Mitochondrion 16 (2014) 83–88

Contents lists available at ScienceDirect

Mitochondrion

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

Investigating the role of site specific synonymous variation in disease association studies

Anshu Bhardwaj

Open Source Drug Discovery Unit, Council of Scientific and Industrial Research (CSIR), Delhi 110001, India

ARTICLE INFO

Available online 14 January 2014

Keywords:
Mitochondrial DNA
Mitochondrial diseases
Rare codon
Protein folding
Translation
Association studies

ABSTRACT

Synonymous codon changes may not always be neutral indicating their significance in disease association studies, which is almost always overlooked. Synonymous substitutions may affect protein-folding rates leading to protein misfolding and aggregation. Genome wide analysis of 2301 mitochondrial genomes is performed to evaluate the significance of synonymous codons in disease association studies. The analysis revealed usage of rare codons at several sites in mitochondrial genes with rare codon usage higher for hydrophobic amino acids. The analysis suggests that variation data in association studies should be analyzed using site-specific codon usage values to infer the potential phenotypic impact of synonymous changes.

1. Introduction

Human mitochondrial DNA (mtDNA) encodes a small set of protein subunits that are critical for respiration. Mutation in mtDNA genome may impair oxidative phosphorylation resulting in mitochondrial dysfunction (Blakely et al., 2005; Chol et al., 2003). A multitude of mtDNA diseases have been correlated with single nucleotide changes and insertion–deletions (Wallace, 1999). Single nucleotide changes leading to change in amino acid, non-synonymous or no change, synonymous, have been observed across different disease phenotypes. A non-synonymous change may be deleterious depending on the properties of the wild type and mutated amino acid, and its position. On the other hand, synonymous changes are expected to be inert and hence are usually ignored when assessing the affect of variations in association studies. However, in this context it is important to note that not all synonymous codons for an amino acid are used equally. Unequal use of synonymous codons has been observed in many organisms, both prokaryotes and eukaryotes (Grantham et al., 1980). It is known that the human mtDNA genetic code is also degenerate but codon assignments are different from the universal genetic code (Anderson et al., 1981; Knight et al., 2001). Earlier studies on functional effects of the six naturally occurring synonymous changes in human DRD2 also indicate that synonymous mutations are not 'silent'. It has been shown that one of these mutations, 957 T altered the predicted mRNA folding which led to a decrease in mRNA stability and translation. This in turn leads to dramatic change in dopamine-induced up-regulation of DRD2 expression. Subsequently it was also shown that C957T was found to be in linkage disequilibrium in a European–American population with the -141C Ins/Del and TaqI 'A' variants. These variants have been reported to be associated with schizophrenia and alcoholism, respectively (Duan et al., 2003). In addition, there are reports on host–pathogen interactions being modulated across different environmental states as an outcome of balancing act between mutational bias and natural selection by modifying codon usage (Goni et al., 2012).

Synonymous codon usage has been implicated in translational control as it relates to movement of the ribosomes and the presence of translational pause. A number of factors, including tRNA availability (Dong et al., 1996), tRNA recharging (Elf et al., 2003), optimal codon/anti-codon pairing (Akashi, 2003) and resemblance to conserved recognition sequences (Kane, 1995; Moszer et al., 1999), have optimized codon usage so that the ribosomal machinery progresses through some mRNA sequences quickly and efficiently while pausing at specific rare codons (Ramachandiran et al., 2000). In case of *Escherichia coli*, it has been observed that the codon GAA (Glu) was translated with a rate of 21.6 codons/s, whereas the synonymous codon GAG (Glu) was translated 3.4-fold slower (6.4 codons/s) (Sørensen and Pedersen, 1991).

It was also report earlier that a synonymous SNP in the MDR1 gene (C3435T: Ile-Ile) results in P-glycoprotein (P-gp) with altered drug and inhibitor interactions. This alteration has been attributed to the differences in protein conformations between the wild type and polymorphic P-gp. It has been hypothesized that the presence of rare codon affected the timing of cotranslational folding and insertion of P-gp into the membrane thereby altering the structure of substrate and inhibitor interaction sites (Kimchi-Sarfaty C Oh et al., 2006; Komar, 2007). Thus, alterations in translation kinetics could influence the *in vivo* protein-folding pathway (Komar et al., 1999; Sorensen Ma Kurland et al.,

E-mail addresses: anshub@osdd.net, anshu@csir.res.in.

1567-7249/$ – see front matter © 2014 Elsevier B.V. and Mitochondria Research Society. All rights reserved.
http://dx.doi.org/10.1016/j.mito.2013.12.005

1989). In this regard, an interesting correlation has been observed in codon usage and protein secondary structures. A preference of the rare codons to encode turns, loops and domain linkers has been reported (Thanaraj and Argos, 1996). Additionally, α-helical sequences tend to have fewer rare codons than β-sheet sequences, and domain boundaries where rare codons are commonly found (Thanaraj Ta Argos and Argos, 1996). These observations suggest a possible relationship between synonymous codon usage and protein folding kinetics: since folding rates correlate with contact order (Plaxco Kw Simons et al., 1998) and chain length (Ivankov Dn Garbuzynskiy et al., 2003). Pause after domains and β-sheets could allow these structures to form co-translationally (Hardesty et al., 1999) preventing protein misfolding. Likewise, silent mutations in a series of codons corresponding to amino acid residues in the active site of chloramphenicol acetyltransferase were found to decrease the activity of the protein. These silent mutations eliminated a specific pause observed during *in vitro* translation of the protein (Cortazzo P Cervenansky et al., 2002).

The discussions above point to the fact that that kinetics of protein translation can affect protein-folding pathway and synonymous changes, from non-preferred to preferred codons, can lead to increased levels of protein misfolding. On the other hand, when a preferred codon at a site changes to a synonymous non-preferred codon, it introduces a pause in the translation machinery, which may again lead to misfolding of the peptide. A lot of experimental data provide evidence that rates of chain elongation during translation of proteins are not uniform (Krasheninnikov et al., 1991). The phenomenon of co-translational ubiquitination (and degradation) of proteins (Sato et al., 1998; Zhou et al., 1998) clearly supports the idea that misfolding of proteins could start during protein translation. This has implications in diseases that may result from conformational changes in the spontaneous protein folding due to the incompatibility of the translation machinery components with the folding-specific translation rate. The misfolded proteins would either adopt a conformation that cannot be proteolyzed, or aggregate, preventing them from entering the proteasome (Solomovici et al., 2004). In addition, there are reports that suggest evolutionary constraints at synonymous sites (Conticello et al., 2000; Shields et al., 1988). More recently, it has been established that codon usage does have a role to play in protein translation rate, which is guided by the secondary structure features in a protein by proposing a new method of normalized translational efficiency (Pechmann and Frydman, 2013). In light of these observations, it seems logical to view certain assumptions on role of synonymous variation in disease association studies where they may have effects of potential pathophysiological and pharmacogenetic importance (Chen et al., 2010). In this study we have made an attempt to evaluate the significance of rare codon usage in mtDNA. We have subsequently discussed and proposed a systematic analysis pipeline to explore the extent to which correlations may be drawn between codon usage and protein folding. In order to understand and evaluate the hypothesis in mtDNA, we have defined the rare codon based on its site-specific usage *vs.* the average codon frequency in the genome (rare/unpreferred codons are defined as per the average genome usage). The analysis revealed usage of rare codons at several sites in mitochondrial genes with rare codon usage higher for hydrophobic amino acids. The analysis suggests that variation data in association studies should be analyzed using site-specific codon usage values to infer the potential phenotypic impact of synonymous changes.

2. Results and discussion

2.1. Results from human mtDNA

Analysis of the Genpat data revealed that certain non-preferred codons (as per the codon usage table) are preferred at certain sites [Supplementary Table 1]. On further analysis we observed that maximum number of these non-preferred codons code for hydrophobic amino acids, namely, isoleucine, phenylalanine, leucine and tyrosine [Fig. 1] (Wimley and White, 1996). Interestingly, it has been shown that the assembly of hydrophobic core is important and perhaps a rate limiting step in protein folding (Agashe et al., 1995). Thus, these non-preferred codons may be of significance in the protein-folding pathway. However, experimental assessment is mandatory for safely concluding the same.

The pyrimidine richness of mitochondrial-encoded membrane protein transcripts is partly driven by U (T in DNA) nucleotides in the second codon position in all species, which yields hydrophobic amino acids (Bradshaw et al., 2005). Thus, we checked for the frequency of T in the unpreferred codons. As can be seen in Fig. 2, T is maximally used at the second codon position, supporting the analysis for human mitochondrial DNA. These values have been normalized for the frequency of the four bases in the protein coding genes on the two-strands, thus these observations are not biased by the genome frequencies of the bases.

We also looked for genes, which might be using non-preferred codons more often than others [Fig. 3]. We found that all genes have a similar usage, except MT-ND6, which lies on the other strand.

To check if the pseudo replication of the data used resulted in the observations mentioned above, we analyzed the Genpat data population-wise and repeated the entire analysis. We still found similar pattern of non-preferred codon usage in all populations with MT-ND6 using non-preferred codons at least 1.5 times more than the average value [Supplementary Fig. 1], thus ruling out the effect of pseudo replication on the results obtained.

We have also applied bootstrapping by selecting 500 individuals at random from the Genpat data for each gene. In 100 bootstraps we found that more than 90% of the time the outcome remains the same.

2.2. Estimation of genomic properties

Biased codon usage may result from a diversity of factors; GC-content and preference for codons with G or C at third nucleotide position (Reyes et al., 1998). We have used an EMBOSS program 'freak' to assess the GC bias across the mitochondrial genome and observed that there is no substantial bias. We also observed that the C-rich mtDNA strand has preference for A and C at the third codon position, as reported earlier (Lafay et al., 1999) [Fig. 4].

2.3. Preliminary analysis of the tim-barrel fold

Knowledge-based analysis of $(\beta/\alpha)_8$ which is the most common fold appearing in ~10% of all known enzyme structures provides some clue towards the role of rare codons in protein folding. It has been reported that in the process of reverse engineering of this fold, β central residues, β-strand stops, inter-subunit salt-bridges (category I) were observed to be significant for protein structure and function, whereas, alpha-helix capping, alpha-stops, turn sequences and residues at interface of alpha-helix and beta-strand are highly mutable (category II) (Silverman et al., 2001). We observed that of the 14 residues that form the part of the hydrophobic core in the first category, 4 are rare codons, whereas none of the 2 in the second category is coded by a rare codon [Supplementary Fig. 2]. This also indicates the role of rare codons in protein-folding event may be driven *via* hydrophobic core.

2.4. Correlation between tRNA abundance, copy number and structural components & rate of protein folding: significance of location-specific codon optimization

Recent studies have reported that there exists correlation between tRNA copy numbers with their expression. Subsequently, models have been built on how the supply–demand has guided codon optimization (Pechmann and Frydman, 2013). These correlations have provided overall analyses of codon optimization, however, it is also significant to study position-specific codon usage, given the examples discussed elsewhere. A more recent observation further substantiates the significance of location-specific codon usage in FREQUENCY (FRQ) protein of Neurospora, which