bioRxiv preprint doi: https://doi.org/10.1101/2020.05.18.099887; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Novel insights into the structure and transport mechanisms of TAPT1

Md Sorwer Alam Parvez[1], Mohammad Mahfujur Rahman[1], Md Niaz Morshed[1], Dolilur Rahman[1], Saeed Anwar[2], Paul J. Coucke[3], Mohammad Jakir Hosen[1,*]

[1]Department of Genetic Engineering & Biotechnology, Shahjalal University of Science & Technology, Sylhet 3114, Bangladesh

[2]Department of Medical Genetics, Faculty of Medicine and Dentistry, University of Alberta, 8440 112 St. NW, Edmonton, AB T6G 2R7, Canada

[3]Center for Medical Genetics, Ghent University Hospital, 9000 Ghent, Belgium

*Corresponding author: E-mail: jakir-gen@sust.edu

Running title: Structure and transport mechanisms of TAPT1

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.18.099887; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Transmembrane anterior-posterior transformation protein 1 (TAPT1), encoded by the *TAPT1* gene expressed in the basal ciliary body, plays a crucial role in cilia formation as well as axial skeletal patterning. Mutations in this gene have been reported to cause several ciliopathies and osteo-related diseases. Unfortunately, the cellular and molecular pathogenic mechanisms are still unclear also due to the lack of X-ray crystallographic structure and further characterization of TAPT1 protein. In this study, we attempted to characterize this protein by *in silico* techniques. A 3D structure of TAPT1 was generated by the *ab initio* method, which was further used for the analysis of the substrate-binding site, to determine pore size and for the prediction of the possible substrate(s). Validation by using different software packages revealed a reliable 3D model of TAPT1. Topology modeling revealed that TAPT1 has eight transmembrane helices with a total number of 27 helices in secondary structure. The amino acid residues H235, R323, K443, N446, S447, L450, K453, S454, Y457, K511, N513, D533, K535, D536, and T538 were found to form the pore surface as well as involved in the binding interaction with the substrate(s). This study predicted flavonoids as the possible substrate for TAPT1, which could further be confirmed by ingenuity pathway analysis. Moreover, our analysis indicated that TAPT1 might localize in the mitochondrial membrane in addition to the ciliary basal body. Our study gives novel insights for TAPT1 structure and its function.

Keywords: TAPT1, topology modeling, substrate-binding site, homology modeling, molecular docking

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.18.099887; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1. Introduction

The evolutionary conserved human trans-membrane anterior-posterior transformation protein 1 (TAPT1) is encoded by the TAPT1 gene, located at Ch4 p15.32. In eukaryotes, TAPT1 plays a role in axial skeletal patterning during early embryonic development [1, 2]. Mutations in TAPT1 causes a lethal congenital syndrome, by which clinical features overlap between skeletal dysplasia and ciliopathies [3]. The cellular and molecular mechanisms causing the syndrome are still unclear. Although expression of the TAPT1 is noticed at most body sites, localization is most common at the centrosome and basal ciliary body [4]. Defective TAPT1 mislocalized in the cytoplasm disrupts the morphology of the Golgi apparatus and intracellular protein trafficking, giving rise to abnormal cilia formation [3, 5].

In mice, Tapt1 is predicted to comprise of several transmembrane domains, and part of the gene is orthologous to alternatively spliced human transcript encoding the cytomegalovirus gH receptor. Besides, Howell et al. also speculated that Tapt1 acts as a downstream effector of Hoxc8 that may transduce extracellular information required for axial skeletal patterning during embryonic development [6]. In zebrafish models, knockdown of tapt1b revealed delayed ossification and abnormal craniofacial skeleton development, linked to abnormal cranial neural crest cell differentiation [3].

To date, several studies endeavored to characterize TAPT1. However, the absence of a high-resolution X-ray crystallographic structure, the unknown substrate(s), and undefined molecular signaling pathway challenged the characterization of the protein. A comprehensive in silico analysis, including 3D homology modeling and other computational approaches, could be an alternative approach to gain insights into the proteins' structure, potential binding sites, and the possible substrate(s). Such studies would help gain envisage the mechanisms of substrate interaction with the protein, as well as to disclose involved molecular signaling pathways [7-9]. The current study aimed to characterize TAPT1, both structurally and functionally, using computational approaches, which may give novel insights into the function of this protein and understand the underlying pathogenic mechanisms of the diseases associated with it.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.18.099887; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2. Materials and Methods

2.1 Prediction of secondary structure and topology model of TAPT1

The primary sequence of human TAPT1 protein (accession no: AAH66899.1) was retrieved from the GenBank database of the National Center for Biotechnology Information (NCBI), which has a sequence length of 567 amino acids [10]. The protein secondary structure was analyzed using PSIPRED secondary structure prediction server, and the topology was predicted by the MEMSAT-SVM of the PSIPRED workbench online server [11-13]. PDBsum was also used to predict the secondary structure from the generated 3D model [14].

2.2 Generation and validation of TAPT1 3D homology model

The Basic Local Alignment Search Tool (BLAST) program of NCBI was used against the Protein Data Bank (PDB) to find the template for the homology modeling [15]. Unfortunately, no significant hit was found. Henceforth, the *ab initio* method was used to generate the 3D model of TAPT1. The Robetta server was used to model TAPT1 by the *ab initio* method [16]. Further, the generated 3D model was validated using RAMPAGE, PROCHECK, and ERRAT [17-19]. For the visualization of the model, a user-sponsored molecular visualization system, PyMOL was used [20].

2.3 Determination of pore diameter and pore-lining residues

We analyzed the transporter channel using the PoreWalker server, a fully automated method designed to detect and characterize transmembrane protein channels from their 3D structure [21]. Further, the PoreLogo server was used to visualize the sequence and conservation of pore-lining residues in transmembrane protein structures [22].

2.4 Subcellular localization and transporter substrate specificity prediction

The subcellular localization of this transporter was predicted by LocTree3 server which predicts the subcellular localization for all proteins in all domains of life where Water-soluble globular and transmembrane proteins are predicted in one of 18 classes in Eukaryota (e.g., chloroplast, chloroplast membrane, cytosol, ER, Golgi, ER membrane, Golgi membrane, extracellular, mitochondria, mitochondria membrane, nucleus, nucleus membrane, peroxisome, peroxisome membrane, plasma membrane, plastid,

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.18.099887; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

vacuole, and vacuole membrane) [23]. Further, the TrSSP server was used to predict the types of substrates specific for this transporter. This server predicts the substrate specificity based on evolutionary information and the AA index [24].

2.5 Protein-protein interaction analysis and prediction of the functions

GeneMANIA was used for the protein-protein interaction analysis. These tools also predict the functions by analyzing a large number of functional association data, including protein and genetic interactions, pathways, co-expression, co-localization, and protein domain similarity [25].

2.6 Molecular docking analysis for substrate prediction

Molecular docking analysis was carried out by AutoDock vina against predicted specific types of metabolites or molecules reported in the Human Metabolome Database (HMDB) [26, 27]. In our previously mentioned method, TrSSP predicted that TAPT1 involved in the transportation of charged molecules or metabolites. For which, only charged molecules or metabolites of HMDB was used as target ligands for molecular docking analysis. During this docking analysis, the grid box size was set to 75, 100, and 75 respectively for X, Y, and Z-axis, where the center was set to 35.254, 3.105 and 22.549 respectively for X, Y, and Z. Further, the binding site was analyzed in the PyMOL, and additionally, CASTp was used to cross-check the biding sites [28].

2.7 Ingenuity pathway analysis for possible pathway prediction

The top predicted substrate was then used for the ingenuity pathway analysis against TAPT1 by Ingenuity Pathway Analysis (IPA) [29]. IPA is an advanced bioinformatics tool that can generate the shortest possible pathway for the interesting data sets through ingenuity knowledge-based or text mining analysis. The shortest possible pathways were generated using the path explorer features of IPA, where the predicted substrate was selected as initial molecules. A hypothetical pathway was illustrated with Biorender.com concluding all the findings in this study.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.18.099887; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3. Results

3.1 Generation and validation of human TAPT1 *ab initio* model

Mining of Protein Data Bank (PDB) revealed no high-resolution X-ray crystallographic structure and the homolog of TAPT1 is available. Thus, Robetta online server was used to generate a 3D homology model of TAPT1. Robetta generated five primary models, which were further validated using several structure assessment methods, including PROCHECK, ERRAT, and Ramachandran plot. Among the five, the best model (Fig. 1) was then selected based on the highest score of various parameters obtained from assessment methods. While the Ramachandran plot (generated by RAMPAGE) showed 99.5% residues of the best model are in the favored regions, PROCHECK showed 93.3% residues of this model were in most favored regions. ERRAT (which analyzes statistics of non-bonded interactions between different atom types based on characteristic atomic interaction) also showed an overall quality factor of the best-selected model is 97.967% (Fig. 2). Thus, the constructed best-selected model was highly reliable for further docking study.

3.2 Secondary structure and topology model of TAPT1

MEMSAT-SVM mediated topology modeling predicted that TAPT1 comprised eight transmembrane helices, and both the N-terminal and C-terminal region located in the extracellular region (Fig. 3). The secondary structure of TAPT1 was predicted by the PDBsum, which revealed the presence of 27 helices, 32 beta turns, and a disulfide bond (Fig. 4).

3.3 Insights of the TAPT1 transport channel pore

This study revealed that 105 amino acids were involved in the formation of transport pore of the TAPT1, where amino acid residues H235, R323, K443, N446, S447, L450, K453, S454, Y457, K511, N513, D533, K535, D536, and T538 were present in the pore surface (Fig 5). Moreover, the pore was found in a diamond shape.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.18.099887; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3.4 Interactome analysis

Interactomics revealed that TAPT1 has interaction with 20 other genes with a total number of 268 links. This gene has physical interactions with MIEF1, MYO10, SQRDL, P4HB, NDUFA13, DERL1, TXNDC15, SUCO, CAV1, RIC3, UBC, and PCNT while co-expressed with DDX6, NDUFA13, NAPA, SUCO, UBC, LYRM5, PTAR1, MED17and RBL2 (Fig 6). Moreover, this interactomics predicted that TAPT1 might involve in the regulation intrinsic apoptotic signaling pathway, cellular response to oxygen level, intracellular transportation of virus and viral life cycle, and multi-organism transportation.

3.5 Subcellular localization and substrate affinity of TAPT1

Prediction of subcellular localization via LocTree3 revealed that TAPT1 possibly localized in the mitochondrial membrane, and the expected accuracy was 84%. Besides, substrate specificity analysis stated that TAPT1 would involve in the transportation of charged (both positive and negative) metabolites.

3.6 Prediction of TAPT1 substrate by molecular docking and virtual screening

We screened all the charged metabolites reported in the Human Metabolome DataBase (HMDB) as target ligands of the TAPT1 model. Molecular docking analysis revealed that flavonoid glycosides (HMDB0124740) have the highest affinity with the lowest binding energy (Table 1). Surprisingly, all the top 10 compounds with the lowest binding energy were found in this subclass. Moreover, the binding site analysis revealed that amino acid A220, W223, T224, I239, R323, K443, S447, L450, N513, and P516 were involved in the interaction with the substrate (Fig 7). The same binding site was also predicted by the CASTp (Supplementary Fig 1).

3.7 Involvement of TAPT1 in different signaling pathways

Ingenuity pathway analysis was performed to understand the relationship between flavonoid and TAPT1, which revealed that flavonoid induced the expression of several genes, including ERK1/2, AHR, HMOX1, CYP1A1/2, and estrogen receptor, which in turn induced CCND1 and CAV1. CCND1 increases the expression of TAPT1, where CAV1 binds to TAPT1 to activate it (Fig 8).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.18.099887; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4. **Discussion**

TAPT1 associated with several diseases, including Osteochondrodysplasia, Complex Lethal, Symoens-Barnes-Gistelinck Type, and Microcephaly caused by a mutation in this protein [3]. Nevertheless, it is still unclear how the mutation in TAPT1 leads to these diseases. It is essential to characterize TAPT1, unravel its substrate(s), and discovering the involved signaling pathways. However, due to the structural complexities, the experimental studies of the membrane proteins are challenging. Computational methods offer an alternative way of studying these proteins. Several studies have so far indicated the efficiency of computational methods in predicting the accurate structures of membrane proteins [30].

Moreover, it is very time consuming and costly to determine the substrates experimentally due to the complexity of the metabolic pathways. Therefore, computational approaches may help predict the potential substrates, which will lead to a valid experimental hypothesis [31]. In this present study, a 3D model was generated for TAPT1, and computational methods were employed to study the protein structurally and functionally. Molecular docking approaches were used to predict the substrates, and a knowledge-based ingenuity pathway was analyzed.

As no crystallographic structure was available for TAPT1, 3D models were generated, and the best one was selected according to the assessment score of validation tools. The results of the validation tools demonstrated that the selected model was reliable. Topology and secondary structure analysis from the primary sequence revealed that the protein had eight transmembrane helices. These helices were also predicted from the 3D structure of the protein. Moreover, it revealed that TAPT1 was localized in the mitochondrial membrane in addition to the basal ciliary body, which was also supported by protein-protein interaction analysis (PPI). The protein was found to interact with proteins localized in both the mitochondria and plasma membrane along with extracellular regions. Also, this interactomics revealed the involvement of TAPT1 with some other functions, e.g., intracellular trafficking and viral transport, which were already reported in several studies. A study conducted in a mouse model stated that TAPT1 provided the entry site or gH receptor for human cytomegalovirus [3, 6].

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.18.099887; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Molecular docking analysis revealed that flavonoids glycosides, as well as flavonoids, would be the potential substrates of TAPT1. The binding site analysis revealed that the amino acid residues involved in the interaction with flavonoids were the same as the pore-lining residues. The amino acid residues H235, R323, K443, N446, S447, L450, K453, S454, Y457, K511, N513, D533, K535, D536, and T538 were found to form the for the surface as well as the binding pocket. Moreover, castP server also predicted the same binding pocket in the TAPT1, which provided strong evidence for this hypothesis. Flavonoids are naturally occurring phenolic compounds that are integral components of the human diet. Several studies revealed that they are involved in the enhancement of bone formation and influence the osteogenic differentiation and mineralization. These compounds also play an essential role in neural crest cell differentiation, proliferation, and survival [32-34]. All of which were disrupted by the mutation in TAPT1 revealed by several studies.

Additionally, mutations in TAPT1 leads to several other diseases related to bone development and disruption in the cilia formation [3]. As flavonoids have a role in bone development and osteogenic differentiation, it was hypothesized that mutations in TAPT1 might disrupt the transfer of flavonoids which caused the osteo-related diseases. Moreover, the ingenuity pathway analysis also provided evidence for the relationship between TAPT1 and flavonoids. IPA revealed that flavonoids could be connected with TAPT1 through several pathways.

5. **Conclusion**

Trans-membrane anterior,-posterior transformation protein 1 (TAPT1) is associated with several ciliopathies and osteogenic diseases. This study provides insights into the structural and functional aspects of the protein. It indicates that flavonoids could be the possible substrate(s) of TAPT1. However, extensive wet-lab experiments are required to validate the outcomes of this study.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.18.099887; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Conflict of interest

The authors declare that they have no competing interests.

Ethical approval

Not required.

Acknowledgments

We wholeheartedly acknowledge the cooperation of Sirazum Nadia Hoque (Dept. Computer Science & Engineering, International Islamic University, Chittagong) for helping us during the project designing.

Funding

No specific grant was received for this study. SA is supported by the (1) Alberta Innovates Graduate Student Scholarship (AIGSS), and the (2) Maternal and Child Health (MatCH) Scholarship programs. MJH receives grant support from the (1) Research Center, Shahjalal University of Science and Technology, (2) Bangladesh Bureau of Educational Information and Statistics, and the (3) Ministry of Education, Government of Bangladesh.

Data availability

All data supporting the findings of this study are available within the article and its supplementary materials

Author contributions

MH conceived the study. MP and SA designed the study. MP, MR, MM, DR conducted the experiments. MP, SA, MR, MM, DR, and PC analyzed and interpreted the data. MP wrote the original draft of the manuscript. MP, SA, PC, and MH reviewed and edited the final manuscript. All authors approved the final manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.18.099887; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Holland, P. W., Booth, H. A. F., & Bruford, E. A. (2007). Classification and nomenclature of all human homeobox genes. *BMC biology*, *5*(1), 47.
2. Kruger, C., & Kappen, C. (2010). Expression of cartilage developmental genes in Hoxc8-and Hoxd4-transgenic mice. *PLoS One*, *5*(2), e8978.
3. Symoens, S., Barnes, A. M., Gistelinck, C., Malfait, F., Guillemyn, B., Steyaert, W., ... & Witten, E. P. (2015). Genetic defects in TAPT1 disrupt ciliogenesis and cause a complex lethal osteochondrodysplasia. *The American Journal of Human Genetics*, *97*(4), 521-534.
4. Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., ... & Olsson, I. (2015). Tissue-based map of the human proteome. *Science*, *347*(6220), 1260419.
5. Baldwin, B. R., Kleinberg, M., & Keay, S. (1996). Molecular cloning and expression of receptor peptides that block human cytomegalovirus/cell fusion. *Biochemical and biophysical research communications*, *219*(2), 668-673.
6. Howell, G. R., Shindo, M., Murray, S., Gridley, T., Wilson, L. A., & Schimenti, J. C. (2007). Mutation of a ubiquitously expressed mouse transmembrane protein (Tapt1) causes specific skeletal homeotic transformations. *Genetics*, *175*(2), 699-707.
7. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, *25*(13), 1605-1612.
8. Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., ... & Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nature methods*, *7*(4), 248-249.
9. Mehmood, M. A., Sehar, U., & Ahmad, N. (2014). Use of bioinformatics tools in different spheres of life sciences. *Journal of Data Mining in Genomics & Proteomics*, *5*(2), 1.
10. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2011). GenBank. *Nucleic acids research*, *39*(Database issue), D32.
11. Jones DT. (1999) Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. 292: 195-202.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.18.099887; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

12. Nugent, T. & Jones, D.T. (2009) Transmembrane protein topology prediction using support vector machines. BMC Bioinformatics. 10, 159. Epub
13. Buchan DWA, Jones DT (2019). The PSIPRED Protein Analysis Workbench: 20 years on. Nucleic Acids Research.
14. Laskowski, R. A., Hutchinson, E. G., Michie, A. D., Wallace, A. C., Jones, M. L., & Thornton, J. M. (1997). PDBsum: a Web-based database of summaries and analyses of all PDB structures. *Trends in biochemical sciences*, *22*(12), 488-490.
15. H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne.(2000) The Protein Data Bank *Nucleic Acids Research*, 28: 235-242.
16. Kim, D. E., Chivian, D., & Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. *Nucleic acids research*, *32*(suppl_2), W526-W531.
17. Laskowski R A, MacArthur M W, Moss D S, Thornton J M (1993). PROCHECK - a program to check the stereochemical quality of protein structures. *J. App. Cryst.*, **26**, 283-291.
18. SC. Lovell, I.W. Davis, W.B. Arendall III, PIW de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson and D.C. Richardson (2002) Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins: Structure, Function & Genetics*. **50**: 437-450.
19. Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein science*, *2*(9), 1511-1519.
20. DeLano, W. L. (2002). Pymol: An open-source molecular graphics tool. *CCP4 Newsletter on protein crystallography*, *40*(1), 82-92.
21. Pellegrini-Calace, M., Maiwald, T., & Thornton, J. M. (2009). PoreWalker: a novel tool for the identification and characterization of channels in transmembrane proteins from their three-dimensional structure. *PLoS computational biology*, *5*(7), e1000440.
22. Oliva, R., Thornton, J. M., & Pellegrini-Calace, M. (2009). PoreLogo: a new tool to analyse, visualize and compare channels in transmembrane proteins. *Bioinformatics*, *25*(23), 3183-3184.
23. Goldberg, T., Hecht, M., Hamp, T., Karl, T., Yachdav, G., Ahmed, N., ... & Bernhofer, M. (2014). LocTree3 prediction of localization. *Nucleic acids research*, *42*(W1), W350-W355.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.18.099887; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

24. Mishra, N. K., Chang, J., & Zhao, P. X. (2014). Prediction of membrane transport proteins and their substrate specificities using primary sequence information. *PLoS One*, *9*(6), e100278.
25. Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., ... & Maitland, A. (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic acids research*, *38*(suppl_2), W214-W220.
26. Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, *31*(2), 455-461.
27. Wishart, D. S., Jewison, T., Guo, A. C., Wilson, M., Knox, C., Liu, Y., ... & Bouatra, S. (2012). HMDB 3.0—the human metabolome database in 2013. *Nucleic acids research*, *41*(D1), D801-D807.
28. Tian, W., Chen, C., Lei, X., Zhao, J., & Liang, J. (2018). CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic acids research*, *46*(W1), W363-W367.
29. Krämer, A., Green, J., Pollard Jr, J., & Tugendreich, S. (2013). Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics*, *30*(4), 523-530.
30. Lehnert, U., Xia, Y., Royce, T. E., Goh, C. S., Liu, Y., Senes, A., ... & Gerstein, M. (2004). Computational analysis of membrane proteins: genomic occurrence, structure prediction and helix interactions. *Quarterly reviews of biophysics*, *37*(2), 121-146.
31. Niu, B., Huang, G., Zheng, L., Wang, X., Chen, F., Zhang, Y., & Huang, T. (2013). Prediction of substrate-enzyme-product interaction based on molecular descriptors and physicochemical properties. *BioMed research international*, *2013*.
32. Weaver, C. M., Alekel, D. L., Ward, W. E., & Ronis, M. J. (2012). Flavonoid intake and bone health. *Journal of nutrition in gerontology and geriatrics*, *31*(3), 239-253.
33. Nones, J., Costa, A. P., Leal, R. B., Gomes, F. C. A., & Trentin, A. G. (2012). The flavonoids hesperidin and rutin promote neural crest cell survival. *Cell and tissue research*, *350*(2), 305-315.
34. Shi, W., Gao, Y., Wang, Y., Zhou, J., Wei, Z., Ma, X., ... & Chen, K. (2017). The flavonol glycoside icariin promotes bone formation in growing rats by activating the cAMP signaling pathway in primary cilia of osteoblasts. *Journal of Biological Chemistry*, *292*(51), 20883-20896.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.18.099887; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1: Top 10 compounds based on the highest binding energy against TAPT1

Sl	HMDB No	Name	Class	Subclass	Binding Affinity (kcal/mol)
1	HMDB0124740	11-[(6-carboxy-3,4,5-trihydroxyoxan-2-yl)oxy]-3-methyl-6-[(3,4,5-trihydroxy-6-{[(3,4,5-trihydroxy-6-methyloxan-2-yl)oxy]methyl}oxan-2-yl)oxy]-7-(3,4,5-trihydroxyphenyl)-2λ⁴,8-dioxatricyclo[7.3.1.0⁵,¹³]trideca-1,3,5(13),6,9,11-hexaen-2-ylium	Flavonoids	Flavonoid glycosides	-11.7
2	HMDB0041762	Peonidin 3-(2-(6-(E)-caffeoyl-beta-D-glucosyl)-beta-D-glucoside) 5-beta-D-glucoside	Flavonoids	Flavonoid glycosides	-11.4
3	HMDB0041161	Cyanidin 3-O-[b-D-Xylopyranosyl-(1->2)-[4-hydroxycinnamoyl-(->6)-b-D-glucopyranosyl-(1->6)]-b-D-galactopyranoside]	Flavonoids	Flavonoid glycosides	-11.4
4	HMDB0041162	Cyanidin 3-O-[b-D-Xylopyranosyl-(1->2)-[(4-hydroxy-3-methoxycinnamoyl)-(->6)-b-D-glucopyranosyl-(1->6)]-b-D-galactopyranoside]	Flavonoids	Flavonoid glycosides	-11.3
5	HMDB0124738	7-{3-[(6-carboxy-3,4,5-trihydroxyoxan-2-yl)oxy]-4,5-dihydroxyphenyl}-11-hydroxy-3-methyl-6-[(3,4,5-trihydroxy-6-{[(3,4,5-trihydroxy-6-methyloxan-2-yl)oxy]methyl}oxan-2-yl)oxy]-2λ⁴,8-dioxatricyclo[7.3.1.0⁵,¹³]trideca-1,3,5(13),6,9,11-hexaen-2-ylium	Flavonoids	Flavonoid glycosides	-11.3
6	HMDB0033021	Pelargonidin 3-O-[4-Hydroxy-3-methoxy-E-cinnamoyl-(->4)-a-L-rhamnopyranosyl-(1->6)-b-D-glucopyranoside] 5-O-b-D-glucopyranoside	Organooxygen compounds	Carbohydrates and carbohydrate conjugates	-11.2
7	HMDB0134843	7-{4-[(6-carboxy-3,4,5-trihydroxyoxan-2-yl)oxy]-3,5-dimethoxyphenyl}-6,11-dihydroxy-3-(4-hydroxyphenyl)-2λ⁴,8-dioxatricyclo[7.3.1.0⁵,¹³]trideca-1(13),2,4,6,9,11-hexaen-2-ylium	Flavonoids	Flavonoid glycosides	-11.2
8	HMDB0134858	7-{4-[(6-carboxy-3,4,5-trihydroxyoxan-2-yl)oxy]-3,5-dimethoxyphenyl}-3-(3,4-dihydroxyphenyl)-6,11-dihydroxy-2λ⁴,8-dioxatricyclo[7.3.1.0⁵,¹³]trideca-1(13),2,4,6,9,11-hexaen-2-ylium	Flavonoids	Flavonoid glycosides	-11.1
9	HMDB0035455	YGM 4B	Flavonoids	Flavonoid glycosides	-11
10	HMDB0038095	Petanin	Flavonoids	Flavonoid glycosides	-11

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.18.099887; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figures

Fig 1: 3D model of TAPT1. (a) The ribbon model where red color represents the helices and green represents the loop. (b) Surface model of TAPT1

Fig 2: 3D model validation assessment results where Ramachandran plot was shown in (a), and error value was shown in (b)

Fig 3: Topology of the TAPT1 protein

Fig 4: Secondary structure prediction from the 3D model where H represents the helices, β represents beta-turn, γ represents the gamma turn, and the yellow line represents the disulfide bond.

Fig 5: Protein-protein interactions of TAPT1. Here, pink lines represent physical interactions. Violet lines represent the co-expression, and gold lines represent the predicted interactions.

Fig 6: Pore of the TAPT1 where (a) shows the pore surface with involved residues (b) pore diameter profile in 3 Armstrong (c) pore line from the surface side and (d) pore line from top to bottom

Fig 7: Docked figures of TAPT1 with its substrates. Here, different colors were used to represent the different pore-lining residues where blue color was used for Lys and His residues and red for Asp and Glu residues. Additionally, green used for Gly, Cys, Ser, Thr, and Tyr residues and gray for the remaining residues.

Fig 8: Hypothetical pathway of TAPT1. Here, arrow line represents the increasing of the expression while dot-headed line represents binding. Additionally, yellow color stated for membrane proteins, pink for receptor, cyan for proteins in cytosol and purple for nucleoplasm.

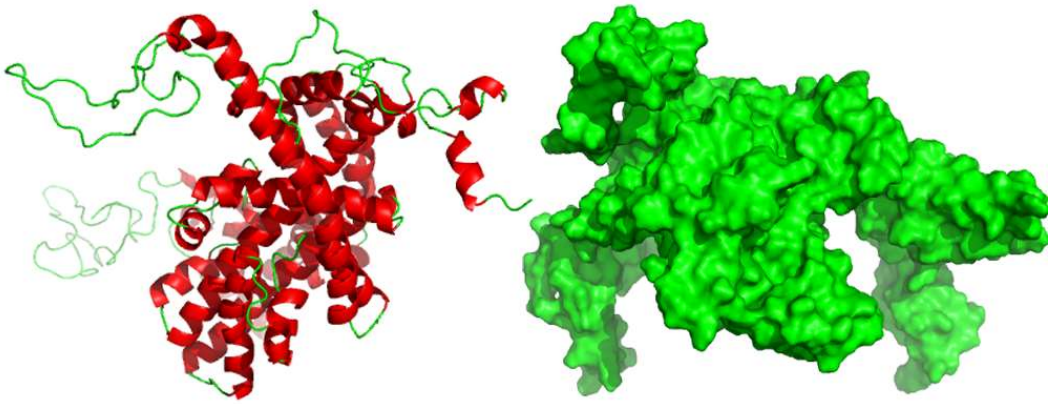
bioRxiv preprint doi: https://doi.org/10.1101/2020.05.18.099887; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary files

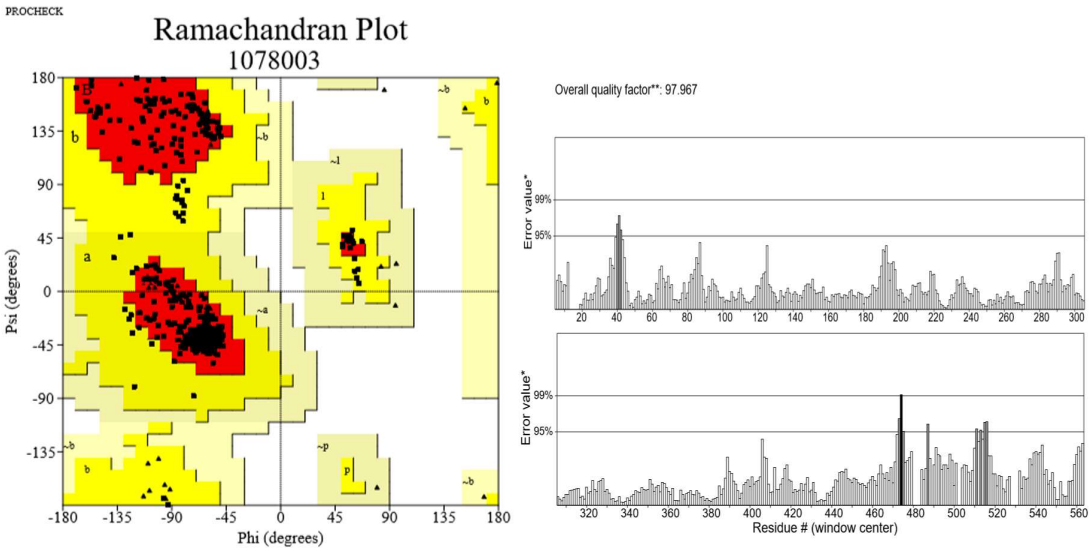
Supplementary Fig 1: The binding pocket of TAPT1 predicted by CASTp server

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.18.099887; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

324



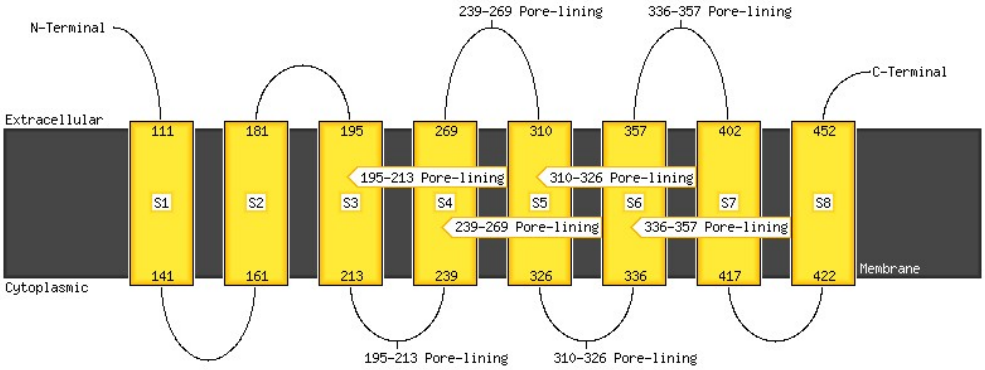
a b

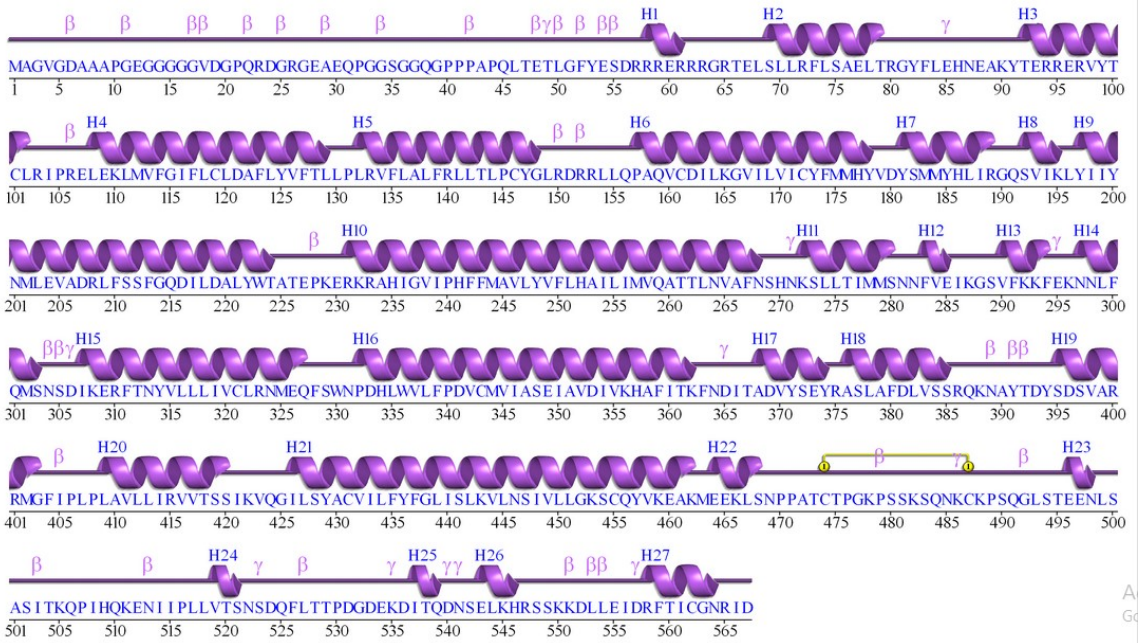


a b

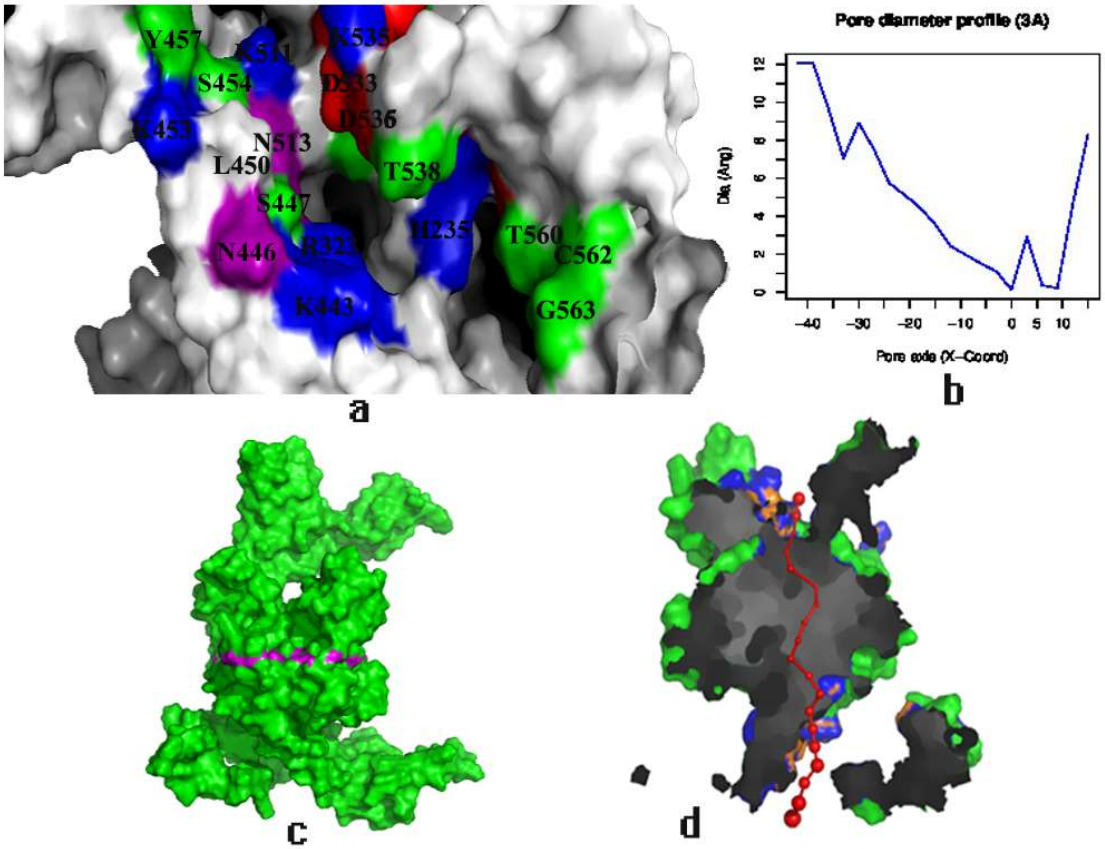
bioRxiv preprint doi: https://doi.org/10.1101/2020.05.18.099887; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

325

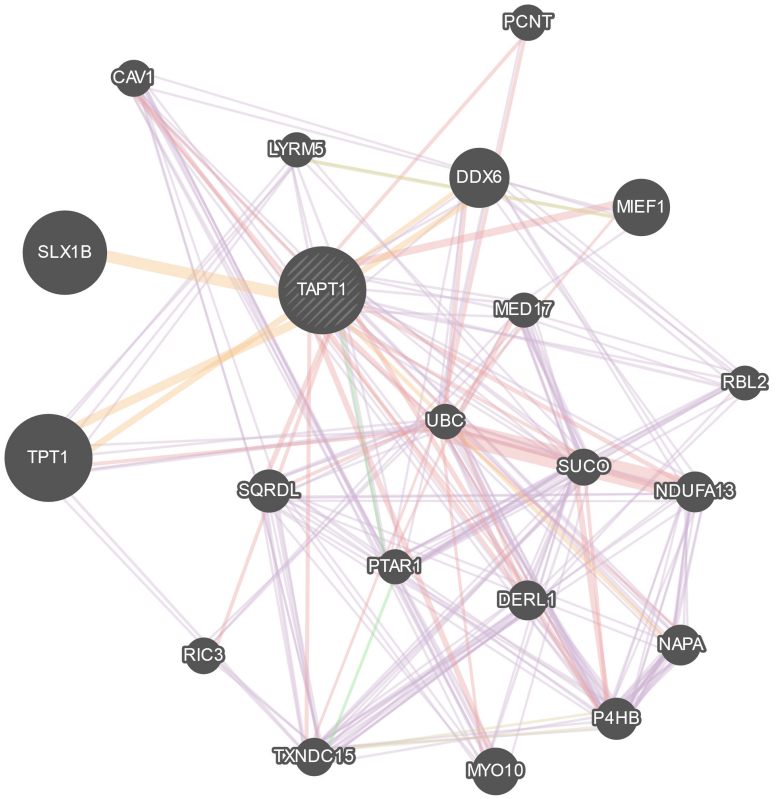


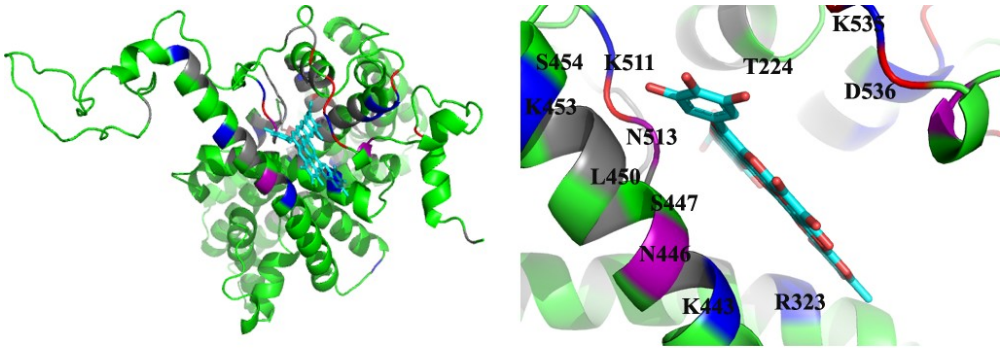
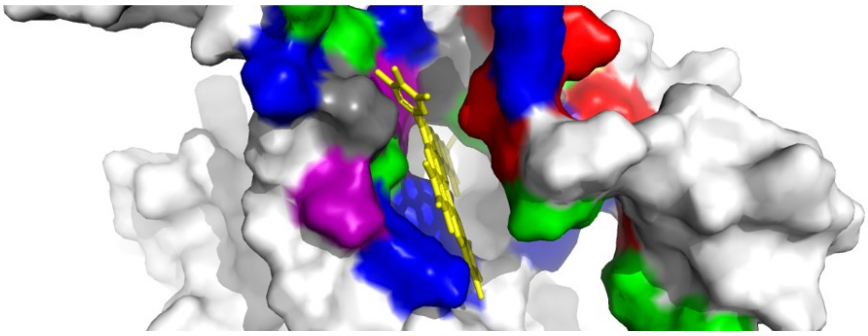
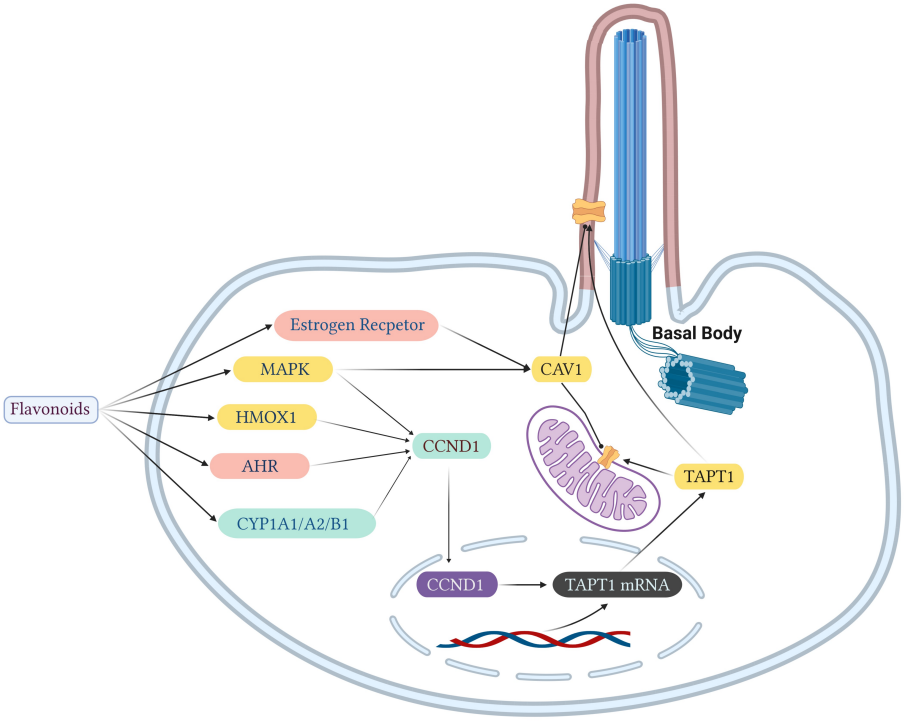


bioRxiv preprint doi: https://doi.org/10.1101/2020.05.18.099887; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



326



bioRxiv preprint doi: https://doi.org/10.1101/2020.05.18.099887; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

327