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
September 1, 2016

Non–polar helices figure prominently in structural biology, from the first protein structure (myoglobin) through Transmembrane (TM) segments, to current work on recognition of protein trafficking and quality control. TM 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 non–polar Transmembrane Helix (TMH) regions. 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. Studies in Frank Eisenhabers group at the A*STAR Bioinformatics Institute have identified types of transmembrane helix, simple are characterised by their hydrophobicity, and complex type by the addition of structural and ancestral features that mediate a role beyond basic membrane insertion 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.
- An up-to-date set of potential Tail Anchor (TA) proteins (a group of post-translationally inserted proteins) is rebuilt.
- A novel Glycosylphosphatidylinositol (GPI) lipid anchor is characterised.
- The hydrophobicity–complexity continuum is investigated in relation to function and recognition.

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

Acknowledgements

I would like to thank all members of both the Eisenhaber research group, as well as the the Curtis and Warwicker research group for discussion, but in particular Jim Warwicker, Frank Eisenhaber, Birgit Eisenhaber, and Wing-Cheong Wong for supervision and guidance during my research. I would also like to thank The University of Manchester and the A*STAR Singapore Bioinformatics Institute for funding the project. Furthermore I would like to extend my gratitude to the research group of Professor Stephen High.

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 Transmembrane Protein Problem

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 [1]. TMHs have been identified as regulators of protein quality control and trafficking mechanisms, shifting the idea away from TMHs broadly simply functioning as anchors [2]. 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 still lack of information regarding their structure, assembly, and the behaviour of TMHs in the lipid bilayer; the native biological environment of TMHs [3].

1.2 Biological Membrane Composition

1.2.1 Lipids of the Membrane

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 [4]. The proteins that allow life to use this biochemical barrier are perhaps equally important. Together, the lipid bilayer and proteins therein allow complex biochemical systems that facillitate life to exist.

It is critical to understand that the lipid bilayer and the transmembrane α helices

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 [5–12].

There is a rich variety of lipid molecules that make up the biological membranes. The majority of lipids in higher organism membranes are phospholipids, sphingolipids, and sterols. 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. 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. Furthermore the bilayer maximises van der Waals interactions between the closely-packed hydrocarbon chains, which contributes to the stability of the bilayer. This can be seen even in relatively early Molecular Dynamics (MD) simulations [13].

1.2.2 Differenes in Membrane Compositions

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 [14]. Furthermore proteinlipid interactions have been shown to be determinants of membrane curvature [11], and undertake complex orientations and conformations to allow for hydrophobic mismatch [15].

1.2.3 Membrane Potential

Nernst:

$$E_m = \frac{RT}{F} \times \ln \frac{c_{out}}{c_{in}}$$

Goldman:

$$E_m = \frac{RT}{F} \times \ln\left(\frac{p_{K^+}.[K^+]_{out} + p_{Na^+}.[Na^+]_{out} + p_{Cl^-}.[Cl^-]_{in}}{p_{K^+}.[K^+]_{in} + p_{Na^+}.[Na^+]_{in} + p_{Cl^-}.[Cl^-]_{out}}\right)$$

Where Em is the membrane potential, z is the ion charge, $[i]$ is the ion concentration p_i is the relative membrane permeability for the actual ion

1.3 α Helices in Membranes

1.3.1 The Importance of Transmembrane Proteins

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 [16]. These proteins allow biochemical pathways that traverse the various biological membranes used in life.

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 -12kcalmol^{-1} [1]; the association of the helix in the membrane is typically spontaneous.

1.3.2 Transmembrane Helix Sequence Composition

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 [19]. 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.

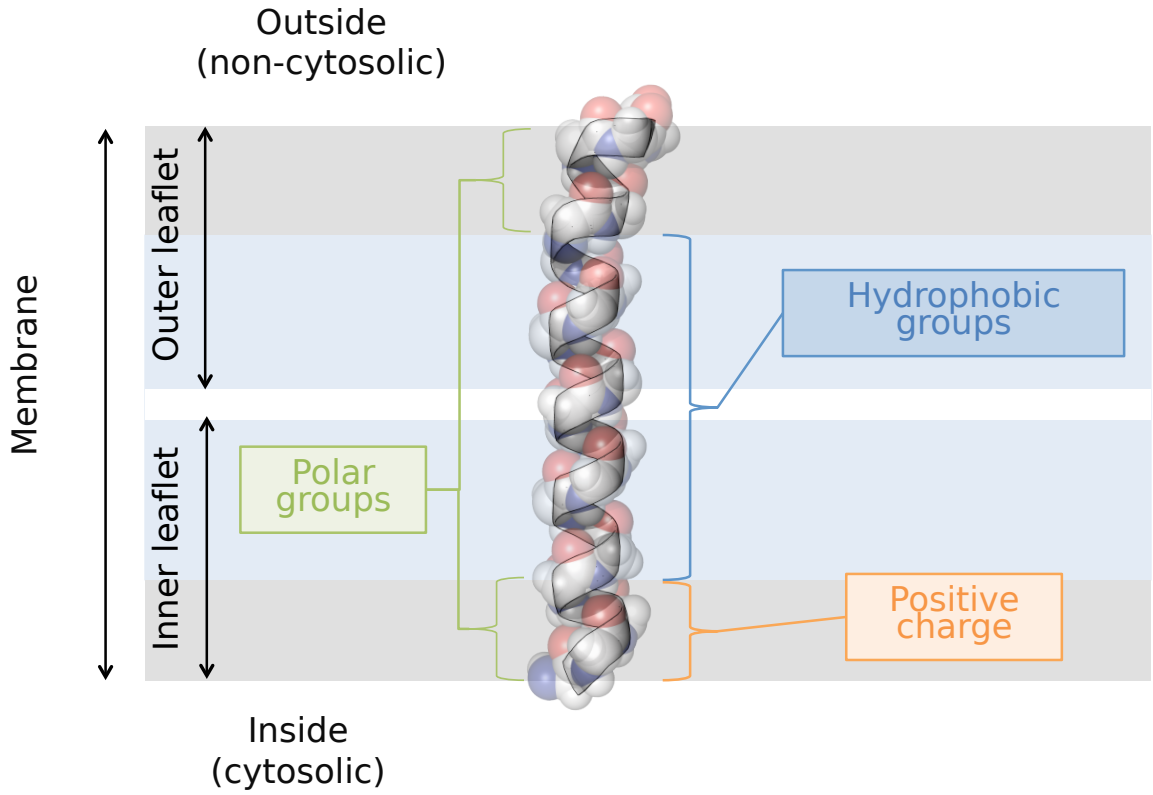

Figure 1.1: A cartoon showing the general components of the membrane and a typical TMH. The example used here for illustrative purposes is transmembrane region of tetherin (PDB 2LK9) [17]. Dark grey areas denote the area of lipid head groups. The residues found in these areas are often described as flanking regions, and are often in contact with the aqueous interface of the membrane. The helix core is mostly composed of hydrophobic residues. More recently the hydrophobic group region has been associated with cell localisation and a broad range of biochemical functions [18, 19]. 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" [14, 20, 21]. 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 [22].

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 [14].

A study by Baeza-Delgado *et al.* from 2013 [20] 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 [20]. It has been noted that transitions between the polar and non-polar groups at the ends of the hydrophobic core occur in a more defined edge on the cytoplasmic side than at the extracytoplasmic face when counting from the middle of the helix outwards [20]. This is probably reflecting the different lipid composition of both leaflets of biological membranes [20]. A larger previous 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 [14]. 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 [14].

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 [14, 20, 21].

The Aromatic Belt

Tyrosine and tryptophan residues commonly are found at the interfacial boundaries of the TMH and this feature is called the "aromatic belt" [14, 20, 26–28]. Not all aromatic residues are not found in the aromatic belt; phenylalanine has no particular preference for this region [27, 29]. However it still remains unclear if this is to do with anchorage or translocon recognition [20].

Snorkeling

Broadly speaking, transmembrane helices are non-polar. However some contain polar and charged residues in the helix itself. Whilst this might seem thermodynamically unstable at first glance, a molecular dynamic feature called the "snorkel" effect explains in part how this is possible [30, 31]. Simply put, the snorkelling effect involves the long flexible side chain of leucine reaching the water interface region to interact with the polar headgroups of the bilayer even when the α helix backbone is pulled into the hydrophobic layer [32]. This has also been suggested to allow helices to adapt to varying thicknesses of the membrane [33].

1.4 Biogenesis of Transmembrane Proteins

1.4.1 Translocation

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 synthesised and localised throughout the cell.

1.4.2 Tail-Anchored Proteins Post Translationally Insert

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.

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

1.4.3 Translocon Independent 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 indepently from the translocon [34–36].

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 [37, 38].

1.5 Choice of Hydrophobicity Values

1.5.1 An Overview of the Different Scales

Throughout this thesis several scales are used to evaluate and estimate hydrophobic values of peptides. All the scales aim for quantifying the hydrophobic values of each residue. There are several key differences in their methodology, assumptions, and aims. Crucially this results in slightly different scores for some residues. Because of this, it's preferable, and typical amongst the literature, to use several scales to verify the patterns observable in one scale. Notably, one of the classic scale, Kyte & Doolittle Hydropathy Scale shows striking similarity to the modern Hessa's Biological Hydrophobicity Scale, and that generally the "better" scales count proline as

hydrophilic, and focus on helix recognition rather than amino acid analogues [39]. Ultimately, all the scales are attempting to allow estimation of ΔG_{whf}; the free energy of a folded helix (f) from the water (w) into the membrane core (h). This free energy measurement is regarded as being currently experimentally inaccessible [1].

Kyte & Doolittle Hydropathy Scale

A scale based on the water–vapour transfer free energy and the interior-exterior distribution of individual amino acids [40].

Hessa's Biological Hydrophobicity Scale

This is arguably the most biologically relevent scale [39]. The scale is based on an experimental method where the free energy exchange during recognition of designed polypeptide helices by the endoplasmic reticulum Sec61 translocon occured [26]. These measurements were then used to calculate a biological hydrophobicity scale.

White and Wimley Octanol – Interface Whole Residue Scale

This scale is calculated from two other scales; the octanol scale, and the interface scale [41]. This scale is fundamentally based on the partitioning of host-guest pentapeptides (acetyl-WL-X-LL-OH) and another set of peptides (AcWLm) between water and octanol, as well as water to Palmitoyloleoylphosphatidylcholine (POPC) .

The Eisenberg Hydrophobic Moment Consensus Scale

The Eisenberg scale is a consensus scale based on the earlier scales from Tanford [42], Wolfenden [43], Chothia [44], Janin [45], and the von Heijne scale [46]. The scales are normalized according to serine [47]. The automatic TRANSMEM annotation currently used in Uniprot is according to TMHMM [48], Memsat [49], Phobius [50] and the hydrophobic moment plot method of Eisenberg and coworkers [47].

1.6 A Brief History of Transmembrane Proteins in Science

1.6.1 Earliest Evidences of Compartmentalisation

1.6.2 Early Models of the Bilayer

1.6.3 The Rise of Crystallography

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 [22, 51] (Figure 1.1). Over the last decade, Nanodiscs have been routinely used to much more easily obtain crystal structures. Nanodiscs overcome some of the major challenges caused by the hydrophobic helices, and a more faithful representation of the biological membranes than alternative model membranes like liposomes [52].

1.7 Aims of This Thesis.

1. Negative not inside rule
2. GPI project
3. SNARE and 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 [53]. 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 [14, 20, 21, 27, 28]. However, in *E. coli* negative residues do experience electrical pulling forces when travelling through the SecYEG translocon indicating that negative charges are biologically relevant [54].

2.3 Methods

2.3.1 Normalisation

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

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

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

2.4 Results

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

Chapter 3

Tail-anchored protein discovery

3.1 Abstract

3.2 Introduction

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 [55].

3.3 Methods

The original list UniProt protein database was queried for records containing "TRANSMEM" annotation on June 15, 2016, totaling 75826 records from swissprot, and 12322000 records from TrEMBL.

Expression

3.3.1 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 [55, 56]. 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 [56]. Duplicate IDs from the previously predicted tail anchored protein were removed from the set. The remaining dataset contained XXX proteins.

3.3.2 Calculating Hydrophobicity

3.3.3 Calculating Sequence Complexity

3.4 Results

3.4.1 An Up To Date Tail-Anchor Dataset

3.4.2 Potential Tail-Anchored SNARE Protein Discovery

3.4.3 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

6.1 Outlook

6.1.1 The hydrophobicity–sequence complexity continuum

We hypothesise that the hydrophobicity–sequence complexity continuum contains nuanced codes for different functions and that such differentiation of sequence and structural properties will allow assignment to these varying functions. Additionally, we suggest probing functional classification of yet uncharacterised membrane proteins by similarities of combinations of complex TM sets to well studied membrane proteins and finding those classes of TM proteins where this principle is most directly applicable.

Bibliography

1. 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).

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

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

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

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. Jard??n-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. White, S. H., Ladokhin, A. S., Jayasinghe, S. & Hristova, K. How Membranes Shape Protein Structure. *Journal of Biological Chemistry* **276,** 32395–32398. ISSN: 00219258 (2001).

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. H??nin, J., Salari, R., Murlidaran, S. & Brannigan, G. A predicted binding site for cholesterol on the GABAA receptor. *Biophysical Journal* **106,** 1938–1949. ISSN: 15420086 (2014).

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

14. 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).

15. De Planque, M. R. R. & Killian*, J. A. Protein\lipid 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).

16. 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).

17. 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).

18. 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).

19. 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).

20. 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).

21. 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).

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

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. Hessa, T. *et al.* Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433,** 377–81. ISSN: 1476-4687 (2005).

27. 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).

28. 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).

29. Braun, P. & Von Heijne, G. The aromatic residues Trp and phe have different effects on the positioning of a transmembrane helix in the microsomal membrane. *Biochemistry* **38,** 9778–9782. ISSN: 00062960 (1999).

30. Chamberlain, A. K., Lee, Y., Kim, S. & Bowie, J. U. Snorkeling preferences foster an amino acid composition bias in transmembrane helices. *Journal of Molecular Biology* **339,** 471–479. ISSN: 00222836 (2004).

31. Strandberg, E. & Killian, J. A. Snorkeling of lysine side chains in transmembrane helices: How easy can it get? *FEBS Letters* **544,** 69–73. ISSN: 00145793 (2003).

32. Krishnakumar, S. S. & London, E. Effect of Sequence Hydrophobicity and Bilayer Width upon the Minimum Length Required for the Formation of Transmembrane Helices in Membranes. *Journal of Molecular Biology* **374,** 671–687. ISSN: 00222836 (2007).

33. Kandasamy, S. K. & Larson, R. G. Molecular dynamics simulations of model trans-membrane peptides in lipid bilayers: a systematic investigation of hydrophobic mismatch. *Biophysical journal* **90,** 2326–2343. ISSN: 00063495 (2006).

34. 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).

35. 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).

36. 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).

37. 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).

38. 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).

39. Peters, C. & Elofsson, A. Why is the biological hydrophobicity scale more accurate than earlier experimental hydrophobicity scales? *Proteins: Structure, Function and Bioinformatics* **82,** 2190–2198. ISSN: 10970134 (2014).

40. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **157,** 105–132. ISSN: 00222836 (1982).

41. White, S. H. & Wimley, W. C. MEMBRANE PROTEIN FOLDING AND STABILITY : Physical Principles. *Annual Review of Biophysics and Biomolecular Structure* **28,** 319–365. ISSN: 1056-8700 (1999).

42. Nozaki, Y. & Tanford, C. The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. Establishment of a hydrophobicity scale. *Journal of Biological Chemistry* **246,** 2211–2217. ISSN: 00219258 (1971).

43. Rose, G. D. & Wolfenden, R. Hydrogen Bonding, Hydrophobicity, Packing, and Protein Folding. *Annual Review of Biophysics and Biomolecular Structure* **22,** 381–415. ISSN: 1056-8700 (1993).

44. Chothia, C. The nature of the accessible and buried surfaces in proteins. *Journal of Molecular Biology* **105,** 1–12. ISSN: 00222836 (1976).

45. Janin, J. Surface and inside volumes in globular proteins. *Nature* **277,** 491–492. ISSN: 0028-0836 (1979).

46. Von Heijne, G. & Blomberg, C. Trans-membrane Translocation of Proteins. The Direct Transfer Model. *European Journal of Biochemistry* **97,** 175–181. ISSN: 0014-2956 (1979).

47. Eisenberg, D. Three-dimensional structure of membrane and surface proteins. *Annual review of biochemistry* **53,** 595–623. ISSN: 00664154 (1984).

48. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305,** 567–580. ISSN: 00222836 (2001).

49. Jones, D. T. Improving the accuracy of transmembrane protein topology prediction using evolutionary information. *Bioinformatics* **23,** 538–544. ISSN: 13674803 (2007).

50. Käll, L., Krogh, A. & Sonnhammer, E. L. L. A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology* **338,** 1027–1036. ISSN: 00222836 (2004).

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

52. Borch, J. & Hamann, T. The nanodisc: A novel tool for membrane protein studies. *Biological Chemistry* **390,** 805–814. ISSN: 14316730 (2009).

53. 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).

54. 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).

55. 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).

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