Gene 503 (2012) 92–100

Contents lists available at SciVerse ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Insight into pattern of codon biasness and nucleotide base usage in serotonin receptor gene family from different mammalian species

J. Febin Prabhu Dass [a], C. Sudandiradoss [a,b,*]

[a] Bioinformatics Division, School of Biosciences and Technology, VIT University, Vellore, Tamil Nadu State, 632014, India
[b] Centre for Nanobiotechnology, VIT University, Vellore, Tamil Nadu State, 632014, India

ARTICLE INFO

Article history:
Accepted 17 March 2012
Available online 27 March 2012

Keywords:
Serotonin receptors
5-HT family
Codon usage bias
Correspondence analysis
Compositional mutation bias
Gene function

ABSTRACT

5-HT (5-Hydroxy-tryptamine) or serotonin receptors are found both in central and peripheral nervous system as well as in non-neuronal tissues. In the animal and human nervous system, serotonin produces various functional effects through a variety of membrane bound receptors. In this study, we focus on 5-HT receptor family from different mammals and examined the factors that account for codon and nucleotide usage variation. A total of 110 homologous coding sequences from 11 different mammalian species were analyzed using relative synonymous codon usage (RSCU), correspondence analysis (COA) and hierarchical cluster analysis together with nucleotide base usage frequency of chemically similar amino acid codons. The mean effective number of codon (ENc) value of 37.06 for 5-HT_6 shows very high codon bias within the family and may be due to high selective translational efficiency. The COA and Spearman's rank correlation reveals that the nucleotide compositional mutation bias as the major factors influencing the codon usage in serotonin receptor genes. The hierarchical cluster analysis suggests that gene function is another dominant factor that affects the codon usage bias, while species is a minor factor. Nucleotide base usage was reported using Goldman, Engelman, Stietz (GES) scale reveals the presence of high uracil (>45%) content at functionally important hydrophobic regions. Our *in silico* approach will certainly help for further investigations on critical inference on evolution, structure, function and gene expression aspects of 5-HT receptors family which are potential antipsychotic drug targets.

1. Introduction

Serotonin or 5-HT is a ubiquitous mono amine neurotransmitter found in both the animal and human nervous systems. Serotonin performs a crucial role in several neurological and physiological conditions such as cardiac development and function, gastrointestinal function and liver regeneration. It is also involved in aggression, substance abuse, addictive behavior, anxiety, depression, obsessive control and learning (Lucki, 1998). The serotonin receptors have been implicated in pathophysiological roles (Meltzer et al., 2003) located in both presynaptic and postsynaptic neurons, which are crucial for binding the endogenous natural neurotransmitter, serotonin. These receptors (excluding the ionotropic 5-HT_3) belong to the metabotropic receptors, constitutes largest family of G-protein coupled receptors (GPCRs). The activated neurons transmit extracellular serotonin signals to the intracellular environments through 5-HT receptors, which are available in multiple classes and subclasses. In mammals, there are six major families of GPCR 5-HT receptors, namely 5-HT_1, 5-HT_2, 5-HT_4, 5-HT_5, 5-HT_6, and 5-HT_7. The 5-HT_1 family contains 5-HT_{1A}, 5-HT_{1B}, 5-HT_{1D}, 5-HT_{1E}, and 5-HT_{1F} subclasses, and 5-HT_2 family contains the 5-HT_{2A}, 5-HT_{2B} and 5-HT_{2C} subclasses that are widely distributed within the central nervous system and smooth muscles. The 5-HT_5 receptor family contains 5-HT_{5A} and 5-HT_{5B} receptor subclasses which are largely distributed in central nervous system. Only the 5-HT_{5A} was included in our analysis due to the non availability of 5-HT_{5B} subclass gene. The 5-HT_4, 5-HT_6 and 5-HT_7 receptors do not contain any subclasses and they are widely distributed in smooth muscles and brain regions (Nichols and Nichols, 2008). Most of the mammalian 5-HT receptor sub families share good amino acid sequence homology. Their signaling function involves by either inhibiting or stimulating the adenylate cyclase via G-proteins.

Genomic information encoded by nucleotides are transformed to their corresponding proteins through triplet genetic code referred as codons. These triplets specifying the same amino acid are called synonymous codons whereas variation in relative codon frequency is termed as the codon usage bias (Grantham et al., 1980). Codon usage is highly variable among different species (Sharp et al., 1988) and is principally related to gene function (Fuglsang, 2003; Liu et al.,

Abbreviations: 5-HT, 5-Hydroxy tryptamine; GPCR, G-Protein Coupled Receptor; 7-TM, 7-Transmembrane; CDS, Coding Sequences; RSCU, Relative Synonymous Codon Usage; COA, Correspondance Analysis; SD, Standard Deviation; ENc, Effective Number of Codons; GES, Goldman Engelman Stietz.

* Corresponding author at: Bioinformatics Division, School of Biosciences and Technology, VIT University, Vellore, Tamil Nadu State, 632014, India. Tel.: +91 9443072826; fax: +91 4162243092.
E-mail address: csudandiradoss@vit.ac.in (C. Sudandiradoss).

0378-1119/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2012.03.057

2005; Ma et al., 2002). However, these variations are also found within the related species and genes (Ma et al., 2009). Research suggests that there are high correlations between codon usage bias and tRNA abundance (Duret, 2000; Ikemura, 1981, 1985; Kanaya et al., 1999), gene length (Duret and Mouchiroud, 1999; Eyre-Walker, 1996) and gene expression level (Carrie et al., 2011; Grosjean and Fiers, 1982; Merkl, 2003; Sharp et al., 1993). Further, codon bias pattern is closely related to protein tertiary structure (Gu et al., 2004). In bacteria, both the mutation bias and natural selection pressure have been reported (Liu et al., 2010) along with the translational efficiency determinants (Gingold and Pilpel, 2011). Moreover, several factors that influence the evolution of codon biasness have been recently reviewed in considerable detail (Palidwor et al., 2010; Plotkin and Kudla, 2011; Sharp et al., 2010). Extensive codon usage studies have been conducted on "deep-branching species" such as viruses (Emily et al., 2010), bacteria (Bailly-Bechet et al., 2006), yeast (Freire-Picos et al., 1994), *Caenorhabditis elegans* (Stenico et al., 1994) and *Arabidopsis thaliana* (Chiapello et al., 1998). However, few such studies have been done on the vertebrate species such as mammals (Eyre-Walker, 1991; Sharp et al., 1995; Sueoka and Kawanishi, 2000) and rodents (Smith and Hurst, 1999). Although, the codon usage patterns in primate species (McWeeney and Valdes, 1999) and mammalian genes (Qin et al., 2009) are available, we tried for the first time the codon bias analysis on serotonin receptors across the mammalian species.

The human 5-HT receptor subtypes share a minimum of 25% and a maximum of 69% sequence identity (Barnes and Sharp, 1999). Regardless of this huge sequence variation, most of the 5-HT receptors are evolutionarily conserved to perform similar function and maintain the structural topology of 7-transmembrane among diverse species ranging from *C. elegans to Homo sapiens*. Because of this reason, it has been speculated that serotonin receptor family may have appeared more than 700–750 million years ago and predates the evolution of other biogenic amine receptors (Peroutka and Howell, 1994). New 5-HT receptors are also being continually added and annotated in the public databases from various genome sequencing projects. This provides us the opportunity to investigate the codon usage of these multi-class receptors within and between species.

Even though evolutionary reports are available based on the pharmacological characterization of serotonin receptors, no one has reported on their patterns of codon usage. In this work, we analyzed in-depth codon and nucleotide usage together with allied factors. Our analysis will surely provide a crucial and novel insight into the function, structural and evolutionary significance of serotonin receptors within mammalian species. Hence, we have conducted a study on 11 mammalian species spanning around 110 peptide coding mRNA sequences. The main objective of our study is to computationally investigate the factors and patterns driving the serotonin receptor genes across mammalian lineages with the aid of most powerful statistical tools.

We applied several methods to measure the synonymous codon usage bias namely relative synonymous codon usage (RSCU), major indices such as GC, GC1, GC2, GC3, effective number of codons (ENc), correspondence analysis (COA), hierarchical cluster analysis and nucleotide base usage by Goldman, Engelman, Stietz (GES) scale. These methods are most widely used since they have numerous advantages over other methods and till date it is considered as one of the best informatics and statistical protocol for analyzing codon biasness among various organisms and mammalian species (Carrie et al., 2011). RSCU is defined as the frequency of a codon divided by the expected frequency if all its synonyms for that amino acid were used equally (Sharp et al., 1986). Thus, RSCU values close to 1.0 indicate a lack of bias for that codon. The ENc is used to measure the degree of departure from the equal use of synonymous codons of coding regions. The values of ENc could theoretically range from 20 (one amino acid using one codon) to 61 (all codons used with equal probability). Hence, we have used more reliable software such as SPSS version 13.0, CodonW version 4.3 and Biopython version 1.53 to evaluate the factors affecting the codon and nucleotide base usage in 110 homologous coding sequences of serotonin receptors from 11 different mammalian species. Our results will certainly help to understand the molecular mechanisms related to serotonin receptors.

2. Materials and methods

2.1. Coding sequence data

A total of 110 serotonin receptors comprising 12 sub-families from 11 mammalian species were used in this study. We retrieved the human 5-HT receptor protein sequences from IUPHAR database (Anthony et al., 2009). Then, we have performed BLASTp (Altschul et al., 1990) search using human protein sequences as reference and obtained the remaining mammalian homologous protein sequences. Protein sequence similarity has revealed better search results with the good e-value across species. The different mammal species included in this study were, human(*Homo sapiens*), chimpanzee(*Pan troglodytes*), rhesus monkey(*Macaca mulatta*), cattle(*Bos Taurus*), horse(*Equus caballus*), pig(*Sus scrofa*), dog(*Canis lupus familiaris*), guinea pig(*Cavia porcellus*), opossum(*Monodelphis domestica*), rat(*Rattus rattus*) and mouse(*Mus musculus*). We have restricted the number of species based on the availability of ≥ 6 serotonin receptor genes in a species. Finally, we obtained the corresponding mRNA coding sequence (CDS) from DDBJ/EMBL/GenBank international nucleotide sequence databanks for the resultant protein sequences from BLASTp. Utmost care was taken in constructing the dataset, which consists of CDS with exclusion of all partial and internal stop codon sequences. The complete detail of the dataset analyzed in this study was given in Table 1.

2.2. Analysis of relative synonymous codon usage in serotonin receptor genes

The RSCU (Sharp and Li, 1987) method was applied to calculate the codon usage for the dataset. It is a simple yet powerful method to identify the high-frequency or preferred codons in serotonin receptor sequences. Apart from reporting the preferred codons, we can also find whether these codons are GC ending or AT ending nucleotide for a specific amino acid. The RSCU values obtained are the number of time the particular codon is observed, relative to the number of times the codon would be observed in the absence of any codon usage bias. This has been calculated by,

$$RSCU_{ij} = \frac{obs_{ij}}{aa_{ij} \Big/ _{k}} = k * \frac{obs_{ij}}{aa_{ij}}.$$

Where $RSCU_{ij}$ corresponds to the relative synonymous codon usage frequency of codon j in sequence i. obs_{ij} denotes the observed number of codon j for amino acid i, which is encoded by aa_i synonymous codons in a specific gene sequence. Similarly, k represents the number of synonymous codons of codon j. RSCU values provide a normalized codon usage value for the complete serotonin receptor family of all the species included in our study. The RSCU result was interpreted as follows, for a corresponding codon, if the RSCU value is >1.0 - indicates the most frequent usage of that particular codon and if the value is <1.0 - vice versa (Paul and Wen-Hsiung, 1986) for a particular amino acid.

2.3. Gene indices for studying heterogeneity and mutation pressure within the serotonin receptor family

Well documented research reports have been available for the existence of codon usage variation within and between genome