

Identification and Classification of Eukaryotic GPCRs by Binary Topology Pattern

Yasuhito Inoue

gs02409@si.hirosaki-u.ac.jp

Masami Ikeda

srikeda@cc.hirosaki-u.ac.jp

Toshio Shimizu

slsimi@si.hirosaki-u.ac.jp

Department of Electronic and Information System Engineering, Faculty of Science
and Technology, Hirosaki University, 3 Bunkyo-cho, Hirosaki 036-8561, Japan

Keywords: G-protein coupled receptor, transmembrane topology, binary topology pattern, functional identification

1 Introduction

G-protein coupled receptors (GPCRs) belong to one of the largest superfamilies in mammals, and important to target for drug design in pharmacology field. The heptahelical fold is a common characteristic to various functional groups (or families) of GPCR with eight loop regions protruding at out sides of the lipid bilayer, of which lengths have a large variation from group to group, indicating the function of each GPCR candidate sequence is inferable from its transmembrane (TM) topology [4, 6]. In this study, a TM topology is treated as a binary topology pattern (BTP, a string ‘0’ and ‘1’) by defining threshold lengths for individual eight loops, where the length of each loop is expressed with a binary number: ‘0’ for short loop or ‘1’ for long loop. By applying the BTPs assigned to individual functional groups, we have detected and classified GPCRs encoded in eukaryotic genomes, i.e., of human, mouse, fugu, fly and worm.

2 Datasets and Methods

The BTP for each GPCR functional group was determined from the TM topology data of 828 mammalian GPCR sequences registered in SWISS-PROT 40.0 [1], which were classified in advance according to GPCRDB 6.0 [3]: class A (717 sequences), class B (52), class C (22), Frizzled/Smoothed family (20) and others (non-GPCR, 17). The determination of BTPs was carried out in a step-wise manner: first, for combined groups of (class A + non-GPCR), (class B + class C) and Frizzled/Smoothed, and then for class B and class C separately in the next step, and so on. The obtained BTPs in the first step, for example, are (0, *, *, *, *, *, *, *), (1, *, *, *, *, *, 0, *) and (1, *, 1, *, *, *, 1, *) with threshold lengths, {91, 13, 25, 18, 25, 25, 24, 55} for (A+non-GPCR), (B+C) and Frizzled/Smoothed, respectively. The asterisk in the BTPs means that the binary loop length couldn’t be determined.

ORFs and full-length cDNAs were obtained from GenBank (human, 33,524 sequences), RIKEN (mouse, 10,465), Ensembl (33,609 fugu sequences), GenBank (fly, 7,424) and GenBank (worm, 17,685). We first selected only TM protein sequences from the proteomes by using SOSUI [2]. After detecting and removing the signal peptide region by DetecSig [5], the sequence was subjected to TM topology prediction by HMMTOP 2.0 [7]. Finally, we had 658 (human), 56 (mouse), 418 (fugu), 133 (fly) and 741 (worm) 7-tms sequences as the GPCR candidates. These sequences were then subjected to the step-wise functional identification/classification procedure by BTP.

3 Results and Discussion

The results are summarized in Table 1, together with the results by BLAST search for comparison. In BLAST search, only the best-hit sequences were counted. Out of 658 (human), 56 (mouse), 416 (fugu), 133 (fly) and 741 (worm) candidate sequences, 585, 37, 337, 111 and 672 sequences were identified, respectively, as GPCR by our method. Our method found 20 and 85 olfactory receptor sequences in fly and worm genomes, respectively, while the BLAST search did nothing.

Table 1: Functional identification/classification of GPCRs in human, mouse fugu, fly and worm, by our method and BLAST search (SWISS-PROT 40.30) (in parenthesis).

Class / Subclass	human (658)	mouse (56)	fugu (416)	fly (133)	worm (741)
Class A Rhodopsin like	499 (320)	39 (14)	295 (119)	98 (23)	660 (28)
Group A1	50 (60)	0 (3)	7 (29)	1 (8)	12 (6)
Group A2	234 (172)	13 (4)	78 (54)	51 (13)	243 (19)
Hormone protein	8 (5)	2 (2)	4 (2)	6 (0)	2 (1)
Olfactory	46 (27)	8 (2)	45 (18)	20 (0)	85 (0)
Nucleotide-like	11 (24)	1 (0)	8 (2)	0 (0)	36 (0)
Platelet activating factor	2 (2)	0 (0)	0 (1)	0 (0)	0 (0)
Gonadotropin-releasing hormone	4 (4)	0 (0)	1 (3)	0 (1)	1 (1)
Lysosphingolipid & LPA (EDG)	7 (12)	1 (1)	3 (4)	1 (0)	11 (0)
<i>Class A / unclassified into subclasses</i>	124 (*)	14 (*)	138 (*)	19 (*)	270 (*)
Class B Secretin like	49 (52)	2 (1)	19 (31)	5 (5)	7 (1)
Group B1	38 (38)	1 (0)	16 (18)	5 (0)	5 (1)
Group B2	11 (11)	1 (1)	3 (13)	0 (5)	2 (0)
Class C Metabotropic glutamate / pheromone	23 (24)	0 (0)	16 (29)	3 (3)	1 (2)
Metabotropic glutamate	10 (12)	0 (0)	13 (12)	1 (1)	1 (1)
Extracellular calcium-sensing	3 (3)	0 (0)	0 (16)	0 (0)	0 (1)
GABA-B	8 (9)	0 (0)	0 (0)	0 (2)	0 (0)
<i>Class C / unclassified into subclasses</i>	2 (*)	0 (*)	3 (*)	2 (*)	0 (*)
Class D Fungal pheromone	* (0)	* (0)	* (0)	* (0)	* (0)
Class E cAMP receptors (Dictyostelium)	* (0)	* (0)	* (0)	* (0)	* (0)
Frizzled/Smoothed family	27 (10)	1 (1)	18 (9)	5 (1)	4 (0)
Other GPCR	* (5)	* (0)	* (2)	* (16)	* (54)
Non-GPCR	50 (199)	14 (27)	45 (174)	16 (66)	32 (210)
Not identified	23 (48)	5 (13)	34 (52)	6 (19)	37 (446)

Group A1: Amine, Cannabis and Melatonin. Group A2: Peptide, (Rhod)opsin, Olfactory, Prostanoid and Thyrotropin-releasing hormone & Secretagogue. Group B1: Calcitonin, Corticotropin releasing factor, Gastric inhibitory peptide, Glucagon, Growth hormone-releasing hormone, Parathyroid hormone, PACAP, Secretin, Vasoactive intestinal polypeptide and Diuretic hormone. Group B2: EMR1 and Brain-specific angiogenesis inhibitor.

References

- [1] Bairoch, A. and Apweiler, R., The SWISS-PROT protein sequence database and its supplement TrEMBL in 1999, *Nucleic Acids Res.*, 28(1):45–48, 2000.
- [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] Horn, F., Vrend, G., and Cohen, E.F., Collecting and harvesting biological data: The GPCRDB and NucleaRDB information systems, *Nucleic Acids Res.*, 29(1):346–349, 2001.
- [4] Inoue, Y., Sugiyama, Y., Ikeda, M., and Shimizu, T., Classification of eukaryotic 7-tms transmembrane proteins by binary topology pattern, *Genome Informatics*, 13:336–337, 2001.
- [5] Lao, D. M. and Shimizu, T., Methods for detecting the signal peptide in transmembrane and globular proteins, *Genome informatics*, 13:340–342, 2001.
- [6] Sugiyama, Y., Arai, A., and Shimizu, T., Comprehensive functional identification of prokaryotic transmembrane proteins by binary topology pattern, *Genome Informatics*, 13:334–335, 2001.
- [7] Tusnady, G.E. and Simon, I., The HMMTOP transmembrane topology prediction server, *Bioinformatics*, 17(9):849–850, 2001.