Gene 590 (2016) 35–43

Contents lists available at ScienceDirect

Gene

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

Research paper

Cytochrome P450 genes in coronary artery diseases: Codon usage analysis reveals genomic GC adaptation

Arup Kumar Malakar, Binata Halder, Prosenjit Paul, Supriyo Chakraborty *

Department of Biotechnology, Assam University, Silchar 788011, Assam, India

ARTICLE INFO

Article history:
Received 27 February 2016
Received in revised form 12 April 2016
Accepted 3 June 2016
Available online 5 June 2016

Keywords:
Codon usage bias
Cytochrome P450
Coronary artery disease

ABSTRACT

Establishing codon usage biases are imperative for understanding the etiology of coronary artery diseases (CAD) as well as the genetic factors associated with these diseases. The aim of this study was to evaluate the contribution of 18 responsible cytochrome P450 (*CYP*) genes for the risk of CAD. Effective number of codon (Nc) showed a negative correlation with both GC3 and synonymous codon usage order (SCUO) suggesting an antagonistic relationship between codon usage and Nc of genes. The dinucleotide analysis revealed that CG and TA dinucleotides have the lowest odds ratio in these genes. Principal component analysis showed that GC composition has a profound effect in separating the genes along the first major axis. Our findings revealed that mutational pressure and natural selection could possibly be the major factors responsible for codon bias in these genes. The study not only offers an insight into the mechanisms of genomic GC adaptation, but also illustrates the complexity of *CYP* genes in CAD.

© 2016 Elsevier B.V. All rights reserved.

1. Background

Coronary artery disease (CAD) is a type of cardiovascular diseases (CVD) that affect the structures or functions of the heart and blood vessel. CAD encompasses numerous problems, many of which are related to a process called atherosclerosis (build-up of fatty acid deposits inside an artery). CVDs are the number one cause of death for both men and women, more people die annually from CVDs than from any other disease. Cytochrome P450 (*CYP*) is a multigene superfamily of heme-thiolate enzymes, therefore hemoproteins, which catalyze the oxidation of endogenous and exogenous compounds. Numerous studies have demonstrated that *CYP*-mediated eicosanoid metabolism may be a viable clinical therapeutic strategy for the management of CVDs (Deng et al., 2010; Schuck et al., 2013). Some *CYP* genes are highly expressed in liver and some in heart. Genetic variations, interactions, or pathophysiological factors can lead to reduced, absent, or increased enzymatic activity. Genetic epidemiology studies have demonstrated that functional polymorphisms in *CYP* genes (e.g., *CYP*2J2 and *CYP*2C8, *CYP*4F2 and *CYP*4A11) are connected with the advancement of CVDs (Spiecker et al., 2004; Ward et al., 2008; Schuck et al., 2013).

It is essential to predict the level of gene expression in such diseases, which is quite challenging, to investigate experimentally in the absence of theoretical guidance owing to a large number of potential effects. Thus, an initial step in such investigations is to explore the potentially significant correlations between synonymous codon usage and the genomic features, which can further guide *in-silico* experimental designs to produce synthetic genes. Codon usage bias (CUB) has been proposed to be associated with diseases, due to synonymous mutation at the silent codon positions. CUB not only affects disease risk, but also plays a major role in how patients respond to medication, whether medications cause adverse effects, and how the disease may progress. CUB or simply codon bias influences gene expression pattern via adaptable codon usage throughout the gene (Chevance et al., 2014). Many studies on CUB confirm the correlation of the complex process of CUB with various biological factors, such as gene expression level (Crameri et al., 1996; Deml et al., 2001; Geddie and Matsumura, 2004; Chevance et al., 2014), gene length (Hooper and Berg, 2000; Qi et al., 2015), gene translation initiation signal (Lobo et al., 2006), protein amino acid composition (Sharp and Li, 1987), protein structure (Wright, 1990), tRNA abundance (Svejstrup, 2002; Fujimori et al., 2005), mutation frequency and patterns (Kyte and Doolittle, 1982), and GC composition (Fox and Erill, 2010).

Nucleotide composition varies greatly between different genomic regions in many eukaryotes. In vertebrates, the variation of G + C content occurs over scales of hundreds of kilobases to megabases, the so-called 'isochore structure' (Bernardi, 2000; Costantini and Bernardi, 2008). The G + C composition resulting from mutational bias and/or translational selection is hypothesized to determine the major trends

Abbreviations: CAD, Coronary artery disease; CVD, Cardiovascular diseases; *CYP*, Cytochrome P450; CUB, Codon usage bias; CAI, Codon adaptation index; Nc, Effective number of codon; RSCU, Relative synonymous codon usage; PCA, Principle component analysis; SCUO, Synonymous codon usage orders.

* Corresponding author.
E-mail addresses: supriyoch2008@gmail.com, supriyoch_2008@rediffmail.com (S. Chakraborty).

http://dx.doi.org/10.1016/j.gene.2016.06.011
0378-1119/© 2016 Elsevier B.V. All rights reserved.

in codon usage by high or low G + C organisms (Knight et al., 2001). The genomes of vertebrates are highly heterogeneous in G + C content and have acquired GC rich regions (Webster et al., 2006). The fluctuation of GC3 along the coding sequences was initially reported from the analysis of a set of phage and bacterial genes, obtained by sliding window analyses (Wada and Suyama, 1985). It is often seen that when GC3 (i.e. the content of G + C at third codon positions) is skewed away from equal usage of GC or AT, the codon bias is the greatest (Mank et al., 2007; Rao et al., 2011). To investigate the patterns and causality of codon bias, many indices have been proposed to determine the degree and direction of codon bias (Sharp and Li, 1987; Wright, 1990).

Logical studies on codon usage assisted in understanding the expression levels of the genes, provided clues about the evolution of the genes and genomes, enhanced immunogenicity of vaccines, etc. (Crameri et al., 1996). Evolutionary investigations on the well documented unevenness of the usage of synonymous codons confirms that codon usage varies between the genes, genomes as well as within a single gene (Hooper and Berg, 2000), which reflect the complex balance among biases generated by mutation, selection and random genetic drift (Lobo et al., 2006; Qi et al., 2015). The codon adaptation index (CAI) was proposed as a measure of codon usage in a gene relative to that in a reference set of genes (Sharp and Li, 1987). CAI correlates better with mRNA expression levels than other codon usage indices, such as the frequency of optimal codons (Fop) (Ikemura, 1985) or the effective number of codons (Nc) (Wright, 1990). Therefore, the study on CUB not only offers an insight into the codon usage bias of the diseases and their mechanisms, but also might help in the possible cures for these diseases. If selection is acting on gene sequences, then we expect them to have been modified to either maximize or minimize the expression efficiency for preventing the disease. For such assumption the effect should be predominantly noticeable in highly expressed genes.

Several studies were conducted using *CYP* genes in CAD (Theken et al., 2012; Schuck et al., 2013). However, till date no studies related to codon usage bias of *CYP* genes in CAD have been carried out. In order to elucidate such findings, in the context of translational selection and mutational bias of *CYP* genes based on bioinformatics approaches, the present study is undertaken.

2. Materials and methods

2.1. Sequence data

The coding sequences (cds) of 18 *CYP*s were retrieved from NCBI (www.ncbi.nl.nih.gov). In order to minimize sampling errors we have taken only those cds which start with the correct initial codon and end with correct termination codon having exact multiple of 3 in total base count.

2.2. Effective number of codon (Nc)

Nc is generally used to measure the codon usage bias of a gene that is independent of the gene length and number of amino acids. Wright, 1990 proposed the measure Nc, which can be used to measure codon usage bias (Wright, 1990). The value of Nc ranges from 20 (when only one codon is used for coding an amino acid) to 61 (when all synonymous codons are used equally to code for an amino acid). It is mathematically expressed as:

$$Nc = 2 + \frac{9}{F2} + \frac{1}{F3} + \frac{5}{F4} + \frac{3}{F6}$$

Where, *Fk* ($k = 2, 3, 4$ or 6) is the average of the *Fk* values for *k*-fold degenerate amino acids. The *F* value denotes the probability that two randomly chosen codons for an amino acid with two codons are identical.

2.3. Relative synonymous codon usage (RSCU)

Sharp et al., in 1986 (Sharp and Li, 1986) proposed that relative synonymous codon usage is the measure of usage of preferred or specified synonymous codons encoding an amino acid and the amino acids may be biased or used randomly. It expresses the relationship between the observed frequency of codons and the expected frequency in case of random usage of preferred synonymous codons. It is mathematically expressed as:

$$RSCU_{ij} = \frac{X_{ij}}{\frac{1}{ni}\sum_{j=1}^{ni} X_{ij}}$$

Where *Xij* is the frequency of the *j*th codon for the *i*th amino acid and *ni* being the number of alternative synonymous codons available for the *i*th amino acid. Codons with RSCU value higher than 1 are used more frequently than expected and are referred to as a positive RSCU, whereas, codons with value less than 1 are used less frequently than expected and are referred to as negative codon usage bias. Codons with value close to 1 follow random usage or used equally.

2.4. Codon Adaptation Index (CAI)

Sharp and Li (1987) proposed that CAI is an effective measure of codon bias in prokaryotes and eukaryotes. CAI is a measurement of the relative adaptiveness of the codon usage of a gene towards the codon usage of more expressed genes. CAI values range from 0 to 1, with higher values towards 1 indicating a higher proportion of the most abundant codons. The CAI is calculated as:

$$CAI = \frac{\exp 1}{L}\sum_{k=1}^{L} \ln Wc(k)$$

Where L is the number of codons in the gene and $Wc(k)$ is the w value for the *k*th codon in the gene. The CAI defines the frequent codons in highly expressed genes as the translationally optimal codons (Sharp and Li, 1987).

2.5. Dinucleotide odds ratio

The odds ratio calculation is generally used to compute the dinucleotides in gene sequences. Odds ratio is the likelihood of observing a dinucleotide in a sequence and is calculated as

$$P_{xy} = \frac{f_{xy}}{f_x f_y}$$

Where *x* and *y* stand for the nucleotides that form dinucleotide *xy*; and *fx*, *fy*, *fxy* denote the frequencies of nucleotide *x*, nucleotide *y*, and dinucleotide *xy* respectively. Karlin et al. showed that dinucleotides with an odds ratio falling outside the range [0.78, 1.25] could be considered as being more under-or over-represented than normal (Karlin et al., 1998).

2.6. Principle component analysis (PCA)

PCA is a multivariate statistical method for simplifying the multidimensional information of the data matrix into a two-dimensional map (Morrison, 1990). This method guarantees that the highest principal components, usually the first and the second, contain as much information as possible. Further components will recover smaller and smaller amounts of the residual variation. Given the eigen values of the data matrix, the percentages of the information recovered by the various components can be calculated.

Download English Version:

https://daneshyari.com/en/article/2814869

Download Persian Version:

https://daneshyari.com/article/2814869

Daneshyari.com