PROTEINS: Structure, Function, and Genetics 28:465–466 (1997)

EDITOR'S CORNER

Are There Dominant Membrane Protein Families With a Given Number of Helices?

In studies of membrane proteins, a focus has developed on groups of proteins possessing a common number of transmembrane helices, e.g., the seven transmembrane helix receptors. Because recent studies have resulted in complete genome sequences for four different organisms, we have scanned these genomes by using a simple hydrophobicity analysis to determine the distribution of groups of proteins sharing a specific number of transmembrane helices. Somewhat surprisingly, we find no obvious domination of any particular protein family; rather, there is a roughly monotone decrease of number of helices from one to fairly substantial numbers. This simple finding suggests that there are many major groups of membrane proteins still to be characterized, and in which commonalities of function may be found as in the seven and twelve TM cases.

We analyzed the genomes of *Mycoplasma genitalium*[1] and *Haemophilus influenzae*[2] from Bacteria, *Methanococcus jannaschii*[3] from Archaea and *Saccharomyces cerevisiae*[4] from Eukarya. Furthermore, because half of the open reading frames from *Caenorhabditis elegans* have been sequenced,[5] we include an analysis of this genome from a multicellular organism. We used an automated hydrophobicity analysis and the GES scale,[6] taking into account signal sequences.[7] This analysis resulted in the distributions shown in Figure 1. Each protein family, defined as a group of proteins sharing a given number of putative transmembrane helices, is found in a roughly continuous descending incidence from single helix proteins to highly polytopic proteins. An interesting difference is seen for *Caenorhaabditis elegans*, in which a substantial population of proteins having many more helical domains is found compared with the simpler organisms. Our finding emphasizes the relatively unexplored character and large abundance of membrane proteins of different classes coded in organismic genomes.

Isaiah T. Arkin
Department of Molecular Biophysics and Biochemistry
Yale University
New Haven, Connecticut 06520

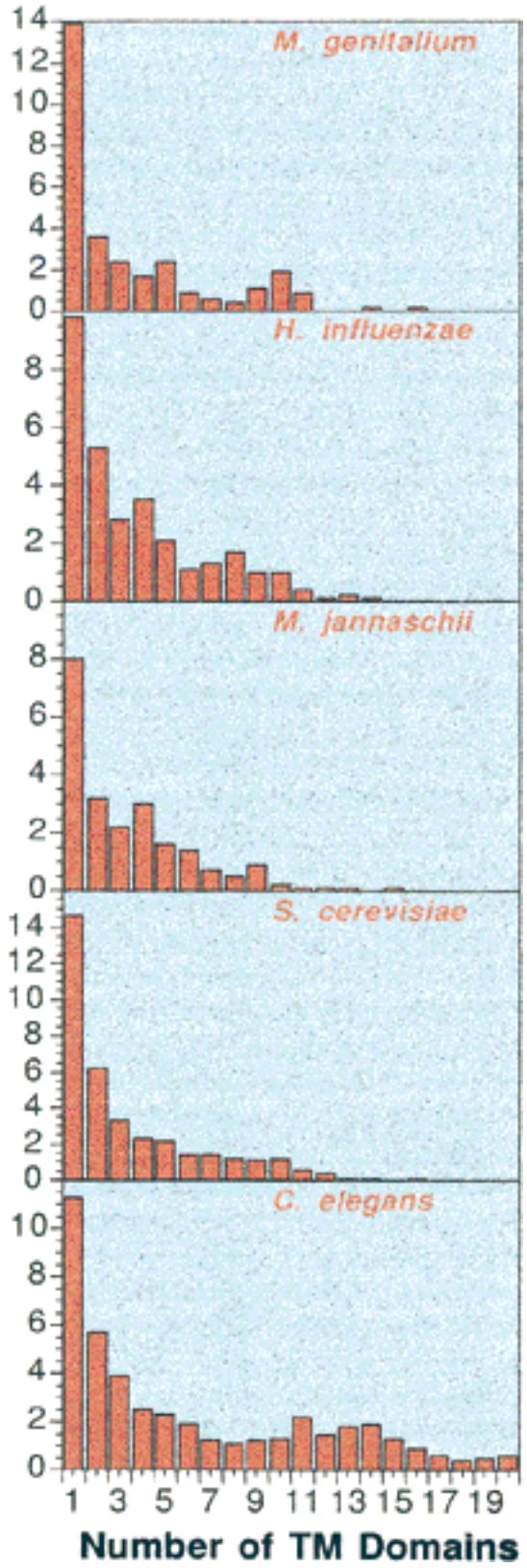
Axel T. Brünger
Department of Molecular Biophysics and Biochemistry
Yale University
New Haven, Connecticut 06520

Donald M. Engelman
Department of Molecular Biophysics and Biochemistry
Yale University
New Haven, Connecticut 06520

REFERENCES

1. Fraser, C.M. Science 270:397–403, 1995.
2. Fleischmann, R.D. Science 269:496–512, 1995.
3. Bult, C.J. Science 273:1058–1072, 1996.
4. Galibert, F. EMBO J. 15:2031–2049, 1996.
5. ftp://ftp.sanger.ac.uk/pub/databases/C.elegans_sequences/
6. Engelman, D.M., Steitz, T.A., Goldman, A. Annu. Rev. Biophys. Biophys. Chem. 15:321–353, 1986.
7. Rusch, S.L., Kendall, D.A. Mol. Membr. Biol. 12:295–307, 1995.

Fig. 1. Relative abundance of different families of membrane proteins, each family defined as containing a common number of putative transmembrane helices. We used a window size of 20 and an energy cutoff of −20 kcal/mol to identify the proteins with the GES scale.[6] Signal sequences, defined as a hydrophobic stretch of 7 or more amino acids within the 25 amino-terminal residues of the protein that is preceded by a net positive charge, were omitted from the analysis.[7]


M. genitalium
H. influenzae
M. jannaschii
S. cerevisiae
C. elegans
1 3 5 7 9 11 13 15 17 19
Number of TM Domains