Protein Design Based on a PHD Scaffold

Ann H. Y. Kwan

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

School of Molecular and Microbial Biosciences
University of Sydney
Sydney, Australia

June, 2004

Declaration

The work described in this thesis was performed between February 2001 and June 2004 in the School of Molecular and Microbial Biosciences (formerly The Department of Biochemistry) at the University of Sydney. The experiments described were performed by the author unless otherwise stated. This work has not been submitted, in part or in full, for the purpose of obtaining any other degree.

Ann H. Y. Kwan

Abstract

The plant homeodomain (PHD) is a protein domain of ~45–100 residues characterised by a Cys_4-His-Cys_3 zinc-binding motif. When we commenced our study of the PHD in. 2000, it was clear that the domain was commonly found in proteins involved in transcription. Sequence alignments indicate that while the cysteines, histidine and a few other key residues are strictly conserved, the rest of the domain varies greatly in terms of both amino acid composition and length. However, no structural information was available on the PHD and little was known about its function. We were therefore interested in determining the structure of a PHD in the hope that this might shed some light on its function and molecular mechanism of action.

Our work began with the structure determination of a representative PHD, Mi2β-P2, and this work is presented in Chapter 3. Through comparison of this structure with the two other PHD structures that were determined during the course of our work, it became clear that PHDs adopt a well-defined globular fold with a superimposable core region. In addition, PHDs contain two loop regions (termed L1 and L3) that display increased flexibility and overlay less well between the three PHD structures available. These L1 and L3 regions correspond to variable regions identified earlier in PHD sequence alignments, indicating that L1 and L3 are probably not crucial for the PHD fold, but are instead likely to be responsible for imparting function(s) to the PHD. Indeed, numerous recent functional studies of PHDs from different proteins have since demonstrated their ability in binding a range of other proteins.

In order to ascertain whether or not L1 and L3 were in fact dispensable for folding, we made extensive mutations (including both insertions and substitutions) in the loop regions of Mi2β-P2 and showed that the structure was maintained. We then went on to illustrate that a new function could be imparted to Mi2β-P2 by inserting a five-residue CtBP-binding motif into the L1 region and showed this chimera could fold and bind CtBP.

Having established that the PHD could adopt a new binding function, we next sought to use combinatorial methods to introduce other novel functions into the PHD scaffold. Phage display was selected for this purpose, because it is a well-established technique

and has been used successfully to engineer zinc-binding domains by other researchers. However, in order to establish this technique in our laboratory, we first chose a control system in which two partner proteins were already known to interact *in vitro*.

We chose the protein complex formed between the transcriptional regulators LMO2 and ldb1 as a test case. We have examined this interaction in detail in our laboratory, and determined its three-dimensional structure. Furthermore, inappropriate formation of this complex is implicated in the onset of T-cell acute lymphoblastic leukemia. We therefore sought to use phage display to engineer ldb1 mimics that could potentially compete against wild-type ldb1 for LMO2, and this work is described in Chapter 4. Using a phage library containing $\sim 3 \times 10^7$ variants of the LMO2-binding region of ldb1, we isolated mutants that were able to interact with LMO2 with higher affinity and specificity than wild-type ldb1. These ldb1 mutants represent a first step towards finding potential therapeutics for treating LMO-associated diseases.

Having established phage display in our laboratory, we went on to search for PHD mutants that could bind selected target proteins. This work is described in Chapter 5. We created three PHD libraries with eight randomized residues in each of L1, L3 or in both loops of the PHD. These PHD libraries were then screened against four target proteins. After four rounds of selection, we were able to isolate a PHD mutant (dubbed L13-FH6) that could bind our test protein Fli-ets. This result demonstrates that a novel function can be imparted to the PHD using combinatorial methods and opens the way for further work in applying the PHD scaffold to other protein design work.

In summary, the work detailed in Chapters 3 and 5 demonstrates that the PHD possesses many of the properties that are desirable for a protein scaffold for molecular recognition, including small size, stability, and a well-characterised structure. Moreover, the PHD motif possesses two loops (L1 and L3) of substantial size that can be remodeled for target binding. This may lead to an enhancement of binding affinities and specificities over other small scaffolds that have only one variable loop. In light of the fact that PHDs are mainly found in nuclear proteins, it is reasonable to expect that engineered PHDs could be expressed and function in an intracellular environment, unlike many other scaffolds that can only function in an oxidizing environment. Therefore, our results together with

other currently available genomic and functional information indicate PHD is an excellent candidate for a scaffold that could be used to modify cellular processes.

Appendices 1 and 2 describe completed bodies of work on unrelated projects that I have carried out during the course of my PhD candidature. The first comprises the invention and application of DNA sequences that contain all N-base sequences in the minimum possible length. This work is presented as a reprint of our recently published paper in Nucleic Acids Research. The second Appendix describes our structural analysis of an antifreeze protein from the shorthorn sculpin, a fish that lives in the Arctic and Antarctic oceans. This work is presented as a manuscript that is currently under review at the Journal of the American Chemical Society.

Acknowledgements

First and foremost, I would like to give the most sincere thanks to my supervisor Joel, for his help, advice, support and never-ending encouragement throughout the years. Joel has been a superb supervisor, but he is also a great mentor and a very good friend. I am especially grateful for his willingness and patience in teaching me MANY things (from NMR to how to slither down steep hills during bushwalks, and everything else in between). I have had a brilliant time in the last few years and I thank him for making my PhD experience such a wonderful one.

I am also very grateful to Jacqui for her help and advice with the molecular biology aspects of my work and introducing me to the world of LMO proteins. I thank her for always giving me time whenever I needed it, whether it was for proofreading or other things. But more importantly, I am overwhelmed by her kindness — from sewing pants for Pokemonk to making birthday cakes on my birthdays.

I would also like to give special thanks to Manuel for his help with the phage display work. Not only has he kindly offered us his vectors and constructs, he has also freely given me many tips, help and advice, which have been absolutely invaluable.

Thanks must also go to Big Chu, Dave, Belle, Janet, Raina and Fionna for their help with experiments and a countless numbers of things throughout the years. During my PhD, I have been extremely fortunate to be able to take part in other people's projects or have other people involved in mine. Thanks to everyone whom have played a role in my project and have allowed me to be a part of theirs (this includes many members of the Mackay, Matthews, Crossley and Harding labs). I am grateful for their generosity and help. And of course, a big thank-you to my fellow lab members (both ex and present) for their help and support, laughs, jokes and their company throughout the years. Many thanks also for their tolerance towards Pokemonk's daily visits. I must also thank Pokemonk, my favorite stuffed monkey for always being there (literally)! Last, but not least, I would like to thank my family for their support throughout the years. I am also grateful for the financial support provided by an Australian Postgraduate Award.

Contents

List of Abbreviations

1D	one-dimensional NMR spectrum
BKLF-F3	basic Krüppel-like factor
BSA	Bovine serum albumin
CD	circular dichroism
CDR	complementarity-determining region of an antibody
cfu	colony forming unit
CV	column volume
DMSO	dimethyl sulfoxide
DQF-COSY	double-quantum-filtered *J*-correlated spectroscopy
dsDNA	double-stranded DNA
DSS	2,2-Dimethylsilapentane-5-sulfonic acid
DTT	dithiothreitol
EDTA	ethylene-diamine-tetraacetic acid (disodium salt)
ELISA	enzyme-linked immunosorbent assay
ESMS	electrospray mass spectrometry
Fli-ets	Ets domain of Fli-1
FN3	fibronectin type III domain
GSH	glutathione (reduced form)
GST	glutathione *S*-transferase
HSQC	heteronuclear single-quantum coherence
IC_{50}	median inhibitory concentration
IPTG	isopropyl (β)-D-thiogalactopyranoside
K_d	dissociation constant

LB	Luria-Bertoni broth
LB-Amp	Luria-Bertoni broth with ampicillin
ldb1-LID	LID domain of ldb
L1	Loop 1 region of Mi2β-P2
L1-lib	PHD library with randomized residues in L1
L13	Loop 3 region of Mi2β-P2
L13-lib	PHD library with randomized residues in L1 and L3
L3-lib	PHD library with randomized residues in L3
LID	ldb1-interacting domain
LMO	LIM-only
LMO2-LIM1	N-terminal LIM domain of LMO2
LMO4-LIM1	N-terminal LIM domain of LMO4
Mi2β-P2	second PHD from Mi2β
MBP	maltose-binding protein
MQW	Milli-Q® water
mRNA	messenger RNA
MW	molecular weight
NMR	nuclear magnetic resonance
NOE	peak in NOESY spectrum resulting from dipolar connectivity
NOESY	nuclear Overhauser enhancement spectroscopy
OD	optical density
p3	coat protein 3 from the filamentous M13 bacteriophage
p8	coat protein 8 from the filamentous M13 bacteriophage
PBS	phosphate-buffered saline

PCR	polymerase chain reaction
PHD	plant-homeodomain
PMSF	phenylmethylsulfonylfluoride
ppm	parts per million
rpHPLC	reverse-phase high pressure liquid chromatography
scFv	single-chain variable fragment of an antibody
ssDNA	single-stranded DNA
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
T-ALL	T-cell acute lymphoblastic leukemia
TACC77	C-terminal coiled coil region of TACC
TCEP	tris(2-carboxyethyl)phosphine
TE	Tris-EDTA
TFA	trifluoroacetic acid
TOCSY	total correlation spectroscopy
Tris	Tris[hydoxymethyl]aminomethane

Publications arising from my PhD work

Journal Articles

1. Newton A. L., Sharpe B. K., Kwan A., Mackay J. P. and Crossley M. The transactivation domain within cysteine/histidine-rich region 1 of CBP comprises two novel zinc-binding modules. (2001) **J. Biol. Chem.** 275 (20): 15128–34.

2. Sharpe B. K., Matthews J. M., Kwan A.H., Newton A. L., Gell D. A., Crossley M. and Mackay J. P. A new zinc-binding fold underlines the versatility of zinc-binding modules in evolution. (2002) **Structure** 10(5): 639–48.

3. Deane J., Mackay J., Kwan A.H., Sum E., Visvader J. and Matthews J. Structural basis for the recognition of ldb1 by the N-terminal LIM domains of LMO2 and LMO4. (2003) **EMBO J.** 22(9): 2224–33.

4. Plambeck C.A., Kwan A.H., Adams D., Westman B.J., van der Weyden L., Medcalf R.L., Morris B. and Mackay J.P. The solution structure of the zinc finger domain from human splicing factor ZNF265. (2003) **J. Biol. Chem.** 278(25): 22805–11.

5. Kwan A.H., Gell D.A., Verger A., Crossley M., Matthews J.M. and Mackay J.P. Engineering a protein scaffold from a PHD finger. (2003) **Structure** 11(7): 803–13.

6. Kwan A.H., Perdomo J., Czolij R., Mackay J.P. and Crossley M. Pentaprobe: a comprehensive sequence for the one-step detection of DNA-binding activities. (2003) **Nucleic Acids Res.** 31(20): e124.

7. Donadini R., Liew C.W., Kwan A.H., Mackay J., and Fields B. Crystal and solution structures of a superantigen from Yersinia pseudotuberculosis reveal a jelly-roll fold similar to viral capsid proteins and TNF-superfamily members. (2004) **Structure** 12(1): 145–56.

8. Kwan A.H., Fairley K., Anderberg P.I., Haymet A.D., Harding M.M. and Mackay J.P. Type I sculpin antifreeze protein: solution structure of recombinant SS3 and an acetylated derivative. Submitted to **J. Am. Chem. Soc.** (May, 2004; manuscript number ja047135f)

Conference proceedings and abstracts

Oral Presentations and Abstracts (only first author presentations listed)

1. Kwan A.H., Fairley K., Westman B.J., Haymet, A.D.J., Anderberg P.I., Harding M.M. and Mackay J.P. SS3 - an antifreeze protein from Arctic Fish. (2003) **East Coast Protein Meeting (Coffs Harbour, Australia).**
2. Kwan A.H., Gell D.A., Crossley M., Matthews J.M. and Mackay J.P. Protein Design using PHD Fingers. (2003) **ComBio2003 Combined Conference (Melbourne, Australia).**

Poster Presentations and Abstracts (only first author presentations listed)

1. Kwan A.H., Gell D.A. and Mackay J.P. Zinc-binding domains that mediate protein interactions in chromatin remodeling. (2001) **26th Lorne Conference on Protein Structure and Function (Lorne, Australia).**
2. Kwan A.H., Gell D.A., Crossley M. and Mackay J.P. PHD domains that mediate protein interactions in chromatin remodeling. (2001) **ASMR Scientific Meeting (Sydney, Australia).**
3. Kwan A.H., Gell D.A., Crossley M. and Mackay J.P. PHD domains that mediate protein interactions in chromatin remodeling. (2001) **ComBio2001 Combined Conference (Canberra, Australia).**
4. Kwan A.H., Gell D.A. and Mackay J.P. PHD Domains from the Dermatomyositis-specific autoantigen (Mi2β). (2002) **27th Lorne Conference on Protein Structure and Function (Lorne, Australia).**
5. Kwan A.H., Gell D.A. and Mackay J.P. PHD Domains from the Dermatomyositis-specific autoantigen (Mi2β). (2002) **Australian and New Zealand Society of Magnetic Resonance Meeting (Taupo, New Zealand).**
6. Kwan A.H., Gell D.A. and Mackay J.P. Versatile PHD zinc fingers. (2002) **XXth International Conference on Magnetic Resonance in Biological Systems (Toronto, Canada).**

7. Kwan A.H., Gell D.A., Matthews J.M. and Mackay J.P. Protein engineering with PHD zinc fingers. (2002) **ComBio2002 Combined Conference (Sydney, Australia).**
8. Kwan A.H., Gell D.A., Crossley M., Matthews J.M. and Mackay J.P. PHD fingers as protein scaffold. (2002) **2nd International Conference on Structural Biology and Functional Genomics (Singapore).**
9. Kwan A.H., Plambeck C.A., Adams D., van der Weyden L., Morris B., and Mackay J.P. Solution structure of the first zinc finger domain from ZNF265 - a human splicing factor. (2003) **28th Lorne Conference on Protein Structure and Function (Lorne, Australia).**
10. Kwan A.H., Czolij R., Mackay J.P. and Crossley M. Pentaprobe: a comprehensive sequence for the one-step detection of DNA-binding proteins. (2003) **BioInfoSummer 2003 – AMSI Summer Symposium in Bioinformatics (Canberra, Australia).**
11. Kwan A.H., Ryan D.P., Deane J.E., Baca M., Mackay J.P. and Matthews J.M. Finding the perfect LID for LIM-only proteins. (2004) **29th Lorne Conference on Protein Structure and Function (Lorne, Australia).**
12. Kwan A.H., Liew C.K. and Mackay J.P. The structure of a zinc finger-zinc finger transcriptional complex. (2004) **Australian and New Zealand Society of Magnetic Resonance Meeting (Adelaide, Australia).**