Biophysical Characterization of a ssDNA Virus

Anneliese J. Morrison
THE FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS & SCIENCES

BIOPHYSICAL CHARACTERIZATION OF A SSDNA BACTERIOPHAGE

By

ANNELIESE JAEHNE MORRISON

A Thesis submitted to the
Department of Chemistry & Biochemistry
in partial fulfillment of the requirements for graduation with
Honors in the Major

Degree Awarded:
Fall, 2014
The members of the Defense Committee approve the thesis of Anneliese Jaehne Morrison defended on December 4, 2014.

Dr. Darin R. Rokyta
Thesis Director

Dr. Michael Blaber
Outside Committee Member

Dr. Brian G. Miller
Committee Member
Acknowledgements

I would like to express my sincere gratitude for the guidance, patience, and opportunities given to me by Dr. Darin Rokyta over the past four and a half years. I extend my gratitude to all members of the Rokyta lab for supporting me and my research. Special thanks to Dr. Lindsey Mcgee for her advice and helping me improve my scientific writing skills.

I also would like to thank Dr. Brian Miller for his support, advice, and help throughout this process and allowing me to participate as a member of his lab. All members of the Miller group, especially Dr. Carl Whittington, have been extremely helpful as well, and I appreciate the time they have taken to answer questions and work on protocol development with me.

I would also like to thank Dr. Michael Blaber for showing his support for my research by participating in my committee.

I sincerely thank Dr. Liam Longo for helping me with all aspects of this project, helping me to become a better scientist, and always being welcoming to scientific discussions.

I extend gratitude to Dr. Claudius Mundoma in the Physical Biochemistry Facility (PBF), Dr. Margaret Seavy in the Biology Core Facility, and Dr. Ewa Bienkiewicz in the Protein Biology Laboratory (PBL) for training me on instruments and offering their advice. Thank you to Dr. Steven Miller in the sequencing facility and everyone in the Molecular cloning facility for their services.
1. Introduction

Mutations are the raw material of evolutionary change within a population. Genetic change is the driving force of adaptation, where beneficial mutations can confer higher fitness in organisms within a novel environment. Protein evolution is contingent upon fixation of beneficial mutations, which allows response to environmental stresses and improvement upon functional attributes. The genetic and molecular determinants of what is and is not permissible in an evolutionary landscape have been shown to be impacted greatly by a protein’s thermodynamic and kinetic stabilities (4, 20, 21, 22, 47, 65, 72). Thermodynamic stability refers to the amount of unfolded or partially unfolded states that are in equilibrium with the native functional protein. Kinetic stability describes the free-energy barrier that separates biologically functional and non-functional entities. While the relation between protein evolution and thermodynamic stability has been detailed extensively, yielding insight into the interplay between stability and evolvability (3), the same relation regarding kinetic stability is lacking. This is presumably due to the relative ease of measurement of thermodynamic quantities through techniques such denaturation experiments and computer modeling that estimates mutational effects on a system’s energetics (57). Understanding kinetic stability in proteins is difficult and is often not possible for complicated protein systems. Developing a more complete understanding of the interaction between protein evolution and both thermodynamic and kinetic stabilities is critical to biotechnological applications, such as increasing shelf-life in pharmaceuticals, understanding diseases caused by misfolding, protein folding, and common trends in protein evolution (57).

1.1 Outline of this thesis

This thesis will detail an investigation of the biophysical determinants of biological function and evolution in a model system: Microvirid bacteriophages. The rest of chapter 1 will focus on defining thermodynamic and kinetic stabilities and summarizing the literature corresponding to protein evolution and how thermodynamic and kinetic stabilities can impact adaptive properties. Small model proteins are often used as experimental systems in studies that investigate these properties, as they typically show reversibility and conform to two-state models. However, these small model proteins are not representative of the vast majority of proteins. Most proteins often exhibit irreversibility and are much more complex. Thus they do not conform to two-state models and cannot be used in equilibrium thermodynamics analyses. They also exist within in vivo environments that are much more complex than their in vitro counterparts. In vivo conditions have been shown to favor formation of irreversible protein states and favor aggregation (57). Viruses are representative of large, self-assembling protein complexes. The biophysical characterization of the viral assembly and dissociation processes will afford a deeper understanding of the physical interactions that occur in these types of proteins as well as the relationship between tightly linked phenotypes and adaptation to extreme environments. A detailed outline of this thesis is as follows:

Chapter 2: Thermodynamic and kinetic stability have roles in the assembly and disassembly of viruses, as well as their evolution. Previously, ssDNA bacteriophages that have undergone a single step in an adaptive walk under selection to withstand low-pH and high heat have been observed by Rokyta and coworkers (38). As is seen in many viral systems, the assembly/disassembly processes are poorly understood from a biophysical standpoint. Similarly, the mechanisms of adaptation to biophysical and biochemical selective pressures, such as high heat and low-pH, are poorly understood in complex multi-protein systems. In order to characterize the biophysical properties of previously observed heat- and pH-adapted bacteriophages, I have optimized protein purification and
fluorescence spectroscopy protocols and characterized the complex temperature-dependent fluorescence curves that have been observed. This protocol development is essential for a systematic analysis and comparison of adapted bacteriophages as well as to understanding biophysical mechanisms that underlie viral adaptation to extreme environments.

Chapter 3: This chapter will contain conclusions from the data collected in this thesis, a discussion of the implications of that data, and the future directions of this project.

1.1 Thermodynamic and kinetic stability in proteins

Thermodynamic stability is generally studied using equilibrium thermodynamics, where simple systems such as small model proteins are used that typically conform to a two-state equilibrium model described by the following:

\[ N \leftrightarrow U \]  

where \( N \) describes the native state of a protein and \( U \) describes the unfolded state, which is an ensemble of fully and partially unfolded states (57). The unfolding equilibrium constant, \( K \), is related to the equation above,

\[ K = \frac{[U]}{[N]} \]

and describes the relative concentration of unfolded to native state protein when the system is at equilibrium. The standard unfolding free energy change associated with a two-state system at equilibrium is given by the Lewis equation,

\[ \Delta G = -RT \ln(K) \]

where \( R \) is the gas constant and \( T \) is the temperature in Kelvins. Experimental data can be fitted to the two-state equilibrium model, giving a value for \( \Delta G \) as a function of an experimental parameter such as temperature or chemical denaturant concentration. A system where \( K<1 \) and a positive value of the unfolding free energy change is indicative of an unfolding equilibrium that is shifted towards the folded state (57). A shift in the unfolding equilibrium towards the folded state indicates that the native state of the protein is favored and that the protein will be in its native, biologically functional state under physiological conditions.

While the protein folding reaction must be thermodynamically favorable to occur, or must decrease in free energy such that the more stable species is formed spontaneously, thermodynamic stability does not guarantee that a protein will remain in the folded, biologically active state for any relevant length of time (57). Kinetic stability describes the free-energy barrier that separates the native and unfolded, non-functional states. A kinetically stable protein can maintain biological function in its native state during a biologically relevant time-scale regardless of its thermodynamic stability in regards to its non-functional state. The Eyring equation describes the relationship between the rate of irreversible denaturation and the free energy barrier (\( \Delta G^\ddagger \)) and is as follows:

\[ k = k_0 e^{-\Delta G^\ddagger/RT} \]

where \( k_0 \) is the front factor and \( k \) is the rate constant for irreversible protein denaturation. A high value for the free-energy of activation will produce a sufficiently slow rate of irreversible denaturation. This scenario produces a kinetically stable protein that will remain in the functional, native state over a sufficiently long time-scale.
1.2 Protein adaptation in light of thermodynamic and kinetic stability

The biophysical basis of protein adaptation is a topic of interest among evolutionary biologists, biophysicists, and bioengineers alike. Molecular evolution and protein biophysics are investigating increasingly complementary questions such as how the physical and chemical effects amino acid substitutions are manifested by specific selective pressures. A discussion of the current understanding of protein stability and its relation to the evolutionary process is therefore necessary to gaining insight into the adaptive evolution of natural proteins.

1.2.1 General trends in protein adaptation

With regards to general trends of protein evolution, two points are of primary importance. These two points describe the role of neutral genetic drift in movement through fitness landscapes and the influence that epistasis exerts on the topology of fitness landscapes. Proteins evolve by a combination of neutral genetic drift and selection on functionally significant amino acid substitutions (3). Mutations that are selectively neutral can become prominent in a population. Selectively neutral mutations are single nucleotide changes that accumulate at an extremely slow rate and are neutral with respect to fitness. Positive selection occurs when selection favors a particular substitution that is functionally beneficial, and is pervasive during the process of adaptation. Functionally neutral amino acid substitutions have been shown to play an important role in adaptation by reshaping the fitness landscape (3, 48). Selectively neutral mutations have no effect on the properties being specifically selected for but can allow another mutation to fix that is beneficial through epistasis. As a couple the two mutations are able to be beneficial, although when alone they are neutral or deleterious. They can effectively provide alternate pathways between fitness optima and enable a more extensive sampling of sequence space by way of producing variants that are especially robust to genetic and environmental changes (48). Neutral mutations also provide a platform for catalytic promiscuity and substrate ambiguity in enzymes, therefore contributing constructively to the evolution of new functions, such as in enzyme recruitment, as well as adaptation to new conditions (59).

Secondly, the influences of epistasis and pleiotropy are prevalent in evolutionary analyses. The case discussed above, where a neutral mutation provides the background for a beneficial but possibly deleterious mutation to fix, is a specific description of epistasis. The ability of an adaptive mutation’s effect to depend on the presence or absence of other non-adaptive mutations is referred to as epistasis. In other words, the presence of two alleles together produces a different phenotype than the expected combined effect. Pleiotropy can be defined in multiple subtly distinct ways and is defined here as when a single allelic change affects more than one phenotype. Pleiotropy will be discussed in the context of genes with high fitness density and few unconstrained sites, such as in viral systems where genomes are very small and genes therefore code for multiple products.

Empirical and computational analyses have shown the pervasiveness of epistasis as a major evolutionary force in adaptation. Epistasis has been shown to constrain the topology of fitness landscapes and has been evidenced to manifest its effects in multiple fashions. Compensatory mutations have been shown to be common in cases where observed mutations restore protein activity or fitness (14). The frequency of the occurrence of such mutations is high, with 10 to 12 compensatory mutations existing for each deleterious mutation observed (14). Antagonistic epistasis has also been observed, where the total fitness effect of multiple allelic changes is less than expected if the effects were assumed to be additive (54). Synergistic epistasis occurs when combined allelic
differences produce effects that are greater than predicted if the effects were assumed to be additive (45). Similar variations of pleiotropy exist and can be important factors in the distribution of beneficial fitness effects (43).

1.2.2 Thermodynamic discussion of protein evolution

The process of folding into a functional, three-dimensional protein is a topic that is not well understood, although the forces that enable proteins to retain structural integrity, and therefore thermodynamically stabilize them, have been heavily investigated. High stability results in structurally rigid, denaturation resistant molecules that are in surplus relative to unfolded states. The forces that stabilize and drive the folding of proteins vary from short-range interactions, such as electrostatic interactions between polar and ionized groups, van der Waals interactions, hydrogen bonding, or the formation of covalent bonds between amino acids such as cysteine, and the global disordering of solvent molecules by packing of hydrophobic residues in the interior cavity. Short-range interactions described above are enthalpic in nature, while the hydrophobic effects are entropic in nature. These relate to the free energy of a system by way of the following equation,

\[ \Delta G = \Delta H - T \Delta S \]  

where \( H \) denotes enthalpy and \( S \) denotes entropy. Protein folding has been shown to be primarily entropy driven, as short-range intermolecular interactions contribute an overall minute quantity to the free energy of folding of the native structure. When selection to extreme conditions are applied, both enthalpic and entropic contributions are affected. Increases in enthalpic interactions necessarily decrease structural flexibility and concomitantly decrease conformational entropy. This proposes a necessary balance between flexibility and rigidity/specificity (27). The fact that evolution has resulted in largely marginally stable proteins, with \( \Delta G \) values between -3 and -10 kcal/mol, is unsurprising (14).

Thermodynamic stability in the context of protein adaptation has been highlighted extensively by directed evolution experiments, accounting for the observation of widespread marginal stability as well as stability-mediated epistasis and stability-function trade-off (3). A few points are necessary to discuss in order to develop an accurate relationship between thermodynamic stability and adaptation.

The observed lower limit to \( \Delta G \), -3 kcal/mol, can be understood in terms of equation 3. A protein that is below the \( \Delta G \) limit will have a decreased concentration of biologically functional molecules at equilibrium (14). Large populations of unfolded or partially unfolded states can favor degradation and aggregation. The observed upper limit to \( \Delta G \), -10 kcal/mol, can be understood structurally, which leads to the concept of stability function trade-off. Increasing the stability of a molecule can reduce dynamic flexibility and concomitantly decrease conformational entropy, which often requires flexibility to bind molecules necessary for catalysis or proper folding (14). The observation that marginal stability was ubiquitous among proteins lead to the conclusion that marginal stability arose as an intrinsic trade-off between thermodynamic stability and function (26).

Contrary to stability-function trade-off, many directed evolution experiments have observed that stability increases do not always result in a cost for protein function (3). Marginal stability can therefore arise neutrally through mutation-selection balance (26). If increases in thermodynamic stability are neutral with respect to fitness, or function, selection will not differentiate between marginal or hyperstable variants (26). Most mutations in the fitness landscape are
thermodynamically destabilizing, as are functionally beneficial mutations (3). It is much more likely that low stability but functionally compatible mutations will be encountered in an adaptive walk, than mutations that are both stabilizing and functionally beneficial. Selective pressure and genetic drift drive proteins through the landscape neutrally with respect to stability and arrive at a marginally stable but functional state (26).

Many directed evolution experiments have observed this phenomenon, whereby a protein can be rescued from being trapped on a fitness peak, thereby enhancing evolvability (3). During these experiments, selection favors stability only in the respect that the protein must fold and be functional, or that the $\Delta G$ of folding is at least above the folding threshold. The protein becomes trapped when a mutation brings it to, or very close to, the folding threshold. By introducing functionally neutral but stabilizing mutations the protein can continue fixing functionally beneficial mutations (3). This phenomenon is known as stability-mediated epistasis, where two alleles interact to produce a mutationally robust variant, rather than a one allele variant trapped on a fitness peak. Stability therefore can enhance evolvability in proteins, rather than come at a cost.

1.2.3 Kinetic discussion of protein evolution

Biological function can also be guaranteed over a relevant timescale if the rate of irreversible denaturation is sufficiently slow, as dictated by equation 4. Kinetic stability can manifest itself as a trait completely independent from thermodynamic stability, as is the case with bacterial $\alpha$-lytic protease, or can show significant correlation with it (57, 63). The generalized Lumry-Eyring model can be used to understand the relationship between thermodynamic and kinetic stabilities:

$$N \leftrightarrow U \rightarrow F$$  \hspace{1cm} (6)

where N, U, and F represent the native, unfolded, and final states of the protein (34, 47, 58). The native state may denature into unfolded and partially-unfolded states, which can further irreversibly denature to a final state by processes such as aggregation or interaction with macromolecules (57). A large free-energy barrier between the native and final protein states can delay the occurrence of irreversible changes to the unfolded states and guarantee that the protein is functional under denaturing conditions, such as crowded and/or harsh environments in vivo (57).

The conditions encountered by proteins both intracellularly and extracellularly are often harsh and favor formation of irreversible states. These conditions include proteolysis, aggregation, macromolecular interactions, autolysis, and covalent modifications of amino acids, which occur readily in harsh extracellular and crowded intracellular environments (47, 57). In vivo stability is likely linked to some degree with natural selection, where proteins are selected to be kinetically stable in response to harsh in vivo conditions that favor aggregation and unfolding processes (20). Kinetic stability can therefore provide a safety mechanism that prevents loss of function over a biologically relevant timescale in an aggressive environment. Kinetic stability has been shown to be important in many complex protein systems such as bacterial $\alpha$-lytic protease, Newcastle disease virus, collagens, SHP viral capsid protein, Cu/Zn superoxide dismutase, pyrrolidone carboxyl peptidase, NAPase, and human low-density lipoprotein, which all show exceedingly slow unfolding kinetics (18, 20, 28, 29, 33, 35, 37, 46, 52, 61, 63).

Evidence for kinetic stabilities role in natural selection is much more indirect than that of thermodynamic stability, presumably due to the relative ease of measuring and estimating thermodynamic quantities. Folding/unfolding kinetics studies on both natural and designed, or “de-
proteins have shown that rapid folding and marginal free energy of folding barriers occur as a consequence of natural selection (57). One can rationalize the role of kinetic stability in natural protein evolution as follows: the unfolded states of a protein are prone to aggregation and degradation and thus a high free-energy barrier separating the unfolded states from the native state indicates that the unfolded states have a higher free-energy and are not thermodynamically favored (57); as mentioned above, the barrier also acts as a safety mechanism, protecting the native state from irreversibly converting to non-functional states (57).

It has also been proposed that the existence of a lower limit to thermodynamic stability reflects natural selection for kinetic stability in E. coli thioredoxin, where a robust correlation was observed between mutational effects of the unfolding activation free-energy and the frequency of the occurrence of residues in a sequence alignment (20). They found that change of a few kJ/mol in the unfolding free-energy only moderately effects the equilibrium denaturation temperature. However, an alteration of the activation free-energy on a similar scale can produce a large effect on the unfolding half-life and time-scale for irreversible denaturation (20, 57). The existence of a widely evolutionarily conserved thermodynamic stability threshold of a few kJ/mol below wild-type therefore supports natural selections actions on kinetic stability (20, 57). It also suggests that purifying selection may operate on kinetic stability thresholds rather than thermodynamic stability, ensuring that the half-life of the functional form is maintained in vivo (57).

Experimental evidence for natural selection on kinetic stability has been indicated by a study on two structural homologues, α-lactalbumin (BLA) and hen-egg-white lysozyme (HEWL), with 38% sequence identity and the same overall folded structure (24). BLA showed a marginal folding-unfolding barrier while HEWL showed a significant barrier (24). The difference in folding-unfolding barriers are easily linked to the biological requirements of each protein and their natural selection. BLA is a membrane binding protein which must partially unfold into a molten globule conformation in order to function. Proteins that function in a molten globule-like state favor marginal folding-unfolding barriers (57). HEWL functions in harsh extracellular conditions as a bactericide. High free-energy barriers to unfolding increase kinetic stability and therefore its half-life in a biologically functional state (24, 57). This experiment shows the clear correlation between protein kinetic stability and biological function and gives strong support for a role of kinetic stability in natural selection.

Finally, kinetic stability has been suggested to aid the evolution of optimal functional properties (28). As detailed in section 1.2.2 above, the optimization of protein function via natural selection is tightly correlated with protein stability. In many systems thermodynamic stability is the physical agent found to be responsible for evolutionary trends, where the native state folds spontaneously and is thermodynamically more stable than the unfolded states. There are systems that display kinetic stability that is completely independent from thermodynamic stability. Bacterial α-lytic protease is such a protein. When denatured, ALP does not spontaneously refold to its native state because it is less thermodynamically stable than the unfolded state (28). It is a proenzyme and thus requires a pro-region to both provide a thermodynamic driving force for protein folding and to decrease the folding free-energy barrier to increase the rate of the folding process (28, 57). ALP is therefore under purely kinetic control, which allows the evolution of properties that would otherwise be incompatible with thermodynamic stability such as restricted native-state conformational dynamics and insensitivity to proteolysis (28, 57).
Chapter 2: Biophysical basis of viral adaptation

2.1 Introduction

Viral systems are extremely complex; they must assemble hundreds of proteins and encapsulate genomic information rapidly with fidelity and specificity, remain viable between hosts, and must be able to release their contents when they reach their next host. Conflicting requirements are necessarily imposed upon viruses by their infection cycle. Viral capsids must be stable enough to survive extracellularly, but must also be sufficiently unstable to release their contents and begin an infection cycle. Mutations that affect capsid stability can therefore have exacerbated effects on the fitness of viruses. Increases in capsid stability can negatively impact assembly processes by promoting kinetic traps, render the virus unable to respond to environmental cues that prompt denaturation and the beginning of an infection cycle, and can lead to the evolution of increased virulence (5, 19). In depth studies on the physical basis of the association and dissociation processes in viral capsids enable the modeling of those processes in large, multimeric protein complexes. As noted above, viral capsids are subjected to conflicting functionality and have evolved diverse structural solutions to fulfill these requirements. They therefore are an attractive model system to study tightly coupled structure-function relationships in large protein complexes and facilitate the development of antiviral strategies.

Prior work by Rokyta and coworkers has focused on studying the first adaptive walk in ssDNA Microvirid bacteriophages in response to biochemical and biophysical selective pressures (38). Because the evolutionary process in viruses necessitates a growth phase when faced with a population bottleneck, our selection protocol more accurately represents natural population dynamics imposed by the viral infection cycle where populations experience fluctuating phases of growth and exposure to extreme environments. Capsid stability is tightly linked to both the assembly and disassembly processes. Thus, adapted variants enable a study of pleiotropy, or the effect of a single allelic change on multiple traits, by characterizing growth and decay rates as well as overall fitness. The results and implications of this study will be discussed in later sections of this chapter. The characterization of adapted mutants thus far by growth and decay rates, as well as computational results, motivated my efforts to empirically analyze the molecular and physical basis of viral adaptation to extreme environments. In order to understand the nature of thermal transitions observed in fluorescence and static light scattering experiments, each was characterized by gel analysis, plaque assays, concentration dependence, and rate dependence using a single standard phage genotype used in adaptation experiments. Assembly kinetics and disassembly kinetics were also performed in order to complete the biophysical characterization. An in depth biophysical characterization of observed thermal transitions is necessary to set the stage for a rigorous systematic biophysical analysis of heat- and pH-adapted phages. Before presenting and discussing experimental results acquired during this project, the relationships and roles of thermodynamic and kinetic stabilities in the function of viruses will be established.

Section 2.1.1 Viral assembly and disassembly: thermodynamic and kinetic effects

There are two major steps in the infection cycle of viruses: assembly and disassembly. Both processes are dependent upon geometry, protein-protein contact energies, global stability, and environmental conditions. Thermodynamic stability is an inherent physical requirement of assembly, as protein-protein interactions must be made spontaneously and each addition of a subunit generating an increasingly more stable assembly intermediate. Thermodynamic stability is
therefore also expected to play a role in disassembly. A globally stable capsid provides a mechanism of survival extracellularly between host cells. Assembly is an inherently kinetic process but does not necessarily or evidently include a role for kinetic stability, in terms of a functional role that a high energetic barrier between intermediate states would serve. However, kinetic stability could play a large role in the disassembly process. Many viral systems exhibit a specific mechanism of kinetic stabilization termed hysteresis, or the failure of opposing reactions to equilibrate, in response to dissociation as well as hysteresis between the assembly and disassembly reactions (62). One can easily articulate a function that kinetic stability would serve in the dissociation of capsids: it allows viruses to remain viable under harsh conditions by increasing the barrier between native and unfolded states while not violating the fundamental requirement of metastability by the assembly reaction. The following portions of this section will focus on common viral structures and geometries and their role in the regulation of the infection cycle, a thermodynamic and kinetic analysis of viral self-assembly, and an analysis of capsid dissociation that focuses on how kinetic stability accounts for experimental observations observed by this study as well as many others.

2.1.1a Viral structure and geometry

Small viruses typically adopt one of two geometries, spherical and rod-like. Bacteriophages, such as those used in this study, often take the former geometry to form an icosahedron. Although geometrically complex, unrelated virus families have adopted an icosahedral geometry, indicating evolutionary convergence upon an optimal capsid geometry (31). This geometry affords small surface area in proportion to volume, allowing small capsid proteins to enclose necessary genomic and infectious components (13). Because geometric and physical principles dictate that one face of an icosahedral capsid must consist of at least three proteins, the simplest icosahedrons must contain 60 repeating subunits that interact via 30 interfacial contacts (31). The number of subunits in an icosahedral face may therefore be multiples of 60 and is described by a T-number, or triangulation number.

The existence of spherical capsid structures that have a T-number higher than one, or have a multiple of 60 subunits that is greater than one, are explained by the theory of quasiequivalence (9). Quasiequivalence also rationalizes the widespread adoption of icosahedral geometries in disparate viral families by providing a mechanism for the construction of large spherical capsid structures (31). Each identical subunit that constitutes the capsid inhabits a non-identical local environment when the lattice expands beyond $T=1$. Due to the slight differences between environments, structural variation arise that allow the otherwise flat hexagonal protein complex to distort and develop a spherical capsid structure while enclosing a large volume (9). Quasiequivalence essentially encodes the ability of structurally rigid capsids to also display flexibility to arrange identical protein subunits into similar, but not identical, functional geometries (9, 31).

Just as with capsid geometry, viral capsid proteins also show preference for certain structural motifs. The capsid proteins of small, non-enveloped icosahedral viruses often adopt an eight-stranded antiparallel $\beta$-barrel fold, which is referred to as a “jelly roll” fold (31). This structural motif is found in diverse viral families such as RNA picornaviruses, DNA paroviruses, RNA plant-viruses, DNA polyomaviruses, DNA papillomaviruses, and DNA bacteriophages such as the Microviridae family that was used in this study (31, 55). Other common motifs include five-strand $\beta$-sheets flanked by two C-terminal $\alpha$-helices in RNA bacteriophages and $\alpha$-helical folds in herpesviruses and bacteriophages (30, 31, 51, 70, 71).
Capsid protein subunits must assemble rapidly with high fidelity and specificity in lieu of the flexibility of intersubunit interactions afforded by quasiequivalence. However, the correct geometry and structural stability are not the only factors necessary to produce an infectious virion. Many viruses also require assembly regulation mechanisms to activate the assembly reaction, ensuring that assembly timing and packaging do not occur aberrantly, as well as disassembly regulation mechanisms (31, 64). These regulators function to enhance the distinction between the assembly and disassembly reactions of viral particles. Hepatitis B virus employs an induced fit allosteric regulation mechanism that requires a conformational switch between active and inactive capsid protein conformations (44). Bacteriophages P22 and ΦX174 use scaffolding proteins that mediate coat protein interactions into the correct geometry. A conformational change causes release of scaffolding proteins to form the final viral particle (17). Molecular dynamic simulations of assembly kinetics have shown that the imposition of a regulatory step to assembly nucleation results in a more robust reaction under broad ranges of conditions (73, 75). Similarly, disassembly regulators are being investigated that trigger release of genomic information within the host. Recently, interdimer disulfide bond formation in Hepatitis B virus was shown to trigger uncoating of capsid particles at the nuclear membrane (60).

Section 2.1.1b Assembly kinetics and thermodynamics

Capsid protein interaction energies are governed by the same interactions as other oligomeric protein structures (i.e. Van der Waals, hydrogen bonding, and disulfide bonding). Electrostatic interactions and bonding between residues within a protein have been shown to contribute a relatively small quantity to the overall free energy of folding/association. Similarly, capsid assembly tends not to be driven by strong covalent or hydrogen bonding interactions but by the interaction between and subsequent burial of interfacial hydrophobic residues (2). Folding, and analogously capsid assembly, has been found to be entropy driven. Ordering of residues within the complex allows for a global disordering of the surrounding aqueous environment and consequently an increase in entropy and decrease in overall free energy of the process (1, 10, 31, 49, 50, 66). Experimental evidence for association by burial of hydrophobic residues has been shown in Hepatitis B virus, where assembly was found to be a temperature dependent reaction with positive enthalpy and entropy, and a negative overall free energy for the reaction (10).

A reversible capsid assembly reaction can be modeled as follows:

\[
\text{Subunits} \leftrightarrow \text{nuclei} + \text{subunits} \leftrightarrow \text{capsids} \tag{7}
\]

and is similar to classical polymerization (31). In classical polymer assembly a single nucleus grows by the addition of an infinite amount of equivalent subunits. The rate-limiting step of the polymer assembly reaction produces a lag phase, and is dependent upon the formation of the nucleus. There is a period of rapid addition of polymer subunits which ends as the reaction reaches equilibrium. In this state the ends of the polymer, which are energetically equal, are in constant flux but achieve no net growth (31). The capsid assembly reaction shares the same major steps as classical polymer assembly but must be treated differently. In capsid assembly, the lag phase observed is caused by the formation of many nuclei that begin the elongation process asynchronously and ends only when progressively more complete intermediates begin to fully form capsids and accumulate. Rapid growth during capsid assembly is not simply due to the addition of subunits to an intermediate state, but the continuation of the nucleation-elongation process and formation of complete capsid
structures. As the population of intermediates are consumed by the production of capsids, the reaction slows and reaches a steady-state.

The capsid assembly reaction therefore requires concurrent nucleation, elongation, and completion which results in a few major differences between classical polymerization and capsid assembly. Following nucleation polymers can elongate infinitely while capsids are discrete endpoints of a nucleation event (76). There is no flux between complete capsids and intermediates because they are closed structures (76). Polymerization also requires that the ends are equivalent, however for capsids this does not apply (76). Completion of the capsid progressively changes the number of ligand per subunit and results in the cooperative nature observed where addition of a subunit is more favorable with every contact that is bound (31, 76).

The closed nature of the capsid structure has certain implications on assembly reaction (73). At the steady-state equilibrium point of assembly, both subunits and complete capsids are present, while intermediates are rare. Both the rates and extent of product formation in assembly are extremely concentration dependent (73). Lastly, intersubunit contact energy and reaction association constants have been found to be surprisingly weak (10, 73, 74). Experimentally measured subunit contact energies for Hepatitis B virus are on the order of -2.9 to -4.4 kcal/mol depending on environmental conditions and association constants are on the order of 1-5 mM (10, 74). These locally weak interactions result in a globally stable capsid, allow flexibility in regulation mechanisms, and promote capsid breathing or other steps that are essential for capsid dissociation, and minimize kinetic trapping (74). Kinetic trapping is a barrier to the assembly process that occurs when subunits bind intermediate structures quickly and deplete the necessary free subunits required to complete the capsids structure (76). It can be stimulated by high intersubunit association energies or environmental conditions such as NaCl in Hepatitis B virus, which is suggested to rapidly induce a conformational change in all free subunits to an assembly active state (10).

Section 2.1.1c Disassembly kinetics and thermodynamics

As mentioned above, intersubunit contact energies are surprisingly weak. Although they result in a globally stable structure, capsids have been observed to be more stable to dissociation than indicated by estimations from thermodynamic measurements alone. However, global thermodynamic stability exhibited by capsid structures only guarantees that the ratio of intermediate and native states at equilibrium energetically favors the native state. In environments where intermediates are not present, such in the extracellular space between hosts or entry into a new host, capsids remain intact under conditions in which they would not assemble (18, 73). Thermodynamic stability alone cannot explain the resistance of viral capsids to harsh environments and to infinite dilution (18).

The contrasting observation of metastable capsid structures that display extraordinary stability in response to severe environments reflects the distinction between the assembly and disassembly reactions, which do not equilibrate and thus exhibit hysteresis (18, 73). Hysteresis is indicative of kinetic stability and does not violate any thermodynamic principles of capsid assembly as it only dictates the rate of interconversion between native and intermediate states (18). The existence of an energetic barrier for the dissociation of viral capsids to unfolded states makes the
unfolding reaction very slow, stabilizing the capsids native state regardless of thermodynamic stability (18, 58). The closed geometry and oligomeric nature of capsids yields a molecular rationale for kinetic stability in lieu of extreme thermodynamic stability—subunits make extensive contacts with multiple neighboring subunits and these interactions can inhibit the concerted motions necessary to begin the denaturation process (18, 62). The widespread adoption of kinetic stability and hysteresis suggests that they play important biological roles in the infection cycles of many viral families and may be indispensable for viral function.

Section 2.1.2 Adaptation of Microvirid bacteriophages to biophysical and biochemical pressures

Microvirid viruses are non-enveloped, tail-less, T=1 icosahedral bacteriophages with small circular ssDNA genomes that are around 3-7 kbp in length (56). There are two subgroups in the Microvirid family with distinctions based upon major coat protein structure, the occurrence of two genes, and host range (6). The microviruses, typified by phages such as ΦX174, G4, and α3, infect Enterobacteria such as E. coli. The gokushoviruses infect intracellular parasitic bacteria such as Chlamydia and contain a more complex coat protein while lacking major spike and scaffolding proteins (6).

Research done in the Rokyta lab has focused on using G4-like microviruses for experimental evolution studies. Genome sizes range from between 4.5 to 6 kbp that encode 11 genes, four of which encode structural proteins F, G, H, and J. These phages are a desirable system for investigating structure/function relationships as the capsid structure has been determined and the conflicting nature of the viral infection cycle has led to the evolution of finely tuned functional relationships in viral components (8). They are also useful in the investigation of the properties of molecular evolution. They are easily cultured, have small genomes amenable to sequence analysis, and have resolved structures that mutations can be mapped to (15, 39, 40, 41).

The evolutionarily refined and contrasting nature of the relationship between viral capsid stability and growth has been a central focus of the Rokyta lab. An increase in capsid stability is expected to deleteriously impact growth rate by way of interfering with the assembly reaction. Physical restrictions introduced by the pleiotropic nature of viral assembly and stability have resulted in a narrow range of optimal intersubunit association energies (74). To investigate the pleiotropic interactions between capsid stability and growth rate, we have developed a selection protocol that mimics the infection cycle of viruses and can be seen in figure 1 below (38). Selection on growth within an excess of host was alternated with selection on increased stability within the phage capsid in response to two different extreme environments, extreme heat (80°C) and low-pH (1.5). Increased growth rate is defined by the term γ and is carried out over a time period τ(g), typically four generations in our experiments (38). The complicated nature of the growth phase renders many evolutionary solutions such as increasing the assembly or attachment rates. Decay rate is defined as δ for an exposure period τ(d) without the host present. Overall fitness, ω, is mathematically defined as follows (25, 38):

$$\omega = \gamma \times \tau(g) - \delta \times \tau(d)$$  \hspace{1cm} (8)

Serial transfers using this two-stage selection regime permitted the observation of phages with single point mutations in structural proteins and therefore an examination of the interaction between viral capsid stability and growth.
Figure 1 Two-stage selection protocol Two complete cycles of the two-stage selection protocol are shown (38). Exponential growth occurs at a rate $\gamma$ during time $\tau(g) = 60/60$ at 37°C for approximately four generations (blue). A portion of the population was removed from host cells and subjected to an extreme selective pressure-80°C or pH 1.5- for a time $\tau(d)$ and decayed at a rate $\delta$ (red). The units of $\gamma$ and $\delta$ are doublings per hour and times are recorded as fractions of an hour. For the heat-shock selection, $\tau(d) = 5/60$; for the pH selection, $\tau(d) = 3/60$. Overall fitness may be improved by increasing $\gamma$ and/or $\delta$. Notation adopted from Handel et. al (25).

Populations were subjected to the two-stage selection protocol until fixation in five independent experiments for both low-pH and extreme heat. Four unique heat-shock mutants and three unique pH-shock mutants were observed with single allelic changes in structural proteins. We have found that our results conflict with a commonly accepted theory in the protein community, or stability-function tradeoff. Evolution acts on mutations that are most likely to increase fitness though fitness landscapes are dominated by functionally neutral and deleterious mutations. It is therefore expected that during an adaptive walk proteins will encounter tradeoffs between traits such as capsid stability and growth rate, as has been observed in many diverse systems (7, 11, 12, 32, 36, 69). Contrary to these expectations, we found that when selected to survive rapidly fluctuating two-stage selective pressures synergistic pleiotropy is observed (38). By comparing the decay rates of phages observed under this two-stage selection protocol with the decay rates of phages that have been subjected to only a selective pressure on growth rate alone, we found that single-stage growth rate mutants showed evidence of tradeoffs between growth rate and decay rate (38, 53). Thus, when selection is challenged to increase two traits simultaneously, synergistic pleiotropy is observed. Otherwise, selection may result in tradeoffs.

Although the phages used in these adaptation experiments have a T=1 capsid geometry and are geometrically "simple", the structure is exceedingly complex with spike proteins, capsid proteins, and both DNA pilot and binding proteins present in the final infectious form. Quantifying interfacial binding strength empirically is thus extremely challenging, if not impossible. In order to quantify capsid stability in terms of binding strength we used molecular-dynamics simulation methods to computationally estimate stability changes between three of the four heat-shock mutants (38). We
found that all mutants tested had significantly increased binding affinities, with changes in intrinsic binding contributions from ancestral to mutant residues ranging between -4.5 and -9.2 kcal/mol (38). These changes are significantly stabilizing when multiplied to represent the change in binding affinity within the entire capsid, which contains 60 copies of the major capsid protein (F) and the spike protein (G). Overall, the computational data suggests that there are spots in the ancestral virus capsid structure that are neutral with respect to interfacial binding energies and therefore are able to be readily selected upon to provide doubly beneficial effects under conflicting two-stage selective pressures (38).

While the computational results used in this study to estimate capsid stability increases between mutants confirmed that it is feasible to use computational methods to predict mutational effects on capsid stability, techniques to empirically determine stability differences have been insufficiently explored for this system. In this honor's thesis project I have focused my efforts on developing biophysical techniques that will allow us to quantitatively characterize the assembly and disassembly processes of Microvirid viruses that have been adapted to extreme environments. In section 2.3 I will describe the optimization of the purification and fluorescence spectroscopy protocols as well as the characterization of both the complex fluorescence melting curves observed for a model genotype, ID8*, and its assembly kinetics.
Section 2.2 Methods

**Bacteriophage purification** A culture of *E. coli* C host cells were grown to a density of 1x10^8 cells/mL (OD$_{660}$: 0.15) in 1 L Lysogeny Broth (10 g Tryptone, 10 g NaCl, 5 g yeast extract, 2 mM CaCl$_2$) within a 2 L Erlenmeyer flask at 37°C in an orbital air shaker set to 200 RPM. The culture was inoculated with about 10^5 phage (MOI of about 0.001) that was isolated from an individual plaque from a single genotype. After inoculation the temperature was decreased to 33°C to optimize phage assembly and delay cell division. The inoculated culture was grown for six hours at 33°C with shaking at 200 RPM. Cells were lysed to release phage by exposure to 5 mL CHCl$_3$ per 250 mL culture. Lysed cultures were then centrifuged at 5,000 RPM for fifteen minutes at 4°C to remove cellular debris. Salt precipitation of the supernatant was done using a final concentration of 2.6 M NaCl. After brief mixing of the NaCl and supernatant, the mixture was stored at 4°C for 60 minutes. Proteins were precipitated by centrifugation at 10,000 RPM for 10 minutes at 4°C. The supernatant was then PEG precipitated using 39 g PEG-8000 per 500 mL of supernatant. PEG was fully dissolved by gentle rocking and then stored at 4°C overnight. Precipitation of phage particles was done by centrifugation at 10,000 RPM for 20 minutes. If viral pellets were not observed, a second centrifugation step at 12,000 RPM for 10 minutes was done. The supernatant was poured off and bleached as waste and the pellets were resuspended in 10mL total of 1X PBS (137 mM NaCl, 2.7 mM KCl, 10 mM NaHPO$_4$, 2 mM KH$_2$PO$_4$, pH 7.4). 100 kDa cutoff 15 mL Centricon filters (Millipore) were used to concentrate samples to ~3-4 mL by centrifugation at 5000 RPM. The concentrated sample was incubated at 37°C with RNase and DNase (70U/mL for both) for 30 minutes followed by a centrifugation step at 7480 RPM for 3 minutes at 4°C to precipitate remaining proteins in the sample. The supernatant was subsequently treated with fungal proteinase K at a final concentration of 1.5 mg/mL for 60 minutes at 37°C to remove contaminating proteins. As a final effort to remove any remaining non-phage proteins from the sample, an equal volume of CHCl$_3$ was added to the sample, vortexed for two minutes at full speed, centrifuged at 370xg for 5 minutes, and then the cap was removed within a chemical hood to evaporate CHCl$_3$ from the aqueous sample for 30 minutes. The top aqueous phase was carefully removed and saved for analysis. Because phage samples of this genotype are known to undergo an initial die-off period within the first day or so of refrigeration, titers were determined only after three days had passed since the purification. Viral titers were determined using plaque assays in *E. coli* C.

The purified phage samples were found to be complex in nature such that protein concentration did not show a 1:1 correspondence with virus concentration due to the persistence of some cellular proteins. When protein concentration was determined spectrophotometric procedures were used using the BCA assay kit (Pierce Biotechnology). Sample purity was determined by SDS-PAGE. Sample genotype was determined by sequencing of the entire genome of each purified phage sample.

**Fluorescence Spectroscopy Measurements** Fluorescence spectra were collected using a Varian Cary Eclipse luminescence spectrophotometer. 450 µl of purified phage preps were added at a concentration of ~3x10^10 pfu/mL in 1X PBS to a stoppered micro square fluorimeter cell with a path length of 4mm (Starna Cells). The excitation wavelength used was 295 nm and the emission wavelength used was 340 nm. The sample was allowed to equilibrate to 4°C for 10 minutes prior to data collection. For intrinsic fluorescence melting curves, the temperature was increased from 4°C to 95°C at different temperature ramp rates, which are labelled on each curve. Data points were collected every 0.10°C with an averaging time of 1.5 s. Static light scattering was monitored at the same time as intrinsic fluorescence using an excitation wavelength of 295 nm and an emission
wavelength of 295 nm. Excitation and emission bandwidths were both set to 5 mm in all experiments on the Cary Eclipse luminescence spectrophotometer.

Kinetic curves monitoring intrinsic fluorescence (ex. 295 nm, em. 340 nm) over time at a constant 80°C were also done using the same phage concentration and fluorimeter cell. Data points were collected every 0.10°C with an averaging time of 1.5s. To identify the correlation between phage viability and the thermal transitions seen in each experiment, samples were removed from a 500 ul phage sample at specified time points and stored at 4°C until viral titers were determined using plaque assays in E. coli C. Plaque assays were performed the same day as the experiment. For titer and thermal transition correlation, the same method was used on the Varian Cary Eclipse luminescence spectrophotometer as was used to acquire intrinsic fluorescence melting curves. All fluorescence data was measured in triplicate or more.

Data Analysis Raw curves from intrinsic fluorescence experiments cannot be directly compared because of differential signal intensities between multiple samples. The denaturation curves were treated as a two-state mixture of intact capsid and denatured intermediates. Normalized datasets therefore express transition data in terms of mass fraction capsid, or MFC, and the temperature at which half capsids are denatured is referred to as T0.5 as has been previously described (62). The data from both static light scattering and intrinsic fluorescence curves were analyzed using the program Datafit (Oakdale Engineering, Oakdale, PA). The program precisely estimated the T0.5 as well as the slopes of the capsid and denatured intermediates by fitting the data to a two-state model.

Mass fraction capsid is defined mathematically as follows:

\[ M_{fc} = \frac{(Signal_{observed} - Signal_{Denatured})}{(Signal_{Capsid} - Signal_{Denatured})} \]  

(7)

where \( Signal_{observed} \) is the experimentally measured value for fluorescence or scattering, \( Signal_{Denatured} \) and \( Signal_{Capsid} \) are mathematically defined as follows:

\[ Signal_{Capsid} = \text{slope}_{Capsid} \times \text{Temperature}_{signal} + \text{Intercept}_{Capsid} \]  

(8)

\[ Signal_{Denatured} = \text{slope}_{Denatured} \times \text{Temperature}_{signal} + \text{Intercept}_{Denatured} \]  

(9)

where \( \text{slope}_{Capsid} \), \( \text{slope}_{Denatured} \), \( \text{Intercept}_{Capsid} \), and \( \text{Intercept}_{Denatured} \) are parameters estimated by the Datafit program by linear extrapolation of the pre- and post-transitition baselines, and \( \text{Temperature}_{signal} \) is the temperature at which the experimental signal was observed in Kelvins. Plots were made using Origin Lab 2015 (OriginLab Corp).

Assembly Kinetics Measurements Assembly kinetics experiments are based on standard bacteriophage kinetics assay protocols (67). The RY7211 strain of E. coli, which confers resistance to E protein-mediated lysis in G4-type phages, was grown to a concentration of 1x10^8 cells/mL in Lysogeny broth. Cells were immediately pelleted and washed three times with 1X HFB-1 buffer (0.06 M NH₄Cl, 0.09 M NaCl, 0.1 M KCl, 0.1 M Tris-HCl (pH 7.4), 1.0 mM MgSO₄, 1.0 mM CaCl₂) to remove carbon sources and stop bacterial growth. The pellet was resuspended in 0.9 mL HFB-2 buffer (1X HFB-1, 10 mM MgCl₂, 5 mM CaCl₂). 0.9 mL of resuspended cells were mixed with 0.1 mL phage (~1x10⁶-10⁸ phage/mL), vortexed, and incubated at 37°C for 20 minutes to allow phage attachment. The cells with attached phage were pelleted by centrifugation and supernatant (unbound phages) were saved for normalization. 1.0 mL of Pre-warmed Lysogeny broth at 37°C was added to resuspend the pellet at t=0. Sample time points consisted of 50ul aliquots added to 0.5 mL of iced CHCl₃
saturated HFB-3 (1X HFB-1, 2 mg/mL lysozyme) containing 2 mg/mL chicken egg white lysozyme. To ensure complete cell lysis, all time points were incubated with shaking at 37°C prior for at least 20 minutes. Samples were centrifuged and the supernatants were plated using standard plaque assays in *E. coli*. The assembly kinetics measurements were replicated in triplicate.

**Fitness Assays**  Fitness was measured at 80°C by calculating the population change rates on a log₂ scale, yielding population doublings during growth, or halvings during decay, per hour (38). Assays were carried out using diluted aliquots of the purified phage preparations. Fitness is treated as a linear combination of growth and decay rates. Growth, $\gamma$, occurs during exposure to the host for a time $\tau(g)$. Decay, $\delta$, occurs during exposure to 80°C for a time $\tau(d)$ in the absence of hosts. Fitness, $\omega$, is defined by $\gamma$ and $\delta$ and the times spent under the respective regimes as follows: $\omega = \gamma \tau(g) - \delta \tau(d)$ (25, 38). Fitness measurements represented in this thesis are of at least 3 replicates.
Section 2.3 Results and Discussion

Purification Optimization  The purification of proteins does not follow a ubiquitous protocol—every protein has unique purification requirements that exploit its physical and chemical characteristics. Viral capsids are no different and thus a large portion of this project was focused on developing a purification scheme that produced phage preparations with large titers, low amounts of contaminating cellular proteins, and large volumes in a reasonable amount of time and in a suitable buffer. Three protocols were quantitatively investigated in order to assess their relative success in fulfilling the requirements listed above and are as follows: 5%-30% sucrose gradients, three step 1.3 to 1.7 g/mL CsCl gradients, both 5%-30% sucrose gradients and CsCl gradients, and a PEG precipitation based protocol.

The first three purification methods investigated were based upon suggestions made by Dr. Bentley Fane through email correspondence. The first steps of the purification process after the growth phase of 1 L of phage involved using buffers and temperatures that favored host attachment and subsequent detachment without host cell lysis. These steps were employed in each of the sucrose and CsCl gradient purification schemes and required a time frame of four days, not including running the gradients, as well as an additional day to dialyze samples out of their respective buffers and into one compatible with biophysical measurement. The use of buffers and temperature to exploit natural host attachment properties of phages was employed in order to avoid cell lysis and the release of endogenous host cell proteins.

Figure 2 shows the results of running the 5%-30% sucrose gradient on a phage sample. Figure 2a shows the composition of each of the 0.5 mL fractions collected from the gradient in a single trial. The fractions corresponding to the highest phage concentrations, 12 through 16, were evaluated by determining protein concentration and purity via a 12.00% SDS-PAGE gel (Figure 2b). Fractions 12 through 16 were combined to give a total phage concentration of 2.10x10^10 pfu/mL, a total protein concentration of 2.047 mg/mL, and a final volume of 2.5 mL (Table 1). The SDS-PAGE gel shows large amounts of contaminating protein present after running the sample through the 5%-30% sucrose gradient, which also yields small sample volumes. The second protocol investigated was a three step CsCl gradient with densities 1.3, 1.5, and 1.7 g/mL as determined by the refractive index of the solution. The CsCl gradient yielded 1.5 mL of an 8.90x10^9 pfu/mL solution with a protein concentration of 2.850 mg/mL (Table 1). The CsCl and sucrose gradient methods alone thus yielded similar phage titers, levels of extraneous protein contamination, sample volumes, and time, both taking around six days total.

Although purification protocols that involve multiple steps are most often successful when the methods used are orthogonal, or exploit different physical and chemical properties, the original method suggested by Dr. Bentley Fane used both CsCl and sucrose gradients to yield pure and high titer phage preps. Both methods exploit buoyant density and are thus not orthogonal but the CsCl gradient was found to separate a bulk amount of contaminating protein by concentrating the phage to a narrow band while the sucrose gradient acted as a finer sieve to separate the phage from contamination. Figure 2c shows the purity of a sample that was subjected to both CsCl and 5%-30% sucrose gradients on a 12.00% SDS-PAGE gel. Using both CsCl and sucrose gradients yielded 2.5 mL of an 3.69x10^10 pfu/mL solution and a protein concentration of 405 μg/mL (Table 1). While there is a much lower level of contaminating protein present in the purified sample, with the protein concentration reduced by approximately 85% from the gradients run separately, the sample volume is still low and the protocol takes an extensive amount of time (seven days).
The final method investigated was a PEG precipitation based purification protocol. PEG precipitations are widely used to isolate/concentrate viruses and has been used in the Rokyta lab to isolate phages DNA from sewer samples. This method does not require manipulating the host attachment process, cutting off an initial four days of the protocol. Following the growth phase of 1 L of phages the sample is treated with CHCl\textsubscript{3} to lyse host cells and release phages and centrifuged to remove cellular proteins. A 2.6 M NaCl precipitation is done using the supernatant to aid the removal of remaining cellular proteins by centrifugation. This is followed by PEG precipitation and a proteinase K treatment step, as G4-type phages are resistant to proteinase K. This method combines multiple orthogonal purification steps and is thus superior to previous methods from a theoretical standpoint. Sample purity was evaluated by the 15.00% SDS-PAGE gel shown in figure 2d. There are barely any visible contaminating proteins, suggesting that they are below the detection limit of the Coomassie stain used (~0.5 μg) and are negligible. The yields of this protocol were also superior with a final phage concentration of 9e11 pfu/mL, protein concentration of 230 μg/mL, and a sample volume of 5 mL. There is a decrease in protein concentration as compared to the methods described above of 92% from the gradients alone and 43% from the gradients together. While there are comparable phage concentrations between the previous three methods described above, this method yields a phage concentration increased by approximately 1600%. The total time necessary to carry out this protocol is two days.

**Table 1** Purification methods investigated

<table>
<thead>
<tr>
<th>Purification Method</th>
<th>Yield (pfu)</th>
<th>Titer (pfu/mL)</th>
<th>Protein Conc. (μg/mL)</th>
<th>Purity (pfu/μg protein)</th>
<th>Volume (mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CsCl Gradient</td>
<td>1.34x10\textsuperscript{10}</td>
<td>8.90x10\textsuperscript{10}</td>
<td>2850</td>
<td>2.81x10\textsuperscript{6}</td>
<td>1.5</td>
</tr>
<tr>
<td>5%-30% Sucrose</td>
<td>5.25x10\textsuperscript{14}</td>
<td>2.10x10\textsuperscript{10}</td>
<td>2047</td>
<td>1.03x10\textsuperscript{7}</td>
<td>2.5</td>
</tr>
<tr>
<td>Both CsCl+Sucrose</td>
<td>9.23x10\textsuperscript{14}</td>
<td>3.69x10\textsuperscript{10}</td>
<td>405</td>
<td>9.11x10\textsuperscript{7}</td>
<td>2.5</td>
</tr>
<tr>
<td>PEG precipitation</td>
<td>9.08x10\textsuperscript{14}</td>
<td>1.80x10\textsuperscript{11}</td>
<td>230</td>
<td>7.83x10\textsuperscript{8}</td>
<td>5</td>
</tr>
</tbody>
</table>

**Table 1 Purification methods investigated** This table contains the yield (pfu), titer (pfu/mL), protein concentration (μg/mL), purity (pfu/μg protein), and volume (mL) of the four purification methods investigated. The PEG precipitation based method gives the best purification efficiency and is ~759% better than the next best protocol.
The yields of total phage (pfu) were normalized to the total protein concentration (μg) in order to quantitatively compare the different purification methods. CsCl gradients alone were by far the least efficient method, with a phage to protein ratio of $3.12 \times 10^6$ pfu/μg. The purification efficiencies of sucrose gradients alone and the combination of CsCl and sucrose gradients together were comparable, with phage to protein ratios of $1.03 \times 10^7$ and $9.11 \times 10^7$ pfu/μg, respectively. The PEG precipitation method was by far the most efficient, with a phage to protein ratio of $7.83 \times 10^8$ pfu/μg. The measured purification efficiency, sample volumes, phage concentrations, and time required indicated that the PEG purification based method would be the most effective for subsequent biophysical characterizations.

**Biophysical characterization of a Microvirid bacteriophage**

Biophysical characterization of viral systems is notoriously difficult because of their inherent complexity. Only groups that study viral systems that have simple protein composition, ability to assemble in the absence of hosts, and established protocols that yields data with unambiguous interpretation are successful in their attempts to empirically evaluate capsid stability and the kinetics that govern their assembly and in some cases their disassembly (10, 16, 23, 62, 74). A major effort of the viral adaptation experiments in the Rokyta lab is to understand capsid stability from an evolutionary standpoint. In this study I will focus on building a repertoire, as well as an understanding, of biophysical techniques that are applicable to our system. The first attempt to characterize the disassembly reaction and monitor capsid stability was focused on using differential scanning calorimetry (DSC), as has been done with phage HK97 (16). Phage samples are complex and a sample with high protein concentration does not correlate with high phage concentration. Although this does become true as sample purity improves, protein concentrations below ~1 mg/mL are not optimal for DSC studies. Thus, DSC signals were weak, uninterpretable, and indicative of sample aggregation. The remaining attempts to characterize capsid stability to dissociation were made using spectroscopy techniques monitoring intrinsic fluorescence as well as static light scattering. Assembly kinetics were also measured using standard kinetics assays. All experimental techniques were done using ID8*, a genotype isolated during purification (Table 2).

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Protein</th>
<th>Protein Name</th>
<th>Nuc position</th>
<th>ΔNuc</th>
<th>Aa position</th>
<th>ΔAa</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID8*</td>
<td>coat</td>
<td>F</td>
<td>3557</td>
<td>A→G</td>
<td>331</td>
<td>K→R</td>
</tr>
<tr>
<td>γ (dblgs/hr)</td>
<td>St. dev.</td>
<td>δ (-dblgs/hr)</td>
<td>St. dev.</td>
<td>ω (dblgs/hr)</td>
<td>St. dev.</td>
<td></td>
</tr>
<tr>
<td>16.60</td>
<td>0.48</td>
<td>-49.41</td>
<td>3.76</td>
<td>20.72</td>
<td>0.63</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2 Genotype used for characterization studies** The characteristics of the genotype used in this study, ID8*. ΔNuc and ΔAa indicate changes from an ancestral genotype that is used as the wild-type in adaptation experiments.
Fluorescence spectroscopy and static light scattering analyses of ID8*

Proteins contain three aromatic residues (tryptophan, tyrosine, and phenylalanine) that contribute to intrinsic fluorescence. Each aromatic residue exhibits characteristic quantum yields, lifetimes, intensities, and excitation/emission wavelengths. Tryptophan is a much stronger fluorophore than both tyrosine and phenylalanine and its emission ranges are highly dependent on its environment. Tryptophan typically absorbs radiation at 280 nm and emits from 300 to 350 nm depending on the environment (42). Tryptophan residues that are buried within the hydrophobic core of a protein can have emission spectra shifted 10 to 20 nm versus those that are solvent exposed (42). Changes in tryptophan fluorescence, and thus changes in tryptophan environment, can be a highly sensitive probe for measurement of conformational states of a protein. In these experiments fluorescence was monitored as heat was increased at a specific rate to yield fluorescence melting curves. The excitation and emission wavelengths used were 295 nm and 340 nm respectively. Using 295 nm as the excitation wavelength ensures that the tryptophan emission spectrum is dominant over both the tyrosine and phenylalanine emission spectra (42). Static light scattering monitors the scattering of photons by particles in solution. This was monitored at excitation and emission wavelengths of 295 nm. Light scattering is indicative of aggregation and loss of macromolecular structure, such as tertiary and quaternary structure.

Figure 3a shows the raw fluorescence (black) and static light scattering (blue) melting curves observed for one repetition for ID8* at a ramp rate of 1 °C/min. When incubated at 4°C for five to ten minutes prior to measurement, a reversible transition is observed at approximately 25.3 ± 2.3°C (table 3). The transition is inconsistent, as evidenced by the large standard deviation, and varies in shape as well as intensity between runs. It is not observed if the sample is incubated and equilibrated to 20°C prior to measurement. The odd shape, general inconsistency between runs, and the static light scattering data, which displays a large increase in scattering intensity, indicates that this transition is likely due to transient aggregation at cold temperatures. It could possibly be a conformational rearrangement, as has been observed in the minute virus of mice, but it is unlikely based on the data observed (8). Figure 3b shows how well the decrease in light scattering and concurrent increase in fluorescence intensity overlay, again suggesting an aggregation event is occurring. There is a second transition that occurs at 69.8 ± 0.3°C (table 3) that appears to follow a two-state model. This transition, unlike the transition that occurs earlier in the melting curve, is consistent and displays a constant shape between runs and different sample preparations. It is likely representative of global capsid dissociation and is irreversible. The transition almost perfectly overlays with the decrease in scattering intensity (figure 3c), which is also consistent with the interpretation that it represents global capsid dissociation. Lastly, figure 4 shows the emission spectrum (an average of three replicates) of a fresh sample prior to analysis and the denatured sample after analysis. There is a shoulder that is observed in all emission spectra and is likely due to the complex nature of the sample and presence of contaminating protein. The denatured emission spectrum (magenta) is substantially red-shifted compared to the fresh emission spectrum (black) with respective emission maxima at 362.92 nm and 347.84 nm. This is also consistent with global capsid dissociation occurring during the melting curve because the exposure of hydrophobic tryptophan residues to aqueous solvent is known to red-shift the emission spectrum (42).
**Figure 3 Raw fluorescence and light scattering melting curves** a.) An overlay of the entire raw fluorescence and static light scattering curves observed at a scan rate of 1°C/min. b.) A view of the first transition observed in 3a. c.) A view of the second transition observed in 3a.

**Figure 4 Emission spectra of fresh and denatured samples** An overlay of the emission spectrum of the fresh sample (black) before being subjected to melting experiments and the emission spectrum of the denatured sample (magenta) after melting. The intensities in each curve were normalized to their respective emission maxima. Each curve represents an average of three curves.
The second transition observed with fluorescence and light scattering methods, although irreversible and unable to be analyzed by equilibrium thermodynamics, was found to fit well to a two-state model using the program Datafit (Oakdale Engineering, Oakdale, PA). The program was used to extract a precise value for the $T_{0.5}$, or the temperature at which half of capsids are dissociated during the transition. $T_{0.5}$ is not a thermodynamic quantity and is not analogous to $T_m$. Figure 5a shows the normalized fluorescence data for five replicates using ID8* and figure 5b shows both normalized fluorescence and scattering overlaid. The data were normalized by converting the native and denatured signals to a ratio that describes the capsids in the native versus dissociated state, mass fraction capsid (mfc), which was previously described in terms of HBV disassembly and is mathematically described in the methods section of this thesis (10). The $T_{0.5}$ at 1°C/min for ID8* was determined to be 69.8 ± 0.3°C (table 3).

**Figure 5 Normalized fluorescence and light scattering data 1°C/min** a.) Normalized fluorescence curves acquired using a scan rate of 1°C/min. b.) Normalized fluorescence curve (black) overlaid with the normalized static light scattering curve (blue) acquired using a scan rate of 1°C/min.

<table>
<thead>
<tr>
<th>Table 3 ID8* $T_{0.5}$ data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate (°C/min)</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>0.25</td>
</tr>
<tr>
<td>0.50</td>
</tr>
<tr>
<td>1.0</td>
</tr>
<tr>
<td>2.0</td>
</tr>
<tr>
<td>6.67</td>
</tr>
</tbody>
</table>

**Transient aggregation $T_{0.5}$**

<table>
<thead>
<tr>
<th>Rate (°C/min)</th>
<th>Avg $T_{0.5}$ (°C)</th>
<th>St. dev. (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>25.3</td>
<td>2.3</td>
</tr>
<tr>
<td>6.67</td>
<td>42.8</td>
<td>2.7</td>
</tr>
</tbody>
</table>

**Table 3 ID8* $T_{0.5}$ Data** The $T_{0.5}$ (°C) data acquired from the Datafit program corresponding to various scan rates for both the major capsid dissociation transition as well as the transient aggregation transition.
The irreversibility of this system prompted an investigation of its rate dependence. For this analysis the temperature scan rates investigated were 0.25, 0.50, 1.00, and 2.00°C/min and are shown in figure 6a. As expected, the system’s T_{0.5} does exhibit scan rate dependence and is characteristic of systems that display hysteresis, or failure of opposing reactions to equilibrate. Hysteresis has been widely observed in viral systems and has been indicated as a mechanism of kinetic stabilization that ensures maintenance of viral activity in extreme environments (18, 57, 73). It also accounts for the contrast between the observation of metastable capsid structures produced by assembly reactions and exceedingly stabilized structures in response to extreme environments (73). Concentration dependence was also investigated (figure 6b). Fluorescence spectra were relatively insensitive to concentration differences. This supports fluorescence spectroscopy as a feasible method to evaluate phage stability as the determination of viral titers are subject to the error associated with plating assays and thus are inherently subject to differences in concentration between preparations.

Fluorescence spectra and scattering data are useful to monitor changes in protein structure but are ambiguous when trying to interpret what specific transitions correspond to within complicated structures such as viruses. In an attempt to link the transitions to specific changes in viral structure, plaque assays were done that required removing phage aliquots at specific temperatures during a melting scan. The normalized plaque assay data was overlaid with the normalized fluorescence curve shown in figures 7a and 7b. Both were performed at a scan rate of 1°C/min. Figure 7a shows a major transition in the phage population surviving at ~57°C, which is well before the T_{0.5} of the major capsid dissociation transition at 69.8±0.3°C. This transition results in the loss of ~99% of the population, or a decrease from \( \sim 10^{10} \) pfu/mL to \( \sim 10^8 \) pfu/mL. Figure 7b shows the ratio of phage surviving on a log_{10} scale. Although one would expect fluorescence emission to be insensitive to 1% of the population, the \( 10^8 \) pfu/mL population is still biologically relevant in terms of adaptation. The \( 10^8 \) pfu/mL phage population decays to essentially zero almost concurrently with the major fluorescence transition. The loss of 100% activity therefore occurs directly after the transition corresponding to global capsid dissociation.
Figure 7 Viral Activity versus major fluorescence transition a.) An overlay of the normalized fluorescence curve obtained using a scan rate of 1°C/min and plaque assay data normalized by dividing the phage surviving over the phage titer at 4°C. Plaque assay data was acquired under the same scan rate. b.) An overlay of the normalized fluorescence curve obtained using a scan rate of 1°C/min and the log\textsubscript{10} scale of the same plaque assay data as 7a. The log\textsubscript{10} scale was used to show the second transition that occurs in a biologically relevant population of \(\sim 1 \times 10^8\) pfu/mL.

Figure 8 shows a schematic that unites the structural transitions that occur in the plaque assays and in the fluorescence melting curves. Below \(\sim 55°C\), the phage population is relatively stable at about \(1 \times 10^{10}\) pfu/mL. A transition in phage activity occurs that is consistent with loss of a majority of the spike protein G (red), which is required for viral activity and host attachment. This is consistent with the decrease in viral activity as well as the lack of signal change in fluorescence and light scattering data. The spike protein is relatively small (19 kDa) versus the major coat protein (48 kDa). Although there is a 99% decrease in the phage population, there is still a biologically relevant population of about \(1 \times 10^8\) pfu/mL. There are 12 spike proteins arranged on the viral capsid and phages only require one spike protein to begin an infection cycle. The survival of this population is therefore justified under the hypothesis that the \(\sim 57°C\) transition corresponds to loss of a majority of the spike protein from viral capsids. The population is stable through the global dissociation transition observed in fluorescence, though the population rapidly decreases to \(\sim 1 \times 10^4\) pfu/mL after the transition is finished, or above 70°C. Above 77°C the population decreases to nearly zero and loss of 100% viral activity. The extreme decreases in viral activity observed after the major fluorescence transition is consistent with it corresponding to major capsid dissociation into constituent F protein pentamers (blue) and possibly monomers, although it is not clear if major capsid dissociation to pentamers and various intermediates occurs concurrently with dissociation/unfolding to monomers.
Figure 8 Structural transitions in response to heat: The hypothesized structural transitions accompanying thermal melting in G4-like phages (PDB code: 1gff). Heating to 57°C causes loss of the major spike protein G and is accompanied by a 99% decrease in the phage population. Heating to 70°C causes major capsid structural loss; above 77°C the remaining phage population decreases to zero rapidly and it can be assumed that all capsid integrity is lost.
Viral assembly kinetics

The assembly kinetics of bacteriophages can be qualitatively assessed for comparison using standard bacteriophage kinetics assays (68). Although alone the assembly kinetics growth curves are not exceptionally enlightening, they are integral to comparisons between mutants and especially so in the case of our adapted phages where there exists pleiotropic interactions between growth rate and capsid stability. For these experiments, an *E. coli* strain (RY7211) was used that confers resistance to E protein mediated cell lysis (67). The viral E protein inhibits peptidoglycan synthesis such that when host cells divide there is not enough peptidoglycan to fill the bacterial cell wall. The cells succumb to osmotic pressure and phage leak out of the cell. In the case of RY7211, the cells are never subjected to osmotic pressure and phage are not released to infect other hosts. Over a 60-minute time period phages are able to assemble within hosts and are only separated during time point acquisition using a CHCl₃ saturated solution containing 2 mg/mL lysozyme to ensure complete cell lysis. Plaque assays are then used to evaluate the extent of the assembly reaction.

There are a few characteristic features shared by bacteriophage growth curves. The growth curve for ID8* is shown below in figure 9. The dip observed between two and six minutes represents the lag phase of assembly, where viruses release genomic information, becoming non-infectious and causing a dip in pfu/cell, and subunits begin to accumulate and form capsid nuclei. The rapid increase in pfu/cell between 10 and 18 minutes represents the elongation phase in which capsids rapidly accumulate. Beyond 18 minutes, available intermediates are depleted and capsid growth plateaus.

![Figure 9 ID8* assembly kinetics](image)

*Figure 9 ID8* assembly kinetics* Growth kinetics of phage genotype ID8*. 

26
Chapter 3 Conclusions and future direction

Section 3.1 Conclusions

The work presented in this thesis sets up a framework for purifying and understanding the biophysical properties associated with Microvirid phages. Although these bacteriophages are extremely complicated in nature, we now know a few key pieces of information that are imperative for carrying out a systematic biophysical analysis of extreme heat and low-pH adapted viruses:

- The PEG precipitation based method is the most efficient purification protocol for Microvirid bacteriophages with a purity value of $7.83 \times 10^8$ pfu/μg protein, which is an increase of approximately 759% over the next best method.
- When samples are pre-equilibrated to 4°C, transient aggregation is observed as an oddly shaped, reversible, and inconsistent fluorescence/light scattering transition.
- An irreversible transition is observed above 60°C that is dependent on the temperature scan rate but is relatively insensitive to phage concentration. The concurrent decrease in light scattering and red-shifting of emission spectra before and after thermal melts are consistent with this representing global capsid dissociation.
- The seemingly conflicting 99% decrease in phage activity that occurs before the major fluorescence transition is remedied by the hypothesis that this represents loss of the majority of spike proteins from the capsid structure, which are required for phage activity.
- Log$_{10}$ scales show that the 1% of the phage population surviving undergo a decrease in activity to essentially zero after the fluorescence transition that corresponds to global capsid dissociation.

Although bacteriophages are an extremely complicated system, this study has shed light on applicable biophysical methods and an understanding of the data yielded from those analyses.

Section 3.2 Future direction

A systematic biophysical characterization of heat- and pH-adapted bacteriophages is the next step in this analysis, which would include collecting fluorescence and light scattering melting curves at multiple rates and assembly kinetics assays. Another necessary experiment that will add weight to the biological relevance of the hysteresis observed in fluorescence melts as well as the hypothesis that the transition in the plaque assay corresponds to loss of the spike protein is to perform the plaque assay at various rates. If the plaque assay transitions shift as the fluorescence transitions do it supports the assertion that kinetic stabilization plays an important role in the survival, and thus evolution, of viruses in extreme environments. Finally, there is ambiguity as to the nature of structural intermediates that exist beyond the global dissociation curve. It is unknown whether global capsid dissociation occurs concurrently with pentamer dissociation and unfolding of monomers or major capsid protein pentamers are stable in solution. Future experiments include using CD spectroscopy to monitor changes in secondary structure at high temperatures. CD spectroscopy has been used in this manner in more structurally simple viruses such as Hepatitis B (62).

Studying the assembly and dissociation processes of viruses affords an opportunity to study biophysical properties of large, self-assembling protein complexes which are currently poorly understood. Viruses are also unique in that their capsids must fulfill conflicting infection cycle requirements. Capsids must assemble rapidly with fidelity and specificity into a structure that is
stabilized to extreme extracellular environments. They must also be able to dissociate and release genomic information to propagate cellular infections. Viruses therefore display evolutionarily fine-tuned structure/function relationships. Biophysical characterization of the phenotypes displayed by our adapted bacteriophages will aid in understanding the physical mechanisms that underlie adaptation to extreme environments as well as the relationship between tightly linked phenotypes in complex protein systems.
Works Cited


