In search of novel folds: Protein evolution via non-homologous recombination

A dissertation presented in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Biochemistry

at Massey University, Albany, New Zealand

Mayank Saraswat

2014

Abstract

The emergence of proteins from short peptides or subdomains, facilitated by the duplication and fusion of the minigenes encoding them, is believed to have played a role in the origin of life. In this study it was hypothesised that new domains or basic elements of protein structure, may result from non-homologous recombination of the genes coding for smaller subdomains.

The hypothesis was tested by randomly recombining two distantly related $(\beta\alpha)_8$-barrel proteins: *Escherichia coli* phosphoribosylanthranilate isomerase (PRAI), and β subunit of voltage dependent K^+ channels (Kvβ2) from *Rattus norvegicus*. The aim was to identify new, folded structures, which may or may not be $(\beta\alpha)_8$-barrels. Incremental truncation (ITCHY), a method for fragmenting and randomly recombining genes, was used to mimic *in vivo* non-homologous recombination and to create a library of chimeric variants. Clones from the library were selected for right reading frame and solubility (foldability) of the recombined chimeras, using the pSALect selection system. Out of the six clones identified as soluble by pSALect, only one (P25K86) was found to be actually soluble. The protein, P25K86, was found to form oligomers and on treatment with a reducing agent, β-mercaptoethanol the multimeric state disappeared. The protein has three cysteines and one of the cysteines (Cys56) was found to mediate in the bond formation, thus giving a dimeric state. An engineered version of P25K86 that has the Cys56 replaced by serine was expressed as a monomer and additionally it was found to be

more stable.

As the pSALect folding selection system reported false positives, i.e. only one of the six chimeras was actually soluble, it was concluded that the *in vivo* solubility selection system was leaky. A series of experiments were conducted so as to improve pSALect that led to the creation of pFoldM – a more stringent selection system, discussed in chapter 4. Comparing the newer improved version with the old, two more interesting chimeras were discovered.

A total of 240,000 non-homologous recombination events were created *in vitro* and three soluble chimeras (evolutionary solutions) were found. Data from circular dichroism spectroscopy (CD) combined with heteronuclear single quantum coherence (HSQC) spectra suggest that the proteins, P24K89 and P25K86, are present in a molten globule state. ITCHY, as a means of mimicking the subdomain assembly model, was applied *in vitro.* The discovery of two interesting chimeras (P25K86 and P24K89) using high-throughput engineering experiments widens the possibilities of exploring the protein structure space, and perhaps offers close encounters with these *never born proteins* that may be trapped in an ensemble of fluctuating (structured and unstructured) states.

Acknowledgements

Boy it has been a philosophical journey! I have learnt a lot about myself and that is because of the people I met and the places I went. When I left the UK for New Zealand, I was passionate to embark on this journey to explore the "protein universe". My supervisor, Dr. Wayne Patrick gave me this opportunity and taught me some really clever techniques. One in particular is my favourite and is called ITCHY. For that I thank him a lot. I also thank him because he coped with my craziness, my bipolarity and my entrepreneurial spirit.

Not long ago, Dr. Jane Allison did rescue me from sinking. I was very close to quit my PhD and move on with my ventures. She motivated me and told me that that there is light at the end of this tunnel. So for that I thank her a lot!

You can't study proteins if you don't have a crystallographer or a structural biochemist on board. My second supervisor, Dr. Gill Norris gave me the chance to work in her lab, for which I am extremely grateful. Trevor Loo has been phenomenal in showing me how things work in the lab and for that I will always be obliged.

Everything comes down to money, so I would like to thank Massey University for offering the scholarship and support to pursue this research. I would also like to thank the New Zealand Society for Biochemistry and Molecular Biology (NZSBMB) and Protein Society (USA) that provided the travel grant, which made my journey to USA possible. I am a huge fan of Marc

Ostermeier and I had the chance to meet him in the conference. Sadly I am leaving academia but I am grateful to Wayne, who arranged a lunch in some food court in Boston where I had the chance to talk to him about my research.

Several members from the lab need special mention: Daniela Quintana, Paulina Hanson-Manful, Matilda Newton (Oh I am so going to miss you pal), Matteo Ferla and Valerie Soo. These folks always make me happy.

I would also like to thank my really good friends: Ralph Grand, Lutz Gehlen, Katrin Hammerschmidt, Matt Woods, Peter Deines, Caroline Cotty, Frazer Nobel, Brig Kreig, Anthony Bloom, Stephen Carr, Jon Atherton, Leon Kapetas, Charlie Kerr, Stephen Wallace and Sharif Zamir.

Being an entrepreneur, I have also met some really encouraging people who have constantly motivated me to finish this thesis. Some of the best people I know are: Jim McNaught (my Guru), John Sullivan, Steve Corbett, John Harrison and lots of other good people (the list is way too long).

Lastly, I would like to thank my Mum and Dad. Mum for always nagging me to finish what I have started and Dad for helping me buy the toys I can't live without. And of course my two sisters, Meetu and Nikki for their endless support and love.

Table of contents

Abbreviations

αs-dNTPs	α-phosphorothioate (αS)-dNTPs (deoxyribonucleotide triphosphates)
c.f.u	colony-forming unit
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
$(His)_6$	Hexa-histidine
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
Kvβ2	β subunit of voltage dependent K^+ channels
MWCO	Molecular Weight Cut Off
OD_{600}	Optical density at 600 nm
NADPH	Reduced form of β-nicotinamide adenine dinucleotide phosphate
NMR	Nuclear magnetic resonance
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
POI	Protein Of Interest
PRAI	Phosphoribosylanthranilate isomerase
SEM	Standard Error of the Mean
SOB	Super Optimal Broth
SOC	SOB with Catabolite repression
Tat	Twin arginine trasnlocase

Tris	Tris (hydroxymethyl) aminomethane
trPRAI	truncated version of PRAI