Life Science Informatics Publications

Research Journal of Life Sciences, Bioinformatics, Pharmaceutical and Chemical Sciences

Journal Home page http://www.rjlbpcs.com/

RJLBPCS

ISSN 2454-6348

Original Research Article **DOI -10.26479/2018.0401.01**

PROTEOMIC CHARACTERIZATION OF BRCA1 GENE AND ITS ROLE IN CANCER PROGRESSION

Kankana Goswami[1], Varshitha. J[2], Naresha. S. S[3], Mahesh Pattabhiramaiah[4]*

Centre for Applied Genetics, Department of Zoology, Janabharathi Campus, Bangalore University, Bengaluru-560056

ABSTRACT: BRCA1 is a human gene which provides instructions for making a protein that acts as a tumor suppressor. The BRCA1 protein is involved in repairing damaged DNA and, therefore, plays critical role in maintaining the stability of a cell's genetic information. When this gene is mutated, its protein product is either not made or functions improperly. As a result, cells are more likely to develop additional genetic alterations that lead to the development of cancer. Our bio-computational analysis of the BRCA1 protein revealed the molecular weight to lie between 107 and 759 Da. The aliphatic index of protein ranged from 57.50 to 82.81 indicating a positive factor for thermostability. Instability index studies predicted that most of the BRCA1 sequences are computationally unstable. Primary structure analysis using CLC workbench indicated the presence of hydrophobic residues and secondary structure of protein using self-optimized prediction method with alignment (SOPMA) tool predicted that most of the helices ranged from 28.10 to 49.64% and coils from 24.46 to 56.41%, respectively. Three-dimensional structure was predicted using phyre2 server. Protein functionality was predicted by SVM method confirms the presence of All DNA binding sites, Metal binding sites etc.

KEYWORDS: BRCA1, CLC, SVM, SOPMA, phyre2

***Corresponding Author: Dr. Mahesh Pattabhiramaiah Ph.D.,**

Centre for Applied Genetics, Department of Zoology, Janabharathi Campus, Bangalore University, Bengaluru-560056 * Email Address: dr.pmahesh@gmail.com

1.INTRODUCTION

Computational biology or Insilco approach is developing day by day with refinement. It is becoming a promising field and with the help of this the time and cost of biological work related to drug discovery, molecular interaction is reducing [1]. Bioinformatics has revolutionized the field of molecular biology. Prediction of protein function is important application of bioinformatics [2]. Sequence analysis and physio-chemical characterization of proteins using bio- computation tools have been done by many researchers and reported [3]; [4]; [5]; [6]. The amino acid sequence provides most of the information required for determining and characterizing the molecule's function, physical and chemical properties. From a protein function standpoint, transfer of annotation from known proteins to a novel target is currently the only practical way to convert vast quantities of raw sequence data into meaningful information. Further Computer-aided techniques for the efficient identification and optimization of novel molecules with a desired biological activity have become a part of the drug discovery process. The integration of wet experiments and the use of bioinformatics analysis have become an indispensable part of the biological and clinical research of this century. The raw sequence information of proteins and nucleic acid can be converted to analytical and relative information with the help of soft computing tools. Cancer is the name for diseases in which cells become abnormal and divide without control [7]. It is the second leading cause of death among populations after cardiovascular disease and one of the most important health problems of the current era [8]. Out of more than 100 different types of cancer, Breast cancer is the second leading cause of cancer death in women only next to lung cancer, affecting millions of people worldwide [8]. Cancer incidence and mortality statistics reported by the American Cancer Society and other resources estimated the annual incidence for 2017 had to be 40,000 cases or more. The most common type of cancer is breast cancer, with more than 255,000 new cases expected in the United States in 2017. The occurrence of breast cancer in India is on the escalation and is quickly becoming the number one cancer in females. More and more numbers of patients being diagnosed with breast cancer to be in the younger age groups (in their thirties and forties). Breast cancer is now the most common cancer in most cities in India, and 2nd most common in the rural areas. Over 16,000 new cases are registered every year with more than 20 % from neighbouring states of Andhra Pradesh and Tamil Nadu and Kerala. Of the estimated 45,000 new cancer cases reported every year in Karnataka, over 8,000 are breast cancer cases. The BRCA1 gene is located on 17q21 and has a total length of about 100 kb. This gene consists of 24 exons, and the coding region starts at the middle of exon [9]. The gene product of BRCA1 is a phosphorylated protein that consists of 1863 amino acids and has a molecular weight of 220kDa [10].BRCA1 is expressed in the cells of breast and other tissues, where it helps to repair damaged DNA, or destroys cells if DNA cannot be repaired. If BRCA1 itself is damaged, damaged DNA is not repaired properly and this increases risks for cancers [11]. The protein encoded by the BRCA1 gene

combines with other tumor suppressors, DNA damage sensors, and signal transducers to form a large multisubunit protein complex known as the BRCA1-associated genome surveillance complex [12]. Women with defects in either the BRCA1 gene have a greater than 80 per cent chance of developing breast cancer [13]. BRCA1 germline mutations also confer increased risks of pancreatic and perhaps uterine and cervical cancer [14]. Recent studies have shown that small deletions, insertions, nonsense mutations and splicing aberrations account for 87% of all pathogenic mutations of the BRCA1 gene, resulting in the generation of BRCA1 protein [15]. Keeping in view the importance and applications of BRCA1 gene in cancer, computational analysis was performed to determine the physicochemical characteristics of BRCA1 so as to pave the way to find out better understanding and novel response. Our research study presents novel insights into the structural, functional, annotational features of BRCA1 gene which might play a significant role in progression and management of the disease.

2. MATERIALS AND METHODS

2.1 BRCA1 protein sequence retrieval: The protein sequence of BRCA1 (8 sequences) in humans were retrieved from NCBI (National Centre for Biotechnology Information) database [16].

2.2 Amino acid frequency: The amino acid frequency of 8 BRCA1 protein sequences were computed using the tool CLC free workbench (CLC bio,2006) (www.clc.bio.com/..../clc-main-workbench/) [16].

2.3 Protein primary structure analysis: Percentages of hydrophobic and hydrophilic residues were calculated from the primary structure analysis [16].

2.4 Physico-chemical parameters: The physio-chemical parameters, isoelectric point (pI), molecular weight, total number of positive and negative residues, extension coefficient [17], half-life [18], instability index [19], aliphatic index [20] and grand average hydropathy (GRAVY) [21] were computed using Expasy protparam (http:/us.expasy.org/tools/protparam.html) prediction server [16].

2.5 SVM prot analysis: The classification and functions of proteins was analyzed using SVM Prot web software (http://jing.cz3.nus.edu.sg/cgi-bin/svmprot.cgi.) [16].

2.6 Protein secondary structure prediction: The secondary structure was predicted by self-optimized prediction method with alignment (SOPMA) [22]. SOPMA was employed for calculating the secondary structural features of the selected protein sequences considered in this study.

2.7 Protein Tertiary structure prediction: Tertiary structure prediction of Breast cancer causing proteins was performed using bioinformatics tool Phyre2 (www.sbg.bio.ic.ac.uk/phyre2/index.cgi).

2.8 RNA structure prediction: The RNA structure was predicted using (www.rna.urmc.rochester.edu/RNAstructureweb/servers/predict1.html).

2.9 Transmembrane region prediction: Transmembrane helices were predicted by the TMPred Expasy software (http://embnet.vital-it.ch/software/TMPRED_form.html) [23].

2.10 Domain architecture analysis: Domain organization and domain composition was analyzed using Simple Modular Architecture Research Tool(SMART) *(*http://smart.embl-heidelberg.de*).*

2.11 Ramachandran Plot Assessment: Ramachandran Plot was used to visualize backbone dihedral angles ϕ against ψ of amino acid residues in protein structure using RAMPAGE. (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) [16].

2.12 DNA **Barcode**: Barcode was generated for the reverse transcribed DNA sequences using Bio-Rad Barcode Generator *(*http://www.biorad-ads.com/BarcodeWeb/).

2.13 Protein QR Code: QR code was generated for the reversed transcribed DNA sequences using QR Code Generator (http://www.goqr.me/).

3.RESULT AND DISCUSSION

3.1 SEQUENCE RETRIEVAL FROM NCBI

Sl. No	ACCESSION NO.	SEQUENCE DESCRIPTION	LENGTH	PROTEIN SEQUENCE
1.	AAI06747.1	BRCA1 protein, partial [*Homo sapiens*].	259aa	RPSPSALGKAACEFSETDVTNTEHHQPSNNDLNTTEKRAAERHPEKYQGSSVS NLHVEPCGTNTHASSLQHENSSLLLTKDRMNVEKAEFCNKSKQPGLARSQHN RWAGSKETCNDRRTPSTEKKVDLNADPLCERKEWNKQKLPCSENPRDTEDVP WITLNSSIQKVNEWFSRSDELLGSDDSHDGESESNAKVADVLDVLNEVDEYS GSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKKASLPN
2.	AKJ84699.1	BRCA1 protein, partial [*Homo sapiens*].	210aa	EQTSKRHDSDTFPELKLTNAPGSFTKCSNTSELKEFVNPSLPREEKEEKLETVK VSNNAEDPKDLMLSGERVLQTERSVESSSISLVPGTDYGTQESISLLEVSTLGK AKTEPNKCVSQCAAFENPKGLIHGCSKDNRNDTEGFKYPLGHEVNHSRETSIE MEESELDAQYLQNTFKVSKRQSFAPFSNPGNAEEECATFSAHSGSLKKQ
3.	ALO20345.1	BRCA1 protein, partial [*Homo sapiens*].	740aa	FSPYLISDNLEQPMGSSHASQVCSETPDDLLDDGEIKEDTSFAENDIKESSAVFS KSVQRGELSRSPSPFTHTHLAQGYRRGAKKLESSEENLSSEDEELPCFQHLLFG KVNNIPSQSTRHSTVATECLSKNTEENLLSLKNSLNDCSNQVILAKASQEHHLS EETKCSASLFSSQCSELEDLTANTNTQDPFLIGSSKQMRHQSESQGVGLSDKEL VSDDEERGTGLEENNQEEQSMDSNLGEAASGCESETSVSEDCSGLSSQSDILTT QQRDTMQHNLIKLQQEMAELEAVLEQHGSQPSNSYPSIISDSSALEDLRNPEQS TSEKAVLTSQKSSEYPISQNPEGLSADKFEVSADSSTSKNKEPGVERSSPSKCPS LDDRWYMHSCSGSLQNRNYPSQEELIKVVDVEEQQLEESGPHDLTETSYLPR QDLEGTPYLESGISLFSDDPESDPSEDRAPESARVGNIPSSTSALKVPQLKVAES AQSPAAAHTTDTAGYNAMEESVSREKPELTASTERVNKRMSMVVSGLTPEEF MLVYKFARKHHITLTNLITEETTHVVMKTDAEFVCERTLKYFLGIAGGKWVV SYFWVTQSIKERKMLNEHDFEVRGDVVNGRNHQGPKRARESQDRKIFRGLEI CCYGPFTNMPTDQLEWMVQLCGASVVKELSSFTLGTGVHPIVVVQPDAWTE

				DNGFHAIGQMCEAPVVTREWVLDSVALYQCQELDTYLIPQIPHSHY
4.	AAB08105.1	BRCA1 protein, partial [*Homo sapiens*].	233aa	SVSREKPELTASTERVNKRMSMVVSGLTPEEFMLVYKFARKHHITLTNLITEE TTHVVMKTDAEFVCERTLKYFLGIAGGKWVVSYFWVTQSIKERKMLNEHDF EVRGDVVNGRNHQGPKRARESQDRKIFRGLEICCYGPFTNMPTDQLEWMVQ LCGASVVKELSSFTLGTGVHPIVVVQPDAWTEDNGFHAIGQMCEAPVVTREW VLDSVALYQCQELDTYLIPQIPHSHY
5.	AKG51647.1	BRCA1 protein, partial [*Homo sapiens*].	139aa	LELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQCPLCKNDITKRSLQESTRFSQ LVEELLKIICAFQLDTGLEYANSYNFAKKENNSPEHLKDEVSIIQSMGYRNRA KRLLQSEPENPSLILHAETSQPEERAFTVSFM
6.	AAC00049.1	BRCA1 protein, partial [*Homo sapiens*].	759aa	MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQ KKGPSQCPLCKNDITKRSLQESTRFSQLVEELLKIICAFQLDTGLEYANSYNFA KKENNSPEHLKDEVSIIQSMGYRNRAKRLLQSEPENPSLQETSLSVQLSNLGTV RTLRTKQRIQPQKTSVYIELGSDSSEDTVNKATYCSVGDQELLQITPQGTRDEI SLDSAKKAACEFSETDVTNTEHRQPSNNDLNTTEKRVAERHPEKYQGEAASG CESETSVSEDCSGLSSQSDILTTQQRDTMQHNLIKLQQEMAELEAVLEQHGSQ PPNSYPSIISDSSALEDLRNPEQSTSEKVLTSQKSSEYPISQNPEGLSADKFEVSA DSSTSKNKEPGVERSSPSKCPSLDDRWYMHSCSGSLQNRNYPSQEELIKVVDV EEQQLEESGPHDLTETSYLPRQDLEGTPYLESGISLFSDDPESDPSEDRAPESAR VGNIPSSTSALKVPQLKVAESAQGPAAAHTTDTAGYNAMEESVSREKPELTAS TERVNKRMSMVVSGLTPEEFMLVYKFARKHHITLTNLITEETTHVVMKTDAE FVCERTLKYFLGIAGGKWVVSYFWVTQSIKERKMLNEHDFEVRGDVVNGRN HQGPKRARESQDRKIFRGLEICCYGPFTNMPTDQLEWMVQLCGASVVKELSS FTLGTGVHPIVVVQPDAWTEDNGFHAIGQMCEAPVVTREWVLDSVALYQCQ ELDTYLIPQIPHSHY
7.	ALO20343.1	BRCA1 protein, partial [*Homo sapiens*].	107aa	NCKHPEIKKQEYEEVVQTVNTDFSPYLISDNLEQPMGSSHASQVCSETPDDLL DDGEIKEDTSFAENDIKESSAVFSKSVRKESLAGVLALSPIHIWLRVTEEGPRN
8.	ABB87061.1	BRCA1 protein, partial [*Homo*	39aa	NDIKESSAVFSKSVQKGELSRSPSPFTHTHLAQGYRRGA

Table 1: BRCA1 protein sequences retrieved from NCBI

The results of Primary sequence analysis of BRCA 1 proteins computed by CLC work bench revealed the sequence length ranging from 39-759 aa (Table 1).

3.2 FREQUENCY OF AMINO ACIDS:

Amino acid	AAI06747.1	AKJ84649.1	ALO20345.1	AAB08105.1	AKG51647.1	AAC00604.1	ALO20343.1	ABB87061.1
Alanine (A)	15	10	38	10	6	38	5	3
Cysteine (C)	7	5	16	6	6	20	2	0
Aspartic acid(D)	18	8	41	9	4	38	8	1
Glutamic acid(E)	27	27	80	20	14	76	13	2
Phenylalanine (F)	4	9	20	10	7	18	3	2
Glycine (G)	9	11	37	14	3	34	4	3
Histidine (H)	10	5	22	9	3	18	3	2
Isoleucine (I)	6	4	24	10	8	32	6	1
Lysine (K)	22	18	34	12	13	43	7	3
Leucine (L)	20	17	64	17	17	66	8	2
Methionine (M)	1	2	15	8	3	16	1	0
Asparagine (N)	20	13	31	7	8	32	5	1
Proline (P)	13	11	36	10	7	40	6	2
Glutamine (Q)	7	8	43	10	8	47	4	2
Arginine (R)	13	7	28	13	6	34	3	3
Serine (S)	33	27	100	13	13	82	13	7
Threonine (T)	14	15	45	18	6	50	5	2
Valine (V)	13	10	44	25	4	49	8	2
Tryptophan (W)	4	0	6	5	0	6	1	0
Tyrosine (Y)	3	3	16	7	3	20	2	1

Table:2 Representation of frequency of amino acids in BRCA1

The frequency of occurrence of amino acid retrieved from NCBI is tabulated in Table 2.The abundant amino acids were serine, (100aa in ALO20343.1), followed by glutamic acid and leucine and the least being tryptophan (W) with 0aa in majority of the sequences under consideration which

are tabulated in CLC work bench (Table 2). Serine provides the essential precursors for the synthesis of proteins, nucleic acids, and lipids that are crucial to cancer cell growth. Increased serine biosynthesis is one of many metabolic changes that have been reported in cancer cells, and serine is a central node for the biosynthesis of many molecules. Serine is a precursor of the nonessential amino acids glycine and cysteine. Moreover, biosynthesis of serine also affects cellular antioxidative capacity, thus supporting tumour homeostasis. Therefore, an increase in serine availability could be valuable for proliferating cancer cells. The antagonists of the enzyme L-glutamine synthetase synthesized from glutamic acid can interfere with the metabolic role of L-glutamine and act as anti-cancer agents[24]. Further work is required to determine the role of serine in tumor growth and to better understand how tumor dependence on the serine can be exploited for therapeutic benefit. Elucidation of these aspects, and others, might provide alternative clinical strategies, potentially enabling the design of personalised therapy for cancer patients.

3.3 PRIMARY STRUCTURE ANALYSIS:

Accession No.	Hydrophobic %	Hydrophilic %
AAI06747.1	31.94	68.06
AKJ84699.1	31.94	68.06
ALO20345.1	34.34	65.66
AAB08105.1	41.92	58.08
AKG51647.1	38.35	61.65
AAC00049.1	35.34	64.66
ALO20343.1	38.15	61.85
ABB87061.1	30.77	69.23

Table No:3 Hydrophobic and Hydrophilic profile of BRCA1 proteins.

The result of primary structure analysis suggests that most of the BRCA1 proteins are hydrophilic in nature due to presence of high polar residues content. The highest hydrophilic count is seen in sequence ABB87061.1 with 69.23% and the least hydrophilic count was seen in sequence AAB08105.1 with 58.08%(Table3).

3.4 PHYSICO-CHEMICAL PARAMETERS:

Sl. NO	ACCESSION NO.	SEQUENCE LENGTH	M.Wt.(Da)	pI	-R	+R	EC REDUCED	EC NON-REDUCED	II(INSTABILITY INDEX)	AI(ALIPHATIC INDEX)	HALF-LIFE (In hours)	GRAVY
1.	AAI06747.1	259	29014.74	5.53	45	35	0.912	0.925	45.17	59.50	1.0	-1.086
2.	AKJ84699.1	210	23253.5	5.10	35	25	0.192	0.203	49.50	57.57	1.0	-0.893

			9									
3.	ALO20345.1	740	82036.26	4.67	121	62	57840	57840	55.93	68.76	1.1	-0.666
4.	AAB08105.1	233	26698.60	6.14	29	25	38305	37930	30.26	80.60	1.9	-0.268
5.	AKG51647.1	139	16012.46	7.76	18	19	4845	4470	58.57	82.81	5.5	-0.494
6.	AAC00049.1	759	85028.07	5.00	114	77	64050	62800	53.50	74.08	30	-0.611
7.	ALO20343.1	107	11950.17	4.52	21	10	8605	8480	49.97	77.38	1.4	-0.642
8.	ABB87061.1	133	15034.97	5.77	15	12	3105	2980	58.68	79.85	0.8	-0.457

Table No 4: Parameters computed using Expasy Protparam (BRCA1).

The length of the 8 BRCA1 protein sequences ranged from 107-759aa.The Molecular weight of BRCA1 proteins ranged from (11950.17-85028.07 Da) (Table 4). Extinction co-efficient of BRCA1 at 280nm ranged from 0.192-64050 M^{-1} Cm^{-1} (Table 4). The isoelectric point (pI) of a protein is the pH where the proteins has no net charge. The results of isolectric point (pI) of BRCA1 proteins ranged from (4.67-7.76) which reveals that most of the BRCA1proteins were found to be acidic in nature except for seq. AKG51647.1 . Computed isoelectric point of proteins > 7 are soluble in basic buffers which are useful for developing buffer system for purification of proteins (Table 4). The Grand Average hydropathy (GRAVY values) showed that all proteins are hydrophilic ranging from -0.268 to -1.086 supports the soluble nature of BRCA1 proteins. Though, it can play a role in substrate recognition. Here the protein sequences showing negative that indicates stability of the protein. (Table 4) The Aliphatic index of a protein is defined as the relative volume occupied by aliphatic side chains which ranges from 57.50 to 82.81(Table 4) which infers positive factor for thermostability [24]. A protein whose instability index is smaller than 40 are predicted as stable, and a value above 40 predicts that the protein may be unstable, [19] only one BRCA1 protein sequence, AAB08105.1 showed stability with instability index of 30.26 (Table 4). Short half-life is one of the key challenges in the field of therapeutic peptides. Most of our BRCA1 protein sequences showed shorter half-life with exception of only one BRCA1 sequence AAC00049.1 (Table 4).

3.5 SVM PROT ANALYSIS:

Accession No.	Metal-binding	Zinc-binding	Actin-binding proteins	Other protein families
	Probability Value (%)	Probability Value (%)	Probability Value (%)	Probability Value (%)
AAI06747.1	N/A	95.2	N/A	58.6
AKJ84699.1	N/A	80.4	58.6	58.6
ALO20345.1	62.2	98.8	N/A	58.6
AAB08105.1	N/A	N/A	N/A	58.6
AKG51647.1	N/A	N/A	N/A	58.6
AAC00049.1	80.4	98.9	58.6	58.6
ALO20343.1	58.6	73.8	58.6	58.6
ABB87061.1	N/A	62.2	58.6	N/A

Table No 5: SVM Prot analysis of Proteins encoded by BRCA1 genes.

Support vector machines method for the classification of proteins with diverse sequence distribution. It has been employed in protein studies including protein–protein interaction prediction, fold recognition, solvent accessibility and structure prediction. Among the selected BRCA1 protein sequences, AAC00049.1 has the highest probability values of metal-binding, zinc binding, actin-binding proteins and other protein families in comparison to the rest, hence resulting in high chances of breast cancer with alterations in this particular gene (Table5). Zinc deficiency can lead to DNA damage and the initiation of cancer. Cancer cells develop mechanisms to shut down zinc efflux and maintain intracellular concentrations when availability is reduced. This demonstrates the importance of intracellular zinc for rapidly dividing cancer cells.

3.6 SECONDARY STRUCTURE PREDICTION:

Sl.No.	Accession No.	α helix (%)	β turn (%)	Extended Strand (%)	Random Coil (%)
1.	AAI06747.1	31.27%	3.09%	10.42%	55.21%
2.	AKJ84699.1	28.10%	10.00%	14.29%	47.62%
3.	ALO20345.1	38.24%	8.11%	14.86%	38.78%
4.	AAB08105.1	38.20%	12.02%	25.32%	24.46%
5.	AKG51647.1	49.64%	5.76%	11.51%	33.09%
6.	AAC00049.1	37.02%	8.04%	15.81%	39.13%
7.	ALO20343.1	41.12%	8.41%	11.21%	39.25%
8.	ABB87061.1	28.21%	12.82%	25.56%	56.41%

Table No:6 Representation of helix, turn, extended strand and coil by online tool SOPMA.

The self-optimized, neural network based alignment tool SOPMA was used for prediction of secondary structural features. This method calculates the content of α-helix, βsheets, turns, random coils and extended strands listed in (Table 6). The secondary structure of BRCA 1 contain more of helices and coils, where helix range from 28.10 -49.64% and coils from 24.46-56.41%. Sheets and turns are in less abundance with maximum percentage being 25.56% and 12.82% respectively. The present investigation throws light on the presence of abundant random coils. This has strengthened our understanding of the temperature, stress adaptation of proteins and results of the present study can be explored further to improve and design the properties of BRCA1 for desired function.

3.7 PROTEIN TERTIARY STRUCTURE PREDICTION:

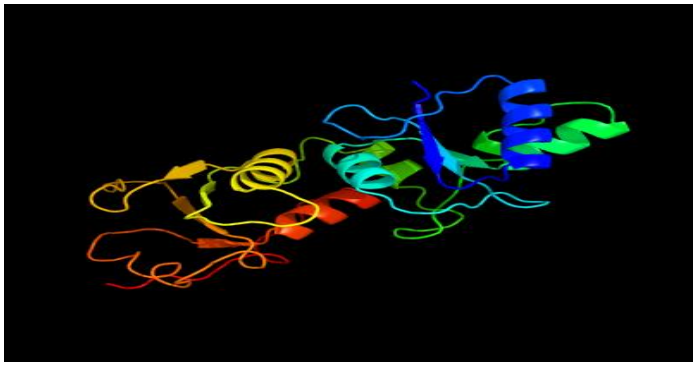


Figure-1 Tertiary structure of BRCA1 protein predicted by PHYRE2 (ALO20345.1)

The primary amino acid sequences of BRCA1 retrieved from NCBI(table1) were submitted to phyre2 server using intensive mode (Fig 1).The phyre2 threading server combined HHsearch for remote homology detection based on pairwise comparison of hidden markov models (HMM) With ab initio and multiple – template modeling. The library of known structures for comparison by phyre2 was from the protein date bank (PDB) and structural classification of proteins (SCOP) data bases.

3.8 RNA STRUCTURE PREDICTION:

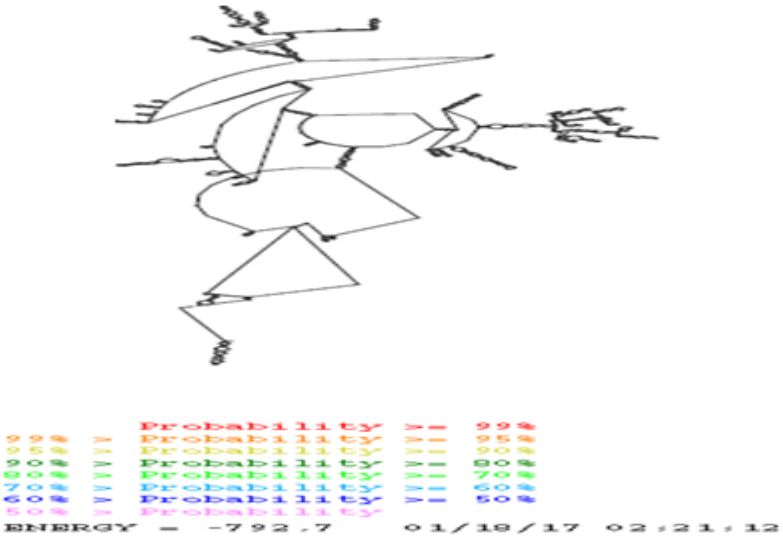


Figure-2 RNA structure of BRCA1 using online server (ALO20345.1)

Sl.No.	Accession No.	Free energy
1.	AAI06747.1	-242.6
2.	AKJ84699.1	-179.7
3.	ALO20345.1	- 792.7
4.	AAB08105.1	-244.9
5.	AKG51647.1	-120.9
6.	AAC00049.1	-812.7
7.	ALO20343.1	-90.1
8.	ABB87061.1	-46.8

Table no. 7: Free energy for the predicted RNA:

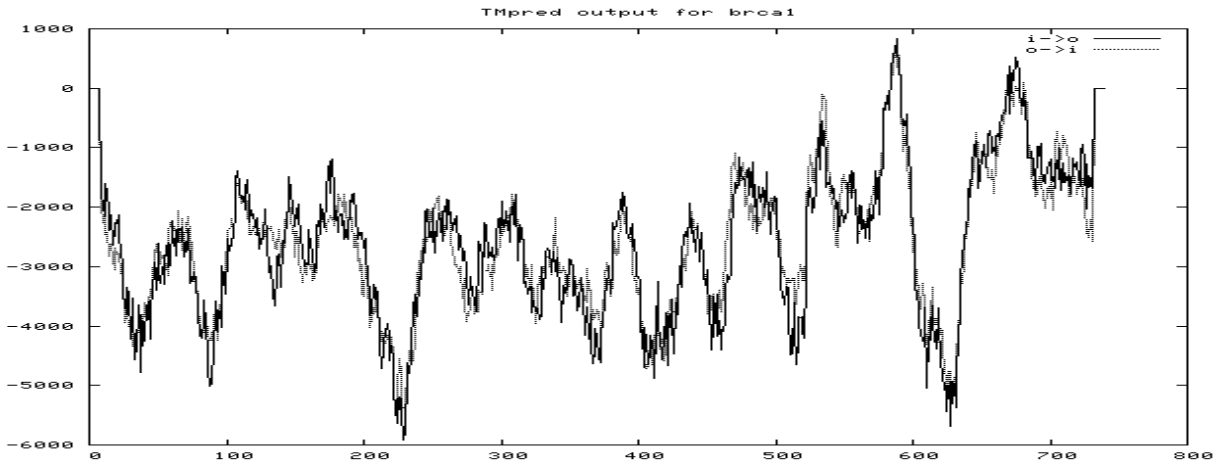
Predicting RNA secondary structures is the first step in predicting RNA tertiary structures from RNA sequences. Current RNA structure was predicted by calculating the free energy (Table and 7 Fig 2). Designing RNAs that form specific secondary structures is enabling better understanding and control of genes regulating breast cancer through RNA-guided silencing, genome editing and protein organization. The folding of RNAs into specific secondary structures is necessary for performing these functions, and numerous methods have been developed to model this folding process.Free energy predicted in human BRCA1 5' untranslated region (UTR) variants ranges from-46.8 to -812.7 for BRCA1 proteins.(Table 7). Experimental studies show that a huge fraction of the human genome is transcribed [26], and computational studies show evidence that thousands of structurally conserved RNAs can be found in the human genome. There exist methods to fold a single RNA sequence either by maximizing base pairing interactions, or by minimizing the free energy of the structure.

3.9 TRANSMEMBRANE PREDICTION ANALYSIS:

Sl. No.	Accession no.	Possible transmembrane helices		Table of correspondences	
		Inside to outside helices	**Outside to inside helices**	**inside >outside**	**outside->inside**
1.	AAI06747.1	0 found	0 found	0 found	0 found
2.	AKJ84699.1	0 found	0 found	0 found	0 found
3.	ALO20345.1	2 found from score to centre 581 (581) 597 (597 831 589 663 (667) 687 (687) 530 676	2 found from score to centre 580 (580) 600 (598) 801 590 672 (672) 692 (690) 111 682	2 found from score to centre 581- 597 (17) 831|663- 687 (25) 530 ++	2 found from score to centre 580- 600 (21) 801 |(672- 692 (21) 111)

4.	AAB08105.1	2 found from score to centre 74 (74) 90 (90) 831 82 156 (160) 180 (180) 530 169	2 found from score to centre 73 (73) 93 (91) 801 83 165 (165) 185 (183) 111 175	74- 90 (17) 831 |156- 180 (25) 530 ++	73- 93 (21) 801|(165- 185 (21) 111)
5.	AKG51647.1	0 found	0 found	0 found	0 found
6.	AAC00049.1	2 found from score to centre 597 (597) 613 (613) 831 605 679 (683) 703 (703) 530692	2 found from score to centre 596 (596) 616 (614) 801 606 688 (688) 708 (706) 111698	597- 613 (17) 831|679- 703 (25) 530 ++	596- 616 (21) 801(688- 708 (21) 111)
7.	ALO20343.1	0 found	0 found	0 found	0 found
8.	ABB87061.1	0 found	0 found	0 found	0 found

Table No 8: Transmembrane region scoring showing helices.



Graph No.1: TMPred Graph of BRCA1 (ALO20345.1)

The transmembrane regions were predicted using TMPred software (Graph no:1 and Table 8) which predicts the transmembrane region and their helical membrane- spanning domains. Possible transmembrane helices, for the accession number ALO20345.1, revealed the presence of 2 helices from inside to outside and for the sequence AAB08105.1, outside to inside 2 helices was found. Transmembrane topology suggestions are purely speculative and should be used with extreme caution since they are based on the assumption that all transmembrane helices have been found. The algorithm is based on the statistical analysis of TM base, a database of naturally occuring transmembrane proteins. The prediction is made using a combination of several weight-matrices for scoring. Some forms of cancer, for instance, are associated with integral membrane protein–protein interactions, which lead to aberrant downstream signal transductions important for cell growth regulation [27].

3.10 DOMAIN ARCHITECTURE ANALYSIS USING SMART:

		LOW COMPLEXITY			BRCT			RING			COILED COIL		
Sl. NO.	ACCESSION NO.	START	END	E-Value	START	END	E-Value	START	END	E-Value	START	END	E-Value
1.	AAI06747.1	180	191	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2.	AKJ84699.1	44	51	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
3.	ALO20345.1	468	483	N/A	540	622	1.82e-7	24	64	1.82e-7	289	321	N/A
4.	AAB08105.1	N/A	N/A	N/A	14	96	8.38e-7	N/A	N/A	N/A	N/A	N/A	N/A
					128	215	8.56e-7						
5.	AKG51647.1	N/A	N/A	N/A	N/A	N/A	N/A	1	37	0.162	N/A	N/A	N/A
6	AAC00049.1	468	483	N/A	N/A	N/A	N/A	24	64	1.82e-7	289	321	N/A
7.	ALO20343.1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
8.	ABB87061.1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table No:9 SMART analysis of BRCA1 coded proteins.

A Web-based tool (SMART) has been designed that makes use of mainly public domain information to allow easy and rapid annotation of signaling multidomain proteins. Many proteins are multidomain in character and possess multiple functions that often are performed by one or more component domains. The tool contains several unique aspects, including automatic detection of repeated motifs or domains, and a protocol for combining domain predictions from homologous subfamilies. The ability of SMART to annotate single sequences or large datasets is exemplified by the cases described in BRCA1 genes. Expect value (E) a parameter that counts the number of hits one can "expect" to see by chance for a database of a particular size. It decreases exponentially as the Score (S) of the match increases. Here it is in the expected range 0.62 to 8.56(Table 9). Most of the proteins do not show presence of any domain region, indicating the absence of conserved sequences.

3.11 RAMACHANDRAN PLOT ANALYSIS:

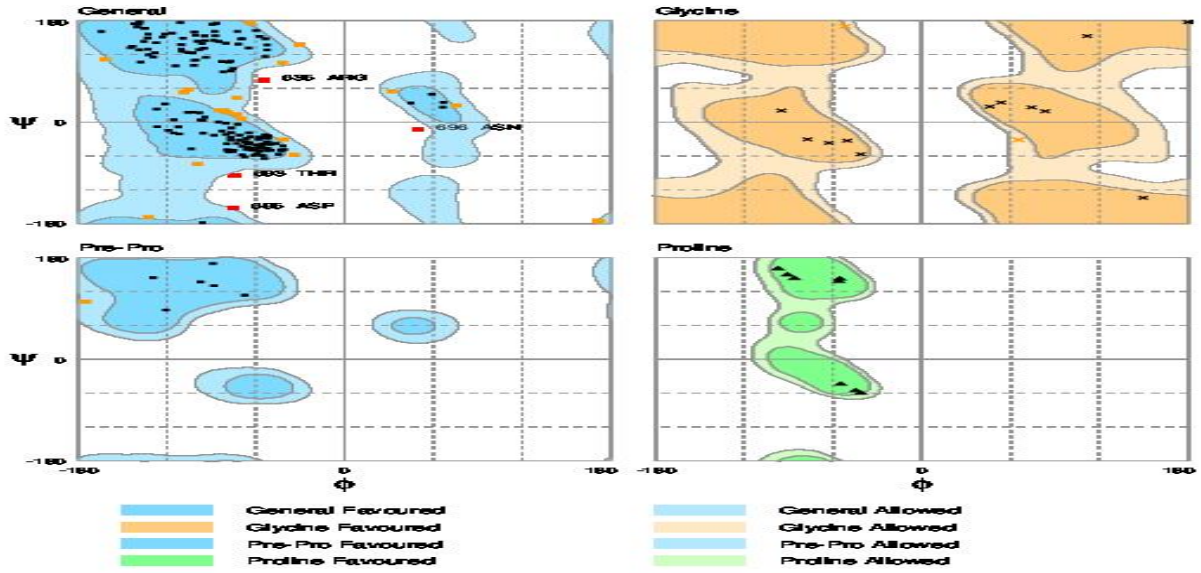


Figure 3: Ramachandran plot of BRCA1 obtained from RAMPAGE (ALO20345.1)

Evaluation of residues

Residue [558 :THR] (-132.70,-168.41) in Allowed region

Residue [572 :PHE] (75.30, 29.65) in Allowed region

Residue [606 :LEU] (-63.81, 175.41) in Allowed region

Residue [632 :SER] (-72.83, 43.16) in Allowed region

Residue [634 :ASP] (-99.51, -74.08) in Allowed region

Residue [637 :ILE] (-41.50, -30.86) in Allowed region

Residue [639 :ARG] (-30.71, 137.00) in Allowed region

Residue [647 :GLY] (-51.52, 169.92) in Allowed region

Residue [652 :MET] (-175.04, 102.08) in Allowed region

Residue [674 :SER] (-81.34, 20.91) in Allowed region

Residue [676 :THR] (-105.90, 57.80) in Allowed region

Residue [680 :GLY] (65.41, -31.20) in Allowed region

Residue [690 :ASP] (-79.28, 17.62) in Allowed region

Residue [698 :PHE] (-72.83, 13.95) in Allowed region

Residue [700 :ALA] (-84.56, 20.80) in Allowed region

Residue [703 :GLN] (-69.72, 6.47) in Allowed region

Residue [705 :CYS] (172.19,-175.37) in Allowed region

Residue [706 :GLU] (-108.65, 53.05) in Allowed region

Residue [722 :TYR] (32.01, 54.62) in Allowed region

Residue [723 :GLN] (-161.71, 111.35) in Allowed region

Residue [724 :CYS] (-42.03, 104.89) in Allowed region

Residue [728 :ASP] (-33.99, -57.84) in Allowed region

Residue [635 :ARG] (-54.18, 74.60) in Outlier region

Residue [693 :THR] (-74.18, -94.75) in Outlier region

Residue [695 :ASP] (-75.33,-152.37) in Outlier region

Residue [696 :ASN] (49.73, -12.97) in Outlier region

Number of residues in favoured region (~98.0% expected): 183 (87.6%)

Number of residues in allowed region (~2.0% expected): 22 (10.5%)

Number of residues in outlier region: 4 (1.9%)

Ramachandran plot of a protein reveals almost exclusively alpha helix and least beta sheets. Ramachandran plot displays the main chain torsion angles phi (φ), psi (ψ); (Ramachandran angles) in a protein of known structure. Each residue is classified according to its regioncore "allowed", "generous" or disallowed. Residues in the generous and disallowed regions are highlighted on the plot. A log-odds score shows how normal or unusual the residues location is on the Ramachandran plot for the given residue type. Glycine is represented by x and other residues are represented by triangle and squares. Results gave us the value of 98% residues in most favoured regions in R-Plot which suggests that they predict Probiotic model of good quality (Fig 4). These structures are allowed as biomarkers and provide a good foundation for finding of new BRCA1 proteins.

3.12 BARCODE

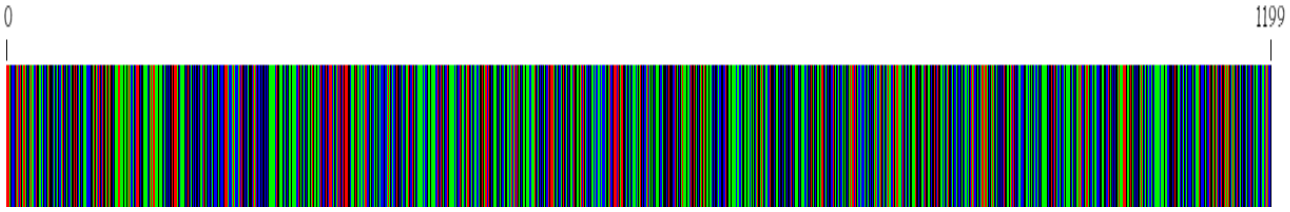


Figure 4: Barcode for reverse translated DNA sequences from the proteins encoded by BRCA1 (ALO20345.1)

DNA Barcode for the BRCA 1 sequence is represented in the figure 4. The DNA barcoding is a molecular technology for rapid identification of animals and plants in species level. It uses sequence comparisons of a standard fragment of DNA sequence ("DNA barcode") to distinguish species. Synthetic DNA sequences attached to the anticancer drugs in advance serve as barcode readers of each drug's activity in the cancer cells. Once the barcode sequence has been obtained, it is placed in the Barcode of Life Data Systems (BOLD) database - a reference library of DNA barcodes that can be used to assign identities to unknown specimens. The DNA barcoding technology uses a short standard piece of DNA sequence for species identification and has gained wide acceptance as a standard and effective method for biodiversity research, conservation genetics, wildlife forensics, and so on.

3.13 QR CODE

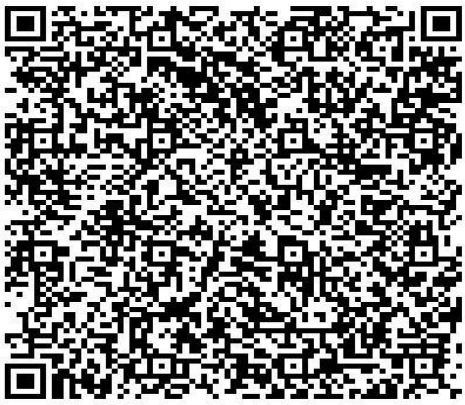


Figure 5: QR code for protein sequences retrieved from NCBI Database (ALO20345.1).

The QR Code for the BRCA1 sequence is represented in the figure5. QR Code Quick Response (QR) Code was the earliest 2D barcode. The codes carry meaningful information in the vertical direction as well as the horizontal. Therefore, the codes can carry up to several hundred times the amount of data carried by ordinary barcodes (which storing a maximum of 20 digit). QR Code consists of black modules arranged in a square pattern on white background. Three big square marks on the three corners of the code are required for positioning the reader, while several smaller marks arranged in some places in the code are for aligning the pattern. Among 2D barcodes, QR Code has the largest capacity to carry 7,089 numeric, 4,296 alphanumeric characters, and 2,953 bytes of binary (8 bits) data. The QR Code has the best compression efficiency in encoding DNA barcode sequences among the other 2D Codes. They are superior to barcodes due to the large amount of programmed information (more than 7000 alphanumeric characters), and the inclusion of error correction patterns. In addition, location markers facilitate scanning and automatic analysis of QR codes.

4. CONCLUSION

Breast cancer is threatening the lives of millions of women in the globe. Researchers were able to identify the major causative protein BRCA1 that can be targeted to control the rate of cell division in the victims. Deleterious mutations in BRCA1 account for a considerable proportion of dominantly inherited breast cancer and have received wide acceptance in diagnostic testing and prevention.

Investigations aimed at deciphering the molecular events that underpin the initiation and progression of disease are primarily targeted towards the profiling of proteins, whose aberrant expression, contributes to alterations in cellular function and ultimately lead to disease. By focusing on the mechanisms of disease, we aimed to identify critical molecular events that can be targeted with novel therapeutic strategies. Our research focused on the proteomic characterization of BRCA1 gene and its role in cancer progression using insilico method. Data analysis methods often rely on the analysis of high-throughput sequence data and they provide understanding of the

relationship between various molecular characteristics of cells. Detailed knowledge of BRCA1 and their properties were revealed through biological and biochemical properties. Eight BRCA proteins have been chosen mainly to study their physico-chemical properties, primary and secondary structures by using computational tools and servers. Primary structure analysis reveals that most of the proteins under study are hydrophobic in nature and contain disulphide linkages. Physico-chemical characterization studies give a good idea about the properties such as pI, EC, AI, GRAVY and stability that are essential and vital in providing data about the proteins and their properties. Secondary structure analysis predicts that most of them contain only α-helices and remaining of them contain mixed structure. The investigation here will pave the path for experimental studies and help build testable hypotheses. The three-dimensional (3D) protein structures provide valuable insights into the molecular basis of protein function, allowing an effective design of experiments. Homology models of proteins are of great interest for planning and analyzing biological experiments when no experimental three dimensional structures are available. Nowadays, 3D structure of protein can be predicted from amino acid sequences by different web based homology modelling servers at different level of complexity. During evolution, the structure is more stable and changes much slower than the associated sequence, so that similar sequences adopt practically identical structures and distantly related sequences still fold into similar structures. The modelling of 3D structure of protein was performed by Swiss model.

Overall, our study contributes to cancer research by providing detailed computational analysis of proteins encoded by BRCA1 genes, which when mutated may alter the protein structure and/or function. Identifying deleterious missense mutations in oncogenes, tumor suppressor genes, or in the genes coding for proteins functioning in cell division and mitosis, is essential in cancer research. All the results of bioinformatics analysis of BRCA will provide the basis for the further study on the progression of diseases. The results obtained will facilitate wet-lab biologists to formulate new experiments.

ACKNOWLEDGEMENT

We wish to express our gratitude to Prof. Usha Anandhi, Co-ordinator, Centre for Applied Genetics, Jnana Bharathi, Bangalore University for her support.

REFERENCES

1. Ambesi-Impiombato A, di Bernardo D. Computational Biology and Drug Discovery: From Single-Target to Network Drugs. Current Bioinformatics. 2006; 1:3-13.
2. Prashant VT, Uddhav SC, Madura SM, Vishal PD, Renuka RK. Secondary Structure Prediction and Phylogenetic Analysis of Salt Tolerant Proteins. Global J. Mol. Sci. 2010; 5 (1): 30-36.
3. Madhu S, Mahesh P. Sequence analysis of Semaphorin in Tumor Progression. Asian Academic Research Journal of Multidisciplinary. 2015; 1: 407-423.
4. Asokan KV and Pillai MM. In silico Characterization of Silk Fibroin Protein using

Computational
Tools and Servers. Asian. J. Exp. Sci. 2008; 22(3) 265-274.
5. Munduganore DS, Mundaganore YD, Ashokan KV. Sequence analysis of protein in peste des petits
ruminants: An in silico approach. International Journal of Food, Agriculture and Veterinary Sciences. 2012; 2 (3):26-30.
6. Swati G, Mayuri B, Shaikh SA. In silico characterization and comparative analysis of PTEN in different species involved in NSCLC. Trends in Life Sciences. 2013; 2(1):2631.
7. Herman JG, Jen J, Merlo A, Baylin, SB. Hypermethylation-associated inactivation indicates a tumor suppressor role for p15INK4B. Cancer Res.15. 1996; 56(4):722–727.
8. Siegel RL, Ward EM, Jemal A. Trends in colorectal cancer incidence rates in the United States by tumor location and stage, 1992-2008. Cancer Epidemiol Biomarkers Prev. 2012; 21:411-416.
9. Miki Y, Swensen J, Shattuck Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, Ding W. A strong candidate for the breast and ovarian cancer susceptibility gene brca1.Science.1994; 266, 66–71.
10. Ruffner H and Verma I M. BRCA1 is a cell cycle-regulate nuclear phosphoprotein. Proc Natl Acad Sci USA. 1997; 94:7138–7143.
11. Friedenson, B. The BRCA1/2 pathway prevents hematologic cancers in addition to breast and ovarian cancers. BMC Cancer. 2007; 7:152.
12. Wang SH, Simcox A, Campbell G. Dual role for Drosophila epidermal growth factor receptor signaling in early wing disc development. Genes Dev. 2000;14(18): 2271-2276.
13. Thakur S and Phadke SR. Familial breast cancer: genetics and counselling. Indian J Surg. 2005; 67:297–301.
14. Thompson D, Easton D. Variation in BRCA1 cancer risks by mutation position. Cancer Epidemiol Biomarkers Prev. 2002; 11:329-336.
15. Couch FJ, Weber BL. Mutations and polymorphisms in the familial early-onset breast cancer (BRCA1) gene. Breast Cancer Information Core. Hum Mutat.1996; 8(1):8-18.
16. Mahesh P, Divya P, Akshatha M, Prathima R and Lava Kumar C. Insilico characterization and phylogenetic analysis of novel probiotic bacteria in honey bees. International Journal of Asian Academic Research of Multidisciplinary. 2015;1(32):337-357.
17. Gill SC and Von Hippel PH. Calculation of protein extinction coefficients from amino acid sequence data. Anal. Biochem. 1989;182: 319-326.
18. Bachmair A, Finley D, Varshavsky A. In vivo half-life of a protein is a function of its amino-terminal residue. Science. 1986; 234:179-186.
19. Guruprasad K, Reddy BVP, Pandit MW. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its

primary sequence. Prot. Eng. 1990; 4(2):155-164.

20. Ikai AJ. Thermostability and aliphatic index of globular proteins. J of Biochem. 1980;88(6):1895-1898.

21. Kyte J, Doolottle RF. A simple method for displaying the hydropathic character of a protein. J of Mol Bio. 1982;157(1):105 -132.

22. Asokan KV, Mundaganur DS, Mundaganur YD Catalase. Phylogenetic Characterization to Explore Protein Cluster. J. Res. Bioinformatics. 2010; 1:001-008.

23. Anusha PK, Mahesh P, Mandeep S, Reddy MS. Computational Annotation of GSK-3 Protein in Homo Sapiens. IJESC. 2016; 6:4720-4733.

24. Zhang Z, Berry EA, Huang LS, Kim SH. Mitochondrial cytochrome bc1 complex. SubCellular Biochemistry. 2000;35, 541–580.

25. ZIkai A. Thermo stability and aliphatic index of globular proteins. J. Biochem.1980; 88:1895-1898.

26.Cheng J. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. Science, vol.
308. 2005; 1149-1154.

27. Dimaio D and Mattoon D. Mechanisms of cell transformation by papillomavirus E5 proteins. Oncogene 20. 2001; 7866–7873.