ABSTRACT

Title of Dissertation: AN INVESTIGATION OF ALLOSTERIC MECHANISMS IN BIOTIN PROTEIN LIGASES USING INTEGRATED BIOPHYSICAL APPROACHES

Jingheng Wang, Doctor of Philosophy, 2019

Dissertation directed by: Professor Dorothy Beckett, Department of Chemistry and Biochemistry

Allostery is a biological process in which action, often ligand binding, at one site of the protein alters the function at another site. It provides a mechanism for modulating protein functions in a variety of cellular events ranging from signaling, metabolism, to transcription regulation. Despite the critical role of allostery in biology and intense research during the past few decades, the mechanism of long-range communication through the protein is still elusive. The *Escherichia coli* biotin protein ligase (BirA) is a bifunctional protein that catalyzes post-translational biotinylation and represses transcription initiation. It serves as a model system to investigate long-range allosteric communication, as binding of the effector molecule, bio-5'-AMP, promotes the repressor complex assembly by enhancing BirA homodimerization occurring at a surface 30Å away. Previous studies have established that disorder-to-order transitions

of several loop segments on the ligand binding and dimerization surfaces contribute to BirA allostery. In this dissertation, integrated structural, functional, and computational approaches were used to investigate the molecular mechanisms of allosteric communication between these transitions. Double-mutant cycle analysis demonstrated reciprocal coupling between residues on two distant surfaces, and results of molecular dynamics simulations indicated that functional coupling occurs via modulation of structure and dynamics of surface loops undergo disorder-to-order transitions. Further structural and simulation-based network analyses revealed that these transitions are linked to formation of a residue network, and alanine substitutions of residues at network positions perturb both input (effector binding) and output (dimerization) of allostery. In addition, Force Distribution Analysis showed that perturbed loop folding is associated with redistribution of mechanical stress experienced by network residues. The combined results indicated a mechanism for BirA allosteric regulation in which disorder-to-order transitions and joint network formation enables long-range communication through the protein. Finally, results of functional measurements indicated a conserved allosteric regulation mechanism among *Escherichia coli (Ec)*, *Staphylococcus aureus (Sa)*, and *Bacillus subtilis (Bs)*, as bio-5'-AMP binding to *Sa* and *Bs*BirA induces homodimerization similar to that observed for *Ec*BirA.

AN INVESTIGATION OF ALLOSTERIC MECHANISMS IN BIOTIN PROTEIN LIGASES USING INTEGRATED BIOPHYSICAL APPROACHES

by

Jingheng Wang

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
2019

Advisory Committee:
Professor Dorothy Beckett, Chair
Associate Professor Douglas Julin
Associate Professor Nicole LaRonde
Associate Professor Paul Paukstelis
Associate Professor Silvina Matysiak

Acknowledgments

I would like to express my gratitude to everyone who provided help during my time at graduate school. Earning a Ph.D. degree is far more challenging than I expected, and I would not have done this without help and support from many people.

I would first like to thank my advisor, Dr. Beckett, for her continuous guidance and advice for all these years. Thank you for providing the opportunities and resources for me to develop scientific skills. You taught me not only professional knowledge but also speaking and writing skills. I am also very grateful for your patience and persistence in teaching, as I would have given up on many occasions. I hope I have become better over these years.

I also thank my committee members, Dr. Julin, Dr. LaRonde, Dr. Paukstelis for their advice and criticism. Thank you for your great suggestions during the committee meetings.

I am grateful to Dr. Matysiak and her group members, Dr. Gregory Custer, Riya Samanta, and Christopher Look, for conducting all the computational analyses on BirA. My thesis would not be complete without results from their studies. Thank you for your hard working.

I would also like to thank my current and former lab members Chenlu He, Calvin Muth, William Cressman, and Dr. Christopher Eginton who helped my research and taught me lab skills. I thank former undergraduate students, especially Rohith Raj, as I really enjoyed working with them.

I thank all my friends, especially Xin Tong, for their support and encouragement which made me more optimistic.

Finally, I would like to thank my parents and girlfriend Yating Zhang for their understanding and support. They are always my backup at home and walked me through the darkest time. I love you.

Table of Contents

Acknowledgments ii
Table of Contents iv
List of Tables vii
List of Abbreviations xi
Chapter 1: Introduction 1
A. Introduction to allostery 1
1. Summary 1
2. A brief history of allostery 1
3. Biophysical methods to investigate allosteric mechanisms 6
4. Application of integrated biophysical methods to investigate allosteric mechanisms 23
B. Allosteric regulation in the Escherichia coli biotin protein ligase 42
C. Biotin regulatory systems in other biotin protein ligases 49
D. Organization of the dissertation 51
Chapter 2: Long distance modulation of disorder-to-order transitions in protein allostery 55
A. Contribution statement 55
B. Introduction 55
C. Materials and Methods 59
1. Chemicals and biochemicals 59
2. Mutagenesis, expression, and purification of BirA double variants 59
3. Isothermal titration calorimetry (ITC) 60
4. Sedimentation equilibrium 60
5. Data analysis 61
6. Molecular dynamics (MD) simulations and analysis 62
D. Results 66
1. Molecular design of BirA variants with double-alanine substitutions 66
2. Bio-5'-AMP binding measurements reveal coupling between residues on the two BirA surfaces 71

3. Homodimerization measurements reveal coupling between G281 and M211 72
4. Molecular origins of functional coupling in BirA 75
5. Simulated structures indicate diversity in holoBirA variant monomers 77
6. Interaction energies on the ligand binding and dimerization surfaces are altered in the holoBirA variants 80
7. Detailed structural and dynamic basis of allosteric modulation of BirA dimerization 82
E. Discussion 87
Chapter 3: Integration of distant disorder-to-order transitions with a residue network in allostery 91
A. Contribution statement 91
B. Introduction 91
C. Materials and Methods 95
3.1 Chemicals and biochemicals 95
3.2 Mutagenesis, expression, and purification of BirA variants 95
3.3 Circular dichroism (CD) spectroscopy 96
3.4 Isothermal titration calorimetry (ITC) 96
3.5 Sedimentation equilibrium measurements 98
3.6 Data analysis 98
3.7 Molecular dynamics (MD) simulations and analysis 99
D. Results 102
1. Computational analysis reveals an extensive residue network in holoBirA 102
2. The network forms in solution 104
3. The network residues function in effector binding 105
4. Function of the network in BirA allosteric output, holoBirA dimerization 111
5. Force Distribution Analysis indicates coupling between disorder-to-order transitions and the network in BirA allostery 114
E. Discussion 118

Chapter 4: A conserved regulatory mechanism in bifunctional biotin protein ligases 123
A. Introduction 123
B. Materials and Methods 127
1. Chemicals and biochemicals 127
2. Expression and purification of *Bs* and *Sa*BirA 127
3. Biotinylation assay 129
4. Sedimentation equilibrium 130
5. Data analysis 131
6. Sequence alignment 132
C. Results 133
1. *Sa* and *Bs* BirA preparations are free of biotin or bio-5'-AMP contamination 133
2. Dimerization measurements on SaBirA indicate coupling between small ligand binding and homodimerization 135
3. Dimerization properties of *Sa*BirA are independent of buffer composition... 140
4. Small ligand effects on *Bs*BirA dimerization mirror those observed for *Ec*BirA 140
5. Coupling between *Sa* and *Bs* BirA dimerization and ligand binding 141
D. Discussion 143
Chapter 5: Summary and future work 151
Bibliography 161

List of Tables

Table 1. Common NMR techniques to study protein dynamics 13
Table 2. Rate constants of THF dissociation from different DHFR complexes 15
Table 3. Bio-5'-AMP binding parameters of BirA variants 68
Table 4. Dimerization properties of wt and variant holoBirA proteins 69
Table 5. Biotin binding parameters of BirA double variants are similar to those of wt 70
Table 6. Thermodynamics of coupling between dimerization and ligand binding surface residues 73
Table 7. Proton release linked to effector binding to BirA 105
Table 8. Bio-5'-AMP binding thermodynamics of BirA variants 110
Table 9. Dimerization energetics of BirA variants 113
Table 10. Maldi-ToF analysis of BCCP samples from biotinylation reactions 134
Table 11. Testing sedimentation equilibrium data for *Sa* and *Bs*BirA against single species and monomer-dimer association models 137
Table 12. Thermodynamic parameters governing *Sa* and *Bs*BirA dimerization 138

List of Figures

Figure 1. Schematic diagrams of the MWC and KNF models 4
Figure 2. Structures of deoxy- and oxy-hemoglobin 4
Figure 3. Structure of TetR dimer bound to its effector molecule, tetracycline 8
Figure 4. Thermodynamic cycle illustrating the four conformational substates of an allosteric protein with two binding sites. 9
Figure 5. Results of joint NMR and ITC analysis on AAC(6')-Ii allostery 12
Figure 6. Schematic representation of the *E. coli* DHFR catalytic cycle. 14
Figure 7. Structures of apo- and holo-glucokinase 19
Figure 8. Cleft-angle profiles of apo- and holo-glocokinase 19
Figure 9. Structure of IGPS heterodimer showing residues identified in the network analysis 22
Figure 10. Dependence of glucokinase activity on glucose concentration 25
Figure 11. Schematic diagrams showing the model for glucose-dependent kinetic cooperativity in glucokinase 28
Figure 12. Holo-glucokinase structure demonstrating locations of amino acid substitutions that lead to α-type activation and β-type activation. 30
Figure 13. Thermodynamic cycle illustrating the negative coupling between Zn and DNA binding to CzrA 33
Figure 14. Structures of inactive and active c-Src kinases 36
Figure 15. Residue networks identified in the simulated kinase domain in its inactive and active conformations 38
Figure 16. Illustration of high betweenness residues identified by network analysis in c-Src kinase 41
Figure 17. Schematic diagram showing the bifunctionality of *E. coli* BirA 43
Figure 18. Thermodynamic cycle showing that the repressor-complex assembly is solely dependent on homodimerization. 45
Figure 19. Bio-5'-AMP binding induces disorder-to-order transitions on both functional surfaces 46
Figure 20. Alanine substitutions at positions on two functional surfaces affect input and output of BirA allostery 48

Figure 21. The dependence of biotin operator occupancy by BirA dimer on biotin concentration 50
Figure 22. Bio-5'-AMP binding to BirA induces disorder-to-order transitions 58
Figure 23. Global root-mean-square deviation (RMSD) time series for backbone positions of the central domain in the apoBirA^{wt} simulation..... 63
Figure 24. HoloBirA^{wt} model showing the positions of dimerization and ligand binding surface residues. 70
Figure 25. Coupling between distant residues in ligand binding and dimerization.... 74
Figure 26. Comparison of wt apo and holo simulations. 76
Figure 27. HoloBirA variant structures differ significantly from that of holoBirA^{wt}. 79
Figure 28. Difference average interaction energy maps for residue pairs in the ligand binding and dimerization surfaces. 81
Figure 29. Alanine substitutions perturb the structure and dynamics of the dimerization surface loops. 84
Figure 30. Root-mean-square deviation (RMSD) time series for C, CA, and N atoms of the central domain of BirA variants 85
Figure 31. Loop packing at the ligand binding surface is influenced by substitutions on the dimerization surface. 86
Figure 32. Structural analysis of disorder-to-order and network formation upon bio-5'-AMP binding to BirA..... 93
Figure 33. Computationally determined residue network in holoBirA 103
Figure 34. CD Spectra for all variants are similar to that of BirA^{wt} 107
Figure 35. Network residues function in bio-5'-AMP binding and holoBirA dimerization 108
Figure 36. ITC measurements of bio-5'-AMP binding to BirA variants…….109
Figure 37. Absorbance versus radius profiles for holoBirA variants……..112
Figure 38. Force Distribution Analysis reveals coupling between disorder-to-order transitions and the network 116
Figure 39. Punctual stress distribution for BirA proteins with substitutions at R213 and G154..... 117
Figure 40. Amino acid substitutions tune the BirA allosteric response..... 122

Figure 41. The *Escherichia coli* Biotin Regulatory System 124
Figure 42. Maldi-ToF spectra of biotinylation products..... 134
Figure 43. Absorbance versus radius profiles from sedimentation equilibrium measurements carried out on holo*Sa*BirA and holo*Bs*BirA 135
Figure 44. Sedimentation equilibrium absorbance versus radius profiles for *Sa* and *Bs*BirA..... 139
Figure 45. Coupling between small ligand binding and dimerization..... 142
Figure 46. Multiple sequence alignment of *Ec*, *Bs*, and *Sa*BirA 146
Figure 47. Sequence and structural comparison of *Ec*, *Bs*, and *Sa*BirA..... 149

List of Abbreviations

3D	Three-Dimensional
AAC(6')-Ii	Aminoglycoside N-(6')-Acetyltransferase-Ii
ABL	Adenylate Binding Loop
AcCoA	Acetyl Coenzyme A
BBL	Biotin Binding Loop
BCCP	Biotin Carboxyl Carrier Protein
Bio-5'-AMP	Biotinoyl-5'-adenylate
BioO	Biotin Operator
BirA	Biotin Protein Ligase
Bs	*Bacillus subtilis*
BtnOH-AMP	Biotinol-5'-adenylate
cAMP	Cyclic Adenosine Monophosphate
CAP	Catabolite Activator Protein
CD	Circular Dichroism
CPMG	Carr–Purcell–Meiboom–Gill
Cyro-EM	Cryogenic Electron Microscopy
CzrA	Chromosomal Zn-regulated Repressor
DBD	DNA Binding Domain
DHF	Dihydrofolate
DHFR	Dihydrofolate Reductase
Ec	*Escherichia coli*
FDA	Force Distribution Analysis

GCK	Glucokinase
ITC	Isothermal Titration Calorimetry
KNF	Koshland-Némethy-Filmer
LIST	ligand-induced slow transition
MD	Molecular Dynamics
MWC	Monod-Wyman-Changeux
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
NMR	Nuclear Magnetic Resonance
PDB	Protein Data Bank
RMSD	Root-Mean-Square Deviation
RMSF	Root-Mean-Square Fluctuation
Sa	*Staphylococcus aureus*
TetR	Tetracycline Repressor
THF	Tetrahydrofolate

This page is intentionally left blank

Chapter 1: Introduction

A. Introduction to allostery

1. Summary

Allostery is the regulation of protein function at one site by action, usually effector binding, at another site. This type of regulation is found in a variety of cellular processes ranging from signaling[1], transcription regulation,[2] to metabolism[3]. Due to its importance in biology, allostery has been the center of intense research since the 1960s. However, how distant functional sites communicate across a protein remains to be elucidated. In this dissertation, combined functional, structural, and computational approaches were employed to investigate the molecular mechanisms of long-range communication in a model system, the *Escherichia coli* biotin protein ligase (BirA). We hope that results of our study can provide insights into understanding general mechanisms of allosteric regulation, which can be utilized to design new drugs that modulate enzyme function through allostery.

2. A brief history of allostery

The first observation of allostery was reported over a century ago by Christian Bohr and his colleagues, who discovered that oxygen binding to hemoglobin yielded a sigmoidal curve instead of a hyperbolic one observed for independent binding events.[4] They also reported that the presence of carbon dioxide weakened oxygen binding to hemoglobin.[4] Unknown to Bohr at that time, both observations are associated with

hemoglobin allostery. The sigmoidal binding curve reflects positive cooperativity among the four hemoglobin subunits, as oxygen binding to one subunit enhances subsequent binding events through conformational changes.[5, 6] The effect of carbon dioxide, also known as the "Bohr effect", is primarily caused by the release of protons as a result of bicarbonate formation in the blood.[7] These protons serve as heterotropic negative effectors of hemoglobin and alter the protein structure to a conformation unfavorable for oxygen binding.[5, 6]

Following the observation by Bohr, a number of phenomenological models were proposed to explain the cooperativity in hemoglobin and other allosteric proteins. These include the Hill[8], Adair[9], Klotz[10] and Pauling[11] equations which provide pure mathematical descriptions of cooperative binding. In the 1960s, two important models, the concerted Monod-Wyman-Changeux (MWC) model[12] and the sequential Koshland-Némethy-Filmer (KNF) model[13], were proposed to explain the mechanism of allosteric regulation in hemoglobin (Figure 1). Both models describe allostery as a conformational exchange between the low-affinity tense (T) and the high-affinity relaxed (R) states. Binding of oxygen shifts the conformation to the R state, while negative allosteric effectors, such as protons, are predicted to stabilize the T state. The difference between the two models is whether the four subunits undergo structural changes simultaneously (MWC) or independently (KNF). In the KNF model, oxygen binding to one subunit induces the T-to-R transition of neighboring subunits, thereby increasing the likelihood of subsequent oxygen binding events.[13] By contrast, the MWC model postulates that all 4 subunits undergo conformational changes

simultaneously. As a result, the protein exists in an equilibrium between the T and R states, and positive cooperativity is generated by shifting the equilibrium towards the R state.[12] Both models are consistent with experimental observations and explain the sigmoidal oxygen binding curve. However, neither MWC nor KNF models address the molecular mechanism of transition between the two states, and how oxygen binding promotes these hypothetical structural changes was unknown.

The development in X-ray crystallography in the 1960s allowed for the elucidation of detailed structural features of hemoglobin allostery for the first time. Although earlier studies had acquired structures of hemoglobin, their resolutions were too low to reveal the details of the T-R state transition.[14, 15] In 1968 and 1970, Max Perutz and his colleagues finally obtained high-resolution (2.8 Å) crystal structures of horse oxy- and deoxy-hemoglobin.[16, 17] Based on these structures, Perutz proposed that the stereochemical change at the heme porphyrin ring is critical for hemoglobin allostery.[5, 6] In this model, oxygen binding to the heme group "pulls" the iron atom into the plane of the porphyrin ring and repositions the histidine attached to the iron (Figure 2). This results in a chain of events including the breakage of inter-subunit salt bridges that are responsible for the stabilization of the deoxy-conformation (T state). The two-state structural transition model proposed by Perutz has now been challenged by the discovery of additional ligand-bound hemoglobin structures,[18, 19] and NMR measurements showed that hemoglobin, like many other proteins, is dynamic in solution.[20, 21] Nonetheless, Perutz's data yielded the first molecular model to explain the allosteric mechanism in hemoglobin.

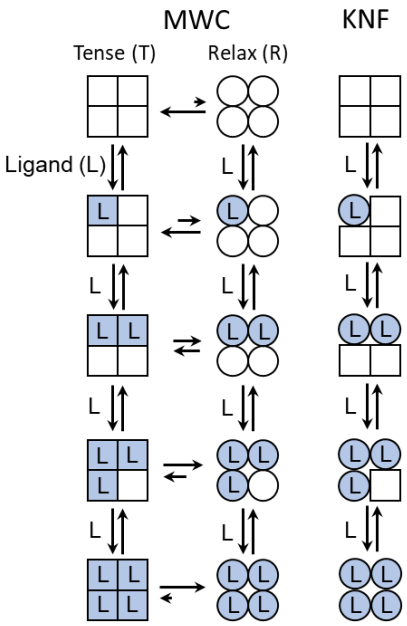


Figure 1. Schematic diagrams of the MWC and KNF models

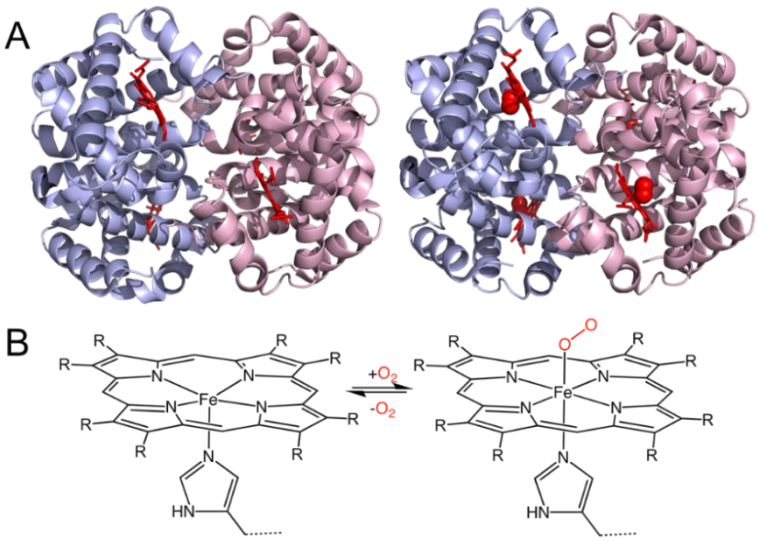


Figure 2. (A) Structure models of deoxy- (left) and oxy- (right) hemoglobin generated in Pymol[22] using PDB files 2DN1 and 2DN2.[23] Heme groups and oxygen molecules are highlighted in red. (B) Structures of the heme group showing that oxygen binding shifts the iron atom into the plane of the porphyrin ring. The figure was taken from the Wikipedia page "Heme".[24]

The idea that allostery occurs via rigid body movements, such as conformational changes in hemoglobin, dominated the field throughout the 1960s and 1970s. However, this view was challenged by Alan Cooper and David Dryden who proposed that allostery can occur solely by altering protein dynamics.[25] The authors demonstrated, using statistical thermodynamics, that changes in frequency and amplitude of protein fluctuations upon effector binding can yield several kJ/mol of cooperative free energy purely through the entropic term. In other words, dynamic changes alone are capable of generating allosteric responses in the absence of structural changes. Following the proposal of this hypothesis, a number of NMR studies reported significant contributions of dynamic changes to allostery,[26-28] and dynamics have now been accepted as an indispensable part of allosteric regulation.

Inspired by results of molecular dynamics simulations,[29, 30] Cooper and Dryden further proposed that allosteric response is the redistribution of pre-existing conformational states.[25, 31] According to their theory, the protein exists as an ensemble of conformations which, based on their probabilities of appearance, form a population distribution. Effector binding is predicted to shift the population towards a specific state, while changes in protein dynamics (flexibility) alter the width of the distribution. The work by Cooper and Dryden was prescient, and the ensemble theory based on their idea has now been widely accepted as a general model for allostery.[32-34] Nevertheless, as a thermodynamic model, the ensemble theory does not reveal the chemical basis of allostery, and the mechanisms underlying the inter-site communication have not been fully resolved.

3. Biophysical methods to investigate allosteric mechanisms

A variety of structural, solution, and computational approaches have been employed to investigate the mechanisms of long-range communication in allosteric regulation. The following section introduces several commonly used biophysical methods and highlights their strengths and weaknesses in elucidating allostery. Due to the limitations of each biophysical technique, combined approaches are usually required to obtain a complete picture of allosteric regulation.

(1) Structure-based approaches

Structural analysis is by far the most frequently used method to study allostery, and the majority of our understanding of allosteric mechanisms comes from analyses of protein structures. X-ray crystallography, cryogenic electron microscopy (cryo-EM), and NMR spectroscopy are the top three techniques to acquire protein structures, each with its unique advantages and drawbacks. As the most popular approach, X-ray crystallography applies to all allosteric proteins provided that they are able to form diffractable protein crystals. By comparison, cryo-EM is a crystal-free technique that has gained much more attention in the past few years due to advancement in data collection and analysis.[35-37] Recent studies have reported several cryo-EM structures with resolutions comparable (≈2Å) to that of crystal structures.[38, 39] However, structure determination of small proteins by cryo-EM remains challenging as the signal-to-noise ratio of data drops with particle size. Finally, NMR spectroscopy is least popular among the three methods because it requires multiple measurements and a substantial amount of effort to solve the structure *de novo*.[40] Moreover, the technique is only applicable to

small proteins, as larger ones experience spectra overlapping and line broadening that complicate peak assignment. Nevertheless, NMR structures better representation of solution conformations, as they are shown as an ensemble of most likely conformations.

It should be noted that structures alone do not tell the full story of allostery, because structural models acquired from X-ray crystallography or cryo-EM are usually static images that represent the end states of allosteric regulation. Therefore, they do not explain how the allosteric signal is transmitted across the protein. In addition, as predicted by Cooper and Dryden, dynamic changes are also integral to allosteric regulation.[25] For instance, structural analysis failed to reveal the dynamic basis of allosteric regulation in tetracycline repressor (TetR). TetR is a transcription repressor in which binding of the effector molecule, tetracycline, impairs DNA binding at a distant functional site (Figure 3).[41, 42] Comparison of TetR-DNA complex crystal structures with and without tetracycline bound suggested that structural transitions in the DNA binding domain (DBD) are responsible for effector-induced de-repression.[43] However, this model fails to explain why free TetR, which binds to DNA tightly, adopts a conformation incompatible with DNA binding.[42] A later study using equilibrium unfolding experiments suggested that tetracycline binding prevents repressor complex assembly by inducing the ordering of the DBD, thereby making it less flexible to sample conformations favorable for DNA binding.[44] These studies indicate that attempts to elucidate allosteric mechanisms using only structure-based approaches are

insufficient and can be misleading, and structural analysis needs to be combined with dynamic measurements to yield a more complete understanding of allostery.

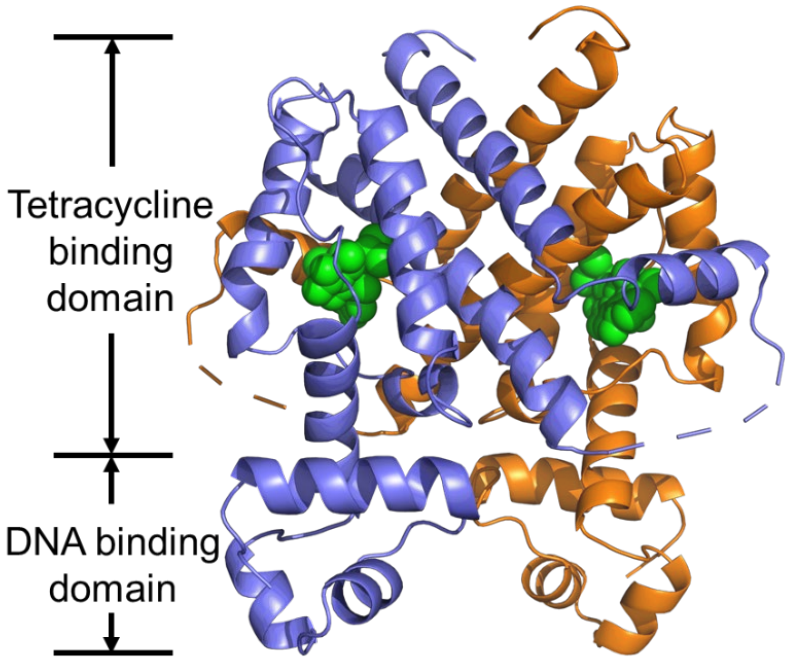


Figure 3. Structure of TetR dimer bound to its effector molecule, tetracycline (green spheres). The figure was generated in Pymol[22] using PDB entry 2TCT.[45]

(2) Nuclear magnetic resonance (NMR) spectroscopy

NMR spectroscopy is another powerful tool to investigate the mechanisms of long-range allosteric communication. In addition to direct structure determination, chemical shift changes provide valuable information regarding structural and/or dynamic transitions associated with external stimuli, such as effector binding or amino acid substitutions. For example, in a recent study by Yunyao *et al.*,[46] the authors proposed a method to predict allosteric participants by determining the number of possible chemical shifts in various states of the protein. The authors reasoned that for

an allosteric protein with two coupled binding sites, there should be four conformational substates, namely the apo, doubly-bound, and two singly-bound states (Figure 4). Non-allosteric residues that respond to binding of one ligand, such as ligand A, are expected to have only two chemical shifts representing either the ligand-free (1 and 3) or bound (2 and 4) state. By comparison, residues involved in allosteric coupling are predicted to have four distinct chemical shifts. Using the potassium ion channel KcsA as the model system, the authors demonstrated that the new method was able to identify potential allosteric residues that were not detected by conventional structural analysis. Their predictions were validated by mutagenesis studies, which showed that that amino acid substitutions of identified residues affect KcsA allostery.[46, 47]

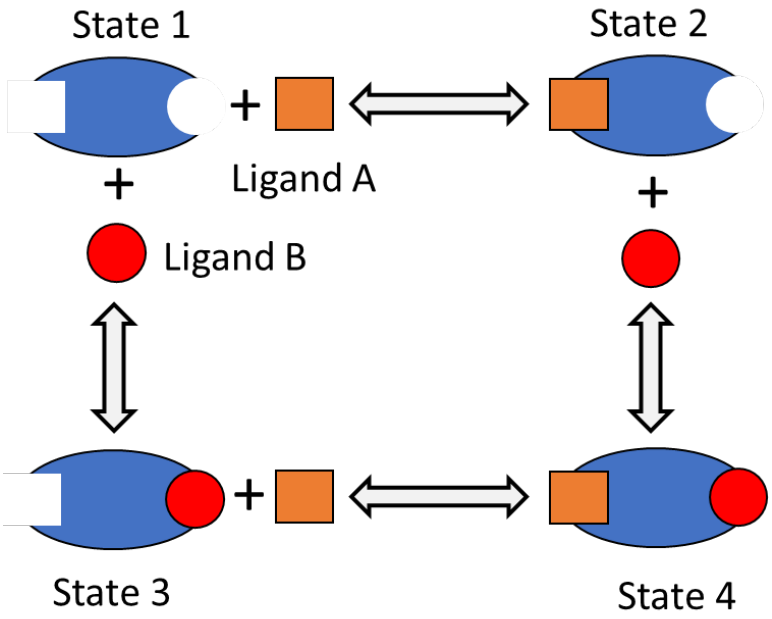


Figure 4. Thermodynamic cycle illustrating the four conformational substates of an allosteric protein with two binding sites.

A more detailed chemical shift analysis carried out by Freiburger *et al.* facilitated the elucidation of allosteric mechanism in aminoglycoside N-(6')-acetyltransferase-Ii (AAC(6')-Ii).[48] AAC(6')-Ii is an enzyme that transfers the acetyl group from acetyl coenzyme A (AcCoA) to the 6' position of aminoglycoside antibiotics to render drugs ineffective against bacteria.[49] Combined structural and functional studies indicated that binding of the first AcCoA molecule to the homodimeric enzyme is positively coupled to the second AcCoA binding.[50, 51] Results of NMR measurements showed that the enzyme undergoes substantial structural and dynamic changes upon AcCoA binding, and allosteric regulation in AAC(6')-Ii is well-characterized by a two-state model.[48] However, the mechanism of this cooperative binding was poorly understood, and it is unclear if AAC(6')-Ii allostery can be described by the MWC or KNF model.[48]

The key to distinguishing between the two models is to identify the conformation of singly-bound AAC(6')-Ii. In the MWC model, the singly-bound enzyme exists in an equilibrium between the free (apo) and fully-bound (holo) states, and AcCoA binding shifts this equilibrium towards the holo state. By contrast, the KNF model predicts that the singly-bound AAC(6')-Ii is a heterodimer, with one monomer resembling the apo conformation and the other mimicking the holo form. Not surprisingly, NMR spectroscopy did not detect any distinct signal for the intermediate state, as it probably overlapped with apo and holo peaks on the spectra. To dissect the signal of the singly-bound AAC(6')-Ii, Freiburger *et al.* developed a joint analysis of NMR and isothermal titration calorimetry (ITC) data.[48] They first determined the

species distribution at different substrate concentrations using ITC, and this information was utilized to calculate the contribution of singly-bound proteins to the signal intensity of each NMR peak. (Figure 5A, B) The KNF model predicts that the NMR signal of the singly-bound protein will be split equally into apo and holo peaks at a 1:1 ratio. On the other hand, if AAC(6')-Ii allostery is consistent with the MWC model, the contributions of singly-bound species to apo and holo peak intensities should be either 0 or 100%, as the species distribution has been taken into consideration in the calculation. Results of the analysis showed that for the majority of the residues, the NMR signal of singly-bound state contributes equally to corresponding apo or holo peaks (Figure 5C, 100% contribution is normalized to 2), thereby supporting the KNF model for AAC(6')-Ii allostery.

The major drawback of NMR chemical shift analysis is that it is difficult to distinguish effects caused by external stimuli, such as effector binding, from those generated by allostery. Therefore, this method usually serves as a tool to generate hypothesis for allosteric mechanisms, and results of chemical shift analysis require validation by other biophysical methods.

Another application of NMR spectroscopy is to characterize the dynamic basis of allosteric regulation.[26, 52, 53] Advances in NMR methods have enabled observation of dynamic events over a range of timescales from picosecond-nanosecond side chain fluctuations to millisecond or longer protein folding and domain movements (Table 1).[54, 55] The most frequently used technique is the Carr–Purcell–Meiboom–Gill (CPMG) relaxation dispersion, which monitors microsecond to millisecond timescale

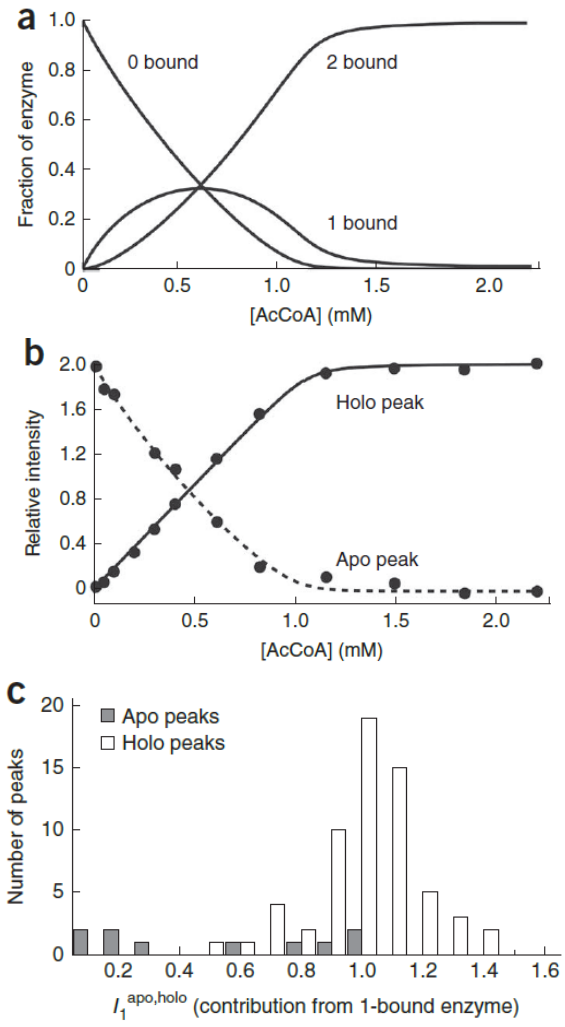


Figure 5. Results of joint NMR and ITC analysis on AAC(6')-Ii allostery. (A) Species distribution plot showing fractions of apo (0-bound), singly-bound, and holo (2-bound) enzyme as a function of AcCoA concentration. (B) Intensity profiles of the apo (dashed line) and holo (solid line) peaks for residue Leu56 as a function of AcCoA concentration. (C) Histograms of the relative contribution of the singly-bound signal to all distinguishable apo and holo peaks. 100% contribution is normalized to 2. Figures were taken from the paper by Freiburger *et al.*[48]

motions that are most relevant to protein conformational exchanges.[56, 57] Analysis of relaxation data yields the exchange rate, the population of each state, and the chemical shift difference between the two states, which enable the characterization of both kinetics and thermodynamic properties of the conformational exchange.

Table 1. Common NMR techniques to study protein dynamics

Technique	Timescale	Reference
EXSY (magnetization exchange spectroscopy)	>100 ms	[58]
CPMG (Carr–Purcell–Meiboom–Gill)	100 ms–100 μs	[56, 57]
RDC (residual dipolar couplings)	<10 ms	[59]
Spin relaxation	ns–ps	[60, 61]

CPMG relaxation dispersion has been employed to investigate the dynamic basis of allosteric regulation in many proteins including dihydrofolate reductase (DHFR). DHFR is a critical metabolic enzyme that catalyzes the nicotinamide adenine dinucleotide phosphate (NADPH)-dependent reduction of dihydrofolate (DHF) to tetrahydrofolate (THF).[62] Under saturating ligand concentrations, the enzyme cycles through five major intermediates in which the product (THF) release serves as the rate-limiting step (Figure 6).[63] Kinetic measurements demonstrated that NADPH binding to the enzyme:THF (E:THF) complex increases the rate of THF dissociation by approximately 8-fold (Table 2).[63] As the two substrates occupy distinct binding sites,[64] NADPH binding probably induces the product release via allostery. Structural analysis suggested that the domain movement upon NADPH binding may contribute to product release,[65] but it does not explain why binding of the almost identical $NADP^+$ fails to

induce a similar effect. Dynamic measurements using CPMG relaxation dispersion demonstrated that NADPH-induced population shift is responsible for elevated THF dissociation.[66, 67] According to the results of NMR measurements, NADPH binding to E:THF populates a higher energy state that resembles the conformation of the E:NADPH complex, the following intermediate in the catalytic cycle (Figure 6).[66] Further investigation of this minor conformational state showed that the two substrates generate a steric clash in the active site that favors the rejection of THF.[67] By contrast, this steric hindrance is not observed in the ground state, and neither E:THF nor E:THF:NADP$^+$ complexes sample the same excited state observed for the E:THF:NADPH complex.[63, 66] In summary, NADPH binding induced-product release occurs by populating a conformation that is unfavorable for simultaneous binding of both THF and NADPH.

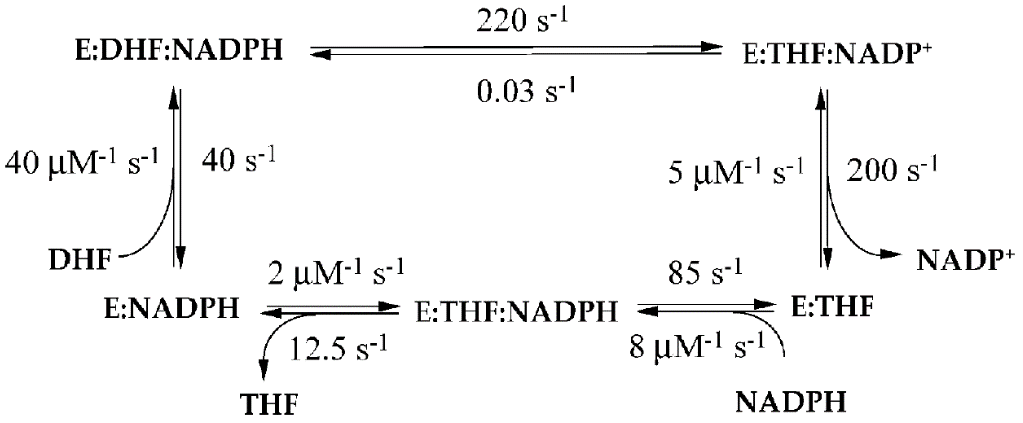


Figure 6. Schematic representation of the *E. coli* DHFR catalytic cycle showing its five primary intermediates and rate constants for each step. The figure was taken from the paper by Schnell *et al.*, 2004.[68]

Table 2. Rate constants of THF dissociation from different DHFR complexes. Values were taken from the paper published by Fierke *et al.* in 1987.[63]

Enzyme species	k_{off} (s^{-1})
E:THF	1.4±0.2
E:$NADP^+$:THF	2.4±0.2
E:NADPH:THF	12±2

Although NMR spectroscopy is powerful in elucidating allosteric mechanisms, it has its own limitations. First, chemical shift changes or dynamic information alone do not reveal the mechanisms of allosteric communication, and interpretation of NMR data usually depends on prior structural and functional studies. Second, spectral analysis relies on the correct assignment of resonance peaks to corresponding nuclei, but spectra of large proteins (>100 kDa) usually contain many overlapping peaks that are difficult to identify with certainty. The slow tumbling of these proteins also results in line broadening and further increases the difficulty of peak assignment. Although isotopic labeling techniques, such as ^{19}F labeling[69] or methyl-specific ^{13}C labeling[70], effectively reduce the number of signals, some proteins are still not suitable for analysis by NMR. Moreover, selective labeling limits the scope of analysis to those labeled residues.

(3) Functional measurements

One critical step to understanding allostery is to link structure and dynamic information to protein function. Studies using structure-based approaches and NMR

measurements can identify structural transitions or dynamic changes that may be important for allostery, but these hypotheses require validation by functional measurements. The classic approach is to use site-direct mutagenesis, as perturbations that disrupt allosterically important transitions are predicted to alter the energetic coupling between effector binding and the functional response. The magnitude of this coupling, also known as the coupling free energy, can be quantitatively determined by thermodynamic or kinetic measurements.

In some cases, functional measurements can provide insights into allosteric mechanisms. For example, ITC measurements on *Staphylococcus aureus* chromosomal Zn-regulated repressor (CzrA) showed that the negative allosteric coupling between Zn and DNA binding is entropic in origin, suggesting the critical role of dynamics in CzrA allosteric regulation.[71] Results of cyclic adenosine monophosphate (cAMP) binding measurements supported dynamically-driven allostery in catabolite activator protein (CAP), as the second cAMP binds CAP with large unfavorable entropy but slightly favorable enthalpy.[26] In both cases, predictions generated by thermodynamic measurements were consistent with results of NMR measurements, which showed that dynamic changes are dominant in allosteric regulation of these proteins. Nevertheless, interpretation of results obtained from functional measurements usually requires additional structural or dynamic information, as thermodynamics alone do not reveal the mechanisms of allosteric regulation.

(4) Molecular dynamics (MD) simulation

MD simulation is a valuable complement to "wet-lab" experiments in studying allostery as it provides an all-atom description of protein motions in both space and time. A typical MD simulation starts with a high-resolution structural model obtained from X-ray crystallography, cryo-EM, or NMR spectroscopy. The structural information is then converted into potential energies using a force field, which is a collection of terms that describe interaction energies, in order to process atomic interactions quantitatively *in silico*. Finally, all atoms and molecules are allowed to interact using Newton's laws of motion, and the position of every atom is recorded at a fixed time interval. In other words, results of MD simulations, also known as trajectories, are a series of snapshots of the protein at different time points.

Compared to "wet-lab" experiments, MD simulation is not restricted by technical limitations such as protein expression, purification or crystallization. This feature allows for easy implementation of external stimuli including ligand binding, amino acid substitutions, and even mechanical forces.[72, 73] These advantages have enabled MD simulation to study some challenging systems including G-protein coupled receptors[74] and ion channels[75].

Analysis of MD simulation data not only provides explanations of experimental observations but also generates testable hypotheses for mechanistic studies. The classic strategy is to compare a parameter of interest between trajectories of different protein states. For example, comparison of correlated motions[76] or root-mean-square deviations (RMSD)[76-78] highlights structural transitions upon effector binding, while

comparing root-mean-square fluctuations (RMSF)[78] may reveal the dynamic basis of allostery. One example is the investigation of the allosteric activation mechanism in human glucokinase, an important metabolic enzyme that regulates glucose hemostasis in the blood. [79-81] It consists of an N-terminal small domain and a C-terminal large domain, with the glucose binding site located in the interdomain cleft. Activation by natural substrate binding induces notable structural changes, including the transition from an open conformation to a more compact closed state. (Figure 7)[82] Previous studies have identified several potent allosteric activators which promote enzyme activity by binding at an allosteric site 20Å away from the active site,[83, 84] but their mechanism of action was unclear. To investigate the mechanism of allosteric activation in glucokinase, MD simulations were performed on the active conformation of glucokinase in the presence and absence of an allosteric effector.[79, 84] Similar to the substrate binding-induced activation, analysis of simulation data showed that activator binding decreases the cleft-angle between the two domains. This yields a more rigid glucokinase conformation and maintains the enzyme in the closed state (Figure 8, Green). By contrast, the free protein is more dynamic and less likely to sample the closed conformation (Figure 8, Red). Results of MD simulation suggest that in addition to structural transitions, dynamic changes also contribute to the activation of glucokinase. This hypothesis is supported by later NMR studies showing that order-disorder transitions of an active site loop and ordering of the small domain play a critical role in glucose binding-induced activation of glucokinase.[85, 86]

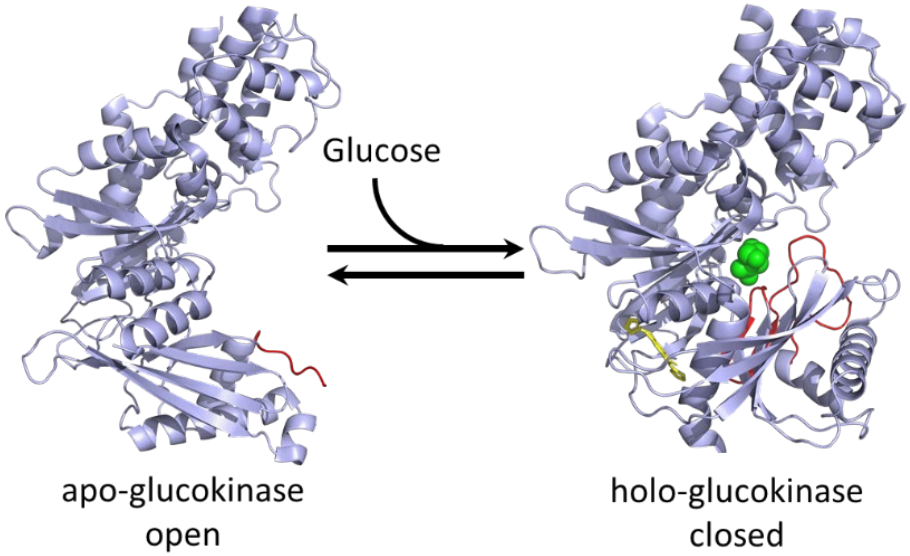


Figure 7. Structures of apo- and holo-glucokinase. Glucose binding induces significant structural changes from an open conformation to a closed form. Color code: Red, the 151-180 loop; Green, glucose; Yellow, an allosteric activator that maintains closed conformation. Models were made using Pymol[22] with PDB files 1V4S and 1V4T.[82]

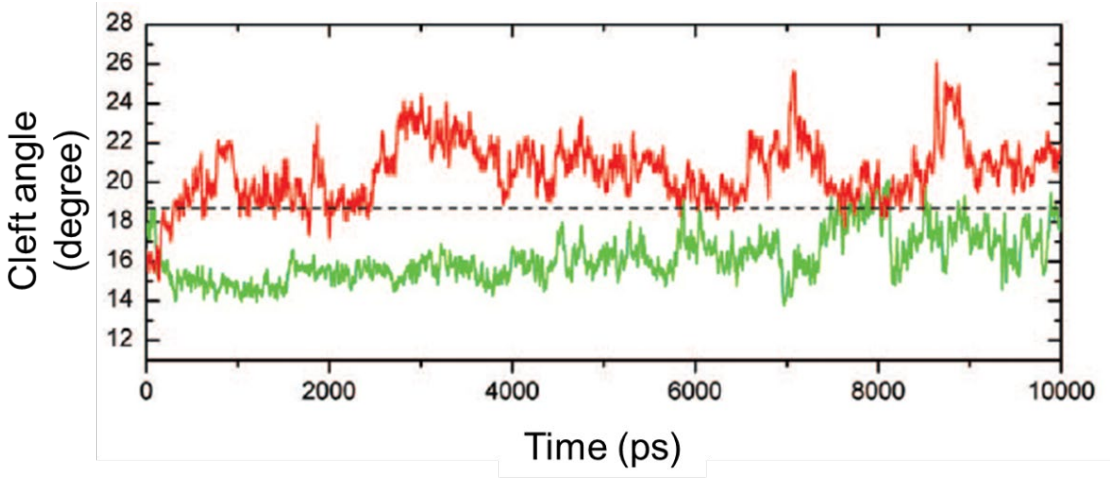


Figure 8. Cleft-angle profiles between two domains of glucokinase in the absence (red) and presence (green) of an allosteric activator. Larger angles indicate more open conformations.[79] The figure was adapted from the paper by Zhang *et al.*, 2006.[79]

MD simulation is also used to identify residue networks that are important for allosteric communication. An increasing number of studies have demonstrated that allosteric communication between distant functional sites occurs through networks of coupled amino acids, rather than a specific pathway that involves only a chain of connecting residues.[87-91] Naturally, identification of putative networks in the protein is important to elucidate the allosteric mechanism. Several structure-based algorithms have been developed to analyze amino acid networks in the protein,[92-94] but the drawback of these approaches is that they rely heavily on static structure models that do not necessarily represent conformations in solution. For instance, residue interactions that form or break transiently due to intrinsic dynamics will not be identified in crystal structures. This results in altered network topology and may affect subsequent identification of allosteric residues.

Multiple simulation-based network approaches have been applied to identify residues and networks that may be important for allostery.[95] One example is the energy-based network analysis proposed by Ribeiro and Ortiz.[96, 97] In this approach, the protein structure is transformed into a mathematical graph where residues are represented as nodes, and node interactions are drawn as edges. Compared to other network approaches, the major difference of this new analysis is that it determines the network topology based on interaction energies. This is based on the assumption that allosteric signal preferentially propagates through the pathway that cost the minimum amount of energy. Therefore, the new network analysis not only reveals what residues are in contact, but also provides information on the most likely (shortest) contacts between

any pair of residues.[96] This feature can be utilized to quantitatively evaluate the participation of a residue in these contacts. Residues that are critical for allostery are predicted to have high participation in inter-residue communication, and top candidates identified by this analysis would form a network through which allosteric signal is propagated.[96, 97]

To test the effectiveness of this new approach, the authors performed the network analysis on an allosteric enzyme, the imidazole glycerol phosphate synthase (IGPS), using different network algorithms. IGPS is a heterodimer that consists of HisH and HisF monomers (Figure 9). Binding of an effector molecule in the HisF monomer, which is one 25Å away from the active site, enhances the catalytic activity of IGPS by about 5000-fold.[98, 99] In the work by Ribeiro and Ortiz, the authors calculated shortest paths between residues in the effector binding and active sites using several different network algorithms. Results of the calculations showed that these approaches yielded different communication pathways, but the energy-based network analysis correctly identified residues important for allosteric communication.[99] These include residues D98f (blue) and K181h (red) which forms a crucial salt bridge connecting the two subunits. By comparison, analysis using well-established Pearson correlation coefficients failed to highlight this interaction.[96] Therefore, the network analysis developed by Ribeiro and Ortiz can effectively identify important network residues with high accuracy.

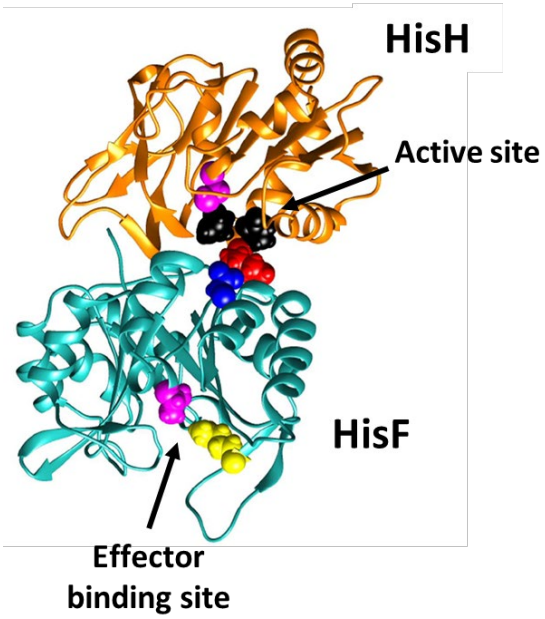


Figure 9. Structure of IGPS heterodimer showing residues identified in the network analysis. The figure was adapted from the paper by Ribeiro and Ortiz.[99]

Despite the power of MD simulation in studying allosteric mechanisms, it should be noted that the results of MD simulations still require validation by experiments. This is mostly because force fields that characterize atomic interactions are far from perfect, and Newton's laws of motion do not take quantum effects into consideration. Another restriction of MD simulation is the limitation in computing power. Common microsecond timescale simulations can take days to complete on supercomputers, while millisecond level simulations are only achieved on a few systems.[100] One alternative is to use faster but less accurate coarse-grained simulation, in which a small group of atoms, such as a residue, are treated as a single particle during

the calculation. This reduces the complexity of the system and lowers the requirement for computing power, albeit at the cost of accuracy.[101]

In summary, this section describes several commonly used structural, dynamic, functional, and computational approaches to study allostery. Each method has its expertise in revealing a particular aspect of allosteric mechanisms, but the complex nature of allostery, as demonstrated in some examples below, requires the use of multiple methods simultaneously to reveal the full story of allosteric regulation.

4. Application of integrated biophysical methods to investigate allosteric mechanisms

As described above, the investigation of allosteric mechanisms by a single approach cannot yield a complete picture of allostery. In this part of the chapter, I will introduce several examples in which allosteric mechanisms were studied using integrated biophysical techniques and discuss the significance of results generated by these methods. In the first example, studies on glucokinase revealed an allosteric mechanism in which substitutions at locations distant from the active site modulate enzyme function by affecting the disorder-to-order transition of an active site loop. The second example of CzrA highlights the importance of dynamics in allosteric regulation and demonstrates the predictive power of MD simulations. In the third system, combined biophysical approaches revealed a distributed allosteric network in Src kinase that links the kinase domain to regulatory regions on the other side of the protein. This example demonstrates the application of network theory in studying allostery.

(1) Human glucokinase

Disorder plays an important role in allosteric regulation, and its significance in the thermodynamics of allostery has been well-characterized.[102, 103] However, the molecular mechanisms of how disorder contributes to allosteric regulation have not been not fully elucidated. Human glucokinase provides a model system to investigate the role of disorder in allostery, as activation of the enzyme is associated with the disorder-to-order transition of an active site loop, and distant amino acid substitutions alter enzyme activity by modulating the dynamics of this loop.[82, 85, 104]

Human glucokinase is a special member of the hexokinase family that catalyzes the conversion of glucose and ATP to glucose-6'-phosphate, the first step of glucose metabolism.[105, 106] It is the predominant hexokinase in the liver and pancreatic β cells and is responsible for maintaining glucose homeostasis in the blood.[107, 108] Dysregulation of enzyme function by genetic mutations in the glucokinase gene is associated with severe metabolic diseases such as maturity-onset diabetes of the young (MODY) or permanent neonatal diabetes mellitus (PNMD).[109] Given its critical role in glucose metabolism, glucokinase has become a potential target to treat diabetes,[110, 111] and studies have identified several allosteric activators that enhance the enzyme activity.[83] Therefore, a better understanding of allosteric activation mechanisms in glucokinase can facilitate the design of novel drugs that treat non-insulin dependent diabetes and metabolic diseases caused by loss-of-function genetic mutations.

As the regulatory function of glucokinase is closely connected to its unique kinetic features, elucidation of allosteric regulation in glucokinase would be impossible

without a thorough understanding of its kinetic cooperativity. Kinetic measurements demonstrated that the activity of glucokinase exhibits a sigmoidal dependence on glucose concentration with a Hill coefficient of 1.7 (Figure 10).[3] As the midpoint of the curve is comparable to physiological blood glucose level (4 to 7 mM), the kinetic cooperativity allows the enzyme to respond sensitively to sugar uptake while maintaining glucose homeostasis in the blood.[112] Allosteric activators or substitutions that generate hyperactive variants also modulate enzyme function by removing the glucose-dependent cooperativity in glucokinase.[83, 104].

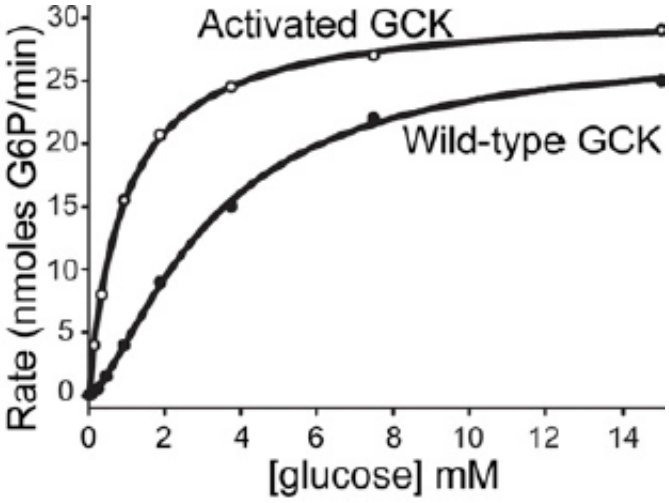


Figure 10. Dependence of glucokinase (GCK) activity on glucose concentration. The figure was adapted from the paper by Whittington *et al.*[104]

The positive kinetic cooperativity observed for glucokinase is characteristic of allosteric regulation, but unlike "classic" allosteric proteins, which have multiple subunits, glucokinase is monomeric and has only one glucose binding site.[82] Two theoretical models have been proposed to explain the kinetic cooperativity in

monomeric enzymes: the mnemonic model[113] and the ligand-induced slow transition (LIST) model[114]. Both models postulate that the enzyme undergoes a slow conformational exchange between the active and inactive conformations. The mnemonic model proposes that the enzyme "memorizes" the active conformation temporarily after turnover, thereby allowing it to skip the slow exchange process at high substrate concentrations. By contrast, the LIST model states that the enzyme in both conformations can catalyze the reaction, albeit at different rates. The positive cooperativity upon effector binding is generated by population shift towards the faster conformation. The major difference is that there is only one catalytically active complex in the mnemonic model, while the LIST model had two separate catalytic cycles. For glucokinase, both models were supported by multiple sources of evidence, and no consensus had been reached for decades.[115] Moreover, like other phenomenological models, neither mnemonic nor LIST models reveal the chemical basis of the cooperativity in glucokinase.

Structural analysis revealed that glucose binding-induced activation is accompanied by large-scale conformational changes. X-ray crystal structures showed that glucokinase adopts a palm-shaped conformation with two domains: the large and small domains (Figure 7).[82] The two domains are connected by a deep cleft, which forms the active site for glucose phosphorylation. Binding of the effector molecule, glucose, leads to the movement of two domains from an open conformation towards a more closed form. In addition, effector binding is also accompanied by the disorder-to-order transition of an active site loop comprised of residues 151-180 in the small

domain.[82] The release of phosphorylated glucose, on the other hand, triggers the order-to-disorder transition and returns the enzyme to the open conformation. However, structures alone do not provide insight into the mechanism of kinetic cooperativity because they only demonstrate equilibrium models of the protein.

To investigate the dynamic basis of glucose-induced allosteric activation of glucokinase, the Brüschweiler group at Ohio State University carried out a series of studies using NMR spectroscopy.[85, 86, 115, 116] Results of their NMR measurements showed that glucokinase samples multiple conformations in solution, and glucose binding alters the dynamics of the protein by stabilizing specific conformations.[116] A further investigation showed that these dynamic changes mostly originated from the ordering of the small domain, including the disorder-to-order transition of the 151-180 loop.[85] These results suggest that kinetic cooperativity may be linked to the dynamics of the small domain. Consistent with the hypothesis, amino acid substitutions that lead to hyperactivity decreased dynamics of the small domain in the unliganded state.[85] Overall, results of combined structural and dynamic measurements indicate that the ordering of the small domain and folding of the 151-180 loop contribute to the glucose-dependent allosteric activation of glucokinase.

To better understand the kinetic cooperativity in glucokinase, the Brüschweiler group characterized conformational exchanges in the protein using CPMG relaxation dispersion experiments. Results of the measurements revealed that the free enzyme exchanges between the active and inactive conformation at a rate of 510±50 s^{-1}, a value that is comparable to the rate of enzyme turnover (220 s^{-1}). By contrast, conformational

exchange on the millisecond timescale was not detected in the substrate-bound glucokinase, suggesting that there is only one active catalytic complex.[86] The slow kinetics of conformational exchange in combination with the absence of second catalytic cycle support the mnemonic model in which the glucokinase "memorizes" the active state shortly after the turnover: At low glucose concentrations, glucokinase undergoes the complete catalytic cycle which results in an enzyme that significantly deviates from Michaelis–Menten kinetics (Figure 11). At high glucose concentrations, the transitions to the disordered, inactive state upon product release is bypassed due to rapid glucose binding, and the enzyme runs through a shorter catalytic cycle with a faster rate of catalysis.

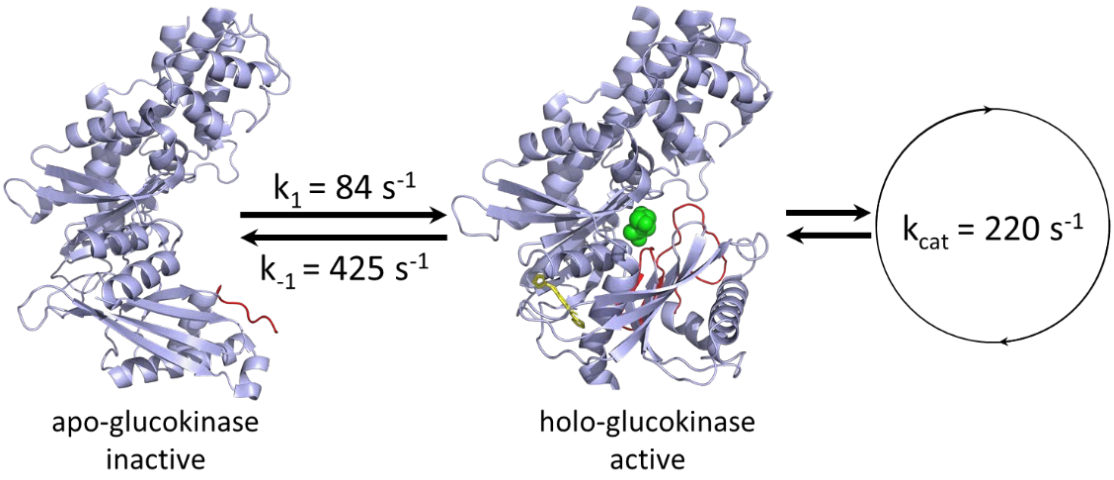


Figure 11. Schematic diagrams showing the model for glucose-dependent kinetic cooperativity in glucokinase. The structure models were generated using the same method as described in Figure 7. The figure is adapted from the paper by Larion *et al.*[86]

Mutagenesis studies highlighted the critical role of the disordered 151-180 loop in long-range allosteric activation. Although prior studies established the model for the kinetic cooperativity in glucokinase, how amino acid substitutions modulate enzyme function from a distant site remained to be elucidated. Combined NMR and proteolysis-based approaches revealed two different mechanisms for substitution-induced allosteric activation, both of which result in highly active and non-cooperative glucokinases.[104] In the α-type mechanism, glucokinase is activated by shifting the ensemble of the free enzyme towards the more ordered glucose-bound state. This is supported by results of NMR measurements which demonstrated a more rigid structure upon substitution. In addition, the susceptibility of the 151-180 loop to proteolysis is decreased, indicating a more ordered active site loop. By comparison, substitutions that activate glucokinase via the β-type mechanism do not affect the overall conformation. Instead, they increase the mobility of the 151-180 loop, which is predicted to enhance the turnover rate by promoting product release. Notably, a number of these substitutions occur at the site where effector binds, suggesting that effector-induced allosteric response may also enhance enzyme activity through the same mechanism (Figure 12). Overall, results of these studies indicate allosteric activation by amino acid substitutions can occur through modulation of the disorder-to-order transition of the active site loop.

In summary, the combined structural, functional, and dynamic studies revealed the mechanism of glucose-dependent kinetic cooperativity in glucokinase and highlighted the importance of the disorder 151-180 loop in allosteric regulation.

Nonetheless, a number of questions remain to be addressed. For example, how the disorder-to-order transition of the active site loop contributes to glucose phosphorylation remains elusive. In addition, studies have demonstrated that residue substitutions at positions distant from the active site loop alter its disorder-to-order transition upon glucose binding, but it is unclear how distant sites communicate across the protein. Finally, the mechanisms of allosteric activation by effector binding remain to be investigated. It is possible that these allosteric activators also enhance glucokinase activity by increasing the mobility of the active site loop.

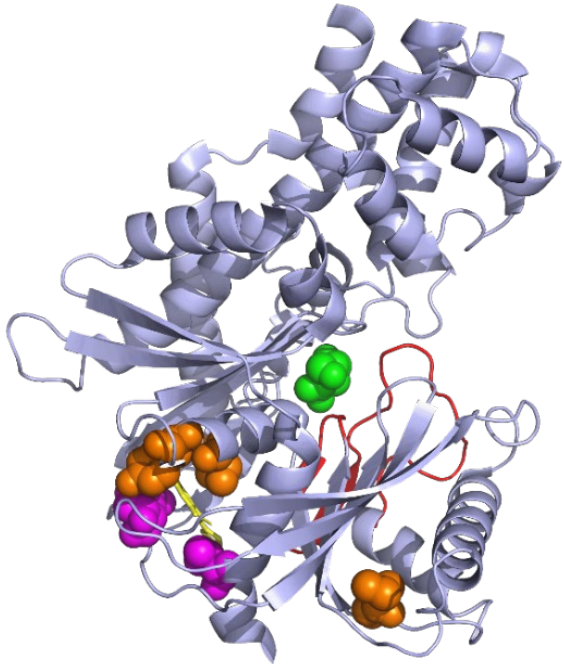


Figure 12. Holo-glucokinase structure (PDB: 1V4S[82]) demonstrating locations of amino acid substitutions that lead to α-type activation (S64P, T65I, V455M) (magenta) or β-type activation (M197V, I211F, Y214C) (orange).[104] The 151-180 loop is colored in red, and the allosteric effector is colored in yellow.

(2) *Staphylococcus aureus* chromosomal Zn-regulated repressor

Regulation of transcription repression by allostery is common in biology, and its mechanisms have been studied extensively in a number of systems.[44, 117, 118] The most important question is how effector binding at one site alters DNA binding affinity at a distant site. The long-range communication between functional sites was once considered to be mediated by structural changes.[43] However, the importance of dynamic changes in allostery cannot be underestimated, and allosteric regulation in some cases is driven entirely by changes in conformational dynamics.[117] In this section, I will briefly describe the investigation of allosteric mechanism in chromosomal Zn-regulated repressor (CzrA) using a combination of biophysical methods. Studies on CzrA also demonstrated the predictive power of computational tools, as results obtained from MD simulations are highly consistent with those of solution measurements.

CzrA is an allosterically regulated Zn(II) sensor protein that controls Zn homeostasis in *S. aureus*. In the absence of the effector molecule, Zn(II), CzrA binds to DNA and represses transcription of the *CzrB* gene which encodes a membrane-bound Zn exporter. When Zn(II) accumulates in the cytoplasm, metal binding releases the protein from the DNA and activates transcription of the *CzrB* gene to export excess Zn(II),[119, 120] which is toxic as it prevents uptake of manganese.[121]

Zn binding-induced allosteric response does not occur via conformational changes. Crystal structures show that CzrA exists as a homodimer, with the metal binding site located at the dimer interface distant from the DNA binding domain

(Figure 13).[122] Despite the strong negative coupling between Zn and DNA binding,[71] both apo- and Zn_2-CzrA share the almost identical "open" conformation, and the structural alignment yielded a global RMSD of 1.6Å.[122] By contrast, the DNA-bound CzrA adopts a more "closed" conformation that is favorable for DNA binding.[52] The lack of structural changes upon Zn binding suggests that dynamics play an important role in CzrA allostery. Consistent with this prediction, ITC measurements show that the +6 kcal/mol coupling free energy between Zn and DNA binding is entropic in origin, indicating a dynamically driven allosteric response.[71]

To better understand the nature of allosteric regulation in CzrA, molecular dynamics (MD) simulations were performed on all four states of CzrA proteins.[123] Consistent with crystal structures, both apo and Zn_2-CzrA adopt an open conformation in simulations, while DNA binding stabilizes the closed conformation. Analyses of simulation data using root-mean-square fluctuations (RMSF) or correlated motions as metrics both indicated that Zn binding quenches internal dynamics near the metal binding site, whereas DNA binding enhances the mobility of these residues. Further investigation revealed a hydrogen bonding pathway that connects Zn and DNA binding sites. As this pathway is only present in the Zn_2-CzrA state, the authors speculated that Zn binding-induced dynamic quenching stabilizes the pathway to prevent CzrA from sampling the high DNA binding affinity conformation. By contrast, the effect of DNA binding may destabilize the hydrogen bonding pathway and ultimately results in increased protein mobility.

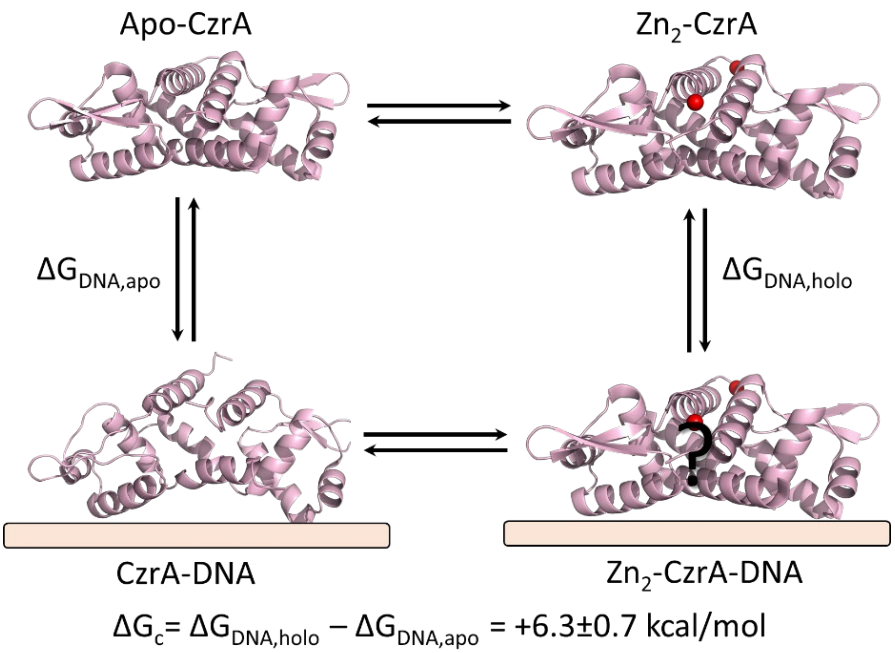


Figure 13. Thermodynamic cycle illustrating the negative coupling between Zn and DNA binding to CzrA.[71] All structure models were generated in Pymol.[22] Apo- and Zn_2-CzrA structures (PDB: 1R1U and 1R1V) were solved by X-ray crystallography,[122] while the structure of CzrA in the DNA-bound form was obtained by NMR spectroscopy (PDB: 2KJB).[52] The structure of Zn_2-CzrA-DNA complex, however, remains to be solved.

The hypothesis generated by MD simulations is supported by results of NMR measurements published 5 years later.[123] NMR measurements performed on apo and Zn_2-CzrA demonstrated that Zn binding is associated with redistribution and quenching of the picosecond to nanosecond (ps-ns) side chain dynamics in the core of the protein. This is consistent with the results of MD simulations and suggests that fast side-chain motions are critical for allostery.[124] As simulations only lasted for 120 ns,[123] Zn-

induced dynamic quenching at microsecond to millisecond (μs-ms) timescales was not detected in the initial study. Nonetheless, mutagenesis experiments showed that changes in slow dynamics alone are insufficient to promote de-repression. For example, double substitutions V66A/L68V and V66A/L68A, which abolish allosteric coupling in CzrA, lead to altered and even complete loss of Zn binding-induced dynamic quenching associated with ps-ns side chain dynamics. By contrast, Zn binding still quenches μs-ms timescale backbone motions identical to that observed in the wild type protein, indicating that slow dynamic changes alone do not modulate allosteric response in CzrA.[124]

Overall, the combined biophysical studies indicate that CzrA allostery occurs through modulation of fast internal dynamics, but not slow motions, upon Zn binding. This work also demonstrates the predictive power of computational approaches, as results of MD simulations are highly consistent with that of NMR measurements. However, more experiments are required to elucidate the mechanism of allosteric communication in CzrA. For instance, MD simulations predicted that the hydrogen bonding pathway is crucial for communication between Zn and DNA binding sites, but evidence supporting this hypothesis is lacking. Moreover, how quenching in side chain dynamics near the Zn binding site alters protein ensemble remains unclear. As μs or even ms timescale simulations are becoming more accessible now, simulations at longer timescales may reveal the linkage between local dynamic changes and alterations in global conformations.

(3) Cellular Src (c-Src) kinase

The mechanisms of long-range communication between functional sites have always been the key information to understanding allosteric regulation. When addressing this challenge, network theory has become increasingly popular as allosteric communication in many proteins has been found to occur through networks of residues.[125-128] Naturally, the identification of residue networks is critical to elucidating the mechanisms of allosteric communication. In this section, combined structural, functional, and computational approaches identified residue networks that may play an important role in inter-domain communication in c-Src kinase.

Cellular Src (c-Src) kinase is a member of mammalian Src family tyrosine protein kinases that function in a large number of cellular signaling pathways including proliferation, differentiation, migration, and survival.[129] Due to its critical role in cellular activities, the enzyme is subjected to tight regulation by kinases and phosphatases,[130] and continuous activation of c-Src kinase has been demonstrated to be oncogenic.[129, 131, 132] However, popular tyrosine kinase inhibitor drugs, such as imatinib (Gleevec)[133] and dasatinib (Sprycel)[134] which target the active site of c-Abl kinase, are not effective against c-Src kinase, even though they share almost identical kinase domains.[135] The failure of conventional anti-cancer drugs suggests that highly conserved active site in c-Src kinase may not be an ideal target. Instead, effector molecules that bind allosteric sites in distant regulatory domains may yield higher selectivity for the protein.

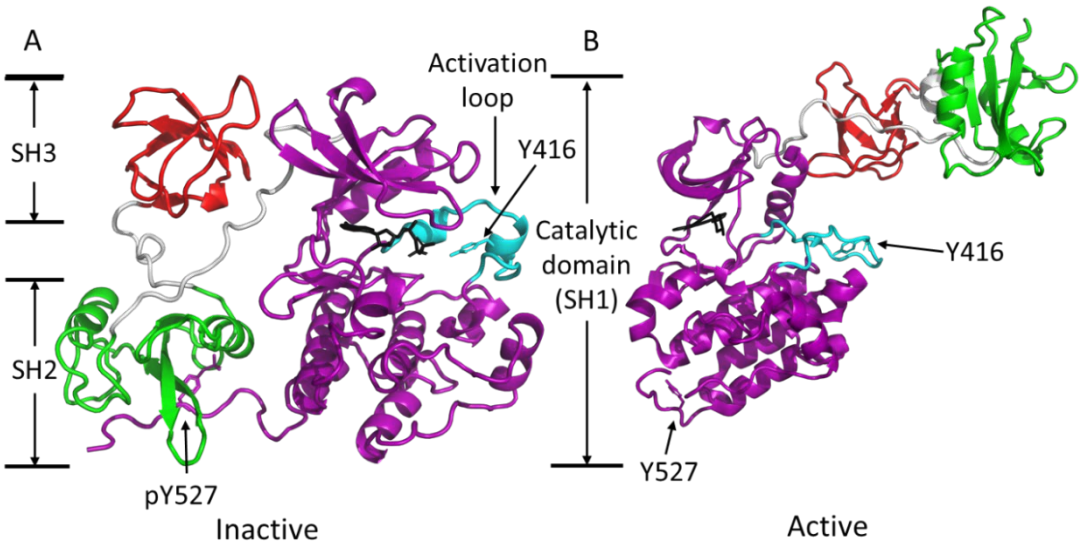


Figure 14. Structures of (A) inactive and (B) active c-Src kinases. Activation of the enzyme by dephosphorylation of pY527 breaks interdomain interactions and relieves the auto-inhibition by SH2 and SH3 regulatory domains. Color code: Purple, the kinase domain; Green, SH2 domain; Red, SH3 domain; Cyan, the activation loop; Black, ligand molecules that stabilize specific conformations. Figures were made in Pymol[22] using PDB files 2SRC[136] and 1Y57[137].

The mechanisms of allosteric regulation in c-Src kinase have been studied extensively using integrated biophysical approaches. The kinase consists of 5 segments from the N-terminus to the C-terminus: four Src-homology (SH4, SH3, SH2, and SH1) domains, followed by a tail at the end of the C-terminus (the C-terminal tail). The SH1 domain is the highly conserved tyrosine kinase domain, while SH2, SH3 domains and the C-terminal tail are important for regulating enzyme activity.[130, 138-140] X-ray crystal structures showed that c-Src kinase activation is accompanied by loss of interdomain

interactions and domain movements. In the native, down-regulated state, c-Src kinase adopts a closed conformation in which regulatory domains are packed and bind to the kinase domain (Figure 14A).[141, 142] Interactions between regulatory domains and the kinase domain are predicted to auto-inhibit the enzyme by stabilizing the inactive conformation. Activation of the enzyme, which usually occurs via dephosphorylation of the residue pY527 or autophosphorylation of residue Y416, leads to the loss of interdomain interactions (Figure 14B). As a result, SH2 and SH3 domains rotate about 130^{o} relative to the inactive conformation, and the kinase domain adopts an open conformation that is active for catalysis.[137]

MD simulations on the kinase domain revealed a residue network that links the active site to the domain interface. Previous studies have established that phosphorylation of residue Y416 in the activation loop, or dephosphorylation of Y527 in the C-terminal tail activates the kinase and results in structural changes from a closed state to an open conformation.[130, 136, 137, 143] However, neither of the two residues are located near the interface between regulatory domains and the kinase domain. To investigate the mechanism of this communication, Foda *et al.* simulated catalytic domain of c-Src kinase in its active and inactive conformations.[91] Direct comparison of simulated structures revealed an extensive network of contacting residues that spans 40Å across the entire kinase domain (Figure 15, left),[91] and activation of the enzyme is accompanied by the rearrangement of network interactions (Figure 15, right).

Functional measurements indicated that perturbations to the network have significant effects on enzyme catalysis. For example, alanine substitution of W260,

which is distant from the active site, greatly enhances the enzyme activity and abolishes allosteric coupling.[91, 138] Protonation of active site residue D404 activates the enzyme and triggers the structural change of the kinase domain to the active conformation.[91, 144] Therefore, Foda *et al.* hypothesized that the residue network identified facilitates the communication between regulatory domains and the kinase domain. However, as simulations were carried out on the kinase domain only, this network only reveals the communication within the kinase domain.

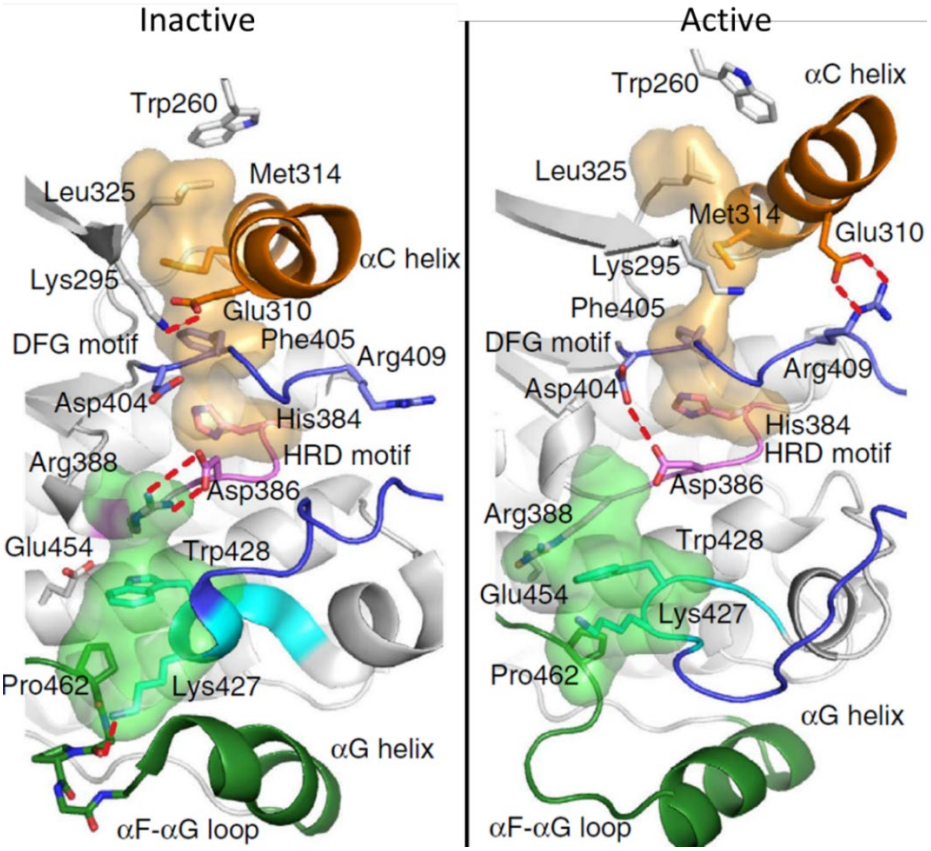


Figure 15. Residue networks identified in the simulated kinase domain in its inactive (left) and active (right) conformations. The figure was adapted from the paper by Foda *et al.*, 2015.[91]

MD simulations performed on the multidomain complex provided more insights into the mechanisms of interdomain allosteric communication in c-Src kinase. In a recent study by Tse and Verkhivker, the authors identified potential allosteric networks in the full complex using a systematic network analysis similar to that described in the previous section.[145] In this approach, the protein is also represented as a 2D graph in which residues are treated as nodes, and interactions between them are drawn as edges. The difference between the two methods is that the edges are weighted by fluctuation cross-correlation between residue pairs rather than interaction energies.[146, 147] Once the graph was established, shortest (most probable) paths between each residue pair were calculated, and node betweenness was determined by the number of shortest paths passing through each node. Nodes with high betweenness values were proposed to have a significant contribution to long-range communication and function in allostery.

Results of the network analysis highlighted a set of high betweenness nodes that may participate in interdomain communication in c-Src kinase (Figure 16).[145] Most of these residues are located in the regulatory domain, but residues in the kinase domain, such as W260, H384, and F405, overlap with the previously identified network.[91] This suggests that allosteric communication within the kinase domain may occur through the network proposed by Foda *et al.*, while interdomain communication is handled by residues identified in this work. Inspection of high betweenness residues showed that they form two networks that bridge SH2 and SH3 regulatory domains to the kinase domain. For example, the network identified in the SH2 domain of the inactive complex

connects the critical pY527 to the active site (Figure 16A). The SH3 domain, which is also important for regulating enzyme function, is predicted to communicate with the kinase domain through linker residues L255 and W260 (Figure 16A). These two residues were hypothesized to act as central nodes in mediating allosteric communication, as they were also identified as bridging residues in the active conformation (Figure 16B). The hypothesis is supported by functional studies showing that alanine substitution of either L255 or W260 results in the enhancement of enzyme activity and loss of allosteric response.[91, 148, 149] Overall, results of MD simulations in this work revealed two residue networks that may be important for allosteric communication between regulatory domains and the kinase domain.

In summary, combined structural and computational analyses identified potential allosteric networks that link regulatory domain with the active site, and functional studies indicated that some of the network residues play a critical role in regulating enzyme function. Nevertheless, the presence of this putative allosteric network requires validation by solution measurements. Site-directed mutagenesis may reveal if these network residues participate in the allosteric response of c-Src kinase, and NMR studies can elucidate if network residues are dynamically coupled.

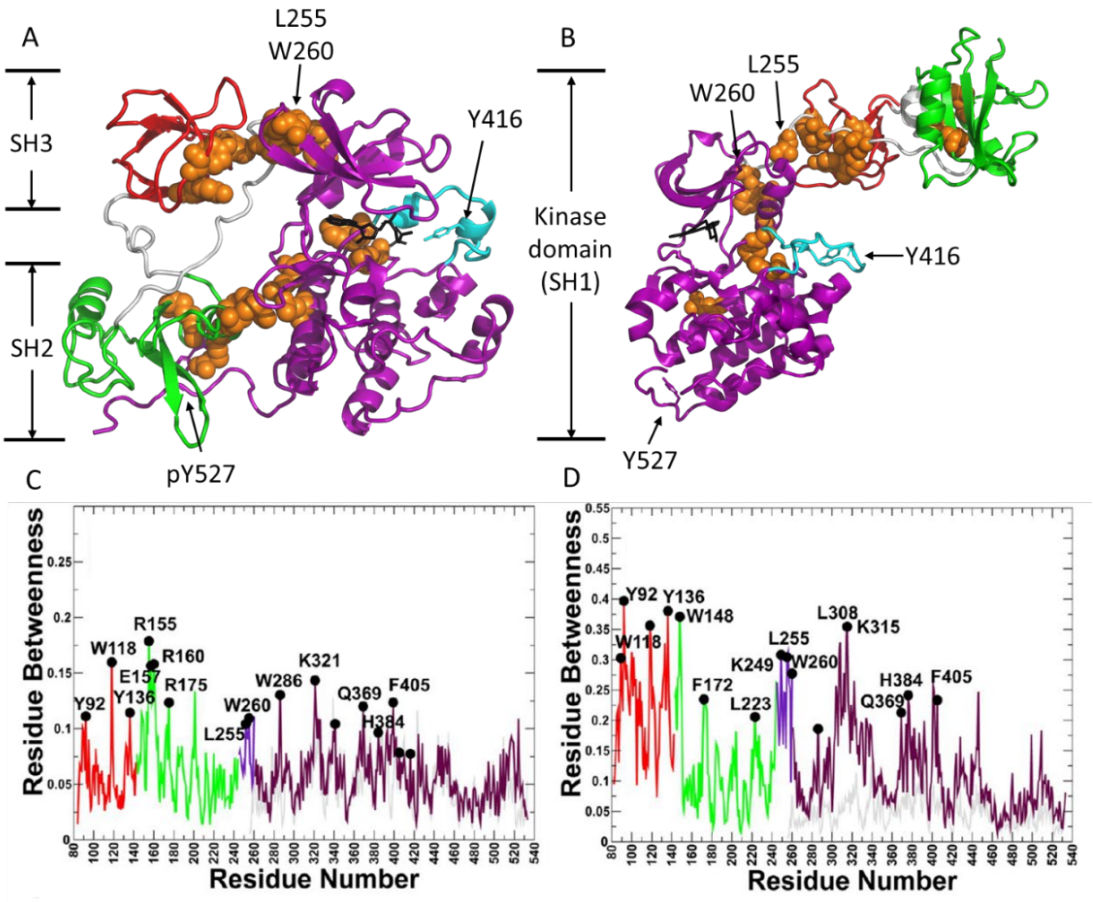


Figure 16. Illustration of high betweenness residues identified by network analysis in c-Src kinase. (A & B) Structures of (A) inactive and (B) active c-Src kinase showing positions of high centrality nodes identified in the analysis. (C & D) Results of network analysis showing the betweenness value for each residue. Figures A and B were made in Pymol[22] using PDB files 2SRC[136] and 1Y57[137]. Figures C and D were adapted from the paper by Tse and Verkhivker, 2015.[145]

B. Allosteric regulation in the *Escherichia coli* biotin protein ligase

Biotin protein ligases, which catalyze biotin transfer to acceptor proteins, are essential for the survival of the majority of organisms. In many bacteria, these enzymes not only participate in post-translational modification but also transcription repression.[150] For example, in the well-studied *E. coli* system, co-repressor binding to its biotin protein ligase (BirA) allosterically activates BirA homodimerization to promote DNA binding and subsequent transcription repression.[151, 152] As allosteric regulation by BirA involves structural and dynamic changes on distant functional surfaces,[153, 154] the protein provides another model system to study the molecular mechanisms of long-range allosteric communication.

As described above, *E. coli* BirA is a bifunctional protein that acts as a metabolic enzyme and a transcription repressor.[152, 155] It first catalyzes the conversion of biotin and ATP to an intermediate, biotinoyl-5'-adenylate (bio-5'-AMP). The resulting BirA·bio-5'-AMP complex, or holoBirA, has two possible fates (Figure 17): It can either homodimerize and bind sequence-specifically to the 40 bp biotin operator (*bioO*) to repress transcription of biotin biosynthetic genes.[156-158] Alternatively, holoBirA can form a heterodimer with the biotin carboxyl carrier protein (BCCP) subunit of acetyl-CoA carboxylase and transfer biotin to the acceptor protein.[155] This activates acetyl-CoA carboxylase to synthesize malonyl-CoA, the starting material for fatty acid synthesis.[159]

Partitioning between the two functions is a kinetic process regulated by the demand for and supply of biotin (Figure 17).[160-163] During the exponential growth when

the demand for biotin is high, increased apoBCCP concentration favors heterodimerization between holoBirA and the acceptor protein, thereby maintaining a constant supply of substrate for fatty acid synthesis. Under conditions when the demand for biotin is low, decreased apoBCCP concentration allows for the accumulation of holoBirA. This leads to elevated homodimerization and transcription repression of biotin biosynthetic genes, which ultimately limits the production of biotin. In summary, the kinetic competition between homo and heterodimerization enables the regulation of a functional switch between metabolism and transcription repression depending on the organism's requirement for nutrients and metabolites.

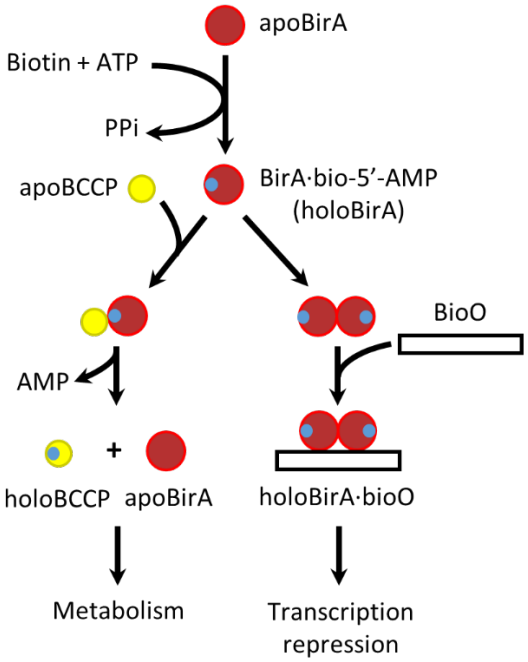


Figure 17. Schematic diagram showing the bifunctionality of *E. coli* BirA. HoloBirA can either heterodimerize with apoBCCP and catalyze biotin transfer, or homodimerize to bind biotin operator (bioO) and repress transcription. The figure is adapted from the paper by Wang and Beckett.[164]

Transcription repression by BirA is an allosterically regulated process. Kinetic studies indicated that assembly of the repression complex occurs in two steps: homodimerization followed by DNA binding (Figure 18).[165] The first step is regulated by allostery, as binding of the effector molecule, bio-5'-AMP, enhances the dimerization equilibrium constant by 1000-fold, or -4 kcal/mol in terms of free energy.[166] In the second step, the resulting homodimer binds to bioO to repress transcription, and DNase I footprinting measurements indicated that both apo and holoBirA dimers have the same affinity for DNA.[166] Given that dimerization is the pre-requisite for DNA binding, the two-step repressor complex assembly is allosterically regulated solely at the dimerization step. From the thermodynamic perspective, the Gibbs free energy of assembling the repressor complex, ΔG_{TOT}, is also enhanced by -4 kcal/mol upon effector binding (Figure 18). These results, combined with prior knowledge on the kinetic partitioning between two dimerization processes, reveal a biotin regulatory system which modulates transcription repression through effects of biotin on BirA dimerization.

Previous studies on the mechanism of allosteric regulation in BirA has established that disorder-to-order transitions on two distant functional surfaces play an important role in the regulation. Crystal structures showed that BirA is composed of three distinct domains: an N-terminal DNA binding domain, a central catalytic domain, and a C-terminal domain that functions in homodimerization and biotin transfer (Figure 19).[153, 154, 167, 168] Comparison of apo and holoBirA structures revealed that most significant structural changes occur on two functional surfaces separated by 30Å

(Figure 19).[153, 154] On the ligand binding surface, the adenylate binding loop (ABL) and the biotin binding loop (BBL) fold around the effector molecule and form a hydrophobic cluster surrounding the adenylate part of bio-5'-AMP. On the dimerization surface, effector binding is coupled to ordering and packing of two surface loops comprised of residues 140-146 and 193-199, as well as the extension of an α-helix from residue A146 to P143 (Figure 19). By contrast, biotin binding, which does not elicit allosteric response, fails to induce disorder-to-order transitions on both surfaces.[169] In the biotin-bound form, the ABL on the ligand binding surface is disordered, and the other surface loops, while more organized compared to apoBirA, are significantly less ordered than those in the holo enzyme.[153, 154, 169]

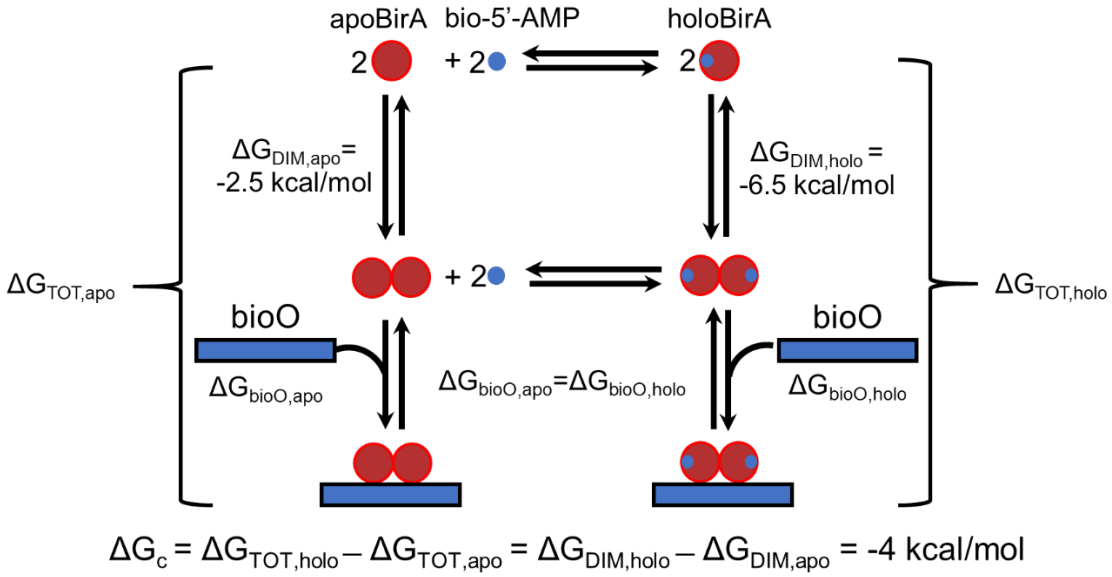


Figure 18. Thermodynamic cycle showing that the repressor-complex assembly is solely dependent on homodimerization. Effector binding enhances the free energies of both dimerization and repressor complex assembly by -4 kcal/mol.

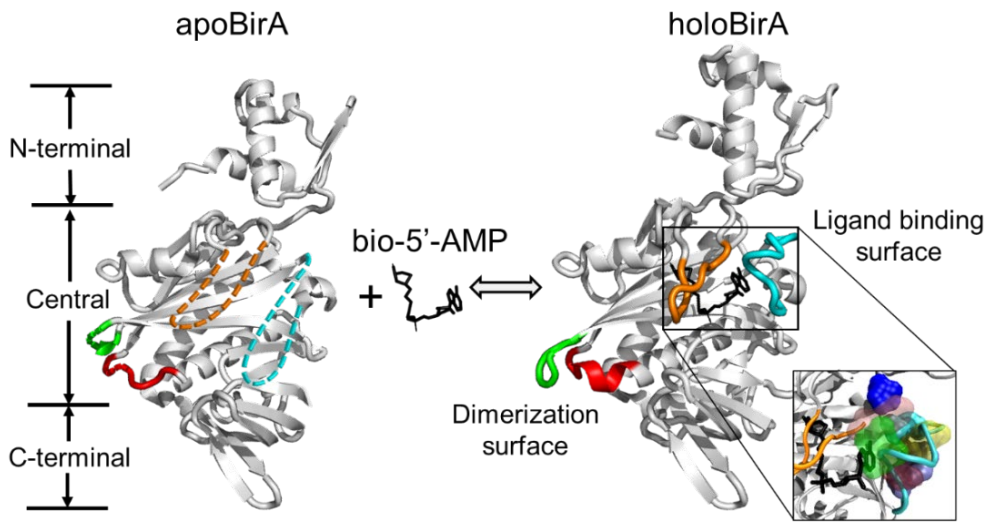


Figure 19. Bio-5'-AMP binding induces disorder-to-order transitions on both functional surfaces. The inset demonstrates the hydrophobic cluster that forms upon effector binding, which includes residues P126 (blue), F124 (pink), M211 (cyan), V214 (yellow), V218 (brown), V219 (green), W223 (lavender). Loops that undergo disorder-to-order transitions are comprised of residues 140-146 (red), 193-199 (green) on the dimerization surface, and 116-124 (orange), 211-222 (cyan) on the ligand binding surface. Figures were adapted from the paper by Wang and Beckett, 2017.[170]

Results of functional measurements support the critical role of disorder-to-order transitions in BirA allostery. To test if disorder-to-order transitions contribute to allosteric regulation, alanine substitutions were made at positions that undergo these transitions on the dimerization and ligand binding surfaces. Functional effects of these substitutions on both input (bio-5'-AMP binding) and output (homodimerization) of BirA allostery were determined by ITC and sedimentation equilibrium, respectively. Results of these measurements showed that most alanine substitutions on the ligand

binding surface, which are predicted to disrupt the hydrophobic cluster, perturb not only bio-5'-AMP binding but also dimerization (Figure 20).[171, 172] These results provide direct evidence that disorder-to-order transitions on the ligand binding surface participate in BirA allostery. Similarly, alanine substitution of G142 on the dimerization surface leads to complete loss of the allosteric response and significantly weaker bio-5'-AMP binding affinity.[173, 174] Moreover, structural analysis indicated that the G142A substitution also results in the loss of disorder-to-order transitions on both functional surfaces,[174] which provides direct evidence supporting the role of disorder-to-order transitions in BirA allostery. Although other alanine substitutions at positions on the dimerization surface only affect holoBirA dimerization (Figure 20),[173, 175] these substitutions probably alter dimerization through their effects on allostery as substituted residues do not contribute directly to the dimer interface. In summary, the combined results of structural and functional studies indicate that allosteric regulation occurs via disorder-to-order transitions on two distant functional surfaces.

Structural and functional studies have demonstrated long-range communication from one functional surface to another in BirA, but its detailed molecular mechanism remained to be elucidated. First, the thermodynamic cycle of BirA allostery predicts reciprocal coupling between distant disorder-to-order transitions, but experimental results supporting this hypothesis were lacking. Second, disorder-to-order transitions play an important role in allostery. However, how these transitions contribute to effector-linked dimerization is unclear. Finally, the mechanisms of long-range communication between the two surfaces remained to be elucidated. In this

dissertation, I will describe the work we have done to address these questions. In the second chapter, double-mutant cycle analysis was used to test reciprocal coupling between functional surfaces, and molecular dynamics simulations revealed structural and dynamic bases of allosteric regulation via disorder-to-order transitions. The mechanism of long-range allosteric communication was investigated in the third chapter, and results of integrated approaches reveal an allosteric network that is coupled to disorder-to-order transitions.

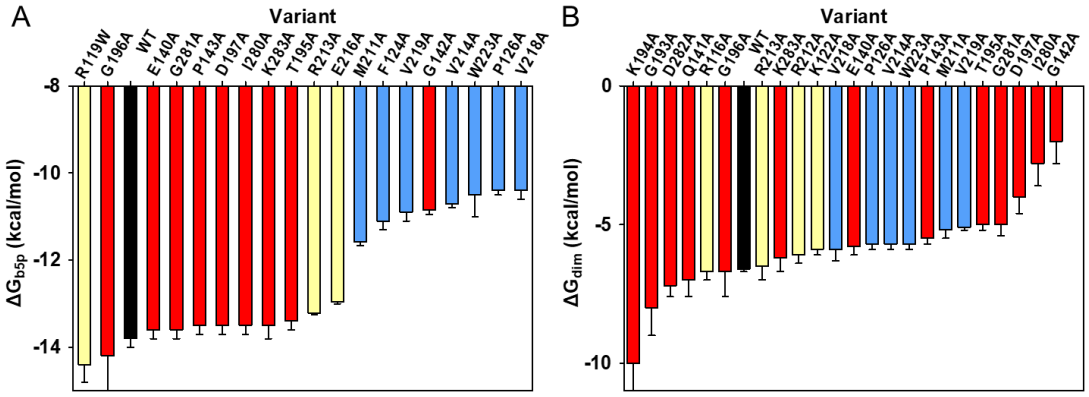


Figure 20. Alanine substitutions at positions on two functional surfaces affect input and output of BirA allostery. Color code: Red, residues on the dimerization surface; Blue: residues in the hydrophobic cluster on the ligand binding surface; Yellow, other residues on the ligand binding surface; Black: wild type. (A) Gibbs free energies of bio-5'-AMP binding to BirA variants. (B) Homodimerization free energies of holoBirA variants.

C. Biotin regulatory systems in other biotin protein ligases

Biotin protein ligases (BPLs) are widespread among bacteria and archaea.[150] These enzymes are classified into two types based on their functionalities: Class I enzymes only catalyze biotin transfer, while Class II BPLs act as transcription repressors and biotin protein ligases. Despite the broad distribution of BPLs among bacteria and archaea, and their importance in the survival of these organisms, few proteins in the family have been subjected to biochemical studies.

Previous studies on *Bacillus subtilis* and *Staphylococcus aureus* BPLs (*Bs* and *Sa*BirA), both of which are class two enzymes, provide some insights into their biotin regulatory system. In *E. coli* BirA (*Ec*BirA), the coupling between bio-5'-AMP binding and dimerization generates a steep dependence of transcription repression on biotin concentration (Figure 21).[176] Similar to *Ec*BirA, both *in vivo* and *in vitro* studies demonstrate that *Bs*BirA is sensitive to biotin concentration, suggesting that *Ec*BirA and *Bs*BirA share the same biotin regulatory system.[177, 178] However, the coupling free energy between bio-5'-AMP binding and dimerization for *Bs*BirA are yet to be determined. By contrast, two independent studies on *Sa*BirA yielded results inconsistent with each other.[178, 179] Sedimentation equilibrium measurements showed that, unlike *Ec*BirA, apo-*Sa*BirA dimerizes with an equilibrium dissociation constant (K_D) similar to that of holo-*Sa*BirA.[179] The measured K_D for apo-*Sa*BirA dimerization is in the micromolar range, suggesting that *Sa*BirA represses transcription regardless of bio-5'-AMP binding. On the other hand, *in vivo* measurements show that transcription repression by *Sa*BirA is dependent on biotin concentration, which

suggests a biotin regulatory system similar to *Ec*BirA.[178] The contradictory results require further investigation of the biotin regulatory systems in *Sa*BirA.

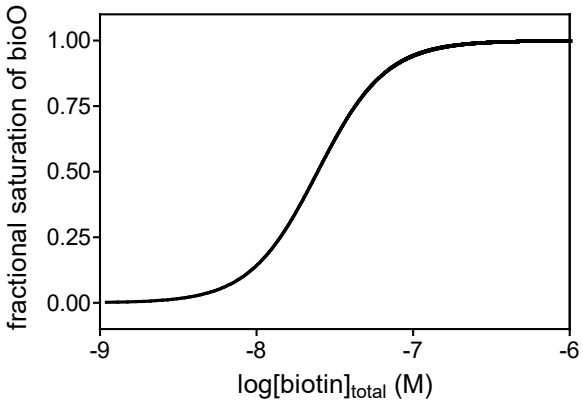


Figure 21. The dependence of biotin operator (bioO) occupancy by BirA dimer on biotin concentration. The figure is adapted from the paper by He and Beckett, 2018.[176]

D. Organization of the dissertation

The first chapter of this dissertation introduces the history of allostery and biophysical approaches to investigate its molecular mechanisms. A literature review is also included to demonstrate how integrated approaches were employed to elucidate allosteric mechanisms in different allosteric systems.

The second chapter describes the work that was published in Biochemistry in 2017. In this work, combined functional and computational approaches were employed to investigate the mechanism of long-range allosteric communication in BirA. Results of functional measurements demonstrated that single alanine substitutions on distant surfaces generate non-additive effects in bio-5'-AMP binding and holoBirA dimerization. This non-additivity provides direct evidence for reciprocal coupling between residues on the ligand binding and dimerization surfaces. Molecular dynamics simulations performed by Dr. Matysiak's lab revealed that these alanine substitutions perturb BirA functions by altering both conformation and dynamics of loops that undergo disorder-to-order transitions. In addition, the coupling in bio-5'-AMP binding is correlated with changes in loop packing on the ligand binding surface. These results indicate that allosteric communication in BirA occurs via modulation of disorder-to-order transitions on two coupled functional surfaces.

The third chapter describes our recent progress in investigating the mechanism of long-range allosteric coupling in BirA. Results of integrated structural, functional, and computational methods demonstrated that disorder-to-order transitions on distant surfaces are communicated via a distributed residue network in the protein. This

network, which was initially identified by structural analysis, was extended by simulation-based network analysis carried out in collaboration with Dr. Matysiak's lab. Proton linkage analysis supported the presence of the network in solution, and functional measurements demonstrated the critical role of this network in BirA allostery. Finally, force distribution analysis, which was also completed by Dr. Matysiak's lab, revealed a connection between disorder-to-order transitions on distant surfaces and residues in the network. The combined results indicate that disorder-to-order transitions and residue network function collectively to mediate long-range allosteric communication in BirA.

The fourth chapter presents the study carried out on biotin protein ligases from other organisms. Previous *in vivo* measurements indicated that *Bacillus subtilis* and *Staphylococcus aureus* biotin protein ligases (*Bs* and *Sa*BirA) repress transcription in a biotin-dependent manner similar to that observed for *E. coli* BirA (*Ec*BirA). Results of sedimentation equilibrium measurements in this study showed that homodimerization of *Bs* and *Sa*BirA, like *Ec*BirA, is significantly enhanced upon bio-5'-AMP binding. Structural alignment of *Ec* and *Sa*BirA also revealed a highly similar allosteric mechanism involving disorder-to-order transitions on distant functional surfaces. However, alignment of the three sequences showed relatively low sequence conservation across the protein, especially for loop segments that undergo disorder-to-order transitions. Overall, these results revealed a conserved allosteric function in *Ec, Bs,* and *Sa*BirA with divergent sequence identity.

The final chapter summarizes the results obtained in this research and their contributions to elucidating allosteric mechanisms in *E. coli* BirA. I will also propose experiments that will further expand our knowledge of BirA allostery and discuss the significance of this work for understanding allosteric regulation in other proteins.

This page is intentionally left blank

Chapter 2: Long distance modulation of disorder-to-order transitions in protein allostery

A. Contribution statement

The work presented in this chapter was completed in collaboration with Dr. Silvina Matysiak's group at the Fischell Department of Bioengineering, University of Maryland, College Park, and it was published in *Biochemistry* in 2017.[170] Dr. Gregory Custer in Dr. Matysiak's group conducted the molecular dynamics simulations and the majority of computational analyses. Our lab performed all non-computational experiments including cloning, protein purification, and functional measurements. We also contributed to the interpretation of simulation results and analysis of their biological significance.

B. Introduction

Allosteric regulation occurs when a signal, typically associated with binding, impinges on one site in a protein to alter function in another, often distant, site. This type of regulation is integral to numerous biological processes, including transmembrane signaling,[180] transcription initiation,[181] and metabolism.[104] Although allostery has been appreciated for well more than a century,[4] the mechanisms by which allosteric signals are transmitted through proteins continue to be a subject of intense research. Protein disorder can play important roles in allostery, and recent experimental results on a number of systems confirm its significance for both the thermodynamics and kinetics of allosteric regulation.[25, 102-104, 117, 124, 182] Approaches that combine

experiment with computation hold great promise for revealing the molecular details of protein allostery.

The *Escherichia coli* BirA protein/biotin repressor provides a model system for elucidation of the molecular basis of allostery. Binding of the small molecule effector bio-5'-AMP is positively coupled to BirA homodimerization, which is a prerequisite for site-specific binding to the biotin operator sequence on DNA (Figure 22B).[165, 183] Operator binding by the holoBirA dimer results in repression of biotin biosynthetic operon transcription initiation.[155, 184] The -4 kcal/mol enhancement of BirA dimerization that accompanies effector binding links metabolic demand for and production of biotin.[160] The BirA structure indicates a distance between the ligand binding and dimerization surfaces of approximately 30 Å (Figure 22A).[153, 154] Moreover, comparison of apo- and holoBirA structures reveals that effector binding is accompanied by disorder-to-order transitions in protein segments on both functional surfaces. Loops on the ligand binding surface containing residues 116-124 [biotin binding loop (BBL)] and 211-222 [adenylate binding loop (ABL)], which are disordered in apoBirA, fold around the bio-5'-AMP ligand to form a hydrophobic cluster in the holoprotein (Figure 22A). Ligand binding is also coupled to the extension of an α-helix by three residues as well as ordering and packing of loops containing residues 140-146 and 193-199 on the dimerization surface.

Experimental studies support critical roles for the disorder-to-order transitions in BirA allosteric signaling. Alanine substitutions at residue positions in the hydrophobic cluster that assembles around the ligand both weaken binding of the BirA

monomer to bio-5'-AMP and compromise coupled dimerization (Figure 22A).[171, 172] Alanine replacements of dimerization surface loop residues, many of which do not contribute directly to the dimer interface, result in altered dimerization.[173] Substitution of glycine at position 142 with alanine in the 140-146 loop on the dimerization surface results in the complete loss of the functional allosteric response and defects in the disorder-to-order transitions on both the dimerization and ligand binding surfaces.[174] Although these experimental results provide evidence supporting the contributions of disorder-to-order transitions to BirA allostery, the mechanism by which these transitions are reciprocally communicated remains to be determined.

Molecular dynamics simulations can provide a fruitful approach to elucidating allosteric mechanisms. For example, coarse-grained and all-atom simulations have yielded insight into the molecular origins of allosteric communication in the GroEL folding chaperone and protein kinase C, respectively.[185, 186] Simulations can provide a detailed molecular understanding of the results of functional measurements and, more importantly, generate hypotheses that can be subjected to further experimental tests.

The thermodynamic cycle relevant to allosteric coupling in BirA (Figure 22B) predicts reciprocal communication between the ligand binding and dimerization surfaces. However, experimental results thus far obtained support only unidirectional communication from the ligand binding to the dimerization surface and vice versa. In this work, results of double-mutant cycle analysis provide direct evidence of reciprocal communication between the two BirA functional surfaces. Results of all-atom molecular dynamics simulations reveal that the communication, and its perturbation by

amino acid substitution, occurs through reciprocal modulation of the conformational and dynamic features of disorder-to-order transitions in the two distant coupled sites.

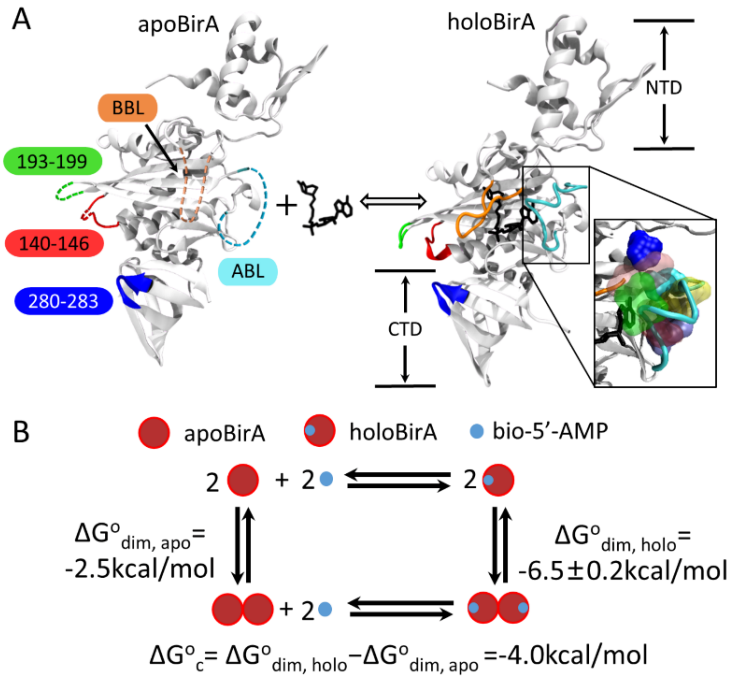


Figure 22. (A) Bio-5′-AMP binding to apoBirA highlighting loops on the ligand binding and dimerization surfaces. Figures were created in VMD[187] with input PDB files 1BIA1[153] for apoBirA and 2EWN[154] for holoBirA. The boxed region illustrates the hydrophobic cluster that forms upon bio-5′-AMP binding: P126 (blue), F124 (pink), M211 (cyan), V214 (yellow), V218 (brown), V219 (green), W223 (lavender). NTD denotes the amino-terminal domain, and CTD denotes the carboxy-terminal domain. (B) Thermodynamic cycle illustrating coupling between bio-5′-AMP binding and BirA dimerization. The coupling free energy, ΔG°_{c}, is defined as the difference between the Gibbs free energy of holoBirA and apoBirA dimerization. No error for $\Delta G^{\circ}_{dim,apo}$ is reported because it is estimated by extrapolation of the value measured in 50 mM KCl.[166]

C. Materials and Methods

1. Chemicals and biochemicals

All chemicals and biochemical were at least reagent grade. The d-biotin (Sigma-Aldrich) solution was prepared quantitatively in standard buffer [10 mM Tris (pH 7.50 ± 0.02 at 20 ºC), 200 mM KCl, and 2.5 mM $MgCl_2$]. Biotinoyl-5'-adenylate (bio-5'-AMP) was synthesized and purified as previously described.[157, 188] Fresh solutions were prepared by dissolving lyophilized powder into Milli-Q H_2O, and bio-5'-AMP concentrations in the resulting solutions were determined by absorption spectroscopy at 259 nm using a molar extinction coefficient of 15400 M^{-1} cm^{-1}.

2. Mutagenesis, expression, and purification of BirA double variants

Coding sequences for BirA double-alanine variants were constructed by site-directed mutagenesis using the Quick-Change II XL Kit (Agilent Technologies) with mutagenic oligonucleotides purchased from Integrated DNA Technologies. In all cases, the mutagenesis template was a pBtac2 plasmid (Boehringer Mannheim) derivative containing the C-terminally $(His)_6$-tagged BirA[189] coding sequence with a mutation at the appropriate codon (P143A, D197A, or M211A). All mutations were verified by sequencing the entire coding sequence (ACGT Inc.).

The BirA variants were expressed and purified as previously reported,[171] with the exception of the addition of a final Q-Sepharose column chromatography step. Protein concentrations were determined using a molar extinction coefficient of 47510 M^{-1} cm^{-1} at 280 nm, which was calculated from the amino acid composition.[190] All

proteins were at least 97% pure as indicated by Coomassie staining of samples subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis.

3. Isothermal titration calorimetry (ITC)

Biotin and bio-5'-AMP binding parameters were measured by ITC using a MicroCal VP-ITC (Malvern Instruments) instrument equipped with a 1.44 mL cell. All titrations were carried out in standard buffer at 20 °C. In direct titrations, 22 × 13 μL of 5-20 μM ligand (biotin or bio-5'-AMP) was titrated into a solution containing the appropriate BirA variant at a concentration of 0.5-2 μM. All *c* values ($c = nK_AC$, where n is the binding stoichiometry, K_A is the equilibrium association constant, and C is the protein concentration) were in the range of 10−1000 required for accurate equilibrium constant determination. The high bio-5'-AMP binding affinity of D197A/F124A BirA [$K_D = (8 \pm 2) \times 10^{-10}$ M] required the use of displacement titrations in which biotin was displaced by bio-5'-AMP.[191] In these titrations, 22 × 12 μL volumes of a 22 μM bio-5'-AMP solution were titrated into a 2.2 μM BirA solution saturated with 4.4 μM biotin.

4. Sedimentation equilibrium

Homodimerization equilibrium association constants of holoBirA variants were measured by sedimentation equilibrium using a Beckman Optima XL-I analytical ultracentrifuge equipped with a four-hole An-60 Ti rotor (Beckman Coulter). Standard 12 mm six-channel cells with charcoal-filled Epon centerpieces and quartz windows were used in all experiments. For all measurements, bio-5'-AMP was combined with protein at a 1.5:1 molar ratio in standard buffer. Samples containing 60, 50, and 40 μM

protein were centrifuged at 18000, 21000, and 24000 rpm, respectively, at 20 °C, and absorbance scans at 300 nm were acquired at each speed after 8 h. An additional scan obtained at each speed 1 h after the first scan confirmed that equilibrium had been achieved.

5. Data analysis

ITC data were analyzed by a nonlinear least-squares method in Microcal Origin 7.0 using a single-site binding model for direct titrations and the competitive model for displacement titrations.

Sedimentation equilibrium data were analyzed using WinNonlin version 1.060.32 Nine absorbance versus radius profiles were globally analyzed using the following equation for a monomer-dimer model to obtain the equilibrium dissociation constant (K_D) for homodimerization:

$$c(r) = c(r_o)e^{\frac{\sigma mon(r^2-r_o^2)}{2}} + \frac{1}{K_D}[c(r_o)]^2 e^{\frac{2\sigma mon(r^2-r_o^2)}{2}} + \delta \quad (1)$$

where $c(r)$ is the protein concentration at position r, δ is the baseline offset, and $c(r_o)$ is the protein concentration at reference radial position r_o. A reduced molecular weight, σ_{mon}, of 1.22 used for the BirA monomer was calculated using equation 2:

$$\sigma_{mon} = \frac{M(1-\bar{\nu}\rho)}{RT}\omega^2 \quad (2)$$

where $\bar{\nu}$ 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. The partial specific volume of 0.755 cm3/g was experimentally determined,[183] and the solvent density was calculated using Sednterp (http://www.jphilo.mailway.com).

6. Molecular dynamics (MD) simulations and analysis

Computational studies described in this work were carried out by Dr. Matysiak's group at the Fischell Department of Bioengineering, University of Maryland, College Park.

MD simulations were performed using the coordinates of a monomer (chain A) of the BirA dimer of Protein Data Bank (PDB) entry 2EWN.[154] This structure, which is of the BirA dimer bound to the corepressor analog biotinol-5'-AMP (btnOH-AMP), was chosen because it is the most complete of the available $BirA^{wt}$ structures, with all residues modeled in the dimerization and ligand binding surfaces. Amino acid substitutions in the protein sequence were made in PyMOL.[192] For simulations, the protein model was placed in a rhombic dodecahedral box with walls extending ~1 nm past the protein and solvated with ~20300 SPC/E34 water molecules. For the R213A simulation, one Na^+ counterion replaced a randomly chosen water molecule to render the system neutral. Prior to production runs, the energy of the system was minimized using the steepest descent method, followed by NVT and NPT equilibration runs of 100 ps each, using position restraints with a force constant of 1000 kJ mol^{-1} nm^{-2} on the protein. Production runs were performed without position restraints, using an NPT ensemble with a temperature of 300 K and a pressure of 1 bar. These runs were carried out for 1 μs, with the last 500 ns of the simulation used for all analysis as, on the basis of global root-mean-square deviation (RMSD) of backbone positions in the central domain relative to starting structure, all simulated variants equilibrate within the first 500 ns of simulation time (Figure 23).

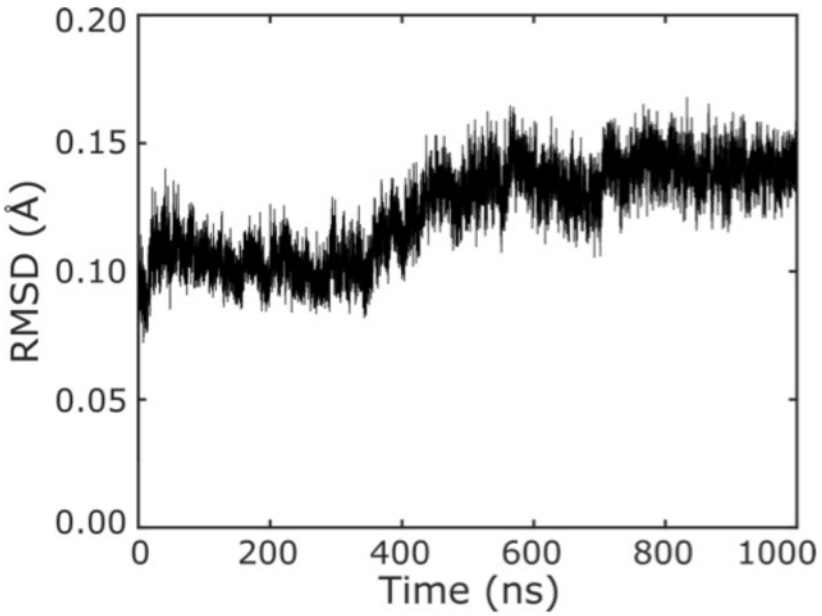


Figure 23. Global root-mean-square deviation (RMSD) time series for backbone positions (C, CA, and N atoms) of the central domain (residues 80-115,127-139, 147-192, 200-210, and 224-269) in the apoBirA^{wt} simulation, which has the longest equilibration time. Global RMSD values were generated by aligning backbone atoms of each frame of simulated structure to the same atoms in the starting structure.

All MD simulations were conducted using the GROMACS 4.6 simulator[193-196] and the OPLS-AA force field.[197] Parameters for btnOH-AMP were constructed by analogy to existing molecule/functional group force fields, as done by others,[198] and are available upon request. Simulations were run using a time step of 2 fs, with neighbor list updates every five steps. Temperatures of protein and water were maintained independently using the V-rescale algorithm[199] with a time constant of 0.1 ps. Ions, where present, were grouped with water for temperature coupling, while btnOH-AMP was grouped with protein. Isotropic pressure coupling in NPT simulations used the

Parrinello-Rahman barostat[200] with a time constant of 2 ps and a compressibility of 4.5 × 10^{-5} bar^{-1}. Bond lengths were constrained using the LINCS algorithm.[201]

MD structure images were rendered using VMD.[187] To identify the representative structure shown for the dimer interface of each variant, a pairwise RMSD matrix was calculated, using the GROMACS tool g_cluster,[193-196] for residues 140−146 and 193−199 in the dimer interface using all structures from the last 500 ns of each trajectory. The structure with the minimum dimer interface RMSD relative to all other structures was then taken as the representative structure. As shown, each representative structure has been aligned to the backbone of the btnOH-AMP-bound representative structure, including residues 135-155 and 189-204.

For pairwise energy difference plots, the Coulombic and Lennard-Jones energies between each pair of residues were output over time using GROMACS.[193-196] The Coulombic and Lennard-Jones energies for each pair were then summed and averaged for the last 500 ns of each trajectory. The energy shown in the difference map is the average energy for each pair minus the average energy for that pair in the holoBirA simulation. As the average energy for all pairs shown was negative (attractive), more negative values would indicate stronger interaction. In the difference maps, this means that negative values are found for a pair when their interaction is stronger than in holoBirA^{wt} and positive values when the interaction is weaker.

Helicity in the simulation of residues 140-152 was quantified by measuring the per-residue average RMSD of an ideal helix (RMSDhx). As both 3_{10}- and α-helices form in the segment, both were considered in this analysis. Ideal helices were

constructed using heavy backbone atoms only (N, CA, C, and O), setting the φ and ψ angles to -49° and -26°, respectively, for the α-helix and to -57° and -47°, respectively, for the 3_{10}-helix.[202] A five-residue ideal α-helix and 3_{10}-helix were used to calculate RMSDα and RMSD310, respectively. For each residue, i, in the range of residues 140-152, the five-residue ideal helices were aligned to a five-residue segment centered on residue i, and RMSD was calculated between the structures. The smaller of either RMSDα or RMSD3_{10} was taken as the RMSDhx for that residue in each structure. The RMSD shown is the average per-residue RMSDhx across the last 500 ns of each trajectory.

Principal component analysis was performed on the last 500 ns of each trajectory. The covariance matrix was calculated for the main-chain backbone atoms (N, CA, and C) of residues 140-146 and 193-199 in the dimer interface, using the GROMACS program g_covar. Root-mean-square fluctuations of these residues along the principal eigenvectors of the covariance matrix were calculated using the GROMACS program g_anaeig.[193-196]

D. Results

1. Molecular design of BirA variants with double-alanine substitutions

Communication between the ligand binding and dimerization surfaces of BirA was investigated using double-mutant cycle analysis. In this approach, the functional effect of introducing alanine substitutions at positions 1 and 2, one on each surface, was compared to the sum of the effect observed for the singly substituted parent variants.[203-205] The effect of substituting two residues simultaneously is quantified using the following equation:

$$\Delta\Delta G^{o}_{12} = \Delta G^{o}_{12} - \Delta G^{o}_{wt} \quad (3)$$

where $\Delta\Delta G^{o}_{12}$ is the difference between the measured free energy of bio-5'-AMP binding or dimerization for the double-alanine variant, ΔG^{o}_{12}, and that of the wild-type protein, ΔG^{o}_{wt}. Functional coupling between two residues is assessed by comparing the magnitude of $\Delta\Delta G^{o}_{12}$ to the sum of the energetic effects of each parent single-alanine substitution on either function, which are expressed as

$$\Delta\Delta G^{o}_{1} = \Delta G^{o}_{1} - \Delta G^{o}_{wt} \quad (4)$$

and

$$\Delta\Delta G^{o}_{2} = \Delta G^{o}_{2} - \Delta G^{o}_{wt} \quad (5)$$

where $\Delta\Delta G^{o}_{1}$ and $\Delta\Delta G^{o}_{2}$ are the energetic consequences of single-alanine substitutions at positions 1 and 2, respectively, for each function, and their sum is

$$\Delta\Delta G^{o}_{1+2} = \Delta\Delta G^{o}_{1} + \Delta\Delta G^{o}_{2} \quad (6)$$

Coupling between two residues is indicated by the following inequality:

$$\Delta\Delta G^{o}_{12} \neq \Delta\Delta G^{o}_{1+2} \quad (7)$$

The magnitude of the energetic coupling, ΔG_c^o, between the two residues is defined as

$$\Delta G_c^o = \Delta\Delta G_{12}^o - \Delta\Delta G_{1+2}^o \quad (8)$$

Previous studies have demonstrated that several single-alanine substitutions on the BirA dimerization and ligand binding surfaces alter bio-5'-AMP binding and/or dimerization.[171-173] In the construction of the BirA double-alanine variants, it was important to select two residues that, when simultaneously replaced with alanine, could yield a protein with dimerization energetics accessible to the sedimentation equilibrium technique. With that restriction in mind, three BirA variants, P143A/M211A, D197A/F124A, and G281A/M211A, were constructed (Figure 24). Single-alanine substitution at M211 or F124 on the ligand binding surface alters the Gibbs free energy of bio-5'-AMP binding by 2-3 kcal/mol, while single-alanine substitutions at P143, D197, and G281 on the dimerization surface have no effect (Table 3).[172, 175] The dimerization penalties for the singly substituted variants range from 0.5 to 2.3 kcal/mol[172, 173] (Table 4).

Biotin binding measurements performed using ITC as described below indicated that all variants bind to the ligand with parameters similar to those measured for $BirA^{wt}$ (Table 5), which indicates that double alanine substitutions do not perturb protein folding.

Table 3. Bio-5'-AMP binding parameters of BirA variants[a]

Protein[b]	K_D (M)	ΔG° (kcal/mol)	ΔH° (kcal/mol)	-TΔS° (kcal/mol)	n
wt[c]	$5\pm2\times10^{-11}$	-13.8±0.2	-15±1	1±1	0.93±0.01
P143A[c]	$8\pm2\times10^{-11}$	-13.5±0.2	-12.20±0.01	-1.3±0.2	0.96±0.01
D197A[c]	$9\pm3\times10^{-11}$	-13.5±0.2	-12.4±0.2	-1.1±0.3	1.21±0.02
G281A[c]	$8\pm3\times10^{-11}$	-13.6±0.2	-12.7±0.2	-0.9±0.3	0.89±0.01
F124A	$5\pm2\times10^{-9}$	-11.1±0.2	-19.3±0.2	8.2±0.3	0.80±0.06
M211A	$2.4\pm0.4\times10^{-9}$	-11.58±0.09	-16.3±0.1	4.7±0.1	1.01±0.02
P143A/M211A	$1.1\pm0.3\times10^{-9}$	-12.0±0.1	-17.4±0.4	5.4±0.4	0.86±0.03
D197A/F124A	$8\pm2\times10^{-10}$	-12.2±0.1	-15.4±0.2	3.2±0.2	0.83±0.02
G281A/M211A	$1.7\pm0.5\times10^{-9}$	-11.8±0.2	-16.9±0.4	5.1±0.4	0.97±0.04

[a]All experiments were performed in Standard Buffer at 20 °C. The reported uncertainties are standard deviations from at least three independent measurements. The Gibbs free energies of and entropic contributions to binding were calculated using the equations $\Delta G° = RT\ln K_D$ and $\Delta G° = \Delta H° - T\Delta S°$, respectively. K_D is the equilibrium dissociation constant of bio-5'-AMP binding. [b]Residue positions on the dimerization surface are shown in bold. [c]Previously published values.[175]

Table 4. Dimerization properties of wt and variant holoBirA proteins[a]

Protein[b]	K_D (M)	ΔG° (kcal/mol)
wt[c]	$14\pm6\times10^{-6}$	-6.4±0.2
P143A[d]	$8\pm2\times10^{-5}$	-5.5±0.2
D197A[d]	$1.0\pm0.6\times10^{-3}$	-4.1±0.3
G281A[d]	$18\pm8\times10^{-5}$	-5.0±0.3
F124A[e]	$2\pm0.8\times10^{-5}$	-6.2±0.2
M211A[e]	$14\pm6\times10^{-5}$	-5.2±0.2
P143A/M211A	$1.0\pm0.5\times10^{-3}$	-4.0±0.2
D197A/F124A	$2\pm1\times10^{-4}$	-3.7±0.2
G281A/M211A	$4\pm1\times10^{-4}$	-4.6±0.2

[a]All measurements were carried out in Standard Buffer at 20 °C. The reported uncertainties are standard deviations from at least three independent measurements. The Gibbs free energy of binding was calculated using the equation $\Delta G° = RT\ln K_D$, where K_D is the equilibrium dissociation constant of dimerization. [b]Residue positions on the dimerization surface are shown in bold. [c]Previously published results from reference 206.[206] [d]Previously published results from reference 173[173]. [e]Previously published results from reference 172.[172]

Table 5. Biotin binding parameters of BirA double variants are similar to those of wt.

Protein	K_D (M)	ΔG° (kcal/mol)	ΔH° (kcal/mol)	-TΔS° (kcal/mol)	n
wt	$4.7\pm0.8\times10^{-8}$	-9.8±0.1	-20.1±0.4	10.3±0.4	0.93±0.01
P143A/M211A	$4.1\pm0.4\times10^{-8}$	-9.91±0.06	-20.4±0.3	10.5±0.3	0.78±0.01
D197A/F124A	$3.7\pm0.4\times10^{-8}$	-9.97±0.06	-22.8±0.4	12.8±0.4	0.90±0.05
G281A/M211A	$3.5\pm0.5\times10^{-8}$	-10.01±0.09	-20.7±0.2	10.7±0.2	0.84±0.03

ITC measurements were performed in Standard Buffer at 20 °C. The reported uncertainties are standard deviations from at least three independent measurements. The Gibbs free energies of and entropic contributions to binding were calculated using the equations $\Delta G° = RT\ln K_D$ and $\Delta G° = \Delta H° - T\Delta S°$, where K_D is the equilibrium dissociation constant of biotin binding.

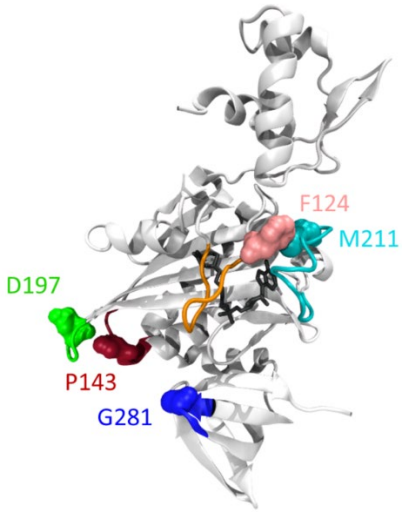


Figure 24. HoloBirA^{wt} model (PDB entry 2EWN[154]) showing the positions of dimerization surface residues P143 (red), D197 (green), and G281 (blue) and ligand binding surface residues F124 (pink) and M211 (cyan).

2. Bio-5'-AMP binding measurements reveal coupling between residues on the two BirA surfaces

Binding of bio-5'-AMP to the three BirA variants was measured using ITC. The range of thermodynamic parameters associated with binding of the effector to the variants necessitated the use of two different approaches for the measurements. The tight binding of bio-5'-AMP to $BirA^{wt}$ requires the use of the competitive binding technique, a method in which bio-5'-AMP is titrated into a BirA solution that is saturated with biotin to compete off the biotin.[191, 207] The resulting data are analyzed using a competitive binding model to obtain the parameters governing bio-5'-AMP binding. In these titrations, the measured heat signal reflects the difference between the heats of binding of bio-5'-AMP to and dissociation of biotin from the protein. For $BirA^{wt}$, this difference of approximately 6 kcal/mol is sufficiently large for accurate measurement.[207] The same is true for the D197A/F124A double variant (Table 3).

By contrast, the molar enthalpies of bio-5'-AMP binding are significantly more negative for the P143A/M211A and G281A/M211A variants. Consequently, the net binding enthalpies for these variants in displacement titrations are too small for reliable measurement, which necessitated the use of direct bio-5'-AMP titrations for these proteins. To achieve acceptable *c* values of 300-400 ($c = nK_AC$, where n is the binding stoichiometry, K_A is the equilibrium association constant, and C is the protein concentration), the titrations were performed at a relatively low protein concentration of 0.5 μM. Nevertheless, data obtained from the measurements are of high quality and described well by a single-site binding model (Figure 25A).

The bio-5'-AMP binding measurements reveal coupling between the ligand binding and dimerization surfaces for two of the variants. For $BirA^{P143A/M211A}$ and $BirA^{D197A/F124A}$, the addition of an alanine substitution to the dimerization surface partially "rescues" the binding defect associated with the alanine substitution in the ligand binding surface as evidenced by the negative or favorable values of the coupling free energies of -0.7 ± 0.3 and -1.4 ± 0.4 kcal/mol, respectively (Figure 25C and Table 6). The detailed thermodynamics of the single- and double-variant binding parameters indicate that for the P143A/M211A variant the coupling is enthalpic in origin while for the D197A/F124A variant it is entropically based. No coupling in bio-5'-AMP binding is observed between residues M211 and G281 (Figure 25C and Table 6).

3. Homodimerization measurements reveal coupling between G281 and M211

The dimerization energetics of the bio-5'-AMP-bound variants were determined by sedimentation equilibrium. For all variants, the relatively modest dimerization free energies necessitated the performance of the measurements at holoBirA concentrations in the range of 40-60 μM. Moreover, even at these high concentrations, a maximal dimerization fraction dimer of approximately 15% was indicated in the analysis of the equilibrium species distributions. Nevertheless, global analysis of the data using a monomer-dimer model yielded equilibrium dissociation constants and Gibbs free energies with good precision. Moreover, the residuals of the analyses indicate excellent agreement between the data and model (Figure 25B). For all doubly substituted variants, the dimerization was significantly weaker than that measured for either singly substituted parent (Table 4). The free energies of

dimerization for the singly and doubly alanine substituted variants for the G281/M211 pair indicate a coupling free energy of -0.9 ± 0.4 kcal/mol. No coupling in dimerization is observed for the P143A/M211A and D197A/F124A variants (Figure 25D and Table 6).

Table 6. Thermodynamics of bio-5'-AMP binding and dimerization properties for BirA variants and coupling between residues on two function surfaces

Protein[a]	$\Delta G°_{b5p}$ (kcal/mol)	$\Delta\Delta G°_{b5p}$ (kcal/mol)	$\Delta G°_{c,b5p}$ (kcal/mol)[b]	$\Delta G°_{dim}$ (kcal/mol)	$\Delta\Delta G°_{dim}$ (kcal/mol)	$\Delta G°_{c,dim}$ (kcal/mol)[c]
wt	-13.8±0.2	-	-	-6.4±0.2	-	-
P143A	-13.5±0.2	0.3±0.3	-	-5.5±0.2	0.9±0.3	-
D197A	-13.5±0.2	0.3±0.3	-	-4.1±0.3	2.3±0.4	-
G281A	-13.6±0.2	0.2±0.3	-	-5.0±0.3	1.4±0.4	-
F124A	-11.1±0.2	2.7±0.3	-	-6.2±0.2	0.2±0.3	-
M211A	-11.58±0.09	2.2±0.3	-	-5.2±0.2	1.2±0.3	-
P143A/M211A	-12.0±0.1	1.8±0.2	-0.7±0.3	-4.0±0.2	2.4±0.3	0.2±0.4
D197A/F124A	-12.2±0.1	1.6±0.2	-1.4±0.4	-3.7±0.2	2.7±0.3	0.1±0.4
G281A/M211A	-11.8±0.2	2.0±0.3	-0.4±0.4	-4.6±0.2	1.8±0.3	-0.9±0.4

[a]Residue positions on the dimerization surface are shown in bold. [b&c]Coupling free energies in [b]bio-5'-AMP binding and [c]holoBirA dimerization calculated using equation 8. The numbers reported were calculated based on unrounded values.

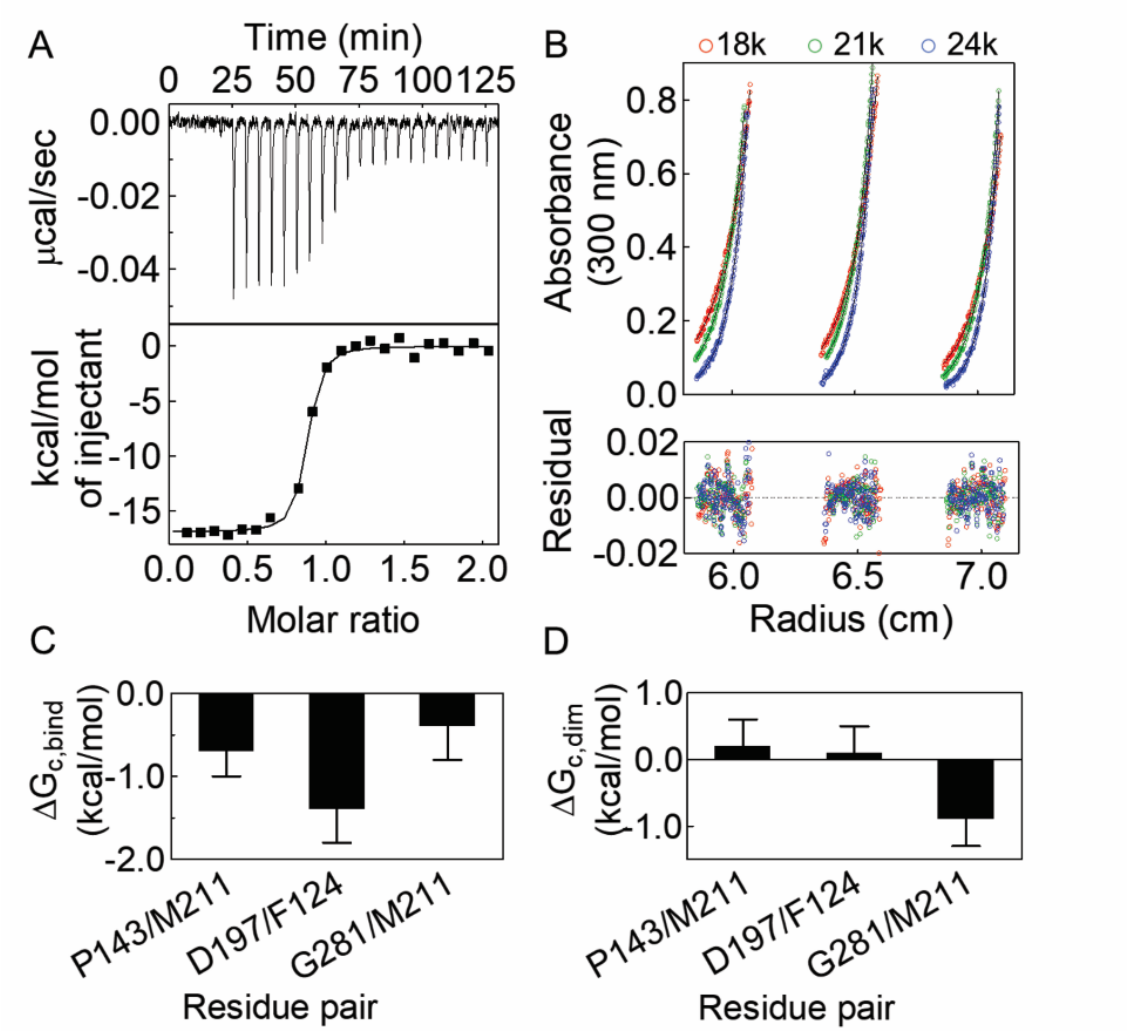


Figure 25. Coupling between distant residues in ligand binding and dimerization. (A) ITC titration of BirA$^{P143A/M211A}$ with bio-5′-AMP at 20 °C (top) and the best-fit curve obtained using a single-site binding model (bottom). (B) Sedimentation equilibrium measurements for holoBirA$^{P143A/M211A}$. Solid lines represent best fits to a monomer-dimer model obtained from global analysis of nine data sets (top) with residuals of the fit (bottom). (C and D) Histograms showing coupling free energies of residue pairs in bio-5′-AMP binding ($\Delta G_{c,bind}°$) and dimerization ($\Delta G_{c,dim}°$).

4. Molecular origins of functional coupling in BirA

The structural and dynamic origins of the observed coupling between residues P143 and M211 were investigated using all-atom molecular dynamics simulations. All simulations were carried out by Dr. Gregory Custer in Dr. Matysiak's lab. Simulations reached equilibrium in 500 ns (Figure 23), and analysis was performed on representative structures from the final 500 ns of each 1 μs trajectory. Simulations of apoBirA^{wt} and holoBirA^{wt} (biotinol-5'-AMP-bound) protein monomers were first carried out to validate the approach. For both species, the simulations were on the monomer that had been extracted from the holoBirA dimer (PDB entry 2EWN), with the apoprotein species generated by removing the biotinol-5'-AMP ligand from the holoBirA monomer structure. Crystal structures indicate that effector binding results in ABL and BBL folding around the bio-5'-AMP in the ligand binding region and disorder-to-order transitions and concomitant packing of the 140-146 and 193-199 loops on the dimerization surface.[153, 154] Therefore, simulation results were first analyzed by comparing pairwise interaction energies for residues in these two regions in the equilibrium structures of the two species. Comparison of the interaction energy maps for holo and apoBirA^{wt} indicates that, consistent with experimentally observed ligand-linked disorder-to-order transitions of surface loops, fewer interactions are present in the ligand binding and dimerization surfaces of the apo-repressor than in the holo-repressor (Figure 26A,C). Thus, for the wild-type protein, the simulations yield results that are consistent with experimentally determined structures.

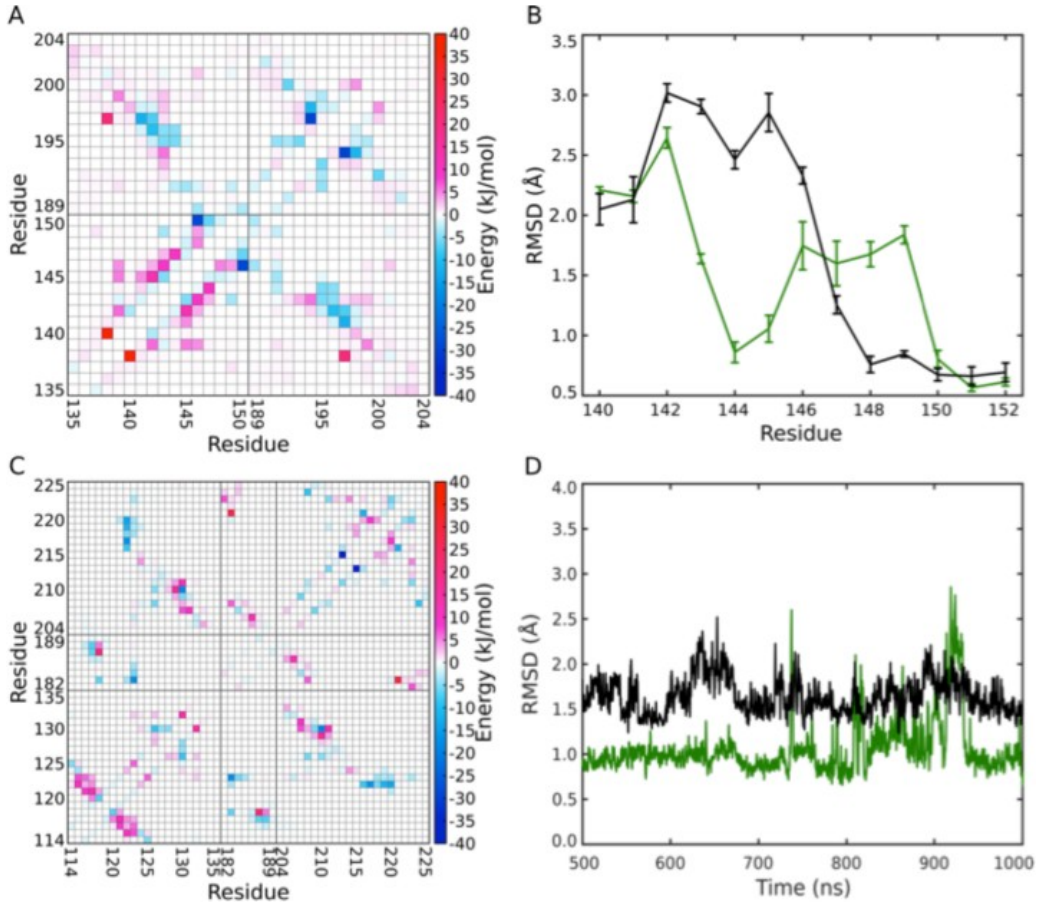


Figure 26. Comparison of wt apo and holo simulations. (A) Difference average interaction energy map for residues pairs in the dimerization surface of the wt apo simulation. (B) Per-residue average RMSD to ideal helix for wt apo (black) and holo (green). Error bars represent 95% confidence intervals. (C) Difference average interaction energy map for residues pairs in the ligand binding region of the wt apo simulation. (D) Root-mean-square deviation (RMSD) time series for C, CA, and N atoms of residues 211- 223, after aligning to C, CA, and N atoms of the central domain (residues 80-115,127-139, 147- 192, 200-210, and 224-269). Values smoothed using a sliding average with a 0.5 ns window. BirA variants shown are wt apo (black) and wt holo (green).

5. Simulated structures indicate diversity in holoBirA variant monomers

Variants with alanine substitutions at M211, P143, and P143/M211 were also subjected to simulations, all in their holo-monomer forms, by Dr. Matysiak's group. In experiments, alanine substitution at M211 results in weaker bio-5'-AMP binding and dimerization.[172] The P143A substitution on the dimerization surface results in no perturbation of bio-5'-AMP binding but does alter homodimerization.[173, 175] Finally, as indicated in the results of double-mutant cycle experiments described above, residues P143 on the dimerization surface and M211 on the ligand binding surface are functionally coupled in bio-5'-AMP binding.

The structure of each variant differs at multiple residue positions from that of holoBirA^{wt}. The average structure from simulations for each variant was first compared to that of holoBirAwt by performing pairwise three-dimensional (3D) alignments using trjconv in GROMACS.[196] The alignments yield backbone RMSD difference plots, which for the sake of clarity include only residues for the protein's central and C-terminal domains (Figure 27). The DNA binding domain was omitted because the available BirA^{wt} structures indicate a high degree of structural variability both within the domain and in its position relative to the central domain.[154] The plots indicate that structural differences between each variant and the wild-type protein are distributed throughout the central and C-terminal domains (Figure 27). Moreover, in the BirA 3D structure, the majority of these differences are located at loop positions.

Residues in the loops that play well-characterized roles in ligand binding and dimerization show variant-specific structural deviations (Figure 27). For example, the

structure of the ABL of only the M211A variant differs significantly from its structure in holoBirAwt. For dimerization surface loops comprised of residues 140-146 and 193-199, structural differences are observed in $BirA^{P143A}$ and $BirA^{P143A/M211A}$. However, while only the 140-146 loop shows a large deviation in the single variant, both loops diverge significantly for the double variant. In all variants, the structure of the BBL differs from its structure in $holoBirA^{wt}$, with the $holoBirA^{M211A}$ loop exhibiting the largest deviations. Discussion of detailed structural and dynamic differences in the ligand binding and dimerization loops of the variants is provided below.

Structural differences are observed for a few additional central and C-terminal domain residues of all variants. The RMSD for residue A166, which is located near the protein's ligand binding surface, is relatively large and roughly equal for the three variants. In the P143A variant, C-terminal residues D282 and I305 show relatively large deviations from their structures in $holoBirA^{wt}$, which are significantly smaller in the other two variants. All three of these residues are far from the locations of amino acid substitutions in the BirA structure.

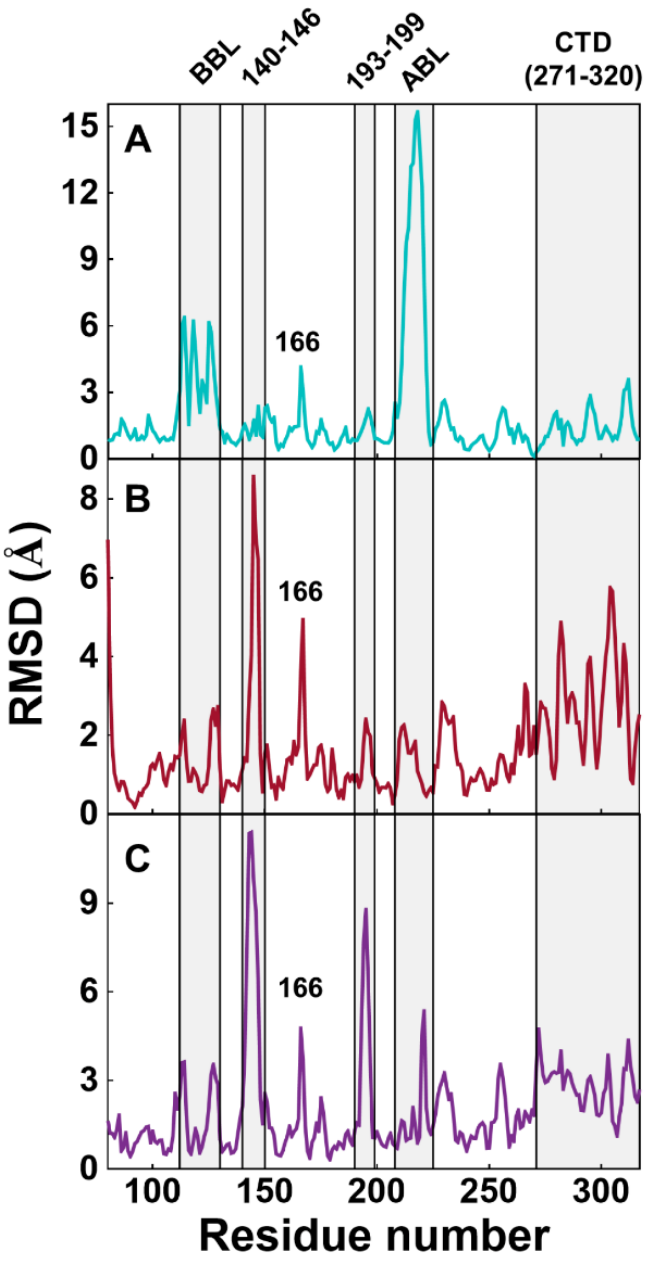


Figure 27. HoloBirA variant structures differ significantly from that of holoBirA^{wt}. The plots show the root-mean-square deviation of backbone atom positions for each holo variant monomer relative wild-type BirA: (A) BirA^{M211A}, (B) BirA^{P143A}, and (C) BirA$^{P143A/M211A}$.

6. Interaction energies on the ligand binding and dimerization surfaces are altered in the holoBirA variants

Comparison of the energy maps of the simulated holoBirA variant structures provides structural insight into energetic perturbations associated with single-alanine substitutions at M211 or P143 and the coupling between these two residues in bio-5'-AMP binding. For each variant, a "difference" average interaction energy map was obtained by subtracting the energy map for holoBirA^{wt} from that of the variant (Figure 28). Consistent with the experimentally observed compromised binding of BirA^{M211A} to bio-5'-AMP (Figure 28A), the difference energy map for the ligand binding region of this variant shows a significant loss of interactions both within and between the ABL and BBL. By contrast, the ligand binding surface energy map for BirA^{P143A}, which binds to bio-5'-AMP with an affinity identical to that of BirA^{wt}, shows only minor differences from that of holoBirA^{wt} (Figure 28B). Combining the P143A substitution on the dimerization surface with the M211A substitution yields interaction energies on the ligand binding surface that are more similar to those of holoBirA^{wt} (Figure 28C), consistent with the experimentally observed ability of the P143A substitution to partially reverse the penalty to bio-5'-AMP binding associated with the single-alanine substitution of M211.

Interaction energy maps also reveal perturbations on the dimerization surface of the alanine variants. These maps provide information about interactions within and between the 140-146 and 193-199 loops. Despite its weaker dimerization, the map for the M211A variant shows little difference from that of holoBirA^{wt} (Figure 28D). The

dimerization surface energy map for holoBirA^{P143A} indicates loss of interactions localized to the 140-146 loop (Figure 28E). The P143A/M211A double variant shows loss of interactions both within the 140-146 loop and between this loop and the 193-199 loop (Figure 28F).

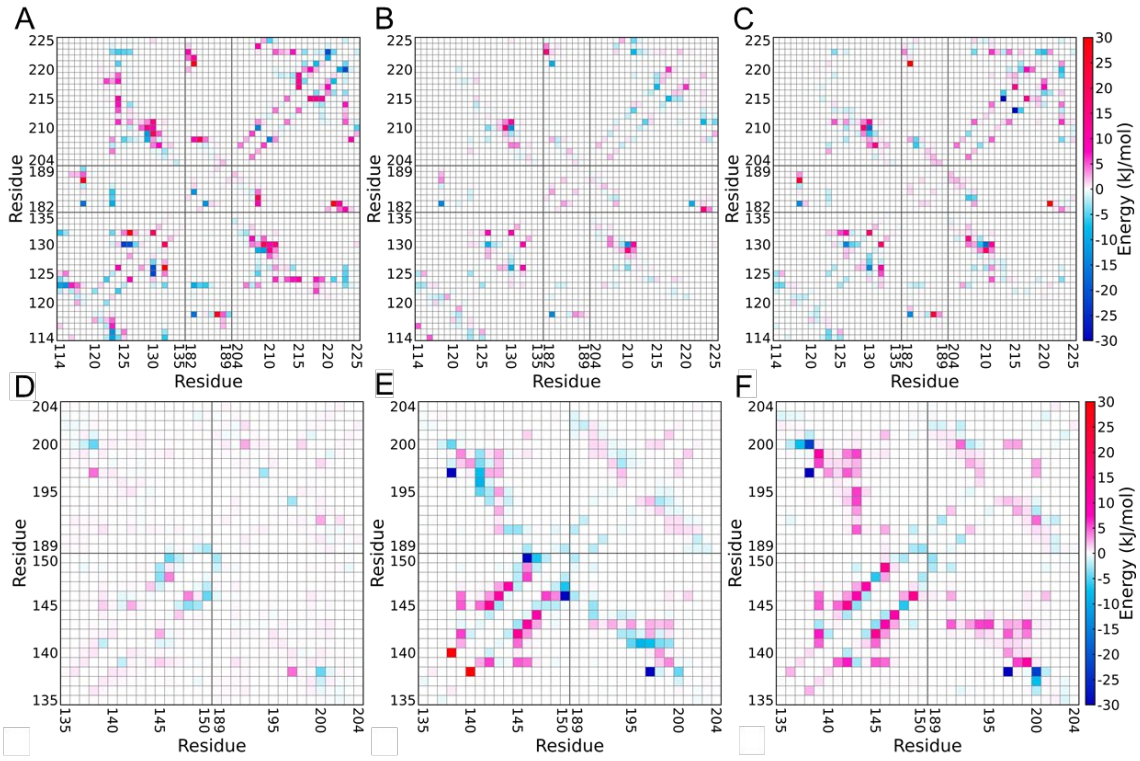


Figure 28. Difference average interaction energy maps for residue pairs in the ligand binding (top panels) and dimerization (bottom panels) surfaces of (A and D) holoBirA^{M211A}, (B and E) holoBirA^{P143A}, and (C and F) holoBirA$^{P143A/M211A}$.

7. Detailed structural and dynamic basis of allosteric modulation of BirA dimerization

Further analysis of simulations of apo- and holoBirA^{wt} indicates detailed structural differences on the dimerization surface that are consistent with experimentally determined structures. In crystal structures of both apo- and holoBirA^{wt}, an α-helix spans the width of the protein from the dimerization to the ligand binding surface (Figure 22).[153, 154] In the apo structure, this helix is comprised of residues 146-164, while in the holo structure, the helix is kinked at residue 146 and extended on its N-terminus to include residues 142-145. MD simulations yield results similar to these experimental observations (Figure 26A, B). In the median structure from the apoBirA^{wt} simulation, residues 148-166 are helical while residues 143-147 are not. By contrast, simulations of the holoBirA^{wt} species indicate a structure with a break in the helix at residues 146-149, preceded at the amino terminus by a short 3_{10}-helix that terminates at P143. Notably, the helical extension is correlated with disruption of the interaction between residues 146 and 150 that is detected in the apoBirA^{wt} energy map (Figure 26A). These structural differences observed in simulated BirA^{wt} structures indicate that the presence of a ligand is transmitted to the dimerization surface in a manner consistent with the differences observed in experimentally determined structures.

The detailed structure of each simulated holoBirA variant dimerization surface indicates variant-specific deviations from the structure of holoBirA^{wt}. Like holoBirA^{wt}, the M211A variant shows the 3_{10}-helical structure for residues 142-145 (Figure 29A, B). However, the structure of the BBL, which is known to function in dimerization,[208] in the M211A variant differs from its structure in holoBirA^{wt}. The analysis was carried

out by first aligning the central domain regions of the equilibrated structures of each holo variant with that of the holoBirA^{wt} median structure, followed by measurement of the RMSD for the BBL backbone atoms. The RMSD values of central domain regions of all variants, which range from 0.7 to 1.1 Å (Figure 30), indicate a high degree of structural similarity over the simulation trajectories. However, the BBL RMSD time course for the M211A variant deviates significantly from those of holoBirA^{wt} and the other two variants. Consistent with the large RMSD, in the median simulated structures, the holoBirA^{M211A} BBL conformation differs from that in holoBirA^{wt} (Figure 29C, D). In the P143A and P143A/M211A variants, the helical conformation of residues 142-145 is absent (Figure 29A,B) but the BBL conformation is identical to that found in the wild-type protein (Figure 29C).

Simulation results indicate that alanine substitutions on either the ligand binding or dimerization surface can alter the dynamics of the 140-146 and 193-199 loops. Principle component analysis of the motions associated with these loops indicates that 60% of the motion is captured in the major component (Figure 29E). In holoBirA^{wt} and holoBirA^{M211A}, the motion, as indicated by the root-mean-square fluctuation along the principle eigenvector, is relatively minor (Figure 29F). By contrast, markedly enhanced motion for the two loops is observed in both P143A and P143A/M211A (Figure 29F).

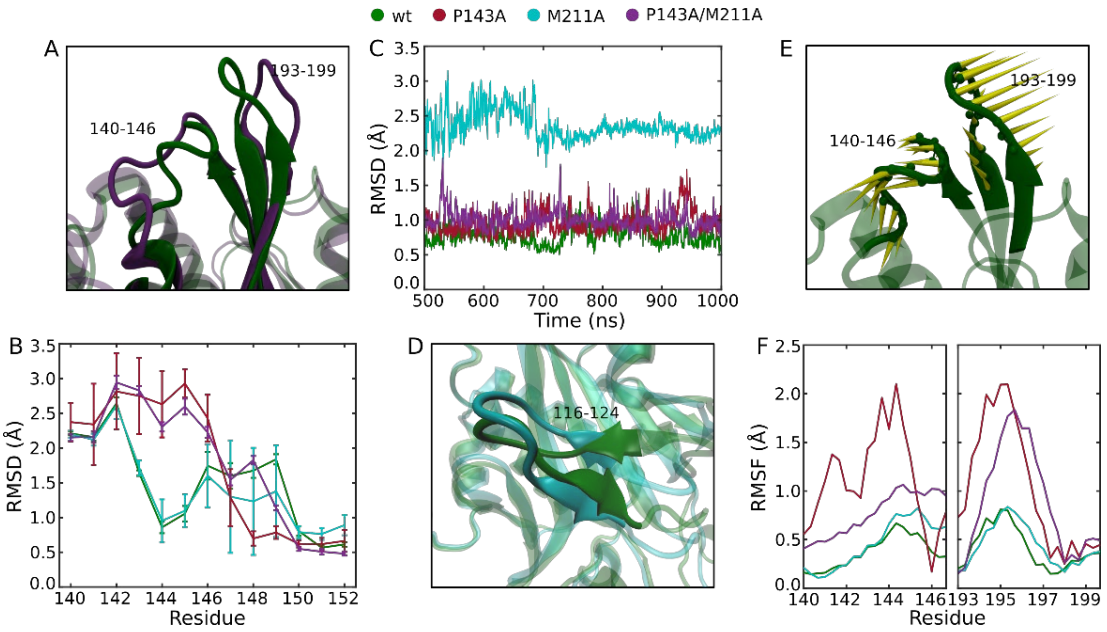


Figure 29. Alanine substitutions perturb the structure and dynamics of the dimerization surface loops. (A) Dimerization surface loops 140−146 and 193−199 for holoBirA^{wt} (green) and holoBirA$^{P143A/M211A}$ (purple). (B) Per-residue average RMSD from an ideal helix for residues 140−152. Error bars represent 95% confidence intervals. (C) RMSD time series for BBL residues 116−124 (C, CA, and N atoms). (D) Alignment of the BBL structures of holoBirA^{wt} (green) and holoBirA^{M211A} (cyan). (E) Dimer interface loops of the median holoBirAwt structure shown with the principal eigenvector of the covariance matrix for residues 140−146 and 193−199. (F) Root-mean-square fluctuation (RMSF) of residues 140−146 and 193− 199 (C, CA, and N atoms) along the principal eigenvector.

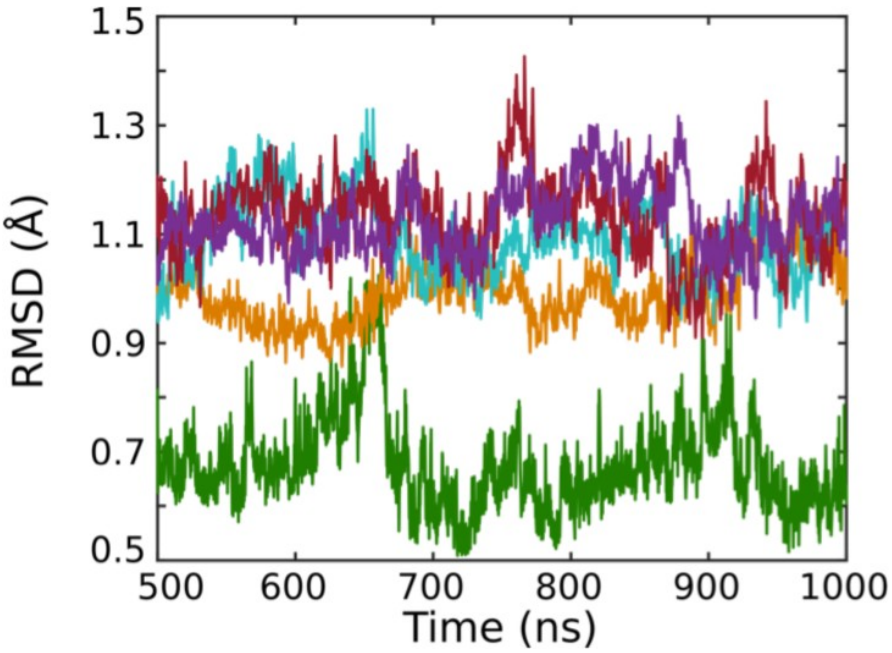


Figure 30. Root-mean-square deviation (RMSD) time series for C, CA, and N atoms of the central domain (residues 80-115, 127-139,147-192, 200-210, and 224-269) of BirA variants. Values smoothed using a sliding average with a 0.5 ns window. BirA variants shown are wt (green), R213A (orange), M211A (cyan), P143A (red), P143A/M211A (purple).

Loop packing on the ligand binding surface is correlated with the Gibbs free energy of bio-5'-AMP binding and is sensitive to amino acid substitution on the dimerization surface. The conformation of the adenylate binding loop (residues 211-222), which packs on the adenine ring of bio-5'-AMP in the holoBirA^{wt} structure (Figure 31A), was analyzed for BirA variant simulation trajectories. The analysis was carried out as described above for the BBL. The RMSD values of the loop segment are similar for holoBirA^{wt} and the P143A variant, both of which bind with similar affinities

to bio-5'-AMP (Figure 31B). By contrast, the loop in the M211A variant, which binds weakly to bio-5'-AMP, exhibits a large RMSD (Figure 31B). Notably, control simulations performed on the R213A variant, which, although it has a substitution in the ABL, binds to bio-5'-AMP with an affinity similar to that of $BirA^{wt}$,[172] indicate an RMSD time course similar to that of the wild-type protein (Figure 31B). Finally, addition of the P143A substitution on the dimerization surface to the ABL variant, $BirA^{M211A}$, results in a loop RMSD similar to that observed for $holoBirA^{wt}$ (Figure 31B).

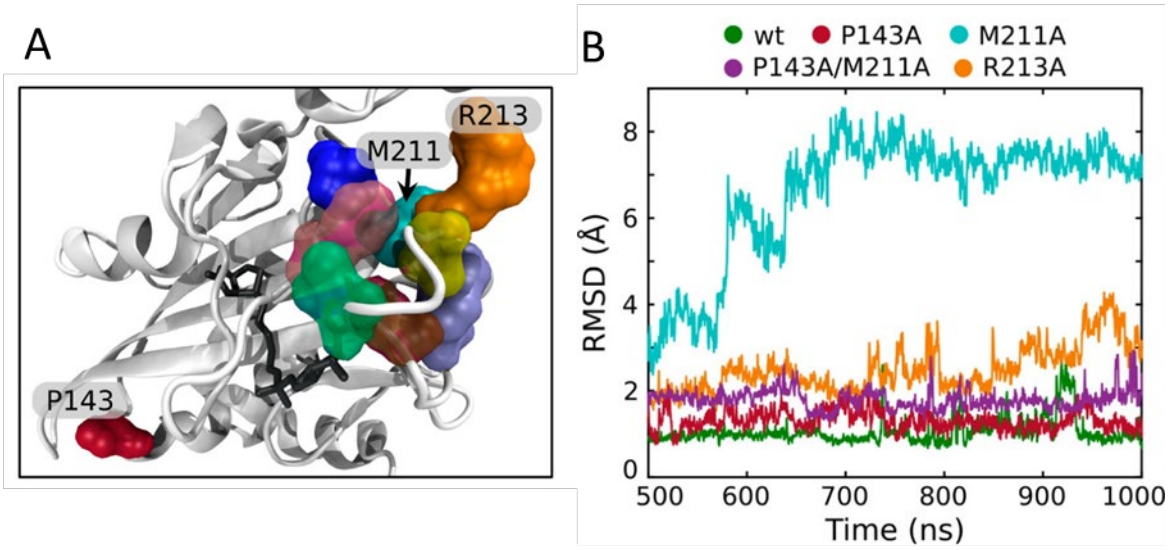


Figure 31. Loop packing at the ligand binding surface is influenced by substitutions on the dimerization surface. (A) HoloBirA central domain. (B) RMSD time series for C, CA, and N atoms of residues 211−223. Values were smoothed using a sliding average with a 0.5 ns window.

E. Discussion

The thermodynamics of allosteric systems are, as indicated in the cycle provided in Figure 22, bidirectional or reciprocal. For BirA, this means that because bio-5'-AMP binding enhances dimerization, dimerization also promotes bio-5'-AMP binding. Previous experimental studies were capable of detecting only unidirectional communication from the ligand binding site to the dimerization surface and vice versa.[171-174] The thermodynamic analysis of the BirA double-alanine variants presented in this work demonstrates long-distance reciprocal coupling between the two functional sites in the protein. For all doubly substituted variants, coupling is energetically favorable. These results do not rule out the existence of residue pairs that are coupled in both functions. For residue pairs that are coupled in effector binding, both substitutions are in the BirA central domain. By contrast, for the G281A/M211A pair that is coupled in dimerization, one substitution is in the central domain and the second in the C-terminal domain. Additional studies are required to determine if this relationship between functional coupling and BirA domain structure is a general property of the system.

The coupling of ligand binding and dimerization observed in the double-alanine variants indicates a context dependence of the functional effect of each single-alanine substitution or epistasis. Protein epistasis is ubiquitous and thought to be important for evolution.[209] The *E. coli* biotin repressor is a member of a large family of biotin ligase proteins that are widely distributed in archaea and eubacteria.[150] Although all of these proteins catalyze the transfer of biotin to biotin-dependent carboxylases, only a subset,

including the *E. coli* protein, also function as transcription repressors. It is likely that the inter-residue interactions detected in this work that function in the coupling of bio-5'-AMP binding to dimerization make up a small fraction of epistatic interactions that accompanied evolution of bifunctionality in this protein family. The results of molecular dynamics simulations indicate that coupling between residues P143 and M211 is exerted through modulation of the disorder-to-order transitions on both the dimerization and ligand binding surfaces. In simulations, introduction of the M211A substitution, which results in a large penalty on effector binding, severely perturbs adenylate binding loop folding and concomitant formation of the hydrophobic cluster. By contrast, introduction of the P143A substitution, which does not alter effector binding, has minor effects on the structure and interaction energetics of the ligand binding loops. Simultaneous alanine replacements at M211 and P143, which yield an effector binding free energy intermediate between those of the M211A and wild-type proteins, also partially reverse the ABL folding defect. Given that the Gibbs free energy of bio-5'-AMP binding for the double variant is still 1.8 kcal/mol less favorable than that of either $BirA^{wt}$ or $BirA^{P143A}$, the detailed structure of the double-variant ligand binding surface still deviates from that of the wild type. Thus, the P143A substitution on the dimerization surface communicates to the ligand binding surface to enhance loop packing and increase bio-5'-AMP binding affinity. A similar phenomenology is observed on the dimerization surface. In addition to its effect on ABL folding, the M211A substitution is accompanied by an alteration in the BBL structure. As the BBL is known to function in both bio-5'-AMP binding and dimerization,[189, 208] the altered

conformation in the M211A variant may be responsible for its weaker dimerization energetics. The P143A substitution results in significant structural and energetic perturbations to the 140-146 loop on that surface. Addition of the M211A substitution on the ligand binding surface, which results in even weaker dimerization, yields a dimerization surface in which the internal structures of the 140-146 and 193-199 loops are significantly perturbed as well as the interloop interaction. These same changes in the double-alanine variant are accompanied by a reversal of the effect of the single M211A substitution on the BBL conformation.

Although the ligand binding and dimerization loops show some of the largest structural and energetic variations, other residues in BirA central and C-terminal domain loops exhibit deviations from their positions in the wild-type protein. One possibility is that these deviations simply reflect the intrinsic loop flexibility. Alternatively, these residues may serve as additional sites of allosteric communication in the protein. For example, the backbone RMSD for residue 166 is relatively large (3−4 Å) for all variants. This residue is at the C-terminus of the α-helix comprised of residues 146-164 in apoBirA that is extended to residue P143 at its N-terminus in the holoprotein,[154] a structural feature that is absent in the weakly dimerizing proteins $BirA^{P143A}$ and $BirA^{P143A/M211A}$. In the model derived from X-ray crystallography, the distance between the A166 side chain and that of residue A229, located in the ABL, is 3.5 Å. Moreover, the pairwise energy maps obtained from MD simulations indicate that this interaction is energetically compromised, albeit by a modest 1−2 kJ/mol, in all three variants. In addition to residue 166, residues in the C-terminal domain loop

positions show structural differences in the variants. Results of X-ray crystallography indicate that the C-terminal domain is rotated 3° relative to the central domain in holoBirA relative to apoBirA, which brings the C-terminal domains of the two monomers in the dimer closer together.[154] Consistent with these structural results, alanine substitutions in loops containing residues 280-283 and 306-316 are known to alter dimerization energetics.[172, 176] The structural perturbations in the loops observed in the variants suggest the existence of communication between residues in the ligand binding and dimerization surfaces with the C-terminus.

The combined results of experiments and simulations reveal that allosteric communication in BirA occurs through reciprocal modulation of the disorder-to-order transitions on two distant functional surfaces. The simulation results obtained with the P143A/M211A pair predict that structural and dynamic correlates to the functional communication between these two residues are primarily localized to the ligand binding and dimerization surfaces. However, the data also provide limited evidence of the distribution of the allosteric response to residues that are far removed from the coupled sites. Additional experimental and computational studies are required to obtain a more comprehensive picture of the sites of allosteric communication in BirA and to define structural networks in the protein core that connects these sites.[95, 97, 210]

Chapter 3: Integration of distant disorder-to-order transitions with a residue network in allostery

A. Contribution statement

The work presented in this chapter was completed in collaboration with Dr. Silvina Matysiak's group at the Fischell Department of Bioengineering, University of Maryland, College Park. All molecular dynamics simulations were conducted by Dr. Gregory Custer, and the network analysis was performed by Christopher Look and Riya Samanta. Riya Samanta also carried out Force Distribution Analysis on BirA wild-type and variants. Our lab performed to all non-computational experiments such as protein purification and functional measurements. We also contributed to the identification of residue networks described in this work and facilitated the analysis of simulation data.

B. Introduction

Allostery, or energetic coupling between events that occur at distinct sites in a protein, is a widespread phenomenon. As "the second secret of life",[211] it is utilized in virtually all biological processes including metabolism,[212] cell signaling,[1] and transcription regulation[213]. Consequently, its mechanism remains the subject of intense research. Although recent studies highlight the importance of disorder for protein allostery,[44, 85, 214] the physico-chemical basis of its contribution to long-range energetic coupling remains to be elucidated.

The *E. coli* biotin protein ligase (BirA) provides an ideal model system for determining how disorder-to-order transitions are communicated in allostery. BirA is a

transcription repressor that is allosterically activated for homodimerization *via* biotinyl-5'-adenylate (bio-5'-AMP) binding.[183] The resulting holoBirA homodimer binds to DNA to repress transcription of the biotin biosynthetic operon.[157, 165] Structural studies reveal that bio-5'-AMP binding is accompanied by disorder-to-order transitions on the coupled ligand binding and dimerization surfaces, which are separated by more than 30Å (Figure 32).[153, 154, 169] The adenylate binding (ABL) and biotin binding loops (BBL) on the ligand binding surface, which are disordered in apoBirA, fold around bio-5'-AMP in holoBirA. On the dimerization surface effector binding is coupled to the extension of an α-helix and ordering/packing of two loop segments. Functional measurements performed on BirA variants with single and multiple amino acid substitutions indicate that the disorder-to-order transitions on each functional surface are critical for BirA allostery[171-173] and that the distant folding processes are coupled.[170, 174] The molecular mechanism of this coupling has yet to be determined.

Comparison of apo and holoBirA structures reveals that the disorder-to-order transitions linked to effector binding are accompanied by formation of an electrostatic network.[176] In the holoBirA structure the network incorporates the effector molecule and several charged/polar amino acid side chains. Only a subset of these interactions can form in apoBirA because the BBL, which contributes residues R118, R119, and R121 to the network, is disordered in the absence of ligand.[153, 169] Despite its presence in the holoBirA structure, the significance of the network in solution and its contribution to allostery are not known.

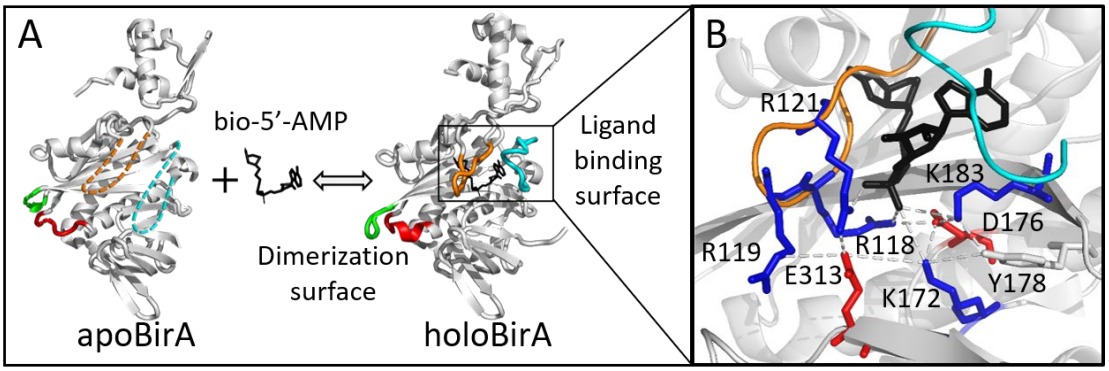


Figure 32. Structural analysis of disorder-to-order and network formation upon bio-5'-AMP binding to BirA. (A) BirA structures with loop disorder-to-order transitions highlighted: 140-146 (Red), 193-199 (Green), BBL (Orange), and ABL (Cyan). The dashed segments indicate disorder. (B) Network interactions in holoBirA with the following side chain color codes: Red, negatively charged; Blue, positively charged; Black, ligand. Models were created in Pymol with 1BIA[153] and 2EWN[154] as input.

Results of computational analysis predict a role for the electrostatic network in allostery. Simulations reveal that an alanine substitution that perturbs both the disorder-to-order transition on the ligand binding surface and bio-5'-AMP-linked dimerization, also disrupts the network.[176] Additionally, analysis of the network in simulated BirA variant structures predicts the correlation of its rearrangement with altered bio-5'-AMP binding-linked dimerization.[176]

In this work, we applied computational and experimental methods to investigate the nature of the electrostatic network, its role in BirA allostery and its relationship to the effector-linked disorder-to-order transitions. Computational analysis reveals an extensive residue network that includes a majority of the residues identified from

structural analysis. Measurements of proton linkage to effector binding support network formation in solution. Results of ITC and sedimentation equilibrium measurements reveal that disruption of the network by alanine substitutions significantly perturbs both allosteric effector binding and effector-linked dimerization. Finally, Force Distribution Analysis reveals the linkage between the network and disorder-to-order transitions on the two functional surfaces. These results support a mechanism in which a residue network and disorder-to-order transitions function in concert to enable long-distance energetic coupling in protein allostery.

C. Materials and Methods

3.1 Chemicals and biochemicals

All chemicals and biochemicals were at least reagent grade. The d-biotin (Sigma-Aldrich) stock solutions were prepared in Standard Buffer [10 mM Tris (pH 7.50 ± 0.02 at 20 °C), 200 mM KCl, and 2.5 mM $MgCl_2$] and stored in -80°C. The biotinoyl-5'-adenylate (bio-5'-AMP) was synthesized and purified as previously described [157, 215] and the bio-5'-AMP analog, biotinol-5'-adenylate (btnOH-AMP), was purchased from RNA-Tech (Leuven, Belgium). The btnOH-AMP and bio-5'-AMP stock solutions, which were stored at -80°C, were prepared by dissolving the dry powder into Milli-Q water. Concentrations were determined by absorption spectroscopy at 259 nm using a molar extinction coefficient of 15400 $M^{-1}cm^{-1}$.

3.2 Mutagenesis, expression, and purification of BirA variants

Mutations in the BirA coding sequence were generated by oligonucleotide-directed mutagenesis using a pBtac2 (Boehringer Mannheim) plasmid derivative that carries the C-terminally $(His)_6$-tagged $BirA^{wt}$ coding sequence. PCR reactions were performed using either Pfu Ultra (Agilent) or KOD (Millipore-Sigma) DNA polymerase and mutations were verified by sequencing the entire coding sequence (ACGT Inc.).

Each BirA variant was expressed in *Escherichia coli* strain HMS174/pMS421 transformed with the appropriate pBtac2-BirA-His derivative plasmid.[189] The variant proteins were purified as previously reported,[171] with the exception of the introduction

of a Q-sepharose chromatography step prior to final chromatography on SP-sepharose. Protein concentrations were determined by absorption spectroscopy at 280 nm using a molar extinction coefficient of 47510 $M^{-1}cm^{-1}$ calculated from the amino acid composition.[190] Protein purity was estimated at >95% based on SDS-PAGE analysis.[216]

3.3 Circular dichroism (CD) spectroscopy

The CD spectra were acquired by a JASCO J-810 (JASCO) spectrophotometer that is equipped with a temperature controller unit. All spectra were collected using a 2mm pathlength quartz cuvette at 20 °C. Data were recorded at 1nm intervals from 260 to 200nm at a scan speed of 50nm/min and a bandwidth of 1nm. To minimize the signal background from salt, spectra for the majority of the BirA variants were acquired in low salt Standard Buffer (10 mM Tris, 50 mM KCl, 2.5 mM $MgCl_2$, pH = 7.5 at 20 °C). Due to its limited solubility in 50 mM KCl, spectra for the $BirA^{E313A}$ variant were acquired Standard Buffer containing 200 mM KCl. Each final spectrum was the average of three measurements.

3.4 Isothermal titration calorimetry (ITC)

All titrations were carried out using a VP-ITC calorimeter (Malvern) equipped with a 1.44mL cell. Proteins were prepared for titration by exhaustive dialysis against the binding buffer, removal of any resulting precipitate by filtration through a 0.22 μm PES syringe filter (SIMSII), and concentration determination by UV absorbance at 280 nm. Bio-5'-AMP binding measurements were carried out in Standard Buffer and proton linkage measurements were performed in Standard Buffer in which Tris was replaced

by 10 mM Bistris, Citrate, or MES. Ligand stock and concentrated protein were diluted into dialysis buffer to the working concentrations immediately before titrations.

Titrations designed to obtain only molar binding enthalpies were carried out under conditions of total association at partial saturation (TAPS),[217] in which the ligand quantitatively binds to the protein in the first few injections. For each measurement, the sample cell contained a BirA variant at a concentration of 2-5 μM and the injection syringe was filled with a 20-50 μM ligand solution. A total of 14 injections were made including an initial 2 μL injection, followed by 6×13μL injections, each of which provides the heat of ligand binding plus ligand dilution. After saturating the protein with a 120 μL ligand injection, 6×13 μL injections were performed to obtain the ligand dilution heat. The net ligand binding heat was obtained by subtracting the ligand dilution heat from the heat of the six initial ligand injections.

Equilibrium binding titrations were carried out using either the direct or displacement method. In direct titrations a bio-5'-AMP solution at a concentration of 5-20 μM was titrated into the sample cell containing the BirA variant at 0.5-2 μM. The titrations were initiated with a 2 μL injection, followed by another 17×16 μL injections.[7] In displacement (competitive) titrations one 2 μL plus 22×13 μL volumes of a bio-5'-AMP solution were injected into a biotin-saturated BirA solution. The reported binding parameters were based on at least two independent titrations, with the majority representing at least three.

3.5 Sedimentation equilibrium measurements

HoloBirA dimerization was measured by sedimentation equilibrium in Standard Buffer at 20 °C using an Optima XL-I analytical ultracentrifuge equipped with a four-hole An-60 Ti rotor (Beckman Coulter). In order to ensure saturation, bio-5'-AMP was added to protein at a 1.5:1 molar ratio under stoichiometric conditions. Samples containing 60, 50, 40 μM holoBirA monomer were loaded into standard 12 mm six channel cells and centrifuged at 18000, 21000, 24000 rpm, respectively. After 8 and 9 hours of centrifugation at each speed absorbance scans (step size=0.001 cm, 5 averages) were acquired at 300 nm. Overlays of the two scans indicated that the system had reached equilibrium. At least two independent measurements of the equilibrium dimerization constant were performed for each variant.

3.6 Data analysis

ITC data were analyzed using the Microcal software suite in Origin 7.0. The heat of ligand dilution, which was calculated from the average of the heats of the final 3-5 injections, was subtracted from each raw injection heat. Injection heats were then normalized to molar enthalpy and the resulting isotherm was analyzed using a single-site binding model for direct titrations and the competitive binding model for displacement titrations.

Sedimentation equilibrium data were analyzed using Nonlin[218] in Heteroanalysis version 1.1.0.58 (https://core.uconn.edu/resources/biophysics#au-software). The equilibrium dissociation constant (K_D) for dimerization was obtained by global nonlinear least squares analysis of nine data sets using a monomer-dimer model:

$$c(r) = c(r_o)e^{\frac{\sigma_{mon}(r^2-r_o^2)}{2}} + \frac{1}{K_D}[c(r_o)]^2 e^{\frac{2\sigma_{mon}(r^2-r_o^2)}{2}} + \delta \quad (1)$$

where c(r) is the protein concentration at position r, $c(r_o)$ is the protein concentration at reference radial position r_o, K_A is the equilibrium association constant governing dimerization, δ is the baseline offset, σ_m is the reduced molecular weight for the BirA monomer calculated using the equation:

$$\sigma_{mon} = \frac{M(1-\bar{\nu}\rho)}{RT}\omega^2 \quad (2)$$

where M is His-tagged BirA monomer molecular weight of 36100 g/mol, $\bar{\nu}$ is the partial specific volume which was experimentally determined as 0.755 cm^3/g,[183] ρ is the buffer density calculated using Sednterp Version 1.09, ω is the angular velocity of the rotor, RT is the gas constant × temperature (K).

3.7 Molecular dynamics (MD) simulations and analysis

All MD simulations described in this chapter were conducted by Dr. Silvina Matysiak's group at the Fischell Department of Bioengineering, University of Maryland, College Park. Simulations were performed on BirA, both wild-type and variant, in complex with the corepressor analog biotinol-5'-AMP (btnOH-AMP), with chain A of the BirA dimer structure in Protein Data Bank (PDB) entry 2EWN[154] used as the starting configuration of the complex. The simulation trajectories used in this study were taken from our previous publications.[170] The GROMACS 4.6 simulator [219-221] with the OPLS-AA force field[222] was used for simulation. For each simulation, the protein was placed in a rhombic dodecahedral box with boundaries extending out ~1 nm from the protein. The system was then solvated with ~20300 SPC/E water

molecules.[223] Random replacement of a water molecule with a Na^+ counterion rendered the system neutral. The duration of the production run, using an NPT ensemble at 300 K and 1 atm, was 1 μs, with the final 500 ns used for analysis.

The shortest communication paths between BirA residues were calculated using a method based on Ribeiro, *et al*. 2014.[96] To calculate shortest paths, a network between BirA residues was first constructed based on the MD simulation, with the final 500 ns of the simulation trajectory used for analysis. Each protein residue is considered a single node in the network. Connections between a pair of nodes (residues) by an edge required contact for at least 20% of the simulation time analyzed. A distance-based cutoff was used to identify contacts between residues, with distances between all pairs of atoms in the residues considered. The distance cutoff for each pair of atoms was 1.7 times the sum of their van der Waals radii. For each residue, i, edges with residues i±1 and i±2 were excluded. The weight (or length), ω_{ij}, of each edge in the network was calculated using the pairwise interaction energy, ε_{ij} between residues i and j connected by that edge. This pairwise interaction energy was defined as the sum of all non-bonded interactions between the residues and was calculated at each time step. The weight, ω_{ij}, was calculated from ε_{ij} as follows:

$$\omega_{ij} = \varepsilon_{ij}^{-1/3} \tag{3}$$

The shortest pathways between all residue pairs, excluding pairs (i, i±1) and (i, i±2), were calculated after every 80 ps using the Floyd-Warshal algorithm,[224] providing an ensemble of shortest paths between all residue pairs. In order to identify residues most likely to be a part of the network, the final 500 ns of the entire trajectory is first

divided into 50 sets. The likelihood of appearance of a particular residue in the shortest path is calculated for the set and normalized to the residue with the highest number of shortest paths. This likelihood was generated for each set and finally averaged across all 50 sets.

Force Distribution Analysis, based on the method outlined in Stacklies, *et al* 2011[225] and Costescu and Grater 2013,[226] was carried out for wild type and variant holoBirA species. Residue pairwise forces, which are based on bonded and non-bonded interactions, were used to calculate the Punctual Stress using the Time-Resolved Force Distribution Analysis (TRFDA) code. These calculations were carried out on reruns of the final 500 ns of each MD trajectory. The Punctual Stress for each residue, i, is the sum of the absolute values of pairwise forces applied by all other residues, j, on that residue.

$$S_i = \sum_{j,i\neq j} |F_{ij}| \quad (4)$$

This calculation generates the Punctual Stress for each residue in the 317 residues that were modeled in the structure for every frame and then averaged across all frames. The averaged values thus obtained were used to compute the absolute difference between punctual stress of variant and wt.

D. Results

1. Computational analysis reveals an extensive residue network in holoBirA

The electrostatic network previously identified in holoBirA^{wt} was based on inspection of the structure, a snapshot of the protein, determined by X-ray crystallography. Computational analysis of all-atom MD equilibrium trajectories was used to obtain a more comprehensive picture of residue networks that may function in BirA allostery. The analysis of holoBirA^{wt} was carried out, as described in Materials & Methods, using a method based on inter-residue interaction energies.[96, 97]

The computational analysis by Dr. Matysiak's group reveals a more extensive network than that obtained from structural analysis. The 317 residues analyzed were ranked according to the magnitudes of their "node betweenness" and those with values in the top 5%, which included 16 total residues, were designated as having a high likelihood of network participation. Of the 16, 6 belong to the network derived from inspection of the structure (Figure 33A). Only residues R119 and R121 of the original network are not included among the high likelihood nodes. The effector, bio-5'-AMP, which was identified from the structure as a network participant, is also a high likelihood node. Included among the computationally identified nodes are residues P126, R235, and W223, which have previously been shown to function either in BirA allostery *in vitro*[171, 172] or in BirA-mediated transcription repression *in vivo*[152, 167, 227]. All three of these residues are in segments that undergo disorder-to-order transitions upon bio-5'-AMP binding. The majority of the residue network in the holoBirA monomer defines a continuous surface on one face of the protein from the N-terminal-

central domain interface to the interface formed between the central and C-terminal domain (Figure 33B).

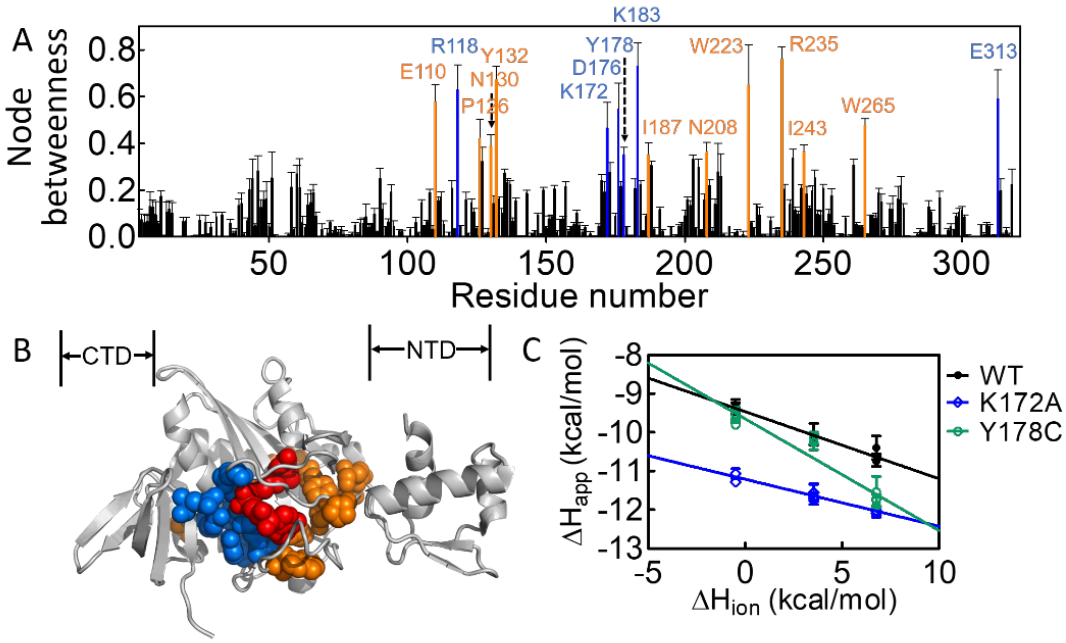


Figure 33. Computationally determined residue network in holoBirA. (A) Node betweenness values for each residue with error bars representing one standard deviation of 50 sets of averaged probabilities (B) Positions of high likelihood network residues on the holoBirA structure. Color code (A & B): Red, btnOH-AMP ligand; Blue, network residues in both the structure and computation; Orange, residues identified solely from computation. (C) Proton release is linked to effector binding. Solid lines result from linear regression using Equation 5 with error bars representing the 67% confidence interval of each individual measurement.

2. The network forms in solution

Although the structural and computational analyses are consistent with network formation upon bio-5'-AMP binding, experimental support of its solution relevance is lacking. Formation of electrostatic interactions upon ligand binding is likely to be linked to proton release from or uptake by the protein and Isothermal Titration Calorimetry (ITC) provides an ideal method to detect this linkage.[228] For example, if ligand binding is linked to proton release, the measured heat signal in ITC reflects the sum of the ligand binding heat and the heat of buffer ionization. Measurement of the apparent heat of binding in buffers characterized by a range of ionization enthalpies yields data that adhere to the following relationship:

$$\Delta H^{o}_{app} = n_{H^+}\Delta H^{o}_{ion} + \Delta H_{int} \quad (5)$$

where $\Delta H_{app}{}^{o}$ is the measured molar heat of binding, ΔH_{int} is the intrinsic binding enthalpy, and $\Delta H_{ion}{}^{o}$ is the buffer ionization enthalpy. Linear regression of the dependence of $\Delta H_{app}{}^{o}$ on $\Delta H_{ion}{}^{o}$ yields n_{H+}, the number of protons released from or absorbed by the protein upon ligand binding.

ITC measurements, which were performed using the TAPS method, indicate effector binding to BirA is accompanied by proton release. Measurements were carried out at pH 6.0 to enhance protonation of ionizable groups on the protein. Additionally, the contribution of the BirA dimerization heat to the measured enthalpy for the wild type protein was minimized by using the relatively weak allosteric activator, btnOH-AMP,[229] for the measurements. Consistent with network formation in solution, btnOH-AMP binding to $BirA^{wt}$ is linked to proton release (Figure 33C, Table 7). Moreover,

amino acid substitutions at residues K172 or Y178, both of which are predicted to contribute to the network, alter the proton linkage (Figure 33C, Table 7). These results support the linkage of effector binding to network formation in solution.

Table 7. Proton release linked to effector binding to BirA

Buffering agent	ΔH^{o}_{app} (kcal/mol)		
	wt	K172A	Y178C
Bistris	-10.6 ± 0.2	-12.1 ± 0.1	-11.8 ± 0.1
MES	-10.1 ± 0.2	-11.6 ± 0.1	-10.2 ± 0.1
Citrate	-9.4 ± 0.1	-11.2 ± 0.1	-9.7 ± 0.2
n_{H+}	-0.166 ± 0.005	-0.12 ± 0.02	-0.35 ± 0.07

Measurements were performed at 20 °C, pH 6.0 in Standard Buffer containing the appropriate buffering agent at 10 mM concentration. The reported errors were propagated from results of at least two independent TAPS measurements. The number of protons released was obtained from linear regression of the data using Equation 5.

3. The network residues function in effector binding

Allosteric effector binding activates the BirA monomer for homodimerization. The contributions of the network residues in this binding/activation process were assessed by measuring bio-5'-AMP binding to eight BirA variants with substitutions at network positions using ITC. Circular dichroism spectra indicated that the folded conformations of all eight variant proteins were identical to that of $BirA^{wt}$ (Figure 34).

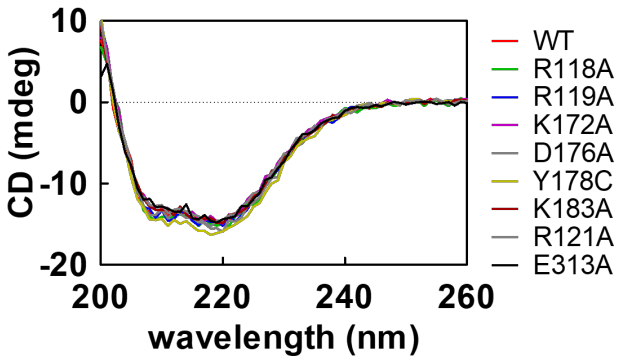


Figure 34. CD Spectra for all variants are similar to that of BirA^{wt}. Spectra are the average of three measurements. Spectra were acquired at 20 °C in Standard Buffer containing 50 mM KCl for all variants except for that of BirA^{E313A} which was acquired in 200 mM KCl.

For variants with alanine substitutions at R118, K183 and E313 both the equilibrium constants and the binding enthalpies were obtained from standard titrations, in which bio-5'-AMP is injected into a solution containing apoBirA. Representative data obtained for BirA^{R118A} reveal that the binding is well described by a single site model and that the binding free energy is 2.2 ± 0.2 kcal/mol less favorable than that measured for BirA^{wt} (Figure 35A, Table 8). Bio-5'-AMP binding to BirA^{wt}, BirA^{R121A} and BirA^{Y178C} occurs in the picomolar concentration range and, consequently, the displacement method, in which addition of bio-5'-AMP competes biotin off the protein, was used for these two variants [191, 207] (Figure 36, Table 8). Although bio-5'-AMP also binds very tightly to variants BirA^{K172A} and BirA^{D176A}, the heat signals in displacement titrations were too small to obtain reliable data. Consequently, the direct titration method was used for these variants (Figure 36). Since

the c-values, the product of the equilibrium constant and the protein concentration, employed for these titrations were large (>1000), the reported equilibrium constants should be considered upper limits. Accurate bio-5'-AMP binding enthalpies for all of the tight-binding variants were obtained using the TAPS method.

Results of the binding measurements reveal that electrostatic network residues contribute significantly to bio-5'-AMP binding. Alanine substitutions of residues R118, R119, K183, and E313 have impacts of +2 to +4 kcal/mol on the binding free energy. Even for variants that bind bio-5'-AMP with free energies similar to that measured for $BirA^{wt}$, the binding enthalpy and/or the entropy is altered (Figure 35B, Table 8).

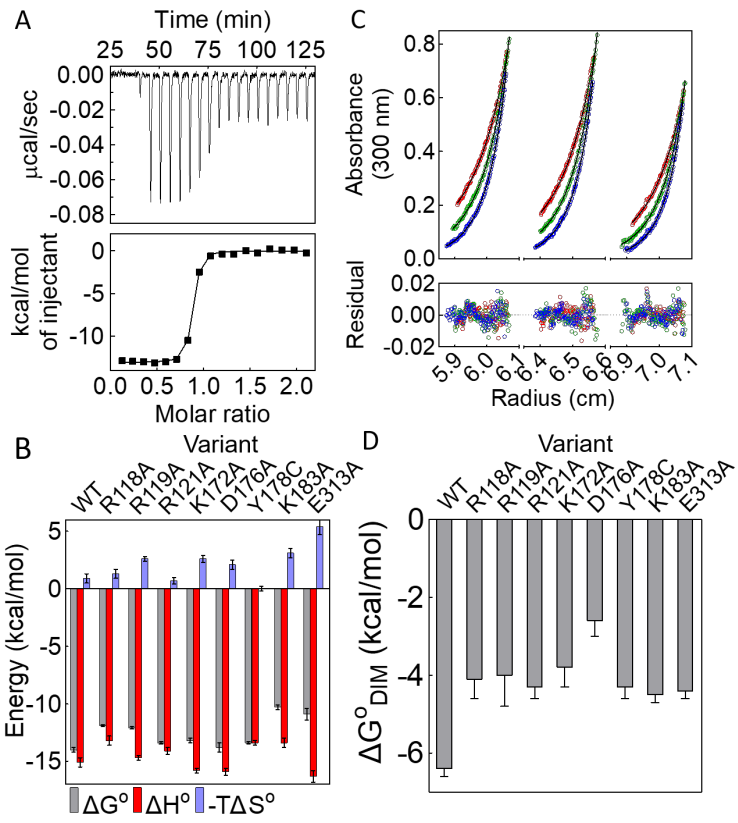


Figure 35. Network residues function in bio-5'-AMP binding and holoBirA dimerization. (A) Titration of $BirA^{R118A}$ with bio-5'-AMP (top) with analysis of the resulting data using a single-site binding model (bottom). (B) Histograms showing thermodynamics of bio-5'-AMP binding, with error bars representing one standard deviation or propagated error calculated from at least two independent measurements. (C) $HoloBirA^{R121A}$ sedimentation equilibrium measurement performed at 18k (red), 21k (green), 24k (blue) rpm with protein samples prepared at 60 µM (left), 50 µM (middle), 40 µM (right) protein concentrations. Top: Absorbance *vs* radius profiles with best-fits to a monomer-dimer model shown as solid lines. Bottom: Residuals of the fit. (D) Dimerization free energies for network variants obtained from at least two independent measurements with error bars representing the 67% confidence intervals.

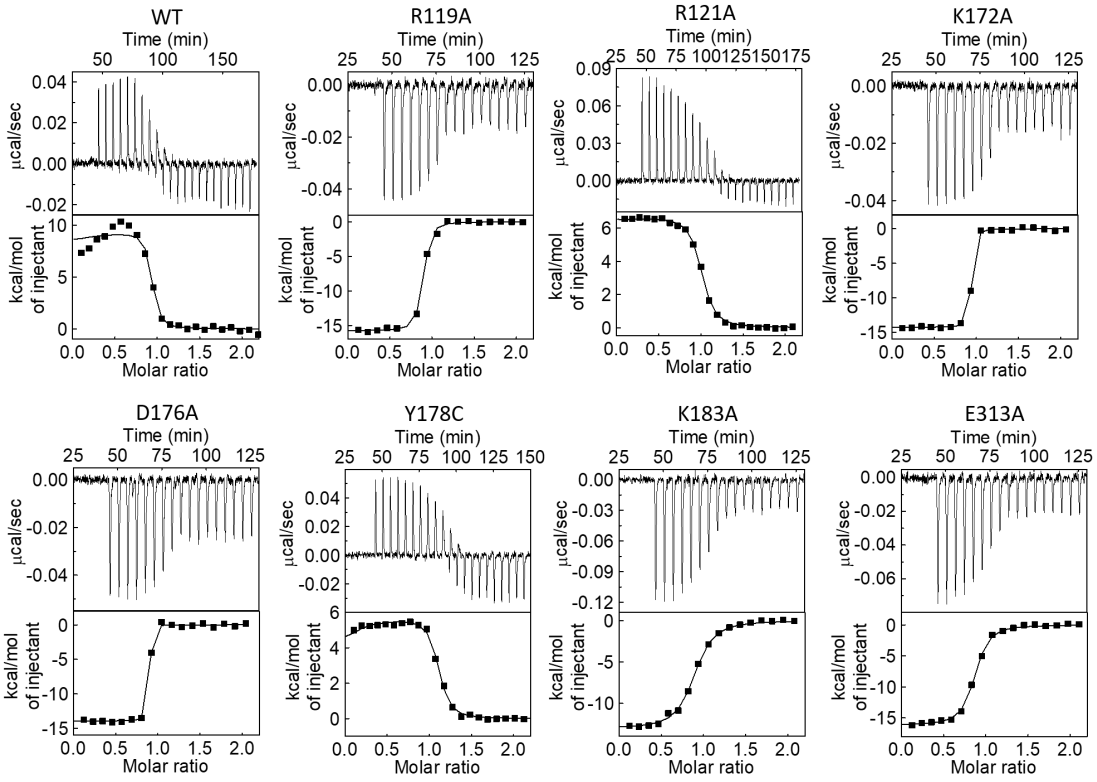


Figure 36. ITC measurements of bio-5'-AMP binding to BirA variants. For each variant the top panel shows the baseline adjusted thermogram and the bottom panel shows the binding isotherm and best-fit curves to a single-site or competitive binding model.

Table 8. Bio-5'-AMP binding thermodynamics of BirA variants

Protein	K_D(M)	ΔG^o (kcal/mol)	ΔH^o (kcal/mol)	$-T\Delta S^o$ (kcal/mol)	n
wt[a,b]	$4(\pm1)\times10^{-11}$	-14.0±0.2	-15.1±0.4	0.9±0.4	0.88±0.02
R118A[c]	$1.3(\pm0.3)\times10^{-9}$	-11.9±0.1	-13.2±0.4	1.3±0.4	0.88±0.06
R119A[b]	$9(\pm2)\times10^{-10}$	-12.1±0.1	-14.7±0.2	2.6±0.2	0.83±0.02
R121A[a,b]	$1.0(\pm0.2)\times10^{-10}$	-13.4±0.1	-14.1±0.3	0.7±0.3	0.94±0.05
K172A[b]	$1.5(\pm0.4)\times10^{-10}$	-13.2±0.2	-15.8±0.2	2.6±0.3	0.90±0.01
D176A[b]	$5(\pm6)\times10^{-11}$	-13.8±0.4	-15.9±0.3	2.1±0.4	0.86±0.02
Y178C[a,b]	$1.0(\pm0.2)\times10^{-10}$	-13.4±0.1	-13.4±0.2	0±0.2	1.0±0.1
K183A[c]	$2.2(\pm0.6)\times10^{-8}$	-10.3±0.2	-13.4±0.4	3.1±0.4	0.83±0.02
E313A[c]	$1(\pm0.9)\times10^{-8}$	-10.9±0.5	-16.3±0.5	5.4±0.7	0.79±0.03

Measurements were carried out in Standard Buffer at 20 °C. The reported errors represent the larger of either the standard deviation or propagated error obtained from at least two independent measurements. The Gibbs free energies and entropic contributions were calculated using equations $\Delta G^o = RT\ln K_D$ and $\Delta G^o = \Delta H^o - T\Delta S^o$, respectively. [a]Equilibrium constants obtained from displacement titrations. [b&c]Binding enthalpies obtained from [b]TAPS or [c]quilibrium titrations.

4. Function of the network in BirA allosteric output, holoBirA dimerization

The output of bio-5'-AMP-linked allosteric activation of BirA is a 1000-fold enhancement of its homo-dimerization[166] and the contributions of network residues to this output were assessed by measuring the dimerization free energies of the variants in their holo (bio-5'-AMP bound) forms using sedimentation equilibrium. For each variant the measurements were performed at three speeds on protein samples prepared at three concentrations and the resulting 9 data sets were globally analyzed using a monomer-dimer model. Results obtained for holoBirA^{R121A}, which indicate excellent agreement between the data and the model, yield an equilibrium dimerization constant of $7(\pm2)\times10^{-4}$ M, 40-fold weaker than that measured for holoBirA^{wt} (Figure 35C). Notably, with this weak dimerization the maximal fraction dimer at highest holoBirA concentration is relatively low, at 16.4% of the total monomer concentration. Nevertheless, analysis of the data using a single-species model yielded an average molecular weight of 42 ± 1 kDa, higher than the 36 kDa expected for the monomer. Moreover, the square root of the variance value associated with a monomer-dimer fit was smaller than that obtained using a single-species model. Results of sedimentation equilibrium measurements performed on all variants with network residue substitutions indicate significantly weaker dimerization than that measured for holoBirA^{wt}, with penalties to dimerization free energy ranging from +2 to +4 kcal/mol (Figure 35D and 37, Table 9).

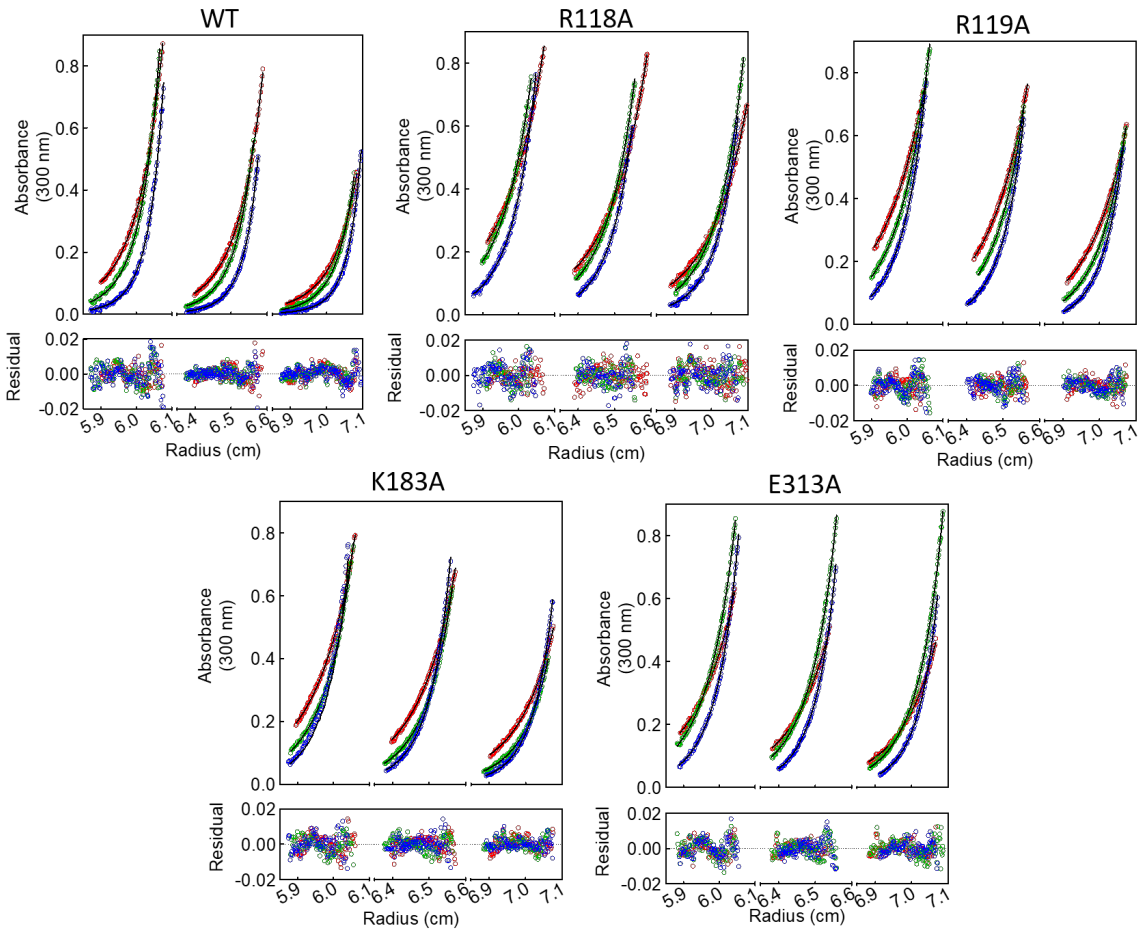


Figure 37. Absorbance versus radius profiles for holoBirA variants prepared at three loading concentrations (left> middle> right) and centrifuged at three rotor speeds. (red< green<blue). For each variant the top panel shows the data and best fit curves to a monomer-dimer association model and the bottom panel shows the residuals of the fit.

Table 9. Dimerization energetics of BirA variants

Variant	K_D (M)	ΔG^o (kcal/mol)
wt	$1.6(\pm0.5)\times10^{-5}$	-6.4±0.2
R118A	$1.0(\pm0.6)\times10^{-3}$	-4.1±0.5
R119A	$1.1(\pm0.8)\times10^{-3}$	-4.0±0.8
R121A	$6(\pm2)\times10^{-4}$	-4.3±0.3
K172A[a]	$1.4(\pm0.8)\times10^{-3}$	-3.8±0.5
D176A[a]	$1(\pm1)\times10^{-2}$	-2.6±0.4
Y178C[b]	$7(\pm3)\times10^{-4}$	-4.3±0.3
K183A	$4(\pm1)\times10^{-4}$	-4.5±0.2
E313A	$5(\pm2)\times10^{-4}$	-4.4±0.2

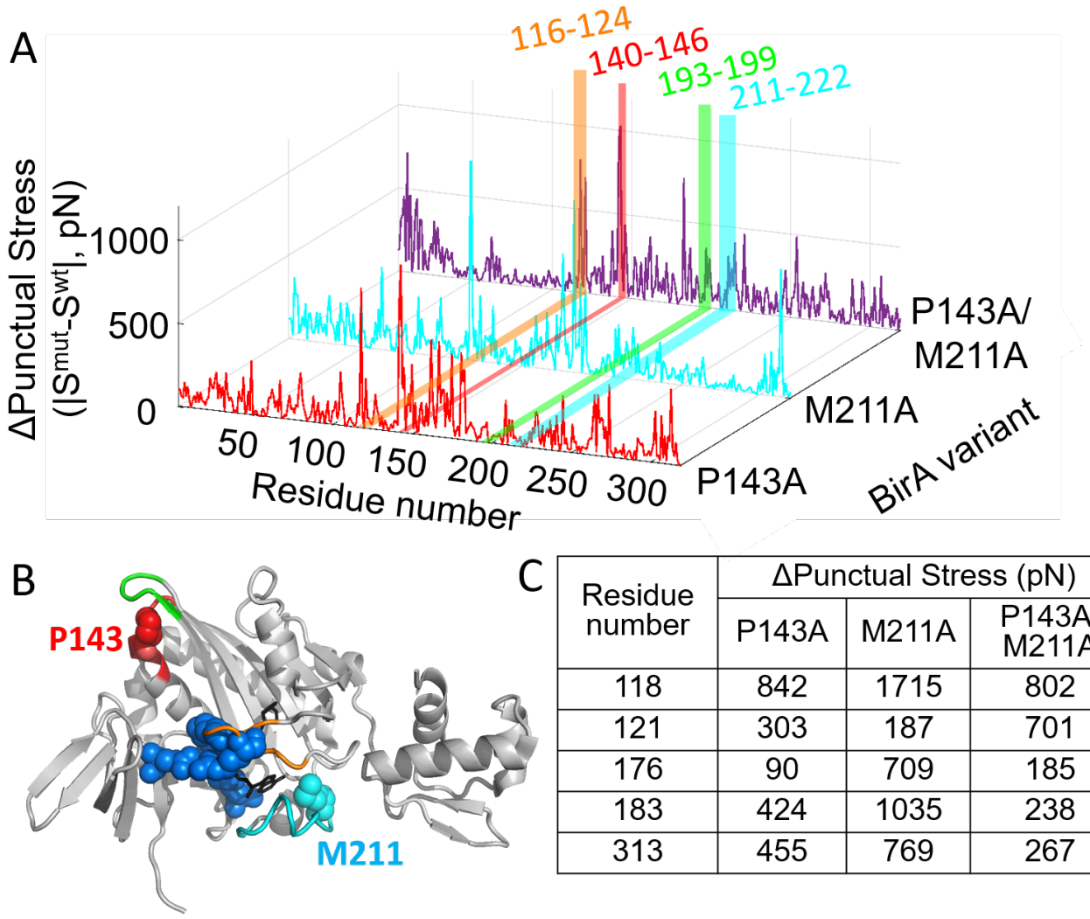
All measurements were carried out in Standard Buffer at 20°C. Errors were propagated from at least two independent measurements. The Gibbs free energies of dimerization were calculated using the equation $\Delta G^o = RT\ln K_D$. [a]From reference 173.[173] [b]From reference 176.[176]

5. Force Distribution Analysis indicates coupling between disorder-to-order transitions and the network in BirA allostery

Network function in coupling distant disorder-to-order transitions in BirA was investigated using Force Distribution Analysis (FDA). The analysis was conducted in collaboration with Dr. Matysiak's group. If the network functions in this coupling, bio-5'-AMP-linked folding is expected to be accompanied by changes in the force experienced by network residues. FDA enables calculation of the sum of the mechanical, or punctual, stress experienced by each amino acid from all pairwise-residue forces. Coupling between the network to disorder-to-order transitions was evaluated by performing FDA on the holo-forms of $BirA^{wt}$ and variants with alanine substitutions at P143, M211, and both P143 and M211 (Figure 38). Single alanine substitutions of residues P143 and M211, which perturb the disorder-to-order transitions at the dimerization and ligand binding surfaces, respectively, alter both effector binding and effector-linked dimerization.[170, 172, 173] Measurements performed on the double variant, $BirA^{P143A/M211A}$, revealed energetically favorable coupling between the two distant residues in effector binding.

The FDA indicates that perturbations to disorder-to-order transitions correlate with alterations in the stress experienced by network residues. The analysis was carried out on the equilibrium portion, the final 500 ns, of the MD trajectories obtained for each holo-variant and the results are presented as the difference in punctual stress at each residue in a variant relative to that experienced by the same residue in $holoBirA^{wt}$. For each single alanine variant significant punctual stress differences (>400pN) are

observed at the substituted residue position as well as at network residues (Figure 38A, B). For example, in BirA^{P143A}, in addition to P143, large perturbations to punctual stress are observed for network residues R118, K183, and E313. In the double variant, which revealed energetically favorable coupling between the distant P143 and M211,[170] the punctual stress perturbations for most residues are significantly smaller than those observed for either singly substituted parent. Control analyses were performed on variants with alanine substitution at residue R213 and an aspartic acid substitution at G154 (Figure 39). For the former substitution, which has minor effects on ligand binding and effector-induced dimerization, relatively small alterations in punctual stress at network residues were observed. By contrast, for BirA^{G154D}, in which allostery is enhanced, the punctual stress values at several network residues show large perturbations. The combined results reveal correlations among perturbation to disorder-to-order transitions, functional effects on allostery, and force distribution in the residue network.



Residue number	ΔPunctual Stress (pN)		
	P143A	M211A	P143A/ M211A
118	842	1715	802
121	303	187	701
176	90	709	185
183	424	1035	238
313	455	769	267

Figure 38. Force Distribution Analysis reveals coupling between disorder-to-order transitions and the network. (A) Absolute punctual stress difference in a variant relative holoBirA^{wt} at each residue position. (B) HoloBirA structural model with positions of amino acid substitutions, P143A and M211A, and network positions that show significant punctual stress changes in the variants (blue). The color scheme for protein segments that undergo disorder-to-order transitions is identical to that used in Figure 32. (C) Numerical values of punctual stress differences for the variants.

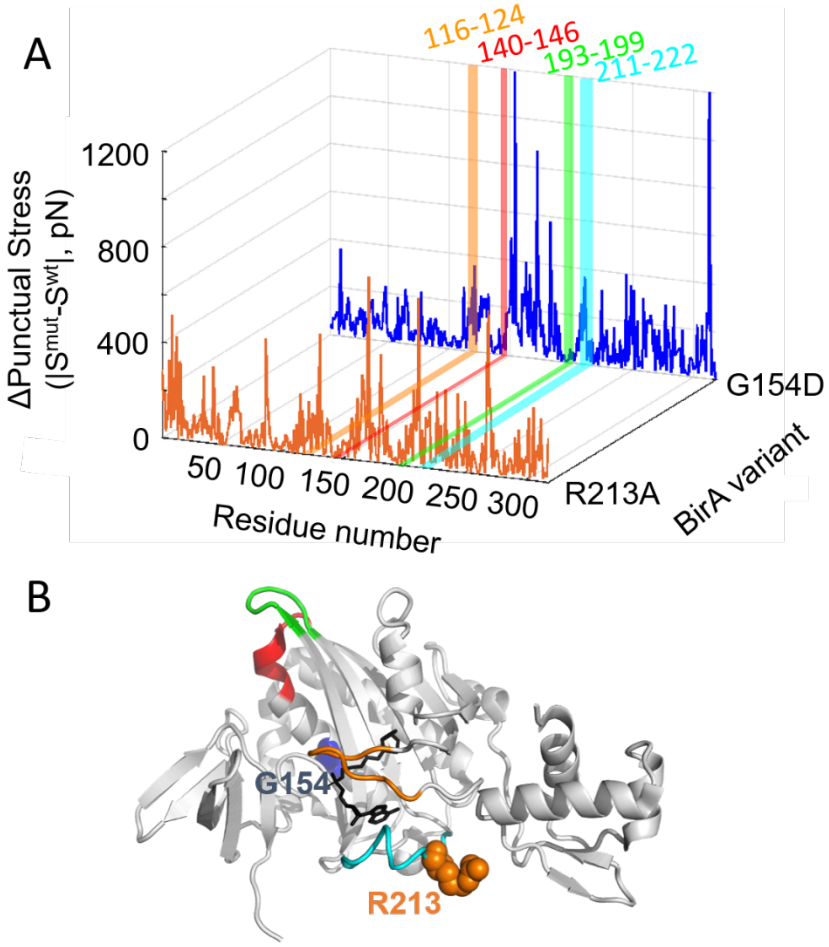


Figure 39. Punctual stress distribution for BirA proteins with substitutions at R213 and G154. (A) Absolute punctual stress difference between the variant and holoBirA^{wt} at each residue position (B) HoloBirA model showing locations of substitutions and loops that undergo disorder-to-order transitions (red, green, orange, cyan). The loop color schemes are identical in A and B.

E. Discussion

Although the significance of disorder for protein allostery is well-documented, the molecular basis of its function in allosteric communication is unknown for many systems. In *E. coli* BirA, coupled disorder-to-order transitions on two distant functional surfaces are critical for activating the protein for dimerization. In this work, combined experimental and computational studies reveal the role of a residue network in communication between the folding transitions.

Computational analysis yields a network that, although more extensive, overlaps significantly with that derived from structural analysis. Effector binding is indispensable to formation of many network interactions, either due to direct network residue contacts or because many network residues are in segments that become ordered concomitant with bio-5'-AMP binding. Notably, residues P126 and W223, both of which participate in disorder-to-order transitions at the ligand-binding site, are among the computationally identified network residues.[171, 172] Comparison of the apo and holoBirA structures reveals that although some network interactions form in the absence of ligand, the bond distances become shorter upon effector binding. For example, in the "hinges" of the two ligand binding loops, the distances between the charged groups on R235 and E110 are decreased by 0.9 angstroms and the hydrogen bond length between the side chain OH of Y132 and the backbone carbonyl of P126 is significantly shortened.[153, 154] Overall, the majority of the network residues in combination with the ligand form a continuous surface from the ligand binding site loop termini to the C-terminal domain (Figure 33). Although the network includes one

C-terminal domain residue, E313, no N-terminal DNA binding domain residues participate in the network. This is consistent with the decoupling of dimerization energetics from the affinity of the resulting holoBirA dimer for DNA observed for a number of BirA variants.[230]

Experimental studies reveal the importance of the network for BirA allostery. In this work measurements performed on BirA variants with alanine substitutions of network residues reveal important functional roles for the network in both bio-5'-AMP binding and coupled dimerization. Additionally, previous genetic studies indicated that residue R235 contributes to BirA-mediated transcription repression. Finally, residues P126 and W223 contribute to both effector binding and holoBirA dimerization.[171, 172] Only two residues, R119 and R121, which were designated as network residues based on structure, were not identified in the computational analysis. Experimental measurements reveal that both of these residues are important for bio-5'-AMP-induced dimerization. While structural data suggest that the R119 may directly contribute to the dimerization interface, no such role for R121 is evident from the structure. Analyses of holoBirA dimer simulation trajectories may yield additional information about these two residues.

The residue network and the disorder-to-order-transitions are integrated in BirA allosteric activation. First, residues that directly participate in disorder-to-order, including P126 and W223, contribute to both the network and to allostery. Second, as mentioned above, many network interactions are enabled by disorder-to-order transitions. Finally, Force Distribution Analysis (FDA) indicates that perturbations of

disorder-to-order in the ligand binding and dimerization surfaces alter punctual stress at network positions. Moreover, the magnitudes of these stress perturbations correlate with the functional effects of amino acid substitutions. However, information regarding changes in stress directionality (compression or expansion) cannot be extracted from the current analysis. The observed alterations in punctual stress do not reveal which residues experience a change in stress directionality or may have the same directionality but a change in magnitude.

The coupled residue network and disorder-to-order transitions likely enhance BirA dimerization by reducing the entropic penalty associated with the process. The modest Gibbs free energy of holoBirA dimerization at 20^{o}C reflects large opposing enthalpic and entropic contributions of +41 and -47.5 kcal/mol, respectively.[206] While solvent release is known to contribute significantly to this thermodynamic signature,[231] the configurational entropy of the BirA monomer could impact the overall dimerization free energy through its effect on the dimerization entropy. Thus, the very weak dimerization (ΔG^{o}= -2.5 kcal/mol)[166] of apoBirA, in which loops on both the dimerization and ligand binding surfaces are disordered and the network is absent, likely reflects the penalty of reducing the protein configurational entropy upon dimerization. Thermodynamic measurements of bio-5'-AMP binding to several BirA variants revealed a linear correlation between the heat capacity change of bio-5'-AMP binding to the monomer and the dimerization free energy of the resulting holo monomer, with less negative heat capacity changes associated with weaker dimerization.[175] These results are consistent with tuning of allosteric activation in these

variants *via* the extent to which effector binding induces folding and/or a reduction of the monomer internal dynamics.[232, 233] In the present work, a number of the network variants, all of which are perturbed in allostery, bind to bio-5'-AMP with entropies that are more unfavorable than that measured for $BirA^{wt}$ (Figure 35B, Table 8). These relatively unfavorable entropies may reflect decreased solvent release due to reduced ligand-linked folding. Direct dynamic measurements will reveal the relationship between allostery, local folding, network formation and BirA internal dynamics.

BirA allostery reflects contributions from several segments that are distributed throughout the protein primary structure (Figure 40). At a biological level this distributed mechanism renders the allosteric response robust to changes in protein sequence. Indeed, amino acid substitutions at 16 different residues that contribute to disorder-to-order transitions and/or the network result in holoBirA dimerization free energies within ±1 kcal/mol of that measured for $holoBirA^{wt}$ (Figure 40). Several of the residues at which substitutions result in modest energetic perturbations to dimerization are located at the network periphery. This suggests that amino acid substitutions preserve allosteric function provided that the network core remains intact to redistribute energetic perturbations that accompany the substitution.

Disorder can play an important role in the thermodynamics of allostery,[102] and manipulation of disorder provides a mechanism for altering allosteric coupling in proteins.[172-174, 234] The results reported in this work illustrate the critical role that residue networks can play in connecting distant disorder-to-order transitions in allostery. The work also highlights the great advantage of integrating computational and experimental

approaches for identifying these networks, testing their functional significance, and deciphering how they facilitate long-distance communication in proteins.

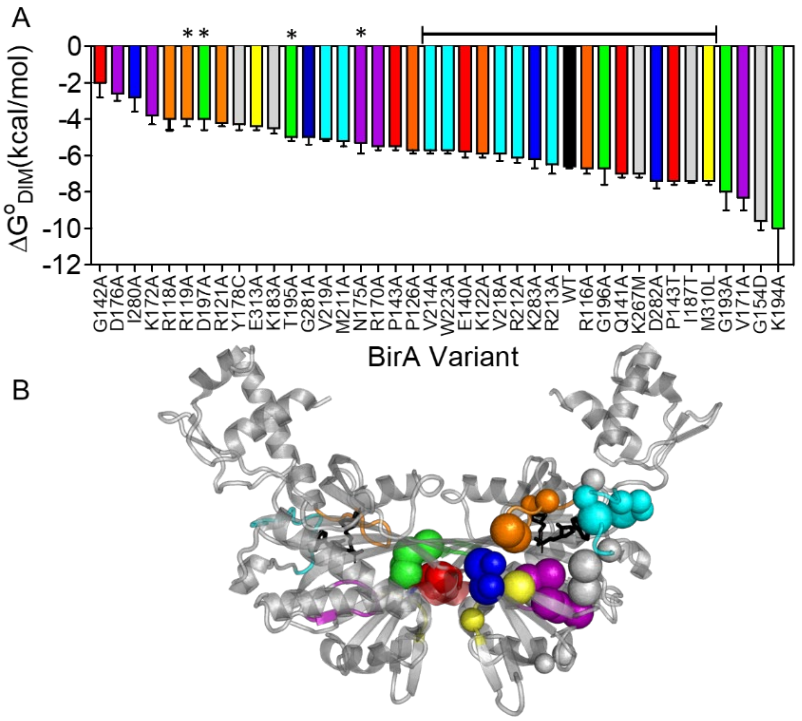


Figure 40. Amino acid substitutions tune the BirA allosteric response. (A) Amino acid substitutions in BirA yield a broad range of holoBirA dimerization free energies. The bracketed line indicates substitutions that alter the dimerization free energy by values within ±1 kcal/mol and asterisks signify residues that contribute directly to the dimer interface. (B) C-alpha carbons of amino acid positions in (A) shown on one subunit in the holoBirA dimer with small spheres signifying modest (within ±1 kcal.mol) and large spheres indicating larger (>±1 kcal/mol) effects on the dimerization free energy. Color code: Loops, Orange: 116-124; Red: 140-146; Purple: 170-176; Green: 193-199; Cyan: 211-222; Blue: 280-283; Yellow: 310-313; Gray: Protein Core. The model was created in Pymol[22] with PDB file 2EWN[154] as input.

Chapter 4: A conserved regulatory mechanism in bifunctional biotin protein ligases

A. Introduction

Cellular homeostasis relies on communication between metabolism and gene expression. The bifunctional biotin protein ligases (BirA) link metabolic demand for and production/uptake of biotin.[150, 160, 235, 236] In the *E. coli* system protein : protein interactions are critical for effecting this linkage. The ligase binds to biotin followed by ATP to produce biotinyl-5'-adenylate (bio-5'-AMP)[237] (Figure 41). Adenylate binding to BirA enhances the equilibrium association constant and Gibbs free energy of dimerization of the protein by 1000-fold and -4 kcal/mol, respectively.[183, 238] Given that homodimerization is a kinetic prerequisite to biotin operator (bioO) binding, transcription repression is regulated by holoBirA dimer availability (Figure 41).[239] When metabolic demand for biotin is high, holoBirA dimerization does not occur because the monomer preferentially binds to the abundant acetyl-CoA carboxylase, resulting in biotin transfer to the biotin carboxyl carrier protein (BCCP) subunit to activate the carboxylase for fatty acid biosynthesis (Figure 41). At low biotin demand, when the apoBCCP concentration is decreased, holoBirA monomer accumulates, dimerizes and binds to bioO to repress transcription of the biotin biosynthetic operon, thereby limiting biotin production when the cellular demand is low.

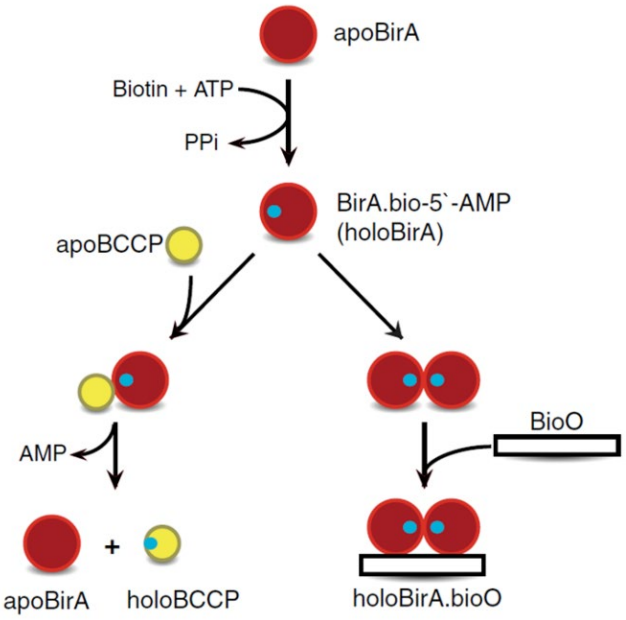


Figure 41. The *Escherichia coli* Biotin Regulatory System: BirA catalyzes the synthesis of bio-50-AMP from biotin and ATP and the resulting holoBirA either transfers biotin to apoBCCP or forms a homodimer that binds to the biotin operator to repress transcription.

Bifunctional ligases are present in a broad range of bacteria and archaea.[150] While some of these ligases regulate only transcription of the biotin biosynthetic genes, others are predicted to regulate expression of biosynthetic genes and/or those that code for biotin transport into the cell.[150] Few of these proteins have been subjected to biochemical studies. Studies of the *Staphylococcus aureus* and *Bacillus subtilis* enzymes provide a mixed picture of the functional energetics of the Biotin Regulatory System in these organisms. Both *in vivo* and *in vitro* measurements performed on the *B. subtilis* BirA (*Bs*BirA) indicate that, like the *E. coli* enzyme (*Ec*BirA), transcription

regulation and biotin operator binding are very sensitive to biotin concentration.[177, 178] However, measurements of *S. aureus* BirA (*Sa*BirA) binding to its cognate operator site using Electrophoretic Mobility Shift Assays (EMSA) indicate a relatively small difference in the overall binding energetics for the bio-5'-AMP-bound and ligand free forms of the protein, a result that predicts little sensitivity of transcription regulation to biotin concentration.[179] By contrast, more recent fluorescence anisotropy measurements of the binding reveal a large difference in the bioO binding energetics of the two species.[178]

Repression complex assembly by the *Ec*BirA occurs in two steps including dimerization followed by bioO binding.[239] Moreover, in the *E. coli* system, the large difference in the overall two-step repression complex assembly energetics in the absence and presence of bio-5'-AMP results from the enhanced dimerization of holoBirA relative to apoBirA.[230, 238] Previous measurements suggest that this is not the case for *Sa*BirA. Homodimerization measurements performed using combined sedimentation velocity and equilibrium methods indicate that apo*Sa*BirA dimerizes in the micromolar range of protein concentration and with energetics identical to those of the biotin-bound species.[179] By contrast, apo- and biotin-bound *Ec*BirA dimerize in millimolar concentration range.[238] No measurements of holo*Sa*BirA dimerization have been published. Given that *Sa*BirA, like *Ec*BirA, represses transcription *in vivo* in a biotin-dependent manner,[178] the apparent relatively tight dimerization of the apo-repressor is difficult to rationalize because it should enable DNA binding and repression in the absence of biotin.

In this work, we have used sedimentation equilibrium to measure coupling between small ligand binding and homodimerization of the bifunctional *Sa* and *Bs* BirA proteins. Results of these measurements indicate that these two proteins behave similarly to the *E. coli* enzyme in coupling between ligand binding and self-association. The one exception is that biotin alone, which has little effect on *Ec*BirA dimerization, enhances *Sa*BirA dimerization. The results support conservation of the mechanism of response to biotin concentration for the *S. aureus*, *B. subtilis*, and *E. coli* Biotin Regulatory Systems.

B. Materials and Methods

1. Chemicals and biochemicals

All reagents used were at least ACS grade. The d-biotin (Sigma-Aldrich) solutions were prepared in Standard Buffer containing 10 mM Tris (pH = 7.50 ± 0.02 at 20 °C), 200 mM KCl, 2.5 mM $MgCl_2$, and stored at -80 °C in 1 mL aliquots. Due to the absence of an absorption signal for biotin, these solutions were prepared by first weighing out the biotin using a microbalance and bringing the final solutions to full volume in a volumetric flask. Bio-5'-AMP was synthesized and purified as previously described and stored desiccated at -80 °C.[240, 241] The bio-5'-AMP solutions were prepared by dissolving lyophilized powder into Milli-Q H_2O, aliquoting the solution into 1 mL volumes and stored at -80ºC. The bio-5'-AMP concentration was determined by absorption spectroscopy at 259 nm using molar extinction coefficient of 15400 $M^{-1}cm^{-1}$. The integrity of the compound was checked by thin layer chromatography.

2. Expression and purification of *Bs* and *Sa*BirA

Both *Bs* and *Sa*BirA were expressed and purified using methods modified from previously published protocols.[177, 178] The pET19b plasmid containing the N-terminally histidine-tagged *BsBirA* coding sequence or the pET28b plasmid containing the C-terminally histidine-tagged *SaBirA* coding sequence, both obtained from J.E. Cronan's laboratory, was transformed into *E. coli* strain BL21 (λDE3). Cells were grown at 37ºC in LB broth containing 100 µg/mL ampicillin (*Bs*BirA) or 50 µg/mL kanamycin

(*Sa*BirA). Once OD_{600} of the culture reached 0.8, protein expression was induced for 12 hours at 30 °C by addition of IPTG to a final concentration of 1 mM.

All purification steps were performed at 4ºC. For *Sa*BirA purification, the cell pellet collected by centrifugation was resuspended in lysis buffer containing 50 mM HEPES (pH = 7.8 at 4ºC), 250 mM NaCl, 0.1 mM dithiothreitol (DTT), 10 mM imidazole, 5% glycerol and lysed by sonication. The crude cell lysate was centrifuged again at 8360 g to remove cellular debris and the resulting supernatant was loaded onto HisPur™ Ni-NTA resin (ThermoFisher scientific). The resin was washed with at least 10 column volumes (CV) of Ni-NTA Wash Buffer (lysis buffer containing 60 mM imidazole), and *Sa*BirA was eluted in the same buffer using a linear 60-250 mM imidazole gradient. Column fractions were analyzed by SDS-PAGE and those containing *Sa*BirA were pooled, dialyzed against SP sepharose starting buffer (50 mM Tris-HCl, 50 mM KCl, 5% glycerol, 0.1 mM DTT, pH = 7.5 at 4 °C), and the resulting sample was loaded onto SP Sepharose Fast Flow resin (GE Healthcare). The resin was washed with at least 10 CV of starting buffer, and *Sa*BirA was eluted using a linear 20-800 mM KCl gradient prepared in the same buffer. To remove residual biotin or bio-5'-AMP, fractions containing pure *Sa*BirA were pooled and dialyzed against lysis buffer, then loaded onto the Ni-NTA resin. At least 20 CV of a solution comprised of 1 mM ATP, 1 µM *E. coli* apo-BCCP, and 0.5 mM $MgCl_2$ in lysis buffer was applied to the column and exposed to the resin-bound *Sa*BirA for at least 2 h to allow for complete biotin transfer. After the incubation, the resin was washed with at least 10 CV of Ni-NTA wash buffer, after which *Sa*BirA was eluted with Ni-NTA elution buffer (lysis

buffer containing 250 mM imidazole). Finally, fractions containing *Sa*BirA were pooled, dialyzed against storage buffer containing 50 mM Tris-HCl (pH = 7.5 at 4 °C), 200 mM KCl, 5% glycerol, 0.1 mM DTT and stored at -80 °C in 1 mL aliquots.

The *Bs*BirA was purified using a method similar to that used for *Sa*BirA with a few modifications. The lysis buffer composition was 50 mM Tris-HCl, 250 mM NaCl, 0.5m M tris(2-carboxyethyl)phosphine (TCEP), 10 mM imidazole, 5% glycerol, pH = 8.7 at 4 °C. The crude cell lysate was centrifuged at 48400 g for 30 min prior to column chromatography. In addition, due to the protein's low solubility in SP sepharose starting buffer, SP sepharose column chromatography was replaced by a second Ni-NTA column chromatography step. The procedure for removing biotin or bio-5'-AMP contamination was identical to that used for *Sa*BirA. Purified *Bs*BirA was dialyzed against storage buffer containing 50 mM HEPES (pH = 7.8 at 4 °C), 250 mM NaCl, 0.5 mM TCEP, 5% glycerol and stored at -80 °C in 1 mL aliquots.

Protein concentrations were determined by UV absorbance using molar extinction coefficients of 39420 $M^{-1}cm^{-1}$ (*Bs*BirA) and 45380 $M^{-1}cm^{-1}$ (*Sa*BirA) at 280 nm calculated from amino acid compositions.[242] All proteins were at least 95% pure as assessed by Coomassie-staining of samples subjected to SDS-PAGE.

3. Biotinylation assay

A biotinylation assay, which is a modification of a previously described method,[243] was performed to ensure that purified *Bs* and *Sa* BirA were not contaminated with biotin or bio-5'-AMP. Solutions containing 20 μM *Bs* or *Sa* BirA, 20 μM *E. coli* apo-BCCP, and 500 μM ATP were prepared in Standard Buffer in the

absence or presence of 40 μM biotin. As a negative control, a reaction containing all components except biotin protein ligase was also prepared. The resulting solutions were incubated overnight for 16 h at 20 °C. To prepare for matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) mass spectrometry analysis, protein solutions were desalted and exchanged into the matrix (50 mM α-Cyano-4-hydroxycinnamic acid in 70% acetonitrile, 30% water) using C-18 Ziptips (Millipore). Spectra were acquired using an Axima-CFR MALDI-TOF (Shimadzu) mass spectrometer in linear mode, with insulin (Sigma-Aldrich) and cytochrome c from horse heart (Sigma-Aldrich) used as calibration standards.

4. Sedimentation equilibrium

All proteins subjected to sedimentation equilibrium measurements were dialyzed exhaustively against the appropriate buffer (Standard Buffer or 20 mM Tris, 150 mM NaCl, pH = 8.0 at 20 °C) prior to measurements. After filtering the protein solutions through 0.22μm PES syringe filters (Simsii), concentrations were determined by UV spectroscopy. Solutions were prepared at three final protein concentrations by dilution into dialysis buffer. For solutions containing biotin or bio-5'-AMP, the ligand was added immediately before centrifugation at a 1.5:1 ligand:protein molar ratio.

Sedimentation equilibrium measurements were performed using a Beckman Optima XL-I Analytical Ultracentrifuge (Beckman Coulter). A four-hole An-60 Ti rotor (Beckman Coulter) and 12-mm six-channel cells equipped with charcoal-filled Epon centerpieces and quartz windows were used in all experiments. Centrifugation was carried out at three speeds of 17000, 20000 and 23000 rpm for 8 h at each speed

and absorbance scans were acquired in step mode with five averages and a spacing of 0.001 cm. Overlay of a second scan acquired after an additional 1 h at each speed with the 8 h scan indicated that equilibrium had been achieved. The wavelengths used for data acquisition were 280, 295 or 300 nm, depending on the particular samples analyzed. At least three independent sedimentation measurements were carried out for each species, apo, biotin, and bio-5'-AMP, of *Sa* and *Bs*BirA.

5. Data analysis

Sedimentation equilibrium data were analyzed using Nonlin[218] in the Heteroanalyisis version 1.1.60 package downloaded from Center for Open Research Resources and Equipment (COR^2E) at the University of Connecticut (http://core.uconn.edu/resources/biophysics). Absorbance versus radius profiles were globally analyzed using a single ideal species model to obtain the reduced molecular weight σ using the following equation:

$$c(r) = c(r_0)e^{\sigma\frac{(r^2-r_0^2)}{2}} + \delta \quad (1)$$

where $c(r)$ is the protein concentration at radius r, $c(r_0)$ is the protein concentration at an arbitrary reference radial position r_0, and δ is the baseline offset. The best-fit reduced molecular weight was used to calculate the weight average molecular weight M of the sample using the following expression[244]:

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

where $\bar{\nu}$ is the partial specific volume of the protein, ρ is the buffer density, ω is the angular velocity of the rotor, R is the gas constant and T is the temperature. Partial

specific volumes of 0.7366 cm^3/g for *Bs*BirA and 0.7335 cm^3/g for *Sa*BirA and buffer densities were calculated using Sednterp (http://www.jphilo.mailway.com/download.htm).

Data were also analyzed using a monomer-dimer self-association model to obtain the equilibrium dissociation constant K_D for homodimerization:

$$c(r) = c(r_o)e^{\sigma_m \frac{(r^2 - r_o^2)}{2}} + \frac{1}{K_D}[c(r_o)e^{\sigma_m \frac{(r^2 - r_o^2)}{2}}]^2 + \delta \quad (3)$$

where σ_m is the reduced molecular weight for *Bs* or *Sa* BirA monomer calculated from Eq. (2), and r, $c(r)$, $c(r_0)$ and δ have the same significance as indicated for equation 1. The reduced molecular weight of the dimer is assumed to be twice that of the monomer.

6. Sequence alignment

Sequences of 34 class II bifunctional protein ligases were selected[150] and alignment was carried out using Clustal Omega.[245] The output file was analyzed using the Sequence Identity And Similarity (SIAS) tool (http://imed.med.ucm.es/Tools/sias.html) to obtain pairwise sequence identities, calculated using the mean length of sequences.

C. Results

1. *Sa* and *Bs* BirA preparations are free of biotin or bio-5'-AMP contamination

Since biotin protein ligases exhibit coupling between small ligand binding and dimerization, unambiguous interpretation of sedimentation equilibrium data obtained on any of these proteins requires knowledge of the ligation state. Previous studies indicate that biotin or bio-5'-AMP may co-purify with biotin protein ligases and a standard protocol developed in this laboratory,[246] which involves exposure of the ligases to a molar excess of apoBCCP in the presence of ATP, can eliminate such contamination.[177, 178, 246] The *Sa* and *Bs* BirA preparations used in the current studies were subjected to this published protocol. Before performing sedimentation measurements the proteins were assayed for the presence of contaminating biotin or bio-5'-AMP.[243] In the assay each purified ligase was incubated with ATP and apoBCCP in the absence and presence of biotin, and the resulting products were subjected to analysis by MALDI-TOF mass spectrometry. The spectra reveal that in the absence of added biotin, no biotinylated BCCP was detected for either *Sa* or *Bs* BirA. By contrast, the expected mass shift of 225 ±1 Da was observed in the control reaction to which biotin had been added (Figure 42, Table 10). Thus, neither the *Sa* or *Bs* BirA preparation used in the measurements reported in this work was contaminated with biotin or bio-5'-AMP.

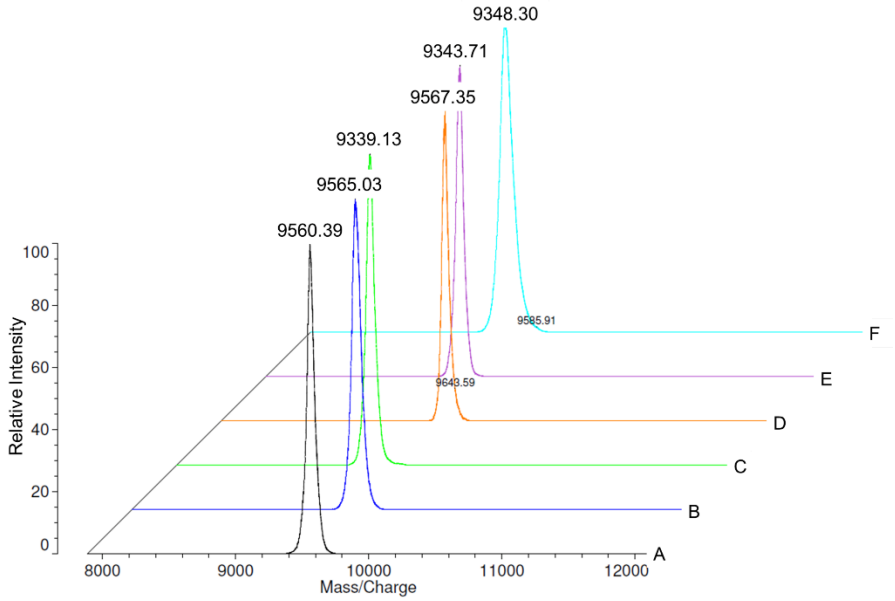


Figure 42. Maldi-ToF spectra of biotinylation products: reactions containing 20 µM *E. coli* apo-BCCP, 500 µM ATP, and (A) 20 µM EcBirA, 40 µM biotin; (B) 20 µM SaBirA, 40 µM biotin; (C) 20 µM SaBirA; (D) 20 µM BsBirA, 40 µM biotin; (E) 20 µM BsBirA; and (F) 40 µM biotin were incubated at 20 ºC overnight for 16 h. The products were subjected to analysis using a SHIMADZU Axima-CFR MALDI-TOF instrument.

Table 10. Maldi-ToF analysis of BCCP samples from biotinylation reactions

BirA species	Mass (daltons)	
	− biotin	+ biotin
*Sa*BirA	9339 ± 1	9565 ± 1
*Bs*BirA	9343 ± 1	9567 ± 1
-Enzyme	9348 ± 1	

MALDI mass spectrometry measurements showing mass-to-charge ratio of BCCP after the biotinylation assay in the presence or absence of biotin.

2. Dimerization measurements on SaBirA indicate coupling between small ligand binding and homodimerization

Coupling between small ligand binding and *Sa*BirA homodimerization was investigated using sedimentation equilibrium measurements. For all species each measurement was carried out on protein solutions prepared at three concentrations that were subjected to centrifugation at three speeds. Moreover, the concentration versus radius profile data obtained for all species were of high quality (Figure 43A).

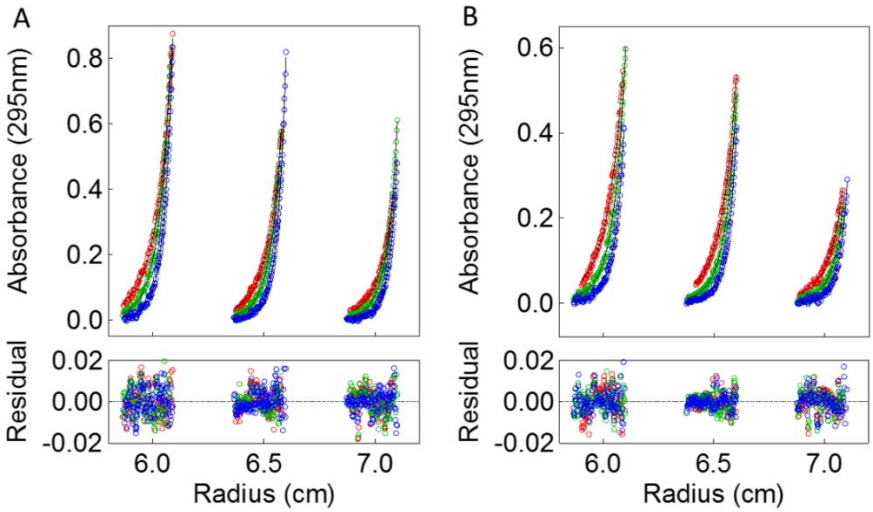


Figure 43. Absorbance versus radius profiles from sedimentation equilibrium measurements carried out on (A) 13, 10, 7 µM holo*Sa*BirA or (B) 12, 9, 6 µM holo*Bs*BirA at 17k (red), 20k (green), and 23k (blue) rpm. Solid lines are best-fits to a monomer-dimer model obtained from global analysis of nine datasets. The lower panels provide the residuals of the fit.

Results of the measurements performed on apo*Sa*BirA indicate that it undergoes no detectable dimerization in the conditions used for the measurements.

Analysis of data obtained at relatively low protein concentrations of approximately 10 μM indicates that oligomerization state of apo*Sa*BirA is well-described by a single species model, with an average molecular weight of 37±1 kDa, consistent with the monomer molecular weight predicted from the sequence. In an effort to obtain an estimate of the apo*Sa*BirA dimerization constant, measurements were performed at protein concentrations up to 100 μM. Analysis of the resulting data indicates that, even at these high concentrations, apo*Sa*BirA is best described as a single species with a molecular weight consistent with that expected of the monomer (Table 11, Figure 44). Attempts to analyze the data using a monomer-dimer model indicate an unrealistically large equilibrium dissociation constant (K_D), and the fit to the model, as judged by the magnitude of the variance of the fit, is not as good as the fit to a single species model (Table 11). The data are consistent with a K_D for apo*Sa*BirA dimerization in the millimolar or higher concentration range, which yields a lower limit for the Gibbs free energy of apo*Sa*BirA dimerization of -4.0 kcal/mol (Table 12).

Biotin and bio-5'-AMP binding significantly enhances *Sa*BirA dimerization. For sedimentation equilibrium measurements performed on small ligand-protein complexes, protein solutions were prepared in Standard Buffer at 1.5 : 1 molar ratio of the appropriate ligand to protein. The data obtained for the biotin-bound protein are well-described by a monomer-dimer model as judged by both the distribution of the residuals and the variance of the fit (Table 11, Figure 44), and the analysis yields an equilibrium dissociation constant for dimerization of 150 ± 20 μM (Table 12). Sedimentation equilibrium measurements performed on holo*Sa*BirA indicate even

tighter dimerization for this species. Again, the data are well described by the monomer-dimer model (Table 11, Figure 43A) and the analysis yields an equilibrium dissociation constant of 3 ± 1 μM. Based on the equilibrium constants, the calculated Gibbs free energies for dimerization of biotin-bound and bio-5'-AMP-bound proteins are -5.1 ± 0.1 kcal/mol and -7.6 ± 0.2 kcal/mol, respectively (Table 12).

Table 11. Testing sedimentation equilibrium data for *Sa* and *Bs*BirA against single species and monomer-dimer association models

Protein	Ligand	BirA loading concentration range (μM)	Single Species		Monomer-Dimer
			Weight average MW (kDa)	Variance[c]	Variance[c]
*Sa*BirA[a]	-	100, 90, 80	37 ± 1	0.0057	0.0060
	biotin	80, 70, 60	54 ± 2	0.0073	0.0070
	bio-5'-AMP	12, 9, 6	70 ± 2	0.0072	0.0060
*Sa*BirA[b]	-	17, 14, 11	37 ± 2	0.0042	0.0042
	bio-5'-AMP	12, 9, 6	60 ± 2	0.0062	0.0058
*Bs*BirA[a]	-	13, 10, 7	43 ± 1	0.0059	0.0057
	biotin	13, 10, 7	43 ± 1	0.0046	0.0043
	bio-5'-AMP	13, 10, 7	68 ± 2	0.0058	0.0054

Measurements were carried out in [a]Standard Buffer (10 mM Tris-HCl, 200 mM KCl, 2.5 mM $MgCl_2$, pH = 7.50±0.02 at 20 °C) or [b] 20 mM Tris, 150 mM NaCl, pH = 8.0 at 20 °C at three different rotor speeds. [c]Square root of the variance of the fit.

Table 12. Thermodynamic parameters governing *Sa* and *Bs*BirA dimerization

Protein	Ligand	BirA loading concentration range (μM)	K_D[c] (μM)	ΔG^o [d] (kcal/mol)	Maximum fraction dimer[e] (%)
*Sa*BirA[a]	-	100-5	N/A	N/A	N/A
	biotin	100-40	150 ± 20	-5.1 ± 0.1	38
	bio-5'-AMP	19-5	3 ± 1	-7.6 ± 0.2	64
*Sa*BirA[b]	-	17-4	N/A	N/A	N/A
	bio-5'-AMP	12-6	9 ± 3	-6.8 ± 0.2	49
*Bs*BirA[a]	-	25-7	300 ± 200	-4.7 ± 0.3	7
	biotin	25-7	300 ± 100	-4.8 ± 0.2	7
	bio-5'-AMP	20-7	4 ± 1	-7.4 ± 0.2	56

Measurements were carried out in [a]Standard Buffer (10 mM Tris-HCl, 200 mM KCl, 2.5 mM $MgCl_2$, pH=7.50±0.02 at 20 °C) or [b]20 mM Tris, 150 mM NaCl, pH = 8.0 at 20 °C at three different rotor speeds. [c]The reported equilibrium dissociation constants and Gibbs free energies for dimerization are the mean and standard deviation of at least three independent measurements. [d]Gibbs free energies were calculated using equation. $\Delta G = RT\ln K_D$. [e]The fraction dimer at the highest protein concentration in the concentration versus radius profile.

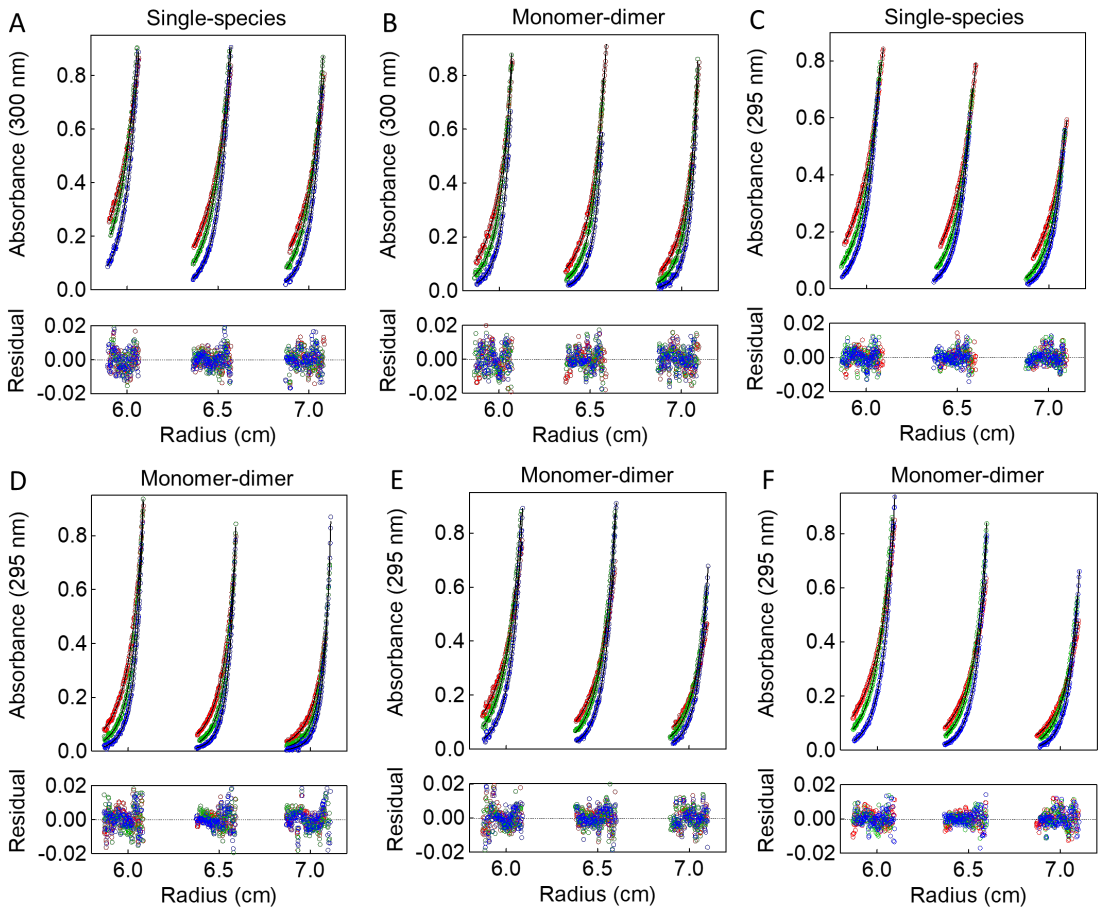


Figure 44. Sedimentation equilibrium absorbance versus radius profiles for *Sa* and *Bs*BirA. Centrifugation was performed at 17k (red), 20k (green) and 23k (blue) rpm using (A) 100, 90, 80 μM apo*Sa*BirA, (B) 80, 70 60 μM biotin-bound *Sa*BirA, (C) 17, 14, 11 μM apo*Sa*BirA, (D) 12, 9, 6 μM holo*Sa*BirA, (E) 13, 10, 7 μM apo*Bs*BirA, and (F) 13, 10, 7 μM biotin-bound *Bs*BirA. Solid lines are best-fits obtained from global analysis of 9 datasets to the model indicated for each panel, with lower panels providing residuals of the fit. Measurements were carried out in Standard Buffer containing 10 mM Tris, 200 mM KCl, 2.5 mM $MgCl_2$, pH = 7.5 at 20 ºC (A, B, E, F) or in 20 mM Tris, 150 mM NaCl, pH 8.0 at 20 °C (C and D).

3. Dimerization properties of *Sa*BirA are independent of buffer composition

Published measurements of *Sa*BirA dimerization yielded results that differ significantly from those reported in the previous section.[179] However, these measurements were carried out in buffer conditions that differ from those used in the current studies. The differences include the use of NaCl versus KCl as the monovalent salt at 150 mM compared to 200 mM in the current study and a 0.5 unit higher pH. To determine if the discrepancy in measured dimerization properties is due to the buffer conditions, sedimentation equilibrium analysis of *Sa*BirA was repeated in a buffer identical to that used in the previous studies. Measurements performed on apo*Sa*BirA indicate that the protein is monomeric (Table 11, Figure 44C). By contrast, data obtained with the bio-5'-AMP bound protein are well-described by a monomer-dimer model. The equilibrium dimerization constant of 9 ± 3 μM obtained from the measurements is similar in magnitude to that obtained in Standard Buffer (Table 11, Figure 44D).

4. Small ligand effects on *Bs*BirA dimerization mirror those observed for *Ec*BirA

Results of *in vivo* measurements of the biotin concentration-dependence of transcription repression suggest that *Bs*BirA and *Ec*BirA share the same regulatory mechanism.[177] To determine the relationship of these *in vivo* results to the *Bs*BirA self-association properties, sedimentation equilibrium measurements were carried out on apo, biotin-bound, and holo*Bs*BirA. Due to the limited solubility of *Bs*BirA in Standard Buffer, relatively low protein concentrations ranging from 7 to 25 μM were used for

all measurements. Nevertheless, the data obtained for all samples were of high quality (Figure 43B).

Sedimentation equilibrium data obtained for apo*Bs*BirA indicate that it dimerizes with a modest Gibbs free energy. The variances of the fits indicate that a monomer-dimer model provides a better description of the data than does a single species model (Table 11, Figure 44E). Moreover, the average molecular weight obtained from the single species fit of 43 ± 1 kDa is greater than that expected of the monomer (Table 11, Figure 44F). The equilibrium dissociation constant obtained from the data analysis is 300 ± 100 μM. The biotin-bound *Bs*BirA also dimerizes weakly with a constant identical to that of apo*Bs*BirA. Thus, consistent with observations made on *Ec*BirA, biotin has no effect on *Bs*BirA dimerization.[183] In contrast, addition of bio-5'-AMP to *Bs*BirA greatly enhances dimerization, yielding an equilibrium dissociation constant of 4 ± 1 μM (Table 11, Figure 43B)

5. Coupling between *Sa* and *Bs* BirA dimerization and ligand binding

The results of sedimentation equilibrium measurements on the *Sa* and *Bs* BirA species can be used to calculate the coupling free energy between small ligand binding and dimerization for the two proteins. This coupling free energy, ΔG^{o}_{c}, is defined as the difference between the Gibbs free energies for dimerization of the ligand-bound and ligand-free proteins (Figure 45A). The value for coupling between biotin binding and dimerization for *Sa*BirA is estimated to be at least -1.1 kcal/mole, based on the lower limit for the Gibbs free energy of apo*Sa*BirA dimerization of -4.0 kcal/mol. By contrast, consistent with *Ec*BirA, no coupling is observed between biotin binding to

*Bs*BirA and its dimerization. Both *Sa* and *Bs* BirA dimerization are significantly enhanced upon bio-5'-AMP binding. Again, using the lower limit estimate for the Gibbs free energy of apo*Sa*BirA dimerization of -4.0 kcal/mol, the estimated coupling free energy associated with bio-5'-AMP binding is at least -3.6 kcal/mol. The calculated value for *Bs*BirA is -2.7 ± 0.3 kcal/mol (Figure 45B).

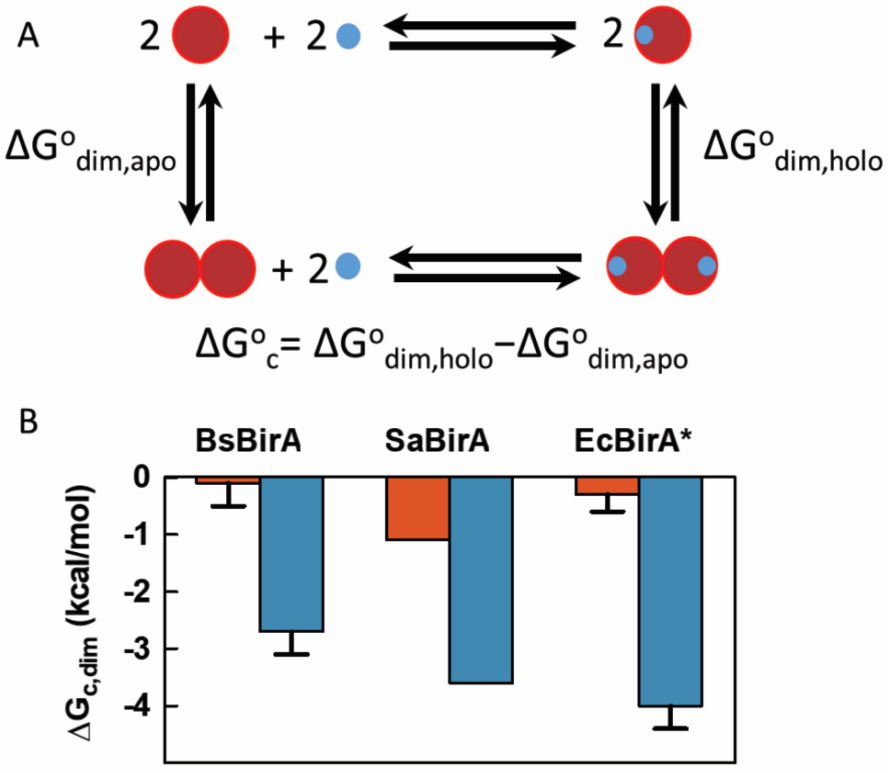


Figure 45. Coupling between small ligand binding and dimerization. (A) Thermodynamic cycle demonstrating coupling between bio-5'-AMP (blue) binding and homodimerization of BirA (red). The coupling free energy, $\Delta G^{o}_{c,dim}$, is the difference between Gibbs free energies of holoBirA, $\Delta G^{o}_{dim,holo}$, and apoBirA, $\Delta G^{o}_{dim,apo}$, dimerization. (B) Coupling free energies between biotin (red) and bio-5'-AMP (blue) binding for *Bs*, *Sa*, and *Ec*BirA. *Previously published data.[166]

D. Discussion

Both *Sa* and *Bs* BirA exhibit large thermodynamic coupling between bio-5'-AMP binding and dimerization. Sedimentation equilibrium analyses were performed on the apo, biotin-bound, and bio-5'-AMP-bound species of the two proteins. All measurements were performed on protein preparations that were free of biotin or bio-5'-AMP contamination, and excellent agreement was observed from the results of nonlinear least squares analysis of all data. In the absence of ligand, no apo*Sa*BirA dimerization was detected, which allows an estimate of a lower limit of 1 mM for the equilibrium dissociation constant governing dimerization of this species. By contrast, *Sa*BirA dimerization is enhanced upon the addition of either biotin or bio-5'-AMP, yielding estimates of the upper limits for the coupling free energies associated with the two ligands of -1.1 and -3.6 kcal/mol, respectively. Dimerization free energies of the apo and biotin-bound species of *Bs*BirA, both of which can be accurately measured, are identical. However, addition of bio-5'-AMP leads to a significant enhancement of dimerization, which results in a coupling free energy of -2.7 ± 0.2 kcal/mol. Thus, as previously demonstrated for *Ec*BirA, dimerization of both *Sa* and *Bs* BirA is significantly enhanced concomitantly with bio-5'-AMP binding.

The dimerization properties obtained for *Sa*BirA in this work differ significantly from those previously reported. The current work indicates no detectable dimerization for apo*Sa*BirA and a K_D of 150 ± 20 μM for the biotin-bound species. By contrast, in the previous study the authors reported K_D values for apo and biotin-bound *Sa*BirA dimerization of 29 ± 2 μM and 30 ± 2 μM, respectively.[179] The buffer

conditions used for these previous studies differ from those used for the current measurement. However, sedimentation equilibrium measurements of apo*Sa*BirA self-association performed in this work in that same buffer also indicate no detectable dimerization. Moreover, consistent with the results summarized above, measurement of holo*Sa*BirA dimerization in that buffer indicates large coupling between bio-5'-AMP binding and dimerization (Table 12). No measurements of holo*Sa*BirA dimerization were included in the previous study. A possible source of the discrepancy in the results is that *Sa*BirA preparations used in previous measurements contained residual biotin or bio-5'-AMP contamination. However, as part of the purification protocol, the authors exposed *Sa*BirA protein to biotin acceptor protein to remove such contamination. Moreover, based on the results of solid phase and immuno-blotting methods, the authors of the previous study concluded that the *Sa*BirA preparations used in their studies were free of small ligand contamination.[179] In this work a more direct mass spectrometry-based method of measuring the shift in the mass of apoBCCP to that of holoBCCP also indicates no residual contamination of either the *Sa* or *Bs* BirA preparation with biotin or bio-5'-AMP. Moreover, the large coupling between bio-5'-AMP binding and *Sa*BirA dimerization observed in the current measurements is consistent with both the *in vivo* biotin concentration-dependence of transcription repression and the recently reported effects of small ligands on sequence-specific DNA binding to the *Sa* biotin operator sequence.[178]

The large coupling free energies between bio-5'-AMP binding and dimerization measured for *Ec*, *Bs* and *Sa* BirA support a conserved molecular mechanism of

allosteric regulation for bifunctional biotin protein ligases/biotin repressors. The bifunctional ligases are widely distributed in eubacteria and archaea.[150] The *Ec*BirA transcription repression function is regulated by combined demand for and supply of biotin.[160, 236] In conditions of high biotin demand the bio-5'-AMP-BirA complex is rapidly consumed in biotin transfer to the BCCP subunit of acetyl-CoA carboxylase.[161, 162] A decrease in biotin demand, which accompanies a decreased growth rate, allows accumulation of both biotin and the holoBirA dimer, which can bind to the biotin operator to repress biotin biosynthetic operon transcription.[163] Henke and Cronan have demonstrated in *in vivo* measurements that, like *Ec*BirA, both *Sa* and *Bs* BirA show biotin-concentration dependent transcription repression at their respective regulatory sequences.[177, 178] They have, moreover, shown that over-expression of the biotin acceptor proteins results in derepression of the transcription. Finally, they demonstrated that for each of these Class II biotin ligases, the overall affinity for the cognate operator site is dramatically increased in the presence of bio-5'-AMP. Results presented in this work indicate that, as previously reported for *Ec*BirA, the enhanced repression complex assembly by *Sa* and *Bs*BirA reflects bio-5'-AMP-promoted dimerization.

The sequences of the three-biotin repressor proteins suggest that conservation is not required to achieve similar allosteric function. Alignment of the three sequences reveals that relative to *Ec*BirA *Sa* and *Bs*BirA show 22% and 27% conservation, respectively (Figure 46). Despite this low sequence identity, the 3-dimensional structures of *Ec* and *Sa*BirA are very similar (Figure 47A). In the *Ec*BirA structure, loops on the dimerization and ligand binding surfaces are known to play critical roles

in bio-5'-AMP binding, dimerization, and coupling between the two processes.[172, 174, 247, 248] Moreover, the disorder-to-order transitions that these loops undergo concomitant with bio-5'-AMP binding are key to allosteric activation of *Ec*BirA dimerization. Nevertheless, with the exception of the glycine-rich segment of the biotin binding loop that is required for biotin binding [189], the sequences of these loop regions in *Sa*, *Bs*, and *Ec* BirA show no conservation (Figure 47B). Thus, it appears that in the bifunctional biotin ligases, similar functional allostery can be achieved with high sequence divergence.

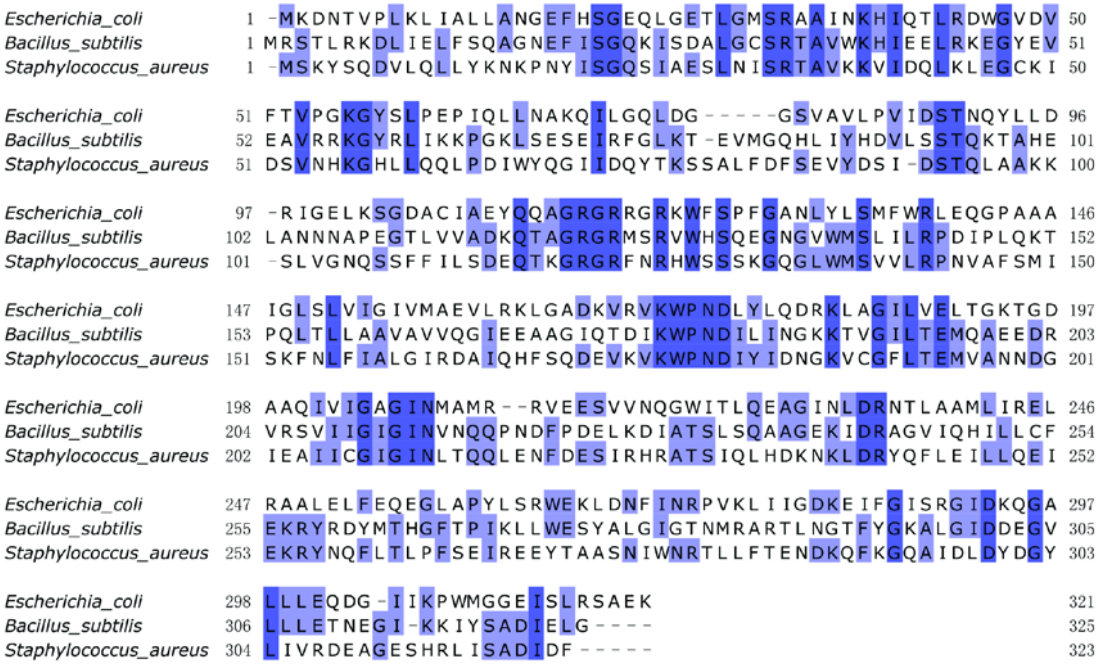
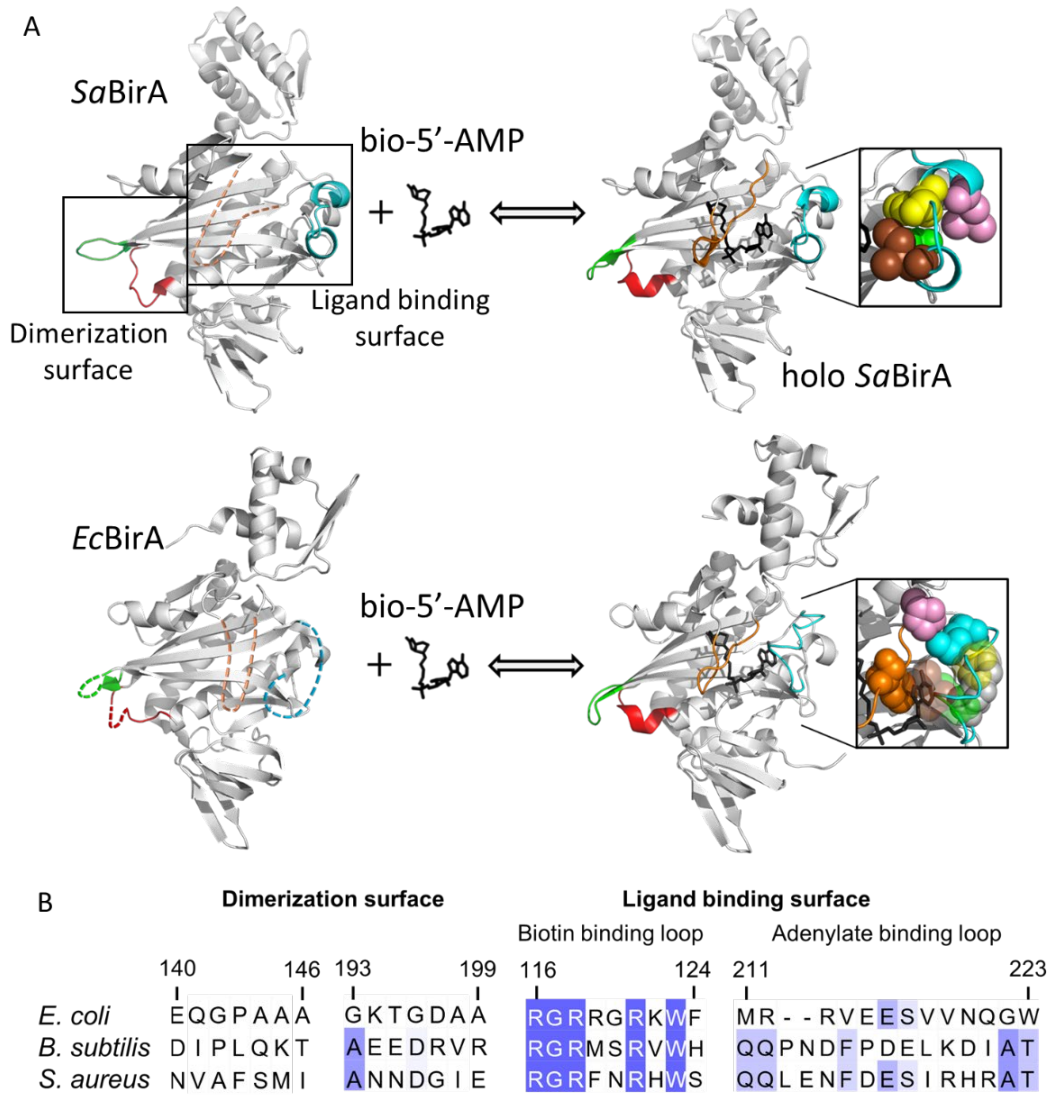


Figure 46. Multiple sequence alignment of *Ec*, *Bs*, and *Sa*BirA. Sequence alignment was performed using Clustal Omega,[245] and the figure was generated in JalView[249]. Residues are colored in blue based on their percentage identities.

Little is known about the relationship of sequence to allosteric function. However, protein disorder is now well-established as important for the thermodynamics of allostery,[102] perhaps because it renders a protein segment poised to respond to an allosteric signal. In the disorder-to-order transition accompanying allosteric activation of *Ec*BirA, hydrophobic side chains in the ligand binding surface loop condense around the bio-5'-AMP ligand to form a cluster (Figure 47A).[250, 251] Perturbation of this cluster through alanine substitution compromises both bio-5'-AMP binding and its coupling to dimerization.[172, 248] A similar clustering of hydrophobic residues in the ligand binding loop around bio-5'-AMP is observed in *Sa*BirA, albeit with a very different sequence (Figure 47B). On the *Ec*BirA dimerization surface the disorder-to-order transition accompanying allosteric activation includes extension of an α-helix, and packing of the two neighboring loops. Although the sequences of the analogous loop segments diverge completely from those of *Ec*BirA, the same helical extension and interloop packing are observed in comparison of the structures of apo and holo*Sa*BirA (Figure 47A). Thus, in *Ec* and *Sa*BirA the detailed structural features of allosteric activation are conserved in the absence of sequence conservation.



(Please see figure legend on the next page)

Figure 47. Sequence and structural comparison of *Ec*, *Bs*, and *Sa*BirA. (A) Disorder-to-order transitions in SaBirA (top) and EcBirA (bottom) upon bio-5'-AMP (black) binding. The models were generated in Pymol[22] using PDB files 3V8J: apo*Sa*BirA,[252] 3V8L: holo*Sa*BirA,[252] 1BIA: apo*Ec*BirA,[153] and 2EWN: holo*Ec*BirA.[154] Color coding for surface loops: *Sa*BirA, 143–149, red; 196–202, green; 119–129, orange; 214–228, cyan; *Ec*BirA, red, 140–146; green, 193–199; 116–124, orange; 211–222, cyan. Color coding for hydrophobic clusters: *Sa*BirA, L216-pink, F219-yellow, I223- brown, A227-green; *Ec*BirA, P126-pink, F124-orange, M211-cyan, V214-yellow, V218-brown, V219-green, W223-white. (B) Sequence alignment of *Ec*, *Bs*, and *Sa*BirA surface loops on the ligand binding and dimerization surfaces. The output of the alignment obtained in Clustal Omega[245] was used to generate the figure in JalView.[249] Residues are colored based on the percentage identity with numbering from the EcBirA sequence.

This page is intentionally left blank

Chapter 5: Summary and future work

The focus of this research is to understand the molecular mechanisms of allostery in biotin protein ligases (BirA). *E. coli* BirA is an allosterically regulated transcription repressor in which binding of the effector molecule, bio-5'-AMP, enhances homodimerization that involves a surface 30Å away from the ligand binding site. Previous studies indicated that disorder-to-order transitions on two distant functional surfaces play important roles in allostery.[153, 154, 171-174] In this dissertation, double-mutant cycle analysis was used to test reciprocal coupling between the two surfaces, and molecular dynamics (MD) simulations revealed structural and dynamic bases of this coupling. The molecular mechanisms of allosteric communication between disorder-to-order transitions were investigated using combined structural, functional, and computational approaches. In addition, allosteric regulation in *Staphylococcus aureus* and *Bacillus subtilis* BirA were also studied by functional measurements.

Results of double-mutant cycle analysis indicated reciprocal communication between the ligand binding and dimerization surfaces. The thermodynamic cycle of BirA allostery predicts reciprocal communication between disorder-to-order transitions on two functional surfaces, but previous studies using single alanine variants only demonstrated unidirectional communication from one surface to another.[171-174] In this work, double-mutant cycle analysis was employed to test the reciprocity.[205] Results of functional measurements showed that for all three residue pairs tested, perturbations to either bio-5'-AMP binding or dimerization by double alanine substitutions, one at each

surface, are smaller compared to the sum of energetic effects of two parent single alanine substitutions. The inequality between these two terms indicates that the two residues on separate functional surfaces are functionally coupled in either effector binding or dimerization, thereby supporting reciprocal communication across the protein.

MD simulations indicated that energetic coupling between distant surfaces occurs through modulation of disorder-to-order transitions. To investigate the molecular mechanism of allosteric communication in BirA, we worked in collaboration with Dr. Matysiak's group who performed simulation on wild type holoBirA and variants. Analyses of simulations data revealed that single alanine substitutions P143A and M211A, which weaken bio-5'-AMP binding and/or dimerization,[172, 173] alter structure and dynamics of surface loops that undergo disorder-to-order transitions. In addition, combining P143A and M211A substitutions reverts the structural and dynamic perturbations of the adenylate binding loop on the ligand binding surface, which is consistent to the positive coupling in bio-5'-AMP binding between the two residues. The combined results showed that alanine substitutions alter BirA allostery through their effects on disorder-to-order transitions.

Further investigation using integrated structural and computational approaches identified a residue network that may participate in allostery. Comparison of apo and holoBirA crystal structures revealed formation of an electrostatic network that is coupled to bio-5'-AMP binding.[176] This network is proposed to function in allostery as several residues in the network play an important role in effector-linked dimerization,

and in simulations perturbations to BirA allostery are correlated with rearrangement of the network.[173, 176] In this dissertation, results of simulation-based network analysis, which were calculated by Christopher Look and Riya Samanta in Dr. Matysiak's lab, identified a distributed residue network which includes the majority of residues in the structurally derived network. In addition, several residues in the newly identified network have been shown to function in either BirA allostery or BirA-mediated transcription repression.[171, 172] These discoveries suggest that the allosteric network in BirA is more extensive than previously anticipated.

The presence of this residue network in solution was tested by proton linkage analysis.[228] As the network involves a number of charged residues, network formation is predicted to alter the protonation state of these residues, which is reflected by proton release/uptake upon effector binding. Consistent with network formation in solution, proton linkage analysis showed that effector binding to wild type BirA is associated with proton release from the protein. Moreover, substitutions of network residues K172 and Y178, which are predicted to contribute to the network, alter the effector-linked protonation change. Therefore, these results support the hypothesis that effector binding is coupled to network formation in solution.

Results of functional measurements indicated that the network is critical for both input (bio-5'-AMP binding) and output (holoBirA dimerization) of BirA allostery. The role of this network in allostery was investigated by measuring the functional effects of residue substitutions at network positions. Results of ITC measurements demonstrated that these substitutions alter the free energy, enthalpy, or entropy of bio-

5'-AMP binding. Analyses of sedimentation equilibrium data showed that all substitutions perturb the dimerization free energy by at least 2 kcal/mol. These results indicate that network residues have significant contributions to BirA allostery.

The relationship between the residue network and disorder-to-order transitions was investigated by Force Distribution Analysis, which was carried out by Riya Samanta in Dr. Matysiak's group. Results of simulations demonstrated that alanine substitutions P143A and M211A, which perturb both allostery and disorder-to-order transitions,[170] lead to the redistribution of punctual stress. Notably, large punctual stress changes upon substitutions are mostly located in the network or loops undergo disorder-to-order transitions. In addition, consistent with functional coupling between residues P143 and M211, P143A substitution lowers the punctual stress changes of network residues induced by the M211A substitution. Overall, results of force distribution analysis indicated a linkage between disorder-to-order transitions and the residue network.

Finally, studies on *Bacillus subtilis* and *Staphylococcus aureus* biotin protein ligases (*Bs* and *Sa*BirA) indicated that allosteric regulation in these proteins is similar to that observed in *E. coli* BirA (*Ec*BirA). Previous *in vivo* measurements showed that *Bs* and *Sa*BirA repress transcription in a biotin-dependent manner,[177-179] suggesting that the two organisms share the same biotin regulatory system as that in *E. coli*. In this work, results of sedimentation equilibrium measurements showed that effector binding to both *Bs* and *Sa*BirA also enhances homodimerization, and the magnitudes of enhancement are similar to the -4 kcal/mol coupling free energy measured for *Ec*BirA.

The energetic coupling observed in *Bs* and *Sa*BirA supports a conserved allosteric regulation mechanism for bifunctional biotin protein ligases. Interestingly, despite the low sequence conservation among the three proteins, especially at disordered surface loops, *Ec* and *Sa*BirA have similar 3D structures and disorder-to-order transitions,[252] indicating that similar function can be achieved even with divergent sequences.

From all these results, we proposed a model for allosteric regulation in *E. coli* BirA, in which effector binding communicates to the dimerization surface through disorder-to-order transitions and a coupled residue network. We further hypothesize that this network, which is formed upon effector binding, functions in allostery by enabling communication between functional surfaces and the core of the protein. In addition, the conserved biotin regulatory systems in *Bs* and *Sa*BirA, as well as the structural similarity between *Sa* and *Ec*BirA, suggest that allosteric communication in *Bs* and *Sa*BirA may also occur through a similar mechanism.

Studies in this work greatly enhanced our understanding of allosteric regulation in *Ec*BirA. Nevertheless, future efforts are still required to fully elucidate the mechanisms of long-range communication between functional surfaces.

First, the mechanism of functional coupling in dimerization is still unknown. Combined functional and computational analyses revealed that coupling in bio-5'-AMP binding, such as that observed between residues P143 and M211, occurs by modulating the conformation of the adenylate binding loop. However, how residues G281 and M211, one on each functional surface, are coupled in dimerization is yet to be determined. MD simulations on holoBirA variants P143A and P143A/M211A

showed that compromised holoBirA dimerization is correlated with loss of helical extension and loop dynamics on the dimerization surface. Therefore, it is possible that coupling in dimerization also occurs by altering disorder-to-order transitions on the dimerization surface. MD simulations on BirA single alanine variants G281A, M211A, and double variant G281A/M211A may elucidate the mechanism of energetic coupling in dimerization.

Second, as indicated by punctual stress analysis, the residue network identified in this work is linked to disorder-to-order transitions on both functional surfaces. Their linkage on the ligand binding surface is supported by crystal structures and MD simulations, which showed that residues on the ligand binding surface directly participate in the network. However, the connection between disorder-to-order transitions on the dimerization surface and the network, which are separated by at least 15Å, is unclear. As effector binding does not elicit any structural change in the region between the dimerization surface and the network, it is likely that allosteric signal is propagated via dynamic changes. Analysis of dynamic information in MD simulations may reveal the communication mechanism to the dimerization surface, and NMR spectroscopy can be used to test the hypothesis generated by computation.

Third, additional tests are required to obtain a more comprehensive understanding of long-range communication between distant sites. In this dissertation, analyses using combined biophysical approaches revealed coupling between distant disorder-to-order transitions and their connection through a residue network. However, amino acid substitutions at positions other than these regions also affect BirA allostery.

For example, V171A and G154D substitutions in the core of the protein significantly enhance holoBirA dimerization (Figure 1),[173, 176] while residues in the C-terminal domain, such as G281, I280, and M310, also participate in allostery.[173, 176] It should be noted that although the loop comprised of residues 280-283 are located on the dimerization surface, it does not undergo the disorder-to-order transition upon effector binding. These observations suggest the presence of additional communication mechanisms that may be independent of the network identified in this work. It is also possible that allosteric regulation in BirA is associated with motions of the entire protein rather than specific networks.

NMR spectroscopy can be helpful in addressing this question, as it is sensitive to structural and dynamic changes upon perturbations, such as effector binding and residue substitutions. Comparison of apo and holoBirA spectra can identify residues that may participate in allostery, and spectra of BirA variants, which have perturbed allostery, can reveal correlated motions that are important in allostery.

Fourth, structural analysis indicated that *Sa*BirA, like *Ec*BirA, also undergo effector-induced disorder-to-order transitions on both functional surfaces.[252] Therefore, *Sa*BirA allostery may also be modulated by altering folding and dynamics of disorder-to-order transitions on two functional surfaces. This hypothesis can be tested by substituting disordered loop residues in *Sa*BirA to alanine and measure their functional effects on bio-5'-AMP binding and dimerization. The network analysis described in this work may reveal if *Sa*BirA utilizes a similar residue network to communicate distant disorder-to-order transitions.

Finally, given the broad distribution of biotin protein ligases in bacteria and archaea,[150] it is interesting to know how the allosteric function in Class II (bifunctional) ligases evolved without compromising their ability to catalyze biotin transfer. Studies on *E. coli* BirA revealed that the allosteric network in *E. coli* BirA includes several highly conserved residues important for biotinylation, such as K172, D176, and K183, and alanine substitutions of these residues lead to compromised biotin transfer.[173] By contrast, most other residues in the network (unpublished results) and loops that undergo disorder-to-order transitions do not contribute to the catalytic function.[173] This overlap between catalytic and allosteric residues suggests that allosteric regulation in Class II ligases may have evolved based on the catalytic function. As these residues are highly conserved among Class I and II biotin protein ligases, we further hypothesize that Class I ligases also possess residue networks that are a subset of that in Class II proteins. Network analysis will be performed on a Class I biotin protein ligase to identify the network in Class I proteins. Bioinformatic approaches, such as mutual information[253] and statistical coupling analyses[254], will be carried out to investigate the evolutionary relationship of residue networks between two classes of biotin protein ligases.

Allosteric regulation in BirA involves coupled disorder-to-order transitions on distant functional surfaces. Thermodynamic studies have established that disorder is beneficial for allosteric coupling,[102, 103] but the molecular mechanism of how disorder contributes to allostery has not been fully understood. Moreover, few studies focus on long-range communication between disordered regions and functional sites. For

example, the disorder-to-order transition of the active site loop in glucokinase is important for its kinetic cooperativity, and distant substitutions regulate glucokinase activity by altering the loop mobility. However, the mechanism of this long-range communication is still under investigation.[85, 104, 255] In tetracycline repressor (TetR), effector binding induces de-repression by promoting ordering of the DNA binding domain,[44] but the mechanism of allosteric communication between distant functional sites is also unknown. In this work, results of combined biophysical approaches demonstrated that BirA allostery can be modulated by altering the structure and dynamics of disordered surface loops, and long-distance communication between disorder-to-order transitions on distant functional surfaces can occur through formation of a coupled residue network. It is possible that in other allosteric proteins that involves effector-induced disorder-to-order transitions, folding of surface loops may be associated with network formation which enables long-range communication between distinct functional sites. In addition, studies in this dissertation support the critical role of residue networks in allosteric communication which has been reported in a number of allosteric systems.[87, 89, 127, 128, 256, 257] Finally, our work highlights the advantage of integrating computational tools and "wet-lab" experiments to deciphering allosteric mechanisms, and demonstrated how predictions made from computational analyses can be used to design subsequent experiments.

This page is intentionally left blank

Bibliography

[1] Knoflach, F., Mutel, V., Jolidon, S., Kew, J. N., Malherbe, P., Vieira, E., Wichmann, J., and Kemp, J. A. (2001) Positive allosteric modulators of metabotropic glutamate 1 receptor: characterization, mechanism of action, and binding site, *Proc Natl Acad Sci U S A 98*, 13402-13407.

[2] Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G., and Lu, P. (1996) Crystal structure of the lactose operon repressor and its complexes with DNA and inducer, *Science 271*, 1247-1254.

[3] Storer, A. C., and Cornish-Bowden, A. (1976) Kinetics of rat liver glucokinase. Co-operative interactions with glucose at physiologically significant concentrations, *Biochemical Journal 159*, 7-14.

[4] Bohr, C., Hasselbalch, K., and Krogh, A. (1904) Über einen in biologischer Beziehung wichtigen Einfluss, den die Kohlensäurespannung des Blutes auf dessen Sauerstoffbindung übt, *Acta Physiologica 16*, 402-412.

[5] Perutz, M. F. (1970) Stereochemistry of cooperative effects in haemoglobin, *Nature 228*, 726-739.

[6] Perutz, M. F., and TenEyck, L. (1972) Stereochemistry of cooperative effects in hemoglobin, In *Cold Spring Harbor symposia on quantitative biology*, pp 295-310, Cold Spring Harbor Laboratory Press.

[7] Perutz, M. F., Muirhead, H., Mazzarella, L., Crowther, R. A., Greer, J., and Kilmartin, J. V. (1969) Identification of residues responsible for the alkaline Bohr effect in haemoglobin, *Nature 222*, 1240-1243.

[8] Hill, A. V. (1910) The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves, *j. physiol. 40*, 4-7.

[9] Adair, G. S. (1925) The hemoglobin system VI. The oxygen dissociation curve of hemoglobin, *Journal of Biological Chemistry 63*, 529-545.

[10] Klotz, I. M. (1946) The application of the law of mass action to binding by proteins; interactions with calcium, *Arch Biochem 9*, 109-117.

[11] Pauling, L. (1935) The oxygen equilibrium of hemoglobin and its structural interpretation, *Proceedings of the National Academy of Sciences 21*, 186-191.

[12] Monod, J., Wyman, J., and Changeux, J. P. (1965) On the Nature of Allosteric Transitions: A Plausible Model, *Journal of molecular biology 12*, 88-118.

[13] Koshland, D. E., Jr., Nemethy, G., and Filmer, D. (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits, *Biochemistry 5*, 365-385.

[14] Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., Will, G., and North, A. (1960) Structure of hæmoglobin: a three-dimensional Fourier synthesis at 5.5-Å. resolution, obtained by X-ray analysis, *Nature 185*, 416.

[15] Perutz, M., Steinrauf, L., Stockell, A., and Bangham, A. (1959) Chemical and crystallographic study of the two fractions of adult horse haemoglobin, *Journal of molecular biology 1*, 402-IN418.

[16] Perutz, M., Muirhead, H., Cox, J., and Goaman, L. (1968) Three-dimensional Fourier synthesis of horse oxyhaemoglobin at 2.8 Å resolution: the atomic model, *Nature 219*, 131.
[17] Bolton, W., and Perutz, M. (1970) Three dimensional Fourier synthesis of horse deoxyhaemoglobin at 2.8 Å resolution, *Nature 228*, 551.
[18] Silva, M. M., Rogers, P. H., and Arnone, A. (1992) A third quaternary structure of human hemoglobin A at 1.7-A resolution, *Journal of Biological Chemistry 267*, 17248-17256.
[19] Safo, M. K., and Abraham, D. J. (2005) The enigma of the liganded hemoglobin end state: a novel quaternary structure of human carbonmonoxy hemoglobin, *Biochemistry 44*, 8347-8359.
[20] Fan, J. S., Zheng, Y., Choy, W. Y., Simplaceanu, V., Ho, N. T., Ho, C., and Yang, D. (2013) Solution structure and dynamics of human hemoglobin in the carbonmonoxy form, *Biochemistry 52*, 5809-5820.
[21] Lukin, J. A., Kontaxis, G., Simplaceanu, V., Yuan, Y., Bax, A., and Ho, C. (2003) Quaternary structure of hemoglobin in solution, *Proc Natl Acad Sci U S A 100*, 517-520.
[22] Schrodinger, LLC. (2015) The PyMOL Molecular Graphics System, Version 2.0.
[23] Park, S. Y., Yokoyama, T., Shibayama, N., Shiro, Y., and Tame, J. R. (2006) 1.25 A resolution crystal structures of human haemoglobin in the oxy, deoxy and carbonmonoxy forms, *Journal of molecular biology 360*, 690-701.
[24] Wikipedia. (2019) Heme, Wikipedia, The Free Encyclopedia.
[25] Cooper, A., and Dryden, D. T. F. (1984) Allostery without Conformational Change - a Plausible Model, *Eur Biophys J Biophy 11*, 103-109.
[26] Popovych, N., Sun, S., Ebright, R. H., and Kalodimos, C. G. (2006) Dynamically driven protein allostery, *Nat Struct Mol Biol 13*, 831-838.
[27] Petit, C. M., Zhang, J., Sapienza, P. J., Fuentes, E. J., and Lee, A. L. (2009) Hidden dynamic allostery in a PDZ domain, *Proc Natl Acad Sci U S A 106*, 18249-18254.
[28] Cameron, C. E., and Benkovic, S. J. (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*, 15792-15800.
[29] Levitt, M. (1983) Molecular dynamics of native protein. I. Computer simulation of trajectories, *Journal of molecular biology 168*, 595-617.
[30] Levitt, M. (1983) Molecular dynamics of native protein. II. Analysis and nature of motion, *Journal of molecular biology 168*, 621-657.
[31] Cooper, A. (1984) Protein fluctuations and the thermodynamic uncertainty principle, *Prog Biophys Mol Biol 44*, 181-214.
[32] Tsai, C. J., Del Sol, A., and Nussinov, R. (2009) Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms, *Mol Biosyst 5*, 207-216.
[33] Hilser, V. J. (2010) An ensemble view of allostery, *Science 327*, 653-654.

[34] Motlagh, H. N., Wrabl, J. O., Li, J., and Hilser, V. J. (2014) The ensemble nature of allostery, *Nature 508*, 331-339.
[35] Bai, X. C., McMullan, G., and Scheres, S. H. (2015) How cryo-EM is revolutionizing structural biology, *Trends Biochem Sci 40*, 49-57.
[36] Cheng, Y. F. (2018) Single-particle cryo-EM-How did it get here and where will it go, *Science 361*, 876-+.
[37] Nogales, E. (2016) The development of cryo-EM into a mainstream structural biology technique, *Nat Methods 13*, 24-27.
[38] Banerjee, S., Bartesaghi, A., Merk, A., Rao, P., Bulfer, S. L., Yan, Y., Green, N., Mroczkowski, B., Neitz, R. J., Wipf, P., Falconieri, V., Deshaies, R. J., Milne, J. L., Huryn, D., Arkin, M., and Subramaniam, S. (2016) 2.3 A resolution cryo-EM structure of human p97 and mechanism of allosteric inhibition, *Science 351*, 871-875.
[39] Bartesaghi, A., Aguerrebere, C., Falconieri, V., Banerjee, S., Earl, L. A., Zhu, X., Grigorieff, N., Milne, J. L. S., Sapiro, G., Wu, X., and Subramaniam, S. (2018) Atomic Resolution Cryo-EM Structure of beta-Galactosidase, *Structure 26*, 848-856 e843.
[40] Billeter, M., Wagner, G., and Wuthrich, K. (2008) Solution NMR structure determination of proteins revisited, *J Biomol Nmr 42*, 155-158.
[41] McMurry, L., Petrucci, R. E., Jr., and Levy, S. B. (1980) Active efflux of tetracycline encoded by four genetically different tetracycline resistance determinants in Escherichia coli, *Proc Natl Acad Sci U S A 77*, 3974-3977.
[42] Orth, P., Cordes, F., Schnappinger, D., Hillen, W., Saenger, W., and Hinrichs, W. (1998) Conformational changes of the Tet repressor induced by tetracycline trapping, *Journal of molecular biology 279*, 439-447.
[43] Orth, P., Schnappinger, D., Hillen, W., Saenger, W., and Hinrichs, W. (2000) Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system, *Nat Struct Biol 7*, 215-219.
[44] Reichheld, S. E., Yu, Z., and Davidson, A. R. (2009) The induction of folding cooperativity by ligand binding drives the allosteric response of tetracycline repressor, *Proc Natl Acad Sci U S A 106*, 22263-22268.
[45] Kisker, C., Hinrichs, W., Tovar, K., Hillen, W., and Saenger, W. (1995) The Complex Formed between Tet Repressor and Tetracycline-Mg2+ Reveals Mechanism of Antibiotic-Resistance, *Journal of molecular biology 247*, 260-280.
[46] Xu, Y., Zhang, D., Rogawski, R., Nimigean, C. M., and McDermott, A. E. (2019) Identifying coupled clusters of allostery participants through chemical shift perturbations, *Proceedings of the National Academy of Sciences*, 201811168.
[47] Xu, Y. Y., Bhate, M. P., and McDermott, A. E. (2017) Transmembrane allosteric energetics characterization for strong coupling between proton and potassium ion binding in the KcsA channel, *P Natl Acad Sci USA 114*, 8788-8793.

[48] Freiburger, L. A., Baettig, O. M., Sprules, T., Berghuis, A. M., Auclair, K., and Mittermaier, A. K. (2011) Competing allosteric mechanisms modulate substrate binding in a dimeric enzyme, *Nat Struct Mol Biol 18*, 288-294.
[49] Wright, G. D., and Ladak, P. (1997) Overexpression and characterization of the chromosomal aminoglycoside 6'-N-acetyltransferase from Enterococcus faecium, *Antimicrob Agents Chemother 41*, 956-960.
[50] Freiburger, L. A., Auclair, K., and Mittermaier, A. K. (2009) Elucidating protein binding mechanisms by variable-c ITC, *Chembiochem 10*, 2871-2873.
[51] Burk, D. L., Xiong, B., Breitbach, C., and Berghuis, A. M. (2005) Structures of aminoglycoside acetyltransferase AAC(6')-Ii in a novel crystal form: structural and normal-mode analyses, *Acta Crystallogr D Biol Crystallogr 61*, 1273-1279.
[52] Arunkumar, A. I., Campanello, G. C., and Giedroc, D. P. (2009) Solution structure of a paradigm ArsR family zinc sensor in the DNA-bound state, *Proc Natl Acad Sci U S A 106*, 18177-18182.
[53] Epstein, D. M., Benkovic, S. J., and Wright, P. E. (1995) Dynamics of the dihydrofolate reductase-folate complex: catalytic sites and regions known to undergo conformational change exhibit diverse dynamical features, *Biochemistry 34*, 11037-11048.
[54] Henzler-Wildman, K., and Kern, D. (2007) Dynamic personalities of proteins, *Nature 450*, 964-972.
[55] Mittermaier, A. K., and Kay, L. E. (2009) Observing biological dynamics at atomic resolution using NMR, *Trends Biochem Sci 34*, 601-611.
[56] Meiboom, S., and Gill, D. (1958) Modified spin‐echo method for measuring nuclear relaxation times, *Review of scientific instruments 29*, 688-691.
[57] Carr, H. Y., and Purcell, E. M. (1954) Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments, *Physical Review 94*, 630-638.
[58] Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1979) Investigation of exchange processes by two‐dimensional NMR spectroscopy, *The Journal of Chemical Physics 71*, 4546-4553.
[59] Tolman, J. R., and Ruan, K. (2006) NMR residual dipolar couplings as probes of biomolecular dynamics, *Chemical Reviews 106*, 1720-1736.
[60] Lipari, G. (1982) Model-free approach to the interpretation of nuclear megnetic relaxation in macro molecules. 2. Analysis of experimental results, *J. Am. Chem. Soc. 104*, 4559-4590.
[61] Lipari, G., and Szabo, A. (1982) Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity, *J Am Chem Soc 104*, 4546-4559.
[62] Futterman, S. (1957) Enzymatic reduction of folic acid and dihydrofolic acid to tetrahydrofolic acid, *The Journal of biological chemistry 228*, 1031-1038.
[63] Fierke, C. A., Johnson, K. A., and Benkovic, S. J. (1987) Construction and evaluation of the kinetic scheme associated with dihydrofolate reductase from Escherichia coli, *Biochemistry 26*, 4085-4092.

[64] Bolin, J. T., Filman, D. J., Matthews, D. A., Hamlin, R. C., and Kraut, J. (1982) Crystal structures of Escherichia coli and Lactobacillus casei dihydrofolate reductase refined at 1.7 A resolution. I. General features and binding of methotrexate, *Journal of Biological Chemistry 257*, 13650-13662.
[65] Sawaya, M. R., and Kraut, J. (1997) Loop and subdomain movements in the mechanism of Escherichia coli dihydrofolate reductase: crystallographic evidence, *Biochemistry 36*, 586-603.
[66] Boehr, D. D., McElheny, D., Dyson, H. J., and Wright, P. E. (2006) The dynamic energy landscape of dihydrofolate reductase catalysis, *Science 313*, 1638-1642.
[67] Oyen, D., Fenwick, R. B., Stanfield, R. L., Dyson, H. J., and Wright, P. E. (2015) Cofactor-Mediated Conformational Dynamics Promote Product Release From Escherichia coli Dihydrofolate Reductase via an Allosteric Pathway, *J Am Chem Soc 137*, 9459-9468.
[68] Schnell, J. R., Dyson, H. J., and Wright, P. E. (2004) Structure, dynamics, and catalytic function of dihydrofolate reductase, *Annu Rev Biophys Biomol Struct 33*, 119-140.
[69] Danielson, M. A., and Falke, J. J. (1996) Use of 19F NMR to probe protein structure and conformational changes, *Annual review of biophysics and biomolecular structure 25*, 163-195.
[70] Kerfah, R., Plevin, M. J., Sounier, R., Gans, P., and Boisbouvier, J. (2015) Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins, *Curr Opin Struct Biol 32*, 113-122.
[71] Grossoehme, N. E., and Giedroc, D. P. (2009) Energetics of allosteric negative coupling in the zinc sensor S. aureus CzrA, *J Am Chem Soc 131*, 17860-17870.
[72] Lee, S. E., Kamm, R. D., and Mofrad, M. R. (2007) Force-induced activation of talin and its possible role in focal adhesion mechanotransduction, *J Biomech 40*, 2096-2106.
[73] Hytonen, V. P., and Vogel, V. (2008) How force might activate talin's vinculin binding sites: SMD reveals a structural mechanism, *PLoS Comput Biol 4*, e24.
[74] Miao, Y., and McCammon, J. A. (2016) G-protein coupled receptors: advances in simulation and drug discovery, *Curr Opin Struct Biol 41*, 83-89.
[75] Jensen, M. O., Borhani, D. W., Lindorff-Larsen, K., Maragakis, P., Jogini, V., Eastwood, M. P., Dror, R. O., and Shaw, D. E. (2010) Principles of conduction and hydrophobic gating in K+ channels, *Proc Natl Acad Sci U S A 107*, 5833-5838.
[76] Moustafa, I. M., Shen, H., Morton, B., Colina, C. M., and Cameron, C. E. (2011) Molecular dynamics simulations of viral RNA polymerases link conserved and correlated motions of functional elements to fidelity, *Journal of molecular biology 410*, 159-181.
[77] Buchenberg, S., Knecht, V., Walser, R., Hamm, P., and Stock, G. (2014) Long-range conformational transition of a photoswitchable allosteric protein: molecular dynamics simulation study, *J Phys Chem B 118*, 13468-13476.

[78] Kumar, A., Rajendran, V., Sethumadhavan, R., and Purohit, R. (2013) Molecular dynamic simulation reveals damaging impact of RAC1 F28L mutation in the switch I region, *PLoS One 8*, e77453.
[79] Zhang, J., Li, C., Chen, K., Zhu, W., Shen, X., and Jiang, H. (2006) Conformational transition pathway in the allosteric process of human glucokinase, *Proc Natl Acad Sci U S A 103*, 13368-13373.
[80] Meglasson, M. D., and Matschinsky, F. M. (1984) New perspectives on pancreatic islet glucokinase, *Am J Physiol 246*, E1-13.
[81] Matschinsky, F. M. (1990) Glucokinase as glucose sensor and metabolic signal generator in pancreatic β-cells and hepatocytes, *Diabetes 39*, 647-652.
[82] Kamata, K., Mitsuya, M., Nishimura, T., Eiki, J., and Nagata, Y. (2004) Structural basis for allosteric regulation of the monomeric allosteric enzyme human glucokinase, *Structure 12*, 429-438.
[83] Grimsby, J., Sarabu, R., Corbett, W. L., Haynes, N. E., Bizzarro, F. T., Coffey, J. W., Guertin, K. R., Hilliard, D. W., Kester, R. F., Mahaney, P. E., Marcus, L., Qi, L., Spence, C. L., Tengi, J., Magnuson, M. A., Chu, C. A., Dvorozniak, M. T., Matschinsky, F. M., and Grippo, J. F. (2003) Allosteric activators of glucokinase: potential role in diabetes therapy, *Science 301*, 370-373.
[84] Brocklehurst, K. J., Payne, V. A., Davies, R. A., Carroll, D., Vertigan, H. L., Wightman, H. J., Aiston, S., Waddell, I. D., Leighton, B., Coghlan, M. P., and Agius, L. (2004) Stimulation of hepatocyte glucose metabolism by novel small molecule glucokinase activators, *Diabetes 53*, 535-541.
[85] Larion, M., Salinas, R. K., Bruschweiler-Li, L., Miller, B. G., and Bruschweiler, R. (2012) Order-disorder transitions govern kinetic cooperativity and allostery of monomeric human glucokinase, *PLoS Biol 10*, e1001452.
[86] Larion, M., Hansen, A. L., Zhang, F., Bruschweiler-Li, L., Tugarinov, V., Miller, B. G., and Bruschweiler, R. (2015) Kinetic Cooperativity in Human Pancreatic Glucokinase Originates from Millisecond Dynamics of the Small Domain, *Angew Chem Int Ed Engl 54*, 8129-8132.
[87] Daily, M. D., and Gray, J. J. (2009) Allosteric communication occurs via networks of tertiary and quaternary motions in proteins, *PLoS Comput Biol 5*, e1000293.
[88] Dima, R. I., and Thirumalai, D. (2006) Determination of network of residues that regulate allostery in protein families using sequence analysis, *Protein science : a publication of the Protein Society 15*, 258-268.
[89] Holliday, M. J., Camilloni, C., Armstrong, G. S., Vendruscolo, M., and Eisenmesser, E. Z. (2017) Networks of Dynamic Allostery Regulate Enzyme Function, *Structure 25*, 276-286.
[90] Wong, K. F., Selzer, T., Benkovic, S. J., and Hammes-Schiffer, S. (2005) Impact of distal mutations on the network of coupled motions correlated to hydride transfer in dihydrofolate reductase, *Proc Natl Acad Sci U S A 102*, 6807-6812.
[91] Foda, Z. H., Shan, Y., Kim, E. T., Shaw, D. E., and Seeliger, M. A. (2015) A dynamically coupled allosteric network underlies binding cooperativity in Src kinase, *Nature communications 6*, 5939.

[92] Sobolev, V., Sorokine, A., Prilusky, J., Abola, E. E., and Edelman, M. (1999) Automated analysis of interatomic contacts in proteins, *Bioinformatics 15*, 327-332.
[93] Pasi, M., Tiberti, M., Arrigoni, A., and Papaleo, E. (2012) xPyder: a PyMOL plugin to analyze coupled residues and their networks in protein structures, *J Chem Inf Model 52*, 1865-1874.
[94] Doncheva, N. T., Assenov, Y., Domingues, F. S., and Albrecht, M. (2012) Topological analysis and interactive visualization of biological networks and protein structures, *Nat Protoc 7*, 670-685.
[95] Feher, V. A., Durrant, J. D., Van Wart, A. T., and Amaro, R. E. (2014) Computational approaches to mapping allosteric pathways, *Curr Opin Struct Biol 25*, 98-103.
[96] Ribeiro, A. A., and Ortiz, V. (2014) Determination of Signaling Pathways in Proteins through Network Theory: Importance of the Topology, *J Chem Theory Comput 10*, 1762-1769.
[97] Ribeiro, A. A., and Ortiz, V. (2015) Energy propagation and network energetic coupling in proteins, *J Phys Chem B 119*, 1835-1846.
[98] Myers, R. S., Jensen, J. R., Deras, I. L., Smith, J. L., and Davisson, V. J. (2003) Substrate-induced changes in the ammonia channel for imidazole glycerol phosphate synthase, *Biochemistry 42*, 7013-7022.
[99] Rivalta, I., Sultan, M. M., Lee, N. S., Manley, G. A., Loria, J. P., and Batista, V. S. (2012) Allosteric pathways in imidazole glycerol phosphate synthase, *Proc Natl Acad Sci U S A 109*, E1428-1436.
[100] Lane, T. J., Shukla, D., Beauchamp, K. A., and Pande, V. S. (2013) To milliseconds and beyond: challenges in the simulation of protein folding, *Curr Opin Struct Biol 23*, 58-65.
[101] Kmiecik, S., Gront, D., Kolinski, M., Wieteska, L., Dawid, A. E., and Kolinski, A. (2016) Coarse-Grained Protein Models and Their Applications, *Chem Rev 116*, 7898-7936.
[102] Hilser, V. J., and Thompson, E. B. (2007) Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins, *Proc Natl Acad Sci U S A 104*, 8311-8315.
[103] Ferreon, A. C., Ferreon, J. C., Wright, P. E., and Deniz, A. A. (2013) Modulation of allostery by protein intrinsic disorder, *Nature 498*, 390-394.
[104] Whittington, A. C., Larion, M., Bowler, J. M., Ramsey, K. M., Bruschweiler, R., and Miller, B. G. (2015) Dual allosteric activation mechanisms in monomeric human glucokinase, *Proc Natl Acad Sci U S A 112*, 11553-11558.
[105] Sharma, C., Manjeshwar, R., and Weinhouse, S. (1963) Effects of Diet and Insulin on Glucose-Adenosine Triphosphate Phosphotransferases of Rat Liver, *The Journal of biological chemistry 238*, 3840-3845.
[106] Katzen, H. M., and Schimke, R. T. (1965) Multiple forms of hexokinase in the rat: tissue distribution, age dependency, and properties, *Proc Natl Acad Sci U S A 54*, 1218-1225.

[107] Iynedjian, P. B., Mobius, G., Seitz, H. J., Wollheim, C. B., and Renold, A. E. (1986) Tissue-specific expression of glucokinase: identification of the gene product in liver and pancreatic islets, *Proc Natl Acad Sci U S A 83*, 1998-2001.
[108] Magnuson, M. A. (1992) Tissue‐Specific regulation of glucokinase gene expression, *Journal of cellular biochemistry 48*, 115-121.
[109] Vionnet, N., Stoffel, M., Takeda, J., Yasuda, K., Bell, G. I., Zouali, H., Lesage, S., Velho, G., Iris, F., Passa, P., and et al. (1992) Nonsense mutation in the glucokinase gene causes early-onset non-insulin-dependent diabetes mellitus, *Nature 356*, 721-722.
[110] Matschinsky, F. M., Magnuson, M. A., Zelent, D., Jetton, T. L., Doliba, N., Han, Y., Taub, R., and Grimsby, J. (2006) The network of glucokinase-expressing cells in glucose homeostasis and the potential of glucokinase activators for diabetes therapy, *Diabetes 55*, 1-12.
[111] Agius, L. (2009) Targeting hepatic glucokinase in type 2 diabetes: weighing the benefits and risks, *Diabetes 58*, 18-20.
[112] Cornish-Bowden, A., and Cárdenas, M. L. (2004) Glucokinase: a monomeric enzyme with positive cooperativity, In *Glucokinase and glycemic disease: From basics to novel therapeutics*, pp 125-134, Karger Publishers.
[113] Ricard, J., Meunier, J. C., and Buc, J. (1974) Regulatory behavior of monomeric enzymes. 1. The mnemonical enzyme concept, *Eur J Biochem 49*, 195-208.
[114] Ainslie, G. R., Jr., Shill, J. P., and Neet, K. E. (1972) Transients and cooperativity. A slow transition model for relating transients and cooperative kinetics of enzymes, *The Journal of biological chemistry 247*, 7088-7096.
[115] Larion, M., and Miller, B. G. (2012) Homotropic allosteric regulation in monomeric mammalian glucokinase, *Arch Biochem Biophys 519*, 103-111.
[116] Larion, M., Salinas, R. K., Bruschweiler-Li, L., Bruschweiler, R., and Miller, B. G. (2010) Direct evidence of conformational heterogeneity in human pancreatic glucokinase from high-resolution nuclear magnetic resonance, *Biochemistry 49*, 7969-7971.
[117] Tzeng, S. R., and Kalodimos, C. G. (2009) Dynamic activation of an allosteric regulatory protein, *Nature 462*, 368-372.
[118] Swint-Kruse, L., and Matthews, K. S. (2009) Allostery in the LacI/GalR family: variations on a theme, *Curr Opin Microbiol 12*, 129-137.
[119] Thelwell, C., Robinson, N. J., and Turner-Cavet, J. S. (1998) An SmtB-like repressor from Synechocystis PCC 6803 regulates a zinc exporter, *Proc Natl Acad Sci U S A 95*, 10728-10733.
[120] Kuroda, M., Hayashi, H., and Ohta, T. (1999) Chromosome-determined zinc-responsible operon czr in Staphylococcus aureus strain 912, *Microbiol Immunol 43*, 115-125.
[121] McDevitt, C. A., Ogunniyi, A. D., Valkov, E., Lawrence, M. C., Kobe, B., McEwan, A. G., and Paton, J. C. (2011) A molecular mechanism for bacterial susceptibility to zinc, *PLoS Pathog 7*, e1002357.

[122] Eicken, C., Pennella, M. A., Chen, X., Koshlap, K. M., VanZile, M. L., Sacchettini, J. C., and Giedroc, D. P. (2003) A Metal–Ligand-mediated Intersubunit Allosteric Switch in Related SmtB/ArsR Zinc Sensor Proteins, *Journal of molecular biology 333*, 683-695.
[123] Chakravorty, D. K., Wang, B., Lee, C. W., Giedroc, D. P., and Merz, K. M., Jr. (2012) Simulations of allosteric motions in the zinc sensor CzrA, *J Am Chem Soc 134*, 3367-3376.
[124] Capdevila, D. A., Braymer, J. J., Edmonds, K. A., Wu, H., and Giedroc, D. P. (2017) Entropy redistribution controls allostery in a metalloregulatory protein, *Proc Natl Acad Sci U S A 114*, 4424-4429.
[125] Agarwal, P. K., Billeter, S. R., Rajagopalan, P. T., Benkovic, S. J., and Hammes-Schiffer, S. (2002) Network of coupled promoting motions in enzyme catalysis, *Proc Natl Acad Sci U S A 99*, 2794-2799.
[126] Fuentes, E. J., Gilmore, S. A., Mauldin, R. V., and Lee, A. L. (2006) Evaluation of energetic and dynamic coupling networks in a PDZ domain protein, *Journal of molecular biology 364*, 337-351.
[127] Gerek, Z. N., and Ozkan, S. B. (2011) Change in allosteric network affects binding affinities of PDZ domains: analysis through perturbation response scanning, *PLoS Comput Biol 7*, e1002154.
[128] Doshi, U., Holliday, M. J., Eisenmesser, E. Z., and Hamelberg, D. (2016) Dynamical network of residue-residue contacts reveals coupled allosteric effects in recognition, catalysis, and mutation, *Proc Natl Acad Sci U S A 113*, 4735-4740.
[129] Thomas, S. M., and Brugge, J. S. (1997) Cellular functions regulated by Src family kinases, *Annu Rev Cell Dev Biol 13*, 513-609.
[130] Roskoski, R., Jr. (2005) Src kinase regulation by phosphorylation and dephosphorylation, *Biochem Biophys Res Commun 331*, 1-14.
[131] Summy, J. M., and Gallick, G. E. (2003) Src family kinases in tumor progression and metastasis, *Cancer Metastasis Rev 22*, 337-358.
[132] Wheeler, D. L., Iida, M., and Dunn, E. F. (2009) The role of Src in solid tumors, *Oncologist 14*, 667-678.
[133] Druker, B. J., Tamura, S., Buchdunger, E., Ohno, S., Segal, G. M., Fanning, S., Zimmermann, J., and Lydon, N. B. (1996) Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells, *Nat Med 2*, 561-566.
[134] Lombardo, L. J., Lee, F. Y., Chen, P., Norris, D., Barrish, J. C., Behnia, K., Castaneda, S., Cornelius, L. A., Das, J., and Doweyko, A. M. (2004) Discovery of N-(2-chloro-6-methyl-phenyl)-2-(6-(4-(2-hydroxyethyl)-piperazin-1-yl)-2-methylpyrimidin-4-ylamino) thiazole-5-carboxamide (BMS-354825), a dual Src/Abl kinase inhibitor with potent antitumor activity in preclinical assays, *Journal of medicinal chemistry 47*, 6658-6661.
[135] Seeliger, M. A., Nagar, B., Frank, F., Cao, X., Henderson, M. N., and Kuriyan, J. (2007) c-Src binds to the cancer drug imatinib with an inactive Abl/c-Kit

conformation and a distributed thermodynamic penalty, *Structure 15*, 299-311.
[136] Xu, W., Doshi, A., Lei, M., Eck, M. J., and Harrison, S. C. (1999) Crystal structures of c-Src reveal features of its autoinhibitory mechanism, *Mol Cell 3*, 629-638.
[137] Cowan-Jacob, S. W., Fendrich, G., Manley, P. W., Jahnke, W., Fabbro, D., Liebetanz, J., and Meyer, T. (2005) The crystal structure of a c-Src complex in an active conformation suggests possible steps in c-Src activation, *Structure 13*, 861-871.
[138] Gonfloni, S., Williams, J. C., Hattula, K., Weijland, A., Wierenga, R. K., and Superti-Furga, G. (1997) The role of the linker between the SH2 domain and catalytic domain in the regulation and function of Src, *Embo Journal 16*, 7261-7271.
[139] Boggon, T. J., and Eck, M. J. (2004) Structure and regulation of Src family kinases, *Oncogene 23*, 7918-7927.
[140] Liu, B. A., Engelmann, B. W., and Nash, P. D. (2012) The language of SH2 domain interactions defines phosphotyrosine-mediated signal transduction, *FEBS Lett 586*, 2597-2605.
[141] Williams, J. C., Weijland, A., Gonfloni, S., Thompson, A., Courtneidge, S. A., Superti-Furga, G., and Wierenga, R. K. (1997) The 2.35 A crystal structure of the inactivated form of chicken Src: a dynamic molecule with multiple regulatory interactions, *Journal of molecular biology 274*, 757-775.
[142] Xu, W., Harrison, S. C., and Eck, M. J. (1997) Three-dimensional structure of the tyrosine kinase c-Src, *Nature 385*, 595-602.
[143] Hubbard, S. R., and Till, J. H. (2000) Protein tyrosine kinase structure and function, *Annu Rev Biochem 69*, 373-398.
[144] Ozkirimli, E., and Post, C. B. (2006) Src kinase activation: A switched electrostatic network, *Protein science : a publication of the Protein Society 15*, 1051-1062.
[145] Tse, A., and Verkhivker, G. M. (2015) Molecular Dynamics Simulations and Structural Network Analysis of c-Abl and c-Src Kinase Core Proteins: Capturing Allosteric Mechanisms and Communication Pathways from Residue Centrality, *J Chem Inf Model 55*, 1645-1662.
[146] Brinda, K. V., and Vishveshwara, S. (2005) A network representation of protein structures: implications for protein stability, *Biophys J 89*, 4159-4170.
[147] Vijayabaskar, M. S., and Vishveshwara, S. (2010) Interaction energy based protein structure networks, *Biophys J 99*, 3704-3715.
[148] LaFevre-Bernt, M., Sicheri, F., Pico, A., Porter, M., Kuriyan, J., and Miller, W. T. (1998) Intramolecular regulatory interactions in the Src family kinase Hck probed by mutagenesis of a conserved tryptophan residue, *The Journal of biological chemistry 273*, 32129-32134.
[149] Gonfloni, S., Frischknecht, F., Way, M., and Superti-Furga, G. (1999) Leucine 255 of Src couples intramolecular interactions to inhibition of catalysis, *Nat Struct Biol 6*, 760-764.

[150] Rodionov, D. A., Mironov, A. A., and Gelfand, M. S. (2002) Conservation of the biotin regulon and the BirA regulatory signal in Eubacteria and Archaea, *Genome Res 12*, 1507-1516.
[151] Prakash, O., and Eisenberg, M. A. (1979) Biotinyl 5'-adenylate: corepressor role in the regulation of the biotin genes of Escherichia coli K-12, *Proc Natl Acad Sci U S A 76*, 5592-5595.
[152] Barker, D. F., and Campbell, A. M. (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, *Journal of molecular biology 146*, 469-492.
[153] Wilson, K. P., Shewchuk, L. M., Brennan, R. G., Otsuka, A. J., and Matthews, B. W. (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*, 9257-9261.
[154] Wood, Z. A., Weaver, L. H., Brown, P. H., Beckett, D., and Matthews, B. W. (2006) Co-repressor induced order and biotin repressor dimerization: a case for divergent followed by convergent evolution, *Journal of molecular biology 357*, 509-523.
[155] Barker, D. F., and Campbell, A. M. (1981) The birA gene of Escherichia coli encodes a biotin holoenzyme synthetase, *Journal of molecular biology 146*, 451-467.
[156] Otsuka, A., and Abelson, J. (1978) The regulatory region of the biotin operon in Escherichia coli, *Nature 276*, 689-694.
[157] Abbott, J., and Beckett, D. (1993) Cooperative binding of the Escherichia coli repressor of biotin biosynthesis to the biotin operator sequence, *Biochemistry 32*, 9649-9656.
[158] Streaker, E. D., and Beckett, D. (1998) A map of the biotin repressor-biotin operator interface: binding of a winged helix-turn-helix protein dimer to a forty base-pair site, *Journal of molecular biology 278*, 787-800.
[159] Ganguly, J. (1960) Studies on the mechanism of fatty acid synthesis. VII. Biosynthesis of fatty acids from malonyl CoA, *Biochim Biophys Acta 40*, 110-118.
[160] Cronan, J. E., Jr. (1989) The E. coli bio operon: transcriptional repression by an essential protein modification enzyme, *Cell 58*, 427-429.
[161] Streaker, E. D., and Beckett, D. (2006) The biotin regulatory system: kinetic control of a transcriptional switch, *Biochemistry 45*, 6417-6425.
[162] Zhao, H., and Beckett, D. (2008) Kinetic partitioning between alternative protein-protein interactions controls a transcriptional switch, *Journal of molecular biology 380*, 223-236.
[163] Li, S. J., and Cronan, J. E., Jr. (1993) Growth rate regulation of Escherichia coli acetyl coenzyme A carboxylase, which catalyzes the first committed step of lipid biosynthesis, *Journal of bacteriology 175*, 332-340.

[164] Wang, J., and Beckett, D. (2017) A conserved regulatory mechanism in bifunctional biotin protein ligases, *Protein science : a publication of the Protein Society 26*, 1564-1573.
[165] Streaker, E. D., and Beckett, D. (2003) Coupling of protein assembly and DNA binding: biotin repressor dimerization precedes biotin operator binding, *Journal of molecular biology 325*, 937-948.
[166] Streaker, E. D., Gupta, A., and Beckett, D. (2002) The biotin repressor: thermodynamic coupling of corepressor binding, protein assembly, and sequence-specific DNA binding, *Biochemistry 41*, 14263-14271.
[167] Buoncristiani, M. R., Howard, P. K., and Otsuka, A. J. (1986) DNA-binding and enzymatic domains of the bifunctional biotin operon repressor (BirA) of Escherichia coli, *Gene 44*, 255-261.
[168] Chapman-Smith, A., Mulhern, T. D., Whelan, F., Cronan, J. E., and Wallace, J. C. (2001) The C-terminal domain of biotin protein ligase from E-coli is required for catalytic activity, *Protein Science 10*, 2608-2617.
[169] Weaver, L. H., Kwon, K., Beckett, D., and Matthews, B. W. (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*, 6045-6050.
[170] Wang, J., Custer, G., Beckett, D., and Matysiak, S. (2017) Long Distance Modulation of Disorder-to-Order Transitions in Protein Allostery, *Biochemistry 56*, 4478-4488.
[171] Naganathan, S., and Beckett, D. (2007) Nucleation of an allosteric response via ligand-induced loop folding, *Journal of molecular biology 373*, 96-111.
[172] Eginton, C., Naganathan, S., and Beckett, D. (2015) Sequence-function relationships in folding upon binding, *Protein science : a publication of the Protein Society 24*, 200-211.
[173] Adikaram, P. R., and Beckett, D. (2012) Functional versatility of a single protein surface in two protein:protein interactions, *Journal of molecular biology 419*, 223-233.
[174] Eginton, C., Cressman, W. J., Bachas, S., Wade, H., and Beckett, D. (2015) Allosteric coupling via distant disorder-to-order transitions, *Journal of molecular biology 427*, 1695-1704.
[175] Cressman, W. J., and Beckett, D. (2016) Heat Capacity Changes and Disorder-to-Order Transitions in Allosteric Activation, *Biochemistry 55*, 243-252.
[176] He, C., Custer, G., Wang, J., Matysiak, S., and Beckett, D. (2018) Superrepression through Altered Corepressor-Activated Protein:Protein Interactions, *Biochemistry 57*, 1119-1129.
[177] Henke, S. K., and Cronan, J. E. (2014) Successful conversion of the Bacillus subtilis BirA Group II biotin protein ligase into a Group I ligase, *PLoS One 9*, e96757.
[178] Henke, S. K., and Cronan, J. E. (2016) The Staphylococcus aureus group II biotin protein ligase BirA is an effective regulator of biotin operon

transcription and requires the DNA binding domain for full enzymatic activity, *Molecular microbiology 102*, 417-429.
[179] Soares da Costa, T. P., Yap, M. Y., Perugini, M. A., Wallace, J. C., Abell, A. D., Wilce, M. C., Polyak, S. W., and Booker, G. W. (2014) Dual roles of F123 in protein homodimerization and inhibitor binding to biotin protein ligase from Staphylococcus aureus, *Molecular microbiology 91*, 110-120.
[180] Christopoulos, A., and Kenakin, T. (2002) G protein-coupled receptor allosterism and complexing, *Pharmacol Rev 54*, 323-374.
[181] Bondos, S. E., Swint-Kruse, L., and Matthews, K. S. (2015) Flexibility and Disorder in Gene Regulation: LacI/GalR and Hox Proteins, *The Journal of biological chemistry 290*, 24669-24677.
[182] Papaleo, E., Saladino, G., Lambrughi, M., Lindorff-Larsen, K., Gervasio, F. L., and Nussinov, R. (2016) The Role of Protein Loops and Linkers in Conformational Dynamics and Allostery, *Chem Rev 116*, 6391-6423.
[183] Eisenstein, E., and Beckett, D. (1999) Dimerization of the Escherichia coli biotin repressor: corepressor function in protein assembly, *Biochemistry 38*, 13077-13084.
[184] Barker, D. F., Kuhn, J., and Campbell, A. M. (1981) Sequence and properties of operator mutations in the bio operon of Escherichia coli, *Gene 13*, 89-102.
[185] Malmstrom, R. D., Kornev, A. P., Taylor, S. S., and Amaro, R. E. (2015) Allostery through the computational microscope: cAMP activation of a canonical signalling domain, *Nature communications 6*, 7588.
[186] Tehver, R., Chen, J., and Thirumalai, D. (2009) Allostery wiring diagrams in the transitions that drive the GroEL reaction cycle, *Journal of molecular biology 387*, 390-406.
[187] Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics, *J Mol Graph 14*, 33-38, 27-38.
[188] Lane, M. D., Rominger, K. L., Young, D. L., and Lynen, F. (1964) The Enzymatic Synthesis of Holotranscarboxylase from Apotranscarboxylase and (+)-Biotin. II. Investigation of the Reaction Mechanism, *The Journal of biological chemistry 239*, 2865-2871.
[189] Kwon, K., and Beckett, D. (2000) Function of a conserved sequence motif in biotin holoenzyme synthetases, *Protein science : a publication of the Protein Society 9*, 1530-1539.
[190] Gill, S. C., and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data, *Anal Biochem 182*, 319-326.
[191] Sigurskjold, B. W. (2000) Exact analysis of competition ligand binding by displacement isothermal titration calorimetry, *Anal Biochem 277*, 260-266.
[192] DeLano, W. L. (2002) The PyMOL molecular graphics system, *http://www.pymol.org*.
[193] Berendsen, H. J. C., Vanderspoel, D., and Vandrunen, R. (1995) Gromacs - a Message-Passing Parallel Molecular-Dynamics Implementation, *Computer Physics Communications 91*, 43-56.

[194] Lindahl, E., Hess, B., and van der Spoel, D. (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis, *J Mol Model 7*, 306-317.
[195] Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. (2005) GROMACS: fast, flexible, and free, *J Comput Chem 26*, 1701-1718.
[196] Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation, *J Chem Theory Comput 4*, 435-447.
[197] Jorgensen, W. L., Maxwell, D. S., and TiradoRives, J. (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, *J Am Chem Soc 118*, 11225-11236.
[198] Lemkul, J. A., Allen, W. J., and Bevan, D. R. (2010) Practical considerations for building GROMOS-compatible small-molecule topologies, *J Chem Inf Model 50*, 2221-2235.
[199] Bussi, G., Donadio, D., and Parrinello, M. (2007) Canonical sampling through velocity rescaling, *The Journal of chemical physics 126*, 014101.
[200] Parrinello, M., and Rahman, A. (1981) Polymorphic transitions in single crystals: A new molecular dynamics method, *Journal of Applied Physics 52*, 7182-7190.
[201] Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997) LINCS: A linear constraint solver for molecular simulations, *Journal of Computational Chemistry 18*, 1463-1472.
[202] Armen, R., Alonso, D. O., and Daggett, V. (2003) The role of α -, 310 -, and π - helix in helix→ coil transitions, *Protein Science 12*, 1145-1157.
[203] Carter, P. J., Winter, G., Wilkinson, A. J., and Fersht, A. R. (1984) The Use of Double Mutants to Detect Structural-Changes in the Active-Site of the Tyrosyl-Transfer Rna-Synthetase (Bacillus-Stearothermophilus), *Cell 38*, 835-840.
[204] Ackers, G. K., and Smith, F. R. (1986) Resolving Pathways of Functional Coupling within Protein Assemblies by Site-Specific Structural Perturbation, *Biophysical Journal 49*, 155-165.
[205] Horovitz, A. (1996) Double-mutant cycles: a powerful tool for analyzing protein structure and function, *Fold Des 1*, R121-126.
[206] Zhao, H., Streaker, E., Pan, W., and Beckett, D. (2007) Protein-protein interactions dominate the assembly thermodynamics of a transcription repression complex, *Biochemistry 46*, 13667-13676.
[207] Brown, P. H., and Beckett, D. (2005) Use of binding enthalpy to drive an allosteric transition, *Biochemistry 44*, 3112-3121.
[208] Kwon, K., Streaker, E. D., Ruparelia, S., and Beckett, D. (2000) Multiple disordered loops function in corepressor-induced dimerization of the biotin repressor, *Journal of molecular biology 304*, 821-833.
[209] Starr, T. N., and Thornton, J. W. (2016) Epistasis in protein evolution, *Protein science : a publication of the Protein Society 25*, 1204-1218.

[210] O'Rourke, K. F., Gorman, S. D., and Boehr, D. D. (2016) Biophysical and computational methods to analyze amino acid interaction networks in proteins, *Computational and structural biotechnology journal 14*, 245-251.
[211] Monod, J. (1971) Chance and Necessity an Essay on the Natural Philosophy of Modern Biology. Translated From the French by Austryn Wainhouse.
[212] Iwata, S., Kamata, K., Yoshida, S., Minowa, T., and Ohta, T. (1994) T and R states in the crystals of bacterial L-lactate dehydrogenase reveal the mechanism for allosteric control, *Nat Struct Biol 1*, 176-185.
[213] JACOB, F., and MONOD, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins, *J Mol Biol 3*, 318-356.
[214] Ampapathi, R. S., Creath, A. L., Lou, D. I., Craft, J. W., Blanke, S. R., and Legge, G. B. (2008) Order-disorder-order transitions mediate the activation of cholera toxin, *J Mol Biol 377*, 748-760.
[215] LANE, M. D., ROMINGER, K. L., YOUNG, D. L., and LYNEN, F. (1964) THE ENZYMATIC SYNTHESIS OF HOLOTRANSCARBOXYLASE FROM APOTRANSCARBOXYLASE AND (+)-BIOTIN. II. INVESTIGATION OF THE REACTION MECHANISM, *J Biol Chem 239*, 2865-2871.
[216] Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature 227*, 680-685.
[217] Bains, G., and Freire, E. (1991) Calorimetric determination of cooperative interactions in high affinity binding processes, *Anal Biochem 192*, 203-206.
[218] Johnson, M. L., Correia, J. J., Yphantis, D. A., and Halvorson, H. R. (1981) Analysis of data from the analytical ultracentrifuge by nonlinear least-squares techniques, *Biophys J 36*, 575-588.
[219] Berendsen, H., van der Spoel, D., and van Drunen, R. (1995) GROMACS - A MESSAGE-PASSING PARALLEL MOLECULAR-DYNAMICS IMPLEMENTATION, *Comp. Phys. Comm. 91*, 43-56.
[220] Lindahl, E., Hess, B., and van der Spoel, D. (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis, *J. Mol. Model. 7*, 306-317.
[221] Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation, *J. Chem.Theory Comput. 4*, 435-447.
[222] Jorgensen, W., Maxwell, D., and TiradoRives, J. (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, *J. Am. Chem. Soc. 118*, 11225-11236.
[223] Berendsen, H. J. C., Grigera, J. R., and Straatsma, T. P. (1987) The Missing Term in Effective Pair Potentials, *J Phys Chem-Us 91*, 6269-6271.
[224] FLOYD, R. (1962) ALGORITHM-97 - SHORTEST PATH, *Communications of the Acm 5*, 345-345.
[225] Stacklies, W., Seifert, C., and Graeter, F. (2011) Implementation of force distribution analysis for molecular dynamics simulations, *Bmc Bioinformatics 12*.

[226] Costescu, B., and Grater, F. (2013) Time-resolved force distribution analysis, *Bmc Biophysics 6*.
[227] Barker, D. F., and Campbell, A. M. (1980) Use of Bio-Lac Fusion Strains to Study Regulation of Biotin Biosynthesis in Escherichia-Coli, *Journal of bacteriology 143*, 789-800.
[228] Baker, B. M., and Murphy, K. P. (1996) Evaluation of linked protonation effects in protein binding reactions using isothermal titration calorimetry, *Biophysical Journal 71*, 2049-2055.
[229] Brown, P. H., Cronan, J. E., Grotli, M., and Beckett, D. (2004) The biotin repressor: modulation of allostery by corepressor analogs, *Journal of molecular biology 337*, 857-869.
[230] Adikaram, P. R., and Beckett, D. (2013) Protein:protein interactions in control of a transcriptional switch, *Journal of molecular biology 425*, 4584-4594.
[231] Eginton, C., and Beckett, D. (2013) A large solvent isotope effect on protein association thermodynamics, *Biochemistry 52*, 6595-6600.
[232] Sturtevant, J. M. (1977) Heat capacity and entropy changes in processes involving proteins, *Proc Natl Acad Sci U S A 74*, 2236-2240.
[233] Cooper, A., Johnson, C. M., Lakey, J. H., and Nöllmann, M. (2001) Heat does not come in different colours: entropy-enthalpy compensation, free energy windows, quantum confinement, pressure perturbation calorimetry, solvation and the multiple causes of heat capacity effects in biomolecular interactions, *Biophys Chem 93*, 215-230.
[234] Saavedra, H. G., Wrabl, J. O., Anderson, J. A., Li, J., and Hilser, V. J. (2018) Dynamic allostery can drive cold adaptation in enzymes, *Nature*.
[235] Barker, D. F., and Campbell, A. M. (1981) The birA gene of *Escherichia coli* encodes a biotin holoenzyme synthetase, *J. Mol. Biol 146*, 451-467.
[236] Barker, D. F., and Campbell, A. M. (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*, *Journal of Molecular Biology 146*, 469-492.
[237] Xu, Y., and 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*, 7354-7360.
[238] Streaker, E. D., Gupta, A., and Beckett, D. (2002) The biotin repressor: thermodynamic coupling of corepressor binding, protein assembly, and sequence-specific DNA binding., *Biochemistry 41*, 14263-14271.
[239] Streaker, E. D., and Beckett, D. (2003) Coupling of protein assembly and DNA binding:biotin repressor dimerization preceeds biotin operator binding., *J. Mol. Biol 325*, 937-948.
[240] Lane, M. D., 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.

[241] Abbott, J., and Beckett, D. (1993) Cooperative binding of the Escherichia coli repressor of biotin biosynthesis to the biotin operator sequence., *Biochemistry 32*, 9649-9656.
[242] Gill, S. C., and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data, *Analytical Biochemistry 182*, 319-326.
[243] Nenortas, E., and Beckett, D. (1996) Purification and characterization of intact and truncated forms of the *Escherichia coli* biotin carboxyl carrier subunit of acetyl-CoA carboxylase, *J. Biol. Chem. 271*, 7559-7567.
[244] Laue, T. M. (1995) Sedimentation equilibrium as thermodynamic tool, *Methods Enzymol 259*, 427-452.
[245] Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, *Mol Syst Biol 7*, 539.
[246] Ingaramo, M., and Beckett, D. (2009) Distinct amino termini of two human HCS isoforms influence biotin acceptor substrate recognition, *The Journal of biological chemistry 284*, 30862-30870.
[247] Kwon, K., Streaker, E. D., Ruparelia, S., and Beckett, D. (2000) Multiple disordered loops function in corepressor-induced dimerization of the biotin repressor., *J. Mol. Biol 304*, 821-833.
[248] Naganathan, S., and Beckett, D. (2007) Nucleation of an allosteric response via ligand-induced loop folding, *Journal of Molecular Biology 373*, 96-111.
[249] Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench, *Bioinformatics 25*, 1189-1191.
[250] Wilson, K. P., Shewchuk, L. M., Brennan, R. G., Otsuka, A. J., and Matthews, B. W. (1992) *Escherichia coli* biotin holoenzyme ynthetase/bio repressor crystal structure delineates the biotin- and DNA- binding domains., *Proc. Natl. Acad. Sci. U.S.A 89*, 9257-9261.
[251] Wood, Z., Weaver L.H., Brown, P.H., Beckett, D., Matthews, B.W. (2006) Co-repressor induced order and biotin repressor dimerization: a case for divergent followed by convergent evolution., *J. Mol. Biol. 357*, 509-523.
[252] Pendini, N. R., Yap, M. Y., Traore, D. A., Polyak, S. W., Cowieson, N. P., Abell, A., Booker, G. W., Wallace, J. C., Wilce, J. A., and Wilce, M. C. (2013) Structural characterization of Staphylococcus aureus biotin protein ligase and interaction partners: an antibiotic target, *Protein science : a publication of the Protein Society 22*, 762-773.
[253] Steuer, R., Kurths, J., Daub, C. O., Weise, J., and Selbig, J. (2002) The mutual information: Detecting and evaluating dependencies between variables, *Bioinformatics 18*, S231-S240.
[254] Suel, G. M., Lockless, S. W., Wall, M. A., and Ranganathan, R. (2003) Evolutionarily conserved networks of residues mediate allosteric communication in proteins, *Nat Struct Biol 10*, 59-69.

[255] Sternisha, S. M., Liu, P., Marshall, A. G., and Miller, B. G. (2018) Mechanistic Origins of Enzyme Activation in Human Glucokinase Variants Associated with Congenital Hyperinsulinism, *Biochemistry 57*, 1632-1639.
[256] Amaro, R. E., Sethi, A., Myers, R. S., Davisson, V. J., and Luthey-Schulten, Z. A. (2007) A network of conserved interactions regulates the allosteric signal in a glutamine amidotransferase, *Biochemistry 46*, 2156-2173.
[257] Boehr, D. D., Schnell, J. R., McElheny, D., Bae, S. H., Duggan, B. M., Benkovic, S. J., Dyson, H. J., and Wright, P. E. (2013) A distal mutation perturbs dynamic amino acid networks in dihydrofolate reductase, *Biochemistry 52*, 4605-4619.