2016

Revisiting Allostery In Lac Repressor

Matthew Alexander Stetz
University of Pennsylvania, mstetz@mail.med.upenn.edu

Recommended Citation

Stetz, Matthew Alexander, "Revisiting Allostery In Lac Repressor" (2016). *Publicly Accessible Penn Dissertations*. 2595.
https://repository.upenn.edu/edissertations/2595

Revisiting Allostery In Lac Repressor

Abstract

Lac repressor (LacI) is an allosterically regulated transcription factor which controls expression of the lac operon in bacteria. LacI consists of a DNA-binding domain (DBD) and regulatory domain (RD), connected by a linker called the “hinge”. Binding of a small molecule inducer to the RD relieves repression through what is presumed to be a series of conformational changes mediated through the hinge. Despite decades of study, our understanding of this allosteric transition remains incomplete—mostly inferred from partial crystal structures and low-resolution scattering studies. In principle, solution-NMR could provide structural and dynamical information unobtainable by X-ray methods. However, due to LacI’s high molecular weight, low solubility, and transient stability, such studies have been limited to the non-allosteric, isolated DBD. Here, we present a solution-NMR study of the changes in structure and dynamics that underlie the allosteric transition of intact LacI. First, an optimized expression system is presented which enables characterization of LacI using NMR methodologies for high molecular weight proteins. Next, alternative NMR data sampling methods are implemented and further extended to overcome the low-solubility and transient stability limitations. Finally, these developments are combined to characterize LacI in each of its functional states. It is shown that the RD but not the DBD of apo LacI exists in an equilibrium between induced and repressed states with exchange occurring on the μs-ms timescale. Inducer binding in the absence of operator mostly quenches exchange but does not result in structural changes in the hinge or DBD. Conformational dynamics detected in the induced state are shown to be localized to a “network” of RD residues previously characterized to be critical for allostery. These dynamics are shown to be quenched in non-allosteric mutants which suggests functional relevance. Operator binding results in globally quenched dynamics and dramatic changes to the structure of the hinge. Inducer binding in the presence of operator results in only minor structural perturbation in the hinge and DBD. However, dynamics are shown to be activated in the RD. These results suggest that conformational dynamics may be critical to the allosteric transition of LacI.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Biochemistry & Molecular Biophysics

First Advisor

A. Joshua Wand

Keywords

Dynamics, NMR, Repressor, Structure

Subject Categories

Biochemistry | Biophysics

REVISITING ALLOSTERY IN LAC REPRESSOR

Matthew Alexander Stetz

A DISSERTATION

in

Biochemistry and Molecular Biophysics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2016

Supervisor of Dissertation

____

A. Joshua Wand

Benjamin Rush Professor of Biochemistry and Biophysics

Graduate Group Chairperson

____

Kim A. Sharp

Associate Professor of Biochemistry and Biophysics

Dissertation Committee

Gregory D. Van Duyne, Ph.D., Jacob Gershon-Cohen Professor of Medical Science (Chair)
Mitchell Lewis, Ph.D., John Morgan Professor of Biomedical Research and Education
Kristen W. Lynch, Ph.D., Professor of Biochemistry and Biophysics
Kim A. Sharp, Ph.D., Associate Professor of Biochemistry and Biophysics
Kenji Murakami, Ph.D., Assistant Professor of Biochemistry and Biophysics
Erik R. P. Zuiderweg, Ph.D., Professor of Biological Chemistry, University of Michigan, Ann Arbor

REVISITING ALLOSTERY IN LAC REPRESSOR

Dedicated to everyone who has ever worked on Lac repressor

ACKNOWLEDGMENT

I would first like to thank my advisor, Prof. A. Joshua Wand for granting me essentially unlimited freedom over the course of my graduate training. His patience, encouragement, and insight were critical to the advancement of this project and my development as an independent scientist. I would also like to acknowledge various members of the Wand lab: Kathy Valentine, Igor Dodevski, Yinan Fu, Christine Jorge, Vignesh Kasinath, Jackwee Lim, Melissa Martinez, John Gledhill, and Nathaniel Nucci. Additional thanks are owed to those who have tried to take this project in new directions: Brian Fuglestad, Zach Belnavis, and Jane Schulte.

I am also grateful for the people outside the Wand lab who helped me over the years: Prof. Mitchell Lewis and Leslie Milk who provided expert advice on Lac repressor. Prof. Garret FitzGerald who gave me my first research job. The various members of the FitzGerald lab who made me want to be a scientist: WenLiang Song, Emanuela Ricciotti, John Lawson, and Susanne Fries. Prof. Casey Londergan at Haverford College and Tony Mancuso at Penn who taught me two very different ways to approach science. Scott Robson at Harvard Medical School who provided early and crucial assistance with the more technical aspects of this work. All of the people in BMB who made things happen: my committee, Ruth Keris, Angie Young, Priya Pidikiti, and the rest of the BMB staff. I also have to acknowledge John Brehm and Lali Pazos for helping me survive two advanced physical chemistry classes.

Special thanks are owed to my friends and family who somehow managed to put up with me throughout grad school: María, Hans, Abdul, Melissa, Jillian, Peter, Jasmine, Yu-San, Joanna, Katie, Jesse, Robbin, Eunkyu, Hana, Eunjung, Youngsam, Megan, Michelle, Eric, Sophie, Ava, Mom, and Dad.

And finally, absolutely none of this work could have been possible without Marie Carter. There are no words to describe how thankful I am for her intellectual contributions to this project and her friendship.

ABSTRACT

REVISITING ALLOSTERY IN LAC REPRESSOR

MATTHEW ALEXANDER STETZ

A. JOSHUA WAND

Lac repressor (LacI) is an allosterically regulated transcription factor which controls expression of the *lac* operon in bacteria. LacI consists of a DNA-binding domain (DBD) and regulatory domain (RD), connected by a linker called the “hinge”. Binding of a small molecule inducer to the RD relieves repression through what is presumed to be a series of conformational changes mediated through the hinge. Despite decades of study, our understanding of this allosteric transition remains incomplete—mostly inferred from partial crystal structures and low-resolution scattering studies. In principle, solution-NMR could provide structural and dynamical information unobtainable by X-ray methods. However, due to LacI’s high molecular weight, low solubility, and transient stability, such studies have been limited to the non-allosteric, isolated DBD. Here, we present a solution-NMR study of the changes in structure and dynamics that underlie the allosteric transition of intact LacI. First, an optimized expression system is presented which enables characterization of LacI using NMR methodologies for high molecular weight proteins. Next, alternative NMR data sampling methods are implemented and further extended to overcome the low-solubility and transient stability limitations. Finally, these developments are combined to characterize LacI in each of its functional states. It is shown that the RD but not the DBD of apo LacI exists in an equilibrium between induced and repressed states with exchange occurring on the μs-ms timescale. Inducer binding in the absence of operator mostly quenches exchange but does not result in structural changes in the hinge or DBD. Conformational dynamics detected in the induced state are shown to be localized to a “network” of RD residues previously characterized to be critical for allostery. These dynamics are shown to be quenched in non-allosteric mutants which suggests functional relevance. Operator binding results in globally quenched dynamics and dramatic changes to the structure of the hinge. Inducer binding in the presence of operator results in only minor structural perturbation in the hinge and DBD. However, dynamics are shown to be activated in the RD. These results suggest that conformational dynamics may be critical to the allosteric transition of LacI.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1: Introduction

1.1 The *lac* operon

In prokaryotes, genes are clustered into discrete functional units called operons. A typical operon consists of a single promoter, one or more structural genes, and various regulatory elements. All structural genes in an operon are expressed or repressed simultaneously. Expression is controlled by the regulatory elements which are often sensitive to environmental stimuli. This sensitivity allows the cell to adjust expression patterns to maximize viability in a particular context and thus ensures an efficient utilization of resources.

The operon was first described by Jacob and Monod to explain the sequential utilization of carbon sources by the bacterium, *Escherichia coli* (*E. coli*) [1, 2]. They observed that when *E. coli* cells were grown in the presence of both glucose and lactose, glucose was preferentially metabolized until its supply was completely exhausted. A critical finding was that lactose metabolism only commenced following a lag period, resulting in biphasic or “diauxic” growth curves (Figure 1.1). The lag period arises due to changes in gene expression levels as cells begin to make the proteins required for lactose transport and catabolism. The genes that encode these proteins are located in the *lac* operon. The *lac* operon has emerged as the paradigmatic example for how genes are regulated in prokaryotes. The basic structure of the operon is illustrated in Figure 1.2a. The structural genes are *lacZ*, *lacY*, and *lacA* [3]. The *lacZ* gene encodes the enzyme β-galactosidase, which cleaves lactose into glucose and galactose. The *lacY* gene encodes β-galactoside permease, which transports lactose into the cell. The *lacA* gene encodes β-galactoside transacetylase which acetylates galactosides but has an unessential role in lactose utilization. The regulatory elements include the promoter, an enhancer-like sequence upstream of the promoter, and an operator sequence downstream of the promoter [4]. Expression of the structural genes is regulated by both glucose and lactose. Glucose levels are sensed by the transcriptional activator, catabolite activator protein (CAP)[1]. CAP binds the enhancer-like sequence and can interact directly with RNA polymerase to facilitate its binding to

---

[1] CAP is also known as cyclic-AMP receptor protein (CRP)

the promoter [5]. CAP can only bind the enhancer-like sequence in the presence of cyclic-AMP (cAMP) and cAMP levels depend inversely on glucose levels [6]. Therefore, the active population of CAP is only high when glucose levels are low. Over 100 different promoters are modulated by CAP so its activity is not specific to the *lac* operon [7].

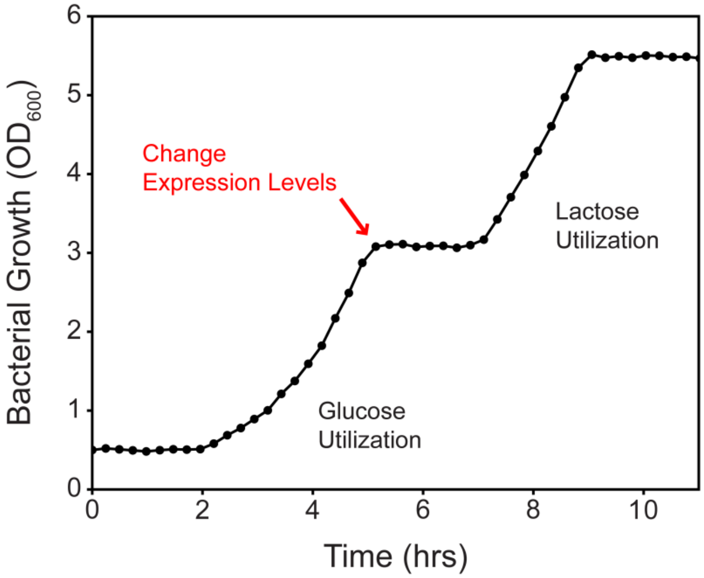


Figure 1.1 Simulated diauxic growth curve for *E. coli* grown in medium containing glucose and lactose. The biphasic pattern arises from sequential utilization of the carbon sources. Adapted from Lewis, 2005 [3].

Lactose levels are sensed by the transcriptional repressor, Lac repressor (LacI). When lactose levels are low, LacI binds to the *lac* operator and physically occludes RNA polymerase from initiating transcription (but does not prevent it from binding the promoter) [8]. The inherent leakiness of the *lac* promoter allows a basal level of β-galactoside permease and β-galactosidase to be present in the absence of lactose. Lactose can then enter the cell via β-galactoside permease and undergo conversion to allolactose by β-galactosidase. Allolactose binds LacI and induces a change such that its affinity for operator drops significantly [9]. Molecules that elicit such an effect in LacI are generically called “inducers.” Allolactose-bound LacI binds non-specific sequences and operator with comparable affinities. Since the concentration of non-specific DNA is much higher than that of operator DNA, allolactose-bound LacI readily dissociates from the operator [10]. This dissociation event relieves the occlusion of RNA polymerase and thus permits transcription initiation.

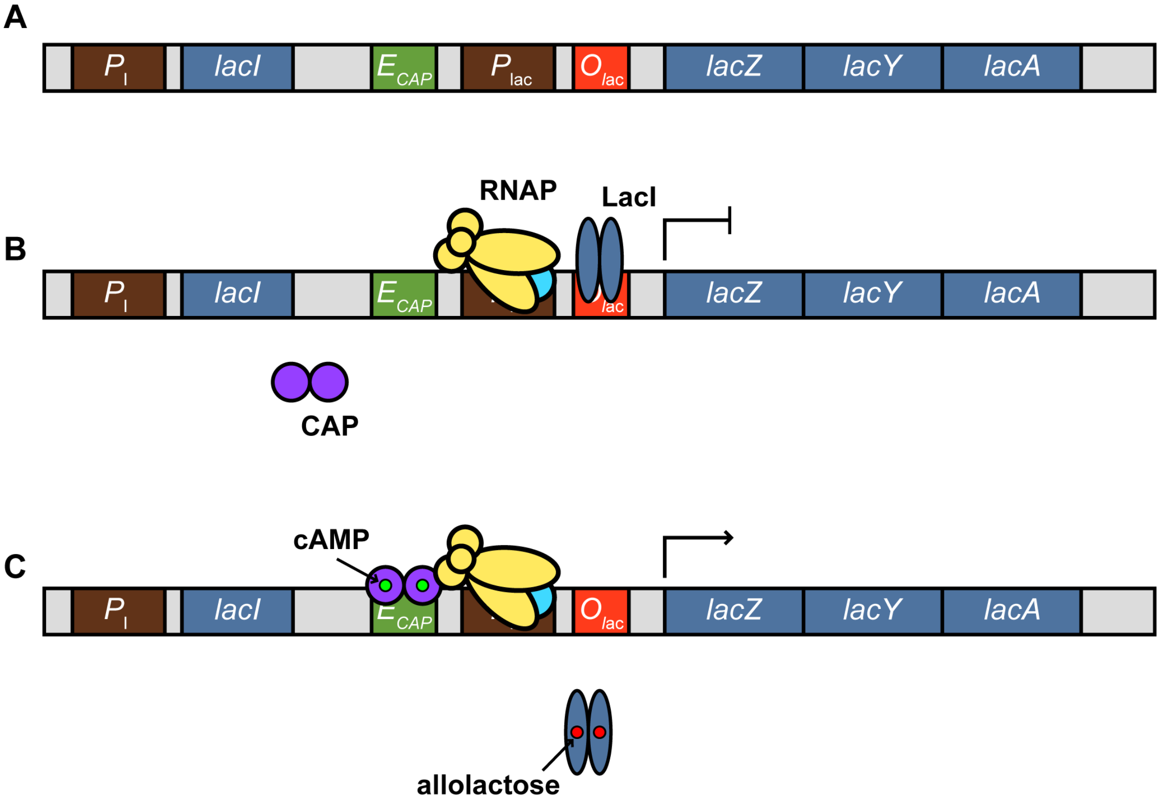


Figure 1.2 Schematic of the *lac* operon. (A) The elements of the operon: P_I is the promoter regulating the gene encoding LacI (*lacI*), E_{cap} is the enhancer-like sequence which CAP binds to, P_{lac} is the promoter regulating the structural genes of the operon. O_{lac} is the operator sequence which LacI binds to, *lacZ* is the gene encoding β-galactosidase, *lacY* is the gene encoding β-galactoside permease, and *lacA* is the gene encoding β-galactoside transacetylase. (B) Protein occupancy when lactose levels are low and glucose levels are high. (C) Protein occupancy when lactose levels are high and glucose levels are low.

Two possible scenarios therefore arise depending on the relative availability of glucose and lactose. In the presence of glucose, CAP is not bound and LacI is bound, resulting in repression (Figure 1.2b). In the absence of glucose and presence of lactose, CAP is bound and LacI is not bound, resulting in transcription (Figure 1.2c). This binary functionality has led some to describe the *lac* operon as a "genetic switch" [3]. The utility of a binary genetic switch is apparent and, as such, the *lac* operon (or its constituent elements) has been appropriated for a wide variety of applications including: recombinant protein expression [11], genetic circuits [12], molecular tethering and localization [13, 14], cost/benefit analysis in evolving populations [15, 16], and even mammalian gene therapy [17, 18]. This ubiquitous utilization suggests that the concept of the genetic switch is well understood. However, the mechanistic details underlying the functionality of the switch have yet to be fully elucidated.

1.2 Functional properties of Lac repressor

The switch-like functionality of the *lac* operon derives largely from the switch-like functionality of LacI. LacI was first isolated in 1966 [19] and was revealed to be a 360 amino acid [20] protein that formed a homotetramer [21]. At the most basic level, LacI has two functions: (1) to bind DNA and (2) to bind a small molecule effector[2]. Understanding these two functions and how they affect one another is central to understanding how LacI functions as a switch.

1.2.1 DNA binding

LacI binds to both operator sequences and to non-specific sequences. The primary *lac* operator (O_1) was first isolated and sequenced by co-purification and DNA-footprinting experiments [22, 23]. The minimal operator is 17 base pairs long and is pseudo-symmetric about a central G:C pair [3]. Two additional operators, O_2 and O_3, were identified based on sequence similarity to the primary operator. The O_2 operator is located in the *lacZ* gene [24] and the O_3 operator is located in the *lacI* gene [25]. These additional operators are not shown in Figure 1.2 for clarity. Of the three natural operators, only O_1 is necessary and sufficient for repression, though the presence of *both* O_2 and O_3 enhances repression by approximately 70 fold [26]. The location of the three natural operators with respect to the transcription initiation site is shown in Figure 1.3a and sequences of their 5' to 3' strands are shown in Figure 1.3b with the G of the central G:C basepair in bold type.

---

[2] In this work, the term "effector" will be used to describe any small molecule that binds specifically to a protein target regardless of its functional effect

The *in vitro* binding affinities for the operators have been measured. As expected, the measured affinities vary depending on the nature and conditions of the assay (particularly the ionic strength). Typical dissociation constants (k_d) for the LacI-O_1 interaction are 0.1 pM at low ionic strength (10 mM KCl) [27] and 30 pM at near physiological ionic strength (150 mM KCl) [28], though one report has argued that these values are overestimates of the true *in vivo* affinity due to the inherent limitations of *in vitro* assays [29]. LacI binds O_2 with 5-31 fold lower affinity than O_1 [4, 24, 30] and binds O_3 with 16-2000 fold lower affinity than O_1 [4, 30, 31].

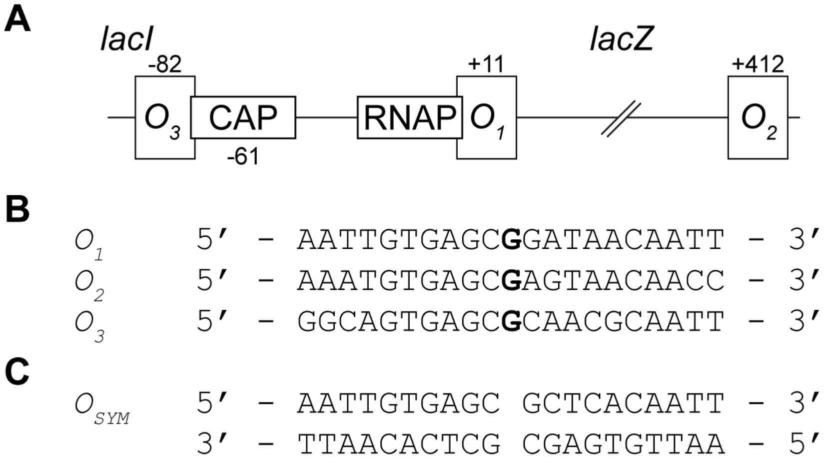


Figure 1.3 Sequences of the *lac* operators. (A) Relative positions of primary and auxiliary operators. The numbering represents the base count relative to the transcription initiation site. (B) Sequences of the primary and auxiliary lac operators. (C) Duplex sequence of idealized symmetric operator. This figure is adapted from Lewis, *et al.*, 1996 [32]

Interestingly, operator mutations that disrupt the interaction with LacI have been shown to increase the asymmetry about the central G:C pair and the most disruptive mutations were found to localize to the left half site [3]. Since the left half site was observed to contribute more to the binding affinity than the right half site, a perfectly symmetric operator was created by removing the central G:C base pair and replacing the right half site with a reflection of the left half site [33]. This symmetric operator (O_{sym}) binds LacI with nearly 10-fold higher affinity relative to O_1 binding.

The sequence of duplexed O_{sym} is shown in Figure 1.3c to illustrate the palindromic nature of operator.

By comparison, LacI binds non-specific sequences with k_d values in the range of 1-40 nm under low ionic strength conditions [34-36] which is approximately 4-5 orders of magnitude weaker than primary operator binding. LacI binding to non-specific DNA is functionally relevant and plays an important role in both operator search/recognition and induction [4, 37-39] but will not be discussed further here.

1.2.2 Effector binding

While LacI senses and responds to lactose levels, lactose itself is not an inducer. The natural inducer, allolactose, is a byproduct of lactose catabolism by β-galactosidase. It has been speculated that the coupling between β-galactosidase activity and induction may have evolved as a way to ensure that upregulation of β-galactosidase only occurs when the enzyme is functional [4].

Allolactose binds LacI with a k_d of 5 μM *in vitro* [9]. However, the interaction between allolactose and LacI is poorly characterized due to the inherent difficulty associated with isolating allolactose. A variety of sugar and synthetic sugar-like molecules which induce expression of the structural genes of the *lac* operon have been identified [40]. One key synthetic molecule, isopropyl β-D-1-thiogalactopyranoside (IPTG), was adopted as the favorite inducer for biochemical studies because it functions similar to allolactose and is not degraded by β-galactosidase. IPTG binds LacI with an affinity of 1-6 μM and induces the lac operon *in vivo* to the same degree as allolactose at millimolar concentrations. However, at micromolar concentrations, allolactose is a more potent inducer of the *lac* operon than IPTG [9]. This observation has been attributed to differences in cellular transport efficiency and thus may not reflect functional differences between the two molecules [9]. The chemical structures of lactose, allolactose, and IPTG are shown in Figure 1.4. It should be noted that while allolactose is a disaccharide, IPTG is a monosaccharide which suggests the overall size of the effector is of modest importance for eliciting a functional effect.

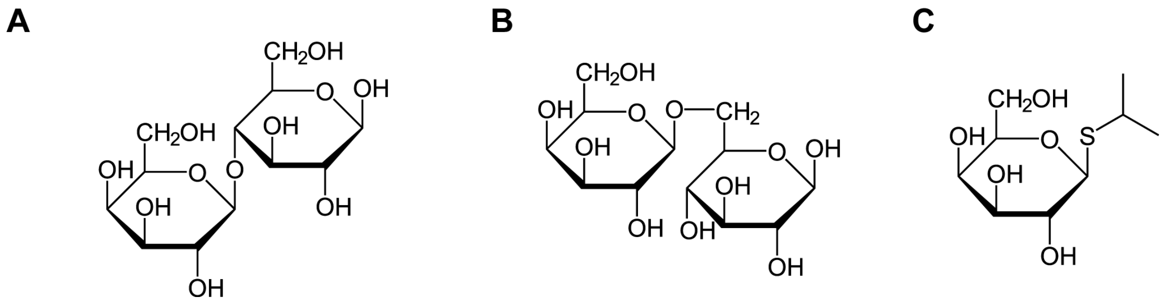


Figure 1.4 Chemical structures of some effectors. (A) Lactose, (B) Allolactose, and (C) IPTG.

Interestingly, a variety of synthetic molecules have been identified which either increase LacI's affinity for operator ("anti-inducers") or bind specifically but produce no functional effect ("neutral ligand" or "neutral agonist"). The most potent anti-inducer identified to date is orthonitrophenyl-β-D-fucoside (ONPF) which binds LacI with a k_d of 160 μM and increases the affinity of the LacI-operator interaction by ~5-fold [41]. Only one neutral agonist has been identified to date, orthonitrophenyl-1-β-D-galactoside (ONPG), which binds LacI with a k_d of 83 μM [41]. While neither ONPF nor ONPG would naturally enter an *E. coli* cell, they have been employed as tools for furthering our understanding of the relationship between effector binding and DNA binding.

1.2.3 The relationship between effector binding and DNA binding

The relationship between inducer binding and O_1 binding has been studied almost exclusively using IPTG as the inducer. Equilibrium dialysis measurements demonstrate that IPTG binding decreases LacI's affinity for O_1 by at least 1000 fold in low ionic strength conditions [42] though nitrocellulose-filter binding measurements at physiological ionic strength report changes in affinity of over 10000 fold [43]. The latter measurement is more consistent with data that shows induced LacI binds operator and non-specific sequences with similar affinity [43]. Importantly, a reciprocal relationship between operator and inducer binding has been demonstrated by fluorescence titrations which showed LacI pre-bound to O_1 binds IPTG with 20-fold lower affinity than free LacI [44, 45]. This suggests inducer and operator binding are linked by a thermodynamic cycle [4]. Curiously, a similar reciprocity has not been demonstrated universally for anti-inducers, with some binding free LacI more tightly and others binding the LacI-operator complex more tightly [46].

Inducer binding also decreases LacI's affinity for other operators. The affinity of the LacI-O_{sym} interaction is diminished by over 10000 fold upon IPTG binding [43]. While no k_d values have been published for the inducer-bound LacI-O_2 or inducer-bound LacI-O_3 interactions, it has been observed that IPTG binding affects the LacI-O_1, LacI-O_2, and LacI-O_3 interactions similarly suggesting a change in affinity by at least several orders of magnitude [47]. Interestingly, the interaction between LacI and non-specific DNA is not affected by inducer binding [36, 43].

In order to understand the determinants of DNA binding, effector binding, and their relationship, the specific parts of LacI required for each function must be delineated. Early insight came from limited proteolysis studies [48-51]. Trypsin digests of LacI yield two fragments: a larger, tetrameric domain capable of inducer binding and four smaller pieces which retain low (mid-micromolar) affinity for operator DNA. These results suggested that different domains were responsible for DNA binding and effector binding. The proteolysis findings were buttressed by genetic studies of missense and nonsense mutations [52-62]. Substitutions in the N-terminal region of the protein primarily disrupt DNA binding [63]. Substitutions in the central region of LacI were largely tolerable but some mutations were shown to disrupt effector binding [64] and oligomerization [65]. Substitutions in the C-terminal region were shown to disrupt oligomerization with little affect on DNA-binding, effector-binding, or inducibility [28, 66].

The limited proteolysis and genetic studies suggested that the relationship between DNA binding and effector binding was allosteric in nature—involving long-range communication between two discrete domains. However, in the absence of high-resolution three dimensional structures, the mechanistic determinants of DNA binding, effector binding, and allosteric communication remained obscured.

1.3 Structural properties of Lac repressor

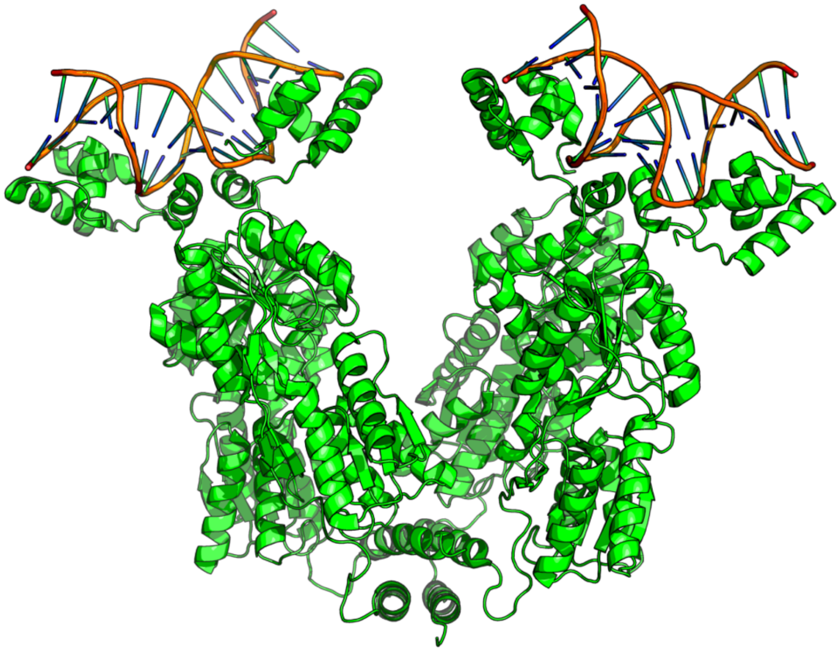


Figure 1.5 Crystal structure of tetrameric Lac repressor bound to operator, O_{sym}. PDB code: 1LBG.

1.3.1 Overall structure

Initial attempts to characterize the three dimensional structure of LacI employed solution scattering methods [67-73]. These studies suggested LacI has an elongated structure with DNA-binding domains located at the ends. No changes in the radius of gyration of LacI could be detected upon IPTG binding. However, the lack of high-resolution structural models precluded detailed interpretation of the scattering data. The complete crystal structure of LacI was not solved until 1996. Structures of LacI in the operator-bound state, inducer-bound state, and apo state were determined to modest/low resolution: 4.8 Å, 3.2 Å, and 4.8 Å, respectively [32]. The operator used was O_{sym} and the inducer molecule used was IPTG. The crystal structure of the LacI tetramer bound to operator is shown in Figure 1.5. Each LacI monomer consists of a DNA-binding domain (DBD, residues 1-45), a regulatory domain (RD, residues 63-329) which includes the dimerization interface and effector binding pocket, and a tetramerization domain (residues 338-358). The DBD and RD are connected by a linker region commonly referred to as the “hinge” (residues 46-62). Curiously, the tetramer is arranged as a V-shaped “dimer of dimers” where one dimer binds one operator sequence. For all practical purposes, the dimer is therefore the minimal functional unit. The DBD consists of a classical helix-turn-helix motif which binds the operator in the major groove. Notably, the center of the hinge (residues 50-58) folds into a helix which intercalates into the minor groove of the operator, resulting in a kinked DNA structure. The RD exhibits a fold similar to bacterial periplasmic sugar binding proteins with the effector-binding pocket located roughly in the middle of the domain. The effector-binding pocket is positioned approximately 40 Å from the DBD, confirming that LacI is allosterically regulated. The tetramerization domain is a helix which associates into a four-helix bundle upon oligomerization. The domain organization is illustrated in Figure 1.6.

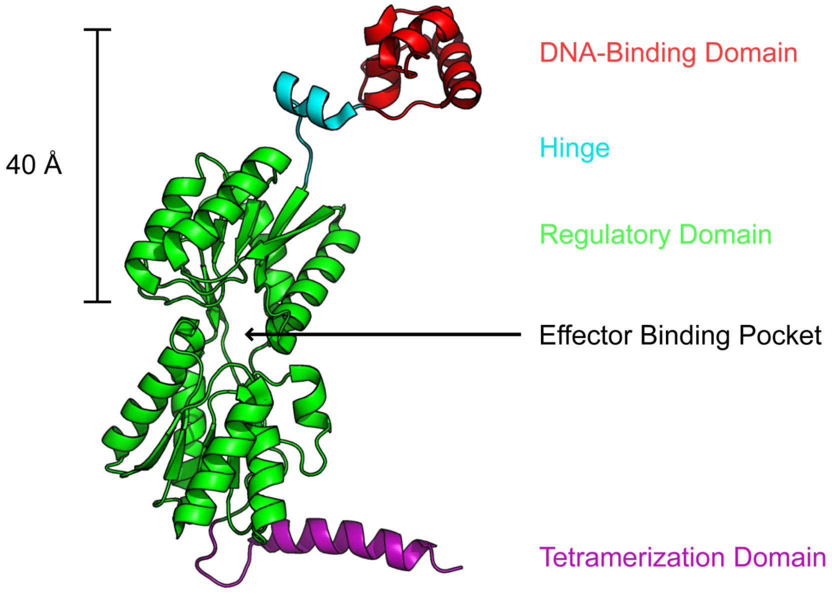


Figure 1.6 The domain architecture of Lac repressor. An isolated monomer from the structure of the tetramer bound to O_{sym} is shown. Each domain is uniquely colored. The structure of the operator is not shown for clarity. PDB code: 1LBG.

The structures of LacI bound to inducer and apo LacI were determined in the absence of operator DNA. Both structures lacked electron density for the hinge and the DBD, suggesting a high degree of mobility. They were similar to previous structures of the isolated RD which had been obtained via proteolytic digestion of LacI [74]. Moreover, the induced and apo states were highly superimposable, suggesting that the apo state may be biased toward the induced conformation. However, it was noted that this may have been because glycerol occupied the inducer binding pocket of the apo state, serving as an effector [75].

1.3.2 DNA binding

The modest/low resolution of the operator-bound structure of LacI prevented site-resolved delineation of the key contacts necessary for operator binding. High resolution insight into the structure of the DBD in the absence of DNA was first provided by nuclear magnetic resonance (NMR) studies of the isolated DBD with a truncated hinge [76-81]. In the absence of DNA, it was shown that a portion of the hinge region was unfolded, potentially explaining the lack of electron density for the DBD and hinge in the crystal structures obtained without operator [82]. Further NMR-based characterization on DBD constructs with intact hinge regions and the eventual determination of a 2 DBD: 1 O_{sym} complex structure revealed that the hinge helices fold upon binding operator, suggesting that the hinge is critical for operator recognition [82, 83]. While this structure was entirely consistent with the LacI-O_{sym} crystal structure, it provided an atomic-level description of the specific contacts made between the DBD and operator. Despite the monomeric nature of the free DBD, the 2 DBD:1 O_{sym} complex was observed to be highly stable and several key protein-protein interactions between the two DBDs were identified. These interactions were localized to the hinge and specifically involved the sidechains of V52, A53, and L56. LacI interactions with operator were shown to consist of both apolar and polar contacts. Key interactions with the major groove of the operator were found to involve Y17 and Q18. Key interactions with the phosphate backbone were found to involve S21, R22, and N25. Residues

Y17, Q18, and R22 had historically been identified by mutagenesis studies to be essential for operator recognition [84, 85]. It should be noted, however, that exhaustive combinatorial mutagenesis studies suggests there are no fundamental rules governing *how* amino acids at these positions recognize specific DNA sequences [85]. Curiously, the hinge helix exhibited numerous apolar protein-DNA interactions. The most notable of these interactions involves L56, which intercalates deeply into the minor groove, consistent with the low resolution crystal structure of LacI bound to O_{sym}.

In order to obtain a higher resolution crystal structure, a dimeric construct of LacI bound to ONPF and O_{sym} was crystallized and the structure was solved to 2.6 Å [86]. This high resolution structure agreed very well with the NMR structure of the 2 DBD: 1 O_{sym} complex (all atom RMS < 1 Å). Though the isolated DBD used in the NMR study binds operator with substantially lower affinity than the intact protein, this difference in affinity is not expected to manifest in detectable structural differences since displacements of only a fraction of an angstrom are required for dramatic changes in the energetic value of a single interaction.

Later NMR studies of the DBD utilized an engineered dimeric construct. A single cysteine was introduced at position 52 (V52C, center of the hinge helix) in order to cross link two DBD molecules via disulfide bond formation [87]. By dimerizing two DBD molecules, high affinity operator binding could be restored. The advantage to this approach was that the 2 DBD:1 O_1 complex could be studied at high resolution which was previously impossible due to the intrinsic dynamics of the DBD and the low affinity nature of the interaction [88]. The solution NMR structure of the 2 DBD: 1 O_1 complex showed that the protein-DNA interactions were asymmetric. The monomers were rotated 48° relative to each other to accommodate the asymmetry in the operator. In this way, differences in the protein-DNA contacts for each half site are minimized. This was in stark contrast to the 4.0 Å crystal structure of a dimeric construct of LacI bound to O_1 which was highly similar to the structure of a dimeric construct of LacI bound to O_{sym} (all atom RMS of 0.4 Å) [89]. The LacI-O_1 structure suggests that the structure of LacI does not rearrange to accommodate the asymmetry of the operator but rather the protein-DNA interactions are shifted as a consequence of the asymmetry [3]. Aligning the DBD from the crystal structure and the NMR structure yields an all atom RMS value of > 3 Å and large deviations are found with respect to the placement of the hinge helices. While the differences between the solution NMR structure of the 2 DBD: 1 O_1 complex and the crystal structure of the dimeric LacI-O_1 complex have not been attributed to any one single explanation, they may originate from restriction in DBD mobility that is imposed by the RD.

The structure of the engineered, dimeric DBD bound to a non-specific DNA sequence revealed that the hinge helix was unfolded, consistent with the proposed role that the hinge helix recognizes the operator specifically [90]. As a result, the DNA is not kinked and remains in canonical B-form. Importantly, it was demonstrated that all apolar interactions previously observed to be critical for operator binding were lost in the non-specific complex. The interaction between the DBD and non-specific DNA is almost entirely electrostatic in nature, relying heavily on charge-charge contacts between R22, H29, K33, R35, and K37 and the phosphate backbone. Of these interactions only K33 and H29 made contacts with the phosphate backbone of O_1 [91]. While these solution NMR studies provide key details about the residue-specific contacts mediating the protein-DNA interaction, the use of a V52C mutant as a model is questionable. Making the same mutation in full length LacI increases operator affinity substantially, diminishes sensitivity to alterations in operator sequence, and abolishes allostery, suggesting the disulfide bond may force the protein into a non-native, “super” repressed conformation [92, 93].

1.3.3 Effector binding

A high resolution description of the structural basis for effector binding was not obtained until 2007, roughly 10 years after the original crystal structures of LacI were presented [75]. The structure of dimeric LacI bound to IPTG revealed that all of the oxygen atoms in the sugar ring of IPTG are hydrogen bonded. Key interactions are made with R197, N246, D274, and A75.The isopropyl group of IPTG also makes a series of apolar contacts with I79, F161, F293, and Leu296. Notably, the O6 hydroxyl moiety of IPTG, a chemical group absent in anti-inducers, facilitates the formation of a rich network of water-mediated hydrogen bonds that crosslink the terminal portions of the RD. The formation of this hydrogen bond network stabilizes the induced conformation and prevents a transition back to the repressed conformation [75]. Importantly, this hydrogen bonding network was not present in the crystal structures of the LacI-ONPF-O_{sym} ternary complex or the LacI-ONPF binary complex [75, 86]. Both structures revealed that ONPF recognizes and interacts with R197, N246, and D274 like IPTG, but exhibits different hydrogen bonding patterns around C6.

Unlike ONPF, the neutral ligand ONPG possesses an O6 hydroxyl group. However, the structure of dimeric LacI bound to ONPG reveals that the water-mediated hydrogen bond network is not formed. ONPG was shown to interact with R197, N246, and D274 like the other effectors but diffuse electron density for its nitrophenyl ring suggested it was not positioned optimally in the effector binding pocket. As a consequence, the O6 hydroxyl was displaced from the orientation necessary to facilitate formation of the water-mediated hydrogen bonding network [75].

1.3.4 The allosteric transition

While it had been known for some time that operator binding and inducer binding were mediated by two discrete domains, the high resolution structure of LacI provided the first mechanistic description for how these two functions are related. Unfortunately, the lack of a complete structure for the induced state precluded a full delineation of the allosteric mechanism. Comparison of the RD in the repressed and induced states shows that the allosteric signal is disseminated through the monomer-monomer interface. The repressed RD and induced RD have nearly identical secondary structures, however they exhibit significant differences in orientation. In the repressed state, the N-terminal portion above the effector-binding pocket (N-terminal subdomain) of the RD is slightly less compact. Upon inducer binding, this region rotates 10° relative to the C-terminal portion below the effector-binding pocket (C-terminal subdomain) of the RD, which remains static during the transition [3, 32]. This conformation is stabilized by the water-mediated hydrogen bond network that was shown to depend greatly on the O6 hydroxyl group of IPTG [75]. An overlay of the repressed and induced states is shown in Figure 1.7.

Since electron density is missing for the hinge and DBD in the induced state structure, the effect of this rotation on protein-DNA interactions is unknown. However, it is plausible that rotation of the N-terminal subdomain of the RD results in displacement of the hinge helices which, in turn, results in diminished operator affinity. The solution NMR structure of the DBD bound to non-specific DNA showed that the hinge is unfolded, which suggests that drastic displacement of the hinge helix from the minor groove would potentially result in local unfolding and thus complete obliteration of key protein-DNA interactions. This hypothesis, however, is inconsistent with biochemical studies of a construct of LacI where the hinge was artificially displaced from the RD using a glycine spacer [43]. The insertion of a glycine spacer between the hinge and the RD does not disrupt allostery. While this observation could be attributed to a variety of explanations, one possibility is that inducer binding functions by simply destabilizing the repressed conformation as opposed to a dramatic structural perturbation [3].

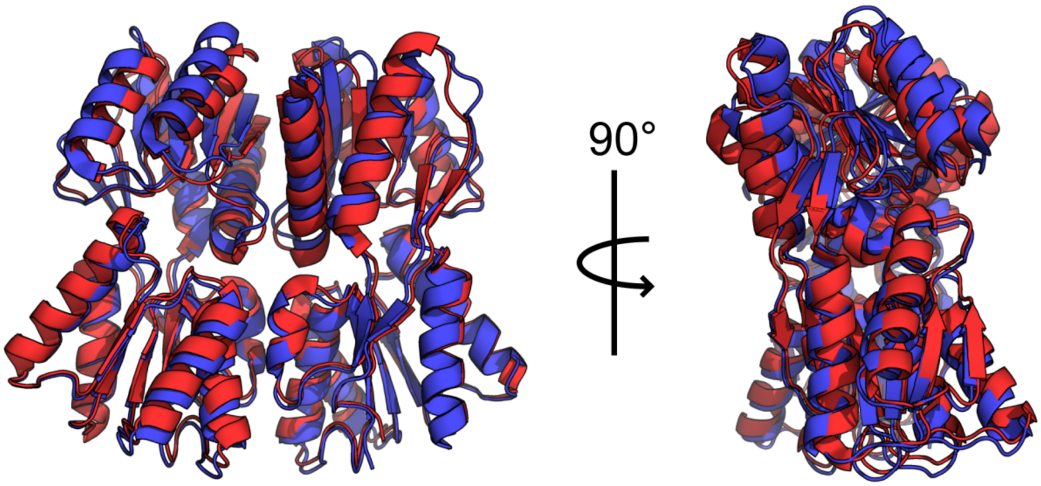


Figure 1.7 Structural basis for the allosteric transition in Lac repressor. Overlay of the structure of the repressed state (blue, PDB code: 1EFA) and the induced state (red, PDB code: 2P9H). Note the rotation of the induced-state structure relative to the repressed-state structure and the preservation of structural symmetry. Effectors are not shown for clarity.

A computational targeted molecular dynamics (TMD) study attempted to elucidate the residue-specific conformational changes involved in the allosteric transition [94]. The starting structure was the RD of the repressed state and the target (final) structure was the RD of the induced state. This study demonstrated that the conformational changes underlying the transition were asymmetric, meaning that each monomer undergoes its own discrete series of changes. Three allosteric pathways were identified: Pathway 1 begins at D149 in the effector-binding pocket of one monomer and conformational changes are propagated across to the other monomer, terminating at K84. Pathway 2 begins at K84 and involves systematic disruption of monomer-monomer contacts, specifically the loss of hydrogen bonding contacts involving Q78 and H74. Pathway 3 involves the propagation of conformational changes from F161 of the effector binding pocket of one monomer to H74 of the other monomer. The resulting conformational changes produce the rotated N-terminal subdomains, characteristic of the induced state. While these findings suggested a possible mechanism for the transition from the repressed state to induced state, they lent little insight into how the DBD could potentially be affected during induction. The details of this study are discussed further in Chapter 4.

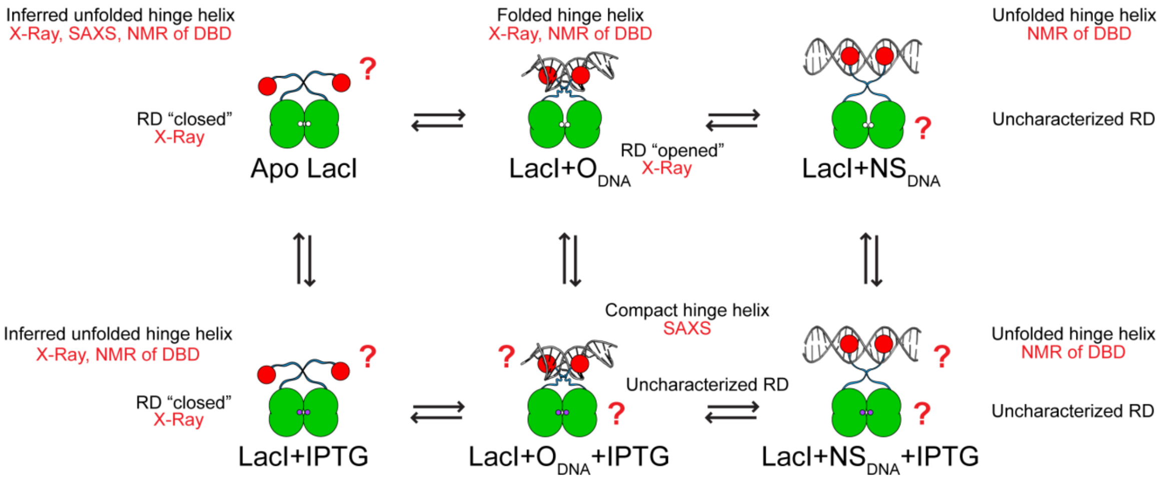


Figure 1.8 Delineation of the functional cycle of Lac repressor and cartoon representation of relevant structural details. O_{DNA} is "operator DNA" and NS_{DNA} is "non-specific DNA." IPTG is shown as the inducer because it has been used in the structural studies. Only limiting[3] states are shown. The transitions between apo lac and non-specific DNA are possible but are not shown for

---

[3] A limiting state is defined here as one in which all subunits are saturated with ligand.

clarity. “RD opened” describes the structure of the RD of the repressed state and “RD closed” describes the structure of the RD in the induced state. Opened and closed are used loosely here to describe the rotation of the N-terminal portion of the RD relative to the C-terminal portion of the RD that occurs upon IPTG binding (see Figure 1.7). Structural details are written in black text and the experimental methods used to obtain those details are written in red text. Inferred or missing structural information is marked with a red question mark.

Perhaps the most informative study to date addressing how the DBD is perturbed upon inducer binding is a SAXS study of tetrameric and dimeric constructs of LacI [95]. Several functional states were characterized including: apo LacI, the LacI-O_1 complex, the LacI-O_1-IPTG complex, and the LacI-non-specific DNA complex. Unfortunately, no data was provided for the LacI-IPTG complex and the LacI-non-specific DNA complex was under-titrated, which precluded analysis. Two major observations were reported. First, LacI was shown to elongate ~20 Å upon operator removal. While not definitive, these data are consistent with unfolding of the hinge helix. Second, IPTG binding to the LacI-O_1 complex did not result in extensive elongation of the molecule, suggesting the hinge helices may remain folded or at least in a compact conformation in the LacI-O_1-IPTG ternary complex. However, given the intrinsically low resolution of the SAXS technique, no mechanistic description of induction was provided.

While the history of structural studies of LacI is extensive, no single set of studies has provided a complete model for how LacI transitions from the repressed state to the induced state. In other words, the switch-like functionality of LacI is still incompletely explained. However, by synthesizing all of the various experimental and computational studies, one can derive a few potential models. The functional cycle of LacI with respect to transitions among limiting states can be described by the simplified diagram in Figure 1.8. The transitions between apo lac and non-specific DNA are, of course, possible but are not shown in Figure 1.8 for clarity. A cartoon representation of the structural changes observed (as well as those inferred) for each transition and the experiments used to characterize those transitions is also shown. The key structural transitions that have been identified to date are the rotation of the RD during the transition from the repressed state to the induced state and the unfolding of the hinge helices upon operator dissociation. It is unclear how these two processes are coupled to one another. An intuitive model is that inducer-binding immediately triggers unfolding of the hinge helix, resulting in the disruption of protein-DNA contacts. This model conveniently explains the dramatic drop in affinity upon induction but is not fully supported (though nor is it fully negated) by SAXS data [95]. Another possible model is that inducer-binding merely results in destabilization of the repressed state by small-scale alterations in the orientation and/or other structural properties of the RD, hinge helices, and DBDs. This would presumably entail some kind of distortion of the protein-DNA interactions such that the binding was no longer optimal. Such a model is consistent with the SAXS data but may not adequately explain the dramatic drop in operator affinity upon induction.

1.4 Classical allostery

The structure of LacI revealed that its function could potentially be explained using the classical allosteric model of Monod, Wyman, and Changeux (MWC)[4]. The MWC model describes the *concerted* transition between two functional states of an allosteric protein. This model assumes that the protein of interest is a symmetric oligomer of multiple, identical subunits[5] and that allosteric effects are correlated with changes in quaternary structure. Additionally, each protomer must possess its own ligand-binding site. The central premise of the model is that at least two states are reversibly accessible by the protein and that each state has a different ligand binding affinity. In the absence of ligand, the protein exists in an equilibrium among all accessible states. Transitions among states must preserve the symmetry of the oligomer. Put in simplistic terms, this means that each protomer is in the same conformational state at all times. Binding of a ligand to the higher affinity state shifts the equilibrium in favor of the higher affinity state, increasing its relative population. By shifting the equilibrium, one dominant conformational state is selected and hence one dominant function is observed.

Structural studies show that LacI meets all of these fundamental criteria with the possible exception of the requirement for maintaining structural symmetry during the transition [94] (though

---

[4] It should be noted that alternative classical models of allostery exist, most famously the Koshland, Némethy, and Filmer (KNF) model in which there are no restrictions on structural symmetry [96]. This model has not been applied to LacI historically so it will not be discussed further.

[5] In the original paper, Monod *et al.* designate functional subunits as “protomers” whereas the word “monomer” is used to describe an individual subunit that has fully dissociated from the oligomer

this has yet to be further verified). LacI was shown to exist in two states: a repressed conformation which binds operator with high affinity and inducer with low affinity and an induced conformation which binds operator with low affinity and inducer with high affinity. The transition between these two states involved a rotation of the protomers relative to one another, *i.e.* a change in quaternary structure. The work of the Lewis and Sharp labs has cast the transition between these two states in terms of the MWC model [97]. In their analysis, the repressed state was designated as R and the induced state was designated as R*. The apo repressor is assumed to exist in an equilibrium between these two states:

$$R \xrightleftharpoons{K_{R^*R}} R^* \tag{1.1}$$

Where K_{R^*R} is the equilibrium constant:

$$K_{R^*R} = \frac{[R^*]}{[R]} \tag{1.2}$$

Both R and R* bind inducer (I) according to:

$$R + I \xrightarrow{K_{RI}} RI \tag{1.3}$$

$$R^* + I \xrightarrow{K_{R^*I}} R^*I \tag{1.4}$$

With equilibrium constants**:**

$$K_{RI} = \frac{[RI]}{[R][I]} \tag{1.5}$$

$$K_{R^*I} = \frac{[R^*I]}{[R^*][I]} \tag{1.6}$$

Equilibrium constants for operator binding can be derived in similar fashion but will not be discussed in what follows. An elegant study utilizing heterodimeric LacI molecules which exhibited either zero, one, or two functional inducer binding pockets and an *in vivo* induction assay based on a GFP reporter determined the value of K_{R^*R} to be 2 ± 0.5 and the ratio K_{R^*I}/K_{RI} to be 15 ± 3 [97]. The value for K_{R^*R} suggests the apo repressor populates mostly the induced state with the free energy difference between states being slightly less than thermal energy (0.4 kcal/mol) at 25˚C. The value of K_{R^*I}/K_{RI} shows that R* binds inducer more favorably than R by about 1.6 kcal/mol at 25˚C. In the presence of one inducer, K_{R^*R} therefore increases to ~ 30 ($K_{RR^*} = e^{(0.4+1.6)/k\beta T} = 29$ at 25˚C). Hence the equilibrium is shifted more in favor of R* by 15 fold, resulting in the reduction of the population of R. Binding a second inducer molecule shifts the value of K_{R^*R} dramatically to 440 ($K_{RR^*} = e^{(0.4+1.6+1.6)/k\beta T} = 436$ at 25˚C) at which point maximal induction levels are achieved. Two inducer molecules are therefore required to achieve full induction. Interestingly, there is essentially no cooperativity (Hill coefficient ~ 1.0) with respect to inducer binding in the absence of operator and only slight positive cooperativity (Hill coefficient ~1.5) when in the presence of operator [98].

The large value of the ratio K_{R^*I}/K_{IR} is significant since it explains how LacI functions as a switch. If this value were equal to unity, inducer would bind to both R and R* with equal probability ($\Delta G = 0$) which would render expression levels insensitive to inducer since no shift in the equilibrium would be observed (in other words, the value of K_{RR^*} does not change upon inducer binding). If the value of K_{R^*I}/K_{IR} were less than unity, then inducer would bind the R state more favorably, resulting in an increase of the population of R and thus increased levels of repression. Of course, this analysis assumes that R binds operator with higher affinity than R* as is the case with LacI.

The thermodynamic analysis above definitively explains how the allosteric functionality of LacI can be described by the MWC model and provides quantitative values for the important equilibrium constants which govern the inducer-triggered shift in populations of the R and R* states. However, a thorough mechanistic description must also include a kinetic analysis. Two distinct mechanisms have been proposed: “induced dissociation” and “conformational selection” [99]. The induced dissociation mechanism requires that inducer-binding triggers the transition from R to R*. The conformational selection mechanism assumes that R and R* are in a dynamic equilibrium and inducer “traps” the R* state. Comprehensive analysis of experimentally determined kinetic data has shown that both mechanisms are involved [99]. The preference for one mechanism over the other is determined by inducer concentration, consistent with studies of other systems [100]. At low inducer concentration, both mechanisms are active whereas the induced dissociation mechanism dominates at high inducer concentration. This is consistent with the requirement that the *lac* operon be leaky in order to facilitate the generation of allolactose in the absence of high concentrations of lactose. Conformational selection therefore establishes a basal amount of β-galactosidase and β-galactoside permease to ensure allolactose is available to induce the switch.

1.5 Non-Classical allostery

The modeling in Section 1.4 provides a quantitative understanding of the transitions between R and R* and how shifting their relative populations can result in repression or induction. However, since the structural descriptions of R and R* are incomplete, it remains difficult to explain the alterations in structural properties that underlie the transitions between R and R*. It is known from X-ray crystallography that the RD of LacI can sample two structurally distinct (albeit modestly distinct) states, which is entirely consistent with the two state modeling above. But what about the structures of the DBD and hinge? Do R and R* exhibit structurally distinct DBDs and hinges?

Suppose, however, that there is little to no change in the average structures of the DBD and hinge region. Would this invalidate our understanding of allostery in LacI? It has been argued that allostery *absolutely* requires a conformational change [101]. This is perhaps unsurprising since classical biochemistry has rationalized the switch-like transition of allosteric proteins in terms of transitions between discrete structural states. Recent experimental studies, however, have shown that allosteric behavior may not require a conformational change if it arises from the modulation of protein dynamics—the time-dependent fluctuations in conformation (motions) about the average structure which are usually not reflected in crystal structures. It has long been recognized that proteins undergo conformational fluctuations of varying magnitude over a wide range of times scales [102]. Furthermore, the nature of these fluctuations has long been known to be sensitive to ligand binding [103]. But what is the relationship between protein dynamics and allostery? What types of motions (timescales) are important for allosteric processes?

1.5.1 Slow motions (μs-ms dynamics) and allostery

Many studies investigating the connection between protein dynamics and allostery have emphasized the contribution of μs-ms timescale dynamics (which will subsequently be referred to as "slow"[6] timescale dynamics) of the protein backbone as these motions have been argued to be an appropriate readout for conformational changes. The vast majority of these studies have employed solution NMR which can report on slow timescale dynamics at residue-level resolution (methods reviewed in [104]).

A landmark study by Kern and co-workers demonstrated that slow timescale dynamics are detected in regions known to undergo conformational changes upon allosteric activation—even in the absence of the allosteric stimulus [105]. They characterized the slow timescale dynamics of a small signaling protein, nitrogen regulatory protein C (NtrC), in the inactive form (unphosphorylated) and the active form (phosphorylated). The major finding was that the inactive state exhibits slow timescale dynamics localized to the exact area where structural differences between the inactive and active state were previously observed. These dynamics were then observed to be quenched in the fully activated state. Ostensibly, this study was nothing more than an experimental confirmation of the MWC model. However, what is unusual about this study is that NtrC is not a symmetric oligomer but rather a single domain monomer. Moreover, the study lends some insight into how transitions between states might occur: extensive slow timescale dynamics in the inactive state reflect transient sampling of the active state, thus priming the protein for its transition. Since the publication of the NtrC study, many other allosteric systems have been shown to exhibit similar behavior (reviewed in [106]).

Slow dynamics have also been implicated in alternative allosteric behavior. A recent study of the enzyme imidazole glycerol phosphate synthase (IGPS) revealed that binding

---

[6] Dynamics that occur on the μs-ms timescale dynamics are sometimes referred to as "intermediate timescale" rather than "slow timescale" as a means of differentiation from dynamics that are slower than ms.

activating allosteric effectors *enhances* slow timescale dynamics rather than quenches them. In the absence of effectors, IGPS is inactive and does not exhibit slow timescale motions. Upon binding allosteric effectors, motions are detected in the active site, nearly 30 Å away from the effector binding site. These motions are thought to break an auto-inhibitory hydrogen bond in the active site which facilitates activation of the enzyme [107].

The NtrC and IGPS studies are just two of many examples of the way μs-ms timescale dynamics can be modulated in allosteric systems (for a recent review of other systems see [108]). It is clear that while the framework provided by the classical models of allostery are useful for facilitating a conceptual understanding, the detailed mechanistic aspects of allosteric transitions is complex and variable.

1.5.2 Fast motions (ps-ns dynamics) and allostery

Another timescale of motion that has been implicated in mediating allosteric processes occurs over the range of ps-ns (which will subsequently be referred to as to as "fast" timescale dynamics). Unlike slow timescale dynamics, which have been treated essentially as a readout for conformational changes, fast timescale dynamics contribute directly to the thermodynamics of allosteric processes [109, 110].

Before delineating the exact relationship between fast timescale dynamics and allostery, the general relationship between these motions and thermodynamics should be clarified. The fast timescale dynamics of proteins has been shown to represent the residual conformational entropy of the folded state [111, 112]. In other words, folded proteins exhibit conformational fluctuations that occur on the ps-ns timescale which contribute to an internal entropy called conformational entropy[7]. These motions can be measured by NMR using classical spin relaxation techniques and related to conformational entropy using a variety of formalisms (see Appendix 1 for more details). Studies of the calcium-sensing signaling protein, calmodulin (CaM), show that the residual conformational entropy of proteins can be substantial and is largely manifest in the

---

[7] Conformational entropy is sometimes referred to as "configurational entropy." This has led it to be conflated mistakenly with the more familiar configurational entropy of proteins which describes the entropy loss upon folding.

motional properties of sidechains [113-116]. The backbone has largely been shown to be a rigid scaffold which exhibits only low amplitude motions on the fast timescale. In the case of CaM, the change in conformational entropy upon ligand binding is large enough to dominate the change in the free energy of binding [115, 116]. This has also been observed for CAP's (the same protein introduced in Section 1.1) interaction with DNA [117].

Like ligand binding, an allosteric transition can be described thermodynamically by the Gibbs-Helmholtz free energy change, (ΔG):

$$\Delta G = \Delta H - T\Delta S \tag{1.7}$$

where ΔH represents the change in enthalpy, T is the temperature, and ΔS is the change in entropy. In the context of allosteric transitions ΔH represents changes in the structure of the protein and ΔS represents a linear combination of a variety of contributions:

$$\Delta S = \Delta S_{solvent} + \Delta S_{conformational} + \Delta S_{RT} \tag{1.8}$$

Where $\Delta S_{\text{solvent}}$ represents the change in solvent entropy which is commonly described in terms of the hydrophobic effect, $\Delta S_{\text{conformational}}$ is the conformational entropy, and ΔS_{RT} which is the change in the entropy of the rotational and translational degrees of freedom and has been shown to be a constant (depending on the dynamical model used for interpretation). Though historically $\Delta S_{\text{conformational}}$ was assumed to be negligible, it is now apparent that it can be large (at least for some systems like CaM and CAP) and may even be the dominant contributor to ΔS. If that is the case, then ΔH describes the change in structure and ΔS (mostly) describes the change in fast timescale dynamics.

In 1984, Cooper and Dryden presented an elegant statistical mechanical argument that the typical ΔG values observed for allosteric processes could be obtained solely from changes in ps-ns motions ($\Delta S_{conformational}$) in the absence of structural changes ($\Delta H = 0$) [118]. For example, they calculated that ΔG values of a few kcal/mol can be obtained by the summation of many small dynamical changes within the protein. Because these changes are minute fluctuations about the mean structure, they would not be detectable by standard methods of structure determination. Of course, the extreme limiting case of purely dynamically driven allostery need not be the rule. One could easily envisage a combination of changes in structure and changes in dynamics underlying an allosteric transition—for example, if the transition from one structural state to another is facilitated by a favorable change in conformational entropy. A simplified illustration of the various thermodynamic scenarios for a two-state allosteric transition like that of LacI is illustrated in Figure 1.9.

The seemingly iconoclastic proposition by Cooper and Dryden was not experimentally observed until 30 years later by Kalodimos and co-workers [119, 120]. Using NMR, they measured the changes in conformational entropy of a mutant form of CAP upon allosteric activation by its small molecule effector, cAMP. The mutant was unable to undergo the necessary conformational change required for activation but was still nonetheless activated by cAMP due to a large favorable change in conformational entropy. This is currently the only published report of experimentally observed dynamically driven allostery in the *complete absence* of conformational change, though it should be noted that allostery mediated by fast timescale dynamics has been described in other contexts and systems [121], including perhaps the most classical model of allostery, Hemoglobin [122].

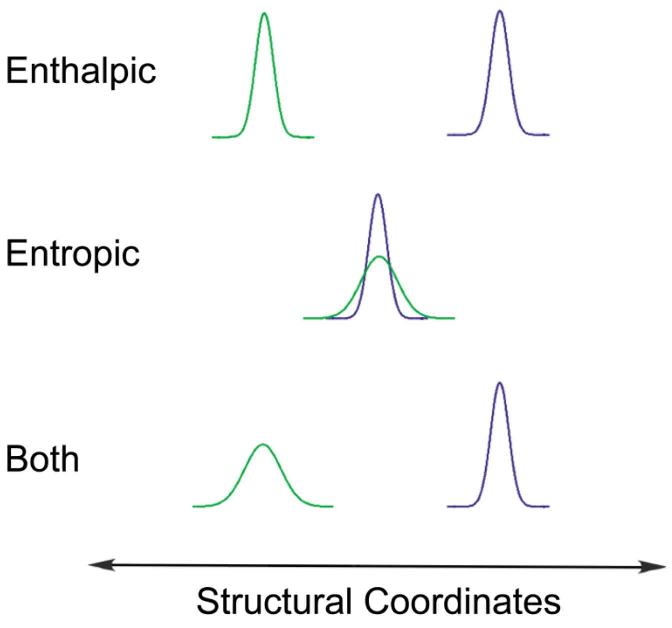


Figure 1.9 Schematic of the possible thermodynamic origins of a two-state allosteric mechanism. The distribution of structural coordinates is shown for one state in green and the other in purple. The “Enthalpic” mechanism relies purely on structural changes as illustrated by the shift in mean structural coordinates. The “Entropic” mechanism illustrates a transition between states that does not involve a change in the mean structural coordinates (no conformational change) but rather a change in the fluctuations about the mean (protein dynamics). “Both” illustrates the case where enthalpic and entropic effects contribute to the allosteric transition. Figure reproduced from Wand, 2001 [109].

1.6 Objectives of dissertation

It is clear that a variety of allosteric proteins operate via modulation of their conformational dynamics on the slow and/or fast timescale. Indeed, recent work has tried to reconcile dynamically-driven allostery with the classical models, leading to a unified “ensemble” model for allostery [123]. In this model, allosteric behavior is taken to be a manifestation of the weighted contribution of all substates in the protein ensemble where both structural and dynamical properties contribute to the sampling of these substates. An understanding of both structure and dynamics with respect to how they relate to allosteric mechanism is therefore necessary. So what does this mean for LacI? To date, no high-resolution characterization of the dynamics of either tetrameric or dimeric LacI has been presented. A few studies have investigated the internal dynamics (on both slow and fast timescales) of the DBD of LacI in the absence and presence of DNA (O_1 and non-operator sequences) but these studies do not offer any insight into allostery which is of interest here [90, 124]. Moreover, the structural characterization of LacI is incomplete which will be necessary to fully delineate the allosteric mechanism. To purpose of this work is to address this dearth of understanding. Specifically, we seek to illuminate the allosteric mechanism of LacI by:

(1) Characterizing the changes in structure and changes in dynamics of LacI upon binding inducer in the absence of operator

(2) Characterizing the changes in structure and changes in dynamics of LacI upon binding inducer in the presence of operator

In order to do this we will employ solution NMR which is the *only* experimental technique capable of providing residue-level structural informational and dynamical informational across both slow and fast timescales. Structural information will primarily be read out using the chemical shift, an NMR observable sensitive to local chemical environment. Dynamics on both slow and fast timescales will be investigated using a variety of methods which will be introduced subsequently.

The experimental system (see Chapter 2 for details) we will employ is carefully designed to minimize complications. It consists of:

Construct:	Dimeric LacI, residues 1-331
Operator:	22-base pair left-half site symmetric operator, O_{sym}
Inducer:	IPTG

The dimeric construct of LacI was chosen because it is the smallest intact functional unit which exhibits inducer-binding, DNA-binding, and allosteric properties similar to the full length tetramer. The motivation for using O_{sym} is two-fold: First, using a completely symmetric operator ensures the symmetry of LacI is intact which minimizes the number of observable NMR cross peaks and simplifies spectral analysis. Second, the bulk of the structural studies have utilized O_{sym}. IPTG was chosen because it has been used in nearly every structural and biochemical study of LacI to date and is commercially available.

Though we have chosen to work with the minimal functional unit of LacI, it is necessary to note that NMR suffers from a notorious size limitation. Traditionally, conventional NMR-based characterization has been limited to proteins with molecular weights of < 20 kDa [125]. Over the last two decades, several novel isotope labeling schemes and technical methodological advances described as **T**ransverse **R**elaxation **O**ptimized **S**pectroscop**Y** (TROSY) have been introduced which greatly extend this limitation [126, 127]. The physical basis of TROSY is rich and discussed briefly in Appendix I. Protein assemblies with molecular weights as large as 1 MDa have been characterized using these methodologies, albeit under highly idealized experimental conditions [128, 129]. Even in light of these technological advances, the vast majority of systems studied by NMR exhibit molecular weights < 40 kDa [130]. With this in mind, we have chosen to study the dimeric construct of LacI which is the minimal functional (allosteric) unit. At ~70 kDa in the free state and ~85 kDa in the 22 bp operator-bound state, dimeric LacI poses a significant technical challenge for NMR both at the level of sample preparation and the level of experiment. Chapter 2 introduces the specific challenges associated with sample preparation and presents a novel recombinant expression system and optimized purification scheme for producing LacI samples amenable to NMR-based characterization. Chapter 3 introduces the specific challenges associated with the implementation of NMR experiments (specifically those necessary for assigning chemical shifts) and presents an adaptation of novel data collection and processing techniques to overcome these challenges. These methodological developments were then applied to characterizing the allosteric mechanism of LacI. Chapter 4 presents the changes in structure and changes in dynamics that occur upon inducer binding in the absence of operator.

Chapter 5 presents the changes in structure and changes in dynamics that occur upon inducer binding in the presence of operator. Though many challenges were overcome over the course of this study, obtaining reliable, quantitative measure of fast time scale dynamics of LacI proved persistently difficult. Chapter 6 introduces improved methodologies for the quantitative characterization of fast timescale dynamics which should enable a more detailed description of the allosteric mechanism of LacI in the future.

CHAPTER 2: Optimized Recombinant Expression and Purification of Lac Repressor and its Constituent Domains for Solution NMR Studies

Most of the work presented in this chapter was done in collaboration with Marie Carter, undergraduate student in Cognitive Science.

The majority of this chapter was published in Stetz MA, Carter MV, and Wand AJ, Protein Expression and Purification **123** (2016) 75-82.*

**Undergraduate student*

2.1 Introduction

An important but often overlooked aspect of NMR-driven structural studies is sample preparation. This is perhaps ironic given that sample preparation is usually the most time-consuming part of any biomolecular NMR project [131]. Like many other biophysical methodologies, NMR requires samples of high purity and high homogeneity. Monodisperse samples are ideal (though not essential) in order to simplify data analysis. Additionally, there are several NMR-specific restrictions that are placed on samples. A prominent limitation of NMR is its intrinsic insensitivity which stems from the small polarization of nuclear spins at accessible static field strengths. This necessitates the use of large volumes (0.3-0.5 mL) of highly concentrated protein solutions (0.1-1.0 mM) for conventional instruments (though small volume “micro-coil” instruments have been introduced recently [132, 133]). Since NMR samples for the experiments proposed in this work also require isotopic enrichment, it is ideal to maximize protein yield in order to minimize cost. Isotopic labeling of proteins is usually achieved via recombinant expression in a host organism cultured in an isotopically enriched growth medium. Though a variety of hosts have been developed for this purpose [134-136], the most cost-effective and efficient is *E. coli*. The cost of protein production increases substantially when high molecular weight proteins are the subject of investigation. As discussed previously in Chapter 1 and further in Appendix I, high molecular weight proteins usually require TROSY methodologies for detailed characterization. In order to properly exploit the TROSY effect, proteins must be highly perdeuterated. This can be achieved via recombinant expression in *E. coli* by replacing H_2O with

D_2O as the growth medium solvent and using perdeuterated carbon sources. These modifications often compromise the viability of *E. coli* and decrease recombinant expression levels substantially. Though other host organisms can also be used to produce perdeuterated proteins [137, 138], the associated costs are much higher and the efficiency of deuterium incorporation is significantly lower. Given that LacI is a native *E. coli* protein, it is reasonable to presume that recombinant expression in *E. coli* should be robust and potentially tolerant of the detrimental effects associated with cell growth in deuterated medium. However, a recent study suggests that high yield recombinant production of LacI is exceptionally difficult to obtain under such conditions [139].

Perhaps the most confounding difficulty that arises when expressing LacI in an *E. coli* host is choosing an appropriate expression vector and strain pair. The vast majority of commercially available inducible expression vectors (such as the wildly popular *lac*-T7 promoter pET series of vectors [140]) and protease-deficient host strains utilize LacI and the *lac* O_1 operator as regulatory elements. Using such systems to express LacI has been reported to result in heterogeneous samples due to heterodimerization between recombinant and regulatory LacI [139]. Additional heterogeneity arises from the use of IPTG to induce expression which co-purifies sub-stoichiometrically with the recombinant product when typical concentrations are utilized [139]. Moreover, expression levels are likely to be reduced since additional repressor molecules are continuously being produced.

Several alternative approaches have been introduced to circumvent contamination. One recent study utilized a pET-derived vector wherein the LacI/O_1 regulatory elements were deleted [139]. However, the recombinant *lacI* gene was kept under control of a T7 promoter and the source of T7 RNA polymerase was the DE3 lysogen of a commonly used protease-deficient *E. coli* strain. In the DE3 lysogen, the gene for T7 RNA polymerase is under control of a *lacUV*5 promoter and expression is regulated by LacI/*O1*. Therefore, IPTG must still be added to the growth medium to induce significant expression. Vectors where LacI is expressed constitutively have also been used widely, however yields are relatively low as most biochemical studies necessitate 12 L of culture—even when highly rich growth medium such as 2xYT is used [28, 141]. Commercially-available inducible expression vectors which do not utilize LacI, such as those featuring arabinose-inducible (pBAD) [142] or rhamnose-inducible (pRham) [143] promoters, require high concentrations of metabolically active sugars to initiate over-expression. In order to prevent scrambling of isotope labels, glucose levels must be kept in excess continuously throughout the growth. This is often not economically feasible, especially when deuterated glucose is required for protein labeling. Moreover, both pBAD and pRham promoters are repressed by high concentrations of glucose [142, 143]. Temperature-inducible expression has also been used for truncated constructs of LacI [144], however, drastic shifts in temperature may have detrimental effects on cells under high stress conditions, such as those imposed by growth in highly deuterated minimal medium.

Complications with sample preparation are also not strictly limited to full length or dimeric constructs of LacI. Indeed, expression of a recombinant dimeric form of the isolated regulatory domain (RD) (residues 60-331) has been shown to be toxic to a variety of commonly used *E. coli* host strains—so much so that even chemical transformations with plasmids harboring the RD gene were shown to be impossible [139]. This toxicity could only be eliminated by co-expression of the isolated RD with full length LacI which resulted in abysmally low yields due to heterodimerization. Recombinant expression of the isolated DNA-binding domain (DBD) in *E. coli* has also proven difficult. It is thought that the isolated DBD becomes more susceptible to proteolysis *in vivo* as its length increases [144]. For example a DBD construct lacking some of the hinge region (residues 1-56) expresses to levels nearly 10-fold higher than those obtained from expressing constructs with an intact hinge region (residues 1-64) [144]. As a result, studies on the isolated DBD often employ fermentation to improve yields [124]. Though the focus of this dissertation is the complete functional unit of LacI, it is often advantageous to carry out preliminary NMR-based characterization (such as resonance assignment) on smaller constructs as a means to improve data quality and simplify data analysis (discussed further in Chapter 3).

Here, novel expression systems for LacI, the isolated RD, and isolated DBD are presented along with optimized and facile purification procedures. Nearly all complications previously associated with the recombinant production of LacI and its constituent domains are eliminated. It is shown that using a derivative of the pASK backbone [145] which employs an anhydrotetracycline-inducible *E. coli tetA* promoter regulated by the Tetracycline repressor (TetR) and the *lacI*-free BLIM strain *E. coli* host [146] eliminates all contamination issues which previously plagued the recombinant production of LacI. Moreover, this expression system can produce the isolated RD without any defects in cell viability. A pET-15b derived expression vector and the common protease-deficient BL21(DE3) host strain are shown to be optimal for expressing the isolated DBD. Sufficient yields of DBD can be obtained using this expression system without the need for fermentation. Facile and semi-automated purification schemes are also introduced which dramatically reduce sample preparation time. The structure and function of the resulting LacI preparations are characterized using limited proteolysis and NMR spectroscopy. It is shown that recombinant LacI and its constituent domains are homogeneous, structured, and functional.

2.2 Material and Methods

2.2.1 Expression vector construction

The sequences for all primers used in this work can be found in Appendix 2.

An expression vector containing the gene encoding dimeric LacI (1-331) under control of an inducible *tetA* promoter regulated by TetR was constructed. The gene encoding wild type *E. coli* LacI (residues 1-331 of LacI, 109A isoform from strain K-12, Swiss-Prot entry: P03023) was amplified by PCR from an arabinose-inducible vector (pBAD Gateway, ThermoFisher/Invitrogen) containing the *LacI* gene in its expression cassette (gift from Prof. Mitchell Lewis, University of Pennsylvania). A Tobacco Etch Virus (TEV) protease cleavage site and EcoRI restriction site were added to the 5’ end of the sequence and an XhoI restriction site was added to the 3’ end of the sequence by primer extension. The gel-purified PCR product was ligated into a high copy number, linearized pGEM-T Easy vector (Promega) and transformed into XL1-Blue cells (Agilent). Cells harboring vectors which contained the insert were identified by blue-white screening by plating cells on X-gal. Vectors were amplified in rich media and subsequently isolated by miniprep. Vectors were then digested using EcoRI and XhoI restriction enzymes (NEB). The insert was isolated by gel purification and then subcloned into a pASK-IBA35plus vector backbone (IBA) using the corresponding restriction sites. The pASK vector backbone confers ampicillin resistance. Ligated vectors were transformed into *E. coli* DH5α cells (ThermoFisher), amplified in rich media, and isolated by miniprep. Ligation of the correct insert was confirmed by DNA sequencing. Vectors harboring the gene for the isolated RD of LacI (residues 60-331) and the isolated DBD of LacI (residues 1-62) were produced using identical procedures. Recombinant LacI constructs expressed from this vector possess an N-terminal 6x-His tag and TEV-protease cleavage site. Following treatment with TEV-protease, a single non-native glycine residue is left at the N-terminus. For all constructs, the first few N-terminal residues are unstructured. The additional glycine is therefore not expected to affect functionality.

An expression vector containing the gene encoding the DBD of LacI, including the entire hinge region (residues 1-62) under control of an inducible T7 promoter was constructed. Residues 1-62 were amplified by PCR from an arabinose-inducible vector (pBAD Gateway, ThermoFisher/Invitrogen) containing the *LacI* gene in its expression cassette (gift from Prof. Mitchell Lewis, University of Pennsylvania). An NdeI restriction site was added to the 5’ end of the sequence and an XhoI restriction site was added to the 3’ end of the sequence by primer extension. The gel-purified PCR product was ligated into a high copy number, linearized pGEM-T Easy vector (Promega) and transformed into XL1-Blue cells (Agilent). Cells harboring vectors which contained the insert were identified by blue-white screening by plating cells on X-gal. Vectors were amplified in rich media and subsequently isolated by miniprep. Vectors were then digested using NdeI and XhoI restriction enzymes (NEB). The insert was isolated by gel purification and then subcloned into a pET-15b vector backbone in-frame with an N-terminal 6x-His tag and Thrombin cleavage site. Ligated vectors were transformed into *E. coli* DH5α cells (ThermoFisher), amplified in rich media, and isolated by miniprep. Ligation of the correct insert was confirmed by DNA sequencing. The DBD construct expressed from this vector possess an

N-terminal 6x-His tag and Thrombin-protease cleavage site. Following treatment with Thrombin, three non-native residues are left at the N-terminus: glycine, serine, and histidine. Again, the first few N-terminal residues of this construct are unstructured so the additional residues are therefore not expected to affect functionality significantly.

2.2.2 Host strains

Because the pASK vector utilizes an *E. coli* promoter, there are no restrictions placed on the choice of host strain. The BLIM strain [146] was chosen as the host strain for the pASK vector. BLIM cells are derived from the BL26 Blue strain (Novagen), a derivative of BL21 cells from which the *lac* operon has been deleted and an F´ episome carrying *lacI* under control of the *lacI*^{q} promoter has been introduced. The BLIM strain was originally made by curing BL26 Blue cells of the F´ episome, resulting in a strain completely devoid of the *lacI* gene. BLIM cells were obtained from Prof. Kathleen Matthews' laboratory (Rice University) via Addgene (Bacterial strain #35609) and the absence of the *lacI* gene was confirmed by colony PCR using the same primers employed to amplify the *LacI* gene. The classical MacConkey colormetric screen was employed to determine if the BLIM strain possessed a functional *lac* operon [147]. In brief, liquid growth medium consisting of peptone (20 g/L), lactose (10 g/L), and phenol red (0.025 g/L) was prepared. Plates of this medium were also cast after the addition of agar powder (15 g/L). Cells which possess functional LacY and LacZ proteins ferment lactose, causing the pH of the medium to fall below 6.8. This results in a yellow-colored medium. Cells which do not possess functional LacY or LacZ proteins metabolize peptone which produces ammonia and raises the pH above 8.2. This results in a pink-colored medium. It should be noted that the original MacConkey screen utilized neutral red as a pH indicator instead of phenol red so the color readout here is reversed. Furthermore, it should be made clear that this test cannot definitively confirm or deny the presence of the *lac* operon structural genes, only the presence of functional forms of the LacY and LacZ proteins.

The BLIM strain cannot be used as a host for the pET-15b vector since this vector utilizes a T7 promoter which is not recognized by *E. coli* RNA polymerase. As a result, T7 RNA polymerase must be expressed separately before recombinant protein expression can commence. This can be accomplished via co-transformation with a separate plasmid [146] or via the use of a genetically modified host strain where the gene for T7 RNA polymerase has been integrated into the chromosome. A variety of commercially-available host strains have been created via the latter approach. We have chosen to use one such strain, BL21(DE3), which is a protease deficient B-strain derivative that has been lysogenized with the DE3 element, a lambda prophage carrying the gene for T7 RNA polymerase under control of an inducible *lacUV5* promoter. The *lacUV5* promoter is regulated by LacI and induction is performed via addition of exogenous IPTG. This vector and strain combination therefore contains multiple copies of the *lacI* gene and *lac* O_1 operator: one copy of the *lacI* gene and *lac* O_1 operator on the vector, one copy of the *lacI* gene and *lac* O_1 operator (as well as O_2 and O_3 operators) on the natural chromosome, and one copy of the *lacI* gene and *lac* O_1 operator on the DE3 element. Though the user manual for the commercially-available BL21(DE3) strain (Novagen pET System Manual) suggests that the *lacI* gene is not included in the DE3 element, DNA sequencing confirms that the *lacI* gene is indeed included (GenBank entry: EU078592.1, enterobacteria phage DE3, complete genome from BL21(DE3)). Native *lacI* is expressed at low levels due to the inherent weakness of the *lacI* promoter, however, it has been documented that heterodimerzation between regulatory LacI and recombinant LacI can still be an issue [139]. As such, the pET-15b vector/BL21(DE3) host strain expression system was only used to express the isolated DBD which does not oligomerize, does not bind IPTG, and binds the *lac* O_1 operator with considerably weaker (~3 orders of magnitude) affinity than full length LacI. Furthermore, the BL21(DE3) strain was also previously demonstrated to be highly sensitive to expression of the isolated RD, so much so that over-expression of the isolated RD resulted in cell death [139].

2.2.3 Expression of LacI and LacI regulatory domain

All growth media were supplemented with ampicillin. Chemically competent BLIM cells were transformed with either the pASK vector harboring the gene encoding LacI or the pASK vector harboring the gene encoding the RD of LacI and selection was performed using LB plates following an overnight (< 16 hour) incubation period at 37˚C. For proteins expressed in rich or H_2O-M9 minimal medium, a single colony was first used to inoculate a 50 mL culture of LB medium and grown in a shake flask at 37˚C overnight to stationary phase (OD_{600} ~ 3-4). The next morning, 1 L cultures of either LB or H_2O-M9 minimal medium were inoculated to an initial OD_{600} of 0.05 and left to grow at 37˚C in shake flasks until an OD_{600} of 0.5 (H_2O-M9 minimal medium) or 1.0 (LB medium) was reached. At this point, cultures were induced with 100 μL of a 2 mg/mL stock of anhydrotetracycline (IBA) prepared in ethanol. The temperature was then reduced to 25˚C and cells were left to grow overnight (< 16 hours). Cells were harvested by centrifugation and frozen at -80˚C as 30 mL suspensions in 50 mM sodium phosphate, 500 mM sodium chloride, pH = 7.8.

Preparation of highly deuterated LacI and RD followed a procedure similar to that previously published for preparing highly deuterated malate synthase G [148]. A single colony was selected from LB plates and used to inoculate a 5 mL LB culture. The culture was grown at 37˚C until cells reached an OD_{600} ~0.8-1.0. At this point, cells were spun down and re-suspended in 20 mL of H_2O-M9 minimal medium and left to grow at 37˚C until an OD_{600} of 0.5 was reached. Cells were then spun down and re-suspended in 75 mL D_2O-M9 minimal medium (> 90% D_2O v/v). Cultures were left to grow overnight at 37˚C. The next morning, cells were harvested and used to inoculate a 900 mL cultures of D_2O-M9 minimal medium to an initial OD_{600} of 0.05. Cells were grown at 37˚C until an OD_{600} of 0.4-0.5 was reached. Cultures were induced with 100 μL of a 2 mg/mL stock of anhydro-tetracycline prepared in ethanol. Induced cultures were left to grow overnight at 30˚C (< 16 hours). Typical OD_{600} values at stationary phase were ~1.0 for LacI and ~2.0 for RD preparations. Cells were harvested by centrifugation and frozen at -80˚C as 30 mL suspensions in 50 mM sodium phosphate, 500 mM sodium chloride, pH = 7.8. In some cases HEPES (20 mM, pH = 7.8) or TRIS (20 mM, pH = 7.8 measured at 25˚C) was used instead of sodium phosphate.

2.2.4 Expression of LacI DNA-binding domain

All growth media were supplemented with ampicillin. Expression of the isolated DBD of

LacI from the pASK/BLIM expression system was compared to that from the pET-15b/BL21(DE3) expression system.

Chemically competent BLIM cells were transformed with the pASK vector harboring the gene encoding the DBD of LacI and selection was performed using LB plates following an overnight (< 16 hour) incubation period at 37˚C. For proteins expressed in rich or H_2O-M9 minimal medium, a single colony was first used to inoculate a 50 mL culture of LB medium and grown in a shake flask at 37˚C overnight to stationary phase (OD_{600} ~ 3-4). The next morning, 1 L cultures of either LB or H_2O-M9 minimal medium were inoculated to an initial OD_{600} of 0.05 and left to grow at 37˚C in shake flasks until an OD_{600} of 0.5 (H_2O-M9 minimal medium) or 1.0 (LB medium) was reached. At this point, cultures were induced with 100 μL of a 2 mg/mL stock of anhydrotetracycline (IBA) prepared in ethanol. A variety of expression temperatures and durations were assessed (*vide infra*), however, the optimal combination was 37˚C and 3-4 hours. Cells were harvested by centrifugation and frozen at -80˚C as 30 mL suspensions in 50 mM sodium phosphate, 500 mM sodium chloride, pH = 7.8. Identical procedures were followed for expression using the pET-15b/BL21(DE3) expression system except using 1 mM IPTG as the inducer. Typical OD_{600} values at stationary phase were ~1.0 for both constructs.

2.2.5 Purification of LacI and LacI regulatory domain

Frozen cell suspensions were thawed in a room-temperature water bath and then transferred to ice. $MgCl_2$ was added to a final concentration of 10 mM, $CaCl_2$ was added to a final concentration of 5 mM, Triton-X 100 was added to a final concentration of 0.2% (v/v), and fresh phenylmethylsulfonyl fluoride (PMSF) was added to a final concentration of 1 mM. One "cOmplete Mini" EDTA-free, broad range protease inhibitor tablet (Roche), 5 mg of hen egg white lysozyme, and 3.5 mg of bovine pancreas DNaseI were also added. The suspension was left to rock gently at 4˚C for one hour. Sonication was then performed on ice using a repeated series of 15 pulses followed by resting on ice. Cell debris was pelleted by centrifugation at 15,000 rpm for 30 minutes at 4˚C and the lysate was further clarified by passage through a 0.4 μm syringe filter.

All purification steps utilized an Äkta Prime Plus FPLC system operating at 4˚C with the exception of the final gel filtration step which was sometimes performed at room temperature. Methods were programmed to automate isolation procedures. Protein was detected by UV/Vis absorbance at 280 nm. Immobilized metal affinity chromatography (IMAC) was performed using a 3-5mL nickel-IDA column (His60 Ni Superflow, Clontech) and two buffers: buffer A (50 mM sodium phosphate, 500 mM sodium chloride, and 3 mM fresh BME, pH = 7.8) and buffer B (50 mM sodium phosphate, 500 mM sodium chloride, 250 mM imidazole, and 3 mM fresh BME, pH = 7.8). In some cases HEPES (20 mM, pH = 7.8) or TRIS (20 mM, pH = 7.8 measured at 25˚C) was used instead of sodium phosphate. The lysate was loaded onto the column using a 50 mL superloop and the loaded column was washed with buffer A then 16% buffer B until the detector baseline was flat. Protein was then eluted isocratically using 100% buffer B. The eluted fractions were analyzed by SDS-PAGE, pooled, and dialyzed against 4 L of buffer A overnight at 4˚C in the presence of 1 mg of recombinant 6x-His tagged TEV protease [149].

Protease-treated protein solutions were then subject to a second round of IMAC using the same nickel IDA column equilibrated with buffer A at 4˚C. The flow through was collected and purity was assessed by SDS-PAGE. Pure fractions were then pooled, concentrated to about 0.2-0.3 mM (for both LacI and LacI regulatory domain), then subject to gel filtration using either a ~120 mL Superdex 75 (at room temperature) or ~65 mL Superdex 200 (at 4˚C) column (GE Healthcare) equilibrated with NMR buffer: either 20 mM sodium phosphate, 150 mM sodium chloride, 3 mM DTT, pH = 7.4 (LacI) or 20 mM sodium phosphate, 50 mM sodium chloride, 3 mM DTT, pH = 7.4 (LacI regulatory domain) at 4˚C. In some cases HEPES or TRIS (20 mM, pH = 7.4 for both) was used instead of sodium phosphate. The final purity of LacI or LacI regulatory domain was typically > 99% as determined by quantification of SDS-PAGE results using the program ImageJ (data not shown) [150].

2.2.6 Purification of LacI DNA-binding domain

Cells were lysed and an initial round of IMAC was performed identically as described for the LacI and RD constructs. Purified fractions were dialyzed overnight against buffer A without added protease. The next morning (<16 hours later), the dialysate was removed and brought to room temperature. A 1/200 v/v dilution of bovine thrombin stock solution (250 U/mL) was added and the protein solution was left to rock gently at room temperature for 2-4 hours. Thrombin was subsequently inhibited by the addition of PMSF to a final concentration of 0.01 mM. The protein solution was then subjected to a second round of IMAC using the same nickel IDA column equilibrated with buffer A at 4˚C. The flow through was collected and purity was assessed by SDS-PAGE. Several high molecular weight impurities which ran at approximately 100 and 30 kDa on an SDS-PAGE gel was consistently retained following the second IMAC run. Preliminary characterization of the purified DBD ignored this contaminant since it was unlikely to be detected by NMR due to its high molecular weight and low concentration relative to the DBD. Fractions containing DBD were buffer exchanged into 20 mM sodium phosphate, 200 mM sodium chloride, pH = 7.4 via serial dilution using a 3.5 kDa molecular weight cutoff centrifugal concentrator. Monitoring of this sample by NMR revealed that it was partially proteolyzed, perhaps due to the high molecular weight impurity. The contaminant was therefore removed in subsequent preparations by gel filtration using a Superdex 200 column using 20 mM sodium acetate, 200 mM sodium chloride, pH = 4.5 as the running buffer which also served as the final NMR buffer.

2.2.7 Preparation of O_{sym}, LacI-O_{sym}, and LacI DBD-O_{sym} samples

The 22 BP O_{sym} operator (5'-GAATTGTGAGCGCTCACAATTC-3') was purchased as HPLC-purified, single-stranded DNA (IDT) and received as lyophilized powder. Operator was dissolved in nuclease-free H_2O (Promega), aliquoted in 200 µM portions, frozen and re-lyophilized to remove residual trifluoracetic acid (TFA). Operator was stored at -20˚C. Duplexed operator was made by solubilizing lyophilized single-stranded aliquots in annealing buffer (10 mM TRIS, 50 mM NaCl, 1 mM EDTA, pH = 8.0) and then heating to 95˚C in a heatblock for 5 minutes. Heat-treated operator was then removed and allowed to slow cool on the bench top for several hours. Operator concentration was determined using the calculated extinction coefficient at 260 nm given in the spec sheet from IDT and assuming negligible hyperchromic effects (i.e. 100% annealing efficiency). Proper duplexing was checked by native agarose gel in TAE buffer (40 mM TRIS, 20 mM acetate, 1mM EDTA, pH = 8.0) using SYBR Green (ThermoFischer) for visualization and by ^{1}H NMR analysis of imino proton resonances using an excitation sculpting sequence [151, 152] for water suppression.

LacI-O_{sym} complexes were prepared by mixing purified LacI with 20% molar excess of duplexed operator. The complex was then buffer exchanged into NMR buffer (20 mM HEPES, 20 mM NaCl, 3 mM TCEP pH = 7.4) using a 2 mL Sephadex-G25 (GE Healthcare) spin column. LacI DBD-O_{sym} and LacI-O_{sym}-IPTG complexes were prepared similarly except with a 50% molar excess of duplexed operator. An alternative NMR buffer (10 mM sodium phosphate, 20 mM sodium chloride, pH = 6.1) was used for the DBD-O_{sym} sample which had been employed previously in many studies by the Kaptein lab [83]. Binding saturation was confirmed by TROSY or conventional ^{1}H-^{15}N HSQC experiments prior to data collection.

2.2.8 Limited proteolysis

In order to assess the structural integrity of recombinant LacI and RD, limited proteolysis was performed using trypsin from bovine pancreas (Sigma-Aldrich). Briefly, 1 mg (minimally 6000 enzyme units) of lyophilizied trypsin was reconstituted in 1 mL of 20 mM sodium phosphate, 200 mM sodium chloride, 3 mM DTT, pH = 7.4. A separate solution containing 50 µM LacI or RD in the same buffer was also made. Following gentle pipetting, 10 µL of trypsin solution was added to 90 µL of the LacI or RD solution. Time points were taken at: 1, 2, 3, 4, 5, 10, 20, and 30 min and quenched in SDS-loading buffer. The time points were then run on a denaturing SDS-PAGE and visualized using coomasie.

2.2.9 Buffer screening for NMR studies

NMR spectroscopy is sensitive to the chemical nature of the buffers used in the solubilization of samples. For experiments which detect amide hydrogen nuclei (the bulk of the experiments proposed in this work), sensitivity is strongly affected by pH due to hydrogen exchange (HX) with H_2O. Accelerated HX results in the attenuation of observable amide proton signals; pH values are therefore minimized in order to slow HX as much as possible. Unfortunately, this presents a severe problem of LacI for two reasons: First, LacI has been documented to precipitate at pH < 6.5 [139]. Second, the bulk of the biochemical characterization has been performed on LacI at pH = 7.4-7.5.

An additional complication specific to our NMR hardware is an acute sensitivity to the ionic strength of the sample solution. In our laboratory, cryogenically cooled NMR probes are used on all instruments which offer a theoretical gain in sensitivity of a factor of 4 [153]. However, the observed gain in sensitivity is strongly dependent on the ionic strength of the sample being analyzed [154]. Since LacI is a DNA-binding protein, it requires relatively high concentrations of salt to maintain solubility when free from DNA—with some reports even suggesting as much as 400 mM salt be used [139]. Such high concentrations of salt largely negate the advantages of using a cryogenically cooled probe (by nearly a factor of 2 for our machines [155]) and will invariably disrupt DNA-binding, making the choice of sample buffer very difficult.

A variety of buffers were therefore screened empirically and in low-throughput fashion. Recombinant LacI prepared in various buffers at various concentrations was monitored for aggregation by visual inspection (for precipitation) or absorbance at 600 nm and at later stages by NMR spectroscopy using ^{1}H-^{15}N TROSY-HSQC experiments (data not shown). Failed screens are not shown for brevity but the following conditions were identified as being optimal:

Table 2.1 Buffer conditions for LacI NMR samples

Functional State	**Protein Conc.**	**Buffer**	**Salt Conc.**	**pH**
Apo LacI	100-200 μM	Phosphate, HEPES, TRIS (20 mM)	150-200 mM	7.4
LacI-IPTG	100-200 μM	Phosphate, HEPES, TRIS (20 mM)	150-200 mM	7.4
LacI-O_{sym}	150-250 μM	HEPES (20 mM)	0-20 mM	7.4
LacI-O_{sym}-IPTG	150-250 μM	HEPES (20 mM)	0-20 mM	7.4
Apo LacI RD	100-200 μM	Phosphate, HEPES, TRIS (20 mM)	50 mM	7.4
LacI RD-IPTG	100-200 μM	Phosphate, HEPES, TRIS (20 mM)	50 mM	7.4
Apo LacI DBD	1 mM	Acetate (20 mM)	200 mM	4.5
LacI DBD-O_{sym}[1]	1 mM	Phosphate (10 mM)	20 mM	6.1

[1] Buffer conditions determined previously by Rob Kaptein's group [83]

2.2.10 NMR spectroscopy

NMR samples were prepared as described in Table 2.1 with 0.1 mM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) and 0.02% w/v sodium azide added as a chemical shift reference and preservative, respectively. For LacI and LacI RD samples, TROSY ^{1}H-^{15}N HSQC [156] experiments were collected at 298 K (for DNA free states) or 308 K (for O_{sym}-bound states) on a Bruker Avance III spectrometer equipped with cryoprobe operating at 750 MHz ^{1}H Larmor frequency. ^{1}H-^{15}N HSQC [157] [158] experiments were collect on LacI DBD samples at 298 K (apo state) or 308 K (O_{sym}-bound state) on a Bruker Avance III spectrometer equipped with cryoprobe operating at 500 MHz ^{1}H Larmor frequency. All NMR data were processed using NMRPipe/NMRDraw [159] after adjusting data for digital oversampling. The solvent signal was suppressed using a low pass boxcar-shaped deconvolution filter with a length of ±16 points. Time domain data were then apodized with a 70˚ squared sine bell function and zerofilled twice (then rounded up to the nearest power of 2) prior to discrete Fourier transform (DFT). In some cases, a polynomial function base line correction was applied after transform to the acquisition ^{1}H dimension to correct for distortions introduced by the H_2O resonance.

2.3 Results and Discussion

2.3.1 Characterization of BLIM strain

The BLIM strain is an attractive host for the recombinant production of LacI and the RD because it is devoid of the *lacI* gene. However, since the strain was derived via curing another strain of an episome, the absence of the *lacI* gene needed to be confirmed. This was done by colony PCR using the same primers used to amplify the *lacI* gene for cloning (see appendix 2 for primer sequences). Figure 2.1a shows the results from the colony PCR. Untransformed BLIM cells were plated on LB-agar and left to grow at 37˚C for approximately 16 hours. BLIM cells transformed with a vector containing the *lacI* gene (residues 1-331) were also plated and left to grow in identical fashion. Single colonies from each plate were then subject to whole colony PCR and the reaction mixtures were subsequently run on an agarose gel. It is clear that all untransformed BLIM colonies do not yield amplicons whereas all transformed BLIM cells yield amplicons which run at ~1 kB, as expected for the 331 residue LacI construct. This confirms that the *lacI* gene is not present in the BLIM strain. Figure 2.1b-c shows the results of a MacConkey colormetric screen which assays for functional LacY and LacZ proteins. Untransformed BLIM and BL21(DE3) cells were grown either in liquid MacConkey medium or on MacConkey agar plates. The BLIM cells uniformly yield pink cultures, indicating that they cannot utilize lactose. The BL21(DE3) cells, however, uniformly yield yellow cultures, indicating that they can utilize lactose. This suggests (though does not necessarily confirm) that the BLIM strain is likely devoid of the *lac* operon as expected from the genotype originally reported of its progenitor BL26 strain. Since the presence of the *lacZ*, *lacY*, and *lacA* genes will not impact the heterogeneity of recombinant LacI or RD produced in BLIM cells, confirmatory genetic studies (such as PCR) were not pursued.

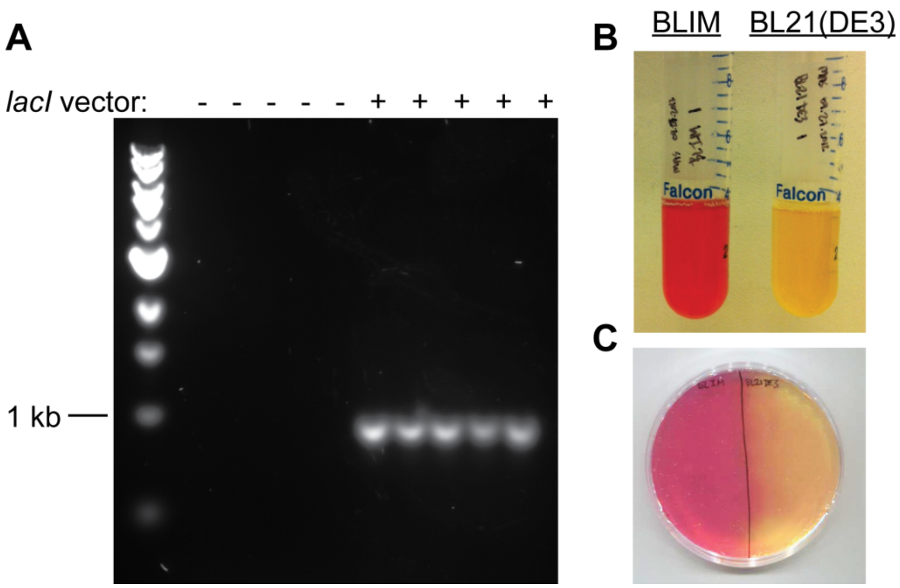


Figure 2.1 Characterization of the BLIM host strain. (A) Colony PCR results for BLIM cells either untransformed (left 5 lanes) or transformed with an expression vector containing the *lacI* gene (right 5 lanes). (B) MacConkey liquid medium screen for untransformed BLIM cells (left) and BL21(DE3) cells. (C) MacConkey solid medium screen for untransformed BLIM cells (left) and BL21(DE3) cells (right).

2.3.2 Expression vector construction

Because LacI is a homo-oligomer, expressing a recombinant copy of *lacI* in the presence of other copies of the *lacI* gene could result in hetero-oligomerization if the dimerization and/or tetramerization domains are intact. In order for significant hetero-oligomerization to occur, the expression levels of each copy of the *lacI* gene need to be somewhat comparable. The native *lacI* gene in *E. coli* is expressed at very low levels and only a few copies of LacI exist in a single *E. coli* cell at any given time [160]. Over-expressing a recombinant copy of *lacI* in this background is unlikely to result in detectable heterodimerization. However, popular commercial *E. coli* host strains used in T7 expression systems (such as BL21(DE3) and derivates thereof) contain an additional copy of the *lacI* gene on the DE3 element which is used to regulate the expression of T7 RNA polymerase [140]. Additionally, the most commonly used vectors possessing LacI-regulated T7 promoters posses their own copy of the *lacI* gene. Most commercially-available pET vectors have a pBR322 origin of replication which yields copy numbers of 15-20 per cell. The potential for heterodimerization becomes even greater when *lacI* promoter mutations, such as the popular *lacI*^{q} [161] or *lacI*^{q1} [162] mutations, are used to increase expression levels of LacI (10-100 fold) for tighter regulation. Despite this, such expression systems are sometimes used to express recombinant LacI and IPTG is used to induce recombinant expression [139, 163-165]. Since the k_d of IPTG binding for LacI and LacI RD is on the order of 1-5 μM [166], contaminating IPTG can be retained throughout purification. Indeed, recombinant LacI produced from an IPTG-inducible expression system has been shown previously to be a mixture of apo and IPTG-bound states, even though the concentration of IPTG used to induce over-expression in that particular study was only 0.5 mM [139].

While procedures could be devised to homogenize LacI, such efforts are laborious and will almost invariably involve some degree of denaturation followed by refolding due to the tight dissociation constants for IPTG binding and dimerization [167]. An alternative approach is presented here which eliminates these complications by simply removing all instances of the *lacI* gene from the expression system. Figure 2.2 shows a diagram of the LacI constructs and TetR-regulated expression vectors built in this work. The backbone is based on the inducible pASK vector [145]. The LacI gene with TEV-protease recognition sequence is inserted in frame with an N-terminal 6x-His tag. Recombinant expression is regulated by the native *E. coli tetA* promoter from transposon 10 (Tn*10*) which is repressed by the TetR protein. Induction is initiated by introducing a small amount of anhydrotetracycline, a non-toxic analog of tetracycline. The use of an *E. coli* promoter eliminates the need for host strains lysogenized with the *lacI*-containing DE3 element.

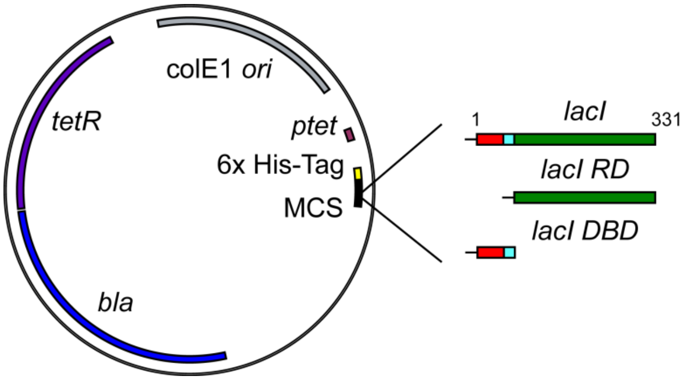


Figure 2.2 TetR regulated expression vector map. The backbone is the pASK35plus vector which possesses an inducible tetracycline promoter (*ptet*) from *E. coli*, regulated by the TetR protein. The vector confers ampicillin resistance (*bla*) and is medium copy number (colE1 origin). The multiple cloning site (MCS) follows a 6x-His tag. Genes encoding either LacI or the isolated LacI regulatory domain (*LacI RD*) with N-terminal TEV protease recognition sites were subcloned into the backbone in frame with the His tag.

Since IPTG co-purification and hetero-oligomerization are not issues when producing the isolated DBD, it could potentially be expressed from a common *lac*-T7 promoter expression system. Therefore, a pET-15b vector harboring the gene encoding the LacI DBD (residues 1-62) was also constructed. The map is shown in Figure 2.3. The motivation for constructing this vector in addition to the pASK vector is simple: T7 RNA polymerase is known to be approximately 5 times more efficient at initiating transcription relative to *E. coli* RNA polymerase [168]. As such, higher yields should be expected from using a T7 expression system. This is particularly important for the recombinant production of a DBD construct which contains the intact hinge region and suffers from low yields due to a presumed susceptibility to proteolysis *in vivo* [144].

2.3.3 Expression and purification of LacI

As described above, uniform perdeuteration is required for solution NMR studies of LacI. In order to express highly deuterated LacI, cells were only minimally adapted to D_2O using slightly modified protocols previously described [148]. Briefly, the procedure involved an initial growth in LB medium, followed by transfer to H_2O-M9 minimal medium, then immediate transfer to D_2O-M9 minimal medium (> 90% v/v). This method was favored over serial adaptation since

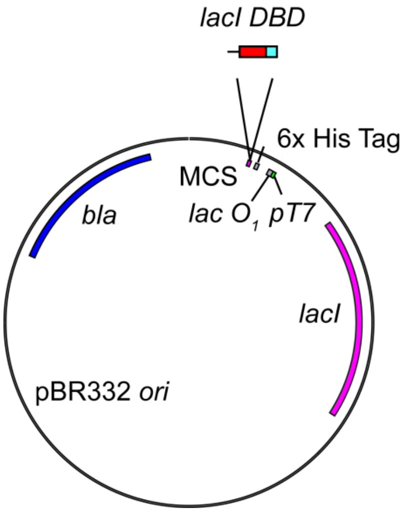


Figure 2.3 LacI regulated expression vector map. The backbone is the pET-15b vector which possesses an inducible T7 promoter (*ptet*), regulated by the LacI. The vector confers ampicillin resistance (*bla*) and is medium copy number (pBR322 origin). The multiple cloning site (MCS) follows a 6x-His tag and thrombin cleavage site and the primary *lac* operator (*lac* O_1). Because expression is induced by the addition of IPTG and a copy of the *lacI* gene is on the vector, this vector should only be used for producing the isolated DBD.

repeated passages over the course of days in an *endA*$^+$/*recA*$^+$ host could result in plasmid instability. In order to qualitatively assess the expression level of LacI in highly deuterated minimal media, several small scale expression tests were performed. Small cultures (~2 mL) of cells that had been transferred to highly deuterated minimal medium were left to grow at 37˚C until densities reached an OD_{600} ~0.4. At this point, a pre-induction sample was taken and the remaining cells were induced. Whole cell pellets, harvested before and after induction were then subjected to SDS-PAGE which is shown in Figure 2.4. Prior to induction, no noticeable bands are evident above background, confirming previous reports that the *tetA* promoter is tightly regulated by TetR, despite being a native *E. coli* promoter [145]. Following induction with anhydrotetracycline, a prominent band appears at about 40 kDa, indicating LacI over-expression. OD_{600} values for cultures in stationary phase following induction were typically ~1.0 for D_2O-M9 minimal medium. These results confirmed that the newly introduced expression system can tolerate growth in highly deuterated minimal medium.

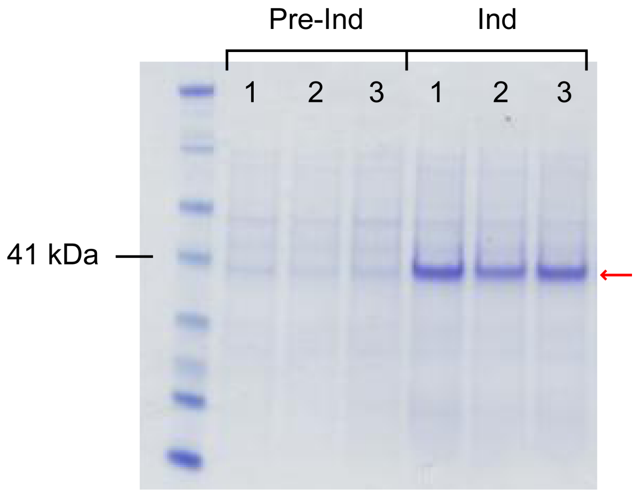


Figure 2.4 Expression of LacI in highly deuterated medium. Whole cell pellets were subject to SDS-PAGE. The results for three separate small scale (~2 mL) cultures are shown. The left most lane is a molecular weight marker sample. “Pre-Ind” lanes were pellets harvested just prior to induction (OD_{600} ~0.4) and “Ind” lanes were pellets harvested following overnight induction at 30˚C (OD_{600} ~1.0). The bands marked with the red arrow correspond to the tagged LacI gene product which has an expected molecular weight of 38.6 kDa.

LacI produced from 1 L cultures was subjected to three purification steps as described in Section 2.2. Protein purity after each step was monitored by SDS-PAGE. The first IMAC step using a nickel IDA resin results in highly pure protein as shown in Figure 2.5. Fractions containing significant amounts of LacI were pooled, treated with His-tagged TEV protease, then subject to IMAC again in order to separate TEV-protease, digested His tags, and any residual LacI that still possessed a His tag (Figure 2.5b). Tag-less LacI elutes in the flow through but is also retained on the column until washed with 16% buffer B. This is because LacI has an inherent affinity for nickel resins [169]. Samples eluted in the flow through and after the 16% wash with buffer B were structurally identical as judged by NMR (data not shown). The final purification step was gel filtration on a Superdex 75 column (Figure 2.6). The elution profile reveals a single, narrow peak which is consistent with a highly homogeneous species. Typical yields of LacI are ~13-14 mg/L of culture in highly deuterated minimal medium. These yields are nearly 2-fold higher than those previously reported for a thermostable mutant of LacI (K84M) expressed from an unregulated vector [139]. Expression of the wild type protein was not presented in that work.

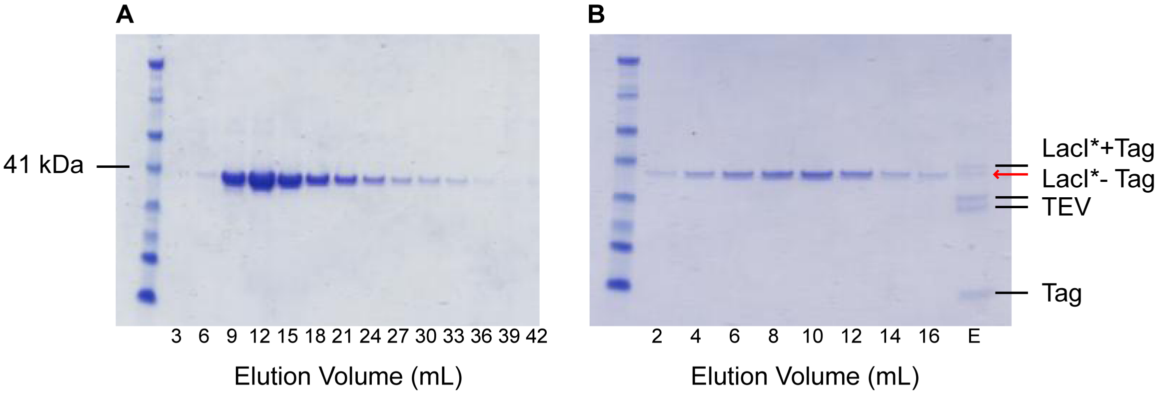


Figure 2.5 SDS-PAGE analysis of purified LacI. The left most lane for each panel is a molecular weight marker sample. (A) Eluted fractions containing tagged LacI from the first IMAC step. The elution volume is relative to the start of the application of 100% buffer B. (B) Flow through containing tag-less LacI from the second IMAC step. The last lane of (B) marked “E” is the eluted fraction obtained from washing the IMAC column with 100% buffer B and contains a small amount of tagged LacI, untagged LacI, TEV (two bands presumably due to TEV protease autodigestion) and the cleaved tag. The red arrow corresponds to the expected molecular weight of the tag-less LacI product (~35.5 kDa).

2.3.4 Expression and purification of LacI regulatory domain

A previous report showed that recombinant production of a thermostable mutant of the isolated RD of LacI (K84M), was toxic to three common *E. coli* host strains [170]. The toxicity was so pronounced that cells could not even be successfully transformed with plasmid DNA. It was observed that the toxicity was reduced only when full length LacI was introduced into the expression system at appreciable levels. As such, an IPTG-inducible expression vector was created to co-express full length LacI with K84M RD. Three major limitations to using this expression system were noted: 1) Heterodimerization between full length LacI and LacI (K84M) RD; 2) IPTG co-purification; 3) The requirement to keep starter cultures at low cell densities. While the first two issues could be addressed by purification and a special IPTG-removal step, the latter issue seemed unavoidable and was attributed to leaky expression of the toxic RD. This was perhaps surprising considering the host strain used for successful recombinant RD production, Rosetta(DE3) pLysS, utilizes a *lacUV5* promoter and expression of T7 lysozyme to regulate the expression and functionality of T7 RNA polymerase. The *lacUV5* promoter should impart very tight regulation since it is insensitive to decreases in catabolite repression as cell densities increase and T7 lysozyme should inhibit T7 RNA polymerase. This would suggest that the toxicity of LacI (K84M) RD is particularly acute.

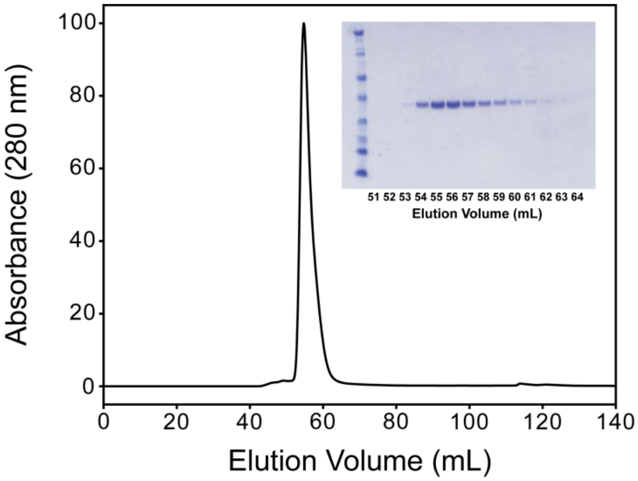


Figure 2.6 Gel filtration chromatography of purified LacI. The inset shows the SDS-PAGE results for the eluted fractions where the left most lane is a molecular weight marker sample.

Recombinant production of the wild type isolated RD was attempted using the pASK vector/BLIM strain combination. Figure 2.7 shows the growth curve for BLIM cells carrying the vector for wild type LacI RD in highly deuterated minimal medium. All measurements were made prior to inducing over-expression. The growth curve for BLIM cells carrying the vector for intact dimeric LacI is also shown for reference. It is clear that cells carrying the LacI RD expression vector grow robustly at a rate nearly identical to that for cells carrying the intact dimeric LacI expression vector. To see if the previously observed toxicity stemmed specifically from the use of the thermostable K84M mutation, growths were monitored for BLIM cells transformed with a vector harboring LacI(K84M) RD and a similar mutant, K84L, which exhibits an even stronger degree of thermostabilization [171]. The growth curves are shown in Figure 2.7b and it is apparent that neither vector is toxic to BLIM cells.

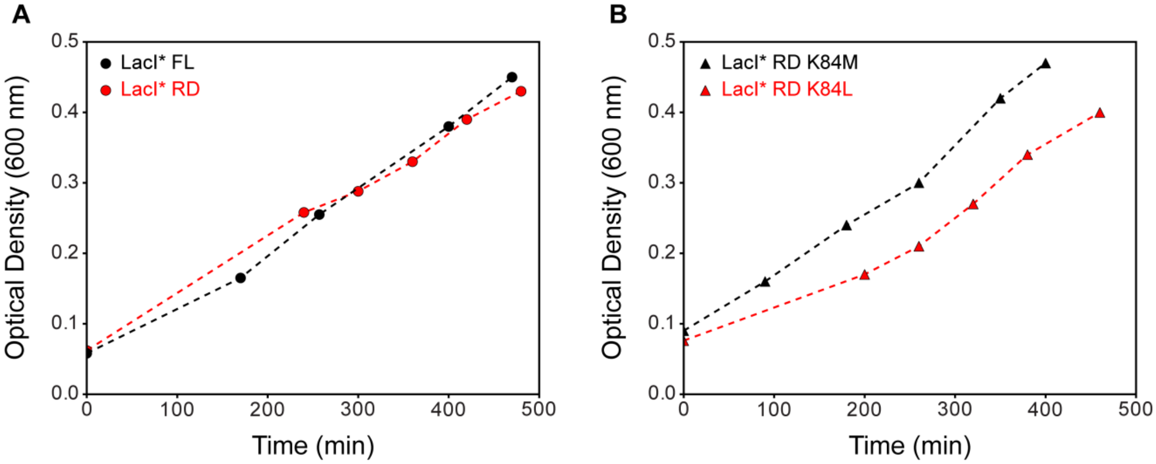


Figure 2.7 Elimination of LacI RD toxicity. Cell growth is monitored by OD_{600}. (A) BLIM cells carrying the full length (FL) LacI* expression vector (black circles) and BLIM cells carrying the LacI* isolated regulatory domain (RD) vector (red circles). (B) BLIM cells carrying the LacI* (K84M) isolated regulatory domain vector (black triangles) and BLIM cells carrying the LacI* isolated regulatory domain vector (red triangles). All cells were cultured in highly deuterated M9 minimal medium and grown at 37˚C in shake flasks.

The effect of over-expressing LacI RD was then examined. Figure 2.8a-c shows an expression test for wild type, K84L, and K84M LacI RD in highly deuterated minimal medium, respectively. It is clear that all three constructs exhibit high levels of over-expression. Final OD_{600} values for induced 1L cultures in stationary phase were ~2.0 for wild type RD and the K84L mutants and ~1.5 for the K84M mutant. Yields obtained from 1 L growths in highly deuterated minimal medium were ~15 mg for wild type, ~10 mg for K84L, and ~9 mg for K84M. The slightly reduced yields for the K84 mutants may not necessarily reflect enhanced toxicity, since it was observed the His tag cleavage was markedly less efficient for these mutants (data not shown). This is perhaps expected considering that the crystal structure of the K84L mutant exhibits a more tightly packed monomer-monomer interface relative to the structure of wild type [172]. This likely reduces flexibility and, in turn, access to the His tag. Regardless, the observed yields are all over 2-fold higher than the previously reported yield [170]. Ostensibly, this increase may not appear that drastic. However, it should be noted that *E. coli* RNA polymerase is five-fold less efficient than T7 RNA polymerase [173], therefore the improvement in yield is actually quite considerable.

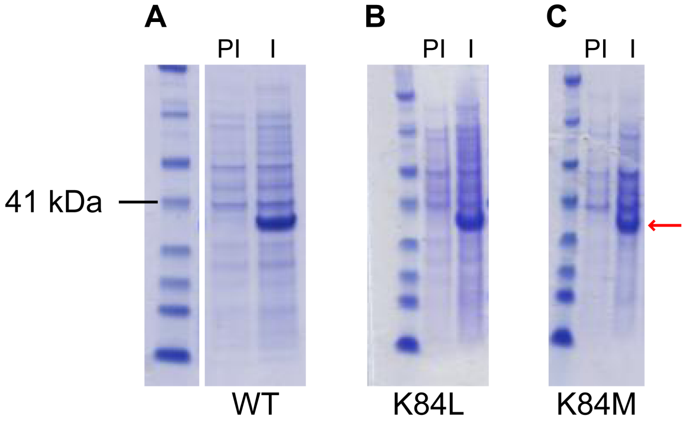


Figure 2.8 Expression profiles for LacI RD (A) wild type LacI* RD, (B) LacI*(K84L) RD, and (C) LacI*(K84M) RD. Whole cell pellets were subject to SDS-PAGE. For each panel, left most lanes are molecular weight marker samples, center lanes are pre-induction (PI) samples (OD_{600} ~0.4) and right most lanes are post-induction (I) samples (OD_{600} ~2.0 for wild type and K84L, ~1.5 for K84M). The red arrow corresponds to the expected molecular weight of the tagged RD (~32.1 kDa).

2.3.5 Expression and purification of LacI DNA-binding domain

Presumably, the recombinant production of the isolated DBD should be straightforward since there is no threat of hetero-oligomerization or co-purification. As described in Chapter 1, the DBD has been the subject of extensive study by solution NMR. Traditional methods for producing the DBD recombinantly have relied on a T7-promoter based system wherein the expression of T7 RNA polymerase is regulated by temperature [124, 144, 174]. However, yields from such expression systems have been reported to be low and often high density fermentation is employed. Two alternative expression vectors were constructed for producing recombinant LacI DBD and compared: one utilizing the pASK/BLIM expression system and another using a pET15-

b/BL21(DE3) expression system.

Because of its small molecular weight, the isolated DBD does not need to be expressed in deuterated growth medium (at least for the specific NMR resonance assignment applications for which this construct will be used). Initial attempts to over-express the isolated DBD using the pASK/BLIM expression system in rich medium (LB) failed. Changing the temperature and expression time had no noticeable effects on expression levels. The exact origin for the low expression yield is unclear, however, it may be due to mild toxicity. The average OD_{600} value of induced cultures in stationary phase was 1.57 ± 0.17 (n = 4) following 4 hours of expression at 37˚C and 2.08 ± 0.61 (n = 4) following 16 hours of expression at 25˚C. This is considerably lower (~2.0x) than typical saturating OD_{600} values for LacI and RD over expressed in LB medium. Previous work had suggested that the DBD was prone to proteolysis *in vivo*, however the strain used in that particular study, HB101, was not protease deficient [144]. The BLIM strain used here, like most B-strain derivatives, is only devoid of lon and OmpT proteases, so the *in vivo* degradation hypothesis is still plausible. Regardless of the exact origin, it is apparent that the pASK/BLIM expression system is not optimal for producing isotopically-labeled DBD.

Similar expression tests were then performed for the pET15-b/BL21(DE3) expression system. Figure 2.9a shows that over-expression is readily detected in rich medium. Importantly, robust over-expression is also detected in H_2O-M9 minimal medium, suggesting that isotopically-labeled NMR samples of the DBD can be made. Though it should be noted that final OD_{600} values for over-expression in H_2O-M9 minimal medium were consistently ~1.0 which is about 2x lower than growths where transformed cells were left uninduced.

The purification of the isolated DBD was straightforward. Figure 2.9b-c shows SDS-PAGE results for a typical preparation. While IMAC was generally sufficient for acceptable purity, faint bands at higher molecular weights (~30 kDA and ~100 kDa) were often observed. The contaminating species could subsequently be removed by gel filtration.

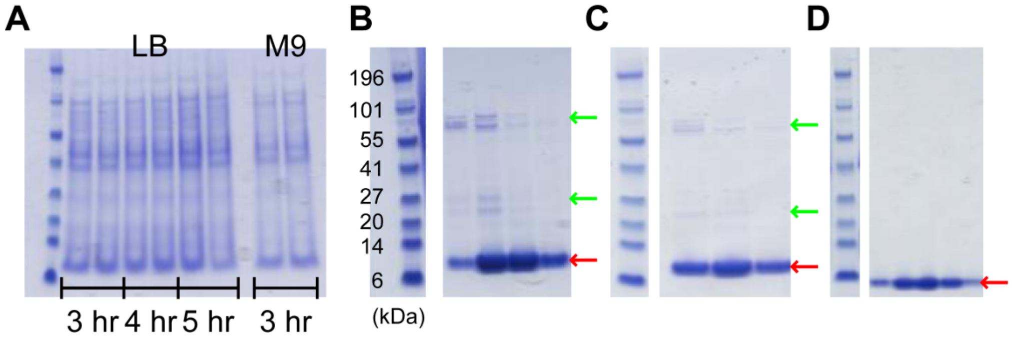


Figure 2.9 Expression and purification of LacI DBD (A) SDS-PAGE of whole cell pellets over-expressing LacI DBD in LB and M9 minimal medium. The units of time refer to expression duration following induction. (B) SDS-PAGE of eluted fractions from 1st IMAC column. (C) SDS-PAGE of eluted fractions from 2nd IMAC column. (D) SDS-PAGE of eluted fractions following gel filtration. The red arrow corresponds to the expected molecular weight of the DBD (~7 kDa) and the green arrows indicate persistent contaminants.

2.3.6 Recombinant LacI is structurally intact

One major concern that arises when producing and isolating LacI is the preservation of the DBD. Early limited proteolysis studies of tetrameric LacI revealed that a small N-terminal fragment, which we now know to be the DBD, is susceptible to proteolytic removal by trypsin. Based on the amino acid sequence and crystal structure, the trypsin cleavage site is located in the hinge region (K59) which explains why proteolyzed DBD fragments retain the ability to bind DNA, albeit weakly. This sensitivity to proteolysis has recently re-emerged as a key concern when producing LacI, so much so that care was taken to retain an N-terminal 6x His tag so as to prevent introducing proteases to the recombinant product [139]. This was an unusual decision since most proteases utilized for the removal of affinity tags are highly specific. While gel-shift assays revealed that the presence of the N-terminal tag did not interfere with DNA-binding, a conflicting report suggested the opposite [163]. Since tags are generally large and mobile, they often produce difficult to interpret NMR spectral signatures. As such, we have sought to remove all affinity tags prior to characterization.

Gel filtration elution profiles and SDS-PAGE suggest that recombinant LacI produced using our new expression system is indeed intact. In order to confirm this, LacI and the isolated

RD were subject to limited proteolysis by trypsin. Figure 2.10 shows that recombinant LacI produced using our expression and purification procedure is indeed intact. Within a minute, the DBD is readily cleaved, yielding the RD which is largely insensitive to trypsin digestion. These results are entirely consistent with the older literature [48, 50, 51].

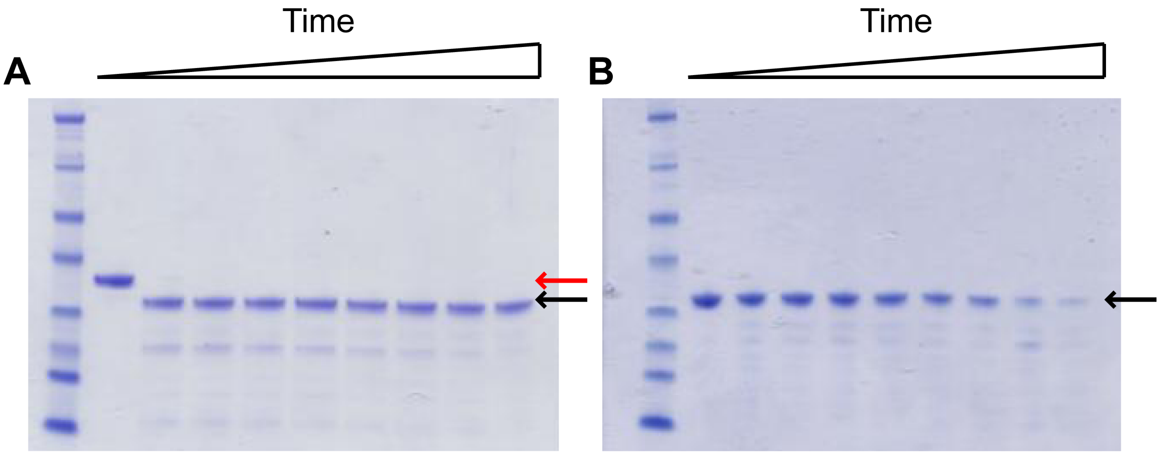


Figure 2.10 Limited proteolysis of LacI and LacI RD (A) SDS-PAGE of LacI samples incubated with trypsin as a function of time. (B) SDS-PAGE of LacI RD samples incubated with trypsin as a function of time. The red arrow corresponds to the molecular weight of full length LacI (~37 kDa). The black arrow corresponds to the molecular weight of the isolated RD (~29 kDa).

2.3.7 Confirmation that O_{sym} is duplexed

The preparation of duplex O_{sym} followed conventional protocols, however, it has been noted that annealing single-stranded O_{sym} can be problematic and inefficient (Liskin Swint-Kruse, personal communication). In order to confirm that annealing O_{sym} resulted in the complete formation of properly duplexed DNA, ^{1}H NMR was performed. Figure 2.11 shows the imino proton region of a ^{1}H NMR spectrum of O_{sym} prepared using the protocol outlined above. The spectrum is identical to the previously published spectrum of the duplexed operator [175], confirming that O_{sym} is indeed properly duplexed in our preparations.

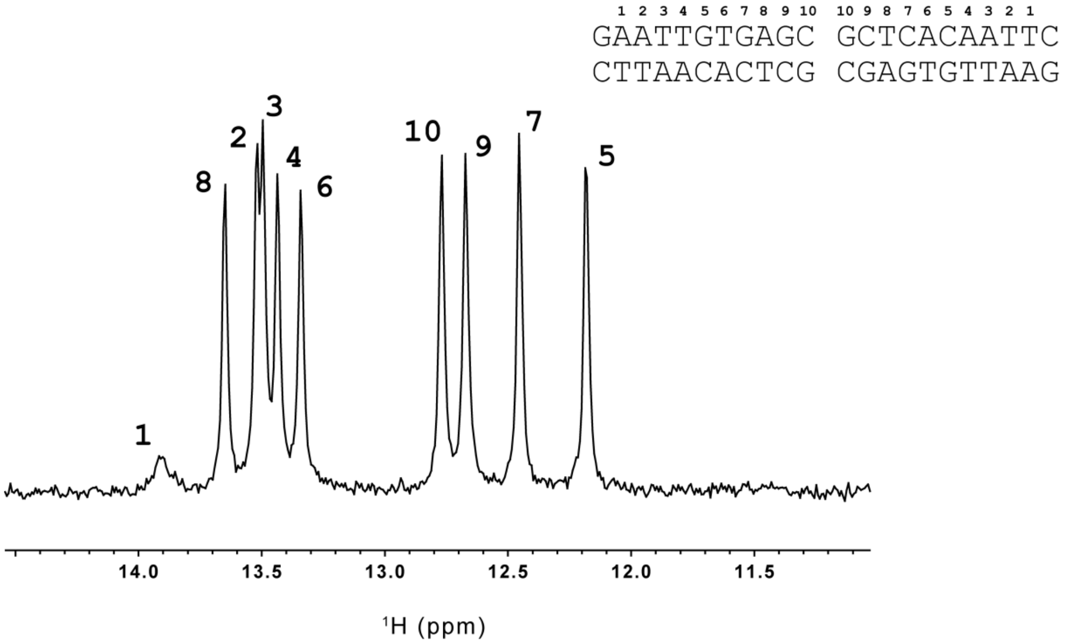


Figure 2.11 ^{1}H NMR spectrum of the imino protons of duplexed O_{sym}. The spectrum is identical to previously published data and resonances assignments are indicated. The terminal base pairs are not observed due to rapid exchange with the solvent as a result of fraying [175].

2.3.8 Preliminary characterization of LacI structure and function by NMR

In order to gain a more detailed structural view of the recombinant products, NMR spectroscopy of isotopically labeled LacI and its constituent domains was performed. The simplest and most informative NMR experiment at this stage is ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC). This spectrum is often referred to as the "fingerprint" of a protein because it yields a resonance for each amide H-N bond vector. However, the identity or "assignment" of each resonance cannot be determined without additional data collection. At the time of this study, resonance assignments were only available for the isolated DBD bound to O_{sym} (via Prof. Rob Kaptein's group deposited in the Biological Magnetic Resonance Bank (BMRB)). However, even without assignments, HSQC spectra can still be interpreted with respect to structure and function.

The frequency of an observed resonance, known as the chemical shift, is exquisitely sensitive to changes in the local chemical environment. As such, it is an extremely valuable readout for conformational changes that may accompany binding events. The functionality of a protein can therefore be determined simply by visual inspection of chemical shift perturbations. With respect to LacI, we are interested in probing IPTG-binding and operator-binding.

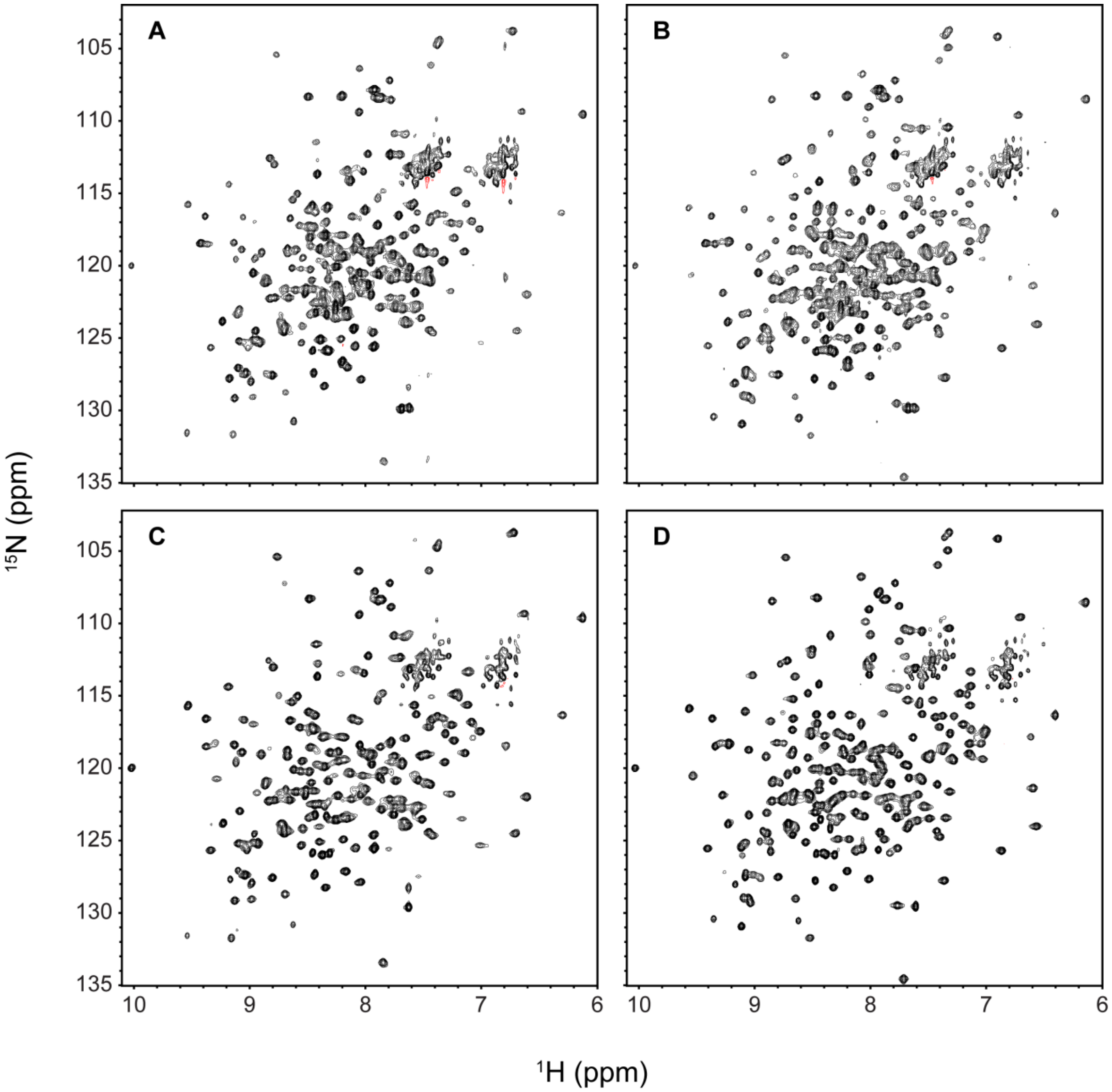


Figure 2.12 ^{1}H-^{15}N TROSY-HSQC spectra of LacI and LacI RD apo and IPTG-bound states (A) Apo LacI. (B) LacI bound to IPTG. (C) Apo RD. (D) RD bound to IPTG. All data were collected at 750 MHz and 25˚C. Negative peaks are shown in red.

As described in Chapter 1, the operator we have chosen for this study is O_{sym} because the perfectly symmetric DNA sequence simplifies NMR spectra. Moreover, the interaction between LacI and O_{sym} has been characterized in the bulk of structural studies performed to date. Both IPTG and O_{sym} bind LacI with sub-micromolar dissociation constants. High affinity binding is typically (though not always [176]) slow on the NMR-timescale, meaning that the rate of exchange between free and bound states is slower than the difference between the free and bound resonance frequencies. In other words, apo and bound states yield unique peaks rather than one peak with a population-averaged frequency. This is beneficial at this preliminary stage of characterization because it allows for the rapid detection of non-functional or under-titrated samples.

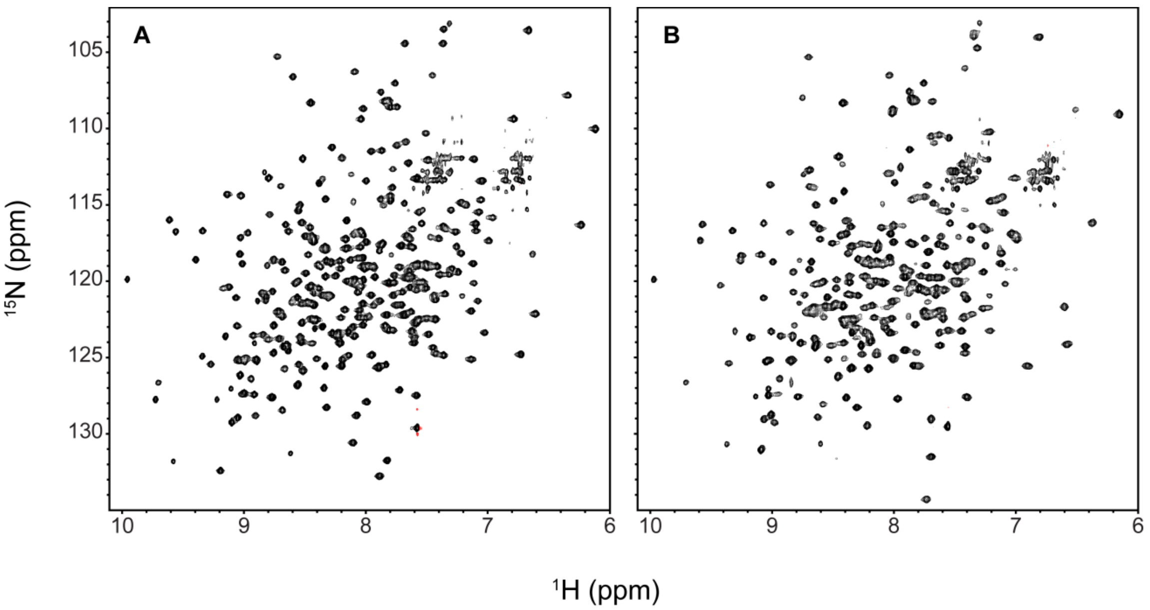


Figure 2.13 ^{1}H-^{15}N TROSY-HSQC spectra of LacI operator-bound states (A) LacI-O$_{sym}$ binary complex. (B) LacI-O$_{sym}$-IPTG ternary complex. All data were collected at 750 MHZ and 35˚C. Negative peaks are shown in red.

Figure 2.12 shows ^{1}H-^{15}N TROSY HSQC spectra for apo and IPTG-bound LacI (2.12a-b, respectively) and LacI RD (2.12c-d, respectively) collected at 25˚C. The temperature was chosen in accordance with previously reported thermal denaturation melting temperatures of dimeric LacI. The LacI RD and DBD have different melting temperatures: 55˚C (measured by us using fluorescence emission, data not shown) and 37˚C [177], respectively. As such, 25˚C was chosen to accommodate the less thermostable DBD. The dispersion of chemical shifts indicates that the proteins are well folded. A single set of resonances is observed, which is to be expected for a symmetric dimer. Importantly, drastic changes in chemical shift are observed upon IPTG-binding which confirms that the proteins are functional. Qualitatively, it is clear that spectra of the RD are highly similar to those of LacI (to be discussed further in Chapter 3) which suggests that the constituent domains will be useful for chemical shift assignment.

Spectra of LacI bound to a duplexed 22-base pair O_{sym} are shown in Figure 2.13. The ^{1}H-^{15}N TROSY HSQC spectrum of the LacI-O_{sym} binary complex is shown in 2.13a and the ^{1}H-^{15}N TROSY HSQC spectrum of the LacI-O_{sym}-IPTG ternary complex is shown if 2.13b. Both were collected at 35˚C to improve molecular tumbling. To our knowledge, thermal denaturation studies have not been reported for the dimeric LacI-O_{sym} binary or LacI-O_{sym}-IPTG ternary complex. However, such studies have been reported for tetrameric LacI bound to non-operator DNA and indicate that DNA-binding increases the melting temperature of LacI [178]. Curiously, it was shown that IPTG-binding has no affect on the melting temperature of LacI, whether free or bound to DNA. As a control, spectra were also collected at 25˚C and no significant differences were observed (data not shown). The spectral quality is excellent. Again, both states yield spectra with a single set of resonances, consistent with a symmetric dimer. Importantly, we present (to our knowledge) the first high resolution view of the ternary LacI-O_{sym}-IPTG complex which is recalcitrant to crystallization.

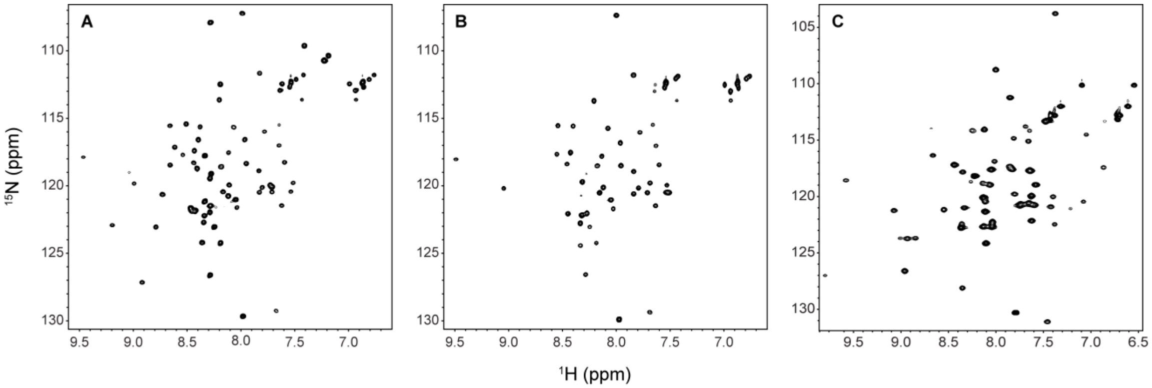


Figure 2.14 ^{1}H-^{15}N HSQC spectra of the isolated LacI DBD (A) Apo DBD at pH = 4.5. (B) Apo DBD at pH = 7.4. (C) DBD-O_{sym} complex. (A-B) were collected at 500MHz and 25˚C whereas (C) was collected at 500 MHz and 35˚C. All data were processed identically.

In order to assess the quality of the recombinant LacI DBD produced using the pET-15b/BL21(DE3) expression system, ^{1}H-^{15}N HSQC spectra were collected at 25˚C. Initial NMR studies were performed at the same pH used in studies of LacI and LacI RD, pH = 7.4. However, it was apparent that many resonances were broadened beyond detection. While line broadening can arise for a variety of reasons, one plausible explanation is accelerated HX. In order to test this, a ^{1}H-^{15}N HSQC spectrum was also collected at pH = 4.5. Figure 2.14a-b shows spectra of the apo DBD at pH = 4.5 and pH = 7.4, respectively. It is clear that the peaks that were not observed at pH = 7.4 are observable at pH = 4.5.

In order to test the functionality of the isolated DBD, a ^{1}H-^{15}N HSQC spectrum of the DBD were collected in the presence of saturating amounts of O_{sym} as shown in Figure 2.14c. The dramatic changes in chemical shift indicate that the isolated DBD is competent to bind O_{sym}, thus confirming its functionality.

2.4 Conclusions

This chapter sought to address several unique complications that are associated with the recombinant production of LacI and its constituent domains for high-resolution NMR studies.

Novel expression vectors and optimized purification methods were introduced which result in the following significant advantages over previous approaches:

- High yield, inducible over-expression of LacI without adding contaminating IPTG or high concentrations of metabolically active sugars.
- High yield over-expression of the isolated LacI RD without defects in cell viability and without the need for co-expression of full length LacI.
- Robust over-expression of LacI and the isolated LacI RD in highly deuterated growth medium.
- Recombinant production of sufficient amounts of the isolated LacI DBD without high density fermentation.
- Increased yield of recombinant LacI, RD, and DBD relative to previous reports.

In addition, suitable buffer conditions were identified for NMR studies of LacI, RD, and DBD which maximize sample lifetime and spectral quality. Importantly, the NMR buffer conditions for LacI are nearly physiological and largely match those used in the bulk of the basic biochemical characterization. This is in contrast to previous attempts to characterize dimeric LacI by high resolution NMR [139]. Our preliminary NMR-based characterization indicates that LacI, RD, and DBD are all well folded and functional. This suggests using NMR to study the structure and function of LacI is feasible.

CHAPTER 3: Resonance Assignment of Lac Repressor and its Constituent Domains

Much of the work presented in this chapter was done in collaboration with Marie Carter, undergraduate student in Cognitive Science and Joanna Giang, high school student at Julia R. Masterman Laboratory and Demonstration School.

3.1 Introduction

In Chapter 2, preliminary ^{1}H-^{15}N TROSY-HSQC and ^{1}H-^{15}N HSQC spectra were presented for LacI and its constituent domains. These spectra revealed that the recombinant products obtained using our optimized expression and purification methods were folded and functional. Moreover, the dispersion of resonances was shown to be favorable, suggesting that high-resolution analysis of the changes in structure and changes in dynamics that accompany the allosteric transition of LacI could be performed. Unfortunately, because there had been no previous attempts to characterize dimeric constructs of wild type LacI, the identity of each resonance was unknown. Only the resonances for the spectrum of the isolated DBD bound to O_{sym} [83] have been made available to the public (BMRB Entry: 4813).

Assigning amide ^{1}H-^{15}N NMR resonances entails identifying the unique ^{1}H and ^{15}N chemical shifts that arise due to the chemical environment of each amide bond vector. This can be done in many ways [179-185] but the conventional approach is to utilize so-called “triple resonance” experiments. These experiments exploit the bonded interactions among nuclei, usually ^{1}H, ^{15}N, and ^{13}C [125]. Magnetically active nuclei that are bonded are coupled (referred to as “scalar-coupling” or “J-coupling”) and this coupling can be utilized to transfer magnetization from one nucleus to another. The efficiency of magnetization transfer through bonds depends on several factors, most notably: the total number of bonds, the magnitude of the “coupling constant” (J) which, in turn, depends on the interaction between nuclei and the electrons in the bond, and the relaxation properties of the nuclei involved in the transfer. Long-range transfers through many bonds are generally inefficient, so most triple resonance experiments involve a series of higher efficiency one-bond transfers. For the conventional assignment of amide ^{1}H and ^{15}N chemical shifts, magnetization is transferred first from the amide ^{1}H nucleus, then to the amide ^{15}N nucleus, then to one of several ^{13}C nuclei and then back to the amide ^{1}H nucleus. This series of magnetization transfers can establish intra-residue and inter-residue connections which can be exploited for resonance assignment. Figure 3.1 shows the major through-bond connections used in the typical suite of triple resonance experiments and their respective coupling constants [125]. Larger values indicate more efficient transfer. Often, triple resonance experiments are designed to correlate a given residue (denoted as residue “i”) with the residue that precedes it in sequence (denoted as residue “i-1”), primarily via carbon nuclei. Since degeneracies in frequency are common, these connections are established redundantly from correlations of several different carbon nuclei: alpha, beta, and carbonyl.

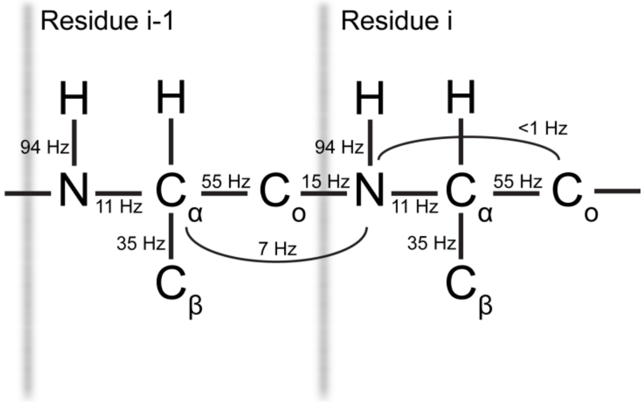


Figure 3.1 Resonance assignment strategy. Illustration of typical through-bond modes of magnetization transfer used in triple resonance experiments to assign amide ^{1}H-^{15}N resonances and their approximate coupling constants. Figure adapted from Sattler, M. *et al.,* 1999 [125]

While Figure 3.1 illustrates the various couplings that can be exploited for establishing sequential connections between residues, such connections yield no information about amino acid type which is required for assigning an identity to a given observable frequency. Identifying amino acid type relies largely on the chemical properties of alpha and beta carbons [186]. A specific subset of residues yield unique alpha and beta carbon frequencies, and assignments are made by simply combining the amino acid sequence, list of sequentially connected frequencies, and alpha and beta carbon frequencies. Though some amino acids also yield unique nitrogen and carbonyl carbon frequencies [181], it is much more difficult to make *unambiguous* assignments using these nuclei alone. Table 3.1 illustrates the unique frequencies commonly used for assigning amino acids types. A consequence of amino acid typing by alpha and beta carbon frequency is that these frequencies must absolutely be observed, meaning the signal must be reliably detectable above the noise floor of the spectrum. However, the experiments which detect these frequencies are highly insensitive due to the sheer number and nature of the magnetization transfer steps required as well as relaxation effects. Table 3.2 lists the relative sensitivities for a variety of experiments. As with nearly all NMR experiments, the sensitivity of triple-resonance experiments becomes limiting when the molecular weight of the molecule of interest increases due to unfavorable relaxation effects [187].

Table 3.1 Unique carbon chemical shifts used for amino acid assignment

Amino Acid	**Average Alpha Carbon Shift**[1]	**Average Beta Carbon Shift**[1]
Glycine	45.38 ± 2.39	N/A
Alanine	53.18 ± 2.98	19.07 ± 3.29
Serine	58.72 ± 2.95	63.72 ± 5.53
Threonine	62.24 ± 2.73	69.60 ± 6.30

[1] Average chemical shifts and standard deviations from BMRB

Once backbone resonance assignments have been obtained, a variety of experiments can be used to assign methyl ^{1}H-^{13}C resonances for methyl-bearing side chains. As discussed in Chapter 1, the methyl-bearing side chains are a crucial probe of protein dynamics and their motional properties contribute significantly to the residual conformational entropy of the folded state which may impact allosteric functionality. The assignment of methyl resonances is traditionally done using total correlation spectroscopy (TOCSY), which utilizes spin-lock sequences to generate correlations among coupled spins [188]. However, such methods are less effective for assigning the methyl resonances of large proteins due to relaxation effects, though some success has been found using modified implementations [189, 190]. Alternative approaches using correlation spectroscopy (COSY)-like transfers have been introduced and shown to be successful for assigning large proteins in high concentration [128, 191].

Table 3.2 Relative S/N of common triple resonance experiments

Experiment	Correlations	Relative S/N (%)
HNCO	H(i), N(i), CO(i-1)	100
HNCA	H(i), N(i), CA(i), CA(i-1)	50/15, CA(i)/CA(i-1)
HN(CO)CA	H(i), N(i), CA(i-1)	71
HNCACB	H(i), N(i), CA(i), CA(i-1), CB(i), CB(i-1)	4/1.7, CA(i)/CB(i) 1.3/0.5, CA(i-1)/CB(i-1)
HN(CA)CO	H(i), N(i), CO(i), CO(i-1)	13/4, CO(i)/CO(i-1)

Initial attempts to assign LacI started with the intact dimeric construct bound to O_{sym}. The reasons for this were as follows: First, the LacI-O_{sym} complex exhibited the highest molecular weight of all functional states and, as such, served as the limiting case for assessing the feasibility of obtaining assignments. Second, perhaps ironically, the LacI-O_{sym} complex yielded the highest quality ^{1}H-^{15}N HSQC spectrum in terms of dispersion of resonances and uniformity of linewidths. A series of TROSY-based triple resonance experiments were attempted on the LacI-O_{sym} under different experimental conditions (such as temperature and salt concentration) and static field strength. The results are summarized in Table 3.3. Unfortunately, only the TROSY-HNCO and TROSY-HNCA yielded reliably interpretable data. These data alone are insufficient for obtaining assignments.

Table 3.3 Summary of initial resonance assignment data for LacI-O_{sym} complex

Functional State	Salt Conc.	Temperature	Static Field[1]	Experiment	Number of Scans	Observable Cross Peaks[2]
LacI-O_{sym}	200 mM	25˚C	750 MHz	HNCO	40	80%
LacI-O_{sym}	200 mM	25˚C	750 MHz	HNCA	40	< 10%
LacI-O_{sym}	200 mM	25˚C	900 MHz[3]	HNCA	40	< 10%
LacI-O_{sym}	20 mM	40˚C	750 MHz	HNCO	24	97%
LacI-O_{sym}	20 mM	40˚C	750 MHz	HNCA	40	70%
LacI-O_{sym}	20 mM	40˚C	750 MHz	HNCACO	40	No Signal
LacI-O_{sym}	20 mM	40˚C	750 MHz	HNCOCA[4]	40	No Signal

[1] Static field is shown as ^{1}H Larmor frequency
[2] Expressed as % of total number of expected peaks based on primary sequence
[3] Data collected at the University of Colorado, Denver
[4] This experiment is known to exhibit very low sensitivity for large proteins at high field due to unfavorable carbonyl carbon relaxation properties [187, 192]

Though the DNA-free states of LacI have lower molecular weights, they are unstable at temperatures above 25˚C and require high salt for solubilization (> 150 mM) as discussed in Chapter 2. Attempts to assign these states resulted in similar failures (data not shown).

The origin of why these experiments failed is simple: insufficient sensitivity. Low sensitivity arises from the slow molecular re-orientation time for the high molecular weight LacI, diminished performance of cyroprobes due to the use of physiological concentrations of salt in sample preparations, and inherently long data collection times associated with triple resonance experiments (*vide infra*). The latter issue limits signal averaging (shown in Table 3.3 as "number of scans"). While increasing the number of scans *N*-fold results in a gain of S/N by $\sqrt{N}$, the maximum gain in S/N achievable by signal averaging is usually small due to the finite amount of time allowed for a single NMR experiment [193]. For our machines, the limit for a single experiment is 7 days in order to accommodate mandatory instrument maintenance.

This chapter introduces two strategies which maximize sensitivity and thus enable resonance assignments of LacI. The first strategy is an implementation of a so-called "divide and conquer" approach, wherein the individual isolated domains of a protein are assigned separately then the resulting assignments are mapped onto spectra of the intact protein [194]. This strategy has been utilized to assign impressively large systems [195], however, it requires that isolated domains and intact proteins exhibit highly similar chemical shifts which is not always guaranteed [196]. The expression, purification, and preliminary NMR-based characterization of the isolated DBD and RD of LacI have already been introduced in Chapter 2. Since the dimerization interface is localized to the RD, the DBD is a mere 7 kDa while the RD is 58 kDa. At 58 kDa, the RD is still extremely challenging to assign using conventional methods [181] so this approach may not enable resonance assignment on its own.

The second strategy attempts to decrease the inherently long data collection times associated with triple resonance experiments using an alternative method of data sampling known as non-uniform sampling (NUS). While NUS methods have a rich history in biomolecular NMR [197], NUS-derived spectra have historically been plagued by spectral artifacts. Recent advances in methodology have dramatically improved the quality of NUS-derived data without the need for laborious post-processing procedures [198], suggesting that the use of NUS could be beneficial to the study of LacI. The next section introduces NUS in the context of conventional NMR data sampling. Because such methods were not previously employed in our laboratory, their introduction, validation, and implementation were a significant undertaking. Novel developments are described in technical detail in the Section 3.2 of this chapter and later in Chapter 6.

3.1.1 Introduction to non-uniform sampling (NUS)

Before describing NUS, the origin of NMR signals should first be introduced. NMR signals are acquired in the time domain then converted to the frequency domain by use of the Fourier transform. In short, detectable NMR signals arise from precessing transverse (x-y plane) magnetization that is generated during the NMR experiment. The spectrometer detects this magnetization as a time-varying current and the amplitude of the signal is damped by the transverse relaxation rate of the nucleus being probed, hence relating rotational correlation time to signal intensity. Both orthogonal transverse components of the magnetization are detectable and the measurement of each is typically done independently to facilitate frequency sign determination. The observable signal for a single frequency is the linear combination of both transverse components which yields a single complex number:

$$S(t) = M_0 \cos(\omega t) e^{-t/T_2}(t) + iM_0 \sin(\omega t) e^{-t/T_2} = M_x(t) + iM_y(t) = M_0 e^{i\omega t} e^{-t/T_2} \tag{3.1}$$

Where M_0 represents the bulk magnetization and the x,y subscripts refer to the two orthogonal transverse components. The choice of the imaginary component is arbitrary. The resonance frequency is denoted by ω and the transverse relaxation rate is given by $1/T_2$. While the signal is analog, it is digitized and thus sampled as a series of discrete points. Because the Fourier transform is used to obtain the frequency components of the signal, the spacing between points must be uniform. In conventional data sampling, the spacing between points is equal to 1/SW, where SW denotes the “spectral width.” SW can be thought of as the bandwidth or “window of frequencies” being probed. This is in accordance with the Nyquist-Shannon sampling theorem which (approximately) states[8] that if all the frequency components of a given signal lie within a fixed range of SW, the entire signal can be characterized by any series of discrete samples collected with spacing less than 1/SW [193]. The resulting damped time series is referred to as the free induction decay or FID. If more than one frequency is observable, equation 3.1 becomes a summation over all individual frequencies. An example FID for two frequencies is shown in Figure 3.2.

---

[8] More formally, the Nyquist-Shannon theorem states that a continuous signal can be fully sampled if and only if that signal does not contain frequency components larger than ½ the sampling rate [199].

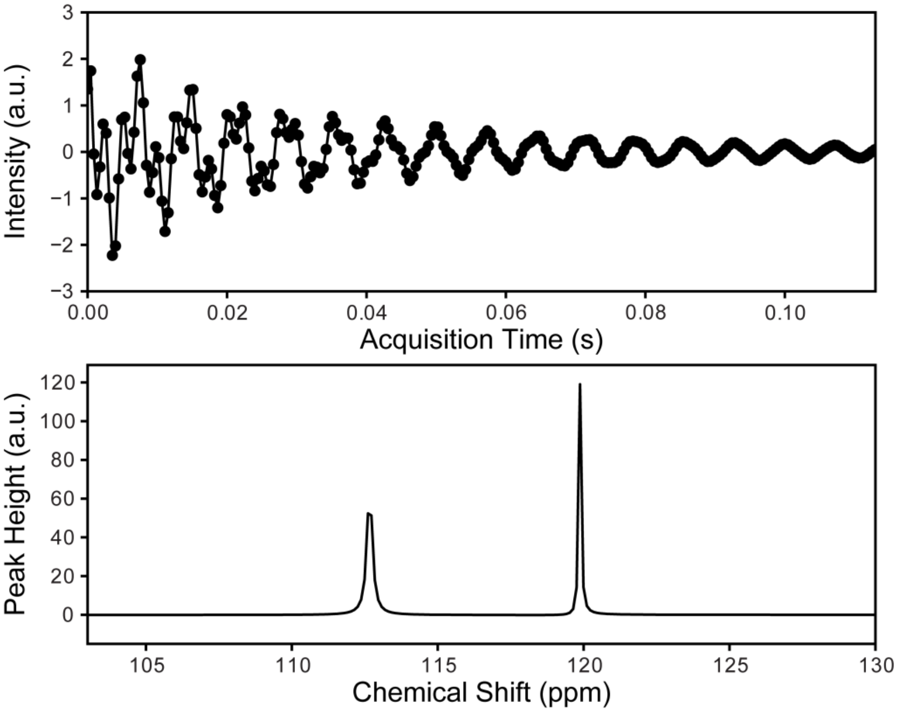


Figure 3.2 Example uniformly sampled FID and spectrum. The top panel shows a simulated FID consisting of 256 real points collected with a uniform spacing of 444 µs. The bottom panel shows the DFT of the top panel. Note that the first point of the FID has been scaled by a factor of 0.5 to prevent a baseline offset in the transformed spectrum [200].

In the case of 1D NMR, sampling of the FID is done directly and the duration of a given experiment is primarily limited by the number of signal averages. However, 1D NMR is limited by resolution which hinders its application to large protein systems that exhibit many observable (and often degenerate) frequencies. Resolution is improved by employing multidimensional methods which evolve couplings between different nuclei. The resulting observable signal is a cross peak defined by multiple chemical shifts thereby decreasing the probability of spectral degeneracy. Multidimensional experiments require so-called “evolution” and “mixing” periods wherein the chemical shift of one nucleus is first evolved then encoded onto another nucleus for detection. For this reason, evolved dimensions are referred to as “indirect” dimensions. Only a single nucleus is detected (for all experiments described in this work, ^{1}H is the directly detected nucleus), however, the obtained signal contains information about all other nuclei whose chemical shifts were allowed to evolve over the course of the experiment. A typical 3D experiment setup is shown in Figure 3.3.

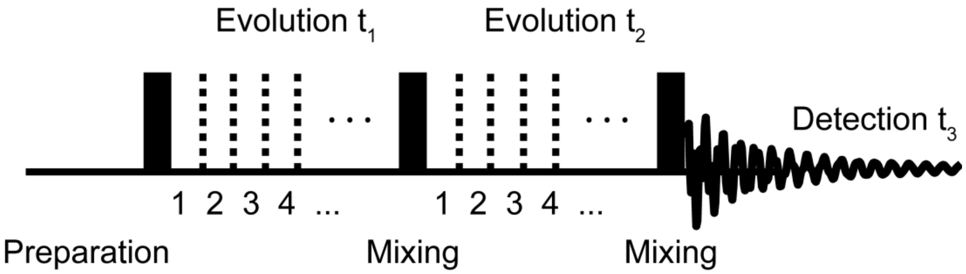


Figure 3.3 Example 3D NMR experiment. The typical multidimensional experiment features a preparation period where magnetization is transferred from nucleus to another.

In order to perform chemical shift evolution, a delay equal to 1/SW must be incremented discretely N times. The value of N is typically determined by the spectroscopist and contributes to the definition of the spectral resolution. Increasing the dimensionality of a given experiment requires increasing the number of evolution periods. For a 3D experiment, which is of interest here, two evolution periods are required and the resulting data is a grid of incremented data samples which are uniformly spaced along each orthogonal axis. Fourier transform of this data results in cross peaks whose positions are within a 3D cube, where the third axis corresponds to the directly detected dimension which is not incremented.

The requirement for uniformly spaced data samples results in protracted data collection times. For example, a 3D experiment consisting of 64 data samples in one indirect dimension and 32 data samples in the other indirect dimension requires collection of 64 x 32 = 2048 FIDs. This number must also be multiplied by the number of signal averages including the duration of time between repetitions (called the recycle delay). The time restriction becomes particularly acute in the case for LacI for the reasons described above and two others:

- The need for long (> 1 s) recycle delays due to amide ^{1}H relaxation properties as a consequence of high molecular weight, perdeuteration, and the use of high static field strength [187]
- The need for large numbers of indirect dimension increments in order to resolve higher density of peaks

We previously showed that the triple resonance experiments required for resonance assignment were too insensitive to be applied to LacI. One could, in principle, improve sensitivity by decreasing the total amount of time required for a given experiment then reinvest the time saved into signal averaging [201]. A general way of reducing the amount of time required for data collection is by bypassing the requirement for uniformly spaced data samples and collecting points non-uniformly. Such non-uniform sampling (NUS) methods have been employed in the context of biomolecular NMR for quite some time [197]. However, it was not until recently that such methods have found routine use. NUS data is intrinsically recalcitrant to the DFT and frequency domain data obtained from NUS time series are obscured by spectral artifacts. A variety of methods have been developed to “reconstruct” data points which were not collected during an NUS experiment, facilitating the use of the DFT [198]. Two critical facets of this approach must be considered:

- How to determine which data points to collect and which to skip
- How to reconstruct the skipped data prior to application of the DFT

Indeed, many methods for addressing the above have been introduced over the last 30 years and the most successful and commonly used are reviewed elsewhere [198]. We favor the sine-weighted Poisson-gap method of data sampling [202] and iterative soft-thresholding (IST) [203, 204] method of reconstruction which will be introduced and discussed briefly below.

3.1.2 Poisson-gap sampling

The quality of spectra obtained from NUS data depends largely on the distribution and intensity of spectral artifacts. While spectral artifacts are unavoidable when using NUS methods, they can be minimized by judicious choice of sampling schedules and reconstruction methodology. The sampling schedule is the list of data samples which are to be collected. It has been shown that several key sampling schedule characteristics promote the minimization of spectral artifacts: (1) randomness in choice of data samples, (2) small gaps between data samples, (3) avoiding large gaps at the beginning or end of FID [202]. To accommodate these criteria, a sine-weighted Poisson distribution of gaps between sampled data points was introduced [202]. The performance of using such sampling schedules was shown to be superior to other sampling methods with respect to the fidelity of reconstructed spectra relative to a uniformly sampled reference. Moreover, the Poisson-gap method of sampling is highly insensitivity to the seed of the random number generator used to satisfy the randomness requirement which results in highly consistent performance across a variety of sampling schedules. It should be noted that schedule-specific performance has been observed in certain extreme cases (e.g. in the limit of very high sampling densities) [205].

3.1.3 Iterative-soft thresholding reconstruction

The success of the DFT with respect to transforming NUS data depends critically on the quality of data reconstruction. Directly applying the DFT to NUS data without reconstruction results in artifact-ridden data which are a poor approximation of what would normally be obtained using uniformly sampled data (Figure 3.4). Typically, peak heights and lineshapes become distorted and noise-like artifacts that arise from the point-spread function contaminate NUS data [206].

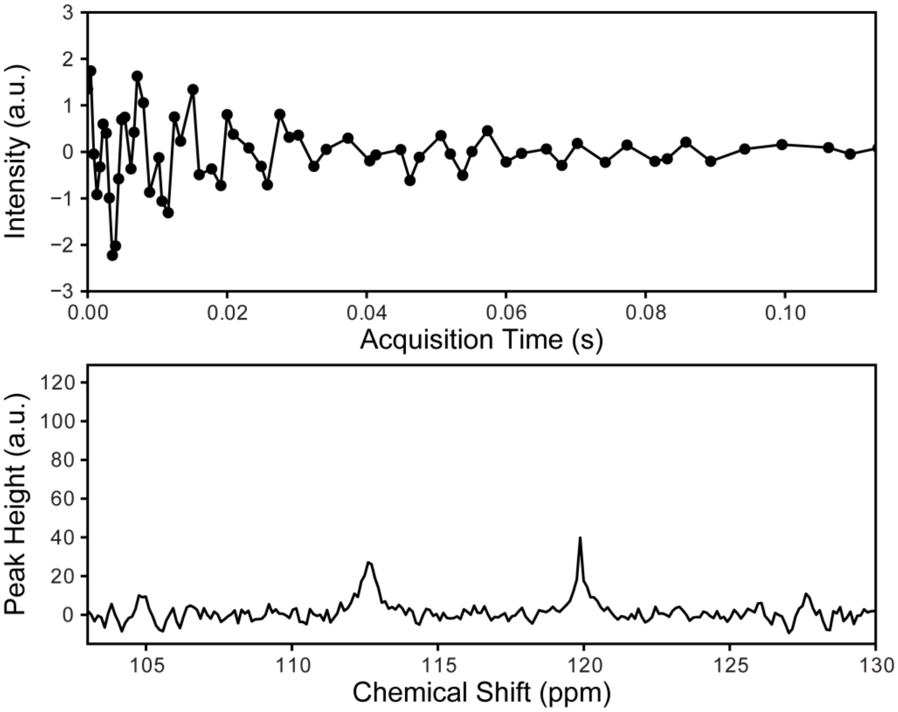
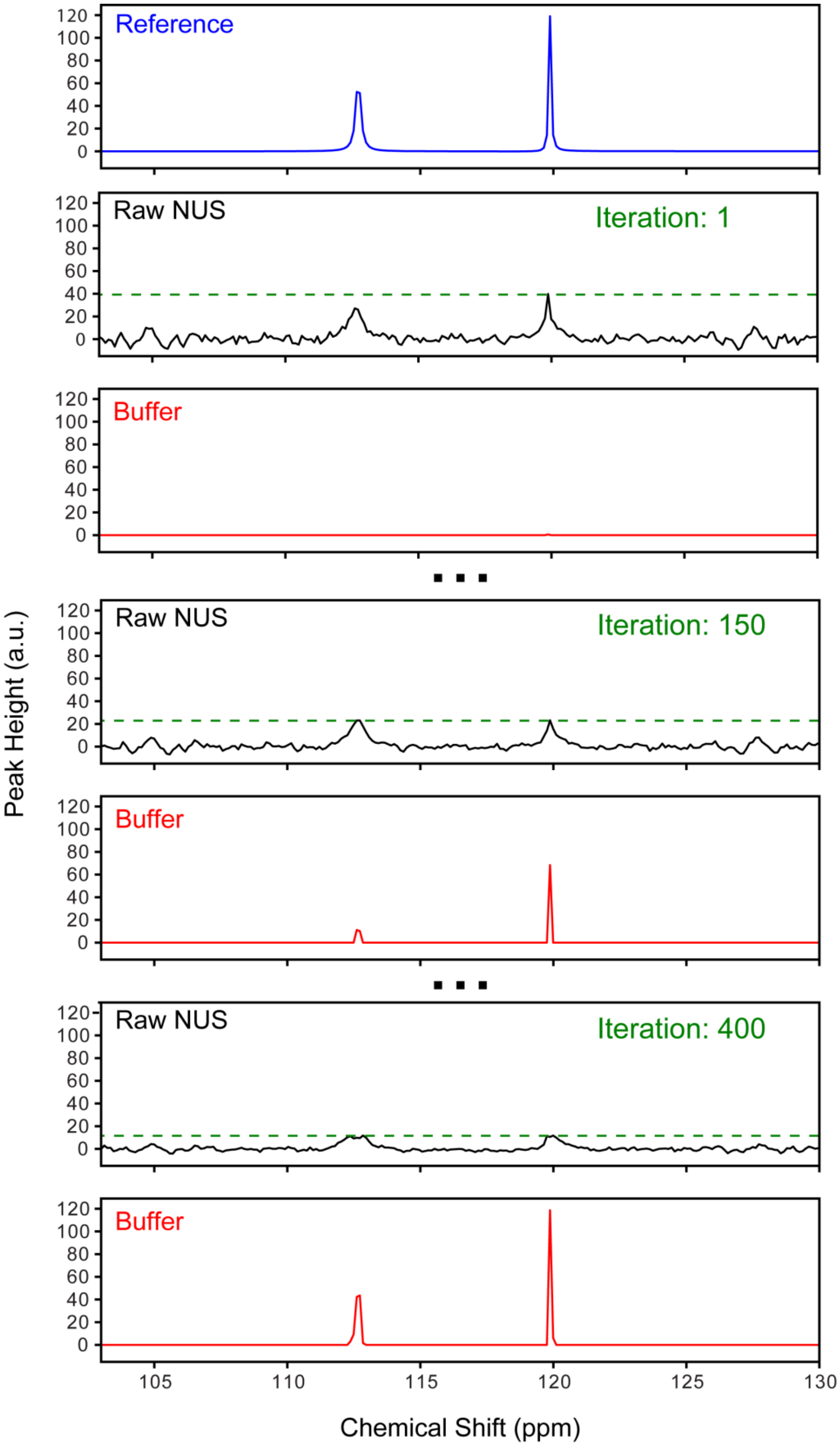


Figure 3.4 Example non-uniformly sampled FID and spectrum. The top panel shows a simulated NUS FID consisting of 64 real points (25% of 256 real points, see Figure 3.2). The points were selected based on a sine-weighted Poisson distribution. The bottom panel shows the DFT of the top panel. The first point of the FID has been scaled by a factor of 0.5 to prevent a baseline offset in the transformed spectrum. The noise-like features in the bottom panel are spectral artifacts.

While many algorithms have been developed to reconstruct missing data points, the most successful have historically been the most computationally expensive. Given the required volume of data necessary for resonance assignment, such methodologies would not be very useful for our purposes. An algorithm recently introduced to the NMR community, Iterative Soft Thresholding (IST), reconstructs data using a series of forward and inverse Fourier transforms. Since computationally efficient or “fast” Fourier transform algorithms are well known [207], the total computational time required for IST reconstruction is minimal.

The general IST procedure is illustrated in Figure 3.5 using synthetic data devoid of instrument noise. The synthetic NUS time series is generated using two unique frequencies with variable amplitude and with variable transverse relaxation rates. The data used in the IST simulation are the same as those presented in Figure 3.4—comprising 25% of 256 real points. The first step of the IST algorithm is DFT of the NUS data. This results in low quality data with prominent artifacts. A threshold is set as a percentage of the maximum intensity of the frequency domain spectrum. For the example used here, the threshold is 98.5%. Every value above the threshold is moved and stored in a buffer that has been initialized to zero. The remaining truncated frequency domain spectrum is then inverse Fourier transformed and the points that were originally skipped are set to zero. The process is then iterated as shown in the bottom four panels of Figure 3.5. In addition to the relative ease of computation, IST provides spectral reconstructions with minimal artifacts. In principle, a lack of high intensity spectral artifacts should benefit NMR studies of LacI which are inherently limited by sensitivity.

The number of iterations performed prior to termination is usually determined by a well-defined convergence criterion. IST belongs to the “compressed sensing” (CS) family of signal reconstruction methods and these methods utilize the minimization of a regularization term such as the l_p -norm (for $p \in (0,1)$), of the preferred solution as a termination criterion [208-210]. In simplistic terms, a regularization term is a constraint used to solve an underdetermined system of linear equations. The details of regularization are beyond the scope of this work and have been described thoroughly elsewhere [211]. In the case of IST, l_1-norm minimization can result in very high-fidelity reconstructions if the data are sufficiently sparse (meaning that only a few values in a given spectrum are non-zero). While NMR data and instrument noise are not formally sparse [212], IST has been demonstrated to produce high quality spectral reconstructions both with respect to peak frequency and height. The relationship between spectral sparseness and peak height is more closely examined in Chapter 6.



Reference
Raw NUS
Iteration: 1
Buffer
Raw NUS
Iteration: 150
Buffer
Raw NUS
Iteration: 400
Buffer
Peak Height (a.u.)
0
20
40
60
80
100
120
Chemical Shift (ppm)
105
110
115
120
125
130

Figure 3.5 Illustration of the IST algorithm. The top panel shows a "Reference" spectrum derived from DFT of a FID sampled with 256 real points with uniform spacing (see Figure 3.2). The remaining panels show "Raw NUS data", i.e. NUS frequency domain data derived using 25% sampling (see Figure 3.4) and the "Buffer" where the reconstructed spectrum is stored. The process follows the description provided in the main text. Following 400 iterations of IST, it is clear the reference spectrum is closely approximated by the buffer and the noise-like artifacts originally present in the raw NUS frequency domain data are eliminated. The simulation was generated in-house using Python and the procedure published by Hyberts SG, *et al*. [204]. The total computation time for this simulation was 10.06 sec using a 2.67 GHz Intel Core i7 desktop CPU.

3.1.4 NUS improves the sensitivity of NMR

Sections 3.1.1-3.1.3 introduce and briefly describe an NUS sampling scheme and reconstruction algorithm that enable robust, computationally efficient, and largely artifact-free spectral reconstructions. However, the question at hand is whether the combination of these methods can be used to improve the sensitivity of triple resonance experiments required for assigning LacI.

In general, the impact of NUS on sensitivity or, more specifically, S/N, has not been documented thoroughly. This stems partly from inconsistencies with respect to how S/N is defined (particularly for reconstructed data where noise and spectral artifacts are convolved) as well as uncertainty regarding how to accurately quantify S/N for biological samples [201]. Previous work has shown that significant gains in S/N can be achieved when 1/k of the uniform grid is sampled using NUS and the number of scans is increased by k-fold. Accordingly, such an approach does not result in time savings. With respect to triple resonance experiments, it was shown that sampling 1/16th the number of indirect data samples and increasing the number of scans by 16-fold resulted in a 22-fold increase in the probability to detect weak peaks (S/N increases are not used as metrics here due to the non-Gaussian nature of noise in NUS-derived spectra) [201]. Even larger gains were demonstrated for more aggressive undersampling. These data suggest that NUS may provide the gain in sensitivity required to enable resonance assignment of LacI.

3.2 Materials and Methods

3.2.1 NMR sample preparation

NMR samples were prepared as described in Chapter 2 and the buffers used to collect NMR data are listed in Table 2.1.

3.2.2 NMR spectroscopy

For all experiments, the temperature was calibrated using methanol [213] prior to data collection and all chemical shifts were referenced using DSS [214].

Modification of pulse sequences for NUS. A variety of conventional and TROSY-triple resonance experiments (described below) were modified for NUS. Our laboratory utilizes TopSpin 2.1 software which does not feature the latest automated NUS feature for data collection on Bruker machines. As such, pulse sequences were manually coded to utilize a non-uniform increment. This involved several core modifications to the pulse sequence code. Sampling schedules were read in as variable pulse ("VP") or variable counter ("VC") lists. Schedules were generated as lists of dwell multiples which describe the increment number. For example a number of "0" in the list corresponds to zero increments (*i.e.* the first point) whereas a number of "3" corresponds to three increments and so on. Phases for coherence and quadrature selection were calculated based on the increment multiple. All modified pulse sequences were verified by comparison with uniformly sampled data for the model protein ubiquitin (data not shown).

Many TROSY-triple resonance experiments had not been executed in our laboratory before and required development independent of the conversion to NUS. Pulse sequences for TROSY-triple resonance experiments from the Bruker pulse sequence library were highly unstable (< 24 hours) due to an incorrect implementation of ^{2}H decoupling during carbon evolution periods. Corrected ^{2}H decoupling schemes were introduced to all TROSY-triple resonance experiments based on previously published work [192, 215] which eliminated this problem.

Resonance assignment of apo LacI DBD. All NMR data were collected at 298 K on

Bruker Avance III operating at 750 MHz ^{1}H Larmor frequency. Since the DBD is small (62 residues) and can be concentrated > 1 mM, all triple resonance experiments were collected with uniform sampling. The following experiments were used for assignment of the backbone ^{1}H-^{15}N resonances: HNCO [216], HN(CA)CO [217], HNCA [216], HN(CO)CA [218], HNCACB [219], CBCACONH [220], and 3D ^{15}N NOESY-HSQC [221]. Methyl ^{1}H-^{13}C resonances were assigned using HCCH-TOCSY [222] and HCC(CO)NH-TOCSY [223].

Resonance assignment of the LacI DBD-O_{sym} complex. Though historically, LacI DBD-O_{sym} complexes were prepared with 2:1 stoichiometry [83], we decided to prepare our complexes in 2:1.5 stoichiometry to ensure saturation of the DBD. This was based on a ligand saturation calculation assuming an approximate binding affinity of ~1 μM which had been reported in the literature previously for DBD (residues 1-62) [87]. Analysis of the ^{1}H-^{15}N HSQC spectrum indicated that DBD was indeed fully saturated with O_{sym} at this stoichiometry. All NMR data were collected at 303 K on Bruker Avance III operating at 500 MHz ^{1}H Larmor frequency. Most triple resonance experiments were collected with uniform sampling with one exception described below. Backbone ^{1}H-^{15}N resonances were assigned using: HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB, CBCACONH, and 3D ^{15}N NOESY-HSQC. Methyl ^{1}H-^{13}C resonances were assigned using a multiple quantum (MQ)-methyl TOCSY [190] optimized for high molecular weight systems. The MQ-methyl TOCSY was collected with 10% NUS and reconstructed with 400 iterations of IST using a threshold of 98%.

Resonance assignment of the LacI RD-IPTG Complex. All NMR data were collected at 298 K on Bruker Avance III operating at 750 MHz ^{1}H Larmor frequency or 600 MHz ^{1}H Larmor frequency. A suite of TROSY triple-resonance experiments [187] were collected: TROSY-HNCO, TROSY-HN(CA)CO, TROSY-HNCA, TROSY-HN(CA)CB, and 3D ^{15}N NOESY-TROSY HSQC were collected at 750 MHz using 10%-15% NUS. Data were reconstructed with 400 iterations of IST using a threshold of 98%. TROSY-HN(CO)CA and TROSY-HN(COCA)CB were collected at 600 MHz and reconstructed identically.

Resonance assignment of the apo LacI RD. All NMR data were collected at 298 K on

Bruker Avance III operating at 600 MHz ^{1}H Larmor frequency. TROSY-HNCO and TROSY-HNCA were collected using 10% NUS. Data were reconstructed with 400 iterations of IST using a threshold of 98%.

Resonance assignment of the LacI-IPTG Complex. All NMR data were collected at 298 K on Bruker Avance III operating at 750 MHz ^{1}H Larmor frequency or 600 MHz ^{1}H Larmor frequency. TROSY-HNCO, TROSY-HNCA, and TROSY-HN(CA)CB were collected using 5%-10% NUS. Data were reconstructed with 400 iterations of IST using a threshold of 98%.

Resonance assignment of the LacI-O$_{sym}$ Complex. All NMR data were collected at 303 K on Bruker Avance III operating at 750 MHz ^{1}H Larmor frequency or 600 MHz ^{1}H Larmor frequency. TROSY-HNCO, TROSY-HN(CA)CO, TROSY-HNCA, and TROSY-HN(CA)CB were collected at 750 MHz using 10% NUS. Data were reconstructed with 400 iterations of IST using a threshold of 98%. TROSY-HN(CO)CA and TROSY-HN(COCA)CB were collected at 600 MHz using 5-10% NUS and reconstructed identically.

3.2.3 Data analysis

All data was processed using the NMRPipe/NMRDraw program [159]. Visualization was performed using SPARKY [224]. Cross peaks were manually picked for all data. Because of the relatively large number of cross peaks, special procedures were enacted to ensure that assignments were performed correctly. First, all assignments were performed independently by two separate people: me and another student. Upon completion of assignments, both of us convened to compare results. Inconsistencies were then investigated more closely and rectified. Once a preliminary set of assignments were established, they were verified using the program PINE [225] and its built-in module, PECAN [226], as well as the program TALOS-N [227]. PINE employs an exhaustive search algorithm for establishing assignments. Both PECAN and TALOS-N use chemical shifts to determine secondary structure propensities using verified relationships. These programs will be introduced formally and described in more detail in the Results section of this chapter. The secondary structure predictions were then compared to the crystal structures (if available). In general, inconsistencies between separate sets of resonance assignments were few and most originated from mistakes in peak annotations.

3.3 Results and Discussion

3.3.1 Assessment of the "divide and conquer" approach

Comparison of the apo DBD and apo RD ^{1}H-^{15}N HSQC spectra to the ^{1}H-^{15}N HSQC spectrum of apo LacI reveal that the spectra of the individual domains and the spectra of the intact protein are superimposable to high degree. The exact same result is observed when comparing the same spectra of the apo DBD, RD-IPTG complex, and LacI-IPTG complex. The spectral comparisons are shown in Figure 3.6a-b. The structural and functional implications of this analysis are discussed in depth in Chapter 4 and will not be addressed further here since resonance assignment is the focus of this chapter.

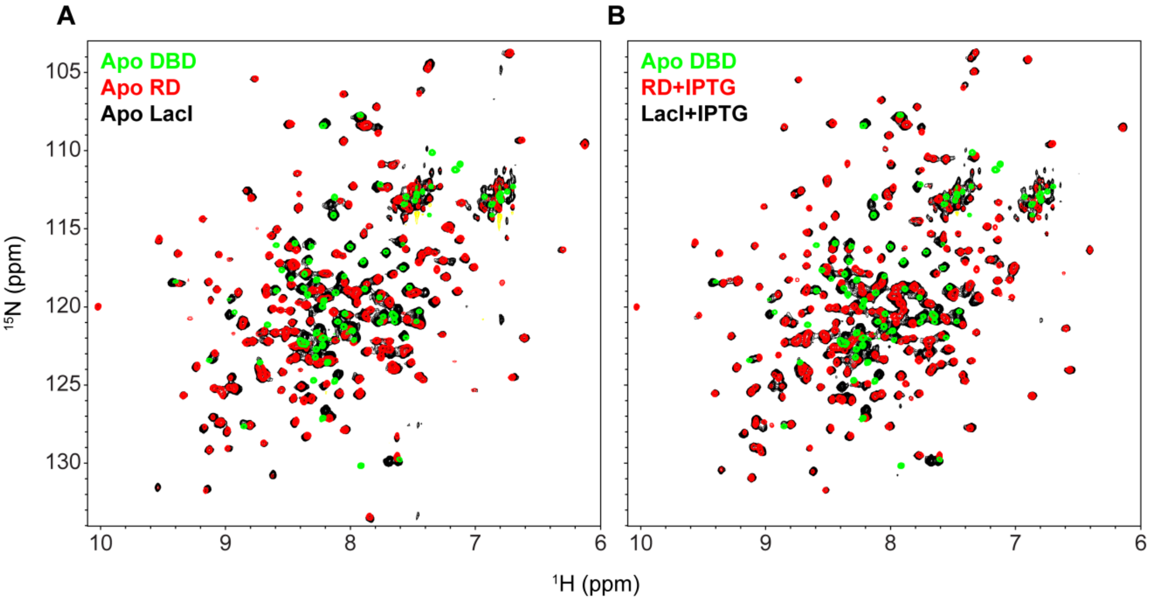


Figure 3.6 Assessment of "divide and conquer" approach for operator-free states. (A) Overlay of ^{1}H-^{15}N HSQC spectra of apo DBD (green), apo RD (red) and apo LacI. (B) Overlay of ^{1}H-^{15}N HSQC spectra of apo DBD (green), RD-IPTG complex (red), and LacI-IPTG complex (black). The ^{1}H and ^{15}N chemical shifts of the DBD were corrected for TROSY-selection. Negative contours are shown in yellow.

Less favorable results are obtained when comparing the operator-bound states. Figure 3.7a shows an overlay of ^{1}H-^{15}N HSQC spectra of the LacI-O_{sym} complex, apo RD, and DBD-O_{sym} complex. Figure 3.7b shows an overlay of ^{1}H-^{15}N HSQC spectra of the LacI-O_{sym}-IPTG complex, RD-IPTG complex, and DBD-O_{sym} complex. Qualitatively, it is apparent the LacI-O_{sym}-IPTG complex is recapitulated better than the LacI-O_{sym} complex. The structural and functional implications of this are discussed in depth in Chapter 5 and will not be addressed here further since the focus of this chapter is resonance assignment.

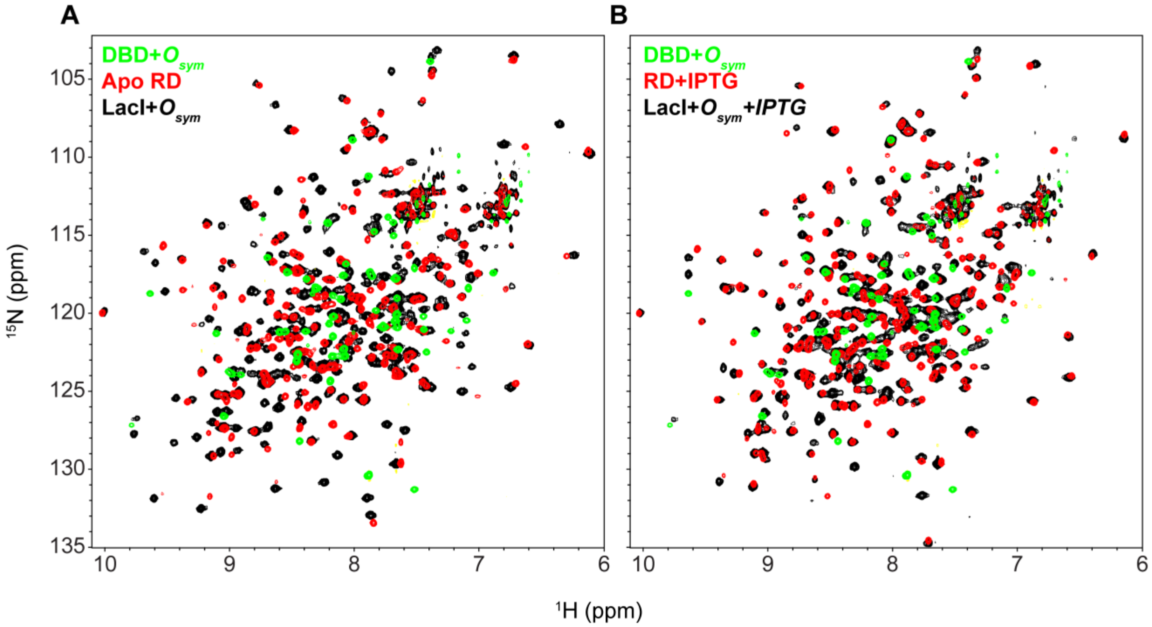


Figure 3.7 Assessment of "divide and conquer" approach for operator-bound states. (A) Overlay of ^{1}H-^{15}N HSQC spectra of DBD-O_{sym} complex (green), apo RD (red) and LacI-O_{sym} complex. (B) Overlay of ^{1}H-^{15}N HSQC spectra of DBD-O_{sym} (green), RD-IPTG complex (red), and LacI-O_{sym}-IPTG complex (black). Negative contours are shown in yellow.

Based on the results above, it is clear that the divide and conquer approach will be more useful when assigning the operator-free states. Importantly, the ternary LacI-O_{sym}-IPTG complex spectrum exhibits a higher degree of similarity to the spectra of its constituent domains than the binary LacI-O_{sym} complex spectrum which suggests it may be amenable to a divide and conquer approach as well. This is fortuitous considering that the ternary complex is poorly characterized with respect to structure. Therefore, we may be able—for the first time—to access the structural properties of the ternary complex at atomic resolution.

3.3.2 Assignment of the apo LacI DBD

Backbone and methyl resonance assignment of apo DBD was straightforward. At ~7 kDa and ideal sample concentrations ~ 1 mM, data could be collected using uniform sampling without the need for additional signal averaging. However, the isolated DBD is not an idealized model system due to unfavorable chemical exchange-induced line broadening which has been characterized previously [90]. Despite this, 100% of ^{1}H-^{15}N amide resonance and 96% of (isoleucine, leucine, and valine) ILV methyl resonance were obtained. Resonance assignments are listed in Appendix 3.

3.3.3 Assignment of the LacI DBD-O_{sym} complex

Backbone and methyl resonances for the LacI DBD-O_{sym} complex were available from the BMRB, however, these assignments were obtained from data collected at a much higher temperature, 42˚C and on a slightly different construct of DBD. Therefore, a full suite of triple resonance experiments were also collected for the LacI DBD-O_{sym} complex. Though the complex is ~29 kDa, 98% of ^{1}H-^{15}N resonance assignments could be obtained using conventional pulse sequences collected with uniform sampling. Methyl assignments were difficult to obtain using the same experiments performed on the apo DBD due to the substantially higher molecular weight of the complex. As such, a separate experiment optimized for high molecular weight proteins was performed to obtain methyl resonance assignments. Using this experiment, 100% of ILV ^{1}H-^{13}C methyl assignments were obtained. Resonance assignments are listed in Appendix 3.

3.3.4 Assignment of the LacI RD-IPTG complex using NUS

While the spectra of both the apo and IPTG-bound RD are dispersed, the apo state exhibits only ~80% of the expected amide cross peaks based on analysis of the TROSY-HNCO which accounts for degeneracies in amide $^{1}H/^{15}N$ frequencies. The origin and functional consequences of this observation are discussed in detail in Chapter 4. With such few resonances, assignment using sequential inter-residue connectivities would be difficult. Similar analysis of the IPTG-bound RD, however, revealed that it exhibited the expected number of resonances (within ~8%) and was therefore a good candidate for assignment.

At 58 kDa, the RD-IPTG complex was among one of the largest proteins to be assigned, particularly for data collected at 25°C and at sample concentrations of ~100 μM. Indeed, low S/N was observed for most uniformly sampled triple-resonance experiments (data not shown). NUS experiments where time savings was reinvested into signal averaging were thus performed in order to improve the quality of data. Table 3.4 shows the specific experimental parameters used for each triple-resonance experiment performed on the RD-IPTG complex.

Table 3.4 Summary of NUS triple-resonance experiments for RD-IPTG complex

Experiment	**Static Field**[1]	**Recycle Delay (s)**	**Scans**	**^{15}N Points (Real)**	**^{13}C Points (Real)**	**NUS (%)**	**Time**	**Uniform Time**[2]
TROSY-HNCO	750	1.8	32	40	64	10.0	18 hours	7 days, 12 hours
TROSY-HN(CA)CO	750	1.8	176	40	64	10.0	4 days, 5 hours	42 days, 2 hours
TROSY-HNCA	750	1.8	80	40	72	10.0	2 days, 3 hours	21 days, 6 hours
TROSY-HN(CO)CA	600	1.5	128	32	56	10.3	1 day, 21 hours	18 days, 5 hours
TROSY-HN(CA)CB	750	1.8	160	40	88	8.97	4 days, 18 hours	52 days, 17 hours
TROSY-HN(COCA)CB	600	1.5	128	32	72	9.5	2 days, 6 hours	23 days, 17 hours
^{15}N 3D NOESY-TROSY HSQC	750	1.5	32	56	72 (^{1}H real points)	15.0	1 day, 16 hours	11 days, 2 hours

[1] Static field strength shown as ^{1}H Larmor frequency
[2] Equivalent data collection time if performed using uniform sampling

We have previously shown failed attempts to assign the *LacI-O_{sym}* complex using uniform sampling. Our strategy there involved maximizing the number of scans and minimizing the number of indirect increments. Even when the number of indirect dimension data samples was minimized, the maximum number of scans that could be performed within the allowable instrument time was 40. As evidenced by Table 3.4, the total number of scans could be increased substantially using NUS with concomitant increases in the number of indirect dimension increments. Every single triple resonance experiment performed on the RD-IPTG complex would have exceeded the maximum allowable data collection time if uniform sampling had been implemented. Using NUS, a total data collection time of just 17.5 days was necessary for the entire suite of high sensitivity and high resolution triple resonance experiments. Had the same experiments been collect with uniform sampling, the total data collection time would have been 177 days.

As mentioned previously, it is difficult to quantify a true S/N ratio for NUS data given the non-Gaussian nature of spectral noise which originates from reconstruction [201]. Moreover, such a comparison could not be done rigorously without a set of uniformly sampled reference data. However, knowledge of the total gain in S/N is not necessary for our purposes. By using NUS, we merely sought to obtain *sufficient* S/N to enable resonance assignment. The only way to determine if this condition was met was by assessing the extent of assignment that could be obtained with the data. Figure 3.7 shows an example strip-plot outlining the inter-residue connections obtained from analysis of a TROSY-HNCA experiment. These plots are commonly used to simplify visualization of the 3D-data objects obtained from triple-resonance experiments. Each strip is a 2D slice obtained from a given point along the third dimension (see figure legend for more details). Similar plots could be generated for other triple-resonance experiments but are not shown for brevity since their information content is redundant. It is clear that the quality of the NUS data is exceptional.

Overall, the NUS-derived triple-resonance data was indeed of sufficient quality to assign the ^{1}H-^{15}N resonances of the RD-IPTG complex because 90% of the expected resonances were assigned. As will be further explored in Chapter 4, the remaining ~10% were not assigned due to line broadening arising from conformational dynamics. Resonance assignments are listed in Appendix 3.

3.3.5 Verifying the assignments of the LacI RD-IPTG complex

The high quality data afforded by NUS enabled resonance assignment of the RD-IPTG complex. However, given the sheer quantity of cross peaks (> 2000) and lack of reference data in the BMRB, we sought a means of verifying the accuracy of the assignments. Previous studies have done this by combining triple resonance data with redundant but less ambiguous four dimensional data [192]. However, given LacI's unfavorable properties, sensitivity was too limiting to employ such an approach. Other means of verification include site-directed mutagenesis or amino acid-specific labeling. While such methods can be highly effective, they are costly and laborious. We did not deem such methods necessary given the relatively clean dispersion of cross peaks. Therefore, a simpler means of verification was applied. First, the data was analyzed using a highly effective automatic assignment algorithm known as "**P**robabilistic **I**nteraction **N**etwork of **E**vidence" or PINE [225]. Briefly, PINE identifies triplets of sequentially connected residues and ranks the assignment of many overlapping triplets based on an energy function. The algorithm essentially works as a computationally feasible exhaustive search. The results from PINE were in fantastic agreement with the manually assigned data. PINE was able to assign 211 of the expected 240 manually assigned non-proline ^{1}H-^{15}N amide resonances with probabilistic confidences > 0.9. Though PINE also made additional assignments of lower probabilities they were excluded from this analysis. Of the 211 high-confidence assignments made by PINE, 206 matched the manually assigned resonances yielding a correspondence of 97.6%. The few discrepancies were resolved by analyzing the ^{15}N 3D NOESY-TROSY HSQC using the crystal structure (PDB: 2P9H) of the LacI RD-IPTG complex to define upper limits for inter-atomic distances.

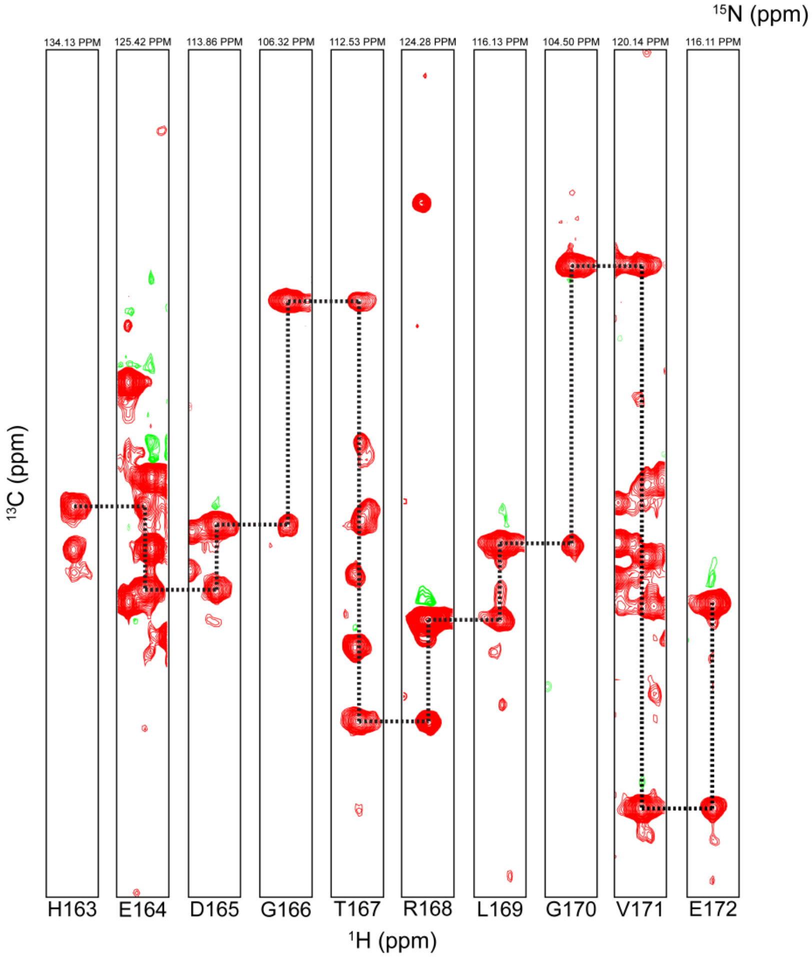


Figure 3.8 Strip plot of NUS-TROSY HNCA collected on the RD-IPTG complex. The ^{1}H-^{13}C 2D planes are shown for a variety of ^{15}N slices in order to illustrate the derivation of sequential connections between residues using alpha carbon chemical shifts which are shown by the black dotted line. Assignments are also shown as labels but formally require additional (though highly similar) data derived from triple-resonance experiments which evolve alternative ^{13}C nuclei. Negative peaks are shown in green and are likely due to artifacts as a consequence of NUS reconstruction and/or a slight first order phase error in the carbon dimension.

A final means of verification utilized the well-characterized relationship between chemical shifts and secondary structure [228]. Two programs were used to analyze our resonance assignments in order to generate secondary structure models: “**P**rotein **E**nergetic **C**onformational **A**nalysis from **N**MR chemical shifts” (PECAN) [226] and “**T**orsion **A**ngle **L**ikeliness **O**btained from **S**hift” using **N**eural networks (TALOS-N) [227]. Both programs combine information from chemical shifts and amino acid sequence. The PECAN algorithm defines an optimal set of limited chemical shifts for a specific residue type whereas the TALOS algorithm simultaneously considers all chemical shifts for each residue type using a knowledge-based approach. Though both programs provide essentially the same information, the primary differences between their outputs are expected to arise due to variations in side-chain conformations and tertiary structure. Using the crystal structure (PDB: 2P9H) of the LacI RD-IPTG complex as a reference, it was found that both PECAN and TALOS-N predicted the correct secondary structure distribution as shown in Figure 3.9. This further confirms the validity of the chemical shift assignments.

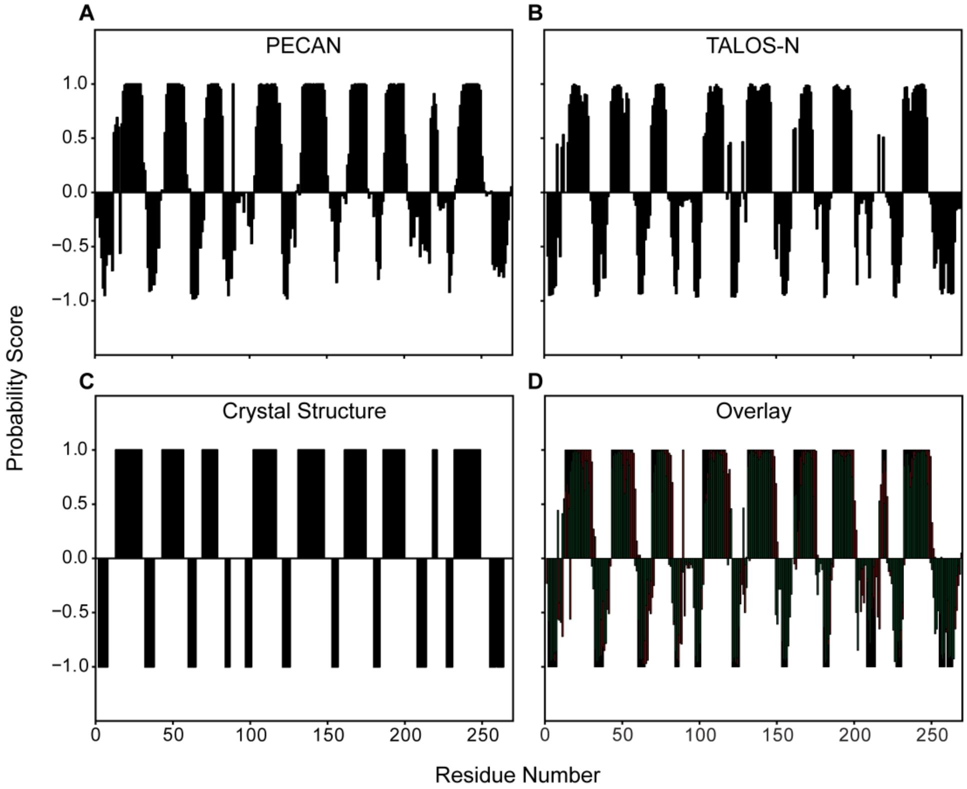


Figure 3.9 Secondary structure prediction for RD-IPTG complex. Secondary structures are shown as a “probability score” where 1.0 indicates a totally helical conformation, -1.0 indicates a totally sheet conformation, and 0.0 indicates totally coil conformation. (A) Predicted secondary structure from resonance assignments derived from the PECAN algorithm. (B) Predicted secondary structure from resonance assignments derived from the TALOS-N algorithm. (C) Observed secondary structure from the crystal structure (PDB: 2P9H). (D) Overlay of A,B, and C. The residue numbering has been offset to start at 1.

3.3.6 Mapping the assignments of the LacI RD-IPTG complex onto spectra of the LacI-IPTG complex

The divide and conquer approach to resonance assignment is only useful if the assigned chemical shifts of the individually parsed domains can be mapped back onto the spectra of the intact multi-domain protein successfully. We have demonstrated that the divide and conquer approach and NUS enabled assignment of both the DBD and RD-IPTG complex to an impressively high degree. Though Figure 3.6 shows an essential identity between the chemical shifts of the individual DBD and RD relative to those of intact LacI, a small set of triple resonance data was collected on the LacI-IPTG complex for confirmation: TROSY-HNCO, TROSY-HNCA, and TROSY-HN(CA)CB: these experiments are the most sensitive with respect to other experiments that evolve the same carbon nuclei. Alone, these experiments would be insufficient for obtaining *de novo* assignments, however, they provide sufficient information for confirming mapped chemical shifts. Relative to the isolated RD, the quality of the data for LacI-IPTG complex is low—even when NUS is employed as illustrated by Figure 3.10. Despite this, of the expected 319 non-proline amide ^{1}H-^{15}N resonances based on the primary sequence, 281 or 88% could be assigned. Missing assignments entirely stemmed from missing DBD peaks, likely due to HX since data collection was performed at pH = 7.4.

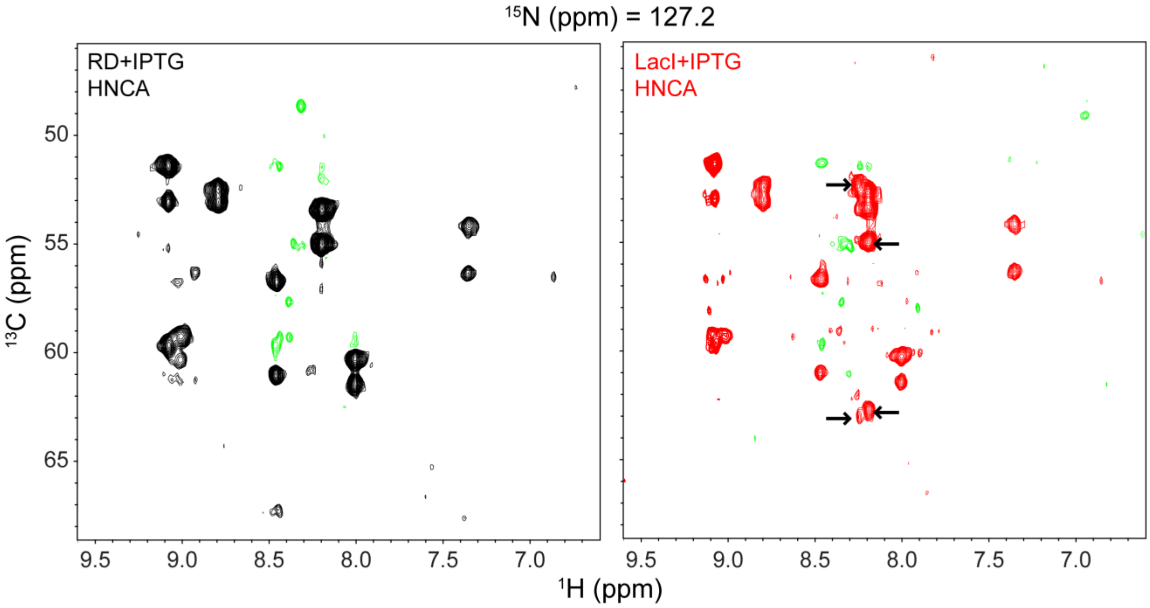


Figure 3.10 Mapping resonance assignments. ^{1}H-^{13}C 2D planes are shown for a single ^{15}N frequency from NUS TROSY-HNCA experiments collect on RD+IPTG (left, black) and LacI+IPTG (right, red). This comparison illustrates the final step of the "divide and conquer" assignment method wherein assigned chemical shifts of an isolated domain are mapped onto spectra of the intact protein. Note the degradation in data quality for the LacI+IPTG complex due to increase salt concentration and molecular weight. Cross peaks that originate from the DBD are easily identifiable and indicated by black arrows. Negative contours are shown in green and arise from noise/NUS-related artifacts.

3.3.7 Assignment of the apo LacI RD

Assignment of the apo state of the RD was performed via assignment mapping. A sparse set of triple-resonance data was collected on the apo RD, analogous to the approach taken to assign the LacI-IPTG complex. Specifically, TROSY-HNCO and TROSY-HNCA experiments were collected at 600 MHz. Assignments were mapped from the RD-IPTG spectra via comparison of intra-reside alpha carbon chemical shifts, inter-residue alpha carbon chemical shifts, and inter-residue carbonyl carbon chemical shifts. In general, most carbon chemical shifts exhibited only slight deviations between the apo and IPTG-bound states (~0.2 ppm). This high degree of correspondence permitted definitive assignment of 92% of the *observable* amide cross peaks of the apo RD or about 74% of the total number of expected amide cross peaks base on amino acid sequence. Resonance assignments are listed in Appendix 3.

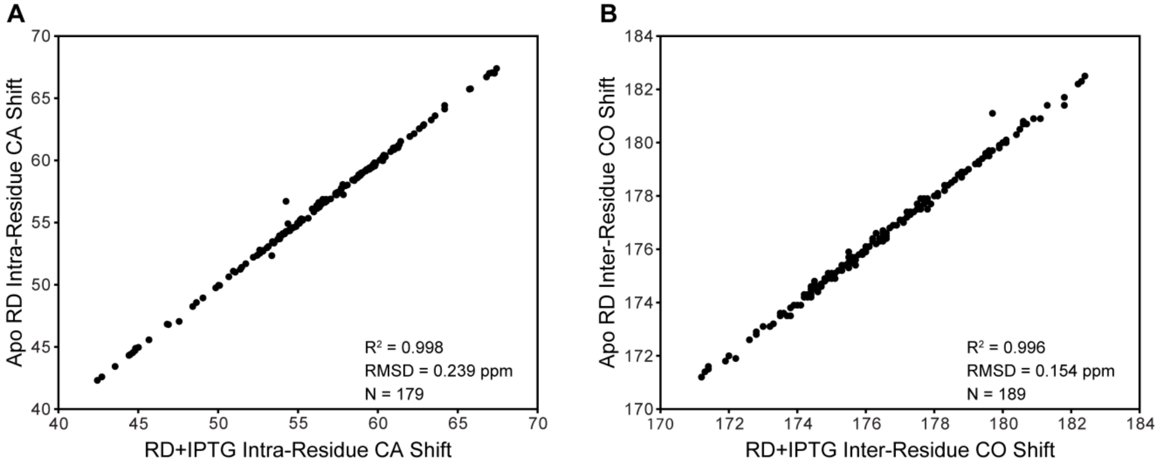


Figure 3.11 Mapping resonance assignments for the apo state (A) Correlation between intra-residue alpha carbon chemical shifts between the apo and IPTG-bound RD spectra. (B) Correlation between inter-residue carbonyl carbon chemical shifts between the apo and IPTG-bound RD spectra.

3.3.8 Assignment of operator-bound states

As previously discussed in this Section 3.3.1, spectra of the operator-bound states of LacI are poorly recapitulated by spectra of the individual RD and DBD. Moreover, since assignments of the apo RD were unobtainable due to its unfavorable relaxation properties, very few assignments could be mapped. Therefore, a full suite of triple resonance experiments was collected on the LacI-O_{sym} complex. The experiments collected included: TROSY-HNCO, TROSY-HN(CA)CO, TROSY-HNCA, TROSY-HN(CO)CA, TROSY-HN(CA)CB, and TROSY-HN(COCA)CB. NUS was utilized in a manner analogous to that presented for assigning the RD-IPTG complex. Unfortunately, the TROSY-HN(CA)CO yielded < 10% of the expected cross peaks and was of little utility. This is unsurprising considering that this experiment is notoriously inefficient for high molecular weight proteins due to unfavorable relaxation effects [187, 192]. Additionally, the TROSY-HN(CA)CB and TROSY-HN(COCA)CB only yielded 58% of the total number of expected cross peaks. Using this limited data set and the chemical shifts of the DBD-O_{sym} complex, and employing a similar chemical shift mapping procedure to the spectra of the other functional states of LacI, ~72% of the ^{1}H-^{15}N amide resonances of the LacI-O_{sym} complex could be assigned. Importantly, of the 62 residues that comprise the DBD, 44 could be assigned, including several key residues in the hinge helix (to be discussed further in Chapter 5). Though this coverage may seem disappointingly low, it is actually entirely consistent with other published assignment statistics. As of 2013, of the 6 proteins with molecular weights > 65 kDa with assignments deposited in the BMRB, none had been assigned > 90% and half were assigned < 60%.

The LacI-O_{sym}-IPTG ternary complex exhibited spectra which were better recapitulated by the spectra of the RD-IPTG complex and the DBD-O_{sym} complex. Therefore, assignments could be made on the basis of spectral comparison with reasonable confidence. Triple resonance experiments were not collected for the ternary complex since the data for the binary complex was of low quality. It will be shown in Chapter 5 that several of the assignments relevant for characterizing the structure and dynamics of the LacI-O_{sym}-IPTG complex could be confirmed simply by comparing the spectra of the LacI-O_{sym} complex, RD-IPTG complex, and DBD-O_{sym} complex. We refrain from assigning a percentage of assignments for this state at the moment.

3.4 Conclusions and Future Directions

The work presented in this chapter sought to assign the ^{1}H-^{15}N amide resonances of LacI in all of its functional states. Initial attempts to assign LacI were hindered due to the low sensitivity of triple-resonance experiments. This was largely overcome by employing the "divide and conquer" technique and NUS. NUS was required because the isolated RD, at 58 kDa and under the experimental conditions necessary for proper protein solubilization, was recalcitrant to conventional NMR experiments.

A complete suite of NUS TROSY-based triple resonance experiments including TROSY-HNCO, TROSY-HN(CA)CO, TROSY-HNCA, TROSY-HN(CO)CA, TROSY-HN(COCA)CB, and 3D ^{15}N NOESY-TROSY HSQC were written (as well as all of the necessary data processing scripts), validated using a test protein with known assignments, and applied to LacI. Experiments which normally would have taken ~200 days to collect were performed in less than three weeks using NUS. The high quality of the NUS-derived triple-resonance data resulted in nearly complete assignments of the RD-IPTG complex. Combining the RD-IPTG complex assignments with the assignments of the DBD and DBD-O_{sym} complex obtained using uniform sampling, enabled the assignment of LacI in all functional states bound to ITPG.

While the "divide and conquer" approach and NUS enabled immense progress toward the assignment of LacI, two restrictions could not be overcome. First, ~20% of all expected amide ^{1}H-^{15}N cross peaks were missing for the apo state. Since chemical shifts change drastically upon binding IPTG and IPTG-binding is slow on the chemical shift timescale, a simple titration experiment could not be performed to follow chemical shift perturbations. Though ostensibly a nuisance, the functional implications of these missing peaks will be discussed further in Chapter 4. Second, despite having very favorable spectral properties, the LacI-O_{sym} complex could only be partially assigned (72% coverage). This may arise from its large size or unfavorable relaxation effects arising from the fully protonated DNA to which it is bound (at least for the sites within and near the DBD). Regardless, many functionally important sites within the LacI-O_{sym} complex were still amenable to assignment as will be further explored in Chapter 5. A table summarizing the extent of assignment for all functional states of LacI and its constituent domains is shown below in Table 3.5.

Table 3.5 Summary of assignment coverage for various states of LacI

Construct	Functional State	Resonance Type	Extent of Assignment
LacI DBD	Apo	Amide	100%
LacI DBD	Apo	Methyl (ILV)	96%
LacI DBD	O_{sym}-bound	Amide	98%
LacI DBD	O_{sym}-bound	Methyl (ILV)	100%
LacI RD	Apo	Amide	74%
LacI RD	IPTG-bound	Amide	90%
LacI	IPTG-bound	Amide	88%
LacI	O_{sym}-bound	Amide	72%

Absent from Table 3.5 is a description of methyl assignments for the RD and full length constructs of LacI. This is because initial methyl assignment studies utilizing the COSY-transfer approach performed on a relatively well-behaved 42 kDa protein indicated that CB assignments were absolutely necessary for unambiguous methyl ^{1}H-^{13}C assignments. Given the paucity of CB shifts obtained for most states of LacI, we decided to forgo methyl assignments for the time being. However, it will be shown in Chapters 5 that key assigned methyl cross peaks obtained from the isolated DBD constructs map unambiguously to those of intact LacI and provide meaningful insight into the changes in structure and changes in dynamics that accompany the IPTG-binding.

Further studies are currently planned to increase the coverage of backbone amide assignments and enable the assignment of methyl-bearing side chains. These include exploration of selective methyl labeling [229] to improve resolution and minimize spectral degeneracy (discussed further in Chapter 5) and the development of four dimensional NUS pulse sequences and processing macros (near completion) which reveal through-space interactions via the Nuclear Overhauser Effect (NOESY). A NOESY approach eliminates the reliance on relaxation-limited, through-bond magnetization transfers which have been shown to be inefficient for several functional states of LacI. Since the efficiency of NOESY experiments improves with increasing molecular weight and static magnetic field strength [230], this approach may be applicable to all functional states of LacI.

CHAPTER 4: Revisiting Allostery in Lac Repressor Part 1: Structural and Dynamical Changes in the Absence of Operator

4.1 Introduction

To date, our understanding of allostery in LacI has been inferred from a synthesis of partial crystal structures [32, 75], NMR structures of an artificially cross-linked variant of the DBD [90], and low resolution SAXS data [95]. No one single study has provided a complete description of the allosteric mechanism [95]. In order to address this, we have performed a detailed characterization of the changes in structure and the changes in dynamics that underlie the allosteric transition of LacI using modern solution NMR techniques optimized for high molecular weight proteins. In this Chapter, we focus on just one part of LacI's functional cycle, the transition between the apo state and the inducer-bound state in the absence of operator.

$$\mathrm{LacI} \underset{-IPTG}{\overset{+IPTG}{\rightleftarrows}} \mathrm{LacI{+}IPTG} \tag{4.1}$$

As previously discussed, biochemical studies on purified LacI have demonstrated that IPTG-binding greatly disrupts operator binding. Specifically, apo LacI binds operator with an affinity 3-5 orders of magnitude higher than that of LacI pre-bound to IPTG. Similar behavior is also observed *in vivo*. Recent in-cell, single-molecule fluorescence studies have shown that LacI spends ~13% of its time dissociated from DNA—diffusing in the cytoplasm between DNA-binding events [231]. IPTG-binding has been shown to "sequester" cytoplasmic LacI from re-associating with operator which contributes to the bursts in gene expression that typify induction [232]. In principle, a detailed understanding of the changes in structure and changes in dynamics that accompany IPTG-binding could explain these phenomena and thus provide a mechanistic understanding of LacI functionality relevant in the cell.

To date, only incomplete structural information is available for apo and IPTG-bound LacI in the absence of operator. The crystal structures of both states lack electron density for the entire DBD and hinge region [75] and the differences between the structures of the RD are slight as shown in Figure 4.1.The overall RMS deviation of the main-chain atoms was determined to be 1.2 Å [75], with most of the change localized to the N-terminal subdomain. IPTG-binding was also shown to result in a ~10% increase in the total number of hydrogen bonds in the N-terminal subdomain which facilitates the formation of a stabilizing water-mediated hydrogen bonding network which was described in Chapter 1 [75].

But how do these changes affect the DBD? Unfortunately, since we lack structural information regarding the DBD there is not enough information currently available to answer this question. A previous SAXS study [95] provided a low-resolution structural description of the apo state with intact DBD, however, the IPTG-bound complex was not characterized due to sample instability. NMR studies of the isolated DBD suggest the hinge region is unfolded in the absence of operator DNA [82], however, this has yet to be observed definitively in the context of the full length protein and remains an issue of debate [233, 234]. Furthermore, studies of the isolated DBD lend no insight into changes that may occur as a consequence of IPTG binding.

The limited structural data have also been used to gain some insight into the dynamical properties of LacI through the use of targeted molecular dynamics (TMD) [94].

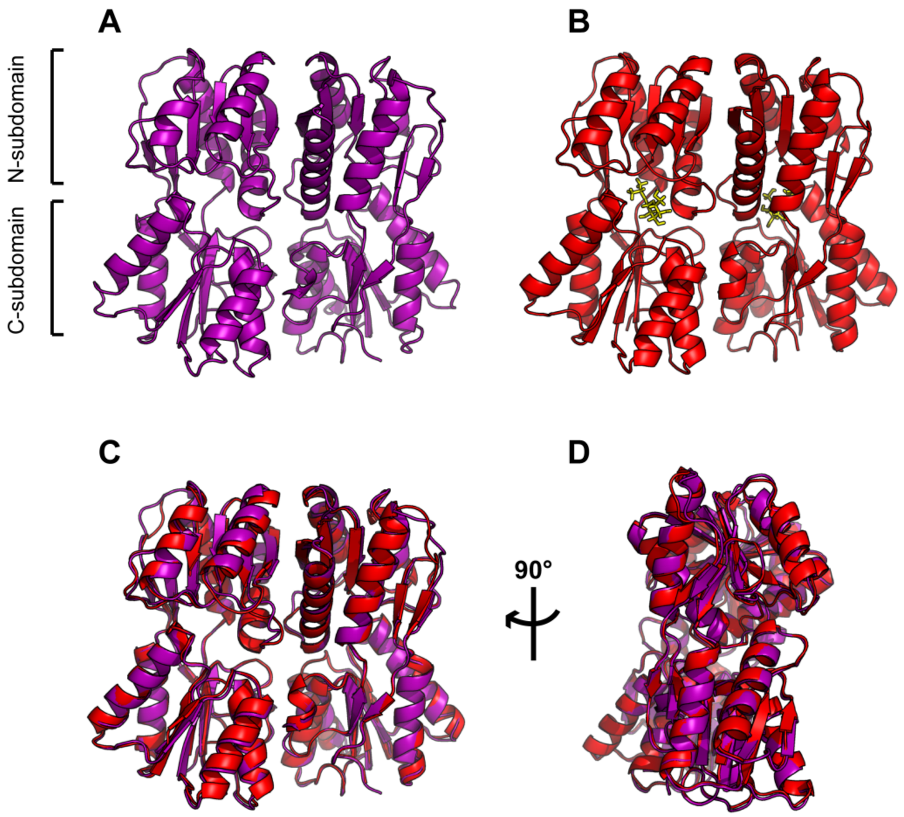


Figure 4.1 Comparison of the crystal structures of apo and IPTG-bound LacI (A) apo LacI PDB: 1LBI (B) IPTG-bound LacI PDB: 2P9H (C) Overlay of the A and B. (D) Rotation of C by 90˚. Note that the IPTG-bound state exhibits a slightly more “closed” conformation in the N-terminal subdomain.

In this simulation, the starting structure was the RD of a dimeric operator-bound complex (PDB: 1EFA) and the target structure was a single dimer from the tetrameric IPTG-bound RD structure (PDB: 1LBH). TMD applies constraints to allow the transition between two states to be sampled in reasonable simulation times. This study suggested that three so-called “allosteric pathways” (described in detail in Chapter 1) exist in the RD which couple inducer-binding to DNA-binding. These pathways were shown to involve structural “alterations” that “move” from the inducer-binding pocket to the monomer-monomer interface then toward the N-terminal subdomain of the RD. Presumably, these alterations also change the DBD but this was not demonstrated by the simulations since the isolated RD structures were used.

Theoretically, solution NMR could illuminate the details inaccessible by other biophysical methods. However, the historical size limit associated with NMR has precluded high-resolution characterization of intact LacI. In Chapters 2 and 3, we presented methods which enabled NMR-based characterization of the intact functional unit of wild type LacI. Here we provide a detailed structural characterization of the DBD in the apo and IPTG-bound states of LacI. We also show compelling evidence that slow timescale (μs-ms) dynamics may play a significant role in mediating allostery.

4.2 Materials and Methods

4.2.1 Site-directed mutagenesis

Two single point mutants of the LacI RD, K84M and K84L, were generated using the QuikChange (Agilent) mutagenesis kit according to standard operating procedures and using the primers listed in Appendix 2.

4.2.2 NMR spectroscopy

NMR samples were prepared as described in Chapter 2 and the buffers used to collect NMR data are listed in Table 2.1. Mutants were prepared and characterized by NMR identically to wild type. NMR experiments were collected as described in Chapter 3. The same suite of resonance assignment experiments conducted on the LacI RD-IPTG complex were also performed on a fractionally deuterated sample of LacI RD K84M-IPTG with the exception of the 3D ^{15}N NOES-TROSY HSQC. Data were collected at 600 MHz and 500 MHz. This sample was grown as described in Chapter 2 but using ~85% D_2O v/v as the growth medium solvent and protonated glucose as the sole carbon source. Even under fractionally deuterated conditions ~20% of the expected amide resonances were either unobserved or too weak to characterize by triple resonance. In total, 75% of the expected amide cross peaks were assigned unambiguously.

Normalized ^{1}H-^{15}N chemical shift perturbations (CSPs) were calculated according to the equation:

$$CSP = \sqrt{(\omega_H \Delta_H)^2 + (\omega_N \Delta_N)^2} \tag{4.2}$$

Where Δ terms are the difference in chemical shift observed between the two states being analyzed and the ω terms are weights which account for the differences in the gyromagnetic ratios of the two nuclei. While there is some variability in the literature as to the choice of these weights, we opted to adapt the approach from the Kay lab where $\omega_H = 1.0$ and ω_N
$= 0.154$ [235].

4.2.3 Determination of the free energy of unfolding for LacI and LacI RD

The free energy of unfolding was determined by monitoring native tryptophan fluorescence-emission at 340 nm of LacI or LacI RD as a function of urea concentration. Dimeric constructs of LacI possess two native tryptophan residues—both localized to the RD. The excitation wavelength for all fluorescence experiments was 280 nm and the protein concentration was ~5-10 μM in NMR buffer. The pH was checked at each denaturant concentration prior to data collection [236]. The temperature of the sample cell was maintained at 25˚C using a water bath and samples were left to equilibrate for ~15 minutes prior to each measurement. The free energy of unfolding was determined from the fluorescence data using both the classical linear extrapolation method [237] and a direct non-linear least squares fit [238] to the raw data with an in-house Python script written by me. In general, both methods yielded similar results (differences typically < $k_\beta T$). Data were analyzed using a two-state model and assuming that unfolding is reversible which has been demonstrated previously for both tetrameric and dimeric constructs of LacI [239]. Though this simplified model does not take into account LacI's oligomerization state [167], it is an accepted method of interpretation and has been used in previous unfolding studies [239]. For intact apo LacI, the $\Delta G_{unfolding}$ = 7.2 kcal/mol•K and $(urea)_{mid}$ = 2.0 M. For apo LacI RD, the $\Delta G_{unfolding}$ = 5.9 kcal/mol•K and $(urea)_{mid}$ = 1.8 M.

4.2.4 *In-vitro* refolding of LacI and LacI RD

A variety of methods to refold purified LacI and LacI RD were assessed. Though refolding studies of LacI have been performed extensively over the course of the last two decades, considerable complications arose when classical approaches [239, 240] were scaled up to NMR concentrations of protein. Since working under dilute protein conditions would have resulted in greatly protracted sample preparation times, an extensive survey of common refolding approaches used in structural studies where protein quantities are large was performed. Two methods were identified which were found to result in consistently high yields of refolded protein.

The first method involved refolding of LacI while immobilized on a chromatography column essentially as described in an application note from GE Healthcare [241]. His-tagged LacI or LacI RD was immobilized on a 3 mL Ni-IDA column then washed with 5 column volumes of equilibration buffer (50 mM sodium phosphate, 500 mM sodium chloride, 3 mM BME, pH = 8.0) at 4˚C. Immobilized LacI was then further washed with 50 column volumes of denaturing buffer (50 mM sodium phosphate, 500 mM sodium chloride, 3 mM BME, 2M urea pH = 8.0). Denaturant concentration was decreased slowly by application of a linear gradient of equilibration buffer over night. After a final wash with 50 column volumes of equilibration buffer, refolded LacI or LacI RD was eluted by washing with elution buffer (50 mM sodium phosphate, 500 mM sodium chloride, 3 mM BME, 250 mM imidazole, pH = 8.0). Once eluted, the His-tag was removed and purification was continued as described in Chapter 2.

The second method was analogous to that described previously for refolding malate synthase G for NMR studies [148]. Briefly, purified LacI or LacI RD was diluted in denaturing buffer (200 mM TRIS, 500 mM sodium chloride, 3-6 M guanidine hydrochloride, 5 mM TCEP, pH = 8.5 measured at 25˚C) to a final concentration of < 50 μM. The protein solution was then incubated at room temperature for 1-2 hours before dropwise addition to refolding buffer (200 mM TRIS, 500 mM sodium chloride, 5 mM TCEP, pH = 8.5 measured at 25˚C) at 4˚C. The volume of refolding buffer was chosen such that the final concentration of denaturant after dilution was < 0.01 M. The protein solution was left at 4˚C overnight. The next morning, the solution was concentrated using 10 kDa molecular weight cutoff spin concentrators (Amicon) and run over a size-exclusion column pre-equilibrated with NMR buffer.

4.2.5 Structure generation using CS-ROSETTA

Structures of the LacI DBD were calculated using CS-ROSETTA [242-244] as implemented by the server provided by the National Magnetic Resonance Facility at Madison (NMRFAM) resource. Typical calculations involved generation of 3000-10000 structures. Two approaches were implemented: the first involved calculation without the hinge region (the "Oliver Lange Method") and the second involved calculation with the hinge region but excluding it from the energy calculation (the "Yang Shen Method"). All input files consisted of complete amide ^{1}H and ^{15}N chemical shifts as well as alpha ^{13}C, beta ^{13}C, carbonyl ^{13}C, and alpha ^{1}H chemical shifts. Alpha ^{1}H chemical shifts were assigned via an HCCH-TOCSY unambiguously for 84% of the non-proline residues. Typical calculations took several hours-overnight depending on the number of structures generated. All structures shown in this work are ensembles of the ten "best" structures identified by the CS-ROSETTA algorithm which ranks structures based on their absolute ROSETTA energy and RMSD from the lowest energy structure calculated.

4.3 Results and Discussion

4.3.1 The DBD is structured and the hinge is unfolded in the absence of DNA

In Chapter 3 (refer to Section 3.3.1 and Figure 3.6), it was shown that the backbone amide ^{1}H and ^{15}N chemical shifts of the isolated DBD were nearly identical to those of the DBD of intact LacI. This correspondence justified the use of the "divide and conquer" approach for obtaining resonance assignments but was not explored further. Here we characterize this correspondence quantitatively and show that it can be used to provide the first high-resolution view of the structure of the DBD of intact LacI.

The chemical shift is an exquisitely sensitive probe of local chemical environment and local structure in proteins. While ^{1}H and ^{15}N shifts are often used as the primary readout in biomolecular NMR studies, changes in these shifts that occur as a result of ligand binding or mutation may not necessarily reflect variations in local or even global structure [245]. Therefore, we used triple resonance data to supplement our analysis with chemical shifts of alpha carbons, carbonyl carbons, and beta carbons which are accurate reporters of structure—particularly secondary structure [228].

Quantitative comparison of the backbone ^{13}C chemical shifts observed in the isolated DBD to those observed in intact LacI bound to IPTG yielded universally strong correlations for all nuclei. The IPTG-bound state of LacI was chosen over the apo state for this initial analysis because it exhibited higher quality spectra and had a higher percentage of unambiguous resonance assignments. The correlations yielded R^2 values of 0.997 for intra-residue CA shifts, 1.0 for inter-residue CA shifts, 0.998 for intra-residue CB shifts, and 0.997 for inter-residue CO shifts. In total, 47 unique DBD residues could be compared which corresponds to ~80% of all DBD residues. The remaining ~20% of residues could not be analyzed because they were unobservable in ^{1}H-^{15}N spectra of LacI. This is likely due to rapid hydrogen exchange with solvent since the data were collected at pH = 7.4 and the unobservable resonances are primarily localized to the termini, regions at the ends of helices which are presumably subject to fraying, and loops which are highly solvent exposed. These data are also consistent with spectra of the isolated DBD at pH = 7.4 (see Figures 2.14a-b in Chapter 2) which exhibits a similar reduction in the number of observable cross peaks relative to data collected at pH = 4.5 where hydrogen exchange is slowed.

Since both amide and backbone carbon chemical shifts are nearly identical for the isolated DBD and the DBD of intact LacI, it follows that their structures are likely highly similar. Using the chemical shifts of the isolated DBD, an ensemble of structural models for the DBD was generated using CS-ROSETTA. CS-ROSETTA is a knowledge-based algorithm which uses a library of NMR chemical shifts from known elements of structure and the robust suite of ROSETTA protein structure prediction methods to generate structural models [242-244]. For proteins in the size regime of the isolated DBD, CS-ROSETTA has been shown to yield highly accurate structural models using chemical shifts as the only input [243]. Figure 4.2a shows an ensemble of the 10 best structures (ranked with respect to ROSETTA energy and alpha carbon RMSD from the lowest energy structure) identified by the CS-ROSETTA algorithm generated using all available chemical shifts. Figure 4.2b shows another ensemble that was generated with the same data but excluding chemical shifts from the hinge region, residues 46-62. The inclusion of the hinge in CS-ROSETTA calculations was found to have no significant effects on the topology of the DBD (Figure 4.2c), although excluding the hinge region from the calculation did result in improved energy minimization profiles (Figure 4.2d-e). The model is consistent with a well folded DBD consisting of three helices and a completely unfolded hinge region that spans residues 47-62. These data are in good agreement (1.2 Å RMSD aligning residues 1-56, 1.1 Å RMSD aligning residues 1-46 only) with the earlier NMR structure of the apo wild type DBD [174], however, that construct was shorter by 6 residues and thus did not have a fully intact hinge.

In order to validate that the structural model derived from analysis of the chemical shifts of the isolated DBD was relevant to the intact protein, peak heights were quantified for the ^1H-^{15}N TROSY HSQC of intact LacI bound to IPTG. As discussed previously in Chapter 3, the intensity of an NMR signal is inversely proportionally to its rotational correlation time due to damping of the FID by the transverse relaxation rate. As a result, disordered regions typically exhibit very strong signals. Figure 4.3 shows that the resonances that arise from the hinge region exhibit very intense peak heights, suggesting that the hinge region exhibits intrinsic disorder in the context of the full length protein.

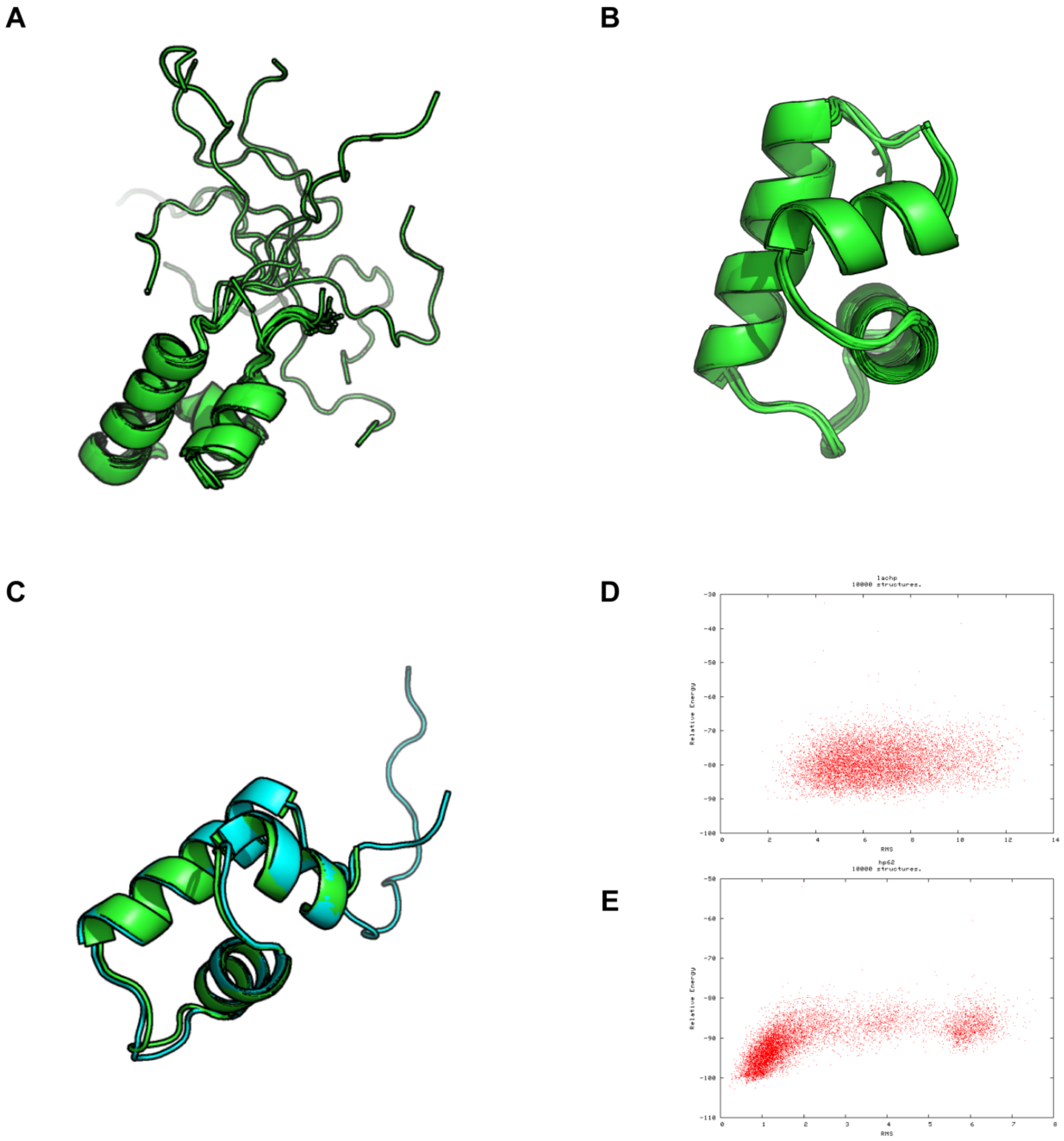


Figure 4.2 CS-ROSETTA modeling of the structure of LacI DBD. (A) Ensemble of the 10 best structures calculated with the hinge region intact. (B) Ensemble of the 10 best structures calculated with the hinge region removed. (C) Overlay of the best structure from A (cyan) and the best structure from B (green). (D) Energy minimization plot for 10000 ROSETTA structures used to generate the ensemble shown in A. (E) Energy minimization plot for 10000 ROSETTA structures used to generate the ensemble shown in B. For D and E the x-axis is RMSD from the best structure (in units of Å) and the y-axis is Relative Energy (in units of ROSETTA energy).

The structured elements of the DBD exhibit peak heights marginally larger than those of other structured regions of LacI. This suggests that the unfolded hinge region imparts additional flexibility on the entire DBD. In other words, the DBD of each monomer is unlikely to dock against the RD or the DBD of the other monomer. This observation is consistent with previous NMR studies of multi-domain proteins where undocked domains tethered by long, unstructured linkers have been shown to tumble semi-independently [246]. These data are also entirely consistent with the structural model derived from CS-ROSETTA. This strongly suggests that in the absence of DNA, the DBD of LacI is mobile but well structured and the hinge region is completely unfolded.

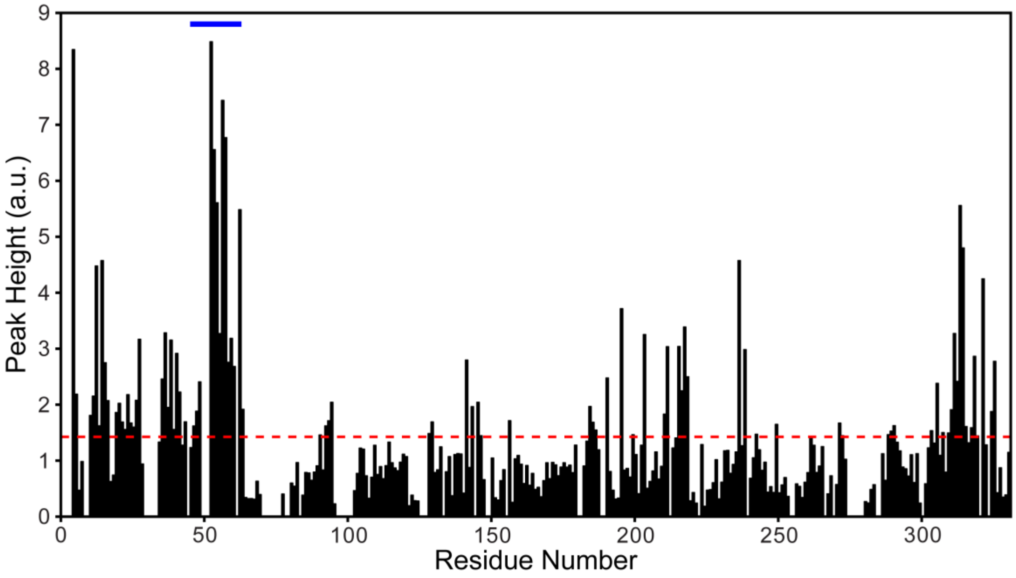


Figure 4.3 Motional properties of the intact LacI-IPTG complex. Peak height calculated at the center point for a given cross peak in the ^{1}H-^{15}N TROSY HSQC is plotted as a function of residue number for all assigned residues. Overlapped peaks are not included in the analysis. The hinge region (residues 46-62) is denoted by the blue bar. The average peak height is shown as a dashed red line.

4.3.2 The structure of the DBD is insensitive to IPTG-binding

The structural modeling and interpretation done in Section 4.3.1 utilized the LacI-IPTG complex rather than the apo state. This was because the LacI-IPTG complex exhibited higher quality spectra and a higher percentage of resonance assignments. In order to gain insight into the structural features of the DBD in the apo state, ^{1}H-^{15}N TROSY HSQC spectra were compared. Interestingly, it was found that the DBD and hinge chemical shifts are identical in both the apo and IPTG-bound states. Figure 4.4 shows a subset of amide resonances from the DBD of the apo and IPTG-bound states of LacI to illustrate this point. Two meaningful conclusions can be drawn from these data: First, the DBD of the apo state is well structured and the hinge is unfolded in the absence of DNA. Second, the DBD does not undergo a conformational change upon DNA-binding.

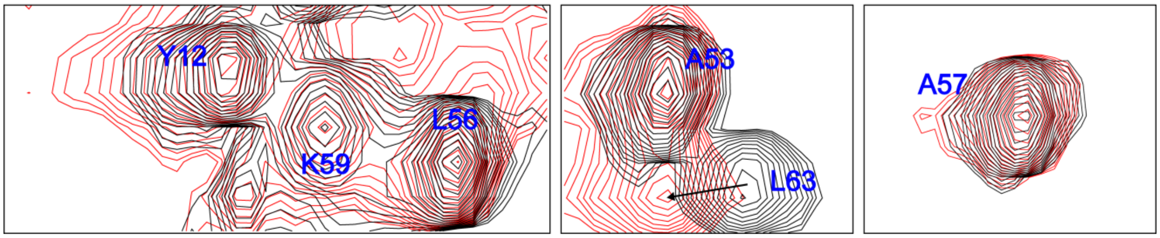


Figure 4.4 The structure of the DBD is insensitive to IPTG-binding. Extracted regions of the ^{1}H-^{15}N TROSY HSQC spectra of apo LacI (black) and IPTG-bound LacI (red). Cross peaks are shown for a subset of DBD and hinge residues. Note that L63, which is the first residue of the RD and part of the N-terminal subdomain, undergoes a change in chemical shift.

These results are ostensibly problematic. As described previously, studies by several groups had demonstrated that apo LacI exists in an equilibrium between induced and repressed states [44, 97]. Since it is already known that the hinge helix is folded in the repressed state, previous NMR studies have suggested that the hinge of the apo state may sample both folded and unfolded states [234]. Our NMR data clearly indicate that the hinge and DBD of the apo state populates an average structure identical to that of the induced state which we have shown has an unfolded hinge. Moreover, a lack of conformational change in the DBD does not adequately explain the functional effect of IPTG-binding. LacI's affinity for operator drops by 3-5 orders of magnitude upon binding IPTG and this is true whether LacI is pre-bound to operator or not [41, 247]. The NMR data effectively rule out the possibility that IPTG-binding results in a change in

DBD structure that compromises operator binding. In order to address these issues and better understand the effect of IPTG-binding, the structure and dynamics of the RD were examined using solution NMR.

4.3.3 The apo RD exhibits slow timescale dynamics

It was previously demonstrated in Chapter 3 that spectra of the isolated RD fully recapitulate spectra of the RD of intact LacI for both apo and IPTG-bound states. The isolated RD produces higher quality NMR spectra than intact LacI which promotes a more detailed analysis. As such, the vast majority of data presented in this Section was collected on the isolated RD.

Unlike the DBD, the RD of the apo state and the RD of the IPTG-bound state exhibited markedly different ^{1}H-^{15}N TROSY HSQC spectra. It was anticipated that IPTG-binding would result in chemical shift perturbations, particularly for residues in the IPTG-binding pocket. However, one unexpected observation was that IPTG-binding resulted in a drastic increase in the total *number* of observable cross peaks. An overlay of the two spectra is shown in Figure 4.5. One simple explanation is that many of the cross peaks in the spectrum of the apo state are simply overlapped and thus cannot be differentiated using a two dimensional experiment. However, this explanation was ruled out upon analysis of the higher-resolution three dimensional TROSY-HNCO and TROSY-HNCA experiments. Based on the triple resonance data, it was confirmed that the ^{1}H-^{15}N TROSY HSQC of apo LacI exhibits only ~80% (211/262) of the total number of expected amide cross peaks based on amino acid sequence. The corresponding spectrum of the IPTG-bound state exhibited ~92% of all expected amide cross peaks (240/262).

There are many potential explanations for why an amide cross peak might be rendered unobservable. One, which has already been discussed above in the context of the DBD, is fast HX with the solvent. Given that spectra of the IPTG-bound and apo states were collected under identical conditions (pH = 7.4 and 25˚C) and that the crystal structures of the two states are highly similar [75], this explanation is unlikely.

Another possible explanation is that not all amide deuterons were back exchanged during sample preparation. Because all LacI NMR samples are perdeuterated, there is some possibility that stably hydrogen bonded amide groups could resist exchange with the protonated solvents used during purification and sample preparation.

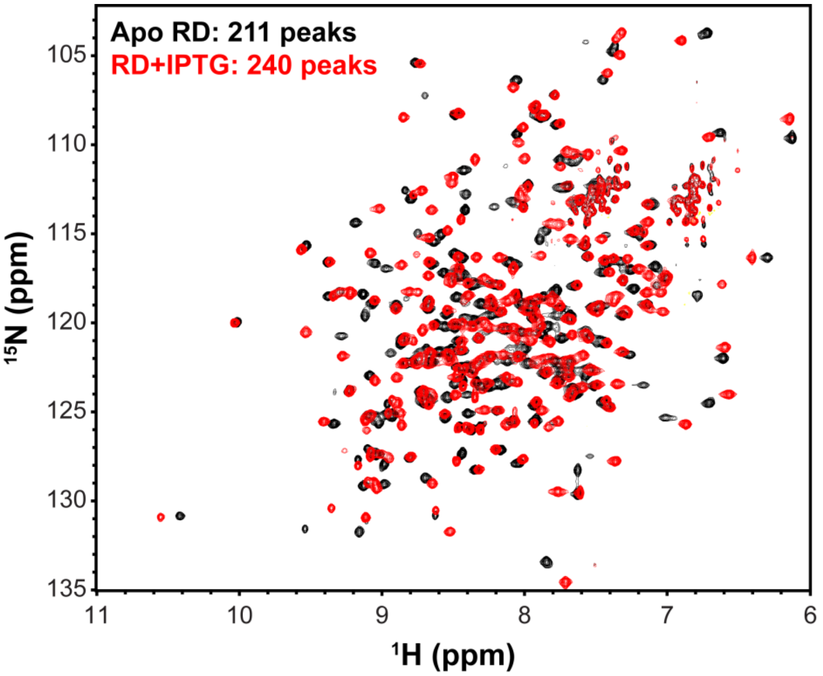


Figure 4.5 Overlay of the ^{1}H-^{15}N TROSY HSQC spectra of apo (black) and IPTG-bound RD (red) The number of observed cross peaks as confirmed via TROSY-HNCO and TROSY-HNCA experiments is listed for each spectrum.

Moreover, since our construct of LacI forms a tight dimer (k_d ~3 nm [240]) and protein concentrations are generally kept high (~100 μM), it is possible amides at the monomer-monomer interface might be resistant to exchange. In order to test this hypothesis, an *in-vitro* refolding protocol was developed for LacI RD. This was necessary because simply adapting refolding protocols from the classical literature on full length LacI failed. A variety of refolding protocols were screened based on previously published approaches and results from classical equilibrium unfolding experiments (see Materials and Methods, data not shown). It was eventually found that high yields of refolded LacI RD could be obtained using either an on-column approach or a rapid-dilution approach. These procedures were also found to be applicable to intact LacI.

^{1}H-^{15}N TROSY HSQC experiments revealed that refolded and non-refolded LacI RD exhibited the exact same number of cross peaks as shown in Figure 4.6a,c. An additional control ^{1}H-^{15}N TROSY HSQC experiment was performed on a fully protonated RD sample (Figure 4.6b). The unfavorable relaxation properties of the fully protonated sample necessitated overnight data collection and resulted in a spectrum with rather broad lines due to the slow rotational correlation time of the 58 kDa dimeric RD. However, the data could still be analyzed. The control data showed unequivocally that the missing amide cross peaks did not arise from a lack of deuterium back exchange.

Another possible explanation is that the missing peaks are broadened due to “chemical exchange”. In a generic sense, chemical exchange can refer to any time-dependent process that exposes a particular NMR probe to at least two different chemical environments [248]. In the context of apo LacI, chemical exchange phenomena would likely result from a conformational exchange process—for example, transitions among structurally distinct functional states (the transition from monomer to dimer has already been ruled out abased on the previously published dimerization constant, the concentration of protein used in our studies, and the consistent presence of a single, narrow peak from size-exclusion chromatography (refer to Chapter 2, Figure 2.6).

But how does conformational exchange render cross peaks unobservable? For simplicity, consider a single NMR exchanging between two distinct conformational states. The probe will exhibit a distinct resonance frequency determined by which state it populates. If the rate of exchange between states is much slower than the difference in resonance frequency, two separate peaks will be observed with intensities that reflect the respective populations of each state. This is known as the “slow exchange” regime. Conversely, if the rate of exchange is much faster than the difference in resonance frequency of the two states, the two peaks are averaged into one peak which will be observed at a population-weighted frequency. This is known as the “fast exchange” regime. However, if the rate of conformational exchange is roughly equivalent to the difference in resonance frequency between the two states, the two peaks will coalesce into a single broadened signal. This is known as the "intermediate exchange" regime. Often, this broadened signal is not detectable, particularly in cases where sensitivity is limited (such as with LacI). This broadening is often described in the literature as "exchange broadening." Conformational exchange typically takes place on the μs-ms timescale and is therefore classified as a slow timescale dynamic process [104].

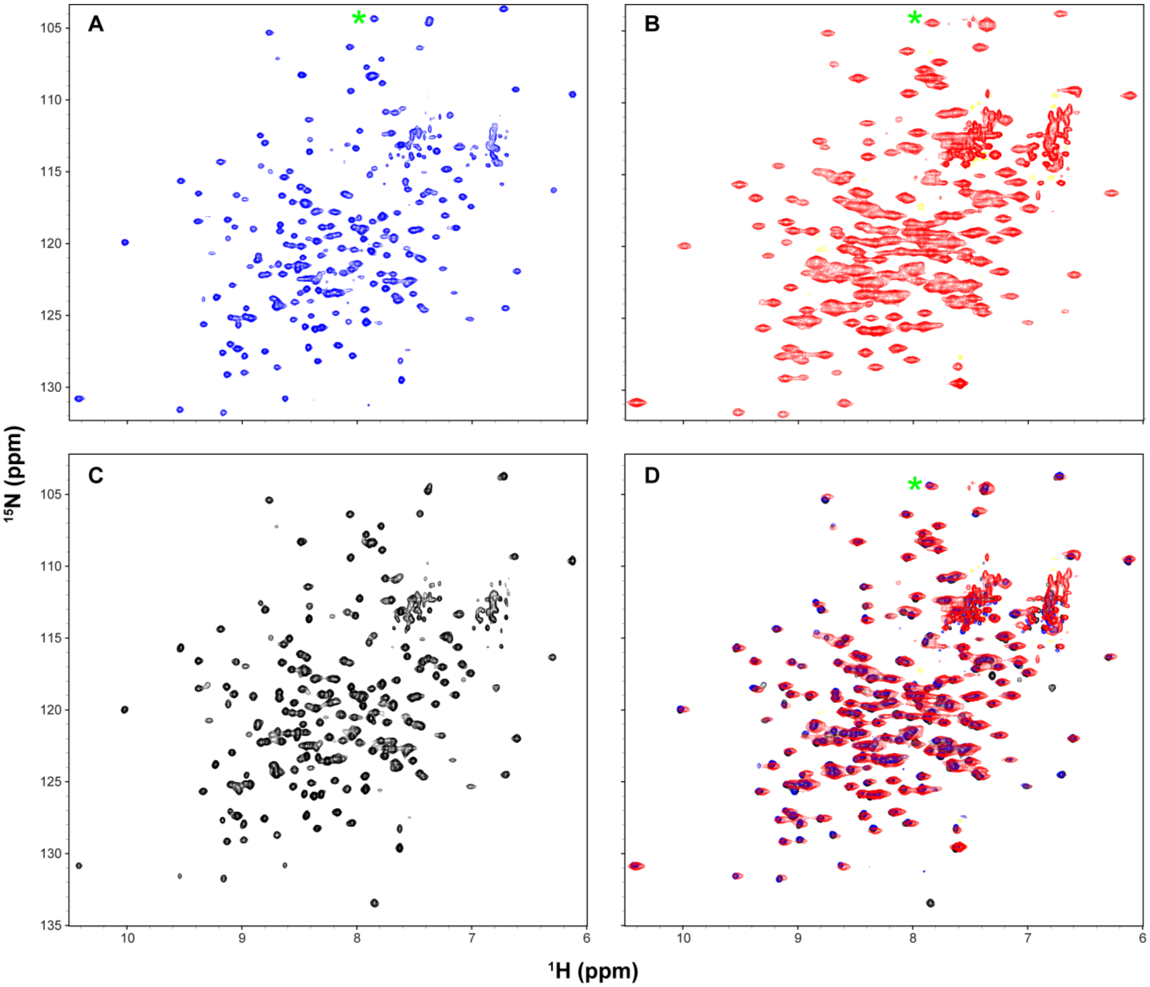


Figure 4.6 Comparison of refolded and fully protonated apo LacI RD. (A) ^{1}H-^{15}N TROSY-HSQC of refolded apo LacI RD. (B) ^{1}H-^{15}N TROSY-HSQC of fully protonated apo LacI RD. (C) ^{1}H-^{15}N TROSY-HSQC of non-refolded apo LacI RD. (D) Overlay of A-C. Negative peaks are shown in yellow. The peak labeled with the green asterisk is aliased in the ^{15}N dimension in A and B.

This scenario is plausible for apo LacI as previous work has suggested that it exists in an equilibrium between repressed and induced states [44, 97]. As discussed previously, our NMR data do not suggest the DBD is in an equilibrium between repressed and induced states based on the equivalence of DBD chemical shifts for both apo and IPTG-bound states. However, analysis of the TROSY HSQC spectrum of apo LacI show that an overwhelming majority of residues in the RD exhibit spectral signatures suggestive of two-state equilibria. Figure 4.7 shows extracts of an overlay of TROSY HSQC spectra of operator-bound, IPTG-bound, and apo LacI. The distinguishing feature of these data is that the resonances of the apo state appear at an intermediate frequency, sandwiched between the resonances of the operator-bound and inducer-bound states. Resonances for all three states fall on an approximate line which indicates that the apo state is interconverting between the induced and repressed states in the fast exchange regime [129, 249, 250]. Some resonances from the apo state, however, show significant line broadening (see G297 and I289 in Figure 4.7) which suggests interconversion on a slower timescale and/or a large chemical shift difference between exchanging states. These data support the hypothesis that the missing cross peaks in spectra of the apo state arise due to exchange broadening, specifically interconversion between repressed and induced states.

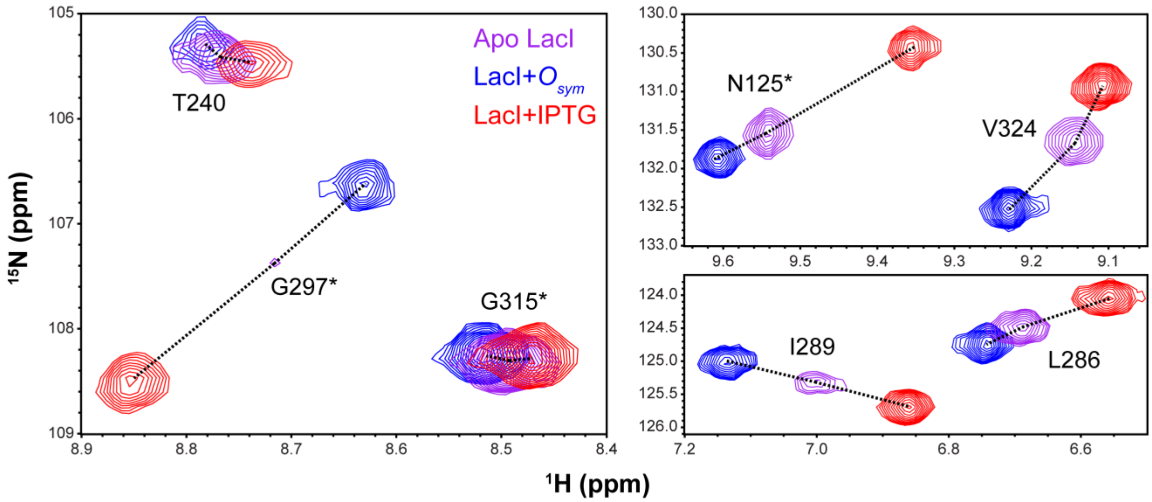


Figure 4.7 The RD of apo LacI is in a two-state equilibrium (A) ^{1}H-^{15}N TROSY-HSQC of refolded apo LacI (purple), LacI-O_{sym} (blue), and LacI-IPTG (red). Resonances for selected RD residues are shown. An asterisk denotes residues localized to the N-terminal subdomain. No asterisk indicates residues localized to the C-terminal subdomain. All data were collected at 25˚C and at the same pH.

An interesting observation about the apparent two-state exchange process exhibited by the apo state is that it is heterogeneous. The variable line broadening observed suggests that multiple timescales of exchange are operative. Additionally, the residues that are in the fast exchange regime may populate the induced and repressed states to different degrees. Studies from the Lewis and Sharp labs suggest that the apo state favors the induced state by a ratio of ~2:1 [97] whereas studies by the Matthews' lab suggest a ratio of ~1:1 [44, 251]. Neither ratio is universally observed in the NMR data, however, it should be noted that changes in chemical shift will reflect any change in local chemical environment. For example, consider the resonance for N125 in Figure 4.7. N125 is in the IPTG-binding pocket, so the chemical shift change observed during the transition from the apo state to the induced state will also reflect changes due modulations in local chemical environment brought about by the IPTG molecule. Therefore, it is very difficult to quantify populations from frequencies alone and the data shown here should not be interpreted as a contradiction of previous work.

Having established that the missing peaks likely arise from conformational exchange, we next sought to identify the particular regions of the RD that exhibit exchange. Figure 4.8a shows that broadened peaks stem from residues that line the monomer-monomer interface with the majority of them localized to the N-terminal subdomain. Interesting, there is also a small cluster of exchange-broadened residues in the C-terminal subdomain beneath the monomer-monomer interface. Historically, the C-terminal subdomain has been identified as a relatively rigid scaffold which remains static while the N-terminal subdomain translocates along the functional cycle of LacI [3, 75]. These data suggest that in the apo state, the C-terminal subdomain may be more flexible than originally thought.

Interestingly, the observed distribution of exchange broadened sites is also qualitatively similar to the areas identified by the TMD study as being salient to allosteric communication. It is important to note that the TMD approach employs specific constraints to enable observation of μs or slower timescale structural transitions on the ns timescale of conventional MD [94]. The simulation and NMR data are therefore qualitatively comparable. Indeed, comparison of the map of residues in Figure 4.8a to a map consisting of the “allosteric pathways” (shown in Figure 4.8b) reveals notable correspondence as well as some differences.

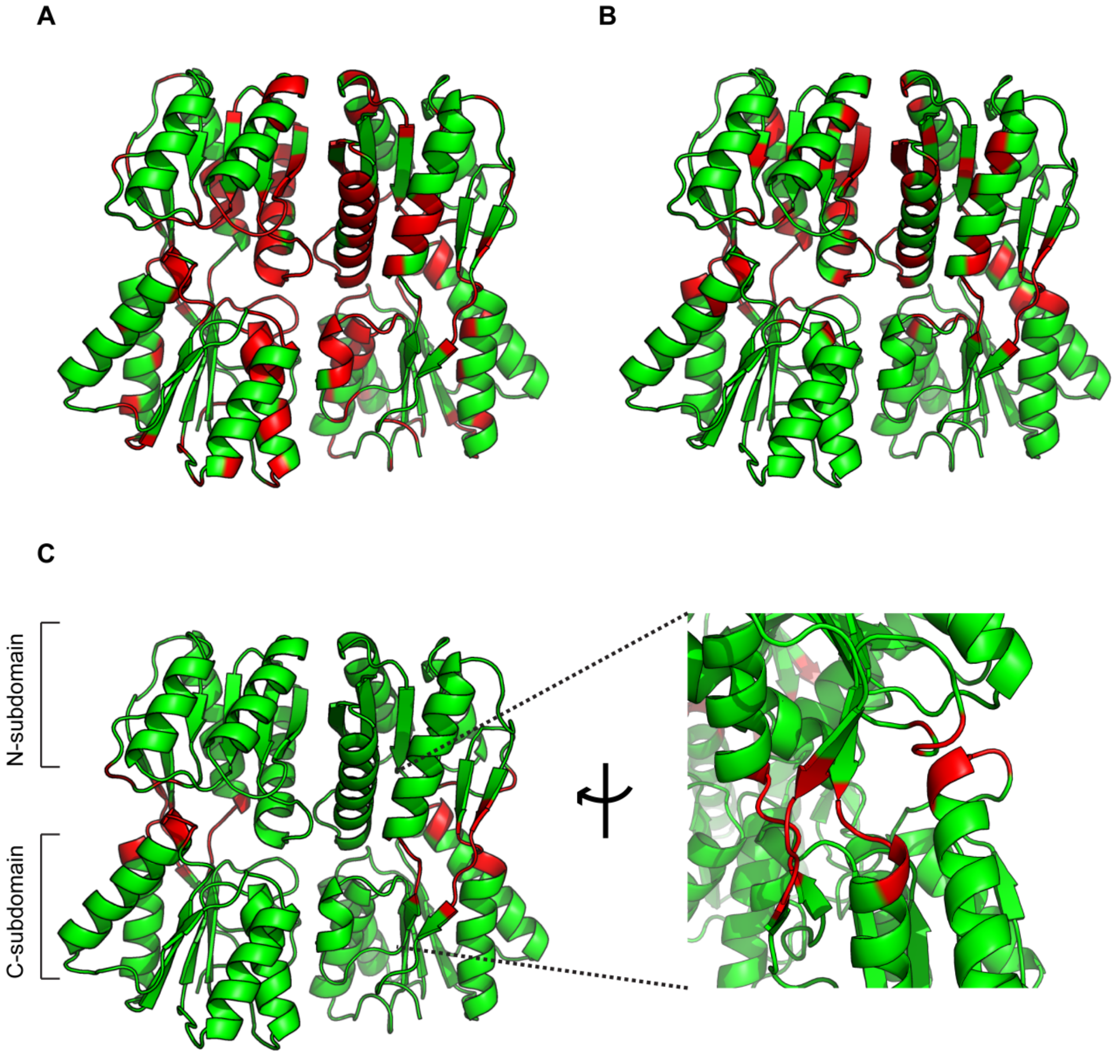


Figure 4.8 Slow timescale dynamics in apo LacI RD. (A) Residues which are broadened by slow timescale dynamics in NMR spectra of apo LacI RD are shown in red (B) Residues involved in the allosteric pathways identified by the TMD study are shown in red. (C) Zoom in on core-pivot region as defined by the Matthews group (shown in red) [251].

In total, ~60% of all residues identified by the TMD study are exchange broadened as detected by NMR. These include all of the key residues at the monomer-monomer interface and most residues in the so-called “core pivot region” where the N- and C-terminal subdomains touch [251]. The most notable discrepancies between experiment and simulation occur at sites involved in hydrophobic packing interactions. The TMD study identified 12 such residues as being dynamic whereas the NMR data show only 3 as being exchange broadened.

It should be noted that the TMD study attempted to delineate a mechanism for how the repressed state transitioned to the induced state. Therefore, we do not expect the motions detected in the simulation to reflect the internal dynamics of the apo state detected by NMR. However, as discussed in Chapter 1, there has been precedent set in the literature that allosteric proteins often sample their various functional states in the absence of allosteric stimulus (consistent with an MWC model of allostery) and this sampling occurs on the slow timescale [105]. Therefore, we deem the comparison valuable.

Our NMR data have shown that the RD of the apo state is likely in a two-state equilibrium between induced and repressed conformations. The timescale of exchange is on the μs-ms timescale but is heterogeneous. Motions are detected in both N- and C-terminal subdomains along the monomer-monomer interface and core pivot region. We next examine the changes in structure and changes in dynamics that result as a consequence of IPTG-binding.

4.3.4 IPTG-binding induces changes localized to the N-terminal subdomain

In order to characterize the changes in local chemical environment that accompany IPTG-binding, normalized ^{1}H-^{15}N chemical shift perturbations (CSPs) [235] were calculated for residues which could be observed and unambiguously assigned in the spectra of both apo and IPTG bound states. In total, 180/262 (~70%) residues from the RD could be compared. Unlike the DBD, which exhibited no changes in chemical shift upon IPTG-binding, the RD shows notable changes. Figure 4.9 shows the CSPs calculated for the RD. The largest changes are localized to the IPTG-binding pocket with R197 and N246 yielding CSPs > 3 standard deviations from the mean CSP. These residues hydrogen bond directly to the IPTG molecule and are essential for inducer binding [75]. A significant CSP is also observed for S77 which is near the IPTG binding pocket but does not contact the molecule. S77 is adjacent to Q78 (not observed in our NMR data) which was previously identified to be involved in a key monomer-monomer interaction that stabilizes the repressed state. The TMD study suggested the cross-monomer interaction involving Q78 is disrupted as a consequence of inducer binding. The CSP of S77 may reflect this event. Large CSPs are also observed throughout the N-terminal subdomain, following an almost linear trajectory away from the IPTG-binding pocket. Overall, the CSP data are in good agreement with previous structural studies and confirm that most perturbations upon inducer binding are indeed localized to the N-terminal subdomain, likely the result of the structural re-arrangements that have been identified to occur there in previous studies [32, 75].

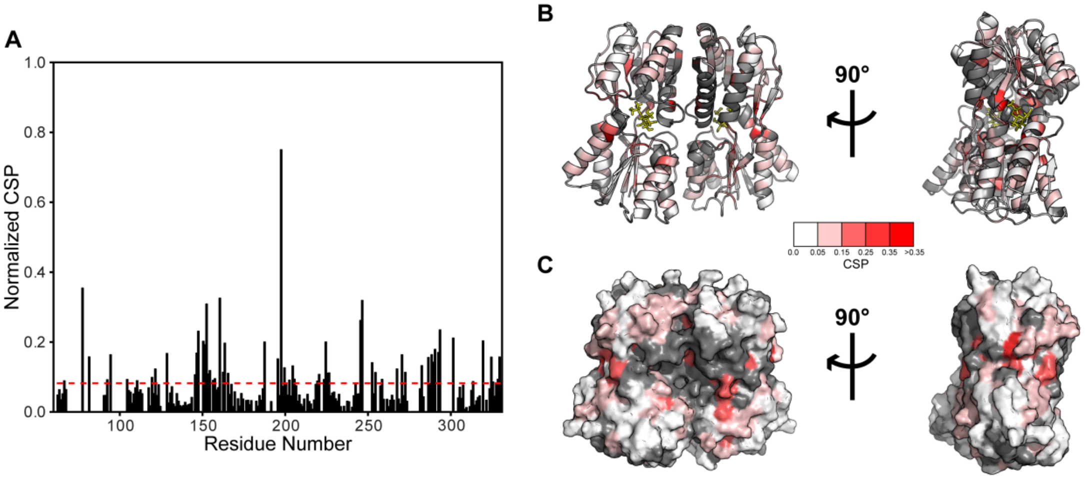


Figure 4.9 Chemical shift perturbations in the RD due to IPTG-binding. (A) Plot of normalized ^{1}H-^{15}N CSPs between apo and IPTG-bound states of LacI RD from ^{1}H-^{15}N TROSY HSQC spectra collected at 750 MHz and at 25˚C. (B) Plot of CSPs on a ribbon representation of the structure of LacI RD bound to IPTG (PDB: 2P9H). (C) Plot of CSPs on a surface representation of LacI RD bound to IPTG (PDB: 2P9H). The scale of CSPs ranges from 0.0 (white) to > 0.35 (dark red). IPTG is shown as yellow sticks. Sites in grey were not analyzable due to exchange broadening in the apo state spectrum.

4.3.5 IPTG-binding only partially quenches slow timescale dynamics

The ^{1}H-^{15}N TROSY spectrum of IPTG-bound LacI RD exhibits nearly 40 additional cross peaks that were unobservable in the corresponding spectrum of the apo state. Since we have shown compelling evidence that apo LacI RD in an in equilibrium of states, we interpret the increase in the number of observable cross peaks as confirmation that IPTG-binding drives the equilibrium toward the induced state.

Chemical shift assignments show the peaks that “appear” in the IPTG-bound state originate from residues at the monomer-monomer interface and the core pivot region. Conformational changes in these two regions have been implicated in the allosteric transition.

Interestingly, a subset of cross peaks was found to be missing from spectra of the IPTG-bound state. Figure 4.10 shows a map of these residues. Similar to the apo state, the missing cross peaks arise from discrete clusters of residues, primarily localized to the monomer-monomer interface. Indeed, nearly all of the sites missing cross peaks in the IPTG-bound state are also missing cross peaks in the apo state. The three clusters of residues consist of S70-Q78, V96-S102, and D275-D278.

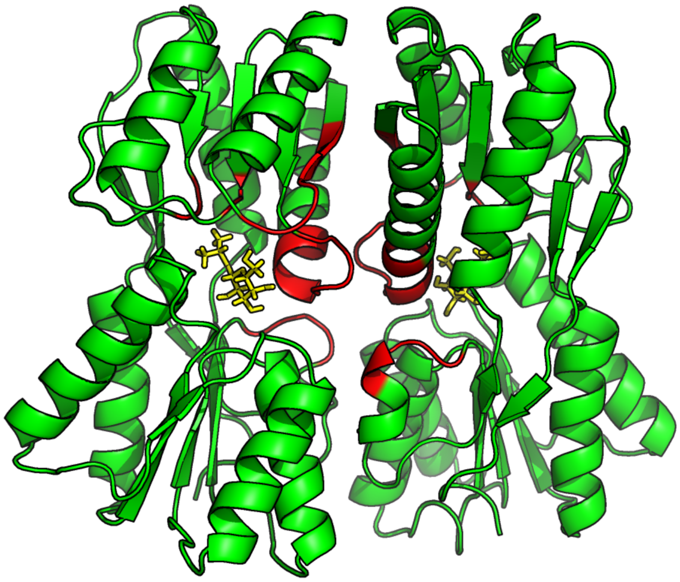


Figure 4.10 Map of residues for which cross peaks could not be identified in the ^{1}H-^{15}N TROSY spectrum of LacI+IPTG RD. Three clusters are localized to the monomer-monomer interface and consist of S70-Q78, V96-S102, and D275-D278. IPTG is shown in stick representation and is colored yellow. PDB code: 2P9H.

Since the clusters consisting of S70-Q78 and D275-D278 are near the IPTG molecule, one possible explanation for the observed line broadening is that the protons of IPTG are accelerating relaxation through a dipolar effect. An alternative explanation is that the line broadening arises from a conformational exchange mechanism like that observed for the apo state. One exciting possibility is that IPTG-binding may fail to quench all slow timescale dynamics in the RD which would further implicate the role of dynamics in the allosteric regulation of LacI. Indeed, all three clusters consist of residues that have been identified by the TMD study as important for allosteric communication. In the next section we show that the line broadening does indeed arise from conformational exchange rather than dipolar relaxation induced by the IPTG molecule and is correlated directly with allosteric functionality.

4.3.6 Mutants with disrupted allostery exhibit totally quenched slow timescale dynamics

A common approach adopted for investigating the relationship between protein dynamics and allostery is the characterization of non-allosteric mutants [117, 120]. This approach posits that mutants which exhibit aberrant functionality will also exhibit aberrant dynamics. We have decided to adopt this approach here by examining mutants of LacI which exhibit disrupted allostery but maintain high-affinity operator binding and inducer binding.

Fortunately, the vast wealth of mutagenesis data on LacI collected by the Miller, Müller-Hill, Matthews, and Swint-Kruse labs made phenotypic screening of non-allosteric mutants unnecessary. Upon thorough examination of the literature, two single point mutants, K84L [171] and K84M [234, 252], were identified as promising subjects. The K84L mutant exhibits slightly diminished affinity for operator, wild-type affinity for IPTG (but with altered kinetics), and severely diminished allostery. The K84M mutant has not been characterized as thoroughly *in vitro* as K84L but has been shown to bind operator with slightly diminished affinity while still exhibiting compromised allostery to about the same degree as K84L. Table 4.1 summarizes data on dimeric constructs of the K84L [247] and K84M [234] mutants as compared to wild type.

Table 4.1 Properties of LacI mutants with hydrophobic substitutions at K84

Mutation	Operator Affinity	IPTG Affinity	Operator Affinity, IPTG-bound
WT	0.63 ± 0.1 nM	1.3 ± 0.1 μM	> 10000 nM
K84L	1.1 ± 0.3 nM	1.4 ± 0.4 μM	4.8 ± 0.2 nM
K84M	2.9 ± 1.0 nM	Same as WT[1]	10.3 ± 3.1 nM

[1] Based on ^{14}C-IPTG binding activity studies which are described relative to WT activity for tetrameric LacI [252]

The K84 site is an attractive option for our study because it is located in the N-terminal subdomain, removed from both the DBD and IPTG-binding pocket. Moreover, K84 had long been recognized as a key mediator of allostery and is often described as a “toggle” [247]. In the repressed state, the charged side chain of K84 is buried in a hydrophobic pocket where it is neutralized by an anion. In the induced state, the sidechain flips out of the hydrophobic pocket and interacts with E100 and D88 (Figure 4.11a). This transition had been identified by the TMD study as being the “juncture” of the various allosteric pathways identified in the RD.

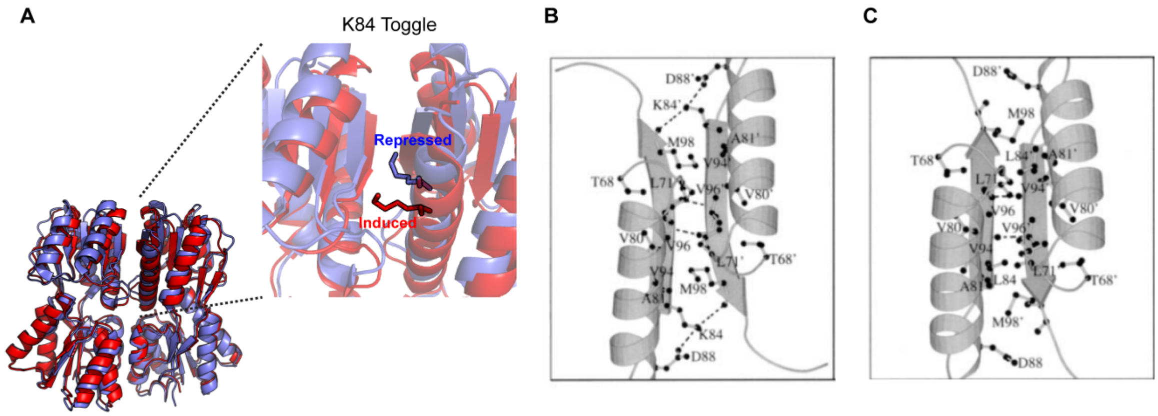


Figure 4.11 Comparison of K84 position in repressed and induced states. (A) Zoom-in of K84 sidechain in the induced state structure (red, PDB: 2P9H), and the repressed state (blue, PDB: 1EFA). K84 is shown in stick representation. (B) Monomer-monomer interface near the K84 site for the wild type protein bound to glycerol. (C) Monomer-monomer interface near the L84 site for the K84L protein bound to glycerol. Figures B and C are reproduced from [253]

Substitution of K84 with hydrophobic amino acids has been shown to dramatically increase the stability of the protein [171, 252]. Indeed, K84L and K84M are resistant to chemical denaturant and exhibit melting temperatures 40˚C higher than that of wildtype. The crystal structure of the free state of K84L mutant suggests that the increased stabilization originates from more efficient packing at the monomer-monomer interface [253] as illustrated in Figure 4.11b-c. Interestingly, the structure of the individual monomers of K84L bears similarity to both apo (with respect to the N-terminal subdomain interface) and repressed structures of LacI (with respect to the N-terminal subdomain orientations) [247]. This is perhaps surprising considering the K84L mutant binds operator with slightly lower affinity than wildtype, despite bearing structural similarity to the repressed state. Unfortunately, while the K84L variant was not crystallized in the presence of IPTG or operator DNA, it was shown to be bound to glycerol which was added in high concentrations as a cryoprotectant. Control operator-binding experiments later showed that the apparent structural bias did not result from glycerol [247].

It is important to note that the hydrophobic K84 variants still exhibit an allosteric response to IPTG-binding, albeit one that is greatly attenuated. A series of biochemical experiments were performed to see if allosteric communication in the K84 variants was a muted version of that exhibited by the wildtype or a completely distinct mechanism [247]. Operator release studies which probe the impact inducer binding has on operator binding clearly indicated that allosteric communication in the hydrophobic K84 variants is damped and not redefined. In other words, the K84 variants undergo the same structural transition from repressed to induced states but exhibit some unique property which damps the functional effect.

Based on the NMR data which indicated that both apo and IPTG-bound LacI exhibit slow dynamics in the RD, we reasoned that the functional damping observed in the K84 variants could arise from a quenching of slow dynamics in the RD.

First, the apo state of the K84 variants was characterized using ^{1}H-^{15}N TROSY HSQC experiments. Samples of perdeuterated K84L RD yielded spectra which exhibited 172/262 expected amide cross peaks which is even less than that observed for wildtype (211/262). However, unlike spectra of the wildtype which exhibited heterogeneous degrees of line broadening, the spectra of K84L samples consistently exhibited uniform linewidths which suggested that the missing peaks did not originate from conformational exchange. A control spectrum was then collected on fully protonated samples of K84L RD since the K84L mutant is resistant to chemical denaturation and therefore cannot be back exchanged via refolding. This spectrum exhibited over 95% of all expected amide cross peaks which confirmed our hypothesis that the slow timescale dynamics (conformational exchange) exhibited by the apo state of the wildtype protein would be quenched in the K84L mutant. The residual 5% of peaks may still be exchanged broadened and may provide a threshold amount of flexibility in order to maintain functionality. However, further studies are required to confirm this hypothesis. To verify that the quenching was not just an anomaly specific to the K84L mutant, the same experiments were performed on the K84M variant. Impressively, the same observation was made. The apo state of K84M exhibited thoroughly quenched slow timescale dynamics. This provides strong evidence that the conformational dynamics observed in the apo state of LacI are indeed quenched in the hydrophobic K84 variants. These data are summarized in Figure 4.12.

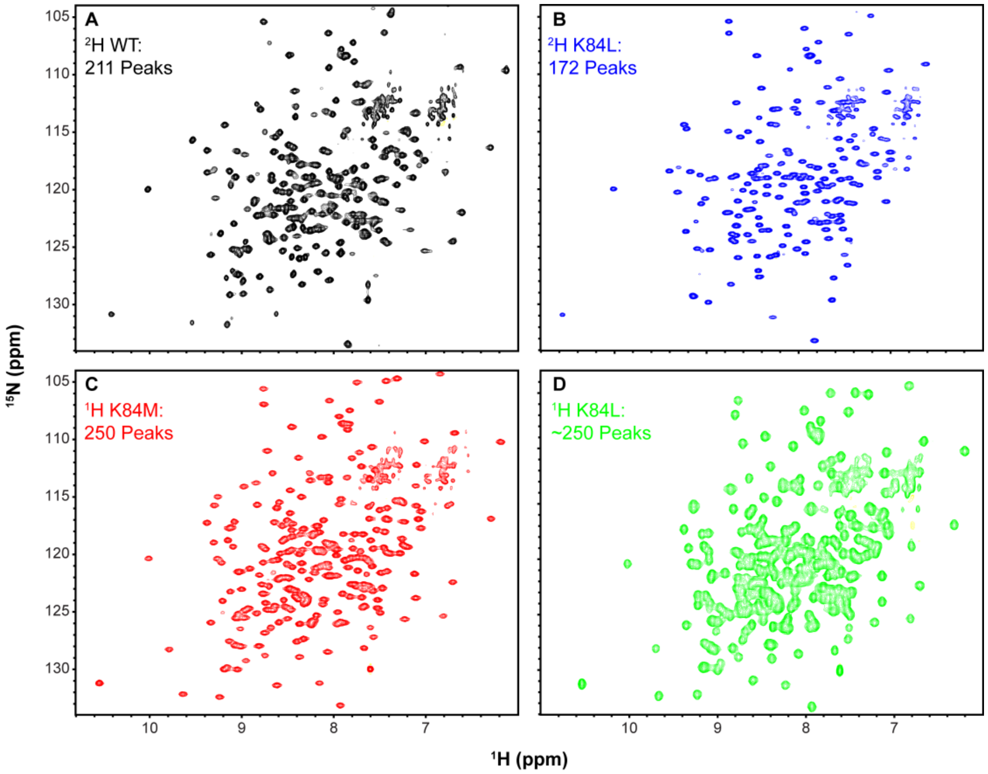


Figure 4.12 Comparison of ^{1}H-^{15}N TROSY HSQC spectra of K84 variants. (A) Apo wild type RD (black), (B) perdeuterated apo K84L RD (blue), (C) fully protonated apo K84M RD (red), (D) full protonated K84L (green). The total number of points sampled in the indirect dimension for D is ½ of that for C. A and B were collected at 750 MHz and 25˚C and C and D were collected at 500 MHz and 40˚C. Control experiments on K84L showed that the number of peaks did not change significantly with increasing temperature (data not shown).

We next investigated the IPTG-bound state. Of interest here is whether the three clusters of missing peaks presented in the previous section arose from conformational exchange or simply from accelerated relaxation due to the close proximity to protonated IPTG molecular. In order to address this we performed a full suite of backbone assignment experiments on a fractionally deuterated K84M sample. Assignments of the K84M mutant were easily transferable to spectra of the K84L mutant by visual inspection due to the high degree of similarity between the two. Comparison of ^{1}H-^{15}N HSQC spectra indicated that both the IPTG-bound K84M and IPTG-bound K84L exhibited cross peaks corresponding to residues that were missing in spectra of the wild type. This confirms that the line broadening observed in the spectra of the IPTG-bound wild-type RD was due to conformational dynamics and not enhanced relaxation effects from IPTG. These data are summarized in Figure 4.13. It should be noted that while differences in the monomer-monomer interface induced by hydrophobic substitutions at position 84 may produce chemical shift changes larger than the small windows shown in Figure 4.13, analysis of the entire spectrum and triple resonance data indicate that they are indeed broadened beyond detection.

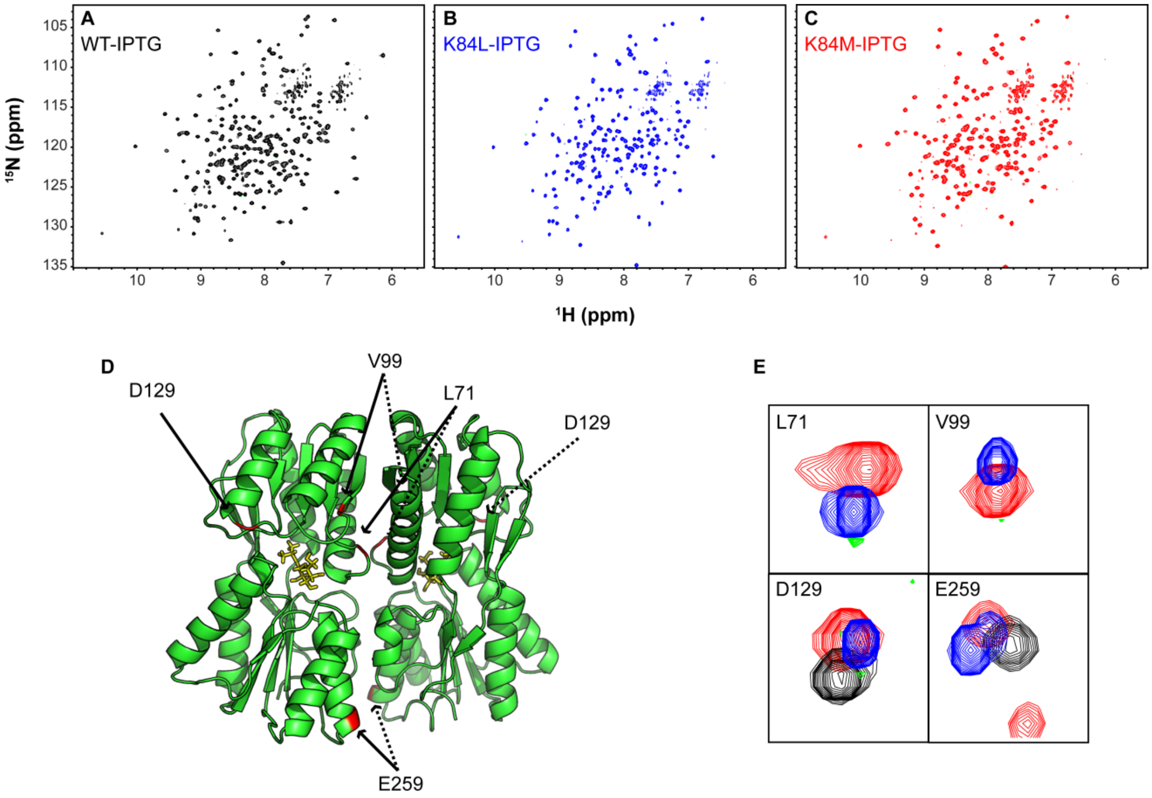


Figure 4.13 K84L and K84M exhibit quenched slow timescale dynamics in the IPTG-bound states. (A) The ^{1}H-^{15}N TROSY HSQC spectrum of WT (black), (B) The ^{1}H-^{15}N TROSY HSQC spectrum of K84L (blue), and The ^{1}H-^{15}N TROSY HSQC spectrum of (C) K84M (red). K84 variants were perdeuterated and non-observable cross peaks originate from a lack of deuterium back exchange as discussed in the main text. (D) Structure of the IPTG-bound wild type LacI RD with several key residues highlighted in red. L71 and V99 are representatives from the clusters of exchange broadened residues exhibited by the wild type protein. Bold lines point to one monomer, dashed lines points to the other. Highlighted cross peaks become observable in spectra of the K84 variants which indicate exchange is quenched. D129 and E259 are sites that did not exhibit exchange broadening in the wild type protein. These sites are also visible in the K84 variants.

4.3.7 Implications for allosteric regulation of LacI

Based on the data presented in the previous section, there is a clear correlation between the presence of residual conformational dynamics in the IPTG-bound state and the magnitude of allosteric response. To reiterate, the wildtype protein exhibits a robust allosteric response to IPTG-binding and also exhibits conformational exchange in three discrete clusters of residues near the monomer-monomer interface: S70-Q78, V96-S102, and D275-D278 in the induced state. Two mutants which exhibit a detectable but greatly diminished allosteric response to IPTG-binding do not exhibit conformational exchange. But why should the presence (or absence) of dynamics at these three clusters have any kind of functional effect on LacI? In order to elucidate this connection, a detailed analysis of the existing structural and biochemical data was performed.

Investigation of the crystal structure of induced LacI reveal that two of these three clusters of residues are structurally connected. Figure 4.14 shows a LIGPLOT/DIMPLOT [254] analysis of the structure. In this representation green dotted lines represent hydrogen bonds and the spoked arcs (eyelash-like figures) represent non-bonded interactions. Red arcs are derived from residues on one monomer while pink arcs are derived from residues on the other monomer. It is striking that a key hydrogen bond interaction is observed between the side chain of H74 of one monomer to side chain of D278 of another monomer. This interaction is abolished in the repressed state [86].

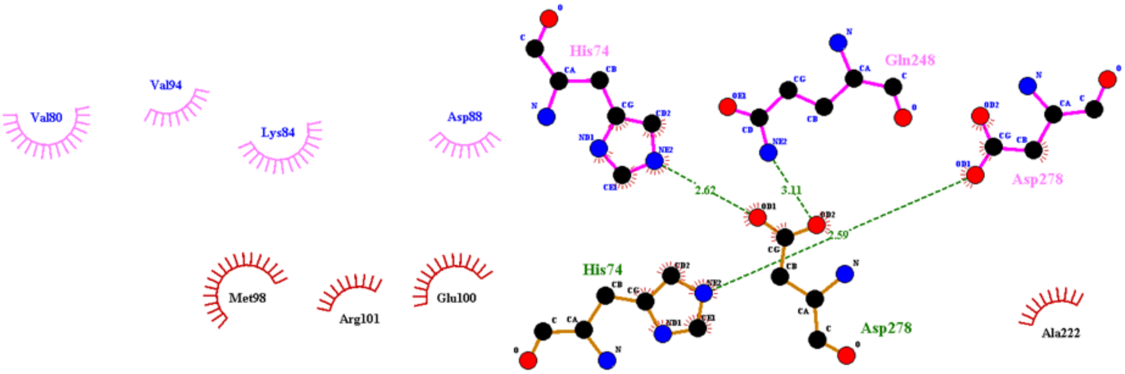


Figure 4.14 LIGPLOT/DIMPLOT diagram illustrating the monomer-monomer interface and cross-monomer interactions in the crystal structure of IPTG-bound LacI RD (PDB: 2P9H). Figures in pink are residues from one monomer and figures in red are residues from the other monomer. Spoked arcs represent non-bonded interactions and green dotted lines represent hydrogen bonds.

The role of the H74-D278 interaction in allostery has been well characterized biochemically [255]. Neither residue is involved in IPTG-binding but mutagenesis of either of the two sites results in disrupted allostery. The nature of the disruption was shown to depend on the nature of the mutation. For example, substitutions of H74 with large, bulky amino acids resulted in a decrease in inducer affinity but increase in operator affinity whereas substitutions with smaller amino acids produced the opposite effect. Mutations at D278 enhance operator binding but diminish IPTG-binding with no apparent size or chemical dependence. Since the mutants produce variable phenotypes, preservation of the actual H74-D278 interaction itself was declared unimportant.

These results can be rationalized in terms of our data. The variable phenotypes may simply reflect how each mutation affects the local motional properties of the protein. A simple interpretation, for example, is that bulky amino acids at position 74 restrict motion more than small amino acids and hence the equilibrium is shifted toward the repressed state for bulky substitutions and toward the induced state for small substitutions. Moreover, networks of spatially separated dynamic residues do not require direct structural connections [256] and hence the actual H74-D278 hydrogen bond would be expected to be of little importance. In order to test this hypothesis, line broadening would have to be measured for the same series of mutants.

The final dynamic cluster V96-S102 features three amino acids (M98, R101, and E100) which interact with the K84 toggle residue in the induced state. It is plausible that flexibility in this cluster may be required to accommodate the flipping motion of the K84 side chain which is abolished when it is replaced with a hydrophobic residue. Since hydrophobic residues at position 84 pack more efficiently with the neighboring apolar residues, motion becomes more restricted.

4.3.8 Model for allosteric regulation

From the NMR data, a simple model for the allosteric mechanism of LacI can be constructed. In the absence of DNA, the apo state of the protein exhibits a well structured DBD, an unfolded hinge, and a RD that exchanges between induced and repressed states. The DBD does not undergo exchange, at least not on the timescale detectable by our measurements. The exchange in the RD is mostly quenched upon binding IPTG as the equilibrium is shifted to the induced state. No structural changes in the hinge or DBD occur upon binding IPTG. We argue that the residual dynamics present in residues S70-Q78, V96-S102, and D275-D278 are a hallmark of the induced state, since mutants that exhibit damped (yet functional) allosteric-responses to IPTG-binding also exhibit quenched dynamics in these regions. Because these mutants are still capable of being induced, they must still be capable of undergoing the transition from the repressed state to the induced state as argued previously [247]. This implicates dynamics may be a key mediator of the magnitude of the allosteric response in LacI.

4.4 Conclusions and Future Directions

In this Chapter, several major conclusions were drawn which either corroborate previous studies or provide new insight into the mechanistic details of LacI functionality.

- In the absence of DNA, the DBD is well structured and the hinge is unfolded

- The structure of the DBD and hinge are both insensitive to IPTG-binding
- The RD is in an equilibrium between repressed and induced states and interconversion takes place on the μs-ms timescale
- Exchange rates of interconversion are heterogeneous across the RD
- Induced LacI exhibits three discrete clusters of residues that are dynamic on the μs-ms timescale: S70-Q78, V96-S102, and D275-D278
- Most residues in these clusters have been identified previously in both biochemical and theoretical studies as being critical to allostery
- Mutants that exhibit greatly diminished allostery also exhibit quenched μs-ms dynamics in the S70-Q78, V96-S102, and D275-D278 clusters

Throughout this Chapter, the number of cross peaks was used as the primary read-out of conformational exchange. While this is a perfectly acceptable readout for the presence of μs-ms timescale dynamics [257, 258], there are NMR methods which quantitatively measure exchange rates for large proteins such as LacI [259, 260]. Of course, these approaches are only applicable to peaks which are observable. Such methods may be useful in the case of IPTG-bound LacI which exhibits only a few missing cross peaks. By having a quantitative measure of exchange rates, one can identify so called synchronous motions [258, 261] which may indicate concerted dynamics across the protein. Such behavior is perhaps anticipated for an MWC model of allostery. At the time this data was collected, the experiments for measuring slow time scale dynamics in large proteins did not exist in our laboratory. We have recently implemented and verified a TROSY-Hahn echo experiment which allows for characterization of dynamics over the μs-ms window and is optimized for large proteins [260]. This will be applied to the various functional states of LacI in the near future.

We have also ignored the role of fast timescale dynamics in this Chapter. As introduced in Chapter 1, both slow and fast time scale dynamics have been implicated in allostery. At the time of data collection, the experiments [262] which quantify fast time scale dynamics were attempted on apo and IPTG-bound LacI. However, the sensitivity of the experiments was severely limited and the data could not be interpreted. In Chapter 6, we introduce a novel approach to collecting these experiments utilizing NUS which permit a 4-fold gain in the number of signal averages. These new experiments may illuminate the role of fast time scale motion and conformational entropy in the allosteric regulation of LacI.

CHAPTER 5: Revisiting Allostery in Lac Repressor Part 2: Structural and Dynamical Changes in the Presence of Operator

5.1 Introduction

In Chapter 4, the changes in structure and changes in dynamics that underlie the transition between apo LacI to IPTG-bound LacI were presented. These studies were done in the absence of operator. While operator-free states of LacI are populated *in vivo*, these states represent a minority of the total ensemble. The majority of LacI is associated with some type of DNA—either diffusing along non-specific sequences in search of operators or specifically bound to operators [37, 231]. LacI-operator complexes are disrupted by inducer binding which lowers LacI's affinity for operator by three to five orders of magnitude. Induced LacI binds operator and non-specific DNA with comparable affinity, so LacI is easily competed off the operator by stoichiometry. This can occur either via stochastic hopping or sliding mechanisms as discussed previously. In this Chapter, we focus on the changes in structure and changes in dynamics that occur upon binding IPTG in the presence of the symmetric *lac* operator, O_{sym}:

$$\mathrm{LacI} + O_{sym} \xrightleftharpoons[-IPTG]{+IPTG} \mathrm{LacI} + O_{sym} + \mathrm{IPTG} \tag{5.1}$$

Throughout this chapter, the LacI+O_{sym} *complex will be referred to as the 'binary' complex and the LacI+*O_{sym}*+IPTG complex will be referred to as the 'ternary' complex.*

While several high-resolution structures of both tetrameric and dimeric LacI bound to O_{sym} exist [32, 75, 86], there is almost no structural information available for the ternary complex. As such, the mechanistic details underlying induction have remained elusive. To date, two dominant models have arisen to explain how IPTG induces such a drastic decrease in operator affinity. In the first model, IPTG-binding causes the N-terminal subdomains of the RD to translocate which, in turn, triggers unfolding of the hinge helix [95, 263]. Since the hinge helix intercalates specifically into the minor groove of the operator, this event would abolish key protein-DNA interactions and result in a lower affinity. An alternative model simply posits that IPTG-binding results in a more subtle destabilization of the repressed state [86], likely through some distortion in the positioning of hinge helices and/or DBD. Unfortunately, it is unclear what the nature of these distortions might be and how only subtle alterations in structure could produce a three-five order of magnitude decrease in operator affinity.

Despite valiant efforts [75], the ternary complex has proven recalcitrant to crystallization. The most detailed structural characterization of the ternary complex to date has come from a low-resolution SAXS study which showed that the binary and ternary complexes had the same radii of gyration [95]. This ostensibly invalidates the model in which the hinge helix unfolds. The reason being that unfolding of the hinge helix would elongate the molecule. While greatly informative, the resolution limitations of SAXS preclude this study from providing mechanistic information.

We have sought to address this dearth of high-resolution structural information by characterizing the binary and ternary complex by solution-state NMR. Unlike the operator-free states of LacI, these complexes exhibit a molecular weight of 85 kDa. This presents an extreme challenge for NMR data collection, even with TROSY methodologies [264]. Here we show that, despite their high molecular weights, both the binary and ternary complexes are amenable to structural characterization by NMR. Compelling evidence is presented that the hinge helix remains folded in the ternary complex, in fantastic agreement with the SAXS data. It is also demonstrated that slow timescale dynamics are likely present in the ternary complex. The potential implications of these observations for the mechanism of induction are also discussed.

5.2 Materials and Methods

5.2.1 NMR spectroscopy

NMR samples were prepared as described in Chapter 2 and the buffers used to collect NMR data are listed in Table 2.1. Data were collected at 750 MHz at either 25˚C, 35˚C or 40˚C.

5.2.2 Isotope labeling of isoleucine, leucine, and valine (ILV) sidechain methyl groups

In order to specifically label isoleucine, leucine, and valine (ILV) sidechain methyl groups, specific metabolic precursor compounds were added to the growth medium following a protocol similar to that previously published for methyl labeling of malate synthase G [148]. Briefly, cells transformed with plasmid harboring the gene for intact dimeric LacI were grown under highly deuterated conditions as described in Chapter 2. When cultures reached an optical density of OD_{600} ~0.3, 70 mg/L of 2-keto-3-d_2-4-^{13}C,-butyrate (Isotec) and 120 mg/L of 2-keto-3-methyl-d_3-3-d_1-4-^{13}C-butyrate (Isotec) were added to the growth medium under sterile conditions. These precursors label isoleucine δ_1 methyl groups as $^{13}C\text{-}H_3$ and valine and leucine methyl groups as ($^{13}C\text{-}H_3$, $^{12}C\text{-}D_3$), respectively. This means only one methyl per isopropyl group is labeled. However, there is equal probability of labeling each stereoisomer so both pro-*R* and pro-*S* methyl groups are observable. Cells were allowed to grow to OD_{600} ~0.4 which typically took about 1 hour following introduction of precursors. At this point, cells were induced as described in Chapter 2. Protein purification and NMR sample preparation procedures were identical to those described in Chapter 2.

5.2.3 Carbon relaxation experiments

Samples for carbon relaxation experiments were specifically labeled at LV sidechains as described in Section 5.2.2 but using 2-(^{13}C)methyl-4-(HD_2)-acetolactate (nmr-bio Technologies). The product is sold as a "single dose" for 1 L of growth medium which upon addition yields a final concentration of (300 mg / L). This precursor specifically labels valine γ_2 and leucine δ_2 methyl groups as ($^{13}C\text{-}HD_2$).This enables isolation of a single ^{13}C-H bond vector which simplifies analysis of relaxation data. ^{13}C T_1 and $T_{1\rho}$ experiments [262] were collected on both binary and ternary complexes at 17.6T and 40˚C. The spectral widths were 10504.202 Hz (F2, ^{1}H) and 2261.920 (F1, ^{13}C), corresponding to 14 ppm and 12 ppm, respectively. The acquisition time was 97.5.0 ms in F2 and 21.2 ms in F1. ^{13}C T_1 delay durations were 0.002 s (twice), 0.160 s, 0.290 s, 0.440 s, 0.620 s (twice), 0.820 s, 1.03 s, 1.26 s (twice), 1.50 s. ^{13}C $T_{1\rho}$ delay durations were 2.00 ms (twice), 6.29 ms, 11.55 ms, 17.78 ms, 24.85 ms (twice), 32.66 ms, 41.16 ms, 50.28 ms (twice), 60.0 ms. For the ^{13}C T_1 experiment, 64 transients/FID were collected and the recycle delay was 1.5 s. For the ^{13}C $T_{1\rho}$ experiment, 64 transients/FID were collected and the recycle delay was 2.2 s. The ^{13}C $T_{1\rho}$ spin lock strength was 3.7 kHz.

Rates were fitted using in-house software written in Python to a two-parameter single exponential decay:

$$I(t) = I_0 \exp(-Rt) \tag{5.2}$$

Where $I(t)$ is the cross peak height at relaxation delay time, t, and I_0 is the initial cross peak height. Best-fit parameters for rates, R, were determined using a Levenberg-Marquardt weighted non-linear least squares algorithm which minimizes χ^2, the sum of the square of the residuals:

$$\chi^2 = \sum_{n=1}^{N} \frac{(I_{calc}(t) - I_{\exp}(t))^2}{\sigma^2} \tag{5.3}$$

Where $I_{calc}(t)$ is the peak height at time t calculated from the fit, $I_{\exp}(t)$ is the experimentally measured peak height at time t, σ is the uncertainty in the peak height and N is the number of planes comprising the relaxation series. The uncertainty was taken to be the standard deviation of the differences in peak height based on replicate measurements scaled by $\sqrt{2}$ [265]. In cases where the data were collected with exceptional S/N, σ was increased by a factor of 2. Errors in the fitted parameters were obtained from the covariance matrix of the fitting routine or from 250-500 Monte Carlo simulations [266] where the sampling bounds were defined by the uncertainty in peak height. Errors obtained from both methods were generally in good agreement.

5.3 Results and Discussion

5.3.1 Operator binding quenches slow timescale dynamics

In Chapter 4, the apo state of LacI was shown to exhibit slow time scale dynamics in areas previously identified as being critical to allosteric communication: the monomer-monomer interface and the core-pivot region. It was shown that these dynamics were consistent with an apo state that exchanges between induced and repressed states on the μs-ms timescale. Comparison of the ^{1}H-^{15}N TROSY HSQC spectra of apo LacI and the binary complex reveal that operator binding also quenches slow timescale dynamics. This is consistent with operator-binding driving the equilibrium toward the repressed state. Of the 319 expected observable amide cross peaks, 310 could be identified based on analysis of triple resonance data. At 97%, this is the highest number of peaks observed for any functional state of LacI. In other words, spectra of the binary complex indicate that LacI does not undergo significant conformational exchange once bound to operator.

5.3.2 The hinge helix of the ternary complex is folded

Since chemical shifts are sensitive reporters of local chemical environment and local structure, they can be used to determine the structural state of the hinge helix in the ternary complex. Unfortunately, the high molecular weight of both LacI-operator complexes precluded comprehensive chemical shift assignment—even with TROSY methodologies and NUS at the limit of sparseness (5% sampling). As described in Chapter 3, only 72% of all expect amide cross peaks could be assigned for the ^{1}H-^{15}N TROSY HSQC of the binary complex. Though this coverage is quite limited, it was sufficient to identify resonances of critical residues in the DBD and hinge helix.

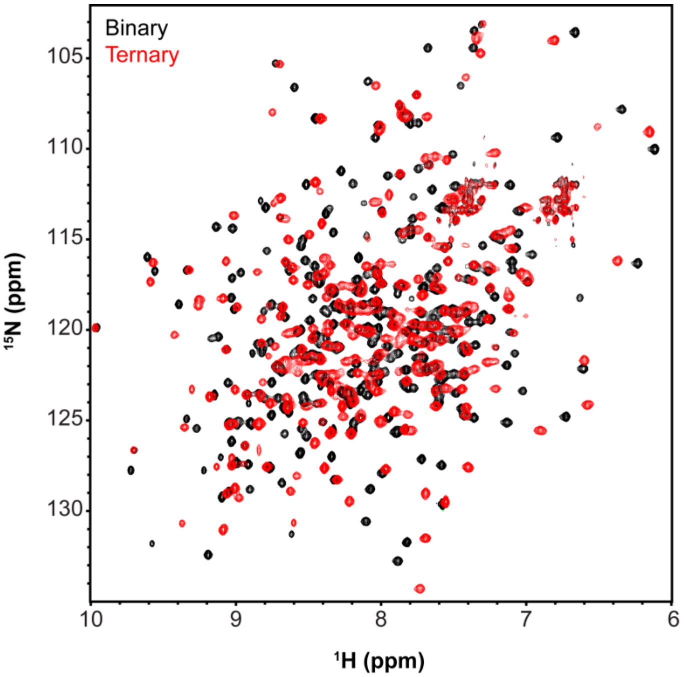


Figure 5.1 Overlay of ^{1}H-^{15}N TROSY HSQC spectra of the LacI-O_{sym} binary complex (black) and LacI-O_{sym}-IPTG, ternary complex (red). All of the data were collected at 750 MHz and 35˚C.

IPTG-binding to the binary complex induces significant perturbations in chemical shift as illustrated in Figure 5.1. Mapping resonance assignments of the binary complex spectrum to the ternary complex spectrum was difficult to do unequivocally. Fortunately, the ^{1}H-^{15}N TROSY HSQC of the ternary complex yielded nearly identical chemical shifts to the corresponding spectrum of the isolated RD bound to IPTG. This indicates that the RD of the ternary complex is structurally identical to that of the isolated RD bound to IPTG. This offers a significant advantage with respect to spectral analysis. Resonances that arise from the RD can be identified immediately and the remaining resonances are likely to arise from the DBD.

In order to facilitate identification of DBD resonances in the spectrum of the ternary complex, additional reference spectra of different constructs of LacI were included in the analysis. Figure 5.2 illustrates how spectra of the RD-IPTG and DBD-O_{sym} complexes could be used to unambiguously identify ternary complex DBD resonances.

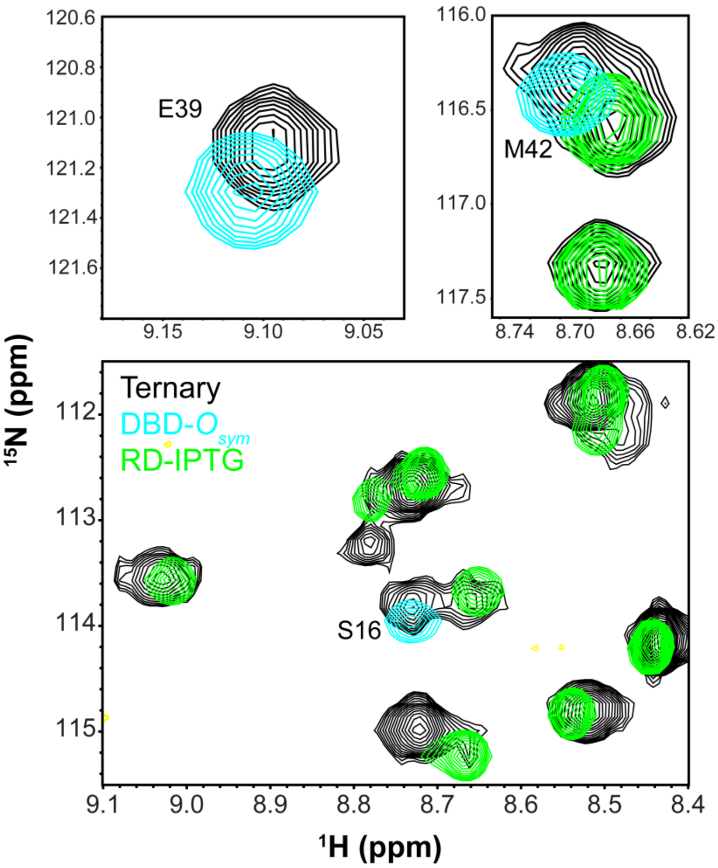


Figure 5.2 Strategy for assigning DBD resonances in the ^{1}H-^{15}N TROSY HSQC spectrum of the ternary complex (black). DBD resonances are identified by comparison of ^{1}H-^{15}N TROSY HSQC spectrum of the RD-IPTG complex (green) and the ^{1}H-^{15}N TROSY HSQC spectrum of the DBD-O_{sym} complex (cyan). Black peaks that overlay with green peaks arise from the RD whereas black peaks that overlay with cyan peaks arise from the DBD. All of the data were collected at 750 MHz and 25˚C. Negative peaks are shown in yellow.

The above strategy enabled identification of DBD resonances in the spectrum of the ternary complex. But it is the hinge region that is of interest here. In order to identify hinge residues and discern secondary structure, chemical shifts of the ternary complex were also compared to those of the LacI-IPTG complex in the absence of operator. In Chapter 4, it was shown that an overwhelming majority of DBD and hinge resonances could be assigned for the ^{1}H-^{15}N TROSY HSQC spectrum of the LacI-IPTG complex. The structural modeling presented in that chapter confirmed that the hinge of the LacI-IPTG complex is unfolded. Therefore, hinge resonances of the LacI-IPTG complex are indicative of an unfolded conformation. If the ternary complex exhibits an unfolded hinge, it should yield resonances consistent with those of the LacI-IPTG complex. If not, it should yield resonances more consistent with the LacI-O_{sym} complex or DBD-O_{sym} complex, where the hinge is known to be folded from previous structural characterization [32, 75, 82, 86].

When comparing spectra of proteins with and without DNA, one potential variable that must be considered is the effect that DNA has on chemical shifts. For example, if a hinge resonance happens to be in close proximity to the operator, it may exhibit a unique chemical shift that is dominated by the chemical properties of the DNA rather than protein secondary structure. This would render the resonance incomparable to those from the spectra of the LacI-IPTG complex. Fortunately, this issue has been addressed by the Kaptein group who characterized the isolated LacI DBD bound to a variant operator in which the hinge helix only formed on one of the two half-sites [87]. It should be clarified that in this complex both half-sites are still bound to protein [87]. The ^{1}H-^{15}N HSQC spectrum of that complex indicated that the unfolded hinge resonances were nearly identical to those from the apo DBD. These data validate our approach. Figure 5.3 shows the amino acid sequence of the hinge helix and Figure 5.4 shows a comparison of ^{1}H-^{15}N HSQC spectra of the ternary complex, LacI-IPTG, RD-IPTG, and DBD-O_{sym}. Note that hinge residues G58 and the strictly conserved A53 of the ternary complex clearly exhibit a resonance frequency more consistent with those from the spectrum of DBD-O_{sym} (folded hinge) rather than those from the spectrum of LacI-IPTG (unfolded hinge). This indicates that the ternary complex exhibits a folded hinge.

Figure 5.3 Amino acid sequence of the hinge region of LacI (residues 46-62). The hinge helix is shown as a blue bar and the corresponding residues involved in helix formation are shown beneath it. Residues in red and marked with asterisks are strictly conserved across the ten nearest relatives of LacI from LacI/GalR family based on sequence alignment using ClustalW [267].

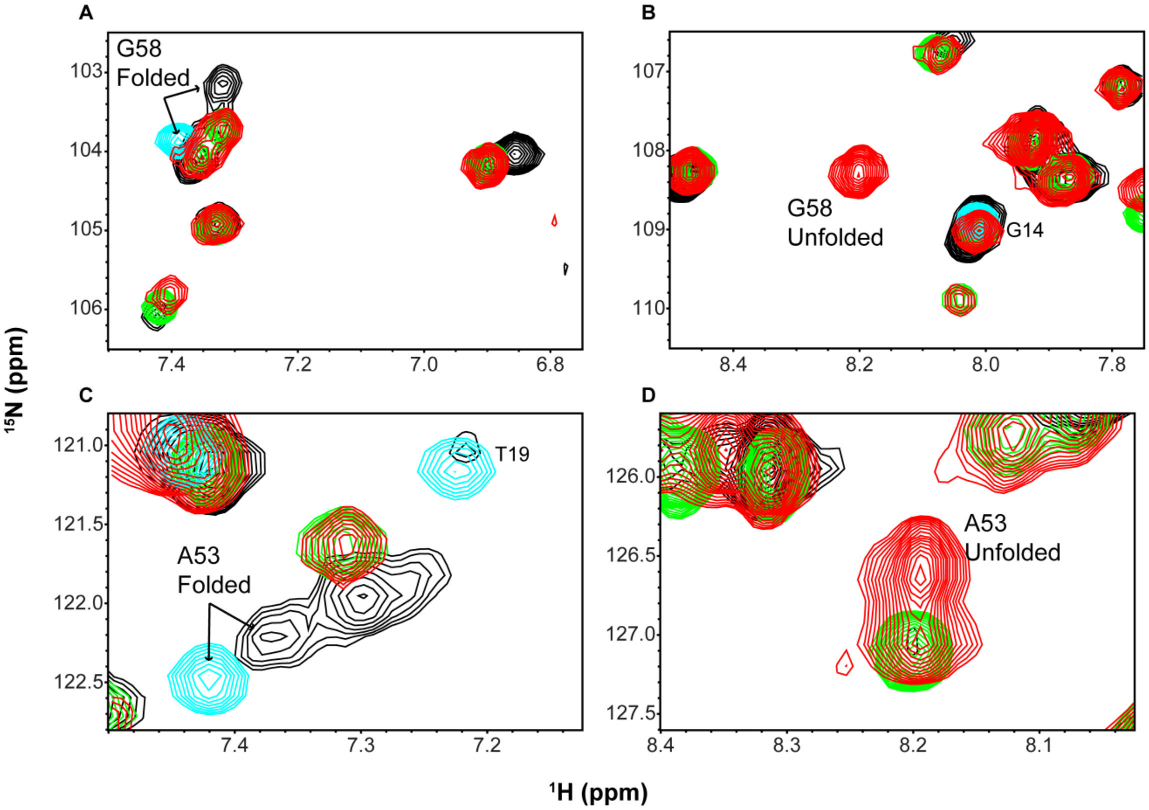


Figure 5.4 Strategy for determining the conformation of the hinge helix of the ternary complex. The ^{1}H-^{15}N TROSY HSQC spectrum of the ternary complex (black) is overlaid with the corresponding spectra for the RD-IPTG complex (green), DBD-O_{sym} complex (cyan), and the LacI-IPTG complex (red). (A) Shift for folded conformation of G58. (B) Shift for unfolded conformation of G58. (C) Shift for folded conformation for A53. (D) Shift for unfolded conformation of A53. All of the data were collected at 750 MHz and 25˚C.

The data above suggest that the hinge helix remains folded in the ternary complex. Like the previous SAXS study, this suggests that the mechanism of induction does not involve local unfolding of the hinge helix.

In order to further characterize the nature of the interaction between the hinge helix and operator in the ternary complex, we focused our attention to L56. As described in Chapter 1, the side chain of L56 intercalates directly into the minor groove of the operator. This interaction is essential for high affinity binding [268] and L56 is strictly conserved across the LacI/GalR family (Figure 5.3). The resonance of L56 could not be assigned unambiguously in the ^{1}H-^{15}N TROSY HSQC of either the binary or ternary complexes. Therefore, NMR experiments which probe the ^{1}H-^{13}C resonances of methyl-bearing side chains were employed to characterize L56. These experiments are advantageous because the side chain of L56 mediates the interaction with operator. The resonances of both methyl groups of L56 were identified from comparison of the stereo-specifically assigned ^{1}H-^{13}C HSQC spectrum of the DBD-O_{sym} complex as shown in Figure 5.5a. In the methyl spectrum of the ternary complex, both resonances are well resolved. Importantly, the resonances exhibit notable line broadening, suggesting that the L56 side chain is indeed in very close proximity to the protonated operator DNA, further corroborating that the assignments are correct.

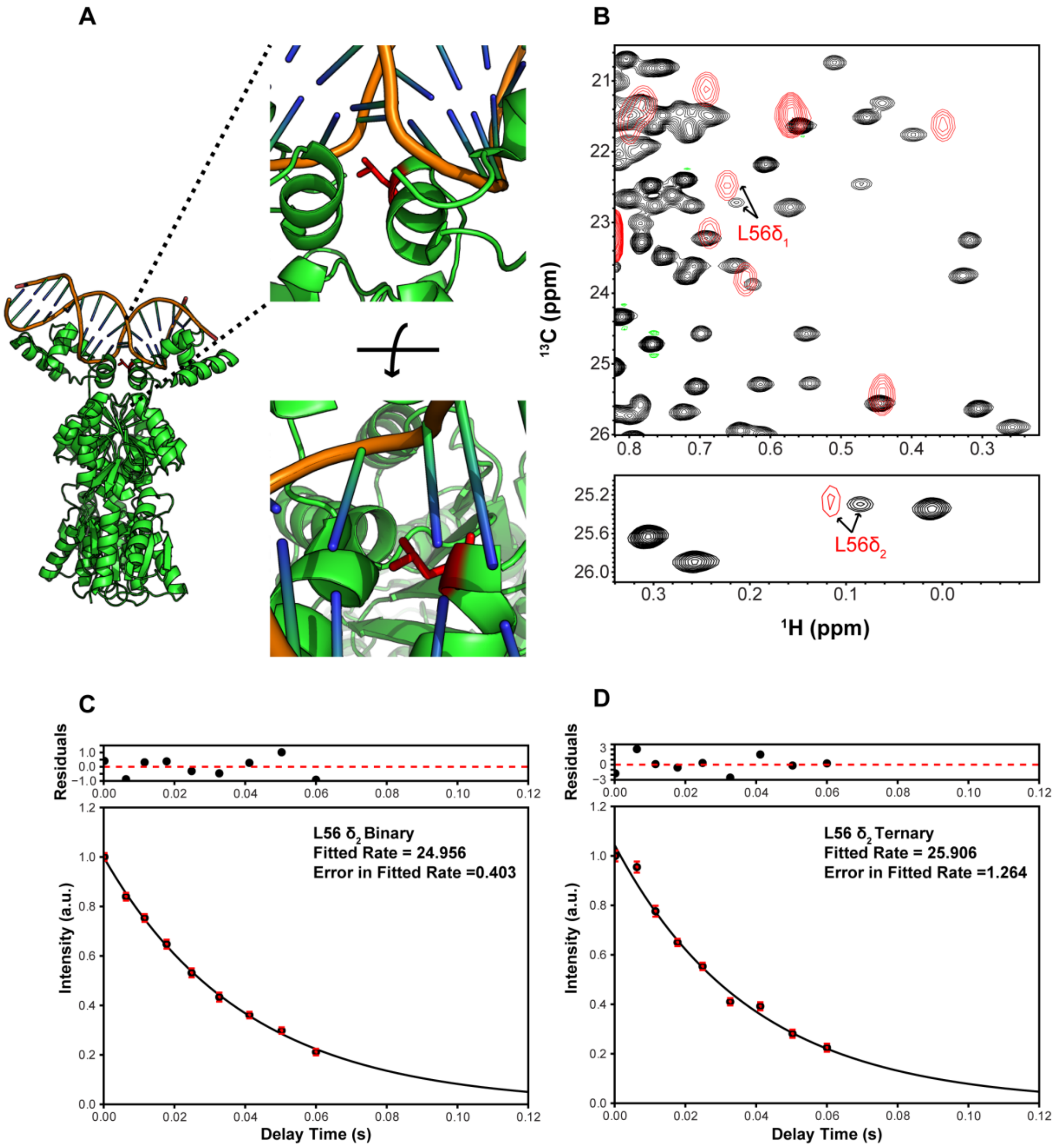


Figure 5.5 L56 is intercalated into the operator in the ternary complex. (A) Crystal structure of Lac-*O*$_{sym}$ complex (PDB: 1EFA). L56 is shown as red sticks. (B) Assignments of the ^{1}H-^{13}C HMQC spectrum of the ternary complex (black) were made via comparison with the assigned ^{1}H-^{13}C HSQC of the DBD-*O*$_{sym}$ complex (red). L56 assignments are shown in red type. (C-D) ^{13}C $R_{1\rho}$ measurements of the well resolved and unambiguously assigned L56 δ_2 side chain resonance for the binary (C) and ternary complex (D).

In order to probe the dynamics properties of the L56 side chain in the binary and ternary complexes, carbon $T_{1\rho}$ relaxation experiments were performed. Relaxation samples were labeled such that only δ_2 and γ_2 methyl groups of leucine and valine, respectively, were observable.

Experiments were performed in D_2O in order to minimize remote protons. This was found to result in stronger peak intensities for L56 relative to samples in H_2O (data not shown) which permitted more quantitatively accurate characterization.

These relaxation measurements are sensitive to the local mobility of the probe (higher mobility probes yield slower relaxation rates) as well as the density of remote proteins nearby. Since our LacI NMR samples are perdeuterated, the vast majority of remote protons that could influence these measurements arise from non-exchangeable sites of the operator. Therefore, it is reasonable to expect that any significant IPTG-induced distortions in the orientation or mobility of the L56 side chain would be detectable. Figure 5.5b shows that the ^{13}C $R_{1\rho}$ rates of L56 in the binary and ternary complexes are within error (24.946 ± 0.403 s^{-1} for binary and 25.906 ± 1.264 s^{-1} for ternary). This suggests that the L56 side chain is not changed significantly as a result of IPTG-binding and further confirms that the hinge helix is folded.

5.3.3 IPTG-binding results in a reduction of observable cross peaks

The data above unambiguously demonstrate that the hinge helix remains folded in the ternary complex. This rules out the possibility that IPTG-binding triggers local unfolding of the hinge helix and thus argues against one of the two potential models for induction. The remaining model posits that IPTG-binding destabilizes the repressed state by a more subtle mechanism. However, even with high-resolution data from NMR, the lack of total resonance assignments precludes the delineation of what may be the subtle distortions and rearrangements that might occur upon IPTG-binding.

One notable difference between the ^{1}H-^{15}N TROSY HSQC spectra of the binary and ternary complexes is that the binary complex spectrum exhibits more cross peaks than the ternary complex spectrum. Unfortunately, no triple-resonance data was obtained on the ternary complex so the possibility of widespread peak degeneracy is still open. However, since there is excellent correspondence between the ternary complex spectrum and the corresponding spectrum of the RD-IPTG complex, it is reasonable to assume the few degeneracies observed in the RD-IPTG spectrum are also observed in ternary complex. Based on this comparison, a conservative estimate of the number of observed ^{1}H-^{15}N cross peaks for the ternary complex is ~280.

To provide a more quantitative confirmation of slow dynamics, ^{1}H line widths were analyzed [257]. Line widths were fitted using the program SPARKY for all resolved peaks in the ^{1}H-^{15}N TROSY spectra of the binary and ternary complexes. A histogram of the trimmed data sets (smallest and largest 5% removed) is shown in Figure 5.6. The binary complex exhibited a trimmed mean ^{1}H line width of 25.0 ± 2.8 Hz where as the ternary complex exhibited a trimmed mean ^{1}H line width of 28.6 ± 8.0 Hz. The ternary complex exhibits a larger mean ^{1}H line width and a much larger standard deviation relative to the binary complex (p-value = 1.54×10^{-9}). These data strongly suggest that IPTG-binding may activate slow timescale dynamics, analogous to what was shown in the absence of DNA. However, further studies are necessary in order to understand the spatial distribution and nature of these dynamics. An alternative explanation for the observed line broadening may be that in the induced state LacI is merely sliding along the operator. However, given the short length of the operator used in these studies (22 bp) and analysis of the crystal structure (PDB: 1EFA), this is unlikely.

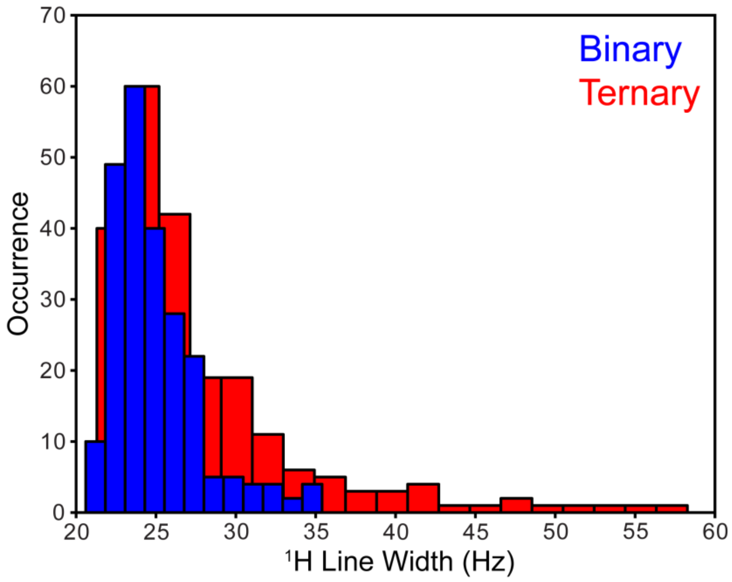


Figure 5.6 Histogram of ^{1}H line widths from spectra of binary and ternary complexes. The smallest 5% and largest 5% values were excluded from analysis. The mean ^{1}H line width for the binary complex is 25.0 ± 2.8 Hz and the mean ^{1}H line width for the ternary complex is 28.6 ± 8.0 Hz.

5.3.4 Implications for allosteric regulation of LacI

These data definitively show that the hinge helix does not unfold upon IPTG-binding. Preliminary analysis suggests the exciting possibility that slow timescale dynamics may also be operative in the ternary complex as was observed in the LacI-IPTG complex in the absence of operator. Though more data is required to confirm this possibility one could envisage that inducer-binding in the presence of operator functions similarly to inducer-binding in the absence of operator.

In the binary complex, LacI does not exhibit significant slow timescale dynamics which suggests that it is not transitioning between functional states. This is consistent with operator-binding driving the LacI equilibrium to the repressed state. Once inducer-binds, little structural perturbation occurs in the DBD and hinge, similar to what was observed in the absence of operator. The RD, however, is shifted toward the induced state which exhibits slow timescale dynamics. It is unclear how dynamic activation of the RD would result in decreased operator affinity. One possible mechanism which has been shown to be operative in Catabolite Activator Protein (CAP) [120] is that the presence of slow timescale dynamics indicate exchange with a functionally distinct, lowly populated state that cannot be detected by ^{1}H-^{15}N HSQC experiments. In the case of LacI, this lowly populated state may exhibit structural features incompatible with high affinity operator binding analogous to what was shown for CAP.

Furthermore, our data do not dismiss a mechanism in which allostery is mediated by modulations in conformational entropy (without structure change) as described in Chapter 1 [118]. While our data indicate that the DBD does not undergo a change in average structure upon IPTG-binding, the changes in conformational entropy are unknown. As previously described in Chapter 4, the necessary experiments required for quantification of conformational entropy could not be performed on LacI in the absence of operator due to low S/N. This was likely due to the requirement for low temperature data collection (25˚C) and high salt in the buffer which degrades the sensitivity of our NMR instruments. However, preliminary methyl carbon relaxation experiments were able to be collected on the binary and ternary complexes. The obtained data show that heterogeneous changes in the fast timescale motion of methyl-bearing side chains accompany IPTG binding (Figure 5.7). However, due to S/N limitations and a lack of assignments, only a small subset (~20%) of all methyl groups were amenable to measurement and analysis. In Chapter 6, we introduce a novel approach to collecting these experiments utilizing NUS which permit a theoretical gain in S/N of at least a factor of 2. These new experiments may illuminate the role of fast time scale motion and conformational entropy in the allosteric regulation of LacI.

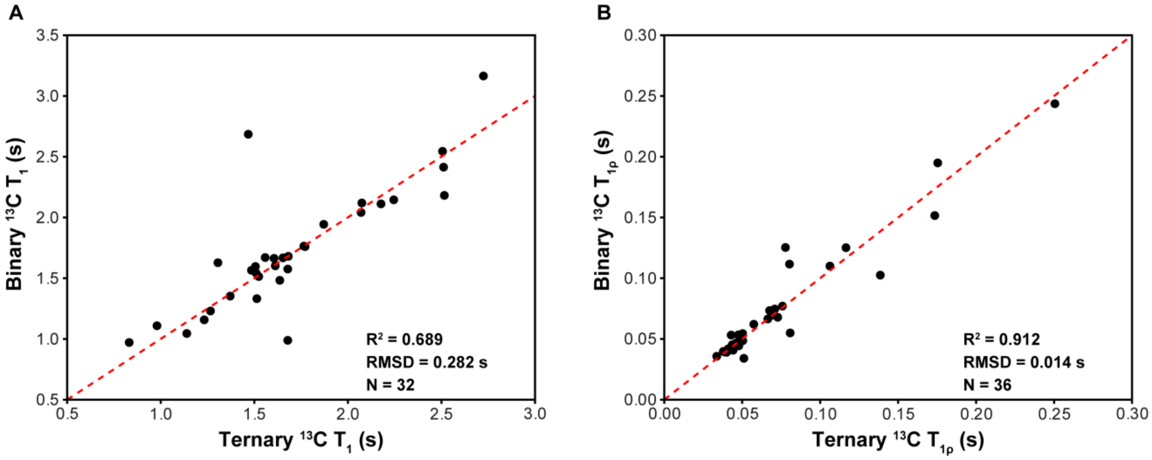


Figure 5.7 Comparison of carbon relaxation times of binary and ternary complexes. (A) ^{13}C T_1 and (B) ^{13}C $T_{1\rho}$. Only ~20% of all methyl groups were analyzable due to the unfavorable properties of the LacI sample and a lack of assignments. Data were collected at 750 MHz and 40˚C. The red dotted line is y=x.

5.4 Conclusions and Future Directions

In this Chapter, we presented compelling evidence that the hinge helix is folded in the LacI-O_{sym}-IPTG ternary complex. These data fully rationalize the data obtained in the previous SAXS study [95]. Moreover, these data support the model that induction likely occurs through a more subtle destabilization of the repressed state rather than a large scale conformational change. It is therefore conceivable that allostery in LacI has a dynamic origin, or at the very least, component.

Like Chapter 4, this Chapter relied heavily on chemical shift assignments and mapping to make inferences about changes in structure and changes in dynamics rather than some sort of quantitative measurement. This is largely derived from the strict limitations imposed on what types of experiments we can reliably perform on LacI due to its high molecular weight and various unfavorable properties. While the operator-bound samples do not require high salt or low temperature data collection, the concomitant increase in molecular weight limited more quantitative experiments to a very small number of observable nuclei. As such, new methodology must be introduced to enable these studies in the future. We address this need in Chapter 6.

Future directions should likely be focused on characterizing the ternary complex for which there is very little structural information. Our data show that this state is stable and yields interpretable NMR spectra. The first priority would have to be expanding resonance assignment of ternary complex spectra by the use of triple resonance and NOESY-type experiments as described in previous chapters. NOESY-type experiments may even be useful in expanding the set of assignments available for the binary complex.

The preliminary dynamics characterization presented here suggest that these measurements will likely be feasible for all available methyl groups (not just the 20% we had access to) via the new methodological development introduced in Chapter 6.

CHAPTER 6: Accurate determination of rates from non-uniformly sampled relaxation data

*The majority of this chapter was published in Stetz MA and Wand AJ, Journal of Biomolecular NMR **65** (2016) 157-70.*

6.1 Introduction

In Chapters 4 and 5, we presented compelling data which suggests that IPTG-binding results in minimal structural rearrangement of the DBD of LacI. Moreover, slow timescale dynamics were shown to be correlated with allosteric communication in the induced state of LacI. These observations open the door to the exciting possibility that the allosteric mechanism of LacI may have a largely dynamic component. Unfortunately, the role of fast timescale dynamics was largely ignored in our studies. This was primarily because the high molecular weight of LacI coupled with its other unfavorable spectroscopic properties rendered the necessary classical spin relaxation experiments too insensitive. In principle, these experiments could also be collected with NUS much like the triple-resonance data presented throughout this work. However, NUS relaxation experiments have historically suffered from a loss of quantitative accuracy for largely unknown reasons [269]. Here we identify the origin of errors in NUS-derived relaxation parameters and introduce a simple procedure for guaranteeing that quantitative accuracy is maintained. We show that reductions in data collection time can be as high as 75% while maintaining accuracy within the inherent reproducibility of the uniformly sampled experiment. This offers the potential for a 4-fold increase in signal averaging with no penalty in total experiment time.

Classical NMR spin relaxation can provide site-resolved, quantitative measures of the amplitudes and timescales of protein motions [256, 270]. These experiments are sensitive to equilibrium fluctuations of interaction (bond) vectors that occur on the ps-ns timescale. The characterization of motions on this “fast” timescale can potentially reveal critical aspects of the thermodynamics and kinetics of molecular recognition and other aspects of protein function such as allostery [109, 271-280]. In structured proteins, relaxation rates are dominated by the molecular tumbling and complete interpretation requires separation of the overall correlation time from internal motions. This is commonly done using the ‘model-free’ formalism [281, 282] and variations thereof [283, 284] wherein internal motion is described with a local correlation time and an amplitude metric called the square of the generalized order parameter. This type of analysis requires collecting multiple relaxation experiments often at more than one static field strength. Traditionally, relaxation phenomena are resolved using two dimensional experiments where cross peak intensity or volume is quantified as a function of an incremented delay period creating a pseudo-three dimensional (pseudo-3D) spectrum. High S/N and high spectral resolution are therefore required to ensure reliable cross peak quantification. As a result, data collection is usually very instrument time intensive and often prohibitively so for unstable or dilute protein preparations. The need to accelerate NMR relaxation experiments has promoted several approaches over the years including reducing the number of relaxation delays [285], reducing the number of NMR observables used for downstream analysis [286], and the recent revival of ‘accordion’ spectroscopy [287, 288].

A potentially more general method for accelerating multidimensional NMR experiments is the use of non-uniform sampling (NUS) strategies in one or more indirectly sampled time domains. Regular sequential sampling (RSS) with uniform increments, which is required for the successful application of the discrete Fourier transform (DFT), has been the dominant approach for decades. However, application of this strategy to multidimensional data acquisition is inherently time intensive. So-called ‘on-grid’ NUS schemes capture only a subset of the uniform-grid and omitted data points are typically reconstructed in processing prior to application of the DFT. NUS methods have existed for quite some time in the context of biomolecular NMR [197] and many different implementations of data sampling and reconstruction exist [198]. To date, these methods have found the widest applicability in contexts where only the resonance frequency is of interest or where peak heights are treated in a semi-quantitative manner. This is largely due to uncertainty regarding the fidelity of reconstructed peak heights and the relationship between sampling density, spectral sparseness, and spectral complexity. The later issues are particularly important for experiments of lower dimensionality such as those typically collected for NMR relaxation. Indeed, among the recent quantitative applications of NUS to date, many have involved the serial collection of higher dimensional 3D (pseudo-4D) experiments [289-291].

To our knowledge, only a few studies have applied NUS to a quantitative pseudo-3D experiment. In the earliest study, ^{15}N R_1 rates were obtained using a variation of maximum entropy (Max-Ent) reconstruction [292-294] where *in situ* calibration was employed to correct for the inherent non-linearity of Max-Ent [295]. Though spectra reconstructed using calibrated Max-Ent were of high quality and largely devoid of artifacts, the resulting relaxation rates were inaccurate relative to reference data. A more recent study demonstrated that the **S**pectroscopy by **I**ntegration of **F**requency and **T**ime domain information (SIFT) method [296, 297] is capable of determining accurate ^{15}N relaxation dispersion parameters using 40% random sampling [298]. However, the generality and limitations of this method were not quantitatively explored. A third study has analyzed the relationship between sampling density, accuracy in peak height, and accuracy in NUS-derived ^{15}N relaxation dispersion parameters for exponentially biased sampling [269]. Various implementations of multi-dimensional decomposition (MDD) [290, 299-301] and iterative re-weighted least squares (IRLS) reconstruction [302] were evaluated and it was shown that incorporation of a single uniformly sampled spectrum in a series of NUS spectra can facilitate the accurate determination of relaxation dispersion parameters from data sampled at impressively low densities (12.5-20%). However, high quality data at such low sampling densities were only obtained for highly resolved spectra of modest complexity and only using recursive MDD reconstruction with co-processing. An additional NUS ^{15}N relaxation dispersion study utilizing MDD reconstruction has also been recently presented, but the fidelity of the NUS-derived relaxation parameters was not discussed [303].

To provide a quantitative evaluation of classical spin relaxation rates derived from pseudo-3D data, a large-scale experimental study comprised of over 60 relaxation experiments (over 30 RSS & NUS pairs) has been undertaken here. The sinusoidally-weighted Poisson-gap sampling method was used for NUS [202] and iterative soft thresholding (IST) [202-204, 208, 304] was used as the primary reconstruction method. Poisson-gap sampling has the advantage of being relatively insensitive to the seed of the random number generator used during schedule generation; though schedule-specific performance has been shown to vary [205]. IST-reconstructed frequency domain data has been shown to exhibit highly linear peak heights with respect to RSS references, even in 2D at lower sampling densities (25-32%) [204]. Finally, these methods are widely used within the NMR community. Here, these approaches are evaluated with respect to their performance in a quantitative context. First, ubiquitin is used as a model system to examine the relationship between the total number of NUS points collected at a fixed sampling density and the accuracy of NUS-derived relaxation rates. Then a comprehensive error analysis is performed to characterize the nature of the inaccuracy of NUS-derived relaxation rates. Finally, the generality of the error analysis is assessed using three alternative reconstruction methods: IRLS as implemented in the MddNMR software package [301, 302], l_1-norm regularization using the NESTA algorithm (NESTA-L1) [305, 306] as implemented in the NESTA-NMR software package [307], and iterative re-weighted l_1-norm regularization using the NESTA algorithm (NESTA-IRL1) [210] as implemented in the NESTA-NMR software package [307], three additional proteins, and eight different relaxation experiments. The protein samples and experiments were chosen to span a wide range of concentrations, feature different numbers of observable cross peaks, and exhibit varying degrees of spectral complexity.

6.2 Materials and Methods

6.2.1 Protein expression and purification

Four different proteins were recombinantly expressed from T7 promoters in the *E. coli* strain BL21(DE3) or derivatives thereof. ^{15}N human ubiquitin was expressed in H_2O-M9 minimal medium and purified as described [308]. Random fractionally deuterated ^{2}H, U-(^{15}N,^{13}C)-ubiquitin was prepared similarly using 55% v/v D_2O-M9 minimal medium. Ubiquitin NMR samples were ~1 mM in protein in 50 mM d_3-acetate, pH 5.0, and 0.02% w/v NaN_3. An additional U-(^{2}H,^{15}N) ubiquitin sample was prepared in the same way using 99.9% v/v D_2O-M9 minimal medium. The

NMR buffer was the same except with 30% glycerol v/v added to simulate the effect of slow tumbling. ^{15}N vertebrate calmodulin (CaM) was prepared as described using H_2O-M9 minimal medium [309]. The NMR sample was ~1 mM in protein in 20 mM imidazole, pH 6.5, 100 mM KCl, 6 mM $CaCl_2$, and 0.02% w/v NaN_3. U-(^{2}H,^{15}N), (ILV-$^{13}CHD_2$) crab arginine kinase (AK) was prepared as described [310] using 99.9% v/v D_2O-M9 minimal medium. Isoleucine, leucine, and valine methyl groups were specifically labeled as (U-^{2}H, Ile-δ_1-$^{13}CHD_2$), (U-^{2}H, Val - $^{13}CHD_2/^{12}CHD_2$), and (U-^{2}H, Leu-$^{13}CHD_2/^{12}CHD_2$) using the precursors and protocol previously described [311]. The NMR sample was ~75 μM in protein in 10 mM sodium citrate, 50 mM sodium chloride, pH 6.0. U-(^{2}H,^{15}N) maltose binding protein (MBP) was prepared as described using 99.9% v/v D_2O-M9 minimal medium [312]. The NMR sample was ~700 μM in protein in 20 mM sodium phosphate, pH 7.5, and 5 mM EDTA.

6.2.2 NMR spectroscopy

All data were recorded on Bruker AVANCE III spectrometers equipped with TXI cryoprobes and z-axis pulse field gradients. TopSpin 2.1 or 3.0 software was used for data collection. The temperature was calibrated prior to data collection using a sample of 99.8% d_4-methanol [313]. All relaxation experiments were collected as interleaved, pseudo-3D datasets where 1/3 of the total number of planes was duplicated for estimating uncertainties in peak height. NUS pulse programs were hardcoded to utilize a non-uniform dwell in TopSpin 2.1 syntax but were otherwise identical to the RSS versions with the exception of ^{2}H relaxation experiments which required a receiver phase flip in the NUS versions for proper quadrature selection. NUS sampling schedules were generated using the PoissonGap2 program [204]. The sinusoidal weight was set to 2. The seed value for generating all schedules was 12321. Sampling schedules were not optimized or ranked prior to use. Experiment-specific parameters and delays for the RSS experiments are shown below. NUS experiments were setup identically to the RSS experiments but utilized 25% sampling in the indirect dimension.

^{15}N T_1 experiments [314] collected on ubiquitin were performed at 11.7 T at 25˚C. The spectral widths were 10504.20 Hz (F2, ^{1}H) and 1417.64 Hz (F1, ^{15}N), corresponding to 21 ppm and 28 ppm, respectively. The acquisition time in F2 was 97.5 ms while the acquisition time in F1 was varied between 33.9 ms and 141.0 ms. The delay durations were: 0.030 s (twice), 0.800 s, 0.084 s, 0.670 s (twice), 0.154 s, 0.549 s, 0.237 s, 0.435 s, 0.331 s (twice). 8 transients/FID were collected and the recycle delay was 2.0 s. ^{15}N T_1 experiments collected on CaM were performed at 14.1 T at 35˚C. The spectral widths were 13157.90 Hz (F2, ^{1}H) and 1700.91 (F1, ^{15}N), corresponding to 22 ppm and 28 ppm, respectively. The acquisition time was 77.9 ms in F2 and 75.3 ms or 108.2 ms in F1. The delay durations were: 0.026 s (twice), 0.700 s, 0.073 s (twice), 0.587 s, 0.135 s, 0.480 s, 0.207 s, 0.381 s, 0.290 s (twice). 8 transients/FID were collected and the recycle delay was 2.0 s.

^{15}N T_2 experiments [314] collected on ubiquitin were performed at 11.7 T at 25˚C. The spectral widths were 10504.20 Hz (F2, ^{1}H) and 1417.64 Hz (F1, ^{15}N), corresponding to 21 ppm and 28 ppm, respectively. The acquisition time in F2 was kept constant at 97.5 ms while the acquisition time in F1 was either 33.9 ms or 84.7 ms. The delay durations were: 0.016 s (twice), 0.164 s, 0.033 s, 0.145 s (twice), 0.049 s, 0.131 s, 0.066 s, 0.099 s, 0.082 s (twice). 8 transients/FID were collected and the recycle delay was 2.0 s. The effective CPMG field was 4.0 kHz.

^{2}H I_ZC_Z compensated $R^Q(D_z)$ and $R^Q(D_+)$ experiments [315] collected on ubiquitin were performed at 17.6 T at 25˚C. The spectral widths were 15756.303 Hz (F2, ^{1}H) and 3713.302 (F1, ^{13}C), corresponding to 21 ppm and 19.7 ppm, respectively. The acquisition time was 65.0 ms in F2 and 17.2 ms in F1. $R^Q(D_z)$ delay durations were: 0.2 ms (twice), 20.0 ms, 2.1 ms, 16.8 ms (twice), 3.9 ms, 13.7 ms, 5.9 ms, 10.9 ms, 8.3 ms (twice). $R^Q(D_+)$ delay durations were: 3.0 ms (twice), 75.0 ms, 8.0 ms, 63.0 ms (twice), 14.0 ms, 51.0 ms , 22.0 ms, 41.0 ms, 31.0 ms (twice). 48 transients/FID were collected and the recycle delay was 2.0 s. The $R^Q(D_+)$ spin lock strength was 1.0 kHz.

^{13}C T_1 and $T_{1\rho}$ experiments [262] collected on AK were performed at 17.6T at 25˚C. The spectral widths were 15756.303 Hz (F2, ^{1}H) and 3015.885 (F1, ^{13}C), corresponding to 21 ppm and 16 ppm, respectively. The acquisition time was 65.0 ms in F2 and 33.1 ms in F1. ^{13}C T_1 delay durations were 0.044 s, 1.200 s, 0.126 s (twice), 1.000 s (twice), 0.231 s, 0.823 s, 0.356 s, 0.653 s, 0.497 s (twice). ^{13}C $T_{1\rho}$ delay durations were 2.00 ms, 100.00 ms, 10.48 ms (twice), 83.81 ms (twice), 19.25 ms, 68.59 ms, 29.63 ms, 54.43 ms, 41.41 ms (twice). For the ^{13}C T_1 experiment, 48 transients/FID were collected and the recycle delay was 2.5 s. For the ^{13}C $T_{1\rho}$ experiment, 64 transients/FID were collected and the recycle delay was 2.5 s. The ^{13}C $T_{1\rho}$ spin lock strength was 3.7 kHz.

TROSY-detected ^{15}N T_1 and $T_{1\rho}$ experiments [316] collected on AK were performed at 17.6T at 25˚C. The spectral widths were 15756.303 Hz (F2, ^{1}H) and 2316.975 (F1, ^{15}N), corresponding to 21 ppm and 30.5 ppm, respectively. The acquisition time was 65.0 ms in F2 and 27.6 ms in F1. ^{15}N T_1 delay durations were 0.00 s, 2.00 s, 0.32 s (twice), 1.44 s (twice), 0.64 s, 0.96 s. ^{15}N $T_{1\rho}$ delay durations were 1.0 ms, 40.0 ms, 7.7 ms (twice), 30.5 ms (twice), 14.1 ms, 22.0 ms. For the ^{15}N T_1 experiment, 128 transients/FID were collected and the recycle delay was 2.3 s. For the ^{15}N $T_{1\rho}$ experiment, 160 transients/FID were collected and the recycle delay was 2.3s. The ^{15}N $T_{1\rho}$ spin lock strength was 1.5 kHz. TROSY-detected ^{15}N T_1 and $T_{1\rho}$ experiments collected on ubiquitin in 30% glycerol were performed at 11.7T at 25˚C. The spectral widths were 11961.72 Hz (F2, ^{1}H) and 1414.358 (F1, ^{15}N), corresponding to 24 ppm and 28 ppm, respectively. The acquisition time was 85.7 ms in F2 and 90.5 ms in F1. ^{15}N T_1 delay durations were 0.00 s, 0.88 s, 0.16 s (twice), 0.64 s (twice), 0.32 s, 0.48 s. ^{15}N $T_{1\rho}$ delay durations were 1.0 ms, 65.0 ms, 12.5 ms (twice), 49.5 ms (twice), 23.0 ms, 35.0 ms. 40 transients/FID were collected and the recycle delay was 2.5 s. The ^{15}N spinlock field strength was 1.5 kHz. TROSY-detected ^{15}N T_1 and $T_{1\rho}$ experiments collected on MBP were performed at 17.6T at 35˚C. The spectral widths were 15756.303 Hz (F2, ^{1}H) and 2354.952 Hz (F1,^{15}N), corresponding to 21 ppm and 31 ppm, respectively. The acquisition time was 65.0 ms in F2 and 84.9 ms in F1. ^{15}N T_1 delay durations were 0.00 s, 2.00 s, 0.32 s (twice), 1.44 s (twice), 0.64 s, 0.96 s. ^{15}N $T_{1\rho}$ delay durations were 1.0 ms, 40.0 ms, 7.7 ms (twice), 30.5 ms (twice), 14.1 ms, 22.0 ms. 8 transients/FID were collected and the recycle delay was 3.0 s. The ^{15}N spinlock field strength was 2.0 kHz.

All data were processed using NMRPipe/NMRDraw [159]. All RSS and NUS data used for comparison were processed identically with the exception of the ^{2}H relaxation experiments which required a sign alternating Fourier transform in the NUS dimension. NUS data were reconstructed using the hmsIST software (v. 211) [204]. A threshold of 98% and 400 iterations of IST were used for all data sets. Increasing the threshold value and number of iterations did not change the results described significantly (data not shown).

A subset of the data was also reconstructed with IRLS [302] as implemented in the MddNMR software package [301]. All IRLS reconstructed data utilized 50 iterations.

A subset of the data was also reconstructed with the NESTA algorithm [305, 306] using L1 or IRL1 regularization [210] as implemented in the NESTA-NMR program [307]. For L1 regularization, 50 iterations were used. For IRL1 regularization 30 iterations were used and the number of re-weighted iterations was set to 5. Increasing the number of iterations and number of re-weighted iterations did not change the results described significantly (data not shown).

The final matrix size for each RSS and NUS data set used in an individual comparison was identical.

6.2.3 Fitting of relaxation data

Relaxation rates were determined by quantification of the maximum cross peak height as a function of relaxation delay period using the built-in peak picking functionality and seriesTab program of NMRPipe/NMRDraw. Rates were fitted using in-house software written in Python to a two-parameter single exponential decay:

$$I(t) = I_0 \exp(-Rt) \tag{6.1}$$

Where $I(t)$ is the cross peak height at relaxation delay time, t, and I_0 is the initial cross peak height. Best-fit parameters for rates, R, were determined using a Levenberg-Marquardt weighted non-linear least squares algorithm which minimizes χ^2, the sum of the square of the residuals:

$$\chi^2 = \sum_{n=1}^{N} \frac{(I_{calc}(t) - I_{\exp}(t))^2}{\sigma^2} \tag{6.2}$$

Where $I_{calc}(t)$ is the peak height at time t calculated from the fit, $I_{\exp}(t)$ is the experimentally measured peak height at time t, σ is the uncertainty in the peak height and N is the number of planes comprising the relaxation series. The uncertainty was taken to be the standard deviation of the differences in peak height based on replicate measurements scaled by $\sqrt{2}$ [265]. In cases where the data were collected with exceptional S/N, σ was increased by a factor of 2. Errors in the fitted parameters were obtained from the covariance matrix of the fitting routine or from 250-500 Monte Carlo simulations [266] where the sampling bounds were defined by the uncertainty in peak height. Errors obtained from both methods were generally in good agreement.

6.2.4 Statistical analysis

Relaxation rates and peak heights derived from RSS and NUS were compared using the square of the Pearson correlation coefficient (R^2) and the RMSD. R^2 values were calculated as:

$$R^2 = \left[\frac{\sum_{k}^{n} (y_k^{RSS} - \overline{y}^{RSS})(y_k^{NUS} - \overline{y}^{NUS})}{\sqrt{\sum_{k}^{n} (y_k^{RSS} - \overline{y}^{RSS})^2 \sum_{k}^{n} (y_k^{NUS} - \overline{y}^{NUS})^2}} \right]^2 \tag{6.3}$$

RMSD was calculated as:

$$RMSD = \sqrt{\frac{\sum_{k}^{n} (y_k^{RSS} - y_k^{NUS})^2}{n}} \tag{6.4}$$

Where y is any NMR observable (e.g. peak height, relaxation rate) for a given cross peak, k, and n is the total number of cross peaks. The over bar denotes the mean value. In cases where RMSD is reported as a percentage, it has been normalized by the mean value of the NUS-derived observables.

Percent error was calculated as:

$$Error = \frac{\left| y_k^{NUS} - y_k^{RSS} \right|}{y_k^{RSS}} \; x\, 100 \tag{6.5}$$

where y is any NMR observable (e.g. peak height, relaxation rate) for a given cross peak, k. In cases where error distributions are shown, the absolute value is not calculated. Comparison of RSS and NUS data was performed only for sites where the quality of fit was sufficiently high in the RSS reference as judged by the reduced chi-squared value.

The interquartile range (IQR) method was used to identify outliers using in-house software as previously described [317]. Briefly, any metric exhibiting > 1.5 times the IQR above the third quartile or below the first quartile was considered an outlier.

In some cases, linear regression was performed where both dependent and independent variables had associated uncertainties. Under these conditions, regression was performed using the orthogonal distance regression method [318]. This was accomplished using the ODRPACK FORTRAN-77 library implemented in SciPy (http://www.scipy.org/) and employs a Levenberg-Marquardt minimization for parameter estimation. Initial guesses for the fitted parameters were the best-fit parameters from a standard un-weighted linear regression.

6.3 Results

6.3.1 Choice of sampling density

IST belongs to a larger family of compressed sensing (CS) techniques that seek to reconstruct data objects from an incomplete number of data samples. Theory indicates that this can be done by minimizing a regularization term, such as the l_p -norm (for $p \in (0,1)$), of the preferred solution provided that the object is sufficiently sparse [208, 209, 211, 319-321]. In general, frequency-domain NMR data and noise do not meet this requirement so the predictive value of CS theory in this context is limited [212]. In qualitative terms, it is well known that the sampling density (percentage of total points collected in the NUS dimension) must increase with decreasing dimensionality and increasing cross peak number and spectral complexity. Formally, the number of cross peaks must be considered independently for each vector. Previous work has recommended that ~20% sampling be used per NUS dimension [322] based on the number and distribution of cross peaks for proteins typically studied using NMR, though it was noted that this recommendation may not be applicable in all contexts. Therefore, the optimal sampling density for an HSQC-type correlation experiment that would typically be employed in a classical pseudo-3D relaxation experiment was determined first. Of interest is the fidelity of the NUS-reconstructed peak height relative to the same peak height derived using RSS. Figure 6.1 shows the dependence of the R^2 and the RMSD in peak height between RSS sampled and NUS ^{15}N HSQC experiments collected on ubiquitin with a moderate number of sampled points in the indirect dimension (120* RSS points, 28 ppm spectral width). The number of transients was kept constant and only sampling density was allowed to vary. The total number of points after reconstruction was the same for all data. Consistent with previous work, IST reconstruction is shown to yield highly linear peak heights with respect to RSS-derived references and 20% sampling yields quantitatively accurate peak heights with a R^2 of 0.982 and a RMSD of 4.97%. The fidelity of NUS-derived peak heights begins to plateau at 25% sampling with a R^2 of 0.996 and a RMSD of 1.15%. Therefore, all subsequent studies employed 25% sampling.

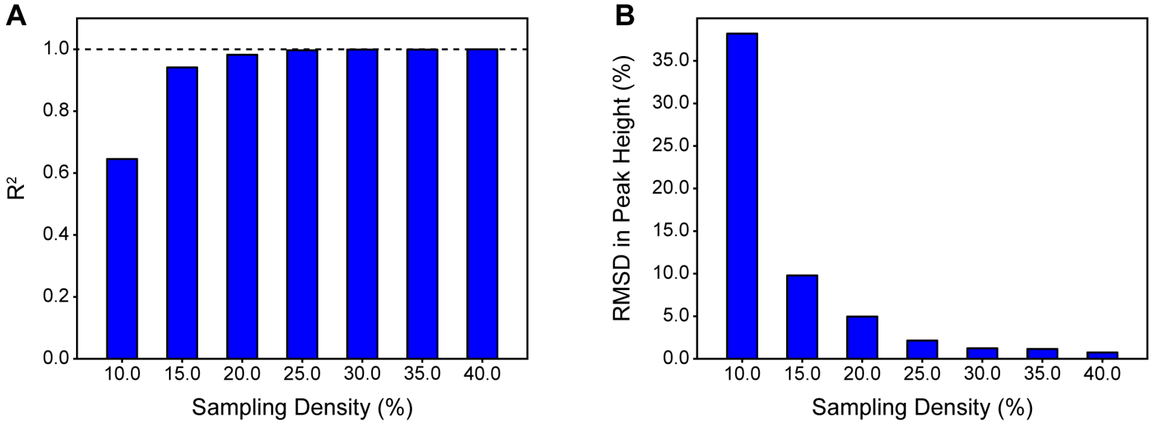


Figure 6.1 The accuracy of NUS-derived peak heights depends strongly on sampling density. Regular sequentially sampled (RSS)-derived and NUS-derived peak heights for a ^{1}H-^{15}N HSQC of ubiquitin are compared quantitatively as a function of sampling density. (A) Dependence of the R^2 for the correlation between RSS-derived and NUS-derived peak heights as a function of sampling density. (B) Dependence of the RMSD of the same correlation as a function of sampling density. The number of transients and spectral width of the NUS dimension were kept constant and the final size of all reconstructed data was identical to that of the RSS data.

6.3.2 Evaluation of NUS-derived relaxation rates

In order to assess the reliability of NUS-derived relaxation rates, RSS and NUS ^{15}N T_1 relaxation experiments [314] were performed on ubiquitin. The NUS data were collected experimentally rather than derived from re-sampling of the RSS reference data set because small but sometimes significant differences between experimentally-determined rates and re-sampled rates were observed (data not shown). The origin of these differences is not clear but it does not seem to be due to spectrometer stability as the data were collected in an interleaved manner and replicates of the entire RSS and NUS time series yield highly similar results (*vide infra*). The data were collected with the same moderate number of points in the indirect dimension (120* RSS points, 30* NUS points, 28 ppm spectral width). In the analysis that follows, the accuracy of an NUS-derived relaxation rate is judged with respect to the same rate derived from a RSS reference data set. Figure 6.2a shows the correlation between R_1 rates derived from RSS and 25% NUS. Relaxation decays obtained from NUS data were nearly identical to those obtained from RSS data (Figure 6.3). Uncertainties in peak heights (estimated from duplicate spectra) and errors in fitted rates were slightly larger for NUS-derived data (Figure 6.4), which indicates that

NUS-derived peak heights are less reproducible than RSS-derived peak heights. The correlation between relaxation rates derived from RSS and NUS is excellent, with a R^2 of 0.995 and a RMSD of 0.89%. A histogram of the errors between RSS-derived and NUS-derived rates indicates that they are normally distributed around a mean value near zero (0.29%) with a standard deviation of 0.82% (Figure 6.2b). It should be noted that the largest error observed (~2.0%) is comparable to the typical reproducibility of RSS data for this particular experiment [265, 323]. This finding is in contrast to a previous report that relaxation rates derived from Poisson-gap sampling and IST reconstruction are systematically offset by ~5% [322]. This suggests that 25% sampling comprised of 30* NUS points is sufficient to obtain quantitatively accurate relaxation rates from reconstructed spectra of ubiquitin.

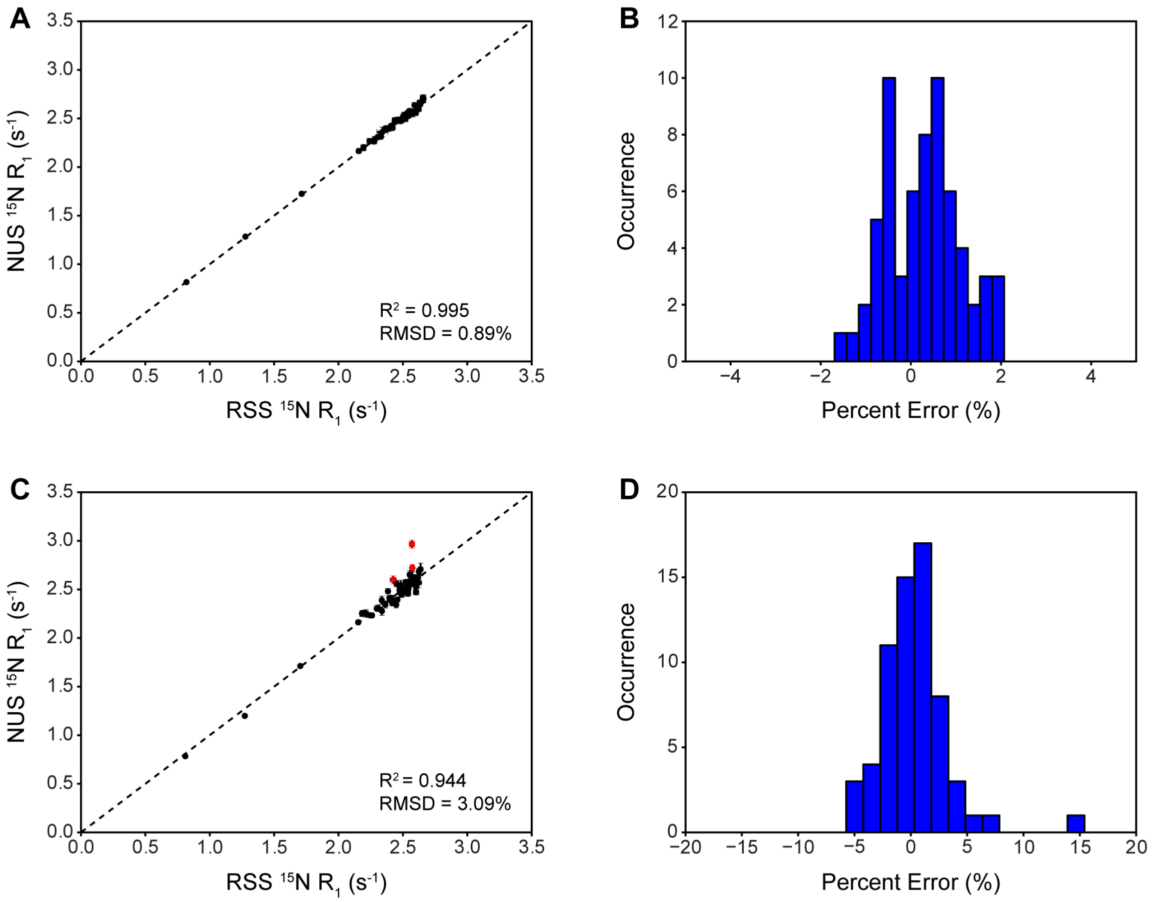


Figure 6.2 Quantitative evaluation of NUS-derived relaxation rates as a function of number of sampled NUS points for a ^{15}N T_1 experiment on ubiquitin. (A) Correlation between RSS-derived and NUS-derived rates for data collected with 30* NUS points. The dotted line is y = x. (B) Distribution of errors in NUS-derived relaxation rates calculated relative to RSS-derived rates for data collected with 30* NUS points. (C) Correlation between RSS-derived and NUS-derived rates for data collected with 12* NUS points. The dotted line is y = x. Outliers identified using the IQR method are shown in red. (D) Distribution of errors in NUS-derived relaxation rates calculated relative to RSS-derived rates for data collected with 12* NUS points.

While sampling densities of 20-25% have been previously reported in the literature to yield quantitatively accurate peak heights [204] and relaxation dispersion parameters [269] for 2D spectra of other proteins, the choice of the total number of points used in such studies is often not fully explained. Indeed, there is no general consensus for how far an indirect dimension should be sampled [201] though the signal-to-noise ratio (S/N) is maximized when the indirect dimension is sampled to 1.26*T_2 [324] and resolution is maximized when the indirect dimension is sampled to

3.14*T_2. However, recent work has shown that weighted NUS can lead to simultaneous improvements in S/N and resolution beyond 1.26*T_2 [325]. Sampling to maximize the resolution or sometimes even the S/N can lead to prohibitively long experiment times for serially collected data. It is also difficult to define a minimum number of sampled NUS points required for faithful reconstruction *a priori* since many confounding issues contribute to the quality of NUS reconstruction. Moreover, it has been noted that the probability of finding a "good" Poisson-gap sampling schedule decreases as the number of desired points decreases, presumably due to the stochastic nature of small sets [326].

The extent to which the total number of points could be decreased was thus explored empirically while maintaining a fixed sampling density. Figure 6.2c shows the correlation between R_1 rates derived from RSS and NUS where the total number of points had been reduced by a factor of 2.5 but a 25% sampling density was maintained (48* RSS points, 12* NUS points, 28 ppm spectral width). Perhaps surprisingly, as few as 12* NUS points yielded interpretable spectra albeit with noticeable artifacts. All resolved cross peaks exhibited typical relaxation decays that still could be fit. It is clear, however, that the accuracy of the rates has been compromised significantly by such drastic undersampling. The correlation has a R^2 of 0.944 and a RMSD of 3.09%. Three statistically significant outliers were identified using the interquartile range (IQR) method and these outliers are shown in red. Figure 6.2d shows the distribution of errors between RSS-derived and NUS-derived rates. Though the errors are still normally distributed around a mean value near zero (0.39%), the spread is much larger, with a standard deviation of 3.06%. The three statistically significant outliers exhibit errors over 5.0%, with a maximum error of 15.4%. Though it is difficult to attribute a single source to the degradation in accuracy, it may arise due to consequences associated with violation of the sparseness condition for CS and/or the increased difficulty associated with generating a proper sine-weighted Poisson distribution of gaps for 12* points.

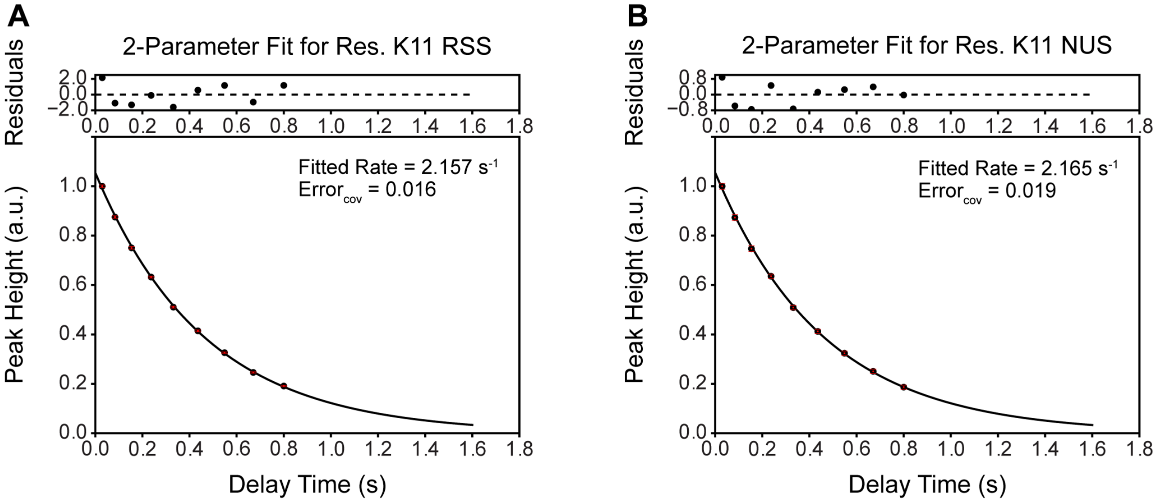


Figure 6.3 Example NUS relaxation decay curves. (A) 2-parameter fit of ^{15}N R$_1$ decay for residue K11 of ubiquitin for RSS data collected with 120* RSS points (28 ppm spectral width). (B) 2-parameter fit of ^{15}N R$_1$ decay for resiude K11 of ubiquitin for data collected with 25% NUS, 30* NUS points (28 ppm spectral width). Error bars representing the uncertainty in peak height are shown in red and are smaller than the symbols used. The fitted rate and error in the fitted rate determined by the covariance matrix of the fit are shown.

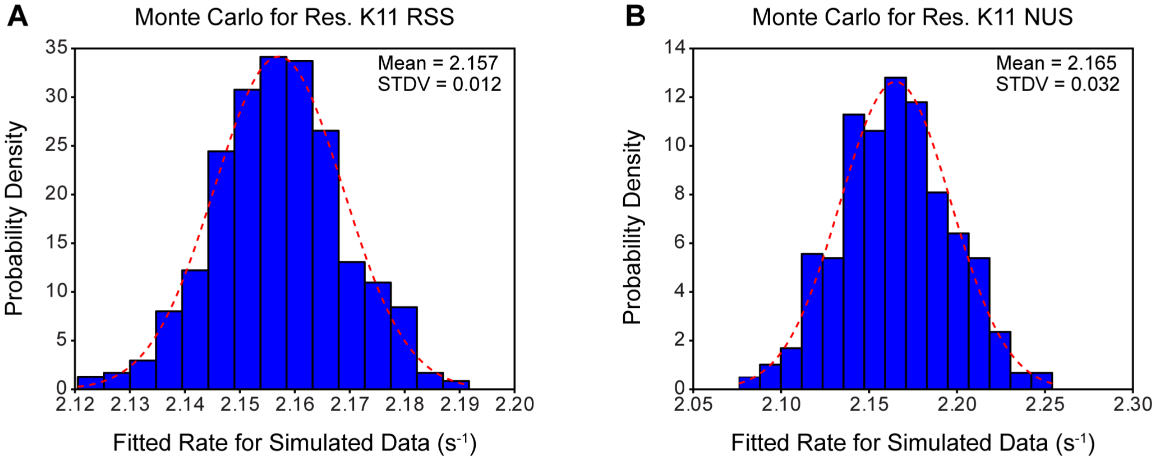


Figure 6.4 Monte Carlo analysis of errors in NUS relaxation rates. (A) Histogram of fitted R_1 rates from 500 Monte Carlo simulations of the data shown in Figure 6.3. (B) Histogram of fitted R_1 rates from 500 Monte Carlo simulations of the data shown in Figure 6.3. In these simulations, peak heights measured at each delay were varied randomly within the boundaries of their uncertainties then the points were fit with a 2-parameter single exponential decay. The red dashed line is a fit to a normalized probability density function for a normal distribution. The normalization is such that the integral of the fitted function is 1. The standard deviation is an estimate of the error in the fitted rate. The error for the NUS data is ~3-fold larger than the error for the RSS data. This reflects an inherently lower reproducibility in peak height for NUS data relative to RSS data.

Based on the stark differences in performance between the data collected with 30* NUS points and 12* NUS points, additional RSS- and NUS-derived ^{15}N T_1 comparisons were collected to gauge the relationship between the number of NUS points and the accuracy in rates empirically. Each experiment utilized a different number of total points but with a constant sampling density of 25%. In order to estimate contributions from experimental variance, each full RSS/NUS comparison was collected in triplicate—with replicate measurements often occurring several months apart. The statistics for each RSS/NUS comparison were determined individually and then averaged. In general, comparisons were highly reproducible. Figure 6.5 shows the relationship between R^2 and RMSD in relaxation rate and total number of sampled points. The accuracy of NUS-derived relaxation rates exhibits a rather steep dependence on the total number of sampled points. The dependence appears to plateau at 30* points where the number of measurements is sufficiently high to accurately reconstruct the spectrum of ubiquitin. Increasing the total number of sampled points beyond this does not lead to improvement.

Since the data shown in Figure 6.5 consisted of variable numbers of NUS points, each bar represents a different sized matrix. To explore if this affected peak height reconstruction, all of the RSS and NUS data were re-processed to a uniform size of 512* points in the indirect dimension via zero-filling. The observed dependencies did not change significantly (data not shown).

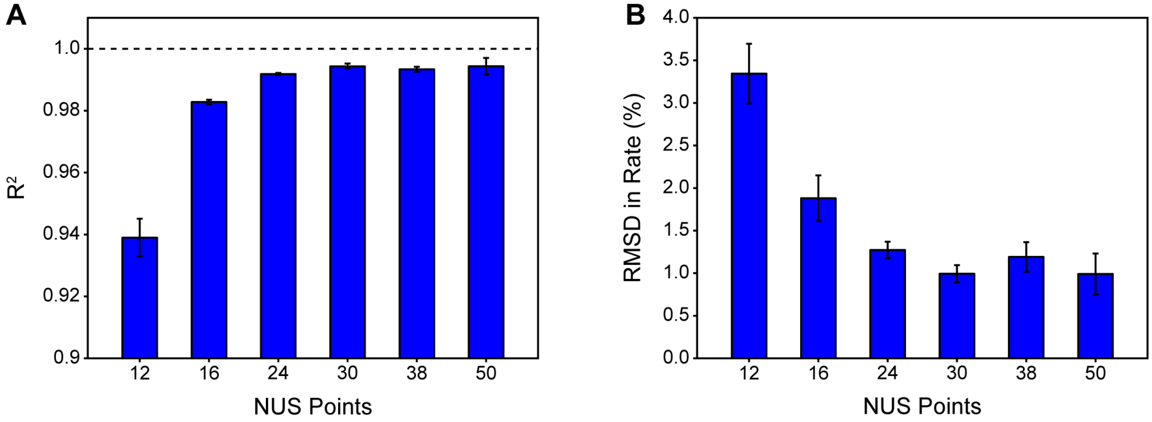


Figure 6.5 The accuracy of NUS-derived relaxation rates depends strongly on the number of sampled NUS points at a fixed sampling density of 25%. RSS-derived and NUS-derived ^{15}N R_1 rates for ubiquitin are compared quantitatively for spectra collected with a variable number of points at a constant spectral width. (A) Dependence of the R^2 on the number of NUS points. (B) Dependence of the RMSD on the number of NUS points. Error bars are one standard deviation from the mean of three independent replicate data sets.

6.3.3 The nature of inaccuracies in NUS-derived relaxation rates

To understand the nature of inaccurate NUS-derived relaxation rates, the fidelity of peak height reconstruction was further characterized for the data set collected with 12* NUS points, which exhibited the largest deviations from the reference RSS-derived rates. Errors in NUS-derived relaxation rates did *not* correlate with errors in NUS-derived peak heights from any plane in the relaxation series (Figure 6.6). Instead, the largest errors in relaxation rate tended to originate from cross peaks with *inconsistent* errors in peak height across the relaxation series. Figure 6.7 illustrates this point by showing the error in reconstructed peak height as a function of relaxation plane number. Data is shown for the cross peak with the highest error in NUS-derived relaxation rate and the cross peak with lowest error in NUS-derived relaxation rate. Both cross peaks exhibited inaccurate peak height reconstruction in the first plane of the relaxation series (~15% error). However, their respective errors in relaxation rate ultimately differed by over 15 fold. The cross peak with the lowest error in relaxation rate maintained nearly the same ~15% peak height error across the entire relaxation series (standard deviation in peak height error ~2%). The cross peak with the highest error in NUS-derived relaxation rate exhibited a peak height error which increased dramatically and in a non-linear fashion across the relaxation series (standard deviation in peak height error ~10%). This demonstrates that changes in the consistency of peak height reconstruction can manifest as large changes in relaxation rate. It is important to note that these data were collected in an interleaved manner so the observed non-linearity is unlikely to arise due to spectrometer stability. Importantly, similar trends were also observed for data that were not as significantly undersampled.

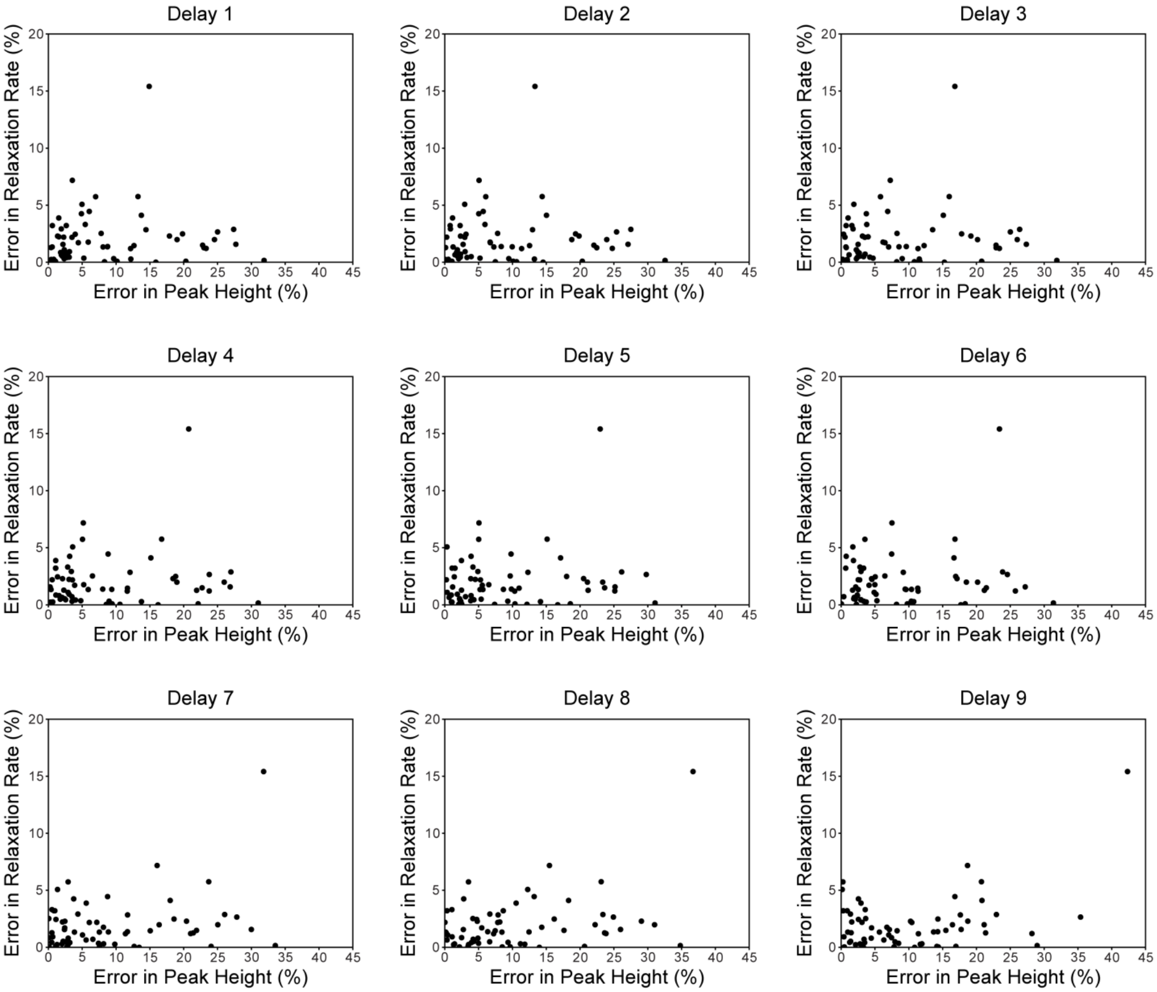


Figure 6.6 The error in NUS-derived peak height is uncorrelated with the error in NUS-derived relaxation rate. NUS-derived peak heights were compared to RSS-derived references for each plane of a ^{15}N T_1 experiment collected on ubiquitin with 48* RSS points, 12* NUS points, and 28 ppm spectral width.

The consistency of peak height reconstruction was then quantitatively evaluated. For each cross peak, the RSS-derived peak height was plotted against the NUS-derived peak height for each plane in the relaxation series and the result was fitted to a straight line. The deviation of the fitted intercept from zero is used an estimate of the non-linearity in peak height across all planes in the relaxation series (the reasoning behind this approach is illustrated in more detail in Figure 6.8). We define the absolute value of the fitted intercept as the “non-linearity factor.” The non-linearity factor is a cross peak specific quantification of how non-linear peak height reconstruction is across a relaxation series. Figure 6.7 shows that the non-linearity factor of each cross peak is strongly correlated with the error in its NUS-derived relaxation rate. This indicates that the consistency of peak height reconstruction across the relaxation series is a strong determinant of the accuracy of NUS-derived relaxation rates. Such behavior could not be detected by just comparing peak heights to a single RSS reference plane as has been done in the optimization of sampling schedule generators [205] or determination of minimum sampling density [269].

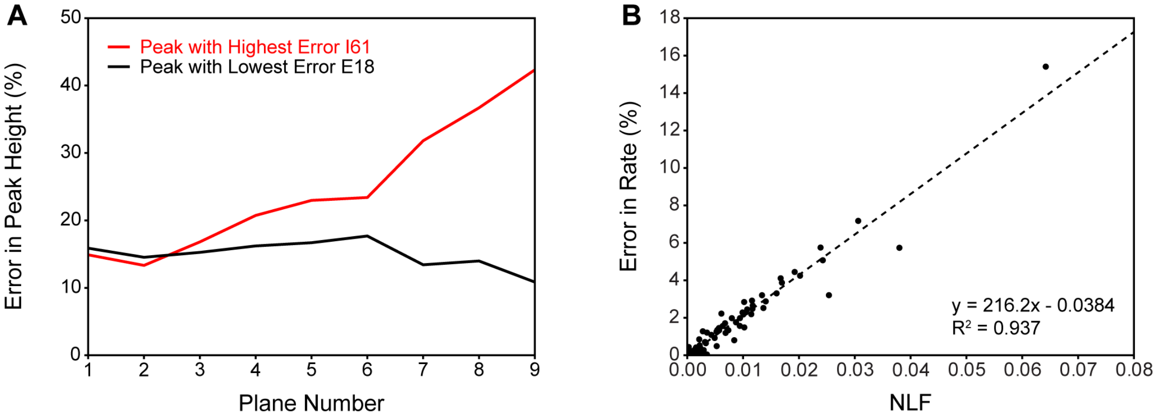


Figure 6.7 Inaccuracy in NUS-derived relaxation rates stem from a lack of peak height reconstruction consistency. Errors in NUS-derived peak heights and relaxation rates calculated relative to RSS-derived references were quantitatively analyzed for ^{15}N R_1 rates of ubiquitin for data collected with 12* NUS points (A) Error in NUS-derived peak height as a function of relaxation plane number (relaxation delay) for the cross peak with the largest error in NUS-derived relaxation rate (red) and the cross peak with the smallest error in NUS-derived relaxation rate (black). (B) The error in NUS-derived relaxation rate as a function of the non-linearity factor (NLF) (see full text and Fig. S6 for a more complete explanation). The NLF is a cross peak specific metric that is correlated with the error in NUS-derived relaxation rate. The dotted line is the best-fit line determined by linear regression: $y = 216.2x - 0.0384$, R^2 = 0.937.

The average of all non-linearity factors reports on the overall accuracy of an entire NUS relaxation series and is useful for comparing the performance of different relaxation series. Figure 6.9 shows that the average non-linearity factor scales with the total number of sampled points and converges once the number of sampled points is sufficiently high to accurately reconstruct the spectrum of ubiquitin. The average non-linearity factor decreases substantially as a function of sampled points, though it never reaches zero, even when as many as 50* NUS points are used.

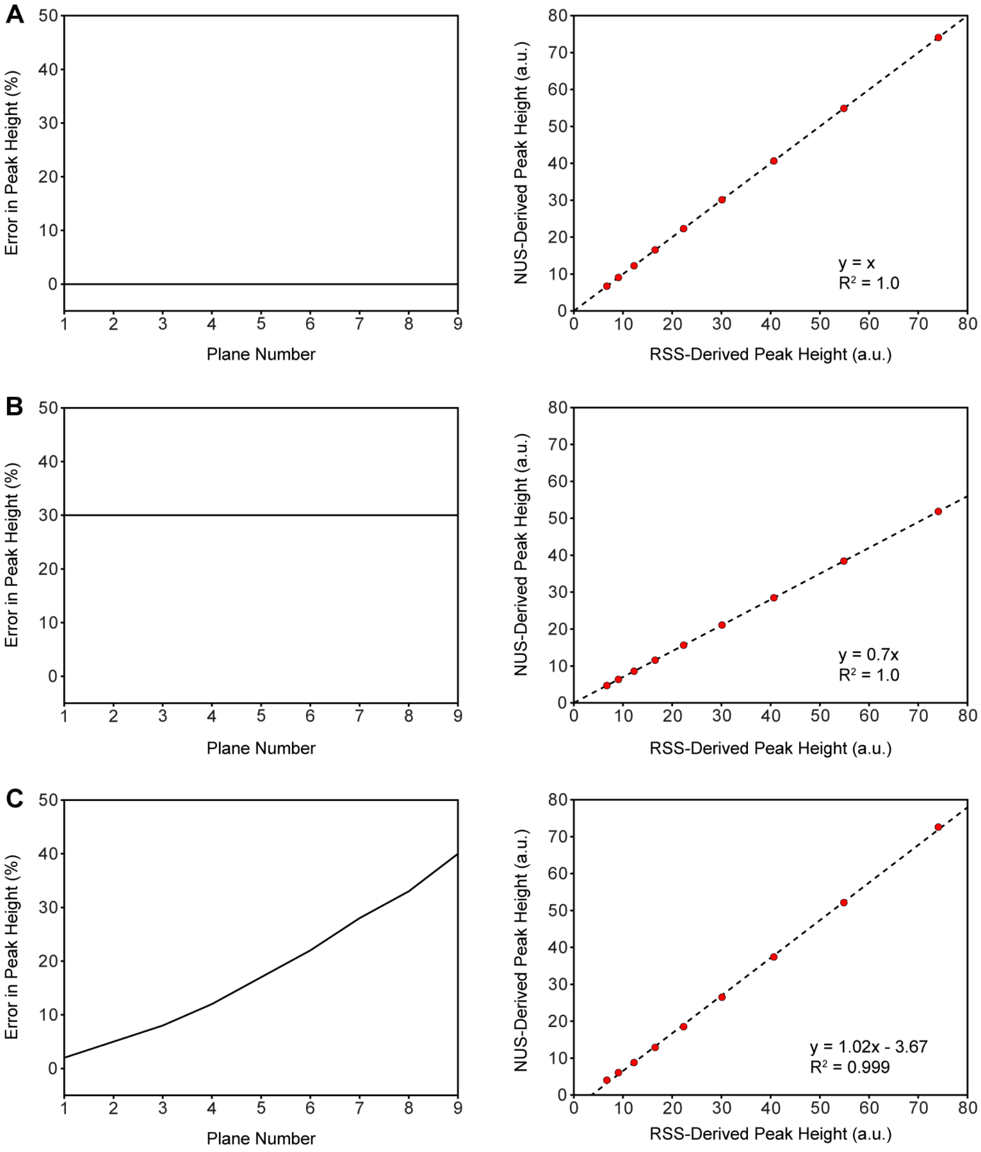


Figure 6.8 Illustration of the derivation of the non-linearity factor. Left panels show the error in NUS-derived peak height for a single cross peak as a function of delay. Right panels show the correlation between RSS-derived and NUS-derived peak heights of the same cross peak where each data point is taken from a single plane in the relaxation series. (A) If peak heights are reconstructed exactly across all planes in the seires then reconstruction is accurate and consistent. Correlations of RSS-derived and NUS-derived peak heights over all planes in the series are straight lines with a slope of 1 and zero intercept. (B) If peak heights are reconstructed with a 30% error across all planes in the series then reconstruction is inaccurate but consistent. Correlations of RSS-derived and NUS-derived peak heights over all planes in the series are straight lines with a scaled slope and zero intercept. (C) If peak heights are reconstructed with variable errors across all planes in the series then reconstruction is neither accurate nor consistent. Correlations of RSS-derived and NUS-derived peak heights are slightly non-linear resulting in a non-zero intercept. All cross peaks are subject to this analysis and the absolute value of the intercept is defined as the non-linearity factor.

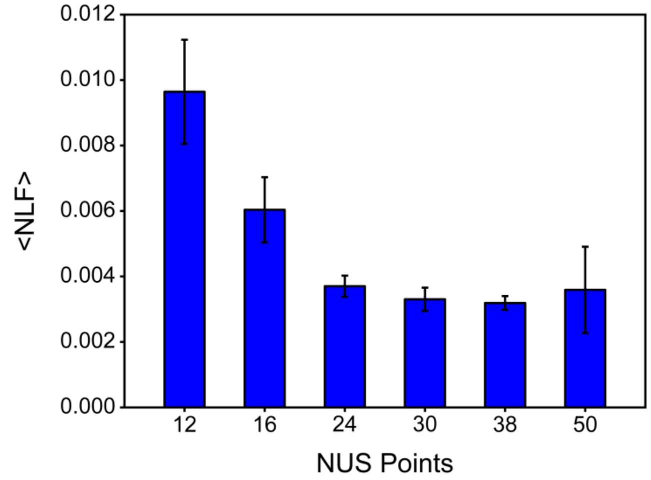


Figure 6.9 Peak height reconstruction linearity improves as a function of the number of NUS points at a fixed sampling density of 25%. The average non-linearity factor decreases with increasing number of NUS points.

6.3.4 Generalization to other reconstruction algorithms

In order to assess if the non-linearity in peak height reconstruction was specific to the IST implementation in hmsIST, the ubiquitin ^{15}N T_1 data sets collected with 12* and 30* NUS points were re-processed with IRLS, NESTA-L1, NESTA-IRL1. While all three algorithms could reconstruct the 30* NUS points data, only IRLS yielded quantifiable reconstructions for the 12* points data. The accuracy of the rates relative to the RSS reference data is shown in Table 6.1. In general, IRLS yielded rates with accuracies comparable to those obtained using hmsIST whereas NESTA-L1 and NESTA-IRL1 yielded slightly less accurate rates.

Figure 6.10a-d shows the error in reconstructed peak height as a function of relaxation plane number for the cross peak which exhibited the highest error in relaxation rate and the cross peak which exhibited the lowest error in relaxation rate. Again, high errors in relaxation rate are shown to arise from non-linearities in peak height across the relaxation series as was demonstrated for data reconstructed with hmsIST, though the nature of the non-linearity seems to be variable. Importantly, errors in NUS-derived relaxation rates are found to be strongly correlated with non-linearity factors as shown Figure 6.10e-h. All correlations yielded R^2 values greater than 0.8, indicating that peak height non-linearity may be a general origin of errors in NUS-derived relaxation rates and not just specific to reconstructions performed using hmsIST.

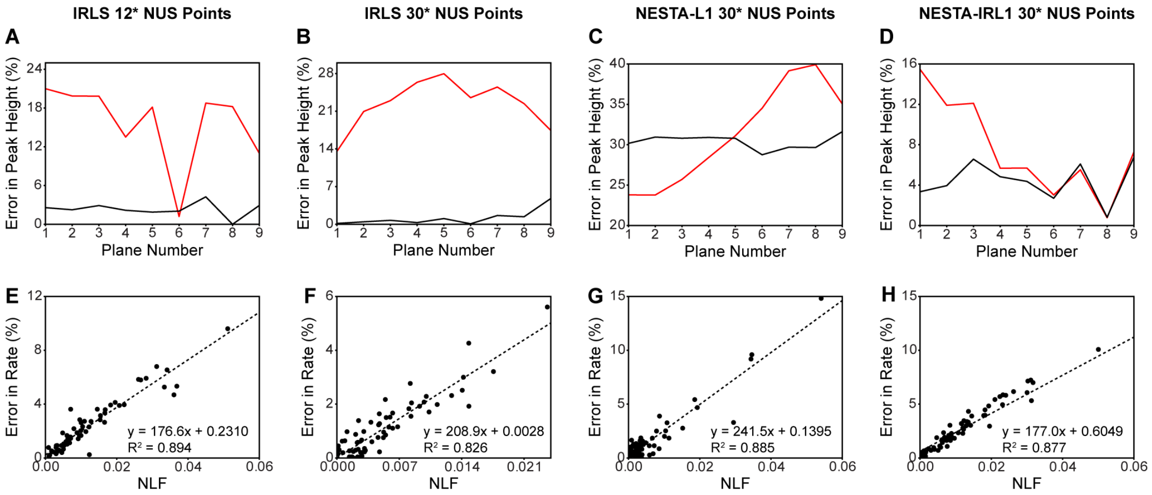


Figure 6.10 The error in NUS-derived peak height as a function of relaxation delay for different reconstruction algorithms (A-D). Lines in red correposond to cross peaks which exhibited the largest error in relaxation rate. Line in black correspond to cross peaks which exhibited the smallest error in relaxation rate. The correlation between errors in relaxation rate and non-linearity factors for different reconstruction algorithms (E-F).

6.3.5 Generalization to other proteins

Additional RSS/NUS comparisons using the proteins calcium-saturated calmodulin (CaM), arginine kinase (AK), and maltose binding protein (MBP) were collected to assess the generality of these findings. Table 6.2 shows the relevant protein details and statistics for each of these experiments. In each experiment, the minimum number of scans to obtain reliable RSS data was used and then the remaining time was invested toward sampling of the indirect dimension. Most experiments yielded high correlations ($R^2 > 0.95$) with reasonable RMSDs (< 4%). Additionally, poorer correlations and RMSDs were again shown to improve dramatically with an increase in NUS points as evidenced by the CaM data which exhibits twice as many cross peaks as ubiquitin and limited spectral dispersion. Notably, all data shown in Table 6.2 exhibit a correlation between errors in NUS-derived relaxation rate and non-linearity factors ($\langle R^2 \rangle = 0.769 \pm 0.141$; n = 14). These data confirm that the conclusions drawn from the ubiquitin data are general to many types of relaxation measurements and to proteins with different properties. Perhaps the most impressive example of this is the ^{13}C relaxation data collected on a very dilute (~75 μM) solution of the 40 kDa monomer arginine kinase (AK). Though ^{13}C relaxation experiments are quite sensitive, the rather dilute sample and room temperature data collection make their execution challenging. In this sample, only ILV methyl groups are labeled so the total number of peaks of low. However, the narrow chemical shift dispersion for ILV methyl groups results in a high density of peaks for each directly detected frequency. Despite this, the R^2 values were ~0.98 - 0.99 and the RMSDs were only ~4 - 5.5% using 26* NUS points.

Since all of the proteins used in this study exhibit different degrees of spectral complexity and data were collected at different static field strengths, utilized different spectral widths, and consist of different numbers of total NUS points, it is extremely difficult to present a truly complete comparison across all data sets. However, a reasonably fair comparison can be made with respect to non-linearity factors. Figure 6.11 shows that for all data presented here, the average non-linearity factor is approximately linearly related to the accuracy of NUS-derived relaxation rates. The relationship in Figure 6.11 is essentially a calibration curve that allows one to identify a required average non-linearity factor necessary to obtain a desired accuracy in relaxation rate. For example, if one desired NUS-derived relaxation rates with less than a 5% RMSD relative to RSS-derived reference rates, then the average non-linearity factor should be ~0.02, or for 2% RMSD then it should be 0.007 or better, and so on.

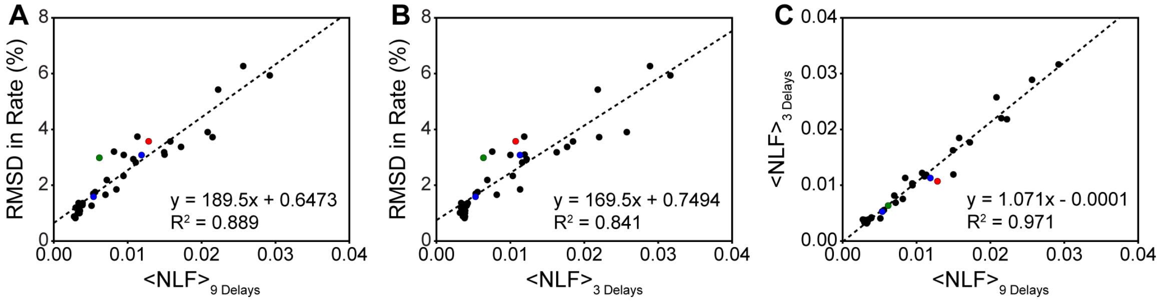


Figure 6.11 Calibration curve relating RMSD in relaxation rate to the normalized sum of non-linearity factors where normalization is done by the number of cross peaks used to calculate the sum of the non-linearity factors. Points in black were obtained by IST reconstruction using hmsIST. Points in blue were obtained by IRLS reconstruction using MddNMR. The point in red was obtained by NESTA-IRL1 reconstruction using NESTA-NMR and the point in green was obtained by NESTA-L1 reconstruction using NESTA-NMR. (A) Correlation between RMSD in relaxation rate and average non-linearity factor (<NLF>) determined from all (nine) delay times. (B) Correlation between RMSD in relaxation rate and <NLF> determined from only three delay times. The goodness of fit indicates that using the three time point duplicates routinely used to estimate intensity uncertainty and in RSS mode are sufficient to confirm the accuracy of obtained relaxation rates. (V) Correlation between <NLF> values determined from nine and three delay times.

6.3.6 Predicting non-linearity factors from reference data

The calibration curve shown in Figure 6.11 is only useful if non-linearity factors can be determined *prior* to full data collection. Ideally, it is most desirable to determine the reliability of a given NUS scheme with little or no penalty in spectrometer time. We therefore explored whether it is possible to estimate non-linearity factors from a smaller set of data that could be collected prior to starting a suite of relaxation experiments. It is apparent that multiple RSS reference planes from the relaxation series must be collected. However, it is unclear how many are required and which planes should be used. Figure 6.12 shows correlations between non-linearity factors derived from a minimal set of data and those derived from a full relaxation series for an example ubiquitin ^{15}N T_1 relaxation series. Figure 6.12a shows that non-linearity factors determined from just two planes are only modestly accurate. However, Figure 6.12b shows that non-linearity factors determined from three planes are quite accurate. This is perhaps unsurprising considering the non-linearity factor is derived by linear regression. It should be noted that the three planes that were chosen were the shortest delay time and the two longest delay times. This was determined empirically but is consistent with the typical nature of peak height non-linearity observed in this study. In order to test the generality of this method, all remaining data sets were subject to the same analysis. Figure 6.12c-d shows that excellent estimations of non-linearity factors can be obtained for the less ideal AK ^{13}C T_1 relaxation data. Figure 6.11 shows that average non-linearity factors estimated from three reference planes are indeed in fantastic agreement with non-linearity factors determined from entire time series for all data presented in this study.

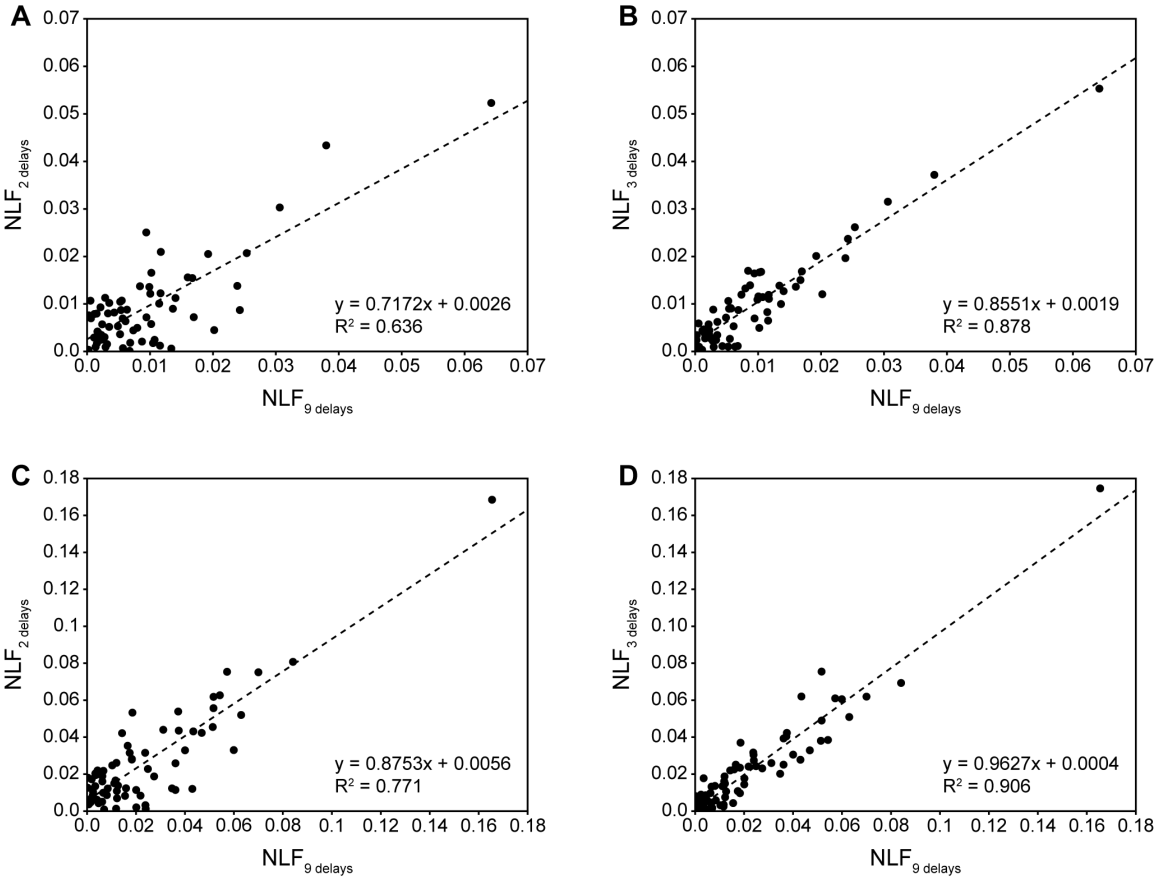


Figure 6.12 Estimation of non-linearity factors from minimal reference data. Correlation between estimated and observed non-linearity factors for ubiquitin ^{15}N T_1 collected with 12* NUS points using (A) 2 RSS reference planes (shortest and longest delays) and (B) 3 RSS reference planes (shortest and two longest delays). Correlation between estimated and observed non-linearity factors for AK ^{13}N T_1 collected with 26* NUS points using (C) 2 RSS reference planes and (D) 3 RSS reference planes.

Table 6.1 Summary of RSS-NUS comparisons using alternative reconstruction algorithms

Algorithm	Protein	NUS Points	R^2	RMSD	<NLF>
IRLS	Ubiquitin	12*	0.950	3.08%	0.01187
IRLS	Ubiquitin	30*	0.983	1.59%	0.00538
NESTA-L1	Ubiquitin	30*	0.958	2.98%	0.00616
NESTA-IRL1	Ubiquitin	30*	0.922	3.57%	0.01284

Table 6.2 Summary of additional RSS-NUS comparisons

Experiment	Protein	Cross Peaks[1]	NUS Points	R^2	RMSD	<NLF>
$^2H\ R^Q(D_Z)$	Ubiquitin	49	16*	0.993	3.10%	0.01499
$^2H\ R^Q(D_+)$	Ubiquitin	49	16*	0.992	3.37%	0.01722
$^{13}C\ T_1$	AK	106	26*	0.990	5.43%	0.02224
$^{13}C\ T_{1\rho}$	AK	106	26*	0.978	3.90%	0.02082
$^{15}N\ T_1$ TROSY	Ubiquitin 30% glycerol	72	32*	0.967	1.66%	0.00699
$^{15}N\ T_1$ TROSY	AK	312	16*	0.881	6.27%	0.02563
$^{15}N\ T_1$ TROSY	MBP	318	50*	0.885	3.57%	0.01577
$^{15}N\ T_{1\rho}$ TROSY	Ubiquitin 30% glycerol	72	32*	0.985	2.30%	0.00943
$^{15}N\ T_{1\rho}$ TROSY	AK	312	16*	0.915	5.94%	0.02923
$^{15}N\ T_{1\rho}$ TROSY	MBP	318	50*	0.942	3.18%	0.01494
$^{15}N\ T_1$	CaM	156	32*	0.919	2.94%	0.01076
$^{15}N\ T_1$	CaM	156	46*	0.975	1.85%	0.00847
$^{15}N\ T_2$	Ubiquitin	72	12*	0.975	3.72%	0.02148
$^{15}N\ T_2$	Ubiquitin	72	30*	0.990	2.82%	0.01109

[1] The number of cross peaks is taken to be the number observed cross peaks identified by automated peak picking

6.4 Conclusions and Future Directions

The goal of this work was to determine if classical NMR spin relaxation experiments could be collected using NUS without a loss in quantitative accuracy. The motivation being that the conventional implementation of these experiments is too insensitive to be performed on LacI. This limitation precludes characterization of fast time scale dynamics and their potential role in allosteric regulation. We have shown that relaxation experiments can indeed be collected using NUS without a loss in quantitative accuracy. Several import conclusions were reached:

- Errors in NUS-derived relaxation rates derive from a non-linearity in peak height across the relaxation series

- Errors become large when data is undersampled. For a given sampling density, the *total* number of NUS points must be chosen judiciously

- Non-linearity in peak height can be quantified using a new metric called the non-linearity factor which can be accurately derived from a sparse set of reference data

- A general linear relationship between non-linearity factor and RMSD in relaxation rate exists

- The accuracy of NUS relaxation experiments can be determined reliably prior to data collection

The potential application of these findings to LacI is clear. The calibration curves relating non-linearity factors and accuracy of relaxation rate (Figure 6.11) allow us to determine the necessary number of NUS points required for a reliable determination of relaxation rates in an efficient manner. This may allow us to finally explore the role of fast timescale motion and the conformational entropy it represents in the allosteric regulation of LacI.

CHAPTER 7: Conclusions and Future Directions

For decades, LacI has served as the paradigmatic example of an allosterically regulated transcription factor. However, from a mechanistic standpoint, many of the fundamental details underlying LacI functionality were poorly characterized. We have sought to characterize the changes in structure and changes in dynamics that accompany IPTG-binding in the absence and presence of operator. Some pressing questions of interest included:

- What is the structure of the DBD in the absence of DNA?
- How does inducer-binding affect the DBD in the absence of DNA?
- What is the structure of the LacI-operator-inducer ternary complex?
- What is the nature of LacI's internal dynamics on the μs-ms timescale?
- What is the nature of LacI's internal dynamics on the ps-ns timescale?
- Are dynamics functionally relevant?

In order to address these issues, we have employed NMR—the only experimental technique capable of providing atomic-level structural and dynamical information for proteins in solution. Ostensibly, NMR was the ideal methodology to address the issues that interested us. However, the high molecular weight of LacI precluded analysis using conventional NMR methods. Our work introduced technical developments both with respect to sample preparation and NMR data collection that ultimately enabled us to address all of the questions above.

In Chapter 2, a novel expression system was introduced that eliminated all of the problems that previously prevented NMR-based characterization of LacI. Using our expression and purification methods, high yields of homogenous, deuterated LacI were consistently obtained which allowed for optimal utilization of TROSY methods. Importantly, our expression platform eliminated the severe toxicity previously associated with recombinant expression of the isolated RD and eliminated the need for fermentation in the production of the isolated DBD. These constructs were critical for proper interpretation of the highly complex NMR spectra of intact LacI.

In Chapter 3, it was shown that the unfavorable properties of LacI rendered resonance assignment impossible using conventional approaches. To overcome this limitation, NUS methods were adapted and in some cases extended to yield a comprehensive library of triple-resonance experiments. By combining NUS methods with the truncated constructs obtained using the methods outlined in Chapter 2, resonance assignment could be obtained for LacI in all of its relevant functionally states.

In Chapter 4, we applied the methods developed in Chapters 2-3 to understand the functional mechanism of LacI in the absence of DNA. Of principal interest was the structure of the DBD and the changes that might occur upon IPTG binding. Our NMR data showed that in the absence of DNA, the DBD is well structured and the hinge is unfolded. Surprisingly, no changes in DBD structure were detected upon IPTG-binding. Instead, extensive changes in the dynamics of the RD were observed. In the absence of IPTG and operator, the RD was shown to be in a dynamic equilibrium between induced and repressed states that exchanged on the μs-ms timescale. Interestingly, exchange rates were found to be heterogeneous across the protein. Dynamics residues were found to cluster in regions of the RD previously implicated in allosteric communication. IPTG-binding largely quenched these dynamics save for three clusters localized to the monomer-monomer interface. The residues in these clusters were previously characterized to be critical to allosteric function. Indeed, two mutants with severely diminished allosteric activity exhibited quenched dynamics in these clusters—suggesting that dynamics are functionally relevant.

In Chapter 5, we characterized the functional mechanism of LacI in the presence of operator DNA. We showed that operator-binding results in widespread quenching of μs-ms dynamics in the RD. We also provided the most definitive evidence to date that the hinge helix of the ternary complex is folded, effectively ruling out a long standing model for LacI's allosteric mechanism which posited that the hinge helix unfolds upon IPTG-binding. These data were ultimately confirmed by quantification of the ps-ns timescale dynamics of L56, a key residue which intercalates into the minor groove of the operator in the repressed state. Upon IPTG-binding, it was found that no significant changes in the motional properties of L56 were detectable. Preliminary analysis suggests that μs-ms motions are activated in the RD upon IPTG-binding. From our data, we can conclude that the RD of the ternary complex is structurally similar to the isolated RD bound to IPTG and that the hinge is folded—thus providing the most comprehensive structural information for the ternary complex to date.

Throughout this work, dynamics were only characterized semi-quantitatively. This stemmed from the fact that the relaxation experiments necessary for quantification of motional processes largely failed on most functional states of LacI. In Chapter 6, we derived the first apparently universal criterion for obtaining highly accurate relaxation rates using NUS: the minimization of the non-linearity factor. Additionally, a simple procedure was introduced for reliably obtaining non-linearity factors from a sparse set of reference data. These methods will likely enable a more quantitative characterization of the dynamical properties of LacI in the near future. Though these methods were developed specifically to enable measurement of ps-ns timescale dynamics, they are applicable to the measurement of dynamic processes on any timescale as long as they are quantified by a serially-collected two-dimensional (pseudo-three-dimensional) experiment.

The work presented here is a significant step forward toward unraveling the mechanism underlying allosteric communication in LacI. However, one can envisage a panoply of follow up experiments. Perhaps of most interest is the characterization of the ternary complex. We have demonstrated that a high-resolution structural characterization of the ternary complex is accessible through NMR. Future work should focus on the expansion of resonance assignments, perhaps through the use of a combination of triple-resonance and NOESY-based experiments as discussed in Chapter 3. These could potentially provide enough restraints for obtaining a backbone fold of the ternary complex. With respect to the other states, methyl resonance assignments may be obtainable with a combination of NOESY and COSY-transfer experiments. To this end, we have recently introduced an NUS 4D NOESY experiment for identifying distances between methyl groups. Coupled with the crystal structures available for the various functional states of the RD, this experiment should enable methyl resonance assignment. Once assignments have been obtained, NUS relaxation experiments can potentially provide a more quantitative description of the dynamic processes operative across the functional cycle of LacI.

APPENDIX 1: NMR Methods

A1.1 Brief introduction to NMR

Nuclear magnetic resonance (NMR) is a spectroscopic technique which detects magnetically active nuclei. These nuclei posses non-zero spin, an intrinsic quantum characteristic. Of interest in most biomolecular (protein and nucleic acid) NMR studies are ^{1}H, ^{13}C, ^{15}N, (all spin ½) and sometimes ^{2}H nuclei (spin 1). Nuclei with non-zero spin are sensitive to external magnetic fields because they possess a magnetic moment, μ, which is proportion to spin, S:

$$\mu = \gamma S \tag{A.1}$$

The proportionality constant, γ, is called the "gyromagnetic ratio." Spin ½ nuclei, can populate two possible quantum states which are degenerate in the absence of an applied magnetic field. Accordingly, both states are populated equally. In the presence of a magnetic field, these states become non-degenerate due to interactions between the magnetic moment and external field—a phenomena known as the "Zeeman effect" or "Zeeman splitting." Since each state now has a different energy (given by $E = \mu B_0$), populations will be skewed toward the lower energy state. For spin ½ nuclei, the two energy states (denoted as α and β) will have populations given by the Boltzmann distribution:

$$\begin{aligned} \alpha \text{ population} &= \frac{N_\alpha}{N_{tot}} = \frac{1}{z} g_\alpha e^{-E_\alpha / k_\beta T} \\ \beta \text{ population} &= \frac{N_\beta}{N_{tot}} = \frac{1}{z} g_\beta e^{-E_\beta / k_\beta T} \end{aligned} \tag{A.2-A.3}$$

Where z is the partition function, g is the degeneracy (1 in this case), and k_β is the Boltzmann constant. The relative population or "fractional polarization" is typically quite small, due to the small difference in energy between the two states. For example, an isolated ^{1}H nucleus will exhibit a fractional polarization of ~4.0 x 10^{-5} at a static field strength of 11.7 T.

The application of an external magnetic field causes spin precession at a frequency known as the "Larmor frequency", ω:

$$\omega = -\gamma B_0 \tag{A.4}$$

Where B_0 is the strength of the magnetic field. The application of electromagnetic radiation at this frequency gives rise to NMR phenomena. These frequencies span the radio frequency (RF) portion of the electromagnetic spectrum. For example, the Larmor frequency of an isolated ^{1}H nucleus at 11.7 T is 500 MHz.

In practice, the Larmor frequency is slightly modulated by the local chemical environment of the nucleus. The degree of modulation is defined as the chemical shift, as discussed extensively in the main text. As such, the chemical shift provides a site-specific readout of the local chemical environment of a given probe. For proteins, this is typically determined by the local structure and presence of ligands.

NMR experiments involve modulating the populations and orientations of spins in order to understand their properties. At equilibrium, spins are oriented along what is called the z-axis as a bulk vector generically called the "magnetization". This axis is defined to be parallel to the direction of the static magnetic field. RF radiation can alter the orientation of the equilibrium magnetization, usually by rotating it to the transverse plane (x,y axes). Over time, spins will begin to return back to their equilibrium orientation in a process called "relaxation." Two processes are operative:

(1) Longitudinal relaxation (T_1): return back to z-axis

(2) Transverse relaxation (T_2): dephasing of the bulk magnetization in the x,y plane

These relaxation processes are strongly influenced by the chemical properties of the nuclei being probed. For proteins, these processes are dependent on the motional properties of the protein.

The theory underlying all of the experiments presented in the main text is well documented [327] and will not be covered in-depth for brevity. Instead, several topics pertinent to the main text will be introduced and describe in the context of how they can provide information about protein structure and protein dynamics.

A1.2 High molecular weight proteins and the TROSY effect

As discussed in Chapter 3, the observable NMR signal is derived from Fourier transform of the FID. The FID decays because it is damped according to the transverse relaxation rate of the nucleus from which it derives. Transverse relaxation rates are proportional to the size of the molecule. In other words, FIDS arising from large molecules are damped much faster than those that arise from small molecules. Fast damping produces broad spectral lines with low intensity and this is the origin of NMR's size limit. Two processes contribute to the observed damping:

1. **Chemical shift anisotropy (CSA)**: CSA describes the anisotropic nature of the local magnetic field surrounding a particular nucleus. It is an orientation dependent effect on the chemical shift. The CSA contribution has a quadratic dependence on the static magnetic field strength which renders it a significant source of relaxation at high field.

2. **Dipole-Dipole interactions (DD)**: DD interactions derive from the dipolar fields that originate from surrounding nuclei. The most significant contribution usually comes from directly bonded nuclei. The efficiency of dipolar relaxation depends on the distance between the inter-atomic distance, r, and scales as $1/r^6$.

The most detrimental DD interactions typically involve directly attached or remote ^{1}H nuclei. These interactions can be greatly diminished by elimination of non-exchangeable ^{1}H nuclei via perdeuteration as described in the main text. However, quite often perdeuteration is not enough to enable NMR-based characterization of high molecular weight proteins.

The principle of **T**ranverse **R**elaxation **O**ptimized **S**pectroscop**Y** (TROSY) is the attenuation of transverse relaxation by cancellation of CSA and DD mechanisms [126]. Consider a ^{1}H-^{15}N bond vector which is a commonly used probe in biomolecular NMR studies. In the absence of decoupling, an HSQC spectrum of this ^{1}H-^{15}N bond vector will yield a multiplet consisting of four components. Each component arises from differential interference between CSA and DD mechanisms. The conventional HSQC collapses the multiplet which essentially averages the relaxation rates of each component. For small molecules, this is often not an issue as all components of the multiplet relax with similar rates. However, as molecular weight increases the, differences in relaxation rate for each component increase which leads to severely degraded observable signals. TROSY selects for the slowest relaxing component in which CSA and DD mechanisms cancel. Despite essentially selecting only a subset of the available signal, TROSY often provides increases in S/N for large proteins since all fast relaxing components are eliminated. The theoretical maximal ^{1}H-^{15}N TROSY effect is observed for a perdeuterated protein at ~1.1 GHz ^{1}H Larmor frequency.

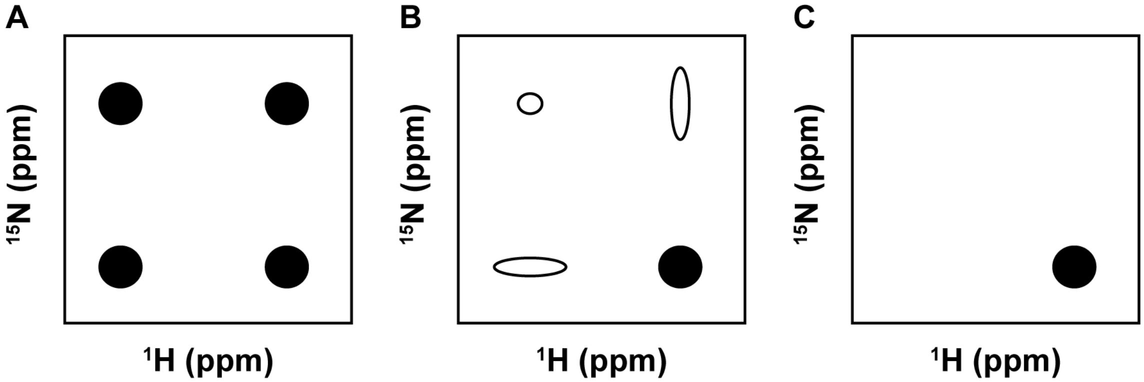


Figure A1.1 Illustration of TROSY-selection. (A) Typical ^{1}H-^{15}N multiplet structure for a fast-tumbling protein. All components relax relatively slowly. (B) Typical ^{1}H-^{15}N multiplet structure for a slow-tumbling protein. Different components relax with different rates depending on the relative contribution of DD interactions and CSA. (C) TROSY-selection of the spectrum shown in B. Only the sharpest peak is retained through the experiment (not mixed with other components) and eventually detected.

A1.3 The relationship between fast timescale dynamics and conformational entropy

The longitudinal and transverse relaxation rates introduced in A1.1 can also report on the motional properties of bond vectors within proteins. These motions occur on the ps-ns timescale and encompass librational motions, rotations, and fluctuations of the bond vectors. Both an amplitude (quantified by the order parameter) and timescale (quantified by τ_e) of motion for a specific bond vector can be obtained from relaxation experiments—although the nature of the motion may not always be apparent. This is commonly done using the model-free formalism of Lipari and Szabo as described in Chapter 6.

Of primary interest is the order parameter because this is the amplitude term. Formally, the order parameter is related to the asymptote of the autocorrelation function of a bond vector's orientation as a function of time. This metric is on a scale of 0 to 1 where 0 represents complete flexibility and 1 represents complete rigidity [281].

Early work from our lab [112] and others [110, 111] recognized that the order parameter of a given bond vector actually reports on the number of states that bond vector has occupied. For example, bonds with high order parameters only visit a few states whereas bonds with low order parameters visit many states. In general, the backbone has been shown to exhibit largely rigid order parameters, whereas methyl-bearing sidechains exhibit heterogeneous behavior ranging from very rigid to very dynamic [256]. Accordingly, sidechains are thought to be the primary contributor to the conformational entropy of proteins.

The formal link between entropy and order parameters comes from the transformation [109]:

$$
\begin{aligned}
O^2 &= \iint d\Omega_1 d\Omega_2 p_{eq}(\Omega_1) P_2(\cos(\vartheta_{12})) p_{eq}(\Omega_2) \\
A &= -kTlnQ \\
H &= kT^2(\partial \ln Q / \partial T) \\
S &= k \ln Z + kT(\partial \ln Q / \partial T)
\end{aligned}
\tag{A.5-A.8}
$$

Where O^2 is the order parameter, Ω represents the number of states, p_{eq} is the angular partition function derived from some potential energy function, A is the Helmholtz free energy, H is the enthalpy, S is the entropy, Q is the partition function defining the thermodynamic parameters, and P_2 is the second Legendre polynomial, k is the Boltzmann constant and T is the temperature.

The choice of the potential energy function is important and most of the early work from our laboratory has utilized a classical harmonic potential [112]. From this, the relationship between internal protein dynamics and conformational entropy can be written as:

$$S_{conf} = \sum p_i S_i^h - k \sum p_i \ln p_i \tag{A.9}$$

Where the first term represents the model-dependent (i.e. harmonic potential) interpretation of order parameters of methyl-bearing side chains and the second term is the classical entropy arising from conformational heterogeneity that can occur over a range of time scales. Studies of calmodulin (CaM) interacting with a variety of peptides showed that the change in conformational entropy upon binding as measured by NMR was linearly related to the total change in entropy as measured by calorimetry for a given CaM binding event [115]. This suggests that conformational entropy may play a significant role in defining the thermodynamics of protein-protein (peptide) interactions.

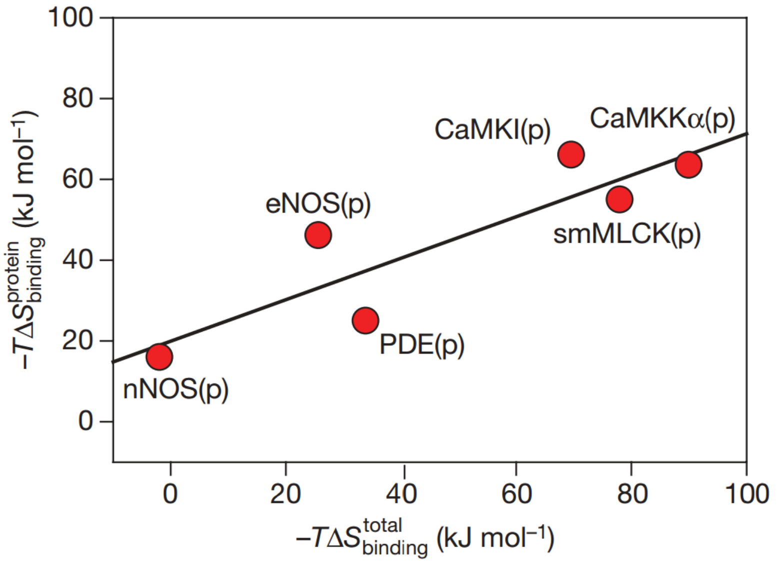


Figure A1.2 The first demonstration that changes in conformational entropy as measured by NMR are correlated with changes in total entropy as measured by calorimetry. The protein is CaM and the ligands are peptides (p) from a variety of native partner proteins. The figure is reproduced from Frederick, K. K., *et al.*, 2007 [115].

Unfortunately, the use of a simple harmonic potential requires assumptions which may not hold true for proteins—for example, an absence of correlated motion [116]. In order to overcome this limitation, our laboratory has proposed an “empirical calibration” which directly relates changes in conformational entropy to changes in the measured methyl sidechain order parameters assuming a linear relationship [116]:

$$\Delta S_{conf} = m[(n^{protein} < \Delta O^{2}_{axis} >^{protein} + n^{ligand} < \Delta O^{2}_{axis} >^{ligand})] + \Delta S_{other} \qquad \text{(A.10)}$$

Where m is the slope of the presumed line and n is the number of amino acids. The ΔS_{other} term includes entropy derived from rotational and translational degrees of freedom as well as any other source of entropy. The use of $O^2{}_{axis}$ is specific for methyl groups which exhibit a symmetry axis. Assuming that the total entropy of binding only has the following components:

$$\Delta S_{total} = \Delta S_{solvent} + \Delta S_{conf} + \Delta S_{other} \qquad \text{(A.11)}$$

and substituting equation A.10 for ΔS_{conf} yields the following relation:

$$\Delta S_{tot} - \Delta S_{sol} = m[(n^{protein} < \Delta O^{2}_{axis} >^{protein} + n^{ligand} < \Delta O^{2}_{axis} >^{ligand})] + \Delta S_{other} \qquad \text{(A.12)}$$

In order to verify this relationship, further studies were performed on CaM. These studies found that the linear relationship shown in equation A.12 did indeed hold true, confirming that the conformational entropy is a critical component of the thermodynamics of protein-protein interactions. The empirical calibration essentially provides an entropy meter wherein a given NMR measurement can be related directly to a change in conformational entropy.

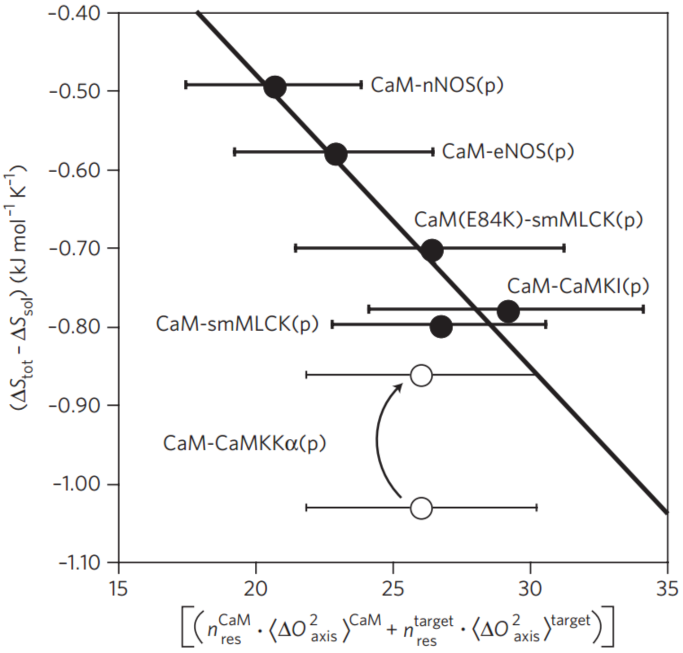


Figure A1.3 Empirical calibration of the entropy meter using CaM and peptides from its native binding partners as a model system. This meter allows changes in NMR observables to be directly related to changes in protein conformational entropy. The figure is reproduced from Marlow, M. *et al.* 2010 [116].

A similar study utilizing a series of mutant variants of Catabolite Activator Protein (CAP) that exhibited different thermodynamic profiles of operator binding also revealed a linear relationship. However, the CAP line had a different slope and intercept. This suggested that perhaps the entropy meter was not universal to all proteins. The utility of such a meter would be quite limited if it had to be calibrated for every individual protein. Follow up work by our lab and the Sharp lab found that the CAP and CaM data exhibit the same slope and intercept if the average change in methyl order parameters is weighted by the sum of the number of torsion angles [328]. This “universal” entropy meter is currently being refined by inclusion of a wide variety of protein interactions (Kasinath V., Harpole K. W., *et al*. in preparation).

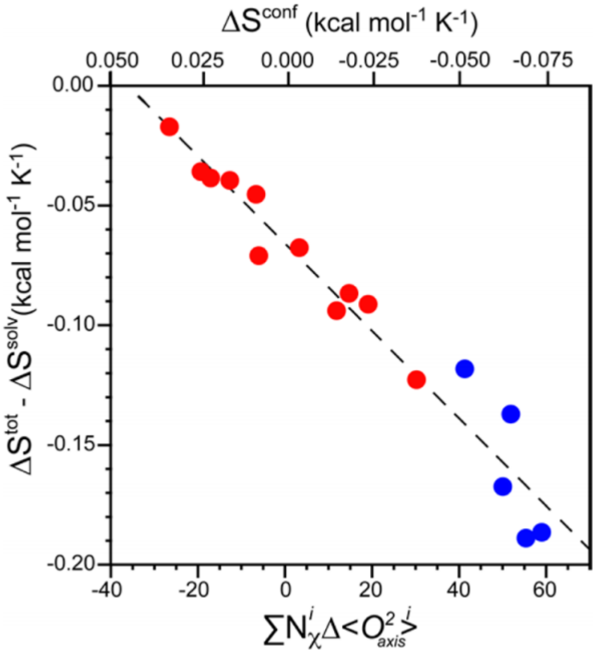


Figure A1.4 Reformulation of the entropy meter to include new points from a study investigating CAP binding to operator DNA [117]. The CAP points are shown in red and the CaM points are shown in blue. The figure is reproduced from Kasinath V., *et al.*, 2013 [328].

APPENDIX 2: List of Primers

A2.1 Primers for amplification of LacI gene and sub-cloning

All primers are written 5’-3’

pASK LacI:

FWD: GGC GCA GAA TTC G GAG AAT CTT TAT TTT CAG GGA ATG AAA CCA GTA ACG

REV: GCT GGA CTC GAG CTA TTA CGC CAG GGT GGT TTT TCT TTT CAC CAG

pASK LacI RD:

FWD: GGC GCA GAA TTC G GAG AAT CTT TAT TTT CAG GGA CAG TCG TTG CTG ATT GGC

REV: GCT GGA CTC GAG CTA TTA CGC CAG GGT GGT TTT TCT TTT CAC CAG

pASK LacI DBD:

FWD: GGC GCA GAA TTC G GAG AAT CTT TAT TTT CAG GGA ATG AAA CCA GTA ACG

REV: GCT GGA CTC GAG CTA TTA CAA CGA CTG TTT GCC

pet15b LacI DBD:

FWD: CAT ATG AAA CCA GTA ACG TTA TAC GAC GTC GCA GAG

REV: GCT GGA CTC GAG CTA TTA CAA CGA CTG TTT GCC

A2.2 Primers for site-directed mutagenesis

All primers are written 5’-3’

LacI K84L:

FWD: CAA ATT GTC GCG GCG ATT CTG TCT CGC GCC GAT CAA CTG

REV:CAG TTG ATC GGC GCG AGA CAG AAT CGC CGC GAC AAT TTG

LacI K84M:

FWD: CAA ATT GTC GCG GCG ATT ATG TCT CGC GCC GAT CAA CTG

REV: CAG TTG ATC GGC GCG AGA CAT AAT CGC CGC GAC AAT TTG

APPENDIX 3: Resonance Assignments

Apo DBD at 25˚C

	HN	N		CM	HM
M1	8.457	121.8	V4	21.46	1.135
K2	8.354	124.1	V4	22.52	1.074
V4	8.333	122.1	L6	25.23	0.763
T5	9.458	117.7	L6	22.55	0.355
L6	9.032	119	V9	22.52	0.579
Y7	7.657	117	V9	22.84	0.728
D8	7.539	120.3	V15	19.62	0.707
V9	7.515	119.7	V15	21.4	0.833
A10	8.039	121.6	V20	23.68	0.804
E11	7.966	116.5	V20	22.73	0.736
Y12	8.174	120.4	V23	24.26	0.993
A13	8.541	117.7	V23	21.62	0.929
G14	7.989	107.2	V24	21.23	0.95
V15	7.83	111.6	V24	22.51	0.977
S16	8.503	115.3	V30	21.41	0.822
Y17	9.19	122.9	V30	21.63	0.88
Q18	8.66	118.4	V38	22.55	0.921
T19	7.945	118.3	V38	24.59	0.9005
V20	7.718	120.4	L45	21.93	0.45
S21	8.203	113.6	L45	26.15	0.486
R22	7.829	120.4	I48	12.48	0.749
V23	7.839	118.8	I48	17.02	0.792
V24	8.116	117.4	V52	21.1	21.1
N25	8.062	115.6	L56	23.43	0.861
Q26	8.19	118.5	L56	24.99	0.916
A27	8.288	121.9	L62	23.7	0.864
S28	8.19	112.4	L62	25.1	0.908
H29	8.612	117.1			

V30	7.712	120
S31	8.786	122.9
A32	8.92	127.2
K33	8.394	116.5
T34	7.593	118.2
R35	8.424	121.9
E36	8.433	117.4
K37	7.733	119.8
V38	8.111	119.9
E39	8.992	119.8
A40	8.35	122.6
A41	7.629	121.4
M42	8.385	115.6
A43	7.811	120.1
E44	8.291	119.4
L45	7.649	115.4
N46	7.788	115.9
Y47	8.443	118.2
I48	7.669	129.1
N50	8.4	118.6
R51	8.282	121.4
V52	8.042	120.9
A53	8.281	126.5
Q54	8.273	119
Q55	8.332	121.1
L56	8.243	123
A57	8.185	124.1
G58	8.282	107.8
K59	8.114	120.7
Q60	8.462	121.5
S61	8.331	117.7

L62	7.974	129.6

DBD-Osym at 35˚C

	HN	N		CM	HM
M1	8.431	121.7	V4	21.03	1.169
K2	8.175	123.5	V4	22.5	1.099
V4	8.183	119.8	L6	24.29	0.6395
T5	9.651	117.9	L6	27.34	0.7322
L6	10.18	121.6	V9	12.84	0.769
Y7	7.695	117.1	V9	23.48	0.6993
D8	7.49	120.2	V15	18.59	0.6776
V9	7.47	119.4	V15	21.33	0.9042
A10	8.205	122	V20	21.51	0.708
E11	7.897	116.9	V20	21.51	0.708
Y12	7.817	120.1	V23	22.01	1.024
A13	8.42	117.2	V23	24.7	1.101
G14	8.072	108.1	V24	21.43	0.934
V15	7.922	110.6	V24	24.14	1.184
S16	8.754	113.3	V30	21.99	0.8123
Y17	8.337	118	V30	21.43	1.007
T19	7.285	120.4	V38	24.55	0.9217
V20	7.148	119.8	V38	23.4	0.9337
S21	7.759	113.1	L45	25.82	0.4515
R22	8.214	118.1	L45	21.91	0.5701
V23	7.879	119.1	I48	12.85	0.7684
V24	8.921	123	I48	16.6	0.7647
N25	7.885	114.1	V52	23.28	0.9047
Q26	7.923	116.7	V52	22.12	0.9533
A27	8.619	120.5	L56	22.89	0.669
S28	8.192	113.4	L56	25.59	0.1343
H29	8.322	113.5	L62	25.36	0.8767

V30	7.65	118.3	L62	23.68	0.8372
S31	9.811	126.3			
A32	9.033	125.9			
K33	8.508	116.5			
T34	7.665	120.1			
R35	9.004	123.1			
E36	8.119	117			
K37	7.695	119.3			
V38	8.402	120.3			
E39	9.146	120.6			
A40	8.112	122			
A41	7.693	121.5			
M42	8.734	115.7			
A43	7.724	119.9			
E44	8.203	119.4			
L45	7.729	114.5			
N46	7.719	117.1			
Y47	8.086	116.3			
I48	7.53	130.4			
N50	9.083	123.1			
R51	8.423	127.4			
V52	7.809	120			
A53	7.452	121.8			
Q54	7.715	113.5			
Q55	8.139	118.3			
L56	8.108	121.7			
A57	6.936	116.8			
G58	7.445	103.1			
K59	8.183	120.7			
Q60	8.436	122.1			
S61	8.295	117.5			

L62	7.86	129.6

Apo RD at 25˚C

	HN	N
L62	8.344	126.1
L63	8.143	127.2
I64	9.151	127.7
G65	8.823	112.6
V66	8.952	125.2
A67	8.607	130.9
S77	9.101	119.8
A81	7.949	122.4
L90	7.434	117.3
G91	7.761	109
A92	8.176	123.5
V94	8.471	122.6
V104	8.664	121.5
E105	8.63	122.2
A106	7.645	123.8
C107	7.547	120.4
K108	8.654	119.3
A109	7.719	120.9
A110	7.228	121.9
V111	7.845	118.6
H112	8.526	119.6
N113	7.928	119.1
Q117	7.403	116.8
R118	7.956	112.3
V119	6.598	109.4
S120	9.514	115.8

RD+IPTG at 25˚C

	HN	N
L62	8.382	125.9
L63	8.199	127
I64	9.168	127.9
G65	8.781	112.8
V66	8.868	125
A67	8.626	130.5
T68	8.457	116.8
S69	8.642	124
I79	8.365	122.7
V80	8.078	118.5
A81	7.821	121.8
A82	7.872	122
K84	8.15	120.2
S85	8.028	112.7
R86	7.289	119.3
A87	8.676	122.3
D88	8.809	118.2
Q89	7.366	118.7
L90	7.398	117.1
G91	7.749	108.7
A92	8.125	123
S93	8.511	117.7
V94	8.319	122.2
V95	8.953	125
G103	8.006	110.7
V104	8.734	121.9

G121	7.359	104.8	E105	8.671	121.9
L122	8.944	124.6	A106	7.597	123.6
I123	8.846	121.2	C107	7.567	120.5
I124	9.086	127.1	K108	8.678	119
L128	7.901	125.6	A109	7.667	120.6
D129	9.206	123.9	A110	7.305	121.6
D130	8.466	120.3	V111	7.902	118.6
Q131	8.615	115.5	H112	8.542	119.2
D132	7.496	122.7	N113	7.953	118.8
A133	8.703	123.9	L114	8.057	120.7
I134	7.234	117.1	L115	8.847	120.9
A135	7.492	123.6	A116	7.692	122.9
V136	8.451	121.1	Q117	7.435	116.4
E137	8.764	122.2	R118	7.974	112.3
A138	8.027	121.9	V119	6.696	109.5
A139	7.525	120.9	S120	9.567	115.8
C140	7.54	116.3	G121	7.356	104
T141	7.597	113.2	L122	8.934	124.3
N142	8.88	119.1	I123	8.774	121
V143	7.733	123.3	I124	9.081	127.2
A145	8.145	123.6	L128	8.068	125.5
L146	8.34	122.8	D129	9.227	123.8
F147	9.014	125.8	D130	8.437	120
V150	6.104	109.8	Q131	8.671	115.2
S151	9.168	114.4	D132	7.513	122.6
D152	9.236	120.8	A133	8.702	123.8
Q153	8.086	116.7	I134	7.191	116.9
T154	7.061	116.9	A135	7.488	123.4
I156	7.758	118.2	V136	8.441	121
N157	9.368	118.6	E137	8.766	122.1
S158	8.284	119	A138	8.046	121.9

I159	7.773	122.2	A139	7.548	120.8
I160	7.866	115.3	C140	7.55	116.2
S162	8.399	111.5	T141	7.612	113.1
H163	7.833	133.5	N142	8.903	119.1
D165	7.191	114.9	V143	7.772	123.4
G166	8.044	106.5	A145	8.049	123.3
T167	8.066	113.3	L146	8.244	121.9
R168	7.396	124.8	F147	8.898	124.5
L169	8.634	116.8	V150	6.135	108.5
V171	8.007	120.5	S151	9.02	113.6
E172	9.361	116.6	D152	9.539	120.4
H173	7.72	119.3	Q153	8.176	116.3
L174	7.203	115	T154	7.011	117.6
A176	8.163	125	I156	7.806	118.7
L177	6.987	117.6	N157	9.295	118.2
G178	7.767	107.3	S158	8.221	119
H179	7.443	119.1	I159	7.705	122.1
I182	9.081	125.3	I160	8.079	116.9
A183	8.324	128.3	F161	8.714	120.4
L184	8.027	118.5	S162	8.503	111.8
L185	8.783	127.6	H163	7.71	134.5
A186	9.018	127.4	E164	8.407	125.8
G187	6.702	103.8	D165	7.13	114.3
S191	8.333	120	G166	8.084	106.7
R195	8.111	120.5	T167	8.016	112.9
L196	8.431	121.7	R168	7.397	124.7
R197	8.09	122.2	L169	8.669	116.5
L198	7.214	118.2	G170	7.329	104.9
A199	8.003	119	V171	7.969	120.5
G200	8.036	109.5	E172	9.368	116.5
W201	8.419	122.6	H173	7.746	119

H202	8.414	116.5	L174	7.224	114.8
K203	8.38	122.5	V175	8.837	120.5
Y204	8.545	117.3	A176	8.175	124.8
L205	9.316	125.7	L177	7.001	117.4
T206	8.46	118.6	G178	7.782	107.2
R207	7.838	122.9	H179	7.472	119
N208	7.359	116.5	I182	9.085	125.1
Q209	7.99	113.5	A183	8.317	128.2
I210	7.926	119.7	L184	8.063	118.3
Q211	8.433	125.9	L185	8.796	127.4
I213	8.211	118.9	A186	9.083	127.5
A214	7.114	119	G187	6.897	104.1
E215	8.495	121.9	L189	8.702	123.8
R216	8.677	124.5	S190	7.457	112.4
E217	8.418	123.3	S191	8.359	120.1
G218	8.394	113.7	V192	8.531	131.7
D219	8.324	117.9	S193	7.022	119.3
W220	7.65	111.1	A194	7.877	124.8
M223	9.316	118.1	R195	8.256	120.2
S224	8.047	116.8	L196	8.442	121.4
G225	7.73	111	R197	7.487	119.3
F226	8.38	125.1	L198	7.142	118.3
Q227	9.046	119	A199	8.121	118.7
Q228	8.567	115.1	G200	8.012	109
T229	7.308	116.6	W201	8.465	122.5
M230	8.299	121	H202	8.335	116.2
Q231	7.629	118	K203	8.337	122.4
M232	7.293	119.7	Y204	8.677	117.3
L233	8.164	118	L205	9.405	125.5
E235	7.662	119.9	T206	8.391	118.4
I237	7.427	124.5	R207	7.842	122.7

V238	7.903	125.7	N208	7.375	116.2
T240	8.746	105.4	Q209	8.006	113.6
A241	7.664	122.3	I210	7.924	119.6
M242	8.427	117.4	Q211	8.467	125.8
L243	7.629	124	I213	8.232	118.6
V244	8.492	121.6	A214	7.14	118.9
A245	8.231	123.6	E215	8.512	121.8
N246	7.679	112.4	R216	8.669	124.5
G252	7.343	104.5	E217	8.434	123.5
A253	8.534	125.4	G218	8.444	114.1
M254	8.957	117.1	D219	8.28	117.4
I257	8.547	120.3	W220	7.654	110.4
T258	7.828	114.9	S221	8.653	113.6
E259	9.068	123	A222	9.097	126
G261	7.863	108.5	M223	9.224	118.2
L262	7.415	121.2	S224	7.878	116.1
R263	10	120	G225	7.702	110.3
V264	9.112	129.2	F226	8.278	124.8
G265	8.648	118.9	Q227	118.7	9.051
D267	6.586	122.1	Q228	8.538	114.8
I268	7.098	115.7	T229	7.274	116.3
S269	9.024	127.4	M230	8.321	120.8
V270	7.986	121.5	Q231	7.625	117.8
V271	8.02	127.9	M232	7.272	119.4
G272	8.786	113.1	L233	8.166	117.9
Y273	7.786	125.6	N234	8.635	120.1
C281	8.349	118	E235	7.668	119.8
Y282	6.267	116.5	G236	7.917	107.7
I283	7.091	113.4	I237	7.423	124.3
L286	6.687	124.6	V238	7.928	125.6
T287	8.964	129.1	T240	8.732	105.4

T288	8.477	123.5	A241	7.671	122
I289	6.997	125.4	M242	8.507	117.4
K290	8.959	127.9	L243	7.678	124.1
D292	8.316	123.4	V244	8.542	121.7
F293	7.6	128.3	A245	8.486	124
V301	7.135	123.5	N246	7.551	110.5
L304	8.215	122	D247	7.354	118.4
L305	8.213	119.6	Q248	7.042	118.3
Q306	7.659	119.6	M249	9.084	116
L307	8.579	121.7	A250	7.588	121.7
S308	8.472	116.2	L251	7.949	120.9
Q309	7.495	119.2	G252	7.315	103.6
Q311	7.939	119.2	A253	8.558	125.1
A312	8.058	125.6	M254	8.862	116.7
V313	7.857	120.6	A256	8.262	123.3
K314	8.286	125.9	I257	8.488	120
G315	8.465	108.4	T258	7.813	114.3
N316	8.813	122.3	E259	9.048	123.2
Q317	7.914	124.7	S260	7.574	115.6
L319	8.912	125.3	G261	7.883	108.3
V321	7.896	107.9	L262	7.416	121.1
L323	8.679	124.2	R263	10.03	119.9
V324	9.145	131.9	V264	9.099	128.9
K325	8.404	127.3	G265	8.663	118.7
R326	8.675	128.7	D267	6.591	121.3
K327	7.81	120.2	I268	7.132	115.6
T328	7.434	106.5	S269	9.007	127.3
T329	7.59	110.7	V270	8.147	121.7
L330	7.664	122.8	V271	8.012	127.6
A331	7.594	129.6	G272	8.72	112.5
			Y273	7.759	125.5

D274	9.341	118.4
S280	7.453	115.2
C281	8.397	117.7
Y282	6.396	116.3
I283	7.132	113.3
L286	6.561	124
T287	9.044	129.2
T288	8.42	122.5
I289	6.862	125.7
K290	9.051	128.9
Q291	7.36	127.7
D292	8.475	123
F293	7.763	129.4
R294	7.904	119.8
L295	7.654	123.8
L296	8.303	122.1
G297	8.846	108.4
Q298	8.076	120.2
T299	8.51	112.1
S300	8.236	117.7
V301	7.346	123.4
D302	7.986	119.2
R303	8.287	119.9
L304	8.241	121.8
L305	8.159	119.3
Q306	7.695	119.7
L307	8.656	121.7
S308	8.464	116.2
Q309	7.508	119.2
G310	7.857	108.2
Q311	7.94	119.3

A312	8.118	125.7
V313	7.816	121.1
K314	8.306	125.9
G315	8.456	108.2
N316	8.852	122.2
Q317	7.917	124.4
L318	8.343	124.1
L319	9.115	125.4
V321	7.93	107.9
S322	7.314	110.3
L323	8.705	123.8
V324	9.114	130.9
K325	8.464	127.7
R326	8.642	128.9
K327	7.891	120.1
T328	7.418	105.9
T329	7.735	111.1
L330	7.694	122.3
A331	7.584	129.4

BIBLIOGRAPHY

[1] Monod J. Recherches sur la croissance des cultures bacteriennes. Theses Doctorat es Sciences. 1942.

[2] Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. 1961;3:318-56.

[3] Lewis M. The lac repressor. C R Biol. 2005;328:521-48.

[4] Matthews KS, Nichols JC. Lactose repressor protein: functional properties and structure. Prog. Nucleic Acid Res. Mol. Biol. 1998;58:127-64.

[5] Zhang Y, Feng Y, Chatterjee S, Tuske S, Ho MX, Arnold E, et al. Structural basis of transcription initiation. Science. 2012;338:1076-80.

[6] Kolb A, Busby S, Buc H, Garges S, Adhya S. Transcriptional regulation by cAMP and its receptor protein. Annu. Rev. Biochem. 1993;62:749-95.

[7] Lawson CL, Swigon D, Murakami KS, Darst SA, Berman HM, Ebright RH. Catabolite activator protein: DNA binding and transcription activation. Curr. Opin. Struct. Biol. 2004;14:10-20.

[8] Schmitz A, Galas DJ. The interaction of RNA polymerase and lac repressor with the lac control region. Nucleic Acids Res. 1979;6:111-37.

[9] Jobe A, Bourgeois S. lac Repressor-operator interaction. VI. The natural inducer of the lac operon. J. Mol. Biol. 1972;69:397-408.

[10] Lin S, Riggs AD. The general affinity of lac repressor for E. coli DNA: implications for gene regulation in procaryotes and eucaryotes. Cell. 1975;4:107-11.

[11] Studier FW, Moffatt BA. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. 1986;189:113-30.

[12] Atkinson MR, Savageau MA, Myers JT, Ninfa AJ. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in Escherichia coli. Cell. 2003;113:597-607.

[13] Janicki SM, Tsukamoto T, Salghetti SE, Tansey WP, Sachidanandam R, Prasanth KV, et al. From silencing to gene expression: real-time analysis in single cells. Cell. 2004;116:683-98.

[14] Bassett EA, DeNizio J, Barnhart-Dailey MC, Panchenko T, Sekulic N, Rogers DJ, et al. HJURP uses distinct CENP-A surfaces to recognize and to stabilize CENP-A/histone H4 for centromere assembly. Dev. Cell. 2012;22:749-62.

[15] Dekel E, Alon U. Optimality and evolutionary tuning of the expression level of a protein. Nature. 2005;436:588-92.

[16] Eames M, Kortemme T. Cost-Benefit Tradeoffs in Engineered lac Operons. Science. 2012;336:911-5.

[17] Cronin CA, Gluba W, Scrable H. The lac operator-repressor system is functional in the mouse. Genes Dev. 2001;15:1506-17.

[18] Sochor MA, Vasireddy V, Drivas TG, Wojno A, Doung T, Shpylchak I, et al. An Autogenously Regulated Expression System for Gene Therapeutic Ocular Applications. Sci. Rep. 2015;5:17105.

[19] Gilbert W, Muller-Hill B. Isolation of the lac repressor. Proc. Natl. Acad. Sci. U. S. A. 1966;56:1891-8.

[20] Farabaugh PJ. Sequence of the lacI gene. Nature. 1978;274:765-9.

[21] Riggs AD, Bourgeois S. On the assay, isolation and characterization of the lac repressor. J. Mol. Biol. 1968;34:361-4.

[22] Bourgeois S, Riggs AD. The lac repressor-operator interaction. IV. Assay and purification of operator DNA. Biochem. Biophys. Res. Commun. 1970;38:348-54.

[23] Gilbert W, Maxam A. The nucleotide sequence of the lac operator. Proc. Natl. Acad. Sci. U. S. A. 1973;70:3581-4.

[24] Reznikoff WS, Winter RB, Hurley CK. The location of the repressor binding sites in the lac operon. Proc. Natl. Acad. Sci. U. S. A. 1974;71:2314-8.

[25] Gilbert W, Majors JM, A. Dahlem Workshop of Chromosomes. Berline: Abakon Verlagsgesellschaff; 1976.

[26] Oehler S, Eismann ER, Kramer H, Muller-Hill B. The three operators of the lac operon cooperate in repression. EMBO J. 1990;9:973-9.

[27] Riggs AD, Suzuki H, Bourgeois S. Lac repressor-operator interaction. I. Equilibrium studies. J. Mol. Biol. 1970;48:67-83.

[28] Chen J, Surendran R, Lee JC, Matthews KS. Construction of a dimeric repressor: dissection of subunit interfaces in Lac repressor. Biochemistry. 1994;33:1234-41.

[29] Sochor MA. In vitro transcription accurately predicts lac repressor phenotype in vivo in Escherichia coli. PeerJ. 2014;2:e498.

[30] Winter RB, von Hippel PH. Diffusion-driven mechanisms of protein translocation on nucleic acids. 2. The Escherichia coli repressor--operator interaction: equilibrium measurements. Biochemistry. 1981;20:6948-60.

[31] Romanuka J, Folkers GE, Biris N, Tishchenko E, Wienk H, Bonvin AM, et al. Specificity and affinity of Lac repressor for the auxiliary operators O2 and O3 are explained by the structures of their protein-DNA complexes. J. Mol. Biol. 2009;390:478-89.

[32] Lewis M, Chang G, Horton NC, Kercher MA, Pace HC, Schumacher MA, et al. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. Science. 1996;271:1247-54.

[33] Sadler JR, Sasmor H, Betz JL. A perfectly symmetric lac operator binds the lac repressor very tightly. Proc. Natl. Acad. Sci. U. S. A. 1983;80:6785-9.

[34] Lin SY, Riggs AD. Lac repressor binding to DNA not containing the lac operator and to synthetic poly dAT. Nature. 1970;228:1184-6.

[35] Dunaway M, Matthews KS. Hybrid tetramers of native and core lactose repressor protein. Assessment of operator and nonspecific DNA binding parameters and their relationship. J. Biol. Chem. 1980;255:10120-7.

[36] Lin SY, Riggs AD. Lac repressor binding to non-operator DNA: detailed studies and a comparison of eequilibrium and rate competition methods. J. Mol. Biol. 1972;72:671-90.

[37] Hammar P, Leroy P, Mahmutovic A, Marklund EG, Berg OG, Elf J. The lac repressor displays facilitated diffusion in living cells. Science. 2012;336:1595-8.

[38] Winter RB, Berg OG, von Hippel PH. Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The Escherichia coli lac repressor--operator interaction: kinetic measurements and conclusions. Biochemistry. 1981;20:6961-77.

[39] Ruusala T, Crothers DM. Sliding and intermolecular transfer of the lac repressor: kinetic perturbation of a reaction intermediate by a distant DNA sequence. Proc. Natl. Acad. Sci. U. S. A. 1992;89:4903-7.

[40] Monod J, Cohen-Bazire G, Cohn M. [The biosynthesis of beta-galactosidase (lactase) in Escherichia coli; the specificity of induction]. Biochim. Biophys. Acta. 1951;7:585-99.

[41] Wilson CJ, Zhan H, Swint-Kruse L, Matthews KS. Ligand interactions with lactose repressor protein and the repressor-operator complex: the effects of ionization and oligomerization on binding. Biophys. Chem. 2007;126:94-105.

[42] Barkley MD, Riggs AD, Jobe A, Burgeois S. Interaction of effecting ligands with lac repressor and repressor-operator complex. Biochemistry. 1975;14:1700-12.

[43] Falcon CM, Matthews KS. Operator DNA sequence variation enhances high affinity binding by hinge helix mutants of lactose repressor protein. Biochemistry. 2000;39:11074-83.

[44] O'Gorman RB, Rosenberg JM, Kallai OB, Dickerson RE, Itakura K, Riggs AD, et al. Equilibrium binding of inducer to lac repressor.operator DNA complex. J. Biol. Chem. 1980;255:10107-14.

[45] Dunaway M, Olson JS, Rosenberg JM, Kallai OB, Dickerson RE, Matthews KS. Kinetic studies of inducer binding to lac repressor.operator complex. J. Biol. Chem. 1980;255:10115-9.

[46] Mueller-Hill B, Rickenberg HV, Wallenfels K. Specificity of the Induction of the Enzymes of the Lac Operon in Escherichia Coli. J. Mol. Biol. 1964;10:303-18.

[47] Eismann ER, Muller-Hill B. lac repressor forms stable loops in vitro with supercoiled wild-type lac DNA containing all three natural lac operators. J. Mol. Biol. 1990;213:763-75.

[48] Platt T, Files JG, Weber K. Lac repressor. Specific proteolytic destruction of the NH 2 - terminal region and loss of the deoxyribonucleic acid-binding activity. J. Biol. Chem. 1973;248:110-21.

[49] Geisler N, Weber K. Escherichia coli lactose repressor: isolation of two different homogeneous headpieces and the existence of a hinge region between residues 50 and 60 in the repressor molecule. FEBS Lett. 1978;87:215-8.

[50] Files JG, Weber K. Limited proteolytic digestion of lac repressor by trypsin. Chemical nature of the resulting trypsin-resistant core. J. Biol. Chem. 1976;251:3386-91.

[51] Friedman BE, Matthews KS. Inducer binding to lac repressor: effects of poly[d(A-T)] and trypsin digestion. Biochem. Biophys. Res. Commun. 1978;85:497-504.

[52] Coulondre C, Miller JH. Genetic studies of the lac repressor. III. Additional correlation of mutational sites with specific amino acid residues. J. Mol. Biol. 1977;117:525-67.

[53] Coulondre C, Miller JH. Genetic studies of the lac repressor. IV. Mutagenic specificity in the lacI gene of Escherichia coli. J. Mol. Biol. 1977;117:577-606.

[54] Schmeissner U, Ganem D, Miller JH. Genetic studies of the lac repressor. II. Fine structure deletion map of the lacI gene, and its correlation with the physical map. J. Mol. Biol. 1977;109:303-26.

[55] Miller JH, Ganem D, Lu P, Schmitz A. Genetic studies of the lac repressor. I. Correlation of mutational sites with specific amino acid residues: construction of a colinear gene-protein map. J. Mol. Biol. 1977;109:275-98.

[56] Calos MP, Galas D, Miller JH. Genetic studies of the lac repressor. VIII. DNA sequence change resulting from an intragenic duplication. J. Mol. Biol. 1978;126:865-9.

[57] Farabaugh PJ, Schmeissner U, Hofer M, Miller JH. Genetic studies of the lac repressor. VII. On the molecular nature of spontaneous hotspots in the lacI gene of Escherichia coli. J. Mol. Biol. 1978;126:847-57.

[58] Schmitz A, Coulondre C, Miller JH. Genetic studies of the lac repressor. V. Repressors which bind operator more tightly generated by suppression and reversion of nonsense mutations. J. Mol. Biol. 1978;123:431-54.

[59] Sommer H, Schmitz A, Schmeissner U, Miller JH. Genetic studies of the lac repressor. VI. The B116 repressor: an altered lac repressor containing amino acid specified by both the trp and lacI leader regions. J. Mol. Biol. 1978;123:457-69.

[60] Miller JH. Genetic studies of the lac repressor. XI. On aspects of lac repressor structure suggested by genetic experiments. J. Mol. Biol. 1979;131:249-58.

[61] Miller JH, Coulondre C, Hofer M, Schmeissner U, Sommer H, Schmitz A, et al. Genetic studies of the lac repressor. IX. Generation of altered proteins by the suppression of nonsence mutations. J. Mol. Biol. 1979;131:191-222.

[62] Miller JH, Schmeissner U. Genetic studies of the lac repressor. X. Analysis of missense mutations in the lacI gene. J. Mol. Biol. 1979;131:223-48.

[63] Weber K, Platt T, Ganem D, Miller JH. Altered sequences changing the operator-binding properties of the Lac repressor: colinearity of the repressor protein with the i-gene map. Proc. Natl. Acad. Sci. U. S. A. 1972;69:3624-8.

[64] Kleina LG, Miller JH. Genetic studies of the lac repressor. XIII. Extensive amino acid replacements generated by the use of natural and synthetic nonsense suppressors. J. Mol. Biol. 1990;212:295-318.

[65] Schmitz A, Schmeissner U, Miller JH. Mutations affecting the quaternary structure of the lac repressor. J. Biol. Chem. 1976;251:3359-66.

[66] Chen J, Matthews KS. Deletion of lactose repressor carboxyl-terminal domain affects tetramer formation. J. Biol. Chem. 1992;267:13843-50.

[67] Maurizot JC, Charlier M, Helene C. Lac repressor binding to poly (d(A-T)). Conformational changes. Biochem. Biophys. Res. Commun. 1974;60:951-7.

[68] Pilz I, Goral K, Kratky O, Bray RP, Wade-Jardetzky NG, Jardetzky O. Small-angle X-ray studies of the quaternary structure of the lac repressor from Escherichia coli. Biochemistry. 1980;19:4087-90.

[69] Charlier M, Maurizot JC, Zaccai G. Neutron scattering studies of lac repressor. Nature. 1980;286:423-5.

[70] Charlier M, Maurizot JC, Zaccai G. Neutron-scattering studies of lac repressor: a low-resolution model. J. Mol. Biol. 1981;153:177-82.

[71] McKay DB, Pickover CA, Steitz TA. Escherichia coli lac repressor is elongated with its operator DNA binding domains located at both ends. J. Mol. Biol. 1982;156:175-83.

[72] Charlier M, Maurizot JC, Zaccai G. Nonspecific binding of lac repressor to DNA. II. A small-angle neutron-scattering study. Biophys. Chem. 1983;18:313-22.

[73] Culard F, Charlier M, Maurizot JC, Tardieu A. Lac repressor-Lac operator complexes. Solution X-ray scattering and electrophoretic studies. Eur. Biophys. J. 1987;14:169-78.

[74] Friedman AM, Fischmann TO, Steitz TA. Crystal structure of lac repressor core tetramer and its implications for DNA looping. Science. 1995;268:1721-7.

[75] Daber R, Stayrook S, Rosenberg A, Lewis M. Structural analysis of lac repressor bound to allosteric effectors. J. Mol. Biol. 2007;370:609-19.

[76] Zuiderweg ER, Kaptein R, Wuthrich K. Sequence-specific resonance assignments in the 1H nuclear-magnetic-resonance spectrum of the lac repressor DNA-binding domain 1-51 from Escherichia coli by two-dimensional spectroscopy. Eur. J. Biochem. 1983;137:279-92.

[77] Zuiderweg ER, Kaptein R, Wuthrich K. Secondary structure of the lac repressor DNA-binding domain by two-dimensional 1H nuclear magnetic resonance in solution. Proc. Natl. Acad. Sci. U. S. A. 1983;80:5837-41.

[78] Zuiderweg ER, Billeter M, Boelens R, Scheek RM, Wuthrich K, Kaptein R. Spatial arrangement of the three alpha helices in the solution conformation of E. coli lac repressor DNA-binding domain. FEBS Lett. 1984;174:243-7.

[79] Kaptein R, Zuiderweg ER, Scheek RM, Boelens R, van Gunsteren WF. A protein structure from nuclear magnetic resonance data. lac repressor headpiece. J. Mol. Biol. 1985;182:179-82.

[80] Zuiderweg ER, Scheek RM, Boelens R, van Gunsteren WF, Kaptein R. Determination of protein structures from nuclear magnetic resonance data using a restrained molecular dynamics approach: the lac repressor DNA binding domain. Biochimie. 1985;67:707-15.

[81] Zuiderweg ER, Scheek RM, Kaptein R. Two-dimensional 1H-nmr studies on the lac repressor DNA binding domain: further resonance assignments and identification of nuclear Overhauser enhancements. Biopolymers. 1985;24:2257-77.

[82] Spronk CA, Slijper M, van Boom JH, Kaptein R, Boelens R. Formation of the hinge helix in the lac repressor is induced upon binding to the lac operator. Nat. Struct. Biol. 1996;3:916-9.

[83] Spronk CA, Bonvin AM, Radha PK, Melacini G, Boelens R, Kaptein R. The solution structure of Lac repressor headpiece 62 complexed to a symmetrical lac operator. Structure. 1999;7:1483-92.

[84] Lehming N, Sartorius J, Kisters-Woike B, von Wilcken-Bergmann B, Muller-Hill B. Mutant lac repressors with new specificities hint at rules for protein--DNA recognition. EMBO J. 1990;9:615-21.

[85] Milk L, Daber R, Lewis M. Functional rules for lac repressor-operator associations and implications for protein-DNA interactions. Protein Sci. 2010;19:1162-72.

[86] Bell CE, Lewis M. A closer view of the conformation of the Lac repressor bound to operator. Nat. Struct. Biol. 2000;7:209-14.

[87] Kalodimos CG, Folkers GE, Boelens R, Kaptein R. Strong DNA binding by covalently linked dimeric Lac headpiece: evidence for the crucial role of the hinge helices. Proc. Natl. Acad. Sci. U. S. A. 2001;98:6039-44.

[88] Spronk CA, Folkers GE, Noordman AM, Wechselberger R, van den Brink N, Boelens R, et al. Hinge-helix formation and DNA bending in various lac repressor-operator complexes. EMBO J. 1999;18:6472-80.

[89] Bell CE, Lewis M. Crystallographic analysis of Lac repressor bound to natural operator O1. J. Mol. Biol. 2001;312:921-6.

[90] Kalodimos CG, Biris N, Bonvin AM, Levandoski MM, Guennuegues M, Boelens R, et al. Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. Science. 2004;305:386-9.

[91] Kalodimos CG, Bonvin AM, Salinas RK, Wechselberger R, Boelens R, Kaptein R. Plasticity in protein-DNA recognition: lac repressor interacts with its natural operator 01 through alternative conformations of its DNA-binding domain. EMBO J. 2002;21:2866-76.

[92] Falcon CM, Swint-Kruse L, Matthews KS. Designed disulfide between N-terminal domains of lactose repressor disrupts allosteric linkage. J. Biol. Chem. 1997;272:26818-21.

[93] Falcon CM, Matthews KS. Engineered disulfide linking the hinge regions within lactose repressor dimer increases operator affinity, decreases sequence selectivity, and alters allostery. Biochemistry. 2001;40:15650-9.

[94] Flynn TC, Swint-Kruse L, Kong Y, Booth C, Matthews KS, Ma J. Allosteric transition pathways in the lactose repressor protein core domains: asymmetric motions in a homodimer. Protein Sci. 2003;12:2523-41.

[95] Taraban M, Zhan H, Whitten AE, Langley DB, Matthews KS, Swint-Kruse L, et al. Ligand-induced conformational changes and conformational dynamics in the solution structure of the lactose repressor protein. J. Mol. Biol. 2008;376:466-81.

[96] Koshland DE, Jr., Nemethy G, Filmer D. Comparison of experimental binding data and theoretical models in proteins containing subunits. Biochemistry. 1966;5:365-85.

[97] Daber R, Sharp K, Lewis M. One is not enough. J. Mol. Biol. 2009;392:1133-44.

[98] Lewis M. Allostery and the lac Operon. J. Mol. Biol. 2013;425:2309-16.

[99] Sharp KA. Allostery in the lac operon: population selection or induced dissociation? Biophys. Chem. 2011;159:66-72.

[100] Hammes GG, Chang YC, Oas TG. Conformational selection or induced fit: a flux description of reaction mechanism. Proc. Natl. Acad. Sci. U. S. A. 2009;106:13737-41.

[101] Nussinov R, Tsai CJ. Allostery without a conformational change? Revisiting the paradigm. Curr. Opin. Struct. Biol. 2015;30:17-24.

[102] Cooper A. Thermodynamic fluctuations in protein molecules. Proc. Natl. Acad. Sci. U. S. A. 1976;73:2740-1.

[103] Weber G. Ligand binding and internal equilibria in proteins. Biochemistry. 1972;11:864-78.

[104] Kleckner IR, Foster MP. An introduction to NMR-based approaches for measuring protein dynamics. Biochim. Biophys. Acta. 2011;1814:942-68.

[105] Volkman BF, Lipson D, Wemmer DE, Kern D. Two-state allosteric behavior in a single-domain signaling protein. Science. 2001;291:2429-33.

[106] Kern D, Zuiderweg ER. The role of dynamics in allosteric regulation. Curr. Opin. Struct. Biol. 2003;13:748-57.

[107] Lipchock JM, Loria JP. Nanometer propagation of millisecond motions in V-type allostery. Structure. 2010;18:1596-607.

[108] Lisi GP, Loria JP. Solution NMR Spectroscopy for the Study of Enzyme Allostery. Chem. Rev. 2016.

[109] Wand AJ. Dynamic activation of protein function: a view emerging from NMR spectroscopy. Nat. Struct. Biol. 2001;8:926-31.

[110] Akke M, Bruschweiler R, Palmer AG. Nmr Order Parameters and Free-Energy - an Analytical Approach and Its Application to Cooperative Ca2+ Binding by Calbindin-D(9k). J. Am. Chem. Soc. 1993;115:9832-3.

[111] Yang D, Kay LE. Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: application to protein folding. J. Mol. Biol. 1996;263:369-82.

[112] Li Z, Raychaudhuri S, Wand AJ. Insights into the local residual entropy of proteins provided by NMR relaxation. Protein Sci. 1996;5:2647-50.

[113] Lee AL, Kinnear SA, Wand AJ. Redistribution and loss of side chain entropy upon formation of a calmodulin-peptide complex. Nat. Struct. Biol. 2000;7:72-7.

[114] Lee AL, Wand AJ. Microscopic origins of entropy, heat capacity and the glass transition in proteins. Nature. 2001;411:501-4.

[115] Frederick KK, Marlow MS, Valentine KG, Wand AJ. Conformational entropy in molecular recognition by proteins. Nature. 2007;448:325-9.

[116] Marlow MS, Dogan J, Frederick KK, Valentine KG, Wand AJ. The role of conformational entropy in molecular recognition by calmodulin. Nat. Chem. Biol. 2010;6:352-8.

[117] Tzeng SR, Kalodimos CG. Protein activity regulation by conformational entropy. Nature. 2012;488:236-40.

[118] Cooper A, Dryden DT. Allostery without conformational change. A plausible model. Eur. Biophys. J. 1984;11:103-9.

[119] Popovych N, Sun S, Ebright RH, Kalodimos CG. Dynamically driven protein allostery. Nat. Struct. Mol. Biol. 2006;13:831-8.

[120] Tzeng SR, Kalodimos CG. Dynamic activation of an allosteric regulatory protein. Nature. 2009;462:368-72.

[121] Petit CM, Zhang J, Sapienza PJ, Fuentes EJ, Lee AL. Hidden dynamic allostery in a PDZ domain. Proc. Natl. Acad. Sci. U. S. A. 2009;106:18249-54.

[122] Yonetani T, Laberge M. Protein dynamics explain the allosteric behaviors of hemoglobin. Biochim. Biophys. Acta. 2008;1784:1146-58.

[123] Motlagh HN, Wrabl JO, Li J, Hilser VJ. The ensemble nature of allostery. Nature. 2014;508:331-9.

[124] Slijper M, Boelens R, Davis AL, Konings RN, van der Marel GA, van Boom JH, et al. Backbone and side chain dynamics of lac repressor headpiece (1-56) and its complex with DNA. Biochemistry. 1997;36:249-54.

[125] Sattler M, Schleucher J, Griesinger C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Progress in Nuclear Magnetic Resonance Spectroscopy. 1999;34:93-158.

[126] Pervushin K, Riek R, Wider G, Wuthrich K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proc. Natl. Acad. Sci. U. S. A. 1997;94:12366-71.

[127] Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE. Cross-correlated relaxation enhanced 1H[bond]13C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. J. Am. Chem. Soc. 2003;125:10420-8.

[128] Sprangers R, Kay LE. Quantitative dynamics and binding studies of the 20S proteasome by NMR. Nature. 2007;445:618-22.

[129] Ruschak AM, Kay LE. Proteasome allostery as a population shift between interchanging conformers. Proc. Natl. Acad. Sci. U. S. A. 2012;109:E3454-62.

[130] Zuiderweg ER, Bagai I, Rossi P, Bertelsen EB. EZ-ASSIGN, a program for exhaustive NMR chemical shift assignments of large proteins from complete or incomplete triple-resonance data. J. Biomol. NMR. 2013;57:179-91.

[131] Teng Q. Structural Biology: Practical NMR Applications. 1 ed. New York, New York, USA: Springer; 2010.

[132] Li Y, Logan TM, Edison AS, Webb A. Design of small volume HX and triple-resonance probes for improved limits of detection in protein NMR experiments. J. Magn. Reson. 2003;164:128-35.

[133] Peti W, Norcross J, Eldridge G, O'Neil-Johnson M. Biomolecular NMR using a microcoil NMR probe - New technique for the chemical shift assignment of aromatic side chains in proteins. J. Am. Chem. Soc. 2004;126:5873-8.

[134] Rodriguez E, Krishna NR. An economical method for (15)N/(13)C isotopic labeling of proteins expressed in Pichia pastoris. J. Biochem. 2001;130:19-22.

[135] Sugiki T, Ichikawa O, Miyazawa-Onami M, Shimada I, Takahashi H. Isotopic labeling of heterologous proteins in the yeast Pichia pastoris and Kluyveromyces lactis. Methods Mol. Biol. 2012;831:19-36.

[136] Wood MJ, Komives EA. Production of large quantities of isotopically labeled protein in Pichia pastoris by fermentation. J. Biomol. NMR. 1999;13:149-59.

[137] Clark L, Zahm JA, Ali R, Kukula M, Bian L, Patrie SM, et al. Methyl labeling and TROSY NMR spectroscopy of proteins expressed in the eukaryote Pichia pastoris. J. Biomol. NMR. 2015;62:239-45.

[138] Kofuku Y, Ueda T, Okude J, Shiraishi Y, Kondo K, Mizumura T, et al. Functional dynamics of deuterated beta2 -adrenergic receptor in lipid bilayers revealed by NMR spectroscopy. Angew. Chem. Int. Ed. Engl. 2014;53:13376-9.

[139] Romanuka J, van den Bulke H, Kaptein R, Boelens R, Folkers GE. Novel strategies to overcome expression problems encountered with toxic proteins: Application to the production of Lac repressor proteins for NMR studies. Protein Expr. Purif. 2009;67:104-12.

[140] Studier FW, Moffatt BA. Use of bacteriophage-T7 RNA-polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. 1986;189:113-30.

[141] Zhan H, Sun Z, Matthews KS. Functional impact of polar and acidic substitutions in the Lactose repressor hydrophobic monomer.monomer interface with a buried lysine. Biochemistry. 2009;48:1305-14.

[142] Guzman LM, Belin D, Carson MJ, Beckwith J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose p-BAD promoter. J. Bacteriol. 1995;177:4121-30.

[143] Giacalone MJ, Gentile AM, Lovitt BT, Berkley NL, Gunderson CW, Surber MW. Toxic protein expression in Escherichia coli using a rhamnose-based tightly regulated and tunable promoter system. Biotechniques. 2006;40:355-64.

[144] Khoury AM, Nick HS, Lu P. In vivo interaction of Escherichia coli lac repressor N-terminal fragments with the lac operator. J. Mol. Biol. 1991;219:623-34.

[145] Skerra A. Use of the tetracycline promoter for the tightly regulated production of a murine antibody fragment in Escherichia coli. Gene. 1994;151:131-5.

[146] Wycuff DR, Matthews KS. Generation of an AraC-araBAD promoter-regulated T7 expression system. Anal. Biochem. 2000;277:67-73.

[147] Macconkey A. Lactose-Fermenting Bacteria in Faeces. J. Hyg. (Lond.). 1905;5:333-79.

[148] Tugarinov V, Kanelis V, Kay LE. Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. Nat. Protoc. 2006;1:749-54.

[149] Tropea JE, Cherry S, Waugh DS. Expression and purification of soluble His(6)-tagged TEV protease. Methods in molecular biology (Clifton, N.J.). 2009;498:297-307.

[150] Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nature Methods. 2012;9:671-5.

[151] Hwang TL, Shaka AJ. Water Suppression That Works - Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. Journal of Magnetic Resonance Series A. 1995;112:275-9.

[152] Stott K, Stonehouse J, Keeler J, Hwang TL, Shaka AJ. Excitation Sculpting in High-Resolution Nuclear-Magnetic-Resonance Spectroscopy - Application to Selective Noe Experiments. J. Am. Chem. Soc. 1995;117:4199-200.

[153] Styles P, Soffe NF, Scott CA, Cragg DA, Row F, White DJ, et al. A High-Resolution Nmr Probe in Which the Coil and Preamplifier Are Cooled with Liquid-Helium. J. Magn. Reson. 1984;60:397-404.

[154] Kovacs H, Moskau D, Spraul M. Cryogenically cooled probes - a leap in NMR technology. Progress in Nuclear Magnetic Resonance Spectroscopy. 2005;46:131-55.

[155] Flynn PF, Mattiello DL, Hill HDW, Wand AJ. Optimal use of cryogenic probe technology in NMR studies of proteins. J. Am. Chem. Soc. 2000;122:4823-4.

[156] Pervushin K, Riek R, Wider G, Wuthrich K. Attenuated T-2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proc. Natl Acad. Sci. USA. 1997;94:12366-71.

[157] Bodenhausen G, Ruben DJ. Natural Abundance N-15 Nmr by Enhanced Heteronuclear Spectroscopy. Chem. Phys. Lett. 1980;69:185-9.

[158] Mandal PK, Majumdar A. A comprehensive discussion of HSQC and HMQC pulse sequences. Concepts in Magnetic Resonance Part A. 2004;20a:1-23.

[159] Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPIPE - a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR. 1995;6:277-93.

[160] Gilbert W, Mullerhi.B. Isolation of Lac repressor. Proc. Natl. Acad. Sci. U. S. A. 1966;56:1891-&.

[161] Mullerhi.B, Crapo L, Gilbert W. Mutants that make more Lac repressor. Proc. Natl. Acad. Sci. U. S. A. 1968;59:1259-&.

[162] Glascock CB, Weickert MJ. Using chromosomal lacI(Q1) to control expression of genes on high-copy-number plasmids in Escherichia coli. Gene. 1998;223:221-31.

[163] Kim S, Brostroemer E, Xing D, Jin J, Chong S, Ge H, et al. Probing allostery through DNA. Science. 2013;339:816-9.

[164] Daber R, Stayrook S, Rosenberg A, Lewis M. Structural analysis of Lac repressor bound to allosteric effectors. J. Mol. Biol. 2007;370:609-19.

[165] Haeusler AR, Goodson KA, Lillian TD, Wang X, Goyal S, Perkins NC, et al. FRET studies of a landscape of Lac repressor-mediated DNA loops. Nucleic Acids Res. 2012;40:4432-45.

[166] Chen J, Surendran R, Lee JC, Matthews KS. Construction of a dimeric repressor - dissection of subunit interfaces in Lac repressor. Biochemistry. 1994;33:1234-41.

[167] Chen J, Matthews KS. Subunit dissociation affects DNA-binding in a dimeric Lac repressor produced by C-terminal deletion. Biochemistry. 1994;33:8728-35.

[168] Boomershine WP, Raj MLS, Gopalan V, Foster MP. Preparation of uniformly labeled NMR samples in Escherichia coli under the tight control of the araBAD promoter: Expression of an archaeal homolog of the RNase P Rpp29 protein. Protein Expr. Purif. 2003;28:246-51.

[169] Owens RM, Grant A, Davies N, O'Connor CD. Copurification of the lac repressor with polyhistidine-tagged proteins in immobilized metal affinity chromatography. Protein Expr. Purif. 2001;21:352-60.

[170] Romanuka J, van den Bulke H, Kaptein R, Boelens R, Folkers GE. Novel strategies to overcome expression problems encountered with toxic proteins: Application to the production of Lac repressor proteins for NMR studies. Protein Express. Purif. 2009;67:104-12.

[171] Nichols JC, Matthews KS. Combinatorial mutations of lac repressor. Stability of monomer-monomer interface is increased by apolar substitution at position 84. J. Biol. Chem. 1997;272:18550-7.

[172] Bell CE, Barry J, Matthews KS, Lewis M. Structure of a variant of Lac repressor with increased thermostability and decreased affinity for operator. J. Mol. Biol. 2001;313:99-109.

[173] Boomershine WP, Raj MLS, Gopalan V, Foster MP. Preparation of uniformly labeled NMR samples in Escherichia coli under the tight control of the araBAD promoter: Expression of an archaeal homolog of the RNase P Rpp29 protein. Protein Express. Purif. . 2003;28:246-51.

[174] Slijper M, Bonvin AM, Boelens R, Kaptein R. Refined structure of lac repressor headpiece (1-56) determined by relaxation matrix calculations from 2D and 3D NOE data: change of tertiary structure upon binding to the lac operator. J. Mol. Biol. 1996;259:761-73.

[175] Lamerichs RM, Boelens R, van der Marel GA, van Boom JH, Kaptein R, Buck F, et al. H NMR study of a complex between the lac repressor headpiece and a 22 base pair symmetric lac operator. Biochemistry. 1989;28:2985-91.

[176] Latham MP, Zimmermann GR, Pardi A. NMR chemical exchange as a probe for ligand-binding kinetics in a theophylline-binding RNA aptamer. J. Am. Chem. Soc. 2009;131:5052-3.

[177] Felitsky DJ, Record MT, Jr. Thermal and urea-induced unfolding of the marginally stable lac repressor DNA-binding domain: a model system for analysis of solute effects on protein processes. Biochemistry. 2003;42:2202-17.

[178] Wang AC, Revzin A, Butler AP, von Hippel PH. Binding of E.coli lac repressor to non-operator DNA. Nucleic Acids Res. 1977;4:1579-93.

[179] Englander SW, Wand AJ. Main-chain-directed strategy for the assignment of 1H NMR spectra of proteins. Biochemistry. 1987;26:5953-8.

[180] Sun ZY, Frueh DP, Selenko P, Hoch JC, Wagner G. Fast assignment of 15N-HSQC peaks using high-resolution 3D HNcocaNH experiments with non-uniform sampling. J. Biomol. NMR. 2005;33:43-50.

[181] Frueh DP. Practical aspects of NMR signal assignment in larger and challenging proteins. Prog Nucl Magn Reson Spectrosc. 2014;78:47-75.

[182] Krishnarjuna B, Jaipuria G, Thakur A, D'Silva P, Atreya HS. Amino acid selective unlabeling for sequence specific resonance assignments in proteins. J. Biomol. NMR. 2011;49:39-51.

[183] Ohki SY, Kainosho M. Stable isotope labeling methods for protein NMR spectroscopy. Progress in Nuclear Magnetic Resonance Spectroscopy. 2008;53:208-26.

[184] Muchmore DC, McIntosh LP, Russell CB, Anderson DE, Dahlquist FW. Expression and nitrogen-15 labeling of proteins for proton and nitrogen-15 nuclear magnetic resonance. Methods Enzymol. 1989;177:44-73.

[185] Trautwein M, Fredriksson K, Moller HM, Exner TE. Automated assignment of NMR chemical shifts based on a known structure and 4D spectra. J. Biomol. NMR. 2016.

[186] Grzesiek S, Bax A. Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. J. Biomol. NMR. 1993;3:185-204.

[187] Salzmann M, Pervushin K, Wider G, Senn H, Wuthrich K. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. Proc. Natl. Acad. Sci. U. S. A. 1998;95:13585-90.

[188] Braunschweiler L, Ernst RR. Coherence Transfer by Isotropic Mixing - Application to Proton Correlation Spectroscopy. J. Magn. Reson. 1983;53:521-8.

[189] Gardner KH, Konrat R, Rosen MK, Kay LE. An (H)C(CO)NH-TOCSY pulse scheme for sequential assignment of protonated methyl groups in otherwise deuterated N-15,C-13-labeled proteins. J. Biomol. NMR. 1996;8:351-6.

[190] Yang D, Zheng Y, Liu D, Wyss DF. Sequence-specific assignments of methyl groups in high-molecular weight proteins. J. Am. Chem. Soc. 2004;126:3710-1.

[191] Tugarinov V, Kay LE. Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. J. Am. Chem. Soc. 2003;125:13868-78.

[192] Tugarinov V, Muhandiram R, Ayed A, Kay LE. Four-dimensional NMR spectroscopy of a 723-residue protein: chemical shift assignments and secondary structure of malate synthase g. J. Am. Chem. Soc. 2002;124:10025-35.

[193] Hoch JC, Stern AS. NMR Data Processing. Hoboken, New Jersey: John Wiley & Sons, INC.; 1996.

[194] Sprangers R, Kay LE. Quantitative dynamics and binding studies of the 20S proteasome by NMR. Nature. 2007;445:618-22.

[195] Gelis I, Bonvin AM, Keramisanou D, Koukaki M, Gouridis G, Karamanou S, et al. Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. Cell. 2007;131:756-69.

[196] Wiegand T, Gardiennet C, Cadalbert R, Lacabanne D, Kunert B, Terradot L, et al. Variability and conservation of structural domains in divide-and-conquer approaches. J. Biomol. NMR. 2016;65:79-86.

[197] Barna JCJ, Laue ED, Mayger MR, Skilling J, Worrall SJP. Exponential sampling, an alternative method for sampling in two-dimensional NMR experiments. J. Magn. Reson. 1987;73:69-77.

[198] Mobli M, Hoch JC. Nonuniform sampling and non-Fourier signal processing methods in multidimensional NMR. Progress in Nuclear Magnetic Resonance Spectroscopy. 2015;86-87:80-.

[199] Smith SW. The Scientist and Engineer's Guide to Digital Signal Processing. San Diego, California: California Technical Publishing; 1997.

[200] Otting G, Widmer H, Wagner G, Wuthrich K. Origin of T1 and T2 Ridges in 2d Nmr-Spectra and Procedures for Suppression. J. Magn. Reson. 1986;66:187-93.

[201] Hyberts SG, Robson SA, Wagner G. Exploring signal-to-noise ratio and sensitivity in non-uniformly sampled multi-dimensional NMR spectra. J. Biomol. NMR. 2013;55:167-78.

[202] Hyberts SG, Takeuchi K, Wagner G. Poisson-Gap sampling and forward maximum entropy reconstruction for enhancing the resolution and sensitivity of protein NMR data. J. Am. Chem. Soc. 2010;132:2145-+.

[203] Drori I. Fast l(1) minimization by iterative thresholding for multidimensional NMR Spectroscopy. Eurasip Journal on Advances in Signal Processing. 2007.

[204] Hyberts SG, Milbradt AG, Wagner AB, Arthanari H, Wagner G. Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson Gap scheduling. J. Biomol. NMR. 2012;52:315-27.

[205] Aoto PC, Fenwick RB, Kroon GJA, Wright PE. Accurate scoring of non-uniform sampling schemes for quantitative NMR. J. Magn. Reson. 2014;246:31-5.

[206] Schuyler AD, Maciejewski MW, Stern AS, Hoch JC. Nonuniform sampling of hypercomplex multidimensional NMR experiments: Dimensionality, quadrature phase and randomization. J. Magn. Reson. 2015;254:121-30.

[207] Cooley JW, Tukey JW. An Algorithm for Machine Calculation of Complex Fourier Series. Mathematics of Computation. 1965;19:297-&.

[208] Stern AS, Donoho DL, Hoch JC. NMR data processing using iterative thresholding and minimum l(1)-norm reconstruction. J. Magn. Reson. 2007;188:295-300.

[209] Kazimierczuk K, Orekhov VY. A comparison of convex and non-convex compressed sensing applied to multidimensional NMR. J. Magn. Reson. 2012;223:1-10.

[210] Candes EJ, Wakin MB, Boyd SP. Enhancing sparsity by reweighted l(1) minimization. Journal of Fourier Analysis and Applications. 2008;14:877-905.

[211] Candes EJ, Romberg JK, Tao T. Stable signal recovery from incomplete and inaccurate measurements. Communications on Pure and Applied Mathematics. 2006;59:1207-23.

[212] Kazimierczuk K, Orekhov VY. Accelerated NMR spectroscopy by using compressed sensing. Angewandte Chemie (International Edition). 2011;50:5556-9.

[213] Findeisen M, Brand T, Berger S. A H-1-NMR thermometer suitable for cryoprobes. Magn. Reson. Chem. 2007;45:175-8.

[214] Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, et al. 1H, 13C and 15N chemical shift referencing in biomolecular NMR. J. Biomol. NMR. 1995;6:135-40.

[215] Krejcirikova A, Tugarinov V. 3D-TROSY-based backbone and ILV-methyl resonance assignments of a 319-residue homodimer from a single protein sample. J. Biomol. NMR. 2012;54:135-43.

[216] Kay LE, Ikura M, Tschudin R, Bax A. 3-Dimensional Triple-Resonance Nmr-Spectroscopy of Isotopically Enriched Proteins. J. Magn. Reson. 1990;89:496-514.

[217] Clubb RT, Thanabal V, Wagner G. A Constant-Time 3-Dimensional Triple-Resonance Pulse Scheme to Correlate Intraresidue H-1(N), N-15, and C-13(') Chemical-Shifts in N-15-C-13-Labeled Proteins. J. Magn. Reson. 1992;97:213-7.

[218] Bax A, Ikura M. An efficient 3D NMR technique for correlating the proton and 15N backbone amide resonances with the alpha-carbon of the preceding residue in uniformly 15N/13C enriched proteins. J. Biomol. NMR. 1991;1:99-104.

[219] Grzesiek S, Bax A. An Efficient Experiment for Sequential Backbone Assignment of Medium-Sized Isotopically Enriched Proteins. J. Magn. Reson. 1992;99:201-7.

[220] Grzesiek S, Bax A. Correlating Backbone Amide and Side-Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance Nmr. J. Am. Chem. Soc. 1992;114:6291-3.

[221] Marion D, Driscoll PC, Kay LE, Wingfield PT, Bax A, Gronenborn AM, et al. Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. Biochemistry. 1989;28:6150-6.

[222] Bax A, Clore GM, Gronenborn AM. H-1-H-1 Correlation Via Isotropic Mixing of C-13 Magnetization, a New 3-Dimensional Approach for Assigning H-1 and C-13 Spectra of C-13-Enriched Proteins. J. Magn. Reson. 1990;88:425-31.

[223] Zimmerman D, Kulikowski C, Wang L, Lyons B, Montelione GT. Automated sequencing of amino acid spin systems in proteins using multidimensional HCC(CO)NH-TOCSY spectroscopy and constraint propagation methods from artificial intelligence. J. Biomol. NMR. 1994;4:241-56.

[224] Goddard TDK, D. G. SPARKY 3. University of California, San Francisco.

[225] Bahrami A, Assadi AH, Markley JL, Eghbalnia HR. Probabilistic interaction network of evidence algorithm and its application to complete labeling of peak lists from protein NMR spectroscopy. PLoS Comput. Biol. 2009;5:e1000307.

[226] Eghbalnia HR, Wang L, Bahrami A, Assadi A, Markley JL. Protein energetic conformational analysis from NMR chemical shifts (PECAN) and its use in determining secondary structural elements. J. Biomol. NMR. 2005;32:71-81.

[227] Shen Y, Bax A. Protein structural information derived from NMR chemical shift with the neural network program TALOS-N. Methods Mol. Biol. 2015;1260:17-32.

[228] Wishart DS, Sykes BD, Richards FM. The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. Biochemistry. 1992;31:1647-51.

[229] Gans P, Hamelin O, Sounier R, Ayala I, Dura MA, Amero CD, et al. Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins. Angewandte Chemie (International Edition). 2010;49:1958-62.

[230] Fischer MWF, Zeng L, Zuiderweg ERP. Use of C-13-C-13 NOE for the assignment of NMR lines of larger labeled proteins at larger magnetic fields. J. Am. Chem. Soc. 1996;118:12457-8.

[231] Elf J, Li GW, Xie XS. Probing transcription factor dynamics at the single-molecule level in a living cell. Science. 2007;316:1191-4.

[232] Choi PJ, Cai L, Frieda K, Xie XS. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. Science. 2008;322:442-6.

[233] Bondos SE, Swint-Kruse L, Matthews KS. Flexibility and Disorder in Gene Regulation: LacI/GalR and Hox Proteins. J. Biol. Chem. 2015;290:24669-77.

[234] Romanuka J. NMR studies of the allosteric effectors of the lac operon: Universiteit Utrecht; 2009.

[235] Evenas J, Tugarinov V, Skrynnikov NR, Goto NK, Muhandiram R, Kay LE. Ligand-induced structural changes to maltodextrin-binding protein as studied by solution NMR spectroscopy. J. Mol. Biol. 2001;309:961-74.

[236] Fersht AR, Petrovich M. Reply to Campos and Munoz: Why phosphate is a bad buffer for guanidinium chloride titrations. Proc. Natl. Acad. Sci. U. S. A. 2013;110:E1244-5.

[237] Greene RF, Jr., Pace CN. Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, alpha-chymotrypsin, and beta-lactoglobulin. J. Biol. Chem. 1974;249:5388-93.

[238] Santoro MM, Bolen DW. Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. Biochemistry. 1988;27:8063-8.

[239] Ramot R, Kishore Inampudi K, Wilson CJ. Lactose repressor experimental folding landscape: fundamental functional unit and tetramer folding mechanisms. Biochemistry. 2012;51:7569-79.

[240] Chen J. Genetic and physical studies of quaternary structure of lactose repressor: Rice University; 1993.

[241] Rapid and efficient purification and refolding of a (His)6-tagged recombinant protein produced in E. coli as inclusion bodies. GE Healthcare Application Note: HiTrap Chelating.

[242] Shen Y, Vernon R, Baker D, Bax A. De novo protein structure generation from incomplete chemical shift assignments. J. Biomol. NMR. 2009;43:63-78.

[243] Shen Y, Lange O, Delaglio F, Rossi P, Aramini JM, Liu G, et al. Consistent blind protein structure generation from NMR chemical shift data. Proc. Natl. Acad. Sci. U. S. A. 2008;105:4685-90.

[244] Shen Y, Bryan PN, He Y, Orban J, Baker D, Bax A. De novo structure generation using chemical shifts for proteins with high-sequence identity but different folds. Protein Sci. 2010;19:349-56.

[245] Huang T, Li J, Byrd RA. Solution structure of lysine-free (K0) ubiquitin. Protein Sci. 2014;23:662-7.

[246] Jankowski W, Saleh T, Pai MT, Sriram G, Birge RB, Kalodimos CG. Domain organization differences explain Bcr-Abl's preference for CrkL over CrkII. Nat. Chem. Biol. 2012;8:590-6.

[247] Swint-Kruse L, Zhan H, Matthews KS. Integrated insights from simulation, experiment, and mutational analysis yield new details of LacI function. Biochemistry. 2005;44:11201-13.

[248] Kaplan JI, Fraenkel G. NMR of Chemically Echanging Systems. New York, New York: Academic Press, INC.; 1980.

[249] Selvaratnam R, Chowdhury S, VanSchouwen B, Melacini G. Mapping allostery through the covariance analysis of NMR chemical shifts. Proc. Natl. Acad. Sci. U. S. A. 2011;108:6133-8.

[250] Falk BT, Sapienza PJ, Lee AL. Chemical shift imprint of intersubunit communication in a symmetric homodimer. Proc. Natl. Acad. Sci. U. S. A. 2016.

[251] Swint-Kruse L, Zhan H, Fairbanks BM, Maheshwari A, Matthews KS. Perturbation from a distance: mutations that alter LacI function through long-range effects. Biochemistry. 2003;42:14004-16.

[252] Gerk LP, Leven O, Muller-Hill B. Strengthening the dimerisation interface of Lac repressor increases its thermostability by 40 deg. C. J. Mol. Biol. 2000;299:805-12.

[253] Bell CE, Barry J, Matthews KS, Lewis M. Structure of a variant of lac repressor with increased thermostability and decreased affinity for operator. J. Mol. Biol. 2001;313:99-109.

[254] Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng. 1995;8:127-34.

[255] Barry JK, Matthews KS. Substitutions at histidine 74 and aspartate 278 alter ligand binding and allostery in lactose repressor protein. Biochemistry. 1999;38:3579-90.

[256] Igumenova TI, Frederick KK, Wand AJ. Characterization of the fast dynamics of protein amino acid side chains using NMR relaxation in solution. Chem. Rev. 2006;106:1672-99.

[257] McElroy C, Manfredo A, Wendt A, Gollnick P, Foster M. TROSY-NMR studies of the 91kDa TRAP protein reveal allosteric control of a gene regulatory protein by ligand-altered flexibility. J. Mol. Biol. 2002;323:463-73.

[258] Srivastava AK, McDonald LR, Cembran A, Kim J, Masterson LR, McClendon CL, et al. Synchronous opening and closing motions are essential for cAMP-dependent protein kinase A signaling. Structure. 2014;22:1735-43.

[259] Loria JP, Rance M, Palmer AG. A TROSY CPMG sequence for characterizing chemical exchange in large proteins. J. Biomol. NMR. 1999;15:151-5.

[260] Wang C, Rance M, Palmer AG, 3rd. Mapping chemical exchange in proteins with MW > 50 kD. J. Am. Chem. Soc. 2003;125:8968-9.

[261] Ihms EC, Zhou M, Zhang Y, Kleckner IR, McElroy CA, Wysocki VH, et al. Gene regulation by substoichiometric heterocomplex formation of undecameric TRAP and trimeric anti-TRAP. Proc. Natl. Acad. Sci. U. S. A. 2014;111:3442-7.

[262] Tugarinov V, Kay LE. Quantitative C-13 and H-2 NMR relaxation studies of the 723-residue enzyme malate synthase g reveal a dynamic binding interface. Biochemistry. 2005;44:15970-7.

[263] Schumacher MA, Choi KY, Lu F, Zalkin H, Brennan RG. Mechanism of corepressor-mediated specific DNA binding by the purine repressor. Cell. 1995;83:147-55.

[264] Mukherjee SP, Borin B, Quintas PO, Dyson HJ. NMR characterization of a 72 kDa transcription factor using differential isotopic labeling. Protein Sci. 2016;25:597-604.

[265] Skelton NJ, Palmer AG, Akke M, Kordel J, Rance M, Chazin WJ. Practical aspects of 2-dimensional proton-detected N-15 spin relaxation measurements. Journal of Magnetic Resonance Series B. 1993;102:253-64.

[266] Kamath U, Shriver JW. Characterization of themotropic state changes in myosin subfragment-1 and heavy-meromyosin by UV difference spectroscopy. J. Biol. Chem. 1989;264:5586-92.

[267] Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23:2947-8.

[268] Suckow J, Markiewicz P, Kleina LG, Miller J, Kisters-Woike B, Muller-Hill B. Genetic studies of the Lac repressor. XV: 4000 single amino acid substitutions and analysis of the resulting phenotypes on the basis of the protein structure. J. Mol. Biol. 1996;261:509-23.

[269] Linnet TE, Teilum K. Non-uniform sampling of NMR relaxation data. J. Biomol. NMR. 2016.

[270] Jarymowycz VA, Stone MJ. Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. Chem. Rev. 2006;106:1624-71.

[271] Frederick KK, Marlow MS, Valentine KG, Wand AJ. Conformational entropy in molecular recognition by proteins. Nature. 2007;448:325-U3.

[272] Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, Kern D. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. Nature. 2007;450:913-U27.

[273] Kay LE, Muhandiram DR, Farrow NA, Aubin Y, Forman-Kay JD. Correlation between dynamics and high affinity binding in an SH2 domain interaction. Biochemistry. 1996;35:361-8.

[274] Li Z, Raychaudhuri S, Wand AJ. Insights into the local residual entropy of proteins provided by NMR relaxation. Protein Sci. 1996;5:2647-50.

[275] Marlow MS, Dogan J, Frederick KK, Valentine KG, Wand AJ. The role of conformational entropy in molecular recognition by calmodulin. Nat. Chem. Biol. 2010;6:352-8.

[276] Stevens SY, Sanker S, Kent C, Zuiderweg ERP. Delineation of the allosteric mechanism of a cytidylyltransferase exhibiting negative cooperativity. Nat. Struct. Biol. 2001;8:947-52.

[277] Tzeng S-R, Kalodimos CG. Dynamic activation of an allosteric regulatory protein. Nature. 2009;462:368-U139.

[278] Tzeng S-R, Kalodimos CG. Protein activity regulation by conformational entropy. Nature. 2012;488:236-40.

[279] Wand AJ, Moorman VR, Harpole KW. A surprising role for conformational entropy in protein function. Dynamics in Enzyme Catalysis2013. p. 69-94.

[280] Zidek L, Novotny MV, Stone MJ. Increased protein backbone conformational entropy upon hydrophobic ligand binding. Nat. Struct. Biol. 1999;6:1118-21.

[281] Lipari G, Szabo A. Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules 1. Theory and range of validity. J. Am. Chem. Soc. 1982;104:4546-59.

[282] Lipari G, Szabo A. Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules 2. Analysis of experimental results. J. Am. Chem. Soc. 1982;104:4559-70.

[283] Clore GM, Szabo A, Bax A, Kay LE, Driscoll PC, Gronenborn AM. Deviations from the simple 2-parameter model-free approach to the interpretation of N-15 nuclear magnetic-relaxation of proteins. J. Am. Chem. Soc. 1990;112:4989-91.

[284] Ryabov Y, Clore GM, Schwieters CD. Coupling between internal dynamics and rotational diffusion in the presence of exchange between discrete molecular conformations. J. Chem. Phys. 2012;136.

[285] Jones JA. Optimal sampling strategies for the measurement of relaxation times in proteins. J. Magn. Reson. 1997;126:283-6.

[286] Lee AL, Wand AJ. Assessing potential bias in the determination of rotational correlation times of proteins by NMR relaxation. J. Biomol. NMR. 1999;13:101-12.

[287] Harden BJ, Frueh DP. SARA: a software environment for the analysis of relaxation data acquired with accordion spectroscopy. J. Biomol. NMR. 2014;58:83-99.

[288] Bodenhausen G, Ernst RR. Direct determination of rate constants of slow dynamic processes by two-dimensional accordion spectroscopy in nuclear magnetic-resonance. J. Am. Chem. Soc. 1982;104:1304-9.

[289] Gledhill JM, Jr., Walters BT, Wand AJ. AMORE-HX: a multidimensional optimization of radial enhanced NMR-sampled hydrogen exchange. J. Biomol. NMR. 2009;45:233-9.

[290] Mayzel M, Rosenlow J, Isaksson L, Orekhov VY. Time-resolved multidimensional NMR with non-uniform sampling. J. Biomol. NMR. 2014;58:129-39.

[291] Long D, Delaglio F, Sekhar A, Kay LE. Probing invisible, excited protein states by non-uniformly sampled pseudo-4D CEST spectroscopy. Angewandte Chemie (International Edition). 2015;54:10507-11.

[292] Laue ED, Skilling J, Staunton J, Sibisi S, Brereton RG. Maximum-Entropy Method in Nuclear Magnetic-Resonance Spectroscopy. J. Magn. Reson. 1985;62:437-52.

[293] Sibisi S, Skilling J, Brereton RG, Laue ED, Staunton J. Maximum-Entropy Signal-Processing in Practical Nmr-Spectroscopy. Nature. 1984;311:446-7.

[294] Hoch JC. Maximum-Entropy Signal-Processing of Two-Dimensional Nmr Data. J. Magn. Reson. 1985;64:436-40.

[295] Schmieder P, Stern AS, Wagner G, Hoch JC. Quantification of maximum-entropy spectrum reconstructions. J. Magn. Reson. 1997;125:332-9.

[296] Matsuki Y, Eddy MT, Griffin RG, Herzfeld J. Rapid three-dimensional MAS NMR spectroscopy at critical sensitivity. Angewandte Chemie (International Edition). 2010;49:9215-8.

[297] Matsuki Y, Eddy MT, Herzfeld J. Spectroscopy by integration of frequency and time domain information for fast acquisition of high-resolution dark spectra. J. Am. Chem. Soc. 2009;131:4648-56.

[298] Matsuki Y, Konuma T, Fujiwara T, Sugase K. Boosting protein dynamics studies using quantitative nonuniform sampling NMR spectroscopy. Journal of Physical Chemistry B. 2011;115:13740-5.

[299] Hiller S, Ibraghimov I, Wagner G, Orekhov VY. Coupled decomposition of four-dimensional NOESY spectra. J. Am. Chem. Soc. 2009;131:12970-8.

[300] Jaravine VA, Zhuravleva AV, Permi P, Ibraghimov I, Orekhov VY. Hyperdimensional NMR spectroscopy with nonlinear sampling. J. Am. Chem. Soc. 2008;130:3927-36.

[301] Orekhov VY, Jaravine VA. Analysis of non-uniformly sampled spectra with multi-dimensional decomposition. Progress in Nuclear Magnetic Resonance Spectroscopy. 2011;59:271-92.

[302] Kazimierczuk K, Orekhov VY. Accelerated NMR spectroscopy by using compressed sensing. Angew. Chem. Int. Ed. Engl. 2011;50:5556-9.

[303] Oyen D, Fenwick RB, Stanfield RL, Dyson HJ, Wright PE. Cofactor-mediated conformational dynamics promote product release from Escherichia coli dihydrofolate reductase via an allosteric pathway. J. Am. Chem. Soc. 2015;137:9459-68.

[304] Donoho DL. De-noising by soft-thresholding. Ieee Transactions on Information Theory. 1995;41:613-27.

[305] Becker S, Bobin J, Candes EJ. NESTA: A fast and accurate first-order method for sparse recovery. Siam Journal on Imaging Sciences. 2011;4:1-39.

[306] Nesterov Y. Smooth minimization of non-smooth functions. Mathematical Programming. 2005;103:127-52.

[307] Sun S, Gill M, Li Y, Huang M, Byrd RA. Efficient and generalized processing of multidimensional NUS NMR data: the NESTA algorithm and comparison of regularization terms. J. Biomol. NMR. 2015;62:105-17.

[308] Wand AJ, Urbauer JL, McEvoy RP, Bieber RJ. Internal dynamics of human ubiquitin revealed by C-13-relaxation studies of randomly fractionally labeled protein. Biochemistry. 1996;35:6116-25.

[309] Kranz JK, Lee EK, Nairn AC, Wand AJ. A direct test of the reductionist approach to structural studies of calmodulin activity - Relevance of peptide models of target proteins. J. Biol. Chem. 2002;277:16351-4.

[310] Dodevski I, Nucci NV, Valentine KG, Sidhu GK, O'Brien ES, Pardi A, et al. Optimized reverse micelle surfactant system for high-resolution NMR spectroscopy of encapsulated proteins and nucleic acids dissolved in low viscosity fluids. J. Am. Chem. Soc. 2014;136:3465-74.

[311] Tugarinov V, Kanelis V, Kay LE. Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. Nat. Protoc. 2006;1:749-54.

[312] Nucci NV, Marques BS, Bedard S, Dogan J, Gledhill JM, Jr., Moorman VR, et al. Optimization of NMR spectroscopy of encapsulated proteins dissolved in low viscosity fluids. J. Biomol. NMR. 2011;50:421-30.

[313] Findeisen M, Brand T, Berger S. A H-1-NMR thermometer suitable for cryoprobes. Magn. Reson. Chem. 2007;45:175-8.

[314] Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, et al. Backbone dynamics of a free and a phosphopeptide-complexed Src homology-2 domain studied by N-15 NMR relaxation. Biochemistry. 1994;33:5984-6003.

[315] Millet O, Muhandiram DR, Skrynnikov NR, Kay LE. Deuterium spin probes of side-chain dynamics in proteins. 1. Measurement of five relaxation rates per deuteron in C-13-labeled and fractionally H-2-enriched proteins in solution. J. Am. Chem. Soc. 2002;124:6439-48.

[316] Lakomek N-A, Ying J, Bax A. Measurement of N-15 relaxation rates in perdeuterated proteins by TROSY-based methods. J. Biomol. NMR. 2012;53:209-21.

[317] Fu Y, Kasinath V, Moorman VR, Nucci NV, Hilser VJ, Wand AJ. Coupled motion in proteins revealed by pressure perturbation. J. Am. Chem. Soc. 2012;134:8543-50.

[318] Boggs PT, Rogers JE. Orthogonal distance regression. Contemporary Mathematics. 1990;112:186.

[319] Candes EJ, Wakin MB. An introduction to compressive sampling. Ieee Signal Processing Magazine. 2008;25:21-30.

[320] Logan BF. Properties of high-pass signals, Ph.D. Thesis: Columbia University; 1965.

[321] Donoho DL, Stark PB. Uncertainty principles and signal recovery. SIAM Journal on Applied Mathematics. 1989;49:906-31.

[322] Hyberts SG, Arthanari H, Robson SA, Wagner G. Perspectives in magnetic resonance: NMR in the post-FFT era. J. Magn. Reson. 2014;241:60-73.

[323] Ferrage F, Piserchio A, Cowburn D, Ghose R. On the measurement of N-15-{H-1} nuclear Overhauser effects. J. Magn. Reson. 2008;192:302-13.

[324] Rovnyak D, Hoch JC, Stern AS, Wagner G. Resolution and sensitivity of high field nuclear magnetic resonance spectroscopy. J. Biomol. NMR. 2004;30:1-10.

[325] Palmer MR, Suiter CL, Henry GE, Rovnyak J, Hoch JC, Polenova T, et al. Sensitivity of nonuniform sampling NMR. J. Phys. Chem. B. 2015;119:6502-15.

[326] Hyberts SG, Arthanari H, Wagner G. Applications of non-uniform sampling and processing. Novel Sampling Approaches in Higher Dimensional NMR2012. p. 125-48.

[327] Cavanagh J, Fairbrother WJ, Palmer AG, Rance M, Skelton NJ. Protein NMR Spectroscopy Principles and Practice. Burlington, MA, USA: Elsevier; 2007.

[328] Kasinath V, Sharp KA, Wand AJ. Microscopic insights into the NMR relaxation-based protein conformational entropy meter. J. Am. Chem. Soc. 2013;135:15092-100.