

Regulatory mechanisms and biological implications of protein complex assembly

Thesis Plan

Jonathan Wells

Outline of thesis chapters

Declaration

Abstract

Acknowledgements

Table of Contents

1. **Introduction and literature review**
2. **Hawk proteins - a paralogous family of Smc-kleisin regulators**
 - 2.1. Smc-kleisin complexes
 - 2.2. A network approach to determining paralogous proteins
 - 2.3. Hawks arose early in eukaryote evolution
 - 2.3.1. Not detected in *Lokiarchaeota sp.*
 - 2.3.2. Are found in all major eukaryotic clades
 - 2.4. Hawks contributed to the evolution of condensin and cohesin
 - 2.4.1. Smc5-6 regulators Nse5-6 are not HEAT proteins, as formerly thought
 - 2.4.2. Similarities between hawks do not reflect cohesin/condensin divide
3. **Operon gene order is optimized for ordered protein complex assembly**
 - 3.1. Protein complexes assemble via ordered pathways
 - 3.2. Encoding complexes in operons enhances assembly efficiency
 - 3.2.1. Operons facilitate co-translational assembly
 - 3.3. Interacting subunits tend to be encoded by adjacent genes
 - 3.4. Operon gene order is optimized for the order of subunit assembly
 - 3.5. Optimization is most important for lowly expressed complexes
 - 3.6. Implications for regulation of assembly in eukaryotes
4. **Protein complex signatures in subunit degradation kinetics**
 - 4.1. Many proteins do not follow first-order rate degradation kinetics
 - 4.2. Non-exponentially degraded proteins are enriched in complexes
 - 4.3. NED subunits are overexpressed relative to ED
 - 4.4. NED proteins imply the existence of 'core' subunits
 - 4.4.1. Alternative explanations of observations
 - 4.4.2. Support for 'core' subunits in other work
5. **Autosomal dosage compensation in aneuploid cells (WIP)**
 - 5.1. Expression of subunits is often attenuated in aneuploid cells
 - 5.2. Are core subunits more likely to be attenuated?
 - 5.3. Adaptation or an unrelated consequence of subunit stability?
6. **Transcriptional regulation of eukaryotic complex assembly (WIP)**
 - 6.1. Members of the same complex are regulated by similar TFs?
 - 6.2. Any interesting patterns in gene location?

7. Analysis of transcript and protein abundance analyses?

7.1 Coexpression calculated via mRNA vs protein. Which is better?

8. Discussion

8.1. Summarise findings, present unifying explanation of results.

8.2. Key findings from project, why they advance the field.

8.3. Limitations of each sub-project, what could be improved.

8.4. What are the major outstanding questions, where next?

9. Methods

9.1. Hawk methods

9.1.1. Network approach to gene family assignment

9.1.2. Structural/Sequence analyses

9.2. Operon order methods

9.2.1. Mapping subunits to operons

9.2.2. Protein abundance data

9.2.3. Assembly order prediction

9.3. Relevant protein degradation methods

9.3.1. Overview of mass spec protocol (not my work)

9.3.2. Coexpression calculation, assembly, etc.

9.4. Aneuploidy methods

9.5. Transcription factor network methods

Bibliography

List of abbreviations

List of figures

Appendices

Notes on chapters

1. Introduction and literature review

I haven't yet decided how to structure this, as it will probably be fairly dependent on how current work-in-progress and future projects go. Joe mentioned that if it's thought out properly then it might be possible to publish in a modified form, which I like the idea of.

2. Hawk proteins - a paralogous family of Smc-kleisin regulators

This section will draw on results from a paper currently undergoing peer-review. It will describe a sub-family of HEAT repeat proteins that regulate Condensin and Cohesins. These were shown to be paralogous through the application of a novel network-based method, built off the back of extensive HMM-profile alignments. Due to their phylogenetic distribution, it appears that they were at least partially responsible for the emergence of novel functions in the branch of Smc-kleisin complexes containing condensin and cohesins. As such it makes a nice point about the way in which neo/sub-functionalization can occur at the level of protein complexes, which I will elaborate on in the discussion.

3. Operon gene order is optimized for ordered protein complex assembly

Chapter 2 will finish by highlighting the importance of subunit gain and loss in driving evolution. This should lead nicely into this chapter, which looks at selection pressures driven by processes that happen on a timescale of seconds to hours - namely the assembly of protein complexes. I will mostly be restating the findings from the paper published on the topic, but will try to integrate any relevant insights from later projects where I can.

4. Protein complex signatures in subunit degradation kinetics

One of the collaborations I've been involved with came up with the important finding that many proteins are degraded non-exponentially, and that these proteins are highly enriched in protein complexes. The work I did for this comprised a fairly small part of a fairly big paper, but I have a few other results and thoughts on the topic which didn't make it in. I will use this space to elaborate on some of these, and to set the stage for the following chapter, which stems directly from the paper.

N.B. The following three chapters are largely hypothetical, and will depend a lot on the direction work goes in.

5. Autosomal dosage compensation in aneuploid cells (WIP)

A major contribution to the deleterious effects of aneuploidy comes from the resulting proteotoxic stress accompanying mass copy number increases. Whilst the majority of proteins show increases in abundance that correspond to copy number, a significant fraction do not. These attenuated proteins are highly enriched in protein complexes. From the paper in the previous chapter, we were able to show that in aneuploid cells the non-exponentially degraded proteins were frequently attenuated in a similar fashion. There is significant overlap between NED proteins, protein complex subunits and attenuated proteins.

I am currently working to try and better define the relationship between protein complex structure and attenuation in autosomal proteins. If this project is fruitful, then I will devote this chapter to the results.

6. Members of the same complex are regulated by similar TFs

As mentioned earlier, this is still very much a hypothetical chapter, but if I can get the necessary work done then it should be a nice demonstration of regulation of assembly in eukaryotes. This would provide a nice counterpoint to chapter 3, which is focused on prokaryotes.

7. Analysis of transcript and protein abundance analyses

Work I've done on this so far has been fairly inconclusive, but the basic idea is that you might expect to see a better correlation in subunit levels when measuring the abundance of protein, rather than mRNA. My feeling is that this is currently confounded significantly by the quality differences in RNA-seq vs the less well developed mass-spec data. However, since data quality seems to be improving, there is a possibility that this will go somewhere, so may be included as a chapter.