, the ϕ1.5 promoter evolved to GAG all four populations by passage 50 and remained fixed through passage 100. These two promoters can be viewed as the limit cases for others; in one case (the ϕ2.5 promoter evolution in a fraction of line 3) virtually no evolution occurred, and in the other (the rapid fixation in all lines at the ϕ1.5 promoter), there was rapid evolution toward a single endpoint, suggesting strong selection for a particular sequence.

The ϕOR promoter first evolved to GCC in all four lines by passage 50, but then proceeded to evolve into a heterogeneous mixture of GCC, GGG, AAG, GAG, and CCC by passage 100 (Figure 1.7).

Interestingly, the non-canonical ϕOL promoter, which contains AAC at positions that correspond to the consensus GAC, followed a different path, evolving to AGC in 39% of the line 4 population by passage 50 (Table 1.2) (Dunn and Studier 1983). By passage 100, the ϕOL promoter in lines 1, 2, and 3 evolved to ACC and additionally to AGC in line 3 (Figure 1.7, Table 1.2). Unexpectedly, the AGC substitution in 39% of the

line 4 population at passage 50 decreased to only 3.2% of the population by passage 100 (Table 1.2). This coincided with and may possibly be explained by widespread duplication of the ϕOL promoter in all lines by passage 100, including duplications with mutant promoters (Figure 1.7, Table 1.2).

The ϕ6.5 promoter seemed to follow a combination of these paths. It was extremely heterogeneous at both passage 50 (GAC, GAG, CCC, and GCC) and 100 (GAG, CCC, GCC, GGG, and AAG), and also eventually duplicated in lines 2, 3, and 4 (Figure 1.7, Table 1.2).

Evolution outside of promoters

Following 50 passages, the four lines each had fixed between 2-7 non-promoter substitutions (Table 1.3). Following 100 passages, the four lines had between 1-6 fixed non-promoter substitutions (Table 1.3). Interestingly, the lines after 100 passages retained only a fraction of the fixed substitutions observed at 50 passages (with one line losing all of the fixed, non-promoter mutations, two lines retaining one mutation, and one line retaining two mutations). The fact that most molecular evolution occurred within promoters is consistent with the expectation that most selection operated on polymerase-promoter binding. Of note, the only fixed non-promoter mutation found in all of the lines was an E34K mutation in the major capsid protein.

In-vitro transcription of unanticipated promoter triplets

The promoter triplets GCC, CCC, and AAG had not been tested for *in-vitro* activity because they were not on the direct evolutionary pathway between wild-type and G78-KIRV promoter sequences. After discovering these triplets in the evolved phages, purified wild-type and G78-KIRV RNAPs were tested for their activities on these new promoters (Figure 1.8 and 1.9). In line with the serial passaging results, G78-KIRV

RNAP was as active as or more active on these three unanticipated promoters than on the three predicted single-step mutants (Figure 1.9).

Discussion

This is the first example in which the evolutionary landscape of a system component has been used to perturb and predict the evolution of an interacting system as a whole, in this case the bacteriophage T7 RNA polymerase and the bacteriophage T7, respectively. The fact that the laboratory-evolved G78-KIRV T7 RNA polymerase variant has no apparent counterparts in nature, and that the promoter utilized by this polymerase requires at least three mutations from the wild-type T7 consensus promoter for optimal transcription activity, means that challenge experiments favor an evolution that should be novel to the phage in several respects. The experiment thereby tests basal issues in how adaptation can occur in complex systems.

At the outset we formed a prediction matrix (Figure 1.1) for the most likely evolutionary path in response to the novel transcription machinery. The T7Δ1 phage used in our selections has 16 promoters that could potentially evolve, yet mutations evolved in only 4 - 5 of these. Interestingly, these were the same promoters that evolved in a previous study in which T7Δ1 was forced to use T3 RNAP. Use of T3 RNAP requires only a single mutational change in a promoter to achieve high transcription activity (Bull, et al. 2007). The similarity between the two studies suggests that these promoters are the primary transcriptional hubs that are under the highest selective pressure.

It is unexpected that so few of the 16 T7 promoters evolved to accommodate the new RNAP, but also not implausible. The relative spacing of the mutant promoters suggests that the inherent processivity of the T7 RNAP is high enough to result in sufficient gene expression of essential phage genes along the T7 genome. The T7

genome has two known terminators, one for *E. coli* RNAP at position 7555 and one for T7 RNAP immediately after gene 10 at position 24,170. Transcription from the 3 *E. coli* promoters in the initial 750bp of the phage genome by the *E. coli* RNAP produces transcripts of up to 7kB (Molineux 2006). The ϕ1.5 and ϕ2.5 phage promoters, which start at nucleotides 7761 and 9090, respectively, are approximately 10kB away from the ϕ6.5 promoter, and in the absence of downstream promoters, the ϕ6.5 promoter would need to transcribe approximately 20kB to produce all of the necessary late genes. While transcripts from these promoters alone would be quite long, T7 RNA polymerase has the ability to transcribe up to 27kB *in-vitro* and up to 32kB *in-vivo* (Mairhofer, et al. 2013; Schelle 2002). Alternatively, G78-KIRV RNAP may still have enough activity on the remaining wild-type promoters to enable limited transcription.

The evolutionary paths chosen made sense in terms of the transcription activities of the G78-KIRV RNAP (Figure 1.8). Three of the six promoters that evolved followed the single and double mutant paths expected. For the three that did not, the mutations that accumulated revealed new, highly active promoter variants (Figure 1.8, 1.9). Overall, the data from *in-vitro* transcription experiments was surprisingly predictive of the evolutionary paths taken by the phage *in-vivo*. Single mutant paths at passage 50 had, by passage 100, continued to accumulate functional mutations (e.g. ϕ6.5 in line 3; Figure 1.7). Some paths seen in one line at passage 50 were reproduced in other lines by passage 100 (e.g. ϕOR in lines 2, 3, and 4; Figure 1.7). It should be noted, however, that no promoter in any of the 4 replicate lines evolved the triplet on which the G78-KIRV RNAP was evolved. We acknowledge that our selection here when the RNAP is already present in the cell may lead to different outcomes than a selection when the RNAP needs to be expressed from the incoming genome.

Beyond changes in promoter identities, promoters in all four lines experienced duplication events (Figure 1.7). Most occurred after transfer 50, under conditions of low expression of the mutant RNAP (Table 1.2), and thus represented a different solution than point mutations to the problem of how to regain adequate transcription activity. All four lines duplicated and mutated their ϕOL promoter, presumably supporting the function of the ϕOL promoter to direct T7 RNAP-mediated genome entry into the host cell (Garcia and Molineux 1995). Similarly, the *E. coli* A2 promoter near the ϕOL promoter at the 5' end of the phage genome duplicated in experimental line 4 (Table 1.2).

Empirical recapitulations of mutational pathways such as those demonstrated in this work can yield insights to how enzymes and organisms traverse fitness landscapes (Poelwijk, et al. 2007). As but one example, the TEM-1 β-lactamase enzyme that is known to have high activity towards ampicillin can also evolve the ability to degrade a structurally different lactam, cefotaxime, via five mutations (Weinreich, et al. 2006). Assaying fitness for all possible combinations of these five mutations revealed that 18 of the 120 possible mutational pathways resulted in stepwise, monotonic fitness increases toward cephalosporinase activity, with 10 pathways being most feasible. This mirrors our own results, in which only a fraction of feasible single-step paths were taken (Figure 1.10). Similarly, other work on the recapitulation of mutational steps in higher-order regulatory networks, such as hormone receptor:binding sites and repressor:operator interactions, have reinforced the notion that multiple different single-step evolutionary paths lead to diverse movements on fitness landscapes (Bridgham, et al. 2006; Lehming, et al. 1990).

One difference between these systems, though, is that unlike the hormone receptor:binding sites and repressor:operator interactions, phage promoters displayed no evidence of so-called sign epistasis in which an initial sampling of a lower-activity

variant was required before subsequent evolution to a higher activity position. This is perhaps because the adaptation of a promoter to a polymerase is something that occurs frequently during evolution; indeed, one can view the promoters in any genomic system as being 'rheostats' that adapt by point mutation to balance the expression of multiple different genes. The sequence changes we observed during phage evolution further emphasize this point, in that they followed 'productive' single mutational paths that likely altered expression levels. Beyond single mutations, the sequence space of promoters within two mutations of the wild-type promoter in the -11 to -9 region is occupied by N=36 possible states (9 single mutants and 27 double mutants) and is populated by all six of the promoter variants from the evolution experiment (Figure 1.10). This suggests that there are several productive paths that in turn further diversify into multiple, viable solutions on a complex fitness landscape. That said, it is unknown at present what portions of the sequence and fitness landscapes may have been restricted to the phage because it was moving only by single mutations.

Assuming that the evolution results speak to natural selection, extensive promoter heterogeneity and duplication in wild phage should be observed. Indeed, single-subunit T7-like RNAPs with wildly different promoter sequences have been identified not just in other bacteriophages but also in eukaryotic mitochondria and plant chloroplasts (Bohne, et al. 2016; Cermakian, et al. 1997), consistent with regulatory fitness landscapes that are only moderately rugged, and that conjoin multiple optima within easy reach of one another.

Conclusion

These enzyme evolution and regulatory interaction studies are useful in displaying critical features of evolutionary pathways, and they thus lead us to the prospect of

predicting the evolution of systems. This prediction of evolutionary paths is what we present here. Our choice to use a heterologous T7 RNAP to evolve T7Δ1 phage was due to the fact that the system is well-characterized down to the promoter bases involved in polymerase recognition, that different RNAPs were available, and that the phage is easily evolved (Bull, et al. 2007; Ellefson, et al. 2014; Rong, et al. 1998). Overall, it has been possible to predict much of the evolutionary trajectory of a bacteriophage in response to a synthetically-created component, and our work points toward the next step in the functional synthesis by a priori predicting rather than post hoc confirming evolutionary trajectories. Even more importantly, the fact that multiple pathways led to a recovery or near-recovery of fitness on a very short evolutionary time-scale may point the way to understanding why different types of biological systems move to position themselves in different local geographies of a fitness landscape.

Materials and Methods

Phage Passaging

Frozen aliquots of BL21-Gold cells (Agilent) transformed with pLUV-G78-KIRV-RNA polymerase (wild-type T7 RNA polymerase with Q744K, L747V, N748H, L749I, R756E, L757M, H772R and E775V mutations, from *Ellefson et al*., 2014) were thawed and used to inoculate 10mL of 2xYT with proper antibiotic and grown at 37° C to achieve a density of 10^8 cells/mL at 60 minutes. For the first 50 passages, isopropyl-L-thio-β-galactoside (IPTG) was added to a concentration of 1mM before the 60 minutes of growth (no IPTG addition for the second 50 passages). After 60 minutes to reach density, 1uL of previous bacteriophage lysate was added (a 10^{-4} dilution, transferring approximately 10^7-10^8 phage). Bacteriophage was allowed to grow under the same

conditions, occasionally allowing the cultures to reach full lysis in order to stimulate recombination.

RNA polymerase purification

RNA polymerases were purified via Ni-NTA N-terminal 6xHis methods in a manner similar to elsewhere (Ellefson, et al. 2014). Briefly, pQE-WT/G78 RNA polymerase variants each respective RNA polymerase under the control of a T5 promoter/Lac operator, were transformed into BL21-Gold cells and grown overnight in 4mLs 2xYT. The next day, cells were diluted into 50mLs of 2xYT media and grown to a density of OD600 ~0.6-0.7, when they were induced with 1mM , isopropyl-L-thio-β-galactoside (IPTG). These cells were allowed to grow for 14-16 hours at 18° C, at which point they were pelleted by spinning in centrifuge at 4000g and re-suspended in 30mLs of binding buffer (500mM NaCl, 5mM imidazole 50mM Tris pH 8). This was sonicated at 50% amplitude for 2 minutes (1s ON, 1s OFF) in an ice bath. Lysed cells were spun at 10,000g for 30 minutes, then the supernatant was added to a 1mL Ni-NTA resin column which had been equilibrated with 2x binding buffer. The cells were washed with 6x column volumes of wash buffer (500mM NaCl, 20mM imidazole, 50mM Tris pH 8) then eluted into 3mLs of elution buffer (500mM NaCl, 250mM imidazole, Tris pH 8). These samples were dialyzed into 1.2L of dialysis buffer (300mM NaCl, 1mM EDTA, 1mM DTT, 50mM Tris pH 8) overnight and again into 1.2L of storage buffer (100mM NaCl, 1mM EDTA, 1mM DTT, 50mM Tris pH 8) overnight. The resulting protein was adjusted to 1mg/mL and added to equal volumes of glycerol.

In-vitro transcription

In-vitro transcription was done in a manner similar to elsewhere (Ellefson, et al. 2014). *In-vitro* transcription reactions of the spinach aptamer consisted of 40mM Tris-HCl pH 7.0, 30mM MgCl2, 6mM spermadine, 6mM each NTP, 10mM DTT, 0.5uM T7

RNA polymerase, 0.5uM DNA template and 0.17mg/mL DFHBI in DMSO. Reactions were incubated at 37° C, and were read every minute at an excitation/emission of 469nm/501nm for 2 hours in a Safire monochromator (Tecan). DNA templates were made by thermal cycling 2uM forward primer and reverse primers (see Table 1.4) with Accuprime Pfx in its buffer using the following, then gel purified: (94 °C:2 min, 12 cycles (94 °C:15 s, 50 °C:30 s, 68 °C:30 s), 68 °C:1 min).

Lysis curves

Cells were grown in a manner identical to those of the passaging scheme (achieve a density of 10^8 cells/mL) and 150uL of cells were added to each well of a clear, 96-well plate. The proper concentration of isopropyl-L-thio-β-galactoside (IPTG) (where required) was added to each well, as, too, was the proper concentration of bacteriophage (determined by previous titer) in order to give the required MOI (MOI for lysis curves here=0.01). The plate were grown in a PowerWave 340 microplate spectrophotometer (BioTek) at 37° C and OD_{600} readings were taken every minute, preceded by shaking for 10 seconds and intensity level 4.

Preparing frozen cell stocks for fitness assay

pLUV-G78-KIRV-RNA polymerase plasmid variants was electroporated into BL21-Gold cells and plated overnight. The next day, the lawn was scraped off and put into 10mL of 2xYT with proper antibiotic and grown for 2 hours, after which point 5mLs was put into 300mLs of 2xYT and grown for an additional hour. The cells were then centrifuged for 20 minutes at 4000g. The 2xYT was decanted and the cell pellets were resuspended in 5mLs of 20% 2xYT glycerol before freezing in -80° C. Cell samples were withdrawn and concentrations were found which gave and OD600 ~1.00 after 1 hour of growth when added to 10mL of 2xYT with proper antibiotic (correlating to 108 cells/mL). All fitness assays were done using the same frozen stocks.

Fitness assays

Less than 50 fresh plaque forming units (passaged and titered within last four days) were added to 4mLs of top agar, 350uL of bacteria grown for 1 hour to an OD600 ~1.00 (a density of 10^8 cells/mL), and 50mg/mL kanamycin before plating on Kanamycin plates. Plates were incubated for 3 hours, top agar was removed with spatula and added to 15mL falcon tubes, to which 4mLs of 2xYT and 1 mL chloroform was added and vortexed, then allowed to sit at 4° C for at least 1 hour. The tubes were then spun at 4000g for 15 minutes at 4° C. The aqueous layer after spinning was 5mL. The phage titer of the aqueous phase was then measured.

Sequencing and Sequence Analysis

Miseq 2x250 paired end reads were taken using purified bacteriophage DNA and aligned to the T7 bacteriophage genome (GenBank V01146.1) using breseq (Deatherage and Barrick 2014). At least $3.6x10^5$ reads were taken for each sample. The breseq output was then analyzed by to identify promoter changes.

Figures

Figure 1.1: Promoters and Single-step prediction matrix from wild-type to G78-KIRV promoter

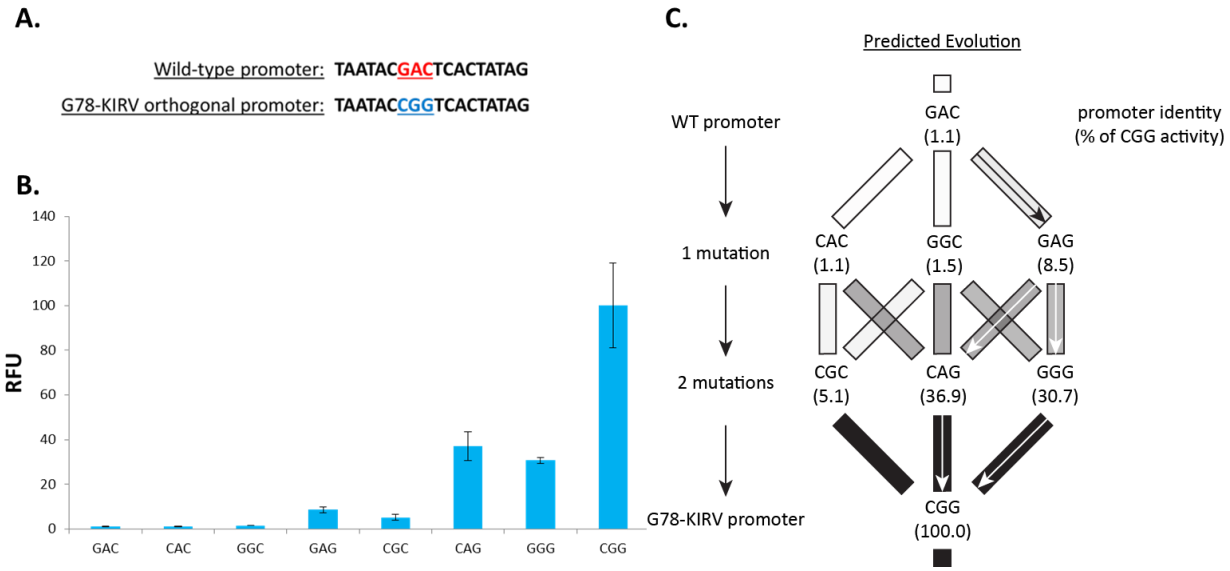


Figure 1.1: A. Wild-type (red) and G78-KIRV (blue) T7 RNA polymerase promoters. B. *In-vitro* activity of G78-KIRV RNA polymerase on wild-type (GAC), G78-KIRV (CGG), and intermediate promoters. G78-KIRV RNA polymerase was purified and used for *in-vitro* transcription using a linear dsDNA spinach template with each of the respective promoters (See Table 1.4). Fluorescence after 1 hour was taken in triplicate shown with standard error as a percentage of G78-KIRV T7 RNA polymerase activity on its cognate promoter (CGG). C. Single-step prediction matrix shows the six intermediate promoters as they take single-steps from the wild-type promoter to the G78-KIRV promoter. The boxes are shaded relative to each respective promoter activity using G78-KIRV RNA polymerase and arrows show the hypothesized most likely steps through mutational space based on based on *in-vitro* transcription activity assay. Percentage activity of G78-KIRV RNA polymerase on the CGG promoter is shown in parentheses.

Figure 1.2: Phage plaques on BL21-Gold *E. coli* cells induced with 1mM pLUV-G78-KIRV RNA polymerase

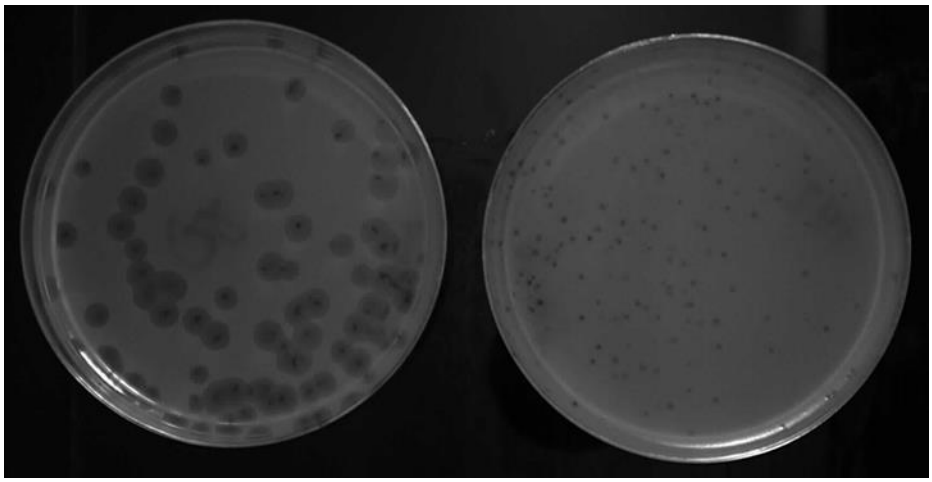


Figure 1.2: Phage plaques on BL21-Gold *E. coli* cells induced with 1mM pLUV-G78-KIRV RNA polymerase. Wild-type T7 bacteriophage population (left) and the ancestor T7Δ1 (right).

Figure 1.3: Passaging scheme

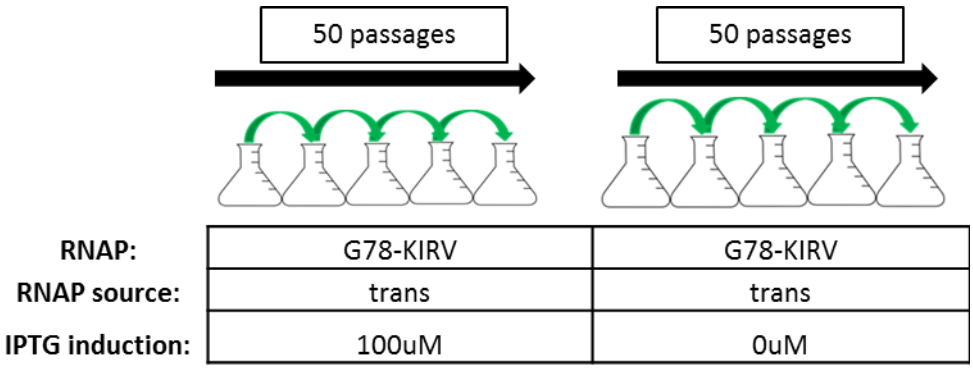


Figure 1.3: Passaging scheme over course of experiment. Four replicate T7Δ1 phage lines were initially subjected to 50 passages on BL21 cells expressing the G78-KIRV RNA polymerase mutant using 100uM IPTG in *trans* (i.e. from a plasmid). Passaging was continued for another 50 passages in the same manner using 0uM IPTG induction.

Figure 1.4: Plate reader-based lysis assays of the four passage 50 isolates

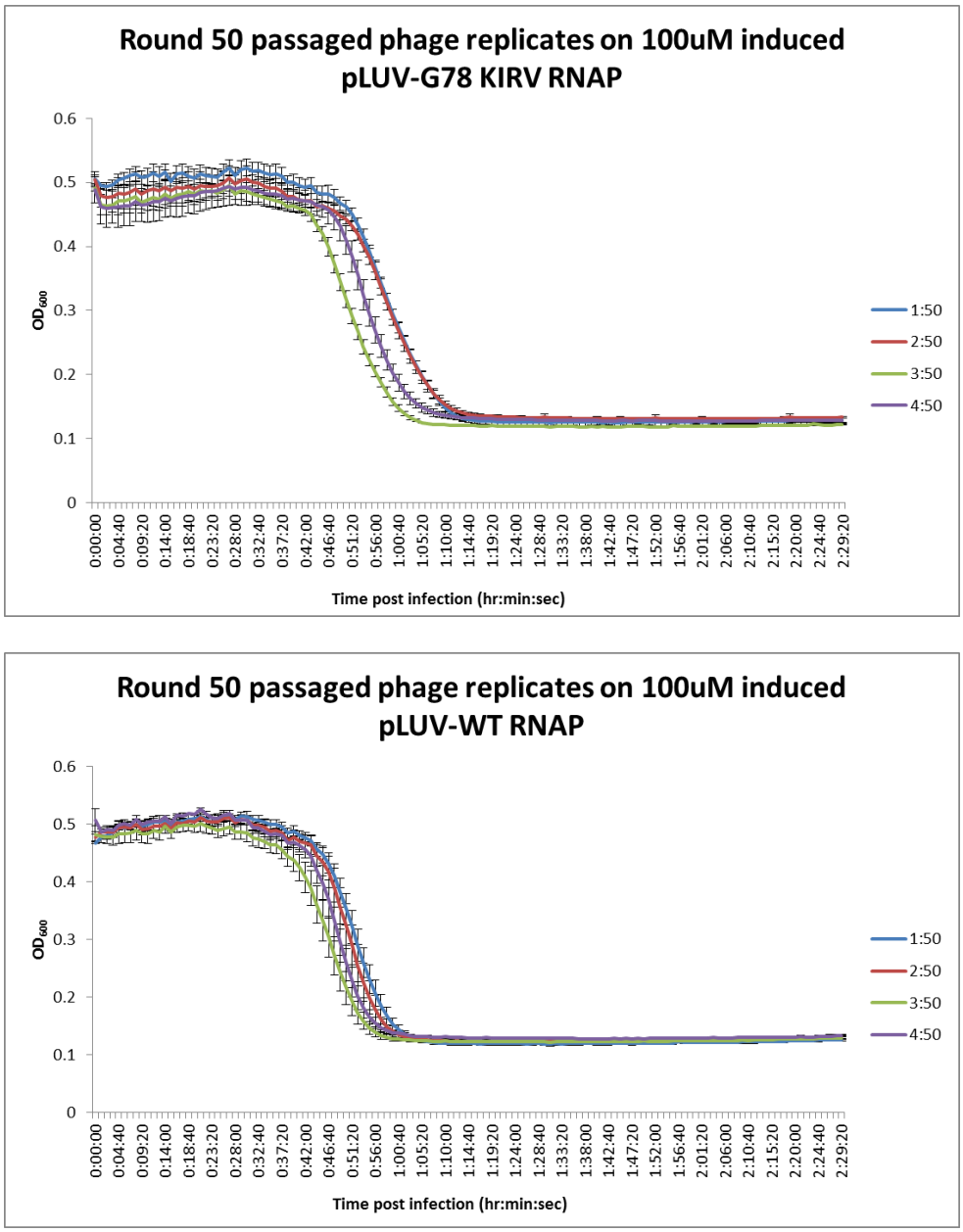


Figure 1.4: Plate reader-based lysis assays of the four passage 50 isolates. Passage 50 isolates were tittered and used to infect either 100uM IPTG induced pLUV-G78-KIRV RNA (A) or 100uM IPTG induced pLUV-WT-RNAP (B) cells at an MOI=0.01. Error bars represent standard deviation.

Figure 1.5: Minutes per passage

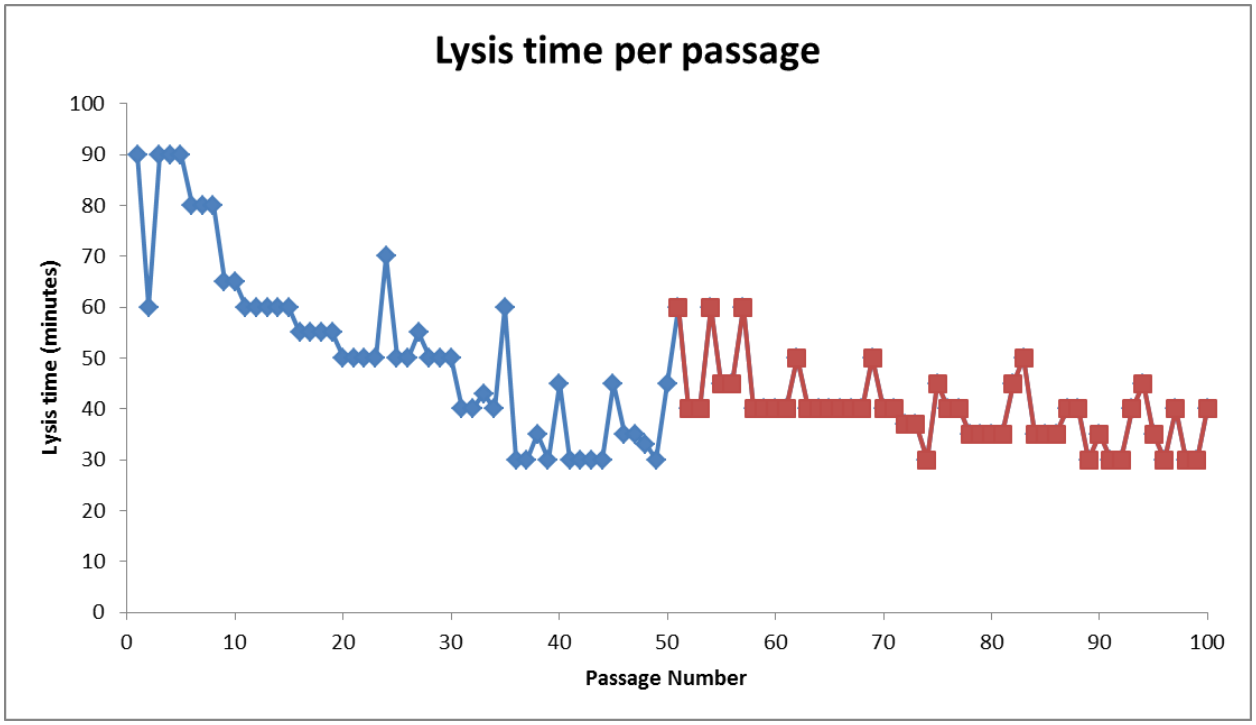
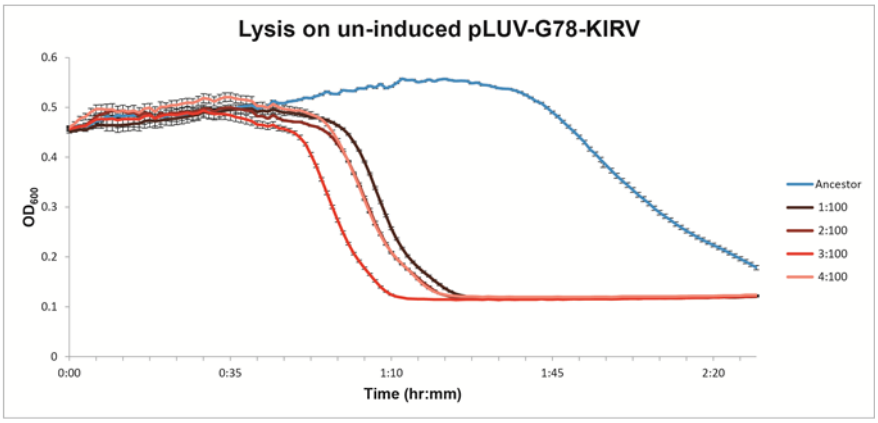


Figure 1.5: Minutes per passage. The four replicate phage populations lysed consistently within two minutes of each other. Phage lysis time, defined here as the amount of time to clear the 10mL flasks of mid-log *E. coli* cells after addition of 1uL of previous lysate, decreased during both the first fifty passages with high expression of the G78-KIRV RNA polymerase (blue) as well as the subsequent fifty passages with low expression of the RNA polymerase (51-100, red).

Figure 1.6: Lysis time and fitness assays after 100 passages

A.



B.

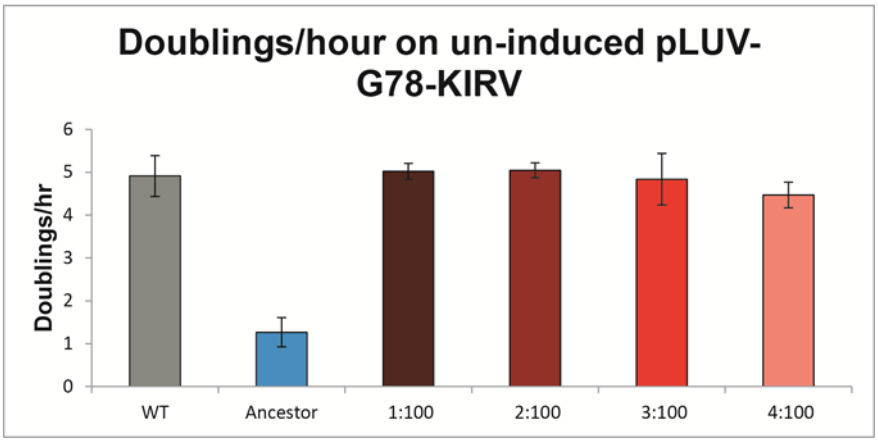


Figure 1.6: A. Whole culture lysis curves (MOI=0.01) for each of the four evolved strains after 100 passages (approximately 72 hours of adaptation) relative to the original T7Δ1 ancestor. B. Plate growth fitness assays (calculated as doublings/hour) comparing WT T7, T7Δ1, and the four evolved strains after 100 passages. The WT T7 phage carries its own RNAP, so the complementation provided by these hosts is not necessary. Note that these assays were conducted from the increase in titer during three hour growth on plates; the growth rate under these conditions is far less than in liquid (Bull, et al. 2011).

Figure 1.7: Promoters mutated from the wild-type T7 promoter in the course of the experiment

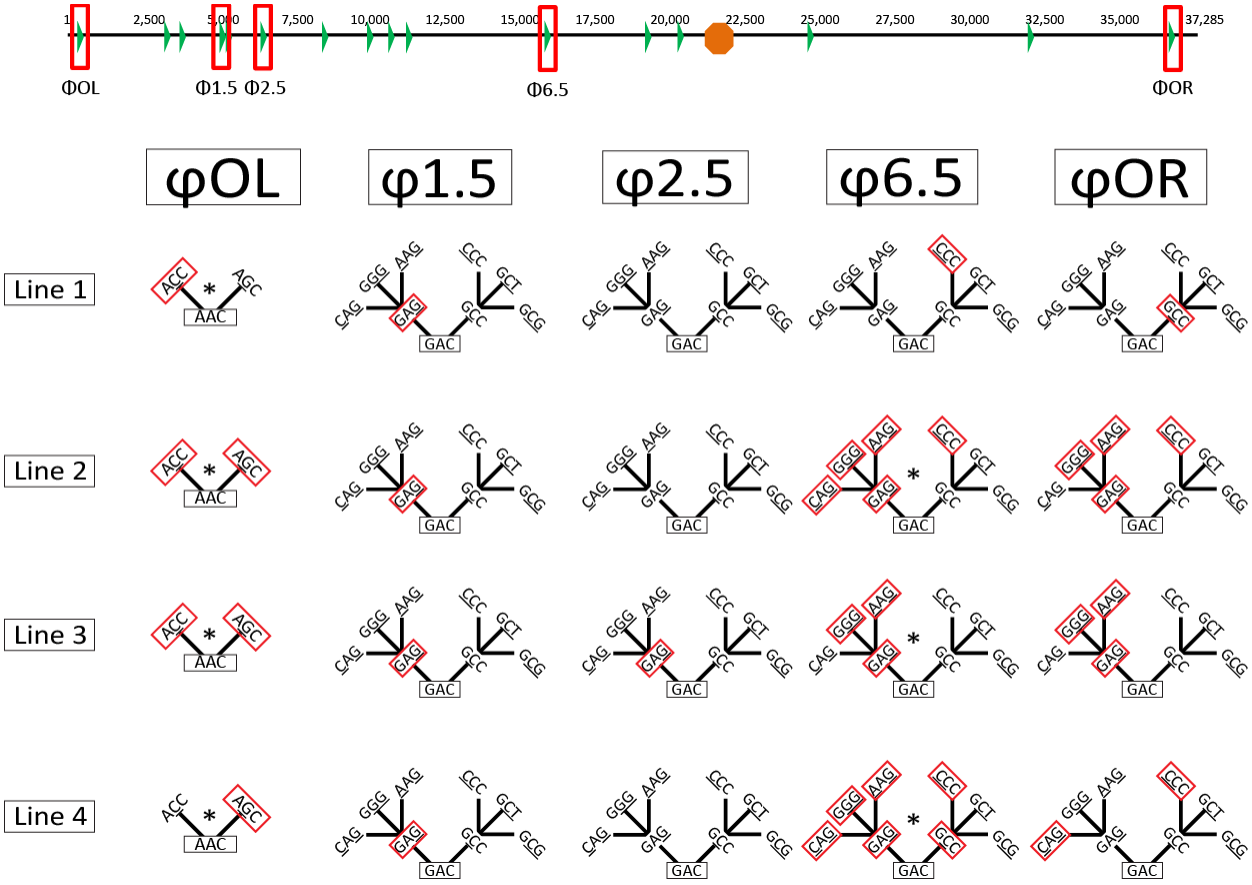


Figure 1.7: Above, map of the T7Δ1 genome with the 16 T7 promoters as green triangles and the T7 terminator as orange octagon. The five promoters mutated during the course of the passaging are shown in red boxes. Below, the five promoters in each of the four lines showing the mutant promoters (red boxes) from the canonical promoters (black boxes) at passage 100. Duplications are identified with an asterisk (*)

Table 1.1: Promoter identities as they occurred during the passaging.

Promoter	WT sequence	1:50	1:100	2:50	2:100	3:50	3:100	4:50	4:100
ϕOL	AAC		ACC		ACC, AGC		ACC, AGC	AGC	AGC
ϕ1.5	GAC	GAG	GAG	GAG	GAG	GAG	GAG	GAG	GAG
ϕ2.5	GAC					GAG	GAG		
ϕ6.5	GAC	CCC	CCC	GAG, CCC, GCC	GAG, GGG, AAG, CCC,CAG	GAG	GAG, GGG, AAG	GAG, GCC, GAC	GAG, GGG, AAG
ϕOR	GAC	GCC	GCC	GCC	GAG, GGG, AAG , CCC	GCC	GAG, GGG, AAG	GCC	GGG, CAG, CCC

Changed promoter
No evidence

Table 1.1: Promoter identities as they occurred during the passaging.

Figure 1.8: Predicted vs. actual promoter changes in the passaging

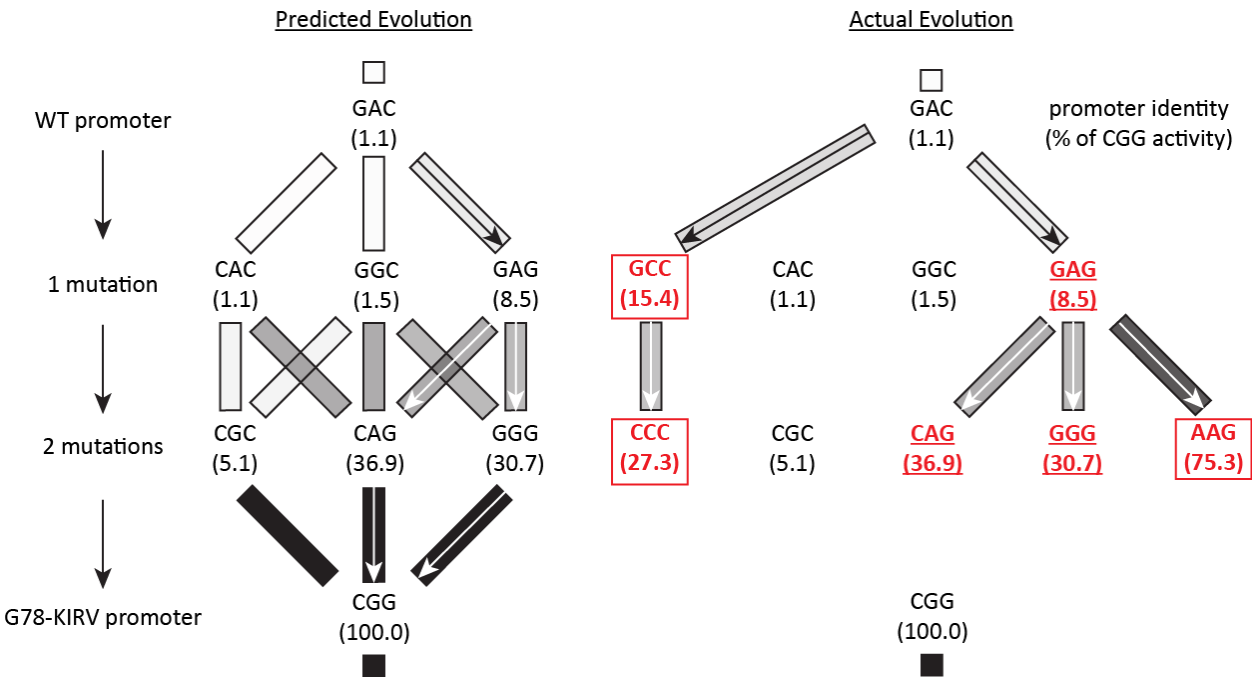


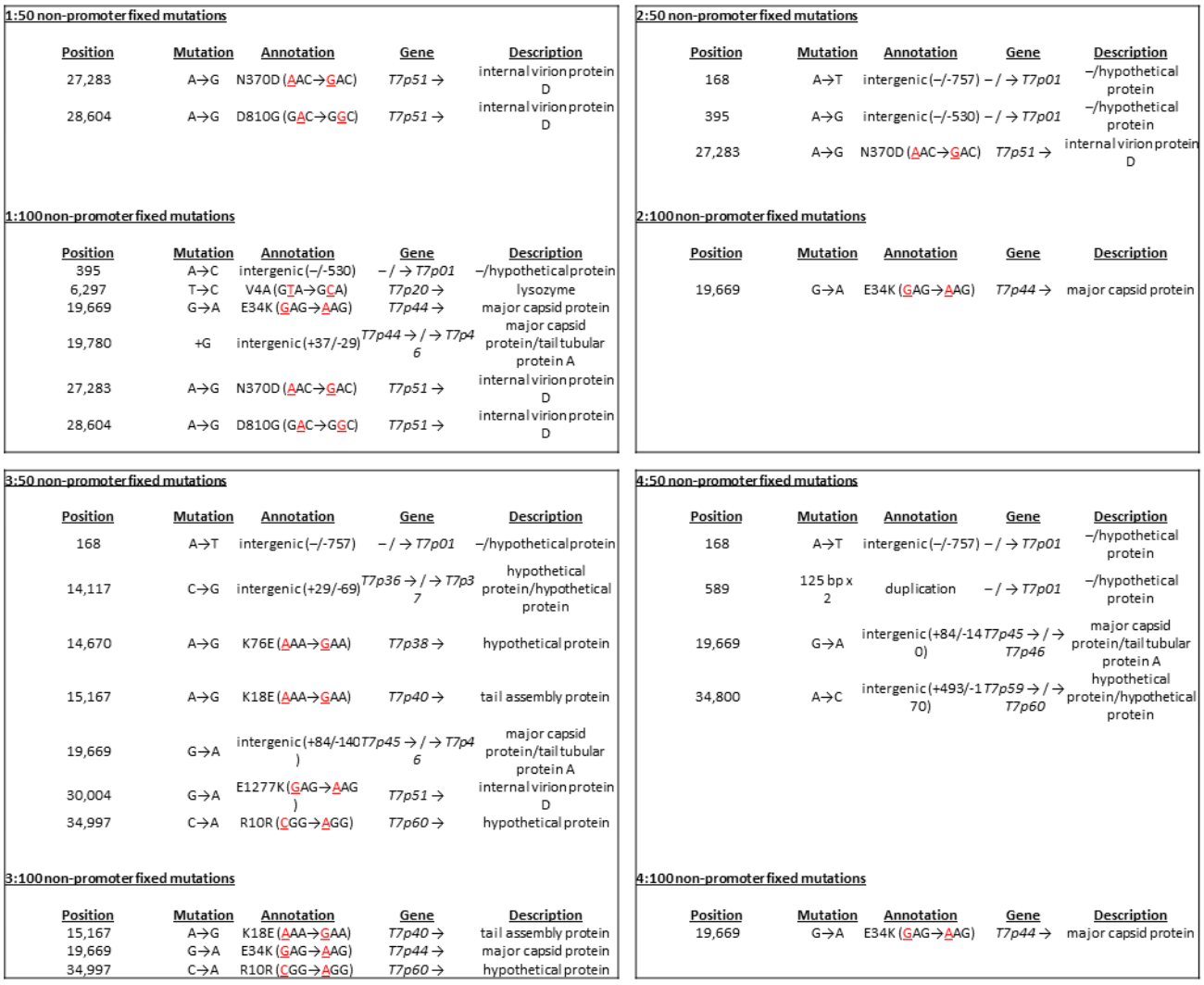
Figure 1.8: Predicted (on left, as before) and actual mutations (on right) found during the passaging. Actual paths (red promoters) on right are shown with arrows and pathway boxes are shaded according to activity. Activity as a percentage of G78-KIRV T7 RNAP on CGG promoter is displayed in parentheses below each triplet (Figure 1.1 and Figure 1.9). Those promoters found that matched the prediction are underlined whereas those unanticipated by the prediction are boxed.

Table 1.2: Promoter mutations and duplications as they occurred during the passaging.

Promoter	Location	1:50	1:100	2:50	2:100	3:50	3:100	4:50	4:100
φOL	CAACT		A→C in duplication (136bp, 94.6%)		A→C,G in duplication (220bp, 94.8%)		A→C,G in duplication (168bp, 91.3%)	A→G (38.8%)	A→G (3.2%), duplication, (136bp, 89.3%)
E. coli A2	→duplication							duplication (125bp)	duplication (57.9%)
φ1.5	GACT	C→G	C→G	C→G	C→G	C→G	C→G	C→G	C→G
φ2.5	GACT					C→G (46.6%)	C→G (25.6%)		
φ6.5	GACT	GA→CC	GA→CC	A→C, G→C (44.6%), C→G (47.9%)	G→A (71.4%), A→G (17.3%), C→G (93.0%)	C→G	G→A, C→G (97.8%)	A→C (13.6%), C→G (78.0%), in duplication (228bp)	G→A (16.7%), A→G (48.8%), C→G (99.0%), duplication (228bp)
φOR	GACT	A→G	A→C	A→C	A→C (16.3%), G→C (17.2%), C→G (89.6%)	A→C	C→G, A→G (56.8%), G→A (34.1%)	A→C	A→G (40.9%), C→G (89.6%)

Table 1.2: Promoter mutations and duplications as they occurred during the passaging. No promoter changes are shown in white boxes, partial percentages are shown in grey boxes, and 100% fixed promoter changes are shown in black boxes.

Table 1.3: Breseq output of all the non-promoter fixed mutations in each of the four populations at passage 50 and 100.



1:50 non-promoter fixed mutations

Position	Mutation	Annotation	Gene	Description
27,283	A→G	N370D (AAC→GAC)	*T7p51* →	internal virion protein D
28,604	A→G	D810G (GAC→GGC)	*T7p51* →	internal virion protein D

1:100 non-promoter fixed mutations

Position	Mutation	Annotation	Gene	Description
395	A→C	intergenic (–/-530)	– / → *T7p01*	–/hypothetical protein
6,297	T→C	V4A (GTA→GCA)	*T7p20* →	lysozyme
19,669	G→A	E34K (GAG→AAG)	*T7p44* →	major capsid protein
19,780	+G	intergenic (+37/-29)	*T7p44* → / → *T7p46*	major capsid protein/tail tubular protein A
27,283	A→G	N370D (AAC→GAC)	*T7p51* →	internal virion protein D
28,604	A→G	D810G (GAC→GGC)	*T7p51* →	internal virion protein D

2:50 non-promoter fixed mutations

Position	Mutation	Annotation	Gene	Description
168	A→T	intergenic (–/-757)	– / → *T7p01*	–/hypothetical protein
395	A→G	intergenic (–/-530)	– / → *T7p01*	–/hypothetical protein
27,283	A→G	N370D (AAC→GAC)	*T7p51* →	internal virion protein D

2:100 non-promoter fixed mutations

Position	Mutation	Annotation	Gene	Description
19,669	G→A	E34K (GAG→AAG)	*T7p44* →	major capsid protein

3:50 non-promoter fixed mutations

Position	Mutation	Annotation	Gene	Description
168	A→T	intergenic (–/-757)	– / → *T7p01*	–/hypothetical protein
14,117	C→G	intergenic (+29/-69)	*T7p36* → / → *T7p37*	hypothetical protein/hypothetical protein
14,670	A→G	K76E (AAA→GAA)	*T7p38* →	hypothetical protein
15,167	A→G	K18E (AAA→GAA)	*T7p40* →	tail assembly protein
19,669	G→A	intergenic (+84/-140)	*T7p45* → / → *T7p46*	major capsid protein/tail tubular protein A
30,004	G→A	E1277K (GAG→AAG)	*T7p51* →	internal virion protein D
34,997	C→A	R10R (CGG→AGG)	*T7p60* →	hypothetical protein

3:100 non-promoter fixed mutations

Position	Mutation	Annotation	Gene	Description
15,167	A→G	K18E (AAA→GAA)	*T7p40* →	tail assembly protein
19,669	G→A	E34K (GAG→AAG)	*T7p44* →	major capsid protein
34,997	C→A	R10R (CGG→AGG)	*T7p60* →	hypothetical protein

4:50 non-promoter fixed mutations

Position	Mutation	Annotation	Gene	Description
168	A→T	intergenic (–/-757)	– / → *T7p01*	–/hypothetical protein
589	125 bp x 2	duplication	– / → *T7p01*	–/hypothetical protein
19,669	G→A	intergenic (+84/-140)	*T7p45* → / → *T7p46*	major capsid protein/tail tubular protein A
34,800	A→C	intergenic (+493/-170)	*T7p59* → / → *T7p60*	hypothetical protein/hypothetical protein

4:100 non-promoter fixed mutations

Position	Mutation	Annotation	Gene	Description
19,669	G→A	E34K (GAG→AAG)	*T7p44* →	major capsid protein

Table 1.3: Breseq output of all the non-promoter fixed mutations in each of the four populations at passage 50 and 100. All four lines picked up the E34K mutation in their major capsid protein by passage 100.

Figure 1.9: *In-vitro* transcription activity of wild-type and G78-KIRV RNA polymerases

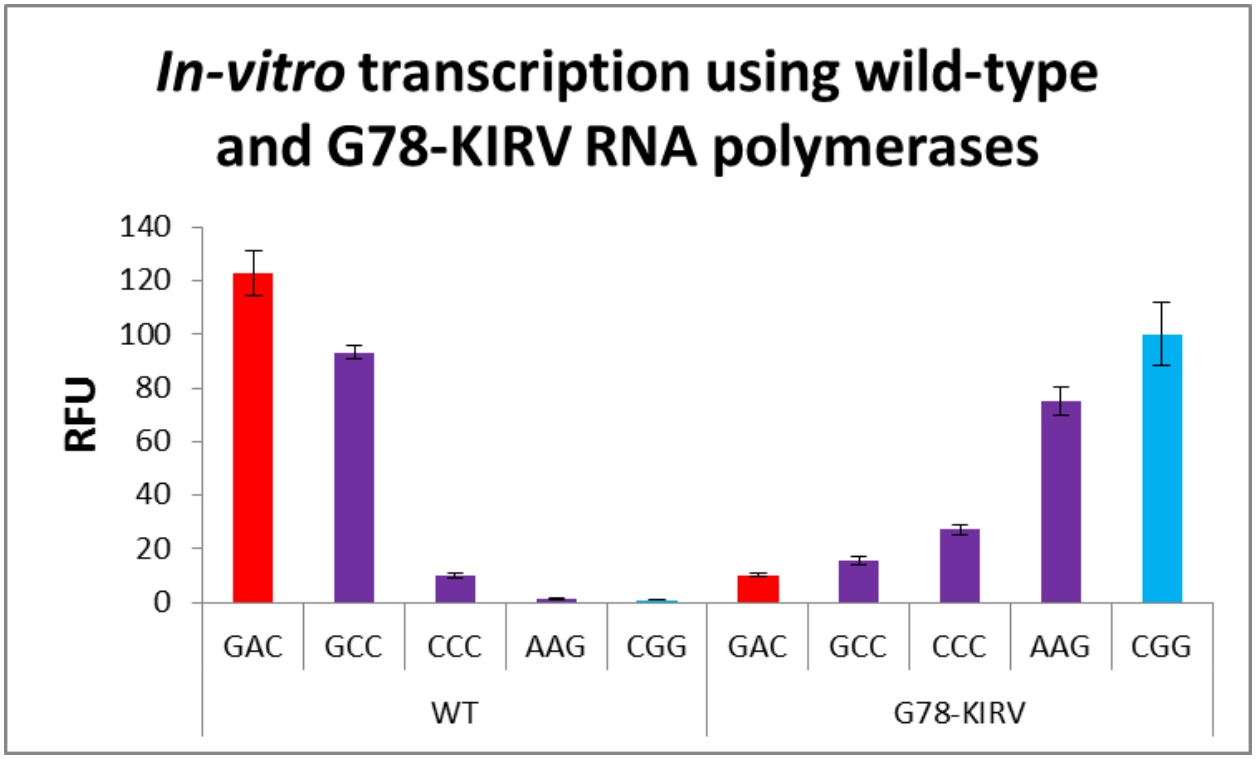


Figure 1.9: *In-vitro* transcription activity of wild-type and G78-KIRV RNA polymerases on wild-type (GAC, in red) and heterologous G78-KIRV promoter (CGG, in blue) as well as additional promoters found in the phage evolution (CCC, AAG, and GCC, in purple). All values normalized to G78-KIRV RNA polymerase activity on G78-KIRV (CGG) promoter defined as 100.

Figure 1.10: Visualization of various sequence spaces.

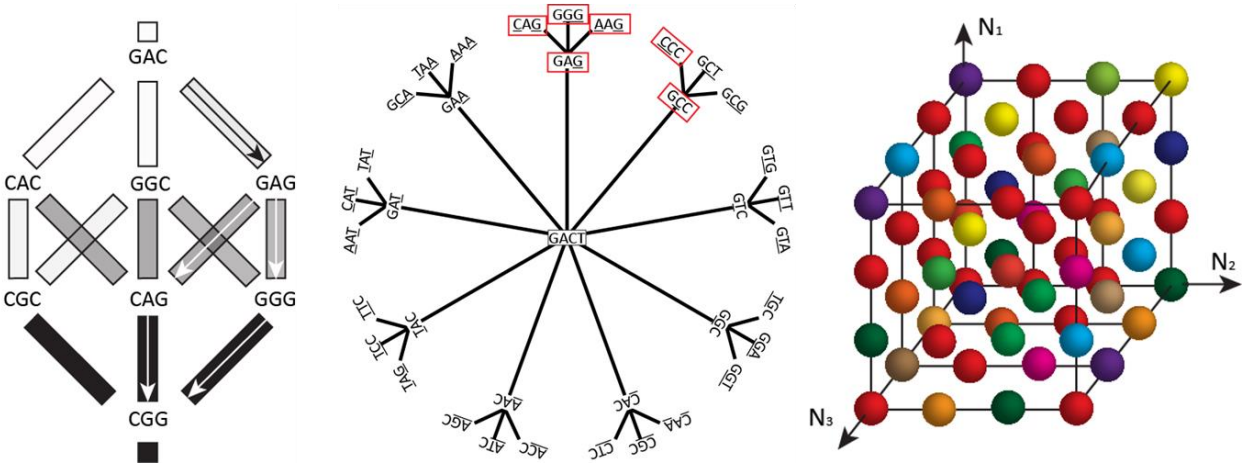


Figure 1.10: Visualization of various sequence spaces. (Left) The single-step map showing the six intermediates between the wild-type and G78-KIRV promoter. (Middle) a representation of all of the 36 promoters within two mutations of the wild-type promoter (those found in the experiment boxed in red). (Right) Representation of the totality of possible sequences in -11 to -9 region of the T7 promoter (N=64). Each axis includes four positions, one for each of the four nucleotides. Colors are used to articulate theoretical differential promoter activity.

Table 1.4: Primers used for *in-vitro* transcription

DJG.GACT.SpMX1.F	AATATAATACGACTCACTATAGAGGAGACTGAAATGGTGAAGGACGGGTCCAGTGCTTCG
DJG.CACT.SpMX1.F	AATATAATACCACTCACTATAGAGGAGACTGAAATGGTGAAGGACGGGTCCAGTGCTTCG
DJG.GGGT.SpMX1.F	AATATAATACGGGTCACTATAGAGGAGACTGAAATGGTGAAGGACGGGTCCAGTGCTTCG
DJG.CAGT.SpMX1.F	AATATAATACCAGTCACTATAGAGGAGACTGAAATGGTGAAGGACGGGTCCAGTGCTTCG
DJG.CGCT.SpMX1.F	AATATAATACCGCTCACTATAGAGGAGACTGAAATGGTGAAGGACGGGTCCAGTGCTTCG
DJG.GAGT.SpMX1.F	AATATAATACGAGTCACTATAGAGGAGACTGAAATGGTGAAGGACGGGTCCAGTGCTTCG
DJG.GGCT.SpMX1.F	AATATAATACGGCTCACTATAGAGGAGACTGAAATGGTGAAGGACGGGTCCAGTGCTTCG
DJG.CGGT.SpMX1.F	AATATAATACCGGTCACTATAGAGGAGACTGAAATGGTGAAGGACGGGTCCAGTGCTTCG
DJG.CCCT.SpMX1.F	AATATAATACCCCTCACTATAGAGGAGACTGAAATGGTGAAGGACGGGTCCAGTGCTTCG
DJG.AAGT.SpMX1.F	AATATAATACAAGTCACTATAGAGGAGACTGAAATGGTGAAGGACGGGTCCAGTGCTTCG
DJG.GCCT.SpMX1.F	AATATAATACGCCTCACTATAGAGGAGACTGAAATGGTGAAGGACGGGTCCAGTGCTTCG
DJG.SpMX1.R	GAAAAGACTAGTTACGGAGCTCACACTCTACTCAACAGTGCCGAAGCACTGGACCCG

Table 1.4: Primers used for *in-vitro* transcription (See Materials and Methods).

Chapter 2: Evolution of the Transcriptional Apparatus in an Evolutionarily Polarized Bacteriophage Population

Abstract

With the emergence of synthetic biology, it remains unclear how far synthetic parts (with altered activity and substrate specificity) can alter the evolutionary trajectory of a host organism. Previously, we had evolved bacteriophage T7 populations (T7Δ1) on a synthetically-evolved T7 RNA polymerase (G78-KIRV) leading to recovery of fitness. Here, we further interrogated these evolutionarily polarized populations by recombining either the wild-type or the synthetic RNA polymerase into the bacteriophage genomes and allowing further adaption to take place in order to discover more about the evolvability of the transcriptional apparatus writ large. The evolved strains were able to accommodate insertion of the synthetic polymerase back into their genomes, whereas the ancestor T7Δ1 was not. Furthermore, the bacteriophage populations subsequently altered the transcriptional apparatus such that the synthetic polymerase was further mutated (rather than promoters) to create two distinct types of RNA polymerase generalists. This work is an example of how evolutionary optimization is both essential for simple design implementation in even the most characterized of biological systems and powerful in enabling evolution to discover unexpected solutions. Additionally, our work raises the question whether synthetic biology and directed evolution can be used for laboratory speciation events and the development of orthogonal viruses.

Introduction

Laboratory-based evolution has long been a tool to study how evolution works in well-defined systems. Laboratory-based evolution studies have been applied to a wide-

variety of applications such as studying the utilization of a new carbon source, replaying the evolutionary "tape-of-life" in order to interrogate evolutionary stochasticity, and even inviting the adoption of unnatural amino acids into essential proteins (Blount, et al. 2012; Dickinson, et al. 2013; Hammerling, et al. 2014). These studies allow empirically-derived conclusions to drive understanding of fundamental aspects of evolution. In parallel, what is now known as synthetic biology has inherited the goal passed on by genetic engineering of being able to design, engineer, and control biological processes. Unlike genetic engineering, however, synthetic biology hopes to achieve this goal of design, engineering, and control of biological processes via standardization of design processes, assays, and genetic parts (Galdzicki, et al. 2014; Gardner and Hawkins 2013). This process of standardization has led to a demand for well-characterized genetic parts in which more complicated genetic circuits can be faithfully and reliably built in a variety of organisms. These parts, acquired by both part mining the genetic treasure trove of genomic sequences or a result of painstaking laboratory manipulation and selection, are desirous for their altered activity in various reaction conditions or for their altered substrate specificity.

One example from our lab of such a well-characterized genetic part was the creation of a synthetically-evolved T7 RNA polymerase (Ellefson, et al. 2014). This synthetically-evolved T7 RNA polymerase, named G78-KIRV RNA polymerase (G78-KIRV RNAP), was the product of a directed evolution scheme which shifted the promoter specificity of the wild-type T7 RNAP away from the wild-type sequence (GAC, in the -11 to -9 promoter specificity region) to CGG. In doing so, the G78-KIRV RNAP: CGG promoter pair became functionally orthogonal to the wild-type RNAP:GAC promoter pair in that each polymerase is uniquely active on its own promoter and inactive on the other promoter. These types of genetic parts are being used to create complex

genetic circuits in a variety of organisms (Brophy and Voigt 2014; Segall-Shapiro, et al. 2014). These parts and circuits often operate outside of the biological contexts from which they originated and are thus rarely used to interrogate the evolution of whatever host genome provided the initial genetic part. Unfortunately, the lack of understanding of even the least complex and most modeled of organisms such as bacteriophages T7 and M13 make it difficult to predict exactly how any synthetically-evolved part will evolutionarily behave in any biological system (Kosuri, et al. 2007; Smeal, et al. 2017a, b). These same difficulties also plague circuits in general in that genetic stability and immutability are far from predictably designable or guaranteed. We realized the orthogonality of G78-KIRV T7 RNAP provided an excellent and unique opportunity to study the evolution of the bacteriophage T7 system from which it was originally derived within the context of a laboratory evolution experiment.

In this study, we used the bacteriophage T7Δ1 populations previously adapted, or evolutionarily polarized, towards using the G78-KIRV RNAP in order to interrogate the totality of the T7 transcriptional apparatus by reinserting back, then further evolving, either the wild-type or the G78-KIRV RNAP into and together with the rest of the T7Δ1 genomes. In doing so, we investigate interesting evolutionary questions such as how will the T7 transcriptional apparatus adapt and at what point has an organism, having gone through a period of adaption to a synthetically-evolved part, adapted enough such that it can be considered a distinct species. As such, we show how the use of laboratory evolutionary adaptation can complement rational design strategies while leading to interesting and unpredicted outcomes.

Results

Reinsertion of wild-type and G78-KIRV RNA polymerases into T7Δ1 genomes

In the previous chapter, we adapted T7Δ1, a bacteriophage T7 lacking its RNAP, for 100 passages on the synthetically-evolved G78-KIRV T7 RNAP, leading to the recovery of overall fitness (Figure 2.1A). Four lines evolved in parallel reached nearly the fitness of wild-type T7 phage. However, the provision of the polymerase in *trans* (i.e. from a plasmid) was obviously an unnatural situation, and wished to determine whether and how the four evolved strains would accommodate reinsertion of either the wild-type T7 RNAP or the orthogonal G78-KIRV RNAP back into the phage (Figure 2.1B).

To accomplish the reinsertion of either the wild-type or altered-specific RNAP into the evolved phage, plasmids were created in which the RNAP genes were flanked with 50-100 base pairs of homology to the phage genome, corresponding to the left and right flanking region around the Δ1 site. By passing the evolved strains in the presence of the plasmid, a small percentage should undergo homologous recombination and thereby allow further viral replication on BL21 *E. coli* cells that lack plasmids that express T7 RNAP.

We successfully recombined both the parental wild-type and the G78-KIRV RNAP into each of the genomes of the four evolved strains and then continued with another 50 passages on BL21 *E. coli* cells. All four of the evolved strains were able to successfully integrate both the wild-type and to G78-KIRV RNAP, respectively, into their genomes. Interestingly, even after repeated efforts, attempts to introduce G78-KIRV RNAP into the wild-type T7Δ1 progenitor would not produce viable phage.

Promoter changes after reinsertion of the wild-type or G78-KIRV RNA polymerase into the bacteriophage genome

Following passaging, the phage populations were sequenced via MiSeq 2x250bp pair end reads, $3.6x10^5$ reads per sample. When the G78-KIRV RNAP was reinserted into the four evolved lines, all of the mutated promoters from the initial evolutionary

adaptation were maintained (Figure 2.2). Interestingly, line 1 lost its ϕOL duplication while gaining a ϕ6.5 duplication, and line 2 lost its ϕ6.5 duplication (Figure 2.2). In contrast, when the wild-type RNAP was reinserted into the lines, five of the mutated promoters reverted back to the wild-type sequence: the ϕ1.5 promoter in line 3, the ϕ6.5 promoter in line 1, and the ϕOR promoter in lines 2, 3, and 4 (Figure 2.2). A number of duplications that arose during the original directed evolution experiment were lost: lines 1 and 3 lost ϕOL duplications and lines 2 and 3 lost ϕ6.5 duplications (Figure 2.2).

RNA polymerase stability after reinsertion and adaptation

Once strains previously adapted to G78-KIRV RNAP had a chance to allow the polymerase sequence to change, two lines (1 and 2) incorporated an E756K mutation in the specificity loop (Table 2.1), while the other two lines (3 and 4) picked up an E222A mutation (100% fixed in line 3, 70% E222A/V in line 4, Table 2.1). The E222A substitution disrupts a salt bridge between E222 and R84, which appears to be otherwise unaltered as the RNAP goes through promoter initiation and elongation (Figure 2.3) (Cheetham and Steitz 1999; Tahirov, et al. 2002). Of the four lines that incorporated the wild-type T7 RNAP, no changes in the polymerase were observed (Table 2.1).

In-vitro transcriptions with wild-type, G78-KIRV, G78-KIRV-E222A, and G78-KIRV-E756K RNA polymerases on the unanticipated promoters found in the evolution

The occurrence of the E222A or E756K mutations to the G78-KIRV RNAP together with the relative stasis of the promoters within the G78-KIRV-inserted lines justified further characterization of the G78-KIRV-E222A and G78-KIRV-E756K RNAPs. Throughout the initial 100 passages of adaption to G78-KIRV in *trans*, the anticipated GAG, CAG, and GGG as well as the unanticipated GCC, CCC, and AAG were the only promoter mutants to occur in more than one promoter in more than one strain, thus justifying further *in-vitro* characterization with the wild-type, G78-KIRV,

G78-KIRV-E222A, and G78-KIRV-E756K RNAP variants. We chose the GCC, CCC, and AAG promoters to use for our *in-vitro* transcription activity assays because they were found in the G78-KIRV-phage populations but lost or diminished in the wild-type populations (Table 2.2) and because they, together with the wild-type (GAC) and cognate G78-KIRV promoter (CGG) provide relatively step-wise activity when transcribed by G78-KIRV RNAP (Figure 2.4). In the same manner as before, we purified the wild-type, G78-KIRV, G78-KIRV-E222A, and G78-KIRV-E756K RNAP mutants and determined their relative activity on the wild-type (GAC), G78-KIRV (CGG), and three unanticipated promoters (GCC, CCC, and AAG) (Figure 2.4).

These data show that both the G78-KIRV-E222A and the G78-KIRV-E756K RNAP mutants found after passage in two each of the four lines after reinsertion of the G78-KIRV RNAP are generalists which have lost the specificity of the original G78-KIRV RNAP. The E222A mutation, in spite of not existing in the RNAP specificity loop, seems to unexpectedly imbue the G78-KIRV with hyperactivity at or above wild-type activity on all of the five promoters tested. Additionally, these data show that the G78-KIRV-E756K mutant is also a generalist with near-wild-type activity levels on the wild-type promoter as well as a much higher activity on the orthogonal promoter (CGG) than that of wild-type RNAP. The G78-KIRV-E756K RNAP is also very active on the three unanticipated promoters.

In-vivo activity assays confirm generalist activity of G78-KIRV-E222A and G78-KIRV-E756K RNAP mutants

We used a simple, anhydrous-tetracycline-inducible T7-promoter driven-GFP-production circuit (Figures 2.5 and 2.6) in order to better understand and characterize the *in-vivo* behavior of the G78-KIRV-E222A and G78-KIRV-E756K RNAP mutants (Figure 2.7). We chose to use the wild-type T7 RNAP promoter (GAC, in the -11 to -9

promoter specificity region) and also, in contrast, a GGC variant because it provides relatively low *in-vitro* activity when used with the G78-KIRV RNAP, as seen in Figure 1.1B in previous chapter. The wild-type and G78-KIRV RNAPs behaved as expected: the wild-type RNAP was extremely active on the wild-type promoter and moderately active on the GGC promoter whereas the G78-KIRV RNAP was had extremely low activity on either promoter (Figure 2.7). Interestingly, the G78-KIRV-E222A and G78-KIRV-E756K RNAPs did indeed operate as generalists, both recovering activity on both promoters relative to G78-KIRV RNAP activity, with the G78-KIRV-E222A variant actually having higher activity on the GGC promoter versus the wild-type promoter.

Of note, however, is that the G78-KIRV-E222A mutant, which had extremely high activity on all the promoters tested in the *in-vitro* activity assay, had a fraction of the *in-vivo* activity relative to the wild-type or G78-KIRV-E756K RNAPs. We presume that this is the result of spurious transcription by the G78-KIRV-E756K of the *E. coli* genome. We decided to test the E222A mutation independently of those in G78-KIRV (that is, the E222A mutation in the wild-type RNAP) and it indeed was enough to give near-wild-type promoter level of *in-vivo* activity on the GGC promoter, thus confirming the generalizability-enabling of the E222A mutation.

Discussion

The synthetically-derived G78-KIRV T7 RNAP and our previous evolutionary polarization of the T7Δ1 bacteriophage populations provided a unique opportunity to enforce an evolutionary bifurcation by respective RNAP reinsertion. In doing so, we hypothesized that placing these bacteriophages in these unique, parallel, and heretofore unexplored sequence space would lead to insights into the nature of transcriptional regulation and evolution in general.

The first insight unexpectedly came from the concurrence of: 1) the successful reinsertion of G78-KIRV RNAP into the evolutionary polarized populations with 2) the failed reinsertion of G78-KIRV RNAP back into the ancestor T7Δ1. This strongly suggests that the evolutionary adaptation through the serial passaging was essential to G78-KIRV RNAP being stably incorporated into the T7Δ1 phage genome. The ability to reinsert the wild-type T7 RNAP back into the evolved lines, however, was not unexpected: indeed in the initial evolutionary adaptation, only 4 or 5 of 16 promoters had mutated away from the wild-type sequence, leaving 10 or 11 of 16 wild-type promoters, presumably with the same transcriptional capacity as they normally would have in the ancestor T7Δ1 phage genome.

The next series of insights came from the mutations which occurred after the reinsertion and adaption. Unlike our previous study, here we had no way to predict the nature by which mutations would occur in the promoter identities, abundances, or duplications, much less any mutations in the RNAPs. Phylogenetic analysis of T7 group phages suggests that, while T7-like promoters co-evolve with their respective RNAPs, they evolve at different rates (Chen and Schneider 2005). It has been hypothesized that as phages begin to diverge polymerases, there is a strong selective pressure on the entirety of the phage transcriptional apparatus (that is, both the polymerase and the promoters) to become distinct (Chen and Schneider 2005). Providing either the synthetic or wild-type RNAPs into these phage populations with a variety of promoter identities allowed us to experimentally rather than bioinformatically explore the relationship between promoter and RNAP co-evolution.

After reinsertion of the wild-type RNAP, four otherwise fixed mutant promoters were reverted back to the canonical wild-type sequence (Figure 2.2, Table 2.2). Beyond these changes, the promoter identities stayed rather consistent as to what they were at the

passage 100 adaptation. These wild-type RNA-containing phages did, however, in four instances lose initial promoter duplications, likely reflecting their redundancy within the context of the wild-type RNAP. In the four lines passaged after reinsertion of the G78-KIRV RNAP, there was no evidence of additional promoter mutation. Unlike those with the wild-type insertion, these lines did, however, maintain those mutated promoters that were there initially. Also in these lines, the pattern of changes in the duplications is a bit more muddled: two lines lost duplications and one line gained a promoter duplication.

The pattern of promoter changes is only one half of the transcriptional system: the other half is the T7 RNAP itself. There were no mutations in the wild-type RNAP after it was reinserted and passaged beyond those necessary to allow viable RNAP expression (Table 2.1, and Figure 2.8). The G78-KIRV RNAP was, however, found to be mutated in all four lines (Table 2.1). Lines 1 and 2 had fixed in their populations an E756K mutation. This mutation, which occurs in the specificity loop of the T7 RNAP, reverts back one of the key mutations in G78-KIRV from a negatively charged glutamic acid to a positively charged arginine, which is more chemically similar to the wild-type residue lysine (Ellefson, et al. 2014). We noticed that an E222A mutant had fixed in line 3 in addition to an E222A/V mixed population reaching 70% of the population in line 4. This E222 residue, though far from the specificity loop, had previously been observed to enable an expanded range of T7 promoter-like sequences (Figure 2.3) (Ikeda, et al. 1993). We suspected that both the E756K as well as the E222A to the G78-KIRV RNAPs by the phage was enabling generalist polymerase functionality.

In order to determine the exact nature of the E756K and E222A mutations in our phage populations we decided to do further *in-vitro* and *in-vivo* characterizations. *In-vitro* activity assays clearly show that both the G78-KIRV-E756K and the G78-KIRV-E222A RNAPs are generalists, with G78-KIRV-E222A in particular with at least 80% of wild-

type level activity on all promoters tested (Figure 2.4). The *in-vivo* activity assays confirm that, the G78-KIRV-E756K and G78-KIRV-222A RNAPs enable activity on promoters that G78-KIRV alone is otherwise inactive (Figure 2.7). We decided to add the E222A mutation back into the wild-type RNAP (creating wild-type-E222A, Figure 2.7) which increased *in-vivo* activity on the GGC promoter, thus suggesting that E222A imbues T7 RNA polymerase with generalist activity and is extensible.

Conclusion

Overall, the bacteriophage populations altered the transcriptional apparatus such that the synthetic polymerase (rather than promoters) picked up subsequent mutations. Interestingly, both the further additions of either E756K or E222A mutations to G78-KIRV resulted in the same outcome: the polymerases because generalists. This unexpected sequence divergence with phenotypic convergence is yet another example of the stochastic nature of evolutionary trajectories. This work is an example of how evolutionary optimization was essential for simple design implementation in even the most characterized of biological systems in that successful recombination of the G78-KIRV was predicated on the previous adaptation to make the evolved populations. Furthermore, this work is an example of the power in enabling evolution to discover unexpected solutions: the creation of two new and distinct RNAP variants may be difficult to create outside of evolutionary-based schemes. Finally, our work raises the question of whether synthetic biology and directed evolution can be used for laboratory speciation events and the development of orthogonal viruses. A phage, having stably maintained the G78-KIRV RNAP while further mutating its promoters, could be considered its own species if it was unable to viably recombine, similar to bacteriophages T7 and T3. Future efforts into large-scale genome assembly and the current existence of

no less than six orthogonal T7 RNAPs may facilitate this laboratory speciation and the creation of orthogonal viruses (Meyer, et al. 2015).

Materials and Methods

Recombination of RNA polymerases back into T7Δ1 phage

To reinsert either the wild-type or G78-KIRV RNAP back into the evolved lineages plasmids were created with approximately 50-100 base pairs flanking on each side of the RNAP corresponding to the left and right side of the T7Δ1 phage genome (see Figure 2.8). These plasmids were transformed into BL21-DE3 cells. Previously evolved T7Δ1 lines from the previous chapter were used to lyse these 10mLs of these cells induced with 1mM IPTG to allow phage propagation. To enrich for those which had successfully recombined either the wild-type or G78-KIRV RNAP into the phage genome all 10mLs of lysate was taken and added to 10mLs of BL21-Gold *E. coli* cells (Agilent) grown to an OD of 1 (approximately 10^8 cells/mL). Successive passaging of 10mLs from the previous flask into 10mLs of BL21-Gold cells eventually led to full lysis of the BL21-Gold cells, at which point 1uL (approximately 10^{10}-10^{11} cfu/mL) of lysate was passaged.

Phage passaging

After recombination, 1uL of complete phage lysis (approximately 10^7 phages) were added into 10mL BL21-Gold *E. coli* cells (Agilent) at a density of approximately 108 cells/mL in 125mL Erlenmeyer flasks, which had previously been diluted from O/N stocks in 2xYT media and allowed to grow for one hour at 37° C. Flasks were shook at 200 RPM at 37° C until complete lysis at which point 1mL of lysate was removed and mixed with 300uL of chloroform to halt any residual cell lysis and kill off any residual *E. coli* cells. At this point, 1uL of this lysate was then transferred to the next flask. Each cycle of 1uL transfer, infection, and lysis is considered one passage.

RNA polymerase purification

RNAP variants were purified using the same protocol as mentioned in the previous chapter. Briefly, IPTG-inducible pQE plasmids containing each respective RNA polymerase variant were transformed into BL21-Gold *E. coli* cells (Agilent) and plated to isolate single colonies. These colonies are grown up overnight in their proper antibiotic and used to seed 1L flasks of 2xYT media (though 100mL cultures are adequate for non-generalist RNA polymerases) which are grown at 37° C until reaching mid-log phase at which point 1mM of IPTG was added and the cells were grown O/N at 18°C then pelleted, frozen at -80° C O/N. Pellets were resuspended in the T7 buffer and subjected to Ni-NTA N-terminal 6x His column purification, as stated previously. The final, purified polymerases were each adjusted to 1mg/mL and stored at -20° C.

Sequencing and analysis

Similar to the methods in the previous chapter, 2x250 paired end read MiSeq runs with a coverage of at least 3.5x105 reads per phage were done and analyzed using breseq (Deatherage and Barrick 2014) as well as Geneious Version 7.1.9 (Kearse, et al. 2012) using the bacteriophage T7 genome (GenBank V01146.1) as reference.

In-vitro transcriptions

In-vitro transcriptions were carried out in a manner similar to elsewhere (Ellefson, et al. 2014). Briefly, 0.5uM DNA template (created from 2uM forward and reverse primer found in figure 2.9 after gel purification following primer extension) is transcribed in 10uL reactions using 0.5uM T7 RNAP in a 40mM Tris-HCl pH 7.0, 30mM MgCl2, 6mM spermadine, and 10mM DTT buffer with 6mM each NTP and 0.17mg/mL DFHBI. Reactions were done at 37° C and readings at 469nm/501nm (excitation/emission) were taken every minute for two hours in a Tecan Safire monochromator.

In-vivo GFP activity assays

pSK constructs containing each respective RNAP variant were transformed into BL21 *E. coli* cells after previous transformation with T7-promoter-driven GFP plasmids (pSK.201 or pSK.205, see Figures 2.5 and 2.6) and plated. Individual colonies were grown up O/N and diluted 1:100 into 1mL 2xYT media supplemented with 50ug/mL kanamycin and 50ug/mL ampicillin and grown for 3 hours. 50uLs of these cells were then diluted again into 50uL 2xYT, 50ug/mL kanamycin, 50ug/mL ampicillin, and anhydrous tetracycline (aTc) (to give final concentration of 2ng/mL or 20ng/mL) and put into clear Corning 96-well plates, covered with Breatheasy sealing cover, and read on a Tecan M200 plate reader which had a kinetic (orbital) shaking cycle of 2.5mm for 430 seconds and took both an absorbance reading at 600nm and fluorescence readings at 485nm/525nm (excitation/emission). All data uses biological triplicates.

Figures

Figure 2.1: Passaging scheme and potentiality diagram

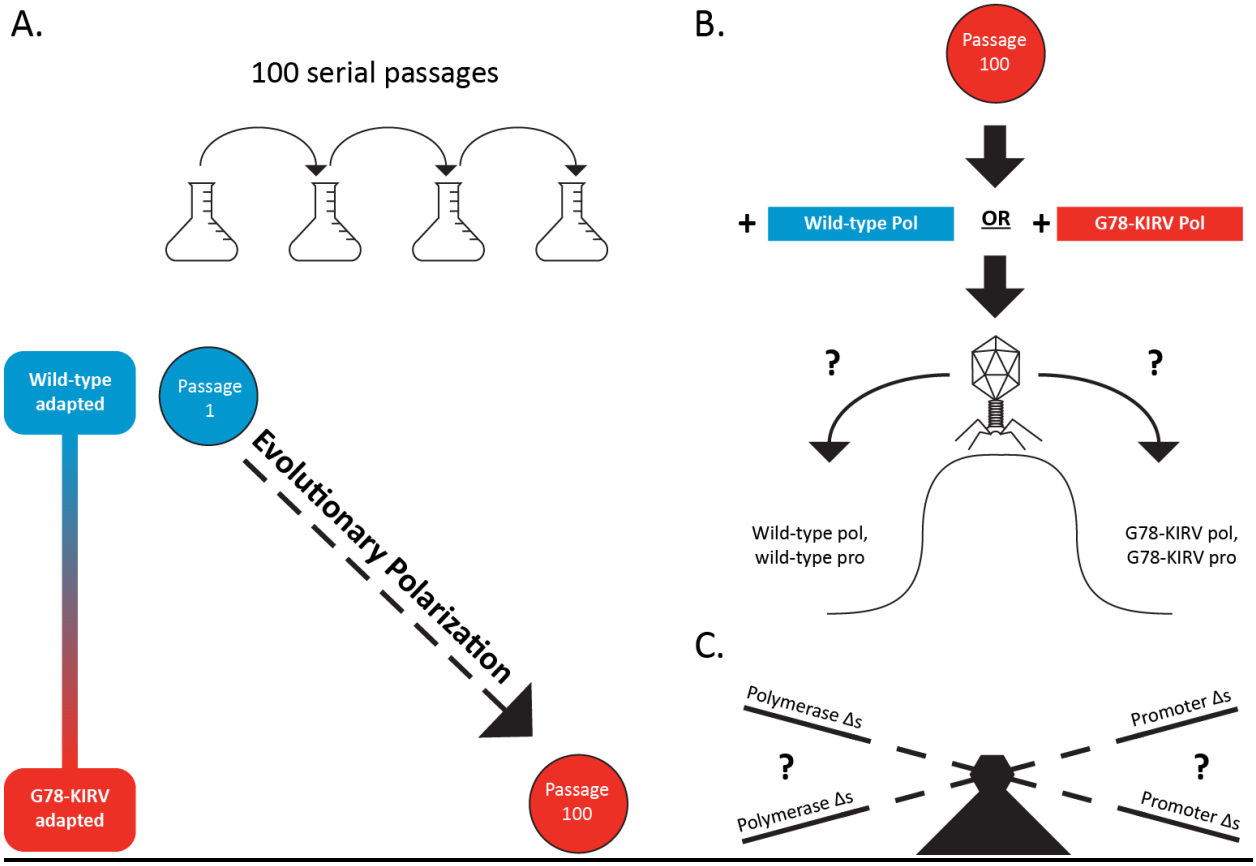
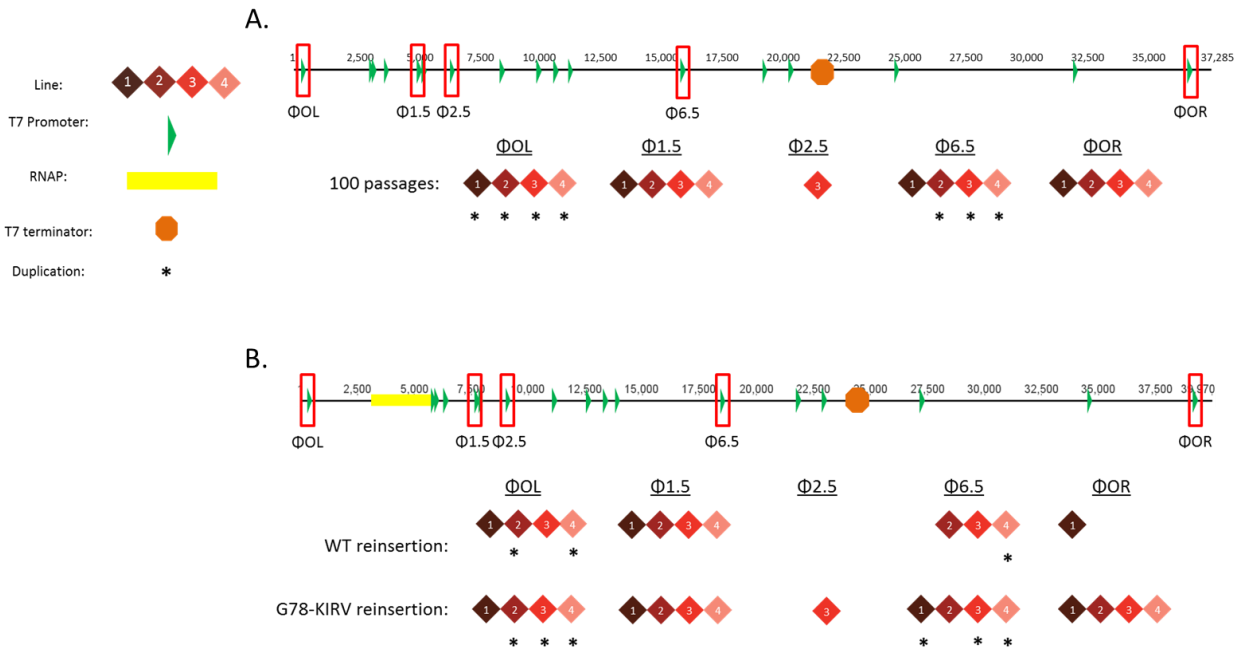


Figure 2.1: Passaging scheme of previous work and outline of this work. (A, previous work) Four T7Δ1 phage populations were serially passaged 100 times on cells expressing the G78-KIRV RNAP in *trans*. Over the course of these passages, four phage replicate phage populations each mutated at least four of their promoters while almost completely regaining fitness thereby polarizing those populations towards the G78-KIRV RNAP. (B, this work) Those four passage 100 phage populations either had the wild-type or the G78-KIRV RNAP, respectively, recombined into them and were passaged another 50 times. (Phage diagram) After recombination, the phage has the opportunity to continue to mutate its promoters in a potential evolutionary bifurcation event. (C, theoretical tradeoff

depiction) We hypothesized that there would be a mutually exclusive tradeoff in the evolution of transcriptional apparatus between further mutations in the polymerases or further mutations in the promoters.

Figure 2.2: Promoter changes after reinsertion and passaging of wild-type or G78-KIRV RNAP

Figure 2.2: Mapping location of all 16 T7 promoters in 40kB phage genome (green) in



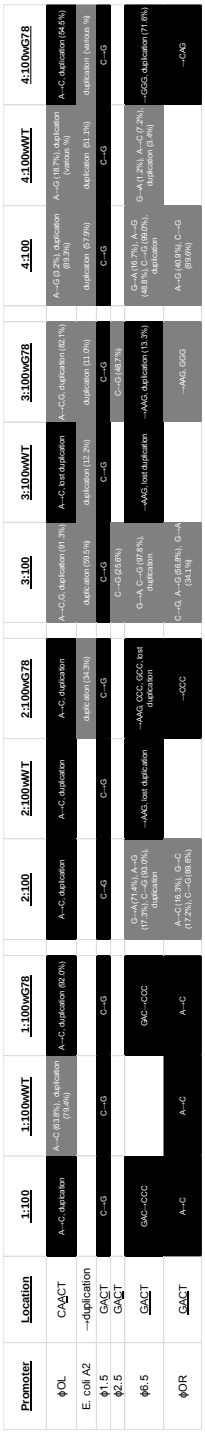
T7Δ1 (in A) and T7 with an RNAP in the Δ1 spot, colored yellow (in B). Mutated promoters are outlined in red boxes. Orange hexagon shows location of T7 terminator. In A, the promoters that changed in the four lines after 100 passages in each of the four lines (decreasing shades of red diamonds). In B, the promoters that changed in the four lines after 50 passages with either the wild-type (WT) or G78-KIRV RNAP were added back to the evolved (passage 100) strains. Asterisks in both represent the presence of promoter duplications in the sequencing data.

Table 2.1: RNA polymerase changes

Line	RNAP Inserted	Mutations Inside Specificity Loop	Mutations Outside Specificity Loop
1	Wild-type	∅	∅
	G78-KIRV	E756K	S4L (77.3%)
2	Wild-type	∅	∅
	G78-KIRV	E756K	S4L (30.7%)
3	Wild-type	∅	∅
	G78-KIRV	∅	S4L (27.9%), E222A
4	Wild-type	∅	∅
	G78-KIRV	∅	R2K (67.4%), M218R (27.8%), E222A/V (40.0%, 30.1%)

Table 2.1: Table displaying the mutations inside and outside the specificity loop in the wild-type or G78-KIRV RNAP inserted into each of the four lines after the 50 passages.

Table 2.2: Promoter mutations after reinsertion of each RNAP



Promoter	Location	1:100	1:100wWT	1:100wG78	2:100	2:100wWT	2:100wG78	3:100	3:100wWT	3:100wG78	4:100	4:100wWT	4:100wG78
ϕOL	CAACT	A→C, duplication	A→C (91.9%), duplication (79.4%)	A→C, duplication (92.0%)	A→C, duplication	A→C, duplication	A→C, duplication	A→CG, duplication (91.3%)	A→C, lost duplication	A→CG, duplication (82.1%)	A→CG (97.2%), duplication (69.3%)	A→G (18.7%), duplication (various %)	A→C, duplication (54.5%)
E. coli A2	→duplication						duplication (24.3%)	duplication (93.5%)	duplication (12.4%)	duplication (11.0%)	duplication (57.9%)	duplication (51.1%)	duplication (various %)
ϕ1.5	GACT	C→G	C→G	C→G	C→G	C→G	C→G	C→G	C→G	C→G	C→G	C→G	C→G
ϕ2.5	GACT							C→G (25.8%)		C→G (48.7%)			
ϕ6.5	GACT	GAC→CCC		GAC→CCC	G→A (71.4%), A→G (17.3%), C→G (93.0%), duplication	→AAG, lost duplication	→AAG, CCC, GCC, lost duplication	G→A, C→G (97.8%), duplication	→AAG, lost duplication	→AAG, duplication (13.3%)	G→A (16.7%), A→G (48.8%), C→G (69.0%), duplication	G→A (1.2%), A→C (7.2%), duplication (3.4%)	→GGG, duplication (71.6%)
ϕOR	GACT	A→C	A→C	A→C	A→C (16.3%), G→C (17.2%), C→G (68.6%)		→CCC	C→G, A→G (56.8%), G→A (34.1%)		→AAG, GCG	A→G (40.9%), C→G (69.8%)		→CAG

Table 2.2: List of promoters which mutated in each of the four lines after insertion of either wild-type or G78-KIRV RNAPs and 50 more passages.

Figure 2.3: Pymol rendering of the salt bridge between TEV protease E222 and R84

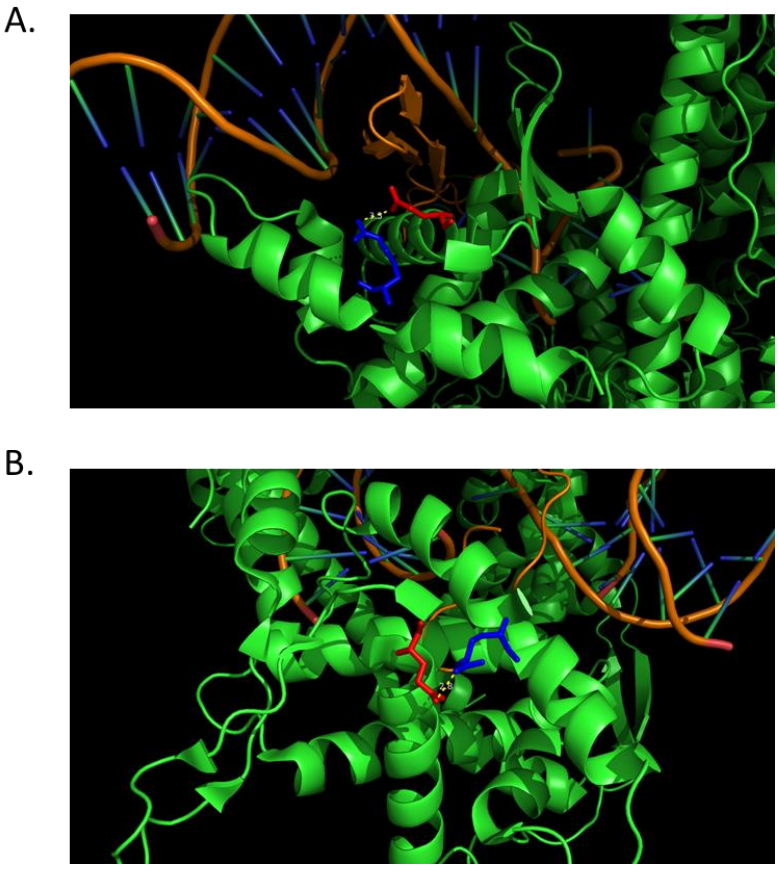


Figure 2.3: Pymol rendering of E222 (in red) and R84 (in blue) in two different crystal structures of T7 RNA polymerase at various enzymatic points. In A, showing T7 RNAP initiation complex, E222A and R84 distance is 3.5 Å (PDB: 1QLN (Cheetham and Steitz 1999)). In B, showing T7 RNAP elongation complex, E222A and R84 distance is 2.8 Å (PDB: 1H38 (Tahirov, et al. 2002)). The E222A mutation should break this salt bridge between E222 and R84.

Figure 2.4: *In-vitro* transcriptions

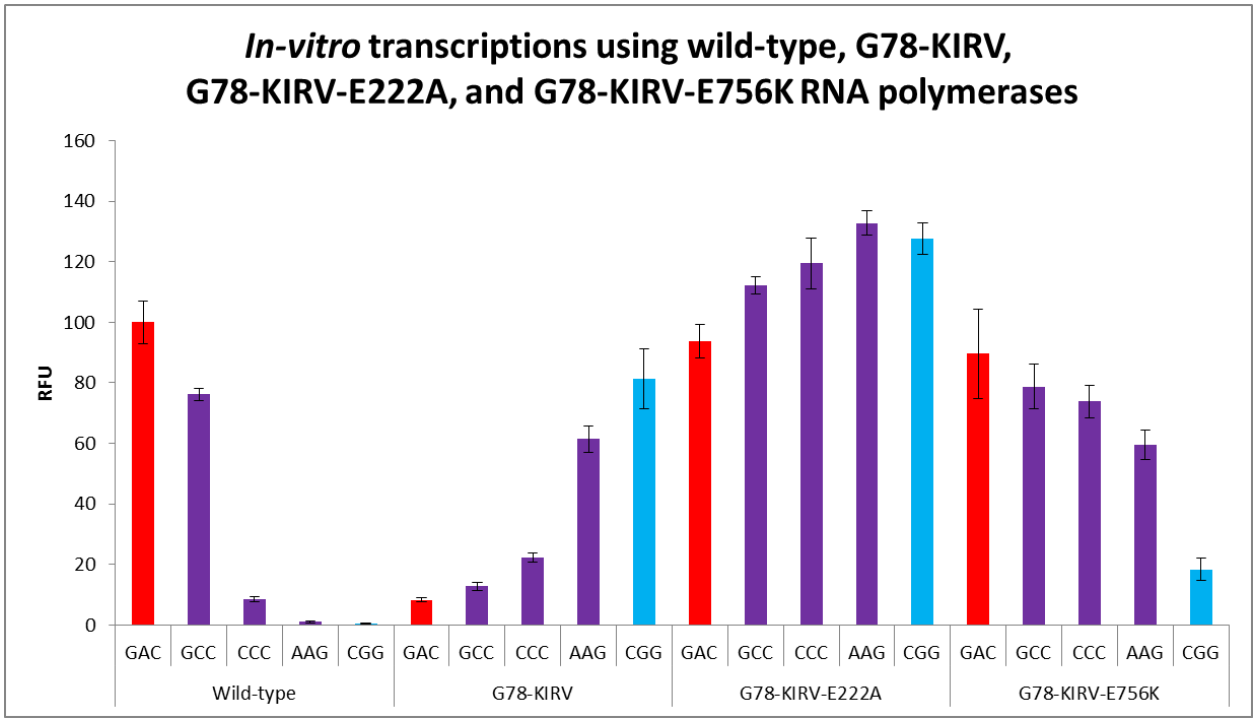


Figure 2.4: *In-vitro* transcription activity assays with two starting RNAPs (wild-type and G78-KIRV) and the two generalist variants that emerged from the passaging (G78-KIRV-E222A and G78-KIRV-E756K) on wild-type promoter (red), three intermediate promoters (purple) and the G78-KIRV promoter (blue). Values normalized to wild-type RNAP activity on wild-type (GAC) promoter. Error bars display standard error.

Figure 2.5: T7 RNAP induction plasmid map for *in-vivo* activity determination

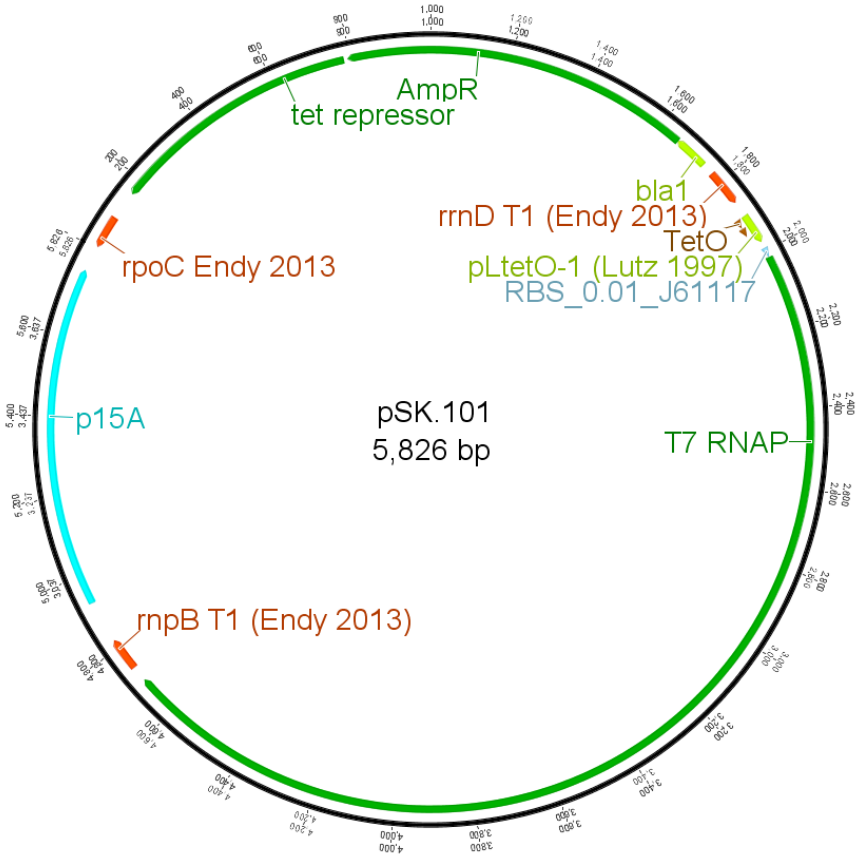


Figure 2.5:

Plasmid map of pSK.101 construct used, in conjunction with pSK.201 or pSK.202, to determine *in-vivo* activity of each respective T7 RNAP variants. In this simple circuit, anhydrous tetracycline drives the expression of the T7 RNAP gene (in pSK.101 plasmid, above) which then will transcribe from the promoter variant on the pSK.201 or the pSK.205 plasmid, respectively, which eventually lead to the production of sfGFP.

Figure 2.6: GFP-plasmid map for *in-vivo* activity determination

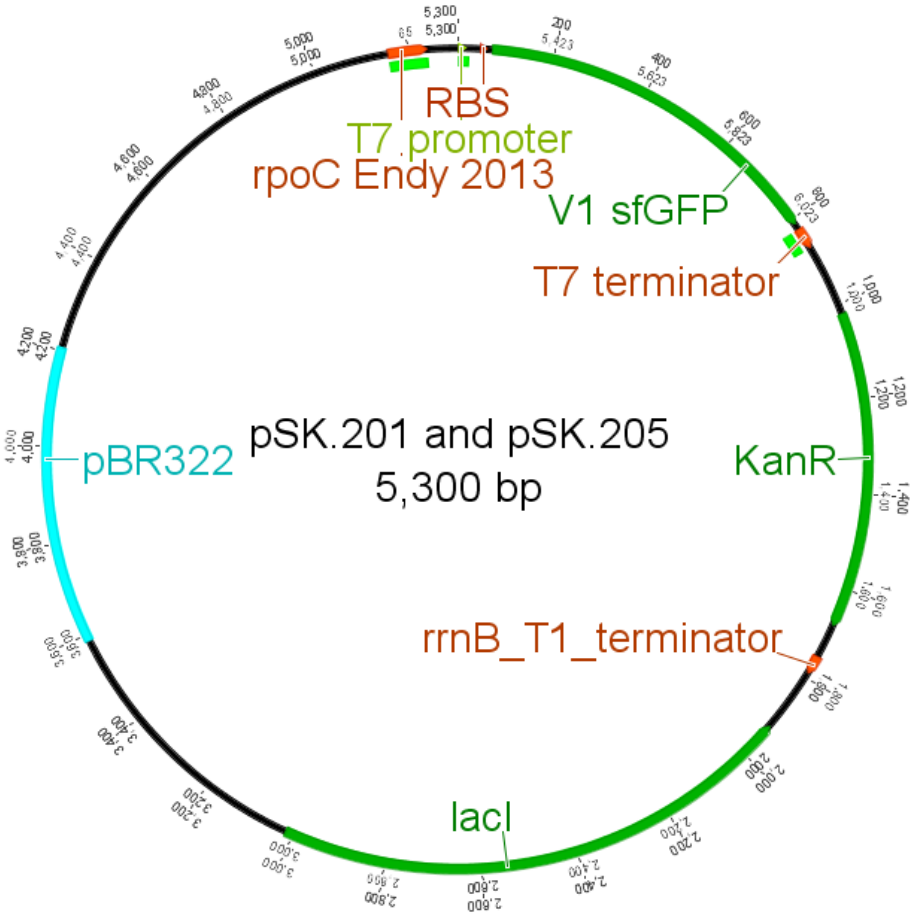


Figure 2.6: Plasmid map of pSK.201 or pSK.205 plasmids which, in conjunction with pSK.101 plasmids, were used to determine *in-vivo* activity of each respective T7 RNAP variants. In this simple circuit, sfGFP gene (in plasmid map above) is transcribed by either the wild-type T7 promoter (TAATACGACTCACTATAG, as in the pSK.201 plasmid) or a low-activity variant promoter (TAATACG**G**CTCACTATAG, as in pSK.205 plasmid).

Figure 2.7: *In-vivo* GFP activity assays

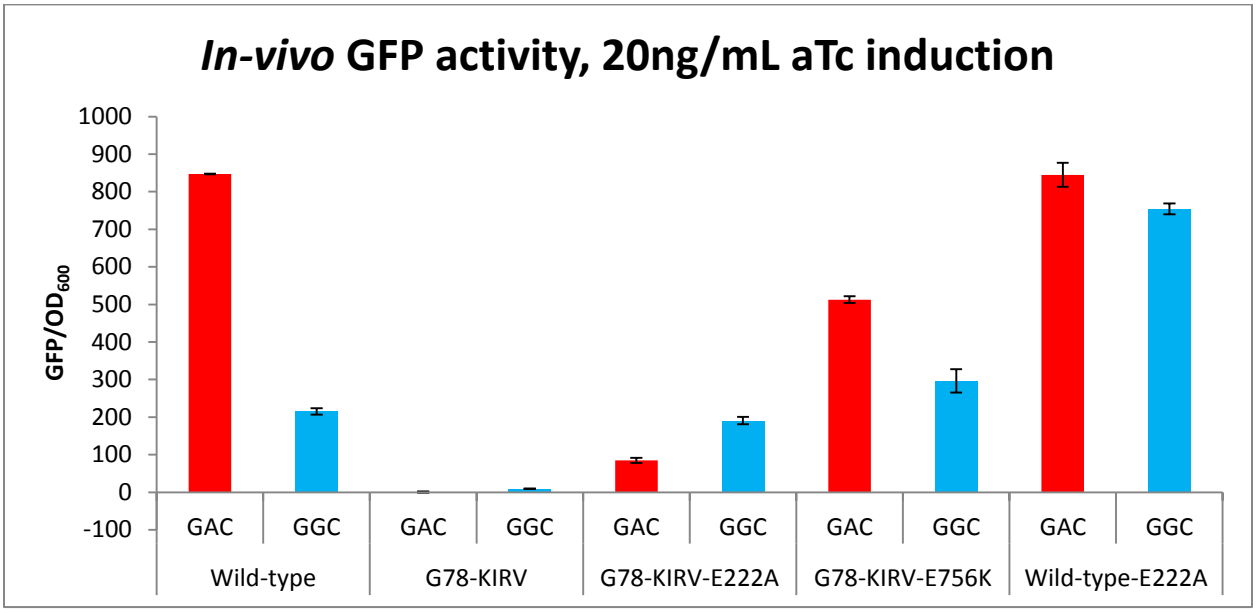


Figure 2.7: *In-vivo* GFP activity assays using wild-type, G78-KIRV, G78-KIRV-E222A, G78-KIRV-E756K, and WT-E222A polymerases on a high (GAC, in red) and low (GGC, in blue) strength T7 promoter-driven GFP at 130 minutes after 20ng/mL aTc induction. Error bars display standard error.

Figure 2.8: Design of recombination constructs to reinsert wild-type and G78-KIRV RNAPs back into T7Δ1 population genomes.

Bacteriophage T7 will recombine with plasmids when provided with 50-100 base pairs of flanking homology around whatever insertion is desired up to a particular size overall phage genome size which otherwise becomes encumbering to genome packaging and capsid formation (Campbell, et al. 1978; Molineux 2006). For RNAP recombination into the T7Δ1 genomes we inserted 51 of 52 identical base pairs to the 5' region immediately in front of the Δ1 deletion onto the 5' end of each RNAP on pLUV-WT-RNAP or pLUV-G78-RNAP (used in previous chapter):
ATCTGCTGAGTGATAGACTCAAGGTCGCTCCTAGCGAGTGGCCTTTTTGATT

This segment had a single base pair change A to T (underlined, above) which was included in order to facilitate reconstitution of a RNase III site which had been destroyed in the creation of the deletion in T7Δ1 while removing the possibility of spurious translation of later codons from an initiating methionine at "ATG". Additionally, 50 identical base pairs immediately downstream of the Δ1 deletion were also added to the 3' end of each RNAP on pLUV-WT-RNAP or pLUV-G78-RNAP:
GACCTTCTTCCGGTTAATACGACTCACTATAGGAGAACCTTAAGGTTTAA

This region includes the T7 ϕ 1.1B promoter (underline, above) which, as part of the 50 base pairs of flanking homology, was essential to clone yet turned out to be problematic when cloning the into pLUV-WT-RNAP but not pLUV-G78-RNAP. We believe that this was the result of creating a runaway positive feedback loop whereas any wild-type RNAP made would initiate transcription at the ϕ 1.1B promoter and transcription of more RNAP in an "autogene"-like circuit, a circuit which quickly succumbs to error catastrophe *in-vitro* (Davidson, et al. 2012) and *in-vivo* (personal observations).

As such, we isolated a mutant pLUV-Flank-WT-RNAP with a single G to T mutation which turned the third amino acid of the His-tagged wild-type RNAP from coding for a GGA (Gly) to TGA (stop). This allowed for recombination of the wild-type

RNAP back into the evolved lines, which all experienced a C to T mutation converting ACG (T3, using no-His tag wild-type T7 RNAP numbering convention) to ATG (Met). This effectively undoes the initial TGA (stop) mutation and allows viable wild-type T7 RNAP (albeit without a His-tag) to be expressed from the phage genome.

Table 2.3: Primers used to create the *in-vitro* templates

DJG.GACT.SpMX1.F	AATATAATACGACTCACTATAGAGGAGACTGAAATGGTGAAGGACGGGTCCAGTGCTTCG
DJG.CGGT.SpMX1.F	AATATAATACCGGTCACTATAGAGGAGACTGAAATGGTGAAGGACGGGTCCAGTGCTTCG
DJG.CCCT.SpMX1.F	AATATAATACCCCTCACTATAGAGGAGACTGAAATGGTGAAGGACGGGTCCAGTGCTTCG
DJG.AAGT.SpMX1.F	AATATAATACAAGTCACTATAGAGGAGACTGAAATGGTGAAGGACGGGTCCAGTGCTTCG
DJG.GCCT.SpMX1.F	AATATAATACGCCTCACTATAGAGGAGACTGAAATGGTGAAGGACGGGTCCAGTGCTTCG
DJG.SpMX1.R	GAAAAGACTAGTTACGGAGCTCACACTCTACTCAACAGTGCCGAAGCACTGGACCCG

Table 2.3: Primers used to create the *in-vitro* templates

Chapter 3: Construction, Exploration, and Optimization of a Designed T7 Life-Cycling System Reliant on Proteases

Abstract

Proteases are a biotechnologically, medically, and economically important class of biomolecules which extend across the three domains of life. This work describes the initial attempts to devise a bacteriophage T7Δ1-based system, as well as the subsequent adoption and optimization of a bacteriophage T7Δ10-based system, constructed in order to facilitate the essential genomic incorporation and use of a foreign protease such that bacteriophage replication is strictly dependent on protease activity. In doing so, work into enhancing the speed of the tobacco etch virus protease (TEV protease) and its plasticity regarding non-canonical target substrate cleavage specificity is explored. Furthermore, in addition to the creation of a T7-TEV bacteriophage-protease-based pair reliant on cleavage on its canonical target substrate for viable bacteriophage life cycling, the creation of a T7-rhinovirus-14 3C protease (3C protease) reliant on its protease activity is also reported. Together, it is shown that these designed T7-TEV and T7-3C phage-protease chimeric viruses show orthogonality which in effect enables selective cell killing of *E. coli* cells dependent on which protease-target they have. The potential use of these *E. coli* cell-specific targeting viruses, together with the overall deficiencies, power, and potential of this T7-protease system, are discussed.

Introduction

Proteases, a class of protein enzymes that cleave other proteins, make up 2% of the human genome (notwithstanding those in other prokaryotic and eukaryotic species) and are important for therapeutic, industrial, and biotechnological applications (Li, et al.

2013). Therapeutic use of protease inhibitors represents a multi-billion dollar per year market (Turk 2006). As such, proteases are an attractive target for protein engineering. In spite of this widespread applicability, rational design and directed evolution of proteases has been difficult on account of toxicity issues inside of whatever cellular chassis is used to produce them (Pogson, et al. 2009).

One strategy developed to get around this issue has been to express the protease for export to the surface of either *E. coli* or yeast cells where their activity is linked to a sortable output, such as fluorescence (Varadarajan, et al. 2008; Yi, et al. 2013). Such strategies, however, are limiting: proteases may not function when placed on the membrane, some may act differently outside of the cellular context based upon differences such as pH, and said strategies require elaborate and potentially noisy methods such as antibody tagging and FACS sorting. Additionally, these strategies have limited dynamic range such that it may be difficult to discern between a "good" variant and a "great" variant.

As we began working on the previous T7Δ1 evolution projects mentioned in the previous chapters, we realized that a T7-based evolution system could be created in which life cycling could be made reliant on protease activity. This system offers advantages over surface display-based systems mentioned above: the protease would work in the context of the cell itself (which is not likely to suffer the same toxicity issues on account of being destroyed by the replicating bacteriophage). Inspired by Phage-Assisted Continuous Evolution, we had anticipated that the speed of replication of bacteriophage T7 (faster than PCR), its rapid rate of phage amplification (a single bacteriophage T7 could lyse the world's oceans-worth of mid-log *E. coli* cells in seven hours), and intense theoretical sampling of sequence space via mutation (a fully-lysed 1mL *E. coli* culture produces a phage population which mutationally samples each

nucleotide position 250 times) would lead to an amenable continuous evolution system (Appendix). Unlike PACE-based evolution systems, however, a T7-based evolution system would require no elaborate steps: all that is required would be to passage the bacteriophage on a series of flasks with target cells (Esvelt, et al. 2011).

There are additional advantages to a T7-based evolution system to specifically evolve proteases. A T7-based system is theoretically easily adaptable to use with multiple variants proteases/protease targets by simple swapping out the protease into the genome and/or any target protease cleavage sight you want to try to target. Additionally, this directed evolution system could be easily adapted to other proteases, the only constraint being that they would need to be able to be expressed and function in the *E. coli* cytoplasm. It is already known that in addition to TEV, the HIV protease, the Hepatitis C protease, and the Human Rhinovirus 3C protease all can be expressed and function in *E. coli* (Dickinson, et al. 2014).

Selection of TEV protease and Initial Efforts to Use T7Δ1 strain

TEV protease was initially chosen as the first protease to use in such a system for a variety of reasons. TEV protease, a 27kDa cysteine protease originally from the genome of the Tobacco Etch Virus, can successfully be expressed in *E. coli*, function *in-vivo*, and can be purified from *E. coli* for *in-vitro* assaying (Parks, et al. 1995; Wei, et al. 2012). TEV protease has also been extensively subjected to engineering and assaying for various purposes using a variety of techniques: solubility optimization, using a whole-cell fluorescence assays; substrate profiling, using antibiotic resistance or bacterial display/flow cytometry sorting; and even increasing substrate tolerance or altering substrate specificity, using *S. cerevisiae* auxotrophy-dependent pigmentation assay or the *S. cerevisiae*-based Yeast ER Sequestration Screening system (Kostallas and Samuelson

2010; Renicke, et al. 2013; Sandersjoo, et al. 2017; Sandersjoo, et al. 2014; van den Berg, et al. 2006; Yi, et al. 2013).

Additionally, the specificity of the TEV protease has been extensively characterized *in-vitro*, *in-vivo* and has been aided by the solving of crystal structures of TEV mutants bound to target substrates (Dougherty, et al. 1989; Phan, et al. 2002). Indeed, the exquisite specificity of the TEV protease towards its 7 amino acid canonical cleavage site, ENLYFQ/S, has led to its extensive use as a biotechnological tool *for in-vivo* and *in-vitro* cleavage of fusion proteins (Phan, et al. 2002; Shih, et al. 2005). Interestingly, in addition to its canonical ENLYFQ/S, the TEV protease has retained intramolecular proteolytic activity on GHKVFM/S, which is the 213th to 219th amino acid itself, in spite of being so different than the canonical cleavage site (Figure 3.1) (Phan, et al. 2002). The crystal structure of the TEV protease together with its substrate provides structural clues which can inform the nature of the interactions between the two and thus inform which residues are fungible (Figure 3.2).

Because no crystal structure has been generated using the self-cleavage site, only those structures with the canonical site were used to understand the nature of the specificity of each residue of the canonical substrate. From these structures and biochemical studies, it has been determined that the P6, P3, P1, and to a lesser degree P1' are the main drivers of specificity determination of the TEV protease for its canonical substrate (Dougherty, et al. 1989; Phan, et al. 2002). In particular, the Glu in P6, the Gln in P1, and the Ser in P1' were identified as the locations by which alterations in specificity could be made and then exploited in order to drive evolution of the TEV protease in this system towards that of this new and hopefully orthogonally targeting, substrate. Indeed, one of the ultimate goals to test the limits of this system as a directed evolution platform to generate multiple orthogonal protease:substrate pairs.

Before reaching for that brass ring, though, there was lower hanging fruit to initially go after. TEV protease is an exceptionally slow enzyme, with a k_{cat} of 0.16 s^{-1} at 30°C (Kapust, et al. 2001). This leads many who use it biotechnologically to use an overnight digestion, usually at 4°C, a temperature chosen in order to preserve the integrity of the desired overexpressed protein to the further detriment of TEV activity. Based on our previous work exploring T7Δ1, we believed it could be used to create a system which would link genotype (in this case, TEV protease genomically integrated into the T7Δ1 locus) to phenotype (in this case, speed of proteolytic activity) by driving creation of something functional that would differentially allow phage life cycling. Because we were using the T7Δ1 system, an obvious way to constrain this phage system and allow differential life cycling was the use of a plasmid-driven T7 RNA polymerase. As noted in the previous chapters, T7Δ1 will not replicate in the absence of T7 RNAP. As such, we decided to create fusions of split T7 RNAP with *E. coli* beta-glucuronidase (GUS) with the canonical TEV substrate as linker(s), aspiring to find a variant fusion protein which lacked T7 RNA polymerase activity without cleavage by TEV yet retained activity when TEV freed the fusion proteins. Our lab's previous work and experience with T7 RNA polymerase in addition to GUS, showed that T7 RNA polymerase was viable as a split protein and that the 614 amino acid GUS protein would be a viable choice for either N-terminus or C-terminus fusion (Flores and Ellington 2002; Segall-Shapiro, et al. 2014).

Three plasmids were created with various fusion proteins, all based on the pLUV chasis, by which inducible expression was driven by the LacUV5 promoter. The first (pLUV-alpha-GUS-beta) had the first 601 base pairs of the T7 RNAP: TEV cleavage site: GUS: TEV cleavage site: rest of T7 RNAP. The second (pLUV-beta-GUS) had, as before, the first 601 base pairs of the T7 RNAP: TEV cleavage site: GUS while allowing

the 601:883 amino acid piece of the T7 RNAP to be independently expressed from another LacUV5 promoter. The third (pLUV-GUS-beta) had GUS: TEV cleavage site: first 601 base pairs of the T7 RNAP then, as the previous construct, the 601:883 amino acid piece of the T7 RNAP was allowed to be independently expressed from another LacUV5 promoter. With these three constructs, each together with a separate plasmid with an anhydrous tetracycline-inducible TEV protease, we were able to determine the relative ability of each to support viable T7Δ1 phage replication in BL21 *E. coli* cells (Table 3.1).

The lysis assays clearly show that when the RNAP-GUS fusions are expressed together with TEV protease the T7Δ1 phage replicate at a faster rate than in any other circumstance. It is worth noting that the alpha-GUS-beta fusion expressing construct, where both pieces of the split T7 RNA polymerase are fused between two canonical TEV cleavage sites on either end of GUS, took the longest of the three constructs to lyse. Because of the need for two cleavage events instead of one (in addition to the rather large size of the alpha-GUS-beta polyprotein and any relative decrease in protein expression with this construct relative to the other two) this result strongly suggests the TEV expression is responsible for cleavage and thus ultimately playing the role in which we expect it to in this system. Although this is promising, the decrease in OD_{600} (and, by proxy, the successful lysis by the T7Δ1 phage of the *E. coli* cells) in the T7Δ1 phage infected -tet/+lac induced cells of both those harboring the beta-GUS and GUS-beta constructs was worrisome: this result, taken together with the other OD_{600} data, indirectly suggests that these constructs, when highly expressed, still allow for some partial T7 RNA polymerase activity and therefore successful the T7Δ1 phage replication.

In spite of this partial T7 RNA polymerase activity and therefore delayed (albeit successful) T7Δ1 phage life cycling, a construct was designed based upon the pLUV-

Flank-wild-type-RNAP plasmid used in the previous chapter to reinsert the T7 RNA polymerase back into the evolved the T7Δ1 phage lines. This plasmid, pLUV-Flank-TEV, was built to completely substitute out the RNA polymerase and contain in its place the TEV protease S219P mutant, which is self-cleavage resistant (Kapust, et al. 2001). This construct was used, in 1mM IPTG induced BL21-DE3 *E. coli* cells, in an attempt to stimulate recombination of the TEV protease into the Δ1 locus of the T7Δ1 phage and thus enable phage replication. These attempts to enrich for potentially recombined now "T7-TEV" phage on all three of the GUS-fusion plasmid populations (after multiple attempts, multiple rounds of passaging on pLUV-Flank-TEV in the hopes of increased recombination, and even after the use of multiple T7 growth strategies such as plates versus liquid culture) led to the same conclusion: no recombinants were found. Though this design strategy was shown to not work, it did show that split T7 RNA polymerase could be constrained by GUS fusion within the activity required of T7Δ1 phage in such a way to at least delay T7Δ1 phage replication. To reach the next step of protease-containing T7 phage in which life cycling is dependent on externally programmed stimuli, a new design modality was needed.

Switching to a T7Δ10 system

At this point, at the suggestion of Ian Molineux, we abandoned the T7Δ1 phage system in favor of a T7Δ10 phage system. In such a T7Δ10 system, the major capsid protein 10A (from T7 gene 10) is exploited in order to link the phenotype (the function of the protease) with its genotype (its DNA code). Gene 10 actually codes for two different versions of the major capsid protein, 10A and 10B, where the 10B protein, though not essential, gets produced at about 10% of the frequency of the 10A protein as a -1 translational shift at the 3' end of 10A (Molineux 2006). Henceforth, all references to the

major capsid protein will refer to the 10A gene product. 415 of these major capsid proteins assemble to make one of the icosahedral capsids of the T7 virus and its production is the limiting step of late T7 replication (Guo, et al. 2014). T7Δ10 strains (where the major capsid protein is deleted) will not lyse cells and complete life cycling unless it is complemented in trans with the gene 10 expressed from a plasmid. As before, the design entailed the creation a fusion protein between the T7 major capsid protein and *E. coli* theoredoxin A (TrxA) linked by a protease site such that the fusion protein will not make viable bacteriophage without cleavage of the fusion protein.

Fusions with the T7 major capsid protein have already been demonstrated for use in phage display such that it is known that it can tolerate up to 50 amino acid fusion to its C-terminus and still be functional (Novagen 2014). TrxA is a 110 amino acid protein which has been used previously as N and C terminal fusion protein expression system to circumvent aggregation in *E. coli* inclusion bodies and also presented an intriguing possibility as adding yet another selective pressure, as TrxA is essential to bacteriophage T7 life cycling (LaVallie, et al. 1993; Molineux 2006). Conceptually, to run this system/selection, our non-functional major capsid protein fusions were to be expressed, as before, from a Lac inducible promoter (LacUV5) from a plasmid (pLUV) in target BL21 *E. coli* cells. T7Δ10 bacteriophages with protease variants inserted into their genomes targeting the protease site linker in the fusion protein were then to be added to these target cells. The viruses with functional proteases will propagate and those that do not will not. A library of protease variants could theoretically be screened against any number of protease site linkers in the major capsid protein simply by swapping out one linker for another (Figure 3.3).

As before, two fusion proteins were designed (a gene 10: TEV cleavage site: TrxA fusion and a TrxA: TEV cleavage site: gene 10 fusion) which, together with a gene

10 only control, was subjected to the same plasmid-based TEV and fusion protein expression assay to determine to what extent these constructs could support successful T7Δ10 life cycling (Table 3.2). All of the cells harboring the gene 10 only control construct and infected with T7Δ10 lysed, which is not surprising given the leakiness of the lacUV5 promoter. In contrast, the gene 10-TrxA containing cells, even with high expression of both the protease and fusion protein did not seem to lyse. Finally, those cells containing the TrxA-gene10 construct only lysed when the protease was highly expressed. The fact that both the +tet/-lac induced cells and the +tet/+lac induced cells both lysed, at first blush, may appear to be worrisome: this worry, however, is needless, as it was shown with the gene 10 only control that the lacUV5 promoter is leaky. As such, the main takeaway is that TEV protease expression was necessary to allow for successful T7Δ10 life cycling.

This led, as before, to attempt to reinsert the TEV S219P protease into the Δ10 locus of T7 through the creation of pPOS-Flank-TEV, a p15A origin Cm resistant plasmid which had 140 bp and 64bp of flanking homology on the left and right side, respectively, of the Δ10 locus. Most importantly, the 140bp of flanking homology on the left side of the pPOS-Flank-TEV contained a T7 promoter which should enable production of TEV protease by any infecting T7Δ10 in a manner similar to pAR5403, a pBR322 origin, Amp resistant plasmid containing a T7 promoter-driven major capsid protein which enables complementation of T7Δ10 (in a similar manner as BL21-DE3 strain was used to complement T7Δ1). As before, a vast array of techniques was employed in an attempt to enrich for T7-TEV phage. Finally, one of them worked: T7Δ10 were passaged for 5 passages on BL21 *E. coli* cells which contained both the pPOS-Flank-TEV plasmid and a pLUV-T7pro-TrxA-TEV-site-gene-10 which was identical to the pLUV-T7pro-TrxA-TEV-site-gene-10 plasmid except the LacUV5

promoter was swapped for a T7 promoter. The reinsertion was confirmed via PCR, population sequencing, as well as individual phage isolate sequencing.

The reasoning behind successful insertion of the protease into T7Δ10 relative to the failure of the same into T7Δ1 was not determined. Speculating, it would seem as if the T7Δ10 system was controlled by the protease more than the T7Δ1 was. This increased dependency or constraint upon the system, together with the use of T7 promoter-driven transcription of both proteases and fusions, was apparently enough for enrichment. It was highly fortuitous that the 140bp of left flanking homology in front of the TEV in the pPOS-Flank-TEV plasmid enabled both recombination as well as expression of the TEV protease. Finally, although noticed later, it would seem as if the plasmid copy number of the plasmid on which flanking homology to recombine into the T7 genome has a direct effect on the efficiency of this relatively low efficiency event. The efficiency of plating after recombination into T7 from a pUC plasmid (mutated pMB1 origin, copy number ~500-700) is roughly 10^{-5} versus roughly 10^{-7} on the pPOS plasmid (p15A origin, copy number 10-12) (Kiro, et al. 2014). As such, at a certain point, the system was switched from using the pPOS plasmid to a pUC plasmid for protease recombination.

Wild-type TEV drift mutations and TEV-Fast explorations

Isolates of the initial TEV-S219P mutant protease in T7 bacteriophage were plated on 1mM induced pLUV-TrxA-TEV-site-gene-10 cells, and single phages were isolated and grown, from which each respective TEV protease was PCR amplified and sequenced. All ten of these isolates were the original TEV-S219P inserted into the T7 genome. One of these isolates was then used to passage, in triplicate, on 1mM induced pLUV-TrxA-TEV-site-gene-10 cells with a transfer volume of 1uL, giving an

approximate MOI of 0.1 for 10 passages. Ten phages from each of the three populations were isolated and their TEV proteases were PCR amplified and sequenced (Table 3.3). By passage 10, all of the thirty isolated bacteriophages had a picked up a P219S mutation which functionally mutates the proline back to the wild-type serine residue and thus renders the protease susceptible to self-cleavage. This in and of itself would not be of interest except for the fact that the wild-type (219S) TEV has a k_{cat} of 0.16 s^{-1} and k_{cat} /Km of 2.62 mM^{-1} s^{-1} whereas the S219P TEV has a kcat of 0.09 s-1 and kcat /Km of 1.36 mM^{-1} s^{-1} (Kapust, et al. 2001). In effect, in the course of only 10 passages, the bacteriophage had mutated its TEV into a faster, more efficient TEV protease. The bacteriophages also all picked up an N177T mutation which, by analysis of the TEV crystal structure, interacts with the P6 Glu of the TEV substrate and is thus hypothesized to stabilize the binding of TEV protease to its substrate (Figure 3.4).

Encouraged by this, another series of passages were done with the original S219P-TEV-containing phage over pLUV-TrxA-TEV-site-gene-10 cells which also contained a Flanked (i.e. flanked with homology to the phage to enable recombination) TEV-Fast protein. “TEV-Fast” is the moniker given to a G79E, T173A, and S219V TEV mutant which emerged from the yeast ER sequestration screening (YESS). TEV-Fast has the fastest published kcat of any TEV protease at 0.30 s^{-1} (Yi, et al. 2013). We wanted to see if the S219P-TEV bacteriophage, a small fraction of which would recombine in the TEV-Fast, would get outcompeted by that small fraction of now TEV-Fast-containing phage. From screening 30 isolates there was no evidence of TEV-Fast although the N177T mutation was picked up by 22/30 phages and the P219S reversion occurred in all 30 bacteriophages (Table 3.4).

We then decided to isolate S219V-TEV and TEV-Fast recombinant T7 bacteriophage and passage them. The S219V mutant, another of the cleavage-resistant

variants, is the second fastest published TEV protease after TEV-Fast, with a k_{cat} of 0.19 s^{-1} (Kapust, et al. 2001). After 10 passages on 1mM induced pLUV-TrxA-TEV-site-gene-10 cells, no mutations occurred in either population. In an effort to both enrich for protease activity as well as for mutants the concentration of IPTG used to induce the pLUV-TrxA-TEV-site-gene-10 cells was decreased to 100uM and the MOI was reduced to 0.01. After another 10 passages, all ten of isolates from the S219V population had a N177I mutation and 8/10 had an R51H mutation (Table 3.5). Ten of ten isolates from the TEV-Fast population had picked up the N177T mutation and one had a S219G mutation (Table 3.6). Taken together within the context of the previous mutation data, these data suggest that mutating the N177 residue (specifically to N177T) is favorable and presumably leads to an increased speed of TEV protease.

It is relatively easy to characterize cell lysis time by bacteriophage T7 through the use of a plate reader. Cell lysis time (together with burst size) is a major contributor towards overall bacteriophage T7 fitness. As such (using the methods described in previous chapters), the cell lysis times of various TEV-phage populations at two different induction strengths were determined (Figures 3.5 and 3.6). These data show very little difference in lysis time among all of the T7-TEV phage populations at 1mM IPTG induction (Figure 3.5). These data also show that, when the inducer concentration is lowered to 10uM IPTG, the later passaged phage populations (that is, passage 10 vs. passage 1 for the S219P phage, passage 20 vs. passage 1 for S219V phage, and passage 20 vs. passage 1 for the TEV-Fast phage, respectively) all lyse at a faster rate than their ancestor phages (Figure 3.6). This suggests that the mutations in the TEV proteases at these later passages, mutations elsewhere in the T7 genomes, or both, are responsible for the rapid decrease in cell lysis time. As no next-generation sequencing was done in order to determine any T7 genome mutations outside of the TEV protease gene, the answer to

this question is currently unknown. Though it is tempting to speculate that the addition of N177T independently in all of the bacteriophage populations has led to faster proteolytic activity by each TEV protease further kinetic characterization would be needed. Having learned a bit about this novel T7 life cycling system, and with a TEV-S219V N177T mutant in hand, we decided to press on towards exploring whether the T7-TEV phage system could be used to expand substrate specificity and, ultimately, lead to orthogonally-acting proteases.

TEV substrate specificity expansion and orthogonality explorations

Somewhat in parallel with the TEV kinetic explorations above the S219V, N177T TEV-containing bacteriophage T7 evolved after 20 passages was subjected to *E. coli* cells containing plasmids encoding for TrxA-gene 10 fusion beyond the canonical TEV substrate (ENLYFQ/S). Specifically, the S219V, N177T TEV-containing bacteriophage were subjected to four non-canonical TEV target populations: EΔW, in the P6 target substrate locus (**W**NLYFQ/S), YΔE in the P3 target substrate locus (ENL**E**FQ/S), YΔH, also in the P3 target substrate locus (ENL**H**FQ/S), and QΔH in the P1 target substrate locus (ENLYF**H**/S). All of these sites (in the P6, P3, and P1 substrate loci) were chosen because of their importance for their TEV protease specificity determinants. The QΔH population, in particular, was chosen on account of the preceding success with its use in the YESS system and nominally served as a basis for comparison between the YESS system and this novel protease evolution methodology (Yi, et al. 2013). Each of the populations, save the QΔH-adapted population, was adapted in the following order of apparent increase in stringency: 10 passage on 1mM induced target at an MOI of 0.1, followed by 10 passages on 1mM induced target at an MOI of 0.01, and finally 10 further passages on 1uM induced target at an MOI of 0.01, giving overall 30 passages on each

non-canonical TEV substrate (Figure 3.7). The QΔH-adapted population was not able to lyse target cells at 1uM IPTG induction and thus the series of passages were omitted. Lysis time with these evolved populations was then determined on *E. coli* cells harboring each of their own adapted TEV substrate target, the other non-canonical TEV substrate targets, as well as the canonical TEV substrate target (Figures 3.8-3.12).

The ancestor N177T, S219V TEV-containing phage (before the 30 passages on the non-canonical TEV substrate target) lysed the canonical TEV substrate target the fastest, followed by the EΔW target substrate locus (WNLYFQ/S), then the YΔH target substrate locus (ENLHFQ/S), and finally the YΔE target substrate locus (ENLEFQ/S) and the QΔH substrate locus (ENLYFH/S), which shown no signs of lysis in this assay (Figure 3.8). The adapted populations, however, lysed each of their respective adapted-to non-canonical TEV substrate targets at or nearly within standard deviation as quickly as they lysed the canonical TEV substrate target (Figures 3.9-3.12). These populations also lysed each of the non-canonical TEV substrate target-containing cells differentially. In particular: 1) the EΔW adapted population lysed the YΔH target cells within deviation of its own target while lysing the YΔE and QΔH cells at a slower rate, 2) the YΔH adapted population lysed both the YΔE and EΔW nearly as quickly as itself while lysing the QΔH cells at a slower rate, 3) the YΔE adapted population lysed both the YΔE and EΔW nearly as quickly as itself while lysing the QΔH cells at a slower rate, and 4) the QΔH adapted population lysed, in order, the YΔE, EΔW, and YΔH at a slower rate than it or the canonical TEV substrate (Figures 3.9-3.12).

Isolates from each of these four non-canonical TEV substrate adapted populations were sequenced (Table 3.7). Although each isolate has quite the variety of further mutations beyond the original S219V and N177T ancestor (ranging from at the minimum three and at the maximum five additional non-synonymous mutations) strikingly none of

the four non-canonical TEV substrate adapted populations have any mutations which they and they alone possess (Table 3.7). This strongly suggests that most of these mutations are generally enabling, that is, allowing either generalist TEV protease activity or allowing enhanced protein expression and translation. Yet an apparent incongruity exists: how do these sequence heterogeneous phage populations lead to differential speed of lysis seen in the lysis assays? One potential explanation is that the apparent heterogeneity among each adapted population (in the apparent absence of exclusivity of any mutations between each population) is specifically tuned in relation to each other such that, on the whole, the population gains the observed differential lysis. Another possibility is that the rare variants within each population are responsible for these differential lysis timing. No work followed in order to tease out these differences, although in hindsight it would be of interest to determine what further evolutionary adaptation would occur with single phage isolates of some of the relatively rare mutations or mutational combination phages. Finally, it is not impossible that adaptation in the rest of the bacteriophage T7 genome beyond the TEV protease is responsible for the differential lysis times. Though theoretically possible, the strong selective pressure on the TEV protease activity itself, born out by its rapid accumulation of mutations, makes this unlikely.

Interestingly, the EΔW, YΔE, and YΔH adapted populations all lysed the QΔH containing target cells the slowest. This suggested that there entailed a potential degree of orthogonality between these populations. As such, an EΔW, QΔH (**W**NLYF**H**/S) double mutant TEV target substrate was created. After a multitude of efforts using both EΔW and QΔH adapted populations (including, but not limited to, passaging on gradually increasing double mutant cell population), there was no successful lysis on this double mutant. As such, it was determined that, though interesting, these gradual steps towards

altered specificity would not ultimately be productive as far as evolving an orthogonally acting TEV protease from its wild-type ancestor and that other alternatives should be pursued.

Rhinovirus 3C and orthogonality between T7-TEV and T7-3C phages

Though the path forward to actually evolving a panel of orthogonally-acting TEV proteases via this T7-based system was limited there was still the potential to use new proteases. The rhinovirus-14 3C protease was chosen after using TEV as the initial proof-of-concept. Rhinovirus 3C protease has many of the same advantages for use in the T7 system that TEV protease does: it can be functionally expressed in *E. coli*, its specificity has been investigated, and it has a published crystal structure (Cordingley, et al. 1990; Cordingley, et al. 1989; Matthews, et al. 1994). Rhinovirus 3C protease, a cysteine protease with exquisite specificity for recognizing and cleaving LEVLFQ/GP, has been used biotechnologically to cleavage fusion proteins and also as a target of chemical inhibitors seeking to provide antiviral activity against rhinovirus life-cycling in humans (Binford, et al. 2007; Wang and Johnson 2001).

Rhinovirus 3C protease was cloned and expressed in *trans* (i.e. from a plasmid) along with the TrxA-3C target site-gene 10 fusion protein as TEV was before in order to determine whether protease activity could enable T7Δ10 life cycling (Table 3.8 and Figure 3.13). The successful lysis when only the protease and target fusion constructs were highly expressed was encouraging and led to the cloning of the 3C protease into a pUC vector with the flanking regions around the T7Δ10 locus. Recombination of the 3C protease into the bacteriophage T7, now generating T7-3C phage, was successful. Efforts to passage this T7-3C phage on both T7 promoter and LacUV5-driven 3C target constructs for 30 passages resulted in the accumulation of no mutations to the 3C

protease. This could be a result of the 3C protease acting as a higher turnover enzyme (with a k_{cat} of 3.4 s^{-1}) relative to TEV protease (k_{cat} of 0.16 s^{-1}) (Kapust, et al. 2001; Wang and Johnson 2001).

Instead of jumping into pursuing the evolution of rhinovirus 3C protease variants towards non-canonical 3C target cleavage sites it was decided to determine the relative orthogonality of the T7-TEV and T7-3C proteases in the context of T7-protease life cycing. Ideally, the T7-TEV phage would only lyse cells containing its ENLYFQ/S canonical target substrate with T7-3C phage only lysing cells containing its LEVLFQ/GP canonical target substrate (Figure 3.14). Lysis curves with each respective T7-TEV and T7-3C phage on the TEV and 3C target substrate cells clearly show this orthogonality (Figures 3.15-3.17). Further work, using tubes of mixes of TEV and 3C target substrate cells subjected to the T7-TEV phage, T7-3C phage, or both, at various concentrations of phage, cell amounts, and inducer, was not able to reach the point of differential lysis, that is, the point at which which only one type of non-targeted cell population was able to survive while the other, targeted type would die. This is to be expected in a static mixture of two types of targeted cells: as one gets targeted and lysed, it releases the approximate burst size of T7 (260 phages/infected cell) which can then infect en masse and kill the non-target *E. coli* cells by membrane destabilization, genomic destruction, or perhaps even basal protease activity on the non-target substrate (De Paepe and Taddei 2006). Differential cell killing via the combination of T7-proteases and target-containing cells is anticipated to be successful not in a static test-tube environment but rather in a chemostatic environment where viruses get washed out instead of building up in this system (for that matter, so could the efforts towards evolving protease variants towards non-canonical targets) (Bull, et al. 2006). Said efforts, if successful and parameterized,

may ultimately enable the use of these T7-proteases for selective *E. coli* cell killing in rodent or human gut model systems (Bull and Gill 2014).

Other efforts to expand this T7-protease system

The realization that this T7-protease system could be used for other purposes both beyond evolving proteases towards non-canonical target cleavage sites and beyond selective cell killing did not go unrecognized. Similar to evolution of protease-drug resistance using PACE, this T7-protease system could also be used to discover which mutations lead to resistance to protease inhibitors (Dickinson, et al. 2014). As such, the T7-3C phage was screened against *E. coli* cells which had been treated with various concentrations up to and beyond the IC_{50} of rupintrivir, an irreversible inhibitor of the 3C protease (Binford, et al. 2007). The T7-3C phage were still able to lyse these rupintrivir-treated *E. coli* cells, presumably because the rupintrivir was not making it into the *E. coli* cells at a concentration by which 3C protease inhibition halted T7-3C life cycling.

Another expansion of this T7-protease system is through using it as a means of profiling non-canonical target cleavage sites. This was done on a proof-of-concept level, by using an NNS library at the codon coding for the P2 residue in the canonical 3C target cleavage site in the pPOS5-TrxA-gene-10-3C constructs to create an N=32 library. After making these now-single-mutant away from the canonical 3C cleavage site (a LEVL**X**Q/GP library), plasmids were transformed into *E. coli* cells, plated, and individual colonies were picked and used to grow up each well in a 96-well plate (a N=32 library can be 95% covered with a sampling of 94). These O/N cultures were used to grow up fresh cells in 1mM IPTG-induced 2xYT, as to stimulate production of the TrxA-cleavage site-gene 10 proteins, and each well was then inoculated with T7-3C phage and placed on a plate reader. Lysis curves were determined and the clonal overnight cells

(from which were derived the freshly induced cells which were successfully lysed) of those that had lysed on the plate reader were then miniprepped and sequenced as to determine the target cleavage sequence. It was found that the 3C target P2 residue in this system could change from a phenylalanine (F) to a glycine (G), valine (V), arginine (R), and glutamic acid (E). Though there are other, higher through-put assays for identifying protease specificity, this method may be particularly useful for helping decide paths towards evolving and targeting protease mutants against non-canonical cleavage targets (O'Donoghue, et al. 2012).

Finally, the efforts here to mutate proteases in response to non-canonical cleavage targets were inherently limited in that all used a single starting protease sequence. In theory, this system would greatly benefit from the use of a library of proteases as a starting point instead of just one. The relatively low recombination rate between bacteriophage T7 and even a high copy-number plasmid makes this route an implausible path forward for library generation (Kiro, et al. 2014). Another strategy used by others to engineer genomic additions to bacteriophage T7 has been the assembly of its genome on a plasmid in yeast, followed by prepping out of said plasmid, and finally transformation into an *E. coli* cell-type specifically adapted for the transformation of large plasmids in which the bacteriophage will "boot" and begin its life cycling (Ando, et al. 2015). Two attempts were made specifically trying to clone wild-type bacteriophage T7 in this manner, using 8 replicates each. These replicates (though having evidence of the bacteriophage T7 genome detected by PCR) nonetheless failed to boot when transformed into *E. coli* cells. Anecdotally, others have had technical issues when attempting to do the same thing. As such, this method in and of itself does not seem like it would be a useful way to generate protease libraries in bacteriophage T7. This scheme may be able to be successfully adapted, however, if one could insert any gene of interest into the

bacteriophage genome within a yeast plasmid using yeast's ability of the yeast I-SceI restriction enzyme to select particularly those T7 genomes which had experienced recombination (Plessis, et al. 1992). These now high-percentage of recombined T7-genome-in yeast plasmids could then be prepped out and successfully booted up when added to a cell-free transcription-translation system, which has already been shown to facilitate bacteriophage T7 genome replication, synthesis, and assembly of infective virons (Shih, et al. 2005).

Conclusion: the deficiencies, power, and potential of this T7-protease system

The creation of this T7-protease system led to multiple lessons regarding the nature and use of split-T7 RNA polymerase, bacteriophage T7 recombination, functional fusions with the major capsid protein, the plasticity of TEV and 3C proteases, and the overall ability to rationally engineer control of bacteriophage life cycling. Though nominally successful in being able to rationally design control over the bacteriophage life cycling, it did not achieve a main goal of acting as a facile directed evolution platform for the directed evolution of proteases towards non-cognate target substrates. Many lessons were learned in this process. Chiefly, it would seem as if the limited sequence constrains which result from using only one protease variant, together with a limited sampling of sequence space as a result of the natural T7 DNA polymerase error rate (which, functionally can sample only those amino acids within point mutational distance of the original/starting sequence). It is tempting to speculate that the results, beyond the implied emergent generalists resulting from the TEV non-canonical substrate selections, would be different had the same or nearly the same library size as is typically brought to bear in *E. coli*-based directed evolution schemes (approximate library size of 10^8) was also used in

this system. The difficulty in creating such a library in bacteriophage as well as a potential solution is stated in the previous section of this chapter.

Beyond library size, an additional deficiency of this T7-protease system as a directed evolution system is a result of its read-out: it is difficult to know in the absence of sequencing which mutations have occurred in the protease of interest and, additionally, difficult to know "what went wrong" when macroscopic lysis does not occur. Indeed, it often takes five or so passages before one can be sure that the T7-protease viruses have or have not enriched microscopically enough to actually be detected macroscopically. These two deficiencies, admittedly, have no very easy solution. One final noteworthy deficiency of this system is inherent to the bacteriophage life cycling: T7 must infect and produce virions while *E. coli* cells are in logarithmic (i.e. growth) phase (Molineux 2006). The exact mechanism behind this limitation is unknown and is also an admitted deficiency of this T7-protease viral life cycling system for the evolution of proteases.

Having conceded some of the limitations of the T7-protease system is it worth noting some of its power and potential. This T7-protease system, unlike many, many other continuous evolution systems, did not cheat, that is, it did not find unintended ways (such as recombination) around of selecting what it was selected on. It is relatively easy, especially in relation to PACE, to turn rounds/passage. A compelling observation speaks to the unique mutational sampling of this system in how it was run: at no point in sequencing any protease variant was there observed a silent mutation. Overall, though a panel of orthogonal proteases derived from the same ancestral protease was not achieved, the creation and orthogonal use of T7-TEV and T7-3C can enable selective cell killing and also suggests that the system in extensible to other proteases. Though unsuccessful in proof of concept, a T7-protease system may enable rapid screening of protease inhibitors or the enrichment of protease mutants in response to protease inhibitors. Beyond

proteases, simple modifications to a T7 life cycle-constraining evolution system can be envisioned where the replication of bacteriophage T7 is reliant upon other genes of interest such as recombinases, tRNAs, and tRNA-synthetases. Overall, this system is another step in the direction of utilizing the rational design and synthesis tools associated with synthetic biology together with the power of evolutionary pressure in order to drive novel viral creation and synthetic viromics.

Figures

Figure 3.1: Depiction of canonical TEV protease cleavage site

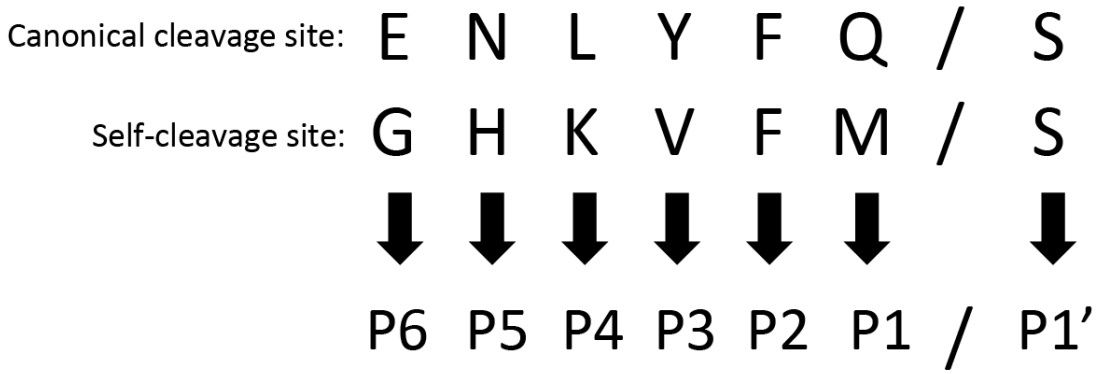


Figure 3.1: Depiction of canonical TEV protease cleavage site and the self-cleavage site on the TEV protease on which it has activity.

Figure 3.2: Ligand-receptor interaction map of TEV protease

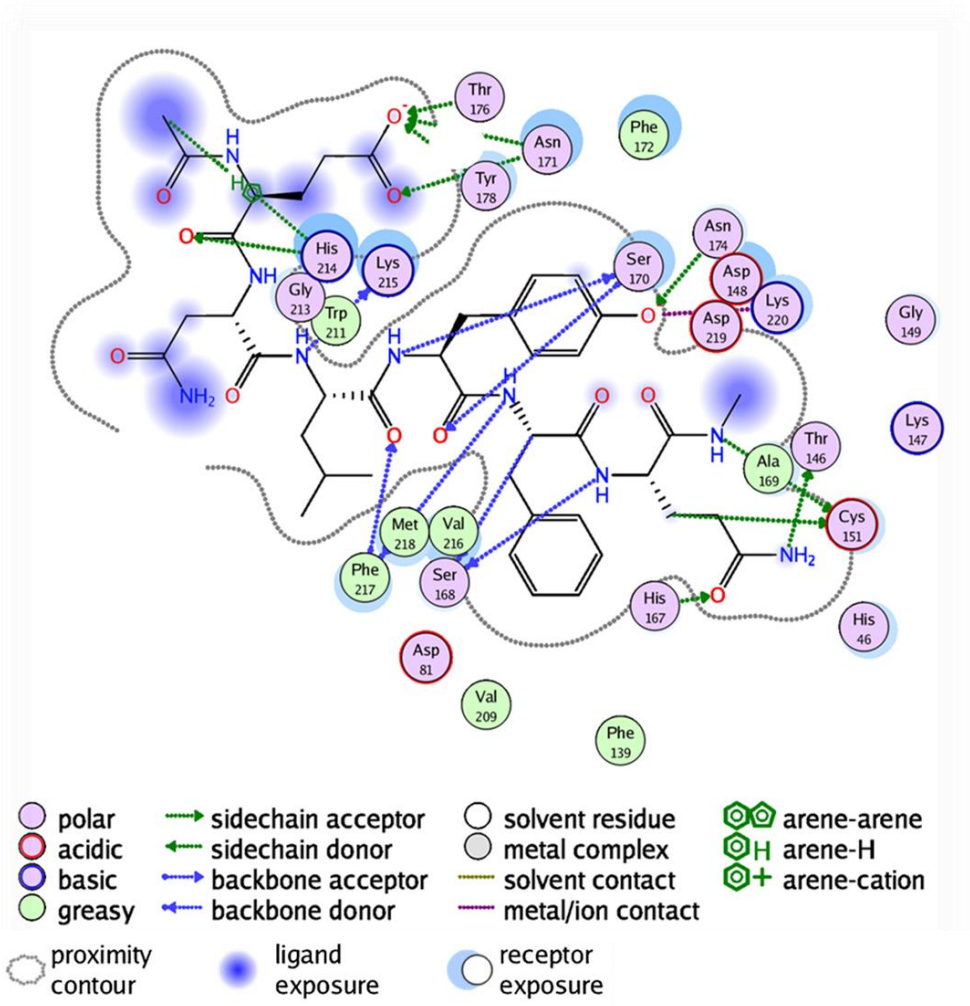


Figure 3.2: Ligand-receptor interaction map of TEV protease complexed with its canonical ENLYFQ product, based on PDB: 1LVM, created using Molecular Operating Environment.

Table 3.1: OD_{600} readings of BL21 *E. coli* cells containing each of the three split-T7 RNAP constructs

		no phage -tet/-lac	+ phage -tet/-lac	+ phage -tet/+lac	+ phage +tet/-lac	+ phage +tet/+lac
pLUV-alpha-GUS-beta	75 min post I	2+	2+	2+	2+	1.86
	105 min post I	2+	2+	2+	2+	1.48
pLUV-beta-GUS	75 min post I	2+	2+	2+	1.98	0.38
	105 min post I	2+	2+	1.72	2+	0.37
pLUV-GUS-beta	75 min post I	2+	2+	1.36	1.96	0.4
	105 min post I	2+	2+	1.1	2+	0.34

Table 3.1: Table showing the OD_{600} readings of BL21 *E. coli* cells containing each of the three split-T7 RNAP constructs under the control of a LacUV5 promoter together with a pTET15A plasmid containing TEV protease under a pTetO promoter. O/N *E. coli* cells were diluted into 2mLs of 2xYT with proper antibiotics and grown for one hour to mid-log phage at which point phage were added. Time recorded in minutes after infection. + phage= MOI 0.1 T7Δ1 phage, +tet = induction of transcription from the pTetO promoter by 200ng/mL anhydrous tetracycline induction, +lac= induction of transcription from the LacUV5 promoter with 1mM IPTG induction.

Figure 3.3: Scheme for T7Δ10-based protease system/selection.

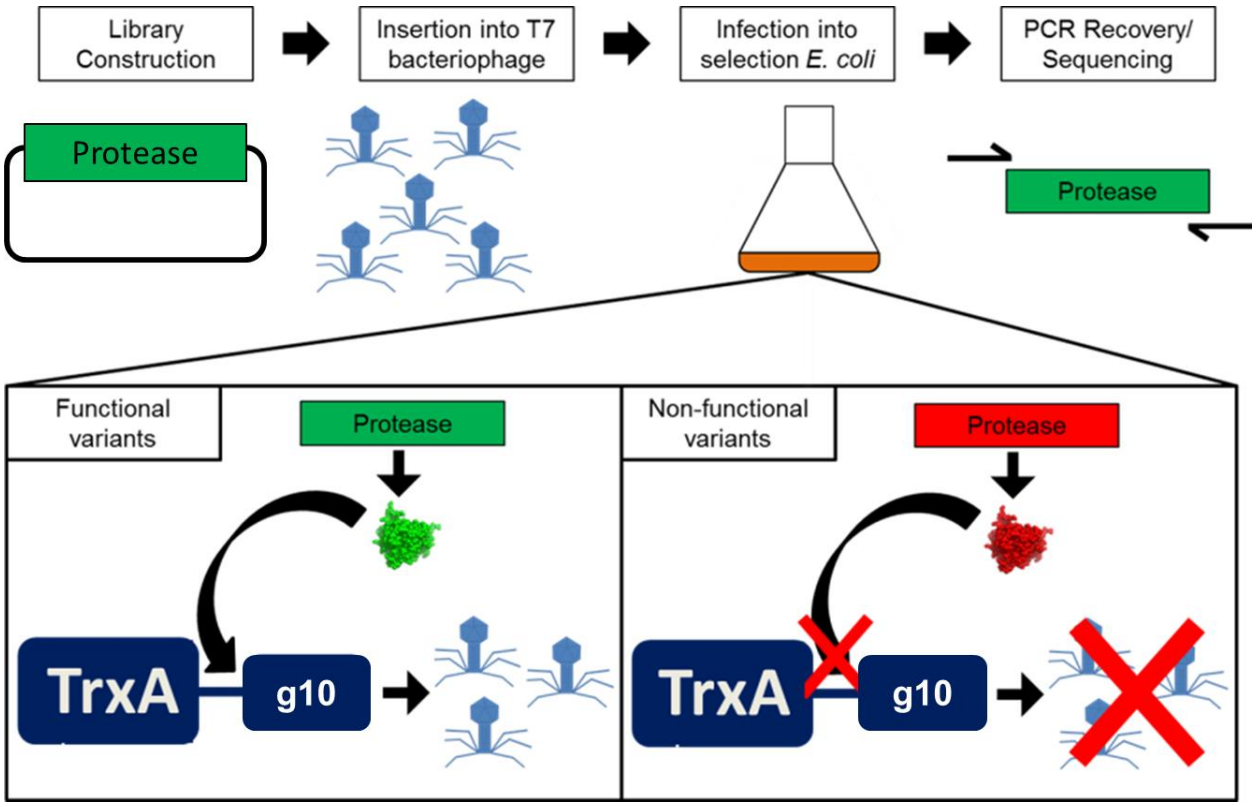


Figure 3.3: Scheme for T7Δ10-based protease system/selection. A protease of interested (or libraries thereof) is cloned into a plasmid with 50-100bp regions flanking each side of the desired insertion site in the T7Δ10 genome. T7Δ10 phages will infect *E. coli* cells permissive to its replication (has gene 10 supplied by another plasmid) and a fraction of these will recombine such that the proteases are inserted into the T7 genome. These resulting cells are used to infect the *E. coli* selection plasmid where one of two events will occur: 1) a functional protease will cleave the protease-cleavage-site linker between the major capsid protein and theoredoxin A (TrxA) allowing successful phage replication to occur or 2) non-functional proteases will not cleave the protease-cleavage-site linker and not allow major capsid proteins to assemble and allow for phage replication. After a number of passages, the lysate (with the successfully replicating T7-protease phages) will

be subjected to PCR using primers flanking the proteases in the T7 genome to recover successful variants for sequencing or additional selection.

Table 3.2: OD_{600} readings of BL21 *E. coli* cells containing the gene 10 only, the gene 10-TEV-cleavage-site-TrxA fusion, or the TrxA-TEV-cleavage-site-gene 10 fusion constructs

		no phage -tet/-lac	+ phage -tet/-lac	+ phage -tet/+lac	+ phage +tet/-lac	+ phage +tet/+lac
pLUV-gene 10	30 min post I	1.47	1.53	1.47	1.51	1.46
	45 min post I	1.67	0.73	0.47	0.68	0.53
	60 min post I	1.82	0.26	0.24	0.19	0.2
	105 min post I	1.9	0.18	0.17	0.14	0.16
pLUV-gene 10-TrxA	30 min post I	1.62	1.68	1.7	1.68	1.75
	45 min post I	1.74	1.74	1.75	1.81	1.87
	60 min post I	1.77	1.91	1.86	1.83	1.93
	105 min post I	1.88	1.98	2+	1.93	2+
pLUV-TrxA-gene 10	30 min post I	1.5	1.65	1.66	1.64	1.75
	45 min post I	1.7	1.78	1.81	1.34	1.54
	60 min post I	1.81	1.87	1.96	0.42	0.61
	105 min post I	1.98	2+	2+	0.2	0.25

Table 3.2: Table showing the OD_{600} readings of BL21 *E. coli* cells containing the gene 10 only, the gene 10-TEV-cleavage-site-TrxA fusion, or the TrxA-TEV-cleavage-site-gene 10 fusion construct under the control of a LacUV5 promoter together with a pTET15A plasmid containing TEV protease under the pTetO promoter. O/N *E. coli* cells were diluted into 2mLs of 2xYT with proper antibiotics and grown for one hour to mid-log phage at which point phage were added. Time recorded in minutes after infection. + phage= MOI 0.1 T7Δ1 phage, +tet = induction of transcription from the pTetO promoter by 200ng/mL anhydrous tetracycline induction, +lac= induction of transcription from the LacUV5 promoter with 1mM IPTG induction.

Table 3.3: Mutations resulting from initial passaging of TEV-S219P

Isolate	Passage 10	Isolate	Passage 10	Isolate	Passage 10
1.1	N177T, P219S	2.1	N177T, P219S	3.1	N177T, P219S
1.2	N177T, P219S	2.2	N177T, P219S	3.2	N177T, P219S
1.3	N177T, P219S	2.3	N177T, P219S	3.3	N177T, P219S
1.4	N177T, P219S	2.4	N177T, P219S	3.4	N177T, P219S
1.5	N177T, P219S	2.5	N177T, P219S	3.5	N177T, P219S
1.6	N177T, P219S	2.6	N177T, P219S	3.6	N177T, P219S
1.7	N177T, P219S	2.7	N177T, P219S	3.7	N177T, P219S
1.8	N177T, P219S	2.8	N177T, P219S	3.8	N177T, P219S
1.9	R50H, N177T, P219S	2.9	N177T, P219S	3.9	N177T, P219S
1.10	N177T, P219S	2.10	N177T, P219S	3.10	N177T, P219S

Table 3.3: Mutations resulting from initial passaging of TEV-S219P on 1mM induced pLUV-TrxA-gene-10 for 10 passages

Table 3.4: Mutations resulting from passaging TEV-S219P in presence of Flanked TEV-Fast

Isolate	Passage 10	Isolate	Passage 10	Isolate	Passage 10
1.1	N177T, P219S	2.1	N177T, P219S	3.1	P219S
1.2	N177T, P219S	2.2	N177T, P219S	3.2	N177T, P219S
1.3	N177T, P219S	2.3	N177T, P219S	3.3	N177T, P219S
1.4	N177T, P219S	2.4	P219S	3.4	P219S
1.5	N177T, P219S	2.5	N177T, P219S	3.5	P219S
1.6	N177T, P219S	2.6	N177T, P219S	3.6	P219S
1.7	N177T, P219S	2.7	P219S	3.7	N177T, P219S
1.8	N177T, P219S	2.8	N177T, P219S	3.8	N177T, P219S
1.9	N177T, P219S	2.9	P219S	3.9	N177T, P219S
1.10	N177T, P219S	2.10	N177T, P219S	3.10	P219S

Table 3.4: Mutations resulting from passaging TEV-S219P in presence of Flanked TEV-Fast for five passages and then five more passages on 1mM induced pLUV-TrxA-gene-10, MOI 0.1

Table 3.5: Mutations resulting from passaging TEV-S219V

Isolate	Passage 10	Isolate	Passage 20
1	N/A	1	N177I
2	N/A	2	R51H, N177I
3	N/A	3	R51H, N177I
4	N/A	4	R51H, N177I
5	N/A	5	N177T
6	N/A	6	R51H, N177I
7	N/A	7	R51H, N177I
8	N/A	8	R51H, N177I
9	N/A	9	R51H, N177I
10	N/A	10	R51H, N177I

Table 3.5: Mutations resulting from passaging TEV-S219V on 1mM induced pLUV-TrxA-gene-10, MOI 0.1 for first 10 passages at MOI 0.1 then 10 more passages on 100uM induced pLUV-TrxA-gene-10, MOI 0.01

Table 3.6: Mutations resulting from passaging TEV-Fast

Isolate	Passage 10	Isolate	Passage 20
1	N/A	1	N177T
2	N/A	2	N177T, S219G
3	N/A	3	N177T
4	N/A	4	N177T
5	N/A	5	N177T
6	N/A	6	N177T
7	N/A	7	N177T
8	N/A	8	N177T
9	N/A	9	N177T
10	N/A	10	N177T

Table 3.6: Mutations resulting from passaging TEV-Fast on 1mM induced pLUV-TrxA-gene-10, MOI 0.1 for first 10 passages at MOI 0.1 then 10 more passages on 100uM induced pLUV-TrxA-gene-10, MOI 0.01

Figure 3.4: PyMOL rendering of the TEV protease

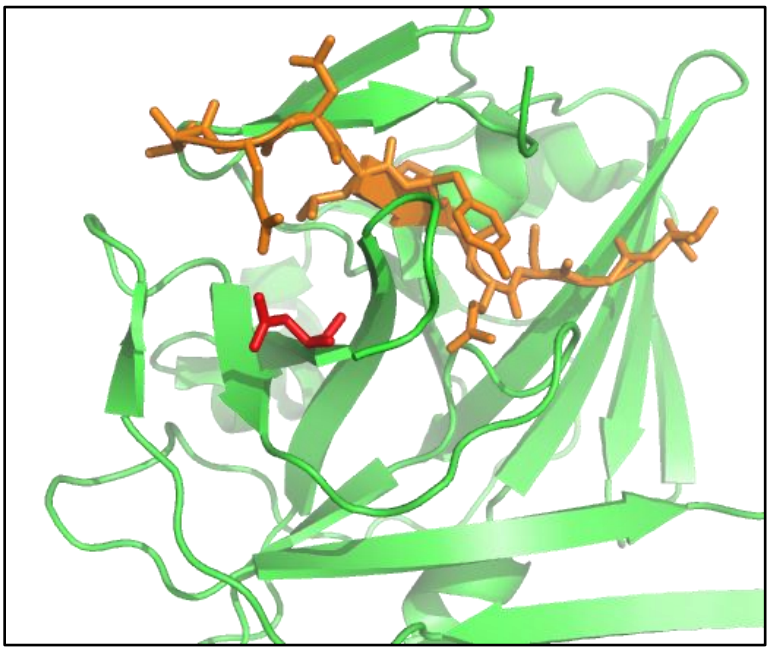


Figure 3.4: PyMOL rendering of the TEV protease (green) complexed with its canonical substrate (orange). The Asn177 residue (red) in TEV protease interacts with the P6 Glu of TEV protease substrate (orange). The N177T mutant is hypothesized to stabilize the binding of TEV protease to its substrate. (PDB ID: 1LVB)

Figure 3.5: Lysis curves comparing various TEV-phage populations at various passages

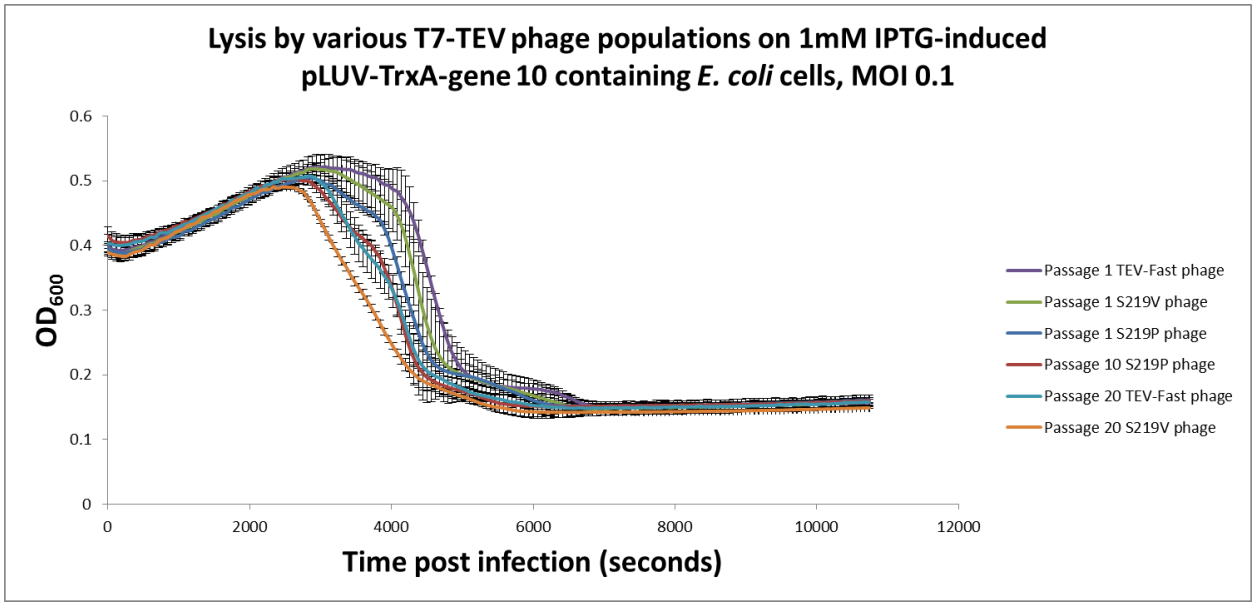


Figure 3.5: Lysis curves comparing various TEV-phage populations at various passages. Above, freshly titered phages used at an MOI of 0.1 to infect 1mM IPTG-induced pLUV-TrxA-gene 10 BL21 *E. coli* cells. Error bars represent standard deviations.

Figure 3.6: Lysis curves comparing various TEV-phage populations at various passages

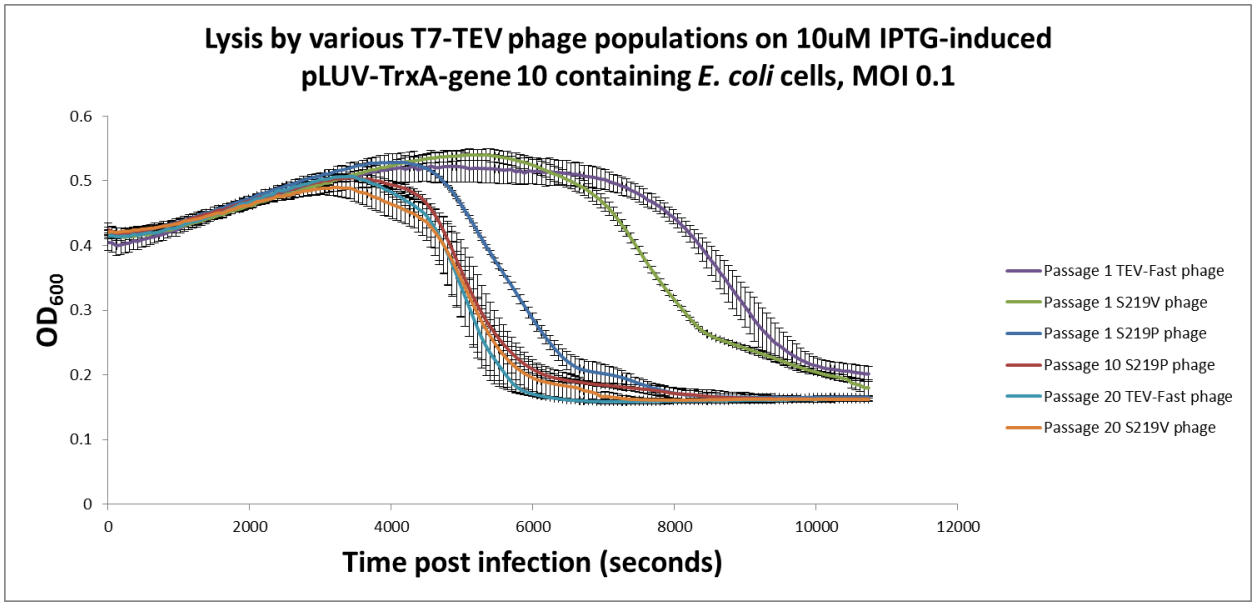


Figure 3.6: Lysis curves comparing various TEV-phage populations at various passages. Above, freshly titered phages used at an MOI of 0.1 to infect 10uM IPTG-induced pLUV-TrxA-gene 10 BL21 *E. coli* cells. Error bars represent standard deviations.

Figure 3.7: Scheme depicting the progression from initial recombination of S219V into bacteriophage T7 to the creation of each EΔW-adapted, YΔE-adapted, YΔH-adapted, and QΔH-adapted T7-TEV phage populations.

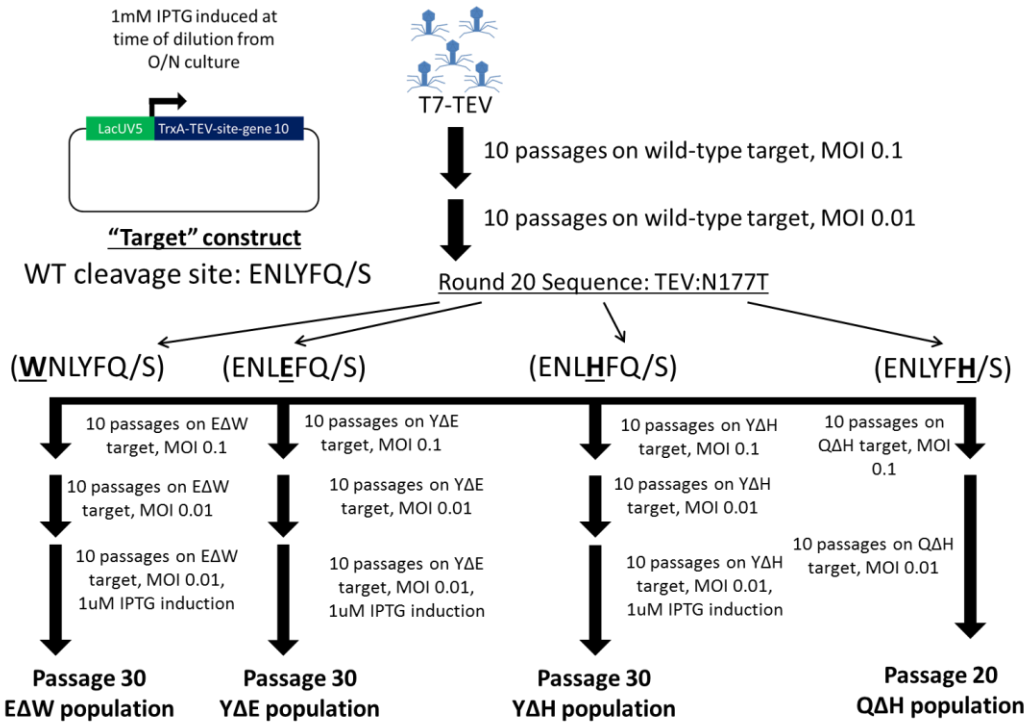


Figure 3.7: Scheme depicting the progression from initial recombination of S219V into bacteriophage T7 to the creation of each EΔW-adapted, YΔE-adapted, YΔH-adapted, and QΔH-adapted T7-TEV phage populations. These populations (with the exception of the QΔH-adapted one) followed the same path whereas 10 passages were done on each respective target-containing cells at 1mM ITPG induction and an MOI of 0.1, then 10 passages at 1mM IPTG induction and an MOI of 0.01, and finally 10 passages at 1uM IPTG induction and an MOI of 0.01. The QΔH-adapted population was not able to lyse target cells at 1uM IPTG induction and thus the series of passages were omitted.

Figure 3.8: Lysis curves of wild-type (S219V) TEV-adapted phage population on various single-point mutation TEV protease targets

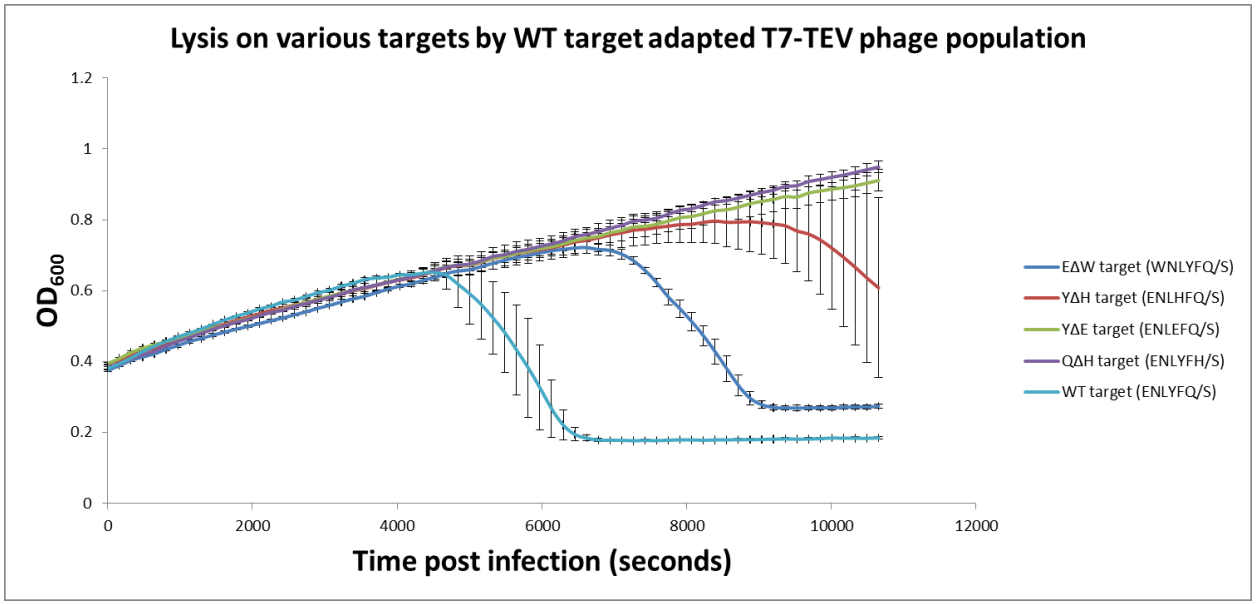


Figure 3.8: Lysis curves of wild-type (here, defined as S219V) TEV-adapted phage population on various single-point mutation TEV protease targets. Freshly titered phages used at an MOI of 0.1 to infect 1mM IPTG-induced pLUV-TrxA-gene 10 BL21 *E. coli* cells. Error bars represent standard deviations.

Figure 3.9: Lysis curves of EΔW TEV-adapted phage population on various single-point mutation TEV protease targets

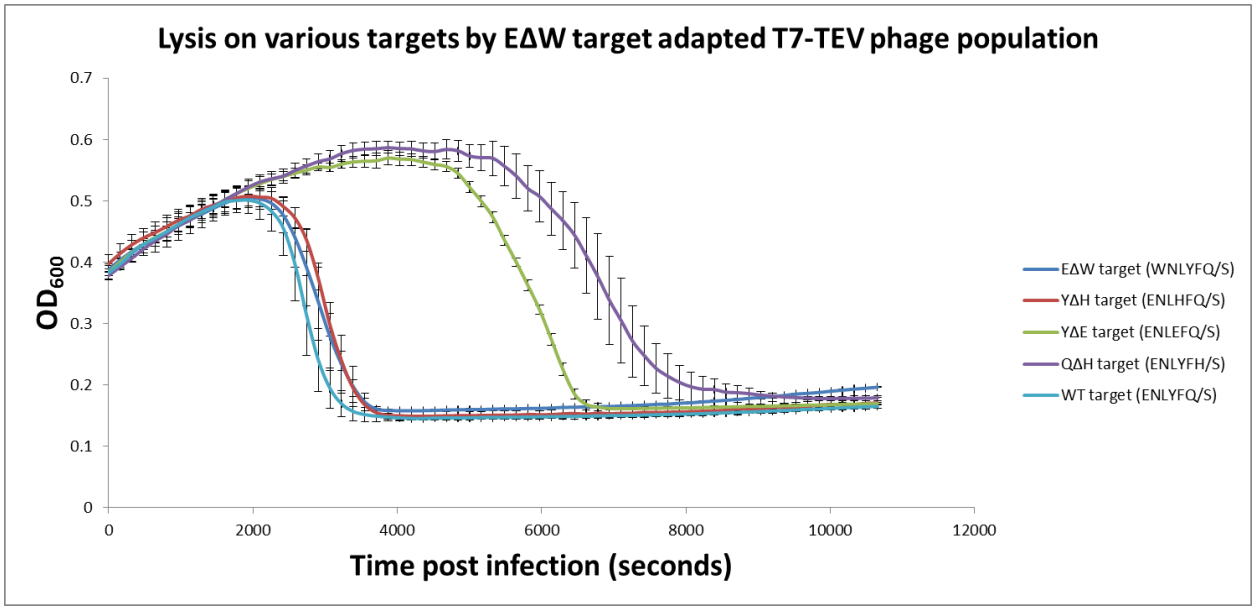


Figure 3.9: Lysis curves of EΔW TEV-adapted phage population on various single-point mutation TEV protease targets. Freshly titered phages used at an MOI of 0.1 to infect 1mM IPTG-induced pLUV-TrxA-gene 10 BL21 *E. coli* cells. Error bars represent standard deviations.

Figure 3.10: Lysis curves of YΔH TEV-adapted phage population on various single-point mutation TEV protease targets

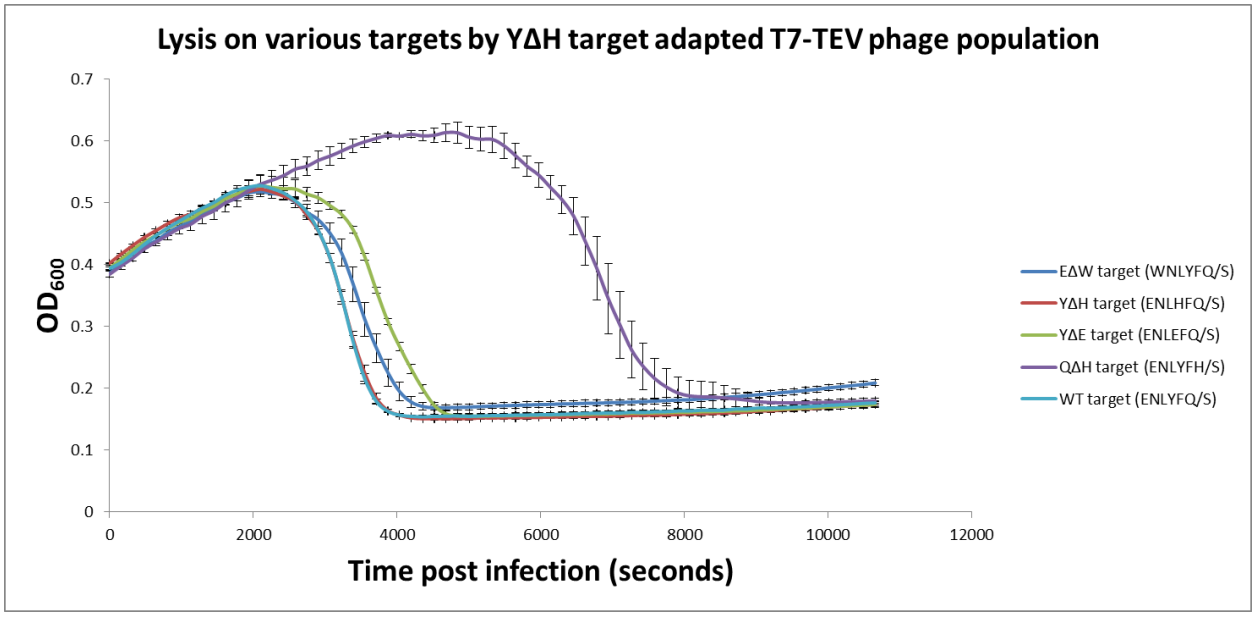


Figure 3.10: Lysis curves of YΔH TEV-adapted phage population on various single-point mutation TEV protease targets. Freshly titered phages used at an MOI of 0.1 to infect 1mM IPTG-induced pLUV-TrxA-gene 10 BL21 *E. coli* cells. Error bars represent standard deviations.

Figure 3.11: Lysis curves of YΔE TEV-adapted phage population on various single-point mutation TEV protease targets

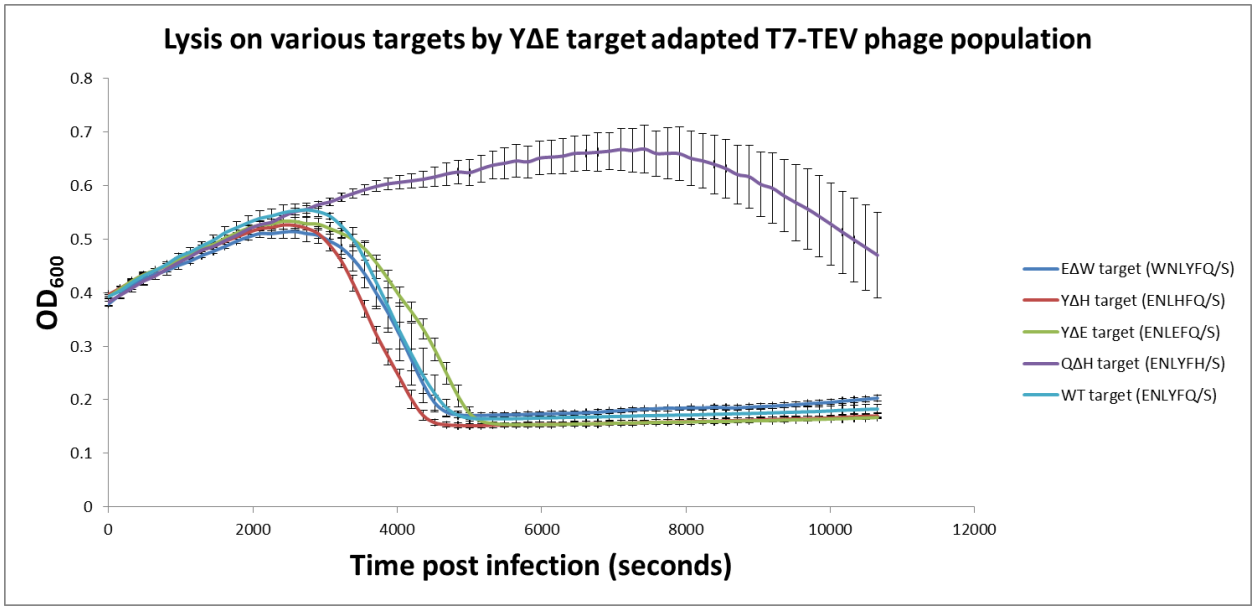


Figure 3.11: Lysis curves of YΔE TEV-adapted phage population on various single-point mutation TEV protease targets. Freshly titered phages used at an MOI of 0.1 to infect 1mM IPTG-induced pLUV-TrxA-gene 10 BL21 *E. coli* cells. Error bars represent standard deviations.

Figure 3.12: Lysis curves of QΔH TEV-adapted phage population on various single-point mutation TEV protease targets

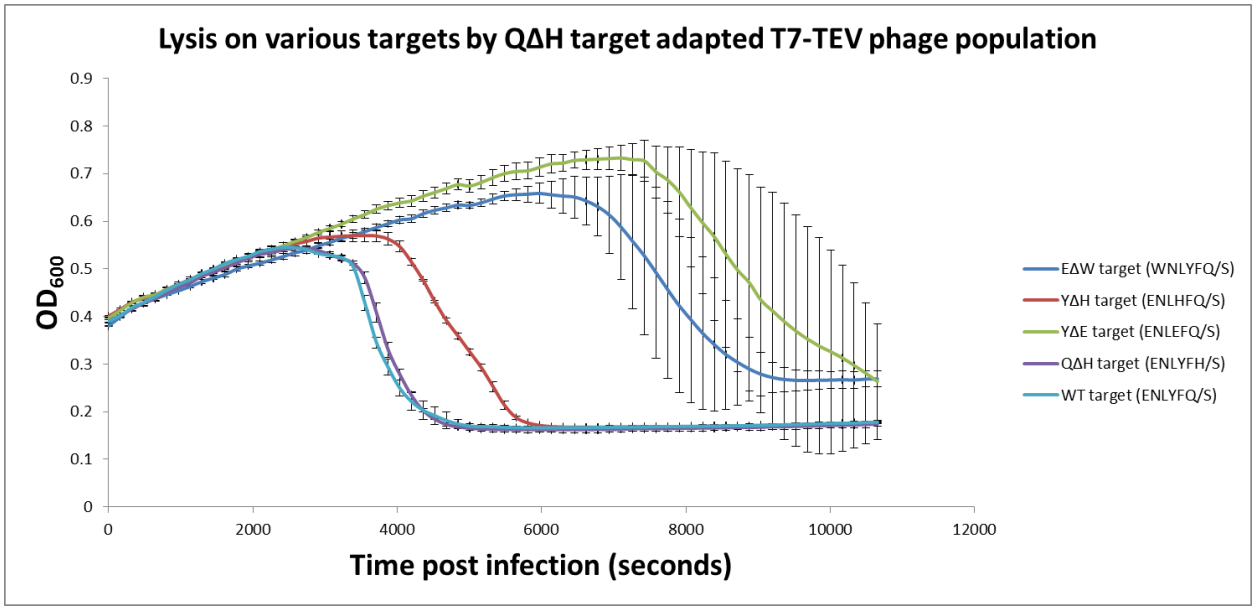


Figure 3.12: Lysis curves of QΔH TEV-adapted phage population on various single-point mutation TEV protease targets. Freshly titered phages used at an MOI of 0.1 to infect 1mM IPTG-induced pLUV-TrxA-gene 10 BL21 *E. coli* cells. Error bars represent standard deviations.

Table 3.7: Mutations in individual T7-TEV phage isolates from EΔW, YΔH, YΔE, and QΔH TEV-adapted populations.

Passage 30 EΔW population isolates

Isolate	Passage 30
1	R50H, S154N, N177T, W212L, S219V
2	R50H, S154N, N177T, W212L, S219V
3	R50H, S154N, S171T, N177T, S219V
4	R50H, I139T, S154N, S171T, N177T, S219V
5	R50H, T147A, D149E, S154N, N177T, W212L, S219V
6	R50H, S154N, N177T, W212L, S219V
7	R50H, S154N, N177T, W212L, S219V
8	R50H, S154H, S171T, N177T, S219V
9	R50H, D149N, S154N, S171T, N177T, S219V
10	bad reads

Passage 30 YΔE population isolates

Isolate	Passage 30
1	R50H, S154N, S171T, N177T, W212L, S219V
2	R50H, I139T, S154N, S171T, N177T, S219V
3	R50H, T147A, D149E, S151N, N177T, W212L, S219V
4	R50H, S154N, S171T, N177T, S219V
5	R50H, S154N, N177T, W212L, S219V
6	R50H, S154N, S171T, N177T, S219V
7	bad reads
8	R50H, T147A, D149E, S151N, N177T, W212L, S219V
9	R50H, T147A, D149E, S151N, N177T, W212L, S219V
10	R51H, T147A, D149E, N177T, W212L, S219V

Passage 30 YΔH population isolates

Isolate	Passage 30
1	bad reads
2	R50H, S154N, N177T, W212L, S219V
3	R50H, S136F, N177I, S219V
4	R50H, S154N, S171T, N177T, S219V
5	R50H, D149N, S154N, S171T, S219V
6	R51H, T147A, D149E, N177T, W212L, S219V
7	R50H, S154N, S171T, N177T, N177T, S219V
8	R50H, S154N, N177T, W212L, S219V
9	bad reads
10	R50H, I139T, S154N, S171T, N177T, S219V

Passage 20 QΔH population isolates

Isolate	Passage 20
1	R50H, D149N, S154N, S171T, N177T, S219V
2	R50H, S154N, N177T, 212L, S129V
3	R50H, S154N, S171T, N177T, S219V
4	R50H, I139T, S154N, S171T, N177T, S219V
5	R50H, T147A, D149E, S154N, N177T, W212L, S219V
6	bad reads
7	R50H, S136F, S154N, N177T, W212L, S129V
8	R50H, S136F, S154N, N177T, W212L, S129V
9	R50H, S154N, N177T, 212L, S129V, W212L, S219V
10	R51H, T147A, D149E, N177T, W212L, S219V

Table 3.7: Mutations in individual T7-TEV phage isolates from EΔW, YΔH, YΔE, and QΔH TEV-adapted populations.

Table 3.8: OD_{600} readings of BL21 *E. coli* cells containing the TrxA-3C-cleavage-site-gene 10 fusion construct

		no phage -tet/-lac	+ phage -tet/-lac	+ phage -tet/+lac	+ phage +tet/-lac	+ phage +tet/+lac
pPOS-TrxA-gene 10-3C	0 min post I	1.2	1.06	1	0.86	0.87
	45 min post I	2+	1.93	1.93	1.69	0.76
	60 min post I	2+	2+	2+	1.85	0.19

Table 3.8: Table showing the OD_{600} readings of BL21 *E. coli* cells containing the TrxA-3C-cleavage-site-gene 10 fusion construct under the control of a LacUV5 promoter together with a pTET15A plasmid containing rhinovirus 3C protease under the pTetO promoter. O/N *E. coli* cells were diluted into 2mLs of 2xYT with proper antibiotics and grown for one hour to mid-log phage at which point phage were added. Time recorded in

minutes after infection. + phage= MOI 0.1 T7Δ10 phage, +tet = induction of transcription from the pTetO promoter by 200ng/mL anhydrous tetracycline induction, +lac= induction of transcription from the LacUV5 promoter with 1mM IPTG induction.

Figure 3.13: Picture of 3C protease and 3C target induced T7Δ10 infected *E. coli* cells

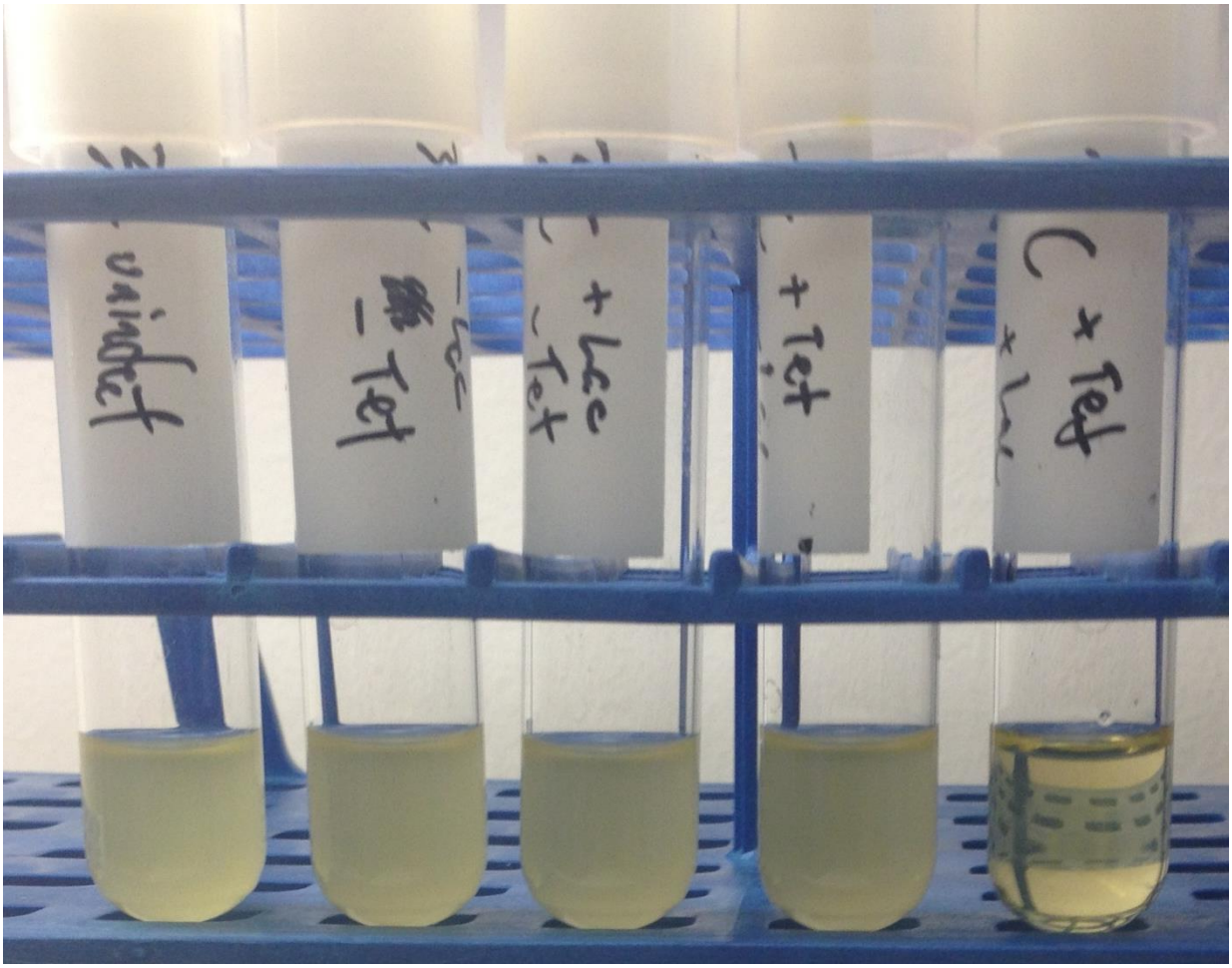


Figure 3.13: Picture of 3C protease and 3C target induced T7Δ10 infected *E. coli* cells in order as listed in the figure above at one hour post infection.

Figure 3.14: Scheme depicting optimal orthogonality of T7-protease phage life cycling

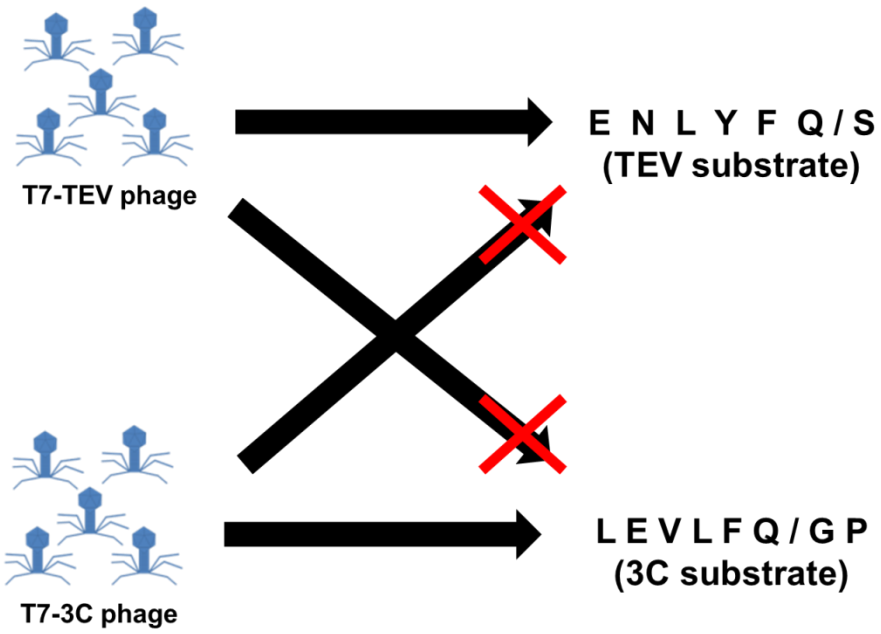


Figure 3.14: Scheme depicting optimal orthogonality of T7-protease phage life cycling. Hypothetically, T7-protease dependent replication can occur in a protease target-dependent manner which allows one T7-protease type to replicate in cells containing its cognate target construct while not replicating in cells containing the other's target construct.

Figure 3.15: Lysis curves of T7-TEV and T7-3C phages on either TEV target or 3C target

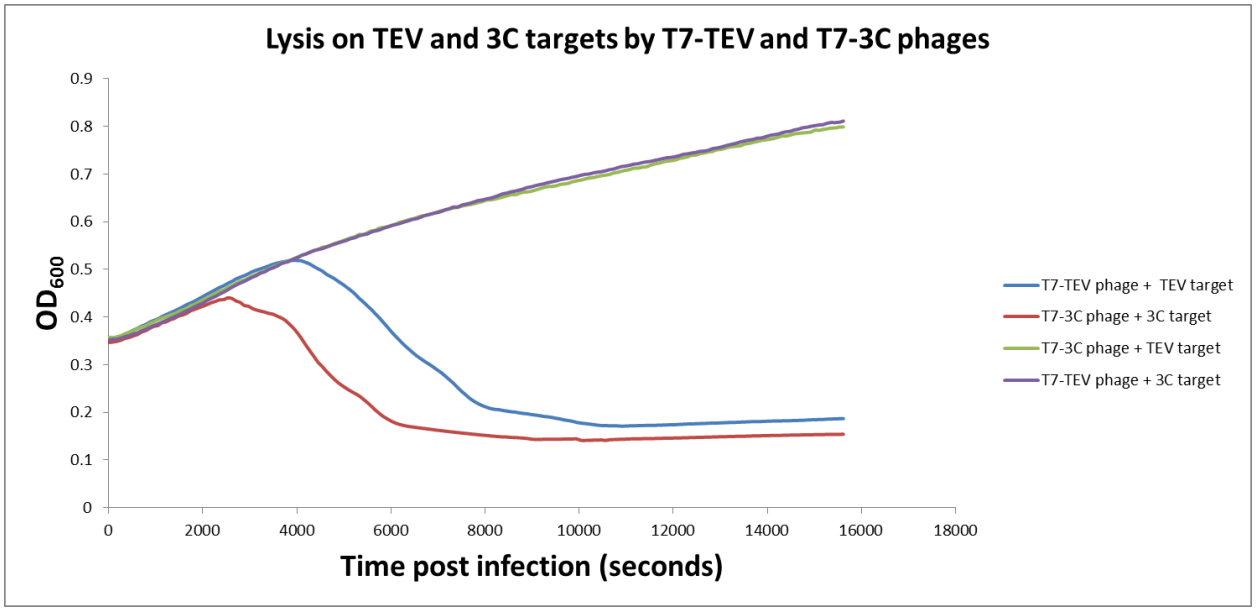


Figure 3.15: Lysis curves of T7-TEV and T7-3C phages on either TEV target or 3C target. Freshly titered phages used at an MOI of 0.1 to infect 1mM IPTG-induced pLUV-TrxA-gene 10 BL21 *E. coli* cells. Results shown here are in singlicate though the same pattern of lysis holds for the range of MOIs (see Figures 3.16 and 3.17).

Figure 3.16: Lysis curves for T7-TEV phage on 1mM induced TEV target or 3C target at various MOIs

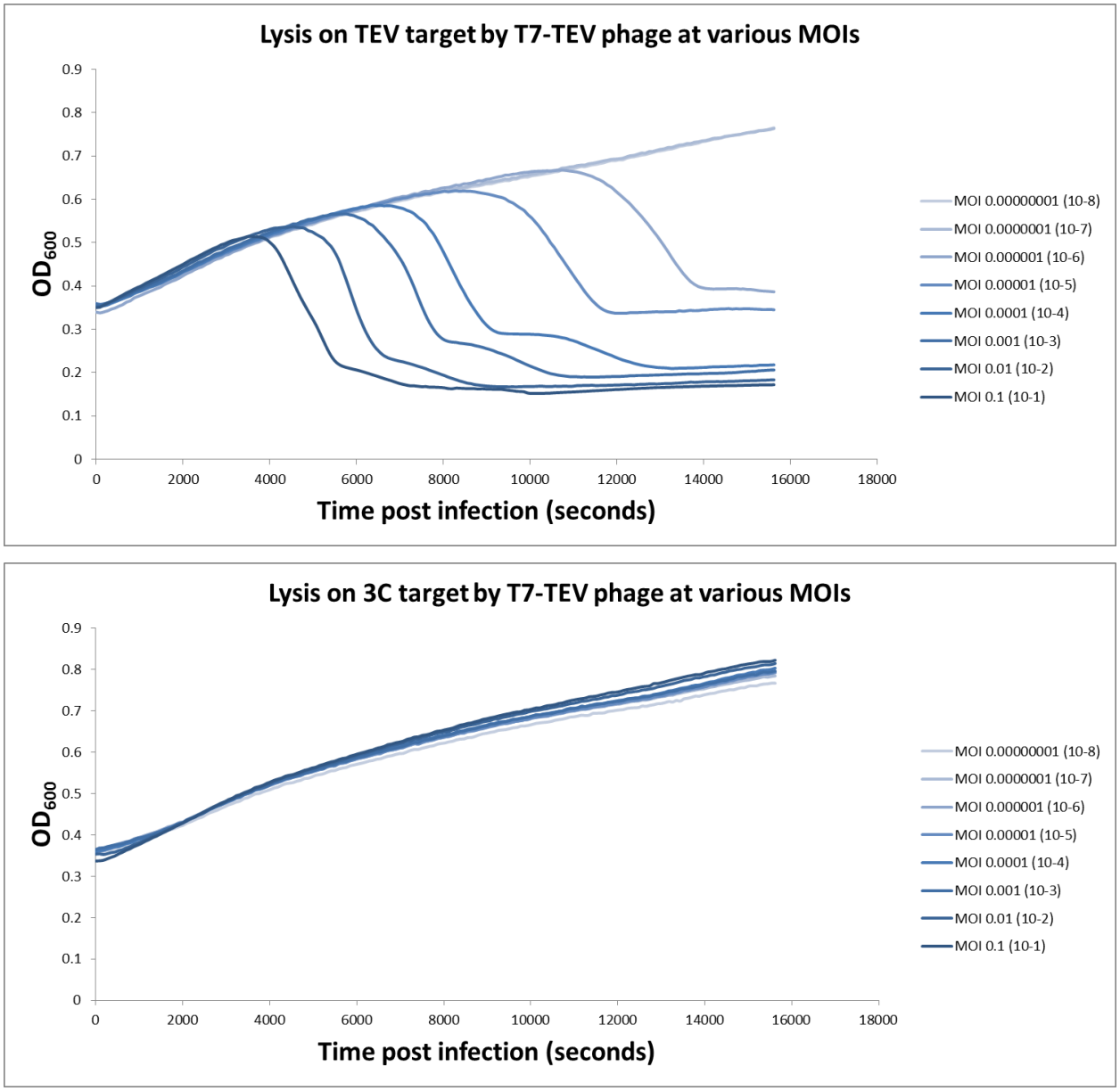


Figure 3.16: Lysis curves for T7-TEV phage on 1mM induced TEV target (above) or 3C target (below) at various MOIs.

Figure 3.17: Lysis curves for T7-3C phage on 1mM induced 3C target or TEV target at various MOIs

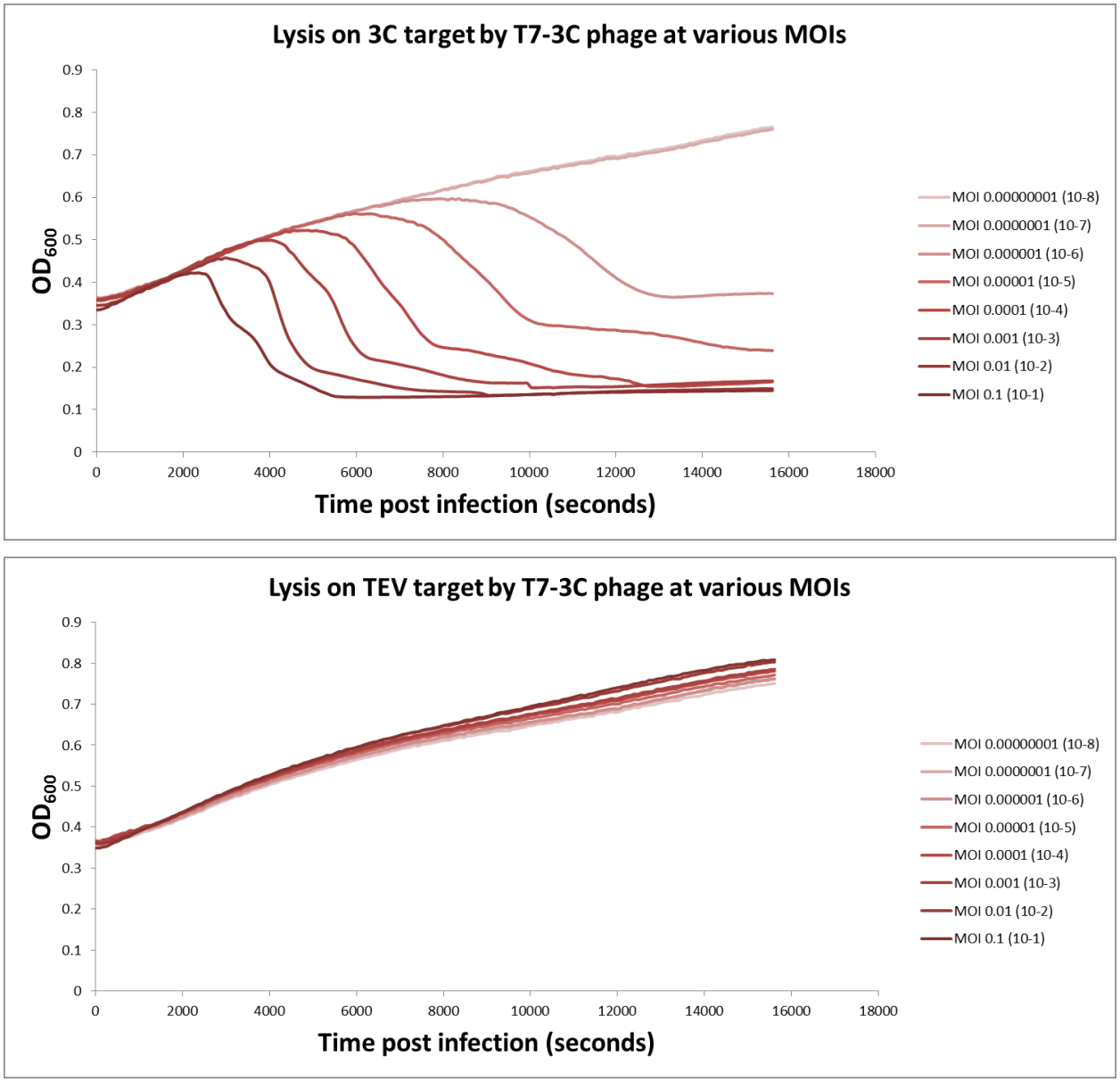


Figure 3.17: Lysis curves for T7-3C phage on 1mM induced 3C target (above) or TEV target (below) at various MOIs.

CONCLUSION

This work describes the findings from three projects in which arbitrarily engineered parts or circuits were used as design constraints to shape and study the evolutionary trajectory of bacteriophage T7. In the first project, we demonstrated that is was possible to predict much of the evolutionary trajectory of bacteriophage promoter identity mutations in response to a synthetically-created RNA polymerase. In the second project, we showed that bacteriophage populations altered their transcriptional apparatus such that the synthetic polymerase (rather than promoters) picked up subsequent mutations and in doing so reached the same outcome of promoter generality in two different ways. In the third project, a system of bacteriophage T7 life-cycling dependent on foreign protease activity was created and explored, eventually showing extensibility which enabled selective cell killing. Overall, these projects are another step in the direction of utilizing the rational design and synthesis tools associated with synthetic biology together with the power of evolutionary pressure in order to drive novel viral creation, evolution, and synthetic viromics.

This work shows how the use of laboratory evolution and directed evolution can lead to unexpected outcomes. This work demonstrates both the power and speed of bacteriophage T7 evolution in the laboratory setting but also some of the limitations to its plasticity. This work is a small addition to the work of many others which demonstrates how bacteriophages, an important biological entity since the founding of molecular biology, together with the newest tools and techniques to read, write, and edit genomes, will continue to be relevant and useful for evolutionary biology and biotechnology well into the future. Finally, in the unending tension between reductionism and holism, this work provides evidence that although the ability to rationally design biology is important,

various ways to rationally control and direct evolution offer a complementary strategy to the systematization of synthetic biology.

Appendix

Bacteriophage T7 by the Numbers

It is sometimes helpful to remind oneself of the quantitative aspects of an organism of study, specifically, regarding information such as its genome size, number of protein coding domains, and life cycle. This section lists a few of the interesting numbers associated with the bacteriophage T7 and derives a few numerical answers to relevant questions which, although assuming much and oversimplifying to allow back-of-the-envelope calculations, lead to interesting and potentially useful conclusions.

Goal: Determine…

1) Rate of T7 DNA replication relative to PCR:

PCR: 15-60 seconds/kB → 17-67 bp/sec, requires thermocycling

T7 bacteriophage: 220 (+/-) 80 bp/sec, isothermal (30-37 °C) (Lee, et al. 2006)

2) How quickly bacteriophage T7 would lyse a mid-log phage culture of E. coli with the volume of the world's oceans:

Assume: Volume of world's oceans = 1.3324×10^9 km3 (Smith 2010), km^3 seawater=10^{12} liters,

Assume: *E. coli* cells @ OD~1= 10^8 cells/mL (10^{11} cells/L)

Assume: Bacteriophage T7 burst size = 260 (De Paepe and Taddei 2006) and lysis time is ~ 30 minutes

Assume: Well-mixed ocean, with no co/super-infection, such that one phage perfectly enters the next uninfected cell at a time immediately after previous cell lysis

1.3324×10^9 km3 x 10^{12} liters = 1.334×10^{21} liters in world's oceans

$1.334x10^{21}$ liters x 10^{11} cells/liter = $1.334x10^{32}$ total *E. coli* cells

The amount of infectious "cycles" needed to lyse these cells can be determined from the following equation:

(Burst size x Number of starting phage)"cycles" of infection = total number of cells to be killed

Starting with one bacteriophage T7,

(260 x 1) "cycles" of infection = $1.334x10^{32}$

"cycles" of infection x log (260) = log ($1.334x10^{32}$)

"cycles" of infection = log ($1.334*10^{32}$) / log (260)

"cycles" of infection = 13.3 → 14 "cycles" of infection

14 "cycles" of infection x (30 minutes / "cycle" of infection) = 420 minutes = 7 hours

Calculation of end-to-end length of all of those T7 genomes

Assume: T7 genome ~ 40kB (Calendar 2006), each base pair is 0.34nm in length (Watson and Crick 1953) and the diameter of the observable universe = 93 billion light years ($8.798x10^{23}$ km) (Bars, et al. 2009)

Number of total T7 genomes = number of total cells x burst size

Number of total T7 genomes = ($1.334x10^{32}$) x 260 = $3.46x10^{34}$

10bp = linear length of 3.4nm

10kB = 3.4um x 4 = 13.6um = 1.36 x 10^{-5} m = length of one T7 genome

Total genomes x length of T7 genome = end-to-end length of T7 genomes from the Earth's oceans volume of lysed *E. coli* cells

($3.46x10^{34}$ genomes) x (1.36 x 10^{-5} m/genome) = $4.71x10^{29}$ m = $4.71x10^{26}$ km

End-to-end length of T7 genomes from the Earth's oceans volume of lysed *E. coli* cells / Diameter of observable universe = ($4.71x10^{26}$ km)/($8.798x10^{23}$ km) = 535 times

3) The theoretical rate of sampling of sequence space as bacteriophage T7 lyses cells and naturally mutates:

Assume: T7 DNA polymerase error rate 10^{-8} (cited values 10^{-6} to 10^{-8}) (Kunkel 2004) and that mutations would occur evenly at all locations in the T7 genome

1 T7 phage enters 1 *E. coli* cell = 260 T7 phage progeny (i.e. burst size). So, 40,000bp x 260 phages = 1.04 x 10^7 bp replicated. Given the T7 error rate of 10^{-8}, there will be errors on the scale of 1 phage in every 10 cells that lyse. In a 10^8 *E. coli* cells/mL culture, complete lysis would generate a pool of approximately 10^{10} phage genomes which, collectively, would mutationally sample each nucleotide position 250 times (10^7 mutant phage with single error / 40,000bp = 250).

Bibliography

Citations in Introduction

Ackermann HW 2003. Bacteriophage observations and evolution. Res Microbiol 154: 245-251. doi: 10.1016/s0923-2508(03)00067-6

Adams MJ, Lefkowitz EJ, King AMQ, Harrach B, Harrison RL, Knowles NJ, Kropinski AM, Krupovic M, Kuhn JH, Mushegian AR, Nibert ML, Sabanadzovic S, Sanfaçon H, Siddell SG, Simmonds P, Varsani A, Zerbini FM, Orton RJ, Smith DB, Gorbalenya AE, Davison AJ 2017. 50 years of the International Committee on Taxonomy of Viruses: progress and prospects. Archives of Virology: 1-6. doi: 10.1007/s00705-016-3215-y

Ando H, Lemire S, Pires DP, Lu TK 2015. Engineering Modular Viral Scaffolds for Targeted Bacterial Population Editing. Cell Syst 1: 187-196. doi: 10.1016/j.cels.2015.08.013

Avery OT, MacLeod CM, McCarty M 1944. STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES : INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III. J Exp Med 79: 137-158.

Barbu EM, Cady KC, Hubby B 2016. Phage Therapy in the Era of Synthetic Biology. Cold Spring Harb Perspect Biol 8. doi: 10.1101/cshperspect.a023879

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315: 1709-1712. doi: 10.1126/science.1138140

Bars I, Krauss L, Nekoogar F, Terning J. 2009. Extra Dimensions in Space and Time: Springer New York.

Bartlett JM, Stirling D 2003. A short history of the polymerase chain reaction. Methods Mol Biol 226: 3-6. doi: 10.1385/1-59259-384-4:3

Bazan J, Calkosinski I, Gamian A 2012. Phage display--a powerful technique for immunotherapy: 1. Introduction and potential of therapeutic applications. Hum Vaccin Immunother 8: 1817-1828. doi: 10.4161/hv.21703

Beadle GW, Tatum EL 1941. Genetic Control of Biochemical Reactions in Neurospora. Proc Natl Acad Sci U S A 27: 499-506.

Bikard D, Euler CW, Jiang W, Nussenzweig PM, Goldberg GW, Duportet X, Fischetti VA, Marraffini LA 2014. Exploiting CRISPR-Cas nucleases to produce

sequence-specific antimicrobials. Nat Biotechnol 32: 1146-1150. doi: 10.1038/nbt.3043

Bikard D, Jiang W, Samai P, Hochschild A, Zhang F, Marraffini LA 2013. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res 41: 7429-7437. doi: 10.1093/nar/gkt520

Biohelix 2017. T7 Gp4A helicase.

Biolabs NE 2017a. Bypassing common obstacles in protein expression.

T7 DNA ligase [Internet]. 2017b [cited 2017. Available from: https://www.neb.com/products/m0318-t7-dna-ligase

T7 DNA Polymerase (unmodified) [Internet]. 2017c [cited 2017. Available from: https://www.neb.com/products/m0274-t7-dna-polymerase-unmodified

T7 Endonuclease I [Internet]. 2017d. Available from: https://www.neb.com/products/m0302-t7-endonuclease-i

T7 Exonuclease [Internet]. 2017e [cited 2017. Available from: https://www.neb.com/products/m0263-t7-exonuclease

T7 RNA Polymerase [Internet]. 2017f [cited 2017. Available from: https://www.neb.com/products/m0251-t7-rna-polymerase

Bosch F, Rosich L 2008. The contributions of Paul Ehrlich to pharmacology: a tribute on the occasion of the centenary of his Nobel Prize. Pharmacology 82: 171-179. doi: 10.1159/000149583

Buchholz K, Collins J 2013. The roots--a short history of industrial microbiology and biotechnology. Appl Microbiol Biotechnol 97: 3747-3762. doi: 10.1007/s00253-013-4768-2

Bull JJ, Cunningham, C.W., Molineux, I.J., Badgett, M.R., Hillis, D.M. 1993. Experimental molecular evolution of bacteriophage T7. International Journal of Organic Evolution 47: 993-1007.

Bull JJ, Molineux IJ 2008. Predicting evolution from genomics: experimental evolution of bacteriophage T7. Heredity (Edinb) 100: 453-463. doi: 10.1038/sj.hdy.6801087

Bull JJ, Molineux IJ, Wilke CO 2012. Slow fitness recovery in a codon-modified viral genome. Mol Biol Evol 29: 2997-3004. doi: 10.1093/molbev/mss119

Bull JJ, Springman R, Molineux IJ 2007. Compensatory evolution in response to a novel RNA polymerase: orthologous replacement of a central network gene. Mol Biol Evol 24: 900-908. doi: 10.1093/molbev/msm006

Cairns J. 2007. Phage and the origins of molecular biology. Cold Spring Harbor, N.Y. : Cold Spring Harbor Laboratory Press.

Calendar R. 2006. The Bacteriophages: Oxford University Press, USA.

Cameron DE, Bashor CJ, Collins JJ 2014. A brief history of synthetic biology. Nat Rev Microbiol 12: 381-390. doi: 10.1038/nrmicro3239

Campbell JL, Richardson CC, Studier FW 1978. Genetic recombination and complementation between bacteriophage T7 and cloned fragments of T7 DNA. Proc Natl Acad Sci U S A 75: 2276-2280.

Chan LY, Kosuri S, Endy D 2005. Refactoring bacteriophage T7. Mol Syst Biol 1: 2005.0018. doi: 10.1038/msb4100025

Citorik RJ, Mimee M, Lu TK 2014. Sequence-specific antimicrobials using efficiently delivered RNA-guided nucleases. Nat Biotechnol 32: 1141-1145. doi: 10.1038/nbt.3011

Crick F 1970. Central dogma of molecular biology. Nature 227: 561-563.

Cunningham CW, Jeng K, Husti J, Badgett M, Molineux IJ, Hillis DM, Bull JJ 1997. Parallel molecular evolution of deletions and nonsense mutations in bacteriophage T7. Mol Biol Evol 14: 113-116.

D'Herelle F 2007. On an invisible microbe antagonistic toward dysenteric bacilli: brief note by Mr. F. D'Herelle, presented by Mr. Roux. 1917. Res Microbiol 158: 553-554. doi: 10.1016/j.resmic.2007.07.005

De Paepe M, Taddei F 2006. Viruses' life history: towards a mechanistic basis of a trade-off between survival and reproduction among phages. PLoS Biol 4: e193. doi: 10.1371/journal.pbio.0040193

Delbruck M 1940a. ADSORPTION OF BACTERIOPHAGE UNDER VARIOUS PHYSIOLOGICAL CONDITIONS OF THE HOST. J Gen Physiol 23: 631-642.

Delbruck M 1940b. THE GROWTH OF BACTERIOPHAGE AND LYSIS OF THE HOST. J Gen Physiol 23: 643-660.

Delbruck M 1970. A physicist's renewed look at biology: twenty years later. Science 168: 1312-1315.

Delbrück M, Bailey W editors. Cold Spring Harbor Symposia on Quantitative Biology. 1946.

Demerec M, Fano U 1945. Bacteriophage-Resistant Mutants in Escherichia Coli. Genetics 30: 119-136.

Dunn JJ, Studier FW 1983. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. J Mol Biol 166: 477-535.

Edgar RS, Feynman RP, Klein S, Lielausis I, Steinberg CM 1962. Mapping experiments with r mutants of bacteriophage T4D. Genetics 47: 179-186.

Ellefson JW, Meyer AJ, Hughes RA, Cannon JR, Brodbelt JS, Ellington AD 2014. Directed evolution of genetic parts and circuits by compartmentalized partnered replication. Nat Biotechnol 32: 97-101. doi: 10.1038/nbt.2714

Ellis EL, Delbruck M 1939. THE GROWTH OF BACTERIOPHAGE. J Gen Physiol 22: 365-384.

Fiers W, Contreras R, Duerinck F, Haegeman G, Iserentant D, Merregaert J, Min Jou W, Molemans F, Raeymaekers A, Van den Berghe A, Volckaert G, Ysebaert M 1976. Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. Nature 260: 500-507.

Forterre P 2006. Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: a hypothesis for the origin of cellular domain. Proc Natl Acad Sci U S A 103: 3669-3674. doi: 10.1073/pnas.0510333103

Gardner TS, Hawkins K 2013. Synthetic biology: evolution or revolution? A co-founder's perspective. Curr Opin Chem Biol 17: 871-877. doi: 10.1016/j.cbpa.2013.09.013

Ghadessy FJ, Ong JL, Holliger P 2001. Directed evolution of polymerase function by compartmentalized self-replication. Proc Natl Acad Sci U S A 98: 4552-4557. doi: 10.1073/pnas.071052198

Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi ZQ, Segall-Shapiro TH, Calvey CH, Parmar PP, Hutchison CA, 3rd, Smith HO, Venter JC 2010. Creation of a bacterial cell controlled by a chemically synthesized genome. Science 329: 52-56. doi: 10.1126/science.1190719

Hammerling MJ, Ellefson JW, Boutz DR, Marcotte EM, Ellington AD, Barrick JE 2014. Bacteriophages use an expanded genetic code on evolutionary paths to higher fitness. Nat Chem Biol 10: 178-180. doi: 10.1038/nchembio.1450

Hatfull GF, Hendrix RW 2011. Bacteriophages and their genomes. Curr Opin Virol 1: 298-303. doi: 10.1016/j.coviro.2011.06.009

Heather JM, Chain B 2016. The sequence of sequencers: The history of sequencing DNA. Genomics 107: 1-8. doi: 10.1016/j.ygeno.2015.11.003

Hendrix RW, Smith MC, Burns RN, Ford ME, Hatfull GF 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc Natl Acad Sci U S A 96: 2192-2197.

Hershey AD 1946. Mutation of Bacteriophage with Respect to Type of Plaque. Genetics 31: 620-640.

Hershey AD, Chase M 1952. Independent functions of viral protein and nucleic acid in growth of bacteriophage. J Gen Physiol 36: 39-56.

Hillis DM, Bull JJ, White ME, Badgett MR, Molineux IJ 1992. Experimental phylogenetics: generation of a known phylogeny. Science 255: 589-592.

Housby JN, Mann NH 2009. Phage therapy. Drug Discov Today 14: 536-540. doi: 10.1016/j.drudis.2009.03.006

Hsu PD, Lander ES, Zhang F 2014. Development and applications of CRISPR-Cas9 for genome engineering. Cell 157: 1262-1278. doi: 10.1016/j.cell.2014.05.010

Hughes RA, Miklos AE, Ellington AD 2011. Gene synthesis: methods and applications. Methods Enzymol 498: 277-309. doi: 10.1016/b978-0-12-385120-8.00012-7

Integrated DNA Technologies I. 2017. gBlock Gene Fragments. In.

Keen EC 2012. Felix d'Herelle and our microbial future. Future Microbiol 7: 1337-1339. doi: 10.2217/fmb.12.115

Kiro R, Shitrit D, Qimron U 2014. Efficient engineering of a bacteriophage genome using the type I-E CRISPR-Cas system. RNA Biol 11: 42-44. doi: 10.4161/rna.27766

Kosuri S, Church GM 2014. Large-scale de novo DNA synthesis: technologies and applications. Nat Methods 11: 499-507. doi: 10.1038/nmeth.2918

Kruger DH, Schneck P, Gelderblom HR 2000. Helmut Ruska and the visualisation of viruses. Lancet 355: 1713-1717.

Krupovic M, Prangishvili D, Hendrix RW, Bamford DH 2011. Genomics of bacterial and archaeal viruses: dynamics within the prokaryotic virosphere. Microbiol Mol Biol Rev 75: 610-635. doi: 10.1128/mmbr.00011-11

Kunkel TA 2004. DNA replication fidelity. J Biol Chem 279: 16895-16898. doi: 10.1074/jbc.R400006200

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann Y, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH,

Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowki J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ, Szustakowki J 2001. Initial sequencing and analysis of the human genome. Nature 409: 860-921. doi: 10.1038/35057062

Lederberg J, Tatum EL 1946. Gene recombination in Escherichia coli. Nature 158: 558.

Lee JB, Hite RK, Hamdan SM, Xie XS, Richardson CC, van Oijen AM 2006. DNA primase acts as a molecular brake in DNA replication. Nature 439: 621-624. doi: 10.1038/nature04317

Lu TK, Collins JJ 2009. Engineered bacteriophage targeting gene networks as adjuvants for antibiotic therapy. Proc Natl Acad Sci U S A 106: 4629-4634. doi: 10.1073/pnas.0800442106

TypeOne Restriction Inhibitor [Internet]. 2017. Available from: http://www.lucigen.com/TypeOne-Restriction-Inhibitor/

Luria SE, Delbruck M 1943. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. Genetics 28: 491-511.

Luria SE, Delbruck M, Anderson TF 1943. Electron Microscope Studies of Bacterial Viruses. J Bacteriol 46: 57-77.

Maddox B 2003. The double helix and the 'wronged heroine'. Nature 421: 407-408. doi: 10.1038/nature01399

Meselson M, Stahl FW 1958. THE REPLICATION OF DNA IN ESCHERICHIA COLI. Proc Natl Acad Sci U S A 44: 671-682.

Nam KT, Kim DW, Yoo PJ, Chiang CY, Meethong N, Hammond PT, Chiang YM, Belcher AM 2006. Virus-enabled synthesis and assembly of nanowires for lithium ion battery electrodes. Science 312: 885-888. doi: 10.1126/science.1122716

Genome information by organism [Internet]. 8600 Rockville Pike, Bethesda MD, 20894 USA: National Center for Biotechnology Information, U.S. National Library of Medicine; 2017 [cited 2017. Available from: https://www.ncbi.nlm.nih.gov/genome/browse/

Nirenberg M 2004. Historical review: Deciphering the genetic code--a personal account. Trends Biochem Sci 29: 46-54. doi: 10.1016/j.tibs.2003.11.009

Norrby E 2008. Nobel Prizes and the emerging virus concept. Arch Virol 153: 1109-1123. doi: 10.1007/s00705-008-0088-8

Ozsolak F, Milos PM 2011. RNA sequencing: advances, challenges and opportunities. Nat Rev Genet 12: 87-98. doi: 10.1038/nrg2934

Packer MS, Liu DR 2015. Methods for the directed evolution of proteins. Nat Rev Genet 16: 379-394. doi: 10.1038/nrg3927

Pauling L, Corey RB, Branson HR 1951. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. Proc Natl Acad Sci U S A 37: 205-211.

Pedulla ML, Ford ME, Houtz JM, Karthikeyan T, Wadsworth C, Lewis JA, Jacobs-Sera D, Falbo J, Gross J, Pannunzio NR, Brucker W, Kumar V, Kandasamy J, Keenan L, Bardarov S, Kriakov J, Lawrence JG, Jacobs WR, Jr., Hendrix RW, Hatfull GF 2003. Origins of highly mosaic mycobacteriophage genomes. Cell 113: 171-182.

Pires DP, Cleto S 2016. Genetically Engineered Phages: a Review of Advances over the Last Decade. 80: 523-543. doi: 10.1128/mmbr.00069-15

Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152: 1173-1183. doi: 10.1016/j.cell.2013.02.022

Rohwer F 2003. Global phage diversity. Cell 113: 141.

Salmond GP, Fineran PC 2015. A century of the phage: past, present and future. Nat Rev Microbiol 13: 777-786. doi: 10.1038/nrmicro3564

Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, Hutchison CA, Slocombe PM, Smith M 1977. Nucleotide sequence of bacteriophage phi X174 DNA. Nature 265: 687-695.

Schoolnik GK, Summers WC, Watson JD 2004. Phage offer a real alternative. Nat Biotechnol 22: 505-506; author reply 506-507. doi: 10.1038/nbt0504-505

Shin J, Jardine P, Noireaux V 2012. Genome replication, synthesis, and assembly of the bacteriophage T7 in a single cell-free reaction. ACS Synth Biol 1: 408-413. doi: 10.1021/sb300049p

Smith MACWHF 2010. The Volume of Earth's Ocean Oceanography 23. doi: http://dx.doi.org/10.5670/oceanog.2010.51

Springman R, Badgett MR, Molineux IJ, Bull JJ 2005. Gene order constrains adaptation in bacteriophage T7. Virology 341: 141-152. doi: 10.1016/j.virol.2005.07.008

Springman R, Keller T, Molineux IJ, Bull JJ 2010. Evolution at a high imposed mutation rate: adaptation obscures the load in phage T7. Genetics 184: 221-232. doi: 10.1534/genetics.109.108803

Springman R, Molineux IJ, Duong C, Bull RJ, Bull JJ 2012. Evolutionary stability of a refactored phage genome. ACS Synth Biol 1: 425-430. doi: 10.1021/sb300040v

Steen H, Mann M 2004. The ABC's (and XYZ's) of peptide sequencing. Nat Rev Mol Cell Biol 5: 699-711. doi: 10.1038/nrm1468

Studier FW 1972. Bacteriophage T7. Science 176: 367-376.

Studier FW 1969. The genetics and physiology of bacteriophage T7. Virology 39: 562-574.

Summers WC 2001. Bacteriophage therapy. Annu Rev Microbiol 55: 437-451. doi: 10.1146/annurev.micro.55.1.437

Summers WC 1993. How Bacteriophage Came to Be Used by the Phage Group. Journal of the History of Biology 26: 255-267.

Tatum EL, Lederberg J 1947. Gene Recombination in the Bacterium Escherichia coli. J Bacteriol 53: 673-684.

Twort FW 2011. An investigation on the nature of ultra-microscopic viruses by Twort FW, L.R.C.P. Lond., M.R.C.S. (From the Laboratories of the Brown Institution, London). Bacteriophage 1: 127-129. doi: 10.4161/bact.1.3.16737

Watson JD 1950. The properties of x-ray inactivated bacteriophage. I. Inactivation by direct effect. J Bacteriol 60: 697-718.

Watson JD, Crick FH 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171: 737-738.

Yi L, Gebhard MC, Li Q, Taft JM, Georgiou G, Iverson BL 2013. Engineering of TEV protease variants by yeast ER sequestration screening (YESS) of combinatorial libraries. Proc Natl Acad Sci U S A 110: 7229-7234. doi: 10.1073/pnas.1215994110

Citations from chapter 1

Bohne AV, Teubner M, Liere K, Weihe A, Borner T 2016. In vitro promoter recognition by the catalytic subunit of plant phage-type RNA polymerases. Plant Mol Biol 92: 357-369. doi: 10.1007/s11103-016-0518-z

Bridgham JT, Carroll SM, Thornton JW 2006. Evolution of hormone-receptor complexity by molecular exploitation. Science 312: 97-101. doi: 10.1126/science.1123348

Bull JJ, Heineman RH, Wilke CO 2011. The phenotype-fitness map in experimental evolution of phages. PLoS One 6: e27796. doi: 10.1371/journal.pone.0027796

Bull JJ, Springman R, Molineux IJ 2007. Compensatory evolution in response to a novel RNA polymerase: orthologous replacement of a central network gene. Mol Biol Evol 24: 900-908. doi: 10.1093/molbev/msm006

Cermakian N, Ikeda TM, Miramontes P, Lang BF, Gray MW, Cedergren R 1997. On the evolution of the single-subunit RNA polymerases. J Mol Evol 45: 671-681.

Dean AM, Thornton JW 2007. Mechanistic approaches to the study of evolution: the functional synthesis. Nat Rev Genet 8: 675-688. doi: 10.1038/nrg2160

Deatherage DE, Barrick JE 2014. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Methods Mol Biol 1151: 165-188. doi: 10.1007/978-1-4939-0554-6_12

Dunn JJ, Studier FW 1983. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. J Mol Biol 166: 477-535.

Ellefson JW, Meyer AJ, Hughes RA, Cannon JR, Brodbelt JS, Ellington AD 2014. Directed evolution of genetic parts and circuits by compartmentalized partnered replication. Nat Biotechnol 32: 97-101. doi: 10.1038/nbt.2714

Garcia LR, Molineux IJ 1995. Rate of translocation of bacteriophage T7 DNA across the membranes of Escherichia coli. J Bacteriol 177: 4066-4076.

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP 2006. A single amino acid mutation contributes to adaptive beach mouse color pattern. Science 313: 101-104. doi: 10.1126/science.1126121

Lehming N, Sartorius J, Kisters-Woike B, von Wilcken-Bergmann B, Muller-Hill B 1990. Mutant lac repressors with new specificities hint at rules for protein--DNA recognition. Embo j 9: 615-621.

Mairhofer J, Scharl T, Marisch K, Cserjan-Puschmann M, Striedner G 2013. Comparative transcription profiling and in-depth characterization of plasmid-based and plasmid-free Escherichia coli expression systems under production conditions. Appl Environ Microbiol 79: 3802-3812. doi: 10.1128/aem.00365-13

Meyer AJ, Ellefson JW, Ellington AD 2015. Directed Evolution of a Panel of Orthogonal T7 RNA Polymerase Variants for in Vivo or in Vitro Synthetic Circuitry. ACS Synth Biol 4: 1070-1076. doi: 10.1021/sb500299c

Molineux I. 2006. The T7 Group. In: Abedon ST, editor. The Bacteriophages. New York, New York: Oxford University Press. p. 277-301.

Newcomb RD, Campbell PM, Ollis DL, Cheah E, Russell RJ, Oakeshott JG 1997. A single amino acid substitution converts a carboxylesterase to an organophosphorus hydrolase and confers insecticide resistance on a blowfly. Proc Natl Acad Sci U S A 94: 7464-7468.

Poelwijk FJ, Kiviet DJ, Weinreich DM, Tans SJ 2007. Empirical fitness landscapes reveal accessible evolutionary paths. Nature 445: 383-386. doi: 10.1038/nature05451

Rong M, He B, McAllister WT, Durbin RK 1998. Promoter specificity determinants of T7 RNA polymerase. Proc Natl Acad Sci U S A 95: 515-519.

T7 RiboMAX™ Express: Generation of 27kb in vitro Transcripts in Minutes [Internet]. Institute of Virology and Immunology, University of Würzburg, 97078 Würzburg, Germany: Promega Corporation Web site; 2002 [cited 2017.

Weinreich DM, Delaney NF, Depristo MA, Hartl DL 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. Science 312: 111-114. doi: 10.1126/science.1123539

Citations from chapter 2

Blount ZD, Barrick JE, Davidson CJ, Lenski RE 2012. Genomic analysis of a key innovation in an experimental Escherichia coli population. Nature 489: 513-518. doi: 10.1038/nature11514

Brophy JA, Voigt CA 2014. Principles of genetic circuit design. Nat Methods 11: 508-520. doi: 10.1038/nmeth.2926

Campbell JL, Richardson CC, Studier FW 1978. Genetic recombination and complementation between bacteriophage T7 and cloned fragments of T7 DNA. Proc Natl Acad Sci U S A 75: 2276-2280.

Cheetham GM, Steitz TA 1999. Structure of a transcribing T7 RNA polymerase initiation complex. Science 286: 2305-2309.

Chen Z, Schneider TD 2005. Information theory based T7-like promoter models: classification of bacteriophages and differential evolution of promoters and their polymerases. Nucleic Acids Res 33: 6172-6187. doi: 10.1093/nar/gki915

Davidson EA, Meyer AJ, Ellefson JW, Levy M, Ellington AD 2012. An in vitro autogene. ACS Synth Biol 1: 190-196. doi: 10.1021/sb3000113

Deatherage DE, Barrick JE 2014. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Methods Mol Biol 1151: 165-188. doi: 10.1007/978-1-4939-0554-6_12

Dickinson BC, Leconte AM, Allen B, Esvelt KM, Liu DR 2013. Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution. Proc Natl Acad Sci U S A 110: 9007-9012. doi: 10.1073/pnas.1220670110

Ellefson JW, Meyer AJ, Hughes RA, Cannon JR, Brodbelt JS, Ellington AD 2014. Directed evolution of genetic parts and circuits by compartmentalized partnered replication. Nat Biotechnol 32: 97-101. doi: 10.1038/nbt.2714

Galdzicki M, Clancy KP, Oberortner E, Pocock M, Quinn JY, Rodriguez CA, Roehner N, Wilson ML 2014. The Synthetic Biology Open Language (SBOL) provides a community standard for communicating designs in synthetic biology. 32: 545-550. doi: 10.1038/nbt.2891

Gardner TS, Hawkins K 2013. Synthetic biology: evolution or revolution? A co-founder's perspective. Curr Opin Chem Biol 17: 871-877. doi: 10.1016/j.cbpa.2013.09.013

Hammerling MJ, Ellefson JW, Boutz DR, Marcotte EM, Ellington AD, Barrick JE 2014. Bacteriophages use an expanded genetic code on evolutionary paths to higher fitness. Nat Chem Biol 10: 178-180. doi: 10.1038/nchembio.1450

Ikeda RA, Chang LL, Warshamana GS 1993. Selection and characterization of a mutant T7 RNA polymerase that recognizes an expanded range of T7 promoter-like sequences. Biochemistry 32: 9115-9124.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647-1649. doi: 10.1093/bioinformatics/bts199

Kosuri S, Kelly JR, Endy D 2007. TABASCO: A single molecule, base-pair resolved gene expression simulator. BMC Bioinformatics 8: 480. doi: 10.1186/1471-2105-8-480

Meyer AJ, Ellefson JW, Ellington AD 2015. Directed Evolution of a Panel of Orthogonal T7 RNA Polymerase Variants for in Vivo or in Vitro Synthetic Circuitry. ACS Synth Biol 4: 1070-1076. doi: 10.1021/sb500299c

Molineux I. 2006. The T7 Group. In: Abedon ST, editor. The Bacteriophages. New York, New York: Oxford University Press. p. 277-301.

Segall-Shapiro TH, Meyer AJ, Ellington AD, Sontag ED, Voigt CA 2014. A 'resource allocator' for transcription based on a highly fragmented T7 RNA polymerase. Mol Syst Biol 10: 742. doi: 10.15252/msb.20145299

Smeal SW, Schmitt MA, Pereira RR, Prasad A, Fisk JD 2017a. Simulation of the M13 life cycle I: Assembly of a genetically-structured deterministic chemical kinetic simulation. Virology 500: 259-274. doi: 10.1016/j.virol.2016.08.017

Smeal SW, Schmitt MA, Pereira RR, Prasad A, Fisk JD 2017b. Simulation of the M13 life cycle II: Investigation of the control mechanisms of M13 infection and establishment of the carrier state. Virology 500: 275-284. doi: 10.1016/j.virol.2016.08.015

Tahirov TH, Temiakov D, Anikin M, Patlan V, McAllister WT, Vassylyev DG, Yokoyama S 2002. Structure of a T7 RNA polymerase elongation complex at 2.9 A resolution. Nature 420: 43-50. doi: 10.1038/nature01129

Citations in Chapter 3

Ando H, Lemire S, Pires DP, Lu TK 2015. Engineering Modular Viral Scaffolds for Targeted Bacterial Population Editing. Cell Syst 1: 187-196. doi: 10.1016/j.cels.2015.08.013

Binford SL, Weady PT, Maldonado F, Brothers MA, Matthews DA, Patick AK 2007. In vitro resistance study of rupintrivir, a novel inhibitor of human rhinovirus 3C protease. Antimicrob Agents Chemother 51: 4366-4373. doi: 10.1128/aac.00905-07

Bull JJ, Gill JJ 2014. The habits of highly effective phages: population dynamics as a framework for identifying therapeutic phages. Front Microbiol 5: 618. doi: 10.3389/fmicb.2014.00618

Bull JJ, Millstein J, Orcutt J, Wichman HA 2006. Evolutionary feedback mediated through population density, illustrated with viruses in chemostats. Am Nat 167: E39-51. doi: 10.1086/499374

Cordingley MG, Callahan PL, Sardana VV, Garsky VM, Colonno RJ 1990. Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro. J Biol Chem 265: 9062-9065.

Cordingley MG, Register RB, Callahan PL, Garsky VM, Colonno RJ 1989. Cleavage of small peptides in vitro by human rhinovirus 14 3C protease expressed in Escherichia coli. J Virol 63: 5037-5045.

De Paepe M, Taddei F 2006. Viruses' life history: towards a mechanistic basis of a trade-off between survival and reproduction among phages. PLoS Biol 4: e193. doi: 10.1371/journal.pbio.0040193

Dickinson BC, Packer MS, Badran AH, Liu DR 2014. A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations. Nat Commun 5: 5352. doi: 10.1038/ncomms6352

Dougherty WG, Cary SM, Parks TD 1989. Molecular genetic analysis of a plant virus polyprotein cleavage site: a model. Virology 171: 356-364.

Esvelt KM, Carlson JC, Liu DR 2011. A system for the continuous directed evolution of biomolecules. Nature 472: 499-503. doi: 10.1038/nature09929

Flores H, Ellington AD 2002. Increasing the thermal stability of an oligomeric protein, beta-glucuronidase. J Mol Biol 315: 325-337. doi: 10.1006/jmbi.2001.5223

Guo F, Liu Z, Fang PA, Zhang Q, Wright ET, Wu W, Zhang C, Vago F, Ren Y, Jakana J, Chiu W, Serwer P, Jiang W 2014. Capsid expansion mechanism of bacteriophage T7 revealed by multistate atomic models derived from cryo-EM reconstructions. Proc Natl Acad Sci U S A 111: E4606-4614. doi: 10.1073/pnas.1407020111

Kapust RB, Tozser J, Fox JD, Anderson DE, Cherry S, Copeland TD, Waugh DS 2001. Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. Protein Eng 14: 993-1000.

Kiro R, Shitrit D, Qimron U 2014. Efficient engineering of a bacteriophage genome using the type I-E CRISPR-Cas system. RNA Biol 11: 42-44. doi: 10.4161/rna.27766

Kostallas G, Samuelson P 2010. Novel fluorescence-assisted whole-cell assay for engineering and characterization of proteases and their substrates. Appl Environ Microbiol 76: 7500-7508. doi: 10.1128/aem.01558-10

LaVallie ER, DiBlasio EA, Kovacic S, Grant KL, Schendel PF, McCoy JM 1993. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the E. coli cytoplasm. Biotechnology (N Y) 11: 187-193.

Li Q, Yi L, Marek P, Iverson BL 2013. Commercial proteases: present and future. FEBS Lett 587: 1155-1163. doi: 10.1016/j.febslet.2012.12.019

Matthews DA, Smith WW, Ferre RA, Condon B, Budahazi G, Sisson W, Villafranca JE, Janson CA, McElroy HE, Gribskov CL, et al. 1994. Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. Cell 77: 761-771.

Molineux I. 2006. The T7 Group. In: Abedon ST, editor. The Bacteriophages. New York, New York: Oxford University Press. p. 277-301.

Novagen. 2014. T7Select System Manual TB178 Rev.B 0203. In: Sigma M, editor.

O'Donoghue AJ, Eroy-Reveles AA, Knudsen GM, Ingram J, Zhou M, Statnekov JB, Greninger AL, Hostetter DR, Qu G, Maltby DA, Anderson MO, Derisi JL, McKerrow JH, Burlingame AL, Craik CS 2012. Global identification of peptidase specificity by multiplex substrate profiling. Nat Methods 9: 1095-1100. doi: 10.1038/nmeth.2182

Parks TD, Howard ED, Wolpert TJ, Arp DJ, Dougherty WG 1995. Expression and purification of a recombinant tobacco etch virus NIa proteinase: biochemical analyses of the full-length and a naturally occurring truncated proteinase form. Virology 210: 194-201. doi: 10.1006/viro.1995.1331

Phan J, Zdanov A, Evdokimov AG, Tropea JE, Peters HK, 3rd, Kapust RB, Li M, Wlodawer A, Waugh DS 2002. Structural basis for the substrate specificity of tobacco etch virus protease. J Biol Chem 277: 50564-50572. doi: 10.1074/jbc.M207224200

Plessis A, Perrin A, Haber JE, Dujon B 1992. Site-specific recombination determined by I-SceI, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus. Genetics 130: 451-460.

Pogson M, Georgiou G, Iverson BL 2009. Engineering next generation proteases. Curr Opin Biotechnol 20: 390-397. doi: 10.1016/j.copbio.2009.07.003

Renicke C, Spadaccini R, Taxis C 2013. A tobacco etch virus protease with increased substrate tolerance at the P1' position. PLoS One 8: e67915. doi: 10.1371/journal.pone.0067915

Sandersjoo L, Jonsson A, Lofblom J 2017. Protease substrate profiling using bacterial display of self-blocking affinity proteins and flow-cytometric sorting. Biotechnol J 12. doi: 10.1002/biot.201600365

Sandersjoo L, Kostallas G, Lofblom J, Samuelson P 2014. A protease substrate profiling method that links site-specific proteolysis with antibiotic resistance. Biotechnol J 9: 155-162. doi: 10.1002/biot.201300234

Segall-Shapiro TH, Meyer AJ, Ellington AD, Sontag ED, Voigt CA 2014. A 'resource allocator' for transcription based on a highly fragmented T7 RNA polymerase. Mol Syst Biol 10: 742. doi: 10.15252/msb.20145299

Shih YP, Wu HC, Hu SM, Wang TF, Wang AH 2005. Self-cleavage of fusion protein in vivo using TEV protease to yield native protein. Protein Sci 14: 936-941. doi: 10.1110/ps.041129605

Turk B 2006. Targeting proteases: successes, failures and future prospects. Nat Rev Drug Discov 5: 785-799. doi: 10.1038/nrd2092

van den Berg S, Lofdahl PA, Hard T, Berglund H 2006. Improved solubility of TEV protease by directed evolution. J Biotechnol 121: 291-298. doi: 10.1016/j.jbiotec.2005.08.006

Varadarajan N, Georgiou G, Iverson BL 2008. An engineered protease that cleaves specifically after sulfated tyrosine. Angew Chem Int Ed Engl 47: 7861-7863. doi: 10.1002/anie.200800736

Wang QM, Johnson RB 2001. Activation of human rhinovirus-14 3C protease. Virology 280: 80-86. doi: 10.1006/viro.2000.0760

Wei L, Cai X, Qi Z, Rong L, Cheng B, Fan J 2012. In vivo and in vitro characterization of TEV protease mutants. Protein Expr Purif 83: 157-163. doi: 10.1016/j.pep.2012.03.011

Yi L, Gebhard MC, Li Q, Taft JM, Georgiou G, Iverson BL 2013. Engineering of TEV protease variants by yeast ER sequestration screening (YESS) of combinatorial libraries. Proc Natl Acad Sci U S A 110: 7229-7234. doi: 10.1073/pnas.1215994110

Citations in Appendix

Bars I, Krauss L, Nekoogar F, Terning J. 2009. Extra Dimensions in Space and Time: Springer New York.

Calendar R. 2006. The Bacteriophages: Oxford University Press, USA.

De Paepe M, Taddei F 2006. Viruses' life history: towards a mechanistic basis of a trade-off between survival and reproduction among phages. PLoS Biol 4: e193. doi: 10.1371/journal.pbio.0040193

Kunkel TA 2004. DNA replication fidelity. J Biol Chem 279: 16895-16898. doi: 10.1074/jbc.R400006200

Lee JB, Hite RK, Hamdan SM, Xie XS, Richardson CC, van Oijen AM 2006. DNA primase acts as a molecular brake in DNA replication. Nature 439: 621-624. doi: 10.1038/nature04317

Smith MACWHF 2010. The Volume of Earth's Ocean Oceanography 23. doi: http://dx.doi.org/10.5670/oceanog.2010.51

Watson JD, Crick FH 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171: 737-738.

Vita

Daniel Joseph Garry was born in Cincinnati, Ohio. After attending Nativity School in Pleasant Ridge, he attended St. Xavier High School, graduating in 2008. While at St. X, he realized he wanted to be a scientist and did work for three summers at Cincinnati Children's Hospital. He attended Arizona State University and the Barrett Honors College, where he helped found the inaugural ASU iGEM team, graduating with a Molecular Biology and Biosciences B.S. and the Barrett Honors College with his undergraduate thesis "An Educational Experience in Synthetic Biology" in 2012. In June of 2012 he entered the Cell and Molecular Biology Ph.D. program at the University of Texas at Austin where he worked in Andy Ellington's lab.

Permanent address: djxgarry@gmail.com

This dissertation was typed by the author, Daniel Joseph Garry.