

Functional Categorization and Comparative 3D Models Study of TMEM16B

Humaira Aziz Sawal¹, Naila Aslam Shad², Rabea Ejaz³, Sharqa Aziz⁴

¹Department of Zoology, Rawalpindi Women University, Rawalpindi, Pakistan.

²Department of Zoology, Rawalpindi Women University, Rawalpindi, Pakistan.

³Department of Zoology, Rawalpindi Women University, Rawalpindi, Pakistan.

Corresponding Author's email: humaira_sawal@yahoo.com
haszoo@f.rwu.edu.pk

Abstract

TMEMs expressed nearly in every part of the body, but its expression is high in brain. TMEM16B is a transmembrane protein perform numerous physiological functions. It encodes a conserved Ca²⁺-activated Cl⁻ channel (CaCC). This protein mediates the regulation of Ca²⁺ levels and is involved in anxiety related disorders. Its distributive pattern shows that it performs an important role in many neurological disorders via multiple signaling pathways. The stability and reactivity of this protein was assessed physio chemically and by determination of its domain. Different bioinformatics tools such as Pfam, ProtParam, and SOPMA were used respectively for protein secondary structures prediction. The drug discovery processes were dependent upon the protein modelling to large extent. This study was based upon the comparative *in-silico* modelling of 3-dimensional (3D) structure prediction through Robetta, I-TASSER, and AlphaFold (bioinformatics tools). The validation of these models were checked by using the SAVESv6.0 (PROCHECK) server. We got the best 3D structure by using the Robetta. The 3D structure prediction of TMEM16B was important because its function was determined by its structure and, after the structure finding, scientists would develop the drugs that could cure the neurological disorders associated with this protein. It can be concluded that this structure will help us in assembly of complete structural proteome profile. Huge data generated through experimentation will help in future machine learning that will assist in increasing the number of structural templates.

Keywords: TMEM16B, Rosetta, Alpha Fold, I-Tasser, CaCC.

Article History: Received: 26th May 2023, Revised: 8th June 2023, Accepted: 4th December 2023, Published: 30th December 2023.

Creative Commons License: NUST Journal of Natural Sciences (NJNS) is licensed under Creative Commons Attribution 4.0 International License.



Introduction

The term protein was predicted by the Swedish chemist Jacob Berzelius for the nitrogen containing compounds formed by the grouping of amino acids connected

through peptide bond [1]. These proteins plays an important role in enzymatic catalysis, cell signaling, cell structural support, immune protection, transcription and translation [2]. The 3D structure is associated with the function of the protein

and is useful for the understanding of biological processes information related to human health and diseases. According to previous reported literature experimentally many protein structures were obtained, but studies showed that there is a big gap between those proteins having tertiary structures and protein sequences placed in Uniprot [3].

The TMEM16 family consist of membrane proteins, called as anoctamins, which were involved in diverse physiological functions which includes transportation of ions and regulation of ion channels. The important functionally characterized members were, TMEM16A (ANO1) and TMEM16B (ANO2), which were related to Ca²⁺-dependent Cl⁻ channels (CaCCs) and plays an important role for phototransduction, trans epithelial ion transport, olfaction, nociception, smooth muscle contraction, cell proliferation and control of neuronal excitability.

The TMEM16 family consist of three different subclasses, channels and scramblases. The structure and function implication of such a functional multiplicity in a single protein family needs to be explained and the associations between TMEM16 functions and human pathologies and physiology need to be inspected [4]. In 2008, 3 groups independently identified two individual members of the TMEM16 family of membrane proteins, TMEM16A (ANO1) and TMEM16B (ANO2), as key constituents of CaCCs [5]. TMEM16B have a sequence of 1003 amino acids with an accession number Q9NQ90 in UniProt [6].

Various functions were regulated by CaCCs mediated via TMEM16 proteins, such as nociception, epithelial secretion, neuronal signaling, smooth muscle contraction, host protection, cell proliferation, sign transduction, and tumor genesis [7]. Protein family of TMEM16 has two types of functionally distinct but structurally conserved membrane

transporters that function as CaCCs. Detailed functional and structural studies have explained the TMEM16 physiological functions and molecular mechanisms. TMEM16A and TMEM16B CaCCs regulate contraction of muscle that is smooth, trans epithelial fluid transport, TMEM16 phospholipid scramblases facilitate the flip-flop structure of phospholipids across the membrane to allow phosphatidylserine externalization, which plays an important role in the increase of processes including blood coagulation, bone development, viral and cell fusion [8].

The study of TMEM16B is significant because they have a role in a number of disease pathways. Scientists can predict the characteristics and structure of proteins that can be used by computational approaches for interaction studies. So, the characterization and *insilico* identification of TMEM16B have been done in the current study.

Secondary structure prediction involves several procedures in bioinformatics. The prediction is based only on amino acid sequences and the main objective is to predict the secondary structure of proteins. The target protein secondary structure was predicted by using an optimized process with help of multiple tools e.g. ALPHAFOLD, ROBETTA and ITASSER. Sequence similarity searching is used to predict protein structure by alpha fold, using the alignment of the target sequence and statistical assessment to homology and transfer the sequence [9]. ROBETTA server is used to predict the 3D structure of a protein, using various parameters such as database searching, template and domain finding [10]. I-TASSER is an online bioinformatics tool which is based on algorithms and provides a platform to the academic users for generating automatically high-quality protein model predictions of 3D structure [11]

Materials and Methods

The methodology used for this comparative study was taken from our previous published study [9]. According to our previous published research three bioinformatics tools (Robetta, I-TASSER, and AlphaFold) were used for comparative *in-silico* 3-dimensional (3D) model structure prediction [9]. The methodology consists of three major steps. The physiochemical properties were determined followed by the depiction of secondary structure and building of protein model.

was retrieved from NCBI's (National Center for Biotechnology Information) protein database (NP_001265525.1) [12]. In UniProt database its ID is Q9NQ90 [13]. For Physiochemical Properties depiction

the server ProtParam, from ExPASy was used. It includes the description of the targeted protein amino acid sequence, the aliphatic index (AI), isoelectric point (pI), and the grand average of hydropathy (GRAVY),

Table 1. The protein predictive models validation by Ramachandran plot

Tools	PDB structure numbers	Error	Ramachandran Most favorable region %	Chat Additional and generously allowed regions %	Results Disallowed Region %
Robetta	• Model 1	94.98 98	89.2	8.9+0.8	1.1
	• Model 2	94.29 15	87.6	9.9+1.4	1.1
	• Model 3	93.05 41	91.1	8.2+0.6	0.1
	• Model 4	93.66 7	87.4	10.4+1.0]	1.3
	• Model 5	96.30 39	85.1	11.6+1.4	1.9
I-Tasser	• Model 1	86.86	71.3	22.7+2.8	3.2
	• Model 2	87	71.5	21.9+4.4	2.2
	• Model 3	80.50	68.8	23.6+4.4	3.2
	• Model 4	51	67.8	24.8+3.8	3.6
	• Model 5	80.10 2 80.40 4 86.72 75	74.4	20.6+2.3	2.7
Alpha fold	• Model 1	95.18 83	91.5	8.5+0.0	0.0

FASTA sequence of TMEM16B protein

Functional Categorization and Comparative 3D Models Study of TMEM16B

the extinction coefficients (EC), the total positive (+R) and negative (-R) residues. Further for protein characterization, the Pfam analysis and the instability index (II), were performed in relation to the protein family [14], [15].

Secondary structures consist of alpha helix, beta turn, beta sheets and random coils. SOPMA (Self-Optimized Prediction Method with Alignment) [16] method was used for determining the secondary structural characteristics. Amino acid

Results

According to NCBI, TMEM16B is a protein consist of 25 exons, and encodes for 1003 amino acids. All PDB files with homologous sequences were searched by Blastp. The detailed study of its homology shows that with Homo sapiens TMEM16B was showing >99% homology. The Physiochemical properties were calculated and it shows that the molecular weight of protein was 113969.38 with 1003 amino acids. The isoelectric point assess the protein electrophoretic separation, its solubility and electrophoresis [19]. The isoelectric point was 6.12 (near to 7), which describes that its electrophoretic separation and protein solubility is in average range. Aliphatic index (AI) is defined as the space filled by aliphatic side chains in a protein, which determines that on a broad range of temperature how stably a protein reacts. 82.85 is the AI value of TMEM16B.

Instability index (II) determines the stability of a protein, if its value is less than 40, it shows that a protein is stable [20]. Our protein was figured out to be 45.46, which shows the nearly stable state of protein. Reported literature shows that if the protein have low GRAVY indices, it interacts with water molecules very competently. If the GRAVY score value is above 0, such proteins have hydrophobic interactions [21]. TMEM16B has the

sequence was used for this procedure. Finally Building of 3D TMEM16B protein model was performed by using different Web tools were used. These tools were Robetta, I-TASSER, and AlphaFold [17].

The model was validated by using SAVESv6.0 (PROCHECK) server [18] (Table 1).

hydrophilic interactions because its GRAVY score is -0.281.

Pfam analysis was performed for determining the functional characterization of protein. This family consist of eight putative transmembrane domains. It is evidently a calcium-activated ionic channel because it has no similarity with other known channel proteins. It is expressed in sensory neurons, secretory epithelia and retina. In diverse physiological processes, it mediates receptor-activated chloride currents [22]. This protein family seems to be the cytoplasmic domain of the calcium-activated chloride-channel, anoctamin, protein. It is responsible for creating the homodimeric architecture of the chloride-channel proteins [23]. SOPMA was used for the determination of protein's secondary structure. The analysis results shows that secondary structural elements consist of extended strands and beta sheets, followed by random coils and alpha helix. According to its analysis protein consist of 1003 amino acids, 10.17% extended strands, 46.06% alpha helix, 41.97% random coils, 1.79% beta sheets.

The protein 3D structure was modelled by an abinitio approach, Robetta, AlphaFold and I-TASSER(15) as shown in the Figure 1A to C. On the basis of quadrangle regions, the residues were characterized in the Ramachandran plot analysis. The more permitted areas were shown by the graph's

red sections while the permitted regions were shown by yellow sections. For models evaluation Ramachandran plot was generated by PROCHECK. By using the calculations of Ramachandran plot with the help of the PROCHECK tool, the predicted model stereochemical quality and the protein models quality were assessed following the refinement process. Our studied proteins result shows that the total number of residues scattered in the most distributed area is greater than 85%, which shows the accuracy and high quality of the modelled structure [24]. The target protein model predicted through I-TASSER has 2.7 residues in the disallowed region and ERRAT value of **86.7275**; it was validated through Ramachandran plot using SAVESv6.0 (Figure 2). The Robetta-predicted model has 1.1 residues in the forbidden area and has an ERRAT score of 94.9898; it was also confirmed using a Ramachandran plot and SAVESv6.0 (Figure 2B). The protein's 3D structure predicted by AlphaFold contains 0.0 residues in the forbidden area and an ERRAT value of 95.1883 (Figure 2). SAVESv6.0 is the programme we used to calculate the ERRAT value and the Ramachandran plot for each model. The model with the best Ramachandran plot was Robetta's. All the conclusions indicated above were extremely comparable to our previously reported comparative analysis research [9].

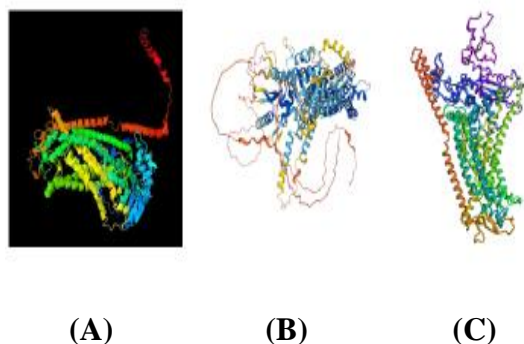


Figure 1. (A) The 3D predictive model through I-TASSER of targeted protein. (B) The 3D predictive model through AlphaFold of targeted protein. (C) The 3D predictive model through Robetta of targeted protein

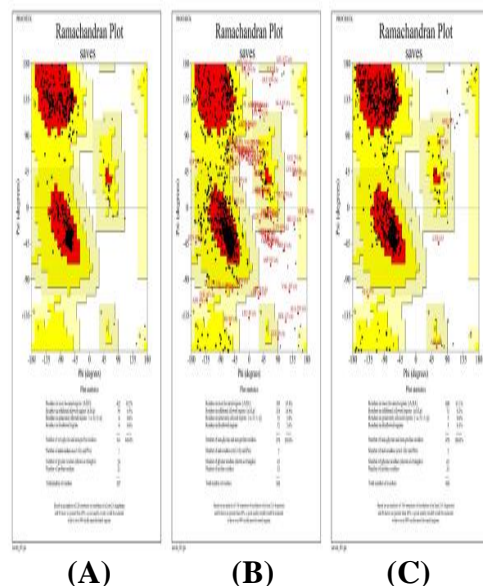


Figure 2. (A) The Ramachandran plot of AlphaFold predicted modelled protein. (B) The Ramachandran plot of I-TASSER predicted modelled protein. (C) The Ramachandran plot of Robetta predicted modelled protein.

Discussion

Proteins are important in the functioning of living cells in organisms. They are important for maintaining the structure and regulates the organs and tissues in the body, and their arrangement can assist in comprehending their purpose. Amino acid sequences dictate the 3D structure prediction of proteins [25]. Tools and methodologies used in computational biology have been developed for studying the 3D structures of proteins and the aim of our study is to focus on different tools for designing a comparative structure of TMEM16B protein.

Transmembrane protein family members (TMEMs), were proteins that cross biofilms [26], and positioned in the plasma membrane specifically in lipid bilayer. Some TMEMs may also present in other membranous organelles of cell. Based on their topological structure, which depends on their C-terminal and N-terminal domains, TMEMs proteins were divided into α -helical and β -barrel proteins [27], [28]. TMEMs were participate in important physiological processes e.g they regulate ion channel transport, signal transduction, cellular chemotaxis, adhesion, programmed cell death, and autophagy [29], due to presence in many types of cells. TMEMs were expressed in nearly every part of the body but highly expressed in brain. Its distributive pattern shows that it performs an important role in many neurological disorders via multiple signaling pathways.

TMEM16B is a transmembrane protein encoded by ANO2 (Anoctamin 2) gene. As it was reported that TMEM16B is a transmembrane protein having many physiologic functions. This protein mediates the regulation of Ca^{2+} levels and that was also involved in anxiety [30]. It encodes a Ca^{2+} -activated Cl^- channel (CaCC) [31]. They were also involved in the metabolism of glucose and lysosomal autophagy. So lysosomal dysfunction and abnormal glucose metabolism is responsible for neurodegenerative changes.

The process encompasses ion transportation, the action of phospholipid scramblases, and the control of additional membrane proteins. It plays a role in diverse physiological activities such as nociception, epithelial fluid secretion, signaling pathways, and cell proliferation. TMEM16B provides an effective way of regulating neuronal interest. It has been reported that over-expression of TMEM16A leads to cell proliferation and causes various types of cancers. The role

of TMEM16B was identified in mice's behavior changes they showed an increase in aggressive behavior with the removal of TMEM16B from the hippocampus of mice's brains. Consequently, any alteration in TMEM16B's structure could potentially induce numerous neurological shifts and might participate in cell growth, mirroring the role previously observed in TMEM16A based on earlier research [6]. The main focus of our investigation, as outlined in this study, centres on protein structures. The model we construct will aid in understanding internal conformational variances and relationships between conformers, thus establishing a foundation for scrutinizing the evolutionary progression of protein. The structural comparison tools explained in this study simplify the examining of other different protein families, which helps in identification of common structural and dynamic features. Such comparative analysis of homologues provides priceless conformational breakthroughs beneficial for evaluating the results of theoretical methods and new crystallographic structures. This includes flexible protein-protein docking and crafting homology models, where exploring principal components might yield plausible alternate arrangements. Another crucial research avenue involves unraveling potential communication networks within proteins, especially in comprehending cellular proliferation mechanisms that seem to persist across distant genetic lineages. Embarking on this path, theoretical examinations coupled with comparative assessments of structural analogs mark the initial stride forward. Large numbers of validation tools were used to evaluate and assess the quality of the models of TMEM16B that are generated by MODLER SAVES v6.0. It is a software tool that is used in this study for evaluation of the ERRAT value and Ramachandran plots generated for all the 11 models to have the best model of it. To run the

program we uploaded the models in PDB format in SAVES one by one and evaluated their ERRATE value and PROCHECK. The best model value we got from validation was of Robetta.

For understanding the structures and functions of membrane proteins, description of classifications of all the membrane proteins will be useful. It will be helpful for membrane proteins studies in providing basic knowledge for numerous predictive research projects.

References

- 1) Voet D, Voet JG, & Pratt CW. Fundamentals of biochemistry: life at the molecular level. John Wiley & Sons. 2016; Feb29.
- 2) Bertoline LM, Lima AN, Krieger JE & Teixeira SK. Before and after AlphaFold2: An overview of protein structure prediction. *Frontiers in Bioinformatics*. 2023; 3: 1120370.
- 3) Picollo A, Malvezzi M, & Accardi A. TMEM16 proteins: unknown structure and confusing functions. *Journal of molecular biology*. 2015; 427(1): 94-105.
- 4) Pifferi S, Dibattista M, & Menini A. TMEM16B induces chloride currents activated by calcium in mammalian cells. *Pflügers Archiv-European Journal of Physiology*. 2019; 458, 1023-1038.
- 5) <https://www.uniprot.org/uniprotkb/Q9NQ90/entry>
- 6) Wang L, Simms J, Peters CJ, Tynan-La Fontaine, M LiK, Gill TM & Jan LY. TMEM16B calcium-activated chloride channels regulate action potential firing in lateral septum and aggression in male mice. *Journal of Neuroscience*. 2019; 39(36): 7102-7117.
- 7) Le SC, & Yang H. Structure–Function of TMEM16 Ion Channels and Lipid Scramblases. In *Ion Channels in Biophysics and Physiology*. Singapore: Springer Nature Singapore. 2022; 87-109.
- 8) Saldaño T, Escobedo N, Marchetti J, Zea DJ, Mac Donagh J, Velez Rueda AJ, Gonik E, García Melani A, Novomisky Nechcoff J, Salas MN, Peters T. Impact of protein conformational diversity on AlphaFold predictions. *Bioinformatics*. 2022; 38(10):2742-2748.
- 9) Sawal HA, Nighat S, Safdar T, Anees L. Comparative In Silico Analysis and Functional Characterization of TANK-Binding Kinase 1–Binding Protein 1. *Bioinform Biol Insights*. 2023; 17:11779322231164828.
- 10) Yang J, Zhang Y. Protein structure and function prediction using I-TASSER. *Current protocols in bioinformatics*. *Curr Protoc Bioinformatics*. 2015;52:5.8.1-5.8.15.
- 11) https://www.ncbi.nlm.nih.gov/protein/NP_001265525.1
- 12) <https://www.uniprot.org/uniprotkb/Q9NQ90/entry>
- 13) ExPASy – ProtParam tool. <https://web.expasy.org/protparam/>
- 14) <https://www.ebi.ac.uk/Tools/pfa/pfamscan/>
- 15) https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.
- 16) <http://robetta.bakerlab.org/>
- 17) <https://saves.mbi.ucla.edu/>
- 18) Audain E, Ramos Y, Hermjakob H, Flower DR, Perez-Riverol Y. Accurate estimation of isoelectric point of protein and peptide based on amino acid sequences. *Bioinformatics*. 2016; 32(6):821-7.

- 19) Sivakumar K, Balaji S, Gangaradhakrishnan. In silico characterization of antifreeze proteins using computational tools and servers. *Journal of Chemical Sciences*. 2007 Sep;119: 571-9.
- 20) Gamage DG, Gunaratne A, Periyannan GR, Russell TG. Applicability of instability index for in vitro protein stability prediction. *Protein Pept Lett*. 2019;26 (5):339-347.
- 21) Magdeldin S, Yoshida Y, Li H, Maeda Y, Yokoyama M, Enany S, Zhang Y, Xu B, Fujinaka H, Yaoita E, Sasaki S. Murine colon proteome and characterization of the protein pathways. *BioData Min*. 2012; 5(1):11.
- 22) <https://www.ebi.ac.uk/interpro/entry/pfam/PF04547>
- 23) <https://www.ebi.ac.uk/interpro/entry/pfam/PF16178>
- 24) Gabler F, Nam SZ, Till S, Mirdita M, Steinegger M, Söding J, Lupas AN, Alva V. Protein sequence analysis using the MPI bioinformatics toolkit. *Curr Protoc Bioinformatics*. 2020; 72(1):e108.
- 25) Cenedese V, Betto G, Celsi F, Cherian OL, Pifferi S, Menini A. The voltage dependence of the TMEM16B/anoctamin2 calcium-activated chloride channel is modified by mutations in the first putative intracellular loop. *J Gen Physiol*. 2012;139(4):285-94.
- 26) Chen Q, Fang J, Shen H, Chen L, Shi M, Huang X, Miao Z, Gong Y. Roles, molecular mechanisms, and signaling pathways of TMEMs in neurological diseases. *Am J Transl Res*. 2021; 13 (12):13273-13297.
- 27) Vinothkumar KR, Henderson R. Structures of membrane proteins. *Q Rev Biophys*. 2010; 43(1):65-158.
- 28) von Heijne G. Membrane-protein topology. *Nat Rev Mol Cell Biol*. 2006; 7(12):909-18.
- 29) Schmit K, Michiels C. TMEM proteins in cancer: a review. *Front Pharmacol*. 2018; 9:1345.
- 30) Subaran RL, Talati A, Hamilton SP, Adams P, Weissman MM, Fyer AJ, Hodge SE. A survey of putative anxiety-associated genes in panic disorder patients with and without bladder symptoms. *Psychiatr Genet*. 2012; 22(6):271-8.
- 31) Wang R, Lu Y, Cicha MZ, Singh MV, Benson CJ, Madden CJ, Chapleau MW, Abboud FM. TMEM16B determines cholecystokinin sensitivity of intestinal vagal afferents of nodose neurons. *JCI Insight*. 2019; 4(5):e122058.

