ISSN No:-2456-2165

In Silico Analysis of the Protein Structure of Tumor Suppressor Genes in Humans

Kooffreh M.E[1], Okon E.A[2]., Ekerete E.E[3] and Simon I[4]
[1,3,4]Dept of Genetics and Biotechnology, University of Calabar, Calabar, Nigeria
[2]Dept of Biological Sciences, Cross River University of Technology, Calabar, Nigeria

Abstract:- Tumor suppressor genes encode for proteins that are involved in inhibition of cell proliferation, induction of apoptosis, DNA damage repair and inhibition of metastasis and these are crucial to normal cell development and differentiation. The study examined the secondary protein structure, tertiary protein structure, the G≡C content of Rb, p53, BRCA1 and BRCA2 human tumor suppressor genes and the phylogenetic relationship between them. The nucleotide and amino acid sequences of the genes were obtained from the GenBank, secondary protein structure of the proteins were determined using Gor4 online software, while the tertiary protein structure was determined using phyre2 online software. Phylogenetic relationship was computed using MEGA 6 software and G≡C content was measured using GENSCAN software. Secondary structure of the genes contained Alpha helix, extended strand and random coiling; with BRCA1 having the highest percentage (46.15%) for Alpha helices and lowest (18.07%) in p53 gene; the percentage for extended strand was highest (19.23%) in BRCA1 and lowest (11.31%) in Rb gene; and random coil percentage was highest (52.59%) in p53 and lowest (31.34%) in BRCA2. The phylogenetic tree showed evolutionary relationship between the genes with BRCA1 gene being the common ancestor of all sampled genes and BRCA2 and Rb arising from the same cluster, the tertiary structure was then determined for each representative of the three (3) clusters resulting from the phylogenetic tree construction. The G≡C content obtained in percentage were 52.59, 50.07, 36.02 and 31.34 in p53, BRAC1, Rb and BRAC2 respectively. These findings may become useful in trying to understand how tumor suppressor genes function and also assist in emerging anti-cancer strategies.

Keywords:- *Tumor Suppressor Genes. Phylogenetic Tree, G≡C Content, Protein Structure.*

I. INTRODUCTION

Tumor suppressor genes are protective genes. Normally, they limit cell growth by monitoring how quickly cells divide into new cells, repairing mismatched DNA, controlling when a cell dies. When a tumor suppressor gene is mutated, cells grow uncontrollably and may eventually form a tumor. BRCA1, BRCA2, Rb and p53 are examples of human tumor suppressor genes. Germ line mutations in BRCA1 or BRCA2 genes increase a woman's risk of developing hereditary breast or ovarian cancers. Suppression of cell division is the main mechanism for most tumor suppressors [1]. Such tumor suppressors include retinoblastoma protein (Rb), adenomatosis polyposis coli (APC), alternate reading frame (ARF), RIZ1, p15, p16, p18, p19, p21, p27, and p53 [2,3].Rb, which is the first discovered tumor suppressor, inhibits the transcription of specific genes required for mitosis through binding to transcription factors such as E2Fs, which are key cell proliferation regulators [4,1]. The most commonly mutated gene in people who have cancer is the p53, which is associated with about 50% of human cancer cases [5,6]. The can trigger DNA repair processes, induce the transcription of other tumor suppressors, such as p21 and p16, and initiate cell apoptosis [7,1,3] However, studying the protein structure of these tumor suppressor genes can be used in many ways to improve results of structure based drug designs and also as an avenue for structure guided rescue of tumor suppressor gene functions in tumors.

II. MATERIALS AND METHODS

Nucleotide and amino acid sequences of human tumor suppressor genes were retrieved from GenBank. This was achieved by obtaining FASTA formats of the nucleotide and amino acid sequences of human tumor suppressor genes from the National Centre for Biotechnology Information (NCBI) database. The GenBank accession number of the nucleotide and amino acid sequences, gene name, and protein name of the gene and sequence length were retrieved and tabulated. Secondary protein structure of the amino acid sequence was determined using Gor4 online software by submitting the non-FASTA format of amino acid sequence of each gene in the workspace of the software. The sequence length for each gene is also shown in the secondary structure and calculation for the percentage of secondary protein structure for each gene was derived from the formula

$$: \frac{Number\ of\ occurence\ of\ secondary\ structure}{Sequence\ length\ of\ secondary\ structure} X\ 100$$

Tertiary protein structures of the genes were determined by also submitting the sequence in FASTA format and searching Phyre2 online software for the tertiary protein structure. The guanine and cytosine content of the genes were determined using GENSCAN online software. The non- FASTA format of each gene was submitted in the workspace of the software to scan for the percentage guanine and cytosine in the sequence. The phylogenetic relationship between human tumor suppressor genes used in this study was determined by constructing phylogenetic trees using the nucleotide sequence of the

ISSN No:-2456-2165

genes retrieved from NCBI. This was achieved using the unweighted pair group method with arithmetic mean (UPGMA) of MEGA software with 1000 bootstrap replications.

III. RESULTS AND DISCUSSION

Retrieved nucleotide and amino acid sequences of human tumor suppressor genes are documented Table 1 where the gene name, the GenBank accession number which enabled direct access to the gene on the database, the protein name of the gene, the GenBank accession number for the protein, the nucleotide sequence length and also the amino acid sequence length are all tabulated. The nucleotide sequence of the RB1 was the longest with 4600bp, then 3798bp, 2451bp and 619bp for BRAC1, p53 and BRAC2 respectively. Also the amino acid sequence length of RB1 was the longest with 928bp and the shortest of 26bp in BRAC1, 393bp and 206bp was observed for p53 and BRAC2 respectively.

The guanine ≡ cytosine content measured in percentage is shown in Table 2 for the tumor suppressor genes, with p53 having the highest percentage of 52.59, followed by BRCA1 with 50.07, RB1 having 36.02 and the least being BRCA2 with 31.34. Genes with higher guanine cytosine content have a longer coding sequence. The secondary protein structures contained in the p53, RB1, BRCA1 and BRCA2 genes, are the Alpha helix (Hh), Extended Strand (Ee) and Random coil (Cc), with differing number of occurrences. The BRCA1 gene has the highest percentage for Alpha helix with 46.15%, followed by Rb1 with 43.64%, BRCA2 with 23.79% and p53 having the lowest alpha helices with 18.07%. The BRCA1 gene also has the highest percentage for extended strand with 19.23%, followed by p53 with 18.58%, BRCA2 with 16.50%, and Rb1 having the lowest percentage with 11.31%. The p53 gene has the highest percentage for random coils with 63.36%, followed by BRCA2 gene with 59.71%, Rb1 with 45.04% and BRCA1 having the lowest percentage for random coil with 34.60%. Tertiary protein structure of tumor suppressor genes are represented in figure 2-5. The Alpha helix is represented by Blue color, the extended strand is represented by Pink color, the Random coil is represented by Yellow color and the Beta turn by Green color. Phylogenetic analysis of tumor suppressor genes in humans showed a phylogenic tree rooted from a common ancestor having three clusters. Each branch or cluster has similar nucleotide sequences, the secondary and tertiary protein sequence structure were model for the representative(s) of each branch or cluster as represented in Figure 2-5. BRCA2 and Rb1 are represented in cluster one, cluster two is p53 and cluster three is BRCA1 which is the common ancestor of all sampled genes.

Subject	Gene Name	GenBank Accession Number	Protein Name	GenBank Accession Number For Protein	Nucleotide Sequence Length	Amino Acid Sequence Length
Homo sapiens	P53mRNA	AB0829231	p53	BAC16799.1	2451	393
Homo sapiens	BRCA1mRNA	U37574.1	BRCA1	AAC50330.1	3798	26
Homo sapiens	BRCA2mRNA	MF769708.1	BRCA2	ASW23229.1	619	206
Homo sapiens	Rb1mRNA	M33647.1	Rb1	AAA69806.1	4600	928

Table 1:- Retrieved Nucleotide and Amino Acid Sequences of Human p53, BRCA1, BRCA2 and Rb1 Tumor Suppressor Genes

Genes	G ≡ C Content (%)
p53	52.59
BRCA1	50.07
BRCA2	31.34
Rb1	36.02

Table 2:- Guanine ≡ Cytosine Content of p53, BRCA1, BRCA2 and Rb1 Tumor Suppressor Genes in Humans

ISSN No:-2456-2165

Genes	Sequence Length	Number of occurrence for alpha helix	Number of occurrence for extended strand	Number of occurrence for random coil	Alpha Helix (%)	Extended Strand (%)	Random coil (%)
p53	393	71	73	249	$\frac{71}{393} X\ 100 = 18.07$	$\frac{73}{393} X\ 100 = 18.58$	$\frac{249}{393} X\ 100 = 63.36$
BRCA1	26	12	5	9	$\frac{12}{26} X\ 100 = 46.15$	$\frac{5}{26} X\ 100 = 19.23$	$\frac{9}{26} X\ 100 = 34.6$
BRCA2	206	49	34	123	$\frac{49}{206} X\ 100 = 23.79$	$\frac{34}{206} X\ 100 = 16.50$	$\frac{123}{206} X\ 100 = 59.71$
Rb	928	405	105	418	$\frac{405}{928} X\ 100 = 43.64$	$\frac{105}{928} X\ 100 = 11.31$	$\frac{418}{928} X\ 100 = 45.04$

Table 3:- Secondary Protein Structures of Tumor Suppressor Genes in Humans

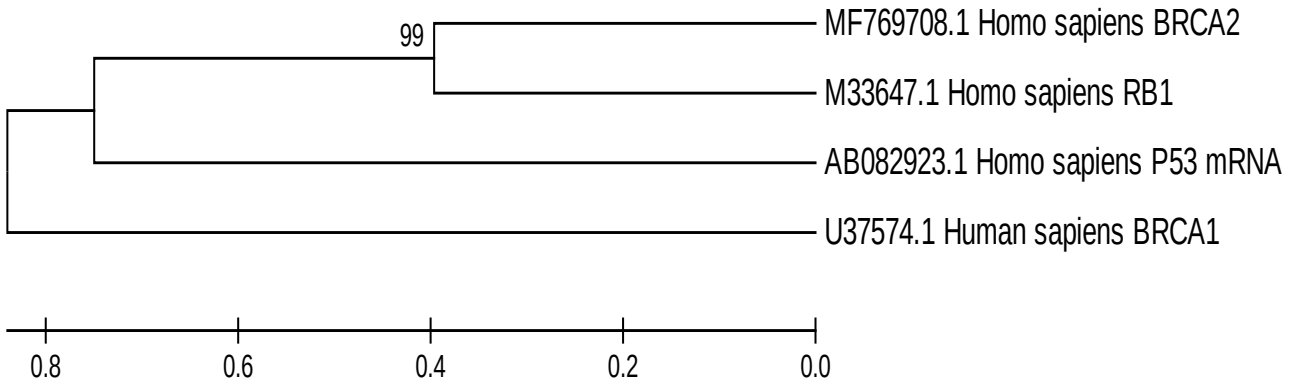

Fig 1:- UPGMA Phylogenetic Tree Showing the Evolutionary Relationship among Tumor Suppressor Genes in Humans

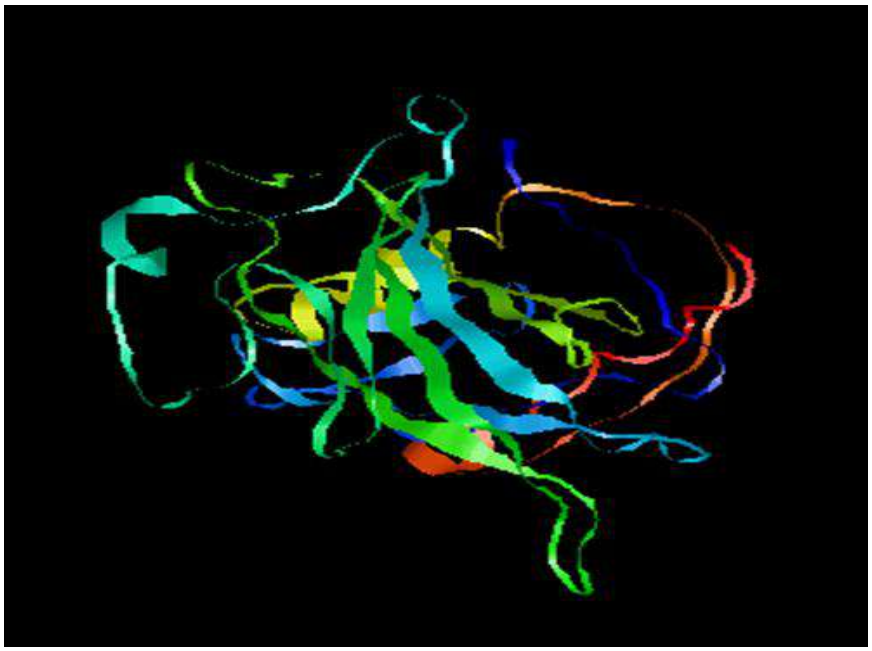

Fig 2:- Tertiary Protein Structure of p53 Tumor Suppressor Gene in Humans

ISSN No:-2456-2165

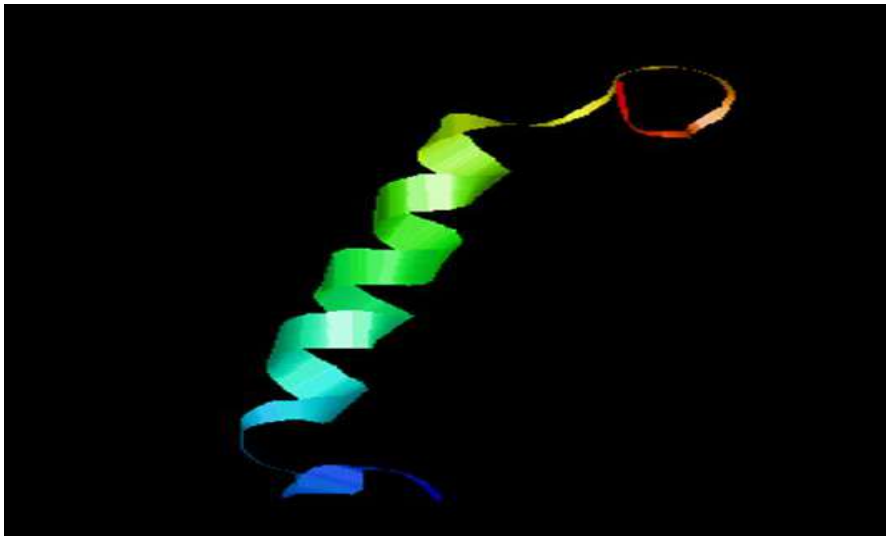

Fig 3:- Tertiary Protein Structure of BRCA1 Tumor Suppressor Gene in Humans

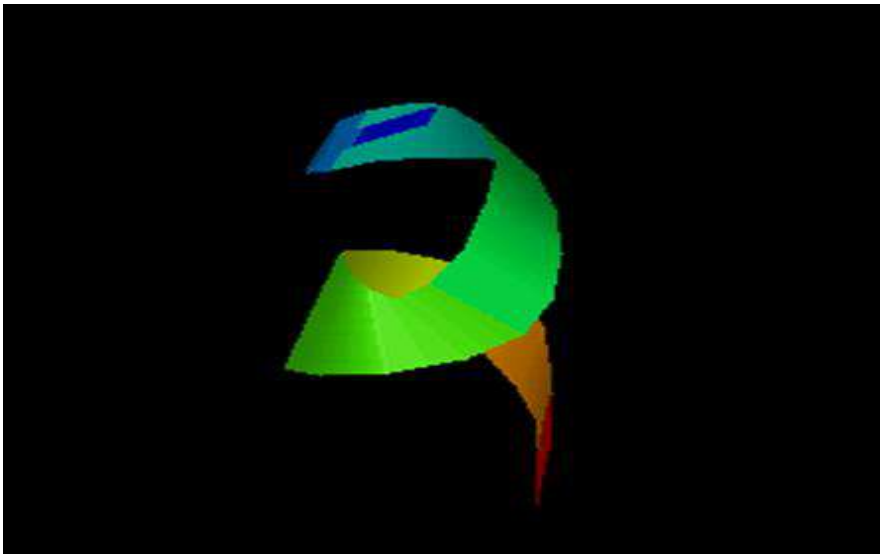

Fig 4:- Tertiary Protein Structure of BRCA2 Tumor Suppressor Gene in Humans

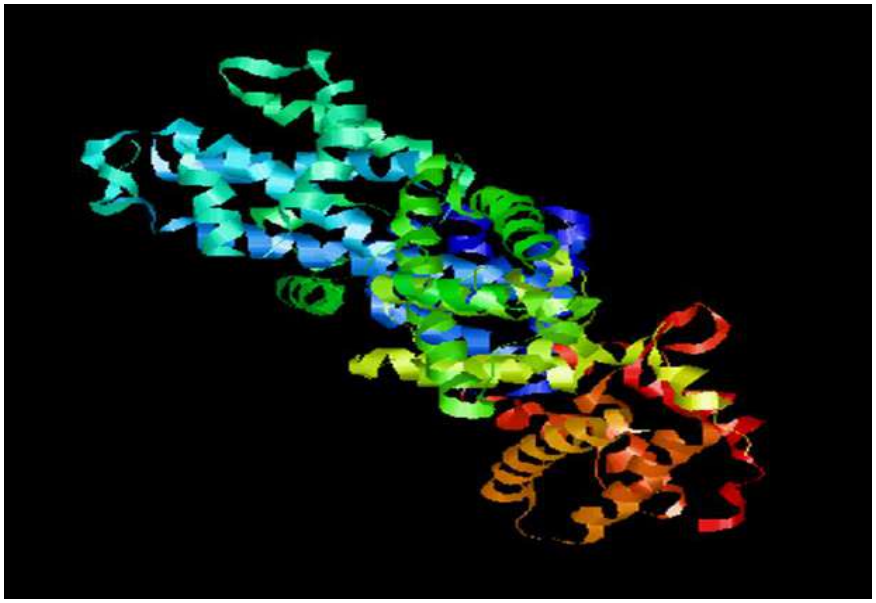

Fig 5:- Tertiary Protein Structure of Rb Tumor Suppressor Gene in Humans

IV. DISCUSSION

In silico analysis of the protein structures of Rb, p53, BRAC1 and BRAC2 genes revealed a high percentage of α-Helix content of the secondary protein structures of the genes; which provides maximal hydrogen bonding for peptide bond components within the interior polypeptides [8]. There was a low percentage of extended strands in the secondary protein structure of the genes, Extended secondary structures however play key roles in the final settlement and preservation of the protein conformation and, by extension, of the structural properties of the protein domains when they are present, or of the intrinsically disordered proteins when they are rare or absent [9]. Random coils were however present in high percentages with p53 having the highest percentage, random coils represent denatured proteins, inferring that the p53 undergoes denaturation more than the other genes which may explain why p53 function is lost in over 50% of human cancers, demonstrating the major role of this protein in tumorigenesis [10] and this is concomitant with previous reports that when subjected to high hydrostatic pressure, wild type p53 acquires a distinct folding resembling that of the R248Q p53C point mutation [11].

The high G-C content recorded in our study in the p53 and BRAC1 genes is believed to confer stability and is significantly associated with gene expression pattern. This implies that the p53 and BRAC1 may be more stable than the Rb and BRAC1 genes however this requires further research. Roa *et al.* [12] demonstrated the impact of GC content on gene expression pattern of vertebrates, the results revealed that the GC content of the gene was significantly and positively correlated with the gene expression level though conflicting reports have been published on the relationships between gene expression and GC content in mammalian chromosomes [13, 14, 15, 16]. It is an established fact that chromosomal regions with high G-C have high gene densities and CpG islands [17]. The phylogenetic analysis of the genes defined their evolutionary history and their relationship by dividing them into three clusters (branches) having BRCA1 as the common ancestral root, the representatives of cluster one are BRCA2 and Rb, the representative of cluster two is p53 and the representative of cluster three being BRCA1. This is similar to the findings of Chakrabarti *et al.* [18] with respect to ancestral background. The homolog modeling of the genes from the representative of each cluster (branch) revealed similar protein structure (tertiary protein structure) in members of the same cluster pointing that all members in each cluster exhibit similar protein function. [19, 20].

V. CONCLUSION

This study revealed the secondary structure, tertiary structure and the GC content of Rb, p53, BRCA1 and BRCA2 human tumor suppressor genes and assessed the phylogenetic relationship between these genes. The p53 and BRAC1 were observed to have a higher GC content than Rb and BRAC2 genes. The phylogenetic tree showed evolutionary relationship between the genes with BRCA1 gene being the common ancestor of all sampled genes and BRCA2 and Rb arising from the same cluster. This study also may become useful in emerging anti-cancer strategies such as selective inhibition of pathways.

REFERENCES

[1]. Müller, H. & Helin, K. (2000). The E2F transcription factors: key regulators of cell proliferation. BiochimBiophysActa, 1470:M1–12.
[2]. Wang, T.J., Huang, M.S., Hong, C.Y., Tse, V., Silverberg, G.D., & Hsiao, M. (2001). Comparisons of tumor suppressor p53, p21,and p16 gene therapy effects on glioblastoma tumorigenicity in situ. BiochemBiophys Res Commun, 287:173–80.
[3]. Levine, A.J., Finlay, C.A., & Hinds, P.W. (2004). P53 is a tumor suppressor gene. Cell, 116, S67–69.
[4]. Vandel, L., Nicolas, E., Vaute, O., Ferreira, R., Ait-Si-Ali, S., & Trouche D. (2001).Transcriptional repression by the retinoblastoma protein through the recruitment of a histone methyltransferase. Molecular Cell Biology, 21, 6484–94.
[5]. Cadwell, C., & Zambetti, G.P. *(*2001*).*The effects of wild-type p53 tumor suppressor activity and mutant p53 gain-of-function on cell growth. Gene, 277, 15–30.
[6]. Sherr, C.J. (2004). Principles of tumor suppression. Cell, 116, 235–246.
[7]. Champe, P. C., Harvey, R. A., & Perrier, D. R. (2008).Lippincott's Illustrated Reviews; Biochemistry, Williams and Wilkins (4th edition), New York, USA.
[8]. Tompa, P. (2011).Unstructural biology coming of age. Current opinions in structural biology, 21, 419-425.
[9]. Levesque, A. A., & Eastman, A. (2007).Carcinogenesis, 28, 13-20.
[10]. Ishimaru, D., Maia, L. F., Maiolino, L. M., Quesado, P. A., Lopez, P. C., Almeida, F. C., Valente, A. P., & Silva, J. L. (2003). Journal of molecular biology, 333, 443-451.
[11]. Roa, Y. S., Chai, X. W., Nie, Q. H., & Zhang, X. Q. (2013). Impact of GC content of gene expression pattern in chicken. Journal of Genetics, Selection and Evolution, 45(1), 2013-2018.
[12]. Goncalves, I., Duret, L., & Mouchiroud, D. (2000) Nature and structure of human genes that generate retropseudogenes. Genome Res.10:672–678.
[13]. Margulies, E. H., Kardia, S. L., & Innis, J.W. (2001) Identification and prevention of a GC content bias in SAGE libraries. Nucl. Acids Res, 29, 60
[14]. Versteeg, R., van Schaik, B. D., van Batenburg, M. F., Roos, M., Monajemi, R., Caron, H., Bussemaker, H. J., & van Kampen, A. H. (2003) The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes. Genome Res, 13, 1998–2004.
[15]. Semon, M., Mouchiroud, D., & Durent, L. (2005) Relationship between gene expression and GC content in Mammals Hum Mol Genet, 14(3), 421-227

[16]. Lercher, M.J., Urrutia, A.O., Pavlicek, A., & Hurst, L.D. (2003) A unification of mosaic structures in the human genome. Hum. Mol. Genet, 12, 2411–2415.
[17]. Chakrabartis, S., Kaur, I., Mandal, A. K., Parikh, R. S., Thomas, R., & Majumder, P. P. (2006). Globally, CYPIBI mutations in primary congenital glaucoma are strongly structured by geographic and haplotype backgrounds. Investigations ophthalmology and visions of Sciences. 47(1), 43-47.
[18]. Sarfarazi M., & Stoilov, I. (2000) "Molecular Genetics of Primary Congenital Glaucoma," Eye 14, 422- 428. doi:10.1038/eye.2000.126
[19]. Acharya, M. S., Reddy, A. B., Chakrabarti, S., Panicker, S. G., Mandal, A. K., Ahmed, N., Balasubramanian, D., Habnian, S. E. and Nagarajaram, H. A. (2006). Disease causing mutation in proteins: structural analysis of the CYP1B1 mutations causing primary congenital glaucoma in humans. Biophysical Journal, 91, 4329-4339.