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

Word count xxxxx

The University of Manchester

James Baker
Doctor of Philosophy
Investigating the Recognition and Interactions of Non-Polar α Helices in Biology
June 9, 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. This is in part due to the very features that make them so biologically influential: their hydrophobic transmembrane helices. What is missing in the current literature is a nuanced understanding of the complexities of the helix composition beyond a hydrophobic region of around 20 residues. 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.

Furthermore we provide an up-to-date dataset of potential Tail-Anchored proteins, a group of post-translationally inserted proteins.

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

Together my supervisory team have instilled in me an excitement of discovery. Furthermore they have taught me a deep value of inductive reasoning and inference over deduction.

Shout out to all the Biopython devs that saved my sanity!

Preamble

In the 1950s and 1960s, the field of biological philosophy was still emerging. David Hull writes on the matter in his 1969 article entitled *"What Philosophy of Biology is not"*:

> "Periodically through the history of biology, biologists have tried to do a little philosophy..." [1]

I think this accurately summarises my attempts at applying philosophy to science for my philosophical doctorate. I can't help but wish my thesis title was "The ins-and-outs of greasy peptides".

So long, and thanks for all the fish!

List of publications

Journal Articles

Posters

Baker, J. and Warwicker, J. A Bioinformatic Method to Identify Potential SNARE Proteins. *40th FEBS Congress* Late Breaker (2015)

Chapter 1

Introduction

1.1 The History of Membranes and Transmembrane Proteins in Science

The insertion and formation of the unusually orientated Transmembrane Helix (TMH)s and of the more traditional TMHs have been shown to be underpinned by complex thermodynamic equilibria [2]. TMHs have been identified as regulators of protein quality control and trafficking mechanisms, shifting the idea away from TMHs broadly simply functioning as anchors [3]. 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 [2, 4].

1.2 Biological Membrane Composition

When looking at the title of this thesis one might be surprised to first be reading not about protein helices, but membrane lipids. However, it is critical to understand that the two are inextricably linked, and often what we observe from the α helices reflect the properties of the much harder to study membranes. The lipid membranes influence the local structure, dynamics, and activity of proteins in the membrane in non-trivial ways [5–9]. It has been known for some time that the outer membranes

of Gram negative bacteria are asymentric in terms of lipid composition. The outer membranes contain lipopolysccharide, whilst the inner is a mixture of approximately 25 phospholipid types. Adding to the membrane asymmetry composition story, a thorough analysis of residue composition in yeast and human TMH regions revealed intra-membrane leaflet composition asymetry in the Endoplasmic Reticulum (ER), but not the Golgi [10]. Furthermore proteinlipid interactions have been shown to be determinants of membrane curvature [11], and undertake complex orientations and conformations to allow for hydrophobic mismatch [12].

Membrane bound proteins underpin almost every biological process directly, or indirectly, from photosynthesis to respiration. Integral Transmembrane Protein (TMP) are encoded by around 30% of the genes in the human genome which reflects their biological importance [13]. These proteins allow biochemical pathways that traverse the various biological membranes used in life. The compartmentalisation of cellular biochemistry is arguably one of the most significant events to have occurred in evolution, and is certainly one of the fundamental prerequisites for life [14]. The proteins that allow life to use this biochemical barrier are perhaps equally important. However we must understand both the membranes and the proteins to get an accurate model of how this complex biological system functions.

There is a rich variety of lipid molecules that make up the biological membranes. Generally speaking, these lipids have a polar headgroup, and a hydrophobic hydrocarbon "tail". Water entropicly drives the self association of the lipid molecules. In other words the bilayer forms from these phospholipid molecules due to the fierce dissociation between the polar water and the hydrophobic tails. This can be seen even in relatively early Molecular Dynamics (MD) simulations [15].

The majority of lipids in the membrane are phospholipids. These are composed of a glycerol molecule. Bonded to the glycerol molecule are two hydrophobic fatty acid tail groups, and a negatively-charged polar phosphate group. The polar phosphate group is modeified with an alcohol group.

1.3 α Helices in Membranes

1.3.1 Transmembrane Helix Sequence Composition

The relationship between the membrane and TMPs is underpinned by complex thermondynamic and electrostatic equilibria. Once inserted the protein doesn't leave the membrane as a result of the transmembrane helix being very hydrophobic. This hydrophobicity, and the hydrophobicity of the lipid tails means that they self associate. A better way of describing it is that they fiercely dissociate from the water. The overall ΔG for a transmembrane helix in the membrane is -12kcal mol1 [2]: the association of the helix in the membrane is typically spontaneous.

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 -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. Although the simple helices are broadly prevelant, 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 [16, 17] (Figure 1.1).

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 [20]. Those findings demonstrated that the 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.

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 [10].

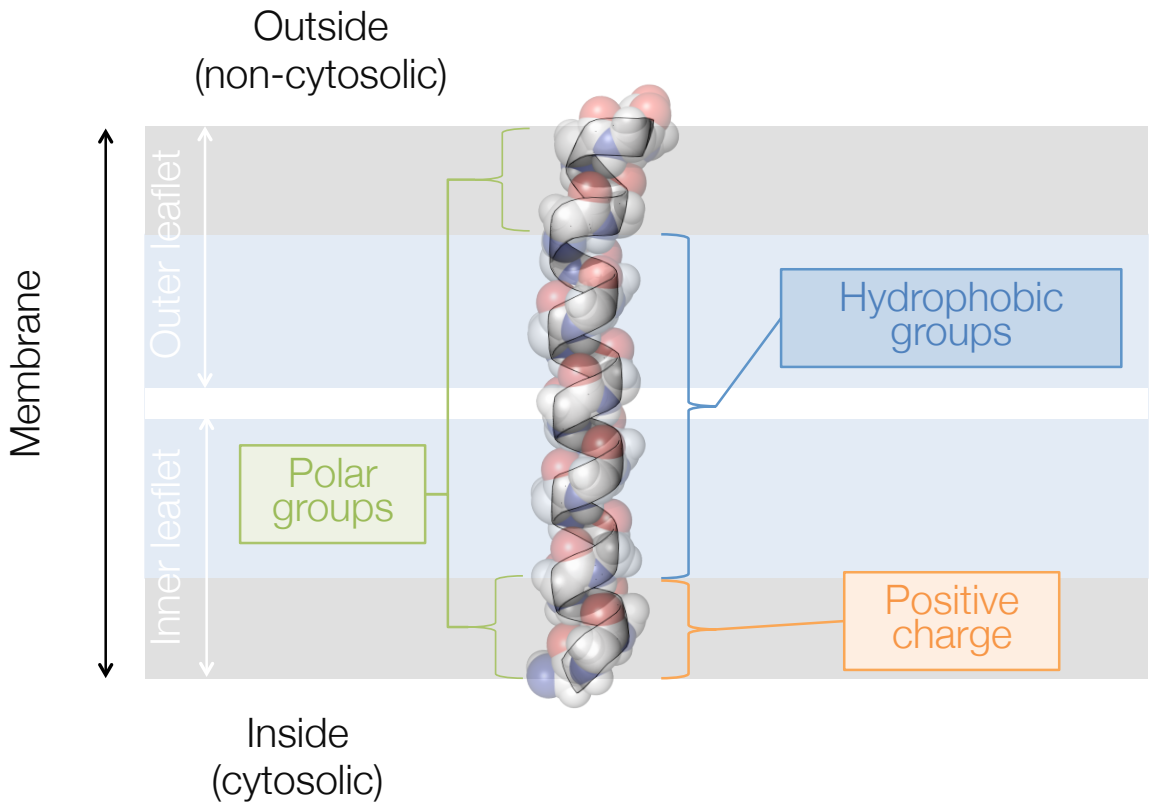

Figure 1.1: A cartoon showing the general components of the membrane and a typical TMH. The example used here for illustrative purposes is transmembrane region of tetherin (PDB 2LK9) [18]. Dark grey areas denote the area of lipid head groups. The residues in these areas are often described as flanking regions, and are often in contact with the aqueous interface of the membrane. The helix core is mostly composed of hydrophobic groups. More recently the hydrophobic group region has been associated with cell localisation and a broad range of biochemical functions [19, 20]. Although the regions labelled here generally hold true in terms of the statistical distribution of polar, non-polar, and charged groups, it is by no means absolute laws and many proteins break these "rules" [10, 21, 22]. Note that the definition of a TM α-helix is not entirely clear; how far the helix rises into the water-interface region to qualify as a TMH for example [16].

A study by Baeza-Delgado *et al.* from 2013 [21] looked at TMHs in 170 integral membrane proteins from a manually maintained database of experimentally confirmed TMPs; MPTopo [23]. 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 [21]. 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 [21]. 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 [10]. 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 lumenal 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 [10].

1.3.2 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 TMHs in bacteria [24, 25]. It was clearly defined and shown that positively charged residues more commonly were found on the "inside" of the cytoplasm rather than the periplasm of *E. coli*. More recently still large scale sequence analysis of TMHs from different organelle membrane surfaces in eukaryotic proteomes, show the clustering of positive charge being cytosolic [10, 21, 22].

1.3.3 The Aromatic Belt

1.3.4 Snorkeling

1.3.5 Biogenesis of Transmembrane Proteins

The "inside" was an imprecise term used to indirectly refer to the cytoplasmic space. To understand why the cytoplasm is the key part, one must recall how the membranes are thought to be synthesised.

1.4 Tail-Anchored Proteins

Tail anchored proteins are a topologically distinct class of intracellular proteins defined by their single carboxy-terminal transmembrane domain with a cytosolic facing amino-terminus. Tail anchored proteins are involved in a range of key cellular functions including protein translocation and apoptosis. Additionally, within the tail anchored class of proteins are a set of vesicle fusion proteins called Soluble N-Ethylmaleimide-Sensitive Factor Attachment Receptor (SNARE) proteins. There is biomedical interest in SNARE drug delivery mechanisms. SNAREs can fuse liposomes containing various drug payloads into the membrane. This study aims to identify SNARE proteins in eukaryotic proteomes by filtering through large datasets using automatically predicted TrEMBL consensus, and manually annotated SWISS-PROT transmembrane regions. The pipeline generates a list of singlepass proteins with a transmembrane domain close to the C terminal, that are not splice isoforms. A previous study predicted 411 tail anchor proteins [26].

Notably, known SNARE transmembrane helices are highly hydrophobic even compared to other tail anchored transmembrane helices.

1.5 Spontaneous membrane insertion

Signal anchored proteins, proteins that contain a single hydrophobic segment that serves as both a mitochondrial targeting signal and a membrane anchor, as well as tail anchored proteins have been shown to be able to spontaneously insert into the membrane [27–29].

It is postulated that there are electrostatic factors in the flanking regions that contribute to this spontaneous membrane insertion. Our experimental collaborators in Stephen Highs group are interested in a small group of tail anchored proteins that have very polar transmembrane domains and are capable of liposome membrane insertion without insertion machinery, also known as spontaneous insertion. They have found that chimeric synaptobrevin, one of the first identified SNARE proteins, is capable of spontaneous insertion if its tail anchor domain is replaced by the transmembrane domains belonging to a protein of known spontaneously inserting domains. Their studies have moved the focus of spontaneous insertion away from the loop regions and

onto the biophysicochemical factors of the TMH itself. The idea that SNARE proteins are modular, and capable of spontaneous insertion has significant implications for both biomedical application in liposome based drug delivery and can aid future research for testing complex biological molecular networks [30, 31].

1.6 Measures of Hydrophobicity

1.7 Aims of This Thesis.

1. Negative not inside rule
2. Glycosylphosphatidylinositol (GPI) project
3. SNARE and Tail Anchor (TA) project
4. Good and bad helices

Chapter 2

The "negative-not-inside" rule

2.1 Abstract

2.2 Introduction

As the idea of positive residues inside the cytoplasm emerged during the late 1980s, 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 [32]. 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 [10, 21, 22, 33, 34]. However, in *E. coli* negative residues do experience electrical pulling forces when travelling through the SecYEG translocon indicating that negative charges are biologically relevant [35].

2.3 Methods

2.3.1 Normalisation

$$c_r = \frac{(a_{K,r}+a_{R,r})-(a_{D,r}+a_{E,r})}{N}$$

$$p_{i,r} = \frac{a_{i,r}}{\max_r(a_r)}$$

$$q_{i,r} = \frac{100a_{i,r}}{a_i}$$

2.4 Results

2.4.1 Biophysicochemical differences in multi-pass and single-pass helices

Chapter 3

Tail-anchored protein discovery

3.1 Introduction

3.2 Methods

3.2.1 Filtering the Uniprot database

Steps carried out by Kalbfleisch *et al.* published in Traffic 2007 (8: 16871694) were recreated using up to date tools. The nonredundant human dataset of 145,715 proteins from SwissProt and TrEMBL [26, 36]. 2,478 singlepass proteins were programmatically extracted according to the TRANSMEM count from that list. Then TMDs not within 15AA of the C terminal were removed, resulting in 455 proteins. No splice isoforms were detected according to searching for NON_TER annotation. 195 proteins of the 411 predicted proteins from the previous study were successfully mapped using the Uniprot mapping tools [26, 36]. Duplicate IDs from the previously predicted tail anchored protein were removed from the set. The remaining dataset contained XXX proteins.

3.2.2 Calculating hydrophobicity

3.3 An up to date tail-anchor dataset

3.4 Potential tail-anchored SNARE protein discovery

3.5 Investigating biology of spontaneously inserting tail anchored proteins

Chapter 4

A novel GPI lipid anchor categorised

4.1 Abstract

4.2 Introduction

4.3 Methods

4.4 Results

Chapter 5

The good, the bad, and the ugly helices

5.1 Abstract

5.2 Introduction

5.3 Methods

5.4 Results

Chapter 6

Conclusions

Bibliography

1. Hull, D. L. What philosophy of biology is not. *Synthese* **20,** 157–184. ISSN: 00397857 (1969).

2. Cymer, F., Von Heijne, G. & White, S. H. Mechanisms of integral membrane protein insertion and folding. *Journal of Molecular Biology* **427,** 999–1022. ISSN: 10898638 (2015).

3. Hessa, T *et al.* Protein targeting and degradation are coupled for elimination of mislocalized proteins. *Nature* **475,** 394–397. ISSN: 0028-0836 (2011).

4. Ladokhin, A. S. Membrane Protein Folding & Lipid Interactions: Theory & Experiment. *The Journal of Membrane Biology* **248,** 369–370. ISSN: 0022-2631 (2015).

5. Bondar, A. N., del Val, C., Freites, J. A., Tobias, D. J. & White, S. H. Dynamics of SecY Translocons with Translocation-Defective Mutations. *Structure* **18,** 847–857. ISSN: 09692126 (2010).

6. Bondar, A. N., del Val, C. & White, S. H. Rhomboid Protease Dynamics and Lipid Interactions. *Structure* **17,** 395–405. ISSN: 09692126 (2009).

7. Jardon-Valadez, E., Bondar, A. N. & Tobias, D. J. Coupling of retinal, protein, and water dynamics in squid rhodopsin. *Biophysical Journal* **99,** 2200–2207. ISSN: 00063495 (2010).

8. Kalvodova, L. *et al.* Lipids as modulators of proteolytic activity of BACE: Involvement of cholesterol, glycosphingolipids, and anionic phospholipids in vitro. *Journal of Biological Chemistry* **280,** 36815–36823. ISSN: 00219258 (2005).

9. Urban, S. & Wolfe, M. S. Reconstitution of intramembrane proteolysis in vitro reveals that pure rhomboid is sufficient for catalysis and specificity. eng. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 1883–8. ISSN: 0027-8424 (2005).

10. Sharpe, H. J., Stevens, T. J. & Munro, S. A Comprehensive Comparison of Transmembrane Domains Reveals Organelle-Specific Properties. *Cell* **142,** 158–169. ISSN: 00928674 (2010).

11. Jensen, M. & Mouritsen, O. G. Lipids do influence protein function - The hydrophobic matching hypothesis revisited. *Biochimica et Biophysica Acta - Biomembranes* **1666,** 205–226. ISSN: 00052736 (2004).

12. De Planque, M. R. R. & Killian*, J. A. Proteinlipid interactions studied with designed transmembrane peptides: role of hydrophobic matching and interfacial anchoring (Review). en. *Molecular Membrane Biology* **20,** 271–284. ISSN: 0968-7688 (2003).

13. Almén, M., Nordström, K. J., Fredriksson, R. & Schiöth, H. B. Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biology* **7,** 50. ISSN: 1741-7007 (2009).

14. Koshland, D. E. The Seven Pillars of Life. en. *Science* **295,** 2215–2216. ISSN: 00368075 (2002).

15. Goetz, R & Lipowsky, R. Computer simulations of bilayer membranes: Self-assembly and interfacial tension. *Journal Of Chemical Physics* **108,** 7397–7409. ISSN: 00219606 (1998).

16. Von Heijne, G. Membrane-protein topology. *Nature Reviews Molecular Cell Biology* **7,** 909–918. ISSN: 1471-0072 (2006).

17. Elofsson, A & von Heijne, G. Membrane protein structure: prediction versus reality. *Annu Rev Biochem* **76,** 125–140. ISSN: 0066-4154 (2007).

18. Skasko, M *et al.* HIV-1 Vpu Protein Antagonizes Innate Restriction Factor BST-2 via Lipid-embedded Helix-Helix Interactions. *J.Biol.Chem.* **287,** 58–67 (2012).

19. Junne, T., Kocik, L. & Spiess, M. The hydrophobic core of the Sec61 translocon defines the hydrophobicity threshold for membrane integration. *Molecular biology of the cell* **21,** 1662–70. ISSN: 1939-4586 (2010).

20. Wong, W.-C., Maurer-Stroh, S., Schneider, G. & Eisenhaber, F. Transmembrane helix: simple or complex. *Nucleic acids research* **40,** W370–5. ISSN: 1362-4962 (2012).

21. Baeza-Delgado, C., Marti-Renom, M. A. & Mingarro, I. Structure-based statistical analysis of transmembrane helices. *European Biophysics Journal* **42,** 199–207. ISSN: 01757571 (2013).

22. Pogozheva, I. D., Tristram-Nagle, S., Mosberg, H. I. & Lomize, A. L. Structural adaptations of proteins to different biological membranes. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1828,** 2592–2608. ISSN: 00052736 (2013).

23. Jayasinghe, S, Hristova, K & White, S. H. MPtopo: A database of membrane protein topology. *Protein Science* **10,** 455–458. ISSN: 0961-8368 (2001).

24. Von Heijne, G. Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. en. *Nature* **341,** 456–458. ISSN: 0028-0836 (1989).

25. Elofsson, A & von Heijne, G. Membrane protein structure: prediction versus reality. *Annu Rev Biochem* **76,** 125–140. ISSN: 0066-4154 (2007).

26. Kalbfleisch, T., Cambon, A. & Wattenberg, B. W. A bioinformatics approach to identifying tail-anchored proteins in the human genome. *Traffic* **8,** 1687–1694. ISSN: 13989219 (2007).

27. Merklinger, E. *et al.* Membrane integration of a mitochondrial signal-anchored protein does not require additional proteinaceous factors. *Biochemical Journal* **442,** 381–389. ISSN: 0264-6021 (2012).

28. Lan, L, Isenmann, S & Wattenberg, B. W. Targeting and insertion of C-terminally anchored proteins to the mitochondrial outer membrane is specific and saturable but does not strictly require ATP or molecular chaperones. *The Biochemical journal* **349,** 611–621. ISSN: 02646021 (2000).

29. Colombo, S. F., Longhi, R. & Borgese, N. The role of cytosolic proteins in the insertion of tail-anchored proteins into phospholipid bilayers. *Journal of cell science* **122,** 2383–92. ISSN: 0021-9533 (2009).

30. Allen, T. M. & Cullis, P. R. Liposomal drug delivery systems: From concept to clinical applications. *Advanced Drug Delivery Reviews* **65,** 36–48. ISSN: 0169409X (2013).

31. Nordlund, G., Brzezinski, P. & von Ballmoos, C. SNARE-fusion mediated insertion of membrane proteins into native and artificial membranes. *Nature Communications* **5,** 4303. ISSN: 2041-1723 (2014).

32. Nilsson, I. & von Heijne, G. Fine-tuning the topology of a polytopic membrane protein: Role of positively and negatively charged amino acids. *Cell* **62,** 1135–1141. ISSN: 00928674 (1990).

33. Granseth, E., Von Heijne, G. & Elofsson, A. A study of the membrane-water interface region of membrane proteins. *Journal of Molecular Biology* **346,** 377–385. ISSN: 00222836 (2005).

34. Nilsson, J., Persson, B. & von Heijne, G. Comparative analysis of amino acid distributions in integral membrane proteins from 107 genomes. *Proteins* **60,** 606–616. ISSN: 1097-0134 (2005).

35. Ismail, N., Hedman, R., Lindén, M. & von Heijne, G. Charge-driven dynamics of nascent-chain movement through the SecYEG translocon. en. *Nature structural & molecular biology* **22,** 145–9. ISSN: 1545-9985 (2015).

36. Bateman, A. *et al.* UniProt: A hub for protein information. *Nucleic Acids Research* **43,** D204–D212. ISSN: 13624962 (2015).