Discovery and Classification of Peptide Family G-Protein Coupled Receptors in the Human Genome Sequence

Maria E. Levchenko maria@kuicr.kyoto-u.ac.jp
Toshiaki Katayama katayama@kuicr.kyoto-u.ac.jp
Minoru Kanehisa kanehisa@kuicr.kyoto-u.ac.jp

Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 611- 0011, Japan

Keywords: GPCR, peptide receptor family, human draft genome

1 Introduction

The G-protein coupled receptor (GPCR) superfamily is one of the largest protein superfamilies, which is involved in a large number of important physiological processes in the cell [1]. The peptide receptor family, which belongs to this superfamily, is in turn a large and diversified group of receptors and represented by a huge variety of receptors such as neuropeptide, chemokine, and melacortin receptors. The peptide receptor family plays an important role in normal and pathological cell physiology and therefore represents an attractive target for pharmaceutical investigation and drug design.

Since the human genome draft sequence became available [2], we started working on to find all possible peptide receptor candidates in the human genome and to fully systematize the finding process. In contrast to previous studies on GPCR candidate prediction [3], the present study focuses on a specific family of GPCR and an increase of selectivity to a maximal degree. Also full and comprehensive classifications were performed based on ligand-specificity of found peptide receptor candidates.

2 Method

First, all known amino acid sequences of peptide receptors for all species were collected from SWISS-PROT (release 39.0) and TrEMBL databases, excluding fragments and duplicate entries. The obtained set of amino acid sequences (498 entries, including 105 human sequences) was used as a query set for subsequent search.

Next, all human amino acid sequences were collected from GenBank (release 124.0) and human draft genome (Ensembl, [4]) databases with lengths of no less than 230 amino acid residues (approximate GPCR minimal length). Then GPCR candidates containing seven transmembrane segments (7-TMS) were selected by using hidden Markov models of Pfam 5.0 (E value < 0.05). A number of 7-TMS proteins appear to be unrelated to GPCRs (for example, the KDEL/HDEL family, etc.), and we used those membrane proteins as a negative control in order to verify the obtained data. As a result of these preparatory procedures we obtained 1333 7-TMS-containing sequences. After merging them with well-known peptide receptor sequences collected earlier from SWISS-PROT and TrEMBL, we calculated similarity score for each pair of sequences by the SSEARCH program. Finally, the hierarchical cluster analysis (complete linkage method) was performed based on found similarity scores.

3 Results

When the cluster analysis was performed for the whole data set, there was a clear tendency of separation among peptide receptors according to their ligand-specificity. So we obtained 23 distinct subgroups of peptide receptors, which corresponded to different classes of peptide ligands. After visual inspection of the obtained cluster dendrogram it became possible not only to find out new peptide

receptor candidates, but also to reveal their probable biological functions in the cell according to their ligand-specificity.

The results of our findings are summarized in Table 1.

Table 1: Functional classification of human peptide receptors according their ligand-specificity.

Ligand class	SP & Tr	Ensembl*	GenBank*	Ligand class	SP & Tr	Ensembl*	GenBank*
Adrenomedullin	1	3/2	3/3	Melanocortin	7	5/1	15/1
Angiotensin	3	2/1	5	Neuromedin U	4	4/2	8/5
APJ like	1	3	1	Neuropeptide Y	7	7/3	17/4
Bombesin	3	4	3	Neurotensin	2	3/1	2
Bradykinin	2	2	5	Opioid	5	3/1	11
C5a anaphylatoxin	2	3/1	5/2	Orexine & NPFF	5	3/1	6
CCK	3	5	7	Proteinase activ.	3	3/1	6
Chemokine	26	22/6	82/30	Somatostatin	6	6	9
Endothelin	4	5/2	8	Tachykinin	6	4	8/1
Fmet-leu-phe	4	3	8	Thrombin	1	1	2
Galanin	3	4/1	6/2	Vasopressin-like	5	5	6
Interleukin-8	2	2	3	*Total*	*105*	*102/23*	*226/48*

*The number of entries that were not annotated before (discovered in the present study) is shown after the slash.

As a result of clustering, we postulate that the total number of functional peptide receptors in the human genome is 350. This number includes 71 newly discovered receptors – 48rom GenBank and 23 from human draft genome – which are not overlapping and consist exclusively of previously unannotated entries.

Acknowledgements

This work was supported by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan, the Japan Society for the Promotion of Science, and the Japan Science and Technology Corporation. The computational resources were provided by the Bioinformatics Center, Institute for Chemical Research, Kyoto University.

References

[1] Bockaert, J. and Pin, J. P., Molecular tinkering of G-protein-coupled receptors: an evolutionary progress, *EMBO J.*, 18(7): 1723–1729, 1999.

[2] Pennisi, E., Human genome. Finally, the book of life and instructions for navigating it, *Science* 288:2304–2307, 2000.

[3] Suwa, M., Sato, T., Okouchi, I., Mikoshiba, M., Akiyama, Y., Asai, K., Matsumoto, S., Tsutsumi, S., and Aburatani, H., Gene discovery of G-protein coupled receptors from human genome, *Genome Informatics,* 11:410–411, 2000.

[4] http://us.ensembl.org