UNIVERSITY OF CALIFORNIA, SAN DIEGO

Single Molecular and Computational Characterizations of Functional Protein Interactions

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Chemistry

by

Amy Davenport Migliori

Committee in charge:

Professor Douglas E Smith, Chair
Professor Simpson Joseph, Co-chair
Professor Gaurav Arya
Professor Michael Burkart
Professor Hector Viadiu

2014

The Dissertation of Amy Davenport Migliori is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Co-chair

_____

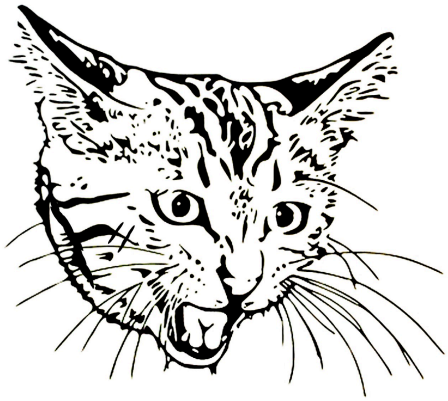
Chair

University of California, San Diego

2014

DEDICATION

This thesis is dedicated to my husband, Ben Migliori, for being my steadier half, always there to cheer me up on a gloomy day, take me for a date shake, indulge me in my incessant singalongs, and for being a genuine equal partner in my success, as I am in his. It is dedicated to my parents, Beth and William Davenport, for instilling in me a love of learning and the sincere curiosity required to undertake my graduate studies. And last, but certainly not least, I dedicate my thesis to the most influential teacher I have ever had : Mr Wood. By creating an environment of acceptance and curiosity, Mr Wood allowed for me to discover my passion for the deeper understanding of biological processes that I build my research around.



EPIGRAPH

Everything must be made as simple as possible. But not simpler.

- Albert Einstein

One never notices what has been done; one can only see what remains to be done.

- Marie Curie

Now. You're looking at now, sir. Everything that happens now, is happening now. What happened to then?

We passed then.

When?

Just now. We're at now now.

Go back to then!

When? Now. **Now?** Now.

I can't. Why?

We missed it.

When? **Just now.**

When will then be now?

Soon.

-Spaceballs

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge the strong support of Dr. Doug Smith and Dr. Gaurav Arya. You have played a major role in helping me to proceed with my research in directions forming me into the scientist that I am, and for that I am eternally grateful. Without your support and guidance, this work would have been impossible.

I would also like to thank my co-chair, Dr Simpson Joseph, and my thesis committee, Dr Hector Viadiu and Dr Michael Burkart for being readily available and willing to give input on the various projects, and to help me view things from different perspectives when needed.

Chapter 2 is adapted from material submitted for publication as of December 20, 2013. The dissertation author was the principal investigator and author of this paper. The author would also like to acknowledge coauthors Nicholas Keller, Tanfis I. Alam, Marthndan Mahalingam, Venaglia B. Rao, Gaurav Arya, and Douglas E. Smith.

Chapters 3 and 4 will be adapted for publication following the author's defense. The author would also like to acknowledge coauthors Gaurav Arya, and Douglas E. Smith.

Chapter 6 is adapted in full from material as published in SPIE Proceedings, 2013. The dissertation author was the principal investigator and author of this paper. The author would also like to acknowledge coauthors Samuel Kung, Danielle Wang, and Douglas E. Smith.

VITA

2006 Bachelor of Science, Biology, University of Massachusetts at Amherst

2008 Master of Science, Biochemistry, University of California, San Diego

2010 Candidate of Philosophy, University of California, San Diego

2014 Ph.D., Chemistry, University of California, San Diego

2003 - 2006 Undergraduate Research Assistant, University of Massachusetts at Amherst

2005 Summer Research Intern, Synta Pharmaceuticals, Bedford Massachusetts

2006 - 2013 Graduate Research Assistant, University of California, San Diego

FIELDS OF STUDY

Major Field: Biochemistry

Studies in Biochemistry:	Professor Elizabeth Komives
	Professor Hector Viadiu
Studies in Physics:	Professor Douglas Smith
Studies in Nanoengineering:	Professor Gaurav Arya

PUBLICATIONS

Migliori, Amy, Nicholas Keller, Tanfis Alam, Marthandan Mahalingam, Venigalla Rao, Gaurav Arya, Douglas E Smith. “Structural basis for force generation by the bacteriophage T4 viral DNA packaging motor” Submitted Dec 20, 2013.

Migliori, Amy, Samuel Kung, Danielle Wang, Douglas E. Smith. “Direct measurement of formation of loops in DNA by a human tumor suppressor protein” *Proc. of SPIE* **8810** 881030-1 (2013).

Keller, Nicholas, Amy Migliori, Gaurav Arya, Venigalla B Rao, Douglas E. Smith. “Testing a structural model for viral DNA packaging motor function by optical tweezers measurements, site directed mutagenesis, and molecular dynamics simulations” *Proc. of SPIE* **8810** 881032-1 (2013).

Migliori, Amy, Gaurav Arya, Douglas E. Smith. “Experimental and computational studies on the DNA translocation mechanism of the T4 viral packaging motor” *Proc. of SPIE* **8458** 84582N-2 (2012).

ABSTRACT OF THE DISSERTATION

Single Molecular and Computational Characterizations of Functional Protein Interactions

by

Amy Davenport Migliori

Doctor of Philosophy in Chemistry

University of California, San Diego, 2014

Professor Douglas Smith, Chair
Professor Simpson Joseph, Co-Chair

Often, proteins are studied using bulk techniques, in which the average properties of large ensembles of molecules are studied. Although this has led to substantial new knowledge, the use of single-molecule biophysics and computational techniques to understand individual molecular actions and dynamics and the role of

each residue allows for a more complete picture of activity in some cases. We present two applications of such techniques, first to determine the relationship between structure and function of the DNA translocation motor gp17 in Bacteriophage T4, and second to study looping of DNA mediated by the tumor suppressor protein p53. Specifically, we studied the role of an interface between the N- and C-terminal subdomains in generation of the high packaging forces and translocation velocities using a dual-beam optical trap. Mutation of charged amino acids located within this interface region confirmed that electrostatic forces play a role in force and velocity generation, with mutants showing a reduction in forward velocity, average velocity, and percentage of time spent packaging at different applied forces. To explain these experimental results, we generated a two-state computational model to calculate the free energy of the translocation step in gp17. We found excellent correlation between experimental data and calculated free energy change of translocation. Decomposition of the free energy change allowed for the identification of key residues involved in gp17-mediated packaging, and the role of each was explained. These results reveal that the power stroke of the motor requires substantial contributions from charged residues, hydrophobic residues, and polar residues instead of charged residues alone. Finally, we propose that several of these key residues may be hot spot residues, contributing a significant portion of the free energy used to package DNA. p53-mediated loop formation in DNA was studied using direct measurement with a dual optical tweezer setup. Looping of DNA by p53 has previously been demonstrated qualitatively using cryo-electron microscopy as well as transcription assays. we demonstrated formation of loops in purified human Col18A promoter containing five

p53 binding sites. This looping may be directly related to p53 activity at transcriptional start sites.

I Introduction

In the study of small molecules, a near-ubiquitous problem arises: the manipulation and measurement of single molecules. This difficulty has spurred the rise of the successful field of bulk biophysics, in which large ensembles of molecules are studied *en masse* and individual properties are inferred. However, when it comes to understanding the minute dynamics that occur during the actions of individual proteins, or how individual atoms influence the behavior of a molecular mechanism, single molecules must be examined alone. In this dissertation, I will demonstrate a series of results that arise from *single molecule biophysics* techniques, in which individual molecules are probed and examined. I accomplish this using a set of tools from both experimental physics, and from computational analysis. The overarching goal is to better understand certain key functional interactions made by two individual proteins. I apply these techniques to two proteins to demonstrate the power of this approach; one, an extremely fast viral DNA packaging motor, capable of generating forces many times that of eukaryotic molecular motors[1], known as gp17 and found in the bacteriophage T4. I also apply this approach to study a potent regulator of the eukaryotic cell cycle found to be mutated in up to 50% of human cancers, the widely known protein p53. This dissertation consists of two parts: first, work we have done to understand the method by which the double-stranded DNA (dsDNA) translocation in Bacteriophage T4 generates high forces and velocities[1], and second an investigation of the little-studied but extremely important protein-protein interaction regulated by p53 that is crucial to full transcriptional control of some genes by loop formation and may be involved in human disease[2]. The common threads linking these two seemingly disparate topics are the tools of single molecule biophysics. Both are studied using a dual optical trap

(also known as an optical tweezer) that allows for direct measurement of forces and dynamics[3]. In the case of gp17, I take the subsequent step of performing computational simulations of the atomic behavior in order to understand the single protein behaviors with finer detail.

Both gp17 and p53 are also involved in forming interactions with DNA and other proteins. In the case of gp17, the T4 dsDNA packaging motor, an interface between two subdomains within the protein forms to generate a ratchet-like motion that pushes DNA into the viral capsid[4]. p53, a multifunctional protein involved in regulation of the cell cycle, interacts directly with DNA to cause changes in gene expression, but evidence also shows that a p53 tetramer bound to DNA forms interactions with a second tetramer to allow for looping of the intervening DNA[5,6]. Understanding the basis for these crucial interactions helps us to better understand how these proteins perform their vital functions. An understanding of how these two proteins interact with DNA opens many possibilities. gp17 presents a potential method for moving DNA rapidly and precisely, which could find utility in next-generation DNA sequencing tools, while understanding p53 may lead to therapeutic interventions for cancers. Further, the single-molecule analysis of these proteins' interactions with DNA will help in developing a more general understanding of protein-protein interfaces, a commonly encountered phenomena in biochemistry.

1.1 The gp17 molecular translocation motor

The mechanism of viral infectivity has evolved to suit the cell type being infected, and the defenses used by the cells have evolved to match. Bacteria have developed a hard cell wall to protect their cellular membrane. For bacteriophages,

viruses that infect bacterial cells, these barriers present a challenge. The phages have developed a unique solution. Rather than invading the host cell through the membrane, as many eukaryotic viruses do, tailed bacteriophages circumvent the problem of a rigid cell wall by injecting their genomic material through the surface of the cell. Densely packaged DNA is expelled from the viral capsid through the portal channel and tail tube into the host cell. Bacteriophage T4 infects *E. Coli* cells in this manner, following surface recognition of a specific receptor and subsequent digestion of the peptidoglycan layer by a tail lysozyme[7]. Upon injection of the genome, the viral DNA hijacks the bacterial machinery to replicate itself, and in the case of T4 produces an average of 200 mature viral particles per infection.

In order to accomplish this powerful attack, bacteriophage T4 needs to build a weapon. Self-assembly of the T4 procapsid occurs around a connector protein that contains a pore through which DNA can pass. After the procapsid is assembled, the motor protein gp17 self-assembles into a pentamer on the portal protein through a helix-loop-helix motif[8]. This ATP-dependent motor protein mediates the genome packaging process by translocating DNA at up to 2000 bp/s and at forces in excess of 60 pN[9]. One the motor is formed, it packages the 171 kilo-basepair (kbp) genome into the capsid at a pressure of over 700 psi - the equivalent of 10x the pressure within a bottle of champagne! By generating such a high pressure, T4 ensures that upon penetration of the cell wall, the entire genome will be injected into the bacteria. The gp17 motor presents a unique opportunity to study a highly evolved and powerful piece of cellular machinery.

Previous studies of the gp17 packaging motor as well as similar motors from phages λ and Φ29 focused on the activity and properties of wild type motors; structural

interrogation based on sequence alignment to known functional motifs showed a conserved ATPase that indirectly changed packaging dynamics in phage λ when mutated[10,11]. Though the ATP binding motif could be studied due to the highly conserved nature of the hydrolysis pocket, further study of these motors was limited due to lack of a molecular structure. The publication of a high-resolution (2.8 Å features) X-ray crystallographic structure of monomeric gp17 substantially broadened the possibilities for functional interrogation of the motor structure[4,12]. The structure of gp17 shows a large N-terminal subdomain containing the ATP hydrolysis pocket, and a C-terminal subdomain connected to the N- via a short, flexible linker termed the hinge. The N- and C-terminal subdomains form an interface between the faces, burying 1000 $Å^2$ of protein surface area with many of the same properties as a protein-protein interface. In addition to this high-resolution structure of monomeric gp17, a second structure of pentameric gp17 complexed with the capsid and stalled during the act of DNA packaging was generated using cryo-electron microscopy. This second structure is much lower resolution at 32 Å but contains information about the physiological positioning of gp17 not available in the x-ray structure. Fitting the high resolution x-ray structure into the electron density of the cryo-EM structure necessitated that gp17 be extended by 7 Å to fit. These two conformations, compact gp17 from the crystal structure, and an extended gp17 from fitting the cryo-EM electron density, implied a mechanism in which these two structures alternated, translocating DNA by 7 Å (2 bp) with each compaction[4]. Further, it was proposed that attraction between charged amino acids on opposite sides of the interface between the gp17 subdomains cause compaction. We tested the role of these charged amino acids in packaging by using a dual optical trap setup to measure the effect of mutations to these charged amino

acids, as well as computational studies to interrogate the structure-function relationship in more detail. This allowed for us to expand on the original electrostatic hypothesis. In the case of gp17, the protein-protein interface being probed is always present (as opposed to separate proteins in the cellular milieu), as it is formed within a single protein. The presence of a hinge allows for the structural change necessary to translocate DNA. By utilizing single molecule measurements of this motor protein, we have revealed that the mechanism behind the ratcheting action is more complicated than initially expected, and relies equally on electrostatic interactions and hydrophobic interactions. Even more surprisingly, only two of the five proposed charged-pair interactions contribute significantly to the motor dynamics by forming ion pairing interactions! This result reinforces the need for careful application of single molecule techniques in the study of protein interactions. It also justifies the application of these techniques to the study of disconnected proteins, such as transcription factors. In the case of transcription factors, proteins need to interface with each other at specific times and at specific locations. We focused our research on the tumor suppressor protein, p53.

1.2 Why study p53?

Proteins form interactions with DNA for many different purposes. Histones wrap DNA into a fiber to compact and control access to genetic material. Molecular motors, such as gp17, make contacts with DNA in order to translocate along its length. Transcription factors bind to DNA to uniquely influence the transcription of gene targets. In order for cells to respond to their environments, intricate patterns of gene and protein expression are formed corresponding to the needs of the cell. Transcription

factors may recruit the machinery necessary to transcribe a gene, or block the start site to cause a gene to be silenced or down-regulated. In some cases, transcription factors can bind sites thousands of base pairs away, thus influencing overall gene expression. In these situations, the intervening DNA is “looped out” as the two distal sites are brought together, causing the correct contacts to form near the start site of the gene. Optical tweezers have been used to study loops in DNA mediated by restriction endonucleases that require two-site binding[13]. Manipulations with optical tweezers allow measurement of loop rupture force (which is related to the binding constant of the loop) and loop size. The direct observation of loop formation between transcription factors allows for a better understanding of the mechanism by which long-distance binders can influence gene expression. p53 is an excellent candidate for studying with optical tweezers because previous studies showed the presence of loops in DNA and implied the importance of these loops in terms of gene expression, but no direct measurements of these loops had been made.

p53 was discovered some 35 years ago as a binding partner of a protein associated with a tumor virus in monkeys[14,15]. Few proteins have been studied as extensively as this transcription factor and cell cycle regulator. p53 has been termed the guardian of the genome[16], and was even named the protein of the year by Science magazine in 1993. Although originally implicated in cancer formation, we now know that p53[17] functions as a tumor suppressor that is found to be mutated in 50% of human cancers[18]. p53 is activated by stresses within a cell and functions to halt the cell cycle such that damage to DNA can be fixed, and can cause cell death or simply stop division indefinitely if the damage is too severe. p53 binds, as two dimers that tetramerize, directly to DNA to influence gene expression via two 10-bp recognition

sequences known as half-sites[19], leading to either expression or repression of a gene. In some cases, p53 binding sites (also known as p53RE for recognition elements) thousands of bases away strongly influences the level of gene expression[20]. Removal of these distal sites has been shown to lead to drastic reduction in mRNA produced[21,22]. This reduction implies that interaction formed by both distal and proximal sites with transcriptional machinery are needed for proper gene expression, an interaction that could occur via distal sites looping back to contact the machinery at the start site. The loops mediated by p53 binding have been observed directly in cryo-electron microscopic images[5] showing stacking of p53 molecules at loop junctions. The structural basis for loop formation by p53 has been proposed to involve the formation of an alternate tetramer allowing for contact of different strands of DNA by the two dimers, but this is controversial[23]. The ability to generate select DNA sequences to probe how p53 forms loops and to measure looping quantitatively using molecular tweezers may lead to a better understanding of this important aspect of gene regulation by this important protein. It also demonstrates the power of single molecule biophysical techniques for studying therapeutically relevant proteins.

1.3 Experimental single-molecule biophysics via optical traps

Single molecule biophysical techniques allow for the observation of properties of dynamic systems that are not always available in bulk studies. To study DNA packaging by gp17 and to visualize the dynamics inherent to the motor[9], as well as to directly visualize the force at which loops in DNA formed by p53 rupture, we use optical tweezers[13,24]. Using focused light, optical tweezers have the ability to trap particles and cells between 50 nm and 10 μm. By moving the focus position (or the sample itself),

manipulation and movement of the trapped particles is possible. When two laser foci are created, interactions between two particles can be measured. Optical traps have been used in a wide array of applications. They have been utilized in the tensile properties of cellular structural elements[25], molecular motor activity[26], the role of riboswitches[27], and the looping of DNA mediated by protein binding[13]. The study of molecular motors has benefitted greatly from the use of optical tweezers, allowing for researchers to study the variability[9] and step size of motors[28], dynamics of RNA transcription[29], and even to uncover the mechanochemical cycles governing motor activity[30].

Our dual optical trap experimental setup utilizes a 3-channeled flow cell within which we perform experiments between two focused laser beams. DNA is attached prior to the experiment to a micron-sized polystyrene bead (the particle which is caught in the trap) and flowed through the top channel into the main channel containing experimental buffer. A second polystyrene bead, with a complementary attachment for the other end of DNA, or with a motor assembly attached, is flowed through the bottom channel into the center. The two focused laser beams are able to trap and manipulate the two beads; by tracking the position of one of the beams, we are able to discern position and deflection of the beam caused by the force induced by the viral motor. With proper calibration[31], we are able to convert this information to packaging force and velocity measurements in base pairs per second. To study p53, sections of DNA containing p53 binding sites are bound to polystyrene beads at either end, and allowed to loop. By measuring the force required to pull these loops apart, and by examining the length changes upon loop rupture, we are able verify that p53 does, in fact, form loops in DNA, and begin to build a more detailed picture of how p53 forms these loops.

The dual optical trap is a powerful tool for the study of molecular motors, DNA-binding proteins, and many other dynamic properties of biological systems. However, to collect information about the atomic structural details underlying these properties, we need even finer measurements then can be accomplished with focused light. Molecular dynamics simulations allow for the details underlying activity to be probed and understood in great detail. The combination of single molecule studies and molecular dynamics, when available, allows for better explanation of the relationship between structure and function than either technique is able to alone.

1.4 Experimental single-molecule biophysics via molecular dynamics

Formation of the three-dimensional structure of a protein is crucial in determining its function. The most common method determining the high-resolution structure of proteins is X-ray crystallography, which uses the diffraction of X-rays by the crystalized atoms to determine their positions. If a protein is sufficiently immobilized within a crystal, the resulting structure is able to determine the three dimensional position of all atoms making up the structure. Although originally it was assumed that proteins had static structures, especially within the core of the protein, it is now known that the amino acids making up a protein structure have substantial structural flexibility even when fully folded. This dynamic information is significantly obscured in the determination of a crystal structure. Further, crystal structures, although extremely useful for initial information about the structure of a protein, suffer from the possibilities of artifacts from packing protein into a crystal (a non physiological conformation in most situations). For proteins that are also enzymes or motors, a crystal structure is a

starting point from which investigations of dynamic motion can be started, but it cannot be the end point. However, imaging on the atomic scale for large systems is not yet possible and it is here that computational methods take over.

Because the interactions between atoms are governed by known physical forces, a macromolecule can be described as a collection of equations governing the connectivity and motion between atoms, stiffness of bonds, interaction with medium, electromagnetic forces, and other physical phenomena. Molecular dynamics simulations solve the equations of motion for each atom in stepwise manner over time, and in doing so over many time steps develop a picture of how the protein moves. The time series formed by these steps is called a simulation. The first step of a simulation need only contain the initial position of each atom, initial velocities, and the acceleration. With these three pieces of information, a continuous trajectory can be calculated, and from this, it is possible to calculate the change in potential and kinetic energy of the system over time. This allows for the calculation of binding energy, change in energy of the system with enforced changes, the energies of folding or unfolding, and more. A number of software suites have been developed to streamline this process and make molecular dynamics and the calculation of free energy readily available for the study of important and interesting biological questions. Simulations can be used to recreate physiological events such as binding events between protein and ligand[32], water molecules passing through an transmembrane pore[33], the movement of a tRNA into a ribosome[34], the folding of a protein[35], or any other system for which a structural model can be created and adequate computing power can be attained. This flexibility in application has been fantastically useful in many fields, and we have used molecular dynamics for our purpose of better explaining the residue-

level workings of the molecular motor gp17. By unifying results from structural, experimental, and computational investigations, it is possible to create a much more complete picture of how an individual protein functions in its natural environment. Although I present results for gp17 and p53 specifically, the general methods presented in this dissertation will hopefully provide guidance to anyone studying the interactions of dynamic, active molecules in any natural environment.

References

1. Fuller, D. N., Raymer, D. M., Kottadiel, V. I., Rao, V. B. & Smith, D. E. Single phage T4 DNA packaging motors exhibit large force generation, high velocity, and dynamic variability. *Proc. Natl. Acad. Sci. U.S.A.* **104,** 16868–16873 (2007).

2. Miled, C. A Genomic Map of p53 Binding Sites Identifies Novel p53 Targets Involved in an Apoptotic Network. *Cancer Research* **65,** 5096–5104 (2005).

3. Smith, S. B., Cui, Y. & Bustamante, C. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science* **271,** 795–799 (1996).

4. Sun, S., Kiran Kondabagil, Bonnie Draper, Tanfis I. Alam, Valorie D. Bowman, Zhihong Zhang, Shylaja Hegde, Andrei Fokine, Michael G. Rossmann, and Venigalla B. Rao. The Structure of the Phage T4 DNA Packaging Motor Suggests a Mechanism Dependent on Electrostatic Forces. *Cell* **135,** 1251–1262 (2008).

5. Stenger, J. E., Peter Tegtmeyer, Gregory A. Mayr, Michael Reed, Yun Wang, Pin Wang, Paul V.C.Hough, and Iris A. Mastrangelo. p53 oligomerization and DNA looping are linked with transcriptional activation. *The EMBO Journal* **13,** 6011–6020 (1994).

6. Jackson, P. Iris Mastrangelo, Michael Reed, Peter Tegtmeyer, Gina Yardley and Julie Barrett. Synergistic transcriptional activation of the MCK promoter by p53: tetramers link separated DNA response elements by DNA looping. *Oncogene* **16,** 283–292 (1998).

7. Rossmann, M. G., Mesyanzhinov, V. V., Arisaka, F. & Leiman, P. G. The bacteriophage T4 DNA injection machine. *Current Opinion in Structural Biology* **14,** 171–180 (2004).

8. Hegde, S., Padilla-Sanchez, V., Draper, B. & Rao, V. B. Portal-large terminase interactions of the bacteriophage T4 DNA packaging machine implicate a molecular lever mechanism for coupling ATPase to DNA translocation. *J. Virol.* **86,** 4046–4057 (2012).

9. Fuller, D. N., Raymer, D. M., Kottadiel, V. I., Rao, V. B. & Smith, D. E. Single phage T4 DNA packaging motors exhibit large force generation, high velocity, and dynamic variability. *Proc. Natl. Acad. Sci. U.S.A.* **104,** 16868 (2007).

10. Tsay, J. M., Sippy, J., Feiss, M. & Smith, D. E. The Q motif of a viral packaging motor governs its force generation and communicates ATP recognition to DNA interaction. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 14355–14360 (2009).

11. Tsay, J. M. Jean Sippy, Damian delToro, Benjamin T. Andrews, Bonnie Draper, Venigalla Rao, Carlos E. Catalano, Michael Feiss and Douglas E. Smith. Mutations altering a structurally conserved loop-helix-loop region of a viral packaging motor change DNA translocation velocity and processivity. *Journal of Biological Chemistry* **285,** 24282–24289 (2010).

12. Sun, S., Kondabagil, K., Gentz, P. M., Rossmann, M. G. & Rao, V. B. The Structure of the ATPase that Powers DNA Packaging into Bacteriophage T4 Procapsids. *Molecular Cell* **25,** 943–949 (2007).

13. Gemmen, G. J., Millin, R. & Smith, D. E. DNA looping by two-site restriction endonucleases: heterogeneous probability distributions for loop size and unbinding force. *Nucleic Acids Research* **34,** 2864–2877 (2006).

14. Kress, M., May, E., Cassingena, R. & May, P. Simian virus 40-transformed cells express new species of proteins precipitable by anti-simian virus 40 tumor serum. *J. Virol.* **31,** 472–483 (1979).

15. Smith, A. E., Smith, R. & Paucha, E. Characterization of different tumor antigens present in cells transformed by simian virus 40. *Cell* **18,** 335–346 (1979).

16. Lane, D. P. p53, guardian of the genome. *Nature* **358,** 15–16 (1992).

17. Finlay, C. A., Hinds, P. W. & Levine, A. J. The p53 proto-oncogene can act as a suppressor of transformation. *Cell* **57,** 1083–1093 (1989).

18. Hollstein, M. K.Rice, M.S.Greenblatt, T.Soussi, R.Fuchs, T.Sorlie1, E.Hovig, B.Smith-Sorensen, R.Montesano and C.C.Harris. Database of p53 gene somatic mutations in human tumors and cell lines. *Nucleic Acids Research* **22,** 3551–3555 (1994).

19. Wang, Y., Schwedes, J. F., Parks, D., Mann, K. & Tegtmeyer, P. Interaction of p53 with its consensus DNA-binding site. *Mol. Cell. Biol.* **15,** 2157–2165 (1995).

20. Miller, S. D., Farmer, G. & Prives, C. p53 inhibits DNA replication in vitro in a DNA-binding-dependent manner. *Mol. Cell. Biol.* **15,** 6554–6560 (1995).

21. Saramaki, A. Regulation of the human p21(waf1/cip1) gene promoter via multiple binding sites for p53 and the vitamin D3 receptor. *Nucleic Acids Research* **34,** 543–554 (2006).

22. Jackson, P. & Yardley, G. Distinct distal and proximal p53-binding sites in the MCK promoter govern the transcriptional response to p53. *FEBS Lett.* **406,** 271–274 (1997).

23. Okorokov, A. L. Michael B Sherman, Celia Plisson, Vera Grinkevich, Kristmundur Sigmundsson, Galina Selivanova, Jo Milner and Elena V Orlova. The structure of p53 tumour suppressor protein reveals the basis for its functional plasticity. *The EMBO Journal* **25,** 5191–5200 (2006).

24. Migliori, A., Kung, S. & Wang, D. Direct measurement of formation of loops in DNA by a human tumor suppressor protein. *SPIE …* (2013).

25. Felgner, H., Frank, R. & Schliwa, M. Flexural rigidity of microtubules measured with the use of optical tweezers. *J. Cell. Sci.* **109 (Pt 2),** 509–516 (1996).

26. Seidel, R. John van Noort, Carsten van der Scheer, Joost G P Bloom, Nynke H Dekker, Christina F Dutta, Alex Blundell, Terence Robinson, Keith Firman and Cees Dekker. Real-time observation of DNA translocation by the type I restriction modification enzyme EcoR124I. *Nature Publishing Group* **11,** 838–843 (2004).

27. Frieda, K. L. & Block, S. M. Direct Observation of Cotranscriptional Folding in an Adenine Riboswitch. *Science* **338,** 397–400 (2012).

28. Svoboda, K., Schmidt, C. F., Schnapp, B. J. & Block, S. M. Direct observation of kinesin stepping by optical trapping interferometry. *Nature* **365,** 721–727 (1993).

29. Bai, L., Santangelo, T. J. & Wang, M. D. Single-molecule analysis of RNA polymerase transcription. *Annu. Rev. Biophys. Biomol. Struct.* **35,** 343–360 (2006).

30. Yu, J., Moffitt, J., Hetherington, C. L., Bustamante, C. & Oster, G. Mechanochemistry of a Viral DNA Packaging Motor. *Journal of Molecular Biology* **400,** 186–203 (2010).

31. Fuller, D. N. Gregory J. Gemmen, John Peter Rickgauer, Aurelie Dupont, Rachel Millin, Pierre Recouvreux and Douglas E. Smith. A general method for manipulating DNA sequences from any organism with optical tweezers. *Nucleic Acids Research* **34,** e15 (2006).

32. Ostermann, A., Waschipky, R., Parak, F. G. & Nienhaus, G. U. Ligand binding and conformational motions in myoglobin. *Nature* **404,** 205–208 (2000).

33. Shrivastava, I. H. & Sansom, M. S. Simulations of ion permeation through a potassium channel: molecular dynamics of KcsA in a phospholipid bilayer. *Biophysical Journal* **78,** 557–570 (2000).

34. Sanbonmatsu, K. Y., Joseph, S. & Tung, C. S. Simulating movement of tRNA into the ribosome during decoding. *Proceedings of the National Academy of Sciences* **102,** 15854–15859 (2005).

35. Duan, Y., Wang, L. & Kollman, P. A. The early stage of folding of villin headpiece subdomain observed in a 200-nanosecond fully solvated molecular dynamics simulation. *Proceedings of the National Academy of Sciences* **95,** 9897–9902 (1998).

II THE MECHANISM OF VIRAL DNA PACKAGING BY gp17

2 Structural basis for force generation by the bacteriophage T4 viral DNA packaging motor.

Amy D. Migliori[a], Nicholas Keller[a], Tanfis I. Alam[b], Marthandan Mahalingam[b], Venigalla B. Rao[b], Gaurav Arya[c], Douglas E. Smith[a]

[a]Department of Physics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093
[b]Dept. of Biology, The Catholic University of America, 620 Michigan Ave. NE, Washington, DC, 20064
[c] Department of Nanoengineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093

2.1 Summary

How viral packaging motors generate enormous forces to translocate DNA into viral capsids remains unknown. Recent structural studies of the bacteriophage T4 packaging motor have led to a proposed mechanism wherein the gp17 motor protein ratchets DNA by transitioning between extended and compact states, orchestrated by electrostatic interactions between complimentarily charged residues across the interface between the N- and C-terminal subdomains. Here, we show that site-directed alterations in these residues cause force dependent impairments of motor function including lower translocation velocity, lower stall force, and higher frequency of pauses and slips. The measured impairments correlate with computed changes in free energy differences between the two states. These findings support the proposed structural mechanism and further suggest an energy landscape model of motor activity that

couples the free energy profile of motor conformational states with that of the ATP hydrolysis cycle.

Keywords: T4, gp17, molecular motor, single molecule biophysics, MM-GBSA

2.2 Introduction

Many double-stranded DNA viruses utilize a molecular motor during assembly to translocate DNA into a preformed capsid shell[1-4]. The motor must do significant mechanical work against forces arising from DNA bending rigidity, electrostatic self-repulsion, and entropy loss, which oppose DNA confinement[5,6]. Measurements of DNA packaging with optical tweezers in bacteriophages *Φ*29, lambda, and T4 show that all three motors can generate forces >50 pN, which is >20x higher than forces generated by skeletal muscle myosin motors[7-9]. The motors differ in DNA translocation rates, which range from up to ~200 bp/s for *Φ*29 to up to ~2000 bp/s for the T4 motor (which has a ~9× longer genome than *Φ29*)[9,10].

A number of studies have begun to examine the microscopic details of how these motors function. Measurements of single-DNA packaging dynamics of *Φ*29 via optical tweezers have enumerated the mechanical and chemical kinetic transitions and rate constants, translocation step size, and strong coordination of motor subunits[7,11-13]. Bioinformatics analyses and mutagenesis studies altering the lambda and T4 motor proteins indicate the presence of several functional motifs that are homologous with other ATPases, including related DNA and RNA translocating motors such as helicase and chromosome segregation transporters[14]. Single-molecule packaging

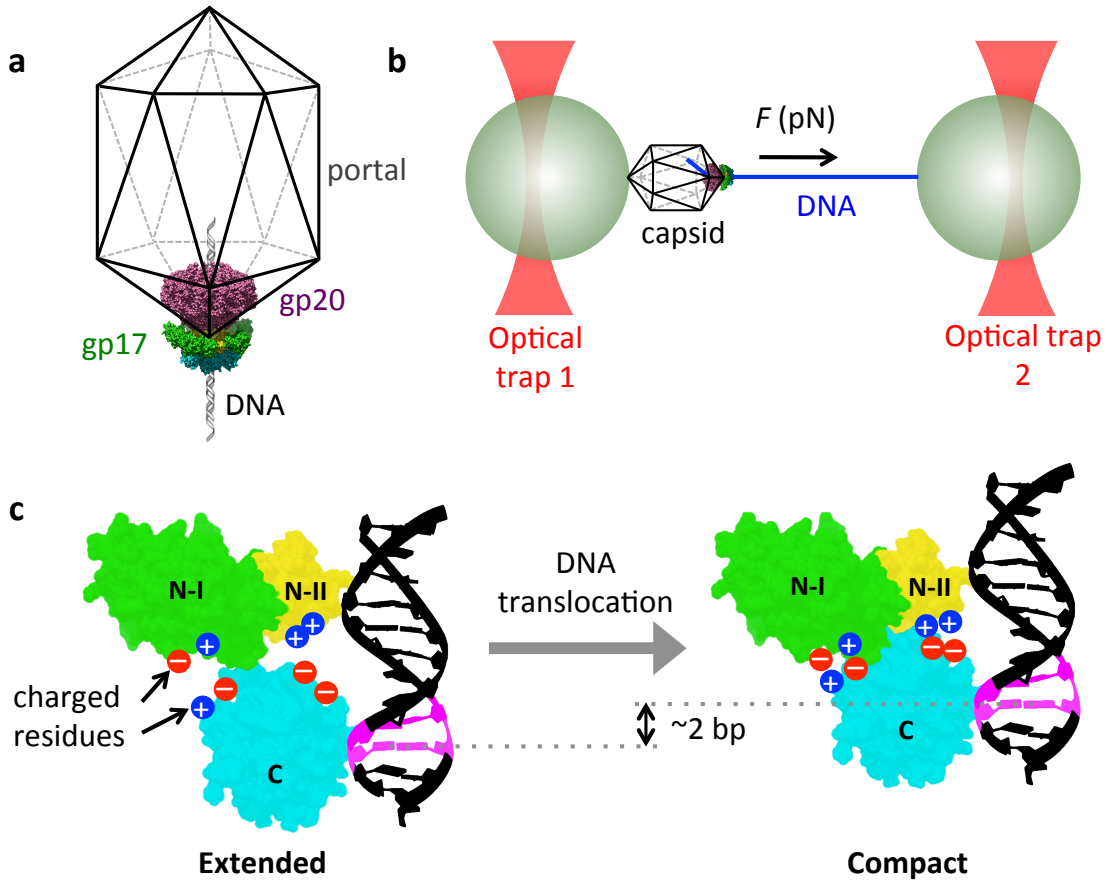


Fig. 2.1. Schematic overview of the ratchet mechanism of DNA packaging in T4 bacteriophages and the single-molecule approach used for testing this mechanism. (A) The T4 packaging motor consists of concentric rings of gp17 and gp20 proteins located at the capsid portal. (B) The gp17 ratchet model proposes that DNA is translocated into the capsid in steps of ~2 bp through a conformational transition in gp17 from an extended to a compact state, triggered by electrostatic attraction between complementary pairs of charges residing at the interface of the N- and C-domains of gp17. (C) The ratchet model is tested through optical tweezers measurements of single DNA molecule packaging dynamics, in conjunction with site-directed mutagenesis and molecular dynamics simulations.

measurements showed that mutations in the Walker A, C-motif, Q-motif, and other structurally conserved regions of the large terminase subunit of the lambda motor cause phenotypic impairments in motor function, including reduced translocation velocity, increased pausing and slipping, and steepening of the force-velocity

relationship[15,16]. The structure-function relationships for these motors, however, have remained unclear due to lack of high-resolution structures.

The recent determinations of the high resolution X-ray crystal structure of the T4 large terminase subunit (gp17), and related structures of the ssRNA phage *Φ*12 and dsDNA phage Sf6 motor proteins, have paved the way to developing a full understanding of the motor mechanism[17-19]. In case of T4, cryo-electron microscopy 3D reconstruction of the entire motor complex during DNA packaging indicates that the motor consists of a pentameric ring of gp17 subunits attached to a dodecameric ring of gp20 proteins, together forming the portal entry channel (Fig 2.1A). gp17 was found to have an N-terminal subdomain, which attaches to the portal and contains the ATP binding site, and a C-terminal subdomain, which was proposed to bind to DNA. The two subdomains, connected by a flexible linker region, interact across a large interface containing five aligned pairs of complimentarily charged residues. Comparison of the X-ray and cryo-EM structures[17] revealed two distinct conformations states: a “compact” state seen in the X-ray structure of the gp17 monomer in which the N- and C-terminal domains are in close contact, and an “extended” state, seen in the cryo-EM structure of packaging complexes in which the N- and C-terminal domains are separated by ~7 Å.

Based on the two different conformations of gp17 seen in the structural data, Sun et al.[17] proposed a model for motor operation in which the DNA is ratcheted in ~7 Å (≅ 2 bp) increments by the transition of gp17 from the extended to compact state. This transition was proposed to be driven by interactions of the complimentarily charged residues across the interface that generate force to translocate the DNA into the prohead (Fig. 2.1B and C). ATP hydrolysis was proposed to regulate DNA

interaction and provide energy to return gp17 to the extended state after each 2 bp DNA ratcheting step. Henceforth, we refer to this model as the “ratchet model”.

Here, we test this model by measuring effects on DNA translocation dynamics of site-directed mutations which alter the charged interfacial residues proposed to be responsible for force generation. To quantify the impact of these residue changes, we also carried out molecular dynamics simulations of gp17 to predict their effects on free-energy differences between the extended and compact states. In support of the model, we find that predicted changes in these free-energy differences correlate strongly with measured impairments of motor function. Further, our findings lead to a simple mechanochemical energy landscape picture of gp17 motor dynamics that can explain the experimental observations and simulations.

2.3 Materials and Methods

Materials

A 25,339 bp dsDNA construct used as a substrate for packaging was prepared by PCR from lambda phage DNA (NEB, Inc.) using primers biotin-5'-CTGATGAGTTCGTGTCCGTACAACTGGCGTAATC and 5'-ATCCGATCTGCGTTACCGAATGGATGGATG (Operon, Inc.) using methods described previously[23]. T4 phage heads were prepared as described in Ref. 24, wild type gp17 motor protein was prepared as described in Ref. 25, and gp17 mutant proteins were prepared as described in Ref. 18. Bulk DNA packaging assays shown in supplemental figure S2.1 were performed as described previously[26].

2.3.1 Single-molecule packaging measurements

T4 head-motor complexes were prepared using a similar protocol as described previously[10] by mixing 1.7×10^{10} heads with 70 picomols of gp17 in a solution containing T4 packaging buffer (80 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 100 mM NaCl), 1 mM gamma-S-ATP (non-hydrolyzable ATP analog), and 450 ng of 120 bp "initiating" DNA (which we found stabilizes the competent head-motor complex). This solution was incubated at room temperature for 45 min. 2 μl of T4 antibody coated 2.2 μm diameter protein-G microspheres (Spherotech) were mixed 10 μl of these stalled T4 complexes and incubated at room temperature for 45 min. The biotinylated dsDNA was attached to 2.2 μm diameter streptavidin coated microspheres (Spherotech). Measurements were carried out in solution containing 50 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 80 mM NaCl, 1 mM ATP, and 0.05 g/L BSA.

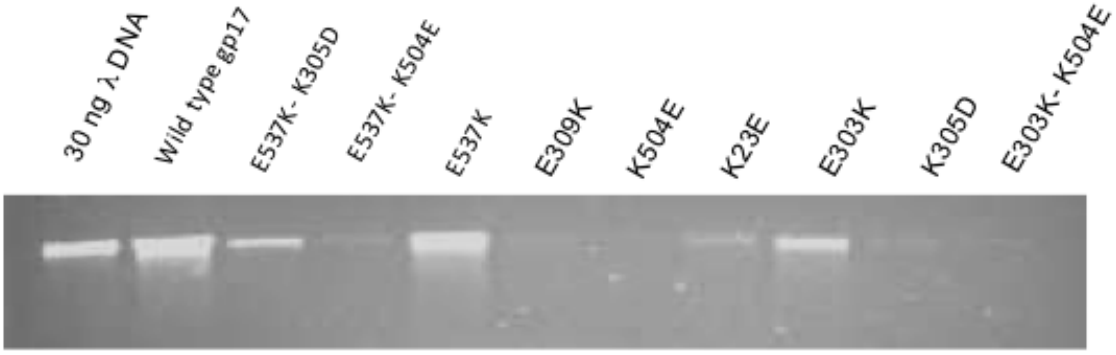


Fig S2.1. Bulk measurements of in vitro DNA packaging using the DNase protection assay as described previously (Ref. 23). Each packaging reaction used 10^{10} proheads, 300 ng of phage λ DNA, 2 µM gp17, and 1 mM ATP. The far left lane is a control sample containing 30 ng λ DNA for comparison.

2.3.2 High force measurements.

Packaging was first measured at low force (5 pN) for 5-10 seconds in force clamp mode as described previously[10], to verify motor activity. The force was then

quickly ramped (within 0.2 s) to a higher value by adjusting the set-point on the force-clamp control system. DNA translocation (packaging and/or slipping) was then tracked in force-clamp mode until the DNA tether was lost. The force was monitored at 2 kHz and was held constant by a feedback control system that adjusts the separation between the two traps in 0.5 nm increments. The optical tweezers system was calibrated as described previously[22]. A depiction of the optical setup can be seen in Appendix A.

2.3.2 Data Analysis.

The motor velocity was calculated using methods similar as those described previously[10] by linear fitting of the DNA length vs. time records in a 1 s sliding window. Slips larger than 100 bp and pauses in which the velocity dropped to < 12 bp/s in the 1 s window were not included in determining "motor velocity" (but were included in determining overall "packaging rate").

Computational Methods Overview

Our aim is to compute the free energy difference $\Delta G_{EC} \equiv G_C - G_E$ between the extended and compact states of gp17 for the wild type and mutant proteins. In principle, ΔG_{EC} can be computed from all-atom molecular dynamics (MD) simulations via approaches like umbrella sampling, though such approaches would entail prohibitive computational costs for systems as large as gp17. To this end, we use the Molecular Mechanics-Generalized Born Surface Area (MM-GBSA) method[27], which provides quick estimates of the binding free energy of molecular complexes by approximating the solvation free energy using implicit solvent mean field models.

However, the application of MM-GBSA to computing intramolecular binding free energies, as opposed to intermolecular binding free energies, requires the N- and C-terminal domains to be treated as separate “molecules”, which assumes, consistent with structural features observed in Ref. 22 that the hinge joining the two domains is flexible and does not contribute much to the free energy difference between the compact and extended states. One can then obtain ΔG_{EC} using the free energy cycle shown in Fig. S2.2, where the MM-GBSA method is used to compute the binding free energies $\Delta G_{bind,E}$ and $\Delta G_{bind,C}$ for assembling the N- and C-terminal domains into extended and compact states from a common isolated state in which the two domains do not interact with each other. The required free energy difference can then be simply calculated as $\Delta G_{EC} = \Delta G_{bind,E} - \Delta G_{bind,C}$.

2.3.3 Preparation of initial configurations.

The MM-GBSA method involves all-atom MD simulations of the extended and compact state complexes in explicit solvent and ions, starting from suitable initial configurations. These initial configurations were prepared based on the high-resolution crystal structure of gp17 in the compact state (PDB code: 3CPE)[18].

To generate the initial configuration of the compact state complex, we solvated the crystal structure in a TIP3P water box extending 10 Å from the surface of the protein in each direction by using the *solvate* function in VMD and then added 150 mM NaCl by using the *autoionize* function in VMD[28]. The solvated structure was energy minimized using 100 steps of steepest gradient, and then equilibrated using MD simulations for 0.5 ns to yield the initial configuration of the compact state complex. The atomic structure of the extended state is not known from X-ray studies but the

cryo-electron microscopy studies show that the C- and N-terminal domains are spaced a further ~7 Å apart as compared to the compact state[18]. To generate an atomic model of this state we generated a vector between the center of mass of the N- and C-terminal domains and used steered MD simulations to pull the C-terminal domain along this vector at a slow rate of 0.0002 Å/fs until a separation of 7 Å between the two center of masses was achieved (the script used for SMD can be found in Appendix B). The resulting structure was equilibrated using MD simulations for 0.5 ns.

To generate the extended and compact state complexes for the three studied mutants, we used the *mutator* function in VMD to replace the chosen residue with the desired residue in the final equilibrated complexes. We then carried out 100 steps of energy minimization via steepest descent to allow the mutated residue as well as the nearby residues and water molecules to relax to their low-energy conformations.

Energy minimization and MD simulations used for preparing the initial configurations were performed in NAMD[29] using the CHARMM 27 force field[30]. The simulations were carried out at constant temperature and pressure consistent with physiological conditions. The temperature was maintained at 310 K using a Langevin thermostat with a damping constant 1/ps and the pressure was maintained at 1 atm using a Langevin piston with a period of 100 fs and damping time constant of 50 fs. We used a time step of 2 fs for integrating the equations of motion and employed the SHAKE algorithm to constrain all bond lengths involving hydrogen atoms. For computing short-range non-bonded interactions, a cutoff distance of 12 Å was employed. The simulations were carried out with periodic boundary conditions with all electrostatic interactions computed using the particle mesh Ewald method with grid spacing of 1 Å.

Although initial equilibration as well as mutant generation were done using namd and the CHARMM force field, the bulk of the computational work was done using AMBER. Where namd uses psf files, AMBER uses topology files, which are also generated using the pdb file. Unfortunately, the residue and atom naming conventions of the two molecular dynamics suites differ quite a bit, and pdb files used in namd must be edited in order to be compatible with AMBER. For example, all amino protons must be renamed as H, in namd they are named HN. After making these changes, it is possible to generate a topology file using either ptraj, a program included with AmberTools (see section about MM-GBSA), or with Chimera Write Prmtop. Either of these processes will output 2 file types : the topology file, and a coordinate file.

We generated three additional initial configurations of the complexes by performing MD simulations of the original initial configurations described above in AMBER with the ff99SB force field. These additional simulations were carried out at three different temperatures (300 K, 320 K, and 330 K) for 50 ps while imposing a weak restraint on all heavy atoms of the protein via a harmonic potential with a spring constant of 2.0 kcal/mol/Å^2. This procedure forced the configurations to depart significantly from the starting configuration while remaining stable at these non-physiological temperatures. We carried out a further 600 ps of MD simulations during which the restraints were removed and the temperatures were brought back to 310 K. Thus, a total of four different initial configurations for each complex were obtained, allowing us to carry out four independent MM-GBSA calculations for improving the accuracy of our free energy estimates.

2.3.4 Preparation for MM-GBSA calculations

The Amber suite has a set of companion programs, called AmberTools, that greatly expands computational utility. AmberTools consists of “several independently developed packages that work well with Amber itself”, including the LEAP program (use of which will be discussed), ptraj (will be discussed), MM-GBSA, and others. To calculate the free energy change associated with the proposed conformational change during packaging in gp17, we utilized MM-GBSA, which is designed to calculate the free energy of each simulation time step and formulate a free energy change average over the course of the simulation. By then calculating the difference in free energies between compact and extended gp17, we are able to find the ΔΔG of conformational change. However, MM-GBSA is traditionally used to calculate binding energies between a receptor protein and a small molecule ligand, and because of this we had to devise a method to use it for our single-protein system.

To use MM-GBSA to probe a single-protein conformational change, it is necessary to divide gp17 into two separate pieces to be considered. In order to satisfy the program’s requirement for separate receptor and ligand molecules, we split gp17 at the hinge to form two artificial molecules. Since we are probing an interaction between two subdomains that we believe operate as a molecular ratchet, we decided to consider the two halves separately and to approximate the interaction as a binding interface. In addition, we wished to measure the importance of the interactions formed at the interface between N- and C-terminal subdomains, which we were able to do by addressing the ratcheting between the two subdomains as a binding interaction.

Separating the N- and C-terminal subdomains into computationally compatible sections is not a trivial problem. MM-GBSA requires that the sum of all atoms contained in the ligand and receptor molecules be equal to that of the complex

(receptor + ligand structure). Because cutting a protein leaves a severed peptide bond that normally becomes hydrolyzed, it is not possible to prepare each section separately; creating topology files requires that the protein be structurally complete, with charges assigned to the atoms, and hydrogens added where necessary. To prep the molecule would add atoms to the split peptide bond, adding 3 atoms total to the ligand and receptor. To circumvent this potential problem (which crashes mm-pbsa/gbsa), we ran the dock prep utility (Chimera) on two identical molecules of the protein being prepped. Dock prep adds hydrogens where needed, and assign charges based on standard residue information. Because the molecule is intact at this point, we are able to standardize the number of atoms. After this dock prep, we save one copy of the entire molecule as a topology file (using the ff99SB force field), needed for MM-GBSA. We then separate the two molecules. To create the two halves, we selected one half of one of the two molecules, noting the number of atoms selected. We then delete the selected atoms. On the second molecule, we make the same selection, making sure the number of atoms is equal to the first selection. We then invert the selection for that molecule and delete the selected atoms. This leaves us with two halves of gp17 that are able to be written to a topology file without adding the missing atoms at the position of the peptide bond split. In this way, we are able to circumvent the problem of needing the #atoms(ligand) + #atoms(receptor) = #atoms(complex) and MM-GBSA is able to calculate the free energy of interaction without error.

To run the actual calculations in MM-GBSA, we created these cut pairs of receptor-ligand molecules for the compact and extended forms of each construct studied. Each molecule was cut at the linkage between T359 and S360 and prepared

as described. We designated the N-terminal subdomain as the receptor, and the C-terminal subdomain as the ligand, though the reverse would yield the same results.

2.3.5 MM-GBSA calculations.

We next performed exhaustive MD simulations of extended state and compact state complexes for wild type and mutant gp17. The binding free energies $\Delta G_{bind,E}$ and $\Delta G_{bind,C}$ were computed according to MM-GBSA via the following relation: $\Delta G_{bind} = \Delta E_{MM} + \Delta G_{solv} - T\Delta S_{conf}$, where ΔE_{MM}, ΔG_{solv}, and $T\Delta S_{conf}$ represent respective changes in molecular mechanics energy, solvation free energy, and configurational entropy of the C- and N-terminal domains as they associate into a complex (extended or compact).

The three free energy contributions were computed from 500 snapshots of each complex recorded during 4 ns MD simulations, carried out using the *mm_pbsa.pl* script in AMBER 10[31]. In these calculations, we defined the N- and C-terminal domains as separate molecules, namely, the “receptor” and “ligand”, respectively, and the full-length gp17 molecule as the “complex”, as explained earlier. Since the flexible hinge region contains residues 358–362, we define the N- and C-terminal domains by residues 1–359 and 360–560, respectively. The ΔE_{MM} contributions were directly computed as the change in the sum of the bonded, electrostatic, and vdW interaction energies. The ΔG_{solv} contributions were further divided into polar and nonpolar contributions. The polar contributions were calculated using the Generalized Born approach[32,33] while the nonpolar contributions were calculated from changes in solvent accessible surface area upon binding multiplied by the surface tension[34]. We used a

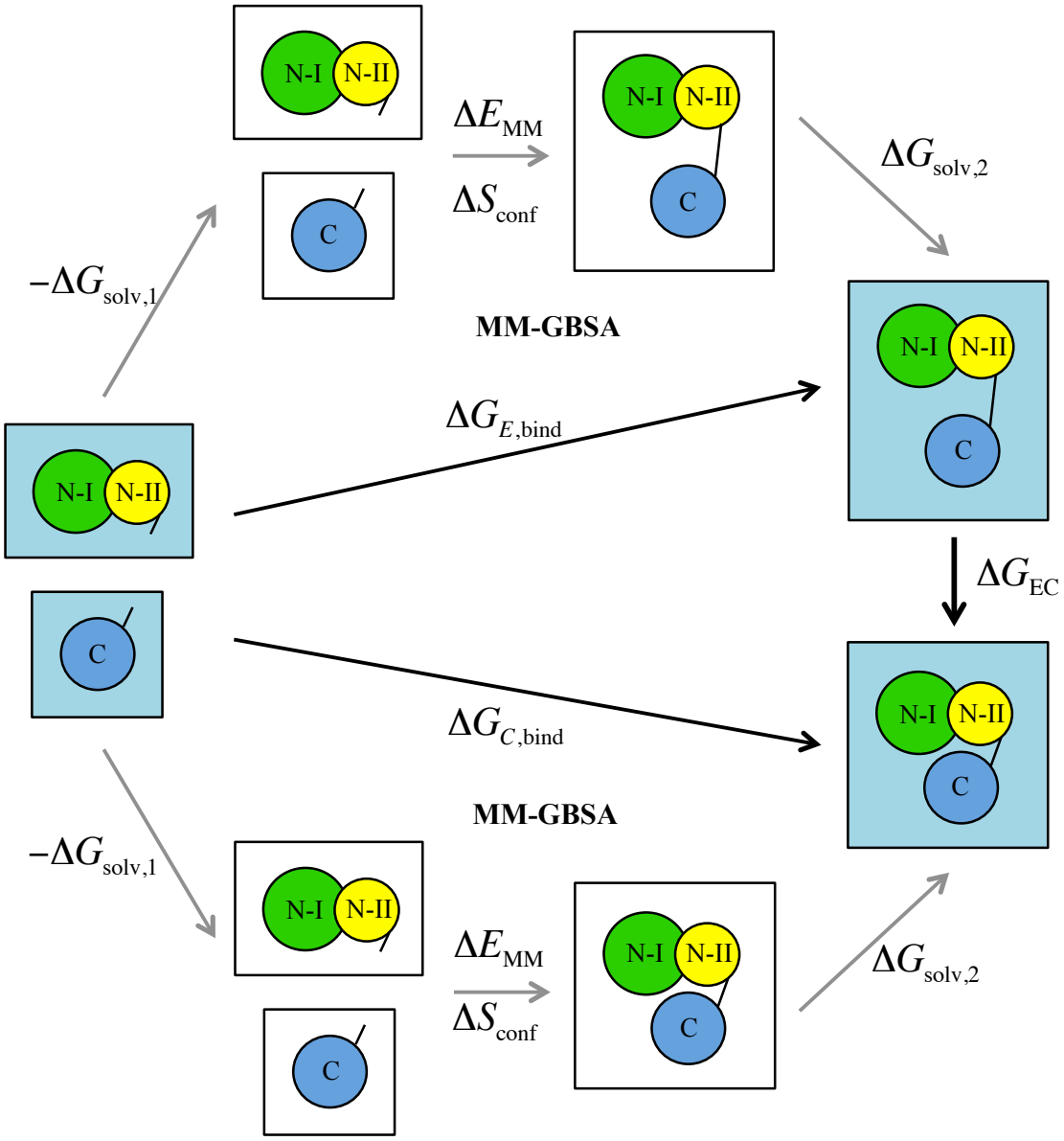


Fig S2.2. Free energy cycle depicting our MM-GBSA-based approach for computing the free energy difference ΔG_{EC} between the extended (E) and compact (C) conformations of gp17 (for wild type and mutant motors) as described in the methods section. Note that the net solvation free energy ΔG_{solv} (see text) is given by $\Delta G_{solv,2}$ - $\Delta G_{solv,1}$.

water probe of radius 1.4 Å and a surface tension of 0.0072 kcal/mol/Å^2 for the latter calculation. The $T\Delta S_{conf}$ contributions are typically computed using normal mode analysis[35], but these calculations were found to converge extremely slowly. Hence, we

neglected its contributions to the free energy difference. However, we believe that these entropic contributions will mostly cancel out when comparing ΔG_{EC} values across wild type and mutant gp17. The algorithms and associated parameters, force field, and external conditions used for these MD simulations were kept identical to those used for generating the initial conditions.

To obtain reliable values of ΔG_{EC} as well as statistical uncertainties, we carried out four sets of MM-GBSA calculations for each wild type and mutant complex, each starting from the four different initial configurations of the extended and compact state complex prepared using the procedures described in the previous section.

2.4 Results

2.4.1 Site-directed mutagenesis

To test the proposal that pairs of aligned, complimentarily charged residues within the interface domain are critical to motor function, we used site-directed mutagenesis to alter these residues. Fig. 2.2A shows the locations of the mutated residues in the motor. We substituted these amino acids with ones with reversed charges (*i.e.*, we changed + to –, or – to +) to putatively weaken the attractive electrostatic interactions between the N- and C-terminal subdomains proposed to drive the conformational change that ratchets the DNA. In particular, six mutants with single residue changes (K23E, E303K, K305D, E309K, K504E, and E537K) and three with double residue changes (E303K/K504E, E537K/K305D, and E537K/K504E) were studied. Of these, six exhibited little or no packaging activity, as measured in bulk packaging reactions by DNase protection assays (see Supplementary Fig. S2.1).

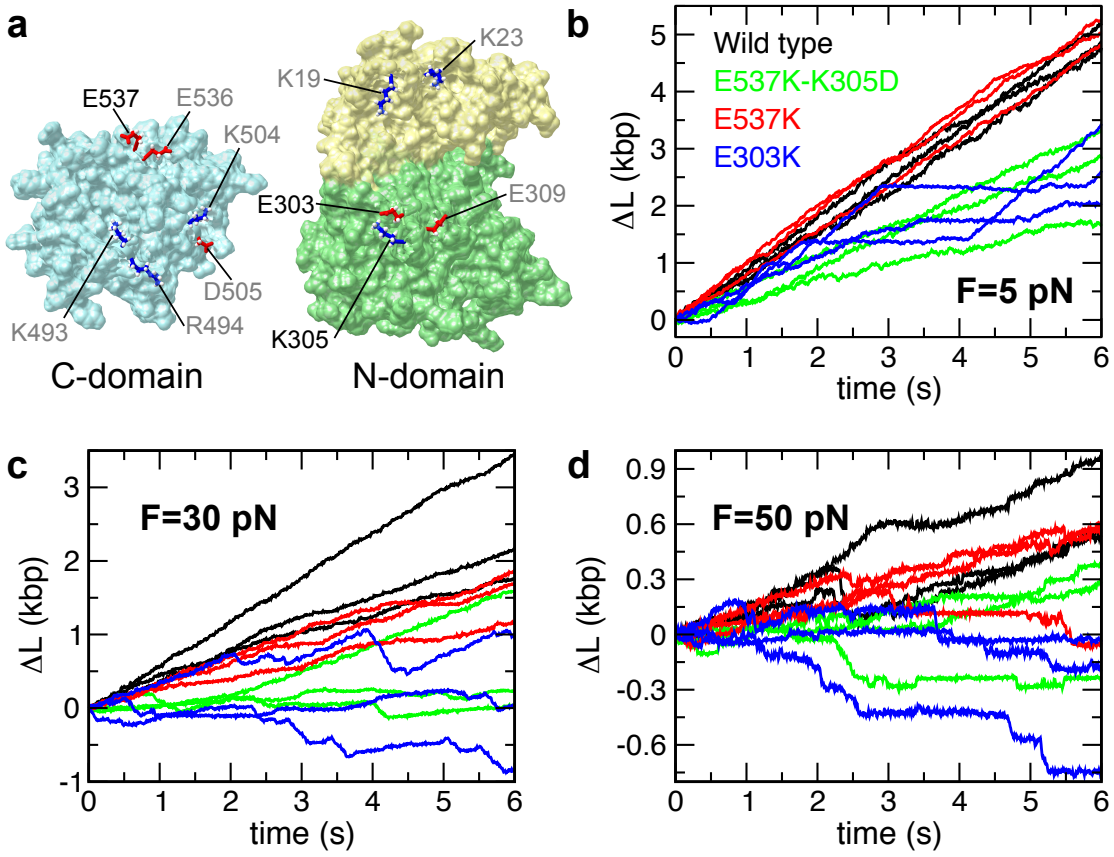


Fig. 2.2. (A) Open book representation of the N- and C-terminal domain of gp17 displaying the charged residues studied in this work. Changes in the residues labeled in grey resulted in zero bulk packaging while those labeled in black resulted in some packaging, leading to their further study through optical traps. Representative measurements of single DNA molecule packaging dynamics (length packaged vs. time) with (B) 5 pN applied load, (C) 30 pN, (D) 50 pN. The colors indicate wild type (black), E537K (red), E537K-K305D (green), and E303K (blue) gp17 motors.

These findings are consistent with the ratchet model, although we note that packaging activity as assessed by these bulk assays can exhibit impairment for multiple reasons, including failure of motor proteins to fold, failure of motor proteins to assemble into an active pentameric complex, or failure to initiate DNA interaction. Three of the mutants (E303K, E537K, and E537K-K305D), however, exhibited sufficient activity to permit detailed measurements of packaging dynamics using the single-molecule optical

tweezers assay. We conducted detailed measurements with these mutants as well as the wild type complex for comparison.

2.4.2 Measurements of packaging dynamics at low force

We first carried out single-molecule packaging measurements using the “low force clamp” technique described previously[9] (Materials and Methods). Briefly, T4 prohead-motor complexes were assembled using purified recombinant gp17, attached to microspheres via antibodies, injected into a fluid chamber containing ATP, and trapped in one optical trap. DNA molecules were biotinylated at one end and attached to streptavidin-coated microspheres and trapped in a second optical trap. Packaging was initiated by bringing the two microspheres into proximity such that the motor binds to and begins reeling in the DNA, thus exerting a force that pulls the two microspheres together (Fig. 2.1D). DNA translocation was tracked using a feedback control system that held the applied load force at a constant value of 5 pN, which is low compared to the maximum force generated by the wild type motor (>60 pN).

Measurements were made with the three mutants and with wild type gp17 for direct comparison (Fig. 2.2A and 2.3). We found that E537K mutants packaged at nearly the same rate (~600 bp/s) on average as wild type, while E303K and E537K-K305D mutants showed a modest reduction to ~390 bp/s. The reductions were partly due to an increase in pausing and slipping and partly due to inherently reduced motor velocity. We considered these results promising for supporting the gp17 ratchet model. However, given that one mutant (E537K) exhibited no change, the other two mutants exhibited only modest changes, and that the double mutant (E537K-K305D) did not

show greater impairment, the support for the ratchet model could be considered ambiguous.

2.4.3 High force measurements

To more rigorously test the model, we investigated the motor response under increasing applied loads, where the impact of residue changes proposed to affect force generation would be expected to be greater. In previous studies of the wild type motor, high force measurements were made by holding the trap positions fixed and allowing the tension in the DNA to build up due to the shortening of the DNA tether accompanying packaging[9]. However, this approach is undesirable for characterizing mutants because they package DNA at much slower rates, which would cause a slower build-up of force. This leads to a bias in the data analysis, as the different complexes would experience different force loading histories. Also, prolonged force build up often leads to rapid detachment of the tethered complexes before the target force is reached, due to the DNA slipping out of the prohead or dissociation of the prohead from the antibody-coated microsphere. Therefore, in the present work we implemented a new protocol in which we first packaged ~5 kbp of DNA at low force (5 pN) to confirm activity, and then quickly ramped the force to a higher value and clamped it via feedback control to measure the response of the motor. Thus, all complexes, wild type and mutants, were subject to the same force history.

Using this new protocol, force clamp measurements were recorded at 5, 15, 30, 40, 50, and 60 pN loads. Examples of individual packaging traces are shown in Fig. 2.2, highlighting the highly variable and stochastic nature of packaging by the gp17 motor. We also note the presence of pauses in translocation (flat regions) and slippage

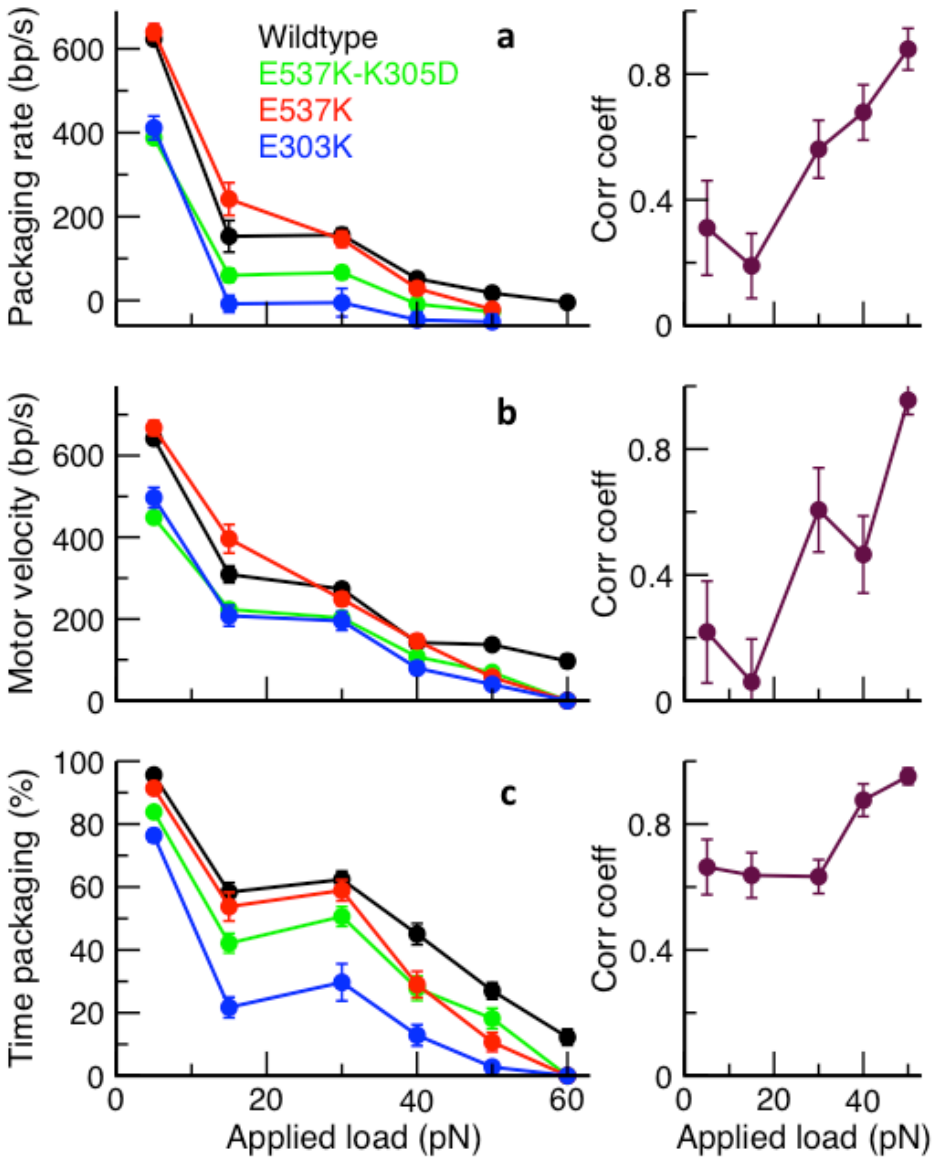


Fig. 2.3. (Left panel) Metrics of motor activity versus applied force for wild type and mutants: packaging rate (A), motor velocity (B), and % time spent packaging (C). The colors indicate wild type (black), E537K (red), E537K-K305D (green), E303K (blue). (Right panel) corresponding Pearson correlation coefficients (min=0, max=1) between experimental metrics of motor function (adjacent plots) and calculated free energy changes between the extended and compact forms of gp17.

(negative slope) interspersed with DNA packaging episodes (positive slope). Higher load forces significantly slow packaging and mutations further slow packaging. The

frequency and length of the pauses and slips also increase with increasing load and mutations.

We analyzed ensembles of packaging trajectories to determine several averaged metrics of motor function (Fig. 2.3). Specifically, we calculated the "packaging rate", defined as the net length of DNA packaged per unit time; the "motor velocity", defined as the rate of DNA translocation during episodes of packaging (*i.e.*, not including pauses and slips); and the "percent time spent packaging", defined as the percentage of time packaging occurs (vs. pausing or slipping). All three metrics show that the residue changes cause more severe impairment at high load than at low load. For example, the packaging rate for mutant E303K, which was ~400 bp/s at 5 pN, drops to near zero above 15 pN. E303K exhibits episodes of packaging, but slips so frequently that there is no net packaging. The packaging rates for E303K and E537K-K305D, which were nearly the same at 5 pN, also exhibited a clear difference above 15 pN. The rate for E537K, which is nearly as large as that of wild type at 5 pN, drops to near zero at 50 pN vs. ~20 bp/s for wildtype. Above 30 pN, the percent time packaging for all the mutants drops clearly below that of wild type, and at 60 pN it drops to 0% for all the mutants vs. ~15% for wild type.

When the force was increased to 50 pN the mutants displayed only brief episodes of highly slowed packaging, and at 60 pN no packaging was detected, whereas the wild type motor was still able to package ~15% of the time and with an average motor velocity of ~134 bp/s. Thus, the mutant motors exhibit a reduced force-generation capacity F_{max} between 50 and 60 pN. Overall, these results consistently support the gp17 ratchet model because residue changes expected to weaken the

electrostatic interactions proposed to drive the extended-to-compact transition indeed cause impairments in the packaging rate and force-generating capacity of the motor.

2.4.4 Molecular dynamics-based free energy calculations

To more rigorously and quantitatively test the model we sought to predict the degree to which different residue changes may be expected to cause impairments. In particular, why does the measured motor impairment increase in the following order E537K-K305D, E537K, and E303K, with the double mutant showing perhaps surprisingly smaller impairment than the corresponding single mutant? To this end, we carried out molecular dynamics (MD) simulations of gp17 to predict the free energy difference ΔG_{EC} between the extended (E) and compact states (C) observed in the structural studies. The ratchet model proposed that DNA translocation is driven by a transition between these states, orchestrated by attractive interactions between charged residues across the N- and C- terminal domains. If the model is correct, we would expect a favorable free energy difference ($\Delta G_{EC} < 0$), sufficient to explain the large forces generated by the motor, and we would also expect the magnitude $|\Delta G_{EC}|$ to be reduced by mutations altering residues involved in generating these forces.

We determined $\Delta G_{EC} = G_C - G_E$ by computing the free energies of the compact (G_C) and extended states (G_E) relative to a common reference state. The relative free energies for each state were calculated as “binding free energies” between the N- and C-terminal domains using the Molecular Mechanics Generalized Born Surface Area (MM-GBSA) approach[20]. In these calculations, the compact state was modeled as the available crystal structure of gp17 in the compact state while the extended state

observed in cryo-EM studies was modeled from the crystal structure by imposing a 7Å separation between the N- and C-terminal domains. See the *Methods* section for more details about the computational approach.

Our calculations yielded ΔG_{EC} = 38.21 ± 4.42 kcal/mol for wild type gp17, -19.95 ± 9.99 kcal/mol for E537K-K305D, -6.14 ± 3.44 kcal/mol for E537K, and -2.79 ± 5.55 kcal/mol for E303K. These values do not include contributions from changes in the configurational (mostly vibrational) entropy of the gp17, as we found that estimates of vibrational entropy from normal mode calculations yielded large uncertainties. We do not, however, expect vibrational entropy to vary significantly across different mutants. Notably the determined free energy values are all negative, consistent with the transition from the extended to compact state being most favorable for the wild type gp17 and favorable to a lesser degree for the mutants. In contrast, we calculated a positive or near zero (unfavorable) ΔG_{EC} = 1.83 ± 10.3 kcal/mol for the mutant E309K, which exhibited no packaging activity in both bulk packaging assays and optical tweezers assays.

For wild type gp17, the calculated magnitude $|\Delta G_{EC}| \cong 38$ kcal/mol is equivalent to ~260 pN·nm, which implies a theoretical maximum average force of ≅ 260/0.7 ≅ 370 pN during the 0.7 nm displacement. Thus, our calculations show that the extended-to-collapsed state ratchet mechanism can indeed produce more than enough force to explain the measured >60 pN forces generated during packaging. The predicted $|\Delta G_{EC}|$ is actually about twice the ~130 pN·nm free energy released by hydrolysis of a single ATP molecule, suggesting hydrolysis would not provide enough energy to return from the compact state to the extended state.

Several caveats regarding the calculations could potentially be responsible for this discrepancy. First, we approximate the extended state structure by simply displacing the center-of-mass of the C-terminal domain away from that of the N-terminal domain by 7 Å. In practice, changes in orientations of the domains may affect ΔG_{EC} values. Although not resolved in the cryo-EM structure, X-ray data for the N-terminal domain crystallized with ATP/ADP suggests it may rotate ~6º relative to the C-terminal domain[21]. Second, our calculations ignore the other components of the motor complex—neighboring gp17 subunits along with the threaded DNA molecule—whose configuration remains undetermined. These components likely occlude portions of the N- and C-terminal domain interface from getting solvated, causing our calculations to either over- or under-predict the solvation free energy. Third, our calculations neglect potential contributions from the configurational entropy of gp17 as discussed above. Finally, like all other biomolecular simulations these calculations are subject to the inherent inaccuracies in the force field representation of protein residues, solvent molecules, and ions. For these reasons, we expect the trend of changes in ΔG_{EC} with residue changes to be more meaningful than absolute ΔG_{EC} values.

2.4.5 Correlation of measured and computed metrics of motor impairment

To compare the measured and calculated metrics of motor function we calculated the Pearson correlation coefficients (0 = “no correlation”, 1 = “perfect correlation”) between each of the three measured quantities (Fig. 2.3), at each of the five increasing values of applied force, and the calculated ΔG_{EC} values for the wild type gp17 and three mutants. As shown in Fig. 2.3, we observe significant correlations in all

cases and the degree of correlation increases with increasing applied force. At 5 pN the packaging rate and motor velocity correlate only weakly with ΔG_{EC}, but both correlation coefficients increase to >0.9 at 50 pN. The correlation coefficient for time spent packaging also rises from ~0.65 to ~0.95 as the force is increased from 5 to 50 pN.

Thus, we have shown that the measured degree of motor impairment due to changes of residues proposed to be important in the model indeed correlates well with the predicted degree of impairment based on our calculations. The strength of correlation increases at high force, where the motor is subject to higher load, consistent with the notion that the E-C transition is responsible for force generation. Several features, such as the double mutant E537K-K305D exhibiting less impairment than single mutant E303K, are reconciled by these calculations. The measured % times spent packaging and motor velocities also asymptote towards zero with increasing force in the order E303K (first), E537K, E537K-K305D, then wildtype (last), consistent with the order of increasing magnitudes of the calculated favorable free-energy changes.

2.4.6 Mechanochemical energy landscape model of gp17 packaging

Based on our findings above we propose a minimal free energy landscape model of gp17 motor mechanochemistry (Fig. 2.4) that provides additional insights into gp17 dynamics. In this model, the free energy landscape (G_{tot}) governing gp17 motor function under applied loads is made up of contributions from the intrinsic conformational free energy landscape of gp17 (G_{gp17}), the chemical free energy of the

ATP hydrolysis cycle (G_{ATP}), and the mechanical work performed by gp17 against the applied load during DNA translocation (W), i.e., $G_{tot} = G_{gp17} + G_{ATP} + W$.

We consider the simplest "tight coupling" model in which ATP hydrolysis is assumed to result in one DNA translocation step. The energetics of ATP hydrolysis by motors with structurally homologous ATPase domains has been well characterized and has been shown to consist of two main steps that release energy: the ATP binding and product release steps[22]. The approximate G_{ATP} profile based on Ref. 22 is sketched as the blue line in Fig. 2.4; in these conditions, the ATP binding and product release steps yield ~15 $k_B T$ and ~10 $k_B T$ energies, respectively.

We propose a minimal model consisting of three conformational states of gp17 (Fig. 2.4A). Before ATP binding, gp17 is in an Apo state (**A**) whose structure is not known. In the first step, ATP binds and provides the energy to drive the transition from the Apo state to the extended state (**E**), whereupon it binds to the DNA. In the second step, ATP is hydrolyzed to ADP and P_i and gp17 transitions from the extended state to the compact state (**C**), and translocates the DNA by 2 bp. In the third step, ADP and P_i are released and the DNA is passed onto the next gp17 subunit in the motor ring (alternatively, P_i could be released earlier during the second step, as suggested by studies of the phage phi29 motor[11]). The original gp17 subunit goes back to the **A** state. This order of conformational-chemical transitions allows for the **E** → **C** transition —that goes "downhill" in free energy, as observed from our free energy calculations, and requires no external energy—to occur during the ATP hydrolysis step that itself releases little or no energy[22]. This then allows for the ATP binding and product release steps, the two energy-producing chemical transitions, to be utilized for pushing gp17 towards the energetically unfavorable extended state (**A** → **E**) and releasing gp17 from

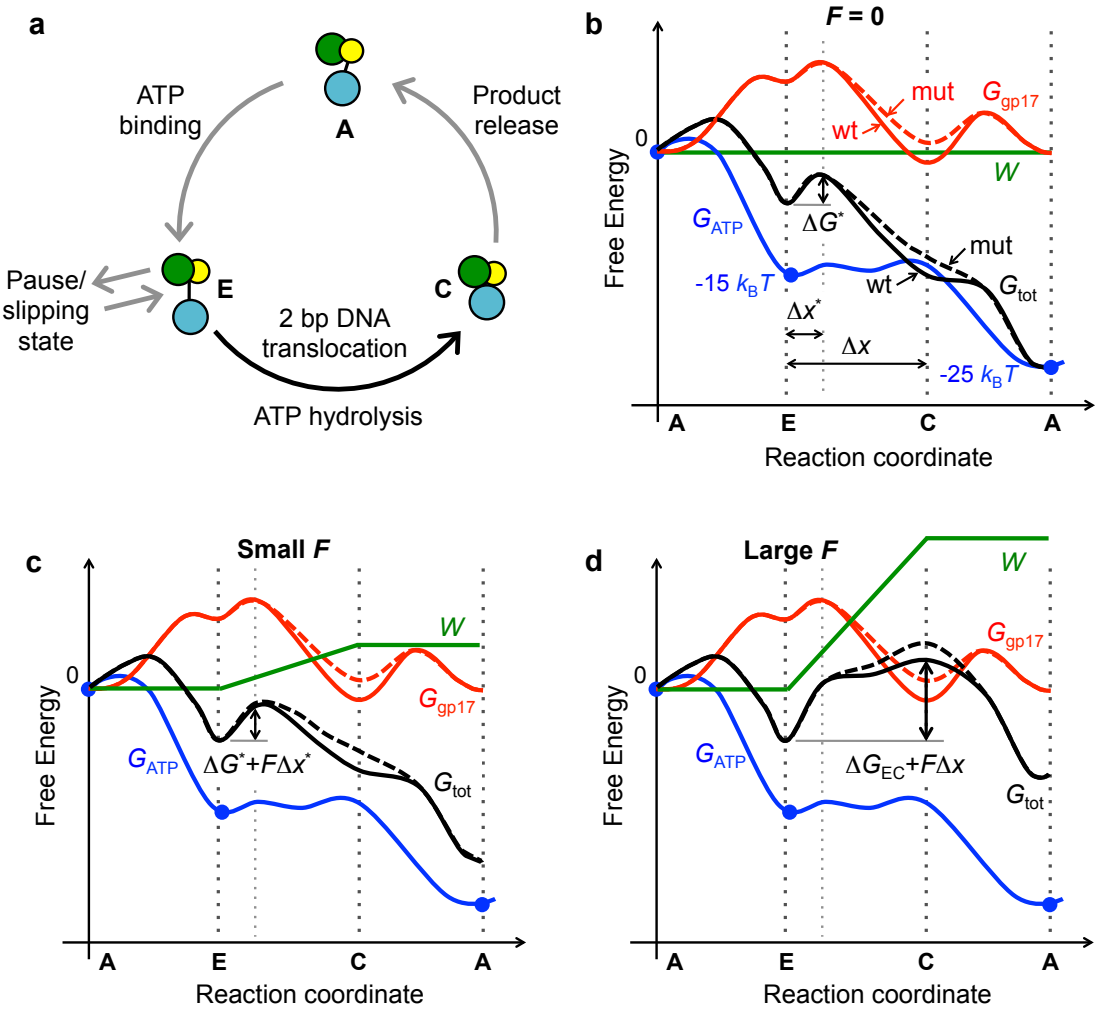


Fig. 2.4. Mechanochemical energy landscape model inferred from experimental and computational observations. (A) Proposed coupling between conformational transitions in gp17 and chemical transitions in ATP during a 2 bp DNA translocation cycle driven by the hydrolysis of one ATP molecule. The pausing and slippage of DNA is represented as an alternative state branching off from and rejoining at the extended state of gp17. (B-D) The conformational energy landscape of gp17 (red line), the ATP hydrolysis free energy profile (blue line), and mechanical work (green) add up to yield the overall mechanochemical energy landscapes. The shapes of the different energy landscape are schematically drawn for three cases: zero load (B), small load (C), and large applied loads (D).

its energetically favorable compact state (**C** → **A**), respectively. This proposed G_{gp17} profile is shown in Fig. 2.4B as the red line. Note that based on our simulations we only know that G_{gp17} decreases during the **E** → **C** transition, with free energy difference

$\Delta G_{EC} \equiv G_{gp17}(\mathbf{C}) - G_{gp17}(\mathbf{E}) < 0$, and increases during the **A** → **E** and **C** → **A** transitions, possibly involving energy barriers, but the magnitudes of the free energy differences and the heights and locations of the energy barriers are not known.

The mechanical work W is given by the applied force F multiplied by the distance the DNA is translocated against the force. During **A** → **E** transition, $W = 0$, as there is no DNA packaging during this transition. During **E** → **C** transition, $W = F \cdot x$, where x is the distance the DNA moves during this transition. At the end of this transition, the net amount of work performed is $W = F \cdot \Delta x$, where $\Delta x = 2$ bp. No additional work is performed during **C** → **A** transition, and the work remains at $W = F \cdot \Delta x$. The resulting W profile is shown as the orange line in Fig. 2.4 B-D.

In the absence of a load force (Fig. 2.4B), the overall free energy profile G_{tot} exhibits a largely downhill path, with the exception of unresolved energy barriers, along each of the three conformational transitions, which facilitates efficient packaging of DNA. As described earlier, this is made possible by the constructive coupling between conformational and chemical transitions, which allows for efficient utilization of the energy released from ATP hydrolysis. Overall, each DNA translocation step of 2 bp leads to a net reduction in the free energy of the system (gp17 + solution) equal to the total free energy released by the hydrolysis of one ATP molecule, i.e., $\Delta G_{ATP} \cong 25\ k_B T$.

The observed reduction in the motor velocity with increasing load force F (Fig. 2.3C) suggests the presence of an energy barrier between the **E** and **C** states. The applied force, due to the mechanical work W, increases this barrier height by an amount $\sim F \cdot \Delta x^*$, where Δx^* is the location of the barrier relative to the **E** state along the force direction. Assuming an Arrhenius-like dependence in transition rates across barriers, the DNA translocation rate k_t (and hence the motor velocity) should then

approximately decrease with increasing F via $k_t \propto \sim\exp[-(\Delta G^* + F{\cdot}\Delta x^*)/k_B T]$, where ΔG^* denotes height of this energy barrier (Fig. 2.4C). However, based on our findings with gp17 mutants we propose that at sufficiently large F, the work term tilts the energy landscape between **E** and **C** states to such a large extent that the free energy of state **C** now becomes higher than that at the original barrier, making the crossing of state **C** the rate limiting step (Fig. 2.4D). The relevant barrier height then becomes equal to the free energy difference between states **E** and **C**, as given by $\Delta G_{EC} + F{\cdot}\Delta x$. The rate of DNA translocation should consequently reduce via $k_t \propto \sim\exp[-(\Delta G_{EC} + F{\cdot}\Delta x)/k_B T]$.

Our measurements show that the mutant motors generate smaller packaging forces than wild type motors. This effect can be understood based on our free-energy calculations, which predict the **E** → **C** transition as being less favorable for mutants, leading to larger state **C** free energies for the mutants as compared to the wild type motor, as depicted in Fig. 2.4B-D. In addition to having lower stall forces (<60 pN), the mutants also exhibit slower rates of packaging as compared to wild type motors, and this difference becomes more significant at high loads. This can be explained by our model, which predicts that the rate of DNA translocation varies as $k_t \propto \sim\exp[-(\Delta G^* + F{\cdot}\Delta x^*)/k_B T]$ at low loads and as $k_t \propto \sim\exp[-(\Delta G_{EC} + F{\cdot}\Delta x)/k_B T]$ at high loads. That the mutants and wild type motors exhibit smaller differences at low forces implies that they both exhibit similar energy barriers ΔG^* (Fig. 2.4C). In contrast, the stronger effect of mutations on the packaging rate at high loads is consistent with the mutations decreasing the magnitude of ΔG_{EC} (Fig. 2.4D).

The observed increase in pausing and slipping of DNA during translocation with high applied force and with mutations (Fig. 2.2) can also be explained by the proposed energy landscape model. A recent study of T4 packaging at low ATP concentration

suggested the presence of a pause-“unpackaging” state in the mechanochemical kinetic cycle that branches off from the DNA translocation pathway. We further propose here that the motor switches to this alternate pathway while residing at the **E** state, which represents the most stretched conformation of the motor resisting the applied load (Fig. 2.4A). It then follows that processes slowing down the **E** → **C** transition rate, such as the applied load and the mutations studied here, would increase the time that the motor resides in the **E** state. This in turn increases the chances of the motor entering the proposed pause-unpackaging state, leading to increased rates of pausing and slipping, which is consistent with our observations.

2.5 Conclusions

Using a combination of single-molecule dynamic packaging measurements with increasing applied load forces, site-directed mutagenesis, and computational analysis we have quantitatively demonstrated the importance of electrostatic interactions between complimentarily charged residues between the N- and C-terminal subdomains in gp17-mediated dsDNA translocation. These findings support the T4 motor force-generation mechanism proposed by Sun et al. on the basis of structural data[17]. We show that our results with both wild type and mutant motors under a wide range of load forces can be rationalized, at least qualitatively, by an energy landscape model which couples the free energy profile of motor conformational states with that of the ATP hydrolysis cycle.

Many molecular motors rely on large-scale structural rearrangements to forcibly translocate cargo. These rearrangements are often coupled to structural changes in the vicinity ATP binding pocket during the ATP hydrolysis cycle, which cause allosteric effects that may be leveraged to produce large-scale motions.

Although coordination of multiple parallel electrostatic interactions across an interface between two subdomains has not previously been demonstrated as a mechanism to directly produce motor force, evidence suggests that both myosin[23] and FoF1-ATP synthase[24] rely on electrostatic interactions between single pairs of charged residues to initiate structural transitions mediated by polar and hydrophobic interactions. As the T4 motor is homologous with many other ring-shaped nucleoside triphosphatases, including enzyme complexes involved in DNA replication and transcription (helicases), protein degradation, and chromosome segregation, the force generation mechanism investigated here may well have broader applications.

2.6 Acknowledgements

This work was carried out as a collaboration between Dr Smith in physics, and Dr Arya in nanoengineering. We would like to thank Dr Dario Meluzzi for help with distributed computing, as well as Vishal Kottadiel (previously of Dr Rao's group) for helpful conversations concerning gp17 mutants and experimental methods. We would also like to thank Damien DelToro for his work on the optical trapping apparatus. This work was supported by NIH grant R01-GM088186, NSF grant PHY-0848905, as well as supported by National Science Foundation grant MCB-0923873 and in part, by the National Institute of Allergy and Infectious Diseases grant R01AI081726. The author would also like to acknowledge coauthors of this work Nicholas Keller, Tanfis I. Alam, Marthandan Mahalingam, Venaglia B. Rao, Gaurav Arya, and Douglas E. Smith.

References

1. Jardine, P.J. & Anderson, D.L. In *DNA packaging in double-stranded DNA bacteriophages.* (ed. Calendar, R.) 49 (Oxford University Press, New York, NY, 2006).

2. Feiss, M. & Rao, V.B. The bacteriophage DNA packaging machine. In *Viral Molecular Machines* 489-509 (Springer, 2012).

3. Morais, M.C. The dsDNA packaging motor in bacteriophage ø29. In *Viral Molecular Machines* 511-547 (Springer, 2012).

4. Smith, D.E. Single-molecule studies of viral DNA packaging. *Current Opinion in Virology* **1**, 134 (2011).

5. Tzlil, S., Kindt, J.T., Gelbart, W.M. & Ben-Shaul, A. Forces and pressures in DNA packaging and release from viral capsids. *Biophys. J.* **84**, 1616-1627 (2003).

6. Harvey, S.C., Petrov, A.S., Devkota, B. & Boz, M.B. Viral assembly: a molecular modeling perspective. *Physical Chemistry Chemical Physics* **11**, 10553-10564 (2009).

7. Smith, D.E. Sander J. Tans, Steven B. Smith, Shelley Grimes, Dwight L. Anderson, and Carlos Bustamante. The bacteriophage phi29 portal motor can package DNA against a large internal force. *Nature* **413**, 748-752 (2001).

8. Fuller, D.N. Dorian M. Raymer, John Peter Rickgauer, Rae M. Robertson, Carlos E. Catalano, Dwight L. Anderson, Shelley Grimes, and Douglas E. Smith. Measurements of single DNA molecule packaging dynamics in bacteriophage lambda reveal high forces, high motor processivity, and capsid transformations. *J. Mol. Biol.* **373**, 1113-1122 (2007).

9. Fuller, D.N., Raymer, D.M., Kottadiel, V.I., Rao, V.B. & Smith, D.E. Single phage T4 DNA packaging motors exhibit barge force generation, high velocity, and dynamic variability. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 16868-16873 (2007).

10. Rickgauer, J.P. Derek N. Fuller, Shelley Grimes, Paul J. Jardine, Dwight L. Anderson, and Douglas E. Smith. Portal motor velocity and internal force resisting viral DNA packaging in bacteriophage phi29. *Biophys. J.* **94**, 159-167 (2008).

11. Moffitt, J.R. Yann R. Chemla, K. Aathavan, Shelley Grimes, Paul J. Jardine, Dwight L. Anderson, and Carlos Bustamante. Intersubunit coordination in a homomeric ring ATPase. *Nature* **457**, 446-450 (2009).

12. Chistol, G. Shixin Liu, Craig L. Hetherington, Jeffrey R. Moffitt, Shelley Grimes, Paul J. Jardine, and Carlos Bustamante. High degree of coordination and division of labor among subunits in a homomeric ring ATPase. *Cell* **151**, 1017-1028 (2012).

13. Chemla, K. Aathavan, Jens Michaelis, Shelley Grimes, Paul J. Jardine, Dwight L. Anderson, and Carlos Bustamante. Mechanism of force generation of a viral DNA packaging motor. *Cell* **122**, 683-692 (2005).

14. Rao, V.B. & Feiss, M. The Bacteriophage DNA Packaging Motor. *Annu. Rev. Genet.* **42**, 647-681 (2008).

15. Tsay, J.M., Sippy, J., Feiss, M. & Smith, D.E. The Q motif of a viral packaging motor governs its force generation and communicates ATP recognition to DNA interaction. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 14355-14360 (2009).

16. Tsay, J.M. Jean Sippy, Damian delToro, Benjamin T. Andrews, Bonnie Draper, Venigalla Rao, Carlos E. Catalano, Michael Feiss and Douglas E. Smith. Mutations altering a structurally conserved loop-helix-loop region of a viral packaging motor change DNA translocation velocity and processivity. *J. Biol. Chem.* **285**, 24282-24289 (2010).

17. Sun, S. Kiran Kondabagil, Bonnie Draper, Tanfis I. Alam, Valorie D. Bowman, Zhihong Zhang, Shylaja Hegde, Andrei Fokine, Michael G. Rossmann, and Venigalla B. Rao. The Structure of the Phage T4 DNA Packaging Motor Suggests a Mechanism Dependent on Electrostatic Forces. *Cell* **135**, 1251-1262 (2008).

18. Mancini, E.J. Denis E. Kainov, Jonathan M. Grimes, Roman Tuma, Dennis H. Bamford, and David I. Stuart. Atomic snapshots of an RNA packaging motor reveal conformational changes linking ATP hydrolysis to RNA translocation. *Cell* **118**, 743-755 (2004).

19. Zhao, H., Christensen, T.E., Kamau, Y.N. & Tang, L. Structures of the phage Sf6 large terminase provide new insights into DNA translocation and cleavage. *Proceedings of the National Academy of Sciences* **110**, 8075-8080 (2013).

20. Massova, I. & Kollman, P.A. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug Discov. Des.* **18**, 113-135 (2000).

21. Sun, S., Kondabagil, K., Gentz, P.M., Rossmann, M.G. & Rao, V.B. The structure of the ATPase that powers DNA packaging into bacteriophage t4 procapsids. *Mol. Cell* **25**, 943-949 (2007).

22. Oster, G. & Wang, H. Reverse engineering a protein: the mechanochemistry of ATP synthase. *Biochim Biophys Acta* **1458**, 482-510 (2000).

23. Fischer, S., Windshügel, B., Horak, D., Holmes, K.C. & Smith, J.C. Structural mechanism of the recovery stroke in the myosin molecular motor. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 6873-6878 (2005).

24. Ito, Y., Oroguchi, T. & Ikeguchi, M. Mechanism of the conformational change of the F1-ATPase β subunit revealed by free energy simulations. *J. Am. Chem. Soc.* **133**, 3372-3380 (2011).

25. Fuller, D. N., Gregory J. Gemmen, John Peter Rickgauer, Aurelie Dupont, Rachel Millin, Pierre Recouvreux and Douglas E. Smith. A general method for manipulating DNA sequences from any organism with optical tweezers. *Nucleic Acids Research* **34,** e15 (2006).

26. Jardine, P. J. & Anderson, D. L. The Bacteriophages - Ohio State University . Stephen T. Abedon Professor of Molecular and Cell Biology Department of Microbiology - Google Books. *The Bacteriophages* (2006).

27. Zhang, Z., Vishal I. Kottadiel, Reza Vafabakhsh, Li Dai, Yann R. Chemla, Taekjip Ha, Venigalla B. Rao. A promiscuous DNA packaging machine from bacteriophage T4. *PLoS Biol* **9,** e1000592 (2011).

28. Leffers, G. & Rao, V. B. Biochemical Characterization of an ATPase Activity Associated with the Large Packaging Subunit gp17 from Bacteriophage T4. *Journal of Biological Chemistry* **275,** 37127-37136 (2000).

29. Kondabagil, K. R., Zhang, Z. & Rao, V. B. The DNA Translocating ATPase of Bacteriophage T4 Packaging Motor. *Journal of Molecular Biology* **363,** 786–799 (2006).

30. Massova, I. & Kollman, P. A. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding - Springer. *Perspectives in Drug Discovery and Design* **18,** 113–135 (2000).

31. Phillips, J., Isgro, T., Sotomayor, M. & Villa, E. NAMD TUTORIAL. (2003).

32. Nelson, M. T. William Humphrey, Attila Gursoy, Andrew Dalke, Laxmikant V. Kalé, Robert D. Skeel and Klaus Schulten. NAMD: a Parallel, Object-Oriented Molecular Dynamics Program. *International Journal of High Performance Computing Applications* **10,** 251–268 (1996).

33. MacKerell, A. D., Banavali, N. & Foloppe, N. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* **56,** 257–265 (2000).

34. Pearlman, D. A. David A. Case, James W. Caldwell, Wilson S. Ross, Thomas E. Cheatham III, Steve DeBolt, David Ferguson, George Seibel, Peter Kollman. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications* **91,** 1–41 (1995).

35. Onufriev, A., Bashford, D. & Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **55,** 383–394 (2004).

36. Jayaram, B., Sprous, D. & Beveridge, D. L. Solvation Free Energy of Biomacromolecules: Parameters for a Modified Generalized Born Model Consistent with the AMBER Force Field. *J. Phys. Chem. B* **102,** 9571–9576 (1998).

37. Sitkoff, D., Sharp, K. A. & Honig, B. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J. Phys. Chem.* **98,** 1978–1988 (1994).

38. Karplus, M., Ichiye, T. & Pettitt, B. M. Configurational entropy of native proteins. *Biophysical Journal* **52,** 1083–1085 (1987).

3 Interactions between N- and C- terminal subdomains in the bacteriophage T4 packaging motor. I. Electrostatic contributions.

Amy Migliori[1], Gaurav Arya[3], and Douglas Smith[2]

3.1 Summary

Bacteriophages, such as T4, infect host cells by injecting their genome through the cell wall of a host bacteria; to allow for successful delivery, the DNA is highly pressurized within the capsid. To pressurize the genome, bacteriophages “package” the DNA into the capsid by pushing it in with a translocation motor capable of high forces. In T4, the motor is composed of five gp17 motor proteins arranged as a pentamer around a pore in the capsid. Each gp17 protein translocates DNA using an ATP-dependent process that is hypothesized to function as a molecular ratchet. For each ATP that is hydrolyzed, 2 base pairs (bp) of the 171 kbp T4 genome are translocated into the capsid by large-scale structural changes in gp17. These structural changes are proposed to be mediated by formation of a complementary interface between two subdomains. This important interface between N- and C-terminal subdomains contains several charged amino acids, proposed to form ion pairs during the “power stroke” of the motor, and generating the force needed to translocate DNA against internal resistance during packaging. We have shown experimentally and computationally that mutation of these amino acids results in a reduction in packaging activity, though the degree of reduction is strongly dependent on position. In this manuscript, we investigate the electrostatic contributions to the subdomain interface in

[1]Department of Biochemistry, University of California, San Diego. La Jolla, CA, 92037.

[2]Department of Physics, University of California, San Diego. La Jolla, CA, 92037.

[3]Department of Nanoengineering, University of California, San Diego. La Jolla, CA, 92037.

gp17. We use computational models to show that only two of five possible ion pairing interactions form the proposed inter-subdomain interactions, while other charged amino acids stabilize secondary structural elements. We also reveal that electrostatic contributions alone do not account for the observed interface behaviors. This strongly suggests that a combination of water-mediated, electrostatic, and hydrophobic interactions work in concert to allow the ratcheting motion that packages DNA into the **T4** capsid. We expand on these non-electrostatic contributions in a later manuscript.

Keywords: MM-GBSA, protein-protein interaction, ion pair, electrostatic

3.2 Introduction

Molecular translocation motors are common within the microscopic world of cells, bacteria, and viruses. Helicase, a ubiquitous class of motors crucial for DNA replication, translocates DNA to unzip the double helix and makes base pairs accessible to replication machinery[1]. Myosin, another motor, allows for the fibrils in muscle to generate a contractile motion[2,3]. Eukaryotic mitochondria employ translocases to import membrane proteins[4,5]. The kinesin family of proteins perform translocation during cellular replication[6]. Most cellular motors are able to generate only the small forces necessary to undertake their particular task; however, few motors are able to generate forces matching those of the genome translocation motors employed by bacteriophages. Because bacteriophages infect a host cell via injection of the genome, it is necessary for the capsid to contain the genetic material at very high internal pressures. Producing these high internal pressures requires a substantial amount of packing force; the molecular motor that generates these forces in

bacteriophage T4 is gp17, an ATP-dependent, motor protein that assembles into a pentamer around a specific vertex of the icosahedral capsid.

To understand the mechanism by which gp17 packages DNA, we combined structural and experimental information to test a proposed model of action that describes the motor as a molecular ratchet[7]. This model describes the mechanism of packaging as an electrostatically-driven conformational change, catalyzed by ATP consumption, that forms a close interface between N- and C-terminal subdomains (named for their proximity to the N- and C- terminals) to translocate DNA. In Chapter 2, we showed that mutation of select charged amino acids located in the interface between the N- and C-terminal subdomains is able to affect both the speed and force generation capabilities of the T4 motor. We also showed an excellent correlation between experimental and computational measurements, suggesting that our computational model accurately describes the motor dynamics. However, we wish to understand the mechanisms of interface formation and stability in more detail, and to explain the balance between different interactions that give rise to the ratcheting mechanism. Specifically, we would like to understand whether electrostatic forces alone account for the observed behavior in the gp17 motor protein.

Table 1 shows the electrostatic and hydrophobic free energy contributions to the overall change in free energy of compaction for the constructs studied in Chapter 2. The nearly equal balance between these two energies in wild type gp17 shows the complexity of the compaction step is far greater than originally thought based on simple ion pairing. This was an indication that electrostatic interactions might not completely describe the protein dynamics.

Table 3.1. Free energy values for wild type as well as studied mutants.
Free energy was calculated using molecular dynamics simulation data discussed in Chapter 2. Each value reported here is the change in free energy experienced upon gp17 going from extended to compact forms. Charged amino acid mutations were also studied experimentally.
The electrostatic contribution was calculated by sum of $\Delta G_{ele} + \Delta E_{GB}$. The hydrophobic contribution was calculated by sum of $\Delta G_{VDW} + \Delta E_{sol}$. ΔG overall was calculated as described in Chapter 2 methods. All numbers are in units of kcal/mol.

	electrostatic (kcal/mol)	**hydrophobic (kcal/mol)**	**ΔG (kcal/mol)**
wild type gp17	-18.37 ± 15.64	-21.67 ± 4.14	-38.21 ± 4.42
E537K	6.27 ± 16.1	-14.23 ± 3.32	-6.14 ± 3.44
E303K	-2.31 ± 17.26	3.46 ± 5.25	-2.79 ± 5.55
E537K-K305D	2.66 ± 17.98	-22.69 ± 9.79	-19.95 ± 9.99

The interface region (defined here as residues with at least one atom within 5Å of the other subdomain) of our computational model of gp17 contains 46 amino acids in total. The residues making up the interface are shown in Figure 3.1. This interface region contains 23 hydrophobic residues, 12 charged residues, and 11 polar residues, any of which could be involved in generating motor activity based on their proximity to the region. To fully understand the interface that gives rise to the properties of this motor, we clarify the role of each participating amino acid. We also explain why some charged amino acids in the interface contribute significantly less to motor activity than others.

To understand the role of each amino acid in the interface, we must analyze interactions between non-sequential amino acids. These interactions are crucial in forming secondary, tertiary, and quaternary structures of proteins. Interactions between

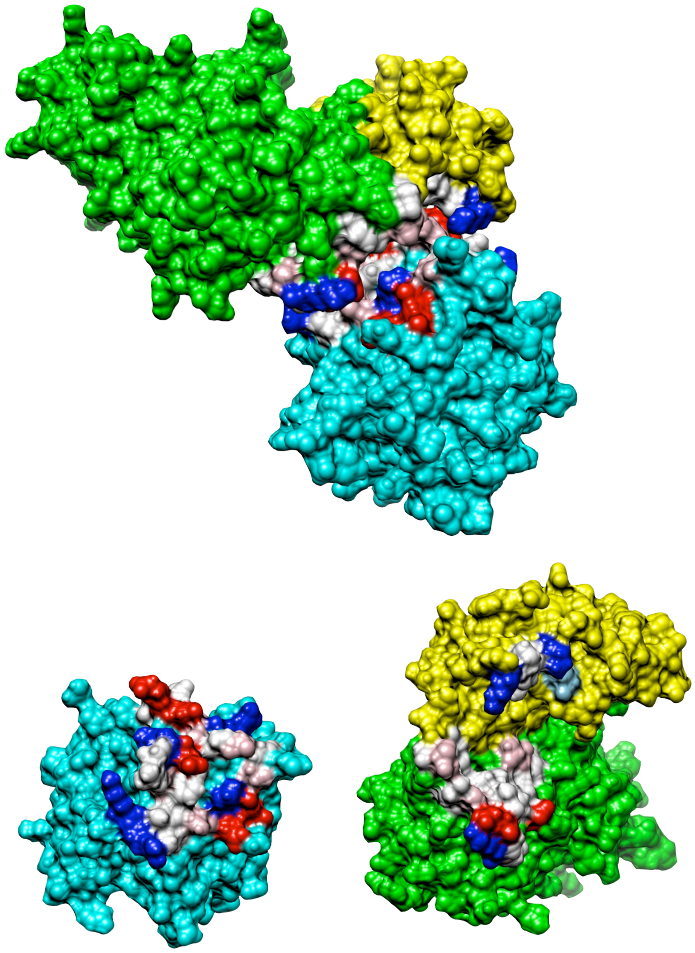


Fig 3.1. The composition of the interface formed between N- and C-terminal subdomains in our computational model. Top : gp17, colored by subdomain and interface residue type. Green, N-I subdomain. Yellow, N-II subdomain. Cyan, C terminal subdomain. Blue residues, basic. Light blue, partial positive polar residue. Red residues, acidic. Pink residues, electronegative polar. White, hydrophobic residues. The interface is defined as residues on a subdomain (ie N or C) having one or more atom falling less than 5Å from the other subdomain. Bottom : open book diagram of top, showing the interaction interface between the two subdomains.
Images generated using Chimera.

amino acids can occur either locally within the surrounding amino acid sequence, or across large distances, and may be specific or non-specific. Specific interactions are caused by the electrostatic forces between charged residues. Non-specific interactions include the van der Waals interaction, which prevents two hydrophobic residues from occupying the same space. In a protein, hydrophobic residues exclude water and are

drawn together via non-specific interaction, much like oil droplets, during the initial stages of folding[8]. At the same time, polar and charged residues form specific interactions as a result of the electrostatic forces. The sum of these interactions work to stabilize structural elements. We examined these interactions to predict which of them, ion-pair or otherwise, are important.

To determine the electrostatic contribution to the formation of the interface, we need to determine what type of interaction is occurring for each possible ion pair. Ion pair interactions in proteins are extremely distance-dependent due to a combination of Coulomb's law and desolvation energy (the energy required to expel water molecules). These two factors working in concert mean that there are three general classes of electrostatic interactions between charged amino acids in a proteins, depending on the number of residues involved and the distances between them. Glutamic acid and aspartic acid, the only negatively charged natural amino acids, are carboxylic acids with two equivalent terminal oxygen atoms, while positively charged amino acids may have one (lysine) or two (histidine and arginine) equivalent terminal nitrogen atoms. Thus, each ion pair interaction involves 3-4 atoms. A salt bridge, the strongest class of ion pair interaction, occurs if the centroids of the charged-group atoms of both residues are within 4Å of each other. These are almost always energetically favorable and structurally stabilizing. Another class, N-O bridges, occur when only one pair of nitrogen and oxygen atoms are within 4Å of each other, and are usually favorable (depending on the conformation and solvation state of the two residues). Last, interactions between charged amino acids that have a distance between both charged atom pairs greater than 4Å are generally considered destabilizing, due to the large change in solvation needed to form the weak interaction. Illustrations of these three

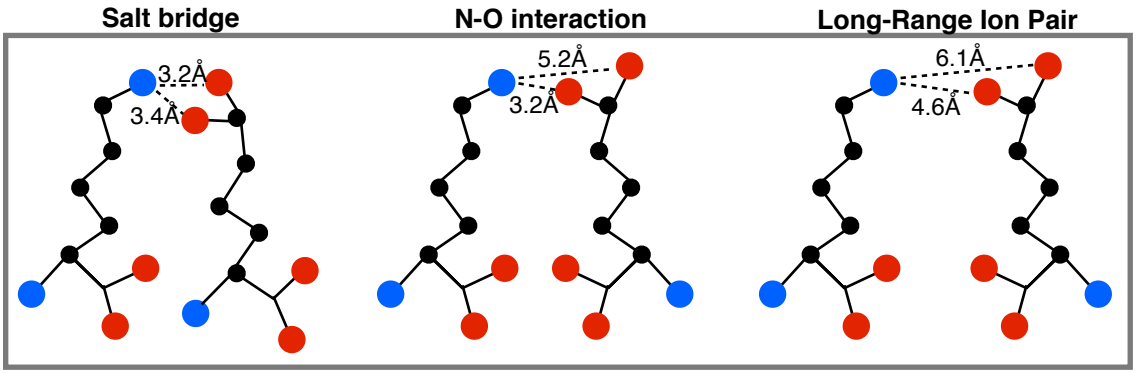


Fig 3.2. Comparison of distances between charged atoms in three types of protein electrostatic interactions. (A) An ion pair interaction is classified as a salt bridge if the centroids of the charged-group atoms of both residues are within 4.0Å of each other and the distance between charged atom and the interatomic distance of at least one pair is less than 4Å. These interactions are nearly always favorable. (B) N-O interaction, the interatomic distance of one pair is less than 4Å, but the distance between centroids is greater than 4Å. These interactions are favorable in many situations. (C) Long-range ion pair, in which the interatomic distance of both pairs is greater than 4Å. Long-range ion pairs are frequently destabilizing and nearly always unfavorable.

Figure adapted with permission from Kumar and Nussinov[8].

types of ion pairing interactions are shown in Figure 3.2[9].

Solvation is also an important factor in the stability of amino acid interactions formed at an interface. Water is a remarkable molecule, and is a crucial part of all interactions between biomolecules. When forming an interface between two proteins, or closing a cleft within a single protein, much of the water must be expelled. This may be energetically favorable, in the case of previously solvent-exposed hydrophobic residues becoming buried, or energetically unfavorable, in the case of polar or charged amino acids losing their associated water molecules. Water also directly impacts the folding of proteins into their native conformation, where it actively helps proteins search for their native state as they navigate the energy landscape[10] by transferring small amounts of energy into the process in a random way. Some water molecules are so

stable in their position around a protein that they show strong densities in X-ray crystallographic structures!

Solvation affects the electrostatic interactions because charged amino acids that are solvent accessible form hydrogen bonds with water. The number of potential hydrogen bonds is related to the number of electrostatic atoms on the side chain. For example, the minimum free energy of solvation for lysine, which has 1 terminal nitrogen atom, occurs in the presence of 3 water molecules, while arginine, which has 3 terminal nitrogen atoms, requires 5. Glutamic acid and aspartic acid, the two negatively charged amino acids both having two equivalent terminal oxygen atoms, ideally form 6 hydrogen bonds[11]. These recent data compare well with classical hydration potential calculations[12]. Surprisingly, the formation of salt bridges in a completely solvated environment has been shown to be universally unfavorable due to the large desolvation penalty required to form the interaction. There is, however, a relationship between the pair that is formed and the favorability of interaction. Interactions between arginine and glutamic acid form the most favorable solvent-exposed salt bridge, while an interaction between arginine and aspartic acid prove the least favorable[11]. An ideal salt bridging relationship occurs when only one water is removed to allow for the ion pairing relationship to occur. Each additional desolvation leads to further energetic costs, making the interaction less and less favorable. An ideal salt bridge therefore contributes a great deal to the free energy while retaining the majority of the original hydrogen bonding with water, while one forming in conjunction with additional desolvation may be destabilizing to the complex.

In our studies of gp17, we have analyzed each of these fundamental interactions to identify the important residues contributing to the formation of the

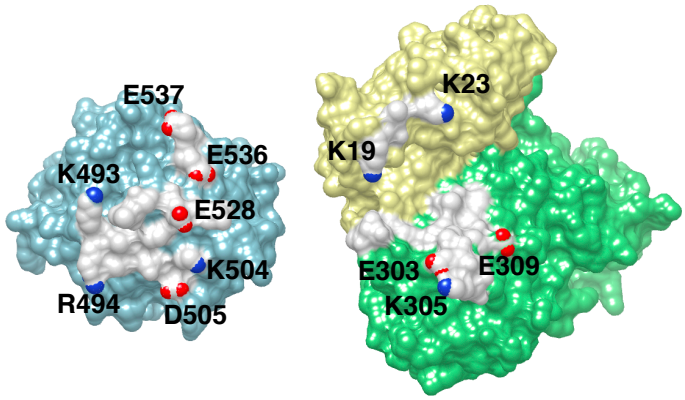


Fig 3.3. Charged residues at the interface between N- and C-terminal subdomains in the gp17 crystal structure. This open book diagram recreates initial structural information from Sun, et al. using the crystal structure, and showing potential ion pairing interactions across the interface. While there appears to be excellent alignment between charged residues at all positions, we find that only two of these pairs form interactions, K19-E537, and E309-R494. Yellow, subdomain N-II; green, subdomain N-I; cyan, C terminal subdomain. White, all residues falling within the interface region. Red, oxygen atoms of acidic amino acids at the interface. Blue, nitrogen atoms of charged amino acids at the interface.

interface via electrostatic interactions of any type. The crystal structure of gp17 shows the interface between the N- and C- subdomains to have an area of ~1000Å^2, which hints at a certain degree of stability in the compacted form. Additional structural information has been obtained by “stalling” viral particles during the process of DNA packaging, and determining the cryo-electron microscopic structure[7] at that point. The stalled structure indicates that during packaging, gp17 is extended 7Å from its crystalline form, defining the “extended” state. Visual inspection of the interface region reveals the presence of 12 charged amino acids arranged in such a way to suggest that five salt bridge interactions are formed between oppositely charged amino acids in the compact structure, shown in Figure 3.3. This proximity suggests that these charged amino acids would be directly involved in force generation via electrostatic forces, forming potential interactions between K19-E536, K23-E537, E303-K504, K305-D505, and E309-K493 or R494.

To probe ion pairing interactions between all residues in the interface, we measured the distances between charged atoms over the course of molecular dynamics simulations and determined the free energy contribution of each residue involved in ion pairing. In addition, we investigated how change in solvation upon compaction effects the interactions between charged and hydrophobic/polar residues in the interface. By combining these methods to describe the electrostatic and solvent interactions in the compact and extended form of gp17, we determine the electrostatic contributions to the effects observed in our mutational studies using computational methods, and reveal which ion-pair interactions contribute substantially to the molecular action of gp17.

3.3 Materials and Methods

Chapter 2 described the relationship between the activity of gp17 and mutations to charged residues, as well as the free energy change of compaction calculated from molecular dynamics simulations of gp17. We showed very strong correlation between activity at increasing forces and free energy in wild type gp17 and charge mutations E537K, E303K, and E537K-K305D, indicating that these charge pairs are important for activity. The data presented in this chapter is based on the wild type molecular dynamics simulations described in detail in Chapter 2.

3.3.1 Free energy decomposition by MM-GBSA

The three free energy contributions were computed from 500 snapshots of each complex recorded during a 4 ns MD simulation of the complex, carried out using the *mm_pbsa.pl* script in AMBER 10[13]. As explained earlier, we defined the two

domains as separate molecules—“ligand” and “receptor”—while carrying out MM-GBSA calculations. Specifically, we defined the N- and C-terminal domains by residues 1–359 and 360–560, respectively. The ΔE_{MM} (change in molecular mechanics energy) contributions were directly computed as the change in the sum of the bonded, electrostatic, and van der Waals interaction energies. The ΔG_{sol} (change in free energy of solvation) contributions were further divided into polar and nonpolar contributions. The polar contributions were calculated using the Generalized Born [14]approach while the nonpolar contributions were calculated from changes in solvent accessible surface area upon binding multiplied by the surface tension. We used a water probe of radius 1.4 Å and a surface tension of 0.0072 kcal/mol/Å^2. The $T\Delta S_{conf}$ (entropic) contributions are typically computed using normal mode analysis[15], but these calculations converge extremely slowly; hence, we neglected entropic contribution to the free energy difference. However, we believe that these entropic contributions will mostly cancel out when comparing ΔG_{EC} (free energy of the change from extended to compact state) values across wild type and mutant gp17. The algorithms, force field parameters, and conditions used for these MD simulations were kept identical to those used for generating the initial conditions.

To obtain reliable values of ΔG_{EC} as well as statistical uncertainties, we carried out four sets of MM-GBSA calculations for each wild type and mutant complex, each starting from a different initial configuration of the extended and compact state complex prepared using the procedures described in the previous section.

Detailed information about molecular mechanics simulations, as well as about whole-protein energy calculations, may be found in Chapter 2 methods. To calculate the per-residue energy contribution, we used the decomposition function (DECOMP) in

MM-GBSA along with molecular mechanics and generalized born free energy calculations (the parameters for which may also be found in Chapter 2 methods). The parameters used were defined by the atoms present in the simulation. NTOTAL can be found in the topology file containing the solvated protein. MM-GBSA requires the input of 3 molecules: a complex (gp17), receptor, and ligand; it was designed to find the free energy of binding between separate molecules. Our system, gp17 compaction, had been cut into two pieces to satisfy this requirement. The number of atoms contained in each was defined by the particular construct (i.e. a mutation at K19 has a different number of N-terminal atoms than wild type). This provides an output that shows the free energy contributed by each residue in total, and is broken down into its components per residue. In this way, we are able to visualize which residues have significant overall free energy changes, and also where electrostatic forces are favorable or unfavorable, where solvent plays a major factor in stability, and which residues experience strong van der Waals forces.

3.3.2 Inter-atomic distance measurements

The distances between charged atoms in proposed salt bridges were measured using two methods. We measured the distance at each simulation time step in order to monitor the fluctuations during the course of a simulation by using a program called ptraj[16], found in AmberTools. In our simulations, we used 600 time steps, spaced at 10,000 fs. We also created a script called alldist.in that takes in our chosen simulation output files, and outputs the distance between a selected pair of atoms at selected intervals during the course of the simulation. An example of this script that has been

modified to measure the distance between any atoms pairwise can be found in Appendix C.

3.3.3 Water occupancy measurements

Some solvent molecules reside in stable positions throughout the course of a simulation. Similar to the water molecules that appear in the x-ray crystallography measurements, these “bound” waters show positions on the protein where water is likely to be involved in a physical setting as well. To find these waters in our computational data, we used Chimera[17]. First, we loaded the simulation coordinate file (.mdcrd) and parameter file (.prmtop) using the MD Movie tool. We fixed gp17 in place so that the water occupancy grid could identify water localizations. To compute the occupancy map, we used the included Analysis/Calculate Occupancy command in the software package Chimera. This command can be used to calculate the relative occupancy of the selected atoms. The output of this program is a pair of maps showing the water occupancy. One map shows the density of the hydrogen atoms, the other the density of the oxygen atoms. We adjusted the contour levels to be above the noise floor such that significant regions were visible. To visualize only high occupancy waters, we adjusted the levels such that only water present in position more than 30% of the time is visible.

3.3.4 Average structure generation

To find the relative locations of hydrophobic and polar residues, we calculated an average structure over the course of our simulations. We loaded our MD output files into Chimera as above, and used the Calculate Trajectory Average Structure tool. We

started at frame 1, and used a step size of 1 to include all snapshots from our simulation in calculating the average. It should be noted that this method was not used to find the distances between charged atoms, as it can introduce structural problems when amino acids have extensive flexibility. Therefore, its use was limited to determining the general locations of hydrophobic residues, and not for calculating exact distances of electrostatic interactions.

3.4 Results

3.4.1 Free energy decomposition reveals per-residue free energy contribution is correlated with experimental data.

In Chapter 2, we showed excellent correlation between the free energy change of compaction for wild type and 3 mutants, and experimental measurements of forward motor velocity, average velocity, percentage of time spent packaging, and the time the motor sustained the load. We wanted to be cross-verify that residues indicated to be important by free energy decomposition also related to experimental measurements. To do this, we analyzed the correlations between the same experimental metrics used in correlations in Chapter 2 and the free energy contributed by the mutated residue. For example, the measurements of E537K were correlated with the free energy change of E537 in wild type. The results of these correlations are shown as filled markers in Figure 3.3. The correlation is shown between experimental data and either the free energy change of the construct itself (open circles), or with the specific residue free energy (closed circles; in this case, residue E537). The observed level of correlation indicates that there is a strong relationship between the per-residue free

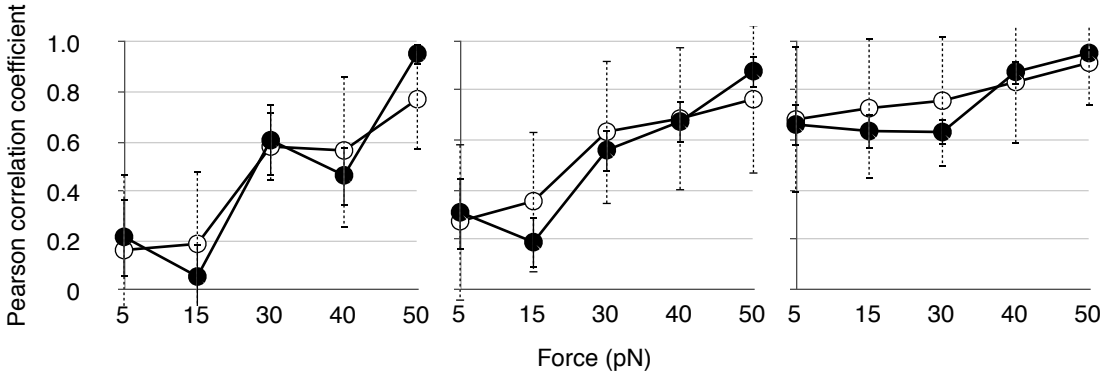


Fig. 3.4. Correlations between experimental metrics of motor function and calculated free energy changes.

Dependence on applied load of Pearson correlation coefficients (min=0, max=1) between free energy changes and (A) forward motor velocity (B) average velocity, and (C) % time packaging. Open markers indicated correlations with $\Delta G_{overall}$, and filled markers indicate correlations with $\Delta G_{residue}$.

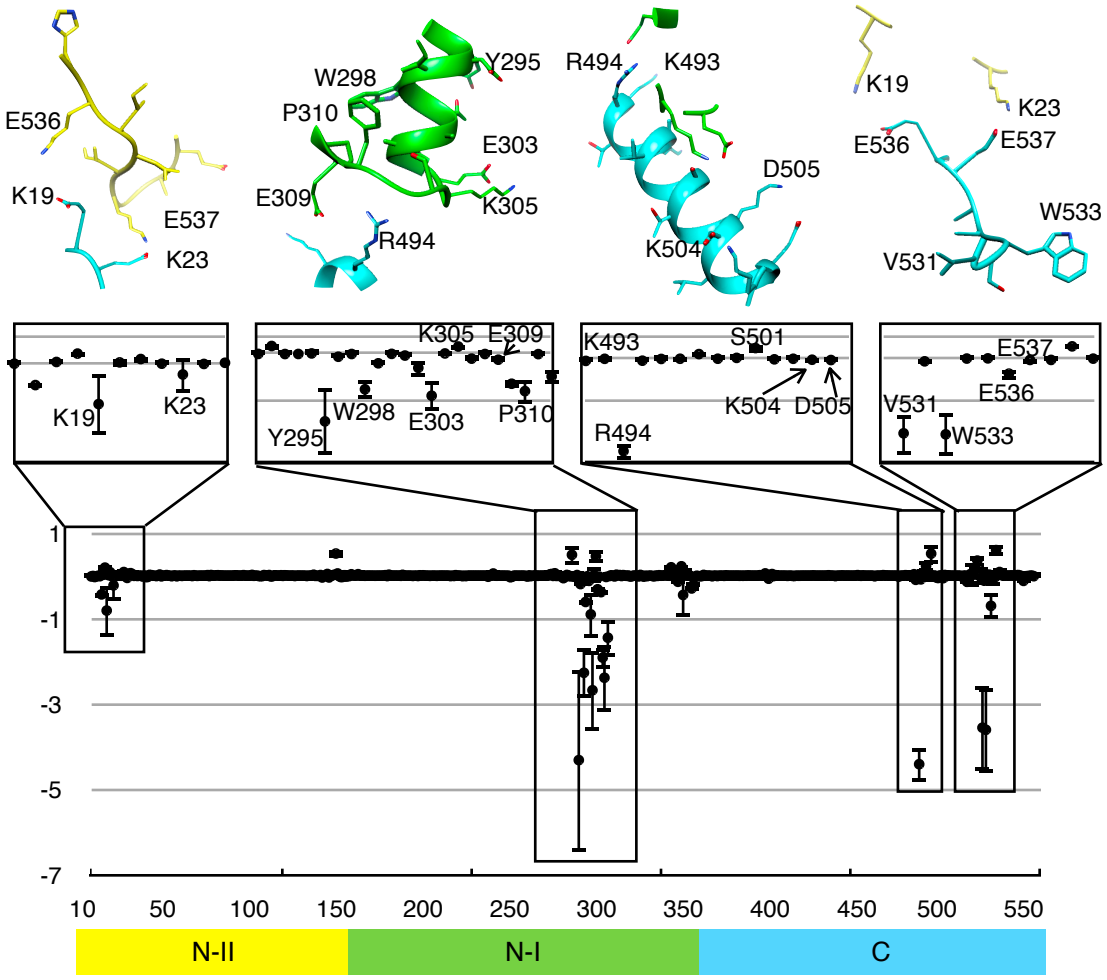
energy change of a mutated residue and the experimental outcome of mutation. Because of this relationship, we are able to deduce the likely experimental effect of mutation to other residues not previously studied by viewing their contribution to the wild type free energy change.

3.4.2 Free energy decomposition reveals important residues in motor compaction.

Fig 3.5. Wild type gp17 free energy decomposition by residue.

Top : Locations of important residues in relation to their binding partners. Important residues have been identified either by being proposed by Sun, et al to participate in salt bridge interactions, or by their large contribution to the overall free energy of compaction.
Top left : K19 and K23 interact with E536 and E537, respectively.
Top center left : Interactions between amino acid 290 and 312. E309 clearly interacts with E494, E303 with K305 (forming a short-range ion pair), and part of the pi-stacking arrangement can be seen round W298.
Top center right : Interactions between amino acids 493 and 505. The interaction between R494 and E309 is much more favorable than between K493 and E309. Also shown is the interaction formed between D505 and a negatively charged amino acid in the next turn of the helix, and a similar interaction at the adjacent amino acid, K504, with a positively charged amino acid in the next turn of the helix.
Top right : W533 interacts with other hydrophobic amino acids (not shown), whereas V531 becomes substantially less solvent exposed upon compaction. As with K19, its binding partner E536 shows a favorable free energy change due to interaction.

Center : per-residue decomposition of the overall free energy change of gp17 compaction as calculated : $\Delta G_{res} = \Delta G_{compact} - \Delta G_{extended}$. For the most part, the free energy change of each residue is zero because, in large part, the structure of gp17 does not change with compaction with the exception of residues at and in close proximity to the interface. Residues that have either a favorable or unfavorable free energy change tend to be located at the interface between N- and C-terminal subdomains. This can include residues such as charged amino acids that were previously proposed to be involved with gp17 compaction, such as R494, which forms a salt bridge with E309. In addition, this decomposition allowed for identification of residues that are likely involved in compaction that were not previously identified. For example, residues V531 and Y295 are both hydrophobic and are among the most favorable upon binding.


E536
K19
E537
K23
W298
Y295
P310
E303
E309
K305
R494
R494
K493
D505
K504
K19
K23
E536
E537
W533
V531
K305
E309
K19
K23
Y295
W298
E303
P310
K493
S501
K504
D505
R494
E537
E536
V531
W533
1
-1
-3
-5
-7
10
50
100
150
200
250
300
350
400
450
500
550
N-II
N-I
C

Free energy decomposition allows for the total calculated free energy to be broken down into per-residue contributions. As our mechanism relies on a conformation change that we probed using molecular dynamics simulations, much of the total free energy of the system arises from amino acids at or close to the interface between the N- and C-terminal subdomains. We show the structural context of residues contributing

Table 3.2. Free energy contributors with ΔG < -0.5. We decomposed the wild type binding free energy to reveal the contribution made by each residue. Of the 553 residues of gp17, eleven residues in wild type exhibit free energy changes in excess of -0.5kcal/mol upon compaction. Of these, 5 are charged amino acids, 1 is polar, and 5 hydrophobic. These five charged residues make up 29% of the total free energy, and the six polar/ hydrophobic residues account for 46% of the total energy change.

charged residues		**hydrophobic and polar residues**	
residue	contribution to total (kcal/mol)	residue	contribution to total (kcal/mol)
K19	-0.73 ± 0.67	Y295	-4.23 ± 2.14
E303	-2.67 ± 0.97	W298	-2.16 ± 0.60
E309	-2.43 ± 0.34	P310	-2.38 ± 0.79
R494	-4.52 ± 0.50	T312	-1.45 ± 0.46
E536	-0.65 ± 0.35	V531	-3.74 ± 1.02
		W533	-3.59 ± 1.02

largely to the free energy change of compaction in Figure 3.5. We find that 85% of the total change in free energy arises from the 46 amino acids at the interface (again

defined as those within 5Å) when all 46 residues are included (this includes both favorable and unfavorable contributors).

Table 3.2 lists all amino acids that exhibit a large (defined as < -0.5 kcal/mol) favorable free energy change upon gp17 compaction. The list includes the five charged amino acids, K19, E303, E309, R494, and E536 that were originally implicated in motor activity by crystallographic studies. In addition, one polar and five hydrophobic residues have very favorable free energy changes : Y295, W298, P310, T312, V531, and W533. These eleven amino acids alone account for 75% of the total free energy of conformational change - equal to the net change of the entire interface. Interestingly, we find that seven of the charged amino acids originally proposed to be involved in generation of the ratcheting mechanism do not exhibit favorable free energy change upon compaction, and are therefore unlikely to be involved in force generation. We have mapped the free energy contributions of individual residues onto the surface of the protein in Figure 3.6. We discuss the structural details behind these observations in sections 3.4.3, 3.4.4, and 3.4.5. and 3.4.5.

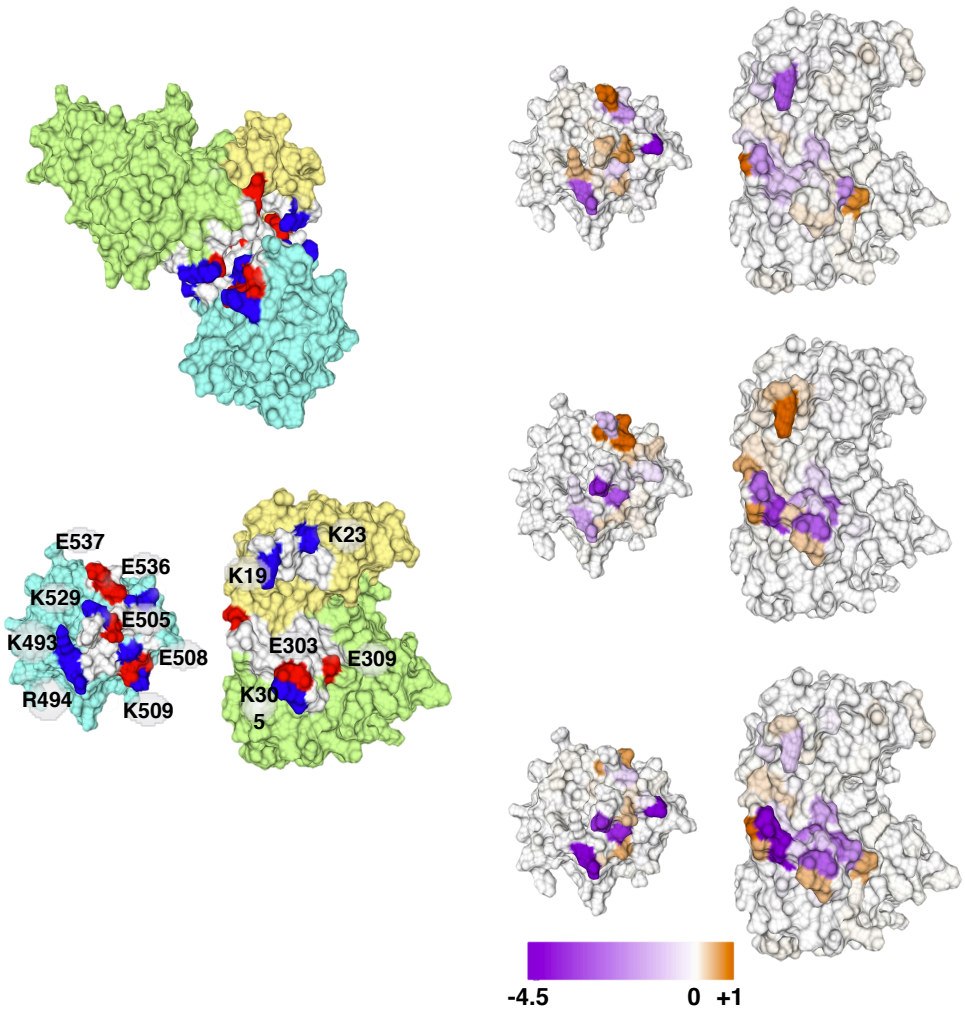


Fig 3.6. gp17 structural map and residue level contributions to overall free energy of binding. A) A model of monomeric gp17 depicting the three subdomains N-I, N-II, and C in green, yellow, and cyan, respectively. Residues at the interface are colored by charge. Positively charged, negatively charged, and hydrophobic residues are shown in blue, red, and white, respectively. Charged residues at the interface are labeled. (B) Open book view of the gp17 interdomain interface region showing charge complementarity between the domains. Charged residues in this contact region defined as atoms less than 5Å apart between the N and C terminal subdomains. (C) gp17 interface colored by per-residue electrostatic contribution to free energy change. Color map indicates contribution in kcal/mol. (D) gp17 interface colored by per-residue hydrophobic contribution to the total free energy change. (E) gp17 interface colored by per-residue overall contribution to the total free energy change.

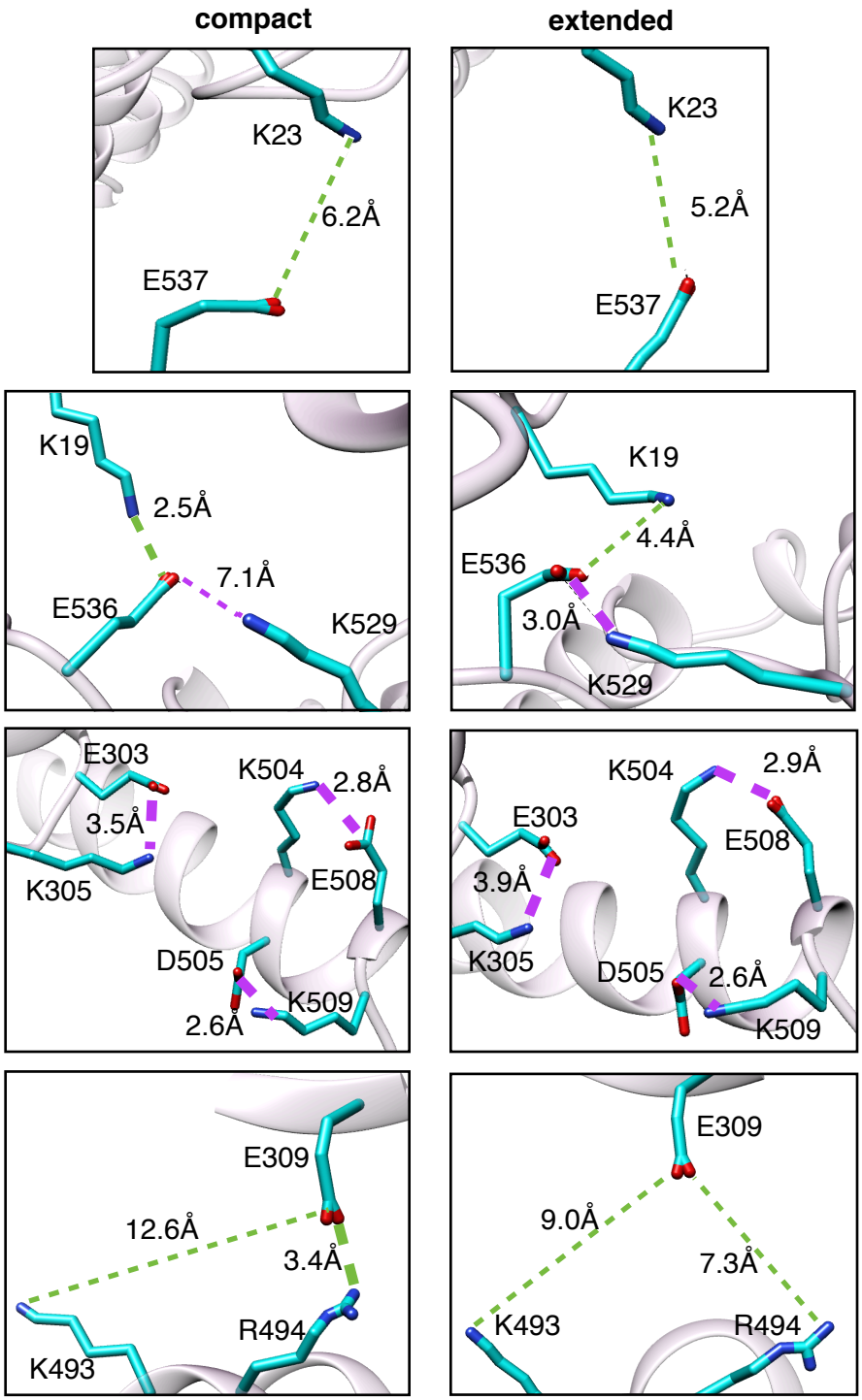
3.4.3 Fewer charged amino acids contribute to interface binding than predicted by X-ray crystallography.

Free energy decomposition suggested to us that ion pairing across the interface between N- and C-terminal subdomains may be less straightforward than originally implied by the crystal structure. The relaxation of gp17 from the crystal structure by molecular dynamics allows for the overall structure to assume a favorable conformation, from which we calculated a free energy change of compaction (ΔG_{EC}) as well as the contribution by each residue. Thus, a salt bridge formed in the extended as well as the compact structure, while very stable and contributing to the free energy of the protein folding, will contribute little to the free energy of compaction. To accompany the free energy calculated per residue, we measured the change in distance between charged amino acid functional groups between extended and compact states throughout 4 ns of our molecular dynamics simulations.

The differences between the free energy contributions of charged residues at the interface are striking and can be seen in Table 3.3. Of the twelve charged amino acids originally proposed to form ion pairs that lead to gp17 compaction, 5 contribute a significant amount to the free energy. Residues K23, K305, K493, K504, E505, E528, and E537 lie within the interface, but collectively contribute very little to the forces leading to gp17 compaction. In most cases, this is due to a combination of the unfavorable desolvation that comes with reducing the volume of water in the interface during compaction (discussed in section 3.4.4) and lack of change in electrostatic interactions (discussed in this section). Five residues contributing the majority of the free energy change derived from charged amino acids (K19, E303, E309, R494, and

Fig 3.7. Wild type ion pairing interactions in compact vs extended conformations.
Using a Chimera structural averaging program included with the molecular dynamics viewer, we calculated the average structure of compact and extended gp17 over the course of several nanoseconds. We then used the Chimera distance tool to measure distance between atoms of interest, in this case charged atoms belonging to charged amino acids proposed to be involved in ion pairing by Sun, et al. We compared the distance between atoms in these proposed ion pairs, as well as with nearby potential binding partners not originally proposed. Left side, compact gp17; right, extended gp17. From top, each ion pair is viewed in context of its surroundings with distances measured. Pink dashed line, intradomain interaction. Green dashed line, interdomain interaction. Thick line, salt bridge interaction (discussed in figure 3.1). Thin line, weaker, long range ion pair interaction (also discussed in figure 3.1).

Top : the interaction between K23 and E537 extends beyond salt bridging capabilities. A weak interaction, it contributes little to the overall free energy change.
Second : the interaction between K19 and E536 is complicated by the proximity of alternate binding partner K529. Compaction of gp17 leads to a lengthening of the interaction between K529 and E536, breaking the 3Å salt bridge between the charged atoms and lengthening it to 7.1Å, while giving rise to a stronger 2.5Å salt bridge between E536 and K19, shortened from 4.4Å in extended gp17.
Third : A complex network of intradomain interactions form between several amino acids in the center of the interface. Although K305 was originally proposed to interact across the interface with D505, and E303 with K504, we find strong salt bridging interactions are formed in both extended and compact conformations within each subunit, likely stabilizing a loop (E303 and K305) or the end of a helix (K504 and D505). Two residues not originally identified as involved in interface formation, E508 and K509, form salt bridges with K504 and D505, respectively, on an adjacent turn of the helix they are all located on.
Bottom : E309 has two possible binding partners which appear equally likely to interact with E309 based on crystal alignment. Upon simulating the equilibrated protein, however, we find that one interaction, E309-R494 is favored over the other. This interaction forms entirely in compact gp17, unlike the interaction between K19 and E536, where a weaker interaction becomes a salt bridge upon compaction. Although E309-R494 have the capability to form interactions between four charged atoms, the geometry of E309 does not allow for this. The reason for this is discussed in detail in section 3.4.5.


compact
extended
K23
6.2Å
E537
K23
5.2Å
E537
K19
2.5Å
7.1Å
E536
K529
K19
4.4Å
E536
3.0Å
K529
E303
3.5Å
K305
K504
2.8Å
E508
D505
K509
2.6Å
K504
2.9Å
E303
E508
3.9Å
K305
D505
2.6Å
K509
E309
12.6Å
3.4Å
R494
K493
E309
9.0Å
7.3Å
K493
R494

E536) exhibit a combination of hydrophobic and electrostatic forces that will also be discussed.

Figure 3.7 shows the relationship between interaction domain ion pairs and the distance between the charged atoms. We find that, though amino acids K19 and K23 appear to be capable of forming interactions with either E536 or E537 (Figure 3.2), a bond is formed exclusively between K19 and E536, and between K23 and E537 (4.22 ± 1.7Å and 5.9 ± 2.5Å, respectively). Both of these interactions become substantially closer (by 2Å and 2.4Å, respectively) when gp17 is in the compact form. There is extremely little interaction between K19 and E537, or between K23 and E536 (with interatomic distances of 8.92 ± 2.3Å and 12.03 ± 2.38Å, respectively) though they are spatially close in the X-ray crystallographic depiction of gp17. We also find that in extended gp17, E536 forms a strong salt bridge with K529 at a distance of 2.9 ± 0.4Å that largely disappears in compact gp17 (4.7 ± 1.7Å), allowing for E536 to interact with K19. This salt bridge helps to make the transition to the extended state favorable.

In the center of the interaction domain, a cluster of closely placed charged amino acids forms an ambiguous region in the crystal structure. The crystal structure of gp17 suggests that amino acids E303 and K305, along with K504 and D505, should form two strong interactions in the center of the interaction domain, shown in Figure 3.2. We find that neither of these interactions is formed at all. E303 and K305 are found to interact very closely (3.5 ± 0.6Å) to form a stable salt bridge that shows very little difference between the extended and compact configurations. K504 forms a salt bridge with E508, on the final turn of the helix containing 4 of the 12 originally proposed charged amino acids. Salt bridging interactions are known to be involved in helix capping, which is what we find here in the two terminal helix salt bridges[8]. The distance

Table 3.3. Contributions of all interface charged residues to the overall free energy change. Free energy decomposition allows for us to distinguish between the charged residues within the interface region and to assign likely importance based on contribution. Out of 12 charged amino acids at the interface, we find that 5 contribute significantly (< -0.5kcal/mol) to the free energy of compaction. These are in bold typeface. Residue names are colored in terms of their charge : blue is positive, red is negative. It is also important to note that the contributions from these residues comes from a mixture of hydrophobic and electrostatic sources, a finding that is discussed in this chapter.

residue	hydrophobic	electrostatic	contribution to total
K19	**2.17 ± 0.41**	**-2.91 ± 4.47**	**-0.73 ± 0.67**
K23	-0.12 ± 0.12	-0.11 ± 3.88	-0.23 ± 0.36
E303	**-2.70 ± 0.93**	**0.02 ± 1.68**	**-2.67 ± 0.97**
K305	0.29 ± 0.11	0.15 ± 1.54	0.45 ± 0.16
E309	**0.11 ± 0.18**	**-2.54 ± 3.25**	**-2.43 ± 0.34**
K493	-0.40 ± 0.06	0.41 ± 2.22	0.002 ± 0.14
R494	**-1.29 ± 0.25**	**-3.23 ± 2.71**	**-4.52 ± 0.50**
K504	0.05 ± 0.05	-0.08 ± 1.46	-0.04 ± 0.11
E505	0.06 ± 0.02	-0.12 ± 0.86	-0.06 ± 0.03
E528	-0.27 ± 0.09	0.62 ± 1.88	0.35 ± 0.15
E536	**1.05 ± 0.21**	**-1.70 ± 4.05**	**-0.65 ± 0.35**
E537	-1.01 ± 0.16	0.97 ± 2.99	-0.05 ± 0.19

between K504 and E508 is stable between both conformations at 2.8 ± 0.9Å. Similarly, D505 forms a salt bridge with K509 in both conformations at a distance of 2.9 ± 0.3Å.

The fifth proposed interaction domain contact was expected to form between E309-K493 or R494. By measuring the distances between charged atoms of each of these proposed pairs, we show that a stronger interaction, E309-R494, forms at a distance of 3.5 ± 1.0 Å. This interaction distance, compared to 8.3 ± 1.9 Å seen in extended gp17, shortens by a massive 4.8 Å when gp17 compacts, explaining the very

large associated free energy changes. Based on the free energy changes as this salt bridge forms, we believe that the interaction between E309 and R494 plays a dominant role in the compaction of gp17; this is confirmed by a complete lack in packaging activity by the construct gp17-E309K.

3.4.4 Solvation plays an important role in the favorability of compaction.

Electrostatic amino acid interactions with water are almost always favorable. Charged and polar atoms, including backbone atoms, make hydrogen bonds with water. The geometry of charged amino acid functional groups allow for water molecules to adopt stable conformations as they create ordered cages surrounding these atoms; the number of water molecules complexed with the residue is directly related to the number of electrostatic atoms. To form the hydrogen bond, each water molecule aligns such that the correct hydrogen bond acceptor is close to the electrostatic atom. The water molecules surrounding each charged amino acid at the interface are shown in Figure 3.8. If a second charged or polar amino acid of opposite charge is nearby, it may also hydrogen bond to these same water molecule. This organization allows for spatially distant residues to effectively create electrostatic interactions via water molecules. The simplest organization, whereby one water molecule is arranged to create an interaction between two residues is called a water bridge[18]. Each water bridge is subject to the same energetic constraints that exist in forming interactions between amino acids, namely loss of two water-water hydrogen bonds in order to allow for bridging, and the entropic cost of fixing a water molecule in place[19].

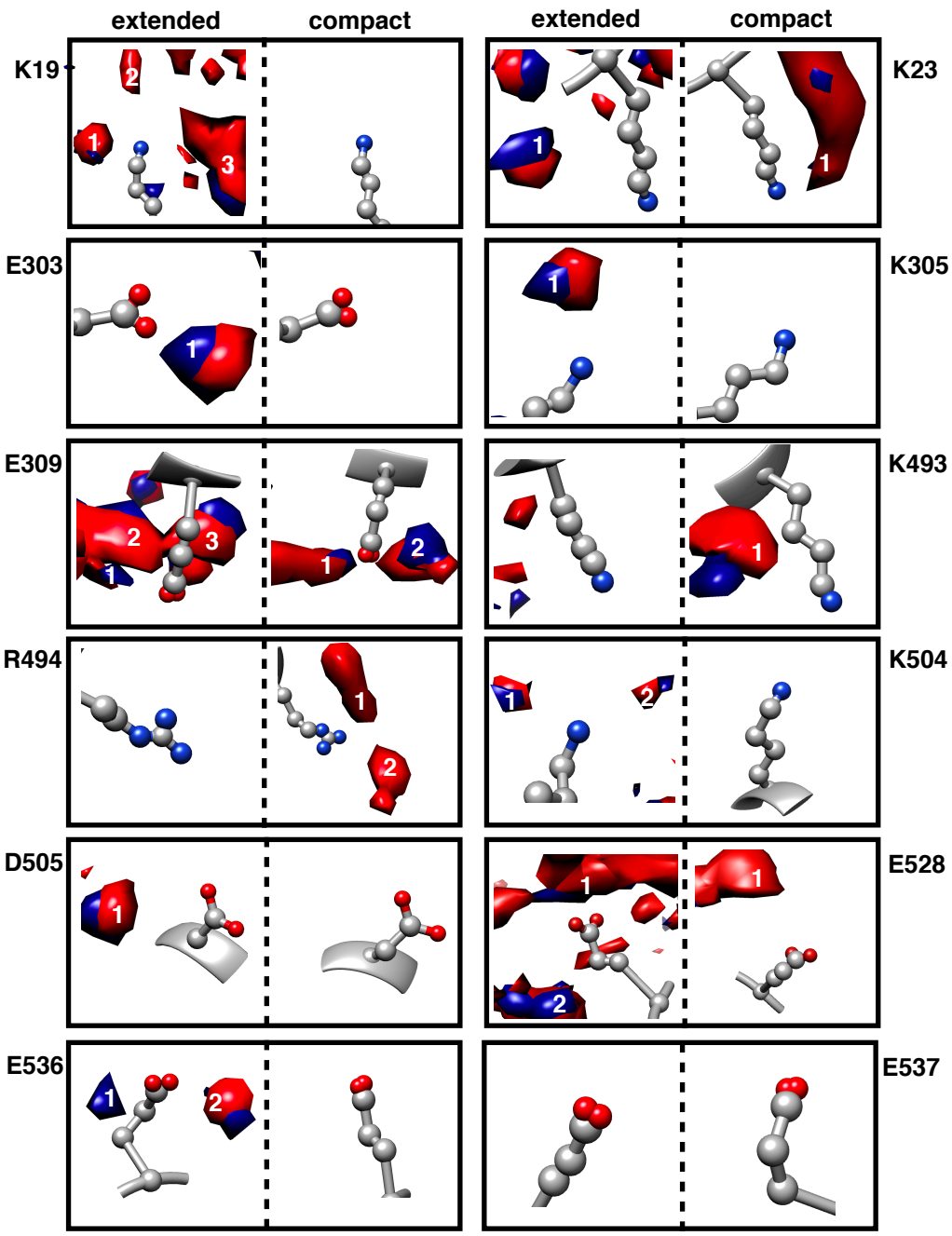


Fig 3.8. The change in charged residue hydration due to compaction.
We calculated the average water density surrounding gp17 in the compact and extended state as discussed in section 3.3.3. Changes in the number of water molecules surrounding the amino acid functional groups are strongly related to the hydrophobic energy term at each position. Each of 11 amino acids is shown in both compact and extended gp17 in relation to surrounding water molecules. We have shown the number of nearby water molecules by numbering those within hydrogen bonding distance in white text. Water molecule densities are shown in blue and red, with red depicting oxygen density, and blue depicting hydrogen density. Likewise, amino acids are labeled in red for oxygen, and blue for nitrogen.

We identified three key ion-pair interactions at the fringes of the interface region (K19-E536, K23-E537, and E309-R494) where water plays a role in the stability. For K19-E536 in the extended conformation, the distance between oxygen and nitrogen is 4.4 Å - too distant to form a very stable ion pair. However, the presence of bridging water molecules between the two allows for an interaction to occur. In the compact structure, this distance is substantially lessened, and the water molecule is expelled. The change in occupancy of water molecules at this position is shown in Figure 3.8; both amino acids become fully desolvated in order to interact strongly. This accounts for the modest free energy decrease upon compaction at this position - a stable water-bridged electrostatic interaction and shared is traded for a direct salt bridge formed in the compact conformation. To gain a total free energy change of -1.37 ± 0.75 kcal/mol, the interaction between K19 and E536 pays a stiff penalty of 3.22 ± 0.46 kcal/mol due to this desolvation as seen in Table 3.3.

The interaction between K23 and E537 is very different in character from K19-E536. In both extended (4.9Å) and compact gp17 (6.2Å), the distance between charged atoms is larger than that of K19-E536. Although water is found between K19 and E536, none is located between the charged atoms of K23 and E537, seen in Figure 3.8. Mutation of E537 to lysine, as studied in Chapter 2, leads to repulsion between K23 and K537. The resulting rearrangement of K537 leads to the interaction between K19 and E536 becoming elongated. In this manner, packaging is affected indirectly by mutation at 537. Thus the interaction between K23-E537 itself likely has little contribution to viral packaging.

E309 and R494 form the strongest salt bridging interaction across the interface. We know from experimental observations that mutation of E309 to lysine is devastating

to the activity of the motor. E309K demonstrated no measurable packaging of DNA in experimental studies. As was mentioned in section 3.3.3, compaction brings the charged atoms in E309-R494 together by nearly 5 Å. In addition to the favorability imparted by the formation of a salt bridge, two water molecules wrap around the back side of the interaction. Because of their structures, arginine and glutamic acid are able to form an interaction between four atoms rather than three, but do not. Instead, additional hydrogen bonding interactions are formed between either of the residues and water. The favorable hydrophobic term in the interaction between E309 and R494 seen in Table 3.3 arises because, in the extended structure, R494 has no stable water molecules surrounding it. This change in hydrogen bonding due to proximity to the waters surrounding E309 can be seen in Figure 3.8. E309 remains surrounded with water molecules, and its change in hydrophobic free energy upon binding is close to zero. In contrast, R494 becomes much more well hydrated when it interacts with E309, and has a hydrophobic free energy change of -1.29 ± 0.25 kcal/mol. This is very different than the water-mediated interaction between K19 and E536 in that the salt bridge formed is independent of hydration state - the water simply stabilizes the interaction by forming hydrogen bonds with the charged atom not involved in the two salt bridges. In this way, the interaction between E309 and R494 does not suffer as greatly from the desolvation penalty generally incurred when a salt bridge is formed.

In a region of the interface that is particularly buried and solvent-exclusive upon compaction, E303 forms an intradomain salt bridge with K305. This interaction exists roughly equally in both compact (3.3Å) and extended (3.7Å). Instead, solvation related term dominates the free energy change of this particular residue. The reason behind this highlights the difference between the interactions of arginine with glutamic acid

and the interaction of lysine with glutamic acid. Because arginine has two terminal nitrogen atoms, it is more flexible in terms of interactions than lysine with its one terminal nitrogen atom. Therefore, the water molecule localized at E303 in the extended structure does not stabilize the salt bridge with K305 as it does the interaction E309-R494. Our results indicate that the expulsion of this water leads to a favorable hydrophobic free energy change upon compaction.

Additionally, as the sum of electrostatic free energy is -18.37 ± 15.64 kcal/mol, and the total free energy change is -38.21 ± 4.42 kcal/mol (as shown in figure 3.1), we know that something other than interactions between charged amino acids (i.e. hydrophobic, polar, or solvation) must be at work in the compaction of gp17.

3.5 Discussion

By approximating the gp17 compaction as a protein-protein interaction (as if the two subdomains were split), we can utilize standard tools more commonly applied when studying binding. This allows us to begin understanding the interactions that occur within a single gp17 protein with several different methods. First, we investigated the per-residue contribution to the overall free energy of binding as presented in Chapter 2. Breaking down the free energy of conformational change per-residue allowed us to understand the importance of each amino acid to the overall ratcheting mechanism (Table 3.2 and Figure 3.5). Although other research has suggested that it is possible to use the Debeye-Waller factor[20] to glean some information about residue mobility, we find that our molecular dynamics simulations give a much more thorough impression of the motion of interface residues without the possibility of including crystallization artifacts.

By then applying free energy decomposition, we identified a total of eleven residues at the interface between N- and C-terminal subdomains that contribute most to the favorable free energy change of the compaction associated with DNA translocation via ratcheting. This allows us to better understand the interface in terms of the electrostatic (charged residues), hydrophobic (nonpolar and polar uncharged residues), and solvent interactions (all residues) of each residue belonging to the interface region.

We also took advantage of techniques used for analysis of non-sequential amino acid interactions. In a large study of ion pairing in proteins, it was shown that ~50% of ion pair interactions form between amino acids separated by greater than 10 amino acids in sequence[21] - implying that long-range interactions play an important role in the understanding electrostatic interactions. Most studies of salt bridge position are based on information from the study of many crystal structures, the structure of which may differ from the free protein; this is certainly the case with gp17, where the extended and compact states have different structures. In order to generate a complete picture of electrostatic interactions at the interface, we measured distances between proposed interdomain ion pairs as well as between amino acids on the same side of the interface that may act to stabilize structural elements. For example, it was proposed that E536 forms a salt bridge with K19. However, by extending interaction candidates to include those on the same interface, we find K529 is also a possible binding partner. This analysis shows that a very strong interaction forms between E309 and R494, contributing 18% of the total free energy change of gp17 compaction. The interaction between E309 and R494 shows distances between charged atoms well within the 4Å cutoff for salt bridging, and has excellent geometry for creating a strong,

stable salt bridge. We showed that mutation of E309 to lysine completely abrogates packaging (Chapter 2), and it is highly probable, based on free energy information, that mutation of R494, K19, and E536 would have similar affect on activity. Residues E536 and K19 form a second salt bridge with a distance of 2.5 Å, though it contributes significantly less energy (~4% of the total energy) of the free energy change because the extended structure shows these two residues separated by only 4.4 Å. In addition, E536 forms a salt bridge with K529 in the extended structure that is severed upon compaction.

Interestingly, we find that the center of the interface, containing residues E303, K305, K504, and E505, forms no interactions stabilizing the interface. In contrast to the two salt bridges outlined above which form between residues on opposite sides of the interface, these residues fall within the 50% of interactions that occur locally. These four residues are located in close proximity to both E309 and R494, and may aid in the positioning of these two amino acids that allow for the crucial salt bridge (between E309 and R494) to form across the interface. Although their stability in both extended and compact gp17 means they contribute a negligible amount of the free energy change of compaction, the mutation of E303K was shown to have a large effect on packaging activity, providing further evidence for this role. Finally, the interaction between K23 and E537 forms a long range ion pair in both extended and compact conformations. However, mutation of E537K shows a marked reduction in packaging; this is likely due to the disruptive effect to the interaction between K19 and E536.

To create the close interactions necessary to form the translocation step by closure of the interface between N- and C-terminal subdomains, it is necessary that resident water molecules be largely expelled from the region. Interestingly, the number

of water molecules found in an interface increases linearly with increasing buried surface area, even though these interfaces generally exclude the majority of water molecules[22]. Water forms ordered hydrogen bonds surrounding solvent-exposed charged and polar residues, interactions that are both electrostatically and entropically favorable. As mentioned in section 3.2, the formation of ion pairing between solvent-exposed, charged amino acids first necessitates the removal of complexed water molecules, an unfavorable step in forming a potentially very favorable interaction. Thus, ion pairing is only favorable if the energy gained by Coulombic forces between charged atoms is greater than that of desolvation. Due to the location of the salt bridge between E309 and R494 on the mostly solvent-exposed back side of gp17, much of the water remains surrounding both amino acids when a salt bridge is formed. The combination of a large distance change upon compaction leading to a very close salt bridge and the comparatively modest desolvation penalty make the ion pair formed between E309 and R494 the most energetically favorable by far. It is known that forming an interaction between glutamic acid and lysine is less favorable than with arginine, and this difference is seen at K19-E536. Salt bridge formation leads to the loss of most hydrogen-bonded water molecules, and as a result contributes substantially less to the overall free energy change.

Most interestingly, we found that electrostatic interactions did not completely explain the free energy change of compaction. Although Sun, et al. proposed that the action of gp17 would be mediated solely by electrostatic interactions, we have found that is not the case. Initial free energy calculations indicated that the source of compaction energy came from equal contributions of electrostatic and hydrophobic free energy. Free energy decomposition further indicated that almost the entirety of the free

energy is derived from the contributions of only 11 amino acids, 5 of which are charged. This stands in stark contrast to the results of traditional analysis of the crystal structure of the compact state, and suggests that computational studies using our methods may reveal additional unusual behavior when applied to other active biomolecules.

In Chapter 2, we showed the results of investigation into the importance of charged amino acids in the interface between N- and C-terminal subdomains. Mutation to the opposite charge revealed that each position affects activity to a differing degree : E309K completely abrogated packaging activity, while E537K had a much more modest affect in activity. These experiments showed a clear relationship between the free energy change of compaction and activity. The analysis of the details of interface formation not only helps to elucidate the details of the interaction, but also suggests mutational studies that can be performed for cross-verification. Finally, it is clear from the magnitude of the electrostatic contribution, as determined via our computational experiments, that in order to fully understand the interface between subdomains of gp17, the importance of hydrophobic interactions can not be ignored. These contributions, as well as other subtle solvation effects, will be discussed in Part II of this analysis.

3.6 Acknowledgments

The author would like to acknowledge coauthors Gaurav Arya and Douglas E. Smith for their contributions to the work presented in this chapter.

References

1. Manosas, M., Xi, X. G., Bensimon, D. & Croquette, V. Active and passive mechanisms of helicases. *Nucleic Acids Research* **38,** 5518–5526 (2010).

2. Cross, R. A. Myosin's mechanical ratchet. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 8911–8912 (2006).

3. Mehta, A. D. Amit D. Mehta, Ronald S. Rock, Matthias Rief, James A. Spudich, Mark S. Mooseker and Richard E. Cheney. Myosin-V is a processive actin-based motor. *Nature* **400,** 590–593 (1999).

4. Rehling, P., Brandner, K. & Pfanner, N. Mitochondrial import and the twin-pore translocase. *Nat Rev Mol Cell Biol* **5,** 519–530 (2004).

5. Baker, M. J., Frazier, A. E., Gulbis, J. M. & Ryan, M. T. Mitochondrial protein-import machinery: correlating structure with function. *Trends in Cell Biology* **17,** 456–464 (2007).

6. Brown, K. D., Wood, K. W. & Cleveland, D. W. The kinesin-like protein CENP-E is kinetochore-associated throughout poleward chromosome segregation during anaphase-A. *J Cell Sci* **109,** 961-969 (1996)

7. Sun, S., Kondabagil, K. R., Draper, B., Alam, T. & Bowman, V. The Structure of the Phage T4 DNA Packaging Motor Suggests a Mechanism Dependent on Electrostatic Forces *Cell* **135**, 1251-1262 (2008).

8. Kumar, S. & Nussinov, R. Close-Range Electrostatic Interactions in Proteins. *ChemBioChem* **3,** 604-617 (2002).

9. Kumar, S. & Nussinov, R. Relationship between ion pair geometries and electrostatic strengths in proteins. *Biophysical Journal* **83,** 1595–1612 (2002).

10. Levy, Y. & Onuchic, J. N. Water mediation in protein folding and molecular recognition. *Annu. Rev. Biophys. Biomol. Struct.* **35,** 389-415 (2006).

11. White, A. D. Andrew J. Keefe, Jean-Rene Ella-Menye, Ann K. Nowinski, Qing Shao, Jim Pfaendtner, and Shaoyi Jiang. Free Energy of Solvated Salt Bridges: A Simulation and Experimental Study. *J. Phys. Chem. B* **117**, 7254-7259 (2013).

12. Wolfenden, R., Andersson, L., Cullis, P. M. & Southgate, C. Affinities of amino acid side chains for solvent water. *Biochemistry* **20,** 849–855 (1981).

13. Massova, I. & Kollman, P. A. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding - Springer. *Perspectives in Drug Discovery and Design* **18,** 113–135 (2000).

14. Srinivasan, J., Trevathan, M. W., Beroza, P. & Case, D. A. Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt

effects. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)* **101,** 426–434 (1999).

15. Hinsen, K. Analysis of domain motions by approximate normal mode calculations. *Proteins Structure Function and Genetics* **33,** 417-429 (1998).

16. Roe, D. R. & Cheatham, T. E., III. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **9,** 3084-3095 (2013).

17. Pettersen, E. F. Thomas Goddard, Conrad Huang, Gregory Couch, Daniel Greenblatt, Elaine Meng, Thomas Ferrin. UCSF Chimera : A visualization system for exploratory research and analysis. *J Comput Chem* **25,** 1605–1612 (2004).

18. Petukhov, M., Cregut, D., Soares, C. M. & Serrano, L. Local water bridges and protein conformational stability. *Protein Sci.* **8,** 1982–1989 (1999).

19. Meenan, N. A. G. Amit Sharmaa, Sarel J. Fleishmanb, Colin J. MacDonaldc, Bertrand Morelc, Ruth Boetzelc, Geoffrey R. Moorec, David Bakerb, and Colin Kleanthous. The structural and energetic basis for high selectivity in a high-affinity protein-protein interaction. *Proceedings of the National Academy of Sciences* **107,** 10080–10085 (2010).

20. Kundu, S., Melton, J. S., Sorensen, D. C. & Phillips, G. N. Dynamics of proteins in crystals: comparison of experiment with simple models. *Biophysical Journal* **83,** 723–732 (2002).

21. Kumar, S. & Nussinov, R. Fluctuations in ion pairs and their stabilities in proteins. *Proteins* **43,** 433–454 (2001).

22. Rodier, F., Bahadur, R. P., Chakrabarti, P. & Janin, J. Hydration of protein-protein interfaces. *Proteins* **60,** 36–45 (2005).

4 Interactions in the interface between N- and C- terminal subdomains in the bacteriophage T4 packaging motor. II. Non-specific interactions.

Amy Migliori[1], Gaurav Arya[3], and Douglas Smith[2]

4.1 Summary

Prior work from our laboratory has demonstrated that mutation of certain charged residues in the T4 bacteriophage translocation motor protein, gp17, affects the ability of the motor to package DNA. Further, computational analysis shows that the degree of motor impairment due to mutation correlates with the free energy change during compaction associated with the residue in wild type gp17. This suggests that electrostatic interactions are an important part of the mechanism by which gp17 packages DNA. However, measurements of the overall free energy change of compaction and of specific residues contributing large portions of the free energy during compaction reveal that electrostatic forces only explain half of the total energy change. Of the eleven residues with free energies ≤ -0.5 kcal/mol, only one is polar; five are hydrophobic. We explain the large hydrophobic contributions by examining the exclusion of water from the interface. Desolvation and subsequent formation of water-excluding interactions is typically mediated by hydrophobic residues. Using a combination of molecular dynamics and analysis of the N- and C- subdomain interface in gp17 to calculate the free energy change of compaction, we find that hydrophobic residues Y295, W298, and W533 aid in the exclusion of water and form extensive

---

[1]Department of Biochemistry, University of California, San Diego. La Jolla, CA, 92037.

[2]Department of Physics, University of California, San Diego. La Jolla, CA, 92037.

[3]Department of Nanoengineering, University of California, San Diego. La Jolla, CA, 92037.

contacts with other residues in the compact form. In addition, residue P310 likely stabilizes a loop (containing E309) that connects a helix (containing Y295 and W298) and a beta sheet. Residue V531 is also indicated to aid in the overall compaction of the interface, and exhibits an increased solvent exclusion upon binding that effects nearby W533 as well. Finally, T312, the only energetically important polar residue, forms twice as many nonspecific interactions across the interface in the compact configuration, although fewer overall than in the extended conformation, allowing for its polar hydroxide to be revealed. This allows T312 to hydrogen bond with a water molecule through its hydroxide, which is occluded in the extended conformation. By combining the energetic analysis with the measurement of local contacts and surrounding solvation, our results strongly suggest that Y295, W298, and W533 represent hot spot residues at the interface. These three residues, due to their large footprint and positioning within the interface, substantially aid in desolvation, allowing for favorable interactions to occur across the interface. The presence of these hot spots help to explain the large hydrophobic contributions to the free energy change of compaction.

Keywords : MM-GBSA, protein-protein interactions, ion pair, hydrophobic, molecular motor

4.2 Introduction

The ubiquity of molecular motors in cellular and viral functions makes understanding the mechanism by which they perform their unique actions especially crucial. Many of these motors translocate payloads to complete molecular rearrangement or assembly of a macromolecular complex. In the extremely crowded

and highly partitioned cellular environment, many functions require a specific subset of components be present at specific times. Diffusion is too slow for such transport by orders of magnitude; without translocation motors, molecular machinery as we know it would cease to exist. Each molecular motor is uniquely suited for the task it undertakes; transporting cargo through the cellular milieu requires a distinctively different set of abilities than does separating the two strands of double-stranded DNA[1] or towing via kinesin[2]. One of the most powerful motors known is found in a subset of viruses, the bacteriophages that infect bacterial cells, which replicate via injection of their genome through the hard, protective cell wall of a bacterium. Releasing 170,000 bp of DNA into a bacterium such that the entire genome reliably enters the cell, ensuring all necessary genes are present for replication, is no simple task. Injection of the genome requires extremely large internal pressure within the viral capsid; the DNA translocation motor gp17 motor has evolved to fill this role. The gp17 motor self-assembles onto a portal protein at a unique 5-fold vertex of the icosahedral capsid and couples the energy released by ATP hydrolysis to DNA translocation, moving 2bp with each hydrolysis cycle[3]. This 2bp translocation has been observed using high-resolution studies measuring the step size for the similar, albeit much slower, genome translocation motor of Φ29[4]. In contrast to the prior suggestion that the motor rotates around the portal protein opening to the capsid, the direct interaction of gp17 using a helix-loop-helix motif close to the ATP binding pocket[5], coupled with structural information showing a gp17 compaction that leads to translocation, suggests a ratchet-like mechanism.

Studies of the crystal structure propose that five ion pairs in the interface region aid in causing the compaction of gp17, and are responsible for generating the high

forces seen experimentally. Our computational experiments showed that only 48% of the overall free energy of compaction results from electrostatic forces, and that only two ion pairs are able to form salt bridges in the gp17 interface domain. One additional weak interaction between K23 and E537 exists both in the extended as well as the compact structure and may be involved in maintaining alignment between the subdomains in the extended configuration, but does not directly assist in compaction force generation. Our analysis suggests that although ion pairing plays a role in gp17 compaction, hydrophobic and solvation/desolvation interactions are equally important to the generation of a metastable interface that allows for packaging to proceed. The relationship demonstrating the connection between the prevalence of hydrophobic residues, particularly aromatic residues, and interface formation has been well-discussed in the literature[6-8].

Once a protein is folded, it generally interacts with other molecules by electrostatic interaction (discussed in Part I), or by making noncovalent bonds. The noncovalent bonds are relatively weak compared to covalent bonding. However, the aggregate effect of many noncovalent interactions can lead to a stable interaction. Noncovalent interactions include van der Waals interactions between closely related hydrophobic residues, and π-stacking interactions between planar portions of large aromatic residues. A salient feature of each of these interactions is that a change in solvation must occur to allow for them to form, as water molecules occlude such interactions. The value of interaction energy between two amino acids is typically measured using calorimetry to determine the free energy change that occurs when two fully solvent-exposed residues form a contact. In the case of the interface between the two subdomains of gp17, the residues are never fully solvent exposed once the protein

is folded. Calorimetric values thus overestimate the free energies of interaction across the gp17 interface. To more accurately measure the free energy change, we used computational simulations to measure the importance of each interaction.

Water molecules make hydrogen bonds with electronegative or electropositive atoms. Polar and charged amino acids form networks of water molecules arranged around nitrogen or oxygen atoms that are very favorable. The aggregate effect of this interaction with water is that most proteins are at least partially water soluble - a property important for existence in the watery environments many proteins reside in. For this reason, the native conformation of a protein tends to favor burial of hydrophobic residues to the center of the protein, where solvation is minimal. Interfaces between proteins typically occur where the protein surface contains a larger percentage of hydrophobic residues. This occurs because the solvent-excluded interface favorably occludes hydrophobic residues from unfavorable interactions with water. In gp17, each amino acid is exposed to solvent in the extended conformation differently, due to the non-homogenous nature of the surface. The noncovalent interactions that aid in compaction are, in many cases, partially formed in the extended state as well. To measure the solvent effect at the position of energetically important residues, we visualize nearby water molecules and ascertain the favorability of their positioning. For example, we found E309-R494 retains nearly all of its hydrogen bonding interactions with water while forming a very close salt bridge. This means that the free energy change of creating the interaction between these two amino acids is very favorable: electrostatic forces are very high, and the desolvation/hydrophobic penalty is minimized due to the minimal desolvation.

As a way of taking solvation into account, protein-protein interactions are typically characterized by the amount of surface area that ends up buried upon binding, the shape complementarity of the surfaces, and the types of specific and non-specific amino acid contacts formed; each of these directly influences the stability of the interface. When the interface forms between non-identical proteins (a heterodimeric interaction), hydrophobic regions on each protein are brought together during binding; this reduces the amount of solvent-exposed hydrophobic area and is thus generally favorable. In the case of homodimeric interfaces, most are highly stable and occur very quickly after the protein is made; burying a hydrophobic surface is less of a concern because the surface is rarely, if ever, exposed due to the immediacy and stability of binding. The interface formed between N- and C-terminal subdomains in gp17 is never fully solvent exposed, but is partially exposed when gp17 is in the extended state. This interface is neither fully heterodimeric, nor homodimeric; in contrast, its dynamic stability allows for packaging to occur, and it is neither highly stable nor highly unstable. Thus, the interface between subdomains in gp17 retains characteristics of both types of dimerization; this means that there is high stability in the compacted state, but with enough flexibility that it can extend under certain circumstances. This action is similar to the way a mechanical ratchet pushes a load, and is called a “molecular ratchet”.

In some cases, protein-protein interfaces include some residues that contribute substantially more to the free energy change of formation than others. The per-residue free energy contribution is generally determined by comparison of alanine scanning mutagenesis products with wild type[9,10]; residues contributing greater than 2 kcal/mol are labeled as “hot spots” which contribute substantially to the free energy change.

Tyrosine, tryptophan, and arginine are commonly found as hot spot residues, partly due to their large size}[11]. Hot spot residues are highly conserved, and are surrounded by a ring of energetically less important residues, called the “O-ring”. The relationship between a hot spot residue and the surrounding O-ring is part of a theory called the water exclusion hypothesis, which describes the stability of a complex as arising from only a small number of energetically crucial residues[12,13]. It is thought that this topological arrangement allows for the crucial hot spot residues to be almost water free, causing a more stable interaction[14].

To better understand the mechanism of gp17-mediated DNA packaging, we sought to understand and classify all important hydrophobic interactions formed at the interface between N- and C-terminal subdomains in gp17. Nonpolar interactions were measured in their surrounding interaction network. We also investigated how change in solvation upon compaction effects the interactions between charged and hydrophobic/ polar residues in the interface. By combining these results with our studies of electrostatic interactions in the compact and extended form of gp17, we are able to explain the effects observed in our mutational studies using computational methods and determine which interactions contribute substantially to the molecular action of gp17. Further, we are able to explain how the remainder of the change of free energy upon compaction arises. Our results indicate that computational modeling of dynamically active proteins reveals much more complex interactions than would be expected based on structural information alone.

4.3 Materials and methods

4.3.1 Molecular dynamics simulations

Molecular dynamics simulations and overall free energy calculation were done as described in Chapter 2. The free energy decomposition including per-residue contributions is described in detail in Chapter 3.

4.3.2 Mapping of inter-residue contacts

To understand the local environment that each residue occupies, we mapped the contacts formed between each atom of the residue to all other residue atoms within the protein, and to residues on the opposite side of the interface. We used the Chimera find clashes/contacts function. We selected each residue and checked against all other atoms within the model. We used a VDW (van der Waals) overlap of -2.75 Å, to simulate residues within the distance of one water molecule. We subtracted nothing from overlap for potentially H-bonding pairs, and we ignored contacts of pairs two or fewer bonds apart. We also ignored intra-residue contacts. This allowed us to construct a map of all atoms potentially interacting with the residue being probed.

To measure the contacts formed by a residue with residues on the other side of the interface, we selected the subdomain, and designated it as a “second set of designated atoms”, meaning contacts would be measured only between the selected residue and atoms within the set distance on the other subdomain. By measuring the overall contacts and the contacts across the interface, at each residue position for the compact and extended forms of gp17, we were able to find the differences and patterns of contacts and how they change with compaction.

4.4 Results

4.4.1 Free energy decomposition reveals hydrophobic contributions to compaction

Free energy decomposition (in Chapter 3) determined which charged amino acids contribute significantly to the overall free energy change of compaction. Although the electrostatic and hydrophobic contributions are roughly equal, this does not mean that either charged nor hydrophobic residues alone contribute 50%. Each interaction has a component of electrostatic and hydrophobic energies. For a charged amino acid to form a salt bridge, desolvation generally leads to an unfavorable hydrophobic contribution, while forming the Coloumbic interaction gives a favorable electrostatic term. Similarly, when a hydrophobic residue becomes more buried, it may contribute more favorably in its hydrophobic term, and if it makes electrostatic interactions through a polar atom, it may also contribute favorably to the electrostatic term. Chapter 3 identified five charged amino acids contributing less than -0.5kcal/mol (K19, E303, E309, R494, and E536), while seven of the 12 charged residues originally implicated[15] in the ratcheting mechanism do not contribute significantly to the free energy. These five charged residues contribute 29% of the total free energy change upon compaction, and we found that mutation of the most favorable salt bridge cripples DNA packaging. However, the large hydrophobic term implies that charged amino acids provide only a partial insight into motor activity. To fully understand the formation of the interface between N- and C-terminal subdomains, interactions between uncharged residues must also be explored.

We identified five nonpolar residues and one polar residue (Y295, W298, P310, T312, V531, and W533) that contribute significantly to the free energy change of

compaction. These six residues collectively contribute 46% of the free energy of compaction, shown in Table 4.1.

Table 4.1. Contributions of important hydrophobic residues to the overall free energy change. Free energy decomposition indicates the energetics of each residue in the interface region. Out of 23 hydrophobic amino acids at the interface, we find that five contribute significantly (< -0.5 kCal/mol) to the free energy of compaction. We also identified one polar residue (T312) contributing a substantial amount of the free energy change. It is important to note that the contributions from these residues comes from a mixture of hydrophobic and electrostatic sources, a finding that is discussed in this chapter. All units are kcal/mol.

residue	hydrophobic	electrostatic	contribution to total
Y295	-3.40 ± 2.12	-0.84 ± 0.64	-4.23 ± 2.14
W298	-1.35 ± 0.58	-0.81 ± 0.56	-2.16 ± 0.60
P310	-2.52 ± 0.78	0.14 ± 0.29	-2.38 ± 0.79
T312	-0.64 ± 0.42	-0.82 ± 0.52	-1.45 ± 0.46
V531	-3.91 ± 1.02	0.17 ± 0.38	-3.74 ± 1.02
W533	-2.84 ± 0.98	-0.75 ± 0.27	-3.59 ± 1.02

4.4.2 The change in nonspecific interactions formed by key hydrophobic residues upon compaction is correlated with free energy change

We have approximated the binding interface between the N- and C-terminal subdomains of gp17 as a protein-protein interaction to calculate free energy of binding with available tools, and to explain the role of important amino acids in leading to the translocation step of gp17 (see Chapter 3). Protein-protein interactions generally increase in stability with increasing buried surface area, the most common sizes being 1200-2000 Å^2. At 1000 Å^2, the interface between N- and C-terminal subdomains in gp17 is less stable than many of these interactions, a necessary structural feature that allows for the interface to contact and extend dynamically as packaging proceeds.

There is a strong correlation between the change in inter-atomic contacts formed by important residues upon compaction and free energy change at the position of the residue, shown in the insert in Figure 4.1. The number and location of contacts formed by hydrophobic residues is related to the degree to which water is occluded from contacting unfavorably. Larger residues tend to form more extensive interaction networks by virtue of their larger molecular footprint; in contrast, small residues require fewer close contacts to keep water from their greasy hydrocarbon sections. Many hydrophobic residues (tyrosine and tryptophan in particular) contain polar groups in addition to large nonpolar regions. These residues are able to hydrogen bond with nearby waters and often make fewer inter-residue contacts with their polar regions, allowing for the favorable water interactions, but occluding the majority of the residue. In this way, nonpolar residues can make both hydrophobic as well as electrostatic contributions to the overall free energy change.

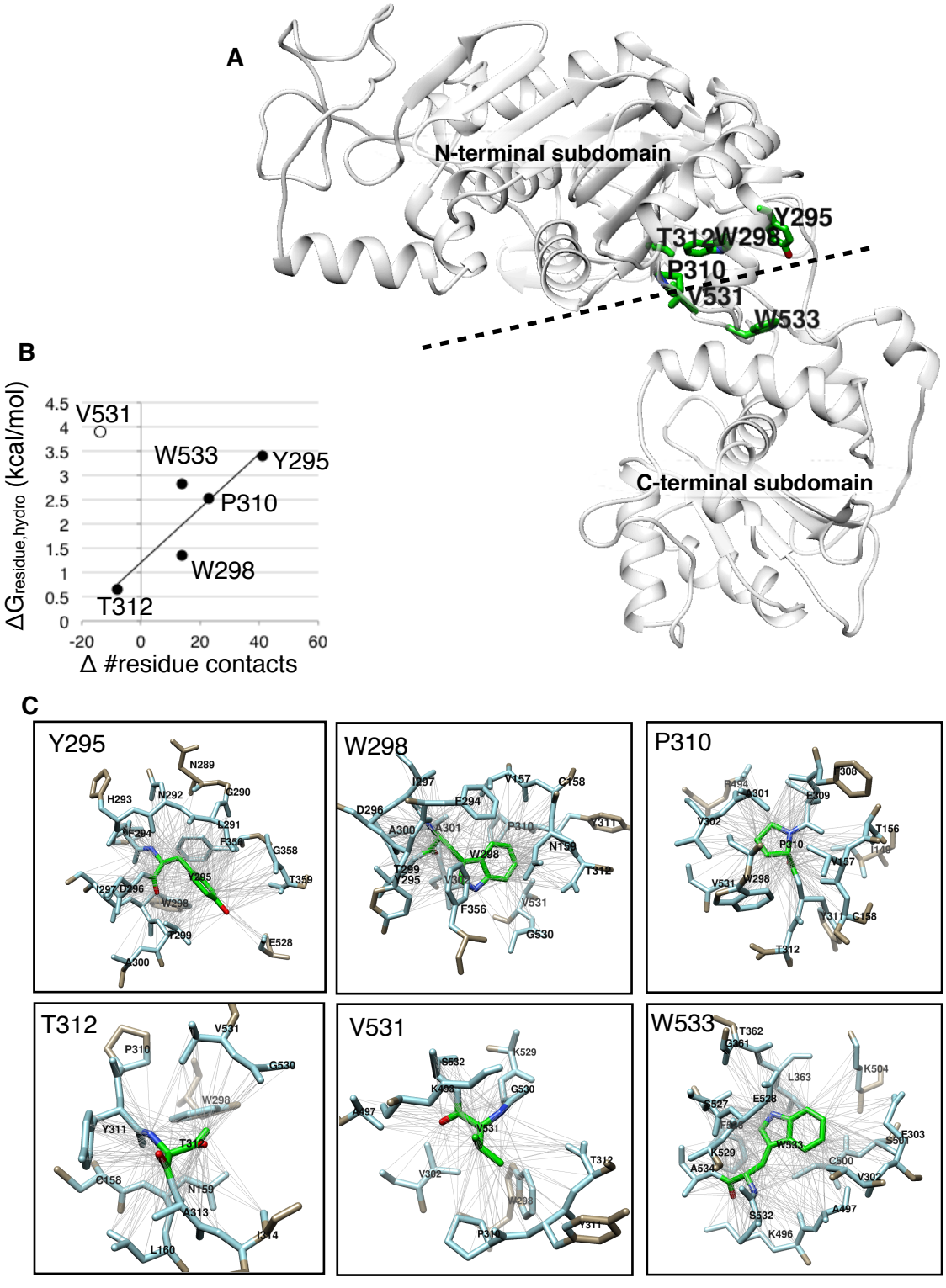
We show the interactions formed by each hydrophobic residue in the compact structure in Figure 4.1C. A depiction of the physical occlusion of each residue is found in Figure 4.2. The structure of the gp17 interdomain interface is such that a narrow cleft is formed between the subdomains that becomes more solvent accessible further away from the hinge region. For this reason, we can think of the hydrophobic residues at the edges of the interface as forming a seal that allows for the formation of optimal interactions within the interface. Indeed, we see far less water in the interface in the compact structure than the extended.

Fig 4.1. Interactions formed by each of the six energetically important hydrophobic or polar uncharged residues in the compact conformation.

(A) The position of each residue, in green, in relation to the interface between N- and C-terminal subdomains in a molecular dynamics averaged gp17 structure. Y295, W298, P310, and T312 lie in the N-terminal subdomain; V531 and W533 lie in the C-terminal subdomain. The dashed line indicates the interface plane between N- and C-terminal subdomains. Red and blue indicate oxygen and nitrogen respectively.

(B) Relationship between the residue hydrophobic free energy change and the change in number of contacts formed by the residue. V531, shown as the white circle and not included in the linear fit, has a very large free energy change caused by large changes in solvent accessibility, instead of additional contact formation.

(C) The interaction network of each residue. The residue probed is shown in green, while contacting atoms are shown in light blue, and noncontacting atoms belonging to residues with contacting atoms are shown in tan. Contacts are shown as gray lines. Atoms are in contact when the distance between them is less than the sum of the van der Waals radii and the diameter of one water molecule (2.75Å).


A
N-terminal subdomain
Y295
T312
W298
P310
V531
W533
C-terminal subdomain
B
ΔGresidue,hydro (kcal/mol)
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
V531
W533
Y295
P310
W298
T312
-20
0
20
40
60
Δ #residue contacts
C
Y295
W298
P310
T312
V531
W533

Tyrosine 295 acts as a gatekeeper to the interface. Behind it lie several other hydrophobic residues, which it shields from water molecules. It forms 406 inter-atomic contacts with 15 other residues. The aromatic axis of Y295 is turned to minimize the solvent contact, and it forms an important hydrogen bond between the ortho hydroxide and water molecules that exists both in the extended and compact conformations. Tyrosine, because of its large size and ability to form important hydrogen bonding interactions, often plays the role of “hot spot” at interfaces. The large free energy change at Y295 suggests that it may be revealed as a hot spot by alanine mutation. Tryptophan 298, in contrast to Y295, is further within the interface. It is located in the center of a helix that spans much of the distance between the ATP binding pocket and the interface. W298 has a free energy change of -2.16 ± 0.60 kcal/mol, of which -1.35 ± 0.58 is derived from hydrophobic forces, and -0.81 ± 0.56 from electrostatics. A channel allowing for the indole nitrogen to hydrogen bond with water exists, though the remainder of the residue is surrounded by other hydrophobic residues, occluding the hydrophobic bulk from solvent interactions. The surrounding molecular crowding can be seen in Figure 4.2. W298 forms 486 contacts with 20 residues. Because of the large number of contacts, as well as the large free energy change and the prevalence of tryptophan residues in the role of “hot spot”, we believe that W298 represents a hot spot within gp17. The surrounding large network of hydrophobic residues may represent the hot spot’s protective o-ring.

Proline 310 exists in a similar orientation between compact and extended gp17, and likely is involved in stabilization of a β-turn[16]. P310 has a free energy change upon compaction of -2.38 ± 0.79 kcal/mol, -2.52 ± 0.78 kcal/mol of that coming from hydrophobic forces (electrostatic contribution is unfavorable). P310 forms 13 additional

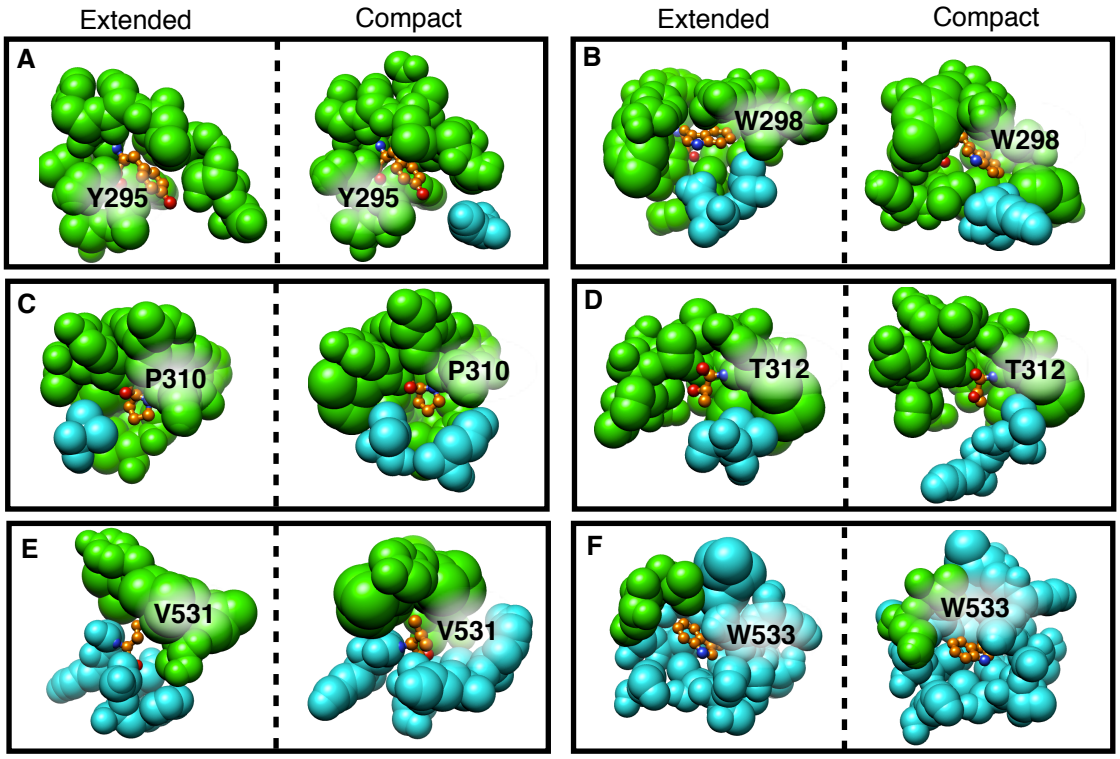


Fig 4.2. Difference in nearby residue environment for six energetically important uncharged residues. Green indicates residues on the N-terminal subdomain, and cyan indicates C-terminal subdomain. The residue being interrogated are shown in orange. Images generated using Chimera.
(A) Upon compaction, the residues surrounding Y295 become closer, helping to reduce interactions with water. This environment leads Y295 to be substantially more energetically favorable in the compact state. It forms many more interactions in the compact state, increasing from 364 in the extended state to 406.
(B) W298 forms 486 contacts with 20 residues in the compact state, an increase of 14 contacts from the extended state. This allows for a channel to become open to a water molecule that forms a favorable interaction with tryptophan's indole nitrogen atom.
(C) Proline commonly stabilizes loops within proteins. P310 forms 23 additional interactions in the compact state and keeps E309 in the correct position to interact with R494.
(D) T312 forms fewer interactions in the compact state to allow for hydrogen bonding with water. In addition, the polar side group rotates by 90° upon compaction, moving oxygen from a position contacting a neighboring hydrophobic residue and freeing it to interact with water.
(E) V531 forms 195 interactions in the extended form, which increases to 209 in compact. This is a substantial increase for a small amino acid. The rearrangement of the hydrophobic side chain upon compaction occludes interaction with water.
(F) W533 makes 14 more interactions when compacted than in the extended conformation. In addition to the favorable van der Waals contacts this imparts, the surrounding residues also greatly reduce the number of water molecules surrounding tryptophan while allowing for hydrogen bonding with the indole nitrogen atom.

contacts upon compaction, for a total of 265. This is fewer than Y295, W298, or W533, which is expected as it is a smaller residue.

Threonine is not generally involved in interface formation between proteins. However, T312 shows a large favorable free energy change, and like Y295, occupies a position occluding the central hydrophobic patch from water molecules. The free energy change upon compaction is distributed equally between hydrophobic and electrostatic interactions, which is expected for a polar residue. T312 actually forms more contacts (218) in the extended conformation than in the compact; it is more able to form hydrogen bonding interactions with water when compacted.

Valine rarely acts as a hot spot within interfaces because of its structural similarity to alanine, especially when compared to the other common hot spots tyrosine and tryptophan. However, V531 is a very important residue within gp17, contributing -3.74 ± 1.02 kcal/mol to the overall free energy, -3.91 ± 1.02 kcal/mol arising solely from hydrophobic forces, with only 0.17 ± 0.38 kcal/mol caused by electrostatic interactions. V531 forms only 195 contacts in the compact form, far fewer than the much larger hydrophobic residues.

Finally, a second important tryptophan exists at position 533. W533 forms 528 contacts in the compact form, the largest number of contacts of these six important residues. W533 also contributes a large amount to the overall free energy change, -3.59 ± 1.02 kcal/mol, -2.84 ± 0.98 kcal/mol arising from hydrophobic forces. It is entirely surrounded by hydrophobic residues aside from a channel formed through to its indole nitrogen atom. W533 also likely represents a hot spot within the protein, and the 22 contacting residues create the protective o-ring surrounding the residue.

4.4.3 Solvation has a large effect on non-specific interactions

We examined the degree to which important nonpolar residues and T312 form interactions with their surroundings by making contacts with residues on the same side of the interface, and to a lesser extent with residues across the interface. Though these van der Waals interactions are favorable, they are quite weak in comparison to interactions with water molecules. The hydrophobic term noted in Table 4.1, showing the breakdown of free energy per residue, is comprised of a sum of van der Waals and the solvation energy, the balance between the two differing depending on the residue.

We measured the distances between W298 and its nearest water neighbor occupying the location. An increase in distance to water shows the solvent exclusion effect of compaction. In Figure 4.3 (left), we show the distance from water to W298 extended by 1 Å upon compaction. The 4 Å distance seen in extended gp17 (Figure 4.3, right) is extremely unfavorable and favorably increases to 5 Å upon compaction. Compact gp17 also has substantially less water in close proximity to W298 than extended, adding to the favorability of the interaction, shown in Figure 4.4.

The distance between the closest resident water molecule and W533, shown in Figure 4.3, increases by 4 Å upon gp17 compaction. The surrounding region, however, has a similar number of water molecules. Examination of inter-residue contacts, shown in Figure 4.2, shows that W533 forms 528 contacts - substantially more than W298, which forms 486 contacts. This indicates that water is more occluded from contact with W533, which can be seen in Figure 4.4. The combination of the hydrophobic o-ring surrounding W533 and the increase in distance to the nearest water molecule shows that W533 is likely to have a greater affect on gp17 compaction.

Tyrosine is important to the formation of interfaces as it fills many of the same roles that tryptophan does. It is large, aromatic, and able to form hydrogen bonds

through the ortho-hydroxide group. Tyrosine 295 lies in a position that is partially solvent exposed in both compact and extended gp17, shown in figure 3.8. The restriction of adjacent hydrophobic amino acids surrounding Y295 in the compact structure (figure 4.2), as well as the increase of distance to the closest water molecule allows for it to undergo a favorable free energy change upon compaction. Figure 4.3 shows the change in distance to water that is contributed to formation of the interface between N- and C-terminal subdomains when DNA is translocated. As the distance between Y295 and water increases from 3 Å to 4 Å, the bulk of the interface water is expelled, and the surrounding residues compact. This may be an example of an "O-ring"[9,11] surrounding a highly-conserved interface residue. If this is the case, it is likely that the immediately surrounding/nearby residues[7], are also conserved, though to a lesser extent than Y295.

Proline is the most commonly found residue near interaction sites. It is thought that the unique structural attributes of proline help to maintain the positioning of key interaction residues, aiding in forming interaction interfaces[17]. In addition, it can commonly be found stabilizing beta turns while interacting with an adjacent aromatic residue[18]. At position proline 310, proline stabilizes a beta turn between the descending helix containing Y295 and W298 and the beta sheet containing T312 and E309. The positioning of these residues is crucial to interface formation and stability; the turn is stabilized both by P310 as well as by the salt bridge between E303 and K305. P310 is not directly involved in interface formation, as it is located several angstroms from the nearest residue on the opposite face, but instead acts to occlude the interface from solvent penetration. P310 may also be involved in creation of a favorable environment to optimize the interaction between E309 and R494 by

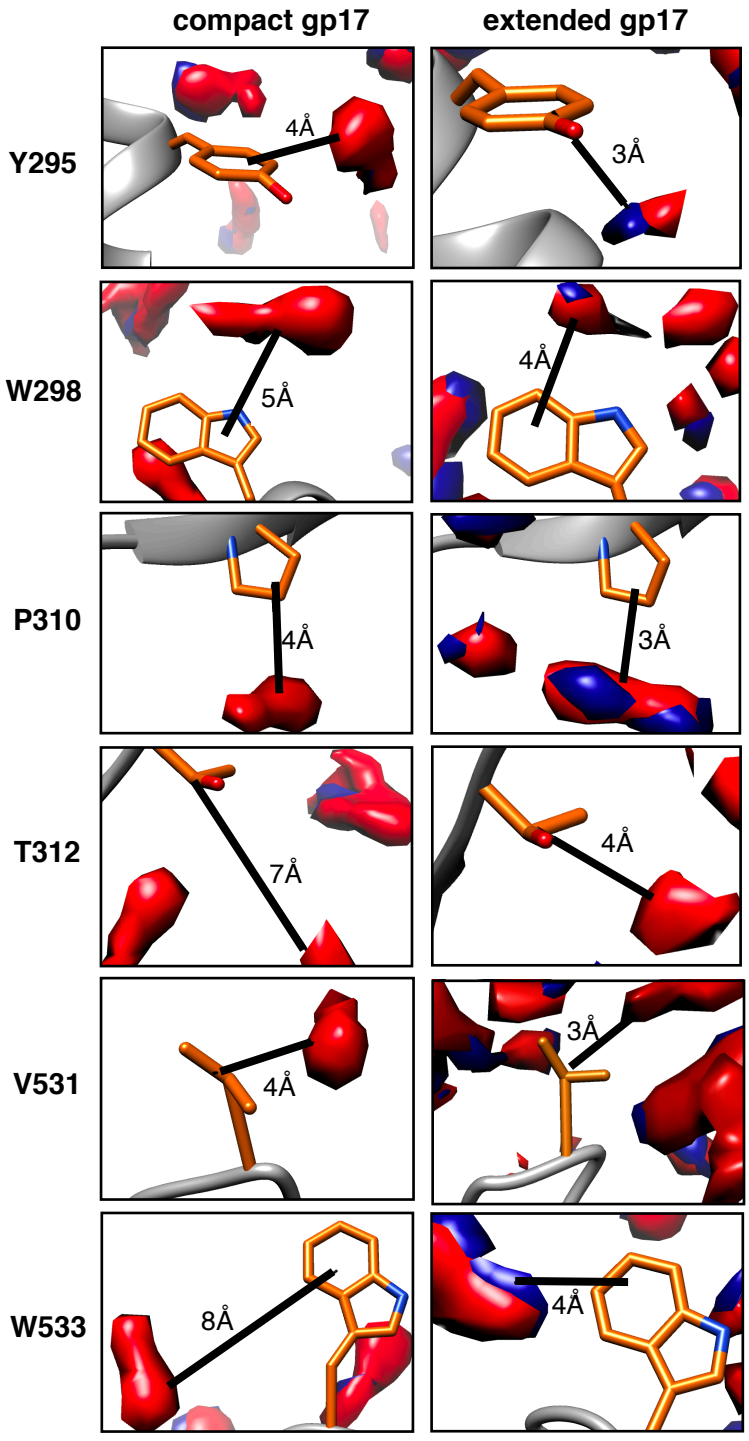


Fig 4.3. Wild type residue-level nearest water distances.
To measure the distance between hydrophobic and polar amino acids and the nearest water molecule, we first calculated an average structure of gp17, and the calculated the water occupancy. Shown here are the uncharged amino acids contributing the most to overall free energy change of compaction, listed in Table 4.1.

occluding solvent from one face of the interaction.

Threonine is unique in the interface in that it is a permanent dipole, able to form hydrogen bonding interactions with water as well as with other dipoles. Extended gp17 shows a close association between the terminal hydroxide of T312 and a nearby water molecule cluster, which also wraps around the entire side chain, offsetting favorable hydrogen bonding interactions with unfavorable hydrophobic contacts. Compaction leads to reorientation of this water, and expulsion of a large portion of it. Water still interacts closely with the hydroxide group, but no longer wraps around the side chain unfavorably.

Valine does not generally appear in interfacial hot spots[9]. Interestingly, V531 exhibits among the most favorable free energy change upon compaction (-3.74 ± 1.02 kcal/mol); here, the entire favorable free energy contribution is derived from hydrophobic forces. This is caused by two factors: the rearrangement of V531 allows for less interaction with fewer water molecules, and the surrounding hydrophobic residues interact via van der Waals forces. The short valine side chain branches before terminating in two equivalent methyl groups. The extended structure shows that the side chain is situated such that both methyl groups are pointed towards water molecules, a very unfavorable arrangement. In contrast, compact gp17 rotates V531 by 90 degrees, allowing for the methyl groups to face away from the methyl groups. This rearrangement, coupled with the overall reduction in water molecules surrounding V531, increases the distance to the closest water molecule from 3Å to 4Å, and the closer placement of surrounding hydrophobic residues allows for V531 to have a favorable free energy change upon compaction. This interaction is shown in the bottom row of Figure 4.4.

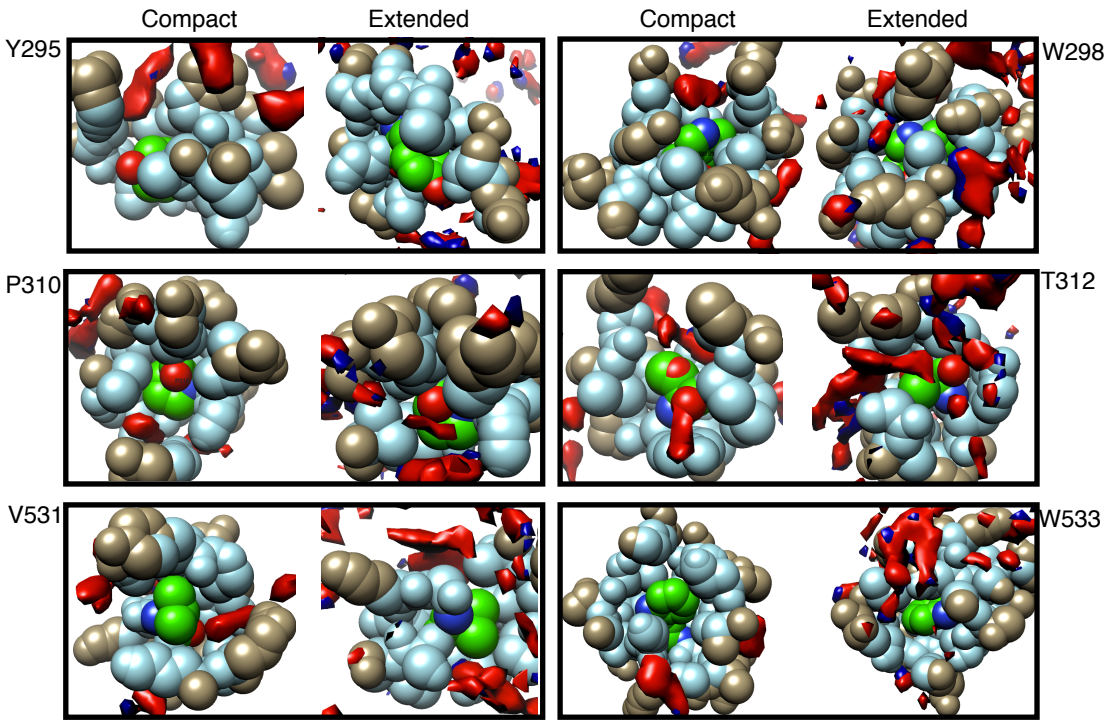


Fig 4.4. Wild type hydrophobic residue surroundings. This depiction of each important residue (in green) with contacting atoms of both subdomains (light blue) shows the amount that water (red and blue blobs) is able to contact each position. The water that is depicted in each image has an occupancy at its position of over 50%, meaning that in over 50% of the time, water is found at that position.
Y295 rotates by 90° to become better buried, as well as better occluded from surrounding water. As it is located at the exterior of the gp17 interface, this rotation makes it substantially less solvent accessible.
W298, in addition to becoming more buried overall, shows access to the indole nitrogen via a channel through which a water molecule is able to hydrogen bond with tryptophan.
P310 excludes far more solvent in the compact state than in the extended, a more favorable arrangement.
T312 is able to rotate by 90° to allow for hydrogen bonding with water through a channel not accessible in the extended state.
V531 reorients such that its side chain faces away from water molecules. In addition, the backbone atoms are able to form hydrogen bonds with water in the compact state.
Compaction substantially reduces the number of water molecules surrounding **W533**. By forming a larger number of contacts with surrounding hydrophobic residues and undergoing a side chain rotation of 180°, W533 becomes much more comfortable in the compact state.

4.4.4 Hot spots may effect gp17 activity

The location, size, and free energy contribution to the wild type conformational change suggest that Y295, W298, and W533 are hot spots. It has been shown that hot spot residues generally cluster together, and we have determined that these 3 residues

interact with each other across the interface. However, because the free energy change we have used to determine potential hot spot residues is that of residue contribution to wild type gp17 compaction, further investigation is needed to find the effect of mutating to alanine. Generally this is done experimentally by purifying a mutated protein and measuring the change in free energy upon binding to its interaction partner[9]. Recently use of computational alanine mutagenesis has been used to find hot spots[19] with good agreement with experimental mutagenesis[20].

The free energy of compaction for wild type, Y295A, W298A, and W533A is listed in Table 4.2. Each residue was simulated for a short time (2.4 - 3.2 ns) in both the compact and extended conformation to calculate the free energy change of compaction, and reveal the difference from wild type that can be attributed to mutation. Free energy values calculated for all three residues replaced with alanine mutations suggest these resides are hot spots. If these three residues are shown to be hot spots experimentally, we will be able to prove the role of hydrophobic residues in motor activity.

Table 4.2. Free energy values for wild type as well as alanine mutations to potential hot spot residues.
Free energy was calculated using molecular dynamics simulation data discussed in Chapter 2 over the course of shorter, diagnostic 8 ns simulations comprised of four parallel 2 ns simulations. Each value reported here is the change in free energy experienced upon gp17 going from extended to compact forms. Mutation of proposed hot spot residues to alanine is used to verify the importance of the residues. The three residues studied here were chosen due to the likelihood of tyrosine and tryptophan acting as hot spots due to their positioning, size, and free energy contribution to wild type gp17. The electrostatic contribution was calculated by sum of $\Delta G_{ele} + \Delta E_{GB}$. The hydrophobic contribution was calculated by sum of $\Delta G_{VDW} + \Delta E_{sol}$. ΔG overall was calculated as described in Chapter 2 methods. All values are in units of kcal/mol.

	electrostatic	**hydrophobic**	**ΔG**
wild type gp17	-18.37 ± 15.64	-21.67 ± 4.14	-38.21 ± 4.42
Y295A	1.57 ± 32.6	-4.47 ± 12.71	-5.3 ± 13.2
W298A	-2.75 ± 28.77	-0.89 ± 8.09	-4.98 ± 8.71
W533A	-3.24 ± 34.28	-5.74 ± 9.57	-9.142 ± 10.1

4.5 Discussion

Hydrophobic and polar residues make up 75% of the interface between N- and C-terminal subdomains in gp17, with the remaining 25% composed of charged residues. Using free energy decomposition to find the free energy contribution to compaction from each amino acid, we identified five hydrophobic residues and one polar residue that contribute -0.5 kcal/mol or more, indicating these residues as important to the compaction step. Each uncharged residue interacts with its surroundings through van der Waals and solvation effects, and occasionally through electrostatic interactions via a polar group. By understanding the change in these effects caused by the compaction of gp17, we have shown why each important

uncharged residue at the interface has the free energy contribution that we found by decomposition.

We analyzed both the local environment of each of these residues as well as solvation changes during compaction and found that a combination of closer packing and solvent exclusion at each position leads to their favorable contribution. We believe that these residues may represent hot spots within the interface, and that surrounding hydrophobic residues that compact to shield them from solvent may represent an “o ring”.

To create the close interactions necessary for the DNA translocation step by closure of the interface between N- and C-terminal subdomains, it is necessary that resident water molecules be largely expelled from the region. Interestingly, the number of water molecules found in an interface increases linearly with increasing buried surface area, even though these interfaces generally exclude the majority of water molecules[21]. Water forms ordered hydrogen bonds surrounding solvent-exposed charged and polar residues, interactions that are both electrostatically and entropically favorable. The formation of ion pairing between solvent-exposed, charged amino acids first necessitates the removal of complexed water molecules, an unfavorable step in forming a potentially very favorable interaction. Thus, ion pairing is only favorable if the energy gained by Coulombic forces between charged atoms is greater than that of desolvation.

In contrast, hydrophobic residues, as well as hydrophobic portions of residues, often do not form favorable interactions with water. Hydrophobic interactions, therefore, are favorable predominantly because of their exclusion of these unfavorable water molecules, though there is also a smaller direct interaction term in that of the van der

Waals forces. Hydrophobic residues not only exclude solvent from themselves, but often “seal” interfaces, protecting the delicate interactions that form between charged and polar residues and creating a reduced desolvation cost. Tyrosine, phenylalanine, and tryptophan are the largest standard amino acid residues. Tryptophan is the most prevalent amino acid found in so-called “hot spots” due to its aromaticity, which allows for π-stacking with other aromatic residues, large size allowing for protection of other hydrogen bonding interactions via solvent exclusion, and ability to participate in hydrogen bonding itself. Indeed, the role of Y295, W298, W533 in interface formation should not be underestimated, and experimental mutation to alanine would likely correlate our computational results and confirm these residues to be “hot spots”[9,11].

Protein-protein interfaces commonly contain a set number of hot spot residues per buried area. Hot spot residues increase the stability of an interface far more than other residues. However, the hot-spot location within the interface is very important. Residues that become more buried than others have a more favorable free energy change. Because large amino acids such as tryptophan have a physically larger footprint, burial leads to a greater overall change. Thus it is more difficult to effect change on a residue that only contains 7 amino acids than one that contains 14 simply because a 14 atom amino acid (such as tryptophan) forms more interactions. Size also matters, as hot spots are nearly always large residues. Shape complementarity is also crucial for the formation of stable interfaces, and altering the landscape by greatly reducing the size of an amino acid at a particular position can have major effects that are not seen with smaller amino acids. These reasons, along with the ability to form hydrogen bonds and ion pairing interactions, explain why tryptophan, tyrosine, and arginine are common hot spots. The presence of these amino acids in the gp17

interface shows the physical basis for much of the hydrophobic contribution to the enhanced stability of the compact state that allows for packaging to occur.

The realization that electrostatic interactions are not the only driving force behind gp17 compaction was unexpected. Ion pair formation between charged amino acids is a compelling way to explain motor activity. It is very tempting to make conclusions about the importance of charged residues based on the placement within a crystal structure, but as we have shown, these positions may be misleading or incorrect once the structure is allowed to equilibrate. This illustrates the importance of taking a combined experimental and computational approach to studying bistable or metastable proteins like gp17.

Interactions between charged atoms within a protein can exert attractive forces at longer distances, based on Coulomb's law. The original model of gp17 packaging implied that charge pairs play a role in pulling the two subdomains together to translocate DNA; the authors were partially correct as two ion pairs are directly involved in force generation by the formation of salt bridges. However, this original hypothesis suffered from the limitation that a static structure can be misleading, and amino acids are not fixed in position. In reality, when the distance between two charged amino acids increase, their side chain extension also becomes greater in order to keep the charged atoms in close proximity to one another. This causes the distance between charged atoms to change less than the 7 Å that the whole structure does, and this strongly effects the energetic contribution of the interaction. We showed, by experimental studies, that disruptive mutation of charged amino acids at the interface between N- and C-terminal subdomains hinders, but does not cripple, packaging in all cases other than E309K. This realization that not all charged amino

acids are created equal lead us to an extensive computational study of the roles of not just charged, but all amino acids important to the compaction of gp17 that translocated DNA.

The interface between N- and C-terminal subdomains is the result of billions of years of evolution that have lead to the development of a transiently stable interface to form in response to ATP binding, and to separate upon ADP + P_i product release. An interface lacking sufficient ion pairing interactions might not be able to stabilize the compact state, making translocation of 2bp of DNA at high forces impossible. If all 12 charged amino acids at the interface formed salt bridges in the compact state and were fully separated in the extended, it would be extremely difficult for the motor to reset following a single translocation step and gp17 would become a one-shot latch. While the force required to cause compaction is initiated by charged amino acids, full compaction is attained by both the charge interactions as well as the formation of large numbers of parallel hydrophobic interactions mediated by, in large part, a few crucial “hot spot” residues. The small size of the interface and the relative paucity of hot spot residues compared to expected values allows for gp17 to finish compacting, but does not permanently lock the protein in the compacted state.

A deeper understanding of viral motors is an important step towards understanding viral development from a basic science perspective. However, this understanding may have practical applications as well. The desire to move and manipulate DNA with temporal and spatial precision, especially in the development of next generation sequencing (NGS) techniques, is becoming a crucial aspect of modern medicine. The ability to pull DNA at a fixed speed through a nanopore is a requirement of many NGS methods, and motors such as gp17 may provide an effective route for

doing so on a massive scale. Although gp17 has the inherent drawback of having variable motor speed (see Chapter 2), we have shown that residue mutation does allow speed tuning. It is not unreasonable to expect that mutations could reduce the inherent variability in packaging speed as well. Further applications may be found in therapeutics by searching for small molecules that disrupt viral interface formation. By preventing the packaging of DNA, viral growth is inhibited. This could be useful in preventing human diseases that employ similar motors, such as herpesvirus.

The information we have collected showing the cooperation between charged and uncharged amino acid interactions may also help to explain other semistable protein interfaces. Most studies of protein-protein interfaces are based on crystal structure information. In the case of gp17, structural information helps to understand the mechanism of packaging because the two faces forming the interaction are covalently connected as a single protein. However, interactions between separate proteins that are weak and transient may not be retained during crystallization. The same kinds of interaction principals, i.e. fewer hot spots, smaller buried area, and few salt bridges to draw the interface together and align the faces, may be involved in these types of interactions as well. By demonstrating a more complete route to understanding gp17 compaction, we hope to enable deeper study of protein interactions wherever they may occur.

4.6 Acknowledgements

The author would like to acknowledge coauthors Gaurav Arya and Douglas E. Smith for their contributions to the work presented in this chapter.

References

1. Manosas, M., Xi, X. G., Bensimon, D. & Croquette, V. Active and passive mechanisms of helicases. *Nucleic Acids Research* **38,** 5518–5526 (2010).

2. Block, S. M. Kinesin Motor Mechanics: Binding, Stepping, Tracking, Gating, and Limping. *Biophysical Journal* **92,** 2986–2995 (2007).

3. Sun, S., Kondabagil, K. R., Draper, B., Alam, T. & Bowman, V. The Structure of the Phage T4 DNA Packaging Motor Suggests a Mechanism Dependent on Electrostatic Forces. *Cell* **135,** 1251-1262 (2008).

4. Chemla, Y. R., K. Aathavan, Jens Michaelis, Shelley Grimes, Paul J. Jardine, Dwight L. Anderson, and Carlos Bustamante. Mechanism of Force Generation of a Viral DNA Packaging Motor. *Cell* **122,** 683–692 (2005).

5. Hegde, S., Padilla-Sanchez, V., Draper, B. & Rao, V. B. Portal-large terminase interactions of the bacteriophage T4 DNA packaging machine implicate a molecular lever mechanism for coupling ATPase to DNA translocation. *J. Virol.* **86,** 4046–4057 (2012).

6. Yan, C., Wu, F., Jernigan, R. L., Dobbs, D. & Honavar, V. Characterization of Protein–Protein Interfaces. *Protein J* **27,** 59–70 (2007).

7. Keskin, O., Gursoy, A., Ma, B. & Nussinov, R. Principles of protein-protein interactions: what are the preferred ways for proteins to interact? *Chem. Rev.* **108,** 1225-1244 (2008).

8. Young, L., Jernigan, R. L. & Covell, D. G. A role for surface hydrophobicity in protein-protein recognition. *Protein Science* **3,** 717–729 (1994).

9. Bogan, A. A. & Thorn, K. S. Anatomy of hot spots in protein interfaces. *Journal of Molecular Biology* **280,** 1–9 (1998).

10. DeLano, W. L. Unraveling hot spots in binding interfaces: progress and challenges. *Current Opinion in Structural Biology* **12,** 14–20 (2002).

11. Keskin, O., Ma, B. & Nussinov, R. Hot Regions in Protein–Protein Interactions: The Organization and Contribution of Structurally Conserved Hot Spot Residues. *Journal of Molecular Biology* **345,** 1281-1294 (2005).

12. Li, X., Keskin, O., Ma, B., Nussinov, R. & Liang, J. Protein–Protein Interactions: Hot Spots and Structurally Conserved Residues often Locate in Complemented Pockets that Pre-organized in the Unbound States: Implications for Docking. *Journal of Molecular Biology* **344,** 781–795 (2004).

13. Li, J. & Liu, Q. 'Double water exclusion': a hypothesis refining the O-ring theory for the hot spots at protein interfaces. *Bioinformatics* **25,** 743–750 (2009).

14. Moreira, I. S., Fernandes, P. A. & Ramos, M. J. Hot spots-A review of the protein-protein interface determinant amino-acid residues. *Proteins* **68,** 803–812 (2007).

15. Sun, S., Kondabagil, K. R., Draper, B., Alam, T. & Bowman, V. The Structure of the Phage T4 DNA Packaging Motor Suggests a Mechanism Dependent on Electrostatic Forces. *Cell* **135,** 1251-1262 (2008).

16. Fu, H., Grimsley, G. R., Razvi, A., Scholtz, J. M. & Pace, C. N. Increasing protein stability by improving beta-turns. *Proteins* **77,** 491–498 (2009).

17. Kini, R. M. & Evans, H. J. A hypothetical structural role for proline residues in the flanking segments of protein-protein interaction sites. *Biochemical and Biophysical Research Communications* **212,** 1115–1124 (1995).

18. Bhattacharyya, R. & Chakrabarti, P. Stereospecific interactions of proline residues in protein structures and complexes. *Journal of Molecular Biology* **331,** 925–940 (2003).

19. Kortemme, T., Kim, D. E. & Baker, D. Computational Alanine Scanning of Protein-Protein Interfaces. *Science Signaling* **2004,** pl2–pl2 (2004).

20. Bradshaw, R. T., Patel, B. H., Tate, E. W., Leatherbarrow, R. J. & Gould, I. R. Comparing experimental and computational alanine scanning techniques for probing a prototypical protein-protein interaction. *Protein Engineering Design and Selection* **24,** 197–207 (2010).

21. Rodier, F., Bahadur, R. P., Chakrabarti, P. & Janin, J. Hydration of protein-protein interfaces. *Proteins* **60,** 36–45 (2005).

5 Concluding Remarks

Our research marks a substantial improvement in the understanding of gp17-mediated DNA translocation by combining structural analysis, single molecule and bulk biophysical techniques, and computational analysis. Experimental studies of mutations to amino acids proposed to participate in inter-subdomain ion pairing shows that these charged residues are involved in packaging - though to a lesser extent than originally expected. Further exploration using a computational model describing gp17 as a molecular ratchet helps to explain the observed magnitudes of the contributions. The 7 Å compaction found by comparison of cryo-EM and X-ray crystallographic data was approximated as a two-state system where gp17 is in either an extended (before translocating DNA) or compact (after translocating DNA) state. By simulating these two states in parallel and then finding the free energy difference between them, we were able to find the free energy change associated with the compaction of gp17 that causes DNA translocation. This different way of looking at the problem makes the analysis much more accessible to the tools used for binding studies. Using this method, we determined that the free energy change of -38.21 ± 4.42 kcal/mol associated with wild type gp17 compaction that becomes less favorable with mutation of certain charged amino acids at the interface. An excellent correlation between experimental measurements and the computed free energy change of the wild type and studied mutants shows that our computational model adequately describes the action of the motor. However, based on experimental work showing that charge-pair mutants retain some of the activity of wild type packaging, it became evident that the compaction of gp17 was mediated by more than just charged amino acids.

We found that the total free energy change in wild type could be broken down into its constituent forces to better understand the different forces acting to cause the compaction of gp17. A mechanism relying on the interaction between charged amino acids would likely have a free energy change comprised of nearly all electrostatic forces. In gp17, we found that of the total free energy of -38.21 ± 4.42 kcal/mol was distributed nearly equally between electrostatic (-18.37 ± 15.64 kcal/mol) and hydrophobic (-21.67 ± 4.14 kcal/mol) forces. This discovery led us to believe hydrophobic amino acids at the interface are involved in packaging as well. Determination of amino acids that contributed < -0.5 kcal/mol to compaction yielded a list of 11 amino acids important to the mechanism : five charged, five nonpolar, and one polar. In-depth analysis allowed us to understand the role of each in the translocation of DNA.

Of the five salt bridges originally proposed to form across the interface, only two form interactions capable of drawing the two subdomains together. E309-R494 forms the most energetically favorable interaction in the entire interface region, contributing a total of -6.95 ± 0.60 kcal/mol to the compaction of gp17 due to a large reduction in separation and the ability to retain much of its original solvation upon compaction. K19-E536 contributes substantially less to the overall free energy of compaction than that between E309-R494, only 1.38 ± 0.76 kcal/mol.This is due to a much smaller distance change to form the salt bridge, as well as the cost of complete desolvation upon compaction. Finally, K23-E537 forms a long-range ion pair in both the extended and compact conformation, perhaps acting as a tether such that the C-terminal subdomain does not rotate and cause ion pairs to become misaligned in the extended state. The remaining two originally proposed ion pairs, E303-K504 and K305-D505 do not interact

at all. As with many other salt bridges within proteins, they make local salt bridges (E303 with D505, K504 with E508, and D505 with K509) to stabilize structural elements. All three of these salt bridges stabilize elements that contain E309 or R494 and likely work to ensure the correct alignment between these two crucial amino acids.

The six uncharged amino acids identified through free energy decomposition are Y295, W298, P310, T312, V531, and W533. With the exception of the polar threonine (T312), these amino acids function to expel and exclude water from the interface, stabilizing the interactions formed between the two subdomains in compacted gp17. We measured the interaction networks of these amino acids by counting the interactions within a range defined as the van der Waal radii plus the radius of a single water molecule, and found an overall increase in the number of interactions formed by these amino acids. This increase is not universal- in some cases the number of interactions decreases, allowing for a polar atom to form a hydrogen bond with a nearby water molecule, as occurs with T312. The overall effect of compaction on these residues is a reduction in surrounding water molecules, allowing for more favorable van der Waals interactions to form.

Hot spots are amino acids that contribute more than -2 kcal/mol to the free energy of association for a given protein binding interaction. Tryptophan, tyrosine, and arginine are most commonly found to be hot spot residues, and the determination is made by the change in free energy when the residue is mutated to alanine. The prevalence of these amino acids as hot spots is due to their large size, and subsequent void left by the mutation to alanine, as well as the ability to form hydrogen bonding interactions or ion pairing interactions in the case of arginine. To test whether Y295, W298, and W533 could function as hot spots, we computationally mutated them to

alanine and found a large free energy reduction in each construct. This lends further credence to the importance of these positions to the stability of the interface.

Based on our experimental and computational work, we have shown that the DNA translocation step of gp17 is caused by a 7 Å compaction that is initiated by ATP binding, pulled together by shortening salt bridges, and stabilized by the formation of an interface between N- and C-terminal subdomains. Although not all charged amino acids at the interface are involved directly in forming inter-subdomain interactions, each plays a role in compaction. E309-R494 and K19-E536 form inter-subdomain salt bridges, while E303, K305, D504, and E505 stabilize structural elements by forming salt bridges within a single subdomain. Several hydrophobic residues function to expel water from the interface, stabilizing all interactions. Instead of a wholly electrostatic mechanism, we propose a similar but far more complete mechanism. ATP binding causes compaction, aided by ion pair formation. The relatively small area buried between N- and C-terminal subdomains allows for the interface to have a degree of instability that allows for the separation of what might otherwise be an interface too stable for the motor to reset using the energy of ADP + P_i product release.

It is tempting to make conclusions about the relationship between structure and function of proteins based on the information gleaned from X-ray crystallographic information. This may lead to valid conclusions about stable or permanent interfaces between co-crystallized proteins, but gp17 relies on dynamic instability at the interface to function. The two-state nature of gp17 combined with the inherent variability of the motor that likely arises from variation in the interface structure means that information from a static structure can be misleading. Alignment of charged amino acids at the interface implies that 5 salt bridges form in the compact state; electrostatic forces

seemed a natural fit to explain force generation. We now know that the dynamic nature of this interface caused an overestimation in the number of ion pairing interactions actually formed. Instead of five salt bridges, we find that two form across the interface as a result of compaction, and only one of these makes a substantial contribution to the free energy of compaction.

Understanding the underlying structural attributes that cause compaction during gp17-mediated DNA packaging allows for possibility of altering activity by changing the properties of the interface. There are many applications for a molecular motor that can translocate a payload quickly and at high forces, such as the assembly of nanoparticle structures and the translocation of DNA through pores necessary for next-generation DNA sequencing applications. It is also possible to think about engineering the speed or amount of force via mutation of residues; for example, an additional salt bridge could be added to gp17 by mutation of S501 to asparagine. This would likely form a salt bridge with K305 that would be limited to the compact structure, due to the close interaction between E303 and K305 that would dominate in the extended structure. This additional salt bridge would be semi-stable due to the proximity of E303, and could lead to higher force generation capabilities. We could also think about lowering the speed variability by tuning the residues, as would be needed for next generation DNA sequencing methods. To limit the variability of the motor, which may be due to the interface coming together in different conformations, each of which having slightly different stabilities, addition of another weak ion pair (like K23-E537) that exists in both compact and extended conformations could help the motor to be more rotationally constrained. The addition of a small molecule that limits the hydrophobic interactions from coming together in the compact structure would allow for the motor to speed up

by making the interface less stable and resetting of the motor to be quicker. Conversely, stabilization of the compact state would likely slow the motor. Regardless of the application, the principals governing the interaction between N- and C-terminal subdomains in gp17 help guide further work and expands the possibilities of selectively changing motor function to suit ones needs.

III p53-MEDIATED LOOP FORMATION IN HUMAN GENOMIC DNA

6 Direct measurement of formation of loops in DNA by a human tumor suppressor protein.

Amy Migliori[1], Samuel Kung[2], Danielle Wang[2], and Douglas Smith[2]

6.1 Summary

In previous work, we developed methods using optical tweezers to measure protein-mediated formation of loops in DNA structures that can play an important role in regulating gene expression. We previously applied this method to study two-site restriction endonucleases[1-3], which were convenient model systems for studying this phenomenon. Here we report preliminary work in which we have applied this method to study p53, a human tumor suppressor protein, and show that we can measure formation of loops. Previous biophysical evidence for loops comes from relatively limited qualitative studies of fixed complexes by electron microscopy[4]. Our results provide independent corroboration and future opportunity for more quantitative studies investigating structure and mechanics.

Keywords: p53, DNA looping, optical trap, protein binding

6.2 Introduction

[1]Department of Biochemistry, University of California, San Diego. La Jolla, CA, 92093.

[2]Department of Physics, University of California, San Diego. La Jolla, CA, 92093.

Interactions between proteins and DNA are crucial to many cellular processes, including gene expression, DNA compaction, and cell division, among others. There are several methods by which protein is able to interact with DNA as well as with other proteins. Protein-DNA interactions may occur non-specifically in order to aid in replication (single strand binding protein)[5] or to search for a specific binding site[6-8] by charged interactions with the DNA backbone or topological features, or can be sequence-specific via specific interactions between amino acids and DNA. Protein-protein interactions generally occur via a combination of hydrophobic interfaces and charge complementarity between the two proteins.

Transcription factors (TFs) are proteins that influence the expression of genes within cells by binding either directly to DNA/associated proteins or otherwise affecting the transcription of genes within a cell. Generally, TFs bind to DNA in a location in the promoter region near the gene being regulated to recruit other factors or affect binding of key transcription machinery in a steric manner, but there are many instances of distant binding sites (hundreds to thousands of basepairs away) that affect gene expression. In some cases of distant binding, the affect of the TF is shown to be due to DNA looping. To adapt to its environmental and nutritional needs, a cell dynamically regulates the TFs present to control what proteins are produced. p53 is a transcription factor that halts cell division if DNA damage is present and can induce apoptosis (cell death) in cells deemed too damaged to fix. p53 has been referred to as the “guardian of the genome” because it helps to ensure that cells containing damaged genomes are not propagated. p53’s critical role in cell proliferation means that its mutation can lead to unregulated cell division and the growth of tumors; indeed, it is mutated in 50% of human cancers.

p53 binds DNA as a dimer of dimers through recognition of two 10bp half sites5, together forming a 20bp full recognition sequence[9]. There are many published structures showing p53 bound at the recognition sequence[10,11] and the protein regions contributing to binding are well established. It is known that a variable length spacer can intercalate these two half sites, and that the intervening DNA bends strongly to allow for tetramerization[12]. This recognition sequence has significant plasticity, as well as variable spacer length between half-sites, and occurs many times in the human genome[13] to allow for transcriptional regulation.

Some genes regulated by p53 contain multiple p53RE, and these sites can be located thousands of base pairs away from the transcriptional start site. It is thought that these distal binding sites affect activity by mediating DNA looping. Although p53 has been studied extensively due to its involvement in cancer development[14], little is known about the physical basis and properties of DNA loop formation allowing for distal binding site association with the transcription complex. Physical evidence for looping is limited and based on electron microscopy images which show some evidence for p53-mediated loop conformations in DNA including p53 recognition elements (p53RE)[4]. Additionally, there have been molecular biology studies showing reduced transcription upon removal of distal p53RE of target genes in cotransfection assays with a reporter gene[15,16]. Taken together, the pictorial and biochemical evidence points to the importance of DNA looping to enhance gene transcription, as seen in Figure 6.1. Using a dual optical trap setup, we measure p53-mediated looping of human genomic DNA containing multiple p53RE. We believe this to be the first direct biophysical measurement of the looping properties of p53 in a physiological substrate.

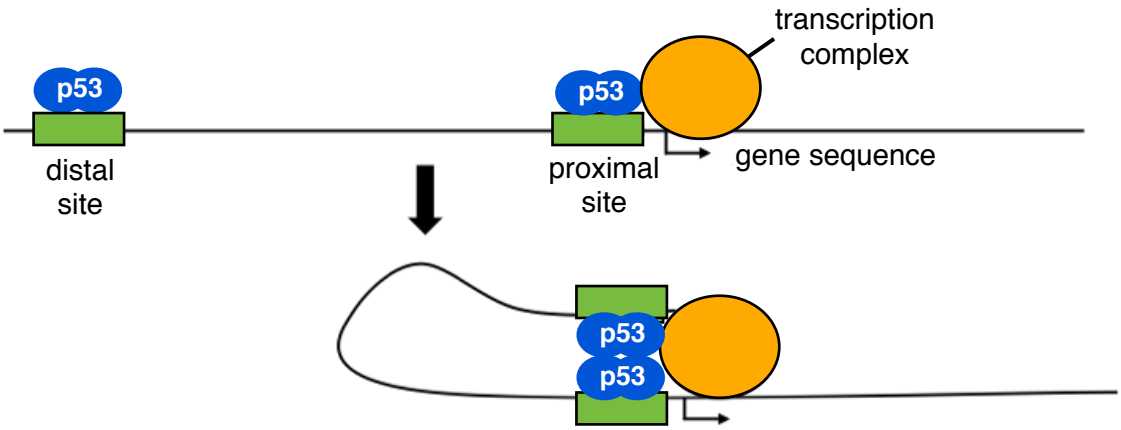


Fig 6.1 Cartoon schematic of p53-mediated DNA looping. This representation shows potential p53-mediated looping as suggested by imaging studies showing stacking of p53 molecules at loop junctions and molecular biology studies showing reduced transcription upon removal of either distal or proximal p53 binding sites. In this mechanism, the interaction between two bound p53 tetramers occurs when the intervening DNA forms a loop stabilized by the protein-protein interaction of the two bound p53s.

6.3 Materials and Methods

6.3.1 Preparation of dual-labeled Col18A DNA

Human DNA was purified using a Qiagen QIAamp DNA Mini kit via buccal swabs and a 10kbp section near the Col18A gene was amplified using PCR with the following primers: TTGAAGGCATGCAGGAATAAC and TGGGCTTTGCTGAGAGAAAC. The primers were tagged with biotin and digoxygenin, similar to the method used previously[17]. This dual-labeled construct comprises the DNA portion of our experimental setup. Figure 6.2 shows the different loop sizes that are possible with the Col18A five p53RE construct.

6.3.2 Manipulation of DNA-p53 complexes with optical tweezers

Purified p53 was generously provided by Prof. Hector Viadiu (UCSD) and was

prepared as described previously[18]. After attaching the DNA to the streptavidin beads, we incubated the SA bead-DNA mixture with the chosen concentration of p53 (1μM-100μM) for 20 minutes at room temperature prior to injection into the experimental fluid chamber. This allowed for p53 to bind to DNA and to form loops between p53RE while limiting the amount of p53 needed for the experiment.

After incubation, the measurements were carried out in 1x PBS buffer. We used a dual-trap optical tweezers apparatus in which one beam was steered relative to the other using an acousto-optic deflector, used and calibrated as described previously[19]. First, a protein G bead was caught in the immobile trap. Second, a SA-DNA bead was caught in the moveable trap. A schematic of the experimental setup can

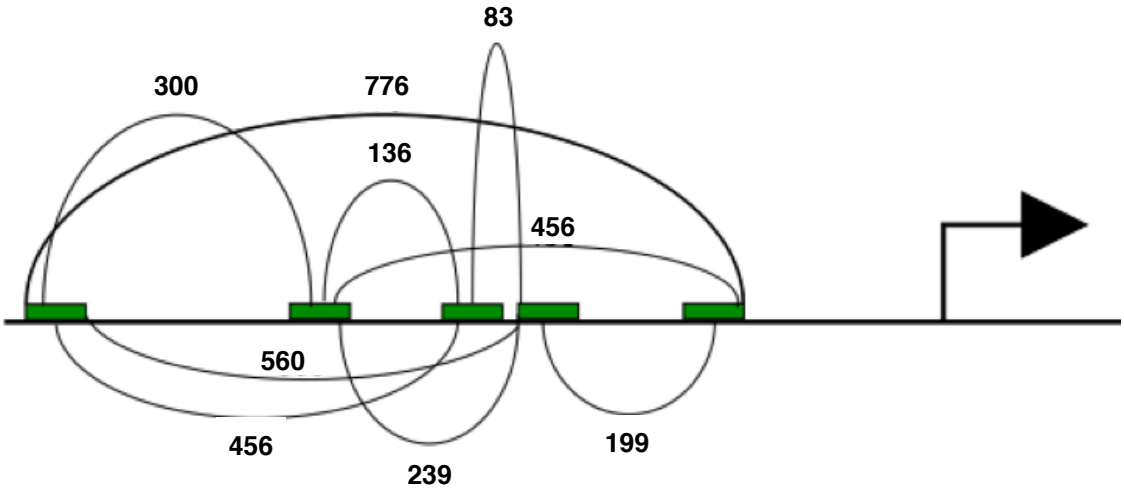


Fig 6.2 Possible sizes of loops formed in Col18A DNA by p53 binding. Purified Col18A contains five p53RE upstream of the transcriptional start site. These five sites can form a total of nine different sized loops. We used these loop sizes to aid in analysis of the data from our optical tweezers experiments.

be seen in figure 6.3.

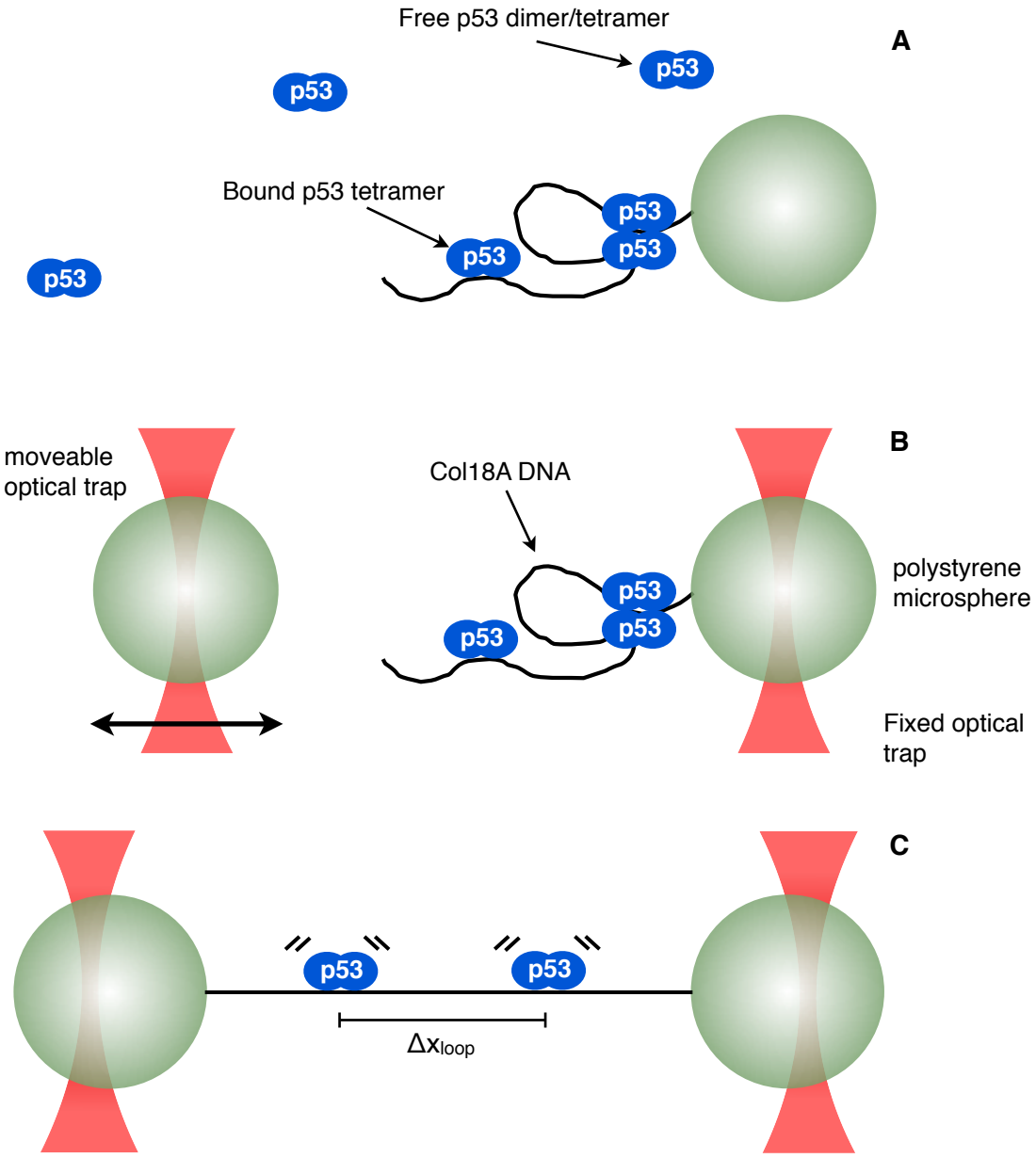


Fig 6.3 Schematic of dual optical trap p53 looping experiments. (A) A 10,000bp region near the Col18A gene was purified from human DNA and double labeled using biotin and digoxygenin to allow for attachment to a 2.3 µm polystyrene bead. Following this, purified p53 was added and allowed to attach to the five p53RE in this segment. (B) The DNA-p53 beads were trapped in one of the two the optical traps (right); the other bead (left) is a second 2.3 µm polystyrene bead coated in anti-digoxygenin antibodies to attach the other end of DNA. The moveable trap brings the two beads together, allowing for DNA to tether at both ends, and apart again to stretch the DNA. (C) At a particular force, the loop ruptures, showing an instantaneous drop in force that allows for the measurement of rupture force and loop size.

6.4 Results

6.4.1 p53 mediated looping of Col18A promoter region DNA

Col18A is a gene that encodes for collagen type XVII in humans. Misregulation of collagen is associated with metastasis of cancer cells through the extracellular matrix17. The promoter region of Col18A contains five p53RE[18] with a spacing of 83-776bp between the closest and furthest p53RE, respectively. Optical tweezers were used to detect loops by stretching the DNA and measuring the force-extension behavior, as described previously[1-3]. We measured loops formed in a ~10kb

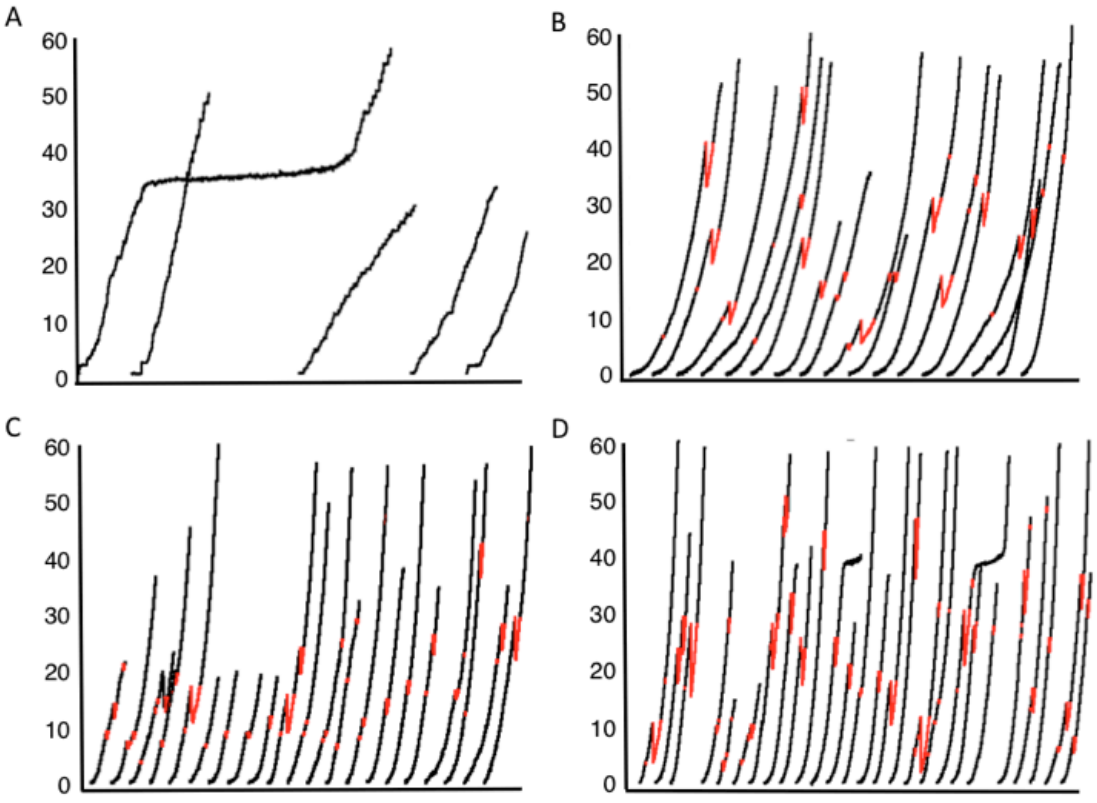


Fig 6.4 Representative experimental traces of DNA with and without p53. (A) 20kb with no binding sites, 100μM full length p53. (B) 10kb with binding sites, 1μM FLp53. (C) 10kb with binding sites, 10μM FLp53 D. 10kb with binding sites, 100μM FLp53

section of the promoter region containing these five p53RE and analyzed loops formed at various concentrations of p53.

At all concentrations of p53 studied, from 1μM to 100μM, loops of various sizes were detected as evidenced by abrupt drops in force while stretching DNA, which correspond to rupture of the loop junction, which releases a length of slack approximately equal to the length of DNA in the loop[1-3]. Figure 6.4 shows, for each concentration of p53, loops formed in the DNA construct rupturing as the force rises. Loops, marked in red, are counted by size in Figure 6.5. Although we did not have sufficient resolution in these measurements to distinguish the exact loop sizes, the measured lengths were mostly consistent with the range of sizes expected for the inter-site spacings of 83-776bp. Instances where different loop sizes were inferred may correspond to cases where the protein is non-specifically bound to the DNA.

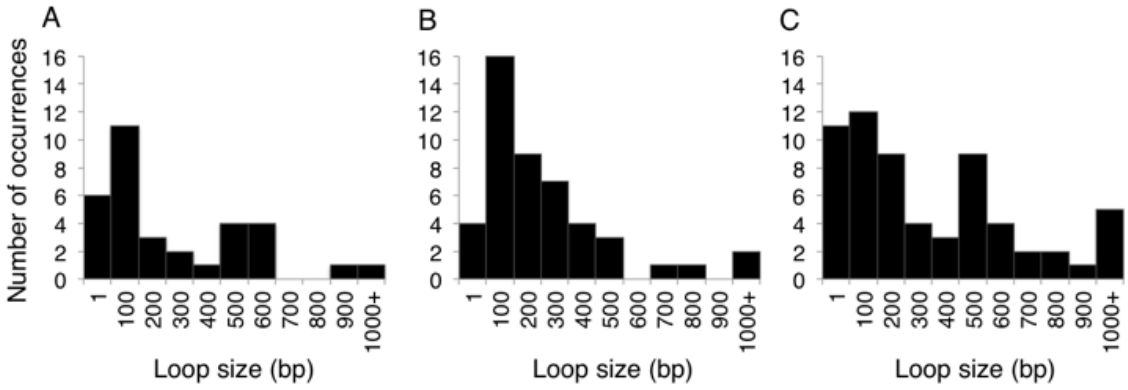


Fig 6.5. Histogram of observed loop sizes at each concentration of p53 studied. (A) 1μM p53, (B) 10μM p53, and (C) 100μM p53.

At the concentrations of p53 studied, there is a strong preference for 1-2 loops per DNA molecule. Figure 6.5 shows histograms of the number of loops observed per experimental measurement at each concentration of p53. Though there is a preference for one or two loop, there are several instances at each concentration of 3, and

sometimes more, loops observed. Increasing p53 concentration seems to increase the probability of formation of larger number of loops.

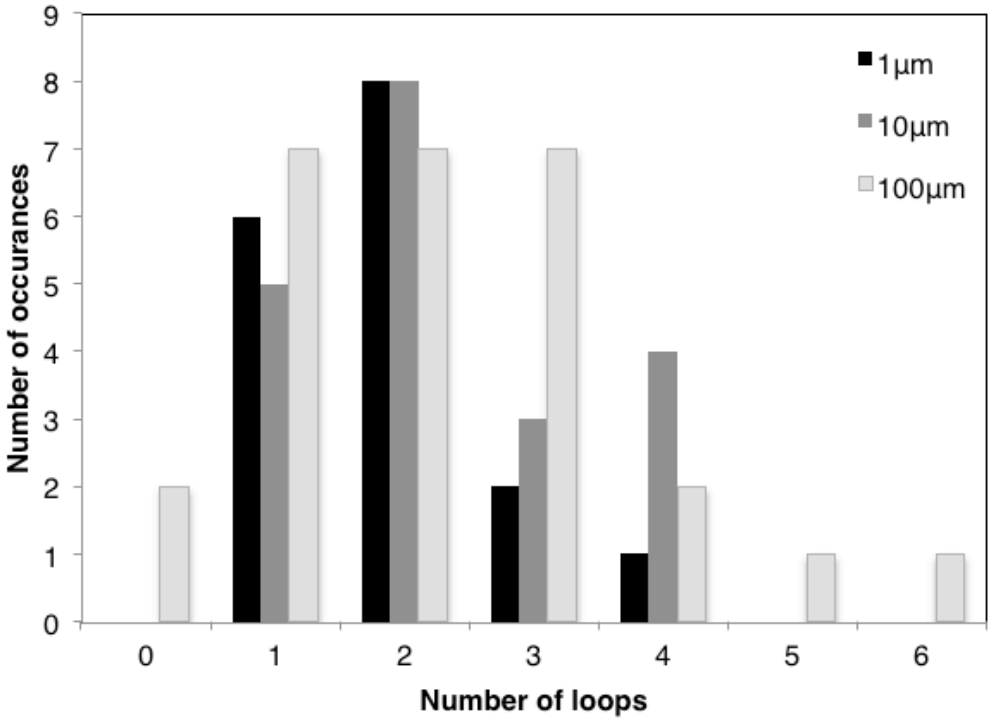


Fig 6.6 Number of loops observed for each experimental measurement at different concentrations of p53. We found that increase in p53 concentration also increased number of loops formed in the Col18A DNA. Although we expected a maximum of 2 loops per DNA, frequently 3 or more were observed.

6.5 Discussion

We present novel evidence for the p53-mediated formation of loops in DNA containing five p53 recognition sequences. Based on a traditional understanding of protein-mediated DNA looping, one would expect each response element (binding site) to potentially mediate one looping interaction with one other site. In this case, given five sites, one might have expected to observe only zero, one, or two loops per DNA. Interestingly, while most stretching events did show two or fewer loops, events were occasionally observed with

three force drops, implying three loops. This finding suggests a non-orthodox DNA looping activity for p53 in which the protein is able to mediate multiple loops between one response element and multiple other response elements, which may have important implications for regulation of gene expression.

p53 forms tetramers on a single full binding site through a well-known tetramerization domain. However, a different and distinct binding mode for tetramerization has been proposed[11] that allows for a tetramer to form spanning two half-sites via a conformational change. This alternate binding geometry could allow for each p53RE to participate in more than one tetramer interaction, and could possibly explain our findings.

Though this preliminary data shows promise in terms of explaining p53-mediated looping in a DNA construct containing multiple p53RE, more work is necessary to obtain quantitative information about interaction affinities and further investigate the apparent multi-site looping.

6.6 Acknowledgements

The author would like to acknowledge coauthors Samuel Kung, Danielle Wang, and Douglas E. Smith for their contributions to the work presented in this chapter.

References

1. Gemmen, G. J., Millin, R. & Smith, D. E. DNA looping by two-site restriction endonucleases: heterogeneous probability distributions for loop size and unbinding force. *Nucleic Acids Research* **34,** 2864–2877 (2006).

2. Gemmen, G. J., Millin, R. & Smith, D. E. Dynamics of single DNA looping and cleavage by Sau3AI and effect of tension applied to the DNA. *Biophysical Journal* **91,** 4154–4165 (2006).

3. Gemmen, G. J., Millin, R. & Smith, D. E. Tension-dependent DNA cleavage by restriction endonucleases: two-site enzymes are 'switched off' at low force. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 11555–11560 (2006).

4. Stenger, J. E., Peter Tegtmeyer, Gregory A. Mayr, Michael Reed, Yun Wang, Pin Wang, Paul V.C. Hough, and Iris A. Mastrangelo. p53 oligomerization and DNA looping are linked with transcriptional activation. *The EMBO Journal* **13,** 6011–6020 (1994)

5. Wold, M. S. Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu. Rev. Biochem.* **66,** 61–92 (1997).

6. Bonnet, I., Andreas Biebricher, Pierre-Louis Porte ́, Claude Loverdo, Olivier Benichou, Raphael Voituriez, Christophe Escude ́, Wolfgang Wende, Alfred Pingoud and Pierre Desbiolles. Sliding and jumping of single EcoRV restriction enzymes on non-cognate DNA. *Nucleic Acids Research* **36,** 4118–4127 (2008).

7. Wang, Y., Austin, R. & Cox, E. Single Molecule Measurements of Repressor Protein 1D Diffusion on DNA. *Phys. Rev. Lett.* **97,** 048302 (2006).

8. Liu, Y. & Kulesz-Martin, M. p53 protein at the hub of cellular DNA damage response pathways through sequence-specific and non-sequence-specific DNA binding. *Carcinogenesis* **22,** 851–860 (2001).

9. Wang, Y., Schwedes, J. F., Parks, D., Mann, K. & Tegtmeyer, P. Interaction of p53 with its consensus DNA-binding site. *Mol. Cell. Biol.* **15,** 2157–2165 (1995).

10. Chen, Y., Dey, R. & Chen, L. Crystal Structure of the p53 Core Domain Bound to a Full Consensus Site as a Self-Assembled Tetramer. *Structure/Folding and Design* **18,** 246–256 (2010).

11. Aramayo, R. Michael B. Sherman, Kathryne Brownless, Rudi Lurz, Andrei L. Okorokov and Elena V. Orlova. Quaternary structure of the specific p53-DNA complex reveals the mechanism of p53 mutant dominance. *Nucleic Acids Research* **39,** 8960–8971 (2011).

12. Nagaich, A. K., Victor Zhurkin, Stewart Durll, Robert Jernigan, Ettore Appella, and Rodney Harrington. p53-induced DNA bending and twisting: p53 tetramer binds on the outer side of a DNA loop and increases DNA twisting. *Proc. Natl. Acad. Sci. U.S.A.* **96,** 1875–1880 (1999).

13. Wei, C.-L., Qiang Wu, Vinsensius B. Vega, Kuo Ping Chiu, Patrick Ng, Tao Zhang, Atif Shahab, How Choong Yong, YuTao Fu, Zhiping Weng, JianJun Liu, Xiao Dong Zhao, Joon-Lin Chew, Yen Ling Lee, Vladimir A. Kuznetsov, Wing-Kin Sung, Lance D. Miller, Bing Lim, Edison T. Liu,

Qiang Yu, Huck-Hui Ng, and Yijun Ruan. A Global Map of p53 Transcription-Factor Binding Sites in the Human Genome. *Cell* **124,** 207–219 (2006).

14. Muller, P. A. J. & Vousden, K. H. p53 mutations in cancer. *Nat. Cell Biol.* **15,** 2–8 (2013).

15. Jackson, P. & Yardley, G. Distinct distal and proximal p53-binding sites in the MCK promoter govern the transcriptional response to p53. *FEBS Lett.* **406,** 271–274 (1997).

16. Jackson, P., Iris Mastrangelo, Michael Reed, Peter Tegtmeyer, Gina Yardley and Julie Barrett. Synergistic transcriptional activation of the MCK promoter by p53: tetramers link separated DNA response elements by DNA looping. *Oncogene* **16,** 283–292 (1998).

17. De Wever, O, An Hendrix, Astrid De Boeck, Wendy Westbroek, Geert Braems, Shahin Emami, Michele Sabbah, Christian Gespach, Marc Bracke. Modeling and quantification of cancer cell invasion through collagen type I matrices. *Int. J. Dev. Biol.* **54,** 887–896 (2010).

18. Miled, C. A Genomic Map of p53 Binding Sites Identifies Novel p53 Targets Involved in an Apoptotic Network. *Cancer Research* **65,** 5096–5104 (2005).

19. Fuller, D. N., Gregory J. Gemmen, John Peter Rickgauer, Aurelie Dupont, Rachel Millin, Pierre Recouvreux and Douglas E. Smith. A general method for manipulating DNA sequences from any organism with optical tweezers. *Nucleic Acids Research* **34,** e15 (2006).

APPENDIX A: Dual optical trap diagram

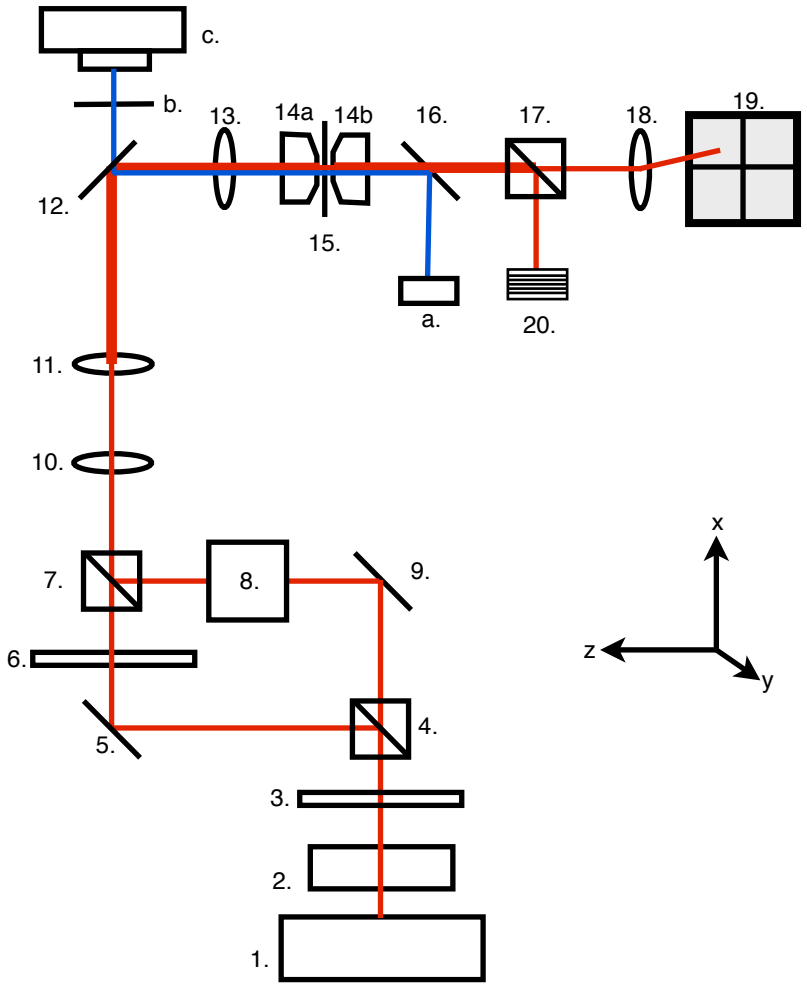


Fig A1. Dual optical trap setup optical diagram.
Numbered : trapping optical pathway. A 1064nm Nd-YAG laser is split and one beam is steered using an acousto-optic deflector, creating a steerable second trap. Beam deflection due to experimental forces is detected on a quadrant photodiode, and is converted to units of force.

1) Nd-YAG fiber laser λ=1064nm. 2) Faraday isolator. 3) 1064 half wave plate. 4) IR beam splitter. 5) Broadband mirror. 6) 1064 half wave plate. 7) IR beam splitter. 8) Acousto-optic deflector. 9) Broadband mirror. 10) AR.18 coated 75mm lens. 11) AR.18 coated 150mm lens. 12) Dichroic. 13) AR.18 coated lens. 14a) Objective: 1.2NA 60x water. 14b) Collector: 1.2NA 60x water. 15) Steerable sample chamber. 16) Dichroic. 17) Polarizing beam splitter 1064-100. 18) 75mm IR lens. 19) Photodiode. 20) Beam Dump

Lettered : imaging optical pathway. A blue LED is passed through the experimental sample chamber and steered away from the trapping pathway to a video camera using a dichroic mirror to provide real-time experimental imaging.

a) Blue LED. b) IR filter c) Video camera

APPENDIX B: Steered Molecular Dynamics

```
#############################################################
## JOB DESCRIPTION                                         ##
#############################################################

# N- C- terminal constant velocity pulling


#############################################################
## ADJUSTABLE PARAMETERS                                   ##
#############################################################

structure          ../common/pullopen/T4_gp17_wb_ion_2.psf
coordinates        ../common/pullopen/gp17equilinwb.pdb
outputName         112211threefixedatoms

set temperature    310

# Continuing a job from the restart files
if {0} {
set inputname      myinput
binCoordinates     $inputname.restart.coor
binVelocities      $inputname.restart.vel  ;# remove the "temperature" entry if you use
this!
extendedSystem           $inputname.xsc
}

firsttimestep      0


#############################################################
## SIMULATION PARAMETERS                                   ##
#############################################################

# Input
paraTypeCharmm          on
parameters          par_all27_prot_lipid.inp

# NOTE: Do not set the initial velocity temperature if you
# have also specified a .vel restart file!
temperature         $temperature


# Periodic Boundary conditions
# NOTE: Do not set the periodic cell basis if you have also
```

```
# specified an .xsc restart file!
if {0} {
cellBasisVector1    100.0  0    0
cellBasisVector2     0   90.0  0
cellBasisVector3     0    0   100.0
cellOrigin          47.30540084838867 -32.351470947265625 6.774631023406982
}
wrapWater           on
wrapAll             on


# Force-Field Parameters
exclude             scaled1-4
1-4scaling          1.0
cutoff              12.0
switching           on
switchdist          10.0
pairlistdist        14.0


# Integrator Parameters
timestep            2.0  ;# 2fs/step
rigidBonds          all  ;# needed for 2fs steps
nonbondedFreq       1
fullElectFrequency  2
stepspercycle       10


#PME (for full-system periodic electrostatics)
if {0} {
PME                 yes
PMEGridSpacing      1.0

#manual grid definition
#PMEGridSizeX        32
#PMEGridSizeY        32
#PMEGridSizeZ        64
}


# Constant Temperature Control
langevin            off    ;# do langevin dynamics
langevinDamping     1     ;# damping coefficient (gamma) of 5/ps
langevinTemp        $temperature
langevinHydrogen    no    ;# don't couple langevin bath to hydrogens


# Constant Pressure Control (variable volume)
```

```
if {0} {
useGroupPressure      yes ;# needed for 2fs steps
useFlexibleCell       no  ;# no for water box, yes for membrane
useConstantArea       no  ;# no for water box, yes for membrane

langevinPiston        on
langevinPistonTarget  1.01325 ;#  in bar -> 1 atm
langevinPistonPeriod  100.0
langevinPistonDecay   50.0
langevinPistonTemp    $temperature
}


restartfreq         500     ;# 500steps = every 1ps
dcdfreq             100
xstFreq             500
outputEnergies      100
outputPressure      100


# Fixed Atoms Constraint (set PDB beta-column to 1)
if {1} {
fixedAtoms          on
fixedAtomsFile      ../common/gp17eqinwater.ref
fixedAtomsCol       B
}


# IMD Settings (can view sim in VMD)
if {0} {
IMDon           on
IMDport         3000    ;# port number (enter it in VMD)
IMDfreq         1       ;# send every 1 frame
IMDwait         no      ;# wait for VMD to connect before running?
}


##############################################################
## EXTRA PARAMETERS                                         ##
##############################################################

# Put here any custom parameters that are specific to
# this job (e.g., SMD, TclForces, etc...)

SMD             on
SMDFile                 ../common/gp17eqinwater.ref
SMDk            1
SMDVel                  0.0002
```

```
SMDDir                  -0.5760420562995319 0.4403363319956876
-0.688679507535102
SMDOutputFreq       10


#############################################################
## EXECUTION SCRIPT                                        ##
#############################################################

# Minimization
if {1} {
minimize            1000
reinitvels          $temperature
}

run 35000 ;# 750ps
```

APPENDIX C : alldist.in Script to Measure Pairwise Atomic Distance

note: this script will measure the distances between proposed ion pairs

trajin simulation_output.mdcrd 1 100000 1

distance end_to_1a :10@NZ :527@OE1 out dist_19_536a.list

distance end_to_1b :10@NZ :527@OE2 out dist_19_536b.list

distance end_to_2a :10@NZ :528@OE1 out dist_19_537a.list

distance end_to_2b :10@NZ :528@OE2 out dist_19_537b.list

distance end_to_3a :14@NZ :527@OE1 out dist_23_536a.list

distance end_to_3b :14@NZ :527@OE2 out dist_23_536b.list

distance end_to_4a :14@NZ :528@OE1 out dist_23_537a.list

distance end_to_4b :14@NZ :528@OE2 out dist_23_537b.list

distance end_to_7a :294@OE1 :495@NZ out dist_303_504a.list

distance end_to_7b :294@OE2 :495@NZ out dist_303_504b.list

distance end_to_8a :296@NZ :496@OD1 out dist_305_505a.list

distance end_to_8b :296@NZ :496@OD2 out dist_305_505b.list

distance end_to_9a :300@OE1 :484@NZ out dist_309_493a.list

distance end_to_9b :300@OE2 :484@NZ out dist_309_493b.list

distance end_to_10a :300@OE1 :485@NH1 out dist_309_494a.list

distance end_to_10b :300@OE2 :485@NH2 out dist_309_494b.list

distance end_to_10c :300@OE1 :485@NH2 out dist_309_494c.list

distance end_to_10d :300@OE2 :485@NH1 out dist_309_494d.list