ABSTRACT

Title of Document: LIGAND-LINKED DISORDER-TO-ORDER TRANSITIONS IN ALLOSTERIC COMMUNICATION

Christopher Eginton, Doctor of Philosophy, 2015

Directed By: Dr. Dorothy Beckett, Department of Chemistry and Biochemistry

Ligand-linked disorder-to-order transitions are integral to the function of numerous cellular proteins. Furthermore, these transitions can contribute to allosteric regulation, a principal mechanism controlling protein function. Despite their ubiquity, the relationship between the sequence and function of these regions and their mechanisms of achieving allosteric regulation are not well understood. The *Escherichia coli* biotin repressor, BirA, is a bifunctional protein that provides a model system to investigate these questions. Binding of the corepressor, biotinyl-5'-AMP, is coupled to a disorder-to-order transition resulting in a complex network of hydrophobic residues packing over the adenylate moiety. Additionally, this binding event is coupled to BirA dimerization, enhancing the self-association free energy by -4.0 kcal/mol. In this work, the sequence-function relationship of the disorder-to-order transition was investigated using several combinations of alanine substitutions in the hydrophobic network. Equilibrium binding and kinetic measurements show that

the full functional response in the disorder-to-order transition is achieved through the appropriate packing of hydrophobic residues in the hydrophobic network. In addition to the disorder-to-order transition on the ligand binding surface, the dimerization interface contains several regions that, while disordered in the unliganded monomer, are folded in the liganded dimer. Through structural and thermodynamic analysis of the G142A variant, long-distance reciprocal communication between disorder-to-order transitions on the ligand binding and dimerization surfaces is identified as central to allostery. Together, these results demonstrate the functional versatility of disorder-to-order transitions and how the sequences of these regions dictate protein function and allosteric regulation.

LIGAND-LINKED DISORDER-TO-ORDER TRANSITIONS IN ALLOSTERIC COMMUNICATION

By

Christopher Ryan Eginton

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2015

Advisory Committee:
Professor Dorothy Beckett, Chair
Professor Catherine Fenselau
Professor David Fushman
Professor George Lorimer
Professor Roy Mariuzza

Dedication

This dissertation is dedicated to my Father Gary L. Eginton and my Grandmother Evelyn M. Eginton.

Acknowledgements

I would first like to thank my advisor Dr. Beckett for her inspiration and support without which I would not be the scientist or person I am today. I am forever grateful for her involvement in my graduate career.

I express my thanks and gratitude to my committee members; Dr. Fenselau, Dr. Fushman, Dr. Lorimer, and Dr. Mariuzza for their support and guidance.

I thank my fellow graduate students and lab mates for their friendship. I would especially like to thank Maria, Poorni, and Ryan for their support and continued friendship. I would also like to thank Bill and Jingheng for being great friends and lab mates.

I would like to express my gratitude to the Department of Chemistry and Biochemistry for the William Bailey Research Fellowship and Sampugna-Keeney Travel Award. I thank the University of Maryland Graduate School for the Jacob K. Goldhaber Travel Award, University of Maryland Summer Research Fellowship, and the Ann G. Wylie Dissertation Fellowship.

Lastly I would like to thank my friends and family. None of this would have been possible without their love and support.

This work was funded, in whole or in part, by the National Institutes of Health Grant R01GM46511 and S10RR15899 to Dr. Beckett.

Table of Contents

List of Tables

List of Figures

Chapter 1: Background and Introduction

1.1 Introduction

Ligand-linked conformational changes, which are central to the physiological function of many proteins (1-3), are frequently manifested as disorder-to-order transitions. This topic has gained considerable interest through the identification of intrinsically disordered proteins (IDPs) involved in a wide array of processes including transcription and neurodegenerative disease (4, 5). In addition, computational studies predict that more than thirty percent of all eukaryotic proteins, many of which are involved in signaling and regulation, contain disordered regions (6, 7). Proteins lacking ordered regions or any three-dimensional structure in the absence of a binding partner often become ordered upon binding to a target molecule (8). Furthermore, ligand binding can be distal to the disorder-to-order transition, connecting these transitions to allosteric regulation (9, 10). Nonetheless, ligand-linked disorder-to-order transitions are poorly understood and the relationship between sequence and function of these regions and their role in allosteric regulation has yet to be established. In this work, the *Escherichia coli* biotin repressor is used as a model system to address these questions.

1.1.1 Ligand-linked conformational change

The prevailing mechanism of binding has shifted over the years starting in the late 1900's when it was proposed by Emil Fischer that enzyme-substrate binding was

analogous to a lock and key (11). The subsequently termed "lock-and-key" hypothesis described enzymes as rigid entities in which the substrate is a precise fit for a given binding site. However, over time this model was not able to explain several functional observations in enzyme catalysis, which led to the development of models that couple ligand binding and conformational change in the form of induced fit and conformational selection. The induced fit theory developed by Daniel Koshland provided an alternative view on how binding can occur between an enzyme and substrate where a precise fit was not initially present (12). In this model, substrate binding precedes a conformational change in the enzyme active site to optimize fit. The induced fit model helps explain anomalies such as why kinases including hexokinase, lack ATPase activity in the absence of substrate (13). By contrast, the conformational selection model proposes that the apo protein samples both conformational states and the ligand stabilizes the bound, lower energy state (14-16). Both the induced fit and conformational selection models have been rigorously applied to catalytic enzymes undergoing conformational change upon binding and more recently to proteins with intrinsically disordered regions in which ligand binding is coupled to folding (17, 18).

1.1.2 Ligand-linked conformational change in catalysis

Enzyme catalysis is required for a broad range of reactions involved in metabolism. The enzyme adenylate kinase (AdK) catalyzes the reversible inter-conversion of adenine nucleotides, a process crucial to ensure energy balance in cells (1). Adenylate kinase has a well-defined three-dimensional structure composed of the

CORE, ATP-LID, and nucleotide monophosphate (NMP) domains (19). Structural data have revealed a large conformational change that joins the ATP-LID and NMP domains upon small molecule binding (Fig. 1) (19, 20). The AdK inhibitor AP_5A, an analog of conjoined ATP and AMP connected with an additional phosphate that simulates both substrates has been used to capture the "closed" conformation of the enzyme that is necessary for catalysis (20).

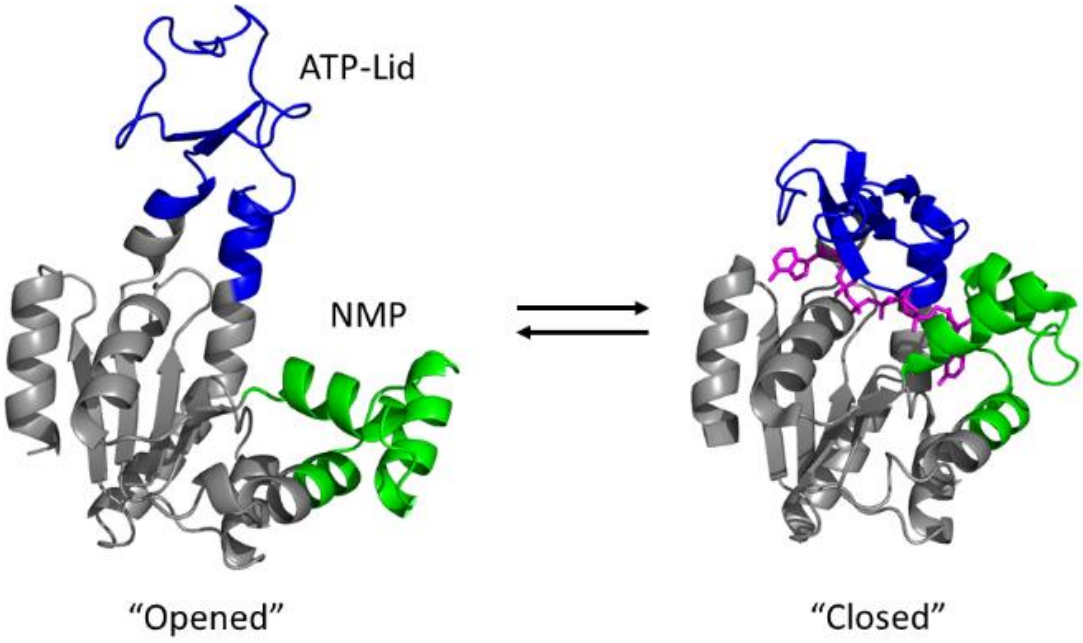

Figure 1. X-ray crystallographic structures of adenylate kinase from *Escherichia coli* in the unliganded "open" conformation (4AKE) and the bound "closed" conformation with the inhibitor Ap5A (1AKE) with the CORE (grey), ATP-LID (blue), and nucleotide monophosphate (NMP) (green) domains highlighted and visualized in PyMOL(19-21).

Furthermore, data obtained from nuclear magnetic resonance (NMR) dispersion have shown that the opening of the ATP-LID is the rate-limiting step for catalytic reaction emphasizing the importance of the conformational transition (22). Molecular dynamic simulations have provided insight into the dynamic nature of adenylate kinase. The apo enzyme appears to sample both open and closed conformations with the closed conformation stabilized through ligand binding in agreement with a conformational selection mechanism (17). Studies on adenylate kinase highlight its dynamic nature and the importance of conformational transitions accompanying substrate binding.

1.1.3 Ligand-linked folding transitions

The notion that a rigid three dimensional structure is required for function has been further challenged in recent years with the discovery of intrinsically disordered proteins. Contrary to their ordered counterparts, IDPs undergo a conformational transition upon binding typified by a shift to a more ordered state. The dynamic nature of IDPs allows them to sample a range of conformational states, often with tenuous secondary and/or tertiary structure that can be stabilized by substrate binding. The extent of these transitions vary, often ranging from a random coil or molten globule state to a well-defined three dimensional structure (23). The absence of defined structure provides IDPs with an advantage with respect to molecular recognition of many partners (24). Furthermore, the increased interaction surface or capture radius has been proposed to increase rates of association accounting for the high number of IDPs involved in signaling (24). While both electrostatic interactions

between IDPs and their partners play important roles in binding, recent evidence suggests that hydrophobic interactions, analogous to those seen in protein folding, are a significant determinant for association (25-27). The association of IDPs has been described as a frustrated energy landscape in which the interaction between partners is not optimized (26). This result may explain the tendency for IDPs to be involved with many functional partners. Conversely, for most proteins with well-defined structure, association is characterized as being minimally frustrated in which the protein in the folded or bound state has all interactions optimized.

Coupled binding linked to folding is illustrated by the interaction between the activator for thyroid hormone and retinoid receptors protein (ACTR) of the p160 nuclear receptor and the nuclear co-activator binding domain (NCBD) of CREB binding protein (CBP). Nuclear receptors function as transcription factors to regulate the expression of genes important for development and cell differentiation (28). The transcription activation is accomplished through the recruitment of the p160 nuclear receptor co-activator and the transcriptional co-activator CBP. ACTR is disordered with ill-defined secondary structure while NCBD has a native-like secondary structure representative of a molten globule state (23, 29). These two intrinsically disordered proteins undergo coupled binding-folding with an equilibrium dissociation constant in the nanomolar range (Fig. 2) (23, 30). Kinetic and structural investigations have led to disagreement about the mechanism of association, specifically whether ACTR and NCBD associate by an induced fit or conformational selection model (25, 29). The binding interface mainly consists of hydrophobic residues rich in leucine

repeats separated by polar residues. Additionally, electrostatic interactions are also significant for the association between the conserved residues aspartate 1068 of ACTR arginine 2105 of NCBD (23). It is clear that ACTR and NCBD associate *via* a large ligand-linked folding transition increasingly seen in disordered proteins. However, there are still many unanswered questions about the relationships between sequence, function and folding in these transitions.

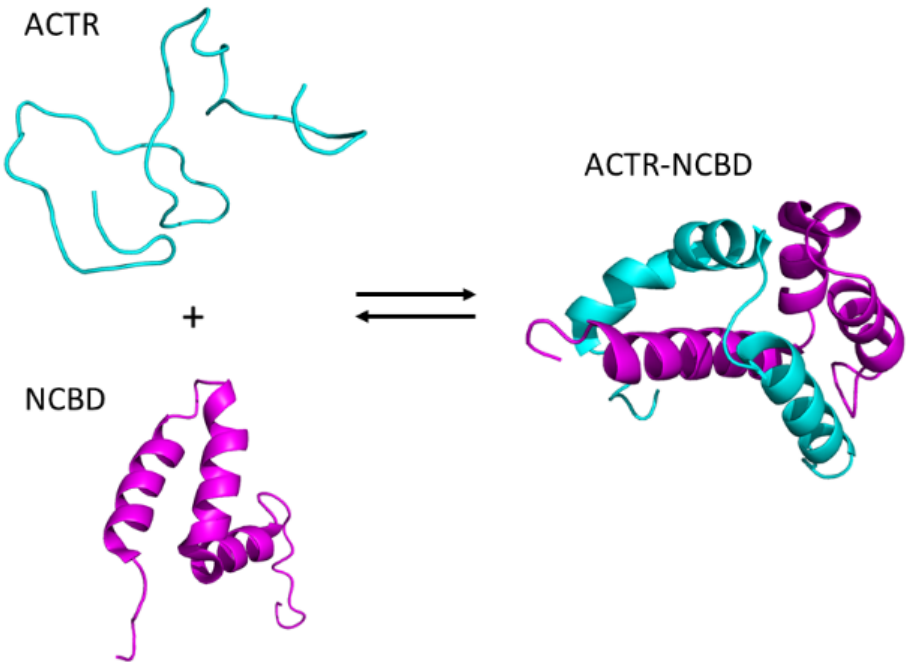

Figure 2. Ligand-linked folding of the thyroid hormone and retinoid receptors protein (ACTR) of the p160 nuclear receptor and the nuclear co-activator binding domain (NCBD) of CREB binding protein (CBP). A conceptualized model of ACTR (cyan) is shown to display the intrinsic disorder of the protein along with NMR solution structures of NCBD alone (magenta, 2KKJ) and bound to ACTR (1KBH) visualized in PyMOL (21, 23, 29).

1.2 Allosteric Communication in Proteins

Control of protein function through allosteric regulation is widespread in biological systems. Allosteric communication is the process by which binding of an effector molecule at one site results in a functional change at another, often distal, site. In the case of allosteric enzymes, effector binding can alter enzymatic activity in two ways. Effector binding can increase of decrease the V_{max} or K_m and referred to as V-type or K-type effectors (31). The concept of allostery emerged in the early 1960's through the study of feedback inhibition of L-threonine deaminase by L-isoleucine (32). The two prevailing allosteric models were the "symmetric" or MWC (Monod–Wyman–Changeux) and the "sequential" or KNF (Koshland–Nemethy–Filmer) that dominated the field for years (33, 34). Both models relied on the importance of conformational change upon binding with identical oligomeric subunits in two distinct end states, tense (T) and relaxed (R), which are characterized by lower and higher affinity for substrate, respectively. In the MWC model, all subunits reside in either the T or R state, which is to say a conformational change in one subunit results in the same change to all other subunits. In the KNF model, binding at one subunit occurs through an induced fit mechanism, transitioning from T to R, though adjacent subunits remain in the T state but are more receptive to ligand binding.

1.2.1 Allostery through rigid body inter-conversion

Allostery communicated through rigid body inter-conversions has been able to accurately describe several systems including aspartate transcarbamoylase (ATCase) (35). ATCase is an allosteric enzyme that catalyzes the condensation of aspartate with

carbamoyl phosphate to form N-carbamoylaspartate, the first committed step in pyrimidine synthesis (36-38). The enzyme is a 310 kDa dodecamer, consisting of three dimers of regulatory subunits and two trimers of catalytic subunits totaling 6 active sites (39, 40). To further understand the mechanism of allosteric regulation, ATCase was crystallized with the bisubstrate analog N-phosphonacetyl-L-aspartate (PALA) (41). The large differences in quaternary structures of apo and PALA-bound ATCase led to the defined T and R states, respectively (Fig. 3) (41, 42). Despite these observations, the question remained as to whether the allosteric response follows the concerted MWC or sequential KNF model. This was investigated by developing a form of ATCase with only one wild type active site and five others containing the R105A substitution, which renders the subunit deficient in substrate binding (35). Small-angle x-ray scattering (SAXS) and crystallographic data showed that upon binding of PALA to the wild type active site, there is a concerted quaternary conversion of all subunits, both active and inactive, to the R state. These results support a concerted model of allosteric regulation for ATCase in which effector binding to one-site shifts the remaining sites to the R state.

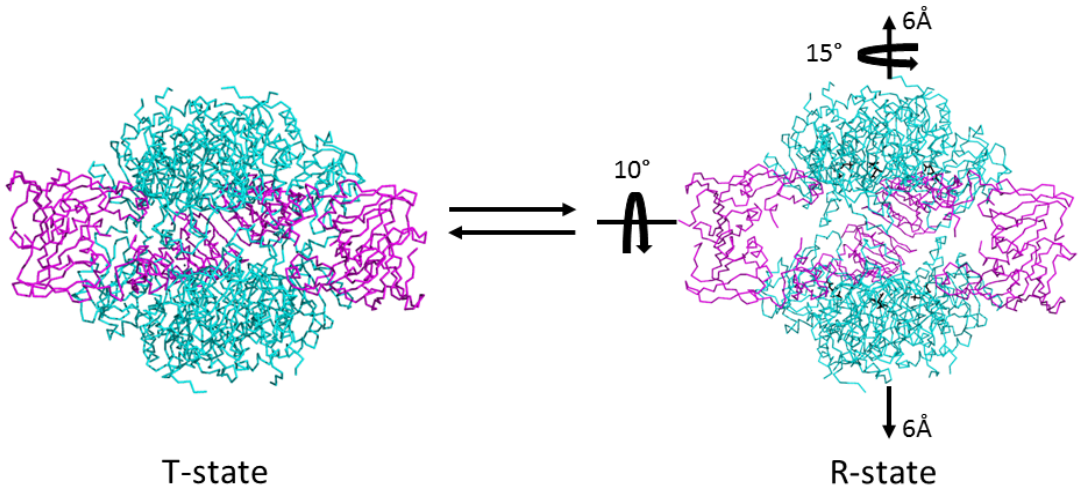

Figure 3. X-ray crystallographic structures of aspartate transcarbamoylase in the T and R-states (6AT1 and 8ATC respectively) (41, 42). Structures are shown with both catalytic subunits (cyan) and regulatory subunits (magenta) highlighted in PyMOL (21). The R-state structure bound to the bisubstrate analog N-phosphonacetyl-L-aspartate (PALA) shown in black sticks.

1.2.2 Allostery through changes in dynamics

Both the MWC and KNF models of allostery are based on a view of rigid conformational transitions focused on distinct end states frequently obtained from x-ray crystallographic structures. However, evidence suggests that these rigid-structural transitions do not explain all of the available data and a growing number of systems have been identified that utilize dynamics or disorder-to-order transitions, thus increasing the mechanistic repertoire for allosteric regulation (Fig. 4) (43). Proteins are dynamic in nature and thus open the possibility to this being involved in allosteric

communication. Cooper and Dryden were the first to propose a general model in which allosteric communication between distant binding sites is facilitated through entropic contributions (44). Conventionally, allosteric communication was described to operate by rigid or mechanical conformational changes to describe cooperative linkage in proteins. Cooper and Dryden derived an expression that showed how cooperative interactions between distant sites can be produced through changes in vibrational modes, dynamics, and thermal fluctuations. This work laid the foundation for allosteric communication achieved through changes in dynamics and has since gained considerable interest through advances in NMR and computational approaches.

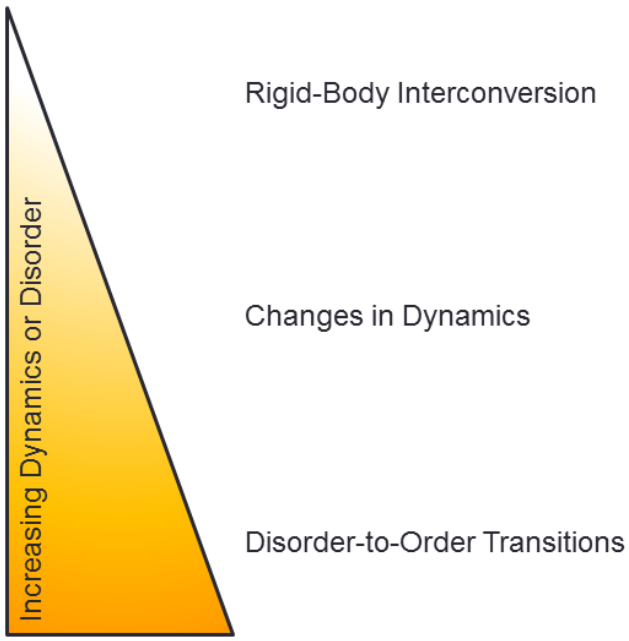

Figure 4. Continuum of recognized mechanisms for allosteric communication. Increasing contributions from dynamics or disorder to the mechanism for allosteric communication is represented by the position on the vertical axis.

One of the best studied systems in which dynamics play a significant role in the allosteric response is the enzyme dihydrofolate reductase (DHFR). DHFR catalyzes the reduction of dihydrofolate to tetrahydrofolate using the cofactor nicotinamide adenine dinucleotide phosphate (NADPH) and is required for the biosynthesis of nucleotides and amino acids (2). The functional role DHFR performs in metabolism has made it an attractive drug target for antiproliferative therapeutics (45). Two functional loops near the DHFR active site have been shown to be particularly important for ligand binding and catalysis and are referred to as the F-G (116-132) and Met20 (9-24) loops (46). Prior to binding NADPH, the Met20 loop protrudes into the NADPH active site and is referred to as the occluded conformation (47). Upon NADPH binding, the Met20 loop assembles over the cofactor while the F-G loop folds over the Met20 loop forming the closed conformation (Fig. 5). The folding and subsequent hydrogen bond formation between the F-G and Met20 loops are necessary to maintain the closed conformation as well as catalysis (47). Moreover, two-dimensional, heteronuclear (^{1}H-^{15}N) NMR studies have implicated the dynamic fluctuations of the F-G loop in the ligand-dependent catalysis (48). Particular focus has been on residue glycine-121 and the variant G121V that resides at the center of the F-G loop, 19Å from the enzyme catalytic site (47, 49). The G121V variant exhibits a 40-fold reduction in NADPH binding (46). Additionally, kinetic measurements show that hydride transfer is reduced 200-fold and that the rate of conformational change that is seen in both mouse and human DHFR are altered (46). While the molecular basis for the observed changes is not known, the replacement of glycine-121 results in the loss of conformational fluctuations of the F-G loop

resulting in less structural interplay between the F-G and Met20 loops to stabilize the closed conformation (47).

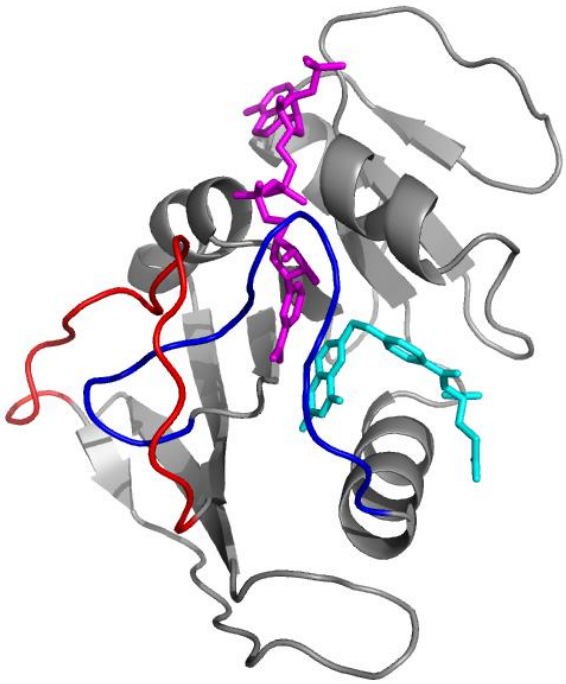

Figure 5. X-ray crystallographic structure of *E. coli* dihydrofolate reductase (1RX2) (47). The structure of DHFR in the closed conformation is shown with the F-G (red) and Met20 (blue) loops highlighted in PyMOL (21). Bound ligands folic acid and nicotinamide adenine dinucleotide phosphate (NADP) (Oxidized) are show in cyan and magenta respectively.

1.2.3 Disorder-to-order transitions in allosteric communication

In addition to rigid conformational changes and dynamics, disorder-to-order transitions have recently been shown to contribute to allosteric regulation.

Furthermore, a theoretical model in which coupling between distant sites is optimized through disorder-to-order transitions has been proposed (10). This mode of allosteric regulation is hypothesized to be particularly prevalent in transcription factors, which contain a high number of disordered regions compared to other classes of proteins (50). Moreover, DNA binding by these proteins is frequently altered by ligand-linked folding/unfolding events (50-53). The tetracycline repressor (TetR) is a homodimeric protein comprised of a DNA-binding domain and a tetracycline-binding and dimerization domain. Tetracycline binding results in a decrease in the affinity of TetR for DNA. An allosteric model was formulated based on comparisons of the structures of free and tetracycline plus DNA-bound TetR (54-56). In this model effector binding resulted in decreased DNA binding affinity by inducing an altered DNA-binding domain conformation that was unable to bind DNA. However, more recently it has been shown that in the unbound state, the DNA-binding domain samples a range of conformation space, which enhances DNA association (9). Binding of tetracycline results in the cooperative folding between the DNA-binding domain and the C-terminal tetracycline binding/dimerization domain which shifts the DNA binding domain to a rigid conformation that is incompatible with DNA binding. While disorder-to-order transitions appear to be involved in allosteric communication, the limited experimental data on the mechanism of communication justifies further study of model systems.

1.3 The E. coli Biotin Regulatory System

The *Escherichia coli* biotin repressor, BirA, is a bifunctional allosteric transcriptional regulatory protein that is the central component of the Biotin Regulatory System (Fig. 6) (57). BirA is both a metabolic enzyme and a transcription repressor of the biotin biosynthetic operon. For both functions, BirA first binds biotin followed by ATP to catalyze the synthesis of the corepressor, biotinyl-5'-AMP (bio-5'-AMP). This is accomplished through anhydride bond formation between the carboxyl group of biotin and the alpha phosphate of ATP with pyrophosphate release (58). The corepressor-bound species can heterodimerize and subsequently biotinylate the biotin carboxyl carrier protein (BCCP) subunit of acetyl-CoA carboxylase (ACC). Biotin plays an integral role in the first committed step of fatty acid anabolism as an intermediate carrier of a carboxyl group from bicarbonate to the acceptor, acetyl-CoA, to produce malonyl-CoA (59, 60). Alternatively, the corepressor-bound BirA can homodimerize and bind site-specifically to the 40-base pair biotin operator, BioO, to repress the transcription of the biotin biosynthetic genes (61-63). The partitioning between the metabolic or transcription regulatory function is dependent on the availability of apoBCCP (64, 65). The apoBCCP concentration is markedly higher during exponential growth due to the increased requirement for membrane synthesis and, in turn, holoBirA partitions toward its metabolic function. However, when growth is slowed reducing the available apoBCCP, corepressor-bound BirA homodimerizes to perform its transcriptional regulatory function.

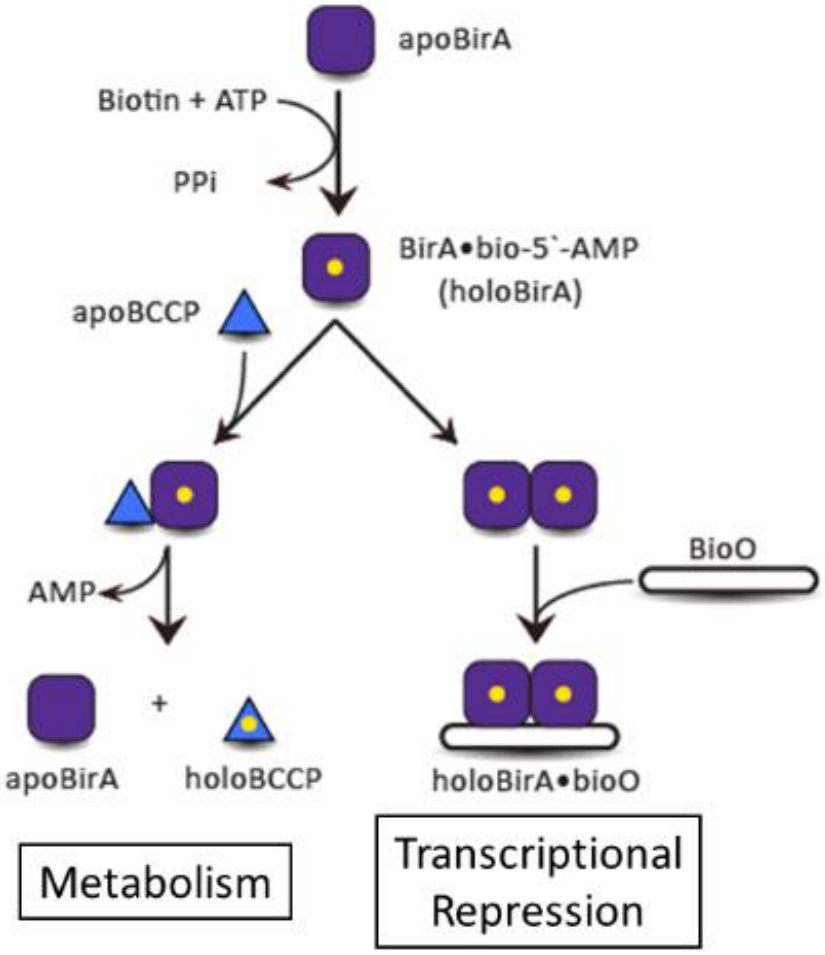

Figure 6. Schematic of the *E. coli.* Biotin Regulatory System. The apo enzyme, BirA, first binds to biotin, then ATP to catalyze the formation of the corepressor, bio-5'-AMP. HoloBirA can then function as either a metabolic enzyme or as a transcription repressor.

Ligand binding and self-association of BirA are coupled processes in which the magnitude of energetic coupling is dependent on the ligand bound. The corepressor, bio-5'-AMP, displays the largest enhancement to dimerization energetics

with -4.0 kcal/mol relative to the apo protein (Fig. 7) (63, 66). The dimerization energetics of BirA complexed with several analogs of bio-5'-AMP were also measured using sedimentation equilibrium (67). While biotin results in no enhancement of dimerization energetics, the analogues 5'-O-[N-(biotinoyl)-sulfamoyl] adenosine (btn-SA) and biotinol-5'-AMP (btnOH-AMP) display -1 and -3.4 kcal/mol enhancement, respectively. The measured equilibrium dissociation constant for the self-association of bio-5'-AMP-bound BirA is 6 μM at 20°C corresponding to a Gibbs free energy of -7.0 kcal/mol (66). The thermodynamic driving forces of the homodimerization reaction were determined by measuring the temperature dependence of holoBirA dimerization. The equilibrium dissociation constant measured at several temperatures was subjected to van't Hoff analysis to obtain the enthalpy of dimerization from which the entropic contribution was calculated at each temperature (68). The analysis yields large opposing enthalpic and entropic contributions of 41 and -48 kcal/mol, respectively, at 20°C. The large opposing thermodynamics parameters of the dimerization process are consistent with solvent release upon homodimerization. The large unfavorable enthalpic contribution is interpreted as the penalty from the removal of water from the dimerization surface whereas the favorable entropic contribution reflects the release of bound water into the bulk. This interpretation is consistent with the minor salt-dependence observed for the self-association reaction indicating that electrostatic interactions do not play a significant role in the process (68).

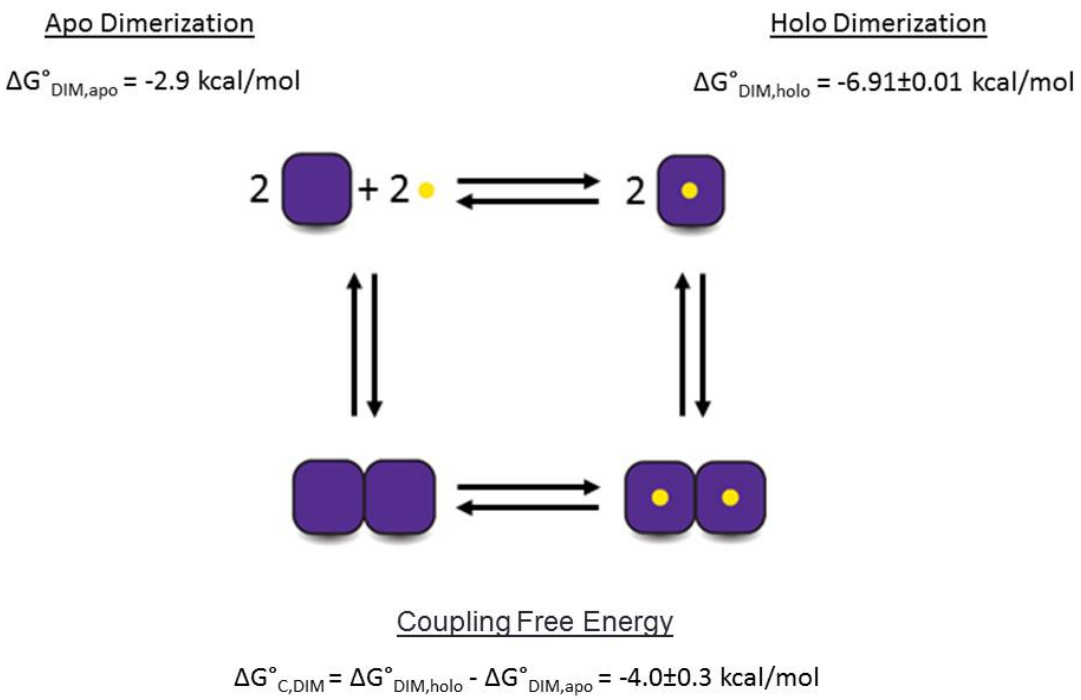

Figure 7. Cycle illustrating the thermodynamic linkage between bio-5'-AMP binding and dimerization of BirA. Binding and dimerization are coupled processes in which corepressor binding enhances dimerization by -4.0 kcal/mol. The free energy of apo BirA dimerization was extrapolated using the value measured in 50 mM KCl and the known salt-concentration dependence on the process (63, 66).

In addition to the enhanced coupling free energy, previous research has revealed evidence for the existence of distinct ligand bound conformational states during the functional cycle (69). Kinetic measurements of the association process of BirA with biotin and bio-5'-AMP indicate that association is at least a two-step process. The first step is the collision complex formation and the second is a

unimolecular conformational transition of the complex. This conformational transition was further investigated through subtilisin-catalyzed proteolysis of the apo enzyme in addition to the two ligand bound states. Subtilisin cleavage occurs in between residues 217 and 218 of a loop in the ligand binding site that is unstructured in apoBirA. The rates of cleavage for the biotin and bio-5'-AMP bound species were decreased 2-fold and 10-fold, respectively, compared to the apo protein. These combined results provided the first evidence of distinct ligand bound conformational states of BirA. X-ray crystallographic studies have further shed light on both the conformational transition in the ligand binding site and the observed increase in coupling free energy for the adenylate-bound complex relative to the unbound or biotin bound species.

1.3.1 The BirA structure

The *E. coli* biotin repressor is a 35.3 kDa globular protein comprised of three distinct domains (70). The structures of apoBirA (1BIA) and holoBirA bound to both biotin (1HXD) and btnOH-AMP (2EWN), a non-hydrolysable analog of the physiological corepressor, were obtained through x-ray crystallography (Fig. 8) (70-72). The N-terminal domain consisting of residues 1-60 is characterized by a winged-helix-turn-helix DNA-binding domain. The N-terminal and central or catalytic domains are connected by a flexible linker of residues 61-79. The catalytic domain, residues 80-269, consists of a 7-stranded mixed beta-sheet packed against 5 alpha helices in addition to six surface loops. Lastly, the C-terminal domain of BirA consists of an SH3-like domain comprised of residues 271-321. In the apo structure,

BirA is monomeric with four of the six surface loops of the catalytic domain partially disordered. On the ligand binding surface two disordered loops consisting of residues 116-126 and 211-234, are referred to as the biotin binding loop (BBL) and adenylate binding loop (ABL), respectively. On the dimerization surface the disordered loops are comprised of residues 140-146 and 193-199. Upon binding biotin, three of the four loops undergo disorder-to-order transitions including the two dimerization surface loops and the biotin binding loop. Although the BirA-biotin complex crystallizes as a dimer, it is monomeric in sedimentation equilibrium measurements performed at protein concentrations as high as 100 μM (73). The btnOH-AMP bound structure crystallizes as a dimer consistent with results from sedimentation equilibrium (63). The dimerization interface contains a combination of non-polar, polar, and charged residues (72). Analysis of the interface reveals 16 potential hydrogen bonds and 4 salt-bridges across the interface (74). Together with the self-association process, all of the previously disordered loops are now ordered including the adenylate binding loop that is packed over the adenylate moiety. Combined, the results from both structural studies and solution measurements highlight the importance of these loop regions and their ligand-induced conformational transitions for BirA function.

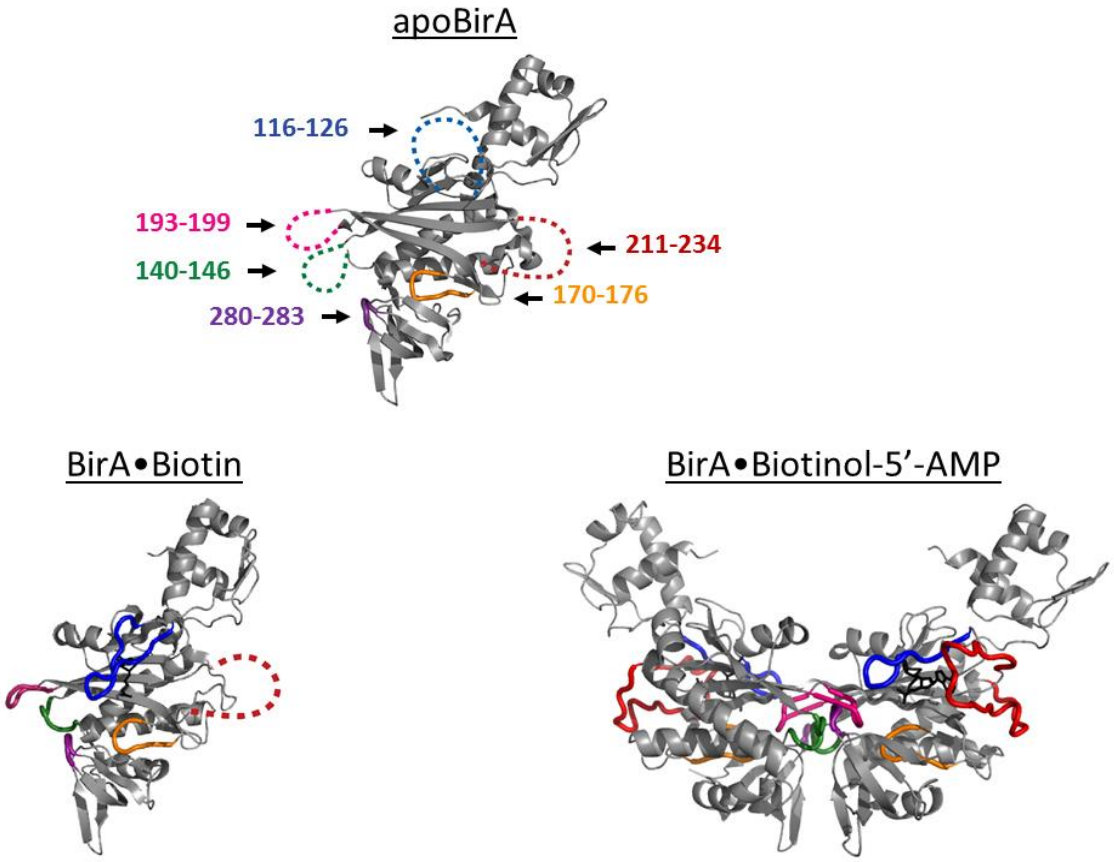

Figure 8. Structures of apo (1BIA), biotin (1HXD), and btnOH-AMP (2EWN) bound BirA obtained from x-ray crystallography (70-72). All loops in the catalytic domain are labeled and color-coded with the partially disordered and folded loops shown as dashed and solid lines, respectively. The model was generated using PyMOL (21).

1.3.2 Investigation of BirA surface loops

Several surface loops have been investigated for their roles in the multiple BirA functions through a combination of mutagenic, thermodynamic, and kinetic measurements. The biotin binding loop has been investigated for its importance in ligand binding and homodimerization. The BBL contains a glycine rich sequence

GRGRXG which is conserved among all biotin protein ligases, was investigated through the variants G115S, R118G, and R119W (75, 76). The results confirm that the loop is important for biotin and bio-5'-AMP binding as well as homodimerization. The findings, particularly the observed functions in biotin and bio-5'-AMP binding, are in agreement with the conservation of these residues among other biotin protein ligases.

The adenylate binding loop has also been investigated for its functional role in multiple BirA functions. Bio-5'-AMP results in the formation of an extensive hydrophobic network involving residues of the adenylate binding loop (ABL) and BBL including F124, P126, M211, V214, V218, V219, and W223. Previous studies have shown that substitution of any of the three ABL residues, V214, V219, and W223 with alanine, results in perturbed ligand binding, ABL folding, and self-association indicating the formation of this network is crucial for BirA function (77). The thermodynamics of ligand binding, measured using isothermal titration calorimetry, indicate penalties to the binding free energy of approximately -3 kcal/mol for each variant. Additionally, subtilisin-catalyzed proteolysis displayed decreased protection from cleavage resulting from bio-5'-AMP binding for all of the variants. Lastly, the variants were deficient in homodimerization with the largest effect observed for V219A with a penalty of -1.5 kcal/mol. Hydrophobic networks over the adenylate are also observed in the crystal structures of BirA homologs, *Staphylococcus aureus* and *Mycobacterium tuberculosis* (78, 79). Despite the

analogous formation of the hydrophobic network in other BPL's, the sequences of these regions and the residues directly involved the network are not conserved.

Four dimerization surface loops were systematically investigated for their importance in the homodimerization reaction through alanine scanning mutagenesis (80). The loops examined include the three variable loops, of residues 140-146, 193-199, and 280-283 that share little sequence similarity with other bifunctional ligases, and the conserved loop comprised of residues 170-176. Upon substitution of residues in the conserved loop, homodimerization was in most cases severely affected. In the variable loops, several alanine substitutions had significant effects on homodimerization. The variable loop residues that are sensitive to substitution make structural contacts with constant loop residues and act to complement these conserved regions. The most significant effects to homodimerization were observed for the G142A, I280, and K194A variants, which have free energy values of dimerization of -2.4, -3.3, and -10.6 kcal/mol, respectively. The dimerization free energy of the adenylate-bound G142A variant, which is identical to that of apoBirA, is consistent with complete loss of coupling between corepressor binding and dimerization.

1.3.3 Biotin protein ligases of *S. aureus* and *M. tuberculosis*

The bifunctional BPL in *S. aureus* acts as a transcriptional repressor and metabolic enzyme (78, 81). Similar to BirA, the transcriptional regulatory role of *S. aureus.* BPL is to repress transcription of biotin biosynthetic genes. By contrast, the metabolic function is two-fold, to transfer biotin to acetyl-coA carboxylase for fatty

acid synthesis, and to transfer biotin to pyruvate carboxylase to replenish the tricarboxylic acid (TCA) cycle with oxaloacetate (82). The three-dimensional structure of the *S. aureus* BPL is similar to other Class II biotin protein ligases (70, 78). The N-terminal domain consists of a winged helix-turn-helix DNA-binding motif. The central or catalytic domain is comprised of 8 beta-strands packed against 7 alpha helices while the C-terminal domain has an SH3-like fold. The structure of apo *S. aureus.* BPL is monomeric with the biotin binding loop (residues 118-129) partially disordered (78). Conversely, the holo structures of *S. aureus.* BPL in complex with both biotin and bio-5'-AMP are dimeric with the BBL ordered (78). In all structures the ordered ABL (residues 220-228) occupies the same position. Structural comparison indicates that, similar to *E. coli* BirA, *S. aureus* BPL undergoes a monomer-dimer transition upon binding either biotin or bio-5'-AMP in addition to ligand-linked folding of the BBL. Analogous to BirA, bio-5'-AMP binding results in formation of a hydrophobic network including residues L216, F219, I223, A227 packed with the adenylate, albeit with different residues. The emergence of methicillin-resistant *Staphylococcus aureus* (MRSA) coupled with the essential function bacterial BPL's perform have made it an attractive drug target (83). Current efforts are encouraging with a focus on exploiting the lack of sequence conservation between various BPL adenylate binding loops. A 1,2,3-triazole analog of bio-5'-AMP developed using click chemistry is only cytotoxic for *S. aureus* BPL and has a >1100 fold selectivity over the human BPL, holocarboxylase synthetase (HCS).

The class I biotin protein ligase in *Mycobacterium tuberculosis* lacks an N-terminal DNA binding domain and therefore only functions to transfer biotin to acetyl-CoA carboxylase (79, 84). The catalytic domain consists of 7 beta-strands packed against 5 alpha helices and similar to both the *E. coli* and *S. aureus* BPL's, the C-terminal domain has an SH3-like fold (79). Both the apo and bio-5'-AMP bound structures are monomeric as confirmed through both solution studies and x-ray crystallography (79, 84). In the apo structure, both the BBL (residues 63-77) and the ABL (residues 162-171) are partially disordered. However, in the bio-5'-AMP structure these loops are ordered forming a hydrophobic network over the adenylate, as seen in the previously mentioned class II BPL's from *E. coli* and *S. aureus*, and consist of residues A75, A76, A162, P163, V166, P168, and A170. Furthermore, the dehydrated apo crystal structure indicates that the ABL is folded in a similar orientation to that found in the bio-5'-AMP-bound structure (85). The exclusion of bulk solvent from the apo structure has shifted the conformation to that of the bound state. As seen in BirA, a disorder-to-order transition occurs upon bio-5'-AMP binding with the formation of a hydrophobic network (70, 72, 79).

1.4 The Experimental Problem

There is a fundamental lack of understanding about the sequence-function relationships for protein segments that undergo disorder-to-order transitions and their mechanism of achieving allosteric regulation. Furthermore, the determining factors of folding in ligand-linked disorder-to-order transitions are not clear. In this work, the *Escherichia coli* biotin repressor is used as a model system in which ligand-linked

folding and subsequent formation of the hydrophobic network are coupled to self-association, enhancing the dimerization free energy by -4.0 kcal/mol. Previous mutational and thermodynamic studies on the hydrophobic cluster formed by residues V214, V219, and W223 of the ABL have shown that replacement of either residue in the cluster significantly compromises corepressor binding, loop folding, and allosteric communication.

Chapter 2 focuses on further investigating the sequence determinants of the extended hydrophobic network formed in the ligand-linked transition for multiple BirA functions. This was investigated through the combination of thermodynamic and kinetic measurements performed on singly and multiply alanine-substituted variants in the network. The results from these experiments display large perturbations to ligand binding and synthesis along with loop folding. Additionally, the functional effects of multiply substituted variants display non-additivity. The results from this study are analogous to those seen in protein folding studies for which full function requires optimal packing of hydrophobic residues.

Chapter 3 describes the allosteric communication in BirA involving two distal disorder-to-order transitions. While disorder-to-order transitions are used by several proteins in allosteric communication, limited experimental data exists on the mechanism of communication. The replacement of glycine 142 with alanine in the dimerization surface loop results in perturbed dimerization energetics of 4 kcal/mol consistent with the abolishment of coupling between ligand-binding and dimerization.

In this chapter, the x-ray crystallographic structure of G142A is described. The structural data in combination with thermodynamic measurements support a mechanism in which allosteric communication is achieved through the coupling of disorder-to-order transitions at distant functional sites on the protein.

Chapter 4 discusses the consequences of replacing H_2O with D_2O on the self-association reaction of BirA. Protein-protein interactions are integral to many biological processes. Nonetheless, there are few studies that detail the contribution of solvent reorganization on association energetics. The self-association of BirA is characterized by a free energy of -7.0 kcal/mol (66). Sedimentation equilibrium was used to monitor the effect of deuterium on the association of wild type BirA along with the surface loop variants T195A and V219A. In each case, the homodimerization free energy was more favorable by approximately -1.5 kcal/mol. Furthermore, the temperature dependence of wild type self-association reaction in deuterated buffer reveals large opposing enthalpic and entropic contributions for the process. The results are consistent with solvent release contributing significantly to the self-association energetics of BirA. Furthermore, these results highlight the need to determine the functional effects of D_2O on biological systems prior to performing any analyses in deuterated medium.

Together, the results presented in this dissertation answer questions about the relationships between sequence, function, and folding of the ligand-linked disorder-to-order transitions in BirA. More specifically, the results demonstrate how the

optimal packing of residues involved in the transition is necessary to achieve full function. Furthermore, evidence for a mechanism of allosteric communication is presented that involves reciprocal disorder-to-order transitions. In combination, these results further our understanding of the complex relationship between sequence and function of ligand-linked disorder-to-order transitions.

Chapter 2: Sequence-Function Relationships in Folding upon Binding

The work in this chapter has been published: Eginton, C., Naganathan, S., Beckett, D. (2014). Sequence-function relationship in a ligand-linked disorder-to-order transition. *Prot. Sci.*

2.1 Abstract

Folding coupled to binding is ubiquitous in biology. Nevertheless, the relationship of sequence to function for protein segments that undergo coupled binding and folding remains to be determined. Specifically, it is not known if the well-established rules that govern protein folding and stability are relevant to ligand-linked folding transitions. Upon small ligand biotinoyl-5'-AMP (bio-5'-AMP) binding the *Escherichia coli* protein BirA undergoes a disorder-to-order transition that results in formation of a network of packed hydrophobic side chains. Ligand binding is also allosterically coupled to protein association, with bio-5'-AMP binding enhancing the dimerization free energy by -4.0 kcal/mol. Previous studies indicated that single alanine replacements in a three residue hydrophobic cluster that contributes to the larger network disrupt cluster formation, ligand binding and allosteric activation of protein association. In this work combined equilibrium and kinetic measurements of BirA variants with alanine substitutions in the entire hydrophobic network reveal large functional perturbations resulting from any single substitution and highly non-additive effects of multiple substitutions. These

substitutions also disrupt ligand-linked folding. The combined results suggest that, analogous to protein folding, functional disorder-to-order linked to binding requires optimal packing of the relevant hydrophobic side chains that contribute to the transition. The potential for many combinations of residues to satisfy this requirement implies that, although functionally important, segments of homologous proteins that undergo folding linked to binding can exhibit sequence divergence.

2.2 Introduction

Coupled binding and folding contributes to the function of many biological macromolecules (3, 9, 86). This phenomenon has gained considerable attention with the discovery of intrinsically disordered proteins (IDPs) that fold upon association with a binding partner (24, 87, 88). Furthermore, disorder-to-order transitions can play critical roles in allosteric regulation (10, 89). Despite their involvement in many cellular processes, little is known about the sequence-function relationship for protein segments that undergo these transitions.

The *Escherichia coli* biotin repressor, BirA, provides a model system to investigate the role of sequence in a coupled ligand-binding and folding process. BirA, a bifunctional protein that is the central component of the Biotin Regulatory System (Fig. 9), functions as both a metabolic enzyme, an activity that is required for viability, and a transcription repressor of the biotin biosynthetic operon (90-92). For both functions, apoBirA binds to biotin followed by ATP to synthesize the corepressor, bio-5'-AMP (93). The resulting holoBirA heterodimerizes with the

biotin carboxyl carrier protein, BCCP, subunit of acetyl-CoA carboxylase, to catalyze post-translational biotin addition (94, 95) or homodimerizes and binds sequence-specifically to the biotin operator, bioO, to repress transcription of the biotin biosynthetic operon (73, 96). Although both dimerization reactions are energetically coupled to bio-5'-AMP binding, the coupling free energy of -4.0 kcal/mol is known only for the homodimerization reaction (66). High-resolution structures of apoBirA (1BIA) and complexes of BirA bound to biotin (1HXD) and a bio-5`-AMP analog, btnOH-AMP (2EWN), reveal disorder-to-order transitions in corepressor binding and allosteric activation (70-72). In apoBirA, two protein segments comprised of residues 116-124 and 206-234, referred to as the biotin binding loop (BBL) and adenylate binding loop (ABL), respectively, are disordered. Although biotin binding is coupled to BBL folding, this folding does not promote homodimerization (67, 73). In the adenylate-bound structure a network of hydrophobic residues from both the ABL and BBL is assembled around the ligand. The network includes a cluster formed by the side chains of residues V214, V219, and W223 of the ABL that assembles over the adenine ring and the side chains of residues F124 and P126 and M211 and V218 that bridge the BBL and ABL (Fig. 9). Charged and polar ABL side chains project away from the protein surface.

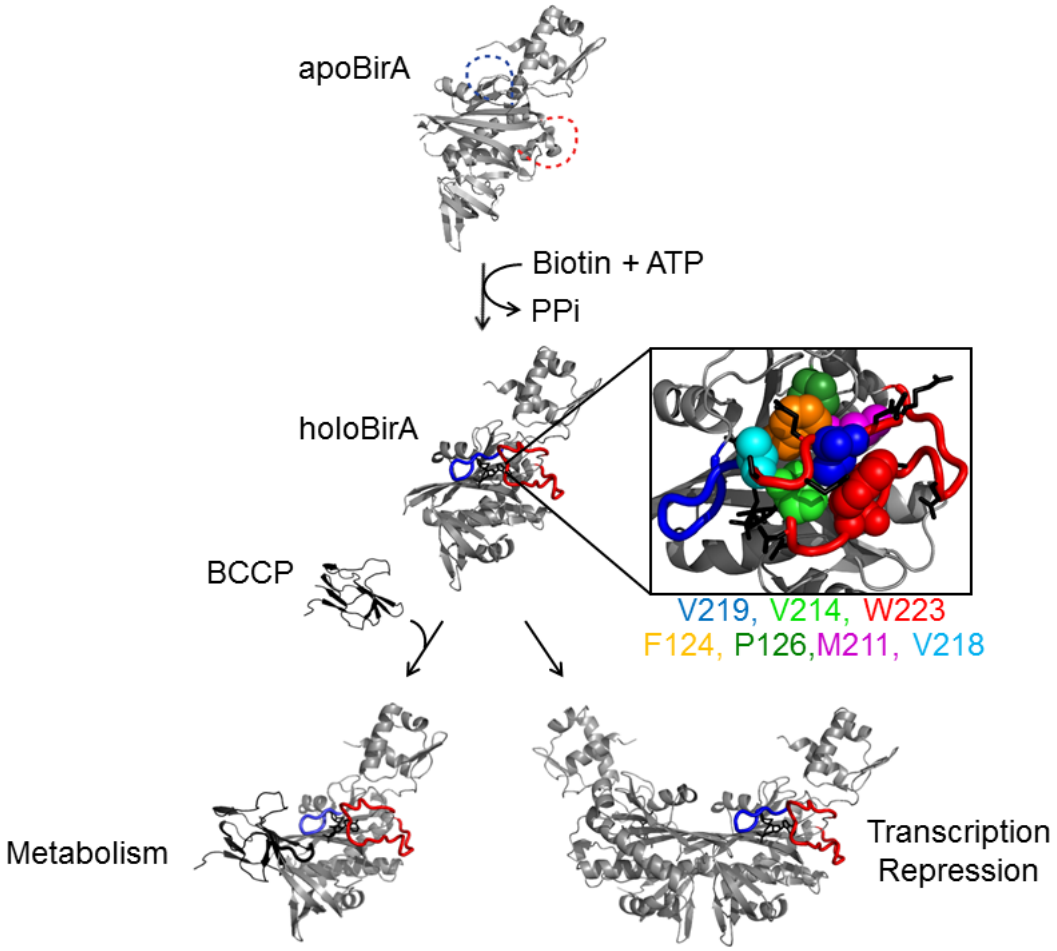

Figure 9. The *E. coli* Biotin Regulatory System illustrating the multiple functions of BirA (see text). Protein segments including the BBL (blue), residues 116-128, and the ABL (red), residues 210-234, are disordered (dashed line) in apoBirA and ordered in holoBirA (70, 72). The boxed figure illustrates the hydrophobic network that assembles upon bio-5'-AMP binding. Color scheme: ABL cluster: V214 (light green), V219 (blue), W223 (red); ABL-BBL bridge: F124 (orange), P126 (dark green), M211 (magenta), V218 (cyan); The charged and polar side chains, which project away from the protein surface, are shown as black sticks. The models were generated using PyMOL (21) with pdb files 1BIA and 2EWN as input.

E. coli BirA is a member of a large enzyme family and alignment of BBL and ABL sequences of *E. coli*, *M. tuberculosis* and *S. aureus* indicates little conservation at the hydrophobic network positions in *E.coli* BirA (Fig. 10A). Moreover, this lack of sequence conservation extends to other bacterial ligases as well as eukaryotic orthologs (Fig. 11). The sequence divergence is difficult to reconcile with the demonstrated importance of the ABL cluster sequence for bio-5'-AMP binding and coupled dimerization observed for the *E. coli* enzyme. However, the high-resolution structures of the *M.tb* (4OP0) and *S.au* (3V8L) enzymes bound to bio-5'-AMP reveal, as observed for *E.coli* BirA, a network of hydrophobic side chains packed around the adenylate ligand (Fig. 10B) (72, 78, 79). The sequence divergence and structural data suggest that, reminiscent of the hydrophobic cores that form upon protein folding, the function of a sequence in folding upon binding in biotin ligases may depend solely on the packing ability of the participating hydrophobic side chains. Protein folding studies have led to the general conclusion that a given fold can tolerate sequence variations provided that hydrophobic core side chain packing is retained (97-101). Moreover, packing between side chains in protein folding can be readily detected by non-additive effects of multiple amino acid substitutions on the folding free energy (102, 103).

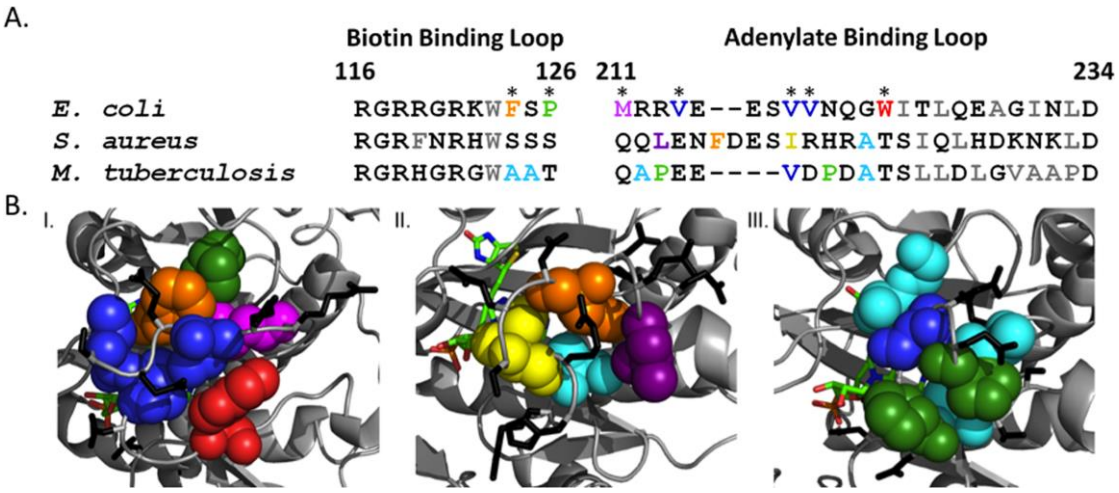

Figure 10: Sequence divergence for hydrophobic residues that assemble around the adenylate ligand in biotin ligase homologs. A. Alignment of the BBL and ABL segments of the *Escherichia coli*, *Mycobacterium tuberculosis* and *Staphylococcus aureus*. The numbering system is from the *E. coli* protein. Color code: polar or charged residues (black), hydrophobic residues (gray), hydrophobic residues that assemble around the adenylate ligand (colored). B. Hydrophobic packing around the adenylate ligand of the I. *E.coli* (2EWN) II. *S.a.* (3V8L) and M.*tb* (4OP0) biotin protein ligases. The color code for the amino acids is identical to that used in A. Polar and charged side chains are shown as black sticks and the adenylate ligand as colored sticks. The alignments were extracted from alignments of the entire sequences of the three proteins using ClustalW (104, 105) and the protein structure figures were prepared using PyMOL (21).

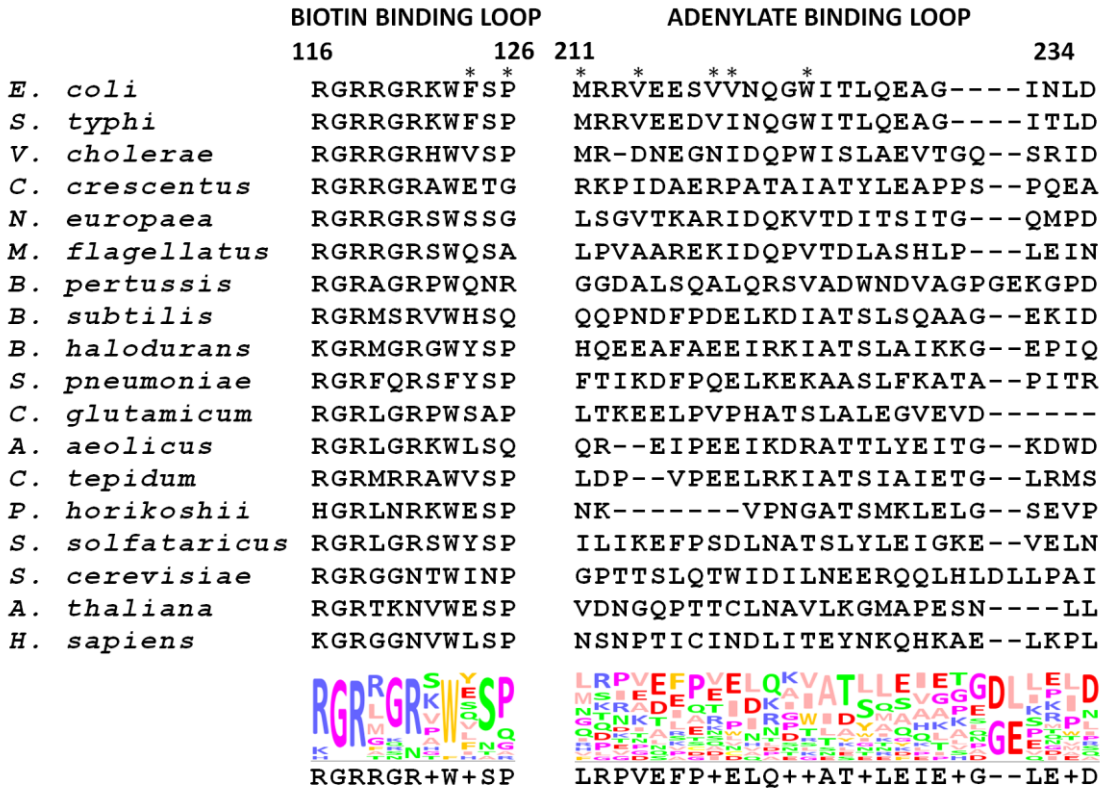

Figure 11: Multiple sequence alignment of biotin protein ligases. Sequence alignments of both the biotin binding loop (BBL) and adenylate binding loop (ABL) were extracted from alignment of the entire sequences proteins performed using ClustalW (104, 105). Sequence conservation analysis at each position displayed in the logo was obtained and visualized using Jalview (106). The conserved region of the BBL containing glycine and arginine residues is utilized in biotin binding (75).

In this work, the sequence-function relationships in BirA ligand-linked folding was investigated by measuring the consequences of introducing single and multiple alanine substitutions into the ABL cluster and bridging residues of the protein. Combined isothermal titration calorimetry (ITC) measurements of bio-5'-AMP binding and stopped- flow kinetic measurements of its synthesis reveal that any single alanine substitution results in large perturbations to adenylate binding and synthesis. Furthermore, effects of multiple substitutions are highly non-additive. Subtilisin-mediated proteolysis indicates that for the majority of the variant proteins the functional defects are accompanied by perturbation to ABL cluster folding. Consequences of the sequence changes for ligand-linked homodimerization, while relatively modest, are also non-additive for proteins with multiple alanine substitutions. These results indicate that, analogous to protein folding, full function of protein segments in binding-linked disorder-to-order transitions in BirA requires appropriate packing of the participating hydrophobic side chains. The results have implications for the evolution of sequences of protein segments that undergo folding coupled to binding.

2.3 Results

2.3.1 Alanine substitutions yield large non-additive effects on adenylate binding

Previous measurements indicate that any single alanine substitution in the ABL hydrophobic cluster significantly perturbs bio-5'-AMP binding while leaving biotin binding intact (77, 107). These measurements were extended to proteins with

alanine substitutions at residues that bridge the ABL and BBL and to variants with multiple alanine substitutions. Alanine substitutions at bridging residues yield large perturbations to bio-5'-AMP binding. In contrast the adenylate binding to wt BirA, which, due to its high affinity, requires use of the ITC displacement titration method (107), direct titrations sufficed for measuring bio-5'-AMP binding to the variants (Fig. 12A). All titrations were performed at a total protein concentration of 2 μM, at which, even in the presence of saturating bio-5'-AMP, the major species is the monomer. For example, the variant that dimerizes most tightly, BirA F124A, is 80% monomer in its adenylate-bound form (Table 4). The bio-5'-AMP binding isotherm for BirAV218A is well-described by a simple binding model and, consistent with 1:1 binding, data analysis yields an n-value of 0.98±0.06 (Fig. 12A, Table 1). Moreover, relative to wild type BirA, the equilibrium dissociation constant and the Gibbs free energy of adenylate binding are perturbed by 450-fold and 3.5 kcal/mol, respectively. Alanine substitutions at the remaining bridge residues, P126, M211, and F124, resulted in changes to adenylate binding similar in magnitude to that measured for the V218A variant (Table 1). By contrast, alanine substitutions of the charged ABL residues, R213 and E216, have minimal effects on bio-5'-AMP binding (Table 1). Finally, consistent with previous studies of alanine substitutions in the ABL cluster residues, with the exception of the P126A variant, substitutions of ABL and bridge residues do not alter biotin binding (Table 2) (77).

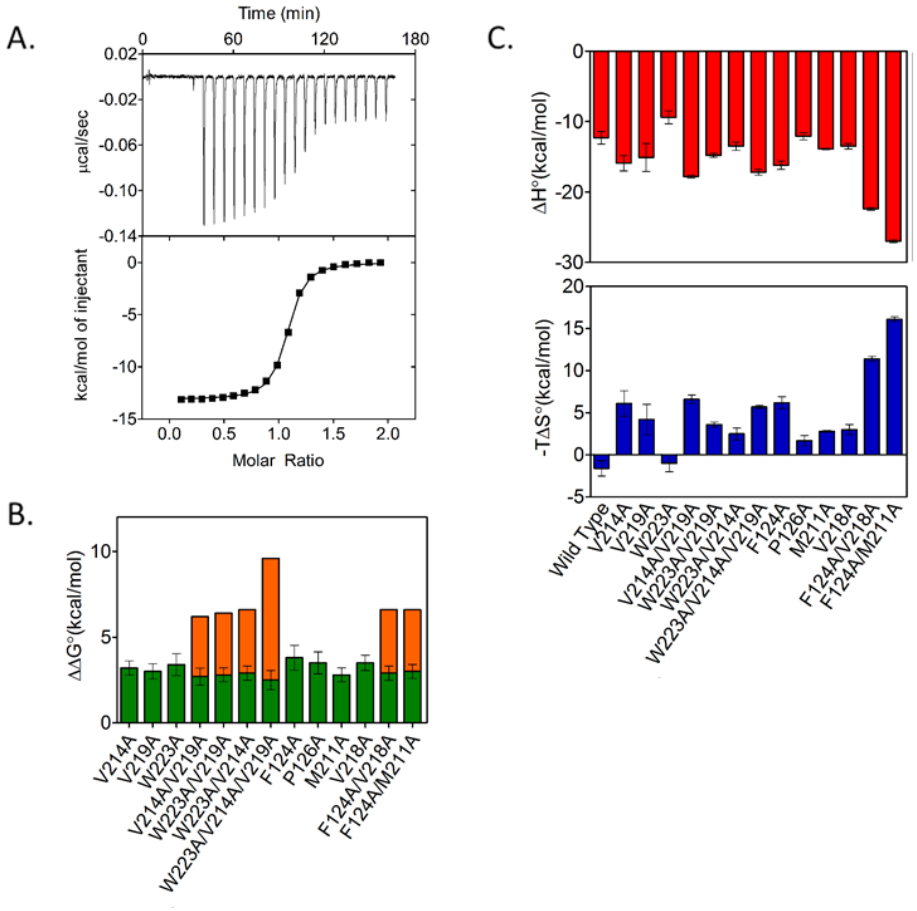

Figure 12. Alanine substitutions result in large, non-additive effects on bio-5'-AMP binding. A. Top panel: ITC trace for V218A BirA with bio-5'-AMP: 19 13 uL volumes of 20 µM bio-5'-AMP were injected into 2 µM V218A. Bottom panel: Nonlinear regression of the binding isotherm obtained from A using a single site binding model. B. Perturbations to the Gibbs free energy of bio-5'-AMP binding to the alanine substituted BirA variants shown as green bars with the expected additive effects of the multiple substitutions shown as the sum of the green plus orange bars. C. Enthalpic (ΔH^o, red) and entropic ($-T\Delta S^o$, blue) contributions to bio-5'-AMP binding for the BirA variants. The error bars were obtained from standard propagation of errors associated with at least two independent measurements.

Table 1: Thermodynamics of bio-5'-AMP binding obtained from isothermal titration calorimetry.

Protein	K_D[a]	ΔG° (kcal/mol)[b,c]	ΔH° (kcal/mol)[a]	–TΔS° (kcal/mol)[b]	n[a]
Wild type[e]	$3.9(\pm1.9)\times10^{-11}$	–13.9±0.4	–12.3±0.9	–1.6±0.9	0.91±0.01
R213A[d]	$1.4(\pm0.1)\times10^{-10}$	–13.21±0.04	–10.16±0.01	–3.05±0.04	0.96±0.17
E216A[d]	$2.1(\pm0.2)\times10^{-10}$	–12.96±0.04	–11.1±0.5	–1.8±0.5	0.84±0.01
V214A[e]	$9.7(\pm1.6)\times10^{-9}$	–10.7±0.1	–15.9±1.1	6.1±1.5	0.88±0.01
V219A[e]	$1.4(\pm0.9)\times10^{-8}$	–10.9±0.2	–15.1±2.0	4.2±1.8	0.90±0.01
W223A[e]	$6.9(\pm2.0)\times10^{-9}$	–10.5±0.5	–9.4±0.9	–1.0±1.0	0.93±0.01
V214A/V219A	$4.8(\pm0.1)\times10^{-9}$	–11.2±0.3	–17.8±0.2	6.6±0.5	0.81±0.01
W223A/V219A	$5.1(\pm2.6)\times10^{-9}$	–11.1±0.1	–14.8±0.3	3.6±0.3	0.88±0.03
W223A/V214A	$6.3(\pm0.8)\times10^{-9}$	–11.0±0.1	–13.5±0.6	2.5±0.7	0.79±0.01
W223A/V214A/V219A	$3.1(\pm0.9)\times10^{-9}$	–11.4±0.2	–17.2±0.4	5.7±0.2	0.84±0.06
F124A	$3.2(\pm0.5)\times10^{-8}$	–10.1±0.1	–16.2±0.6	6.2±0.7	1.00±0.01
P126A	$1.9(\pm0.4)\times10^{-8}$	–10.4±0.1	–12.1±0.5	1.7±0.6	0.96±0.04
M211A	$5.1(\pm0.6)\times10^{-9}$	–11.1±0.1	–13.9±0.1	2.8±0.1	1.00±0.02
V218A	$1.7(\pm0.6)\times10^{-8}$	–10.4±0.2	–13.5±0.4	3.0±0.6	0.98±0.06
F124A/V218A	$5.8(\pm0.8)\times10^{-9}$	–11.0±0.1	–22.4±0.2	11.4±0.3	0.86±0.01
F124A/M211A	$7.5(\pm1.1)\times10^{-9}$	–10.9±0.1	–27.0±0.2	16.1±0.3	0.85±0.01

All measurements were performed at 20°C in standard buffer (10 mM Tris HCl, 200 mM KCl, 2.5 mM $MgCl_2$, pH 7.5 at 20°C).

[a] Reported values, which are the average of at least two independent measurements, were obtained from the best-fit curves of non-linear least squares analysis using a single-site binding model in Origin 7.0 (MicroCal).

[b] Gibbs free energy values, ΔG°, were obtained using the relationship $\Delta G^\circ = -RT\ln(1/K_D)$. Values of $-T\Delta S^\circ$ were calculated using the relationship $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$.

[c] The reported uncertainties were calculated by propagating the confidence intervals of the independent measurements used to calculate the average reported value. Comparison of wild type with variant binding free energies using the unpaired t-test yielded P values<0.05.

[d] Measured using the displacement method as described in Materials & Methods

[e] Previously published results (77, 107).

Table 2: Biotin binding thermodynamics obtained from isothermal titration calorimetry.

Protein	K_D (M) [a]	ΔG° (kcal/mol)[b,c]	ΔH° (kcal/mol)[a]	−TΔS° (kcal/mol)[b]	n[a]
Wild type[d]	$4.3(\pm0.3)\times10^{-8}$	–9.84±0.04	–19.4±0.2	9.4±0.2	0.91±0.01
R213A	$4.1(\pm0.6)\times10^{-8}$	–9.9±0.1	–19.1±0.5	9.2±0.6	0.88±0.04
E216A	$4.2(\pm0.2)\times10^{-8}$	–9.9±0.1	–19.6±0.3	9.7±0.3	0.78±0.01
V214A[d]	$3.8(\pm0.8)\times10^{-8}$	–9.9±0.1	–19.5±0.2	9.6±0.2	0.85±0.01
V219A[d]	$3.8(\pm1.0)\times10^{-8}$	–10.0±0.2	–19.4±0.5	9.3±0.2	0.85±0.01
W223A[d]	$4.7(\pm0.4)\times10^{-8}$	–9.8±0.1	–19.6±0.2	9.8±0.2	0.83±0.01
V214A/V219A	$5.3(\pm0.6)\times10^{-8}$	–9.7±0.1	–20.1±0.1	10.4±0.1	0.89±0.02
W223A/V219A	$4.1(\pm0.5)\times10^{-8}$	–9.9±0.1	–20.4±0.2	10.4±0.1	0.85±0.02
W223A/V214A	$4.9(\pm0.7)\times10^{-8}$	–9.8±0.1	–20.6±0.1	10.8±0.2	0.84±0.03
W223A/V214A/V219A	$4.1(\pm0.5)\times10^{-8}$	–9.9±0.1	–18.6±0.3	8.8±0.2	0.92±0.01
F124A	$3.2(\pm0.2)\times10^{-8}$	–10.04±0.04	-22.05±0.03	11.98±0.03	0.89±0.01
P126A	$3.9(\pm0.1)\times10^{-7}$	–8.58±0.02	–16.3±0.1	7.7±0.1	0.81±0.01
M211A	$4.5(\pm0.2)\times10^{-8}$	–9.8±0.1	–20.2±0.7	10.3±0.7	0.89±0.01
V218A	$3.2(\pm0.4)\times10^{-8}$	–10.0±0.1	–20.4±0.6	10.3±0.7	0.87±0.01
F124A/V218A	$4.0(\pm0.6)\times10^{-8}$	–9.9±0.1	–19.7±0.4	9.8±0.5	0.96±0.04
F124A/M211A	$4.1(\pm0.3)\times10^{-8}$	–9.9±0.1	–20.2±0.1	10.3±0.1	0.93±0.02

All measurements were performed at 20°C in standard buffer (10 mM Tris HCl, 200 mM KCl, 2.5 mM MgCl2, pH 7.5 at 20°C).

[a] Reported values are the average of at least two independent measurements obtained from the best-fit curves of non-linear least squares analysis using a single-site binding model in Origin 7.0 (MicroCal).

[b] Gibbs free energies, ΔG°, were obtained using the relationship $\Delta G° = RT\ln K_b$ and -TΔS° values were calculated using the relationship $\Delta G° = \Delta H° - T\Delta S°$.

[c] The reported uncertainties were calculated by propagating the confidence intervals of the independent measurements used to calculate the average reported value. Application of the t-test to compare the Gibbs free energies of variant and wt BirA binding to biotin indicates no significant differences, $P>0.05$, with the exception of P126A.

[d] Previously published results (77).

Adenylate binding measurements performed on variants with multiple alanine substitutions reveal functional coupling between hydrophobic side chains that participate in the adenylate-linked disorder-to-order transition. Consistent with coupling, the free energy of adenylate binding to each multiply substituted variant is similar to that measured for each singly substituted parent (Fig. 12B). The interaction energy between two residues in bio-5'-AMP binding can be calculated from the measured binding free energies for the wild type, each singly substituted protein and the doubly substituted protein. For example, interaction between V219 and W223 is calculated using the following expression:

$$\Delta G^{o}_{\mathrm{int},V219/W223} = \Delta G^{o}_{wt} - \Delta G^{o}_{V219A} - \Delta G^{o}_{W223A} + \Delta G^{o}_{V219A/W223A}$$

in which $\Delta G^{o}_{int,V219/W223}$ represents interaction energy between the V219 and W223 in adenylate binding and ΔG^{o}_{wt} , ΔG^{o}_{V219A} , ΔG^{o}_{W223A} and $\Delta G^{o}_{V219A/W223A}$ are the Gibbs free energies for binding of the wt, V219A, W223A and V219A/W223A variants to bio-5'-AMP. For the doubly substituted variants the coupling free energies range from -3 to -4 kcal/mol, while for the triply substituted variant, W223A/V219A/V214A, it is -7 kcal/mol. The absolute values of these coupling free energies are illustrated by the gray bars in Figure 12B.

The partitioning of energetic perturbations to adenylate binding for the variants into enthalpic and entropic contributions is complex. However, for a majority of the variants the enthalpic contribution to adenylate binding is more favorable than

that measured for the wild type protein and the entropic contribution less favorable (Fig. 12C). Furthermore, with the exception of the binding enthalpy measured for the V214A/W223A doubly substituted variant, the enthalpic and entropic perturbations to adenylate binding are non-additive for multiply substituted variants.

2.3.2 Alanine substitutions compromise ligand-linked ABL folding

Previous studies indicated that alanine substitutions in the BirA ABL hydrophobic cluster are accompanied by perturbations to ligand-linked ABL folding. Measurements of subtilisin-catalyzed BirA proteolysis, which initially occurs between residues 217 and 218 of the ABL (69, 77), were used to determine if this is also the case for variants with substitutions in the bridging residues as well as for those with multiple alanine substitutions. BirA, variant or wild type, in the presence or absence of saturating bio-5'-AMP was combined with subtilisin and the resulting time-dependent decrease in the amount of intact protein was obtained from densitometric analysis of samples subjected to SDS-PAGE. For each time course analysis of the optical density versus time data using a single exponential model yields the rate of loss of intact protein (Fig. 13A). The ratio of the rates measured in the absence and presence of bio-5'-AMP, k_{apo}/k_{holo}, for each protein indicates the magnitude of protection from digestion afforded by adenylate binding, which provides a measure of ABL folding upon binding (77). For the majority of the variants the k_{apo}/k_{holo} value is significantly decreased relative to that obtained for wild type BirA (Fig. 13B). Furthermore, alanine substitution at M211 and V218, which do not contribute directly to the ABL cluster, alter the ratio. Finally, consistent with the

observed functional non-additivity, multiple alanine substitutions do not disrupt ABL folding any further than single substitutions. Indeed, the variant with the combined F124A (k_{apo}/k_{holo}=7.0±0.2) and M211A (k_{apo}/k_{holo}=2.9±0.1) substitutions has a k_{apo}/k_{holo} value similar to that measured for wild type BirA.

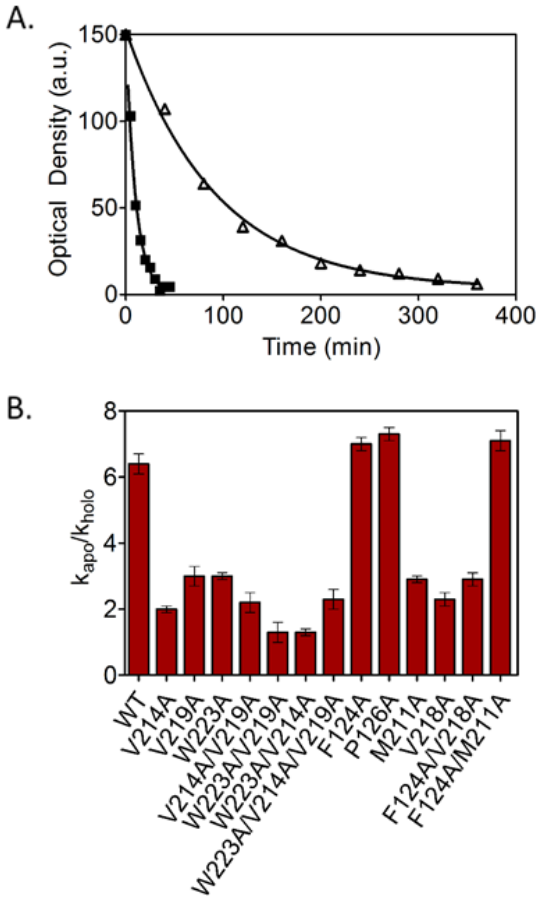

Figure 13: Ligand-linked folding is perturbed for the majority of the BirA variants. A. Time courses of subtilisin-catalyzed cleavage of P126A BirA measured in the absence (■) and presence (Δ) of saturating bio-5'-AMP. The lines represent the best-fit of the data to a single exponential model. B. Ratio of the pseudo first-order rates of subtilisin-mediated cleavage (red) in the absence, k_{apo}, and presence, k_{holo}, of saturating bio-5'-AMP.

2.3.3 Alanine substitutions perturb BirA-catalyzed bio-5'-AMP synthesis

Since BirA catalyzes adenylate synthesis from biotin and ATP, the observed perturbations to bio-5'-AMP binding and loop folding for the alanine substituted proteins suggested that they may also be altered in catalyzing its synthesis. The time-dependence of adenylate synthesis is measured by monitoring the decrease in intrinsic BirA fluorescence that occurs upon rapid mixing of a solution containing the BirA·biotin complex with excess ATP. The resulting transient is analyzed using a single exponential model to obtain the apparent rate of adenylate synthesis (Fig. 14) (93). For all variants, the measurements were performed at a single biotin and ATP concentration, with biotin present at sub-saturating but stoichiometric concentration and ATP at 1 mM, close to the K_m value measured for wild type BirA (75, 93). The measurements reveal that bio-5'-AMP synthesis is compromised for all variants (Table 3) with rate decreases ranging from 2-fold for V214A to 190-fold for the F124A/M211A double variant. For variants W223A and V219A/V214A, which catalyze adenylate synthesis too slowly for detection in stopped-flow measurements, synthesis was indirectly confirmed by mass spectrometric detection of biotinylated apoBCCP following overnight incubation with enzyme, biotin and ATP (Data not shown). The multiply substituted variants V214A/W223A and V219A/W223A, and V219A/V214A/W223A exhibited no detectable bio-5'-AMP synthesis. While all BirA variants are compromised in bio-5'-AMP synthesis, the measured rates show no correlation with the bio-5'-AMP binding affinities.

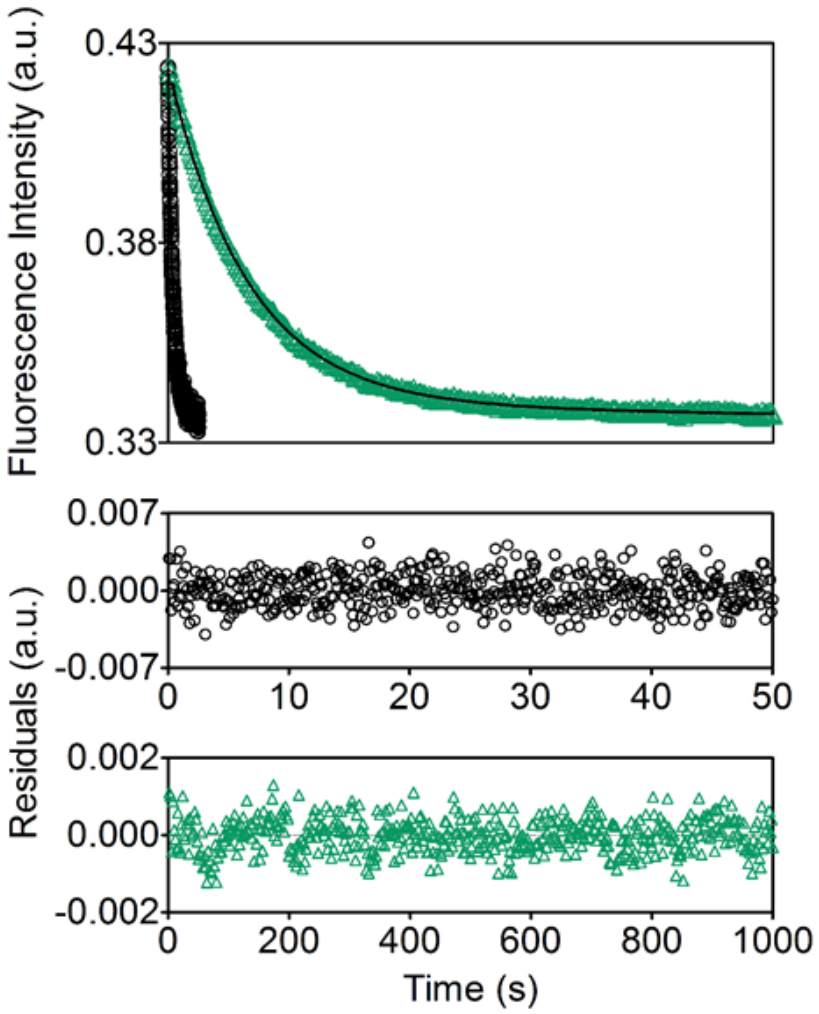

Figure 14: Alanine substituted variants are defective in catalyzing bio-5'-AMP synthesis. Kinetic transients showing the time-dependent decrease in the intrinsic protein fluorescence associated with bio-5'-AMP synthesis by wild type (**O**, black) and F124AV218A BirA (Δ, green) with the lines representing the best-fits of the data to a single exponential equation and the residuals of the fits provided in the bottom panels.

Table 3: Variants are defective in catalyzing bio-5'-AMP synthesis[a].

BirA Variants	Apparent Rate k_{app} (s^{-1})[b]	Fold Decrease[d]
Wild type	0.136±0.004	-
V214A	0.0537±0.0004	2.53±0.03
V219A[c]	0.0064±0.0003	21.25±0.06
F124A	0.0218±0.0004	6.24±0.03
P126A	0.052±0.001	2.62±0.04
M211A[c]	0.0043±0.0003	31.63±0.08
V218A	0.043±0.001	3.16±0.04
F124A/V218A	0.0081±0.0002	16.79±0.04
F124A/M211A[c]	0.00072±0.00002	188.88±0.04

[a] Measurements were performed in Standard Buffer at 20°C as described in Materials and Methods.

[b] The apparent rates are the average of the best fit rates obtained from nonlinear least squares analysis of at least five (stopped-flow) or three (fluorimeter) traces with the standard error of the mean. Comparison of wild type with variant rates using the unpaired t-test yielded P values<0.05. With the exception of those obtained for the wild type protein, all transients were analyzed using the model that included a linear term for photobleaching along with the single exponential.

[c] Measured using the fluorimeter.

[d] Fold decrease is the rate measured for wild type BirA divided by that measured for the variant.

2.3.4 Alanine substitutions yield modest but non-additive effects on homodimerization

Dimerization of BirA bound to bio-5'-AMP, holoBirA, is energetically more favorable than apoBirA dimerization by -4.0 kcal/mol and measurements performed on adenylate-bound variants with single alanine substitutions in the ABL cluster indicate penalties to this coupling ranging from 1 to 1.5 kcal/mole (Table 4) (66, 77). In this work the dimerization measurements were extended to proteins with single alanine substitutions in the BBL-ABL bridging residues and those with multiple alanine substitutions in the hydrophobic network.

Dimerization measurements were performed on the bio-5'-AMP bound variants using sedimentation equilibrium. Centrifugation of M211A variant was carried out at three speeds on protein prepared at three concentrations. Analysis of the data using a single species model indicated average molecular weights higher than expected for the monomeric protein. Global analysis using a monomer-dimer model, which exhibited excellent agreement between the best-fit curves and the data (Fig. 15A), indicated that the alanine replacement results in a 1.5 kcal/mol penalty to the dimerization free energy (Table 4). Alanine substitutions of the remaining bridging residues resulted in perturbations to dimerization energetics ranging from 0.5 to 1.0 kcal/mol.

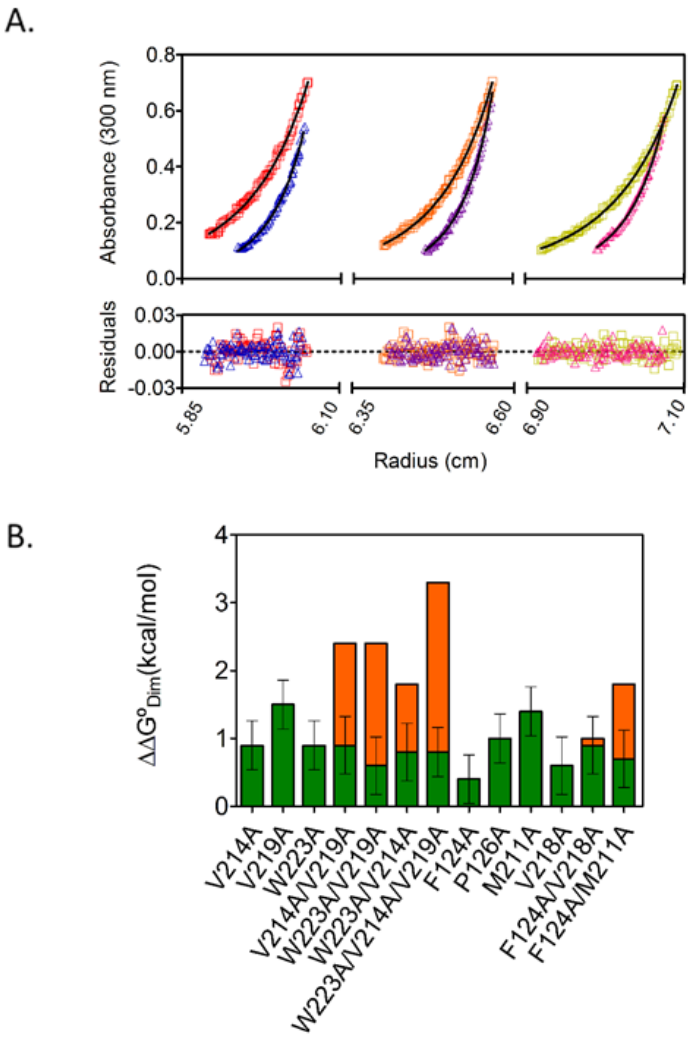

Figure 15: Effects of alanine substitutions on ligand-linked dimerization are modest but non-additive. A. Absorbance *versus* radius profiles for bio-5'-AMP bound M211A variant prepared at 40, 50, and 60 μM and centrifuged at 18,000 (□), and 24,000 (Δ) rpm with the best-fit curves obtained from the global analysis of six data sets to a monomer-dimer model using WinNonLin (108). In order to clearly see the best-fit curves, only 6 data sets are shown. The lower panel shows the residuals of the fits for each data set. B. Energetic penalties to holoBirA dimerization indicated (green) with anticipated additive effects of multiple alanine substitutions (orange).

Table 4: Dimerization energetics of BirA variants bound to bio-5'-AMP.

BirA Variant	K_{DIM} (M)[a]	ΔG^{o}_{DIM}(kcal/mol)[b]
WT[c]	$6(\pm2)x10^{-6}$	–7.0±(0.3)
V214A[c]	$3(\pm1)x10^{-5}$	–6.1±(0.2)
V219A[c]	$8(\pm1)x10^{-5}$	–5.5±(0.2)
W223A[c]	$3(\pm1)x10^{-5}$	–6.1±(0.2)
V214A/V219A	$3(\pm2)x10^{-5}$	–6.1±(0.3)
W223A/V219A	$2(\pm1)x10^{-5}$	–6.4±(0.3)
W223A/V214A	$2(\pm1)x10^{-5}$	–6.2±(0.3)
W223A/V214A/V219A	$2(\pm1)x10^{-5}$	–6.2±(0.3)
F124A	$1(\pm0.4)x10^{-5}$	–6.6±(0.2)
P126A	$3(\pm1)x10^{-5}$	–6.0±(0.2)
M211A	$7(\pm3)x10^{-5}$	–5.6±(0.2)
V218A	$2(\pm1)x10^{-5}$	–6.4±(0.3)
F124A/V218A	$3(\pm2)x10^{-5}$	–6.1±(0.3)
F124A/M211A	$2(\pm1)x10^{-5}$	–6.3±(0.3)

[a] Equilibrium dimerization constant are the average of at least two independent measurements performed as described in Materials & Methods. The errors were obtained by propagation the 65% confidence intervals associated with the equilibrium constant obtained from each individual measurement. In comparing values for the variants with wild type the P values obtained from the unpaired t-test are <0.05 for all variants with the exception of F124A/M211A.

[b] The Gibbs free energies were calculated using the equation $\Delta G^{o}_{DIM}=RTlnK_{DIM}$.

[C] Previously published results (77).

The effects on dimerization resulting from multiple alanine substitutions in the hydrophobic network were also measured. Data obtained on all proteins with two or three alanine substitutions indicated good agreement with a monomer-dimer model. The dimerization free energies for the variants indicate non-additive effects of the multiple substitutions (Fig. 15B). For multiple substitutions in the hydrophobic cluster the non-additivity is large, as indicated by the 1-3 kcal difference between the anticipated additive effects and the measured values of the dimerization free energies. For the bridging residues addition of the substitution at position 124 to that at 211 reverts the dimerization free energy of the latter protein to a value closer to that measured for wild type BirA. Only the combination of the V218A and F124A substitutions show additive effects on dimerization.

2.4 Discussion

The hydrophobic network in BirA that assembles concomitant with bio-5'-AMP binding provides an opportunity to examine the relationship between the sequence and function in a ligand-linked disorder-to-order transition. The measurements performed on the alanine-substituted BirA variants indicate that, reminiscent of the hydrophobic cores that assemble in protein folding, full function requires cooperative packing of all contributing side chains. These results imply that although the sequence of the network for a single biotin ligase from a particular species is, due to packing constraints, conserved, the network residues for ligases from distant homologs are free to diverge so long as packing can be achieved.

The large perturbations to adenylate binding resulting from single alanine substitutions in the hydrophobic network coupled with the non-additivity of multiple substitutions indicate that the disorder-to-order transition in BirA adheres to the rules that govern hydrophobic packing in protein folding. All variants with both single and multiple alanine substitutions exhibit similar large perturbations to adenylate binding free energy. Moreover, the majority of the variants, as judged by the loss of adenylate-mediated protection from subtilisin cleavage, are defective in adenylate binding loop folding. Even the variant with an alanine substitution at position M211, which does not form part of the ABL cluster, shows decreased ligand-mediated protection. These results are consistent with cooperative assembly of the hydrophobic network in which any single substitution affects the integrity of the entire network.

The perturbations to the detailed thermodynamics of adenylate binding to the alanine substituted variants are consistent with disruption of folding. All alanine substitutions in the network result in Gibbs free energies of bio-5'-AMP binding similar in magnitude to the measured free energy of biotin binding by wild type BirA, which is accompanied by BBL folding only. Moreover, the patterns of enthalpic and entropic contributions to adenylate binding by the variants are more similar to those measured for wtBirA binding to biotin, not bio-5'-AMP. Wild type BirA binds to biotin with a relatively modest Gibbs free energy of -10 kcal/mol, a highly favorable enthalpy of -20 kcal/mol and an unfavorable entropic contribution to binding at 20^{o}C of +10 kcal/mol. By contrast adenylate binding occurs with a more favorable Gibbs free energy of -14 kcal/mol but a less favorable -12 kcal/mol enthalpy and more

favorable -1.6 kcal/mol entropy at 20ºC. The differences in the energetic signatures for binding of the two ligands is consistent with the additional folding that accompanies adenylate binding with the less favorable enthalpy reflecting dehydration of the hydrophobic side chains that form the network and the more favorable entropy release of the water to the bulk. With the exception of W223ABirA, the thermodynamic signatures of adenylate binding to the variants, with their more favorable enthalpies and unfavorable entropies, are more consistent with wtBirA binding to biotin than to bio-5'-AMP.

In contrast to adenylate binding, the impacts of single alanine substitutions on homodimerization are relatively small, ranging from 0.5 to 1.5 kcal/mol penalties to the Gibbs free energy for the process. Moreover, unlike the effects on bio-5'-AMP binding which are consistently in the +4 kcal/mol range, the perturbations to dimerization vary for the proteins studied. In general, multiple alanine substitutions show non-additive effects on dimerization, with the V218A/F124A double variant the only exception. Given the distance of the hydrophobic cluster from the dimerization surface it is reasonable to assume that all of the measured energetic penalties reflect perturbations to the coupling free energy between corepressor binding and dimerization. The modest effects of alanine substitutions on dimerization reflects the fact that (Fig. 9), although optimal packing of the hydrophobic network is required for achieving the full -4.0 kcal/mol coupling between ligand binding and dimerization, the coupling free energy provides only a fraction of the total free energy of holoBirA dimerization.

The results of these studies are consistent the lack of sequence conservation for residues that assemble around the adenylate ligand in BirA homologs (Fig. 10). Like the amino acid residues that form the hydrophobic cores in protein folding, the requirement for optimal packing provides the major selective pressure for these segments (101). In principle, a number of combinations of hydrophobic side chains can pack sufficiently well to support function. Nonetheless, the lack of sequence conservation in the cluster residues does not imply for any single ligase that the sequence of this region is unimportant. Indeed, kinetic measurements reveal that even single alanine substitutions in the hydrophobic network of *E.coli* BirA can have large consequences for bio-5'-AMP synthesis (Table 3). The requirement of biotin transfer to BCCP for viability predicts that for a ligase from a single species the identity of the hydrophobic network sequence is highly constrained. This prediction was born out in the results of genetic screens for BirA mutants, which yielded no mutations in the coding sequences for the hydrophobic network residues (90, 109). The requirement of biotin transfer for viability precluded survival mutants belonging to this class.

The results of this work indicate that the sequences of regions that undergo ligand-linked disorder-to-order transitions in BirA are, like the hydrophobic cores in protein folding, constrained by the requirement for optimal packing. This result predicts that although proteins segments that fold upon binding may be of great functional importance, their sequences need not be evolutionarily conserved.

2.5 Material and Methods

2.5.1 Chemicals and biochemicals

All chemicals were at least reagent grade. The d-biotin, ATP, isopropyl β-d-thiogalactoside (IPTG), phenylmethanesulfonyl fluoride (PMSF), and polyethyleneimine (PEI) were purchased from Sigma-Aldrich and 1,4-dithio-DL-threitol (DTT) was obtained from Research Organics. The bio-5'-AMP was synthesized and purified as previously described (58, 61) and d-biotin stock solutions were prepared by dissolving the dry powder in Standard Buffer (SB: 10 mM Tris HCl, 200 mM KCl, 2.5 mM $MgCl_2$, pH to 7.5 at 20°C). The ATP was dissolved in water, the solution pH was adjusted to 7.5, and after spectrophotometric determination of the concentration 250 μL aliquots were stored at -70°C.

2.5.2 Site-directed mutagenesis, expression, and purification of BirA variants

Site-directed mutagenesis of the C-terminally $(His)_6$-tagged BirA coding sequence in the pBtac2 vector was carried out using the QuikChange II XL Site-Directed Mutagenesis Kit (Stratagene) according to the manufacturer's instructions. The C-terminal $(His)_6$ tag has previously been shown to have no effect on BirA function (75). Sequences were verified by dideoxy sequencing at the University of Maryland College Park DNA Sequencing Facility and ACGT Inc. Each protein was expressed in *E. coli* JM109 cells transformed with the appropriate plasmid. Cultures were grown at 30°C in LB media supplemented with 100 μg/mL ampicillin to an $O.D._{600}$ of 0.6-0.9, at which point protein expression was induced by addition of IPTG

to a final concentration of 1 mM. After 4 hours the cells were harvested by centrifugation, lysed by sonication and the resulting cell debris was removed by centrifugation at 5,000 rpm at 4°C for 30 minutes. Polyethyleneamine (PEI) was added to the cell extract at a 0.2% v/v ratio and the precipitate was pelleted by centrifugation. After chromatography on Ni-NTA resin (Qiagen) and followed by SP Sepharose resin (77) (GE Healthcare) the pure protein was dialyzed against storage buffer (10 mM Tris HCl, 200 mM KCl, 2.5 mM $MgCl_2$, 5% (v/v) glycerol, pH 7.5 at 20°C) and stored in 1 mL aliquots at -70°C. Protein purity was estimated to be >95% based on coomassie brilliant blue staining of samples subjected to electrophoresis on SDS-polyacrylamide gels. The fractional activity in bio-5'-AMP binding was determined by stoichiometric titrations monitored by steady-state fluorescence spectroscopy (61).

2.5.3 Isothermal titration calorimetry (ITC)

BirA binding to biotin and bio-5'-AMP binding was measured by isothermal titration calorimetry (ITC) using a VP-ITC microcalorimeter (MicroCal Inc. Northampton, MA). Each protein was first exchanged into Standard Buffer (10 mM Tris HCl, pH 7.50±0.01 at 20°C, 200 mM KCl, 2.5 mM $MgCl_2$) either by dialysis or chromatography using a Micro Bio-Spin 6 column (Bio-Rad), filtered through a 0.45 µm PTFE membrane syringe filter (PALL), and subjected to spectrophotometric concentration determination. Final ligand and protein solutions were prepared by dilution of concentrated stocks into SB that had been filtered through a 0.22 µm PVDF membrane syringe filter (Millipore). All samples were degassed for at least 10

minutes at 20°C using a Thermovac (MicroCal Inc.) prior to sample cell and syringe loading. Titrations were performed at 20°C by injecting 19-25, 10-13 μL volumes of a 20 μM biotin or bio-5'-AMP solution into 1.4 mL of a 2 μM protein solution at a stirring rate of 310 rpm. Displacement titrations were performed by injecting 29-10 μL volumes of 20 μM bio-5'-AMP into a solution containing 2 μM BirA saturated with biotin.

2.5.4 Sedimentation equilibrium

Protein dimerization was measured by equilibrium analytical ultracentrifugation using an Optima XL-I Analytical Ultracentrifuge (Beckman Coulter). Each protein was exchanged into SB by dialysis or chromatography on a Micro Bio-Spin 6 column (Bio-Rad). Samples were then filtered through 0.45 μm PTFE membrane syringe filters (PALL) before determining the concentration spectrophotometrically. Protein samples at three concentrations, each combined with bio-5'-AMP in a 1:1:15 molar ratio under stoichiometric conditions, were centrifuged in cells equipped with 12 mm six-hole or 3 mm two-hole charcoal-filled Epon centerpieces in a four-hole An-60 rotor (Beckman Coulter) at 3 speeds ranging from 18,000 to 24,000 rpm.

2.5.5 Kinetic measurements of bio-5'-AMP synthesis

Single turnover measurements of BirA-catalyzed bio-5'-AMP synthesis were carried out using either a Kintek SF-2001 stopped-flow instrument equipped with

fluorescence detection or an ISS-PC1 fluorimeter. For all measurements the excitation wavelength was set at 295 nm and emission was monitored above 340 nm using a cutoff filter (Corion Corp.) for the stopped-flow measurements and at 333 nm for measurements performed using the fluorimeter. The time-dependent decrease in the intrinsic protein fluorescence was monitored upon mixing BirA (wt or variant) combined with biotin in SB at a 1:0.9 molar ratio with ATP to achieve final BirA:biotin and nucleotide concentrations of 0.9 μM and 1 mM.

2.5.6 Subtilisin-catalyzed proteolysis

The rates of subtilisin-catalyzed cleavage of BirA (wt or variant) were determined at 20°C in SB in the presence and absence of saturating bio-5'-AMP. Solutions of protein or protein and ligand were first equilibrated for 30 minutes at 20˚C. Subtilisin prepared in SB was then added to the protein to achieve a final BirA:subtilisin (w:w) ratio of 41.5:1. Aliquots of 20 μL were removed from each reaction at regular time intervals and combined with 1 μL freshly prepared 100 mM PMSF in absolute ethanol and 12 μL of Laemmli sample loading buffer to stop proteolysis. Digestion products were then separated by electrophoresis in a 15% SDS-polyacrylamide gel. After staining with coomassie brilliant blue the optical density of bands corresponding the intact BirA were quantified using a Molecular Dynamics Laser Scanning Personal Densitometer (GE Healthcare). The rates of subtilisin-catalyzed cleavage of BirA were obtained by nonlinear least squares analysis using the following equation:

$$BirA_{OD,t} = A * e^{kt} + C \qquad (1)$$

in which $BirA_{OD,t}$ is the integrated optical density for the band corresponding to intact BirA at time t in minutes, and k is the pseudo-first order rate of cleavage.

2.5.7 Data analysis

ITC: Analysis of the bio-5'-AMP binding data was carried out using the following single-site binding model in Origin 7.0 (MicroCal):

$$Q = \frac{nM_t \Delta H V_o}{2}[1 + \frac{X_t}{nM_t} + \frac{1}{nKM_t} - ((1 + \frac{X_t}{nM_t} + \frac{1}{nKM_t})^2 - \frac{4X_t}{nM_t})^{1/2}] \quad (2)$$

in which Q is the total heat content of the solution contained in V_o, the active cell volume, ΔH is the molar binding enthalpy, K is the equilibrium association constant, n is the binding stoichiometry, and M_t and X_t are the bulk concentration of macromolecule and ligand respectively. Binding free energies and entropies were calculated using the relationships $\Delta G° = -RT\ln K$, $\Delta G° = \Delta H° - T\Delta S°$ and the parameters obtained from the analysis. For the displacement titrations the equilibrium constants and enthalpies obtained from the data analysis are apparent values, which can be related to the true values for bio-5'-AMP binding using the following equations:

$$K_{APP} = \frac{K_A}{1 + K_B[B]} \qquad (3)$$

$$\Delta H_{APP} = \Delta H_A^o - \Delta H_B^o \frac{K_B[B]}{1 + K_B[B]} \qquad (4)$$

in which the A and B subscripts are relevant to the parameters associated with bio-5'-AMP and biotin binding, respectively. Thus, the parameters associated with adenylate binding are obtained from the apparent values obtained from analysis of the displacement titrations and the values obtained for biotin binding in independent titrations (Table 2).

Sedimentation equilibrium: The absorbance versus radius profiles obtained for each scan were first analyzed in WinNonLin (108) using a single species model to obtain σ, the reduced molecular mass, from which the weight average molecular weight was calculated using the following equation:

$$\sigma = \frac{M(1-\bar{v}\rho)}{RT}\omega^2 \qquad (5)$$

where M is the molecular weight, $\bar{v}$ is the partial specific volume of the protein, ρ is the density of the buffer, ω is the angular velocity of the rotor, R is the gas constant and T is the temperature in Kelvin. Values of the BirA monomer partial specific volume and the buffer density are 0.755 mL/g and 1.007g/mL, respectively (73).

Absorbance versus radius profiles were also globally analyzed to obtain the equilibrium association constant for dimerization, K_a, using the following monomer-dimer model:

$$c_t(r) = \delta + c_m(r_o)e^{\sigma_m(\frac{r^2-r_o^2}{2})} + K_a(c_m(r_o))^2 e^{2^{\sigma_m(\frac{r^2-r_o^2}{2})}} \qquad (6)$$

in which C_t is the total concentration at each radial position r, δ is the baseline offset, $c_m(r_o)$ is the monomer concentration at reference radial position, r_o. The reduced molecular weight of the monomer, σ_{mon}, of 1.22 was previously determined by sedimentation equilibrium measurements performed on apoBirA. The quality of the fits was assessed from the magnitude of the square root of the variance and the distribution of the residuals of the fit about zero.

Kinetic measurements of bio-5'-AMP synthesis: Analysis of kinetic transients was carried out either using the software provided by KinTek Corporation or in Prism. Apparent rates of bio-5'-AMP synthesis were obtained by nonlinear regression of the fluorescence versus time data using a single exponential model. For the variants that catalyzed bio-5'-AMP with rates significantly slower than wild type BirA the time-dependent decrease in fluorescence included a contribution from photobleaching. In these cases the data were analyzed using a model that included a linear term for the photobleaching along with the single exponential decrease associated with adenylate synthesis. Inclusion of this additional phase yielded more random distribution of the residuals of the fits but did not significantly change the best-fit values of the apparent rates.

Acknowledgements

I thank Dr. Naganathan for her work on the hydrophobic cluster and the purification and characterization of the V214A, V219A, W223A, R213A, and E216A variants. I

also thank Colef Talbert for the purification and initial characterization of the W223A/V214A and W223A/V214A/V219A variants.

Chapter 3: Allosteric Communication *via* Distant Disorder-to-Order Transitions

3.1 Abstract

While disorder-to-order transitions are known to be utilized in allosteric regulation, limited experimental data exists on the mechanisms by which they function. The *Escherichia coli* biotin repressor provides a model system in which ligand-linked folding is allosterically coupled to self-association resulting in a -4.0 kcal/mol enhancement of homodimerization. Replacement of glycine 142 with alanine on the dimerization surface results in the complete abolishment of the coupling. In this work, the x-ray crystallographic structure of G142A variant bound to a corepressor analog provides insight into the structural origins of the loss of allosteric coupling. The structural data in combination with thermodynamic measurements, support a mechanism in which allosteric communication is achieved through the coupling of two distal disorder-to-order transitions.

3.2 Introduction

Regulation of protein function through allosteric control is pervasive in biological systems. Over the years, our view of allosteric mechanisms has shifted from rigid structural inter-conversion to a more dynamic view and more recently incorporates roles for protein segments that undergo ligand-linked folding/unfolding events (43). One of the most significant contributions to our understanding of

allosteric regulation has been the identification of several proteins, mainly transcription factors, which employ ligand-linked folding/unfolding events in allosteric regulation utilizing intrinsically unstructured regions (9, 50-53). These coupled binding and folding/unfolding events have experimentally been shown to be key for the regulation of gene expression in several systems and have been proposed to operate in many others. Further theoretical studies have also proposed that the coupling between distant sites can be optimized through disorder-to-order transitions (10). Despite the current experimental and theoretical observations in the literature there is still inadequate characterization and understanding of model systems utilizing disorder-to-order transitions in allosteric communication.

The *Escherichia coli* biotin repressor, BirA, is a bifunctional allosteric transcriptional regulatory protein and provides a model system to investigate allosteric communication involving disorder-to-order transitions (57). Binding of the corepressor, bio-5'-AMP, is coupled to BirA self-association, enhancing the dimerization free energy by -4.0 kcal/mol relative to the apo protein (Fig. 16A) (63). Allosteric communication involving disorder-to-order transitions was suggested from solution measurements and the high-resolution crystallographic structures of BirA in various liganded states (69-72). Comparison of the structure of apoBirA with that of BirA complexed with biotin and a bio-5'-AMP non-hydrolyzable analog, biotinol-5'-AMP, reveals loop folding events on the two functional surfaces. (Fig. 16B). On the ligand binding surface, two protein segments consisting of residues 116-124 and 211-234, referred to as the biotin binding loop (BBL) and adenylate binding loop (ABL),

are disordered in the apo protein but folded over the ligand in the btnOH-AMP structure. In the adenylate-bound structure the folding transition results in a hydrophobic network between the ABL and BBL that assembles around the adenylate. Previous studies have shown that single alanine substitutions in this cluster perturb the self-association of BirA by as much as 1.5 kcal/mol linking the disorder-to-order transition on the ligand binding surface to the allosteric response involving the dimerization surface (77). In addition to the ligand-linked folding on the ligand binding surface, three flexible loops on the dimerization surface undergo disorder-to-order transitions upon corepressor binding and dimerization. Residues in these loops have been systematically investigated through alanine substitutions to determine their importance in the self-association process (80). Upon substitution, large effects to the homodimerization free energy are observed ranging from -3 kcal/mol to +4 kcal/mol relative to the wild type protein. One of the largest effects on homodimerization was observed for the variant G142A in the 140-146 loop which dimerizes with a free energy of -2.4 kcal/mol. The dimerization free energy of the adenylate-bound G142A variant is identical to that of apoBirA, consistent with a complete loss of coupling between ligand binding and dimerization.

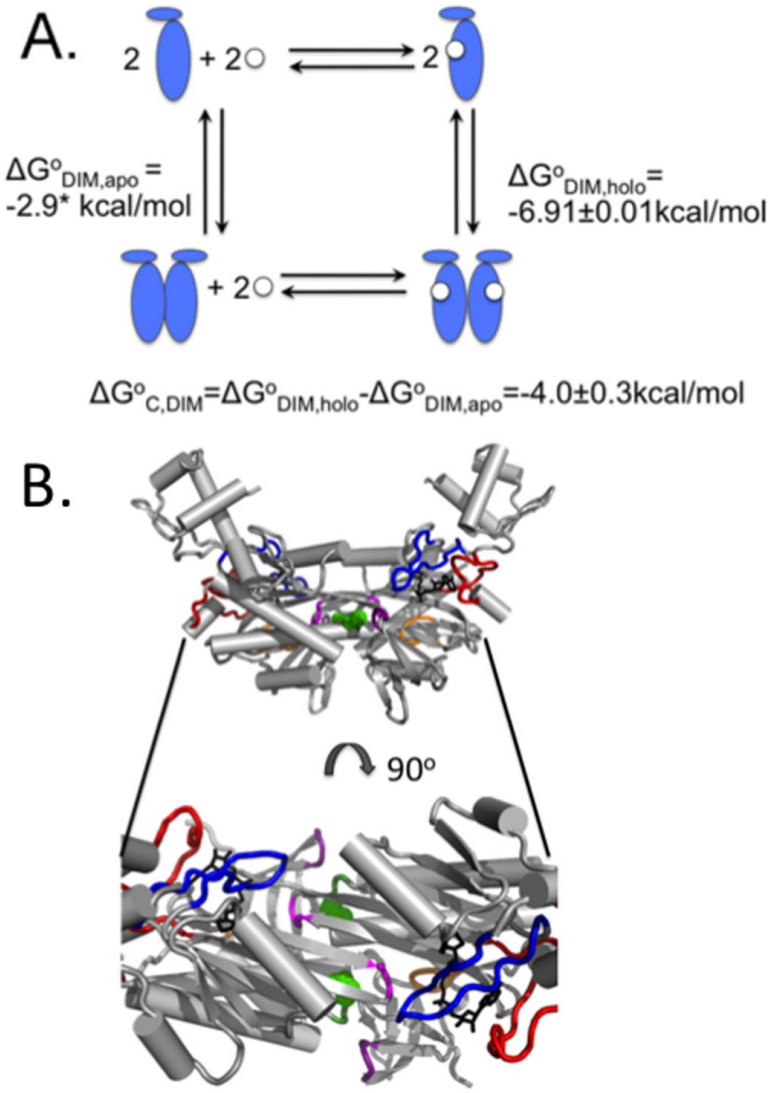

Figure 16: A. Thermodynamic cycle illustrating coupling between small ligand, bio-5'-AMP (circles), binding and BirA dimerization. B. Model of the holoBirA (biotinol-5'-AMP-bound) dimer structure with loops highlighted in color. Color coding: Blue: 116-128, Green: 140-146, Orange: 170-175, Pink: 193-199, Red: 206-233, Purple: 280-282. The ligand, biotinol-5'-AMP, is shown in black sticks and the G142A residue as a green sphere. The model was created in PyMOL (21) using pdb file 2EWN as input.

The availability of high-resolution x-ray crystallographic structures can provide detailed information about the mechanism of allosteric regulation. The crystallographic structure of the BirA variant G142A provides support for a model in which disorder-to-order transitions participate in the allosteric communication. Consistent with previous sedimentation equilibrium measurements the structure is monomeric. The structural origins of the loss of dimerization are the loss of folding/packing of two dimerization surface loops, consisting of residues 140-146 and 190-193. Furthermore, substitution of glycine 142 with alanine on the dimerization surface results in disorder 33Å away on the ligand binding surface in the adenylate binding loop. This observation coincides with decreased binding affinity for the corepressor as measured by isothermal titration calorimetry (ITC). The combination of structural and thermodynamic data provide support for an allosteric mechanism involving reciprocal communication between two disorder-to-order transitions on distant protein surfaces.

3.3 Results

3.3.1 Structural overview of the G142A variant

The model structure was solved through molecular replacement using the monomeric model of the wild type structure bound to biotinol-5'-AMP (2EWN) as the starting model (72, 110, 111). Both the ligand and waters were removed prior to using the model for molecular replacement. Following the refinement process, the structure reveals a monomeric protein-ligand complex in accordance with the previously reported sedimentation equilibrium measurements (Fig. 17A, Table 5)

(80). The G142A variant crystallizes in the $P4_3$ space group, unlike the wild type protein bound to biotinol-5'-AMP, which crystallizes in the $P4_32_12$ space group. The final model is in excellent agreement with the diffraction data as evidenced by the overlay of the structure and ligand with the electron density (Fig. 17B and C).

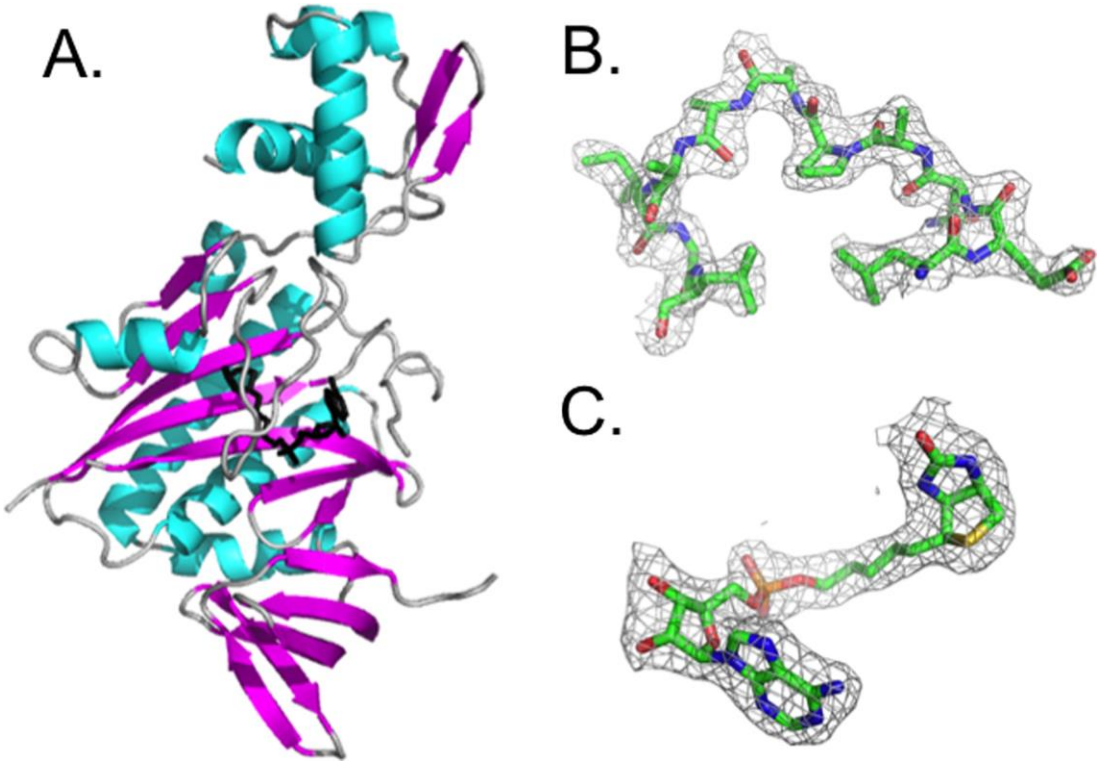

Figure 17: A. Model of the G142A monomer structure created in PyMOL (21). The ligand, biotinol-5'-AMP, is shown in black sticks. B & C. Overlays of the model for residues 139-149 and the ligand, biotinol-5'-AMP, with the electron density from the 2Fo-Fc map contoured at 1 sigma.

Table 5. Crystallographic data and refinement statistics.

Data Collection	
Resolution Range (Å)	30.19-2.31 (2.39-2.31)
Space Group	$P4_3$
Cell dimensions	
a, b, c (Å)	46.2, 46.2, 157.1
α, β,γ (°)	90.00, 90.00, 90.00
Mean I/sigma(I)	11.68 (1.71)
Unique Reflections	14338 (1323)
Completeness (%)	99.42 (94.30)
Wilson B-factor ($Å^2$)	34.42
Refinement	
R-work	0.172 (0.218)
Free R-factor	0.218 (0.277)
No. water molecules	135
Total non-H atoms	2535
Protein + Ligand	2400
Solvent	135
RMS Deviations	
Bond length r.m.s.d. (Å)	0.005
Bond angle r.m.s.d. (°)	0.84
Coordinate error (Å) (Maximum-Likelihood Based)	0.24
Ramachandran favored (%)	98.0
Ramachandran outliers (%)	0
Clashscore	2.68
Average B-factor ($Å^2$)	38.4
Protein	38.3
Solvent	39.7

Statistics for the highest-resolution shell are shown in parentheses.

3.3.2 Comparison of the wild type and G142A structured regions

Comparison of the biotinol-5'-AMP bound complexes of wild type and G142A proteins was accomplished by overlaying the structures using Multiseq in VMD (Fig. 18A) (112, 113). The structural alignment reveals an overall backbone RMSD of 1.15Å, indicating significant agreement between the two structures. Despite this close agreement there are four regions of divergence with RMSD values

> 2Å (Fig. 18B). The first region comprises residues 53-56 of the β-hairpin in the winged helix-turn-helix motif of the N-terminal domain. This region is known for its conformational heterogeneity in the various liganded forms of BirA and is characterized by high temperature factors relative to the other structured regions of BirA (70-72). In addition to the deviation at the N-terminal domain, the wild type and G142A structures also diverge at glycine 78, which is at the C-terminal end of a flexible linker joining the N-terminal and central domains. This linker is also characterized by relatively high temperature factors.

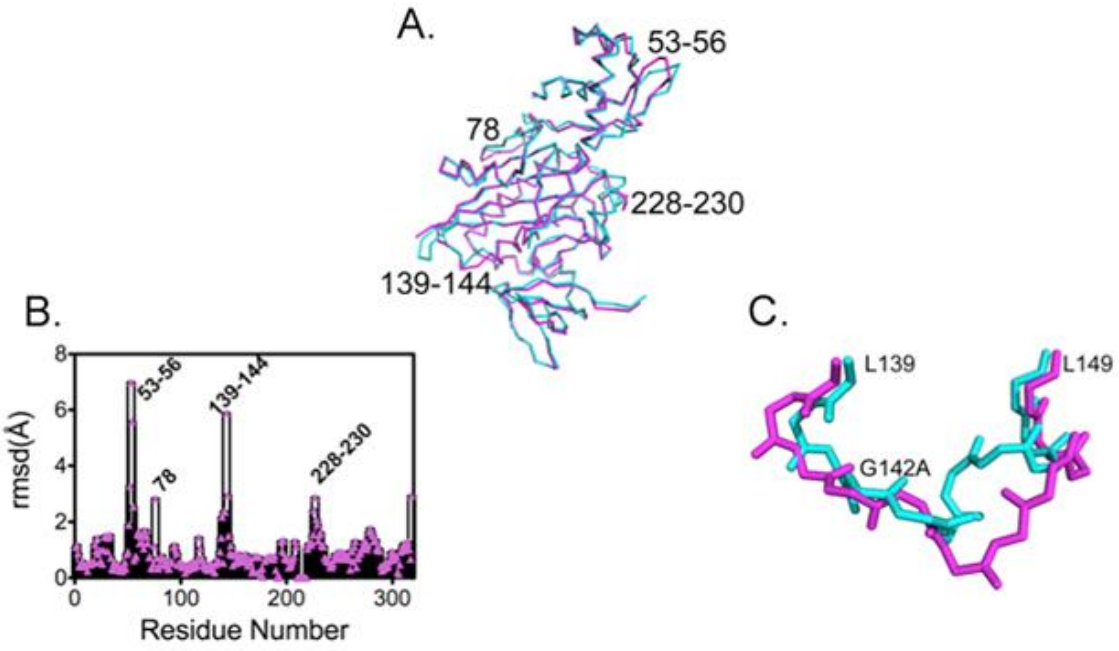

Figure 18: Comparison of the G142A and wild type structures. A. 3D alignment of the backbone structures for G142A (magenta) and wild type (teal) BirA. B. Dependence of the root mean square deviation of backbone atom positions between G142A and wild type BirA on amino acid position. C. Overlay of the backbone region of residues 139-149 for G142 (magenta) and wild type (teal) BirA.

In the comparison of the wt and G142A BirA structures the central domain contains two regions of significantly higher RMSD values including residues 139-144 on the dimerization surface and 228-230 on the ligand binding surface. The region consisting of residues 139-144 contains the alanine replacement at position 142 on the dimerization surface. The substituted position has the smallest deviation from the wild type structure while the remaining residues are quite divergent (Fig. 18C). Surprisingly, the last region of significant divergence lies in residues 228-230 of the adenylate binding loop on the opposite side of the protein monomer 33Å from the alanine substitution at position 142. While structural differences were observed between the liganded wild type and G142A structures in both the N-terminal and central domains, the C-terminal domains were comparable.

3.3.3 Disordered regions in the G142A structure

Analysis of the G142A structure reveals two unstructured regions. The first region is part of the dimerization surface containing residues 193-196 while the second is part of the ligand binding surface consisting of residues 212-221. These regions are folded in the liganded wild type dimer and unfolded in the apo wild type structure. For both of the unstructured regions, manual building coupled to refinement was first unsuccessfully attempted in Coot and further attempts using Autobuild resulted in poor agreement between these regions of the model and density (114-116). The poor agreement obtained after the Autobuild refinement is evident in the real-space correlation coefficients and temperature factors for the two regions (Fig. 19). Furthermore, the disorder does not result from intermolecular crystal lattice contacts.

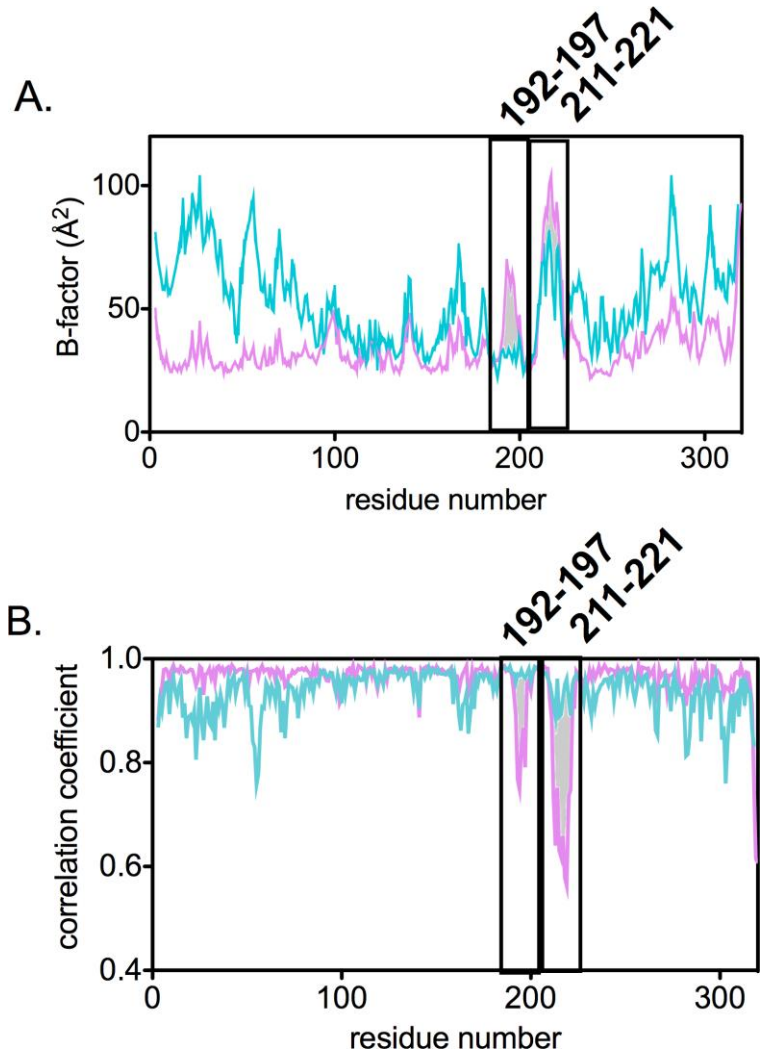

Figure 19. Temperature factors and real-space correlation coefficients for the wild type and G142A structures. A. Temperature factors for wild type BirA (cyan) and G142A (magenta) with the two regions 192-197 and 211-221 highlighted. B. Real-space correlation coefficients for these same regions indicate the agreement between the model and experimental electron density is significantly better for the wild type (cyan) than the G142A (magenta) structure.

The structural origins for the loss of dimerization free energy of the G142A variant appear to originate from the unstructured region on the dimerization interface consisting of residues 193-196 in the 193-199 loop. This region is folded in the biotinol-5'-AMP bound wild type structure with several interactions with the neighboring 140 loop. Using CMView with a 4Å cutoff, the interactions between residues of the 140-146 loop and 193-199 loop in the wild type protein include: E140 with D97 and A198, G142 with D197 and A199, and lastly P143 with K194 and A199 (Fig. 20A). However, in the G142A structure many of these interactions are no longer present including the loss of intramolecular contacts between P143 and residues K194 and A199 which appear to prevent the folding of residues 193-199. In fact, further comparison of the wild type and G142A structure highlights the orientation of the proline at position 143 in the intramolecular interaction. In the wild type structure, P143 protrudes away from the core structure enabling interaction with the alpha carbon of K194. By contrast, P143 is flipped approximately 180° in the G142A structure toward the protein core precluding any interaction with 190 loop residues.

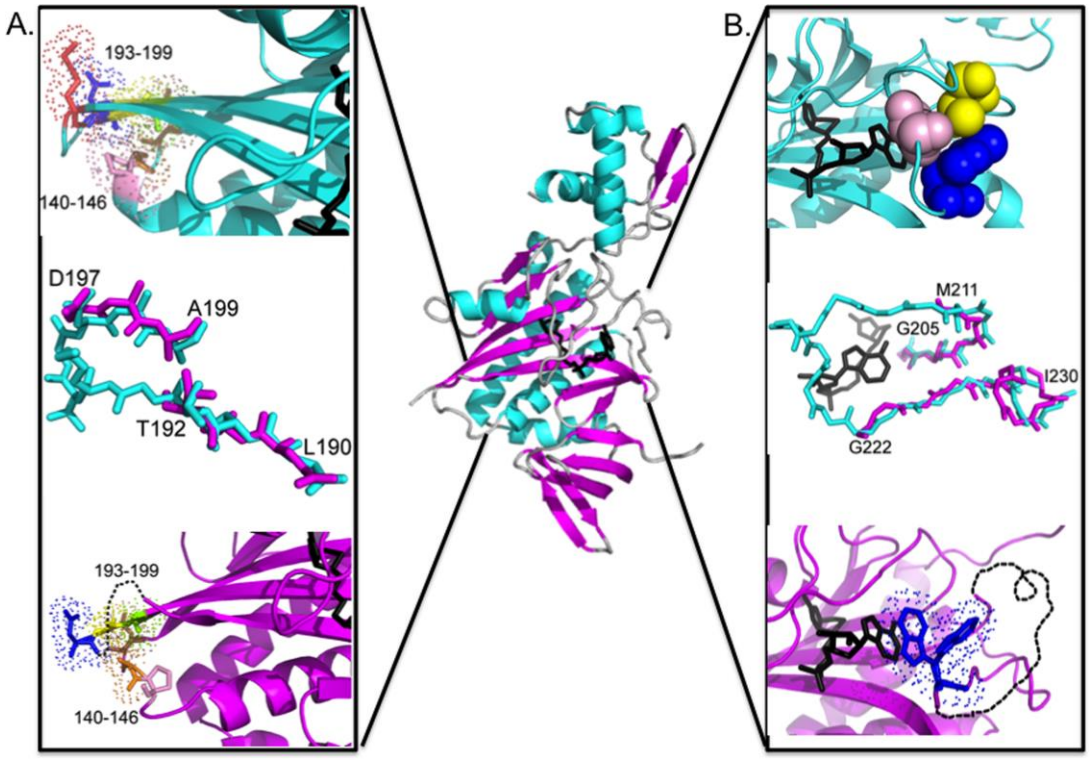
A.
193-199
140-146
D197
A199
T192
L190
193-199
140-146
B.
M211
G205
I230
G222

Figure 20: Disorder on the dimerization and bio-5'-AMP binding faces of the G142A variant. A. Comparison of the dimerization surface loop structures for G142A and wt BirA. Center: Alignment of backbone structures of the 193-199 loop of G142A (magenta) and wt (teal) BirA. Top: Interaction of the 140-146 and 193-199 loops in adenylate-bound wt BirA. Bottom: Inter-loop interactions in the G142A variant with disordered 193-196 segment shown as a dashed line. Residues of the two structures that participate in inter-loop interactions based on contact map analysis (see text) are shown in sticks with dots representing the atomic surfaces. Color code: 140-146 loop: E140: brown, G or A142: orange, P143: pink, 193-199 loop: K194: red, D197: blue, A198: yellow, A199: chartreuse B. Loop structures for G142A and wt BirA on the ligand binding surface. Center: Overlay of backbone structures for residues 205-230 of G142A (magenta) and wt (teal) BirA. Top: Folding of the 211-222 segment over the adenylate in the wt complex. The V214 (pink), V219 (yellow), and W223 (blue) residues of the hydrophobic cluster are shown as spheres. Bottom: The disordered 211-221 segment (dashed line) in adenylate-bound G142A BirA with partial occupancy of the W223 side chain (blue) in two orientations, stacked and unstacked, relative to the adenine ring of the adenylate.

In addition to the disordered region on the dimerization surface, 33Å away on the opposite side of the protein monomer the ligand binding surface contains an unstructured region comprised of residues 212-221. This region is part of the larger adenylate binding loop consisting of residues 211-234. Folding of the ABL in the wild type structure coincides with the formation of a hydrophobic cluster of comprised of residues V214, V219, and W223 over the adenylate moiety of the biotinol-5'-AMP (Fig. 20B) (72). In the G142A structure residues 212-221 of the ABL are disordered prohibiting the formation of the hydrophobic cluster over the adenylate. Moreover, the tryptophan at position 223, now unable to participate in the cluster, partially occupies a stacked position over the adenylate in a similar fashion to V219 in the biotinol-5'-AMP bound wild type structure. Additionally, while the position of the biotin component of biotinol-5'-AMP is highly conserved in the two structures, the adenine and ribose portion of the ligand deviate (Fig. 21). This is not surprising when comparing the liganded structures of the wild type and G142A variant. The biotin binding loops of both structures overlay extremely well, while portions of the adenylate binding loop are folded in the wild type structure and completely disordered in G142A.

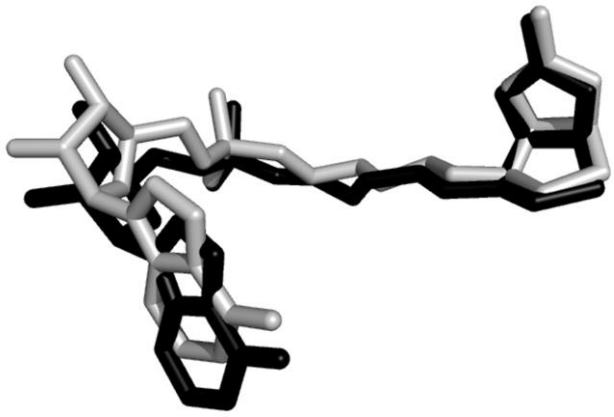

Figure 21. Overlay of biotinol-5'-AMP ligands from wild type and G142A structures. Ligands were extracted from the aligned models with bioinol-5'-AMP from wild type (black) and G142A (gray).

To determine if the disorder on the ligand binding surface correlates with decreased binding affinity for the corepressor, binding measurements were performed using isothermal titration calorimetry (Fig. 22). In agreement with the structural similarities of the biotin binding loop for the wild type and G142A proteins the biotin binding affinities are identical (Table 6). By contrast, the bio-5'AMP binding free energy of the G142A is perturbed by 3.5 kcal/mol relative to the wild type protein. While corepressor binding to the wild type protein, due to its high affinity, required the use of a displacement titration method, direct titrations were adequate for measuring bio-5'-AMP binding to G142A (107). The decreased binding affinity of the G142A is in direct agreement with the observed structural perturbation to the adenylate binding loop. Additionally, this result is consistent with previous binding

measurements in which perturbation to ABL folding by single alanine substitutions confirmed by partial proteolysis resulted in decreased binding affinity (77).

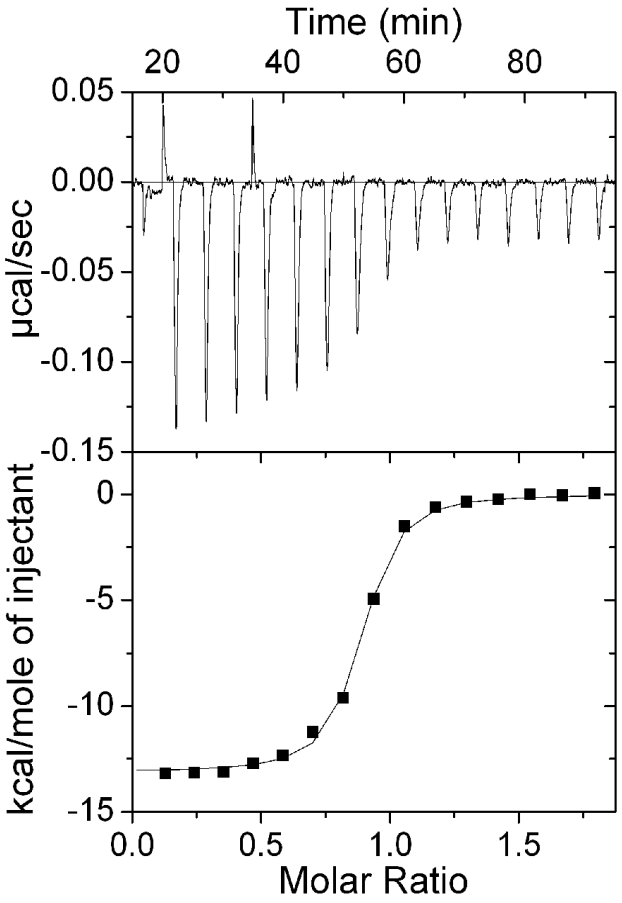

Figure 22. Isothermal titration calorimetry trace for G142A BirA with bio-5'-AMP. Titrations were performed with 15 13 uL injections of 20 µM bio-5'-AMP into a sample cell containing 2 µM G142A. Titration data were analyzed using a single-site binding model in Origin 7.0.

Table 6: Comparison of ligand binding thermodynamics.

Ligand/BirA variant	wild type	G142A
Biotin		
K_D(M)[a]	4.0(±0.7)x10^{-8}	2.2(±0.2)x10^{-8}
ΔG^o(kcal/mol)	-9.9±0.1	-10.3±0.1
ΔH^o(kcal/mol)	-19.0±0.5	-19.7±0.3
$-T\Delta S^o$ (kcal/mol)	9.1±0.5	9.5±0.3
n	0.85±0.04	0.86±0.01
Bio-5'-AMP		
K_D(M)[a]	3.2(±0.3)x10^{-11}	1.8(±0.7)x10^{-8}
ΔG^o(kcal/mol)	-14.1±0.6	-10.4±0.3
ΔH^o(kcal/mol)	-11.7±0.1	-13.4±0.2
$-T\Delta S^o$ (kcal/mol)	-2.34±0.02	2.9±0.5
n	1.01±0.01	0.89±0.05

[a] The errors were propagated from the 95% confidence intervals associated with at least two independent titrations.

3.4 Discussion

Despite the predicted role of protein disorder in allosteric regulation, limited experimental data exists on the mechanisms by which it functions. In this work the x-ray crystallographic structure of G142A BirA combined with functional thermodynamic measurements highlight how replacement of glycine 142 with alanine in a dimerization surface loop perturbs both protein folding and function. Furthermore, the work, in combination with previous studies, demonstrates a mechanism of allosteric communication that involves coupled disorder-to-order transitions at two distant sites.

3.4.1 Structural origins for the loss of dimerization energetics

Previous work has demonstrated that dimerization surface loops in BirA contribute to self-association energetics (80). In particular, replacement of glycine 142 with alanine perturbs dimerization energetics by 4 kcal/mol, which is consistent with a complete loss of coupling. To determine the structural origins of this loss of coupling, the crystal structure of the BirA variant G142A (4WF2) bound to the non-hydrolysable corepressor analog, biotinol-5'-AMP, was determined at a resolution of 2.39Å. Comparison of the G142A structure with that of the wild type protein in different liganded states the structural origins for the loss of dimerization energetics becomes clear. In the monomeric apo protein, residues 194-196 are disordered and the remaining residues of the 193-199 loop on the dimerization surface display high temperature factors (70). Additionally, the neighboring loop comprised of residues 140-146 is highly disordered with no electron density present for the G142 residue or the side chains of both E140 and Q141. Furthermore, high temperature factors are observed for the remainder of the loop. The monomeric unit of the wild type biotinol-5'-AMP bound structure indicates folding and rigidification in both of these loops (72). For both the 140-146 and 193-199 loops all residues are visible and are characterized by low temperature factors. In the G142A variant the structure of this region reveals the structural origins of the loss of dimerization free energy. The dimerization surface is analogous to that of apo wild type structure with the absence of interaction between the 140-146 and 193-199 loops. As mentioned previously, comparison of the wild type and G142A structure highlights the rotated orientation of the proline at position 143 which precludes interaction with 190 loop and concomitant

disorder of residues 193-196. The structural comparisons combined with the previously determined importance of these loops in dimerization indicate that folding and subsequent interaction of the 140-146 and 193-199 loops are critical for coupling.

Further comparison of the 140 loop region of the dimerization surface of the G142A and wild type structures reveals additional differences. The apo and biotin-bound wild type structures contain an alpha helix composed of residues 147-164 and 146-164, respectively, which spans the width of the monomer from the dimerization to the ligand binding surface (Fig. 23). In the adenylate-bound wild type protein this alpha helix is extended by an additional 4 residues, 143-146, on the dimerization surface. In the adenylate bound G142A structure, similar to the apo and biotin-bound wild type proteins, residues 143-145 are not helical preventing interaction with the 193-199 loop. Analogous to the α-helical shortening seen in G142, previous mutational studies have shown that deletion of the residue A146 perturbs the dimerization free energy by 2 kcal/mol (76). The extension of the alpha helix allows the 140-146 loop to interact with the 193-199 loop thereby enhancing dimerization.

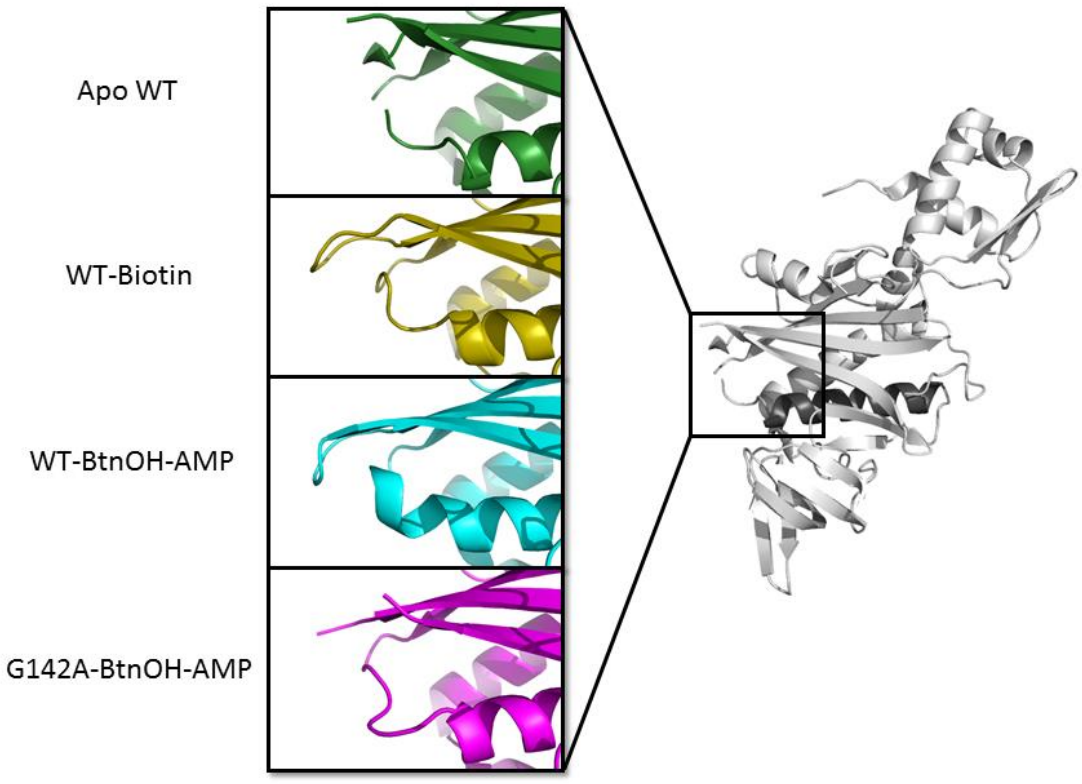

Figure 23. Absence of alpha helical extension on the G142A dimerization surface. Right: In the full apo structure, the alpha helix (black) extends from the dimerization to ligand binding surface. Left: The alpha helical segment on the dimerization surface is highlighted in wild type apo (green), biotin-bound (yellow), BtnOH-AMP-bound (cyan), and G142A BtnOH-AMP-bound (magenta) structures.

3.4.2 Coupling between two disorder-to-order transitions

The structural and thermodynamic measurements of the G142A variant are consistent with coupling of corepressor binding and dimerization communicated through disorder-to-order transitions on the two distant functional surfaces. The

G142A structure clearly shows disorder on the dimerization surface in residues 193-196 and also 33Å away on the ligand binding surface in residues 212-221. Extensive investigation of the adenylate binding loop, which is comprised of residues 211-234, has revealed the importance of its proper folding for corepressor binding and allosteric activation (77, 117). Single and multiple alanine substitutions in the ABL hydrophobic cluster consisting of the residues V214, V219, and W223 result in perturbed loop folding, adenylate binding, and dimerization energetics. Consistent with these previous studies, perturbed ABL folding in the G142A variant is associated with an approximately 450-fold increase in the equilibrium dissociation constant for bio-5'-AMP binding compared to the wild type protein. These results indicate that folding and intramolecular interactions between the 140-146 and 193-199 loops and folding of residues 211-221 are reciprocally coupled.

3.4.3 The mechanism of communication between distant disorder-to-order transitions

Allosteric communication in BirA occurs through reciprocal coupling of disorder-to-order transitions on the dimerization and ligand binding surface. However, the mechanism of communication between these distant sites remains unknown. Analysis of the protein core connecting these two regions does not reveal any obvious structural mode of communication. The two beta strands connecting the disordered loops in the G142A variant overlay exceptionally well with the corresponding segment of the wild type structure, with backbone RMSD values of 0.79 and 0.66 Å. Further, the adjacent alpha helix that spans the width of the BirA

monomer from the ligand binding surface to dimerization surface was examined. With the exception of the 4-residue extension on the dimerization surface in the adenylate-bound wild type structure, no divergence is observed in the G142A variant with a backbone RMSD value of 0.92 Å. The absence of information about the mechanism of communication from the static structural information available from x-ray crystallographic structures opens the possibility that mechanism is dynamic in nature. In the case in dihydrofolate reductase (DHFR) replacement of glycine-121 in a loop distal to the active site decreases the rate of hydride transfer by 200-fold (46). Comparative NMR analysis of the wild type and G121V variant has revealed that decreased dynamic fluctuations of the distal loop correlate with impaired enzyme function (49). Further analysis with techniques that capture the dynamic properties of the protein could potentially shed light on the mechanism of communication in BirA.

The work presented demonstrates allosteric coupling of two distal disorder-to-order transitions. Furthermore this work highlights the functional importance of glycine residues in flexible loop regions. While the mechanism by which these transitions are communicated is unknown, their involvement in the allosteric response is clearly demonstrated.

3.5 Materials and Methods

3.5.1 Protein preparation and crystallization

The BirA variant, G142A, was over-expressed, purified and characterized as previously described (80). Prior to crystallization the protein was exchanged into

buffer containing 100 mM NaH_2PO_4/K_2HPO_4 (pH 6.5 (±0.02) at 20.0 (±0.1) °C), 5% (v/v) glycerol. The protein and the ligand, biotinol-5'-AMP (RNA-TEC), were diluted into this same buffer to achieve final concentrations of 10 mg/mL (280 μM) and 1 mM, respectively, at which the protein is saturated with ligand. A one microliter volume of the protein:ligand complex was combined with an equal volume of precipitant containing 0.1 M Tris HCl, pH 8.0 and 12.5% (w/v) PEG 8K (Hampton) at 20°C and crystallization was carried out using the hanging drop vapor diffusion method in Linbro Boxes over a well containing 300 μL of precipitant solution. Crystals appeared in 24 hours and grew to a size of 200 μm. The BirAG142A-btnOH-AMP complex crystallized in the $P4_3$ space group with unit cell dimensions a=b=46.2Å and c=157.1Å, and 1 molecule per asymmetric unit.

3.5.2 Data acquisition and structural solution

Prior to data collection the crystals were dipped into a solution containing 30% ethylene glycol, 8%PEG 8K, 40 mM Tris HCl pH 8.0 for 20 sec and rapidly plunged into liquid nitrogen. Data were acquired on a Rigaku FR-E SuperBright rotating anode X-ray generator with a Saturn 944+ CCD detector and processed to 2.3-Å resolution using HKL2000 (118). A total of 180° (1°/frame) were collected and indexed in the space group.

The structure of the G142A BirA:btnOH-AMP complex was solved by molecular replacement (MR). Initial MR was performed using MolRep (110) in the CCP4 6.3.0 program suite (119). The starting monomer model used for in the search

was extracted from the Protein Data Bank file for the wild type BirA protein dimer bound to btnOH-AMP (2EWN, (72)). Both the ligand and water molecules were removed prior to using the model in the search. Immediately following MR, 10 cycles of restrained refinement in REFMAC5 (120) yielded a free R_{FREE} of 27%. The ligand, btnOH-AMP, was added at this stage and final refinement was carried out iteratively using COOT (121) and REFMAC5 in the CCP4 environment. In an attempt to define the structures of the two loops segments comprised of amino acid residues 193-196 and 212-221 model building was performed using Autobuild (122) in Phenix (122-124). Incorporation of TLS refinement using five groups in the final stages yielded a significantly improved R_{free}-R_{work} value (125, 126). The quality of the final model was assessed using the comprehensive validation tools in Phenix (127-130). All structure figures were generated using PyMOL (21). The pairwise 3D alignment of the wild type and variant structures was performed using Multiseq (112) in VMD (113). Contact map analysis was performed in CMview (131). Two-dimensional representations of the protein-ligand interactions were produced using LigPlot (132). The table of refinement statistics was generated using the tool in Phenix.

3.5.3 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) was carried out using a VP-ITC microcalorimeter (G.E. Healthcare). Protein was dialyzed extensively against Standard Buffer (10 mM Tris HCl, 200 mM KCl, 2.5 mM $MgCl_2$, pH 7.5 at 20°C), filtered through 0.22 μm PTFE membrane syringe filters and its concentration was determined by UV absorption using an extinction coefficient of 47510 $M^{-1}cm^{-1}$ (133)

Titrations of wt with biotin and G142A BirA with biotin and bio-5'-AMP were performed by injecting 15 to 25, 8 to 13 μL volumes of 20 μM ligand solution prepared in Standard Buffer into the sample cell containing 1.4 mL of 2 μM protein. Titration data were analyzed using a single-site binding model in Origin 7.0 to obtain the equilibrium association constant for binding, the standard binding enthalpy and the stoichiometry of the binding interaction. Due to the very high affinity of the interaction, adenylate binding to wtBirA was measured using the displacement method in which a 2 μM protein saturated with biotin (6 μM) was titrated with 21-12 μL volumes of a 20 μM bio-5'-AMP solution. The resulting data were subjected to analysis using the competitive binding model in Origin 7.0. For all titrations the stirring rate was set at 310 rpm.

Acknowledgements

I would like to thank Dr. Wade and Dr. Bachas for their support and allowing me to use the diffractometer at the Johns Hopkins University School of Medicine. Thank you to Dr. Beckett for the solution of the G142A structure and for providing Table 5 and Figures 16-21 and their corresponding legends in addition to the materials and methods section. I thank Dr. Paukstelis, Dr. LaRonde, and Dr. Moulaei for their assistance in solving the structure of G142A. Thank you to Dr. Adikaram for the preparation and purification of the G142A. Lastly, I thank William Cressman for the wild type-bio-5'-AMP binding measurements by ITC. My contributions to this work include the crystallization of G142A, development of a suitable cryo-protectant, data acquisition, and the measurements of biotin and bio-5'-AMP binding using ITC.

Chapter 4: A Large Solvent Isotope Effect on Protein Association Thermodynamics

The work in this chapter has been published: Eginton, C., Beckett, D. (2013). A large solvent isotope effect on protein association thermodynamics. *Biochemistry* 52(38):6595-6600.

4.1 Abstract

Solvent reorganization can contribute significantly to the energetics of protein:protein interactions. However, our knowledge of the magnitude of the energetic contribution is limited, in part, by a dearth of quantitative experimental measurements. The biotin repressor forms a homodimer as a prerequisite to DNA binding to repress transcription initiation. At 20^{o}C the dimerization reaction, which is thermodynamically coupled to binding of a small ligand, bio-5'-AMP, is characterized by a Gibbs free energy of -7 kcal/mole. This modest net dimerization free energy reflects underlying very large opposing enthalpic and entropic driving forces of 41 ± 3 and -48 ± 3 kcal/mole, respectively. The thermodynamics have been interpreted as indicating coupling of solvent release to dimerization. In this work, this interpretation has been investigated by measuring the effect of replacing H_2O with D_2O on the dimerization thermodynamics. Sedimentation equilibrium measurements performed at 20^{o}C reveal a solvent isotope effect of -1.5 kcal/mole on the Gibbs free energy of dimerization. Analysis of the temperature dependence of the reaction in D_2O indicates enthalpic and entropic contributions of 28 and -37 kcal/mole,

respectively, considerably smaller than the values measured in H_2O. These large solvent isotope perturbations to the thermodynamics are consistent with a significant contribution of solvent release to the dimerization reaction.

4.2 Introduction

Protein:protein interactions are central to a broad range of biological processes including signal transduction, transcription regulation and morphogenesis (134-136). Although the intrinsic chemistry of the interacting protein partners is important in determining the strength of an interaction, water, either through its release into the bulk upon interface formation or through its direct participation in the interface, is also integral to protein:protein interactions. Moreover, simulations predict that the interplay of protein sequence with solvent profoundly influences protein assembly reactions (137). Despite the widespread appreciation of its significance, experimental determination and computational prediction of the energetic contribution of solvent reorganization remain a challenge for understanding both the physical chemistry of protein:protein interactions and for design of novel interactions (138).

The *E. coli* protein BirA forms a homodimer prior to binding to DNA to regulate transcription initiation (63). The dimerization reaction is thermodynamically coupled to small molecule bio-5'-AMP binding, which renders the protein:protein interaction more favorable by -4 kcal/mole (Fig. 24) (66, 73). The equilibrium dimerization constant for bio-5'-AMP bound BirA, holoBirA, is 6 μM at 20^{o}C, 200 mM KCl, pH 7.5, which corresponds to a modest Gibbs free energy of -7 kcal/mole

(77). By contrast, van't Hoff analysis of the temperature-dependence of the reaction indicates very large opposing enthalpic (ΔH^{o}_{DIM}) and entropic ($-T\Delta S^{o}_{DIM}$) driving forces of 41 and -48 kcal/mole, respectively, at 20°C (68). Furthermore, the linearity of the van't Hoff plot obtained from analysis of the temperature-dependence of the dimerization reaction is consistent with the absence of a heat capacity change in the reaction.

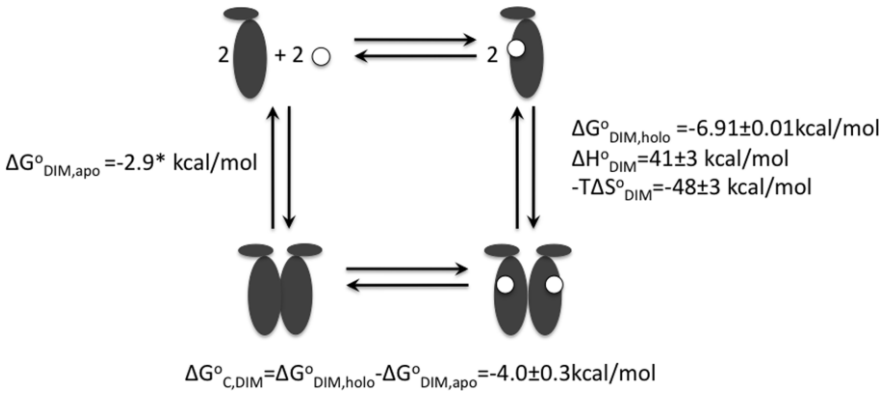

Figure 24. Thermodynamic cycle illustrating the linkage between BirA dimerization and bio-5'-AMP binding. The coupling, ($\Delta G^{o}_{C,DIM}$), is calculated from the difference in the measured dimerization free energy of bio-5'-AMP-bound or holoBirA, and unliganded or apoBirA, ($\Delta G^{o}_{DIM,holo}-\Delta G^{o}_{DIM,apo}$). The thermodynamic parameters for holoBirA dimerization are shown on the right: (white circle) bio-5'-AMP and (gray shape) BirA. The free energy of apoBirA dimerization in 200 mM KCl is calculated from the value measured in 50 mM KCl and the measured effect of salt concentration on holoBirA dimerization (66, 68).

The large opposing enthalpies and entropies of holoBirA dimerization are consistent with solvent release upon dimer formation (68, 139). The enthalpic penalty may reflect the cost of water removal from the dimerization surface concomitant with formation of the interface (140). Likewise, the favorable entropy is consistent with release of bound water to the bulk upon dimerization. The chemistry of the dimerization surface, which contains a number of polar and charged groups, is consistent with a large enthalpic penalty of removing surface-bound water during the course of dimerization (72). Analysis of the dimer interface, in which 40% of the buried groups are polar, indicates the presence of 16 hydrogen bonds and 4 salt bridges (74, 141). Furthermore, few water molecules are located in the interface. However, given the net unfavorable dimerization enthalpy of 41 kcal/mole, this structural interpretation of the thermodynamics assumes a desolvation penalty that is much larger than the favorable enthalpy associated with formation of combined intersolvent hydrogen and protein:protein interface bonds.

A number of experimental approaches have been used to investigate the contribution of water reorganization to biomolecular interactions. Addition of osmolytes allows measurements of the dependence of the reaction on water activity, thus providing information about the number of waters released or taken up in the course of binding. However, a potential complication to interpretation of the results lies in the preferential interaction of the osmolyte with one or more of the interacting partners (142). The volume change obtained from measuring the response of the equilibrium constant for an interaction to changes in hydrostatic pressure can be

interpreted in terms of the linkage of water binding or release to the equilibrium process (143). Pressure perturbation calorimetry also provides a means of estimating the role of water reorganization in biomolecular interactions (144). Finally, because of the distinct hydrogen bonding properties of the two, measurements of the solvent isotope effect of replacing H_2O with D_2O have been used to estimate the contribution of water reorganization to protein:protein association (145).

In this work the role of solvent reorganization in holoBirA dimerization has been probed by measuring the consequences for the reaction of replacing H_2O with D_2O. Sedimentation equilibrium measurements performed on wild type and single-amino acid variants of BirA reveal that dimerization is consistently more favorable by approximately -1.5 kcal/mole in D_2O than in H_2O. van't Hoff analysis of the temperature-dependence of the dimerization equilibrium measured in heavy water indicates linear behavior that yields an unfavorable dimerization enthalpy of 28 kcal/mole, 13 kcal/mole less than that measured in H_2O. This enthalpy decrease is accompanied by similarly large decrease in the favorable dimerization entropy. The results are consistent with a contribution of solvent reorganization to the holoBirA dimerization energetics.

4.3 Results

4.3.1 The holoBirA dimerization reaction is more favorable in D_20 than in H_2O

The equilibrium dimerization constant of wild-type holoBirA at 20°C in standard buffer (10 mM Tris HCl, 200 mM KCl, and 2.5 mM $MgCl_2$) prepared with

D_2O (SB:D_2O) was measured using sedimentation equilibrium. Initial analysis of the concentration versus radial position curves acquired at three loading concentrations and three speeds indicated weight-average molecular weights higher than that expected for the monomer (Fig. 25). Global analysis using a monomer-dimer model yielded a best-fit equilibrium dissociation constant for dimerization of $5 \pm 5 \times 10^{-7}$ M, indicating 12-fold tighter dimerization in SB:D_2O than in SB:H_2O (Table 7).

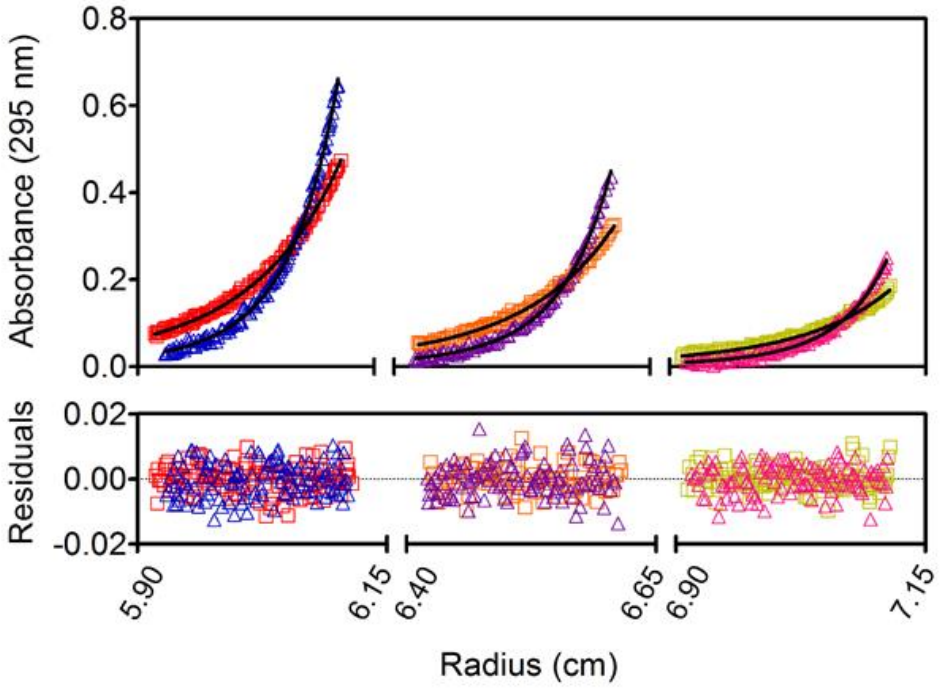

Figure 25. Sedimentation equilibrium measurements of wild-type holoBirA in SB:D_2O at 20°C. HoloBirA prepared at 23, 15 and 7.5 μM (from left to right) was centrifuged at 18,000 (□), and 24,000 (Δ) rpm. The lines correspond to the best-fit of six data sets to a monomer-dimer model. In order to clearly see the best-fit curves, only 6 data sets are shown. The lower panel shows the residuals of the fits for each data set.

Table 7: Solvent isotope effects on BirA dimerization.

	H_2O		D_2O		
BirA variant	K_{Dim} (M)[a]	ΔG^o_{Dim} (kcal/mol)[b]	K_{Dim} (M)[a]	ΔG^o_{Dim} (kcal/mol)[b]	$\Delta\Delta G^o_{Dim}$ (kcal/mol)[b,c]
Apo wild-type		(-2.8)[d]	$6 \pm 1 \times 10^{-4}$	–4.3± 0.1	(-1.5)[d]
Holo wild-type[e]	$6 \pm 2 \times 10^{-6}$	–7.0 ± 0.3	$5 \pm 5 \times 10^{-7}$	–8.5± 0.5	–1.5 ± 0.6
V219A[e]	$8 \pm 1 \times 10^{-5}$	–5.5 ± 0.2	$7 \pm 4 \times 10^{-6}$	–6.9± 0.3	–1.4 ± 0.4
T195A[e]	$10 \pm 4 \times 10^{-5}$	–5.3 ± 0.2	$9 \pm 7 \times 10^{-6}$	–6.8± 0.4	–1.5 ± 0.4

[a] Standard errors reported are from two independent experiments.
[b] Standard error propagation methods were used to determine the uncertainties for each reported value.
[c] $\Delta\Delta G^o_{DIM} = \Delta G^o_{DIM}$ (SB:D_2O) - ΔG^o_{DIM} (SB:H_2O).
[d] The values provided for apoBirA were calculated using the equilibrium dissociation constant measured in SB:H_2O containing 50 mM KCl and assuming identical dependencies of apo- and holoBirA dimerization on KCl concentration (66, 68).
[e] The equilibrium constant and Gibbs free energy of dimerization in H_2O for the holo wild-type and T195A and V219A variants were previously reported (77, 80).

4.3.2 The magnitude of the D_2O effect on holoBirA self-association is conserved

The generality of the effect of D_2O on holoBirA self-association energetics was investigated by performing measurements on the single-alanine-substituted variants, T195A and V219A. These particular variants were chosen based on the distinct locations of the substituted residues in the dimer structure and their altered dimerization energetics in H_2O relative to wild-type BirA. The T195 residue is located at the core of the dimer interface, and V219 is on the dimer surface in a loop that folds over the adenylate moiety of bio-5'-AMP (Fig. 26A). The equilibrium dissociation constants for dimerization of the T195A and V219A variants in SB:H_2O at 20^{o}C are 80 and 100 μM, respectively (77, 80).

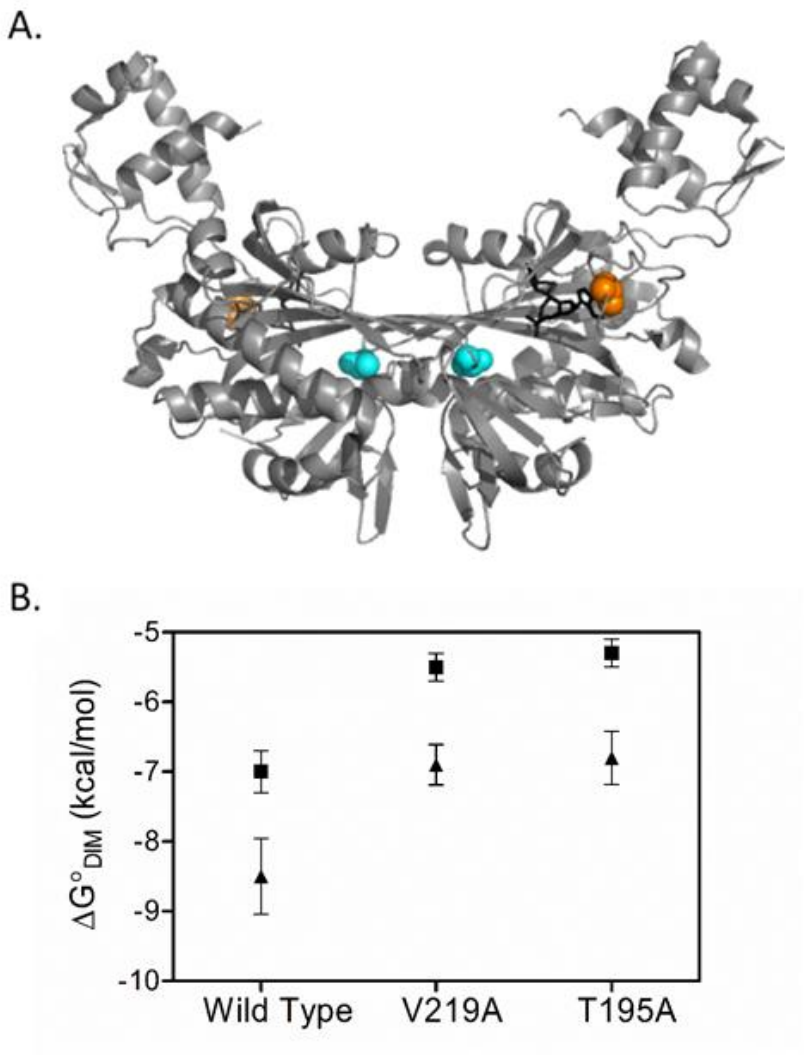

Figure 26. A. Locations of alanine substitutions in the BirA structure. The T195A (cyan) substitution is at the dimerization interface, and the V219A (orange) is in a loop that folds over the adenosine moiety and is on the outside surface of the dimer. The model was constructed in PyMOL using the input file 2EWN (21, 72) B. Energetic effects of transfer of the dimerization reaction mixture from SB:H_2O to SB:D_2O: The Gibbs free energies of dimerization in H_2O (■) and D_2O (▲) indicate a consistent enhancement of -1.5 kcal/mole.

Sedimentation measurements for the bio-5'-AMP-bound forms of T195A and V219A in SB:D_2O were performed as described for wild-type BirA. The high affinities of both proteins for the ligand ensure that they are quantitatively in the holo form at the concentrations employed for the measurements. Global nonlinear least-squares analysis of the data yielded equilibrium dissociation constants for dimerization in D_2O buffer that are 10-fold smaller than those measured in H_2O buffer. Thus, similar to the effect observed for wild type holoBirA dimerization, the Gibbs free energies calculated form the resolved equilibrium constants indicate -1.4 to -1.5 kcal/mole enhancements in dimerization of the variants resulting from transfer of the reaction from SB:H_2O to SB:D_2O (Fig. 26B).

4.3.3 The magnitude of the coupling of ligand, bio-5'-AMP, binding to BirA dimerization is preserved in D_2O

BirA dimerization is thermodynamically coupled to bio-5'-AMP binding with a difference in dimerization free energy for the liganded, holoBirA, and unliganded, apoBirA, forms in SB:H_2O of -4 kcal/mole (66). The coupling in D_2O was investigated by performing sedimentation measurements on apoBirA prepared in SB:D_2O. In contrast to the holoBirA samples, which were prepared at low micromolar concentrations, the weak apoBirA dimerization necessitated measurements on samples prepared in the 100 micromolar concentration range. Global analysis of the data using a monomer-dimer model yielded an equilibrium dissociation constant of 600 μM, corresponding to a Gibbs free energy of -4.3 kcal/mole, -1.5kcal more favorable than the value estimated for apoBirA in SB:H_2O

(66, 68). Calculation of the coupling free energy or $\Delta G^{o}_{C,DIM}$ from the dimerization free energies obtained for apo and holoBirA in SB:D_2O yields a value of -4.2 ± 0.5, identical, within error, to that obtained in SB:H_2O.

4.3.4 Enthalpic and entropic driving forces for dimerization in D_2O differ significantly from those measured in H_2O

In SB:H_2O at 20°C the modest holoBirA dimerization free energy results from large opposing enthalpic (ΔH^{o}) and entropic ($-T\Delta S^{o}$) contributions of 41 and -48 kcal/mole, respectively (68). The solvent isotope effect on this thermodynamic signature was investigated by performing sedimentation measurements in SB:D_2O at temperatures ranging from 5 to 20°C. The tight dimerization at higher temperatures precluded measurements above 20°C. At each temperature the data were acquired at three rotor speeds on samples prepared at three concentrations. At all temperatures, the data are well-described by the monomer-dimer model and, as observed in SB:H_2O, the dimerization becomes tighter with increasing temperature (Fig. 27) (68). The dimerization free energies calculated from the resolved equilibrium constants reveal that increasing the temperature from 5 to 20°C renders the reaction more favorable by -1 kcal/mole. van't Hoff analysis of the data indicates a linear relationship (Fig. 27) that yields a temperature-independent enthalpy, ΔH°_{DIM}, of 28 ± 3 kcal/mol. The entropic contributions to the dimerization free energy, calculated using the expression $\Delta G^{\circ}_{DIM} = \Delta H^{o} - T\Delta S^{o}$, are large and favorable over the entire temperature range (Fig. 28). As observed in SB:H_2O, holoBirA dimerization in SB:D_2O is characterized by a large unfavorable enthalpy and large favorable entropy.

However, the absolute values of the two energetic parameters in D_2O are markedly more modest than in H_2O.

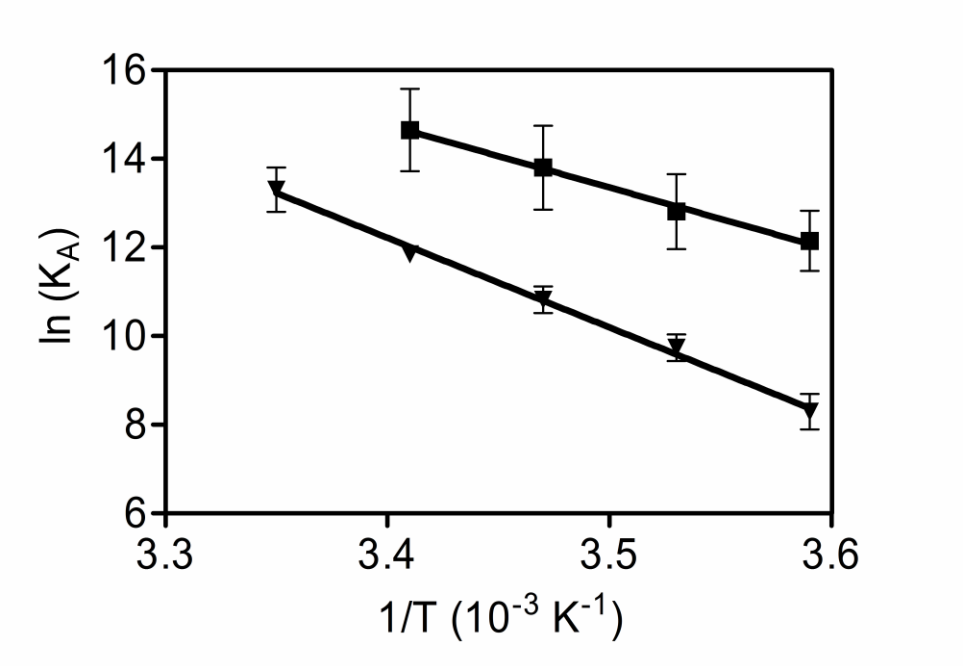

Figure 27. van't Hoff analysis of the temperature dependence of the equilibrium constant for holoBirA dimerization in SB:D_2O (■) and SB:H_2O (▼). The lines correspond to the best fits of the data to the van't Hoff equation. The data obtained in SB:H_2O were previously described (68).

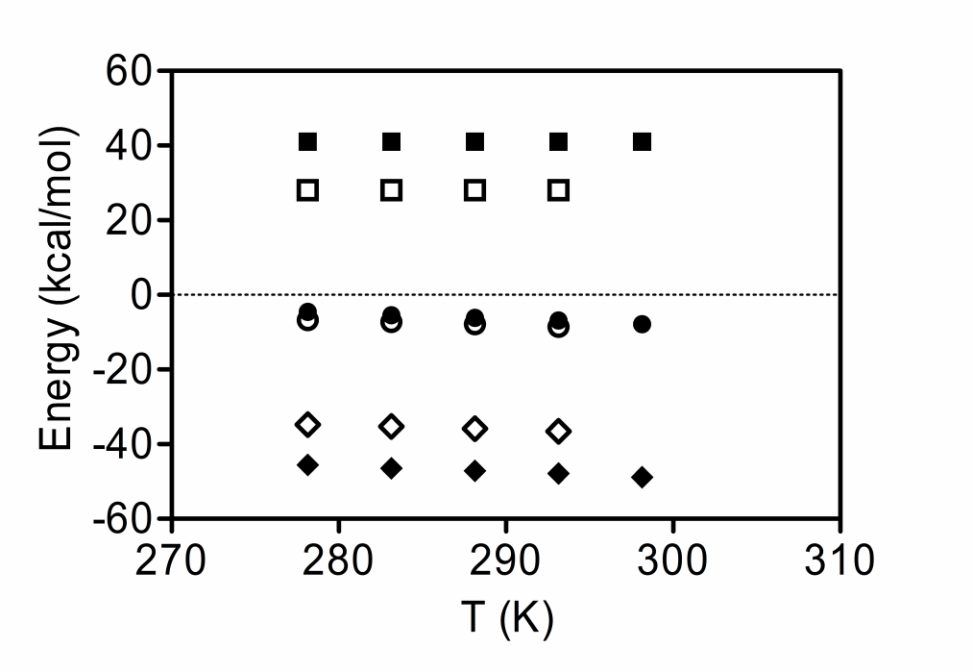

Figure 28. Thermodynamic signatures for holoBirA dimerization in SB:H_2O and SB:D_2O. The energetic terms in H_2O and D_2O are ΔG^o_{DIM} (● and ○), ΔH^o_{DIM} (■ and □) and $-T\Delta S^o_{DIM}$ (◆ and ◇). The values in SB:H_2O were previously published (68). On the scale used for the graph the estimated errors in the values are within the symbols.

4.4 Discussion

4.4.1 BirA dimerization is enhanced in D_2O relative to that in H_2O

HoloBirA self-association in D_2O is energetically more favorable than in H_2O. Sedimentation equilibrium measurements of wild-type holoBirA dimerization indicate a 12-fold decrease in the equilibrium dissociation constant in D_2O relative to

H_2O, corresponding to a -1.5 kcal/mole enhancement in the dimerization free energy. Measurements performed on two BirA variants with single alanine substitutions as well as apoBirA indicate enhancements of identical magnitude. The distinct locations of the two substitutions in the three-dimensional structure of BirA underscore the general effect of D_2O on the reaction. The T195A substitution is located in a loop in the dimerization interface. By contrast, the V219A substitution is in a surface loop that folds over the adenylate moiety of the ligand bio-5'-AMP, 24 Å away from the interface (72).

Although holoBirA dimerization in D_2O is enhanced relative to the reaction in H_2O no effect on the linkage between corepressor binding and dimerization is observed. Bio-5'-AMP, enhances BirA dimerization in both H_2O and D_2O with a coupling free energy in of -4.0 ± 0.3 kcal/mol, a result consistent with the general effect of the heavier solvent on dimerization. However, the coupling free energy in H_2O was determined in buffer containing 50 mM KCl rather than the 200 mM KCl used in the current studies (66). Thus, the conclusion that coupling free energies are equivalent in the two solvents assumes no influence of salt concentration on the coupling. The measured modest effect of monovalent salt concentration on holoBirA dimerization supports this assumption (68).

Previous studies indicate increases in self-association in D_2O relative to H_2O for several proteins including phycocyanin, glutamate dehydrogenase, and α-chymotrypsin (146-148). For the limited number of studies that provided quantitative

information, the magnitude of the effect varies. For example, at 4^{o}C self-association of tobacco mosaic virus coat protein is modestly enhanced by 3-fold in D_2O versus H_2O (149). A 10-fold enhancement of β-lactoglobulin A self-association is observed in heavy water, similar in magnitude to the effect on BirA dimerization (150).

4.4.2 The thermodynamics in H_2O and D_2O are consistent with coupling of solvent release to holoBirA dimerization

The measurements of holoBirA dimerization as a function of temperature in heavy water provide additional support for a significant contribution of solvent release to the energetics of the process. The dimerization enthalpies obtained in SB:H_2O and SB:D_2O are 41 ± 3 and 28 ± 3 kcal/mol, respectively. At 20^{o}C the entropic contributions, -TΔS^{o}, to dimerization free energy, are -48 and -36 kcal/mole, respectively, in the two buffers. As observed in H_2O, the net modest dimerization free energy of -8.5 kcal/mole in D_2O reflects large opposing enthalpies and entropies, albeit significantly more modest than those obtained in H_2O. These differences in the dimerization thermodynamics exist over the temperature range employed for the measurements (Fig. 28). In both solvents the net unfavorable dimerization enthalpy reflects the penalty of desolvating the polar dimerization surface. However, this removal is energetically offset, in part, by formation of intersolvent hydrogen bonds that accompanies its release to the bulk. The more modest unfavorable enthalpy measured for the reaction in D_2O than in H_2O reflects the stronger bonding between the heavy water molecules (151). Likewise, this stronger bonding yields a less favorable dimerization entropy in D_2O than in H_2O (151). The very slow holoBirA

dimerization kinetics estimated from analysis of sedimentation velocity data provides additional support for a significant role of solvent release in the process (152). Simulations of protein aggregation indicate a high kinetic barrier to solvent release from polar side chains (153). The favorable interaction of water with the charged and polar residues on the BirA dimerization surface should result in a similarly high barrier to dimer formation.

The solvent isotope effect on holoBirA dimerization thermodynamics is significantly larger than that measured for other protein association reactions. For example, measurements of the dimer-octamer equilibrium of β-lactoglobulin yielded enthalpies of -64 and -69 kcal/mol in H_2O and D_2O respectively (150). Indeed, solvent isotope effects of the magnitude observed in this work have been reported only for protein folding, which may indicate a more significant role for solvent release in the holoBirA dimerization reaction than for other protein association reactions (154). Solvent isotope effects on protein folding and association reactions have previously been ascribed to stronger bonding resulting from exchange of deuterium for hydrogen in hydrogen bonding partners (151, 155). Alternatively, the decreased flexibility observed for some proteins in D_2O may alter interaction thermodynamics (156, 157). For BirA the absence of a solvent isotope effect on coupling between adenylate binding and dimerization points to solvent release as the major source of the distinct thermodynamic profiles measured in the two solvents.

4.4.3 Practical implications for hydrogen-deuterium exchange studies

Hydrogen-deuterium exchange measurements, detected by either mass spectrometry or NMR spectroscopy, are used to study protein folding, dynamics and interactions (158, 159). The potential for solvent isotope perturbation of protein function has previously been suggested (160, 161). Results reported in this work, which indicate that these effects can be large in magnitude, underscore the importance of performing appropriate controls to determine the functional effects of transferring a biomolecular system from H_2O to D_2O.

4.5 Materials and Methods

4.5.1 Chemicals and biochemicals

All chemicals used in buffer preparation were at least reagent grade. The bio-5'-AMP was synthesized and purified as previously described (61, 162). The standard buffer (10 mM Tris HCl, 200 mM KCl, and 2.5 mM $MgCl_2$) prepared in D_2O was adjusted to a pD of 7.5 using a meter reading of 7.1 to correct for the effect of deuterium on glass electrodes (163).

4.5.2 Protein preparation

BirA protein variants were prepared and purified as previously described and the purity of each was estimated to be >95% based on coomassie brilliant blue staining of samples subjected to SDS-polyacrylamide gel electrophoresis (77, 80).

Protein concentrations were determined spectrophotometrically at 280 nm using a molar extinction coefficient of 47,510 M^{-1} cm^{-1} calculated from the amino acid composition (133). The fractional activity of each protein is >90% as determined by stoichiometric binding titrations with bio-5'-AMP monitored by steady-state fluorescence spectroscopy (61).

4.5.3 Sedimentation equilibrium

The self-association of each BirA variant complexed with bio-5'-AMP was measured by equilibrium analytical ultracentrifugation using a Beckman Coulter Optima XL-I Analytical Ultracentrifuge. Proteins were first exchanged into SB:D_2O using Micro Bio-Spin 6 chromatography columns (Bio-Rad) as described by the manufacturer. For each measurement, protein prepared at three different concentrations was combined at stoichiometric conditions with bio-5'-AMP at a final molar ratio of 1:1.5. Equilibrium dissociation constants for binding of the ligand to the BirA variants range from picomolar to nanomolar (77, 80). For each variant and/or solvent condition the protein concentrations employed varied depending on the strength of the dimerization reaction. Lower concentrations were employed for tightly dimerizing systems and higher concentrations for weakly dimerizing systems, with the goal of optimizing representation of the dimer and monomer species in the concentration versus radial distance profiles. Samples were centrifuged in cells equipped with 12 mm six-hole or 3 mm two-hole charcoal-filled Epon centerpieces with quartz windows in a four-hole An-60 rotor (Beckman Coulter). Prior to centrifugation, the filled sample cells and rotor were incubated at the specified

temperature for 1 h. Centrifugation was carried out at three rotor speeds ranging from 18,000 to 24,000 rpm. After reaching equilibrium, scans were acquired with a step size of 0.001 cm with 5 averages at either 295 or 300 nm. At these wavelengths the contribution from the absorbance of the adenosine moiety of bio-5'-AMP was avoided and, for samples prepared at very high protein concentrations, the total absorbance was in the linear range of the detection system.

4.5.4 Data analysis

The absorbance *versus* radius profiles obtained for each scan were analyzed with WinNONLIN (108) using a single species model to obtain σ, the reduced molecular weight, from which the weight average molecular weight was calculated using the following equation:

$$\sigma = \frac{kM(1-\frac{\bar{\nu}}{k}\rho)\omega^2}{RT} \quad (1)$$

where M is the molecular weight, k is the proportionality constant for the associated increase in molecular weight and decrease in the partial specific volume that occurs as a result of deuterium exchange into the protein (164), $\bar{\nu}$ is the protein partial specific volume, ρ is the buffer density, ω is the angular velocity of the rotor, R is the gas constant and T is the temperature in kelvin. On the basis of a buffer composition of 95% (v/v) D_2O, a value for k of 1.0147 was used (165). The partial specific volume of the BirA monomer is 0.755 mL/g (73) and the density was calculated from

the buffer composition at the appropriate temperature using Sednterp (http://sednterp.unh.edu/).

Absorbance *versus* radius profiles were also globally analyzed to obtain the equilibrium association constant for dimerization, K_a, using the following monomer-dimer model:

$$C_t(r) = \delta + C_{mon}(r_o)e^{\sigma_{mon}(\frac{r^2}{2}-\frac{r_o^2}{2})} + K_a(C_{mon}(r_o))^2 e^{2\sigma_{mon}(\frac{r^2}{2}-\frac{r_o^2}{2})} \quad (2)$$

in which C_t is the total concentration at each radial position r, δ is the baseline offset, which was allowed to float in the analysis, and $C_{mon}(r_o)$ is the monomer concentration at reference radial position r_o. For measurements performed at 20°C, the σ_{mon} value used in calculations performed for wild-type and variant proteins is the measured value in D_2O buffer for wild-type apoBirA. In analysis of data acquired at other temperatures, the σ_{mon} value was adjusted for accompanying changes in the solvent density. In all analyses the reduced molecular weight of the dimer was assumed to be twice that of the monomer. The quality of each fit was assessed from the magnitude of the square root of the variance and the distribution of the residuals of the fit about zero.

Acknowledgements

Thank you to Dr. Adikaram for the preparation, purification, and characterization in H_2O of the T195A variant. Thank you to Dr. Naganathan for the preparation, purification, and characterization in H_2O of the V219A variant.

Chapter 5: Summary and Future Directions

The work presented in this manuscript focuses on characterizing the sequence-function relationships of the ligand-linked folding transitions of the *E. coli* biotin repressor. The importance of the hydrophobic network in several BirA functions was systematically investigated using multiple combinations of alanine substitutions. Both equilibrium and kinetic measurements were used to analyze the effects of sequence variation on synthesis, ligand binding, and self-association. All alanine substituted proteins exhibited large effects on the synthesis and binding of bio-5'-AMP. In agreement with these results, folding of the ABL was compromised for the majority of network variants. Additionally, there were modest effects on the homodimerization energetics. Combined, the results reveal that in order to achieve the optimum functional response, the appropriate packing of side chains in the hydrophobic network is required. This suggests that the number of combinations of hydrophobic side chains in this region that can achieve the functional responses is potentially large, in agreement with the observed sequence variation among biotin protein ligases. Despite the perceived flexible sequence requirements, it is clear from the results that the sequence of this region for a single biotin protein ligase is highly constrained. This is particularly evident in the large perturbation to bio-5'-AMP synthesis observed for the majority of variants studied. Corepressor synthesis and the subsequent transfer of biotin to BCCP are necessary for cell viability, thus highlighting the importance of this region.

As previously mentioned, there is no sequence conservation in the cluster region between BirA and the BPLs from *Staphylococcus aureus* and *Mycobacterium tuberculosis*. In spite of the lack of sequence conservation a similar structural architecture is observed in the hydrophobic network formed by the ABL and BBL (72, 78, 79). This observation suggests that the hydrophobic network residues for various ligases are free to diverge as long as the appropriate packing is achieved. To determine if this is a general phenomenon applicable to other biotin protein ligases, and therefore other systems, further mutagenic and functional studies are needed. This could be accomplished through analogous experiments presented in this work on the *S. aureus* or *M. tuberculosis* BPL. The resulting studies would further our understanding of the sequence requirements of looped regions involved in structural transitions. This could also further drug development focused on inhibitors targeting the ABL by taking advantage of the sequence divergence observed between bacterial homologs and the human ligase.

Although allosteric regulation has been proposed to operate through disorder-to-order transitions, limited experimental evidence of the phenomenon has been presented. The structural and functional analysis of the G142A variant supports a mechanism for long-distance allosteric communication achieved through the disorder-to-order transitions on the ligand binding and dimerization surfaces. Perturbations to dimerization energetics upon substitution of glycine 142 with alanine are consistent with loss of coupling between adenylate binding and homodimerization. The resulting adenylate-bound monomer displays disorder in two loops comprised of residues 193-

199 and 211-234 that are ordered in the wild type liganded dimeric structure. The loss of coupling between corepressor binding and self-association appears to be caused by the disruption of disorder-to-order transitions on distant surfaces of the protein monomer. However, what is not known is if the disruption of ABL folding through alanine substitution perturbs dimerization by disrupting folding on the dimerization surface. This can be investigated- through structural studies of alanine substituted proteins on the ligand binding surface for which coupling is disrupted. The BirA variants V219A and M211A display the largest defects in coupling free energy of 1.5 kcal/mol. Preliminary data on V219A at ~3.5 Å resolution has been obtained. However, crystallization conditions must be further optimized to improve the resolution. Crystallographic data obtained for either variant could potentially shed light on the structural origins of the loss of ligand binding free energy possibly arising from alternative packing of the hydrophobic network. In all, these results highlight the sensitivity of loop regions to changes in sequence and how these, sometimes subtle, variations can have dramatic effects on the functional and dynamic properties of the protein as a whole.

The mechanism of allosteric communication in BirA is not readily apparent upon examination of the crystallographically determined structures. Comparison of the structures of the wild type protein in the different liganded states with the G142A structure does not reveal any obvious conformational changes connecting the two coupled sites. This suggests that the mechanism for allosteric communication is not a rigid-structural transition but one involving dynamics. Determination of the

mechanism requires analysis using techniques that can provide structural information over various timescales. Hydrogen-deuterium exchange mass spectrometry (HX-MS) or NMR spectroscopy present powerful tools to reveal information about protein dynamics (166, 167). Hydrogen-deuterium exchange detected by either mass spectrometry or by NMR rely on the spontaneous exchange of amide backbone hydrogens with deuterium. While in deuterium buffer the self-assembly energetics of BirA have been shown to be enhanced approximately 10-fold with the equilibrium dissociation constant for dimerization in H_2O and D_2O of approximately 6 μM and 500 nM, respectively (168). The enhanced assembly energetics do not allow for the direct comparison of G142A and wild type BirA in the monomeric state. Alternatively, G142A can be compared with a weakly dimerizing variant that displays no defect in ligand binding such as I280A. With the appropriate experimental setup hydrogen-deuterium exchange detected by either method offers a range of dynamic information in various liganded states from milliseconds to seconds (167, 169). Further, the timescales can be extended from picoseconds to seconds using a range of NMR spectroscopy techniques (170). Additionally, spectral crowding frequently encountered for larger proteins such as BirA can be mitigated through various isotopic labeling techniques such as selective methyl labeling with ^{13}C (171).

Computational methods have been used in structural biology to complement experimental methods (172). Molecular dynamic (MD) simulations enable one to maintain a network of atomic positions and forces to accurately characterize the structure and dynamics of a model system. This is a powerful approach as the

mechanisms of allosteric communication involve dynamic motions on a shorter timescale, often beyond the limited resolution of experimental techniques that are amenable to a particular system. This technique used in conjunction with the extensive experimental knowledge on BirA could provide insight into the mechanism by which these disorder-to-order transitions communicate. More specifically, MD simulations could be used to monitor the correlated loop motions in BirA and determine the population of ensemble members (172). Further, these simulations could help map the allosteric pathway and provide insight into the motion of the structured region that connects the two loops (173).

The work presented in this thesis emphasizes the versatile roles that loops play in a range of processes including binding, catalysis, and allostery. This versatility can be partly attributed to the conformational heterogeneity that loop regions enjoy. However, because of this flexibility significant challenges remain in the design and utilization of looped regions to modulate function. Even with the increasing success researchers have had in protein structure prediction, the accurate prediction of loop structures remains challenging (174, 175). These studies have shed light on the sequence determinants for binding, catalysis, and protein:protein interactions. Furthermore, they have shown that the principles that govern protein folding can be applied to loop regions that undergo a binding-linked folding transition. Lastly, these results demonstrate that distal disorder-to-order transitions can be involved in long-distance allosteric communication and the modulation of sequence in these regions can dramatically affect function and energetic coupling.

Bibliography

1. Dzeja P & Terzic A (2009) Adenylate kinase and AMP signaling networks: metabolic monitoring, signal communication and body energy sensing. *Int J Mol Sci* 10(4):1729-1772.
2. Schnell JR, Dyson HJ, & Wright PE (2004) Structure, dynamics, and catalytic function of dihydrofolate reductase. *Annu Rev Biophys Biomol Struct* 33:119-140.
3. Malabanan MM, Amyes TL, & Richard JP (2010) A role for flexible loops in enzyme catalysis. *Curr Opin Struct Biol* 20(6):702-710.
4. Joerger AC & Fersht AR (2010) The tumor suppressor p53: from structures to drug discovery. *Cold Spring Harb Perspect Biol* 2(6):a000919.
5. Stefanis L (2012) α-Synuclein in Parkinson's disease. *Cold Spring Harb Perspect Med* 2(2):a009399.
6. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, & Jones DT (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* 337(3):635-645.
7. Uversky VN, *et al.* (2014) Pathological unfoldomics of uncontrolled chaos: intrinsically disordered proteins and human diseases. *Chem Rev* 114(13):6844-6879.
8. Wright PE & Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* 293(2):321-331.
9. Reichheld SE, Yu Z, & Davidson AR (2009) The induction of folding cooperativity by ligand binding drives the allosteric response of tetracycline repressor. *Proc Natl Acad Sci U S A* 106(52):22263-22268.
10. Hilser VJ & Thompson EB (2007) Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. *Proc Natl Acad Sci U S A* 104(20):8311-8315.
11. Fischer E (1894) Einfluss der configuration auf die wirkung der enzyme. *Berichte der deutschen chemischen Gesellschaft* 27(3):2985–2993.
12. Koshland DE (1958) Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc Natl Acad Sci U S A* 44(2):98-104.
13. Koshland DE (1994) The key-lock theory and the induced fit theory. *Angew. Chem. Inl. Ed. Engl.* 33:2375-2378.
14. Tsai CJ, Ma B, & Nussinov R (1999) Folding and binding cascades: shifts in energy landscapes. *Proc Natl Acad Sci U S A* 96(18):9970-9972.
15. Tsai CJ, Kumar S, Ma B, & Nussinov R (1999) Folding funnels, binding funnels, and protein function. *Protein Sci* 8(6):1181-1190.
16. Kumar S, Ma B, Tsai CJ, Sinha N, & Nussinov R (2000) Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci* 9(1):10-19.
17. Arora K & Brooks CL (2007) Large-scale allosteric conformational transitions of adenylate kinase appear to involve a population-shift mechanism. *Proc Natl Acad Sci U S A* 104(47):18496-18501.

18. Rogers JM, Wong CT, & Clarke J (2014) Coupled folding and binding of the disordered protein PUMA does not require particular residual structure. *J Am Chem Soc* 136(14):5197-5200.
19. Müller CW, Schlauderer GJ, Reinstein J, & Schulz GE (1996) Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. *Structure* 4(2):147-156.
20. Müller CW & Schulz GE (1992) Structure of the complex between adenylate kinase from Escherichia coli and the inhibitor Ap5A refined at 1.9 A resolution. A model for a catalytic transition state. *J Mol Biol* 224(1):159-177.
21. DeLano WL (2002) The PyMOL Molecular Graphics System (Schrödinger, LLC), Version 1.2r3pre.
22. Wolf-Watz M, *et al.* (2004) Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nat Struct Mol Biol* 11(10):945-949.
23. Demarest SJ, *et al.* (2002) Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature* 415(6871):549-553.
24. Turoverov KK, Kuznetsova IM, & Uversky VN (2010) The protein kingdom extended: ordered and intrinsically disordered proteins, their folding, supramolecular complex formation, and aggregation. *Prog Biophys Mol Biol* 102(2-3):73-84.
25. Dogan J, Schmidt T, Mu X, Engström Å, & Jemth P (2012) Fast association and slow transitions in the interaction between two intrinsically disordered protein domains. *J Biol Chem* 287(41):34316-34324.
26. Jemth P, Mu X, Engström Å, & Dogan J (2014) A frustrated binding interface for intrinsically disordered proteins. *J Biol Chem* 289(9):5528-5533.
27. Lawrence CW, Kumar, S., Noid, W.G., Showalter, S.A. (2014) Role of ordered proteins in the folding-upon-binding of intrinsically disordered proteins. (. *Phys. Chem. Lett.*), pp 833-838.
28. Mangelsdorf DJ, *et al.* (1995) The nuclear receptor superfamily: the second decade. *Cell* 83(6):835-839.
29. Kjaergaard M, Teilum K, & Poulsen FM (2010) Conformational selection in the molten globule state of the nuclear coactivator binding domain of CBP. *Proc Natl Acad Sci U S A* 107(28):12535-12540.
30. Dogan J, Mu X, Engström Å, & Jemth P (2013) The transition state structure for coupled binding and folding of disordered protein domains. *Sci Rep* 3:2076.
31. Fenton AW (2008) Allostery: an illustrated definition for the 'second secret of life'. *Trends Biochem Sci* 33(9):420-425.
32. Changeux JP (1961) The feedback control mechanisms of biosynthetic L-threonine deaminase by L-isoleucine. *Cold Spring Harb Symp Quant Biol* 26:313-318.
33. Monod J, Wyman J, & Changeux JP (1965) On the nature of allosteric transitions: a plausible model. *J Mol Biol* 12:88-118.
34. Koshland DE, Némethy G, & Filmer D (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* 5(1):365-385.

35. Macol CP, Tsuruta H, Stec B, & Kantrowitz ER (2001) Direct structural evidence for a concerted allosteric transition in Escherichia coli aspartate transcarbamoylase. *Nat Struct Biol* 8(5):423-426.
36. Jones ME, Spector, L., Lipmann, F. (1955) Carbamyl phosphate, the carbamyl donor in enzymatic citrulline synthesis. *J. Am. Chem. Soc.* 77:819-820.
37. Lowenstein JM, Cohen, P.P. (1956) Studies on the biosynthesis of carbamylaspartic acid. *J. Biol. Chem* 235:57-78.
38. Reichard P, Hanshoff, G. (1956) Aspartate carbamyl transferase from escherichia coli. *Acta. Chem. Scand.* 10:548-560.
39. Gerhart JC & Schachman HK (1965) Distinct subunits for the regulation and catalytic activity of aspartate transcarbamylase. *Biochemistry* 4(6):1054-1062.
40. Wiley DC & Lipscomb WN (1968) Crystallographic determination of symmetry of aspartate transcarbamylase. *Nature* 218(5147):1119-1121.
41. Ke HM, Lipscomb WN, Cho YJ, & Honzatko RB (1988) Complex of N-phosphonacetyl-L-aspartate with aspartate carbamoyltransferase. X-ray refinement, analysis of conformational changes and catalytic and allosteric mechanisms. *J Mol Biol* 204(3):725-747.
42. Stevens RC, Gouaux JE, & Lipscomb WN (1990) Structural consequences of effector binding to the T state of aspartate carbamoyltransferase: crystal structures of the unligated and ATP- and CTP-complexed enzymes at 2.6-A resolution. *Biochemistry* 29(33):7691-7701.
43. Motlagh HN, Wrabl JO, Li J, & Hilser VJ (2014) The ensemble nature of allostery. *Nature* 508(7496):331-339.
44. Cooper A & Dryden DT (1984) Allostery without conformational change. A plausible model. *Eur Biophys J* 11(2):103-109.
45. Meyer LM, Miller FR, Rowen MJ, Bock G, & Rutzky J (1950) Treatment of acute leukemia with amethopterin (4-amino, 10-methyl pteroyl glutamic acid). *Acta Haematol* 4(3):157-167.
46. Cameron CE & Benkovic SJ (1997) Evidence for a functional role of the dynamics of glycine-121 of Escherichia coli dihydrofolate reductase obtained from kinetic analysis of a site-directed mutant. *Biochemistry* 36(50):15792-15800.
47. Sawaya MR & Kraut J (1997) Loop and subdomain movements in the mechanism of Escherichia coli dihydrofolate reductase: crystallographic evidence. *Biochemistry* 36(3):586-603.
48. Epstein DM, Benkovic SJ, & Wright PE (1995) Dynamics of the dihydrofolate reductase-folate complex: catalytic sites and regions known to undergo conformational change exhibit diverse dynamical features. *Biochemistry* 34(35):11037-11048.
49. Boehr DD, *et al.* (2013) A distal mutation perturbs dynamic amino acid networks in dihydrofolate reductase. *Biochemistry* 52(27):4605-4619.
50. Liu J, *et al.* (2006) Intrinsic disorder in transcription factors. *Biochemistry* 45(22):6873-6888.
51. Zhao D, Arrowsmith CH, Jia X, & Jardetzky O (1993) Refined solution structures of the Escherichia coli trp holo- and aporepressor. *J Mol Biol* 229(3):735-746.

52. Zheng Z, Czaplicki J, & Jardetzky O (1995) Backbone dynamics of trp repressor studied by 15N NMR relaxation. *Biochemistry* 34(15):5212-5223.
53. Popovych N, Tzeng SR, Tonelli M, Ebright RH, & Kalodimos CG (2009) Structural basis for cAMP-mediated allosteric control of the catabolite activator protein. *Proc Natl Acad Sci U S A* 106(17):6927-6932.
54. Kisker C, Hinrichs W, Tovar K, Hillen W, & Saenger W (1995) The complex formed between Tet repressor and tetracycline-Mg2+ reveals mechanism of antibiotic resistance. *J Mol Biol* 247(2):260-280.
55. Orth P, *et al.* (1998) Conformational changes of the Tet repressor induced by tetracycline trapping. *J Mol Biol* 279(2):439-447.
56. Orth P, Schnappinger D, Hillen W, Saenger W, & Hinrichs W (2000) Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nat Struct Biol* 7(3):215-219.
57. Beckett D (2007) Biotin sensing: universal influence of biotin status on transcription. *Annu Rev Genet* 41:443-464.
58. Lane MD, Rominger KL, Young DL, & Lynen F (1964) The enzymatic synthesis of holotranscarboxylase from apotranscarboxylase and (+)-biotin. II. Investigation of the reaction mechanism. *J Biol Chem* 239:2865-2871.
59. Knowles JR (1989) The mechanism of biotin-dependent enzymes. *Annu Rev Biochem* 58:195-221.
60. Attwood PV & Wallace JC (2002) Chemical and catalytic mechanisms of carboxyl transfer reactions in biotin-dependent enzymes. *Acc Chem Res* 35(2):113-120.
61. Abbott J & Beckett D (1993) Cooperative binding of the Escherichia coli repressor of biotin biosynthesis to the biotin operator sequence. *Biochemistry* 32(37):9649-9656.
62. Otsuka A & Abelson J (1978) The regulatory region of the biotin operon in Escherichia coli. *Nature* 276(5689):689-694.
63. Streaker ED & Beckett D (2003) Coupling of protein assembly and DNA binding: biotin repressor dimerization precedes biotin operator binding. *J Mol Biol* 325(5):937-948.
64. Li SJ & Cronan JE (1993) Growth rate regulation of Escherichia coli acetyl coenzyme A carboxylase, which catalyzes the first committed step of lipid biosynthesis. *J Bacteriol* 175(2):332-340.
65. Cronan JE (1988) Expression of the biotin biosynthetic operon of Escherichia coli is regulated by the rate of protein biotination. *J Biol Chem* 263(21):10332-10336.
66. Streaker ED, Gupta A, & Beckett D (2002) The biotin repressor: thermodynamic coupling of corepressor binding, protein assembly, and sequence-specific DNA binding. *Biochemistry* 41(48):14263-14271.
67. Brown PH, Cronan JE, Grøtli M, & Beckett D (2004) The biotin repressor: modulation of allostery by corepressor analogs. *J Mol Biol* 337(4):857-869.
68. Zhao H, Streaker E, Pan W, & Beckett D (2007) Protein-protein interactions dominate the assembly thermodynamics of a transcription repression complex. *Biochemistry* 46(47):13667-13676.

69. Xu Y, Nenortas E, & Beckett D (1995) Evidence for distinct ligand-bound conformational states of the multifunctional Escherichia coli repressor of biotin biosynthesis. *Biochemistry* 34(51):16624-16631.
70. Wilson KP, Shewchuk LM, Brennan RG, Otsuka AJ, & Matthews BW (1992) Escherichia coli biotin holoenzyme synthetase/bio repressor crystal structure delineates the biotin- and DNA-binding domains. *Proc Natl Acad Sci U S A* 89(19):9257-9261.
71. Weaver LH, Kwon K, Beckett D, & Matthews BW (2001) Corepressor-induced organization and assembly of the biotin repressor: a model for allosteric activation of a transcriptional regulator. *Proc Natl Acad Sci U S A* 98(11):6045-6050.
72. Wood ZA, Weaver LH, Brown PH, Beckett D, & Matthews BW (2006) Co-repressor induced order and biotin repressor dimerization: a case for divergent followed by convergent evolution. *J Mol Biol* 357(2):509-523.
73. Eisenstein E & Beckett D (1999) Dimerization of the Escherichia coli biotin repressor: corepressor function in protein assembly. *Biochemistry* 38(40):13077-13084.
74. Saha RP, Bahadur RP, Pal A, Mandal S, & Chakrabarti P (2006) ProFace: a server for the analysis of the physicochemical features of protein-protein interfaces. *BMC Struct Biol* 6:11.
75. Kwon K & Beckett D (2000) Function of a conserved sequence motif in biotin holoenzyme synthetases. *Protein Sci* 9(8):1530-1539.
76. Kwon K, Streaker ED, Ruparelia S, & Beckett D (2000) Multiple disordered loops function in corepressor-induced dimerization of the biotin repressor. *J Mol Biol* 304(5):821-833.
77. Naganathan S & Beckett D (2007) Nucleation of an allosteric response via ligand-induced loop folding. *J Mol Biol* 373(1):96-111.
78. Pendini NR, *et al.* (2013) Structural characterization of Staphylococcus aureus biotin protein ligase and interaction partners: an antibiotic target. *Protein Sci* 22(6):762-773.
79. Ma Q, Akhter Y, Wilmanns M, & Ehebauer MT (2014) Active site conformational changes upon reaction intermediate biotinyl-5'-AMP binding in biotin protein ligase from Mycobacterium tuberculosis. *Protein Sci* 23(7):932-939.
80. Adikaram PR & Beckett D (2012) Functional versatility of a single protein surface in two protein:protein interactions. *J Mol Biol* 419(3-4):223-233.
81. Rodionov DA, Mironov AA, & Gelfand MS (2002) Conservation of the biotin regulon and the BirA regulatory signal in Eubacteria and Archaea. *Genome Res* 12(10):1507-1516.
82. Attwood PV & Keech DB (1984) Pyruvate carboxylase. *Curr Top Cell Regul* 23:1-55.
83. Soares da Costa TP, *et al.* (2012) Selective inhibition of biotin protein ligase from Staphylococcus aureus. *J Biol Chem* 287(21):17823-17832.
84. Purushothaman S, Gupta G, Srivastava R, Ramu VG, & Surolia A (2008) Ligand specificity of group I biotin protein ligase of Mycobacterium tuberculosis. *PLoS One* 3(5):e2320.

85. Gupta V, *et al.* (2010) Structural ordering of disordered ligand-binding loops of biotin protein ligase into active conformations as a consequence of dehydration. *PLoS One* 5(2):e9222.
86. Risueno RM, Gil D, Fernandez E, Sanchez-Madrid F, & Alarcon B (2005) Ligand-induced conformational change in the T-cell receptor associated with productive immune synapses. *Blood* 106(2):601-608.
87. Wright PE & Dyson HJ (2009) Linking folding and binding. *Curr Opin Struct Biol* 19(1):31-38.
88. Tompa P (2011) Unstructural biology coming of age. *Curr Opin Struct Biol* 21(3):419-425.
89. Larion M, Salinas RK, Bruschweiler-Li L, Miller BG, & Bruschweiler R (2012) Order-disorder transitions govern kinetic cooperativity and allostery of monomeric human glucokinase. *PLoS Biol* 10(12):e1001452.
90. Barker DF, and Campbell, A. M. (1981) The birA gene of *Escherichia coli* encodes a biotin holoenzyme synthetase. *J. Mol. Biol* 146:451-467.
91. Barker DF & Campbell AM (1981) Genetic and biochemical characterization of the birA gene and its product: evidence for a direct role of biotin holoenzyme synthetase in repression of the biotin operon in Escherichia coli. *J Mol Biol* 146(4):469-492.
92. Cronan JE, Jr. (1989) The *E.coli* bio operon:transcriptional repression by an essential protein modification enzyme. *Cell* 58:427-429.
93. Xu Y & Beckett D (1994) Kinetics of biotinyl-5'-adenylate synthesis catalyzed by the Escherichia coli repressor of biotin biosynthesis and the stability of the enzyme-product complex. *Biochemistry* 33(23):7354-7360.
94. Weaver LH, Kwon K, Beckett D, & Matthews BW (2001) Competing protein:protein interactions are proposed to control the biological switch of the E. coli biotin repressor. *Protein Science* 10(12):2618-2622.
95. Bagautdinov B, Matsuura Y, Bagautdinova S, & Kunishima N (2008) Protein biotinylation visualized by a complex structure of biotin protein ligase with a substrate. *J Biol Chem* 283(21):14739-14750.
96. Streaker ED, and Beckett, D. (2003) Coupling of protein assembly and DNA binding:biotin repressor dimerization preceeds biotin operator binding. *J. Mol. Biol* 325:937-948.
97. Lim WA & Sauer RT (1989) Alternative packing arrangements in the hydrophobic core of lambda repressor. *Nature* 339(6219):31-36.
98. Lim WA & Sauer RT (1991) The role of internal packing interactions in determining the structure and stability of a protein. *J Mol Biol* 219(2):359-376.
99. Katz B & Kossiakoff AA (1990) Crystal structures of subtilisin BPN' variants containing disulfide bonds and cavities: concerted structural rearrangements induced by mutagenesis. *Proteins* 7(4):343-357.
100. Sondek J & Shortle D (1990) Accommodation of single amino acid insertions by the native state of staphylococcal nuclease. *Proteins* 7(4):299-305.
101. Lesk AM & Chothia C (1980) How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. *J Mol Biol* 136(3):225-270.

102. Chen J & Stites WE (2001) Energetics of side chain packing in staphylococcal nuclease assessed by systematic double mutant cycles. *Biochemistry* 40(46):14004-14011.
103. Hurley JH, Baase WA, & Matthews BW (1992) Design and structural analysis of alternative hydrophobic core packing arrangements in bacteriophage T4 lysozyme. *J Mol Biol* 224(4):1143-1159.
104. Larkin MA, *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21):2947-2948.
105. Goujon M, *et al.* (2010) A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res* 38(Web Server issue):W695-699.
106. Waterhouse AM, Procter JB, Martin DM, Clamp M, & Barton GJ (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25(9):1189-1191.
107. Brown PH & Beckett D (2005) Use of binding enthalpy to drive an allosteric transition. *Biochemistry* 44(8):3112-3121.
108. Johnson ML, Correia JJ, Yphantis DA, & Halvorson HR (1981) Analysis of data from the analytical ultracentrifuge by nonlinear least-squares techniques. *Biophys J* 36(3):575-588.
109. Buoncristiani MR, Howard PK, & Otsuka AJ (1986) DNA-binding and enzymatic domains of the bifunctional biotin operon repressor (BirA) of Escherichia coli. *Gene* 44(2-3):255-261.
110. Vagin A & Teplyakov A (1997) MOLREP: an automated program for molecular replacement. *J Appl Crystallogr* 30:1022-1025.
111. Vagin A & Teplyakov A (2010) Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* 66(Pt 1):22-25.
112. Roberts E, Eargle J, Wright D, & Luthey-Schulten Z (2006) MultiSeq: unifying sequence and structure data for evolutionary analysis. *Bmc Bioinformatics* 7.
113. Humphrey W, Dalke A, & Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph Model* 14(1):33-38.
114. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126-2132.
115. Terwilliger TC, *et al.* (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr* 64(Pt 1):61-69.
116. Murshudov GN, *et al.* (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 67(Pt 4):355-367.
117. Eginton C, Naganathan S, & Beckett D (2014) Sequence-function relationships in folding upon binding. *Protein Sci*.
118. Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol* 276:307-326.
119. Winn MD, *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D* 67:235-242.
120. Vagin AA, *et al.* (2004) REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D* 60:2184-2195.

121. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D* 60:2126-2132.
122. Terwilliger TC, *et al.* (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D* 64:61-69.
123. Adams PD, *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213-221.
124. Afonine PV, *et al.* (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* 68(Pt 4):352-367.
125. Painter J & Merritt EA (2005) A molecular viewer for the analysis of TLS rigid-body motion in macromolecules. *Acta Crystallogr D Biol Crystallogr* 61(Pt 4):465-471.
126. Painter J & Merritt EA (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr* 62(Pt 4):439-450.
127. Davis IW, *et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35:W375-W383.
128. Chen VB, *et al.* (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D* 66:12-21.
129. Urzhumtseva L, Afonine PV, Adams PD, & Urzhumtsev A (2009) Crystallographic model quality at a glance. *Acta Crystallogr D* 65:297-300.
130. Laskowski RA, Macarthur MW, Moss DS, & Thornton JM (1993) Procheck - a Program to Check the Stereochemical Quality of Protein Structures. *J Appl Crystallogr* 26:283-291.
131. Vehlow C, *et al.* (2011) CMView: Interactive contact map visualization and analysis. *Bioinformatics* 27(11):1573-1574.
132. Wallace AC, Laskowski RA, & Thornton JM (1995) Ligplot - a Program to Generate Schematic Diagrams of Protein Ligand Interactions. *Protein Eng* 8(2):127-134.
133. Gill SC & von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Analytical Biochemistry* 182:319-326.
134. Hodkinson PS, *et al.* (2007) Mammalian NOTCH-1 activates beta1 integrins via the small GTPase R-Ras. *J Biol Chem* 282(39):28991-29001.
135. Dickinson DJ, Robinson DN, Nelson WJ, & Weis WI (2012) alpha-catenin and IQGAP regulate myosin localization to control epithelial tube morphogenesis in Dictyostelium. *Dev Cell* 23(3):533-546.
136. Layer JH, Miller SG, & Weil PA (2010) Direct transactivator-transcription factor IID (TFIID) contacts drive yeast ribosomal protein gene transcription. *J Biol Chem* 285(20):15489-15499.
137. Thirumalai D, Reddy G, & Straub JE (2012) Role of water in protein aggregation and amyloid polymorphism. *Acc Chem Res* 45(1):83-92.
138. Stranges PB & Kuhlman B (2013) A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds. *Protein Sci* 22(1):74-82.

139. Khalil MT & Lauffer MA (1967) Polymerization-depolymerization of tobacco mosaic virus protein. X. Effect of D20. *Biochemistry* 6(8):2474-2480.
140. Baldwin RL (2010) Desolvation penalty for burying hydrogen-bonded peptide groups in protein folding. *J Phys Chem B* 114(49):16223-16227.
141. Krissinel E & Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372(3):774-797.
142. Courtenay ES, Capp MW, Anderson CF, & Record MT, Jr. (2000) Vapor pressure osmometry studies of osmolyte-protein interactions: implications for the action of osmoprotectants in vivo and for the interpretation of "osmotic stress" experiments in vitro. *Biochemistry* 39(15):4455-4471.
143. Kornblatt MJ, Kornblatt JA, & Hui Bon Hoa G (1993) The role of water in the dissociation of enolase, a dimeric enzyme. *Arch Biochem Biophys* 306(2):495-500.
144. Zhai Y, Okoro L, Cooper A, & Winter R (2011) Applications of pressure perturbation calorimetry in biophysical studies. *Biophys Chem* 156(1):13-23.
145. Chervenak MC & Toone EJ (1994) A Direct Measure of the Contribution of Solvent Reorganization to the Enthalpy of Ligand-Binding. *J Am Chem Soc* 116(23):10533-10539.
146. Lee JJ & Berns DS (1968) Protein aggregation. The effect of deuterium oxide on large protein aggregates of C-phycocyanin. *Biochem J* 110(3):465-470.
147. Woodfin BM, Henderson RF, & Henderson TR (1970) Effects of D2O on the association-dissociation equilibrium in subunit proteins. *J Biol Chem* 245(15):3733-3737.
148. Aune KC, Goldsmith LC, & Timasheff SN (1971) Dimerization of alpha-chymotrypsin. II. Ionic strength and temperature dependence. *Biochemistry* 10(9):1617-1622.
149. Paglini S & Lauffer MA (1968) Polymerization-depolymerization of tobacco mosaic virus protein. XI. Osmotic pressure studies of solutions in water and in deuterium. *Biochemistry* 7(5):1827-1835.
150. Baghurst PA, Nichol LW, & Sawyer WH (1972) The effect of D 2 O on the association of -lactoglobulin A. *J Biol Chem* 247(10):3199-3204.
151. Nemethy G & Scheraga HA (1964) Structure of Water + Hydrophobic Bonding in Proteins .4. Thermodynamic Properties of Liquid Deuterium Oxide. *J Chem Phys* 41(3):680-&.
152. Zhao H & Beckett D (2008) Kinetic partitioning between alternative protein-protein interactions controls a transcriptional switch. *J Mol Biol* 380(1):223-236.
153. Reddy G, Straub JE, & Thirumalai D (2010) Dry amyloid fibril assembly in a yeast prion peptide is mediated by long-lived structures containing water wires. *Proc Natl Acad Sci U S A* 107(50):21459-21464.
154. Makhatadze GI, Clore GM, & Gronenborn AM (1995) Solvent isotope effect and protein stability. *Nat Struct Biol* 2(10):852-855.
155. Cho Y, *et al.* (2009) Hydrogen bonding of beta-turn structure is stabilized in D(2)O. *J Am Chem Soc* 131(42):15188-15193.
156. Cioni P & Strambini GB (2002) Effect of heavy water on protein flexibility. *Biophys J* 82(6):3246-3253.

157. Tehei M, Madern D, Pfister C, & Zaccai G (2001) Fast dynamics of halophilic malate dehydrogenase and BSA measured by neutron scattering under various solvent conditions influencing protein stability. *Proc Natl Acad Sci U S A* 98(25):14356-14361.
158. Konermann L, Pan J, & Liu YH (2011) Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem Soc Rev* 40(3):1224-1234.
159. Lee YH & Goto Y (2012) Kinetic intermediates of amyloid fibrillation studied by hydrogen exchange methods with nuclear magnetic resonance. *Biochim Biophys Acta* 1824(12):1307-1323.
160. Sackett DL, Chernomordik V, Krueger S, & Nossal R (2003) Use of small-angle neutron scattering to study tubulin polymers. *Biomacromolecules* 4(2):461-467.
161. Prasannan CB, Artigues A, & Fenton AW (2011) Monitoring allostery in D2O: a necessary control in studies using hydrogen/deuterium exchange to characterize allosteric regulation. *Anal Bioanal Chem* 401(3):1083-1086.
162. Lane MD, Rominger, K. L., Young, D. L., and Lynen, F. (1964) The enzymatic synthesis of holotranscarboxylase from apotranscarboxylase and (+)-biotin. *J. Biol. Chem* 239:2865-2871.
163. Glasoe PK & Long FA (1960) Use of Glass Electrodes to Measure Acidities in Deuterium Oxide. *J Phys Chem-Us* 64(1):188-190.
164. Martin WG, Winkler, C.A. and Cook, W.H. (1959) Partial Specific Volume Measurements by Differential Sedimentation. *Can. J. Chem.* 37:1662-1670.
165. Edelstein SJ & Schachman HK (1967) The simultaneous determination of partial specific volumes and molecular weights with microgram quantities. *J Biol Chem* 242(2):306-311.
166. Iacob RE & Engen JR (2012) Hydrogen exchange mass spectrometry: are we out of the quicksand? *J Am Soc Mass Spectrom* 23(6):1003-1010.
167. Kleckner IR & Foster MP (2011) An introduction to NMR-based approaches for measuring protein dynamics. *Biochim Biophys Acta* 1814(8):942-968.
168. Eginton C & Beckett D (2013) A large solvent isotope effect on protein association thermodynamics. *Biochemistry* 52(38):6595-6600.
169. Hu W, *et al.* (2013) Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. *Proc Natl Acad Sci U S A* 110(19):7684-7689.
170. Tzeng SR & Kalodimos CG (2011) Protein dynamics and allostery: an NMR view. *Curr Opin Struct Biol* 21(1):62-67.
171. Velyvis A, Yang YR, Schachman HK, & Kay LE (2007) A solution NMR study showing that active site ligands and nucleotides directly perturb the allosteric equilibrium in aspartate transcarbamoylase. *Proc Natl Acad Sci U S A* 104(21):8815-8820.
172. Fenwick RB, Esteban-Martín S, & Salvatella X (2011) Understanding biomolecular motion, recognition, and allostery by use of conformational ensembles. *Eur Biophys J* 40(12):1339-1355.
173. Feher VA, Durrant JD, Van Wart AT, & Amaro RE (2014) Computational approaches to mapping allosteric pathways. *Curr Opin Struct Biol* 25:98-103.

174. Danielson ML & Lill MA (2012) Predicting flexible loop regions that interact with ligands: the challenge of accurate scoring. *Proteins* 80(1):246-260.
175. Subramani A & Floudas CA (2012) Structure prediction of loops with fixed and flexible stems. *J Phys Chem B* 116(23):6670-6682.