Structural, Biophysical, and Functional Studies on the Head-Tail Joining Reaction of Bacteriophage λ

Karen Lee Maxwell

Thesis submitted in conformity with the requirements for the Degree of Doctor of Philosophy, Graduate Department of Molecular and Medical Genetics in the University of Toronto

© by Karen Lee Maxwell 2001
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.
Structural, Biophysical, and Functional Studies on the Head-Tail Joining Reaction of Bacteriophage λ

Doctor of Philosophy 2001
Karen Lee Maxwell
Graduate Department of Molecular and Medical Genetics
University of Toronto

ABSTRACT

Many crucial processes occurring within cells are mediated by macromolecular complexes. The bacteriophage λ provides an excellent system with which to characterize a macromolecular assembly process. The λ morphogenetic pathway is very complicated, involving the assembly of 22 different proteins and DNA. In this thesis, I examine a single step in the pathway, the head-tail joining reaction mediated by gpW and gpFII, at a molecular level.

gpW is composed of only 68 residues, yet must interact with at least two other proteins in the phage, and probably with DNA. I found that fusion proteins with hexahistidine-tag sequences at either the N- or C-terminus are stably folded and biologically active. The protein is monomeric as judged by equilibrium ultracentrifugation, and appears to unfold by a cooperative two-state mechanism. Circular dichroism studies indicate that the protein is 47% helical, with a $T_m$ of 71.3 °C, and a $\Delta G_u$ of 3.01 kcal/mol at 25 °C. Mutagenesis of the three hydrophobic C-terminal residues of gpW showed that they are critical for activity, even though they do not contribute to the thermodynamic stability of the protein. I determined the solution structure of gpW using NMR spectroscopy and found it possesses a novel
fold consisting of two α-helices and a single two-stranded β-sheet arranged around a single, well-packed hydrophobic core. gpW displays an intriguing structural similarity with the "sliding clamp" proteins associated with DNA polymerases. Using a sliding clamp structure, gpW was modeled into a structurally plausible hexameric ring, which is consistent with its known roles in bacteriophage λ morphogenesis.

The NMR solution structure of gpFII was determined to be a novel fold comprised of 7 β-strands arranged in a twisted β-sandwich structure. A remarkable feature of this protein is the presence of two large unstructured regions, which we postulate may play a role in viral assembly. My results suggest that the two head-tail joining proteins from bacteriophage λ, gpW and gpFII, may both employ a disorder-to-order transition to control protein assembly in the virion.
Acknowledgements

I would like to thank my supervisor, Dr. Marvin Gold, for his support and encouragement during my three years in his lab. I am grateful to him for allowing me the opportunity to develop as an independent thinker and scientist. I would like to thank Helios Murialdo for bringing gpW into my life, and Andy Becker for his support of my research. Thanks also to my committee members Cheryl Arrowsmith and Jun Liu for their guidance over the course of this project, and to the members of the Sadowski and Funnell laboratories for many stimulating lab meetings.

My NMR work would never have been completed without the help of the Arrowsmith lab, especially Adelinda Yee and Valerie Booth who patiently answered my never-ending questions about NMR, structure calculations, computers programming, etc...

Finally, I would like to thank Vanita for her good cheer throughout our graduate student years, and Alan for everything he's done.
# TABLE OF CONTENTS

Chapter 1: *Introduction*  
The Bacteriophage λ  
- Assembly of the Bacteriophage λ Particle  
- Protein W  
- Protein FII  
- DNA Injection  
- Atomic Structures of Bacteriophage λ Heads and Tails  
- Atomic Structures of Bacteriophage Head-Tail Connectors  
Protein-Protein Interactions  
Thermodynamic Analysis of Protein Stability  
Solving Protein Structures Using NMR  
Thesis Overview  
References  

Chapter 2: *Thermodynamic and Functional Characterization of Protein W from Bacteriophage λ: The Three C-terminal Residues are Critical for Activity*  
Introduction  
Materials and Methods  
- Plasmid Construction  
- Protein Expression and Purification  
- Determination of Thermodynamic Parameters by CD Spectroscopy  
- Molecular Weight Determination  
- *In Vitro* Assay of gpW Activity  
Results  
- Secondary Structure Prediction and Analysis  
- Functional Characterization of His-tagged gpW Constructs  
- Determination of the Native Molecular Weight of gpW  
- Thermodynamic Characterization of His-tagged gpW Constructs  
- Mutagenesis of the C-terminus of gpW  
- Destabilizing Substitutions in gpW  
Discussion  
- Structural Features of gpW  
- Functional Characterization of gpW  
- The Role of the Three C-Terminal Residues
Chapter 3: *The novel fold of the bacteriophage λ protein W reveals structural similarity to the DNA sliding clamps.*

Introduction 53

Materials and Methods
- Cloning, Expression, and Purification of gpW 53
- NMR Spectroscopy 54
- Structure Determination 54
- Modeling of the Hexameric Ring 55

Results and Discussion
- The solution structure of monomeric gpW 55
- The Structure of gpW Resembles DNA Sliding Clamps 56
- Modeling gpW Into a Sliding Clamp-Like Ring 59
- Functional Implications of the gpW Ring 61
- Conclusion 64

References 65

Chapter 4: *The Solution Structure of the Bacteriophage λ Head-Tail Joining Protein, gpFII*

Introduction 70

Materials and Methods
- Sample preparation 71
- Determination of thermodynamic parameters by CD spectroscopy 72
- *In Vitro* Assay for gpFII Activity 72
- NMR Spectroscopy and Structure Calculations 72

Results
- Functional Characterization of gpFII 73
- Biophysical Characterization of gpFII 74
- NMR Solution Structure 75
- Mixing Experiments with gpW and gpFII 79

Discussion
- Structural Features of gpFII 82
- The Genetic Source of gpFII Determines the Tail Specificity 85
- Conclusion 86

References 88
LIST OF FIGURES

Figure 1-1. Assembly pathway of the bacteriophage λ virion. 5
Figure 1-2. Crystal structure of the φ29 connector. 13

Figure 2-1. Sequence alignment of gpW and protein constructs. 37
Figure 2-2. In vitro activity of gpW(NT) as a function of time and protein concentration. 39
Figure 2-3. Analytical ultracentrifugation of gpW(NT). 40
Figure 2-4. CD spectra of folded and unfolded gpW. 41
Figure 2-5. Unfolding of the three WT gpW constructs. 42
Figure 2-6. ΔG_u versus temperature plot for gpW(NT). 43

Figure 3-1. NMR solution structure of gpW. 57
Figure 3-2. gpW shows structural similarity to the DNA sliding clamp 60
Figure 3-3. gpW can be modeled into a hexameric ring. 62
Figure 3-4. Alignment of the C-terminus of gpW with sliding clamps. 63

Figure 4-1. Size exclusion chromatography of gpFII. 74
Figure 4-2. Biophysical characterization of gpFII. 75
Figure 4-3. NMR solution structure of gpFII. 78
Figure 4-4. Sequence alignment of gpFII. 83
Figure 4-5 NMR mixing experiments with gpFII and gpW. 80
Figure 4-6 NMR mixing experiments with gpW (CT) and gpFII. 81
Figure 4-7 Analysis of gpFII structure reveals two binding surfaces. 87

Figure 5-1. HSQC of gpU, the bacteriophage λ tail-terminator protein. 95
LIST OF TABLES

Table 1-1. The properties of nuclei commonly used in NMR. 21

Table 2-1. Characterization of the three WT gpW constructs. 36
Table 2-2. Effects of amino acid substitutions on the stability and biological activity of gpW. 45

Table 3-1. Structural statistics for the family of 15 gpW structures 58

Table 4-1. Structural statistics for the family of 10 gpFII structures. 77
LIST OF ABBREVIATIONS

λ: bacteriophage lambda
CD: circular dichroism
ΔCp: change in heat capacity upon unfolding
ΔGu: free energy of unfolding in water at 25 °C
ΔH: enthalpy of unfolding
ΔS: entropy of unfolding
gpFII: gene product FII
gpW: gene product W
gpW(CT): C-terminally 6-His tagged gpW
gpW(NT): N-terminally 6-His tagged gpW
GuHCl: guanidine hydrochloride
HSQC: heteronuclear single quantum correlation
IPTG: isopropyl-β-D-thiogalactopyranoside
kDa: kilodaltons
MAD: multiple anomalous dispersion (phasing)
NMR: nuclear magnetic resonance
NOE: nuclear Overhauser effect
pfu: plaque forming units
Se-Met: selenomethionine
Tm: temperature at the midpoint of the thermal unfolding transition
WT: wild type
Chapter One

General Introduction
Genomic era research will soon progress from the identification and characterization of components of large multi-subunit assemblies to an understanding of how they integrate to produce the molecular machines that generate basic cellular functions. Detailed knowledge of the structure and dynamics of these macromolecular complexes and how the structures of the individual components change upon assembly will be of crucial importance. As the protein complexes that underlie cellular processes like signal transduction and DNA replication are transient in nature, model systems that can be examined at a molecular level are necessary. Because viruses are complex assemblies of protein and nucleic acids, information about their structures provides not only an understanding of viruses themselves, but also knowledge that is more widely applicable to macromolecular complexes in general. Thus, viral structures can serve as a paradigm for solutions to problems concerning the assembly and function of large multi-subunit structures within a cell.

The Bacteriophage λ

The bacteriophage λ presents a unique opportunity for structural biologists to intensively study a complex macromolecular system. The well characterized nature of this phage, and the ease with which it can be genetically manipulated makes it a useful model for studying macromolecular structure, protein-protein interactions, and multi-subunit assembly. Furthermore, the enormous body of mutational data from λ studies, and sequence information from a large number of related bacteriophages provide guides for experiments linking structure with function. Bacteriophage λ is the prototype of a group of temperate phages, referred to as the lambdoid phages, that are capable of productive genetic recombination with λ. These phages have three properties in common: they recombine when crossed, their DNA molecules possess identical pairs of cohesive ends, and their prophages are induced by UV irradiation. All members of the family of lambdoid phages have a similar genetic map, in the sense that the order of the genes relative to the functions specified by those genes is always the same. The order of functions in the late operon is lysis, followed by DNA maturation and head assembly, and finally tail assembly, adsorption, and injection. The morphogenetic genes are present in non-overlapping gene clusters, and the proteins encoded by the genes within these clusters that interact most intimately in the phage particle are coded by adjacent genes (Casjens & Hendrix, 1974). The invariant order of the genes is maintained even in cases where the actual functions encoded at each position are radically different. This
arrangement may have evolved to conserve groups of genes that interact by minimizing the possibility of recombination within such groups. This phenomenon results in the formation of groups of genes coding for functional units such as prohead assembly, DNA packaging, and tail assembly that can be exchanged among related phages and then be allowed to diverge. Such recombination events could occur in nature, as evidenced by the isolation of a viable λ phage that had acquired only the head genes from phage 21 (Feiss et al., 1981). This idea is also supported by the existence of the same gene order for homologous genes in other lambdoid phages and the divergence of lambdoid phage proteins in terms of molecular weight and functional specificity.

Bacteriophage λ is one of the most well characterized viruses, with all major aspects of its life-cycle having been intensively studied over the past 50 years. λ was identified in 1951 (Lederberg, 1951) as a prophage in the K12 strain of E. coli, which had been in laboratory culture since its isolation from a human patient in California in 1922. The phage is an obligatory parasite, and its life cycle begins when a particle adsorbs to the surface of an E. coli host cell and it injects its viral DNA. Upon infection, λ can grow lytically, which results in the eventual lysis of the cell and release of approximately one hundred phage particles, or it can insert its genome into the host genome and form a prophage, which may remain dormant for many generations. The small size of the bacteriophage λ belies the complexity of the functions that it must perform in order to propagate. To successfully infect the host E. coli cell, the phage must be structurally stable in the potentially harsh conditions of the extracellular environment. It must be able to recognize and bind specifically to the maltose receptor on the surface of an E. coli cell, and should not bind non-specifically to other receptors or other bacteria. After adsorption to the cell the phage must change conformation to allow the DNA to exit from the head, pass through the tail, and into the host cytoplasm. Once the bacteriophage DNA has entered the host cell, it must enlist the host cell's machinery in order to replicate its genome and subsequently produce progeny DNA and phage proteins. The newly synthesized viral proteins and concatameric genomes must then be assembled into phage particles, and the host cell must be lysed to allow their release. This whole process, which involves at least 46 different phage-encoded proteins and a large number of host factors, must be carried out with great accuracy and efficiency in a matter of a few minutes.
Assembly of the Bacteriophage \(\lambda\) Particle

The \(\lambda\) virion is comprised of a double stranded DNA-filled icosahedral head that is 0.05 mm in diameter, and a 0.15 mm long non-contractile tail. The 48.5 kb mature \(\lambda\) genome contains twelve nucleotide self-complementary single stranded DNA overhangs at its 5'-terminal ends known as cohesive ends. Following genome injection into the host cell, these overhangs, with the help of host-supplied DNA ligase, anneal to form a closed circular DNA molecule. During a lysogenic infection, the \(\lambda\) DNA molecule is then inserted in the \(E.\ coli\) genome by a \(\lambda\) encoded site specific recombinase and is passively replicated as part of the bacterial chromosome. In lytic growth the \(\lambda\) proteins gpO and gpP initiate the replication of the \(\lambda\) DNA molecule by sequestering the host DNA replication machinery at the \(\lambda\) origin of replication. The first rounds of DNA replication proceed by the theta model and generate more circular DNA molecules. At a later stage the replication switches to a rolling-circle model, which results in the production of linear concatamers consisting of many \(\lambda\) genomes joined end-to-end.

As DNA replication continues in the host cell, other \(\lambda\) genes are also expressed. During the late stages of infection, the 22 genes involved in morphogenesis of the viral particle are transcribed from a single promoter, PR'. The products of ten of these genes and groEL and groES, which are supplied by the host cell, are required for the production of heads. The head and tail of the virus are assembled in separate pathways. Head morphogenesis initiates with the assembly of the preconnector, a doughnut-shaped structure composed of 12 subunits of uncleaved gene product (gp) B, in a reaction that requires the presence of gpNu3, a scaffolding protein, and host-supplied groEL and groES proteins (Figure 1-1) (Murialdo & Becker, 1978; Murialdo, 1979). The connector has a 2-2.5 nm hole in its center which is believed to be the entry point for DNA during packaging (Kochan et al., 1984), and the exit point for DNA upon infection of a host cell. The preconnector interacts with gpC, and acts as the starting point for the polymerization of 405 molecules of gpE into the head (Murialdo & Becker, 1978). Next, about three quarters of the B molecules are cleaved, all of the Nu3 is degraded, and each of the C molecules participates in a fusion/cleavage reaction with gpE to form the proteins pX1 and pX2. Terminase, a hetero-oligomer composed of gpNu1 and gpA, then binds to the connector and to the concatameric DNA, and packages linear monomers into the empty head using a hole in the connector as an entry point (Murialdo, 1991). gpFI plays an accessory role in the packaging process by facilitating the interaction of the terminase-DNA complex with empty heads (Davidson &
Gold, 1987). The packaged DNA is stabilized within the head by expansion of the capsid by approximately 20%, and the incorporation of 405 molecules of gpD into the surface lattice (Imber et al., 1980), which thereby completes the capsid.

The tail of bacteriophage λ assembles in a morphogenetic pathway that takes place simultaneously, but is separate from the formation of heads. Tail formation is controlled by at least 12 genes (Katsura, 1983) which are clustered in the left end of the λ genome, immediately adjacent to the head genes. If one of the tail genes is defective, completed heads and tail precursors are produced instead of whole phage. The heads produced are active, and can spontaneously bind to free tails produced by phage mutants in the head genes and yield infectious phage particles (Weigle, 1968; Harrison et al., 1973), illustrating that heads and tails are assembled independently and join at the last step of assembly.

Tail assembly occurs in three major events, beginning with the formation of the initiator complex (Figure 1-1). This first step involves the assembly of two to four copies of gpJ, the tail fiber that is located at the end of the tail distal to the head. This allows the bacteriophage to attach to the surface of the cell by binding to the outer membrane protein lamB. The products of genes I, L, K, H, G, and M then act sequentially on the tail fiber as determined by in vitro complementation experiments and complete the initiator complex. The second event is the polymerization of the major tail protein, gpV into a tube-like structure comprised of 32 stacked disks. Each disk is formed from six copies of gpV, arranged so that there is a central hole 3 nm in diameter (Katsura & Hendrix, 1984; Katsura, 1987). This allows the third event, termination and maturation to occur. Following the polymerization of the gpV rings, gpH (92 kDa) is cleaved to form a 78 kDa protein called pH* (Tsui & Hendrix, 1980), and a hexamer of gpU and gpZ add to the head proximal end of the tail (Katsura & Tsugita, 1977). gpU, the terminator protein, stops the polymerization at the correct tail length and yields a normal looking, but defective, tail which is activated by gpZ. In the absence of gpU, gpV continues polymerization and forms a polytail (Mount et al., 1968). Cleavage of gpH occurs at a step in the assembly pathway between the action of gpU and gpZ. Unattached tails contain pH*, but neither pH* nor gpH is found in polytails made by gpU− bacteriophage (Katsura, 1976).

The final step in the assembly of bacteriophage λ involves the attachment of the DNA-filled capsid with the tail at the ring-shaped connector located at one vertex of the head (Harrison et al., 1973). The connector is formed from 12 copies of the head protein gpB, and its cleaved form gpB*, which are assembled into a ring
Figure 1-1. The morphogenetic pathway of bacteriophage λ.
with 12-fold symmetry. The products of genes $W$ (gpW) and $FII$ (gpFIII) mediate this reaction which can be carried out \textit{in vitro}, where gpW must act before gpFII (Casjens \textit{et al.}, 1972). \textit{In vivo}, infection by phages carrying amber mutants in genes $W$ or $FII$ results in the accumulation of normal heads containing mature DNA molecules, and fully assembled tails, which cannot be joined. The head-tail connector is an essential structure as the hole through the ring-shaped connector provides the only egress for the DNA contained within the phage head. During infection, the DNA must pass out of the head, and through the length of the tail before it is ejected from the tail’s distal tip.

\textit{Protein $W$}

gpW, the 68 residue head stabilizing protein, has a molecular weight of 7,614, and a calculated pI of 10.8. Phage lacking gpW produce heads of the same diameter as wild type $\lambda$, DNA molecules of monomer length with cohesive ends, and active tails that are capable of efficiently complementing heads from a $J^-$ infection (Casjens \textit{et al.}, 1972; McClure \textit{et al.}, 1973). Katsura and Tsugita (1977) detected a gpW complementing activity of approximately 10 kDa in guanidine hydrochloride dissociated phage ghosts, suggesting that it is a structural protein, although its precise stoichiometry is unknown. Partially purified gpW has been shown to interact with DNA-filled proheads \textit{in vitro} after the incorporation of gpD. Although no direct evidence has pinpointed the localization of gpW within the phage particle, it is required for the incorporation of gpFII, which has been conclusively shown to be bound at the connector (Tsui & Hendrix, 1980). Thus, in the completed virion gpW probably interacts with both gpFII and some other component of the head. The most likely candidate is gpB, since this forms the portal collar where the tail attaches. The DNA inside heads in $W$- phage has an increased sensitivity to DNase treatment (Perucchetti \textit{et al.}, 1988), and is more likely to spontaneously come out of the head, suggesting that gpW may play a role either as a structural protein that forms a plug at the portal prohead vertex, or as a DNA-binding protein that interacts directly with the right end of the $\lambda$ DNA molecule which protrudes into the tail in fully assembled phage. The high pI (10.8) of gpW is consistent with a direct DNA interaction. The apparent ability of this very small protein to mediate interactions with two different proteins and DNA is remarkable.
Protein FII

The importance of gpFII in attaching the connector and the tail has been demonstrated both biochemically and genetically. When cells are infected with phage mutants lacking heads, the connector is seen to become attached to the end of tails, a phenomenon that does not occur if gpFII is missing (Tsui & Hendrix, 1980). gpFII is a 117 residue protein with a molecular weight of 11.5, and an isoelectric point of 4.75 (Casjens, 1974). Phage lacking gpFII produce stable DNA-filled heads, and completed tails, but fail to join them together into a mature, infectious virion. Tsui and Hendrix (1980) discovered that the biological activity of gpFII could be isolated from two different peaks when phage treated with 3M GuHCl were separated on a sucrose gradient. The two peaks correspond to the position of phage ghosts (\(\lambda\) particles with no DNA) and knobbed tails (tails dislocated from heads which still have the connector attached), with the activity divided equally between the two positions. These results confirmed that gpFII is a structural protein that is present in the head-tail connector in mature phage. The observation that gpFII that has been isolated from phage particles has the same molecular weight, isoelectric point, and purification properties as gpFII that has not been assembled suggests that no significant covalent modifications take place during viral assembly (Casjens, 1974). E. coli cell lysates produced from \(\lambda\)am lysogens contain normal amounts of gpFII, which cannot be incorporated into the phage particle (Casjens, 1974), suggesting that gpFII binds specifically to the head structure, and interacts directly with gpW. These results confirm that the incorporation of gpFII is the last step in head assembly, and that its incorporation requires a specific, preformed structure. Fewer than ten molecules of gpFII are present in each phage particle, and many have been tempted to speculate that the number is actually six, which matches the symmetry of the connector and the tail (Casjens, 1974).

It has previously been shown that the genetic source of gpFII determines the specificity of the head-tail joining reaction (Casjens et al., 1972). While gpFII and gp8, the bacteriophage 21 homologue, are fully interchangeable (Smith & Feiss, 1993), the \(\phi 80\) gpFII homologue is not (Casjens, 1974). *In vitro*, when purified \(\lambda\) heads (with \(\lambda\) gpFII) are mixed with a 50/50 mixture of \(\lambda\) and \(\phi 80\) tails, both types of tails are able to join to the head and form a mature virion. When \(\phi 80\) gpFII is added to gpFII- \(\lambda\) heads, it is able to bind to the head structure. However, when this intermediate is then mixed with a 50/50 mixture of \(\lambda\) and \(\phi 80\) tails, only \(\phi 80\) tails are able to attach and complete the virion. Because the \(\lambda\) gpFII protein is able to interact with \(\phi 80\) tails,
and the φ80 gpFII is able to interact with the λ head, both the head and tail binding surfaces must be similar for the two proteins.

**DNA Injection**

The bacteriophage λ connector, located at the head-tail junction, is required for egress of the viral DNA from the capsid upon infection of a host cell. The delivery of the bacteriophage λ genome into the host cell takes place in two stages; attachment of the virus to the host cell, and injection of the chromosome into the cell. DNA injection is a complex process that includes triggering of ejection of the chromosome from the capsid, ejection through the tail, and routing of the DNA into the host cell. Successful injection of the DNA requires that the virion be properly assembled. In mature virions, the right end of the DNA molecule extends into the tail for approximately 130 bp (Thomas, 1974). It appears that the placement of the right end of the chromosome into the tail involves specific interactions between the DNA and a tail protein, and that these specific interactions are critical for DNA ejection. This has been demonstrated experimentally by two classes of aberrant λ particles; docL which contain chromosomes that lack the right chromosome end (Sternberg & Weisberg, 1975), and Z- particles, which lack the last protein to act upon the tail (Katsura & Kuhl, 1975). In both of these mutant phage particles the right end of the DNA is not placed in the tail, and both types are ejection-defective (Thomas, 1978).

DNA ejection is triggered by a signal generated by the interaction of the adsorbed phage and the host cell enzyme II component of the mannose transport system (Xu & Feiss, 1991). It has been suggested that protein H* is probably ejected with the DNA and may act as a pilot protein and enter the cell. Protein H* may have a role in the formation of a transmembrane channel, or the association with an existing channel, necessary for DNA transport through the cell membrane. The right end of the chromosome is thought to be ejected through the tail tube first, which is opposite to the direction that it proceeds during packaging (Katsura, 1983). It is unknown which part of the tail the chromosome is ejected from, but it is postulated that it passes through the tip of the tail, possibly through the end of the tail fiber. It is possible that the tail tip undergoes a structural rearrangement before, or during injection.
Atomic Structures of Bacteriophage λ Heads and Tails

Cryo-electron micrographs of the bacteriophage λ head show that has a hexagonally-shaped surface with a diameter of 63 nm. The head contains 405 copies each of the major capsid proteins gpE and gpD arranged on a T=7 laevo lattice. At 34 Å resolution the dominant features of the λ virion are the strongly protruding, thimble-shaped trivalent spikes which consist of trimers of gpD. The gpE hexamers are spaced approximately 13 nm apart and appear as flat, skewed cartwheels with a depression in the middle. The spokes radiating from three adjacent hexamers join underneath the protruding spikes. Each hexamer is thus surrounded by six trimers. The hexamers and pentamers together form a thin network that defines the shell of the capsid. No icosahedrally arranged structures can be seen inside this shell. The three subunits that comprise a thimble are not distinguishable, and form a compact, invariant unit. The fivefold capsomers protrude more from the surface of the virion than the hexamers due to the different geometry of this position. The connections between a pentamer and the neighboring hexamers are essentially the same as in the inter-hexameric connections, except for an additional crossbridge linking the arms of the pentamer and hexamers. If phage are disrupted so they release their DNA, the head becomes a deformed, empty shell, with a stain-excluding knob located at the vertex to which the tail attaches (Harrison et al., 1973). This knob-like structure is the connector, and is assembled as part of the head. When virions are disrupted under certain conditions the connector remains bound to the tail rather than the head shell. This head-tail connector has gpB, pB*, and gpFII activity (Tsui & Hendrix, 1980), and it seems likely that gpW is also present here.

The bacteriophage λ tail is a thin, flexible tube that is 135 nm in length, with a 15 nm long conical part on one end, and a single tail fiber 23 nm in length attached to the conical part at the distal end of the tail. Electron microscopy of phage particles show a tail that is not penetrated by negative stain, while particles lacking DNA show a stained inner lumen, revealing that the tail is a long, cylindrical tube. Treatment of the tail with chemicals causes it to dissociate into ring-like structures that correspond to single- or multi-layered disks. The conical and tubular parts dissociate into 3 (possibly 4) disks and 32 disks, respectively. An end view of the ring-like structure from the tail tube shows that the individual disks consist of an annular ring that is 9 nm in diameter, with a central hole that is 3 nm in diameter. The overall diameter of the rings is approximately 18 nm, due to the presence of six small knobs arranged around the core. Electron microscopy and biochemical data
indicate that the single disks are composed of six subunits of the major tail protein gpV \( \text{II} \) (Buchwald et al., 1970; Casjens & Hendrix, 1974), each consisting of two separate folding domains. Mutants in which approximately 30% of the carboxy terminus of the protein is missing have been isolated. Electron microscopy and hydrodynamic measurements show that these mutant phage lack the outer knobs (Katsura, 1981). Electron microscopy studies of phage resulting from other tail mutants show that \( Z^- \) phage accumulate free heads, inactive tails, and noninfectious phage-like particles (Casjens & Hendrix, 1974; Katsura & Kuhl, 1975). \( U^- \) lysates contain polytails that have random tail tube lengths many times longer than normal (Kemp et al., 1968; Mount et al., 1968). There have been no tail precursors observed by electron microscopy for other \( W^- \) mutants even though biochemical evidence shows that they do accumulate. They have likely escaped detection by electron microscopy due to their small size or possession of shapes that are difficult to recognize as tail-related.

**Atomic Structures of Bacteriophage Head-Tail Connectors**

The portal vertex, or connector, where gpFII is known to assemble (Tsui & Hendrix, 1980), and gpW is postulated to assemble, is the region located between the icosahedral head and the long, cylindrical tail. This is a key viral structure that is involved in a number of steps of the bacteriophage life cycle. Head morphogenesis begins with the assembly of the connector, a ring-shaped structure composed of 12 molecules of gpB, which acts as a nucleation site for polymerization of the other head proteins. Subsequently, in conjunction with terminase, the connector controls the sizing and translocation of the DNA which passes through a hole in the center of the ring. After the DNA has been packaged, the connector undergoes structural changes that allow the tail, which has been formed in a separate pathway, to bind to the head. Finally, when the phage adsorbs to a host cell, the connector must receive a signal for the DNA to exit through the channel connector and infect the host (Katsura, 1983).

Structural information about a number of unrelated bacteriophage connectors, including \( \lambda, \phi29, T3, P22, T4, P2/P4 \) and SPP1 has been obtained from electron microscopy and imaging processing studies giving resolution of 20 to 25 Å. Each of these portal proteins is able to form an oligomeric structure in the absence of head assembly. All of these connectors display the same general morphology with a propeller-like structure composed of a wide domain that is inserted into the prohead vertex, and a narrow region that interacts with the tail. Running through
the center of this structure is a channel through which the DNA is hypothesized to enter into the viral capsid during DNA packaging, and exit during the infection of a host cell. The differences in the various bacteriophage connector structures resides in the number of copies of the homo-oligomeric protein in the connector. The connectors from bacteriophage \( \lambda \) (Kochan et al., 1984), T4 (Driedonsks, 1981), P22 (Bazinet et al., 1988), and P2/P4 show clear 12-fold symmetry. In the case of the SPP1 connector the electron microscopy projection structure reveals a clear 13-fold symmetry (Dube et al., 1993). The symmetry of the phage connectors from \( \phi 29 \) (Carazo et al., 1986), T3 (Valpuesta et al., 1992; Valpuesta et al., 2000), and T7 (Cerritelli & Studier, 1996) are more controversial, with reports of both 12- and 13-fold symmetry. However, X-ray crystallographic data to 3.2 Å has cleared up the \( \phi 29 \) controversy, and shows clear 12-fold symmetry. The discrepancy in the results have been attributed to two causes: the non-controlled overexpression of the recombinant connector protein in bacteria, which could lead to the production of different oligomers, or the result of the image processing methods used, which could prevent the identification of the 13-fold symmetric components.

Since there is no atomic resolution data available for the bacteriophage \( \lambda \) connector, the connector from \( \phi 29 \) must be examined. Although there is no sequence homology between \( \lambda \) and \( \phi 29 \) connectors, proheads of bacteriophage \( \lambda \) which carry the connector of \( \phi 29 \) instead of that of \( \lambda \) can be produced \textit{in vitro} by mixing and incubating extracts containing complementary gene products (Donate et al., 1990). The extracts were prepared from induced lysogens in which the \textit{in vivo} prohead assembly had been blocked by mutations in the prohead genes B, C, Nu3 and E. Sucrose gradient centrifugation and electron micrographs of these hybrid proheads showed that their structure was similar to that of normal \( \lambda \) proheads. These chimeric proheads can package both \( \lambda \) and \( \phi 29 \) DNA, showing that the connector domains involved in both head assembly and DNA packaging are functionally similar. The DNA-containing \( \lambda \)-\( \phi 29 \) proheads can be complemented \textit{in vitro} with \( \phi 29 \) tails to yield infective particles capable of DNA transfer. These results show that in spite of the lack of sequence identity, the \( \phi 29 \) connector can mimic the function of the \( \lambda \) connector during the assembly of \( \lambda \) proheads and establish the proper contacts with the \( \lambda \) head proteins. Additionally, the \( \phi 29 \) connector-containing hybrid proheads can associate with and package both \( \lambda \) and \( \phi 29 \) DNA. Because the \( \lambda \) packaging complex can recognize the p10 domains of the \( \phi 29 \) connector as compatible with the \( \lambda \) connector, it is likely that the \( \phi 29 \) connector protein (gp10) closely resembles gpB from bacteriophage \( \lambda \).
Figure 1-2. Crystal structure of the φ29 connector. (A) View of the connector looking from the top down shows the 12-fold symmetry. The inner channel through which the DNA passes has a minimum diameter of 36 Å. (B) Side view of the connector (rotated 90° about the x-axis) show the three regions of the connector; the wide end, the central region, and the narrow end.
The crystal structure of the \( \Phi 29 \) connector shows that it is composed of three approximately cylindrical regions; the narrow end, the central region, and the wide end (Figure 1-2). These regions have external radii of 33, 47, and 69 Å, and heights of 25, 28, and 22 Å, respectively. The internal channel through which the DNA passes has a diameter ranging from approximately 36 Å at the narrow end, to 60 Å at the wide end. Residues 1-11, 229-246, and 287-309 are not observed in the electron density. The second and third disordered regions are both located on the inside of the channel, and have positively charged sequences, suggesting that they may interact with the DNA backbone. The first 11 residues are also highly basic, which may also facilitate interaction with nucleic acid.

The connector proteins represent a class of structures that are capable of fully surrounding the DNA in a sequence non-specific manner, analogous to the ring-like sliding clamp components of DNA replication machinery. One intriguing aspect of the connector's role in light of its 12-fold symmetry is the ability to attach to one vertex of the head, which itself possesses five-fold rotational symmetry. This symmetry mismatch has been postulated to play an important role in DNA packaging (Hendrix, 1978). The 12-fold symmetry of the connector may be important for its complementarity to the six-fold rotational symmetry of the tail, and should allow a strong interaction between the two.

**Protein-Protein Interactions**

The assembly pathways of bacteriophage provide an interesting system with which to address the factors that govern protein interactions. Protein-protein interactions are biophysical phenomena that occur on the surfaces of proteins. Although the properties that govern the affinity and specificity of these interactions are not fully understood, the rapidly expanding database of structures of proteins and protein complexes has revealed a number of generally applicable factors that govern these reactions. Protein interactions are controlled by the same ubiquitous elements that contribute to the stability and specificity of protein structures; van der Waals forces, electrostatic interactions, hydrogen bonding, and the hydrophobic effect. However, these elements contribute to protein-protein interactions in a very different manner. While protein folding allows the polypeptide chain to sample all configurations in its search for the optimal folded structure, protein interactions do not sample as large a variety. For most reversible interactions, the structures of the proteins do not change significantly when the complex is formed.
Van der Waals interactions are weak and close range, and result in all atoms and molecules being attracted to each other, even in the absence of charged groups, as a result of mutual interactions related to induced polarization effects. These forces also specify the minimum distance that must be present between atoms and molecules due to the repulsion that eventually occurs as they approach one another and their electron orbitals begin to overlap. The volume that specific atoms occupy are modeled as spheres, and are defined as the van der Waals volume. A second type of noncovalent attractions, electrostatic interactions, occur between charged particles. If the charges on two separate atoms are of opposite sign, the energy decreases as they approach each other, and the interaction is favourable. If the charges are of the same sign, there is repulsion between them. The interaction between two oppositely charged groups in a protein is known as a salt bridge, and usually consists of both electrostatic interactions and hydrogen bonding, which occurs when two electronegative atoms compete for the same hydrogen atom. The hydrogen atom is formally bonded covalently to one of the atoms (donor), but also interacts favourably with the other (acceptor). The electrostatic and covalent aspects of the hydrogen bond cause the most common, and presumably most energetically favourable, hydrogen bonds to keep the three bonded atoms collinear. The lengths and strengths of hydrogen bonds depend on the electronegativities of the acceptor and donor; the greater the electronegativities, the shorter the distance between them and the stronger the hydrogen bond. The most commonly observed hydrogen bond in proteins involves the C=O and N-H groups of the polypeptide backbone. The chemical groups in proteins that most commonly serve as hydrogen bond donors are N-H, O-H, and less commonly S-H and C-H groups. The most common acceptors are O=, -O-, -N=, and less frequently -S-, -S-, and the π electrons of aromatic groups. Electrostatic interactions and hydrogen bonding constraints play a critical role in the specificity with which two proteins come together to form a complex, while the thermodynamic energy necessary for complex formation is provided by the hydrophobic effect.

The protein interface that is involved in complex formation is usually composed of a number of closely packed hydrophobic residues, similar to those found in the core of a protein. Recent studies show that the hydrophobic effect, widely accepted as the driving force for protein folding (Dill, 1990), also plays a significant role in protein-protein interactions even though the hydrophobic nature of residues at interaction surfaces is intermediately dispersed between those in the protein core and those at non-interacting surfaces. The large non-polar component
provides hydrophobic free energy in favour of association as it becomes buried in the complexes. The interfaces involved in transient protein-protein interactions are composed differently than those found between subunits of oligomeric proteins such as those found in viral particles. The interfaces found in protein heterocomplexes are half-way between hydrophobic core and surface, while the subunit interfaces in homo-oligomers is much closer in composition to hydrophobic cores (Lo Conte et al., 1999). This difference is likely due to the differences in the roles that these two types of complexes play. As homo-oligomers rarely function as monomers, their hydrophobic surfaces are permanently buried within the protein complex. However, the proteins found in many heterocomplexes may function as monomers in solution, and thus cannot have large exposed hydrophobic surfaces as it would be energetically unfavourable. The amino acid composition of the surfaces of a protein that are involved in interactions differ from those that remain exposed to solvent. A study of 75 protein-protein complexes by Lo Conte et al., (1999) showed that interfaces are much richer in the aromatic residues His, Tyr, Phe, and Trp (21% versus 8%), and somewhat richer in the aliphatic residues Leu, Ile, Val, and Met (17% versus 11%). They are depleted in the charged residues Asp, Glu, and Lys, but not Arg, which is the residue type that makes the largest single overall contribution to interfaces.

Comparison of the structures of both free and complexed proteins which involve standard size interaction interfaces shows that only small changes take place upon association. These include shifts in the surface loops, movements of short segments of the polypeptide chain by up to 1.5 Å, and the rotation of surface side chains, which may move by 5 Å or more (Janin & Chothia, 1990; Davies & Cohen, 1996). The proteins appear to assemble as rigid bodies, with local adjustments of the side chains. In a few cases where there is a much larger protein interface involved in the interaction, there are more substantial changes in conformation of one or both of the components upon complex formation. There are three major types of structural changes observed; large movements of the main-chain as loops switch to quite different conformations; changes in the relative positions of the domains in multi-domain proteins; and disorder-to-order transitions as regions of the polypeptide chain that are disordered in the free protein become ordered in the complex (Lo Conte et al., 1999).

The most striking of these three categories of interactions involves complexes where one or both components undergo major disorder-to-order transitions. The interaction of the serine protease thrombin and hirudin, an anti-clotting agent,
provides an example of this. Hirudin is a 65 residue protein composed of a compact N-terminal domain and a 26-residue C-terminal domain. The NMR solution structure has been determined for two uncomplexed recombinant hirudins proteins (Folkers et al., 1989; Haruyama & Wuthrich, 1989), and both show a stably-folded N-terminal domain and a disordered region encompassing approximately 16 residues at the C-terminus of the protein. There is also a crystal structure of hirudin complexed with thrombin that shows the first 48 residues organized into a compact, globular domain similar to that observed by NMR. However, the C-terminus of hirudin, which is disordered in solution, is found in a well defined conformation in the complex where it extends across the surface of the thrombin molecule for 40 Å (Rydel et al., 1991). This unusually long extended conformation allows hirudin to interact with a large number of residues on the thrombin surface. In all, 12 of the 17 C-terminal residues of hirudin (70%) are involved in interactions with thrombin.

**Thermodynamic Analysis of Proteins**

A multifaceted approach to protein research includes functional, structural and stability studies. The atomic resolution three-dimensional structure of a protein allows the construction of hypotheses about interactions that may be crucial for a protein's stability or function. We can then test these hypotheses by using molecular biology techniques to create mutant proteins with amino acid substitutions at positions we believe to be important. The *in vivo* activity and *in vitro* properties of these mutant proteins can then be assessed using functional assays and biophysical methods such as circular dichroism spectroscopy, tryptophan fluorescence spectroscopy, and nuclear magnetic resonance spectroscopy. In this way, we can determine exactly how particular amino acid substitutions affect the protein's activity and thermodynamic stability. By combining *in vitro* folding and protein mutagenesis studies, the contribution of individual amino acids to stability, structure and activity can be determined. This type of analysis is possible as most substitutions do not affect the two-state behavior of a protein, but result only in local structural rearrangements (Lattman & Rose, 1993).

The thermodynamic parameters that can be derived from thermal and chemical denaturation experiments provide a means to quantitatively compare the folding properties of various mutant proteins as they are precise measures of the thermodynamic stability of a protein. To study protein folding, simple methods for following the reaction are needed. In our laboratory we use both circular dichroism (CD) spectroscopy and fluorescence spectroscopy to assess the conformation of a
protein, and monitor its unfolding. CD spectroscopy can be used to gain structural information about proteins since these molecules exhibit circular dichroism by absorbing light that is circularly polarized in one rotation differently than the opposing rotation, due to fixed asymmetry at chiral centers. In the far U.V. region (below 250 nm) the spectral characteristics of a protein are primarily determined by the conformation of its polypeptide backbone, especially its secondary structure. Most native proteins also display significant optical activity in the near-U.V region (250-300 nm) resulting from the presence of aromatic side chains in asymmetric environments. There are several advantages to using CD as a monitor for protein structure, including the speed of data collection, relatively simple interpretation of the spectra, the ability to collect spectra under a wide variety of conditions, the small amount of sample required, and the ability to recover the sample. The major disadvantage is that the technique gives only a global "average" view of the protein. Although deconvolution of the spectra can approximate the amount of secondary structure present, it cannot be related to the exact structure of the protein.

Fluorescence spectroscopy is also used to characterize proteins in solution by measuring the fluorescence emission of the natural fluorophores tryptophan, tyrosine and phenylalanine after excitation at 280 nm. Because the absorbance wavelengths of the aromatic side chains are in the order of Trp>Tyr>Phe, proteins containing all three amino acids generally display fluorescence spectra characteristic of Trp, those without Trp display spectra characteristic of Tyr, and only in the absence of both Trp and Tyr is Phe fluorescence observed. The fluorescence of these aromatic side chains is highly sensitive to the polarity of the immediate surroundings of these residues, but it varies in an unpredictable manner. Folding of a protein can lead to either an increase or decrease in the quantum yield, and the magnitude of the fluorescence is not very informative in itself. However, any differences in the observed magnitude can serve as a sensitive probe for perturbations in the folded structure of the protein. The fluorescence spectrum of a protein becomes complex when there is greater than one aromatic side chain due to the very efficient energy transfer between them.

Since the unfolding of a small globular protein is a highly cooperative process, its denaturation by heat or chemical denaturant can, in many cases, be modeled as a two-state equilibrium, in which that folded and unfolded forms of the protein interconvert without the formation of populated intermediates. Thus, the unfolding reaction can be stated simply as:
where $F$ is folded protein and $U$ is unfolded protein. The equilibrium constant of unfolding ($K_U$) can then be stated as:

$$\frac{[U]}{[F]}$$

A typical GuHCl denaturation curve plots the protein signal (ellipticity) as a function of the concentration of GuHCl in the sample. As the protein unfolds, the amount of ellipticity observed decreases. The resulting curve consists of a folded and unfolded baseline separated by a sharp unfolding transition. The fraction of unfolded protein present at any point within the unfolding transition can be calculated by extrapolating the folded and unfolded baselines into the transition region. The fraction of unfolded protein at any given GuHCl concentration can then be calculated by

$$\text{fraction unfolded} = \frac{y - \text{folded baseline}}{\text{unfolded baseline} - \text{folded baseline}}$$

where $y$ is the ellipticity at any given concentration of GuHCl, and

$$\text{fraction unfolded} = \frac{[U]}{[F] + [U]}$$

By substituting in the equation of $K_U$, the fraction unfolded can also be expressed as:

$$\text{fraction unfolded} = \frac{K_U}{K_U + 1}$$

so that $K_U$ can be calculated as each point in the transition region. Once the value of $K_U$ at each point is known, $\Delta G_U$ at each point can be calculated from the equation

$$\Delta G_U = -RT\ln K_U$$

The $\Delta G_U$ values calculated at numerous GuHCl concentrations are plotted, and the data points are fit by linear regression. Since $\Delta G_U$ decreases linearly with increasing concentrations of GuHCl, the $\Delta G_U$ at 0 M GuHCl can be determined by simple linear extrapolation. The slope of this line, $m$, represents the GuHCl concentration dependence of $\Delta G_U$. In general, for an average protein domain, the difference in the
free energy states of folded and unfolded proteins is relatively small, with typical 
$\Delta G_U$ values ranging from 5 to 15 kcal/mol (Pace, 1990).

This type of analysis can be used with thermal denaturation curves as well, 
but with slightly more complex equations as the relationship between $\Delta G_U$ and 
temperature is not linear. This non-linearity is a direct result of the differences in 
heat capacity for the folded and unfolded states. Analysis of thermal denaturation 
curves must take into account the change in heat capacity between the folded and 
unfolded states ($\Delta G_U$) and the heat required to actually unfold the protein ($\Delta H_U$). 
The following equation relates these quantities:

$$\Delta G_U (T) = \Delta H_{Tm} \left(1 - \frac{T}{T_m}\right) - \Delta C_p \left[(T_m - T) + T \ln \left(\frac{T}{T_m}\right)\right]$$

where $\Delta G_U (T)$ is the $\Delta G_U$ at a given temperature ($T$), $T_m$ is the temperature at the 
midpoint of the unfolding transition (i.e. where $\Delta G_U = 0$), and $\Delta H_{Tm}$ is the change 
in enthalpy of unfolding at the $T_m$.

Solving Protein Structures with NMR

Nuclear magnetic resonance and X-ray crystallography are currently the only 
techniques capable of determining the structures of biological macromolecules like 
proteins and nucleic acids at atomic resolution. Although the resolutions of NMR 
solution structures are not as high as well-determined crystal structures, NMR 
allows the study of time dependent phenomena, such as intramolecular dynamics 
in macromolecules, reaction kinetics, molecular recognition or protein folding 
while crystallography provides only a static image. Historically, NMR of 
macromolecules was limited by both the low inherent sensitivity of the technique, 
and the complexity of the spectra. These limitations have been partially alleviated by 
the development of stronger magnets, more sensitive spectrometers, and advances 
in labeling of proteins using NMR active isotopes.

The first step in the analysis of a molecule by NMR is the assignment of a 
unique chemical shift to each of the NMR active nuclei. The nuclei of all elements 
carry a charge, and when the spins of the protons and neutrons comprising a 
nucleus are not paired, the overall spin of the nucleus generates a magnetic dipole 
along the spin axis. All nuclei with unpaired protons have spin angular 
momentum (I), which is a quantum mechanical property of matter. Nuclei with an 
angular momentum quantum number of zero (e.g. $^{12}$C) are not observable by NMR, 
but can be replaced by $^{13}$C nuclei, which have I=1/2. Nuclei with I=1, such as $^{14}$N, are
also generally not useful for NMR, as they have quadrupole moments, which shortens the lifetime of their magnetic states, and result in broad resonance lines that are difficult to study. Thus, for structural studies using NMR the $^{14}\text{N}$ nuclei are replaced with $^{15}\text{N}$ nuclei in which $I=1/2$. NMR manipulates the magnetic field of the nuclei by placing the protein in a static magnetic field that is applied along the z-axis. The nuclei inhabit one of two energy levels (Zeeman levels), determined by the allowed quantum states of spin $1/2$ nuclei ($m=\pm 1/2$), with the majority of the nuclei aligning themselves with the magnetic field. In Fourier Transform pulse NMR one sends a radio frequency pulse that simultaneously excites all nuclei of a given type ($^1\text{H}$ for example) and the absorbed energy is detected as an induced current in the detector. The protons then absorb at different frequencies depending on the local variations in the field. Because of the geometry of the detector (the probe head) the resonance frequencies are detected as the sum of the sinusoidal signals whose frequency corresponds to the energy absorbed and whose decay rate is a function of the relaxation rate of each signal back to the equilibrium state. Fourier transformation of this time domain free induction decay (FID) signal yields the NMR spectrum.

<table>
<thead>
<tr>
<th>Nucleus</th>
<th>$I$</th>
<th>$\gamma(Ts)\text{-1}$</th>
<th>Natural Abundance $\nu$ at 11.74T</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^1\text{H}$</td>
<td>1/2</td>
<td>$2.675 \times 10^8$</td>
<td>99.98</td>
</tr>
<tr>
<td>$^2\text{H}$</td>
<td>1</td>
<td>$4.107 \times 10^7$</td>
<td>0.02</td>
</tr>
<tr>
<td>$^{13}\text{C}$</td>
<td>1/2</td>
<td>$6.728 \times 10^7$</td>
<td>1.11</td>
</tr>
<tr>
<td>$^{15}\text{N}$</td>
<td>1/2</td>
<td>$-2.712 \times 10^7$</td>
<td>0.36</td>
</tr>
</tbody>
</table>

NMR instruments excite protons and make them precess in phase using an oscillating magnetic field, without which the protons would lose both phase coherence and excitation energy. Two relaxation processes are involved in these losses, spin-lattice and spin-spin relaxation. In spin-lattice relaxation, a proton in the excited state interacts with fluctuating magnetic fields generated by other atoms in the molecule (the lattice). This is an enthalpic relaxation, because it generates heat.
while bringing the proton to the ground state. It is this relaxation that maintains the excess of protons in the ground state. The second process, spin-spin relaxation, involves dipole-dipole interactions between excited state and ground state spins for the same type of proton. This interaction leads to an exchange of energy making the system more random (entropic relaxation), and eroding the phase coherence. However, entropic spin-spin relaxation does not change the net population of the excited state. The relaxation time $T_2$ is the time for $1/e$ of the protons to relax via this process.

Two types of NMR interactions, through-bond spin-spin effects ($J$ couplings) and through-space nuclear Overhauser effects (NOEs), are used to determine the three-dimensional structure of a protein. Spin-spin interactions are responsible for splitting the absorption of energy by a nucleus. The magnetic dipoles of one nucleus affect the magnetic dipoles of another nucleus through the electrons of the bonds between them. This splitting of the absorption is often referred to as a through-bond interaction, and is dependent only on the nature of the bond between the two nuclei. The spin-spin $J$-coupling constant, which measures the splitting between two different nuclei in a protein in Hz, is dependent on the relative orientation of the two nuclei. The splittings from spin-spin interactions can be removed with double resonance techniques, whereby the excitation of one group of nuclei (for example protons) with one source of electromagnetic radiation removes the spin-spin interaction with a second group of protons (decoupling), which allows the resonances of these protons to be viewed with a second source. This permits the determination of peaks from protons that are close enough together to interact with another proton through a bond. NOEs are a result of dipole-dipole couplings between different nuclear spins, which causes spin polarization to be transferred from one residue to any nearby nucleus. The atoms do not have to be in the same residue, they must simply be near one another in space. Since this dipole-dipole interaction is through space, it is sensitive to both nuclei that are close together because of bonding structure, and those that are within approximately 5Å as a result of the secondary and tertiary structure of the protein.

The usefulness of NMR comes from the fact that given a certain strength of externally applied magnetic field, each unique type of nucleus will absorb a particular frequency of electromagnetic wave. Protons resonate at ~500 MHz in an 11.74 tesla magnetic field, whereas nitrogens resonate at ~50 MHz, and carbons at ~125 MHz (Table 1-1). The resonance frequency of a nucleus ($^1H$, $^{13}C$, or $^{15}N$) is referred to as its chemical shift and is given in units of parts per million (ppm).
relative to the resonance signal from a reference compound with very shielded protons that resonate at a particularly high frequency (usually DSS for proteins). The chemical shift of a nucleus is modulated by the environment surrounding it. For example, the protons found in amide groups along the backbone of a protein have a characteristic range of resonance frequencies which is different from the range for protons attached to sidechain methyl groups. In addition, the chemical shift is modulated by all of the neighboring atoms in the protein through spin dipolar interaction and thus is sensitive to secondary and tertiary structure.

Thus, protein structure determination begins with identifying which frequency corresponds to each NMR active nucleus in the protein. Once this assignment process is complete, cross peaks observed in three-dimensional spectra must be assigned and interpreted geometrically as constraints on interatomic distances. Information about the torsion angles about single bonds provided by J-coupling experiments and chemical shift values which help predict the position of secondary structure elements are combined with these constraints and are used as input for distance geometry calculations which provide molecular models that are consistent with the constraints. Finally, these models are then refined using energy minimization programs to eliminate structural irregularities while maintaining consistency with the experimental constraints.

Triple resonance experiments performed using double-labeled ($15N$, $13C$) proteins are the choice method for the sequential assignment of proteins. The use of triple resonance experiments greatly simplifies the assignment problem due to a marked reduction in spectral overlap and ability to perform scalar type experiments with magnetization transferred through the peptide bond. A number of experiments that take advantage of through bond scalar couplings between $^1H$, $^{15}N$, and $^{13}C$ atoms linked by covalent bonds are utilized. The general assignment scheme used for the protein structures solved in this work is as follows. First, the observed HN, Cα, and Cβ resonances are placed in sequential order using the HNCACB and HN(CO)CACB spectra, a pair of experiments which correlate resonances from neighboring amino acids. The HNCACB spectrum has frequency axes for $^1H$, $^{15}N$, and $^{13}C$. It correlates the amide proton with the frequencies of the Cα and Cβ atoms for itself and the residue preceding it in the primary sequence of the protein. The HN(CO)CACB experiment allows the discrimination of the intra- and inter-residue cross peaks. The magnetization in this experiment is transferred via the CO atom, thus allowing only the inter-residue cross signal to be observed. When these two spectra are examined side by side, it is possible to discriminate
between the intra- and inter-residue peaks for the Cα and Cβ atoms. Additional information is available, as certain residues such as glycine, alanine, serine, and threonine are readily identifiable from the frequencies of their Cα and Cβ atoms. These residues are then used to place the sequentially ordered strips in register with the known amino acid sequence of the protein. Once the backbone HN, Cα, and Cβ resonances have been assigned, the complete side chain resonances are determined using the CCC-TOCSY (carbon resonances), HCCH-TOCSY (proton resonances), and TOCSY-HSQC. The assignment of side chain resonances helps to confirm the primary assignment as the TOCSY experiments identify individual spin systems, and allows the correct type of amino acid to be determined from the peak pattern of each spin system.

Once the unique chemical shifts for each residue have been assigned, as many cross peaks as possible in the NOESY spectra are assigned to pairs of resonances. These peaks provide the most important structural information in the form of the distance information that is inherent in the NOESY cross peaks. While the volumes of the cross peaks in 3D-NOESY spectra are, in principal, proportional to the inverse sixth power of their internuclear distance \( r_{ij}^{-6} \) (where \( i \) and \( j \) are two protons) they are actually influenced by local dynamics and the averaging of the ensemble of different conformations undergoing rapid exchange in solution. Thus, a conservative approach has been developed whereby the volumes of the cross peaks are interpreted in terms of upper and lower bounds on the interatomic distances, and the peaks are classified as strong, medium, or weak. These distance boundaries are then input into distance geometry computer programs and are used to compute conformational ensembles that satisfy all of the input restraints.

**Thesis Overview**

My research objective was to further understand how large multi-subunit structures assemble using the head-tail joining reaction of the bacteriophage \( \lambda \) virion as a paradigm. To improve our understanding of how bacteriophage \( \lambda \) performs the complex functions that are needed for the head-tail joining reaction, I characterized the proteins gpW and gpFII using a combination of biochemical and biophysical techniques. I chose the "divide and conquer" approach whereby these two individual components of the virion were characterized biochemically, as well as at atomic resolution using NMR spectroscopy. I have demonstrated that gpW can be tagged at both the N-and C-termini, and that these tagged proteins display reversible two-state unfolding behaviour that is indistinguishable from the
untagged protein. The N-terminal affinity tag has no effect on the biological activity of gpW as measured in an in vitro assembly assay, while the C-terminal tag decreases, but does not abolish activity. A number of site-directed mutants in gpW were purified and characterized thermodynamically and biochemically. The results suggest that the C-terminal region of gpW is not associated with the folded part of the protein, but may be making crucial interactions with other proteins in the viral particle upon assembly. To gain more insight into the structure-function relationships of gpW, I determined its three-dimensional structure using NMR spectroscopy. gpW possesses a novel fold comprising two α-helices packed against a single two-stranded β-sheet. As expected from the mutagenesis studies, the 14 C-terminal residues of gpW were found to be unstructured. We speculate that this region may become structured when gpW is incorporated into the viral particle. Structural similarity of gpW with the DNA processivity factor from DNA polymerases allowed me to model gpW into a hexameric ring that is consistent with its known roles in λ morphogenesis. Finally, I characterized gpFII biochemically and structurally, and found that it can be hexahistidine-tagged on its N-terminus with no effect on its in vitro biological activity. The gpFII structure, which I solved using NMR spectroscopy, comprises 7 β-strands arranged in a twisted β-sandwich fold formed by two orthogonally packed antiparallel β-sheets. A remarkable feature of the structure is the presence of two large disordered regions, which we postulate may play a role in assembly and become ordered during this process. My results suggest that the two head-tail joining proteins from bacteriophage λ, gpW and gpFII, may share a common mechanism, employing a disorder-to-order transition, to control protein assembly in the virion.
References


CHAPTER 2

Thermodynamic and Functional Characterization of Protein W from Bacteriophage λ: The Three C-terminal Residues are Critical for Activity

A version of this chapter has been published as:

I performed all of the experiments in this chapter.
Introduction

In this chapter I describe the initial functional and thermodynamic characterization of gpW. The bacteriophage λ gpW is a 68 residue protein that is required for the final steps in λ morphogenesis, including head-tail attachment. Phage lacking gpW produce heads of the same diameter as wild type λ, DNA molecules of monomer length with cohesive ends, and active tails that are capable of combining with heads from a tail minus infection.

For the detailed analysis of the in vitro activities and biophysical properties of gpW, it was necessary to isolate significant amounts of this protein in a purified form. Untagged gpW had previously been purified in our laboratory from an overexpression vector, but the protocol was labour intensive (unpublished results). As we wished to characterize a number of mutant forms of gpW to elucidate its structure and activities, a faster, more efficient protocol was required. This chapter describes the construction of vectors that produce gpW fused to a polyhistidine sequence to allow rapid purification with Ni²⁺-affinity chromatography. We show here that large amounts of active gpW can be easily purified using N- or C-terminal hexahistidine tags and that these tagged proteins display reversible two-state unfolding behaviour indistinguishable from the untagged protein. It also describes the purification and detailed biophysical characterization of the wild type protein, as well as mutants at seven different positions in the protein.

Materials and Methods

Plasmid Construction

Gene W was amplified by PCR from bacteriophage λ DNA. Residues 1-68 were cloned into the expression vector pET21d (Novagen), which contains a T7 RNA polymerase promoter, in such a way that the second residue was changed from Thr to Val, and sixteen extra residues encoding the FLAG epitope (Hopp et al., 1988) and hexahistidine sequence used for protein purification (RLDYKDDDDK- HIS₆) were fused to the C-terminus. Changing the second residue from Thr to Val was considered a conservative mutation as Val is observed at this position in gene 3 from the closely related bacteriophage 21 (Siegele et al., 1983). An N-terminal tagged fusion construct was also created which has a native C-terminus, and twenty-five additional residues (HIS₆-DYDIPITENLYFQ*G) fused to the amino terminus that
encode the hexahistidine tag used for purification of the protein, a protease cleavage site, and a seven residue linker (Gibco-BRL). The protease that cleaves this tag is isolated from the tobacco etch virus (TEV) and specifically cleaves the sequence Glu-Asn-Leu-Tyr-Phe-Gln*Gly, leaving a single Gly fused to the N-terminus. This vector also has the T2V substitution. Mutations were introduced into the N-terminal tagged vector by PCR-mediated site-directed mutagenesis using Vent® DNA polymerase (NEB). The DNA sequences of the wild type and mutant constructs used in this study were verified using the Sequenase® Version 2.0 kit (United States Biochemical).

Protein Expression and Purification

All proteins were expressed in the E. coli strain KM1502, a derivative of GJ1158 (Bhandari & Gowrishankar, 1997), the genotype of which is ompT hsdS gal dcm ΔmalAp510 malP::(proUp-T7 RNAP) malQ::lacZhyb11 Δ(zhf-900::Tn10dTet) ΔslyD (kan'). This strain contains the gene for T7 RNA polymerase under the control of a salt-inducible promoter, and the slyD deletion which prevents the expression of a 21 kD histidine-rich E. coli protein that binds strongly to the Ni-affinity resin used in this purification. Protein expression was induced by the addition of 200 mM NaCl to a culture with an optical density at 600 nm of approximately 1.0, followed by incubation at 37 °C for an additional 3 hours. Unstable gpW mutants that were quickly degraded in KM1502 were expressed in the E. coli strain SG1146A (Gottesman et al., 1998), and induced with 200 μg/ml IPTG for three hours at 37 °C. The cells were harvested and lysed in 6 M GuHCl, 100 mM NaH₂PO₄, 10 mM TrisHCl, 10 mM imidazole, pH 8.0, and purified in the same buffer via batch method using NiNTA-agarose resin (Qiagen). The pure proteins were eluted with 6 M GuHCl, 0.2 M acetic acid, and were refolded by dialysis into 10 mM TrisHCl, pH 8.0, 0.2 mM EDTA, 200 mM NaCl. All other experiments described in this work were performed in this buffer. After dialysis of the N-terminal tagged protein, TEV cleavage was performed according to the manufacturer's directions (Gibco-BRL), and the proteins were loaded onto a second NiNTA column in the TEV buffer, and were shaken at 4 °C for one hour. The TEV-protease is 6-His tagged, so the NiNTA resin binds the TEV protease, any uncleaved gpW(NT), and the cleaved 6-His tag. The unbound protein was dialyzed against 10 mM TrisHCl, pH 8.0, 0.2 mM EDTA and was loaded onto an FPLC MonoS column (Pharmacia). A gradient from 0 to 1 M NaCl was run, with the major protein peak eluting at 0.4 M NaCl. All proteins were determined to be greater than 98% pure by SDS-PAGE electrophoresis followed by
Coomassie staining, and their concentrations were determined by UV absorbance at 280 nm using a molar extinction coefficient of 4500 M⁻¹cm⁻¹ for gpW(CT), 9000 M⁻¹cm⁻¹ for gpW(NT), and 3000 M⁻¹cm⁻¹ for gpW(+G). Extinction coefficients were calculated from the number of tyrosine residues present in each construct (Pace et al., 1995). Crude extracts of gpW(NT) containing gP were prepared by sonication of cells resuspended in our standard buffer. The extracts were centrifuged at 15,000 rpm for 20 minutes to remove insoluble protein.

**Determination of Thermodynamic Parameters by CD Spectroscopy**

Thermal and chemical denaturation experiments were performed in an Aviv 62A DS circular dichroism spectrometer. Fractional helicity of gpW(+G) in buffer was calculated by the formula \( ([\theta]_{222} + 2340)/30,300 \) (Chen et al., 1972). Thermal denaturation was monitored by measuring the CD signal at 222 nm in a 0.1 cm cuvette on samples with protein concentrations ranging from 10 μM to 200 μM. The proteins were heated from 25 °C to 109 °C in 2 °C increments, with a 1 minute equilibration time and a 15 second averaging time for the CD measurement. Near UV denaturation experiments were performed in a 1 cm cuvette, monitoring the change in signal at 280 nm. The equilibration time was increased to 2 minutes, and the other parameters were unchanged. Urea denaturation experiments were carried out in a 1 cm cuvette using a Microlab 500 series automated titrator, and the software program Igor Pro™. A 4 μM protein solution in buffer was mixed in step-wise fashion into a 4 μM protein solution in a high concentration of urea (8.5 to 9.0 M) at 25 °C. After each injection, the sample was mixed for 60 seconds, and the CD measurement at 225 nm was averaged over 30 seconds. Both thermal and urea denaturation curves were completely reversible under all conditions. These data were fit to the standard thermodynamic equations for a monomeric protein. A minimum of 5 thermal and 3 urea denaturation experiments were performed on each of the wild type constructs. The average error associated with these experiments was 0.7 °C and 0.24 kcal/mol respectively.

The change in heat capacity upon unfolding (ΔC_p) of gpW(NT) was determined according to the method of Pace and Laurents (1989). Averaged ΔG_u values determined from the thermal unfolding transition zone of two temperature melts, as well 10 ΔG_u values determined by performing urea denaturation experiments at temperatures between 15 °C and 35 °C were used to determine ΔG_u values over a range of temperatures. A plot of ΔG_u versus temperature was fit using the following equation:
\[ \Delta G(T) = \Delta H_m (1 - T/T_m) - \Delta C_p [(T_m - T) + T \ln (T/T_m)] \] (1)

where \( \Delta G(T) \) is the \( \Delta G \) at a temperature \( T \), \( T_m \) is the midpoint of the thermal denaturation curve, \( \Delta H_m \) is the unfolding enthalpy measured at the \( T_m \), and \( \Delta C_p \) is the change in heat capacity upon unfolding (Pace & Laurents, 1989). Fitting was performed using Kaleidograph with \( \Delta C_p \) as the only free parameter. The \( T_m \) and \( \Delta H_m \) values used were 72.4 °C and 36.4 kcal/mol, respectively, as averaged from the manual fitting of two temperature induced unfolding curves of gpW(NT) by the method of Breslauer (1995).

**Molecular Weight Determination**

Sedimentation equilibrium centrifugation experiments were performed at 20 °C in a Beckman model XL-I analytical ultracentrifuge equipped with UV-visible absorbance optics. Samples of gpW(CT) at concentrations of 40, 100, and 300 μM in buffer (10 mM Tris, pH 8.0, 0.2 mM EDTA, 200 mM NaCl) were centrifuged at 25,000, 29,000 and 30,000 rpm until equilibrium was reached at each speed, and the absorption profiles at the appropriate wavelength was then recorded. The oligomerization state of gpW(NT) was also examined by equilibrium centrifugation. This protein was examined at a single concentration, 20 μM, as at lower concentrations the scatter in the measurement was too high, and at higher concentrations the protein began to aggregate in the cuvette. Data was collected at 25,000, 30,000 and 35,000 rpm. The molecular weights of the proteins were calculated by fitting the resulting data sets using non-linear least squares regression (using the program Kaleidagraph) to equation 2:

\[ A = A(0) \exp \left[ \omega^2 M_r (1 - \nu \rho) (r^2 - r(0)^2) / 2RT \right] \] (2)

where \( A \) is the absorbance at 280 nm at radius \( r \), \( A(0) \) is the absorbance at 280 nm at reference radius \( r(0) \), \( R \) is the gas constant, \( T \) is the temperature, \( \rho \) is the buffer density, \( \nu \) is the partial specific volume of the protein calculated from the amino acid composition, \( \omega \) is the angular velocity, and \( M_r \) is the molecular weight of the protein. The partial specific volumes as calculated from the amino acid compositions of gpW(CT) and gpW(NT) were 0.7030 cm³g⁻² and 0.7115 cm³g⁻², respectively, and the solution density was measured to be 1.011 g/ml.

Gel filtration chromatography was performed at room temperature by FPLC on a 25 ml Superdex-75 column (Pharmacia) calibrated with BSA (67 kDa), ovalbumin.
(44 kDa), carbonic anhydrase (29 kDa), myoglobin (17 kDa), lysozyme (14 kDa), and insulin (3.5 kDa). The column was washed with 2 volumes of buffer with a 1 ml/min flow rate. A 2 mM sample of gpW(CT) was loaded on the column and eluted at 1 ml/min, using absorbance at 280 nm to detect protein.

In Vitro Assay of gpW Activity

Virion extracts lacking gpW were prepared from the lysogenic E. coli strain 594(λWam403cI857Sam7). 50 ml of LB was inoculated with 1 ml of overnight culture and the cells were grown to an optical density at 600 nm of approximately 0.4, heat induced at 45 °C for 15 minutes, and then grown one hour at 37 °C. At the end of the induction period the cells were collected by centrifugation, and resuspended in 1 ml of RRM buffer (16.8 mM NH₄Cl, 0.9 mM MgSO₄, 18 mM KCl, 2.7 μM FeCl₃, 44.5 mM Na₂HPO₄, 19.8 mM KH₂PO₄, 0.25 mM CaCl₂, 0.36% D-maltose, 10% glycerol, 0.1% β-mercaptoethanol, 10 mM putrescine) and 150 μL chloroform to induce lysis of the cells. Cellular debris was collected by centrifuging the tubes for 5 minutes at 15,000 rpm at 4 °C, and the upper aqueous phase was collected and stored at −80 °C. Equal volumes of acceptor extract and purified gpW at known concentrations were mixed in an Eppendorf tube and incubated at 37 °C for one hour, then placed on ice, and serial dilutions of the reaction were titered for plaque-forming units on the indicator strain QD5003 (Yanofsky & Ito, 1966). The WT protein activity assays were repeated at least 5 times, and the results averaged. Each mutant assay was repeated at least twice, and, as the number of pfu/ml is dependent on the preparation and age of the extract used, wild type protein was assayed each time as a control. The wild type activity was then used to normalize the activities of the mutant gpW proteins. Repetitions of the activity assays for the mutant gpW proteins resulted in an average error of ± 30%.

Results

Secondary Structure Prediction and Analysis

The sequence of gpW is shown in Figure 2-1A, aligned with homologous proteins from phage 21 (gene 3) and bacteriophage N15 (gene 3) which share sequence identity with gpW of 48% and 85%, respectively. This three sequence alignment was subjected to the PHD algorithm (Rost & Sander, 1994), which uses multiple-sequence alignments to predict secondary structure. The 68 residue protein
was predicted to have two \( \alpha \)-helical regions, spanning residues 3-18 and 40-53, and two \( \beta \)-strands from residues 22-27 and residues 32-36.

**Functional Characterization of His-tagged gpW Constructs**

We constructed a vector to produce gpW fused to a polyhistidine sequence to allow rapid purification with Ni-affinity chromatography. Using this construct, gpW was purified with an N-terminal 6-His tag that was subsequently removed by cleavage with tobacco etch virus (TEV) protease, leaving only one extra Gly residue at the N-terminus of the protein (Figure 2-1B). A purified preparation of this protein, which was named gpW(+G), displayed a high level of activity in an *in vitro* assay of gpW function. Since its level of activity was similar to that observed with untagged wild-type gpW (H. Murialdo, personal communication), we concluded that this purification strategy would provide an effective means to study gpW structure and function. Surprisingly, the uncleaved 6-His tagged construct, gpW(NT), displayed equivalent activity to gpW(+G), even though it possesses 25 extra residues fused to its amino-terminus (Table 2-1). Since gpW(NT) could be produced even more easily than gpW(+G), further functional studies were performed with this protein.

**Table 2-1**

*Characterization of the three WT gpW constructs.*

<table>
<thead>
<tr>
<th>Construct</th>
<th>( T_m (^\circ C) )</th>
<th>( \Delta G_U ) (kcal/mol)</th>
<th>( m ) (kcal mol(^{-1}) M(^{-1}))</th>
<th><em>In vitro activity (pfu/ml)</em></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Protein concentration</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5 ( \mu M )</td>
</tr>
<tr>
<td>gpW(CT)</td>
<td>71.6 ± 0.4(^a)</td>
<td>2.91 ± 0.11(^c)</td>
<td>647 ± 25</td>
<td>8.5 × 10(^3)</td>
</tr>
<tr>
<td>gpW(NT)</td>
<td>71.8 ± 0.7(^b)</td>
<td>3.27 ± 0.00(^c)</td>
<td>687 ± 17</td>
<td>5.0 × 10(^8)</td>
</tr>
<tr>
<td>gpW(+G)</td>
<td>70.8 ± 0.7(^b)</td>
<td>2.77 ± 0.11(^c)</td>
<td>648 ± 42</td>
<td>5.0 × 10(^8)</td>
</tr>
<tr>
<td>Average</td>
<td>71.5 ± 0.7</td>
<td>3.01 ± 0.24</td>
<td>662 ± 30</td>
<td>5.0 × 10(^8)</td>
</tr>
</tbody>
</table>

\(^a\) averaged from seven experiments performed at protein concentrations of 10 to 300 \( \mu M \)

\(^b\) averaged from 5 experiments at 10 to 50 \( \mu M \)

\(^c\) averaged from 3 experiments at 10 to 50 \( \mu M \)

\(^d\) averaged from the N-terminal tagged constructs only
Figure 2-1. Sequence alignment of gpW and protein constructs. (A) Alignment of the amino acid sequences of gpW (accession P03727) and homologues from phage 21 (gp3; accession P36271) and N15 (gp3; accession AAC19039). The positions of the two α-helices and two β-strands predicted by PHD (Rost & Sander, 1994) are indicated. The putative hydrophobic core positions identified by the heptad repeat in the second helix that were substituted to reduce the stability of the protein are boxed, and position 52, which was predicted to be a non-core residue is underlined. In addition, the three C-terminal residues that were studied in this work are shaded. (B) Representation of the fusion proteins of wild type gpW used in this study. The complete amino acid sequence of the 68 residue protein is fused to a histidine affinity tag on the C-terminus (gpW(CT)), or the N-terminus (gpW(NT)). After cleavage of the N-terminal tag, a single Gly remains fused (gpW(+G)).
The reaction kinetics and concentration dependence of the gpW-mediated in vitro reaction were studied in more detail. Figure 2-2A shows that the maximal yield of phage is observed within 12 minutes of incubation of purified gpW with the extract, while an increase of four orders of magnitude is observed in only 2 minutes. In Figure 2-2B it can be seen that as the concentration of gpW increases, the number of plaque forming units (pfu) produced increases in a logarithmic fashion, reaching a plateau at a concentration of 10 μM protein. To ensure that the concentration of gpW was truly limiting when gpW was present at low concentrations, and that the results observed were not due to the phage components becoming inactivated before the reaction had time to proceed to completion, we added more phage extract to reaction mixtures after they had been incubated for 1 hour (data not shown). When extra phage extract was added to reactions in which gpW was under limiting conditions (i.e. < 10 μM gpW), no increase in the number of phage particles produced was observed. Conversely, when extra phage extract was added to the reactions in which gpW was in excess (> 10 μM gpW), there was an increase in the number of phage produced.

To ensure that the protein purification performed under denaturing conditions was not inactivating a large percentage of the molecules in our gpW preparations, we performed packaging assays on crude E. coli extracts containing gpW(NT). gpW(NT) was found predominantly in the soluble fraction of these extracts. E. coli cells expressing gpW(NT) were diluted and assayed for in vitro activity. The intensity of the Coomassie blue stained band corresponding to gpW(NT) in the extract, as visualized by SDS-PAGE electrophoresis, was compared to known concentrations of purified gpW(NT) to estimate the concentration. As shown in Figure 2-2B, the level of activity in the crude gpW(NT) extract and its concentration dependence was identical to that of the purified gpW(NT), demonstrating that the purification protocol is not inactivating the protein.

To assess whether the C-terminus of gpW could also be tagged, we created a second construct, gpW(CT), which encodes full-length gpW fused to the FLAG epitope and six histidines (Figure 1B). This tag had previously been found to affect neither the structure or function of an another protein studied in our laboratory (Maxwell & Davidson, 1998). However, the in vitro activity of gpW(CT) was decreased by 80-fold compared to gpW(NT) (Table 2-1) and its activity reached a plateau at 1.0 x 10^7 pfu/ml even at protein concentrations as high as 170 μM. Although gpW(CT) is not as active as the N-terminal tagged constructs, it is soluble
at protein concentrations as high as 2 mM, while gpW(+G) and gpW(NT) are soluble only to concentrations of approximately 40 μM.

**Figure 2-2.** In vitro activity of gpW(NT) as a function of time and protein concentration. Activity was monitored as a function of the number of pfu/ml formed by the addition of purified gpW to a phage extract prepared from a Wam lysogenic strain. (A) 20 μM gpW(NT) was mixed with a W- extract and incubated at 37 °C. Aliquots were removed at various intervals, and plated on QD5003 cells. The activity leveled at approximately 15 minutes, and did not increase further by 60 minutes. (B) gpW(NT) (○) was added to the phage extracts at varying concentrations between 0.125 μM and 28 μM, and the reaction was allowed to proceed for 60 minutes before plating on QD5003 cells. The soluble fraction of crude E. coli extract containing induced gpW(NT) (□) was also tested for activity in the in vitro assay. The concentration of gpW(NT) in the extract was estimated by visualizing the band corresponding to gpW(NT) on an SDS-PAGE gel stained with Coomassie Blue.

**Determination of the Native Molecular Weight of gpW**

The high order dependence of the gpW mediated assembly reaction on protein concentration (Figure 2-2B) implies that multiple copies of gpW are required for the production of one phage particle. For this reason, we hypothesized that gpW may be required to oligomerize for its assembly into phage, as do many structural proteins in viruses. To address this issue, analytical ultracentrifugation experiments were used to determine whether gpW exists as a multimer in solution. Both gpW(CT) and gpW(NT) were used for these experiments. Although gpW(CT) is somewhat less active than the N-terminally tagged gpW constructs, it possesses high solubility allowing native molecular weight to be determined at a wide range of protein concentrations. Sedimentation equilibrium experiments were performed at protein concentrations of 40, 100, and 300 μM (Figure 2-3A). When the data from
these experiments were fit as a single species, a molecular weight of 9,235 ± 299 was calculated. The actual molecular weight of the monomer calculated from its amino acid sequence is 9,436, a deviation of only 2%. These experiments demonstrate that gpW(CT) remains monomeric even at concentrations 20-fold higher than that used in the in vitro gpW assay. Furthermore, gel filtration experiments confirmed that gpW(CT) exists in solution as a monomer at concentrations as high as 2 mM (data not shown). Since gpW(CT) is not fully active, we also determined the native molecular weight of gpW(NT). These experiments could be performed only at low protein concentration (20 μM) due to the low solubility of this construct. When the data were fit as a single species, a molecular weight of 10,738 ± 132 was calculated (Figure 2-3B). The expected molecular weight of a gpW(NT) monomer is 10,707, showing that gpW(NT) is monomeric at 20 μM, which is higher than the concentration required for maximal specific activity in the in vivo assay. These results imply that gpW activity does not require prior oligomerization of the protein.

Figure 2-3. Analytical ultracentrifugation of gpW(NT). Sedimentation equilibrium ultracentrifugation data are shown for purified gpW(CT) at a protein concentration of 300 mM (A), and gpW(NT) at a protein concentration of 20 μM (B). Data points were collected at 25,000 rpm (20 °C) by measuring absorbance at 250 nm and 280 nm, respectively. The lines joining the points in the graphs are the theoretical fits to the data using a one-species function. Residual deviations from the theoretical fits for these data sets are shown in the upper panels. The residuals in panel B appear less scattered because the OD of the sample is much lower.
Thermodynamic Characterization of His-tagged gpW Constructs

To determine whether the N- or C-terminal tags had any effect on the structure or stability of gpW, circular dichroism (CD) studies were undertaken. The far UV folded spectrum of gpW(+G) is typical of a helical protein (Figure 2-4A), with minima observed at 222 and 208 nm. Identical spectra were observed for gpW(NT), gpW(+G), and gpW(CT), suggesting that all three assume the same structure with no contribution from the affinity tags (data not shown). Near UV CD spectra were also collected for gpW(CT). As the signal in the near UV region is much weaker than far UV gpW(CT) was the only protein that was soluble enough to be examined by this technique. gpW(CT) displays significant near UV protein spectrum, with a single maximum observed at 280 nm (Figure 2-4).

![Figure 2-4. CD spectra of folded (•) and unfolded (○) gpW. (A) These spectra of gpW(+G) were obtained from a 25 μM sample of protein in a 1 mm cuvette. The mean residue ellipticity value of ~28.6 at 222 nm indicates that the protein is approximately 47% helical. (B) Near UV CD spectra of gpW(CT) at a protein concentration of 80 μM in a 1 cm cuvette.](image)

Since a large difference was seen between its folded and unfolded spectra (Figure 2-4A and B), CD provided a means to monitor gpW unfolding. All three tagged gpW variants showed fully reversible thermal denaturation curves that were independent of protein concentration (Figure 2-5A). The three proteins displayed almost identical transition midpoint temperature ($T_m$) values, which averaged to 71.5 °C. Additionally, the thermal denaturation of gpW(CT) following both near and far UV CD signal gives identical curves (Figure 2-5A), illustrating that the thermal unfolding transition is independent of the monitoring method.
Urea-induced denaturation curves were obtained for each of the tagged gpW constructs and in each case a single transition between the folded and unfolded states was observed (Figure 2-5B). The free energy of unfolding in water ($\Delta G_u$) and the dependence of the free energy of unfolding on urea concentration ($m$) were found to be independent of protein concentration (data not shown) and were similar for each of the wild type tagged proteins (Table 2-1), giving average values of 3.01 kcal mol$^{-1}$ and 0.66 kcal mol$^{-1}$ M$^{-1}$ respectively. All thermodynamic data were fitted with the assumption, based on the native molecular weight determination described above, that gpW folds and unfolds as a monomer. The conclusion that gpW folds as a monomer is also supported by the observation that its stability, as measured by thermal and chemical denaturation, was independent of protein concentration.

To fully analyze the data from the thermal melts, the molar heat capacity change upon unfolding ($\Delta C_p$) was determined using the method of Pace and Laurents (1989). For this purpose, eight urea melts were performed at various temperatures between 15 °C and 35 °C. The $\Delta G_u$ values calculated from these melts were plotted and fit in combination with $\Delta G_u$ values calculated from the transition regions of thermal denaturation experiments (Figure 2-6). A $\Delta C_p$ value of 0.570 kcal mol$^{-1}$ K$^{-1}$ was derived by fitting of this curve.
Figure 2-6. $\Delta G_u$ versus temperature plot for gpW(NT). These data were gathered from the unfolding free energies calculated in the transition zone of a thermal melt (○), as well as eight urea denaturation experiments (●) performed at various temperatures between 15 and 35 °C. Fitting of this curve as described in Materials and Methods yielded a $\Delta C_p$ of 0.570 kcal mol$^{-1}$ K$^{-1}$.

Mutagenesis of the C-terminus of gpW

Since gpW(CT) was less active than both untagged and N-terminal tagged gpW, we postulated that the C-terminus may be important for protein function. To further investigate the role of the C-terminus in gpW stability and function, each of the last three positions (residues 66-68) of gpW(NT) were replaced with several different amino acids and an amber codon. These N-terminal tagged mutant proteins were subsequently purified and characterized. Each variant displayed the same CD spectrum as WT (data not shown), and retained WT thermal stability (Table 2-2), indicating that no large structural rearrangements occurred upon substitution.

While the substitutions at the C-terminus of gpW caused no reduction in thermodynamic stability, dramatic effects on biological activity were observed (Table 2-2). A single substitution of the terminal residue (V68E) completely abolished activity, as did truncations of the protein (resulting from amber mutations) by more than one residue. Greater than 1000-fold decreases in activity were displayed by single substitutions at all three C-terminal residues. The activity of the mutant proteins was measured at two concentrations: 5 μM, where the concentration of WT gpW is limiting in the in vitro reaction; and 15 μM, where the concentration of WT gpW is in excess (Figure 2-2A). The relative activity of most of the mutants compared to WT increased markedly when they were added to the reaction in
excess. For example, the activity of the Y67S mutant is reduced by 10⁴-fold compared to WT when assayed at 5 μM, but is less than 3-fold reduced when assayed at 15 μM, indicating that this mutant is able to function at close to normal level when present in the reaction in excess. The increased relative activity of mutants when present in excess may indicate that these have a reduced affinity for one or more of the components in the assembly reaction; thus, a higher concentration of protein can partially or completely compensate for the defect. Additionally, some mutants may have an intrinsically reduced activity once incorporated into the phage, which could lead to a lower plateau of infectivity at saturating concentrations of protein.

*Destabilizing Substitutions in gpW*

To provide a contrast to the C-terminal gpW substitutions, which all retained WT thermodynamic stability, we set out to design mutations that would decrease the stability of gpW. In the strongly predicted helix including residues 40-53 (Figure 2-1A), the periodicity of hydrophobic residues suggested that the residues at positions 40, 43, 47, and 50 would form part of the hydrophobic core of the protein. Three of these residues, Val 40, Leu 43, and Leu 50, were substituted with Ser or Trp, two of the substitutions made in the C-terminal residues. If these positions were buried in the hydrophobic core of the protein, their substitution with Ser, a polar residue, or Trp, a very bulky aromatic residue, would be expected to decrease stability. Substitutions at these positions did indeed lead to large decreases in stability (Table 2-2). Most dramatically, the L43W substitution resulted in a protein that was completely unfolded at room temperature. The V40S and L50W mutants displayed close to 20 °C decreases in their T_m values. In contrast, substitution of Val 52, which is not predicted to form part of the hydrophobic core, caused no decrease in thermostability (Table 2-2).

The V40S, V52L, and V52W mutants all retained close to WT *in vitro* activity, even when assayed at a protein concentration of 5 μM. The high activity of the V40S mutant demonstrates that a large destabilization of gpW does not necessarily lead to a reduction in biological activity. The Val 52 position does not appear to play a crucial role in either the structure or function of gpW. The L50W mutant activity is reduced by approximately 20-fold, indicating that structural rearrangements induced by this putative hydrophobic core substitution cause some alteration of the functional surface of the protein. Surprisingly, the L43W mutant retains some activity even though it is unfolded under the conditions of the assay. This observation suggests that the assembly of this mutant into phage helps to stabilize
its native structure. Its failure to display an increase in relative activity when assayed at high concentration could indicate that the phage particles containing this unstable mutant protein are themselves unstable (i.e. the ability of this mutant to produce plaque forming units is limited because the phage produced are too unstable to infect cells efficiently).

Table 2-2

<table>
<thead>
<tr>
<th>Protein</th>
<th>Tm (°C)</th>
<th>Normalized Biological Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Protein concentration</td>
<td>5 μM</td>
</tr>
<tr>
<td>WT</td>
<td>71.8</td>
<td>100</td>
</tr>
<tr>
<td>V40S</td>
<td>54.0</td>
<td>65</td>
</tr>
<tr>
<td>L43W</td>
<td>unfolded</td>
<td>0.011</td>
</tr>
<tr>
<td>L50W</td>
<td>52.4</td>
<td>4.5</td>
</tr>
<tr>
<td>V52L</td>
<td>71.8</td>
<td>91</td>
</tr>
<tr>
<td>V52W</td>
<td>73.2</td>
<td>28</td>
</tr>
<tr>
<td>F66Am</td>
<td>71.7</td>
<td>&lt; 10⁻⁵</td>
</tr>
<tr>
<td>F66L</td>
<td>71.0</td>
<td>&lt; 10⁻⁵</td>
</tr>
<tr>
<td>F66W</td>
<td>70.5</td>
<td>0.0024</td>
</tr>
<tr>
<td>Y67Am</td>
<td>73.1</td>
<td>&lt; 10⁻⁵</td>
</tr>
<tr>
<td>Y67L</td>
<td>72.5</td>
<td>0.046</td>
</tr>
<tr>
<td>Y67S</td>
<td>73.2</td>
<td>0.001</td>
</tr>
<tr>
<td>Y67W</td>
<td>73.2</td>
<td>0.018</td>
</tr>
<tr>
<td>V68Am</td>
<td>73.6</td>
<td>&lt; 10⁻⁵</td>
</tr>
<tr>
<td>V68L</td>
<td>72.7</td>
<td>34</td>
</tr>
<tr>
<td>V68S</td>
<td>73.6</td>
<td>0.088</td>
</tr>
<tr>
<td>V68W</td>
<td>72.7</td>
<td>0.020</td>
</tr>
<tr>
<td>V68E</td>
<td>71.8</td>
<td>&lt; 10⁻⁵</td>
</tr>
</tbody>
</table>

*a not determined

When the V40S, L50W, and L43W mutant proteins were expressed in WT *E. coli*, they did not accumulate to a high level and could not be purified. Since thermodynamically unstable proteins are generally degraded rapidly in *E. coli*
(Parsell & Sauer, 1989), this result was not surprising. Unstable proteins spend a greater amount of time in the unfolded state, and thus are better targets for intracellular proteases than stable proteins. In order to obtain purified protein for *in vitro* studies, we expressed the unstable mutants in the *E. coli* strain SG1146A (Gottesman *et al.*, 1998), which lacks the ClpP protease, which is responsible for selectively degrading proteins with hydrophobic C-termini (Levchenko *et al.*, 1997). Using this strain we were able to purify the unstable mutants at levels comparable to the wild type protein.

**Discussion**

Although gpW was partially purified previously, its structural, biophysical, and functional properties had been largely unstudied. The construction of 6-His tagged forms of gpW described here and the demonstration that these constructs possess thermodynamic stability and *in vitro* activity at levels similar to the untagged protein has allowed a detailed analysis of gpW structure and function.

**Structural Features of gpW**

Our results indicate that gpW is a monomeric, helical protein. The degree of ellipticity seen at 222 nm (Figure 2-4) corresponds to a protein with approximately 47% helical content, which is similar to the 43% helical content predicted by the PHD structure prediction program. The validity of this structure prediction is also supported by our ability to successfully predict destabilizing substitutions in gpW (Table 2-2).

Denaturation experiments of the tagged and untagged proteins show cooperative, fully reversible unfolding curves with a single transition, suggesting that gpW folds by a two-state process in which only the native and denatured states are significantly populated. Thermal denaturation curves of gpW(CT) were monitored by the change in CD signal in both the near and far UV wavelengths. As the far UV CD signal reports primarily on secondary structure, and near UV CD signal provides tertiary structure information (Towell & Manning, 1994), monitoring protein denaturation in these two regions provides distinct probes with which to assess protein unfolding. The fact that the near and far UV CD thermal denaturation curves are coincident supports a two-state unfolding model.
It is generally assumed that protein affinity tags will be unstructured in solution and thus have little effect on the stability or folding properties of the tagged protein. This assumption holds true with gpW, as the values calculated for various thermodynamic parameters obtained for temperature and urea melts of the two tagged proteins and the cleaved protein shows that the thermodynamic stability of gpW is not affected by the presence of the tags. Thus, the stability data obtained from the tagged versions of the protein is comparable to the results that would be obtained from untagged protein.

Earlier work from our laboratory led to the conclusion that gpW is a small, heat-stable protein (Perucchetti et al., 1988). Although the protein is only 68 residues long and has no disulfide bonds, temperature induced unfolding experiments on the three WT constructs monitored by CD show that it is indeed thermostable, with a denaturation midpoint of 71.5 °C. ΔG_u values obtained for the three proteins were similar, with an average value of 3.0 kcal/mol at 25 °C. This property of possessing high thermal stability accompanied by moderate stability at ambient temperatures is typical of small proteins as they have a low ΔC_p of unfolding. Other small proteins such as the SH3 domain and the B1 IgG-binding domain (approximately 60 residues each) display similar behavior (Alexander et al., 1992; Lim et al., 1994).

Functional Characterization of gpW

The most striking feature of the in vitro activity of gpW is its extreme concentration dependence (Figure 2-2B). The most likely explanation for this phenomenon is that many molecules of gpW are required for the formation of a single phage particle. The time-dependent measurements of the gpW mediated reaction support this supposition in that the reaction occurs very quickly at first when the concentration of free gpW is highest, but decreases rapidly as the components of the reaction are incorporated irreversibly into phage particles. This assay does not provide a clear answer to which of the components becomes limiting. The incorporation of gpW into the phage particle is assumed to be an irreversible reaction as once the phage are formed, they are stable for long periods of time. In addition, large dilutions of the reactions required for plating the phage do not cause the particles to dissociate. The demonstration by analytical ultracentrifugation and size exclusion chromatography that free gpW is monomeric even at very high concentrations suggests that if gpW oligomerizes, it must do so as it is being incorporated into phage particles.
An enigmatic aspect of the *in vitro* gpW reaction is the necessity to add gpW in vast excess compared to the other reaction components. A typical reaction contains approximately $10^{15}$ molecules/ml of gpW to produce a maximum of only $10^8$-$10^9$ phage/ml. In spite of this excess of gpW, the concentration of gpW is clearly the limiting factor in the yield of the reaction when it is present at a concentration of 5 μM or below. One explanation for this observation is that a large proportion of gpW molecules are inactive. However, gpW(NT) in a crude *E. coli* extract produces the same number of phage particles as purified gpW(NT), showing that the protein is not becoming inactivated during the purification process. A second possibility, that a large proportion of the protein molecules are not folded, is discounted by the CD data, which shows that almost all of the protein is folded at 37 °C. A third possibility is that the protein is subtly proteolyzed in *E. coli*, while it is being expressed. This theory can also be ruled out, as the mutagenic data showing distinct differences in the C-terminal mutations shows that the C-terminus of the protein is not being proteolyzed, and the purification of the protein using the hexahistidine tag shows that the N-terminus is intact. For these reasons we think that it is unlikely that a large percentage of gpW molecules are inactive. Thus, the most likely explanation for the data is that the *in vitro* assembly of each phage particle requires many molecules of gpW. The requirement for a large excess of gpW *in vitro* appears to differ from that inside phage-infected cells, where the level of gpW appears to be relatively low. It was never possible to identify a band on SDS polyacrylamide gels corresponding to gpW in lysates from phage-infected cells or from plasmid-transformed cells expressing gpW under the control of its natural translation initiation site (Murialdo & Siminovitch, 1972; unpublished results). It is known that the low expression level of certain other λ morphogenetic genes is mediated by poor translation initiation (Murialdo et al., 1987). The low yields of gpW purified from phage-infected cells also indicate that gpW is present at relatively low levels *in vivo*. (8). Clearly, some distinct properties of the phage assembly process *in vivo* must exist to allow gpW to function efficiently at low concentrations.

**The Role of the Three C-Terminal Residues**

The sequence alignment of the gpW homologues from phage 21 and N15 shows that the hydrophobic nature of the C-terminus of the protein is conserved. In all three proteins positions 66 and 68 are hydrophobic, while position 67 is hydrophobic in λ and N15, and is an Arg, which has significant aliphatic character,
in phage 21. Although conserved, substitution of the three C-terminal residues of gpW does not decrease the stability of the protein, nor does truncation of the protein by insertion of an amber codon at any of these positions. This suggests that the C-terminus is not buried in the interior of the protein as might be expected for a hydrophobic region, but is solvent exposed. This idea is supported by our observation that the gpW constructs with wild type C-termini (gpW(NT) and gpW(+G)) are much less soluble than gpW(CT), which has charged residues at its C-terminus. When the native hydrophobic C-terminus is present, low solubility likely results from the exposure of these hydrophobic residues, which could cause protein aggregation. In examining the gpW sequence, it can be seen that the last 12 residues consist of mostly polar and charged residues with no large hydrophobic residues except the last three, suggesting that this region may either be unstructured (as predicted by the PHD program), or posses only local structure that does not contribute to the stability of the whole protein. The possibility that gpW may contain some unstructured regions is supported by the $\Delta C_p$ and $m$ values calculated from our unfolding data, which are considerably lower than would be expected for a fully folded 68 residue protein (Myers et al., 1995).

The results of our mutagenesis experiments clearly demonstrate that the last three residues of gpW are crucial for its activity (Table 2-2). Even relatively conservative substitutions at these positions (e.g. F66W, Y67W) cause large decreases in activity, and other substitutions totally abrogate function (e.g. V68E). The dependence of a protein's function upon the character of its last few residues is quite unusual. We hypothesize that the C-terminus of gpW is disordered, and may serve as a binding site for another protein involved in $\lambda$ morphogenesis (e.g. gpFII, which is incorporated into the head after gpW addition). The hydrophobic and functionally critical nature of the C-terminal residues of gpW are reminiscent of proteins bound by PDZ domains in eukaryotic cells (Fanning & Anderson, 1996). PDZ domains bind to hydrophobic C-termini of target proteins using the terminal carboxylate group as part of their recognition site. In $E. coli$, proteins with hydrophobic C-termini have been shown to be specifically degraded by a variety of proteases present in the cytoplasm and periplasm (19,20). The ClpP-containing proteases, which we have found are able to degrade our destabilized gpW mutants, comprise a group of these tail-specific proteases. Interestingly, this protease is thought to recognize its substrates using PDZ-like domains (Levchenko et al., 1997). Future studies will be directed at determining whether a PDZ-like interaction may be important in the function of gpW.
References


CHAPTER 3

The novel fold of the bacteriophage $\lambda$ protein W reveals structural similarity to the DNA sliding clamps.


Adelinda Yee collected the NMR spectra for gpW and Valerie Booth assisted in training me to use the structure analysis software. I completed all of the data analysis and modeling.
Introduction

The 68 residue protein W from bacteriophage λ is required for the stabilization of DNA within the phage head and for interaction with gpFII, which subsequently allows the attachment of tails onto the head during morphogenesis. Since gpFII and tails are known to be attached at the connector (Casjens, 1974), gpW is also likely to assemble at this site. The addition of gpW to filled heads increases the DNase resistance of the packaged DNA, suggesting that gpW either forms a plug at the connector to prevent ejection of the DNA, or binds directly to the DNA (Perucchetti et al., 1988). The large number of positively charged residues in gpW (its calculated pI is 10.8) is consistent with a role in DNA interaction. Thus, in the mature λ virion, gpW likely interacts with gpB in the connector, gpFII, and DNA.

The multiplicity of functions performed by this small protein make it an intriguing subject for detailed structural investigation. To better understand the mechanism by which gpW interacts with other head proteins and DNA, I determined its atomic structure. This chapter describes the solution structure of gpW determined using NMR spectroscopy. gpW possesses a novel fold consisting of two α-helices and a single two-stranded β-sheet arranged around a well-packed hydrophobic core. It displays an intriguing structural similarity with the "sliding clamp" proteins associated with DNA polymerases. Using a sliding clamp structure, gpW was modeled into a structurally plausible hexameric ring, which is consistent with its known roles in bacteriophage λ morphogenesis.

Materials and Methods

Cloning, Expression, and Purification of gpW

The sequence encoding full length gpW from bacteriophage λ was cloned into the pET15b vector (Novagen), as described in Maxwell et al. (2000) and expressed with the FLAG epitope (IBI Technologies) and a hexa-histidine tag fused to the C-terminus. The fusion protein was overexpressed in Escherichia coli BL21 (λDE3) cells by growing at 37°C in M9 minimal medium enriched with 0.7 g/liter of 15N-NH₄Cl and 2.5 g/liter of 13C-glucose to an OD₆₀₀ of 0.8, inducing with 200 μg/mL IPTG, then incubating at 37 °C for an additional 3 hours. Purification of the protein was performed as described in Maxwell et al. (2000). A final purification step using a Superdex 75 FPLC column was performed, and the protein was dialysed into buffer
containing 10 mM NaH₂PO₄, 200 mM NaCl. The protein was concentrated by ultrafiltration to approximately 1.5 mM. Ten percent D₂O was added to provide NMR lock signal. For NMR experiments that require D₂O as the solvent, the samples were lyophilized and resuspended in D₂O.

**NMR Spectroscopy**

All NMR spectra were collected at 25 °C on either a Varian 500-MHz or 600-MHz Inova spectrometer equipped with pulse field gradient units and actively shielded triple-resonance probes. NMR data were processed using nmrPipe software (Delaglio *et al.*, 1995) and were analyzed using NMRView 3.0 (Johnson & Blevins, 1994). Sequence-specific backbone resonance assignments were achieved using ¹⁵N-HSQC (Pascal *et al.*, 1994), CBCA(CO)NH (Gronenborn & Clore, 1996), and HNCACB (Kay *et al.*, 1994) spectra. Side chain ¹H and ¹³C resonances were assigned using HCCH-TOCSY (Bax *et al.*, 1990; Kay *et al.*, 1993), CCC-TOCSY (Montelione *et al.*, 1992), and ¹⁵N-edited TOCSY (Marion *et al.*, 1989) spectra. Side chain ¹H resonances of aromatic residues were assigned using homonuclear two-dimensional NOESY (Jeener *et al.*, 1979) and TOCSY (Braunschweiler & Ernst, 1983) spectra in D₂O. ¹³C, ¹⁵N-edited NOESY (Pascal *et al.*, 1994) spectra in H₂O and homonuclear NOESY in D₂O (mixing times 150 ms) were used to identify H-H nuclear Overhauser effects (NOE). ³JNH-Hα couplings were measured using an HNHA experiment (Kay & Bax, 1990). In total 86% of the backbone resonances were assigned (residues 3-74 of 82).

**Structure Determination**

Structure calculations were performed using version 3.851 of XPLOR with ambiguous restraints for iterative assignment (ARIA) (Nilges *et al.*, 1997). The initial input for ARIA consisted of the ten lowest energy structures calculated by XPLOR using 324 unambiguous, manually assigned NOE, dihedral angle, and hydrogen bond restraints. Dihedral angle restraints were derived from ³JHNHα couplings, and supplemented with TALOS (Cornilescu *et al.*, 1999) predictions when they were in agreement. Hydrogen bond restraints were added for residues that were clearly in the α-helices or β-sheet as evidenced by NOE patterns and chemical shift values.

The NOE peak lists for ARIA were generated from manual peak picking the NOESY spectra using NMRView 3.0 (Johnson & Blevins, 1994). The frequency window tolerance for assigning NOEs using ARIA was ±0.03 ppm for the proton and ±0.5 ppm for nitrogen and carbon dimensions. The ARIA parameters p, Tₓ and Nₓ were as described in Nilges *et al.* (Nilges *et al.*, 1997). Twenty structures were refined in

54
each iteration, and the ten lowest energy structures were used for the purpose of residue specific NOE assignments. In the final (eighth) iteration, 50 structures were refined, and the 15 lowest energy structures were retained for analysis. Residues 59-82 were disordered and were not included in the structural models. The residues that comprise the hydrophobic core of the protein were determined by analysis of the structure using the program ENVIRONMENTS (Bowie et al., 1991).

Modeling of the Hexameric Ring

The α-helices from the gpW structure (residues 4-17 and 40-52) were compared to pairs of helices from each of the sliding clamp structures. The DNA-binding helices from the human PCNA (hPCNA) protein overlaid with those of gpW with the lowest r.m.s.d. over the greatest number of atoms. The six pairs of DNA-binding helices from the hPCNA structure were then used as a scaffold upon which the backbone atoms in the helices from six gpW monomers were superimposed using Swiss PDB Viewer. No further energy minimization steps were performed. The resulting PDB file was then analyzed for steric clashes using Swiss PDB Viewer, and the burial of residues in the monomer and hexamer was calculated using the program ENVIRONMENTS (Bowie et al., 1991).

Results and Discussion

The Solution Structure of Monomeric gpW

NMR data were collected for a biologically active (Maxwell et al., 2000) recombinant gpW fusion protein with a C-terminal FLAG epitope followed by a hexa-histidine tag. The fusion was necessary because untagged gpW is soluble only to 50 μM, while the C-terminally tagged protein remains soluble and monomeric at a concentration of 2 mM. The three dimensional structure of gpW was determined in aqueous solution at pH 7.0, 30 °C, using a total of 1249 experimental constraints.

The ensemble of 15 low-energy structures calculated for gpW is presented in Figure 3-1A, and the statistical parameters of the structure determination are summarized in Table 3-1. gpW is composed of a two-stranded, antiparallel β-sheet packed against two α-helices. The two β-strands, comprised of residues 23-27 (β1) and residues 31-35 (β2), are separated by a tight turn (Figure 3-1B). The two α-helices are packed in an anti-parallel orientation, and consist of residues 4-17 (α1) and 40-52 (α2). The protein is stabilized by a single, well packed hydrophobic core with residues contributed by all four secondary structural elements. The core is comprised
of Leu 7, Ala 8, Ala 10, Leu 14, Leu 17, Ala 24, Val 33, Val 40, Leu 43, Ile 47, and Leu 50, which are all greater than 94% buried (Figure 3-1C).

The 10 C-terminal residues of gpW and the C-terminal tag displayed random-coil-like backbone chemical shifts, and possessed no medium- or long-range NOEs, consistent with a disordered conformation for this region of the protein. The finding that the last ten residues of gpW are unstructured is consistent with our previous thermodynamic characterization of this protein, which showed that its $\Delta C_p$ of unfolding was considerably lower than would be expected for a fully folded 68 residue protein (Maxwell et al., 2000). The unstructured nature of the gpW C-terminus was also presaged by our observation that the last three residues of gpW, though hydrophobic, did not contribute to the stability of the domain.

A search of the protein structure database for folds similar to that of gpW using the program DALI (Holm & Sander, 1993) yielded no matches with both $\alpha$-helices and $\beta$-strands. A manual search of the SCOP (Murzin et al., 1995) and CATH (Orengo et al., 1997) databases also failed to detect any similar folds. Thus, we conclude that gpW represents a novel fold. With a folded region of only 50 residues, gpW appears to be the smallest known fold containing both $\alpha$-helices and a $\beta$-sheet, but lacking disulfide bonds and cofactors.

*The Structure of gpW Resembles DNA Sliding Clamps*

Although the overall fold of gpW is novel, manual inspection of various structures revealed an intriguing similarity to a portion of the "sliding clamp" protein complexes responsible for the processivity of many DNA polymerases (Kuriyan & O'Donnell, 1993). The structures of five different sliding clamp proteins have been solved by X-ray crystallography. The three dimensional folds of these proteins are all very similar, despite a lack of sequence similarity between some of them. The sliding clamps are homo-oligomeric complexes with pseudo six-fold symmetry. Each complex forms a toroid-like structure with an inner pore formed by six pairs of $\alpha$-helices supported by a $\beta$-sheet scaffold. The $\alpha$-helices are the most structurally similar part of the sliding clamp molecules, while the $\beta$-sheets are more variable (Figure 3-2A). These proteins do not form base-specific contacts with DNA, but rather encircle it and thereby form a topological link. The inner diameter of the ring formed by the sliding clamps is 30-35 Å, which can easily accommodate double stranded DNA with no steric repulsion, allowing the clamp to slide freely along the DNA molecule.
Figure 3-1. NMR solution structure of gpW. (A) Stereo view showing the best-fit superposition of the backbone atoms (N, Ca, and C') of the final 15 structures of gpW. The structures are superimposed against the average structure using residues 4-54. (B) Ribbon representation of residues 1-58 of the lowest energy solution structure. The secondary structure elements are labeled. (C) Ribbon representation of the lowest energy structure showing the packing of the hydrophobic core residues. The structure is rotated 90° about the y-axis with respect to (B). Structures are displayed using the molecular graphics program SETOR (Evans, 1993).
Table 3-1

*Structural statistics for the family of 15 gpW structures* \(^1,2\)

<table>
<thead>
<tr>
<th>Experimental restraints:</th>
<th>Number of restraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOE restraints (total)</td>
<td>1171</td>
</tr>
<tr>
<td>Unambiguous</td>
<td>883</td>
</tr>
<tr>
<td>Intraresidue (i-j = 0)</td>
<td>280</td>
</tr>
<tr>
<td>Sequential (</td>
<td>i-j</td>
</tr>
<tr>
<td>Medium range (2&lt;</td>
<td>i-j</td>
</tr>
<tr>
<td>Long range (</td>
<td>i-j</td>
</tr>
<tr>
<td>Ambiguous</td>
<td>288</td>
</tr>
<tr>
<td>Hydrogen bond restraints</td>
<td>51</td>
</tr>
<tr>
<td>Dihedral angle restraints</td>
<td>72</td>
</tr>
</tbody>
</table>

Mean r.m.s. deviations from the experimental restraints

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance (Å)</td>
<td>0.034 ± 0.002</td>
</tr>
<tr>
<td>Dihedral angle (°)</td>
<td>0.895 ± 0.057</td>
</tr>
</tbody>
</table>

Mean r.m.s. deviations from idealized covalent geometry

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bond (Å)</td>
<td>0.037 ± 0.002</td>
</tr>
<tr>
<td>Angle (°)</td>
<td>0.558 ± 0.020</td>
</tr>
<tr>
<td>Improper (°)</td>
<td>0.464 ± 0.024</td>
</tr>
</tbody>
</table>

Mean energies (kcal mol-1)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ENOE</td>
<td>68.5 ± 10.1</td>
</tr>
<tr>
<td>Ecdih</td>
<td>3.63 ± 0.47</td>
</tr>
</tbody>
</table>

Ramachandran analysis (residues 4-53)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Residues in favored regions (%)</td>
<td>84.2</td>
</tr>
<tr>
<td>Residues in additional allowed regions (%)</td>
<td>12.0</td>
</tr>
<tr>
<td>Residues in generously allowed regions (%)</td>
<td>3.8</td>
</tr>
<tr>
<td>Residues in disallowed regions (%)</td>
<td>0</td>
</tr>
</tbody>
</table>

\(^1\)None of the structures exhibits distance violations greater than 0.3 Å or dihedral angle violations greater than 5°.

\(^2\)The program PROCHECK was used to assess the overall quality of the structures.

A surprisingly high degree of similarity was noted in the length and packing angle of the helices of gpW and those of the sliding clamps. A comparison of the helices of gpW with those from a single domain of a sliding clamp (the human PCNA structure was used in this case) shows that they are the same length, and 108 backbone atoms (27 residues) can be overlaid with an r.m.s. deviation of 1.57 Å (Figure 3-2B). In addition, the β-sheet from gpW packs against the outer side of the protein at an angle similar to one of the β-sheets in the human PCNA. To assess whether the similarity of the gpW and sliding clamp helices exists simply because
this particular helical arrangement is very common in proteins, we compared the gpW structure to 27 different proteins identified by DALI as having helical structures similar to gpW. Of the 27, only 3 had helices that were similar to gpW in both length and packing angle. However, the helices in these three proteins were buried in the interior of the protein, and lacked the solvent exposed surface observed in the helices of gpW and the sliding clamps. This point is significant because the interface between the sliding clamps and DNA is provided by their helices.

Because of the known roles and site of action of gpW, we believe that the structural similarity between gpW and the sliding clamps may have functional significance. As mentioned above, the sliding clamps are composed of six structurally similar domains forming a ring with approximate six-fold rotational symmetry. gpW likely assembles on the connector, which is composed of 12 molecules of gpB assembled into a ring with 12-fold rotational symmetry. In addition, the mature phage particle contains six copies of gpFII (Tsui & Hendrix, 1980), the other binding partner of gpW, and the phage tail possesses six-fold rotational symmetry (Buchwald et al., 1970). For these reasons, it seems very plausible that gpW forms a six membered ring when it is assembled into the phage particle.

Modeling gpW Into a Sliding Clamp-Like Ring

Using the human PCNA sliding clamp structure as a model, a hexameric ring of gpW was generated (Figure 3-3A, B). The ring has an outer diameter of approximately 120 Å, and the inner channel approximately 30 Å. Since gpW is located at the head-tail junction, the size of the gpW ring must be compatible with both the phage connector and tail. No high resolution structural data is available for the mature connectors, but electron microscopy (EM) has been performed on the preconnector, an intermediate in prohead assembly composed of 12 uncleaved gpB molecules. Although a number of fusion and cleavage reactions take place to form the mature connector, the overall shape and dimensions of the preconnector are in agreement with the dimensions of the head-tail connector from disrupted phage (Harrison et al., 1973; Tsui & Hendrix, 1980; Kochan et al., 1984b). Negatively stained electron micrographs show that preconnectors are doughnut shaped molecules with an outer diameter of 140 to 150 Å, and an axial hole of 22 to 25 Å (Kochan et al., 1984a). Additionally, EM studies of bacteriophage λ tails have shown that they are thin, flexible tubes, composed of layered disks made up of six subunits of the major tail protein gpV (Buchwald et al., 1970). An endview of the hexameric subunits shows
Figure 3-2. gpW shows structural similarity to the DNA polymerase sliding clamp proteins. (A) Structural similarity of gpW to single domains from four DNA sliding clamp structures. (B) Overlay of the helices of gpW (green; residues 4-17 and 40-52) and the human PCNA domain (yellow; residues 9-23 and 71-81). The PCNA DNA-contacting helices are similar to gpW helices in both length and packing angle, and 108 backbone atoms in the helices overlay with an r.m.s. deviation of 1.57 Å.
that the tail has an outer diameter of approximately 180 Å and a central hole 30 Å in diameter. Thus, the size and symmetry of the gpW ring is compatible with both connector and tail structures.

A space filling representation of the hexameric gpW ring (Figure 3-3C, D) shows that a number of residues are more highly buried upon formation of the ring, including Ala 9, Ala 12, and Asp 16 from one monomer, and Lys 45 and Val 52 from the adjacent monomer. Examination of the model reveals only one steric clash, involving the side chains of Asp 16 and Lys 45. The NMR data indicate that Lys 45 is exposed on the surface of the monomer, and is highly mobile in solution as evidenced by the absence of inter-residue NOEs. It is therefore likely that the Lys 45 side chain could assume an alternative conformation, not subject to steric problems in the assembled phage particles. However, the proximity of Asp 16 and Lys 45 suggests that they could form a salt bridge and thereby contribute to the specificity with which the monomers come together to form the ring structure. The construction of this model demonstrates that gpW can be formed into a structurally plausible ring using an extremely simple modeling procedure (e.g. no energy minimization was used).

Functional Implications of the gpW Ring

The functional relevance of the hexameric gpW ring presented here is supported by a number of features. Four basic residues, Arg 11, His 15, Lys 44, and Lys 45 are located on the surface of the two helices, and form a positively charged ring on the inside of the hexamer. Calculation of the electrostatic charge dispersion of the hexameric ring structure shows that the inner surface of the ring structure is predominantly positively charged, providing a favourable surface for interaction with DNA (Figure 3-3C, D). The charge dispersion in the ring is similar to that observed in the DNA sliding clamps, with the inside of the ring being predominantly positively charged. In addition to the favourable electrostatic environment, the dimensions of the inner channel (~ 30Å diameter) of the gpW ring are compatible with a double stranded DNA molecule. In the completed phage, the DNA protrudes into the tail by approximately 200 base pairs, and upon infection it must pass through the tail as it is injected into the host cell. Thus, as DNA must pass through the part of the connector where gpW is likely positioned, the ring of gpW must be large enough to accommodate it, and positively charged in order to allow it to pass through. An exposed hydrophobic surface formed at the top of the gpW ring by the Met 18 and Ala 31 residues provides a potential site of interaction.
Figure 3-3. *gpW can be modeled into a hexameric ring.* (A) Six *gpW* monomers can be modeled into a ring based upon the structure of the human PCNA DNA sliding clamp. (B) Side view of the hexameric ring, with the upper model (A) rotated 90° about the x-axis. (C) Surface representation and electrostatic potential generated with GRASP (Nicholls et al., 1991) showing positively charged (basic) regions in blue (+9.351 kT), and negatively charged (acidic) regions in red (-7.870 kT). (D) Side view of the center channel of the ring. This view is a 90° rotation about the x-axis relative to (C), and the front three monomers have been removed.
with other phage proteins, such as gpFII or gpB.

We previously demonstrated that substitutions at the last three residues of gpW had dramatic effects on biological activity. The C-terminal region of gpW, which has been shown here to be unstructured, may serve as a binding site for another phage protein, likely either gpB in the connector, or gpFII. This feature of gpW provides another parallel with the sliding clamps. Substitutions within the last four residues of both the yeast PCNA protein and E. coli DNA polymerase β abrogate the function of these proteins (Kelman et al., 1999). These conserved regions in both proteins have been shown to be critical for interactions with components of the replication machinery. As shown in Figure 3-4, the C-terminal sequences of the DNA polymerase β subunits from various bacteria are similar to those at the C-termini of gpW and its homologue from phage 21 in that they all possess hydrophobic residues at the last and third to last positions. Although gpW does not interact with the same proteins as these sliding clamp proteins, the mode of interaction mediated through the C-termini of gpW and the sliding clamps may be the same.

| λ gpW | R G P A G F Y V |
| Phage 21 gp3 | R R P L G V R L |

Bacterial β-Subunits:
- E. coli: V M P - - M R L
- A. aeolicus: I M P - - M R V
- C. crescentus: L M P - - M R V
- H. pylori: M M P - - I T L
- R. prowazekii: I M P - - V K V

Figure 3-4. Sequence alignment of the C-terminus of gpW with sliding clamps. The sequences of the C-termini of the DNA Polymerase III β-subunits from various bacteria are compared to the C-termini of gpW and its homologue from bacteriophage 21. Conserved positions are noted in bold.

The fact that gpW remained monomeric at the high protein concentrations used for the NMR structure determination suggests that the forces driving multimerization may be quite weak and the presence of DNA and/or the connector proteins may be required for gpW to assemble into a hexamer. This is consistent with the order in which the phage particle is known to assemble. Several phage and viral proteins are known to be monomeric at high concentrations under certain conditions, and multimeric under others. For example, the λ gpD protein is
monomeric at high concentrations in solution, but assembles onto the head surface after DNA packaging. Under the conditions used for its crystallization, it assembled into the trimeric form that is actually observed in high resolution EM micrographs of the phage head (Yang et al., 1999). The capsid protein of Rous Sarcoma Virus also remains monomeric in solution at high protein concentration, even though it is known to assemble into a large multimeric structure in the viral particle. In this case, lowered pH and high salt caused the protein to polymerize (Kingston et al., 2000). We have not yet been able to identify conditions under which gpW is able to form soluble multimers, but these cases illustrate that this may be possible.

Conclusion

We have shown that gpW possesses a novel fold consisting of two α-helices and a β-sheet. The secondary structure of the protein was found to correlate highly with the secondary structure predicted by PHD (Figure 2-1). The β-strands were identical to the prediction, and the α-helices are only one residue longer in the prediction. We propose that gpW assumes a hexameric structure similar to that of the DNA sliding clamps when incorporated into phage particles. The existence of this structure is supported by the similarity between the gpW helices and those of the sliding clamp proteins and by our demonstration that a structurally plausible gpW ring can be easily constructed using a sliding clamp structure as a model. Furthermore, the known functions of gpW are entirely consistent with hexameric ring-shaped structure.
References


CHAPTER 4

The Solution Structure of the Bacteriophage λ Head-Tail Joining Protein, gpFII

I purified the protein, and performed all of the structural analysis, biochemical and biophysical characterization. Adelinda Yee collected the NMR data.
Introduction

The bacteriophage λ is a member of a group of temperate phages, including bacteriophages 21, N15, and φ80, that share a common organization of their genes and are capable of productive genetic recombination. Heteroduplex DNA analysis reveals that recombination events in the heads and tails of these bacteriophage are very rare, likely due to the intimate interactions of the proteins in these structures. It is improbable that mixing head and tail genes from two different phages will produce a functional group of proteins. The exception is a sequence transition that often occurs within the structural genes roughly at the boundary between the head and tail genes, where the FII gene lies (Simon, 1971).

The λ gpFII homologues in bacteriophage 21, φ80, and N15 have sequence identities of 65%, 64%, and 68%, respectively, and it has previously been shown that it is the genetic source of gpFII that determines the specificity of the head-tail joining reaction (Casjens et al., 1972). While gpFII and gp8, the bacteriophage 21 homologue, are fully interchangable (Smith and Feiss, 1993), the φ80 gpFII homologue is not (Casjens, 1974). In vitro, when purified λ heads (with λ gpFII) are mixed with a 50/50 mixture of λ and φ80 tails, both types of tails are able to join to the head and form a mature virion. When φ80 gpFII is added to gpFII- λ heads, it is able to bind to the head structure. However, when this intermediate is then mixed with a 50/50 mixture of λ and φ80 tails, only φ80 tails are able to attach and complete the virion. This illustrates that the surface of gpFII in both λ and φ80 that recognizes the λ head is the same, as both are able to bind to it. However, because the φ80 gpFII is able to interact only with φ80 tails, their tail-binding surfaces must differ. As the proteins are 64% identical, it is unlikely that they bind tails with different surfaces. It is more likely that a small number of substitutions in the same area of the protein are responsible for the tail-binding differences.

To better understand how gpFII is able to discriminate between the different binding partners we solved its three dimensional structure using NMR spectroscopy. This system provides an interesting test case with which to investigate the structural determinants of protein-protein interactions. The ability of proteins to associate with multiple ligands, and to discriminate between binding partners using the same surface residues is of great interest as associations between proteins are essential to the regulation of most biological processes.
**Materials and Methods**

**Sample Preparation**

Gene FII was amplified by polymerase chain reaction from bacteriophage λ DNA. Residues 1-117 were cloned into the expression vector pET15b (Novagen) as a fusion with an N-terminal 6-His affinity tag and a thrombin cleavage site. The fusion protein was overexpressed in the E. coli strain KM1603, a derivative of BL21 (λDE3) cells with a slyD deletion, which prevents the expression of a 21-kDa histidine-rich protein which binds strongly to the nickel affinity resin used in this purification. The cells were grown at 37 °C in LB or M9 minimal media enriched with either 2.5 g/L of ^13^C-glucose, 0.7 g/L ^15^N-NH$_4$Cl, or both to an OD$_{600}$ of 1.0. Protein expression was induced by the addition of isopropyl-β-D-thiogalactopyranoside (final concentration 175 μg/mL), followed by incubation at 37 °C for an additional 3 hours. The cells were harvested and lysed in 6M GuHCl, 100 mM NaH$_2$PO$_4$, 10 mM Tris-HCl, 10 mM imidazole, pH 8.0, and purified in the same buffer via batch method using nickel-nitrilotriacetic acid-agarose resin (Qiagen). The pure proteins were eluted with 6 M GuHCl, 0.2 M acetic acid, and were refolded by dialysis into 50 mM Tris-HCl, pH 7.0, 200 mM NaCl. The 6-His tag was cleaved by incubation of the purified proteins overnight with thrombin in a cutting buffer consisting of 50 mM Tris-HCl, pH 7.0, 200 mM NaCl, and 2.5 mM CaCl$_2$. The sample was then passed over a second nickel column to remove the cleaved tag and any uncleaved protein. Gel filtration, used as a final purification step, was performed at room temperature by FPLC on a 25-mL Superdex-75 column (Pharmacia). The column was washed with 2 volumes of buffer with a 1 mL/min flow rate. A 2 mM sample of gpFII was loaded on the column and eluted at 1 mL/min, using absorbance at 280 nm to detect the protein. The labelled samples were then dialyzed into NMR buffer consisting of 150 mM NaCl and 25 mM phosphate (pH 7.0) and were concentrated by ultrafiltration to approximately 1.5 mM. Ten percent D$_2$O was added to provide NMR lock signal. All proteins were determined to be greater than 98% pure by SDS-PAGE analysis followed by Coomassie staining. The concentration of the protein samples was determined by UV absorbance at 280 nm using a molar extinction coefficient of 13,890 M$^{-1}$ cm$^{-1}$ calculated from the number of tyrosine and tryptophan residues present (Pace et al., 1995).
Determination of Thermodynamic Parameters by Circular Dichroism Spectroscopy

Thermal denaturation experiments were performed in an Aviv 62A DS circular dichroism spectrometer. Thermal denaturation was monitored by measuring the CD signal at 226 nm in a 0.1 cm cuvette. The proteins were heated from 25 °C to 109 °C in 2 °C increments, with a 1 minute equilibration time, and a 15 second averaging time for the CD measurement. Thermal denaturation experiments were fully reversible under all conditions. The data were fit to the standard thermodynamic equations for a monomeric protein obeying two-state behaviour.

In Vitro Assay for gpFII Activity

Virion extracts lacking gpFII were prepared from the lysogenic E. coli strain 594 (λFIIm423, cl857, Sam7). 1 mL of overnight culture was used to inoculate 50 mL of LB and the cells were grown to an optical density at 600 nm of 0.4. The culture was transferred to a 45 °C waterbath for 15 minutes, and then to a 37 °C incubator for 1 hour. Following incubation the cells were collected by centrifugation and resuspended in RRM buffer (16.8 mM NH₄Cl, 0.9 mM MgSO₄, 18 mM KCl, 2.7 mM FeCl₃, 44.5 mM Na₂HPO₄, 19.8 mM KH₂PO₄, 0.25 mM CaCl₂, 0.36% D-maltose, 10% glycerol, 0.1% β-mercaptoethanol, 10 mM putrescine) and 150 mL chloroform to induce lysis of the cells. Cellular debris was collected by centrifugation for 5 minutes at 15,000 rpm at 4 °C, and the aqueous phase was collected and stored at -80 °C. 20 mL of acceptor extract and purified gpFII at known concentrations were mixed together and incubated at 37 °C for one hour, then placed on ice. Serial dilutions of the reaction mixtures were titered for plaque forming units by plating on the indicator strain QD5003 (Yanofsky & Ito, 1966).

NMR Spectroscopy and Structure Calculations

All NMR spectra were collected at 25 °C on either a Varion 600-MHz or 500-MHz Inova spectrometer equipped with pulse field gradient units and actively sheilded triple-resonance probes. NMR data were processed using nmrPipe software (Delaglio et al., 1995), and were analysed using XEASY (Bartels, 1995). Sequence-specific backbone resonance assignments were attained using ¹⁵N-HSQC (Gronenborn & Clore, 1996), CBCA(CO)NH (Pascal et al., 1994), and HNCACB (Kay et al., 1994) spectra. Side chain ¹H and ¹³C resonances were assigned using HCC-TOCSY (Bax et al., 1990; Kay et al., 1993) and CCC-TOCSY (Montelione et al., 1992) spectra. Side chain ¹H resonances of aromatic residues were assigned using
homonuclear two-dimensional NOESY (Jeener et al., 1979) and TOCSY (Braunschweiler & Ernst, 1983) spectra in D$_2$O. H-H nuclear Overhauser effects (NOE) were identified using $^{15}$N, $^{13}$C-edited NOESY (Pascal et al., 1994) spectra in H$_2$O and homonuclear NOESY in D$_2$O (mixing time 150 ms). An HNHA experiment (Kay & Bax, 1990) was used to measure $^3$J$_{NH-H\alpha}$ couplings.

Structure calculations were completed using version 3.851 of XPLOR with ambiguous restraints for iterative assignment (ARIA) (Nilges et al., 1997). The initial structures input into ARIA consisted of the ten lowest energy structures calculated by XPLOR using 462 unambiguous, manually assigned NOE, dihedral angle, and hydrogen bond restraints. The NOE peak lists for ARIA were generated from manual peak picking of the NOESY spectra using XEASY. The frequency window tolerance for assigning NOEs using ARIA was ± 0.5 ppm for nitrogen and carbon, and ± 0.03 ppm for proton dimensions. The ARIA parameters $p$, $T_v$, and $N_v$ were as described (Nilges et al., 1997). Twenty structures were refined in each iteration, and the ten lowest energy structures were used for the purpose of residue specific NOE assignments. In the final iteration 50 structures were refined, and the 10 lowest energy structures were retained for analysis.

**Results**

Functional Characterization of gpFII

Active gpFII has been previously purified from two different sources: cells infected with phage mutants in the head genes, which allows the purification of gpFII that has not yet been incorporated into phage, and from phage particles that have been dissociated with GuHCl (Casjens, 1974). However, to facilitate the purification of large amounts of pure gpFII necessary for NMR spectroscopy, we cloned the gene into an expression vector that produces the full length protein with an N-terminal thrombin-deavable hexahistidine affinity tag. Since we are examining a recombinant form of the protein purified by a denaturing protocol, we characterized the activities of both the His-tagged protein, and the thrombin-cleaved protein, which has three extra residues (GlySerHis) fused to the amino terminus. The activity of gpFII can be easily determined using a complementation assay in which the purified protein is incubated with a phage extract lacking gpFII, and the number of plaque forming units determined by plating dilutions of this reaction mixture on a reporter strain. Both of these recombinant gpFII proteins exhibited identical activities, and formed $10^{10}$ phage/mL when incubated at concentrations
between 0.4 μM and 100 μM at 37 °C with the acceptor extract. These activities agree those obtained in previous experiments characterizing the activity of native gpFII purified from phage (Casjens, 1974).

**Biophysical Characterization of gpFII**

Before beginning NMR structure determination, experiments were performed to confirm that gpFII is a folded, monomeric protein. Fast protein liquid chromatography (FPLC) gel filtration experiments, where the protein eluted as a single, sharp peak, verified that gpFII is monomeric in solution at a concentration of 2 mM (Figure 4-1). To determine if gpFII possesses stably folded structural elements, we examined it using circular dichroism (CD) spectroscopy, which provides a measure of the amount of secondary structure present. Although the CD spectrum of folded gpFII is unusual, reaching a single minimum at approximately 203 nm rather than 217 nm as expected for a protein with significant amounts of β-sheet structure, and resembles that of an unfolded protein, there is a clear difference between the folded and unfolded states (Figure 4-2A). gpFII shows a fully reversible thermal denaturation curve that is independent of protein concentration, with a single transition between the folded and unfolded states (Figure 4-2C). Fitting the data using a non-linear least squares regression assuming a two-state unfolding mechanism gives a transition midpoint temperature (Tm) value of 65.8 °C.

![Figure 4-1. Size exclusion chromatography of gpFII.](image)

Figure 4-1. **Size exclusion chromatography of gpFII.** Size exclusion chromatography was performed using a Pharmacia Superdex-75 column as described in Materials and Methods. Purified gpFII (x) elutes as a single peak with a molecular mass of approximately 13 kDa. The column was standardized with the following markers (○) as indicated in the above figure; BSA (67 kDa), ovalbumin (43.5 kDa), carbonic anhydrase (29 kDa), myoglobin (16.9 kDa), abp1 SH3 domain (7.9 kDa) and insulin (3.5 kDa).
Fluorescence spectroscopy, which reports the hydrophobicity of the environment surrounding aromatic residues, can also be used as a method of assessing whether a protein is stably folded. gpFII has tryptophan residues at positions 89 and 104. Figure 4-2B shows the folded fluorescence spectra of gpFII in buffer and unfolded in 6 M GuHCl. The unfolded spectrum shows a maximum tryptophan fluorescence at 355 nm, which indicates that the tryptophan residues are exposed to a hydrophilic environment, which causes quenching of their signal. Upon folding, the absorption maximum is shifted to 342 nm, and there is an increase in the maximum fluorescence emission, implying an increase in the hydrophobicity of the environment surrounding the tryptophan residues in the native protein. This shows that the environment of the tryptophan residues changes upon folding of the protein, but the small increase in fluorescence suggests that they are not highly buried in the interior of the protein.

NMR Solution Structure

NMR data were collected for the biologically active thrombin-cleaved gpFII protein. A combination of two and three dimensional NMR experiments were used to identify and sequentially assign greater than 95% of the $^1$H, $^{13}$C, and $^{15}$N resonances of the backbone and side chains atoms for residues 1-117. The secondary structure elements were defined by several NMR criteria. Seven $\beta$-strands were predicted by negative $\Delta C\alpha$ shifts, strong $d\alpha N$ (i, i+1) NOEs and $^3$JHN coupling constants. These combined data allowed the identification of extended $\beta$-strands for residues 26-31 (strand 1), 37-40 (strand 2), 42-44 (strand 3), 64-67 (strand 4), 80-83 (strand 5), 87-91 (strand 6) and 104-108 (strand 7). A short $\alpha$-helix (residues 71-74) was identified by coincidence of strong $dNN$ (i, i+1) NOEs and by NOEs to hydrogens three and four residues distant. The $\beta$-sheets were identified by the presence of strong H$\alpha$-H$\alpha$ and weak H$\alpha$-HN and HN-HN interstrand NOEs characteristic of antiparallel $\beta$-sheet structures.

The overall topology of gpFII was determined from NMR data using XPLOR and ARIA structure-generating programs. Three types of experimental distance constraints were used as input for these calculations: (i) 1865 NOE-derived $1H-1H$ distance constraints, both manually assigned and those assigned by ARIA, (ii) 22 constraints to maintain interstrand hydrogen bonds of the $\beta$-sheets, and (iii) 54 constraints on the backbone dihedral angles using CSI, TALOS, and $^3$JHN coupling data.
Figure 4-2. Biophysical characterization of gpFII. Circular dichroism (A.) and fluorescence (B.) spectra of folded (O) and unfolded (●) gpFII. The CD spectra were obtained from a 50 mM sample in a 1 mm pathlength CD cuvette, while the fluorescence spectra were obtained from a 50 mM sample in a 1 cm pathlength fluorescence cuvette. (C.) Thermal denaturation experiments performed on 50 mM gpFII as monitored by far UV CD (226 nm).

The ensemble of 10 low energy structures calculated for gpFII is presented in Figure 4-3A, and the statistical parameters of the structure determination are summarized in Table 4-1. The calculated set of structures provided identical topologies, with a backbone root mean square deviation of 0.36 Å for the structured region of the protein composed of residues 25-45, 63-91 and 104-108. The structure of gpFII consists of a single domain, with two large disordered regions; one at the N-terminus (residues 1-24), and one in the middle of the protein (residues 46-62), both of which are characterized by no medium or long range NOEs. The seven β–strands are organized in a twisted β–sandwich structure formed by two orthogonally packed antiparallel sheets. Sheet 1 contains strands 1, 2, 5, 6, and 7, while sheet 2 is formed by strands 3 and 4, which are separated by the unstructured region from residues 46-62. The short α–helix is present between strands 4 and 5. The amphipathic β–sheets
**Table 4-1**

*Structural statistics for the 10 lowest energy simulated annealing structures*  \(^1,2\)

<table>
<thead>
<tr>
<th><strong>Unambiguos distance restraints</strong></th>
<th>1274</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intraresidue</td>
<td>425</td>
</tr>
<tr>
<td>Sequential</td>
<td>404</td>
</tr>
<tr>
<td>Medium range</td>
<td>77</td>
</tr>
<tr>
<td>Long range</td>
<td>346</td>
</tr>
<tr>
<td>Hydrogen bonds</td>
<td>22</td>
</tr>
<tr>
<td><strong>Ambiguous distance restraints</strong></td>
<td>613</td>
</tr>
<tr>
<td><strong>Total distance restraints</strong></td>
<td>1887</td>
</tr>
</tbody>
</table>

| **Dihedral angle restraints**     | 52   |

<table>
<thead>
<tr>
<th><strong>Mean r.m.s. deviations from the experimental restraints</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance (Å)</td>
<td>0.036 ± 0.003</td>
</tr>
<tr>
<td>Dihedral angle (°)</td>
<td>1.48 ± 0.325</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Mean r.m.s. deviations from idealized covalent geometry</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bond (Å)</td>
<td>0.003 ± 0.000</td>
</tr>
<tr>
<td>Angle (°)</td>
<td>0.447 ± 0.016</td>
</tr>
<tr>
<td>Improper (°)</td>
<td>0.449 ± 1.18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Mean energies (kcal mol(^{-1}))</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ENOE</td>
<td>76.1 ± 12.9</td>
</tr>
<tr>
<td>Ecdih</td>
<td>8.57 ± 4.11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Ramachandran plot</strong>  (^3)**</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>% residues in most favourable regions</td>
<td>86.4</td>
</tr>
<tr>
<td>Additional allowed regions</td>
<td>12.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Atomic r.m.s. deviations (Å)</strong>  (^3)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Backbone atoms (N, Cα, C' and O)</td>
<td>0.36</td>
</tr>
<tr>
<td>Heavy atoms</td>
<td>0.94</td>
</tr>
</tbody>
</table>

\(^1\)None of the structures exhibits distance violations greater than 0.3 Å or dihedral angle violations greater than 5°.

\(^2\)The program PROCHECK was used to assess the overall quality of the structures.

\(^3\)Residues 25-45, 63-92, 105-108
Figure 4-3. NMR solution structure of gpFII. (A) Stereo view showing the best fit superposition of the backbone of the folded nucleus (residues 25-45, 63-91, 104-108) of the ten lowest energy structures. (B) Ribbon representation of residues 23-117 of the lowest energy solution structure showing the packing of the hydrophobic core residues. The figures were generated using the molecular graphics program SETOR (Evans, 1993).
form a tightly packed hydrophobic core (Figure 4-3B) consisting of 13 residues that are greater than 94% buried (Ala 27, Ile 29, Ile 39, Gly41, Leu 65, Val 72, Leu 75, Leu 81, Ile 83, Phe 88, Val 90, Leu 103, and Leu 105). Each of the secondary structure elements contributes residues to the hydrophobic core.

Mixing Experiments with gpW and gpFII

Since gpW and gpFII likely interact in the phage head, we examined mixtures of the two proteins by NMR to determine if we could detect any interaction between the proteins in solution. Figure 4-5 shows the HSQC of 15N-labeled gpFII alone (A), and mixed with equal amounts of gpW with a hexahistidine affinity tag on the N-terminus (gpW(NT); panel B), or on the C-terminus (gpW(CT); panel C). We previously demonstrated that the gpW(NT) displayed higher in vitro activity, but lower solubility, than gpW(CT) (Maxwell et al., 2000). gpW(NT), which has the native C-terminus ending in three large hydrophobic residues, is soluble only to concentrations of 40 μM, which is unsuitable for NMR experiments. However, we postulated that the C-terminus of gpW may interact with gpFII in the phage and, if they associated in solution, possibly gpW(NT) would be more soluble due to the protection of the hydrophobic C-terminal residues from solvent. When gpW(NT) was mixed with gpFII and co-concentrated, large amounts of precipitate were formed. The aggregation of gpW(NT) caused gpFII to also come out of solution, as evidenced by the decreased intensity of the peaks in the spectrum due to a lower concentration of protein remaining in solution (Figure 4-5B). However, examination of the spectrum of gpFII remaining in solution shows no change in chemical shift for any of the peaks, indicating that there are no interactions forming between the two proteins that change the local environment of any of the gpFII backbone amides. While gpFII mixed with the gpW(CT) did not aggregate, there was no change in this HSQC either (Figure 4-5C). To ensure that we were not overlooking an interaction that involved only the side chains of gpFII, we set up the reverse experiment using 15N-labeled gpW(CT) and unlabelled gpFII (Figure 4-6). There was no change in the HSQC of gpW(CT), again suggesting that gpW and gpFII are unable to form a complex when they are in a monomeric state in solution.
Figure 4-5. NMR mixing experiments with gpFII and gpW. HSQC of (A.) 15N-labelled thrombin-cleaved gpFIL and labelled gpFII mixed with unlabelled gpW with a N-terminal hexahistidine tag (B.), or unlabelled gpW with an C-terminal hexahistidine tag.
Figure 4-6. *Mixing experiments with gpW and gpFII.* HSQC of (A.) 15N-labeled gpW with a C-terminal hexahistidine tag, and (B.) labeled gpW mixed with equimolar concentrations of unlabeled gpFII.
Discussion

Structural Features of gpFII

gpFII is a monomeric, single domain, twisted β–sandwich protein comprised of 7 β–strands and a short α–helix. A search of the protein structure database for folds similar to gpFII using the program DALI (Holm & Sander, 1993) did not identify any other proteins with similar three-dimensional structures, implying that this structure describes a new fold. The secondary structure of the protein is quite irregular, and there are a large number of disordered residues in two large unstructured regions, one at the N-terminus of the protein, and a second one from residues 46 to 62. There are several lines of evidence that confirm that the lack of restraints in the disordered regions is not merely due to experimental problems. First, although regular secondary structural elements account for only 30% of the total sequence, they are distributed over the entire primary structure of the protein, and the structured regions of the protein are very well determined with a backbone r.m.s. deviation of 0.36 Å. Second, the overall packing of the hydrophobic core of the protein is very tight, and without any internal cavities, and examination of the sequence alignment shows that the hydrophobic core positions in gpFII are occupied exclusively by hydrophobic residues in gpFII homologues from related lambdoid phages (Figure 4-4). Finally, fluorescence data which shows only a modest change between the folded and unfolded protein spectrum is consistent with the structure, where the Trp residues are partially exposed on the surface of the protein, not buried in the hydrophobic core.

An unusually large number of residues not in regular secondary structure has previously been observed in another bacteriophage λ morphogenetic protein, gpD. This 109 residue protein is made up of approximately 65% coils and loops, with several short β–strands and one α–helix. While 30% of the gpFII primary sequence is in regular secondary structure, the gpD crystal structure, solved at 1.1 Å resolution, shows that only 27% of its residues are in β–strands or α–helices. In addition, the first 14 residues in gpD are disordered, and are not visible in the crystal structure.

The CD spectrum of folded gpFII, which resembles an unfolded protein, was predictive of significant amounts of unstructured protein. A typical β–sheet protein has a folded CD spectrum characterized by a single minimum at approximately 217 nm. gpFII reaches its minimum at approximately 203 nm, which is closer to the
Figure 4. Sequence alignment of 8pII. Alignment of the amino acid sequences of 8pII and homologues

8pII homologues are shaded. The positions in the structured region of the protein that are conserved among all of the
but not in 80 and 80, Thr residues are boxed. The positions in the structured region of the protein that are conserved in 8 and 8, and the positions in the structured region that are conserved in 8 and phase 2, from phage 2, 115, and 80. The positions of the P-stretches and the a-helices are highlighted above the sequence.
minimum expected for a random coil structure. Unusual CD spectra have previously been observed in other β-sheet proteins, such as SH3 domains, which have a maximum at approximately 220 nm, instead of the characteristic minimum at 217 nm (Maxwell & Davidson, 1998). It has been suggested that the unusual CD spectra sometimes observed in β-sheet proteins may be a result of β-turn conformations (Chang et al., 1978; Shin et al., 1993), or fixed conformations of buried tyrosine side chains (Chakrabartty et al., 1993). The CD spectrum of the bacteriophage λ protein gpD also was found to resemble that of a random coil, which is consistent with the high percentage of irregular secondary structure elements found in the crystal structure (Yang et al., 2000). Although the CD spectrum of gpFII resembles an unfolded protein, the change in ellipticity upon denaturation of the protein by heat is consistent with the presence of folded structure within this protein (Figure 4-2A). This protein can be reversibly denatured by heat, and displays a Tm of 65.8 °C. It has previously been reported that gpFII is very stable as illustrated by its resistance to heat and chemical denaturation (Casjens, 1974). In fact, with a Tm of 65.8 °C, it is only moderately stable, but it is able to refold, and thus regain activity when it is returned to room temperature or dialysed from guanidine or urea.

An interesting feature of gpFII is that it is monomeric in vitro at high concentrations, yet is believed to have the ability to assemble into a stable multimeric structure when presented with the correct interaction surface provided by bacteriophage λ heads. In addition, although it is known that the incorporation of gpW is required for the subsequent activity of gpFII, NMR experiments described above show that the two monomeric proteins do not interact in vitro, even at concentrations greater than 100-fold higher than are present in the cell. The bacteriophage λ has a strictly ordered morphogenetic pathway, in which which gpW must first bind to the head-tail connector, which then allows gpFII to interact first with connector, and subsequently with tails. This is likely crucial to avoid the formation of dead-end intermediates, and suggests that there must be a change in gpFII upon binding to the head that allows it to bind to other phage proteins. This change may be the juxtaposition of gpFII with some other protein in the head, or a conformational change in the secondary or tertiary structure of gpFII itself. One method that gpFII may use to control these protein-protein interactions is by folding the large number of residues that are unstructured in the monomeric state into an ordered array only when the correct target surface is contacted. These newly folded regions could provide a binding site for the next protein in the morphogenetic pathway. For example, the amino terminus of gpFII, which is unstructured in the
monomeric state, has high helical propensity as predicted by PHD (Rost & Sander, 1994), and could form a helix in the completed phage. This type of phenomenon has previously been observed in the bacteriophage λ protein N, which is responsible for activating the transcription of the delayed early genes, and the P22 connector protein. NMR and circular dichroism spectroscopic analysis of the 107 residue N protein are consistent with the existence of N in solution as an essentially unstructured molecule containing a small amount of fluctuating and transient secondary structure (Van Gilst et al., 1997). When N protein is mixed with boxB RNA, its cognate RNA target, the structure of the protein changes from approximately 18% to 35% helical, corresponding to the formation of about 16-18 amino acids of α-helical structure. The P22 connector protein also undergoes a significant conformational change when it is assembled into the 12-membered ring that is present in the phage head. The structure of the connector protein is highly α-helical in its monomeric and ring form (Moore & Prevelige, 2001). However, there is a significant increase in the amount of α-helix in the ring structure, with the protein changing from 41% to 49% helix upon ring formation. CD data shows that this increase in α-helix, which corresponds to approximately 60 amino acids, is accompanied primarily by a loss in random coil.

The Genetic Source of gpFII Determines the Tail Specificity

Since λ gpFII is fully interchangeable with its phage 21 but not its φ80 homologue, sequence analysis may provide insight into the surface(s) of the protein that interact with the various phage components. Both phage 21 and φ80 gpFII proteins can bind to λ heads (and vice versa), thus, the amino acid side chains that are responsible for these interactions would be expected to be conserved between the three homologues. Conversely, because the φ80 gpFII protein is unable to interact with λ tails, while phage 21 gpFII is able to exchange fully with the λ homologue, exposed amino acid positions that are conserved between λ and phage 21, but not φ80, may identify the surface that is responsible for tail specificity.

A sequence alignment (Figure 4-4) of gpFII and homologues from phage 21, N15 and φ80 shows that the proteins have 65%, 68%, and 64% sequence identity, respectively. Examination of the sequence alignment reveals 7 strictly conserved residues, 6 hydrophobic and one Arg, that are exposed on the surface of the protein (Figure 4-4, shaded residues). Since the energy needed for protein-protein interactions is provided primarily by the burial of solvent-exposed hydrophobic residues, and Arg residues were found to make up 10% of the amino acid
population in protein interfaces in a large study of crystal structures (Lo Conte et al., 1999), these residues are more likely than polar residues to play an important role in forming interactions with other phage components, such as gpB and/or gpW in the connector. Figure 4-7 (green residues) shows the positioning of these amino acids in gpFII. Intriguingly, all of these residues are located in one face of the protein. The amino acids are spread along a cleft in the protein, and provide an extended hydrophobic pocket that would be ideal for binding to other proteins.

The sequence alignment also reveals six surface-exposed positions in λ gpFII at which the bacteriophage 21 homologue has the same amino acid, while the φ80 homologue has a non-conservative substitution (Figure 4-4, boxed residues). Of the six positions, two involve polar or glycine residue substitutions with an amino acid with significant hydrophobic character (Gln34Leu and Gly78Leu), while the remaining three result in differences in the charge dispersion on the surface (Ser26Glu, Asp70Ser, Arg76Glu, and Glu86Arg). The locations of these six positions are also displayed in Figure 4-7 (red residues). Satisfyingly, these residues are also clustered on one face of the protein structure, separate from the conserved residues. This suggests that these two regions may be the important regions for making protein interactions with the connector (conserved green residues), and the tail (non-conserved red residues).

Conclusion

We have shown that gpFII possesses a novel fold consisting of 7 β-strands and a short α-helix. The large amount of unstructured protein present in two regions, and the presence of conserved residues within these unstructured regions, suggests that they may play an important role in regulating the assembly of gpFII. The structure has also allowed us to identify two potential protein interaction interfaces on the surface of gpFII, which may provide binding sites for the bacteriophage λ head and tail. These positions suggest possible targets for site-directed mutagenesis, and other studies to determine the importance of these residues.
Figure 4-7. Analysis of the gpFII structure reveals two putative binding surfaces. Hydrophobic residues that are conserved in all gpFII homologues are shown in green. Residues that are conserved between λ and phage 21, but have a non-conservative substitution in φ80, and thus many play an important role in tail specificity, are shown in red. All panels are rotated with respect to panel A; panel B is rotated 180 degrees about the y-axis, panel C is rotated +90 degrees about the x-axis, and panel D is rotated -90 degrees about the x-axis. The N- and C-termini are labelled.
References


Thesis Summary

In this thesis I have examined the head-tail joining reaction of the bacteriophage \( \lambda \) morphogenetic pathway in atomic detail. Using a number of biophysical techniques, including circular dichroism and fluorescence spectroscopy, I have thermodynamically characterized gpW and gpFII, the proteins that carry out this assembly step. In addition, I solved their three dimensional structures using NMR spectroscopy.

The combination of biophysical and mutagenesis studies was a powerful tool in helping to characterize the mechanism by which gpW operates. The key to interpreting the mutagenesis results was the CD analysis of gpW, which enabled us to differentiate properly folded, stable mutants from mutants with major disruptions in their three dimensional structure. The mutational studies revealed that the C-terminus, specifically the last three residues, of gpW is crucial for the activity of the protein. The last three positions were individually substituted with several different amino acids and an amber codon, and the stabilities and activities of the resulting 12 proteins were determined. Each of the mutant proteins displayed the same CD spectrum as the wild type protein and retained wild type thermal stability, indicating that there were no significant structural rearrangements taking place. Surprisingly, although stability was unaffected, even relatively conservative amino acid substitutions at these positions resulted in a large decrease in the \textit{in vitro} activity of the protein. Non-conservative substitutions at these positions, for example Val68Glu, were found to totally abrogate activity, as was an amber mutation at the second last position.

To further probe the characteristics of gpW I solved its three-dimensional structure using NMR spectroscopy. The protein fold consists of a two-stranded, antiparallel \( \beta \)-sheet that is packed against two \( \alpha \)-helices, which surround a single, well-packed hydrophobic core. The 10 C-terminal residues of gpW were found to be disordered. This is consistent with both the results from the thermodynamic characterization of the wild type gpW, where it was discovered that the \( \Delta C_p \) was considerably lower than expected for a 68 residue protein, and the mutagenesis experiments which showed that the thermodynamic stability of gpW was unaffected by mutations at the C-terminus. Although the structure of gpW appears to be a novel fold, the \( \alpha \)-helices show similarity in both length and packing angle to the \( \alpha \)-helices found in the sliding clamp proteins that are responsible for the processivity of many DNA polymerases. Using the human PCNA sliding clamp structure as a
template, gpW was modelled into a hexameric ring which is consistent with its known roles in bacteriophage λ morphogenesis.

We postulated that the upper surface of the hexameric ring of gpW might interact with the connector region of the phage head, leaving the C-termini of the six gpW monomers free to interact with gpFII, the next component in the bacteriophage assembly pathway. The disordered structure of the C-terminus of the gpW monomer, combined with its functionally critical nature, is reminiscent of proteins bound by PDZ domains in eukaryotic cells. We postulated that a PDZ-like interaction might be important in the function of gpW, and that gpFII might provide this interaction. Thus, we decided to solve the three dimensional structure of gpFII by NMR spectroscopy. gpFII is a monomer composed of 7 β-strands arranged in two orthogonally-packed β-sheets, and a short α-helix. Although gpFII is thought to interact with gpW in the phage particle, we were unable to detect any interaction between the two proteins in solution. The most striking feature of the gpFII structure is the unusually large number of unstructured residues that are present in two regions, one from residues 1-23, and the other from residues 42-62. We postulate that these large unstructured regions may play an important role in gpFII assembly into the virion, by modulating its activity through a disorder-to-order transition.

**Future Directions**

I have examined the head-tail joining reaction of the bacteriophage λ in hopes of revealing generally applicable rules that apply to the assembly of macromolecular complexes. During λ morphogenesis a DNA-filled head binds to gpW, then gpFII binds, and then tails are able to bind to complete the virion. This process sounds simple, but gpW and gpFII do not interact in solution on their own. Does gpW change conformation upon binding to the head, and thus become competent to bind gpFII, or does the multimerization of gpW on the connector create the gpFII binding site, or does gpFII interact with the connector and gpW simultaneously, or do gpW and gpFII bind to entirely different sites? There are a number of experiments, outlined below, that may help answer these types of questions, and in the process also help elucidate some general principles that govern macromolecular interactions.
Identify and Characterize the Phage Proteins Bound Directly by gpW and gpFII

Both gpW and gpFII are likely to possess multiple protein-protein interaction interfaces that allow them to assemble onto the phage head and, in turn, become the sites of assembly for other phage components. To elucidate the surfaces of these proteins that are responsible for specific protein-protein interactions, one could use a combination of chemical cross-linking and site-directed mutagenesis experiments. To perform the cross-linking experiments, \( \lambda \) lysogens with amber mutations in either gene W or gene FII could be induced in the presence of plasmids expressing fully functional 6-His tagged variants of gpW or gpFII. This will produce viral particles that have incorporated the 6-His tagged proteins, which we know from previous in vitro experiments do not interfere with assembly. Ni-affinity chromatography could then be used to purify the phage particles. The same experiment can also be performed with \( \lambda \) lysogens that also carry mutations in the tail genes, creating phage heads with only tagged gpW, or tagged gpW and untagged gpFII, or untagged gpW and tagged gpFII.

The whole phage (or phage heads) containing the 6-His tagged proteins could then be treated with a reversible cross-linking agent, such as SPDP. This is a heterobifunctional compound that cross-links cysteine residues to nearby primary amine groups (lysine residues and protein N-termini). After cross-linking, the particles can be dissociated by treatment with \( \text{GuHCl} \), and Ni-affinity chromatography used to isolate the 6-His tagged protein along with other proteins to which it has become cross-linked. Once these complexes have been isolated, the cross-linking reaction can be reversed by the addition of a reducing agent such as DTT. SDS-PAGE analysis can then be used to identify phage proteins that were cross-linked to the 6-His tagged proteins. Purified phage head and tail proteins can be run on the gel as standards, making the identification of the bands straightforward.

Since gpW has no cysteine residues, and gpFII has only one, this protocol would provide a means to identify proteins that are close to the surface of the protein being tested. One could determine which phage proteins are interacting with different surfaces of gpW and gpFII by replacing single residues with cysteine at several positions in each protein. This approach has been used to map protein-protein interaction interfaces in a number of other systems (Hansen & Barklis, 1995; Herzig et al., 1996; Lee et al., 1996; Villarreal & Lee, 1998). The amino acid positions in gpW and gpFII that could be changed to cysteine should be exposed on the protein surface, and dispersed on different faces of the protein. Both the thermodynamic
stability, as monitored by CD, and biological activity, as measured by the in vitro assembly reaction, should be tested to ensure that the substitutions are not highly disruptive. The single cysteine in gpFII could be substituted with serine, threonine, or alanine (whichever is not disruptive) to simplify the interpretation of cysteine substitutions at other positions in the protein. Fortunately, besides gpFII, gpB is the only protein in the head-tail joining region that contains cysteine residues. Thus, the cross-linked proteins that are observed in these experiments should only involve the proteins of interest.

Investigate the 3-D Structures of Proteins at the Head-Tail Junction

It would be of significant interest to solve the atomic resolution structure of other proteins located at the head-tail junction, such as gpU, gpZ, and gpB. gpU is small enough to solve in its monomeric form using NMR spectroscopy, and we have already shown that it is amenable to structure determination by this method (Figure 5-1). Due to their large molecular weights, the other two proteins will need to be solved by X-ray crystallography.

An interesting aspect of these structural studies would be that one may be able to solve these proteins in both the monomeric form and the multimeric form in which they exist in the phage particle. For example, gpB requires the λ scaffolding protein, gpNu3, to assemble into the dodecameric connector structure (Kochan et al., 1984). If gpB were expressed on its own, it would likely remain monomeric, but if it were co-expressed with gpNu3, sufficient amounts of dodecameric material may be produced to solve the structure of the connector ring. In addition, gpU, which could be solved by NMR in its monomeric form, can be made into a hexameric ring structure simply by the addition of Mg²⁺ ions (Katsura & Tsugita, 1977). By comparing the monomeric and multimeric structures of these proteins, one could assess the interactions and conformational changes that are required for their assembly.
Figure 5-1. HSQC of the monomeric form of gpL, the bacteriophage λ tail-terminator protein.
References


