Beyond BLASTing: Ribonucleoprotein evolution via structural prediction and ancestral sequence reconstruction

A thesis presented in partial fulfilment
of the requirements for the degree of

Doctor of Philosophy in
Genetics

at Massey University, Manawatū Campus

Toni K. Daly

2016

Abstract

Primary homology in DNA and protein sequence has long been used to infer a relationship between similar sequences. However gene sequence, and thus protein sequence, can change over time. In evolutionary biology that time can be millions of years and related sequences may become unrecognisable via primary homology. This is demonstrated most effectively in chapter 4a (figure 10). Conversely the number of possible folds that proteins can adopt is limited by the attractions between residues and therefore the number of possible folds is not infinite. This means that folds may arise via convergence between evolutionarily unrelated DNA sequences.

This thesis aims to look at a process to will wring more information from the primary protein sequence than is usually used and finds other factors that can support or refute the placement of a protein sequence within the family in question. Two quite different proteins; the Major Vault Protein whose monomers make up the enigmatic vault particle and the argonaute family of proteins (AGO and PIWI) that appear to have a major hand in quelling parasitic nucleic acid and control of endogenous gene expression, are used to demonstrate the flexibility of the workflow.

Principally the method relies on prediction of three-dimensional structure. This requires at least a partially solved crystal structure but once one exists this method should be suitable for any protein. Whole genome sequencing is now a routine practice but annotation of the resultant sequence lags behind for lack of skilled personnel. Automated pipeline data does a good job in annotating close homologs but more effort is needed for correct annotation of the exponentially growing data bank of uncharacterised (and wrongly characterised) proteins. Lastly, in deference to budding biologists the world over, I have tried to find free stable software that can be used on an ordinary personal computer and by a researcher with minimal computer literacy to help with this task.

Acknowledgments

I have been a student since 1988 but now is the time to thank everyone and of course over the years there have been so many people that have inspired and encouraged me and I will miss some people that I shouldn't have.

Trevor Kitson made me realise that it was possible to be smart and funny, Mike Hardman for having the foresight to grab a box of tissues whenever I walked through his office door when I eventually became an internal student. Although Massey could bend the rules in those pre-studylink days a student loan required a minimum of seven papers…….but I only had one semester! So I physically couldn't attend all my classes, I just grabbed the notes, sat the exams and came back the following year to do the lab work. I want to thank Mark Patchett for just being the best tonic one could ever need when things looked glum, Rosie Bradshaw for her kindness, and Kathryn Stowell, a newly minted lecturer when I was an internal student, has encouraged me ever since.

Lesley Collins for allowing a complete unknown to contribute to her book, Austen Ganley from Albany for helping to prepare me for Vienna when Palmerston North was just too far away and lastly Andrew Sutherland-Smith and David Penny who have stoically listened to my troubles academic and personal once a week for six years. In retrospect I do not know how they coped and without them I surely would have quit. I also thank David for organising payment of my fees, and Massey for paying the bill.

It is an extraordinarily lonely thing to attempt a PhD without the camaraderie and the vicarious learning opportunities of watching fellow students give presentations etc. I want to especially thank Bruce White from the library for his help and patience and Tim White for helping me navigate computer-speak.

I set out to show the students of Northland New Zealand, that you can achieve your goals without the financial, geographical and educational advantages that you perceive everyone else to have. Protein annotation in particular needs help and people all over the world with access to a computer and an internet link can join in.

My biggest thanks though must go to my husband Dan Daly for his continual support despite landing him with three children as well as a full time job while I did my undergraduate stints, and I know that I have been a serious drain on finances and I know that we haven't been out in the kayaks (or out anywhere) for a long time and now I promise that our time will be spent together.

Preface

This thesis is written according to the regulations of the latest version of the Handbook for Doctoral Study (2016), published online by the Doctoral Research Committee. This thesis complies with the format of a thesis based on publication as described in the handbook and includes both published and unpublished chapters. The chapters do not follow the order of publication in order to better demonstrate the flow of the development of the work. Chapters 1 and 2 have been written by Toni Daly as an introduction and literature review and are not intended for publication.

Chapter 1. General Introduction.

Chapter 2. Literature review of the Major Vault Protein.

Chapter 3a. Toni K Daly *et al.* (2013) Beyond BLASTing: Tertiary and quaternary structure analysis helps identify Major Vault Proteins. Genome Biology and Evolution 5: 217-232.

Chapter 3b. Toni K Daly *et al*. (2013) In silico resurrection of the Major Vault Protein suggests it is ancestral in modern eukaryotes. Genome Biology and Evolution 5 (8): 1567-1583.

Chapter 4. Toni Daly, X. Sylvia Chen and David Penny (2011) How old are RNA networks? (L J Collins ed. RNA Infrastructure and Networks) Landes BioScience and Springer Science.

Chapter 4a. Toni Daly, *et al*. (2016) Long Long AGO: The evolutionary history of Argonaute and PIWI in metazoa by ancestral protein inference and structure prediction. (submitted).

Chapter 4b. Toni K Daly *et al*. (2016) Argonaute gain and loss during fungal evolution. (submitted).

Chapter 4c. Toni Daly *et al*. Argonautes origins in eukaryotes. (in preparation).

Chapter 6. Conclusion.

David Penny is co-author on all of the published and prepared papers and Andrew Sutherland-Smith is co-author on five of them. Contributions to each paper are described in Appendix IV.

Contents

List of Figures

Chapter one:

Chapter Two (published papers)

Preface

Chapter 2a:

Chapter 2b:

Chapter Three:

Chapter Four (submitted papers)

Preface

Chapter 4a: