

Low thymine content in *PINK1* mRNAs and insights into Parkinson's disease

Perumal Anandagopu, Shamima Banu, Jinyan Li*

Bioinformatics Research Center, School of Computer Engineering, Nanyang Technological University, NS4-04-33a, 50 Nanyang Avenue, Singapore 639798, Jinyan Li-E-mail id: jyli@ntu.edu.sg, Phone: +656790 6253, *corresponding author

Received March 19, 2010; accepted April 09, 2010; published April 30, 2010

Abstract:

Thymine is the only nucleotide base which is changed to uracil upon transcription, leaving mRNA less hydrophobic compared to its DNA counterpart. All the 16 codons that contain uracil (or thymine in gene) as the second nucleotide code for the five large hydrophobic residues (LHRs), namely phenylalanine, isoleucine, leucine, methionine and valine. Thymine content (i.e. the fraction of XTX codons, where X = A, C, G, or T) in *PINK1* mRNA sequences and its relationship with protein stability and function are the focus of this work. This analysis will shed light on *PINK1*'s stability, thus a clue can be provided to understand the mitochondrial dysfunction and the failure of oxidative stress control frequently observed in Parkinson's disease. We obtained the complete *PINK1* mRNA sequences of 8 different species. The distributions of XTX codons in different frames are calculated. We observed that the thymine content reached the highest level in the coding frame 1 of the *PINK1* mRNA sequence of *Bos Taurus* (*Bt*), that is peaked at 27%. Coding frame 1 containing low thymine leads to the reduction in LHRs in the corresponding proteins. Therefore, we conjecture that proteins from the other organisms, including *Homo sapiens*, lost some of their hydrophobicity and became susceptible to dysfunction. Genes such as *PINK1* have reduced thymine in the evolutionary process thereby making their protein products potentially being susceptible to instability and causing disease. Adding more hydrophobic residues (thymine) at appropriate places might help conserve important biological functions.

Keywords: thymine distribution; *PINK1*; sequence analysis; protein stability; frame analysis

Background:

Parkinson's disease (PD) is the most prevalent neurodegenerative movement disorder and affects about 1% of the elderly population. Cell death of midbrain dopaminergic neurons is the key pathological feature, and clinical symptoms include bradykinesia, tremor and rigidity among others [1]. The disease etiology remains largely unknown although aberrant protein degradation and mitochondrial dysfunction have been highlighted as probable causes. PD has a much higher proportion of sporadic cases as compared to familial or inherited cases. Yet "significant advances in our understanding of PD have stemmed directly from the study of [those] genes associated with a small proportion of familial cases" [2] and mitochondrial dysfunction has been implicated as a common pathogenic mechanism in both familial and sporadic PD. Mitochondrial dysfunction in PD can occur along three pathogenic pathways [3]. However for many PD genes the precise mechanism through which mitochondrial integrity is maintained remains yet unknown. Six PD associated genes namely α -synuclein, *Parkin*, *PINK1* (PTEN-induced putative kinase 1), *DJ-1*, *LRKK2* (Leucine-rich repeat kinase 2) and *HTRA2* (high temperature requirement protein A2) have been identified to show involvement in mitochondrial dysfunction and oxidative stress. Intensive work on mutational studies demonstrated their importance for normal mitochondrial function [4]. Of these 6 genes we chose *PINK1* for a study of thymine levels as ample evidence exists that this gene product is targeted to mitochondria and is located upstream of *Parkin* in the pathogenic pathways leading to mitochondrial dysfunction [5], [6], [7]. Recessive mutations of *PINK1* have caused mitochondrial failure and an inability of the mitochondrion to cope with oxidative stress. Maguire Zeiss *et al.* [8] & Pridgeon *et al.*, [9] have conducted an analysis on an inherited form of early-onset PD which has been linked to mutations in both copies of *PINK1*. It was also reported that *PINK1* phosphorylates tumor necrosis factor receptor-associated protein 1 (TRAP1), and in turn TRAP1 prevents oxidative stress by preventing mitochondria from releasing cytochrome c --- a crucial step in cell death. Such mutational studies seek functional insights by implicating genes and proposing models of PD pathways. However it may also be useful to study those genes from a wider perspective by performing a comparison of sequence analysis across different organisms. PD gene

studies have sought to identify possible loci of implicated genes, to understand gene function through deletions, knock-out mutations and other studies. However to the best of our knowledge no study on sequence analysis of PD genes comparing nucleotide differences across different organisms has been carried out. This is important as besides site specific or targeted gene mutations, naturally differing levels of certain nucleotides may affect hydrophobicity of proteins when compared across differing organisms. This provides a possible reason as to why certain proteins are more likely to be unstable despite the fact that no disease causing mutations or hereditary disease causes can be identified. Here we use a macro approach to PD mechanisms by focusing on *PINK1* mRNA sequences to compare the differing distribution of thymine content across species in an attempt to identify protein instability due to the loss of hydrophobic residues.

Methodology:

Dataset:

PINK1 mRNA sequences in FASTA format for *Homo sapiens* (*Hs*), *Pan troglodytes* (*Pt*), *Bos taurus* (*Bt*), *Mus musculus* (*Mm*), *Macaca mulatta* (*Mmu*), *Gallus gallus* (*Gg*), *Danio rerio* (*Dr*) and *Caenorhabditis elegans* (*Ce*) are taken from NCBI Genbank [10]. The total number of bases in the mRNA sequence for each organism is listed in Table 1 (see supplementary material).

The frames 1, 2 and 3 represent the 5'-3' mRNA sequences and frames 4, 6 and 5 are antisense strand. Each mRNA sequence is read in six different frames and the number of XTX occurrences in each frame is counted (i.e., Thymine nucleotide 2 in the second position of the codons is considered the beginning of frame 1. The next two immediate nucleotides 3 and 4 are considered the beginnings of frames 2 and 3 respectively. The antisense strand nucleotides of 2, 3 and 4 are considered frames 4, 6 and 5, respectively). The number of XTX occurrences for each frame is subsequently expressed as a fraction of the total number of codons for all six frames. This procedure is repeated for all 8 organisms and the results shown in Table 2. The proportion of XTX content thus obtained for each of the eight organisms for all six frames is plotted and shown in Figure 1.

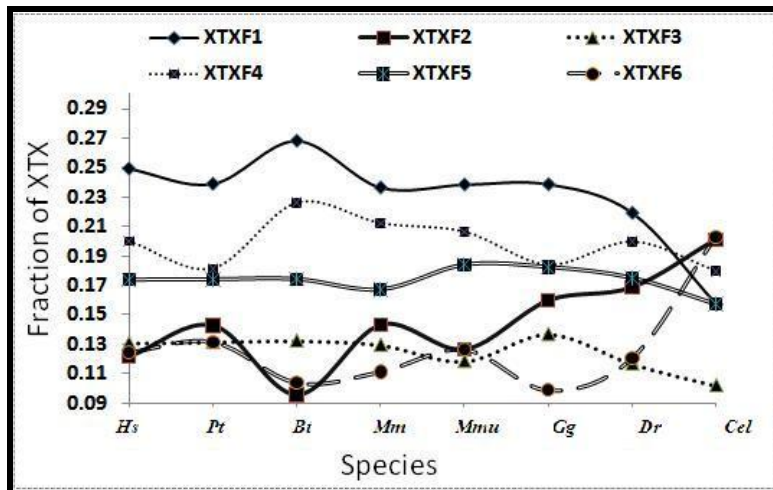


Figure 1: The fraction of thymine (XTX) in the coding frames of *PINK1* mRNA sequences in eight different species found in **Table 2** (see supplementary material).

Discussion:

It is observed that for frame 1 the thymine content is lower than 27% in all species studied here except for *Bos taurus* which has reached the expected genome wide thymine content level of 27% [11]. The coding frames of mRNA sequences contain a greater amount of thymine in frame 1 than frames 2, 3, 4, 5 and 6 (except for *Caenorhabditis elegans*). Frames 3 and 5 have similar fractions of XTX for *Homo sapiens*, *Pan troglodytes*, *Bos taurus* and *Mus musculus*, except *Macaca mulatta*, *Gallus gallus*, *Danio rerio* and *Caenorhabditis elegans*. However, frames 2 and 6 show inconsistent fractions of thymine as seen in **Figure 1**.

As highlighted above, the thymine content in Frame 1 and Frame 4 of the *PINK1* mRNA sequence in *Homo sapiens* is lesser than that of *Bos taurus*. Therefore, the coding frames of these mRNA sequences are more hydrophilic and may be more likely to contribute to impaired protein activity whereas the hydrophobicity is maintained in the corresponding coding frames of *Bos taurus*. From **Figure 1**, for *Homo sapiens*, *Pan troglodytes*, *Mus musculus*, *Macaca mulatta*, *Gallus gallus*, *Danio rerio* and *Caenorhabditis elegans*, we also noticed that almost all frame 1 coding mRNA sequences code for more hydrophilic residues, which increases exposure to water and thus contributes to protein instability.

A recent study [12] has reported that hydrophobic interactions comprise the dominant force in most biochemical reactions which take place in water. Carbon is the main element which contributes to this interaction. This study also analyzed the carbon content at the protein functional sites of *PINK1* which prefer to have 31.44% of total carbon along the chain. The *PINK1* amino acid sequence in *Homo sapiens* seems to be less stable in comparison to that of *Bos taurus*. They suggested that the first 267 amino acids are hydrophilic and therefore do not contribute to protein stability as they lack the required carbon content. However in *Bos taurus* the hydrophobicity is maintained within this range of 267 amino acids. From our results and their findings, we suggest that the *PINK1* protein loses large hydrophobic residues due to low thymine content in *Homo sapiens* rendering it more hydrophilic and thus susceptible to losing its stability and activity. This shows that mutations to the genetic sequence might have occurred suggesting evolutionary changes.

As a factor to balance the thymine content and distribution of LHR, more numbers of small hydrophobic residues (SHR) such as glycine, alanine, proline and cysteine are found in the *PINK1* mRNA

sequences. With the addition of these residues, the length of protein sequences also increases. This is the reason why the length of animal coding mRNA sequence increases and in turn protein length increases [12]. We expect that these genes and corresponding longer mRNA sequence of proteins contain low thymine content.

The statistical distribution of hydrophobic residues along the protein chains and its implications to protein folding and evolution were reported by White and Jacobs in 1990 [14]. They tested for randomness of hydrophobic residues in proteins but only on a limited number of sequences. They suggested that the folding of proteins into compact structures may be much more permissive with less sequence specificity than previously thought and that the clusters of hydrophobic residues along chains which are revealed by hydrophobicity plots are a natural consequence of a random distribution. Currently most of the protein sequences of several model organisms are available for a more thorough understanding which we have carried out here. White and Jacobs in 1993 [15] further argued that the distribution of hydrophobic residues along a sequence cannot be distinguished from that expected for a random distribution and suggested that functional proteins may have originated from random sequences.

Conclusion:

Our study strongly suggests that coding frames of mRNA sequences for *Homo sapiens* are closely associated with *Pan troglodytes*, *Mus musculus*, *Macaca mulatta* and *Gallus gallus*. The *Homo sapiens* coding frame 1 of the mRNA sequence in *PINK1* has a thymine distribution of only 25% which is less than the random percentage of 27%. As introduced, the XTX thymine codons are responsible for coding LHRs in proteins. This thymine that has one extra methyl group compared to its RNA Globular proteins is expected to follow the 27% distribution profile. A protein's stability, function and interaction with other proteins are primarily determined by LHRs. Thus, tracing a protein's hydrophobicity via the thymine content levels can serve as a useful tool in identifying the coding frames of mRNA sequences that might contribute to diseases.

Large numbers of hydrophilic residues are found among seven of the eight organisms analyzed. This is especially true of *Homo sapiens*. Any reduction of hydrophilic amino acid residues or addition of hydrophobic residues can increase the functional activity of *PINK1*. Therefore, adding more thymine at appropriate places in mRNA or in the gene will improve the amount of hydrophobic residues which

could contribute to better biological activity, and defects in the disease-causing sequence could be eradicated. Applying such a strategy for *PINK1* might prevent mitochondrial dysfunction and oxidative stress.

References:

- [1] JA Obeso *et al. Trends Neurosci.* (2000) **23**: S8
- [2] PM Abou-Sleiman *et al. Nat. Rev. Neurosci.* (2006) **7**: 207 [PMID: 16495942]
- [3] Y Lee, *Interdisciplinary Bio Central.* (2009) doi: 10.4051/ibce.2009.2.0011
- [4] C Henchcliffe *et al., Nat Clin Pract Neurol.* (2008) **4**: 600 [PMID:18978800]
- [5] IE Clark *et al. Nature.* (2006) **441**: 1162 [PMID: 16672981]
- [6] J Park *et al. Nature.* (2006) **441**: 1157 [PMID: 16672980]
- [7] Y Yang *et al. Proc Natl Acad Sci U S A.* (2006) **36**: 10793 [PMID: 16818890]
- [8] KA Maguire-Zeiss & HJ Federoff, *Ann NY Acad Sci.* (2003) **991**: 152 [PMID: 12846984]
- [9] JW Pridgeon *et al. PLoS Biol.* (2007) **5**: 1494 [PMID: 17579517]
- [10] <ftp://ftp.ncbi.nlm.nih.gov/genbank/>
- [11] P. Anandagopu *et al. Bioinformation.* (2008) **2**: 304 [PMID: 18478084]
- [12] R Suhanya *et al. Int.Journal Bioinfo.* (2008) **1**: 9
- [13] V Jayaraj *et al. Bioinformation.* (2009) **3**: 409 [PMID: 19759817]
- [14] SH White & RE Jacobs, *Biophys J.* (1990) **57**: 911 [PMID: 2188687]
- [15] S.H. White & RE Jacobs, *J. Mol. Evol.* (1993) **36**: 79 [PMID: 8433379]

Edited by P. Kanguane

Citation: Anandagopu et al. Bioinformation 4(10): 452-455 (2010)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

Supplementary Materials

Table 1: Total number of base pairs in *PINK1* mRNA sequences.

Organisms	Genbank ID Number	mRNA base pairs
<i>Homo sapiens</i> (Hs)	NM_032409	1746
<i>Pan troglodytes</i> (Pt)	XM_001164912	1476
<i>Bos taurus</i> (Bt)	NM_001099701	1767
<i>Mus musculus</i> (Mm)	NM_026880	1743
<i>Macaca mulatta</i> (Mmu)	XM_001096957	1626
<i>Gallus gallus</i> (Gg)	XM_423139	1575
<i>Danio rerio</i> (Dr)	NM_001008628	1725
<i>C. elegans</i> (Ce)	NM-062616	1926

Table 2: Counts and fractions of thymine content (XTX) in the 6 coding frames of mRNA sequences of PINK1 (PTEN Induced Putative Kinase1) in eight different species.

Species	No. of XTX in all six frames						Total		Fraction of the XTX in all six frames					
	XTXF1	XTXF2	XTXF3	XTXF4	XTXF5	XTXF6	TotXTX	Totbp	XTXF1	XTXF2	XTXF3	XTXF4	XTXF5	XTXF6
<i>Homo sapiens</i> (Hs)	162	79	85	130	113	81	650	1746	0.2492	0.1215	0.1308	0.2	0.1738	0.1246
<i>Pan troglodytes</i> (Pt)	129	77	71	98	94	71	540	1476	0.2389	0.1426	0.1315	0.1815	0.1741	0.1315
<i>Bos taurus</i> (Bt)	160	57	79	135	104	62	597	1767	0.2680	0.0955	0.1323	0.2261	0.1742	0.1039
<i>Mus musculus</i> (Mm)	157	95	86	141	111	74	664	1743	0.2364	0.1431	0.1295	0.2123	0.1672	0.1114
<i>Macaca mulatta</i> (Mmu)	149	79	74	129	115	79	625	1626	0.2384	0.1264	0.1184	0.2064	0.184	0.1264
<i>Gallus gallus</i> (Gg)	157	105	90	121	120	65	658	1575	0.2386	0.1596	0.1368	0.1839	0.1824	0.0988
<i>Danio rerio</i> (Dr)	169	130	90	154	135	93	771	1725	0.2192	0.1686	0.1167	0.1997	0.1751	0.1206
<i>Caenorhabditis</i>	182	233	118	208	182	235	1158	1926	0.1572	0.2012	0.1019	0.1796	0.1572	0.2029