

# Identification and Characterisation of Novel Substrates and Binding Partners of the Asparaginyl Hydroxylase, FIH

Rachel Jane Hampton-Smith

B. Sc. (Hons)

A thesis submitted in fulfilment of the requirements of the degree of Doctor of Philosophy (Science).

September 2015

Discipline of Biochemistry

School of Biological Sciences



THE UNIVERSITY  
*of* ADELAIDE



# Contents

Contents.....	iii
Figures quick reference .....	viii
Tables quick reference.....	x
Abstract.....	xi
Declaration .....	xiii
Acknowledgements.....	xv
Author contributions.....	xvii
<b>1 Introduction .....</b>	<b>1</b>
1.1 Oxygen and the electron transport chain .....	3
1.2 Fulfilling the cellular demand for oxygen.....	5
1.3 Adaptive mechanisms to hypoxia .....	5
1.4 HIF hydroxylases and regulation of Hypoxia-inducible Factor.....	7
1.5 FIH substrate recognition mechanisms.....	11
1.5.1 Structural basis for FIH's interaction with HIF-1 $\alpha$ .....	11
1.5.2 Substrate binding is predominantly mediated by backbone interactions.....	15
1.6 ARD-containing proteins as substrates of FIH.....	16
1.7 The FIH knockout mouse.....	18
1.7.1 FIH knockout mice are hypermetabolic, and have increased insulin sensitivity .....	18
1.8 Summary .....	19
1.9 Aims and approach .....	20
<b>2 Materials and Methods.....</b>	<b>25</b>
2.1 Abbreviations .....	27
2.2 Primers .....	30
2.2.1 Cloning.....	30
2.2.2 qPCR .....	34
2.3 Plasmids .....	36
2.3.1 Plasmids described elsewhere .....	36
2.3.2 Plasmids cloned in this work.....	38

2.4	Chemicals.....	45
2.5	Growth mediums.....	46
2.6	Antibodies.....	47
2.6.1	Primary antibodies.....	47
2.6.2	Secondary antibodies .....	48
2.7	Cloning.....	48
2.7.1	Restriction digestion.....	48
2.7.2	Sequence amplification from plasmid DNA or cDNA .....	48
2.7.3	Site directed mutagenesis .....	48
2.7.4	Overlap extension PCR.....	49
2.7.5	Gel extraction and purification of DNA fragments.....	49
2.7.6	Ligations.....	49
2.7.7	PCR product insertion into pGEM-T Easy .....	49
2.7.8	PCR product insertion into pENTR-TOPO .....	50
2.7.9	Gateway recombination .....	50
2.7.10	Bacterial transformation.....	50
2.8	DNA plasmid preparation .....	50
2.8.1	Plasmid miniprep.....	50
2.8.2	Plasmid midiprep.....	50
2.9	DNA sequencing.....	51
2.10	Electrocompetent cell preparation .....	51
2.11	Yeast 2-hybrid assay .....	51
2.12	Y2H prey plasmid identification.....	52
2.12.1	Plasmid isolation from yeast .....	52
2.12.2	Bacterial electroporation.....	52
2.13	Yeast colony hybridisation.....	52
2.13.1	Ferritin probe preparation.....	52
2.13.2	Filter preparation.....	53
2.13.3	Ferritin probe hybridisation.....	54
2.14	Protein expression and purification .....	55
2.14.1	Culture growth.....	55
2.14.2	Protein purification, batch.....	55
2.14.3	Protein purification, Profinia™ .....	56
2.15	Protein buffer exchange .....	56

2.16	Purified protein yield determination .....	56
2.17	Protein concentration .....	56
2.18	Preparation of purified tcHIF (780-879) for NMR .....	57
2.18.1	Expression, purification and concentration .....	57
2.18.2	TEV cleavage of 6His tag and tag/AcTEV removal.....	57
2.18.3	Buffer exchange and concentration.....	57
2.19	NMR .....	58
2.19.1	1D <sup>1</sup> H.....	58
2.19.2	2D NOESY.....	58
2.20	Collation and analysis of ankyrin repeat sequences across species .....	58
2.21	<i>In vitro</i> hydroxylation assay .....	59
2.22	Mammalian cell culture .....	60
2.23	Mammalian cell transfection .....	60
2.24	Mammalian whole cell extract preparation.....	60
2.25	SDS PAGE.....	61
2.25.1	Gel preparation .....	61
2.25.2	Electrophoresis.....	61
2.26	Coomassie protein staining.....	61
2.27	Western blot .....	62
2.28	Western blot band quantitation .....	62
2.29	9E10 hybridoma S/N preparation .....	62
2.30	Anti-myc co-immunoprecipitation.....	63
2.30.1	WCE preparation .....	63
2.30.2	Co-immunoprecipitation .....	63
2.31	Dox-inducible FIH MEF cell line generation .....	64
2.32	Reporter assays.....	65
2.33	qPCR analysis of G9a target genes.....	65
2.33.1	Cell treatments.....	65
2.33.2	qPCR .....	66
2.34	IkB $\alpha$ biophysical characterisation methods .....	66
<b>3</b>	<b>Results – Part 1 Isolation and characterisation of novel FIH binding partners</b>	
	67	
3.1	Introduction .....	69
3.2	Y2H potential positive identification.....	69

3.3	ARD-containing proteins are common substrates of FIH .....	74
3.4	The FIH “consensus sequence” as a predictive tool for novel substrates.....	77
3.5	Both substrate and non-substrate Y2H positives interact with FIH .....	83
3.6	Effect of ARD protein over-expression on FIH repression of the CAD .....	87
3.7	Structurally constrained Gankyrin can bind FIH.....	93
3.8	Effect of non-ARD-containing Y2H positive over-expression on FIH repression of the CAD.	96
3.9	The C-terminus of 4E-T displays DMOG-inducible binding to FIH.....	99
3.10	Stress-inducing agents do not influence the interaction between 4E-T and FIH.....	100
3.11	Discussion .....	102
3.11.1	ARD-containing proteins comprise a novel class of substrates for FIH.....	102
3.11.2	FIH’s predicted substrate repertoire – a true indication of ankyrin repeat hydroxylation within cells? .....	103
3.11.3	Ankyrin repeat substrates are effective competitors for FIH binding in cells – implications for HIF regulation .....	107
3.11.4	Functional effects of ARD hydroxylation .....	108
3.11.5	FIH interacts with cell stress signalling molecules.....	112
3.11.6	4E-T, FIH and regulation of translation.....	114
3.11.7	Conclusions .....	115
<b>4</b>	<b>Results – Part 2 An evolutionary biology approach to understanding FIH substrate recognition .....</b>	<b>117</b>
4.1	Introduction.....	119
4.2	FIH-CAD relationships in divergent species .....	119
4.3	FIH-ankyrin repeat relationships across species .....	122
4.3.1	Proteome-wide analysis of ankyrin repeat conservation.....	127
4.4	The <i>T. castaneum</i> CAD is not stably folded in solution .....	134
4.5	<i>T. castaneum</i> FIH more efficiently targets mammalian ankyrin repeats over the CAD compared to hFIH .....	136
4.6	Substrate specificity of tcFIH cannot be conferred to hFIH through binding pocket-targeted tcFIH “mimic” mutations .....	139
4.7	Discussion and Conclusions .....	143
4.7.1	Comparison of <i>T. castaneum</i> and human FIH highlight the complexity of FIH substrate recognition mechanisms.....	144
4.7.2	Were ARDs the first FIH substrates? .....	144
4.7.3	Structural diversity in FIH substrates: was FIH substrate recognition originally based on the ARD fold?.....	145

4.7.4	Functional relevance of ARD hydroxylation in higher eukaryotes.....	146
<b>5</b>	<b>Results – Part 3 Effect of FIH on G9a function .....</b>	<b>149</b>
5.1	Introduction .....	151
5.1.1	G9a and GLP: ARD-containing lysine methyltransferases.....	151
5.2	G9a and GLP are novel substrates of FIH.....	155
5.3	FIH activity does not alter expression of a subset of putative G9a-regulated target genes.....	157
5.3.1	Construction of a cell line with dox-inducible FIH expression .....	157
5.3.2	Effect of FIH manipulation on G9a target gene expression .....	161
5.4	Discussion.....	165
<b>6</b>	<b>Results – Part 4 Does asparaginyl hydroxylation of IκBα affect its stability and function? .....</b>	<b>167</b>
6.1	Introduction .....	169
6.2	Published manuscript: “Consequences of I kappa B alpha hydroxylation by the factor inhibiting HIF (FIH)” .....	169
6.3	Discussion.....	171
<b>7</b>	<b>Final Discussion and Future Perspectives .....</b>	<b>173</b>
7.1	FIH and ARDs: many targets but few consequences?.....	175
7.2	FIH and non-ARD substrates/interacting partners .....	179
7.3	The function of FIH: clues from the FIH KO mouse.....	180
7.3.1	The FIH KO mouse phenotype – is HIF involved?.....	180
7.3.2	Deciphering the FIH KO mouse – foci for future study .....	181
7.4	FIH as a therapeutic target: motivation for better understanding FIH function .....	183
7.4.1	FIH in tumour growth.....	183
7.4.2	FIH and treatment of ischaemic disease .....	184
7.4.3	FIH inhibitors: a new diabetes therapy? .....	185
7.5	Final summary.....	185
<b>8</b>	<b>Appendices.....</b>	<b>187</b>
<b>9</b>	<b>References.....</b>	<b>209</b>

# Figures quick reference

Figure 1.1 Hypoxia adaptation mechanisms..	6
Figure 1.2 Domain structure of the HIF-1 and HIF-2 transcription factors..	8
Figure 1.3 Mechanisms of HIF target gene activation in mild and severe hypoxia.....	11
Figure 1.4 Crystal structure of FIH complexed with HIF-1 $\alpha$ peptide.....	12
Figure 1.5 Multi-species alignment of FIH.....	15
Figure 1.6 Conservation of HIF- $\alpha$ CAD residues in HIF-1 $\alpha$ and HIF-2 $\alpha$ .....	16
Figure 3.1 Colony hybridisation analysis of Y2H potential positive clones.....	70
Figure 3.2 ARD-containing proteins are among positives isolated in the Y2H screen.....	73
Figure 3.3 Assessment of Y2H positive proteins as substrates of FIH by <i>in vitro</i> hydroxylation assay.	75
Figure 3.4 Effect of FIH on transcriptional activity of Gal DBD-FGIF.....	77
Figure 3.5 Determination of a “preferred binding sequence” for FIH.....	80
Figure 3.6 TCPTP and GRK2 contain a surface accessible predicted FIH target motif.....	82
Figure 3.7 Co-IP of endogenous FIH with Myc-tagged Y2H positives.....	85
Figure 3.8 Testing of non-ARD-containing proteins for interaction with endogenous FIH.....	87
Figure 3.9 Effect of Y2H positive over-expression on HIF- $\alpha$ CAD repression.....	90
Figure 3.10 Comparative structure of a human Notch1 target sequence when bound to FIH or as part of the Notch1 ARD.....	92
Figure 3.11 Assessment of binding of constrained and free Gankyrin to FIH.....	96
Figure 3.12 Effect of 4E-T fragments on FIH-mediated repression of the HIF- $\alpha$ CAD.....	98
Figure 3.13 Interaction of FIH with truncation mutants of 4E-T.....	100
Figure 3.14 Effect of cell stress inducing agents on the 4E-T-FIH interaction.....	101
Figure 3.15 I $\kappa$ B $\alpha$ foldedness is a key determinant of its ability to repress NF $\kappa$ B.....	111
Figure 4.1 Conservation of FIH and the HIF- $\alpha$ CAD across species.....	121
Figure 4.2 Conservation of FIH target sequences in Notch and I $\kappa$ B $\alpha$ across species.....	126
Figure 4.3 Analysis of ankyrin repeat repertoire characteristics across species.....	129
Figure 4.4 Proteome-wide analysis of FIH target motif conservation.....	131
Figure 4.5 1D $^1$ H and 2D NOESY NMR analysis of <i>Tribolium castaneum</i> HIF- $\alpha$ CAD.....	135
Figure 4.6 Differential substrate specificity of human and Tribolium FIH.....	139
Figure 4.7 Effect of “Tribolium mimic” mutations on the substrate specificity of hFIH.....	142
Figure 4.8 Alignment of known and predicted HIF- $\alpha$ CAD homologs.....	146



Figure 4.9 Comparison of FIH motif prevalence vs absolute repeat number in species with and without FIH/CAD. ....	148
Figure 5.1 Structure and likely target asparagines of G9a and GLP.....	155
Figure 5.2 Analysis of G9a and GLP as FIH substrates.....	156
Figure 5.3 Characterisation of dox-inducible FIH MEF cell lines.....	160
Figure 5.4 Effect of FIH over-expression on G9a target genes.. ....	164
Appendix 8.1 Hits from a Prosite scan of the UniProtKB database (limited to the Mus musculus taxonomy) using the motif L-X(5)-[DE]-[IV]-N-[AV]. ....	189
Appendix 8.2 Predicted nucleotide sequences for novel FIH homologs. ....	201
Appendix 8.3 Alignment of known and predicted FIH homologs. ....	207

# Tables quick reference

Table 1.1 List of MS-MS-verified ARD-containing substrates of FIH. ....	17
Table 3.1 In-frame cDNAs isolated from the Y2H screen. ....	71
Table 3.2 Hydroxylation efficiency of target sites within MS-verified FIH substrates.. ....	106
Table 4.1 Analysis of FIH and control motif frequency in species subgroups. ....	133
Table 5.1 G9a target genes analysed for expression in dox-inducible FIH MEF cell lines. ....	165
Table 7.1 Large ARDs with multiple FIH target residues in the human proteome. ....	178

# Abstract

The ability of cells to sense and respond to sub-optimal levels of oxygen is a key requirement for organism survival. Many cellular and physiological signalling pathways have been identified as being sensitive to oxygen levels, although remarkably, few of these pathways have been successfully linked with a genuine “oxygen sensing” molecule. Discovery of the 2-oxoglutarate-dependent asparaginyl hydroxylase, Factor Inhibiting HIF (FIH), as an oxygen sensitive regulator of the Hypoxia-inducible Factor (HIF) transcription factors has therefore led to considerable interest in the enzyme as a potential regulator of multiple oxygen-regulated processes. In this work, the known substrate repertoire of FIH was expanded using both yeast 2-hybrid (Y2H) and bioinformatics-based approaches. Potential positives identified in the Y2H screen included a number of proteins which contain an ankyrin repeat structural domain (ARD), and subsequent characterisation of these proteins by *in vitro* hydroxylation assay suggest that both Fetal Globin Inducing Factor (FGIF) and Serine/threonine-protein phosphatase 6 regulatory ankyrin repeat subunit B (PP6-ARS-B) are both novel substrates of FIH. Concomitant with this discovery, a number of other ARD-containing proteins have been reported as substrates of FIH in the literature, thus suggesting that hydroxylation of ARDs by FIH is common. Comparison of the target sites in these substrates reveals an “FIH preferred sequence” of LXXXXX[-]φN, however, it was discovered that FIH can also bind an ARD constrained in its folded state, suggesting that the tertiary fold of ankyrins may also participate in enzyme recruitment. Thus far, hydroxylation of ARDs reported in the literature has not been found to have a significant functional effect on ARD biology. Largely consistent with this, an assessment of the influence of FIH-mediated hydroxylation on IκBα stability and interaction with NFκB in this work suggested that the modification has only subtle effects. Furthermore, FIH was found to have no clear effect on the methyltransferase activity of the novel ARD-containing substrate, G9a. In addition to ARD-containing proteins, the Y2H also identified a number of non-ARD-containing proteins which displayed weak interactions with FIH that were inducible by the FIH inhibitor, DMOG. *In vitro* hydroxylation assays suggest that these proteins are not FIH substrates, and further study will be required to establish the biological significance of these interactions. Overall, this work suggests that FIH interfaces with many partners, and it remains to be determined how these interactions influence the function of FIH, as well as that of its substrates and binding proteins.



# Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name in any university or other tertiary institution 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. In addition, I certify that no part of this work will, in the future, be used in a submission in my name for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

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 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 Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Rachel Hampton-Smith

Date

8/10/2015

Published work contained herein:

Devries, I.L., Hampton-Smith, R.J., Mulvihill, M.M., Alverdi, V., Peet, D.J., and Komives, E.A. (2010). Consequences of IkappaB alpha hydroxylation by the factor inhibiting HIF (FIH). *FEBS Lett* 584, 4725-4730.



# Acknowledgements

Since this is the part of the thesis that EVERYONE reads, I'd better do a decent job (and since I'm writing it at 1 in the afternoon, instead of at 3 in the morning (like my honours thesis), hopefully it will be a little more comprehensible).

It's customary to thank one's supervisor first, and I see no reason to break with that tradition. Dan, to use an American expression, I "lucked out" when I rather randomly selected your lab to do honours. You were friendly, funny, actually gave a toss about what your students were doing, and always had a cool head on your shoulders no matter what crisis was going on around you. As a student, I valued your (always very sensible) scientific advice, which would invariably inspire me to consider 14 other aspects of my data which I hadn't yet considered. More broadly speaking, I continue to aspire to your sense of responsibility, whether it be cleaning up people's used dishes in the kitchen, or putting your best forward in the teaching arena. In a world where tertiary education is very much a commodity, you have helped me to retain my faith that learning new things is not just something I had grudgingly doled out to me because I purchased it, but something which is fun, inspiring, and may actually (dare I say it) have an impact on the wider world.

While I'm riding the philosophical wave, I'd also like to thank my collaborators. As a researcher who had some of my most interesting findings "scooped" by another group, I learnt rather early why scientists tend to keep their cards very close to their chest. For that reason, I have great appreciation for those who quickly got on board when approached about investigating a new idea, or were happy to approach me about some of my work. Special thanks goes to Betsy Komives, Ingrid Devries and the rest of the crew at UCSD, Kian Leong Lee, Lorenz Poellinger, Jolene Caifeng Ho and Jia You at the Cancer Science Institute of Singapore, Marie Bogoyevitch at the University of Melbourne, and Grant Booker, Kate Wegener and Iain Murchland at the University of Adelaide. In this category, I also have to mention Dan again, as his response whenever I queried him about sharing information or reagents was always an immediate (and refreshing) "send them whatever they need".

When it comes to thanking all the people who have shaped my time in the Peet Lab, there is an embarrassingly long list. First off, I'll mention our extended family in the Whitelaw lab: Murray, Susi, Fi, Alix, Anne C-S, Anne R, Adrienne and Veronica, you've all made life easier for me at some point! Now for the Peets. In a very approximate order of when they cohabited the lab with me, cheers to: Cam and Anthony for general cheeriness, and Sarah Linke (Karttunen) for your unforgettable impact

– you guys got to experience newbie scientist Rachel, and handled it all with aplomb. Karolinee, my comrade in arms, I'm so glad I got to struggle through it all with you! Sam, for his quiet but capable presence, and Sarah Wilkins, who, to this day remains one of the scariest (in a good way, if that's possible) people I've ever met. Words out of your mouth always had unpredictable effects, and never ceased to be entertaining. You were one of the biggest generators of the Peet Lab's social and teaching culture, a feature which I always admired and appreciated. Bec, my moral compass, a genuinely inspiring gal with a tremendous sense of responsibility (all the more amazing because it comes with a healthy dose of cynical humour). TERESA! If there's something that every lab must have, it's a Teresa. Nothing but fun and good humour, I'm greatly looking forward to your return to Oz. Briony, one of those rare creatures who brightens the world around them – I love your curiosity and optimism, 2 things which you share in buckets with everyone else. Our many honours students over the years, including Freya, Michael, Lauren (I will never forget that exchange with Pete ("why are you such a knob?")), and Wai Li (another whose mouth played host to many a surprising sentence). Two of my "own" students, Jackie Lu and Max Tollenaere, both of whom contributed work which ended up in this thesis. As all teachers know, there's nothing better than having a student who's actually engaged in what they're doing, and both put in a great effort while in the lab. Plus, Max, rifling through your lab book to find various experimental details was very funny – I didn't even need Dutch-to-English translation for those rude words! Lastly, to those who currently occupy the Peet Lab as I make my exit: Erin – a rockin' result this year – go girl! Navdeep, someone with a wonderful, cheeky sense of humour who has taught me many interesting things about Indian culture. NatandJay (think, Richard and Amy from Zits) the feisty Spanish/Italian duo that always ensure that the lab is filled with noise of some description, not to mention good food. They are lovely individuals – Nat with her humour and incomparable story telling skills (I've seen many an audience in thrall), and Jay, with her own (usually self-aimed) sense of humour, which never fails to make me laugh, even on the crap days. Finally, Ice, another wonderful, upbeat gal who shares many a giggle with me, usually over our inability to pronounce various words in the English or Thai language. Always good fun.

Last but not least, thanks to my family, who managed only minimal derogatory comments about how long it's been taking me to finish writing my thesis.



# Author contributions


The following author contribution statement details the contribution of each author to the published article, “Consequences of IkappaB alpha hydroxylation by the factor inhibiting HIF (FIH)”, which can be found in section 6.2.



# Statement of Authorship

Title of Paper	Consequences of IkappaB alpha hydroxylation by the factor inhibiting HIF (FIH)	
Publication Status	<input checked="" type="checkbox"/> Published <input type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication <input type="checkbox"/> Unpublished and Unsubmitted work written in manuscript style	
Publication Details	Devries, I.L., Hampton-Smith, R.J., Mulvihill, M.M., Alverdi, V., Peet, D.J., and Komives, E.A. (2010). Consequences of IkappaB alpha hydroxylation by the factor inhibiting HIF (FIH). FEBS Lett 584, 4725-4730. Please note: the 1 <sup>st</sup> 3 authors contributed equally	

## Co-1st Author

Name of Co-1st Author (Candidate)	Rachel Hampton-Smith	
Contribution to the Paper	Co-conceived the project, planned experiments, generated data, interpreted experiments, discussed data, edited the paper.	
Overall percentage (%)	20%	
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the joint primary author of this	
Signature		Date 30/9/2015

## Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis, and
- iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Contribution to the Paper
Ingrid Devries	Planned experiments, generated data for Fig 3, interpreted experiments, discussed data.
Melinda Mulvihill	Planned experiments, generated data for Figs 4, 5 and 6, interpreted experiments, discussed data.
Vera Alverdi	Planned experiments, generated data for Fig 2, interpreted experiments.
Daniel Peet	Co-conceived the project, planned experiments, interpreted experiments, discussed data, edited the paper.
Elizabeth Komives	Co-conceived the project, planned experiments, interpreted experiments, discussed data, wrote and edited the paper, acted as corresponding author.

Ingrid Devries Ingrid Devries

Signature		Date 10/5/15
-----------	---	--------------

# Statement of Authorship

Title of Paper	Consequences of IkappaB alpha hydroxylation by the factor inhibiting HIF (FIH)		
Publication Status	<input checked="" type="checkbox"/> Published	<input type="checkbox"/> Accepted for Publication	
	<input type="checkbox"/> Submitted for Publication	<input type="checkbox"/> Unpublished and Unsubmitted work written in manuscript style	
Publication Details	Devries, I.L., Hampton-Smith, R.J., Mulvihill, M.M., Alverdi, V., Peet, D.J., and Komives, E.A. (2010). Consequences of IkappaB alpha hydroxylation by the factor inhibiting HIF (FIH). FEBS Lett 584, 4725-4730. Please note: the 1 <sup>st</sup> 3 authors contributed equally		

## Co-1st Author

Name of Co-1st Author (Candidate)	Rachel Hampton-Smith		
Contribution to the Paper	Co-conceived the project, planned experiments, generated data, interpreted experiments, discussed data, edited the paper.		
Overall percentage (%)	20%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the joint primary author of this		
Signature		Date	30/9/2015

## Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Contribution to the Paper
Ingrid Devries	Planned experiments, generated data for Fig 3, interpreted experiments, discussed data.
Melinda Mulvihill	Planned experiments, generated data for Figs 4, 5 and 6, interpreted experiments, discussed data.
Vera Alverdi	Planned experiments, generated data for Fig 2, interpreted experiments.
Daniel Peet	Co-conceived the project, planned experiments, interpreted experiments, discussed data, edited the paper.
Elizabeth Komives	Co-conceived the project, planned experiments, interpreted experiments, discussed data, wrote and edited the paper, acted as corresponding author.

Melinda Mulvihill

Signature	Date	10/10/15
-----------	------	----------

# Statement of Authorship

Title of Paper	Consequences of IkappaB alpha hydroxylation by the factor inhibiting HIF (FIH)
Publication Status	<input checked="" type="checkbox"/> Published <input type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication <input type="checkbox"/> Unpublished and Unsubmitted work written in manuscript style
Publication Details	Devries, I.L., Hampton-Smith, R.J., Mulvihill, M.M., Alverdi, V., Peet, D.J., and Komives, E.A. (2010). Consequences of IkappaB alpha hydroxylation by the factor inhibiting HIF (FIH). FEBS Lett 584, 4725-4730. Please note: the 1 <sup>st</sup> 3 authors contributed equally

## Co-1st Author

Name of Co-1st Author (Candidate)	Rachel Hampton-Smith		
Contribution to the Paper	Co-conceived the project, planned experiments, generated data, interpreted experiments, discussed data, edited the paper.		
Overall percentage (%)	20%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the joint primary author of this		
Signature	<table border="1" style="float: right;"> <tr> <td>Date</td> <td>2/10/2015</td> </tr> </table>	Date	2/10/2015
Date	2/10/2015		

## Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Contribution to the Paper
Ingrid Devries	Planned experiments, generated data for Fig 3, interpreted experiments, discussed data.
Melinda Mulvihill	Planned experiments, generated data for Figs 4, 5 and 6, interpreted experiments, discussed data.
Vera Alverdi	Planned experiments, generated data for Fig 2, interpreted experiments.
Daniel Peet	Co-conceived the project, planned experiments, interpreted experiments, discussed data, edited the paper.
Elizabeth Komives	Co-conceived the project, planned experiments, interpreted experiments, discussed data, wrote and edited the paper, acted as corresponding author.

**Vera Alverdi**

Signature	<table border="1" style="float: right;"> <tr> <td>Date</td> <td>3/10/2015</td> </tr> </table>	Date	3/10/2015
Date	3/10/2015		

# Statement of Authorship

Title of Paper	Consequences of IkappaB alpha hydroxylation by the factor inhibiting HIF (FIH)
Publication Status	<input checked="" type="checkbox"/> Published <input type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication <input type="checkbox"/> Unpublished and Unsubmitted work written in manuscript style
Publication Details	Devries, I.L., Hampton-Smith, R.J., Mulvihill, M.M., Alverdi, V., Peet, D.J., and Komives, E.A. (2010). Consequences of IkappaB alpha hydroxylation by the factor inhibiting HIF (FIH). FEBS Lett 584, 4725-4730. Please note: the 1 <sup>st</sup> 3 authors contributed equally

## Co-1st Author

Name of Co-1st Author (Candidate)	Rachel Hampton-Smith		
Contribution to the Paper	Co-conceived the project, planned experiments, generated data, interpreted experiments, discussed data, edited the paper.		
Overall percentage (%)	20%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the joint primary author of this		
Signature	<table border="1" style="float: right;"> <tr> <td>Date</td> <td>30/9/2015</td> </tr> </table>	Date	30/9/2015
Date	30/9/2015		

## Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Contribution to the Paper
Ingrid Devries	Planned experiments, generated data for Fig 3, interpreted experiments, discussed data.
Melinda Mulvihill	Planned experiments, generated data for Figs 4, 5 and 6, interpreted experiments, discussed data.
Vera Alverdi	Planned experiments, generated data for Fig 2, interpreted experiments.
Daniel Peet	Co-conceived the project, planned experiments, interpreted experiments, discussed data, edited the paper.
Elizabeth Komives	Co-conceived the project, planned experiments, interpreted experiments, discussed data, wrote and edited the paper, acted as corresponding author.

**Daniel Peet**

Signature	<table border="1" style="float: right;"> <tr> <td>Date</td> <td>8/10/2015</td> </tr> </table>	Date	8/10/2015
Date	8/10/2015		

# Statement of Authorship

Title of Paper	Consequences of IkappaB alpha hydroxylation by the factor inhibiting HIF (FIH)
Publication Status	<input checked="" type="checkbox"/> Published <input type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication <input type="checkbox"/> Unpublished and Unsubmitted work written in manuscript style
Publication Details	Devries, I.L., Hampton-Smith, R.J., Mulvihill, M.M., Alverdi, V., Peet, D.J., and Komives, E.A. (2010). Consequences of IkappaB alpha hydroxylation by the factor inhibiting HIF (FIH). FEBS Lett 584, 4725-4730. Please note: the 1 <sup>st</sup> 3 authors contributed equally

## Co-1st Author

Name of Co-1st Author (Candidate)	Rachel Hampton-Smith	
Contribution to the Paper	Co-conceived the project, planned experiments, generated data, interpreted experiments, discussed data, edited the paper.	
Overall percentage (%)	20%	
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the joint primary author of this	
Signature	Date	30/9/2015

## Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- the candidate's stated contribution to the publication is accurate (as detailed above);
- permission is granted for the candidate to include the publication in the thesis; and
- the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Contribution to the Paper
Ingrid Devries	Planned experiments, generated data for Fig 3, interpreted experiments, discussed data.
Melinda Mulvihill	Planned experiments, generated data for Figs 4, 5 and 6, interpreted experiments, discussed data.
Vera Alverdi	Planned experiments, generated data for Fig 2, interpreted experiments.
Daniel Peet	Co-conceived the project, planned experiments, interpreted experiments, discussed data, edited the paper.
Elizabeth Komives	Co-conceived the project, planned experiments, interpreted experiments, discussed data, wrote and edited the paper, acted as corresponding author.

Elizabeth Komives

Signature	Date	30/9/2015
-----------	------	-----------





# 1 Introduction



What is the essence of life? Sex, drugs and rock 'n' roll! Health, wealth and wisdom? Let's face it: when you boil it down, the basic triad for life should read "food, water and oxygen". Mundane as this may sound, it pays to remember that maintaining an organism in a healthy, fully functional state is exceptionally hard work. At any point in time, different cells within an organism will be undergoing division, growth, differentiation, repair, and migration (to name just a few processes) and constantly monitoring and maintaining their intra- and extracellular environment in order to preserve effective tissue function. Performance of each of these tasks requires the constant synthesis and maintenance of appropriate molecular tools, as well as a source of energy to drive these molecules' functions. Essential to production of the latter in mammals, at least in a majority of situations, is molecular oxygen. Indeed, without the reduction of oxygen in the mitochondrion, efficient energy production ceases, and death is quick to follow. Thus, oxygen is a vital molecule for survival, and much of an organism's physiology and molecular biology is devoted to ensuring the supply is adequate for maintenance of homeostasis.

I shall begin this thesis with a short review of oxygen homeostasis, and the methods by which it is maintained. This shall serve as an ideal introduction for the real subject for this thesis, the oxygen-sensing enzyme, Factor Inhibiting HIF (FIH), one of a currently small number of oxygen sensors believed to contribute to preserving the knife-edge balance between oxygen supply and demand.

## 1.1 Oxygen and the electron transport chain

In the twenty-first century, a myriad of mechanical and electrical machines can be found which require a constant source of energy in order to exist in a "working" state. Living organisms are no different, but rather than topping up on petrol or plugging into a power point, mammals must make their own fuel "on the fly" by regularly consuming food, water and oxygen. Each of these three ingredients is equally important for ensuring an organism's survival. However, the need for oxygen is lent particular notoriety due to the rapid death that occurs in its absence. A significant proportion of emergency and resuscitation medicine is geared towards maintaining adequate tissue oxygenation, as mere minutes without a supply of oxygen to the brain can result in severe brain injury and death. This vulnerability has prompted much study into the reasons for this rapid loss of tissue vitality, and methods by which such injury can be lessened, or avoided altogether.

Fundamental to this research, is an understanding of the physiological functions of inhaled oxygen, of which there are many. According to the Gene Ontology database, there are at least 190 human enzymes which utilise molecular oxygen during catalysis [Amigo version 1.8, database release

29/11/11]. It is without question, however, that cytochrome c oxidase (COX) of the mitochondrion takes precedence as the most important oxygen-requiring enzyme complex in the mammalian organism. COX is responsible for transferring electrons along the mitochondrial electron transport chain (ETC) to the final electron acceptor, oxygen, reducing it to water. The energy released by the ETC redox reactions is harnessed to pump protons across the inner mitochondrial membrane, which then return along a concentration gradient to the mitochondrial matrix through the mitochondrial ATP synthase. This final step results in the conversion of adenosine diphosphate (ADP) and inorganic phosphate ( $P_i$ ) into adenosine triphosphate (ATP), the predominant energy currency of the cell (Sherratt, 1991). Without oxygen, the ETC machinery ceases to function, and this has a number of detrimental effects on the cell.

The most obvious of these effects is, of course, a lack of ATP production by the mitochondrion. The energy released by conversion of ATP to ADP is required to drive thousands of different processes. For example, DNA synthesis, protein synthesis, and ubiquitin-dependent protein degradation are all significant energy consumers. Also of considerable note, is the cost of maintaining cellular ion gradients. An average of 36 per cent of ATP production goes towards maintaining the  $Na^+/K^+$  gradients across the plasma membrane (Hand and Menze, 2008). Indeed, ATP-driven maintenance of cellular polarity across both plasma and mitochondrial membranes is critical for avoiding increases in intracellular calcium concentration, which may trigger apoptotic or necrotic cell death (Hand and Menze, 2008).

A second significant problem presented by ETC cessation results from the cell's attempts to maintain ATP production in the absence of oxidative phosphorylation. Whilst ATP can be produced anaerobically, the process is far less efficient than oxidative phosphorylation, and can cause a hazardous drop in pH due to lactic acid production. In combination with ATP depletion, acidification can be a significant contributor to initiating cell death processes (Lipton, 1999).

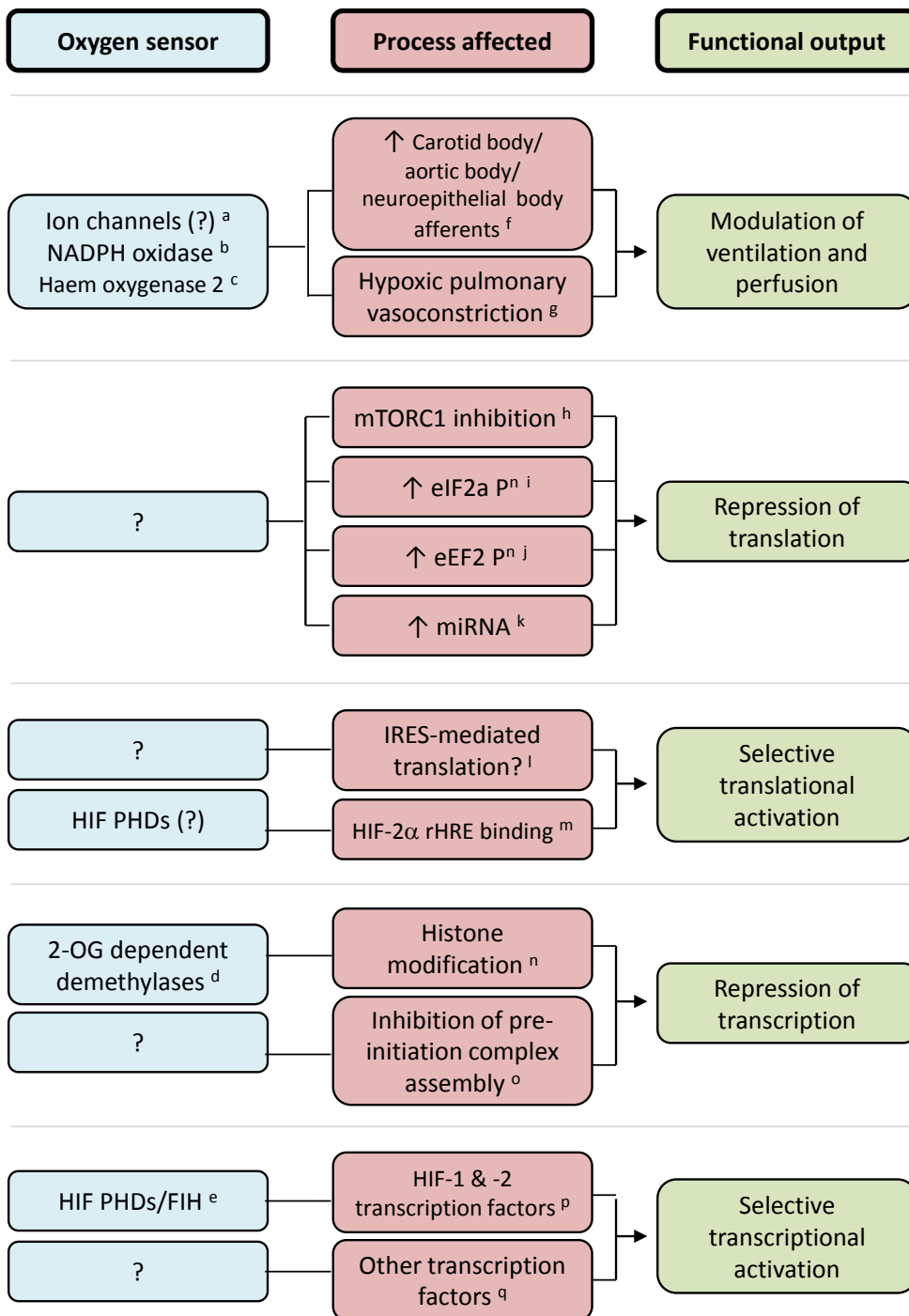
The third danger of oxygen deprivation concerns the by-products generated by the ETC itself. Fluctuations in available oxygen have been shown to increase the formation of reactive oxygen species (ROS), and the damage caused by these highly reactive molecules also likely contributes to the initiation of cell death when oxygen supply is compromised (Lipton, 1999). Clearly, a functional ETC is an essential guard against cell death, and it is therefore of great importance that cells constantly receive adequate oxygen to maintain this energy production system.

## 1.2 Fulfilling the cellular demand for oxygen

How much oxygen, then, is considered “adequate” for a human body? It has been estimated that the average human male at rest produces  $9 \times 10^{20}$  molecules of ATP via oxidative phosphorylation every second - a feat requires approximately 380 litres of oxygen every day (Rich, 2003)! In essence, every day, hundreds of litres of oxygen must be inhaled by the lungs, taken up by the pulmonary capillaries, and then faultlessly delivered to each and every oxidative phosphorylation-utilising cell in the body (Rich, 2003). Usually, physiological systems perform this activity admirably well. However, acute injury, disease and even normal growth and development can perturb efficient oxygen delivery such that *cellular demand for oxygen exceeds supply*. Such an “oxygen-insufficient” environment is referred to as *hypoxia*. Many pathologies can contribute to systemic or local hypoxia, such as chronic obstructive pulmonary disease and emphysema, which inhibit blood diffusion across the blood gas barrier in the lungs, sickle cell or haemolytic anaemias, which reduce the oxygen-carrying capacity of the blood, and blood vessel obstruction, which prevents adequate blood perfusion. Thus, hypoxia has many different causes, and may rapidly prove fatal if not ameliorated. Indeed, despite the clear advantages of using oxygen for more efficient ATP production, a reliance upon it is also something of an “Achilles’ heel”. As such, in order to minimise physiological vulnerability, the body is armed with an array of adaptive mechanisms aimed at minimising the impact of oxygen deprivation.

## 1.3 Adaptive mechanisms to hypoxia

The ability of an organism to sense and respond to a potentially life-threatening drop in oxygen supply is of great importance, and has resulted in the evolution of a variety of adaptive mechanisms. Broadly speaking, these mechanisms aim to improve oxygen transport and delivery, generate ATP by oxygen-independent means, reduce ATP expenditure, and up-regulate hypoxia-protective pathways within cells (Figure 1.1, green boxes). For example, upon sensing hypoxia, the peripheral chemoreceptors (including the carotid and aortic bodies) increase afferent nerve impulses to the medullary respiratory centre and the vasomotor centre, thus modulating ventilation and blood perfusion (Seeley, 2006; West, 2008). Hypoxia is also known to repress “temporarily expendable” cellular processes, such as protein translation and gene transcription, thus conserving ATP for the processes most critical for cell survival (such as maintenance of cellular ion gradients, see section 1.1) (Buttgereit and Brand, 1995; Fahling, 2009; Johnson et al., 2008; Wieser and Krumschnabel, 2001). Perhaps counterintuitively, hypoxia also has the ability to up-regulate the production or activity of a



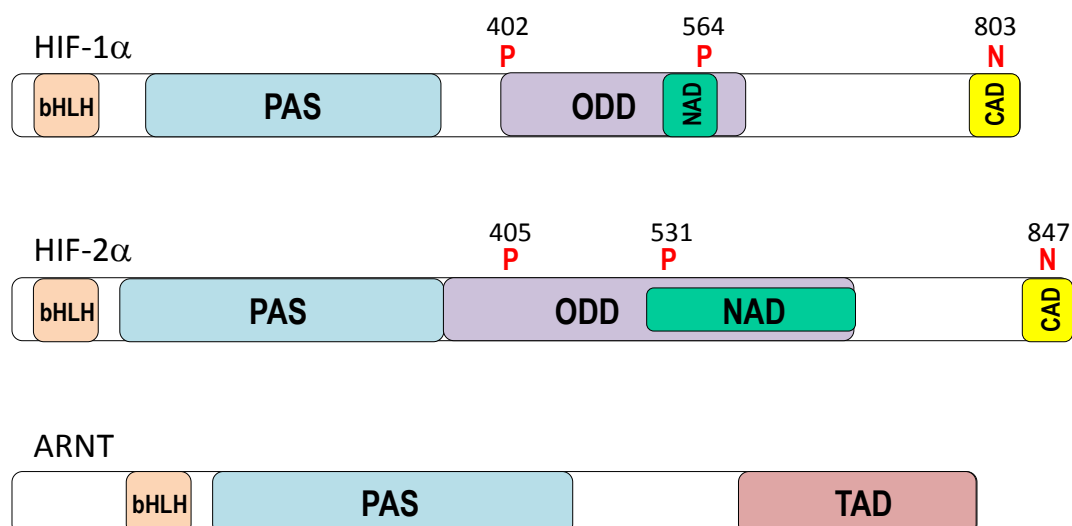
**Figure 1.1 Hypoxia adaptation mechanisms.** The functional outputs of major hypoxia adaptation mechanisms are shown (green boxes), together with known or proposed effectors and oxygen sensors for each process (pink and blue boxes, respectively; question marks denote unknown sensors, or controversial/untested candidates). a, (Shimoda and Polak, 2011); b, (O'Kelly et al., 2000; Wang et al., 1996); c, (Adachi et al., 2004; Williams et al., 2004); d, (Mimura et al., 2011); e, (Epstein et al., 2001; Lando et al., 2002a); f, (Seeley, 2006); g, (Sylvester et al., 2012); h, (Liu et al., 2006); i, (Fahling, 2009); j, (Horman et al., 2002); k, (Pocock, 2011); l, (Akiri et al., 1998; Lang et al., 2002; Young et al., 2008); m, (Uniacke et al., 2012); n, (Mimura et al., 2011); o, (Denko et al., 2003); p, (Semenza, 2003); q, (Cummins and Taylor, 2005).

select number of genes and proteins which help to alleviate the severity of oxygen deprivation, or protect against cell damage and death. Indeed, a significant number of transcription factors are known to be activated by oxygen depletion (for examples, see (Cummins and Taylor, 2005)). Of these, the Hypoxia-inducible Factors (HIFs, which includes the 3 paralogs HIF-1, HIF-2 and HIF-3) are by far the most potent, with over 70 hypoxia-inducible genes confirmed to have direct HIF binding sites, and hundreds more implied as targets of HIF by microarray and ChIP-Seq studies (Elvidge et al., 2006; Mole et al., 2009; Wenger et al., 2005; Zhang et al., 2014). The HIF transcriptional response (predominantly mediated by HIF-1 with some contribution from HIF-2, with the influence of HIF-3 remaining uncertain) serves many different purposes. For instance, HIF-regulated gene products have considerable influence over blood perfusion and oxygen-carrying capacity, glycolytic flux, and cell proliferation (reviewed in (Semenza, 2003)). Additionally, some HIF target genes also play important roles in protecting cells against apoptosis-inducing molecules such as ROS, which are often produced in response to oxygen fluctuations (see (Fukuda et al., 2007; Kim et al., 2006a; Zhang et al., 2008) for examples).

Despite a comparatively robust knowledge of the “functional outputs” of hypoxia adaptation mechanisms, surprisingly few of these processes have been conclusively linked to an oxygen sensor (see Figure 1.1, blue boxes). Notable exceptions, are the HIF transcription factors, which are known to be regulated by a group of iron and 2-oxoglutarate (2-OG)-dependent oxygenases, namely Prolyl hydroxylase domain-containing 1-3 (PHDs 1-3), and Factor Inhibiting HIF (FIH). The method by which these hydroxylases regulate HIF activity is detailed below.

#### 1.4 HIF hydroxylases and regulation of Hypoxia-inducible Factor

HIF-1 and HIF-2, regarded as the primary regulators of oxygen-dependent gene transcription, are heterodimers made up of 2 basic helix-loop-helix Per ARNT Sim (bHLH-PAS) domain-containing transcription factors. The  $\alpha$  subunit is interchangeable, and can be either of the 2 paralogs HIF-1 $\alpha$  or HIF-2 $\alpha$ , whilst the  $\beta$  subunit is the common partner factor, ARNT (Figure 1.2). Early studies of HIF-1 revealed that it was purely the  $\alpha$  subunit which rendered HIF-1 activity sensitive to oxygen levels. In fact, the  $\alpha$  subunit was observed to be regulated by hypoxia in 2 distinct ways. Firstly, the normoxic half-life of HIF-1 $\alpha$  was determined by pulse chase to be exceptionally short (less than 5



**Figure 1.2 Domain structure of the HIF-1 and HIF-2 transcription factors.** Domain structures of human HIF-1 $\alpha$ , HIF-2 $\alpha$  and ARNT are shown. Prolyl and asparaginyl residues hydroxylated by the PHDs (Pro402 and Pro564 in HIF-1 $\alpha$ , Pro405 and Pro531 in HIF-2 $\alpha$ ) and FIH (Asn803 in HIF-1 $\alpha$  and Asn847 in HIF-2 $\alpha$ ) are shown in red type along with their aa number above the relevant schematic. bHLH, basic helix loop helix; PAS, Per ARNT Sim homology domain; ODD, oxygen dependent degradation domain; NAD, N-terminal transactivation domain; CAD, C-terminal transactivation domain; TAD, transactivation domain.

minutes), whilst exposure of cells to the hypoxia mimetic chemical, desferrioxamine (DFO), prolonged this half-life to approximately 30 minutes (Huang et al., 1996; Huang et al., 1998). This normoxic lability was eventually attributed to prolyl hydroxylation of Pro402 and Pro564 within the oxygen dependent degradation domain (ODD, aa 401-603 (unless otherwise stated, all numbering refers to human HIF-1 $\alpha$ )) by the 2-OG-dependent oxygenases, PHDs 1-3. These modifications can each independently promote recognition of HIF-1 $\alpha$  by the Von Hippel Lindau (VHL) E3 ubiquitin ligase, thus promoting HIF-1 $\alpha$ 's ubiquitination and proteasomal degradation.

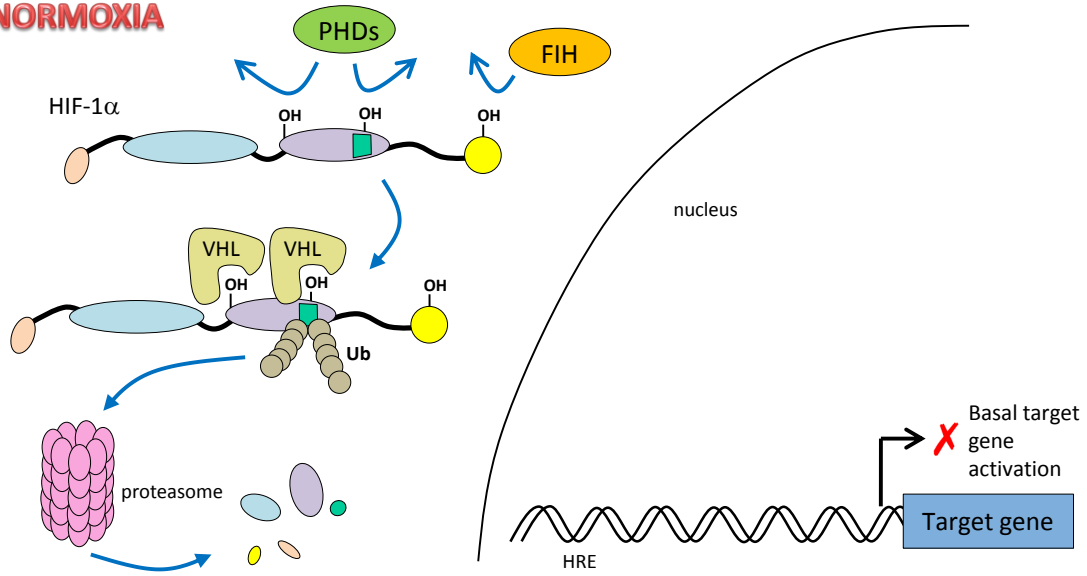
In addition to changes in protein stability, hypoxic conditions were also observed to strongly induce transactivation from the C-terminal transactivation domain (CAD, aa 727-826) of HIF-1 $\alpha$  (for clarity, it should be noted that the minimally functional CAD was originally defined as the last 40 aa of HIF-1 $\alpha$ , however, in much of the literature, and henceforth in this thesis, the name "CAD" is used to refer to the C-terminal 90 or 100 aa of HIF-1 $\alpha$  or -2 $\alpha$ , both of which include the minimal transactivation domain as well as an adjacent region which is required for the CAD's oxygen sensitivity (see section 1.5.1 for details)) (Jiang et al., 1997; O'Rourke et al., 1999; Pugh et al., 1997; Sang et al., 2002). In a similar manner to the ODD, the CAD was also found to be hydroxylated in normoxia, but in this case



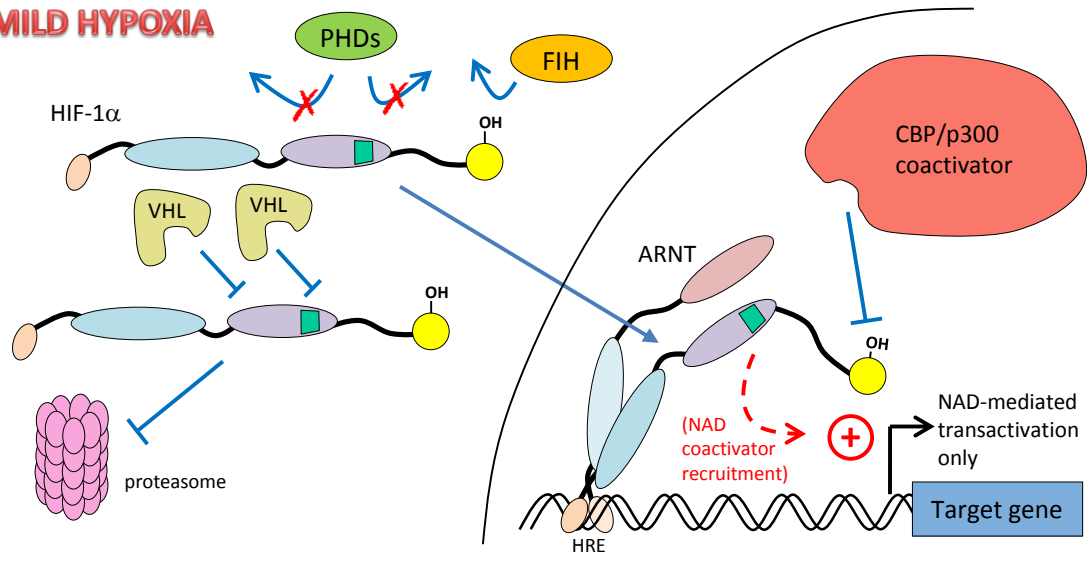
on an asparaginyl residue (Asn803, Figure 1.2) by another 2-OG-dependent oxygenase, FIH (Hewitson et al., 2002; Lando et al., 2002a; Lando et al., 2002b). Hydroxylation of this residue was found to block the binding of p300 and CBP coactivators to HIF-1 $\alpha$ , likely through a mechanism involving steric hindrance, and unfavourable polar/hydrophobic clashes at the CAD/coactivator binding interface. Thus, hydroxylation of the CAD is an effective “hypoxic switch” for CAD transactivation (Dames et al., 2002; Freedman et al., 2002; Lando et al., 2002b; McNeill et al., 2002a). Importantly, both the PHDs and FIH directly utilise dioxygen in their modification of HIF-1 $\alpha$  (Hewitson et al., 2002; McNeill et al., 2002b), and have  $K_m$ s for oxygen (between 100 and 250  $\mu$ M for the PHDs, and between 50 and 90  $\mu$ M for FIH) in a range which is well above oxygen concentrations likely to be found in tissues (Carreau et al., 2011; Hirsila et al., 2003; Koivunen et al., 2004; Mik, 2013; Wilkins et al., 2009). Thus, upon onset of hypoxia, enzymatic modification of HIF-1 $\alpha$  would be reduced, thus stabilising the protein and facilitating its interaction with coactivators, which in turn promotes activation of HIF-1 target genes (Figure 1.3).

Given that the HIF hydroxylases have proven roles as oxygen sensors, and that many oxygen-sensitive pathways are yet to be paired with a genuine oxygen sensor (see section 1.3), it would seem plausible for the enzymes in this family to have multiple substrates, thus expanding their oxygen-sensing capabilities to pathways beyond the HIFs. Indeed, since the initial characterisation of these enzymes, multiple substrates (both confirmed or likely) have been identified for both the PHDs (e.g. Rpb1, IKK $\beta$ ,  $\beta_2$ AR, Sprouty2, MYBBP1A, and ALS-2) and FIH (see section 1.6) (Abu-Farha et al., 2005; Anderson et al., 2011; Cummins et al., 2006; Kuznetsova et al., 2003; Lai et al., 2011; Xie et al., 2009). However, at the commencement of the work for this thesis, only one potential non-HIF target for the PHDs had been discovered (Rpb1) (Kuznetsova et al., 2003), and HIF-1 $\alpha$  and -2 $\alpha$  remained the only known targets for FIH. **Therefore, it was decided to search for novel substrates of the asparaginyl hydroxylase, FIH, thus potentially expanding the known role of this enzyme in oxygen sensing.**

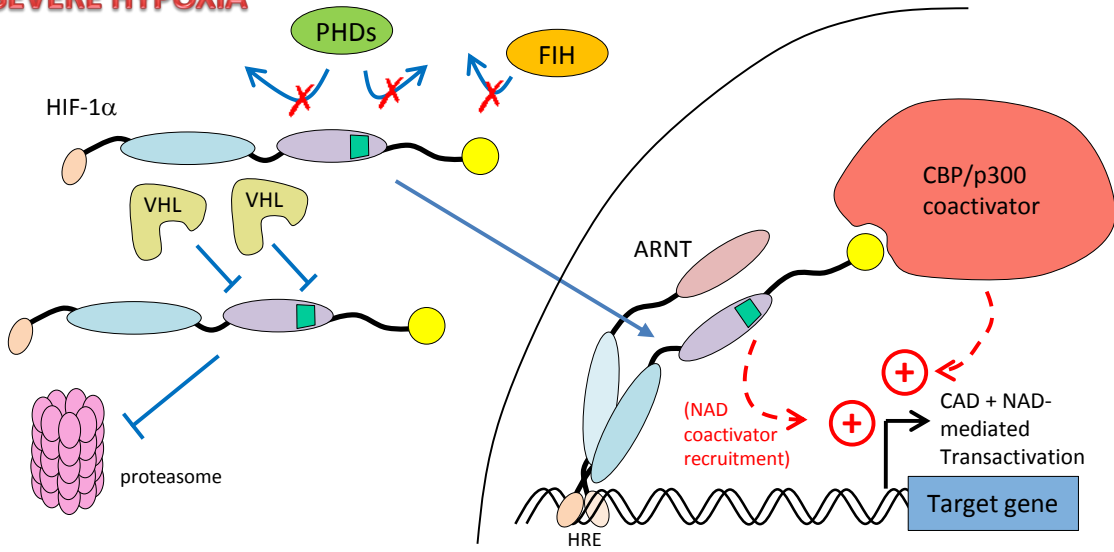
### NORMOXIA



### MILD HYPOXIA



### SEVERE HYPOXIA



(Figure 1.3)

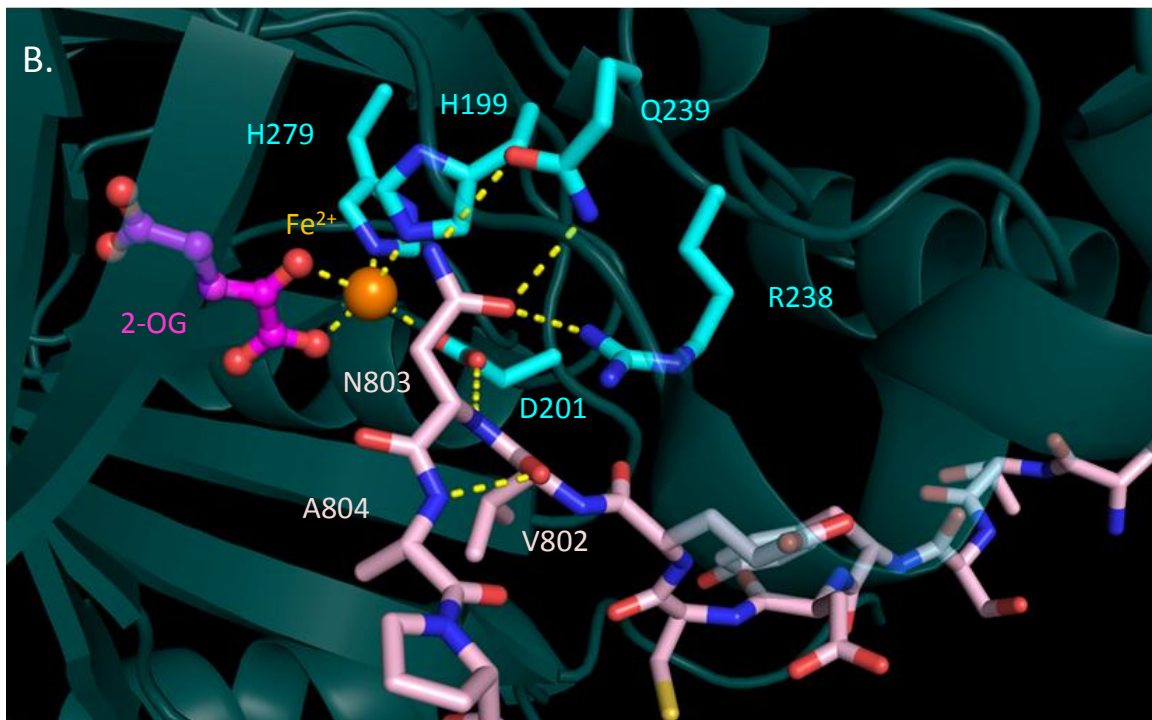
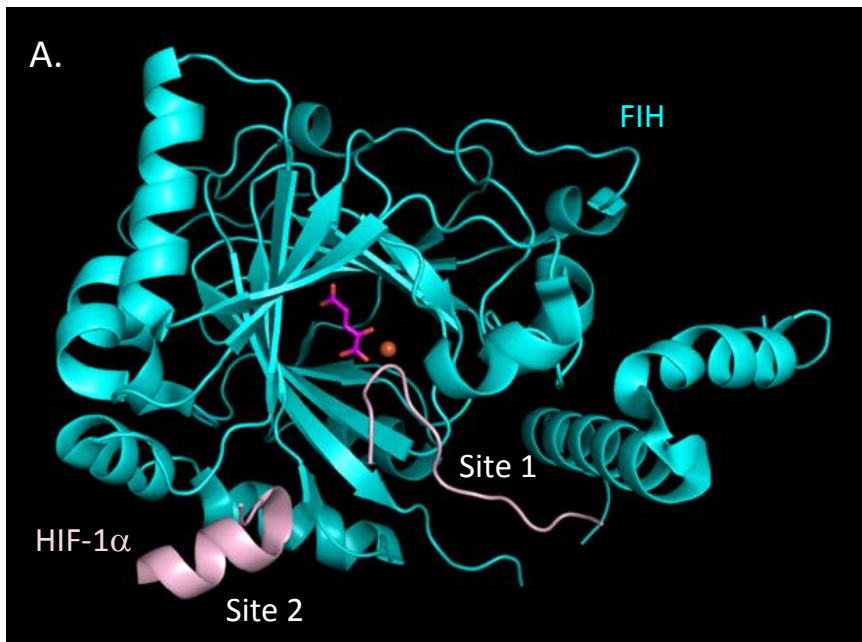
**Figure 1.3 Mechanisms of HIF target gene activation in mild and severe hypoxia.** In normoxia, the PHDs hydroxylate 2 prolyl residues in the ODD (the purple domain in HIF-1 $\alpha$ ), allowing binding of HIF-1 $\alpha$  by the E3 ubiquitin ligase, VHL, which in turn promotes ubiquitination and degradation of HIF-1 $\alpha$ . In mild hypoxia, PHD activity is inhibited (thus promoting stabilisation of HIF-1 $\alpha$ ) while hydroxylation of Asn803 in the CAD is maintained. Thus, when HIF-1 $\alpha$  translocates to the nucleus and pairs with ARNT on the hypoxia response element (HRE), CAD-mediated transactivation is blocked, leaving only the NAD (a second, constitutively active transactivation domain in HIF-1 $\alpha$ ) to drive activation of HIF target genes. In severe hypoxia, both PHD and FIH-mediated hydroxylation ceases, such that HIF-1 $\alpha$  may drive target gene expression via the NAD, CAD or both. See Figure 1.2 for an explanation of HIF-1 $\alpha$  domain colouration.

## 1.5 FIH substrate recognition mechanisms

From the outset, discovery of FIH as an enzyme with oxygen-sensing properties developed interest in two separate areas: firstly, the specific role of FIH in HIF regulation, and secondly, the possible contribution of FIH to other oxygen-sensitive signalling pathways. To assist study of the latter idea, a clearer understanding of FIH substrate recognition was required, which in turn could assist with bioinformatic prediction of novel FIH substrates.

### 1.5.1 Structural basis for FIH's interaction with HIF-1 $\alpha$

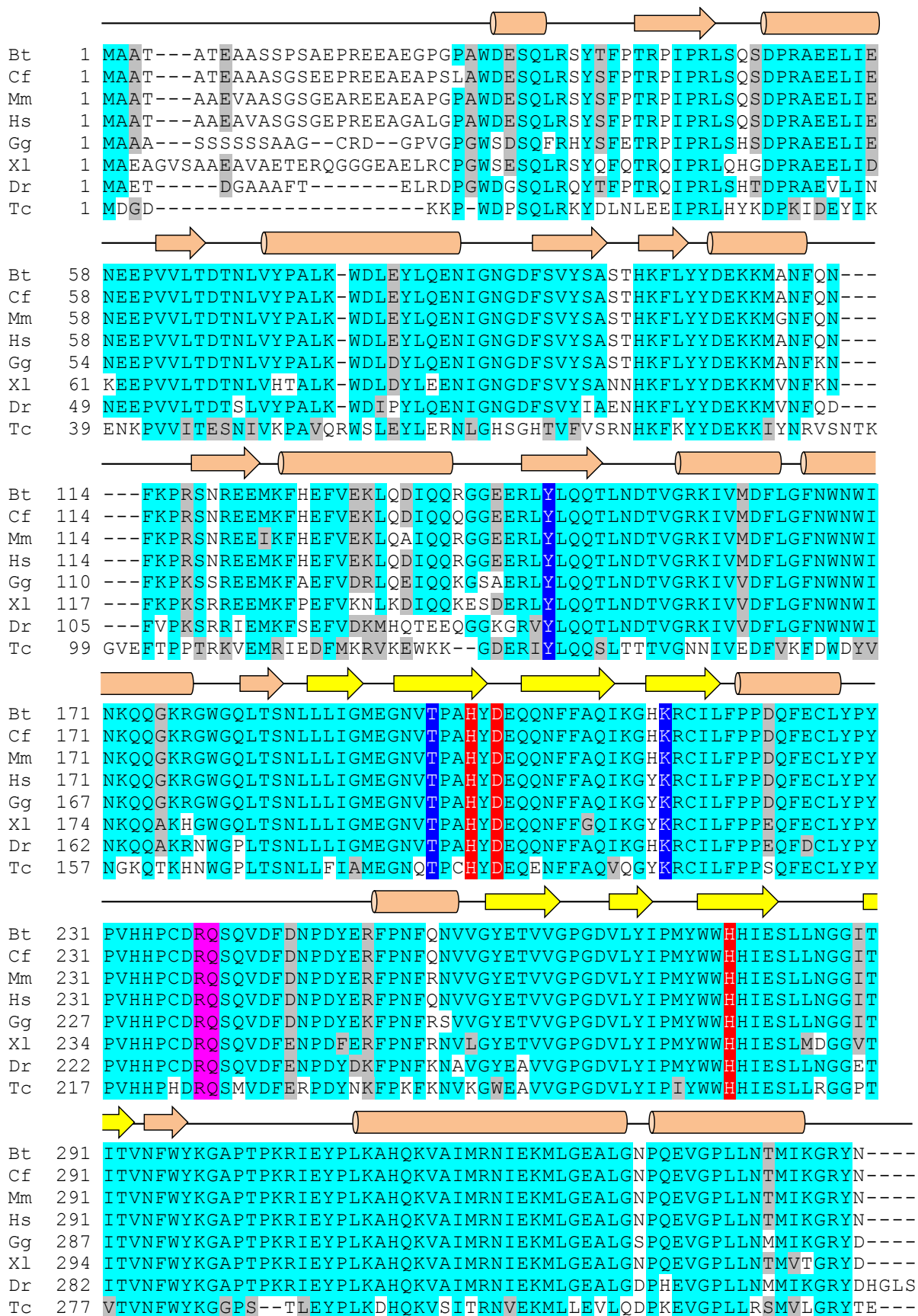
Sequence analysis of FIH post its discovery as a HIF repressor suggested that the protein contained a HXD/E...H "facial triad", a motif characteristic of the Fe<sup>2+</sup> and 2-OG-dependent family of dioxygenases (Hewitson et al., 2002; Lando et al., 2002a). Subsequent publication of 2 independent crystal structures of FIH in 2002 confirmed this to be the case, with the enzyme adopting a central double-stranded beta helix (DSBH) which supported the HXD/E...H motif in a conformation suitable for Fe<sup>2+</sup> binding (Figure 1.4A and B). Further inspection of the crystal structure allowed identification of the amino acids employed for 2-OG cofactor binding, and also revealed that FIH operates as a dimer by interacting through its C-terminal alpha helices (Figure 1.4B, Figure 3.10 and see Figure 1.5 for residue identities). In support of the important role of these residues in FIH function, comparison of FIH homologs from different species shows strong conservation of key residues of the catalytic core, the DSBH and dimerisation helices, even in species as distantly related as the *Tribolium castaneum* beetle (Figure 1.5).



**Figure 1.4 Crystal structure of FIH complexed with HIF-1 $\alpha$  peptide.** (A) An FIH monomer (shown in cyan) is depicted bound to HIF-1 $\alpha$  aa 795-826 (light pink; note that no electron density is shown for aa 807-812 of HIF-1 $\alpha$ ). The active site Fe<sup>2+</sup> (orange circle) and 2-OG cofactor (purple) are also shown. “Site 1” and “Site 2” indicate the 2 locations where the HIF- $\alpha$  CAD interacts with FIH. (B) The active site of FIH is shown, with FIH residues responsible for coordinating the Fe<sup>2+</sup> (orange) or positioning the HIF-1 $\alpha$  substrate peptide (light pink) shown in cyan. Hydrogen bonds and Fe<sup>2+</sup> coordination interactions are indicated with dotted yellow lines, and the surrounding FIH structure is shown in semi-transparent cartoon form. *Structure PDB reference 1H2L. Structure images generated using Pymol version 1.3.*

Taking this work a step further, the group of Christopher Schofield reported the crystallisation of FIH with a HIF-1 $\alpha$  substrate peptide (aa 775-826), thus showing that the C-terminus of HIF-1 $\alpha$  contacts FIH in two discrete locations (Elkins et al., 2003). "Site 1" (employing residues 795-806 of the CAD) binds along a groove in an extended conformation, while "site 2" (residues 813-822) forms an alpha helix which rests in a hydrophobic patch on the surface of FIH (Figure 1.4A) (Elkins et al., 2003). No electron density was observed for either the N-terminal 20 amino acids of the HIF-1 $\alpha$  peptide, or for the 6 amino acid linker between site 1 and 2, suggesting that these regions have minimal contact with FIH, and little definitive structure (Elkins et al., 2003). The lack of interaction of FIH with the N-terminus of the peptide was somewhat surprising, as a number of studies have indicated that amino acids 775-785 of HIF-1 $\alpha$  play a crucial role in promoting normoxic repression of CAD activity (O'Rourke et al., 1999; Sang et al., 2002). In particular, mutation of an RLL motif (residues 781 to 783) was found to greatly increase normoxic activity of a GalDBD-HIF-1 $\alpha$  (775-826)-driven reporter construct, and reduce its hypoxic inducibility by almost 10-fold (O'Rourke et al., 1999). In 2012, these findings were extended to show that an intact di-leucine motif (residues 782 and 783) can greatly improve capture of FIH by HIF-1 $\alpha$  fragments *in vitro* using pulldown assays. Intriguingly, the positive effect of the motif was reliant upon conditions of molecular crowding, as demonstrated by a loss of efficacy when interaction assays were performed in dilute buffers (Wilkins et al., 2012). Exactly how the di-leucine motif and molecular crowding work together to promote an interaction between FIH and HIF-1 $\alpha$  is unknown, but it has been proposed that crowded conditions may promote di-leucine-dependent structure formation within the CAD (which is normally unstructured in dilute solution), and this in turn participates in FIH-HIF binding interactions (Wilkins et al., 2012).

An additional notable finding from the FIH/HIF-1 $\alpha$  crystal structure is that a backbone hydrogen bond is formed between the residues on either side of Asn803 in HIF-1 $\alpha$ , thus forming a tight turn which propels the target residue towards the active site iron. Once there, the Asn side chain is held in place by hydrogen bonding (H-bonding) with FIH Asp201, Arg238 and Gln239 (Figure 1.4B), thus carefully rendering the *pro-S* hydrogen of the  $\beta$ -carbon available for attack by iron-bound superoxide (Elkins et al., 2003; McNeill et al., 2002a). Indeed, this precise Asn803 positioning is critical, as mutation of Gln239 vastly increases the rate of non-productive autohydroxylation (rather than substrate hydroxylation) by FIH (Saban et al., 2011). Thus, FIH's catalytic mechanism and precise positioning of the CAD ensures its utility as an oxygen sensor by tightly coupling oxygen consumption to substrate hydroxylation.



(Figure 1.5)

**Figure 1.5 Multi-species alignment of FIH.** Clustal W2 was used to align the FIH protein from multiple species, and residues strongly or partially conserved are shown with cyan and grey highlights, respectively. Residues involved in Fe<sup>2+</sup> coordination (red), 2-OG binding (dark blue), and target asparagine positioning (pink) are all completely conserved. The secondary structure of hFIH is depicted above the alignment, with yellow arrows indicating the beta strands which make up the DSBH. Amino acid numbers are shown to the left of the alignment. Alignment shading performed using the Boxshade server. Bt = *Bos taurus*, Cf = *Canis lupus familiaris*, Mm = *Mus musculus*, Hs = *Homo sapiens*, Gg = *Gallus gallus*, Xl = *Xenopus laevis*, Dr = *Danio rerio*, Tc = *Tribolium castaneum*.

### 1.5.2 Substrate binding is predominantly mediated by backbone interactions

Given the vital role of CAD positioning in FIH-mediated hydroxylation, one might anticipate the involvement of a high degree of side chain specificity in CAD binding as a whole. Rather, the opposite seems to be the case. Of the 6 hydrogen bonds predicted to form at site 1 (excluding those involved in Asn803 orientation and tight turn formation), and the 2 predicted for site 2, only 2 hydrogen bonds (both involving Tyr798) are mediated by CAD side chain atoms (Elkins et al., 2003). A study by Linke et al (2004) using HIF-1 $\alpha$  737-826 supports this apparent lack of specificity in FIH substrate binding, wherein a selection of residues (including hydrogen bond contributor Tyr798, Asp799-Asn803, Pro805 and G808), were mutated individually to alanine, and the mutant substrates then tested by *in vitro* hydroxylation assay. Among the variants assessed, only alteration of Val802 (aside from the target Asn itself) significantly affected HIF-1 $\alpha$  737-826 hydroxylation by FIH, and this was predicted to be caused by altered catalysis rather than substrate binding (Linke et al., 2004). Thus, despite considerable conservation of CAD residues between HIF- $\alpha$  paralogs and also across species (Figure 1.6 and see also Figure 4.8), few residues can autonomously influence CAD binding and hydroxylation. This observation raised several important points. Firstly, the lack of side chain involvement in FIH substrate binding suggested that FIH may be capable of accommodating structurally diverse proteins in its substrate-binding groove, thus strengthening the case for the existence of a broader substrate repertoire. However, the initial inability to assign a critical role to a majority of the CAD's FIH-binding residues precluded the definition of a strong "consensus sequence" and called into question use of such a consensus to predict novel FIH targets. In light of this, alternative techniques, such as a non-directed screen for FIH binding partners, would likely have a greater chance of success, and hence was utilised in the initial screen included in this thesis. Notably, while the work for this thesis was being conducted, an approach of this kind was also successfully

Hs_HIF-1a	776	SDLACRLLGQSMDESGLPQLTSYDCEVNAPIQGSRNLLQGEELLRALDQVN
Mm_HIF-1a	786	SDLACRLLGQSMDESGLPQLTSYDCEVNAPIQGSRNLLQGEELLRALDQVN
Hs_HIF-2a	820	SGMASRLLGPSFESYLLPELTRYDCEVNVFVLSSTLLQGGDLLRALDQAT
Mm_HIF-2a	824	SGVASRLLGPSFEPYLLPELTRYDCEVNVFVPGSSTLLQGRDLLRALDQAT

**Figure 1.6 Conservation of HIF- $\alpha$  CAD residues in HIF-1 $\alpha$  and HIF-2 $\alpha$ .** The target asparaginyl residue hydroxylated by FIH is shown in red. Strongly or partially conserved residues are indicated in turquoise or grey, respectively. Amino acid numbers are shown to the left of each sequence. Sequences aligned using Clustal Omega. Hs = *Homo sapiens*, Mm = *Mus musculus*.

used by another research group to identify novel FIH substrates (Cockman et al., 2006). In the publication detailing this research, it was reported that a number of ankyrin repeat domain (ARD)-containing proteins were targeted by FIH. Indeed, subsequent to publication of these findings, a large number of ARD-containing proteins have been reported to be hydroxylated by FIH (see Table 1.1). A brief summary of these findings shall be presented in the following section. However, for convenience, a more in depth discussion of ARD-containing proteins as substrates of FIH shall be included in the results sections of this thesis, as ARD-containing proteins were also identified in the research presented herein.

## 1.6 ARD-containing proteins as substrates of FIH

The first ARD-containing proteins to be reported as novel substrates of FIH were the NF $\kappa$ B signalling pathway-associated factors, p105 and I $\kappa$ B $\alpha$ . For both proteins, the asparaginyl residues targeted (see Table 1.1) were determined by MS/MS analysis to be located within the long loops connecting adjacent ankyrin repeats (Figure 3.2B(i) and (ii)). Since this initial finding, a substantial number of other ARD-containing proteins have also been verified by similar methods as FIH substrates (Table 1.1). Interestingly, despite this influx of novel substrates, a report of a clear, functionally important effect for any one of these modifications is lacking, with most functional assays demonstrating no effect (e.g. MYPT1) (Webb et al., 2009), conflicting results (I $\kappa$ B $\alpha$ ) (Cockman et al., 2006), or lacking sufficient follow-up to verify interesting *in vitro* results in an *in vivo* setting (Ankyrin R, ASPP2, TRPV3) (Janke et al., 2013; Karttunen et al., 2014; Yang et al., 2011b). Additionally, studies examining the interaction between FIH and its substrate, Notch1, suggest that FIH's ability to repress Notch1 signalling in myoblast cells is independent of its catalytic activity, thus adding further complexity to the proposed roles of FIH regarding ARD protein modulation (Zheng et al., 2008). In



Substrate	Identification method	Hydroxylated Asns (aa no.)	Source of protein for MS/MS	Ref
p105	Y2H, FIH bait	678	Over-expressed p105 in 293T cells	(Cockman et al., 2006)
IκBα	Literature-based prediction	210, 244	Endogenous protein, various cell lines	(Cockman et al., 2006)
Notch1	Literature-based prediction	1945, 2012 (mouse) 2023 (human)	Bacterially expressed ARD (mouse); Endogenous protein, 293T cells (human)	(Coleman et al., 2007; Zheng et al., 2008)
Asb4	Y2H, Asb4 bait	246	Over-expressed mouse protein in COS7 cells	(Ferguson et al., 2007)
Rabankyrin-5	FIH IP, DMOG stimulated/SILAC	316, 485, 752, 797	Over-expressed human protein in 293T cells	(Cockman et al., 2009)
RNase L	FIH IP, DMOG stimulated	196	Over-expressed human protein in 293T cells	(Cockman et al., 2009)
Tankyrase-2	SILAC	586, 706, 739	Over-expressed human protein in 293T cells	(Cockman et al., 2009)
MYPT1	Bioinformatic prediction	67, 100, 226	Turkey gizzard	(Webb et al., 2009)
Notch2	Bioinformatic prediction	1902, 1969	Bacterially expressed mouse ARD	(Wilkins et al., 2009)
Notch3	Bioinformatic prediction	1867, 1934	Bacterially expressed mouse ARD	(Wilkins et al., 2009)
Ankyrin R	Bioinformatic prediction	105, 138, 233, 728	Human blood	(Yang et al., 2011b)
Ankyrin B	Bioinformatic prediction	58, 124, 656	Bacterially expressed human ARD	(Yang et al., 2011b)
TRPV3	Bioinformatic prediction	242	Over-expressed human protein in 293T cells	(Karttunen et al., 2014)
ASPP2	Bioinformatic prediction	986	Over-expressed human protein in 293T cells	(Janke et al., 2013)

**Table 1.1** List of MS-MS-verified ARD-containing substrates of FIH.

summary, whilst hydroxylation of ARD-containing proteins is clearly a common task for FIH, determining the importance of these modifications, and how they shape the overall role of FIH in cells is an ongoing area of investigation.

Indeed, during the latter stages of this project, the lack of understanding of FIH's functions was made particularly apparent following generation of an FIH knockout mouse, the phenotype of which has stimulated further interest in FIH's substrate repertoire (Zhang et al., 2010).

## 1.7 The FIH knockout mouse

Although findings from the generation of the FIH knockout (KO) mouse were not made available until well after the commencement of this project, a brief mention of the KO mouse phenotype will be made here for 2 reasons: (1) FIH-deficient mouse embryonic fibroblasts (MEFs) derived from the KO mouse proved a useful biochemical tool in a number of experiments, and (2) the phenotype itself could not easily be explained by regulation of known substrates of FIH, thus reaffirming the goals of this work.

Considering the role of PHDs 1-3 and FIH in negative regulation of HIF, knockout (KO) mouse models for the hydroxylases were largely anticipated to reflect excessive normoxic HIF signalling. In the case of the PHDs, this has generally proven to be true. Phenotypes developed following global or tissue-specific PHD deletion vary widely. However, many of the effects observed (e.g. increased serum EPO levels (PHD2<sup>-/-</sup>), resistance to ischaemia in heart (PHD2<sup>-/-</sup>, PHD1<sup>-/-</sup>) or skeletal muscle (PHD1<sup>-/-</sup>), and altered neutrophil and sympathoadrenal neuron survival (PHD3<sup>-/-</sup>) are likely to be (at least partially) mediated by dysregulation of HIF1 or HIF2 (Adluri et al., 2011; Aragonés et al., 2008; Bishop et al., 2008; Holscher et al., 2011; Takeda et al., 2008; Walmsley et al., 2011). In contrast, parallels between FIH deletion and canonical symptoms of HIF derepression have been somewhat harder to find.

### 1.7.1 FIH knockout mice are hypermetabolic, and have increased insulin sensitivity

Unlike PHD KO mice, global FIH deletion did not appear to affect tissue vascularisation, erythropoiesis, or the contribution of glycolysis to energy production. Rather, the predominating phenotype was one of hypermetabolism. The transgenic mice had a 15% increase in heart rate, were observed to consume more food and water, were less active and generated 20% more body heat when compared to their wild-type counterparts. In addition, the mice consumed 20% more oxygen, with a similar increase seen in carbon dioxide production (Zhang et al., 2010). Closer inspection of

the KO mice's breathing dynamics revealed this increased consumption of oxygen to be achieved through an increase in tidal volume (volume of air inhaled per breath). Reasoning that this altered breathing may be an attempt to compensate for hypermetabolism-induced acidosis, Zhang et al measured blood pH levels in the mice. Interestingly, the pH was raised rather than lowered (or unchanged), suggesting that under normoxic atmospheric conditions, FIH KO mice are programmed to breathe more deeply than their WT counterparts (Zhang et al., 2010). However, the molecular mechanisms which promote this increased normoxic tidal volume in FIH KO mice are yet to be elucidated.

Further studies of the KO mice demonstrated irregularities in both glucose and fat homeostasis. In particular, the mice had greater insulin sensitivity, and were more resistant to weight gain when placed on a high fat diet. These alterations were correlated with a slight reduction in liver mRNA expression of genes involved in lipogenesis, as well as a significant elevation in white adipose tissue (WAT) of adiponectin (an adipokine with roles in insulin sensitisation and suppression of inflammation and apoptosis) and one of its major regulators, peroxisome proliferator-activated receptor  $\gamma$  (PPAR $\gamma$ ) (Turer and Scherer, 2012; Zhang et al., 2010). Additionally, of great interest was the discovery that a neuronal-specific KO of FIH replicates the altered oxygen consumption and insulin sensitivity of global FIH KO mice, thus implicating a previously unappreciated role for FIH in neuron-like cells (Zhang et al., 2010).

Whether or not FIH's effects on glucose/lipid homeostasis or ventilation are direct, the challenge remains to determine the molecular details of FIH's cellular role(s). Even though the FIH KO mice do not display "classical" symptoms of HIF up-regulation, a potential role for the HIFs in mediating the FIH KO phenotype warrants careful consideration. Importantly, in some cases, elevated HIF signalling produces effects which are at odds with those reported in the FIH KO mice, a subject which will be discussed further in section 7.3.1. On the whole, the generation of the FIH KO mouse has illustrated the complexity of FIH's biological role, and most importantly for this work, reinforces the idea that FIH's substrate repertoire may extend beyond the HIFs.

## 1.8 Summary

The ability of oxygen-sensitive species to maintain effective oxygen homeostasis is of critical importance, as mere minutes without this important molecule can inhibit ETC function, resulting in cell death. Hypoxia adaptation mechanisms are many and varied, although remarkably, a majority of

these processes are yet to be definitively linked to a genuine oxygen sensing molecule. One exception is the group of transcription factors responsible for a large proportion of hypoxic target gene regulation, the HIFs. Both HIF-1 $\alpha$  and -2 $\alpha$  are subjected to hydroxylation by the oxygen- and 2-OG-dependent oxygenases, PHDs 1-3 and FIH, a modification the presence or absence of which acts as a switch for induction of HIF activity. Given the oxygen sensing properties of these enzymes, a key question has concerned the possible role of these enzymes in other oxygen-sensitive processes beyond regulation of the HIFs. Indeed, characterisation of FIH's substrate binding methods suggest that the enzyme can bind diverse substrate sequences, thus adding weight to the idea that multiple FIH substrates exist. **Thus, the aim of thesis is to isolate and characterise novel substrates of FIH, with a view to better understanding the biological roles of this enzyme, and its participation in cellular oxygen sensing.** The finer details of these aims and how they will be tackled experimentally are described below.

## 1.9 Aims and approach

At the commencement of this project, the only known substrates of the FIH asparaginyl hydroxylase were HIF-1 $\alpha$  and HIF-2 $\alpha$ . Studies of FIH's substrate binding mechanisms, however, have suggested that FIH may have the flexibility to engage a number of different substrates. As a result, it was hypothesised that FIH may have a larger role in oxygen sensing than currently appreciated, and as such, isolation of novel FIH substrates would be a highly worthwhile pursuit. Thus, the following experimental aims are proposed.

### **Aims:**

- 1) To isolate and characterise novel substrates and interacting proteins of the asparaginyl hydroxylase, FIH**
- 2) To determine whether hydroxylation or binding by FIH changes the functional output of the substrate/interacting factor**
- 3) To further define the sequence requirements used by FIH to bind and hydroxylate substrates, thus improving bioinformatic substrate prediction.**

The means by which each aim is intended to be addressed is explained briefly below.

### Aim 1: Novel substrate discovery and characterisation

Approach at a glance:

- a) Yeast 2-hybrid (Y2H) screen with FIH as bait
- b) Literature-based prediction of possible FIH substrates
- c) Substrate candidacy assessed via *in vitro* hydroxylation assays
- d) Binding to FIH assessed via co-immunoprecipitation (co-IP) and ability to compete out HIF measured via reporter assay
- e) Compare the structural features of substrates via sequence alignment to define a consensus sequence for FIH

Rationale:

- a) Detailed analysis of the means by which FIH recognises substrate peptides suggests that very few of the substrate residues which occupy the active site are likely to be indispensable for substrate recognition. As a consequence, bioinformatic approaches are currently of limited use for identification of novel candidate substrates. Instead, an unbiased screen using FIH as bait was chosen, as this method has the added benefit of isolating non-substrate interacting factors of FIH which may also play a role in modulating FIH activity, and vice versa.
- b) Although at times a “hit and miss” exercise, the benefits of function-based prediction of novel substrates cannot be overlooked. Proteins which may be linked to FIH function and are also associated with sensitivity to oxygen levels are of particular interest.
- c) The *in vitro* hydroxylation assay is still the simplest and easiest method by which proteins can be assessed as potential substrates, provided that the candidate protein can be expressed in and purified from bacteria. The exact residue hydroxylated can also be inferred through point mutations of suspected target residues. As the *in vitro* assay is an indirect measure of hydroxylation, further verification that the activity observed is truly due to hydroxylation can be obtained through mass spectrometric analysis of the substrate post incubation with FIH.
- d) Co-IP was chosen to assess the binding capabilities of any proteins identified in the Y2H screen as it has the added benefit of ensuring that candidate interactors have been synthesised, folded and post-translationally modified in the environment of a mammalian cell, thus maximising the chances of detecting FIH:binding partner interactions which are sensitive to these parameters. In addition, it is desirable to verify that the interaction can occur in the context of a living cell. In this context, it has been found that the transcriptional readout from the HIF-1 $\alpha$  CAD is a useful measure of FIH’s availability in the cell, thus if over-

expression of a candidate substrate/binding protein alters the CAD's transactivation activity (as assessed by reporter assay), it can be inferred that the two engage in a genuine binding interaction.

- e) If additional substrates are isolated by the Y2H or function-based prediction, comparison of the targeted sequences via simple alignment may provide additional information regarding the structural features used by FIH for substrate targeting. Ultimately, if a consensus sequence can be built it may be a useful bioinformatic tool for further substrate prediction.

### Aim 2: Functional consequences of hydroxylation by/ interaction with FIH

Naturally the exact nature of these experiments will depend on the identity of the substrate/interacting partner involved. However, a number of options are available for testing the ability of FIH to change a substrate/binding protein's function, and vice versa. The simplest of these include (1) over-expression or knockdown of FIH (or use of cells which lack FIH) and analysis of the effect of these changes on substrate protein activity, (2) essentially the reciprocal version of experiment (1), wherein overexpression of the substrate/binding factor is manipulated to assess its effect on FIH function, and (3) mutagenesis of the target asparagine in a substrate protein, thus making it refractive to modification by FIH, which may alter its response to modulation of FIH levels, or application of related stimuli such as hypoxia.

### Aim 3: Define the structural requirements for hydroxylation

Approach at a glance:

Compare the ability of human and *Tribolium castaneum* FIH orthologs to hydroxylate HIF- $\alpha$  CAD and ARD-containing substrates, with a view to better defining structural features of FIH which may influence hydroxylation of these different substrates.

Rationale:

The similar structures, but considerably different hydroxylation properties displayed by the human and beetle FIH enzymes towards human HIF-1 $\alpha$  (see sections 4.4 and 4.5 for details) may provide a unique opportunity to determine which amino acids of the substrate binding pocket are critical for substrate binding and modification. This, in turn, may provide insight into which features of a substrate protein may render it more likely to be a genuine target of FIH hydroxylation. In order to compare the properties of the two enzymes, residues in beetle FIH which are predicted to hamper its ability to hydroxylate human HIF-1 $\alpha$  shall be "transferred" to the structure of human FIH via site-directed mutagenesis. The mutant human FIH shall then be tested by *in vitro* hydroxylation assay to

determine if the mutation is, in fact, inhibitory. It is hoped that this analysis will provide an additional method to define critical features of FIH substrate selection, thus improving the chances of successfully predicting novel substrates in the future.

Collectively, it is predicted that experimental engagement of these aims will significantly improve understanding of FIH's substrate recognition properties, and the functional consequences of FIH-mediated hydroxylation in cells.





# 2 Materials and Methods



## 2.1 Abbreviations

2-OG	2-oxoglutarate
4E-T	eIF4E transporter
aa	amino acids
ADP	adenosine diphosphate
ALS-2	5-aminolevulinate synthase 2
APS	ammonium persulphate
ARD	ankyrin repeat domain
ARE	AU-rich element
ARNT	aryl hydrocarbon receptor nuclear translocator
ASB4	ankyrin repeat and SOCS box protein 4
ASPP2	apoptosis-stimulating p53-binding protein
ATP	adenosine triphosphate
$\beta_2$ AR	beta-2 adrenergic receptor
bHLH-PAS	basic helix-loop-helix Per ARNT Sim
BSA	bovine serum albumin
CA9	carbonic anhydrase 9
CAD	C-terminal transactivation domain
cDNA	complementary DNA
Co-IP	co-immunoprecipitation
COX	cytochrome c oxidase
CP	Chuvash Polycythemia
CV	column volumes
Da	Daltons
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
ddH <sub>2</sub> O	deionised water
DFO	desferrioxamine
dGTP	deoxyguanosine triphosphate
DMOG	dimethyloxalylglycine
DTT	dithiothreitol
dTTP	deoxythymidine triphosphate
EDTA	ethylenediaminetetraacetic acid
eIF4E	eukaryotic initiation factor 4E

eIF4F	eukaryotic initiation factor 4F
Epo	erythropoietin
ETC	electron transport chain
EYFP	enhanced yellow fluorescent protein
FGIF	Fetal globin-inducing factor
FIH	Factor Inhibiting HIF
GRK2	G protein-coupled receptor kinase 2
HEK 293T	human embryonic kidney 293T
HIF	Hypoxia-inducible Factor
HO-1	haem oxygenase-1
HPAC	Human pancreatic adenocarcinoma
IDA	iminodiacetic acid
I $\kappa$ B $\alpha$	NF-kappa-B inhibitor alpha
IKK $\beta$	I $\kappa$ B kinase beta
IMAC	immobilised metal affinity chromatography
IPTG	isopropyl-beta-D-thiogalactopyranoside
JNK1	c-Jun N-terminal kinase 1
KO	knockout
kV	kilovolts
LB	luria broth
LDH-A	lactate dehydrogenase A chain
MBP	maltose binding protein
MCPH	autosomal recessive primary microcephaly
MEF	murine embryonic fibroblast
mRNA	messenger RNA
MS	mass spectrometry
MW	molecular weight
MYBBP1A	Myb binding protein (p160) 1a
MYPT1	Myosin phosphatase-targeting subunit 1
NF $\kappa$ B	Nuclear factor kappa-light-chain enhancer of B cells
NMR	nuclear magnetic resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
N-OG	N-oxalylglycine
O/N	overnight
ODD	Oxygen-dependent degradation domain

p105	Nuclear factor NF-kappa-B p105 subunit
PB	Processing body
PGS	protein G sepharose
PHD	prolyl hydroxylase domain-containing protein
PMSF	Phenylmethanesulfonyl Fluoride
PP1R12C	Protein phosphatase 1 regulatory subunit 12C
PP6-ARS-B	Serine/threonine-protein phosphatase 6 regulatory ankyrin repeat subunit B
PPAR $\gamma$	Peroxisome proliferator-activated receptor gamma
qPCR	quantitative real-time PCR
RAM	Rbp-associated molecule
RBM4	RNA binding motif protein 4
Rbp1	RNA polymerase II subunit B1
RNASEL	Ribonuclease L
ROS	Reactive oxygen species
rpm	revolutions per minute
RT	room temperature
S/N	supernatant
SD	Standard deviation
SDS PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEM	Standard error of the mean
SET	Suppressor of variegation 3-9, Enhancer of zeste, Trithorax
shRNA	short hairpin RNA
SOC	Super Optimal broth with Catabolite repression
T2DM	Type 2 Diabetes Mellitus
TCPTP	T cell protein tyrosine phosphatase
TEMED	tetramethylethylenediamine
TEV	tobacco etch virus
TRPV3	Transient receptor potential cation channel, subfamily V, member 3
VEGF	Vascular endothelial growth factor
VHL	von Hippel Lindau
WAT	white adipose tissue
WDR62	WD repeat domain 62
Y2H	Yeast 2-hybrid
YPED	Yeast Extract Peptone Dextrose
YT	Yeast extract and Triptone

## 2.2 Primers

All primer sequences are depicted 5'-3'.

### 2.2.1 Cloning

#### mFERRLCfor

TTGTCGACCATGACCTCTCAGATTCG

Primer amplifies from the start codon. Used in conjunction with mFERRLCrev to amplify full-length ferritin.

#### mFERRLCrev

TTGCGCCGCTAGTCGTGCTTGAG

Primer amplifies from the stop codon. Used in conjunction with mFERRLCfor to amplify full-length ferritin.

#### sal-linkLO

6GA TCC TGC AGG TCG ACC TG (6 denotes 5' phosphate)

Generates BamHI ends when annealed with sal-linkUP. Inserted into BamHI-digested pET32a.

#### sal-linkUP

6GA TCC AGG TCG ACC TGC AG

Generates BamHI ends when annealed with sal-linkUP. Inserted into BamHI-digested pET32a.

#### 4E-T867-983for

CACCGTGTGCACTCCTGGGTCAACC

Primer amplifies from aa 867. Used in conjunction with 4E-T867-983rev to amplify c-terminus of 4E-T.

#### 4E-T867-983rev

AAG CGG CCG CTC TGC CCT CTT CAC TG

Primer amplifies from stop codon. Used in conjunction with 4E-T867-983for to amplify c-terminus of 4E-T.

#### 4E-T(1-240)for

C ACC GTG TCG ACC CAA GAA GCC ATG GAG

Primer amplifies from just before start codon. Used in conjunction with 4E-T240rev to amplify first 240 aa of 4E-T.

4E-T240rev

AAG CGG CCG CTA CTT ATC ATC AAA GCC AG

Used in conjunction with 4E-T(1-240)for to amplify first 240 aa of 4E-T.

4E-T491rev

AAG CGG CCG CTA CTT CAT GGT GTT CAC

Used in conjunction with 4E-T(241-729)for to amplify aa 241-491 of 4E-T.

4E-T(241-729)for

C ACC GTG TCG ACG ATT CTA GAG GAG GAT C

Used in conjunction with 4E-T729rev to amplify aa 241-730 of 4E-T.

4E-T729rev

AAG CGG CCG CTA CTG AGT ATC CTC TTT AC

Used in conjunction with 4E-T(241-729)for to amplify aa 241-730 of 4E-T.

4E-T(731-983)for

C ACC TTG TCG ACG AAG ACC AGC GAA GAG

Used in conjunction with 4E-TUTRrev to amplify aa 731-983 of 4E-T.

4E-TUTRrev

TTG CGG CCG CAC AGG TCC AGG TAT GG

Used in conjunction with 4E-T(731-983)for to amplify aa 731-983 of 4E-T.

FGIFfor2

CGG TCG ACA AGA GTA GTA AAA ATG G

Primer amplifies from a little before the start codon of mouse FGIF. Used in conjunction with FGIFrev2 to amplify full-length FGIF from mouse 10.5 cDNA.

FGIFrev2

TTG CGG CCG CTG TAT TAC CTG TGT CC

Primer amplifies from approximately 250 bp into the 3'UTR. Product includes stop codon. Will be used in conjunction with FGIFfor2 to amplify full-length FGIF.

FIHNcolfor

TTCCATGGCGGCGACAG

Used in conjunction with NTKGVEinsREV to generate the upstream product later used in overlap extension PCR to insert the sequence NTKGVE from aa 96-101 of tcFIH after aa 113 of hFIH.

FIHXholrev

TTCTCGAGCTAGTTGTATC

Used in conjunction with NTKGVEinsFOR to generate the downstream product later used in overlap extension PCR to insert the sequence NTKGVE from aa 96-101 of tcFIH after aa 113 of hFIH.

gankN67Afor

CTTGGAGTGCCAGTGGCTGATAAAGACGATGCAG

Used in conjunction with gankN67Arev to mutate Asn67 to Ala via Quikchange.

gankN67Arev

CTGCATCGTCTTTATCAGCCACTGGCACTCCAAG

Used in conjunction with gankN67Afor to mutate Asn67 to Ala via Quikchange.

gankN100Afor

GGAAAAGGTGCTCAAGTGGCTGCTGTCAATCAAATG

Used in conjunction with gankN100Arev to mutate Asn100 to Ala via Quikchange.

gankN100Arev

CATTTTGATTGACAGCAGCCACTTGAGCACCTTTTCC

Used in conjunction with gankN100Afor to mutate Asn100 to Ala via Quikchange.

gankN166Afor

CTACAAAGCATCCACAGCCATCCAAGACTGAGG

Used in conjunction with gankN166Arev to mutate Asn166 to Ala via Quikchange.



gankN166Arev

CCTCAGTGTCTTGGATGGCTGTGGATGCTTTGTAG

Used in conjunction with gankN166Afor to mutate Asn166 to Ala via Quikchange.

GLPN869Afor

CAAGGGCTCTGACATCGCCATCCGAGACAACGAG

Used in conjunction with GLPN869Arev to mutate GLP Asn 900 to Ala.

GLPN869Arev

CTCGTTGTCTCGGATGGCGATGTCAGAGCCCTTG

Used in conjunction with GLPN869Afor to mutate GLP Asn 900 to Ala.

hCADshortf

TGCTCCTATACAAGGCAACCTACTGCAGGGTG

Used in conjunction with hCADshortr to delete aa 809 and 810 from the CAD of hHIF-1 $\alpha$ .

hCADshortr

CACCCTGCAGTAGGTTGCCTTGTATAGGAGCA

Used in conjunction with hCADshortf to delete aa 809 and 810 from the CAD of hHIF-1 $\alpha$ .

hGLPARD739for

C ACC CAC CAG AAT AAG CGC TCT CC

Primer amplifies from aa 739 of human GLP. Used in conjunction with hGLPARD979rev to amplify the ARD of GLP.

hGLPARD979rev

T TGC GGC CGC TTA CAC TAT CCT CTC CAC

Primer amplifies from aa 979 and contains stop codon. Used in conjunction with hGLPARD739for to amplify the ARD of GLP.

hG9aARD682for

C ACC CAG CAG AGC AAG CGC ACG

Primer amplifies from aa 682 of human G9a. Used in conjunction with hG9aARD913rev or hG9a-ARD1-4Rv to amplify the ARD or ankyrin repeats 1-4 in G9a.

hG9a-ARD1-4Rv (designed by Max Tollenaere)

GC GGC CGC CTA GAT GTT CTC CTC GTT

Primer amplifies from aa 820 of human G9a. Used in conjunction with hG9aARD682for to amplify ankyrin repeats 1-4 in G9a.

hG9aARD913rev

T TGC GGC CGC TTA ATT TCC CAC CCC AAG

Primer amplifies from aa 913 of human G9a. Used in conjunction with hG9aARD682for to amplify the ARD of G9a.

hG9a-preARD446Fw (designed by Max Tollenaere)

CACCGAGTCAGAGAGGCGGAAG

Primer amplifies from aa 446 of human G9a. Used in conjunction with hG9aARD913rev to amplify the pre ARD and ARD of G9a.

NTKGVEinsFOR

AACACCAAGGGGGTCGAGTTTAAGCCGAGGTCC

Used in conjunction with FIHXholrev to generate the upstream product later used in overlap extension PCR to insert the sequence NTKGVE from aa 96-101 of tcFIH after aa 113 of hFIH.

NTKGVEinsREV

CTCGACCCCTTGGTGTGTTCTGGAAATTGGCC

Used in conjunction with FIHNcolfor to generate the downstream product later used in overlap extension PCR to insert the sequence NTKGVE from aa 96-101 of tcFIH after aa 113 of hFIH.

## 2.2.2 qPCR

<b>Gene</b>	<b>Forward</b>	<b>Reverse</b>
m4930486L24Rik	GGAGAAGAATCAGATGGCAACAG	TGTATCCCTTCATGCCCCATT
mAass	CTCGGCAATGGCTAAAACACTGT	AGGCCTTTGGCTTCAATTCA

mAdh4	GGTGACCCATGCCCTACCTT	TGCTCCTGGCATTTCAGAA
mAkr1c12	CCCTGAGGACATGAAAACACTAGA	GGATACTCTGGGTGGTCAGCAA
mAkr1c18	GATGGCCTGGACAGAAATCTTC	GGCCTCCATCTTAATATTCATCGA
mAkr1cl	GAAGCGGATCAAGGAGAACCT	CGGTAGTTTCTGTTCAGCCTATCTATAA
mApom	TGCTTGGACTTCAAAGCCTTCT	TCTAGGACACGGGTGAAGTCACT
mAsz1	GGGCTTCTCCCAGAAATTTACTT	GGGTCATTTTCCCGCTCATT
mAU040829	CCCCATGTTAGACGGCATGT	TGACCCCAGCTCTATGTGGAT
mBC099439	CCTGCTGAGGACAACCTGAGA	GTGGTCTTCTGTACCGGATCAT
mBdh2	CAAAAGACGGGAAGGTTTGC	GACAGGGTTGCCAGTTACATAGG
mC80008	GGACAGCTCAGAGGACTTGGGA	CAATAGGTGTGGCTGAGACATGA
mCcdc114	GAGTTGCTAAGCCAAGTCATGAAG	TGTGGACTGGAGAGAGTCAAGCT
mCd3d	GCTGAAGAATGAGCAGCTGTATCA	CACTTTGAGAAACCTCCATCTTAAGA
mCd3e	CCTGTTCCCAACCCAGACTATG	GCTCTCTGATTGAGCCAGAA
mCd3g	TCCAAGGAAACCAACTGAGGAA	TGATTCTGGGTGCTGGATAGAA
mCox7b2	AGAACTTCTGAGGAGAGCGTCTGT	GGCCAAGGGAAACATCATGA
mDefb42	TTCATCCTGGTGACATCATTGC	CGGTGCACGTCTCAAAGGATA
mEdg2	GGAATCGGGACACCATGATG	CCGGAGTCCAGCAGACAATAA
mEfhc2	TGAACTGGAGAATGAATCCAACAC	TCCCACCCAAACGTCTTCAC
mEG638695	GCACTCACCTGTGTCTCCTACAAG	CTGCAGAAGCTCAAACAGATCAG
mGlrx	ACTCGGCTGAAGCAGATTGG	TGGTTAGACAGCTGTGTCAGCAT
mGm773	GTCCACAGCCTGAAAGCACTTA	CACTGCTATTGTCTTTGTCCAAGTTC
mHerc3	CACGAGTTCCCATTTGGAGAAG	CCATGCCATAGATGGGAATCC
mIgbp1b	AACCGACAGAACATGGGCTAA	TCCTTGCTGGGAAGGTGAGA
mKlb	CTAGAGGATGATCAGATCCGAAAAGT	AGCCTTTGATTTTGACCTTGTC
mKrt23	AAACCCACCTGGAGAAGGAAA	TTGACTCAGATCCATCCATCGT
mLOC545238	CTCCGGACAACGGATTAAGG	CCGACATCCGTATTCGACACT
mMagea4	ACGCAGAGGCTTTGAGAGATG	GAGCTGTAGAACTATCTGCTGGATGA
mMagea6	AGGGAGGATCTCAGGAGTGTCA	AACAGTAGGCAGGATAGTATTCCAATC
mMdfic	AAATGCTGTCTTTTCCCCAAAA	AGAACAAGCATGCCAGGATACA
mNckap1l	CGTGAATCCATTTCTCTGCTCAT	CAGCATGTCCAGGGTCAAGA
mOTTMUSG00000025408	CCTGCTGAGGACAACCTGAGA	GTGGTCTTCTGTACCGGATCAT
mPde4dip	GCATACCCCGGTCCACATTA	CTTGGCGTTGCTCAGTTCTTT
mPkd2l1	TGAAGATGCTGGAGAGGAAAGG	GGTTCTCCAAACAGCTGGTT
mSftpd	GTTGCCTTCTCCACTATCAGAA	GTCCTGAAGATCTTGTCTCCAACA
mSgcb	AACGAAGGCGTCTTCATCATG	GTGCCGTTGAGGATGATGCT

mSlc5a4b	AGAAGAAGCTGGTGGTGCTCTT	AGCAGCATGCCCTCTTGAGA
mSpic	GCTGAAAGCCAGCTGGTACAA	TCAAACAGCCGAAGCTTTCTC
mSpink10	CCCAGTCTGTGGGACTAATGGA	GACCAAAATGCCGATAATCAATTT
mTph2	CCTTTCAGGACGCTTACTTTGTTT	CCGAGAAGGGACGGGTAATT
mUbe2dn1	CCTAAAGCGACAGGAGGGAAT	CACTGAGTATGGCAGCTGGTAGA
16S	GTGGGGAGCAAAYAGGATTAGA	GGCATGATGATTTGACGTCRT
mHMBS	GTAGCAGTGCATACAGTGATGAAAGA	GCATGCTATCTGAGCCATCTAGAC
mPPIA	CTGCACTGCCAAGACTGAATG	CACAATGTTCATGCCTTCTTTCAC
mTBP	GCTTCTGAGAGCTCTGGAATTGTAC	GTGCAATGGTCTTTAGGTCAAGTTT
UBC2	AGGAGGCTGATGAAGGAGCTTGA	TGGTTTGAATGGATACTCTGCTGGA
YWHAZ	CGTTGTAGGAGCCCGTAGGTCAT	TCTGGTTGCGAAGCATTGGG

## 2.3 Plasmids

### 2.3.1 Plasmids described elsewhere

#### pEF-6myc-AdipoR1OOF (pEF-6myc-PP9)

Provided by Dan Peet. The construct encodes a 6myc tag fused to out of frame Adiponectin receptor

1. The amino acid sequence is as follows:

MVHGGSSSEQKLISEEDLNGEQKLISEEDLNGEQKLISEEDLNGSSRGEQKLISEEDLNGEQKLISEEDLNGEQKL  
 ISEEDLNGSSSNMPSTHADRAPLGRPPPPVPRDVAPGPAAPSRXSQPGLSTITGRSLPSGAERPQFFKLLKLP  
 TEVPARCLPTKALPGHKAMGLLLVTEKLTQWSWLSWGPCWRRRASGQPAAQPRLRKKIKHARCLRKRRRRRCGC .

#### pEF-6myc-mRPS2 (pEF-6myc-PP13)

Provided by Dan Peet. The construct encodes a 6myc tag fused to murine ribosomal protein S2.

#### pEF-GalDBD

Constructed by Dan Peet (Lando et al., 2002b).

#### pEF-GalDBD-gateway

Constructed by Alex Colella (Richards Laboratory, School of Molecular and Biomedical Science, Adelaide University) from pEF-GalDBD (Lando et al., 2002b).

#### pEF-myc-6His-FIH

Constructed by Karolina Lisy (Lisy, 2011).

pE-GFP-N3-WDR62

Previously described in (Bogoyevitch et al., 2012).

pENTR11

Product of Invitrogen (<http://www.lifetechnologies.com>).

pENTR11+

Provided by Sarah Karttunen.

pENTR-D/TOPO

Product of Invitrogen (<http://www.lifetechnologies.com>).

pET32a

Product of Novagen (<http://www.merckmillipore.com>)

pET32a-mHIF-1 $\alpha$  747-836

Previously described in (Wilkins et al., 2012).

pET32a-mHIF-2 $\alpha$  774-874

Previously described in (Lando et al., 2002a).

pET32a-mNotch1 ARD 1862-2104

Previously described in (Zheng et al., 2008).

pET32a-RAM 1753-1874

Previously described in (Linke et al., 2007).

pET32a-TEV-FIH

Previously described in (Linke et al., 2007).

pGEM-T Easy

Product of Promega ([www.promega.com](http://www.promega.com))

pGRE-luciferase

Previously described in (Pongratz et al., 1998).

#### pMBP-FIH

Previously described in (Lando et al., 2002a).

#### pRL-TK

Product of Promega ([www.promega.com](http://www.promega.com))

#### pTRDC-EYFP

Provided by Anthony Fedele (Peet Laboratory, School of Molecular and Biomedical Science, Adelaide University).

#### pEF-puro-TETON

Provided by Anthony Fedele (Peet Laboratory, School of Molecular and Biomedical Science, Adelaide University).

### 2.3.2 Plasmids cloned in this work

#### pEF-6myc-BOS destination vector

pBS(KS)6myc was digested with Kpn/HindIII, the 6myc fragment purified and then ligated into similarly digested pBS-RfcB-For. The resulting vector, pBS-6myc-RfcB-For, was then digested with KpnI/SpeI, the 6myc-RfcB-For fragment purified and then ligated into similar digested pEF-BOS-CS.

#### pEF-6myc-4E-T

pPC86-4E-T was digested with Sall/NotI, the 4E-T fragment purified and then ligated into similarly digested pEF-6myc-#24.

#### pEF-6myc-4E-T (1-240)

4E-T (1-240) was transferred to the pEF-6myc-BOS destination vector via Gateway LR reaction with pENTR-4E-T (1-240).

#### pEF-6myc-4E-T (241-491)

4E-T (241-491) was transferred to the pEF-6myc-BOS destination vector via Gateway LR reaction with pENTR-4E-T (241-491).

pEF-6myc-4E-T (241-730)

4E-T (241-730) was transferred to the pEF-6myc-BOS destination vector via Gateway LR reaction with pENTR-4E-T (241-730).

pEF-6myc-4E-T (731-983)

4E-T (731-983) was transferred to the pEF-6myc-BOS destination vector via Gateway LR reaction with pENTR-4E-T (731-983).

pEF-6myc--4E-T (867-983)

pENTR-4E-T (867-983) was digested with Sall/NotI, the 4E-T (867-983) fragment purified, and ligated into similarly digested pEF-6myc-#24.

pEF-6myc-hHIF-1 $\alpha$  (727-826)

pPC86-HIF-1 (727-826) was digested with Sall/NotI enzymes, the HIF1 fragment purified and ligated into similarly digested pEF-6myc-#24.

pEF-6myc-ferritin

pGEM-T Easy-ferritin was digested with Sall/NotI, the ferritin fragment purified and then ligated into similarly digested pEF-6myc-#24.

pEF-6myc-mFGIF

pGEM-T Easy-mFGIF was digested with Sall/NotI, the mFGIF fragment purified and then ligated into similarly digested pEF-6myc-#24.

pEF-6myc-mFGIF(153-238)

pGEM-T Easy-mFGIF(153-238) was digested with Sall/NotI, the mFGIF(153-238) fragment purified and then ligated into similarly digested pEF-6myc-#24.

pEF-6myc-I $\kappa$ B $\alpha$  (136-314)

pPC86-I $\kappa$ B $\alpha$  (136-314) was digested with Sall/NotI, the I $\kappa$ B $\alpha$  (136-314) fragment purified and then ligated into similarly digested pEF-6myc-#24.

pEF-6myc-PP1R12C

pPC86-PP1R12C was digested with Sall/NotI, the PP1R12C fragment purified and then ligated into similarly digested pEF-6myc-#24.

pEF-6myc-PP6-ARS-B ((341-486))

pPC86-PP6-ARS-B (341-486) was digested with Sall/NotI, the PP6-ARS-B (341-486) fragment purified and then ligated into similarly digested pEF-6myc-#24.

pEF-6myc-WDR62 (1131-1523)

pPC86- WDR62 (1131-1523) was digested with Sall/NotI, the WDR62 (1131-1523) fragment purified and then ligated into similarly digested pEF-6myc-#24.

pEF-Gal DBD-hHIF1 (727-826)

hHIF1 (727-826) was transferred to the pEF-Gal DBD-BOS destination vector via Gateway LR reaction with pENTR11+-hHIF1 (727-826).

pEF-Gal DBD-mFGIF

pGEM-T Easy-mFGIF was digested with Sall/NotI, the mFGIF fragment purified and then ligated into similarly digested pEF-Gal DBD-BOS.

pENTR-4E-T (1-240)

4E-T (1-240) was amplified by PCR from pPC86-4E-T using primers 4E-T(1-240)for and 4E-T240rev, then incorporated into pENTR-D-TOPO via a TOPO reaction.

pENTR-4E-T (241-491)

4E-T (241-491) was amplified by PCR from pPC86-4E-T using primers 4E-T(241-729)for and 4E-T491rev, then incorporated into pENTR-D-TOPO via a TOPO reaction.

pENTR-4E-T (241-730)

4E-T (241-730) was amplified by PCR from pPC86-4E-T using primers 4E-T(241-729)for and 4E-T729rev, then incorporated into pENTR-D-TOPO via a TOPO reaction.

pENTR-4E-T (731-983)

4E-T (731-983) was amplified by PCR from pPC86-4E-T using primers 4E-T(731-983)for and 4E-UTRrev, then incorporated into pENTR-D-TOPO via a TOPO reaction.



#### pENTR-4E-T (867-983)

4E-T (867-983) was amplified by PCR from pPC86-4E-T using primers 4E-T867-983for and 4E-T867-983rev, then incorporated into pENTR-D-TOPO via a TOPO reaction.

#### pENTR11+-hHIF-1 $\alpha$ (727-826)

pPC86-HIF-1 (727-826) was digested with Sall/NotI enzymes, the HIF1 fragment purified and ligated into similarly digested pENTR11+.

#### pET32a destination vector

pBS-RfcA-For was digested with KpnI/SacI, the RfcA-For fragment purified and ligated into similarly digested pET32a.

#### pET32a-2x\_link

pET32a was digested with BamHI and then ligated with annealed sal-linkUP and sal-linkLO oligos. The vector contains 2 copies of the oligo.

#### pET32a-link

pET32a-2x\_link was digested with Sall and then religated to remove half of the first and all of the second sal-link oligo, as well as part of the pET32a MCS. The vector contains a single Sall site.

#### pET32a-4E-T (1-355)

pPC86-4E-T was digested with NcoI/BamHI, the 4E-T (1-355) fragment purified and ligated with similarly digested pET32a.

#### pET32a-4E-T (354-757)

pPC86-4E-T was digested with BamHI/HindIII, the 4E-T (354-757) fragment purified and ligated with similarly digested pET32a.

#### pET32a-Gankyrin

For details see (Wilkins et al., 2012).

#### pET32a-Gankyrin N67A, N100A, N166A, NN67/100AA, NN100/166AA

pGEX-Gankyrin (provided by Ian Murchland, Booker Laboratory, School of Molecular and Biomedical Science, Adelaide University) was subjected to site-directed mutagenesis using the following primer pairs: N67A, gankN67Afor and gankN67Arev; N100A, gankN100Afor and gankN100Arev; N166A, gankN166Afor and gankN166Arev. To create double asparagine mutants, pGEX-Gankyrin N67A and pGEX-Gankyrin N166A were subjected to a further round of mutagenesis using gankN100Afor and gankN100Arev primers. pGEX-Gankyrin mutant clones were then digested with EcoRI and XhoI and ligated into similarly digested pET32a.

#### pET32a-S6-Gankyrin

For details see (Wilkins et al., 2012).

#### pET32a-link-4E-T (756-983)

pPC86-4E-T was digested with HindIII/NotI, the 4E-T (756-983) fragment purified and ligated with similarly digested pET32a.

#### pENTR-G9a (cloned by Max Tollenaere)

The coding sequence for the ARD region of G9a (aa 682-913) was amplified from HPAC (human pancreatic adenocarcinoma) cDNA using primers hG9aARD682for and hG9aARD913rev, and then cloned into pENTR/D-TOPO as per the manufacturer's instructions.

#### pENTR-GLP (cloned by Max Tollenaere)

The coding sequence for the ARD region of GLP (aa 770-1010) was amplified from HEK 293T cDNA using primers hGLPARD739for and hGLPARD979rev and then cloned into pENTR/D-TOPO as per the manufacturer's instructions.

#### pET32a-G9a 1-4 (cloned by Max Tollenaere)

The coding sequence for G9a (aa 446-913), which includes the "pre ARD" region, was amplified from HPAC cDNA using primers hG9a-preARD446Fw and hG9aARD913rev, then cloned into pENTR/D-TOPO as per the manufacturer's instructions. G9a (446-913) was then gateway cloned into pET32a destination vector using an LR reaction. Subsequent to this, pET32a-G9a (446-913) was used as a template to amplify G9a 1-4 (aa 682-820) using primers hG9aARD682for and hG9a-ARD1-4Rv, and the product cloned into pENTR/D-TOPO. G9a (682-820) was then gateway cloned into pET32a destination vector using an LR reaction.

#### pET32a-GLP

pENTR-GLP was combined with pET32a destination vector in an LR recombination reaction.

#### pET32a-GLP N900A

pET32a-GLP was digested with KpnI and SacI, and ligated into similarly digested pBS(KS)6myc. Quikchange mutagenesis was then performed on pBS(KS)6myc-GLP using the primers GLPN869Afor and GLPN869Arev to mutate Asn 900 to Ala. Subsequent to this, pBS-6myc-GLP N900A was digested with KpnI and SacI and ligated into similarly digested pET32a.

#### pET32a-hHIF-1 $\alpha$ - $\Delta$ SR

Quikchange mutagenesis with the primers hCADshortf and hCADshortr was used to delete aa Ser809 and Arg810 from the HIF construct in pGEM7-HIF CAD 737-826. pGEM7-HIF CAD 737-826  $\Delta$ SR was then digested with NcoI and NotI and ligated into similarly digested pET32a-TEV.

#### pET32a-TEV-hFIH-NTKGVE/TTP

Primer pairs FIHNcolfor and NTKGVEinsREV, and NTKGVEinsFOR and FIHXholrev were employed to produce first round PCR products for overlap extension PCR with template pET-TEV-hFIH dbl mut (NDQ/TTP). The second round PCR product (generated with FIHNcolfor and FIHXholrev) was then ligated into pGEM-T Easy. pGEM-T Easy-hFIH-NTKGVE/dbl mut was then digested with NcoI and XhoI and ligated into similarly digested pET32a.

#### pET32a-TEV-tcHIF- $\alpha$ (780-879)

pGEM-T Easy-tcHIF- $\alpha$  (780-879) was digested with HindIII/XhoI, the tcHIF- $\alpha$  fragment purified and ligated into similarly digested pET32a-TEV.

#### pET32a-link-ferritin

pGEM-T-easy-ferritin was digested with Sall/NotI, the ferritin fragment purified and ligated with similarly digested pET32a-2x\_link.

#### pET32a-link-PP6-ARS-B (341-486)

pPC86-PP6-ARS-B (341-486) was digested with Sall/NotI, the PP6-ARS-B (341-486) fragment purified and ligated with similarly digested pET32a-2x\_link.

#### pET32a-link-WDR62 (1131-1523)

pPC86-WDR62 (1131-1523) was digested with Sall/NotI, the WDR62 (1131-1523) fragment purified and ligated with similarly digested pET32a-2x\_link.

#### pET32a-link-mFGIF

pGEM-T-Easy-mFGIF was digested with Sall/NotI, the mFGIF fragment purified and ligated with similarly digested pET32a-2x\_link.

#### pET32a-link-mFGIF (153-238)

pPC86-mFGIF (153-238) was digested with Sall/NotI, the mFGIF fragment purified and ligated with similarly digested pET32a-2x\_link.

#### pET32a-S6-Gankyrin

For details, see (Wilkins et al., 2012).

#### pGEM-T Easy-mFerritin

Full length ferritin was amplified from mouse 10.5 dpc embryonic cDNA using mFERRLCfor and mFERRLCrev primers, A-tailed, then ligated into pGEM-T-easy.

#### pGEM-T Easy-mFGIF

Full length mFGIF was amplified from mouse 10.5 dpc embryonic cDNA using FGIFfor2 and FGIFrev2 primers, A-tailed, then ligated into pGEM-T-easy.

#### pGEM-T Easy-hFIH-BglII

hFIH was amplified from pcDNA3.1-FIH using hFIHBglIIPOfor and hFIHBglIIrev primers, A-tailed, then ligated into pGEM-T Easy.

#### pGEM-T Easy-hFIH H199A-BglII

hFIH was amplified from pcDNA3.1-FIH H199A using hFIHBglIIPOfor and hFIHBglIIrev primers, A-tailed, then ligated into pGEM-T Easy.

#### pGEM-T Easy-V5-tcFIH-BglII

V5-tagged tcFIH was amplified from pGEM-tcFIH (constructed by Briony Davenport) using tcFIHV5BglIIIMet and tcFIHBglIIrev primers, A-tailed, and ligated into pGEM-T Easy.

#### pGEM-T Easy-tcHIF-a (780-879) HindIII

tcHIF- $\alpha$  (780-879) was PCR amplified from pET32a-tcHIF- $\alpha$  using tcHIFCADHindIIIfor and rfbHIF-a 790-879 rev primers, A-tailed and ligated into pGEM-T Easy.

#### pTRDC-ARE-EYFP

AREBgIIIfor and AREBgIIIrev oligos were annealed and then ligated into BgIII-digested pTRDC-EYFP.

#### pTRDC-ARE-FIH-EYFP

pGEM-T Easy-hFIH-BgIII was digested with BgIII, the hFIH fragment purified and ligated into similarly digested pTRDC-ARE-EYFP.

#### pTRDC-ARE-hFIH H199A-EYFP

pGEM-T Easy-hFIH HH199A-BgIII was digested with BgIII, the hFIH H199A fragment purified and ligated into similarly digested pTRDC-ARE-EYFP.

#### pTRDC-ARE-V5-tcFIH-EYFP

pGEM-T Easy-V5-tcFIH-BgIII was digested with BgIII, the V5-tcFIH fragment purified and ligated into similarly digested pTRDC-ARE-EYFP.

## 2.4 Chemicals

$[\alpha\text{-}^{32}\text{P}]$ dATP	Perkin Elmer (NEG512H)
$\alpha\text{-}[1\text{-}^{14}\text{C}]$ 2-OG	Perkin Elmer (NEC5970)
2-OG	Sigma (K1875)
APS	Sigma (A3678)
aprotinin	Sigma (A1153)
Sodium arsenate dibasic heptahydrate	Sigma (455857)
ascorbate	Sigma (A4403)
ATP	Invitrogen (18330-019)
bestatin	Sigma (B8385)
BSA	Sigma (A4919)
$\text{Ca}(\text{OH})_2$	BDH (103042)
Coomassie blue R-250	Sigma (B0149)

cycloheximide	Sigma (C1988)
DMOG	Cayman Chemical (71210)
DMSO	Sigma (D2650)
doxycycline	Sigma (D9891)
DTT	Biovectra (1365)
FeSO <sub>4</sub>	Sigma (F7002)
Fugene 6	Promega
imidazole	Sigma (56749)
IPTG	Biovectra (1882)
leupeptin	Sigma (L9783)
Lipofectamine 2000	Invitrogen
lyticase	Sigma (L2524)
maltose	Sigma (M5885)
Ni IDA agarose	Scientifix
pepstatin	Sigma (P5318)
PMSF	Sigma (7626)
protein g sepharose	Zymed, Life technologies
Restriction enzymes	NEB
TEMED	Sigma (T9281)
UltimaGold XR scintillation fluid	Packard Bioscience

## 2.5 Growth mediums

### 2x YT

1.6% tryptone, 1% yeast extract, 86 mM NaCl in MQ H<sub>2</sub>O (pHed to 7.2 with NaOH).

### LB

5 g bacto-yeast extract, 10 g peptone, 10 g NaCl (dissolved in MQ H<sub>2</sub>O to 1L)

### SOC

2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl<sub>2</sub>, 10 mM MgSO<sub>4</sub>, 0.2% glucose, pH 7.5, autoclaved

#### Selective yeast growth media

800 mL yeast nitrogen base, 100 mL 20% glucose, 100 mL appropriate 10X selective amino acid mix (for plates, 15 g agar added per L of medium)

#### 10X selective amino acid mix

(for 100mL) Appropriate amino acids dissolved in 80mL MQ H<sub>2</sub>O, with heat (<60°C) and stirring. Volume made up to 100 mL and filter sterilised. Volumes scaled up as required. Masses of amino acids required to make a 100mL 10X stock are as follows: 20 mg His, 30 mg Ile, 150 mg Val, 150 mg adenine, 20 mg Arg, 100 mg Leu, 60 mg Trp, 20 mg uracil, 30 mg Lys, 20 mg Met, 50 mg Phe, 200 mg Thr, 30 mg Tyr.

#### YPED

10 g bacto-yeast extract, 20 g bacto-peptone, 20 g dextrose (dissolved in MQ H<sub>2</sub>O to 1L; for plates, 10 g agar added per L of medium)

## 2.6 Antibodies

### 2.6.1 Primary antibodies

#### 9E10 hybridoma

See section 2.29. Mouse monoclonal antibody. For westerns, used neat or at up to a 1 in 4 dilution in PBS, depending on antibody concentration in harvested S/N. For IPs, used neat.

#### α-myc clone 4A6 (Millipore)

Monoclonal mouse antibody used at 1:5000 in PBS for 3hrs at RT or O/N at 4°C.

#### α-FIH (Novus, NB100-428)

Polyclonal rabbit antibody used at 1:1000 in PBS O/N at 4°C.

#### α-FIH (#8)

Polyclonal rabbit antibody generated in-house. Used at 1:1000 in PBS-T O/N at 4°C.

### $\alpha$ -tubulin (Sigma, T8203)

Monoclonal mouse antibody used at 1:10000 in PBS O/N at 4°C.

### $\alpha$ -HIF-1 $\alpha$ (Novus, NB100-449)

Polyclonal rabbit antibody used at 1:1000 in PBS O/N at 4°C.

## 2.6.2 Secondary antibodies

### Goat $\alpha$ -rabbit IgG (Pierce, 31460)

HRP-conjugated, used at 1:20000 in PBS for 1 hr at RT.

### Goat $\alpha$ -mouse IgG (Pierce, 31430)

HRP-conjugated, used at 1:20000 in PBS for 1 hr at RT.

## 2.7 Cloning

### 2.7.1 Restriction digestion

Reactions were set up based on the recommendations of the New England Biolabs handbook, and contained 1-2  $\mu$ g of DNA, 1x NEB buffer, 100 ng/ $\mu$ L BSA (if recommended), and 0.2-0.5  $\mu$ L of each enzyme to be used.

### 2.7.2 Sequence amplification from plasmid DNA or cDNA

25  $\mu$ L PCR reactions were set up containing 1x Pfu Turbo buffer, 62.5 ng each of forward and reverse primers, 0.6 mM dNTPs (containing 0.15 mM of dATP, dGTP, dCTP and dTTP each), 0.5  $\mu$ L of 293T cDNA or 25 ng plasmid DNA template, 1.25 U Pfu Turbo enzyme (Stratagene), and made up to 25  $\mu$ L with MQ H<sub>2</sub>O. Selected reactions also contained 3% DMSO for optimisation. Cycling parameters: 95°C for 1 min, then [95°C for 30 s, 45-65°C depending on primer design for 15 s, and 72°C for 30 s per kb to be amplified] repeated 30-40x, then 72°C for 10 min.

### 2.7.3 Site directed mutagenesis

PCR reactions contained 5  $\mu$ L 10x Pfu buffer, 50 ng of plasmid template, 125 ng of forward and reverse primers, 3  $\mu$ L 10 mM dNTPs, 1  $\mu$ L Pfu Turbo (Stratagene), and were made up to 50  $\mu$ L with



MQ H<sub>2</sub>O. Cycling parameters were as follows: 95°C for 5 min, 18x [95°C for 30 sec, 55°C for 30 sec, 68°C for 18 min].

#### 2.7.4 Overlap extension PCR

Overlap extension PCR first involved the production of 2 overlapping PCR products (products “A” and “B”) in separate PCR reactions. First round amplification reactions contained 5 µL 10x Pfu buffer, 50 ng of plasmid template, 125 ng of the appropriate primers (forward flanking primer and reverse mutagenesis primer for product A, and forward mutagenesis primer with reverse flanking primer for product B), 3 µL 10 mM dNTPs, 1 µL Pfu Turbo (Stratagene), and were made up to 50 µL with MQ H<sub>2</sub>O. Cycling parameters were as follows: 95°C for 3 min, 30x [95°C for 30 sec, 55°C for 15 sec, 72°C for 1 min]. PCR products were purified as detailed in section 2.7.5, then employed in a single second round PCR amplification containing the following: 5 µL 10x Pfu buffer, 5 µL product A, 5 µL product B, 125 ng of forward and reverse flanking primers, 3 µL 10 mM dNTPs, 1 µL Pfu Turbo (Stratagene), and MQ H<sub>2</sub>O up to a volume of 50 µL. Cycling parameters were as for round one. The resulting PCR product was gel purified and cloned into pGEM-T Easy as described in section 2.7.7.

#### 2.7.5 Gel extraction and purification of DNA fragments

DNA to be purified was run on a 1% agarose gel in TBE buffer for 20 min to 1 hr depending on fragment size. Fragments were excised from the gel with a scalpel blade, then purified using a QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer’s instructions.

#### 2.7.6 Ligations

10 µL reactions were set up containing vector DNA and insert fragment DNA in a molar ratio of 1:3.5, 1 mM ATP, 1x T4 ligase buffer and 400 U T4 ligase (NEB), then left at RT for 2 hrs, or placed at 4°C overnight.

#### 2.7.7 PCR product insertion into pGEM-T Easy

The PCR product was first A-tailed by combining 10.5 µL of PCR product, 1.5 µL 10x Taq polymerase buffer, 0.9 µL 25 mM MgCl<sub>2</sub>, 1 µL 0.25 mM dATP, and 1 µL of 5000 U/mL Taq polymerase (NEB), and incubating the mixture at 72°C for 30 min. Then, the A-tailed PCR product was ligated into pGEM-T Easy by combining 5.5 µL of A-tailed PCR product, 7.5 µL 2x ligation buffer, 1 µL of ligase and 1 µL of

pGEM-T Easy from the pGEM-T Easy kit (Promega), and incubation of the reaction either for 2 hrs at RT or overnight at 4°C.

### 2.7.8 PCR product insertion into pENTR-TOPO

Reactions were set up containing 20 ng of PCR product (designed with a CACC sequence at the 5' end as per the manufacturer's instructions), 20 ng of pENTR-TOPO vector (Invitrogen), 1 µL of salt solution and made up to 6 µL with MQ H<sub>2</sub>O. Reactions were incubated at RT for 30 min.

### 2.7.9 Gateway recombination

Reactions contained 1µL 300 ng/µL entry vector, 1 µL 150 ng/µL destination vector, 7 µL of 1x TE, and 1 µL of LR clonase II (Invitrogen). Reactions were incubated at RT for 1 hr before inactivation with 2 µL proteinase K at 37°C for 10 min.

### 2.7.10 Bacterial transformation

5 µL of ligation mixture or Gateway reaction was added to 20-100 µL of chemically competent DH5α *E. coli* bacteria, left on ice for 10 min, then heat shocked in a heating block (42°C for 1 min). Cells were then returned to ice for 5 min before the addition of 500 µL SOC medium and incubation at 37°C for 1 hr. Cells were then plated on 1.5% LB agar containing a suitable selection antibiotic (100 µg/mL ampicillin or 50 µg/mL kanamycin), and left at 37°C overnight.

## 2.8 DNA plasmid preparation

### 2.8.1 Plasmid miniprep

5 mL LB cultures containing a suitable antibiotic (100 µg/mL ampicillin or 50 µg/mL kanamycin) were grown overnight at 37°C before being spun down (3000 g, 15 min, RT), and the S/N removed. The pellet was then lysed for plasmid purification using the QIAprep Spin Miniprep Kit (Qiagen) according to the manufacturer's instructions.

### 2.8.2 Plasmid midiprep

100 mL LB cultures containing a suitable antibiotic (100 µg/mL ampicillin or 50 µg/mL kanamycin) were grown overnight at 37°C before being spun down (3000 g, 15 min, RT), and the S/N removed.

The pellet was then lysed for plasmid purification using the Nucleobond Xtra Midi Plus Kit according to the manufacturer's instructions.

## 2.9 DNA sequencing

PCR amplification reactions were prepared containing 400 ng of plasmid template, 50 ng of sequencing primer, 3  $\mu$ L of 5x sequencing buffer, and 1  $\mu$ L BigDye Terminator v3.1 (Applied Biosystems), and then made up to 15  $\mu$ L using MQ H<sub>2</sub>O. Tubes were then placed in a thermal cycler and subjected to the following cycling parameters: 95°C for 3 min, then 25x [95°C for 30 sec, 50°C for 15 sec, 60°C for 4 min], after which reactions were cooled to 4°C. The reaction was transferred to a microcentrifuge tube along with 80  $\mu$ L 75% isopropanol and 1  $\mu$ L of glycogen and left at RT for 30 min. Tubes were spun (14000 rpm, 20 min, RT), the S/N removed, and the pellet washed with 250  $\mu$ L 75% isopropanol. Tubes were spun as before, the S/N removed and then the pellet air dried for 10 min at 42°C. Reactions were then sent to the Australian Genome Research Facility for analysis.

**5x sequencing buffer:** 200 mM Tris HCl pH9, 5mM MgCl<sub>2</sub>

## 2.10 Electrocompetent cell preparation

30 mL 2x YT medium was inoculated with DH5 $\alpha$  cells and allowed to grow overnight at 37°C. The following morning, 2.4 mL of overnight culture was used to inoculate each of 4 250 mL 2x YT medium cultures which were then grown to an OD of 0.8. Cells were transferred to centrifuge bottles and left on ice for 1hr, then spun down (4500 rpm, 15 min, 4°C). Pellets were resuspended 2x in 150 mL cold 10% glycerol each with spins (4500 rpm, 15 min, 4°C) in between, after which the pellet was resuspended in 50 mL 10% glycerol, transferred to a 50 mL tube and spun (3400 rpm, 15 min, 4°C). The pellet was resuspended in 2 mL 10% glycerol, then split into 40  $\mu$ L aliquots and snap frozen.

## 2.11 Yeast 2-hybrid assay

Refer to (Hampton-Smith, 2004).

## 2.12 Y2H prey plasmid identification

### 2.12.1 Plasmid isolation from yeast

A single yeast colony was used to inoculate 5 mL of SC-Leu-Trp medium and grown overnight, shaking at 30°C. Cultures were spun down (3 min, 2000 rpm, 4°C) and transferred to screw cap eppendorfs in 0.5 mL of medium, then spun again to remove the S/N. Pellets were resuspended in Back extraction buffer, and enough glass beads were added to sit just below the surface of the liquid. 200 µL of phenol:chloroform:isoamyl alcohol (25:24:1) was added, the tube vortexed for 1.5 min, cooled on ice for 30 sec, then vortexed for a further 1.5 min. Tubes were spun (5 min, 14000 rpm), and the aqueous layer removed and transferred to a new eppendorf tube. To precipitate the DNA, 1/10 of a volume of 3M NaAc and 2 volumes of 95% EtOH were added, and the tube placed at -80°C for 20 min. Tubes were spun (10 min, 14000 rpm, 4°C), the S/N removed, and the pellet washed with 300 µL 70% EtOH before again spinning (1 min, 14000 rpm, 4°C) and the S/N being removed. Pellets were air dried at 42°C and then resuspended in 15 µL of MQ H<sub>2</sub>O.

### 2.12.2 Bacterial electroporation

5 µL of yeast plasmid DNA was added to 50 µL electrocompetent cells in a microcentrifuge tube and incubated on ice for 20 min. Cells and DNA were then transferred to a pre-chilled electroporation cuvette and electroporated (2.5 kV), straight after which 1 mL of SOC medium was added and the cells incubated at 37°C for 1 hr. Tubes were spun (1 min, 2500 rpm), 800 µL of S/N removed, and the remaining 200 µL used to resuspend the cells for plating on LB agar plus 100 µg/mL ampicillin. After incubation of plates O/N at 37°C, colonies were used to inoculate 5 mL LB cultures, then DNA extracted and plasmids sequenced as detailed in sections 2.8.1 and 2.9.

## 2.13 Yeast colony hybridisation

### 2.13.1 Ferritin probe preparation

8 µg of pGEM-T Easy-mFerritin was digested with SphI and NotI enzymes, the reaction run on a 1% agarose gel and the 200bp band corresponding to the 3' end of the ferritin coding sequence excised and purified using the Qiagen Gel extraction purification kit. The purified DNA solution was then placed at 70°C to evaporate the elution buffer and concentrate the DNA. Next, 12 µL of 3' ferritin probe, 1 µL of MQ H<sub>2</sub>O and 1 µL of 300 ng/µL random hexamer primer (Geneworks) were combined

in a tube, boiled for 10 min and then immediately placed on ice. 2.5  $\mu\text{L}$  of dGTP/dCTP/dTTP (at 5 mM each), 2.5  $\mu\text{L}$  of 10x TM buffer, 5  $\mu\text{L}$  of [ $\alpha$ - $^{32}\text{P}$ ] dATP (3000 Ci/mmol, Perkin Elmer) and 2  $\mu\text{L}$  of 2u/mL Klenow enzyme (NEB) were added and the reaction and incubated at 37°C for 1 hr. 25  $\mu\text{L}$  of STE buffer was added to the reaction after which it was added to a ProbeQuant G50 Micro Column (GE Healthcare), and spun (2 min, 3000 rpm). The eluate was diluted in 350  $\mu\text{L}$  of 1x TE, and 2  $\mu\text{L}$  subjected to scintillation counting to verify radioactive [ $^{32}\text{P}$ ] dATP incorporation.

**10x TM buffer:** 100 mM Tris-HCl pH 8, 100 mM  $\text{MgCl}_2$  in MQ  $\text{H}_2\text{O}$ .

**STE buffer:** 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 150 mM NaCl (pHed with NaOH).

**1x TE buffer:** 10 mM Tris HCl pH 7.5, 1 mM EDTA.

### 2.13.2 Filter preparation

Note: filter preparation steps were adapted from the protocol suggested in the ProQuest Two-Hybrid System manual (GibcoBRL).

Potential positives to be tested as well as positive (containing prey plasmid encoding ferritin) and negative control yeast (containing empty pPC86, pPC86-HIF-1 $\alpha$  (727-826) or pPC86-HIF-2 $\alpha$  (565-871)) were streaked out in an asymmetrical grid on SC-Leu-Trp agar and allowed to grow for 4 days at 30°C. Yeast streaks were then transferred onto a nylon membrane, the membrane placed yeast-side-up on YPED agar, and allowed to grow overnight at 30°C. To aid with orientation, a pin was used to punch a hole adjacent to each streak. The membrane was allowed to dry for 5 min at RT, and then placed yeast-side-up onto a Whatman 3MM paper pre-soaked in freshly prepared sorbitol/EDTA/DTT solution for 30 min. The membrane was then freeze-thawed 5 times at -80°C for 5 min. Next, the membrane was placed on a fresh Whatman paper soaked with 10 mL of sorbitol/EDTA solution containing 200 U/mL of lyticase and incubated at 37°C for 4 hrs. The membrane was then sequentially placed on top of Whatman papers soaked with the following solutions: 0.5 M NaOH for 10 min, 0.5M Tris-HCl (pH 7.5)/6x SSC for 5 min (repeated with a second soaked sheet), then 2x SSC for 5 min, then allowed to dry on clean Whatman paper for 10 min. 1 ng of purified plasmid DNA of pPC86-ferritin, pPC86-HIF-1 $\alpha$  (727-826) or pPC86-HIF-2 $\alpha$  (565-871) was boiled for 10 min, immediately cooled on ice and then spotted onto the membrane prior to UV crosslinking the membrane 2x with a Stratalinker UV crosslinker using the "Autocrosslink" function.

**Sorbitol/EDTA/DTT solution (prepare fresh):** 1M sorbitol, 20mM EDTA, 50mM DTT.

**20x SSC (pH 7):** 3 M NaCl, 300mM trisodium citrate (pHed to 7.2 with citric acid).

**SC-Leu-Trp medium:** 800 mL yeast nitrogen base, 100 mL 20% glucose, 100 mL 10X amino acid mix lacking leucine and tryptophan (for plates, 15 g agar added per L of medium).

**10x amino acid mix:** (for 100mL) Appropriate amino acids dissolved in 80 mL MQ H<sub>2</sub>O, with heat (<60°C) and stirring. Volume made up to 100 mL and filter sterilised. Volumes scaled up as required. Masses of amino acids required to make a 100 mL 10X stock are as follows: 20 mg His, 30 mg Ile, 150 mg Val, 150 mg adenine, 20 mg Arg, 100 mg Leu, 60 mg Trp, 20 mg uracil, 30 mg Lys, 20 mg Met, 50 mg Phe, 200 mg Thr, 30 mg Tyr.

**YPED medium:** 10 g bacto-yeast extract; 20 g bacto-peptone; 20 g dextrose (dissolved in MQ H<sub>2</sub>O to 1L; for plates, 10 g agar added per L of medium).

### 2.13.3 Ferritin probe hybridisation

Membranes were soaked with 5 mL each of hybridisation solution and stacked with mesh in between. Membranes were transferred to a hybridisation bottle and generously covered with hybridisation solution (20 mL), then incubated at 65°C for 1 hr. In a screw-cap tube, 2 mg sonicated herring sperm DNA and sufficient radioactive ferritin probe to achieve 15 ng/mL in the total hybridisation solution were combined, boiled for 10 min then immediately cooled on ice. 2 mL of hybridisation solution was then added to the boiled probe before transferral into the hybridisation bottle and incubation overnight at 65°C. Membranes were then briefly washed 2x with low stringency wash buffer at RT, then 2x quick washes with high stringency wash buffer (pre-warmed to 65°C), followed by 2x longer 20 min washes with high stringency wash buffer at 65°C. Filters were allowed to dry, then wrapped in Glad wrap and left on a phosphorimager for 14 days before scanning by a Typhoon imager (100 µm resolution, best sensitivity; GE).

**Hybridisation solution (can be stored at RT):** 0.5 M NaHPO<sub>4</sub> buffer (pH 7.2), 1 mM EDTA, 1% BSA, 7% SDS.

**Low stringency wash buffer:** 0.5% BSA, 1 mM EDTA, 40 mM NaHPO<sub>4</sub> buffer (pH 7.2), 5% SDS, up to 1L with ddH<sub>2</sub>O.

**High stringency wash buffer:** 1 mM EDTA, 40 mM NaHPO<sub>4</sub> buffer (pH 7.2), 1% SDS, up to 1L with ddH<sub>2</sub>O.

## 2.14 Protein expression and purification

### 2.14.1 Culture growth

A 50 mL LB culture containing 100 µg/mL ampicillin was inoculated with a single colony from transformed BL21 DE3 bacteria and grown overnight, shaking at 37°C. The following morning, the overnight culture was spun down and resuspended in fresh LB, then a 500 mL LB culture plus 100 µg/mL ampicillin was inoculated such that the OD was 0.1. The culture was grown to an OD of 0.4 for bacteria transformed with pET32a expression constructs, or 0.6-0.8 for pMBP expression constructs, then induced with 0.2mM IPTG and placed at 30°C for 3-5 hrs. Cultures were spun down, the medium removed and pellets stored at -20°C until protein preparation.

### 2.14.2 Protein purification, batch

Note: the buffers used for Ni<sup>2+</sup> affinity and MBP purifications are shown below.

Pellets were resuspended in 30 mL cold lysis buffer, passed 3 times through a cell disruptor (Microfluidics M-110L Pneumatic Microfluidizer), then spun (15000g, 45 min, 4°C). The S/N was added to 1 mL of Ni-IDA agarose (Ni<sup>2+</sup> affinity purification; Scientifix) or amylose agarose (MBP purification; Scientifix) and incubated rocking at 4 °C for 1 hr. The resin-S/N mixture was then poured into an empty PD-10 column (GE healthcare) and the resin allowed to settle before being washed with 50 mL lysis buffer, 50 mL wash buffer, and then eluted with 2.5 mL of elution buffer.

**Lysis buffer (Ni<sup>2+</sup> affinity):** 20 mM Tris-HCl pH 8, 150 mM NaCl, 5 mM imidazole, 1 mM PMSF (added fresh), 0.5 mM DTT (added fresh).

**Wash buffer (Ni<sup>2+</sup> affinity):** 20 mM Tris-HCl pH 8, 150 mM NaCl, 10 mM imidazole.

**Elution buffer (Ni<sup>2+</sup> affinity):** 20 mM Tris-HCl pH 8, 150 mM NaCl, 250 mM imidazole.

**Lysis buffer (MBP):** 20 mM Tris-HCl pH 8, 150 mM NaCl, 1 mM PMSF (added fresh).

**Wash buffer (MBP):** 20 mM Tris-HCl pH 8, 150 mM NaCl, 1 mM PMSF.

**Elution buffer (MBP):** 20 mM Tris-HCl pH 8, 150 mM NaCl, 10 mM maltose.

### 2.14.3 Protein purification, Profinia™

Bacterial cell lysates were prepared as for “Protein purification, batch”, then loaded onto a 1 mL Bio-scale Mini Profinity IMAC cartridge using the Profinia protein purification system (Bio-rad). The column was washed with 6 column volumes (CV) lysis buffer, 6 CV mL wash buffer, and then eluted in 3.5 CV elution buffer. 2.5 mL of eluate was then loaded onto a PD-10 column and buffer exchanged as for “Protein purification, batch”.

### 2.15 Protein buffer exchange

A PD-10 desalting column (GE) was equilibrated with 25 mL of the desired exchange buffer (for all applications except purification of Trx-6H-tcHIF (780-879) (see section 2.18), the exchange buffer was protein storage buffer). 2.5 mL of purified protein was then applied to a PD-10 column, and eluted with 3.5 mL exchange buffer. In most cases, the first 0.5 mL fraction of the eluate was discarded to maximise protein concentration.

**Protein storage buffer:** 20 mM Tris-HCl pH 8, 150 mM NaCl.

### 2.16 Purified protein yield determination

Protein concentration was calculated by measuring the absorbance at 280 nM using an Eppendorf Biophotometer, and utilising the protein’s extinction coefficient calculated using the EXPASY Prot param server (<http://web.expasy.org/protparam/>) in the equation  $C = A/\epsilon$ .

### 2.17 Protein concentration

Protein purifications that resulted in poor recovery of soluble protein were concentrated via centrifugation in an Amicon Ultra 4 mL concentration column (MW cut-off 3000 or 10000 Da, depending on the size of the protein to be concentrated; Millipore). The column was first spun containing 4 mL of the buffer in which the protein was dissolved, which was then removed and replaced with protein sample to be concentrated. Protein samples were spun at 2000 g for 5 min, the



liquid above the filter gently mixed using a pipette, and this procedure repeated until the volume above the filter was typically between 0.5-1 mL. This liquid was then removed and transferred to a clean LoBind™ eppendorf tube.

## 2.18 Preparation of purified tcHIF (780-879) for NMR

### 2.18.1 Expression, purification and concentration

Trx-6H-tcFIH (780-879) was purified from 1.5 L of LB medium supplemented with 100 µg/mL ampicillin as documented in section 2.14.2, and then buffer exchanged as per section 2.15. The 10.5 mL of buffer exchanged eluate was then concentrate to ~2.3 mL using a 15 mL Amicon Ultra column (MW cut-off 10000 Da) equilibrated with protein storage buffer.

### 2.18.2 TEV cleavage of 6His tag and tag/AcTEV removal

A 2.5 mL TEV cleavage reaction containing 72 mg purified protein were set up according to the manufacturer's instructions and employed 33U of AcTEV protease (Life Technologies) per mg of protein digested. Reactions were left at 30°C for 2 hrs, then left at 4°C O/N. Next, 0.7 mL of Ni<sup>2+</sup> IDA agarose, 100 µL of 4M NaCl, and 2 µL of 6M imidazole were added to the reaction, and the mixture incubated at 4°C with mixing for 1 hr. The tube was then spun (500g, 3 min, 4°C), and the S/N transferred to a fresh tube.

### 2.18.3 Buffer exchange and concentration

The S/N from above was buffer exchanged into low salt phosphate buffer as per section 2.15. The eluate was then concentrated down to ~ 500 µL using an Amicon Ultra 4 mL column with a cut-off of 3000 Da as detailed in section 2.17.

**Phosphate buffer:** 20 mM phosphate buffer pH 6.8 (Na<sub>2</sub>HPO<sub>4</sub> and NaH<sub>2</sub>PO<sub>4</sub> combined in a ratio of 0.98:1.02), 50mM NaCl

## 2.19 NMR

### 2.19.1 1D $^1\text{H}$

Spectra were obtained using an Agilent 600 MHz spectrometer with cryo-probe at 25°C. 1D – pulse sequence was first increment of Watergate NOESY. Spectral width, 12019 Hz (20 ppm); number of points, 2048; 256 transients.

### 2.19.2 2D NOESY

Spectra were obtained using an Agilent 600 MHz spectrometer with cryo-probe at 25°C. Sw1 = sw2 = 12019.23 Hz; 2k data points in first dimension; 680 increments in second dimension; 32 scans per increment.

## 2.20 Collation and analysis of ankyrin repeat sequences across species

Species were selected to obtain a variety of representatives from multiple different phyla. Identification of FIH homologs was performed by BLASTing the human FIH sequence (Refseq accession NP\_060372.2) with NCBI's BLASTp and tBLASTn algorithms. Partial matches to the human FIH sequence were, where possible, extended to include additional theoretical exons via manual translation and searching of genomic sequence adjacent to the identified match. If no BLAST hit was found which contained clear matches to human FIH residues Tyr145, Thr196, His199, Asp201, K214, Arg238, Gln239 and His279, the species was classified as lacking FIH. Upon selection of species for analysis, ankyrin repeats encoded by each species' proteome were identified using 3 mechanisms: (1) a HMMER3.0-based search (Finn et al., 2011) of the relevant proteome using the Pfam database ankyrin repeat hidden Markov model (Pfam accession PF00023, <http://pfam.xfam.org/> ; proteomes downloaded from the UNIPROT database <http://www.uniprot.org/>), (2) species-filtered searches of UNIPROT proteomes using Scan Prosite (de Castro et al., 2006) and Prosite motif PS50088 (<http://prosite.expasy.org/scanprosite/>), and (3) direct download of ankyrin repeats for each species from the SMART database (Schultz et al., 1998). Sequences for each species were collated, extended (where necessary) to 33 aa in length (or longer if the repeat was predicted to contain insertions) using the Uniprot sequence database, and then filtered to remove duplicate sequences (note: this process may have resulted in the removal of genuine occurrences of duplicate sequences in different ankyrin repeat domains, but was primarily intended to avoid inclusion of the same sequences with different identifiers). Sequences aligned for WebLogo representation had gaps included, or sequence

insertions deleted (“alignment-adjusted” repeats, as predicted by the Prosite and HMMER3.0 algorithms), in order to more accurately align key regions of the ankyrin repeat consensus (Mosavi et al., 2004). Note that insertions predicted to occur between aa 21 and 29 of the 33 aa repeat were not removed in order to preserve the sequence spacing in the FIH-binding region. Analysis of FIH target and control motif frequencies was performed on alignment-adjusted repeats using Excel. Graphs were drawn using GraphPad Prism 6. Linear regression and  $r^2$  values were calculated using Prism’s straight line fitting algorithms, while  $\chi^2$  values were calculated using Excel. Probabilities for F and  $\chi^2$  distributions were calculated using Excel’s FDIST and CHIDIST functions, respectively.

### 2.21 *In vitro* hydroxylation assay

For each substrate to be tested, 3 positive reactions (containing FIH, concentrations as specified in the text) and 2 negative control reactions (containing protein storage buffer) were set up. Filters of Whatman paper were cut to fit in the cylindrical portion of 1.5 mL, conical, air-tight screw-cap eppendorf, such that they would stay suspended above a 40  $\mu$ L reaction in the bottom of the tube, then soaked in a ~30 mM solution of  $\text{Ca}(\text{OH})_2$ . For each reaction, 5  $\mu$ L of MBP- or Trx-6H-tagged FIH or protein storage buffer was added to a screw-cap eppendorf tube and placed on ice. 25  $\mu$ L of relevant substrate was then added to all tubes (substrate concentrations as specified in the text) and mixed. Next, a 320  $\mu$ M mastermix of 2-OG was prepared (5  $\mu$ L per reaction), containing  $\alpha$ -[1- $^{14}\text{C}$ ] 2-OG (specific activity 54.5 mCi/mmol, Perkin Elmer) and cold 2-OG in a ratio of 1:1. This was followed by preparation of a cofactor mastermix (5  $\mu$ L per reaction) containing 0.8% BSA, 160 mM Tris-HCl (pH 7 for HIF substrates or pH 7.5 for ARD substrates unless otherwise stated), 4 mM DTT, 8 mM ascorbate,  $\text{FeSO}_4$  to a concentration 80x that of FIH in the reaction, and MQ  $\text{H}_2\text{O}$ . Just prior to beginning the assay, the 2-OG and cofactor solutions were mixed, then each reaction was initiated by the addition of 10  $\mu$ L of 2-OG/cofactors, resulting in final concentrations of 0.1% BSA, 20 mM Tris-HCl, 0.5 mM DTT, 1 mM ascorbate,  $\text{FeSO}_4$  at 10x the concentration of FIH in the reaction, and 40  $\mu$ M 2-OG (specific activity 18 Ci/mmol). The tube was mixed, a  $\text{Ca}(\text{OH})_2$ -soaked filter blotted dry and then placed into the tube, the lid closed, and then the tube placed into a 37°C heating block for 30 min. Reactions were terminated by removal of the filter, which was allowed to dry for at least 30 min before the addition of 200  $\mu$ L of UltimaGold™ XR scintillation fluid, and reading of counts per minute using a MicroBeta<sup>2</sup> plate reader (Perkin Elmer).

## 2.22 Mammalian cell culture

All cell lines were maintained in a humidified incubator at 37°C and 5% CO<sub>2</sub>. HEK 293T cells (obtained from ATCC (Vancouver, USA)) and 9E10 hybridoma cells were grown in DMEM (Gibco cat no. 12430), supplemented with 10% FCS and 2 mM glutamine, and passaged 1 in 10 every 2-3 days using trypsinisation. Murine embryonic fibroblasts (MEFs) with a floxed FIH allele were generously provided by Randall Johnson (Wellcome Trust, Cambridge, UK), and maintained in DMEM (Gibco, cat no. 11995) supplemented with 10% FCS and 2 mM glutamine, and passaged 1 in 10 or 1 in 20 using trypsinisation every 2-3 days. KO FIH MEFs were generated by S. Karttunen by infection of FIH-floxed MEFs with adenovirus expressing Cre recombinase. Hypoxia treatments were performed by sealing culture vessels in an air-tight container with an AnaeroGen sachet (2.5 L).

## 2.23 Mammalian cell transfection

HEK 293T cells were transfected using Lipofectamine 2000 (Invitrogen) at 30-60% confluency, and MEFs were transfected using Fugene 6 (Promega) at 30-80% confluency, each according to the manufacturer's instructions.

## 2.24 Mammalian whole cell extract preparation

Cells were harvested using trypsin, then spun down (200g, 3 min, RT), and then transferred to a microcentrifuge tube using 1 mL of ice cold PBS. Cells were spun again (1400 rpm, 5 min, 4°C) and the S/N removed, after which ice cold WCE buffer (100-150 µL for a T75 culture flask) was added, and the mixture rocked at 4°C for 40 min. Tubes were then spun (14000 rpm, 30 min, 4°C), the S/N removed and transferred to a fresh microcentrifuge tube. Extracts were stored at -80°C prior to use.

**WCE buffer:** 20 mM HEPES, 0.42 mM NaCl, 0.5% NP-40, 25% glycerol, 1.7 mM EDTA, 1.5 mM MgCl<sub>2</sub>, with freshly added 1 mM PMSF, 1mM DTT, 2 µg/mL aprotinin, 4 µg/mL bestatin, 5 µg/mL leupeptin, 1 µg/mL pepstatin.

## 2.25 SDS PAGE

### 2.25.1 Gel preparation

For a 10% Tris-glycine separating gel, 2.5 mL of 4x separating gel buffer, 3.33 mL of Accugel (30%) 29:1 acrylamide:bis (National Diagnostics), 4.17 mL MQ H<sub>2</sub>O, 8 µL of TEMED and 80 µL of 10% APS were mixed, then quickly transferred to a Mini-PROTEAN glass plate mould (Biorad), layered on top with 70% ethanol, and allowed to set for at least 20 min. For the separating gel, the 70 % ethanol from the previous step was poured off, then 2.5 mL of 4x stacking gel buffer, 1.67 mL of Accugel (30%) 29:1 acrylamide:bis, 5.83 mL MQ H<sub>2</sub>O, 8 µL of TEMED and 80 µL of 10% APS were mixed, then quickly transferred to the glass plate mould, and a well comb applied. The stacking gel was allowed to set for at least 20 min prior to use of the gel.

**4x separating gel buffer:** 181.1 g Tris, 40 mL 10% SDS, up to 1 L with MQ H<sub>2</sub>O, pHed to 8.8 with HCl.

**4x stacking gel buffer:** 60.5 g Tris, 40 mL 10% SDS, up to 1 L with MQ H<sub>2</sub>O, pHed to 6.8 with HCl

### 2.25.2 Electrophoresis

Gel were separated in a tank containing 1x GTS running buffer at 160 V for approximately 1 hr depending on separation required.

**1x GTS:** 14.4 g glycine, 30 g Tris, 10 g SDS, made up to 1L with MQ H<sub>2</sub>O.

## 2.26 Coomassie protein staining

Gels were transferred to Coomassie stain and incubated at RT O/N with gentle shaking, then placed in destain the following day until bands were visible and a colourless background were obtained (often performed O/N).

**Coomassie stain:** 0.03% Coomassie blue R-250 (Sigma), 8.75% acetic acid, 50% methanol

**Destain:** 5% acetic acid, 50 % methanol

## 2.27 Western blot

All western blots were performed using a wet transfer technique. Transfer sandwiches were prepared immersed in transfer buffer and contained, in order from cathode to anode, a fibre pad, Whatman paper, nitrocellulose membrane (PALL BioTrace NT), SDS PAGE gel to be transferred, Whatman paper, fibre pad. Gel holder cassettes containing transfer sandwiches were placed in the electrode rack in a Mini Trans Blot tank (Bio-Rad) containing transfer buffer, and transferred (2 hrs, 250 mA) in a 4°C room. Post transfer, the nitrocellulose membrane was removed and blocked in 10% w/v skim milk for at least 30 min. Primary antibody was then applied for either 3hrs at RT or O/N at 4°C. Next, the membrane was washed 3x (5 mins each) with PBS-T, and then the secondary antibody applied for 1 hr at RT. The membrane was again washed 3x, the excess moisture removed by blotting on tissue paper and then Immobilon Western Chemiluminescent HRP Substrate (Millipore) applied as per the manufacturer's instructions. Chemiluminescence was visualised by either exposure of the blot to x-ray film (AGFA) followed by development with an AGFA CP1000 processor, or via imaging of the blot using a Chemidoc MP Imager (Bio-Rad).

**Transfer buffer:** 50 mM Tris, 2.85% glycine

**PBS-T:** 0.2 g  $\text{KH}_2\text{PO}_4$  (anhydrous), 1.15 g  $\text{Na}_2\text{HPO}_4$  (anhydrous), 8g NaCl, 0.2 g KCl, 0.1% TWEEN 20 (Sigma-Aldrich) made up to 1 L with MQ  $\text{H}_2\text{O}$ .

## 2.28 Western blot band quantitation

The Volume Tools in Image Lab software (Bio-Rad) were used to select rectangular areas of equivalent size around each protein band to be quantified. An additional set of rectangular areas of the same size were drawn at a consistent height above or below the protein band of interest to serve as a background control. Band intensities were calculated by subtracting the background "volume" from the protein band "volume".

## 2.29 9E10 hybridoma S/N preparation

9E10 hybridoma cells were seeded at 10% confluency in T175 flasks containing 100 mL of culture medium and allowed to grow until the medium just started to gain a noticeable orange hue (typically 3-4 days), after which the growth medium was removed, transferred to 50 mL falcon tubes, and spun

(1000 g, 15 min, RT) to remove cell debris. The S/N was then removed, placed in a fresh tube and sodium azide added to a concentration of 0.02% before storage at 4°C.

## 2.30 Anti-myc co-immunoprecipitation

### 2.30.1 WCE preparation

HEK 293T cells were seeded in T75 flasks at a confluency of 10%, and then transfected with 10-20 µg of the relevant pEF-6myc-BOS expression vector the following day. In experiments requiring chemical treatments, these were administered at least 4 hrs post transfection at the indicated times and concentrations. Approximately 24 hrs later, cells were harvested for WCE preparation as detailed in section 2.24. To ascertain extract concentration, extracts were diluted 1 in 50 in WCEB, then a standard curve generated using one of the extracts by adding between 3 and 20 µL of extract to 200 µL of 1x Bradford reagent (Biorad). Absorbance at 600 nM was determined using a plate reader (Multiskan Ascent), then absorbance values were plotted against volume of extract added, and a line of best fit determined using Excel's trendline fitting function. For the remaining extracts, 10 µL of each extract was added to Bradford reagent, and relative concentrations of extract determined from the standard curve equation.

### 2.30.2 Co-immunoprecipitation

Protein G Sepharose (PGS) 4B (Zymed Laboratories, Invitrogen) at 30 µL per IP to be performed, was first washed with IP binding buffer 2x, then resuspended in 500 µL of 9E10 hybridoma S/N per IP and incubated O/N at 4°C. The following day, 500 µL of PGS/9E10 S/N mixture was aliquotted into microcentrifuge tubes (such that each tube received the equivalent of 30 µL of PGS slurry), the tubes spun down (500g, 1 min, 4°C), and the 9E10 S/N removed. Next, equivalent amounts of WCE were added to the PGS (typically, 100 µL of the least concentrated WCE was added, and suitable volumes to achieve an equivalent amount of protein was added of the other WCEs employed in the assay (see section 2.30.1 above). The total volume in each tube was made up to 300 µL with IP binding buffer, and the IPs incubated for 2 hrs at 4°C. Tubes were spun down (500g, 1 min, 4°C), the S/N removed, and the PGS washed 2x by adding 1 mL IP binding buffer and inverting the tube 10 times followed by centrifugation. Bound complexes were eluted by the addition of 25 µL of 4x sample buffer, boiling for 10 min at 95°C, and then removal of 20 µL of S/N. Eluates were then analysed by SDS PAGE followed by western blotting for endogenous FIH.

**IP binding buffer:** 250 mM NaCl, 20 mM HEPES pH 8, 0.1% IGEPAL, 1mM EDTA pH8 plus freshly added 1mM PMSF, and 1x protease inhibitors.

**4x sample buffer:** 32 % glycerol, 0.01% bromophenol blue, 5% SDS, 50 mM Tris pH 6.8, and freshly added 100 mM DTT.

### 2.31 Dox-inducible FIH MEF cell line generation

80% confluent FIH KO MEFs in a T75 culture flask were first transfected with 30 µg of pEF-Puro-TETON, split 1 in 2 the following day, then after another 24 hrs, treated with 2 µg/mL puromycin. Once colonies were established, cells were trypsinised and combined to generate a polyclonal cell line stably expressing the TETON regulator (Clontech), and the puromycin concentration was increased to 10 µg/mL over a period of 2 weeks. Next, monoclonal cell lines were generated by diluting the cells such that single cells were plated in each well of a 96-well tray. To assist with cell growth during the dilution process, cells were grown in a 3:1 mixture of fresh and conditioned medium (growth medium previously used to feed MEFs for at least 24 hrs which was then centrifuged and filtered with a 0.45 micron filter (Sartorius Minisart) to remove contaminating cells). To test for successful stable integration of the pEF-Puro-TETON plasmid, monoclonal lines were plated in 6 well trays, transiently transfected with 2 µg pTRDC-ARE-FIH-EYFP the following morning. After 4 hrs, cells were treated with 2 µg/mL doxycycline for at least 24 hrs prior to lysis for analysis of FIH expression via western blot.

The pEF-Puro-TETON stable cell line, A10, was then plated in T75 flasks at 10% confluency, and transfected the following day with 15 µg of either pTRDC-ARE-FIH-EYFP or empty pTRDC-ARE-EYFP. Cells were then treated for 36 hrs with 2 µg/mL dox, starting 4 hrs after the transfection, after which cells were subjected to FACS sorting for EYFP positive cells. Collected cells were then expanded and cultured for 2 weeks prior to a dox treatment and a second round of FACS sorting for EYFP positive cells, during which positive cells were aliquotted into 96 well trays (pre-seeded with FIH KO MEF “helper cells” at a confluency of 5%) at a concentration of 1 cell per well. After 4 days of monoclonal growth with helper cells, puro was added to the medium at a concentration of 2 µg/mL to remove helper cells. Surviving colonies were expanded, and then tested for dox inducibility of the pTRDC-EYFP construct via plating in 6 well trays, treatment with 2 µg/mL dox for 24 hrs, and analysis of FIH or V5-tcFIH expression via western blot. Due to poor survival of monoclonal lines generated by FACS,



additional monoclonal lines were generated via the same procedure used to generate pEF-Puro-TETON monoclonal lines, and tested for dox inducibility of FIH/V5-tcFIH as for the FACS-generated lines.

## 2.32 Reporter assays

For assays assessing effect of myc-tagged Y2H positives on HIF-1 $\alpha$  (727-826) transactivation: 293T cells or WT and KO MEFs were seeded at 30% or 10 % confluency, respectively, in 24-well trays, left O/N, then transfected in triplicate the following day with 10 ng pRL-TK, 150 ng pGRE, 100 ng of either pEF-GalDBD-empty or pEF-GalDBD-HIF-1 $\alpha$  (727-826) and 200 ng of the indicated pEF-6myc Y2H positive vectors. After 24 hrs, cells were lysed with 100  $\mu$ L of Passive lysis buffer (Promega), then 20  $\mu$ L of sample was assayed for luciferase production using the Dual luciferase reporter assay system (Promega).

For assays assessing effect of FIH expression on FGIF transactivation:

Assays were performed as described above for 293Ts, except that 200 ng of pEF-GalDBD-empty or pEF-GalDBD-FGIF was transfected, and 200 ng of pcDNA3.1-V5-FIH or pcDNA3.1-V5-FIH H199A was substituted for pEF-6myc Y2H positive vectors.

For assays with dox inducible cell lines:

Cell lines were seeded at 10 % confluency, then transfected as above minus the pEF-6myc constructs. 5 hrs was allowed to elapse before chemical treatments were applied. For assays including DMOG treatments, relevant wells were treated with 2  $\mu$ g/mL dox and 1 mM DMOG for 20 hrs prior to cells being lysed and assayed as above. For assays involving dox titrations, relevant wells were treated with the indicated concentrations of dox for 20 hrs before cell lysis and analysis as above.

## 2.33 qPCR analysis of G9a target genes

Experiments performed by Jolene Caifeng Ho and Jia You under the supervision of Kian Leong Lee (Cancer Science Institute of Singapore, National University of Singapore).

### 2.33.1 Cell treatments

FIH<sup>-/-</sup> Tet-hFIH cells or FIH<sup>-/-</sup> Tet-empty vector control MEFs cultured in 96-well trays were split into 4 groups and were treated with 200 ng/mL dox at 0, 24 or 48 hrs, or left untreated (resulting in dox

treatment times of 72, 48, 24 or 0 hrs, respectively). Care was taken to avoid the cells becoming too overcrowded at the end of the experiment to avoid any peri-cellular hypoxic effects.

### 2.33.2 qPCR

RNA extractions were performed using a 96 well format, and utilised the ZR-96 RNA Clean & Concentrator kit (Zymo Research Corporation) according to manufacturer recommendations. 200ng of total RNA was reverse transcribed into cDNA using the High Capacity RNA-to-cDNA Master Mix (Applied Biosystems). The cDNA was preamplified with a 95°C hot start for 10 mins followed by 16 cycles of 15 secs denaturation at 95°C and 4 mins elongation at 60°C using the TaqMan PreAmp Master Mix (Applied Biosystems). Subsequently, all samples were quantified by qPCR using the Power SYBR Green PCR Master Mix (Applied Biosystems) supplemented with AmpliTaq Gold DNA polymerase (Applied Biosystems) on the BioMark real-time PCR system (Fluidigm Corporation) according to the manufacturer's specifications. Genes selected for qPCR analysis were chosen from a list of the most highly dysregulated targets between WT and G9a KO mouse ES cells identified by microarray. Selected genes were also chosen based on their lack of response to hypoxia (Kian Leong, person communication). Changes in gene expression were determined using the delta-delta Ct method comparing each dox treatment against the average of the 0h controls for each cell line and dox concentration after normalisation against the mouse Hmbs housekeeping gene (n = 3, 8 technical replicates in each biological replicate). Primer sequences used can be found in section 2.2.2.

### 2.34 I $\kappa$ B $\alpha$ biophysical characterisation methods

For all I $\kappa$ B $\alpha$  methods, please refer to (Devries et al., 2010).

# 3 Results – Part 1

Isolation and characterisation of novel

FIH binding partners



### 3.1 Introduction

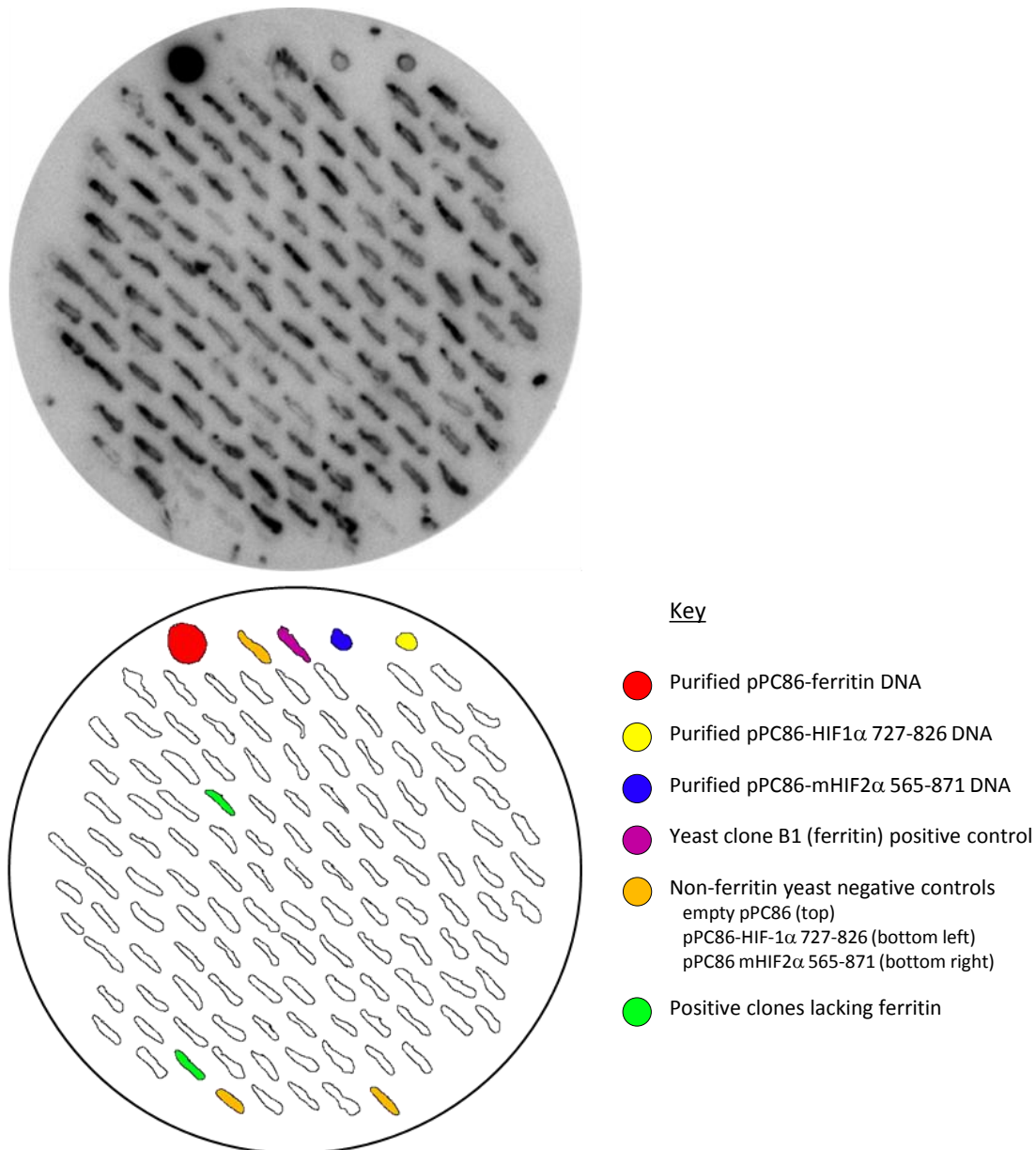
Since its characterisation as a member of the Fe<sup>2+</sup> and 2-OG-dependent oxygenase family, FIH has garnered considerable interest as a candidate cellular oxygen sensor. At the commencement of this project, the only known substrates of FIH were the oxygen-sensitive transcription factors, HIF-1 $\alpha$  and HIF-2 $\alpha$ , the hydroxylation of which serves to repress transactivation activity of the CAD (Lando et al., 2002b). However, FIH appears to interact with a HIF-1 $\alpha$  substrate peptide largely through backbone interactions (Elkins et al., 2003), suggesting that a large array of peptide sequences could be accommodated by FIH's substrate binding groove. Hence, the primary goal of this work is to identify novel substrates of FIH, with the ultimate aim being to increase understanding FIH's role in oxygen sensing.

When faced with the challenge of finding new substrates, a number of methods, both directed and non-biased, may yield results. Unfortunately, due to FIH's seeming lack of specificity when binding substrate residues (Elkins et al., 2003), bioinformatic prediction of substrates was not considered the most viable approach. However, of the various "non-biased" options available, a yeast 2-hybrid (Y2H) assay was chosen to permit screening of a large number of potential substrate proteins. The Y2H has a number of advantages, in that it is an *in vivo*, eukaryotic system, and it also permits the identification of non-substrate binding proteins of FIH, which may be equally important as substrates for FIH's biological role(s). The sections that follow detail the results of this Y2H screen (it should be noted that the screen itself was performed as part of my Honours project, and that this PhD project commenced with identification of the potential positives generated by the screen) and basic characterisation of the candidate interacting proteins as substrates and interacting proteins of FIH.

### 3.2 Y2H potential positive identification

As part of my Honours project, a Y2H assay using full length human FIH as bait was performed, with a view to isolating as yet undiscovered interacting partners of FIH. The screen employed a mouse embryonic day 10.5 cDNA library, and generated a total of  $9.8 \times 10^5$  transformants, which in turn yielded a total of approximately 950 potential positive clones. Due to time constraints, the screen was not analysed further during the Honours project, thus, this thesis work commenced with identification of the prey constructs which permitted growth in the positive yeast clones. Initial restriction digestion and sequencing of a random selection of positive clones suggested that a disproportionately large number of colonies were transformed with prey vector encoding ferritin light chain (hereafter referred to as ferritin; data not shown). Consequently, to rapidly identify both

ferritin-containing and non-ferritin-containing positives amongst the remaining clones, a colony hybridisation screen was performed. Yeast clones were grown on complete medium plates, transferred to a nylon membrane, then lysed via a combination of freeze thawing and sequential incubation of the membrane with  $\beta$ -glucuronidase and NaOH. After ultraviolet crosslinking, the membrane was hybridised with a radiolabelled ferritin cDNA probe corresponding to the region encoding aa 117-183 and then exposed to autoradiograph film (Figure 3.1). Clones which did not test positive for ferritin cDNA were cultured, lysed for plasmid purification and the plasmid analysed by



**Figure 3.1 Colony hybridisation analysis of Y2H potential positive clones.** Yeast clones were transferred to nitrocellulose prior to lysis, UV cross-linking and overnight hybridisation with 15 ng/mL [ $^{32}$ P]-labelled ferritin cDNA. Washed filters were then exposed to film for 14 days, and clones lacking a positive ferritin signal were selected for further analysis by sequencing. A representative of 8 blots is shown.

sequencing. Of the approximately 50 clones found not to contain ferritin, 9 contained in frame cDNAs, representing 6 different cDNA constructs. Thus including ferritin, the Y2H returned 7 different potential positives (Table 3.1). cursory examination of the potential positives did not reveal any obvious trends in function of the encoded proteins, however of particular note, was the finding that 4 out of the 7, including Foetal Globin Inducing Factor (FGIF), Serine/threonine-protein phosphatase 6 regulatory ankyrin repeat subunit B (PP6-ARS-B), Protein phosphatase 1 regulatory subunit 12C (PP1R12C) and NF-kappa-B inhibitor alpha ( $\text{I}\kappa\text{B}\alpha$ ), contained an ankyrin repeat domain (ARD) structural motif (Figure 3.2A, and see Figure 3.2B for an explanation of the ankyrin repeat structure). While this work was being carried out, studies in our laboratory had simultaneously identified the

Potential Positive	Region isolated (aa)	No. clones isolated	Primary function	Ankyrin repeats
Fetal globin-inducing factor (FGIF)	153-238	1	Gamma globin inducer <sup>a</sup>	Yes
Protein phosphatase 1 regulatory subunit 12C (PP1R12C)	242-782	2	PP1c regulatory subunit <sup>b</sup>	Yes
NF-kappa-B inhibitor alpha ( $\text{I}\kappa\text{B}\alpha$ )	136-314	2	Inhibitor of NF $\kappa$ B <sup>c</sup>	Yes
Serine/threonine-protein phosphatase 6 regulatory ankyrin repeat subunit B (PP6-ARS-B)	341-486	1	Putative PP6 regulatory subunit <sup>d</sup>	Yes
4E-Transporter (4E-T)	Full length	2	eif4E binding protein <sup>e</sup>	No
WD repeat-containing protein 62 (WDR62)	1131-1523	1	Spindle orientation, centrosome integrity <sup>f</sup>	No
Ferritin light chain	Full length	~ 900	Iron storage <sup>g</sup>	No

a (Yang et al., 2001)

b (Tan et al., 2001)

c (Li and Verma, 2002)

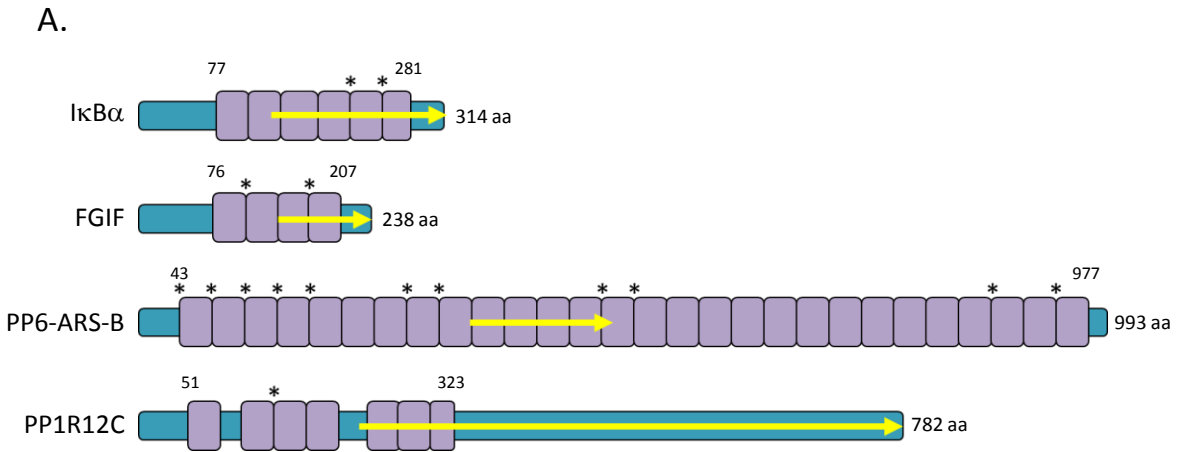
d (Stefansson et al., 2008)

e (Dostie et al., 2000)

f (Bogoyevitch et al., 2012)

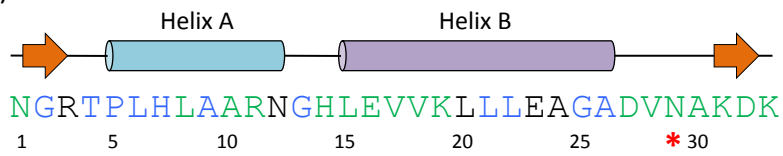
g (Wang et al., 2006)

**Table 3.1 In-frame cDNAs isolated from the Y2H screen.**

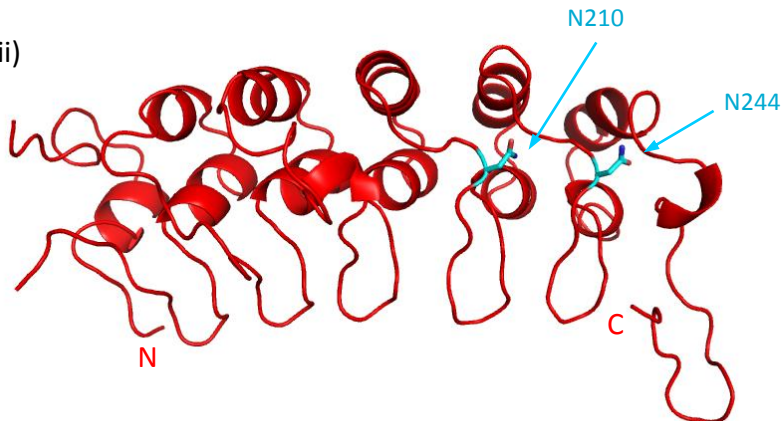


**B.**

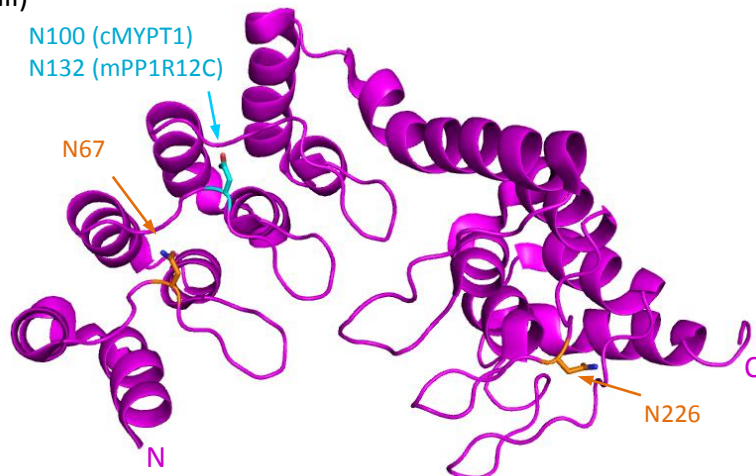
(i)



(ii)

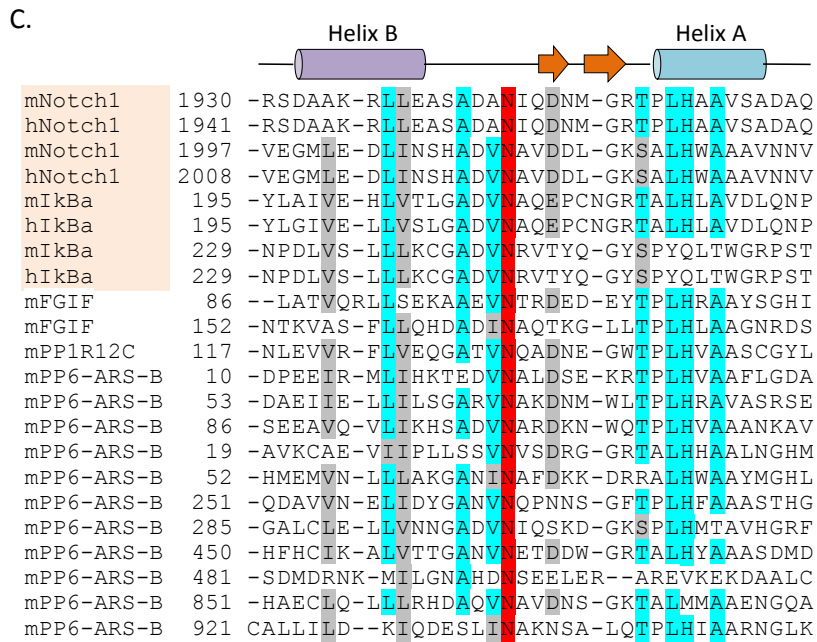


(iii)



(Figure 3.2)



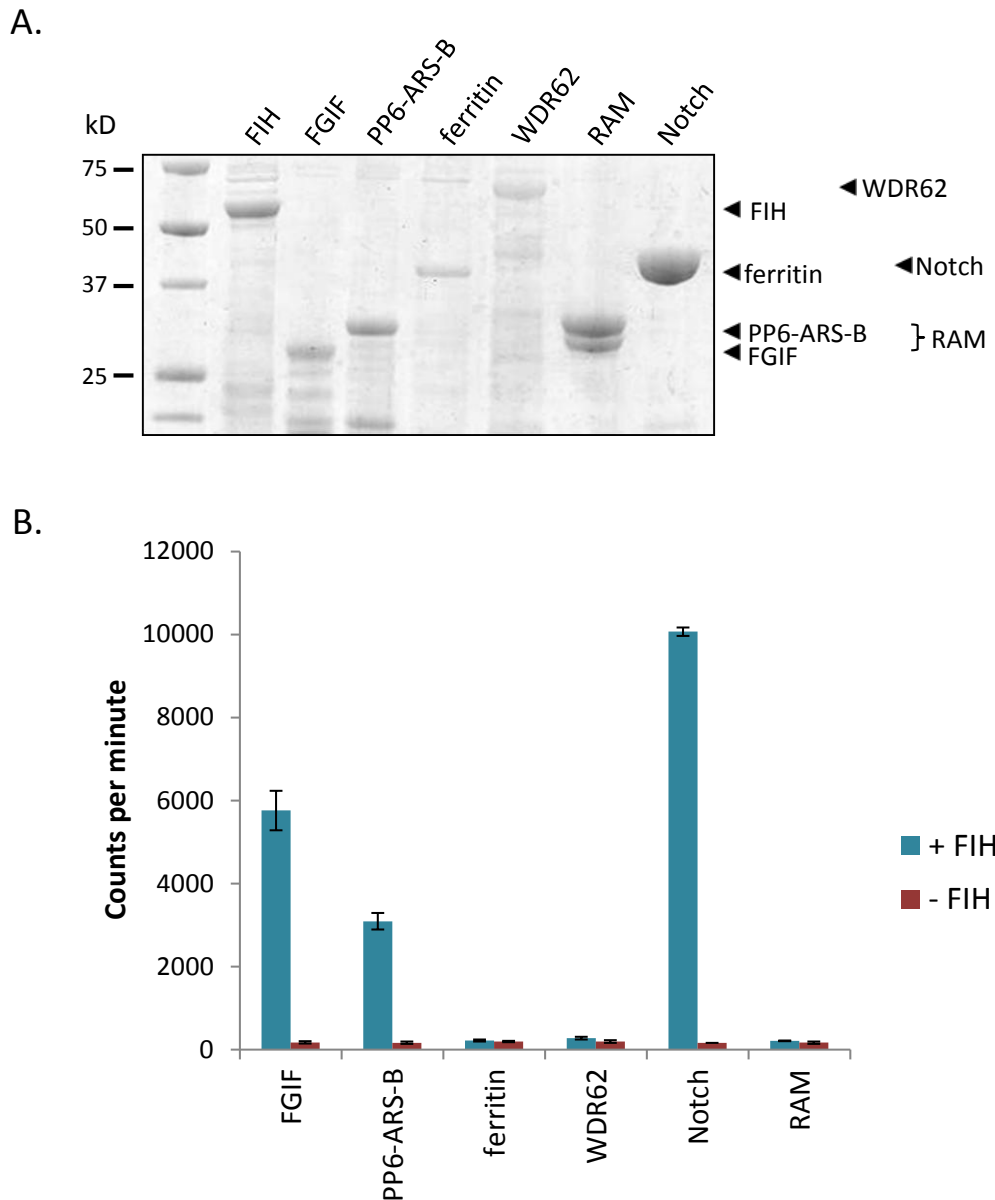


**Figure 3.2 ARD-containing proteins are among positives isolated in the Y2H screen.** (A) Schematic representation of the 4 ARD-containing proteins isolated in the screen. Purple rectangles represent individual ankyrin repeats within the domains, while yellow arrows signify regions of each protein isolated in the yeast clones. For each protein, ARD sequences were compared with known ARD substrates of FIH to locate appropriately positioned asparaginyl residues in the long loops between ankyrin repeats that may serve as targets of FIH (asterisks, and see also part (C)). Start and end amino acid numbers for each ARD are indicated above each protein. (B) (i) The 33aa ankyrin repeat consensus sequence according to Mosavi et al is depicted, with strongly conserved aa shown in blue, semi-conserved in green, and non-conserved in black (Mosavi et al., 2004). Amino acid (aa) number is shown below the consensus, while the typical secondary structure of the repeat, consisting of 2 helices (cylinders) and a long loop containing a  $\beta$ -turn (arrows) is shown above. The aa of the repeat typically hydroxylated by FIH is shown with a red asterisk. (B)(ii) and (iii) The crystal structures of human  $\text{I}\kappa\text{B}\alpha$  (ii) and chicken MYPT1 (a protein with a high degree of structural and functional homology to PP1R12C) (iii) ARDs are shown, with target asparaginyl residues, either known (for  $\text{I}\kappa\text{B}\alpha$ ) or predicted (PP1R12C; the position of the likely target in PP1R12C is shown mapped onto the MYPT1 structure) shown in blue. Note: 3 asparaginyl residues in MYPT1, including a residue positioned equivalently to Asn132 in mouse PP1R12C, are hydroxylated in MYPT1 (shown in orange). (C) Alignment of possible target asparaginyl residues (highlighted red) in Y2H positives which are equivalently positioned to those found in the known substrates,  $\text{I}\kappa\text{B}\alpha$  and Notch1 (highlighted in orange at the top of the alignment). The approximate position of sequences relative to helix A and B and the long loop of the canonical ankyrin repeat structure is indicated above the alignment. Strongly or moderately conserved residues between repeats are highlighted in turquoise or grey, respectively. Aa numbers are shown to left of each sequence. Sequences aligned using Clustal W2 and shaded using the Boxshade server. Crystal structures were obtained from the PDB database, identifiers:  $\text{I}\kappa\text{B}\alpha$ , 1IKN; MYPT1, 1S70. Structures generated in Pymol version 1.3.

ARD-containing protein Notch1 as a novel substrate of FIH (later published in PNAS, (Zheng et al., 2008)), and in addition, a report was published in the literature detailing a similar finding for I $\kappa$ B $\alpha$  (Cockman et al., 2006), thus validating the results of the screen presented in this thesis. The identification of hydroxylated asparagines in these molecules immediately prompted a search for the existence of similarly positioned target asparagines in the newly isolated Y2H positives. As can be seen in Figure 3.2C, comparison of the screen-isolated ARD-containing proteins with the targeted sequences in I $\kappa$ B $\alpha$  and Notch1 revealed promising target asparaginyl residues in all of the ARD fragments isolated except for PP1R12C (it should be noted, however that PP1R12C contains a potential target asparaginyl residue in a region further N-terminal to that isolated in the Y2H). Given these intriguing findings, it was next sought to determine which of the isolated positives were, in fact, *in vitro* substrates of FIH.

### 3.3 ARD-containing proteins are common substrates of FIH

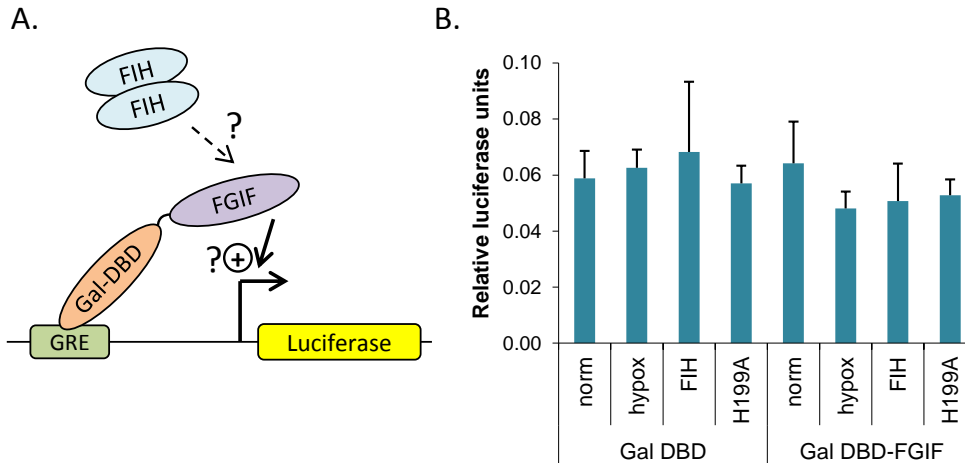
To assess the candidacy of the proteins as substrates of FIH, each of the cDNA fragments from the yeast positives (with the exception of I $\kappa$ B $\alpha$ ) was first cloned into the pET32a vector for protein expression in bacteria with an N-terminal Thioredoxin-6-Histidine tag. Test purifications revealed limited solubility and stability of FGIF (153-238), ferritin, WDR repeat protein 62 (WDR62) (1131-1523) and PP6-ARS-B (341-486), whereas eIF4E-Transporter (4E-T) and PP1R12C (242-783) were completely insoluble (Figure 3.3A and data not shown). Samples of soluble protein were then tested with bacterially expressed MBP-FIH by *in vitro* hydroxylation assay, with CO<sub>2</sub> released in the reaction a surrogate measure for levels of hydroxylation of each protein (Linke et al., 2007). Intriguingly, both FGIF (153-238), and PP6-ARS-B (341-486), stimulated CO<sub>2</sub> release significantly above background levels in the presence of FIH and cofactors (Figure 3.3B), suggesting that these proteins were genuine substrates of the enzyme. This finding, combined with the data regarding p105, I $\kappa$ B $\alpha$  and Notch1 as novel targets of FIH confirms the supposition made by our laboratory (Zheng et al., 2008) and others' (Cockman et al., 2006) that ARD-containing proteins comprise a novel class of substrates for FIH. Indeed, since these assays were performed, at least 11 other ARD-containing proteins have been validated as genuine targets of FIH by MS/MS analysis of bacterial-, cell- or tissue-derived ARD-containing proteins (see Table 1.1), putting this conclusion beyond doubt.



**Figure 3.3 Assessment of Y2H positive proteins as substrates of FIH by *in vitro* hydroxylation assay.** (A) Thioredoxin-6-histidine-tagged Y2H positive protein constructs were expressed and purified from bacteria using  $\text{Ni}^{2+}$  affinity chromatography and analysed for solubility and stability by SDS-PAGE followed by Coomassie staining. (B) Samples of each protein (equating to final concentrations in the reaction as follows: FGIF (153-238), 38.75  $\mu\text{M}$ , PP6-ARS-B (341-486), 38.75  $\mu\text{M}$ , Ferritin, 13.7  $\mu\text{M}$ , WDR62 (1131-1523), 27.75  $\mu\text{M}$ , mNotch1 (1862-2104; positive control), 38.75  $\mu\text{M}$ , RAM (mNotch1 1753-1847; negative control), 38.75  $\mu\text{M}$ , were incubated with 1  $\mu\text{M}$  bacterially expressed MBP-FIH and cofactors in an *in vitro* hydroxylation assay for 30 min at 37°C (B). Activity is expressed as the mean counts per minute generated by  $^{14}\text{C}$   $\text{CO}_2$  evolved during the reaction +/- standard deviation of 3 (for + FIH samples) or 2 (for - FIH samples) replicates. Assay is representative of 3 independent experiments.

Unlike the proteins containing ankyrin repeats, the two non-ARD-containing proteins tested in the hydroxylation assay (ferritin and WDR62) did not promote the production of CO<sub>2</sub> (Figure 3.3B), suggesting that, at least in the context of this *in vitro* assay, they are not FIH substrates. Nonetheless, it is important to note that bacterially expressed proteins may be poorly folded or lack appropriate post-translational modifications, and would additionally lack contact with accessory proteins which may be required for successful FIH-mediated hydroxylation. As a result, ferritin and WDR62 cannot be ruled out as substrates *in vivo*. In order to test 4E-T as a substrate (which is insoluble in bacteria as a full length protein), the protein coding sequence was divided into sections (4E-T (1-355), 4E-T (354-757) and 4E-T (756-983)) using conveniently placed restriction sites already present in the 4E-T cDNA, and the resulting fragments cloned into pET32a. Whilst these new constructs permitted the production of soluble protein, preliminary testing of the fragments by *in vitro* hydroxylation assay did not produce any activity above background (data not shown), thus providing no evidence that any of these fragments are substrates.

Given the identification of FGIF and PP6-ARS-B as novel FIH substrates, it was of interest to know if asparaginyl hydroxylation influenced either protein's function. At the time of their identification as substrates, PP6-ARS-B (then known as Ankyrin repeat domain 44) was still an uncharacterised ARD-containing protein. FGIF, on the other hand, had been identified as a factor capable of promoting  $\gamma$ -globin expression in human fetal erythroblast-mouse erythroleukemia (HFE-MEL) hybrid cells, and evidence was presented in the same paper to suggest that FGIF may have some capacity to function as a transcriptional coactivator (Yang et al., 2001). In this work, however, fusion of full length FGIF to a Gal DBD transcription factor showed no evidence of enhanced transcriptional activation relative to the Gal DBD alone in a Gal DBD response element-driven reporter gene assay (Figure 3.4). Since the poor functional characterisation of these proteins limited design of "ARD-specific" functional assays, it was decided instead to further characterise their interaction with FIH and assess the effects of this interaction on HIF-1 $\alpha$  transactivation activity, as it had been proposed that ARD-mediated sequestration of FIH may serve to regulate HIF- $\alpha$  CAD activity (Cockman et al., 2006). In addition, it was of interest to ascertain if the structural information provided by the novel substrates could better define FIH substrate recognition properties, therefore allowing us to improve bioinformatic FIH substrate prediction.



**Figure 3.4 Effect of FIH on transcriptional activity of Gal DBD-FGIF.** (A) Schematic representation of the reporter assay. (B) HEK 293T cells were transfected with a Gal response element-driven firefly luciferase reporter gene, renilla luciferase under the control of a constitutive promoter, Gal DBD alone or fused to full length FGIF, and V5-tagged FIH or FIH H199A (a catalysis-defective mutant). 4 hrs post transfection, cells were either left in normoxia (“norm” “FIH” and “H199A” samples), or placed in hypoxia (<1% oxygen, “hypox” sample) for 16 hrs prior to cell lysis and analysis by dual luciferase assay. Data is shown as relative luciferase units (RLU) +/- standard deviation for each sample triplicate.

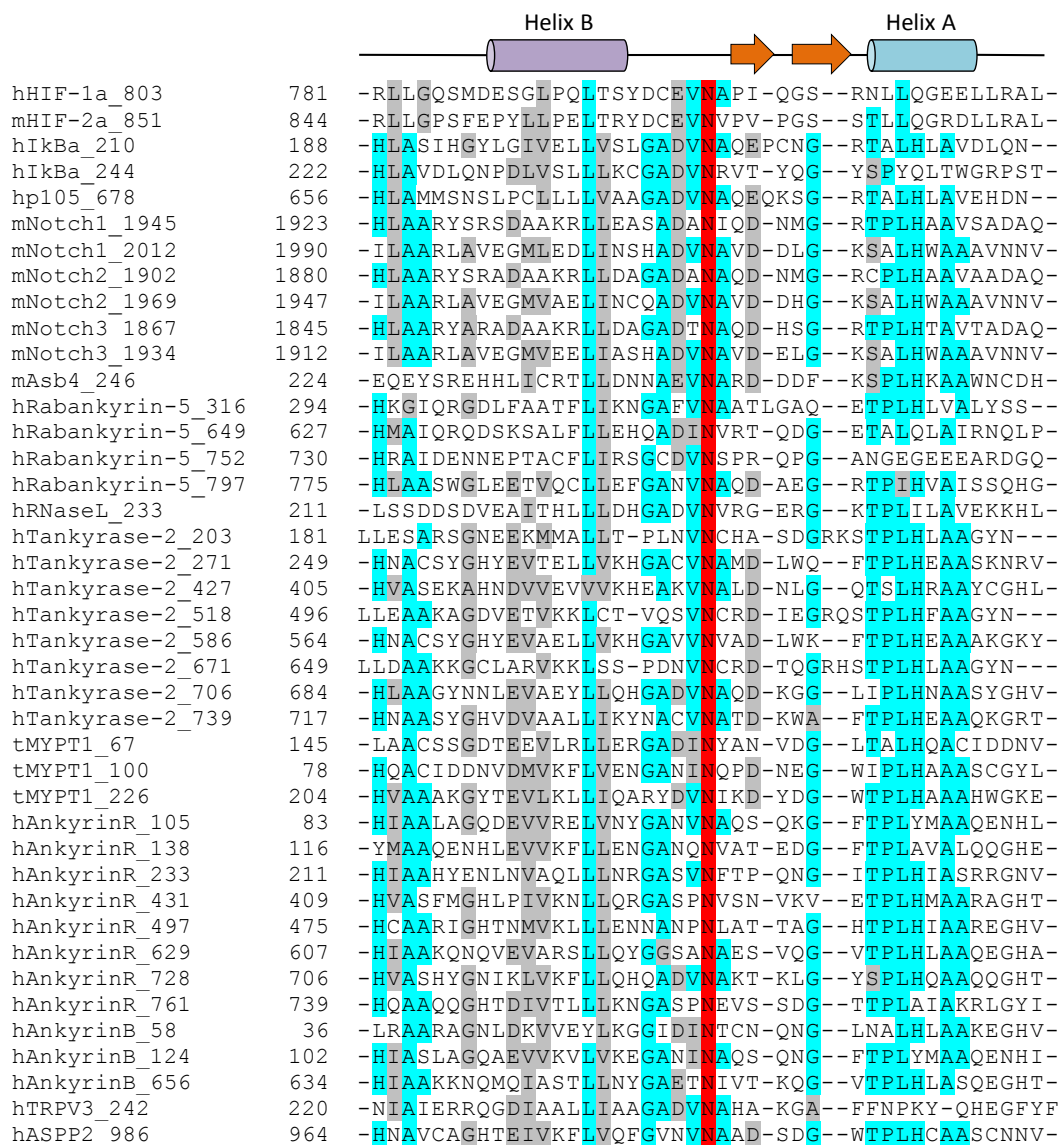
### 3.4 The FIH “consensus sequence” as a predictive tool for novel substrates

In order to determine the level of sequence specificity required by FIH for targeting substrates, the C-terminal region of HIF-1 $\alpha$  and HIF-2 $\alpha$ , as well as target-asparaginyl containing ankyrin repeats which have been verified in the literature as genuine targets of FIH via MS/MS analysis were aligned using Clustal W2 (Figure 3.5A). Whilst a significant number residues appear to be strongly conserved within the ankyrin substrates (Figure 3.5B), this may well reflect amino acid constraints typically found within the ankyrin repeat structural motif (depicted in the consensus below the WebLogo), rather than sequence requirements for FIH hydroxylation per se. However, closer inspection of the aligned sequences shows that a small number of residues are conserved across both HIF and ARD classes of substrates, including a leucine in the -8 position relative to the target asparagine, an acidic residue at the -2 position, and a hydrophobic residue at the -1 position (Figure 3.5C).

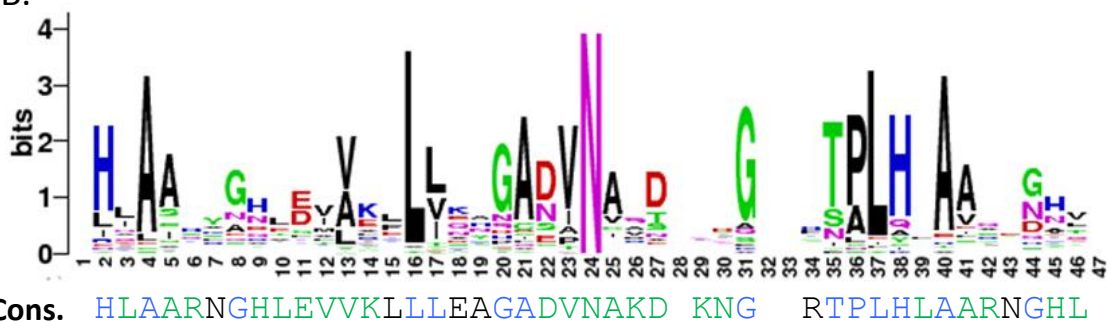
Although the sequence denoted in Figure 3.5C is clearly a preference rather than an absolute requirement for FIH-mediated hydroxylation, it was of interest to determine if this pattern of amino acids could be used as a predictive tool for finding more novel FIH substrates. As such, the sequence

{L-X(5)-[DE]-[IV]-N-[AV]} (where X(5) represents 5 aa of any identity, and letters within brackets denote alternative amino acids for a given position within the sequence) was used in a ScanProsite pattern search (<http://prosite.expasy.org/scanprosite/>) to isolate database protein sequences which contained this target motif. The search retrieved 404 matches in 383 mouse sequences, many of which (approximately 85%) were ARD-containing proteins (see Appendix 8.1). As ankyrin repeats were rapidly becoming well established as novel FIH substrates, it was decided to instead investigate the candidacy of some non-ARD containing proteins as FIH targets. Accordingly, two different proteins were selected from the list returned by the ScanProsite search based upon (1) their possible functional relevance to the phenotype of the FIH KO mouse, and (2) the likely accessibility, based on crystal structure evidence, of the target asparagine for hydroxylation by FIH. The first candidate, T cell protein tyrosine phosphatase (TCPTP), has been shown to regulate the intensity and duration of insulin receptor signalling in MEFs, and therefore could conceivably affect insulin sensitivity as is observed in the FIH KO mice (Galic et al., 2005). The second, G protein-coupled receptor kinase 2 (GRK2), is also believed to influence insulin sensitivity (Mayor et al., 2011), and additionally may have

A.



B.



C.

LXXXXXX[-]φN

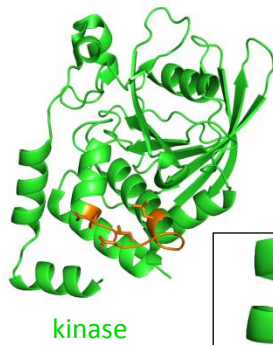
(Figure 3.5)

**Figure 3.5 Determination of a “preferred binding sequence” for FIH.** (A) FIH substrate ankyrin repeats reported in the literature which have been verified via MS/MS were aligned using Clustal W2. Strongly conserved amino acids, or moderately-conserved amino acids of a similar type are shown highlighted in turquoise or grey, respectively, whilst the target asparagine is shown in red. The secondary structure of the Notch1 protein is shown above the alignment. Lowercase “h”, “m” or “t” preceding protein names refers to human, mouse or turkey sequences, respectively. Numbers directly after protein names refer to the position of the target asparagine. Numbers before the alignment are amino acid numbers of the presented sequence. (B) Sequence logo representation of the aligned sequences. Letter stack heights correspond to the level of conservation at each position in the sequence, and individual letter heights within a stack indicate the relative frequency with which each amino acid occurs. The ankyrin repeat consensus as reported by Mosavi et al (2004) is shown below the logo, with strongly conserved amino acids shown in blue, semi-conserved in green, and non-conserved in black. A single 33 aa ankyrin repeat within the consensus is underlined. *Sequence logo generated using the WebLogo tool at <http://weblogo.berkeley.edu/logo.cgi>.* (C) Sequences from the alignment in (A) were analysed for the presence of residues which were strongly conserved between both ankyrin and HIF classes of substrates. These residues are shown using the amino acid nomenclature proposed by Aasland et al (2002), wherein [-] and  $\phi$  represent negatively charged and hydrophobic amino acids, respectively (Aasland et al., 2002). The target asparagine is shown in red.

the ability to modulate cardiovascular effects, such as heart rate, via its phosphorylation of G-protein-coupled receptors (GPCRs) (Lodowski et al., 2003). The matching motifs generated by the ScanProsite search are shown mapped onto the relevant crystal structures in Figure 3.6A and B, with the motif for TCPTP located on a surface-exposed extended loop (Iversen et al., 2002), and that for GRK2 found on a flexible loop believed to regulate nucleotide entry into GRK2’s catalytic pocket (Lodowski et al., 2003).



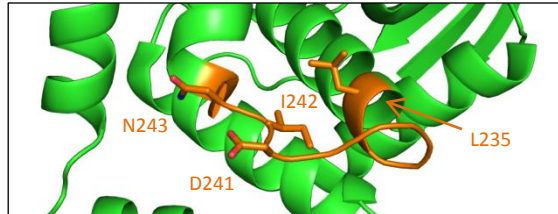
A.  
(i)



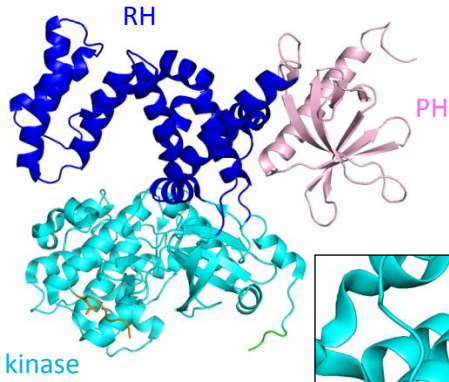
kinase

(ii)

mTCPTP	235	LmekgeD	VNV
hTCPTP	235	Lmekgd	DINI



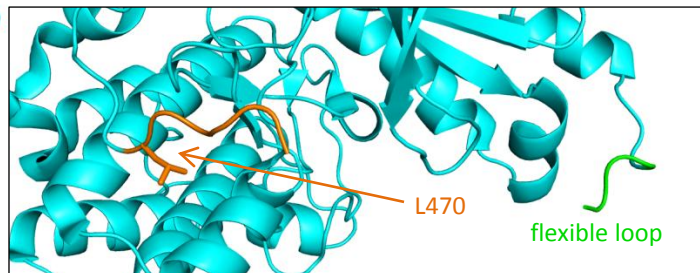
B.  
(i)



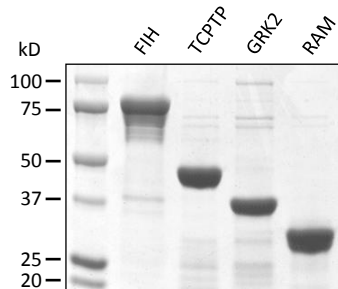
kinase

(ii)

mGRK2	470	LipprgEV	NA
hGRK2	470	LipprgEV	NA
bGRK2	470	LipprgEV	NA



C.



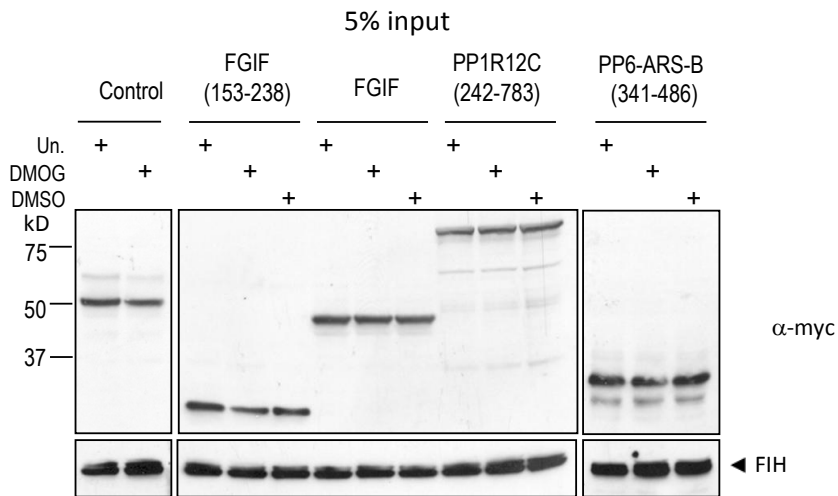
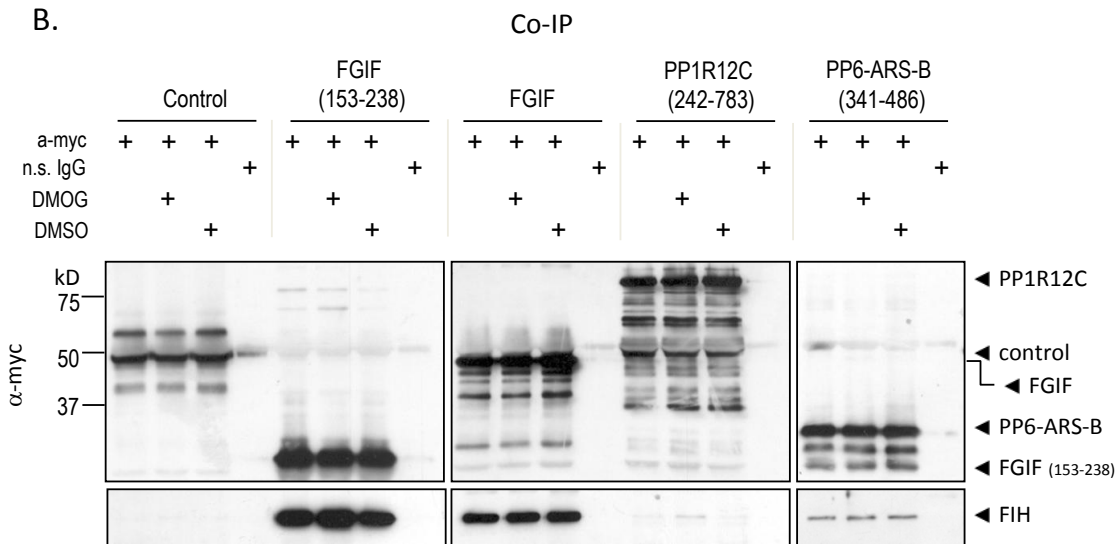
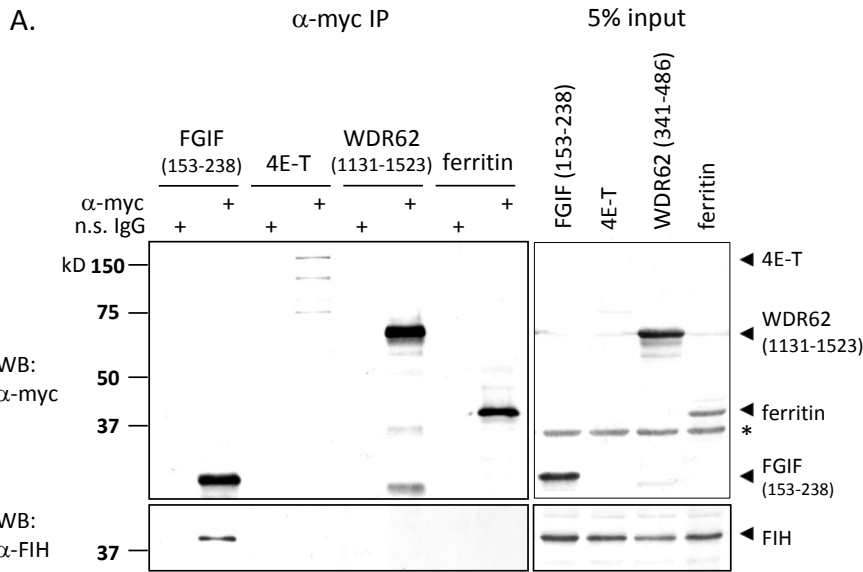
(Figure 3.6)

**Figure 3.6 TCPTP and GRK2 contain a surface accessible predicted FIH target motif.** (A) The TCPTP FIH target motif found by the ScanProsite database search tool (<http://prosite.expasy.org/scanprosite/>; motif L-X(5)-[DE]-[IV]-N-[AV]) is shown (i) in orange mapped onto the human TCPTP kinase domain crystal structure (and magnified in the inset), and (ii) in text form, with conservation across mouse and human sequences indicated with turquoise highlighting. Capital letters indicate amino acids which were constrained during the database search. (B) The FIH target motif for GRK2 is, as for (A), (i) mapped onto the bovine crystal structure of GRK2, with a zoomed-in version shown in the inset, and (ii) shown in text form aligned with human and bovine GRK2. Note: some of the motif is present on a flexible loop (the edge of which is shown in green in the inset) for which no electron density was observed in the structure. The PH domain which was utilised in order to obtain a soluble bacterially expressed GRK2 protein fragment containing the FIH target motif (see main text) is shown in pink in (i). (C) MBP-FIH and Thioredoxin-6His-tagged TCPTP 1-312, GRK2 455-660, and Notch RAM domain (a negative control substrate) were expressed in bacteria and purified by amylose (FIH) or Ni<sup>2+</sup> resin. 8 µg of each were then analysed by SDS-PAGE and Coomassie staining prior to utilisation of the proteins in a hydroxylation assay to assess TCPTP and GRK2 as substrates of FIH (data not shown).

To test the proteins as substrates of FIH, a region of the protein containing the target motif (aa 1-312 for TCPTP, aa 186-525 for GRK2, both of which comprise the catalytic domains of both enzymes), was selected for amplification by PCR and cloned into the pET32a expression vector. Initial attempts to express and purify the kinase domain of GRK2 were hampered by the complete insolubility of the protein, thus a new construct was prepared (aa 455-660) incorporating the potential FIH target site and regions further C-terminal to the motif, including the GRK2 PH domain, which has previously been shown to be soluble post bacterial expression (Yang et al., 2003). Soluble protein was then successfully generated for both TCPTP as well as the new GRK2 construct (Figure 3.6C), and they were subsequently tested with bacterially expressed FIH by *in vitro* hydroxylation assay. However, neither of the constructs were observed to promote evolution of CO<sub>2</sub> by FIH (data not shown), suggesting that the proteins are unlikely to be substrates. Nonetheless, it cannot be ruled out that post-translational modifications applied or accessory proteins present in a cellular context may be required for FIH-mediated hydroxylation to take place. Together, the results likely reflect the importance of tertiary structure context in substrate selection by FIH. However, even GRK2 (aa 455-660), which, based on the crystal structure, would form a folded PH domain attached to an unfolded loop containing the target asparaginyl residue, was not hydroxylated by FIH. This is a somewhat surprising result given that the FIH substrate, the HIF-1α CAD, is documented to be unfolded in solution, suggesting that FIH can recognise target sequences in this form. Overall, these data reiterate that there are complex substrate selection criteria for FIH which are yet to be fully understood.

### 3.5 Both substrate and non-substrate Y2H positives interact with FIH

While *in vitro* hydroxylation assays have demonstrated only a subset of the positives isolated in the Y2H could be shown to be substrates of FIH, this naturally does not rule out the possibility that the remainder may participate in other sorts of functional interaction with FIH. To test this idea, HEK293T cells were transfected with pEF-6myc potential positive plasmid constructs and then extracts prepared for use in co-immunoprecipitation (co-IP) assays. In preliminary experiments, endogenous FIH was effectively co-IPed with 6myc-FGIF (153-238), a result which was not replicated by a type-matched non-specific antibody (Figure 3.7A). However, under these conditions, none of the non-ARD containing proteins tested, including 4E-T, WDR62 (1131-1523) and ferritin, showed any interaction with FIH, although it should be noted that 4E-T input levels were significantly lower than that of the other proteins examined in this initial assay. In similar experiments, both a 6myc-tagged full length FGIF construct and 6myc-PP6-ARS-B (341-486) were able to co-IP FIH with varying efficiency, and in the case of FGIF (153-238), the interaction was slightly enhanced by addition to the cell culture media of the FIH inhibitor, DMOG (dimethyloxallylglycine) (Figure 3.7B, and published in (Linke et al., 2007)). In contrast, pulldown of 6myc-PP1R12C was only able to enrich for detectable amounts of FIH after treatment of the cells with DMOG (Figure 3.7B). Unexpectedly, addition of DMOG was also able to promote a very weak interaction between FIH and the non-ARD-containing proteins 4E-T and WDR62 (1131-1523), despite these proteins having shown no activity in hydroxylation assays with FIH (Figure 3.8A). Exactly why this would be the case is not clear. It is possible that N-OG binding (DMOG is processed into N-OG in cells) to FIH may somehow promote a structural conformation which is non-specifically “sticky”. Alternatively, it is conceivable 4E-T or WDR62 may be a binding partner of an FIH substrate, such that the addition of DMOG may promote binding of this substrate, and therefore may also promote capture of WDR62/4E-T, albeit weakly due to the secondary nature of the interaction.

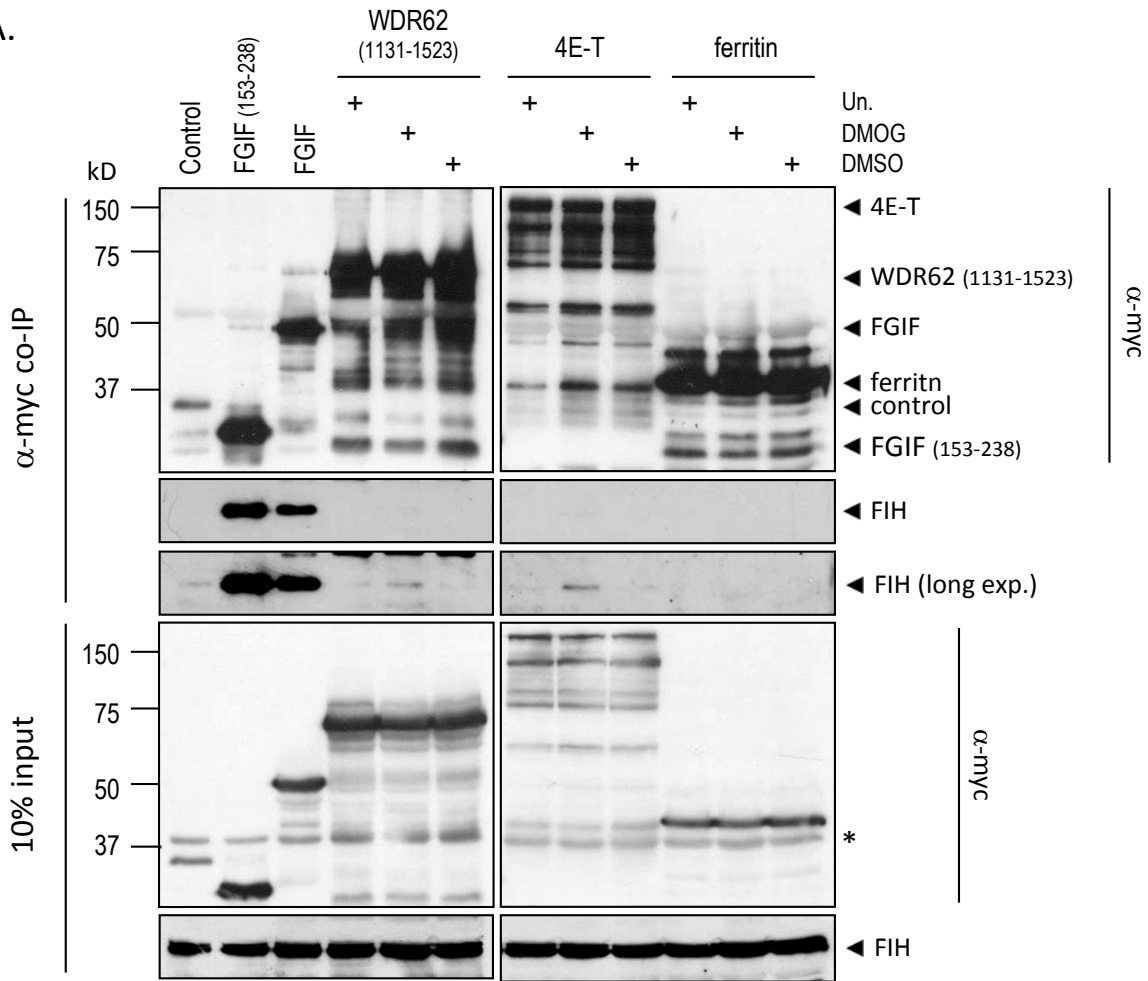


(Figure 3.7)

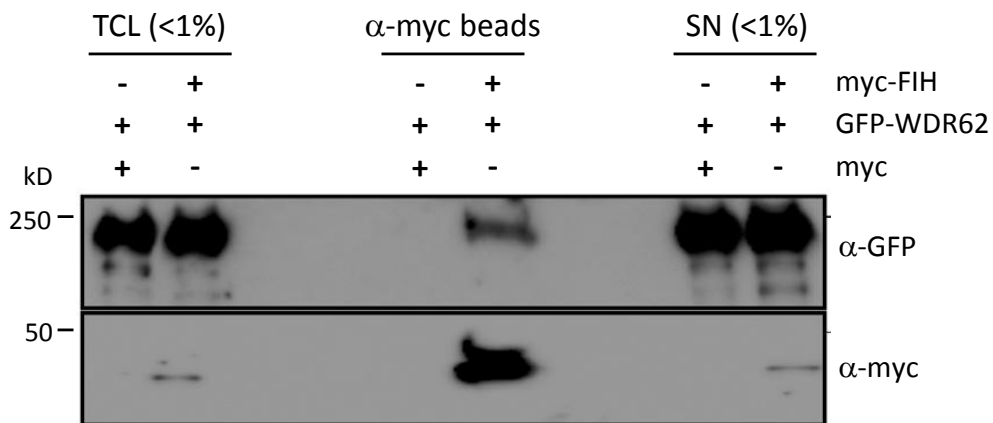
**Figure 3.7 Co-IP of endogenous FIH with Myc-tagged Y2H positives.** (A) HEK 293T cells were transfected with pEF-6myc expression plasmids encoding a selection of 6-myc tagged Y2H positive constructs, and cell lysates were subjected to IP using 9E10 anti-myc Ab or a type matched negative control Ab. IP efficiency was determined by blotting with 9E10 Ab, while co-IP of endogenous FIH protein was assessed by western blot using #8 polyclonal FIH antibody. (B) HEK 293T cells were transfected with pEF-6myc plasmids encoding 6myc-tagged ARD-containing proteins or a 6myc-tagged control protein (mouse ribosomal protein S2) and then left untreated, or treated with DMSO or 1mM DMOG for 20 hrs. Cell lysates were then prepared and 6myc-tagged proteins IPed using 9E10 anti-myc Ab, or subjected to a control IP using a non-specific type matched Ab. IP of myc-tagged proteins and co-IP of endogenous FIH was assessed as for (A). Experiment representative of 2 independent experiments.

The inefficient nature of the interaction between FIH and WDR62 (1131-1523) also led to the hypothesis that a full length construct of WDR62 might display a stronger interaction with FIH than just the C-terminus alone. To test this idea, a plasmid encoding Myc-6his-tagged FIH was sent to the laboratory of Marie Bogoyevitch at the University of Melbourne, wherein the construct was employed in a co-IP experiment with GFP-tagged full length WDR62. As can be seen in Figure 3.8B,  $\alpha$ -myc antibody-mediated pulldown of FIH resulted in the co-IP of GFP-WDR62, although, as for the C-terminal fragment of WDR62, the interaction was of low efficiency, with less than 1% of input GFP-WDR62 being co-IPed (compare this with the FIH-FGIF interaction (Figure 3.7B and Figure 3.8A), where >5% of input FIH was co-IPed by the  $\alpha$ -myc antibody). Even so, a significant functional effect for this interaction cannot be ruled out, hence important future work will be to determine if the interaction is direct (for example, is the N-terminus of WDR62 a substrate?), or if FIH and WDR62 may indirectly interact through a 3<sup>rd</sup> party accessory protein to regulate an as-yet unidentified cellular process.

A.



B.



Part (B) performed by Nicholas Lim Rui Yuan.

(Figure 3.8)

**Figure 3.8 Testing of non-ARD-containing proteins for interaction with endogenous FIH.** (A) Co-IPs and analysis of FIH pulldown were performed as for Figure 3.7B, except that the control plasmid encoded the 6myc tag upstream of an out of frame (OOF) coding sequence of murine Adiponectin receptor1 (hereafter referred to as AdipoR1OOF; see Materials and methods for details), and the non-specific Ab control was not included. A long exposure of the  $\alpha$ -FIH blot was required in order to detect the low efficiency interactions between FIH and 4E-T/WDR62 (1131-1523). Experiment representative of 2 independent experiments. (B) Assessment of the interaction of FIH with full length WDR62. Plasmids encoding Myc-6His-FIH and GFP-WDR62 were co-transfected into AD293 cells, followed by cell lysate preparation and IP with  $\alpha$ -myc antibodies. Co-IP was assessed by western blot with  $\alpha$ -GFP (for WDR62) antibodies. Note: the experiment in part (B) was performed by Nicholas Lim Rui Yuan of the Bogoyevitch Lab. *WB = western blot; TCL = total cell lysate; SN = depleted supernatant from IP incubation. Asterisks indicate non-specific background bands.*

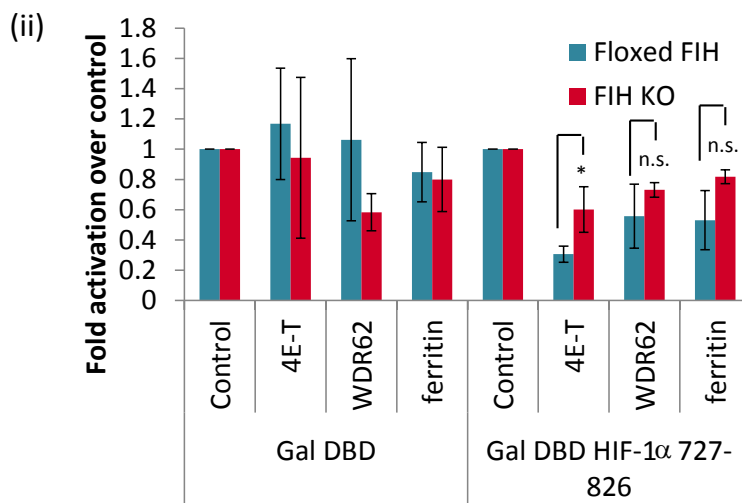
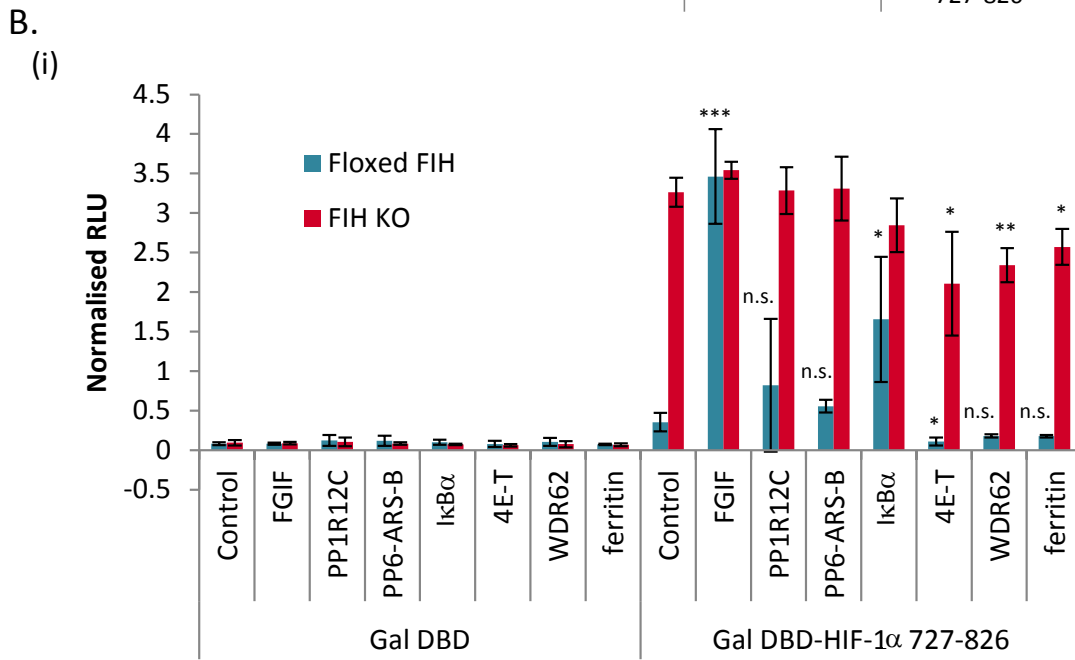
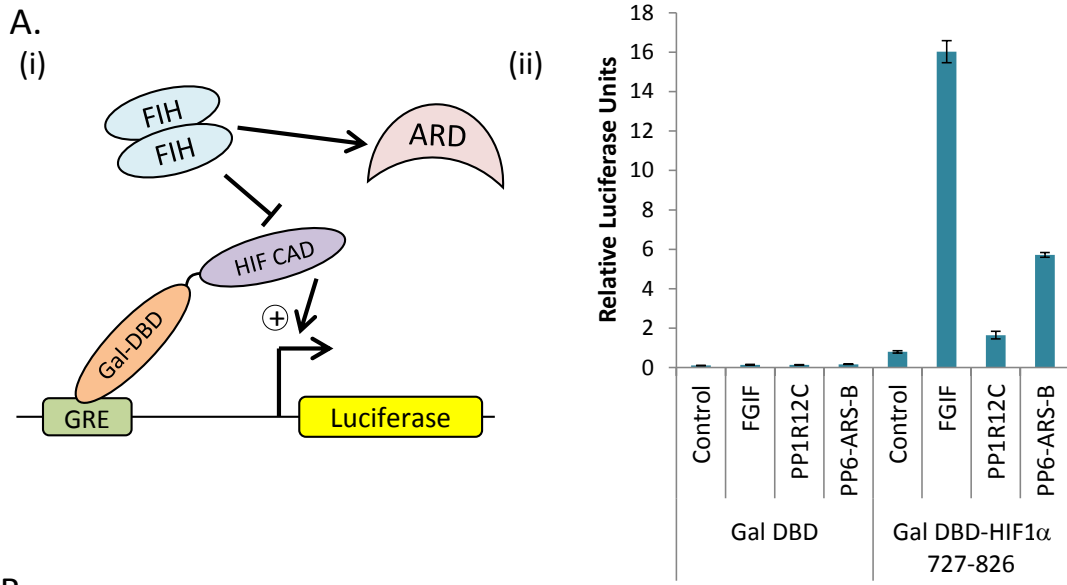
### 3.6 Effect of ARD protein over-expression on FIH repression of the CAD

To verify that the FIH-substrate interactions observed above occurred in a cellular context, advantage was taken of the well-characterised interaction between FIH and the CAD of HIF-1 $\alpha$ . Proteins which are substrates of FIH, should, when over-expressed, be able to compete with HIF for hydroxylation by FIH (Figure 3.9A(i)). The effectiveness as a competitor can then be measured through the ability of the CAD to drive production of a reporter gene, with more effective competitors producing greater derepression of the CAD. Accordingly, each of the ARD-containing Y2H positives were over-expressed in HEK 293T cells in the presence of a Gal DBD luciferase reporter driven by Gal DBD-HIF- $\alpha$  CAD (727-826). As seen in Figure 3.9A(ii), each of the ARD-containing potential positives found to show activity in the *in vitro* hydroxylation assay also generated derepression of the CAD. In fact, even PP1R12C, which did not show activity in the hydroxylation assay, also produced a small increase in reporter gene activity, which is suggestive of a low efficiency interaction with FIH, and agrees with the co-IP result in Figure 3.7B. To demonstrate that these increases in reporter activity were dependent on FIH, the experiment was repeated in FIH knockout mouse embryonic fibroblasts (MEFs), with resultant activity of the reporter compared to the effect in FIH floxed MEFs (Zhang et al., 2010). As seen in HEK 293T cells, over-expression of FGIF and PP6-ARS-B (341-486) in the FIH floxed MEFs increased expression of the reporter gene to varying extents (although the increase did not quite

reach statistical significance for PP6-ARS-B (341-486);  $p = 0.069$ ) as did known substrate I $\kappa$ B $\alpha$  (136-314) (Figure 3.9B(i)). PP1R12C (242-786) expression, however, had little effect, although this likely reflects the poor transfection efficiency of these cells compared to HEK 293Ts, which may mask small increases in reporter activation. Notably, these variations in reporter activation cannot be entirely explained by differences in expression level of the different myc-tagged constructs. As can be seen in Figure 3.9B(iii), I $\kappa$ B $\alpha$  (136-314) and PP6-ARS-B (341-486) expression levels were very similar by western blot, but I $\kappa$ B $\alpha$  (136-314) generates a 3-fold greater increase in reporter activity compared to PP6-ARS-B (341-486), suggesting that the I $\kappa$ B $\alpha$  fragment has a greater efficiency of interaction with FIH. Importantly, in the MEFs lacking functional FIH, over-expression of FGIF, I $\kappa$ B $\alpha$  (136-314) and PP6-ARS-B (341-486) did not produce any further increases in luciferase production (Figure 3.9B(i)), suggesting that inhibition of FIH is responsible for the effects seen in the FIH floxed MEFs.

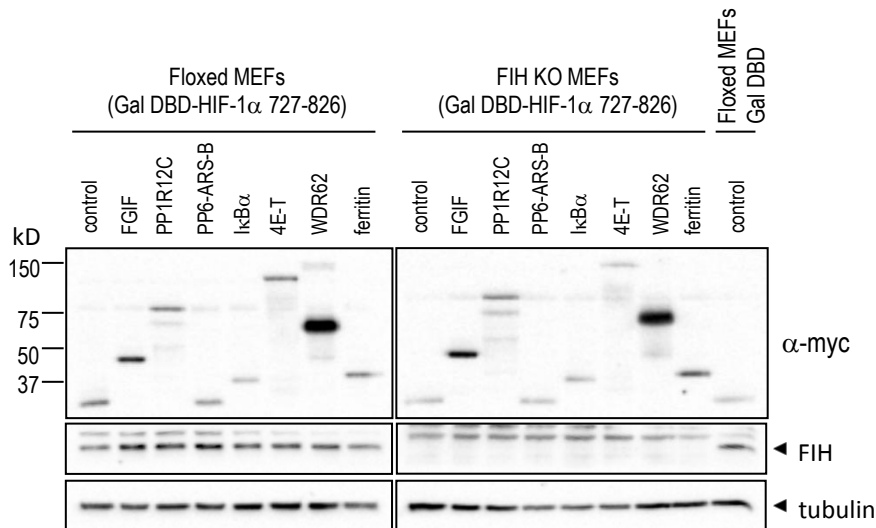
A similar activation of HIF- $\alpha$  CAD-driven reporter activity has been demonstrated following over-expression of a number of other ARD-containing proteins (Coleman et al., 2007; Webb et al., 2009), and it is likely that many as yet untested proteins containing the ARD structural motif (and suitably located asparaginyl residues) have a similar ability to compete with HIF for FIH binding. Indeed, a 2010 paper from Schmierer et al suggested that 166 ankyrin repeats in 105 human proteins contain the “FIH preferred” sequence, L<sup>-8</sup> X4 A/C<sup>-3</sup> D/E/N<sup>-2</sup> I/V<sup>-1</sup> N (Schmierer et al., 2010). However, exactly how many of these candidate FIH substrates are truly capable of being recognised by FIH is a much more difficult question to answer. In this respect, it should be reiterated that understanding of the “rules of engagement” for FIH and its substrates is still limited. For example, although a preferred binding sequence for FIH has been established (see section 3.4), the tertiary structure that this sequence adopts is quite different in HIF-1 $\alpha$  (disordered in solution (Dames et al., 2002)) compared to ARD-containing proteins (which adopt the classical ankyrin fold). One might anticipate,





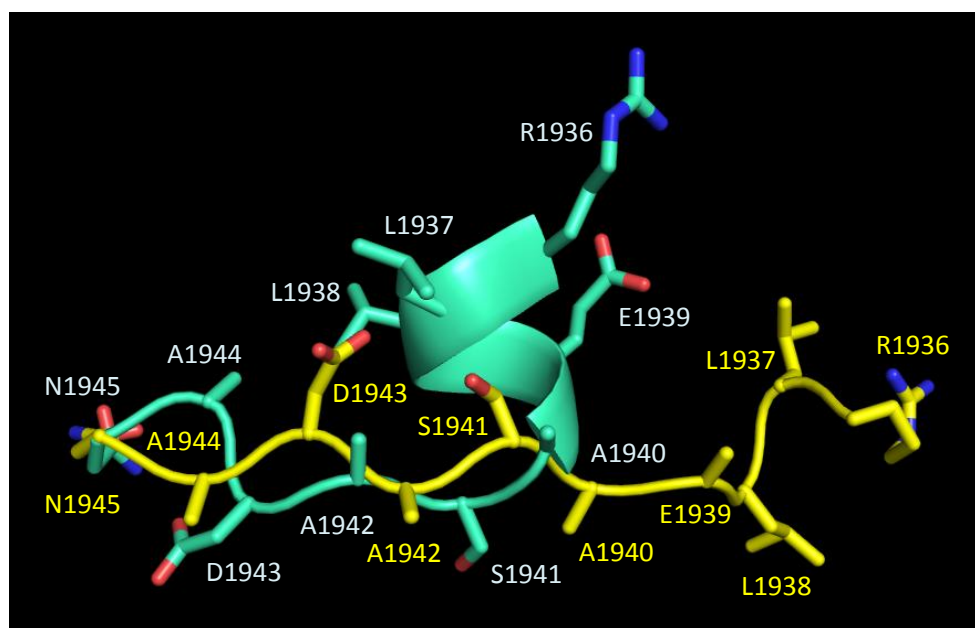
(Figure 3.9)

B.  
(iii)

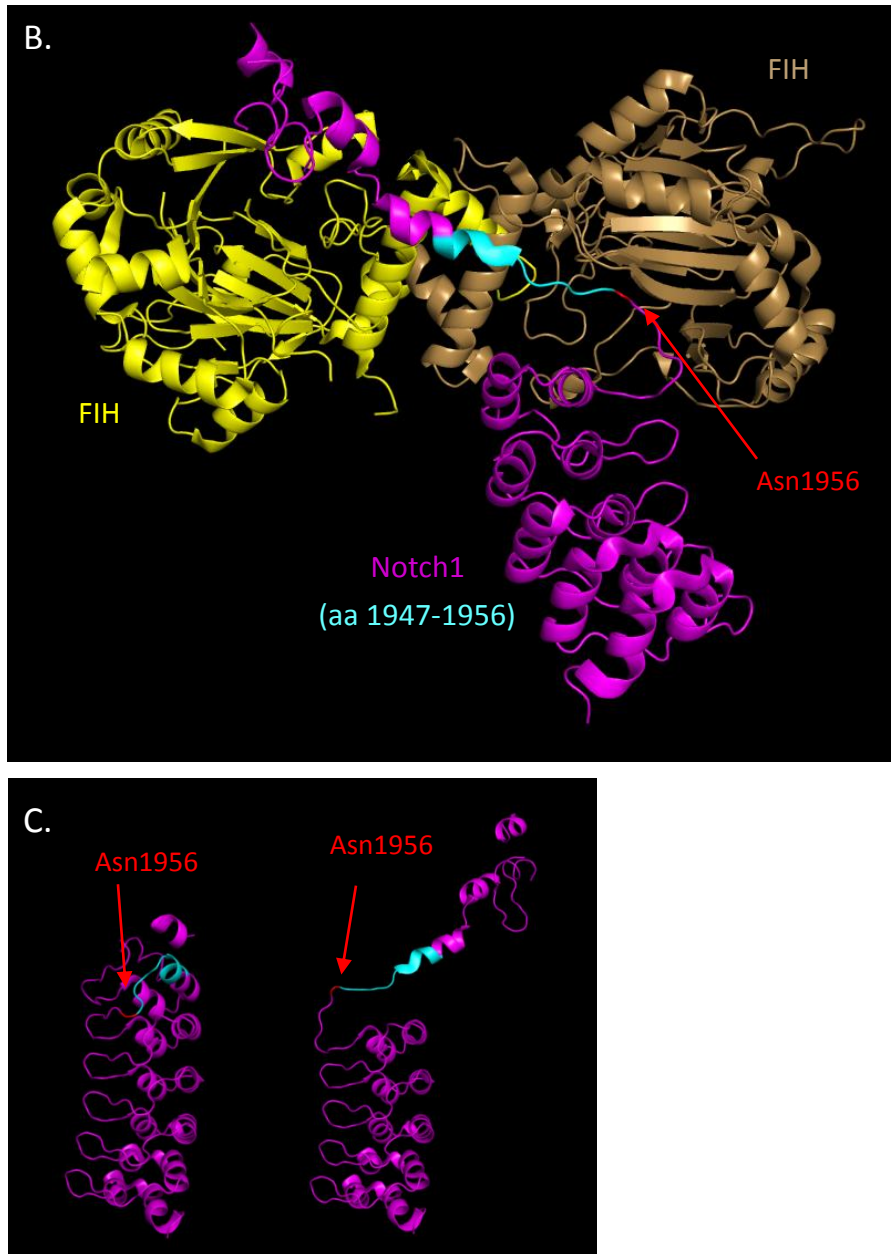


**Figure 3.9 Effect of Y2H positive over-expression on HIF- $\alpha$  CAD repression.** (A) HEK 293T cells were transfected with pEF-6myc expression plasmids encoding 6myc-tagged ARD-containing potential positives (full length FGIF, PP1R12C (242-782), and PP6-ARS-B (341-486)) or a myc-tagged control (Adiponectin receptor1 coding sequence out of frame (OOF) with the myc-tag, hereafter referred to as AdipoR1OOF; see Materials and methods for details), a Gal response element-driven firefly luciferase reporter gene, renilla luciferase under the control of a constitutive promoter, and Gal DBD alone or fused to HIF-1 $\alpha$  (727-826). Proteins which compete with the HIF-1 $\alpha$  CAD for FIH-mediated hydroxylation are anticipated to derepress the CAD and up-regulate reporter gene expression, diagrammatically represented in (i). (ii) Post transfection, cells were incubated for 20 hrs prior to cell lysis and analysis by dual luciferase assay. Data is shown as relative luciferase units (RLU) +/- standard deviation for each sample triplicate. Figure is representative of 2 independent experiments. (B) (i) The experiment in (A), with the additional inclusion of expression plasmids encoding 6myc-tagged  $\text{I}\kappa\text{B}\alpha$  (136-314), 4E-T, WDR62 (1131-1523) and ferritin, was performed in cells containing or lacking FIH (floxed FIH MEFs or FIH KO MEFs, respectively). For each individual experiment, RLU for each sample were divided by the mean of all RLU sample values across the experiment, then the average of these values graphed +/- SEM. Graph is the combined data of 3 independent experiments. p values calculated using a Student's 2-tailed t-test, where floxed FIH samples are compared to the floxed FIH control, and FIH KO samples are compared with the FIH KO control. \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001, n.s. not significant (p > 0.05). (ii) Data from the 4E-T, WDR62 and ferritin samples in part (i) are shown as fold change in RLU compared to the control +/- SEM to permit a direct comparison of repression of firefly luciferase activity in floxed FIH vs FIH KO cells. Graph is the combined data of 3 independent experiments. p values calculated for the Gal DBD-HIF-1 $\alpha$  (727-826)-containing samples using a Student's 2-tailed t-test comparing 4E-T Floxed FIH with 4E-T FIH KO, WDR62 Floxed FIH with WDR62 FIH KO, and ferritin floxed FIH with ferritin FIH KO. \* p < 0.05, n.s. not significant (p > 0.05). (iii) Cell lysates from (B)(i) were separated by SDS-PAGE and then analysed for myc-tagged positive and FIH expression levels using  $\alpha$ -myc (4A6, Millipore) and  $\alpha$ -FIH (Novus) Abs. Equal loading was assessed using  $\alpha$ -tubulin Ab (Sigma). Western representative of 2 blots performed for different experiments.

then, that substrate peptides derived from HIF-1 $\alpha$  versus an ARD-containing protein would display distinct binding mechanisms to FIH. However, crystallisation of FIH in complex with a substrate peptide from Notch1 argues against this, as the conformation of the peptide is virtually indistinguishable from that seen with a HIF- $\alpha$  CAD-derived peptide (Coleman et al., 2007). Whether or not a peptide would bind in the same manner as the full ARD is unknown, as attempts to generate a crystal structure with FIH and the full Notch1 ARD were unsuccessful (Coleman et al., 2007). Nonetheless, if the mode of binding demonstrated by the peptide is indeed faithful to that adopted by the whole molecule, this would have considerable consequences for the FIH:Notch1 interaction. Most notably, the ARD's modular structure would need to be unfolded in dramatic fashion



(Figure 3.10)



**Figure 3.10 Comparative structure of a human Notch1 target sequence when bound to FIH or as part of the Notch1 ARD.** (A) Overlay of the crystal structures of Notch1 (1936-1945) in the context of an FIH-bound peptide (yellow), or as part of the folded Notch1 ARD (turquoise). Red and blue colouring denote oxygen and nitrogen atoms, respectively. (B) Theoretical diagram to illustrate the unfolding of the Notch1 ARD required to align the Notch1 Asn 1956 target site (equivalent to mNotch1 Asn1945) in the active site of one monomer of the FIH dimer. The region of the ARD which corresponds to the Notch1 peptide crystallised with FIH is shown in turquoise. The target asparaginyl residue is shown in red. The Notch1 ARD, and the two monomers of the FIH dimer are shown in purple, and yellow and brown, respectively. (C) The structure of folded (left structure) and “unfolded” (FIH-bound) (right structure, from part (B)) Notch1 ARD. Structures were generated using Pymol version 1.3. PDB identifiers of structures used as follows: Notch1 ARD, 3V79; FIH, 1H2K; FIH-bound Notch1 peptide, 3P3N.

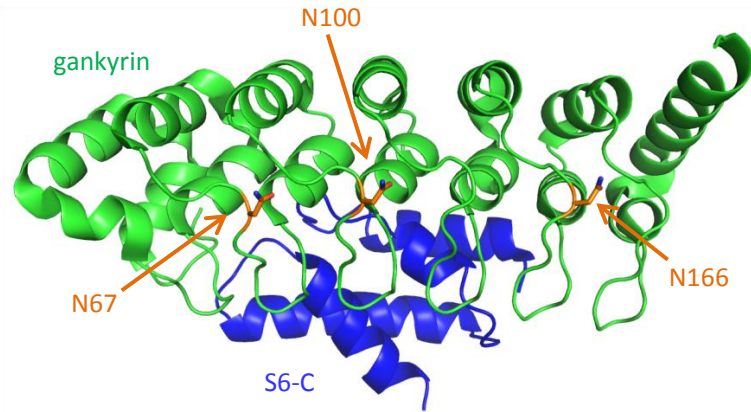
(Figure 3.10) in order to fit FIH's active site. Given the extreme nature of this structural rearrangement, it seems more likely that FIH would take advantage of natural structural fluctuations in the ARD (which may be a part of normal conformational flexibility of the ARD, or occur during synthesis/folding, dissociation from partner proteins, or post damage) (Barrick et al., 2008; Ferreiro et al., 2007; Ferreiro et al., 2005) in order to perform hydroxylation. Indeed, for the FIH substrate,  $\text{I}\kappa\text{B}\alpha$ , partial "unfoldedness" is an important facet of  $\text{I}\kappa\text{B}\alpha$  function (Cervantes et al., 2009; Ferreiro et al., 2007), and could, in theory, facilitate interaction with FIH. Even so, it cannot be ruled out that FIH somehow instigates the unfolding of its substrates itself. If this latter idea is true, this implies that FIH might first need to recognise and bind to a folded ARD before the structural rearrangements required for hydroxylation can take place. In order to test this idea, experiments were designed to determine if FIH was capable of interacting with a structurally constrained ARD.

### 3.7 Structurally constrained Gankyrin can bind FIH

To assess the importance of "unfoldability" in FIH recognition of ARD-containing proteins, the tight binding affinity (observed through robust co-purification of the two proteins, I. Murchland, unpublished observations) between the C-terminal domain of the proteasomal S6 ATPase (S6-C, aa 337-418), and the ARD-containing protein, Gankyrin, was exploited (detailed in the publication (Wilkins et al., 2012)). Crystal structure data indicates that Gankyrin retains the fold of a classical ARD when bound to S6 ATPase (Nakamura et al., 2007), and S6 ATPase binds on the opposite face to the asparaginyl residues that are typically targeted by FIH (Figure 3.11A). In light of this observation, it was anticipated that S6 ATPase would bind to and stably hold Gankyrin in a folded ARD conformation, thus making the S6 ATPase-Gankyrin complex a useful model of a "constrained" ARD which FIH could potentially bind, but not unfold for modification.

To first assess the relevance of Gankyrin as an FIH substrate and binding partner, full length Gankyrin was expressed and purified from bacteria using  $\text{Ni}^{2+}$  affinity chromatography, and then tested as a substrate of FIH by *in vitro* hydroxylation assay. Incubation with MBP-FIH stimulated approximately 3-fold the release of  $\text{CO}_2$  seen with the control, and mutagenesis of likely target asparaginyl residues within the Gankyrin sequence pinpointed Asn100 as the major hydroxy-acceptor residue, with a small additional contribution from Asn166 (Figure 3.11B and (Wilkins et al., 2012)). Next, the comparative FIH binding capacity of S6 ATPase-bound versus unbound Gankyrin was assessed using *in vitro* pulldown assays. To ensure that the samples containing S6 ATPase-bound Gankyrin were not contaminated with free Gankyrin which might skew the results of the assay, S6 ATPase-Gankyrin complexes were purified prior to the experiment using the 6H tag fused to the S6 ATPase. This allowed for purification of solely Gankyrin-bound S6 ATPase, as the S6 ATPase C-terminal domain is insoluble in the absence of the ARD (I. Murchland, personal communication, and (Nakamura et al., 2007)). Likewise, attachment of the tag to the S6 ATPase makes certain that only complexed (and therefore soluble) S6 ATPase/Gankyrin will be available for pulldown by the  $\text{Ni}^{2+}$  resin during assessment of its FIH binding characteristics. Intriguingly, Gankyrin pulled down similar amounts of FIH, regardless of the presence of S6 ATPase, suggesting that FIH can indeed bind to a folded ankyrin (Figure 3.11C and (Wilkins et al., 2012)). The similar efficiency with which FIH binds to free and constrained Gankyrin also raises the possibility that much of the binding affinity may come from FIH's interaction with the folded form of Gankyrin, rather than the open form presumably adopted during hydroxylation. However, a more quantitative method than *in vitro* pulldowns would be required to verify that this is the case. In summary, this finding adds further complexity to the understanding of FIH substrate selection. In particular, if target asparaginyl residues are not part of the interface which facilitates interaction of FIH with a classically folded ARD, this suggests that ARD proteins which are not substrates might also be members of the FIH-binding ARD protein pool. Indeed, this "alternative interface" may help facilitate the interaction of FIH with the ARD-containing protein Notch4, which, despite containing a targetable asparaginyl residue, is not a substrate of FIH (Coleman et al., 2007; Wilkins et al., 2009). Considerable further work will be required to unravel the rules of ARD protein interaction, as well as how FIH's promiscuity affects its regulation of HIF and other substrates.

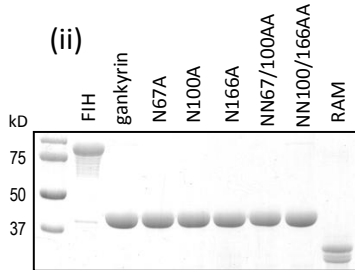
A.



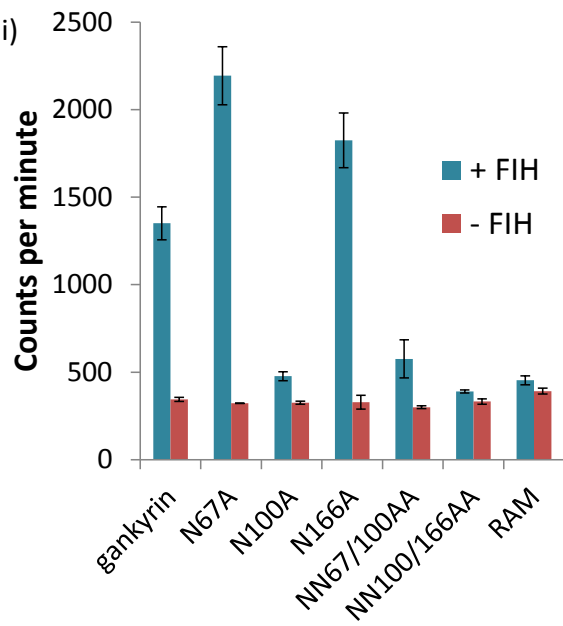
B. (i)

		L	D	Φ	N
hGankyrin	59	L	L	L	L
hGankyrin	92	L	L	L	L
hGankyrin	158	L	L	L	L

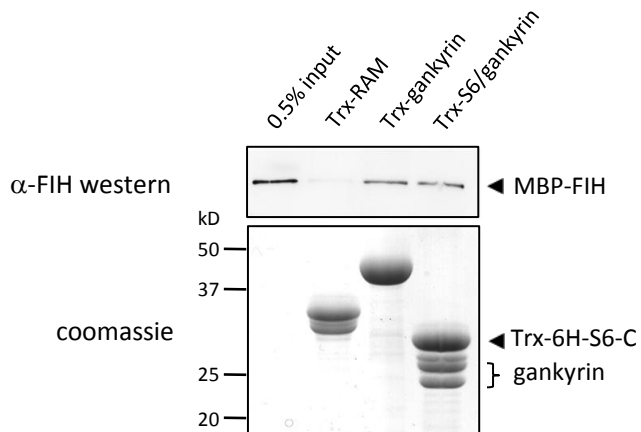
(ii)



(iii)



C.



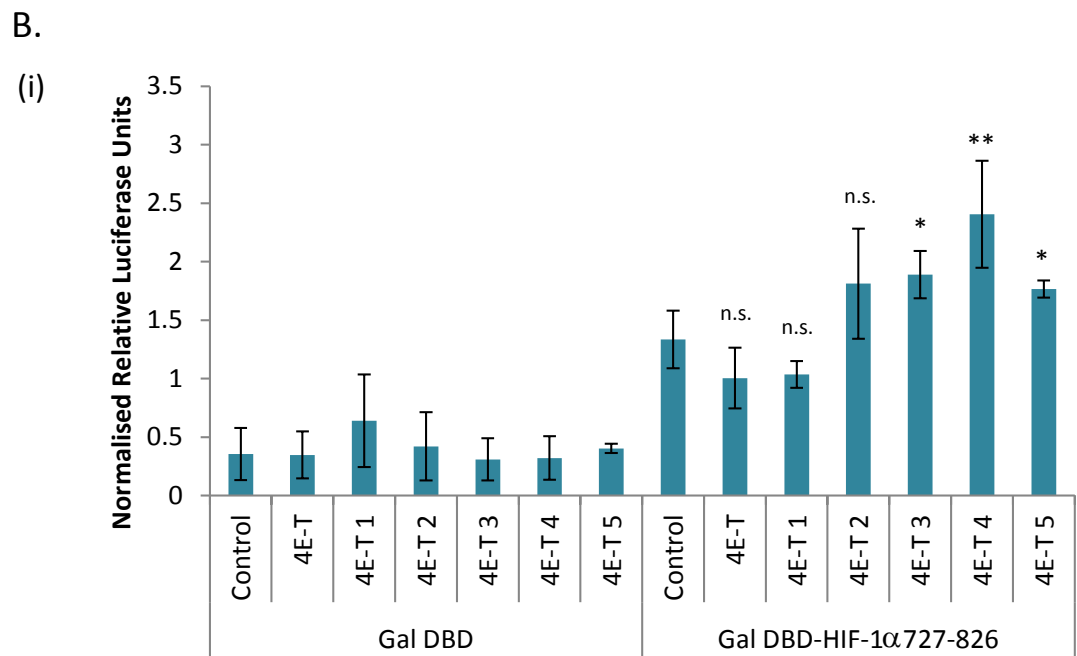
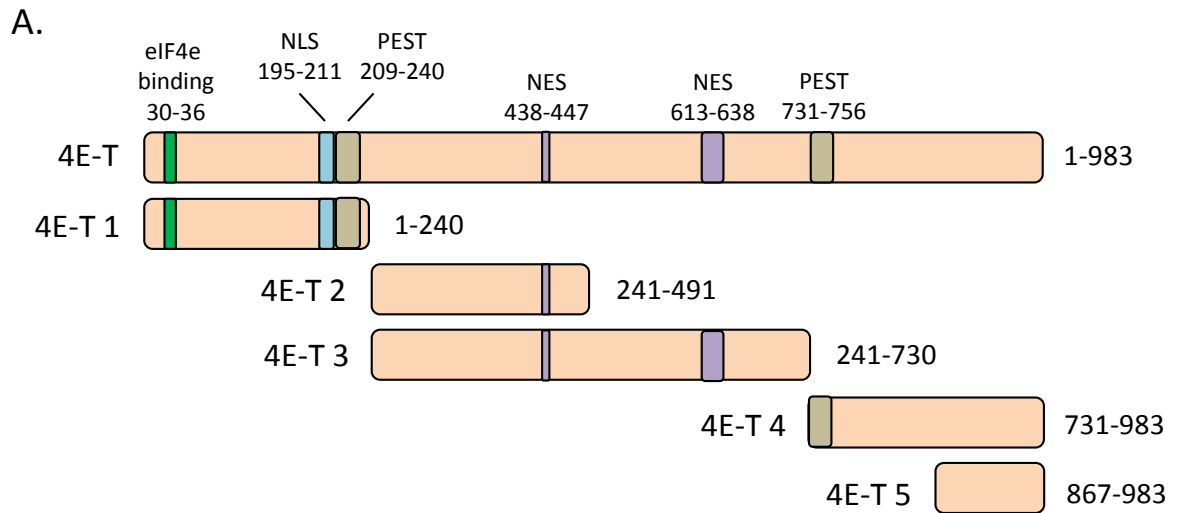
(Figure 3.11)

**Figure 3.11 Assessment of binding of constrained and free Gankyrin to FIH.** Gankyrin was first assessed as a substrate of FIH prior to comparison of the FIH-interaction properties of free versus S6-C-bound Gankyrin. (A) Crystal structure of the Gankyrin S6-C complex (PDB identifier 2DVW) with Gankyrin shown in green, S6-C in blue, and potential target asparagines (based on their location within the long loops between repeats) in Gankyrin shown in orange. (B)(i) Human Gankyrin Asn67, Asn100 and Asn166 are shown with their surrounding sequence, where the target asparagine is shown in red, and residues conforming to the “FIH preferred sequence” are highlighted in green. Aa numbers are shown to the left of each sequence. (ii) and (iii) Full length Gankyrin, as well as Gankyrin containing single or double asparagine-to-alanine mutations of potential target residues were expressed and purified from bacteria using Ni<sup>2+</sup> affinity chromatography. Stability and purity of the proteins was verified by SDS-PAGE followed by Coomassie staining (ii) prior to combination of 65 μM of Gankyrin and its mutants (as well as the negative control protein, RAM) with 2.45 μM MBP-FIH and cofactors in a hydroxylation assay (iii). Data are represented as counts per minute of released [<sup>14</sup>C] CO<sub>2</sub> +/- standard deviation of 3 (for + FIH samples) or 2 (for – FIH samples) replicates. Assay is representative of 3 independent experiments. (C) Gankyrin binding to FIH was assessed by pulldown assay in the presence or absence of the Gankyrin binding protein S6-C. Equimolar concentrations of Trx-6H-RAM, Trx-6H-Gankyrin or a 1:1 complex of Gankyrin:Trx-6H-S6-C (expressed and purified from bacteria) were bound to Ni<sup>2+</sup> resin and then incubated with bacterially expressed MBP-FIH for 2 hrs. The resin was washed, and bound complexes eluted, after which equal volumes of eluate were run on an SDS-PAGE gel. The top half of the gel was transferred to nitrocellulose for western blot with an α-FIH Ab (Novus), while the bottom half was stained with Coomassie. Note: Gankyrin purified bound to S6-C was consistently found to run as a doublet on SDS-PAGE (bands were verified as Gankyrin via MS), unlike Trx-6H-Gankyrin, which runs as a single species. Experiment representative of >3 independent experiments.

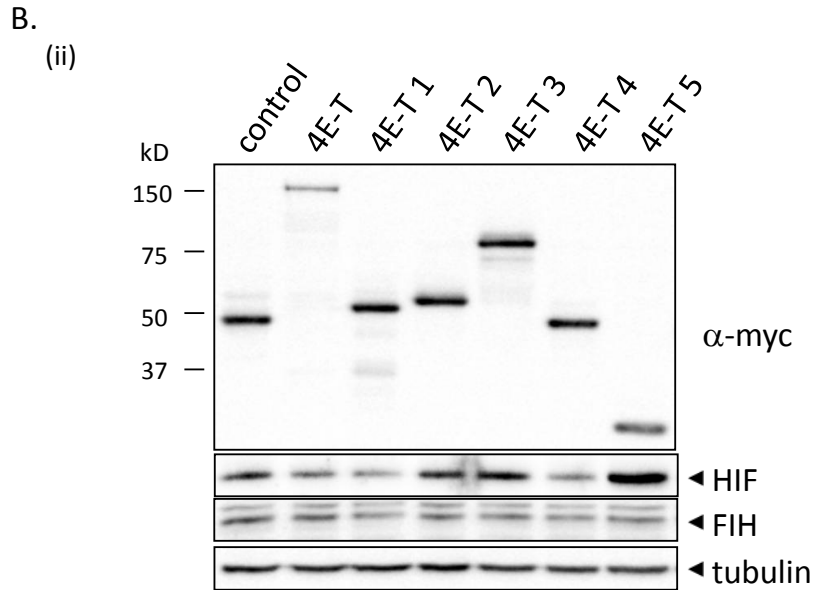
### 3.8 Effect of non-ARD-containing Y2H positive over-expression on FIH repression of the CAD

Whilst exogenous expression of ARD-containing proteins clearly has the ability to titrate FIH away from HIF (see section 3.6), it was important to assess if the non-ARD-containing positives from the Y2H screen had similar properties. Thus, 4E-T, WDR62 and ferritin were also assessed for their ability to influence FIH-mediated repression of the CAD. Unlike most of the ARD-containing positives which activated luciferase production, the non-ARD-containing positives produced a slight repression of reporter gene production in the FIH floxed MEFs (Figure 3.9B(i) and (ii)). A similar but less pronounced effect was seen in the FIH KO MEFs, making the importance of FIH in the effect difficult to ascertain. Indeed, only the difference in repression between Floxed FIH and FIH KO MEFs





(Figure 3.12)



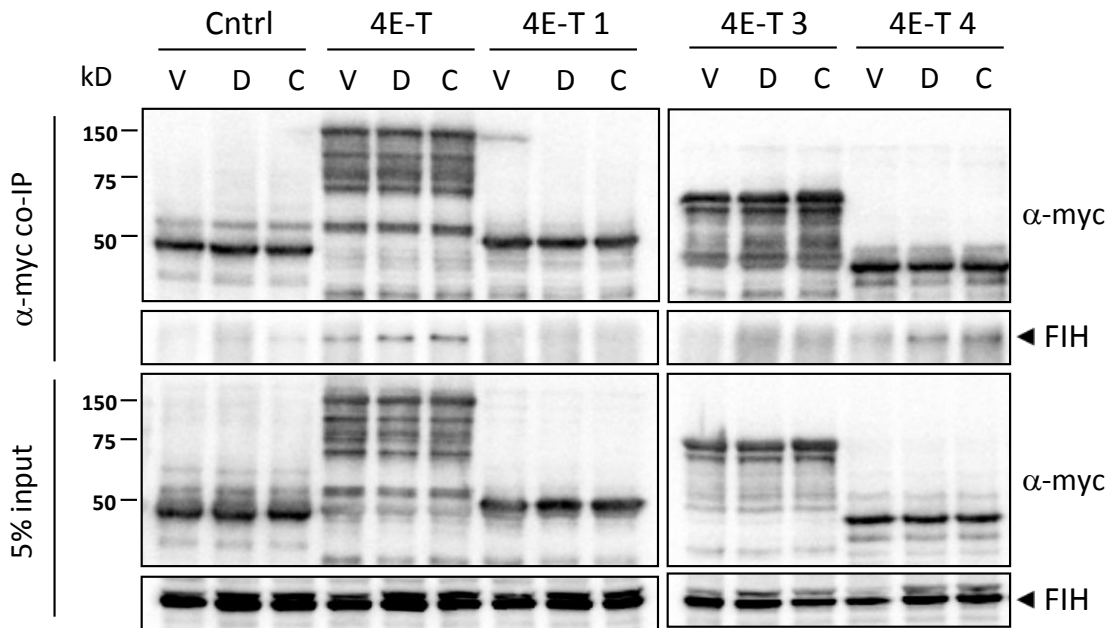
**Figure 3.12 Effect of 4E-T fragments on FIH-mediated repression of the HIF- $\alpha$  CAD.** (A) Schematic representation of the 4E-T fragments and associated functional elements tested in (B). Numbers above the schematics indicate aa positions of functional elements, while those to the right are aa numbers of the fragment. (B) (i) HEK 293T cells were transfected with expression plasmids encoding 6myc-tagged 4E-T constructs (full length 4E-T or 4E-T fragments 1-5) or a myc-tagged control (mouse Ribosomal protein S2, mRPS2), a Gal response element-driven firefly luciferase reporter gene, renilla luciferase under the control of a constitutive promoter, and Gal DBD alone or fused to HIF-1 $\alpha$  (727-826). After 20 hrs cells were lysed and analysed by dual luciferase assay. For each individual experiment, relative luciferase units for each sample were divided by the mean of all RLU sample values across the experiment, then the average of these values graphed +/- SEM. Graph is the combined data of 4 independent experiments. p values calculated using a Student's 2-tailed t-test. \* p < 0.05, \*\* p < 0.01, n.s. not significant (p > 0.05). (ii) Reporter assay lysates from samples transfected with Gal DBD-HIF-1 $\alpha$  (727-826) were separated by SDS-PAGE and analysed for levels of myc-tagged proteins, FIH, and Gal DBD HIF-1 $\alpha$  (727-826) using  $\alpha$ -myc (4A6, Millipore),  $\alpha$ -FIH (NB100-428, Novus) and  $\alpha$ -HIF-1 $\alpha$  (NB100-449, Novus) Abs, respectively. Equal loading was assessed using  $\alpha$ -tubulin Ab (T8203, Sigma).

produced by 4E-T was calculated to be statistically significant (p = 0.033), thus 4E-T was singled out for more detailed analysis. Although characterisation of structure-function relationships for 4E-T are still in their early stages, 4E-T is known to contain an N-terminal eIF4E binding site, as well as a number of nuclear import and export sequences (Figure 3.12A) (Dostie et al., 2000). Additionally, 4E-T was predicted (by epestfind, <http://emboss.bioinformatics.nl/cgi-bin/emboss/epestfind>) to contain 2 PEST motifs, which may potentially contribute to the instability of 4E-T observed by western blot (see Figure 3.8A). Thus it was reasoned that removal of these regions may improve expression of 4E-

T protein and enhance the 4E-T-mediated effects observed. Accordingly, fragments of 4E-T were cloned to isolate various combinations of these functional elements (designated 4E-T 1-5, Figure 3.12A), and then tested for their ability to modulate FIH-mediated repression of Gal DBD HIF-1 $\alpha$  727-826 in 293T cells. Among the new constructs, only 4E-T 1 (aa 1-240), which contained the eIF4E binding site, was able to replicate the repressive effects of full length 4E-T in the reporter assay system, although neither repression reached statistical significance (Figure 3.12B(i)). In contrast, a more C-terminally located region (aa 731-866) present in 4E-T 4 appeared to contribute to a small 1.8-fold activation of reporter gene activity. Given that 4E-T can bind the important translation regulator, eIF4E, the results of this assay must be interpreted with caution. For example, it is plausible that over-expression of 4E-T constructs containing the eIF4E binding site may reduce relative luciferase ratios through a comparatively greater inhibition of translation of firefly versus renilla luciferase. Likewise, inhibition of translation of, for example, the Gal DBD-HIF-1 $\alpha$  (727-826) fusion construct may also reduce firefly luciferase production in a manner independent of FIH-mediated repression of the CAD. Indeed, anti-HIF-1 $\alpha$  western blots (although somewhat inconsistent in some samples across replicates) suggested that both 4E-T and 4E-T 1 produced a trend of reduced expression of the CAD fusion construct (Figure 3.12B(ii)). Finally, the observation that 4E-T may still repress HIF-1 $\alpha$  CAD activity even in the absence of FIH (albeit not to the same extent as when FIH is present) further suggests that the repressive effects of 4E-T are likely not due to a direct influence of this protein on FIH-mediated hydroxylation of HIF-1 $\alpha$ .

### 3.9 The C-terminus of 4E-T displays DMOG-inducible binding to FIH

Although 4E-T over-expression seemed to have minimal effect on FIH-mediated repression of HIF-1 $\alpha$  (727-826) transactivation, this naturally does not preclude a HIF-independent role for the FIH-4E-T interaction. To gain further insight into the regions of 4E-T responsible for the observed co-IP with FIH, myc-tagged 4E-T fragments 1, 3 and 4 were expressed in HEK 293T cells, after which cell lysates were immunoprecipitated with  $\alpha$ -myc antibody, and the efficiency of FIH pulldown compared with full length 4E-T (Figure 3.13). Among the three fragments, only 4E-T 4 (aa 731-983) showed an interaction with FIH, and this interaction was of similar efficiency to that observed with full length 4E-T. As seen in Figure 3.8A, DMOG treatment of the cells prior to extract preparation slightly enhanced the interaction between FIH and 4E-T/4E-T 4. The finding that 4E-T 4 contains the region responsible

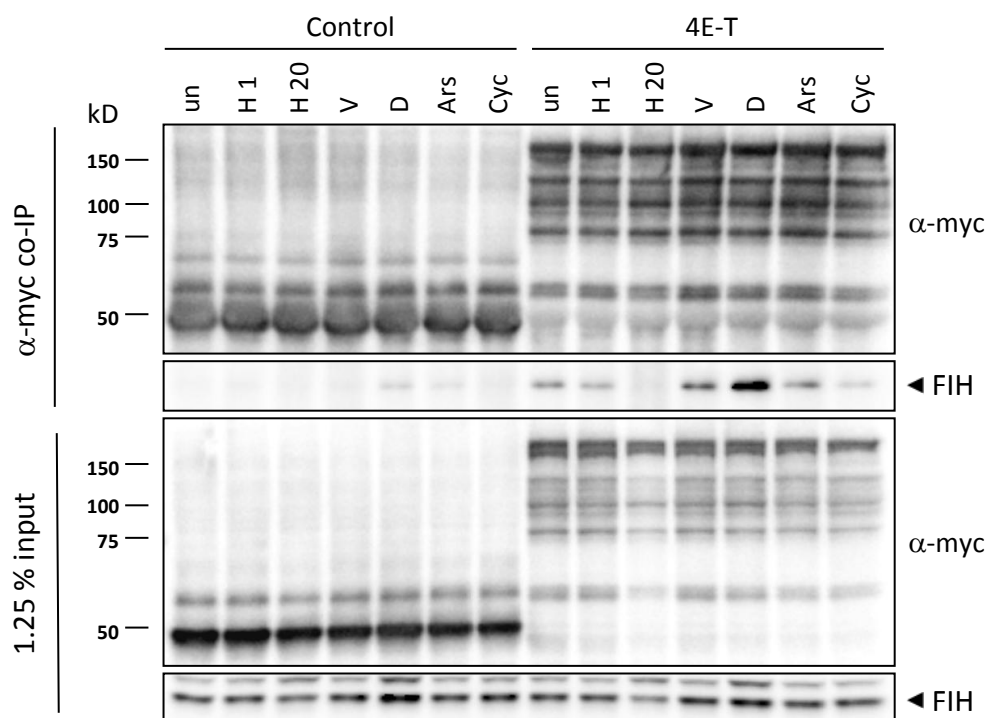


**Figure 3.13 Interaction of FIH with truncation mutants of 4E-T.** HEK 293T cells were transfected with 6-myc-tagged 4E-T, 4E-T 1, 4E-T 3 4E-T 4 or mRPS2 (used as a negative control), and then incubated for 4 hrs prior to a 20 hr treatment with DMSO (1 sample) or 1mM DMOG (2 samples). Then, 1 DMOG-treated sample was additionally incubated in the presence of 1 mM arsenate for 1 hr immediately followed by cell lysate preparation. Lysates were then immunoprecipitated using  $\alpha$ -myc bound protein G sepharose, and co-IP of endogenous FIH was assessed using western blot with  $\alpha$ -FIH Abs (NB100-428, Novus). Lower panels indicate 5% of total WCE used in the IP. V = DMSO, D = DMOG, C = DMOG + arsenate. Figure representative of 2 independent experiments.

for interaction with FIH may go some way towards explaining the reporter assay results in Figure 3.9B, in that 4E-T 4 produced the strongest (albeit comparatively weak) up-regulation of reporter activity, suggesting that the weak FIH-4E-T 4 interaction may interfere slightly with FIH's ability to repress the CAD.

### 3.10 Stress-inducing agents do not influence the interaction between 4E-T and FIH

Preliminary characterisation of 4E-T suggests that it is a nucleo-cytoplasmic shuttling protein, with a significant portion of the protein pool localising to cytoplasmic granules known as processing bodies (PBs). These granules are believed to house non-translating mRNAs as well as many of the proteins involved in mRNA degradation, and accordingly depletion of 4E-T has been shown to increase the



**Figure 3.14 Effect of cell stress inducing agents on the 4E-T-FIH interaction.** HEK 293T cells were transfected with 6-myc-4E-T or 6myc-mRPS2 (used as a negative control), and then incubated for 4 hrs after which samples were left either untreated or treated with 1mM DMOG, DMSO or hypoxia (<1%). After 19 hours, 3 of the untreated samples were incubated with either 1 mM arsenate, 10  $\mu$ g/mL cycloheximide, or treated with hypoxia for 1 hr, which was immediately followed by cell lysate preparation. Lysates were then immunoprecipitated using  $\alpha$ -myc bound protein G sepharose, and co-IP of endogenous FIH was assessed using western blot with  $\alpha$ -FIH Abs (NB100-428, Novus). Lower panels indicate 1.25% of total WCE used in the IP. Un=untreated, H 1 = hypoxia 1hr, H 20 = hypoxia 20 hrs, V = DMSO, D = DMOG, Ars = arsenate, Cyc = cycloheximide. Figure representative of 2 independent experiments.

half-life of mRNA (Ferraiuolo et al., 2005). A variety of cellular stresses are known to increase both PB number and size (Kedersha and Anderson, 2007), and recently it was shown that JNK1-mediated phosphorylation of 4E-T plays a significant role in stimulating PB aggregation in response to the oxidative stress inducer, arsenite (Cargnello et al., 2012). In contrast to 4E-T, FIH is localised primarily to the cytoplasm, with no evidence thus far for concentration within PBs or similar cytoplasmic granules (Linke et al., 2004; Metzen et al., 2003; Stolze et al., 2004). It was of interest, therefore, to determine if cellular stresses which may potentially alter the localisation of 4E-T, or its phosphorylation status, had any influence on the interaction between 4E-T and FIH. As seen in Figure 3.13, treatment of HEK 293T cells with 1 mM arsenate (a cellular stress inducer which has been documented to activate JNK signalling (Porter et al., 1999; Zhang et al., 2002)) for 1 hour in

combination with overnight treatment with DMOG had no significant influence on the binding of 4E-T, or any of 4E-T 1, 3 or 4, to FIH. In efforts to further investigate the influence of PB integrity and 4E-T localisation on 4E-T-FIH interactions, HEK 293T cells were also treated with varying durations of hypoxia (known to promote 4E-T relocation to PBs, as well as an increase in PB numbers (Koritzinsky et al., 2006)), or the PB formation inhibitor, cycloheximide (Figure 3.14). Compared to untreated cells, neither a short (1 hour) hypoxic incubation nor addition of cycloheximide altered the efficiency of interaction between 4E-T and FIH. A longer period in hypoxic conditions (20 hours) did decrease the amount of FIH co-IPed, but this likely reflected the poor health of the cells following both 4E-T overexpression and prolonged hypoxia, and it can be seen that relative input of FIH and 4E-T under these conditions is somewhat reduced. Overall, these results suggest that alteration of 4E-T association with PBs has no effect on the interaction of 4E-T with FIH. However, it must be kept in mind that these experiments utilised over-expressed 4E-T, and it is possible that the treatments used above do not significantly alter the pool of 4E-T available to interact with FIH in the cytoplasm or nucleus. Alternatively, it is possible the interaction may occur both within and outside of PBs. Whatever the case, it is clear that only a small proportion of the total available of each protein interacts, and it remains of interest to determine the functional role of this minority.

### 3.11 Discussion

Following its discovery as a HIF hydroxylase in 2002 (Hewitson et al., 2002; Lando et al., 2002a) and its establishment as an enzyme with genuine oxygen sensing properties, it has been of interest to gain a clearer understanding of the overall role of FIH in oxygen sensing and hypoxia adaptation. Many processes within cells are known to be hypoxia-sensitive, but despite this, the number of characterised oxygen sensors is small. This may suggest one of 2 things: firstly that many oxygen sensors are awaiting identification and characterisation, or alternatively, that a smaller number of oxygen sensors are responsible for regulating multiple processes. Thus, in this work, the aim was to identify novel substrates and binding proteins of FIH, with a view to extending knowledge of FIH's role in hypoxia adaptation.

#### 3.11.1 ARD-containing proteins comprise a novel class of substrates for FIH

A yeast 2-hybrid screen using FIH as bait isolated 7 novel potential interacting factors for FIH, 4 of which contained the structural motif known as the ankyrin repeat domain. Subsequent testing of 2 of the 4 by *in vitro* hydroxylation assay provided supporting evidence that both FGIF and PP6-ARS-B are, in fact, novel substrates of FIH. In addition, co-IP assays further validated the interaction of these

proteins with endogenous FIH in the context of cell lysates. Concurrent with this work being carried out, studies by our group and in other laboratories have established that asparaginyl hydroxylation of ankyrin repeats by FIH is a common event, with at least 13 ARD-containing proteins confirmed as substrates of FIH by mass spectrometry analysis (see Table 1.1). Moreover, 2 of the ARD proteins tested from this group (MYPT1 and Ankyrin R) were purified from tissue or blood samples, thus demonstrating that the modification occurs *in vivo* (Webb et al., 2009; Yang et al., 2011b).

Sequence alignment of target sites within these 13 substrates with the HIF- $\alpha$  proteins suggests that a small number of residues N-terminal to the target asparagine are strongly conserved across all substrates, resulting in the definition of an “FIH-preferred” target sequence, LXXXXX[-] $\phi$ N (see Figure 3.5). Notably, each of the residues which comprise this motif are either strongly (for the -8 Leu relative to the target asparagine) or semi-conserved within the canonical 33 aa ankyrin repeat (Mosavi et al., 2004), suggesting that FIH’s theoretical substrate repertoire may be substantial. It came as no surprise, then, when a comprehensive analysis of all human ankyrin repeat-containing sequences (1505 ankyrin repeats in 252 proteins total) revealed that a suitably located FIH target motif matching the constraint (L- X(4)-[AC]-[DEN]-[IV]-N), was present in 166 ankyrin repeats in 105 proteins (Schmierer et al., 2010). Further broadening FIH’s potential substrate repertoire, was the discovery that the enzyme’s target residue is not limited to asparagine. Specifically, MS analysis of Ankyrin R and Tankyrase-2 has indicated that both aspartyl and histidinyl residues, respectively, may be hydroxylated in a cellular context, although both were hydroxylated less efficiently than asparaginyl residues in each protein (Yang et al., 2011a; Yang et al., 2011b). Importantly, the FIH preferred sequence {L-X(5)-[DE]-[IV]-N-[AV]} is also found in approximately 60 mouse non-ARD containing proteins (see section 3.4). However, testing of several of these candidate substrates by *in vitro* hydroxylation assay did not stimulate FIH activity, suggesting that target site context is important. Nonetheless, discovery of such a plethora of potential target sequences leads to 2 important questions. Firstly, are these predictions truly representative of FIH’s activity within cells? And secondly, what are the functional consequences of these modifications?

### 3.11.2 FIH’s predicted substrate repertoire – a true indication of ankyrin repeat hydroxylation within cells?

In consideration of the first question, some perspective can be gained from examination of FIH’s verified ankyrin repeat targets. In a number of studies, the relative amount of hydroxylation of FIH-targeted sites has been calculated using MS analysis (Table 3.2), with the results suggesting that

hydroxylation efficiency of different ankyrin repeats varies considerably. In some cases, the variation in percentage hydroxylation from one protein to another may reflect the relative amounts of ARD versus FIH protein in a given experimental setting, as increasing FIH levels via over-expression invariably promotes a higher percentage of hydroxylation for any given site (Table 3.2). However, even within a single analysis, some sites are clearly better targets than others (e.g. efficiency of hydroxylation of sites within over-expressed Tankyrase-2 in the absence of FIH over-expression has been observed to be between near background to approximately 49%). The reasons for this inconsistency could be numerous. For example, it is unlikely that all target motifs will be consistently accessible for FIH, as some may be masked by ARD interacting factors. Furthermore, even if a target site is not hidden by binding partners, studies suggest that “ARD unfolding” (as is believed to be required in order to fit the ankyrin repeat target motif into the FIH active site, see Figure 3.10), may prove a further obstacle. A study by Hardy et al (2009) showed that engineered consensus ARDs with high thermal stability (higher than that found in any natural ARDs) were completely refractory to FIH-mediated hydroxylation, despite the presence of a suitable target motif (Hardy et al., 2009). In addition, ARDs for which folding kinetics have been analysed (e.g. Notch and  $\text{I}\kappa\text{B}\alpha$ ) have shown that repeats which are more poorly folded, or fold in the latter stages of ARD formation are generally more efficiently hydroxylated (Bradley and Barrick, 2005; Croy et al., 2004). Thus, the relative structural stability of the ankyrin repeats surrounding a target sequence is an important determinant of hydroxylation efficiency.

With regards to structural stability, it becomes apparent that the tertiary structure of ARDs also warrants consideration as a factor determining hydroxylation efficiency. In the current work, it has been shown that the ARD-containing protein, Gankyrin, can bind FIH even when constrained in its folded ARD structure by association with S6-C (see Figure 3.11). Indeed, a similar finding has been made for an artificial construct termed 4CA (4 consensus ankyrin), a highly thermostable ARD which, despite not being an FIH substrate, has nonetheless been shown to bind FIH by NMR (Hardy et al., 2009). Together, these findings imply that the façade presented by a folded ARD may, rather than concealing a potential hydroxylation site, actually act as part of the recruitment mechanism for FIH. In support of this idea, the  $K_m$  of FIH for the Notch1 ARD is more than 3-fold lower than that observed for a Notch1 peptide encompassing “site 1” and “site 2” interacting residues (Wilkins et al., 2009). Exactly which residues are involved in this recruitment are yet to be determined, although presumably they include residues distinct from the FIH preferred sequence, as some of the amino acids comprising the latter (e.g. the -8 Leu relative to the target asparagine) are typically not present on the surface of the ARD fold, and as such would be unavailable for interaction with FIH.



Substrate	Site	% modified under different hydroxylation conditions				
		Bacterially expressed ARD and FIH	Cell culture			<i>In vivo</i>
			o/e ARD, end. FIH	o/e ARD, o/e FIH	end. ARD, end. FIH	
hIκBα	Asn210	60	10-20	-	-	-
	Asn244	95	25-45	100	5-20 (HeLa), 50 (U2OS, 293T)	-
mNotch1	Asn1945	>95	35	98	30 (HeLa), 96 (293T)	-
	Asn2012	>95	3	81	X	-
hAsb4	Asn246	-	25	89	-	-
hTankyrase-2	Asn203	-	X	-	-	-
	His238	-	<4	30	-	-
	Asn427	-	17	70	-	-
	Asn518	-	X	-	-	-
	His553	-	15	70	-	-
	Asn586	-	46	-	-	-
	Asn671	-	5	-	-	-
	Asn739	-	49	-	-	-
tMYPT1	Asn67	-	5-10	90	-	48
	Asn100	-	5-10	90	-	43
	Asn166	-	5-10	90	-	-
mNotch3	Asn1867	-	50	-	-	-
hAnkyrin R	Asn105	-	-	-	-	26-34
	Asn138	-	-	-	-	20-27
	Asn233	-	-	-	-	35-54
	Asn431	50	X	54	-	X
	Asn464	95	-	-	-	-
	Asn497	7	-	19	-	-
	Asn563	X	-	X	-	-
	Asn629	88	X	94	-	X
	Asn662	8	X	-	-	X
	Asp695	15	-	16	-	-
	Asn728	69	3	95	-	6-21
	Asn761	23	X	30	-	X
hAnkyrin B	Asn58	68	-	-	-	-
	Asn124	74	-	-	-	-
	Asn656	71	-	-	-	-
	Asp491	15	-	-	-	-

**Table 3.2 Hydroxylation efficiency of target sites within MS-verified FIH substrates.** “x” indicates that the tryptic peptide was detected, but found not to be hydroxylated. “-” indicates that the relevant tryptic peptide was not detected, not analysed, or the % hydroxylation was not specified. o/e, over-expressed; end., endogenous. “h”, “m” and “t” before protein names refers to human, mouse and turkey, respectively. See Table 1.1 for references for each substrate.

Finally, although an FIH preferred target sequence has been defined, it is difficult to determine whether this sequence reflects the natural frequency with which these amino acids appear in ankyrin repeats to facilitate ARD folding and tertiary structure, or are true structural preferences of FIH for substrate binding and hydroxylation. Notably, none of the strongly conserved residues amongst FIH substrates differ from the ankyrin repeat consensus described by Mosavi et al (2004) (see Figure 3.5) (Mosavi et al., 2004). However, for each amino acid defined in the FIH preferred binding sequence, there exist examples of efficient hydroxylation of ankyrin repeats (see Figure 3.5A and Table 3.2) which diverge from the preferred residue at that position (see, for example Asn58 of Ankyrin B, which has a Tyr in place of the typical Leu found at the -8 position relative to the target Asn, or Asn431 in Ankyrin R, which has the -2 and -1 residues (most commonly Asp and Val) occupied by Ser and Pro residues, respectively) (Yang et al., 2011b). In essence, the FIH preferred binding sequence is likely to be overly restrictive when attempting to predict the likelihood of hydroxylation of an ankyrin repeat substrate. Indeed, it may be more useful to define a set of markers which are inhibitory to hydroxylation, such as the D<sup>-2</sup>P<sup>-1</sup>N motif (target asparagine in bold type). This sequence, when introduced into the mNotch1 Asn1945 hydroxylation site, was found to generate -1 Pro-dependent unfavourable interactions of the -2 Asp with FIH Lys107, thus preventing Asn1945 hydroxylation (Wilkins et al., 2012). However, on the whole, the number of permissive sequences (i.e. those matching the ankyrin repeat consensus, which by definition, constitute the majority of ankyrin repeats) are likely to far outweigh the number which prevent hydroxylation, which in turn suggests that ankyrin repeat hydroxylation will depend more greatly on enzyme recruitment and target asparagine accessibility than primary sequence determinants.

### 3.11.3 Ankyrin repeat substrates are effective competitors for FIH binding in cells – implications for HIF regulation

In order to determine if ARD-containing proteins could compete with HIF-1 $\alpha$  for FIH-mediated hydroxylation, the Y2H positives containing this domain were over-expressed in HEK 293Ts or MEFs, and hydroxylation efficiency of HIF-1 $\alpha$  assessed using a HIF-1 $\alpha$  (727-826)-driven reporter gene. Results of the assays showed that the elevated levels of ARD-containing proteins do indeed effectively sequester FIH away from HIF-1 $\alpha$ , with ARDs containing 2 hydroxylation sites (FGIF and I $\kappa$ B $\alpha$  (136-314)) performing better than those with only one (PP6-ARS-B (341-486)). Notably, the PP1R12C fragment used in the assay, which does not contain a likely target residue for FIH hydroxylation, could also generate a small induction in reporter activity in HEK 293T cells, suggesting that either the individual ankyrin repeats, or perhaps the folded domain may itself have some affinity for FIH. These data, which agree with results obtained in similar experiments with other ARD-containing proteins (Coleman et al., 2007; Webb et al., 2009), suggest that ARD-containing proteins as a group may act as effective manipulators of FIH availability. Even so, it must be noted that CAD hydroxylation in normoxia is, nonetheless, highly efficient (Tian et al., 2011). But what if the affinity of ARDs for FIH was to increase? Indeed, it has been hypothesised in the literature that cessation of hydroxylation of ARDs, as would occur in hypoxia, may cause a global increase in the binding affinity between ARDs and FIH (hereafter referred to as the “ARD pool theory”), thereby sequestering FIH from HIF and providing a mechanism by which ARDs could indirectly manipulate CAD hydroxylation (Schmierer et al., 2010). In support of this idea, a selection of hydroxylated ARD substrates has been documented to have a reduced affinity for FIH compared to their unmodified equivalents, as would be expected for a typical enzyme-substrate interaction (Coleman et al., 2007; Hardy et al., 2009; Zheng et al., 2008). With this in mind, ARD proteins would be expected to be a more effective “sink” for FIH binding in hypoxia when a larger proportion of ARDs are unmodified. However, this theory is reliant upon the relative oxygen sensitivities of HIF-1 $\alpha$  versus ARD hydroxylation - if oxygen-dependent HIF- $\alpha$  CAD hydroxylation has already ceased before this pool of unmodified ARDs can be generated, it can be assumed that the hydroxylation status of ARD proteins as a whole will have little influence on the kinetics of CAD derepression during decent into hypoxia. Indeed, measurements of FIH's  $K_m$  for oxygen when modifying HIF or Notch substrates (90  $\mu$ M for the CAD,  $\sim$ 12  $\mu$ M or  $\sim$ 90  $\mu$ M for the N- and C-terminal Notch1 target sites, respectively (Wilkins et al., 2009)) suggests that a subset of ankyrin repeats (likely those which are more efficient substrates) would still be effectively hydroxylated by FIH well past the point when CAD hydroxylation would be inhibited. In direct contrast to these *in vitro* measurements, measurement of hydroxylation levels of Notch1 versus the HIF-1 $\alpha$  CAD (albeit in the context of an artificial Notch1 ARD-HIF-1 $\alpha$  CAD fusion protein) in cell

lysates suggests that the hydroxylation of Asn1945 in Notch1 is considerably more sensitive to reductions in oxygen concentration than is the target asparaginyl residue in the HIF- $\alpha$  CAD (Singleton et al., 2011). Exactly why *in vitro* predictions of hydroxylation efficiency of the 2 substrates vary so starkly from “in cell” observations in this scenario is not clear. One possibility is the presence of the di-leucine motif in the HIF- $\alpha$  CAD, which has a strong influence on CAD modification in the crowded cell environment, but fails to influence CAD hydroxylation kinetics *in vitro* ((O'Rourke et al., 1999; Sang et al., 2002; Wilkins et al., 2012) and S. Karttunen, unpublished data). Thus, the current view of HIF- $\alpha$  CAD hydroxylation kinetics may vary considerably from the reality *in vivo*. More importantly, the data reported by Singleton et al (2011) lends credence to the “ankyrin pool theory”, wherein hydroxylation of ARDs as a whole may serve to regulate FIH availability, and thus fine-tune HIF- $\alpha$  CAD activation during fluctuations in oxygen levels (Singleton et al., 2011).

#### 3.11.4 Functional effects of ARD hydroxylation

For the FIH substrate, HIF-1 $\alpha$ , hydroxylation of Asn803 has a well-defined role in perturbing interaction of the CAD with obligate coactivator proteins, thus preventing involvement of the CAD in HIF target gene regulation in normoxia (see Figure 1.3). In contrast, despite various reports of interesting preliminary findings, clear functional effects for ARD hydroxylation have been more elusive. At the time of writing this document, at least 14 separate ARD-containing proteins have been experimentally verified (via mass spectrometry) as recipients of one or more hydroxylation modifications by the FIH enzyme (see Table 1.1). In most instances, the authors reporting these findings have made efforts to determine the functional effects of the modification, with mixed results. For example, in a recent study by Yang et al (2011), over-expression of FIH (but not a catalytic mutant of the enzyme) was documented to reduce the interaction between Ankyrin R and its binding partner, erythrocyte band 3 (Yang et al., 2011b). It was hypothesised that this may alter the morphology and mechanical stability of the erythrocyte membrane, but an influence of endogenous levels of FIH on this structure is unfortunately yet to be demonstrated (Yang et al., 2011b). Conversely, depletion of FIH reduces the binding between ASPP2 (apoptosis-stimulating p53-binding protein) and Par-3 (Partitioning defective 3 homolog), an interaction which is proposed to be important for the maintenance of the polarity and proliferation of neural progenitor cells (Janke et al., 2013). However, again, the functional consequences of this altered binding was not examined further. Studies of the FIH-Notch1 interaction have also produced interesting findings, in that it appears that binding of FIH, rather than hydroxylation of the ARD per se can repress Notch1 signalling in myoblasts (although it should be noted that a majority of Zheng et al's data were obtained using overexpressed proteins, rendering the results of questionable *in vivo* relevance)

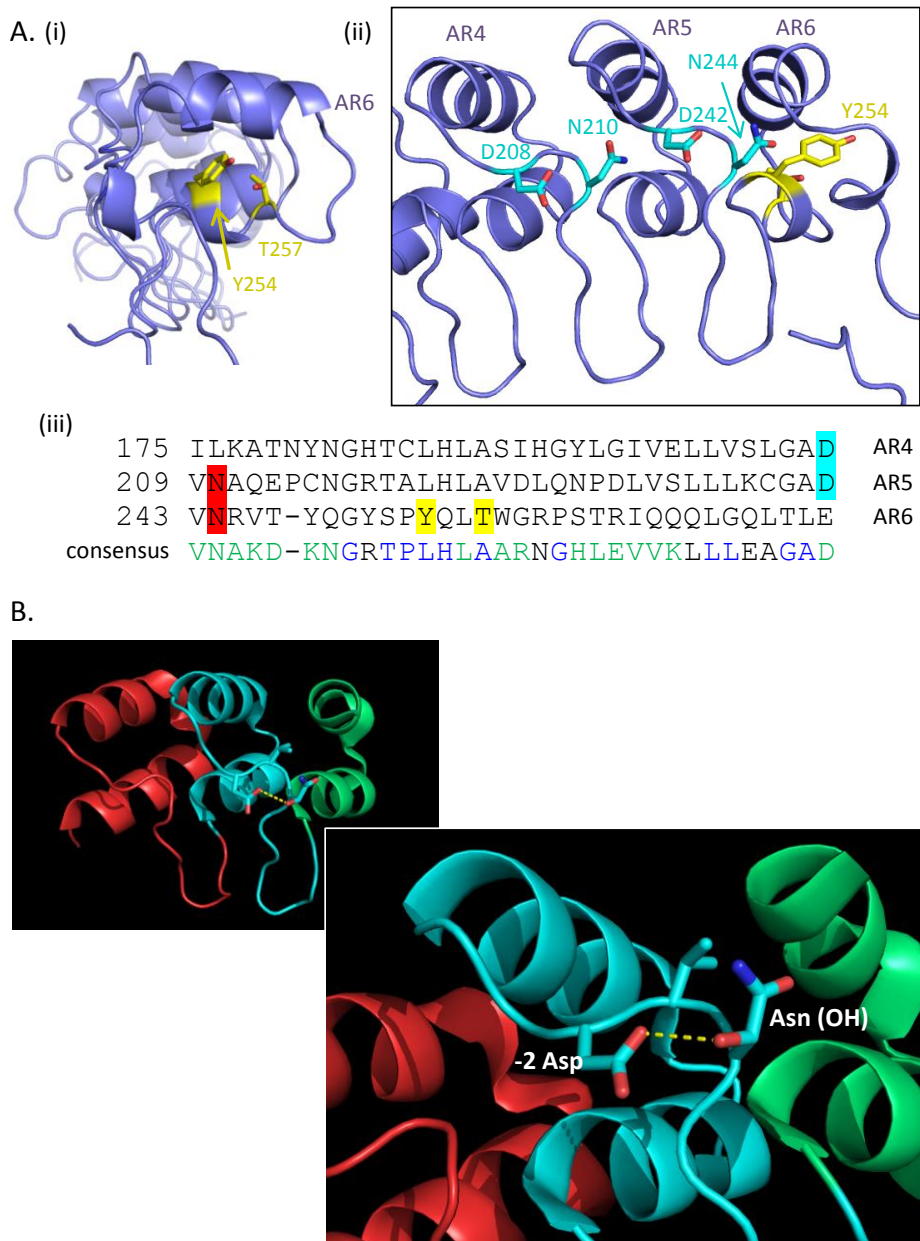
(Zheng et al., 2008). A similar binding-mediated mechanism of action may also apply for the FIH-dependent repression of LRRK1 (leucine-rich repeat kinase 1). FIH KO mice have been found to have delayed wound healing of skin epithelium, and studies in human epidermal keratinocytes (HEKs) and human corneal epithelial keratinocytes (HCEKs) suggest that this may occur through FIH-mediated inhibition of the epidermal growth factor receptor inhibitor, LRRK1 (Peng et al., 2014). Although the authors of this work cite LRRK1 as a potential substrate of FIH, examination of LRRK1's 6 ankyrin repeats does not reveal any likely target residues (data not shown). Thus, although hydroxylation cannot be ruled out as an effector of the observed decrease in EGFR-LRRK1 complex formation upon FIH over-expression (Peng et al., 2014), it is more likely that LRRK1-FIH complex formation is the more important event. If this is the case, successful regulation of LRRK1 would require that the levels of available FIH be dynamically regulated, and indeed, Peng et al report that FIH levels are substantially upregulated in response to wounding of the skin or cornea in mice (Peng et al., 2014).

Perhaps the most convincing data for a genuine role for FIH-mediated hydroxylation has come from analysis of the hypoxia-sensitivity of the FIH substrate, TRPV3 (TRP vanilloid 3 channel). Studies by our group have shown that TRPV3 calcium currents stimulated by the TRPV3 agonist, 2-aminoethoxydiphenyl borate (2-APB) are increased by both sustained (24 hrs, 0.1% oxygen) and acute (5 mins, 3% oxygen) hypoxic treatments, and these effects are dependent on the integrity of FIH hydroxylation target, Asn242 (Karttunen et al., 2014). Notably, Asn242 exists on an extended loop which has been reported to be critical for TRPV3 function, thus further supporting a role for FIH in TRPV3 regulation (Shi et al., 2013). In light of these intriguing findings, a possible role for FIH in TRPV3 modulation *in vivo* is currently under investigation (D. Peet, personal communication).

Although a variety of studies have shown some level of influence of FIH on ARD function, a number of other manuscripts have reported less encouraging results. In the case of FIH substrate MYPT1, manipulation of FIH levels was found to have no effect on MYPT1 protein levels or phosphorylation of the MYPT1 target, myosin light chain (Webb et al., 2009). FIH's first-identified ARD substrate, I $\kappa$ B $\alpha$ , was also assessed on a number of levels to determine if its hydroxylation status affected the activity of its primary target, NF- $\kappa$ B. Results obtained in an *in vitro* setting indicated that hydroxylated I $\kappa$ B $\alpha$  could (1) improve binding of I $\kappa$ B $\alpha$  to bind the p50 subunit of the NF- $\kappa$ B dimer in pulldown assays, and (2) improve I $\kappa$ B $\alpha$ 's ability to inhibit NF- $\kappa$ B DNA binding (as assessed by EMSA) (Cockman et al., 2009). However, treatment of cells with FIH siRNA instead indicated that lack of hydroxylation of I $\kappa$ B $\alpha$  had the opposite effect on NF- $\kappa$ B activity to that implied by the *in vitro* results, leaving considerable doubt as to the true effect of FIH on I $\kappa$ B $\alpha$  activity (Cockman et al., 2009). In

summary, despite the discovery of a wealth of new ARD substrates, initial studies would suggest that many of FIH's enzymatic interactions may have underwhelming functional consequences.

Although the publication regarding FIH-I $\kappa$ B $\alpha$  functional interactions documented findings that were less than clear-cut, intriguing data were recently generated by the laboratory of Betsy Komives describing the critical role of I $\kappa$ B $\alpha$  ARD stability in regulation of NF $\kappa$ B signalling. In free I $\kappa$ B $\alpha$ , repeats 5 and 6 of the ARD are only loosely folded due to significant deviation of the ankyrin repeat sequences from the consensus ankyrin repeat structure (Figure 3.15A) (Truhlar et al., 2008). In this 2008 publication, Truhlar et al reported that this partially unfolded nature was a critical determinant of NF $\kappa$ B binding affinity, as well as the rate of degradation of free I $\kappa$ B $\alpha$  protein. Artificial stabilisation of the I $\kappa$ B $\alpha$  ARD by mutating ankyrin repeat 6 to a more "consensus-like" sequence resulted in a reduction in NF $\kappa$ B binding affinity, as well as a slower rate of degradation of free I $\kappa$ B $\alpha$ . Collectively, these changes resulted in an elevated level of nuclear NF $\kappa$ B in resting cells, and a poorer degree of NF $\kappa$ B activation upon stimulation (Truhlar et al., 2008). Interestingly, hydroxylation of ankyrin repeats by FIH has also been reported to have a stabilising effect on ARD structure, possibly due to the formation of a hydrogen bond between the newly applied beta carbon hydroxyl group on the target asparagine and the carboxyl group of the neighbouring acidic residue at position -2 (Figure 3.15B) (Hardy et al., 2009; Kelly et al., 2009). Given that one of I $\kappa$ B $\alpha$ 's 2 characterised hydroxylation sites resides in the poorly folded 5<sup>th</sup> ankyrin repeat, experiments were designed to determine if FIH-mediated hydroxylation could influence I $\kappa$ B $\alpha$  "foldedness". Indeed, given that classical activation of NF $\kappa$ B involves degradation of NF $\kappa$ B-bound I $\kappa$ B $\alpha$ , it is conceivable that the level of oxygen availability at the time of I $\kappa$ B $\alpha$  resynthesis may dictate I $\kappa$ B $\alpha$ 's hydroxylation state, and in turn alter the ability of I $\kappa$ B $\alpha$  to strip NF $\kappa$ B from DNA post cessation of the activating stimulus. In line with this, there are numerous reports supporting a role for hypoxia in modulation of NF $\kappa$ B signalling (Culver et al., 2010; Scholz and Taylor, 2013; Taylor and Cummins, 2009). To avoid the complicating factors associated with analysing these processes in a cellular context, it was decided to undertake a detailed analysis of



**Figure 3.15 IκBα foldedness is a key determinant of its ability to repress NFκB.** (A) The loosely folded ankyrin repeats 5 and 6 in IκBα can be stabilised through mutation of Tyr254 and Thr257 (depicted in yellow in the “end-on” (i) and “side-on” (ii) structures of IκBα, or highlighted in yellow in the alignment (iii)) to the residues typically found in these positions in the ankyrin repeat consensus (Mosavi et al., 2004). Target asparagines in IκBα which may hydrogen bond with the neighbouring -2 Asp residues post hydroxylation are shown in turquoise in (ii) or turquoise and red (-2 Asp and target Asns, respectively) in (iii). The consensus in (iii) is coloured as for Figure 3.5. (B) Structure of 3CA (3 consensus ankyrin) is shown with consecutive repeats depicted in red, turquoise or green. The hydroxylated target asparagine in 3CA is oriented to hydrogen bond (yellow dotted line) with the -2 aspartic acid residue relative to the target asparagine. AR = ankyrin repeat. Structure diagrams were generated using Pymol version 1.3. PDB identifiers of structures used: IκBα, 1IKN; 3CA, 2ZGD.

the effect of hydroxylation on I $\kappa$ B $\alpha$  function using well-established *in vitro* protocols utilised by the Komives laboratory. The results of this collaborative study can be found in Chapter 6.

In the work presented here, both FGIF and PP6-ARS-B were identified as likely FIH substrates, but lack of functional characterisation of the proteins at the time of their discovery hampered further examination of the consequences of hydroxylation. PP6-ARS-B is, based on its homology to PP6-ARS-A, predicted to be an adaptor protein for Protein phosphatase 6, although this is yet to be verified in an experimental setting (Stefansson et al., 2008). FGIF, is marginally better characterised in that it is believed to promote  $\gamma$ -globin expression in human fetal erythroblast-mouse erythroleukemia (HFE-MEL) hybrid cells. However, few mechanistic details regarding how this effect was achieved were provided (Yang et al., 2001), and attempts to corroborate their claim that FGIF functions as a coactivator were unsuccessful (see section 3.3). Given its potential role in regulation of  $\gamma$ -globin expression, it would be of interest to determine if manipulation of FIH levels in transgenic mice bearing the a human  $\gamma\beta^A$  globin knock-in (unlike WT mice, which switch directly from embryonic to adult haemoglobin expression during development, the  $\gamma\beta^A$  mice have a human-like progression of beta globin gene expression, wherein embryonic globin expression is followed by a period of human  $\gamma$ -globin expression prior to the final switch to adult  $\beta$ -globin (McConnell et al., 2011)) displayed altered kinetics of globin switching, a process known to be sensitive to oxygen levels (Bichet et al., 1999; Narayan et al., 2005). However, it should be noted that no phenotype associated with anaemia or altered haemoglobin production was reported for the FIH KO mice (Zhang et al., 2010). Thus, it can be assumed that the effects of FIH on FGIF signalling (at least in the case of globin expression), if any, would be subtle.

### 3.11.5 FIH interacts with cell stress signalling molecules

In addition to ARD-containing proteins, the Y2H also identified a number of non-ARD-containing proteins (4E-T, WDR62 and ferritin) as potential FIH binding partners. Unlike the ARD containing proteins, none of these factors were found to be hydroxylated by FIH, nor could their over-expression significantly perturb FIH-mediated repression of the HIF-1 $\alpha$  CAD. However, co-IP assays demonstrated that full length constructs of both 4E-T and WDR62 could pull down low levels of either endogenous or over-expressed FIH, respectively, although the former interaction was only detectable by western blot post treatment of cells with FIH inhibitor, DMOG. Similarly, a C-terminal fragment of WDR62 was found to weakly interact with endogenous FIH in a DMOG-inducible manner. Given the lack of hydroxylation of either of these proteins, the finding that a “substrate



trapping” inhibitor of FIH can promote their binding to FIH is puzzling. Conceivably, the interaction between FIH and 4E-T/WDR62 may be secondary rather than direct, wherein the intermediary factor may be either an FIH substrate (the binding of which is promoted by DMOG), or possibly even another 2-OG-dependent oxygenase, although little evidence exists thus far to support such an idea (see discussion of HIF-2 $\alpha$  below). Indeed, the requirement for an accessory factor for the interactions may also explain the poor efficiency of the observed co-IP for both 4E-T and WDR62. Alternatively, in some experiments, it was observed that in DMOG treated samples the levels of FIH protein in cell lysates seemed to be subtly increased (see for example, Figure 3.14 and Figure 3.15), which may have enhanced pulldown of FIH in these experiments. Due to the subtlety of the increase, this enhancement was not always visible in input samples of the co-IPs (indeed, vastly overexposed western blots examining FIH levels post DMOG treatment in the literature also preclude assessment of the consistency of the effect), and further experimentation will be required to ascertain if (1) DMOG is a genuine effector of FIH protein expression, and (2) if this effect contributes to improved 4E-T/WDR62-mediated pulldown of FIH.

With regards to the functional relevance of the interaction between FIH and WDR62, a cellular role for WDR62 has only recently been defined. In 2010, protein interaction screens initially identified WDR62 as a scaffolding protein for JNK (c-Jun N-terminal kinase) pathway kinases (Wasserman et al., 2010). Subsequent to this, defects in the WDR62 coding sequence were identified as the second most common cause of autosomal recessive primary microcephaly (MCPH), a neurodevelopmental disorder phenotypically characterised by a small head and brain (Nicholas et al., 2010). More detailed study of WDR62 function revealed it to localise to the centrosome during mitosis, and knock-down of the protein results in delayed mitotic progression and spindle orientation defects, which likely contribute to disturbed mitosis in the neural precursors involved in MCPH development (Bogoyevitch et al., 2012; Nicholas et al., 2010). When considering a possible role for FIH in modulation of this aspect of WDR62 function, it was apparent that the FIH KO mice do not display any obvious symptoms of neurodevelopmental defects, thus making the likelihood of FIH involvement in this process remote. That said, FIH localisation during mitosis has never been specifically examined, thus intracellular localisation studies of synchronised cell populations may help to clarify if a proportion of FIH relocates to spindle poles during M-phase. Moreover, the lack of a neurodevelopmental phenotype in the FIH KO mice does not preclude a role for FIH in other aspects of WDR62 biology, or the possibility that FIH, which is a known phosphoprotein (Lisy, 2011) may function as a substrate of JNK.

### 3.11.6 4E-T, FIH and regulation of translation

4E-T was originally identified as a nucleocytoplasmic shuttling protein which could mediate transport of the translation initiation factor, eIF4E, via a canonical Tyr-x-x-x-Leu- $\phi$  eIF4E binding motif found in the N-terminus of the protein (Dostie et al., 2000). Subsequent to this, 4E-T has been found to be an important component of P-bodies (PBs), and can recruit eIF4E to these cytoplasmic foci wherein non-translating mRNAs are stored for future reinitiation of translation or degradation (Ferraiuolo et al., 2005). The precise role that 4E-T plays in PBs is unclear, but interaction of 4E-T with eIF4E prevents the latter's association with eIF4F, thus presumably promoting a repressed state for mRNAs engaged in the eIF4E-4E-T complex (Dostie et al., 2000). In addition, knock-down of 4E-T has been shown to promote disassembly of PB aggregates, as well as increase mRNA half-life (Andrei et al., 2005; Ferraiuolo et al., 2005).

Repression of translation in response to cell stresses, such as hypoxia and oxidative stress, is an important adaptive response which aids to maintain energy homeostasis (see section 1.3). In 2006, it was shown that hypoxia promoted the accumulation of 4E-T in PBs, suggesting that 4E-T may play a part in global translational repression in response to oxygen deprivation (Koritzinsky et al., 2006). In addition, oxidative stress-induced phosphorylation of 4E-T by the stress-activated protein kinase JNK (c-Jun N-terminal Kinase) has also been documented to increase 4E-T self-association and PB size (Cargnello et al., 2012). Post identifying the oxygen sensor FIH as a 4E-T interacting protein, it was of interest to determine if the interaction between 4E-T and FIH might be modulated in response to alterations in PB formation and integrity. However, preliminary experiments involving treatment of HEK 293T cells with arsenate (a JNK activator), hypoxia and cycloheximide (which induce or disperse PBs, respectively) had minimal effects on the binding efficiency of FIH and 4E-T (see section 3.10). A 20 hour hypoxic treatment of HEK 293T cells did reduce the amount of FIH pulled down by 4E-T in co-IPs, but this result may have reflected reduced input caused by long-term hypoxic stress. Thus, a time course of hypoxic treatments may prove more informative as to the true effect of hypoxia on 4E-T-FIH binding interactions. Furthermore, although experiments examining the localisation of FIH in normoxia and hypoxia has not revealed any dramatic changes in the enzyme's localisation (Stolze et al., 2004), assessment of colocalisation of FIH with 4E-T or PB markers (such as Dcp1a (Ferraiuolo et al., 2005)) during stress would further help to determine if FIH plays a role in PB biology.

If FIH is a true modulator of 4E-T (and by extension eIF4E) function, it is plausible that FIH may have some influence on the efficiency of translation initiation. At present, there is little evidence to support such an idea. However, in 2012, the group of Stephen Lee published a report showing that

the FIH substrate, HIF-2 $\alpha$ , was part of a regulatory complex comprising eIF4E2 (an eIF4E homolog), RNA-binding motif protein 4 (RBM4), and RNA helicase eIF4A which was required for effective activation of translation in hypoxia (Uniacke et al., 2012). Translation initiation by this method requires the presence of an RBM4-bound "rHRE" in the 3'-UTR of the translated mRNA, to which HIF-2 $\alpha$  binds in a hypoxia-inducible manner, thereafter promoting the recruitment of eIF4E2 and polysomes (Uniacke et al., 2012). As normoxic translation typically utilises eIF4E and the eIF4F complex (Jackson et al., 2010), the onset of hypoxia would likely trigger an exchange of these factors for either the eIF4E2/HIF-2 $\alpha$ /RBM4 complex for those mRNAs which are preferentially translated during hypoxic stress, or repressive factors (e.g. 4E-T), which would down-regulate translation of non-essential transcripts. The translational status of mRNAs is believed to be highly dynamic, with rapid exchange of mRNAs between actively translating pools and mRNA "triage" centres such as PBs. Thus, it is likely that factors associated with the 5' cap would also need to be amenable to exchange. The finding that FIH may interact with both translationally repressive (4E-T) and activating (HIF-2 $\alpha$ ) factors reveals the intriguing possibility that FIH could modulate the transition of mRNAs from active to inactive translation, or vice versa. With this in mind, manipulation of FIH levels followed by polysome profiling or qPCR of transcripts isolated from polysome cell fractions would be worthwhile experiments to assess the effects of FIH on hypoxic translation.

### 3.11.7 Conclusions

The discovery of FGIF and PP6-ARS-B as likely novel substrates, combined with reports of other targets in the literature, has defined ARD-containing proteins as a novel class of substrates for the 2-OG-dependent oxygenase, FIH. Sequence analysis of these substrates suggests that FIH is highly flexible when it comes to substrate choice, and hydroxylation efficiency of individual ankyrin repeats in cells will likely be influenced by ankyrin repeat accessibility, unfoldability of the ARD, as well as primary sequence determinants. The substantial number of predicted ankyrin repeat substrates for FIH has led to the proposal of an "ARD pool theory", wherein FIH availability for HIF- $\alpha$  hydroxylation is hypothesised to be tuned by the hydroxylation status (which in turn regulates affinity for FIH) of ARD substrates as a whole. The theory suggests that the effectiveness of this tuning will be dependent on the concentration of ankyrin repeat substrates, and the effectiveness with which they are hydroxylated. However, at present, accurate prediction of which ankyrin repeats are likely to make effective substrates is not feasible. Furthermore, the finding in this work that a folded ARD may also bind FIH adds a further level of complexity, and it is yet unknown whether hydroxylation is also able to regulate folded-ARD/FIH binding interactions. Finally, FIH was also found to interact with several non-ARD containing proteins, although the efficiency of these interactions was considerably

lower than that observed for ARD-containing proteins. It remains to be determined if these interactions are functionally relevant – indeed, preliminary experiments which manipulated PB integrity had minimal influence on 4E-T-FIH interactions. Nonetheless, the fact that FIH interacts with both positive and negative regulators of hypoxic translation flags the enzyme as a potential modulator of this process, an observation which warrants further investigation.

# 4 Results – Part 2

An evolutionary biology approach to understanding FIH substrate recognition



## 4.1 Introduction

For many fundamental biological processes, the key signalling pathways are evolutionarily conserved across a wide variety of organisms (Hardison, 2003). Moreover, the degree of conservation of any given molecule can be a useful marker of its functional contribution to a certain biological output. Indeed, this idea is the basis for an ongoing area of research by our laboratory to assess the relative importance of FIH in HIF- $\alpha$  CAD regulation versus ARD hydroxylation. It is hypothesised that the order in which a genuine FIH homolog, a targetable HIF and targetable ARD substrates appeared in evolutionary history will shed some light on the “original” function of FIH. More specifically, if FIH evolved before the oxygen-sensitive HIF- $\alpha$  CAD, an important functional role for FIH in hydroxylating ARD-containing proteins (or indeed some other non-HIF substrate) would seem more plausible. In the current work, conservation of FIH across phyla has been investigated, and evolutionary comparisons have also been used to study the conservation and structural properties of FIH substrates in different species, with a view to (1) improving understanding of FIH’s flexible substrate binding properties, and (2) determining if FIH exerts “evolutionary pressure” on these substrates, as may be anticipated if the interaction is of functional importance.

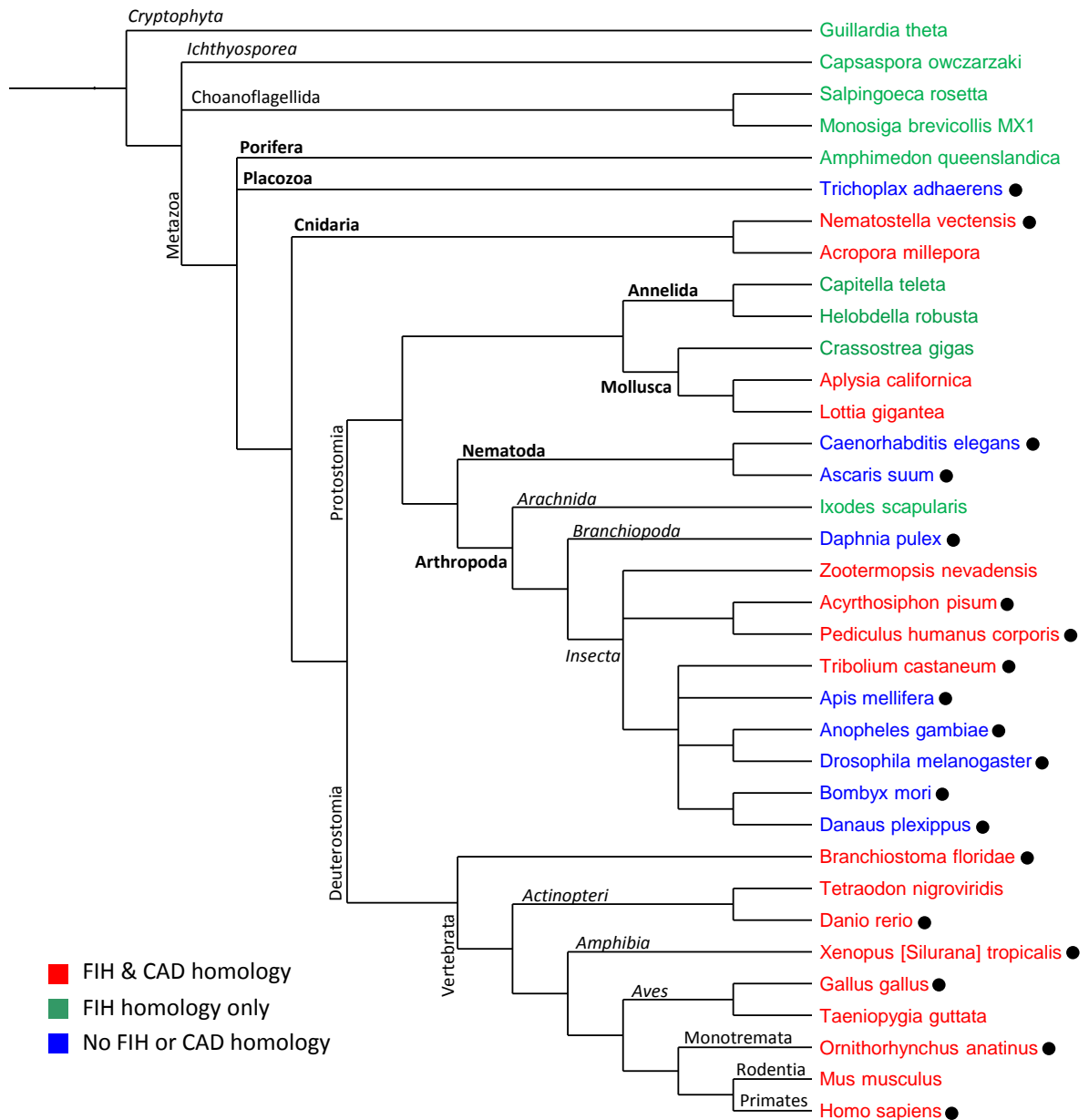
## 4.2 FIH-CAD relationships in divergent species

Work carried out in our laboratory and others’ has revealed FIH and a CAD-containing HIF- $\alpha$  protein to be conserved in a wide variety of species, including 2 members of the ancient Cnidaria phylum, *Acropora millepora* (coral) and *Nematostella vectensis* (Starlet Sea anemone) (B. Davenport, unpublished data and (Loenarz et al., 2011)). Indeed, the bioinformatic analysis performed by Loenarz et al (2011) predicted perfect co-conservation of putative FIH and CAD homologs in the 20 species examined, which in turn suggests that an important functional link between FIH and the CAD was developed early in metazoan evolution (Loenarz et al., 2011).

To investigate this link between FIH and the CAD further, genomes from a range of species were first searched for the presence of FIH using the pBLAST and tBLASTn algorithms and the human FIH sequence. In some cases, these searches revealed only partial or intermittent matches to an FIH-like sequence. However, these partial matches could often be extended or revised by reannotation of intron-exon boundaries in nearby genomic sequences to the identified BLAST match, or concatenation with database EST sequences (see Appendix 8.2). If BLAST results from a species lacked clear homology to human FIH, in particular the key residues Tyr145, Thr196, His199, Asp201, Lys214, Arg238, Gln239 and His279, the species was classified as lacking FIH. Post classification of

their FIH status, genomes were similarly BLAST-searched for a CAD- containing HIF- $\alpha$  homolog using the human HIF-1 $\alpha$  sequence. For searches which identified a likely HIF- $\alpha$  homolog lacking a CAD domain, the hit's coding sequence was located within the relevant genome, and then at least 20 kb of downstream sequence was retrieved, translated and searched using a HIF- $\alpha$  CAD hidden Markov model (HMM, generated using HMMER3.0 from known CAD homologs found in *Tribolium castaneum*, *Acyrtosiphon pisum*, *Strongylocentrotus purpuratus*, *Danio rerio*, *Xenopus tropicalis*, *Homo sapiens*, *Gallus gallus*, *Ornithorhynchus anatinus*, *Branchiostoma floridae*, *Acropora millepora* and *Nematostella vectensis*) to locate as yet unannotated CAD-containing exons. Using this method, CAD-like sequences were located adjacent to partially annotated HIF- $\alpha$  genes for 2 members of the class Insecta, *Zootermopsis nevadensis* (termite), and *Pediculus humanus corporis* (human body louse). An alignment of known and predicted FIH and CAD homologs including these search results, as well as the evolutionary relationships of the host organisms can be found in Appendix 8.3 (FIH) and Figure 4.8 (HIF- $\alpha$ ), and Figure 4.1, respectively. As for the results of Loenarz et al's study, all species which contained a HIF- $\alpha$  CAD also contained an FIH homolog, supporting the supposition that FIH is an integral part of CAD regulation (Loenarz et al., 2011). Notably however, a number of metazoan species, including *Ixodes scapularis* (deer tick), *Helobdella robusta* (Californian leech), *Capitella teleta* (Polychaete worm), *Crassostrea gigas* (Pacific oyster), and *Amphimedon queenslandica* (sponge) were found to contain a likely FIH homolog, but no discernible CAD equivalent within their predicted HIF- $\alpha$  homolog. As genome assembly and annotation is a long-term process, it is possible that further sequencing and refinement of these genomes may yet reveal the presence of a CAD in some or all of these organisms. However, of considerable interest was the discovery that a number of pre-metazoan eukaryotic genomes, including those of *Monosiga brevicollis*, *Salpingoeca rosetta*, *Capsaspora owczarzaki* and *Guillardia theta*, also yielded either partial or full length BLAST hits to a human FIH query (Appendix 8.3). As for the partial FIH matches retrieved for the metazoan species, additional FIH-homologous sequence for each of the non-metazoan fragments could be found via manual examination of BLAST hit-proximal genomic sequence (Appendix 8.2). Given that the HIF transcription factors are believed to have first evolved in multicellular animals (and indeed, BLAST searches using human HIF-1 $\alpha$  in the non-metazoan species revealed little similarity (data not shown)), the finding of likely FIH homologs in these simple organisms suggests that the FIH enzyme evolved prior to its HIF- $\alpha$  CAD substrate. In contrast, these species are each host to an abundance of FIH's other known targets, ARD-containing proteins.





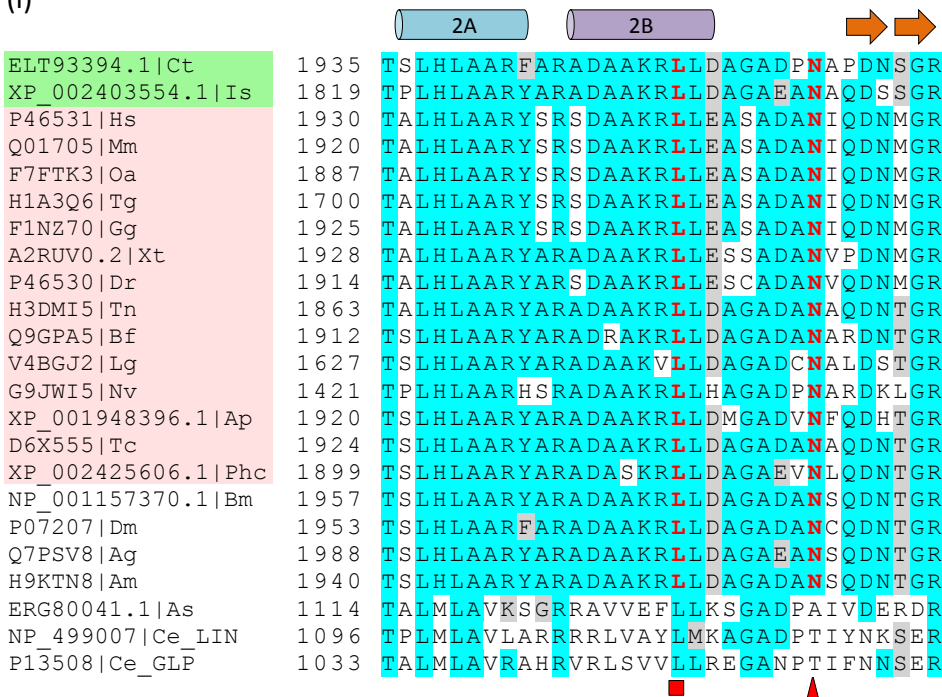
**Figure 4.1 Conservation of FIH and the HIF- $\alpha$  CAD across species.** Results of BLAST searches were combined with existing data (B. Davenport, unpublished data, and see also (Loenarz et al., 2011)) to generate an evolutionary tree wherein the presence (known or predicted) or absence of an FIH homolog and an “FIH-targetable” HIF- $\alpha$  CAD is indicated via red, green or blue text. Horizontal branch labels show either phyla (bold text), classes (italics), or orders (normal text). Black circles after species’ names indicate those that were included in the proteomic analysis of ankyrin repeats described in Section 4.3.1. Evolutionary tree generated using phyloT (<http://phylot.biobyte.de/index.html>) and Interactive Tree of Life (<http://itol.embl.de/index.shtml>).

### 4.3 FIH-ankyrin repeat relationships across species

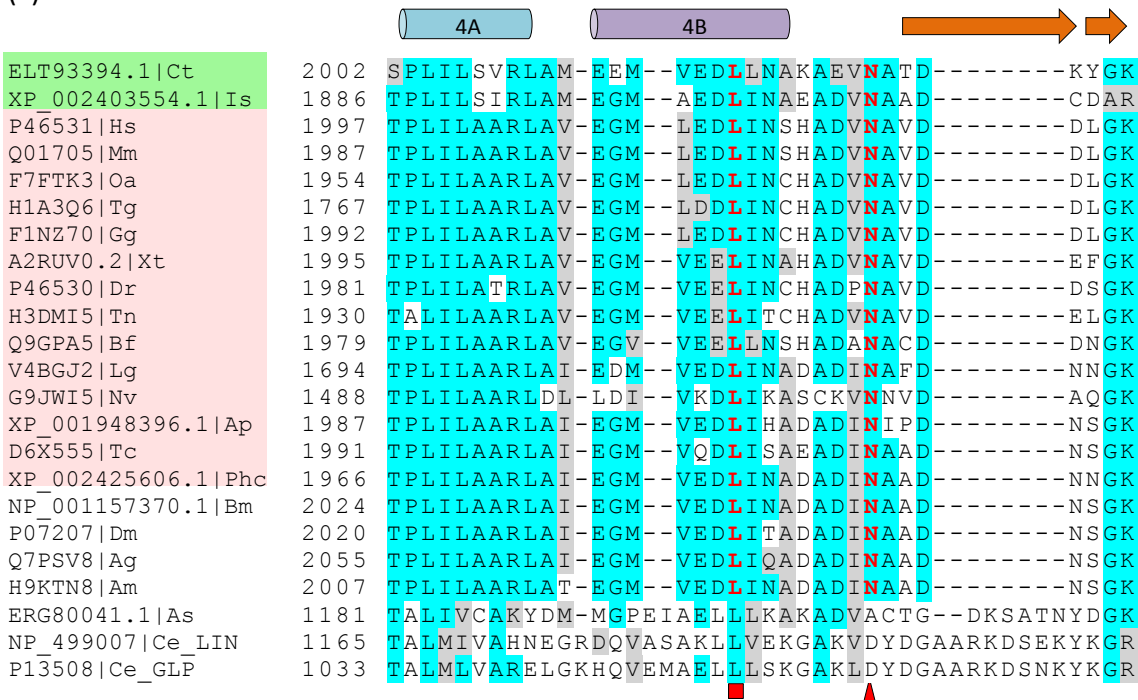
The observation that many species which lack a HIF- $\alpha$  CAD also lack FIH suggests that there is an important functional connection between these elements. By this reasoning, if an important functional relationship exists between FIH and hydroxylation of one of its ankyrin repeat substrates, it would be anticipated that loss of FIH would also influence the conservation of its FIH target sequence. Testing this idea in practice requires that the substrate to be analysed is conserved in a wide variety of organisms, including the evolutionary more basal species which lack FIH. Fortunately, among FIH's ankyrin repeat-containing targets, Notch homologs are believed to be present in phyla as primitive as Porifera, while the *cactus* gene product is considered to be a direct arthropod homolog of vertebrate  $\text{I}\kappa\text{B}\alpha$  (Basith et al., 2013; Gilmore and Wolenski, 2012; Richards and Degnan, 2009). Thus, these proteins were selected for analysis. For each protein, the human homolog (Notch1 was chosen from the 4 human Notch homologs) was BLASTed against the non-redundant protein sequence database, and known or predicted homologs documented for the species listed in Figure 4.1 (where available) were collected. Due to the notable size and similarity of the  $\text{I}\kappa\text{B}\alpha$ -like gene family (at least in vertebrates), the BLAST hits to human  $\text{I}\kappa\text{B}\alpha$  were subjected to additional filtering before further analysis: hits with least 34% similarity to human  $\text{I}\kappa\text{B}\alpha$  were excluded if they contained a Rel homology domain (as assessed by Scan Prosite analysis using motif IDs PS50254 and PS01204) or lacked at least one N-terminal  $\beta\text{TrCP}$  "phosphodegron" (a regulatory feature of both  $\text{I}\kappa\text{B}\alpha$  and *cactus*) conforming to the sequence DSGxx[S/T/D/E] (Daigneault et al., 2013; Low et al., 2014), and sequences were additionally "reverse BLASTed" to verify that human  $\text{I}\kappa\text{B}\alpha$  was the best match to the hit in the human proteome. This filtering process identified a number of species, both vertebrate and invertebrate, which contained multiple  $\text{I}\kappa\text{B}\alpha$ -like proteins. Following isolation of likely Notch/ $\text{I}\kappa\text{B}\alpha$  homologs, the ARDs of each group were aligned using Clustal Omega. As seen in Figure 4.2A(i), conservation of an  $\text{L}^{\text{-8}}\text{N}$  motif equivalent to Leu1937/Asn1945 in *M. musculus* Notch1 is 100% in species containing FIH, but only 57% in species lacking FIH, as the species from the Nematoda phylum contain either alaninyl or threoninyl residues at the hydroxylation position. The situation is similar for the  $\text{L}^{\text{-8}}\text{N}$  motif equivalent to Leu2004/Asn2012 in *M. musculus* Notch1 (Figure 4.2A(ii)), as in this case, the Nematoda homologs contain aspartyl or alaninyl residues in the target position. It should be noted, however, that evidence exists for hydroxylation of aspartyl residues by FIH (namely Asp695 in human Ankyrin R (Yang et al., 2011b)), although this residue type is hydroxylated with a much lower efficiency. Therefore, it would appear that the presence of FIH is strongly correlated with the existence of an  $\text{L}^{\text{-8}}\text{N}$  motif in a variety of Notch homologs, while in its absence, the residue identities vary more widely. In the case of the  $\text{I}\kappa\text{B}\alpha$ /Cactus homologs, it was found that ARs 4 and 5

(which contain the hydroxylated asparaginyl residues in human  $\text{I}\kappa\text{B}\alpha$ ) had somewhat variable “long loop” lengths between different species. Nonetheless, precedent exists for hydroxylation of extended long loops (e.g. Asn242 in human TRPV3). In addition, it was noted that a number of the sequences contained  $\text{L}^{\delta}\text{N}$  motifs which were (based on primary sequence inspection alone) displaced slightly from that typically observed for known substrates (i.e. Leu at position 21 of the AR, Asn at position 29). These putative “atypical” motifs are indicated by pink text in Figure 4.2B. Among the FIH-containing species, the  $\text{L}^{\delta}\text{N}$  motif equivalent to Leu202/Asn210 in human  $\text{I}\kappa\text{B}\alpha$  (Figure 4.2B(i)) is conserved in 14 out of 16 of the homologs, with 1 of the 2 exceptions (*D. rerio* NP\_998349.2 ( $\text{L}^{\delta}\text{D}$ )) existing in a species which contains a second “conforming” homolog. For the FIH-lacking species, all but one (*B. mori*,  $\text{L}^{\delta}\text{D}$ ), have at least one putative  $\text{I}\kappa\text{B}\alpha$ /Cactus homolog containing an  $\text{L}^{\delta}\text{N}$  motif, suggesting that FIH target motif conservation is quite similar in the 2 species groups at this site. Remarkably, the situation is quite different for the  $\text{L}^{\delta}\text{N}$  motif equivalent to human  $\text{I}\kappa\text{B}\alpha$  Leu236/Asn244. Examination of the sequences from the FIH-containing species revealed that 12 of the 14 organisms have preserved at least one  $\text{L}^{\delta}\text{N}$  motif-containing sequence (Figure 4.2B(ii)). Those that deviate from these amino acids contain either  $\text{L}^{\delta}\text{D}$  (*C. teleta* ELU11258.1 and *Aplysia californica* XP\_005111546.1), or  $\text{L}^{\delta}\text{H}$  (*D. rerio* NP\_955923.1) motifs. Once again, it should be pointed out that, as for aspartyl residues, histidinyl residues are also documented as lower efficiency FIH targets (Yang et al., 2011a). In contrast to the FIH-containing species, only 1 of the 5 FIH-lacking species (*D. melanogaster*) has an  $\text{I}\kappa\text{B}\alpha$ /Cactus homolog containing a possible FIH target motif, with the other sequences coupling a -8 leucinyl or valinyl residue with glycinylnyl, serinylnyl, aspartyl, or glutamyl residues (again, of these, only low efficiency aspartyl residue hydroxylation has been reported in cells). Thus, with the exception of *C. teleta* and *A. californica*, FIH-containing species have maintained an  $\text{I}\kappa\text{B}\alpha$ /Cactus homolog containing the most efficiently hydroxylated FIH target residue, while FIH-lacking species have largely (but not completely) lost this characteristic. In summary, 3 of the 4 target sites in Notch and  $\text{I}\kappa\text{B}\alpha$ /Cactus more frequently contain  $\text{L}^{\delta}\text{N}$  motifs in species with an FIH homolog, a bias that is most apparent for  $\text{I}\kappa\text{B}\alpha$  Leu236/Asn244 equivalents. Given this observation, this site may represent an FIH target which has functional relevance, and indeed, analysis of the functional effects of hydroxylation on  $\text{I}\kappa\text{B}\alpha$  in this work suggests hydroxylation may have subtle effects on  $\text{I}\kappa\text{B}\alpha$  stability (see Chapter 6 for details). Alternatively, if the “ankyrin pool theory” is correct, maintenance of Leu236/Asn244 equivalents as an  $\text{L}^{\delta}\text{N}$  motif in all but 1 of the CAD-containing species may signify that this ARD participates in regulation of FIH availability. Indeed, the ankyrin pool theory proposes that the degree of interaction of FIH with ARD-containing proteins would be widespread and require global adjustment in order to effectively regulate CAD activation. According to the model proposed

A.  
(i)

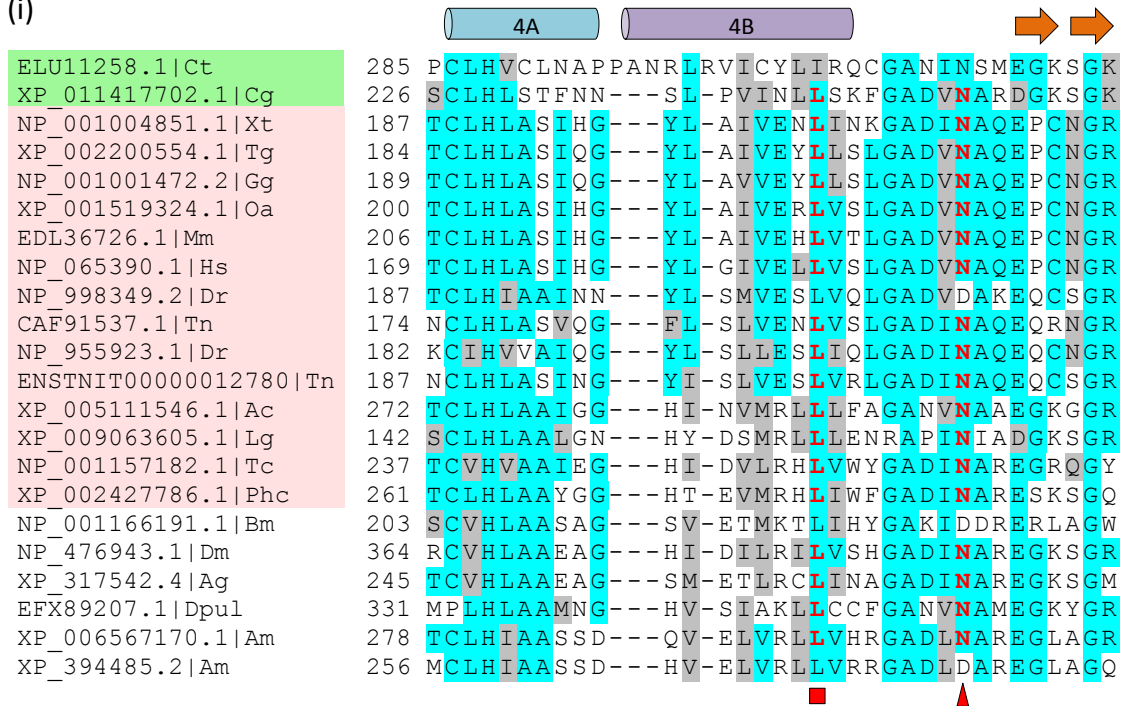


(ii)

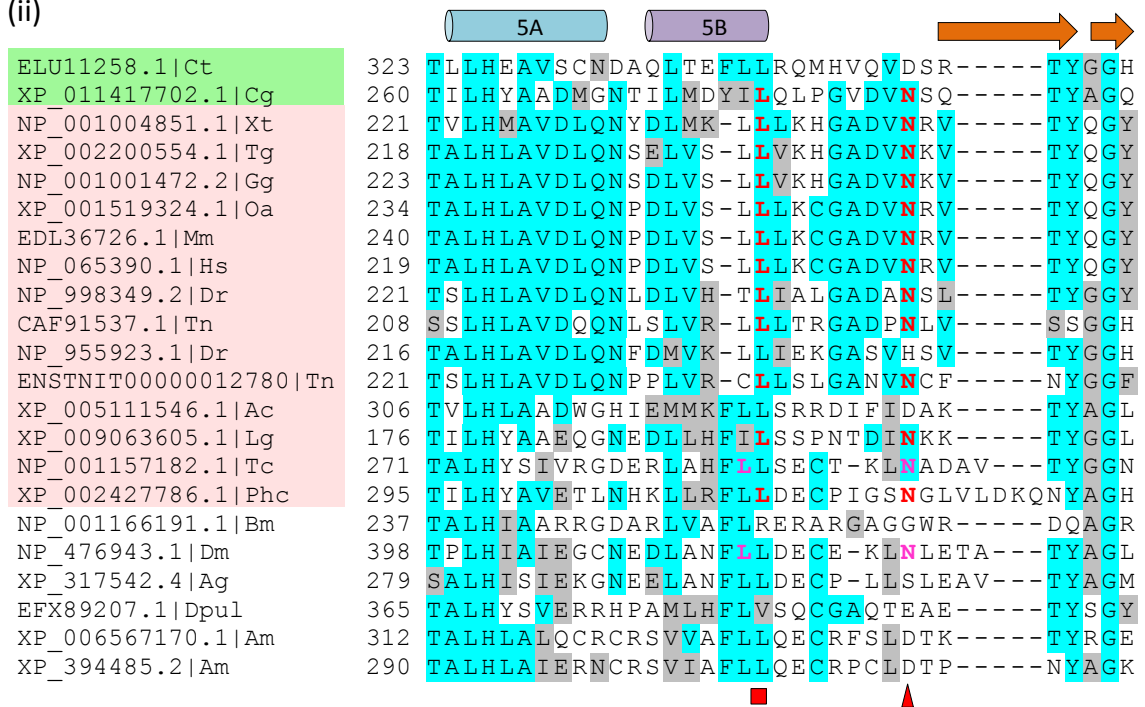


(Figure 4.2)

B.  
(i)



(ii)



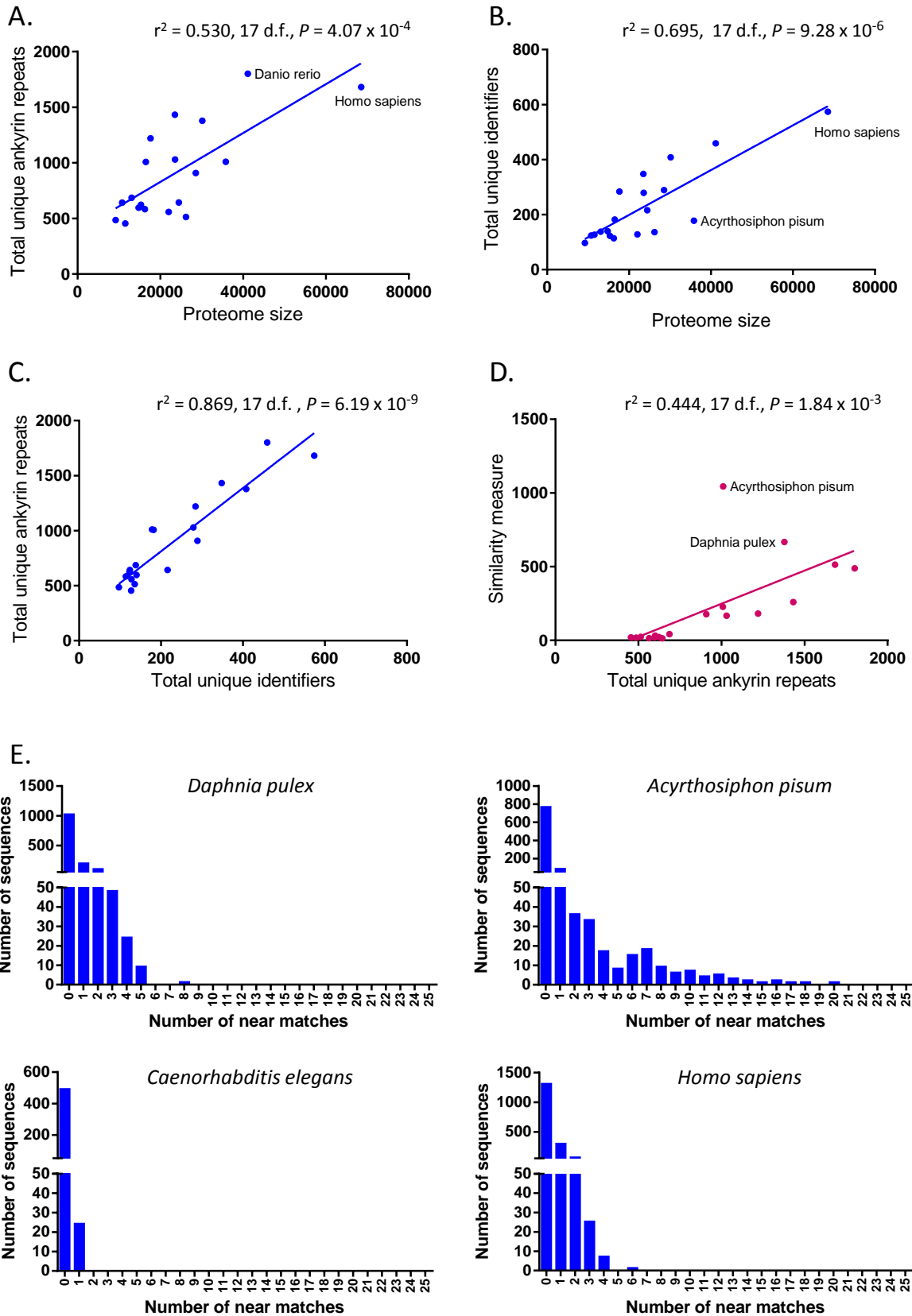
(Figure 4.2)

**Figure 4.2 Conservation of FIH target sequences in Notch and I $\kappa$ B $\alpha$  across species.** (A) Notch AR 2 (i) or AR 4 (ii) from species containing FIH and the CAD (highlighted in pink), FIH alone (green), or neither were aligned using Clustal Omega. Strongly conserved residues are highlighted in cyan, while similar residues are shown in grey. Positions of known target L<sup>-8</sup>N motifs in the mouse or human homologs are indicated below the alignment by a red box (L) and arrowhead (N), while putative target L<sup>-8</sup>N motifs which positionally match these known motifs are shown with red text. Locations of the helices and  $\beta$ -turns which make up each ankyrin repeat (based on published crystal structure 2F8X (Nam et al., 2006)) are indicated above the alignment. Uniprot or Refseq protein IDs for each sequence are shown to the left, as well as the host species. Hs = *Homo sapiens*; Mm = *Mus musculus*; Oa = *Ornithorhynchus anatinus*; Tg = *Taeniopygia guttata*; Gg = *Gallus gallus*; Xt = *Xenopus tropicalis*; Dr = *Danio rerio*; Tn = *Tetraodon nigroviridis*; Bf = *Branchiostoma floridae*; Lg = *Lottia gigantea*; Nv = *Nematostella vectensis*; Ap = *Acyrtosiphon pisum*; Ct = *Capitella teleta*; Ix = *Ixodes scapularis*; Tc = *Tribolium castaneum*; Phc = *Pediculus humanus corporis*; Bm = *Bombyx mori*; Dm = *Drosophila melanogaster*; Ag = *Anopheles gambiae*; Am = *Apis mellifera*; As = *Ascaris suum*; Ce\_LIN = *Caenorhabditis elegans* LIN-12; Ce\_GLP = *Caenorhabditis elegans* GLP-1. (B) AR 4 (i) or AR 5 (ii) from known or predicted I $\kappa$ B $\alpha$ /Cactus homologs were aligned and annotated as for (A). Ankyrin repeat secondary structure is shown based on crystal structure 1IKN (Huxford et al., 1998). A number of the sequences contained L<sup>-8</sup>N motifs that were slightly offset compared to known target motifs in other proteins of the alignment, and these are shown in pink text. Species abbreviations are as for (A) with the addition of: Cg = *Crassostrea gigas*; Dpul = *Daphnia pulex*.

by Schmierer et al (2010), the presence of ARD-containing substrates within cells has a number of theoretical “benefits”, including providing a means for focussing CAD activation into a defined range of oxygen tensions, and rendering CAD activation particularly sensitive to a certain oxygen concentration threshold (Schmierer et al., 2010). The values of these oxygen ranges and thresholds were dependent on a number of variables, among them ARD-containing protein concentration, their affinity for FIH, as well as their efficiency of hydroxylation. By this reasoning, maintaining a “suitable” pool of ankyrin repeats to facilitate this regulation may include large-scale adjustment of ankyrin repeat sequence composition, thereby altering the kinetics of FIH-ARD (and by extension FIH-CAD) interactions. To determine if conservation of an FIH and CAD homolog has a detectable influence on ankyrin repeat primary sequence, it was decided to analyse these sequences on a proteomic scale in a variety of metazoan species.

#### 4.3.1 Proteome-wide analysis of ankyrin repeat conservation

Based on the analysis shown in Figure 4.1, species either with or without FIH and CAD homologs were chosen from a range of phyla (indicated by black circles in Figure 4.1), and their proteomes searched for ankyrin repeats using 3 different methods: (1) a Scan Prosite-based search using Prosite motif PS50088, (2) HMMER3.0 analysis utilising the ankyrin repeat HMM from the Pfam database (Pfam accession PF00023), and (3) direct download of ankyrin repeat entries for each species from the SMART database. Although all species selected had undergone complete genome sequencing, these genomes vary in terms of their level of assembly and annotation. Thus, to assess the quality of the ankyrin repeat data retrieved, proteome size was compared with the total number of unique ankyrin repeats isolated for each species, revealing a significant positive correlation ( $r^2 = 0.530$ ,  $P = 4.07 \times 10^{-4}$ ) (Figure 4.3A). A similar significant positive correlation was found between proteome size and the number of unique identifiers retrieved (Figure 4.3B), while the average number of repeats per ARD was found to be highly consistent across species (Figure 4.3C). As an additional quality control step, each species' ankyrin repeat repertoire was subjected to a "similarity measurement" to identify repeat sets with an unusually high level of near matches, which may reflect inaccurate genome assembly. For this, each unique ankyrin repeat in a species' set was compared with every other, and the number of similar repeats with 4 or fewer amino acid mismatches (an arbitrarily chosen cut-off) recorded for each sequence. The scores for each repeat were then summed to give an overall score for the species' repeat repertoire, and plotted against the total number of



(Figure 4.3)

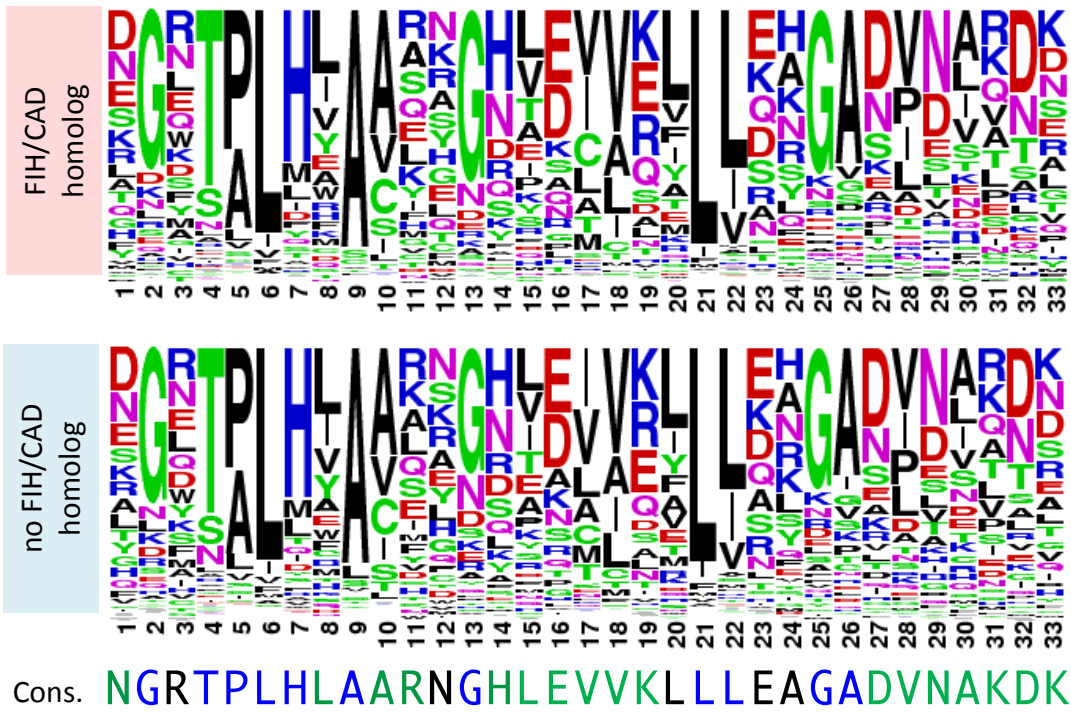


**Figure 4.3 Analysis of ankyrin repeat repertoire characteristics across species.** (A) Comparison of the relationship between proteome size (obtained from Uniprot) and the number of unique ankyrin repeat sequences identified for each species. The trendline was fitted using Prism's straight line least squares equation, while  $r^2$  was calculated using Excel's FDIST function (19 degrees of freedom (d.f.)). (B) A significant positive correlation is seen between the proteome size of each species and the number of unique ankyrin repeat-containing identifiers for each species (intended as close approximation of the number of ARD-containing proteins). (C) The average number of ankyrin repeats per ARD-containing protein was analysed by plotting the number of unique ankyrin repeats for each species against the total unique identifiers found, revealing a significant, strong correlation between these values. (D) A "similarity measure" was first calculated for each species' repertoire of ankyrin repeats using Excel's Fuzzy Lookup Add-In, and then these values compared with the number of unique ankyrin repeats identified for each species. (Trendlines and  $r^2$  values were calculated as for (A) (18 d.f.)). (E) Histograms documenting the number of "near matches" for each sequence in the *D. pulex*, *A. pisum*, *C. elegans* and *H. sapiens* ankyrin repeat repertoires.

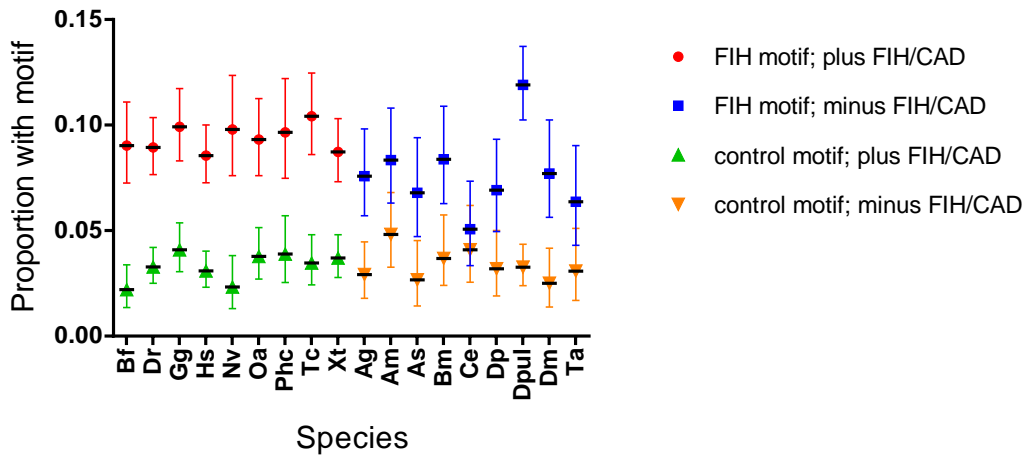
unique ankyrin repeats for each species (Figure 4.3D). Analysis of these data showed a significant positive correlation between the data points ( $r^2 = 0.444$ ,  $P = 1.84 \times 10^{-3}$ ), with only *Acyrtosiphon pisum*, varying substantially from the observed trend (see also Figure 4.3E for example similar sequence frequency distributions). As it was unable to be determined if this deviation was due to the presence of genuine gene homologs or greater erroneous assembly of sequencing reads in this species, the data from *A. pisum* was removed from further analyses, as this repetitiveness may skew future analyses of species-specific ankyrin repeat composition.

Next, ankyrin repeat sequences from the remaining 18 species were grouped into those from species with (10366 repeats total) or without FIH and the CAD (5860 repeats total) to facilitate sequence alignment and comparison of FIH target sequence prevalence. For the alignment, sequence gaps and insertions predicted by the Scan Prosite and HMMER3.0 search algorithms were included or deleted, respectively, with the exception of those predicted to occur between aa 21 and 29 of the canonical ankyrin repeat, as such an alteration may provide a false representation of the aa sequence in the FIH binding region. Visualisation of these alignments using the WebLogo sequence logo generator (Crooks et al., 2004) suggested only minor differences in amino acid frequencies across the 2 groups (Figure 4.4A). To gain a quantitative measure of the appearance of an FIH target motif in FIH/CAD-containing and lacking species, the frequency of the aa sequence LXXXXX[D/E][V/I/A]N occurring at

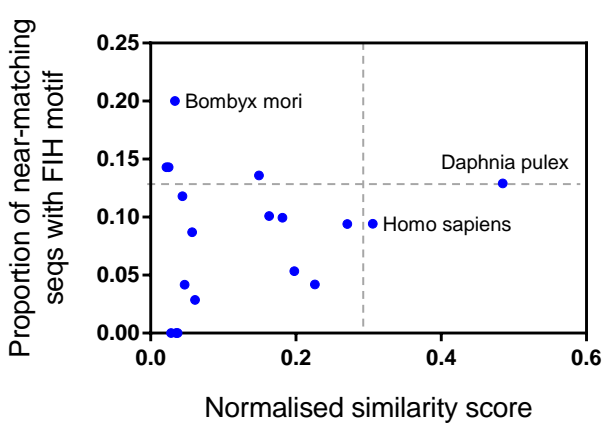
A.



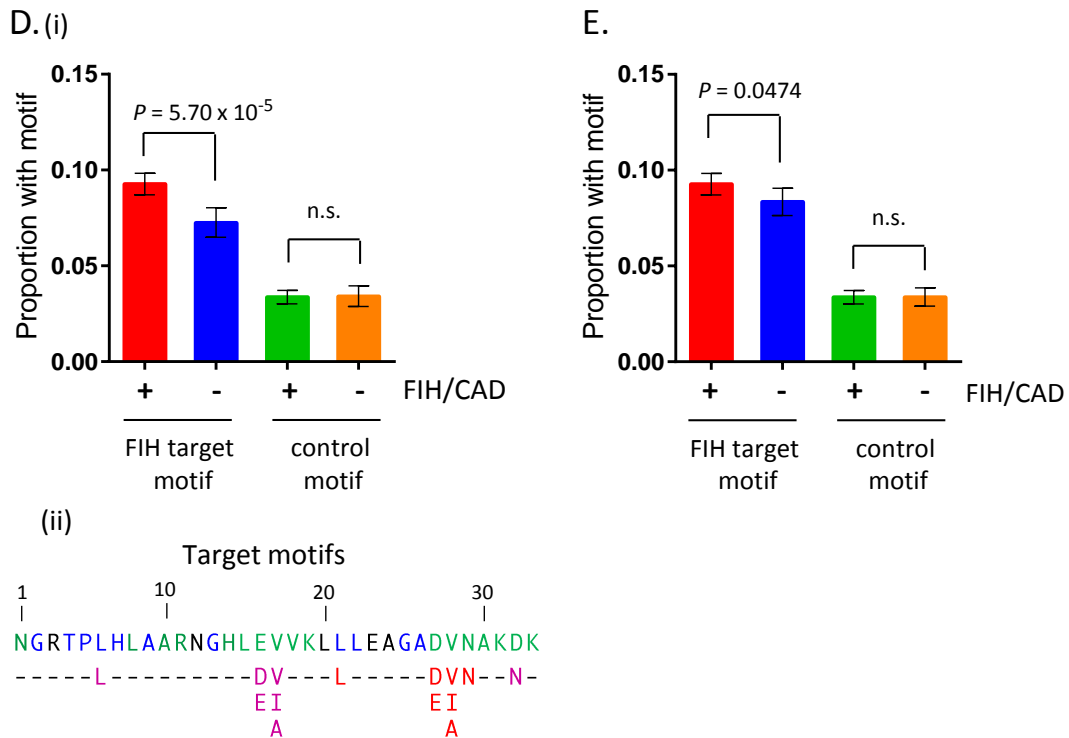
B.



C.



(Figure 4.4)



**Figure 4.4 Proteome-wide analysis of FIH target motif conservation.** (A) Unique repeats identified for FIH-containing and -lacking species were analysed using the WebLogo Sequence logo generator, and the results represented using a frequency plot, where the height of each letter within the stack corresponds to the relative abundance of the amino acid at that position. The Mosavi et al (2004) consensus is shown below the frequency plots, and is coloured as for Figure 3.5 (Mosavi et al., 2004). (B) The proportion of FIH target or control motifs (see parts (D) and (E)) in different species containing or lacking FIH/CAD equivalents is shown, with error bars representing 95% confidence intervals. Species abbreviations as for Figure 4.2, plus Ta = *Trichoplax adharens*, Dp = *Danaus plexippus*. (C) Comparison of the contribution of “near-match” ankyrin repeats to overall calculation of FIH motif frequency. Each species’ normalised similarity score (the similarity score calculated in Figure 4.3D normalised to the total number of ankyrin repeats for that species – higher numbers indicate a greater influence of ankyrin repeat “repetitiveness” on amino acid frequencies observed in the ankyrin repeat repertoire) was plotted against the proportion of near-matching sequences which were found to contain an FIH target motif. Species with higher values for both measurements are more likely to have a higher proportion of FIH target motifs due to a high level of gene paralogs or erroneous genome assembly. (D)(i) Ankyrin repeat repertoires for each species were first assessed for the frequency of appearance of an FH target or control motif (indicated in (ii) relative to the Mosavi et al (2004) consensus in red or purple text, respectively). The observations were then pooled into FIH-containing or FIH-lacking species (with the exception of *D. pulex* (see main text for details)), and then analysed for significant differences in motif frequency using Person’s  $\chi^2$  statistic. See also Table 4.1 for parameter details. Data represent motif frequency of pooled datasets +/- 95% confidence intervals. n.s. = not significant. (E) Analysis of FIH target or control motif abundance as for (D), except that the *D. pulex* dataset was included in the pooled observations for the FIH-lacking subgroup.

positions 21-29 of the 33 aa repeat (shown in Figure 4.4D(ii)) was first calculated for each individual species. Next, it was sought to compare these motif frequencies both in terms of (1) their overall level of variation within the FIH/CAD-containing species as well as the FIH/CAD-lacking species, and (2) the comparative frequency of the motif in the FIH/CAD-containing species vs the FIH/CAD-lacking species. If the presence of FIH/CAD homologs influences a species' ankyrin pool, it could be envisaged that the prevalence of an FIH target motif would be maintained within a certain range in order to achieve effective CAD regulation. The species without an FIH/CAD system, however, would be free from this restriction, thus permitting ankyrin repeat sequences to be shaped by other variables, which may be reflected by a different typical frequency of FIH motif in these species, or perhaps a wider range of frequencies. To gain a measure of the heterogeneity of FIH motif prevalence within the FIH/CAD-containing species alone or FIH/CAD-lacking species alone, the calculated frequencies were compared using Pearson's  $\chi^2$  statistic. For the FIH/CAD-containing species, the calculated  $\chi^2$  value returned a probability of 0.827, suggesting that the FIH motif frequencies were quite similar, while for the FIH/CAD-lacking species, the probability was far lower, at  $1.592 \times 10^{-5}$  (Table 4.1), implying that the frequencies were more widely spread. However, removal of a single organism from the FIH/CAD-lacking group, namely *Daphnia pulex*, vastly altered the probability of obtaining the calculated  $\chi^2$  value to 0.436, suggesting that this dataset is an "outlier" from an otherwise narrowly distributed set of frequencies (graphically depicted in Figure 4.4B). Whether or not this deviation is artefactual is not clear. *D. pulex* is among the species with a higher "similarity measure" relative to the size of its ankyrin repeat repertoire (see Figure 4.3D and E), and was also among the species with a higher relative proportion of FIH-motif containing ankyrin repeats among its near-matching sequences (Figure 4.4C). Thus, if the "over-represented" repeats are a result of inaccurate genome assembly this may have contributed to the unusually high frequency observed. On the other hand, *D. pulex* was noted to have a rate of gene duplication approximately 3 times higher than that observed in fly, nematode or humans (Colbourne et al., 2011), suggesting the inflation may instead be due to the existence of genuine gene homologs. As the current data do not permit distinction between these different scenarios, the frequency variation across FIH/CAD lacking species could be interpreted in several different ways. Firstly, if the motif frequency observed in *D. pulex* is genuine, this organism is a clear indication that typical aa frequencies within the canonical ankyrin repeat can (at least in non-FIH/CAD-containing species) vary substantially. If spurious, it would appear that FIH motif frequency is maintained in a narrower window in both FIH/CAD-containing and -lacking groups. With regards to the second possibility, it was next sought to determine if the frequencies observed across the 2 groups differed significantly.

Comparison	motif	$\chi^2$	d. f.	P value
FIH-containing-species	FIH target	4.318	8	0.827
FIH-lacking species (all)	FIH target	36.233	8	1.592 x 10 <sup>-5</sup>
FIH-lacking species (- <i>D. pulex</i> )	FIH target	6.936	7	0.436
FIH-containing species (pooled) and FIH-lacking species (pooled, - <i>D. pulex</i> )	FIH target	16.213	1	5.659 x 10 <sup>-5</sup>
FIH-containing species (pooled) and FIH-lacking species (pooled)	FIH target	3.933	1	0.0474
FIH-containing species (pooled) and FIH-lacking species (pooled, - <i>D. pulex</i> )	control	0.0113	1	0.915
FIH-containing species (pooled) and FIH-lacking species (pooled)	control	0.00025	1	0.987

**Table 4.1 Analysis of FIH and control motif frequency in species subgroups.**

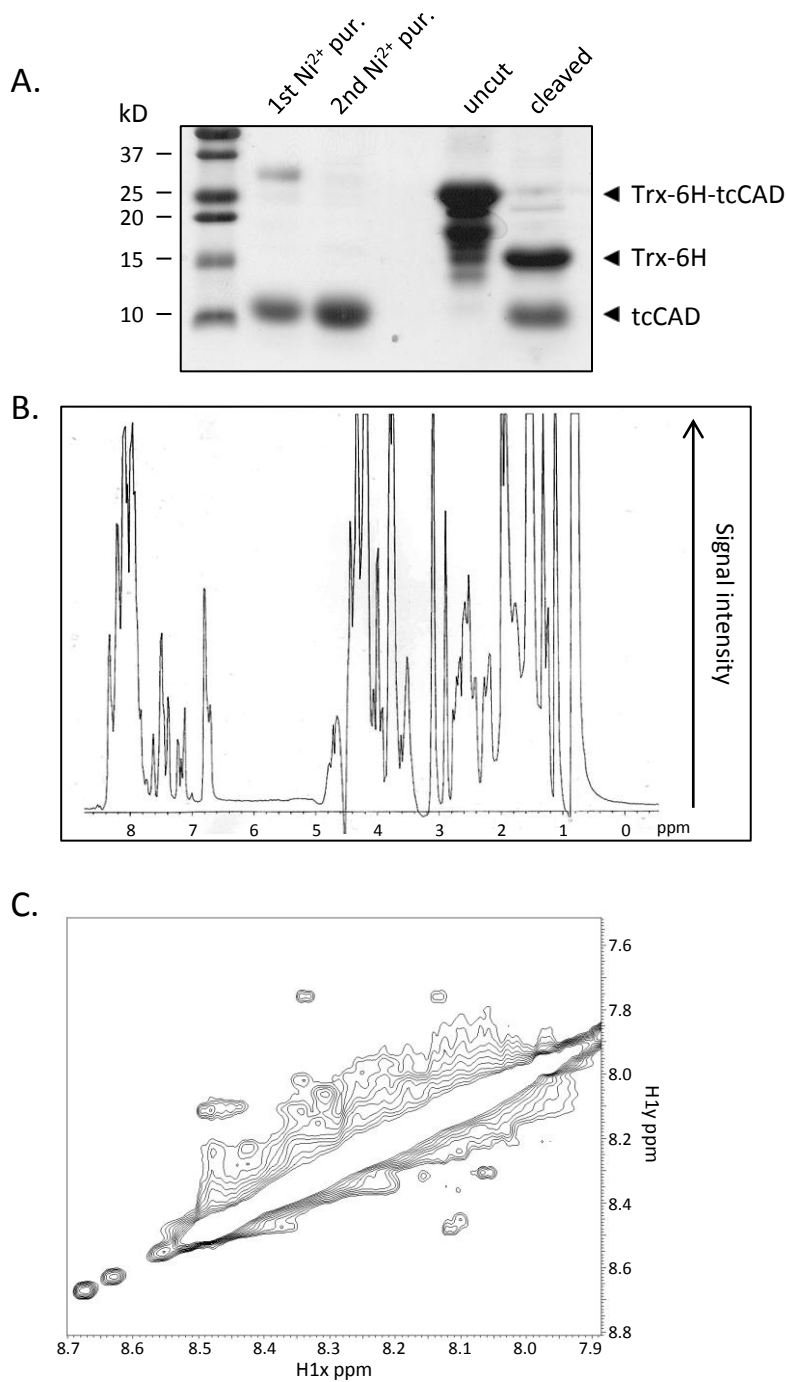
d.f. = degrees of freedom

Since the heterogeneity of FIH motif prevalence within FIH/CAD-containing and -lacking groups was found (with the exception of *D. pulex*) to be reasonably small, it was decided that pooling of frequency observations within the 2 groups would provide a good approximation of FIH motif prevalence for each, thus facilitating a simplified cross-group comparison. Analysis of cross-group FIH target motif frequency minus the *D. pulex* dataset revealed a statistically significant decrease in the species lacking FIH (Table 4.1), although the size of this reduction was only 1.28-fold (Figure 4.4D(i)). Notably, the difference in FIH motif frequency between species pools was still significant following inclusion of the *D. pulex* dataset, but only marginally so (Figure 4.4E). A similar analysis was then carried out for a control motif comprising Leu at position 6, Asp/Glu at position 16, Val/Ile/Ala at position 17, and Asn at position 32 of the 33 aa ankyrin repeat (see Figure 4.4D(ii)), positions chosen based on their similar aa type and level of conservation to the FIH target motif. Examination of the heterogeneity of observed frequencies between either FIH/CAD-containing species alone or FIH/CAD-lacking species alone returned non-significant *P*-values irrespective of the inclusion or exclusion of the dataset from *D. pulex* (Table 4.1), suggesting that pooled datasets would be a good approximation of control motif frequency in FIH/CAD-containing and -lacking species. However, unlike for the FIH target motif, cross-group control motif frequency was not significantly different (Table 4.1 and Figure 4.4D(ii)), and inclusion of the *D. pulex* dataset had little impact on this result

(Figure 4.4E). In summary, current data suggest that the majority of FIH/CAD-containing and -lacking species have a comparable within-group FIH target motif prevalence, while cross-group analysis suggests that the observed frequencies are slightly but significantly lower in species lacking an FIH/CAD system. Moreover, this frequency decrease is specific to the region of the repeat known to interact with FIH.

#### 4.4 The *T. castaneum* CAD is not stably folded in solution

In addition to considering the functional interplay between FIH, ARDs and HIF-1 $\alpha$ , the question of how FIH evolved to hydroxylate such structurally distinct substrates (see section 3.6) remains unanswered. One possibility is that, despite the observed structural disparity between the human homolog of the HIF- $\alpha$  CAD and the classical ankyrin fold, the structures of more ancient homologs of the two classes may have been more similar. That is, earlier incarnations of the HIF- $\alpha$  CAD may have been ankyrin repeat-like in structure, or at the very least, may have contained significant  $\alpha$ -helical content reminiscent of that observed when HIF-1 $\alpha$  is bound to FIH or p300 (Dames et al., 2002; Elkins et al., 2003; Freedman et al., 2002). To test this possibility, the model organism *Tribolium castaneum* was utilised. *T. castaneum* has an FIH homolog which is capable of promoting 2-OG decarboxylation in the presence of both tcHIF- $\alpha$  or tcNotch (B. Davenport, unpublished data). However, Coleoptera diverged from a common ancestor with humans approximately 780 million years ago (compare this with humans, which diverged from a common ancestor with mice approximately 90 million years ago (Hedges et al., 2006)), and as such, it was hypothesised that if the folding properties of the HIF- $\alpha$  CAD had changed during evolution, some of this change may be evident in this more ancient species. To investigate this idea, the last 90 aa of tcHIF- $\alpha$  was expressed in bacteria and purified by Ni<sup>2+</sup> affinity chromatography (Figure 4.5A) and subsequently analysed via 1D <sup>1</sup>H and 2D NOESY NMR. Assessment of the 1D spectrum showed that there was minimal dispersion of amide proton peaks, suggesting that stable structure formation within the *T. castaneum* CAD is unlikely (Figure 4.5B). However, preliminary 2D NOESY analysis showed



**Figure 4.5**  $1D$   $^1H$  and  $2D$  NOESY NMR analysis of *Tribolium castaneum* HIF- $\alpha$  CAD. (A) Samples of protein from the tcHIF- $\alpha$  CAD (aa 780-879) purification process were separated by SDS PAGE and then stained with Coomassie. Trx-6H-TEV-tcHIF- $\alpha$  CAD (780-879) fusion protein was expressed and purified from bacteria (lane 5) and then incubated with AcTEV protease (Life Technologies) to facilitate cleavage of the Trx-6H tag (lane 6). The cleaved protein was then subjected to 2 further rounds of  $Ni^{2+}$  affinity purification (lanes 2 and 3) for removal of the Trx-6H tag and the His-tagged AcTEV protease. The purified CAD protein was then subjected to  $1D$   $^1H$  NMR analysis (B) or  $2D$  NOESY (C) for assessment of secondary structure content.

some evidence for Nuclear Overhauser Effects (NOEs) within the sample, particularly between amide protons (Figure 4.5C), raising the possibility that some regions of the tcCAD may have some level of secondary structure formation. Nonetheless, the peaks were quite weak, thus making firm conclusions about structural content difficult. Even so, the lack of stable  $\alpha$ -helical content within the tcHIF- $\alpha$  CAD would suggest that the tcFIH enzyme is not limited to targeting proteins which present a stably folded “ARD-like” binding interface. Furthermore, if the CAD and ankyrin repeats were initially structurally more similar, it can be assumed that a change in the CAD structure towards random coil must have occurred prior to divergence of the Coleoptera order.

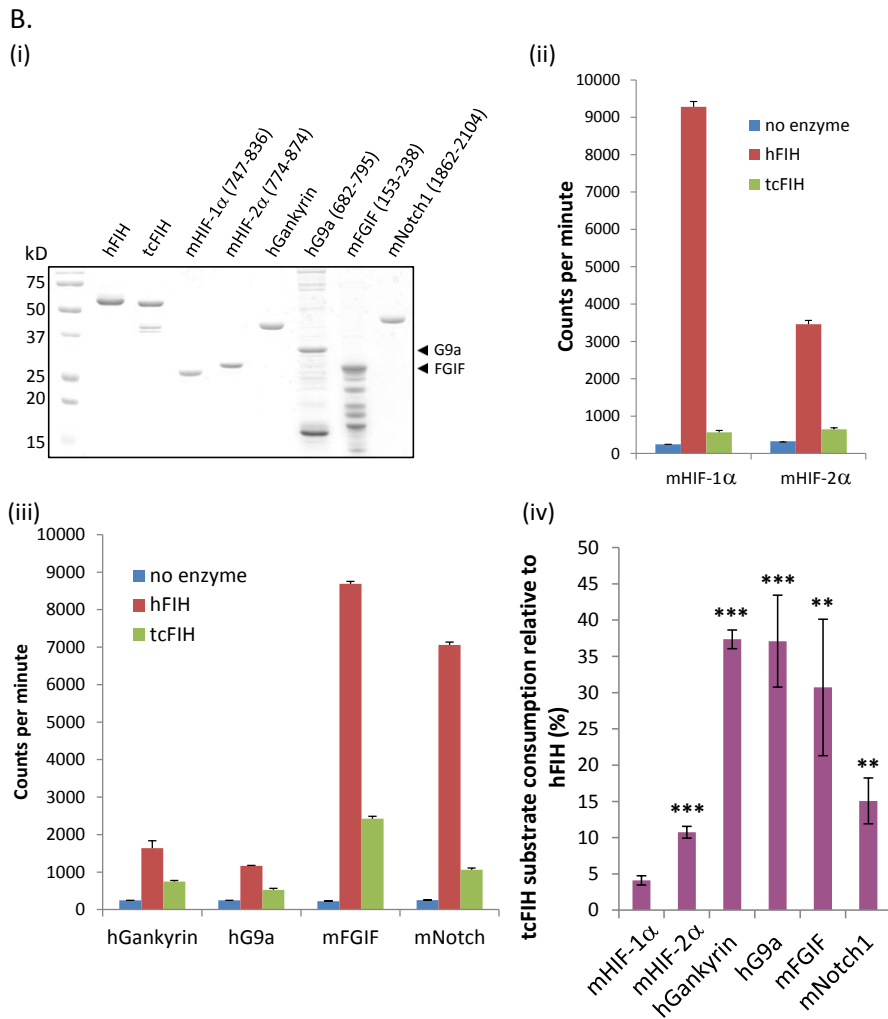
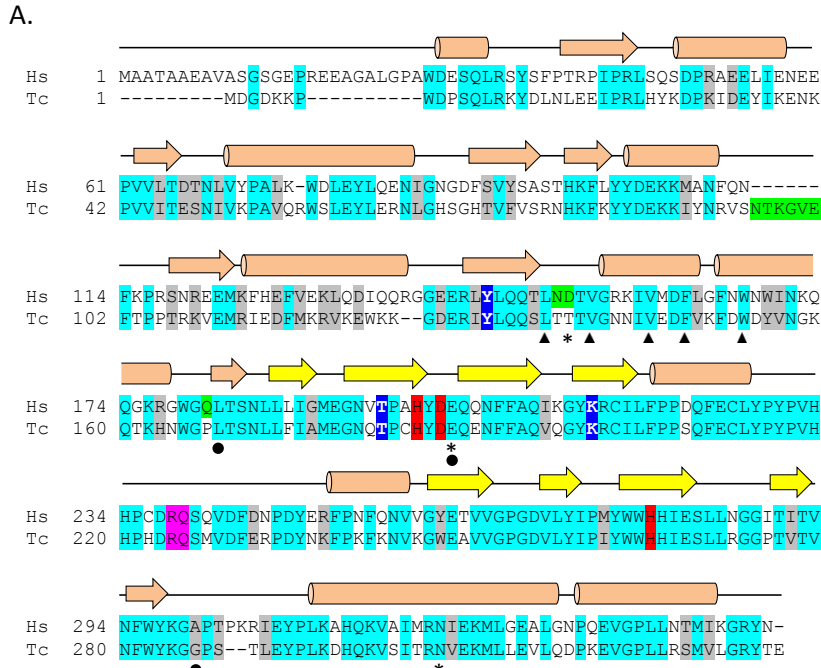
#### 4.5 *T. castaneum* FIH more efficiently targets mammalian ankyrin repeats over the CAD compared to hFIH

In addition to understanding the role of secondary and tertiary structure in FIH substrate recognition, the molecular details required in the primary structure are also of considerable interest. Discovery of numerous FIH substrates has permitted the definition of an “FIH preferred sequence” for FIH hydroxylation (see Figure 3.5). However, FIH substrate recognition may well be considerably more complex than the crystal structure of FIH bound to the HIF- $\alpha$  CAD (or indeed FIH bound to ankyrin repeat peptides) would suggest (Coleman et al., 2007; Elkins et al., 2003; Yang et al., 2013). The fact that FIH can bind folded ankyrin repeat domains (see Figure 3.11C), perhaps as a precursor to unwinding and hydroxylation, is one example of this. The involvement of the “di-leucine motif” in HIF- $\alpha$  hydroxylation (at least under conditions of molecular crowding, see section 1.5.1) is another. Thus, much remains to be clarified before the efficiency of HIF- $\alpha$  and ARD hydroxylation by FIH can be predicted with any certainty. With this in mind, investigation of *T. castaneum* FIH has serendipitously provided new avenues of investigation. Preliminary studies with *T. castaneum* FIH revealed that its substrate repertoire includes both tcHIF- $\alpha$  and tcNotch. However, analysis of cross-species substrate recognition by tcFIH produced some unexpected findings. Whilst tcFIH displayed similar activity towards murine or *T. castaneum* Notch, activity in the presence of human and *T. castaneum* HIF homologs was noticeably different, with tcHIF- $\alpha$  CAD promoting approximately 8 times more 2-OG decarboxylation than hHIF-1 $\alpha$  CAD by *in vitro* hydroxylation assay (B. Davenport, unpublished findings). Intriguingly, human FIH showed no such bias, with human and *T. castaneum* HIF- $\alpha$  CAD homologs promoting similar activity from hFIH. Thus, despite a high degree of similarity between the two enzymes (58% sequence identity, 75% similarity, Figure 4.6A), particularly in residues predicted to be involved in catalysis, these findings suggest that the substrate specificity of



human and tcFIH is somewhat different. To explore these observations further, attempts were made to express and purify hFIH and tcFIH concurrently, so that the substrate specificity of the two enzymes could be compared side by side. However, during initial tests with tcFIH, it was found that the absolute activity of the enzyme was somewhat variable between protein preparations, which may reflect incomplete folding of the protein when expressed in *E. coli*, variations in handling during the purification process, or the affinity tag used. For example, MBP-tagged tcFIH was found to have greater absolute activity than Trx-6H-tagged tcFIH (data not shown). However, to minimise differences in manual handling of hFIH and tcFIH during the purification process, it was decided to use Trx-6H-tagged enzymes which could be purified by the automated Profinia protein purification system (see Materials and Methods for details). It was also found that use of the Trx-6H-tagged tcFIH allowed for far more efficient purification of tcFIH than the MBP-tagged version of the enzyme, although it was found that high concentrations of Trx-6H-tagged tcFIH would precipitate during storage. This problem could be circumvented, however, if the enzyme was diluted immediately post preparation.

Once suitable purification conditions were established, the activity of the two enzymes was compared in the presence of bacterially expressed and purified Trx-6H-tagged substrates from both HIF and ARD protein classes (Figure 4.6B(i)). Example experiments utilising Trx-6H-tagged human and *T. castaneum* FIH are shown in Figure 4.6B(ii) and (iii), wherein, at least under the shown assay conditions, tcFIH showed considerably less absolute activity in the presence of all substrates tested than human FIH. More importantly, however, irrespective of the absolute activity of the tcFIH and hFIH enzymes, tcFIH activity normalised to that of hFIH was always lowest when hydroxylating mHIF-1 $\alpha$  (Figure 4.6B(iv)). (For technical purposes, most replicates of the experiments in part (ii) and (iii) tested the HIF and ARD substrates separately (but with the same batch of enzyme) and with differing pH conditions (pH 7.0 for HIF, pH 7.5 for ARDs) to optimise hydroxylation conditions (Linke et al., 2007). However, HIF-1 $\alpha$  and Notch1 were tested side by side and using the same pH (pH 7.5) on multiple occasions (see Figure 4.7B(i), and data not shown) to verify that comparison of tcFIH/hFIH



(Figure 4.6)

**Figure 4.6 Differential substrate specificity of human and *Tribolium* FIH.** (A) Alignment of hFIH (Hs) and tcFIH (Tc). Secondary structure of hFIH ((Elkins et al., 2003)) is shown above the sequence. Yellow arrows indicate beta strands that make up the DSBH. Conserved or similar residues are shown in turquoise and grey, respectively, and residues required for iron chelation (red), 2-OG coordination (dark blue), target asparagine orientation (pink) are highlighted. Residues which contribute side chain atoms to H-bonding of the HIF-1 $\alpha$  CAD, or are part of the hydrophobic pocket proposed to bind “site 2” of HIF-1 $\alpha$  ((Elkins et al., 2003)) are indicated by an asterisk or closed triangles, respectively, below the alignment. Residues mutated to generate “*Tribolium* mimic” mutants are shown in green. (B)(i) Using the relevant pET32a plasmid constructs, Trx-6H-tagged versions of mHIF-1 $\alpha$  (747-836) and mHIF-2 $\alpha$  (774-874), hGankyrin, hG9a (682-795), mFGIF (153-238) and mNotch1 (1862-2104) were expressed and purified from bacteria using Ni<sup>2+</sup> affinity chromatography and analysed by SDS PAGE for stability and purity. Both G9a (35.3 kDa) and FGIF (27.9 kDa, indicated by arrowheads) are consistently found to have poor stability post purification. (ii) and (iii) Proteins were then analysed by *in vitro* hydroxylation assay with bacterially expressed Trx-6H-hFIH or Trx-6H-tcFIH using the following substrate concentrations: 76  $\mu$ M mHIF-1 $\alpha$  (747-836), 76  $\mu$ M mHIF-2 $\alpha$  (774-874), 48  $\mu$ M hGankyrin, ~12  $\mu$ M hG9a (682-795), ~51  $\mu$ M mFGIF (153-238), and 73  $\mu$ M mNotch1 (1862-2104). For HIF, FGIF and Notch, reactions contained 200 nM hFIH or tcFIH. For G9a and Gankyrin, reactions contained 500 nM hFIH or tcFIH. Each graph is an example experiment (see main text) where data is represented as the mean +/- SD of samples tested in triplicate (hFIH and tcFIH samples), or duplicate (no enzyme samples). (iv) The results from (ii) and (iii) are expressed as substrate consumption by tcFIH as a percentage of that of hFIH (mean +/- SEM, n = 3). A student’s 2-tailed t-test was employed to compare the SEM of HIF-1 $\alpha$  with the other substrates. \*\* p < 0.01; \*\*\* p < 0.001.

activity across the HIF and ARD substrate-containing experiments was valid.) In very broad terms, it would seem that tcFIH substrate specificity (at least for mammalian ARD and HIF substrates) is geared further towards ankyrin repeat hydroxylation than that of hFIH. Analysis of tcFIH substrate recognition may therefore have two distinct benefits: (1) extending knowledge of the molecular details of FIH substrate hydroxylation, and (2) permitting the engineering a “tcFIH-like” enzyme which could preferentially target ARDs over HIF-1 $\alpha$ , which in itself may prove a valuable biochemical tool.

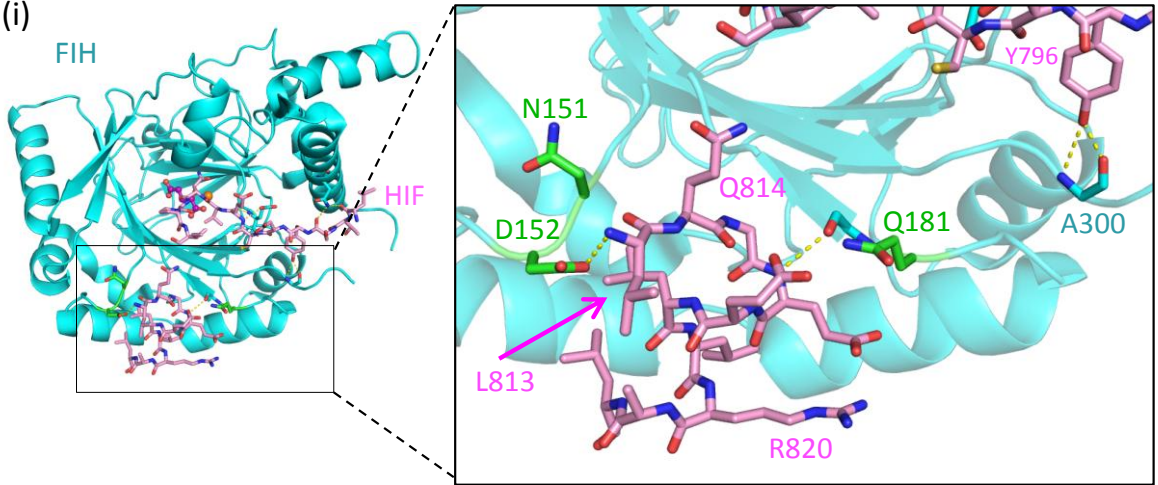
#### 4.6 Substrate specificity of tcFIH cannot be conferred to hFIH through binding pocket-targeted tcFIH “mimic” mutations

In efforts to better understand tcFIH’s reduced ability to target hHIF-1 $\alpha$ , the substrate binding regions of human and tcFIH, as well as residues involved in catalysis, were analysed in more detail. Residues predicted to be involved in Fe<sup>2+</sup> coordination (His199, Asp201 and His279 (numbers refer to

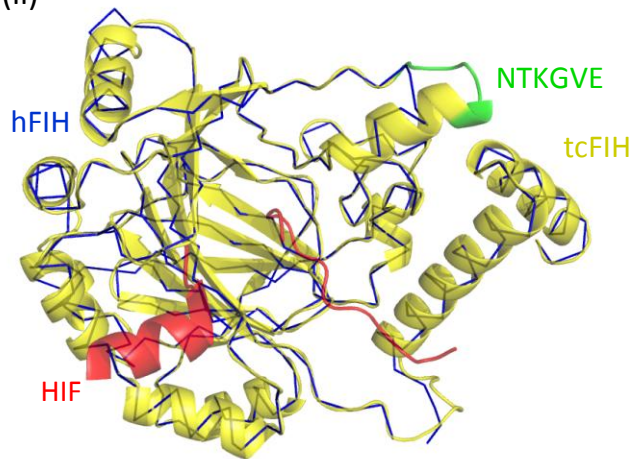
human FIH)), 2-OG binding (Tyr145, Thr196, Lys214), or enzyme dimerisation (Elkins et al., 2003) were well conserved (Figure 4.6A), thus attention was focussed instead on the FIH-CAD interaction. The crystal structure of hFIH bound to a human CAD fragment indicates that a total of 10 H-bonding contacts are likely across two sites of interaction (site 1 and site 2), with an additional hydrophobic patch at site 2 suggested to contribute to binding affinity ((Elkins et al., 2003), and see Figure 4.6A for the location of these residues in the FIH sequence). In agreement with the observed flexibility in FIH substrate choice, a majority of the H-bonding contacts made between FIH and the CAD are mediated by backbone atoms. Nonetheless, a small number of FIH and HIF side chains are among the proposed H-bond donors/acceptors ((Elkins et al., 2003), Figure 4.6A and Figure 4.7C(i)). Among hFIH's predicted H-bonding side chains, only Asp152 is not conserved in tcFIH, although the replacement residue, Thr138, may still be capable of acting as an H-bond donor/acceptor. During the search for differing binding pocket residues between human and tcFIH, it was also noted that Gln181 is a prolyl residue in tcFIH. Although Gln181 has not been directly implicated in CAD binding, it is situated in close proximity to site 2, and it was also hypothesised that, due to proline's unique structure, its introduction may subtly perturb the FIH surface in this region, thus altering CAD binding. Finally, it was observed that there was a six residue "NTKGVE" insertion after Asn113 of human FIH, which is predicted to extend the C-terminal end of  $\alpha$ -helix 4 (Figure 4.6A, and Figure 4.7A(ii)). Although not situated near to the substrate binding pocket of the FIH monomer, the sequence could potentially be in close proximity to substrate proteins bound on the opposite side of the FIH dimer. In order to evaluate the importance of these residues in FIH substrate recognition, a mutant human enzyme was constructed (hFIH NTKGVE/TTP) which contained the NTKGVE insertion, as well as conversions of hFIH Asp152 (plus the adjacent non-conserved residue, Asn151) and Gln181 to their corresponding residues in tcFIH. If hFIH NTKGVE/TTP has a more "tcFIH-like" substrate selectivity, it would be expected that the activity of hFIH NTKGVE/TTP relative to hFIH would be greater for the ARD-containing substrate than for mHIF-1 $\alpha$ . To test this idea, the hFIH NTKGVE/TTP enzyme was expressed in bacteria with a Trx-6H tag and purified (Figure 4.7B(i), and its activity relative to Trx-6H-tagged hFIH and tcFIH compared in a hydroxylation assay. As can be seen in Figure 4.7B(ii), the absolute activity of hFIH NTKGVE/TTP was intermediate between that of hFIH and tcFIH for both mHIF-1 $\alpha$  and mNotch1 substrates. However, this differential activity of hFIH NTKGVE/TTP relative to hFIH was the same for both substrates, unlike tcFIH, which again showed greater hFIH-normalised activity towards mNotch1 than mHIF-1 $\alpha$  (Figure 4.7B(iii)). In summary, these non-conserved residues between hFIH and tcFIH are not responsible (at least when mutated in isolation) for the differential substrate specificity of the two enzymes.

A.

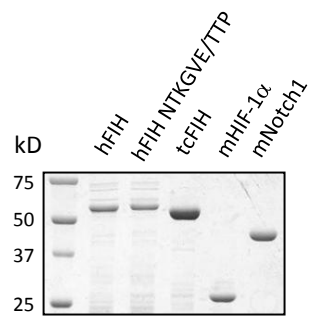
(i)



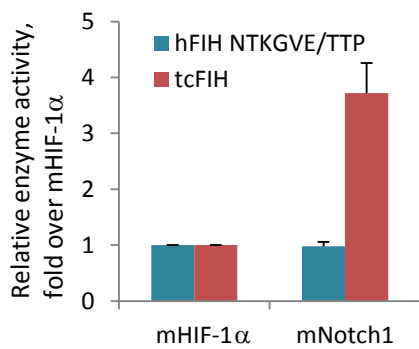
(ii)



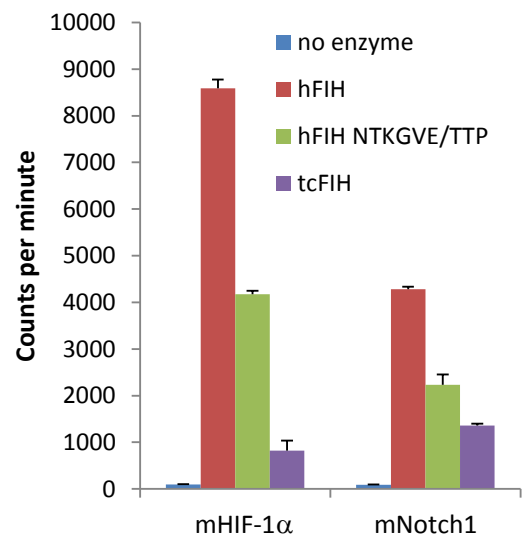
B. (i)



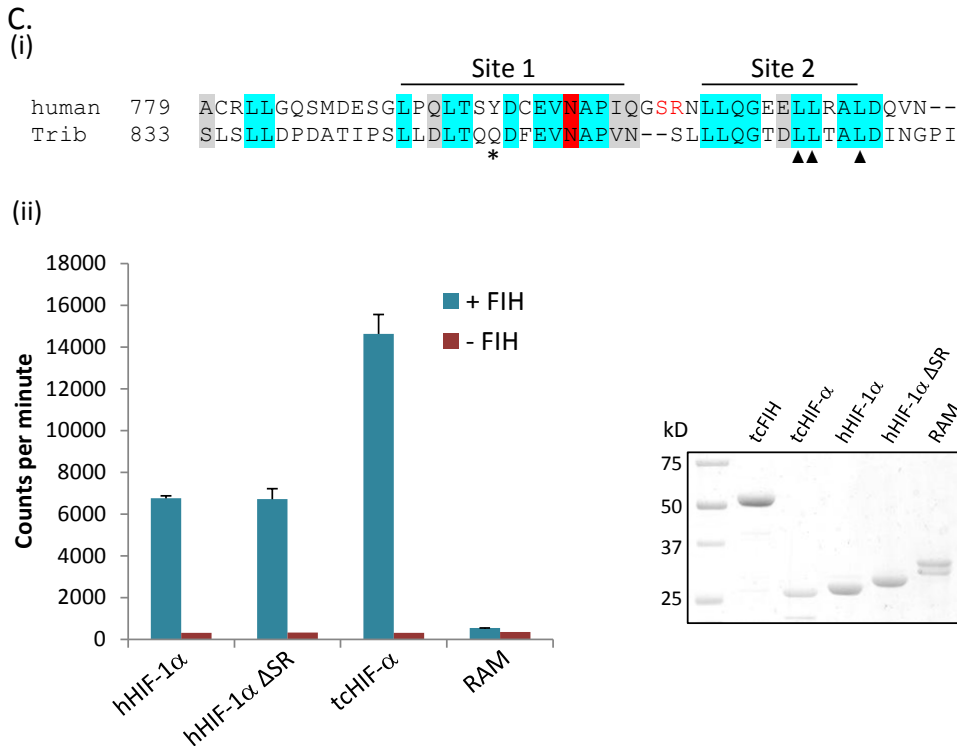
(iii)



(ii)



(Figure 4.7)



**Figure 4.7 Effect of “Tribolium mimic” mutations on the substrate specificity of hFIH.** (A) (i) The structure of hFIH is shown (cyan) with bound hHIF-1 $\alpha$  (795-823) (pink). Amino acids substituted for their tcFIH counterparts are shown in green. H-bonds are shown as yellow dotted lines. (ii) The structure of tcFIH (yellow cartoon representation) was modelled using the Phyre2 Protein Fold Recognition Server, and then overlaid with hFIH (dark blue ribbon representation; PDB identifier 1H2L) bound to hHIF-1 $\alpha$  (red). The NTKGVE insert sequence in tcFIH is highlighted in green. (B) (i) hFIH, hFIH NTKGVE/TTP (incorporating mutations ND151/152TT, Q181P and the NTKGVE insertion) and tcFIH enzymes as well as substrates mHIF-1 $\alpha$  (747-836) and mNotch1 (1862-2104) were expressed and purified from bacteria using Ni<sup>2+</sup> affinity chromatography and analysed by SDS PAGE. (ii) equimolar concentrations of enzyme were tested for their ability to hydroxylate both substrates (at 108  $\mu$ M each) by *in vitro* hydroxylation assay. Activity is expressed as the mean +/- SD of triplicate (hFIH, hFIH NTKGVE/TTP, and tcFIH) or duplicate (no enzyme) samples. Assay representative of 3 independent experiments. (iii) The enzyme activity of hFIH NTKGVE/TTP and tcFIH seen in (ii) was normalised to that of hFIH for each substrate, and then these relative activities represented as fold activity over that seen with mHIF-1 $\alpha$  (mean +/- SEM, n = 3). (C) (i) Alignment of human HIF-1 $\alpha$  and Tribolium HIF- $\alpha$ . The alignment is annotated as for Figure 4.6A, with the exception that the red highlight is indicative of the FIH target asparagine within the CAD. “Site 1” and “site 2” indicate regions of hHIF-1 $\alpha$  involved in interaction with hFIH, and red text denotes residues deleted from the hinge between site 1 and site 2 in hHIF-1 $\alpha$  (737-826) to generate HIF-1 $\alpha$  (737-826)  $\Delta$ SR. (ii) Trx-6H-tagged hHIF-1 $\alpha$  (737-826), hHIF-1 $\alpha$  (737-826)  $\Delta$ SR, tcHIF- $\alpha$  (780-879) and the negative control substrate mNotch1-RAM (1753-1847) were expressed and purified as above and analysed by SDS-PAGE (right panel) before being tested as substrates (at 83  $\mu$ M each) of bacterially expressed Trx-6H-tcFIH (400 nM, left panel). Activity is expressed as for (B)(i). Assay is representative of 2 independent experiments.

Thus far, attempts to identify mutations capable of conferring tcFIH substrate specificity onto hFIH have been unsuccessful. However, preliminary studies of human and *T. castaneum* FIH substrate hydroxylation suggested that the 2 enzymes also differed in terms of their ability to target human versus *T. castaneum* HIF homologs (in particular, tcFIH was initially observed to hydroxylate tcHIF- $\alpha$  considerably more effectively than hHIF- $\alpha$ ; see section 4.5). In light of this, it seemed that comparing human and *T. castaneum* HIF- $\alpha$  sequences might also reveal features of the two enzymes which permit their divergent substrate selectivity. Alignment of hHIF-1 $\alpha$  and tcHIF- $\alpha$  revealed a high degree of conservation in many of the residues believed to be important for hydroxylation by hFIH *in vitro*, including the residues on either side of the target asparagine, the -8 leucine, the -2 acidic residue, as well as residues occupying the hydrophobic groove at site 2 (Figure 4.7C(i)). One obvious difference, however, was a truncation of the connecting loop between site 1 and site 2 in tcHIF from 6 residues to 4. To assess the importance of this shortened loop, a human HIF- $\alpha$  CAD mutant (hHIF- $\alpha$  (737-826)  $\Delta$ SR) was generated which lacked Ser809 and Arg810, thus creating a 4 aa hinge between site 1 and site 2 as is found in tcHIF. If loop length is a significant determinant of substrate binding for tcFIH, then the truncated mutant should show improved hydroxylation over WT hHIF when tested as a substrate of tcFIH. However, hydroxylation assays utilising bacterially expressed substrates and enzyme demonstrated this was not the case (Figure 4.7C(ii)), suggesting that neither conformational flexibility of site 1 and site 2 nor the bulk of the connecting loop are factors contributing to tcFIH's substrate selectivity.

## 4.7 Discussion and Conclusions

The establishment of ARD-containing proteins as a new class of substrates for FIH has raised a number of interesting questions with regards to FIH function in cells. Despite the large number of new substrates identified, clear, biologically important functional effects for ARD hydroxylation are yet to be found. Furthermore, the stark differences in structure between HIF- $\alpha$  and ARD groups of substrates has added further interest to understanding FIH's substrate recognition mechanisms. Since FIH is conserved in many diverse species, an evolutionary biology approach was utilised to gain further insight into the comparative biological significance of ARD versus HIF- $\alpha$  hydroxylation, and the structural features of FIH which permit hydroxylation of both groups of substrates.

#### 4.7.1 Comparison of *T. castaneum* and human FIH highlight the complexity of FIH substrate recognition mechanisms

Considering the strong degree of primary sequence conservation between hFIH and tcFIH, the finding that, relative to hFIH, tcFIH substrate hydroxylation is more heavily weighted towards ARDs than HIF-1 $\alpha$  is somewhat surprising. Efforts to create a “*T. castaneum* mimic” hFIH via mutation of hFIH residues proximal to the HIF-1 $\alpha$  CAD binding site had little effect on hFIH substrate specificity, suggesting that the observed differences in activity of the two enzymes is a product of differential folding, rather than a localised change in the substrate binding interface. As an alternative explanation, it was also found that FIH can bind a folded ARD (see section 3.7), which in turn suggests that an as yet unknown set of FIH residues may be involved in folded ARD recruitment. Consequently, it cannot be ruled out that these residues may be different between the two species.

Given that the physiological relevance of FIH-mediated ARD hydroxylation is still largely speculative, the ability to analyse the effects of ARD hydroxylation independent of concurrent HIF- $\alpha$  modification would be of considerable value. From this perspective, the tcFIH enzyme may provide a launching pad for engineering of an FIH enzyme which preferentially targets ARDs over HIF- $\alpha$ . Indeed, to further this goal, a collaboration has recently been established with Christopher Schofield’s group in Oxford to perform a more in-depth analysis of tcFIH substrate binding using crystallography.

#### 4.7.2 Were ARDs the first FIH substrates?

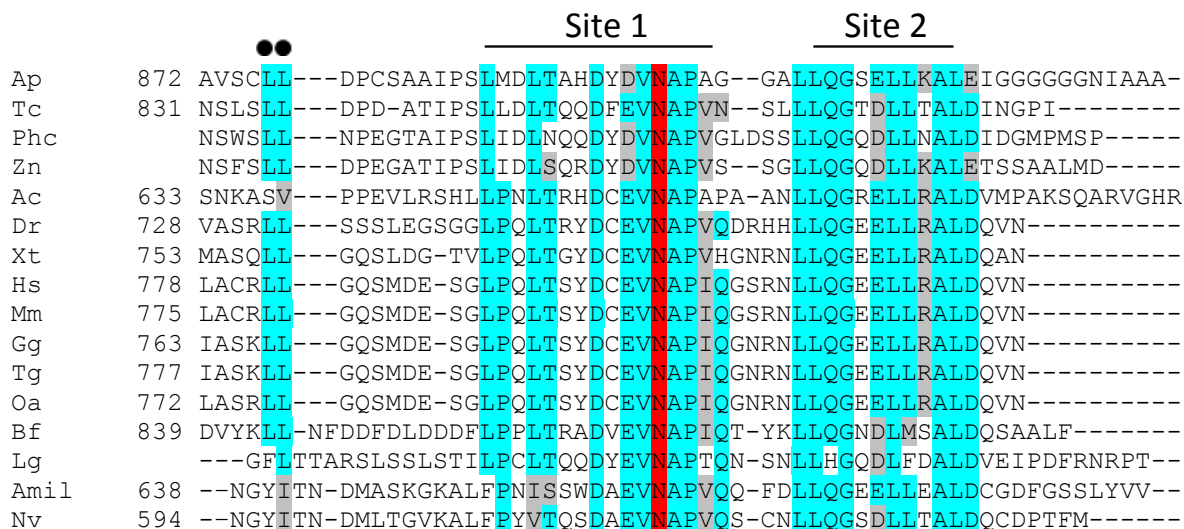
To further examine the functional relevance of ankyrin repeat hydroxylation, the evolutionary origin of the modification has also been an area of interest. The discovery of likely FIH homologs in pre-metazoan eukaryotes which lack HIF- $\alpha$  (see section 4.2) makes it highly probable that FIH’s first functional role was not oxygen-dependent regulation of the CAD. The identity of FIH’s “first” substrate is yet to be discovered. However, given that each of *G. theta*, *C. owczarzaki*, *S. rosetta* and *M. brevicollis* are host to numerous ARD-containing proteins, a protein of this type is a clear front-runner. An important next step will be to examine the catalytic activity and substrate specificity of these ancient FIH homologs. In particular, if ankyrin repeats are found to be poor targets of these enzymes, it adds credence to the existence of a non-HIF non-ARD FIH substrate. Alternatively, if ARD hydroxylation is a characteristic pre-metazoan FIH shares with its higher homologs, the notion that there exists at least 1 functionally relevant ARD modification by FIH in these species should not be discarded. More importantly, the simplified biology of organisms such as choanoflagellates may well offer an opportunity to identify this mystery substrate. In theory, this would require that both



forward and reverse genetics approaches be applicable to the organisms chosen for study, and also that disruption of FIH function has a measurable phenotype. Notably, a successful forward genetics approach was recently reported by Nicole King's group to identify the gene responsible for a defect in *S. rosetta* colony formation (Levin et al., 2014). However, at present, reverse genetics approaches in *S. rosetta* are noticeably absent from the literature, but, according to King, are currently under development.

#### 4.7.3 Structural diversity in FIH substrates: was FIH substrate recognition originally based on the ARD fold?

A consistently fascinating aspect of substrate hydroxylation by FIH has been the clear contrast in structure between the canonical ARD fold, and the considerably more flexible CAD. How are such diverse structures accommodated by the enzyme? The most obvious possibility, i.e. that FIH is "opportunistic", and only targets loosely folded ARDs, is at odds with data showing that folded ARDs can bind FIH, and thus may be involved in substrate recruitment (see section. 3.7, and (Hardy et al., 2009)). Another hypothesis is that the secondary structure proposed to be induced by the di-leucine motif of the HIF- $\alpha$  CAD (Wilkins et al., 2012) induces ankyrin repeat-like architecture, thus facilitating enzyme binding. Whatever the case in higher eukaryotes, it would be of interest to know if this structural variation was present from the outset of CAD/ARD-containing protein coexistence. In particular, if ankyrin repeats were in fact the first substrates of FIH, it is possible that early versions of FIH were geared towards recognition and hydroxylation of (at least partially folded) ARDs. From this perspective, evolution of a new CAD substrate might require that it also have ARD-like structural properties, a feature which may be evident in ancient homologs of this domain. In the present work, structural analysis of the CAD homolog from *T. castaneum* (a species which diverged from a common ancestor approximately 780 million years ago) revealed no evidence for stable structure content (section 4.4), thus arguing against this idea. However, it should be noted that, like the human CAD, the beetle version of this domain also contains a di-leucine motif (Figure 4.8), suggesting that it might also exhibit molecular crowding-induced secondary structure. In contrast, predicted CAD homologs from the more basal Cnidaria phylum lack an obvious equivalent of this motif, and therefore may be interesting candidates for future structural analysis.



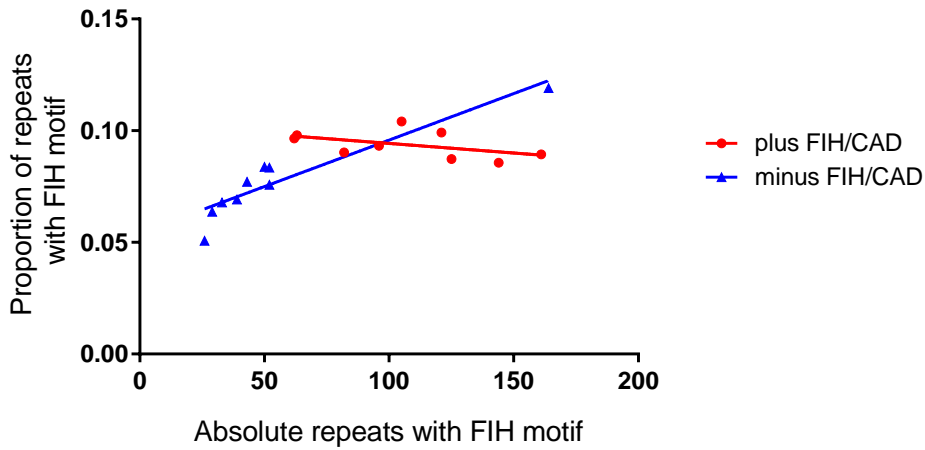
**Figure 4.8 Alignment of known and predicted HIF- $\alpha$  CAD homologs.** HIF- $\alpha$  CAD homologs isolated through BLAST and HMM searching (see main text for details) were aligned using Clustal Omega, and residues strongly or partially conserved shown with cyan and grey highlights, respectively. The Asn targeted by FIH is shown in red. Amino acid numbers are shown to the left of the alignment except for Phc, Zn and Lg, for which complete HIF sequences have not yet been annotated. Site 1 and Site 2 indicates the regions of the human HIF-1 $\alpha$  CAD that bind to FIH. The di-leucine motif is indicated above the alignment with closed circles. Alignment shading was performed using the BoxShade server. Species abbreviations and sequence IDs as follows: Bf = *Branchiostoma floridae*, AGX25238.1; Amil = *Acropora millepora*, JR993532.1; Nv = *Nematostella vectensis*, AII22158.1; Dr = *Danio rerio*, Q6EHI4; Xt = *Xenopus tropicalis*, F6VTE9; Gg = *Gallus gallus*, Q9YIB9; Hs = *Homo sapiens*, Q16665; Oa = *Ornithorhynchus anatinus*, XP\_007659416.1; Ap = *Acyrtosiphon pisum*, XP\_008186097.1; Tc = *Tribolium castaneum*, D6WMG7; Phc = *Pediculus humanus corporis*, XP\_002431762.1; Zn = *Zootermopsis nevadensis* KDR13744.1; Ap = *Aplysia californica*, XP\_012941668.1; Mm = *Mus musculus*, XP\_006515541.1; Tn = *Taeniopygia guttata*, XP\_002200394.2.

#### 4.7.4 Functional relevance of ARD hydroxylation in higher eukaryotes

Although the co-existence of FIH and ARD-containing proteins in pre-metazoans increases the likelihood that ankyrin repeats were the first FIH substrate, it is still not clear what level of functional importance ankyrin repeat hydroxylation has maintained throughout evolutionary history. The fact that there is a noticeable correlation between FIH conservation and that of the HIF- $\alpha$  CAD would suggest that there is a strong functional link between the two. Indeed, if hydroxylation of ARDs is of minimal functional importance (or if the sole purpose of ARD hydroxylation was to fine tune CAD activity), then loss of the CAD from HIF- $\alpha$  would remove evolutionary pressure to retain FIH, thus also permitting loss of the enzyme. However, it cannot be ruled out that negative selection for FIH

was the result of a change in an ARD-containing (or perhaps some as yet unidentified) FIH substrate, such that the loss of the CAD is a consequence rather than a cause of FIH deletion. In essence, the observed co-conservation of FIH and the CAD does not necessarily discredit a similar functional link between FIH and ankyrin repeat hydroxylation.

Comparison of the prevalence of an FIH target motif in different organisms (see section 4.3.1) demonstrated that the prevalence of such a motif in species which have FIH and CAD equivalents was noticeably less variable than that observed in FIH/CAD-lacking species, and furthermore these frequencies were generally higher than those observed in a majority of FIH/CAD deficient species. It is therefore tempting to speculate that the evolutionary pressure exerted by FIH/CAD system on ankyrin repeat sequences is not wholly insignificant. The small magnitude of this difference is not surprising, as the presence of an FIH motif does not guarantee that a repeat will be hydroxylated to a level which would be substantially impacted by loss of FIH (further discussed in section 3.11.2). In addition, FIH motif frequency would likely be balanced against the number and expression levels of different ankyrin repeat proteins in order to facilitate effective CAD regulation (Schmierer et al., 2010), an idea in part supported by finding that, as the absolute number of FIH motif-containing ankyrin repeats within a species decreases, the proportion of the total repeats which contain the motif tends to increase slightly, although this correlation did not reach significance (Figure 4.9). (Notably, this trend is the opposite in species which lack FIH/CAD equivalents.) Nevertheless, a



Sample	slope	r <sup>2</sup>	d.f.	P
Plus FIH/CAD	-8.60 x 10 <sup>-5</sup>	0.232	7	0.189
Minus FIH/CAD	4.16 x 10 <sup>-4</sup>	0.866	7	2.72 x 10 <sup>-4</sup>

**Figure 4.9 Comparison of FIH motif prevalence vs absolute repeat number in species with and without FIH/CAD.** Comparison of the variance of absolute FIH motif-containing repeat number with the proportion of total repeats that contain the motif. Each data point represents a different species, with red and blue points representing those containing and lacking FIH/CAD equivalents, respectively. The slope of the regression lines, the r<sup>2</sup> values and P values (the former 2 calculated using Prism, the latter calculated using Excel's FDIST function) are shown in the table below the graph. d.f. = degrees of freedom.

clearer measure of FIH/CAD-dependent differences in ankyrin repeat sequences will become possible as new genomes are sequenced and annotated. In particular, comparison of ankyrin repeats from more closely related species (e.g. those within the class Insecta, which include both FIH/CAD-containing and -lacking species) will provide a greater opportunity to compare the effect of FIH/CAD status on the ankyrin repeat sequences of genuine gene homologs, and may even offer a means to detect functional FIH-ARD interactions, as suggested by the analysis of IκBα hydroxylation sites in Figure 4.2. In summary, FIH-mediated ARD hydroxylation remains an important area of study, and the challenge remains to identify these substrates, and to determine their contribution to organism physiology.

# 5 Results – Part 3

Effect of FIH on G9a function



## 5.1 Introduction

The findings presented here and by others that ARD-containing proteins comprise a new class of substrate for FIH has raised many questions regarding the downstream consequences of these hydroxylation events. Thus far, functional studies on the effects of hydroxylation have suggested that many of these modifications have little influence on the biology of the recipient ARD (see section 3.11.4). Because of this, it has been proposed that ARD hydroxylation could be of “large-scale utility” e.g. by being a general component of ARD folding (unlikely, given that many species that contain ARDs lack an FIH homolog (see section 4.2)) or by regulating HIF via the “ARD pool theory” (Schmierer et al., 2010). Nonetheless, the possibility that large-scale ARD hydroxylation is the trade-off for evolution of a single, or discrete subset of, functionally important ARD protein hydroxylation events in a distant ancestor cannot be ignored. This idea, which is of course not mutually exclusive with the notion that other functionally important hydroxylation events may have evolved at later time points, has ensured that functional characterisation of FIH-mediated ARD protein hydroxylation remains a strong research interest. Clearly, there are few clues to predict which FIH-ARD protein interactions, among the myriad possible, might produce functionally important output. However, for convenience in the current work, it was logical to focus on ARD-containing proteins for which the functional role of the ARD is known. Moreover, potential ARD-containing substrates of FIH which might have links with sensitivity to oxygen levels were also given higher priority, as hydroxylation by FIH (or, as hypothesised in the ARD pool theory, oxygen-regulated interaction with FIH) would provide a connection between oxygen levels and oxygen-dependent functionality of the substrate. The following chapter details the preliminary analysis of one such substrate, the histone methyltransferase, G9a.

### 5.1.1 G9a and GLP: ARD-containing lysine methyltransferases

As the study of chromatin has evolved from the first concepts of the “histone code” to genome-wide mapping of histone modifications, appreciation of the importance of epigenetic modulation in the regulation of gene expression has never been greater. Among the best studied of the histone tail modifications is that of lysine methylation, which has been documented to have both positive (e.g. histone3 lysine4, H3K4) or repressive (e.g. H3K27, H3K9) effects on gene transcription (reviewed in (Bannister and Kouzarides, 2011)). Dynamic lysine methylation is achieved through the combined action of lysine demethylases, which work in opposition to the enzymatic action of lysine methyltransferases. One such methyltransferase is the SET (Suppressor of variegation 3-9, Enhancer of zeste, Trithorax) and ARD-containing enzyme, G9a. Together with the closely-related enzyme, GLP

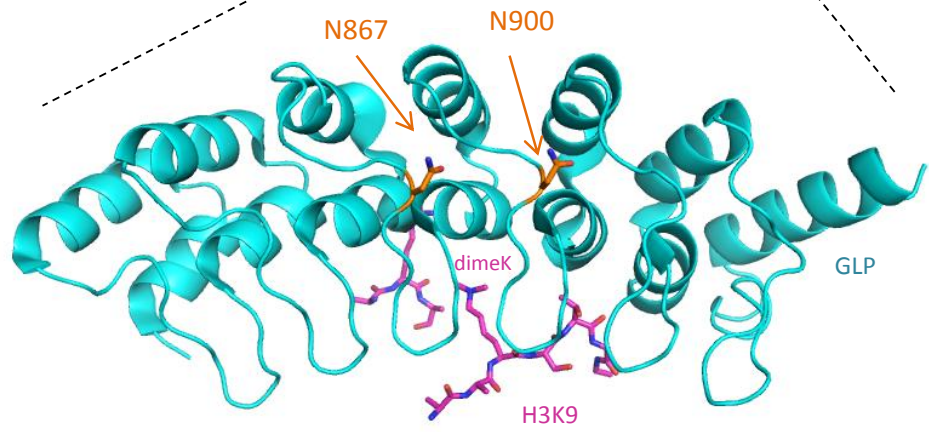
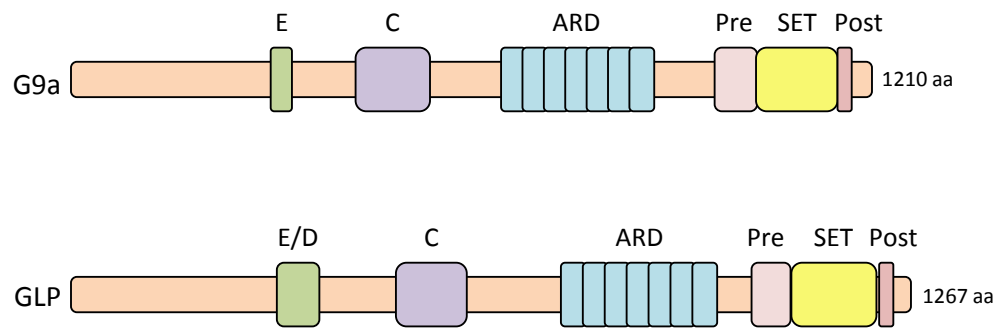
(G9a-like protein), with which it forms a heteromeric complex, G9a has a prominent role in maintaining the mono- and dimethylation states of H3K9 in euchromatin (reviewed in (Shinkai and Tachibana, 2011)).

An interesting property of G9a is its ability to both apply and “read” mono- and dimethylation, the first occurring via the catalytic SET domain, and the second through binding of methylated lysines to G9a’s ARD (Figure 5.1A) (Collins et al., 2008). The mutation of residues required for ARD H3K9me binding do not appear to affect the efficiency of histone methylation by G9a *in vitro*, however it has been suggested that the interaction may be important for efficient propagation of mono/dimeH3K9 modifications to neighbouring histones within a chromatin domain (Brent and Marmorstein, 2008; Collins et al., 2008; Shinkai and Tachibana, 2011). Given FIH’s ability to hydroxylate ARD domains, this led us to hypothesise that FIH can influence G9a’s mono/dimeH3K9 reading abilities. Notably, both G9a and GLP contain a target asparagine within one of the ankyrin repeats employed to bind dimeH3K9 (Figure 5.1A and B, and see section 5.2 for details), thus adding credence to this hypothesis.

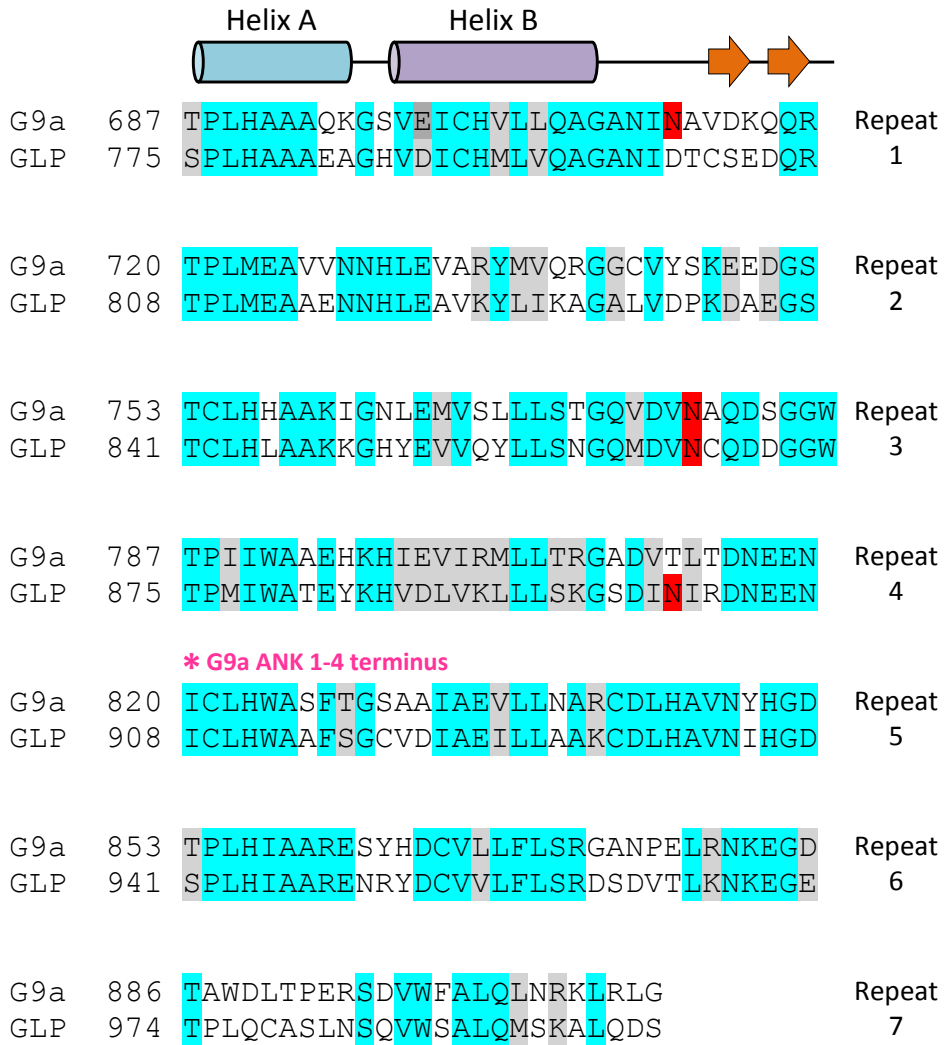
Adding further interest to G9a as an FIH substrate was a 2006 publication showing that treatment of a number of different cell lines with 0.5% hypoxia resulted in a global increase in H3K9 di- and trimethylation levels (Chen et al., 2006). Subsequent investigation of the effect demonstrated that protein levels of G9a were elevated post 6 hrs of hypoxia treatment in a transcription-independent manner (Chen et al., 2006). Moreover, since this discovery, a small number of hypoxia-repressed genes have now been shown to be targets of promoter H3K9 dimethylation by G9a, thus suggesting a possible role for this methyltransferase in hypoxic target gene regulation (Lee et al., 2009; Wang et al., 2011). In summary, given the inferred importance of the G9a ARD to methyl-lysine domain



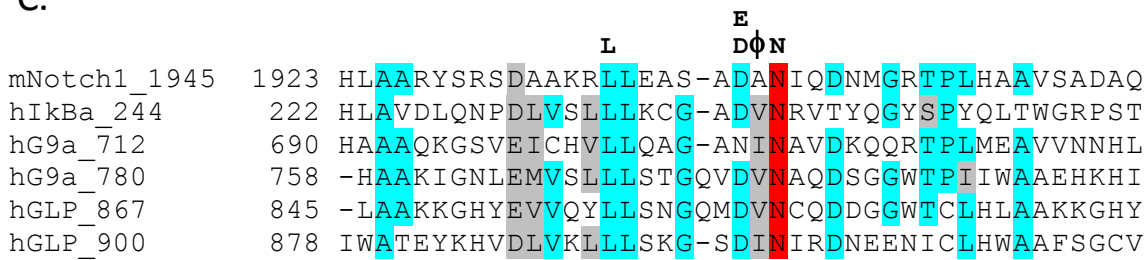
A.



B.



C.



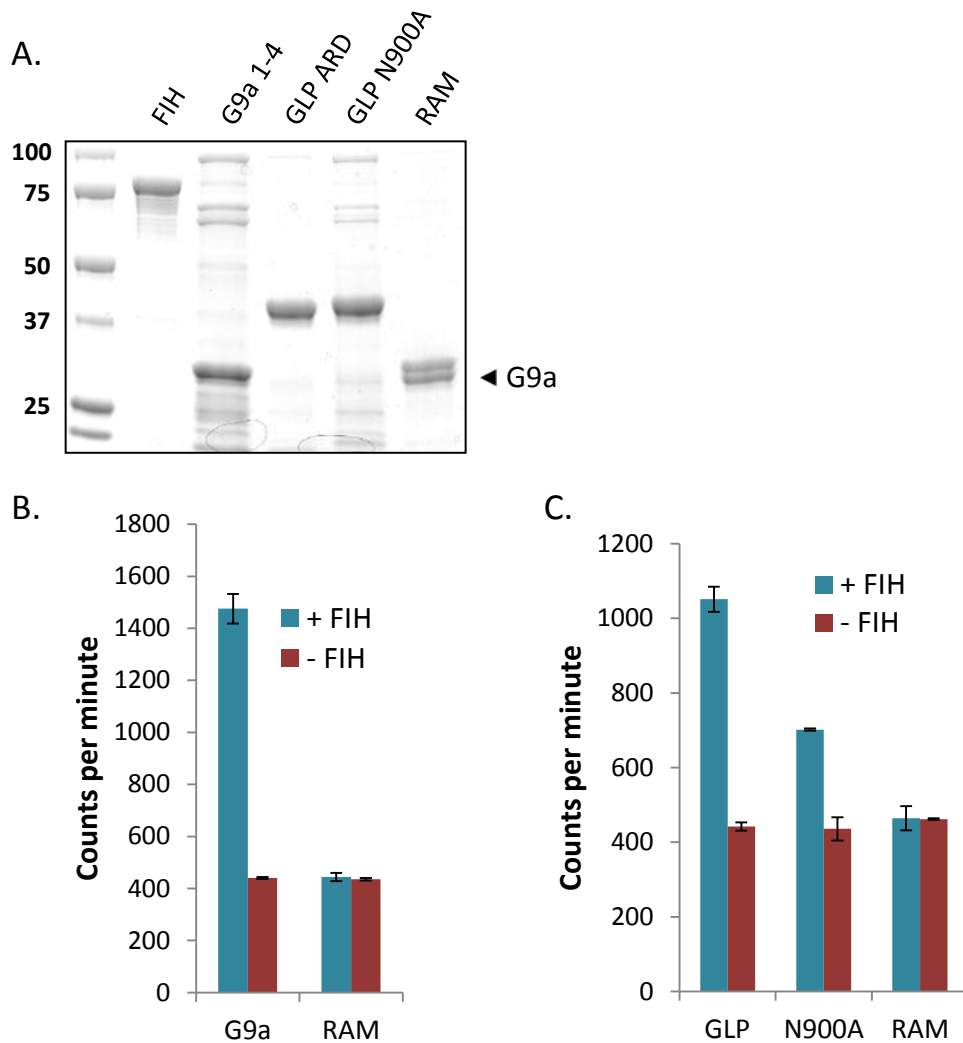
(Figure 5.1)

**Figure 5.1 Structure and likely target asparagines of G9a and GLP.** (A) Domain representation of human G9a and GLP. Numbers to the right indicate lengths of the 2 methyltransferase constructs. The crystal structure of the GLP ARD (aa 743 – 967) is shown (cyan) bound to a dimeH3K9 peptide (aa 1-13, purple. Note: no electron density is shown for aa 4-6 of the H3 peptide. *PDB reference number 3B95*). Likely target asparagines within the GLP ARD are shown in orange. E, Glu-rich region; E/D, Glu/Asp-rich region; C, Cys-rich region; ARD, ankyrin repeat domain; Pre, pre-SET domain; SET, SET domain; Post, post-SET domain. (B) Alignment of human G9a and GLP ankyrin repeats. The location of the likely target asparagines within the long loop of the repeats is shown in red. Conserved or similar residues are shown in turquoise or grey, respectively. Secondary structure of each repeat is shown above the alignment. In order to obtain soluble protein expression of G9a, a fragment of the ARD was cloned, the C-terminus of which is indicated by a pink asterisk. (C) Alignment of likely target asparagines with known targets in mNotch1 and hIκBα. The alignment is coloured as for (B), with the “FIH preferred sequence” indicated above the alignment. Numbers adjacent to protein names or shown to the left of sequences refer to target Asn positions, or aa number of the 1<sup>st</sup> residue of the repeat, respectively.

maintenance, and FIH’s propensity for binding and hydroxylation of ARDs, it was of considerable interest to determine two things: firstly, if G9a was a substrate of FIH, and secondly, the influence of this binding/hydroxylation on G9a-dependent gene regulation.

## 5.2 G9a and GLP are novel substrates of FIH

Perusal of the amino acid sequences of the G9a and GLP ARDs suggested the presence of two likely target asparaginyll residues for hydroxylation by FIH in each protein (Figure 5.1B and C). Thus, sequences for the ARD regions of G9a (aa 682-913) and GLP (aa 770-1010) were amplified from HPAC (human pancreatic adenocarcinoma) or HEK 293T cDNA, respectively, and subsequently cloned into



**Figure 5.2 Analysis of G9a and GLP as FIH substrates.** (A) hG9a 1-4 (682-820), hGLP (770-1010), hGLP (770-1010) N900A and mNotch1 RAM (1753-1847) were expressed in *E. coli* and purified using  $\text{Ni}^{2+}$  affinity chromatography, and then analysed by SDS PAGE. G9a 1-4, which was typically found to have poor stability, is indicated by an arrowhead. Hydroxylation assays were then performed using  $0.85 \mu\text{M}$  bacterially expressed MBP-FIH, and either  $4.5 \mu\text{M}$  G9a and RAM (B), or  $8.7 \mu\text{M}$  GLP, GLP N900A and RAM (C). Activity (measured as counts per minute of  $^{14}\text{C}$   $\text{CO}_2$  liberated during the reaction) is expressed as the mean  $\pm$  SD of triplicate (+ FIH) or duplicate (- FIH) samples. Graphs are representative of 3 independent experiments.

the pET32a vector for bacterial expression. Initial assessment of soluble protein yield indicated that, unlike GLP, the full length G9a ARD was insoluble under the conditions employed. As there is existing evidence for successful purification of a partial ARD from the Notch1 protein (Zheng et al., 2008), a number of shortened G9a ARD constructs were analysed, as well as one expressing the ARD and SET

domains together, for improved solubility. Among these, a construct encompassing the first 4 ankyrin repeats of G9a (G9a 1-4, aa 682-820) facilitated the recovery of a small amount of soluble protein by Ni<sup>2+</sup> affinity purification (Figure 5.2A, and data not shown). Importantly, the first 4 ankyrin repeats contain both potential hydroxy-acceptor residues present in G9a (see Figure 5.1B). Following successful purification of both protein fragments, samples of each were tested along with bacterially expressed MBP-FIH by *in vitro* hydroxylation assay. As can be seen in Figure 5.2B and C, both G9a and GLP stimulated approximately 3- and 2-fold the release of CO<sub>2</sub>, respectively, seen with the control substrate, the mNotch1 RAM domain. In order to ascertain if the predicted hydroxy-acceptor residues were in fact the genuine targets of FIH's enzymatic activity, a number of individual Asn to Ala mutant constructs were synthesised by site-directed mutagenesis (G9a N712A and N779A, and GLP N867A and N900A). However, mutagenesis of all but one of the target asparagines (GLP N900A) rendered the constructs completely insoluble (data not shown), suggesting that the asparaginyl residues in these loops are important for maintenance of ARD structural integrity. Nonetheless, the GLP N900A mutant was tested as a substrate alongside the WT GLP sequence, and a clear reduction in CO<sub>2</sub> release was observed (Figure 5.2C), but not to control levels, implying that Asn900 is only partially responsible for the stimulation of CO<sub>2</sub> release in the assay. In summary, both G9a and GLP are likely substrates of FIH, and although not all likely hydroxy-acceptor sites could be tested in hydroxylation assays, the reduction in activity seen with the N900A GLP mutant suggests that FIH is targeting asparaginyl residues positioned in the long connecting loops between repeats, as has been observed in other ARD-containing substrates (Cockman et al., 2006; Cockman et al., 2009; Zheng et al., 2008).

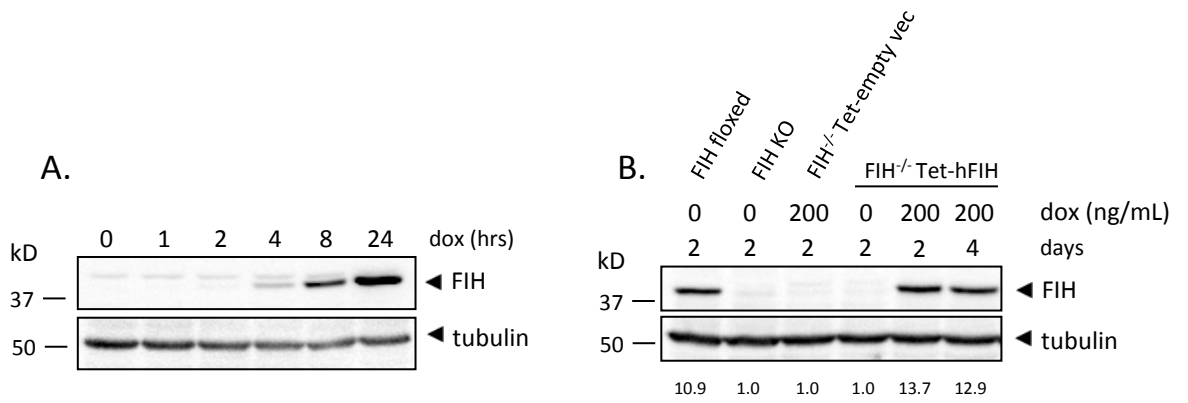
### 5.3 FIH activity does not alter expression of a subset of putative G9a-regulated target genes

Having obtained strong correlative evidence that both methyltransferase ARDs are hydroxylated by FIH, the next step was to determine the effect of this interaction with FIH on G9a/GLP's repression of gene transcription. To this end, it was aimed to disrupt FIH activity in cells, and then examine the consequences of this inhibition for expression of known G9a target genes.

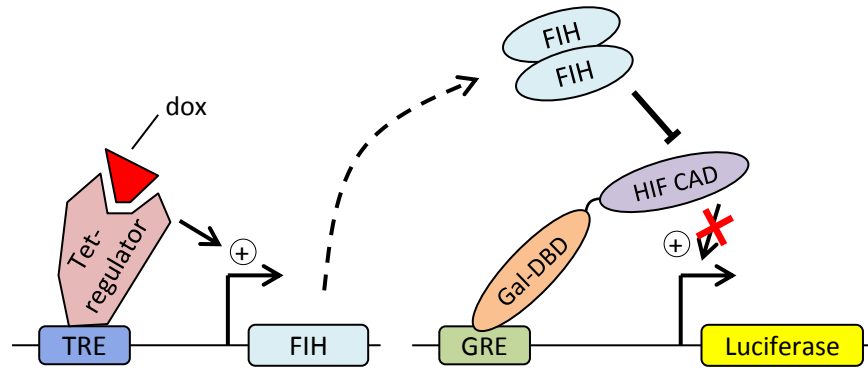
#### 5.3.1 Construction of a cell line with dox-inducible FIH expression

In order to be able to manipulate FIH activity in a cellular system, a doxycycline (dox)-inducible FIH cell line was constructed using FIH deficient MEFs (Zhang et al., 2010). FIH KO MEFs were first

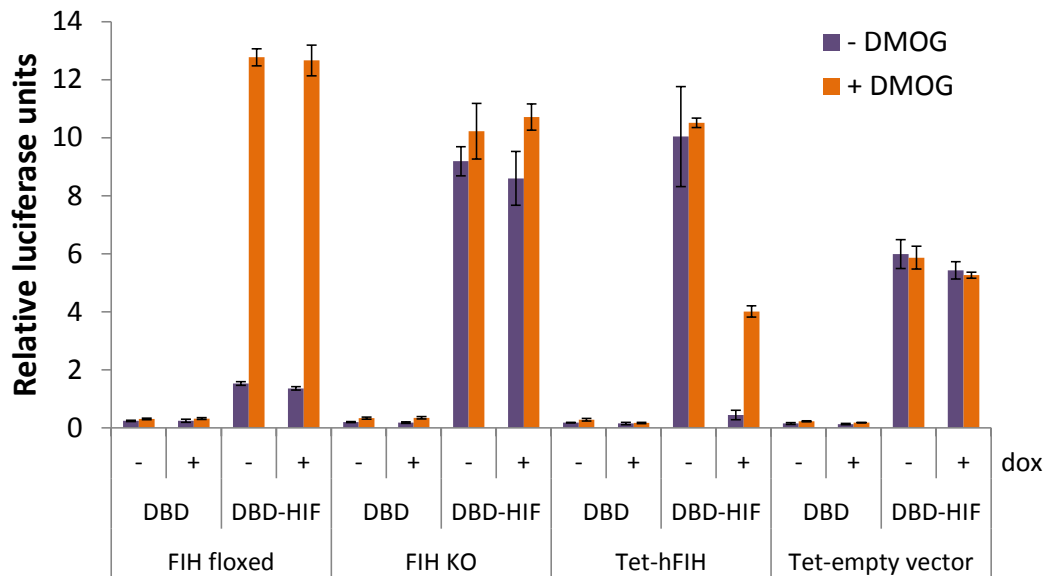
transfected with pEF-puro-TETON vector, selected for successful uptake of plasmid via culturing in puromycin, and then monoclonal cell lines generated via limiting dilution. The pEF-puro-TETON monoclonal line, "A10", was then transfected with pTRDC-ARE-FIH-EYFP or pTRDC-ARE-EYFP empty vector, and successful stable integrants selected via FACS using EYFP expression. Double stable monoclonal cell lines were then generated for each cell line via FACS or limiting dilution, resulting in isolation of several independent clones of the MEF cell lines  $FIH^{-/-}$  Tet-hFIH (2 independent clones), and  $FIH^{-/-}$  Tet-empty vector (>5 clones) (see section 2.31 for details). Analysis of the effects of dox addition showed robust induction of WT FIH protein in the  $FIH^{-/-}$  Tet-hFIH cells, with no obvious leakiness in the absence of dox (Figure 5.3A). By adjusting the concentration of dox added to the culture medium it was found that 200 ng/mL dox could induce levels of FIH expression which were comparable to that seen in the control FIH-floxed MEFs (Figure 5.3B, and data not shown), and this induction was stable over a 4 day period. To investigate the catalytic efficacy of the FIH in the  $FIH^{-/-}$  Tet-hFIH cell line, the ability of dox to repress a Gal DBD-HIF-1 $\alpha$  CAD reporter gene was assessed. As can be seen in Figure 5.3C, addition of 200 ng/mL dox strongly reduced the activity of the reporter gene in the  $FIH^{-/-}$  Tet-hFIH MEFs, but had little effect in  $FIH^{-/-}$  Tet-empty vector MEFs. The repression of CAD activity in the dox-treated  $FIH^{-/-}$  Tet-hFIH MEFs could also be partially reversed by the FIH inhibitor, DMOG, thus providing further evidence that the FIH expressed from the Tet response element retained similar characteristics to that of endogenous FIH (Figure 5.3C). A more thorough examination of the effect of dox addition in the  $FIH^{-/-}$  Tet-hFIH cell line indicated that a dox concentration between 250 ng/mL and 125 ng/mL would produce levels of HIF-1 $\alpha$  CAD repression comparable to that seen in the floxed FIH MEFs (Figure 5.3D). This observation is in agreement with Figure 5.3B, which shows that the intermediate concentration of 200 ng/mL dox produces "endogenous-like" levels of exogenous WT FIH by western blot. Based on these initial tests, a concentration of 200 ng/ml dox would be used to simulate endogenous levels of WT FIH expression in all subsequent experiments.



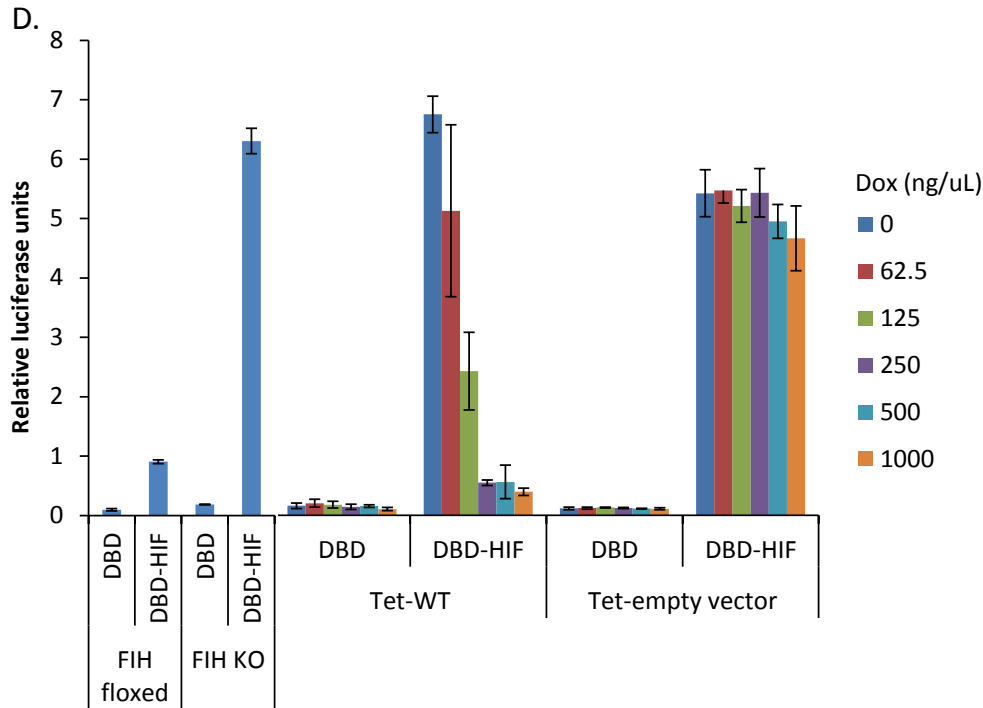
**C.**  
(i)



(ii)



(Figure 5.3)



**Figure 5.3 Characterisation of dox-inducible FIH MEF cell lines.** (A)  $FIH^{-/-}$  Tet-FIH MEFs were seeded in 6-well trays, allowed to recover for at least 4 hours, then treated with 2  $\mu$ g/mL dox for the indicated times. Cells were subsequently lysed and analysed by western blot for FIH expression ( $\alpha$ -FIH, Novus). (B) FIH floxed, FIH KO,  $FIH^{-/-}$  Tet-empty vector or  $FIH^{-/-}$  Tet-FIH MEFs were seeded in T25 flasks and then left untreated or treated with 200 ng/mL dox for 2 or 4 days prior to cell lysis. Equal amounts of extract for each condition were then analysed by western blot to compare FIH levels with those found in FIH floxed MEFs. Numbers below the panels represent tubulin-normalised signal intensity for FIH bands relative to the intensity of the FIH KO MEF FIH band. Blot is representative of 2 independent experiments. (C) (i) Schematic of the reporter assay system used to analyse catalytic activity of the dox inducible FIH in  $FIH^{-/-}$  Tet-FIH or  $FIH^{-/-}$  Tet-empty vector MEFs. (ii) FIH floxed, FIH KO,  $FIH^{-/-}$  Tet-FIH or  $FIH^{-/-}$  Tet-empty vector MEFs were transfected with a Gal response element-driven firefly luciferase reporter gene, renilla luciferase under the control of a constitutive promoter, and Gal DBD alone or fused to HIF-1 $\alpha$  (727-826). Five hours post transfection, cells were either left untreated, treated with 2  $\mu$ g/mL dox alone, treated with 1 mM DMOG alone, or treated with both chemicals for 20 hrs. Cells were then lysed and analysed by dual luciferase assay. Data is represented as mean relative luciferase  $\pm$  SD, with each sample tested in triplicate. Data is representative of 3 independent experiments for FIH floxed, FIH KO, and  $FIH^{-/-}$  Tet-FIH MEF samples, and 2 independent experiments for the  $FIH^{-/-}$  Tet-empty vector MEF sample. (D) FIH floxed, FIH KO,  $FIH^{-/-}$  Tet-FIH, or  $FIH^{-/-}$  Tet-empty vector MEFs were transfected as for (C), allowed to recover for 4 hrs, then treated with the indicated concentrations of dox for 20 hrs. Cells were subsequently lysed and analysed by dual luciferase assay. Data is represented as mean relative luciferase units  $\pm$  SD, with each sample tested in triplicate.  $n = 1$  for concentrations of dox of 125 ng/mL or lower, however, a similar experiment was performed utilising concentrations of 250 -2000 ng/mL dox which showed only slight derepression of HIF-CAD activity at 250 ng/mL dox (data not shown).



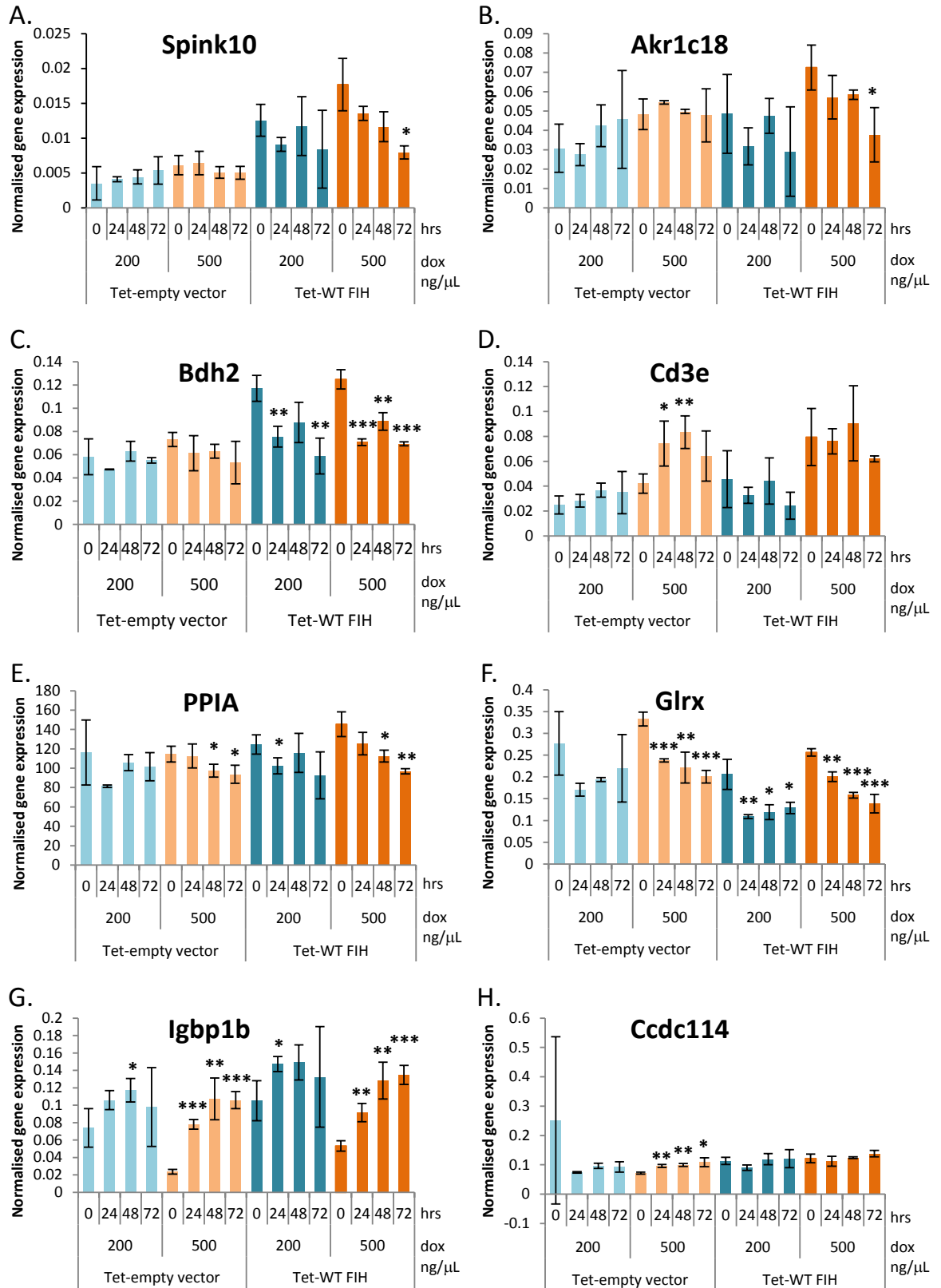
### 5.3.2 Effect of FIH manipulation on G9a target gene expression

Given that G9a (in cooperation with GLP) is responsible for a large proportion of H3K9 dimethylation, great potential exists for G9a to have wide-spread repressive effects on multiple target genes. Indeed, work by the group of Lorenz Poellinger has identified a large number of genes which are differentially expressed in WT versus G9a<sup>-/-</sup> murine ES cells (Kian Leong Lee, personal communication). The work presented in this thesis has shown that G9a (and GLP) are likely targets of the FIH enzyme, thus it was of interest to know if modulation of FIH activity, via dox-inducible expression, might alter the magnitude of G9a binding and/or hydroxylation, and in turn alter its methyltransferase function. A change in G9a target gene expression, if observed, could be the downstream consequence of covalent modification of G9a with a hydroxyl group (analogous to what is seen with the FIH substrate, HIF-1 $\alpha$  (Lando et al., 2002b)), or alternatively may result from the binding of FIH to G9a, irrespective (to some degree) of hydroxylation (see, for example, the effect of FIH on Notch1 activity (Zheng et al., 2008)). Thus, it is conceivable that FIH may regulate G9a in an oxygen-dependent manner, or it may have more generalised effects on G9a activity. Notably, although a number of the G9a targets identified by the Poellinger lab were also found to be significantly hypoxia-sensitive, it was decided to exclude these from the initial analysis, as many of these genes may additionally be targets of the HIF-1 $\alpha$  transcription factor, which is also hydroxylated and repressed by FIH. Thus, by excluding HIF target genes, any effects observed would be more likely due to a direct effect of FIH on G9a.

To assess the effect of FIH on G9a target gene expression, FIH<sup>-/-</sup> Tet-hFIH or FIH<sup>-/-</sup> Tet-empty vector control MEFs were seeded in 96 well trays, and then left untreated, or treated with 200 or 500 ng/mL dox for 24, 48 or 72 hrs. Based on the experiments presented in Figure 5.3, 200 ng/mL dox was anticipated to simulate endogenous levels of FIH, while 500 ng/mL dox would serve as a system for “overexpressed” FIH, which may amplify any subtle effects seen with the “endogenous-like” dox treatment. Following the treatment period, cells were lysed and RNA prepared ready for cDNA

synthesis and qPCR analysis. In addition, RNA yield was used as a surrogate measure of cell health during the treatment period. While 200 ng/mL dox appeared to have no effect on RNA yield from either FIH<sup>-/-</sup> Tet-hFIH or FIH<sup>-/-</sup> Tet-empty vector control MEFs, 500 ng/mL dox had a clear detrimental effect on RNA yield (and by extension, cell proliferation) which increased over time (data not shown). As such, any effects seen with the higher concentration of dox would need be interpreted with caution.

A total of 42 G9a target genes which were previously identified as sensitive to G9a activity in mES cells (Kian Leong Lee, unpublished data) were analysed by qPCR (see Table 5.1). Perhaps unsurprisingly, amplification reactions of 14 of the G9a-sensitive genes from mES cells displayed Ct values comparable to the negative control, the mycoplasma 16S ribosomal RNA, in all conditions tested, likely indicating that these genes were not expressed in MEFs (Table 5.1, and data not shown). Among the remaining genes, a number of different effects on the G9a targets were observed. For Spink10, Akr1c18, Bdh2, Cd3e, Cox7b2 and Defb42, expression of WT FIH produced a trend towards repression of gene expression which often increased over time, while addition of dox to the FIH<sup>-/-</sup> Tet-empty vector control MEFs either had no effect on gene expression (Spink10, Akr1c18, and Bdh2) or caused an increase in target gene expression (Cd3e, Cox7b2 and Defb42; Figure 5.4A-D, and data not shown). However, it should be noted that a similar trend (albeit smaller in magnitude) was observed in the housekeeping gene, PPIA (Figure 5.4E), suggesting that some of



(Figure 5.4)

**Figure 5.4 Effect of FIH over-expression on G9a target genes.** FIH<sup>-/-</sup> Tet-FIH or FIH<sup>-/-</sup> Tet-empty vector MEFs were treated with the indicated concentrations of dox for 24, 48 or 72 hrs, after which cells were lysed, and cDNA prepared for qPCR analysis. A dox concentration of 200 ng/mL was chosen to simulate endogenous levels of FIH expression, while 500 ng/mL dox was used to potentially amplify any effects seen through over-expression of FIH. (A-D) Genes which demonstrated a decrease in expression upon dox addition which was specific to the FIH<sup>-/-</sup> Tet-FIH cell line, an effect which was also observed to a lesser extent in the housekeeping gene, PPIA (E). (F-H) Genes which were non-specifically repressed (F), activated (G), or unchanged (H) by treatment with dox. Gene expression data were normalised to HMBS gene expression for each treatment sample, and then plotted as mean +/- SEM (n = 3). p values were calculated by comparing the 0 hr time point for each cell line and dox concentration with the corresponding 24, 48 and 72 hr time points of that same cell line/dox concentration using a 2-tailed Student's T-test. \* p ≤ 0.05, \*\* p ≤ 0.01, \*\*\* p ≤ 0.001.

these changes may be due to non-specific side effects of dox treatment. Indeed, non-specific effects were clearly visible for a number of other genes, with the addition of dox causing FIH-independent repression (Herc3, Edg2, Glrx, Mdfic and Pde4dip) or activation (BC099439, LOC545238, Tph2 and Igbp1b) of target gene expression (Figure 5.4F and G, and data not shown). For the remainder of genes analysed, expression of FIH in response to 200 or 500 ng/mL dox had no consistent influence on G9a target gene expression (Figure 5.4H, and data not shown), suggesting that FIH is not a global regulator of hypoxia-insensitive G9a target genes in MEFs. However, it remains to be determined if FIH may influence a subset of G9a target genes which are responsive to oxygen levels.

Gene	Expressed	Gene	Expressed
m4930486L24Rik	Y	mGlrx	Y
mAass	N	mGm773	N
mAdh4	Y	mHerc3	Y
mAkr1c12	N	mIgbp1b	Y
mAkr1c18	Y	mKlb	N
mAkr1cl	Y	mKrt23	N
mApom	Y	mLOC545238	Y
mAsz1	N	mMagea4	Y
mAU040829	Y	mMagea6	N
mBC099439	Y	mMdfic	Y
mBdh2	Y	mNckap1l	Y
mC80008	N	mOTTMUSG00000025408	Y
mCcdc114	Y	mPde4dip	Y
mCd3d	N	mPkd2l1	Y
mCd3e	Y	mSftpd	N
mCd3g	N	mSgcb	Y
mCox7b2	Y	mSlc5a4b	Y
mDefb42	Y	mSpic	N
mEdg2	Y	mSpink10	borderline
mEfhc2	N	mTph2	Y
mEG638695	Y	mUbe2dnl1	N

**Table 5.1 G9a target genes analysed for expression in dox-inducible FIH MEF cell lines.**

## 5.4 Discussion

Research over the past 8 years has seen a substantial increase in the number of verified ARD-containing substrates for the asparaginyl hydroxylase, FIH, which has in turn revitalised interest in the establishment of new functional roles for the enzyme. Indeed, generation of an FIH KO mouse has added further strength to the idea that FIH has HIF-independent functional roles. Here, the data presented show that the ARD-containing methyltransferases G9a and GLP are both likely novel substrates of FIH, with alanine mutagenesis of the GLP substrate suggesting that FIH targets asparaginyl residues located in the long loops between repeats, as has been found for other ARD-containing substrates. Interestingly, the ARDs of G9a and GLP have been shown to bind mono and dimeH3K9 (Collins et al., 2008), thus prompting us to question if binding or hydroxylation of the ARDs by FIH may alter their ability to maintain methylated H3K9 domains in cells. However, whilst manipulation of FIH levels in MEFs caused repression of a small subset of the G9a target genes

analysed, data interpretation was complicated by low level (or absence of) expression of many of the genes, as well as non-specific effects generated by the addition of doxycycline. Nevertheless, although FIH had little effect on most of the genes analysed, independent follow-up of e.g. Bdh2 may yet be worthwhile given the consistent repression of gene expression seen in response to FIH (over)expression. Given that some of the genes chosen for analysis appeared to be silenced in MEFs, it is likely that G9a-sensitive target genes in these cells may differ considerably from those in ES cells. With this in mind, deletion of G9a in MEFs may permit the identification of more relevant G9a target genes for analysis in response to manipulation of FIH activity. Alternatively, synthesis of a dox-inducible system for FIH in ES cells would facilitate a more conclusive analysis of the effect of FIH on the G9a target genes examined in the present study. With regards to the latter idea, the release of the Tet-ON 3G generation of dox-inducible vectors (Clontech), which have a considerably improved sensitivity to doxycycline, would allow the use of smaller amounts of dox for FIH induction, thus reducing non-specific effects and providing clearer evidence as to whether the gene repression observed in this work is a genuine FIH-dependent effect. Additionally, the development of CRISPR/Cas technology (reviewed in (Mali et al., 2013)) now allows for the rapid deletion of individual genes in cultured cells, and could be used to efficiently generate FIH-deficient ES cells. Furthermore, the finding that G9a activity is sensitive to oxygen levels suggests that this enzyme may have an important role in hypoxic target gene repression. Therefore, expansion of G9a targets analysed to include those that are activated or repressed by hypoxia may help to determine if FIH activity modulates an oxygen-sensitive subset of G9a regulated genes.

Finally, it is also noteworthy that G9a is known to methylate non-histone proteins, including p53, G9a (automethylation), Wiz (widely interspaced zinc finger motifs protein), CDYL1 (chromodomain Y-like protein), CSB (Cockayne syndrome group B protein), Pontin and Reptin (Lee et al., 2011; Rathert et al., 2008). Thus far, the biological significance of most of these modifications is not well understood. However, for the chromatin remodelling factors Pontin and Reptin, this methylation is known to occur in a hypoxia-dependent manner, and in turn promotes binding to HIF-1 $\alpha$  and recruitment of Pontin/Reptin to HIF target gene promoters (Lee et al., 2011; Lee et al., 2010). As such, assessment of the effect of FIH on Pontin/Reptin methylation and target gene modulation (necessarily in the context of FIH-insensitive HIF-1 $\alpha$ , as may be achieved through use of an N803A mutant) may also prove an interesting avenue of investigation for FIH modulation of G9a function.

# 6 Results – Part 4

Does asparaginyl hydroxylation of I $\kappa$ B $\alpha$  affect its stability and function?





## 6.1 Introduction

Despite a large number of ARD-containing proteins having been identified as substrates of FIH, the functional importance of the modification, either in the context of individual ankyrin repeat domains, or for ARD-containing proteins as a whole, is a matter still under investigation. Conceptually, hydroxylation of ARDs could be envisaged to alter the chemical properties of a binding surface (much as hydroxylation of HIF-1 $\alpha$  abrogates binding to p300/CBP), or alternatively, it may influence intramolecular interactions and protein folding. In addition, these modifications have the potential to be applied in an oxygen-dependent manner, due to FIH's direct utilisation of dioxygen for hydroxylation. As such, the current search for ARD-containing proteins which are likely to be functionally influenced by FIH have focussed on those with oxygen-sensitive functions, or those for which structural stability is known to be mechanistically important. As discussed in section 3.11.4, the strong influence of I $\kappa$ B $\alpha$  "foldedness" on its ability to regulate NF $\kappa$ B function, coupled with the knowledge that FIH-mediated hydroxylation of ankyrin repeats can increase ARD stability led us to embark on a collaborative project with the laboratory of Betsy Komives to assess the effect of I $\kappa$ B $\alpha$  hydroxylation on its stability and its inhibitory properties towards NF $\kappa$ B. Section 6.2 contains the published manuscript detailing this work. In addition, a brief discussion of the manuscript's results in the context of the work of this thesis is included in section 6.3.

## 6.2 Published manuscript: "Consequences of I kappa B alpha hydroxylation by the factor inhibiting HIF (FIH)"

Please refer to the manuscript on the following pages. Author contributions for the manuscript presented herein can be found in the Author contributions section.





## Consequences of I $\kappa$ B $\alpha$ alpha hydroxylation by the factor inhibiting HIF (FIH)

Ingrid L. Devries<sup>a,1</sup>, Rachel J. Hampton-Smith<sup>b,1</sup>, Melinda M. Mulvihill<sup>a,1</sup>, Vera Alverdi<sup>a</sup>, Daniel J. Peet<sup>b</sup>, Elizabeth A. Komives<sup>a,\*</sup>

<sup>a</sup>Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive La Jolla, CA 92093-0378, United States

<sup>b</sup>School of Molecular and Biomedical Science, University of Adelaide, Adelaide SA 5005, Australia

### ARTICLE INFO

#### Article history:

Received 7 August 2010

Revised 28 September 2010

Accepted 28 October 2010

Available online 5 November 2010

Edited by Zhijie Chang

#### Keywords:

Hypoxia-inducible factor

Ankyrin repeat

Post-translational modification

Protein folding

Proteasome degradation

### ABSTRACT

**The factor inhibiting HIF-1 (FIH-1) hydroxylates many ankyrin repeat-containing proteins including I $\kappa$ B $\alpha$ . It is widely speculated that hydroxylation of I $\kappa$ B $\alpha$  has functional consequences, but the effects of hydroxylation have not been demonstrated. We prepared hydroxylated I $\kappa$ B $\alpha$  and compared it to the unhydroxylated protein. Urea denaturation and amide H/D exchange experiments showed no change in the “foldedness” upon hydroxylation. Surface plasmon resonance measurements of binding to NF $\kappa$ B showed no difference in the NF $\kappa$ B binding kinetics or thermodynamics. Ubiquitin-independent proteasomal degradation experiments showed no difference in the half-life of the protein. Thus, it appears that hydroxylation of I $\kappa$ B $\alpha$  by FIH-1 is inconsequential, at least for the functions we could assay in vitro.**

#### Structured summary:

MINT-8051494: *NF-kappa-B p65* (uniprotkb:Q04207) physically interacts (MI:0915) with *NF-kappa-B p50* (uniprotkb:P25799) and *I-kappa-B alpha* (uniprotkb:O15111) by surface plasmon resonance (MI:0107)

© 2010 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

### 1. Introduction

The hypoxia-inducible transcription factors (HIFs) mediate the genomic response to oxygen deficiency (hypoxia) in multicellular organisms. These transcription factors are  $\alpha/\beta$  heterodimers that bind DNA and activate transcription of over 70 target genes during cellular hypoxia [1]. HIFs “sense” the level of oxygen by being modified by four oxygen-sensitive hydroxylases; three prolyl hydroxylases (PHDs) and one asparaginyl hydroxylase, factor inhibiting HIF-1 (FIH) [2]. These hydroxylases are all members of the 2-oxoglutarate (2OG)-dependent dioxygenase superfamily and their activity depends on the cellular oxygen concentration. Under normoxia, the HIF- $\alpha$  subunit is hydroxylated by the PHDs leading to ubiquitination and rapid degradation. In addition, FIH-mediated hydroxylation regulates the transcriptional activity of the HIF- $\alpha$  proteins by inhibiting binding of the obligate CREB-binding protein (CBP)/p300 coactivator proteins to the HIF- $\alpha$  C-terminal transactivation domain (CAD), further repressing transcriptional activity [3,4]. At the onset of hypoxia, hydroxylase activity is greatly reduced, resulting in HIF protein accumulation and derepression of CAD activity, and consequently robust transcriptional activation.

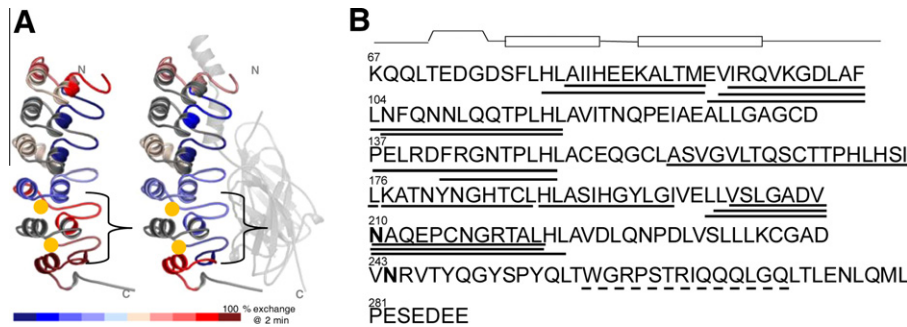
Recent work has demonstrated that, in addition to hydroxylating HIF- $\alpha$ , FIH can also hydroxylate the ankyrin domains of a wide range of proteins. Ankyrin repeat domains contain a consensus sequence that has been shown to bind to, and be a good substrate for, FIH [5]. Searches for alternative substrates of FIH identified I $\kappa$ B, Notch family members and ASB4 (ankyrin repeat and SOCS box protein 4) as substrates of FIH [5–10]. These intracellular proteins all contain ankyrin repeat domains (ARDs), and in each case the target asparagine residues lie within the ARD. Ankyrin repeat-containing proteins are found in all three phyla, and are present in some 6% of eukaryotic protein sequences [11,12]. Thus the consequences of FIH-dependent hydroxylation could be wide-ranging if modification of these alternate substrates has functional significance. Although these alternative substrates have been shown to be hydroxylated in vitro, in cultured cells, and in vivo, it remains difficult to establish functional consequences of the hydroxylation event in a rigorous manner.

Since hydroxylation of HIF- $\alpha$  potentiates its degradation, and the degradation rate of I $\kappa$ B $\alpha$  is a sensitive parameter in control of the NF $\kappa$ B signaling pathway [13,14], we were particularly interested in the possible functional role of hydroxylation in regulating the stability and/or degradation of I $\kappa$ B $\alpha$ . Indeed, human I $\kappa$ B $\alpha$  is hydroxylated by FIH at positions 210 and 244 (Fig. 1A) with Asn 244 being the preferred site [10]. The two hydroxylated asparagines are at the interfaces between AR4-AR5 and AR5-AR6, which is the weakly-folded part of I $\kappa$ B $\alpha$  [15] that folds on binding of NF $\kappa$ B

\* Corresponding author. Fax: +1 858 534 6174.

E-mail address: [ekomives@ucsd.edu](mailto:ekomives@ucsd.edu) (E.A. Komives).

<sup>1</sup> These authors all contributed equally.



**Fig. 1.** (A) Structure of IκBα from the NFκB–κBα complex showing the relative amide exchange in free IκBα vs. NFκB-bound IκBα. The red regions exchange nearly all of their amides within 2 min, whereas the blue regions exchange less. Gold spheres mark the positions of the two hydroxylated asparagines, Asn 210 and Asn 244. (B) Sequence of IκBα showing the two hydroxylated asparagines in bold-face type. Above the sequence is the secondary structural representation of each ankyrin repeat. The lines underneath the sequence depict the peptides generated upon pepsin cleavage during the amide exchange experiment that were analyzed. The solid lines indicate peptides for which amide exchange can be quantitatively determined whereas the dashed lines indicate peptides for which the amide exchange can only be qualitatively assessed.

(Fig. 1A) [16]. In previous work, we showed that stabilization of AR5–AR6 in IκBα altered its intracellular half-life and had drastic functional consequences [17]. Given that hydroxylation has been shown to stabilize consensus ARD proteins, it is possible that hydroxylation might increase the stability of the IκBα ARD, and therefore influence its functional activity [18]. Here we present a full biophysical characterization of FIH-hydroxylated IκBα, including binding of the hydroxylated protein to NFκB and proteasomal degradation rates.

## 2. Materials and methods

### 2.1. IκBα expression and purification

Human IκBα<sub>67–287</sub> protein was expressed and purified as described previously [17]. Residues 67–287 comprise the entire ARD and the PEST sequence. This protein is truncated at the C-terminus, and is missing residues 288–317, but we have shown previously that it has identical folding and binding properties to the full-length ARD [19]. The final purification step was on a Superdex-75 gel filtration column (GE Healthcare) to ensure the absence of any aggregated species. The proteins were stored at 4 °C and used within 3 days of gel filtration. Protein concentrations were determined by spectrophotometry, using extinction coefficient of 12950 M<sup>-1</sup> cm<sup>-1</sup> for IκBα.

### 2.2. In vivo hydroxylation

In vivo hydroxylation was achieved by co-expressing IκBα behind the T7 promoter in a kanamycin resistant vector and Trx-6H-FIH [20] behind the T7 promoter in an ampicillin resistant vector in BL21 (DE3) cells. Hydroxylated IκBα was purified from the co-expression system as described above. Although some of the IκBα remained in a high molecular weight complex with FIH, monomeric, hydroxylated IκBα could be obtained after a final purification step on Superdex-75 gel filtration column.

### 2.3. Mass spectrometry

Hydroxylated IκBα samples were analyzed by liquid chromatography (LC)–MS with electrospray ionization. All nanospray ionization experiments were performed by using a QSTAR-XL hybrid mass spectrometer (ABSciex) interfaced to a nanoscale reversed-phase high-pressure liquid chromatograph (Tempo) using a 10 cm-180 ID glass capillary packed with 5-μm Magic C-4 beads (Michrom). Buffer A was 98% H<sub>2</sub>O, 2% ACN, 0.2% formic acid, and 0.005% TFA and buffer B was 100% ACN, 0.2% formic acid, and 0.005% TFA. Peptides were eluted from the C-4 column into the

mass spectrometer using a linear gradient of 25–80% Buffer B over 60 min at 400 μl/min. LC–MS data were acquired in TOF scan mode (400–2000). The parameters for ESI-MS were: IS 2.3 KV; DP 65 V; CAD 5 V. The data were processed with Analyst 2.0 software using Bayesian Protein Reconstruction tool in the mass range 20000–30000 with step mass of 1 Da and S/N threshold of 20.

### 2.4. Equilibrium folding experiments

Equilibrium folding experiments were performed on an Aviv 202 spectropolarimeter (Aviv Biomedical, Lakewood, NJ, USA) with a Hamilton Microlab 500 titrator (Hamilton, Reno, NV, USA). Urea (Fisher Scientific, Pittsburg, PA, USA) was specially purified by dissolving it at a nominal concentration of 8M in water, and then treating with AG 501-X8 (D) resin (BioRad Laboratories, Hercules, CA, USA) for 1 h to remove cyanate contaminants [21]. Resin was filtered out with a 0.2 μm filter and buffer salts were added to the purified urea. Urea concentrations were checked using refractometry [22]. Urea was used within 2 days of resin treatment to prevent re-accumulation of cyanate. A 1 cm fluorescence quartz cuvette containing 2.0 ml of 0.5–4 μM of the native protein in buffer (25 mM tris, 50 mM NaCl, 1 mM DTT, 0.5 mM EDTA, pH 7.5) and was titrated with denatured protein (7.3–8.4 M urea in buffer), in 30–40 injection steps. After each injection, samples were equilibrated with constant stirring at 80 rpm for 180 s prior to data collection. The CD signal was collected at 225 nm, averaged 10 s, and the fluorescence signal was collected through a 320 cut-off filter with an excitation wavelength of 280 nm, averaged over 2–5 s. Experiments were performed at 5 °C unless otherwise stated.

Folding curves were fit to a two state folding model, assuming a linear dependence of the folding free energy on denaturant concentration [22]. The pre (native) and post (unfolded) transition baselines were treated as linearly dependent on denaturant concentration. The data was globally fit to Eq. (1).

$$S_{\text{obs}} = (a_1 + p_1[\text{Urea}]) + (a_2 + p_2[\text{Urea}]) \exp(-(\Delta G - m[\text{Urea}]/RT)) / (1 + \exp(-(\Delta G - m[\text{Urea}]/RT))) \quad (1)$$

where  $S_{\text{obs}}$  is the observed signal,  $p_1$  and  $p_2$  are the pre and post transition baselines,  $a_1$  and  $a_2$  their corresponding y-intercepts;  $\Delta G$  is the folding free energy in water and  $m$  is the cooperativity parameter ( $m$ -value). The data were fit using a non-linear least square fitting algorithm in Kaleidagraph (Synergy Software, Reading, PA, USA).

### 2.5. Amide exchange experiments

Native state backbone amide exchange was measured as described previously [23]. To specifically probe the weakly-folded

parts of the ARD, I $\kappa$ B $\alpha$  was incubated for 2 min in deuterated buffer at pH 7.4 and then the amide exchange was quenched by dilution into ice cold 0.1% TFA solution to bring the pH down to 2.2 and the temperature down to 0 °C. The sample was immediately digested with an excess of immobilized pepsin and the digest mixture was frozen in liquid N<sub>2</sub>. All samples were analyzed on the same day by MALDI-TOF mass spectrometry [24]. Eighteen peptides that cover 60% of the sequence were analyzed for both wild-type and hydroxylated I $\kappa$ B $\alpha$ . The centroids of the mass envelopes were measured and compared to undeuterated controls and corrected for back exchange as described previously [15].

### 2.6. Surface plasmon resonance binding experiments

Sensorgrams were recorded on a Biacore 3000 instrument using streptavidin chips as described [19]. NF $\kappa$ B was biotinylated and immobilized as described; 150 RU, 250 RU, and 350 RU of NF $\kappa$ B (p50<sub>248–321</sub>/p65<sub>190–321</sub>) were immobilized. Wild type I $\kappa$ B $\alpha$  (0.02–1.6 nM) was injected for 5 min and dissociation was measured for 20 min at 25 °C at 50  $\mu$ L/min. Regeneration was achieved by a 1 min pulse of 3 M urea in running buffer. Data were fit using the BiaEvaluation 2.0 software and plotted in Origin 7.0.

### 2.7. Proteasome degradation experiments

Proteasome degradation assays were performed basically as described [17]. I $\kappa$ B $\alpha$  (1  $\mu$ M), freshly purified by size-exclusion chromatography, was incubated with 100 nM bovine 20S proteasome (a gift from G. Ghosh) for 0, 15, 30, 60, 90, or 120 min at 25 °C in 20 mM Tris, 200 mM NaCl, 10 mM MgCl<sub>2</sub>, 1 mM DTT (pH 7.0). Degradation reactions were quenched by boiling with SDS-PAGE sample buffer. Intact I $\kappa$ B $\alpha$  was separated by SDS-PAGE (13% polyacrylamide gel) and visualized using Western blots probed with sc-847 (Santa Cruz Biotechnologies) followed by anti-rabbit HRP conjugate.

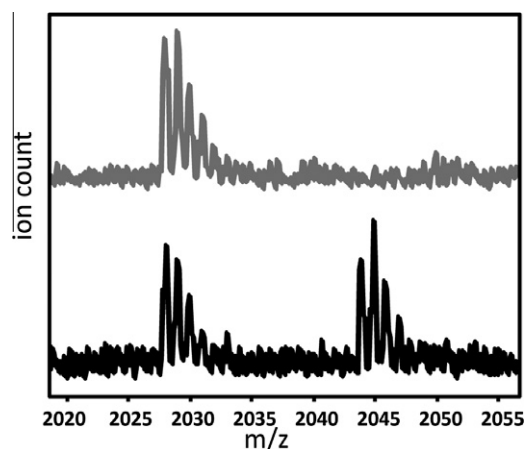
## 3. Results

### 3.1. Preparation and characterization of FIH-1-hydroxylated I $\kappa$ B $\alpha$

Several different approaches to obtaining fully-hydroxylated I $\kappa$ B $\alpha$  were attempted, including in vitro incubation with either purified FIH or lysates from *Escherichia coli* cells expressing FIH. The FIH and I $\kappa$ B $\alpha$  appeared to form a tight complex prior to hydroxylation, and only partial hydroxylation was observed under initial in vitro hydroxylation conditions tested (data not shown). Co-expression of the I $\kappa$ B $\alpha$  with FIH in *E. coli* finally resulted in sufficient yields of hydroxylated protein. Mass spectrometric analysis of the hydroxylated protein compared to unhydroxylated controls showed that the protein was fully hydroxylated with only one predominant peak at 24427 Da after deconvolution of the multiply charged species, as compared to the unhydroxylated control, which had a mass of 24394 Da. Others have shown that FIH hydroxylates I $\kappa$ B $\alpha$  at Asn 210 and Asn 244, but the hydroxylation at 244 occurs more readily [10]. Pepsin digestion of the hydroxylated protein gave four peptides spanning residues 203–221 (Fig. 1). Analysis of these peptides showed that at least 50% of the I $\kappa$ B $\alpha$  was hydroxylated also at Asn 210 (Fig. 2).

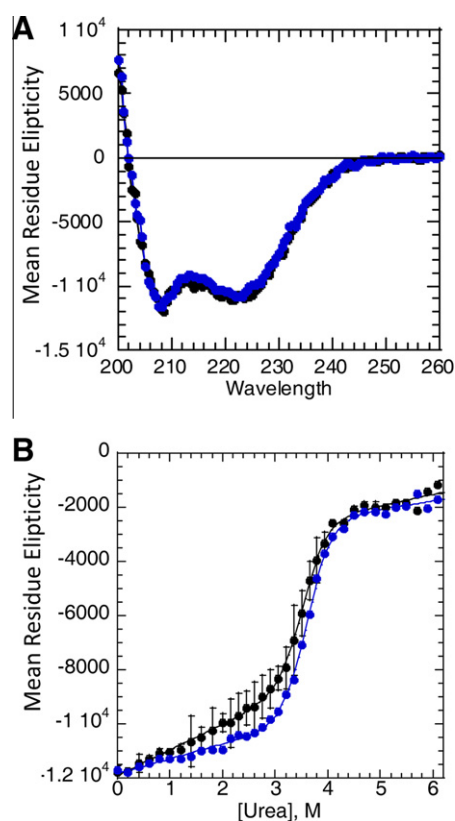
### 3.2. Equilibrium unfolding experiments

The hydroxylated protein displayed high  $\alpha$ -helical secondary structure content, as indicated by the far-UV circular dichroism (CD) spectrum, which was not different from the wild type (Fig. 3A). To probe the stability of the FIH hydroxylated I $\kappa$ B $\alpha$ ,



**Fig. 2.** MALDI-ToF mass spectra of the peptide corresponding to residues 202–221 ( $m/z$  208) from wild type I $\kappa$ B $\alpha$  (grey) and hydroxylated I $\kappa$ B $\alpha$ . Although quantitation is not possible due to the expected difference in ionization between the hydroxylated and unhydroxylated peptides, it is clear that a substantial amount of Asn 210 is hydroxylated.

titration experiments were performed to unfold the protein in urea monitoring the CD signal at 222 nm. Although I $\kappa$ B $\alpha$  is prone to aggregation and is not thermally denaturable, at pH 7.5, in 50 mM NaCl, 1 mM DTT, 25 mM Tris-HCl buffer, the refolding upon chemical denaturation was >95% reversible, as shown previously [23]. The CD signal changed upon addition of denaturant and displayed



**Fig. 3.** (A) Circular dichroism (CD) spectra of wildtype (black) and hydroxylated (blue) I $\kappa$ B $\alpha$ . (B) Equilibrium urea denaturation of I $\kappa$ B $\alpha$  (black) and hydroxylated I $\kappa$ B $\alpha$  (blue) at 3  $\mu$ M total protein concentration at 5 °C, (conditions described in Section 2) followed by the CD signal at 225 nm. The standard error of the mean from several experiments is also plotted. The data were fit assuming a sloping baseline giving for the wild type protein  $\Delta G = -9.1 \pm 0.7$  kcal/mol,  $m = 2.6 \pm 0.2$  kcal mol<sup>-1</sup> M<sup>-1</sup> and for the hydroxylated protein  $\Delta G = -9.3 \pm 0.5$  kcal/mol,  $m = 2.6 \pm 0.13$  kcal mol<sup>-1</sup> M<sup>-1</sup>.

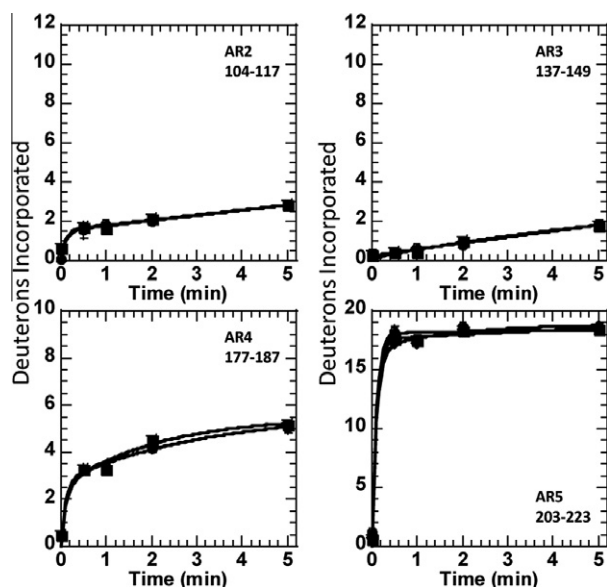
a sharp transition, typical of a cooperatively folded unit, but with a sloping pre-transition baseline which we have previously assigned to the weakly-folded AR5–6 region [23]. Multiple experiments on the hydroxylated versus wild type I $\kappa$ B $\alpha$  were performed, however the equilibrium denaturation properties of the hydroxylated protein were not different from those of wild type I $\kappa$ B $\alpha$  (Fig. 3B).

### 3.3. Amide hydrogen exchange experiments

Native state amide exchange monitored by MALDI-TOF mass spectrometry is a sensitive probe of the folded state of proteins in general, and is highly informative in I $\kappa$ B $\alpha$  because pepsin digestion after exchange produces fragments of similar secondary structure from each repeat that can be directly compared [15]. We have previously shown that mutation of only two residues in AR6 pre-fold the weakly-folded part, and this is most clearly shown by a dramatic decrease in amide exchange (a difference of some five deuterons) in AR5 and AR6 in this pre-folded mutant [17]. Fig. 4 shows plots of the amide exchange of several peptides corresponding to regions of AR4–AR5. Comparison of the hydroxylated I $\kappa$ B $\alpha$  to the wild type protein shows no difference in amide exchange. This mass spectrometry experiment is particularly important because here the peptides corresponding to the hydroxylated protein are separated (by virtue of their difference in mass) from those of the unhydroxylated protein, even if they are present in the same mixture. Thus, this experiment allows us to compare the fully hydroxylated protein to the protein with only a hydroxyl at Asn 244 in the same experiment. It is clear from the plots in Fig. 4 that again, no differences were observed in the “foldedness” of I $\kappa$ B $\alpha$  due to FIH hydroxylation. Although the peptide of mass 1767 from AR6 could not be analyzed quantitatively, this region also did not differ as assessed by qualitative comparisons of the mass envelopes.

### 3.4. Binding to NF $\kappa$ B by surface plasmon resonance

To probe whether the hydroxylation affects NF $\kappa$ B binding, surface plasmon resonance binding kinetics experiments were per-



**Fig. 4.** Plots of the amide H/D exchange over five minutes into regions of I $\kappa$ B $\alpha$  or hydroxylated I $\kappa$ B $\alpha$ . In the case of the peptides covering the Asn 210, only the peptide of mass 2165 is presented, but all the peptides from this region gave the same results. For this peptide, three data sets are presented. Data for the unhydroxylated protein was obtained from a separate unhydroxylated sample as well as from the unhydroxylated protein present in the sample of hydroxylated protein.

formed. As shown previously, I $\kappa$ B $\alpha$  binds extremely tightly to NF $\kappa$ B [19]. The binding experiments showed that hydroxylation had little, if any, effect on NF $\kappa$ B binding ( $K_D$  of the wild type protein was  $62 \pm 12$  pM whereas that for the hydroxylated protein was  $88 \pm 14$  pM) (Fig. 5).

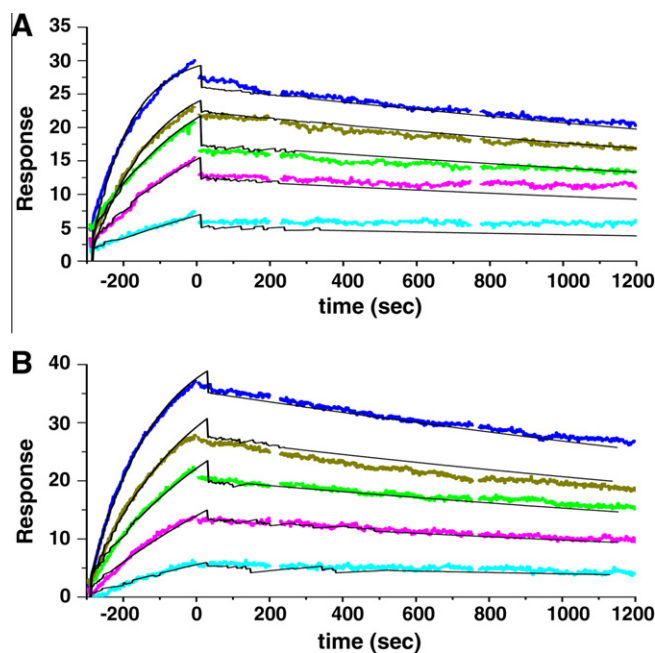
### 3.5. Proteasomal degradation

We have previously shown that I $\kappa$ B $\alpha$  is rapidly degraded in vitro and in vivo by the 20S proteasome [17]. This process is ubiquitin independent, but rather depends on a degron in the C-terminal part of the protein that might be sensitive to FIH hydroxylation [25]. To probe whether FIH hydroxylation affected the degradation rate of I $\kappa$ B $\alpha$  by the 20S proteasome, similar in vitro experiments were performed. Fig. 6 shows that the degradation of FIH-hydroxylated I $\kappa$ B $\alpha$  is not significantly different from that of wild type I $\kappa$ B $\alpha$ .

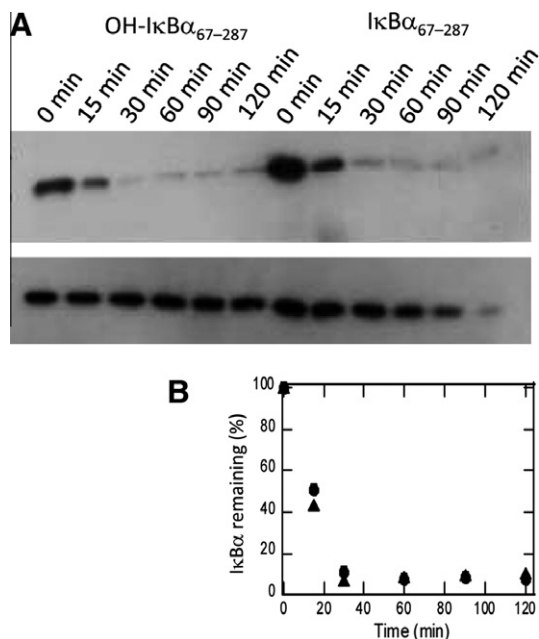
Quantitation of the bands by densitometry gives a half-life of 15 min for both the hydroxylated and unhydroxylated protein.

## 4. Discussion

The NF $\kappa$ B signaling pathway is controlled by several inhibitor proteins, but I $\kappa$ B $\alpha$  is responsible for the rapid response after acute stress signals [26]. Properties of I $\kappa$ B $\alpha$  including its “foldedness”, intracellular half-life, and binding affinity for NF $\kappa$ B dimers are all exquisitely balanced in order to maximize pathway response and control [14]. In addition, NF $\kappa$ B signaling has previously been shown to be induced by hypoxia although the mechanism of induction is not fully understood [27,28]. Previous reports indicated a slightly enhanced inhibitory action of hydroxylated I $\kappa$ B $\alpha$  (that would presumably be suppressed by hypoxia) but siRNA-



**Fig. 5.** Surface plasmon resonance (Biacore) binding kinetics of the interaction between NF $\kappa$ B and either wild type (A) or hydroxylated (B) I $\kappa$ B $\alpha$  was performed as previously described [19]. NF $\kappa$ B(p50<sub>248–321</sub>/p65<sub>190–321</sub>) was biotinylated at the N-terminus of p65 and immobilized on a streptavidin sensor chip. I $\kappa$ B $\alpha$  (0.12, 0.34, 0.56, 0.95, 1.56 nM) was the flowing analyte. The data were fit using a simple 1:1 binding model yielding for the wild type protein  $k_a = 6 \times 10^6$  M<sup>-1</sup>s<sup>-1</sup>,  $k_d = 2.3 \times 10^{-4}$ ,  $R_{MAX} = 28.3$ ,  $X^2 = 0.55$ , and  $K_D = 62 \pm 12$  pM and for the hydroxylated protein  $k_a = 4 \times 10^6$  M<sup>-1</sup>s<sup>-1</sup>,  $k_d = 2.8 \times 10^{-4}$ ,  $R_{max} = 44.3$ ,  $X^2 = 0.74$ , and  $k_d = 88 \pm 14$  pM.



**Fig. 6.** Proteasome degradation experiments were performed as described in Section 2. (A) Samples were taken at 0, 30, 60, 90, and 120 min and quenched by addition of gel loading buffer. The amount of remaining IκBα was analyzed by SDS-Page and Western blotting by comparing the sample containing proteasome (top gel) with the sample that did not contain proteasome (bottom gel). (B) Densitometric quantitation of the gels shown in A confirms the similar degradation rates of IκBα (●) and hydroxylated IκBα (▲).

mediated suppression of FIH actually reduced NFκB activity [10]. Thus, although a functional significance would make a lot of sense, the evidence was tenuous at best. In previous work, we have demonstrated that we can subtly perturb the “foldedness” of the IκBα ARD and consequently its various functions, including NFκB binding and intracellular half-life through the use of stabilizing mutations [17,23]. Here we used the same experimental approaches to analyze the consequences of FIH hydroxylation on the “foldedness” and function of IκBα.

Others have shown that IκBα is hydroxylated by FIH at two asparagine residues, Asn 210 and 244, the latter of which is more readily hydroxylated [10]. In attempting to hydroxylate IκBα with FIH in vitro, we observed that a tight complex forms between the two proteins as has been seen for FIH and other ARD-containing proteins [18]. This complex was so tight as to make it difficult to separate the IκBα from the FIH to obtain reasonable yields of fully hydroxylated protein from in vitro reactions. Instead, from co-expression, we were able to obtain monomeric hydroxylated IκBα in reasonable yields. Electrospray Q-TOF mass spectrometry of the intact hydroxylated IκBα showed a single peak with a mass approximately 32 Da higher than IκBα suggesting that the protein was 100% hydroxylated at two sites. Pepsin digestion followed by MALDI mass spectrometry, which is not quantitative and is expected to cause at least some fragmentation of the hydroxyl group, showed that at least 50% of Asn 210 was hydroxylated. Taken together, these data strongly suggest that the protein is 100% hydroxylated at Asn 244 and at least 50% hydroxylated at Asn 210.

Equilibrium unfolding experiments showed equal stability of the FIH hydroxylated IκBα and consistent with this result, native state amide H/D exchange to probe the backbone dynamics also showed no difference between unhydroxylated and hydroxylated IκBα across all of the ankyrin repeats. Importantly, even though the protein may not have been completely hydroxylated, the hydroxylated and unhydroxylated protein fragments separate by mass and could be simultaneously analyzed from the same sample.

Previous indirect DNA binding activity (EMSA assays) of NFκB from nuclear extracts had suggested addition of FIH reduced the binding of NFκB [10]. We previously showed that EMSA assays do not accurately recapitulate the intracellular binding affinity of the NFκB-IκBα interaction, but that surface plasmon resonance with specifically biotinylated NFκB does [19]. We therefore measured the binding affinity of the hydroxylated IκBα in the same manner and compared it to the unhydroxylated protein. The results clearly showed that hydroxylation had minimal effects on the binding affinity.

Finally, we previously showed that the intracellular half-life of IκBα is exquisitely sensitive to the “foldedness” of AR5 and AR6 [17]. We also showed that intracellular half-life of free IκBα is short because of proteasomal degradation of IκBα that is ubiquitin independent and likely involves some form of the 20S proteasome [25]. As few as two mutations in AR6 dramatically altered both the intracellular half-life and the in vitro half-life in the presence of 20S proteasome preparations [17]. Since intracellular half-life is one of the sensitive parameters that controls NFκB signaling, we were anticipating that FIH hydroxylation would alter the proteasomal sensitivity of IκBα. Again, however, both proteins had the same rate of degradation by 20S proteasomal preparations.

Thus, for every function we can rigorously and quantitatively measure in vitro, FIH hydroxylation of IκBα is inconsequential. It is possible that many of the ARD hydroxylations that have been observed are similarly inconsequential and that the consensus sequence that is being hydroxylated is an accidental substrate. On the other hand, we have shown that the FIH orthologue from *Tribolium castaneum* (red flour beetle) can also hydroxylate the *Tribolium* Notch analogue [29]. This implies that this “accidental” modification is highly conserved and may have some functional consequence other than alteration of stability or NFκB binding. FIH plays an important role in regulating metabolism, with the phenotype of the FIH knockout mice displaying elevated metabolism, including higher respiration and heart rate, as well as energy expenditure [30]. It is unclear at this stage whether this phenotype is mediated by the HIF or ankyrin repeat substrates, although there is no obvious disruption to NFκB signalling.

It is interesting that a similar conundrum also plagues asparagine hydroxylation in the EGF-like domains of most extracellular proteins. Here too, no evidence that hydroxylation alters the structure or function of the EGF-like domain-containing proteins has been forthcoming. Yet, the beta aspartyl/asparaginyl hydroxylase knockout mice display a distinct developmental phenotype, which the authors concluded is due to altered hydroxylation of Jagged and probably also its receptor, Notch [31]. So, although the molecular consequences of EGF hydroxylation are yet to be defined, it obviously has an important physiological role. As mass spectrometry advances and more post-translational modifications are found, it will likely become more and more common that the functionally relevant substrates are few in a mix of unintentional ones. In addition, the functional consequences of such modifications may require years of experimentation to establish.

## Acknowledgements

NIH P01GM071862-04 to E.A.K. and funding from the National Health and Medical Research Council of Australia to D.J.P.

## References

- [1] Semenza, G.L. (2003) Targeting HIF-1 for cancer therapy. *Nat. Rev. Cancer* 3, 721–732.
- [2] Schofield, C.J. and Ratcliffe, P.J. (2004) Oxygen sensing by HIF hydroxylases. *Nat. Rev. Mol. Cell Biol.* 5, 343–354.

- [3] Lando, D., Peet, D.J., Gorman, J.J., Whelan, D.A., Whitelaw, M.L. and Bruick, R.K. (2002) FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia-inducible factor. *Genes Dev.* 16, 1466–1471.
- [4] Hewitson, K.S. et al. (2002) Hypoxia-inducible factor (HIF) asparagine hydroxylase is identical to factor inhibiting HIF (FIH) and is related to the cupin structural family. *J. Biol. Chem.* 277, 26351–26355.
- [5] Cockman, M.E., Webb, J.D., Kramer, H.B., Kessler, B.M. and Ratcliffe, P.J. (2009) Proteomics-based identification of novel factor inhibiting hypoxia-inducible factor (FIH) substrates indicates widespread asparaginyl hydroxylation of ankyrin repeat domain-containing proteins. *Mol. Cell. Proteomics* 8, 535–546.
- [6] Zheng, X. et al. (2008) Interaction with factor inhibiting HIF-1 defines an additional mode of cross-coupling between the Notch and hypoxia signaling pathways. *Proc. Natl. Acad. Sci. USA* 105, 3368–3373.
- [7] Coleman, M.L. et al. (2007) Asparaginyl hydroxylation of the Notch ankyrin repeat domain by factor inhibiting hypoxia-inducible factor. *J. Biol. Chem.* 282, 24027–24038.
- [8] Cockman, M.E., Webb, J.D. and Ratcliffe, P.J. (2009) FIH-dependent asparaginyl hydroxylation of ankyrin repeat domain-containing proteins. *Ann. NY Acad. Sci.* 1177, 9–18.
- [9] Ferguson 3rd, J.E., Wu, Y., Smith, K., Charles, P., Powers, K., Wang, H. and Patterson, C. (2007) ASB4 is a hydroxylation substrate of FIH and promotes vascular differentiation via an oxygen-dependent mechanism. *Mol. Cell. Biol.* 27, 6407–6419.
- [10] Cockman, M.E. et al. (2006) Posttranslational hydroxylation of ankyrin repeats in I $\kappa$ B proteins by the hypoxia-inducible factor (HIF) asparaginyl hydroxylase, factor inhibiting HIF (FIH). *Proc. Natl. Acad. Sci. USA* 103, 14767–14772.
- [11] Björklund, A.K., Ekman, D. and Elofsson, A. (2006) Expansion of protein domain repeats. *PLoS Comput. Biol.* 2, e114.
- [12] Mosavi, L.K., Cammett, T.J., Desrosiers, D.C. and Peng, Z.Y. (2004) The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci.* 13, 1435–1448.
- [13] O’Dea, E.L., Barken, D., Peralta, R.Q., Tran, K.T., Werner, S.L., Kearns, J.D., Levchenko, A. and Hoffmann, A. (2007) A homeostatic model of I $\kappa$ B metabolism to control constitutive NF- $\kappa$ B activity. *Mol. Syst. Biol.* 3, 111.
- [14] Ferreira, D.U. and Komives, E.A. (2010) Molecular mechanisms of system control of NF- $\kappa$ B signaling by I $\kappa$ B $\alpha$ . *Biochemistry* 49, 1560–1567.
- [15] Croy, C.H., Bergqvist, S., Huxford, T., Ghosh, G. and Komives, E.A. (2004) Biophysical characterization of the free I $\kappa$ B $\alpha$  ankyrin repeat domain in solution. *Protein Sci.* 13, 176777.
- [16] Truhlar, S.M., Torpey, J.W. and Komives, E.A. (2006) Regions of I $\kappa$ B $\alpha$  that are critical for its inhibition of NF- $\kappa$ B. DNA interaction fold upon binding to NF- $\kappa$ B. *Proc. Natl. Acad. Sci. USA* 103, 18951–18956.
- [17] Truhlar, S.M.E., Mathes, E., Cervantes, C.F., Ghosh, G. and Komives, E.A. (2008) Pre-folding I $\kappa$ B $\alpha$  alters control of NF- $\kappa$ B signaling. *J. Mol. Biol.* 380, 67–82.
- [18] Hardy, A.P., Prokes, I., Kelly, L., Campbell, I.D. and Schofield, C.J. (2009) Asparaginyl beta-hydroxylation of proteins containing ankyrin repeat domains influences their stability and function. *J. Mol. Biol.* 392, 994–1006.
- [19] Bergqvist, S., Croy, C.H., Kjaergaard, M., Huxford, T., Ghosh, G. and Komives, E.A. (2006) Thermodynamics reveal that helix four in the NLS of NF- $\kappa$ B p65 anchors I $\kappa$ B $\alpha$ , forming a very stable complex. *J. Mol. Biol.* 360, 421–434.
- [20] Linke, S., Hampton-Smith, R.J. and Peet, D.J. (2007) Characterization of ankyrin repeat-containing proteins as substrates of the asparaginyl hydroxylase factor inhibiting hypoxia-inducible transcription factor. *Methods Enzymol.* 435, 61–85.
- [21] Street, T.O., Courtemanche, N. and Barrick, D. (2008) Protein folding and stability using denaturants. *Methods Cell Biol.* 84, 295–325.
- [22] Pace, C.N. (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* 131, 266–280.
- [23] Ferreira, D.U., Cervantes, C.F., Truhlar, S.M., Cho, S.S., Wolyne, P.G. and Komives, E.A. (2007) Stabilizing I $\kappa$ B $\alpha$  by “consensus” design. *J. Mol. Biol.* 365, 1201–1216.
- [24] Mandell, J.G., Falick, A.M. and Komives, E.A. (1998) Measurement of amide hydrogen exchange by MALDI-TOF mass spectrometry. *Anal. Chem.* 70, 3987–3995.
- [25] Mathes, E., O’Dea, E.L., Hoffmann, A. and Ghosh, G. (2008) NF- $\kappa$ B dictates the degradation pathway of I $\kappa$ B $\alpha$ . *EMBO J.* 27, 1357–1367.
- [26] Hoffmann, A., Levchenko, A., Scott, M.L. and Baltimore, D. (2002) The I $\kappa$ B-NF- $\kappa$ B signaling module: temporal control and selective gene activation. *Science* 298, 1241–1245.
- [27] Schmedtje, J.F., Ji, Y.S., Liu, W.L., DuBois, R.N. and Runge, M.S. (1997) Hypoxia induces cyclooxygenase-2 via the NF- $\kappa$ B p65 transcription factor in human vascular endothelial cells. *J. Biol. Chem.* 272, 601–608.
- [28] Zampetaki, A., Mitsialis, S.A., Pfeilschifter, J. and Kourembanas, S. (2004) Hypoxia induces macrophage inflammatory protein-2 (MIP-2) gene expression in murine macrophages via NF- $\kappa$ B: the prominent role of p42/p44 and PI3 kinase pathways. *FASEB J.* 18, 1090–1092.
- [29] Hampton-Smith, R.J. and Peet, D.J. (2009) From polypos to people: a highly familiar response to hypoxia. *Ann. NY Acad. Sci.* 1177, 19–29.
- [30] Zhang, N. et al. (2010) The asparaginyl hydroxylase factor inhibiting HIF-1 $\alpha$  is an essential regulator of metabolism. *Cell Metab.* 11, 364–378.
- [31] Dinchuk, J.E. et al. (2002) Absence of post-translational aspartyl beta-hydroxylation of epidermal growth factor domains in mice leads to developmental defects and an increased incidence of intestinal neoplasia. *J. Biol. Chem.* 277, 12970–12977.



### 6.3 Discussion

In 2009, it was reported that hydroxylation of ARDs by FIH could promote the thermal and chemical stability of artificial ARDs designed using the ankyrin repeat consensus sequence (Hardy et al., 2009). Based on these findings, it was of interest to know if similar increases in stability were conferrable to naturally occurring ARDs, and whether this increase in stability was sufficient to alter ARD function. With this in mind, reports that the ARD-containing protein  $\text{I}\kappa\text{B}\alpha$  harbours an ARD which is loosely folded at its C-terminal end, and that this conformational flexibility is important for regulating  $\text{I}\kappa\text{B}\alpha$  stability as well as its binding to  $\text{NF}\kappa\text{B}$  were intriguing (Truhlar et al., 2008). Importantly, this flexible region also contains an asparaginyl residue hydroxylated by FIH (Asn244) (Cockman et al., 2006). Given that this targeted residue resides on the opposite face to that used by  $\text{I}\kappa\text{B}\alpha$  to bind to p50/p65 (Huxford et al., 1998), it is unlikely that the hydroxylation modifications would directly interfere with  $\text{I}\kappa\text{B}\alpha/\text{NF}\kappa\text{B}$  complex formation. As such,  $\text{I}\kappa\text{B}\alpha$  was considered to be a useful test case for studying the effects of hydroxylation on ARD stability and stability-regulated biological function. In the manuscript presented in section 6.2, it was found that hydroxylation by FIH produced only a very minor decrease in  $\text{NF}\kappa\text{B}$  binding affinity (for comparison, hydroxylation of  $\text{I}\kappa\text{B}\alpha$  produced a 1.5-fold increase in dissociation constant with  $\text{NF}\kappa\text{B}$ , while stabilising  $\text{I}\kappa\text{B}\alpha$  through conversion of ankyrin repeat 6 to a more “consensus-like” sequence produced a 28-fold increase in this property (Truhlar et al., 2008)). It is probable that this hydroxylation-induced increase in stability was too small to be detected using H/D exchange (Figure 4), and it was likewise found that the stability of  $\text{I}\kappa\text{B}\alpha$  calculated from equilibrium unfolding CD spectra was unchanged (Figure 3B). Interestingly however, the sloping pre-transition baseline (an effect believed to be due to a linear dependence of ankyrin repeat 6 folding on denaturant concentration) observed for the hydroxylated  $\text{I}\kappa\text{B}\alpha$  in Figure 3B was flatter than that measured for the unhydroxylated  $\text{I}\kappa\text{B}\alpha$ , thus raising the possibility that this ankyrin repeat alone has subtly enhanced stability. In light of this, it would be of value to examine the fluorescence of Trp258 (located within ankyrin repeat 6) during denaturant titration (as has been done previously (Truhlar et al., 2008)) to verify if ankyrin repeat 6 is indeed stabilised by FIH-mediated hydroxylation. Regardless of the exact mechanism, the data are a “proof of principle” demonstrating that hydroxylation by FIH can produce a measurable, albeit subtle, change in the interaction properties of a non-synthetic substrate. Furthermore, it should be noted that the residues in  $\text{I}\kappa\text{B}\alpha$  targeted by FIH reside in repeats 4 and 5, while it is both repeats 5 and 6 which lack compact folding. As such, it would be of interest to determine if rational design of an “FIH-stabilisable” ankyrin repeat 6 in  $\text{I}\kappa\text{B}\alpha$  (e.g. by generating target asparaginyl and -2 aspartyl residues in appropriate positions in the repeat) would increase the magnitude of the stabilising effect. If this is found to be the case, this would suggest that ARDs which contain target asparagines in the context of

“consensus-divergent” ankyrin repeats may be more likely to be functionally altered by FIH-mediated hydroxylation. In turn, bioinformatic prediction of these sites may assist in finding functionally important FIH-ARD interactions.

Although the data point to a mild increase in I $\kappa$ B $\alpha$  stability in response to hydroxylation, exactly how great an increase in this property would be required to have a physiologically relevant effect on NF $\kappa$ B function is not known. Of note, a paper examining I $\kappa$ B $\alpha$ / NF $\kappa$ B complex formation (post removal of TNF- $\alpha$ , which promotes resynthesis of I $\kappa$ B $\alpha$ ) in the presence or absence of FIH found there to be no difference in the timing or efficiency of interaction, as judged by co-immunoprecipitation (Cockman et al., 2006). However, it can be assumed that only a proportion of NF $\kappa$ B would be bound to DNA during active signalling, and it is possible that “mopping up” of free NF $\kappa$ B after cessation of the activating signal is unaffected by I $\kappa$ B $\alpha$ 's hydroxylation status, while removal of NF $\kappa$ B from DNA is compromised. If the ratio of free to DNA-bound NF $\kappa$ B is high, then techniques such as co-immunoprecipitation may lack the sensitivity required to detect a reduction in hydroxylated I $\kappa$ B $\alpha$ 's ability to compete with DNA for NF $\kappa$ B binding. More encouragingly, in the same publication it was found that FIH siRNA treatment could subtly repress luciferase expression from an NF $\kappa$ B reporter plasmid in HEK 293T cells (Cockman et al., 2006), which is in agreement with the finding that an absence of hydroxylation may destabilise ankyrin repeat 6 of I $\kappa$ B $\alpha$ , thus leading to improved NF $\kappa$ B binding and repression. To better determine if this is the case, a more sensitive technique, such as kinetic ChIP analysis, may more conclusively determine if FIH can influence NF $\kappa$ B DNA occupancy post stimulus removal and I $\kappa$ B $\alpha$  resynthesis. Moreover, the I $\kappa$ B $\alpha$ /NF $\kappa$ B signalling pathway is responsive to many stimuli, including TNF- $\alpha$ , interleukin-1, lipopolysaccharide, UV light, H<sub>2</sub>O<sub>2</sub> and, of particular relevance to the current study, hypoxia (notably however, hypoxia is considered an activator of NF $\kappa$ B signalling, thus FIH's contribution to I $\kappa$ B $\alpha$  signalling under these conditions, if any, is likely fine tuning during oxygen level flux; reviewed in (Perkins, 2007)). Thus, while current data suggest that FIH is, at best, a minor regulator of TNF- $\alpha$ -regulated I $\kappa$ B $\alpha$  function, it cannot be ruled out that a non-TNF- $\alpha$  stimulus may alter I $\kappa$ B $\alpha$ / NF $\kappa$ B post-translational modifications or binding partners in such a way that FIH-mediated hydroxylation has a more substantial effect on I $\kappa$ B $\alpha$  biology. Thus, future studies examining the influence of FIH on I $\kappa$ B $\alpha$ / NF $\kappa$ B interactions should be expanded to include other NF $\kappa$ B stimuli.

# 7 Final Discussion and Future Perspectives



Since the 18<sup>th</sup> century discovery of oxygen as a gas critical for the sustenance of life (Springer, 2009), a myriad of observations have been made to advance the understanding of oxygen uptake, usage, and how organisms adapt when oxygen levels are insufficient (see section 1.3). Nonetheless, although the “end point” of many hypoxia responses are known, exactly how changes in oxygen levels are communicated to elicit changes in physiological function is still poorly understood. In 2002, the discovery of FIH as an oxygen-utilising enzyme which could oxygen-dependently regulate the activity of the HIF-1 $\alpha$  transcription factor generated interest in this enzyme as a potential oxygen sensor (Hewitson et al., 2002; Lando et al., 2002a). Indeed, the flexible nature of FIH’s substrate binding mechanism supported the notion that FIH could bind and hydroxylate asparaginyl residues in numerous proteins (Elkins et al., 2003). Thus, in this study it was sought to expand FIH’s known substrate repertoire, with a view to better defining FIH’s participation in oxygen sensing.

## 7.1 FIH and ARDs: many targets but few consequences?

Although it is now clear that FIH has a substantial array of ARD-containing substrates and binding proteins, the true extent of the targeted group remains to be clarified. To put this in perspective, amino acid motifs based upon the “FIH preferred” target sequence are found in more than 100 human or mouse ARD-containing proteins (section 3.4, and (Schmierer et al., 2010)). However, mass spectrometric analysis of a number of different ARD-containing substrates purified from cells has shown that the presence of an FIH preferred target sequence alone is not sufficient to explain the substantial variation in ankyrin repeat asparaginyl hydroxylation efficiency. Clearly, variable target asparagine accessibility due to binding of ARD partner proteins, as well as ARD structural stability are also important contributing factors. Moreover, given the considerable flexibility of FIH in terms of primary sequence hydroxylation (see Figure 3.5), ARD folding dynamics may prove a better prediction tool for successful hydroxylation by FIH than use of the “preferred” target sequence (see also discussion of p19INK4d below). With regards to structural stability, the finding that FIH hydroxylates Gankyrin (which presumably requires unfolding of Gankyrin’s prototypical ARD structure) but can also bind Gankyrin when it is constrained as a folded ARD (section 3.7) has added a new dimension to the concept of enzyme recruitment. For example, if it is assumed that the residues involved in FIH’s interactions with folded ARDs are different to those utilised for hydroxylation, this suggests that not all of FIH’s ARD binding partners need have an appropriate “hydroxylatable” (i.e. Asn, Asp or His) residue, thus further expanding FIH’s theoretical catalogue of ARD binding partners. By the same reasoning, these data imply that FIH’s catalytic activity (as opposed to ARD binding

capacity) may be expendable for recruitment of FIH to some of its binding partners. Indeed, this idea could be tested further using catalytically inactive mutants of FIH (discussed below in section 7.3).

Investigation of the effect of hydroxylation on a number of FIH's novel targets has revealed that, in many cases, the modification has little to no influence on the activity of the substrate (see section 3.11.4). Indeed, studies into the effect of hydroxylation on I $\kappa$ B $\alpha$  and G9a in the present work are in agreement with this. As a result, the reason for the existence of wide-spread ARD hydroxylation is still an open question. The sheer number of possible substrates for FIH has generated several theories, including (1) that asparaginyl hydroxylation is a general part of ARD structure (Kelly et al., 2009), and (2) the "ARD pool theory", whereby sequestration of FIH by ankyrin repeats is predicted to fine tune activation of the HIF transcription factors during fluctuation of oxygen levels (Schmierer et al., 2010). The former suggestion was initially deemed unlikely since ARD hydroxylation is often incomplete (Kelly et al., 2009), and this conclusion also clearly supported by the discovery that numerous species lack an FIH homolog (see section 4.2). The latter idea, although difficult to test in practice, is supported by the finding that Notch1 and Rabankyrin-5 ARD hydroxylation is more sensitive to reductions in oxygen concentration than is the HIF-1 $\alpha$  CAD (Singleton et al., 2011), such that altered interactions between FIH and ARDs could adjust the pool of FIH available at a time when HIF-1 $\alpha$  hydroxylation is still occurring. Irrespective of ARD involvement in HIF activity modulation, this theory does not rule out the possibility that large-scale ankyrin repeat hydroxylation is "collateral damage" from the evolution (possibly in pre-metazoan species) of a single, functionally important ARD hydroxylation event. If this were true, one might expect the "collateral" modifications to have been lost over time. The fact that large-scale hydroxylation of ankyrin repeats still exists may indicate that selective pressure for maintenance of the ARD consensus sequence is stronger than any negative selective pressure resulting from "incidental" hydroxylation of a large number of ankyrin repeats (which in turn, implies that hydroxylation has little impact on a majority of ARD's functions, in agreement with current data). Following through on the idea of a single initiator ARD hydroxylation event, the key question becomes: what is the identity of this factor? And equally importantly, have additional functionally important modifications evolved since ARD hydroxylation first began? Given the hundreds of ARD modifications predicted for FIH, finding functionally relevant target proteins is a daunting task. In light of this, if the putative FIH homologs identified in pre-metazoan species prove to be functional enzymes, the simplified biology of these organisms may prove invaluable in the future for identification of ancient homologs of FIH substrates in higher eukaryotes. Moreover, for studies within higher eukaryotes, a number of considerations may assist in focussing future studies. Firstly, in light of the finding that FIH-mediated asparaginyl hydroxylation

can stabilise ARD structure (Hardy et al., 2009; Kelly et al., 2009), it could be hypothesised that ARDs which are the recipient of many hydroxylation modifications are amenable to stabilisation-mediated modulation. Indeed, Yang et al's 2011 study of Ankyrin R D34 (a C-terminal construct of Ankyrin R containing 12 ankyrin repeats and at least 7 hydroxylated residues) showed an increase in thermal stability from 50.3°C for unhydroxylated D34 to 53.5°C for hydroxylated D34 (Yang et al., 2011b). Furthermore, the hydroxylated D34 was shown to have a reduced interaction with cytoplasmic domain of band 3 (CDB3), although the functional significance of this reduction was not investigated further. As for Ankyrin R, a number of other human ARD-containing proteins exist with 12 or more ankyrin repeats and 5 or more hydroxylatable residues (Table 7.1), thus rendering them interesting candidates as functionally relevant substrates of FIH.

Also worthy of consideration as FIH-modulated targets are ARDs for which folding dynamics play an important part in regulating their biological roles. Indeed, precisely this reasoning prompted the presented analysis of I $\kappa$ B $\alpha$  hydroxylation by FIH (see section 3.11.4). However, the most loosely folded ankyrin repeat in I $\kappa$ B $\alpha$  was not targeted by FIH, which may explain the finding that FIH has only a minor influence on I $\kappa$ B $\alpha$  stability and NF $\kappa$ B binding. Nonetheless, other ARDs are known to adopt partially folded states, such as a phosphomimetic substituted mutant of p19INK4d (Low et al., 2009). This mutant, which is designed to replicate p19INK4d phosphorylated at residue Ser 76, displays poor folding of the first 2 of its 5 ankyrin repeats, which may potentially disrupt its interaction with cyclin dependent kinases 4 and 6 (CDK4/6) (Low et al., 2009). Notably, the loops between ankyrin repeats 1-2 and 2-3 contain appropriately positioned aspartyl and asparaginyl residues, respectively, thus it would be of interest to determine the effect of FIH activity on CDK4/6 binding by p19INK4d.

A third characteristic which may aid in the prediction of functionally relevant FIH substrates is the requirement for the pool of hydroxylated substrate to be dynamically regulated. Thus far, there is strong evidence to suggest that no asparaginyl dehydroxylase exists for either HIF or ARD-containing substrates of FIH (Singleton et al., 2011), which in turn implies that effective switching of a pool of substrate from hydroxylated to dehydroxylated would require either hydroxy-substrate degradation followed by resynthesis (as is seen with HIF- $\alpha$ ), or localisation of hydroxylated substrate (e.g. receptor internalisation) such that it is no longer accessible to a particular signalling pathway. Notably, this latter method has been suggested as a means to regulate the prolyl hydroxylation status of the TRPA1 receptor (Takahashi et al., 2011), and a similar mechanism may be responsible for the observed oxygen sensitivity of the FIH substrate, TRPV3 (Karttunen et al., 2014). In essence,

Uniprot ID	Name	Repeats	Target residues <sup>a</sup>
Q9P2R3	Ankyrin repeat and FYVE domain-containing protein 1	21	11
P16157	Ankyrin-1 (Ankyrin-R) **	23	17
Q12955	Ankyrin-3 (Ankyrin-G) **	23	16
Q01484	Ankyrin-2 (Ankyrin-B) **	24	18
Q8NFD2	Ankyrin repeat and protein kinase domain-containing protein 1	12	9
Q8IWZ3	Ankyrin repeat and KH domain-containing protein 1 **	25	11
O15084	Serine/threonine-protein phosphatase 6 regulatory ankyrin repeat subunit A	28	18
Q8N8A2	Serine/threonine-protein phosphatase 6 regulatory ankyrin repeat subunit B ***	28	14
Q9ULJ7	Ankyrin repeat domain-containing protein 50	19	16
Q8NB46	Serine/threonine-protein phosphatase 6 regulatory ankyrin repeat subunit C	28	15
Q9Y283	Inversin	16	6
Q9ULH0	Kinase D-interacting substrate of 220 kDa	12	7
Q9H2K2	Tankyrase-2 *	15	10
O95271	Tankyrase-1	17	11
O75762	Transient receptor potential cation channel subfamily A member 1	15	5
O75179	Ankyrin repeat domain-containing protein 17 **	25	12

a Number of repeats containing an easily identifiable asparaginyl, aspartyl or histidinyl residue in a suitable position for hydroxylation by FIH

Identified as FIH substrates in

\* (Yang et al., 2011a)

\*\* (Cockman et al., 2009; Yang et al., 2011b)

\*\*\* this work

**Table 7.1 Large ARDs with multiple FIH target residues in the human proteome.** Human ARD-containing proteins from the Uniprot database which contain greater than or equal to 12 ankyrin repeats are shown, as well as the number of ankyrin repeats within each that are potential substrates of FIH.

focussing studies on either highly labile proteins or those amenable to “differential compartmentalisation” may also yield functionally relevant FIH substrates.

Lastly, the utility of function-based substrate selection should not be underestimated. Through analysis of HIF- $\alpha$  hydroxylation, FIH has been shown to have a clear, albeit comparatively subtle role



in oxygen-dependent HIF- $\alpha$  regulation (Hewitson et al., 2002; Lando et al., 2002a). As such, other oxygen-sensitive ARD-containing proteins would make logical choices for further investigation as FIH substrates.

## 7.2 FIH and non-ARD substrates/interacting partners

In addition to a number of ARD-containing proteins, the Y2H screen presented here also identified several proteins lacking ankyrin repeats, including ferritin, 4E-T, and WDR62. Although none of these stimulated FIH activity by *in vitro* hydroxylation assay, both 4E-T and WDR62 displayed a weak (DMOG-dependent, in the case of 4E-T) interaction with FIH following co-immunoprecipitation (section 3.5), prompting us to examine further the functional relevance of FIH's interaction with these proteins. WDR62 has recently been characterised as an important regulator of mitotic progression and spindle orientation in neuronal precursors, and defects in its function are a primary cause of autosomal recessive primary microcephaly. As FIH has thus far not been linked to neuronal defects, it is unlikely that FIH is a significant modulator of this facet of WDR62 function. Nonetheless, WDR62 is also documented to scaffold JNK pathway kinases, and an interesting future experiment would involve examining any effects of FIH on binding of these kinases by WDR62, or alternatively, the effect of WDR62 on the phosphorylation status of FIH.

A clearer candidate for further analysis as an FIH interacting protein is the translation modulator, 4E-T. This factor, as well as translation itself, are known to be regulated by oxygen availability (Koritzinsky et al., 2006). While a lack of hydroxylation of 4E-T *in vitro* does not rule out 4E-T being a substrate in cells, it is more likely that any involvement of FIH in oxygen-dependent 4E-T regulation would be modulated indirectly through FIH's availability (see discussion of the "ARD pool theory" in section 3.11.3), or through a multi-protein complex involving FIH, 4E-T, and an as-yet-undefined FIH substrate. Whatever the mechanism, determining if FIH has a role in translation regulation, using e.g. ribosomal profiling with and without modulation of FIH/4E-T, would be of considerable value. Moreover, microarray analysis of genes misregulated in the absence of FIH did not reveal obvious candidates which may account for the metabolic phenotype seen in the FIH KO mice (Zhang et al., 2010). However, an alternative explanation is that FIH's influence is less at the gene level, and more post-transcriptional. As a result, analysis of protein expression via techniques such as SILAC-MS or 2D electrophoresis may also help to ascertain if FIH's influence extends to translational regulation.

As study of FIH's substrate and binding protein repertoire continues, it is important to note that the ARD-containing proteins isolated in the Y2H screen presented herein display a considerably stronger interaction with FIH than those seen with WDR62 or 4E-T (see, for example, Figure 3.8A). Indeed, this bias in interaction efficiency is a clear obstacle to identification of non-ARD-containing FIH substrates, as many screening processes rely on binding strength to isolate novel interacting proteins (Cockman et al., 2006; Cockman et al., 2009). To partially address this difficulty, Cockman et al (2009) utilised the FIH inhibitor DMOG paired with SILAC-MS in an attempt to identify novel targets of FIH through "enzyme stalling". Interestingly, in addition to the 10 ARD-containing proteins identified by this method, the authors revealed that a number of non-ARD-containing proteins were also found, however their identities and status as FIH substrates was not reported (Cockman et al., 2009). Whilst the use of DMOG clearly improved the efficiency of substrate identification, this approach could be further enhanced by utilising chemical crosslinking (e.g. by coupling chemical crosslinking agents with DMOG, similar to e.g. Caprotec's Capture Compounds (Koster et al., 2007)) prior to purification of FIH-substrate complexes, thus ensuring that even the most transient of enzyme-substrate interactions are preserved, and the fullest picture of FIH's interaction network is obtained.

### 7.3 The function of FIH: clues from the FIH KO mouse

As discussed in section 1.7, the predominant phenotype of the FIH KO mice is one of hypermetabolism, with the mice displaying an elevated heart rate, greater food, water and oxygen consumption, and altered glucose and lipid metabolism including a higher sensitivity to insulin (Zhang et al., 2010). However, thus far the molecular details which lead to these changes are largely uncharacterised.

#### 7.3.1 The FIH KO mouse phenotype – is HIF involved?

Despite the phenotype of the FIH KO mouse bearing little obvious resemblance to the phenotypes of other KO mice bearing HIF up-regulation (Adluri et al., 2011; Aragonés et al., 2008; Bishop et al., 2008; Holscher et al., 2011; Takeda et al., 2008; Walmsley et al., 2011), HIF is clearly an important regulator of many aspects of metabolism, and as such, this factor deserves close scrutiny as a possible mediator of the effects seen in the FIH KO mice. For example, the excessive cardiopulmonary function seen in the absence of FIH bears similarity to that observed in sufferers of Chuvash Polycythemia (CP), an autosomal recessive disorder characterised by an R200W mutation in the VHL ubiquitin ligase. This single amino acid mutation in VHL hinders its ability to ubiquitinate and degrade the HIFs, thus permitting accumulation of the HIF transcription factors in normoxic

conditions. Consequently, individuals with CP have significantly elevated serum levels of HIF target genes such as Epo and VEGF, which in turn contributes to thrombosis, vascular abnormalities, pulmonary hypertension, and extensive polycythemia (reviewed in (Gordeuk and Prchal, 2006)). Notably, CP patients are also known to hyperventilate and have an elevated heart rate in normoxia, and these symptoms are believed to be independent of the observed polycythemia (Smith et al., 2006). Together, these observations implicate VHL, and by extension the HIFs, in regulation of basal cardiopulmonary activity. In fact, this idea is supported by more focussed studies of HIF activity in the peripheral and central chemoreceptors. Removal of HIF-1 $\alpha$  from either the carotid body and CNS, or regions of the brain controlling reflex stimulation of breathing alone produces a clear deficit in ventilatory acclimatisation to chronic hypoxia (reviewed by (Prabhakar and Semenza, 2012)). Overall, further investigation of FIH-mediated repression of HIF in these key control centres would be of considerable value.

Much like cardiopulmonary regulation, considerable evidence exists for HIF involvement in modulation of glucose and lipid metabolism. A large proportion of this evidence has been obtained through the study of mouse models which either remove or overexpress HIF-1 $\alpha$  (and HIF-2 $\alpha$  in the case of the overexpression studies) in the liver, adipose tissue, or both (Halberg et al., 2009; Jiang et al., 2011; Kim et al., 2006b; Shin et al., 2012). Whilst the phenotypic details of each model are quite variable, some general trends are apparent. On the whole, HIF overexpression appears to promote weight gain and liver steatosis, and reduces glucose tolerance. Conversely, reduction of HIF levels tends to improve glucose tolerance and insulin sensitivity, provides resistance to weight gain on a high fat diet, and shifts energy production towards fatty acid oxidation (Halberg et al., 2009; Jiang et al., 2011; Kim et al., 2006b). When comparing these trends to those observed in the FIH KO mice, ablation of FIH activity appears to echo the effects of reduced HIF signalling. However, given that FIH's enzymatic activity is known to repress the activity of HIF- $\alpha$  CAD, this result is contrary to expectation. While it is conceivable that such an effect may result from feedback due to chronic HIF activation, it is also possible that FIH's role extends beyond direct hydroxylation of HIF- $\alpha$  to modulation of other proteins which are integral to metabolism.

### 7.3.2 Deciphering the FIH KO mouse – foci for future study

Thus far, few in-roads have been made regarding identifying the molecular pathways in the KO mice which are perturbed in the absence of FIH activity. A microarray study looking at altered gene expression in KO mouse-derived MEFs revealed little that would explain the metabolic changes observed in the mice (Zhang et al., 2010). However, direct analysis of mouse tissues showed that a

small number of genes have altered expression, including reduced *Scd1* in liver (involved in lipogenesis), and adiponectin (involved in insulin sensitisation, and suppression of inflammation and apoptosis) and its regulator PPAR $\gamma$  in WAT (Zhang et al., 2010). Among these, adiponectin is an attractive candidate for contributing to a number of the responses observed in the KO mice, as it has documented roles in promoting pancreatic  $\beta$ -cell survival, decreasing liver-mediated glucose output and lipogenesis, and also contributes to CNS regulation of energy expenditure (reviewed in (Turer and Scherer, 2012)). Whether FIH contributes directly to modulation of adiponectin expression remains to be determined (notably, HIF-1 is documented as a repressor of adiponectin in WAT, thus suggesting that any FIH-mediated effect would not occur through modulation of CAD transactivation (Jiang et al., 2013)). However, it has been observed that hyperoxia (which may be present in KO mouse tissues due to elevated partial pressures of oxygen in arterial blood) promotes PPAR $\gamma$  mRNA expression in WAT (Quintero et al., 2012; Zhang et al., 2010). Thus, it is tempting to speculate that the elevated adiponectin secretion from WAT is a secondary consequence of the hyperventilatory phenotype manifested in the FIH KO mice. In support of an indirect influence of FIH on WAT gene expression, several of the phenotypic effects observed in the global FIH KO mouse (including increased oxygen consumption and insulin sensitivity, and resistance to weight gain on a high fat diet) can be reproduced in a conditional FIH KO mouse wherein gene deletion is achieved via Cre expression from the nestin gene promoter (Zhang et al., 2010). Although nestin-driven Cre is traditionally used for the purposes of generating “neuronal-specific” knockouts (Tronche et al., 1999), it is becoming increasingly recognised that, in addition to strong expression in neural progenitor cells, the nestin promoter has some activity in a number of other tissues, including kidney, heart, lung, muscle, intestine, pancreas, spleen, testis and thymus (Harno et al., 2013). Thus, the notion that the FIH KO phenotype is manifested by abnormal signalling in neuronal-like cells should be treated with caution. Likewise, it should also be noted that nestin-Cre mice themselves have a metabolic phenotype, including a reduced body weight, resistance to weight gain on a high fat diet, and increased insulin sensitivity (Briancon et al., 2010). As a result, the phenotype of the FIH KO mice should additionally be compared to nestin-Cre mice (as well as WT mice, as was done in Zhang et al’s study (Zhang et al., 2010)) in order to ascertain which phenotypic effects are truly due to the loss of FIH. Once FIH-dependent effects in the mice have been verified, it will be important to determine the key changes in protein/gene expression (using techniques such as SILAC/RNA-Seq) in the tissues wherein the nestin promoter is most active, such as brain, kidney and lung (The\_Jackson\_Laboratory, 2014). This, in turn, will assist with isolation of the molecules susceptible to modulation by FIH.

As a complementary technique to specific pathway analysis within the FIH KO mice, knock-in of FIH variants with altered functional capabilities (a highly feasible prospect with the advent of CRISPR/Cas technology) may also provide a means for narrowing down the type of interacting factor(s) which are responsible for generation of the phenotypic effects in the mice. For example, knock-in of a catalytically inactive mutant of FIH (e.g. FIH with a His199 conversion to Ala) would help to determine if the absence of substrate hydroxylation is responsible for the FIH KO mouse phenotype, or if complex formation (as opposed to substrate hydroxylation) is the more important mechanism of action. Likewise, it was found in the current work that the *Tribolium castaneum* ortholog of FIH more effectively hydroxylates ARDs than HIF-1 $\alpha$  compared to its human counterpart (see section 4.5), and as such may be a useful starting point for engineering an ARD-hydroxylation-specific enzyme. Subsequent knock-in into mice of such an enzyme would shed light on the comparative contribution of HIF-1 $\alpha$  versus ARD hydroxylation to FIH's various functional outputs, thus focussing future research on the appropriate substrate class(es).

## 7.4 FIH as a therapeutic target: motivation for better understanding FIH function

Further motivating the need for a clearer understanding of FIH's functions, is the potential of FIH as a therapeutic target. In fact, research has demonstrated that there are at least three separate scenarios in which inhibition of FIH may be considered a worthwhile pursuit. The first of these concerns a possible role for FIH in potentiation of tumour growth, the second exploits the FIH's repression of HIF for treatment of ischaemic disease, while the last stems from the intriguing anti-diabetic effects seen in the FIH KO mice. However, as is always the case, an improved understanding of FIH's cellular roles will be required before the suitability of FIH as a drug target can be properly assessed.

### 7.4.1 FIH in tumour growth

With regards to the first scenario, the HIF transcription factors, as well as their regulatory hydroxylases, have received a great deal of attention as likely modulators of cell metabolism and survival in the tumour microenvironment. Activation of HIF may occur in several different ways. For example, rapidly expanding tumours can sometimes outpace the growth of blood vessels, thus limiting their oxygen and nutrient uptake which in turn activates HIF signalling. Alternatively, it has been recognised that a common feature of tumour cells is the activation of glycolytic metabolism in

the absence of hypoxia (also known as the Warburg effect) which is likely the result of hypoxia-independent activation of HIF (reviewed in (Semenza, 2010)). Conceivably, inactivation of any of HIF- $\alpha$ 's negative regulators, including FIH, could result in aberrant activation of HIF signalling and enhanced tumour growth. For this reason, a potential role for FIH in cancer progression has been examined in multiple cancer types, including breast, renal, colon and pancreatic malignancies, as well as in a number of cancer cell lines. In agreement with FIH's proposed role in HIF-1 $\alpha$  repression, up-regulation of FIH levels via repression of its regulatory miRNAs prevents HIF-1 $\alpha$ -mediated angiogenesis in hypoxic non-small cell lung carcinoma tumours (Chen et al., 2014). Unexpectedly however, FIH activity appears to be pro-tumourigenic both in renal cell culture (Khan et al., 2011) and in mouse tumour xenografts grown from colon carcinoma cells (Pelletier et al., 2012), such that knockdown of FIH could promote apoptosis or reduce tumour growth, respectively. Based on these findings, pharmacological inhibition of FIH may prove to be a useful therapy, but only for specific tumour types.

#### 7.4.2 FIH and treatment of ischaemic disease

Whilst HIF activity is sometimes considered a pathophysiological phenomenon in the context of cancer growth, activation of HIF target genes in tissues affected by ischaemic disease can often be highly desirable. The ischaemia caused by events such as stroke and heart attack can result in substantial cell death, and survival of the affected tissue is often reliant on altered ATP production and usage, rapid reestablishment of functional vasculature, minimisation of ischaemia-reperfusion injury, and recruitment of stem cells to replace damaged tissue (see for example (Eltzschig and Collard, 2004; Kavanagh and Kalia, 2011) and section 1.3). The HIF transcription factors, which transcribe multiple genes contributing to angiogenesis, ATP regulation, erythropoiesis and progenitor cell recruitment and differentiation, are instrumental in this process. Hence, HIF activation may be a useful therapy both after the onset of ischaemic disease, as well as in patients who are risk of ischaemic injury.

This potential has not gone unnoticed in the pharmaceutical industry, and several companies are working towards the development of PHD inhibitors for the treatment of multiple ischaemia-related pathologies. Indeed, positive effects for these inhibitors in animal models of myocardial infarction, or in patients with chronic kidney disease have been reported by multiple pharmaceutical companies (Bao et al., 2010; Besarab, 2010; FibroGen, 2013). Given these encouraging results with PHD inhibitors, it is worthwhile considering the contribution that FIH derepression could make to this process. Although FIH seems to have little effect on HIF target gene activation without prior HIF

stabilisation (see for example (Stolze et al., 2004)), it cannot be ruled out that inhibition of FIH in combination with the PHDs may heighten the effects seen with PHD inactivation alone. In support of this idea, microarray analysis of VHL and/or FIH-deficient MEFs has shown that combinatorial (rather than individual) removal of these two repressors improves normoxic activation of multiple constructs relevant to ischaemic injury, including adrenomedullin, HO-1, VEGFA, and liver glycogen phosphorylase (Zhang et al., 2010). If a similar effect is seen in animal models of ischaemic disease, inhibitors of FIH may prove to be a useful adjunct to the PHD inhibitors currently in development.

#### 7.4.3 FIH inhibitors: a new diabetes therapy?

The third, and newest motivation for better characterisation of FIH stems from studies of the FIH KO mouse. Among the phenotypic effects observed, the knockouts demonstrated increased insulin sensitivity, as well as improved serum lipid and cholesterol homeostasis. Indeed, both of these traits are critical goals in the treatment of type 2 diabetes mellitus (T2DM), as many diabetes patients have dyslipidaemia, and the associated lipotoxicity is believed to play a part in insulin resistance, cardiovascular atherosclerosis and pancreatic  $\beta$ -cell dysfunction and death (Gupta et al., 2012; Imrie et al., 2010). Medications designed to assist with T2DM management include drugs which reduce liver gluconeogenesis, increase insulin secretion, alter starch digestion, or promote the activation of the PPAR $\gamma$  transcription factor (reviewed in (Krentz and Bailey, 2005)). Notably, knockout of FIH in mice produces a significant up-regulation of PPAR $\gamma$  and adiponectin specifically in white adipose tissue, and it is entirely conceivable that this increase contributes to the anti-diabetic phenotype of the mice (see section 1.7.1). Importantly, the KO mice do not appear to have any of the adverse effects (including plasma volume expansion, oedema, increased adiposity, and weight gain (Heald and Cawthorne, 2011)) typically associated with the use of thiazolidinediones or other standard T2DM treatments. As a result, FIH inhibitors are well worth consideration as candidate anti-diabetic therapies, and it will be important to gain a better understanding of FIH's cellular activities in order to ascertain the suitability and safety of inhibition of FIH.

### 7.5 Final summary

The discovery and analysis of the novel ARD-containing substrates FGIF, PP6-ARS-B, Gankyrin, G9a and GLP, combined with analysis of those isolated by other laboratories has confirmed that FIH is highly flexible when binding and hydroxylating ARD targets. Although it was found that I $\kappa$ B $\alpha$  stability and NF $\kappa$ B binding was subtly affected by FIH hydroxylation, this is but one example of many where the modification has only minor (or indeed no) functional effect. Clearly, a more in-depth

understanding of FIH's overall biological role will be required in order to focus future studies on functionally relevant substrates and binding partners. In this respect, the discovery of likely FIH homologs in ancient pre-metazoan eukaryotes has provided a new means for investigating the evolutionary origins of FIH function. In higher eukaryotes, further characterisation of the FIH KO mouse will be key to defining FIH functions. Such a goal may be aided by additional mouse-based studies using a substrate class-specific FIH enzyme, a "prototype" of which was characterised in the *T. castaneum* FIH enzyme as part of this thesis. Finally, the discovery that FIH associates with the PB factor 4E-T suggests a possible role for FIH in post-transcriptional gene regulation, which will be an important consideration when assessing alterations in biological signalling networks in relevant mouse tissues. Given the considerable therapeutic potential of FIH, it is hoped that this work will facilitate further understanding of FIH's biological role in oxygen sensing and metabolism.



# 8 Appendices



**Appendix 8.1** Hits from a Prosite scan (<http://prosite.expasy.org/scanprosite/>) of the UniProtKB database (limited to the *Mus musculus* taxonomy) using the motif L-X(5)-[DE]-[IV]-N-[AV].

Uniprot accession	Gene symbol/UniProt accession	Protein name	Region (aa)	Sequence
Q99LJ2	ABTB1	repeat and BTB/POZ domain-containing protein 1	22-31	LeqrdvEVNV
Q05921	RN5A	2-5A-dependent ribonuclease	225-234	LiqhgaDVNV
Q5SSL4	ABR	Active breakpoint cluster region-related protein	707-716	LmlsdmDINA
Q14DH7	ACSS3	Acyl-CoA synthetase short-chain family member 3, mitochondrial	596-605	LcvlkkDVNA
A2AE25	A2AE25	Adenosine monophosphate deaminase 2 (Isoform L)	38-47	LqefvaDVNV
A2AE26	A2AE26	Adenosine monophosphate deaminase 2 (Isoform L)	99-108	LqefvaDVNV
A2AE27	A2AE27	Adenosine monophosphate deaminase 2 (Isoform L)	298-307	LqefvaDVNV
Q7TS64	Q7TS64	Adrenergic receptor kinase, beta 1	428-437	LipprgEVNA
O08523	TECTA	Alpha-tectorin	1156-1165	LwvkqvDVNV
P97449	AMPN	Aminopeptidase N	746-755	LmeqynEINA
Q9DBT5	AMPD2	AMP deaminase 2	272-281	LqefvaDVNV
B7ZW98	B7ZW98	Ank1 protein	724-733	LlqhqaDVNA
Q0VGY9	Q0VGY9	Ank1 protein	716-725	LlqhqaDVNA
B2RY01	B2RY01	Ankrd11 protein	187-196	LisegaDVNV
B7ZWM7	B7ZWM7	Ankrd11 protein	153-162	LisegaDVNV
Q05DP4	Q05DP4	Ankrd11 protein	187-196	LisegaDVNV
Q99LB5	Q99LB5	Ankrd22 protein	121-130	LlnagvEVNA
Q9EQ29	Q9EQ29	Ankrd36 protein	53-62	LefgdiDVNV
B7ZP20	B7ZP20	Ankyrin repeat and FYVE domain containing 1	641-650	LlehqaDINV
Q810B6	ANFY1	Ankyrin repeat and FYVE domain-containing protein 1	641-650	LlehqaDINV
Q6GQX6	ANKS6	Ankyrin repeat and SAM domain-containing protein 6	121-130	LldhgaDVNA
Q8K0L0	ASB2	Ankyrin repeat and SOCS box protein 2	388-397	LlaarfDVNA
Q9WV71	ASB4	Ankyrin repeat and SOCS box protein 4	238-247	LldnnaEVNA
Q9D1A4	ASB5	Ankyrin repeat and SOCS box protein 5	252-261	LlefgaDINA
Q91ZU1	ASB6	Ankyrin repeat and SOCS box protein 6	155-164	LldlgaDVNA
Q91ZT9	ASB8	Ankyrin repeat and SOCS box protein 8	72-81	LlekgaEVNA
Q8CDP8	Q8CDP8	Ankyrin repeat and SOCS box-containing 15	327-336	LiengfDVNA
Q2VPQ8	Q2VPQ8	Ankyrin repeat and SOCs box-containing 5	252-261	LlefgaDINA

Q6BCN6	Q6BCN6	Ankyrin repeat and SOCS box-containing protein 6	155-164	LldlgaDVNA
A2ANT9	A2ANT9	Ankyrin repeat and sterile alpha motif domain containing 6	121-130	LldhgaDVNA
Q5QNR3	Q5QNR3	Ankyrin repeat domain 36	53-62	LefgdiDVNV
A6H6J9	A6H6J9	Ankyrin repeat domain 50	38-47	LiachaDVNA
Q99NH0	ANR17	Ankyrin repeat domain-containing protein 17	316-325	LlahkaDVNA
Q9D3J5	ANR22	Ankyrin repeat domain-containing protein 22	121-130	LlnagvEVNA
Q8R3P9	ANR32	Ankyrin repeat domain-containing protein 32	823-832	LslpgiDINV
Q8VE42	ANR49	Ankyrin repeat domain-containing protein 49	159-168	LlqhdaDINA
Q9D504	ANKR7	Ankyrin repeat domain-containing protein 7	199-208	LlkkgaDVNA
Q6PD92	Q6PD92	Ankyrin repeat, SAM and basic leucine zipper domain containing 1	168-177	LvahgaEVNA
Q8VD46	ASZ1	Ankyrin repeat, SAM and basic leucine zipper domain-containing protein 1	168-177	LvahgaEVNA
Q02357	ANK1	Ankyrin-1	716-725	LlqhqaDVNA
B2RUJ4	B2RUJ4	Atad5 protein	1082-1091	LgfkifEVNA
Q6PG79	Q6PG79	Atp2b1 protein	133-142	LmqllfnEINA
Q8K314	Q8K314	Atp2b1 protein	675-684	LmqllfnEINA
Q4QY64	ATAD5	ATPase family AAA domain-containing protein 5	1136-1145	LgfkifEVNA
Q32ME1	Q32ME1	ATPase, Ca <sup>++</sup> transporting, plasma membrane 4	970-979	LmqllfnEINA
Q6KAU8	A16L2	Autophagy-related protein 16-2	336-345	LdahlsEVNA
P18572	BASI	Basigin	289-298	LtisnlDVNV
O55107	O55107	Basigin	175-184	LtisnlDVNV
O55108	O55108	Basigin	157-166	LtisnlDVNV
Q99MK8	ARBK1	Beta-adrenergic receptor kinase 1	470-479	LipprgEVNA
Q3UYH7	ARBK2	Beta-adrenergic receptor kinase 2	470-479	LipprgEVNA
Q80XK2	Q80XK2	Bromodomain and PHD finger containing, 1	787-796	LlerldEVNA
B7ZNI3	B7ZNI3	Brpf1 protein	793-802	LlerldEVNA
Q8K2E3	Q8K2E3	Brpf1 protein	498-507	LlerldEVNA
Q99JV4	Q99JV4	Brpf1 protein	281-290	LlerldEVNA
Q08460	KCMA1	Calcium-activated potassium channel subunit alpha-1	429-438	LhkdrdDVNV
Q60649	CLPB	Caseinolytic peptidase B protein homolog	153-162	LlsegaDVNA
Q3URM1	Q3URM1	ClpB caseinolytic peptidase B homolog (E coli), isoform CRA_c	153-162	LlsegaDVNA
Q8CDI7	CC150	Coiled-coil domain-containing protein 150	514-523	LaeskeDINV
Q61330	CNTN2	Contactin-2	515-524	LapssaDINV
P70194	CLC4F	C-type lectin domain family 4 member F	140-149	LgshleDVNA

Q3UW12	CNGA4	Cyclic nucleotide-gated cation channel alpha-4	473-482	LkmnklDVNA
Q60773	CDN2D	Cyclin-dependent kinase 4 inhibitor D	93-102	LvehgaDVNA
Q91YV3	Q91YV3	Cyclin-dependent kinase inhibitor 2D (P19, inhibits CDK4)	93-102	LvehgaDVNA
Q9JHU4	DYHC1	Cytoplasmic dynein 1 heavy chain 1	4534-4543	LeelclEVNV
Q45VK7	DYHC2	Cytoplasmic dynein 2 heavy chain 1	768-777	LnenlpEINV
			1774-1783	LlgkevEINA
Q9R160	ADA24	Disintegrin and metalloproteinase domain-containing protein 24	244-253	LfllsvDINV
Q5U476	Q5U476	Dmd protein	245-254	LrkqaeEVNA
P49717	MCM4	DNA replication licensing factor MCM4	149-158	LviwgtDVNV
Q8K135	K319L	Dyslexia-associated protein KIAA0319-like protein	790-799	LveiildVNV
P11531	DMD	Dystrophin	2897-2906	LrkqaeEVNA
A2A9Z0	A2A9Z0	Dystrophin, muscular dystrophy	2897-2906	LrkqaeEVNA
Q80TP3	UBR5	E3 ubiquitin-protein ligase UBR5	2671-2680	LvngcgEVNV
Q543J1	Q543J1	Elongation of very long chain fatty acids (FEN1/Elo2, SUR4/Elo3, yeast)-like 2	4-13	LkafdnEVNA
Q9JLJ4	ELOV2	Elongation of very long chain fatty acids protein 2	4-13	LkafdnEVNA
Q6PEU2	Q6PEU2	Endothelial PAS domain protein 1	843-852	LtrydcEVNV
P97481	EPAS1	Endothelial PAS domain-containing protein 1	843-852	LtrydcEVNV
Q61304	Q61304	Erythroid ankyrin	317-326	LlqhgaDVNA
Q3V1V3	ESF1	ESF1 homolog	742-751	LvedddfEVNV
Q8K089	Q8K089	Esf1 protein	41-50	LvedddfEVNV
B2RXU2	B2RXU2	ESF1, nucleolar pre-rRNA processing protein, homolog (S cerevisiae)	742-751	LvedddfEVNV
A2CG73	A2CG73	Euchromatic histone lysine N-methyltransferase 2	733-742	LstgqvDVNA
A2CG74	A2CG74	Euchromatic histone lysine N-methyltransferase 2	767-776	LstgqvDVNA
A2CG76	A2CG76	Euchromatic histone lysine N-methyltransferase 2	790-799	LstgqvDVNA
A2CG77	A2CG77	Euchromatic histone lysine N-methyltransferase 2	345-354	LstgqvDVNA
Q91X85	FLVC2	Feline leukemia virus subgroup C receptor-related protein 2	360-369	LhfpgeEVNA
B2RRW5	B2RRW5	Fem-1 homolog c (Celegans)	514-523	LiecgaDVNV
A2RSY9	A2RSY9	G protein-coupled receptor 128	493-502	LktsdsDINV
Q6WJ09	Q6WJ09	GA binding protein subunit beta1	53-62	LlkhgaDVNA
Q3UT69	Q3UT69	GA repeat binding protein, beta 1	90-99	LlkhgaDVNA
Q00420	GABP1	GA-binding protein subunit beta-1	90-99	LlkhgaDVNA
P81069	GABP2	GA-binding protein subunit beta-2	90-99	LvrsgaDVNA
Q3TWB2	Q3TWB2	Glypican 3	153-162	LyilgsDINV
Q9QZF1	Q9QZF1	Glypican-3	153-162	LyilgsDINV

Q99LI8	HGS	Hepatocyte growth factor-regulated tyrosine kinase substrate	95-104	LlkrqvEVNV
B1ATZ1	B1ATZ1	HGF-regulated tyrosine kinase substrate	95-104	LlkrqvEVNV
Q3U4Z7	Q3U4Z7	High density lipoprotein (HDL) binding protein, isoform CRA_d	410-419	LegpteDVNV
Q9Z148	EHMT2	Histone-lysine N-methyltransferase EHMT2	824-833	LstgqvDVNA
Q61221	HIF1A	Hypoxia-inducible factor 1-alpha	805-814	LtsydcEVNA
Q1RLJ1	Q1RLJ1	IA-2beta protein-tyrosine phosphatase	985-994	LtavaeEVNA
O55222	ILK	Integrin-linked protein kinase	86-95	LlqykaDINA
Q8VCY7	Q8VCY7	Josd2 protein	63-72	LgtgnyDVNV
Q9DBJ6	JOS1	Josephin-1	77-86	LgngnyDVNV
Q9CR30	JOS2	Josephin-2	63-72	LgtgnyDVNV
Q6P8V1	Q6P8V1	Kank3 protein	81-90	LleccaDVNV
Q9Z1P7	KANK3	KN motif and ankyrin repeat domain-containing protein 3	699-708	LleccaDVNV
C3VLD3	C3VLD3	Large conductance Ca2+-activated potassium channel DEC variant 1	364-373	LhkdrdDVNV
E3VRZ0	E3VRZ0	Large conductance Ca2+-activated potassium channel DEC variant 10	364-373	LhkdrdDVNV
E3VRZ1	E3VRZ1	Large conductance Ca2+-activated potassium channel DEC variant 11	364-373	LhkdrdDVNV
E3VRY5	E3VRY5	Large conductance Ca2+-activated potassium channel DEC variant 2	310-319	LhkdrdDVNV
E3VRY8	E3VRY8	Large conductance Ca2+-activated potassium channel DEC variant 7	364-373	LhkdrdDVNV
E3VRY9	E3VRY9	Large conductance Ca2+-activated potassium channel DEC variant 8	364-373	LhkdrdDVNV
E3VRZ2	E3VRZ2	Large conductance Ca2+-activated potassium channel ERL variant 12	364-373	LhkdrdDVNV
E3VRZ4	E3VRZ4	Large conductance Ca2+-activated potassium channel ERL variant 14	364-373	LhkdrdDVNV
E3VRZ5	E3VRZ5	Large conductance Ca2+-activated potassium channel ERL variant 15	364-373	LhkdrdDVNV
E3VRZ6	E3VRZ6	Large conductance Ca2+-activated potassium channel ERL variant 16	364-373	LhkdrdDVNV
E3VRZ9	E3VRZ9	Large conductance Ca2+-activated potassium channel ERL variant 22	364-373	LhkdrdDVNV
E3VRY6	E3VRY6	Large conductance Ca2+-activated potassium channel ERL variant 4	364-373	LhkdrdDVNV
E3VRY7	E3VRY7	Large conductance Ca2+-activated potassium channel ERL variant 6	364-373	LhkdrdDVNV
E3VRY4	E3VRY4	Large conductance Ca2+-activated potassium channel VYR variant 1	364-373	LhkdrdDVNV
E3VRZ3	E3VRZ3	Large conductance Ca2+-activated potassium channel VYR variant 13	364-373	LhkdrdDVNV
E3VRZ7	E3VRZ7	Large conductance Ca2+-activated potassium channel VYR variant 20	364-373	LhkdrdDVNV
E3VRZ8	E3VRZ8	Large conductance Ca2+-activated potassium channel VYR variant 21	364-373	LhkdrdDVNV
E3VS00	E3VS00	Large conductance Ca2+-activated potassium channel VYR variant 23	364-373	LhkdrdDVNV

E3VS01	E3VS01	Large conductance Ca <sup>2+</sup> -activated potassium channel VYR variant 24	364-373	LhkdrdDVNV
E3VS02	E3VS02	Large conductance Ca <sup>2+</sup> -activated potassium channel VYR variant 25	364-373	LhkdrdDVNV
E3VS03	E3VS03	Large conductance Ca <sup>2+</sup> -activated potassium channel VYR variant 26	364-373	LhkdrdDVNV
E3VS04	E3VS04	Large conductance Ca <sup>2+</sup> -activated potassium channel VYR variant 27	364-373	LhkdrdDVNV
Q8CDN9	LRRC9	Leucine-rich repeat-containing protein 9	762-771	LkktgeEINV
A2AC66	A2AC66	Lipopolysaccharide binding protein	433-442	LnslypDVNA
Q61805	LBP	Lipopolysaccharide-binding protein	433-442	LnslypDVNA
Q61114	LPLC1	Long palate, lung and nasal epithelium carcinoma-associated protein 1	316-325	LqmdikeINA
Q80ZU7	LPLC3	Long palate, lung and nasal epithelium carcinoma-associated protein 3	131-140	LlqlaaEVNV
B0V2W1	B0V2W1	MCG14251	1926-1935	LiashaDVNA
D3YWP0	D3YWP0	MCG5197	204-213	LaqrdeDVNV
Q9D4J9	Q9D4J9	MCG52616	116-125	LwrrgcEINV
Q9CQI9	MED30	Mediator of RNA polymerase II transcription subunit 30	163-172	LrnliwDINA
Q6KAP1	Q6KAP1	MFLJ00246 protein	1211-1220	LldmgsDINA
			1277-1286	LldkgaDVNA
Q571E1	Q571E1	MFLJ00392 protein	160-169	LldagaDVNA
Q921D5	Q921D5	Minichromosome maintenance deficient 4 homolog ( <i>S cerevisiae</i> )	149-158	LviwgtDVNV
Q80U36	Q80U36	MKIAA0325 protein	1889-1898	Leelc1EVNV
Q5DTZ5	Q5DTZ5	MKIAA1223 protein	171-180	LiachaDVNA
Q5DTQ3	Q5DTQ3	MKIAA4064 protein	978-987	LtavaeEVNA
Q5DTP0	Q5DTP0	MKIAA4087 protein	643-652	LrnlrpEVNV
P62774	MTPN	Myotrophin	54-63	LllkgaDINA
Q01705	NOTC1	Neurogenic locus notch homolog protein 1	2003-2012	LinshaDVNA
Q35516	NOTC2	Neurogenic locus notch homolog protein 2	1961-1970	LincqaDVNA
Q61982	NOTC3	Neurogenic locus notch homolog protein 3	1926-1935	LiashaDVNA
Q8CH77	NAV1	Neuron navigator 1	1696-1705	LvesdsDVNA
Q9Z1E3	IKBA	NF-kappa-B inhibitor alpha	202-211	LvtlgaDVNA
Q8CC86	PNCB	Nicotinate phosphoribosyltransferase	366-375	LaqkgsEVNV
Q8K428	Q8K428	Notch 1 protein	2004-2013	LinshaDVNA
Q7TQ50	Q7TQ50	Notch gene homolog 1 ( <i>Drosophila</i> )	2004-2013	LinshaDVNA
Q6PCM9	Q6PCM9	Notch2 protein	20-29	LincqaDVNA
Q3TSJ9	Q3TSJ9	Novel protein	66-75	LrlihaDINV
A2ASY2	A2ASY2	Novel protein (4931423N10Rik)	116-125	LwrrgcEINV
P25799	NFKB1	Nuclear factor NF-kappa-B p105 subunit	666-675	LvaagaEVNA

B2RRQ6	B2RRQ6	Nuclear factor of kappa light chain gene enhancer in B-cells 1, p105, isoform CRA_e	666-675	LvaagaEVNA
Q99JX7	NXF1	Nuclear RNA export factor 1	468-477	LpktqhDVNA
Q3TJA5	Q3TJA5	Nuclear RNA export factor 1 homolog (S cerevisiae), isoform CRA_b	468-477	LpktqhDVNA
Q5SRW7	Q5SRW7	Pellino 1	268-277	LealrqEINA
Q9R269	PEPL	Periplakin	840-849	LaakftEVNA
Q6Q477	Q6Q477	Plasma membrane Ca <sup>++</sup> transporting ATPase 4 splice variant b	970-979	Lmq1fnEINA
D1FNM8	D1FNM8	Plasma membrane Ca <sup>++</sup> transporting ATPase 4 variant x/e	970-979	Lmq1fnEINA
Q8C102	GALT5	Polypeptide N-acetylgalactosaminyltransferase 5	33-42	LrlsfsEINA
Q8CFW6	Q8CFW6	Ppp1r16a protein	123-132	LldagaDVNA
Q8BM96	GP128	Probable G-protein coupled receptor 128	493-502	LktsdsDINV
Q91V51	TTLL1	Probable tubulin polyglutamylase TTLL1	320-329	LkpwliEVNA
A2APC3	TTLL9	Probable tubulin polyglutamylase TTLL9	355-364	LkpwllEVNA
Q9WU78	PDC6I	Programmed cell death 6-interacting protein	523-532	LckpepEVNA
P08003	PDIA4	Protein disulfide-isomerase A4	460-469	LsesgeDVNA
Q9Z2G0	FEM1B	Protein fem-1 homolog B	518-527	LldcgaEVNA
Q8CEF1	FEM1C	Protein fem-1 homolog C	514-523	LiecgaDVNV
Q8CHG5	K0317	Protein KIAA0317	703-712	LmcgtgDINV
Q8C669	PELI1	Protein pellino homolog 1	268-277	LealrqEINA
Q923M0	PP16A	Protein phosphatase 1 regulatory subunit 16A	123-132	LldagaDVNA
Q9D394	RUFY3	Protein RUFY3	189-198	LlsefyEVNA
A6MDD2	A6MDD2	Protein tyrosine phosphatase receptor type N	965-974	LtavaeEVNA
Q925J6	Q925J6	Protein tyrosine phosphatase receptor type N	857-866	LtavaeEVNA
Q925J7	Q925J7	Protein tyrosine phosphatase receptor type N	860-869	LtavaeEVNA
Q6NVE2	Q6NVE2	Protein tyrosine phosphatase, receptor type, N	969-978	LtavaeEVNA
A3KN68	A3KN68	Protein tyrosine phosphatase, receptor type, N polypeptide 2	985-994	LtavaeEVNA
B1AQN4	B1AQN4	Protein tyrosine phosphatase, receptor type, T	200-209	LrlqnvEVNV
B2RRF0	B2RRF0	Ptprk protein	198-207	LrlgdvEVNA
Q6PDN0	Q6PDN0	Ptprm protein	189-198	LriqnvEVNA
Q6P3E6	Q6P3E6	Ptprn protein	342-351	LtavaeEVNA
B2RX73	B2RX73	Putative uncharacterized protein	301-310	LilgtgEVNV
Q3T992	Q3T992	Putative uncharacterized protein	534-543	LiatggDINA
Q3TB53	Q3TB53	Putative uncharacterized protein	123-132	LldagaDVNA



Q3TB69	Q3TB69	Putative uncharacterized protein	482-491	LmeqynEINA
Q3TJA0	Q3TJA0	Putative uncharacterized protein	268-277	LealrqEINA
Q3TJB1	Q3TJB1	Putative uncharacterized protein	471-480	LqdivsEINA
Q3TLL4	Q3TLL4	Putative uncharacterized protein	95-104	LlkrqvEVNV
Q3TN93	Q3TN93	Putative uncharacterized protein	534-543	LiatggDINA
Q3TRT7	Q3TRT7	Putative uncharacterized protein	38-47	LiachaDVNA
Q3TSK3	Q3TSK3	Putative uncharacterized protein	51-60	LmqfnEINA
Q3TVB8	Q3TVB8	Putative uncharacterized protein	227-236	LkqqkyEINV
Q3TXD4	Q3TXD4	Putative uncharacterized protein	153-162	LlsegaDVNA
Q3TZW2	Q3TZW2	Putative uncharacterized protein	371-380	LinshaDVNA
Q3U017	Q3U017	Putative uncharacterized protein	937-946	LelfdgEVNV
Q3U109	Q3U109	Putative uncharacterized protein	788-797	LlerldEVNA
Q3U1V3	Q3U1V3	Putative uncharacterized protein	470-479	LipprgEVNA
Q3U1W7	Q3U1W7	Putative uncharacterized protein	951-960	LincqaDVNA
Q3U274	Q3U274	Putative uncharacterized protein	66-75	LrlihaDINV
Q3U3U6	Q3U3U6	Putative uncharacterized protein	153-162	LlsegaDVNA
Q3U3Y1	Q3U3Y1	Putative uncharacterized protein	703-712	LmcgtgDINV
Q3U3Y2	Q3U3Y2	Putative uncharacterized protein	632-641	LinshaDVNA
Q3U4G9	Q3U4G9	Putative uncharacterized protein	790-799	LstgqvDVNA
Q3U9W9	Q3U9W9	Putative uncharacterized protein	202-211	LvtlgaDVNA
Q3UA65	Q3UA65	Putative uncharacterized protein	149-158	LviwgtDVNV
Q3UA73	Q3UA73	Putative uncharacterized protein	97-106	LeelclEVNV
Q3UB40	Q3UB40	Putative uncharacterized protein	202-211	LvtlgaDVNA
Q3UCW2	Q3UCW2	Putative uncharacterized protein	805-814	LtsydcEVNA
Q3UES3	Q3UES3	Putative uncharacterized protein	121-130	LlqhgaDVNA
Q3UF47	Q3UF47	Putative uncharacterized protein	189-198	LrnqaeEVNA
Q3UGL0	Q3UGL0	Putative uncharacterized protein	703-712	LmcgtgDINV
Q3UGS9	Q3UGS9	Putative uncharacterized protein	74-83	LlsegaDVNA
Q3UGZ2	Q3UGZ2	Putative uncharacterized protein	362-371	LapssaDINV
Q3UH39	Q3UH39	Putative uncharacterized protein	515-524	LapssaDINV
Q3UI28	Q3UI28	Putative uncharacterized protein	172-181	LiachaDVNA
Q3UJN3	Q3UJN3	Putative uncharacterized protein	90-99	LlkhgaDVNA
Q3UL56	Q3UL56	Putative uncharacterized protein	85-94	LhfpgeEVNA
Q3UMA3	Q3UMA3	Putative uncharacterized protein	95-104	LlkrqvEVNV
Q3UNR3	Q3UNR3	Putative uncharacterized protein	503-512	LvdfveEVNV
Q3UNT7	Q3UNT7	Putative uncharacterized protein	93-102	LlahsaDVNA
Q3UP74	Q3UP74	Putative uncharacterized protein	746-755	LmeqynEINA
Q3UQ23	Q3UQ23	Putative uncharacterized protein	429-438	LhkdrdDVNV
Q3US51	Q3US51	Putative uncharacterized protein	72-81	LvehgaDVNA
Q3UTI6	Q3UTI6	Putative uncharacterized protein	572-581	LtavaeEVNA
Q3UZE1	Q3UZE1	Putative uncharacterized protein	205-214	LldnnaEVNA
Q3V070	Q3V070	Putative uncharacterized protein	115-124	LldnnaEVNA
Q3V1K4	Q3V1K4	Putative uncharacterized protein	371-380	LqdivsEINA
Q3V1M8	Q3V1M8	Putative uncharacterized protein	410-419	LegpteDVNV

Q3V260	Q3V260	Putative uncharacterized protein	265-274	LlecgaDVNV
Q3V2J6	Q3V2J6	Putative uncharacterized protein	171-180	LiatggDINA
Q8BIY0	Q8BIY0	Putative uncharacterized protein	225-234	LiqhgaDVNV
Q8BM07	Q8BM07	Putative uncharacterized protein	262-271	LlldhsEINA
Q8BM19	Q8BM19	Putative uncharacterized protein	265-274	LlldhsEINA
Q8BP38	Q8BP38	Putative uncharacterized protein	176-185	LstgqvDVNA
Q8BP60	Q8BP60	Putative uncharacterized protein	468-477	LpktqhDVNA
Q8BQI6	Q8BQI6	Putative uncharacterized protein	471-480	LqdivsEINA
Q8BTR9	Q8BTR9	Putative uncharacterized protein	155-164	LldlgaDVNA
Q8BVT9	Q8BVT9	Putative uncharacterized protein	428-437	LipprgEVNA
Q8BX62	Q8BX62	Putative uncharacterized protein	505-514	LlehgaDVNA
Q8BXZ6	Q8BXZ6	Putative uncharacterized protein	248-257	LldmgsDINA
			315-324	LldkgaDVNA
Q8BYT6	Q8BYT6	Putative uncharacterized protein	27-36	LihkteDVNA
			93-102	LikhsaDVNA
			228-237	LgveidEINV
Q8C036	Q8C036	Putative uncharacterized protein	93-102	LlahsaDVNA
Q8C075	Q8C075	Putative uncharacterized protein	45-54	LisegaDVNV
Q8C103	Q8C103	Putative uncharacterized protein	724-733	LldmgsDINA
			791-800	LldkgaDVNA
Q8C1Z0	Q8C1Z0	Putative uncharacterized protein	149-158	LviwgtDVNV
Q8C2T8	Q8C2T8	Putative uncharacterized protein	90-99	LvrsgaDVNA
Q8C462	Q8C462	Putative uncharacterized protein	198-207	LrlgdvEVNA
Q8C565	Q8C565	Putative uncharacterized protein	349-358	LiengfDVNA
Q8C6T5	Q8C6T5	Putative uncharacterized protein	371-380	LeelclEINV
Q8C905	Q8C905	Putative uncharacterized protein	787-796	LlerldEVNA
Q8CB55	Q8CB55	Putative uncharacterized protein	357-366	LvesdsDVNA
Q8CCN6	Q8CCN6	Putative uncharacterized protein	153-162	LlsegaDVNA
Q8CDF0	Q8CDF0	Putative uncharacterized protein	72-81	LlekgaEVNA
Q8CDV3	Q8CDV3	Putative uncharacterized protein	244-253	LfllsvDINV
Q9CTJ8	Q9CTJ8	Putative uncharacterized protein	106-115	LkqqkyEINV
Q9CTY1	Q9CTY1	Putative uncharacterized protein	66-75	LlerldEVNA
Q9CU54	Q9CU54	Putative uncharacterized protein	7-16	LrmgtrEINA
Q9D077	Q9D077	Putative uncharacterized protein	149-158	LviwgtDVNV
Q9D1V0	Q9D1V0	Putative uncharacterized protein	14-23	LrlccvDVNA
Q9D294	Q9D294	Putative uncharacterized protein	124-133	LqdivsEINA
Q9D4Y2	Q9D4Y2	Putative uncharacterized protein	72-81	LlekgaEVNA
Q9DA85	Q9DA85	Putative uncharacterized protein	153-162	LlsegaDVNA
P35822	PTPRK	Receptor-type tyrosine-protein phosphatase kappa	198-207	LrlgdvEVNA
P28828	PTPRM	Receptor-type tyrosine-protein phosphatase mu	189-198	LriqnvEVNA
P80560	PTPR2	Receptor-type tyrosine-protein phosphatase N2	985-994	LtavaeEVNA
Q99M80	PTPRT	Receptor-type tyrosine-protein phosphatase T	200-209	LrlqnvEINV

Q60673	PTPRN	Receptor-type tyrosine-protein phosphatase-like N	963-972	LtavaeEVNA
B2RQP1	B2RQP1	Ribonuclease L (2', 5'-oligoadenylate synthetase-dependent)	225-234	LiqhgaDVNV
Q05DT6	Q05DT6	Rnasel protein	225-234	LiqhgaDVNV
Q5GIG6	TNI3K	Serine/threonine-protein kinase TNI3K	186-195	LlkfgaDVNV
Q505D1	ANR28	Serine/threonine-protein phosphatase 6 regulatory ankyrin repeat subunit A	93-102	LlkhsaDVNA
B2RXR6	ANR44	Serine/threonine-protein phosphatase 6 regulatory ankyrin repeat subunit B	27-36	LihkteDVNA
			93-102	LlkhsaDVNA
			228-237	LgveideEINV
Q8BTI7	ANR52	Serine/threonine-protein phosphatase 6 regulatory ankyrin repeat subunit C	93-102	LlahsaDVNA
			877-886	LlqhqaEVNA
Q91WV6	Q91WV6	Solute carrier family 27 (Fatty acid transporter), member 2	503-512	LvdfveEVNV
Q6DIC9	Q6DIC9	Sparc/osteonectin, cwcv and kazal-like domains proteoglycan 1	265-274	LlldhsEINA
B2RXY1	B2RXY1	Spock1 protein	264-273	LlldhsEINA
Q6PFX9	Q6PFX9	Tankyrase, TRF1-interacting ankyrin-related ADP-ribose polymerase	849-858	LleghaDVNA
Q8K0F1	TBC23	TBC1 domain family member 23	437-446	LqghlaDINV
B7ZWA6	B7ZWA6	Tecta protein	1156-1165	LwvkqvDVNV
Q62288	TICN1	Testican-1	265-274	LlldhsEINA
Q8BYY4	TT39B	Tetratricopeptide repeat protein 39B	301-310	LilgtgEVNV
B2RTJ7	B2RTJ7	TNNI3 interacting kinase	186-195	LlkfgaDVNV
B7ZNV4	B7ZNV4	Tnnt1 protein	216-225	LkqqkyEINV
Q8R4D5	TRPM8	Transient receptor potential cation channel subfamily M member 8	601-610	LakvknDINA
Q8K424	TRPV3	Transient receptor potential cation channel subfamily V member 3	234-243	LiaagaDVNA
Q148W9	Q148W9	Transient receptor potential cation channel, subfamily M, member 8	601-610	LakvknDINA
Q32MT9	Q32MT9	Transient receptor potential cation channel, subfamily V, member 3	234-243	LiaagaDVNA
B9EI42	B9EI42	Transmembrane protein 136	119-128	LgesgtEVNA
Q7TQ52	Q7TQ52	Transmembrane receptor Notch1 B	1989-1998	LinshaDVNA
Q7TQ51	Q7TQ51	Transmembrane receptor Notch1 D	1999-2008	LinshaDVNA
O88346	TNNT1	Troponin T, slow skeletal muscle	228-237	LkqqkyEINV
E2RUG8	E2RUG8	TRP receptor like protein	234-243	LiaagaDVNA
A7E1Z2	A7E1Z2	Ttll9 protein	291-300	LkpwllEVNA
A4Q9F4	TTL11	Tubulin polyglutamylase TTL11	432-441	LkpwllEVNA
A6PW21	A6PW21	Tubulin tyrosine ligase-like 1	320-329	LkpwllEVNA
B1AW61	B1AW61	Tubulin tyrosine ligase-like family, member 11	96-105	LkpwllEVNA
Q06180	PTN2	Tyrosine-protein phosphatase non-	235-244	LmekgeDVNV

		receptor type 2		
Q9Z315	SNUT1	U4/U6U5 tri-snRNP-associated protein 1	130-139	LglkplEVNA
B1AY62	B1AY62	Ubiquilin 2	618-627	LiatggDINA
Q8R317	UBQL1	Ubiquilin-1	562-571	LiatggDINA
Q9QZM0	UBQL2	Ubiquilin-2	618-627	LiatggDINA
Q99NB8	UBQL4	Ubiquilin-4	576-585	LiatggDINA
Q66K07	Q66K07	Ubqln1 protein	562-571	LiatggDINA
Q14B51	Q14B51	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 5	33-42	LrlsfsEINA
D3YTV8	D3YTV8	Uncharacterized protein	724-733	LlqhqaDVNA
D3YZA5	D3YZA5	Uncharacterized protein	86-95	LlqykaDINA
D3Z025	D3Z025	Uncharacterized protein	768-777	LnenlpEINV
D3Z322	D3Z322	Uncharacterized protein	93-102	LvehgaDVNA
D3Z425	D3Z425	Uncharacterized protein	265-274	LlldhsEINA
D3Z4K0	D3Z4K0	Uncharacterized protein	53-62	LefgdiDVNV
D3Z5M4	D3Z5M4	Uncharacterized protein	753-762	LlqhqaDVNA
D3Z653	D3Z653	Uncharacterized protein	315-324	LdahlsEVNA
D3Z6W2	D3Z6W2	Uncharacterized protein	235-244	LmekgeDVNV
D6RG98	D6RG98	Uncharacterized protein	186-195	LlkfgaDVNV
E0CYD9	E0CYD9	Uncharacterized protein	30-39	LspdwqEINA
E9PUD3	E9PUD3	Uncharacterized protein	121-130	LldhgaDVNA
E9PUD9	E9PUD9	Uncharacterized protein	293-302	LkpwllEVNA
E9PUE7	E9PUE7	Uncharacterized protein	661-670	LmlsdmDINA
E9PUL2	E9PUL2	Uncharacterized protein	763-772	LkktgeEINV
E9PUP5	E9PUP5	Uncharacterized protein	270-279	LldmgsDINA
			337-346	LldkgaDVNA
E9PUR0	E9PUR0	Uncharacterized protein	1212-1221	LldmgsDINA
			1279-1288	LldkgaDVNA
E9PWK5	E9PWK5	Uncharacterized protein	88-97	LrdrileVNA
E9PX75	E9PX75	Uncharacterized protein	65-74	LlkhgaDVNA
E9PXP1	E9PXP1	Uncharacterized protein	788-797	LlerldEVNA
E9PXS7	E9PXS7	Uncharacterized protein	208-217	LisegaDVNV
E9PY58	E9PY58	Uncharacterized protein	153-162	LlsegaDVNA
E9PYT7	E9PYT7	Uncharacterized protein	95-104	LlkrqvEVNV
E9PYW3	E9PYW3	Uncharacterized protein	265-274	LlldhsEINA
E9Q187	E9Q187	Uncharacterized protein	121-130	LlnagvEVNA
E9Q2H1	E9Q2H1	Uncharacterized protein	2677-2686	LvngcgEVNV
E9Q4F7	E9Q4F7	Uncharacterized protein	187-196	LisegaDVNV
E9Q4F8	E9Q4F8	Uncharacterized protein	153-162	LisegaDVNV
E9Q688	E9Q688	Uncharacterized protein	215-224	LkqqkyEINV
E9Q6Z2	E9Q6Z2	Uncharacterized protein	14-23	LrlccvDVNA
E9Q719	E9Q719	Uncharacterized protein	1156-1165	LwvkqvDVNV
E9Q746	E9Q746	Uncharacterized protein	963-972	LtavaeEVNA

E9Q7E1	E9Q7E1	Uncharacterized protein	514-523	LaeskeDINV
E9Q804	E9Q804	Uncharacterized protein	315-324	LlahkaDVNA
			1232-1241	LldmgsDINA
			1299-1308	LldkgaDVNA
E9Q828	E9Q828	Uncharacterized protein	970-979	Lmq1fnEINA
E9Q9A2	E9Q9A2	Uncharacterized protein	698-707	LlqhgaDVNA
E9Q9N4	E9Q9N4	Uncharacterized protein	762-771	LkktgeEINV
E9Q9R6	E9Q9R6	Uncharacterized protein	793-802	LlerldEVNA
E9Q9S6	E9Q9S6	Uncharacterized protein	189-198	LriqnvEVNA
E9QAI4	E9QAI4	Uncharacterized protein	585-594	LvfhlkDVNA
E9QAZ5	E9QAZ5	Uncharacterized protein	762-771	LkktgeEINV
E9QKG6	E9QKG6	Uncharacterized protein	316-325	LlahkaDVNA
			982-991	LldmgsDINA
			1049-1058	LldkgaDVNA
E9QKU4	E9QKU4	Uncharacterized protein	189-198	LriqnvEVNA
E9QL68	E9QL68	Uncharacterized protein	244-253	LfllsvDINV
E9QL96	E9QL96	Uncharacterized protein	239-248	LlsefyEVNA
E9QLM8	E9QLM8	Uncharacterized protein	716-725	LlqhqaDVNA
E9QM70	E9QM70	Uncharacterized protein	1754-1763	LincqaDVNA
E9QM71	E9QM71	Uncharacterized protein	4534-4543	Leelc1EVNV
E9QMC0	E9QMC0	Uncharacterized protein	514-523	LaeskeDINV
			93-102	LikhsaDVNA
			228-237	LgveidEINV
E9QNN8	E9QNN8	Uncharacterized protein	666-675	LvaagaEVNA
E9QNQ2	E9QNQ2	Uncharacterized protein	27-36	LihkteDVNA
E9QNR3	E9QNR3	Uncharacterized protein	1156-1165	LwvkqvDVNV
E9QNT8	E9QNT8	Uncharacterized protein	716-725	LlqhqaDVNA
E9QNU8	E9QNU8	Uncharacterized protein	349-358	LiengfDVNA
E9QP40	E9QP40	Uncharacterized protein	1702-1711	LinshaDVNA
E9QPP1	E9QPP1	Uncharacterized protein	716-725	LlqhqaDVNA
E9QPU1	E9QPU1	Uncharacterized protein	937-946	Le1fdgEVNV
E9QPV4	E9QPV4	Uncharacterized protein	334-343	LlqhqaDVNA
Q3TYS2	CQ062	Uncharacterized protein C17orf62 homolog	103-112	LlkdiqDVNV
			1279-1288	LldkgaDVNA
E9Q165	E9Q165	Uncharacterized protein{E11,E12,E13}	601-610	LakvknDINA
E9Q1M6	E9Q1M6	Uncharacterized protein{E11,E12}	1212-1221	LldmgsDINA
O35488	S27A2	Very long-chain acyl-CoA synthetase	503-512	LvdfveEVNV
Q8VDJ3	VIGLN	Vigilin	410-419	LegpteDVNV
Q8CIZ8	VWF	von Willebrand factor	934-943	Le1fdgEVNV
Q2IOJ8	Q2IOJ8	VWF	934-943	Le1fdgEVNV
Q05AH2	Q05AH2	Zcwpw2 protein	437-446	LigrlyeEVNV
Q9Z150	Q9Z150	Zinc finger and BTB domain containing 12	447-456	LeagvaEINV
Q68FG0	ZC4H2	Zinc finger C4H2 domain-containing	66-75	LrlihaDINV

		protein		
Q8C208	IKZF4	Zinc finger protein Eos	341-350	LsdlpyDVNA
Q504M3	Q504M3	Zinc finger, C4H2 domain containing	66-75	LrlihaDINV
A2A7T8	A2A7T8	Zinc finger, MYM domain containing 1	373-382	LqdivsEINA
Q05CB3	Q05CB3	Zmym1 protein	373-382	LqdivsEINA
Q6P214	Q6P214	Zmym1 protein	146-155	LqdivsEINA
Q6PG63	Q6PG63	Zmym1 protein	386-395	LqdivsEINA

**Appendix 8.2 Predicted nucleotide sequences for novel FIH homologs.** Shown below are the predicted nucleotide sequences for FIH coding regions that were constructed via analysis of hFIH BLAST-matched ab initio database mRNAs and genomic/EST sequence data. Yellow and blue highlights indicate predicted alternating exons, while regions of red text are novel exons identified in this work based on homology to hFIH and the presence of likely splicing sites. Unhighlighted red text is derived from EST data which could not be matched to genomic sequence, and therefore cannot be divided into exons. "X"s within the *M. Brevicollis* sequence indicate a region of uncertain sequencing quality. Grey text indicates portions of the FIH BLAST match ab initio mRNA which were predicted to be absent from the final FIH coding sequence. The green highlight in *G. theta* FIH represents a the location of a thymidine nucleotide which was removed during construction of *G. theta* FIH.

>Branchiostoma\_floridae\_FIH\_predicted\_coding\_sequence; gi|260802762|  
and manual inspection of JGI 1.0 braFlol

ATGGCTGGTGCAGAGGGCTGGTCGTCAGCTCAACTCCGACAGTATTCCTTCCCTTGGAGACGATTCC  
CCGTCTCTCATGCACGGATCCTGAAGCAGACAGACTCATAAGCGAAGAGAAACCTGTAGTTCTAACAG  
ACACCCACCTAATAGACAGTGTCTCTAAAATGGGACCTGGGTTACCTCCGGGCAAATTTGGGAGGAGGC  
CTGTGCTCCGTGTACGAGTCGGACAATCACAAGTTCATGTACTTTGATGAGAAAAAGGCAAAGGATAG  
AAAAGACTACACGCCATGTACAAAAAGAATGGAAATGCCATTTGAAGAGTTTGTCAAACAGTTGGAAA  
GCTCTGCAGAAACAGAAGGAAAGCGCGTGTACCTGCAGCAGACCCTGAATGATACAGTTGGTAAGAAC  
ATTGTGGTGGACTTCCTTCAGTTCAACTGGGACTGGATCAACAAAACAGCAGCAGAAGCAGAACTGGGG  
GCACCTCACCTCCAACCTCTTGCTGATTGGCATGGCAGGAAATGTAACACCAGTACACTATGATGAAC  
AGCAGAACTTCTTTGCACAAGTAAAAGGATTCAAGCGATGCATCCTGTTTGCACCTGAACAGATTGAG  
TGCTTGTATCCCTACCCTGTGCACCACCCATGTGACAGACAAAGCCAGGTTGACTTTTGATAACCCTGA  
CTACAACAGATTTTCTAAGTTCAGATATGCGTTCAGGGTTGGAGGCAGTGGTGGGCCCAGGGGACGTGC  
TGTACATCCCCATGTACTGTTGGCACCACATTGAGTCCTTGCTAGAGGGAGGGTATACGACATCTGTC  
AATTTCTGGTACAAGGCTGGTCCCACACCCAGAAAACATTGTGTATCCACTGAAGCCCCTGCAGAAGAT  
GGCCATCATGAGAAACCTTGAGAAAATGCTGGCTGAAGCACTAGGAGATTACAGAGAAGTTGGGCCCTT  
TGATAAACACAATGATTCAGGTGCTACACTCTGATGGACATGACGGAGGTCAACGGAAACAAGGTC  
CTGTACACTGAC

>gi|156353101|ref|XM\_001622865.1| Nematostella vectensis predicted  
protein (NEMVEDRAFT\_v1g231221) partial mRNA and manual inspection of  
NEMVEscaffold\_471

ATGGCTGCTGTGTGTCGGATTTGTACACGGATTTGGACGGCCATTTGGTCCGCAAATATTCCTTGATGT  
AATTCAAGTCCCGAGACTGCATTACCAGGACCCTAAAGCCACAGAAATAATTAAGAATGAGTGCCCAG  
TTGTTTTAACCGACTCCGACATAATATCAAGCGCTATGAAATGGGACATTGAGTATTTGAGGGAAAAC  
ATTGGTGATGGAGATTTTGCTGTCTATGCTTCCGAGACAACAAGTTTATGTACTATGACGAGAAGCG  
TGTA AAAA ACTGGCCTCACTTCAGGCCACCAACTGTTAGGATAGACATGAAGTTTGGCAGTTTTACA  
ACAAGTGAACAAGTTTGATCCAACCTATGGCAAAGCAGGAAATTTGAGATTCTATCTTCAGCAGATG  
TTAAATGATCAAGTTGGAAGGAATATTGTTGCAGATTTTCTTGGTTTCAATTGGGCGTGGCTGAACAC  
GATGCAGAAAGAGATGGACTGGGGCCCACTTACCTCTAATCTCCTGTTGATTGGTCTCCCTGGCAACA  
TAACACCTGTCCACTACGATGAGCAGCAAACCTTTTTTGGCAAGTTACAGGATGCAAACGAGTGCTA  
CTGTTCCACCCAGACAAGTTCAAATGTCTTTATCCATTCCCTGTCCATCACCTTGTGATCGGCAAAG  
CCAGGTTGACTTTGATTGCCCTGACTACATAAGATTTCCACTGTTCAAGGAAATCTGTGGTATGGAAG  
CTATGGTCAAGCCTGGTGATGTGCTGTATATCCCAATGTACTGTTGGCATTACGTGGAATCAACTTTG  
AATGGTGAATTACCACATCTGTCAACTTCTGGTACAAGGCCGCAAACCTCCAAGTGAGATATCATT  
CCCTTTGTCCAGCCAACAGAAGATTGCAATTATGAGGAACATCGAGAGAATGCTAGGGGACGCCCTTG

GGAGTCATACTGAGGTTGGGCCCTGCTCAATTTGATGGTGAATGGGAGATATAATTTACCAGAGACA  
GAATCCAAGAAAGGAGAAACCCACAGCTCACATAGT

>gi|762108117|ref|XM\_011438379.1| PREDICTED: Crassostrea gigas  
hypoxia-inducible factor 1-alpha inhibitor-like (LOC105334801),  
mRNA; prefaced by EST 313310952

ATGGCCAAGGAAAAGAAAAATTTGATCTTAAAGAATACCCATTTTCCACGGAAGAAATCCAAGACT  
TAATATAGACAATCCAGAGACAGAAAAACGTATCGCAGCTGGGCTACCAGTTGTGATCACAGATTCCG  
GTTTAGTGCAGGTCAGCGTTGCATTGGGACTTAGATTATCTGCAAGGAAACATTGGAGATGGAAAATTC  
ACTGTGTACAAGTCAAAGAACAAAAGATTTTCAAGTATTTTGGATGATAAAAAAGTTGACACTTTCAAGGA  
TTTTGAAAAACCTATGGAACATTTGGATTTAACCTTTCCAGAGTTTGTCAAAAAATTAAGACAGCAA  
AACCTGGGAAAGACAAGGTGTATTTGCAACAAGCTCTAAATGAAGGTGTTGGTAAGAGGATTGTCGCA  
GATTTTCTTGGATTTAATTGGAGCTGGGTAACGGAGCAGCAGAAGAAGAACAAGTTTGGAGCCCTCAC  
CTCAAATCTTCTCCTCATTGGGATGGAGGGAAACGTAACCTCCTGCCATTATGATGAACAGGAGA  
ACTTCTTTGCACAGATACGAGGATACAAGAGGTTTCTTGTTCATCCAGACCAGTTAAGTGCATGTAT  
CCTTATCCAACCTACCACCCATGTGATAGGCAGAGCCAGGTTGATTTTGGATAATCCAGATTACAAGAG  
ATTTCCAAAGTTAAGGATGTTAAAGGATACGAAACAGTGGTTGGTCCCGGTGATGTGCTGTTTTTGC  
CAATGTACTGTTGGCACCAAGTGAATCTCTCCCTGACCACGGGCACACAATATCTGTGACCTTCTGG  
TACAAGGCTGGACCATTGGGAACGTTGTCTATCCCTTTCTCCACAACAGAAGGTGTCCATGATGAG  
AAACCTGGAGAAGATGATTCATCAGGCCCTGAATAACACTGATGAGATTCGGAATTCATGCAGAACA  
TGGTACTTGGTCGATACACTGATCCTCCTGCT

>Monosiga\_brevicollis\_FIH\_predicted\_coding\_sequence; gi|167524635  
plus manual inspection of MONBRscaffold\_13\_Cont651 3735-5610

ATGGCGTCCAATCCGCGGACCCTTCCCCGCGCTCATGGTCCGCTCCCTCACCAGCAGCCAATGGCCC  
TGCGCCGGTTCGAGCCGACCTTGAGCGCGAGATCTGTGCAGCCTTCCGTCCCCCGTCCCCGTGACCT  
TTCCCTTTTCGGACACGCGAGCTTCCCCGTATTTCCGCGGATGACCAAGACAAGTTGATGCGCCATATG  
CAGGCCGGTTCGCCCCGTCGTCATCACCACCTCATCCCTTTGTGCCCCAGCCGCACGCTGGGACCTGGC  
GTACCTGGGCAAGCACATGGGTAGCGGGCCCTGCCACGTGTACACGGGGCCCAGTCCGACTTTCAGGG  
ACTTTGACCCTGCAAAGAATCAAGGCGGGTACCCTTTGATACCCCCATCACCTCTTCCCAACGCACT  
TTTGC GCGATTTCATCGATATGCTCCGCAAAGCCGGCGGCCAGAACAAGTTTCACTATCTTCAGCAGGC  
CTTGAATGACACAGTTGGGGCTCGCCTGGTTCAAGATTACCTCGGTTTTGACTGGGAGTGGCTCAGGT  
CCATCGTGGTAGCTCTCAAGTGGGGCGCACTCAACTCCAACACGCTCTTTGTTGGACCCCGGTGTG  
CACACCCCATGCACTACGATGAATACCAAAATCTCTTTGCTCAGGGATCCAAGCAATGCACTCTGGT  
TCCGCTTCCAGTTTGGAGCCATGCGCGCATACCGGGTTGGGCATCCGTACGATCGACAGAGCCAAAX  
XXXCTG  
CAACCCGGCGAGGTCTTGTACATCCCACGTAGCTGGTGGCATCACATTCAATCCCTGGAGCAAGTGAC  
GGTTTCGCTCAACTTCTGGTACCACGCTCCCTCAACCCCACTGGGCGCTATTCATCAGCCCTTGGCAG  
CACCCACAAGGTGGCGCTTTTGC GCAACGTTGAGCAGATGCTGTTGACAGCCCTTGGATCACCGACG  
GAGGTTGAGGATTTTCTGCGCCTGTTGGTCAAGGGTGCCTATGATACACGTTGACAATCACCAGTTGGC  
TACATCCAGAGACGTC AAT

> Capsaspora\_owczarzaki\_FIH\_predicted\_coding\_sequence; manual  
inspection of NW\_011887294.1 and predicted mRNA XM\_004349232.2

AGTTGTTTTGCTTGTCAATTTCTCAAAATCCACTCACTGCTGGTGTGAGTTTTGTTGGTTGTTCTGCTC  
CTGCTTTTTCAGTCAATGGATCAAGTCGCGTCGCACATTTTCGCAGCAGATTGAACGCGGAAGGCTGAGAT  
TGCTTGGAGTGAAGCTCGCCGGTGGTCAACTGCCGCCAATTGGGCGTTTGCAGCGCAGCCCATCCCCGCGC  
TACCGTGCCACCGACGATGCTGCCAGGCTCGCCCTGAGAAGCAAGACTCCGTTTGTACTGCTAGACTC  
GAGTATCGCAGCACCAGCAGTCGAGCGCTGGGATTTGGAGTTTCTCAAGGAGCATTGTTGGGAATGTCCA  
ACACGGTGTGTTGTGTCGAGCACTGGCAAAGTCAAGTACTTTGATGAGACTCGCCTGACACCGGGACAG



GAAGCATCACATCATGATGTGCGGAAGCTCACCATGCGATTTCGCCGAATTTCTCGAGTGCATGGCGGT  
CGCAGCACGTGACACCGAACGATTTGCTCCACGACCACAATCGAGTTGCAACGAACCAACGACAAGCA  
CACAACGTGATTCTTGTGTCAACCGACAGCAGCTGCTGCTGTTGCTCCTGCTGATGCTGCTGCTTTT  
GATTCAAATTCGCAA CAACCGCCAGCTGCCAGCGACGACTCTGCAGCACGCATCGCGTGCCAATTTGG  
CATTTCATTCGGGGATGTCGTCTACTTGCAGCAAGGCCTCTCGAACAGCGTCCGGCTTGAAAAATAGTCG  
AATCCTTTCTCCAGTTCAACTGGAAATGCTGGCAGAGCTCCAGCGCGACGCGAATTGGAGCGACTTG  
ACCACCAACCTCTTGTGGTCAGCATGGCTGGTGCAGTCACTCCCGCCATTACGACGAGCAGGAAAA  
CCTCTTTGCTCAGGTGCGCGGTGCCAAGCGGTGTGTGCTCTTTGCTCCCGATCGCTTTCCGTGCTTGT  
ATCCGTATCCAGTGCACCATCCGTGCGATCGGCCAAAGTCAGGTGATTTTCGACAACCCGGATTGGCG  
CGGTTTCCAGATTTTCCGAGCTGCATGGATGGGAATGCATTCTGGAGCCTGGCGAAGTGTCTACAT  
TCCCGGTACTGGTGGCATCACGTTGAGTGCCTCACCGATTCCGTTTCCGTCAATTTTTGGTATCTCG  
TTGGCCCTCCCGAGATTGCCATCCCCTGACCTACCAGCAGAAAAATCCGTGATGCGCAACATTGAA  
AAGATGCTGGGCGAGGCTCTAGGCGATCCTACGCAGCTGTCCAGCTTGCTGTGCGATTTAGTATTAGG  
GCGGTACACGCAC

>Guillardia\_theta\_FIH\_predicted\_coding\_sequence; gi|551662010 and  
manual inspection of Guith1\_scaffold\_29:693332-695003

ACAGGAGGAAAGAAGAGCAGTGTTGGTGGTTGAGAAGGAACGCTCAGATTGGTTGCTGGAAAAAGGCA  
AATCAGCAAGTCTTCTCAAAGGCATTGATCGGCTGAAACAAAAGACGGAAGGCTCACCGTCACGTCTC  
CTGGGAGCGTCAAGATATCTCATCTTACCCCTAAACTATCAGGAGGCATCGGGAGAAAGCCTAGACAG  
CTGCTGTGACGAGCTAACAAGTCACTGAGACTGTAATCGCTAGCTCAAGAAATATACTAAAAACAAG  
ATCAACATACCCAGCAACAAGAATTCGTTACAGGATTTCTCTTTCAAATTGTTATTTCTGTGGCCTG  
TGTAGCTTGGTCAACAGCAGTGCACCTCGAGACAGCACCTCTGCACAGTCTGGATTGGTTTATCCAA  
AGATGGACGAGCACAGCATGGCGTGCCTTCTCTTTTCCATGCCAGCCTGTTTCAGCGCATCTCTTTCGATGA  
TCCAGCTGTGATAGATTTTCTTAGAAGGGAGGAACAGTCATTCTCTCCAATGTCTCTCTCGTTAGGCCCTCTCGT  
TGGGAAGTGAATGTCGATTACTTAAAAGAAATTTCCCAATGATTTCTCTTCCACCGTTCTTCAGAAAGCAA  
CTTTTTCCAATACTGGGACGAAGAGAAGAATGCAGGGCAATACGAATACACACCACCAACATCGAAGATTGAGAT  
GAACTTCCATGACTTCGTCGAAACAATCCGACGGCCGCGAGTCTCATGCGGGAAGAAGGCATTACCTACAGCAGAT  
GTTGGTGACAGGAATGGGAGAGAGGATTCTAGACGATTTTAAGCATGTTGACTGGGGCACTGTGTACGGGTGGAA  
AGCGCAGTTGGGGTTTGGTGATTTGACGACCAATCTTCTGCTGGTTGGAACGACGGGCCATGTGACCCCCGCACA  
CTATGACGAGCAACACAACCTTCTTCTGCCAGCTTCAGGGACGGAAACAAGTCATGTTGTTTCAGCCCCGGCCGAGTG  
GGGAAGATGTATCCGTTTCTCTCCATCATGCATGTGATAGGCAGTCCATGTTGGACGTGTTTCGGATGGGT  
TGGCCAACGTTGGACAACAATGGAAATTTATTTCCGGTTCAGATGCCAGCCACGATCGAATGCTCTTTCCAAA  
CTTCGATCGAGCCTCTGGATATGAGGCCATCGTACATCCCGGTGAGATGTTGTATATTCCAAGCTACTGGTGGCA  
TCATGTGATCAACCTGGAAGAGACTGTCAGCCTGACGTTTTGGTTCAAATGTCCTCTGTTGGCGACCTTCGCCT  
GCCATTTACAAGTACAACGAGTTGCGCTGAGCGGAAACATTGAGAATATGTTCCGGAAGCATGCTGGGCAATCG  
CAACGTGCAGGGACTCATGCTGAATCTGCTGCATGGAGCACTCACGCCGAACAGACGAGAGCGGTAGA  
GGAAGTGAAGAAGCTTTTGCGGCATGTTTTACCTCCTGAGGAAATCAGCGCATTTCTGCAGGAGCTTGCGA  
GCAATCGATTTAATCTTCCAGTGAGAATG




Gt 1 -----MACVPPF-PCQPVQRI  
Mb 1 ---MASNSADPSPRSWSAPSPAANGPAPVEPDLEREICAAFPSPVPVTFPF-RTRELPRRI  
Co 1 MDQVASHIS-----QQIER-----GRLRLPGVSSPVVNCRQLGV-ATQPIPRY  
Sr 1 -----MS-GGG-----EWWQCPC-PLESVPRV  
Hr 1 -----MMAEWNAWTDDQLKTYNF-PIHPLPIC  
Aq 1 -----MAEPKGPSSSKIEMKKYSF-PVKPIQRL  
Ap 1 -----MAGWNESQLRKYNI-KVDKIPTL  
Tc 1 -----MDGDKKPWDPQLRKYDL-NLEEIPRL  
Ph 1 -----MAENNEKNWDQSQLRKYNI-QCDQIPRL  
Zn 1 -----MADKKEWDQSQLRKYDI-QLDEIPRL  
Lg 1 -----MA-----TSVER-----SPESGEDAYSTIRGELKQYSF-QTEPVPRL  
Cg 1 -----MAKEKKNFDLKEYPF-STEEIPRL  
Ac 1 -----MATENTVNSGEKVFVTKKYPF-STQPIPRRL  
Is 1 -----M-----AQSECEGCTGCSSLDFQDYGI-PLEAVPRRL  
Nv 1 -----MAAVSDLYTDLDGHLVRKYSL-DVIQVPRRL  
Amil 1 -----MASV-----VSDKRGTYVDLEGHKVRKYDFLQLSKIPRL  
Ct 1 -----MATVEGGTDGWDENQLRKYTF-PTTQLQRM  
Bf 1 -----MAGAEGWSSAQLRQYSF-PLETIPRL  
Dr 1 -----MAETDGA-----AAFTELRDPGWGDSQLRQYTF-PTRQIPRL  
Xt 1 MAEAGVSAAEVAETERQG-----GGEAELRCPGWSESQ LRSYQF-QTRQIPRL  
Oa 1 ---MV-----IKQLCV-----RLQYTCIKPGVMERAFFR-----  
Hs 1 ---MAATAAEAVASGSSEP-----REEAGALGPAWDESQ LRSYSF-PTRP IPRL  
Gg 1 ---MAAASSSSSSAA-----GCRDGPVGPWSDSQFRHYSF-ETRP IPRL

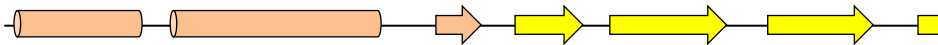


Gt 16 SFDDP-AVIDFLRREEPVILSNVSLVRPLVGKWNVDYLKEIFPMISSPFTSESNFQYW  
Mb 57 SADDQDKLMRHMQAGR PVVITNSSLCAP-AARWDLAYLGKHMGS GPCHVYTGPSPTFRDF  
Co 43 RATDD-AARLALRSKTPFVLLDSSIAAPAVERWDL EFLKEHLGMS-NTV FVSSSTGKVKYF  
Sr 21 RVCDP-KLLDFIREQRPIVITDSKLVES-ARHWDLEYMEKHM GDCKFTVYESKERVFMFS  
Hr 27 SVN DP-QAEELIKEQKPVVIKGS ELIST-AQKWDLEYLREHIGDGEFTVYESDSESFKYY  
Aq 29 CWTEA-TADHLMTHEEPVVLTD TDLVQS-ALHWDLSYLEQNMGP AKHTVYVSKSRYFMYY  
Ap 23 QYD DP-KVDSLLSNKPVLIKGSKLV SQ-VLKWDL DYLAEHMNSICCNVLVSKNHKFKYY  
Tc 27 HYK DP-KIDEYIKENKPVVIT ESNIKPAVQRWSLEYLERNLGHSGHTV FVSRNHKFKYY  
Ph 28 RFD DP-KVLELILQNKPVVILGSDLVKS-TEKWDLEYLEKNMGDSDFTVFQSRNHLFKFF  
Zn 26 SVK DP-KVDEFIADNKPVII TDTRLVAS-AERWDL DYLEQNMGN GDFSVFLSRNHKFKYF  
Lg 37 HHL DP-RVEALMRAEKPVVITD TNLVKS-AFHWDADY LSSNIGSAKQTVYSSYKNKFLYY  
Cg 24 NID NP-ETEKRI AAGLPVVITDSGLVRS-ALHWDLDY LQGNIGDGKFTVYKSKNKRQYF  
Ac 29 YYR DP-KAEHLIQSGQPVILTGSNLVGS-AYHWDMDY LCSNLGCGQYTVY TSSSGKFKYY  
Is 31 SHT DP-EADRLIANMMPVVLTDTGLVAP-ALKWDL DYLSEHLGEGSCTVYQSDTPYFKYY  
Nv 30 HYQ DP-KATEI IKNECPVVLTDSDI ISS-AMKWDIEYLRENIGDGF AVYASEDNKFMYY  
Amil 35 EYD DP-NVESIANEE PVLLVNSNI ISS-ALKWDL DYLKKNLGAGSFSVYSSKTCKFMYY  
Ct 30 SCK DP-RLNDVIAREEPVVITDCN LASS-ASHWSLEYLSSNIGNGTFSVYESD SHLFKYF  
Bf 26 SCT DP-EADRLISEEKPVVLTDT HLIIDS-ALKWDLGYLRANLGGGLCSVYESDNHKFMYY  
Dr 37 SHT DP-RAEVLINNEEPVVLTDTSLVYP-ALKWDLI PYLQENIGNGDFSVYIAENHKFLYY  
Xt 49 QHG DP-RAEELIDKEEPVVLTDTNLVHT-ALKWDL DYLEENIGNGDFSVYSANSHKFLYY  
Oa 27 ---YP--PRSC LICGEPVVLTDTNLVYP-ALKWDL DYLQENIGNGDFSVYSAGTHKFLYY  
Hs 46 SQS DP-RAEELIENEE PVVLTDTNLVYP-ALKWDL EYLQENIGNGDFSVYSASTHKFLYY  
Gg 42 SHS DP-RAEELIENEE PVVLTDTNLVYP-ALKWDL DYLQENIGNGDFSVYSASTHKFLYY

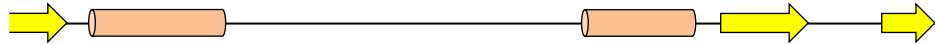
(Appendix 8.3)



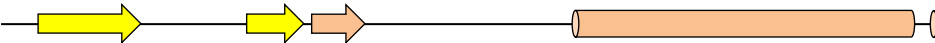
Gt	75	DEEK	NAGQ	-----	YEY	T	P	P	T	S	K	I	Q	M	N	F	H	D	F	V	E	T	I	R	R	P	----	Q	S	H	A	G	R	R	H	Y	L	Q	Q	M	L	V	T	G	M										
Mb	116	DPAK	NQGG	-----	YH	F	D	T	P	I	T	S	S	Q	R	T	F	A	R	F	I	D	M	L	R	K	A	----	G	Q	N	K	F	H	Y	L	Q	Q	A	L	N	D	T	V											
Co	101	DE	T	R	L	T	P	G	-----	Q	Q	F	P	A	A	S	D	D	S	A	-	A	R	I	A	S	Q	F	----	G	I	H	S	G	D	V	Y	L	Q	Q	G	L	S	N	S	V									
Sr	79	DE	E	K	N	L	G	N	-----	Y	K	F	T	P	T	A	T	K	R	S	M	T	F	Q	E	F	A	S	E	F	R	E	A	----	M	K	S	K	D	K	F	L	Y	L	Q	Q	G	L	N	D	T	V			
Hr	85	DD	K	K	V	I	D	G	-----	F	T	P	P	M	K	Q	R	R	M	I	F	S	E	F	Y	D	A	I	K	N	K	----	N	R	T	K	K	Y	Y	L	Q	Q	P	L	N	D	K	V							
Aq	87	DE	R	K	V	N	D	F	-----	P	S	F	K	E	P	M	K	K	R	E	W	T	F	S	K	F	V	D	E	Y	K	S	----	E	G	R	T	D	E	Y	L	Y	Q	D	A	L	T	D	T	V					
Ap	81	D	Q	K	K	I	T	P	N	-----	M	T	F	K	P	I	S	R	P	S	P	M	I	F	S	E	F	A	K	K	I	K	D	W	----	K	K	G	D	S	R	L	Y	M	Q	Q	V	L	N	N	S	I			
Tc	86	DE	K	K	I	Y	N	R	V	S	N	T	K	G	V	E	F	T	P	T	R	K	V	E	M	R	I	E	D	F	M	K	R	V	K	E	W	----	K	K	G	D	E	R	I	Y	L	Y	Q	Q	S	L	T	T	V
Ph	86	DD	K	K	V	N	Q	L	L	S	-	N	Y	K	M	E	F	T	P	T	K	R	I	D	I	K	L	S	E	F	C	Q	R	L	R	E	W	----	K	K	G	D	D	R	L	Y	L	Q	Q	L	N	N	T	V	
Zn	84	DD	K	K	I	L	P	T	S	G	-	D	S	R	L	D	F	T	P	T	R	R	V	D	M	K	L	P	E	F	T	R	K	L	R	Q	W	----	R	R	G	E	R	M	L	Y	L	Q	Q	A	L	N	N	T	V
Lg	95	DD	K	K	L	P	Q	F	-----	P	D	L	Q	L	A	T	K	H	E	E	M	T	F	K	E	F	T	E	K	V	K	N	H	----	K	-	K	G	D	K	I	Y	L	Y	Q	Q	A	L	N	D	S	V			
Cg	82	DD	K	K	V	D	T	F	-----	K	D	F	E	K	P	M	E	H	L	D	L	T	F	P	E	F	V	K	K	L	K	T	A	----	K	-	P	G	K	D	K	V	Y	L	Q	Q	A	L	N	E	G	V			
Ac	87	DE	K	K	M	P	E	N	-----	K	D	F	K	P	T	S	H	T	D	M	T	F	Q	E	F	A	A	K	L	K	Q	E	----	R	N	N	E	K	E	R	M	Y	L	Q	Q	P	L	N	D	T	V				
Is	89	DE	T	K	V	R	D	H	R	L	----	T	D	F	R	A	P	T	L	R	H	E	W	K	I	G	T	F	A	Q	R	L	R	S	A	----	G	H	N	K	Y	Y	L	Q	Q	M	L	N	E	T	V				
Nv	88	DE	K	R	V	K	N	W	-----	P	H	F	R	P	T	V	R	I	D	M	K	F	E	Q	F	Y	K	V	N	K	F	D	P	T	M	A	K	A	G	N	L	R	F	Y	L	Q	Q	A	L	N	D	Q	V		
Amil	93	DE	K	R	A	K	Q	W	-----	P	Y	F	I	P	T	Q	R	I	H	M	K	F	E	E	F	C	S	R	V	R	N	L	----	K	A	S	D	S	R	V	Y	L	Q	Q	M	L	D	D	S	V					
Ct	88	DE	K	K	I	P	G	H	-----	K	D	F	R	P	E	M	R	R	K	E	M	K	F	N	D	F	V	A	L	M	Q	N	P	----	E	----	Q	R	R	F	Y	L	Q	Q	P	L	N	D	T	V					
Bf	84	DE	K	K	A	K	N	R	-----	K	D	Y	T	P	C	T	K	R	M	E	M	P	F	E	F	V	K	Q	L	E	S	S	----	A	E	T	E	-	G	K	R	V	Y	L	Q	Q	T	L	N	D	T	V			
Dr	95	DE	K	K	M	V	N	F	-----	Q	D	F	V	P	K	S	R	R	I	E	M	K	F	S	E	F	V	D	K	M	H	Q	T	----	E	E	Q	G	G	K	R	V	Y	L	Q	Q	T	L	N	D	T	V			
Xt	107	DE	K	K	M	V	N	F	-----	K	N	F	K	P	K	S	S	R	E	E	M	K	F	P	E	F	V	N	K	L	K	D	I	----	Q	E	R	E	S	D	E	R	L	Y	L	Q	Q	T	L	N	D	T	V		
Oa	81	DE	K	K	M	A	N	F	-----	K	N	F	K	P	R	S	N	R	E	E	M	K	F	H	E	F	V	E	K	L	Q	D	I	----	E	R	R	G	S	E	E	R	L	Y	L	Q	Q	T	L	N	D	T	V		
Hs	104	DE	K	K	M	A	N	F	-----	Q	N	F	K	P	R	S	N	R	E	E	M	K	F	H	E	F	V	E	K	L	Q	D	I	----	Q	Q	R	G	G	E	E	R	L	Y	L	Q	Q	T	L	N	D	T	V		
Gg	100	DE	K	K	M	A	N	F	-----	K	N	F	K	P	K	S	S	R	E	E	M	K	F	A	E	F	V	D	R	L	Q	E	I	----	Q	Q	K	G	S	A	E	R	L	Y	L	Q	Q	T	L	N	D	T	V		



Gt	125	GER	I	L	D	D	F	K	H	V	D	W	G	T	V	Y	G	W	K	A	Q	L	G	F	G	D	L	T	N	L	L	L	V	G	T	T	G	H	V	T	P	A	H	Y	D	E	Q	H	N	F	F	C	Q	L	Q	R	K		
Mb	165	GAR	L	V	Q	D	Y	L	G	F	D	W	E	L	R	S	I	V	V	A	L	K	W	G	A	L	N	S	N	T	L	F	V	G	P	P	G	V	H	T	P	M	H	Y	D	E	Y	Q	N	L	F	A	Q	--	G	S	K		
Co	145	GLK	I	V	E	S	F	L	Q	F	N	W	K	L	L	A	E	L	Q	R	D	A	N	W	S	D	L	T	N	L	L	L	V	S	M	A	G	A	V	T	P	A	H	Y	D	E	Q	E	N	L	F	A	Q	V	R	G	A	K	
Sr	129	GPT	I	V	K	D	F	V	G	F	N	W	T	W	L	T	D	V	V	K	A	L	N	W	G	H	L	T	N	M	L	F	V	S	M	P	H	L	V	T	P	C	H	Y	D	E	Q	E	N	F	F	A	Q	V	R	G	T	K	
Hr	131	GPQ	I	V	V	D	F	I	N	F	N	W	K	W	I	K	Q	Q	M	N	N	N	W	G	P	L	T	S	N	L	L	L	I	S	E	E	G	N	V	T	P	C	H	Y	D	E	Q	E	N	F	Y	S	Q	A	N	G	Y	K	
Aq	137	GSG	I	K	H	D	F	V	H	F	N	W	H	V	S	E	R	K	K	R	Y	N	W	G	L	T	S	N	L	L	L	I	S	Q	P	G	N	I	T	P	C	H	Y	D	E	Q	H	N	F	F	C	Q	V	R	G	L	K		
Ap	130	SQR	I	V	S	D	F	V	M	F	N	W	D	W	I	M	S	K	Q	K	V	C	K	W	S	L	T	S	N	L	L	L	I	A	Q	E	G	N	V	T	P	C	H	Y	D	E	Q	Q	N	M	F	A	S	I	R	G	Y	K	
Tc	141	GNN	I	V	E	D	F	V	K	F	D	W	D	Y	V	N	G	K	Q	T	K	H	N	W	G	P	L	T	S	N	L	L	F	I	A	M	E	G	N	Q	T	P	C	H	Y	D	E	Q	E	N	F	F	A	Q	V	Q	G	Y	K
Ph	140	GPA	I	V	D	D	F	L	H	F	K	W	D	L	S	N	I	Q	K	L	S	N	W	G	P	L	T	S	N	L	L	L	I	S	M	E	G	N	V	T	P	C	H	Y	D	E	Q	Q	N	L	F	A	Q	I	T	G	Y	K	
Zn	138	GPA	I	V	Q	D	F	L	N	F	R	W	E	W	I	N	S	K	Q	K	M	H	S	W	G	P	L	T	S	N	L	L	L	I	A	M	E	G	N	V	T	P	C	H	Y	D	E	Q	Q	N	F	F	A	E	V	R	G	F	K
Lg	144	GKS	L	V	L	D	F	I	K	F	N	W	S	W	A	T	E	Q	Q	K	K	Y	W	G	A	L	T	S	N	L	L	L	V	G	M	D	G	V	I	T	P	A	H	Y	D	E	Q	Q	N	F	F	A	I	G	F	K			
Cg	131	GKR	I	V	A	D	F	L	G	F	N	W	S	W	T	E	Q	Q	K	N	K	F	G	A	L	T	S	N	L	L	L	I	G	M	E	G	N	V	T	P	A	H	Y	D	E	Q	E	N	F	F	A	I	R	G	Y	K			
Ac	137	GKW	I	V	M	D	F	L	T	F	N	W	E	W	A	N	Q	Q	K	T	N	R	W	G	P	L	T	S	N	L	L	L	V	G	M	E	G	V	I	T	P	A	H	Y	D	E	Q	E	N	L	F	A	Q	I	Q	G	Y	K	
Is	139	GQS	I	M	L	D	F	L	R	F	N	W	D	W	N	R	Q	K	H	N	G	W	G	P	L	T	S	N	L	L	L	V	G	T	G	N	V	T	P	A	H	Y	D	E	Q	Q	N	F	F	A	H	L	R	G	H	K			
Nv	142	GRN	I	V	A	D	F	L	G	F	N	W	A	L	N	T	M	Q	K	E	M	D	W	G	P	L	T	S	N	L	L	L	I	G	L	P	G	N	I	T	P	V	H	Y	D	E	Q	Q	N	L	F	C	Q	V	T	G	C	K	
Amil	142	GKR	I	V	K	D	F	L	E	F	K	W	Q	L	T	E	M	Q	R	K	M	N	W	S	L	T	S	N	L	L	L	I	G	L	P	G	N	I	T	P	V	H	Y	D	E	Q	Q	N	F	F	G	Q	L	T	G	F	K		
Ct	135	GPQ	V	V	K	D	F	L	Q	F	N	W	D	E	A	K	Q	Q	K	R	N	G	W	G	P	L	T	S	N	L	L	L	I	G	M	A	G	N	V	T	P	C	H	Y	D	E	Q	E	N	L	F	A	Q	V	R	G	Y	K	
Bf	134	GKN	I	V	V	D	F	L	Q	F	N	W	D	W	I	N	K	Q	Q	K	Q	N	W	G	H	L	T	S	N	L	L	L	I	G	M	A	G	N	V	T	P	V	H	Y	D	E	Q	Q	N	F	F	A	Q	V	K	G	F	K	
Dr	146	GRK	I	V	V	D	F	L	G	F	N	W	N	W	I	N	K	Q	Q	A	K	R	N	W	G	P	L	T	S	N	L	L	L	I	G	M	E	G	N	V	T	P	A	H	Y	D	E	Q</											




Gt	185	QVMLFSPAEW	GKMYPFPLH	HACDRQ	SMMP--SH	DRMLFPNF	DRASGYE	AI	VHPGEMLYIP
Mb	223	QCTLVPPSQ	FEAMRAYRV	GHPYDRQ	SQVXXXXXXXXXXXXXXXXXXXXXXXXX	LQ	PGEVLYIP		
Co	205	RCVLFAPDR	FPCLYPYPV	HHPCDRQ	SQVDFDNP	LARFPR	FSELHGWE	CILE	PGEVLYIP
Sr	189	RVILFHDPN	FRCLYPYR	YGHPCDRQ	SQVDFDNP	DYERFPK	FKDARGLE	AILR	PGDVLYIP
Hr	191	RVILFPPSQ	YKYLPHPV	YHPHCDRQ	SQVDFDEP	NLKLFPK	FKKVVGF	GTVL	PGDVLYIP
Aq	197	RCLLFAPDQ	YDKLYPV	VAHPCDRQ	SQVDFDS	PDFERFPK	FKEIEGYE	CILS	PGDVLYIP
Ap	190	RFILFPPSE	FECLYPHP	VHHPYDRQ	SQVDFDNP	DIKFPK	FKEACGYE	VI	VGPEDEVLYIP
Tc	201	RCILFPPSQ	FECLYPYV	HHPHCDRQ	SMVDFER	PDYKFPK	FKNVK	GWEAV	VGPGDVLYIP
Ph	200	RCILFPPSQ	FECLYPYV	HHPHCDRQ	SQVDFENP	DLKFPK	FSQVKGQ	ETVL	GP
Zn	198	RCILFPPSQ	FECLYPYV	HHPHCDRQ	SQVDFER	PDYTRFPK	FREARGQ	EAI	VGPGDVLYIP
Lg	204	RFLLFDP	SMFECLYP	PVYHPHCDRQ	SQVDFDNP	HERFPN	FRNL	LYGQ	EAVGPGDVLYIP
Cg	191	RFILFHDP	QFKCMYP	PTYHPCDRQ	SQVDFDNP	YKRFPK	FKDV	KGYE	TVVGPGDVLYIP
Ac	197	RFLLFDP	QFECLYP	PIYHPHCDRQ	SQVDFDNP	DFEKFPK	FADL	KGYE	GIVGPGDVLYIP
Is	199	RFLFSPDQ	YGCLYP	HPVHHPHCDRQ	SQVDFSD	PDL	SRFPE	FAHL	RGWETVLGPGDVLYIP
Nv	202	RVLLFHPD	KFKCLYP	FVHHPCDRQ	SQVDFD	CPDI	IRFPL	FKEIC	GMEAMV
Amil	202	RVILFHDP	QFKCLYP	FPLYHPCDRQ	SQVDFDNP	DFKRFPK	LKEL	KGYE	VI
Ct	195	RVILFPPSQ	FSCLYPY	VHHPHCDRQ	CQVDFEN	PDYERFP	PLFKD	VAGQ	EAVL
Bf	194	RCILFAPQ	IECLYP	PVHHPCDRQ	SQVDFDNP	YNRFPK	FRYAS	GLEAV	VGPGDVLYIP
Dr	206	RCILFPPSQ	FDCLYPY	PVHHPCDRQ	SQVDFEN	PDYDKFPN	FKN	AVGYE	AVVGPGDVLYIP
Xt	218	RCILFPPSQ	FECLYP	PVHHPCDRQ	SQVDFEN	PDFERFPN	FRNL	LYGQ	EAVGPGDVLYIP
Oa	192	RCILFPPD	QFECLYP	PVHHPCDRQ	SQVDFDNP	YERFPN	FQNV	VGYE	TVVGPGDVLYIP
Hs	215	RCILFPPD	QFECLYP	PVHHPCDRQ	SQVDFDNP	YERFPN	FQNV	VGYE	TVVGPGDVLYIP
Gg	211	RCILFPPD	QFECLYP	PVHHPCDRQ	SQVDFDNP	YEFKFPN	FRS	VVGYE	TVVGPGDVLYIP



Gt	243	SYWWHHV	INLE---	ETVSLTF	FWFKC--	PPVGD	LRLP	ISQV	QRVAL	RNIEN	NMFG	SMLG	NR
Mb	283	RSWWHHI	QSLE---	QVTVSLN	FWYHAP	STPLGA	THQPL	AAPH	KVALL	RNV	QMLL	TALG	SP
Co	265	AYWWHHV	ESLTD---	SVSVNF	WYLV--	GPP-E	IAHPL	TYQQ	KISV	MRN	IEKML	GEAL	GD
Sr	249	RCWWHLV	RSLD---	ELSVSN	FWYMS	PPIAED	IRFPL	TASQ	KVAM	MRN	IEQMLL	TALG	SA
Hr	251	SYWFFHY	FETL	KNSGI	ATAIT	FWYKA--	APVGN	ITYPL	TAVQ	L	MAMRN	TEKML	EAL
Aq	257	MYWWHTI	ETSP-	GELSI	SIT	FWYRG-	GPVPS	KITYPL	SSQ	KVAI	TRN	IEKML	HEAL
Ap	250	MYWFFH	VESLM	HGCT	TVSVNF	WFKA--	GSVEK	IEYPL	LDH	QK	MVIMRN	VEKML	AEL
Tc	261	IYWWHHI	ESLLR	GGPT	TVT	VNFYK	--GPS-	TLEYPL	KDH	QK	VSI	TRN	VEKML
Ph	260	IYWWHHI	ESLMR	GGYT	TF	SIN	FWYKA--	GPTGP	ITYPL	KSR	QKVA	IMRN	VEKML
Zn	258	IYWWHHI	ESIMR	GGYT	TVSVNF	WYKA--	GPTGQ	IVYPL	KDH	QKVA	IMRN	VEKML	VEAL
Lg	264	IYWWHQ	VESVP	NHGM	TVSVNF	WYKS--	GTEEK	IVYPL	KPQ	KVAM	MRN	IEKML	VTSA
Cg	251	MYWWHQ	VESLP	DHGT	ISVTF	WYKA--	GPIGN	VVYPL	SPQ	KVSM	MRN	LEKML	IHQAL
Ac	257	MYWWHQ	VESLS	HGET	ISIT	FWYKA--	AQTMK	VVYPL	TAG	QKVA	IMRN	IEKML	IAEAL
Is	259	MYWWHQ	VESAP	GKDY	TVSVNF	WYKA--	APVDK	VIYPR	MGH	QKVA	IMRN	IEKML	VLEAL
Nv	262	MYWWHY	VESL	NGGIT	TVSVNF	WYKA--	GQTP	SEI	PLSS	QKVA	IMRN	IERML	GDAL
Amil	262	MYWWHHV	ESTM	NSGIT	TVSVNF	WFKA--	GPTPS	QITH	PLTA	QKVA	IMRN	IERML	GEAL
Ct	255	MYWFFH	HFESL	LDGGL	TVT	SVTF	WYKA--	PPVGK	VEYPL	KPQ	KVAM	MRN	IEKML
Bf	254	MYWWHHI	ESLLE	GGYT	TVSVNF	WYKQ	AGPT	PRN	IVYPL	KPV	QKVA	IMRN	LEKML
Dr	266	MYWWHHI	ESLLN	GGIT	TVSVNF	WYKQ	APT	PKR	IEYPL	KAH	QKVA	IMRN	IEKML
Xt	278	MYWWHHI	ESLMD	GGIT	TVSVNF	WYKQ	APT	PKR	IEYPL	KAH	QKVA	IMRN	IEKML
Oa	252	MYWWHHI	ESLLN	GGIT	TVSVNF	WYKQ	APT	PKR	IEYPL	KAH	QKVA	IMRN	IEKML
Hs	275	MYWWHHI	ESLLN	GGIT	TVSVNF	WYKQ	APT	PKR	IEYPL	KAH	QKVA	IMRN	IEKML
Gg	271	MYWWHHI	ESLLN	GGIT	TVSVNF	WYKQ	APT	PKR	IEYPL	KAH	QKVA	IMRN	IEKML

(Appendix 8.3)



Gt	298	NEISAF	LQELAS	NR	F	NLPVRM	-----
Mb	341	TEVEDF	LRLLVK	GRY		DTRDNHQLATSRDVN	-----
Co	319	TQLSSL	SDLV	LGRY	TH	-----	-----
Sr	307	GEVSDF	LRMLV	SGRY		DTDANHTFRQLEM	-----
Hr	309	KQVDIF	WNMLAA	GRY		CKDVOQDDTDEEDAGVGN	DASVVAKLGDIGRCTGHGDEAKNV
Aq	315	NEVGPLL	HMICD	GRY		-----	-----
Ap	308	NEVGNML	NTIVL	GRY	TSDTN	-----	-----
Tc	318	KEVGPLL	RSMLV	LGRY	TE	-----	-----
Ph	318	REVGPL	FRALV	LGRY	AEASSKID	NN	-----
Zn	316	HEVGPLL	RALV	LGRY	TD	-----	-----
Lg	322	EEVPHLL	KTMV	LGRY	T	-----	-----
Cg	309	DEIPEFM	QNMV	LGRY	TDPPA	-----	-----
Ac	315	SEMSPF	MSNMV	LGRY	T	-----	-----
Is	317	NEVGPLL	RAMV	LGRY	T	-----	-----
Nv	321	TEVGPLL	NLMVN	GRY	NLPETESK	KGETHSSHS	-----
Amil	321	HEVGPLL	NCMVN	GRY	TENNDEKDY	GADVLS	-----
Ct	313	HEVAPFM	RNMV	LGRY	TSDDL	-----	-----
Bf	314	REVGPL	INTMI	QGRY	TLMDMTEV	NGNKVLYTD	-----
Dr	325	HEVGPLL	NMMIK	GRY	DHGLS	-----	-----
Xt	337	QEVGPLL	NTMV	KGRY	D	-----	-----
Oa	311	QEVGPLL	NTMI	KGRY	D	-----	-----
Hs	334	QEVGPLL	NTMI	KGRY	N	-----	-----
Gg	330	QEVGPLL	NMMIK	GRY	D	-----	-----

**Appendix 8.3 Alignment of known and predicted FIH homologs.** FIH homologs were aligned using Clustal Omega, and residues strongly or partially conserved are shown with cyan and grey highlights, respectively. Residues involved in Fe<sup>2+</sup> coordination (red), 2-OG binding (dark blue), and target asparagine positioning (pink) are indicated. The secondary structure of hFIH is depicted above the alignment, with yellow arrows indicating the beta strands which make up the DSBH. Amino acid numbers are shown to the left of the alignment. Alignment shading was performed using the BoxShade Server. “x”s in the *Monosiga brevicollis* sequence were included due to a region of questionable sequence quality in the predicted *M. brevicollis* FIH gene. Species abbreviations and sequence IDs as follows (for those without IDs, see Appendix 8.2): Gt = *Guillardia theta*, Co= *Capsaspora owczarzaki*, Mb = *Monosiga brevicollis*, Sr = *Salpingoeca rosetta*, XP\_004998739.1; Hr = *Helobdella robusta*, XP\_009025637.1; Aq = *Amphimedon queenslandica*, XP\_011405699.1; Ap = *Acyrtosiphon pisum*, J9JSB5; Tc = *Tribolium castaneum*, XP\_975469.1; Phc = *Pediculus humanus corporis*, XP\_002428892.1; Zn = *Zootermopsis nevadensis*, KDR18274.1; Lg = *Lottia gigantea*, XP\_009061887.1; Cg = *Crassostrea gigas*, Ac = *Aplysia californica*, XP\_005092336.1; Is = *Ixodes scapularis*, XP\_002412378.1; Nv = *Nematostella vectensis*, Amil = *Acropora millepora*, Ct = *Capitella teleta*, ELU06865.1; Bf = *Branchiostoma floridae*, Dr = *Danio rerio*, NP\_958904.1; Xt = *Xenopus tropicalis*, NP\_988915.1; Oa = *Ornithorhynchus anatinus*, XP\_001514980.2; Hs = *Homo sapiens*, NP\_060372.2; Gg = *Gallus gallus*, XP\_426507.2.



# 9 References





- Aasland, R., Abrams, C., Ampe, C., Ball, L.J., Bedford, M.T., Cesareni, G., Gimona, M., Hurley, J.H., Jarchau, T., Lehto, V.P., *et al.* (2002). Normalization of nomenclature for peptide motifs as ligands of modular protein domains. *FEBS Lett* *513*, 141-144.
- Abu-Farha, M., Niles, J., and Willmore, W.G. (2005). Erythroid-specific 5-aminolevulinate synthase protein is stabilized by low oxygen and proteasomal inhibition. *Biochem Cell Biol* *83*, 620-630.
- Adachi, T., Ishikawa, K., Hida, W., Matsumoto, H., Masuda, T., Date, F., Ogawa, K., Takeda, K., Furuyama, K., Zhang, Y., *et al.* (2004). Hypoxemia and blunted hypoxic ventilatory responses in mice lacking heme oxygenase-2. *Biochem Biophys Res Commun* *320*, 514-522.
- Adluri, R.S., Thirunavukkarasu, M., Dunna, N.R., Zhan, L., Oriowo, B., Takeda, K., Sanchez, J.A., Otani, H., Maulik, G., Fong, G.H., *et al.* (2011). Disruption of hypoxia-inducible transcription factor-prolyl hydroxylase domain-1 (PHD-1<sup>-/-</sup>) attenuates ex vivo myocardial ischemia/reperfusion injury through hypoxia-inducible factor-1 $\alpha$  transcription factor and its target genes in mice. *Antioxid Redox Signal* *15*, 1789-1797.
- Akiri, G., Nahari, D., Finkelstein, Y., Le, S.Y., Elroy-Stein, O., and Levi, B.Z. (1998). Regulation of vascular endothelial growth factor (VEGF) expression is mediated by internal initiation of translation and alternative initiation of transcription. *Oncogene* *17*, 227-236.
- Anderson, K., Nordquist, K.A., Gao, X., Hicks, K.C., Zhai, B., Gygi, S.P., and Patel, T.B. (2011). Regulation of cellular levels of Sprouty2 protein by prolyl hydroxylase domain and von Hippel-Lindau proteins. *J Biol Chem* *286*, 42027-42036.
- Andrei, M.A., Ingelfinger, D., Heintzmann, R., Achsel, T., Rivera-Pomar, R., and Luhrmann, R. (2005). A role for eIF4E and eIF4E-transporter in targeting mRNPs to mammalian processing bodies. *Rna* *11*, 717-727.
- Aragones, J., Schneider, M., Van Geyte, K., Fraisl, P., Dresselaers, T., Mazzone, M., Dirx, R., Zacchigna, S., Lemieux, H., Jeoung, N.H., *et al.* (2008). Deficiency or inhibition of oxygen sensor Phd1 induces hypoxia tolerance by reprogramming basal metabolism. *Nat Genet* *40*, 170-180.
- Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res* *21*, 381-395.
- Bao, W., Qin, P., Needle, S., Erickson-Miller, C.L., Duffy, K.J., Ariazi, J.L., Zhao, S., Olzinski, A.R., Behm, D.J., Pipes, G.C., *et al.* (2010). Chronic inhibition of hypoxia-inducible factor prolyl 4-hydroxylase improves ventricular performance, remodeling, and vascularity after myocardial infarction in the rat. *J Cardiovasc Pharmacol* *56*, 147-155.
- Barrick, D., Ferreira, D.U., and Komives, E.A. (2008). Folding landscapes of ankyrin repeat proteins: experiments meet theory. *Curr Opin Struct Biol* *18*, 27-34.
- Basith, S., Manavalan, B., Gosu, V., and Choi, S. (2013). Evolutionary, structural and functional interplay of the I $\kappa$ B family members. *PLoS One* *8*, e54178.
- Besarab, A. (2010). FG-4592, A Novel Oral HIF Prolyl Hydroxylase Inhibitor, Elevates Haemoglobin in Anemic Stage 3/4 CKD Patients.

Bichet, S., Wenger, R.H., Camenisch, G., Rolfs, A., Ehleben, W., Porwol, T., Acker, H., Fandrey, J., Bauer, C., and Gassmann, M. (1999). Oxygen tension modulates beta-globin switching in embryoid bodies. *Faseb J* 13, 285-295.

Bishop, T., Gallagher, D., Pascual, A., Lygate, C.A., de Bono, J.P., Nicholls, L.G., Ortega-Saenz, P., Oster, H., Wijeyekoon, B., Sutherland, A.I., *et al.* (2008). Abnormal sympathoadrenal development and systemic hypotension in PHD3<sup>-/-</sup> mice. *Mol Cell Biol* 28, 3386-3400.

Bogoyevitch, M.A., Yeap, Y.Y., Qu, Z., Ngoei, K.R., Yip, Y.Y., Zhao, T.T., Heng, J.I., and Ng, D.C. (2012). WD40-repeat protein 62 is a JNK-phosphorylated spindle pole protein required for spindle maintenance and timely mitotic progression. *J Cell Sci* 125, 5096-5109.

Bradley, C.M., and Barrick, D. (2005). Effect of multiple prolyl isomerization reactions on the stability and folding kinetics of the notch ankyrin domain: experiment and theory. *J Mol Biol* 352, 253-265.

Brent, M.M., and Marmorstein, R. (2008). Ankyrin for methylated lysines. *Nat Struct Mol Biol* 15, 221-222.

Briancon, N., McNay, D.E., Maratos-Flier, E., and Flier, J.S. (2010). Combined neural inactivation of suppressor of cytokine signaling-3 and protein-tyrosine phosphatase-1B reveals additive, synergistic, and factor-specific roles in the regulation of body energy balance. *Diabetes* 59, 3074-3084.

Buttgereit, F., and Brand, M.D. (1995). A hierarchy of ATP-consuming processes in mammalian cells. *Biochem J* 312 (Pt 1), 163-167.

Cargnello, M., Tcherkezian, J., Dorn, J.F., Huttlin, E.L., Maddox, P.S., Gygi, S.P., and Roux, P.P. (2012). Phosphorylation of the eukaryotic translation initiation factor 4E-transporter (4E-T) by c-Jun N-terminal kinase promotes stress-dependent P-body assembly. *Mol Cell Biol* 32, 4572-4584.

Carreau, A., El Hafny-Rahbi, B., Matejuk, A., Grillon, C., and Kieda, C. (2011). Why is the partial oxygen pressure of human tissues a crucial parameter? Small molecules and hypoxia. *J Cell Mol Med* 15, 1239-1253.

Cervantes, C.F., Markwick, P.R., Sue, S.C., McCammon, J.A., Dyson, H.J., and Komives, E.A. (2009). Functional dynamics of the folded ankyrin repeats of I kappa B alpha revealed by nuclear magnetic resonance. *Biochemistry* 48, 8023-8031.

Chen, H., Yan, Y., Davidson, T.L., Shinkai, Y., and Costa, M. (2006). Hypoxic stress induces dimethylated histone H3 lysine 9 through histone methyltransferase G9a in mammalian cells. *Cancer Res* 66, 9009-9016.

Chen, S., Xue, Y., Wu, X., Le, C., Bhutkar, A., Bell, E.L., Zhang, F., Langer, R., and Sharp, P.A. (2014). Global microRNA depletion suppresses tumor angiogenesis. *Genes Dev* 28, 1054-1067.

Cockman, M.E., Lancaster, D.E., Stolze, I.P., Hewitson, K.S., McDonough, M.A., Coleman, M.L., Coles, C.H., Yu, X., Hay, R.T., Ley, S.C., *et al.* (2006). Posttranslational hydroxylation of ankyrin repeats in I kappa B proteins by the hypoxia-inducible factor (HIF) asparaginyl hydroxylase, factor inhibiting HIF (FIH). *Proc Natl Acad Sci U S A* 103, 14767-14772.

Cockman, M.E., Webb, J.D., Kramer, H.B., Kessler, B.M., and Ratcliffe, P.J. (2009). Proteomics-based identification of novel factor inhibiting hypoxia-inducible factor (FIH) substrates indicates widespread

- asparaginyl hydroxylation of ankyrin repeat domain-containing proteins. *Mol Cell Proteomics* **8**, 535-546.
- Colbourne, J.K., Pfrender, M.E., Gilbert, D., Thomas, W.K., Tucker, A., Oakley, T.H., Tokishita, S., Aerts, A., Arnold, G.J., Basu, M.K., *et al.* (2011). The ecoresponsive genome of *Daphnia pulex*. *Science* **331**, 555-561.
- Coleman, M.L., McDonough, M.A., Hewitson, K.S., Coles, C., Mecinovic, J., Edelman, M., Cook, K.M., Cockman, M.E., Lancaster, D.E., Kessler, B.M., *et al.* (2007). Asparaginyl hydroxylation of the Notch ankyrin repeat domain by factor inhibiting hypoxia-inducible factor. *J Biol Chem* **282**, 24027-24038.
- Collins, R.E., Northrop, J.P., Horton, J.R., Lee, D.Y., Zhang, X., Stallcup, M.R., and Cheng, X. (2008). The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules. *Nat Struct Mol Biol* **15**, 245-250.
- Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190.
- Croy, C.H., Bergqvist, S., Huxford, T., Ghosh, G., and Komives, E.A. (2004). Biophysical characterization of the free IkappaBalpha ankyrin repeat domain in solution. *Protein Sci* **13**, 1767-1777.
- Culver, C., Sundqvist, A., Mudie, S., Melvin, A., Xirodimas, D., and Rocha, S. (2010). Mechanism of hypoxia-induced NF-kappaB. *Mol Cell Biol* **30**, 4901-4921.
- Cummins, E.P., Berra, E., Comerford, K.M., Ginouves, A., Fitzgerald, K.T., Seeballuck, F., Godson, C., Nielsen, J.E., Moynagh, P., Pouyssegur, J., *et al.* (2006). Prolyl hydroxylase-1 negatively regulates IkappaB kinase-beta, giving insight into hypoxia-induced NFkappaB activity. *Proc Natl Acad Sci U S A* **103**, 18154-18159.
- Cummins, E.P., and Taylor, C.T. (2005). Hypoxia-responsive transcription factors. *Pflugers Arch* **450**, 363-371.
- Daigneault, J., Klemetsaune, L., and Wasserman, S.A. (2013). The IRAK homolog Pelle is the functional counterpart of IkappaB kinase in the *Drosophila* Toll pathway. *PLoS One* **8**, e75150.
- Dames, S.A., Martinez-Yamout, M., De Guzman, R.N., Dyson, H.J., and Wright, P.E. (2002). Structural basis for Hif-1 alpha /CBP recognition in the cellular hypoxic response. *Proc Natl Acad Sci U S A* **99**, 5271-5276.
- de Castro, E., Sigrist, C.J., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P.S., Gasteiger, E., Bairoch, A., and Hulo, N. (2006). ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res* **34**, W362-365.
- Denko, N., Wernke-Dollries, K., Johnson, A.B., Hammond, E., Chiang, C.M., and Barton, M.C. (2003). Hypoxia actively represses transcription by inducing negative cofactor 2 (Dr1/DrAP1) and blocking preinitiation complex assembly. *J Biol Chem* **278**, 5744-5749.
- Devries, I.L., Hampton-Smith, R.J., Mulvihill, M.M., Alverdi, V., Peet, D.J., and Komives, E.A. (2010). Consequences of IkappaB alpha hydroxylation by the factor inhibiting HIF (FIH). *FEBS Lett* **584**, 4725-4730.

- Dostie, J., Ferraiuolo, M., Pause, A., Adam, S.A., and Sonenberg, N. (2000). A novel shuttling protein, 4E-T, mediates the nuclear import of the mRNA 5' cap-binding protein, eIF4E. *Embo J* 19, 3142-3156.
- Elkins, J.M., Hewitson, K.S., McNeill, L.A., Seibel, J.F., Schlemminger, I., Pugh, C.W., Ratcliffe, P.J., and Schofield, C.J. (2003). Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1 alpha. *J Biol Chem* 278, 1802-1806.
- Eltzschig, H.K., and Collard, C.D. (2004). Vascular ischaemia and reperfusion injury. *Br Med Bull* 70, 71-86.
- Elvidge, G.P., Glenny, L., Appelhoff, R.J., Ratcliffe, P.J., Ragoussis, J., and Gleadle, J.M. (2006). Concordant regulation of gene expression by hypoxia and 2-oxoglutarate-dependent dioxygenase inhibition: the role of HIF-1alpha, HIF-2alpha, and other pathways. *J Biol Chem* 281, 15215-15226.
- Epstein, A.C., Gleadle, J.M., McNeill, L.A., Hewitson, K.S., O'Rourke, J., Mole, D.R., Mukherji, M., Metzen, E., Wilson, M.I., Dhanda, A., *et al.* (2001). *C. elegans* EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. *Cell* 107, 43-54.
- Fahling, M. (2009). Surviving hypoxia by modulation of mRNA translation rate. *J Cell Mol Med* 13, 2770-2779.
- Ferguson, J.E., 3rd, Wu, Y., Smith, K., Charles, P., Powers, K., Wang, H., and Patterson, C. (2007). ASB4 is a hydroxylation substrate of FIH and promotes vascular differentiation via an oxygen-dependent mechanism. *Mol Cell Biol* 27, 6407-6419.
- Ferraiuolo, M.A., Basak, S., Dostie, J., Murray, E.L., Schoenberg, D.R., and Sonenberg, N. (2005). A role for the eIF4E-binding protein 4E-T in P-body formation and mRNA decay. *J Cell Biol* 170, 913-924.
- Ferreiro, D.U., Cervantes, C.F., Truhlar, S.M., Cho, S.S., Wolynes, P.G., and Komives, E.A. (2007). Stabilizing IkappaBalpha by "consensus" design. *J Mol Biol* 365, 1201-1216.
- Ferreiro, D.U., Cho, S.S., Komives, E.A., and Wolynes, P.G. (2005). The energy landscape of modular repeat proteins: topology determines folding mechanism in the ankyrin family. *J Mol Biol* 354, 679-692.
- FibroGen (2013). Cardioprotective HIF-PHI.
- Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 39, W29-37.
- Freedman, S.J., Sun, Z.Y., Poy, F., Kung, A.L., Livingston, D.M., Wagner, G., and Eck, M.J. (2002). Structural basis for recruitment of CBP/p300 by hypoxia-inducible factor-1 alpha. *Proc Natl Acad Sci U S A* 99, 5367-5372.
- Fukuda, R., Zhang, H., Kim, J.W., Shimoda, L., Dang, C.V., and Semenza, G.L. (2007). HIF-1 regulates cytochrome oxidase subunits to optimize efficiency of respiration in hypoxic cells. *Cell* 129, 111-122.
- Galic, S., Hauser, C., Kahn, B.B., Haj, F.G., Neel, B.G., Tonks, N.K., and Tiganis, T. (2005). Coordinated regulation of insulin signaling by the protein tyrosine phosphatases PTP1B and TCPTP. *Mol Cell Biol* 25, 819-829.

GibcoBRL. ProQuest Two-Hybrid System Instruction Manual, L. Technologies, ed.

Gilmore, T.D., and Wolenski, F.S. (2012). NF-kappaB: where did it come from and why? *Immunol Rev* 246, 14-35.

Gordeuk, V.R., and Prchal, J.T. (2006). Vascular complications in Chuvash polycythemia. *Semin Thromb Hemost* 32, 289-294.

Gupta, D., Krueger, C.B., and Lastra, G. (2012). Over-nutrition, obesity and insulin resistance in the development of beta-cell dysfunction. *Curr Diabetes Rev* 8, 76-83.

Halberg, N., Khan, T., Trujillo, M.E., Wernstedt-Asterholm, I., Attie, A.D., Sherwani, S., Wang, Z.V., Landskroner-Eiger, S., Dineen, S., Magalang, U.J., *et al.* (2009). Hypoxia-inducible factor 1alpha induces fibrosis and insulin resistance in white adipose tissue. *Mol Cell Biol* 29, 4467-4483.

Hampton-Smith, R.J. (2004). Searching for FIH-1 substrates and Interacting Proteins. In School of Molecular and Biomedical Science (Adelaide, University of Adelaide).

Hand, S.C., and Menze, M.A. (2008). Mitochondria in energy-limited states: mechanisms that blunt the signaling of cell death. *J Exp Biol* 211, 1829-1840.

Hardison, R.C. (2003). Comparative genomics. *PLoS Biol* 1, E58.

Hardy, A.P., Prokes, I., Kelly, L., Campbell, I.D., and Schofield, C.J. (2009). Asparaginyl beta-hydroxylation of proteins containing ankyrin repeat domains influences their stability and function. *J Mol Biol* 392, 994-1006.

Harno, E., Cottrell, E.C., and White, A. (2013). Metabolic pitfalls of CNS Cre-based technology. *Cell Metab* 18, 21-28.

Heald, M., and Cawthorne, M.A. (2011). Dual acting and pan-PPAR activators as potential anti-diabetic therapies. *Handb Exp Pharmacol*, 35-51.

Hedges, S.B., Dudley, J., and Kumar, S. (2006). TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* 22, 2971-2972.

Hewitson, K.S., McNeill, L.A., Riordan, M.V., Tian, Y.M., Bullock, A.N., Welford, R.W., Elkins, J.M., Oldham, N.J., Bhattacharya, S., Gleadle, J.M., *et al.* (2002). Hypoxia-inducible factor (HIF) asparagine hydroxylase is identical to factor inhibiting HIF (FIH) and is related to the cupin structural family. *J Biol Chem* 277, 26351-26355.

Hirsila, M., Koivunen, P., Gunzler, V., Kivirikko, K.I., and Myllyharju, J. (2003). Characterization of the human prolyl 4-hydroxylases that modify the hypoxia-inducible factor. *J Biol Chem* 278, 30772-30780.

Holscher, M., Silter, M., Krull, S., von Ahlen, M., Hesse, A., Schwartz, P., Wielockx, B., Breier, G., Katschinski, D.M., and Ziesenis, A. (2011). Cardiomyocyte-specific prolyl-4-hydroxylase domain 2 knock out protects from acute myocardial ischemic injury. *J Biol Chem* 286, 11185-11194.

- Horman, S., Browne, G., Krause, U., Patel, J., Vertommen, D., Bertrand, L., Lavoigne, A., Hue, L., Proud, C., and Rider, M. (2002). Activation of AMP-activated protein kinase leads to the phosphorylation of elongation factor 2 and an inhibition of protein synthesis. *Curr Biol* *12*, 1419-1423.
- Huang, L.E., Arany, Z., Livingston, D.M., and Bunn, H.F. (1996). Activation of hypoxia-inducible transcription factor depends primarily upon redox-sensitive stabilization of its alpha subunit. *J Biol Chem* *271*, 32253-32259.
- Huang, L.E., Gu, J., Schau, M., and Bunn, H.F. (1998). Regulation of hypoxia-inducible factor 1alpha is mediated by an O<sub>2</sub>-dependent degradation domain via the ubiquitin-proteasome pathway. *Proc Natl Acad Sci U S A* *95*, 7987-7992.
- Huxford, T., Huang, D.B., Malek, S., and Ghosh, G. (1998). The crystal structure of the I $\kappa$ B $\alpha$ /NF- $\kappa$ B complex reveals mechanisms of NF- $\kappa$ B inactivation. *Cell* *95*, 759-770.
- Imrie, H., Abbas, A., and Kearney, M. (2010). Insulin resistance, lipotoxicity and endothelial dysfunction. *Biochim Biophys Acta* *1801*, 320-326.
- Iversen, L.F., Moller, K.B., Pedersen, A.K., Peters, G.H., Petersen, A.S., Andersen, H.S., Branner, S., Mortensen, S.B., and Moller, N.P. (2002). Structure determination of T cell protein-tyrosine phosphatase. *J Biol Chem* *277*, 19982-19990.
- Jackson, R.J., Hellen, C.U., and Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* *11*, 113-127.
- Janke, K., Brockmeier, U., Kuhlmann, K., Eisenacher, M., Nolde, J., Meyer, H.E., Mairbaurl, H., and Metzen, E. (2013). Factor inhibiting HIF-1 (FIH-1) modulates protein interactions of apoptosis-stimulating p53 binding protein 2 (ASPP2). *J Cell Sci* *126*, 2629-2640.
- Jiang, B.H., Zheng, J.Z., Leung, S.W., Roe, R., and Semenza, G.L. (1997). Transactivation and inhibitory domains of hypoxia-inducible factor 1alpha. Modulation of transcriptional activity by oxygen tension. *J Biol Chem* *272*, 19253-19260.
- Jiang, C., Kim, J.H., Li, F., Qu, A., Gavrilova, O., Shah, Y.M., and Gonzalez, F.J. (2013). Hypoxia-inducible factor 1alpha regulates a SOCS3-STAT3-adiponectin signal transduction pathway in adipocytes. *J Biol Chem* *288*, 3844-3857.
- Jiang, C., Qu, A., Matsubara, T., Chanturiya, T., Jou, W., Gavrilova, O., Shah, Y.M., and Gonzalez, F.J. (2011). Disruption of hypoxia-inducible factor 1 in adipocytes improves insulin sensitivity and decreases adiposity in high-fat diet-fed mice. *Diabetes* *60*, 2484-2495.
- Johnson, A.B., Denko, N., and Barton, M.C. (2008). Hypoxia induces a novel signature of chromatin modifications and global repression of transcription. *Mutat Res* *640*, 174-179.
- Karttunen, S., Duffield, M., Scrimgeour, N.R., Squires, L., Lim, W.L., Dallas, M.L., Scragg, J.L., Chicher, J., Dave, K.A., Whitelaw, M.L., *et al.* (2014). Oxygen-dependent hydroxylation by Factor Inhibiting HIF (FIH) regulates the TRPV3 ion channel. *J Cell Sci*.
- Kavanagh, D.P., and Kalia, N. (2011). Hematopoietic stem cell homing to injured tissues. *Stem Cell Rev* *7*, 672-682.

- Kedersha, N., and Anderson, P. (2007). Mammalian stress granules and processing bodies. *Methods Enzymol* 431, 61-81.
- Kelly, L., McDonough, M.A., Coleman, M.L., Ratcliffe, P.J., and Schofield, C.J. (2009). Asparagine beta-hydroxylation stabilizes the ankyrin repeat domain fold. *Mol Biosyst* 5, 52-58.
- Khan, M.N., Bhattacharyya, T., Andrikopoulos, P., Esteban, M.A., Barod, R., Connor, T., Ashcroft, M., Maxwell, P.H., and Kiriakidis, S. (2011). Factor inhibiting HIF (FIH-1) promotes renal cancer cell survival by protecting cells from HIF-1alpha-mediated apoptosis. *Br J Cancer* 104, 1151-1159.
- Kim, J.W., Tchernyshyov, I., Semenza, G.L., and Dang, C.V. (2006a). HIF-1-mediated expression of pyruvate dehydrogenase kinase: a metabolic switch required for cellular adaptation to hypoxia. *Cell Metab* 3, 177-185.
- Kim, W.Y., Safran, M., Buckley, M.R., Ebert, B.L., Glickman, J., Bosenberg, M., Regan, M., and Kaelin, W.G., Jr. (2006b). Failure to prolyl hydroxylate hypoxia-inducible factor alpha phenocopies VHL inactivation in vivo. *Embo J* 25, 4650-4662.
- Koivunen, P., Hirsila, M., Gunzler, V., Kivirikko, K.I., and Myllyharju, J. (2004). Catalytic properties of the asparaginyl hydroxylase (FIH) in the oxygen sensing pathway are distinct from those of its prolyl 4-hydroxylases. *J Biol Chem* 279, 9899-9904.
- Koritzinsky, M., Magagnin, M.G., van den Beucken, T., Seigneuric, R., Savelkoul, K., Dostie, J., Pyronnet, S., Kaufman, R.J., Wepler, S.A., Voncken, J.W., *et al.* (2006). Gene expression during acute and prolonged hypoxia is regulated by distinct mechanisms of translational control. *Embo J* 25, 1114-1125.
- Koster, H., Little, D.P., Luan, P., Muller, R., Siddiqi, S.M., Marappan, S., and Yip, P. (2007). Capture compound mass spectrometry: a technology for the investigation of small molecule protein interactions. *Assay Drug Dev Technol* 5, 381-390.
- Krentz, A.J., and Bailey, C.J. (2005). Oral antidiabetic agents: current role in type 2 diabetes mellitus. *Drugs* 65, 385-411.
- Kuznetsova, A.V., Meller, J., Schnell, P.O., Nash, J.A., Ignacak, M.L., Sanchez, Y., Conaway, J.W., Conaway, R.C., and Czyzyk-Krzeska, M.F. (2003). von Hippel-Lindau protein binds hyperphosphorylated large subunit of RNA polymerase II through a proline hydroxylation motif and targets it for ubiquitination. *Proc Natl Acad Sci U S A* 100, 2706-2711.
- Lai, Y., Qiao, M., Song, M., Weintraub, S.T., and Shiio, Y. (2011). Quantitative proteomics identifies the Myb-binding protein p160 as a novel target of the von Hippel-Lindau tumor suppressor. *PLoS One* 6, e16975.
- Lando, D., Peet, D.J., Gorman, J.J., Whelan, D.A., Whitelaw, M.L., and Bruick, R.K. (2002a). FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia-inducible factor. *Genes Dev* 16, 1466-1471.
- Lando, D., Peet, D.J., Whelan, D.A., Gorman, J.J., and Whitelaw, M.L. (2002b). Asparagine hydroxylation of the HIF transactivation domain a hypoxic switch. *Science* 295, 858-861.

- Lang, K.J., Kappel, A., and Goodall, G.J. (2002). Hypoxia-inducible factor-1 $\alpha$  mRNA contains an internal ribosome entry site that allows efficient translation during normoxia and hypoxia. *Mol Biol Cell* *13*, 1792-1801.
- Lee, J.S., Kim, Y., Bhin, J., Shin, H.J., Nam, H.J., Lee, S.H., Yoon, J.B., Binda, O., Gozani, O., Hwang, D., *et al.* (2011). Hypoxia-induced methylation of a pontin chromatin remodeling factor. *Proc Natl Acad Sci U S A* *108*, 13510-13515.
- Lee, J.S., Kim, Y., Kim, I.S., Kim, B., Choi, H.J., Lee, J.M., Shin, H.J., Kim, J.H., Kim, J.Y., Seo, S.B., *et al.* (2010). Negative regulation of hypoxic responses via induced Reptin methylation. *Mol Cell* *39*, 71-85.
- Lee, S.H., Kim, J., Kim, W.H., and Lee, Y.M. (2009). Hypoxic silencing of tumor suppressor RUNX3 by histone modification in gastric cancer cells. *Oncogene* *28*, 184-194.
- Levin, T.C., Greaney, A.J., Wetzel, L., and King, N. (2014). The Rosetteless gene controls development in the choanoflagellate *S. rosetta*. *Elife* *3*.
- Li, Q., and Verma, I.M. (2002). NF-kappaB regulation in the immune system. *Nat Rev Immunol* *2*, 725-734.
- Linke, S., Hampton-Smith, R.J., and Peet, D.J. (2007). Characterization of ankyrin repeat-containing proteins as substrates of the asparaginyl hydroxylase factor inhibiting hypoxia-inducible transcription factor. *Methods Enzymol* *435*, 61-85.
- Linke, S., Stojkoski, C., Kewley, R.J., Booker, G.W., Whitelaw, M.L., and Peet, D.J. (2004). Substrate requirements of the oxygen-sensing asparaginyl hydroxylase factor-inhibiting hypoxia-inducible factor. *J Biol Chem* *279*, 14391-14397.
- Lipton, P. (1999). Ischemic cell death in brain neurons. *Physiol Rev* *79*, 1431-1568.
- Lisy, K. (2011). Investigating potential post-translational modification of factor-inhibiting HIF (FIH-1). In *Biochemistry* (Adelaide University).
- Liu, L., Cash, T.P., Jones, R.G., Keith, B., Thompson, C.B., and Simon, M.C. (2006). Hypoxia-induced energy stress regulates mRNA translation and cell growth. *Mol Cell* *21*, 521-531.
- Lodowski, D.T., Pitcher, J.A., Capel, W.D., Lefkowitz, R.J., and Tesmer, J.J. (2003). Keeping G proteins at bay: a complex between G protein-coupled receptor kinase 2 and Gbetagamma. *Science* *300*, 1256-1262.
- Loenarz, C., Coleman, M.L., Boleininger, A., Schierwater, B., Holland, P.W., Ratcliffe, P.J., and Schofield, C.J. (2011). The hypoxia-inducible transcription factor pathway regulates oxygen sensing in the simplest animal, *Trichoplax adhaerens*. *EMBO Rep* *12*, 63-70.
- Low, C., Homeyer, N., Weininger, U., Sticht, H., and Balbach, J. (2009). Conformational switch upon phosphorylation: human CDK inhibitor p19INK4d between the native and partially folded state. *ACS Chem Biol* *4*, 53-63.
- Low, T.Y., Peng, M., Magliozzi, R., Mohammed, S., Guardavaccaro, D., and Heck, A.J. (2014). A systems-wide screen identifies substrates of the SCFbetaTrCP ubiquitin ligase. *Sci Signal* *7*, rs8.



- Mali, P., Esvelt, K.M., and Church, G.M. (2013). Cas9 as a versatile tool for engineering biology. *Nat Methods* *10*, 957-963.
- Mayor, F., Jr., Lucas, E., Jurado-Pueyo, M., Garcia-Guerra, L., Nieto-Vazquez, I., Vila-Bedmar, R., Fernandez-Veledo, S., and Murga, C. (2011). G Protein-coupled receptor kinase 2 (GRK2): A novel modulator of insulin resistance. *Arch Physiol Biochem* *117*, 125-130.
- McConnell, S.C., Huo, Y., Liu, S., and Ryan, T.M. (2011). Human globin knock-in mice complete fetal-to-adult hemoglobin switching in postnatal development. *Mol Cell Biol* *31*, 876-883.
- McNeill, L.A., Hewitson, K.S., Claridge, T.D., Seibel, J.F., Horsfall, L.E., and Schofield, C.J. (2002a). Hypoxia-inducible factor asparaginyl hydroxylase (FIH-1) catalyses hydroxylation at the beta-carbon of asparagine-803. *Biochem J* *367*, 571-575.
- McNeill, L.A., Hewitson, K.S., Gleadle, J.M., Horsfall, L.E., Oldham, N.J., Maxwell, P.H., Pugh, C.W., Ratcliffe, P.J., and Schofield, C.J. (2002b). The use of dioxygen by HIF prolyl hydroxylase (PHD1). *Bioorg Med Chem Lett* *12*, 1547-1550.
- Metzen, E., Berchner-Pfannschmidt, U., Stengel, P., Marxsen, J.H., Stolze, I., Klinger, M., Huang, W.Q., Wotzlaw, C., Hellwig-Burgel, T., Jelkmann, W., *et al.* (2003). Intracellular localisation of human HIF-1 alpha hydroxylases: implications for oxygen sensing. *J Cell Sci* *116*, 1319-1326.
- Mik, E.G. (2013). *Measuring Mitochondrial Oxygen Tension: From Basic Principles to Application in Humans*. *Anesth Analg*.
- Mimura, I., Tanaka, T., Wada, Y., Kodama, T., and Nangaku, M. (2011). Pathophysiological response to hypoxia - from the molecular mechanisms of malady to drug discovery: epigenetic regulation of the hypoxic response via hypoxia-inducible factor and histone modifying enzymes. *J Pharmacol Sci* *115*, 453-458.
- Mole, D.R., Blancher, C., Copley, R.R., Pollard, P.J., Gleadle, J.M., Ragoussis, J., and Ratcliffe, P.J. (2009). Genome-wide association of hypoxia-inducible factor (HIF)-1alpha and HIF-2alpha DNA binding with expression profiling of hypoxia-inducible transcripts. *J Biol Chem* *284*, 16767-16775.
- Mosavi, L.K., Cammett, T.J., Desrosiers, D.C., and Peng, Z.Y. (2004). The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci* *13*, 1435-1448.
- Nakamura, Y., Nakano, K., Umehara, T., Kimura, M., Hayashizaki, Y., Tanaka, A., Horikoshi, M., Padmanabhan, B., and Yokoyama, S. (2007). Structure of the oncoprotein gankyrin in complex with S6 ATPase of the 26S proteasome. *Structure* *15*, 179-189.
- Nam, Y., Sliz, P., Song, L., Aster, J.C., and Blacklow, S.C. (2006). Structural basis for cooperativity in recruitment of MAML coactivators to Notch transcription complexes. *Cell* *124*, 973-983.
- Narayan, A.D., Ersek, A., Campbell, T.A., Colon, D.M., Pixley, J.S., and Zanjani, E.D. (2005). The effect of hypoxia and stem cell source on haemoglobin switching. *Br J Haematol* *128*, 562-570.
- Nicholas, A.K., Khurshid, M., Desir, J., Carvalho, O.P., Cox, J.J., Thornton, G., Kausar, R., Ansar, M., Ahmad, W., Verloes, A., *et al.* (2010). WDR62 is associated with the spindle pole and is mutated in human microcephaly. *Nat Genet* *42*, 1010-1014.

O'Kelly, I., Lewis, A., Peers, C., and Kemp, P.J. (2000). O<sub>2</sub> sensing by airway chemoreceptor-derived cells. Protein kinase c activation reveals functional evidence for involvement of NADPH oxidase. *J Biol Chem* 275, 7684-7692.

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, 2060-2071.

Pelletier, J., Dayan, F., Durivault, J., Ilc, K., Pecou, E., Pouyssegur, J., and Mazure, N.M. (2012). The asparaginyl hydroxylase factor-inhibiting HIF is essential for tumor growth through suppression of the p53-p21 axis. *Oncogene* 31, 2989-3001.

Peng, H., Kaplan, N., Yang, W., Getsios, S., and Lavker, R.M. (2014). FIH-1 Disrupts an LRRK1/EGFR Complex to Positively Regulate Keratinocyte Migration. *Am J Pathol* 184, 3262-3271.

Perkins, N.D. (2007). Integrating cell-signalling pathways with NF-kappaB and IKK function. *Nat Rev Mol Cell Biol* 8, 49-62.

Pocock, R. (2011). Invited review: decoding the microRNA response to hypoxia. *Pflugers Arch* 461, 307-315.

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, 4079-4088.

Porter, A.C., Fanger, G.R., and Vaillancourt, R.R. (1999). Signal transduction pathways regulated by arsenate and arsenite. *Oncogene* 18, 7794-7802.

Prabhakar, N.R., and Semenza, G.L. (2012). Adaptive and maladaptive cardiorespiratory responses to continuous and intermittent hypoxia mediated by hypoxia-inducible factors 1 and 2. *Physiol Rev* 92, 967-1003.

Pugh, C.W., O'Rourke, J.F., Nagao, M., Gleadle, J.M., and Ratcliffe, P.J. (1997). Activation of hypoxia-inducible factor-1; definition of regulatory domains within the alpha subunit. *J Biol Chem* 272, 11205-11214.

Quintero, P., Gonzalez-Muniesa, P., Garcia-Diaz, D.F., and Martinez, J.A. (2012). Effects of hyperoxia exposure on metabolic markers and gene expression in 3T3-L1 adipocytes. *J Physiol Biochem* 68, 663-669.

Rathert, P., Dhayalan, A., Murakami, M., Zhang, X., Tamas, R., Jurkowska, R., Komatsu, Y., Shinkai, Y., Cheng, X., and Jeltsch, A. (2008). Protein lysine methyltransferase G9a acts on non-histone targets. *Nat Chem Biol* 4, 344-346.

Rich, P. (2003). Chemiosmotic coupling: The cost of living. *Nature* 421, 583.

Richards, G.S., and Degnan, B.M. (2009). The dawn of developmental signaling in the metazoa. *Cold Spring Harb Symp Quant Biol* 74, 81-90.

Saban, E., Chen, Y.H., Hangasky, J.A., Taabazuing, C.Y., Holmes, B.E., and Knapp, M.J. (2011). The second coordination sphere of FIH controls hydroxylation. *Biochemistry* 50, 4733-4740.

- Sang, N., Fang, J., Srinivas, V., Leshchinsky, I., and Caro, J. (2002). Carboxyl-terminal transactivation activity of hypoxia-inducible factor 1 alpha is governed by a von Hippel-Lindau protein-independent, hydroxylation-regulated association with p300/CBP. *Mol Cell Biol* 22, 2984-2992.
- Schmierer, B., Novak, B., and Schofield, C.J. (2010). Hypoxia-dependent sequestration of an oxygen sensor by a widespread structural motif can shape the hypoxic response--a predictive kinetic model. *BMC Syst Biol* 4, 139.
- Scholz, C.C., and Taylor, C.T. (2013). Hydroxylase-dependent regulation of the NF-kappaB pathway. *Biol Chem* 394, 479-493.
- 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, 5857-5864.
- Seeley, R., Stephens, T., Tate, P. (2006). *Anatomy and Physiology*, 7th Edition edn (McGraw Hill).
- Semenza, G.L. (2003). Targeting HIF-1 for cancer therapy. *Nat Rev Cancer* 3, 721-732.
- Semenza, G.L. (2010). Defining the role of hypoxia-inducible factor 1 in cancer biology and therapeutics. *Oncogene* 29, 625-634.
- Sherratt, H.S. (1991). Mitochondria: structure and function. *Rev Neurol (Paris)* 147, 417-430.
- Shi, D.J., Ye, S., Cao, X., Zhang, R., and Wang, K. (2013). Crystal structure of the N-terminal ankyrin repeat domain of TRPV3 reveals unique conformation of finger 3 loop critical for channel function. *Protein Cell* 4, 942-950.
- Shimoda, L.A., and Polak, J. (2011). Hypoxia. 4. Hypoxia and ion channel function. *Am J Physiol Cell Physiol* 300, C951-967.
- Shin, M.K., Drager, L.F., Yao, Q., Bevans-Fonti, S., Yoo, D.Y., Jun, J.C., Aja, S., Bhanot, S., and Polotsky, V.Y. (2012). Metabolic consequences of high-fat diet are attenuated by suppression of HIF-1alpha. *PLoS One* 7, e46562.
- Shinkai, Y., and Tachibana, M. (2011). H3K9 methyltransferase G9a and the related molecule GLP. *Genes Dev* 25, 781-788.
- Singleton, R.S., Trudgian, D.C., Fischer, R., Kessler, B.M., Ratcliffe, P.J., and Cockman, M.E. (2011). Quantitative mass spectrometry reveals dynamics of factor-inhibiting hypoxia-inducible factor-catalyzed hydroxylation. *J Biol Chem* 286, 33784-33794.
- Smith, T.G., Brooks, J.T., Balanos, G.M., Lappin, T.R., Layton, D.M., Leedham, D.L., Liu, C., Maxwell, P.H., McMullin, M.F., McNamara, C.J., *et al.* (2006). Mutation of von Hippel-Lindau tumour suppressor and human cardiopulmonary physiology. *PLoS Med* 3, e290.
- Springer, E. (2009). *O2 Measurement Guide*.
- Stefansson, B., Ohama, T., Daugherty, A.E., and Brautigan, D.L. (2008). Protein phosphatase 6 regulatory subunits composed of ankyrin repeat domains. *Biochemistry* 47, 1442-1451.

Stolze, I.P., Tian, Y.M., Appelhoff, R.J., Turley, H., Wykoff, C.C., Gleadle, J.M., and Ratcliffe, P.J. (2004). Genetic analysis of the role of the asparaginyl hydroxylase factor inhibiting hypoxia-inducible factor (HIF) in regulating HIF transcriptional target genes. *J Biol Chem* 279, 42719-42725.

Sylvester, J.T., Shimoda, L.A., Aaronson, P.I., and Ward, J.P. (2012). Hypoxic pulmonary vasoconstriction. *Physiol Rev* 92, 367-520.

Takahashi, N., Kuwaki, T., Kiyonaka, S., Numata, T., Kozai, D., Mizuno, Y., Yamamoto, S., Naito, S., Knevels, E., Carmeliet, P., *et al.* (2011). TRPA1 underlies a sensing mechanism for O<sub>2</sub>. *Nat Chem Biol* 7, 701-711.

Takeda, K., Aguila, H.L., Parikh, N.S., Li, X., Lamothe, K., Duan, L.J., Takeda, H., Lee, F.S., and Fong, G.H. (2008). Regulation of adult erythropoiesis by prolyl hydroxylase domain proteins. *Blood* 111, 3229-3235.

Tan, I., Ng, C.H., Lim, L., and Leung, T. (2001). Phosphorylation of a novel myosin binding subunit of protein phosphatase 1 reveals a conserved mechanism in the regulation of actin cytoskeleton. *J Biol Chem* 276, 21209-21216.

Taylor, C.T., and Cummins, E.P. (2009). The role of NF-kappaB in hypoxia-induced gene expression. *Ann N Y Acad Sci* 1177, 178-184.

The\_Jackson\_Laboratory (2014). B6.Cg-Tg(Nes-cre)1Kln/J.

Tian, Y.M., Yeoh, K.K., Lee, M.K., Eriksson, T., Kessler, B.M., Kramer, H.B., Edelmann, M.J., Willam, C., Pugh, C.W., Schofield, C.J., *et al.* (2011). Differential sensitivity of hypoxia inducible factor hydroxylation sites to hypoxia and hydroxylase inhibitors. *J Biol Chem* 286, 13041-13051.

Tronche, F., Kellendonk, C., Kretz, O., Gass, P., Anlag, K., Orban, P.C., Bock, R., Klein, R., and Schutz, G. (1999). Disruption of the glucocorticoid receptor gene in the nervous system results in reduced anxiety. *Nat Genet* 23, 99-103.

Truhlar, S.M., Mathes, E., Cervantes, C.F., Ghosh, G., and Komives, E.A. (2008). Pre-folding I $\kappa$ B alters control of NF-kappaB signaling. *J Mol Biol* 380, 67-82.

Turer, A.T., and Scherer, P.E. (2012). Adiponectin: mechanistic insights and clinical implications. *Diabetologia* 55, 2319-2326.

Uniacke, J., Holterman, C.E., Lachance, G., Franovic, A., Jacob, M.D., Fabian, M.R., Payette, J., Holcik, M., Pause, A., and Lee, S. (2012). An oxygen-regulated switch in the protein synthesis machinery. *Nature* 486, 126-129.

Walmsley, S.R., Chilvers, E.R., Thompson, A.A., Vaughan, K., Marriott, H.M., Parker, L.C., Shaw, G., Parmar, S., Schneider, M., Sabroe, I., *et al.* (2011). Prolyl hydroxylase 3 (PHD3) is essential for hypoxic regulation of neutrophilic inflammation in humans and mice. *J Clin Invest* 121, 1053-1063.

Wang, D., Youngson, C., Wong, V., Yeger, H., Dinauer, M.C., Vega-Saenz Miera, E., Rudy, B., and Cutz, E. (1996). NADPH-oxidase and a hydrogen peroxide-sensitive K<sup>+</sup> channel may function as an oxygen sensor complex in airway chemoreceptors and small cell lung carcinoma cell lines. *Proc Natl Acad Sci U S A* 93, 13182-13187.

- Wang, Z., Li, C., Ellenburg, M., Soistman, E., Ruble, J., Wright, B., Ho, J.X., and Carter, D.C. (2006). Structure of human ferritin L chain. *Acta Crystallogr D Biol Crystallogr* 62, 800-806.
- Wang, Z., Yang, D., Zhang, X., Li, T., Li, J., Tang, Y., and Le, W. (2011). Hypoxia-induced down-regulation of neprilysin by histone modification in mouse primary cortical and hippocampal neurons. *PLoS One* 6, e19229.
- Wasserman, T., Katsenelson, K., Daniliuc, S., Hasin, T., Choder, M., and Aronheim, A. (2010). A novel c-Jun N-terminal kinase (JNK)-binding protein WDR62 is recruited to stress granules and mediates a nonclassical JNK activation. *Mol Biol Cell* 21, 117-130.
- Webb, J.D., Muranyi, A., Pugh, C.W., Ratcliffe, P.J., and Coleman, M.L. (2009). MYPT1, the targeting subunit of smooth muscle myosin phosphatase, is a substrate for the asparaginyl hydroxylase factor inhibiting hypoxia inducible factor (FIH). *Biochem J*.
- Wenger, R.H., Stiehl, D.P., and Camenisch, G. (2005). Integration of oxygen signaling at the consensus HRE. *Sci STKE* 2005, re12.
- West, J.B. (2008). *Respiratory Physiology: The essentials*, 8th edn (Lippincott Williams & Wilkins).
- Wieser, W., and Krumschnabel, G. (2001). Hierarchies of ATP-consuming processes: direct compared with indirect measurements, and comparative aspects. *Biochem J* 355, 389-395.
- Wilkins, S.E., Hyvarinen, J., Chicher, J., Gorman, J.J., Peet, D.J., Bilton, R.L., and Koivunen, P. (2009). Differences in hydroxylation and binding of Notch and HIF-1alpha demonstrate substrate selectivity for factor inhibiting HIF-1 (FIH-1). *Int J Biochem Cell Biol* 41, 1563-1571.
- Wilkins, S.E., Karttunen, S., Hampton-Smith, R.J., Murchland, I., Chapman-Smith, A., and Peet, D.J. (2012). Factor inhibiting HIF (FIH) recognizes distinct molecular features within hypoxia-inducible factor-alpha (HIF-alpha) versus ankyrin repeat substrates. *J Biol Chem* 287, 8769-8781.
- Williams, S.E., Wootton, P., Mason, H.S., Bould, J., Iles, D.E., Riccardi, D., Peers, C., and Kemp, P.J. (2004). Hemoxygenase-2 is an oxygen sensor for a calcium-sensitive potassium channel. *Science* 306, 2093-2097.
- Xie, L., Xiao, K., Whalen, E.J., Forrester, M.T., Freeman, R.S., Fong, G., Gygi, S.P., Lefkowitz, R.J., and Stamler, J.S. (2009). Oxygen-regulated beta(2)-adrenergic receptor hydroxylation by EGLN3 and ubiquitylation by pVHL. *Sci Signal* 2, ra33.
- Yang, M., Chowdhury, R., Ge, W., Hamed, R.B., McDonough, M.A., Claridge, T.D., Kessler, B.M., Cockman, M.E., Ratcliffe, P.J., and Schofield, C.J. (2011a). Factor-inhibiting hypoxia-inducible factor (FIH) catalyses the post-translational hydroxylation of histidyl residues within ankyrin repeat domains. *Febs J* 278, 1086-1097.
- Yang, M., Ge, W., Chowdhury, R., Claridge, T.D., Kramer, H.B., Schmierer, B., McDonough, M.A., Gong, L., Kessler, B.M., Ratcliffe, P.J., *et al.* (2011b). Asparagine and aspartate hydroxylation of the cytoskeletal ankyrin family is catalyzed by factor-inhibiting hypoxia-inducible factor. *J Biol Chem* 286, 7648-7660.

Yang, M., Hardy, A.P., Chowdhury, R., Loik, N.D., Scotti, J.S., McCullagh, J.S., Claridge, T.D., McDonough, M.A., Ge, W., and Schofield, C.J. (2013). Substrate selectivity analyses of factor inhibiting hypoxia-inducible factor. *Angew Chem Int Ed Engl* 52, 1700-1704.

Yang, X.L., Zhang, Y.L., Lai, Z.S., Xing, F.Y., and Liu, Y.H. (2003). Pleckstrin homology domain of G protein-coupled receptor kinase-2 binds to PKC and affects the activity of PKC kinase. *World J Gastroenterol* 9, 800-803.

Yang, Y., Duan, Z., Skarpidi, E., Li, Q., Papayannopoulou, T., and Stamatoyannopoulos, G. (2001). Cloning and characterization of a potential transcriptional activator of human gamma-globin genes. *Blood Cells Mol Dis* 27, 1-15.

Young, R.M., Wang, S.J., Gordan, J.D., Ji, X., Liebhaber, S.A., and Simon, M.C. (2008). Hypoxia-mediated selective mRNA translation by an internal ribosome entry site-independent mechanism. *J Biol Chem* 283, 16309-16319.

Zhang, H., Bosch-Marce, M., Shimoda, L.A., Tan, Y.S., Baek, J.H., Wesley, J.B., Gonzalez, F.J., and Semenza, G.L. (2008). Mitochondrial autophagy is an HIF-1-dependent adaptive metabolic response to hypoxia. *J Biol Chem* 283, 10892-10903.

Zhang, H., Shi, X., Zhang, Q.J., Hampong, M., Paddon, H., Wahyuningsih, D., and Pelech, S. (2002). Nocodazole-induced p53-dependent c-Jun N-terminal kinase activation reduces apoptosis in human colon carcinoma HCT116 cells. *J Biol Chem* 277, 43648-43658.

Zhang, N., Fu, Z., Linke, S., Chicher, J., Gorman, J.J., Visk, D., Haddad, G.G., Poellinger, L., Peet, D.J., Powell, F., *et al.* (2010). The asparaginyl hydroxylase factor inhibiting HIF-1alpha is an essential regulator of metabolism. *Cell Metab* 11, 364-378.

Zhang, P., Yao, Q., Lu, L., Li, Y., Chen, P.J., and Duan, C. (2014). Hypoxia-inducible factor 3 is an oxygen-dependent transcription activator and regulates a distinct transcriptional response to hypoxia. *Cell Rep* 6, 1110-1121.

Zheng, X., Linke, S., Dias, J.M., Gradin, K., Wallis, T.P., Hamilton, B.R., Gustafsson, M., Ruas, J.L., Wilkins, S., Bilton, R.L., *et al.* (2008). Interaction with factor inhibiting HIF-1 defines an additional mode of cross-coupling between the Notch and hypoxia signaling pathways. *Proc Natl Acad Sci U S A* 105, 3368-3373.