



ISSN Print: 2394-7500
 ISSN Online: 2394-5869
 Impact Factor: 5.2
 IJAR 2016; 2(2): 342-344
 www.allresearchjournal.com
 Received: 06-12-2015
 Accepted: 09-01-2016

Partha Majumder
 Biomedical Scientist,
 Former Head of Department of
 Applied Biotechnology &
 Bioinformatics, Sikkim
 Manipal University, CC: 1637,
 Kolkata, India.

Anjana Mazumdar
 Professor, Department of Oral
 Pathology, Dr. R. Ahmed
 Govt. Dental College, Kolkata,
 India

Amitabha Kar
 Associate Professor,
 Department of Physiology,
 Women's College, Govt. of
 Tripura, Agartala, Tripura,
 India

Sufal Halder
 Senior Consultant Physician,
 Acharya Pranavananda
 Sevashram, Po: Purba
 Nischintapur, Budge Budge,
 Kolkata, India

Correspondence
Partha Majumder
 Biomedical Scientist,
 Former Head of Department of
 Applied Biotechnology &
 Bioinformatics, Sikkim
 Manipal University, CC: 1637,
 Kolkata, India.

Analysis of codon and amino acid usage pattern in four replicons of radiation-resistant bacterium *Deinococcus radiodurans*- A challenging factor for waste cleanup & new industrial approach

Partha Majumder, Anjana Mazumdar, Amitabha Kar, Sufal Halder

Abstract

The complete genetic code of the world's most radiation resistant organism - the bacteria, *Deinococcus radiodurans* - has been detailed by researchers at The Institute for Genomic Research (TIGR). This strain of pink bacteria can survive in 1.5 million rads of gamma irradiation - 3,000 times of the amount that would kill a human. The organism's ability to repair the DNA damage in a day and go on living offered inspiration to carry on research to know the mechanisms of cellular repair. Advances in this area could improve the understanding of cancer, which can be caused by unrepaired DNA damage. Genetically engineering of this microbe could lead to improved ways to clean up pollution (especially, waste cleanup) and to new industrial processes. Besides the insights into the way cells work, this new research may help to provide a new safe and inexpensive tool for some of the Nation's most difficult clean up challenges. In this paper of research work we have analyzed how really the high radiation resistant of *D. radiodurans* strain R1 makes it an ideal candidate for bioremediation of the sites, which are contaminated with radiation and toxic chemicals, it is also naturally transformable and therefore amenable to genetic manipulation. On cultivation of with the bacteria *D. radiodurans*, we found that hydrophobicity, expressivity and aromaticity are the major sources of variation in amino acid usage among the genes in the organism makes it compatible to survive itself even in adverse condition of nature compared to other organisms.

Keywords: Gamma radiation, Radiation resistant, DNA repair, DNA damage, Waste clean-up,

1. Introduction

All species in the genus *Deinococcus*, in particular *Deinococcus radiodurans*, are extremely resistant to a number of agents and conditions that damage DNA, including ionizing and ultraviolet (UV) radiation and hydrogen peroxide [1, 13]. The whole genome sequence of *Deinococcus radiodurans* strain R1 became available in 1999 (White, O. 1999) [24]. The genome of *Deinococcus radiodurans* is composed of four replicons: chromosome I (2.65Mb), a megaplasmid -1 or possible "chromosome II" (412Kb), a megaplasmid-2 (177.5Kb), and a usual size plasmid-3 (45.7Kb). The number of predicted protein coding sequences for the chromosome-I, chromosome-II (megaplasmid-1), megaplasmid (megaplasmid-2) and small plasmid (plasmid-3) are 2633, 369, 145 and 40 respectively (White, O.1999) [24].

D. radiodurans is the most radiation-resistant organism known to date; exponentially growing cells are 200 times more resistant to ionizing radiation and 20 times as resistant to UV irradiation (as measured by survival) as *Escherichia coli* (J. R. Battista, 1997) [8]. The high radiation resistant of *D. radiodurans* strain R1 makes it an ideal candidate for bioremediation of the sites, which are contaminated with radiation and toxic chemicals, it is also naturally transformable and therefore amenable to genetic manipulation.

Synonymous codon usage patterns in microbial organisms are usually shaped by two major forces, the directional mutational bias (Sueoka 1988) [23] and natural selection operative at the level of translation (Ikeumura 1981; Gouy and Gautier, 1982 [7] Sharp and Li 1987; Pan, Dutta and Das 1998 [19]; Kanaya *et al.* 1999) [7, 9, 19, 22]. There relative strength and direction of these forces vary within and across the organisms, depending on their life-styles and/or overall G+C bias. While the codon usage patterns in base-equipotent *Escherichia coli* and many other free-living microbial organisms reflects the selection-mutation balance (Andersson and Kurland 1990; Pan, Dutta and Das 1998) [1, 19], synonymous codon selection in intracellular

pathogens and/or in genomes with highly skewed GC – composition such as *Mycoplasma genitalium* (GC%=31.6), *Rickettsia prowazekii* (GC% =28.9), *Buchnera* (GC% =25.2) or *Micrococcus luteus* (GC% = 72) are primarily governed by local GC- composition with little or no effect of translation selection Ohama, Muto and Osawa 1990; Anderson & Sharp 1996, Melnamey, 1997; Waranagreen and Morgan, 1999 ^[1, 22]. Apart from mutational bias and translation selection, there are other factors like context-dependence (Berg and Silva, 1997) ^[2], codon-anticodon interaction (Grosjean and Fiers, 1982; West, Iglewski ^[5, 6, 14, 16] physical location of each gene on the chromosome (Kerr *et al.* 1997) ^[11], replication-translation selection (Lafay *et al.* 1999; Romero *et al.* 2000a) ^[14, 21], hydrophobicity of each gene (de Miranda *et al.* 2000) ^[3] etc. that may influence the synonymous codon selection in various microbial organisms. Recent genomic studies have demonstrated that variations in nucleotide composition have significant effects not only on the patterns of codon usage (Frank and Lobry, 1999; Kanaya *et al.* 2001 Grocock and Sharp, 2002) ^[4, 5], but also on the amino acid composition of encoded proteins (Lobry, 1997; Knight *et al.*, 2001; Palacios and Wernegreen, 2002) ^[15, 12, 18]. Apart from mutational bias, the amino acid usage in bacterial proteins may be influenced by various other factors. For example, amino acid usage in *E. coli* is influenced by the hydrophobicity, expressivity and aromaticity of the respective proteins (Lobry and Gauter, 1994) ^[16]. In *Buchnera*, the A+T- rich endosymbiotic bacteria of aphids, the amino acid usage in putative highly expressed genes is characterized by avoidance of aromatic amino acids and by greater conservation and resistance to A+T enrichment (Risper *et al.*, 2004) ^[20], while in another endosymbiotic bacterium *Blochmannia floridanus*, conservation of GC-rich amino acids in highly expressed gene-products and the average hydrophobicity of the proteins are reported to be the two major sources of variation in amino acid usage . In the spirochaetes *Borrelia burgdorferi* and *Treponema pallidum*, the strand-specific mutational biases have strong effects not only on the codon usage, but also on the amino acid selection in genes/proteins of the organisms (Lafay *et al.* 1999) ^[14].

In the present study, the codon and amino acid usage in four replicons of *D. radiodurans* were extensively analyzed; a comparison between four replicons may provide insights into the distinct compositional characteristics of the genes/proteins playing significant roles in radiation resistance.

2. Materials and Methods

The complete genome sequence of *D. radiodurans* R1 was downloaded from NCBI GenBank. To minimize sampling error we had taken only those reading frames (ORFs), which are greater than or equal to 300 bp and have correct initial and termination condos. The presumed duplicates, transposes genes and recent horizontally transferred genes were also excluded. Finally the sequences selected for the four replicons were as follows: 2403 sequences for chromosome-I, 353 sequences for chromosome-II, 112 sequences for megaplasmid and 27 sequences for small plasmid. The analysis of codon and amino acid usage patterns and the nucleotide usage at the synonymous site were carried out using CODONW 1.4.2 (available from www.molbiol.ox.ac.uk/win95.codonW.zip). Parameters like number of each codon, the relative synonymous codon usage (RSCU) (Sharp and Li 1987) ^[22], the codon adaptive index (CAI) (Sharp and Li 1987) ^[22], the effective number of codons(Nc) and the G+C content at third codon position

(GC₃) were calculated for individual ORF. The relative amino acid usage (RAAU), average hydrophobicity (Gravy score) and aromaticity for each gene product were estimated. The CAI, which is widely used to estimate the potential gene expression level, was calculated by taking the reference of codon usage in highly expressed genes (i.e. ribosomal proteins, transcription and translation processing factors etc.). The corresponding analysis (COA) available in CODONW program was carried out on RSCU values and amino-acid frequencies to look for the factors affecting the codon and amino-acid usage respectively. In order to detect the differences between the highly and lowly expressed genes (for all four replicons), if any, codon abundance were calculated in 2x2 contingency tables and to find out the differences of mean CAI values for four replicons, the STATISTICA (Version 6.0) was used.

3. Results and Discussion

Compositional bias of genes in four replicons

The G+C content might be one of the most important factors in the evolution of genomic structure ^[8, 12, 13, 15]. The genome of *D. radiodurans* is compositionally biased towards high G+C content (G+C=66.6%). The GC distribution at third codon position (GC₃) of genes and effective number of codon (Nc) values of different genes of four replicons have been shown in the table below.

Table 1: The GC distribution at third codon position (GC₃) of genes and Nc values of different genes of four replicons

Replicons		Mean	Max value	Min value	S.d.
Chromosome-I	GC ₃	83.9%	94.4%	28.2.9%	6.74%
	Nc	36.53	61	23.89	5.237
Chromosome-II	GC ₃	82.4%	92.3%	28.4%	9.12%
	Nc	37.82	59.1	29.1	5.45
Megaplasmid	GC ₃	76.8%	92.5%	41.8%	10.47%
	Nc	42.39	61	25.26	6.94
Small plasmid	GC ₃	60.04%	73.9%	42.6%	7.79%
	Nc	54.88	61	45.22	4.285

From Table.1 it can be inferred that large variations in GC₃ and Nc indicate that there is a marked heterogeneity in compositional bias as well as in the codon usage pattern within the members of the four replicons, suggesting that the codon usage in genes might be governed by some additional selection forces that are independent of the overall mutational bias of the organism. Corresponding analysis on relative synonymous codon usage in chromosome-I, chromosome-II, megaplasmid of the bacteria *D. radiodurans* suggest a clear version of variations in codon selection is very strong than that in small plasmid which depicts the trend in codon selection is not as strong as in previous three replicons.

Table 2: The mean CAI values of the four replicons

Replicons	Mean	S.d.
Chromosome-I	0.513	± 0.104
Chromosome-II	0.504	± 0.107
Megaplasmid	0.432	± 0.107
Small plasmid	0.265	± 0.048

From the above data we can say the mean CAI values of the first three replicons (Chromosome-I> Chromosome-II> Megaplasmid) are significantly different from the mean CAI values of fourth replicons (small plasmid). It is known that greater is the CAI value of a gene, stronger is the bias in the synonymous codon selection in the same (i.e., higher is the

usage of optimal codons). So the higher value of average CAI in the first three replicons than that in fourth replicons suggests that the extent of synonymous codon bias, on an average, is more in the first three replicons than in the small plasmid. In other words, the influence of translation selection is stronger in the first three replicons genes than the genes in small plasmid.

The overall codon usage in the coding sequences of four replicons

The overall codon usage in the coding sequences of four replicons of *D. radiodurans* expected to be biased towards GC-rich codons. C-ending synonyms are the most preferred codons and the G-ending codons usually appeared in the second choice. There are 20 codons, the usage of which significantly increases ($p < 0.01$) in highly expressed genes of the chromosome-I, as compared to the lowly expressed genes of the same. Out of these 20 codons, 19 codons tend to have higher frequencies in highly expressed genes than in lowly expressed ones of chromosome-II and megaplasmid, although such increase in frequencies is found to be statistically significant ($p < 0.01$) for 16 codons only in chromosome-II and megaplasmid. On the other hand in case of small plasmid only one codon (GAG), which codes for Glu (Glutamic acid) is significantly higher ($p < 0.01$) in frequencies in the high bias dataset.

This observation indicates that the optimal codon preferences are almost similar in the genes of chromosome-I, chromosome-II and megaplasmid, whereas small plasmid is the exception. This suggests that the events of horizontal transfer are old enough to make the amelioration of the acquired gene near completion so that the signature of the so-called mother genome in the codon usage of chromosome-II and megaplasmid have been almost wiped out, but in case of small plasmid the picture is different, this may be due to the recent horizontal transfer of the genes.

Major trends in the amino acid usage in proteins coded by four replicons

To understand the trends in amino acid usage in the proteins coded by the four replicons of *D. radiodurans*, we performed COA separately on amino acid frequencies of each protein in four datasets. In *E. coli*, Lobry and Gautier (1994) [16] demonstrated that hydrophobicity, expressivity and aromaticity are the major sources of variation in amino acid usage of 999 chromosome-encoded proteins. In *D. radiodurans* too, these are the major sources of variation in the amino acid usage, but the relative importance of these parameters in amino acid selection varies significantly.

4. Conclusion

The present study demonstrates that the pattern of synonymous codon usage and amino acid frequencies in *D. radiodurans* are the result of several factors. Variation in the codon usage is determined not only by the compositional constraints but also by the translational selection. The overall codon usage of the organism shows expected bias towards G+C-rich codons. It was also observed that, C-ending codons are always preferred than G-ending codons. The comparative multivariate analysis of synonymous codon usage patterns in four replicons of *D. radiodurans* suggests that the event of horizontal gene transfer is old enough for chromosome-I and megaplasmid genes but genes of small plasmid are recently horizontally transferred. The cumulative amino acid usage of first three replicons shows similar nature. In case of amino

acid usage, expressivity is one of the major sources of variation. Hydrophobicity of genes is the second major factor in differentiating the amino acid usage variation among the genes in the organism.

5. Acknowledgements

All of the authors achieved extreme guidance favoring the in depth cultivation with a positive output from Dr. D.N. Tibarewala, Professor, School of Biosciences and Engineering, Jadavpur University, Kolkata, India. Dr. D.N. Tibarewala contributed a pioneer role to the design of the study, data analysis, and revision of the manuscript. And for that, both of the authors are grateful for his extreme support to make the endeavor successful. Dr. Partha Majumder is particularly indebted for constant & evergreen inspiration received from his mother, Late Gita Majumder (passed away on 29.10.2007 in MIOT Hospitals, Chennai) who is in the Heavenly abode of LORD SHIVA & MAA TARA and also like to dedicate this valuable paper to his father, Late Anil Chandra Majumder (passed away on 3rd August, 2014, at Nirmal Ashram Hospital, Rishikesh), from whom he has constantly achieved the patience of tolerance, graciousness of heart and attitude to serve for mankind.

6. References

- Andersson SGE, Kurland CG. *Microbiol Rev.* 1990; 54:198-210.
- Berg OG, Silva PJN. *Nucleic Acids Res.* 1997; 25:1397-1404.
- DE Miranda AB, Alvarez-Valin F, Jabbari K, Degraeve WM, Bernardi G. *J Mol Evol.* 2000; 50:45-55.
- Frank AC, Lobry JR. *Gene*, 1999; 238:65-77.
- Grocock RJ, Sharp PM. *Gene*, 2002; 289:131-139.
- Grosjean H, Fiers W. *Gene*, 1982; 18:199-209.
- Ikeumura T. *J Mol Biol.* 1981; 151:389-409.
- Battista JR. *Annu. Rev. Microbiol.* 1997; 51:203.
- Kanaya S, Yamada Y, Kudo Y, Ikemura T. *Gene*, 1999; 238:143-155.
- Karlin S, Mrazek J. *PNAS.* 2001; 98:5240-5245.
- Kerr AR, Peden JF, Sharp PM. *Mol. Microbiol.* 1997; 125:1177-1179.
- Knight RD, Feeland SJ, Landweber LF. *Genome Biol.* 2001; 2:research0010.10010.13.
- Milton KW. *Mol. Microbiol.* 1994; 13:9.
- Lafay B, Lloyd AT, Mc Lean M, Devine KM, Sharp PM, Wolfe KH. *Nucleic Acids Res.* 1999; 27:1642-1649.
- Lobry JR. *Gene*, 1997; 205:309-316.
- Lobry JR, Gautier C. *Nucleic Acids Res.* 1994; 22:3174-3180.
- Minton KW. *Mutat. Res.* 1996; 363:1-7.
- Palacios C, Wernegreen JJ. *Mol. Biol. Evol.* 2002; 19:575-1584.
- Pan A, Dutta C, Das J. *Gene*, 1998; 215:405-413.
- Rispe C, Delmotte F, van Ham RC, Moya A. *Genome Res.* 2004; 14:44-53.
- Romero H, Zavala A, Musto H. *Nucleic Acids Res.* 2000; 28:2084-2090.
- Sharp PM, W-H Li. *Nucleic Acids Res.* 1987; 15:1281-1295.
- Sueoka N. *Proc. Natl. Acad. Sci. USA*, 1988; 85:2653-2657.
- White O, Eisen JA, Heidelberg JF, Hickey EK, Peterson JD, Dodson RJ *et al.* *Science.* 1999; 286:1571-1577.