

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

*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

© 2014
Mayank (Mack) Saraswat
All rights reserved.

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 (Kv β 2) 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

1.	Introduction - The Protein Universe	1
1.1	Protein domains and folds defined	2
1.2	Protein Evolution	3
1.2.1	The role of subdomains in protein evolution	5
1.2.2	The primordial minigenes coding short peptides	8
1.3	Non homologous recombination and its role in generating novelty	10
1.3.1	Non-homologous recombination: definition and examples	10
1.3.2	Generating novelty <i>via</i> NHR	12
1.4	The subdomain assembly model: a central theme of this research project-	16
1.5	Aims of this study	18
2.	Mimicking non-homologous recombination	20
2.1	Premise of the chapter	21
2.2	Introduction	22
2.2.1	TIM barrels – A proof-of-principle	22
2.2.1.1	The two distantly related barrels	24
2.2.2	Protein chimeragenesis: an engineering approach to mimic non - homologous recombination	27
2.2.2.1	Making the THIO-ITCHY library	35
2.3	Results	37
2.3.1	Library design	37
2.3.1.1	Making the long fusion PCR product	37
2.3.1.2	Making a unidirectional library	38
2.3.2	Crossover plots	42
2.3.3	The distribution of crossovers	43
2.3.4	Solubility validation	44
2.3.4.1	Proteins from cluster A	45
2.3.4.2	Proteins from cluster B	49
2.4	Discussion	52
2.4.1	pSAlect is not robust enough	52
2.4.2	The outcome of random recombination	53
2.5	Materials and methods	56

- 2.5.1 Construction of pSALect-trPRAI and pSALect-Kvβ2 56
- 2.5.2 ITCHY library construction 57
- 2.5.3 Screening the clones 62
- 2.5.4 Scaling up and harvesting the library 63
- 2.5.5 Folding selection 63
- 2.5.6 Sequencing 64
- 2.5.7 Construction of the expression vector pLAB101 64
- 2.5.8 Expression and purification of chimeric proteins 66

3. Characterisation of P25K86 68

- 3.1 Premise of the chapter 69
- 3.2 Introduction 70
 - 3.2.1 Borrowing some tools from biophysics to study P25K86 70
 - 3.2.2 Circular dichroism spectroscopy 71
 - 3.2.2.1 Secondary structure analysis 72
 - 3.2.3 Nuclear magnetic resonance spectroscopy 74
 - 3.2.3.1 Chemical shift 74
 - 3.2.4 Protein NMR 75
- 3.3 Results 77
 - 3.3.1 Oligomerisation 77
 - 3.3.1.1 Size exclusion chromatography 77
 - 3.3.1.2 High performance liquid chromatography (HPLC) 80
 - 3.3.1.3 Mass spectrometry (MS) 81
 - 3.3.2 Investigating the role of cysteine in oligomerisation and a quick test for function 82
 - 3.3.3 The P25K86_CCS transition 84
 - 3.3.3.1 Labelling attempts for NMR spectroscopy 85
 - 3.3.3.1.1 Far-UV circular dichroism spectroscopy 87
 - 3.3.4 NMR spectroscopy of P25K86_CCS 92
- 3.4 Discussion 97
 - 3.4.1 Switching from P25K86 to P25K86_CCS 97
 - 3.4.2 What can we learn from this experiment? 100
- 3.5 Materials and methods 101
 - 3.5.1 Expression and purification of P25K86 & its variants 101
 - 3.5.2 Buffer for biophysical experiments 102
 - 3.5.2.1 Exchanging the buffer *via* concentrator 104
 - 3.5.3 Measuring concentration 104
 - 3.5.4 Gel filtration & HPLC 105
 - 3.5.5 Mass Spectrometry 106
 - 3.5.6 Making the P25K86 mutants 106
 - 3.5.7 Immunoblotting 108
 - 3.5.8 Labelling P25K86_CCS 109
 - 3.5.8.1 Preparing M9 minimal media base 110

- 3.5.8.2 Modified Media 110
- 3.5.9 Circular dichroism spectroscopy 111
- 3.5.10 NMR spectroscopy 112

4. Improving the folding selection system 113

- 4.1 Premise of the chapter 114
- 4.2 Introduction 115
 - 4.2.1 Switch to pInSAlect 115
 - 4.2.2 Protein translocation via Tat and Sec pathways 116
 - 4.2.2.1 Tat *versus* Sec 117
 - 4.2.2.2 Export promiscuity 118
 - 4.2.3 Other improvements 120
 - 4.2.3.1 Improving the desalting step prior to transformation 120
 - 4.2.3.2 Switching to time-dependent ITCHY 121
 - 4.2.4 Focus of the chapter 122
- 4.3 Results 124
 - 4.3.1 Attempts to improve the folding selection system 124
 - 4.3.1.1 Step 1 of the two-step strategy to reengineer pSAlect 124
 - 4.3.1.2 Step 2 of the two-step strategy to reengineer pSAlect 126
 - 4.3.2 Initial tests with pFoldM-KRK and pFoldM-KR 129
 - 4.3.3 A new library 133
 - 4.3.3.1 Test library results – a disappointment 134
 - 4.3.4 Time-dependent ITCHY of PRAI-Kvβ2 135
 - 4.3.4.1 Optimized time-dependent ITCHY protocol 136
 - 4.3.5 A library that pushed things forward 138
 - 4.3.6 Comparing the three folding selection systems 139
 - 4.3.6.1 Is temperature the key to filtering out false positives? 141
 - 4.3.6.2 Crossover distribution of PRAI-Kvβ2 in the three-selection systems 142
 - 4.3.7 An overall ITCHY comparison 145
 - 4.3.8 Solubility validation 147
 - 4.3.8.1 P19K93: a quirky selection 149
 - 4.3.8.2 The protein P24K89 150
 - 4.3.8.2.1 Purification attempts for biophysical characterization 151
 - 4.3.8.2.2 CD spectrum of P24K89 153
 - 4.3.8.2.3 Impediments in NMR trials 155
 - 4.3.9 Support from our collaborators 156
 - 4.3.10 Comparing P24K89 and P25K86_CCS
- 4.4 Discussion 162
 - 4.4.1 An improved selection for protein folding? 162
 - 4.4.2 Comparing the three folding selection systems 164
 - 4.4.3 Comparing and contrasting the three soluble chimeras 166

4.5	Materials and methods	170
4.5.1	Construction of pFoldM-KR and pFoldM-KRK	170
4.5.1.1	Step I of the strategy	170
4.5.1.2	Step II of the strategy	172
4.5.2	Sub-cloning P25K86 and P69K149 in both versions of pFoldM	174
4.5.2.1	Screening the clones	175
4.5.3	Sub-cloning PRAI and Kv β 2 in the vector backbone pInSAlect	176
4.5.3.1	Constructing the test library	176
4.5.4	THIO-ITCHY library	177
4.5.5	Screening the clones	177
4.5.6	Time-dependent ITCHY library	178
4.5.6.1	Steps of the optimized protocol	178
4.5.7	Column <i>versus</i> drop dialysis	180
4.5.8	Scaling up and harvesting the library	181
4.5.9	Frame selection	181
4.5.10	Sequencing	182
4.5.11	Comparing the fold selection system	182
4.5.11.1	Plating experiment	183
4.5.12	Expression and purification of chimeric proteins	184
4.5.13	Exchanging the buffer via concentrator and dialysis	185
4.5.14	Labelling P24K89	186
4.5.15	Circular dichroism spectroscopy	187
4.4.16	Nuclear magnetic resonance spectroscopy	187

5. Conclusions 161

5.1	The findings	189
5.2	Combinatorial assembly: a way to innovate	191
5.3	Engineering functions on artificially created folds	193
5.4	In the future	194

Appendix I. General materials and methods 197

AI.1	Reagents	197
AI.2	Growth media and antibiotics	197
AI.3	Storing strains	198
AI.4	Bacterial strain	198
AI.5	Plasmids	198
AI.6	Oligonucleotides	200
AI.7	Software	201
AI.8	Agarose gel electrophoresis	201
AI.9	SDS-PAGE	201
AI.10	Electrocompetent cells	202
AI.11	Sequencing	203

Appendix II. Statement of contributions 204

Appendix III. Publication arising from this work 206

Appendix IV. Supplementary data 210

- AIV.1 Protein sequence of chimeras 210
- AIV.2 DNA and protein sequence of Kv β 2 and trPRAI 211
- AIV.3 Multiple sequence alignment of all chimeras 213
- AIV.4 Secondary structure elements in trPRAI; PDB-2KZH 214
- AIV.5 Secondary structure elements in Kv β ; PDB-1EXB 215
- AIV.6 Deconvolution of Kv β 2 216
- AIV.7 Deconvolution of trPRAI 216
- AIV.8 NADPH-binding subdomain experiment 217
- AIV.9 SDS-PAGE gel of P25K86 218

References. 219

Abbreviations

α S-dNTPs	α -phosphorothioate (α S)-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
Kv β 2	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 translocase

Tris	Tris (hydroxymethyl) aminomethane
trPRAI	truncated version of PRAI