

PAS DOMAIN-MEDIATED DIMERIZATION OF THE ARYL HYDROCARBON  
RECEPTOR NUCLEAR TRANSLOCATOR (ARNT) IN THE  
HYPOXIA RESPONSE PATHWAY

APPROVED BY SUPERVISORY COMMITTEE

---

---

---

---

## DEDICATION

I would like to thank my mentor Kevin H. Gardner for all that he has taught me, the members of my graduate committee (Jose Rizo-Rey, Betsy Goldsmith and Steve McKnight) for their inspired guidance, my parents Lewis and Gretchen for their endless support, my wife Melissa for everything she does for our family, and my daughter Vanessa for making it all worthwhile.



PAS DOMAIN-MEDIATED DIMERIZATION OF THE ARYL HYDROCARBON  
RECEPTOR NUCLEAR TRANSLOCATOR (ARNT) IN THE HYPOXIA RESPONSE  
PATHWAY

by

PAUL B. CARD

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

December, 2005

Copyright

by

Paul B. Card, 2005

All Rights Reserved

## PAS DOMAIN-MEDIATED SIGNALING IN THE HYPOXIA RESPONSE

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

Paul B. Card, Ph.D

The University of Texas Southwestern Medical Center at Dallas, 2005

Supervising Professor: Kevin H. Gardner, Ph.D

PAS (Per-ARNT-SIM) domains are versatile protein-protein interaction domains that are often used as regulatory modules in a variety of important biological pathways. Although their importance in several of these pathways has been well established, only sparse structural data exists that could help elucidate a general mode of PAS-PAS interaction. As such, more examples of these domains must be studied using a variety of techniques to understand how these domains carry out viable functions within the cell.

The Aryl Hydrocarbon Receptor Nuclear Translocator (ARNT) is a constituent of heterodimeric transcriptional activation complexes used in several important biochemical pathways. One such complex is the hypoxia inducible transcription factor (HIF), which

allows mammalian cells to respond to changes in oxygen availability. In HIF, ARNT dimerizes with HIF $\alpha$  to upregulate genes involved in the hypoxia response. The dimerization of ARNT and HIF $\alpha$  involves interactions between PAS domains in both proteins, but details regarding how these domains interact in this and other heterodimeric complexes have not previously been well characterized.

To investigate these interactions within the context of the HIF/ARNT heterodimer, we expressed the C-terminal PAS domains from both proteins and characterized the complex using NMR-based methods. Solution structures of each domain are presented, as well as a model of the ARNT/HIF heterodimeric PAS complex. This model was used to identify of key interfacial residues, and the roles of these were tested in a variety of ways by site-directed mutagenesis. In addition, extended constructs from ARNT that include other components of the full-length protein were investigated to establish the validity of a reductionist approach in the study of individual PAS domains from this system.

In many biological systems, PAS domains bind small molecules to regulate protein-protein interactions. With this in mind, we also subjected PAS domains from HIF $\alpha$  and ARNT to an NMR-based screen against an in-house library of 800 compounds to determine the potential of these domains to bind small molecules. We have used this approach to identify compounds that can disrupt PAS-PAS interactions in the hypoxia pathway in order to help elucidate some of the molecular details of PAS domain-mediated signaling.

## TABLE OF CONTENTS

<u>I. Introduction - PAS domain-mediated signaling</u>	1
A. PAS domains	1
i. PAS domain structure	2
ii. PAS domains as environmental sensors	4
a. Redox state and oxygen sensing	5
b. PAS domains as light sensors	7
iii. PAS-mediated protein-protein interactions	8
a. PAS domains in the activation of kinases	8
b. Variable stoichiometry of PAS complexes	10
B. PAS domains as modular components of transcription factors	11
i. bHLH PAS transcription factors	12
a. The Aryl hydrocarbon receptor	14
b. The Hypoxia Inducible Factor	16
c. The Aryl hydrocarbon Receptor Nuclear Translocator	18
C. Ligand binding in PAS domains	20
 <u>II. Domain optimization for NMR studies – White Collar</u>	 22
A. The <i>Neurospora</i> circadian clock	22
B. Domain optimization for NMR studies	28
i. Computational methods	29
ii. Domain prediction	30
a. White Collar-1 and White Collar-2 domain prediction	32
iii. Secondary structure prediction	33
a. White Collar secondary structure prediction	34
iv. Parallel cloning and expression	37
a. Cloning and expression of WC PAS domains	39
v. Whole cell lysate screening	40
a. Cell lysate screening of WC PAS domains	41
b. Expression and purification of WC-1 PAS-B	44
vi. Solubility screening	48
a. Solvent screening of WC-1 PAS-B	49

<u>III. ARNT in the Hypoxia Response</u>	52
A. The Aryl hydrocarbon Receptor Nuclear Translocator	52
B. Solution structure of ARNT PAS-B	55
i. Cloning, expression and purification of ARNT PAS-B for structure determination	55
ii. Solution Structure of ARNT PAS-B	56
a. Materials and methods	56
b. Results	59
iii. ARNT PAS-B Forms Heterodimers via the central $\beta$ -sheet	62
C. Model of the ARNT PAS-B/HIF PAS-B heterodimer	65
i. Model of HIF-2 $\alpha$ /ARNT PAS-B Complex	66
ii. Spin-labeling of ARNT PAS-B	70
a. Paramagnetic broadening enhancement theory	71
b. Spin labeling materials and methods	72
c. Analysis of paramagnetic broadening enhancement data	76
iii. Relevance to other systems	83
<u>IV. ARNT PAS-B Homodimerization and Mutagenesis</u>	89
A. ARNT PAS-B homodimerization	89
i. Characterization of ARNT PAS-B homodimer	90
B. ARNT PAS-B Mutagenesis	93
i. Generation of ARNT PAS-B mutants	94
ii. Multiple conformations of Y456T mutation	95
iii. F444Q shifts the equilibrium of the conformational transition	98
iv. Further shifting of the equilibrium	99
v. Proposed mechanism for the conformational change	100
<u>V. Ligand Screening of ARNT PAS-B</u>	105
i. Ligand screening strategy and methods	109
ii. Dissociation constant measurement methods	112
iii. Choice of library compounds	112
a. Compound specificity	113

iv. Lead compounds identified by the initial ligand screen of ARNT PAS-B	114
a. Determination of dissociation constants	120
v. <sup>19</sup> F NMR of compound KG-548	120
vi. KG-548 and KG-655 attenuate ARNT PAS-B dimerization	124
a. Disruption of the ARNT PAS-B/HIF-2 $\alpha$ PAS-B heterodimer	125
b. Effects of ligand binding on complex formation of full-length proteins in vivo	128
vii. Mapping the binding site of KG-548 in ARNT PAS-B	130
viii. Alternative explanation for heterodimer disruption	133
ix. Structure activity relationships (SAR) of ARNT PAS-B small molecule ligands	136
x. Conclusions	139
 <u>VI. Larger ARNT Constructs</u>	 141
A. Limitations of the reductionist approach	141
i. Initial expression studies of larger ARNT constructs	142
B. ARNT PAS-A domain	146
i. Expression and purification of ARNT PAS-A (144-350)	147
ii. NMR spectroscopy of ARNT PAS-A (144-350)	149
iii. Relevance to other multi-PAS proteins	155
C. Multi-PAS domain construct: ARNT PAS-AB	156
i. Expression and purification of ARNT PAS-AB (155-470)	157
ii. NMR of ARNT PAS-AB (155-470)	158
D. C-terminal extensions of ARNT PAS-B and PAS-AB constructs	160
i. Extended ARNT PAS-B (356-540)	161
a. NMR of C-terminal extended ARNT PAS-B (356-540)	161
b. Rotational correlation time determination	165
ii. Extended ARNT PAS-AB (155-540)	167
a. Purification and size exclusion chromatography of ARNT PAS-AB (155-540)	167
b. NMR of ARNT PAS-AB (155-540)	170
iii. Conclusions	174

## PRIOR PUBLICATIONS

Paul B. Card, Paul J.A. Erbel, and Kevin H. Gardner (2005) “Structural basis of ARNT PAS-B dimerization: Use of a common beta-sheet interface for hetero- and homodimerization” *Journal of Molecular Biology*, **353**: 664-677

Paul B. Card and Kevin H. Gardner (2005) “Identification and optimization of protein domains for NMR studies” *Methods in Enzymology*, **394**: 3-18

Paul J.A. Erbel, Paul B. Card, Ozgur Karakuzu, Richard K. Bruick, and Kevin H. Gardner (2003) “Structural basis for PAS domain heterodimerization in the bHLH-PAS transcription factor HIF” *Proc. Natl. Acad. Sci.*, **100**: 15504-15509



## LIST OF FIGURES

### Chapter I: introduction

Figure 1-1: PAS domain fold.....	4
Figure 1-2: AhR pathway .....	14
Figure 1-3: The hypoxia response pathway .....	17

### Chapter II: Domain optimization for NMR studies – White Collar

Figure 2-1: <i>N.crassa</i> circadian clock .....	24
Figure 2-2: White Collar domain prediction.....	33
Figure 2-3: JPred domain prediction.....	36
Figure 2-4: Parallel expression .....	37
Figure 2-5: White Collar constructs.....	39
Figure 2-6: Whole cell lysate screening .....	43
Figure 2-7: WC-1 PAS-B expression and purification.....	46
Figure 2-8: <sup>15</sup> N/ <sup>1</sup> H HSQC spectra of WC-1 PAS-B .....	47

### Chapter III: ARNT in the hypoxia response

Figure 3-1: ARNT PAS-B structure.....	58
Figure 3-2: ARNT sequence alignment.....	61
Figure 3-3: HIF-2 $\alpha$ /ARNT PAS-B association .....	63

Figure 3-4: HIF-2 $\alpha$ /ARNT PAS-B titration .....	64
Figure 3-5: Common $\beta$ -sheet interface.....	65
Figure 3-6: HIF-2 $\alpha$ /ARNT PAS-B complex model.....	69
Figure 3-7: Electrostatic surface potential mapping .....	70
Figure 3-8: HIF-2 $\alpha$ /ARNT two-dimensional model.....	70
Figure 3-9: ARNT PAS-B wild-type cysteines.....	73
Figure 3-10: CMTSL-spin label.....	74
Figure 3-11: $^{15}\text{N}/^1\text{H}$ HSQC spectra.....	75
Figure 3-12: Spin label control experiment .....	77
Figure 3-13: Alternate spin labels .....	78
Figure 3-14: Alternate spin label analysis.....	80
Figure 3-15: Spin labeling results.....	82
Figure 3-16: BjFixL dimer – two orientations.....	84
Figure 3-17: Per homodimer structure .....	85
Figure 3-18: New model for HIF/ARNT interaction .....	88

#### Chapter IV: ARNT PAS-B homodimerization and mutagenesis

Figure 4-1: ARNT homo- and heterodimer interface .....	92
Figure 4-2: ARNT PAS-B mutations.....	95
Figure 4-3: $^{15}\text{N}/^1\text{H}$ HSQC spectra of ARNT PAS-B mutations.....	97
Figure 4-4: ARNT PAS-B mutations disrupt homodimer.....	99

Figure 4-5: ARNT PAS-B mutations – proposed mechanism.....	102
--	-----

## Chapter V: Ligand screening of ARNT PAS-B

Figure 5-1: ARNT mutations – CMTSL reactivity.....	108
Figure 5-2: Screening protocol .....	111
Figure 5-3: Compound specificity.....	114
Figure 5-4: Screening results .....	115
Figure 5-5: Compound titrations.....	116
Figure 5-6: Titration analysis.....	119
Figure 5-7: <sup>19</sup> F NMR analysis .....	122
Figure 5-8: Heterodimer disruption.....	127
Figure 5-9: Co-immunoprecipitation assays.....	129
Figure 5-10: Methyl groups involved in binding.....	131
Figure 5-11: Control experiments .....	135
Figure 5-12: SAR studies.....	137
Figure 5-13: Similar binding of compounds.....	138

## Chapter VI: Larger ARNT constructs

Figure 6-1: ARNT constructs – secondary structure.....	145
Figure 6-2: SDS-PAGE analysis – ARNT PAS-A .....	148
Figure 6-3: 1D <sup>1</sup> H NMR analysis.....	150

Figure 6-4: $^{15}\text{N}/^1\text{H}$ HSQC spectra of ARNT PAS-A .....	153
Figure 6-5: SDS-PAGE analysis – ARNT PAS-AB .....	157
Figure 6-6: $^{15}\text{N}/^1\text{H}$ HSQC spectra of ARNT PAS-AB .....	158
Figure 6-7: $^{15}\text{N}/^1\text{H}$ HSQC spectra of ARNT PAS-B extension.....	162
Figure 6-8: ARNT PAS-B extension – complex with HIF-2 $\alpha$ PAS-B .....	163
Figure 6-9: ARNT PAS-B extension – serial dilutions .....	164
Figure 6-10: SDS-PAGE analysis – ARNT PAS-AB extension .....	168
Figure 6-11: Size exclusion - ARNT PAS-AB extension .....	169
Figure 6-12: $^{15}\text{N}/^1\text{H}$ HSQC spectra of ARNT PAS-AB extension.....	171
Figure 6-13: Complexes with HIF-2 $\alpha$ PAS-B.....	173

## List of Tables

Table 3-1: ARNT PAS-B structure statistics .....	59
Table 3-2: ARNT/HIF-2 $\alpha$ PAS-B complex model statistics .....	68
Table 6-1: ARNT PAS-B rotational correlation times .....	166

## I. Introduction - PAS domain-mediated signaling

It is well understood that cell survival often depends on robust adaptive responses that allow cells to sense and react to external stimuli in order to maintain physiological homeostasis over a wide range of environmental conditions. In addition, biochemical pathways are often modulated by exogenous signals to maintain cyclic behaviors, cellular metabolism, and/or growth patterns of the organism as a whole. Regulation of these mechanisms is strictly controlled within the cellular environment, and perturbations of these controls can result in a wide variety of disease states. The elucidation of signal transduction mechanisms is therefore an important field of biological study, and one that benefits from wide varieties of methodologies and collaborations between scientists working with both *in vivo* and *in vitro* systems. The purpose of my research is to understand the structural basis for signal transduction in a particular class of protein domains involved in many biochemical pathways, and through collaborative efforts with scientists of other disciplines, discover methods for modulating these signaling pathways for the possible development of therapeutic strategies.

### *A. PAS domains*

PAS domains are modular protein-protein interaction domains that are involved in signal transduction pathways in all three kingdoms of life: *Bacteria*, *Archaea*, and *Eucarya* (9). The PAS domain nomenclature is derived from the first three proteins in which they were identified: The *Drosophila* period clock protein (PER) was discovered by virtue of its involvement in the regulation of circadian rhythms (10, 11). The

vertebrate Aryl Hydrocarbon Receptor Nuclear Translocator (ARNT) was identified as an essential component for AHR signaling (12), and *Drosophila* single-minded protein (SIM) is a regulatory protein controlling midline cell lineage (13, 14). These domains are combined with other types of regulatory or effector modules to make many different proteins with a wide variety of functions including chemoreceptors and photoreceptors used in the regulation of movement toward or away from external stimuli, ion channels, kinases, phosphodiesterases, and transcriptional activators. Extensive sequence alignments of PAS-like sequences from many different types of organisms revealed that PAS domains are not confined to any specific phylogenetic group. The first PAS-containing proteins found in eukaryotes (PER, ARNT and SIM) were found to contain two PAS domains, which may have diverged in origin (15). However, proteins from microbial organisms with fully sequenced genomes sometimes contain multiple PAS domains of variable evolutionary origin. For example, the cyanobacterial protein Slr0222 contains six PAS domains, each of which has more similarity to PAS domains from archaeal and human proteins than to each other (15). This implies that the presence of multiple PAS domains in many proteins is not merely a consequence of simple domain duplication for the purpose of signal amplification, but could instead represent a more complicated regulatory system with multiple input signals all contained within the context of a single sensory protein (15).

#### *i. PAS domain structure*

The discovery of PAS domains as a distinct type of protein module that exists in a large variety of prokaryotic and eukaryotic protein sequences can be attributed to vast

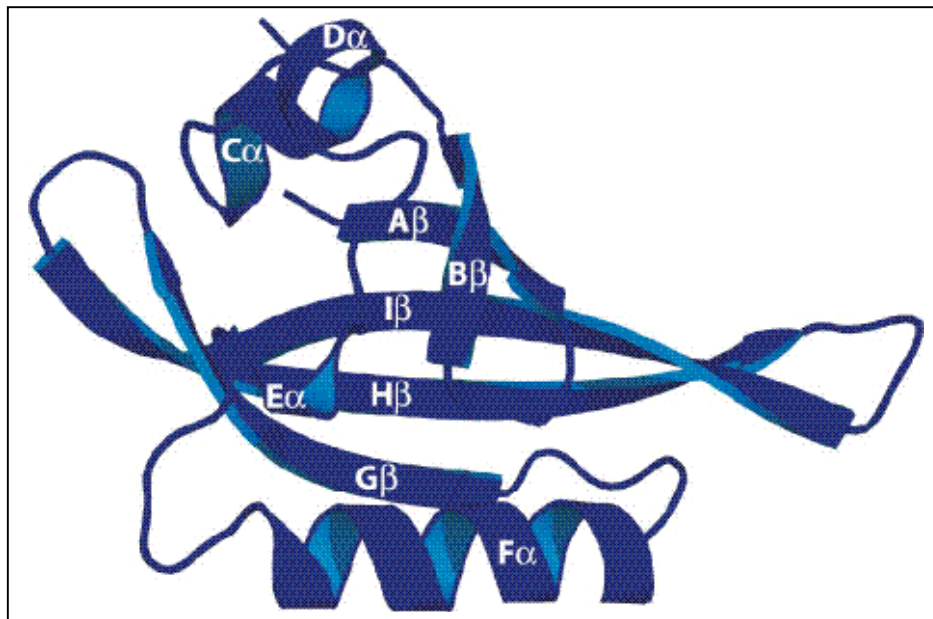
improvements made in computer-based sequence analysis. Multiple analyses have been used for the identification of these domains, due to the common confusion that has historically plagued the prediction of PAS domain sequences before the development of the Gapped- and PSI-BLAST programs (16). For example, the PAS domain contained in the N-terminal region of the Aer signal transducing protein was originally identified from sequence alignments with the known PAS sequences of ARNT and Sim, and evidence that this region served as a functional domain came from comparing results from the PhD secondary structure prediction server against the known three-dimensional structures of prototypical PAS domains such as PYP, FixL, and HERG (15, 17).

The PAS fold is characterized by a five-stranded anti-parallel  $\beta$ -sheet that is flanked by a series of  $\alpha$ -helices that define the core of the domain on one side, and exposed to solvent on the other (Fig. 1-1). The solvent-exposed face of the central  $\beta$ -sheet has been implicated as a likely interaction interface that PAS domains from several systems use to associate with other PAS domains or other components of their respective signaling network (8) (discussed in more detail for other systems in following sections). Alignment of many PAS domains reveals that the more highly conserved N-terminal and C-terminal regions of identified PAS domains, previously termed S1- and S2-boxes (or PAS and PAC), are separated by sequences that are highly variable in both composition and length (18). The PAS and PAC nomenclatures correspond to different internal segments within the  $\alpha/\beta$  fold. The N-terminal of these (PAS or S1 box) include the A $\beta$ -F $\alpha$  segments, while the C-terminal regions (PAC or S2 box) provide the three final G $\beta$ -I $\beta$  strands of the  $\beta$ -sheet. The sequence variability between these elements could help



explain the highly differentiated specificities that PAS domains exhibit for their binding partners in many different biochemical pathways.

**Fig 1-1. The overall PAS domain fold is represented with distinctive secondary structural elements labeled. The structure shown is that of the ARNT PAS-B domain (discussed in Chapter III).**



## *ii. PAS domains as environmental sensors*

Biological switches used for the regulation of biochemical pathways are important for any complex organism, and most aspects of cellular function in metazoans are dependent upon external signaling mechanisms (19). Intracellular processes are therefore dependent upon robust and specific responses to extracellular stimuli in the form of signaling switches. Molecular switches transduce signals from external sources to affect biological responses, and must have the capacity to both sense the signal and transmit information to downstream biochemical components. There are several forms of signal

that can be used for this purpose. For example, post-translational modifications such as hydroxylation, acetylation, methylation, ubiquitination, and phosphorylation can serve to modify a protein binding site which affects its ability to interact with other components in multiprotein complexes (20). An alternative strategy is the binding of small molecule ligands that can serve as sensory components within the signaling protein.

PAS domains are often used as sensor modules that monitor changes in redox potential, energy levels, light and oxygen tension in the cell. In this way, PAS domains are important components of regulatory pathways that monitor and restore homeostasis. The methods by which PAS domains sense and transduce signals include both post-translational modifications and the binding of small molecule ligands, although the mechanism of PAS domain sensory function is unknown for many systems.

#### *a. Redox state and oxygen sensing*

It is thought that decreased electron transport is an important first indicator of cellular energy level depletion, and as such there have evolved many mechanisms to sense changes in redox potential inside the cell (15, 21). Oxygen is a terminal acceptor for oxidative phosphorylation, which due to its high yield of ATP, is a crucial pathway for the storage of cellular energy. The *E.coli* protein Aer takes advantage of this relationship to indirectly sense oxygen using an FAD cofactor for the purpose of aerotaxis, instead of the more commonly known methylation-dependent mechanisms for chemotaxis in bacteria and archaea (22). Aer contains an N-terminal cytoplasmic PAS domain that detects oxygen-dependent cellular redox changes through its FAD prosthetic group (23). Alternatively, the *E.coli* Arc (anoxic redox control) two-component signal

transduction system involves a transmembrane sensor kinase (ArcB), that does not contain a bound cofactor, but senses aerobic conditions through the quinone-dependent oxidation of two cytosol-located redox-active cysteine residues that mediate intermolecular disulfide bond formation to regulate its kinase activity (24).

The sensing of oxygen is also vital for the fixation of nitrogen which is crucial for the survival of higher organisms, but is carried out by various symbiotic and free-living bacteria. Oxygen sensors that contain PAS domains are responsible for the regulation of nitrogen-fixation genes in these bacteria, and the fixing of nitrogen is obligately dependent on anaerobic conditions due to the involvement of an oxygen-labile nitrogenase (15). For example, FixL is a PAS domain containing protein that binds a heme cofactor that directly senses oxygen and controls the phosphorylation of a classic two component response regulator, FixJ, to transduce the signal. The heme cofactor is coordinated by a histidine residue on the F $\alpha$  helix, which is directed into the core of the PAS domain. Upon oxygen binding, the heme molecule becomes more planar, resulting in a conformational change in the FG-loop of the domain which is thought to allow an interaction with the histidine kinase domain, thereby inactivating it (25).

Even in biochemical pathways that utilize identical effector molecules, such as those that use heme for oxygen sensing, it is clear that PAS domains exhibit a wide variety of strategies and mechanisms for exerting their effect upon cells and cellular processes. In addition, another pathway uses PAS domains for responding to changing oxygen concentrations despite the lack of evidence for a required cofactor that directs its activity. Although it is not known to have any oxygen sensing activity per se, the Hypoxia Inducible Factor (HIF) is nevertheless a central player in the adaptation response

of cells to lower oxygen tension in humans and many other organisms. This pathway is the major focus of the work described in this research project, and will therefore be described in more detail later in this chapter and in Chapter III.

*b. PAS domains as light sensors*

PAS domains are used as protein modules in systems designed to sense light in pathways dedicated to circadian clock function or phototaxis. For example, the first structurally characterized PAS domain was the photoactive yellow protein (PYP), a bacterial blue-light photoreceptor that consists of an isolated PAS domain (26, 27). The protein acts as a receptor for negative phototaxis, and the cycle involves a trans-cis isomerization of a thioester-linked p-coumaric acid chromophore (28-30). One aspect of the structure of this protein is an N-terminal lariat-like structure including two  $\alpha$ -helices joined by a loop which bends into the space between the two helices. This area associates with the solvent exposed side of the  $\beta$ -sheet, a region which evidence provided in section *iii* suggests is used by many PAS domains for both intra- and intermolecular interactions.

Phototropins are a group of light-activated kinases that regulate activities in plants such as stomatal opening, phototropism, chloroplast relocation and leaf opening (31, 32). They consist of a serine/threonine kinase domain preceded by two PAS photosensors called light, oxygen, voltage (LOV) domains, LOV1 and LOV2. The *Avena sativa* (oat) phototropin LOV2 domain (AsLOV) binds the flavin chromophore FMN, which upon photoactivation forms a covalent cysteine adduct with Cys450 that remains stable for tens of seconds before relaxing back to the non-covalent dark state (33, 34). Mutation of this

highly conserved and reactive cysteine prevents both adduct formation and kinase activation (35). It was found that a C-terminal  $\alpha$ -helical extension (J $\alpha$ ) associates with the solvent exposed side of the central  $\beta$ -sheet, and under lit conditions, becomes unstructured and dissociates, thereby activating the kinase domain (36). The use of this interface for the regulation of PAS-containing proteins might be a prevalent mechanism used in PAS domain-mediated cellular activities, although more examples need to be studied to test the veracity of this assertion.

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

#### *a. PAS domains in the activation of kinases*

The sensing of environmental stimuli using bound cofactors such as those described so far are often vital functions carried out by PAS domains for the regulation of a variety of biochemical pathways, but the effects of these events are believed to result in the modulation of interactions between the sensory PAS domains themselves and other protein elements. An example of this is the FixL system discussed in section *ii.a.*, where a heme cofactor changes conformation upon oxygen binding which is thought to modulate interactions between the PAS and histidine kinase domains (25). This mode of kinase inactivation is also implicated in PAS kinase, a serine/threonine kinase where a similarly unstructured FG-loop is thought to change its conformation, resulting in the association and subsequent downregulation of the activity of its kinase domain (2).

These proposed mechanisms are centered on a PAS domain interacting *in cis* with a C-terminal kinase domain, which results in the inactivation of the enzymatic activity.

However, it should be noted that the kinase domains of histidine kinases such as FixL and serine/threonine kinases such as PAS kinase/phototropins share no structural homology (37, 38). This might imply that identical mechanisms of deactivation via *in cis* interactions with their structurally similar PAS domains would be unlikely. Similar mechanisms for these enzymes could be trivially explained, however, based on the dimerization of kinase domains that has been reported as a requirement for full activation in histidine kinases (reviewed in (39)). In this case, the PAS domains could serve as self-interacting dimerization elements of full-length proteins that help bring the kinase domains together for the reciprocal cross-phosphorylation events that have been implicated in the activation of these enzymes (39). Isolated PAS domains from PAS kinase and phototropin studied in our lab have not been shown to dimerize, however, and full length PAS kinase purified in the lab of Steve McKnight also seems to be active in the monomeric state (40). A variety of mechanisms could therefore be used in the PAS-mediated regulation of kinases. It is possible, for example, that phototropins differ significantly in their central mechanism of activation. One proposed model involves the interaction between the AsLOV2 domain and its C-terminal J $\alpha$ -helix, which could prevent the interaction with the kinase domain, keeping it in the inactive state. Upon illumination with blue light, however, this interaction between the PAS domain and the C-terminal helix is disrupted, which according to this particular model, allows the PAS domain to associate with and activate the kinase (36). Although evidence supporting any of these proposed models has not yet been obtained, it seems likely that protein-protein interactions modulated by PAS domains will emerge as important regulatory events in the

activation of these kinases, as well as other proteins involved in many other important pathways.

*b. Variable stoichiometry of complexes involving PAS domains*

PAS domains transduce environmental signals to other proteins or domains through protein-protein interactions that can vary widely in specificity and stoichiometry. For example, *in vitro* studies have shown that the isolated N-terminal PAS domain of HERG can form tetramers (41), and the PAS domain-containing RmFixL protein was shown to form a 2:1 complex with RmFixJ in solution (42). Studies of PAS domains from the Trachealess (Trh) and Single-minded (Sim) proteins from *Drosophila* also provide strong evidence that PAS domains are instrumental in providing specificity for the interactions between binding partners that form functional transcriptional activators. Chimeric proteins made by replacing of the Trh PAS domain with the analogous domain from Sim produced a protein with the functional specificity of Sim as shown by monitoring target gene activation in embryos (43). Most PAS domain-containing proteins form heterodimers with other PAS proteins, although PAS domains from RmFixL, BjFixL, PER, and EcDOS, a bacterial heme-containing oxygen sensor, have also been shown to homodimerize in structural studies (5, 6, 25, 44).

Despite the relatively well-characterized ability of PAS domains to interact with each other, the biophysical basis for these interactions are not well understood. The crystal structure of the PAS homodimer from the *E.coli* redox sensor protein EcDOS reveals a binding interface located on the solvent-exposed side of the central  $\beta$ -sheet (44, 45). Mutagenesis data from HERG also implicates this interface as evidenced by genetic

analysis of patients suffering from a cardiac disorder known as long QT syndrome (46). The crystal structure of an RmFixL homodimer shows that the  $\beta$ -sheet is involved in forming the dimer, but also that an N-terminal helix-loop-helix motif seems to provide the majority of the hydrophobic surface area for the interaction (6). Although this type of HLH-mediated interaction is an attractive prospect for a generalized model of PAS/PAS associations due to the additional buried surface area provided by the leucine zipper-like interaction of the N-terminal  $\alpha$ -helices, not all PAS domains involved in putative homo or heterodimers have predicted helical regions flanking the PAS core. In addition, different relative orientations between the binding partners have been observed for a homodimer crystal structure of the FixL PAS domain from *Bradyrhizobium japonicum*, so it remains an open question whether the interaction observed for RmFixL is physiologically relevant (discussed further in Chapter III). Prior to this work, no structures were available of any PAS domain complexes in solution, nor any PAS heterodimeric structures by any method. Therefore, one of the major goals of this project was to solve a heterodimeric PAS domain structure using solution NMR and modeling techniques to determine if any of the current models of PAS/PAS interaction were supported.

#### *B. PAS domains are often modular components of transcription factors*

Some common uses of PAS domains in biological pathways are as modular components of transcriptional activation proteins. There are many examples of these domains individually, or in tandem with other PAS regions, being coupled to DNA binding motifs for the purpose of regulating gene expression in response to exogenous



environmental signals. One example relates not only to photoreception, but also the control of circadian rhythms in *Neurospora crassa*. The White Collar proteins (WC-1 and WC-2) are transcriptional activators that upregulate the expression of a feedback repressor (frq) upon forming a higher order complex and binding to light response elements (LRE) via their zinc finger DNA binding domains (47, 48). One of the earliest efforts in this project was to characterize the PAS domains involved in this system, so a more thorough description will be presented in Chapter II.

*i. bHLH PAS transcription factors*

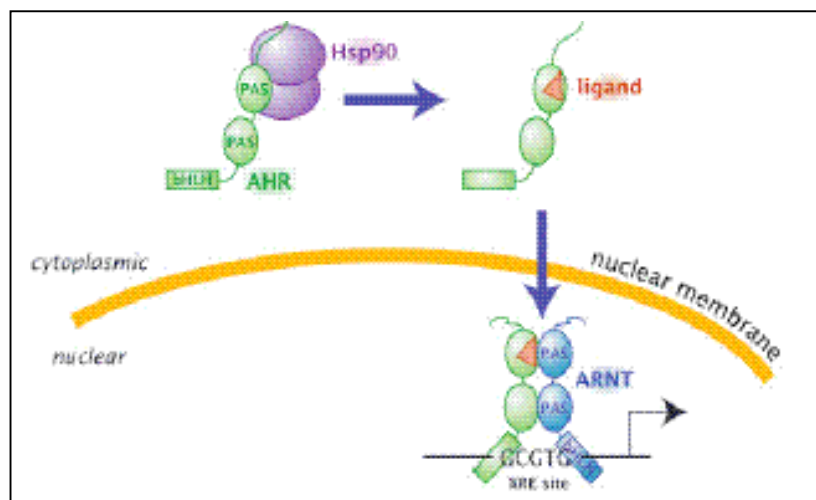
The most common uses of PAS domains in transcription factors, however, are as modular components of basic helix-loop-helix (bHLH) domain-containing proteins. bHLH/PAS proteins are critical regulatory components used in a number of important gene expression networks that are important for many developmental and physiological processes. For example, the regulation of neural development in many organisms is mediated by the bHLH/PAS protein SIM, which in *Drosophila* is proposed to act as a master regulator in midline cell development, and is required for nervous system midline cell lineage (14). This is an example of an even more exclusive category of bHLH proteins, those that form functional, heterodimeric complexes with another bHLH/PAS protein, the Aryl hydrocarbon Receptor Nuclear Translocator (ARNT). ARNT is involved in several different transcriptional activation pathways, and is the main focus of this research. As such, its role in two more important systems will be discussed in more detail later in this chapter and following chapters.

bHLH proteins must dimerize to form fully functional DNA-binding complexes, which bind DNA by virtue of basic regions N-terminal to the Helix-loop-helix dimerization domains (3). bHLH/PAS proteins are distinguished from other members of the larger bHLH superfamily by the inclusion of PAS domains C-terminal to the DNA binding regions. These domains confer specificity to the interaction with their dimerization partners, and are required for the assembly of functional complexes. For example, the isolated bHLH domain of the AhR is capable of forming homodimers alone, but is restricted to a heterodimeric configuration with ARNT when the PAS A domain is included (49). bHLH/PAS proteins tend to be ubiquitous, and their dimerization and transcriptional activity is often mediated by signals transduced by bound cofactors. They recognize variant forms of the classic E-box enhancer elements (ACNNGT) that are also recognized by the wider class of bHLH proteins, and domains C-terminal to the PAS domain region are often used as activation domains that mediate associations with other transcriptional activation cofactors (15, 50, 51). Studies such as these have shown that despite the high sequence conservation of the DNA binding basic regions and helix-loop-helix dimerization domains, bHLH/PAS proteins display high degrees of specificity with respect to the formation of active complexes. In addition, the PAS domain regions seem to be essential for this specificity, and contain highly specialized interaction-mediating information within their amino acid sequence despite the highly conserved nature of their tertiary structures. Structural information describing how these complexes form is lacking, however, and a greater understanding of the differences inherent to their formation could provide valuable information about the nature of biophysical specificity in the daunting myriad of complex biochemical pathways.

*a. The Aryl hydrocarbon receptor*

Exposure of vertebrates to the environmental pollutant dioxin (2,3,7,8-tetrachlorodibenzo-p-dioxin, TCDD) and other structurally related halogenated aromatic hydrocarbons results in a variety of biochemical responses including epidermal hyperplasia and metaplasia, severe wasting syndrome, thymic involution and tumourgenesis (52, 53). Responses to these toxins are mediated through the Aryl hydrocarbon Receptor (AhR), one of the most studied bHLH/PAS proteins. A schematic representation of this pathway is depicted in Fig 1-2.

**Fig 1-2. Molecular components of the Aryl hydrocarbon Receptor (AhR) pathway. AhR serves as a heterodimeric partner of ARNT to form a functional transcription factor. The details of this event are outlined in the text.**



Beyond the N-terminal bHLH domain are two tandem PAS domains (PAS-A and PAS-B), in addition to a C-terminal transactivation domain. In its latent state, AhR is

found in the cytoplasm associated with various molecular chaperones including heat shock protein 90 (Hsp90), which is proposed to keep the protein in a soluble state that is competent to bind ligand (reviewed in (3)). Upon ligand binding to the C-terminal PAS domain (PAS-B), the AhR/Hsp90 complex is translocated to the nucleus, where the chaperone is exchanged for ARNT to form a functional heterodimeric transcriptional activation complex (53, 54). The complex recognizes and binds to xenobiotic response elements (XREs) upstream of genes that respond to TCDD exposure such as cytochrome P4501A1, which metabolizes the toxin in an attempt to attenuate cellular damage. One gene that is targeted by this complex expresses a protein that has high sequence conservation to AhR, but lacks the C-terminal ligand-binding and transactivation domains. It is proposed that this protein, the aryl hydrocarbon receptor repressor (AhRR), serves as a negative feedback component that sequesters ARNT and binds to XREs as a non-functional complex to downregulate the expression of AhR-mediated genes products (55, 56). Dioxin is an industrial pollutant that has not existed long enough to affect evolutionary pathways, although natural products from the incomplete combustion of organic material can also serve as AhR-activating ligands. Although this could provide a potential evolutionary pressure to develop a pathway for toxin metabolism, the highly conserved nature of this system between species makes this unlikely (57).

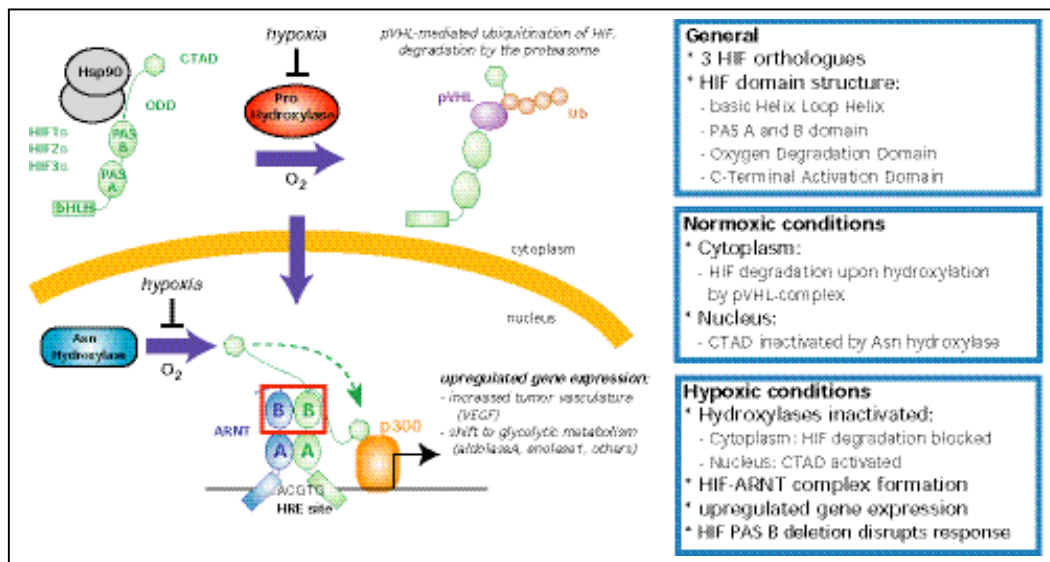
Further evidence based on knockout mice suggests that there might also be a role for AhR in development. Gene targeting has revealed that mice deficient in the AhR protein have defects in liver development as well as other problems (55, 56, 58). It has therefore been proposed that an endogenous ligand for this receptor exists, although such

a factor has not yet been identified. The ligand-binding characteristics of the PAS-B domain, coupled to the formation of a heterodimeric transactivation complex makes AhR and attractive system for studying the biophysical characteristics of bHLH/PAS mediated transcription. Unfortunately, the inherent instability of this protein without its natural chaperones make structural studies on this system extremely problematic. Nevertheless, the ability of PAS domains to bind a variety of ligands in this and other previously described systems offers the opportunity for domain-directed drug design which could hold therapeutic promise in this, and many other biologically relevant pathways.

#### *b. The Hypoxia Inducible Factor*

Stringent regulation of oxygen homeostasis in mammals is crucial for survival. Excessive O<sub>2</sub> tension in the cell (hyperoxia) can create reactive oxygen intermediates that can damage DNA and membranes, and low oxygen tension (hypoxia) can result in ATP production levels that are insufficient for proper cellular function (3). Responses to low oxygen concentrations in the cell are controlled by proteins called the hypoxia inducible factors, of which there are three known isoforms (HIF-1 $\alpha$ , HIF-2 $\alpha$  and HIF-3 $\alpha$ ). HIF-1 $\alpha$  and HIF-2 $\alpha$  show similar modes of regulation and share 48% sequence identity (59). Both of these proteins consist of the bHLH DNA binding domain, two tandem PAS domains (PAS-A and PAS-B), N-terminal and C-terminal transactivation domains (N-TAD and C-TAD), and oxygen degradation domain (ODD). Fig.1-3 shows a schematic representation of this pathway as it is understood so far.

**Fig 1-3. Molecular components of the hypoxia response pathway. Like AhR, HIF- $\alpha$ (1,2,3,) serves as a heterodimeric partner for ARNT to form a functional transcriptional activation complex. The details of this pathway are described in the text.**



With normal intracellular oxygen concentrations, HIF- $\alpha$  is selectively hydroxylated at two proline residues within the ODD (Pro402 and Pro564), one of which serves as the binding site for the Von Hippel-Lindau tumor suppressor (VHL) (60-63). In conjunction with other proteins, this becomes part of a functional E3 ubiquitin-protein ligase complex that marks the protein for degradation by the proteasome (64-66). Another mode of regulation exists in the nucleus during normoxic conditions, preventing transcriptional activation from any remaining low levels of HIF- $\alpha$ . Asparaginyl hydroxylase factor (FIH-1) hydroxylates an asparagine residue (N803 in HIF-1 $\alpha$  and N851 in HIF-2 $\alpha$ ) located within the C-TAD region, preventing the recruitment of the necessary co-activator, CBP/p300 (67, 68).

During periods of hypoxia, however, both hydroxylases are deactivated, stabilizing the protein, which is then imported into the nucleus where it binds its

heterodimeric partner ARNT (Aryl Hydrocarbon Receptor Nuclear Translocator) (3, 69). This forms a transcriptional activation complex which binds to specialized E-box-like regions in the genome called hypoxia response elements (HREs), resulting in the recruitment of CBP/p300 and upregulating a variety of genes responsible for helping the cell adapt to low oxygen concentrations (3). Some examples of these genes include those responsible for switching the cell to glycolytic metabolism, increasing vasculature to the region via the vascular endothelial growth factor (VEGF), and triggering apoptosis in the event that normal oxygen homeostasis cannot be restored (69, 70).

The function of the PAS domains in this system is largely unknown, except that both PAS domains have been shown to be required for the formation of a robust transcriptional activation complex (71). One of the main goals of this project was to understand how these domains are involved in this pathway, by revealing the structural and ligand-binding characteristics of the HIF and ARNT PAS domains both individually and in the context of their dimeric and/or multimeric associations.

### *c. The Aryl hydrocarbon Receptor Nuclear Translocator (ARNT)*

The Aryl Hydrocarbon Receptor Nuclear Translocator (ARNT) is a ubiquitous, constitutively nuclear localized protein that acts as a general binding partner for a variety of bHLH/PAS proteins, such as SIM, AhR, and HIF- $\alpha$  (reviewed in (3)). It is comprised of an N-terminal nuclear localization sequence (residues 39-61), which controls nuclear localization through interactions with components of the nuclear pore targeting complex, a bHLH domain and two PAS domains (PAS-A and PAS-B). It also contains a strong C-terminal activation domain that has also been shown to interact with CBP/p300 in a yeast

2-hybrid screen (72, 73). Knockout studies in mice and embryonic stem cells have demonstrated how important this protein is for both an effective response to hypoxic conditions and development in mammals. Embryonic stem cells with both copies of endogenous ARNT knocked out (ARNT<sup>-/-</sup>) show no transcriptional activation of genes responsible for the hypoxia response (74, 75). In addition, mouse embryos with similar knockouts do not develop past day 10.5, primarily due to vascularization defects, mimicking the phenotype for VEGF knockouts (75-77). It is believed that developing embryonic organs create local regions of hypoxic conditions, necessitating the induction of the hypoxia response in order to recruit new vasculature to the area (3). This is analogous to the metabolic stress of immortalized cells during solid tumor formation, during which the HIF transcription factor is activated for similar purposes (78).

The role for ARNT as a heterodimeric partner for the transcriptional activators discussed so far has been well characterized. ARNT is also known to form homodimers, although any physiological purpose for this interaction remains a mystery. ARNT is a constitutively present, nuclear localized protein that is widely expressed in many mammalian tissues (79). ARNT does contain a C-terminal transactivation domain, and electrophoretic mobility shift assays have demonstrated that baculovirus-expressed ARNT binds the CACGTG consensus sequence (80, 81). Other studies have shown that ARNT can also activate transcription from promoters driven from this E-box element in transient transfections (82, 83). The question remains, however, if ARNT has any role in regulating gene activation on its own under natural conditions in the cell. It might be that ARNT homodimerizes in the nucleus in order to stably sequester itself until its heterodimeric partner appears in the nucleus for the purpose of transcriptional activation.



Such a mechanism would imply that the homodimer, and perhaps the individual domain components that mediate the interaction, would dimerize with a weaker association constant to allow competitive displacement by the heterodimeric partner. The formation of homodimer in this case might also impart a degree of regulation in itself, by tuning down the system until high enough concentrations of stabilized sensor proteins such as HIF or AhR are present in the nucleus. Studies of ARNT and other PAS domain related systems should therefore benefit by structural and quantitative studies that could reveal not only the details of how these domains interact, but also how the relative strengths of these interactions could be related to physiological function. It was therefore a major focus of this work to characterize the structural details and *in vitro* interactions of PAS domains from this protein with itself and similar domain components from HIF- $\alpha$ , in an effort to understand potential structural bases for regulation in bHLH/PAS transcriptional activation systems.

### *C. Ligand binding in PAS domains*

PAS domains often bind ligands, covalently or non-covalently, to transduce signals to downstream effectors. They can bind heme non-covalently in the case of FixL, cyclic halogenated toxins in AhR, and various types of cyclic chromophores for light sensing in PYP, White Collar-1, and phototropin (reviewed in (15)). Although sequence homology between most PAS domains for which there are structures available is poor, they appear to be very similar with regard to tertiary fold. The wide variety in types of ligands and modes of binding exhibited by the PAS domains implies that this type of fold must be quite flexible with respect to the types of small molecules that are potentially

capable of binding. PAS domains are involved in biological pathways that are implicated in a variety of disease states that involve misregulation of important biochemical pathways. For example, the hypoxia pathway is involved in several examples of disease progression, such as enabling the growth of large solid tumors as described earlier. Since the roles of these modules as specificity and affinity enhancers in multi-domain proteins is becoming clearer, it seems reasonable that they could potentially be important targets for optimized small molecule inhibitors and/or enhancers for the treatment of diseases such as cancer. Toward the investigation of this possibility, another aspect of the work described therein focuses on developing a rational NMR-based screening approach for the identification of optimized lead compounds that could bind and subsequently mediate the function of PAS domains involved in the hypoxia response.

## II. Domain optimization for NMR studies – White Collar

(Parts of this chapter adapted from Card et al, 2005 (84))

### *A. The Neurospora circadian clock*

Circadian rhythms are important for the regulation of a wide variety of an organism's activities, both at the molecular level and with regard to behavioral responses. The entrainment of these molecular clocks by external stimuli such as light and temperature allow organisms to adapt to environmental changes and synchronize daily activities via the rhythmic oscillation of gene expression (reviewed in (85)). Many studies have been pursued in order to discover the genetic and molecular bases for circadian clock assembly and entrainment in several model systems (86, 87). Among these, studies of the circadian clock in the mold *Neurospora crassa* have advanced our understanding of how the cyclic expression of genetic components in this system function to maintain the endogenous clock, and how light serves as an entrainment stimuli to permit a robust and adaptive circadian system (88-91).

*Neurospora crassa* is a good model system for circadian rhythm investigations due to many factors including its ease of culture and genetic tractability. The production of asexual spores (conidiation) was previously shown to be controlled by circadian rhythms (92), facilitating their measurement by observing the timing between successive waves of conidiation as it grows laterally in partially agar-filled "race tubes". The tubes are first inoculated with *Neurospora* at one end, and then light or temperature is applied to reset the circadian clock (85). Developmental process which are responsible for conidia production occurs in the late subjective night and a characteristic band appears

within a certain period each day, providing an accessible metric for the periodicity and quality of the circadian clock in this organism. The relative ease of genetic manipulation allowed the first *Neurospora* clock mutant strains to be identified as early as 1971 (93), and resulted in pioneering work on the molecular function of specific clock components (94). Cloning the clock gene *frq* (95), and completion of the first global screens for clock-controlled genes (ccgs) for any organism was accomplished on *Neurospora* (96). The use of the race tube assay and other analytical tools have helped identify more than a dozen genetic loci that affect clock function (88). Although the products of many of these genes remain unknown, three of the major components of the *Neurospora* clock have been characterized (*frq*, white collar-1 (*wc-1*), and white collar-2 (*wc-2*)), which has led to the development of a model for the *Neurospora* circadian oscillator. Following the example of circadian systems in other organisms, this model involves interacting autoregulatory feedback elements whose expression and/or stability are modified in a regular manner to generate a robust and entrainable circadian clock. Particularly notable for the subject of this thesis, the WC-1 and WC-2 proteins contain PAS domains, and both the ligand binding and protein-protein interaction capabilities of these domains play important roles in the circadian clock of *Neurospora* and other organisms.

A model that depicts part of the current understanding of feedback loops involved in the *Neurospora* circadian clock is shown in Fig 2-1.

WC-1 and WC-2 are positive elements of this pathway, forming a heterodimeric complex in constant darkness that binds to two light response elements (LREs) in the *frq* promoter, upregulating the expression of the FRQ protein (91, 97, 98). This is supported by the greatly reduced levels of *frq* mRNA and FRQ protein found in *wc-1* or *wc-2* mutant strains which result in a defective clock under normal conditions (90, 99-101). Unlike WC-2, which is constitutively present and mediates a WC-1/FRQ interaction, WC-1 is the limiting element in the formation of functional heterodimer and is typically expressed out of phase with FRQ. Following transcription, two variant forms of FRQ are expressed which form homodimeric complexes (98, 102, 103). When FRQ levels reach a certain point, it interacts with the *frq* activating WC-1/WC-2 complex (WCC) to inhibit

its own transcription (94, 104-108). Soon afterwards, WC-1 levels decline, reducing the amount of viable transcriptional activation complex. This event, coupled to the repressor activity of FRQ with regard to its own transcription, is thought to provide a temporal and physical relationship that maintains a functional clock in *Neurospora* (98, 105).

Phosphorylation events also destabilize FRQ and seem to be important for regulating the interaction of FRQ with the WCC which serves to close the negative feedback loop (107, 109). The progressive destabilization of FRQ results in relief of the inhibition of the WCC complex, which causes reactivation of *frq* transcription and initiates a new cycle to maintain robust circadian oscillation.

It has been demonstrated that blue light rapidly induces the transcription of *frq*, and that a light pulse applied in *Neurospora*'s subjective night will result in a phase shift (advance) in the circadian clock (89, 110). Both WC-1 and WC-2 are essential for all light responses in this organism, including the light-mediated induction of *frq*, and all of these responses are eliminated in null mutant strains of either protein (110, 111). These proteins bind to the promoter elements of light-inducible genes via Zn-finger DNA-binding domains, and contain PAS domains which have been shown to be important in the formation of the WC-1/WC-2 heterodimeric complex *in vivo* (97, 101, 108).

Interestingly, it was also found that two different WC complexes exist that each bind to LRE's under different circumstances. A smaller WC complex that is consistent in size with a WC-1/WC-2 heterodimer, binds to LRE's in the dark, and this association is reduced after light exposure *in vivo*. A larger WC complex then replaces the smaller complex after light induction. Therefore, the smaller complex seems to be responsible

for *frq* activation in the dark, and light-induced activation is apparently modulated by the larger White Collar complex (91).

WC-1 contains three tandem PAS domains, each of which plays an important role in the circadian clock. The N-terminal WC-1 PAS domain (PAS-A) belongs to a special class of PAS modules known as LOV (light, oxygen, voltage) domains, sharing sequence similarity with a FMN-binding LOV domain present in the phototropins, the plant blue-light receptors responsible for the phototropic response in plants (112, 113). Deletion mutants of WC-1 revealed that the PAS-A domain was necessary for all light responses in *Neurospora*, including the light induction of *frq* (90). WC-1 isolated from *N. crassa* binds the chromophore FAD (90), which is necessary for the replacement of the smaller WC complex responsible for *frq* activation in the dark with the larger complex that mediates light-induced *frq* expression (85). This leads to many intriguing questions as to how the light-sensing capabilities of PAS domains are coupled to the formation of higher order complexes that have variable functions and oligomerization states. To address these issues, detailed biophysical studies on constructs of isolated and/or tandem PAS domains from WC-1 and WC-2 will have to be done to learn how the putative light induced conformational changes of the LOV domain in this system modulates the association of WC-1 with WC-2 and possibly itself in the formation of the small and large complexes involved in the regulation of the *Neurospora* circadian clock.

In addition to the N-terminal LOV domain (PAS-A), WC-1 contains two other PAS domains (PAS-B and PAS-C) that play a significant role in the *Neurospora* circadian oscillator (97, 101, 108). The putative function of these domains is to mediate protein-protein interactions, but their specific involvement in oligomer formation has yet

to be fully characterized. The PAS-B domain is essential for WC function in the circadian clock and for light responses, but not required for the WC-1/WC-2 interaction (108). PAS-C, in addition to a region C-terminal to this domain, was found to be necessary for complex formation (101, 108). WC-2, in addition to its Zn-finger DNA-binding domain, contains only a single PAS domain (PAS-A), which like the PAS-C domain of WC-1, is necessary for complex formation (97). Detailed biophysical characterization of the particular contributions of these domains to protein-protein interactions in higher order complexes involved in the regulation of the *Neurospora* clock will certainly lead to a greater understanding of how various components in this and other systems cooperate within biochemical pathways to regulate a large number of biological responses.

The importance of PAS domains in this and other circadian clock systems for photosensing and downstream signal transduction mediated through protein-protein interactions has led us to pursue the characterization of PAS domains and their interactions in this and other systems using NMR-based strategies. However, investigations of this kind require relatively large amounts of well-folded, stable protein constructs. The studies on PAS domains described in the rest of this document required significant optimization of solution conditions and construct design in order to ensure success. Methods we used to address these issues were therefore an important aspect of this work, and before further details regarding the characterization of PAS domains from the hypoxia response pathway are presented, a brief overview of some of these techniques will be discussed in the context of initial studies on PAS domains from White Collar-1 and White Collar-2.



### *B. Domain optimization for NMR studies*

The success of genomic sequencing projects in recent years has presented protein scientists with a formidable challenge in characterizing the vast number of gene products that have subsequently been identified. NMR has proven to be a valuable tool in the elucidation of various properties for many of these proteins, allowing versatile studies of structure, dynamics and interactions in the solution state. But the characteristics needed for proteins to be amenable to this kind of study, such as folding capability, long-term stability and high solubility, require robust and expeditious methods for the identification and optimization of target protein domains. Here we present a variety of computational and experimental methods developed for these purposes, and show that great care must often be taken in the design of constructs intended for NMR-based investigations.

Although there have recently been major advances in the sequencing and annotation of genomic information from a wide variety of organisms, a large percentage of identified genes have no known function (e.g. > 40% of human genes initially sequenced were of unknown function (*114*)). Structural biology has played an important role in addressing this shortcoming, as structural homology between proteins can provide valuable clues toward the elucidation of unknown protein function (*115*). Whether practiced on individual systems or on a proteomic scale, a common challenge to structural biologists is the need to rapidly identify, express and purify proteins that are amenable to NMR spectroscopy, X-ray crystallography and other biophysical techniques.

Solution NMR has played an integral role in these studies for a variety of reasons. These include the abilities to investigate protein structure in a solution state more closely resembling physiological conditions, monitor binding processes and interfaces, extract

dynamic properties from relaxation and exchange parameters, and provide important avenues to study crystallization-resistant targets. However, the inherently low sensitivity of NMR requires that proteins be soluble to high concentrations and relatively stable over long periods of time. In addition, although many advances have been made for NMR studies of larger proteins and complexes using deuterium labeling (116), TROSY-based pulse sequences (117) and higher magnetic field strengths, NMR methods have been most routinely applied to proteins under approximately 20 kDa to avoid problems associated with spectral overlap and broad linewidths.

The method most often used to overcome this high molecular weight problem is the “divide-and-conquer” approach where constituent domains of the full-length protein are individually cloned, expressed, and purified. Although this strategy helps keep resonance signals sharp by increasing the tumbling rate and greatly simplifies the interpretation of NMR spectra by reducing the number of correlation peaks, it requires the identification of well-folded, soluble fragments that retain both their native structure and function when removed from the context of full-length protein. Here we review a variety of computational and experimental approaches that can be used to identify well-behaved protein domains for NMR studies.

#### *i. Computational methods*

Computational algorithms have proven invaluable to molecular biologists by helping to identify open reading frames, homologous proteins, and in the classification of gene families. In a similar manner, they are often used for the prediction and annotation of protein and domain families based on statistical models derived from multiple

sequence alignments that are generated from the constantly increasing storehouse of genomic data. Identification of these relationships is particularly important for structural biologists because many structurally similar members of a particular family of proteins can exhibit relatively low degrees of sequence homology. For example, PAS (Per-ARNT-Sim) protein-protein interaction domains have a high degree of structural conservation despite notoriously low similarities between their primary sequences (15, 118). For this reason, robust algorithms designed for the identification and annotation of protein domains and secondary structural elements are often used to locate independently folded, stable protein constructs that can be cloned and expressed for structural studies.

## *ii. Domain prediction*

An important development in computational tools designed for these purposes was the incorporation of hidden Markov models (HMMs) (119). Statistical methods for modeling, database searching, and generating multiple alignments are based on the assumption that even when structurally similar protein domains have widely divergent sequences, there should be conserved elements within the primary sequence that define the folding, structure and function of the target. This information is encoded into an HMM with a training process involving representative members of a domain family, generating a matrix that describes the likelihood of finding any given amino acid (or insertions or deletions) at each location throughout the domain. These matrices can be used to generate a probability distribution for all possible sequences, allowing one to score the probability that a given query sequence is a member of the family described by the HMM. As such, these are powerful tools to search sequence databases for previously

unidentified representatives of HMM-encoded domains, and can be used to define common aspects of various members, including domain boundaries. One particularly successful implementation of HMM-based methodologies to define protein members, sequence alignments and domain boundaries is Pfam (<http://pfam.wustl.edu>) (120).

Unlike many contemporary techniques, Pfam used the entire span of the domain to search sequence databases for new members. Previous strategies typically included only well-conserved motifs, but HMM-profile methods allow less conserved regions, as well as insertions and deletions, to be dealt with and annotated in a more accurate manner.

Another important HMM-based tool, particularly for researchers interested in the study of domains used in signaling processes, is SMART (Simple Modular Architecture Research Tool, <http://smart.embl-heidelberg.de>) (121). Currently, SMART is based on multiple sequence alignments of 667 domain families that have been updated to include additional homologues predicted by computational methods and verified by their experimentally-determined biological context. Taken together, these sequences were then used to make HMMs for the identification and annotation of novel or uncharacterized domains. Although Pfam demonstrated an increased ability to find incomplete domain sequences relative to SMART, the latter was designed to identify full-length domains, which is clearly important for structural biologists intending to study fully functional constructs. Another advantage of SMART is the inclusion of integral HMM-based tools to identify compositionally biased regions such as transmembrane segments, which can aid the rational design of protein fragments that are less likely to aggregate under the high concentrations necessary for NMR studies.

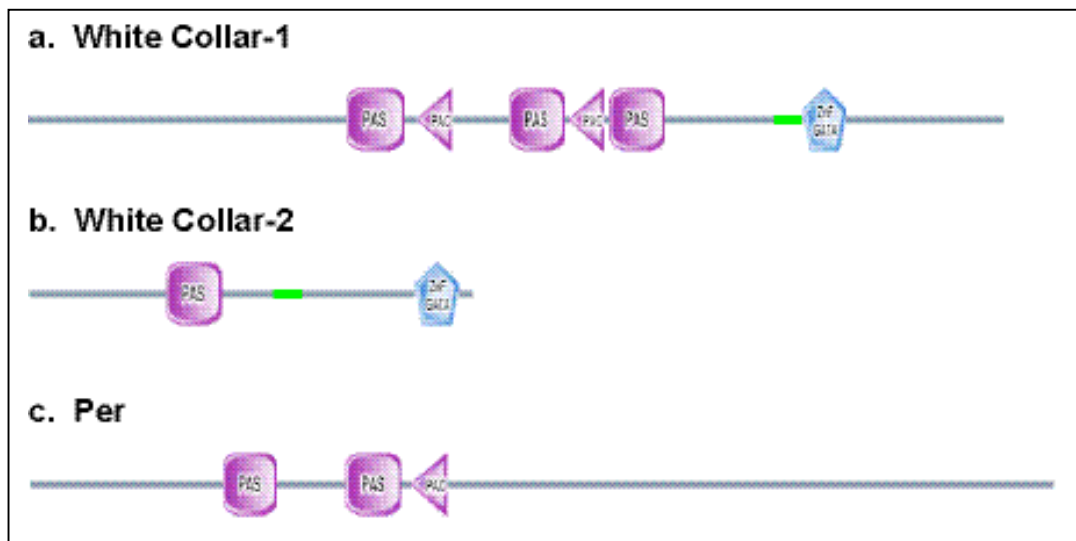
*a. White Collar-1 and White Collar-2 domain prediction*

The first example of a PAS domain containing protein that was extensively characterized was the product of the *Drosophila* period (*per*) gene, a protein involved in regulating the circadian rhythms of locomotor activity (10, 11, 13), so the involvement of PAS domains in circadian systems has been well established. Identification of the existence and locations of specific PAS modules was an important first step in the experimental design strategy for NMR-based studies on the PAS domains involved in the *Neurospora* circadian oscillator, so the domain architectures of the white collar proteins (WC-1 and WC-2) predicted by the HMM-based SMART domain prediction algorithm are shown in Fig 2-2 a,b. This method accurately predicts the three PAS domains in WC-1 and the single PAS module in WC-2, in addition to the Zn-finger DNA binding domains that support the direct transcriptional activation activity of these proteins. For comparison, the domain prediction for *Per* is also shown in Fig 2-2 c, revealing a conspicuous lack of DNA-binding elements. It should be noted that HMM-based methods used to predict PAS domains originally used an archaic definition, which assumed a size of only about 60 residues for these protein modules. As such, many descriptions of PAS domains in various databases still show an artificially short “PAS” domain followed by a “PAC” module (Fig 2-2). Work in our lab and elsewhere has helped redefine the definition of the canonical PAS unit to include both of these elements in a single protein domain which typically contains a minimum of approximately 100 residues.

The present model of *Per* activity suggests that it acts as a dominant negative inhibitor of its own transcription, perhaps via PAS-mediated interactions with the bHLH-

PAS transcriptional activators SIM and ARNT (5). This serves as a good example of the ubiquitous nature of these domains, perhaps revealing how similar domains can be hijacked over time by new regulatory elements during the gradual evolution of analogous systems in more advanced organisms. If so, characterization of the specific interactions between the domains used between species in similar systems might yield important insights with regard to the evolution of regulatory mechanisms. Once the prediction of individual domains has been established, the next step in the NMR-based investigation of the *Neurospora* circadian clock was to identify accurate domain boundaries for the cloning and expression of well-behaved domains.

**Fig 2-2. Results of the HMM-based SMART domain prediction algorithm. Shown are the results for a. WC-1, b. WC-2, and c. Per protein from *D. melanogaster*.**



### iii. Secondary structure prediction

There have been many techniques developed that utilize different heuristics for the prediction of protein secondary structure. For example, the PHD method (17) uses a multiple sequence alignment of the query sequence as input for a neural network initially

trained on a non-redundant set of 130 protein chains, integrating evolutionary information into the prediction process. Alternatively, the PREDATOR algorithm (122) is based on the recognition of potential hydrogen bonding between residues in a query amino acid sequence, using database-derived statistics on residue type occurrences in  $\alpha$ -helical and  $\beta$ -strand secondary structural elements. Studies have shown that the shortcomings of these and other computational methods can be addressed by combining multiple methods to improve the accuracy of *de novo* predictions (123). This approach is well-implemented by the JPred secondary structure prediction server

(<http://www.compbio.dundee.ac.uk/~www-jpred>) (124). JPred further improves its accuracy by generating secondary structure predictions not only for the initial query sequence but also for a series of non-redundant homologues found with BLAST searches. In addition, a consensus is derived with scoring functions for each residue to give a weighted average prediction among the different methods and sequences that provides a more accurate identification of secondary structure and domain boundaries.

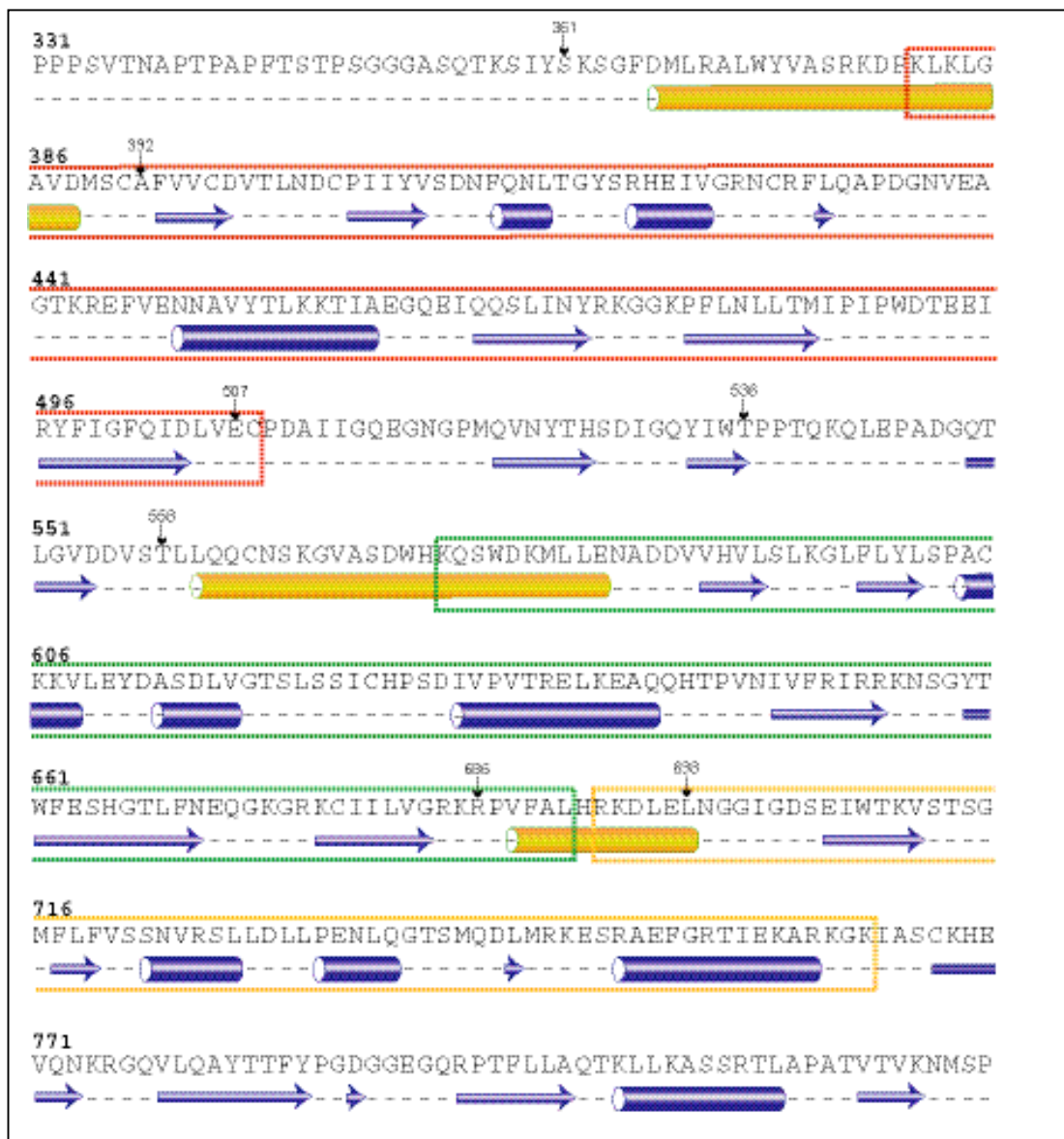
#### *a. White Collar secondary structure prediction*

The JPred secondary structure prediction algorithm was used to identify specific domain boundaries for the cloning of PAS domains in WC-1 and WC-2. It should be noted that although functional data has emerged that implicates the PAS-C domain of WC-1 as mediating interactions between the heterodimeric partners of the transcriptional activation complex (101, 108), at the time of this investigation the presence of this third PAS domain was not firmly established and so was not included in the initial characterization of WC-1. Therefore, the studies described herein will only discuss

initial progress for the optimization of WC-1 PAS-A, WC-1 PAS-B, and WC-2 PAS-A. Results of the JPred method as applied to the WC-1 protein is shown in Fig 2-3. For the sake of simplicity, the homologues used in the sequence alignments for secondary structure prediction are not shown. The consensus for the secondary structure prediction from the analysis is represented by arrows ( $\beta$ -strand) and tubes ( $\alpha$ -helix).



**Fig 2-3. Results of the JPred secondary structure prediction of the White Collar-1 PAS domain containing region.** Secondary structural elements are shown as blue arrows ( $\beta$ -strands) and tubes ( $\alpha$ -helix), with orange helices assumed to be external helical elements relative to the standard PAS fold. Boxed regions correspond to the domain boundaries predicted by the prediction algorithm SMART (red – PAS/PAC-A, green – PAS/PAC-B, orange – PAS-C). Residue numbers that correspond to domain boundaries used for cloning of PAS domain constructs in the following sections are also shown by arrows above the sequence.

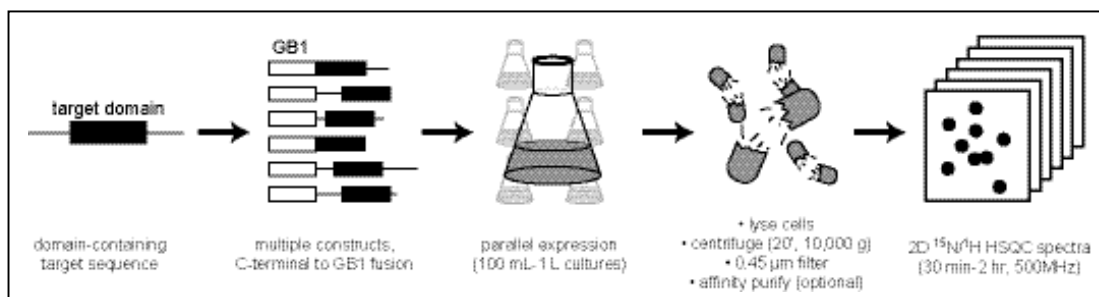


#### iv. Parallel cloning and expression

While computational methods have often demonstrated their ability to identify the locations of domains and secondary structure elements within protein sequences, this still leaves a range of challenges for an NMR spectroscopist intending to translate these predictions into large quantities of soluble domains for study. These complications stem from several sources, including poor sequence conservation and incomplete HMM-based domain models, as well as the many difficulties with predicting complex solution behavior directly from sequence.

To address these problems, methods have been developed to experimentally identify protein domains from within larger sequences. One of the most common approaches is based on limited proteolysis (125), where the domain-containing sequence and flanking regions are treated with small quantities of proteases. Protease-resistant fragments, typically identified and mapped within the protein sequence by mass spectrometry, are interpreted to be well-folded domains. While this approach has often been used successfully, it is potentially susceptible to artifacts from proteolysis of solvent-exposed loops and other extended segments located within domains of interest.

**Fig 2-4. Overview of parallel expression approach.**



As an alternative strategy, we have relied on an NMR-based screen of related protein constructs with different N- and C-termini flanking a domain of interest identified by computational approaches (Fig 2-4). Typically, 6-8 constructs are designed with unique sets of termini based upon comparisons between orthologous proteins and secondary structure predictions. In parallel, these fragments are cloned as fusions to the C-terminal side of small, highly-soluble auxiliary proteins, expressed in *E. coli* cultures and used for 2D  $^{15}\text{N}/^1\text{H}$  HSQC experiments as detailed below. Obtaining such spectra at an early stage provides qualitative assessments of protein folding from amide proton chemical shift dispersion. In addition, these data provide early indicators of spectral quality that might reveal issues with peak doubling or heterogenous linewidths that might complicate experiments relying on amide proton detection.

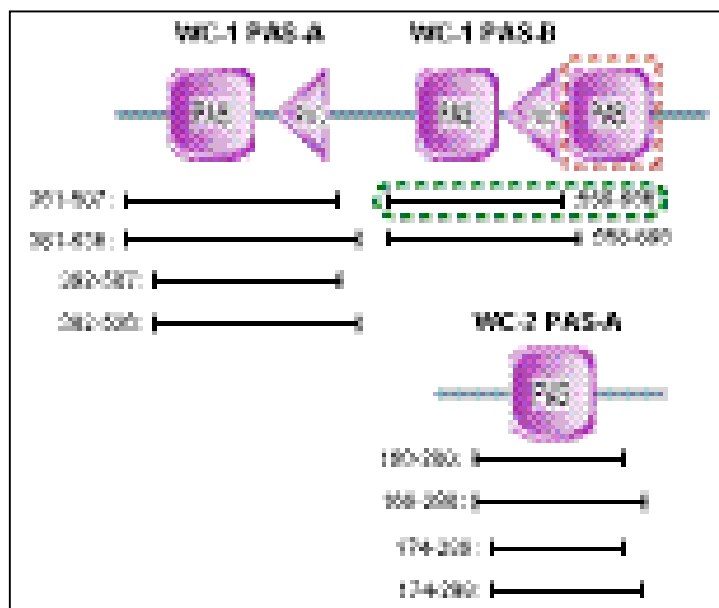
Fusion proteins can aid structural studies of native protein domains by increasing expression levels, improving solubility, and providing opportunities to use affinity purification methods. While these goals are accomplished to varying degrees depending on both the choice of fusion and the target (126), there are several options widely available including maltose binding protein, glutathione S-transferase and a variety of multi-histidine tags.

We have routinely used a point mutant of the Streptococcus protein G beta 1 domain (mutant T2Q, here referred to as G $\beta$ 1) because it meets several of these goals, and provides additional benefits for cell lysate screening (Section 3.3). Foremost, G $\beta$ 1 fusions typically express to very high levels and are quite soluble (126-128). G $\beta$ 1-containing proteins can be affinity purified on IgG resins, although the relatively harsh conditions needed to elute such proteins (pH < 4.0 or chemical denaturant) have led us to

generate His<sub>6</sub>-tagged versions of Gβ1 for Ni(II) affinity purification (36). Gβ1 is also very small (56 amino acid residues) and separated from the target protein by a flexible linker of approximately 12 residues in our expression vectors (2, 36). As such, Gβ1 produces clear, well-resolved peaks in <sup>15</sup>N/<sup>1</sup>H HSQC spectra, conveniently providing a positive control for establishing that fusions are present at high concentration and not grossly aggregated.

*a. Parallel cloning and expression of WC-1 and WC-2 PAS domains*

Constructs were designed based on JPred results (shown in Fig 2-3 for WC-1), and fragments were cloned encompassing the predicted PAS domain region with variable-length extensions at the N- and C-terminus to screen for both stability and folded states. Fig 2-5 shows a few of the constructs that were cloned and screened using the procedure described in Fig 2-4.



**Fig 2-5. Diagram outlining the constructs used for initial cloning and screening for folded PAS domains within the *Neurospora crassa* circadian clock system. Cloned constructs for WC-1 PAS-A, WC-2 PAS-A and WC-1 PAS-B are shown below the results from the SMART-derived domain prediction algorithm for the PAS domain regions. The WC-1 PAS-C domain (red box) was not screened due to the low confidence of the prediction at that time.**

The choice of constructs, primer design and some follow up work was done by Lori Neil. Standard PCR reactions were used to generate fragments which were then inserted into the G $\beta$ 1-parallel cloning vector using standard restriction sites present in the polylinker region. All constructs were designed with an N-terminal G $\beta$ 1 fusion tag, for solubility and to encourage a folded state as described in section D. All fusion proteins contained a TEV protease cleavage site (Glu-Asn-Leu-Tyr-Phe-Gln-↓-Gly) present in the open reading frame of the vector for removal of the fusion tag after soluble well-folded domains were identified, expressed and purified.

As will be discussed in further detail in following chapters, PAS domains sometimes require additional  $\alpha$ -helical segments either before or after the predicted domain for a variety of functions involving intra- or intermolecular associations (5, 6, 36, 44, 45). Because of this, several constructs were designed that included additional residues N- or C-terminal to the PAS domain boundaries, due to the prediction of  $\alpha$ -helical elements by JPred in these regions (see Fig 2-3 for construct ends in the context of additional secondary structural elements).

#### *v. Whole cell lysate screening*

To confirm the folded state of proteins up to approximately 20 kDa in molecular weight, we have found it useful to obtain 2D  $^{15}\text{N}/^1\text{H}$  HSQC spectra directly on *E. coli* cell lysates from cultures over-expressing the proteins of interest (129). For this, we use 100mL-1L cultures of *E. coli*, grown in uniformly  $^{15}\text{N}$  labeled media, containing plasmids overexpressing the target fusion protein under the control of a T7-based promoter. After overnight expression at 20°C, cell pellets from these growths are lysed, clarified with

high speed centrifugation (20 min, 10,000 g) and filtered through a 0.45 $\mu$ m filter.

Lysates are then concentrated to approximately 1mL, or can alternatively be subjected to a rapid affinity-based purification step (130). With or without initial purification, this general approach can be straightforwardly used to manually screen 6-8 samples, or potentially larger numbers with the aid of automation.

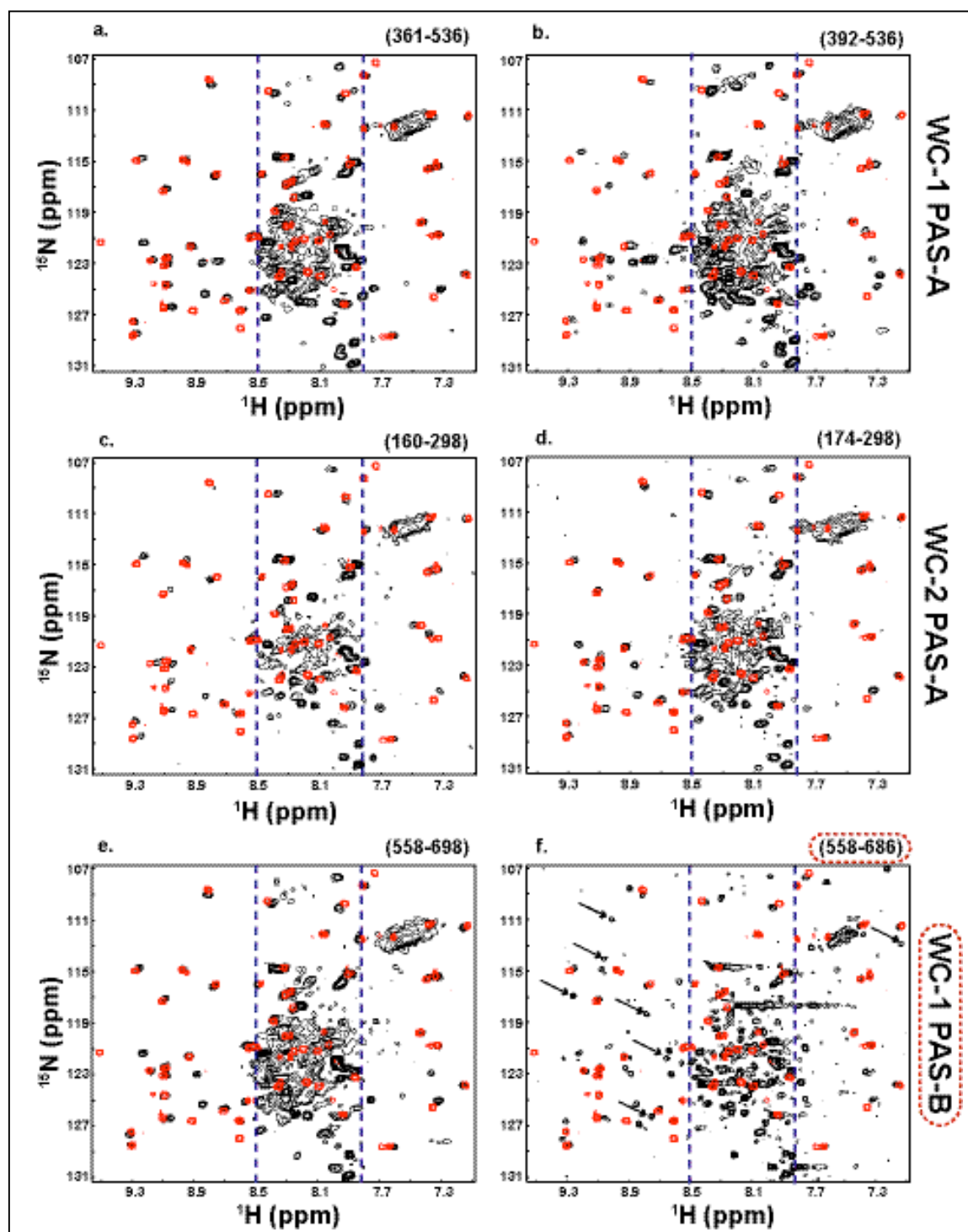
Lysates prepared in this manner typically generate 2D  $^{15}\text{N}/^1\text{H}$  HSQC spectra with reasonable signal-to-noise in a fairly short amount of time (30 minutes to 2 hours) on commonly available 500MHz spectrometers. It has been shown that spectra recorded from *E. coli* lysates or intact cells overexpressing small proteins are dominated by signals originating from the protein of interest (129, 131). This stems from the ability of common bacterial expression systems to often produce target proteins as the single most abundant protein in *E. coli*. In contrast, many host proteins are either present in low abundance or participate in large complexes, hampering their detection by NMR. Signals from a number of small nitrogen-containing metabolites, such as cyclic enterobacterial common antigen (8, 132) can often be seen in these spectra, although the small number of these peaks typically do not interfere with analysis.

#### *a. Whole cell lysate screening of WC-1 and WC-2 PAS domains*

To demonstrate this approach, Fig 2-6 shows a comparison of  $^{15}\text{N}/^1\text{H}$  HSQC spectra of lysates from cells expressing G $\beta$ 1 fusions of several constructs for the N-terminal PAS domains of WC-1, WC-2 and the middle PAS domain of WC-1 (WC-1 PAS-A, WC-2 PAS-A and WC-1 PAS-B). Spectra from each of the fusion proteins are superimposed with a spectrum of free G $\beta$ 1 recorded under comparable conditions

(pH=7.0, 25°C). Although some Gβ1-derived crosspeaks show chemical shift changes due to the presence of the fusion partner, it was straightforward to distinguish peaks originating from Gβ1 and the PAS fragments. Examination of spectra obtained from Gβ1 fused to many constructs of WC-1 PAS-A, WC-2 PAS-A, and WC-1 PAS-B (Fig 2-6 a-e) reveals that signals originating from the PAS domain in these cases have poor amide proton chemical shift dispersion ( $7.8\text{ppm} < \delta(^1\text{H}) < 8.5\text{ppm}$ ) and heterogeneous linewidths. Poor chemical shift dispersion is consistent with a fragment not adopting a well-ordered tertiary structure, because of the relatively homogenous chemical environment sampled by backbone amide protons that exist in a random coil state. Heterogeneous linewidths can indicate a disordered state due to the wide range of transverse relaxation rates that can exist in this type of unfolded state.

**Fig 2-6. Utility of *E.coli* cell lysate screening for determining domain boundaries.** Fragments of WC-1 and WC-2 were expressed in *E.coli* BL21(DE3) as C-terminal fusions to Gβ1 using a T7 RNA polymerase expression vector (2). Each spectrum was generated from a lysate generated by a 1L culture grown in uniformly  $^{15}\text{N}$ -labeled M9 media, which was concentrated to 1mL before acquiring 2D  $^{15}\text{N}/^1\text{H}$  HSQC spectra (40 min, 37°C). Construct boundaries from WC-1 PAS-A (a,b), WC-2 PAS-A (c,d) and WC-1 PAS-B (e,f) are shown above each spectra. Spectra from Gβ1-White Collar PAS fusions are shown with black contours; those from isolated Gβ1 samples are drawn with red contours. One construct (WC-1 PAS-B (558-682) – red dotted ovals) showed good dispersion in the proton dimension, indicating a folded protein.





However, when the WC-1 PAS-B construct was shortened from the C-terminal side from residue 698 to residue 686 (Fig 2-6 f), many new peaks appear in more disperse regions of the spectrum (arrows), and the peak density in the unfolded region is greatly diminished. This result could indicate that many backbone amide resonances not associated with the fusion tag are now experiencing a more heterogeneous chemical environment and therefore are in a folded state. A more likely explanation, however, is that the extra residues in the larger construct were unfolded and causing aggregation that prevented the observation of signals from the folded domain. Regardless of the reason, the smaller construct (558-686) was chosen for large-scale expression, purification and further optimization for structural studies.

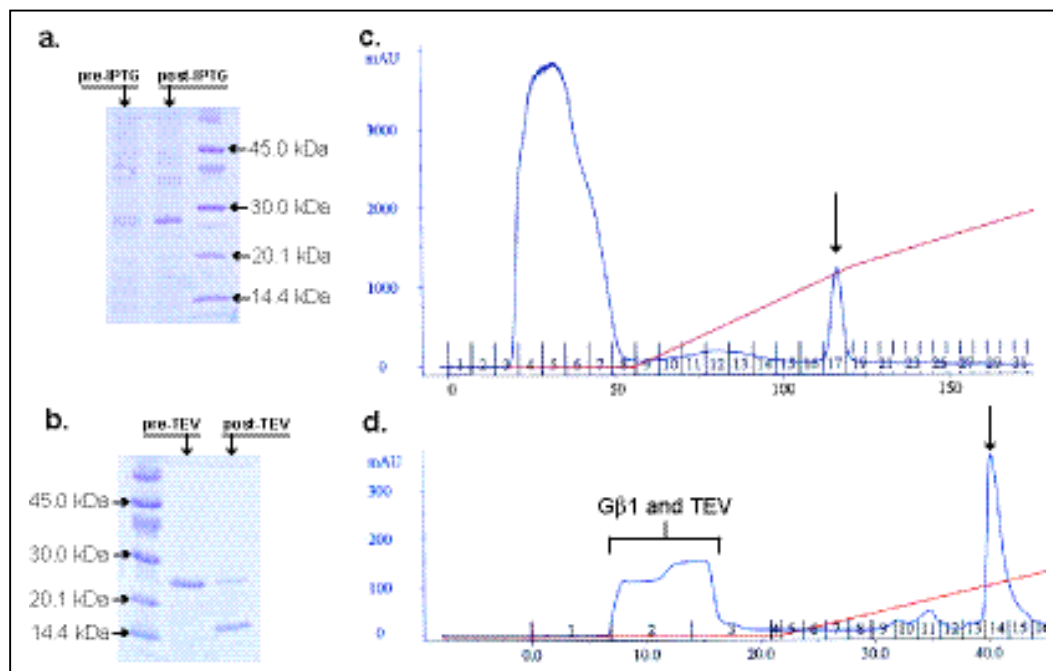
*b. Expression and purification of WC-1 PAS-B (558-686)*

Although initial results for this construct were very encouraging, limitations of the whole cell lysate method often require that purification protocols be performed on a protein of interest before high quality spectra can be obtained. Typical purification procedures for this protein proceeded as follows: The plasmid encoding the G $\beta$ 1 fusion of the PAS-B domain of WC-1 (558-686) was transformed into *E.coli* BL21(DE3) cells, plated onto LB-agar plates containing 0.1 mg/ml ampicillin and grown overnight at 37°C. 5 ml of LB/Amp was then inoculated with a single colony, grown to an OD<sub>600</sub>=0.6, spun down and transferred into 50 ml LB/Amp. This was grown again to OD<sub>600</sub>=0.6, and centrifuged again. The pellet was then introduced into 1L M9 minimal media enriched with 1g <sup>15</sup>NH<sub>4</sub>Cl for isotope labeling. After growth to OD<sub>600</sub>=0.6, 0.12g IPTG was added for the induction of expression, and the culture was then grown overnight (12-16 hr) at

20°C. Cell cultures typically reached OD<sub>600</sub> values of 3.0-3.5 post-induction. A sample gel showing the typical induction of expression is shown in Fig 2-7 a. The cells were then harvested, resuspended in buffer (50mM NaPi, 20mM NaCl, 5mM DTT, pH=6.0), subjected to high pressure extrusion for lysis, clarified by centrifugation (30 min @10,000g) and syringe filtered (0.45µm). The cell lysate was then injected over a Source-S ion-exchange column, and eluted with a linear NaCl gradient (20mM-500mM). A typical elution profile for this initial step is shown in Fig 2-7 b. Fractions were collected, concentrated and buffer exchanged into another buffer (50mM NaPi, 15mM NaCl, pH=7.5, final concentration of protein=25µM) for cleavage of the fusion tag (Gβ1) with TEV protease (1mg TEV/30mg protein, overnight at 25°C).

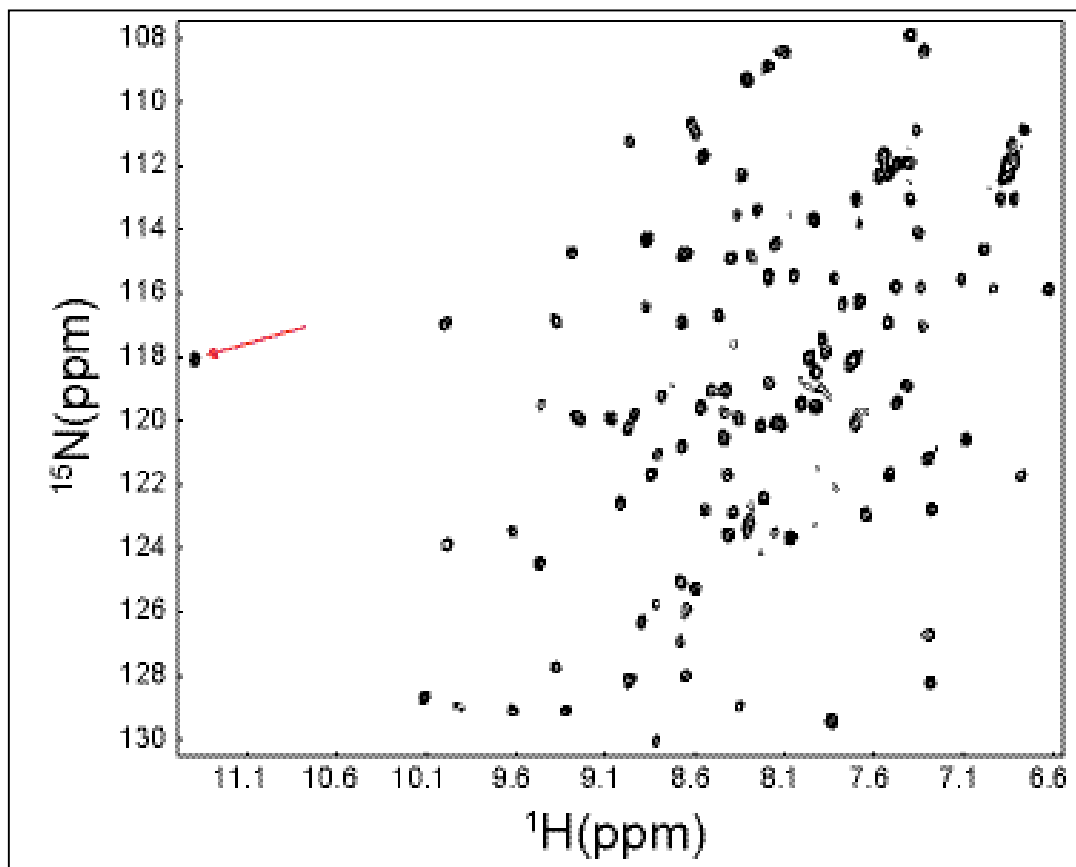
After overnight cleavage of the tag (Fig 2-7 c), another ion-exchange step was performed, this time to separate the cleaved tag from the WC-1 PAS-B domain and the TEV protease. The mixture is then injected over a mono-S column and a linear gradient of 20mM-350mM NaCl is applied for elution of the protein (Fig 2-7 d). From our experience, the cleaved Gβ1 tag does not bind the cation exchange columns, so it is effectively separated from the PAS domain at this step. The final purified protein from this procedure was then exchanged into the final buffer conditions (50mM NaPi, 50mM NaCl, pH=6.0). Occasionally, an additional gel filtration step was performed for further purification, and results from these experiments revealed that this protein eluted as a monomeric species (data not shown).

**Fig 2-7. a. Induction of expression of G $\beta$ 1-WC1-PAS-B(558-686). b. Cleavage of the G $\beta$ 1 fusion tag. c. First ion exchange chromatography step of G $\beta$ 1-WC1-PAS-B(558-686) purification. Lysate of 1L of cells induced for 12hr injected over a SourceS ion exchange column. Arrows denotes elution of fusion protein. d. Injection of TEV-cleaved protein over monoS ion exchange column. Cleaved fusion tag and TEV protease in flow through.**



$^{15}\text{N}/^1\text{H}$  HSQC spectra were obtained on the purified WC-1 PAS-B domain (558-686), and an example of typical results from these experiments is shown in Fig 2-8. The spectrum shows that the protein is well folded, as confirmed by the well-dispersed backbone amide resonances. In addition, the far downfield peak at around 11.2 ppm (red arrow) is of particular interest. A single peak near this location is often seen for well-folded PAS domains, the reasons for which will be discussed in more detail in Chapter III, section D.

**Fig 2-8.**  $^{15}\text{N}/^1\text{H}$  HSQC spectrum of purified WC-1 PAS-B (558-686). The concentration of protein used was  $250\mu\text{M}$  in  $20\text{mM NaP}_i$ ,  $50\text{mM NaCl}$ ,  $\text{pH}=6.0$  and spectra was acquired in a 20 min experiment ( $\text{nt}=8$ ). A single peak (red arrow) is in a location that is often diagnostic for the presence of a PAS domain fold.



Stability tests were then performed by concentrating the protein, allowing the samples to sit at room temperature for various periods of time, centrifuging the samples for 10 minutes at 10,000g and re-measuring the absorbance of the sample at 280nm. This revealed that the domain is soluble up to a final concentration of only  $160\mu\text{M}$  for extended periods of time ( $\sim 3$  days). Although this degree of solubility and stability is sufficient to obtain high quality spectra in an  $^{15}\text{N}/^1\text{H}$  HSQC, high resolution structural studies on proteins require experiments that involve multiple transfers of magnetization

through bonds via scalar couplings, as well as through-space transfers via Nuclear Overhauser Effects (NOEs), all of which can be prone to signal attenuation through inefficient magnetization transfer, exchange and relaxation effects. For these reasons, the modest concentrations achieved so far for this construct were deemed insufficient for high-resolution structural studies, given the limitations of the instrumentation present at that time (no cryogenically cooled probes available which can provide a 2-3 fold sensitivity gain). Further efforts were therefore made to optimize solvent conditions for this construct in an effort to achieve greater solubility and stability for the purpose of pursuing further NMR-based studies on this domain.

#### *vi. Solubility screening*

After a well-folded and aggregation-free construct is identified, solvent screening is often required to find conditions that maximize long-term solubility. Ionic strength, temperature, amphipathicity, reducing capacity, and pH are all important parameters that can greatly affect the behavior of proteins in solution. There is a wide range of buffers, stabilizers, and affinity tags that can be tested for their ability to keep proteins soluble for long periods of time. Several methods have been used to identify optimal solvent conditions for NMR studies of particular proteins, two of the best documented of which are the microdialysis button test (133) and the microdrop screen (134). Both of these approaches require a small quantity of protein (~0.1-0.2 mg/sample) that is distributed into different solvent conditions and visually assessed for precipitation over time. In the microdialysis button technique, very small aliquots of protein are placed in specially designed reservoirs covered by dialysis membranes, allowing the gradual exchange into

new solvent conditions while ensuring that protein concentration remains unchanged. The microdrop screening method instead uses a setup common to many crystallization trials, where small drops of protein solution on slides are suspended over many different types of buffer in sealed chambers. Vapor diffusion brings each sample to equilibrium, providing an effective screen for conditions while allowing the protein concentration to change during the screening process. An extensive review of useful solvent conditions and the screening methods mentioned here is presented elsewhere (135).

*a. Solvent screening of WC-1 PAS-B*

To optimize the solvent conditions required to increase the solubility of the WC-1 PAS-B domain, we chose to employ the microbutton dialysis method. During the initial efforts to characterize this domain, it was found that under acidic conditions (pH = 4.5), this construct was soluble for short periods of time to concentrations up to about 700 $\mu$ M. Although these concentrations were sufficient for high-resolution structural studies, the question of physiological relevance under these relatively extreme conditions convinced us to screen for other factors that would allow similar solubilities at a higher pH. The initial procedure used was as follows: Purified protein was concentrated in (20mM NaP<sub>i</sub>, 50mM NaCl, pH=5.0 – final concentration=540 $\mu$ M), 10 $\mu$ l of this solution was placed in each reservoir of 24 microdialysis buttons which were sealed with dialysis membranes (10 $\mu$ m pore size), and each was then placed in a 50 ml conical tube with a particular set of buffer conditions. The conical tubes were then placed at 25°C for three days with gentle shaking on a platform rocker. All possible combinations of the following buffer variables were screened:

- 50mM NaP<sub>i</sub> or 50 mM NaOAc
- 10mM or 50mM NaCl
- pH = 4.5, 5.0, 5.5, 6.0

The results of the initial screen showed that significant precipitation occurred over the time period used for any pH above 5.0, regardless of the buffering agent or NaCl concentration used. A follow up screen was performed that reduced the concentration of NaP<sub>i</sub> to 10mM or 20mM, and also tested the effects of 5mM dithiothretal (DTT) to see if intra- or intermolecular disulfide bond formation could be responsible for poor long-term solubility. The results showed again that the most important factor in the solubility of this construct was pH, although the presence of DTT also seemed to be detrimental for stability. The most favorable conditions identified were subjectively identified to be 10mM NaPi, 10mM NaCl, 0mM DTT, pH=4.5 (580μM final concentration over three days), although the same conditions at a slightly higher pH (5.0) also resulted in an acceptable increase in long-term solubility (560μM final concentration over three days).

Although the screening procedures described in this chapter served us well in identifying a well-folded construct of the White Collar-1 PAS-B domain, and further helped improve the behavior of this construct in solution, the difficulties involved in obtaining similar results for other PAS domains from the *Neurospora crassa* circadian oscillator prompted us to temporarily abandon this project in order to pursue studies on another system that will be the subject of the next chapter. Nevertheless, development of the screening and optimization procedures described therein have proven invaluable to

our lab for preparing small protein domains for study in a variety of systems, and will undoubtedly continue to be an important first step in many NMR-based investigations.



### III. ARNT in the Hypoxia Response

(Parts of this chapter adapted from Card et al. 2005 (136))

#### *A. The Aryl hydrocarbon Receptor Nuclear Translocator*

The Aryl Hydrocarbon Receptor Nuclear Translocator (ARNT) is a promiscuous bHLH-PAS (Per-ARNT-Sim) protein that forms heterodimeric transcriptional regulator complexes with several other bHLH-PAS subunits to control a variety of biological pathways, some of which are centrally involved in disease initiation and/or progression. One of these is the hypoxia response pathway, which allows eukaryotic cells to respond to low oxygen tension via the formation of a heterodimeric complex between ARNT and another bHLH-PAS protein, the hypoxia inducible factor alpha (HIF- $\alpha$ ). Basic helix-loop-helix (bHLH)/PAS domain proteins are essential transcriptional regulators of a wide range of eukaryotic cellular and developmental processes (reviewed in Kewley *et al.* (3)). As with other bHLH-containing proteins, this regulation is dependent upon the formation of homo- and heterodimeric DNA-binding complexes. While it is clear that specific residues in the bHLH regions of an assembled dimer mediate DNA binding, dimerization itself appears to involve a more complex interplay between the bHLH and two PAS domains found in each monomeric partner. An important distinction between proteins in this family is that one group is capable of forming both homo- and heterotypic interactions, while others can only form heterodimeric complexes with members of the former group. Because these dimerization preferences affect the distinct DNA-binding characteristics of resulting protein complexes, thereby determining the regulation of individual genes, it is important to understand the structural basis of this specificity.

The best characterized member of the first class of bHLH/PAS proteins is ARNT (Aryl Hydrocarbon Receptor Nuclear Translocator), which forms both homodimers and heterodimers with several other bHLH/PAS factors including AHR (Aryl Hydrocarbon Receptor), various isoforms of HIF- $\alpha$  (Hypoxia Inducible Factor) and Sim (single-minded) (54, 79, 137-140). Although the *in vivo* importance of ARNT homodimers remains unclear, each of the heterodimeric ARNT-containing complexes plays a critical function in pathways used to respond to environmental conditions (e.g. exposure to xenobiotic compounds for AHR, hypoxic conditions for HIF- $\alpha$ ) or developmental processes (e.g. HIF- $\alpha$ , Sim – reviewed in (3)). The biological importance of ARNT in these pathways is emphasized by the embryonic lethality of murine ARNT knock-out mutations (74). In addition, a (1;12(q21;p13) translocation that results in a fusion between the translocated ETS leukemia gene (TEL) and ARNT was found in an acute myeloblastic leukemia (AML-M2), and is thought to promote the disease through impairment of the cellular response to extracellular hypoxic stimuli (141).

The role ARNT plays in these signaling pathways is well illustrated by the hypoxia response pathway, which serves to upregulate the expression of a variety of genes that facilitate cell survival during periods of low oxygen availability. Central to this pathway is the carefully regulated formation of HIF, a heterodimeric transcriptional activation complex comprised of ARNT and one of the three isoforms of HIF- $\alpha$  (HIF-1 $\alpha$ , HIF-2 $\alpha$  and HIF-3 $\alpha$ ) (142-144). A description of this pathway was presented earlier in the introductory chapter and Fig.1-3. The essential role of ARNT in this signaling process has recently been highlighted by the ability of a liver-specific Cre/Lox knockout

of ARNT to suppress tumor formation in a transgenic mouse line that maintains constitutively high levels of HIF- $\alpha$  under normoxic conditions (145).

It has been shown that while the isolated bHLH domains of ARNT, AHR and HIF-1 $\alpha$  can form heterodimeric complexes and bind E-box like enhancer elements *in vitro*, the inclusion of the adjacent PAS domains enhances the affinity and specificity of the protein/DNA interaction (146, 147). Using biophysical and biochemical approaches, we have previously extended these studies with investigations of the C-terminal PAS domain of human HIF-2 $\alpha$  (HIF-2 $\alpha$  PAS-B). Our lab has demonstrated that PAS-B domains from HIF-2 $\alpha$  and ARNT interact *in vitro* via solvent-exposed residues on the  $\beta$ -sheet surface of HIF-2 $\alpha$ , and showed that this interaction is vital for robust transcriptional activation of an HRE-driven luciferase reporter gene in living cells under hypoxic conditions (8). Further work in the lab of Rick Bruick, a close collaborator on this project, has indicated that both PAS domains in HIF-2 $\alpha$  are required for heterodimer formation, and that the same mutations on the solvent-exposed  $\beta$ -sheet surface found to attenuate the hypoxic response in living cells disrupt heterodimer formation in cell lysates by co-immunoprecipitation (71).

Although many PAS domains involved in biological signaling pathways have been identified, only limited structural information is available regarding how they interact with each other in multiprotein complexes. It has been shown that both PAS domains of HIF-2 $\alpha$  are required for HIF activity in living cells, and also that both contribute to the heterodimerization of full length HIF subunits as assayed by co-immunoprecipitation in the lysates of hypoxic cells (71). Point mutations on the solvent-exposed side of the HIF-2 $\alpha$  PAS-B domain that disrupt the interaction with ARNT PAS-

B *in vitro* were also shown to attenuate the hypoxia response in living cells. Although this evidence indicates that the PAS-B interaction between HIF-2 $\alpha$  and ARNT is important for a robust response to low oxygen tension, and that the central PAS domain  $\beta$ -sheet is involved, specific details of the domain/domain interactions had not been previously elucidated.

### *B. Solution structure of ARNT PAS-B*

A main goal of the work described here was to further explore the role of the ARNT PAS-B domain in the formation of the heterodimeric HIF complex through structural studies of the domain alone and in complex with its binding partner, HIF PAS-B. The high-resolution solution structure of HIF-2 $\alpha$  PAS-B had been previously solved, so the first step was to complement this information with the structure of the corresponding PAS domain in ARNT.

#### *i. Cloning, expression and purification of ARNT PAS-B for structure determination.*

A PCR-generated DNA fragment encoding human ARNT PAS-B (residues 356-470) was subcloned into the pHis-parallel bacterial expression vector (148). The plasmid was then transformed into *Escherichia coli* BL21(DE3) cells that were then grown in M9 minimal media containing 1g/L  $^{15}\text{NH}_4\text{Cl}$  for U- $^{15}\text{N}$  samples (and 3g/L  $^{13}\text{C}_6$  glucose for U- $^{15}\text{N}/^{13}\text{C}$  labeled samples). Cultures were grown at 37°C to an  $A_{600}$  of 0.6-1.0 and induced at 20°C for 12-14 hours by the addition of 0.5mM isopropyl  $\beta$ -D-thiogalactoside. Subcloning and expression of HIF-2 $\alpha$  PAS-B was done as described previously (8).

Purification of His<sub>6</sub>-tagged ARNT PAS-B proceeded as follows. Cell pellets were harvested and resuspended in 20ml of 50mM sodium phosphate buffer (pH 7.5)-15mM NaCl, 20mM imidazole, lysed by high-pressure extrusion, and centrifuged. The supernatant was then passed through 0.45µm syringe filters and injected over a Ni<sup>2+</sup>-loaded NTA column, eluted using a linear 20-500 mM imidazole gradient, and exchanged into imidazole-free buffer. The His<sub>6</sub>-tag was cleaved by incubating overnight with His<sub>6</sub>-TEV protease. Both the His<sub>6</sub>-TEV and free His<sub>6</sub>-tag were removed by a combination of Ni<sup>2+</sup>-NTA affinity and Superdex 75 size-exclusion chromatography. The final ARNT PAS-B construct contains four vector-derived N-terminal residues (GAMD), and the proper molecular weight was confirmed by electrospray ionization mass spectrometry.

## *ii. Solution Structure of ARNT PAS-B.*

### *a. Materials and methods*

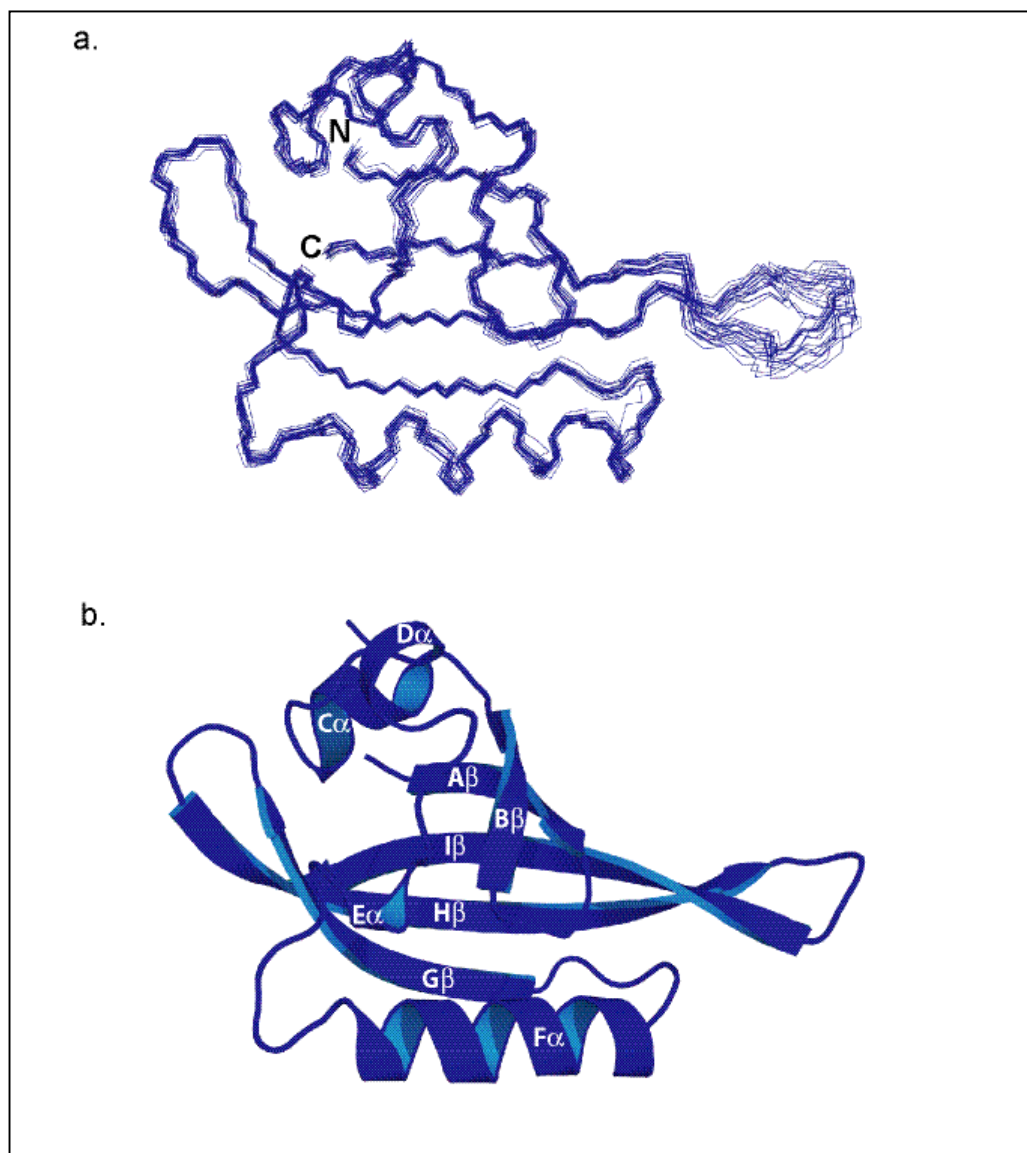
We determined the solution structure of residues 356-470 of human ARNT (ARNT PAS-B) using standard double- and triple resonance NMR experiments performed on uniformly <sup>15</sup>N and <sup>15</sup>N/<sup>13</sup>C labeled protein samples. In the determination of this structure, we used >4500 geometric constraints derived from interproton distances, dihedral angles and amide N-H bond alignments obtained from partially aligned samples (Table 1). These data are well satisfied by the ensemble of 20 lowest energy structures obtained from restrained molecular dynamics calculations in ARIA/CNS (Fig 3-1 a).

NMR samples of ARNT PAS-B used for structure determination consisted of 1mM protein in 50mM sodium phosphate buffer, 15mM NaCl, 5mM DTT, 5mM NaN<sub>3</sub>,

and a protease inhibitor mixture (Sigma) at pH 6.5 (90% H<sub>2</sub>O/10% D<sub>2</sub>O). NMR data were recorded at 25°C with Varian Inova 500 and 600 MHz spectrometers using NMRPipe for data processing (*149*) and NMRView for analysis (*150*). Backbone chemical-shift assignments were obtained from 3D HNCACB, CBCA(CO)NH, HNCO, and HNHA spectra. Aliphatic side chain resonances were assigned from C(CO)-NH total correlation spectroscopy (TOCSY), H(CCO)-NH TOCSY, and HCCH-TOCSY experiments. Chemical shift assignments have been deposited with BioMagResBank (accession #6597). Deuterium exchange reactions were initiated by resuspending a lyophilized <sup>15</sup>N-labeled ARNT PAS-B in 99% D<sub>2</sub>O (800 μM concentration, pH 7.5) and sequentially acquiring <sup>15</sup>N/<sup>1</sup>H (HSQC) spectra every 15 minutes at 25°C. Observed <sup>2</sup>H exchange rates were converted into protection factors using standard methods (*151*).

Interproton distance constraints were obtained from 3D <sup>15</sup>N edited NOESY ( $\tau_m = 150$  ms), <sup>15</sup>N, <sup>13</sup>C edited NOESY ( $\tau_m = 100$  ms), and 2D NOESY ( $\tau_m = 150$  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 D<sub>2</sub>O exchange for > 30 minutes (25°C, pH 7.5).  $\phi$  and  $\psi$  dihedral angle constraints were generated by chemical-shift analysis using TALOS (*152*), with two times the standard deviation of TALOS predictions as the bounds (minimum bounds:  $\pm 15^\circ$ ). Also, 86 <sup>15</sup>N-<sup>1</sup>H residual dipolar coupling constraints were obtained from a partially aligned sample in 5% (wt/vol) DPMC/DHPC ratio of 3:1 (Avanti Polar Lipids) and 5mM cetyltrimethylammonium bromide (CTAB) at 35°C.

**Fig 3-1. a.** Ensemble of 20 lowest energy structures obtained from restrained molecular dynamics calculations in ARIA/CNS. N- and C-termini are labeled accordingly. **b.** Ribbon diagram of the structure closest to the mean from the calculations, with typical PAS secondary structural elements labeled.



**Table 3-1. Statistics for ARNT PAS-B solution structure determination**List of Constraints

NOE distance restraints	
Unambiguous	3421
Ambiguous	975
Hydrogen bond restraints	37
Dihedral angle restraints	140
<sup>15</sup> N- <sup>1</sup> H residual dipolar couplings	86

Structural Analysis

Mean r.m.s. d. from experimental restraints

NOE (Å)	0.0179 +/- 0.0015
Dihedral angles (deg)	0.452 +/- 0.078

Average number of:

NOE violations > 0.5 Å	0
NOE violations > 0.3 Å	1.80 +/- 1.17
dihedral violations > 5°	0

Mean r.m.s. from idealized covalent geometry

bonds (Å)	0.0042
angles (deg.)	0.54
impropers (deg.)	1.59

Geometric analysis of residues 361-445 and 456-463

rmsd deviation to mean	0.50 +/- 0.07 Å (backbone)
	1.17 +/- 0.08 Å (all heavy)

Ramachandran analysis (procheck) 83.6% most-favored region  
14.1% additionally allowed region  
1.46% generously allowed region  
0.81% disfavored

*b. Results*

Initial structures were obtained with automated assignments of NOESY spectra using ARIA 1.2 (153, 154) and refined with the incorporation of both automated and manual assignments. From the final 1,000 structures, the 20 with the lowest energies were analyzed with MOLMOL (155) and PROCHECK-NMR (156). This ensemble has been deposited with the RCSB/PDB database (accession code 1X0O).

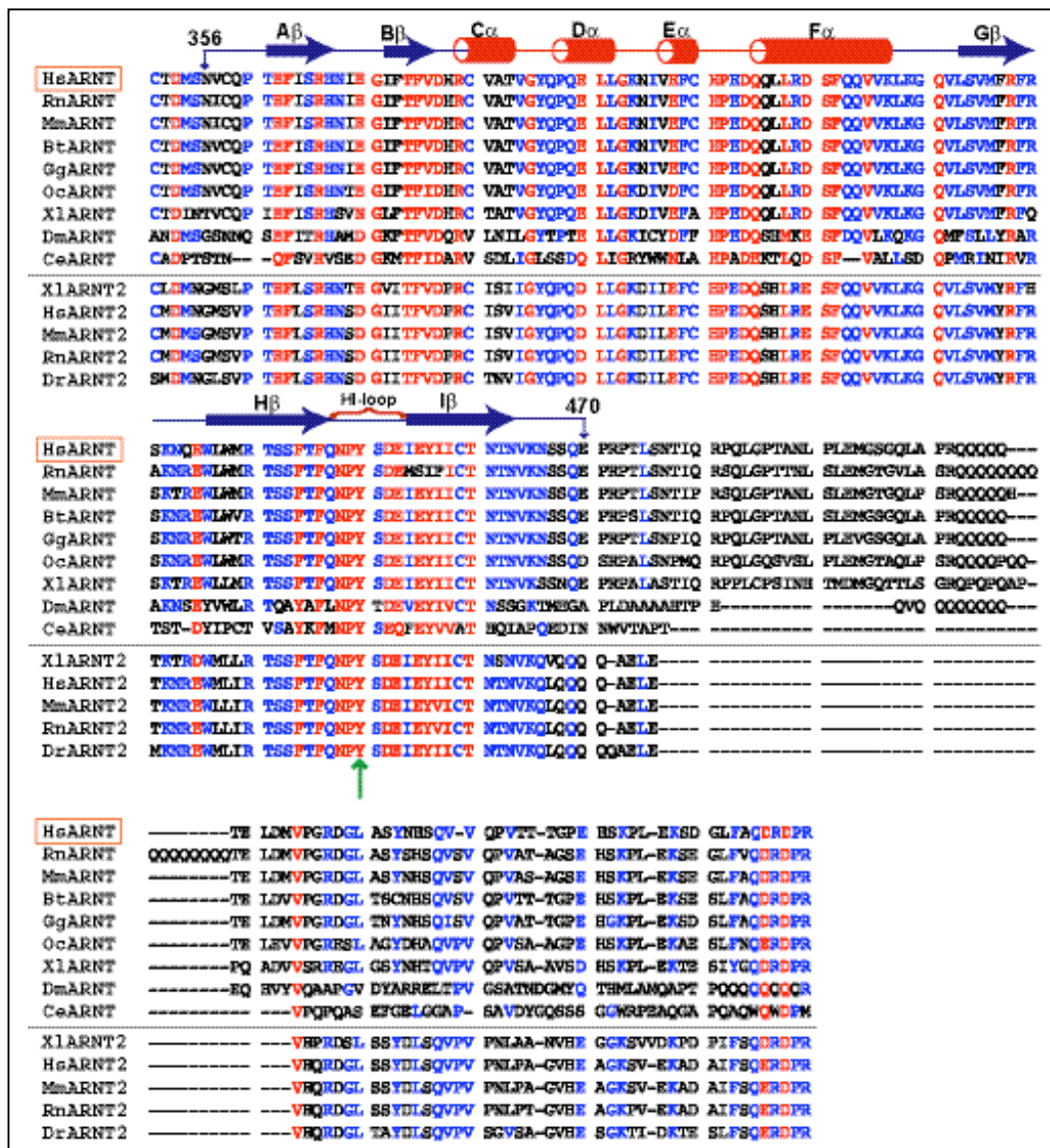


From this ensemble, the structure closest to the mean was used for structural alignment of ARNT PAS-B with eight other PAS domains using the MASS software package (157), based on secondary-structure based fitting and alignment. Additional PAS domains included in this analysis with their respective Research Collaboratory for Structural Bioinformatics Protein Data Bank IDs were from the proteins: HERG (1BYW), EcDOS (1S67), PHY3 (1G28), PASK (1LL8), individual PAS A and PAS B domains from PER (1WA9), PYP (1D7E), and HIF-2 $\alpha$  PAS-B (1P97). The MASS program automatically identified a group of 49 residues for this alignment, most of which came from the central  $\beta$ -sheet with the exception of the G $\beta$  strand. Notably, the long F $\alpha$  helix was also not used for the alignment, presumably due to the well-established variability of the position of this helix with respect to the core in many PAS domain structures (158). The average r.m.s.d. of C $\alpha$  atoms for this alignment was found to be 1.5 Å using this analysis.

From the calculated structures it is clear that ARNT PAS-B has a standard PAS domain fold, including a five-stranded anti-parallel  $\beta$ -sheet flanked by several  $\alpha$ -helices that lie on one face of the protein (Fig 3-1). The representative structure of this ensemble is very similar to other PAS domain structures as evidenced by the low r.m.s.d. obtained by optimized superposition of secondary structure elements in ARNT PAS-B and other PAS domain structures. Notable in the ARNT PAS-B structure is a ten-residue long HI loop connecting the H $\beta$  and I $\beta$  strands that exhibits some intermediate exchange behavior on the NMR timescale, indicative of dynamic motion occurring in this region. Similar flexible regions have been observed in other PAS domains, including those of murine NPAS2 PAS-A and *Drosophila* Period (5, 159), possibly suggesting a functional role for

this region. This idea is also supported by multiple sequence alignments of human ARNT with a variety of homologues which shows nearly complete conservation of residues in this HI loop across many organisms (Fig 3-2).

**Fig 3-2. Multiple sequence alignment of human ARNT and some of its homologues. The alignment shows good sequence alignment in the HI loop, a region possibly important for function.**



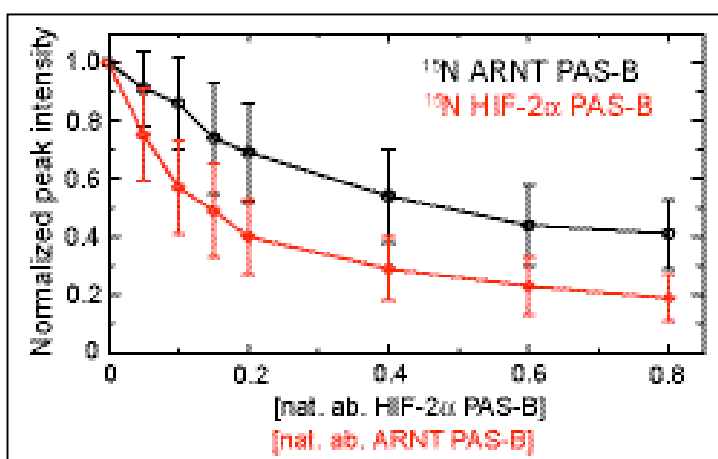
*iii. ARNT PAS-B Forms Heterodimers via the central  $\beta$ -sheet.*

Previously, we reported the solution structure of HIF-2 $\alpha$  PAS-B and demonstrated that residues on the solvent-exposed side of the central  $\beta$ -sheet are used to bind ARNT PAS-B (8). We therefore wanted to obtain similar information from the ARNT PAS-B domain. Confirmation of complex formation of the HIF-2 $\alpha$  PAS-B/ARNT PAS-B heterodimer was accomplished by titration analysis using stepwise addition of natural abundance HIF-2 $\alpha$  PAS-B (0mM to 1mM over 8 steps) into 250  $\mu$ M  $^{15}$ N-labeled ARNT PAS-B, and observing changes in chemical shift and/or peak intensities of ARNT PAS-B backbone amide resonances in  $^{15}$ N/ $^1$ H HSQC experiments. Backbone amide resonances of residues that exhibited a significant change in chemical shift ( $\Delta\delta_{\text{TOT}} > 0.05$  ppm, using Eq. 3-1) with  $\Delta\delta_{\text{TOT}}$  normalized for proton with the scale factor  $\chi=0.17$  established from estimates of atom-specific chemical shift ranges in a protein environment (160) were interpreted as being involved in complex formation.

$$\text{Eq. 3-1: } \Delta\delta_{\text{TOT}} = [(\Delta\delta^1\text{H})^2 + (\chi * \Delta\delta^{15}\text{N})^2]^{1/2}$$

Data from resonances displaying both intermediate exchange linebroadening and chemical shift changes (Fig 3-3 a) exhibited concentration dependence consistent with the 30 $\mu$ M dissociation constant that we previously determined by monitoring the effect of ARNT addition into  $^{15}$ N-labeled HIF-2 $\alpha$  (8). This interaction is specific by several lines of evidence. For example, titration analysis done by Paul Erbel and myself on the HIF-2 $\alpha$  and ARNT PAS-B domains respectively, showed that normalized peak intensity loss upon complex formation became saturated, indicating that corresponding changes in peak

intensities and chemical shift changes were not the result of non-specific aggregation events (Fig 3-3). Additionally, previous work in our lab showed that addition of a non-related PAS domain (PAS-A from human PAS kinase) to  $^{15}\text{N}$ -labeled HIF-2 $\alpha$  PAS-B did not result in significant broadening or changes in chemical shifts (data not shown). This is also supported by mutagenesis studies on ARNT PAS-B (discussed in Chapter IV) that indicate very minimal spectral changes occurring during the titration of non-interacting ARNT mutants, despite the relatively high concentrations used (up to 1mM).

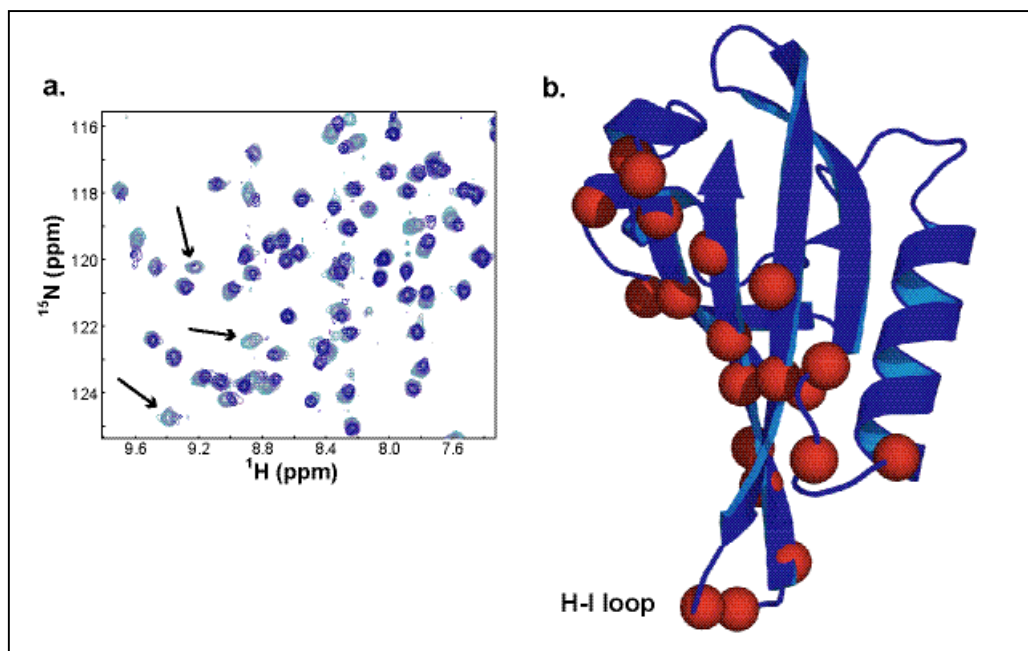


**Fig 3-3. Association of HIF-2 $\alpha$ /ARNT PAS-B domains is not due to non-specific aggregation.** Normalized  $^{15}\text{N}/^1\text{H}$  HSQC peak intensities of HIF-2 $\alpha$  (black) and ARNT PAS-B (red) domains are plotted as increasing amounts of unlabeled partner are titrated in. Saturation is indicated for both, and the overall lower intensities of HIF-2 $\alpha$  could indicate that ARNT is partially involved in a homodimeric interaction at the beginning of the titration.

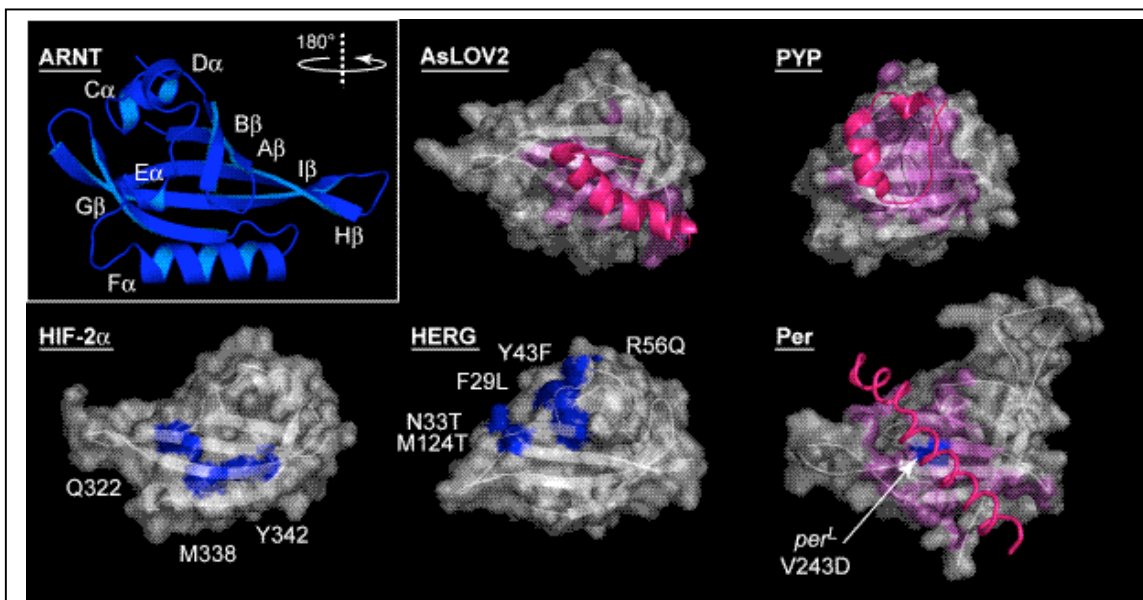
Backbone amide resonances that represent significantly perturbed residues upon complex formation were then mapped onto the high-resolution structure of the ARNT PAS-B domain (Fig 3-4 b), and a very similar interface was found for ARNT PAS-B that was previously described for HIF-2 $\alpha$  PAS-B in the same complex. Several groups, including our own, have previously observed that this  $\beta$ -sheet interface is widely used by PAS domains for intra- and intermolecular protein/protein interactions (8, 161), suggesting that it may represent a unifying link among some members of the family (Fig 3-5). Understanding the basis for the specificity that guides the formation of variable

complexes that utilize this interaction interface is one of the main goals of this study, so the next step was the structural characterization of the HIF-2 $\alpha$ /ARNT PAS-B heterodimer.

**Fig 3-4. a.** Tiration of unlabeled HIF-2 $\alpha$  PAS-B (0 $\mu$ M, 100 $\mu$ M, 500 $\mu$ M, 1mM from light to dark blue) into 250 $\mu$ M  $^{15}$ N-labeled ARNT PAS-B. Arrows highlight backbone amide resonances that broaden due to intermediate exchange. **b.** Representative structure of ARNT PAS-B with red spheres indicating resonances displaying both intermediate exchange linebroadening and significant chemical shift changes according to the criteria outlined in the text.



**Fig 3-5. Solvent exposed  $\beta$ -sheet surface is a common interface for many PAS domains. Several PAS domains for which structures have been determined are shown. The LOV-PAS domain from phototropin (AsLOV) interacts with a helical C-terminal extension ( $J\alpha$ ) at a very similar interface (pink) described for the interaction of PYP with its own N-terminal  $\alpha$ -helix, and for Per with the C-terminal helix ( $\alpha F$ ) of its homodimeric binding partner (discussed in more detail in section iii). This correlation also extends to interfaces identified through mutagenesis, such as in the HIF-2 $\alpha$  PAS-B domain, shown with mutations that disrupt the hypoxia response in luciferase reporter assays, and HERG, where  $\beta$ -sheet mutations in patients have resulted in cardiac disorders. Also shown is the ribbon diagram of the solution structure of ARNT for comparison of secondary structural elements. For a more details regarding the HIF-2 $\alpha$  and ARNT PAS-B interfaces, see chapter IV.i and Fig. 4-1.**



### C. Model of the ARNT PAS-B/HIF PAS-B heterodimer

While several crystal structures of PAS domain homodimers have been solved to date (5, 6, 25, 44, 45), structures of PAS domain heterodimeric complexes have been elusive. Elucidation of the structural details of PAS domain heterodimerization in bHLH-PAS transcription factors could provide important clues as to how the PAS modules contribute to the binding affinities and/or provide the basis of specificity in these complexes. A major effort in the course of this work was to use the high-resolution

solution structures solved for the HIF-2 $\alpha$  and ARNT PAS-B domains as a basis for the structure determination of the heterodimeric complex between these two domains.

*i. Model of HIF-2 $\alpha$ /ARNT PAS-B Complex.*

Upon formation of the HIF-2 $\alpha$ /ARNT PAS-B complex, we typically observed significant broadening of signals from HIF-2 $\alpha$  PAS-B, and to a much lesser extent, ARNT PAS-B. We attribute this to both the slower tumbling of the complex and a likely contribution from intermediate timescale chemical exchange broadening. While this effect aided in the identification of interfacial residues, it also complicated the acquisition and assignment of intermolecular NOEs typically used to obtain distance restraints for determining the structure of multiprotein complexes. To address these difficulties, we developed a model of the HIF-2 $\alpha$ /ARNT PAS-B heterodimer using the program HADDOCK, which assembles the complex from high-resolution structural ensembles of the isolated partners (162). This is achieved via an energy minimization approach that progressively reduces violations of experimentally-derived restraints, including intermolecular distances between residues on the interacting surfaces of the two partners, while also optimizing van der Waals and electrostatic interactions. Starting from our ensembles of 10 structures each of HIF-2 $\alpha$  and ARNT PAS-B, this approach produced a single cluster of complex structures based on the correlation of overall backbone r.m.s.d. values of individual structures within the ensemble (backbone r.m.s.d. = 1.21Å).

Models of the HIF-2 $\alpha$ /ARNT PAS-B complex were generated using HADDOCK1.2 (163) in conjunction with CNS(153). HADDOCK1.2 provides support for docking ensembles of structures, allowing us to start these calculations with

ensembles of the ten lowest energy NMR structures of the PAS-B domains of HIF-2 $\alpha$  (8) and ARNT (above). The active residues used to define the ambiguous interaction restraints (AIRs) were solvent accessible (as calculated by NACCESS(164)) and demonstrated significant chemical shift perturbations or peak disappearance in  $^{15}\text{N}/^1\text{H}$  HSQC titration experiments. Only residues having an average solvent accessibility (plus standard deviation) of over 50% were retained for the definition of active and passive residues. Passive residues were chosen as those that were surface neighbors of the active residues as identified by visual inspection. A complete list of the residues defined at the interface of each PAS-B domain are: HIF-2 $\alpha$  residues 242-262, 299-308, 318-348 and ARNT residues 358-368, 372-384, 420-426, 440-465. The N- and C-termini of both proteins were allowed to be fully flexible at all stages of the modeling, as was the HI loop of ARNT PAS-B (residues 448-453). AIRs were implemented with a 2.0 Å distance cutoff. Docking proceeded in two stages, beginning with the generation of an initial set of 1000 rigid-body docking solutions (10 trials per combination of HIF-2 $\alpha$  and ARNT starting structures). This set was sorted by the intermolecular energy of each complex (sum of van der Waals, electrostatic, AIRs), with the 200 lowest energy complexes submitted for a second semi-flexible simulated annealing stage. The fifty lowest energy structures from this stage were subjected to a final simulated annealing refinement in water, and clustered using a 2.0 Å R.M.S.D. cut-off criterion. The twenty lowest energy structures from this final group were used for all analyses and have been deposited with the RCSB/PDB database (accession #2A24).

The resulting complex is dominated by intermolecular interactions between residues located on the  $\beta$ -sheet surfaces of both PAS domains, burying an average of



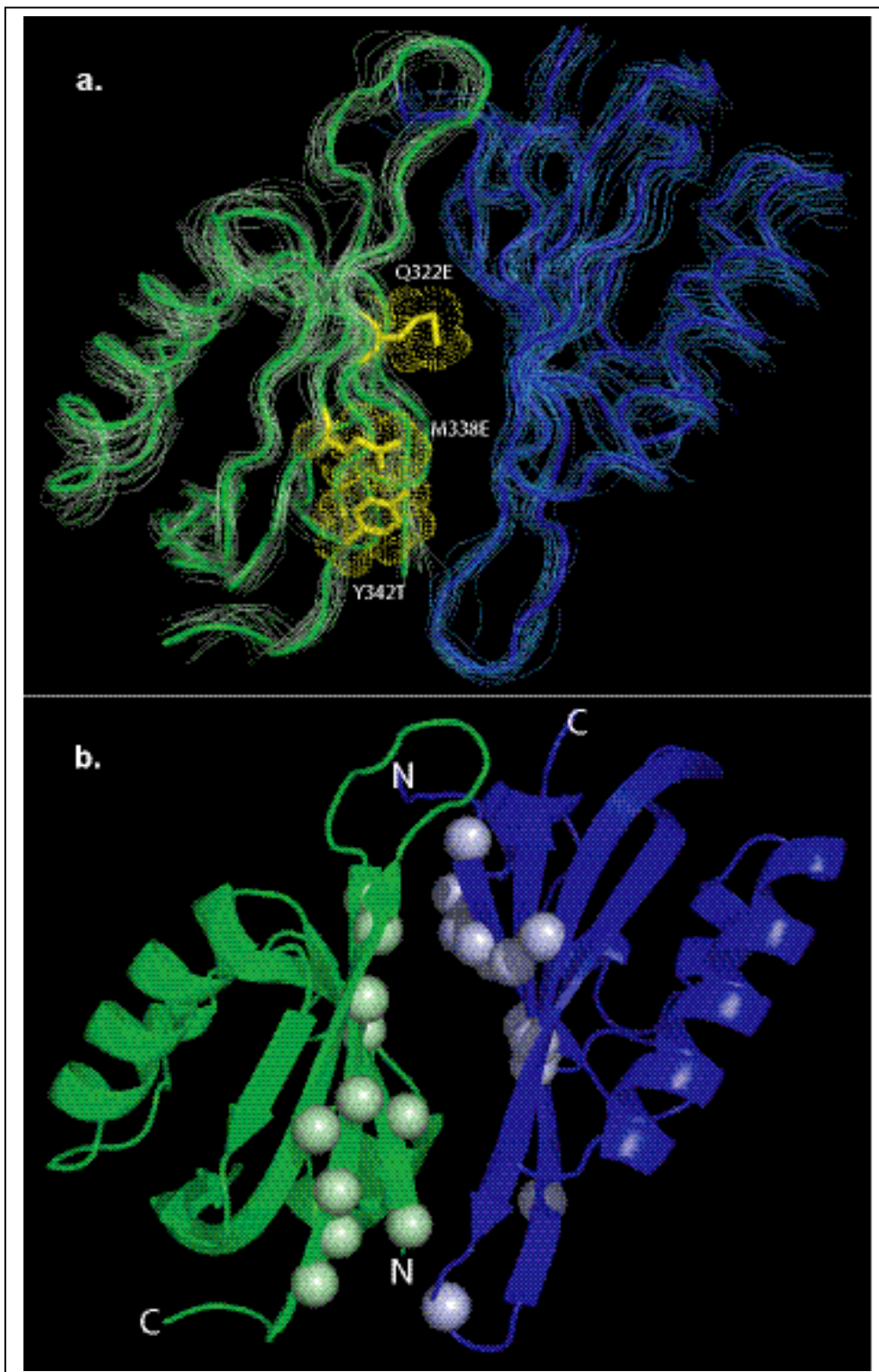
2200 Å<sup>2</sup> in total surface area (Fig. 3-6, Table 3-2). This arrangement also buries three residues on HIF-2α (Gln322, Met338 and Tyr342) which we have previously found to be critical for the HIF-2α/ARNT PAS-B interactions *in vitro* and in living cells (8).

Intriguingly, this model clearly shows an anti-parallel association, placing the HIF-2α N- and C-termini at opposite ends of the complex from those of the ARNT subunit. This arrangement is supported by the examination of electrostatic surface potentials calculated for both PAS-B domains in the complex (Fig 3-7). These calculations indicate that positively and negatively charged surface residues on each partner interact with appropriately charged residues on the corresponding interface, providing an orientational specificity to the interaction that otherwise involves hydrophobic surfaces on each monomeric unit.

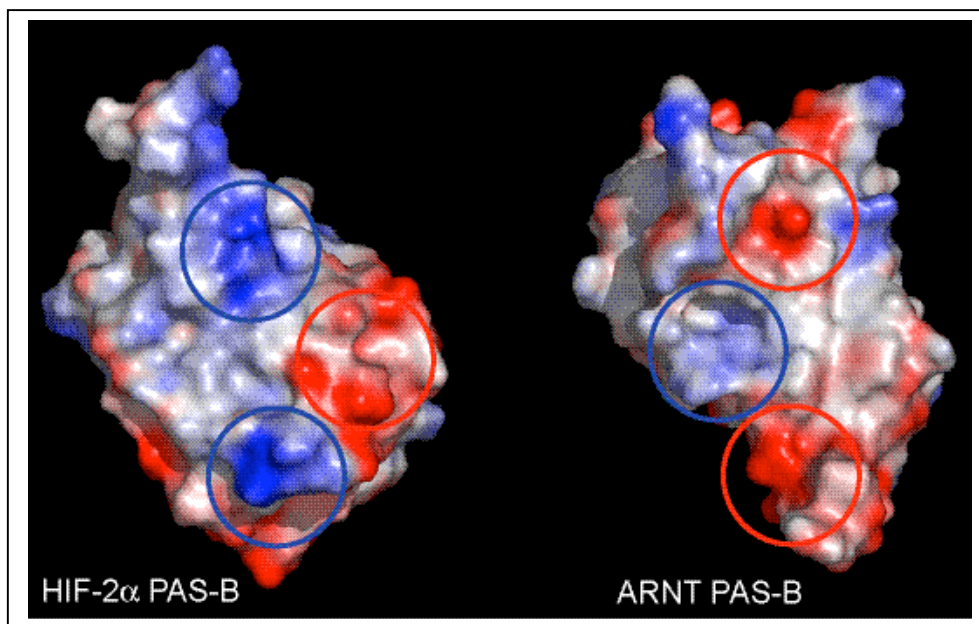
**Table 3-2. Statistics for ARNT PAS-B/HIF2α PAS-B complex model structures**

Number of structures in analysis	20
Backbone rmsd (Å) with respect to mean	
Flexible interface	1.21 ± 0.32
All backbone	1.25 ± 0.31
Number of ambiguous interaction restraints (AIRS)	
From ARNT PAS-B	13
From HIF-2α PAS-B	11
Total AIRs	24
CNS intermolecular energies after water refinement	
E <sub>ext</sub> (kcal mol <sup>-1</sup> )	-72.7
E <sub>int</sub> (kcal mol <sup>-1</sup> )	-497.9
Buried surface area (Å <sup>2</sup> )	2205
Average number of:	
AIR violations > 0.3 Å	0
dihedral violations > 5°	0.4
Mean r.m.s. from idealized covalent geometry	
bonds (Å)	0.003
angles (deg.)	0.410
impropers (deg.)	0.380
Ramachandran analysis (procheck)	81.4% most-favored region
	17.1% additionally allowed region
	0.8% generously allowed region
	0.7% disfavored

**Fig 3-6. Results of the modeling of the HIF-2 $\alpha$ /ARNT PAS-B interaction using HADDOCK. a. The ensemble of the ten lowest energy structures with the structure closest to the mean highlighted. The three mutations shown previously to disrupt the interaction are shown in yellow. b. Ribbon diagram of the representative complex structure with the backbone amide protons from residues used as AIRS in the HADDOCK calculation shown as spheres.**



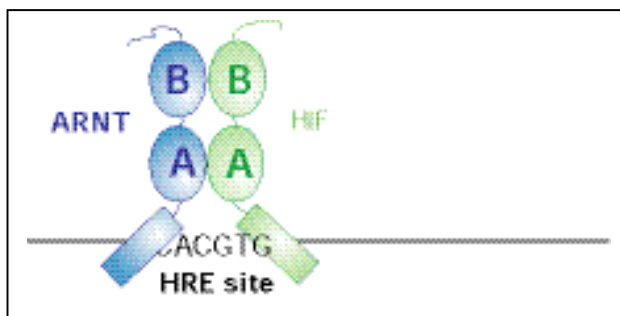
**Fig 3-7.** Starting with the view shown in Fig 3-4, the monomeric components of the HADDOCK-derived complex structure were rotated 90° in opposite directions to reveal their interfacial surfaces, with HIF-2 $\alpha$  on the left and ARNT on the right. Surface electrostatic potentials were calculated using APBS (4) and displayed on each structure using a color gradient ranging from negative (red = -4kT) to neutral (white) to positive (blue = +4kT).



*ii. Spin-labeling of ARNT PAS-B.*

Although the HADDOCK calculations clearly favor an anti-parallel association between the HIF-2 $\alpha$  and ARNT PAS-B domains, this result was unexpected in light of the simple two-dimensional model where the PAS domains of each protein interact in a reciprocal manner with the corresponding domains on the other partner (Fig 3-8).

**Fig 3-8.** Two-dimensional model of the interaction between HIF and ARNT in the formation of the intact heterodimer. The model shows a parallel interaction between the individual PAS domains involved in the complex.



To confirm the anti-parallel arrangement of the heterodimeric PAS-B complex, we used site-directed spin labeling as an independent biophysical method to measure long (< 20 Å) distances between corresponding regions of HIF-2 $\alpha$  and ARNT.

*a. Paramagnetic broadening enhancement theory.*

The relaxation of a proton that exists in the vicinity of an unpaired electron is generally dominated by the electron-nuclear dipole-dipole interaction, due to the fact that the electron magnetic moment is ~700 times greater than that of a proton. Given the Larmor precession frequencies of the proton ( $\omega_I$ ) and electron ( $\omega_S$ ), relaxation rates for the interaction between the nuclear spin and the paramagnetic ion are represented quite well by the Solomon-Bloembergen equations given in Eq. 3-2 (165, 166).

Eqs. 3-2

$$\frac{1}{T_1} = R_1 = \frac{2}{15} S(S+1) \frac{(\gamma_I)^2 g^2 \beta^2}{r^6} \left\{ \frac{3\tau_c}{1+(\omega_I)^2(\tau_c)^2} + \frac{7\tau_c}{1+(\omega_S)^2(\tau_c)^2} \right\}$$

$$\frac{1}{T_2} = R_2 = \frac{1}{15} S(S+1) \frac{(\gamma_I)^2 g^2 \beta^2}{r^6} \left\{ 4\tau_c + \frac{3\tau_c}{1+(\omega_I)^2(\tau_c)^2} + \frac{13\tau_c}{1+(\omega_S)^2(\tau_c)^2} \right\}$$

Where  $T_1$  is the spin-lattice (longitudinal) relaxation time,  $T_2$  is the spin-spin (transverse) relaxation time,  $R_1$  is the transverse relaxation rate,  $R_2$  is the transverse relaxation rate,  $S$  is the electron spin number (1/2 for stable nitroxide radicals),  $\gamma$  is the gyromagnetic ratio (proton),  $g$  is the electronic  $g$  factor,  $\beta$  is the Bohr magneton,  $r$  is the distance between the electron and nuclear spins, and  $\tau_c$  is the correlation time. The assumption of a single correlation time is reasonable because in our case the spin label is attached to a relatively

large macromolecule (ARNT PAS-B), so the distinction between individual correlation times that reflect a dependence on the longitudinal ( $\tau_{C1}$ ) and transverse ( $\tau_{C1}$ ) relaxation times of the electron are not necessary for this level of analysis. Also, terms that describe the dependence of  $T_1$  and  $T_2$  relaxation from scalar coupling are not included in Eqs. 3-2, because experiments described in this chapter involve inter-molecular paramagnetic broadening enhancements from the unpaired electron on a nitroxide spin label. Since the unpaired electron and nuclear spin are on different molecules, these terms can be neglected.

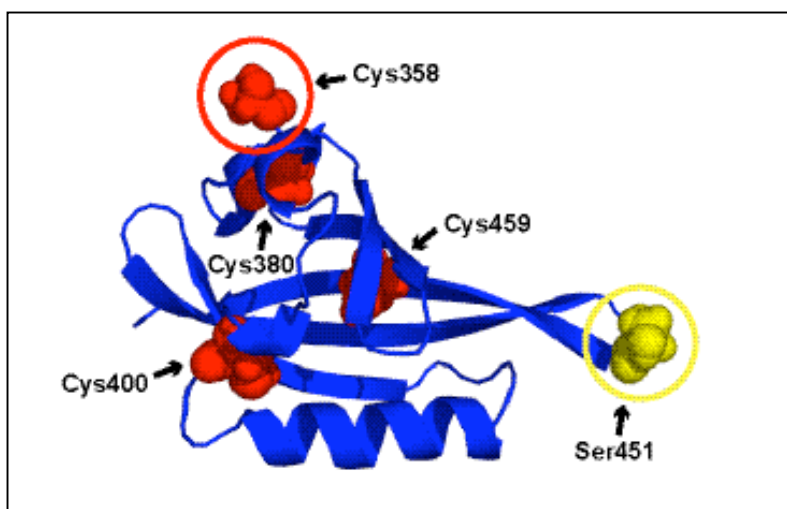
The large magnetic moment of the unpaired electron means that there will be significant effects on the relaxation rates of nuclei that lie within relatively long distances (up to  $\sim 20$  Å) of the spin label. For this reason, spin labels can be used to probe distances between binding partners in hetero- or homodimeric protein associations in cases where normal dipole-dipole interactions between nuclei (NOE) cannot be used because of weak affinities or long distances between the nuclei used as probes for the interaction. Because of the problems encountered in obtaining intermolecular NOE's between the PAS-B domains of HIF-2 $\alpha$  and ARNT in the heterodimeric complex, we therefore decided to use spin labeling to try and confirm the HADDOCK-derived model of the heterodimer.

*b. Spin labeling materials and methods.*

We utilized paramagnetic broadening enhancement to confirm the anti-parallel orientation of the HIF-2 $\alpha$ /ARNT PAS-B complex through thiol-reactive spin labeling reagents that contain a stable nitroxide radical. We began by replacing all four ARNT

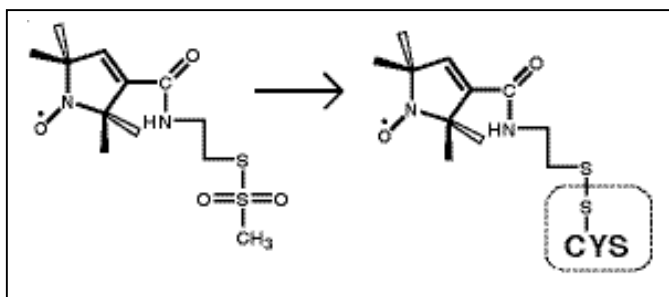
PAS-B wildtype cysteines with serine, and reintroducing single cysteines into two strategic locations on opposite sides of the domain. Constructs containing single cysteine residues were generated using site-directed mutagenesis to convert existing cysteines to serine (Cys358Ser, Cys380Ser, Cys400Ser, Cys459Ser) or to introduce cysteines into a desired location on the opposite end of the molecule (Ser451Cys). For this study, two constructs containing lone cysteines at positions 358 and 451 were used (Fig 3-9). All mutations were confirmed by DNA sequencing over the entire open reading frame, and were found to be properly folded and capable of binding HIF-2 $\alpha$  PAS-B as described above for wildtype ARNT.

**Fig 3-9. Wild type cysteines of ARNT PAS-B that were mutated to serine for the spin labeling experiments are shown in red, with the exception of Cys358 which was retained for labeling in one set of experiments due to its N-terminal location. A wild type serine (Ser451) was also mutated to cysteine in another set of experiments to get obtain site-specific orientational information.**



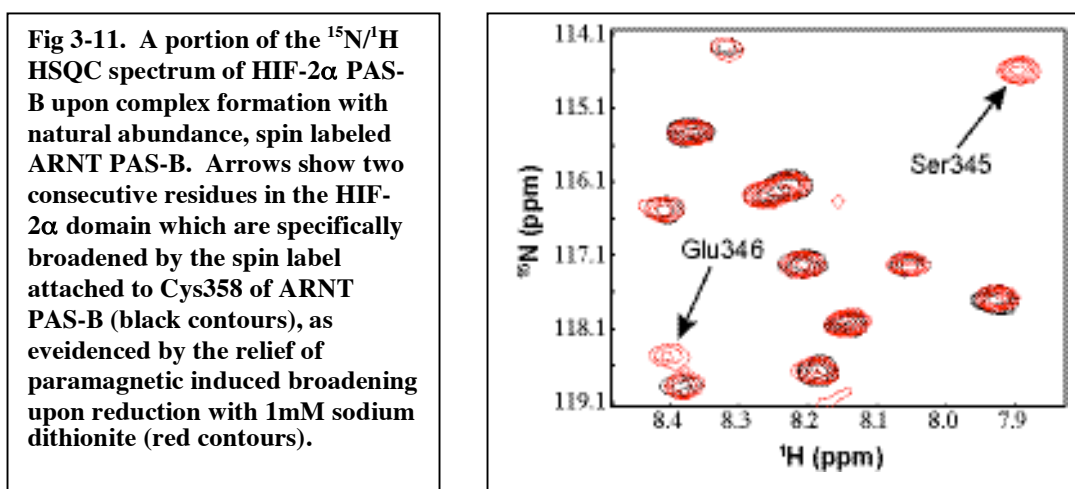
Natural abundance, purified samples of these single Cys-containing proteins were reacted with (1-oxyl-2,2,5,5-tetramethylpyrroline-3-yl) carbamidoethyl methanethiosulfonate (here referred to as CMTSL) (Fig 3-10), a stable nitroxyl radical spin label. CMTSL was added to 100  $\mu$ M samples of ARNT PAS-B mutants at a 0.9:1 CMTSL:protein ratio and incubated overnight at room temperature in 50 mM Tris, 25mM NaCl, pH 7.5. Unreacted spin label and unwanted disulfide-bridged dimers of mutant ARNT PAS-B were subsequently removed by injection over a Superdex75 gel filtration column. Electrospray ionization mass spectrometry performed on protein labeled in initial studies at a higher spin label concentration (3:1 CMTSL:ARNT PAS-B) revealed a high labeling efficiency (~90%), with the majority of non-reacted protein sequestered as disulfide-bridged homodimers (data not shown).

**Fig 3-10. Reaction of the CMTSL spin label with cysteine.**



After spin labeling of natural abundance ARNT PAS-B, complexes were formed between this domain and  $^{15}$ N-labeled HIF-2 $\alpha$  PAS-B. Spin-labeled ARNT PAS-B mutants were added to  $^{15}$ N-labeled HIF-2 $\alpha$  PAS-B at a ratio of 1.5:1 ARNT PAS-B (112.5  $\mu$ M):HIF-2 $\alpha$  PAS-B (75  $\mu$ M) and standard  $^{15}$ N/ $^1$ H HSQC experiments were acquired at 25°C on a cold-probe equipped Varian Inova 600 MHz spectrometer. All experiments were then repeated after reduction of the spin label with 1mM sodium dithionite, adding 5 $\mu$ l of a 100mM stock in DMSO to a 500 $\mu$ l sample under anaerobic

conditions. Comparisons of  $^{15}\text{N}/^1\text{H}$  HSQC spectra recorded on these samples before and after reducing the nitroxyl radical with sodium dithionite showed that most amide resonances in HIF-2 $\alpha$  PAS-B were unaffected by this treatment, but a subset had peak intensities that were selectively lowered in the presence of the unquenched radical (Fig 3-11), consistent with those sites being located near the spin label attached to ARNT PAS-B when complex was formed.



Spectra were processed and analyzed using the programs nmrPipe (149) and nmrView (150). Peaklists were manually inspected after peak picking, omitting peaks with any significant spectral overlap from further evaluation (30/114 peaks omitted). Peak intensities were obtained within nmrView to determine the maximum peak heights in both the oxidized (paramagnetic) and reduced (diamagnetic) spectra. Ratios of the oxidized/reduced spectra were then calculated to obtain quantitative measurements of peak broadening due exclusively to the paramagnetic spin label. Amide signals that were broadened beyond detection in the oxidized spectra were assigned a ratio value of zero.



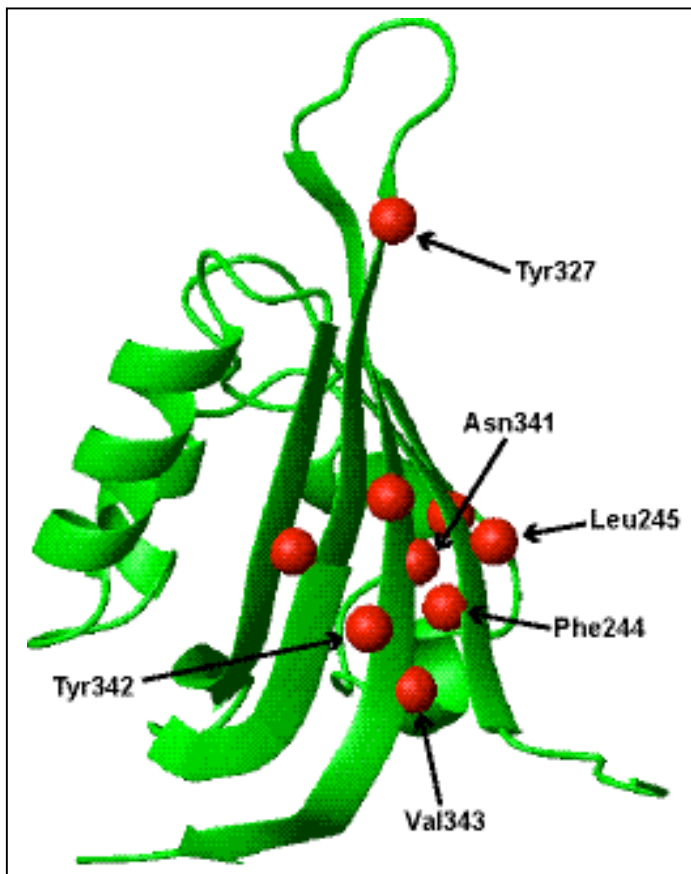
*c. Analysis of paramagnetic broadening enhancement data.*

By examining the ratio of HIF-2 $\alpha$   $^{15}\text{N}/^1\text{H}$  HSQC peak intensities under standard and quenched conditions for HIF-2 $\alpha$ /ARNT PAS-B complexes assembled with two different single spin-labeled variants of ARNT, we were able to clearly identify HIF-2 $\alpha$  amides that were located near each of the spin label locations. These analyses also revealed several amides that were perturbed regardless of the site of spin label attachment on ARNT, suggesting non-specific interactions between HIF-2 $\alpha$  and the spin label. These effects were predominantly localized in a hydrophobic patch near residues Phe244, Leu245, Asn341, and Tyr342, Val343 and a single tyrosine residue at a relatively distant location, Tyr327. Similar non-specific effects have been observed elsewhere, presumably due to the hydrophobic nature of PROXYL spin labels (167).

To confirm that this non-specific paramagnetic broadening was in fact a result of interactions between the spin label and the HIF-2 $\alpha$  PAS-B domain and not an indication of a non-specific PAS domain interaction, free L-cysteine was reacted with the spin label under similar conditions as the labeling of ARNT PAS-B, but with 2-fold excess L-cysteine present to reduce the presence of free spin label still in solution. This mixture was then added to  $^{15}\text{N}$  HIF-2 $\alpha$  PAS-B, and  $^{15}\text{N}/^1\text{H}$  HSQC experiments were performed under oxidizing and reduced conditions to see if similar residues were affected that corresponded to the residues that are being non-specifically broadened upon complex formation. The results are shown in Fig 3-12, where residues that show ox/red ratios less than 0.65 are displayed as red spheres. The hydrophobic patch and the single tyrosine residue described earlier (Tyr327) are being affected by spin-labeled free cysteine in a similar manner than upon complex formation with spin-labeled ARNT PAS-B, indicating

that direct interactions between the spin label and HIF-2 $\alpha$  PAS-B are the source of the non-site specific paramagnetic broadening.

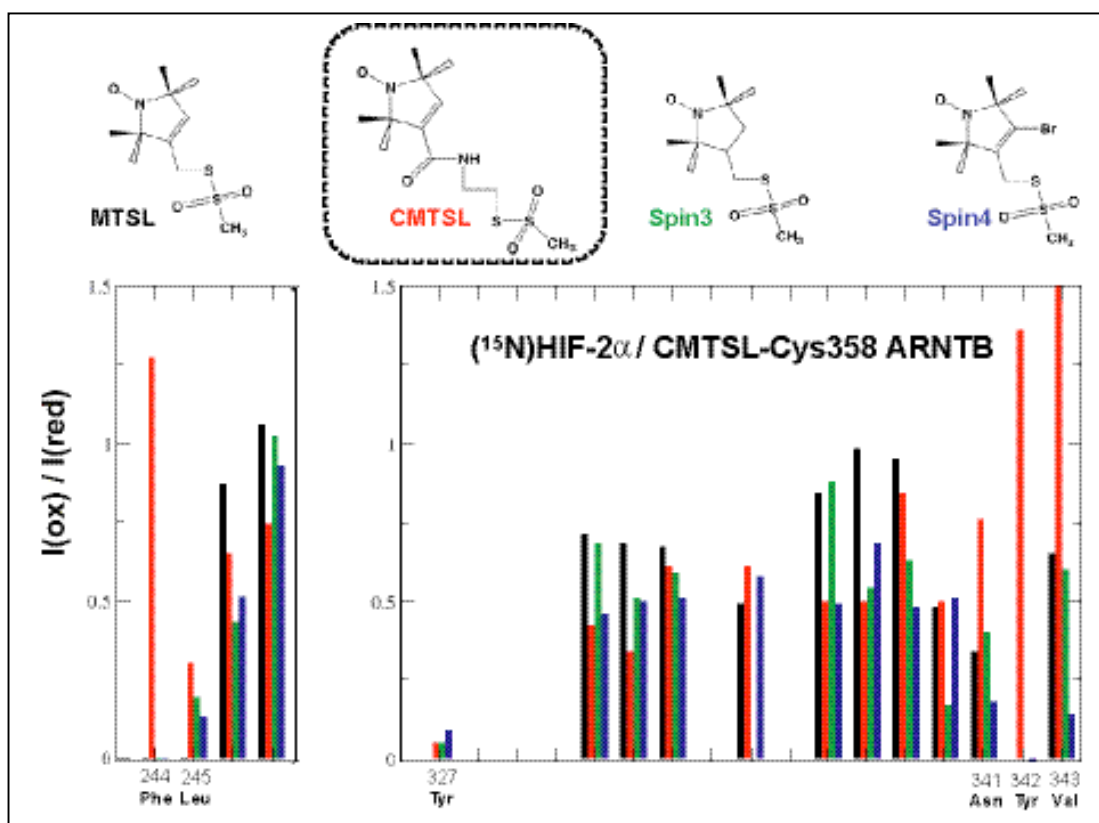
**Fig 3-12. Ribbon diagram of the HIF-2 $\alpha$  PAS-B domain, with backbone amide resonances that show significant L-Cys-CMTSL dependent broadening ( $I_{ox}/I_{red} < 0.65$ ) in the control experiments described in the text.**



We first addressed this problem by evaluating several different spin labels of similar type to minimize the extent of non-specific broadening effects. Ratios of oxidized vs. reduced backbone amide resonances from HIF-2 $\alpha$  PAS-B residues affected by spin labeled free cysteine in the control experiments were analyzed after complex was formed with ARNT PAS-B spin labeled with purchased variants of MTSL (Fig 3-13). Based on the results, CMTSL was decided on as an alternative over the more commonly used MTSL spin label. Compared to MTSL, CMTSL contains an additional amide group in

the linker between the protein and the PROXYL ring, introducing a degree of polarity into an otherwise hydrophobic compound. This reduced, but did not completely eliminate, the non-specific broadening we observed in the hydrophobic surface patch on the HIF-2 $\alpha$  PAS-B  $\beta$ -sheet (Fig 3-12).

**Fig 3-13. Alternative spin labels reduce the non-specific broadening effects seen for MTSL.** The use of the longer, more polar linker in the CMTSL spin label reduces the interaction of spin label attached to free L-Cys with the HIF-2 $\alpha$  PAS-B domain. This is seen by an increase in the ratios of the peak intensities of affected backbone amide resonances in the oxidized and reduced state when CMTSL is used (red bars) relative to the three other spin labels tried (black, green and blue bars).



To control for any residual non-specific interactions between the spin label and HIF-2 $\alpha$  PAS-B, we developed a new comparative analysis based on the ratios of the peak intensities of the oxidized and reduced state spectra ( $I_{ox}/I_{red}$ ) of two different complexes. In this analysis, these standard ratios obtained from complexes formed with spin-labeled ARNT PAS-B at Cys358 and Cys459 were divided again to give information on paramagnetic broadening effects that are specific for individually labeled sites according to equation 3-2:

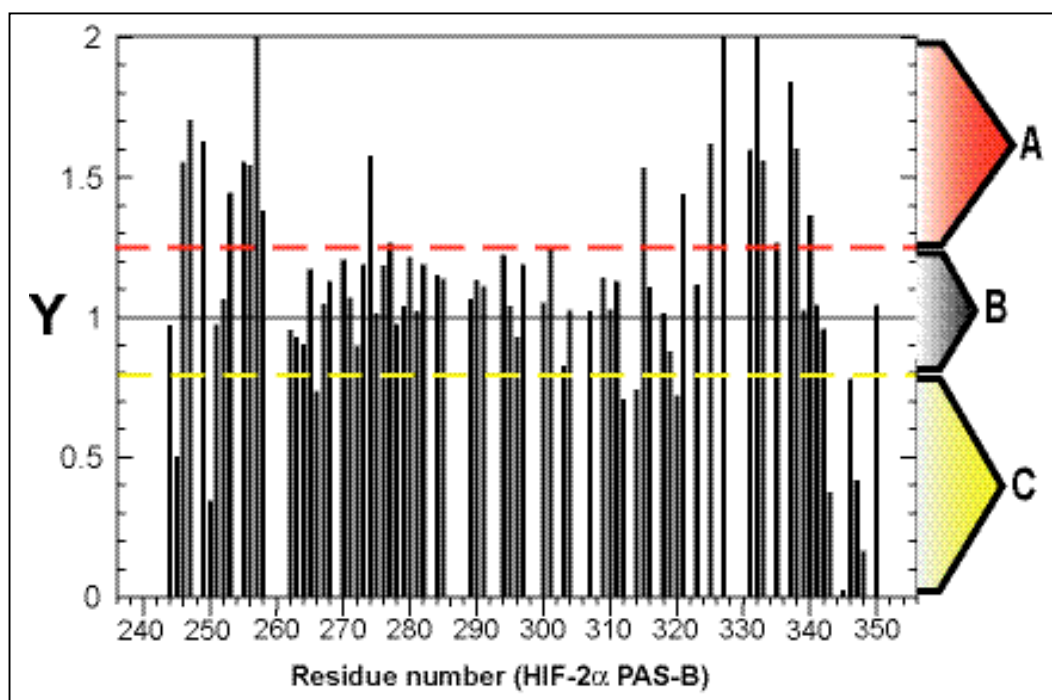
**Equation 3-2:**

$$Y = \frac{I_1(ox) / I_1(red)}{I_2(ox) / I_2(red)}$$

**Where:**  $I_1$ = HIF-2 $\alpha$  PAS-B peak intensity with C451 of ARNT PAS-B spin labeled  
 $I_2$ = HIF-2 $\alpha$  PAS-B peak intensity with C358 of ARNT PAS-B spin labeled

We also analyzed the NMR peak intensity changes as a ratio of the effects observed in the two complexes, each of which contained ARNT labeled with CMTSL at opposite ends of the domain. By taking the ratio of the ratio of spin label-mediated changes in peak intensities from these complexes, non-specific effects were readily identified by yielding a value near one (Fig 3-14).

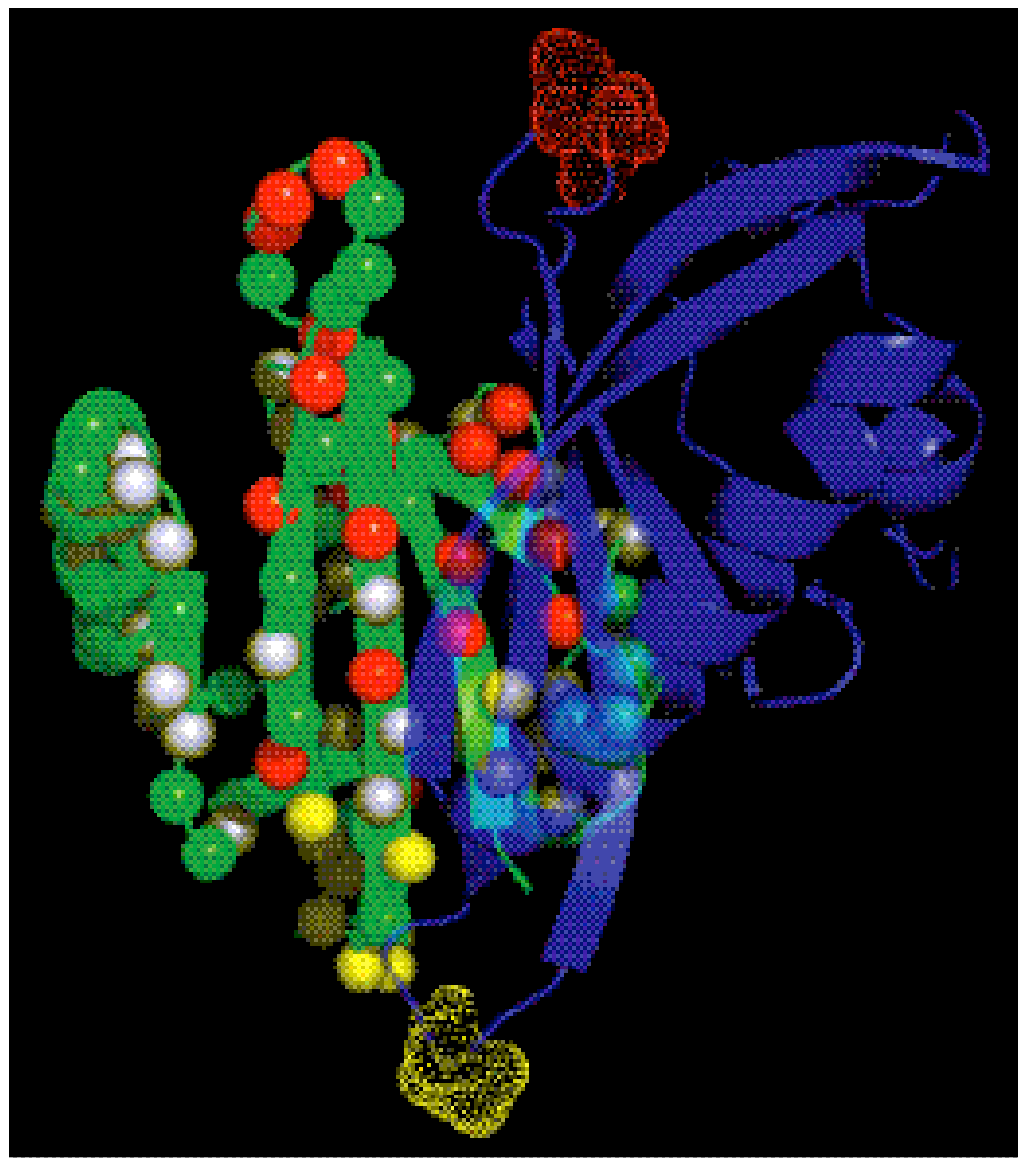
**Fig 3-14.** Alternative analysis used to control for non-specific interactions between the spin label (CMTSL) and HIF-2 $\alpha$  PAS-B. The ratio of ratios approach described in the text is interpreted graphically by plotting the values obtained from applying equation 3-2 (Y) against residue number. Resulting values from the analysis that reside in the red area (A) indicate residues specifically affected by spin label attached to ARNT PAS-B-Cys358, and values in the yellow area indicate those affected by the spin label on the opposite end (Cys451). Values in the grey region represent residues that are either not affected by the spin label or similarly affected regardless of labeling site.



Using these approaches, we prepared HIF-2 $\alpha$ /ARNT PAS-B complexes labeled with CMTSL at residues 358 and 451 (Fig 3-9), and mapped any residues showing significant site-specific broadening effects onto the HADDOCK-generated model of the complex (Fig 3-15). This analysis clearly confirms that the predominant solution form of this complex has the PAS-B domains in an antiparallel arrangement, placing HIF-2 $\alpha$  residues distant from the N- and C-termini near the ARNT termini and vice-versa. To further test the consistency of this model with regard to our spin label data, the

HADDOCK calculation was repeated with site-specific paramagnetic broadening effects incorporated as loose distance restraints. This was achieved by restricting the separation between the gamma position of a CMTSL-attached residue to be less than 25Å from the backbone amide protons of the residues most specifically affected by spin label attachment. The ensemble of HADDOCK-generated structures (not shown) was as well defined as the original set obtained without spin label restraints (average pairwise r.m.s.d. = 1.14Å with restraints, 1.21Å without). More importantly, the inclusion of these restraints did not significantly alter the complex structure, as evidenced by an equivalently low pairwise backbone r.m.s.d. between members of the two ensembles (r.m.s.d. = 1.15Å). These findings firmly establish the accuracy of the modeled complex at low resolution by direct comparison with data obtained from an independent approach.

**Fig 3-15. Orientation of HIF-2 $\alpha$ /ARNT PAS-B complex, shown in a view that is approximately 60 degrees rotated clockwise about the vertical axis as compared to Figure 3-4. Backbone amide proton resonances of HIF-2 $\alpha$  PAS-B that were significantly broadened by spin labeled ARNT PAS-B according to the analysis detailed in the text are shown as spheres, with red sites specifically broadened by spin label at C358 and yellow sites broadened by spin label at C451. Grey spheres designate residues that were equivalently affected by spin label at both locations, and green spheres indicate sites that were not included in the analysis due to spectral overlap.**



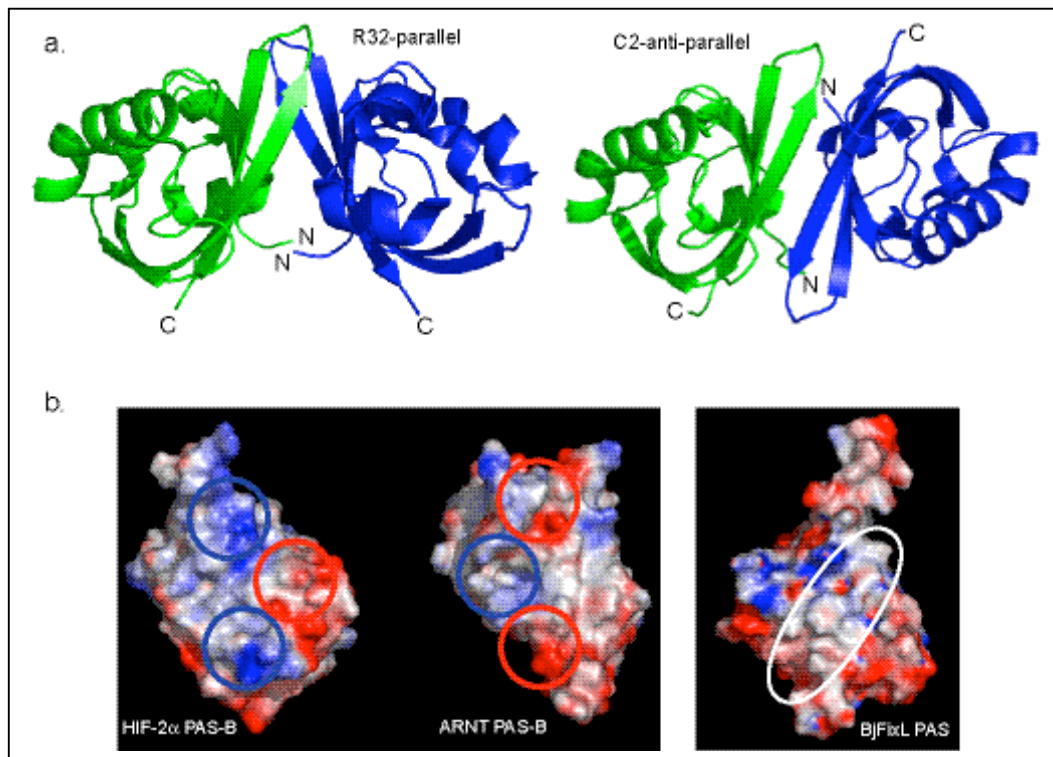
### iii. Relevance to other systems

It should be noted that homodimeric PAS domain crystal structures have been solved that reveal significant variations on the observed  $\beta$ -sheet interaction of our antiparallel HIF-2 $\alpha$ /ARNT PAS-B complex. For example, the structure of the PAS domain homodimer from the bacterial chemotaxis sensor EcDOS shows an orthogonal association with respect to the  $\beta$ -strands (44), and the PAS homodimer from the FixL nitrogen fixation regulatory protein of *Sinorhizobium meliloti* (SmFixL) demonstrates an essentially parallel interaction with respect to the N- and C-termini (6). However, the PAS domain subunits that comprise these homodimers include N-terminal  $\alpha$ -helical regions that significantly contribute to the buried surface area. Although N-terminal helices are predicted to precede the PAS-A domains of both HIF-2 $\alpha$  and ARNT, they are conspicuously absent from the PAS-B domains by secondary structure analysis.

Interestingly, crystal structures of the homodimeric FixL PAS domain from another organism (*Bradyrhizobium japonicum* - BjFixL) reveal either a parallel or an antiparallel association depending on the conditions used for crystallization (Fig 3-16 a) (25). Examination of the BjFixL  $\beta$ -sheet interface shows that it is somewhat neutral compared to the PAS-B domains from HIF-2 $\alpha$  and ARNT, suggesting that it might lack the electrostatic complementarity that could provide the specificity for the anti-parallel HIF-2 $\alpha$ /ARNT PAS-B interaction (Fig 3-16 b). In addition, these constructs lacked the N-terminal helices corresponding to those involved in the SmFixL homodimer, which suggests that the presence or absence of these helical elements might provide the basis for specific modes of PAS-PAS interactions in some systems, both in isolated domains and within the context of full-length proteins.



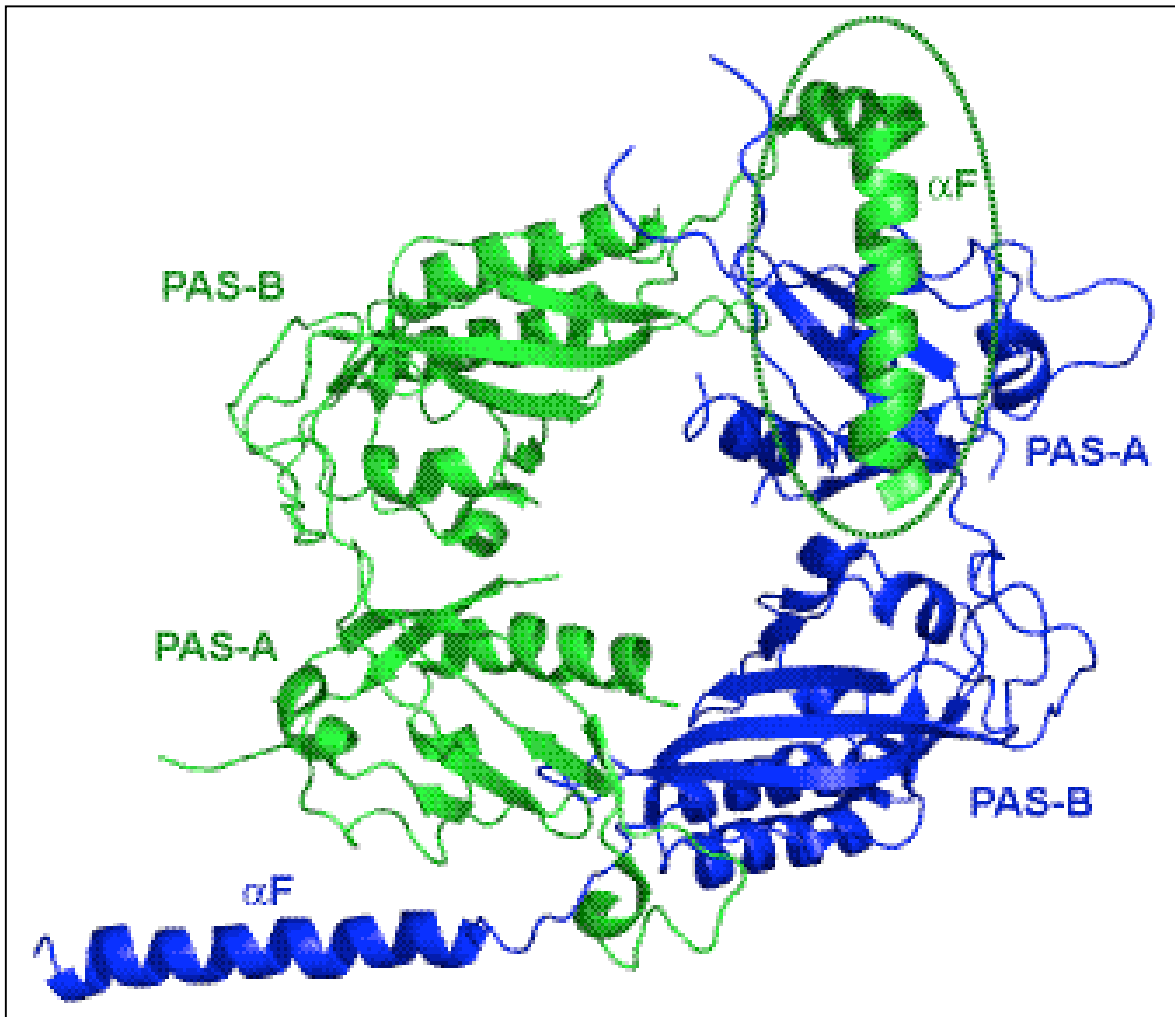
**Fig 3-16. BjFixL crystallizes in two orientations. a.** The crystal structure of BjFixL in the R32 crystal form shows a parallel homodimerization between the monomer units (green and blue), and the C2 form shows an anti-parallel association (6). **b.** Electrostatic surface potential mapping of the BjFixL structure reveals a somewhat more neutral, hydrophobic  $\beta$ -sheet surface (white circle) compared to that of HIF-2 $\alpha$  PAS-B and ARNT PAS-B.



A recent investigation that sheds some light on potential interactions between multi-PAS proteins involves the homodimeric crystal structure of a tandem PAS domain construct from the *Drosophila* period (Per) protein. In this case, the homodimer exhibits reciprocal interactions between the PAS-A domain of one monomer with PAS-B of the other chain (5). Intriguingly, a key set of interprotein contacts used to hold the homodimer together is provided by an additional element located outside of the canonical PAS domain. In this case, an  $\alpha$ -helical region C-terminal to the PAS-B domain of one monomer (named  $\alpha$ F by Yildiz *et al.*(5)) makes extensive contact with the exposed  $\beta$ -

sheet face of the PAS-A domain of the other subunit (Fig 3-17). This interaction, which resembles the intramolecular helix/sheet interactions seen in the photosensory AsLOV2 and PYP PAS domains (36, 168), is essential for stable Per homodimerization. This is demonstrated by the abilities of either point mutations in the  $\beta$ -sheet (V243D, per<sup>L</sup>) or removal of the  $\alpha$ F helix to significantly weaken the homodimer affinity of Per (5).

**Fig 3-17. Crystal structure of the Per homodimer (5). The homodimer is characterized by reciprocal PAS-A/PAS-B interactions, one of which involves interactions between an  $\alpha$ -helical element ( $\alpha$ F) C-terminal to the PAS-B domain of one monomer that makes extensive contact with the  $\beta$ -sheet surface of the PAS-A domain from the other monomeric subunit.**

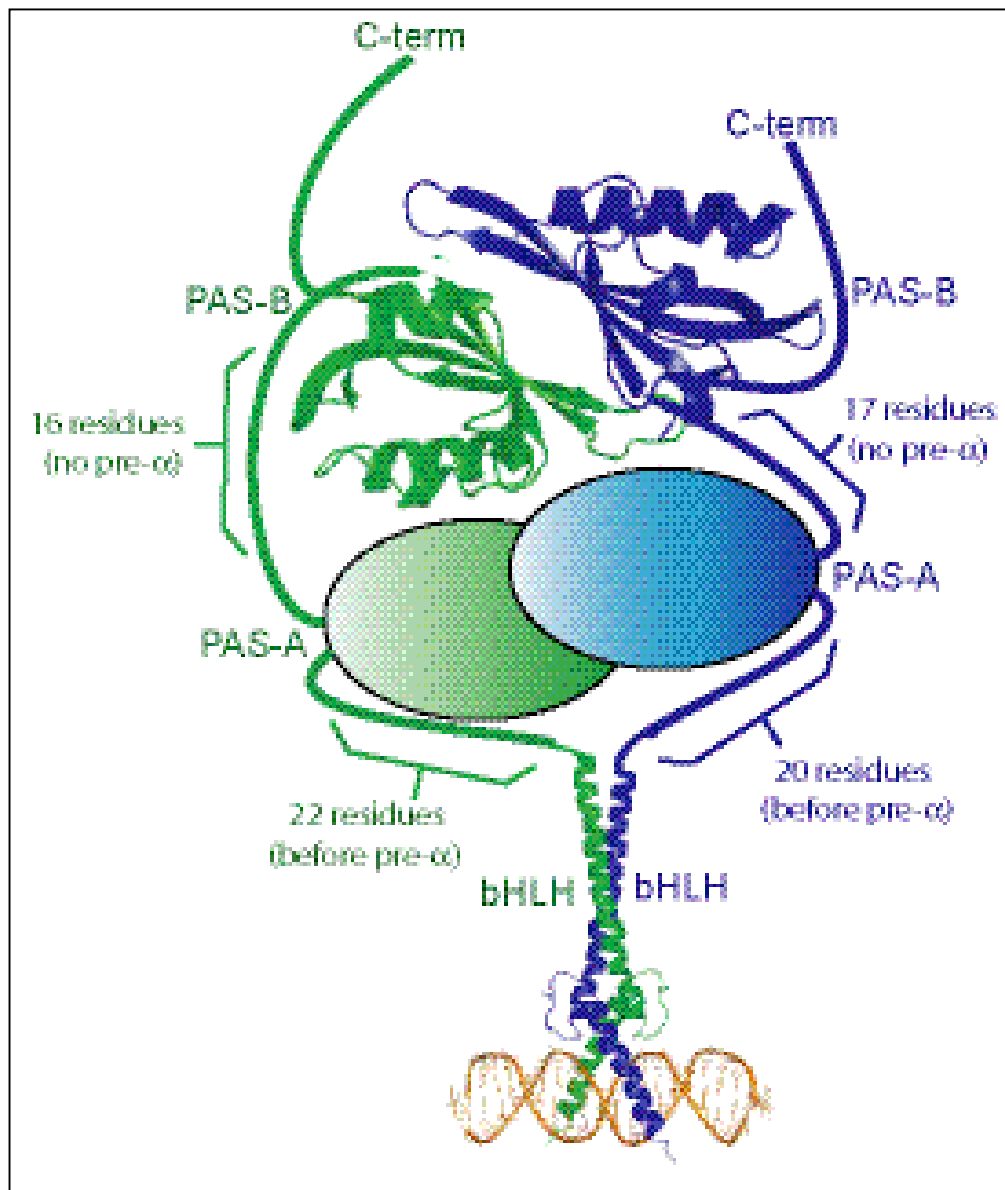


While the Per homodimer structure provides an interesting view of how a tandem PAS domain protein may associate, it should be emphasized that the heterodimeric partner of Per is the circadian clock protein Timeless (TIM), which does not contain a PAS domain (169, 170). Consistent with this, several pieces of data also suggest that HIF- $\alpha$  and ARNT interact in a different manner. If the  $\alpha$ F helix/ $\beta$ -sheet interaction interface is neglected, the buried surface area of the tandem PAS-A/B heterodimer of Per (2390 Å<sup>2</sup>) is similar to that of the HIF-2 $\alpha$ /ARNT PAS-B heterodimer alone (2205 Å<sup>2</sup>), and the interaction affinities of the two complexes are correspondingly similar ( $K_d$  ~30  $\mu$ M), despite the much smaller size of the HIF-2 $\alpha$ /ARNT PAS-B complex. Additionally, sequence alignments of ARNT show that the region where an  $\alpha$ F helix would expect to be found C-terminal of PAS-B are not conserved (Fig 3-2) between ARNT and its functionally redundant ARNT2 homolog (171), and is also not predicted to be helical by secondary structure prediction algorithms (data not shown). The Per PAS-B domain also makes critical intermolecular interactions using a Trp residue in the HI loop, which has been replaced with an absolutely invariant tyrosine in ARNT (Fig 3-2). Overall, these variations in binding partner preference, secondary structure content, and sequence conservation suggest that bHLH-PAS proteins could interact differently than other PAS domain-containing systems.

These differences between Per and ARNT are not surprising in light of the relationships that their PAS domains presumably have with additional domains in full-length proteins. Notably, Per lacks bHLH DNA binding domains, relieving it of topological restraints imposed on bHLH/PAS factors by their need to form DNA-bound dimer. This observation may have significant implications for the nature of

domain/domain interactions found in functional protein complexes. This could include the antiparallel organization of the HIF-2 $\alpha$ /ARNT PAS-B complex, which came somewhat of a surprise in light of two-dimensional models (Fig 3-8) and the relatively short linkers found between domains in bHLH/PAS proteins. However, these antiparallel associations are easily accommodated by the available linker lengths in a complex that could simultaneously engage all four PAS domains within the heterodimer (Fig 3-18). The PAS domains could form a cooperative, intermolecular "tetramer" of PAS components that interact to convert an inherently weak set of protein/protein and protein/DNA interactions into a stronger binding force that permits the formation of a functionally viable complex. Further investigations of larger constructs of PAS domain containing proteins that include additional structural elements will undoubtedly be required to fully understand these interactions, and will be instrumental in designing strategies for artificially controlling biological pathways for the treatment of a wide variety of diseases that involve transcriptional misregulation as an important part of their pathology.

**Fig 3-18. New model for the interaction of HIF-2 $\alpha$  and ARNT PAS domains in the context of the functional transcription factor.** The model shows a potential cooperative interaction between the PAS domains of both proteins that could explain the anti-parallel orientation, and provide additional affinity and/or specificity for the complex. The bHLH domain binding the hypoxia response element (HRE) is represented by the crystal structure of the Max protein binding its cognate DNA (7), followed by a schematic representation of a putative PAS-A/PAS-A interaction. The antiparallel HIF-2 $\alpha$ /ARNT interaction is shown by the representative structure from our HADDOCK calculations.



#### IV. ARNT PAS-B Homodimerization and Mutagenesis

##### *A. ARNT PAS-B homodimerization*

While the homodimerization of full length ARNT has been well-documented *in vitro*, the biological significance of this interaction is not yet understood. ARNT is constitutively localized in the nucleus, and shows widespread expression in mammalian tissues (3, 79). Electrophoretic mobility shift assays have demonstrated that a baculovirus-expressed ARNT homodimer binds an E-box element with a CACGTG consensus sequence (81) and can activate transcription from promoters containing this element in transient transfection assays (82, 83). The physiological significance of this has not been fully explored, however, and there is no structural information that distinguishes between this type of association and its heterodimeric interactions with other bHLH-PAS transcription factors.

A primary goal was to see if the  $\beta$ -sheet surface used for the interaction between the ARNT and HIF-2 $\alpha$  PAS-B domains could also facilitate homodimeric interactions between ARNT PAS-B domains. The promiscuous binding behavior of the ARNT protein has raised many questions regarding the biophysical basis for the multiple specificities exhibited by this protein in the formation of many different transcriptional activation complexes. Another goal of this investigation was to use a reductionist approach to determine the possible basis for the multiple binding partners of ARNT through the biophysical characterization of complexes involving the PAS-B domain from this protein. With the heterodimeric complex between the HIF-2 $\alpha$ /ARNT PAS-B domains solved to low resolution, the next obvious step was to determine any similarities

and/or differences between this complex and the putative ARNT PAS-B homodimer. Complementing a prior demonstration of ARNT bHLH domain homodimerization (80), results outlined below show that ARNT PAS-B weakly self-associates in a concentration-dependent manner. The role of these weak interactions may be minimal for DNA-bound ARNT homodimers, as indicated by a recent report showing that neither of the ARNT PAS domains contribute significantly to the DNA-binding affinity of ARNT homodimers (146). It should be noted that these findings are exclusively based on DNA gel shift experiments, preventing the parsing of the observed  $K_d$  into separate components for protein dimerization and DNA binding (as noted by the authors).

As such, we suggest that these interactions may also play a role in maintaining ARNT homodimeric complexes independent of DNA, perhaps when there are no heterodimerization partners of ARNT available at high concentrations in the nucleus. This could serve to keep ARNT sequestered in a more stable form until one of its heterodimeric binding partners arrive to activate transcription, but more information will be needed to discern the binding characteristics of the variable complexes in which ARNT participates.

#### *i. Characterization of ARNT PAS-B homodimer*

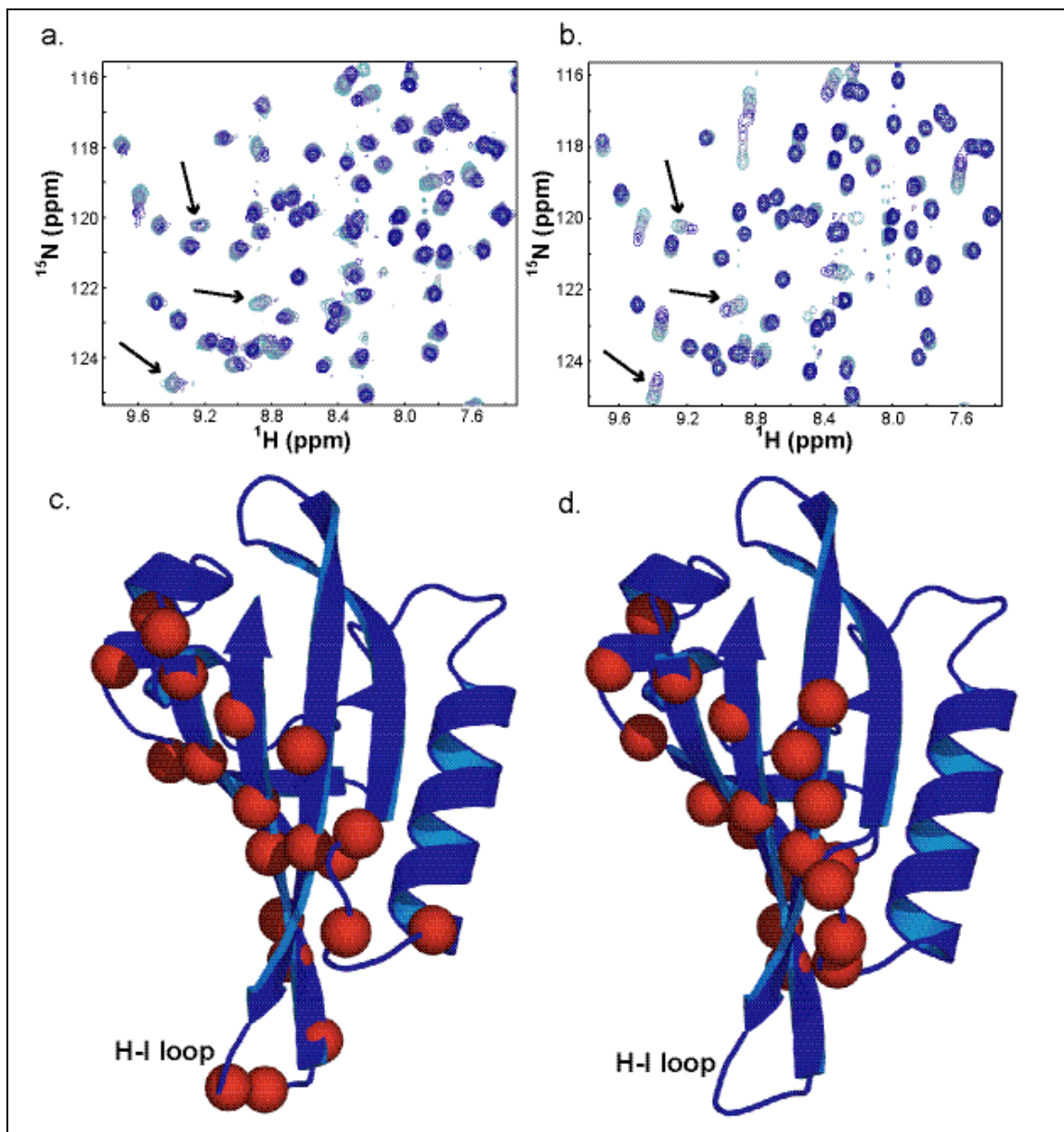
Homodimer formation of ARNT PAS-B was identified by serial two-fold dilutions of a 1mM sample of  $^{15}\text{N}$ -labeled protein to a final concentration of 125  $\mu\text{M}$  in four steps, while increasing the number of scans in concurrent  $^{15}\text{N}/^1\text{H}$  HSQC experiments to control for the reduction in protein concentration. Significantly perturbed resonances

of ARNT PAS-B for which  $\Delta\delta_{\text{TOT}} > 0.03$  (homodimer) were then mapped onto the structure in order to identify binding interfaces involved in homodimer formation.

In contrast to the case for the heterodimer, which exhibited both fast and intermediate exchange, the ARNT dilution series demonstrated only fast exchange behavior, consistent with a lower affinity interaction (Fig 4-1 b). Although full saturation was not achieved due to the high protein concentrations required, titration curves for the interactions indicate an association that is approximately 15-fold weaker than that for the heterodimer ( $K_d \sim 500 \mu\text{M}$ ). Mapping the ARNT PAS-B residues perturbed by hetero- and homodimerization onto the solution structure shows a very similar group of residues distributed fully across the  $\beta$ -sheet being affected (Fig 4-1 c,d), indicating that a comparable interface is being used for both complexes.



**Fig 4-1. ARNT heterodimerizes and homodimerizes using a similar  $\beta$ -sheet interface.** a. Titration of unlabeled HIF-2 $\alpha$  PAS-B (0mM to 1mM, light to dark blue) into 200 $\mu$ M  $^{15}$ N ARNT PAS-B. The interaction results in many peaks being broadened due to intermediate exchange events. b. Serial dilution series of  $^{15}$ N ARNT PAS-B (0mM to 1mM, light to dark blue). Unlike the heterodimer, homodimerization is characterized by fast exchange, evidenced by progressive chemical shift changes during the dilution. c. ARNT PAS-B resonances that are significantly perturbed (shifted ( $\Delta\delta \geq 0.05$  ppm) or broadened) upon addition of unlabeled HIF-2 $\alpha$  PAS-B are mapped onto the high resolution structure of ARNT PAS-B, which clearly identifies the solvent-exposed  $\beta$ -sheet surface as the interaction interface. d. A very similar interface is observed for the ARNT PAS-B homodimer when significantly shifted resonances ( $\Delta\delta \geq 0.03$  ppm) from the serial dilution are mapped onto the structure.



Although our data clearly suggests that the ARNT PAS-B homodimer also utilizes the exposed  $\beta$ -sheet of the domain, it seems unlikely that such a complex would adopt the same orientation as the heterodimer given the resulting electrostatic repulsion that would occur as suggested by Fig 3-5. In addition, the conserved residues in the HI loop of ARNT PAS-B also show a greater degree of chemical shift perturbation in the HIF-2 $\alpha$ /ARNT heterodimer compared to the ARNT PAS-B homodimer (Fig 4-1 c,d), suggesting that they could be more involved in the interaction. This observation, in addition to the lack of electrostatic complementarity, could help explain the lower affinity of the homodimeric complex.

### *B. ARNT PAS-B Mutagenesis*

During the study of the heterodimeric complex between the PAS-B domains of HIF-2 $\alpha$  and ARNT, the interpretation of data collected from several experiments was complicated by the multiple equilibria present due to homodimeric associations in addition to heterodimerization. For this reason, it was decided that mutants designed to disrupt the various interactions could aid further structural characterization on the ARNT PAS-B domain. Future investigations that could benefit from establishing a constitutively monomeric PAS subunit include the characterization of potential intramolecular associations with other domains in ARNT, which would presumably involve a different binding interface than that used for homo-/heterodimerization, and the identification and structural characterization of ligand binding sites within the core of the ARNT PAS-B domain. Preliminary efforts to meet these goals will be discussed in following chapters. It was also recognized that another potential benefit of mutagenesis

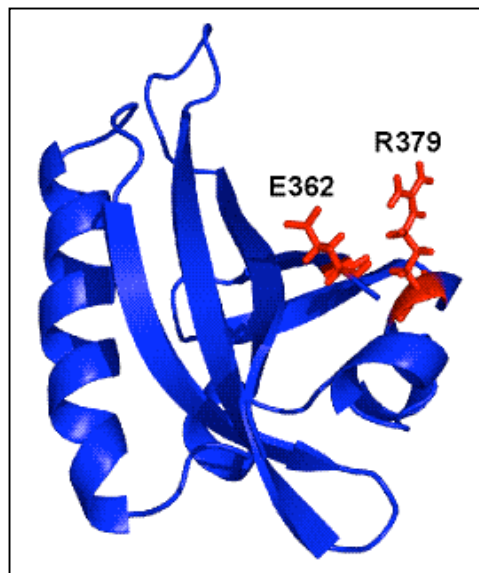
studies on ARNT PAS-B could be the discovery of variations in the biophysical properties of particular residues or groups of residues at the binding interface that could provide a basis for the promiscuous binding behavior of ARNT in the variety of biological pathways discussed in the introduction.

*i. Generation of ARNT PAS-B mutants*

Mutagenesis studies were performed on the PAS-B domain of ARNT in a manner analogous to the previous efforts on HIF-2 $\alpha$  PAS-B (8) in order to identify mutations that disrupt hetero- and homodimeric associations of ARNT PAS-B, and to further study the mode of PAS domain interactions. Mutations were made using the Stratagene QuickChange<sup>®</sup> II XL Site-Directed Mutagenesis Kit in a similar manner to that used for the cysteine mutagenesis in the spin label-based studies described in Chapter III. It should be noted that most sample preparations of mutant variants of ARNT PAS-B were done by Daniel Buehler and Matthew Evans, and their help on this project was greatly appreciated. The selection of residues for mutagenesis was based on chemical shift information which identified amino acids with solvent-exposed side chains on the central  $\beta$ -sheet that were potentially involved in the homo- and heterodimeric interactions of ARNT PAS-B. Targeted residues included Glu362, Arg366, Phe444, and Tyr456. Mutagenesis of Glu362 to arginine, designed to alter the electrostatic charge of part of the central  $\beta$ -sheet, resulted in a completely insoluble protein, most likely because of aggregation due to unfolding induced by electrostatic repulsion with another nearby arginine, Arg379 (Fig 4-2). Although other mutations were made on Arg379 and Arg366 that resulted in a folded, well-behaved domain, these did not significantly alter the

binding properties of ARNT PAS-B to HIF-2 $\alpha$  PAS-B so they will not be discussed further.

**Fig 4-2. Representative solution structure of ARNT PAS-B, with location of E362 shown relative to R379. Mutation of E362 to Arginine results in an insoluble protein, most likely due to electrostatic repulsion from R379 disrupting the folding of the domain.**

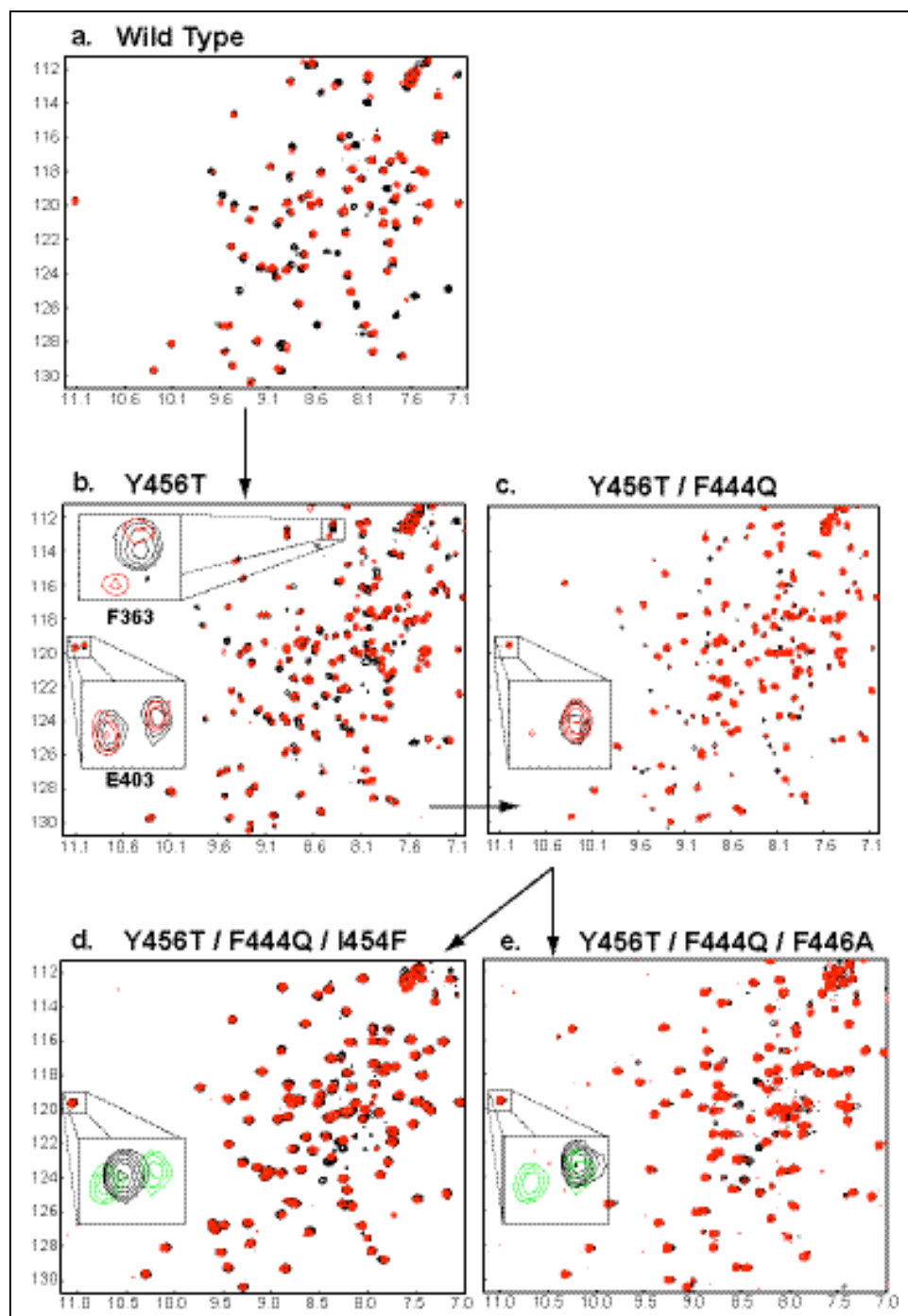


#### *ii. Multiple conformations of Y456T mutation*

Interestingly, another mutation (Y456T) resulted in the doubling of many ARNT PAS-B signals, indicating not only that two conformations of the domain were present, but also that they were in slow exchange (Fig 4-3 b). This is best illustrated by the presence of two very downfield peaks at approximately 11.0 – 11.1 ppm that both represent backbone amide resonances of Glu403 in the two conformations (Fig 4-3 b – lower insert). This extremely downfield peak has been observed in several other PAS domains studied in our lab, and is most likely due to a single backbone amide proton being positioned at the very end of the macrodipole created by the relatively long F $\alpha$  helix. In addition, the two peaks originating from the backbone amide of this residue showed similar peak intensities and peak volumes, indicating that the two conformations exist in approximately equimolar amounts under the sample conditions used. It should

also be noted that this residue is located far from the site of mutation ( $\sim 26$  Å) (Fig. 4-5), so a change in the chemical shift of this residue most likely indicates a global change in the conformation of this domain. Within the spectrum of this variant, many peaks appeared to correspond very well in their chemical shifts with those of the wild type protein. There was also evidence that peaks corresponding to the more wild type conformation continued to shift and/or broaden upon the addition of natural abundance HIF-2 $\alpha$  PAS-B, whereas some peaks representing the new conformation showed no changes upon addition of the HIF PAS domain (Fig. 4-3 b – upper insert). This suggested that the new conformation might not be binding its heterodimeric partner, and gave us reasonable cause to continue mutating residues in this region of the protein.

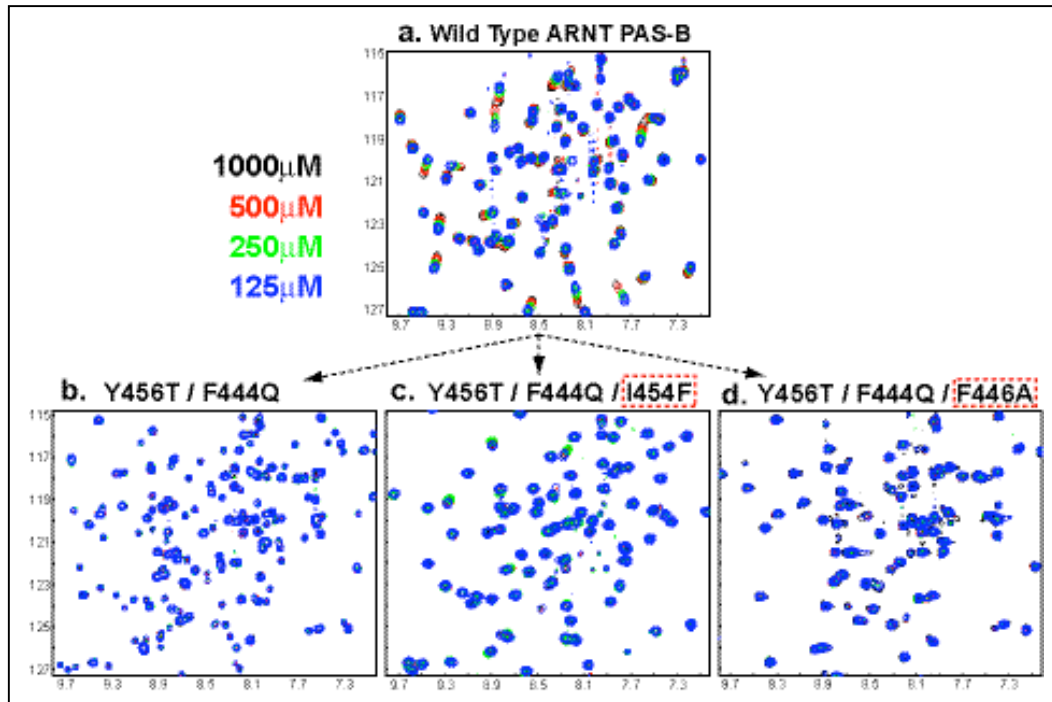
**Fig 4-3.  $^{15}\text{N}/^1\text{H}$  HSQC spectra of various mutants of ARNT PAS-B.** a. Spectra of 250 $\mu\text{M}$   $^{15}\text{N}$ -ARNT PAS-B with (red) and without (black) 750 $\mu\text{M}$  HIF-2 $\alpha$  PAS-B. b. Spectra of Y456T mutant of ARNT PAS-B (200 $\mu\text{M}$ ) with (red) and without (black) 600 $\mu\text{M}$  HIF-2 $\alpha$  PAS-B. Two conformations of the protein are evident (E403 insert). c. Same experiment as b. after an additional mutation (F444Q). The equilibrium is now shifted toward the new conformation (insert). d. Spectra of one triple mutant of ARNT PAS-B (Y456T/F444Q/I454F - 125 $\mu\text{M}$ ) with (red) and without (black) 660 $\mu\text{M}$  HIF-2 $\alpha$  PAS-B. A single, non-binding conformation is present, as evidenced by the insert which shows an overlay of the E403 resonance from the Y456T mutant alone (green) with that of the triple mutant (black) e. Same experiment as in d, but with another mutation (F446A) substituted for I454F. Again, a single conformation that no longer binds HIF-2 $\alpha$  PAS-B is observed.



*iii. F444Q mutation shifts the equilibrium of the conformational transition*

Our next goal was to shift the equilibrium of this slow exchange between discrete structural states toward the new, non-binding conformation. When a second mutation (F444Q) was introduced, this equilibrium was shifted significantly (~90%) to the new conformation, as evidenced by a shift in the relative peak intensities originating from both forms of the domain (Fig. 4-3 c). Upon addition of natural abundance HIF-2 $\alpha$  PAS-B, significant disruption of the heterodimer was evident for the most prevalent conformation (Fig. 4-3 c). In addition, these mutations completely eliminated ARNT PAS-B homodimerization in serial dilution experiments similar to those used to identify the homodimer interface (Fig. 4-4 a). We suggest that both mutations result in the relief of steric hindrance allowing a greater range of motion of the HI loop with respect to the central  $\beta$ -sheet, and that this effectively disrupts the wild-type  $\beta$ -sheet binding surface of the ARNT PAS-B domain.

**Fig 4-4. Mutations effectively disrupt ARNT PAS-B homodimerization. a. Serial dilution of wild type ARNT PAS-B shows significant chemical shift changes. Double (Y456T/F444Q – b.) and triple (Y456T/F444Q/I454F – c. and Y456T/F444Q/F446A – d.) mutations of ARNT PAS-B significantly disrupt self-association as evidenced by similar serial dilution experiments. The double mutant (Y456T/F444Q) shows extensive peak doubling, indicative of slow exchange between two conformations. Unlike initial indications for the heterodimer, however, where only the new conformation showed significant disruption of the interaction, both conformations appear to no longer homodimerize. This is likely due to the weaker association constant of the homodimer relative to the heterodimer.**



*iv. Further shifting of the equilibrium*

To test this hypothesis, we designed further mutations in an attempt to shift the equilibrium entirely to the new non-binding conformation. If disruption of the binding interface resulted from allowing greater flexibility of the HI loop, we thought it should be possible to introduce larger side chains pointed toward the core on the opposite side of the loop, thereby forcing the loop over into the newly allowed conformation. Toward this goal, an Isoleucine residue at position 454 was mutated to Phenylalanine and its ability to dimerize was tested with  $^{15}\text{N}/^1\text{H}$  HSQC experiments. The results with respect to heterodimerization potential are shown in Fig 4-3 d, which shows that a single conformation is present, and that its ability to form heterodimers with HIF-2 $\alpha$  PAS-B has been significantly compromised. Likewise, its ability to form homodimers was tested in



serial dilution experiments which reveal that the domain's ability to self-associate has also been significantly disrupted.

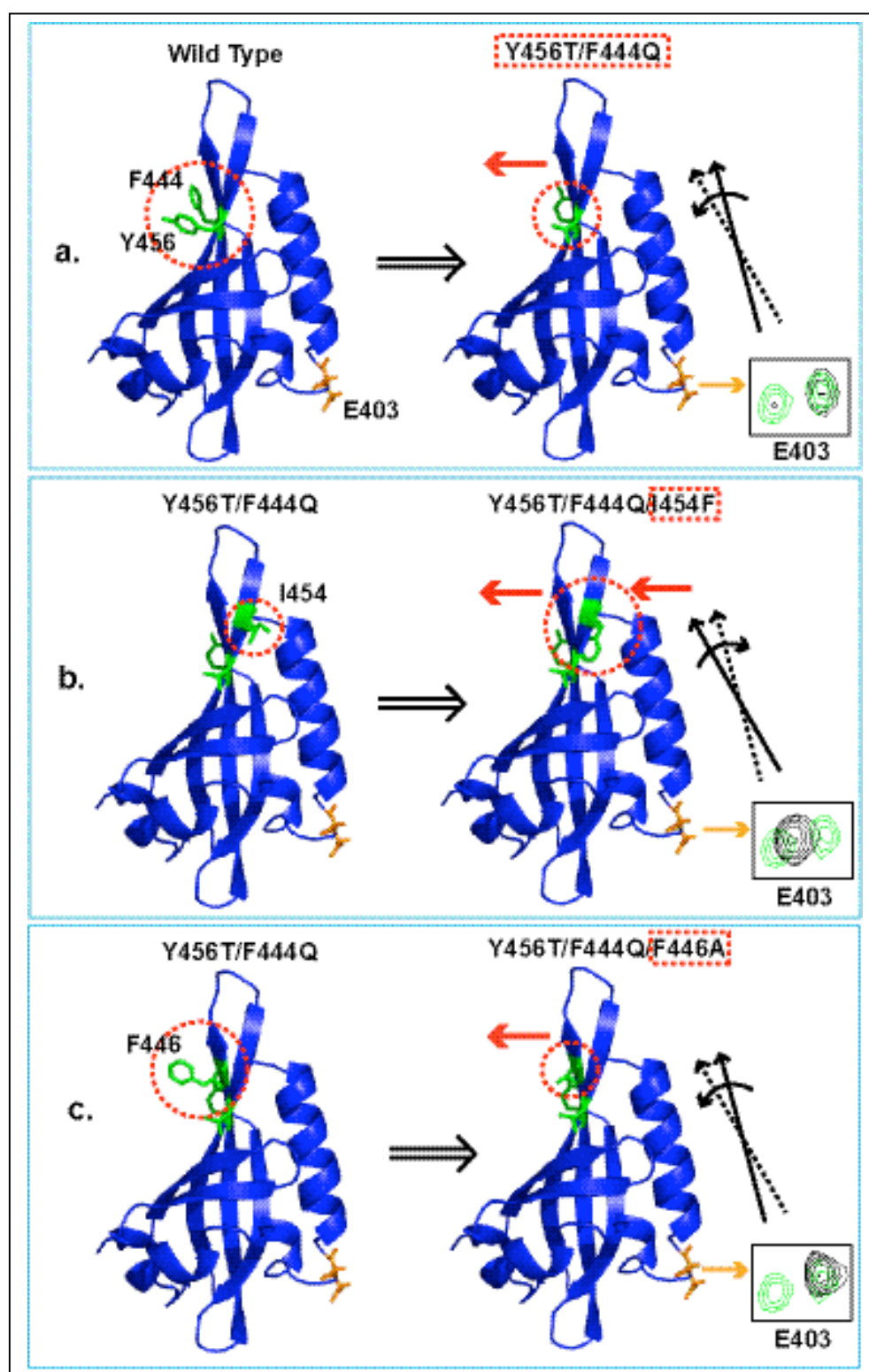
One further mutation was designed by Matthew Evans to further test the idea that a relief of steric hinderance of the HI loop could completely shift the equilibrium to a new, non-binding conformation. An additional relatively large, solvent-exposed residue (F446) near the HI loop was mutated to Alanine, and this triple mutant was also tested for its ability to form dimeric complexes. It appears that a further reduction of large bulky side chains in this region can shift the equilibrium to the new conformation, as evidenced by the chemical shift of the far downfield peak representing the distant E403 residue, which has fully shifted to the new position (Fig 4-3 e - insert). It is also clear from this experiment that its ability to heterodimerize (Fig 4-3 e), and homodimerize (Fig 4-4 d), has been significantly disrupted.

*v. Proposed mechanism for the conformational change*

A model that represents a possible structure-based mechanism for the effect of mutagenesis on the dimerization interface is also described in Fig 4-5. As mentioned previously, a relief in the steric hinderance of the HI loop could allow it to have a greater flexibility that could disrupt the solvent exposed  $\beta$ -sheet binding interface. Fig 4-5 (a) shows how this could be occurring for the double mutant (Y456T/F444Q) which originally shifted the equilibrium toward the new conformation indentified by the Y456T mutant. In addition to the proposed effect on the HI loop, relief of steric hiderance in this region could also allow the F $\alpha$  helix on the other side of the protein to shift position as described by the vector diagrams drawn next to the structure representations in Fig 4-5

a,b and c. This could easily explain how the distant E403 residue is experiencing a change in its backbone amide chemical shift to a more downfield position in the HSQC experiments, because allowing it to move slightly out of position with respect to the macrodipole created by the long  $\alpha$ -helical secondary structural element would reduce the large degree of deshielding normally experienced by this backbone amide, thereby reducing the magnitude of its chemical shift. This is shown in Fig 4-5 a by spectral peaks from the E403 backbone amide resonances representing the two conformations observed in the single, Y456T mutant (green contours), overlaid with those observed in the double mutant (Y456T/F444Q – black contours). The hypothesis is further supported by the additional mutation of F446 to alanine, which continues to relieve steric hinderance in this region and effectively shifts the equilibrium of this peak entirely to the new position, as shown by the spectral overlays in Fig 4-5 c.

**Fig 4-5. Proposed mechanism for the formation of a new, non-dimerizing conformation(s) of ARNT PAS-B. Transitions in each figure from larger to smaller red circles indicate mutations from larger to smaller residues and vice-versa. Black line vector diagrams to the right of the structure representations indicate the proposed relative movements of the F $\alpha$  helix (from solid to dashed orientations). Red arrows indicate either HI movement due to a relief of steric hinderance (arrow on the left (a and c)) or introducing steric bulk on the opposite side (arrow on the right, b) Peaks representing the E403 residue (orange sticks) are shown as green contours originating from the spectrum of the single Y456T mutant, to show representative environments felt by this residue in the two conformations, and black contours for the indicated mutations. a. Structural representation of the double mutant (Y456T/F444Q), which shifts the equilibrium predominantly to the new, non-dimerizing conformation. b. Similar diagram for the triple mutant (Y456T/F444Q/I454F), which suggests an intermediate environment experienced by residue E403. c. Diagram for the triple mutant (Y456T/F444Q/F446A), which should have the most HI loop flexibility, and suggests that a shift entirely to the new, non-dimerizing conformation has occurred.**



Another result that could support this proposed mechanism is shown in Fig 4-5 b, which describes the attempt to force the HI loop over into the new position by introducing a larger side chain (I103F) in the region which points more toward the core. Interestingly, although the backbone amide resonance representing E403 also indicates a single conformation, it appears to be in an intermediate chemical environment with respect to the two previously identified conformations (Fig 4-5 b - green and black contours). This would actually be expected if the proposed mechanism is correct, because introduction of the larger residue within the core should not only force the HI loop over to disrupt the dimeric interactions, but should also help push the F $\alpha$  helix back toward its original position as shown by the vector diagrams to the right of the structures in Fig 4-5 b.

The proposed mechanism does not take into account effects on the dimerization behavior of ARNT PAS-B due to changes in residue type (such as the switch from a nonpolar to a polar side chain in the F444Q mutation). Although these types of changes could certainly have an effect on the dimerization potential of the domain, one of our goals in this investigation was to identify regions within the core of the ARNT PAS-B domain that could be used to transduce an effect on its binding activity. It is possible that the binding of small molecule ligands within the core of a small molecule domain could affect its ability to interact with other proteins in important biological pathways, and that these effects might be used as a strategy for the discovery of lead compounds for the development of therapeutic agents to treat a variety of disease states. Mutagenesis experiments such as these could help identify relatively accessible regions within this domain that could be targeted by small molecule ligands for this purpose. The next

chapter will describe initial efforts to identify small molecule ligands that are capable of binding within the core of the ARNT PAS-B domain.

### V. Ligand Screening of ARNT PAS-B

It has been well established that PAS domains found in proteins from many systems bind small molecule ligands for a variety of sensory and activation functions, and that molecular structures of these cofactors vary considerably depending on their particular function (reviewed in (15)). For example, the FixL protein coordinates a heme cofactor covalently within its core to sense oxygen for the regulation of nitrogen fixation pathways (25). Oxygen binding to the heme moiety increases its planarity, triggering protein conformational changes which deactivate the FixL kinase activity. FAD, FMN and p-coumaric acid are also used by PAS domains to sense light and/or oxygen for a variety of biological functions (28-30, 90, 113) (discussed in more detail in chapter I).

The wide range of ligand binding capabilities demonstrated by PAS domains implies that members of this family have the potential to bind many types of small molecules, while retaining a high degree of structural conservation between individual representatives of this family. This flexibility with respect to ligand binding, in addition to the wide range of protein-protein interactions previously discussed, provides an interesting structural problem, as well as an evolutionary one. PAS domains are protein-protein interaction domains, and these associations are often coupled to the sensory functions of bound ligands. However, many examples of these domains seem not to bind naturally occurring ligands, and could therefore act merely as additive interaction modules for the purpose of modulating the affinity and/or specificity of multi-protein complexes.

This leads to a central question: Which evolved first in PAS domains, ligand binding or protein-protein interaction activity. The early prototype of this domain type

might have bound a particular ligand, and as the primary sequence diverged, protein-protein interaction activity gradually developed. For modern PAS examples for which no ligand binding activity has been observed, the stability imparted by such a small molecule could have been substituted by the incorporation of well-placed hydrophobic side chains within the binding core. But other than ligand binding and protein-protein interactions, PAS domains have not yet been identified that serve other functions (such as catalysis, direct DNA-binding, or sequestering naturally occurring small molecules). With this in mind, it is likely that the ability to bind small molecule ligands developed originally as a means to modulate already existing protein-protein interaction activities.

Regardless of which activity evolved first, it is intriguing how a relatively small domain managed to develop such a wide range of ligand binding specificities while maintaining a conserved structure. The presence of multiple, tandem PAS domains in many proteins could mean that duplication events occurring during the course of a proteins evolution allowed redundant copies of these modules to sample different naturally occurring ligands as their primary sequence gradually diverged from that of their ancestor domain, which eventually led to the development of a useful regulatory function. In this way, ligand binding could have also developed independently of any previous function, since such a redundant protein module could theoretically exist essentially without a function until one evolved.

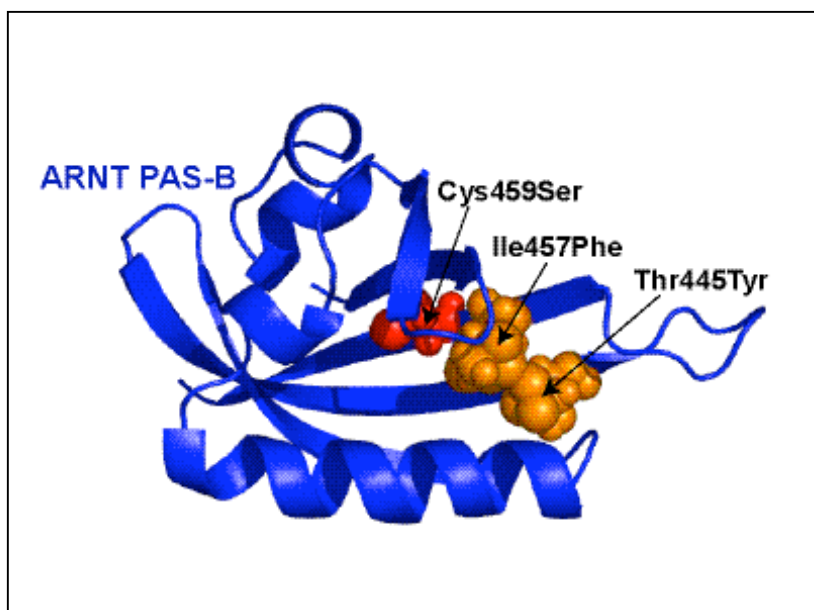
These domains could sometimes (or always) have a flexible core capable of accommodating many types of small molecule ligands, often without disruption of the overall PAS structure. This would allow an already functional domain to sample a variety of existing metabolites and/or exogenous small molecules until a sufficiently

beneficial regulatory mechanism of the existing function developed. From an evolutionary perspective, the widespread use of PAS domains in many (>3000) proteins from all kingdoms of life could be a direct result of the great malleability of these domains. A protein fold that is amenable to many modifications without disruption of an already essential function would be very useful for the evolution of more sophisticated and/or diverse activities. This idea is partially supported by the poor sequence conservation of PAS domains, which greatly complicated the identification of many members of this domain family until sophisticated HMM-based algorithms were developed (119).

Current work in our lab by Qiong Wu involving relaxation studies of partially deuterated methyl groups implies that this idea of a particularly accommodating PAS domain might be possible. Using these experiments, it was observed that side chain methyl groups buried in the core of the PAS-B domain of HIF-2 $\alpha$  exhibit a great degree of flexibility on the nanosecond-picosecond timescale while maintaining an overall robust folded state. In addition, several mutations have been made within or near the core of the PAS-B domain of ARNT for the purpose of spin labeling or modulating protein-protein interactions (Cys459→Ser, Thr445→Tyr, Ile457→Phe – Fig 5-1, Chapter IV), which still permit stable folding of the domain. Also, during the course of the spin label studies, one wild type cysteine revealed itself to be much more reactive with the spin label than was expected. This residue (Cys459), was not mutated to serine in the early stages of the project because, although spin labeling of this residue was undesirable, the high resolution structure of the ARNT PAS-B domain showed that it was buried deep within the core and should therefore have been relatively inaccessible to the thiol-reactive



spin label (Fig 5-1). However, mass spectrometry indicated that this residue was spin labeled with a high efficiency after incubation for 3 hours at 25°C with a 3-fold excess of MTSL spin labels (> 90% - data not shown). This indicates that although the ARNT PAS-B domain is well folded and stable at high concentrations in solution (soluble up to at least 3mM for short periods, and 1mM for weeks), there is a high degree of accessibility within its core.



**Fig 5-1. Mutations within or near the core still permit folding of the ARNT PAS-B domain. The side chain of Cys459 (located in the I $\beta$  strand) was shown by mass spectrometry to be highly reactive with MTSL spin labels, despite its location buried deep within the core.**

PAS domains are capable of binding a wide variety of ligands, and are thought to use these cofactors to initiate small conformational changes in the domain, which can then modulate their protein-protein binding activity (36). We have shown through mutagenesis of the PAS-B domains of both HIF-2 $\alpha$  and ARNT that relatively minor changes in the surface complementarity and/or conformation of these domains (as evidenced by small chemical shift changes of backbone amides) can disrupt protein-protein interactions *in vitro*, and also have large effects on the hypoxia response pathway *in vivo* (8). Therefore, we wanted to see if small, artificial ligands could be identified that

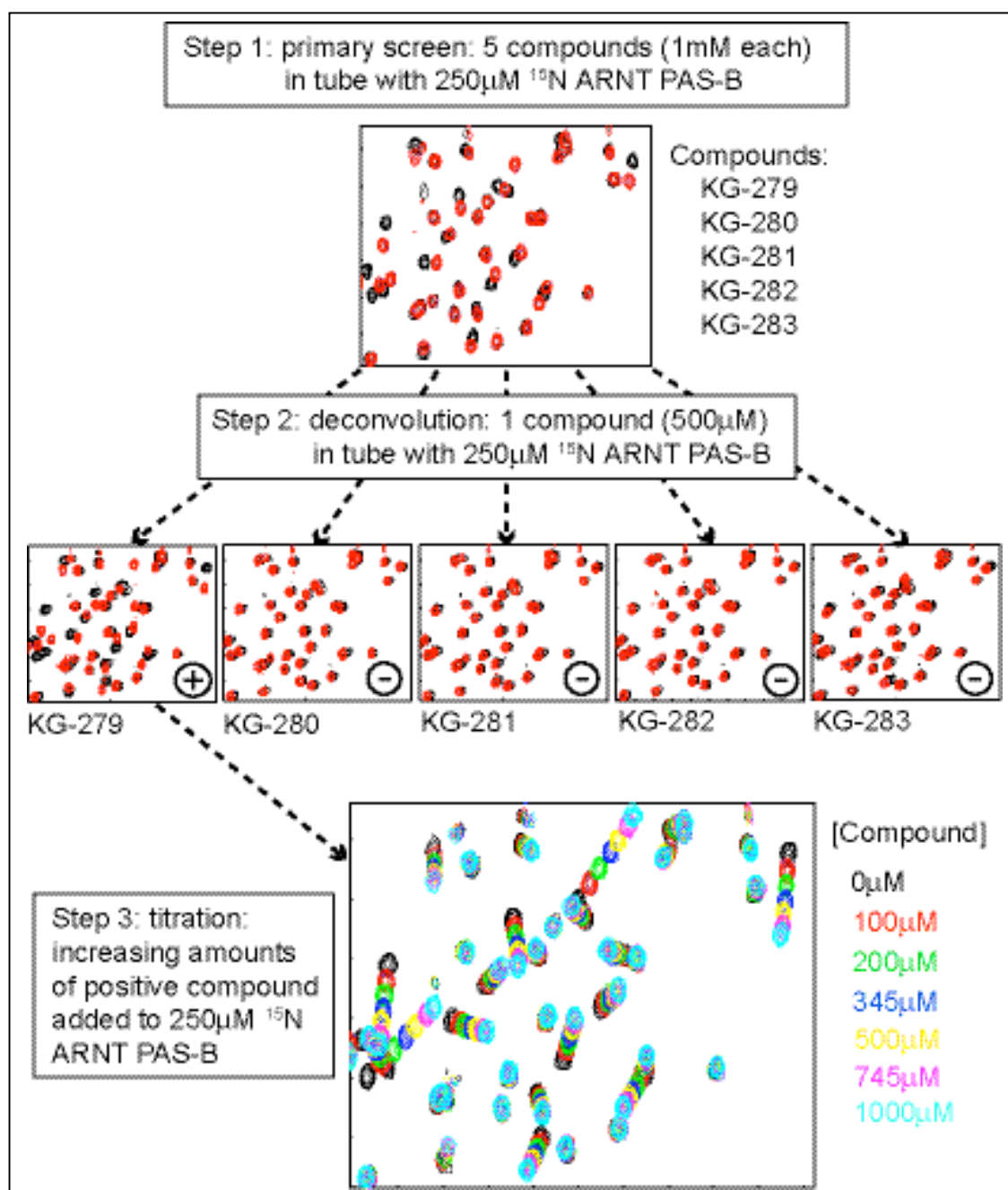
bind into the cores of PAS domains and exert an effect on PAS-mediated protein-protein interactions involved in the hypoxia response pathway.

*i. Ligand screening strategy and methods*

An in-house small molecule compound library assembled by Dr. Carlos Amezcua consisting of over 750 compounds was used for the initial screen of ARNT PAS-B. The overall strategy involved a primary screen of compound mixtures followed by a secondary deconvolution step to identify specific compounds that bound to the PAS domain. The first step of the primary screen was to make solutions consisting of five compounds at 50mM each in DMSO. Purified protein samples at a concentration of 250 $\mu$ M were prepared, 11 $\mu$ l of the compound mixtures were added (1mM final concentration for each compound), and  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra were then acquired. Comparative overlays of these spectra and with that of a control sample (containing only 11 $\mu$ l DMSO added) were used to identify the presence of one or more binding compounds. The observation of peak broadening or significant changes in chemical shift was considered indicative of a potential binder, so the determination of a potential “positive” in the initial screen was somewhat subjective. Initial screening was aided by the use of a Zymark Zymate-XP robot module along with in-house software written by Dr. Kevin Gardner that automatically switched samples after data acquisition and began the next screening experiment with a series of previously prepared samples. Whenever significant spectral changes were observed for a given mixture, a deconvolution step was then performed by running similar experiments with individual compounds at a concentration of 500 $\mu$ M to identify the particular small molecule(s) of interest. Finally,

individual compounds that were determined to result in significant chemical shift perturbations were subjected to titration analysis. Sixteen deconvolution steps were performed (80 spectra total), and from these, 22 compounds were identified as positive hits. Four of the deconvolutions did not yield any compounds that were deemed a positive hit, and five deconvolutions yielded multiple positives from the five compounds present in the same tube during the primary screen. The overall screening protocol is outlined in Fig. 5-2.

**Fig 5-2. Overview of screening protocol of ARNT PAS-B demonstrated by a primary screening mixture containing compound KG-279.**



### *ii. Dissociation constant measurements*

After the identification of a potentially interesting compound, attempts were made to obtain a binding curve for estimation of the dissociation constant ( $K_d$ ). These curves were based on changes in the chemical shifts of backbone amide resonances as compound was titrated into uniformly  $^{15}\text{N}$ -labeled protein in a series of  $^1\text{H}$ - $^{15}\text{N}$  HSQC experiments. This analysis was performed using titration analysis software developed in our lab by Dr. Carlos Amezcua, Brad Holmes and Dr. Kevin Gardner, and was incorporated into the nmrView software package (150). A representative binding curve is shown in Fig. 5-6 c for one of the primary hits that was chosen for further characterization, along with a description of the  $K_d$  determination from the titration data.

When possible and needed,  $K_d$ 's were also determined using isothermal titration calorimetry (ITC), both as an independent method for affinity assessment and as justification for the chemical shift-based binding curve method. A Microcal ITC machine available from the lab of Dr. Steve Sprang was used for the measurements, and data was obtained by titrating compound (1mM over 30 injections) into 50 $\mu\text{M}$  protein in the sample cell. In certain cases, these parameters were optimized to obtain better data and/or to test the results under different conditions.

### *iii. Choice of library compounds*

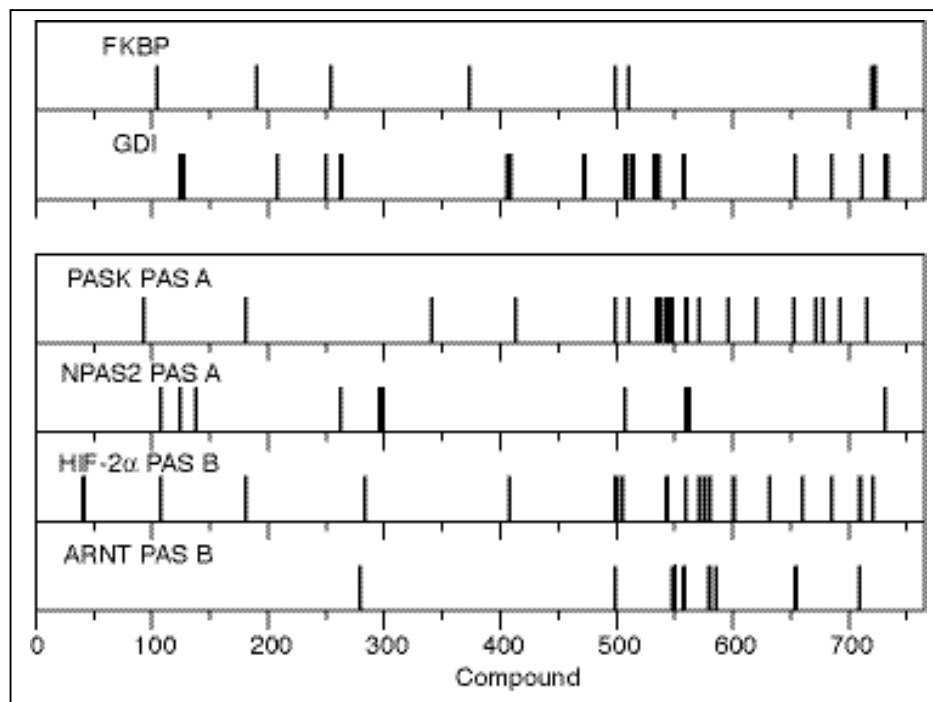
Dr. Amezcua's choice of compounds used for building the primary screening library was based on several factors. Small compounds (average MW=200 Da) were chosen so that they could bind within the core of a relatively small protein such as a PAS domain with minimal steric hindrance. Both heterocyclic and aromatic moieties were

used as the structural basis of the library due to the precedence established for these types of compounds in past drug discovery efforts (172, 173), and also because similar moieties are often present in naturally occurring PAS domain ligands (e.g. heme, FMN, dioxin). In addition, compounds were chosen that contained at least one hydrogen bond acceptor or donor to improve the solubility of these molecules in aqueous buffers.

*a. Compound specificity*

Certain aspects of the small compound library were a source of concern at the onset of this screening effort. First, compound solubility in aqueous buffer was difficult to predict, due to the types of compounds used to construct the library. Also, it was not certain *a priori* whether positive hits could be identified that bound to PAS domains with a high degree of specificity, given the limited complexity of compounds in this molecular weight range. Positive hits were identified from the initial screens of four PAS domains studied in our lab, and two other domains obtained from the lab of Dr. Michael Rosen that were established previously to contain pre-formed surface pockets (FK-506 binding protein, FKBP (174), and the C-terminal domain of RhoGDI, a Rho GDP-dissociation inhibitor which binds isoprenes (175)). These data indicate an acceptable level of specificity for these proteins, as evidenced by the diversity of initial hits identified for each domain (Fig. 5-3).

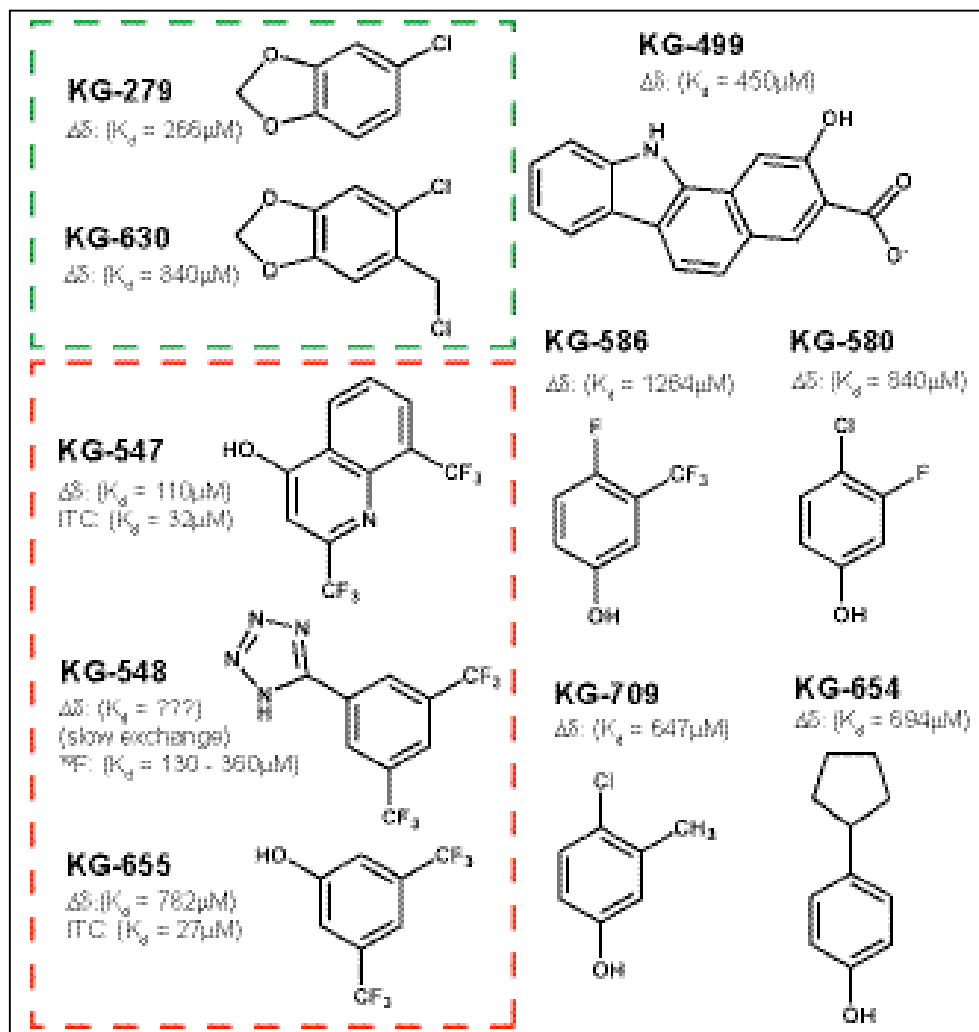
**Fig 5-3. Compounds from the small molecule library demonstrate a reasonable degree of specificity to individual targets. Four PAS domains in addition to FKBP and GDI were screened and compound numbers considered to be “positive” hits are shown as bars ( $K_d < 1\text{mM}$ ). The relatively disperse patterns of hits across all the targets indicates that certain proteins are specific for individual compounds.**



*iv. Lead compounds identified by the initial ligand screen of ARNT PAS-B*

The primary screening of the ARNT PAS-B domain (356-470) found ten compounds that resulted in significant spectral perturbations upon deconvolution of the initial mixtures and were thus deemed worthy of further investigation. These were reasonably soluble, and titration analysis indicated  $K_d$ 's for each compound to be below 1mM. The chemical structures of these initial “hits” are shown in Fig. 5-4.

**Fig 5-4.** Top ten compounds identified by the NMR-based screening protocol for ARNT PAS-B. The green and red boxes indicate groups of compounds that are structurally related, and cause similar patterns of chemical shift perturbations when added to ARNT PAS-B. Titrations were performed for all compounds as described in Fig 5-2, and the resulting  $K_d$  estimates using the chemical shift method (when possible) are reported.  $K_d$  measurements for two compounds using ITC (KG-655 and KG-547), and estimates from the 1D  $^{19}\text{F}$  NMR-based method (KG-548) are also shown.



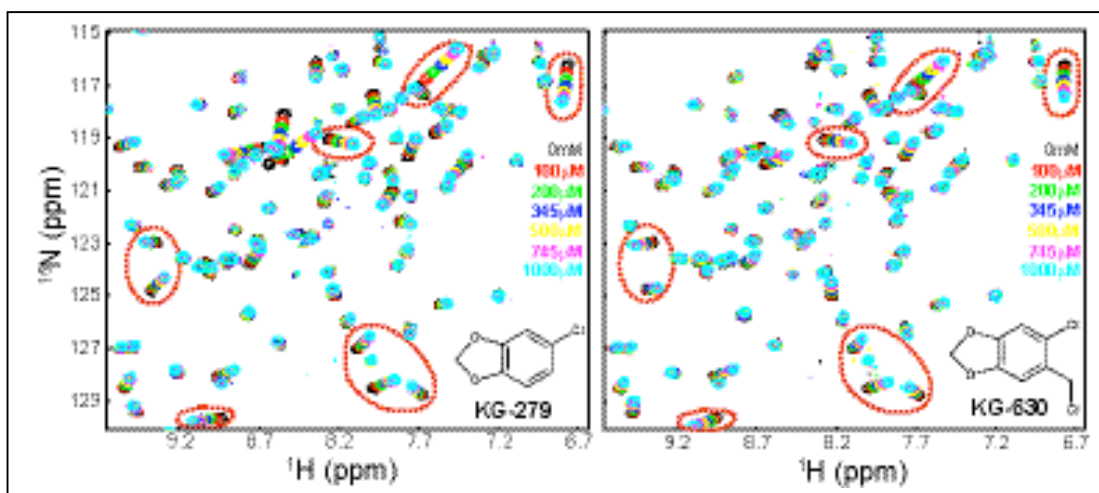
Two of these compounds, KG-279 and KG-630 (Fig. 5-4 – green box), are very similar in that they both include a mono- or di-substituted benzyl ring attached to a dioxole group. Relationships such as these between pairs or groups of binding compounds were considered important for the identification of structure activity



relationships (SAR) that could help identify important moieties that allow specific binding within the core of a PAS domain. Another criteria for potential lead compounds destined for more extensive characterization was the presence of chemical moieties that were amenable to modification for the optimization of binding affinity.  $^{15}\text{N}/^1\text{H}$  HSQC spectra of ARNT PAS-B with and without KG-279 and KG-630 are shown in Fig. 5-5 as examples of positive hits in the primary screen after subsequent deconvolution.

Compound binding is demonstrated by significant chemical shift changes of backbone amide resonances or peak broadening due to intermediate exchange. In the case of KG-279 and KG-630, very similar patterns of chemical shift changes are observed with both compounds (Fig 5-5), indicating that both of these bind to a similar and specific location within the protein, and could be optimized for tighter binding by substitutions on the benzyl ring.

**Fig 5-5. Titration of compounds KG-279 and KG-630 into ARNT PAS-B.** Compounds were added incrementally into 360mM  $^{15}\text{N}$ -ARNT PAS-B (see figure for concentrations), and perturbations of chemical shifts monitored. Similar patterns of chemical shift changes are shown with red circles, indicating a similar mode of binding for the two related compounds.



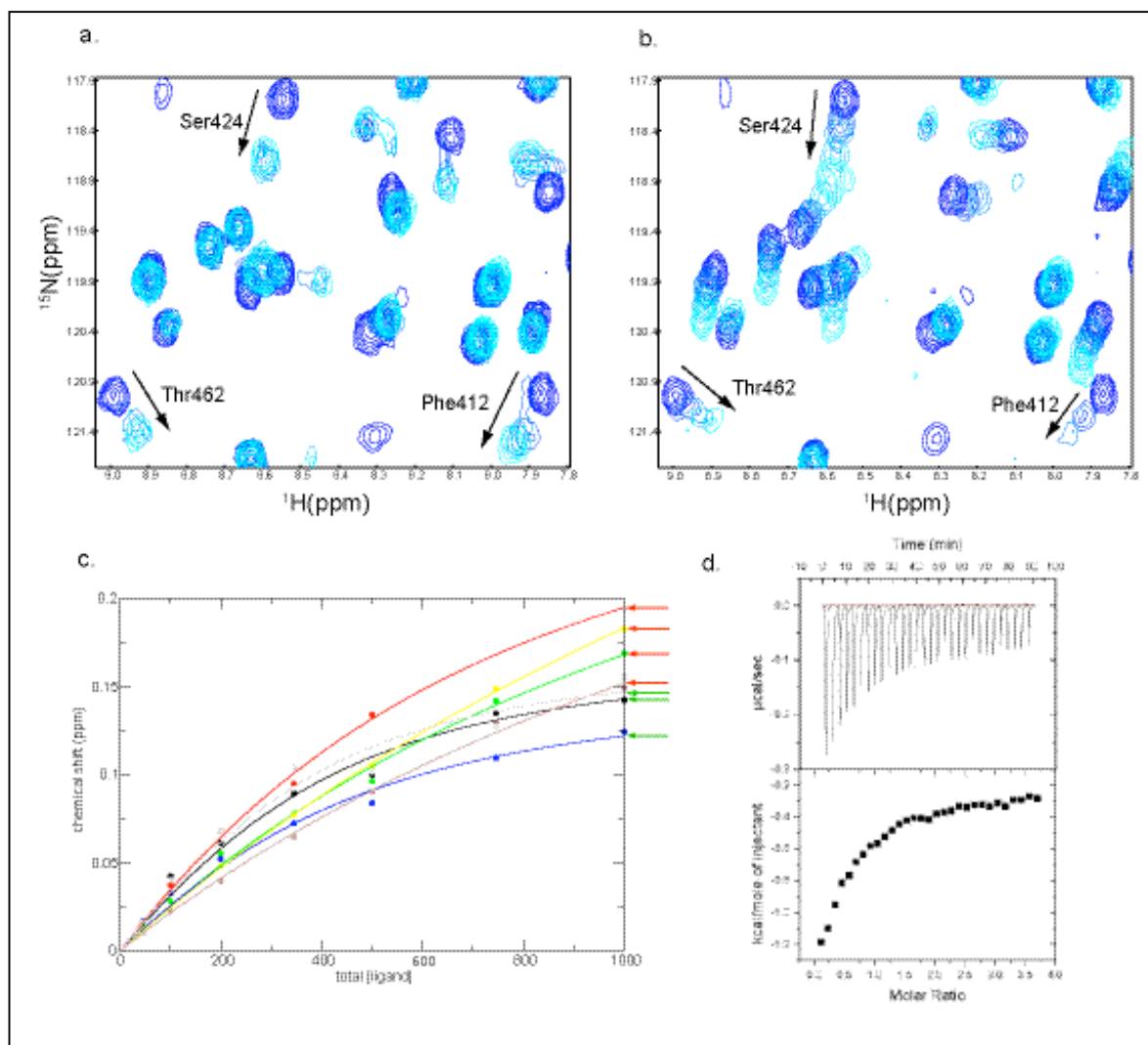
Another set of compounds identified from the initial screen that possessed a different type of structural similarity (Fig. 5-4 – red box) was eventually chosen for further study, however, due to indications that they might be more likely to bind with high affinity. Two representatives from this group (KG-548 and KG-655) are substituted trifluoromethyl toluenes, differing only in the groups attached to the C-1 position of the central moiety.

KG-548 was particularly interesting because titration analysis indicated that it was exhibiting slow exchange behavior. Such slow exchange during a titration is characterized by the loss of signal intensity at the chemical shifts of the apo protein, followed by the progressive reappearance of peaks at new locations. This behavior is typically observed for relatively tight complexes (approximately less than 1  $\mu$ M in most cases), because it implies that the exchange rate,  $k_{ex}$ , is much lower than the change in resonance frequencies between the two states ( $k_{ex}(s^{-1}) \ll \Delta\nu(s^{-1})$ ). This situation contrasts with both fast exchange behavior ( $k_{ex}(s^{-1}) \gg \Delta\nu(s^{-1})$ ), where signals gradually shift to their new frequencies as the titration progresses (signals appear at locations representing the population-weighted average of the free and bound states), and intermediate exchange ( $k_{ex}(s^{-1}) \approx \Delta\nu(s^{-1})$ ), which causes signal loss due to continuous dephasing of the magnetization in the transverse plane. It should be noted, however, that the existence of slow exchange behavior in an NMR experiment indicates only that the off-rate ( $k_{off}$ ) is slow relative to the difference in the resonance frequencies between the bound and unbound states. Therefore, this behavior can sometimes be observed when the  $K_d$  is much higher, if the on-rate ( $k_{on}$ ) is not diffusion limited or if other exchange

processes limit the accessibility of the ligand to its protein target (discussed further in section v. of this chapter).

Fast and/or intermediate exchange behavior was apparent for three other compounds in this group (KG-547, KG-586 and KG-655). Examples of these phenomena are shown by the titration series of compounds KG-548 and KG-655 in Fig. 5-6 a,b. Also shown in Fig. 5-6 c is the curve fitting for several backbone amide shifts during the course of the titration of KG-655, which yields an estimate of 782 $\mu$ M for the  $K_d$  based on the average shifts of all the resonances used for the analysis. The addition of compound KG-586 resulted in only minor changes in the spectra of ARNT PAS-B, and was excluded from further characterization. Further studies focused on compounds KG-548 and KG-655, due to the identical arrangement of their tri-fluoromethyl groups, in order to test the general efficacy and optimization potential of this particular class of small molecules identified by the initial screening protocol.

**Fig 5-6.** a. Titration analysis of compound KG-548. Compound added incrementally from 0-1mM (dark to light blue) into  $^{15}\text{N}$ -ARNT PAS-B (360 $\mu\text{M}$ ). Signal loss followed by concomitant appearance of new signals indicates slow exchange is occurring. b. Same experiment as a. for compound KG-655. Gradual changes in chemical shifts indicate fast exchange. c. Titration analysis for  $K_d$  estimates of KG-655. Changes in chemical shifts of a few residues are fit to a binding curve, obtaining a  $K_d$  estimate for this compound (782 $\mu\text{M}$ ). d. An example of ITC data for the titration of KG-655 into ARNT PAS-B. The curve indicates a dissociation constant of 27 $\mu\text{M}$ , well below that obtained from the chemical shift data.



*a. Determination of dissociation constants*

ITC was used to acquire the dissociation constant for KG-655 under two different sets of conditions (50 $\mu$ M or 100 $\mu$ M protein in the cell and 1mM compound titrated in 30 injection steps). An example of these data is shown in Fig. 5-6 d. The results for both experiments matched very well ( $K_d$ =25.5 $\mu$ M and 27.0 $\mu$ M, respectively), although this value is 30-fold lower than that obtained from the chemical shift data. This can partially be explained by the variable curves that were fit for individual residues in the protein (red arrows and green arrows in Fig. 5-6 c), which might indicate two state binding.  $K_d$  estimates for the residues shown by green arrows alone (the curves showing a faster rate of saturation) range from 112 $\mu$ M to 256 $\mu$ M, which implies an affinity below that determined by the ITC experiments. These data shows that although the chemical shift method can be useful for initial  $K_d$  estimates, care should be taken in the interpretation due to unknown factors that could result in unusual titration curve fitting, such as poor ligand solubility or multiple equilibria caused by dimerization, non-specific binding or aggregation.

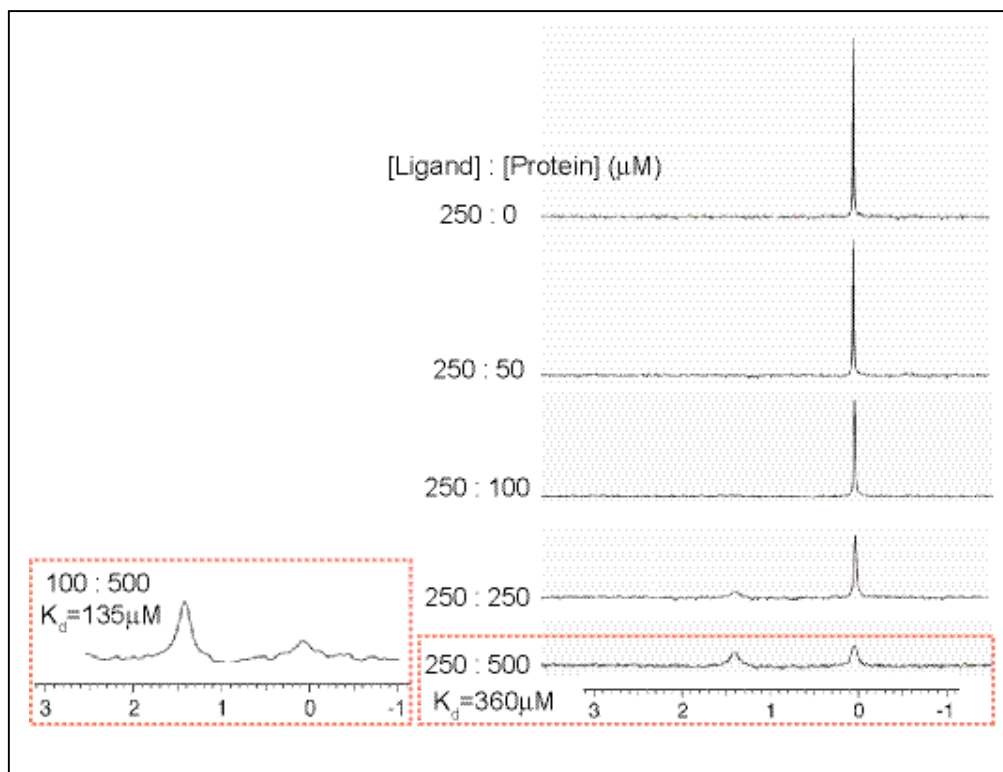
*v.  $^{19}\text{F}$  NMR of compound KG-548*

It should be explicitly noted that accurate  $K_d$  estimates from titration curves based on changes in chemical shift are often difficult to obtain for compounds that exhibit slow exchange, due to the fact that only two resonance frequencies are observed during the course of the titration. For this reason,  $K_d$  estimates based on  $^{15}\text{N}/^1\text{H}$  HSQC chemical shift changes could not be obtained for the compound KG-548. Although ITC was also attempted for this compound under many conditions (including reversing the strategy by

titrating protein into the measurement cell containing compound), measured heats of binding could not be fit to interpretable binding curves. This was presumably due to similar problems as described previously for the chemical shift method such as compound aggregation or multiple equilibria, which can also be problematic for ITC-based  $K_d$  measurements.

An alternate method was therefore used to get an estimate for the dissociation constant of KG-548. This method takes advantage of the spin=1/2 state of fluorine atoms ( $^{19}\text{F}$ ) which allows  $K_d$  measurements to be obtained directly from one-dimensional  $^{19}\text{F}$  NMR spectra of the compound, where signals arising from the two tri-fluoromethyl groups in the 3,5 positions on the benzyl ring can be used as probes for the interaction. For this series of experiments, spectra were obtained on a 250 $\mu\text{M}$  sample of compound KG-548 using a locally available 300MHz Varian instrument equipped with a fluorine probe. Samples were also prepared that contained varying amounts of natural abundance ARNT PAS-B (356-470) in addition to compound, to acquire a titration series for dissociation constant measurements. The results for KG-548 are shown in Fig. 5-7, and the presence of two distinct chemical shifts near the end of the series recapitulates the slow exchange behavior observed in the  $^{15}\text{N}/^1\text{H}$  HSQC experiments. The peak that appears for the unbound state (arbitrarily set to 0 ppm) represents all six fluorine atoms in the two tri-fluoromethyl groups, and shows up as a single peak due to the high degree of symmetry in the molecule.

**Fig 5-7.  $^{19}\text{F}$  NMR showing the binding of KG-548 to ARNT PAS-B. Protein was titrated into a  $250\mu\text{M}$  sample of compound to obtain a  $K_d$  estimate by calculating the areas under each peak using the procedure outlined in the text. Two different  $K_d$  estimates were obtained, depending on the conditions used in the experiments outlined in red.**



It should be noted that these spectra were referenced with the chemical shift representing the unbound state being assigned to 0 ppm, so the reported ppm values are only for comparison purposes. The significant peak broadening observed for the bound state peak at 1.4 ppm was expected, due to the longer rotational correlation times (and subsequently increased transverse relaxation rates) for the probe molecule (KG-548) when bound to the ARNT PAS-B domain. Since integration of a particular peak in the slow exchange case can provide an estimate of the relative concentrations of compound in the free [L] and bound [PL] states, equation 5.1 can be used to get an estimate of the dissociation constant:

Eq. 5.1

$$K_d = \frac{[P][L]}{[PL]}$$

where: [P] = concentration of free protein (500 $\mu$ M – [PL])

[L] = concentration of free ligand

[PL] = concentration of ligand-protein complex

Using this method, estimates for the  $K_d$  were obtained (135 $\mu$ M or 360 $\mu$ M) depending on the concentration of compound used in the titration experiments (100 $\mu$ M or 250 $\mu$ M, respectively). Although these values are estimates that depend to some extent on the particular conditions used during the titration, the overall dissociation constant is much higher than would typically be expected for a binding event exhibiting slow exchange behavior (approximately < 1 $\mu$ M in most cases, see section *iv* in this chapter).

One explanation for this result could lie in the behavior of the ligand itself. Slow exchange behavior as an indicator of a tight interaction is not only dependent on the magnitude of the change in chemical shift upon ligand binding, but also on the assumption that the on-rate of ligand binding to the protein,  $k_{on}$ , is diffusion limited. Therefore, one explanation for slow exchange behavior despite a relatively high  $K_d$  is that some aggregation event or non-specific association of the compound and the experimental media is preventing the ligand from interacting with the protein. Ligand aggregation could be irreversible, or it might exist in an equilibrium between the protein-available form and aggregate. Either way, the amount of ligand available for complex formation at any given moment would be less than the concentration added to the NMR sample, and could explain the unexpectedly high dissociation constant.



Another possible interpretation is that the protein exists in two distinct conformations, one that is capable of binding ligand and another this is not. Since the free form of the protein exhibits only a single set of NMR peaks when no ligand is present, this conformational “switch” would have to be occurring in the fast exchange regime (section *iv*). However, since the diffusion constant for a small molecule in water is approximately  $10^8 \text{ M}^{-1}\text{s}^{-1}$ , and given the concentrations of ligand used in the binding experiments ( $\sim 100\text{-}1000\mu\text{M}$ ), this corresponds roughly to an average collision rate between the ligand and protein of  $10^5\text{-}10^6 \text{ s}^{-1}$ . This value still greatly exceeds the conformational exchange rate between the two forms of the protein (ligand binding vs. non-ligand binding) required for fast exchange behavior (approximately  $\geq 100 \text{ s}^{-1}$  with a 0.2 ppm average proton chemical shift on a 500MHz spectrometer). As such, a conformational “switch” that underlies ligand to protein accessibility might result in slow exchange behavior upon ligand binding despite a relatively high dissociation constant.

*vi. KG-548 and KG-655 attenuate ARNT PAS-B dimerization*

Although the dissociation constant for compound KG-548 was higher than initially expected, the presence of benzyl-attached tri-fluoro methyl groups in several of the lead compounds identified suggested that this could be a good starting place for determining structure activity relationships for the purpose of optimizing the affinity by screening related compounds not included in our original library. The ultimate goal was to discover and/or develop small molecule ligands that could modulate the hypoxia response *in vivo*, which was previously accomplished through mutation of the central  $\beta$ -sheet binding interface of HIF-2 $\alpha$  PAS-B (8). Toward this end, we first assessed the

capabilities of the lead compounds to modulate the dimerization efficiency of the ARNT PAS-B domain *in vitro*.

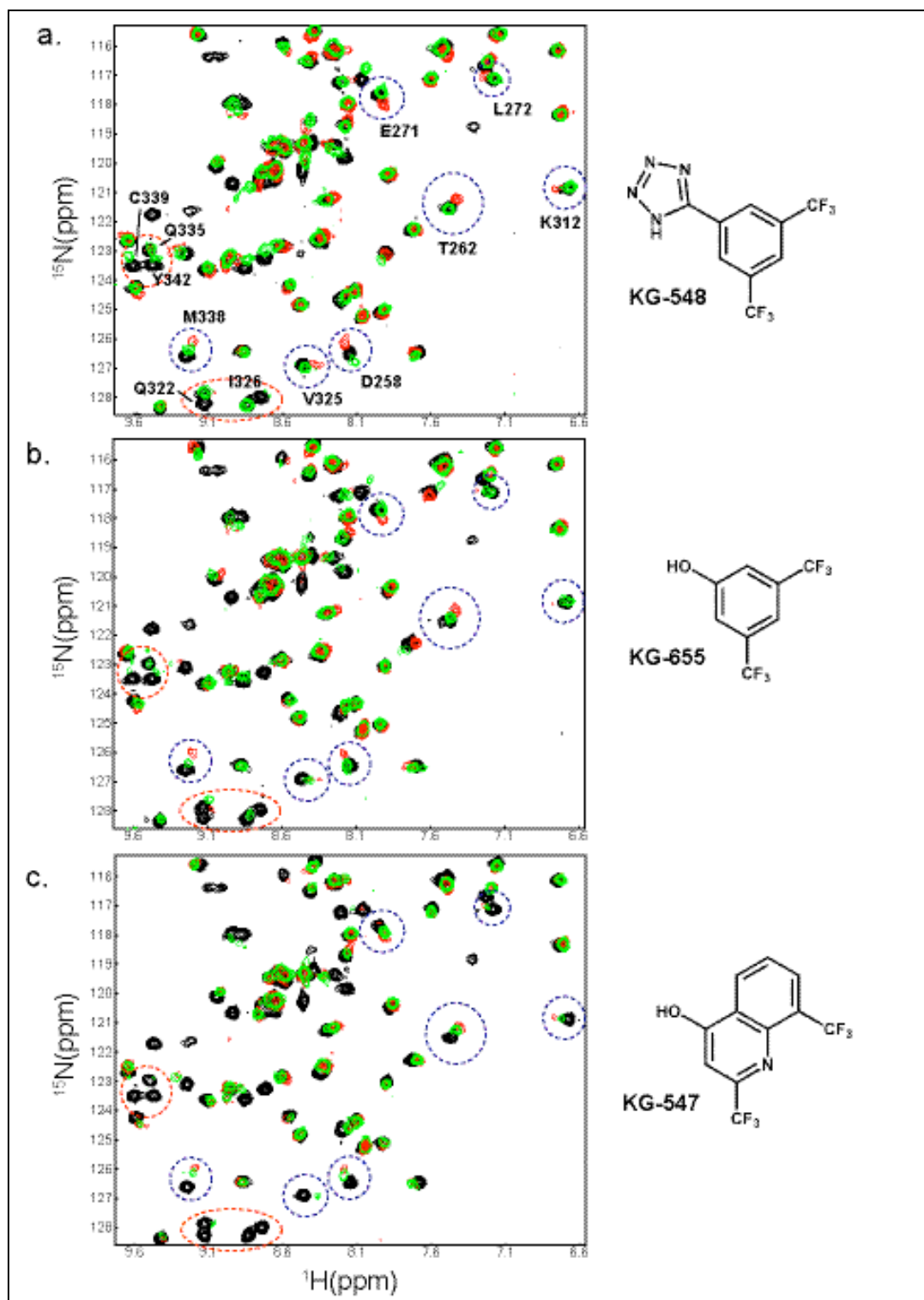
*a. Disruption of the ARNT PAS-B/HIF-2 $\alpha$  PAS-B heterodimer*

To test the ability of these compounds to disrupt the ARNT/HIF-2 $\alpha$  PAS-B heterodimeric complex, we monitored backbone amide resonances from HIF-2 $\alpha$  PAS-B in  $^{15}\text{N}/^1\text{H}$  HSQC experiments of HIF-2 $\alpha$  PAS-B alone (250 $\mu\text{M}$ ), in the presence of a three-fold excess of natural abundance ARNT PAS-B (750 $\mu\text{M}$ ), and again on the same sample with compound added to the complex (1mM). In the initial screen of HIF-2 $\alpha$  PAS-B performed by Dr. Paul Erbel, the compounds KG-548 and KG-655 were not identified as significant binders, so signals from HIF-2 $\alpha$  PAS-B were monitored to avoid the problem of deconvoluting chemical shift changes that occur from ligand binding from those arising from complex formation.

Results from the experiments on these two compounds are shown in Fig. 5-8 a,b. Reversion of several backbone amide signals back to their original chemical shift positions, in addition to the reappearance of some peaks that disappear upon complex formation due to intermediate exchange, suggests that these compounds are partially disrupting the complex between the ARNT and HIF-2 $\alpha$  PAS-B domains (Fig 5-8 a,b). This effect seems to be more pronounced for compound KG-548 relative to KG-655. In contrast, another compound from this group of primary hits, KG-547, shows much less of a disrupting influence on the complex (Fig. 5-8 c – blue circles). This result was somewhat surprising, because this compound binds to ARNT PAS-B with very similar affinity as measured by ITC analysis ( $K_d=32\mu\text{M}$ ) than that previously determined for

KG-655 ( $K_d=27\mu\text{M}$ ) (Fig. 5-8 b). In addition, titration of compound KG-547 into  $^{15}\text{N}$ -labeled ARNT PAS-B reveals that many backbone amide resonances experience similar chemical shift changes and broadening relative to similar experiments on KG-548 and KG-655, although significant differences are also apparent (data not shown). Similarly affected residues could indicate a common binding pocket for all three compounds, and the similar orientations of the 3,5 trifluoromethyl groups could be affecting the same backbone amides upon binding to ARNT PAS-B (Fig 5-4 – red box).

**Fig 5-8. Disruption of the heterodimeric complex with small molecule ligands.** Black spectra is from  $^{15}\text{N}$  HIF-2 $\alpha$  PAS-B alone, red spectra were taken after the addition of unlabeled ARNT PAS-B, and green spectra were obtained after the addition of various compounds. a. The addition of 1mM KG-548 partially disrupts complex formation between  $^{15}\text{N}$ -HIF-2 $\alpha$  and ARNT PAS-B domains, as evidenced many backbone amide peaks returning to their original locations upon addition of compound (blue circles), and/or peaks returning from intermediate exchange broadening (red circles) in  $^{15}\text{N}/^1\text{H}$  HSQC experiments. b. A similar affect is seen for KG-655, but not as pronounced as KG-548 in many cases despite the higher affinity for this compound to ARNT PAS-B. c. Addition of compound KG-547 appears to be ineffective at disrupting the complex. Experiments were performed as described in the text.



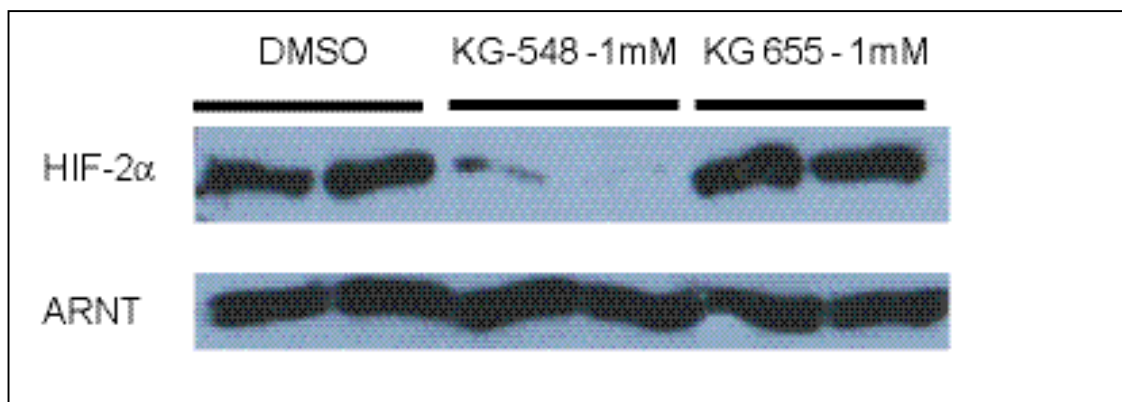
*b. Effects of ligand binding on complex formation of full-length proteins in vivo*

Because of the molecular weight restrictions inherent to NMR-based experiments on proteins in solution, our efforts so far have focused on purified, small domains in an *in vitro* environment. However, through close collaborative efforts with members of Dr. Rick Bruick's lab, progress has been made toward integrating our initial results with studies done in a more physiological context. As previously discussed, Dr. Paul Erbel designed mutations on the solvent-exposed  $\beta$ -sheet surface of HIF-2 $\alpha$  PAS-B that, when introduced into full-length HIF and transfected into a HIF-deficient cell line, significantly attenuated the transcription levels of a luciferase reporter gene under the control of an HRE-containing promoter (8). One of our main goals was to discover and/or develop small molecule ligands that could similarly inhibit (or activate) the hypoxia response *in vivo*. Unfortunately, initial results from studies done by Carolyn Michnoff showed that many of the compounds found in the initial screen of ARNT PAS-B suffered from an unreasonable degree of cytotoxicity at the concentrations needed for complex disruption *in vitro*. Nevertheless, work has continued in Dr. Bruick's lab to assess the potential of these compounds to disrupt dimerization of full-length HIF and ARNT in an *in vivo* setting. Since this work has been done by another lab (using their compounds in addition to ours), most of the details of these studies will not be discussed here. But the results of one experiment will be briefly described, for the purpose of placing our work in a more physiologically relevant context, and to highlight the virtues of productive collaborative efforts in large-scale scientific investigations.

In short, Dr. Bruick's lab developed co-immunoprecipitation assays using antibodies to full-length HIF and ARNT for the detection of complex formation in

nuclear extracts, to assess the abilities of small molecule ligands to disrupt these interactions.

**Fig 5-9. Co-immunoprecipitation assays show a variation in the disrupting influence of compounds KG-548 and KG-655 on the formation of heterodimer formation between full-length HIF-1 $\alpha$  and ARNT in nuclear extracts.**



The results of one experiment is shown in Fig. 5-9, and show that an antibody specific to full-length ARNT is capable of pulling down significant amounts of HIF in the absence of compound (1<sup>st</sup> two lanes). However, this ability is severely compromised in the presence of 1mM KG-548, suggesting that this compound successfully disrupts heterodimer formation between full-length proteins (2<sup>nd</sup> two lanes). Although KG-655 showed a similar effect on heterodimer formation of isolated PAS domains in vitro, it seems unable to mimic the effect observed for KG-548 in the context of this experiment. There are many possible explanations for this result. These include variable solubility of the two compounds, multiple binding sites for KG-655 that could serve to lower the effective concentration in the binding site of ARNT PAS-B, or perhaps the larger size of KG-548 makes it a more effective disruptor of dimerization by causing larger

conformational changes within the protein. Another likely explanation for this variable effect will be discussed in section *v.e.* Whatever the reason, KG-548 seems to be a more efficient dimerization inhibitor despite its lower affinity for the PAS domain as determined by ITC and  $^{19}\text{F}$ -NMR.

*vii. Mapping the binding site of KG-548 in ARNT PAS-B*

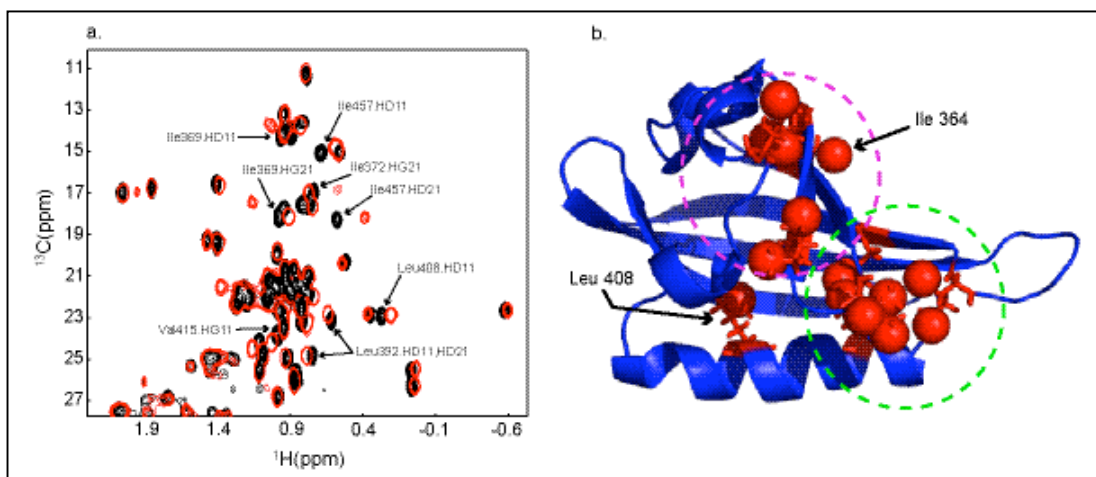
The compounds KG-548 and KG-655 seem to be capable of modulating the dimerization affinity of the ARNT PAS-B domain. The question remained, however, if these compounds merely stick to the exterior of a PAS domain, thereby disrupting the  $\beta$ -sheet binding interface, or if the signal is transduced from the core of the domain in a manner similar to that described for the light-sensing PAS domains involved in PYP and phototropin signaling. In phototropin, light activated chemistry of an internally bound cofactor results in relatively minor conformational changes in the protein that effectively disrupts the domain's interaction with a C-terminal  $\alpha$ -helix by transducing the signal through the central  $\beta$ -sheet binding interface (36). Therefore, we wanted to test whether our primary lead compound, KG-548, was binding within the core of ARNT PAS-B to affect the dimerization potential of this domain.

Side chain methyl groups are often good probes for interactions that occur within protein cores for several reasons. The nonpolar character of most residues that contain methyl groups mean that they often lay buried within the hydrophobic cores of folded proteins. They also have the benefit of three equivalent protons, which aid in the sensitivity of NMR experiments. In addition, they freely rotate about the covalent bonds to their attached methylene or methine carbons, reducing the effective rotational

correlation time, which lowers the transverse relaxation rates. For these reasons, we ran  $^{13}\text{C}/^1\text{H}$  HSQC experiments on the ARNT PAS-B domain with and without compound to observe which methyl groups were perturbed upon binding resulting in subsequent changes in their chemical shifts. Natural abundance ARNT PAS-B was used due to the lack of uniformly  $^{13}\text{C}$ -labeled protein present in the lab at that time. This was possible because at the concentration used (1mM), enough naturally occurring  $^{13}\text{C}$  is present in a natural abundance sample (1.1%) to obtain reasonable signals from methyl groups, primarily due to the many desirable characteristics of methyl groups discussed above.

The compound KG-548 was added in equimolar amounts after a reference spectra was obtained, and previously assigned resonances in the methyl region of the spectrum that exhibited significant changes in chemical shift were then mapped onto the high resolution structure of the ARNT PAS-B domain. Results of these experiments are shown in Fig. 5-10

**Fig 5-10. Methyl groups perturbed by KG-548 binding.** a.  $^{13}\text{C}/^1\text{H}$  HSQC spectra of natural abundance ARNT PAS-B shows changes in the chemical shifts arising from particular methyl groups upon addition of KG-548 (red) compared to ARNT PAS-B alone (black). b. Perturbed methyl groups were mapped onto the high-resolution structure of ARNT PAS-B (red spheres), revealing a likely binding site in the core of the domain.





Perturbed methyl groups seem to map to two general regions in ARNT PAS-B, one of which lies near the C-terminal end of the F $\alpha$ -helix (green circle). Another general region (magenta circle) centralized to specific areas on the A $\beta$ - and B $\beta$ -strands is also being significantly affected. In addition, of the residues comprising the F $\alpha$ -helix that contain methyl groups, only ones with side chains pointed directly into the core of the PAS domain (Leu408 and Val415) are significantly affected by compound binding. This suggests that KG-548 is binding within the core of the domain, although an alternative explanation is possible. It was shown that the binding of KG-548 effectively disrupts homodimerization of ARNT PAS-B, so within the context of this experimental setup, there exist multiple equilibria that could affect the chemical environments felt by methyl groups both internal and external to the core. This is perhaps demonstrated best by Ile364, whose methyl group is also being affected even though it is solvent exposed on the central  $\beta$ -sheet. Since this surface has been well characterized as the interaction interface for both homodimerization and heterodimerization of ARNT PAS-B, it is likely that this methyl group is not being affected by direct ligand binding, but rather by the subsequent disruption of the homodimer.

In light of this, more studies will be required to accurately map the binding site of these compounds, most likely including intermolecular NOE measurements to get direct distance restraints between the PAS domain and the ligand. An alternative strategy is possible, however, as the availability of mutant versions of ARNT PAS-B that are incapable of homodimerizing (Chapter IV) will allow a more accurate determination of residues that interact directly with ligand without having to deconvolute the effects of homodimerization. Further studies are underway to determine if the ligand binding

capacity of this and other PAS domains are strongly coupled to the protein-protein interaction activities of the ARNT PAS-B domain.

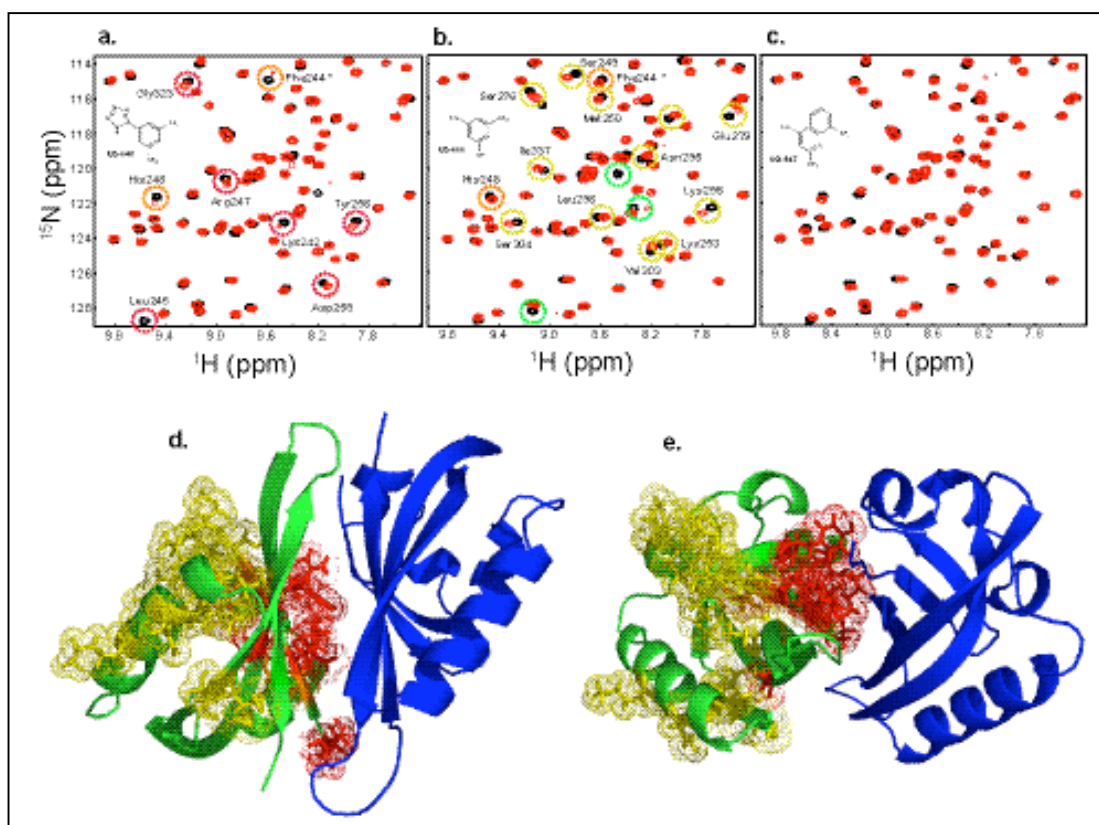
*viii. Alternative explanation for heterodimer disruption*

The compounds shown in Fig 5-8 (KG-548, KG-655 and KG-547) were not identified as positive hits to HIF-2 $\alpha$  PAS-B in the initial screening protocol, so deconvolution of the screening mixtures and subsequent titrations were not performed. Nevertheless, control  $^{15}\text{N}/^1\text{H}$  HSQC experiments of 250 $\mu\text{M}$   $^{15}\text{N}$ -labeled HIF-2 $\alpha$  PAS-B with and without 1mM of KG-548, KG-655 and KG-547 were recently performed to make sure that the disruption effect was due exclusively to compound binding to ARNT PAS-B, and not because of transient interactions with the HIF-2 $\alpha$  PAS-B domain. These experiments showed that at high concentrations, the compounds KG-548 and KG-655 do cause minor changes in the spectrum of HIF-2 $\alpha$  PAS-B alone, as evidenced by broadening of certain peaks and changes in the chemical shifts of others (Fig 5-11 a,b). Interestingly, addition of compound KG-547 does not significantly perturb HIF-2 $\alpha$  PAS-B backbone amide resonances, especially with regard to peak broadening from intermediate exchange events (Fig 5-11 c).

Changes to the spectra of HIF-2 $\alpha$  PAS-B upon addition of KG-548 and KG-655 at first seemed relatively inconsequential, and did not cause problems with the deconvolution of spectral changes originating from the addition of ligand and from the formation of complex. Nevertheless, affected HIF-2 $\alpha$  PAS-B residues defined as either broadening significantly upon addition of compound or exhibiting chemical shift changes (equal to at least half of the peak linewidth) were mapped onto the high-resolution

structure of HIF-2 $\alpha$  PAS-B to see if localized regions affected by the two compounds could be identified. The results are shown in Fig 5-11 d,e, which shows two views of the HIF-2 $\alpha$ /ARNT PAS-B complex determined by HADDOCK, with residues affected by KG-548 shown in red, and those affected by KG-655 shown in yellow. Surprisingly, there was very little overlap in residues affected by these compounds, with only two similarly perturbed backbone resonances identified by the previously described criteria (Phe244 and His248 –shown in orange in Fig 5-11 d,e).

**Fig 5-11. Compounds KG-548 and KG-655 bind transiently to HIF-2 $\alpha$  PAS-B.** a.  $^{15}\text{N}/^1\text{H}$  HSQC spectra of 200 $\mu\text{M}$  HIF-2 $\alpha$  PAS-B with (red) and without (black) 1mM KG-548. Transient interactions are evidenced by peak shifting and/or broadening. b. Same experiment with KG-655 showing similar effects as KG-548, but affecting different residues on HIF-2 $\alpha$  PAS-B. c. Same experiment with KG-547, showing very few spectral changes, and significantly less peak broadening indicating much less interaction with the domain. d and e. Two views of the HADDOCK generated HIF-2 $\alpha$ /ARNT PAS-B complex, with KG-548 residues perturbed by KG-548 shown in red, and those affected by KG-655 shown in yellow. Two similarly affected residues (Phe244 and His248) are shown in orange. Both compounds affect heterodimerization (to a greater extent than KG-547), but bind to HIF-2 $\alpha$  PAS-B in markedly different locations, perhaps suggesting multiple modes of action (see text).



These results show that caution should be used when interpreting NMR data that reports on the binding of heterocyclic compounds to proteins. Although KG-548, KG-655 and KG-547 all bind to ARNT PAS-B in a similar manner as evidenced by comparable chemical shift changes of particular backbone amide resonances, their varying effects on heterodimeric formation led us to pursue more carefully designed

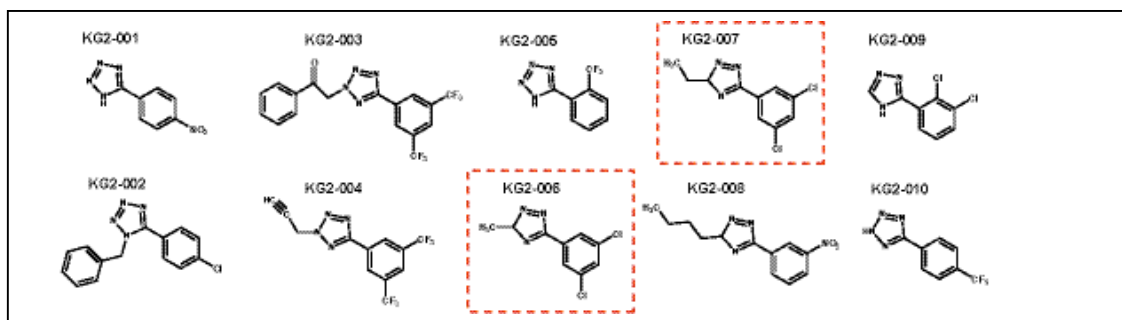
control experiments before assuming a particular mode of action for these compounds. The transient interactions that the compound KG-548 appears to have with the binding interface of the HIF-2 $\alpha$  PAS-B domain is a likely mechanism by which it attenuates complex formation. However, this explanation doesn't seem to hold for KG-655, whose interactions with HIF-2 $\alpha$  PAS-B, are centralized to regions on the other side of the binding interface, unless these interactions are causing a global conformational change in the protein. This model does explain the variable effects seen for the nuclear extract Co-IP experiments done in Rick Bruick's lab described in section *v.b.*. It was shown in Fig 5-9 that although KG-548 was effective at disrupting the complex between full-length HIF-2 $\alpha$  and ARNT, the compound KG-655 had little or no effect. This could easily be explained by the variable modes of compound binding to the HIF-2 $\alpha$  PAS-B domain with KG-548 attaching itself to an important binding interface within the full-length molecule, and KG-655 interacting in a functionally distant location. This is hard to reconcile with the *in vitro* NMR data, however, because both compounds seem relatively effective at partially disrupting the interaction. However, the higher affinity for KG-655 for ARNT PAS-B relative to KG-548 could amplify any ARNT-mediated effects on heterodimer disruption, so the overall perturbation of complex formation in both isolated PAS domains and in full-length proteins might be the result of contributions from both types of PAS/compound interactions.

#### *ix. Structure activity relationships (SAR) of ARNT PAS-B small molecule ligands*

The compounds KG-548 and KG-655 have been shown to bind the PAS-B domain of ARNT, and to varying extents disrupt homo- and heterodimerization of

isolated domains and full-length proteins. However, lead compounds meant for development as therapeutics require many characteristics that must be optimized in order for them to be efficacious, including high affinities for their targets. In an attempt to increase the affinities of ARNT PAS-B binding compounds and to understand in more detail the characteristics of our two lead compounds that allow association with the domain, we ordered more commercially available compounds based on the presence of one or more benzyl rings with a variety of substitutions. Structures of ten compounds screened for binding to ARNT PAS-B using methods described earlier are shown in Fig. 5-12.

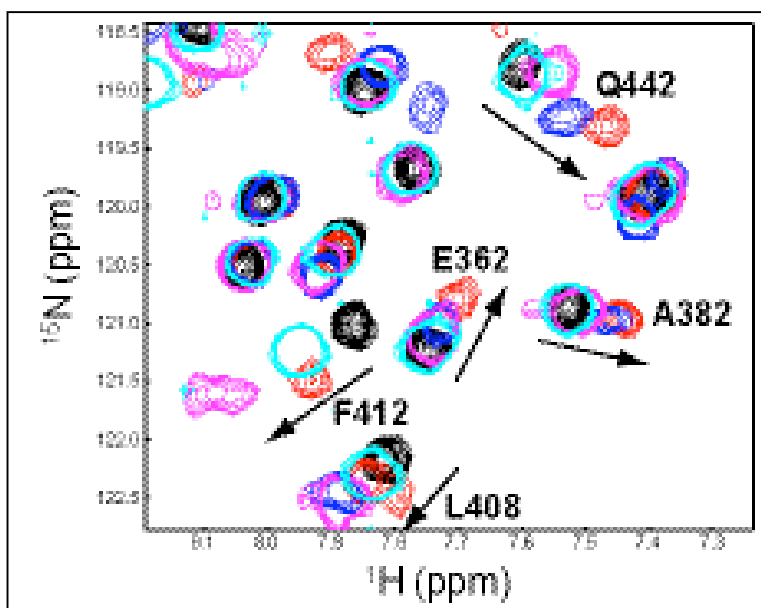
**Fig 5-12. Initial SAR screen of ARNT PAS-B using commercially available compounds whose structures are based on the 3,5 di-substituted benzyl ring present in compounds KG-548 and KG-655. The compounds with red boxes (KG2-006 and KG2-007) were the only compounds shown to appreciably bind to ARNT PAS-B (see Fig 5-13).**



From the ten compounds above, some interesting information was obtained regarding the structural basis of KG-548 and KG-655 binding, and provided some insight into the steric restrictions required for binding to ARNT PAS-B in this type of small molecule ligand. Only two of the compounds shown above were found to cause any significant changes in the spectrum of ARNT PAS-B upon the addition of 500 $\mu$ M

compound to 200  $\mu\text{M}$   $^{15}\text{N}$  labeled ARNT PAS-B in  $^{15}\text{N}/^1\text{H}$  HSQC experiments (KG2-006 and KG2-007 - red boxes). Both molecules have two chlorine atoms substituted for the tri-fluoro methyl groups of KG-548 and KG-655, along with minor additions to the tetrazole ring. This suggests that the tri-fluoro methyl groups are not an absolute requirement. However, the placement of groups on the central benzyl ring does appear to be important, because molecules that contain chloro- or tri-fluoro methyl groups located in alternate positions on the aromatic ring (KG2-005, KG2-009 and KG2-010) do not seem to bind, despite their otherwise identical structures relative to KG-548. The sizes of additional groups added to the tetrazole ring are also important, because although small additions such as methyl (KG2-006) or ethyl (KG2-007) groups appear to be tolerated, the larger substitutions found in KG2-003 and KG2-004 seem to preclude binding. This provides further indirect evidence that the compounds are acting within the core of ARNT PAS-B, since there is a clear correlation between steric hinderance and binding. An overlay of a small region of the spectra of ARNT PAS-B alone and with each of the four related compounds so far discussed is shown in Fig. 5-13.

**Fig 5-13. Compounds found with the SAR screen bind to ARNT PAS-B in a similar manner to KG-548 and KG-655, based on comparable chemical shift changes.  $^{15}\text{N}/^1\text{H}$  HSQC spectra were acquired on 250  $\mu\text{M}$  of ARNT PAS-B alone (black), and after addition of 500  $\mu\text{M}$  of compounds: KG-548 (red), KG-655 (blue), KG2-006 (magenta), and KG2-007**



It is clear that many peaks are shifting upon addition of all four compounds, and that many of the chemical shift changes are similar, although it should be noted that the interactions are not necessarily saturated. This further suggests that we have identified a specific interaction mediated by common structural elements on each compound. It also further supports the idea that very little additional steric hinderance is allowed beyond the original compound KG-548, because the addition of a single additional carbon bond in KG2-007 (cyan peaks) relative to KG2-006 (magenta peaks) reduces its affinity for ARNT PAS-B, as suggested by the generally smaller chemical shift changes (Fig 5-13).

#### *x. Conclusions*

Although these results are encouraging with regard to the potential of small molecules to modulate protein-protein interactions, it is clear that additional studies are necessary from both structural biology and biochemistry efforts to further characterize these interactions and their effects. Significant changes in the spectra of ARNT PAS-B are seen with the two new compounds KG2-006 and KG2-007, but their effects on the dimerization of the PAS domain are minimal when compared with the original two compounds (KG-548 and KG-655 – data not shown). This could be due to issues with overall solubility, or steric hinderance due to the introduction of additional carbons to the tetrazole ring. Nevertheless, general guidelines regarding the important aspects of these chemicals for binding to the ARNT PAS-B domain can be drawn from the structure/activity relationships discovered through this screening process. Further studies must also be done to improve the optimal conditions of chemical compounds such as these in order to establish any potential as lead compounds for the development of drugs



that could be used *in vivo* to modulate the hypoxia response for the treatment of diseases such as cancer. In the future, studies will move toward more directed structure-based modifications through the organic synthesis of new compounds, and to investigate other classes of compounds identified by this, and future screening procedures.

## VI. Larger ARNT Constructs

### *A. Limitations of the reductionist approach*

One of our main goals in this work was to reconcile the promiscuous activity of ARNT with its ability to form functional dimeric complexes with multiple bHLH-PAS proteins. So far, this investigation has focused on the PAS-B domain, due to its stable behavior in solution and the relative ease in which its hetero- and homodimeric associations have been identified and characterized. Although the use of this reductionist approach has yielded intriguing clues about how this domain might interact with itself and the PAS-B domain of HIF-2 $\alpha$ , the question remains whether the associations studied for the isolated PAS-B domains are identical to interactions occurring in the context of full-length protein in a cellular environment.

The crystal structure of the Per PAS-AB homodimer discussed earlier (5) could provide important clues regarding PAS-mediated interactions involving proteins with multiple PAS subunits. But several lines of evidence suggest that different modes of association are possible for these ubiquitous domains within the context of different biological pathways. Although high-resolution heterodimeric PAS domain complex structures involving PAS domains have been elusive, it seems likely that they participate in and/or regulate interactions with many structurally diverse targets. Their widespread use in many types of proteins of varying structure and function from organisms in all three kingdoms of life implies that evolution has developed many different types of associations utilizing these domains, despite their high degree of structural conservation. In addition, the existence of many eukaryotic proteins that contain multiple tandem PAS domains suggests that function could have evolved through the addition of modular

components that could participate in intra- or intermolecular interactions for the enhancement (or attenuation) of protein function. There is also growing evidence in many systems such as the phototropins and PYP (27, 36), that additional structural elements flanking core PAS domains are often required for proper interactions and/or function (other examples discussed in the Chapter I). For these reasons, it is reasonable to assume that many insights into the specific functions of PAS domain modules within the context of full-length proteins will require studies on constructs that extend beyond individual domains.

Although many different types of complexes likely exist between PAS-containing proteins, the very similar domain architectures of the bHLH-PAS transcription factors imply that generalizations about their formation and their basis for specificities in the upregulation of transcriptional activation might be possible from the study of one or two isolated systems. In addition, the fact that ARNT is involved in the formation of several of these complexes supports the idea that many transcription factors in the bHLH-PAS family could form in much the same way. Because the ARNT protein is a promiscuous central player in these systems, and is also known to homodimerize, it represents a good subject for studying a variety of bHLH-PAS interactions.

*i. Initial expression studies of larger ARNT constructs*

To better understand the interactions that are important for ARNT-mediated regulation of biochemical pathways such as the hypoxia response, we have begun to investigate larger constructs that include the bHLH and PAS-A domains both alone and in tandem with PAS-B. Initially, efforts were made to express the PAS-A domain of

ARNT alone and in tandem with the PAS-B domain. The final results of these efforts to date are outlined in more detail below, but initial results showed that any construct containing the PAS-A region suffered much lower expression levels than PAS-B alone. The ARNT PAS-B construct encoding residues 356-470 used for structure determination (Chapter III) showed post-purification expression levels in excess of 30 to 60 mg/L in  $^{15}\text{N}$ -enriched minimal or LB media respectively, and had strong induction of protein expression upon addition of IPTG. In contrast, PAS-A and PAS-A/B constructs showed no discernable IPTG-induced expression on an SDS-PAGE gel (data not shown). As such, we initially focused on increasing the expression levels of PAS-A containing constructs, starting by cloning various PAS domain regions which included the bHLH domain with the hope that this would increase expression levels and/or stabilize the PAS-A domain to impart protection from proteolysis.

PCR fragments containing the bHLH domain, in addition to the predicted PAS-A region of ARNT (residues 1-350, 85-350, 90-350, 95-350) were cloned into G $\beta$ 1, His $_6$ -G $\beta$ 1 N-terminal fusion vectors and transformed into HMS 174 (DE3) *E.coli* cells for expression. Initial studies showed undetectable expression levels by SDS-PAGE (data not shown). These constructs were also transformed and expressed in the Rosetta and RIL cell lines, which are enriched in several eukaryotic tRNAs that are normally at low abundance in standard *E.coli* cells. The use of these cell lines showed no appreciable improvement in protein expression (data not shown). These constructs were also extended to include the PAS-B domain, and fragments encoding residues 90-470 (bHLH-PAS-AB) and 105-470 (HLH-PAS-AB) were screened for expression in the cell lines described above. The choice of the N-terminal residue 90 was made based on previous

studies that showed functional bHLH dimerization with constructs beginning at this residue (80) and constructs beginning at residue 105 (beyond the basic DNA binding region) were also tested. None of these constructs showed any improvement in expression levels, leading us to believe that a likely explanation was the inherent instability and subsequent proteolysis of constructs that include the PAS-A domain.

JPred secondary structure prediction reveals that the PAS-A regions of all three isoforms of HIF (HIF-1,2,3 $\alpha$ ) and ARNT are significantly larger in size than their PAS-B domains. Although all the secondary structural elements expected for PAS domains are present in the PAS-A regions of these two proteins, they are separated by longer, presumably unstructured loop elements relative to the PAS-B domains (Fig. 6-1 b,c).

**Fig. 6-1. a. Diagram of all ARNT constructs cloned and expressed for the identification of optimal domain boundaries. The PAS-B construct for which the high-resolution structure was determined is circled in red, and PAS-A and PAS-A/B constructs that were further characterized by NMR are circled in green. b/c. Secondary structure alignments of PAS-B regions (as determined from high-resolution solution structures (8)) and PAS-A regions (as determined by JPred analysis) from b. HIF-1,2,3 $\alpha$  and c. ARNT. Gaps in the PAS-B sequences indicate extended loop regions in PAS-A relative to PAS-B, which can be as large as 24-28 residues in the F $\alpha$ -G $\beta$  and G $\beta$ -H $\beta$  loops, respectively.**



This is especially true for ARNT PAS-A, where an additional 60 residues are predicted to exist within unstructured loop regions between the F $\alpha$ , G $\beta$ , H $\beta$ , and I $\beta$  secondary structural elements. The possible function (if any) for these regions is unknown, but could easily provide the basis for non-specific proteolysis and/or aggregation due to unstably folded segments of constructs containing the PAS-A domain of both ARNT and the HIF-2 $\alpha$ .

Although the results will not be discussed in detail here, HIF-2 $\alpha$  constructs containing the PAS-A region were successfully expressed in our lab at high levels by Dr. Paul Erbel, but  $^{15}\text{N}/^1\text{H}$  HSQC experiments revealed them to be unfolded as evidenced by poor amide  $^1\text{H}$  chemical shift dispersion and heterogenous linewidths (data not shown). Because of the need to address issues related to long-term stability and/or folding of PAS-A domains for effective structural studies, we decided to focus initially on constructs of ARNT that contain only the PAS-A and PAS-B regions before continuing to address larger constructs containing the DNA-binding bHLH domains.

#### *B. ARNT PAS-A domain*

The ARNT PAS-A domain is immediately preceded by approximately 12 residues which are predicted to be  $\alpha$ -helical by JPred analysis. Such N-terminal helices are often found flanking either termini of PAS domains, and have been implicated in other systems as key structural components that provide much of the buried surface area needed for homodimeric PAS domain complexes (discussed in detail in Chapter III). Constructs were therefore designed that include this region in addition to those predicted to form the standard PAS domain secondary structural elements.

Despite results that indicated a much lower expression yield of protein fragments containing the PAS-A domain relative to PAS-B alone, we continued to characterize PAS-A constructs for the purpose of optimizing the stability and folded state of this domain for future studies. Of the ARNT PAS-A constructs screened (covering residues 144-350, 144-355, 151-350, 151-355, 155-350, 155-355), the best results obtained so far were from the 144-350 construct, which will be described in more detail.

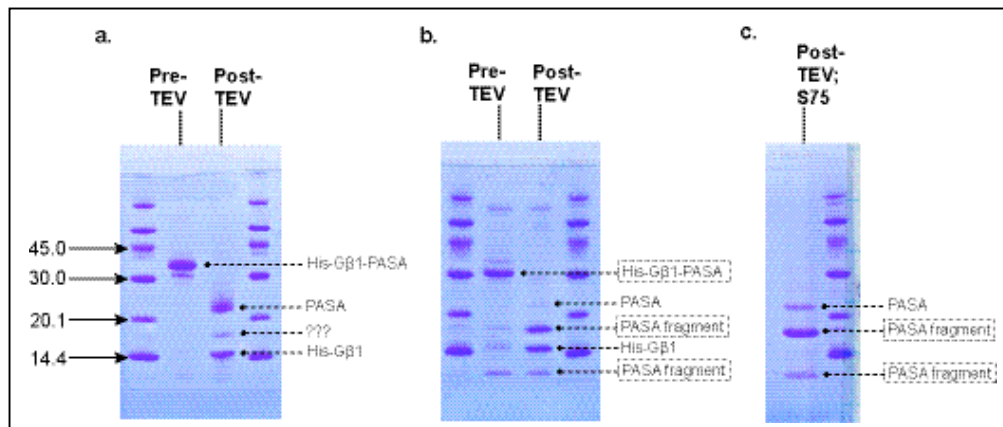
*i. Expression and purification of ARNT PAS-A (144-350)*

PCR fragments encoding this region were cloned into the His<sub>6</sub>-Gβ1 vector and transformed into HMS174(DE3) cells. The His<sub>6</sub>-Gβ1 vector was chosen to allow affinity purification schemes to be used, while taking advantage of the added solubility and expression that often accompanies inclusion of the Gβ1 protein fusion tag (126). As with all vectors used in this investigation, a TEV cleavage site is also present C-terminal to the Gβ1 domain, allowing cleavage and subsequent removal of the tag by affinity chromatography. After overnight growth of a 5ml LB starter culture, the cells were collected and inoculated into 1L LB and grown to OD<sub>600</sub> = 0.55, induced with 120 mg IPTG and incubated overnight (14 hr.) at 20°C to a final OD<sub>600</sub> = 0.75. Despite the relatively low cell densities (cultures expressing PAS-A typically showed OD<sub>600</sub> = 1.5-2.0 after overnight induction), we harvested the cells to attempt an initial purification. The cells were cracked by extrusion and clarified by high-speed centrifugation before injection on a Ni<sup>2+</sup>-charged chelating resin column. The protein was eluted with a linear gradient of 50mM Tris, 25mM NaCl, 20-250mM imidazole, pH=7.5, and further purified by injection over a Superdex S-75 gel filtration column. Fractions collected after this



step were concentrated to 1ml, 50  $\mu$ g of previously purified TEV protease was added, and the mixture was incubated at room temperature for 3 hr. SDS-PAGE analysis shows that the purification was reasonably successful, and that cleavage of the fusion tag yielded products similar in size to those predicted for the PAS-A domain and cleaved tag (Fig 6-2 a). The presence of a smaller product at approximately 16 kDa was not expected, but the majority of the protein appeared to be close to the ~23 kDa predicted mass of the ARNT (144-350) construct. After the final injection over the Superdex S-75 column, however, the final sample appeared to be mostly in the proteolytically cleaved state (Fig 6-2 c). In addition, the final yield from this preparation was only about 4mg/L, well below the amounts typically obtained from of ARNT PAS-B (30-40 mg/L), but sufficient for NMR studies.

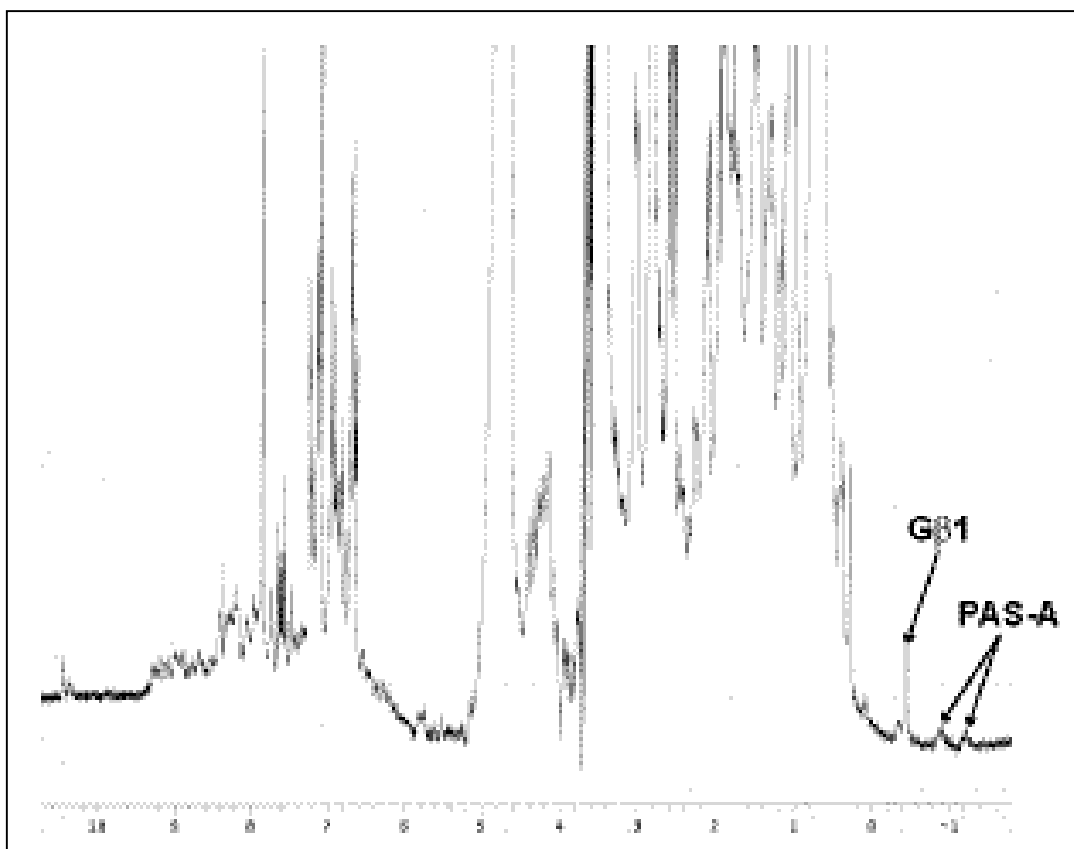
**Fig 6-2. SDS-PAGE analysis of ARNT PAS-A (144-350) samples after purification and cleavage of the His<sub>6</sub>-G $\beta$ 1 expression tag. a. Natural abundance sample of ARNT PAS-A. b. uniformly <sup>15</sup>N-labeled ARNT PAS-A. c. Same sample as b. after cleavage and further purification using size-exclusion chromatography.**



*ii. NMR spectroscopy of ARNT PAS-A (144-350)*

One-dimensional proton NMR spectroscopy was performed on the unlabeled sample before cleavage of the His<sub>6</sub>-Gβ1 tag to see if any evidence of folded protein was detected. Chemical shifts of sidechain methyl protons are often a good indicator of the folded state of a protein because of several factors including: the large signal-to-noise ratio of sidechain methyl protons due to fast rotational correlation times, the heterogeneous environments often sampled by methyl groups within the folded cores of proteins, and the lack of crowding in the region of the spectrum typically populated by folded methyls (-1 to 1 ppm <sup>1</sup>H). The resulting spectrum is shown in Fig. 6-3, which reveals peaks around 0-1 ppm, indicating methyl groups that exist within the context of a folded protein environment and that are not derived from the His<sub>6</sub>-Gβ1 fusion tag.

**Fig 6-3. 1D  $^1\text{H}$  NMR spectrum of the natural abundance ARNT PAS-A sample shown in Fig. 6-2 a. (pre-TEV). The arrows indicate peaks originating from methyl protons from the expression tag His<sub>6</sub>-Gβ1 and PAS-A. The upfield locations of these peaks indicate that the methyl groups are in a folded environment, indicating that some tertiary structure is present in the PAS-A domain.**



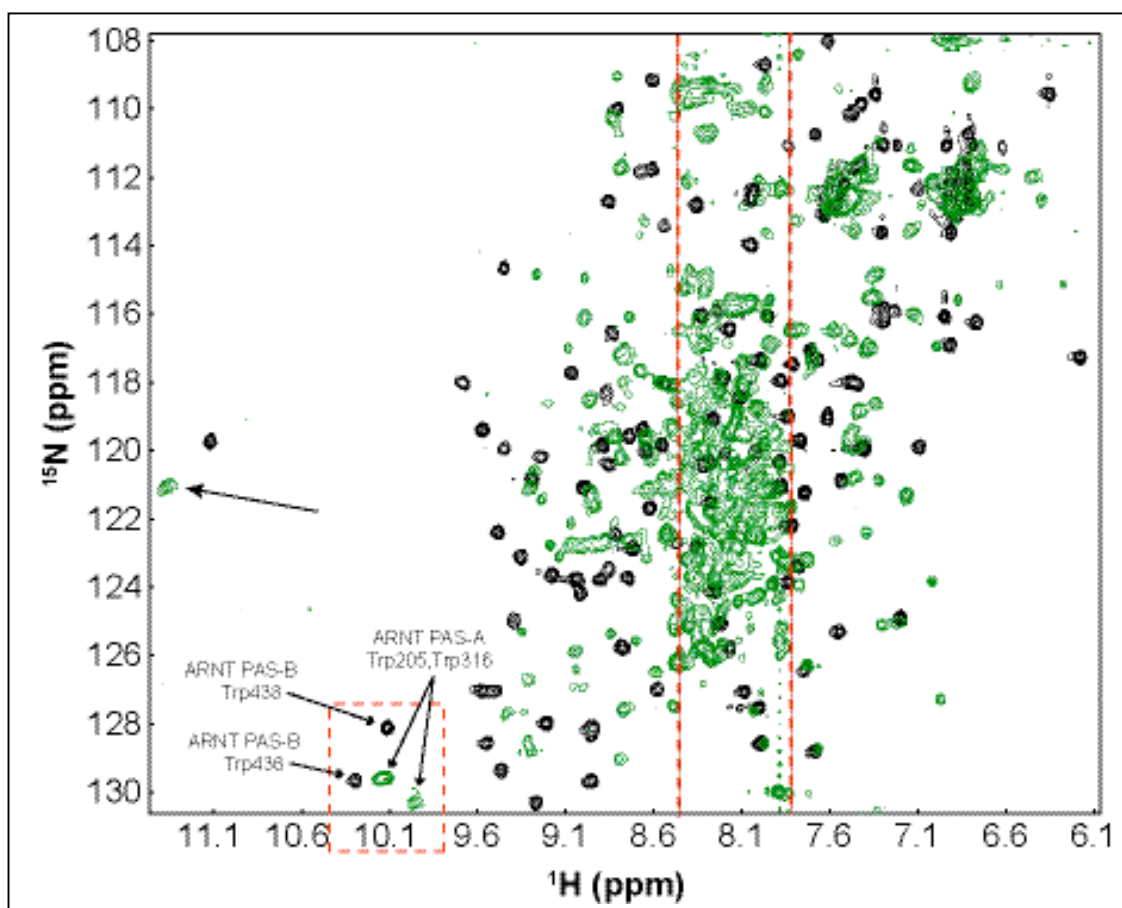
This result encouraged us to produce  $^{15}\text{N}$ -labeled ARNT PAS-A to further investigate its folded state. Protein was grown and prepared in the same manner as before, but cells from the 5ml starter cultures were used to inoculate 4L M9 minimal growth media enriched with  $^{15}\text{NH}_4\text{Cl}$  for uniform labeling. The results of the initial purification step before and after cleaving the His<sub>6</sub>-Gβ1 fusion tag are shown in Fig 6-2 b. Although the overall protein preparation is not as clean as shown for the natural abundance sample in Fig. 6-2 a, this is to be expected because in this case the gel

filtration step was not used prior to NMR studies. Another difference that was evident in this preparation relative to the previous one was the presence of only a very faint band representing the correct size for the PAS-A construct, and a much darker band at the ~16kDa position observed in the natural abundance preparation (data not shown). This indicates that proteolysis is occurring, and at a greater rate than the natural abundance sample. This faster rate for the proteolysis is reasonable since the  $^{15}\text{N}$  sample had not yet been additionally purified over the gel filtration column, but further characterization of the natural abundance sample, however, revealed that although the initial rate of proteolysis that occurred with this sample was lower, the degradation continued until nearly all of the proper size natural abundance product was gone after a single night at 4°C (data not shown). These data suggest not only that the PAS-A domain is subject to proteolysis, but also that a naturally occurring protease in this strain is likely co-purifying with the PAS domain.

Further attempts to relieve this unwanted cleavage including expression in the BL21(DE3) *E.coli* cell line (which has been genetically engineered to remove certain proteases), purification of the PAS domain in the presence of high salt and/or protease inhibitors, and further ion-exchange chromatography to remove the protease have yet to be fully effective at attenuating this problem. Nevertheless, this cleavage event appears to be occurring in a localized area within the protein, most likely in one of the extended loops, and the final cleavage products seem to be stable over reasonable periods of time (days). This was determined after a further gel filtration step on the  $^{15}\text{N}$ -labeled sample, after which the sample sat overnight and was run again on an SDS-PAGE gel (Fig 6-2 c). In addition, the cleavage products seem to be co-eluting with the residual uncleaved

protein after injection on both S75 and S200 gel filtration columns. This could indicate that the cleaved products are interacting and folding together properly into its native PAS domain confirmation. Precedence for the proper folding of cleaved fragments was previously observed for the N-terminal SH2 domain from the p85 $\alpha$  subunit of phosphatidylinositol 3' kinase, where two cleaved fragments of the domain (9 and 5 kDa) were demonstrated by NMR to be properly folded as compared to the uncleaved domain (176, 177). Unfortunately, attempts to obtain mass spectrometry data in an effort to map the cleavage site have so far been unsuccessful, with no unambiguous peaks yet identified that can be added together to obtain the proper mass of the full domain. Edmond degradation-based sequencing of the cleaved fragments will probably therefore be necessary to identify the specific site of proteolysis. Mutations can then be made in the primary sequence to prevent site-specific proteolysis if this issue cannot be resolved by variation of the buffer conditions.

**Fig 6-4.**  $^{15}\text{N}/^1\text{H}$  HSQC spectra of ARNT PAS-B (356-470 –black) for which a high-resolution structure was determined, and ARNT PAS-A (144-350 - green). The spectra are overlaid for comparison, and the region of the spectra typically associated with the chemical shifts of backbone amides in a random coil conformation is shown by red dashed lines. A far downfield peak that has often been observed in other PAS domains is marked by an arrow, indicating that a typical PAS domain fold might be occurring despite proteolysis, and the presence of two peaks in the region of the spectra where tryptophan sidechain indole NH groups typically reside (red box) matches the predicted number of tryptophans present in the primary sequence of the ARNT PAS-A (144-350) construct.



To determine if the  $^{15}\text{N}$ -labeled PAS-A domain exhibited folded properties despite the unwanted proteolytic cleavage, the final sample (Fig. 6-2 c) was concentrated to approximately 160  $\mu\text{M}$  and  $^1\text{H}$ - $^{15}\text{N}$  HSQC experiments were performed. The results are shown in Fig. 6-4 as an overlay of the well-behaved spectra of the PAS-B domain of ARNT (black) with that of the proteolyzed PAS-A preparation (green). It is clear that

many peaks arising from the PAS-A sample appear in the folded region of the spectrum (outside red region), and that certain resonances correlate well with those expected for this construct. For example, two tryptophans (Trp205, Trp316) are contained within the predicted ARNT PAS-A domain, and the two peaks at around 10.0 ppm (red box) are consistent with the chemical shifts typically observed for tryptophan side-chain indole HN resonances (compare with two black peaks in the same region representing the PAS-B tryptophan side-chain amide signals). This suggests that the observed spectrum originates from the proper construct (144-350). More importantly, the far downfield peak at ~11.3 ppm gives further evidence that this in fact represents a relatively well-folded PAS domain. A single peak in this region has been observed for several PAS domains in our own lab and by other groups, and originates from a single backbone amide that is positioned directly over the macrodipole created by the F $\alpha$  helix, shifting this peak far downfield of where most backbone amide resonances typically reside. While we cannot unambiguously identify which residue gives rise to this peak without experiments to assign the backbone chemical shifts, an alignment of predicted secondary structural elements suggests that this peak most likely represents the backbone amide of residue Asp217.

Examination of the  $^{15}\text{N}/^1\text{H}$  HSQC of the PAS-A construct shows that many peaks are found with poor  $^1\text{H}$  chemical shift dispersion and variable intensities, suggesting that they originate from unfolded residues. This was expected given the large number of residues that make up the presumably unstructured loops between secondary structural elements (Fig 6-1 c). It is possible that the peaks representing residues in the folded state might arise from the small population of protein that has not been proteolytically cleaved,

and in fact many of these show significantly lower peak intensity than the corresponding PAS-B peaks (Fig 6-4). However, several peaks including the diagnostic PAS domain peak at ~11.3 ppm also show sufficiently large peak intensities relative to the PAS-B spectra, so it would be unlikely that these peaks arise solely from the small percentage of uncleaved protein that is still present according to SDS-PAGE analysis (Fig 6-2 c, Fig 6-4).

The PAS-A construct also appears to suffer some solubility issues at high concentrations, and the highest estimates of the concentrations of NMR samples for this construct never exceeded 180-200 $\mu$ M. Although this concentration is more than sufficient to obtain quality spectra, other issues such as the larger size of PAS-A relative to PAS-B, dimerization and/or aggregation due to the large preponderance of unfolded residues could all contribute to peak broadening and lower signal intensities.

### *iii. Relevance to other multi-PAS proteins*

These results are encouraging in light of previous results in our lab, which indicate that proteins with two distinct PAS domains in tandem often have one PAS domain that is well-folded and soluble whereas the other shows poor solution behavior. For example, White Collar-1 contains three PAS domains, WC-1A, WC-1B and WC-1C. Of the two that were initially characterized in our lab (WC-1A and WC-1B), WC-1B showed reasonable solution behavior after solvent screening, and the other, WC-1A, was consistently unfolded. This also applies to HIF-2 $\alpha$ , where the PAS-B domain proved amenable to high-resolution structure determination, and the PAS-A domain always gave spectra indicating a more or less random coil structure (data not shown). It should also



be pointed out that the N-terminal domain in multi PAS proteins is not always problematic. In the initial characterization of murine NPAS2 done in our lab, the PAS-A domain (78-240) was sufficiently well-behaved to obtain backbone chemical shift assignments, and the C-terminal PAS-B domain was unfolded (159). If this trend is prevalent among many multi-PAS proteins, it could pose serious problems to structural biologists who would like to characterize intramolecular interactions between PAS domains in addition to the intermolecular associations involved in the formation of multi-protein complexes. Although our ARNT PAS-A domain construct still suffers from site-specific proteolysis, which might adversely affect its overall behavior in solution, we continued to investigate the possibility of these higher order interactions by cloning larger protein constructs that contain both the PAS-A and PAS-B domains of ARNT.

### *C. Multi-PAS domain construct: ARNT PAS-AB*

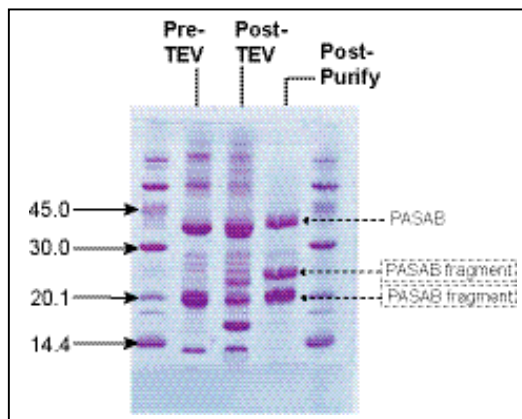
Initial attempts have been made to obtain information on an ARNT PAS-A/PAS-B construct that was cloned into the His<sub>6</sub> vector using methods described earlier. The choice of this vector was made due to an EcoRI site that is present in the ARNT PAS-B domain, precluding the effective use of the His-Gβ1 vector because of fewer acceptable cloning sites. The construct used for these studies also begins at residue 155 instead of 144 because positive results were obtained for this construct initially, and also because the choice of this N-terminus reduces the number of predicted unstructured residues preceding the pre- $\alpha$  helix of ARNT PAS-A.

*i. Expression and purification of ARNT PAS-AB (155-470)*

After sequencing confirmed that the cloned construct was in frame and covered the desired residues, uniformly  $^{15}\text{N}$  labeled protein was expressed in 2 X 1L of  $^{15}\text{N}$ -enriched M9 minimal media by IPTG induction of log-phase cultures ( $\text{OD}_{600} = 0.5$  and  $0.8$  respectively) at  $20^\circ\text{C}$  overnight. Affinity-based protein purification proceeded as previously described. TEV cleavage of the affinity tag produced a fragment roughly corresponding in size to the expected 36,902 Da PAS-AB (155-470) protein, suggesting that the desired product had in fact been isolated (Fig 6-5). For initial studies, however, the  $\text{His}_6$  tag was retained due to the unlikelihood that this small tag would create any serious spectral overlap issues, and also to obtain spectra as soon as possible in an attempt to avoid proteolysis.

Injection of the  $\text{His}_6$ -PAS-AB construct over a Superdex S75 gel filtration column removed many impurities, but again revealed two prevalent protein bands that, when added together, corresponded roughly in size to the predicted full-length construct (Fig 6-5). This indicates that although inclusion of the well-behaved PAS-B domain did not completely protect the PAS-A from proteolytic cleavage, the fragments were fairly stable and amenable to characterization using NMR methods.

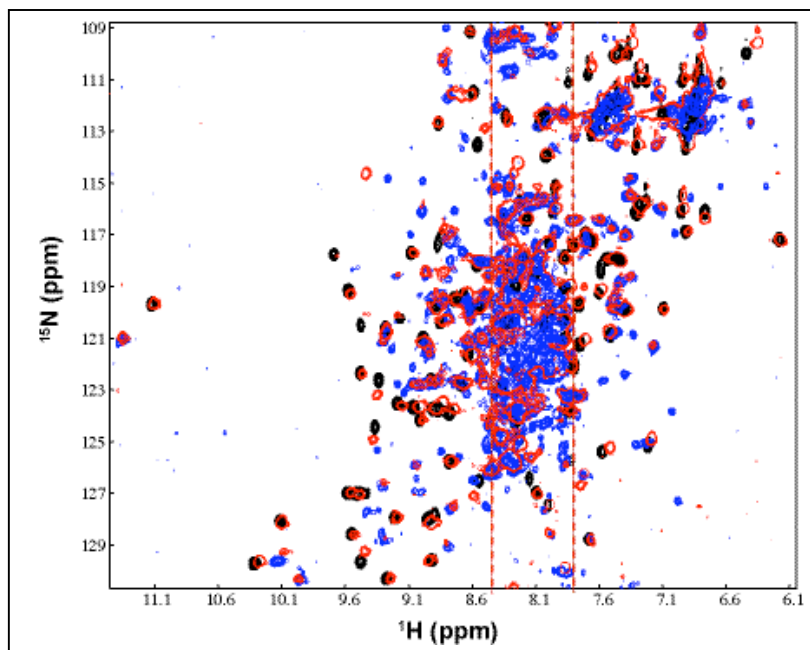
**Fig 6-5.** Uniformly  $^{15}\text{N}$ -labeled samples of ARNT PAS-AB (155-470) were expressed, purified and characterized by SDS-PAGE analysis. A small sample of the Superdex S75-purified sample was TEV digested to remove the tag in order to confirm vector-derived expression of the PAS-AB construct, and the rest of the uncleaved protein was further purified for NMR studies. Yields were similar to those obtained from the ARNT PAS-A studies (4-6mg/L).



ii. NMR of ARNT PAS-AB (155-470)

After concentrating the purified, His<sub>6</sub>-PAS-AB (155-470) construct (third lane, Fig 6-5), <sup>15</sup>N/<sup>1</sup>H HSQC spectra of this protein was obtained, which indicated the presence of both PAS-A and PAS-B domains (Fig 6-6). The peaks previously observed for the PAS-A domain (144-355) are present at the same locations, indicating that the folded state of PAS-A in the multi-PAS construct (155-470) is the same despite lacking the additional 11 N-terminal amino acids that were present in the isolated PAS-A domain (N-terminus = residue 144). Similarly, peaks that represent specific backbone amide resonances in the spectra of ARNT PAS-B alone are also well represented and do not shift significantly. This data suggests that both domains are in similar conformations whether they are expressed alone or in tandem, confirming their identities as individually folded domains.

**Fig 6-6. <sup>15</sup>N/<sup>1</sup>H HSQC spectra of ARNT PAS-B (356-470 – black), ARNT PAS-A (144-350 – blue) and His<sub>6</sub>-ARNT PAS-AB (155-470 – red).**



In addition to the obvious differences in their primary sequences, the similar chemical shift patterns of the PAS-A and PAS-B domains observed either isolated or in tandem supports the idea that the individual domains could have distinct functions. We wanted to investigate the basis for the ability of ARNT to form dimeric complexes with a variety of bHLH-PAS proteins. The group of bHLH-PAS transcription factors that form functional complexes with ARNT (AhR, Sim and HIF $\alpha$ ) do not appear to form homodimeric complexes, unlike ARNT itself. This is one reason why studies of ARNT might yield clues regarding the basis of this promiscuity, by identifying sequence and/or structural characteristics of ARNT that differ from those of other family members.

Perhaps the PAS-B domain was originally added to the most ancient bHLH-PAS protein through a domain duplication event, which eventually allowed a greater range of specificities to develop for the formation of several ARNT-containing transcription factors with variable function. The specialized ligand binding activity of the AhR PAS-B domain implies that C-terminal PAS domains have developed a wide range of functions, and could provide much of the requirements needed for the formation of specific complexes. The PAS-A domains of bHLH-PAS proteins contain extended loop regions (Fig. 6-1), which also could be a likely place for specialized interactions to occur. Alternatively, the PAS-B (and PAS-A) domains of ARNT might merely add interaction affinities, with the formation of specific complexes being regulated solely by the availability of heterodimeric binding partners in the nucleus. This model, however, would not account for the well-characterized, highly specialized ligand binding/activation function of the PAS-B domain in AhR. Whatever the answer, interactions between the PAS-B (and PAS-A) domains of several proteins involved in the formation of these types

of transcription factors will have to be explored to fully understand the biophysical mechanism for specificity in these complexes.

#### *D. C-terminal extensions of ARNT PAS-B and PAS-AB constructs*

As discussed in Chapter III, the crystal structure of the tandem PAS-A/PAS-B homodimeric complex of the Per protein was solved (5), which reveals quite a different type of interaction between the PAS domains than is proposed here for the HIF-2 $\alpha$ /ARNT complex. This structure shows a reciprocal PAS-A/PAS-B association between the subunits, and involves a  $\alpha$ -helical region C-terminal to the PAS-B domain that wraps around the PAS-A domain of the heterodimeric partner, providing a large part of the dimerization interface. Although many lines of evidence suggest that this type of interaction might not apply to the HIF-2 $\alpha$ /ARNT complex (also discussed in Chapter III), we decided to further test this model in our system through the cloning and expression of C-terminal extensions of the ARNT PAS-B domain alone (residues 356-520 and 356-540) and in tandem with the PAS-A domain (residues 155-520 and 155-540). These constructs were expressed as His<sub>6</sub>-tagged fusions and were purified using methods outlined previously. Preliminary results for the shorter of the C-terminal extended constructs (356-520 and 155-520) were very similar to those obtained from the longer ones (356-540 and 155-540), so only the results of the longer constructs will be discussed. The majority of the preparatory work and much of the data analysis was done by Matthew Evans in our lab.

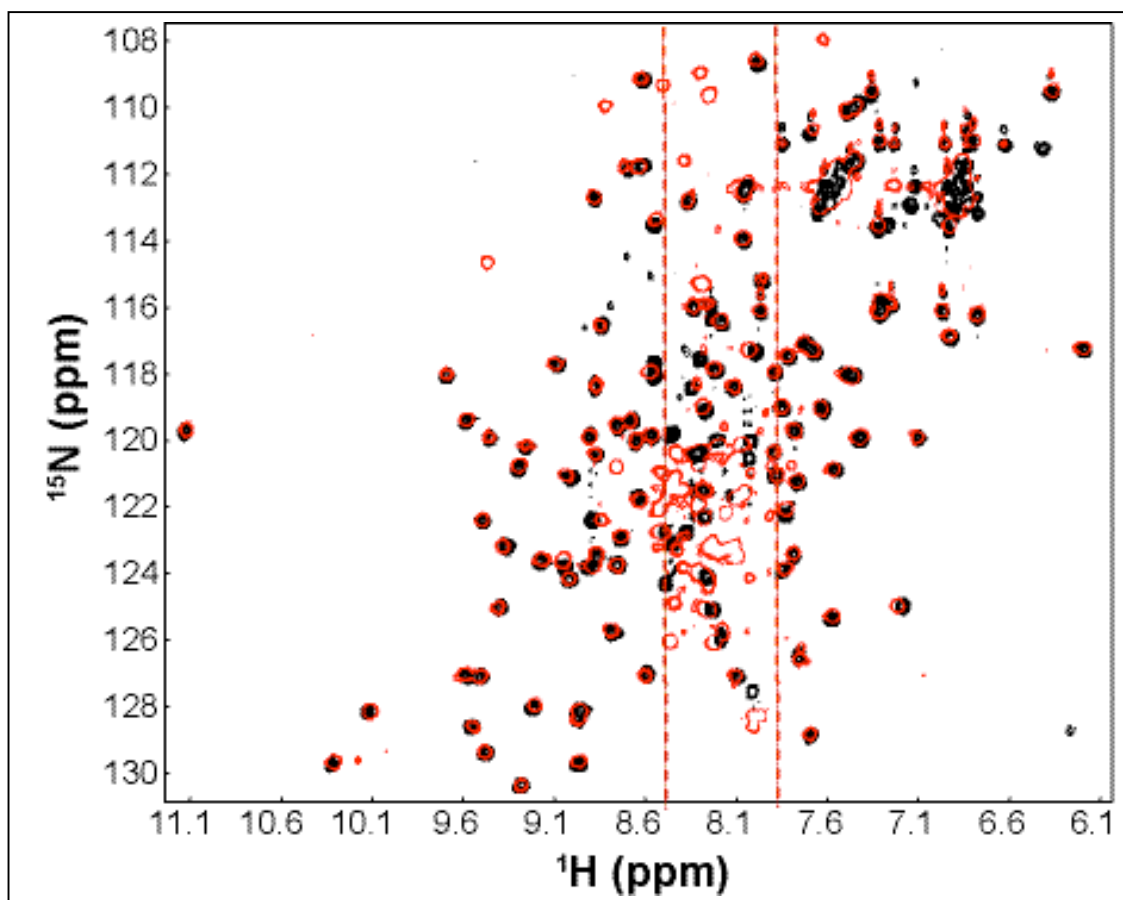
*i. Extended ARNT PAS-B (356-540)*

The cloning and expression of the extended ARNT PAS-B construct (356-540) was carried out using methods outlined previously, and so will not be discussed in detail here. The generation of PCR fragments for cloning into the His6-parallel expression vector was performed using the same 5' primer used for cloning of the shorter PAS-B domain (356-470). New 3' primers were ordered for the generation of fragments that extended the PAS-B domain of ARNT an additional 70 residues compared to the shorter construct to see if additional secondary structural elements might exist that could mediate the interactions seen for the ARNT PAS-B domain.

*a. NMR of C-terminal extended ARNT PAS-B (356-540)*

After the expression and purification of  $^{15}\text{N}$ -labeled samples of the C-terminal extensions of the PAS-B domain alone (construct 356-540),  $^1\text{H}$ - $^{15}\text{N}$  HSQC experiments were performed to determine if any significant changes in the structural conformation of the PAS-B domain could be detected. Results for the 356-540 construct are shown in Fig 6-7, which clearly indicates that the environments of the backbone amide, tryptophan side chain indole, and side chain amine resonances for both of these extended domains have not been significantly altered, suggesting that the overall conformation of the PAS-B domain is similar to that of the shorter constructs. Furthermore, most of the additional signals arising from the extra C-terminal residues (50 more for the 356-520 construct and 70 more for 356-540) appear to lie in the unfolded region of the spectrum, indicating that additional secondary structural elements have not been introduced that form an integral part of the PAS domain fold.

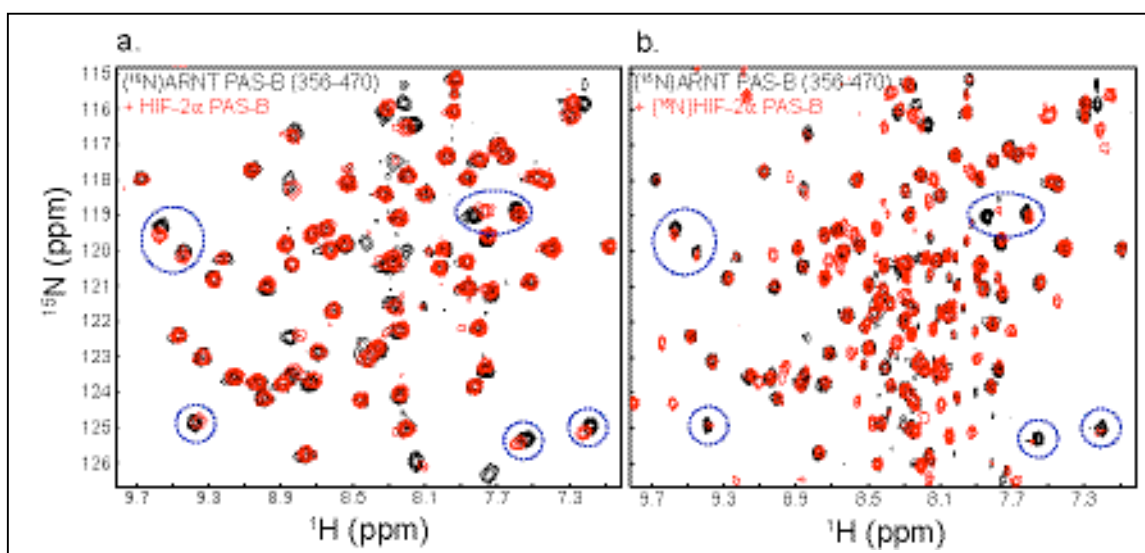
**Fig 6-7. Overlays of the spectrum from the original ARNT PAS-B construct (356-470 – black contours) and the new C-terminal extension PAS-B (356-540 – open red contours). The majority of peaks overlap very well, indicating that no significant perturbation of the ARNT PAS-B structure has occurred due to the 70 extra residues.**



Next we determined if these extended constructs formed similar dimeric complexes compared with the shorter PAS-B domain (356-470). First, complexes were made between the two extended constructs and natural abundance HIF-2 $\alpha$  PAS-B, to determine if similar heterodimeric associations were occurring relative to the shorter ARNT PAS-B domain (356-470). The results are shown in Fig 6-8, and it is clear not only that similar ARNT resonances are being affected, but that the electronic environments of the backbone amides are experiencing similar changes relative to the

original complex formed with the smaller domain, as evidenced by similar patterns of chemical shift changes and/or intermediate exchange peak broadening upon complex formation.

**Fig 6-8. Extended ARNT PAS-B construct (356-540) forms similar complex with HIF-2 $\alpha$  PAS-B.**  
**a.**  $^{15}\text{N}/^1\text{H}$  HSQC of 250 $\mu\text{M}$  ( $^{15}\text{N}$ )ARNT PAS-B (356-470) before (black), and after (red) adding 250 $\mu\text{M}$  unlabeled HIF-2 $\alpha$  PAS-B. **b.** 250 $\mu\text{M}$  ( $^{15}\text{N}$ )ARNT PAS-B (356-540) before (black), and after (red) adding 250 $\mu\text{M}$  ( $^{15}\text{N}$ )HIF-2 $\alpha$  PAS-B. Similar peak broadening and shifting is seen for ARNT PAS-B resonances in both experiments. Additional peaks in (b)(red) observed due to the presence of ( $^{15}\text{N}$ )HIF-2 $\alpha$  PAS-B (to concurrently monitor ARNT PAS-B binding to HIF-2 $\alpha$  PAS-B), and smaller linewidths are observed in (b) due to a larger magnetic field strength (600MHz in (a), 800MHz in (b)).



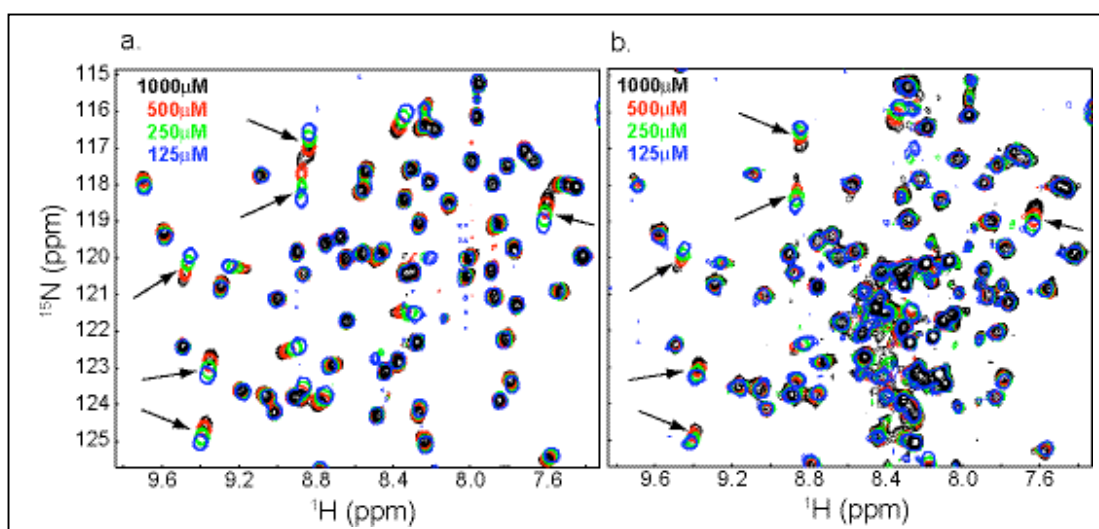
We also wanted to determine if homodimeric complexes are being formed by these extended constructs, and if they are similar to those formed by the shorter PAS-B domain. Serial dilution experiments were performed for the extended construct (356-540) similar to those used to characterize the ARNT PAS-B (356-470) homodimer. The results are shown in Fig 6-9, and reveal again that similar residues on the solvent exposed side of the central  $\beta$ -sheet are being utilized for a homodimeric association, and are being



perturbed in a similar manner to the dimer originating from the original shorter construct (Fig 6-9 a,b).

However, while investigating both the hetero- and homodimeric complexes, it was observed that many signals were broadening to a much greater extent for the larger construct (356-540) relative to the shorter one (356-470), sometimes beyond the range of detection (Fig 6-9 b). This might be explained trivially by the larger size of the larger protein, which will have a larger transverse relaxation rate resulting from a longer rotational correlation time.

**Fig 6-9. Serial dilutions of: a. ARNT PAS-B (356-470) and b. ARNT PAS-B (356-540). The  $^{15}\text{N}/^1\text{H}$  HSQC spectra reveal that similar residues are being affected in both constructs upon serial dilution, indicating that similar homodimeric complexes are being formed.**



Another explanation might be an increase in the affinity of the hetero- and homodimeric associations with the extended constructs relative to the shorter domain, resulting in a shift in the exchange rate of the interaction. In fast exchange, the observed chemical shifts represent the average electronic environments of the entire population of backbone amides that exist in varying proportions of both the apo and bound states

depending on the overall concentration of the sample in a particular experiment. Intermediate exchange is characterized by the broadening of peaks due to the continuous dephasing of magnetization in the transverse plane as a result of the variable Larmor frequencies of individual backbone amide resonances in the apo and bound states. A transition from fast to intermediate exchange could therefore mean that the interaction has become one of higher affinity, assuming that the on-rate ( $k_{on}$ ) remains diffusion limited. However, if the new longer constructs are aggregating or participating in non-relevant interactions that artificially keep  $k_{on}$  lower than that expected from free diffusion in solution, then intermediate exchange broadening effects can be observed even if the affinity of the interaction has not changed (or even if it has been reduced). This is quite possible in our case, because the presence of long, presumably unfolded extensions C-terminal to the putative PAS-B domain could be artificially reducing the on-rate due to non-specific electrostatic and/or hydrophobic interactions between the unfolded residues.

#### *b. Rotational correlation time determination*

To determine if the C-terminal extensions could be involved in the ARNT PAS-B homodimerization resulting in a tighter interaction, rotational correlation time estimates were determined from longitudinal ( $T_1$ ) and transverse ( $T_2$ ) relaxation time measurements taken on the original ARNT PAS-B (356-470) and the longer of the two C-terminal extension constructs (356-540) at two different concentrations (250 $\mu$ M and 950 $\mu$ M). Pulse sequences that contain a delay time parameter were used to quantify the effects of  $T_1$  and  $T_2$  relaxation effects by taking a series of experiments with variable relaxation delays prior to acquisition. Using a program written by Qiong Wu based on the Lipari-

Szabo method (178, 179), the ratio of the  $T_1$  and  $T_2$  relaxation times were used to obtain estimates of the overall isotropic rotational correlation times, and the results are shown in Table 6-1.

<b>Table 6-1</b>		<b>(356-470) - 250<math>\mu</math>M</b>		<b>(356-470) - 1mM</b>	
trim factor	trimmed	trim factor	trimmed	trim factor	trimmed
5%	7.69 nsec	5%	11.42 nsec	5%	11.42 nsec
10%	7.67 nsec	10%	11.43 nsec	10%	11.43 nsec
untrimmed	7.72 nsec	untrimmed	11.46 nsec	untrimmed	11.46 nsec

<b>(356-540) - 250<math>\mu</math>M</b>		<b>(356-540) - 1mM</b>	
trim factor	trimmed	trim factor	trimmed
5%	6.58 nsec	5%	15.50 nsec
10%	6.68 nsec	10%	15.48 nsec
untrimmed	6.38 nsec	untrimmed	15.66 nsec

Dr. Wu's program allows variable trimming of any data that lies outside of the normal distribution of observed  $T_1/T_2$  ratios, which removes data originating from signals that might be excessively affected by other influences, such as intermediate exchange-mediated broadening or a high degree of localized motion that can result in artificially narrow linewidths. The data shown in table 6-1 indicates that both constructs show significant differences in their correlation times at the two concentrations, indicative of a monomer – homodimer transition as the concentrations increase. Given the similar changes in the rotational correlation times for both constructs, it appears that the longer construct does not significantly increase the affinity of the homodimeric interaction,

suggesting that regions C-terminal to residue 470 are unlikely to change the association behavior of ARNT PAS-B.

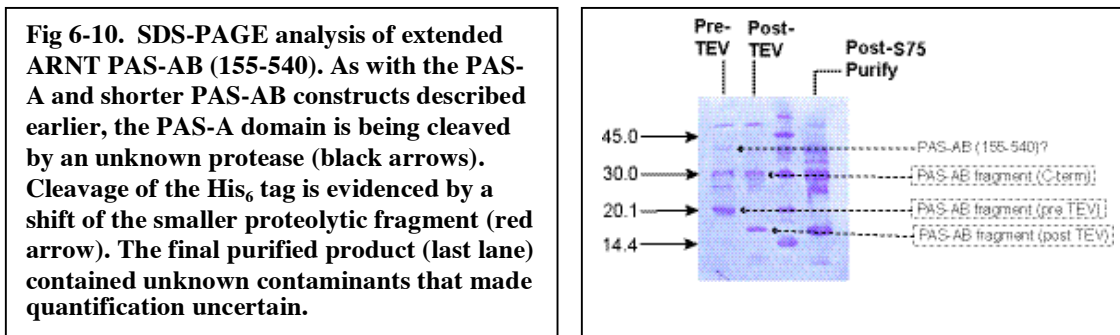
*ii. Extended ARNT PAS-AB (155-540)*

Results from the extended ARNT PAS-B construct (356-540) show that there are no significant changes in the structure relative to the shorter domain (356-470), and suggest that the extra 70 residues are unstructured and do not participate in the heterodimeric and homodimeric interactions (Figs 6-7, 6-8, 6-9). However, these data do not address the possibility of reciprocal PAS-A/PAS-B interactions as seen in the tandem PAS domain homodimer of the Per protein (5). To address this possibility, C-terminal extensions of the previously described ARNT PAS-AB construct (155-540) were also cloned, expressed and purified by Matthew Evans for characterization using similar techniques used for the C-terminal extended ARNT PAS-B (356-540).

*a. Purification and size exclusion chromatography of ARNT PAS-AB (155-540)*

The expression and purification of the extended ARNT PAS-AB construct (155-540) proceeded as previously described for the shorter construct (155-470). As observed for all previously expressed ARNT constructs containing the PAS-A domain, proteolysis occurred either prior to cell extrusion or soon after (Fig 6-10). After the initial round of purification, the His<sub>6</sub> tag was cleaved by overnight incubation with 1mg/30mg TEV/protein, and the tag subsequently removed from the mixture by another round of affinity chromatography. The protein was then injected over a Superdex S75 size exclusion column for further purification and assessment of its dimeric state. The results

of the cleavage of the His<sub>6</sub> tag and the final, post S-75 purified protein destined for NMR experimentation is shown by SDS-PAGE analysis in Fig 6-10.

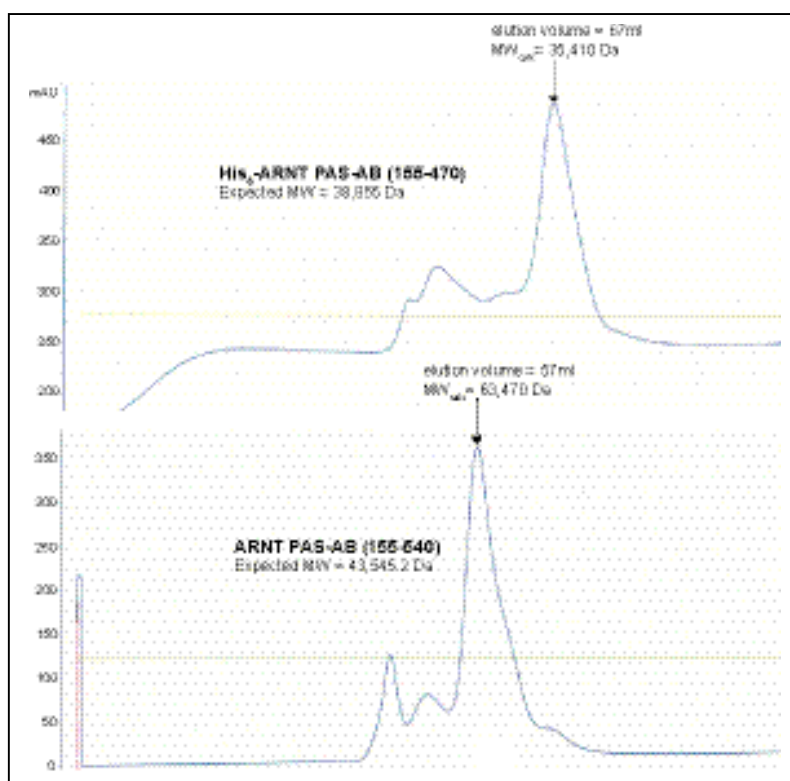


Analysis of all PAS-A domain constructs characterized so far seems to indicate that a single, non-TEV cleavage event is occurring, and that the resulting fragments are stable for days. The bottom band in lane 1 of Fig 6-10 (black arrow) corresponds well to the size of the bottom PAS-AB fragment in Fig 6-5. In the second lane of Fig 6-10 (post TEV), the bottom band is slightly smaller than the bottom band of the post-TEV PAS-A construct in Fig 6-2 b, which is expected because the construct used in that experiment contained an extra 11 residues N-terminal to the PAS-A domain (residues 144-350). All of these data confirm that somewhere in the PAS-A domain is a site particularly accessible to a naturally occurring *E.coli* protease, and that the resulting fragments are relatively stable. Although attempts to map this site by mass spectrometry have been unsuccessful so far, the relative sizes of the fragments indicate that a likely site for this cleavage event is somewhere in the G-H or H-I loops of ARNT PAS-A.

After cleavage and subsequent removal of the His<sub>6</sub> tag, samples were injected over a Superdex S75 size exclusion column to assess their monomeric/dimeric state and

as an additional purification step. The results for the extended ARNT PAS-AB construct (155-540) are shown in Fig 6-11 b, and are directly compared with a similar experiment on the shorter ARNT PAS-AB protein discussed previously (155-470).

**Fig 6-11. Superdex S200 size exclusion chromatography of ARNT PAS-AB constructs.**  
**a. Elution profile of His<sub>6</sub>-ARNT PAS-AB (155-470).** The major peak corresponds to an estimated MW of 35,410 Da, which is consistent with the expected MW of 38,854 Da, within experimental error. **b. Elution profile of ARNT PAS-AB (155-540).** The major peak corresponds to an estimated MW of 63,473 Da, somewhat greater than the expected 43,545 Da predicted for this construct.



The extended construct (155-540) elutes from the size exclusion column at a position that indicates a larger molecular weight than expected (see figure legend). Although this could reveal a shift in the equilibrium of the monomer-dimer transition relative to the shorter PAS-AB construct (155-470), the molecular weight indicated by

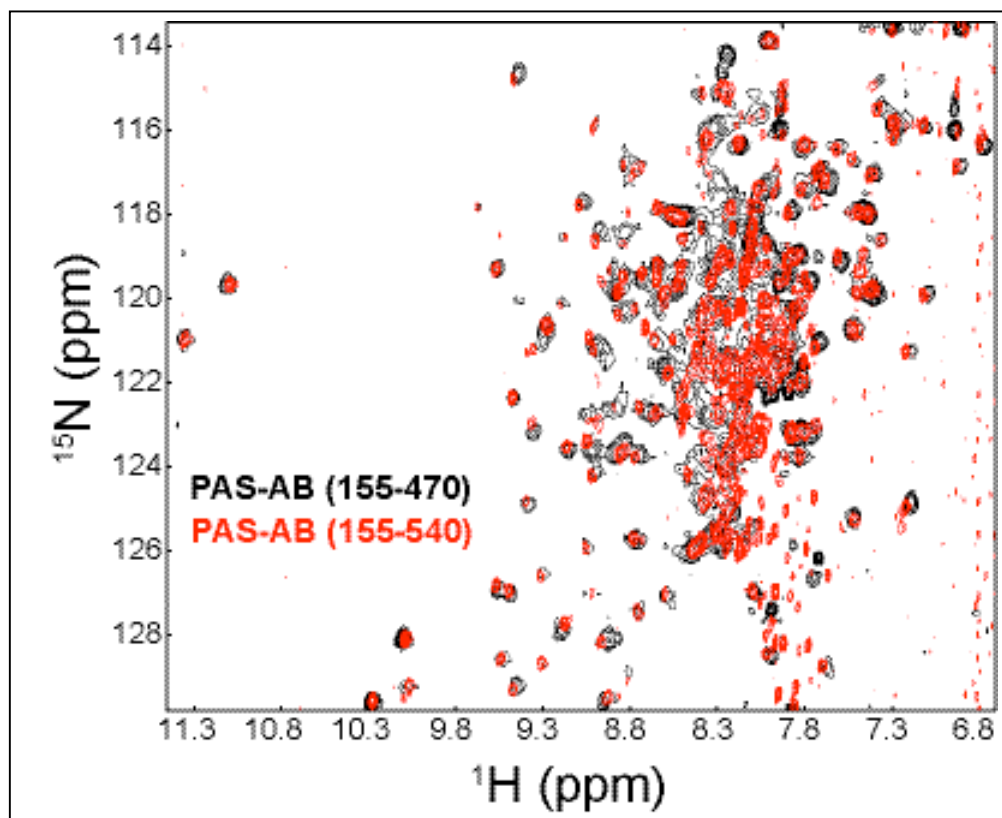
the elution profile is still well below that expected for a stable homodimer (~87,000 Da). In addition, previous results showed that the C-terminal extensions of the PAS-B domains alone appeared unstructured (Fig 6-7), which could affect the behavior of the protein within the gel matrix.

SDS-PAGE analysis of the sample after elution from the column also indicated that significant amounts of unidentified bands were present (Fig 6-10), which means that other proteins could be co-eluting with the extended ARNT PAS-AB protein. This might occur through aggregation mediated by the unstructured C-terminal residues, and could cause a shift in the elution profile of the protein. It should also be noted that the cleavage products from the proteolysis also appear to be co-eluting on the size exclusion column in a similar manner to that reported earlier for the PAS-A domain alone. This is further evidence that the cleaved fragments from ARNT PAS-A are associating, and perhaps folding together into a functional domain.

*b. NMR of ARNT PAS-AB (155-540)*

<sup>15</sup>N-labeled samples of ARNT PAS-AB (155-540) were prepared, and <sup>15</sup>N/<sup>1</sup>H HSQC spectra were obtained to see if any chemical shift changes were detected in either the assigned PAS-B resonances or originating from the unassigned resonances arising from the PAS-A domain. The results are shown in Fig 6-12, and reveal good overlap between the signals in the folded region of the spectra originating from the PAS-A and PAS-B domains of all constructs characterized so far.

**Fig 6-12. Overlay of the  $^{15}\text{N}/^1\text{H}$  HSQC spectra of His<sub>6</sub>-ARNT PAS-AB (155-470) and ARNT PAS-AB (155-540). Although a change in the reference frequency in the  $^{15}\text{N}$  dimension was required due to reasons outlined in the text, no significant changes in the observed chemical shifts are present. This indicates that no significant conformational changes and/or previously unobserved homodimeric interactions are occurring.**



Although there are some minor changes in the chemical shifts of a few peaks, it should be noted that these spectra were obtained under quite different sets of conditions. Quality spectra for the extended PAS-AB (155-540) construct has so far only been obtained with a TROSY-based pulse sequence, which selects a particular multiplet component that exhibits slower transverse relaxation due to interference between the dipole-dipole (DD) relaxation and chemical shift anisotropy (CSA) interactions at high magnetic field strengths (800 MHz in our case) (117). Although this results in a much



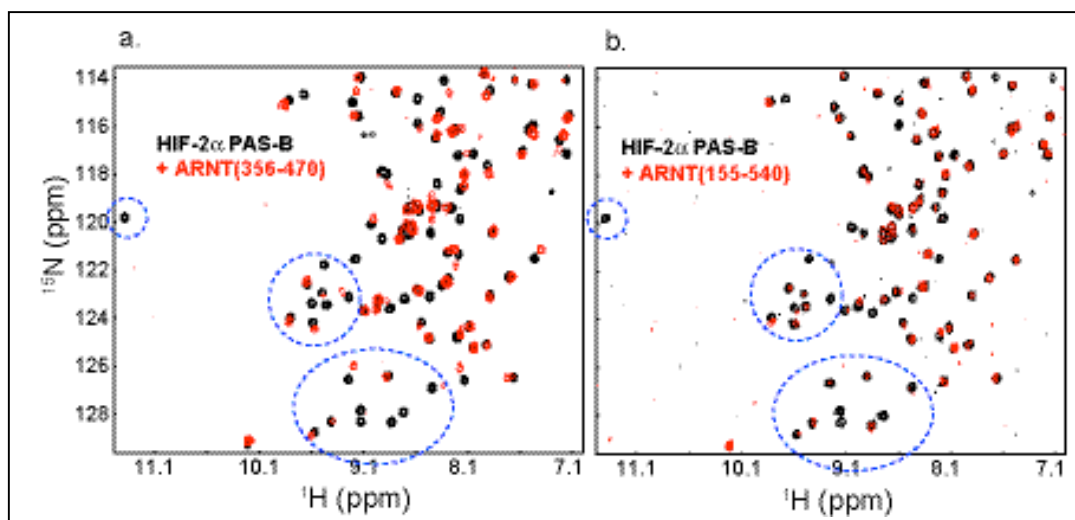
narrower linewidth which can help both with the signal to noise ratio and the resolution of peaks, the positions of peaks in the spectrum no longer reflects an average of the chemical shifts from the multiplet components. The spectra of the original ARNT PAS-AB construct (155-470), on the other hand, was not obtained with a TROSY-based pulse sequence, so a direct overlay of the two required a shift of one spectra relative to the other. In addition, the spectra of the two constructs were obtained at different temperatures (25°C for (155-47) and 35°C for 155-540)), which can be expected to result in chemical shift changes.

Nevertheless, a reasonable overlay of spectra from the two variable-length constructs was obtained by applying a 0.65ppm (52Hz) shift in the reference frequency of the indirect dimension of the spectra originating from the longer construct (155-540). The results shown in Fig 6-10 indicate that there are no major changes in the conformations of the PAS-A or PAS-B domains of the extended ARNT PAS-AB construct (155-540) relative to the individual PAS-A (144-350) or PAS-B (356-470) domains, or as compared to the shorter tandem PAS-AB protein (155-470). This implies that an analogous C-terminal  $\alpha$ -helical region like the one found to be important for Per homodimerization is unlikely to be involved in homodimeric interactions between the ARNT PAS-A and PAS-B domains.

An additional experiment was conducted to see if the extended ARNT PAS-AB (155-540) construct was interacting with the HIF-2 $\alpha$  PAS-B domain in a similar manner to the ARNT PAS-B domain alone (356-470). The results are shown in Fig 6-13, and although peak broadening is increased for the interaction with the longer ARNT

construct, similar residues are being perturbed, indicating that a similar type of interaction is taking place.

**Fig 6-13. Comparison of complexes formed between HIF-2 $\alpha$  PAS-B and either ARNT PAS-B (356-470) or extended ARNT PAS-AB (155-540)** a.  $^{15}\text{N}/^1\text{H}$  HSQC spectra of 250 $\mu\text{M}$  HIF-2 $\alpha$  PAS-B before (black) and after (red) addition of 500 $\mu\text{M}$  ARNT PAS-B (356-470). b.  $^{15}\text{N}/^1\text{H}$  HSQC spectra of 200 $\mu\text{M}$  HIF-2 $\alpha$  PAS-B before (black) and after (red) addition of ~200mM ARNT PAS-AB (155-540). Similar patterns of peak broadening are observed for both ARNT constructs, indicating that similar interactions are occurring.



The enhancement of peak broadening seen for the larger complex could have several explanations. It could be that the greater degree of broadening is simply due to the greater size of the heterodimer, resulting in longer rotational correlation times and subsequent  $R_2$  enhancement. Although it was difficult to obtain an accurate concentration of the extended ARNT PAS-AB construct due to the relative impurity of the sample (Fig 6-10), it is reasonable to assume that the solubility issues we have encountered with the shorter PAS-AB constructs (155-470) would also be an issue with

the extended constructs. As such, the highest concentration of the binding partner of the 200 $\mu$ M HIF-2 $\alpha$  PAS-B sample probably did not exceed 150-200 $\mu$ M.

It was somewhat surprising that the broadening effect was more pronounced for the interaction of HIF-2 $\alpha$  PAS-B with the extended ARNT PAS-AB construct. It is possible that the interaction with HIF-2 $\alpha$  PAS-B is tighter for the extended construct than for the ARNT PAS-B domain alone. This might be supported by the apparent lack of resonances that exhibit chemical shift changes in Fig 6-13 b. If the interaction has moved more into the intermediate and/or slow exchange regime for many residues, then broadening of previously shifting peaks would be observed instead of changes in the chemical shifts. Involvement of the regions C-terminal to the PAS-B domain in the formation of homodimer might also be supported by the longer retention time of the shorter ARNT PAS-AB construct (155-470) on the size exclusion column (Fig 6-11), which could indicate that it is less likely to exist as a homodimer relative to the tandem PAS-AB with the C-terminal extensions (155-540).

### *iii. Conclusions*

Although there appears to be some evidence that extension of the C-terminus of tandem ARNT PAS-AB constructs might increase the affinities of both homodimeric and heterodimeric interactions, a model for a PAS-AB interaction such as that seen for the Per protein does not seem to be supported by most of the experimental data obtained so far. The homodimeric interaction observed in the Per crystal structure is quite different with respect to the solvent-exposed  $\beta$ -sheet interface observed for many PAS domains. It is therefore unlikely that the apparent increase in dimerization affinities observed for the

extended PAS-AB tandem ARNT domains is due to a similar type of reciprocal PAS-A/PAS-B interaction, because the patterns of chemical shifts observed for the various constructs correlate very well, indicating that similar interactions are occurring regardless of whether the PAS domains are expressed alone or in tandem. It is possible that an additional 70 residues existing in a random coil configuration might be participating in non-specific interactions with the hetero- or homodimeric binding partner, which could result in an apparent higher affinity on a size exclusion column or in greater peak broadening due to longer rotational correlation times. Such non-specific interactions between the unstructured regions could also serve to increase the effective concentrations of the attached PAS domains, thereby increasing the apparent affinities of the same  $\beta$ -sheet mediated interactions that were the basis of the HADDOCK-derived HIF-2 $\alpha$ /ARNT PAS-B heterodimer model. In addition, the presence of the ARNT PAS-A domain could be acting in concert with the HIF-2 $\alpha$ /ARNT PAS-B interactions in a cooperative manner, as described in our PAS tetramer model presented in Chapter III, section *B.iii* (Fig 3-18).

Regardless of the actual mode of PAS domain associations in the formation of bHLH-PAS complexes in the hypoxia response pathway, more work will have to be done in order to stabilize ARNT constructs containing the PAS-A domain. In addition, extensive characterization of functional complexes that include the bHLH DNA-binding regions will probably have to be performed to fully understand how PAS domains are involved in this and other systems to contribute to affinity and/or specificity in the formation of viable transcriptional activation complexes.

## References

- (1) Dunlap, J. C., and Loros, J. J. (2005) Analysis of circadian rhythms in neurospora: overview of assays and genetic and molecular biological manipulation. *Methods Enzymol* 393, 3-22.
- (2) 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 (Camb)* 10, 1349-61.
- (3) 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.
- (4) Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* 98, 10037-41.
- (5) 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.
- (6) Miyatake, H., Mukai, M., Park, S. Y., Adachi, S., Tamura, K., Nakamura, H., Nakamura, K., Tsuchiya, T., Iizuka, T., and Shiro, Y. (2000) Sensory mechanism of oxygen sensor FixL from Rhizobium meliloti: crystallographic, mutagenesis and resonance Raman spectroscopic studies. *J Mol Biol* 301, 415-31.
- (7) Ferre-D'Amare, A. R., Prendergast, G. C., Ziff, E. B., and Burley, S. K. (1993) Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature* 363, 38-45.
- (8) 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, 15504-9.
- (9) Huang, Z. J., Edery, I., and Rosbash, M. (1993) PAS is a dimerization domain common to Drosophila period and several transcription factors. *Nature* 364, 259-62.
- (10) 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, 42-7.
- (11) 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.
- (12) 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, 954-8.
- (13) 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, 185-8.

- (14) 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, 1157-67.
- (15) 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.
- (16) Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25, 3389-402.
- (17) Rost, B., and Sander, C. (1993) Prediction of protein secondary structure at better than 70% accuracy. *J Mol Biol* 232, 584-99.
- (18) 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, 331-3.
- (19) Pawson, T., and Nash, P. (2000) Protein-protein interactions define specificity in signal transduction. *Genes Dev* 14, 1027-47.
- (20) Lim, W. A. (2002) The modular logic of signaling proteins: building allosteric switches from simple binding domains. *Curr Opin Struct Biol* 12, 61-8.
- (21) Taylor, B. L., Zhulin, I. B., and Johnson, M. S. (1999) Aerotaxis and other energy-sensing behavior in bacteria. *Annu Rev Microbiol* 53, 103-28.
- (22) 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, 3730-7.
- (23) 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, 5830-5.
- (24) 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, 13318-23.
- (25) 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, 15177-82.
- (26) 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, 206-9.
- (27) 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, 12689-99.
- (28) 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, 14671-8.
- (29) Hoff, W. D., Dux, P., Hard, K., Devreese, B., Nugteren-Roodzant, I. M., Crielaard, 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, 13959-62.

- (30) 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, 14369-77.
- (31) Sakamoto, K., and Briggs, W. R. (2002) Cellular and subcellular localization of phototropin 1. *Plant Cell* 14, 1723-35.
- (32) Briggs, W. R., and Christie, J. M. (2002) Phototropins 1 and 2: versatile plant blue-light receptors. *Trends Plant Sci* 7, 204-10.
- (33) 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, 9401-10.
- (34) 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, 36493-500.
- (35) 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, 205-19.
- (36) Harper, S. M., Neil, L. C., and Gardner, K. H. (2003) Structural basis of a phototropin light switch. *Science* 301, 1541-4.
- (37) Hanks, S. K., and Hunter, T. (1995) Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *Faseb J* 9, 576-96.
- (38) Tanaka, T., Saha, S. K., Tomomori, C., Ishima, R., Liu, D., Tong, K. I., Park, H., Dutta, R., Qin, L., Swindells, M. B., Yamazaki, T., Ono, A. M., Kainosho, M., Inouye, M., and Ikura, M. (1998) NMR structure of the histidine kinase domain of the E. coli osmosensor EnvZ. *Nature* 396, 88-92.
- (39) Robinson, V. L., Buckler, D. R., and Stock, A. M. (2000) A tale of two components: a novel kinase and a regulatory switch. *Nat Struct Biol* 7, 626-33.
- (40) Rutter, J., Michnoff, C. H., Harper, S. M., Gardner, K. H., and McKnight, S. L. (2001) PAS kinase: an evolutionarily conserved PAS domain-regulated serine/threonine kinase. *Proc Natl Acad Sci U S A* 98, 8991-6.
- (41) 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, 705-8.
- (42) 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, 449-55.
- (43) Zelzer, E., Wappner, P., and Shilo, B. Z. (1997) The PAS domain confers target gene specificity of *Drosophila* bHLH/PAS proteins. *Genes Dev* 11, 2079-89.
- (44) Kurokawa, H., Lee, D. S., Watanabe, M., Sagami, I., Mikami, B., Raman, C. S., and Shimizu, T. (2004) A redox-controlled molecular switch revealed by the crystal structure of a bacterial heme PAS sensor. *J Biol Chem* 279, 20186-93.
- (45) Park, H., Suquet, C., Satterlee, J. D., and Kang, C. (2004) Insights into signal transduction involving PAS domain oxygen-sensing heme proteins from the X-ray crystal structure of *Escherichia coli* Dos heme domain (Ec DosH). *Biochemistry* 43, 2738-46.

- (46) 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, 649-55.
- (47) Ballario, P., Vittorioso, P., Magrelli, A., Talora, C., Cabibbo, A., and Macino, G. (1996) White collar-1, a central regulator of blue light responses in *Neurospora*, is a zinc finger protein. *Embo J* 15, 1650-7.
- (48) Linden, H., and Macino, G. (1997) White collar 2, a partner in blue-light signal transduction, controlling expression of light-regulated genes in *Neurospora crassa*. *Embo J* 16, 98-109.
- (49) Pongratz, I., Antonsson, C., Whitelaw, M. L., and Poellinger, L. (1998) Role of the PAS domain in regulation of dimerization and DNA binding specificity of the dioxin receptor. *Mol Cell Biol* 18, 4079-88.
- (50) Crews, S. T. (1998) Control of cell lineage-specific development and transcription by bHLH-PAS proteins. *Genes Dev* 12, 607-20.
- (51) Crews, S. T., and Fan, C. M. (1999) Remembrance of things PAS: regulation of development by bHLH-PAS proteins. *Curr Opin Genet Dev* 9, 580-7.
- (52) 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, 517-54.
- (53) Safe, S. H. (1986) Comparative toxicology and mechanism of action of polychlorinated dibenzo-p-dioxins and dibenzofurans. *Annu Rev Pharmacol Toxicol* 26, 371-99.
- (54) 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, 428-38.
- (55) 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, 645-54.
- (56) Baba, T., Mimura, J., Gradin, K., Kuroiwa, A., Watanabe, T., Matsuda, Y., Inazawa, J., Sogawa, K., and Fujii-Kuriyama, Y. (2001) Structure and expression of the Ah receptor repressor gene. *J Biol Chem* 276, 33101-10.
- (57) Schrenk, D. (1998) Impact of dioxin-type induction of drug-metabolizing enzymes on the metabolism of endo- and xenobiotics. *Biochem Pharmacol* 55, 1155-62.
- (58) Schmidt, J. V., Su, G. H., Reddy, J. K., Simon, M. C., and Bradfield, C. A. (1996) Characterization of a murine Ahr null allele: involvement of the Ah receptor in hepatic growth and development. *Proc Natl Acad Sci U S A* 93, 6731-6.
- (59) O'Rourke, J. F., Tian, Y. M., Ratcliffe, P. J., and Pugh, C. W. (1999) Oxygen-regulated and transactivating domains in endothelial PAS protein 1: comparison with hypoxia-inducible factor-1 $\alpha$ . *J Biol Chem* 274, 2060-71.
- (60) 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) HIF $\alpha$  targeted for VHL-mediated destruction by proline hydroxylation: implications for O<sub>2</sub> sensing. *Science* 292, 464-8.



- (61) 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 O<sub>2</sub>-regulated prolyl hydroxylation. *Science* 292, 468-72.
- (62) Yu, F., White, S. B., Zhao, Q., and Lee, F. S. (2001) HIF-1 $\alpha$  binding to VHL is regulated by stimulus-sensitive proline hydroxylation. *Proc Natl Acad Sci U S A* 98, 9630-5.
- (63) Masson, N., Willam, C., Maxwell, P. H., Pugh, C. W., and Ratcliffe, P. J. (2001) Independent function of two destruction domains in hypoxia-inducible factor- $\alpha$  chains activated by prolyl hydroxylation. *Embo J* 20, 5197-206.
- (64) 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, 271-5.
- (65) 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, 4298-309.
- (66) 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, 1822-33.
- (67) 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, 858-61.
- (68) 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, 2984-92.
- (69) 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, 1376-93.
- (70) Bruick, R. K. (2000) Expression of the gene encoding the proapoptotic Nip3 protein is induced by hypoxia. *Proc Natl Acad Sci U S A* 97, 9082-7.
- (71) Yang, J., Zhang, L., Erbel, P. J. A., Gardner, K. H., Ding, K. M., Garcia, J. A., and Bruick, R. K. (2005) Functions of the Per/ARNT/Sim (PAS) domains of the hypoxia inducible factor (HIF). *Submitted*.
- (72) 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, 8343-55.
- (73) 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 (Tokyo)* 122, 703-10.
- (74) 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, 403-7.

- (75) Kozak, K. R., Abbott, B., and Hankinson, O. (1997) ARNT-deficient mice and placental differentiation. *Dev Biol* 191, 297-305.
- (76) Carmeliet, P., Ferreira, V., Breier, G., Pollefeyt, S., Kieckens, L., Gertsenstein, M., Fahrig, M., Vandenhoek, A., Harpal, K., Eberhardt, C., Declercq, C., Pawling, J., Moons, L., Collen, D., Risau, W., and Nagy, A. (1996) Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele. *Nature* 380, 435-9.
- (77) Ferrara, N., Carver-Moore, K., Chen, H., Dowd, M., Lu, L., O'Shea, K. S., Powell-Braxton, L., Hillan, K. J., and Moore, M. W. (1996) Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene. *Nature* 380, 439-42.
- (78) Maxwell, P. H., Dachs, G. U., Gleadle, J. M., Nicholls, L. G., Harris, A. L., Stratford, I. J., Hankinson, O., Pugh, C. W., and Ratcliffe, P. J. (1997) Hypoxia-inducible factor-1 modulates gene expression in solid tumors and influences both angiogenesis and tumor growth. *Proc Natl Acad Sci U S A* 94, 8104-9.
- (79) Hirose, K., Morita, M., Ema, M., Mimura, J., Hamada, H., Fujii, H., Saijo, Y., Gotoh, O., Sogawa, K., and Fujii-Kuriyama, Y. (1996) cDNA cloning and tissue-specific expression of a novel basic helix-loop-helix/PAS factor (Arnt2) with close sequence similarity to the aryl hydrocarbon receptor nuclear translocator (Arnt). *Mol Cell Biol* 16, 1706-13.
- (80) Huffman, J. L., Mokashi, A., Bachinger, H. P., and Brennan, R. G. (2001) The basic helix-loop-helix domain of the aryl hydrocarbon receptor nuclear transporter (ARNT) can oligomerize and bind E-box DNA specifically. *J Biol Chem* 276, 40537-44.
- (81) Swanson, H. I., Chan, W. K., and Bradfield, C. A. (1995) DNA binding specificities and pairing rules of the Ah receptor, ARNT, and SIM proteins. *J Biol Chem* 270, 26292-302.
- (82) Antonsson, C., Arulampalam, V., Whitelaw, M. L., Pettersson, S., and Poellinger, L. (1995) Constitutive function of the basic helix-loop-helix/PAS factor Arnt. Regulation of target promoters via the E box motif. *J Biol Chem* 270, 13968-72.
- (83) 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, 1936-40.
- (84) Card, P. B., and Gardner, K. H. (2005) Identification and optimization of protein domains for NMR studies. *Methods Enzymol* 394, 3-16.
- (85) Liu, Y. (2003) Molecular mechanisms of entrainment in the *Neurospora* circadian clock. *J Biol Rhythms* 18, 195-205.
- (86) Dunlap, J. C. (1999) Molecular bases for circadian clocks. *Cell* 96, 271-90.
- (87) Young, M. W., and Kay, S. A. (2001) Time zones: a comparative genetics of circadian clocks. *Nat Rev Genet* 2, 702-15.
- (88) Loros, J. J., and Dunlap, J. C. (2001) Genetic and molecular analysis of circadian rhythms in *Neurospora*. *Annu Rev Physiol* 63, 757-94.
- (89) Crosthwaite, S. K., Loros, J. J., and Dunlap, J. C. (1995) Light-induced resetting of a circadian clock is mediated by a rapid increase in frequency transcript. *Cell* 81, 1003-12.

- (90) He, Q., Cheng, P., Yang, Y., Wang, L., Gardner, K. H., and Liu, Y. (2002) White collar-1, a DNA binding transcription factor and a light sensor. *Science* 297, 840-3.
- (91) Froehlich, A. C., Liu, Y., Loros, J. J., and Dunlap, J. C. (2002) White Collar-1, a circadian blue light photoreceptor, binding to the frequency promoter. *Science* 297, 815-9.
- (92) Pittendrigh, C. S., Bruce, V. G., Rosenzweig, N. S., and Rubin, M. L. (1959) A biological clock in *Neurospora*. *Nature* 184, 169-170.
- (93) Feldman, J. F., and Waser, N. (1971) *New mutations affecting circadian rhythmicity in Neurospora*, Natl. Acad. Sci, Washington, D.C.
- (94) Aronson, B. D., Johnson, K. A., and Dunlap, J. C. (1994) Circadian clock locus frequency: protein encoded by a single open reading frame defines period length and temperature compensation. *Proc Natl Acad Sci U S A* 91, 7683-7.
- (95) McClung, C. R., Fox, B. A., and Dunlap, J. C. (1989) The *Neurospora* clock gene frequency shares a sequence element with the *Drosophila* clock gene period. *Nature* 339, 558-62.
- (96) Loros, J. J., Denome, S. A., and Dunlap, J. C. (1989) Molecular cloning of genes under control of the circadian clock in *Neurospora*. *Science* 243, 385-8.
- (97) Talora, C., Franchi, L., Linden, H., Ballario, P., and Macino, G. (1999) Role of a white collar-1-white collar-2 complex in blue-light signal transduction. *Embo J* 18, 4961-8.
- (98) Cheng, P., Yang, Y., and Liu, Y. (2001) Interlocked feedback loops contribute to the robustness of the *Neurospora* circadian clock. *Proc Natl Acad Sci U S A* 98, 7408-13.
- (99) Crosthwaite, S. K., Dunlap, J. C., and Loros, J. J. (1997) *Neurospora* wc-1 and wc-2: transcription, photoresponses, and the origins of circadian rhythmicity. *Science* 276, 763-9.
- (100) Collett, M. A., Garceau, N., Dunlap, J. C., and Loros, J. J. (2002) Light and clock expression of the *Neurospora* clock gene frequency is differentially driven by but dependent on WHITE COLLAR-2. *Genetics* 160, 149-58.
- (101) Cheng, P., Yang, Y., Gardner, K. H., and Liu, Y. (2002) PAS domain-mediated WC-1/WC-2 interaction is essential for maintaining the steady-state level of WC-1 and the function of both proteins in circadian clock and light responses of *Neurospora*. *Mol Cell Biol* 22, 517-24.
- (102) Garceau, N. Y., Liu, Y., Loros, J. J., and Dunlap, J. C. (1997) Alternative initiation of translation and time-specific phosphorylation yield multiple forms of the essential clock protein FREQUENCY. *Cell* 89, 469-76.
- (103) Liu, Y., Garceau, N. Y., Loros, J. J., and Dunlap, J. C. (1997) Thermally regulated translational control of FRQ mediates aspects of temperature responses in the *neurospora* circadian clock. *Cell* 89, 477-86.
- (104) Cheng, P., Yang, Y., Heintzen, C., and Liu, Y. (2001) Coiled-coil domain-mediated FRQ-FRQ interaction is essential for its circadian clock function in *Neurospora*. *Embo J* 20, 101-8.
- (105) Denault, D. L., Loros, J. J., and Dunlap, J. C. (2001) WC-2 mediates WC-1-FRQ interaction within the PAS protein-linked circadian feedback loop of *Neurospora*. *Embo J* 20, 109-17.

- (106) Mellow, M., Franchi, L., Dragovic, Z., Gorl, M., Johnson, J., Brunner, M., Macino, G., and Roenneberg, T. (2001) Circadian regulation of the light input pathway in *Neurospora crassa*. *Embo J* 20, 307-15.
- (107) Yang, Y., Cheng, P., and Liu, Y. (2002) Regulation of the *Neurospora* circadian clock by casein kinase II. *Genes Dev* 16, 994-1006.
- (108) Cheng, P., Yang, Y., Wang, L., He, Q., and Liu, Y. (2003) WHITE COLLAR-1, a multifunctional *neurospora* protein involved in the circadian feedback loops, light sensing, and transcription repression of *wc-2*. *J Biol Chem* 278, 3801-8.
- (109) Liu, Y., Loros, J., and Dunlap, J. C. (2000) Phosphorylation of the *Neurospora* clock protein FREQUENCY determines its degradation rate and strongly influences the period length of the circadian clock. *Proc Natl Acad Sci U S A* 97, 234-9.
- (110) Linden, H., Ballario, P., and Macino, G. (1997) Blue light regulation in *Neurospora crassa*. *Fungal Genet Biol* 22, 141-50.
- (111) Lakin-Thomas, P. L., Cote, G. G., and Brody, S. (1990) Circadian rhythms in *Neurospora crassa*: biochemistry and genetics. *Crit Rev Microbiol* 17, 365-416.
- (112) Christie, J. M., Salomon, M., Nozue, K., Wada, M., and Briggs, W. R. (1999) LOV (light, oxygen, or voltage) domains of the blue-light photoreceptor phototropin (*nph1*): binding sites for the chromophore flavin mononucleotide. *Proc Natl Acad Sci U S A* 96, 8779-83.
- (113) Briggs, W. R., and Huala, E. (1999) Blue-light photoreceptors in higher plants. *Annu Rev Cell Dev Biol* 15, 33-62.
- (114) Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., et al. (2001) The sequence of the human genome. *Science* 291, 1304-51.
- (115) Zarembinski, T. I., Hung, L. W., Mueller-Dieckmann, H. J., Kim, K. K., Yokota, H., Kim, R., and Kim, S. H. (1998) Structure-based assignment of the biochemical function of a hypothetical protein: a test case of structural genomics. *Proc Natl Acad Sci U S A* 95, 15189-93.
- (116) Gardner, K. H., and Kay, L. E. (1998) The use of <sup>2</sup>H, <sup>13</sup>C, <sup>15</sup>N multidimensional NMR to study the structure and dynamics of proteins. *Annu Rev Biophys Biomol Struct* 27, 357-406.

- (117) Pervushin, K. (2000) Impact of transverse relaxation optimized spectroscopy (TROSY) on NMR as a technique in structural biology. *Q Rev Biophys* 33, 161-97.
- (118) 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, 1198-208.
- (119) 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, 1501-31.
- (120) Sonnhammer, E. L., Eddy, S. R., Birney, E., Bateman, A., and Durbin, R. (1998) Pfam: multiple sequence alignments and HMM-profiles of protein domains. *Nucleic Acids Res* 26, 320-2.
- (121) Schultz, J., Milpetz, F., Bork, P., and Ponting, C. P. (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* 95, 5857-64.
- (122) Frishman, D., and Argos, P. (1996) Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence. *Protein Eng* 9, 133-42.
- (123) Livingstone, C. D., and Barton, G. J. (1996) Identification of functional residues and secondary structure from protein multiple sequence alignment. *Methods Enzymol* 266, 497-512.
- (124) Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M., and Barton, G. J. (1998) JPred: a consensus secondary structure prediction server. *Bioinformatics* 14, 892-3.
- (125) Cohen, S. L., Ferre-D'Amare, A. R., Burley, S. K., and Chait, B. T. (1995) Probing the solution structure of the DNA-binding protein Max by a combination of proteolysis and mass spectrometry. *Protein Sci* 4, 1088-99.
- (126) Hammarstrom, M., Hellgren, N., van Den Berg, S., Berglund, H., and Hard, T. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*. *Protein Sci* 11, 313-21.
- (127) Huth, J. R., Bewley, C. A., Jackson, B. M., Hinnebusch, A. G., Clore, G. M., and Gronenborn, A. M. (1997) Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR. *Protein Sci* 6, 2359-64.
- (128) Zhou, P., Lugovskoy, A. A., and Wagner, G. (2001) A solubility-enhancement tag (SET) for NMR studies of poorly behaving proteins. *J Biomol NMR* 20, 11-4.
- (129) Gronenborn, A. M., and Clore, G. M. (1996) Rapid screening for structural integrity of expressed proteins by heteronuclear NMR spectroscopy. *Protein Sci* 5, 174-7.
- (130) Woestenenk, E. A., Hammarstrom, M., Hard, T., and Berglund, H. (2003) Screening methods to determine biophysical properties of proteins in structural genomics. *Anal Biochem* 318, 71-9.
- (131) Serber, Z., Straub, W., Corsini, L., Nomura, A. M., Shimba, N., Craik, C. S., Ortiz de Montellano, P., and Dotsch, V. (2004) Methyl groups as probes for proteins and complexes in in-cell NMR experiments. *J Am Chem Soc* 126, 7119-25.

- (132) Erbel, P. J., Seidel, R., Macintosh, S. E., Gentile, L. N., Amor, J. C., Kahn, R. A., Prestegard, J. H., McIntosh, L. P., and Gardner, K. H. (2004) Cyclic enterobacterial common antigen: potential contaminant of bacterially expressed protein preparations. *J Biomol NMR* 29, 199-204.
- (133) Bagby, S., Tong, K. I., Liu, D., Alattia, J. R., and Ikura, M. (1997) The button test: a small scale method using microdialysis cells for assessing protein solubility at concentrations suitable for NMR. *J Biomol NMR* 10, 279-82.
- (134) Lepre, C. A., and Moore, J. M. (1998) Microdrop screening: a rapid method to optimize solvent conditions for NMR spectroscopy of proteins. *J Biomol NMR* 12, 493-9.
- (135) Bagby, S., Tong, K. I., and Ikura, M. (2001) Optimization of protein solubility and stability for protein nuclear magnetic resonance. *Methods Enzymol* 339, 20-41.
- (136) 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. *the Journal of Molecular Biology In Press*.
- (137) 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, 5510-4.
- (138) 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, 5811-22.
- (139) 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, 5865-75.
- (140) 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, 4933-47.
- (141) 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, 6757-62.
- (142) Ema, M., Taya, S., Yokotani, N., Sogawa, K., Matsuda, Y., and Fujii-Kuriyama, Y. (1997) A novel bHLH-PAS factor with close sequence similarity to hypoxia-inducible factor 1alpha regulates the VEGF expression and is potentially involved in lung and vascular development. *Proc Natl Acad Sci U S A* 94, 4273-8.
- (143) Ryan, H. E., Lo, J., and Johnson, R. S. (1998) HIF-1 alpha is required for solid tumor formation and embryonic vascularization. *Embo J* 17, 3005-15.
- (144) Tian, H., McKnight, S. L., and Russell, D. W. (1997) Endothelial PAS domain protein 1 (EPAS1), a transcription factor selectively expressed in endothelial cells. *Genes Dev* 11, 72-82.
- (145) 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. *Molecular and Cellular Biology* 25, 3163-72.
- (146) 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, 5353-62.
  - (147) Jiang, B. H., Rue, E., Wang, G. L., Roe, R., and Semenza, G. L. (1996) Dimerization, DNA binding, and transactivation properties of hypoxia-inducible factor 1. *J Biol Chem* 271, 17771-8.
  - (148) 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, 34-9.
  - (149) 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, 277-93.
  - (150) Johnson, B. A., and Blevins, R. A. (1994) NMRView: a computer program for the visualization and analysis of NMR data. *Journal of Biomolecular NMR* 4, 603-614.
  - (151) Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) Primary structure effects on peptide group hydrogen exchange. *Proteins* 17, 75-86.
  - (152) 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.
  - (153) 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 (Pt 5), 905-21.
  - (154) Nilges, M., and O'Donoghue, S. I. (1998) Ambiguous NOEs and automated NOE assignment. *Progress in Nuclear Magnetic Resonance Spectroscopy* 32, 107-139.
  - (155) Koradi, R., Billeter, M., and Wuthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* 14, 51-5, 29-32.
  - (156) Laskowski, R. A., Rullmann, 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, 477-86.
  - (157) Dror, O., Benyamini, H., Nussinov, R., and Wolfson, H. (2003) MASS: multiple structural alignment by secondary structures. *Bioinformatics* 19 Suppl 1, i95-104.
  - (158) Pellequer, J. L., Brudler, R., and Getzoff, E. D. (1999) Biological sensors: more than one way to sense oxygen. *Current Biology* 9, R416-8.
  - (159) Holdeman, T. C., and Gardner, K. H. (2001) <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N chemical shift assignments of the N-terminal PAS domain of mNPAS2. *J Biomol NMR* 21, 383-4.
  - (160) Farmer, B. T., 2nd, Constantine, K. L., Goldfarb, V., Friedrichs, M. S., Wittekind, M., Yanchunas, J., Jr., Robertson, J. G., and Mueller, L. (1996) Localizing the NADP<sup>+</sup> binding site on the MurB enzyme by NMR. *Nat Struct Biol* 3, 995-7.

- (161) Cusanovich, M. A., and Meyer, T. E. (2003) Photoactive yellow protein: a prototypic PAS domain sensory protein and development of a common signaling mechanism. *Biochemistry* 42, 4759-70.
- (162) Dominguez, C., Boelens, R., and Bonvin, A. M. (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* 125, 1731-7.
- (163) Dominguez, C., Bonvin, A. M., Winkler, G. S., van Schaik, F. M., Timmers, H. T., and Boelens, R. (2004) Structural model of the UbcH5B/CNOT4 complex revealed by combining NMR, mutagenesis, and docking approaches. *Structure (Camb)* 12, 633-44.
- (164) Hubbard, S. J., and Thornton, J. M. (1993), Department of Biochemistry and Molecular Biology, University College London, U.K.
- (165) Bloembergen, N. (1957) Proton relaxation times in paramagnetic solutions. *Journal of Chemical Physics*, 572-573.
- (166) Solomon, I. (1955) Relaxation processes in a system of two spins. *Physical reviews*, 559-565.
- (167) Battiste, J. L., and Wagner, G. (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry* 39, 5355-65.
- (168) Pellequer, J. L., Wager-Smith, K. A., Kay, S. A., and Getzoff, E. D. (1998) Photoactive yellow protein: a structural prototype for the three-dimensional fold of the PAS domain superfamily. *Proc Natl Acad Sci U S A* 95, 5884-90.
- (169) Gekakis, N., Saez, L., Delahaye-Brown, A. M., Myers, M. P., Sehgal, A., Young, M. W., and Weitz, C. J. (1995) Isolation of timeless by PER protein interaction: defective interaction between timeless protein and long-period mutant PERL. *Science* 270, 811-5.
- (170) Saez, L., and Young, M. W. (1996) Regulation of nuclear entry of the Drosophila clock proteins period and timeless. *Neuron* 17, 911-20.
- (171) Keith, B., Adelman, D. M., and Simon, M. C. (2001) Targeted mutation of the murine arylhydrocarbon receptor nuclear translocator 2 (Arnt2) gene reveals partial redundancy with Arnt. *Proc Natl Acad Sci U S A* 98, 6692-7.
- (172) Fejzo, J., Lepre, C. A., Peng, J. W., Bemis, G. W., Ajay, Murcko, M. A., and Moore, J. M. (1999) The SHAPES strategy: an NMR-based approach for lead generation in drug discovery. *Chem Biol* 6, 755-69.
- (173) Hajduk, P. J., Bures, M., Praestgaard, J., and Fesik, S. W. (2000) Privileged molecules for protein binding identified from NMR-based screening. *J Med Chem* 43, 3443-7.
- (174) Michnick, S. W., Rosen, M. K., Wandless, T. J., Karplus, M., and Schreiber, S. L. (1991) Solution structure of FKBP, a rotamase enzyme and receptor for FK506 and rapamycin. *Science* 252, 836-9.
- (175) Gosser, Y. Q., Nomanbhoy, T. K., Aghazadeh, B., Manor, D., Combs, C., Cerione, R. A., and Rosen, M. K. (1997) C-terminal binding domain of Rho GDP-dissociation inhibitor directs N-terminal inhibitory peptide to GTPases. *Nature* 387, 814-9.



- (176) Ojennus, D. D., Fleissner, M. R., and Wuttke, D. S. (2001) Reconstitution of a native-like SH2 domain from disordered peptide fragments examined by multidimensional heteronuclear NMR. *Protein Sci* 10, 2162-75.
- (177) Ojennus, D. D., Lehto, S. E., and Wuttke, D. S. (2003) Electrostatic interactions in the reconstitution of an SH2 domain from constituent peptide fragments. *Protein Sci* 12, 44-55.
- (178) Lipari, G., and Szabo, A. (1982) Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. *Journal of the American Chemical Society* 104, 4546-4559.
- (179) Lipari, G., and Szabo, A. (1982) Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 2. Analysis of Experimental Results. *Journal of the American Chemical Society* 104, 4559-4570.

## VITAE

Paul Brandon Card was born on October 5<sup>th</sup>, 1968 in Dallas, Texas, the son of Lewis and Gretchen Card. After graduating from Plano Senior High School in 1987, he attended Stephen F. Austin State University for two years before transferring to the University of North Texas in Denton, Texas. While there he studied a wide variety of subjects and took several semesters off for work and travel, including a five month hiatus spent hitchhiking and working throughout Alaska and six months working in Boston. While in Denton he also joined a non-profit student organization, of which he later became president and a member of the Board of Directors. Although he changed majors twice, he received a Bachelor of Science degree with a major in biology in 1996. After graduation he spent two years as a lab assistant working on the Human Genome Project at the University of Texas Southwestern Medical Center at Dallas before beginning his graduate studies at the same institution, where he was awarded the degree of Doctor of Philosophy in Molecular Biophysics in December, 2005. He lives near Denton with his wife of nine years, Melissa, and their daughter Vanessa.

Permanent Address: 3608 Leeds, Ct.  
Corinth, Texas 76210