Regulation of Eukaryotic Transcription by bHLH/PAS Transcription Factors AhR and Arnt

Nan Hao

B.Sc. (Biomedical Science) (Honours)

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Discipline of Biochemistry

School of Molecular and Biomedical Science

The University of Adelaide, Australia

August, 2012

Table of Contents

Abstract iii

Original Publications v

Declaration vi

Acknowledgements vii

Abbreviations ix

Chapter 1: Introduction 1

1.1 History at a glance 1

1.2 The modular nature of bHLH/PAS transcription factors 2

1.2.1 The bHLH domain 2

1.2.2 The PAS domain 3

1.2.3 The transactivation/transrepression domain 7

1.3 Mammalian bHLH/PAS proteins 7

1.3.1 Activation and regulation of AhR 8

1.3.2 The endogenous function of AhR 13

1.3.2.1 Cross talk between AhR and ER 16

1.3.2.2 AhR and immunity 19

1.3.3 Evolutionary diversity of AhR-- What have we learnt? 26

1.3.4 AhR ligands 31

1.3.5 The LBD of AhR 35

1.4 Aims and Approaches 41

Chapter 2: Results 43

2.1 Discovery that the dimerization specificity bHLH/PAS factor Arnt resides in its N-terminal PAS domain (Paper II) 43

2.2 Discovery that Xenobiotic and Suspension Activation of AhR regulates both overlapping and distinct sets of gene expression (Paper III) 44

2.3 Discovery that the mode of interaction between AhR and atypical ligand YH439 is distinct from that of HAH/PAH ligands (Paper I) 46
Chapter 3: Discussion 49
Chapter 4: References 59

Abstract

The Aryl hydrocarbon Receptor (AhR) is a Class I basic Helix-Loop-Helix/Per-Arnt-Sim (bHLH/PAS) transcription factor essential for adaptive response to xenobiotics. As for all bHLH/PAS proteins, it has a uniform molecular design that is composed of three functional domains, including a highly conserved N-terminal bHLH domain, a pair of degenerate PAS repeats (designated PAS.A and PAS.B), and a poorly conserved C-terminal transactivation domain. However, in contrast to the other proteins in the family, AhR is the only member known to bind xenobiotic ligand found in nature.

The ligand binding domain (LBD) of AhR resides in its PAS.B region. Structure activity relationship analysis suggested that the ligand binding pocket of AhR is promiscuous, which can accommodate a large number of planar and hydrophobic compounds. Polycyclic Aromatic Hydrocarbons (PAHs) and Halogenated Aromatic Hydrocarbons (HAHs) are by far the most common classes of AhR ligands. In addition, pharmaceutical compounds such as the hepatoprotective agent YH439 that fall outside the aromatic classification, have also been shown to activate AhR, presumably by functioning as AhR agonists.

To better characterize the LBD of mouse AhR (mAhR), rational site-directed mutagenesis was performed based on the LBD sequences of zebrafish (*Danio rerio*). Unexpectedly, the mAhR H285Y mutant as well as previously identified A375I mutant were found to discriminate between ligands, suggesting that in contrast to the PAH/HAH ligands, the atypical ligand YH439 has a novel mode of interaction that does not require full access of the ligand binding cavity.

All bHLH/PAS proteins function as obligate dimers. In order to form an active, DNA binding complex, the AhR have to dimerize with Class II bHLH/PAS protein Aryl hydrocarbon receptor nuclear translocator (Arnt). This is mediated primarily via the N terminal bHLH and PAS.A domains. Furthermore, the data presented in this thesis suggest that both the α-helical connector and β-strand structure of the PAS.A domains are required for AhR/Arnt heterodimerization, which is distinct from the β-scaffold surfaces proposed for dimerization between the PAS.B domains of HIF-2α (Hypoxia Inducible Factor-2α) and Arnt. Intriguingly, interaction between Arnt and other

Class I bHLH/PAS proteins were found to occur via the same dimerization interface, suggesting that the dimerization selectivity of common partner factor Arnt resides on a small number of key amino acids within a single dimerization interface of Arnt.

In addition to the canonical AhR activation by xenobiotics, switching cells from adherent to suspension culture also activates the AhR, representing a non-xenobiotic, physiological activation of AhR signaling. This is further supported by the observation that AhR is recruited to the xenobiotic response element (XRE) of prototypical AhR target genes *Cyp1a1*, *Cyp1b1* and *Tiparp* following both xenobiotic and suspension culture induced AhR activation. However, genome wide microarray analysis revealed significant differences between the two activation mechanisms in modulating target gene expression, implying the existence of a fine-tuning control to define the target gene specificity of AhR.

Taken together, the work presented in this thesis explores the various mechanisms underlying AhR regulation with a special emphasis on specificity, which not only advances our current understanding on the non-canonical pathways of AhR activation, but also lends novel insights into how eukaryotic genes are regulated at the transcriptional level.

Original Publications

This thesis is based on the following publications, which will be referred to by their roman numerals in the text:

Paper I Whelan F, **Hao N**, Furness SG, Whitelaw ML, Chapman-Smith A. (2010) Amino acid substitutions in the aryl hydrocarbon receptor ligand binding domain reveal YH439 as an atypical AhR activator. *Mol. Pharmacol.* 77(6):1037-1046.

Paper II **Hao N**, Whitelaw ML, Shearwin KE, Dodd IB, Chapman-Smith A. (2011) Identification of residues in the N-terminal PAS domains important for dimerization of Arnt and AhR. *Nucleic Acids Res.* 39(9):3695-3709.

Paper III **Hao N**, Lee KL, Furness SG, Poellinger L, Whitelaw ML. (2012) Xenobiotics and Loss of Cell Adhesion Drives Distinct Transcriptional Outcomes by Aryl Hydrocarbon Receptor Signaling. *Mol. Pharmacol.* Manuscript under peer-review

Declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution to Nan Hao and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that copyright of published works contained within this thesis (as listed in the previous page) resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue, and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Nan Hao

July 2012

Acknowledgements

Time flies, four years of my PhD study has now come to an end. Doing a PhD is by no means an easy task, and I could not have done it without help from many wonderful people.

First and foremost, I want to thank my principle supervisor A/Prof. Murray Whitelaw, for welcoming me to work in your lab, and for always being supportive and encouraging. You are a truly inspirational model to many students like me, both as a passionate scientist and as a human being.

I also want to acknowledge my co-supervisor Dr. Anne Chapman-smith, for her patience, guidance and encouragement, and Dr. Dan Peet, for his support and friendship. It is a great opportunity to learn from you.

I am also grateful to my collaborators Dr. Keith Shearwin, Dr. Ian Dodd at School of Molecular and Biomedical Science (Biochemistry), The University of Adelaide, Dr. Kian Leong Lee, Prof. Lorenz Poellinger at Cancer Science Institute of Singapore, National University of Singapore, and Prof. Jeff Gorman at Protein Discovery Centre, Queensland Institute of Medical Research. Thank you all for your advice and valuable insights on the project.

My Thanks also go to my fellow colleagues in Whitelaw and Peet labs – Adrienne, Alix, Anne R., Bec, Colleen, Dave, Fiona, Freya, Jacqueline, Karolina, Margo, Matt, Natalia, Navdeep, Nick, Rachel, Sam, Sarah L., Sarah W., Scott, Sebastian, Shweetha, Teresa, Tom, Veronica, and Wei Li. Thank you for your thoughtful discussions and advice throughout my PhD.

To all members in Thomas and Jensen labs, for being good neighbors, and to all other people in Biochemistry discipline and school support staffs, for providing such an enjoyable and inspiring environment to learn.

To all my friends, especially Ang Zhou, Jennifer Liu, Peter Feng, Fay Qiao, and Danni Yang for their friendship and support, and for making my life in Adelaide colorful and enjoyable.

To my wife, for her love, support, encouragement, and understanding.

And finally, to my parents for their love and support. I am forever grateful for them in supporting me to pursue my interest.

Abbreviations

3',4'-DMF	3',4'-Dimethoxyflavone
3MC	3-Methylcholanthrene
6-MCDF	6-Methyl-1,3,8-trichlorodibenzofuran
AhR	Aryl hydrocarbon Receptor
AhRE-II	AhR Response Element like sequence
AhRR	AhR Repressor
AIP	AhR-Interacting Protein
Aldh3a1	Aldehyde dehydrogenase 3 family, member a1
APC	Antigen Presenting Cell
Arnt	Aryl hydrocarbon receptor nuclear translocator
α-NF	α-Naphthoflavone
B[α]P	Benzo[α]pyrene
bHLH	basic Helix-Loop-Helix
bHLH.Zip	basic Helix-Loop-Helix Zipper
β-NF	β-Naphthoflavone
CAT	Chloramphenicol Acetyltransferase
CD	Circular Dichroism
Cdk	Cyclin dependent kinase
cDNA	complementary DNA
C/EBPβ	CCAAT/Enhancer-Binding Protein β
CH-223191	2-Methyl-2H-pyrazole-3-carboxylic Acid (2-methyl-4-o-tolylazo-phenyl)-amide
CRM1	Chromosome Region Maintenance
CRP	C-Reactive Protein
CTL	Cytotoxic T Lymphocyte
Cyp	Cytochrome P450
DBD	DNA Binding Domain
DC	Dendritic Cell
DIM	Diindolylmethane

DiMNF	3’,4’-Dimethoxy-α-naphthoflavone
DMSO	Dimethyl Sulfoxide
DN	Double Negative
DNA	Deoxyribonucleic Acid
DV	Ductus Venous
E1	Estrone
E2	17β-Estradiol
EAE	Experimental Autoimmune Encephalomyelitis
E-box	Enhancer Box
EMSA	Electrophoretic Mobility Shift Assays
ER	Estrogen Recepter
ERE	Estrogen Responsive Elements
FCS	Fetal Calf Serum
FICZ	6-formylindolo[3,2-b]carbazole
FoxP3	Forkhead box P3
GNF351	N-(2-(1H-indol-3-yl)ethyl)-9-isopropyl-2-(5-methylpyridin-3-yl)-9H-purin-6-amine
GST-Ya	Glutathione-S-transferase
GVH	Graft-vs-Host
HAH	Halogenated Aromatic Hydrocarbon
HDC	Histidine Decarboxylase
HES-1	Hairy and Enhancer of Split homology-1
HIF	Hypoxia Inducible Factor
H-NOXA	Heme-Nitric oxide/Oxygen binding Associated
HP	Haptoglobin
HSC	Hematopoietic Stem Cell
Hsp90	Heat shock protein 90
H/W	Han/Wistar
IDO	Indoleamine-2, 3-dioxygenases
iDRE	inhibitory Dioxin Response Element
IFN-γ	Interferon-γ

IGFBP-1	Insulin-like Growth Factor Binding Protein-1
IL	Interleukin
IMDM	Iscove's Modified Dulbecco's Medium
KA	Kynurenic Acid
kb	kilo base
Kd	Dissociation constant
KLF2	Kruppel-Like Factor 2
KO	Knock Out
LBD	Ligand Binding Domain
LD50	Lethal Dose 50
LDL	Low Density Lipoprotein
L-E	Long-Evans
LMB	Leptomycin B
LOV	light, oxygen, voltage
LPS	Lipopolysaccharide
MD	Molecular Dynamics
MHC-II	Major Histocompatibility Complex class II
MNF	3’-methoxy-4’-nitroflavone
MOG35-55	Myelin Oligodendrocyte Glycoprotein peptide 35-55
NES	Nuclear Export Sequence
NF-κB	Nuclear Factor-kappa B
NK	Natural Killer
NLS	Nuclear Localization Sequence
NMT2	N-Myristoyltransferase 2
NOD	Nonobese Diabetic
NPAS	Neuronal PAS protein
Nqo1	NAD(P)H quinone oxidoreductase 1
NR	Nuclear Hormone Receptor
PAH	Polycyclic Aromatic Hydrocarbons
Pai-2	Plasminogen activator inhibitor-2

PAS	Per-Arnt-Sim homology
pCA	p-Coumaric Acid
PCDD	Polychlorinated Dibenzo-p-dioxins
PCDF	Polychlorinated Dibenzofuran
pentaCB	3,3',4,4',5-pentachlorobiphenyl
PIA	Pancreatic Islets Allotransplantation
pM	picomolar
Por	P450 (cytochrome) oxidoreductase
PPAR α	Peroxisome proliferator-activated receptor α
pRb	Retinoblastoma protein
PTM	Post Translational Modification
PYP	Photoactive Yellow Protein
qRT-PCR	quantitative Real Time Polymerase Chain Reaction
RAR	Retinoic Acid Receptor
RevB2H	Reverse Bacterial two Hybrid
RMSD	Root-Mean-Square Deviation
RNA	Ribonucleic Acid
RNAPII	RNA polymerase II
RPMI	Roswell Park Memorial Institute
SAA	Serum Amyloid Associated
SAhRM	Selective AhR Modulator
SAR	Structure Activity Relationship
SGA360	1-allyl-3-(3,4-dimethoxyphenyl)-7-(trifluoromethyl)-1H-indazole
Sim	Single minded
SNP	Single Nucleotide Polymorphism
SSD	Signal Sensing Domain
Stat1	Signal transducer and activator of transcription 1
TAD	Transactivation Domain
TCDD	2,3,7,8-tetrachlorodibenzo-p-dioxin
TCDF	2,3,7,8-tetrachlorodibenzofuran

TDO	Tryptophan-2,3-dioxygenase
TGF-β	Tansforming Growth Factor-β
Tiparp	TCDD-inducible poly(ADP-ribose) polymerase
TK	Thymidine Kinase
TNF-α	Tumour Necrosis Factor-α
TRAMP	Transgenic Adenocarcinoma of the Mouse Prostate
Treg	Regulatory T cell
Trh	Trachealess
Ugt1a6	UDP glucuronosyltransferase 1 family, polypeptide A6
UTR	Untranslated Region
Way-169916	4-[1-allyl-7-(trifluoromethyl)-1H-indazol-3-yl] benzene-1
wt	wild type
XRE	Xenobiotic Response Element
YH439	isopropyl-2-(1,3-dithietane-2-ylidene)-2-[N-(4-methylthiazol-2-yl)carbamoyl]acetate