8-23-2013

Exploring the Solution Behavior of the Translational GTPase BipA Using Molecular Dynamics

Gilman L. Dionne
gilman.dionne@uconn.edu

Recommended Citation
Dionne, Gilman L., "Exploring the Solution Behavior of the Translational GTPase BipA Using Molecular Dynamics" (2013). Master's Theses. 496.
http://digitalcommons.uconn.edu/gs_theses/496

This work is brought to you for free and open access by the University of Connecticut Graduate School at DigitalCommons@UConn. It has been accepted for inclusion in Master's Theses by an authorized administrator of DigitalCommons@UConn. For more information, please contact digitalcommons@uconn.edu.
Exploring the Solution Behavior of the Translational GTPase BipA Using Molecular Dynamics

Gilman L. Dionne

B.S., University of Connecticut, 2012

A Thesis
Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science
At the
University of Connecticut
2013
Masters of Science Thesis

Exploring the Solution Behavior of the Translational GTPase BipA Using Molecular Dynamics

Presented by
Gilman L. Dionne, B.S.

Major Advisor_____________________________________________________

Dr. Victoria L. Robinson

Associate Advisor___________________________________________________

Dr. Eric R. May

Associate Advisor___________________________________________________

Dr. Daniel J. Gage

University of Connecticut
2013
Acknowledgements

I would like to thank Dr. Victoria Robinson for being a fantastic mentor and continually helping me grow and develop.

I would like to thank the members of the Robinson Lab, past and present, for making my time in the lab rewarding and enjoyable.

I would like to thank Dr. Eric May for all his help and guidance with Molecular Dynamics.

I would like to thank the University of Connecticut School of Engineering and the Booth Engineering Center for Advanced Technology for the use of the HORNET high-performance cluster.
Table of Contents

Approval Page ii
Acknowledgments iii
Table of Contents iv
Abstract v

Chapter 1: Introduction 1
1.1 Background on BipA (TypA) 1
   1.1.1 General Cellular Rolls 1
   1.1.2 Structural and Biochemical Overview of BipA 2
1.2 Purpose of Study 3

Chapter 2: Classical Molecular Dynamics 13
2.1 Introduction to Molecular Dynamics 13
2.2 Materials and Methods 15
   2.2.1 MD Simulations 15
   2.2.2 Multiple Sequence Alignment 16
2.3 Results 17
   2.3.1 Identification and Description of States 24
   2.3.2 Identification of the Mechanisms of Conformational Change 24
   2.3.3 ppGpp Recognition Mechanism 26
2.4 Conclusions 27

Appendices 51
A-1 Salmonella enterica serovar Typhumirium BipA R41A Mutagenesis 51

References 52
Abstract

BipA is a conserved prokaryotic GTPase necessary for securing bacterial survival and successful invasion of the host. Structural and biochemical studies indicate that GTP and ppGpp compete for binding to BipA to promote differential association of BipA to either the 70S or 30S ribosomal species. Exactly how guanine nucleotide binding to BipA prompts a change in the association of this protein with the ribosome is not understood. Crystallographic models show local structural rearrangements occur near the nucleotide-binding pocket but unexpectedly the overall domain arrangement, and therefore intramolecular contacts are similar in the various guanine nucleotide bound states. One explanation is that the lattice contacts in the crystal restrict the conformational space available to the protein. Another is that the BipA is metastable and binding of GTP or ppGpp trap intermediate states poised to bind the ribosome. This model would account for a reverse flow of information evidenced by the increase in BipA’s rate of GTP hydrolysis upon ribosome binding. We have monitored the solution dynamics of the various nucleotide bound states of BipA using amide hydrogen/deuterium exchange mass spectrometry (HDXMS). These data indicate that GTP and ppGpp binding lead to large scale conformational changes that are propagated throughout BipA, underscoring the idea that BipA is a metastable molecule where mutually exclusive association of GTP or ppGpp drive equilibria to alternate distinct conformations resulting in differential binding to the ribosome. Molecular dynamics simulations support this observation and point to dynamic allostery between the GTPase and novel C-terminal domain.
Chapter 1  Introduction

1.1  Background on BipA (TypA)

1.1.1  General Cellular Rolls

BipA is a member of the translation factor family of GTPases (trGTPases). (Leipe, Wolf, Koonin, & Aravind, 2002) It has been implicated as a global regulator of stress responses and virulence pathways in prokaryotes. (Scott, Diggle, & Clarke, 2003) While the absence of BipA has no apparent effects on bacteria grown under normal conditions, BipA is vital for growth under adverse conditions. In enteropathogenic Escherichia coli (EPEC), E. coli K-12, Sinorhizobium meliloti and Bacillus subtilis, BipA is required for growth at low temperatures (Beckering, Steil, Weber, Volker, & Marahiel, 2002; A. J. Grant, Haigh, Williams, & O'Connor, 2001; Kiss, Huguet, Poinso, & Batut, 2004; Pfennig & Flower, 2001) A S. meliloti bipA deletion strain showed increased sensitivity to environments of low pH and exposure to sodium lauryl sulfate. (Kiss et al., 2004) In EPEC and Salmonella enterica serovar Typimurium, BipA expression is upregulated in the presence of an antimicrobial protein produced by neutrophils, bactericidal/permeability increasing protein. (Barker, Kinsella, Jaspe, Friedrich, & O'Connor, 2000; Qi et al., 1995)

BipA also plays a role in the regulation of virulent phenotypes. An EPEC ΔbipA strain is hyper-motile. (Farris, Grant, Richardson, & O'Connor, 1998) In addition, BipA is required in EPEC for proper actin pedestal formation and microcolony formation on host epithelial cells. (A. J. Grant et al., 2003; Qi et al., 1995) In S. meliloti, BipA may be involved in infection and host maturation. (Kiss et al., 2004) Both EPEC and enterohemorrhagic E. coli require BipA for the expression of genes from espC
pathogenicity islands. (A. J. Grant et al., 2003) This same study showed that BipA null strains of EPEC and Pseudomonas aeruginosa are avirulent. (Farris et al., 1998; A. J. Grant et al., 2003; Neidig et al., 2013) BipA also provides resistance to S. enterica and EPEC to some host immune responses. (Barker et al., 2000; Qi et al., 1995)

1.1.2 Structural and Biochemical Overview of BipA

Similar to other trGTPases, BipA likely exerts its influence on cellular processes by binding to the ribosome. However, its ribosome binding properties differ drastically from other family members. When S. enterica is grown under normal growth conditions, BipA co-sediments with 70S ribosome, yet when S. enterica are grown under adverse growth conditions, such as nutrient limited growth, BipA binds to the 30S ribosomal subunit (Figure 1). (deLivron & Robinson, 2008) A ribosome-pelleting assay done with purified components proved that the GTP-bound form of BipA associates with the 70S ribosome, while the apo and GDP-bound form exhibited no ribosomal associations (Figure 2). (deLivron & Robinson, 2008) During stress events, the predominant guanine nucleotide species is neither GDP nor GTP, but rather ppGpp. (Cashel & Gallant, 1969; Potrykus & Cashel, 2008) Structurally, ppGpp resembles GDP with an additional pyrophosphate group esterified to the 3’ hydroxyl group. When ppGpp was incubated with purified components in a ribosome-pelleting assay, BipA associated with 30S ribosomal subunit (Figure 2). (deLivron & Robinson, 2008)

BipA is a ~67 kDa protein comprised of five domains (Figure 3). Four of these five domains are also present in close homologs EF-G and EF-4. (Leipe et al, 2002) Following the nomenclature established for the trGTPase family, BipA has a G domain at its N terminus, domain II is a β barrel, domains III and IV are α/β domains and the C
terminal domain (CTD) possesses a novel fold. (A.E. Every et al., unpublished) GTPases are called molecular switches based on their ability to interconvert between a guanosine diphosphate (GDP)- and guanosine triphosphate (GTP)-bound state. (Sprang, 1997) The surface properties of the GTPase domain are different in these two states thereby promoting interactions with various cellular partners. All GTPases share the same fold and basic mechanism of action but diverge greatly in their sequence size and shape. (Leipe et al, 2002) The GTPase core contains six-stranded β-barrel surrounded by five α helices. (Sprang, 1997) Conformational changes within this core are primarily localized to two regions of the domain termed the switch I and switch II (SWI and SWII). The GTPase cycle that involves binding, hydrolysis and release of guanine nucleotides, is unique for each GTPase.

The unique CTD is necessary but not sufficient for interaction with 70S ribosome. (deLivron et al., 2009) Two features of the CTD that are of interest are a highly flexible distal loop, and a C-terminal helix (Figure 3). This ultimate helix is necessary for interaction with 70S ribosome, and insertion of the mini-mariner transposon in this region affects SDS-resistance and the colonization of mouse large intestine in E. coli MG1655. (deLivron et al., 2009; Moller et al., 2003) The G domain and domain II are not necessary for 70S ribosome binding, but are required for interaction with the 30S ribosomal subunit. (deLivron et al., 2009) It is suspected that domain III, domain V, and the CTD form a binding surface for the 70S ribosome. (deLivron et al., 2009)

1.2 Purpose of Study

Despite BipA’s documented involvement in the regulation of a variety of stress responses and virulence pathways, a defined cellular role for this protein has remained
elusive. An understanding of how BipA affects its influence on these events is important for the validation of BipA as a potential target for anti-microbial therapeutics. As stated previously, the interaction between BipA and the ribosome is pivotal to its function in the cell. (deLivron & Robinson, 2008; Owens et al., 2004) Formation of these distinct species, however, is complex. One study suggests that the interaction of BipA:GTP with the 70S ribosomal species may involve mRNA and tRNA. (Owens et al., 2004) Virtually nothing is known about the 30S:BipA:ppGpp complex. In order to begin to understand how BipA can differentiate between the 70S ribosome and the 30S ribosomal subunit, Every et al. (unpublished) carried out crystallographic studies of apo-BipA as well as BipA in the presence of GDP, GMPPNP (a non-hydrolysable GTP analogue) and ppGpp. Models have been completed for apo-BipA and BipA:GDP. The models for BipA:GMPPNP and BipA:ppGpp are still undergoing refinement. A preliminary comparison of all four BipA structures suggests these models are remarkably alike, differing only by local rearrangements near the guanine nucleotide-binding site. The similarity between the models may be due to artificial restraints being put on the protein by lattice contacts.

Because little explanation for the differential ribosome binding by BipA could be extrapolated from the crystallographic models, isothermal titration calorimetry (ITC) was used to characterize the thermodynamics of BipA binding to various guanine nucleotides. (Every et al., unpublished) The affinities measured were 22 µM, 29 µM, and 11 µM for GDP, GMPPNP and ppGpp, respectively. The comparable affinities for GTP and ppGpp indicate association of these compounds is driven by the relative amounts of GTP to ppGpp in the cell. The ΔC_p values for GTP and ppGpp are -22 cal·mol⁻¹·K⁻¹ and -89
The small magnitude of these values suggest that very subtle structural rearrangements occur upon guanine nucleotide binding, which would agree with the crystallographic data. For comparison, EF-G, one of BipA’s closest homologues, has a $\Delta C_p$ value of -270 cal·mol$^{-1}$·K$^{-1}$: the result of a rather pronounced conformational change that exposes ~600 Å$^2$ of nonpolar surface area. (Hauryliuk et al., 2008)

In collaboration with Dr. Ganesh Anand’s group and the National University of Singapore, hydrogen/deuterium exchange mass spectrometry (HDXMS) experiments were used to further probe the solution behavior of BipA in the presence of various nucleotides. (Hoofnagle, Resing, & Ahn, 2003; Wales & Engen, 2006) The GTP bound form of BipA had a large amount of amide hydrogen deprotection compared to apo-BipA (Figure 4a). This deprotection could indicate increased flexibility of BipA:GTP or changes in solvent accessible surface area. (Hoofnagle et al., 2003; Wales & Engen, 2006), In contrast, the exchange patterns for BipA:ppGpp show large regions of amide protection when compare to apo-BipA (Figure 4b). These data demonstrate that the GTP and ppGpp bound forms of BipA have different solution behaviors. The crystallographic and ITC data suggests that BipA exists in a relatively flat energy landscape. It is likely that the binding of guanine nucleotides is governed by conformational selection. That is, Apo-BipA exists as an ensemble of similar but distinct conformations in solution and the binding of the ligand shifts the equilibrium toward a subset of these various conformations.

This work seeks to address some of the issues that are currently unresolved; what defines the solution states of the various nucleotide bound forms of BipA, how does GTP and ppGpp influence the solution behavior of BipA, and how does BipA recognize
ppGpp. Classical molecular dynamics (MD) were used to model BipA and its various nucleotide bound states. There are three facets of this work that will have a great impact on understanding the mechanism of action of BipA and cellular stress events in general.

As stated previously, the cellular roles of the 70S:BipA and 30S:BipA complexes have yet to be delineated. Two biologically active small molecules, GTP and ppGpp dictate these ribosomal binding events. Our research would be greatly facilitated if we could significantly stabilize either the 70S:BipA or 30S:BipA ribosomal complex. This would enhance not only our ongoing cryoEM, crystallographic and biochemical endeavors but also provide new tools to decipher the in vivo function of the protein. Right now, we are dependent upon stressing the bacteria almost to the point of death, in order to carry out our analysis. MD simulations can provide theoretical structural models of BipA:GTP and BipA:ppGpp. These models can then be used for the development of mutants and/or small molecules that are specific for either the 30S or 70S ribosome interaction.

In addition, thermodynamic data from the ITC studies suggests that GTP and ppGpp promote limited conformational changes. However, HDXMS experiments reveal that BipA:GTP and BipA:ppGpp have vastly different solution behaviors. Therefore, comparison of the MD simulations may provide insights into the solution behavior of BipA, and reveal how GTP and ppGpp influence the solution dynamics of BipA.

And finally, a model describing the BipA:ppGpp complex is of interest. To date, ppGpp has only be crystallized with two other GTPases, Obg and Release Factor 3 (RF3). (Buglino et al., 2002; Kihira et al., 2012) Obg has no interactions between the protein and the 3′ pyrophosphate of ppGpp (Figure 5a), while RF3 has a lysine that interacts with
the 3′ pyrophosphate (Figure 5b). (Laskowski et al., 1997; Wallace, Laskowski, & Thornton, 1995) However, this lysine has low conservation: only ~60% of Obg homologues contain a lysine at the same location. (Kihira et al., 2012) Thus, there is a limited amount of information for how ppGpp can be recognized by GTPases. The BipA:ppGpp model derived from MD may identify potential ppGpp recognition elements or motifs which would allow us to then identify other GTPases that interact with ppGpp.
Figure 1: BipA adopts an alternate ribosome binding mode under adverse growth conditions. *S. enterica* SB300A cells containing the plasmid encoding His-tagged BipA were grown in minimal medium to mid-log phase and induced with 0.2% arabinose. After 90 min, SHX was added to the cells to 0.1 mM to trigger the stringent response. The cells were grown for an additional 30 min and then harvested. Lysates were clarified and sedimented through 7- to 47% sucrose gradients. The resulting ribosomal UV profile measured at 254 nm (solid line, with SHX; dashed line, without SHX) is shown. Gradient fractions were TCA precipitated and analyzed by SDS-PAGE and Western blotting against His-tagged BipA (-SHX). A parallel experiment without SHX (-SHX) demonstrated that under normal growth conditions BipA associates with 70S ribosome. Adapted from (deLivron & Robinson, 2008)
Figure 2: The bacterial alarmone, ppGpp, promotes 30S ribosome binding. Purified His-tagged BipA and ribosomes were incubated with 20-fold molar excess of GDP, GMPPNP or ppGpp. Mixtures were placed on top of a 1.1 M sucrose cushion and centrifuged. Fractions corresponding to the 30S, 50S, and 70S as well as fractions from the top and bottom of the gradient were analyzed via western blot with a HisDetector Western blot kit. Apo-BipA and BipA:GDP remained at the top of the gradient. BipA:GMPPNP cosedimented with the 70S Ribosome. BipA:ppGpp cosedimented with the 30S ribosome. Adapted from (deLivron & Robinson, 2008)
FIGURE 3: BipA has five domains. Ribbon diagram depicting the crystal structure of apo-BipA solved to 2.8 Å resolution (Every and Robinson, unpublished). The G domain is highlighted in red, with the nucleotide-binding site highlighted in grey. Domain II (blue) is a β-barrel. Domain III (green) and Domain V (magenta) are α/β domains. The C-terminal Domain (cyan) contains a novel fold and is functionally important for BipA’s association to the 70S ribosome.
Figure 4: Hydrogen/deuterium exchange rates in BipA:GTP and BipA:ppGpp. In collaboration with Dr. Ganesh Anand’s group at the National University of Singapore, mass spectrometry was used to measure rates of hydrogen/deuterium exchange of the backbone amide hydrogen. Samples of apo-BipA, BipA:GTP, and BipA:ppGpp were exposed to a deuterated solvent (pD 7.5) for 10 minutes. The reaction was quenched, and the protein digested by pepsin. Peptide fragments were separated by liquid chromatography and analyzed by ESI-MS. Analyzed fragments provide 90% sequence coverage. Regions with no coverage are represented in grey. Peptides with an increased exchange rate, amide deprotection, (compared to apo-BipA) are represented in red, while peptides with decreased exchange rates, amide protection, are in blue. No change is depicted in green. (a) BipA:GTP experiences a significant amount of deprotection or increased hydrogen exchange across all domains except for the CTD. (b) In contrast, BipA:ppGpp exhibits protection in all five domains. Some increase is observed at the start of a loop region in the G domain, parts of domain II, parts of domain III and the CTD.
Figure 5: Examples of ppGpp bound to GTPases. LIGPLOT diagrams of (a) Obg from *Bacillus subtilis* and (b) RF3 from *Desulfovibrio vulgaris*. Obg has no interaction with the 3′ pyrophosphate group on ppGpp. A lysine was observed interacting with the 3′ pyrophosphate group in RF3. (Buglino, Shen, Hakimian, & Lima, 2002; Kihira et al., 2012)
Chapter 2 Molecular Dynamics Simulations and Analysis

2.1 Introduction to Molecular Dynamics

Classical molecular dynamics (MD) calculates the energy, motions and interacting forces of atoms in a defined system. MD simulations use a differentiable equation to represent the potential energy of an atom. From this we can extract the forces on that atom and subsequently determine the velocities and change in positions over a fempto-scale time-step (Guvench & MacKerell Jr, 2008) There are multiple independently developed representations (?) of this equation (referred to force fields hereafter), each one taking different approaches to define the energies in the system. While different force fields contain similar terms that describe bonded and non-bonded interactions, they differ in how these terms are parameterized and in the usage of additional terms that attempt to rectify discrepancies that arise between MD data and quantum mechanical calculations or experimental data. (MacKerell et al., 1998) These force fields have successfully been used in simulations to fold (and refold) elongated peptides into their experimentally determined structures. (Bowman, Voelz, & Pande, 2011; Lindorff-Larsen, Piana, Dror, & Shaw, 2011; Seibert, Patriksson, Hess, & van der Spoel, 2005; van der Spoel & Seibert, 2006)

The CHARMM force field, which as been constantly developed and improved since the 1980’s, was utilized for all MD simulations in this study. (B. R. Brooks et al., 2009; Bernard R Brooks, Brucoleri, Olafson, Swaminathan, & Karplus, 1983; MacKerell et al., 1998) The potential energy function for charm is as follows:
While this equation has elements that are present in other force fields, CHARMM has terms to address Urey-Bradley interactions, planarity of improper dihedrals, and inaccuracies in the protein backbone energetics (CMAP). (B. R. Brooks et al., 2009; MacKerell et al., 1998) The energy constants \( (K_b, K_\theta, K_\phi, K_\omega, K_{UB}) \) have been parameterized for proteins by fitting the equations to gas-phase vibrational spectra of organic molecules of comparable structures. (MacKerell et al., 1998) The developers of CHARMM have also developed a water model in order to improve protein-solvent interactions in a computationally efficient manner. (MacKerell et al., 1998) The CHARMM force field has recently been used fold 11 proteins, ranging in size from 10 to 80 residues, to their experimentally determined structure. (Lindorff-Larsen et al., 2011)

In brief, a classical MD simulation starts with a high-resolution structural model. This model is then analyzed in order to generate a topology file that lists all the connections in the model and contains all relevant energy constants for the force field. The model is then centered in a box that defines the boundary of the system. Periodic boundary conditions (pbc) allow the system to simulate an infinite solvent and maintain a constant mass. Particles that cross the boundary on one side of the box are replaced by an equivalent particle on the opposite side. In order to prevent the protein interacting with itself across the boundary, the box must be sufficiently larger than protein. The box is then solvated with water models. Water typically comprises the majority of the system.
Ions are also added to the system to mimic a particular buffer condition as well as neutralize the net charge within the box.

Before beginning the simulation, an energy minimization allows the system to resolve any energetically unfavorable conditions, e.g. steric clashing or unrealistic bond measurements that would otherwise produce unnatural artifacts or cause the simulation to crash. Typically, after minimization a short (100’s of ps) equilibration is carried out with the protein restrained. This allows the water models to relax and solvate the protein.

The simulations in this study had initial velocities for atoms randomly generated according to a Boltzmann distribution for atom velocities at 300 K. The potential energy equation (force field) for each atom is then integrated to find positions and velocities after each time-step.

2.2 Materials and Methods

2.2.1 MD Simulations

Atomic coordinates for apo-BipA were obtained from the crystallographic structures discussed previously. Missing loop regions of the protein were modeled in using the MMTSB toolset loopmodel script. (Feig, Karanicolas, & Brooks, 2004; Sali & Blundell, 1993) The GROMACS 4.5 software package was used for preparation of the system, running of the simulation and analysis of simulation data. (Pronk et al., 2013) GDP and GTP were parameterized using values from existing topologies of ADP, ATP, and guanine. ppGpp was parameterized using the same values as well as by comparison to a parameterization by SwissParam. (Foloppe & MacKerell Jr, 2000; MacKerell & Banavali, 2000; Pavelites, Gao, Bash, & Mackerell, 1997; Zoete, Cuendet, Grosdidier, & Michielin, 2011) Hydrogen atoms were added on to the atomic coordinates of BipA. A
topology was generated for the CHARMM force field. A cubic box was defined with a length 20 Å larger than the largest length of BipA. BipA was placed in the center of this box. The box was then solvated by a 3-atom water model (the CHARMM water topology was used). Sodium chloride (to an effective concentration of 200 mM), magnesium chloride (to an effective concentration of 5 mM) and excess sodium were added to remove any net charge in the box. The system then went through two stages of energy minimization, first using a steepest decent algorithm, then using a conjugate gradient. The solvent was then allowed to relax for 100 ps while the protein is restrained. Simulations were run on 192 Intel Xeon X5650 Westmere cores on the University of Connecticut School of Engineering and Booth Engineering Center for Advanced Technologies’ Hornet HPC Cluster. Simulations were run for a minimum of 100 ns, with extensions given subjectively to trajectories that were still exploring different conformations (based on RMSD).

2.2.2 Multiple Sequence Alignment (MSA)

A multiple sequence alignment was constructed in order to identify residues that have high conservation. A PSI-BLAST search was carried out on the sequence of BipA from S. enterica. (Altschul, Gish, Miller, Myers, & Lipman, 1990; Altschul et al., 1997) The top 10,000 hits were then screened to remove any sequence that was not annotated as BipA, was a putative or theoretical protein, or if they were too short or too long in length. The resulting list of ~ 1400 sequences was then aligned in JalView using MUSCLE for large alignments. (Edgar, 2004a, 2004b; Waterhouse, Procter, Martin, Clamp, & Barton, 2009)
2.3 Results

2.3.1 Identification and Description of States

The trajectories of apo-BipA, BipA:GTP and BipA:ppGpp were analyzed in order to discern the conformation(s) that each of these different forms adopt. Analysis of root mean squared deviation (RMSD) of the trajectories after least squares fitting (LSQ fitting) to an averaged structure, obtained from all trajectories, can provide an idea of stability and conformational variability. (Börjesson & Hünenberger, 2004; Brigo, Lee, Iurcu Mustata, & Briggs, 2005) The apo-BipA trajectory(Figure 6), which spans 140 ns, varies in RMSD (fitted to a computed average structure for the trajectory) from ~1.7 Å to ~4.2 Å, with an average RMSD of 2.7 Å and a standard deviation of 0.4 Å. The average RMSD value indicates that BipA is not experiencing any large structural rearrangements, but local regions of apo-BipA are deviating in position. So then this does really match/support the ITC data and visa versa.

The trajectory of the BipA:GTP simulation, which is 100 ns in duration is more stable than the apo-BipA trajectory (Figure 7). The RMSD (fitted to a computed average structure) of the entire trajectory ranges from ~1.2 Å to ~4.9 Å. The average RMSD and standard deviation for the BipA:GTP trajectory are 1.8 Å and 0.4 Å, respectively. However, in first few nanoseconds of the trajectory, the RMSD drops dramatically and stabilizes. When the RMSD values from the first 4 ns are removed, the range narrows to ~1.2 Å to ~2.8 Å. This truncation also changes the average RMSD to 1.7 Å and the standard deviation to 0.2 Å. Thus, BipA:GTP undergoes an initial equilibration period and then settles into a relatively stable conformation.
The BipA:ppGpp trajectory is 150 ns in duration (Figure 8a). It ranges in RMSD from \(-1.5\) Å to \(-5.1\) Å, with an average RMSD of 2.6 Å and a standard deviation of 0.5 Å. However, a visual inspection of the trajectory shows that BipA:ppGpp adopted a stable conformation at approximately 75 ns. To investigate this further, the trajectory was truncated to the last 75 ns and analyzed. The RMSD values (fitted to a computed average structure) from the truncated trajectory range from \(-1.2\) Å to \(-2.9\) Å (Figure 8b). This is comparable to the range measured for BipA:GTP. The average RMSD for the truncated BipA:ppGpp trajectory is 1.7 Å with a standard deviation of 0.3 Å. These values indicate that during the first 75 ns of the BipA:ppGpp trajectory, BipA is undergoing a transition but then assumes a stable conformation and remains in that conformation for the remainder of the simulation.

By analyzing the frequency distribution of RMSD values between every structure in the trajectory, it is possible to extract information about the number of conformational states explored by the protein during the simulation (Figure 9). (Fenwick, Esteban-Martin, & Salvatella, 2011; Henriksen, Roe, & Cheatham, 2013) The RMSD distribution for apo-BipA exhibits two clear peaks, with maxima at \(-3.2\) Å and \(-4.5\) Å, with a possible third peak that’s partially masked at \(-2.5\) Å. The presence of three peaks suggests that the apo form of BipA may sample two to three similar, yet distinct conformations in solution. The RMS distribution of the GTP bound form of BipA has one major peak at 2.7 Å. The second, minor peak is most likely a result of the transition period in the first 4 nanoseconds of the trajectory. This singular peak, at a relatively low RMSD, indicates that BipA:GTP exists in solution as a single stable state. The full trajectory for BipA:ppGpp has four peaks, with maxima at \(-2.1\) Å, \(-2.7\) Å, \(-3.7\) Å and \(-4.8\) Å.
suggesting that BipA:ppGpp may have several different solution structures. However, the truncated trajectory for BipA:ppGpp has only two peaks, with maxima at ~2.1 Å and ~2.7 Å. Their low RMSD values indicate that they have similar in conformations.

Root mean square fluctuation (RMSF) values can be calculated to get a sense of the flexibility of the different regions of the protein. (Brigo et al., 2005; Franco-Gonzalez, Cruz, Ramos, & Martinez-Salazar, 2013; McGowan & Hamelberg, 2013) In the apo-BipA trajectory, two regions of the protein corresponding to the SWI region in the G domain and the distal loop in the CTD are revealed from this analysis (Figure 10a). These residues are expected to be highly mobile as they are not resolved in the crystal density. In contrast, the BipA:GTP trajectory, indicates that not unexpectedly, the SWI region in the G domain has become ordered, and there is decreased flexibility in the distal loop (Figure 10b). Comparison of crystal structures of apo and GTP-bound Ras-like GTPases show an ordering of both the SWI and SWII regions when guanine nucleotide is present (Vetter & Wittinghofer, 2001). The differences in the RMSF profiles for apo and GTP-bound BipA shown in figure 11, demonstrate the extent to which these residues have stabilized, with the decreases as large as 10 Å in the G domain loop.

The RMSF of the full trajectory of BipA:ppGpp is similar to that of apo-BipA (Figure 10c). The RMSF plot of the truncated trajectory for BipA:ppGpp, however, has large reductions in the loop regions in both the G domain and the CTD (Figure 10d). The RMSF of the loop region in the G domain decreases as much as 12 Å, the distal loop in the CTD decreases by about 5 Å and the C-terminal helix experiences an increase of RMSF of ~2 Å (Figure 12). As with BipA:GTP, the ordering of residues in the G domain coincides with the nucleotide binding. Surprisingly, our analysis indicates that the distal
loop in the CTD experiences a greater decrease in flexibility than it does in the BipA:GTP trajectory. Perhaps it is being locked into a particular conformation that allows for 30S ribosomal subunit association. As stated before, a truncated BipA protein with domains III, V and the CTD can bind to the 70S ribosome. (deLivron et al., 2009) Therefore, communication between the G domain and the CTD may not be essential for 70S ribosome binding. However, 30S ribosomal subunit association, however, is abolished by removal of the G domain. Thus communication between the G domain and the CTD is important for 30S ribosomal subunit association, and the reduction in flexibility of the CTD loop may be the result of this inter-domain communication.

Solvent accessible surface area (SASA) measurements can be used to assess changes in the surface features of the protein and provide a limited correlation to the HDXMS data. (Chik & Schriemer, 2003; Eisenhaber, Lijnzaad, Argos, Sander, & Scharf, 1995; Shan, Arkhipov, Kim, Pan, & Shaw, 2013; Yan, Watson, Ho, & Deinzer, 2004) While some parallels exist between parameters measured from MD simulations and the rate of amide exchange in HDXMS, these methods measure events on vastly different time scales (ns-µs for MD, seconds-hours for HDXMS), which make direct correlations difficult.

The change in SASA (ΔSASA) between the apo-BipA trajectory and the BipA:GTP or BipA:ppGpp (truncated) trajectories were calculated for hydrophobic atoms and hydrophilic atoms of the entire protein, as well as each of the five domains. In addition, absolute (the sum of the absolute values of ΔSASA_{hydrophilic} and ΔSASA_{hydrophobic}) and total changes in SASA were also calculated (Table 1). Globally, the BipA:ppGpp trajectory has an total decrease of 245 Å² in SASA, compared to a total
decrease of 38 Å² in the BipA:GTP trajectory. However, the absolute changes in average SASA for the BipA:GTP and BipA:ppGpp trajectories are remarkably similar, with values of 812 Å² and 879 Å², respectively. The absolute changes in average SASA for each domain show where the largest changes in surface area features are occurring. For the BipA:GTP trajectory, the G domain, domain III and the CTD have the largest absolute changes in SASA features. The G-domain has an overall increase in surface area, which agrees with the deprotection observed in the HDXMS data. The CTD experiences an overall decrease in surface area. Domain III does not change much in terms of net surface area, but rather presents a slightly more hydrophobic surface. The BipA:ppGpp trajectory, shows the largest absolute changes in average SASA for the G domain, domain III and domain V. Contrasting the behavior observed in the BipA:GTP trajectory, the G domain of the BipA:ppGpp trajectory exhibits a decrease in SASA. This decrease may explain the protection that is observed in G domain in the HDXMS experiments. As in the BipA:GTP trajectory, domain III changes its surface to become more hydrophobic in character. Domain V changes to become a bit more hydrophilic.

Principle component analysis (PCA) can extract and quantify correlated motions of atoms observed in a trajectory of a protein. (Brigo et al., 2005; Franco-Gonzalez et al., 2013; McGowan & Hamelberg, 2013; Munz, Hein, & Biggin, 2012; van Aalten et al., 1995) These motions are expressed as a set of eigenvectors and eigenvalues, which explain the significant correlated motions, principle components (PCs), of the protein through multidimensional space. When used in conjuncture with clustering analysis, two-dimensional plots of the first and second PCs can be used to extract structural information about the various conformations a protein samples in a trajectory. (Munz et al., 2012)
PCA was carried out by concatenating the trajectories for apo-BipA, BipA:GTP and BipA:ppGpp. This trajectory was used to construct a covariance matrix. Eigenvectors (PCs) and corresponding eigenvalues can be calculated by diagonalizing the covariance matrix. The first two PCs effectively describe ~43% and ~16%, respectively, of the correlated motions observed in the three trajectories.

A projection of the Apo-BipA trajectory on to the 1st and 2nd eigenvectors produces a plot with a relatively dispersed set of points (Figure 13a). The RMS distribution for apo-BipA indicated three possible clusters in the apo-BipA trajectory. K-means clustering analysis, using the “clara” function in the CRAN clustering packages, was carried out to find three clusters in this plot. (Kaufman & Rousseeuw; Maechler, 2005; Team, 2008) The clustering analysis assigned different points to one of three clusters (Figure 13b). The structures assigned to a particular cluster were extracted from the trajectory and compared to one another in order to find a centroid structure, which is the structure that has the lowest average RMSD after LSQ fitting to the entire set of structures.

The projection of the BipA:GTP trajectory onto the same eigenvectors produces a plot with one main cluster (Figure 14). The points outside this grouping correspond to the first few nanoseconds of the trajectory. Since there is only one cluster for BipA:GTP, all structures were used to find a centroid structure.

The full trajectory for BipA:ppGpp as well as the truncated trajectory were analyzed in the same manner (Figure 15a and 15b). The plot of the full BipA:ppGpp trajectory is dispersed, with approximately 3 clusters. However, one of these groups
corresponds to the points of the truncated BipA:ppGpp trajectory. Therefore, all the structures from the truncated trajectory were used to calculate a centroid structure.

The centroid structures for BipA:GTP and BipA:ppGpp were used as representative structures for each of these complexes. A comparison of these models revealed some of the larger differences between these states. In the BipA:ppGpp structure, the loop in the G domain is closer to domain III and the CTD moves towards the G domain. To quantify this, the difference in the distance between the centers of mass (COM) of each domain was calculated (Table 2). The interdomain distance between the G domain and both domain III and the CTD are shorter in the BipA:ppGpp structure. The distance between domain V and the CTD is also shorter in the BipA:ppGpp structure. This shift in domain position means that BipA:ppGpp is slightly more compact than BipA:GTP, with average radii of gyrations of 27.8 Å and 28.8 Å, respectively.

A list of salt-bridges for the centroid structures of the various forms of BipA was calculated using VMD. (Humphrey, Dalke, & Schulten, 1996) Only interactions involving conserved residues were considered and have been listed in Table 3. In particular, there are significant differences in salt bridge interactions between the G domain and domain III. In the BipA:GTP model, Lys50 and Glu51 extend from the G domain BipA:GTP and make contacts with Glu 317 and Arg 375 in domain III (Figure 17a). In BipA:ppGpp, the interaction between these domains occurs between Asp 44, Asp 47, and Arg 52 on the G domain and Arg 375 and Glu 376 on domain III (Figure 17d). These altered inter-domain contacts may represent separate communication pathways between the SWI of the G domain and the rest of BipA. The nucleotide species present in
the binding pocket grounds the SWI in a particular conformation, dictating which inter-domain contacts are made.

2.3.2 Identification of the Mechanisms of Conformational Change

To understand how BipA transitions from an apo to a guanine nucleotide bound conformation, the trajectories of BipA:GTP and BipA:ppGpp were compared to the trajectory of apo-BipA. Matrices were constructed by calculating RMSD values after LSQ fitting for every combination of structures between two trajectories. Figure 18 is a visual representation of the RMSD matrix between apo-BipA and BipA:GTP. In this matrix, there is a region of low RMSD which reveals that apo-BipA samples conformations that are also accessible to BipA:GTP. This does not seem to be the case in the RMSD matrix between apo-BipA and BipA:ppGpp (Figure 19). In fact, along the BipA:ppGpp axis, there seems to be a trend of incrementally increasing RMSD. At approximately 75 ns on the BipA:ppGpp axis, there is an increase in RMSD, as BipA adopts the stable ppGpp conformation discussed earlier. We can conclude then that unlike BipA:GTP, structures in the BipA:ppGpp trajectory have lower similarity to structures in the apo-BipA trajectory.

PCA can be used to examine the shared, or unshared, conformational space exhibited by each form of BipA. It has been used previously to examine at the conformational spaces explored by GTPases. (Bucher, Grant, & McCammon, 2011; Gorfe, Grant, & McCammon, 2008; B. J. Grant, Gorfe, & McCammon, 2009; B. J. Grant, McCammon, & Gorfe, 2010; Long & Bruschweiler, 2011; van Aalten et al., 1995) Presented in Figure 22 are the projections of all the trajectories (apo-BipA, BipA:GTP, BipA:ppGpp, and the truncated BipA:ppGpp) onto the 1st PC against projections of all
the trajectories onto the 2\textsuperscript{nd} PC (Figure 20a), 3\textsuperscript{rd} PC (Figure 20b), 4\textsuperscript{th} PC (Figure 20c), and 5\textsuperscript{th} PC (Figure 20d). These 5 PCs represent \(\sim78\%\) of the motion in observed in the trajectories. In the plot of PC1 and PC2, there is a small area of overlap between the apo-BipA projection and the BipA:GTP projection. This region of overlap between the two trajectories is present in the plots with PCs 3, 4 and 5 and so represent the conformational space shared between apo-BipA and BipA:GTP. This suggests that BipA interacts with GTP by conformational selection. However, between the apo-BipA and BipA:ppGpp projections, there is little to no overlap and where it does occur, it happens with points early on in the BipA:ppGpp trajectory. Later on, the points move away from PC space visited by apo-BipA until they reach the region occupied by the truncated BipA:ppGpp trajectory. This suggests that BipA interacts with ppGpp to some degree by conformational selection, but then ppGpp induces a conformational change.

To examine guanine nucleotide binding, principle component analysis was carried out on residues within 10 Å of GTP in the BipA:GTP centroid structure. (McGowan & Hamelberg, 2013) Trajectories of this subset of residues were extracted from the trajectories of apo-BipA, BipA:GTP, and BipA:ppGpp. Eigenvectors and corresponding eigenvalues were calculated in the same manner as described above. The trajectories of the subset of residues for Apo-BipA, BipA:GTP, BipA:ppGpp and the truncated BipA:ppGpp were projected onto the first 5 eigenvectors (which explain \(\sim83\%\) of the motions in the system) The projections of the 1\textsuperscript{st} eigenvector were plotted against the projections of the 2\textsuperscript{nd} (Figure 21a), 3\textsuperscript{rd} (Figure 21b), 4\textsuperscript{th} (Figure 21c), and 5\textsuperscript{th} (Figure 21d) eigenvectors. The projections of this subset of residues tend to mimic the behaviors observed in projection of the whole protein (Figure 20).
The apo-BipA and BipA:ppGpp projections are dispersed. In contrast, those for BipA:GTP and the truncated BipA:ppGpp are relatively condensed. As with the full length protein trajectories, the BipA:GTP trajectory shares some conformational space with apo-BipA trajectory. This would indicate that the nucleotide-binding pocket of apo-BipA is flexible and sampling different conformations, and the binding of GTP stabilizes a particular state of BipA. The slight differences in conformational space occupied by apo-BipA and BipA:GTP is probably the result of some minor induced rearrangements. The projections of apo-BipA and BipA:ppGpp, however, only sample similar conformation space early on in the BipA:ppGpp trajectory. As the BipA:ppGpp trajectory continues, it moves further away the conformational space sampled by apo-BipA towards the space that is occupied by the truncated BipA:ppGpp trajectory. This interaction is best described by an initial stage of conformational selection followed by an induced fit mechanism toward a stable BipA:ppGpp conformation.

2.3.3 ppGpp Recognition Mechanism

Analysis of the BipA:ppGpp trajectory were done to see if any information about how BipA recognize ppGpp could be obtained. Approximately 75 ns into the trajectory, a large loop in the G domain experiences a dramatic decrease in flexibility. One residue, Arg 41, makes contact with the 3’ pyrophosphate group (Figure 22). According to our MSA, this arginine is conserved in >83% of BipA homologues. This interaction is probably driven by electrostatics, as the addition of the 3’ pyrophosphate group adds an additional charge of -3e, compared to GDP with a formal charge of -3e. The positive charge of the arginine would help to stabilize this excessive negative charge.
Arginine has also been observed interacting with the 3’ pyrophosphate group of ppGpp in other protein structures. (Artsimovitch et al., 2004; Mechold, Potrykus, Murphy, Murakami, & Cashel, 2013; Zuo, Wang, & Steitz, 2013) RNA polymerase from *Thermus thermophilus*, *E. coli* DH1, and *E. coli* K-12 have been crystalized in the presence of ppGpp; the structural mechanisms of ppGpp recognition in these structures all involve an arginine (Figure 23a,b,c). (Artsimovitch et al., 2004; Laskowski et al., 1997; Mechold et al., 2013; Wallace et al., 1995; Zuo et al., 2013) Arginine also plays a role in the recognition of ppGpp in the structure of guanosine pentaphosphate phosphohydrolase in *Aquifex aeolicus* (Figure 23d). (Kristensen, Ross, & Gajhede, 2008; Laskowski et al., 1997; Wallace et al., 1995) The stabilization of the 3’ pyrophosphate by arginine appears to be a general feature of proteins which productively bind ppGpp.

2.4 Conclusions

Analysis of the trajectories of BipA:GTP and BipA:ppGpp have allowed for the extraction of centroid structures representative of the solution conformations of BipA:GTP and BipA:ppGpp. SASA measurements show these two states of BipA have different surface characteristics. Differences in relative domain positions give BipA:ppGpp a slightly more compact form compared to BipA:GTP. The two structures also form differing salt-bridges that allow for modified inter-domain communications.

RMS distribution and principle component analysis of the apo-BipA trajectory show that apo-BipA exists in a few metastable solution states. RMSD calculations and RMSF profiles reveal these states do not reflect large global rearrangements, but rather are due to local rearrangements of SWI in the G domain and the flexible loop in the CTD.
The two current models used to describe ligand-dependent conformational change are the induced fit model and the conformational change model. Induced fit describes a system in which the protein exists in an unbound conformation in solution; ligand entering the binding pocket then prompts the protein to adopt a new conformation: the bound conformation (Figure 24). (Boehr, Nussinov, & Wright, 2009; Csermely, Palotai, & Nussinov, 2010) Conformational selection, however, starts with an ensemble of various protein conformations in solution. The ligand is able to bind to a particular conformation in the ensemble. Binding of the ligand shifts the equilibrium from the multiple solution conformations to the single ligand bound conformation (Figure 24). (Boehr et al., 2009; Csermely et al., 2010)

RMSD matrices and PCA suggest at some point, apo-BipA and BipA:GTP sample very similar global conformations. PCA show significant overlap in the structures of the nucleotide-binding pocket between the apo and GTP bound forms. Apo-BipA exists in solution as a set of metastable conformations, of which one conformations favors GTP association. While GTP does induce some subtle changes in the binding pocket, globally the structure is very similar to conformations observed in the apo-BipA trajectory. Thus GTP binding alters the energy landscape of apo-BipA and shifts the equilibrium to the observed BipA:GTP conformation. This BipA:GTP conformation is then able to interact with the 70S ribosome. Therefore, the transition from apo-BipA to BipA:GTP is an example of conformational selection.

The story is slightly different for BipA:ppGpp and apo-BipA which adopt dissimilar conformations, as evident by the RMSD matrices and PCA. PCA analysis of the nucleotide-binding pocket of apo-BipA indicates that it occupies the same
conformational space that BipA:ppGpp does early in its trajectory but then migrates away from an apo-BipA like conformation. This suggests that ppGpp modulates the structure and dynamics of BipA through a two-stage mechanism, conformational selection followed by induced fit. That is, ppGpp binds to a particular state that apo-BipA samples in solution. Once bound, ppGpp reconfigures the energy landscape, stabilizing a particular conformation of the nucleotide-binding pocket. The SWI region of the G domain then undergoes a rearrangement that causes BipA to transitions to a new area of conformational space, where it can adopt the BipA:ppGpp conformation represented by the centroid structure. RMSF profiles indicate that the highly flexible loop regions of apo-BipA have now adopted new stable conformations. This BipA:ppGpp conformation has the ability to bind to the 30S ribosome.

The formation of the BipA:ppGpp conformation seems to be dependent on Arg 41. This residue contacts with the 3’ pyrophosphate group of ppGpp. The interaction between Arg 41 and ppGpp appears to stabilize SWI of the G domain, allowing it to make contacts on domain III. Arg 41 is therefore proposed to be BipA’s ppGpp recognition mechanism.
Figure 6: The trajectory for Apo-BipA samples a variety of conformational space. A plot of RMSD values calculated for the apo-BipA trajectory after LSQ fitting to a calculated average structure. The plot was smoothed with a rolling average of 3000 points. On average, the RMSD values are high, indicating apo-BipA samples numerous conformations in the trajectory.
Figure 7: The trajectory of BipA:GTP is relatively stable. A plot of RMSD values calculated for the BipA:GTP trajectory after LSQ fitting to a calculated average structure. The plot was smoothed with a rolling average of 3000 points. The low RMSD values indicate that BipA:GTP is in the same state for the majority of the simulation.
Figure 8: The BipA:ppGpp trajectory starts off highly dynamic, but eventually reaches a stable conformation. (a) A plot of RMSD values calculated for the BipA:ppGpp trajectory after LSQ fitting to a calculated average structure. The plot was smoothed with a rolling average of 3000 points. The trajectory of BipA:ppGpp has a relatively high RMSD value, with the highest values occurring early on in the trajectory. (b) A plot of RMSD values calculated for a truncated BipA:GTP trajectory (last 75 ns of BipA:ppGPp trajectory) after LSQ fitting as described for (a).
Figure 9: RMSD Distributions for guanine nucleotide bound forms of BipA. RMSD values are calculated between every structure in a trajectory. RMSD values are grouped into 101 separate bins (spaced evenly between 0 and the maximum RMSD value). RMS distributions for the a) apo-BipA, b) BipA:GTP, c) BipA:GTP and (d) BipA:ppGpp trajectory.
Figure 10: BipA contains two regions that undergo change upon nucleotide binding. Root mean square fluctuation (RMSF) values per residue. High values of RMSF indicate regions of high mobility/flexibility. (a) RMSF profile for apo-BipA. Two peaks of high RMSF (>5 Å) are present. These correspond to SWI in the G domain and a loop region in the CTD. (b) RMSF profile for BipA:GTP. The same regions that have large RMSF values in apo-BipA undergo a decrease in flexibility. (c) RMSF profile for BipA-ppGpp. Profile appears similar to apo-BipA. However, on average the RMSF values are higher, with a modest decrease in RMSF observed in the loop in the CTD. (d) RMSF profile for the truncated BipA:ppGpp trajectory. There is an increase in RMSF values only in the C-terminal helix.
Figure 11: GTP binding greatly reduces flexibility of the loop in the G domain. The difference in RMSF profiles from apo-BipA and BipA:GTP. The loop region in the G domain exhibits a large decrease in the flexibility. The loop region in the CTD sees a mild decrease in flexibility, but retains most of its flexibility.
Figure 12: ppGpp binding greatly reduces flexibility of SWI in the G domain and a loop region in the CTD. The difference in RMSF profiles from apo-BipA and BipA:ppGpp. The loop region in the G domain exhibits a large decrease in the flexibility. The loop region in the CTD also sees a significant decrease in flexibility. The C-terminal helix sees a modest increase in flexibility.
<table>
<thead>
<tr>
<th></th>
<th>$\Delta S_{\text{ASA}}_{\text{Hydrophobic}}$ ($\AA^2$)</th>
<th>$\Delta S_{\text{ASA}}_{\text{Hydrophilic}}$ ($\AA^2$)</th>
<th>$\Delta S_{\text{ASA}}_{\text{Absolute}}$ ($\AA^2$)</th>
<th>$\Delta S_{\text{ASA}}_{\text{Total}}$ ($\AA^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GTP:G Domain</strong></td>
<td>103</td>
<td>137</td>
<td>240</td>
<td>240</td>
</tr>
<tr>
<td><strong>GTP:Domain II</strong></td>
<td>31</td>
<td>-60</td>
<td>91</td>
<td>-29</td>
</tr>
<tr>
<td><strong>GTP:Domain III</strong></td>
<td>113</td>
<td>-69</td>
<td>182</td>
<td>44</td>
</tr>
<tr>
<td><strong>GTP:Domain V</strong></td>
<td>-81</td>
<td>3</td>
<td>84</td>
<td>-78</td>
</tr>
<tr>
<td><strong>GTP:CTD</strong></td>
<td>-82</td>
<td>-133</td>
<td>215</td>
<td>-215</td>
</tr>
<tr>
<td><strong>GTP Total</strong></td>
<td>84</td>
<td>-122</td>
<td>812</td>
<td>-38</td>
</tr>
<tr>
<td><strong>ppGpp:G Domain</strong></td>
<td>-63</td>
<td>-156</td>
<td>219</td>
<td>-219</td>
</tr>
<tr>
<td><strong>ppGpp:Domain II</strong></td>
<td>43</td>
<td>-54</td>
<td>97</td>
<td>-11</td>
</tr>
<tr>
<td><strong>ppGpp:Domain III</strong></td>
<td>174</td>
<td>-136</td>
<td>310</td>
<td>38</td>
</tr>
<tr>
<td><strong>ppGpp:Domain V</strong></td>
<td>-54</td>
<td>90</td>
<td>144</td>
<td>36</td>
</tr>
<tr>
<td><strong>ppGpp:CTD</strong></td>
<td>10</td>
<td>-99</td>
<td>109</td>
<td>-89</td>
</tr>
<tr>
<td><strong>ppGpp Total</strong></td>
<td>110</td>
<td>-355</td>
<td>879</td>
<td>-245</td>
</tr>
</tbody>
</table>

Table 1: BipA:GTP and BipA:ppGpp present slightly more hydrophobic surface than apo-BipA. Table of $\Delta S_{\text{ASA}}$ values (compared to apo-BipA) for hydrophobic and hydrophilic atoms, as well as absolute (sum of the absolute values of $\Delta S_{\text{ASA}}_{\text{Hydrophilic}}$ and $\Delta S_{\text{ASA}}_{\text{Hydrophobic}}$) and net change in SASA. Both BipA:GTP and BipA:ppGpp display decreases in hydrophilic SASA and an increase in hydrophobic SASA, with BipA:ppGpp displaying a larger amount of change.
Figure 13: Examining the metastable states of apo-BipA by PCA and Clustering Analysis. a) 2D projection of the apo-BipA trajectory onto PC1 and PC2. Points are dispersed unevenly. (b) The result of K-means clustering into 3 different groups, represented by different colors.
Figure 14: The single solution conformation of BipA:GTP shown in the 2D projection of the BipA:GTP trajectory. The vast majority of points cluster in a single ellipse, indicating a single stable conformation. The few points outside the ellipse are from the first few nanoseconds of the trajectory, in which an initial transition occurs.
Figure 15: BipA:ppGpp undergoes a conformational change. (a) 2D projection of the BipA:ppGpp trajectory. BipA:ppGpp samples a variety of conformational space throughout the trajectory, with point unevenly distributed. (b) 2D projection of the truncated BipA:ppGpp trajectory. Points in the truncated BipA:ppGpp trajectory concentrate into an ovoid observed in the full trajectory.
Figure 16: BipA:GTP and BipA:ppGpp have different interdomain distances. A cartoon representation of the centroid structures of BipA:GTP (red) and BipA:ppGpp (blue). The largest differences in position occur in the SWI of the G domain and the loop region in the CTD. The CTD is closer to the G domain in the BipA:ppGpp. Globally, the relative positions of the domains are shift between both figures.
<table>
<thead>
<tr>
<th>Distance Between:</th>
<th>BipA:GTP</th>
<th>BipA:ppGpp</th>
<th>Δ Inter-Domain Distance (GTP to ppGpp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G Domain-Domain II (Å)</td>
<td>27.2</td>
<td>26.0</td>
<td>-1.2</td>
</tr>
<tr>
<td>G Domain-Domain III (Å)</td>
<td>37.3</td>
<td>33.8</td>
<td>-3.5</td>
</tr>
<tr>
<td>G Domain-Domain V (Å)</td>
<td>30.6</td>
<td>31.7</td>
<td>1.1</td>
</tr>
<tr>
<td>G Domain-CTD (Å)</td>
<td>55.6</td>
<td>50.3</td>
<td>-5.3</td>
</tr>
<tr>
<td>Domain II-Domain III (Å)</td>
<td>30.1</td>
<td>31.0</td>
<td>0.9</td>
</tr>
<tr>
<td>Domain II-Domain V (Å)</td>
<td>39.1</td>
<td>37.3</td>
<td>-1.8</td>
</tr>
<tr>
<td>Domain II-CTD (Å)</td>
<td>54.7</td>
<td>55.6</td>
<td>0.9</td>
</tr>
<tr>
<td>Domain III-Domain V (Å)</td>
<td>26.5</td>
<td>25.4</td>
<td>-1.1</td>
</tr>
<tr>
<td>Domain III-CTD (Å)</td>
<td>25.1</td>
<td>25.3</td>
<td>0.2</td>
</tr>
<tr>
<td>Domain V-CTD (Å)</td>
<td>38.3</td>
<td>33.9</td>
<td>-4.4</td>
</tr>
</tbody>
</table>

**Table 2:** Relative domain positions are shifted between BipA:GTP and BipA:ppGpp. Distances, in Å, between the centers of mass of the different domains. Shorter inter-domain distances are observed in BipA:ppGpp.
<table>
<thead>
<tr>
<th>Salt-Bridge</th>
<th>Centroid Structures Present In</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASP74-LYS18</td>
<td>All</td>
<td>G Domain:G Domain</td>
</tr>
<tr>
<td>ASP93-ARG25</td>
<td>All</td>
<td>G Domain:Domain II</td>
</tr>
<tr>
<td>GLU338-ARG334</td>
<td>All</td>
<td>Domain III:Domain III</td>
</tr>
<tr>
<td>GLU398-ARG446</td>
<td>All</td>
<td>Domain V:Domain V</td>
</tr>
<tr>
<td>GLU401-ARG361</td>
<td>All</td>
<td>Domain V:Domain III</td>
</tr>
<tr>
<td>GLU576-ARG483</td>
<td>All</td>
<td>CTD:CTD</td>
</tr>
<tr>
<td>GLU579-ARG586</td>
<td>All</td>
<td>CTD:CTD</td>
</tr>
<tr>
<td>GLU84-ARG87</td>
<td>All</td>
<td>G Domain:G Domain</td>
</tr>
<tr>
<td>ASP47-ARG52</td>
<td>BipA:GTP only</td>
<td>G Domain:G Domain</td>
</tr>
<tr>
<td>GLU371-LYS50</td>
<td>BipA:GTP only</td>
<td>Domain III:G Domain</td>
</tr>
<tr>
<td>GLU51-ARG375</td>
<td>BipA:GTP only</td>
<td>G Domain:Domain III</td>
</tr>
<tr>
<td>ASP44-ARG375</td>
<td>BipA:ppGpp Only</td>
<td>G Domain:Domain III</td>
</tr>
<tr>
<td>ASP47-ARG375</td>
<td>BipA:ppGpp Only</td>
<td>G Domain:Domain III</td>
</tr>
<tr>
<td>GLU49-ARG41</td>
<td>BipA:ppGpp Only</td>
<td>G Domain:G Domain</td>
</tr>
<tr>
<td>GLU376-ARG52</td>
<td>BipA:ppGpp Only</td>
<td>Domain III:G Domain</td>
</tr>
<tr>
<td>GLU570-ARG384</td>
<td>BipA:ppGpp Only</td>
<td>CTD:Domain III</td>
</tr>
<tr>
<td>GLU304-ARG346</td>
<td>All but BipA:GTP</td>
<td>Domain III:Domain III</td>
</tr>
<tr>
<td>GLU576-LYS589</td>
<td>All but BipA:GTP</td>
<td>CTD:CTD</td>
</tr>
<tr>
<td>ASP532-LYS589</td>
<td>All but BipA:GTP</td>
<td>CTD:CTD</td>
</tr>
</tbody>
</table>

Table 3: BipA:GTP and BipA:ppGpp form salt bridges between the G domain and Domain III. A table describing the salt-bridges between conserved residues in the centroid structures of apo-BipA, BipA:GTP and BipA:ppGpp. Interactions observed in all three structures may be important to the stability of BipA. Those unique to BipA:GTP and BipA:ppGpp occur primarily between the G domain and Domain III.
Figure 17: BipA:GTP and BipA:ppGpp use different residues for contact between the Switch I of the G domain and domain III. Ribbon diagrams of the G domain/domain III interface in the centroid structures of (a, c) BipA:GTP (green) and (b, d) BipA:ppGpp (cyan). Conserved residues that form salt bridges between the G domain and domain III are represented as sticks. BipA:ppGpp is present for comparison.
Figure 18: Conformational selection drives the transition from apo-BipA to BipA:ppGpp. A matrix of RMSD values calculated after LSQ fitting every structure in the Apo-BipA trajectory to every structure in the BipA:GTP trajectory.
Figure 19: BipA transitions to a ppGpp bound state by an induced fit mechanism. A matrix of RMSD values calculated after LSQ fitting every structure in the Apo-BipA trajectory to every structure in the BipA:ppGpp trajectory.
Figure 20: Principle component analysis of the various nucleotide bound forms of BipA. 2D projections of the apo-BipA (black), BipA:GTP (red), BipA:ppGpp (green), and truncated BipA:ppGpp (blue) trajectories. All graphs have PC1 on x-axis. Y-axes are (a) PC2, (b) PC3, (c) PC4, (d) and PC5.
Figure 21: Principle component analysis of residues within the nucleotide-binding pocket for the various forms of BipA. Trajectories were extracted for residues within 10 Å of GTP in the BipA:GTP centroid structure. These trajectories (apo-BipA, BipA:GTP and BipA:ppGpp) were used to construct a covariance matrix, from which eigenvectors and values were extracted. 2D projections of the apo-BipA (black), BipA:GTP (red), BipA:ppGpp (green), and truncated BipA:ppGpp (blue) nucleotide binding pocket trajectories. All graphs have PC1 on x-axis. Y-axes are (a) PC2, (b) PC3, (c) PC4, (d) and PC5.
Figure 22: Structural mechanism for ppGpp recognition. Ribbon diagram of the BipA:ppGpp centroid structure. BipA recognizes the 3’ pyrophosphate group of ppGpp through an interaction with arginine 41.
Figure 23: Involvement of arginine in ppGpp recognition. LIGPLOT schematics for ppGpp interactions in structures of non-GTPases. (a) RNA polymerase holoenzyme from *Thermus thermophiles* (b) RNA polymerase from *E. coli* DH1 (c) *E. coli* K-12 (d) guanosine pentaphosphate phosphohydrolase from *Aquifex aeolicus*
Figure 24: Comparison of induced fit and conformational selection models. The induced fit model states that the binding of a ligand to the protein promotes a conformational change resulting in higher affinity of that ligand for the protein. Conformational selection postulates that the unbound protein exists in many energetically accessible conformations in solution. The ligand binds to a given conformation and shifts the equilibrium towards this stabilized bound state. ("Model Binding Systems," 2010)
Appendix 1: *Salmonella enterica* serovar Typhimurium R41A mutagenesis

Following these analyses, an R41A mutation was introduced into *Salmonella enterica* serovar Typhimurium strain of BipA in a pET28a vector. This vector was transformed into *Escherichia coli* BL21 for expression. This mutant is ready for analysis of nucleotide binding affinity and ribosomal association.
References


