Current Topics in Peptide & Protein Research
Vol. 23, 2022

Original Article

Functional and clinical importance of charged amino acids in protein transmembrane domains

Jordan Duong[1,*], Nicholas Manolios[2,3] and Erdahl Teber[4]

[1]Westmead Clinical School, University of Sydney, Westmead, NSW, Australia;
[2]Faculty of Medicine and Health, University of Sydney, Sydney, NSW, Australia;
[3]Department of Rheumatology, Westmead Hospital, Westmead, NSW, Australia;
[4]Bioinformatics Unit, Children's Medical Research Institute, Westmead, NSW, Australia.

ABSTRACT

Transmembrane (TM) proteins perform many essential biological roles. The presence of charged amino acids within the TM domain of human proteins has been shown to be functionally relevant. The objective of this research was to characterise this functional importance and gain a greater understanding of how electrostatic interactions within TM domains can influence protein structure, interactions, and overall functionality. Protein sequence and topological information was extracted from UniProtKB/Swiss-Prot to obtain a list of all human proteins containing charged TM amino acids. A heat map was generated to visualise the distribution of these residues along the TM. Information relating to rare genetic variants known to occur in humans was derived from ClinVar to evaluate the clinical significance of TM charge disturbances. Of 5,177 human TM proteins, 3,096 (59.80%) contained at least one charged amino acid within their TM sequence. There were 20,911 unique TM domains among these proteins, 9,678 (46.28%) of which contained a charged residue. These residues were distributed relatively sparsely in TM sequences compared to their hydrophobic counterparts. 34 proteins were identified sharing a total of 84 pathogenic point mutations involving a TM charged residue. Most TM proteins with reported variants were transport proteins, with the most clinically significant variants occurring in the cystic fibrosis transmembrane conductance regulator. Despite scarce distribution, there are many cases of TM charges playing key functional roles. This has implications for our understanding of health and disease processes, providing a basis for therapeutic targeting of these residues.

KEYWORDS: transmembrane protein, charged amino acid, electrostatic interaction, amino acid distribution, bioinformatics, peptidomimetics.

INTRODUCTION

Approximately 30% of all proteins encoded by the human genome are membrane proteins [1]. Integral membrane proteins are those that are permanently embedded within the phospholipid bilayer, and transmembrane (TM) proteins are an important subset that have a segment spanning the entire length of the cell membrane. Beyond serving as hydrophobic anchors of these proteins to the membrane, TM segments are widely recognised to play key roles in protein structure determination and, by extension, function. They achieve this through functioning as ligand receptors, transporters, and signal transducers to facilitate intracellular communication. Their biological significance is reflected by the fact that many currently approved drugs target TM proteins, with G protein-coupled receptors being the most highly represented followed by ion channels [2, 3].

*Corresponding author: jordan.duong@gmail.com

Charged amino acid residues are known to be less prevalent in TM domains due to the prevailing hydrophobicity of this environment [4]. Yet, there are numerous examples of intramembrane charges that are necessary for normal biological function. The T-cell receptor (TCR) is a prime example of electrostatic protein-protein interactions within TM domains, where charged amino acids are fundamental to the structure-function relationship that underlies T-cell assembly and activation. Charged residues in the TM domains facilitate interactions between protein subunits, which subsequently lead to surface expression and appropriate functionality of the TCR complex [5, 6]. Conversely, genetic mutations that introduce unfavourable charges into TM regions can result in aberrant peptide folding and oligomerisation [7]. From a molecular level, these changes can translate into disease states as a consequence of impaired protein trafficking, reduced levels of functional protein or gain of toxic function [8]. This knowledge provides an exciting foundation for the development of pharmacological interventions to modulate electrostatic interactions within and between TM sequences. Peptidomimetics are an evolving class of therapeutics which mimic biological peptides to favourably alter molecular function, either by stabilising or disrupting protein interactions [9]. In the case of the TCR, this approach could be utilised to downregulate inflammation in the setting of autoimmune diseases such as rheumatoid arthritis, systemic lupus erythematosus and multiple sclerosis [10].

The aim of this study was to examine the distribution of charged amino acids in TM domains and investigate how electrostatic interactions between these residues can influence protein structure, interactions and overall functionality to meaningfully impact on human health and disease processes.

MATERIALS AND METHODS

Identifying all proteins in nature that contain a charged amino acid within their TM domains

UniProt Knowledgebase (UniProtKB)/Swiss-Prot [11], a manually annotated database of protein sequences and functional information, was downloaded as a text file. Swiss-Prot is a non-redundant database in the sense that all products encoded by one gene – including alternative splicing isoforms, polymorphic variants, and other sequence conflicts – are represented by a single entry. Sequence analysis was performed based on the canonical sequence of each protein, chosen above other naturally occurring isoforms by way of being the most prevalent or the most similar sequence among orthologous species. This approach was necessary because all positional information, including TM domain positions, relates to the canonical sequence. From the raw data set, all proteins with annotated TM domains were extracted to obtain information of interest including entry ID, organism, topological information (*e.g.* delineation between extracellular, TM and cytoplasmic domains) and complete amino acid sequence. Secondly, any TM domain that contained a charged amino acid was selected. Arginine and lysine were defined as positively charged, while aspartate and glutamate were defined as negatively charged. For preliminary analysis and to determine basic trends in the prevalence of charged residues in TM domains, data from several organisms were extracted for comparison. In addition to humans, TM proteins in the mouse, rat, bovine, nematode, and chimpanzee were examined. For further detailed analysis of the distribution of charged residues in TM domains and to examine the clinical significance of these charges, only human proteins were considered.

Generating a heat map to visualise trends in TM amino acid distribution

TM domain lengths of 21 amino acids were included for this component of the analysis, as these constituted the vast majority (73.45%) of all database-defined TM domains. From a practical point of view, an identical sequence length also allowed for precise sequence alignment. Prior to alignment, TM sequences were oriented to ensure that they travelled in the same direction from cytoplasmic to extracellular domain. With reference to the SwissProt topology annotations, the sequences of TM domains with a preceding cytoplasmic domain were unaltered, while those with a preceding extracellular domain had their sequences reversed. This was needed to account for the influence of different subcellular environments on the predilection for certain amino acids throughout the cell membrane. Once aligned, the percentage frequency of every amino acid at each position throughout the cell membrane was calculated, and a heat map was

generated to visualise their relative distribution [12]. Cytoplasmic and extracellular 'flank' sequences of 5 amino acids on either side of the database-defined TM domains were also included because the positioning of TM domains is a dynamic process such that the amino acid representing the precise boundary point is not always fixed. These neighbouring residues may also influence membrane insertion, orientation, structure, and interactions [13]. Amino acids were presented in the order of decreasing hydrophobicity according to the Kyte-Doolittle scale, which is widely used to detect hydrophobic regions within protein sequences and hence predict the presence of TM domains [14].

Generating a list of rare genetic variants affecting charged residues in TM domains

A data set compiling all human genetic variations with interpretations of their clinical significance and associated disease phenotypes was downloaded from ClinVar [15], a database housed by the National Centre for Biotechnology Information (NCBI). Variations were only included if they were clinically significant (*i.e.* interpreted as either pathogenic or likely pathogenic by submitters to ClinVar) and had an adequate level of review supporting this assertion of clinical significance. Only those variations with a review status rating of two stars (criteria provided, multiple submitters, no conflicts), three stars (reviewed by expert panel) or four stars (practice guideline) were included. The web-based BioMart tool was used to 'translate' different gene IDs, allowing entries corresponding to identical genes across the ClinVar and UniProt databases to be merged [16]. Because representative gene transcripts are not always matched across the two databases, ClinVar entries that did not have an identical gene transcript version in the UniProt data set were omitted. While this was a constraining factor, it was necessary to frame the analysis around the UniProt data set as this is where sequence and TM positional information is derived from. UniProt topological annotation was used to extract only those genetic variations known to occur in TM proteins, before further filtering to obtain a subset of genes with documented mutations involving the gain or loss of charged residues within their TM domains.

RESULTS

Basic trends and prevalence of charged residues in TM domains

There were 5,177 human proteins with annotated TM domains on Swiss-Prot, 3,096 (59.80%) of which contained at least one charged amino acid in any one of their TM domains. This proportion was comparable across the different organisms, with larger differences likely due to limited representation on Swiss-Prot (Figure 1). Of these proteins, 2,364 (45.66%) contained at least one positive amino acid, while 2,582 (49.87%) contained at least one negative amino acid. In general, negatively charged residues were slightly more prevalent than positively charged residues, a finding consistent across different organisms even as the sample size decreased (Figure 2). There was clearly a considerable overlap, with many TM proteins containing both positive and negative charges in their TM domains.

Among the 5,177 identified TM proteins in humans, there were a total of 20,911 individual TM domains. Of these, 9,678 (46.28%) contained a charged amino acid. 4,871 (23.3%) contained at least one positive amino acid, while 6,413 (30.67%) contained at least one negative amino acid. Similarly, there was considerable overlap, with many of these individual domains containing both positive and negative charges. Most of these TM domains (6,733) contained just a single charged residue (median = 1). The number of TM domains containing two charges was markedly reduced (2,131). As expected, there were progressively fewer occurrences as the number of charges increased. According to Swiss-Prot, however, there were single TM domains containing 9, 10, 11 and 13 charged residues which were all significant outliers. These TM domains were identified as belonging to potassium voltage-gated channel subfamily H member 4, microsomal glutathione S-transferase 1, cytochrome c oxidase subunit 6C and sodium-dependent glucose transporter 1, respectively. Notably, the TM domains containing 10 and 11 charged residues had significantly longer database-defined TM sequences than average (34 and 41 residues, respectively).

Heat map visualising amino acid distribution in TM domains spanning 21 residues

The 20,911 individual TM domains had lengths ranging from 6 to 46 amino acids. This was normally

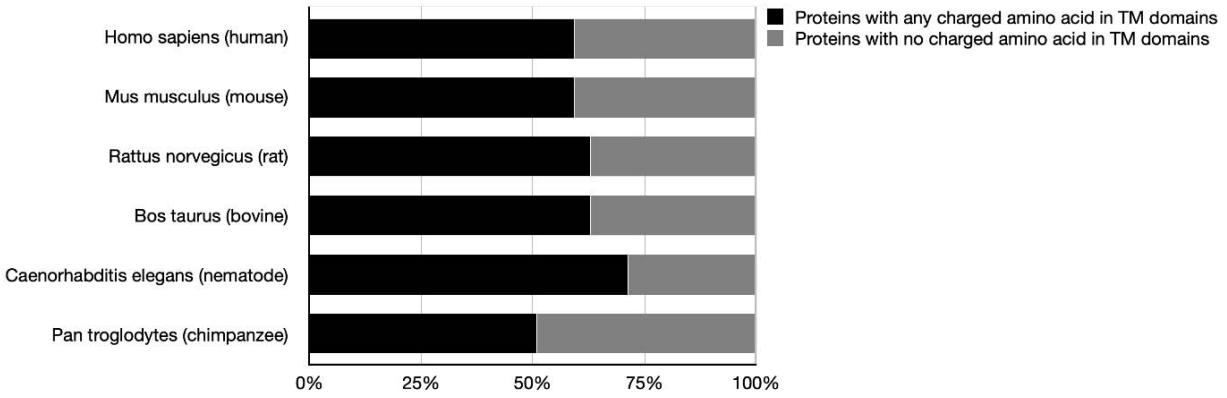

Figure 1. Percentage bar chart depicting the proportion of TM proteins that contain at least one charged amino acid in one of their TM domains. Data for several organisms are shown for comparison with the human data set.

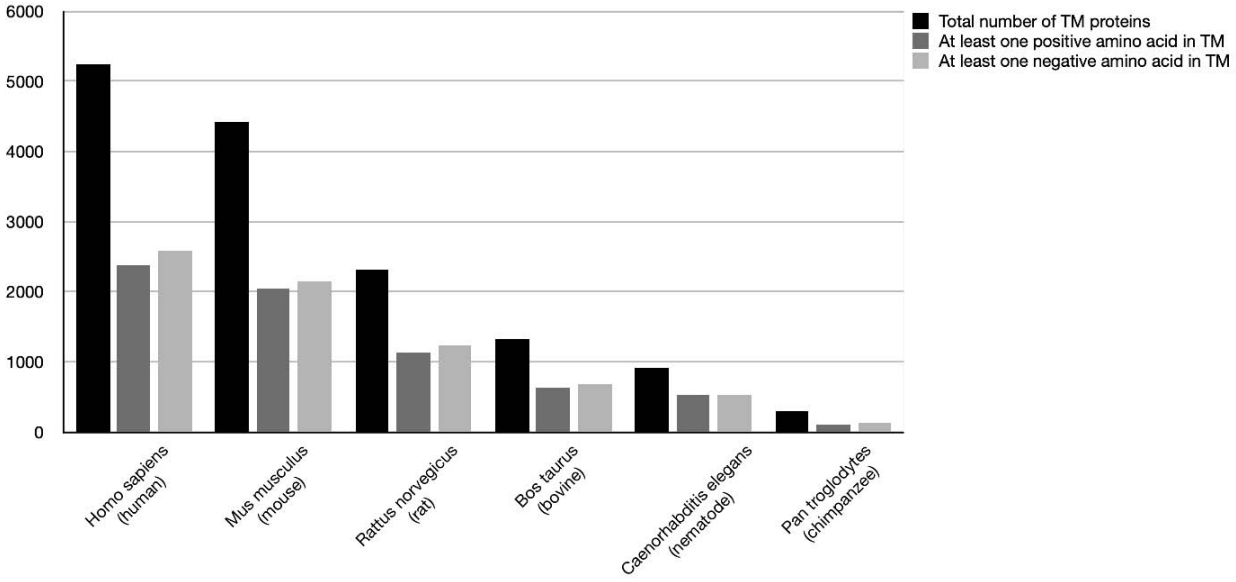

Figure 2. Bar chart of the total number of TM proteins and the number of these that contain either at least one positively or negatively charged amino acid within a TM domain. Data for several organisms are shown for comparison with the human data set.

distributed with a peak at 21 amino acids constituting 15,359 (73.45%) of all TM domains. These 15,359 TM domains were included in the representative heat map of amino acid distribution (Figure 3). In general, the more hydrophobic residues appear with greater frequency, while there is a progressive decline in frequency as hydrophobicity decreases. Leucine (L) was found to be the most frequent amino acid appearing in TM sequences by a large margin (average frequency of 20.54%). Conversely, the hydrophilic residues including all four charged amino acids were distributed quite sparsely. The positively charged residues lysine (K) and arginine (R) had average frequencies of 0.55% and 0.52%, respectively, and the negatively charged residues glutamate (E) and aspartate (D) had average frequencies of 0.83% and 0.74%, respectively. There was a relative clustering of charged amino acids in isolated flanking domains. This was particularly noticeable for the positively charged residues in cytoplasmic flanks (combined average frequency of 8.44% in cytoplasmic flanks compared

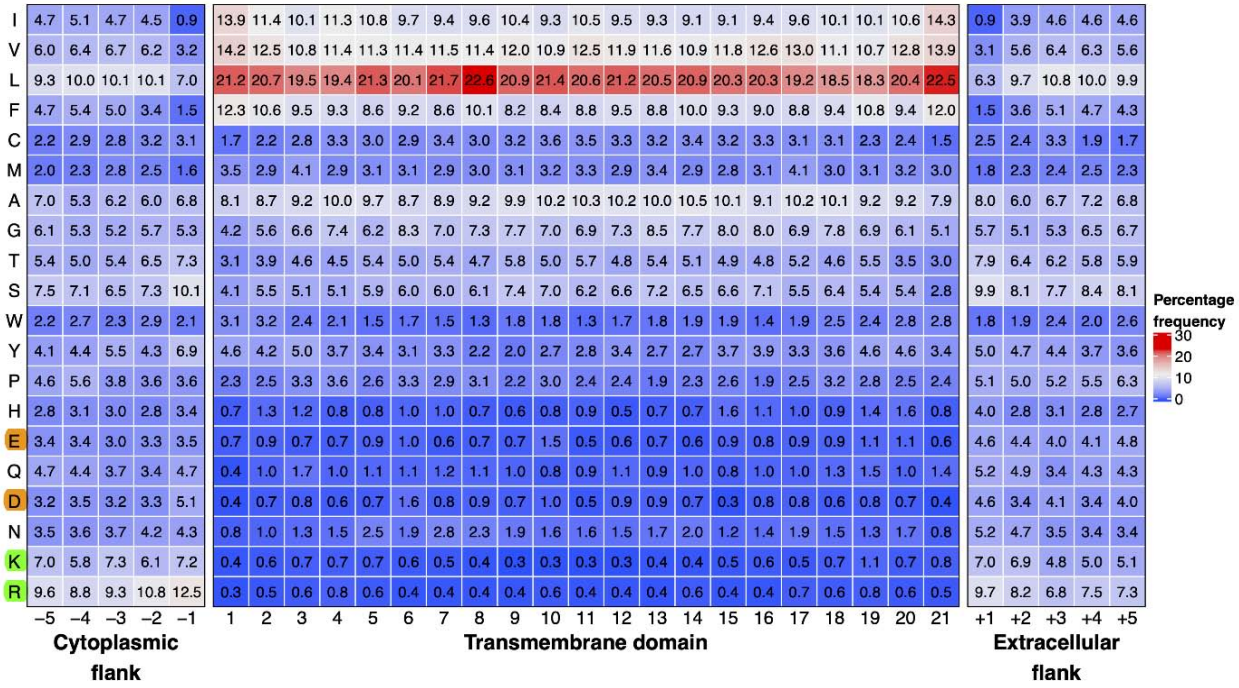

Figure 3. Heat map visualising the percentage frequency distribution of every amino acid along the length of TM domains spanning 21 amino acids. All TM sequences are oriented such that they travel from cytoplasmic to extracellular domain. Cytoplasmic and extracellular flanking sequences of five additional amino acids are included. Amino acids are listed in the order of decreasing hydrophobicity according to the Kyte–Doolittle scale. The positively charged residues, lysine (K) and arginine (R), are highlighted in orange. The negatively charged residues, glutamate (E) and aspartate (D), are highlighted in green.

with 6.85% in extracellular flanks) and for negatively charged amino acids in extracellular flanks (combined average frequency of 4.15% in extracellular flanks compared with 3.48% in cytoplasmic flanks). While positive residues appeared less frequently in isolated TM domains, negative residues were comparatively rarer when taking into account the short flanking sequences.

Rare functional variants involving charged residues in TM domains

In total, there were 235 TM proteins with a total of 2,670 clinically significant, rare genetic variants reported in ClinVar, representing 4.54% of the 5,177 TM proteins in Swiss-Prot. 69 of these proteins shared 255 variants which were localised to TM domains. After complete filtering, 34 proteins were identified which shared a total of 84 documented point mutations involving a charged residue within a TM domain (Table 1). There were more variants involving positively charged residues than negatively charged. Notably, there were no reported variants where a charged residue was substituted with another carrying the same charge. The most highly represented disease phenotype, and the only phenotype with a review status of three stars or greater, was cystic fibrosis that had 18 associated mutations involving TM charge disturbances in the cystic fibrosis transmembrane conductance regulator (CFTR). Other commonly observed phenotypes included early infantile epileptic encephalopathy, Brugada syndrome and Wilson disease. A list of all reported phenotypes and their frequency is shown in Table 2. Many variants were associated with more than one phenotype.

DISCUSSION

Transmembrane biology and physicochemical significance of charged residues

Of the 20 amino acids in the standard genetic code, four carry an electric charge at physiological pH, approximately 7.4. Arginine and lysine contain a

Table 1. Frequency of clinically significant rare genetic variants involving TM charge disturbances.

	Number of rare genetic variants
Involving any charged amino acid	84
Change from a positively charged amino acid	34
to another positively charged amino acid	0
to a negatively charged amino acid	0
Change from a negatively charged amino acid	18
to another negatively charged amino acid	0
to a positively charged amino acid	3
Change to a positively charged amino acid	26
Change to a negatively charged amino acid	9

Table 2. List and frequency of disease phenotypes arising from rare genetic variants involving TM charge disturbances.

Phenotype	**Number of rare genetic variants**
Cardiac	
Brugada syndrome	10
Long QT syndrome	6
Dilated cardiomyopathy	3
Atrial fibrillation, familial	2
Progressive familial heart block type 1A	2
Sick sinus syndrome	2
Arrhythmogenic right ventricular dysplasia/cardiomyopathy	1
Atrial standstill	1
Atrioventricular block	1
Paroxysmal familial ventricular fibrillation	1
Dermatological	
Autosomal recessive congenital ichthyosis	1
Developmental	
Autism	1
Endocrine	
Glycogen storage disease type 1A	3
Familial hypercholesterolaemia	2
Neuronal ceroid lipofuscinosis	2
Niemann-Pick disease type C1	2
Gonadotropin-independent familial sexual precocity; Leydig cell agenesis	1
Menkes kinky-hair syndrome	1
Pendred syndrome	1
Sitosterolaemia	1

Table 2 continued..

Gastrointestinal	
Wilson disease	6
Hereditary pancreatitis	2
Genetic	
Craniosynostosis	2
Achondroplasia	1
Crouzon syndrome with acanthosis nigricans	1
Joubert syndrome	1
Lethal multiple pterygium syndrome	1
Meckel syndrome type 2	1
Megaloblastic anaemia, thiamine-responsive, with diabetes mellitus and sensorineural deafness	1
Zimmermann-Laband syndrome	1
Neurological	
Early infantile epileptic encephalopathy	11
Benign familial neonatal seizures	7
Alternating hemiplegia of childhood	4
Dystonia 12 (rapid-onset dystonia-parkinsonism)	4
Cerebellar ataxia, areflexia, pes cavus, optic atrophy and sensorineural hearing loss (CAPOS syndrome)	3
Myokymia	2
Charcot-Marie-Tooth disease	1
Hereditary spastic paraplegia	1
Hyperekplexia	1
Megalencephalic leukoencephalopathy with subcortical cysts	1
Spinocerebellar ataxia type 42	1
Oncological	
Gastrointestinal stromal tumour	2
Hereditary paraganglioma-phaechromocytoma syndromes	2
Cowden syndrome	1
Ophthalmological	
Cone/cone-rod dystrophy; retinitis pigmentosa	1
Respiratory	
Cystic fibrosis	18
Bronchiectasis with or without elevated sweat chloride	2
Urological	
Congenital bilateral absence of the vas deferens	2

flexible side chain with positive charge, while aspartate and glutamate contain a hydrophilic acid group with strong negative charge. These residues therefore confer a unique potential for electrostatic interactions among the proteins in which they are present. Although histidine contains an ionisable side chain that can become positively charged depending on local environmental factors, it generally

remains uncharged at physiological pH [17]. Analysis was therefore limited to arginine, lysine, aspartate, and glutamate for the purpose of this research. Given that four of 20 amino acids are electrically charged, there is a probability that one in every five residues is charged by random chance. For any given sequence of 21 amino acids (the predominant sequence length of database-defined TM domains) there is a likelihood of $1 - (16/20)^{21} \geq 99\%$ that at least one of the four charged residues is present by random chance alone. The finding that only 46.28% of TM domains contain a charged residue therefore reveals that charges are underrepresented in these domains. It also aligns with prior knowledge that the presence of charges in TM domains is relatively unfavourable due to the predominant hydrophobicity of these sequences, lending support to the view that charges exist in TM sequences to play specific functional roles. As expected, the amino acids with hydrophobic side chains are the most prevalent in TM sequences (Figure 3). Their relative hydrophobicity confers a predilection for the lipid-exposed surface of the TM environment [4]. While hydrophobic residues predominate TM segments, they have a high relative mutability. In contrast, polar and charged regions are highly conserved and often highly specific as they impart unique structural and functional properties to TM proteins, such as correct orientation of the protein in the membrane [18].

Functional significance of charged residues

Many TM proteins function as transport proteins, where charged residues can have influential roles in coordinating conformational changes to facilitate the movement of substrates across the membrane. A major interaction that is unique to charged amino acids is salt bridge formation, involving the combination of electrostatic interactions and hydrogen bonding between oppositely charged residues within close proximity of each other. Salt bridges not only direct protein folding and structure, but also contribute to the specificity of protein interactions with other molecules [19]. When present within TMs, they may be essential for normal protein functionality such as in the CFTR protein where two separate salt bridges work in conjunction to maintain an open ion channel [20]. Furthermore, there are proven examples of functionally relevant charges and electrostatic interactions in the core of TM sequences beyond that of structure determination. The TCR is a multisubunit protein where charged TM residues in the α chain, β chain and all CD3 subunits are critical for receptor assembly and interactions [5, 21]. Germline elimination of these charges on the pre-TCRα chain in mice has been demonstrated to impair autonomous signalling for T-cell developmental checkpoints including β-selection, impairing overall TCR repertoire formation [22].

The functional significance of charge clustering localised to TM flanking regions has also previously been observed. According to the 'positive-inside rule', first postulated by von Heijne, the cytoplasmic edges of TM domains generally have a net positive charge as a result of being enriched in positively charged arginine and lysine residues [23]. Interestingly, positive charges at these sites act as important topological determinants for these proteins, helping to correctly orient their insertion into the membrane [24, 25]. An apparent lack of conservation in these topogenic sequences is suggestive of an electrostatic mechanism rather than, for example, a sequence motif [26]. Our results reflect this by demonstrating that positively charged residues are relatively more abundant in the cytoplasmic flank of TM domains (Figure 3). The 'negative inside depletion/outside enrichment' rule has more recently been proposed to complement the 'positive-inside' trend, though negative TM charges may not be as topologically influential as their positive counterparts [13]. This phenomenon can be seen in the relative clustering of negative charges at the extracellular flank of TM domains (Figure 3).

Clinical significance of charged residues

A relative abundance of positive residues in the CFTR pore is suggestive of their involvement in ion transport [27]. Three identified variants involve substitution of the positively charged arginine in TM helix 6 of CFTR (Arg347Pro, Arg347His and Arg352Gln), while another involves the loss of arginine in TM helix 11 (Arg1102Ter). Arg347 and Arg352 are critical players in the formation of two separate salt bridges required for normal ion permeation [20]. Missense mutations in TM 6 resulting in the loss of these positive charges therefore negatively impact ion flow through the channel [28]. Significantly, the Arg347Pro mutation

is part of the current core mutation panel for general population cystic fibrosis carrier screening recommended by the American College of Medical Genetics and Genomics [29]. In addition to the CFTR, most TM proteins that were found to have clinically significant rare genetic variants were also characterised by transporter activity. In the context of epileptic encephalopathy, potassium currents are key regulators of neuronal excitability. In the potassium voltage-gated channel protein Kv7.2 (encoded by the KCNQ2 gene), an electrostatic interaction between glutamate (Glu140) in the S2 TM helix and arginine (Arg210) in the S4 TM helix is required to stabilise the voltage-sensing domain. A novel variant involving the substitution of this glutamate residue (Glu140Gln) prevents this interaction from occurring, contributing to the pathogenesis of epileptic disease [30, 31].

Based on the knowledge that charged amino acids confer physicochemical properties to proteins, it is not surprising to find that there were no reported disease phenotypes associated with missense mutations where a charged amino acid is substituted for one with the same charge (Table 1).

Therapeutic utility of charged residues

Even though many drug targets are membrane proteins, specific targeting of charged TM residues to modulate electrostatic interactions is still a developing field. Manolios *et al.* established proof of concept for the design of therapeutic peptides to inhibit T-cell function, demonstrating that conjugation by synthetic lipopeptides of a region within the α-subunit TM domain containing two positive charges (arginine and lysine) can disrupt TCR assembly, preventing T-cell proliferation [21]. Novel peptides engineered to specifically alter receptor function in this way have wide potential application. For example, a proposed mechanism of reducing rheumatoid arthritis-mediated inflammation involves downregulation of the TCR response by targeting the TM charges that are critical for its function [32]. In human immunodeficiency virus (HIV), a potentially charged residue in the Gp41 protein modulates T-cell activation *via* electrostatic interactions due to sequence similarity with the TCR-α TM domain [33, 34]. This provides a basis for the ongoing design of drugs, particularly therapeutic peptides, to inhibit viral fusion and entry into host cells [35].

Clinically important non-human examples have also been demonstrated by the presence of evolutionarily conserved TM charges in the proteins of pathogenic microorganisms. For instance, the multidrug transporter protein MdfA in *Escherichia coli* contains a positively charged arginine residue in the TM region which is implicated in multidrug efflux, a key mechanism of antimicrobial resistance [36]. A similar outcome has been demonstrated in the multidrug resistance protein Pdr5p of *Saccharomyces cerevisiae*, where the substitution of negatively charged glutamate residues with a positively charged lysine in TM helix 2 impairs its efflux function [37].

Limitations

This research encountered several limitations that warrant discussion. Firstly, although the Swiss-Prot database is manually annotated and reviewed, positional information including TM domain determination is largely derived from predictive tools based on amino acid sequence. While sequence information makes analysis of very large data sets more accessible, structural analysis by way of crystallisation remains the gold standard for investigating protein structure, function, and interactions. However, there is a relative scarcity of high-resolution three-dimensional protein structures, with membrane proteins constituting only 2% of solved protein structures in the Protein Data Bank [38]. This is also the major obstacle for advancing the therapeutic targeting of these proteins [39]. Secondly, there was some difficulty in matching protein data from UniProt and ClinVar due to the databases using gene transcripts that were updated to different degrees. This resulted in possible failure to capture relevant genetic variants. For example, a variant may have been reported in ClinVar for one transcript, while UniProt used a different transcript version to report sequence and positional information. However, it was necessary to retain the UniProt data as this is where information relating to TM domains was derived. This was therefore a constraining factor in our analysis. Thirdly, organelle-specific features may also influence TM sequence due to differences in physicochemical properties of the bilayers [40]. Similar distributional analysis has been performed for different subcellular membrane locations such as endoplasmic reticulum and Golgi apparatus [13]. However, no distinction

was made in this study which opted for an overarching analysis of TM domains in general.

CONCLUSION

The presence of charged amino acids in TM domains of human proteins is relatively uncommon. Despite scarce distribution, there are many documented cases of TM charges performing key functional roles that range from membrane insertion and structure determination to directing specific interactions with other peptides and molecules. This has many implications for our understanding of health and disease states and provides a basis for therapeutic targeting of these residues. Further detailed investigation into specific TM proteins using three-dimensional structures will help to elucidate the unique significance of charged amino acids.

ACKNOWLEDGEMENTS

The authors acknowledge the assistance of members of the Bioinformatics Unit at the Children's Medical Research Institute, Ted Wong and Simon Cai, for their invaluable guidance relating to bioinformatics and computational biology, particularly the use of programming languages to extract data from online sources. In addition, we would like to thank Dr Bret Church for his expert insight and input relating to protein structure-function relationships.

CONFLICT OF INTEREST STATEMENT

The authors confirm that this article content has no conflict of interest.

REFERENCES

1. Wallin, E. and von Heijne, G. 1998, Protein Sci., 7, 1029.
2. Bull, S. C. and Doig, A. J. 2015, PLoS One, 10, e0117955.
3. Rask-Andersen, M., Almén, M. S. and Schiöth, H. B. 2011, Nat. Rev. Drug Discov., 10, 579.
4. Ulmschneider, M. B. and Sansom, M. S. 2001, Biochim. Biophys. Acta, 1512, 1.
5. Manolios, N., Bonifacino, J. S. and Klausner, R. D. 1990, Science, 249, 274.
6. Call, M. E., Pyrdol, J., Wiedmann, M. and Wucherpfennig, K. W. 2002, Cell, 111, 967.
7. Ng, D. P., Poulsen, B. E. and Deber, C. M. 2012, Biochim. Biophys. Acta Biomembr., 1818, 1115.
8. Sanders, C. R. and Nagy, J. K. 2000, Curr. Opin. Struct. Biol., 10, 438.
9. Stone, T. A. and Deber, C. M. 2017, Biochim. Biophys. Acta Biomembr., 1859, 577.
10. Gokhale, A. S. and Satyanarayanajois, S. 2014, Immunotherapy, 6, 755.
11. The UniProt Consortium. 2019, Nucleic Acids Res., 47, D506.
12. Gu, Z., Eils, R. and Schlesner, M. 2016, Bioinformatics, 32, 2847.
13. Baker, J. A., Wong, W. C., Eisenhaber, B., Warwicker, J. and Eisenhaber, F. 2017, BMC Biol., 15, 66.
14. Kyte, J. and Doolittle, R. F. 1982, J. Mol. Biol., 157, 105.
15. Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Jang, W., Karapetyan, K., Katz, K., Liu, C., Maddipatla, Z., Malheiro, A., McDaniel, K., Ovetsky, M., Riley, G., Zhou, G., Holmes, J. B., Kattman, B. L. and Maglott, D. R. 2018, Nucleic Acids Res., 46, D1062.
16. Kinsella, R. J., Kähäri, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., Staines, D., Derwent, P., Kerhornou, A., Kersey, P. and Flicek, P. 2011, Database (Oxford), 2011, bar030.
17. Pace, C. N., Grimsley, G. R. and Scholtz, J. M. 2009, J. Biol. Chem., 284, 13285.
18. Jones, D. T., Taylor, W. R. and Thornton, J. M. 1994, FEBS Lett., 339, 269.
19. Bosshard, H. R., Marti, D. N. and Jelesarov, I. 2004, J. Mol. Recognit., 17, 1.
20. Cui, G., Freeman, C. S., Knotts, T., Prince, C. Z., Kuang, C. and McCarty, N. A. 2013, J. Biol. Chem., 288, 20758.
21. Manolios, N., Collier, S., Taylor, J., Pollard, J., Harrison, L. C. and Bender, V. 1997, Nat. Med., 3, 84.
22. Ishikawa, E., Miyake, Y., Hara, H., Saito, T. and Yamasaki, S. 2010, Proc. Natl. Acad. Sci. USA, 107, 19979.
23. von Heijne, G. 1989, Nature, 341, 456.
24. Lerch-Bader, M., Lundin, C., Kim, H., Nilsson, I. and von Heijne, G. 2008, Proc. Natl. Acad. Sci. USA, 105, 4127.

25. Boyd, D. and Beckwith, J. 1990, Cell, 62, 1031.
26. Harley, C. A. and Tipper, D. J. 1996, J. Biol. Chem., 271, 24625.
27. van Willigen, M., Vonk, A. M., Yeoh, H. Y., Kruisselbrink, E., Kleizen, B., van der Ent, C. K., Egmond, M. R., de Jonge, H. R., Braakman, I., Beekman, J. M. and van der Sluijs, P. 2019, Life Sci. Alliance, 2, e201800172.
28. Cotten, J. F. and Welsh, M. J. 1999, J. Biol. Chem., 274, 5429.
29. Grody, W. W., Cutting, G. R., Klinger, K. W., Richards, C. S., Watson, M. S., Desnick, R. J., Subcommittee on Cystic Fibrosis Screening, Accreditation of Genetic Services Committee, ACMG. American College of Medical Genetics. 2001, Genet. Med., 3, 149.
30. Soldovieri, M. V., Ambrosino, P., Mosca, I., Miceli, F., Franco, C., Canzoniero, L. M. T., Kline-Fath, B., Cooper, E. C., Venkatesan, C. and Taglialatela, M. 2019, Int. J. Mol. Sci., 20, 3382.
31. Millichap, J. J. and Cooper, E. C. 2012, Epilepsy Curr., 12, 150.
32. Ali, M. and Manolios, N. 2007, Future Rheumatol., 2, 287.
33. Perrin, J., Bary, A., Vernay, A. and Cosson, P. 2018, BMC Cell Biol., 19, 3.
34. Cohen, T., Cohen, S. J., Antonovsky, N., Cohen, I. R. and Shai, Y. 2010, PLoS Pathog., 6, e1001085.
35. Reuven, E. M., Dadon, Y., Viard, M., Manukovsky, N., Blumenthal, R. and Shai, Y. 2012, Biochemistry, 51, 2867.
36. Sigal, N., Vardy, E., Molshanski-Mor, S., Eitan, A., Pilpel, Y., Schuldiner, S. and Bibi, E. 2005, Biochemistry, 44, 14870.
37. Dou, W., Zhu, J., Wang, T., Wang, W., Li, H., Chen, X. and Guan, W. 2016, FEMS Yeast Res., 16, fow031.
38. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. 2000, Nucleic Acids Res., 28, 235.
39. Yin, H. and Flynn, A. D. 2016, Annu. Rev. Biomed. Eng., 18, 51.
40. Sharpe, H. J., Stevens, T. J. and Munro, S. 2010, Cell, 142, 158.