

Transmembrane Protein Evolution by Internal Gene Duplication

Hironori Mitsuke¹

mitsuke@si.hirosaki-u.ac.jp

Masafumi Arai^{1,2}

d01603@si.hirosaki-u.ac.jp

Keisuke Noto¹

gs02414@si.hirosaki-u.ac.jp

Toshio Shimizu¹

slsimi@si.hirosaki-u.ac.jp

¹ Department of Electronic and Information System Engineering, Faculty of Science and Technology, Hirosaki University, Hirosaki 036-8561, Japan

² Department of Developmental Biology and Neuroscience, Graduate School of Life Science, Tohoku University, Sendai 980-8577, Japan

Keywords: internal gene duplication, transmembrane protein, transmembrane topology, partial sequence

1 Introduction

Recent studies reveal that transmembrane (TM) topology (the number of TM segments (TMSs), the TMS position and the orientation of TMS to the membrane lipid bilayer) evolves by internal gene duplication [1, 6]. A large number of TM proteins, such as Bacteriorhodopsin, Na⁺/Ca²⁺ exchanger and the major facilitator superfamily, have been recognized to have internal repeats that presumably has arisen from the duplication event [5, 6, 7]. In this study, we search comprehensively for sequences with internal repeat by comparing partial sequences (PSs) within a single sequence in prokaryotic genomes, and investigate evolutionary pathways of TM topologies by analyzing duplication patterns obtained.

2 Materials and Methods

Out of 239,359 ORFs in 87 prokaryotic (72 bacterial and 15 archaean) genomes registered in GenBank, we predicted 52,686 TM protein sequences (22%) and their TM topologies by using SOSUI [2], DetecSig [4] and ConPred [3]. Next, PSs containing one or consecutive 2-6 TMSs (1-6tms) were extracted from all the predicted TM proteins. For detecting candidate sequences with internal repeats in these TM proteins, threshold values of sequence identities with enough statistical significance for homologous relationship were determined for 1-6tms PSs (detailed procedure not described here). The thresholds thus obtained are 48.4, 33.8, 29.8, 28.2, 27.1 and 26.2% for 1tms, 2tms, 3tms, 4tms, 5tms and 6tms, respectively. The same-sized PSs were compared one another within a single TM protein sequence to find internal repeats which satisfy the defined conditions, i.e., larger sequence identity between them than the thresholds determined above.

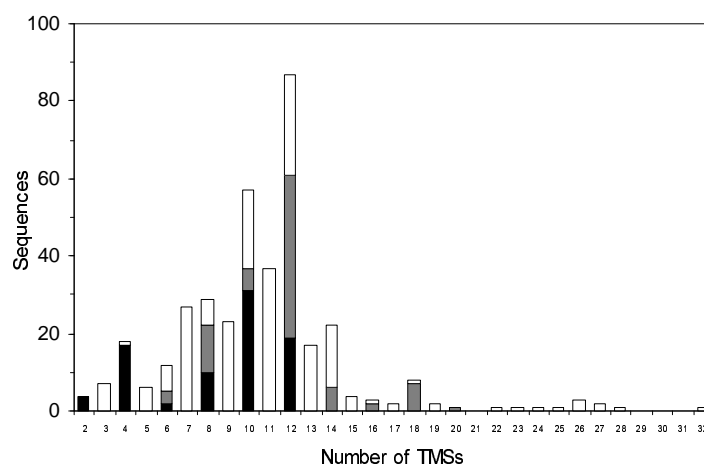


Figure 1: The distribution of the 377 detected TM protein sequences with internal repeat over the number of TMSs, black bars, “diploid-type”; gray bars, “quasi-diploid-type”; white bars, others.

3 Results and Discussion

Finally, 377 TM protein sequences seemed to have evolved by internal duplication were obtained, which were distributed over 70 out of 87 genomes (Fig. 1). Among the detected sequences, various duplication patterns are recognized. One duplication pattern is “diploid-type”, which holds 99.4, 8.3, 34.4, 54.3 and 21.8% in the detected TM protein sequences with 4, 6, 8, 10 and 12 TMSs, respectively. Putting “diploid-type” and “quasi-diploid-type” together, these fractions increase to as much as 99.4, 41.6, 75.8, 64.9 and 70.1%.

Other duplication patterns than “diploid-type” are prevailing in the TM protein sequences with the odd numbers of TMSs. The case of 7tms TM protein is interesting in particular. There are mainly three duplication patterns recognized: from 4tms primordial sequence with the duplication of 3tms element, from 3tms one with the duplication of 3tms element and the generation of a new TMS, and from 5tms to 7tms, as depicted schematically in Figure 2.

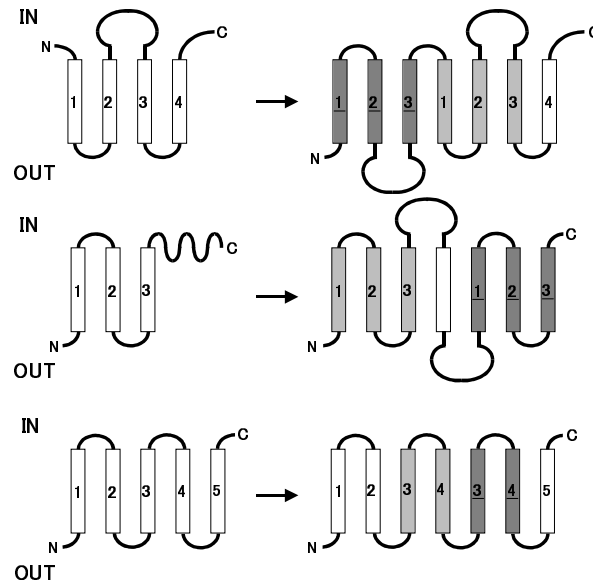


Figure 2: Schematic presentation of three duplication patterns for the evolution of 7TM proteins. The function of the sequences with the first pattern is assigned to “rhodopsin pump” [7], and the second is reported as “zinc transporter”. A similar evolutionary pathway to the second one was reported earlier for “lysosomal cystine transporter” [8]. The function of the third one is not known yet at this stage, being remained for future research.

References

- [1] Arai, M., Ikeda, M., and Shimizu, T., Comprehensive analysis of transmembrane topologies in prokaryotic genomes, *Gene*, 304:77–86, 2003.
- [2] Hirokawa, T., Boon-Chieng, S., and Mitaku, S., SOSUI: classification and secondary structure prediction system for membrane proteins, *Bioinformatics*, 14(4):378–379, 1998.
- [3] Ikeda, M., Arai, M., Lao, D.M., and Shimizu, T., Transmembrane topology prediction method: a re-assessment and improvement by consensus method using a dataset of experimentally-characterized transmembrane topologies, *In Silico Biol.*, 2(1):19–33, 2002.
- [4] Lao, D.M. and Shimizu, T., Methods for detecting the signal peptide in transmembrane and globular proteins, *Genome Informatics*, 12:340–342, 2001.
- [5] Sääf, A., Baars, L., and von Heijne, G., The internal repeats in the $\text{Na}^+/\text{Ca}^{2+}$ exchanger-related *Escherichia coli* protein YrbG have opposite membrane topologies, *J. Biol. Chem.*, 276(22):18905–18907, 2001.
- [6] Saier, M.H., Jr., Tracing pathways of transport protein evolution, *Mol. Microbiol.*, 48(5):114–1156, 2003.
- [7] Taylor, E.W. and Agarwal, A., Sequence homology between bacteriorhodopsin and G-protein coupled receptors: exon shuffling or evolution by duplication?, *FEBS Lett.*, 325(3):161–166, 1993.
- [8] Zhai, Y., Heijne, W.H.M., Smith, D.W., and Saier, M.H., Jr., Homologues of archaeal rhodopsins in plants, animals and fungi: structural and functional predictions for putative fungal chaperone protein, *Biochem. Biophys. Acta*, 1151(2):206–223, 2001.