

INVESTIGATING THE RECOGNITION
AND INTERACTIONS OF NON-POLAR
 α HELICES IN BIOLOGY.

A THESIS SUBMITTED TO THE UNIVERSITY OF MANCHESTER
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN THE FACULTY OF LIFE SCIENCES

2016

James Baker

Contents

Abstract	5
Declaration	6
Copyright Statement	7
Acknowledgements	8
1 Introduction	9
1.1 The importance of transmembrane proteins.	9
1.2 The positive inside rule.	11
2 The “negative-not-inside” rule	12
3 Conclusions	13
A Big tables	16

List of Tables

List of Figures

The University of Manchester

James Baker

Doctor of Philosophy

Investigating the Recognition and Interactions of Non-Polar α Helices in Biology.

May 25, 2016

Transmembrane α helix containing proteins make up around a quarter of all proteins, as well as two thirds of drug targets, and contain some of the most critical proteins required for life as we know it. Yet they are fundamentally difficult to study experimentally due to the very features that make them so biologically influential: their hydrophobic transmembrane helices. What is now missing is a complex, nuanced understanding of this helix composition. Currently it is known that the properties of transmembrane protein α -helices underpin membrane protein insertion mechanisms and furthermore can be used to predict presence of function in the transmembrane helix itself. By leveraging large datasets of transmembrane proteins, this thesis is focussed on characterising features of α helices en masse, particularly regarding their topology, membrane-protein interactions, and intra-membrane protein interactions.

Herein we expand on the core understanding of the biophysicochemical properties of these helices. We find evidence of a universal, “negative-not-inside” rule that complements the famous “positive-inside rule” as well as intramembrane leucine propensity for the inner leaflet.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

- i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.
- ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.
- iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.
- iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see <http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487>), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see <http://www.manchester.ac.uk/library/aboutus/regulations>) and in The University’s Policy on Presentation of Theses.

Acknowledgements

So long! And thanks for all the fish.

Chapter 1

Introduction

1.1 The importance of transmembrane proteins.

Transmembrane Proteins (TMPs) underpin almost every biological process directly, or indirectly, from photosynthesis to respiration. Integral TMPs are encoded by around 30% of the genes in the human genome which reflects their biological importance (1). In the 1990s and early 2000s the Transmembrane Helix (TMH) was a rather simple topological story. TMPs were thought to contain membrane spanning bundles of non-polar α -helices, with each TMH having a consistent orientation of being perpendicular to the membrane surface. Although this is broadly true, hundreds of high quality membrane structures have elucidated that TMHs can adopt a plethora of lengths and orientations within the membrane. TMHs are capable of partial spanning of the membrane, spanning using oblique angles, and even lying flat on the membrane surface (2, 3).

Because of the experimental hinderence, the story of transmembrane proteins has been relatively slow to emerge. In the 1990s and early 2000s the story was seemingly uncomplicated. There were membrane-spanning bundles of non-polar α -helices of roughly 20 residues length, with a consistent orientation of being perpendicular to the membrane surface. Since the mid-2000s the elucidation of many more intramembrane helix structures implied a far richer variety of transmembrane helices existed than previously thought, with a range of orientations and intra-membrane biophysical variations.

The language used to describe TMHs varies somewhat across the literature, primarily due to a changing understanding of TMH general structure and relevance to function over the last 15 years or so. There is a general composition of a TMH despite specific protein and membrane constraints (4). Figure 1 outlines the basic vocabulary used herein to describe different aspects of the TMH and membrane.

More recently, the insertion and formation of the unusually orientated TMHs and of the more traditional TMHs have been shown to be underpinned by complex thermodynamic equilibria (5). TMHs have been identified as regulators of protein quality control and trafficking mechanisms, shifting the idea away from TMHs broadly simply functioning as anchors (6). The story is not as simple as originally thought. There is a contingency in the field of biological membranes that despite progress over the last decade, there is a lack of information regarding their structure, assembly, and the behaviour of TMHs in the lipid bilayer; the native biological environment of TMHs (5, 7).

Properties that can be analysed by bioinformatics, the sequence complexity and hydrophobicity, of the TMH have been used to predict the role of the TMH as either functional or structural, and as a discrete cluster from other SCOP annotated helices (8)(Figure 2). Those findings demonstrated that sequence of the TMH holds valuable information regarding biological roles, and forms the basis of our interest in the link between the polarity of a helix and functional activity beyond structural anchorage.

A study by Baeza et al. from 2013 (9) looked at TMHs in 170 integral membrane proteins from a manually maintained database of experimentally confirmed TMPs; MPTopo (10). The group examined the distribution of residues along the TMHs. As expected, half of the natural amino acids are equally distributed along Transmembrane (TM) helices whereas aromatic, polar, and charged amino acids along with proline are biasedly near the flanks of the TM helices (9). Transitions between the different types of amino acid at the ends of the hydrophobic core occur in a more defined region on the cytosolic side than at the extra cytosolic face. This is probably reflecting the different lipid composition of both leaflets of biological membranes (9). A larger study using 1192 human and 1119 yeast predicted TMHs that were not structurally validated further explored the difference in TMH and leaflet structure by exploiting the evolutionarily conserved sequence differences between the TMH in the inner and outer

leaflets (4). TMHs from vertebrates and invertebrates were found to be reasonably similar compositionally. The differences in consensus TMH structure implies that there are general differences between the membranes of the golgi and ER. The abundance of serines in the region following the luminal end of golgi TMDs probably reflects the fact that this part of many golgi enzymes forms a flexible linker that tethers the catalytic domain to the membrane (4).

1.2 The positive inside rule.

Two publications by von Heijne coined the “Positive-Inside rule” demonstrated the practical value of positively charged residue sequence clustering in topology prediction of transmembrane helices in bacteria (11, 12). It was clearly defined and shown that positively charged residues more commonly were found on the “inside” of the cytoplasm rather than the periplasm. More recently still large scale sequence analysis of transmembrane helices from different organelle membrane surfaces in eukaryotic proteomes, show the clustering of positive charge being cytosolic (4, 9). As the idea of positive residues inside the cytoplasm emerged, so did the idea of negative residues working in concert with TMH orientation. It was shown that removing a single lysine residue reversed the topology of a model *Escherichia coli* protein, whereas much higher numbers of negatively charged residues are needed to reverse topology (13). One would also expect to see a skew in negatively charged distribution if a cooperation between oppositely charged residues orientated a TMH, however there is no conclusive evidence in the literature for an opposing negatively charged skew (4, 9, 14, 15). However, in *E.coli* negative residues do experience electrical pulling forces when travelling through the SecYEG translocon indicating that negative charges are biologically relevant (16).

Chapter 2

The “negative-not-inside” rule

Chapter 3

Conclusions

And here is the final chapter showing how clever you are

Bibliography

1. M. Almén, K. J. Nordström, R. Fredriksson, H. B. Schiöth, *BMC Biology* **7**, 50, ISSN: 1741-7007 (Jan. 2009).
2. G. von Heijne, *Nature Reviews Molecular Cell Biology* **7**, 909–918, ISSN: 1471-0072 (Dec. 2006).
3. A. Elofsson, G. von Heijne, *Annu Rev Biochem* **76**, 125–140, ISSN: 0066-4154 (Jan. 2007).
4. H. J. Sharpe, T. J. Stevens, S. Munro, *Cell* **142**, 158–169, ISSN: 00928674 (July 2010).
5. F. Cymer, G. Von Heijne, S. H. White, *Journal of Molecular Biology* **427**, 999–1022, ISSN: 10898638 (Sept. 2015).
6. T. Hessa *et al.*, *Nature* **475**, 394–397, ISSN: 0028-0836 (July 2011).
7. A. S. Ladokhin, *The Journal of Membrane Biology* **248**, 369–370, ISSN: 0022-2631 (June 2015).
8. W.-C. Wong, S. Maurer-Stroh, G. Schneider, F. Eisenhaber, *Nucleic acids research* **40**, W370–5, ISSN: 1362-4962 (July 2012).
9. C. Baeza-Delgado, M. A. Marti-Renom, I. Mingarro, *European Biophysics Journal* **42**, 199–207, ISSN: 01757571 (Mar. 2013).
10. S. Jayasinghe, K. Hristova, S. H. White, *Protein Science* **10**, 455–458, ISSN: 0961-8368 (2001).
11. G. von Heijne, *en*, *Nature* **341**, 456–458, ISSN: 0028-0836 (Oct. 1989).
12. A. Elofsson, G. von Heijne, *Annu Rev Biochem* **76**, 125–140, ISSN: 0066-4154 (May 2007).

13. I. Nilsson, G. von Heijne, *Cell* **62**, 1135–1141, ISSN: 00928674 (Sept. 1990).
14. E. Granseth, G. Von Heijne, A. Elofsson, *Journal of Molecular Biology* **346**, 377–385, ISSN: 00222836 (Feb. 2005).
15. J. Nilsson, B. Persson, G. von Heijne, *Proteins* **60**, 606–616, ISSN: 1097-0134 (2005).
16. N. Ismail, R. Hedman, M. Lindén, G. von Heijne, en, *Nature Structural & Molecular Biology* **22**, 145–149, ISSN: 1545-9993 (Feb. 2015).

Appendix A

Big tables

.....