

A FRAGILE NATIVE STATE STRUCTURE: AN  
ARYL HYDROCARBON RECEPTOR NUCLEAR TRANSLOCATOR (ARNT)  
VARIANT EXHIBITS SLOW INTERCONVERSION KINETICS  
BETWEEN TWO DIFFERENT FOLDS

APPROVED BY SUPERVISORY COMMITTEE

---

Kevin Gardner, Ph.D.

---

Elizabeth Goldsmith, Ph.D.

---

Jose Rizo-Rey, Ph.D.

---

Philip Thomas, Ph.D.

## DEDICATION

This work could not have been completed without the assistance of a number of individuals. First of all, I would like to greatly thank my mentor, Kevin Gardner, for his considerable guidance, support and training to become an independent thinker. I would also like to thank my thesis committee, Elizabeth Goldsmith, Jose Rizo-Rey and Philip Thomas for their guidance and encouragement throughout this process.

I would also like to thank those individuals who assisted this project in the lab. Paul Card started this project and passed on his knowledge on a number of techniques. He was a joy to work with and truly missed when he left the lab. I would also like to thank several students that worked on various aspects of the project. Leanna Steier spent countless hours screening compounds, Kyle Brewer investigated the unfolding pathways, and Amy Zhou created a number of mutations. I could not have completed this project without this assistance and I wish them all the best of luck.

Lastly, I would like to extend a special thanks to my parents, Kimberly and Curtis, for their endless support and daily prayers, to my grandparents, Helen and Roland Larson, for their constant encouragement, and to my beautiful wife, Amanda, who made this journey all worthwhile.

A FRAGILE NATIVE STATE STRUCTURE: AN  
ARYL HYDROCARBON RECEPTOR NUCLEAR TRANSLOCATOR (ARNT)  
VARIANT EXHIBITS SLOW INTERCONVERSION KINETICS  
BETWEEN TWO DIFFERENT FOLDS

by

MATTHEW RYAN EVANS

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

August, 2009

Copyright

by

Matthew Ryan Evans, 2009

All Rights Reserved



A FRAGILE NATIVE STATE STRUCTURE: AN  
ARYL HYDROCARBON RECEPTOR NUCLEAR TRANSLOCATOR (ARNT)  
VARIANT EXHIBITS SLOW INTERCONVERSION KINETICS  
BETWEEN TWO DIFFERENT FOLDS

Publication No. \_\_\_\_\_

Matthew Ryan Evans, Ph.D

The University of Texas Southwestern Medical Center at Dallas, 2009

Supervising Professor: Kevin H. Gardner, Ph.D

The aryl hydrocarbon receptor nuclear translocator (ARNT) is a promiscuous basic helix-loop-helix Period/ARNT/Single-minded (bHLH-PAS) protein that controls various biological pathways by forming heterodimeric transcriptional regulator complexes with several other bHLH-PAS proteins via the  $\beta$ -sheet surfaces of its two PAS domains. The  $\beta$ -sheets of PAS domains are

involved in many intermolecular interactions with other proteins and natural cofactors in order to detect environmental changes in sensor PAS proteins.

As part of a study of the HIF-2 $\alpha$ :ARNT PAS-B heterodimer, site-directed mutagenesis was performed on the ARNT PAS-B domain. Interestingly, one point mutation on the ARNT  $\beta$ -sheet surface (Y456T) resulted in a new conformation of the domain that existed in equimolar concentrations with the wild-type conformation. Subsequent studies demonstrated both that the two conformations are in equilibrium and that relative populations of the two conformations can be perturbed by additional mutations. Using solution NMR spectroscopy, we solved the high resolution solution structure of a mutant ARNT PAS-B domain in the new conformation, demonstrating that it contains a three-residue slip in register and accompanying inversion of the central I $\beta$ -strand. In addition, this new conformer has a greater than hundred-fold reduction in affinity for HIF-2 $\alpha$  PAS-B, disrupting the hypoxia response pathway.

Solution NMR measurements of the interconversion kinetics have let us establish that these two conformations interconvert slowly ( $\tau = 40$  min at 25°C) with a linear Arrhenius temperature-dependence of the interconversion rate. Stopped-flow unfolding experiments using GdmHCl on Y456T, revealed a similarly slow unfolding rate ( $\tau \sim 25$  min at 25°C) and an energy barrier to unfold of  $\sim 13$  kcal/mol, which is nearly identical to that for the interconversion process. These data indicate that the protein must undergo a global unfolding process in

order to interconvert between conformations. Lastly, these relative populations of Y456T can be affected by compound preferentially binding into the core of one of these conformations.

This discovery highlights the malleability of PAS  $\beta$ -sheets and suggests ARNT may act as a regulatory switch to control different biological pathways. Furthermore, this system presents a great opportunity to further understand the structural and kinetic impact of  $\beta$ -strand slips observed in nature.

## TABLE OF CONTENTS

<u>I. Introduction - PAS domain-mediated signaling</u>	1
A. PAS domains	1
i. PAS domain structure and diversity	
ii. PAS domains as environmental sensors	
a. Redox state and oxygen sensing	
b. PAS domains as light sensors	
iii. PAS-mediated protein-protein complexes	
B. bHLH-PAS domains role as components of transcription factors	12
i. The Aryl hydrocarbon Receptor Nuclear Translocator	
a. ARNT mediated pathways	
1. Single-minded Protein	
2. The Aryl hydrocarbon receptor	
3. The Hypoxia Inducible Factor	
b. Biological significance of ARNT	
C. One common $\beta$ -sheet to bind various ligands/proteins	23
D. Questions addressed in this thesis	28
<u>II. Conformational flexibility observed in nature</u>	29
A. The protein folding problem	29
B. One sequence equals one structure	29
i. Challenges to Anfinsen's postulate	
a. Arc repressor mutant: switch Arc	
b. Serpins	
c. Lymphotactin	
<u>III. A second conformation of ARNT PAS-B nearby in sequence space</u>	41
A. Introduction	42

B. Results	46
i. Point mutations reveal second conformation	
ii. Interconversion rates between the two conformations	
iii. Solution structure of the new conformation 2	
iv. X-Pro isomerization associated with the two conformations	
v. Effect of slipped conformation on domain stability	
vi. Functional effect of I $\beta$ slip	
C. Discussion	64
i. Structural features that allow this transition	
ii. PAS domains flexible central $\beta$ -sheet	
iii. $\beta$ -strand slips in nature	
D. Materials and Methods	71
i. Mutagenesis	
ii. Protein expression and purification	
iii. Biochemical purification of mutant conformation in Y456T	
iv. Structure determination of ARNT PAS-B F444Q/F446A/Y456T	
v. Global stability measurements	
E. Acknowledgements	75
<u>III. Transition between two <math>\beta</math>-strand registers of ARNT PAS-B Y456T goes through a chiefly unfolded state</u>	76
A. Introduction	76
B. Results for ARNT PAS-B Y456T interconversion and unfolding kinetics	77
i. Kinetics of interconversion for ARNT PAS-B Y456T	
ii. Rates of folding and unfolding for ARNT PAS-B mutants	
iii. Model for ARNT PAS-B Y456T interconversion	
C. Effect of a N448-P449 peptide bond isomerization on the kinetics	86
i. Results	
ii. Stopped-flow experiments	
D. Materials and Methods	93

E. Acknowledgements	96
<u>IV. Factors affecting the equilibrium of ARNT PAS-B Y456T</u>	97
A. Factors that affect equilibrium	97
i. Solution conditions affect equilibrium	
ii. Temperature dependence of the equilibrium	
iii. Denaturant studies suggests both conformations are equally stable	
B. Effects of various point mutations on the equilibrium constant	104
i. Mutations at Y456T	
ii. Mutations in the HI-loop	
iii. Length of HI-loop affects equilibrium	
C. Ligand binding to shift equilibrium of ARNT PAS-B Y456T	112
i. Ligand binding of PAS domains	
ii. Initial setup of compound library	
iii. Thermal shift assay to detect ligand binding	
a. Ligand screening strategy and methods	
b. Lead compound hits for the initial ligand screen	
iv. NMR to detect ligand binding	
a. NMR screen on positive hits by thermal shift assay	
b. New NMR screen of the 850 compound library	
v. Different assays, different results	
vi. Preferential binding for KG-548 against Y456T	
vii. KG-548 drives equilibrium towards the wildtype	
D. Conclusions	128
E. Materials and Methods	129
i. NMR analysis for Y456T equilibrium constant	
ii. Ligand screening strategy and methods	
iii. NMR analysis of positive hits by thermal shift assay	
iv. Structure determination of ARNT PAS-B F444Q/F446A/Y456T	
v. Global stability measurements	
F. Acknowledgements	133

<u>IV. Conclusions and Future Directions</u>	134
A. Conclusions	134
B. Future Directions	135

## PRIOR PUBLICATIONS

Matthew R. Evans, Paul B. Card, and Kevin H. Gardner (2009) “ARNT PAS-B has a fragile native state structure with an alternative beta-sheet register nearby in sequence space” *Proc Natl Acad Sci.* 106(8): 2617-22.

Junko Kajimura, Arifur Rahman, James Hsu, Matthew R. Evans, Kevin H. Gardner, and Paul D. Rick (2006) “O Acetylation of the Enterobacterial Common Antigen Polysaccharide Is Catalyzed by the Product of the yiaH Gene of *Escherichia coli* K-12” *J. Bacteriol.* 188: 7542-7550.



## LIST OF FIGURES

### Chapter I: Introduction

Figure 1-1: PAS domain fold .....	4
Figure 1-2: Mode of regulation for Arc .....	6
Figure 1-3: Mode of regulation for FixJ .....	8
Figure 1-4: Model of PYP conformational changes .....	9
Figure 1-5: bHLH-PAS transcription factors .....	13
Figure 1-6: Domain architecture of bHLH-ARNT .....	15
Figure 1-7: Solution structure of ARNT PAS-B .....	15
Figure 1-8: Crystal structure of ARNT/HIF PAS-B heterodimer .....	16
Figure 1-9: ARNT/AhR pathway .....	19
Figure 1-10: Hypoxia response pathway .....	21
Figure 1-11: One common $\beta$ -sheet surface .....	24
Figure 1-12: Two different dimers of KinA .....	26
Figure 1-13: Crystal structure of a mutant HIF-2 $\alpha$ PAS-B .....	27

### Chapter II: Conformational flexibility observed in nature

Figure 2-1: Anfinsen's experiment on ribonuclease A.....	31
Figure 2-2: Two structures of switch Arc .....	35
Figure 2-3: Two structures of serpin PAI-1 .....	37

Figure 2-4: Structures of lymphotactin.....	39
Chapter III: A second conformation of ARNT PAS-B nearby in sequence space	
Figure 3-1: Architecture of human ARNT.....	45
Figure 3-2: Two conformations of ARNT PAS-B Y456T .....	48
Figure 3-3: $^{15}\text{N}/^1\text{H}$ HSQC spectrum of both conformations .....	52
Figure 3-4: Twenty lowest energy structures of conformation <b>2</b> .....	54
Figure 3-5: Solution structure of ARNT PAS B conformation <b>2</b> .....	54
Figure 3-6: I $\beta$ strand adopts similar conformations in both structures.....	56
Figure 3-7: Solvent exposed residues in both conformations are similar.....	57
Figure 3-8: Core residues adopt a similar conformation in both structures.....	57
Figure 3-8: Evidence for N448-P449 peptide bond isomerization.....	59
Figure 3-9: Monitoring the unfolding process using Trp fluorescence .....	61
Figure 3-10: Both conformations have similar global stabilities .....	62
Figure 3-11: HIF-2 $\alpha$ PAS-B selectively binds conformation <b>1</b> .....	64
Figure 3-12: Sequence alignment of human ARNT with homologues .....	68
Chapter IV: Kinetics associated with the $\beta$ -strand slip	
Figure 4-1: Two conformations of ARNT PAS-B .....	77
Figure 4-2: Rates of interconversion for ARNT PAS-B Y456T.....	79
Figure 4-3: Eyring plot for the interconversion of Y456T .....	80

Figure 4-4: Chevron plot for ARNT PAS-B wildtype and mutants.....	82
Figure 4-5: Temperature dependence for the unfolding of Y456T .....	83
Figure 4-6: Summary of rates .....	85
Figure 4-7: Equilibrium denaturation of ARNT PAS-B P449A/Y456T .....	86
Figure 4-8: ARNT PAS-B P449A/Y456T does not exchange.....	88
Figure 4-9: Chevron plot for ARNT PAS-B mutants.....	90
Figure 4-10: Elution profile of Y456T injected onto a MonoQ column .....	94

## Chapter V: Factors affecting the equilibrium of ARNT PAS-B Y456T

Figure 5-1: Effects of temperature on the equilibrium .....	100
Figure 5-2: $^{15}\text{N}/^1\text{H}$ HSQC of ARNT PAS-B Y456T at various temps .....	101
Figure 5-3: GdmHCl induced unfolding transition monitored by NMR.....	103
Figure 5-4: Multiple conformations of ARNT PAS-B Y456 mutants .....	106
Figure 5-5: Structures of the HI loop for both conformations .....	109
Figure 5-6: Multiple conformations of ARNT PAS-B P449A .....	110
Figure 5-7: Sequence alignment of TEV site incorporated .....	111
Figure 5-8: Typical thermal shift assay results .....	115
Figure 5-9: Summary of compound binding by thermal shift assay .....	117
Figure 5-10: Titration series of ARNT PAS-B wildtype with KG-077 .....	118
Figure 5-11: Compound binding by monitored by $^{15}\text{N}/^1\text{H}$ HSQC spectra...	121
Figure 5-12: Structure of KG-548 .....	123

Figure 5-13: KG-548 preferentially binds ARNT PAS-B wildtype.....	125
Figure 5-14: Binding curve of KG-548 with ARNT PAS-B Y456T .....	127
Figure 5-15: Scatchard plot of KG-548 binding to ARNT PAS-B Y456T ..	127
Figure 5-16: Diagram showing a four state ligand binding model.....	128

## List of Tables

Table 3-1: Effect of ARNT PAS-B mutations on relative populations .....	50
Table 3-2: Statistics for ARNT PAS-B F444Q/F446A/Y456T structure .....	53
Table 4-1: Thermodynamic parameters for ARNT PAS-B Y456T .....	80
Table 4-2: Rates of folding and unfolding for ARNT PAS-B mutants .....	83
Table 5-1: Solvent effect on the relative populations of ARNT PAS-B Y456 ...	98
Table 5-2: Relative populations of ARNT PAS-B Y456 mutations .....	106
Table 5-3: Compounds preferentially binding to wildtype conformation .....	119
Appendix I: Relative populations of all ARNT PAS-B mutations .....	138
Appendix II: Additional mutations that do not affect equilibrium .....	139

## I. Introduction - PAS domain-mediated signaling

It is well understood that cell's survival depends on the ability to detect and respond to fluctuations in the environment, in order to maintain physiological homeostasis. As such, exogenous signals often modulate biochemical pathways in order to maintain cyclic behaviors, cellular metabolism, growth patterns, and many other responses of the cell. These pathways are often closely regulated as deviations can lead to a variety of diseases and even cell death. Therefore, a greater understanding of signal transduction mechanisms is essential to prevent these diseases from occurring. The purpose of my research is to investigate the flexibility inherent to a certain class of protein domains involved in signal transduction; most specifically how the Aryl hydrocarbon Receptor Nuclear Translocator (ARNT) protein modulates multiple biochemical pathways using a simple Per-ARNT-Sim (PAS) fold.

### *A. PAS domains*

One such family of domains that mediate protein-protein interactions involved in signal transduction pathways in all three kingdoms of life (*Bacteria*, *Archaea*, and *Eucarya*) are PAS domains (20, 21). PAS domains are named after the three proteins in which they were first identified: the *Drosophila* period clock protein (PER) was discovered as a regulatory protein controlling circadian

rhythms (22, 23); the vertebrate Aryl hydrocarbon Receptor Nuclear Translocator (ARNT) was identified as an essential component for Aryl Hydrocarbon Receptor (AHR) signaling (24); and lastly, the *Drosophila* single-minded protein (SIM) was found as a protein involved in regulating midline cell development (21, 25). PAS domains often bind biological cofactors, such as heme, flavin adenine dinucleotide, and flavin mononucleotide, and conferring sensitivity to environmental conditions including redox potential, blue light, and oxygen levels on these functions. Many proteins have been found to contain multiple PAS domains with quite different origins. For instance, the cyanobacterial protein Slr0222 contains six different PAS domains, each more similar to PAS domains from archaeal and human proteins than to each other (20). This discovery suggests that these multiple PAS-containing sensory proteins may be able to perform complex integration of multiple input signals.

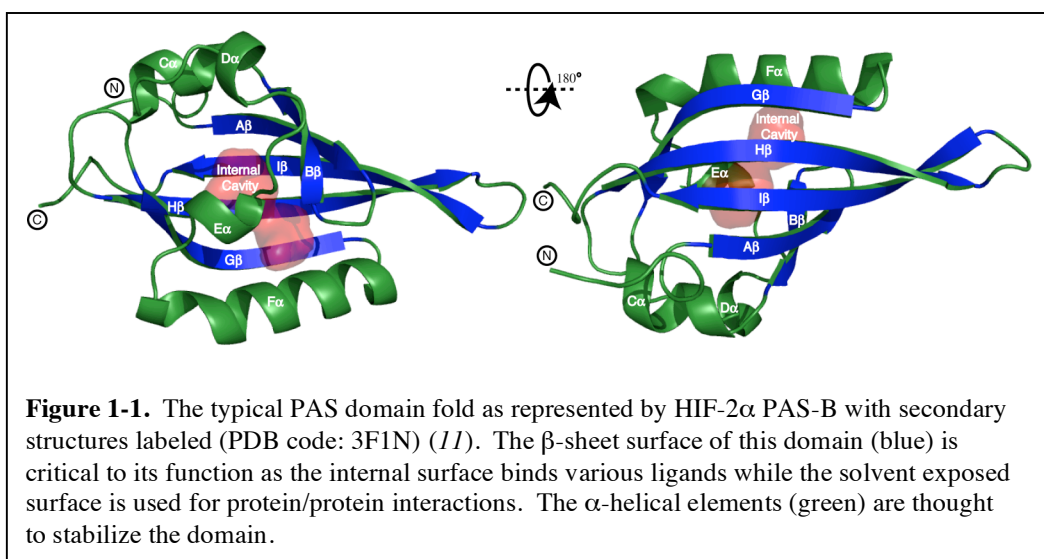
*i. PAS domain structure and diversity*

Despite their diverse biological functions, PAS domains demonstrate a high degree of structural similarity. They adopt a mixed  $\alpha/\beta$  fold of approximately 100 residues in length and are characterized by several  $\alpha$ -helices packing against one side of a five-stranded anti-parallel  $\beta$ -sheet (Figure 1-1) (20). This  $\beta$ -sheet plays a particularly critical role in PAS structure and function as it is involved in a wide variety of intra- and intermolecular interactions, ranging from

binding environmentally sensitive cofactors to other PAS-containing proteins (12, 18, 26-28). While the  $\alpha$ -helical surface is generally thought to help stabilize the domain, in some PAS domains, such as NcoA-1 PAS-B (29), these secondary elements form a critical dimerization interface with other proteins. Interestingly, despite their conserved common structure, PAS domains demonstrate a high degree of sequence diversity. Sequence alignment of more than fifty previously identified PAS domains reveals a conserved N-terminal PAS box, containing the A $\beta$ -F $\alpha$  segments, as well as a C-terminal PAC box, which includes the last three strands G $\beta$ -I $\beta$  (30). The PAS and PAC boxes, which were previously identified as S1 and S2 boxes (30), respectively, are linked by sequences that are quite diverse in composition and length, which form specific interactions with the hydrophobic core and bound cofactors. The variety of sequences in this region may help explain the number of different cofactors and other proteins PAS domains bind in many biochemical pathways. At the time of which these two sequence “boxes” were identified in 1997, the only PAS domain structure known was photoactive yellow protein (PYP). Since that time, dozens of PAS domain structures have been identified making the PAS/PAC and S1/S2 box nomenclature irrelevant. The identification of PAS domains was also assisted by the development of Hidden Markov Models (HMMs), a statistical method which identified conserved residues important for folding and function when the overall primary sequence greatly varies (31). Over the past decade, additional structural



and sequence information of PAS domains have helped advance HMMs to find the full domain instead of the individual S1/S2 or PAS/PAC boxes (32). To date, over 7000 PAS-containing proteins have been identified in all three kingdoms of life that perform functions as diverse as DNA binding, phosphorylation, and ion conduction.



## *ii. PAS domains as environmental sensors*

Critical for cell survival, cell must respond to fluctuations in the external environment around them. Therefore, intracellular processes are highly dependent upon robust and specific responses to extracellular signals in order to maintain homeostasis. Cells respond to environmental changes using molecular switches to not just sense but also turn on the correct biological pathway to re-establish equilibrium. Well-studied molecular switches that cells use are post-translational modifications, including hydroxylation, acetylation, methylation,

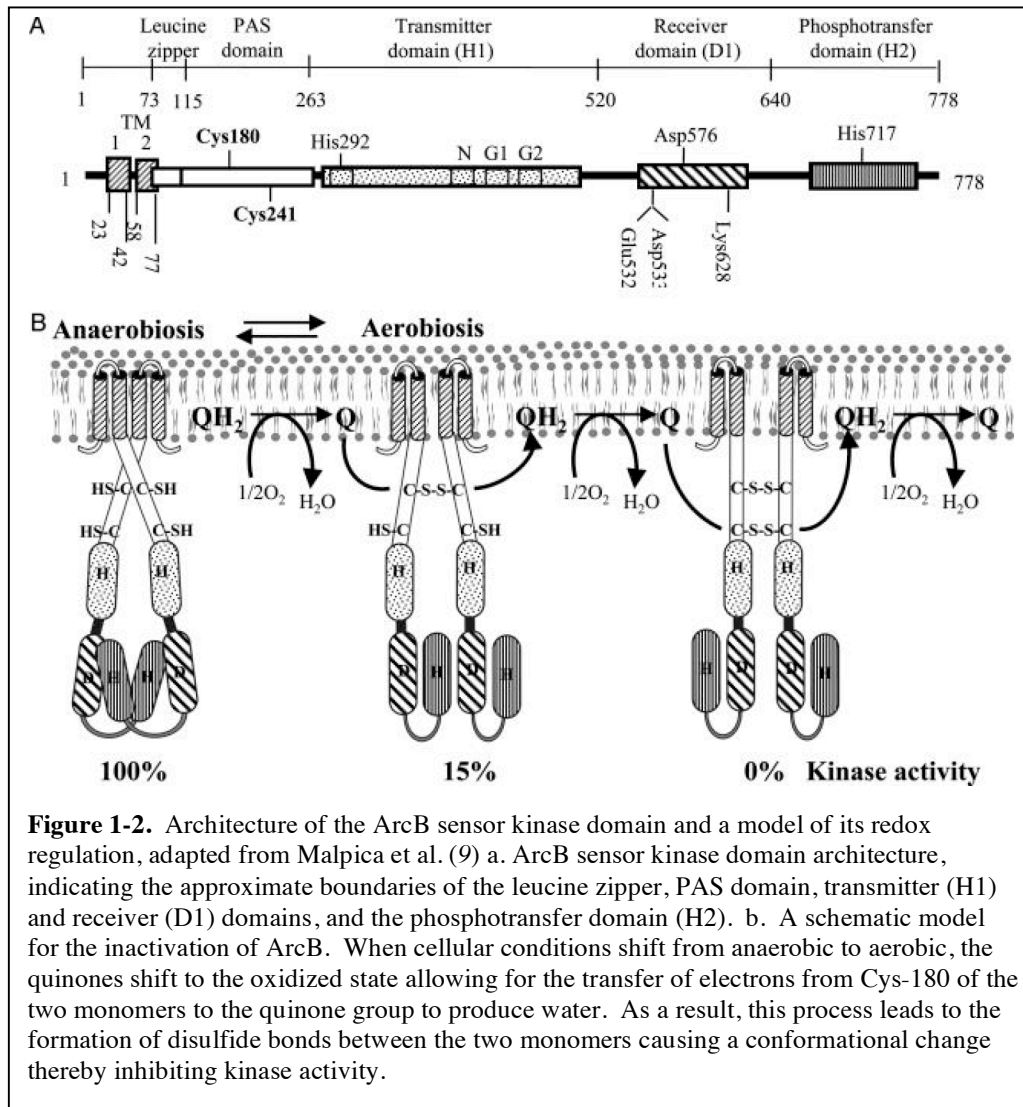
ubiquitination, and phosphorylation (33). These modifications can change a protein's binding interface, thereby modulating protein/protein affinities.

One such family of domains that mediate protein-protein interactions involved in signal transduction pathways are PAS domains. This family of domains is able to sense and transmit signals by monitoring changes in binding occupancy and/or configuration of environmentally-sensitive cofactors. These conformational changes have been well documented for PAS domains that control responses to various environmental stimuli, such redox potential, light, and oxygen tension in the cell.

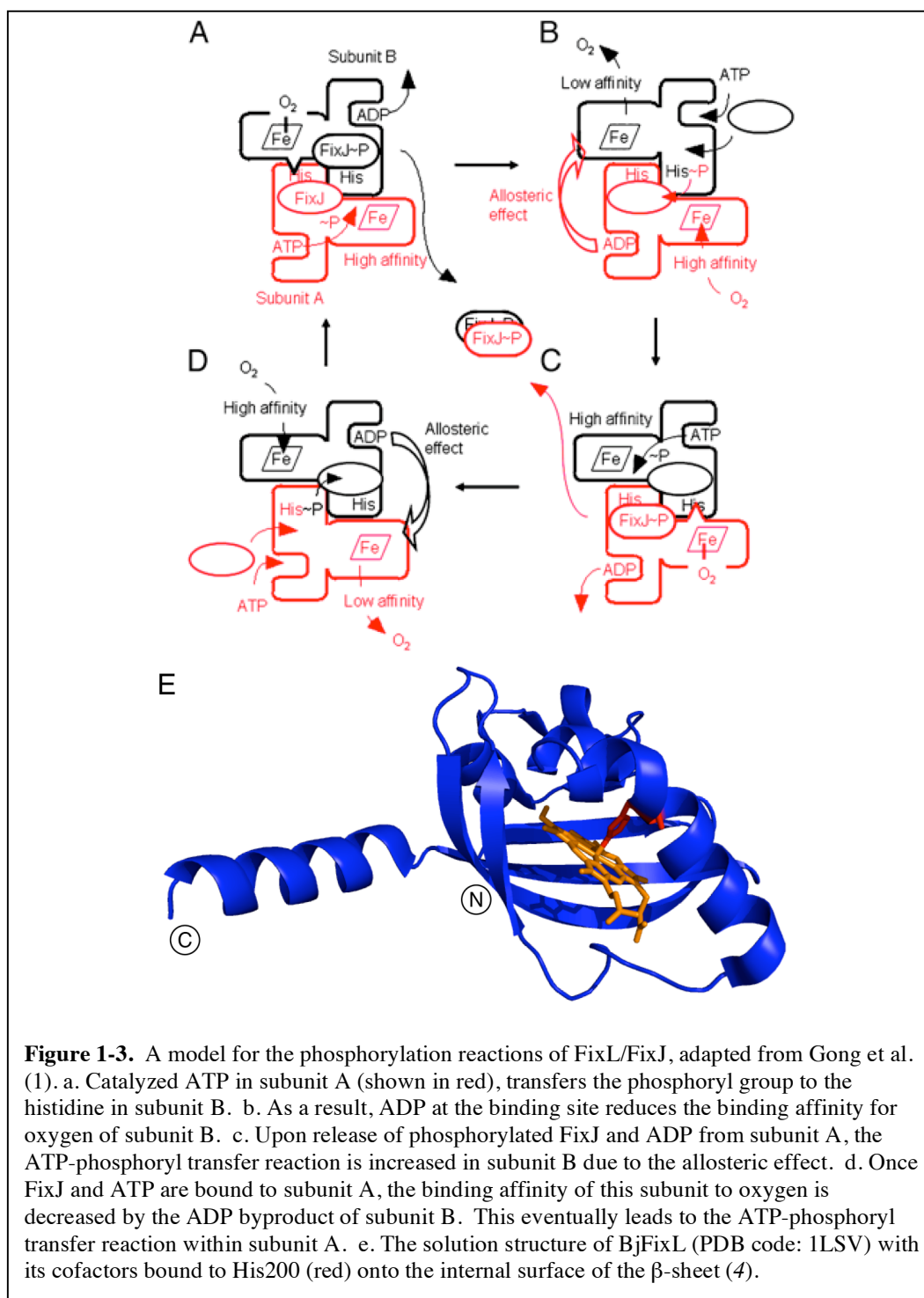
*a. Redox state and oxygen sensing*

One of the first indications of cellular energy depletion shows up as a decrease in the electron transport chain activity due to the close association between the electron transport chain and the production of new ATP. During oxidative phosphorylation, oxygen is the terminal electron acceptor and thus, is vital for cells to closely monitor intracellular oxygen concentrations (34). The *E. coli* aerotaxis receptor, Aer has an N-terminal cytoplasmic PAS domain that binds an FAD cofactor to detect oxygen-dependent cellular redox changes (35, 36). Upon changes in redox potential, Aer rapidly transmits the necessary signals to guide the cell into regions where the oxygen concentration is optimal for growth. Alternatively, *E. coli* can use a two-component system involving the sensor

kinase Arc (anoxic redox control) that does not require a bound cofactor. Instead, under aerobic conditions, ArcB oxidizes two redox-active cysteine residues that then form an intermolecular disulfide bond that causes a conformation change of the ArcB dimer, thereby inhibiting kinase activity (Figure 1-2) (9).

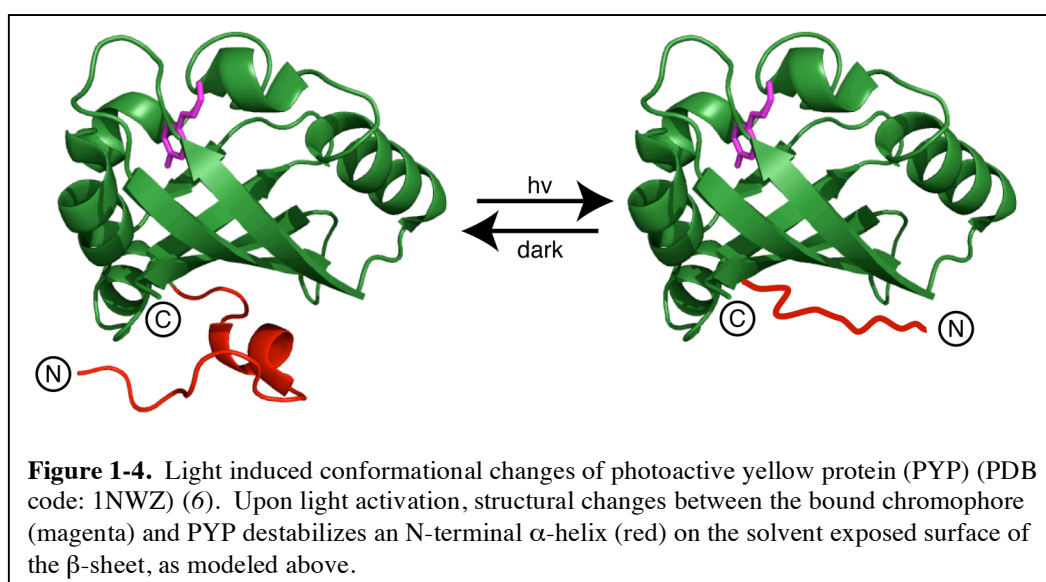


Oxygen sensing is also important for nitrogen fixation, a process by which atmospheric nitrogen is converted to ammonia by an oxygen-labile nitrogenase. This process is vital for cell survival as ammonium derivatives are used as reducing agents for oxidative phosphorylation (20). One bacterial PAS-containing protein that is able to directly sense oxygen through an internally bound heme cofactor is FixL. The heme molecule becomes more planar as oxygen binds, resulting in a conformational change in the FG-loop of the domain (1). This subtle change, which now allows FixL to bind a histidine kinase, is proposed as the mechanism by which the histidine kinase domain is inhibited (Figure 1-3).



*b. PAS domains as light sensors*

Many PAS domains sense light and direct themselves into optimal growth conditions. For example, certain halophilic bacteria sense blue light by using photoactive yellow protein (PYP), a photosensory protein consisting only of a single PAS domain, which has been suggested to act as a photoreceptor that controls for negative phototaxis (37, 38). Light induced conformational changes involve a *cis-trans* isomerization of a thioester-linked p-coumaric acid chromophore bound in the hydrophobic core (39-41). This interaction destabilizes an N-terminal helix-loop-helix motif from the solvent exposed surface of the  $\beta$ -sheet, which has been implicated in many PAS domains as a surface for both intra- and intermolecular interactions (Figure 1-4).



A subset of the PAS domain family that responds to blue light and use a flavin bound chromophore are called light, oxygen, voltage (LOV) domains. These domains were initially identified in phototropins, proteins that regulate plant activity such as stomata openings, phototropism, chloroplast relocations, and leaf opening (42, 43). Phototropins consist of two LOV domains N-terminal to a serine/threonine kinase domain. These LOV domains use light-sensitive flavin cofactors (flavin mononucleotide - FMN or flavin adenosyl-dinucleotide - FAD) that bind noncovalently within the core of the protein. Upon exposure to blue light, a covalent bond is formed between the C4a position of the flavin moiety and a conserved cysteine, leading to a global conformational change resulting in an active signaling state (44, 45). Without additional illumination, this active state exists for a finite period of time, ranging from seconds to minutes before spontaneously decaying back to the dark, non-active state where another round of photocycling can occur. This process has been well documented for the *Avena sativa* phototropin 1 LOV2 domain (AsLOV2), which binds an FMN chromophore. In this system, blue light causes a covalent cysteine adduct to form between the S $\gamma$  of the conserved Cys450 and the isoalloxazine ring C(4a) position of the flavin moiety and lasts for tens of seconds (44, 45). Cysteine at position 450 is absolutely necessary for adduct formation, and ultimately kinase activation, as mutagenesis of this residue abolishes the cycling completely (46). As a result of the adduct formation, the protein changes its conformation and hydrogen-

bonding patterns with the internally bound flavin chromophore. As such, distortions in the central  $\beta$ -sheet cause a C-terminal helix, termed  $J\alpha$ , to dissociate from the solvent exposed side of the  $\beta$ -sheet and eventually unfold, thereby exposing the  $\beta$ -sheet and ultimately leading to kinase activation (28, 47). While the solvent exposed  $\beta$ -sheet surface is not covered by intramolecular helical elements in all PAS-containing proteins, what is common among the PAS family of proteins is the use of the  $\beta$ -sheet for PAS-directed protein regulation.

### *iii. PAS-mediated protein-protein complexes*

As previously noted, the  $\beta$ -sheet of PAS domains plays a critical role for the function of PAS domains, as it is the binding interface for most PAS-bound protein targets. Many PAS domains control the binding affinity of this site by changes in occupancy of environmentally-sensitive cofactors bound in the core of the domain, such as those described thus far. However, not all PAS domains have been identified as binding endogenous ligands. Instead, some of these PAS proteins transmit environmental signals to other proteins through a wide variety of apparently constitutive protein-protein interactions. For example, LuxQ contains two tandem PAS domains, neither of which have a known endogenous ligand. Instead, these two domains homodimerize and bind to LuxP forming a 1:1 complex (48). However, upon ligand binding to LuxP, this complex is reorganized. Higher order multimeric complexes formed by PAS domains have

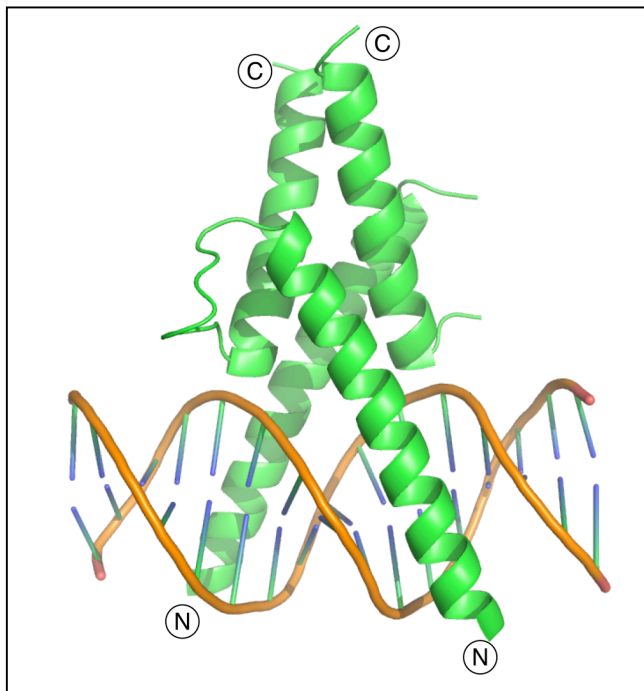


also been observed in nature. Examples of these include PAS domains from the RmFixL protein that forms a 2:1 complex with RmFixJ (49), as well as the isolated N-terminal PAS domain of HERG that can form tetramers *in vitro* (50).

### *B. bHLH-PAS domains role as components of transcription factors*

Within higher eukaryotes, PAS domains are found in several classes of DNA binding proteins, one of which is the basic helix-loop-helix (bHLH) domain. PAS proteins that contain an N-terminal bHLH domain, termed bHLH/PAS, control the gene expression activity for many biological processes, including those involved in cell development, maintenance, and homeostasis (51-53). One dimerization interface in bHLH/PAS proteins is the helix-loop-helix domain (Figure 1-5). Dimerization of this domain must occur to form a fully functional DNA-binding complex. This functional complex binds DNA by virtue of the basic region N-terminal to the helix-loop-helix domain (54). bHLH/PAS proteins can be subdivided into two classes. Class I neither homodimerizes nor heterodimerizes with other class I proteins. Examples include the aryl hydrocarbon receptor (AhR), the hypoxia inducible factors (HIF-1 $\alpha$ , HIF-2 $\alpha$  and HIF-3 $\alpha$ ), and the single minded proteins (SIM). However, to be an active complex, they must dimerize with a class II protein, such as the aryl hydrocarbon receptor nuclear translocator (ARNT), which promiscuously homo- and heterodimerizes (7, 55, 56). Another dimerization interface is provided by the

PAS domains, which can selectively determine the binding partners and ultimately target gene recognition (57). A well-characterized bHLH-PAS protein involved in several developmental and physiological processes is the Aryl Hydrocarbon Receptor Nuclear Translocator.



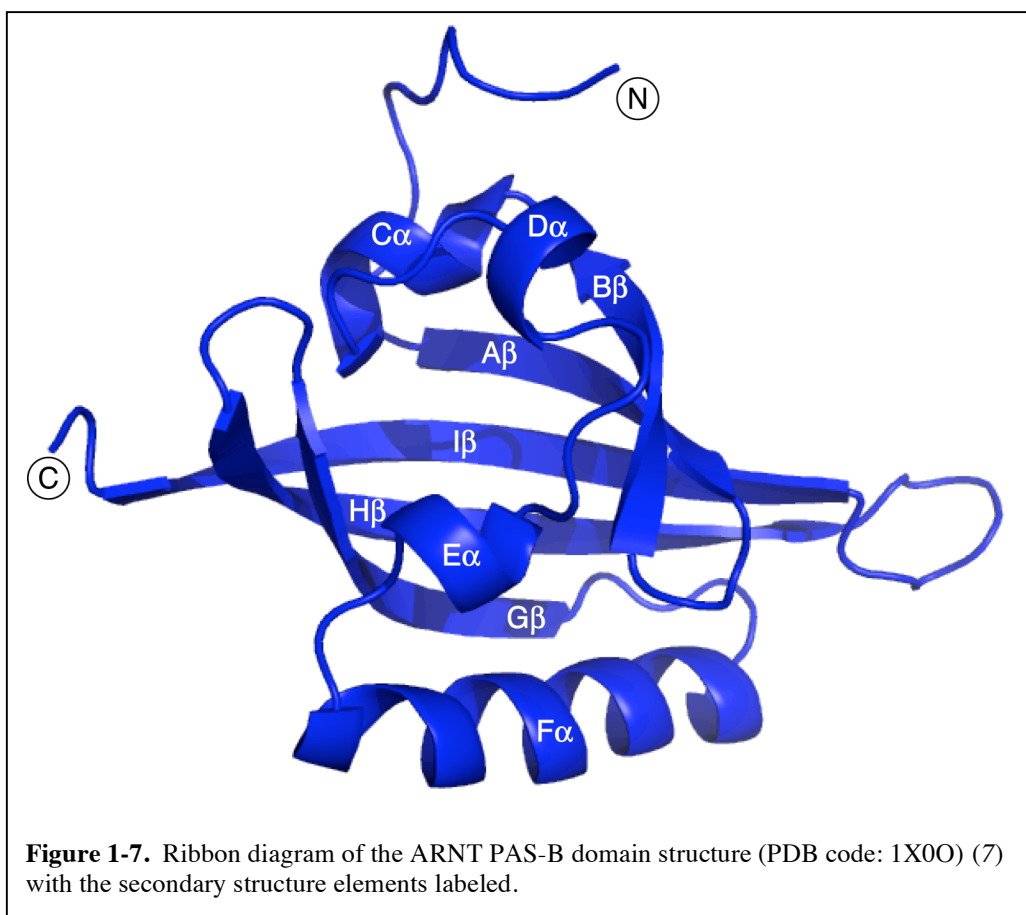
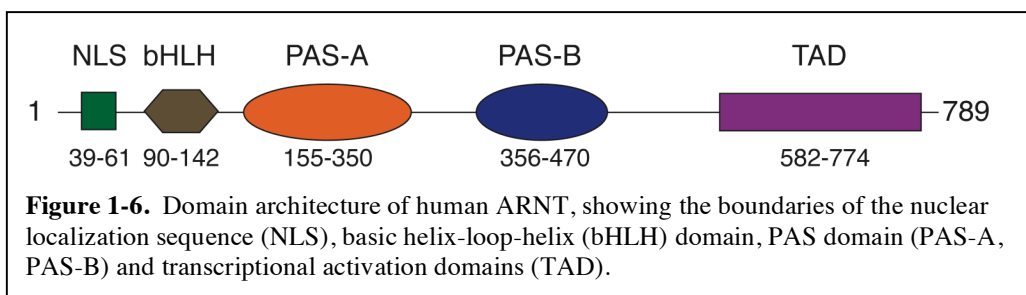
**Figure 1-5.** The crystal structure of the basic helix-loop-helix domains of the heterodimer E47/NeuroD1 bound to DNA (PDB code: 2QL2) (2).

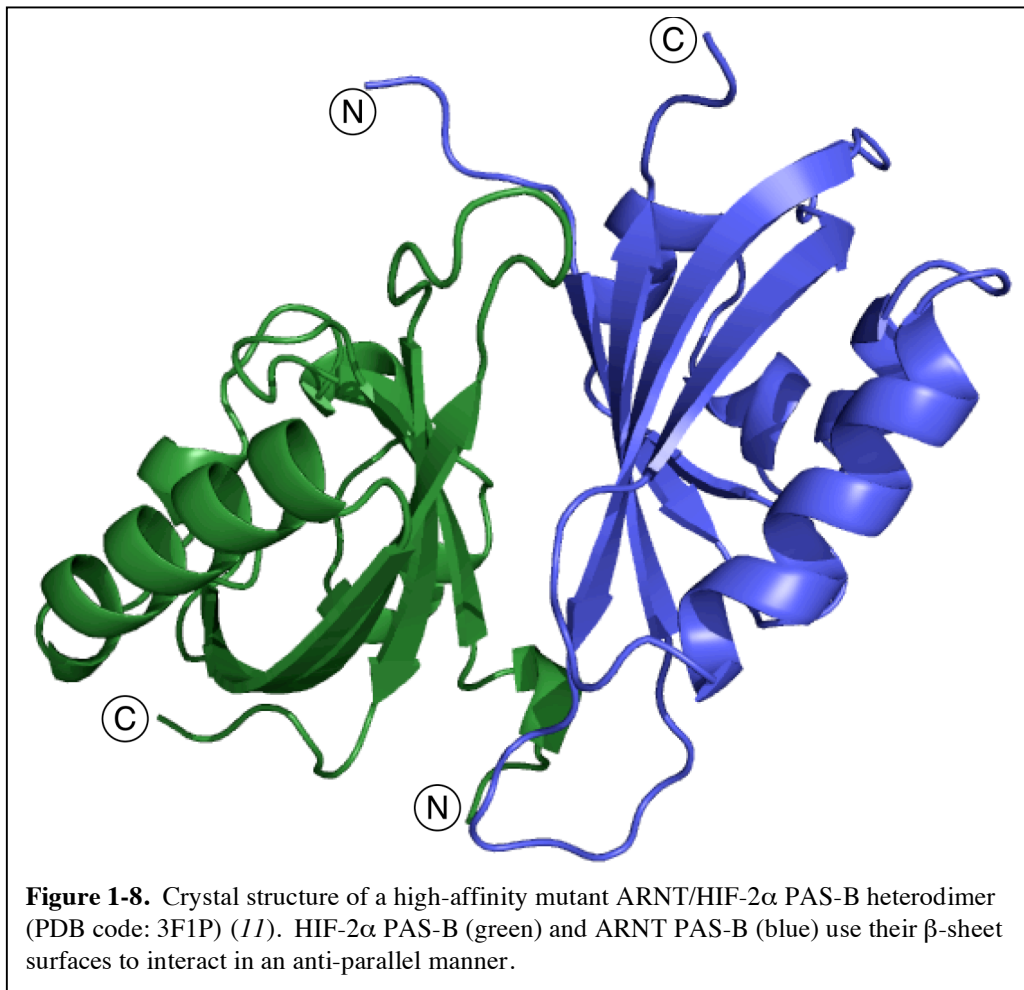
*i. The Aryl hydrocarbon Receptor Nuclear Translocator*

The Aryl Hydrocarbon Receptor Nuclear Translocator (ARNT) is a promiscuous bHLH-PAS protein that controls various biological pathways by forming heterodimeric transcriptional regulator complexes with several other bHLH-PAS proteins and is constitutively localized in the nucleus. Studies have identified three isoforms of ARNT, (ARNT, ARNT-2 and bMAL), which are greater than 30% homologous to each other. In addition, ARNT and ARNT-2 are

functionally exchangeable whereas brain and muscle aryl hydrocarbon receptor nuclear translocator ARNT-like (bMAL) interacts with a limited number of bHLH-PAS proteins, such as the circadian regulator Clock (58, 59). ARNT consists of an N-terminal nuclear localization sequence, which mediates nuclear localization through interactions with components of the nuclear pore targeting complex. Typical of members of this protein family, ARNT contains a bHLH and two PAS (PAS-A and PAS-B) domains (Figure 1-6). At the C-terminus, a transactivation domain has been identified, which interacts with the transcriptional co-activator CREB Binding Protein (CBP) and p300, as shown by a yeast 2-hybrid screen (60, 61). The NMR solution and X-ray crystal structures of ARNT PAS-B reveals a typical PAS domain fold, consisting of several  $\alpha$ -helices packed against one face of a five-stranded, anti-parallel  $\beta$ -sheet (Figure 1-7) (7, 11). Like many PAS domains, the  $\beta$ -sheet surface is critical for the function of the domain as ARNT uses this interface for homo- and heterodimerization with several other bHLH-PAS proteins (Figure 1-8) (7, 11). The potential for homodimerization of full length ARNT *in vitro* and *in vivo* has been repeatedly described to be critical for embryonic survival as seen by murine ARNT knock-out mutations (62, 63), although the biological relevance of this interaction is unclear (7, 55, 56). In contrast, the biological importance of ARNT heterodimeric complexes, including Aryl hydrocarbon Receptor, Hypoxia Inducible Factor, and

Single-minded, are well understood (26, 64). These heterodimeric complexes that form in response to changing environmental factors is crucial for cell survival.





*a. ARNT mediated pathways*

*i. Single-minded Protein*

A member of bHLH-PAS protein family that forms a heterodimeric transcriptional regulator complex with ARNT is the *Drosophila* Single-minded (dSIM) protein. Originally identified from *Drosophila melanogaster*, dSIM acts

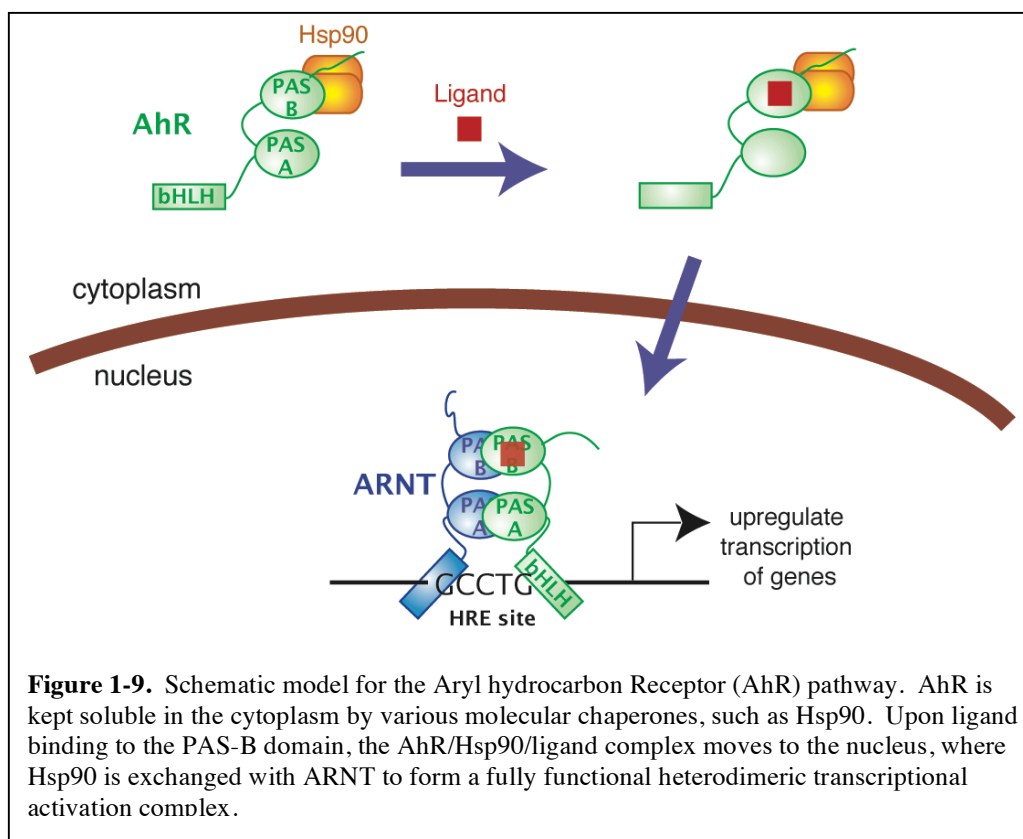
as a master regulator of midline cell development and is thought to be required for all developmental stages of the central nervous system midline cell lineage (21). dSIM contains a C-terminal transactivation domain and is a transcriptional activator when bound to ARNT (65). However, homologs from other organisms, including mouse, show a divergence in the sequence. mSIM1 and mSIM2 do not contain transactivation domains, and thus act as repressors when dimerized with ARNT by sequestering it away from other transactivating partner proteins (66, 67). In addition to indirectly repressing other proteins through competition for ARNT, mSIM2 also directly represses the transactivation domain of ARNT. Homozygous mutant (mSIM1<sup>-/-</sup>) mice exhibit defects in the terminal differentiation of certain neuroendocrine lineages and die perinatally (68). Similar neuroendocrine lineages are affected by loss of function of the ARNT homologue, suggesting their importance for cell development. In addition, SIM1 also regulates obesity as heterozygous alleles in mice and humans have resulted in profound early onset obesity and hyperphagia (54). Such results show the importance of selectively disrupting ARNT-mediated complexes.

#### *ii. The Aryl hydrocarbon receptor*

The Aryl hydrocarbon Receptor (AhR), a member of the bHLH-PAS family, mediates transcriptional responses to environmental pollutants by upregulating genes, such as cytochrome P450A1, that help metabolize toxins.

Exposure to one such pollutant, dioxin (2,3,7,8-tetrachlorodibenzo-p-dioxin, TCDD), causes a biochemical response resulting in severe wasting syndrome, epidermal hyperplasia, metaplasia, and tumor promotion (69, 70). Cells respond to dioxin and other structurally related halogenated aromatic hydrocarbons through the AhR, also known as the dioxin receptor. AhR consists of a N-terminal bHLH domain preceding two tandem PAS domains (PAS-A and PAS-B) and a C-terminal transactivation domain. In its non-DNA binding state, AhR is kept soluble by several molecular chaperones including heat shock protein 90 (Hsp90) in the cytoplasm (71, 72). Hsp90 is also thought to mask AhR's nuclear localization sequence keeping AhR in a ligand binding state. Upon ligand binding, the AhR/Hsp90 complex dissociates allowing AhR is translocated to the nucleus where it associates with ARNT (Figure 1-9) (70, 73). This association allows ARNT to form a heterodimeric transcriptional activation complex with AhR that binds to xenobiotic response elements (XREs) upstream of TCDD-responsive genes. One target by this complex is another bHLH-PAS protein, the aryl hydrocarbon receptor repressor (AhRR). AhRR also heterodimerizes with ARNT, forming a non-functional complex which binds to XRE's. This complex acts as a negative feedback loop to down-regulate the expression of AhR-ARNT target genes (74-76). Studies of AhR deficient mice suggest AhR may not only be responsive to environmental pollutants but also may be involved in liver development, suggesting that an endogenous ligand for this receptor may exist,

although one has not yet been identified (77-79). Regardless, it is important to understand the specific details of how PAS domains bind various ligands and form heterodimeric complexes in order to respond to environmental conditions.



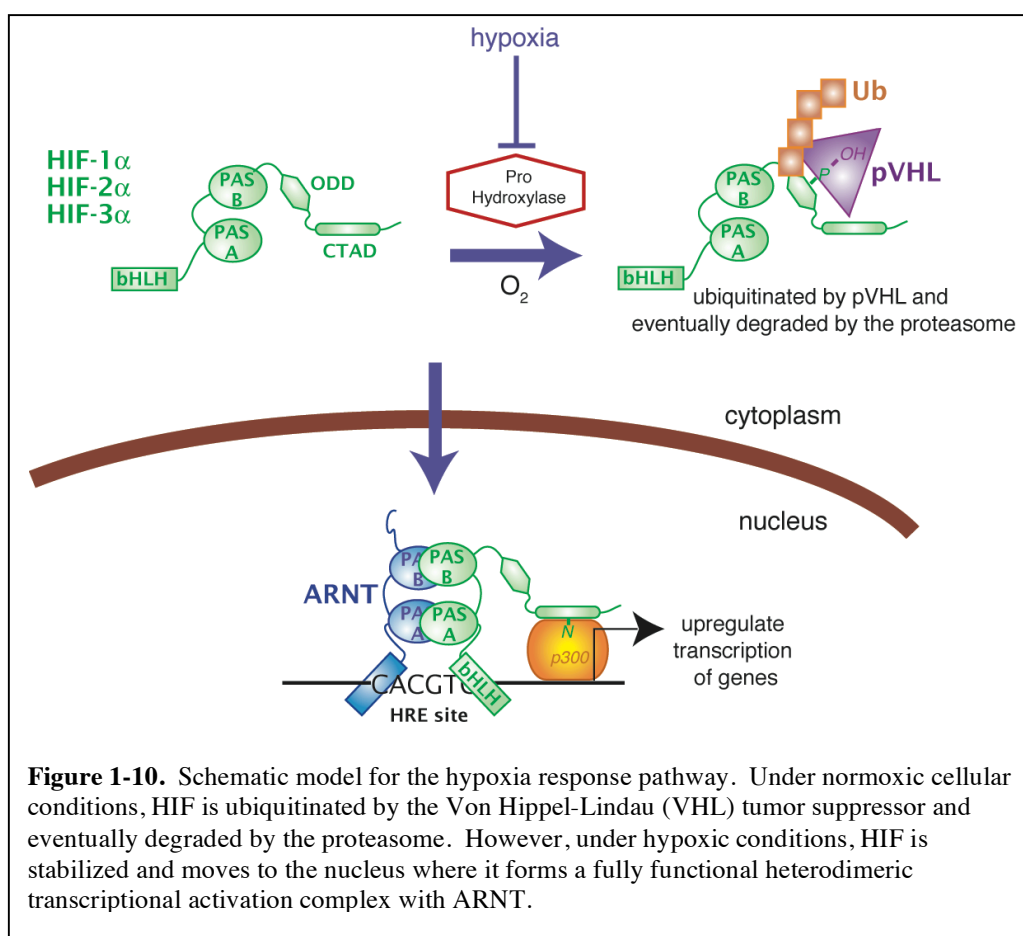
### *iii. The Hypoxia Inducible Factor*

Another bHLH-PAS domain that mediates cell responses to environmental stimuli through the formation of a heterodimeric transcriptional regulator complex is found in the hypoxia response pathway. This pathway allows mammalian cells to maintain oxygen homeostasis. During high oxygen tension in the cell



(hyperoxia), reactive oxygen intermediates are generated that can lethally damage cell membranes and DNA (54). Meanwhile, cells under low oxygen tension (hypoxia), cannot maintain essential cellular functions due to insufficient levels of ATP. Responses to hypoxia in mammalian cells are mediated by proteins called the hypoxia inducible factors, of which three known isoforms exist (HIF-1 $\alpha$ , HIF-2 $\alpha$ , and HIF-3 $\alpha$ ). HIF $\alpha$  contains a bHLH DNA binding and dimerization domain, two tandem PAS domains (PAS-A and PAS-B), a C-terminal transactivation domains (C-TAD), and an oxygen degradation domain (ODD). During normoxia, HIF- $\alpha$  is hydroxylated at two proline residues (Pro402 and Pro564) within the ODD. The Von Hippel-Lindau tumor suppressor E3 ligase complex recognizes these hydroxylated residues (80-82), resulting in ubiquitination and subsequent degradation of HIF- $\alpha$  by the proteasome (83-85). In addition, one asparagine residue (N803) found within the C-TAD domain is hydroxylated by an aspararginyl hydroxylase factor, which prevents the recruitment of necessary co-activator CBP/p300 (86, 87). Under hypoxic conditions, proline hydroxylases are deactivated, stabilizing HIF- $\alpha$  which moves into the nucleus and bind its heterodimeric partner ARNT, to form the active HIF complex (Figure 1-10) (88, 89). This complex binds hypoxia response elements (HREs), recruiting transcriptional co-activators, such as CBP/p300, which activates transcription of over 70 genes. These genes are essential for switching the cell from oxidative phosphorylation to glycolytic anaerobic metabolism as

well as promoting new vasculature growth, and triggering a countdown to apoptosis. The hypoxic state is found in many diseases, such as cancer and ischemias, emphasizing the importance of understanding PAS-mediated protein:protein associations in response to environmental conditions (62, 90, 91).



### *b. Biological Significance of ARNT*

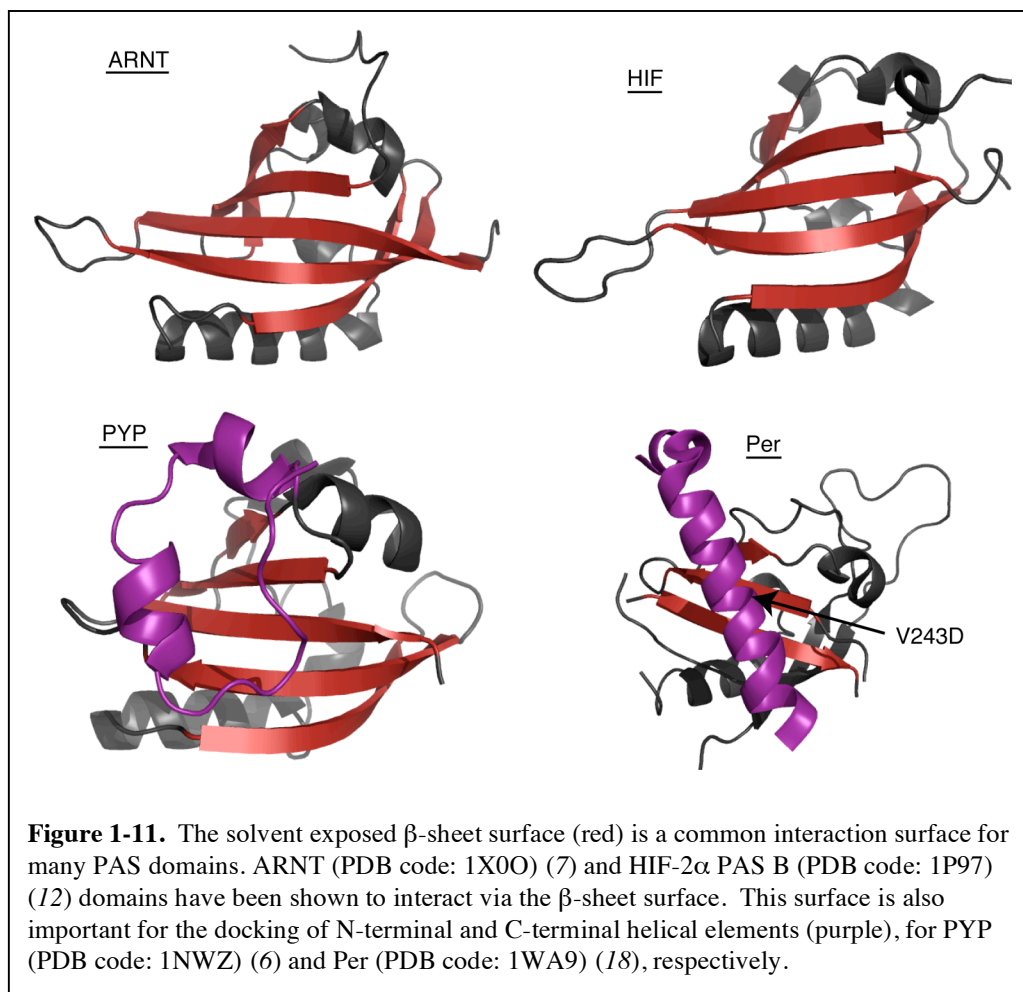
As previously mentioned, ARNT mediates a number of different pathways by forming heterodimeric complexes in responses to various environmental

factors. Misregulation of these signaling pathways by inappropriate up-regulation or activation of transcription factors involved, can lead to a variety of human diseases, such as helping cancer propagate (54). This correlation is clearly seen in the AhR and HIF pathways. For AhR, exposure to environmental pollutants lead to inappropriate activation of gene transcription, leading to liver and skin tumor formation whereas, for HIF, a hypoxic state activates gene transcription promoting new vasculature growth, leading to the spread of brain and breast tumors (92, 93). In addition, over-expression of ARNT by itself has also been correlated with cancer, as evidenced by the high expression levels of ARNT in blood cells of patients with acute myeloid leukemia (94, 95).

Several other PAS containing proteins are critical for a variety of biomedically-relevant signaling pathways. One example, human ether-a-go-go-related gene (HERG) is a voltage-dependent potassium channel found in the human heart and forms a homotetramer with a central pore. This protein contains a PAS domain at the N-terminus, which regulates the closing of the channel with the proper kinetics (96, 97). Interestingly, several human polymorphisms in the HERG PAS domain correlate with long QT syndrome, a genetic or pharmacologically induced condition causing familial cardiac arrhythmia and sudden death (98, 99). These PAS mediated diseases show the importance of further understanding the mechanism by which these proteins interact with each other and ultimately block disease progression.

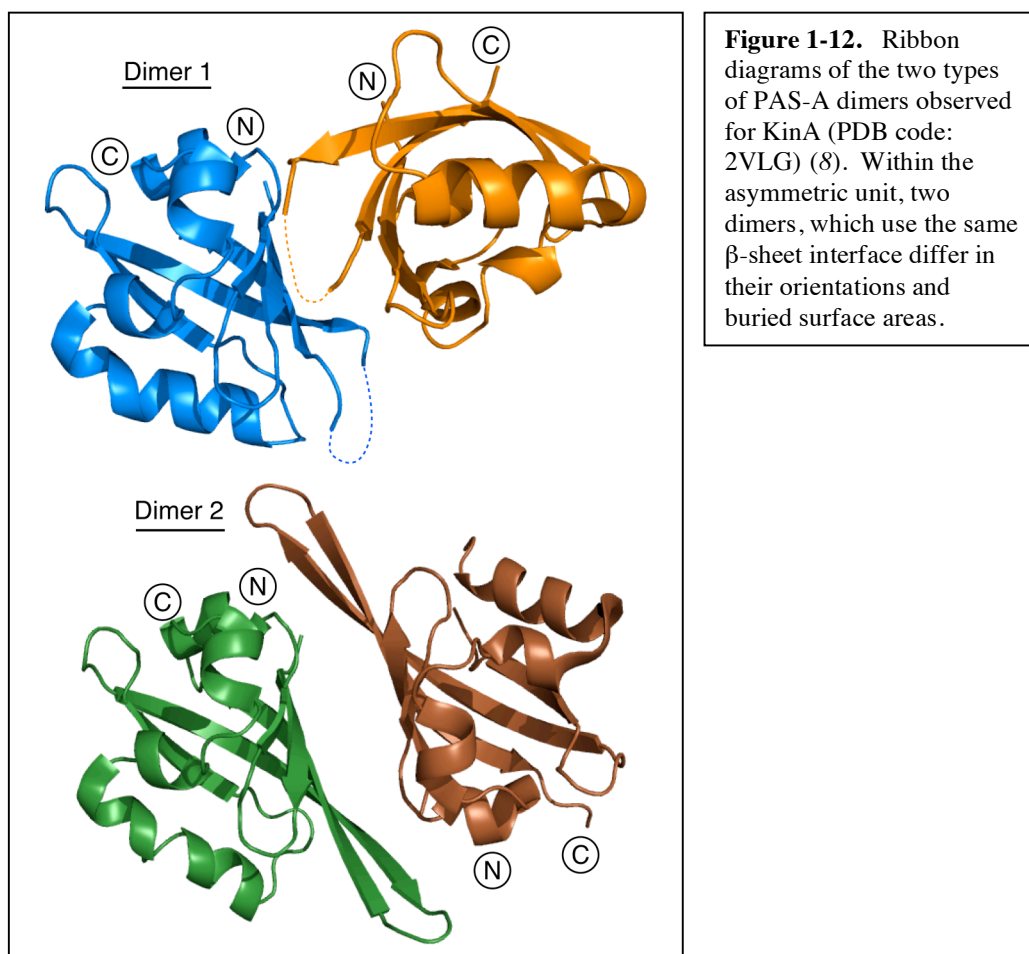
*C. One common  $\beta$ -sheet to bind various ligands/proteins*

More broadly, the  $\beta$ -sheet surface of many PAS domains has been found to be important for a wide array of inter- and intramolecular interactions (Figure 1-11). Using this common interface, PAS domains bind various cofactors such as: (a) heme covalently linked to FixL (*1*); (b) 4-hydroxycinnamic acid covalently attached to photoactive yellow protein (*39-41*); and (c) FMN covalently or non-covalently bound to LOV domains (*44, 100*). For LOV domains, it has been accepted that the cofactor binds on the internal surface of the  $\beta$ -sheet, likely causing a conformational change directly impacting the  $\beta$ -sheet. Assuming that PAS proteins bind other cofactors in a similar manner as observed in LOV domains, it is reasonable to assume cofactor binding causes similar conformational changes across the  $\beta$ -sheet surface.



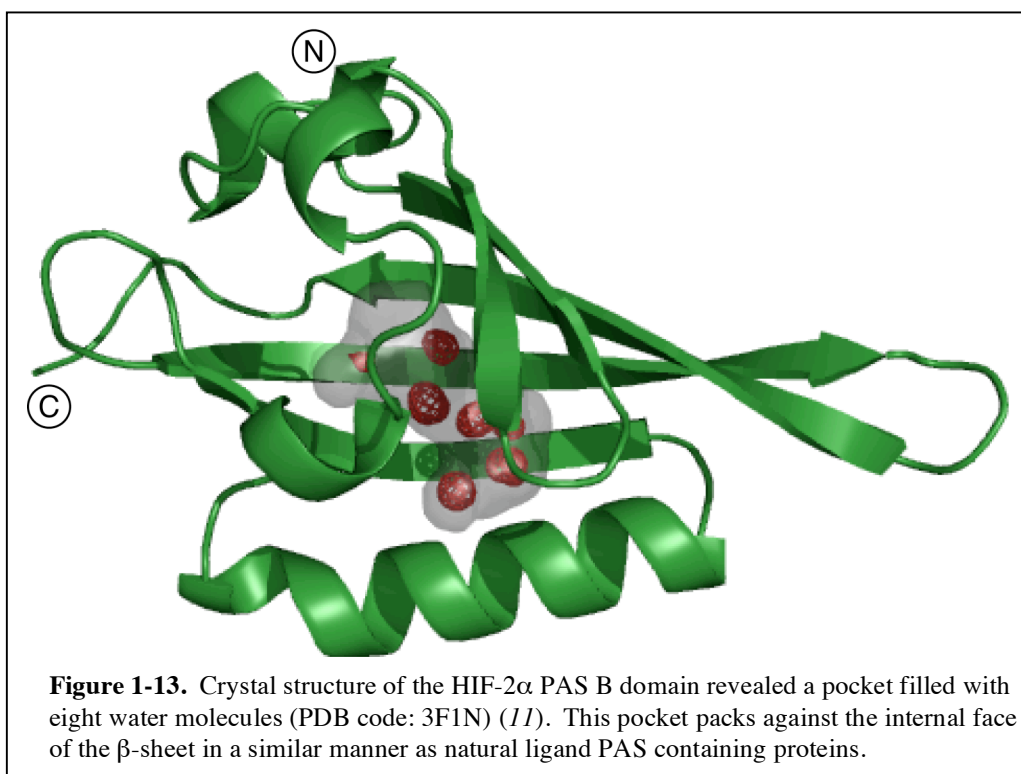
The wide variety of cofactor-binding properties, as well as, the multiple modes of binding to other PAS domains suggests this domain fold must exhibit a large conformational flexibility. For example, the plasticity of PAS domains to accommodate different modes of binding is clearly demonstrated by the *Bacillus subtilis* KinA sporulation kinase and *Bradyrhizobium japonicum* FixL sensor kinase. The bacterial KinA sensor kinase contains three tandem PAS domains

and is responsible for regulating the initiation of sporulation in response to nutrient deprivation (*101, 102*). Intriguingly, the X-ray crystal structure of the N-terminal PAS domain within KinA resulted in two dimers in the asymmetric unit with different tertiary and quaternary structures (Figure 1-12) (*8*). While both dimers utilized the same  $\beta$ -sheet surface, they differed in their orientations and buried surface areas, thus suggesting an inherent flexibility of this domain. Similarly, the PAS domain from the oxygen sensor kinase, FixL, also resulted in two different dimer orientations when the crystallization conditions varied (*103*). While the biological importance of these two complexes remains unclear, it is important to note these domains to be quite flexible to allow such distinct interactions to occur.



Another example of the flexibility of PAS domains is demonstrated by hypoxia inducible factor (HIF) protein for which no known endogenous ligand has been reported. As previously discussed, HIF dimerizes with ARNT by virtue of its two PAS domains to respond to hypoxic conditions. An X-ray crystal structure of a mutant, high-affinity HIF-2 $\alpha$  PAS-B to ARNT PAS-B heterodimer revealed a large (approximately 290 Å<sup>3</sup>), water-filled cavity within the core of HIF (Figure 1-13) (11). Eight water molecules in this pocket pack against the  $\beta$ -

sheet surface in a similar manner as other PAS domain cofactors. In conjunction with this discovery was the identification of small-molecule ligands that bind in the core of the HIF-2 $\alpha$  PAS-B domain and disrupt heterodimerization to ARNT PAS-B *in vitro* (11). The ability of the HIF PAS-B domain to bind ligands within a completely internal cavity typically occupied by water molecules provides further evidence that some PAS domain can be inherently flexible.



Lastly, additional data supporting the malleability of PAS  $\beta$ -sheets comes from single-molecule atomic force microscopy experiments of photoactive yellow protein (PYP). Under mechanical stress, the A $\beta$ -strand slips towards the N-terminus (104, 105). Interestingly, even though there is a  $\beta$ -strand slip, the rest of



the protein remains intact. More broadly, these results suggest that the PAS fold is built to withstand some conformational flexibility when binding various ligands or other proteins and that this flexibility may be needed for the diverse set of functions PAS domain conduct.

#### *D. Questions addressed in this thesis*

Thus, the plasticity of the PAS fold may be related to how ARNT PAS-B utilizes its  $\beta$ -sheet to interact with multiple binding partners with low sequence identity. To address this question, we generated a series of point mutations on the solvent exposed  $\beta$ -sheet surface of ARNT PAS-B and examined the structural and functional effects on heterodimer formation. Intriguingly, despite the high stability this domain exhibits, we found this specific fold is fragile to the point that a single point mutation enables the domain to slip into a different, non-functional  $\beta$ -strand register. This discovery reflects an inherent flexibility of the ARNT PAS-B domain and provides further evidence of the conformational flexibility of PAS domains as a whole.

## II. Conformational flexibility observed in Nature

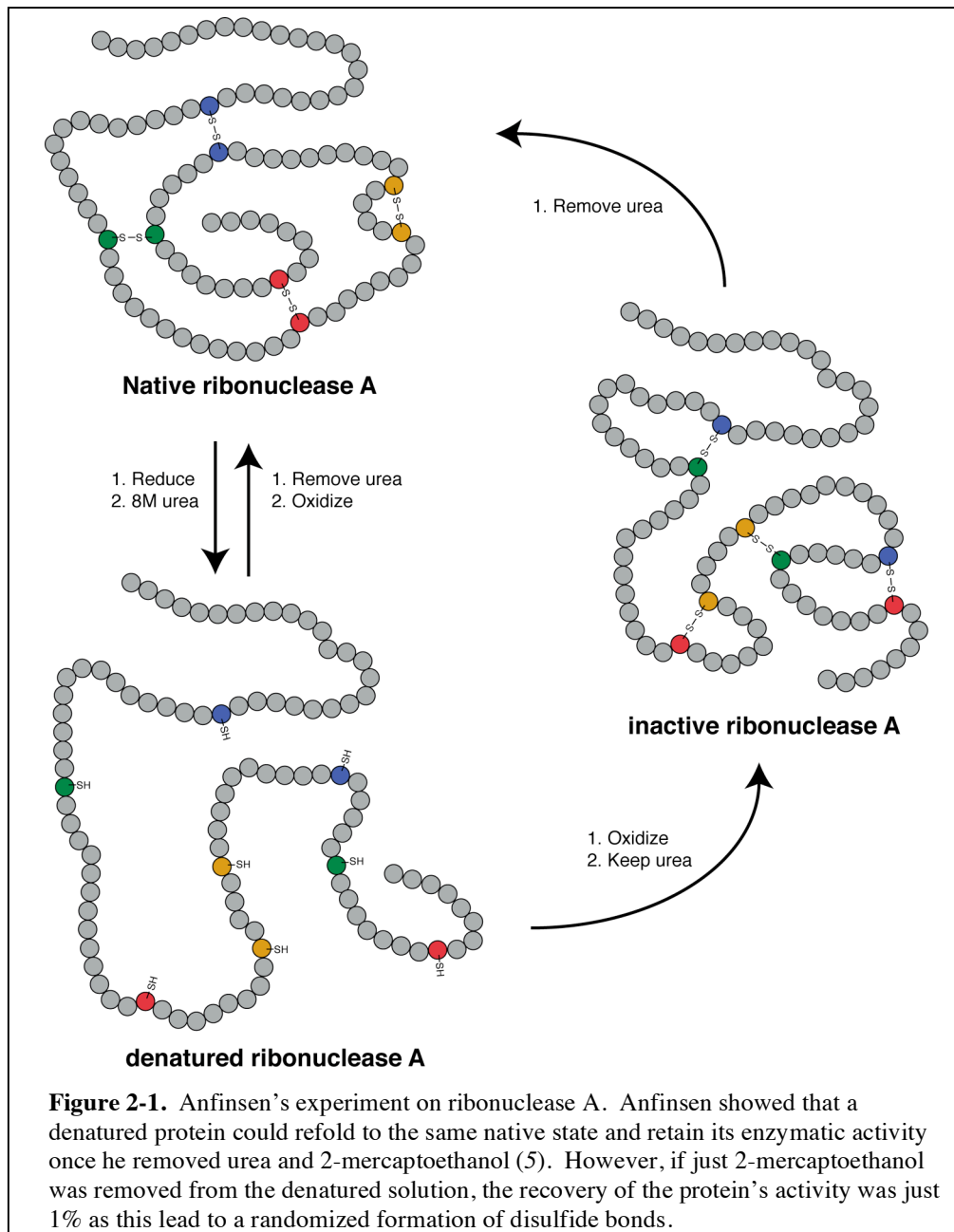
### *A. The protein folding problem*

A fundamental, unresolved problem in structural biology, despite extensive research, is to understand how the amino acid sequence of a protein folds into its functional three-dimensional structure (5). Precise understanding of this process has eluded scientists for years and has been termed “The Protein Folding Problem”. It consists of three closely related issues: (a) the folding code: the thermodynamic question as to how interatomic forces help determine the native structure for a specific amino acid sequence; (b) protein structure prediction: the computational problem of how to take an amino acid sequence and predict the native structure; and (c) the folding rates: the kinetic question as to how proteins fold so quickly on a biologically relevant timescale considering many the degrees of freedom that are possible (106, 107).

### *B. One sequence equals one structure*

One of the greatest achievements made towards the protein folding problem was work done by Christian B. Anfinsen. He found the connection between the amino acid sequence of a protein and its functional three-dimensional structure, a discovery that won him the 1972 Nobel Prize in Chemistry (5). Work prior to Anfinsen suggested that protein folding always depended on the kinetic

folding pathway. Anfinsen suggested otherwise by showing ribonuclease could refold to the same native state and retain its enzymatic activity after being denatured (Figure 2-1) (*108-110*). From his studies of ribonuclease, Anfinsen concluded that the amino acid sequence alone contains all the necessary information to fold into its fully active tertiary structure (*5*). With this milestone discovery, the field of protein folding expanded as the idea that the native structure would remain the same regardless if it was synthesized on a ribosome or refolded in a test tube (*107*). Amazingly, after 40+ years of protein structure work, thousands of structures have been solved and Anfinsen's hypothesis of one sequence equals one structure still holds true for many structures, with occasional exceptions.



Advancements in the protein structure field have suggested that the Anfinsen paradigm that has been the foundation of protein structure for many

years should be a guideline for protein structure work and not taken for granted, as a few exceptions have occasionally been found to this hypothesis. In a broader context, the idea that one sequence can adopt two conformations has been observed in several biological systems, where they can switch between a folded and a (partially) unfolded state (e.g. drk SH3 domain (*111*) and interleukin-2 (*112*)), an inactive and an active native state (e.g. serpins (*113*), Mad2 spindle checkpoint protein (*114*, *115*), and prions (*116*)) or two distinct active states with different functional consequences (e.g. lymphotactin (*117*)). While it is known that many proteins show conformational changes in response to ligand or protein binding or even in the absence of a binding event, these are not considered to contradict Anfinsen's hypothesis. The rare exceptions to Anfinsen's hypothesis come from those proteins that in the absence of a binding event can adopt different tertiary folds that result from multiple energetically favorable minima. Therefore, to consider a protein an exception to Anfinsen's hypothesis, it must conflict with two conditions. First, it must have a structure that is unique, in that no other configurations may exist with a comparable free energy minimum. Second, the protein must be quite stable, such that any small change in the surrounding environment should not give rise to changes in configuration. Therefore, the free energy landscape around the native state must be steep, with the native state at the bottom as a means to provide stability. Only those proteins,

which contradict both of these conditions, will be considered an exception to Anfinsen's postulate.

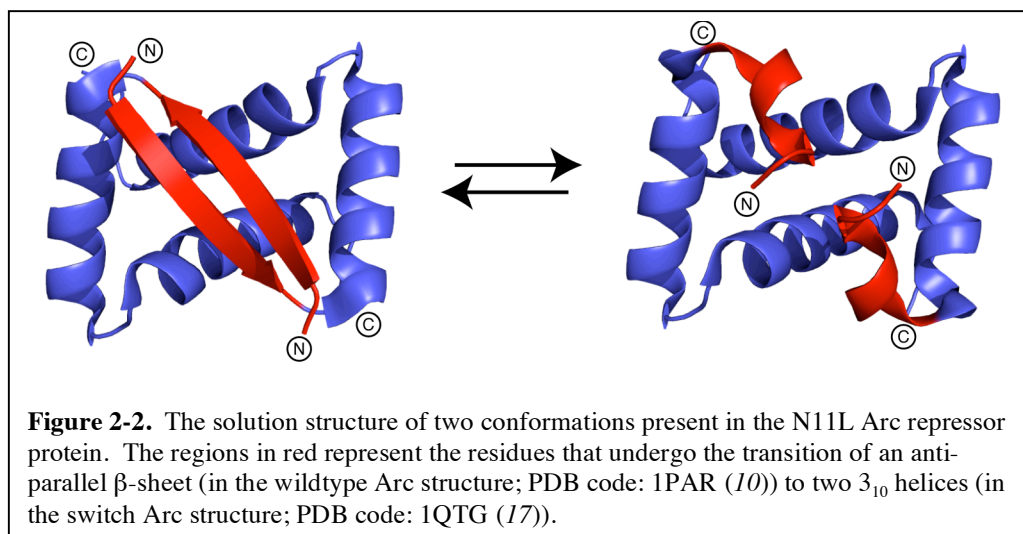
*i. Challenges to Anfinsen's postulate*

An early challenge to Anfinsen's theory of protein folding came in the 1980's with the discovery of proteins that helped in the folding process. These small proteins, called molecular chaperones, serve to help other proteins fold and assemble. The first discovered chaperone, nucleoplasmin, helped fold histones and DNA assemble into a nucleosome (118). For this reason, molecular chaperones were initially thought to directly fold proteins, thereby disproving Anfinsen's hypothesis. However, upon further studies, chaperone proteins were found to safely sequester the newly synthesized polypeptide chains and assembled subunits from improperly folding and aggregating into a nonfunctional state. This result eventually led to the idea that chaperones act as a cage to protect the protein from aggregation while it properly folds and is referred to as an Anfinsen cage (119, 120). After years of research, chaperones were found not to affect the final structure of the protein and the idea that molecular chaperones violated Anfinsen's postulate was dismissed (121).

*a. Arc repressor mutant: switch Arc*

One such protein found to adopt more than one stable conformation was a mutant form of the *Salmonella* Arc repressor. This protein is a dimeric sequence-specific, DNA binding protein that is involved in the switch between lysis and lysogeny of *Salmonella* bacteriophage P22 (122-124). The solution structure reveals a strong intertwined dimer with an antiparallel  $\beta$ -sheet, which is the DNA-binding motif formed by residues of both monomers (124). A mutagenesis study of these  $\beta$ -strands found a simple interchange of Asn11 and Leu12 residues resulted in a significant chemical shift differences for those residues on the  $\beta$ -sheet (17). These large differences in chemical shift suggested the  $\beta$ -sheet is replaced with a different type of secondary structure. By switching a surface residue with an adjacent core residue the  $\beta$ -strands are replaced with two right-handed  $3_{10}$  helices (Figure 2-2). This new conformation, termed “switch Arc” is equally stable as the wild-type Arc, yet no longer binds DNA (17, 125). Intriguingly, mutating just N11L allowed Arc to fold into either conformation, with a 67:33 (helix:sheet) equilibrium at 25°C, and could rapidly interconvert between these two structures with an exchange rate of 2,200 per second, resulting in significant line broadening for residues on the  $\beta$ -sheet surface (126). In addition, when DNA was added to the Arc N11L mutant, DNA bound only the wild-type conformation, which shifted the equilibrium almost completely towards this binding conformation (126). The discovery of this distinct conformation

nearby in sequence space suggested that structural plasticity may be more common than previously thought and that different protein structures could lead to certain types of disease, as thought to be the case with the prion protein PrP (127, 128).



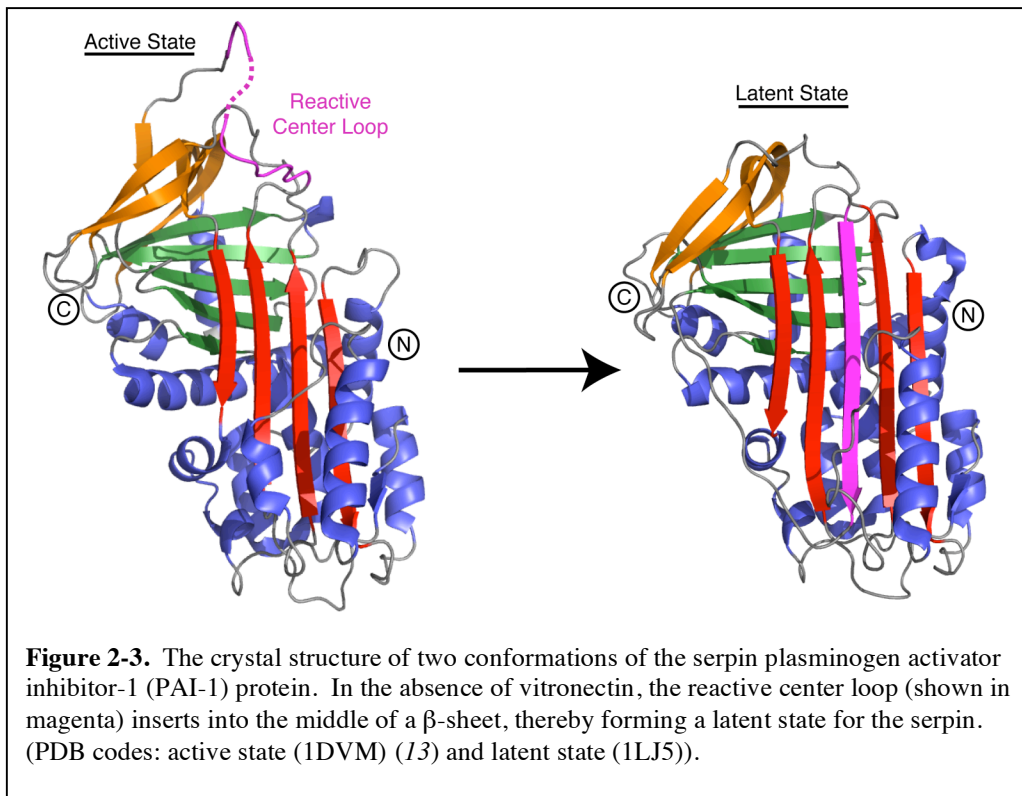
### *b. Serpins*

A family of proteins that can adopt more than one conformation is serine protease inhibitors, termed serpins. Members of this functionally diverse superfamily rely on a unique structural rearrangement to inhibit target proteases (129, 130). Inhibitory serpins use an exposed loop, known as the reactive center loop (RCL) to interact with a target protease. Upon binding, serpins transition from a stressed (S) to relaxed (R) conformation by inserting the RCL into the middle of a  $\beta$ -sheet, forming an extra  $\beta$ -strand (131, 132). This new conformation remains covalently bound to the target protease with a half-life of weeks (133).



After cleavage, this conformation is greatly stabilized compared to the native conformation, resulting in a metastable folding intermediate (134).

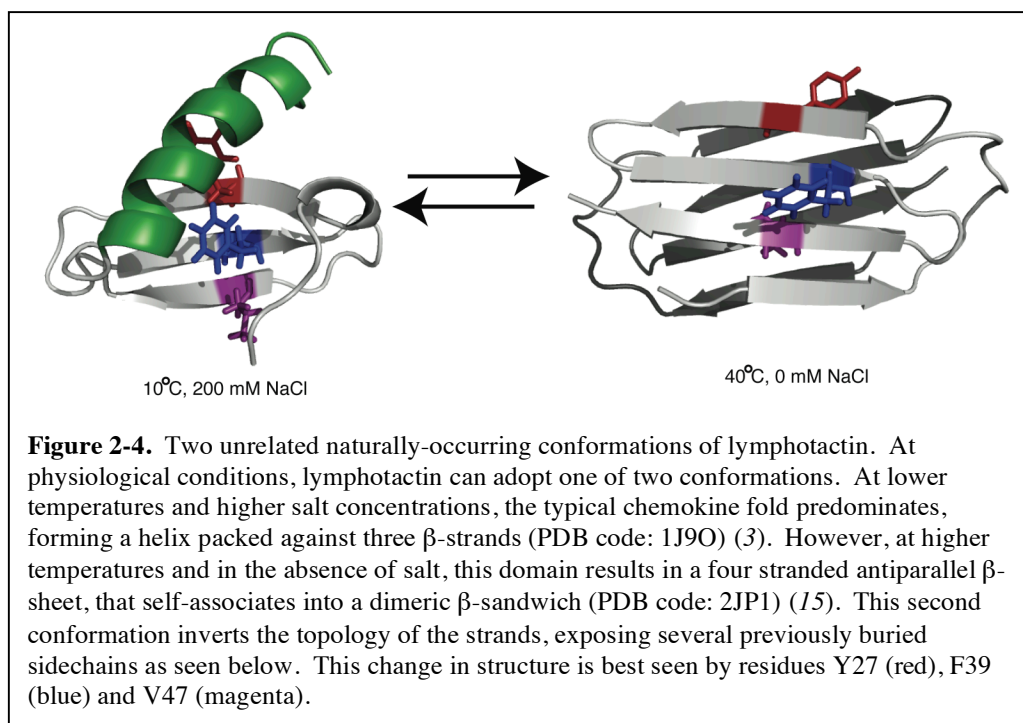
The conformational flexibility exhibited by serpins is best shown by latent serpins that can spontaneously undergo the S to R transition, forming a latent state (135). Serpins in the latent state are unable to interact with different proteases and thus provide a means to modulate its activity. For example, structural studies of the serpin plasminogen activator inhibitor-1 (PAI-1) have shown the protein is able to adopt either the native inhibitory or latent, inactive folds unless it is bound to the cofactor vitronectin (Figure 2-3) (136). This intrinsic flexibility of serpins to adopt multiple folds allows for precise control of active-state protein levels in the cell, thus providing another exception to Anfinsen's hypothesis.



### *c. Lymphotactin*

The ability of a native, wildtype protein to interconvert between two different, naturally-occurring folds with distinct functions by changing only the solution conditions make lymphotactin a notable exception to Anfinsen's paradigm. Lymphotactin is a small-secreted protein that induces chemotaxis by binding cell surface glycosaminoglycans and activating certain G-protein-coupled receptors (GPCRs) (137). Initial structural studies of lymphotactin near physiological solution conditions (37°C, 150 mM NaCl) yielded two, equally stable conformations of the domain in equilibrium that slowly exchanged on the

100ms timescale (15, 117). Intriguingly, changing the solution conditions by varying the temperature and salt concentrations varied the relative populations between conformations. Under certain solution conditions (10°C, 200 mM NaCl), lymphotactin adopted the conserved chemokine fold, which is designated Ltn10. However, at 40°C in the absence of salt, an alternative conformation, termed Ltn40, predominates. Solution NMR studies yielded stark structural differences between the two conformations (Figure 2-4). The structure of Ltn10 consists of a N-terminal  $\alpha$ -helix packed against a three-stranded antiparallel  $\beta$ -sheet. However, upon converting to Ltn40, the N-terminal  $\alpha$ -helix unfolds and a new  $\beta$ -strand is formed, generating a four-stranded antiparallel  $\beta$ -sheet, that self-associates into a dimeric  $\beta$ -sandwich. This new Greek-key motif is accompanied by a one-residue slip in register of strands  $\beta$ 1 and  $\beta$ 3, which inverts the topology of the strands, burying several previously exposed sidechains and generates an entirely new hydrogen bonding network. In addition, the functional difference of these two conformations is just as significant as the structural differences shown as only the Ltn10 fold adheres to cell-surface glycosaminoglycans, and only Ltn40 binds heparin, a polysaccharide found in the extracellular matrix. These data show that in order for lymphotactin to successfully complete its biological function, it must convert between these conformations, implying a switching mechanism for regulation.



Lymphotactin presents an example of how nature can efficiently use a single polypeptide chain and generate multiple folds of a domain, which can be used as a switching mechanism for regulating various pathways. This discovery leads to the idea that other proteins may adopt multiple structures in order to keep a protein inactive until needed by the cell. In addition, the identification of alternative, stable conformations just one point mutation away in sequence space may give rise to more disease states than previously thought. We suggest that the barriers which establish normal Anfinsen sequence:structure relationships can be fragile for a wider variety of natural proteins, or trivial modifications thereof, as

our studies of the ARNT PAS-B indicate that one point mutation allow the same primary sequence to explore multiple stable folded conformations.

### III. A second conformation of ARNT PAS-B nearby in sequence space

(published as ARNT PAS-B has a fragile native state structure with an alternative

$\beta$ -sheet register nearby in sequence space;

Matthew R. Evans, Paul B. Card, and Kevin H. Gardner

*Proc Natl Acad Sci*, **2009**, 106(8): 2617-2622.)

#### *Abstract*

The aryl hydrocarbon receptor nuclear translocator (ARNT) is a basic helix-loop-helix Period/ARNT/Single-minded (bHLH-PAS) protein that controls various biological pathways as part of dimeric transcriptional regulator complexes with other bHLH-PAS proteins. The two PAS domains within ARNT, PAS-A and PAS-B, are essential for the formation of these complexes, as they mediate protein/protein interactions via residues located on their  $\beta$ -sheet surfaces. While investigating the importance of residues in ARNT PAS-B involved in these interactions, we uncovered a point mutation (Y456T) on the solvent-exposed  $\beta$ -sheet surface that allowed this domain to interconvert with a second, stable conformation. While both conformations are present in equivalent quantities in the Y456T mutant, this can be shifted almost completely to either endpoint by additional mutations. A high-resolution solution structure of a mutant ARNT PAS-B domain stabilized in the new conformation revealed a three-residue slip in register and accompanying inversion of the central I $\beta$ -strand. We have demonstrated that the new conformation has over 100-fold lower *in vitro* affinity

for its heterodimerization partner, HIF-2 $\alpha$  PAS-B. We speculate that the pliability in  $\beta$ -strand register is related to the flexibility required of ARNT to bind to several partners, and more broadly, to the abilities of some PAS domains to regulate their activities in response to small molecule cofactors.

### *A. Introduction*

It is well understood that cell survival depends on the ability to detect and respond to fluctuations in the environment. Environmental factors, such as redox potential, blue light, and oxygen levels, trigger changes in signal transduction pathways that rely on protein-protein interaction domains. One family of such domains are the Per-ARNT-Sim (PAS) domains (20), named after the three proteins where they were first identified: Period (Per), Aryl hydrocarbon Receptor Nuclear Translocator (ARNT), and Single-minded (Sim) (21). Since being recognized in these three eukaryotic transcription factors, PAS domains have been found in proteins from all three kingdoms of life that perform functions as diverse as DNA binding, phosphorylation, ion conduction and others (20).

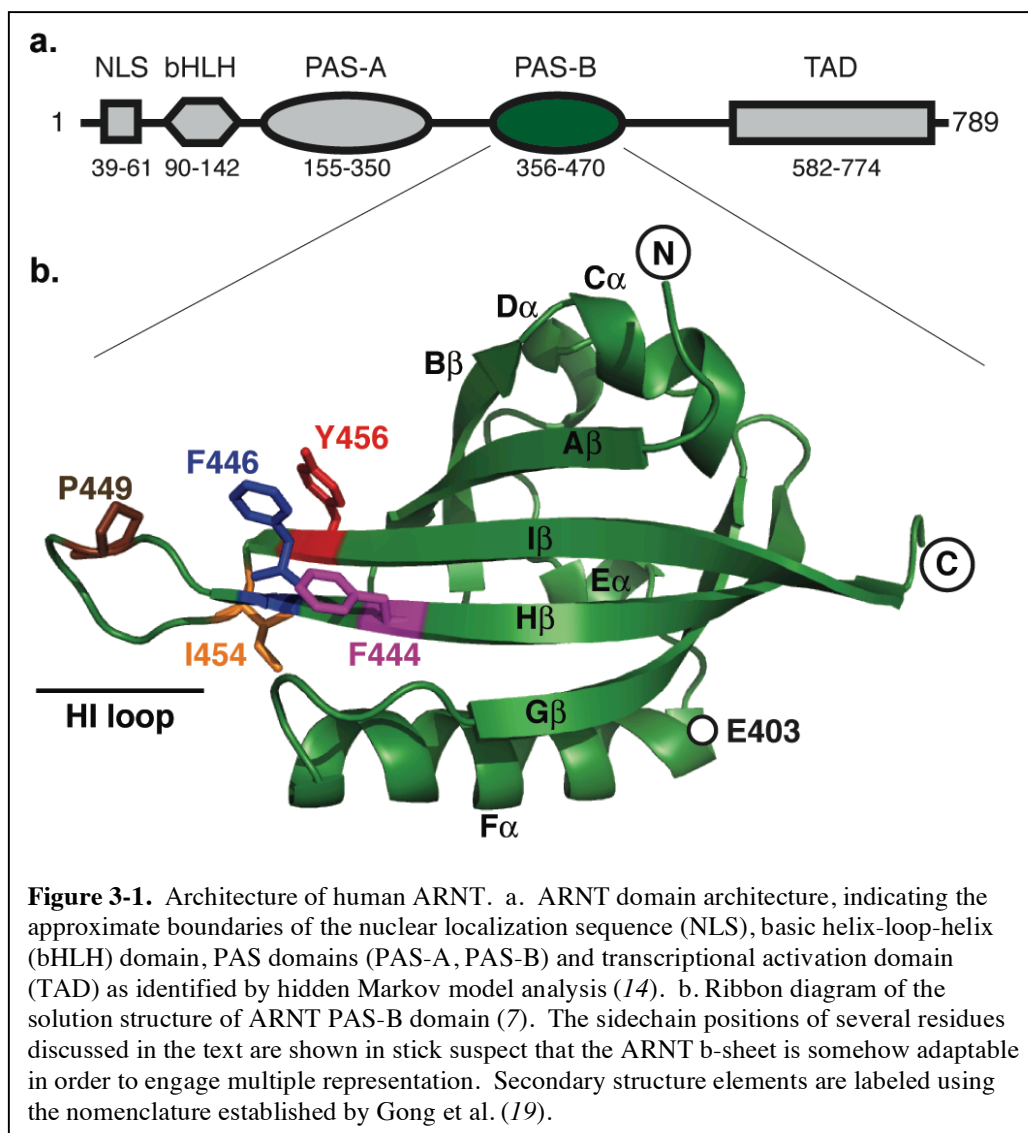
Despite these differences, PAS domains share a common role as protein:protein interaction domains, binding PAS or non-PAS partners either *in cis* or *in trans*. To achieve this, PAS domains have chiefly conserved a common structure composed of mixed  $\alpha/\beta$  fold of approximately 110-140 amino acids featuring several  $\alpha$ -helices packed against one side of a five-stranded, anti-

parallel  $\beta$ -sheet (20). This  $\beta$ -sheet plays a particularly critical role in PAS structure and function, as its solvent-exposed side is the binding site for a wide range of PAS-bound protein targets (12). Notably, some PAS domains control the binding affinity of this interface *via* changes in the conformation or occupancy of environmentally sensitive cofactors bound on the opposite, internal side of the sheet. This principle is exemplified by a group of PAS-based photosensors that bind flavin chromophores in their cores. These cofactors are non-covalently bound in the dark, but form a protein-flavin bond when exposed to blue light (44), generating conformational changes within the domain. As shown in the *A. sativa* phototropin1 LOV2 domain (28) and the *N. crassa* Vivid protein (138), this change is transmitted through the  $\beta$ -sheet, displacing or rearranging other protein elements bound to the exterior.

While these data suggest that PAS domain  $\beta$ -sheet surfaces are inherently flexible in ligand-regulated PAS domains, it is unclear if this holds for constitutively active PAS domains lacking internal cofactors. An example of such a domain is provided by ARNT, one of the three founding members of the PAS family (21). First identified as an essential component in the xenobiotic-sensing Aryl Hydrocarbon Receptor (AHR) pathway (24), ARNT also plays a key role in developmental processes and response to environmental factors including hypoxia (66, 67, 139, 140). ARNT is a member of the bHLH-PAS family of eukaryotic transcription factors, named for the presence of a N-terminal basic helix-loop-



helix (bHLH) DNA binding domain and two PAS domains (PAS-A, PAS-B) (Figure 3-1). ARNT constitutively dimerizes with other bHLH-PAS family members which themselves are regulated, including AHR and hypoxia inducible factors (HIFs). PAS domains are essential for stable heterodimer formation, as supported by biochemical studies that show that deletion of PAS domains from ARNT (64, 141) or its partners (*e.g.* HIF-2 $\alpha$  (26)) impairs the formation and activity of regulatory complexes. Biophysical studies of PAS domains from HIF-2 $\alpha$  and ARNT have shown that the C-terminal PAS-B domains from these proteins bind each other via an antiparallel  $\beta$ -sheet: $\beta$ -sheet packing (7). Using structure-based point mutations in these  $\beta$ -sheets, we have demonstrated that this interaction is relevant for full-length proteins in cells (26). A broader survey of PAS-mediated protein-protein interactions further supports the role of this  $\beta$ -sheet surface as a general site for PAS domains to interact with a wide variety of intra- and intermolecular partners (12).



Assuming that ARNT PAS-B binds other bHLH-PAS proteins similarly as observed in the ARNT/HIF-2 $\alpha$  complex, it is reasonable to binding partners with low sequence identity (approximately 10% among AHR, three HIF $\alpha$  isoforms and Sim). To test this, we generated a series of point mutations in the exposed  $\beta$ -sheet

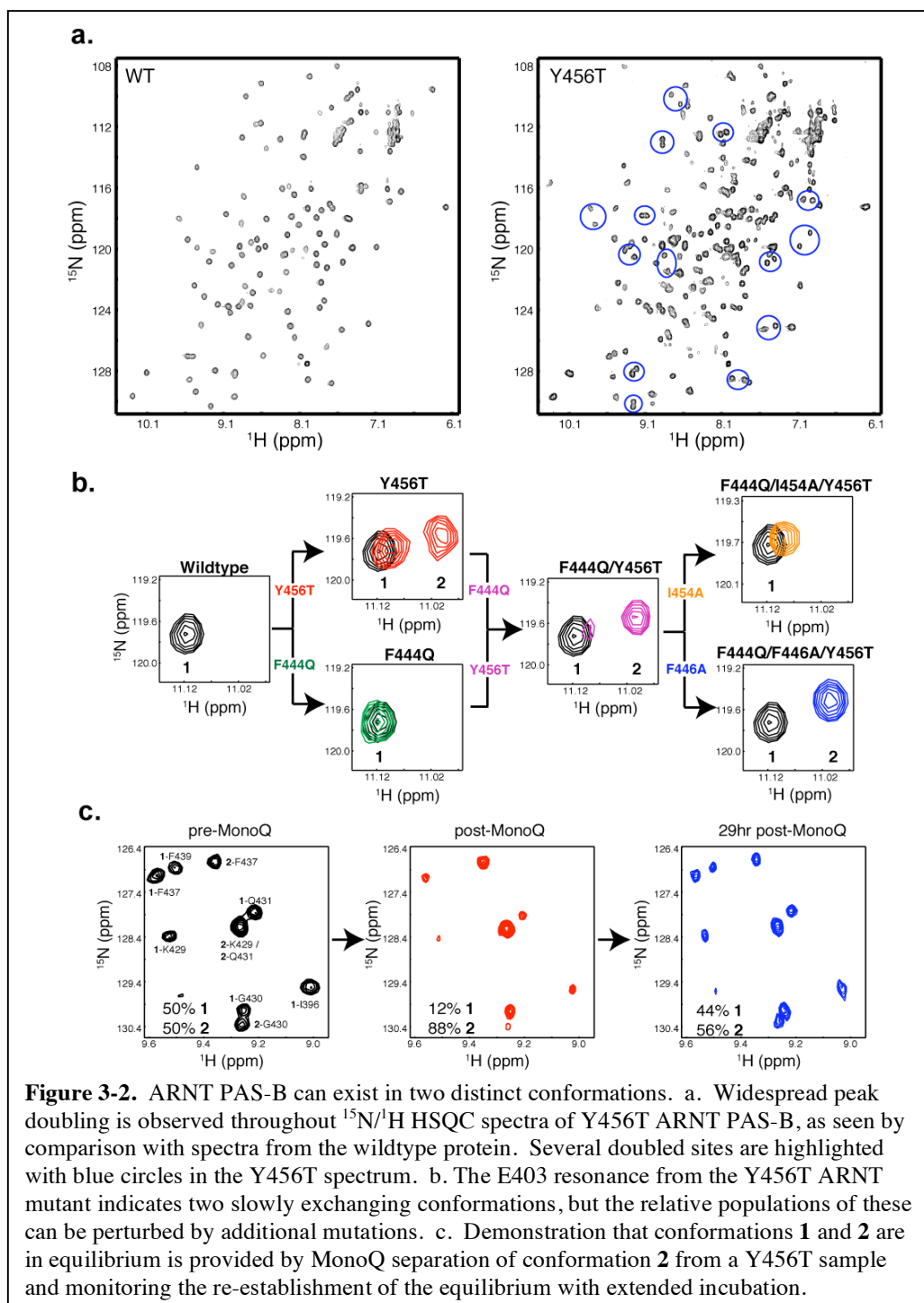
surface of ARNT PAS-B and examined the structural and functional effects on heterodimer formation. Strikingly, we discovered a dramatic demonstration of the flexibility possible in this domain as we exposed a new conformation that coexisted in approximately equimolar concentrations with the wild type conformation. The equilibrium between the wild type and the new conformation was affected by additional mutations, providing us with a library of mutants that span a wide range of populations between the two. A high-resolution solution structure of the new conformation reveals that it was achieved by slipping the central I $\beta$  strand by three residues, inverting the topology of this essential protein/protein interface. Accordingly, we found that the new conformation bound one of its heterodimerization partners, HIF-2 $\alpha$  PAS-B, with over 100-fold lower affinity than the wildtype conformation. We suggest that this conformational change reflects an inherent adaptability to the PAS domain fold that is used for a wide range of protein-binding activities and switchability in many systems.

## *B. Results*

### *i. Point mutations reveal second conformation of ARNT PAS-B.*

Our previous studies of ARNT PAS-B indicated that the solvent-exposed surface of the  $\beta$ -sheet is used both in the heterodimeric complex with HIF-2 $\alpha$

PAS-B and a weaker homodimer (7, 12). To examine the importance of specific ARNT residues in these complexes, we designed mutations on this surface to disrupt protein/protein interactions from the  $\beta$ -sheet surface. Unexpectedly, a change to a solvent-exposed tyrosine on the I $\beta$  strand (Y456T, Figure 3-1b) generated a protein that exhibited crosspeak doubling in  $^{15}\text{N}$ - $^1\text{H}$  HSQC spectra for approximately thirty backbone amides (Figure 3-2a). This doubling generated two crosspeaks of roughly equivalent intensity for each of these sites, one of which remained near the chemical shifts of the wildtype ARNT PAS-B peak. This is illustrated by the peak from the E403 backbone amide (Figure 3-2b), which has one peak near the wildtype (peak **1**) and another shifted upfield in  $^1\text{H}$  by  $\sim 0.1$  ppm (peak **2**). Notably, E403 is located 24 Å from the site of mutagenesis, consistent with a *bona fide* structural change. Coupled with the excellent  $^1\text{H}$  chemical shift dispersion of the new conformer **2** peaks, these data suggested that the Y456T point mutant adopted two distinct, well-folded, and equally populated conformations under these solution conditions. Sample heterogeneity, *e.g.* proteolysis, could have provided a trivial explanation for these results but was inconsistent with mass spectrometry data which revealed a single MW = 13881.7 Da species, (consistent with the predicted mass of 13880.6 Da, data not shown).



Examining spectra for other ARNT PAS-B domains containing point mutations at other solvent-exposed residues near Y456, we found that these proteins adopted differing ratios of the same two distinct conformations. This is demonstrated by the F444Q mutation, which affects a position on the H $\beta$  strand adjacent to Y456 (Figure 3-1b). On its own, the F444Q change does not reveal a second conformation in the wildtype context (Figure 3-2b). However, it significantly changes the relative populations of wildtype **1** and novel **2** crosspeaks when combined with Y456T (Y456T: **1:2** ratio = 50:50; F444Q/Y456T = 19:81). Further mutations at nearby sites exhibited similar behavior, generating a group of seventeen proteins (containing three or fewer point mutations) which have a range of preferences for one conformation or the other spanning >99:<1 for both conformations **1** and **2** (Table 3-1).

Mutation	Percent WT	Percent MUT
Wild-Type	>99	<1
F444Q	>99	<1
F444Q/I454A/Y456T	>98	<2
F444Q/I454F/Y456T	>98	<2
F444Q/Y456T/I458E	>98	<2
Y456A	>98	<2
Y456S	81	19
Y456T	50	50
P449A/Y456T	32	68
F444Q/T445A/Y456T	31	69
F444Q/Y456T	19	81
F444Q/Y456T/T460V	15	85
F444Q/T445Y/Y456T	14	86
F444Q/N448T/Y456T	10	90
F444Q/N448W/Y456T	<2	>98
F444Q/Y456T/I457F	<2	>98
F444Q/F446A/Y456T	<1	>99

**Table 3-1.** Effect of ARNT PAS-B mutations on relative populations in conformations. ~8% relative error on any given ratio.

*ii. Interconversion rates between the two conformations*

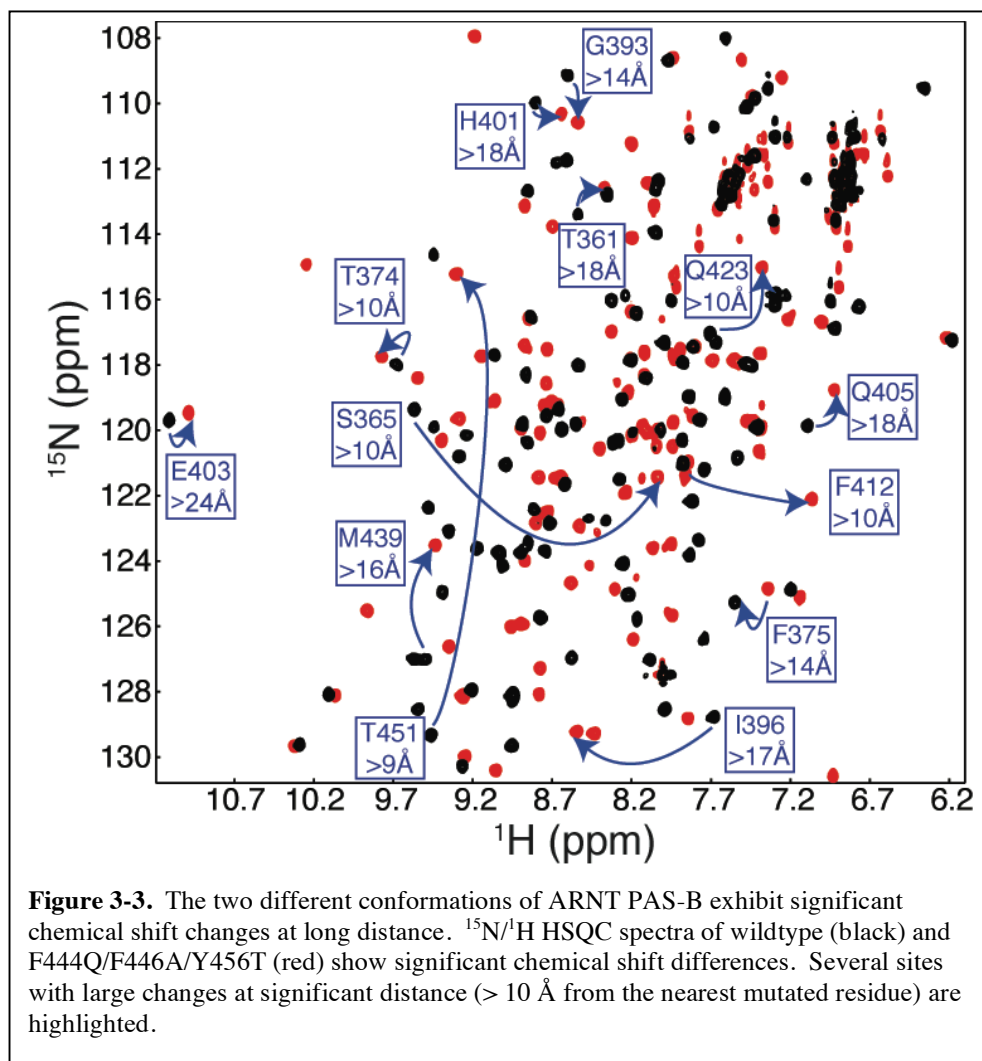
While these data indicate that these domains can adopt two distinct conformations, we set out to establish that these states interconverted in an equilibrium process. The presence of two distinct sets of peaks in our spectra indicates that such interconversion must be slow ( $\tau > \text{ms}$ ) (142). While interconversion could not be directly detected in the Y456T mutant with ZZ-exchange experiments (143), we found that this process was slow enough to allow chromatographic separation of the conformers at 4°C.  $^{15}\text{N}$ - $^1\text{H}$  HSQC spectra

collected on a Y456T sample before and after this chromatographic step demonstrated that we could significantly enrich conformation **2** to a **1:2** ratio of 12:88 starting from a Y456T sample equally populated in both states (Figure 3-2c). After incubating this sample at room temperature for 29 hr, peak intensities had nearly restored to their original values. These data are consistent with a slow interconversion occurring *in vitro* with a first-order exponential time constant of 16 hr or shorter.

*iii. Solution structure of the new conformation 2*

To understand the structural changes between the two conformations we observed in these domains, we determined the solution structure of a variant of human ARNT PAS-B that was stabilized in conformation **2** (>99%, Table 3-1) by three point mutations, F444Q/F446A/Y456T. Using standard triple resonance NMR experiments performed on uniformly  $^{15}\text{N}$ ,  $^{13}\text{C}$ -labeled protein samples, we assigned  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shifts for over 95% of NMR-observable sites in the protein. Comparing the chemical shifts of the wildtype and triple mutant proteins, we observed numerous changes at distances over  $>20 \text{ \AA}$  from the locations of the mutated residues in the native structure (Figure 3-3), thus reinforcing the prospect of a global change between states **1** and **2**.



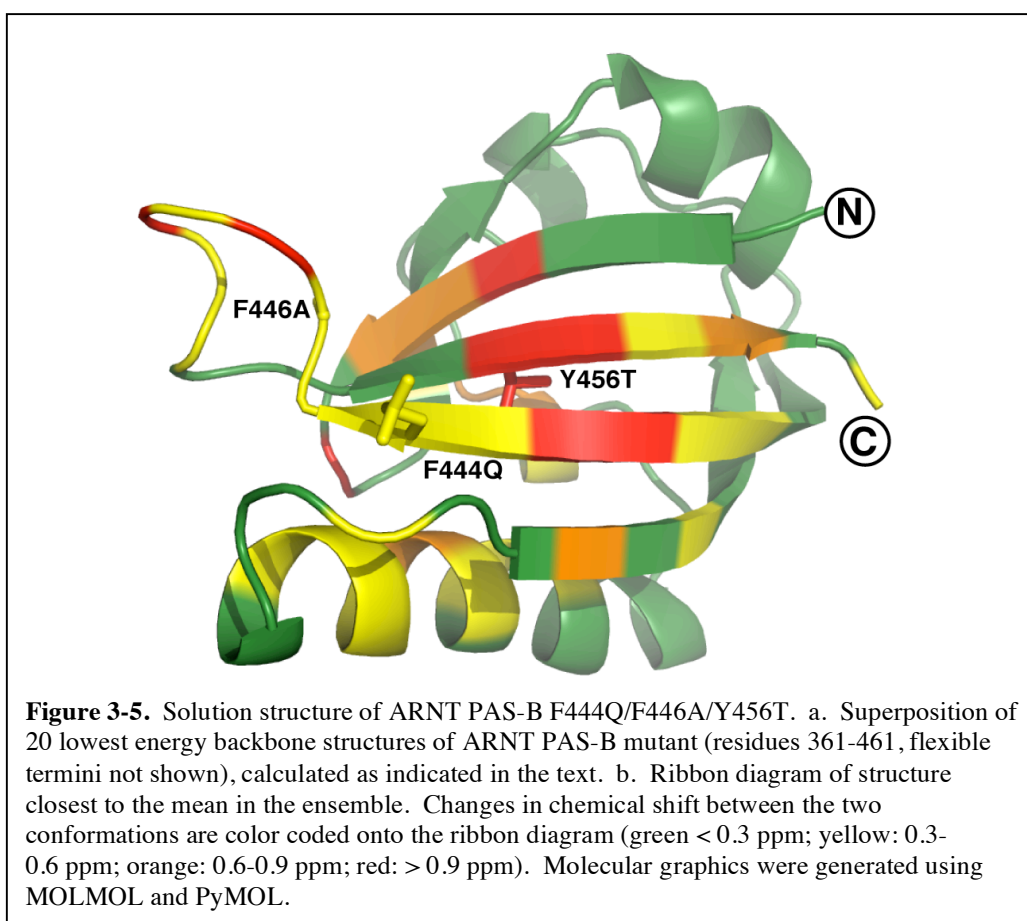
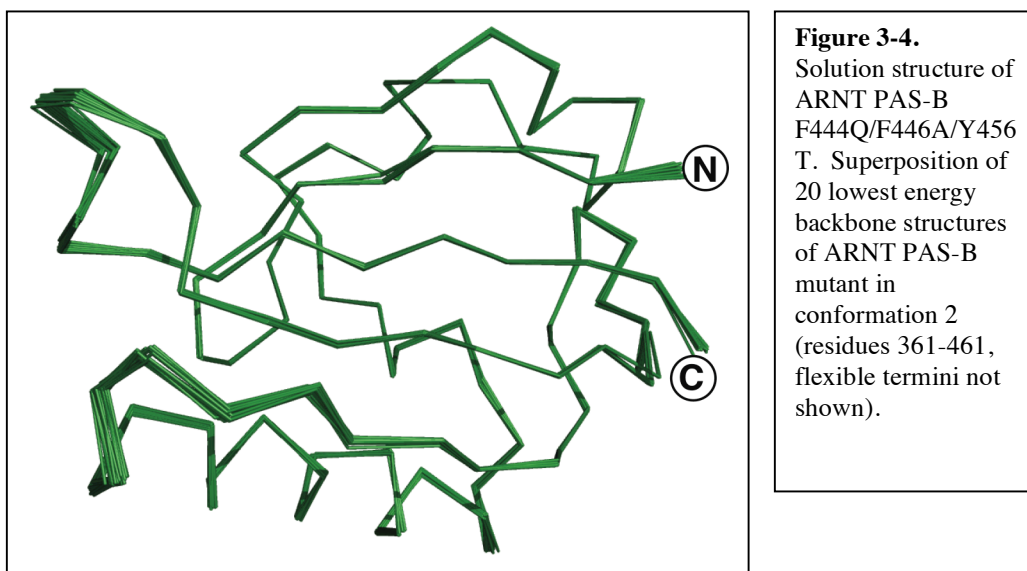


For structure determination, we used over 3,000 interproton distance, dihedral angle and hydrogen bond constraints (Table 3-2). Using ARIA/CNS (144, 145), we obtained a high-precision ensemble of the 20 lowest energy structures that were all consistent with these data (Figure 3-4). From this group, it is clear that the mutant ARNT PAS-B domain retains the standard PAS topology

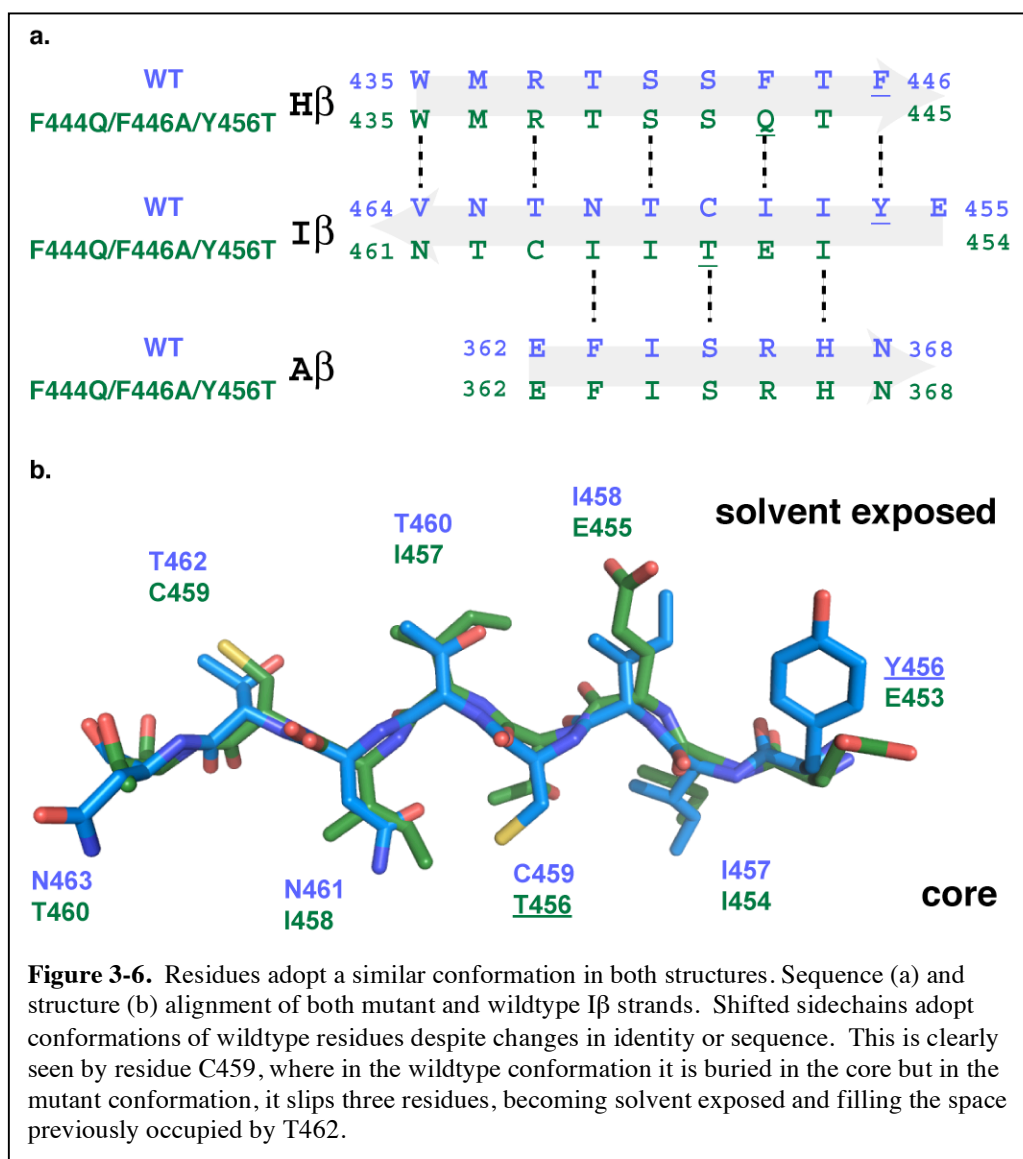
while exhibiting extensive backbone amide chemical shift differences at long range ( $> 10 \text{ \AA}$ ) in the  $\beta$ -sheet and adjacent F $\alpha$  helix (Figure 3-5).

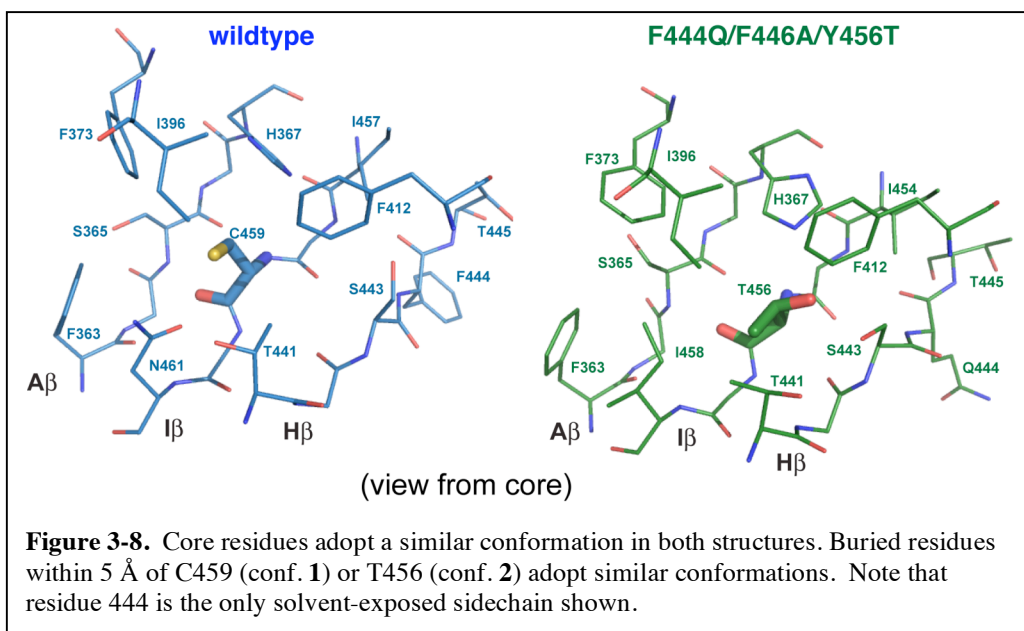
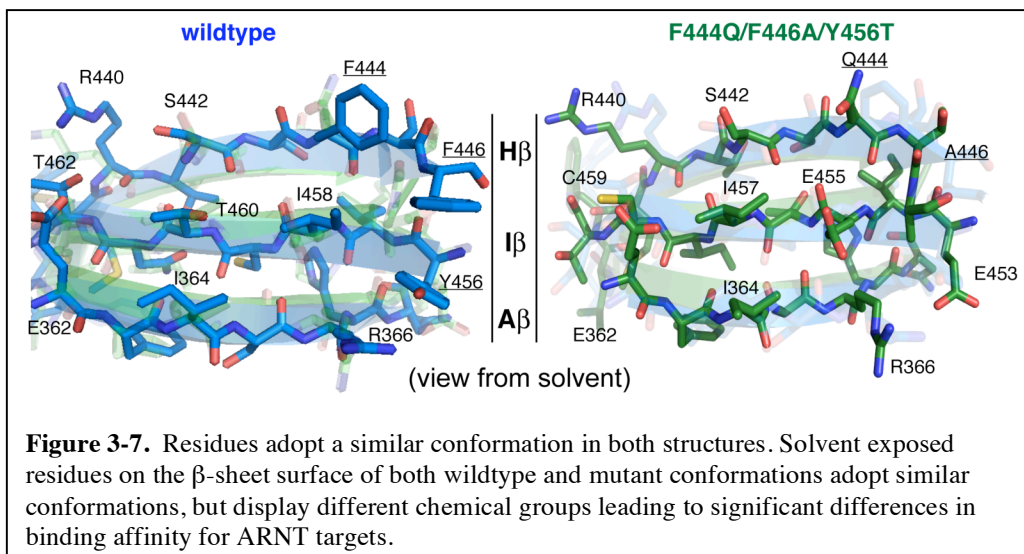
<i>Structural analysis:</i>	
NOE distance restraints	
Unambiguous	2688
Ambiguous	490
Hydrogen bond restraints	50
Dihedral angle restraints	133
Mean r.m.s.d. from experimental restraints	
NOE ( $\text{\AA}$ )	0.033 $\pm$ 0.001
Dihedral angles (deg.)	0.0044 $\pm$ 0.0321
Average number of:	
NOE violations $> 0.5 \text{ \AA}$	0
NOE violations $> 0.3 \text{ \AA}$	0.45 $\pm$ 0.50
Dihedral violations $> 5^\circ$	0
Mean r.m.s.d. from idealized covalent geometry	
Bond lengths ( $\text{\AA}$ )	0.0039
Bond angles (deg.)	0.54
Impropers (deg.)	0.47
<i>Geometric analysis of residues 360-462</i>	
r.m.s.d. from the mean	
Backbone atoms ( $\text{\AA}$ )	0.28 $\pm$ 0.08
All heavy atoms ( $\text{\AA}$ )	0.68 $\pm$ 0.07
Ramachandran analysis (PROCHECK)	
Most-favored region (%)	71.8
Additionally allowed region (%)	24.5
Generously allowed region (%)	2.7
Disfavored region (%)	0.8

**Table 3-2.** Statistics for ARNT PAS-B F444Q/F446A/Y456T solution structure determination.



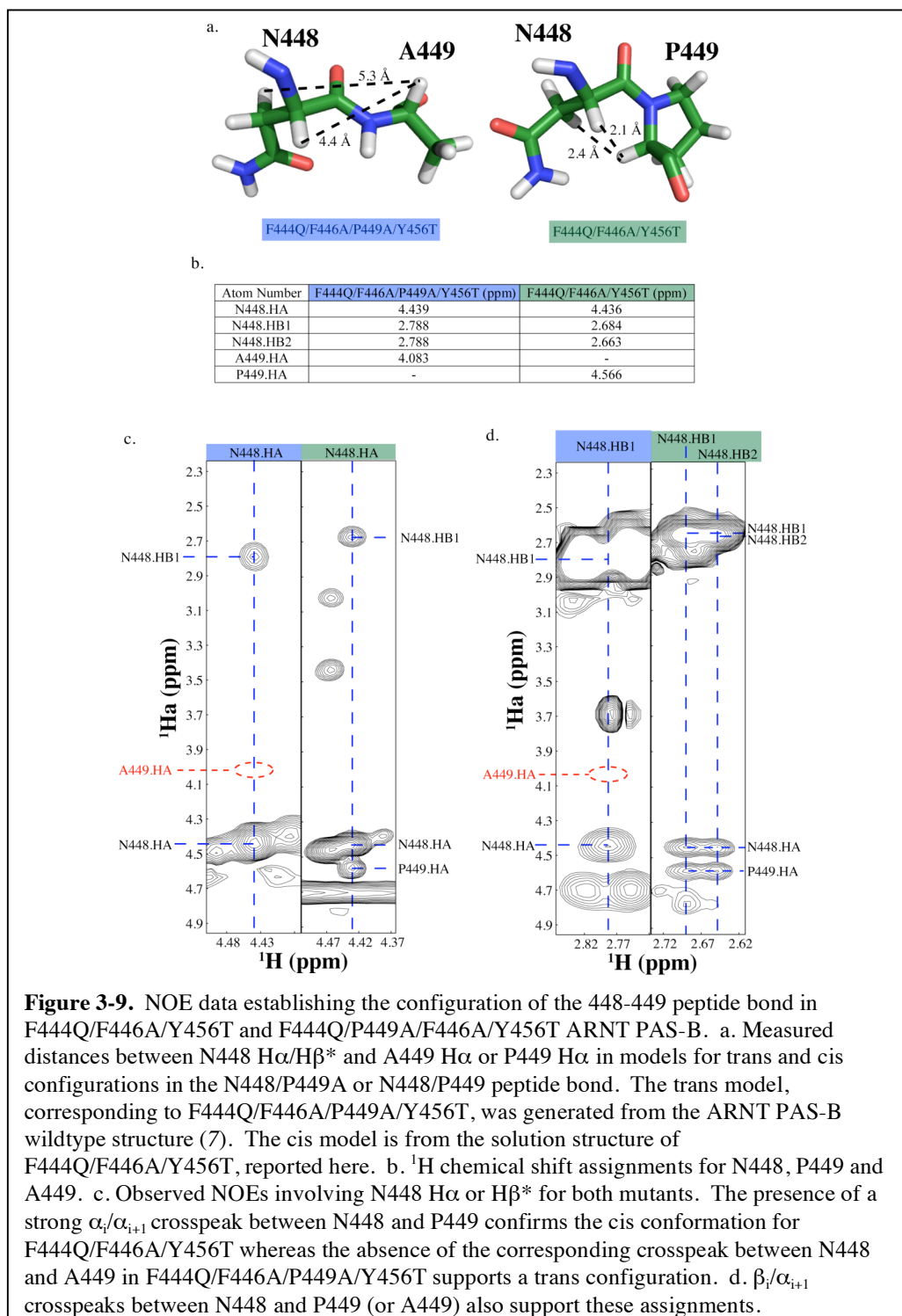
On closer inspection, we found that the source of our long-range changes was a three-residue register shift of the I $\beta$  strand, corresponding to a movement of about 10 Å towards the C-terminus (Figure 3-6a). As expected by the odd-numbered change in register, this shift inverts the topology of the I $\beta$  strand, burying several previously exposed sidechains (and vice-versa; Figure 3-6b). Notably, only one position has the same residue type in both conformations (conformer **1** I457 = **2** I454), necessitating the packing of different sidechains into similar conformations (Figure 3-6b, 3-7, 3-8). These changes include the mutated Y456T, whose sidechain adopts an interior position previously held by C459, placing it in an environment with a mix of polar and hydrophobic neighbors. This process also significantly remaps residues presented by ARNT at the HIF- $\alpha$  binding surface, suggesting changes in binding affinity (*vide infra*). Aside from this, the remainder of the protein retains a very similar overall structure to the wildtype domain (7), with a 1.08 Å pairwise backbone rmsd between the wildtype and triple mutant ensembles for residues 360-443 (omitting the HI loop and I $\beta$  strand).





*iv. N448-P449 peptide bond isomerization*

To accommodate the I $\beta$  strand register shift, the preceding HI loop shortens from eight to five residues. This is associated with a *trans* to *cis* isomerization of the peptide bond between N448 and P449 within the HI loop. This bond, which adopts a *trans* configuration for the wildtype domain, switches to *cis* in the triple mutant as established by strong N448 H $\alpha$ -P449 H $\alpha$  NOEs (Figure 3-9) and large ( $\delta(^{13}\text{C}\beta) - \delta(^{13}\text{C}\gamma) > 9$  ppm) differences between P449  $^{13}\text{C}\beta$  and  $^{13}\text{C}\gamma$  chemical shifts (146) in the mutant domain, compared to 5 ppm for the wildtype. Two lines of evidence suggest that this is a secondary effect to the I $\beta$  strand slip and not causative. First, a P449A point mutation in a Y456T background led to only minor changes in the **1:2** equilibrium (P449A/Y456T: 32:68, Y456T: 50:50 corresponding to  $\Delta\Delta G \sim 1.5$  kcal mol $^{-1}$ ; Table 3-1). Second, when the same P449A mutation was made in a F444Q/F446A/Y456T background, we observed that the N448-A449 peptide bond was in a *trans* configuration (Figure 3-9) while remaining predominantly in the +3 slipped I $\beta$  strand conformation (**1:2** ratio: 7:93). Taken together, these data suggest that the ability to slip the I $\beta$  strand is primarily established by residues in the  $\beta$ -sheet itself rather than in the neighboring loops.

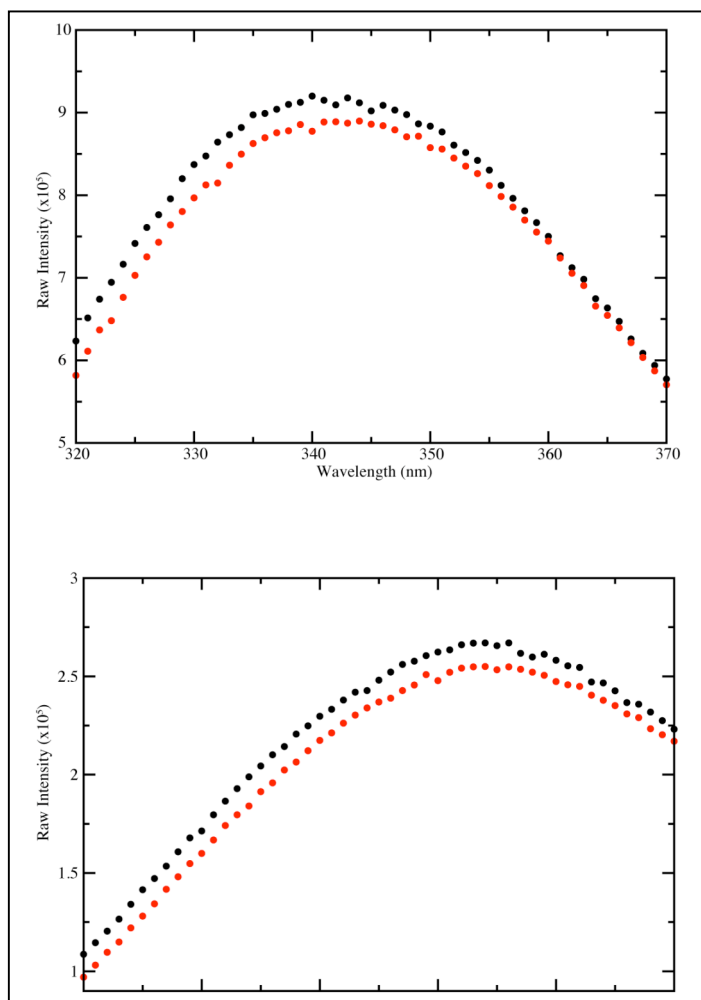




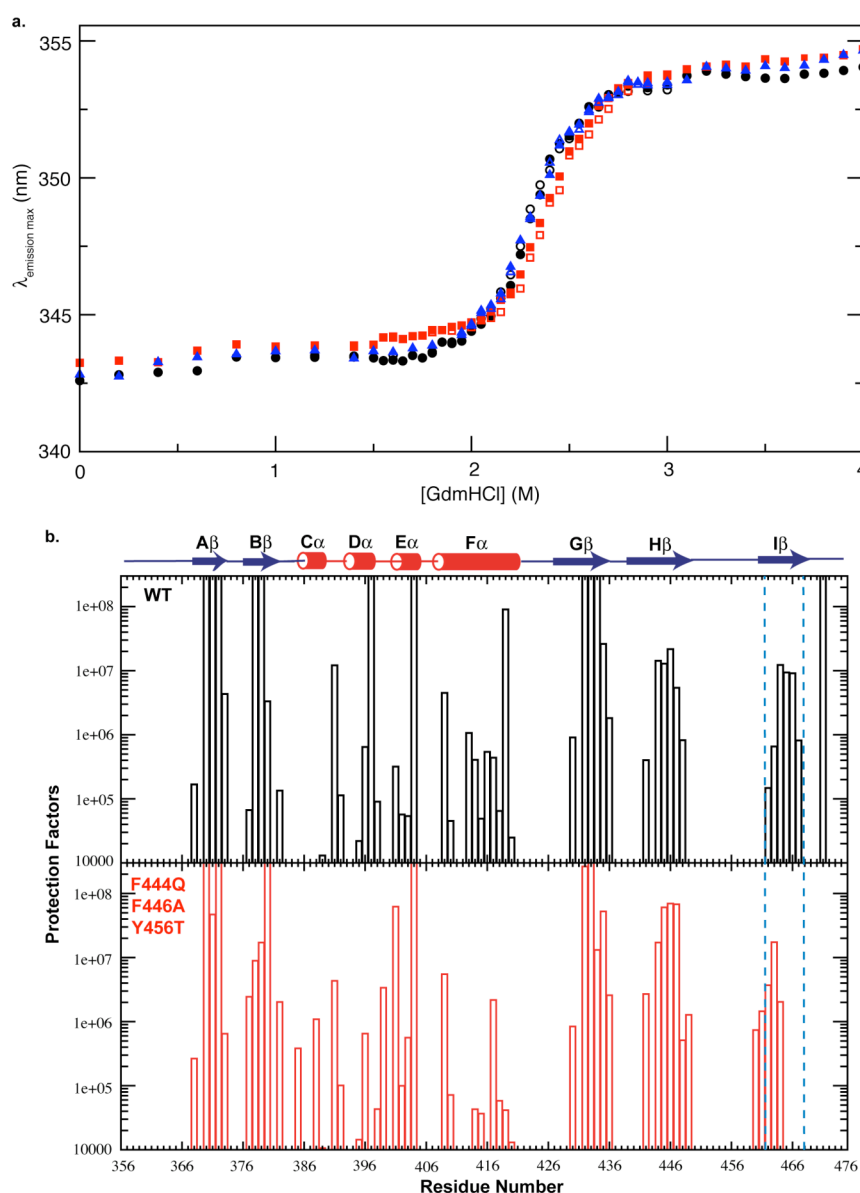
*iv. Effect of slipped conformation on domain stability*

Complementing these studies of the structural changes in ARNT PAS-B introduced by this  $\beta$ -strand slip, we characterized the effect on protein stability using equilibrium denaturation and  $^2\text{H}$  exchange measurements. Using the intrinsic fluorescence properties of two tryptophan residues (W436, W438) in the H $\beta$  strand, we monitored the equilibrium denaturation behavior of the wildtype, Y456T and F444Q/F446A/Y456T ARNT PAS-B domains using guanidinium hydrochloride (GdmHCl). As expected for denaturant-induced unfolding, increasing GdmHCl concentrations were accompanied by shifts of fluorescence emission spectra to longer wavelengths (344 to 354 nm) (Figure 3-10, Figure 3-11). All three proteins exhibited similar cooperative and reversible unfolding behaviors with transitions centered near 2.3 M GdmHCl. While these titrations formally involve three states (folded states **1** and **2** plus a denatured state), the similarities of the emission spectra of proteins which populate **1** or **2** reduces these data to being sensitive to the relative populations of only two states, a combined native state (**1** + **2**) and an unfolded state. Combined with the similarities of the denaturation profiles (Figure 3-11) of all three proteins, we do not anticipate significant differences in the global stabilities of conformations **1** and **2**, allowing us to interpret these data in a pseudo two state model involving a folded and an unfolded state. Such analyses indicated that the wildtype, Y456T and F444Q/F446A/Y456T domains all had very similar global stabilities,

$\Delta G_u = 9.0, 9.4$  and  $9.2 \text{ kcal mol}^{-1}$ , respectively. In parallel,  $^2\text{H}$  exchange measurements conducted under EX2 conditions demonstrated comparable protection factors for the wildtype, Y456T and F444Q/F446A/Y456T, with the only significant changes being the expected changes in residues in the I $\beta$  strand due to the +3 register shift (Figure 3-11). These data indicate proteins in either conformation **1** or **2** (or interconverting between both) are equivalently stable, despite the significant change in  $\beta$ -sheet register.



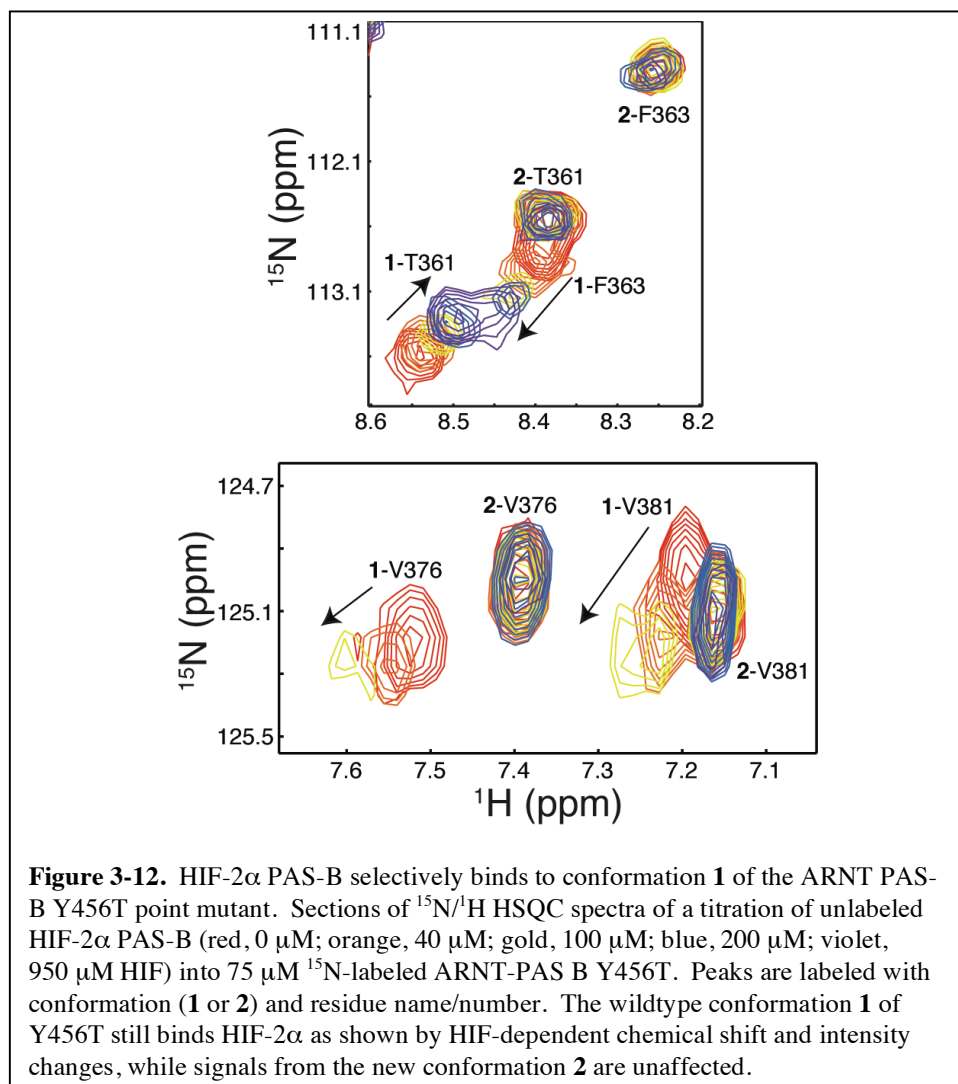
**Figure 3-10.** Two tryptophan residues within ARNT PAS-B have similar local environments throughout the unfolding process in both conformations. a. Tryptophan fluorescence emission spectra reveal similar local environments for the two ARNT PAS-B tryptophan residues for the wildtype (black circles) and F444Q/F446A/Y456T (red circles) proteins under both native (top) and denaturing conditions (3 M GdmHCl, bottom).



**Figure 3-11.** ARNT PAS-B mutants have similar global stabilities as wildtype. a. Equilibrium denaturation behavior of ARNT PAS-B wildtype (black circles), Y456T (blue triangles) and F444Q/F446A/Y456T (red squares), monitored by shifts in the wavelength maximal intensity in tryptophan fluorescence emission spectra. Guanidinium hydrochloride denaturation is reversible as shown by the similarity of denaturation (filled symbols) and renaturation (open symbols) profiles. b.  $^2\text{H}$  exchange protection factors of wildtype and F444Q/F446A/Y456T ARNT PAS-B.

*v. Functional effect of I $\beta$  slip*

To examine functional differences between conformations **1** and **2**, we took advantage of the favorable interconversion kinetics and equivalent populations of Y456T ARNT PAS-B to simultaneously monitor binding of both conformations to HIF-2 $\alpha$  PAS-B. Complex formation can be monitored by chemical shift and peak intensity changes in the  $^{15}\text{N}$ - $^1\text{H}$  HSQC spectra of one domain as the other partner is added (7, 12). Titrating HIF-2 $\alpha$  PAS-B into a  $^{15}\text{N}$ -labeled sample of Y456T, we observed dimerization-associated spectral changes only with the wildtype conformation **1** (Figure 3-12). Examining the concentration dependence of these changes, we estimate that conformation **1** binds HIF-2 $\alpha$  PAS-B with a  $K_d=90\ \mu\text{M}$  affinity. In contrast, conformation **2** shows no binding when HIF-2 $\alpha$  PAS-B was added at concentrations up to 950  $\mu\text{M}$ . Given these data and the minimum dimer concentration that would give a detectable spectral change, we suggest a minimum  $K_d\sim 10\ \text{mM}$  for this interaction, showing that the inversion of the I $\beta$  strand topology effectively abolishes the HIF-2 $\alpha$ /ARNT PAS-B interaction. Notably, we did not observe a significant change in the **1:2** equilibrium during these titrations after the samples were allowed to equilibrate at room temperature for 20 hours, perhaps related to the slow interconversion between conformations and low affinity of the complex (Figure 3-2).



### C. Discussion

The  $\beta$ -sheet surface of many PAS domains bind a variety of inter- (7, 8, 18) or intramolecular (28, 138, 147) partners, critical to their function as protein/protein interaction domains (12). Notably, environmental stimuli modulate the affinity of these binding events, harnessing changes in the

conformations of cofactors inside of the domain itself (28, 138, 148, 149). These proteins suggest that intramolecular signaling can be achieved by ligand control of the  $\beta$ -sheet structure or stability, which in turn alters PAS/protein binding equilibria (150). This is supported by site-directed mutagenesis studies highlighting the critical role played by I $\beta$  strand residues in signal transduction in a subset of photosensory PAS domains via internally-bound flavin chromophores (138, 151, 152).

In this context, it is perhaps surprising that the structure and function of this critical  $\beta$ -sheet is fragile to the point that a single point mutation (Y456T) enables the domain to slip into a different, non-functional I $\beta$  strand register. Notably, this change allows ARNT PAS-B to slowly interconvert between both of these two possible I $\beta$  registers, contrary to the usual behavior linking sequence and a single structure as encapsulated in Anfinsen's hypothesis (5). Subsequent mutations (Table 3-1) adjust the relative populations of the two states, shifting some forms almost completely into one conformation while others continue to sample both of the two structures. It should be emphasized that these changes occur without compromising the integrity of the domain itself, which remains remarkably unchanged in its overall structure and stability. As such, our data are consistent with mutations shifting the relative stabilities of conformations **1** or **2** within the context of a structurally and energetically stable fold, where the final I $\beta$  strand effectively completes the domain topology in a manner reminiscent of

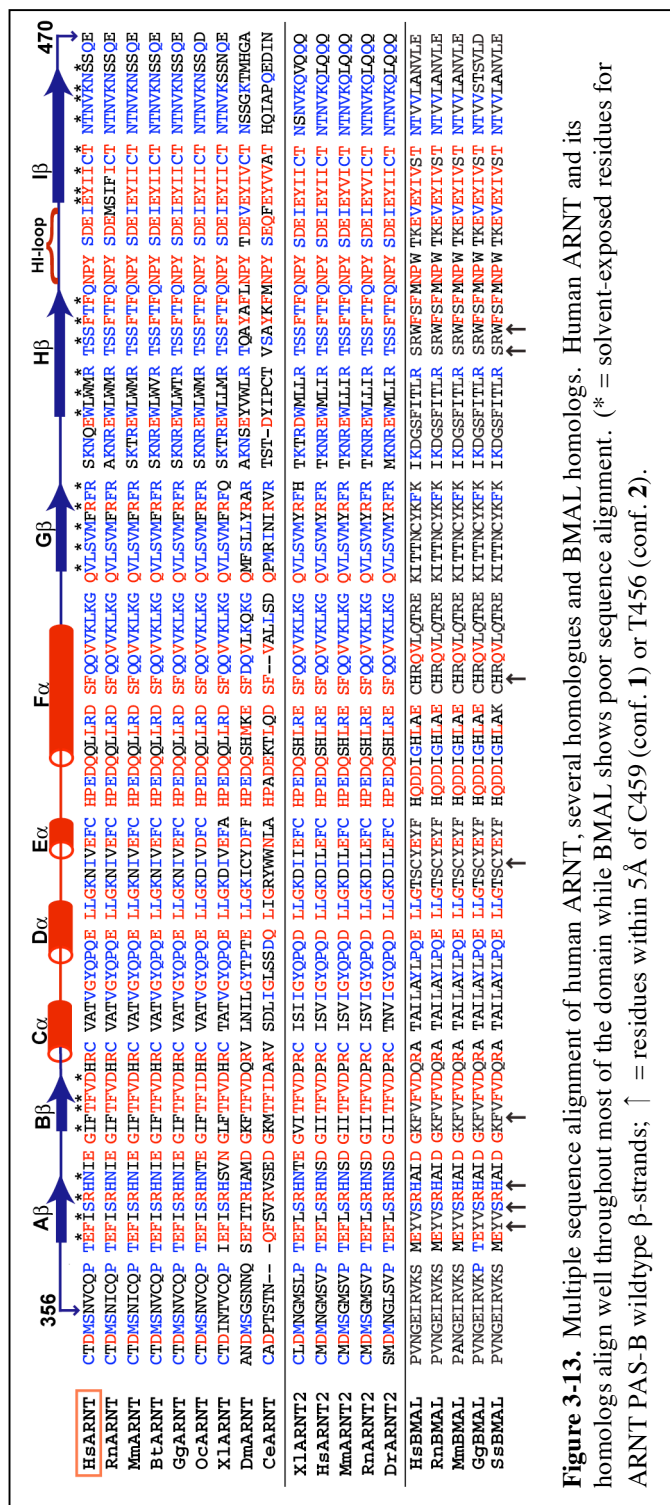
receptor/ligand interactions. These thermodynamic effects on the stabilities may also be complemented by kinetic changes via alterations in the energetics of the transition state between **1** and **2**, but characterization of such effects must await more precise measurements of interconversion rates between these species.

*i. Structural features that allow this transition.*

What structural features of ARNT PAS-B enable such a transition to occur? An obvious candidate is the sequence of the I $\beta$  strand, which is well conserved among all homologs and consists of a mix of similarly-sized hydrophobic and polar residues (Figure 3-13). The key exception to this is Y456, suggesting that the Y456T point mutation removes a local impediment to the wildtype protein undergoing the +3 I $\beta$  strand slip. We suggest that the bulk of the Y456 sidechain itself plays a significant role in preventing the slip, as we expect significant steric occlusion to hinder the burial of this sidechain in the wildtype sequence (Figure 3-6b). Notably, the removal of steric hindrance at this position is not the sole determinant of slip propensity, as indicated by the differences in **1:2** equilibrium among the Y456T > Y456S > Y456A mutants (Table 3-1). This suggests that interactions between this site (occupied by a polar residue, C459, in conformation **1**) and neighboring residues also influence register choice (Figure 3-8). As expected, other changes in the newly-slipped I $\beta$  or neighboring H $\beta$  strands can also modulate the relative populations of the two conformations we observed,

as demonstrated by point mutations at F444, I454 or I458 in addition to Y456 (Table 3-1). As expected, changes that result in the burial of charged or large groups diminish the population of the slipped conformation **2** (*e.g.* I458E), while mutations that remove unfavorable interactions in the new form promote its formation (*e.g.* F444Q, which is adjacent to a newly-exposed E455). These data indicate that sidechain-sidechain tertiary interactions play a critical role in establishing  $\beta$ -strand register, as they do with  $\beta$ -strand propensity itself (153, 154).





**Figure 3-13.** Multiple sequence alignment of human ARNT, several homologs and BMAL homologs. Human ARNT and its homologs align well throughout most of the domain while BMAL shows poor sequence alignment. (\* = solvent-exposed residues for ARNT PAS-B wildtype β-strands; ↑ = residues within 5Å of C459 (conf. 1) or T456 (conf. 2)).

As noted above, slippage of the I $\beta$  strand shortens the preceding HI loop from eight to five residues and leads to a *trans*- to *cis*- isomerization of the N448-P449 peptide bond in this loop. Given the minor effects this has on the equilibrium, and our confirmation that a P449A point mutant retains a *trans* configuration at the preceding peptide bond while in a slipped register, it is clear that the I $\beta$  strand sequence and interactions are the primary determinants of the strand register at equilibrium. We note that this leaves open the potential for this bond isomerization to have significant kinetic effects as seen in several biologically-relevant systems (155), but this remains to be fully characterized.

*ii. PAS domains flexible central  $\beta$ -sheet.*

Looking more broadly in the context of PAS domain signaling, it is clear that some degree of flexibility in the central  $\beta$ -sheet is essential. This has been most vividly demonstrated among a class of light-sensitive PAS domains, known as LOV (Light-Oxygen-Voltage) domains (156), which regulate the conformations of various regulatory elements on the external side of the sheet in response to light (28, 138). Photoactivation is dependent on a highly-conserved glutamine (Q513 in *A. sativa* phototropin 1), which is located in the I $\beta$  strand. Upon the formation of the photochemical adduct, this residue changes its conformation and hydrogen-bonding patterns with the internally-bound flavin chromophore. These changes are a key element of the photodetection process as

established by conformational (138, 151) and functional changes (152) caused by mutations to this residue in several LOV proteins. While a signaling mechanism has yet to be fully detailed, the most direct path is *via* distortions in the central  $\beta$ -sheet. Additional data supporting the malleability of PAS  $\beta$ -sheets comes from a combination of experiments and simulations with single molecule pulling studies of photoactive yellow protein (PYP) (104, 156). Under mechanical stress, PYP slips its A $\beta$  strand while the remainder of the protein remains intact, entirely analogous to our +3 shift in  $\beta$ -strand register that spontaneously occurs in Y456T ARNT PAS-B. While the biological role of such motions remains to be established, it is tempting to speculate that this flexibility may somehow reflect inherent flaws or properties in the PAS domain fold that are equally useful for ARNT to bind different bHLH-PAS partners as they are for LOV and PYP to distort the sheet in response to photoactivation.

### *iii. $\beta$ -strand slips in nature*

More globally,  $\beta$ -strand slips have been observed in a variety of biological systems, where they can be triggered by a variety of stimuli including ligand binding (*e.g.* ARF (157)), changes in solvent conditions (*e.g.* lymphtactin (15)) or proteolysis (*e.g.* clotting factor VIIa (158)). Their functional role has been detailed in enzymatic regulation (158, 159) and in the conversion of  $\beta$ -2-microglobulin and transthyretin from their soluble forms into disease-associated

amyloid fibrils (160, 161). These cases exploit the long-range conformational changes that can be generated by  $\beta$ -strand slips, which can be particularly significant for odd-numbered shifts (as observed for all of the systems above except ARF). We suggest that the barriers which establish normal Anfinsen sequence:structure relationships (5) can be fragile for a wider variety of natural proteins, susceptible to modifications in sequence (*vide supra*) or solution conditions (15), enabling the same sequence to explore multiple stable folded conformations. In addition to providing valuable models to explore the sequence dependence of determining the equilibrium positions of strand register and interconversion kinetics, these may also reveal a greater role for such slips in other systems that affect protein structure and function (162).

#### *D. Materials and Methods*

##### *i. Mutagenesis*

Point mutants of ARNT PAS-B were created from wildtype DNA and primers including the desired mutation(s), with the Quick Change Site-Directed Mutagenesis Kit (Stratagene). ARNT PAS-B mutants were obtained by PCR amplification of the corresponding sequence and digestion of the parental template with *DpnI*, as directed by the manufacturer. The modified plasmid was

transformed into the pHis-parallel bacterial expression vector (163) and sequenced to confirm the presence of the desired mutations.

*ii. Protein expression and purification*

DNA-encoding fragments of ARNT PAS-B (residues 356-470) were subcloned into the pHis-parallel expression vector and transformed into *E. coli* BL21(DE3) cells. Cultures were grown in M9 minimal medium in the presence of 1 g/L  $^{15}\text{NH}_4\text{Cl}$  for U- $^{15}\text{N}$  samples (and 3 g/L of  $^{13}\text{C}_6$ -glucose for U- $^{15}\text{N}$ ,  $^{13}\text{C}$  labeled samples) at 37°C to an  $A_{600}$  of 0.6-0.8, upon which the cells were induced with 0.5 mM isopropyl  $\beta$ -D-thiogalactoside for 15 hr at 20°C. Following expression, cell pellets were harvested, resuspended in 20 mL of 50 mM Tris (pH 7.5), 15 mM NaCl and 20 mM imidazole, lysed by high-pressure extrusion, centrifuged, and filtered (0.45  $\mu\text{m}$  pore size). The supernatant was then injected over a  $\text{Ni}^{2+}$ -NTA affinity column and eluted using a linear 20 mM-500 mM imidazole gradient. His<sub>6</sub>-tagged ARNT PAS-B mutants eluted from the column were buffered exchanged into an imidazole-free buffer and followed by His<sub>6</sub>-tag cleavage overnight using His<sub>6</sub>-TEV protease. Another pass over a  $\text{Ni}^{2+}$ -NTA column was used to remove His<sub>6</sub>-TEV and free His<sub>6</sub>-tag and the remaining protein was concentrated in an Amicon pressure-driven ultrafiltration cell with YM10 10 kDa filters. The molecular weight of the purified protein, which contains a four residue vector-derived N-terminal cloning artifact (GAMD) followed by

residues 351-470 of human ARNT, was confirmed for the wildtype protein by electrospray ionization mass spectrometry and for the mutants by DNA sequencing.

*iii. Biochemical purification of mutant conformation in Y456T*

ARNT PAS-B Y456T protein suspended in 50 mM Tris (pH 6.6), 17 mM NaCl buffer was injected over a 1 mL MonoQ column. The chromatogram resulted in one of three peaks that was enriched ~88% for the mutant conformation, as established by  $^{15}\text{N}$ - $^1\text{H}$  HSQC spectra.

*iv. Structure determination of ARNT PAS-B F444Q/F446A/Y456T*

NMR experiments for structure determination were carried out at 25°C on a Varian Inova 600 and 800 MHz spectrometers with 0.8 mM samples of uniformly  $^{15}\text{N}$ - or  $^{15}\text{N}$ ,  $^{13}\text{C}$ -labeled ARNT PAS-B F444Q/F446A/Y456T mutant in 50 mM Tris (pH 7.5), 17 mM NaCl and 5 mM DTT solutions. Chemical shift assignments were obtained from 3D HNCACB, CBCA(CO)NH, HNCO and HCCH-TOCSY data. Interproton distance restraints were obtained from 3D  $^{15}\text{N}$  edited NOESY (120 ms), 3D  $^{15}\text{N}$ ,  $^{13}\text{C}$  simultaneous-edited NOESY (120 ms), and 2D NOESY (120 ms) spectra. Hydrogen bond constraints ( $1.3 \text{ \AA} < d_{\text{NH-O}} < 2.5 \text{ \AA}$ ,  $2.3 \text{ \AA} < d_{\text{N-O}} < 3.5 \text{ \AA}$ ) were defined for backbone amide protons protected from  $^2\text{H}$  exchange for more than 5 hr (25°C, pH 7.5) (164).  $\phi$  and  $\psi$  dihedral angle

constraints were obtained from analysis of backbone chemical shifts using TALOS, with restraints set at twice the standard deviation yielded by TALOS (165) (15° minimum). Initial structures were calculated with automatic assignments of NOESY spectra using ARIA 2.1 (144) and later redefined using both automated and manual assignments. A total of 1,000 structures were obtained, and the 20 structures with the lowest NOE energy were evaluated with MOLMOL (14) and PROCHECK-NMR (19).

*v. Global stability measurements*

Global stability measurements were made using standard equilibrium denaturation measurements (monitored by tryptophan fluorescence) and  $^2\text{H}$  exchange measurements (monitored by sequentially acquired 2D  $^{15}\text{N}/^1\text{H}$  HSQC). All experiments were carried out at 25°C in 50 mM sodium phosphate (pH 7.5), 17 mM NaCl. GdmHCl-induced equilibrium denaturation/renaturation experiments were performed after 24 hr of equilibrating protein samples in buffers containing different concentrations of denaturant. Equilibrium measurements were monitored using a Fluorolog-3 spectrofluorimeter (JY Horiba). Intrinsic tryptophan fluorescence emission spectra were recorded between 290 and 380 nm with excitation at 295 nm in a 4 mm pathlength cell. Best-fit curves were generated using nonlinear regression, and the locations of

maximum emission intensity were identified by calculating the first derivative of these spectra and finding the position where the first derivative is zero.

Initially,  $^{15}\text{N}$ -labeled ARNT PAS-B samples (in 50 mM Tris (pH 7.5), 17 mM NaCl) were lyophilized and resuspended in 99%  $\text{D}_2\text{O}$ , starting the deuterium exchange reactions.  $^{15}\text{N}/^1\text{H}$  HSQC spectra were sequentially recorded every 20 min. Afterwards, the pH was recorded and deuterium exchange rates were calculated using nmrview, converting these into protection factors using the approach of Bai *et al.* (164).

To confirm that these measurements were made under EX2 conditions, we examined the pH dependence of the  $^2\text{H}$  exchange rates for wildtype ARNT PAS-B. As expected, we observed a significant decrease in these rates at lower pH. Comparing the exchange rates at pH 6.76 and 7.63 (uncorrected, measured pH after exchange), we found an average ratio of exchange rates of 0.163 ( $k_{\text{ex}}[\text{pH } 6.76]/k_{\text{ex}}[\text{pH } 7.63]$ ) for  $n=19$  sites. This compares favorably to a calculated ratio of 0.135 given these pH values, confirming EX2 conditions.

#### *E. Acknowledgements*

We thank Daniel Buehler, Leanna Steier, Amy Zhou and Brian Zoltowski for their contributions to this work and members of the Gardner lab for constructive comments and suggestions. This research was supported by a grant from the NIH (R01 GM081875) to K.H.G.



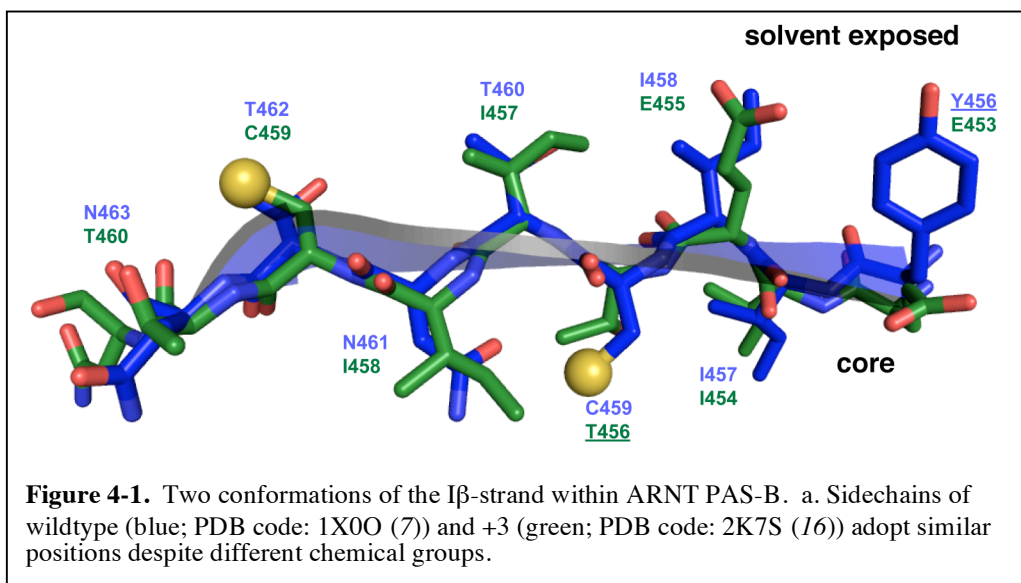
#### IV. Transition between two $\beta$ -strand registers of ARNT PAS-B Y456T

goes through a chiefly unfolded state

##### *A. Introduction*

The aryl hydrocarbon receptor nuclear translocator (ARNT) is a promiscuous basic helix-loop-helix Period/ARNT/Single-minded (bHLH-PAS) protein that controls various biological pathways as part of dimeric transcriptional regulator complexes with other bHLH-PAS proteins (66, 67, 139, 140). These complexes utilize two PAS domains within ARNT, PAS-A and PAS-B, for protein/protein interactions via residues located on their  $\beta$ -sheet surfaces (7). Previous mutagenesis studies of ARNT PAS-B have demonstrated the plasticity of this domain to adopt different conformations (16). As shown in Chapter 3, one point mutation (Y456T) on the ARNT PAS-B solvent exposed  $\beta$ -sheet surface resulted in two conformations (“wt” and “+3”) of the domain that coexists in approximately equimolar concentrations (51:49, wt:+3). The difference between the two structures is a +3 shift in register and inversion of a central I $\beta$ -strand accompanied with an isomerization of N448-P449 peptide bond in the preceding HI loop (Figure 4-1). Intriguingly, we found the two conformations of Y456T interconvert slowly enough that the two forms could be separated by anion exchange chromatography, and that they could reestablish the equilibrium over time. However, the kinetics were initially only roughly characterized. Thus, we

set out to directly detect the rates of interconversion for ARNT PAS-B Y456T, and determine how this might compare to the rates of folding/unfolding that might influence this process.



## B. Results for ARNT PAS-B Y456T interconversion and unfolding kinetics

### i. Kinetics of interconversion for ARNT PAS-B Y456T

To measure the kinetics of interconversion, we used time-resolved NMR spectroscopy to record a series of  $^{13}\text{C}$ -edited 1D  $^1\text{H}$  experiments on an enriched fraction for the +3 conformation (88:12, +3:wt), as generated by MonoQ ion exchange chromatography at 4°C, from an initial sample of Y456T (51:49, wt:+3). Preliminary spectra showed peaks resulting primarily from the +3

conformation, as illustrated by the peak at -0.348 ppm, which corresponds to Leu391  $\delta^1$  methyl group in the +3 conformation (Figure 4-2a). This same methyl group in the wt conformation is located at -0.628 ppm. By recording a series of sequential spectra collected approximately four minutes apart, the process of interconversion between the two conformations as the system relaxes back to equilibrium (51:49, wt:+3) was monitored. As a result, time-dependent peak intensity changes could be fit to a two-state interconversion with a single exponential decay, allowing the extraction of the kinetic rates. By plotting normalized peak intensity over time (Figure 4-2b) this slow interconversion can be observed as equilibrium is finally reached after 600 minutes at 284 K. In addition, we established the temperature dependence of this process with comparable experiments conducted between 278 K and 291 K and found a linear Arrhenius dependence, with rates between  $k = 5.49 \times 10^{-5} \text{ s}^{-1}$  (278 K) and  $k = 2.13 \times 10^{-4} \text{ s}^{-1}$  (291 K), respectively (Figure 4-2c). The activation energy barrier can be calculated from the slope of the Arrhenius plot using Equation 4-1:

$$E_a = b \times R \quad (\text{Eq. 4-1})$$

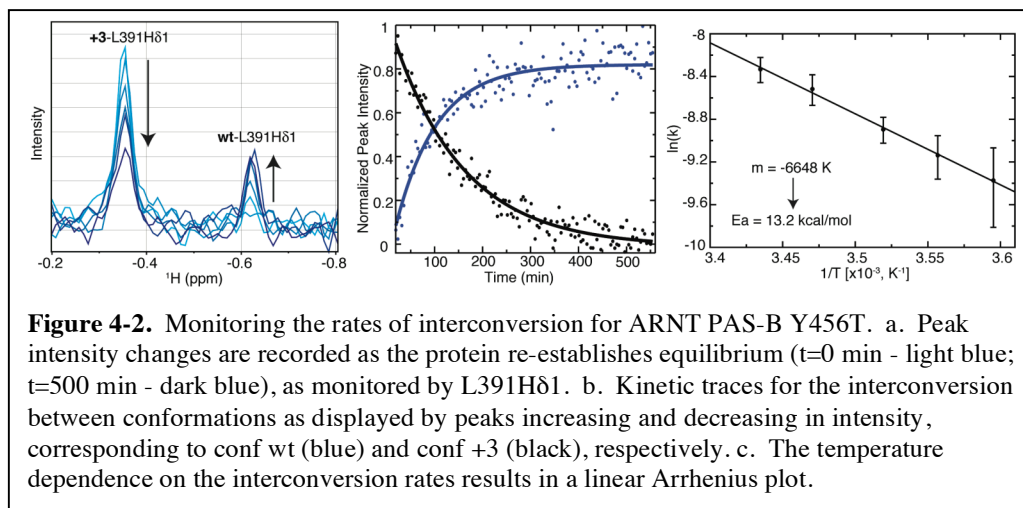
Therefore, analysis of the Arrhenius plot yields an activation energy barrier of 13.2 kcal/mol. In addition, we established the temperature dependence on the equilibrium (detailed in Chapter 5) and the rate of interconversion (Eyring plot) between the two conformations of Y456T (Figure 4-3). Analysis of the Eyring

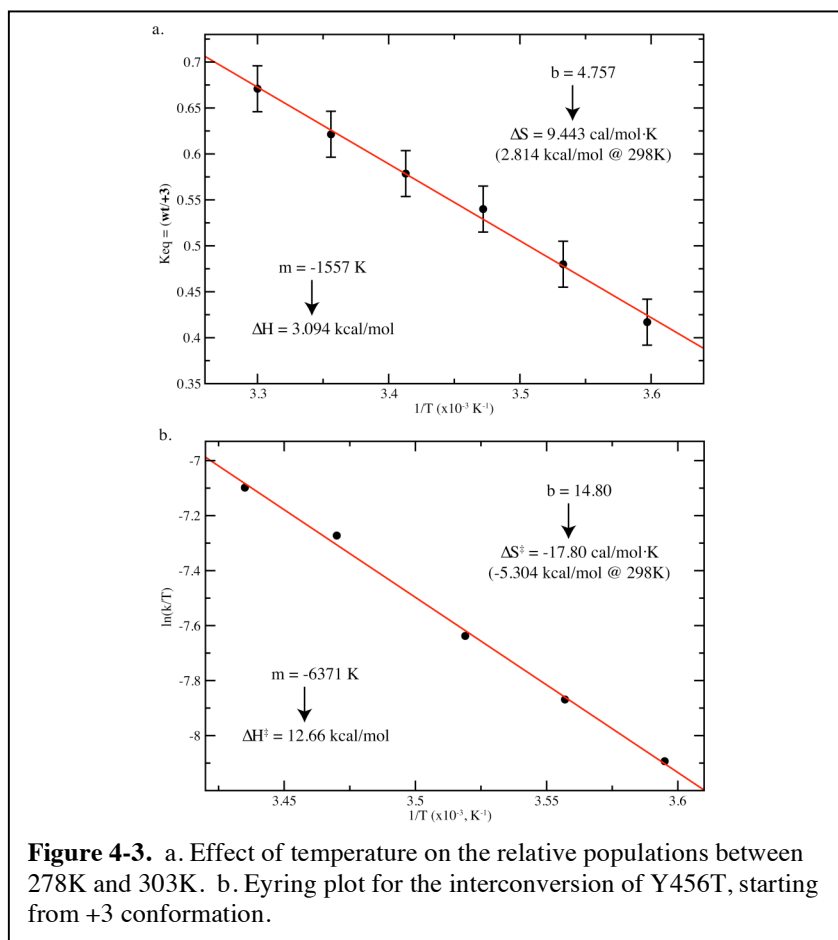
plot yields thermodynamic parameters for the interconversion process, including enthalpy and entropy values derived from Equations 4-2 and 4-3, respectively:

$$\Delta H = -m \times R \quad (\text{Eq. 4-2})$$

$$\Delta S = b \times R \quad (\text{Eq. 4-3})$$

As a result, the Eyring plot suggests that a large enthalpic barrier (12.7 kcal/mol) must be crossed in order to interconvert (Table 4-1). The breaking of hydrogen bonds is a major source of enthalpy barriers. Therefore, we set out to determine the mechanism of interconversion, hypothesizing the domain may partially unfold to interconvert.





**Figure 4-3.** a. Effect of temperature on the relative populations between 278K and 303K. b. Eyring plot for the interconversion of Y456T, starting from +3 conformation.

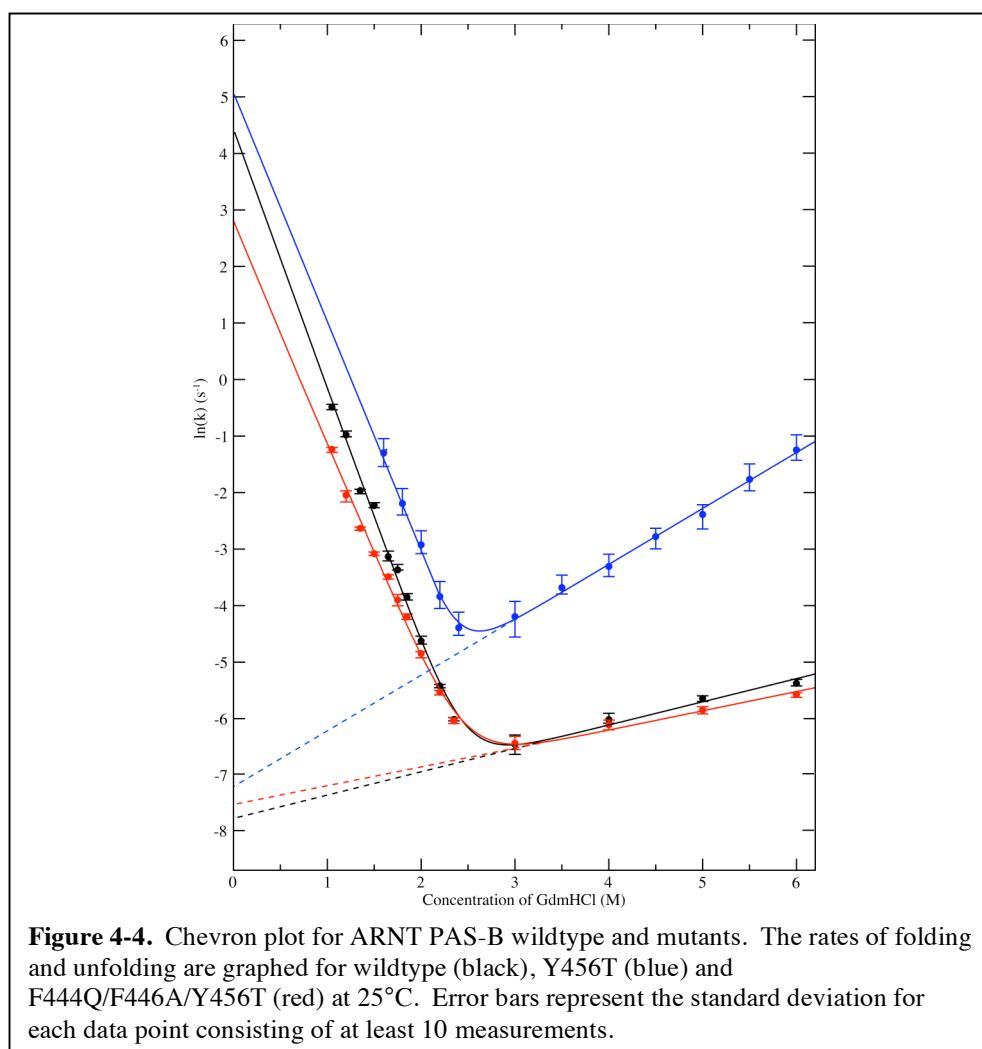
	Thermodynamic Parameter	Value
Equilibrium MUT $\rightarrow$ WT (Fig 4-3a)	$\Delta G$	0.3 kcal/mol
	$\Delta H$	3.1 kcal/mol
	$\Delta S$	9.4 cal/mol·K (2.8 kcal/mol @ 298 K)
Interconversion MUT $\rightarrow$ WT (Fig 4-3b; Eyring)	$\Delta G^\ddagger$	18.0 kcal/mol
	$\Delta H^\ddagger$	12.7 kcal/mol
	$\Delta S^\ddagger$	-17.8 cal/mol·K (-5.3 kcal/mol @ 298 K)
Unfolding (Fig 4-4)	$\Delta G^\ddagger$	21.6 kcal/mol
	$\Delta H^\ddagger$	11.8 kcal/mol
	$\Delta S^\ddagger$	-32.9 cal/mol·K (-9.8 kcal/mol @ 298 K)

**Table 4-1.** Thermodynamic parameters for the ARNT PAS-B Y456T mutant at 25°C.

*ii. Rates of global folding and unfolding for ARNT PAS-B mutants*

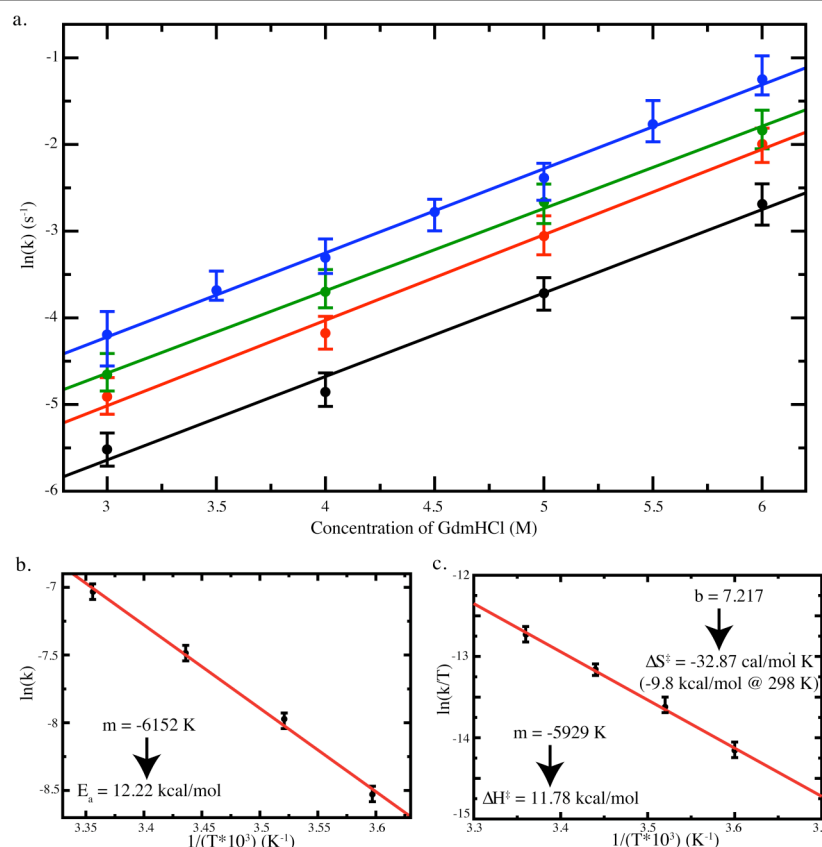
To complement this study of the interconversion rates between conformations, we investigated the rates of global folding and unfolding using stopped-flow fluorescence in combination with Gdn·HCl. As expected, these kinetic rates of folding and unfolding for either conformation resulted in a chevron plot (Figure 4-4), with a denaturant midpoint (~2.5 M Gdn·HCl) that corresponded well to our previously determined equilibrium denaturation midpoint (16). Unexpectedly, for proteins that existed greater than 99% in either the wt or +3 conformations, the unfolding kinetics were relatively slow,  $k_u = 5.53 \times 10^{-4} \text{ s}^{-1}$  and  $k_u = 7.47 \times 10^{-4} \text{ s}^{-1}$  (Table 4-2), respectively, when compared to dozens of other similarly-sized proteins that have been reported with kinetics in the  $k_u = 5.85 \times 10^{-6} \text{ s}^{-1}$  to  $89 \text{ s}^{-1}$  range (166). It should be noted that the rates of folding, extracted back to zero denaturant, for the wt conformation ( $k_f = 52.9 \text{ s}^{-1}$ ) are slightly faster than for the +3 conformation ( $k_f = 10.6 \text{ s}^{-1}$ ) at 298 K and that these values correspond well to protein folding rates of other comparably sized proteins. Repeating these experiments for ARNT PAS-B Y456T yielded similar kinetics ( $k_f = 158 \text{ s}^{-1}$  and  $k_u = 7.99 \times 10^{-4} \text{ s}^{-1}$ ) at 298 K with zero denaturant and suggest the two conformations of Y456T fold and unfold comparably to the rates of proteins locked into either conformation. These experiments were repeated at various temperatures ranging from 278 to 298 K to determine the temperature dependence of the unfolding process (Figure 4-5a).

Arrhenius (Figure 4-5b) and Eyring plot (Figure 4-5c) analysis for the unfolding rates of Y456T resulted in an activation energy barrier of 12.5 kcal/mol with a large enthalpic (11.8 kcal/mol) and entropic (-9.8 kcal/mol at 298 K) barriers to unfold (Table 4-1). The large entropic barrier is presumably due to a drop in solvent entropy upon exposure of non-polar core residues. Interestingly, these thermodynamic values for unfolding are very similar to those measured for the interconversion of Y456T.



Construct	Unfolding Rate ( $s^{-1}$ )	Folding Rate ( $s^{-1}$ )	$\Delta G$ (kcal/mol)
Wildtype	$5.5 \times 10^{-4}$	53	6.9
Y456T	$8.0 \times 10^{-4}$	158	7.3
P449A/Y456T	$2.1 \times 10^{-4}$	193	8.2
F444Q/F446A/Y456T	$7.5 \times 10^{-4}$	11	5.8

**Table 4-2.** Rates of folding and unfolding for ARNT PAS-B wildtype and mutants extracted back to denaturant-free conditions at 25°C and the associated  $\Delta G$  values, which were calculated from the unfolding and refolding rates.



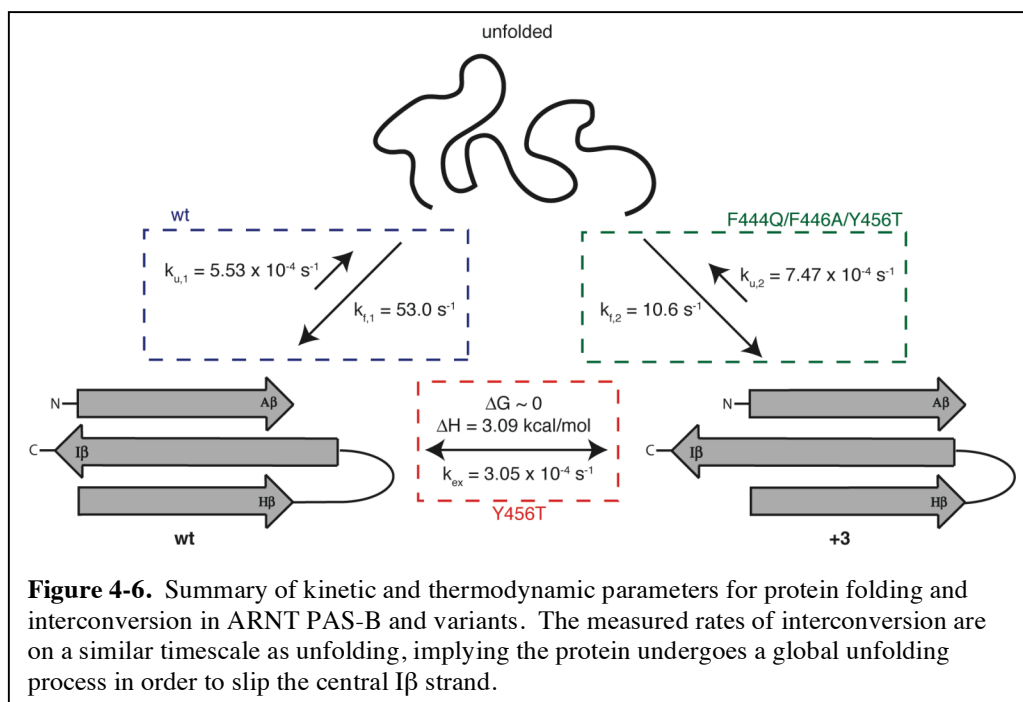
**Figure 4-5.** Measuring the temperature dependence on unfolding rates for ARNT PAS-B Y456T. a. The rates of unfolding over a range of temperatures (black, 278 K; red, 284 K; green, 291 K; and blue, 298 K). Error bars represent the standard error for each data point consisting of at least 10 measurements. b. Arrhenius plot for the unfolding process, which results in a energy of activation barrier of 12.2 kcal/mol. c. Analysis of the Eyring plot for the unfolding process yields an enthalpic barrier of 11.8 kcal/mol and an entropic barrier of -9.8 kcal/mol at 298 K.



### *iii. Model for ARNT PAS-B Y456T interconversion*

Comparing the rate of interconversion to the rates of unfolding, one can develop a simple model for the process by which the two conformations of Y456T interconvert. We have demonstrated that ARNT PAS-B Y456T is able to interconvert between two structures on the timescale of minutes ( $\tau \sim 40$  min at 298 K, Figure 4-2). Additionally, we have shown that this same mutant unfolds on a similar timescale in denaturant-free conditions ( $\tau \sim 22$  min at 298 K, Figure 4-4). Upon further inspection, these rates of unfolding are approximately two times faster than the rate of interconversion (Figure 4-6). These data are consistent with the protein unfolding to a chiefly disordered state as it interconverts between conformations. This result fits well with our model, as the unfolded state of ARNT PAS-B Y456T has an equal probability of refolding into either folded conformation, as established by the 51:49, wt:+3 equilibrium. This equal probability of refolding into either conformation makes the rates of interconversion appear to be twice as slow than if the interconversion was limited to a single direction. Additional evidence to support this model is found by examining the temperature dependence on the ARNT PAS-B Y456T mutant interconversion and unfolding rate. This study shows a similar temperature dependence on the interconversion and unfolding rates, and equivalently comparable energies of activation ( $E_a = 13.2$  kcal/mol and 12.5 kcal/mol,

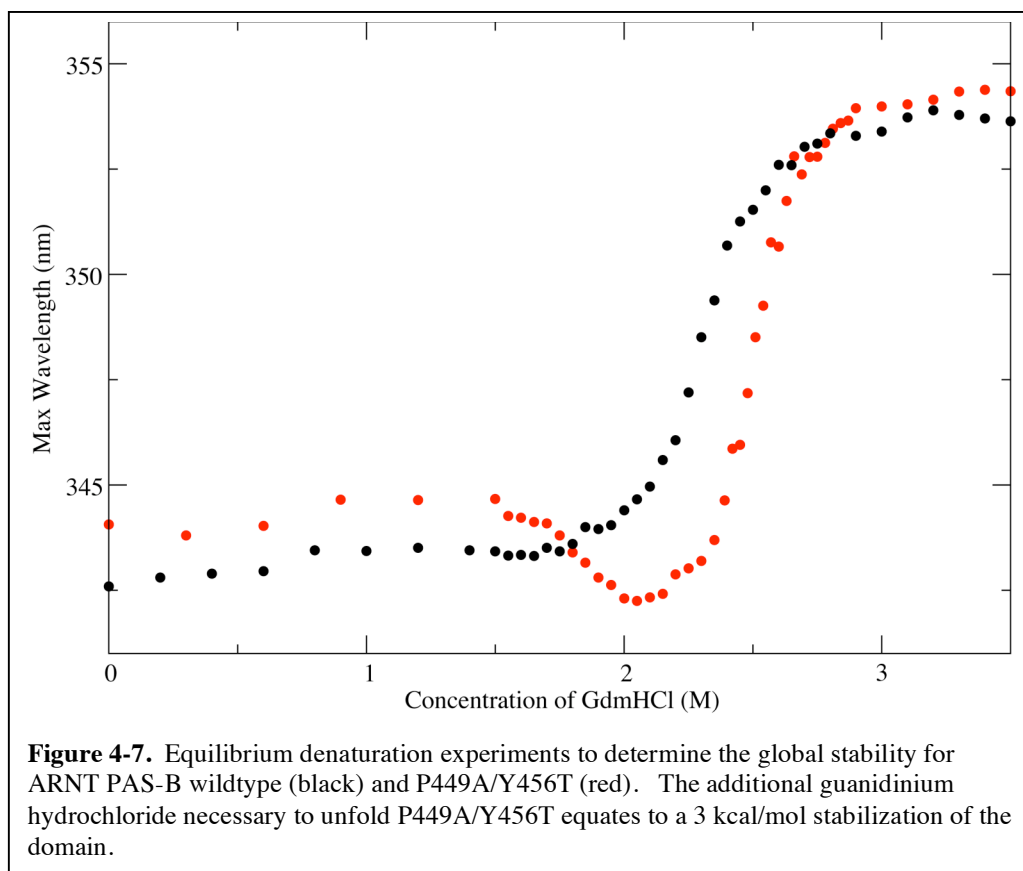
respectively), further supporting the idea that both processes (interconversion and unfolding) transition through the same chiefly, unfolded state.



#### iv. Stability of ARNT PAS-B mutants

Previously reported equilibrium denaturation and  $^2\text{H}$  exchange protection factors indicated the wildtype and +3 conformations were relatively similar, review Chapter 3. Therefore, we determined the free energy change for ARNT PAS-B wild-type, Y456T and F444Q/F446A/Y456T mutant proteins to unfold is approximately 9 kcal/mol. This value corresponds well to reported values between 5 and 10 kcal/mol for small, globular proteins (167). Interestingly, the P449A/Y456T mutant stabilized the folded structure over the wildtype protein by

3 kcal/mol (Figure 4-7). These results suggest the HI loop of ARNT PAS-B plays an important role on the stability of the entire domain.



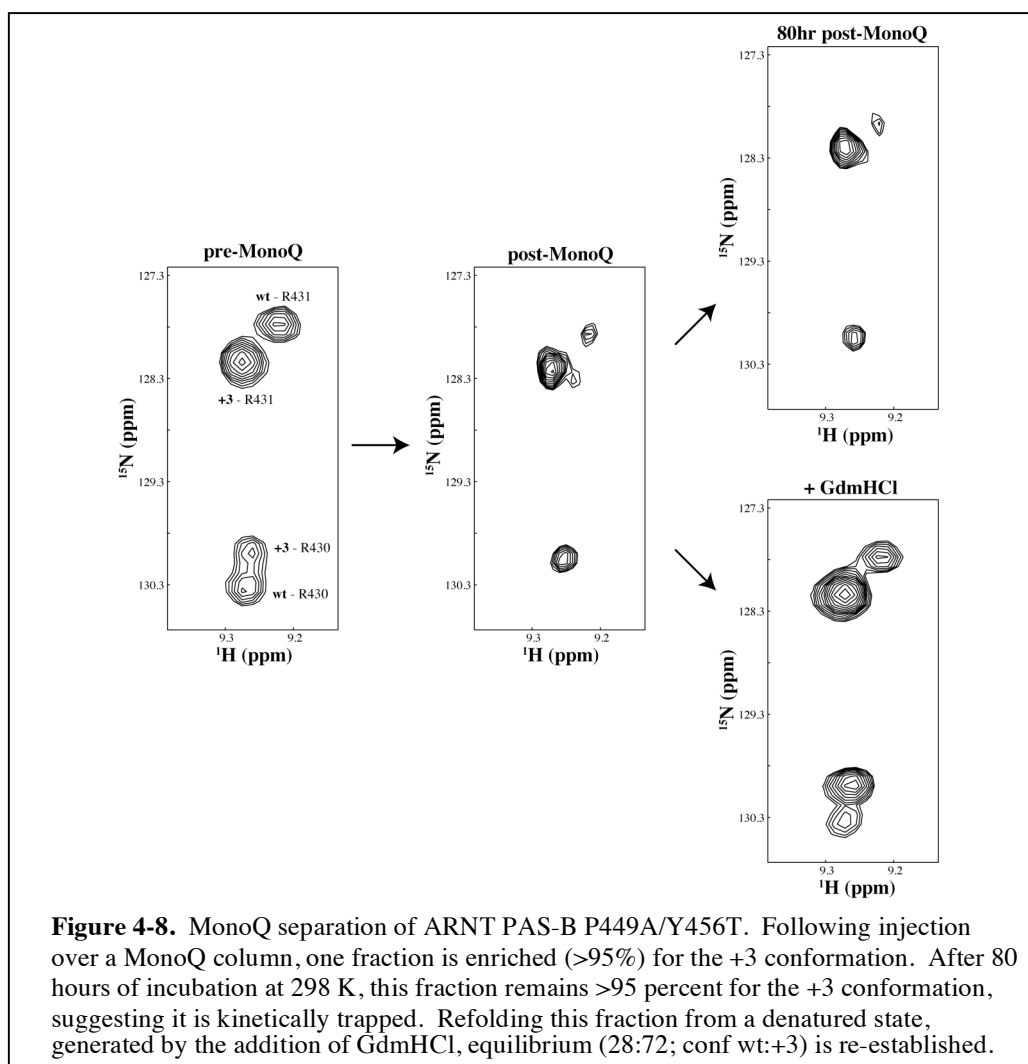
### C. The effect of a N448-P449 peptide bond isomerization on the kinetics

#### i. Results

Previously, we hypothesized that the N448-P449 peptide bond in the HI loop preceding I $\beta$  may have a significant kinetic effect on the rates of interconversion between conformations as this bond isomerizes from trans to cis

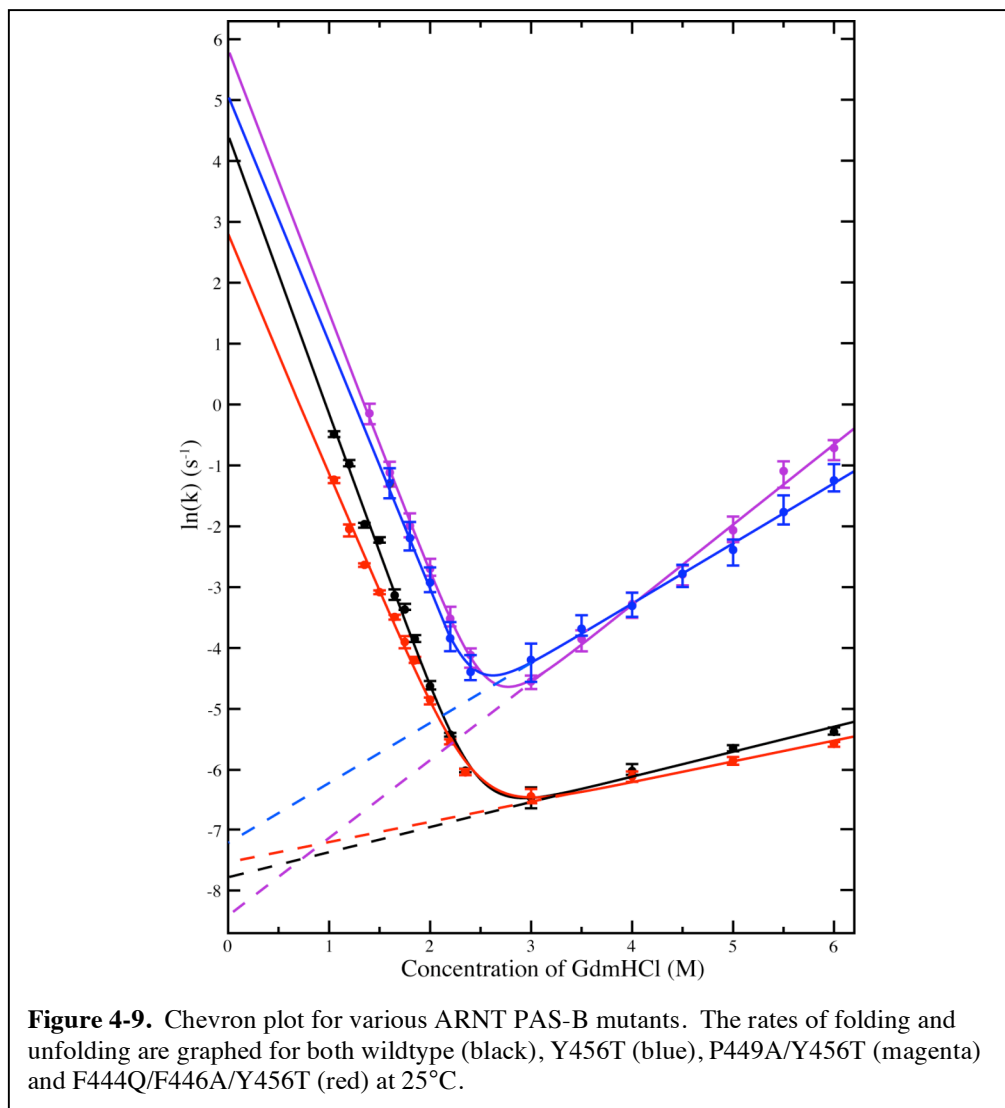
in the wt and +3 conformations, respectively (16). The cis-trans isomerization of X-Pro peptide bonds have been well documented for their key role in the rate-limiting step for various protein folding processes (168). Intriguingly, the energy barrier for interconversion in Y456T is close to the barrier of a X-Pro peptide bond isomerization between the cis-trans configurations, which has been reported for peptides in the range of 13-20 kcal/mol (169). Therefore, we speculated that isomerization of the N448-P449 peptide bond was the rate-limiting step for the interconversion between the two conformations of Y456T. To test this, we generated a P449A/Y456T variant that resulted in a well-folded domain and slightly shifted the equilibrium (32:68, conf wt:+3, Appendix Table 1). Thus, to test our model, we measured the rates of interconversion for P449A/Y456T. We repeated our anion exchange chromatography and purified a fraction (>95%) of the +3 conformation and monitored interconversion using time-resolved 1D  $^{13}\text{C}$  edited  $^1\text{H}$  NMR experiments. Surprisingly, the enriched fraction did not exchange back to equilibrium after over 300 hr suggesting this fraction is kinetically trapped (Figure 4-8). To our surprise, this ARNT PAS-B mutant started to slowly precipitate out of solution 100 hr post separation. Despite this precipitation, total peak intensities in  $^{15}\text{N}/^1\text{H}$  HSQC spectra decreased but the relative populations remained the same, which indicated the protein remained >95% in the +3 conformation. These data suggest once the protein folds into either conformation, it becomes kinetically trapped into that state. However, this kinetic trap can be

escaped by completely unfolding the protein into 3.5 M GdmHCl and refolding the protein by dialyzing the mutant to a final guanidium hydrochloride concentration of 1 M. As such, the equilibrium is restored to 28:72, conf wt:+3, which is near the initial 32:68 equilibrium (Figure 4-8). This experiment confirms that the two conformations of P449A/Y456T remain in equilibrium but are somehow kinetically trapped from interconverting between conformations.



Since P449A/Y456T has a large impact on the equilibrium and kinetics of interconversion, we wanted to determine the effect this mutation may have on the kinetics of folding and unfolding. Therefore, we repeated our stopped-flow experiments to experimentally determine the folding and unfolding rates.

Analysis of the chevron plot for P449A/Y456T (Figure 4-9) supports our model as this mutant shows an increase in the folding rate ( $k_f = 193 \text{ s}^{-1}$ ) and a decrease in the unfolding rate ( $k_u = 2.08 \times 10^{-4} \text{ s}^{-1}$ , Table 4-2). The change in rates associated with the P449A/Y456T mutant equates to a stabilization of 1 and 2 kcal/mol over the wildtype and F444Q/F446A/Y456T proteins, respectively (Table 4-2). These data suggest the proline at position 449 slows down the unfolding rate four-fold and increases the folding rate 23% when comparing P449A/Y456T to Y456T. These data are consistent with proline isomerization having an effect on the rates of interconversion between conformations potentially by slowing the folding rate and increasing the unfolding rate.



## ii. Discussion

For many proteins, the reorganization of secondary structures between different conformations is relatively quick, generally on the millisecond timescale. For example, after illumination of *Arabidopsis thaliana* phototropin

LOV2 domain with light, a small volume expansion for the domain was observed with a time constant of 9 ms (170). Some secondary structural changes have even been recorded on the picosecond timescale. For instance, myoglobin reorganizes about the docked CO ligand with a time constant of 1.6 ps (171). For this reason, one would expect the rates of interconversion to be relatively fast. However, the slow interconversion rates between ARNT PAS-B “wildtype like” and +3 conformations observed for ARNT PAS-B Y456T was expected, due to the absence of cross peaks in ZZ-exchange studies (Chapter 3).

To better understand the mechanism by which the two conformations interconvert, we considered two general models. The first model proposes only a local unfolding occurring around the I $\beta$ -strand while most of the protein remains unchanged. As such, the protein breaks 10 hydrogen bonds between the I $\beta$  and neighboring strands, slips down three residues and reforms these hydrogen bonds. The second model proposes a global unfolding process in order to interconvert. In this case, the protein must undergo a global unfolding process to a chiefly unfolded state and that the sequence of the protein allows for equal probability of refolding into either conformation for the Y456T point mutant, and presumably different probability for other variants. To test these models, stopped-flow unfolding experiments in the presence of GdmHCl were conducted to determine the rates of unfolding. Interestingly, the unfolding rates were roughly two-fold faster than the rates of interconversion, suggesting it may unfold in order to



interconvert. This result fits well with our second model, as the unfolded state of ARNT PAS-B Y456T has an equal probability of refolding into either folded conformation for the Y456T variant, which makes the rates of interconversion appear to be twice as slow than if the interconversion was limited to a single direction. Therefore, we concluded from these data that most likely ARNT PAS-B Y456T unfolds to a chiefly unfolded state, and refolds into both conformation with equal probability, re-establishing the 51:49 (wt:+3) equilibrium.

The proline residue at position 449, located in the HI-loop of ARNT PAS-B, has an effect on the rate of interconversion that may reveal a slow folding/unfolding pathway. These isomerization events have been described as molecular switches as they have shown to be crucial for opening/closing membrane channels (172), mediating conformer-specific ligand recognition (173), and infection of *E. coli* by phage (174), all of which occur on the seconds to minutes timescale. While the physiological relevance of this Asn-Pro isomerization in ARNT PAS-B Y456T remains to be determined, it is clear this peptide bond isomerization has some drastic effects on the rates of the reactions in a similar manner as observed in other biological systems.

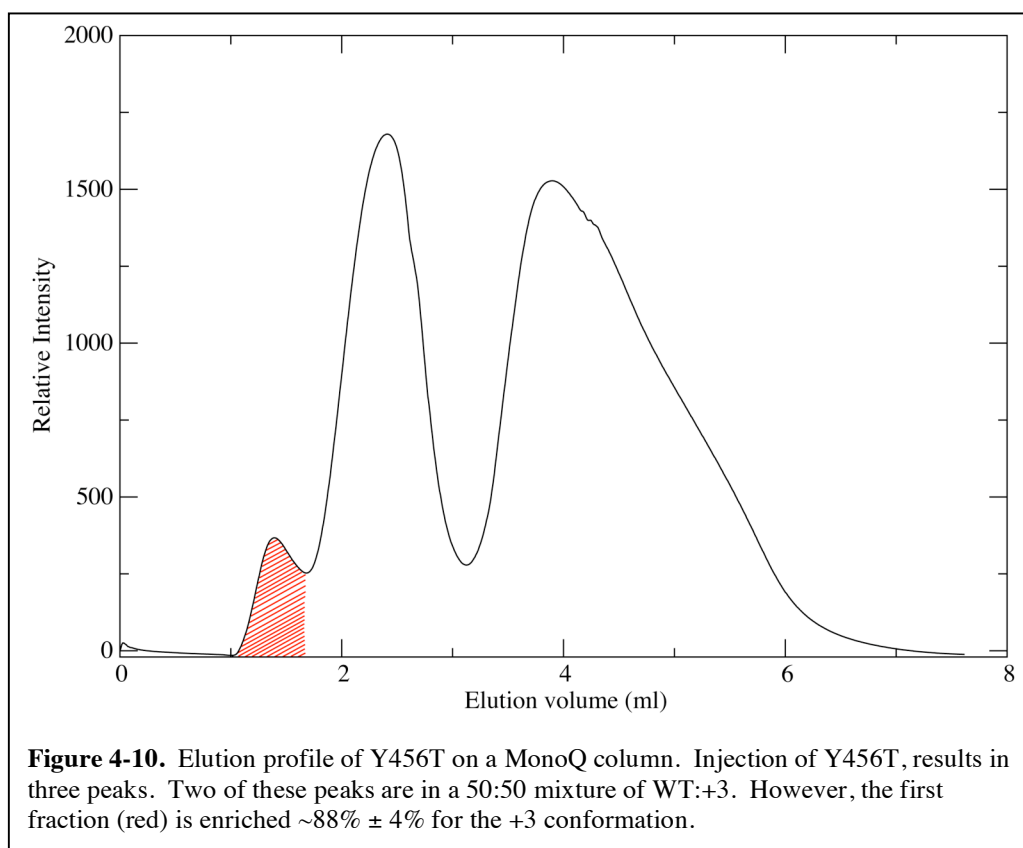
#### *D. Materials and Methods*

##### *i. Measuring the interconversion rate of ARNT PAS-B Y456T*

Point mutants of ARNT PAS-B were created, expressed and purified using methods described previously (16). 400  $\mu$ L of 1.2 mM ARNT PAS-B Y456T protein suspended in 50 mM Tris (pH 6.6), 17 mM NaCl and 5 mM DTT buffer was injected over a 1 mL MonoQ column located in a 4°C refrigerator at a rate of 0.5 ml per minute. The chromatogram resulted in three peaks, in the absence of any salt gradient (Figure 4-10). The first eluted peak fraction yielded 350  $\mu$ L of a 40  $\mu$ M sample, which was enriched ~88% for the +3 conformation, as established by  $^{15}\text{N}$ - $^1\text{H}$  HSQC spectra. The two peaks that eluted after this first fraction remained in an equal mixture. Due to the low protein concentration of the first fraction, multiple injections were required to obtain a reliable measurement of the relative intensities. Upon elution from the column, the protein was placed on an ice slurry to slow any interconversion while an additional two FPLC runs were conducted. The total time to run all three FPLC experiments was approximately 15 minutes. The first fraction from all three runs were pooled together and concentrated to a final volume of 500  $\mu$ L, which typically equated to a 140  $\mu$ M concentration. To directly detect the rates of interconversion, NMR experiments were carried out on a Varian Inova 600 MHz NMR spectrometer with samples of uniformly  $^{15}\text{N}$ ,  $^{13}\text{C}$ -labeled ARNT PAS-B

Y456T mutant in 50 mM Tris (pH 6.6), 17 mM NaCl and 5 mM DTT solutions.

A series of 1D  $^{13}\text{C}$  edited proton experiments were recorded until equilibrium was established, with the total time set to four times the expected time constant for each temperature. Peak intensities over time were plotted using the DosyView utility in NMRDraw (175) and the respective rates were measured. Chemical shift assignments for ARNT PAS-B Y456T were transferred from previously solved NMR structures for the wt (7) and +3 conformations (16).



*ii. Stopped-flow Experiments*

Rapid mixing experiments were conducted using a Bio-Logic SFM-3 stopped flow apparatus equipped with a 1 cm path length quartz cell. All experiments were carried out at 25°C in 50 mM sodium phosphate buffer (pH 7.5) and 17 mM NaCl. A monochromator was used for excitation at 280 nm and the intensity of fluorescence emission was recorded. Experiments were controlled by BioKine 16 V 3.03 software and had an estimated dead time of 6 ms. Kinetic rates were measured by fitting the average of 10-18 traces per sample point to a single exponential curve, using BioKine 16 software. Twenty-four hours prior to recording the refolding rates, the proteins were denatured to 4M Gdn·HCl and placed at 25°C to ensure equilibrium was achieved.

*iii. Measuring the interconversion rate of ARNT PAS-B P449A/Y456T*

The procedure by which the +3 conformation of ARNT PAS-B P449A/Y456T was separated using ion exchange column is the same as that detailed for Y456T under the same solution conditions. Since this enriched fraction was kinetically trapped for at least 100 hours, the experiment was repeated and this time, after 24 hours at 298 K, GdmHCl was added. The protein was completely denatured for two hours in the presence of 3.5 M GdmHCl. The sample was diluted to a final concentration of 1.0 M GdmHCl and re-concentrated down to 500 µL and analyzed by <sup>15</sup>N/<sup>1</sup>H HSQC spectra on the refolded fraction.

*E. Acknowledgements*

We thank John Richardson and Nicholas Malmquist for their technical assistance as well as Elliott Ross for his laboratory space and equipment. This research was supported by a grant from the NIH (R01 GM081875) to K.H.G.

## V. Factors affecting the equilibrium of ARNT PAS-B Y456T

Equilibrium is defined as the condition of a system in which two opposing processes occur simultaneously at the same rate. If the balance of this system is altered, such as a change in temperature, the equilibrium constant shifts to accommodate this change. Henry Le Chatelier studied this process and tried to predict how chemical equilibrium is affected as a result of a change in solvent conditions. Le Chatelier's principle states that if a system at chemical equilibrium experiences a change in temperature or total pressure, the equilibrium shifts in order to counter-act that change. Therefore, we set out to determine what factors could affect the 51:49 (wt:+3) equilibrium of the ARNT PAS-B Y456T mutant, including solution conditions, temperature, and binding of small molecule ligands which preferentially associate with one conformation. Additionally, we found that not only temperature but also compounds that preferentially bound the "wildtype-like" conformation of Y456T could affect the wt:+3 equilibrium.

### *A. Factors that affect equilibrium*

#### *i. Solution conditions affect equilibrium*

Since ARNT PAS-B Y456T exists in a roughly equal distribution between the wt and +3 conformations, it presents a simple system to test the effects that

different solution conditions have on this equilibrium. To this extent, ARNT PAS-B Y456T was buffer exchanged into various buffers over a 5.5 - 9.0 pH range. We found that pH effects on the equilibrium were minimal, with only a slight stabilization of the wildtype conformation at lower pH values (Table 5-1). However, no significant (greater than a 5% deviation) changes were observed for the equilibrium constant, which was measured as the average peak intensity for a minimum of eight amides. This result is not surprising considering the slight change in solvent exposed sidechains between the two conformations (Figure 3-6).

Buffer	pH values	Percent WT	Percent +3
Tris	6.5	53	47
Tris	7.0	51	49
Tris	7.5	51	49
Tris	8.0	51	49
Tris	8.5	53	47
Tris	9.0	54	46
PIPES	5.5	56	44
PIPES	6.0	52	48
PIPES	6.5	50	50
PIPES	7.0	48	51
PIPES	7.5	47	53
PIPES	8.0	47	53

**Table 5-1.** Relative populations for ARNT PAS-B Y456 in Tris or PIPES over a broad range of pH values. Approximately 6% relative error on any given ratio for an average of eight sites.

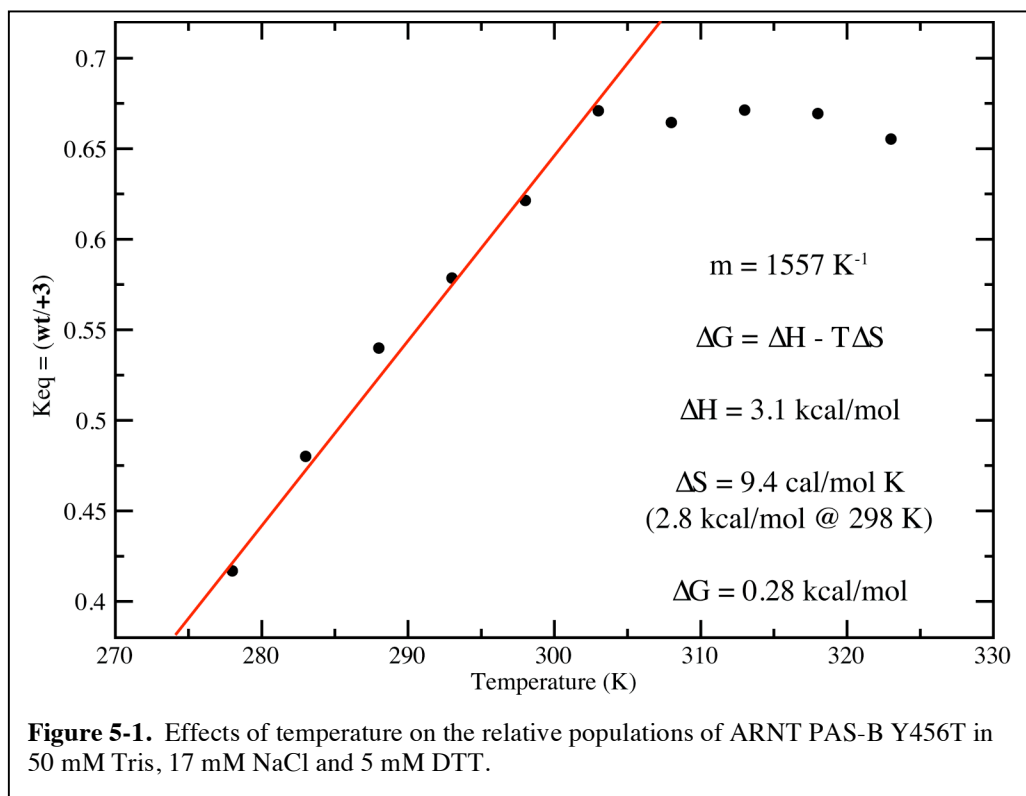
*ii. Temperature dependence of the equilibrium*

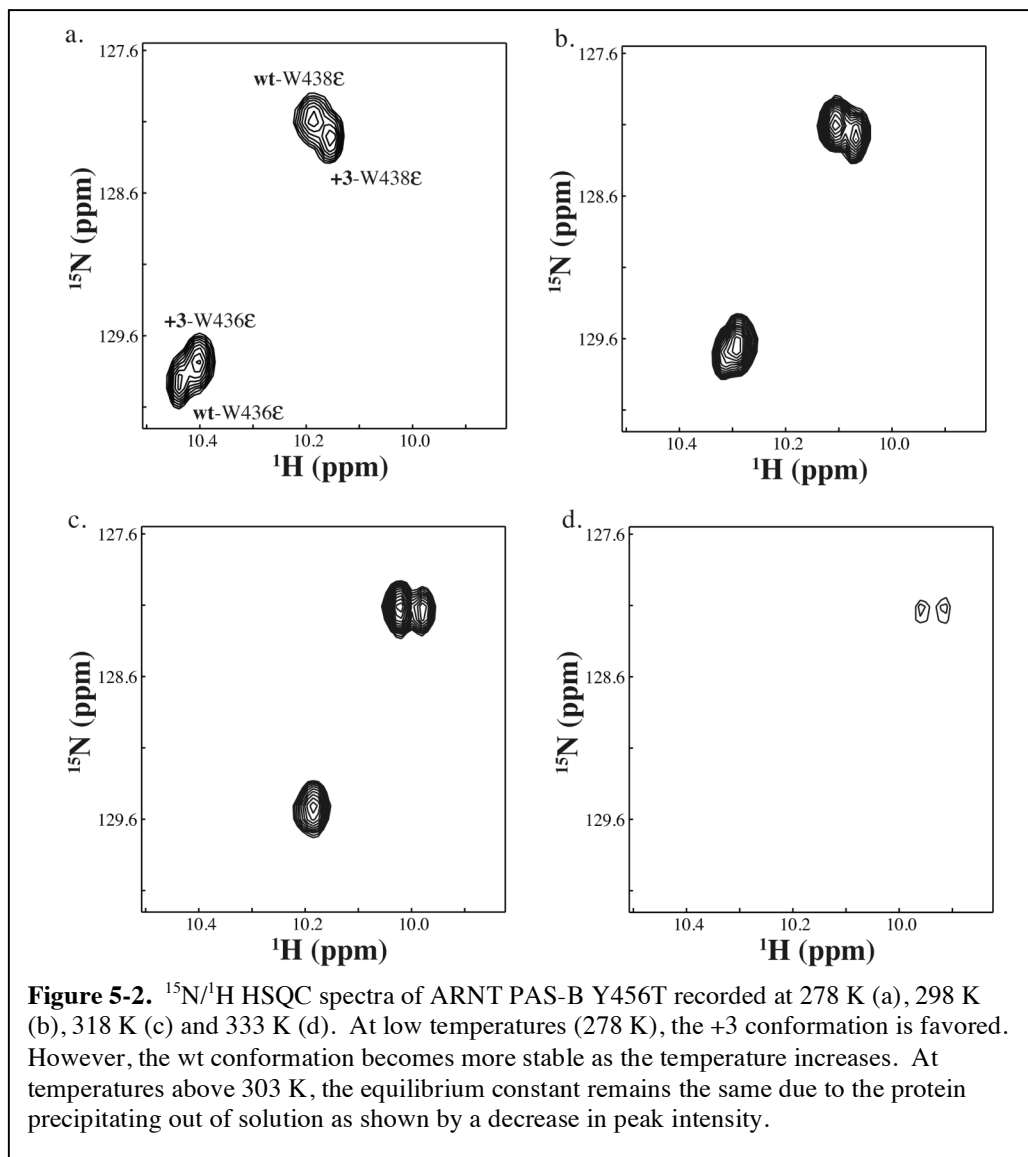
Previous experiments indicate the interconversion from the +3 conformation to the wildtype is an endothermic process ( $\Delta H = 3.1$  kcal/mol, Table 4-1), as indicated by a positive free energy change and enthalpy change (Table 4-2). Therefore, one would predict at higher temperatures, the wildtype conformation would be favored and vice-versa at lower temperatures. For this reason,  $^1\text{H}$ - $^{15}\text{N}$  HSQC experiments of ARNT PAS-B Y456T were recorded between 278 K and 333 K, to determine the temperature dependence on the equilibrium. I found there is a linear dependence of temperature on the equilibrium values from 275 to 300 Kelvin. As expected, at temperatures below 285 K the +3 conformation is favored, whereas at higher temperatures the wildtype conformation is favored (Figure 5-1). This linear dependence on the equilibrium values in this region also provides information on the thermodynamic parameters. As such, the enthalpy and entropy values were derived from Equations 4-2 and 4-3, respectively. As a result, the enthalpy change is calculated as 3.0 kcal/mol and the entropy change is 9.44 cal/mol\*K (2.8 kcal/mol at 298 K), which equate to a free energy change of 0.28 kcal/mol at 25°C (described in further detail in Chapter 4). Above 300 K, the equilibrium values level off and remain constant at 2:1 (wt:+3). This result is most likely due to the protein aggregation, as seen by a decrease in peak intensity across the spectrum,



preventing the protein from establishing its true equilibrium between states

(Figure 5-2).



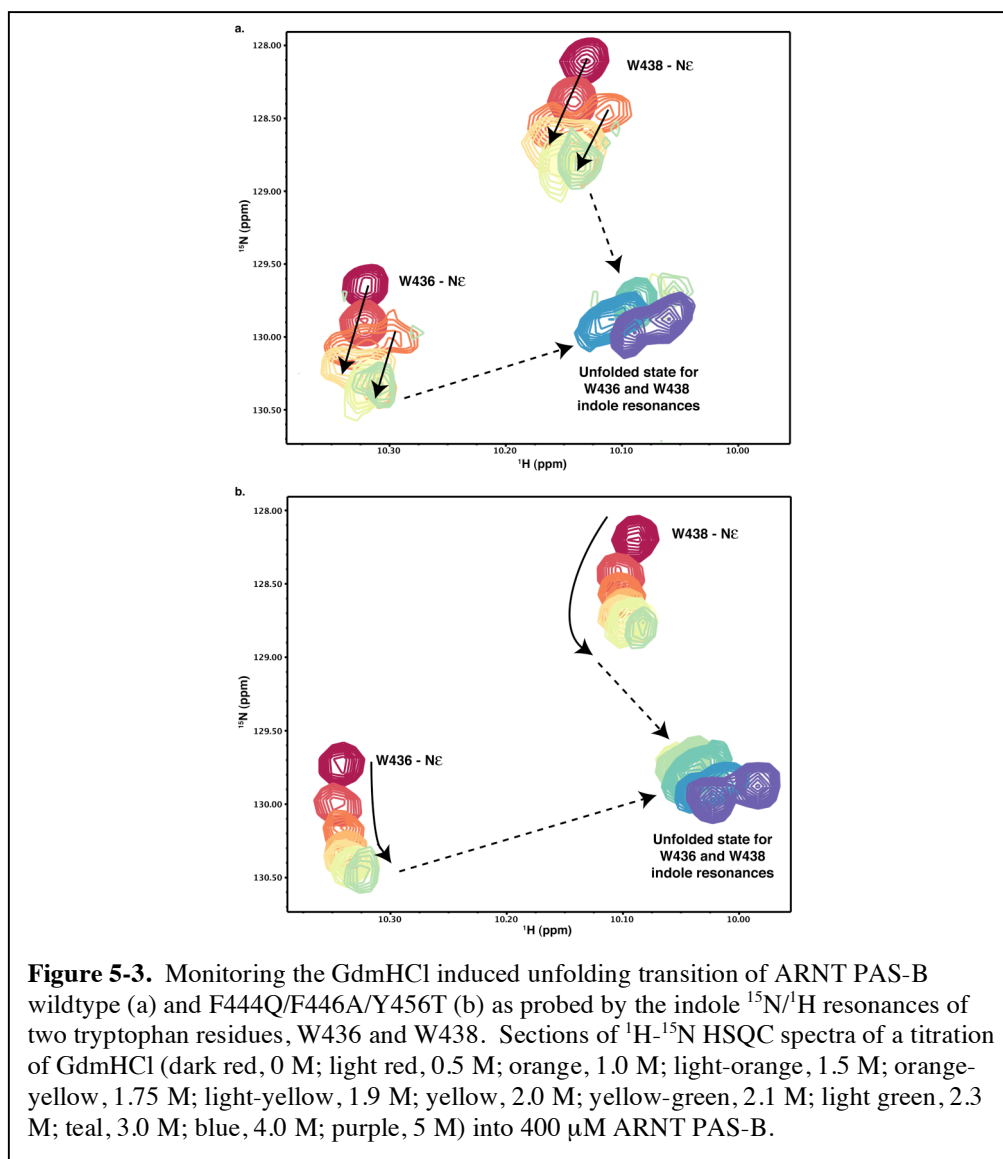


*iii. Denaturant studies suggests both conformations are equally stable*

Previous equilibrium and kinetic guanidinium denaturation studies revealed the two individual conformations, wildtype and +3, undergo very similar unfolding transitions (described in further detail in Chapter 3). With the help of

Kyle Brewer, we set out to examine the denaturant dependence using  $^{15}\text{N}$ - $^1\text{H}$  HSQC experiments to get site-specific information about the unfolding transitions of both conformations. As expected, both wildtype and triple mutant proteins showed similar chemical shift changes throughout the unfolding process. For both proteins, the addition of guanidinium hydrochloride to start the unfolding transition (1.9 M GdmHCl) resulted in chemical shift changes in fast exchange (Figure 5-3). Chemical shift differences between wildtype and the +3 conformations are thought to result from solvent effects and/or the protein fold loosening up from its rigid, packed structure. Interestingly, the wildtype protein showed peak doubling for nearly every amide resonance in the presence of 0.5 to 1.5 M GdmHCl. This result suggests the addition of GdmHCl allows the structure to adopt two different conformations, which are likely in a slow equilibrium with one another, as observed by the two tryptophan indole sites. At higher (>2.1 M) GdmHCl concentrations, the apparent peak doubling is lost but the proton chemical shifts remain well dispersed, which suggests the protein remains folded. In contrast, this peak doubling is not observed for the +3 conformation, suggesting that it starts to unfold along a single pathway. As guanidinium hydrochloride was titrated through the midpoint of the unfolding transition (2.3 M GdmHCl), both the “wildtype like” and +3 conformations exhibit slow exchange behavior with an unfolded state. This unfolded state is evident by the lack of amide  $^1\text{H}$  dispersion in the  $^{15}\text{N}/^1\text{H}$  HSQC spectrum

suggesting the local environment around each residue is similar, as expected for an unfolded protein. Since both the “wildtype like” and +3 conformations unfold with similar concentrations of GdmHCl, it is reasonable to conclude that neither conformation would be favored with different amounts of guanidium hydrochloride present.



*B. Effects of various point mutations on the equilibrium constant*

*i. Additional mutations at Y456*

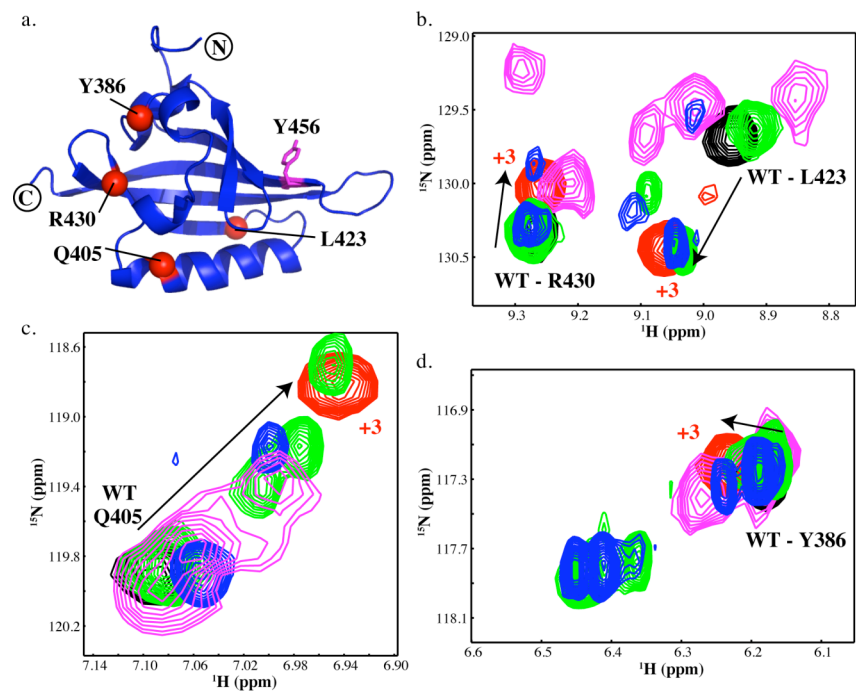
The I $\beta$ -strand of ARNT PAS-B consists of a mixture of polar and hydrophobic residues of similar size, except for Y456 (Figure 3-6). By mutating this tyrosine residue to threonine, the steric hindrance experienced by this residue in the +3 conformation is removed, allowing the strand to slip. However, while removal of this steric hindrance at Y456 is necessary for the strand to slip, it is not sufficient for maximal slippage, as seen in the Y456S (wt:+3; 81:19) and Y456A (wt:+3; >98:<2) mutants (Table 3-1, Appendix 1-1). Both of these mutants favor the wildtype conformation compared to Y456T despite having sidechains smaller than threonine. These data suggest that interactions from neighboring residues to Y456 also are critical to determine the equilibrium position between these two.

To better understand the interactions at position 456 that determine the relative populations of both conformations, this residue was mutated to a number of different amino acids of varying size and charge. Notably, with the help of a rotation student, Amy Zhou, we discovered that placing a charged or larger polar large amino acid at this position (Y456D and Y456Q) resulted in the protein staying in the wildtype conformation (Table 5-2). This is not unexpected, since should the strand slip and adopt the +3 conformation, it would force a larger sidechain residue to take the place of a previously occupied cysteine residue in the

core of the domain. Surprisingly, mutating this residue to a small residue (Y456A, Y456C) also existed solely in the wildtype conformation. This result is most likely due to the lack of side-chain interactions with neighboring residues that could stabilize the +3 conformation. Furthermore, when Y456 was mutated to a non-polar beta-branched residue (Y456F, Y456I, Y456L and Y456V), multiple conformations existed for nearly every residue, as observed in  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra (Figure 5-4). These spectra show peaks for both the wildtype and +3 conformations, as well as additional peaks that arise from other conformations of the domain. The total number of peaks observed for Y456V and Y456I doubled compared to the wildtype spectra and quadrupled for Y456F and Y456L mutants. This is clearly seen for residues Q405 and Y386, which are located at greater than  $18\text{\AA}$  from the sites of mutations. This suggests these new peaks may arise from intermediate conformations between the two states.

Mutation	Percent WT	Percent MUT	Percent Additional Peaks
Y456A	>98	<2	-
Y456C	>97	<3	-
Y456D	>97	<3	-
Y456W	>97	<3	-
Y456Q	>96	<4	-
Y456S	81	19	-
Y456T	50	50	-
Y456F	53	27	13, 8
Y456I	41	-	59
Y456L	40	37	14, 10
Y456V	37	-	42, 21

**Table 5-2.** Relative populations for all ARNT PAS-B Y456 mutations. Approximately 6% relative error on any given ratio, based on 8-12 measurements.



**Figure 5-4.** Mutations that place a beta-branched residue at position Y456 produce multiple conformations of ARNT PAS-B. a. A ribbon diagram of the solution structure of wildtype ARNT PAS-B domain (PDB code: 1X0O) (7). Mutations at position Y456, which the sidechain is displayed in magenta, causes global conformational changes. Widespread peak doubling is observed throughout  $^{15}\text{N}/^1\text{H}$  HSQC spectra as probed by Y386 (d), Q405 (c), L423 (b) and R430 (b). Proteins: wildtype (black), F444Q/F446A/Y456T (red), Y456F (green), Y456L (blue) and Y456V (magenta).

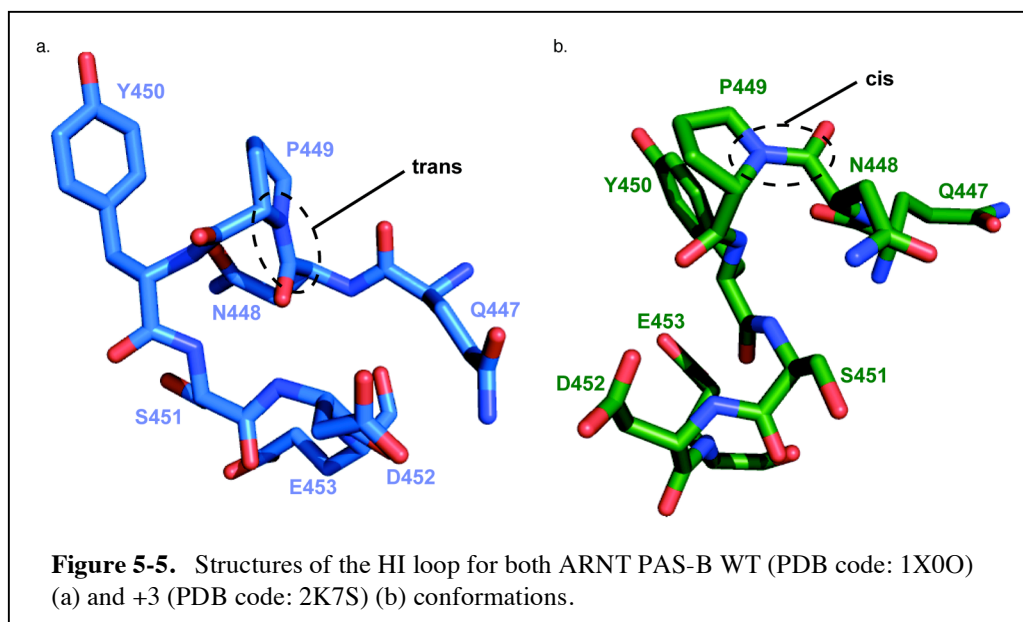
Upon further inspection, the presence of multiple peaks suggests exchange processes are occurring among several conformations. When an NMR-active nucleus exchanges among different environments, it will exhibit different chemical shifts. The impact of this exchange on peak lineshapes is dependent on the rate of this exchange process ( $k_{ex}$ , expressed in  $s^{-1}$ ) compared to the difference in the resonant frequencies of these states ( $\Delta\nu$ , measured in  $Hz = s^{-1}$ ). As such, the exchange regime for this resonance can be defined as fast, intermediate or slow. If the rate of exchange is fast,  $k_{ex} \gg \Delta\nu$ , the signal leads to a single sharp peak, representing the population-weighted average of the two chemical shift resonances. On the other hand, if the exchange process is slow,  $k_{ex} \ll \Delta\nu$ , the signal splits into two separate peaks with amplitudes determined by the population of both states. In intermediate exchange, where the rate of the exchange process is roughly equal to the difference in resonance frequencies,  $k_{ex} \cong \Delta\nu$ , and the spectrum results in two significantly broadened peaks. These peaks are often broadened to the point that no detectable signal can be measured. The presence of multiple peaks for proteins containing beta-branched residues at position 456 suggests slow exchange processes occur in these proteins. The closer the chemical shifts of the intermediate peak are to the wildtype like peak, the more similar the intermediate structure is to the wildtype structure. While the structures of these intermediate conformations remain to be determined, it is intriguing to see how flexible the domain is to subtle mutations on the surface of



the protein, stabilizing nearby conformations that are apparently not the native fold.

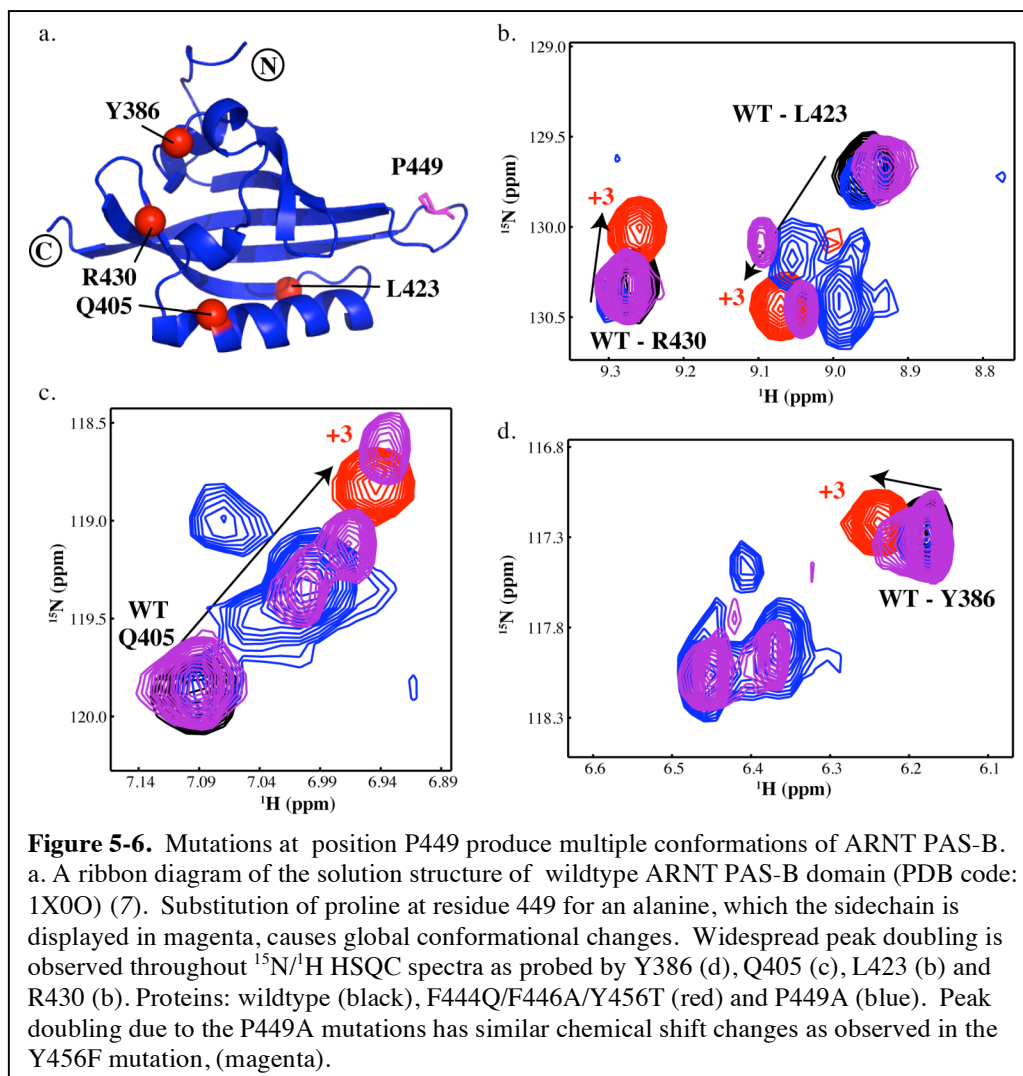
*ii. Mutations in the HI-loop: Role of P449*

While residues in the  $\beta$ -sheet play a major role in determining the relative populations between the two conformations of ARNT PAS-B, it is not the only determinant. In addition to the slipped strand, the preceding HI-loop shortens three residues, which is accompanied by an isomerization of the N448-P449 peptide bond in this region (*16*) (Figure 5-5). To further determine the role of the loop, and more specifically the P449 residue, on the equilibrium between the two conformations, a mutagenesis study was conducted. It has been well documented that the X-Xnp peptide bond, where Xnp represents any non proline residue, only occurs in the cis conformation less than 1% of the time in the PDB (*176, 177*). There have been only a few X-Xnp peptide bonds identified as existing in the cis conformation, one of which is found in the P202A variant of human carbonic anhydrase II. In this structure, the Pro201-Pro202 peptide bond exists in the cis conformation, and substitution of Pro202 for a alanine residue, retains the cis-peptidyl linkage (*178*). On the other hand, X-Pro peptide bonds have been observed to adopt the cis conformation approximately 5 percent of the time (*179, 180*).



To determine the importance of a proline residue at position 449 on the equilibrium of various conformations of ARNT PAS-B, proline was mutated to an alanine to keep the N448-A449 bond in trans while also providing flexibility in the HI loop. As a result, multiple peaks existed for each amide residue for a total of three times as many peaks as the wildtype protein (Figure 5-6). Furthermore, the new peaks that arise due to the mutation overlay well with the ARNT PAS-B Y456F mutation (Figure 5-6). This suggests that both P449A and Y456F mutations each allow the ARNT PAS-B domain to adopt a new, stabilized conformation. In addition, mutating P449 to either Ala or Gly in the Y456T background shifted the equilibrium from the initial 50:50 (wt:+3) equilibrium towards the +3 conformation with P449A/Y456T at 32:68 and even further for P449G/Y456T mutant, 9:91. These results suggested that the proline to alanine

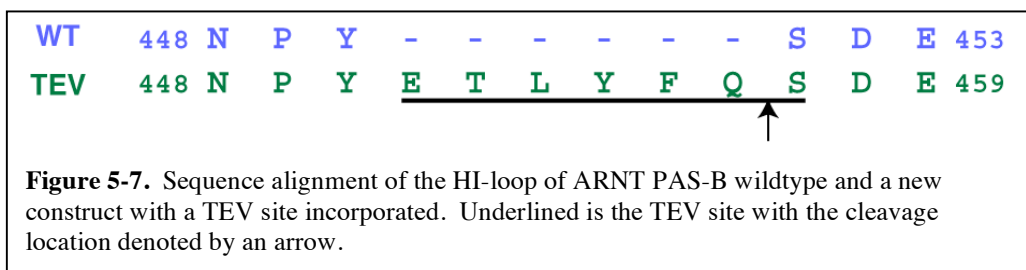
mutation either destabilizes the wildtype conformation and/or stabilizes the +3 conformation, in the presence or absence of the Y456T mutation.



### iii. Length of HI-loop affects equilibrium

In addition to finding that the identity of the proline (P449) affected the relative populations of the wt:+3 conformations, it was discovered that the length of the HI loop also affected the equilibrium. For one experiment, the initial goal

was to design a TEV site in the middle of the HI loop to examine if removing the covalent linkage would allow the I $\beta$ -strand to slip while still keeping the protein folded. This required the addition of six residues into ARNT PAS-B Y456T (Figure 5-7). Once the I $\beta$ -strand was cut with TEV, it would be free to bind in either conformation, enabling us to remove any effects the HI loop may impose on the I $\beta$ -strand register. As expected with a longer HI loop, before the protein was cut with TEV, the equilibrium between wt and the +3 conformation was shifted to 20:80. This result suggested the addition of six residues in the elongated HI loop was more energetically favorable to exist in the +3 conformation and that this long loop eliminated the effect the P449 residue had on the equilibrium. After cutting the TEV site, the equilibrium between the two conformations remained unchanged. Immediately after injecting this sample over a MonoQ column to separate the two conformations, a precipitant is formed. The precipitant is likely due to the cleaved fraction since no protein precipitant was observed prior to cleavage and is not observed for the wildtype. These results support previous studies that suggest the length of the HI loop in PAS domains may decide binding conformations and affinities to other proteins/ligands (138).



*C. Ligand binding to shift equilibrium of ARNT PAS-B Y456T*

*i. Ligand binding of PAS domains*

An intrinsic property of many PAS domains is their ability to internally bind small molecule ligands, enabling them to sense various environmental factors (181). As previously noted, PAS domains have been found to bind a wide variety of ligands, including FMN, FAD, p-coumaric acid and heme. The binding capabilities of PAS domains are quite diverse, suggesting the potential for these proteins to bind a variety of small molecules without disrupting the overall PAS fold. The extent of this flexibility can be best illustrated for ligands binding to those PAS domains for which no known endogenous cofactors has been found. For example, to date, no endogenous ligands have been found to bind HIF-2 $\alpha$  PAS-B, yet this domain has been shown to bind many types of small molecules in its core, such as ethylene glycol and various artificial ligands up to 380 daltons in molecular weight (11). For those PAS-containing proteins that do bind various cofactors, the binding of ligands induces a small conformational change in the domain, which in turn modulates protein-protein interactions with multiple downstream affects. In the case of HIF-2 $\alpha$  PAS-B binding an artificial ligand, this event leads to a subtle conformational change within the domain, which decreases the affinity to ARNT PAS-B *in vitro* (11). As for ARNT PAS-B, it has been demonstrated that different ARNT variants can adopt one of two

conformations, WT or +3, which results in slightly different core residues (16). Therefore, I wanted to test if small, artificial ligands could preferentially bind into the core of either one of these conformations.

*ii. Initial setup of compound library*

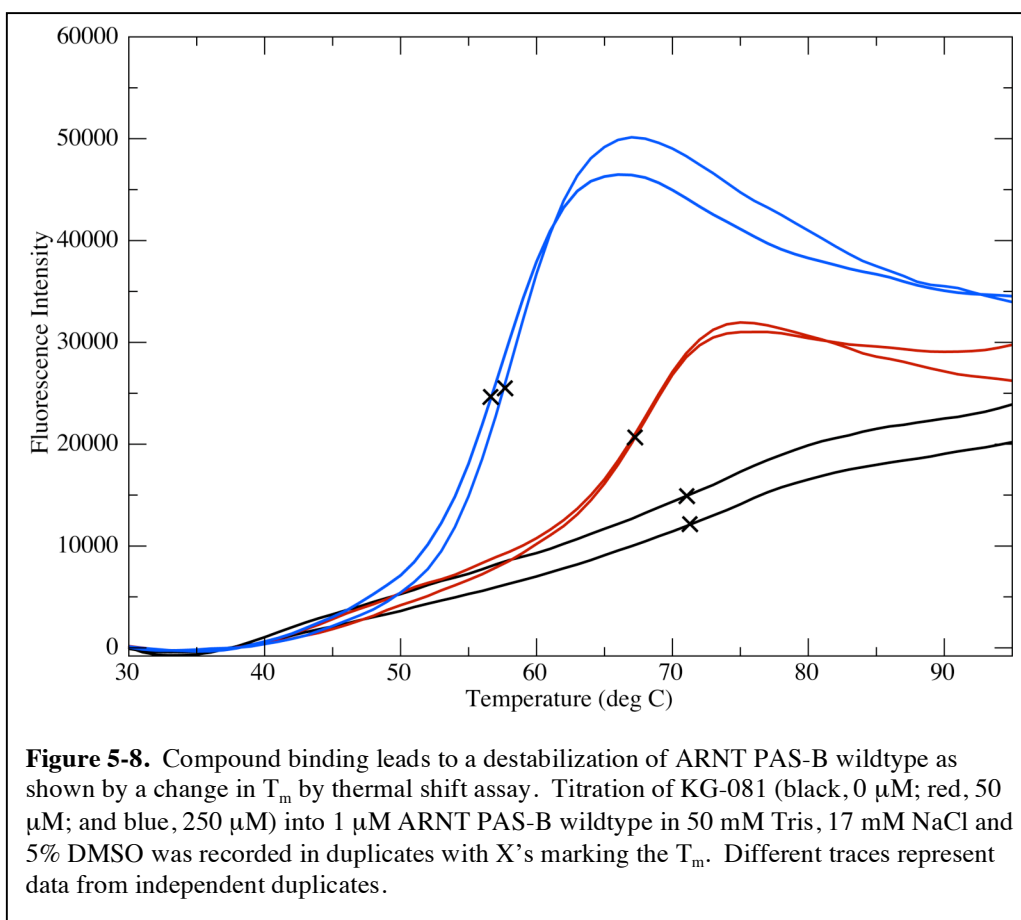
The construction of a small molecule compound library (~800 compounds) was hand-picked for the initial primary screen by Dr. Carlos Amezcua with the following criteria. First, compounds were kept small (MW = 202 +/- 73 Da) as they aimed to bind within the hydrophobic core of relatively small PAS domains. Second, to optimize the solubility of these compounds, compounds were selected based on their number of hydrogen bond acceptors and donors. In addition, the selected compounds were chosen with a bias towards aromatic or heterocyclic containing rings as to try to mimic naturally occurring cofactors that exist in other PAS domains (181, 182).

*iii. Thermal shift assay to detect ligand binding*

A goal of screening is to quickly determine if a ligand binds its receptor and if so, what the relative binding affinity is for this interaction. It has been well documented that ligand binding to proteins causes a conformational change in the protein, which can stabilize or destabilize the overall fold (181). This idea of measuring the protein stability to determine ligand binding is the basis of a

medium/high-throughput ligand screening assay to detect small molecule binding to various proteins, called the thermal shift assay. This conformational stabilization or destabilization of the domain upon ligand binding is due to the energetic coupling between the ligand binding event and protein unfolding transition (183). Therefore, the midpoint for temperature induced unfolding transition ( $T_m$ ) of the ligand-protein complex will shift compared to apo-protein. In addition, the magnitude of the shift in  $T_m$  is directly proportional to the ligand binding affinity ( $K_a$ ). To monitor these ligand-induced conformational changes, the thermal shift assay takes advantage of an environmentally sensitive fluorescent dye, such as Sypro Orange (183). When Sypro Orange is added to a solution of well-folded proteins, the signal is quenched as it is exposed to an aqueous environment. However, as the temperature rises, the protein will start to unfold, exposing the hydrophobic residues normally found in the protein interior. Sypro Orange binds these hydrophobic residues and, as a result, the fluorescence signal increases as the dye is unquenched (Figure 5-8). Ligand binding affects this process by changing the stability of the protein, which results in a change in melting temperature. Ultimately, the thermal shift assay allows for the identification of compounds that bind various receptors without any prior knowledge of the binding site. One of the biggest advantages of the thermal shift assay over other screening techniques, such as NMR, is its ability to be used in conjunction with a 96 or 384-well plate. This advantage greatly reduces the total

amount of protein necessary for the assay down to as little as 100  $\mu\text{L}$  of 1  $\mu\text{M}$  protein per well in a 96-well plate or 35  $\mu\text{L}$  of 1  $\mu\text{M}$  protein per well in a 384-well plate, compared to 500  $\mu\text{L}$  of 200  $\mu\text{M}$  protein per tube for NMR, eliminating one of the biggest obstacles for high-throughput drug discovery.



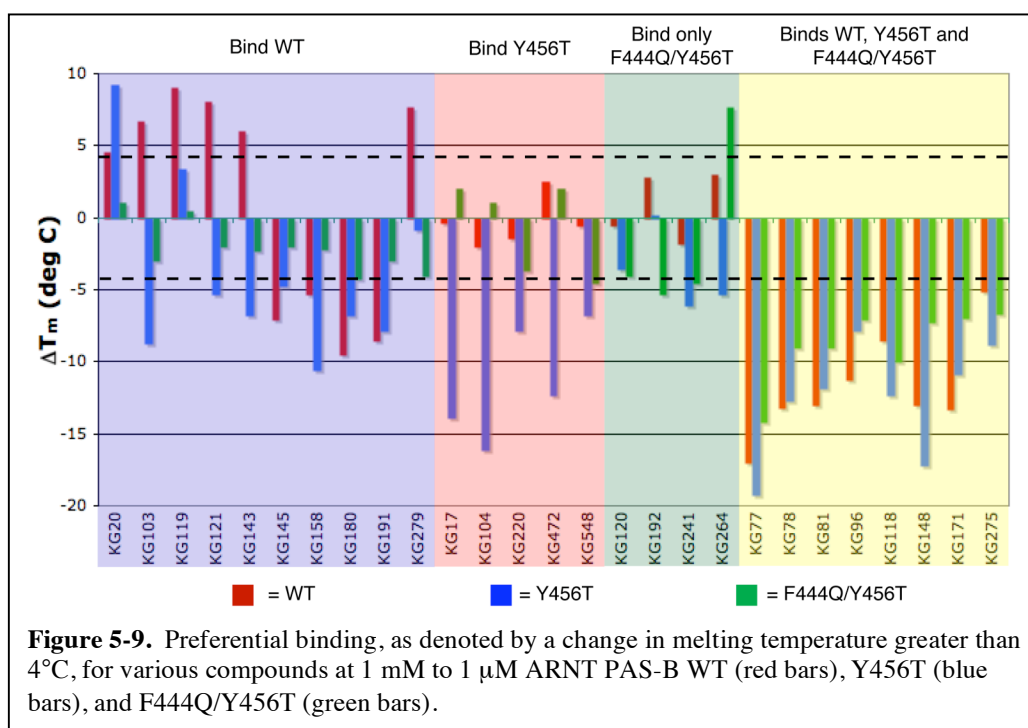
*a. Lead compound hits identified for the initial ligand screen*

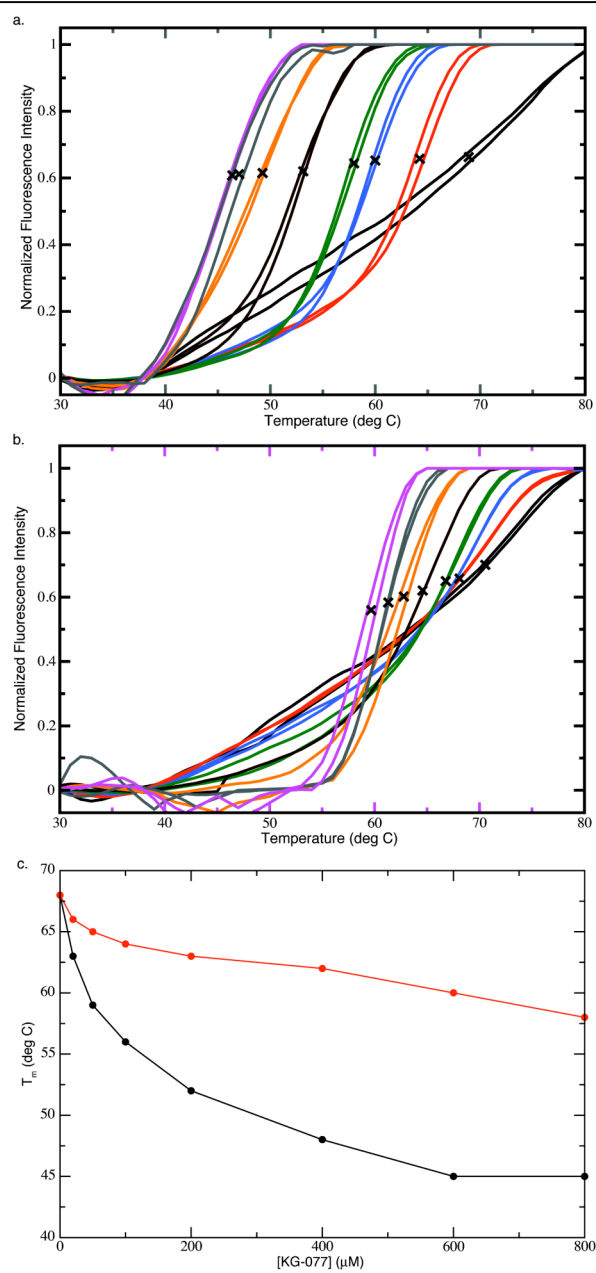
From the primary screen of ARNT PAS-B Y456T, twenty-seven compounds resulted in significant  $T_m$  changes ( $\Delta T_m > 4^\circ\text{C}$ ), suggesting that they



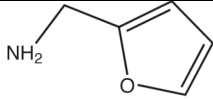
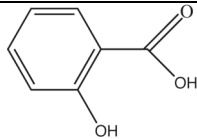
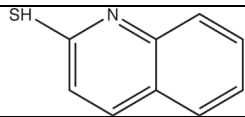
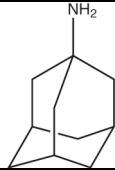
bound either one or both wt and +3 conformations, and thus were further investigated (Figure 5-9). These initial hits were deconvoluted further against either ARNT PAS-B WT or F444Q/Y456T, as these proteins were stabilized in only one of the two conformations. As a result, five of the twenty-seven initial hits were found to only bind ARNT PAS-B Y456T and not the WT or the double mutant proteins. These initial hits were removed from further analysis as they proved to be false positives or bound specifically to only the single mutant, neither of which proved to be useful for this study. As such, ten remaining compounds preferentially bound the wildtype protein, while four compounds were identified as binding the ARNT PAS-B F444Q/Y456T variant. The remaining eight other compounds were identified to bind all three proteins. Intriguingly, the  $T_m$  for these eight compounds that bound all three proteins decreased upon ligand binding, implying the well-packed core of ARNT PAS-B becomes less stable when compounds bind. Lastly, a titration series of compounds, ranging from 20  $\mu$ M to 800  $\mu$ M, with the three different protein receptors was performed on four compounds that were identified as binding all three proteins, to determine the relative binding affinities for each receptor (Figure 5-10). The binding affinities for each of these compounds for each protein are tabulated in Table 5-3. Intriguingly, compound KG-077 showed a 44  $\mu$ M binding affinity to the wildtype conformation and only a 413  $\mu$ M binding affinity to the +3 conformation. This

experiment proved that ligands could preferentially bind one conformation over the other.





**Figure 5-10.** Titration series of KG-077 (black, 0  $\mu$ M; red, 20  $\mu$ M; blue, 50  $\mu$ M; green, 100  $\mu$ M; brown, 200  $\mu$ M; orange, 400  $\mu$ M; grey, 600  $\mu$ M; and purple, 800  $\mu$ M) against 1  $\mu$ M ARNT PAS-B wildtype (a) or F444Q/Y456T (b), with the  $T_m$  denoted with an X. c. The concentration dependence of the  $T_m$ 's are plotted.

Compound	Structure	K <sub>d</sub> WT (μM)	K <sub>d</sub> Y456T (μM)	K <sub>d</sub> F444Q/Y456T (μM)
KG-077		44	154	413
KG-081		363	381	N.D.
KG-148		374	422	759
KG-171		345	266	N.D.

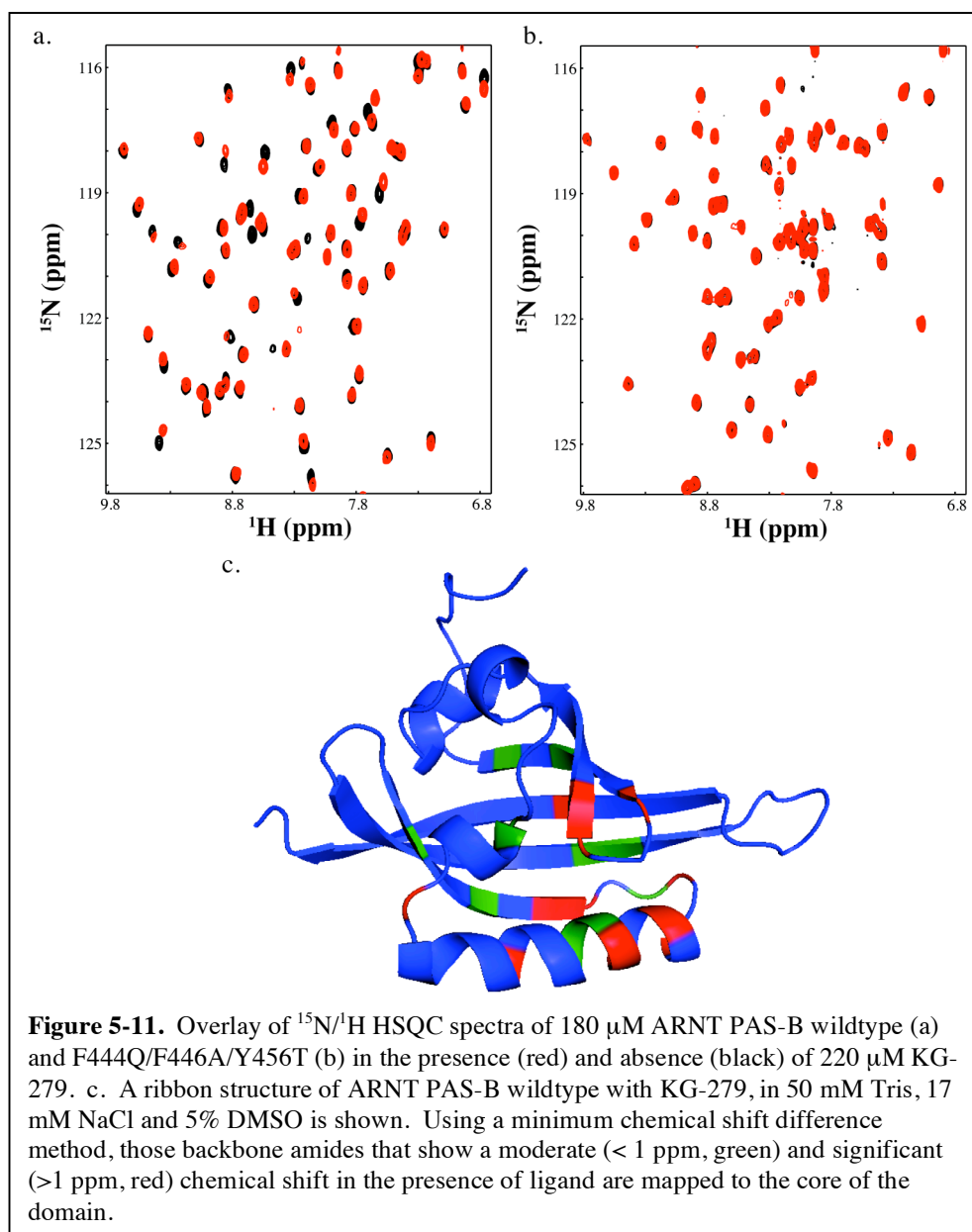
**Table 5-3.** Several compounds show preferential binding to one conformation over another. The midpoint for some transitions could not be determined, due to an increasing baseline before the transition curve, and are denoted with a N.D.

*iv. NMR to detect ligand binding*

*a. NMR screen on positive hits denoted by thermal shift assay*

Once compounds were identified as positive hits by thermal shift assay, we used NMR to study which protein residues were interacting with the compounds. Surprisingly, the top twenty-seven compounds that showed the greatest shifts in  $T_m$  by the thermal shift assay resulted in very minor chemical shift or peak intensity changes by NMR (Figure 5-11; discussed in further detail in section v.). For those few compounds that exhibited minor chemical shift

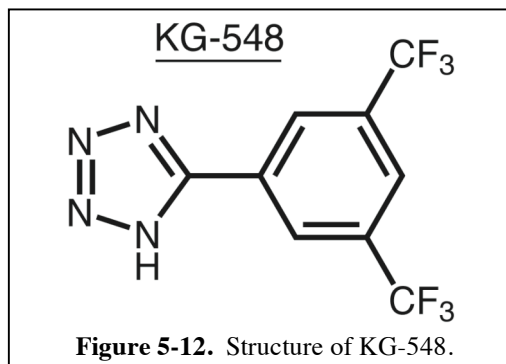
changes in either conformation, we wanted to confirm our compounds were binding into the core of the domain. Using the backbone amides from  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra as probes, those amides that were significantly broadened were mapped onto a high resolution ARNT PAS-B structure (Figure 5-11). As a result, perturbed amide reporters in response to compound binding mapped their sidechains to the core of the domain, particularly to those residues on the  $\beta$ -sheet surface. This result was encouraging as those PAS domains that contain a naturally bound cofactor bind on the same internal side of the  $\beta$ -sheet.



*b. New NMR screen of the 800 compound library*

In a separate experiment, conducted by Dr. Paul Card, the same 800 compound library was screened against ARNT PAS-B wildtype using  $^1\text{H}$ - $^{15}\text{N}$

HSQC spectra to determine binding as a first step in an effort to disrupt the ARNT/HIF PAS-B heterodimer. Initially, the compounds were screened as a mixture of five compounds per experiment to efficiently make use of time and protein stocks. Those mixtures that resulted in significant chemical shift changes were subjected to a secondary screen consisting of just one compound per sample to deconvolute which compounds within the mixture bound. As a result, several lead compounds were identified that bound into the core of the domain with  $K_d$  values below 300  $\mu$ M. One of these compounds, KG-548, exhibited slow exchange behavior, which can be indicative of a tight binding compound (Figure 5-12). Therefore, this compound was subjected to further experiments in an effort to disrupt the ARNT/HIF PAS-B heterodimer. As a result of *in vitro* experiments, Paul showed KG-548 could bind ARNT PAS-B, presumably by binding into the core of the domain with an affinity in the range of 130-350  $\mu$ M, depending on the technique used. Mapping the ligand-induced chemical shift changes onto the structure of ARNT PAS-B revealed the largest changes occur on the internal face of the  $\beta$ -sheet surface. This suggests the ligand may bind onto this domain in a similar manner as other cofactors bind PAS domains. This binding resulted in a conformational change of the domain and ultimately significantly disrupts the PAS-B heterodimer complex at high concentrations *in vitro*.



*v. Different assays, different results*

Surprisingly, the top twenty-seven compounds found to bind ARNT PAS-B wildtype by the thermal shift assay were not considered significant binders using NMR spectroscopy (Figure 5-10). In addition, the top compounds Paul found to bind the PAS-B domain of ARNT using NMR did not result in any significant change ( $\Delta T_m > 4^\circ\text{C}$ ) in  $T_m$  by thermal shift assay. These results suggest the two different methods, thermal shift and NMR, report on different properties of the protein. NMR spectroscopy reports on the local chemical environment around each NMR-active nucleus of the protein. On the other hand, the thermal shift assay reports on the melting temperature of the protein as a whole. When a compound binds and results in a significant chemical shift or peak intensity changes, this binding event does not necessary guarantee a change in global protein stability. Also, perhaps for those compounds that result in a change in  $T_m$  but lack any significant chemical shift in the core of the protein, may have certain properties that slightly change the buffer conditions, which in turn stabilizes or

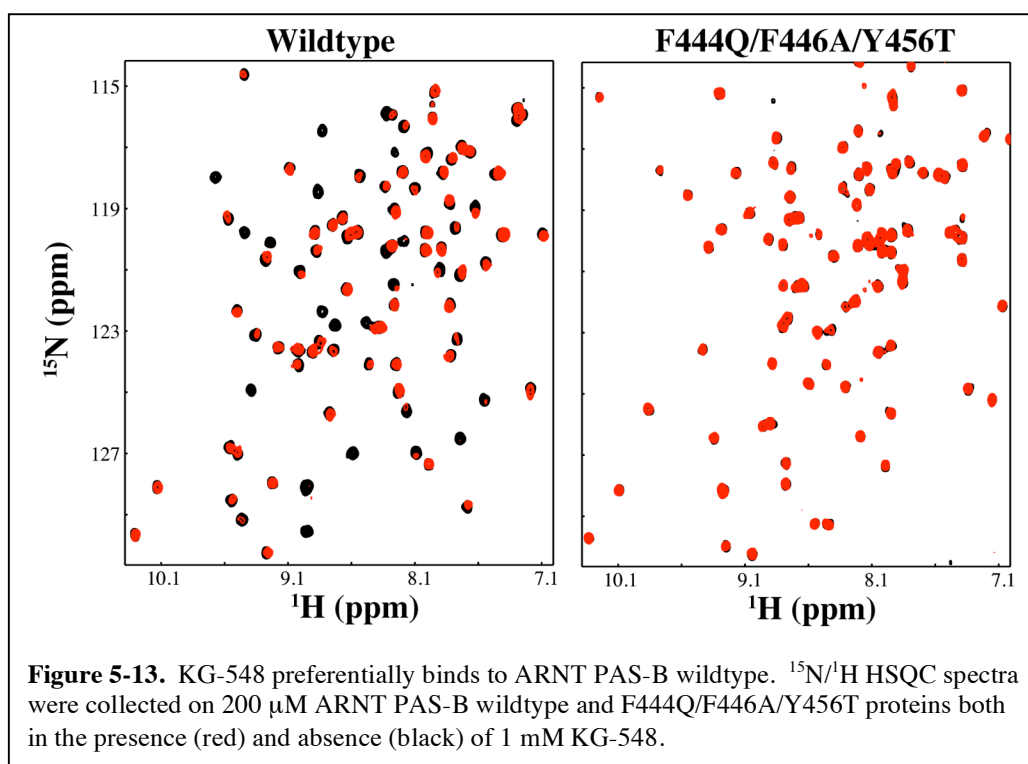


destabilizes the protein without any binding event occurring. Regardless of the method being used, it is intriguing that the core residues of ARNT PAS-B wildtype and the +3 conformation, which are very similar between both conformations, differ enough to preferentially bind different ligands by either technique.

*vi. Preferential binding for KG-548 against ARNT PAS-B conformers*

With the evidence that KG-548 binds into the core of the ARNT PAS-B wildtype conformation, we questioned its binding affinity to the +3 conformation. As such,  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of 180  $\mu\text{M}$  of ARNT PAS-B F444Q/F446A/Y456T were recorded in the presence and absence of 220  $\mu\text{M}$  KG-548. For the control sample in which no compound was present, DMSO was added to the same final concentration as the solution with compound. Compound stocks were initially dissolved and therefore stored in DMSO to keep as stable as possible. This control experiment makes any chemical shift or peak intensity changes due to compound only. Interestingly, ARNT PAS-B F444Q/F446A/Y456T did not show any significant changes compared to the reference spectrum, thereby indicating a very weak interaction with KG-548, at best (Figure 5-13). This finding is in complete contrast to the binding of KG-548 with wildtype protein and also to thermal shift data (Figure 5-9). In this spectra, greater than 30 amide sites are significantly broadened across the  $\beta$ -sheet surface.

This discovery supports the findings by thermal shift that compounds can have significant different binding affinities to either the wildtype or +3 conformations.



*vii. KG-548 drives the equilibrium towards the wildtype conformation*

Since KG-548 has been shown to preferentially bind the wildtype protein over the +3 conformation of ARNT PAS-B, the addition of this compound to Y456T should drive the equilibrium of the 51:49 (wt:mut) mixture towards the wildtype conformation. As expected, the binding of KG-548 to the wildtype protein (Figure 5-14) shifted the relative populations of the two conformations towards wildtype. Due to the broadening of spectral peaks in the presence of

compound, only those sites that did not shift upon ligand binding were analyzed.

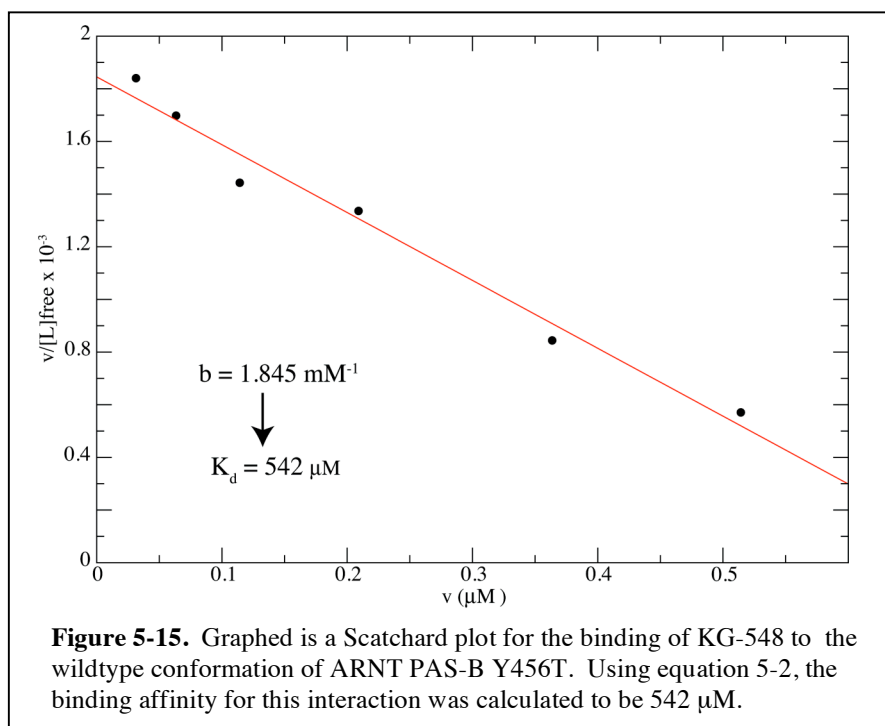
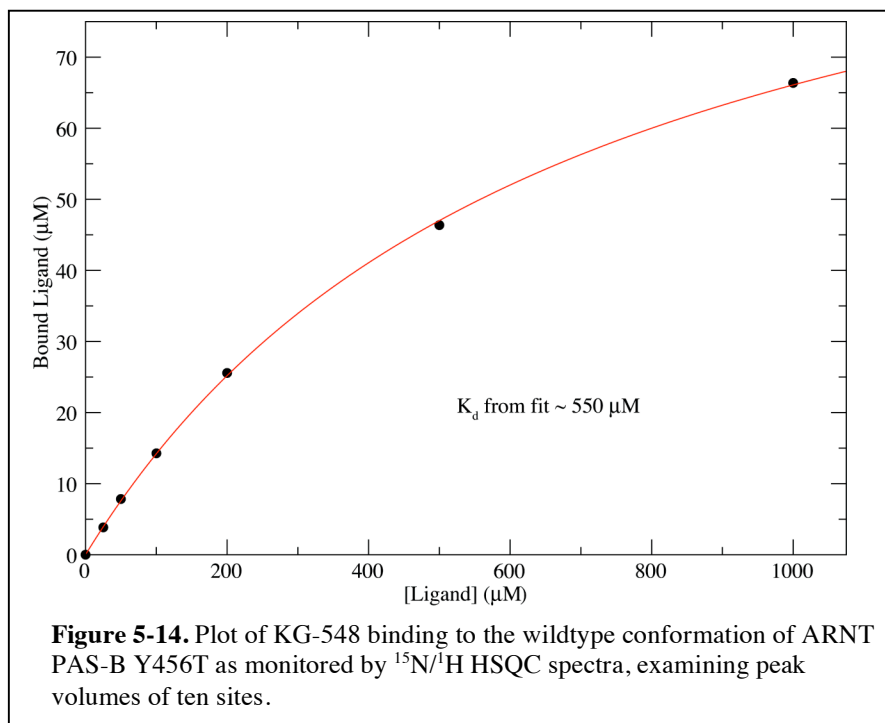
As a result, the binding of KG-548 to the wildtype conformation of ARNT PAS-B Y456T, affects the equilibrium of wildtype and +3 populations. Due to the size of KG-548 and the small core pocket of ARNT PAS-B, it is likely that the PAS domain only has one binding site. To confirm this suspicion, a Scatchard plot was graphed. The number of binding sites per molecule was determined from equation 5-1:

$$N = x\text{-intercept} \quad (\text{Eq. 5-1})$$

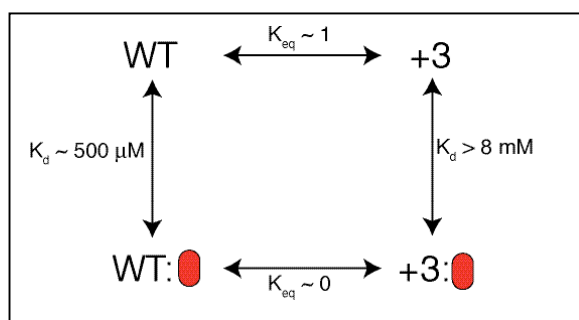
As a result, the number of binding sites per binding molecule of ARNT PAS-B was calculated to be .716. During the titration, a precipitant was formed in the tube at higher ligand concentrations. Assuming all ligand is bound, the amount of precipitant is approximately thirty percent of the protein, which equates well with the number of available binding sites per molecule of 0.716. In addition, the binding affinity of KG-548 for the wildtype conformation of Y456T can be extracted from the Scatchard plot using equation 5-2:

$$K_d = 1/b \quad (\text{Eq. 5-2})$$

Using this equation, a binding affinity of 500  $\mu\text{M}$  was calculated for the wildtype protein to KG-548 (Figure 5-15). These data confirm our previous findings that compounds can preferentially bind into the core of either ARNT PAS-B wildtype or +3 conformations.



These data support a four state model with the following states: wildtype, wildtype bound, +3, and +3 bound (Figure 5-16). In theory, the compound also binds the +3 conformation but our NMR data shows that the affinity between the two is minor compared to compound bound to wildtype, and that the dissociation constant is at best 8 mM. As such, with the addition of compound to Y456T, the compound binds the wildtype conformation creating a new wildtype bound state. This process depletes the wildtype unbound state, thereby driving the +3 conformation towards wildtype to re-establish equilibrium.



**Figure 5-16.** Diagram showing a four state model, with values presented for ARNT PAS-B Y456T. Red oval represents compound binding.

#### D. Conclusions

As previously noted, an intrinsic property of many PAS domains is their ability to bind small molecule ligands in order to regulate various environmental factors (181). For one PAS protein, for which no known endogenous cofactors has been found, HIF-2 $\alpha$  PAS-B has been shown to bind many types of small molecules in its core (11). For those PAS-containing proteins that do bind various cofactors, the binding of ligands induces a small conformational change in the

domain, which in turn modulates protein-protein interactions with multiple downstream affects. For this reason, I wanted to determine if a small molecule could preferentially bind into the core of either the ARNT PAS-B wildtype or +3 conformations and drive the relative populations of ARNT PAS-B Y456T towards one conformation. As a result, we identified several lead compounds by thermal shift assay to preferentially bind either conformation. However, analyzing these hits by NMR spectroscopy did not produce any significant chemical shift changes or peak broadening, implying no binding of the ligand to the protein. These data imply that both thermal shift assay and NMR spectroscopy report on different events of the binding process. It is intriguing that some of these compounds that preferentially bind the wildtype conformation could drive the equilibrium of ARNT PAS-B Y456T towards the wildtype state. Overall, the idea that a domain, for which no endogenous cofactors has been identified, can bind a small molecule and cause a conformational change to disrupt its interactions with other domains could lead to novel therapeutic agents.

### *E. Materials and Methods*

#### *i. NMR analysis of ARNT PAS-B Y456T equilibrium constant*

Structural studies of ARNT PAS-B and its mutants are typically conducted in 50 mM Tris, 17 mM NaCl and 5 mM DTT (12, 16). To determine the effects

buffer and pH have on the stability of each conformation, ARNT PAS-B Y456T was buffer exchanged into either 50mM Tris or PIPES over the pH range of 7-9 and 5-8, respectively.  $^{15}\text{N}$ - $^1\text{H}$  HSQC experiments of 200  $\mu\text{M}$  protein were recorded more than 20 hours post-exchange to allow adequate time for equilibrium to be established. For those experiments to determine the effects on the equilibrium constant from temperature,  $^1\text{H}$ - $^{15}\text{N}$  HSQC experiments of 200  $\mu\text{M}$  ARNT PAS-B Y456T in 50 mM Tris, 17 mM NaCl and 5 mM DTT were recorded after a period of time to ensure equilibrium was established in 5 Kelvin increments between 278 K and 333 K.

#### *ii. Ligand Screening Strategy and Methods*

In order to determine if compounds can preferentially bind either ARNT PAS-B WT or “+3” conformations, an initial screen of small molecule library assembled by Dr. Carlos Amezcua was used (181). The overall strategy involved a primary screen of all 800 compounds using the thermal shift assay with the assistance of Leanna Steier against the receptor, ARNT PAS-B Y456T. ARNT PAS-B Y456T was chosen as the receptor for the primary screen as this variant exists in both conformations, allowing only one round of screening instead of two against ARNT PAS-B WT and “+3” conformations. The first step of the primary screen was to load each well of a 96-well plate with 1  $\mu\text{M}$  ARNT PAS-B Y456T and 1 mM compound in a final volume of 100  $\mu\text{L}$  in 50 mM Tris, 17 mM NaCl

and 5% DMSO. All experiments were conducted in duplicate trials to reduce false hits. Using an Applied Biosystems 7500 Real-Time PCR machine and an excitation wavelength of 475 nm specifically for the Sypro Orange dye, the fluorescence emission intensity at 590 nm was recorded initially at 25°C. Throughout the course of the experiment, the temperature was incremented by 1°C (up to 95°C) and the fluorescence intensity was recorded at each interval after a short time delay of thirty seconds, to ensure the samples were at the proper temperature. Using a Boltzmann sigmoidal curve with a 95% confidence band, the  $T_m$  for each well was determined. The observation of a ligand-induced shift in  $T_m$  of 4°C or greater was considered indicative of a potential binder. Once a potential binder is identified, that compound was re-screened against ARNT PAS-B WT and F444Q/Y456T to determine which conformation the compound binds. ARNT PAS-B F444Q/Y456T is a mutant that exists in a 19:81 (wt:+3) equilibrium. This mutant was used to determine if the compound binds the “+3” conformation instead of ARNT PAS-B F444Q/F446A/Y456T since the triple mutant is more stable at higher temperatures ( $T_m > 80^\circ\text{C}$ ) under certain solution conditions. Calculating an accurate  $T_m$  at these higher temperatures was challenging as the Applied Biosystems 7500 RT-PCR machine’s highest temperature is limited to 95°C, which does not allow for the entire transition curve to be recorded.



*iii. NMR analysis of positive hits by thermal shift assay*

Once compounds were identified as positive hits by thermal shift assay, a further study as to which protein residues were interacting with the compound was investigated. Purified protein samples, either WT, Y456T, or F444Q/F446A/Y456T, at a concentration of 180  $\mu\text{M}$  was combined with a single compound with a final concentration of 220  $\mu\text{M}$  and a  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra was recorded. Samples remained at room temperature for twenty-four hours to ensure equilibrium had been reached. Those samples that showed either chemical shifts and peak intensity changes were determined as binders by NMR.

*iv. Scatchard plot analysis*

To determine the binding affinity, a titration series of increasing concentrations of KG-548 (0, 25, 50, 100, 200, 500, and 1000  $\mu\text{M}$ ) against a constant 200  $\mu\text{M}$  ARNT PAS-B Y456T was conducted.  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra were recorded 24 hours after samples were mixed to allow sufficient time to equilibrate. Measuring the peak intensities to get the relative populations of each conformation was challenging due to the peak broadening of the wildtype conformation in the presence of KG-548. For this reason, peak volumes were measured for those sites that showed no chemical shift changes, which suggest the bound state has the same chemical shifts as the unbound state. Using the relative peak volumes, the concentrations of all species (total wildtype and +3

conformations) were determined. Over the course of the experiment, the wildtype peak volume, which was the sum of the wildtype and wildtype bound conformations increased upon increasing amounts of KG-548. The wildtype bound state was calculated to be the difference between the total wildtype concentration and the +3 concentration, since the equilibrium between the wildtype unbound and +3 unbound conformations are at a 51:49 equilibrium. The amount of precipitant at each data point was calculated as the total amount of protein (200  $\mu$ M) minus the wildtype and +3 concentrations. Lastly, the measure of unbound ligand was calculated as the difference between the total amount of ligand present and the wildtype bound concentration.

#### *F. Acknowledgements*

We thank Kyle Brewer, Leanna Steier and Amy Zhou for their technical assistance with various aspects of this research. This research was supported by a grant from the NIH (R01 GM081875) to K.H.G.

## VI. Conclusions and Future Directions

### *A. Conclusions*

The  $\beta$ -sheet surface of many PAS domains is critical to their function as protein/protein interaction domains. For this reason, it is intriguing that this binding surface for ARNT PAS-B is flexible enough to adopt a new conformation in the presence of a single point mutation (Y456T). This new conformation differs from that of the native wildtype state by a three-residue slip, and inversion of the central I $\beta$ -strand. Overall, this mutation allows the domain to interconvert slowly as it samples both states equally. This discovery puts the ARNT PAS-B Y456T protein into a unique class of proteins that can adopt multiple folds, such as lymphotactin (15). In this system, the interconversion between two novel states of lymphotactin provides a novel mechanism for switching between two functionally different conformations. As for ARNT, it interacts with multiple different proteins, many of which presumably bind the  $\beta$ -sheet surface. However, it remains to be determined if the wildtype ARNT PAS-B domain can adopt this second conformation, and even if so, what the functional consequences would be, *in vivo*.

With this discovery, we suggest that a wider range of proteins, including those susceptible to genetic or post-translational modifications in sequence, may also explore different folded states. The low number of proteins found to adopt

multiple states might be due, in part, to the structural techniques used.

Crystallization may not be best suited to capture these low populated, folded states as a solution with multiple states would not produce a well, packed crystal. On the other hand, the ability to study these low-populated folding intermediates has been aided by the advancements made in NMR. Using a CPMG experiment, a train of spin echo pulses are used to slow down the decay of signal intensity, allowing data to be collected on these states (184-186). This technique has been used to get relative populations of very low populated states. By using advanced NMR methods, we speculate that a wider variety of natural protein structures will be found to adopt multiple conformations with distinct functions.

### *B. Future Directions*

Future experiments should try to address whether the wildtype protein samples this new shifted conformation or if we just stumbled upon a flaw (Y456) within the ARNT PAS-B fold. First of all, we should continue to screen small molecule compounds that preferentially bind the +3 conformation with properties similar to those initial hits found by thermal shift and NMR. As such, the addition of this compound to the wildtype protein, could shift the equilibrium towards the +3 conformation to a significant level, thereby allowing the detection of this species by standard  $^1\text{H}$ - $^{15}\text{N}$  HSQC experiments. The finding of such a compound that preferentially bound the +3 conformation would help address if

this second slipped strand conformation was present at low concentration in hARNT. However, this experiment is a long shot since such a conformational change would require burying Y456 into the core of the protein.

Secondly, it should also be noted that a very limited number of experiments have investigated the stability of both conformations in different solution conditions. For the majority of the structural studies of ARNT PAS-B wildtype and various mutants in this thesis, the following solution was used: 50 mM Tris, 17 mM NaCl and 5 mM DTT. This solution was initially chosen for the structural studies of ARNT PAS-B wildtype as it stabilized the wildtype protein and resulted in a well folded protein domain, as determined by NMR experiments. Perhaps under different solution conditions, the +3 conformation may prove to be more stable and provide a means to experimentally detect this conformation in wildtype protein. In this regard,  $^1\text{H}$ - $^{15}\text{N}$  HSQC experiments should be collected on ARNT PAS-B Y456T under many different solution conditions, in which the type of buffer, salt concentration and pH values significantly vary. Such experiments may discover a new solution condition in which the +3 conformation is detectable in the wildtype protein.

Lastly, as noted in Chapter 3, the sequence of the I $\beta$ -strand is nearly identical across all ARNT homologs, as well as ARNT2 and bMAL (Figure 3-13). The notable exceptions are the residues, MSIF, in the HI loop of RnARNT, where in the hARNT the sequence is IEYI. Interestingly, when the Y456 residue of

hARNT was mutated to various beta-branched residues, multiple conformations of the domain appeared, Chapter 5. One would expect mutating the four residues in hARNT to those found in RnARNT would also result in a different conformation and possibly multiple conformations of the domain, as observed for the single Y456I point mutation. Such a result would suggest RnARNT is able to adopt multiple conformations of the domain, with potentially distinct biological roles.

Overall, these additional experiments will help address if the +3 conformation, initially found by mutagenesis, is present in the wildtype protein. Should such experiments conclude that the wildtype hARNT and/or RnARNT can adopt either conformation, it could imply ARNT may utilize its different folds to selectively interact with its multiple binding partners in an effort to control various biological pathways.

## Appendix

Mutation	Percent WT	Percent MUT	Percent additional Peaks
Wild-Type	>99	<1	
F444Q	>99	<1	
F444Q/I454A/Y456T	>98	<2	
F444Q/I454F/Y456T	>98	<2	
F444Q/Y456T/I458E	>98	<2	
Y456A	>98	<2	
Y456W	>98	<2	
Y456C	>97	<3	
Y456D	>97	<3	
Y456Q	>97	<3	
N448W	>97	<3	
P449W	>97	<3	
Y456M	>96	<4	
Y456Q	>96	<4	
Y456S	81	19	
Y456F	53	27	13, 8
Y456T	50	50	
Y456L	40	37	14, 10
Y456V	37	-	42, 21
P449A/Y456T	32	68	
F444Q/T445A/Y456T	31	69	
P449A/Y456T refolded post-GdmHCl	28	72	
F444Q/Y456T/T460V	25	75	
Y456T + TEV Site Cleaved	20	80	
Y456T + TEV Site Uncleaved	19	81	
F444Q/Y456T	19	81	
F444Q/N448T/Y456T	18	82	
F444Q/Y456T/T460V	15	85	
P449A	14	-	60, 14, 12
F444Q/T445Y/Y456T	14	86	
F444Q/N448T/Y456T	10	90	
P449G/Y456T	9	91	
F444Q/F446A/P449A/Y456T	7	93	
F444Q/N448T/Y456T/I457F	<3	>97	
F444Q/N448W/Y456T	<3	>97	
F444Q/N448W/Y456T	<2	>98	
F444Q/Y456T/I457F	<2	>98	
F444Q/F446A/Y456T	<1	>99	

**Appendix Table 1-1.** Relative populations for all ARNT PAS-B mutations measured in 50 mM Tris, 17 mM NaCl and 5 mM DTT. The relative populations were calculated from peak intensities of  $^{15}\text{N}/^1\text{H}$  HSQC spectra, based on 8-14 measurements. Each measurement carries approximately 6% relative error on any given ratio.

Mutation	Percent WT	Percent MUT	Percent additional Peaks
E362K	>97	<3	
R366E	>97	<3	
F444S	>96	<4	
F444Y	>96	<4	
R366A	>94	<6	

**Appendix Table 1-2.** Listed are those mutations that did not produce a second conformation as determined by 6-10 sites in the  $^{15}\text{N}/^1\text{H}$  HSQC spectra and were not discussed in this dissertation. Each measurement carries approximately 6% relative error on any given ratio.



## References

1. Gong, W., Hao, B., Mansy, S. S., Gonzalez, G., Gilles-Gonzalez, M. A., and Chan, M. K. (1998) Structure of a biological oxygen sensor: a new mechanism for heme-driven signal transduction, *Proc Natl Acad Sci U S A* 95, 15177-15182.
2. Longo, A., Guanga, G. P., and Rose, R. B. (2008) Crystal structure of E47-NeuroD1/beta2 bHLH domain-DNA complex: heterodimer selectivity and DNA recognition, *Biochemistry* 47, 218-229.
3. Kuloglu, E. S., McCaslin, D. R., Kitabwalla, M., Pauza, C. D., Markley, J. L., and Volkman, B. F. (2001) Monomeric solution structure of the prototypical 'C' chemokine lymphotactin, *Biochemistry* 40, 12486-12496.
4. Hao, B., Isaza, C., Arndt, J., Soltis, M., and Chan, M. K. (2002) Structure-based mechanism of O<sub>2</sub> sensing and ligand discrimination by the FixL heme domain of *Bradyrhizobium japonicum*, *Biochemistry* 41, 12952-12958.
5. Anfinsen, C. B. (1973) Principles that govern the folding of protein chains, *Science* 181, 223-230.
6. Getzoff, E. D., Gutwin, K. N., and Genick, U. K. (2003) Anticipatory active-site motions and chromophore distortion prime photoreceptor PYP for light activation, *Nat Struct Biol* 10, 663-668.
7. Card, P. B., Erbel, P. J., and Gardner, K. H. (2005) Structural basis of ARNT PAS-B dimerization: use of a common beta-sheet interface for hetero- and homodimerization, *J Mol Biol* 353, 664-677.
8. Lee, J., Tomchick, D. R., Brautigam, C. A., Machius, M., Kort, R., Hellingwerf, K. J., and Gardner, K. H. (2008) Changes at the KinA PAS-A dimerization interface influence histidine kinase function, *Biochemistry* 47, 4051-4064.
9. Malpica, R., Franco, B., Rodriguez, C., Kwon, O., and Georgellis, D. (2004) Identification of a quinone-sensitive redox switch in the ArcB sensor kinase, *Proc Natl Acad Sci U S A* 101, 13318-13323.

10. Raumann, B. E., Rould, M. A., Pabo, C. O., and Sauer, R. T. (1994) DNA recognition by beta-sheets in the Arc repressor-operator crystal structure, *Nature* 367, 754-757.
11. Scheuermann, T. H., Tomchick, D. R., Machius, M., Guo, Y., Bruick, R. K., and Gardner, K. H. (2009) Artificial ligand binding within the HIF2alpha PAS-B domain of the HIF2 transcription factor, *Proc Natl Acad Sci U S A* 106, 450-455.
12. Erbel, P. J., Card, P. B., Karakuzu, O., Bruick, R. K., and Gardner, K. H. (2003) Structural basis for PAS domain heterodimerization in the basic helix--loop--helix-PAS transcription factor hypoxia-inducible factor, *Proc Natl Acad Sci U S A* 100, 15504-15509.
13. Stout, T. J., Graham, H., Buckley, D. I., and Matthews, D. J. (2000) Structures of active and latent PAI-1: a possible stabilizing role for chloride ions, *Biochemistry* 39, 8460-8469.
14. Koradi, R., Billeter, M., and Wuthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures, *J Mol Graph* 14, 51-55, 29-32.
15. Tuinstra, R. L., Peterson, F. C., Kutlesa, S., Elgin, E. S., Kron, M. A., and Volkman, B. F. (2008) Interconversion between two unrelated protein folds in the lymphotactin native state, *Proc Natl Acad Sci U S A* 105, 5057-5062.
16. Evans, M. R., Card, P. B., and Gardner, K. H. (2009) ARNT PAS-B has a fragile native state structure with an alternative {beta}-sheet register nearby in sequence space, *Proc Natl Acad Sci U S A*.
17. Cordes, M. H., Walsh, N. P., McKnight, C. J., and Sauer, R. T. (1999) Evolution of a protein fold in vitro, *Science* 284, 325-328.
18. Yildiz, O., Doi, M., Yujnovsky, I., Cardone, L., Berndt, A., Hennig, S., Schulze, S., Urbanke, C., Sassone-Corsi, P., and Wolf, E. (2005) Crystal structure and interactions of the PAS repeat region of the Drosophila clock protein PERIOD, *Mol Cell* 17, 69-82.
19. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for

- checking the quality of protein structures solved by NMR, *J Biomol NMR* 8, 477-486.
20. Taylor, B. L., and Zhulin, I. B. (1999) PAS domains: internal sensors of oxygen, redox potential, and light, *Microbiol Mol Biol Rev* 63, 479-506.
  21. Nambu, J. R., Lewis, J. O., Wharton, K. A., Jr., and Crews, S. T. (1991) The *Drosophila* single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development, *Cell* 67, 1157-1167.
  22. Citri, Y., Colot, H. V., Jacquier, A. C., Yu, Q., Hall, J. C., Baltimore, D., and Rosbash, M. (1987) A family of unusually spliced biologically active transcripts encoded by a *Drosophila* clock gene, *Nature* 326, 42-47.
  23. Reddy, P., Jacquier, A. C., Abovich, N., Petersen, G., and Rosbash, M. (1986) The period clock locus of *D. melanogaster* codes for a proteoglycan, *Cell* 46, 53-61.
  24. Hoffman, E. C., Reyes, H., Chu, F. F., Sander, F., Conley, L. H., Brooks, B. A., and Hankinson, O. (1991) Cloning of a factor required for activity of the Ah (dioxin) receptor, *Science* 252, 954-958.
  25. Jackson, F. R., Bargiello, T. A., Yun, S. H., and Young, M. W. (1986) Product of per locus of *Drosophila* shares homology with proteoglycans, *Nature* 320, 185-188.
  26. Yang, J., Zhang, L., Erbel, P. J., Gardner, K. H., Ding, K., Garcia, J. A., and Bruick, R. K. (2005) Functions of the Per/ARNT/Sim domains of the hypoxia-inducible factor, *J Biol Chem* 280, 36047-36054.
  27. Rubinstenn, G., Vuister, G. W., Mulder, F. A., Dux, P. E., Boelens, R., Hellingwerf, K. J., and Kaptein, R. (1998) Structural and dynamic changes of photoactive yellow protein during its photocycle in solution, *Nat Struct Biol* 5, 568-570.
  28. Harper, S. M., Neil, L. C., and Gardner, K. H. (2003) Structural basis of a phototropin light switch, *Science* 301, 1541-1544.
  29. Razeto, A., Ramakrishnan, V., Litterst, C. M., Giller, K., Griesinger, C., Carlomagno, T., Lakomek, N., Heimbürg, T., Lodrini, M., Pfitzner, E.,

- and Becker, S. (2004) Structure of the NCoA-1/SRC-1 PAS-B domain bound to the LXXLL motif of the STAT6 transactivation domain, *J Mol Biol* 336, 319-329.
30. Zhulin, I. B., Taylor, B. L., and Dixon, R. (1997) PAS domain S-boxes in Archaea, Bacteria and sensors for oxygen and redox, *Trends Biochem Sci* 22, 331-333.
  31. Krogh, A., Brown, M., Mian, I. S., Sjolander, K., and Haussler, D. (1994) Hidden Markov models in computational biology. Applications to protein modeling, *J Mol Biol* 235, 1501-1531.
  32. Hefti, M. H., Francoijs, K. J., de Vries, S. C., Dixon, R., and Vervoort, J. (2004) The PAS fold. A redefinition of the PAS domain based upon structural prediction, *Eur J Biochem* 271, 1198-1208.
  33. Lim, W. A. (2002) The modular logic of signaling proteins: building allosteric switches from simple binding domains, *Curr Opin Struct Biol* 12, 61-68.
  34. Mitchell, P. (1961) Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism, *Nature* 191, 144-148.
  35. Bibikov, S. I., Miller, A. C., Gosink, K. K., and Parkinson, J. S. (2004) Methylation-independent aerotaxis mediated by the Escherichia coli Aer protein, *J Bacteriol* 186, 3730-3737.
  36. Bibikov, S. I., Barnes, L. A., Gitin, Y., and Parkinson, J. S. (2000) Domain organization and flavin adenine dinucleotide-binding determinants in the aerotaxis signal transducer Aer of Escherichia coli, *Proc Natl Acad Sci U S A* 97, 5830-5835.
  37. Genick, U. K., Soltis, S. M., Kuhn, P., Canestrelli, I. L., and Getzoff, E. D. (1998) Structure at 0.85 Å resolution of an early protein photocycle intermediate, *Nature* 392, 206-209.
  38. Dux, P., Rubinstenn, G., Vuister, G. W., Boelens, R., Mulder, F. A., Hard, K., Hoff, W. D., Kroon, A. R., Crielaard, W., Hellingwerf, K. J., and Kaptein, R. (1998) Solution structure and backbone dynamics of the photoactive yellow protein, *Biochemistry* 37, 12689-12699.

39. Xie, A., Hoff, W. D., Kroon, A. R., and Hellingwerf, K. J. (1996) Glu46 donates a proton to the 4-hydroxycinnamate anion chromophore during the photocycle of photoactive yellow protein, *Biochemistry* 35, 14671-14678.
40. Hoff, W. D., Dux, P., Hard, K., Devreese, B., Nugteren-Roodzant, I. M., Crielard, W., Boelens, R., Kaptein, R., van Beeumen, J., and Hellingwerf, K. J. (1994) Thiol ester-linked p-coumaric acid as a new photoactive prosthetic group in a protein with rhodopsin-like photochemistry, *Biochemistry* 33, 13959-13962.
41. Baca, M., Borgstahl, G. E., Boissinot, M., Burke, P. M., Williams, D. R., Slater, K. A., and Getzoff, E. D. (1994) Complete chemical structure of photoactive yellow protein: novel thioester-linked 4-hydroxycinnamyl chromophore and photocycle chemistry, *Biochemistry* 33, 14369-14377.
42. Sakamoto, K., and Briggs, W. R. (2002) Cellular and subcellular localization of phototropin 1, *Plant Cell* 14, 1723-1735.
43. Briggs, W. R., and Christie, J. M. (2002) Phototropins 1 and 2: versatile plant blue-light receptors, *Trends Plant Sci* 7, 204-210.
44. Salomon, M., Christie, J. M., Knieb, E., Lempert, U., and Briggs, W. R. (2000) Photochemical and mutational analysis of the FMN-binding domains of the plant blue light receptor, phototropin, *Biochemistry* 39, 9401-9410.
45. Swartz, T. E., Corchnoy, S. B., Christie, J. M., Lewis, J. W., Szundi, I., Briggs, W. R., and Bogomolni, R. A. (2001) The photocycle of a flavin-binding domain of the blue light photoreceptor phototropin, *J Biol Chem* 276, 36493-36500.
46. Christie, J. M., Swartz, T. E., Bogomolni, R. A., and Briggs, W. R. (2002) Phototropin LOV domains exhibit distinct roles in regulating photoreceptor function, *Plant J* 32, 205-219.
47. Harper, S. M., Christie, J. M., and Gardner, K. H. (2004) Disruption of the LOV-J $\alpha$  helix interaction activates phototropin kinase activity, *Biochemistry* 43, 16184-16192.

48. Neiditch, M. B., Federle, M. J., Miller, S. T., Bassler, B. L., and Hughson, F. M. (2005) Regulation of LuxPQ receptor activity by the quorum-sensing signal autoinducer-2, *Mol Cell* 18, 507-518.
49. Tuckerman, J. R., Gonzalez, G., and Gilles-Gonzalez, M. A. (2001) Complexation precedes phosphorylation for two-component regulatory system FixL/FixJ of *Sinorhizobium meliloti*, *J Mol Biol* 308, 449-455.
50. Li, X., Xu, J., and Li, M. (1997) The human delta1261 mutation of the HERG potassium channel results in a truncated protein that contains a subunit interaction domain and decreases the channel expression, *J Biol Chem* 272, 705-708.
51. Thomas, J. B., Crews, S. T., and Goodman, C. S. (1988) Molecular genetics of the single-minded locus: a gene involved in the development of the *Drosophila* nervous system, *Cell* 52, 133-141.
52. Isaac, D. D., and Andrew, D. J. (1996) Tubulogenesis in *Drosophila*: a requirement for the trachealess gene product, *Genes Dev* 10, 103-117.
53. Wilk, R., Weizman, I., and Shilo, B. Z. (1996) trachealess encodes a bHLH-PAS protein that is an inducer of tracheal cell fates in *Drosophila*, *Genes Dev* 10, 93-102.
54. Kewley, R. J., Whitelaw, M. L., and Chapman-Smith, A. (2004) The mammalian basic helix-loop-helix/PAS family of transcriptional regulators, *Int J Biochem Cell Biol* 36, 189-204.
55. Levine, S. L., and Perdew, G. H. (2002) Okadaic acid increases ARNT homodimer transactivation potential, *Cell Biol Toxicol* 18, 109-120.
56. Sogawa, K., Nakano, R., Kobayashi, A., Kikuchi, Y., Ohe, N., Matsushita, N., and Fujii-Kuriyama, Y. (1995) Possible function of Ah receptor nuclear translocator (Arnt) homodimer in transcriptional regulation, *Proc Natl Acad Sci U S A* 92, 1936-1940.
57. Zelzer, E., Wappner, P., and Shilo, B. Z. (1997) The PAS domain confers target gene specificity of *Drosophila* bHLH/PAS proteins, *Genes Dev* 11, 2079-2089.

58. Gekakis, N., Staknis, D., Nguyen, H. B., Davis, F. C., Wilsbacher, L. D., King, D. P., Takahashi, J. S., and Weitz, C. J. (1998) Role of the CLOCK protein in the mammalian circadian mechanism, *Science* 280, 1564-1569.
59. Darlington, T. K., Wager-Smith, K., Ceriani, M. F., Staknis, D., Gekakis, N., Steeves, T. D., Weitz, C. J., Takahashi, J. S., and Kay, S. A. (1998) Closing the circadian loop: CLOCK-induced transcription of its own inhibitors *per* and *tim*, *Science* 280, 1599-1603.
60. Whitelaw, M. L., Gustafsson, J. A., and Poellinger, L. (1994) Identification of transactivation and repression functions of the dioxin receptor and its basic helix-loop-helix/PAS partner factor Arnt: inducible versus constitutive modes of regulation, *Mol Cell Biol* 14, 8343-8355.
61. Kobayashi, A., Numayama-Tsuruta, K., Sogawa, K., and Fujii-Kuriyama, Y. (1997) CBP/p300 functions as a possible transcriptional coactivator of Ah receptor nuclear translocator (Arnt), *J Biochem* 122, 703-710.
62. Maltepe, E., Schmidt, J. V., Baunoch, D., Bradfield, C. A., and Simon, M. C. (1997) Abnormal angiogenesis and responses to glucose and oxygen deprivation in mice lacking the protein ARNT, *Nature* 386, 403-407.
63. Abbott, B. D., and Buckalew, A. R. (2000) Placental defects in ARNT-knockout conceptus correlate with localized decreases in VEGF-R2, Ang-1, and Tie-2, *Dev Dyn* 219, 526-538.
64. Chapman-Smith, A., Lutwyche, J. K., and Whitelaw, M. L. (2004) Contribution of the Per/Arnt/Sim (PAS) domains to DNA binding by the basic helix-loop-helix PAS transcriptional regulators, *J Biol Chem* 279, 5353-5362.
65. Sonnenfeld, M., Ward, M., Nystrom, G., Mosher, J., Stahl, S., and Crews, S. (1997) The *Drosophila* tango gene encodes a bHLH-PAS protein that is orthologous to mammalian Arnt and controls CNS midline and tracheal development, *Development* 124, 4571-4582.
66. Ema, M., Morita, M., Ikawa, S., Tanaka, M., Matsuda, Y., Gotoh, O., Saijoh, Y., Fujii, H., Hamada, H., Kikuchi, Y., and Fujii-Kuriyama, Y. (1996) Two new members of the murine Sim gene family are transcriptional repressors and show different expression patterns during mouse embryogenesis, *Mol Cell Biol* 16, 5865-5875.

67. Moffett, P., Reece, M., and Pelletier, J. (1997) The murine Sim-2 gene product inhibits transcription by active repression and functional interference, *Mol Cell Biol* 17, 4933-4947.
68. Michaud, J. L., Rosenquist, T., May, N. R., and Fan, C. M. (1998) Development of neuroendocrine lineages requires the bHLH-PAS transcription factor SIM1, *Genes Dev* 12, 3264-3275.
69. Poland, A., and Knutson, J. C. (1982) 2,3,7,8-tetrachlorodibenzo-p-dioxin and related halogenated aromatic hydrocarbons: examination of the mechanism of toxicity, *Annu Rev Pharmacol Toxicol* 22, 517-554.
70. Safe, S. H. (1986) Comparative toxicology and mechanism of action of polychlorinated dibenzo-p-dioxins and dibenzofurans, *Annu Rev Pharmacol Toxicol* 26, 371-399.
71. Morales, J. L., and Perdew, G. H. (2007) Carboxyl terminus of hsc70-interacting protein (CHIP) can remodel mature aryl hydrocarbon receptor (AhR) complexes and mediate ubiquitination of both the AhR and the 90 kDa heat-shock protein (hsp90) in vitro, *Biochemistry* 46, 610-621.
72. Bell, D. R., and Poland, A. (2000) Binding of aryl hydrocarbon receptor (AhR) to AhR-interacting protein. The role of hsp90, *J Biol Chem* 275, 36407-36414.
73. Pollenz, R. S., Sattler, C. A., and Poland, A. (1994) The aryl hydrocarbon receptor and aryl hydrocarbon receptor nuclear translocator protein show distinct subcellular localizations in Hepa 1c1c7 cells by immunofluorescence microscopy, *Mol Pharmacol* 45, 428-438.
74. Hahn, M. E. (1998) The aryl hydrocarbon receptor: a comparative perspective, *Comp Biochem Physiol C Pharmacol Toxicol Endocrinol* 121, 23-53.
75. Denison, M. S., and Nagy, S. R. (2003) Activation of the aryl hydrocarbon receptor by structurally diverse exogenous and endogenous chemicals, *Annu Rev Pharmacol Toxicol* 43, 309-334.
76. Mimura, J., Yamashita, K., Nakamura, K., Morita, M., Takagi, T. N., Nakao, K., Ema, M., Sogawa, K., Yasuda, M., Katsuki, M., and Fujii-



- Kuriyama, Y. (1997) Loss of teratogenic response to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) in mice lacking the Ah (dioxin) receptor, *Genes Cells* 2, 645-654.
77. Walisser, J. A., Bunger, M. K., Glover, E., Harstad, E. B., and Bradfield, C. A. (2004) Patent ductus venosus and dioxin resistance in mice harboring a hypomorphic Arnt allele, *J Biol Chem* 279, 16326-16331.
  78. Hosoya, T., Harada, N., Mimura, J., Motohashi, H., Takahashi, S., Nakajima, O., Morita, M., Kawauchi, S., Yamamoto, M., and Fujii-Kuriyama, Y. (2008) Inducibility of cytochrome P450 1A1 and chemical carcinogenesis by benzo[a]pyrene in AhR repressor-deficient mice, *Biochem Biophys Res Commun* 365, 562-567.
  79. Frericks, M., Temchura, V. V., Majora, M., Stutte, S., and Esser, C. (2006) Transcriptional signatures of immune cells in aryl hydrocarbon receptor (AHR)-proficient and AHR-deficient mice, *Biol Chem* 387, 1219-1226.
  80. Ivan, M., Kondo, K., Yang, H., Kim, W., Valiando, J., Ohh, M., Salic, A., Asara, J. M., Lane, W. S., and Kaelin, W. G., Jr. (2001) HIFalpha targeted for VHL-mediated destruction by proline hydroxylation: implications for O2 sensing, *Science* 292, 464-468.
  81. Yu, F., White, S. B., Zhao, Q., and Lee, F. S. (2001) HIF-1alpha binding to VHL is regulated by stimulus-sensitive proline hydroxylation, *Proc Natl Acad Sci U S A* 98, 9630-9635.
  82. Jaakkola, P., Mole, D. R., Tian, Y. M., Wilson, M. I., Gielbert, J., Gaskell, S. J., Kriegsheim, A., Hebestreit, H. F., Mukherji, M., Schofield, C. J., Maxwell, P. H., Pugh, C. W., and Ratcliffe, P. J. (2001) Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation, *Science* 292, 468-472.
  83. Maxwell, P. H., Wiesener, M. S., Chang, G. W., Clifford, S. C., Vaux, E. C., Cockman, M. E., Wykoff, C. C., Pugh, C. W., Maher, E. R., and Ratcliffe, P. J. (1999) The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis, *Nature* 399, 271-275.

84. Tanimoto, K., Makino, Y., Pereira, T., and Poellinger, L. (2000) Mechanism of regulation of the hypoxia-inducible factor-1 alpha by the von Hippel-Lindau tumor suppressor protein, *EMBO J* 19, 4298-4309.
85. Lisztwan, J., Imbert, G., Wirbelauer, C., Gstaiger, M., and Krek, W. (1999) The von Hippel-Lindau tumor suppressor protein is a component of an E3 ubiquitin-protein ligase activity, *Genes Dev* 13, 1822-1833.
86. Lando, D., Peet, D. J., Whelan, D. A., Gorman, J. J., and Whitelaw, M. L. (2002) Asparagine hydroxylation of the HIF transactivation domain a hypoxic switch, *Science* 295, 858-861.
87. Sang, N., Fang, J., Srinivas, V., Leshchinsky, I., and Caro, J. (2002) Carboxyl-terminal transactivation activity of hypoxia-inducible factor 1 alpha is governed by a von Hippel-Lindau protein-independent, hydroxylation-regulated association with p300/CBP, *Mol Cell Biol* 22, 2984-2992.
88. Bracken, C. P., Whitelaw, M. L., and Peet, D. J. (2003) The hypoxia-inducible factors: key transcriptional regulators of hypoxic responses, *Cell Mol Life Sci* 60, 1376-1393.
89. Vinson, C. (2005) A rationally designed small molecule that inhibits the HIF-1alpha-ARNT heterodimer from binding to DNA in vivo, *Sci STKE* 2005, pe23.
90. Gibson, G. (1989) Causes of cell damage in hypoxia/ischemia, aging and Alzheimer's disease, *Neurobiol Aging* 10, 608-609; discussion 618-620.
91. Rankin, E. B., Higgins, D. F., Walisser, J. A., Johnson, R. S., Bradfield, C. A., and Haase, V. H. (2005) Inactivation of the arylhydrocarbon receptor nuclear translocator (Arnt) suppresses von Hippel-Lindau disease-associated vascular tumors in mice, *Mol Cell Biol* 25, 3163-3172.
92. Stebbins, C. E., Kaelin, W. G., Jr., and Pavletich, N. P. (1999) Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function, *Science* 284, 455-461.
93. Shimizu, Y., Nakatsuru, Y., Ichinose, M., Takahashi, Y., Kume, H., Mimura, J., Fujii-Kuriyama, Y., and Ishikawa, T. (2000) Benzo[a]pyrene

- carcinogenicity is lost in mice lacking the aryl hydrocarbon receptor, *Proc Natl Acad Sci U S A* 97, 779-782.
94. Salomon-Nguyen, F., Della-Valle, V., Mauchauffe, M., Busson-Le Coniat, M., Ghysdael, J., Berger, R., and Bernard, O. A. (2000) The t(1;12)(q21;p13) translocation of human acute myeloblastic leukemia results in a TEL-ARNT fusion, *Proc Natl Acad Sci U S A* 97, 6757-6762.
  95. Elvert, G., Kappel, A., Heidenreich, R., Englmeier, U., Lanz, S., Acker, T., Rauter, M., Plate, K., Sieweke, M., Breier, G., and Flamme, I. (2003) Cooperative interaction of hypoxia-inducible factor-2alpha (HIF-2alpha ) and Ets-1 in the transcriptional activation of vascular endothelial growth factor receptor-2 (Flk-1), *J Biol Chem* 278, 7520-7530.
  96. Wang, X., Li, L., Zou, A., Tu, D., and Liao, Y. (2008) [Analysis and analyzing mechanisms of HERG channel kinetics], *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi* 25, 1068-1073.
  97. Paulussen, A. D., Raes, A., Jongbloed, R. J., Gilissen, R. A., Wilde, A. A., Snyders, D. J., Smeets, H. J., and Aerssens, J. (2005) HERG mutation predicts short QT based on channel kinetics but causes long QT by heterotetrameric trafficking deficiency, *Cardiovasc Res* 67, 467-475.
  98. Chen, J., Zou, A., Splawski, I., Keating, M. T., and Sanguinetti, M. C. (1999) Long QT syndrome-associated mutations in the Per-Arnt-Sim (PAS) domain of HERG potassium channels accelerate channel deactivation, *J Biol Chem* 274, 10113-10118.
  99. Morais Cabral, J. H., Lee, A., Cohen, S. L., Chait, B. T., Li, M., and Mackinnon, R. (1998) Crystal structure and functional analysis of the HERG potassium channel N terminus: a eukaryotic PAS domain, *Cell* 95, 649-655.
  100. Crosson, S., and Moffat, K. (2001) Structure of a flavin-binding plant photoreceptor domain: insights into light-mediated signal transduction, *Proc Natl Acad Sci U S A* 98, 2995-3000.
  101. Perego, M., Cole, S. P., Burbulys, D., Trach, K., and Hoch, J. A. (1989) Characterization of the gene for a protein kinase which phosphorylates the sporulation-regulatory proteins Spo0A and Spo0F of *Bacillus subtilis*, *J Bacteriol* 171, 6187-6196.

102. Burbulys, D., Trach, K. A., and Hoch, J. A. (1991) Initiation of sporulation in *B. subtilis* is controlled by a multicomponent phosphorelay, *Cell* 64, 545-552.
103. Zhong, X., Hao, B., and Chan, M. K. (2003) Structure of the PAS Fold and Signal Transduction Mechanisms, in PAS Proteins, *Regulators and Sensors of Development and Physiology*, 1-16.
104. Zhao, J. M., Lee, H., Nome, R. A., Majid, S., Scherer, N. F., and Hoff, W. D. (2006) Single-molecule detection of structural changes during Per-Arnt-Sim (PAS) domain activation, *Proc Natl Acad Sci U S A* 103, 11561-11566.
105. Nome, R. A., Zhao, J. M., Hoff, W. D., and Scherer, N. F. (2007) Axis-dependent anisotropy in protein unfolding from integrated nonequilibrium single-molecule experiments, analysis, and simulation, *Proc Natl Acad Sci U S A* 104, 20799-20804.
106. Dill, K. A., Ozkan, S. B., Weikl, T. R., Chodera, J. D., and Voelz, V. A. (2007) The protein folding problem: when will it be solved?, *Curr Opin Struct Biol* 17, 342-346.
107. Dill, K. A., Ozkan, S. B., Shell, M. S., and Weikl, T. R. (2008) The protein folding problem, *Annu Rev Biophys* 37, 289-316.
108. Sela, M., Anfinsen, C. B., and Harrington, W. F. (1957) The correlation of ribonuclease activity with specific aspects of tertiary structure, *Biochim Biophys Acta* 26, 502-512.
109. White, F. H., Jr. (1961) Regeneration of native secondary and tertiary structures by air oxidation of reduced ribonuclease, *J Biol Chem* 236, 1353-1360.
110. Anfinsen, C. B., and Haber, E. (1961) Studies on the reduction and re-formation of protein disulfide bonds, *J Biol Chem* 236, 1361-1363.
111. Zhang, O., and Forman-Kay, J. D. (1995) Structural characterization of folded and unfolded states of an SH3 domain in equilibrium in aqueous buffer, *Biochemistry* 34, 6784-6794.

112. Bondensgaard, K., Breinholt, J., Madsen, D., Omkvist, D. H., Kang, L., Worsaae, A., Becker, P., Schiodt, C. B., and Hjorth, S. A. (2007) The existence of multiple conformers of interleukin-21 directs engineering of a superpotent analogue, *J Biol Chem* 282, 23326-23336.
113. Franke, A. E., Danley, D. E., Kaczmarek, F. S., Hawrylik, S. J., Gerard, R. D., Lee, S. E., and Geoghegan, K. F. (1990) Expression of human plasminogen activator inhibitor type-1 (PAI-1) in *Escherichia coli* as a soluble protein comprised of active and latent forms. Isolation and crystallization of latent PAI-1, *Biochim Biophys Acta* 1037, 16-23.
114. Luo, X., Tang, Z., Rizo, J., and Yu, H. (2002) The Mad2 spindle checkpoint protein undergoes similar major conformational changes upon binding to either Mad1 or Cdc20, *Mol Cell* 9, 59-71.
115. Luo, X., Tang, Z., Xia, G., Wassmann, K., Matsumoto, T., Rizo, J., and Yu, H. (2004) The Mad2 spindle checkpoint protein has two distinct natively folded states, *Nat Struct Mol Biol* 11, 338-345.
116. Wagner, A. C., Wishart, M. J., Mulders, S. M., Blevins, P. M., Andrews, P. C., Lowe, A. W., and Williams, J. A. (1994) GP-3, a newly characterized glycoprotein on the inner surface of the zymogen granule membrane, undergoes regulated secretion, *J Biol Chem* 269, 9099-9104.
117. Kuloglu, E. S., McCaslin, D. R., Markley, J. L., and Volkman, B. F. (2002) Structural rearrangement of human lymphotactin, a C chemokine, under physiological solution conditions, *J Biol Chem* 277, 17863-17870.
118. Laskey, R. A., Honda, B. M., Mills, A. D., and Finch, J. T. (1978) Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA, *Nature* 275, 416-420.
119. Ellis, R. J. (1996) Discovery of molecular chaperones, *Cell Stress Chaperones* 1, 155-160.
120. Shtilerman, M., Lorimer, G. H., and Englander, S. W. (1999) Chaperonin function: folding by forced unfolding, *Science* 284, 822-825.
121. Yon, J. M. (2002) Protein folding in the post-genomic era, *J Cell Mol Med* 6, 307-327.

122. Knight, K. L., Bowie, J. U., Vershon, A. K., Kelley, R. D., and Sauer, R. T. (1989) The Arc and Mnt repressors. A new class of sequence-specific DNA-binding protein, *J Biol Chem* 264, 3639-3642.
123. Pabo, C. O., and Sauer, R. T. (1984) Protein-DNA recognition, *Annu Rev Biochem* 53, 293-321.
124. Breg, J. N., van Opheusden, J. H., Burgering, M. J., Boelens, R., and Kaptein, R. (1990) Structure of Arc repressor in solution: evidence for a family of beta-sheet DNA-binding proteins, *Nature* 346, 586-589.
125. Cordes, M. H., Walsh, N. P., McKnight, C. J., and Sauer, R. T. (2003) Solution structure of switch Arc, a mutant with 3(10) helices replacing a wild-type beta-ribbon, *J Mol Biol* 326, 899-909.
126. Cordes, M. H., Burton, R. E., Walsh, N. P., McKnight, C. J., and Sauer, R. T. (2000) An evolutionary bridge to a new protein fold, *Nat Struct Biol* 7, 1129-1132.
127. Prusiner, S. B. (1994) Molecular biology and genetics of prion diseases, *Philos Trans R Soc Lond B Biol Sci* 343, 447-463.
128. Huang, Z., Gabriel, J. M., Baldwin, M. A., Fletterick, R. J., Prusiner, S. B., and Cohen, F. E. (1994) Proposed three-dimensional structure for the cellular prion protein, *Proc Natl Acad Sci U S A* 91, 7139-7143.
129. Komiyama, T., Ray, C. A., Pickup, D. J., Howard, A. D., Thornberry, N. A., Peterson, E. P., and Salvesen, G. (1994) Inhibition of interleukin-1 beta converting enzyme by the cowpox virus serpin CrmA. An example of cross-class inhibition, *J Biol Chem* 269, 19331-19337.
130. Schick, C., Pemberton, P. A., Shi, G. P., Kamachi, Y., Cataltepe, S., Bartuski, A. J., Gornstein, E. R., Bromme, D., Chapman, H. A., and Silverman, G. A. (1998) Cross-class inhibition of the cysteine proteinases cathepsins K, L, and S by the serpin squamous cell carcinoma antigen 1: a kinetic analysis, *Biochemistry* 37, 5258-5266.
131. Stein, P. E., Leslie, A. G., Finch, J. T., and Carrell, R. W. (1991) Crystal structure of uncleaved ovalbumin at 1.95 Å resolution, *J Mol Biol* 221, 941-959.

132. Whisstock, J. C., Skinner, R., Carrell, R. W., and Lesk, A. M. (2000) Conformational changes in serpins: I. The native and cleaved conformations of alpha(1)-antitrypsin, *J Mol Biol* 296, 685-699.
133. Plotnick, M. I., Samakur, M., Wang, Z. M., Liu, X., Rubin, H., Schechter, N. M., and Selwood, T. (2002) Heterogeneity in serpin-protease complexes as demonstrated by differences in the mechanism of complex breakdown, *Biochemistry* 41, 334-342.
134. Huber, R., and Carrell, R. W. (1989) Implications of the three-dimensional structure of alpha 1-antitrypsin for structure and function of serpins, *Biochemistry* 28, 8951-8966.
135. Declerck, P. J., De Mol, M., Alessi, M. C., Baudner, S., Paques, E. P., Preissner, K. T., Muller-Berghaus, G., and Collen, D. (1988) Purification and characterization of a plasminogen activator inhibitor 1 binding protein from human plasma. Identification as a multimeric form of S protein (vitronectin), *J Biol Chem* 263, 15454-15461.
136. Mottonen, J., Strand, A., Symersky, J., Sweet, R. M., Danley, D. E., Geoghegan, K. F., Gerard, R. D., and Goldsmith, E. J. (1992) Structural basis of latency in plasminogen activator inhibitor-1, *Nature* 355, 270-273.
137. Rossi, D., and Zlotnik, A. (2000) The biology of chemokines and their receptors, *Annu Rev Immunol* 18, 217-242.
138. Zoltowski, B. D., Schwerdtfeger, C., Widom, J., Loros, J. J., Bilwes, A. M., Dunlap, J. C., and Crane, B. R. (2007) Conformational switching in the fungal light sensor Vivid, *Science* 316, 1054-1057.
139. Lees, M. J., and Whitelaw, M. L. (1999) Multiple roles of ligand in transforming the dioxin receptor to an active basic helix-loop-helix/PAS transcription factor complex with the nuclear protein Arnt, *Mol Cell Biol* 19, 5811-5822.
140. Wang, G. L., Jiang, B. H., Rue, E. A., and Semenza, G. L. (1995) Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O<sub>2</sub> tension, *Proc Natl Acad Sci U S A* 92, 5510-5514.

141. Chapman-Smith, A., and Whitelaw, M. L. (2006) Novel DNA binding by a basic helix-loop-helix protein. The role of the dioxin receptor PAS domain, *J Biol Chem* 281, 12535-12545.
142. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., Rance, M., and Skelton, N. J. (2007) *Protein NMR Spectroscopy: Principles and Practice*, Elsevier, San Diego, CA.
143. Farrow, N. A., Zhang, O., Forman-Kay, J. D., and Kay, L. E. (1994) A heteronuclear correlation experiment for simultaneous determination of <sup>15</sup>N longitudinal decay and chemical exchange rates of systems in slow equilibrium, *J Biomol NMR* 4, 727-734.
144. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr D Biol Crystallogr* 54, 905-921.
145. Linge, J. P., O'Donoghue, S. I., and Nilges, M. (2001) Automated assignment of ambiguous nuclear overhauser effects with ARIA, *Methods Enzymol* 339, 71-90.
146. Schubert, M., Labudde, D., Oschkinat, H., and Schmieder, P. (2002) A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on <sup>13</sup>C chemical shift statistics, *J Biomol NMR* 24, 149-154.
147. Borgstahl, G. E., Williams, D. R., and Getzoff, E. D. (1995) 1.4 Å structure of photoactive yellow protein, a cytosolic photoreceptor: unusual fold, active site, and chromophore, *Biochemistry* 34, 6278-6287.
148. Key, J., and Moffat, K. (2005) Crystal structures of deoxy and CO-bound bJFixLH reveal details of ligand recognition and signaling, *Biochemistry* 44, 4627-4635.
149. Lee, B. C., Pandit, A., Croonquist, P. A., and Hoff, W. D. (2001) Folding and signaling share the same pathway in a photoreceptor, *Proc Natl Acad Sci U S A* 98, 9062-9067.



150. Yao, X., Rosen, M. K., and Gardner, K. H. (2008) Estimation of the available free energy in a LOV2-J alpha photoswitch, *Nat Chem Biol* 4, 491-497.
151. Nozaki, D., Iwata, T., Ishikawa, T., Todo, T., Tokutomi, S., and Kandori, H. (2004) Role of Gln1029 in the photoactivation processes of the LOV2 domain in adiantum phytochrome3, *Biochemistry* 43, 8373-8379.
152. Jones, M. A., Feeney, K. A., Kelly, S. M., and Christie, J. M. (2007) Mutational analysis of phototropin 1 provides insights into the mechanism underlying LOV2 signal transmission, *J Biol Chem* 282, 6405-6414.
153. Minor, D. L., Jr., and Kim, P. S. (1994) Context is a major determinant of beta-sheet propensity, *Nature* 371, 264-267.
154. Smith, C. K., Withka, J. M., and Regan, L. (1994) A thermodynamic scale for the beta-sheet forming tendencies of the amino acids, *Biochemistry* 33, 5510-5517.
155. Lu, K. P., Finn, G., Lee, T. H., and Nicholson, L. K. (2007) Prolyl cis-trans isomerization as a molecular timer, *Nat Chem Biol* 3, 619-629.
156. Crosson, S., Rajagopal, S., and Moffat, K. (2003) The LOV domain family: photoresponsive signaling modules coupled to diverse output domains, *Biochemistry* 42, 2-10.
157. Goldberg, J. (1998) Structural basis for activation of ARF GTPase: mechanisms of guanine nucleotide exchange and GTP-myristoyl switching, *Cell* 95, 237-248.
158. Eigenbrot, C., Kirchhofer, D., Dennis, M. S., Santell, L., Lazarus, R. A., Stamos, J., and Ultsch, M. H. (2001) The factor VII zymogen structure reveals reregistration of beta strands during activation, *Structure* 9, 627-636.
159. Maun, H. R., Eigenbrot, C., Raab, H., Arnott, D., Phu, L., Bullens, S., and Lazarus, R. A. (2005) Disulfide locked variants of factor VIIa with a restricted beta-strand conformation have enhanced enzymatic activity, *Protein Sci* 14, 1171-1180.

160. Eakin, C. M., Berman, A. J., and Miranker, A. D. (2006) A native to amyloidogenic transition regulated by a backbone trigger, *Nat Struct Mol Biol* 13, 202-208.
161. Eneqvist, T., Andersson, K., Olofsson, A., Lundgren, E., and Sauer-Eriksson, A. E. (2000) The beta-slip: a novel concept in transthyretin amyloidosis, *Mol Cell* 6, 1207-1218.
162. Kinch, L. N., and Grishin, N. V. (2002) Evolution of protein structures and functions, *Curr Opin Struct Biol* 12, 400-408.
163. Sheffield, P., Garrard, S., and Derewenda, Z. (1999) Overcoming expression and purification problems of RhoGDI using a family of "parallel" expression vectors, *Protein Expr Purif* 15, 34-39.
164. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) Primary structure effects on peptide group hydrogen exchange, *Proteins* 17, 75-86.
165. Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology, *J Biomol NMR* 13, 289-302.
166. Maxwell, K. L., Wildes, D., Zarrine-Afsar, A., De Los Rios, M. A., Brown, A. G., Friel, C. T., Hedberg, L., Horng, J. C., Bona, D., Miller, E. J., Vallee-Belisle, A., Main, E. R., Bemporad, F., Qiu, L., Teilum, K., Vu, N. D., Edwards, A. M., Ruczinski, I., Poulsen, F. M., Kragelund, B. B., Michnick, S. W., Chiti, F., Bai, Y., Hagen, S. J., Serrano, L., Oliveberg, M., Raleigh, D. P., Wittung-Stafshede, P., Radford, S. E., Jackson, S. E., Sosnick, T. R., Marqusee, S., Davidson, A. R., and Plaxco, K. W. (2005) Protein folding: defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins, *Protein Sci* 14, 602-616.
167. Privalov, P. L. (1979) Stability of proteins: small globular proteins, *Adv Protein Chem* 33, 167-241.
168. Brandts, J. F., Halvorson, H. R., and Brennan, M. (1975) Consideration of the Possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues, *Biochemistry* 14, 4953-4963.

169. Pal, D., and Chakrabarti, P. (1999) Cis peptide bonds in proteins: residues involved, their conformations, interactions and locations, *J Mol Biol* 294, 271-288.
170. Eitoku, T., Nakasone, Y., Zikihara, K., Matsuoka, D., Tokutomi, S., and Terazima, M. (2007) Photochemical intermediates of Arabidopsis phototropin 2 LOV domains associated with conformational changes, *J Mol Biol* 371, 1290-1303.
171. Lim, M., Jackson, T. A., and Anfinrud, P. A. (1997) Ultrafast rotation and trapping of carbon monoxide dissociated from myoglobin, *Nat Struct Biol* 4, 209-214.
172. Lummis, S. C., Beene, D. L., Lee, L. W., Lester, H. A., Broadhurst, R. W., and Dougherty, D. A. (2005) Cis-trans isomerization at a proline opens the pore of a neurotransmitter-gated ion channel, *Nature* 438, 248-252.
173. Mallis, R. J., Brazin, K. N., Fulton, D. B., and Andreotti, A. H. (2002) Structural characterization of a proline-driven conformational switch within the Itk SH2 domain, *Nat Struct Biol* 9, 900-905.
174. Eckert, B., Martin, A., Balbach, J., and Schmid, F. X. (2005) Prolyl isomerization as a molecular timer in phage infection, *Nat Struct Mol Biol* 12, 619-623.
175. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J Biomol NMR* 6, 277-293.
176. Stewart, D. E., Sarkar, A., and Wampler, J. E. (1990) Occurrence and role of cis peptide bonds in protein structures, *J Mol Biol* 214, 253-260.
177. Weiss, M. S., Metzner, H. J., and Hilgenfeld, R. (1998) Two non-proline cis peptide bonds may be important for factor XIII function, *FEBS Lett* 423, 291-296.
178. Tweedy, N. B., Nair, S. K., Paterno, S. A., Fierke, C. A., and Christianson, D. W. (1993) Structure and energetics of a non-proline cis-peptidyl linkage in a proline-202-->alanine carbonic anhydrase II variant, *Biochemistry* 32, 10944-10949.

179. MacArthur, M. W., and Thornton, J. M. (1991) Influence of proline residues on protein conformation, *J Mol Biol* 218, 397-412.
180. Jabs, A., Weiss, M. S., and Hilgenfeld, R. (1999) Non-proline cis peptide bonds in proteins, *J Mol Biol* 286, 291-304.
181. Amezcua, C. A., Harper, S. M., Rutter, J., and Gardner, K. H. (2002) Structure and interactions of PAS kinase N-terminal PAS domain: model for intramolecular kinase regulation, *Structure* 10, 1349-1361.
182. Pellequer, J. L., Brudler, R., and Getzoff, E. D. (1999) Biological sensors: More than one way to sense oxygen, *Curr Biol* 9, R416-418.
183. Pantoliano, M. W., Petrella, E. C., Kwasnoski, J. D., Lobanov, V. S., Myslik, J., Graf, E., Carver, T., Asel, E., Springer, B. A., Lane, P., and Salemme, F. R. (2001) High-density miniaturized thermal shift assays as a general strategy for drug discovery, *J Biomol Screen* 6, 429-440.
184. Skrynnikov, N. R., Dahlquist, F. W., and Kay, L. E. (2002) Reconstructing NMR spectra of "invisible" excited protein states using HSQC and HMQC experiments, *J Am Chem Soc* 124, 12352-12360.
185. Ishima, R., and Torchia, D. A. (2003) Extending the range of amide proton relaxation dispersion experiments in proteins using a constant-time relaxation-compensated CPMG approach, *J Biomol NMR* 25, 243-248.
186. Wang, C., Grey, M. J., and Palmer, A. G., 3rd. (2001) CPMG sequences with enhanced sensitivity to chemical exchange, *J Biomol NMR* 21, 361-366.