



STRUCTURE BASED DRUG DESIGNING ON BREAST CANCER GENE (BRCA 1)

Rajeswara Reddy. Erva*

Vijaya Kumar S**

Abstract: *Breast Cancer is the world's second death causing disease. BRCA1 is the gene responsible for causing this Breast Cancer. Actual function of BRCA1 is to control the cell division at optimal levels. If there is any abnormality in the gene then the division of breast cells will be altered and as a result the cancerous growth of cells will occur in breast. BRCA1 codes for a protein of 1863 amino acids, producing a protein of about 220 kd. The encoded protein combines with other tumor suppressors, DNA damage sensors, and signal transducers to form a large multi-subunit protein complex known as BASC for BRCA1-associated genome surveillance complex. This gene product associates with RNA polymerase II, and through the C-terminal domain, also interacts with histone deacetylase complex. Mutations in this gene are responsible for approximately 40% of inherited breast cancers and more than 80% of inherited breast and ovarian cancers. Even though there are some drugs to cure this disease, this is an attempt to design a new drug, which is more specific for the gene product causing the Breast Cancer.*

Keywords: *Breast Cancer, BRCA1, BASC, RNA polymerase II , C-terminal domain, histone deacetylase complex, Mutations.*

*Department of Biotechnology, Jawaharlal Nehru Technological University Anantapur
College of Engineering Pulivendula, Pulivendula, Andhra Pradesh, INDIA.

**Department of Biotechnology, Madanapalle Institute of Technology & Science,
Madanapalle, Andhra Pradesh, INDIA.



INTRODUCTION

“Cancer is a disease in which abnormal malignant body cells uncontrollably grow and multiply”. “Breast Cancer happens when cells in the breast begin to grow out of control enabling them to invade nearby tissues or even spread through the body”. There are two main places in the breast where cancer can occur: the lobules (the milk-producing tissue) and the ducts (which carry the milk to the nipple). Ductal carcinoma in situ, Invasive ductal carcinoma, Lobular carcinoma in situ, Invasive lobular carcinoma. To date, most inherited cases of breast cancer have been associated with two genes: BRCA1, which stands for BREast CAncer gene one, and BRCA2, or BREast CAncer gene two. The function of these genes is to keep breast cells growing normally and to prevent any cancer cell growth. But when these genes contain abnormalities, or mutations, they are associated with an increased breast cancer risk. Abnormal BRCA1 and BRCA2 genes may account for up to 10% of all breast cancers. The interactions that have been described for BRCA1 includes BRCA1 has been shown to be part of a genome surveillance complex termed BASC, which contains a number of proteins that are important for efficient DNA repair (the MRE11/RAD50/NBS1 repair complex, proteins involved in mismatch repair such as the MSH2/MSH6 and PMS2/MLH1 heterodimers, the DNA-damage-signalling protein kinase ATM, BLM helicase and replication factor-C) [1]. Additionally, BRCA1 associates with RNA polymerase II holoenzyme, the cell-cycle-checkpoint kinase ATR, a SWI/SNF chromatin-remodelling complex, the Fanconi's anaemia protein FANCD2, and a number of other important proteins. [2 & 3].

Another protein that is known to interact with the BRCA1/BACH1 complex is BARD1 [4]. Both BARD1 and BRCA1 contain a conserved RING finger motif, which is often linked to either DNA binding or ubiquitination [5]. The RING finger motif of BRCA1 has been shown to possess ubiquitin ligase activity, which is greatly stimulated by binding to BARD1 [6]. Significantly, a breast-cancer-associated *BRCA1* mutation within the RING finger motif abolished the ubiquitin ligase activity of BRCA1, indicating that this activity is important for the action of BRCA1 in tumour suppression [7]. The BRCA1 gene is located on chromosome 17 at the locus 17q21 and encodes a protein of 1863 amino acids. The BRCA 1 locus spans >100 kb comprising 24 exons [8]. Expression of wild-type BRCA 1 inhibits colony function and tumor growth in vivo, whereas tumor derived mutations of BRCA 1 abolish this growth



suppression. Germline mutations in BRCA 1 appear to account for 50% of familial breast cancers and essentially all families with 17q21-linked inherited susceptibility to ovarian and breast cancer [9, 10, 11, 12]. Kindreds segregating constitutive BRCA1 mutations show a lifetime risk of 40-50% for ovarian cancer and > 80% for breast cancer [13, 11]. The classification of BRCA 1 as a highly penetrant, autosomal dominant tumor suppressor gene has been genetically confirmed by the finding of frequent loss or mutation (LOH) of the wild-type allele in breast tumors from mutation carriers [10,8,14]. Surprisingly, BRCA 1 mutations in sporadic breast cancer including those which show LOH have yet to be found and are extremely rare in sporadic ovarian cancer [15,16]. The carboxyl-terminal region of BRCA1 contains two BRCT motifs, which are important for protein/protein interactions. Interestingly, many cancer-associated mutations in BRCA1 result in the formation of a truncated BRCA1 protein, in which the carboxyl-terminal region has been deleted. It is therefore possible that proteins that interact with the BRCT motifs of BRCA1 may play important roles in BRCA 1-dependent tumor suppression [17].

BRCA 1 is a transcription factor expressed in a variety of cell types. Loss of expression of BRCA 1 leads to the development of breast and ovarian cancer. Recent investigations by my group indicated that BRCA1 is induced by estrogen. This steroid along with other sex hormones such as progesterone, regulate growth and the development of mammary epithelial cells. Our findings [18], which have been recently cited in review articles indicate that exposure to environmental xenobiotics can reduce significantly cell viability and the expression of BRCA 1. These observations are important because they provide new evidence that exposure to environmental polycyclic aromatic hydrocarbons (PAHs) can be a risk factor in the onset of breast and ovarian cancer [19].

METHODOLOGY:

Sequence retrieval alignment and Protein structure prediction:

Various **genomic & proteomic online tools** were used to evaluate the BRCA 1. The amino acid sequence of BRCA 1 was retrieved from NCBI (NP 009234). The BRCA1 protein was modeled by using the **SPDBV** (Swiss PDB Viewer) software. The templates were taken by running **BLAST** tool against BRCA 1. The constructed model was further validated by Ramachandran plot. Q-Site Finder [20] server was used to predict the structure based protein functional site. The modeled protein is then docked against the ligand database by **ARGUSLAB**. The Protein



Ligands interactions were performed using ArgusLab. Virtual Screening method was used to find the best ligand interaction and we can consider the lowest free energy molecules, which shows high affinity to the protein. The Physico-chemical properties of the Ligand were performed using HyperChem, ChemSketch and CACHE.

RESULTS AND DISCUSSION:

Genomic Tools:

BioEdit:

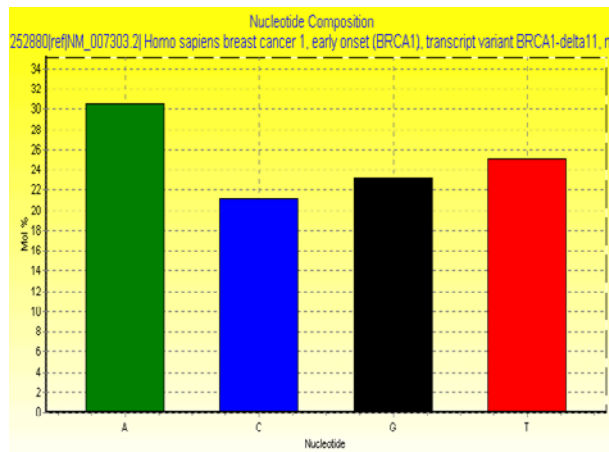


Figure 1 : Nucleotide composition of asnA gene

Nucleotide composition of asnA has been predicted by Bioedit & the results reveals that the sequence is purine rich with the molecular weight of more than 1143571.00 Daltons (single stranded). The CG & AT content was found to be **44.36%** and **55.64%** respectively.

Genscan:

Predicted genes/exons:

Gn.Ex Type S .Begin ...End .Len Fr Ph I/Ac Do/T CodRg P.... Tscr..

Gn.Ex	Type	S	.Begin	...	End	.Len	Fr	Ph	I/Ac	Do/T	CodRg	P....	Tscr..
1.01	Sngl+		201	2366	2166	2	0	59	42	1219	0.995	108.50	
1.02	PlyA+		3724	3729	6							1.05	

GenScan predicts the presence of the exons and the poly A tail region in the direct strand from the position 201 to 2366 & 3724 to 3729 respectively. The GC content of the sequence is given as 44.36%. The probability of the predicted output for the exon is 0.99. The predicted first exon is categorized as a strong exon as the exon score is greater than 100 i.e.,



108.50. But the second exon is categorized as a weak exon because the exon score is less than 50 i.e., 1.05.

Proteomic Tools:

PROT PARAM:

The amino acid composition is self-explanatory. The molecular weight of the protein is given as 80590.3 and the isoelectric point as 4.99. The protein was acidic in nature, as it has more negatively charged residues than the positively charged ones. The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro), >20 hours (yeast, in vivo) and >10 hours (Escherichia coli, in vivo). The protein is classified as unstable, as the instability index is computed to be 54.54. The aliphatic index predicts the volume occupied by the aliphatic residue side chains and the index is 76.37. The protein is highly hydrophilic as the Grand Average of Hydropathicity value is -0.552 , which is very much lesser than 0.05.

NetPhos 2.0:

Phosphorylation sites predicted: Ser: 42 Thr: 10 Tyr: 7

GOR 4:

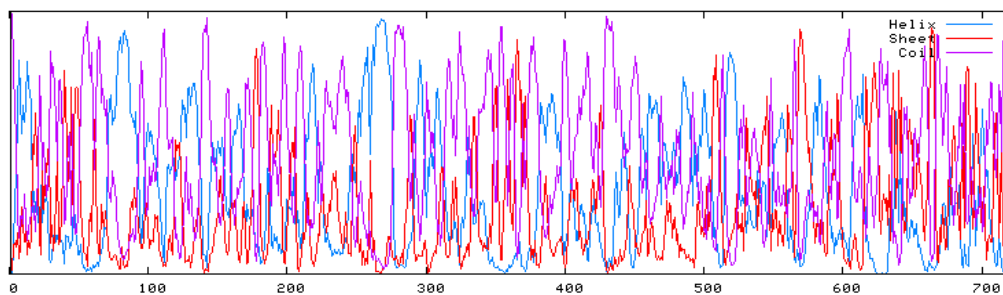


Figure 2 : Secondary structure of peptide encode by asnA gene by GOR 4

The secondary structure prediction of BRCA1 was predicted using GOR (Garnier Method). It was found that 32.45% of amino acids fall on Alpha helix region, 17.84% of amino acid was found to be lays in beat sheet and remaining 50.21% tends to form random coil.

SPDBV (SWISS pdb Viewer):

SwissPDB Viewer is used to do the homology modeling and to carry out the active site prediction for the modeled protein. The query protein is modeled using the two Swiss PDB template namely, 1t29A and 1y98A. The RMSD value is 0.99. This modeled protein when checked for the Ramachandran plot the amino acid residue Tyr1845 is present in the disallowed region. So the residues from 1843 to 1847 are selected for loop building. Finally



the modeled protein is subjected to energy minimization and thus again checked for the presence of any amino acid residues in the disallowed regions.

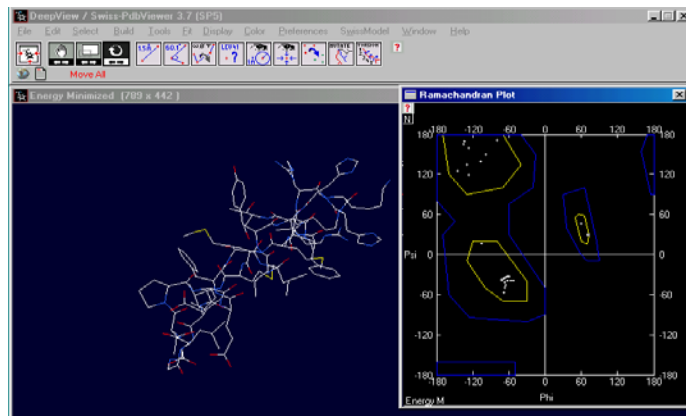


Figure 3: Modeled Protein by SPDBV

Active site prediction (Q-Site Finder):

Ligand Binding Site Prediction:

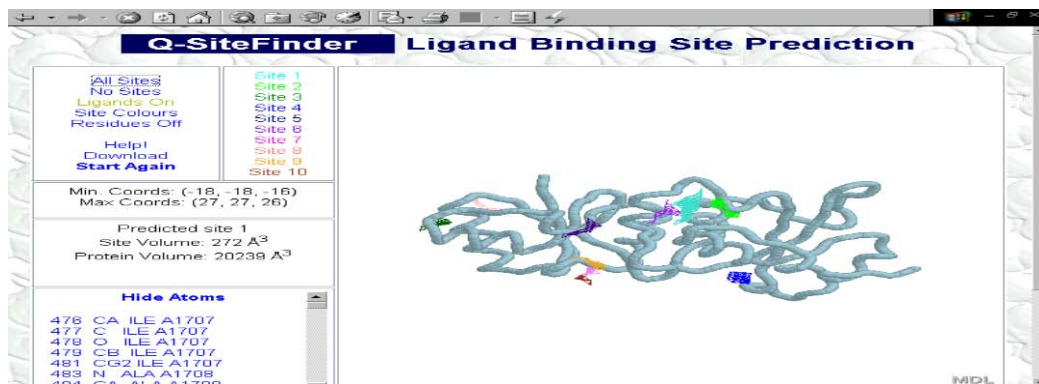


Figure 4: Active sites predicted in Modeled Protein by Q site finder

Q site finder is a tool used for the active site prediction. For the query totally ten potential active sites were predicted.



Protein Ligand Interaction:

Table 1: Protein Ligand Interaction

ACTIVE SITE	FREE ENERGY	AMINO ACID	LIGAND	GRID SIZE (Maximum)		
				X	Y	Z
1	-13.772881	LEUSINE 1786	41	14	6	4
2	-14.304344	CYSTEIN 1847	13	16	6	4
3	-15.190476	HISTIDINE 1672	18	27	22	26
4	-15.218418	ARGININE 1835	15	27	22	26
5	-13.395705	PHENYLE ALANINE 1695	17	27	22	26
6	-12.384316	TYROSINE 1845	13	27	22	26
7	-15.429602	ARGININE 1699	17	27	22	26
8	-15.530199	TYROSINE 1666	17	27	27	26
9	-12.730552	PHENILE ALANINE 1695	17	27	27	26
10	-14.604001	THREONINE 1691	17	27	27	26

Most of the amino acids were docked with 17th ligand.

Ligands Interaction:

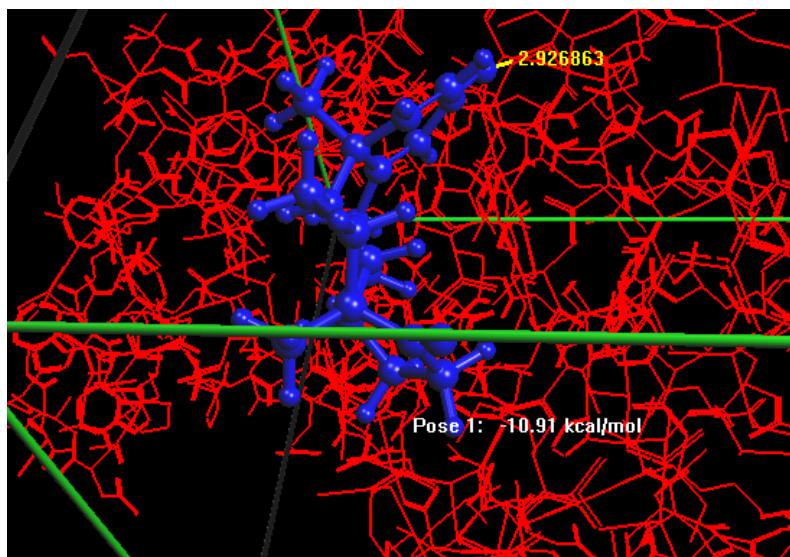


Figure 5: Protein-Ligand Interactions

Best Ligand Pose: energy = -10.9092 kcal/mol

The Protein Ligands interactions were performed using ArgusLab. In our study we created a database of ligands contains both Aromasin and Arimidex (41 molecules), and found the high affinity ligand to the protein. Virtual Screening method was used to find the best ligand interaction, out of 41 molecules only 35 ligand shows best interaction to all the ten potential



active sites. We consider the lowest free energy molecules, which shows high affinity to the protein. We have found the interaction of ligand and protein, which shows one hydrogen bond interaction (2.9268763Å) the interaction is between the nitrogen atom of leucine and the oxygen atom of the ligand, and the ligand was found to be 10,13-dimethyl-7,8,9,11,12,14,15,16-octahydro-6H-cyclopenta[a]phenanthrene-3,17-dione (IUPAC Name). The Physico-chemical properties of the Ligand were performed using HyperChem, ChemSketch and CAChe.

MODELLING USING HYPERCHEM:

IUPAC Name:

10,13-dimethyl-7,8,9,11,12,14,15,16-octahydro-6H-cyclopenta[a]phenanthrene-3,17-dione

MW: 284.393 g/mol

Molecular

Formula: C₁₉H₂₄O₂

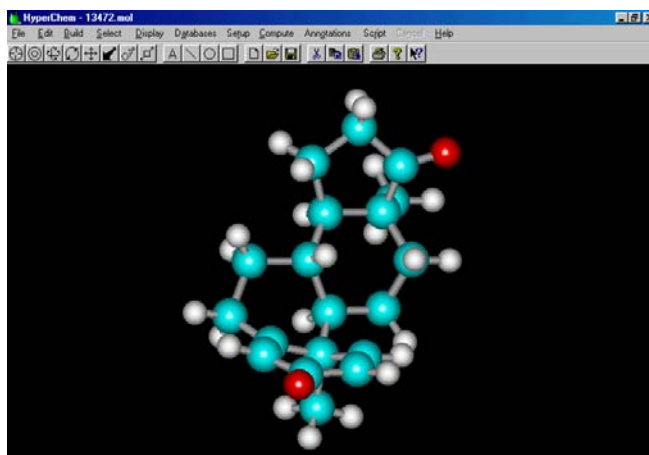


Figure 6: Protein-Ligand Interactions



Single Point (Molecular Mechanics)	Geometry center....	: -0.0361
Total Energy =110.049217	0.3990 0.1554	
Gradient =19.579498.		
Bond =19.2337	Mass center.....	: -0.3160
Angle =4.74805	0.4399 -0.1230	
Dihedral =46.9461	Appx. dimensions...	: 11.8894
Vdw =39.2305	9.4331 8.2787	
Stretch-bend =0.254086	Surf. area (1.40)..	: 493.2 Å ²
Electrostatic =-0.363168.	(ds=12.5 Å)	
Geometry optimization (Molecular Mechanics) FletcherReeves optimizer	Polar area (PSA)...	: 85.3 Å ²
Energy=53.460434 Gradient=0.095513	(apolar=407.8 Å ²)	
Converged=YES (122 cycles 276 points).	Volume.....	: 280.1 Å ³
Bond =2.14943	(dv=8.1 Å)	
Angle =8.79727	Ovality.....	: Not
Dihedral =29.5842	available (use -g 0)	
Vdw =12.9415	logP (Crippen)....	: 4.9643
Stretch-bend =0.46325	Lipole (Crippen)...	: 0.0548
Electrostatic =-0.475272.	logP (Broto).....	: 1.8070
Ligand Properties:	Lipole (Broto)....	: 2.0997
Atoms.....	Virtual logP.....	: 1.9570
Heavy atoms.....		
Centroids.....		
Residues.....		
Molecules....		
Waters.....		
Formula.....		
C19H24O2		
Molecular weight...		: 284.393
Daltons		
Gyration radius....		: 3.2517 Å



QSAR properties :

Table 2: QSAR Properties of DRUG

PROPERTY	VALUE
Partial charges	0.00e
Surface area (approx)	108.07 Å ²
Surface area (grid)	3450.03 Å ²
Volume	577.86 Å ³
Hydration energy	-2.16kcal/mol
Log P	0.78
Refractivity	16.39Å ³
Polarizability	20.97 Å ³
Mass	260.21 amu

Among the entire QSAR properties log P plays a major role in predicting the type and the nature of the drug. As log P value is the partition co-efficient, ratio of the concentration of the compound between the immiscible solvent and the aqueous phase. As the log P value is 0.78, the ratio is 1:10 between the organic solvent and the aqueous phase and this drug can also be inject safely due to its less log P value.

CACHE:

UV Transition:

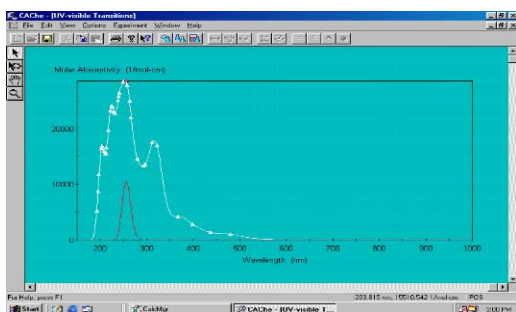


Figure 7: Graph showing UV Transitions

IR Transition:

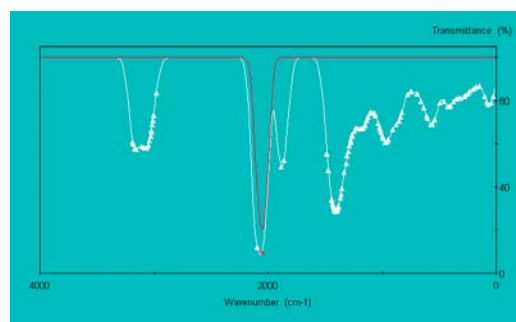


Figure 8: Graph showing IR Transition



The UV Visible and IR Transition for the molecule 10,13-dimethyl-7,8,9,11,12,14,15,16-octahydro-6H-cyclopenta[a]phenanthrene-3,17-dione has found to have a maximum uv-visible absorbance around 10000 /mol-cm at a wavelength of 260-270 nm. This has a minimum IR transmittance of about 10% at the wave number of about 1400/cm.

CONCLUSION:

The protein, which causes Breast Cancer, is an unstable protein. The gene, which codes for this protein is located on the 17th chromosome at a loci 17q21.31. As the signal peptide is absent the entire sequence is taken up for the protein modeling. The raw sequence is modeled using two appropriate templates. The RMSD value of the protein is 0.60. The ligands were taken from the most effective drugs Aromasin & Arimidex, the Aromatase inhibitors. Ten potential active sites are predicted, the residue TYR1666 of eighth active site was docked against the modeled ligand 10,13-dimethyl-7,8,9,11,12,14,15,16-octahydro-6H-cyclopenta[a]phenanthrene-3,17-dione with a very low free energy of -15.530199 kcal/mol. This docked molecule belongs to the class of Aromasin (exemestane) drugs. These will effectively prevent production of estrogen in adrenal glands. The results demonstrate that investigational compounds in which estrogen is blocked by chemical modification can display strong activity, low toxicity and excellent bioavailability, making them suitable for further exploration. Further the ADME\T property and activity has to be performed for all the ligands to get a potential drug for Breast Cancer.

REFERENCES:

- [1] Wang Y, Cortez. D, Yazdi. P, Neff. N, Elledge SJ, Qin J. (2000): BASC, A super complex of BRCA1 associated proteins involved in recognition & repair of aberrant DNA structures, Genes Dev14:927-939.
- [2] Kerr P, Ashworth A. (2001): New complexities for BRCA1 and BRCA2. Curr Biol 11:668-676.
- [3] Deng CX, Brodie SG, (2000): Roles of BRCA1 and its interacting proteins. Bioessays 22:728-737.
- [4] Cantor SB, Bell DW, Ganesan S, Kass EM, Drapkin R, Grossman S, Wahrer DCR, Sgroi DC, Lane WS, Haber DA, Livingston DM. (2001): BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. Cell 105:149-160.



- [5] Wu LC, Wang ZW, Tsan JT, Spillman MA, Phung A, Xu XL, Yang MC, Hwang LY, Bowcock AM, Baer R.(1996): Identification of a RING protein that can interact *in vivo* with the *BRCA1* gene product *Nat Genet* 14: 430-440.
- [6] Hashizume R, Fukuda M, Maeda I, Nishikawa H, Oyake D, Yabuki Y, Ogata F, Ohta T. (2001): The RING heterodimer BRCA1-BARD1 is a ubiquitin ligase inactivated by a breast cancer-derived mutation.*J Biol Chem* 276:14537-14540.
- [7] Ruffner H, Joazeiro CAP, Hemmati D, Hunter T, Verma IM. (2001): Cancer-predisposing mutations within the RING domain of BRCA1: loss of ubiquitin protein ligase activity and protection from radiation hypersensitivity.*Proc Natl Acad Sci USA* 98:5134-5139.
- [8] Miki et al. (1994): A strong candidate for breast and ovarian susceptibility gene BRCA 1, *Science*, 266: 66-71
- [9] Szabo et al. (1995): *Hum. Mol. Genet.*, 4 : 1811-1817.
- [10] Hall et al. (1990): *Science*, 250: 1684- 1689.
- [11] Easton et al, *Am. J.* (1995): *Hum. Genet.*, 56 : 265-27.
- [12] Narod et al.,*Am. J. Hu.* (1995): 67: 156-165.
- [13] Easton et al, *Am. J.* (1993):Genetic linkage analysis in familial breast and ovarian cancer:results from 214 families. The Breast Cancer Linkage Consortium, *Hum. Genet.*, 52 : 678-701.
- [14] Smith et al. (1992): Allele losses in the region 17q12-21 in familial breast and ovarian cancer involve the wild type chromosome,*Nature Genetics*, 2 : 128-131.
- [15] Futreal et al. (1994): *Science*, 266: 120- 122
- [16] Merajver et al. (1995): *Nature Genetics*, 9 : 439-443
- [17] Zhang XD, Morera S, Bates PA, Whitehead PC, Coffey AI, Hainbucher K, Nash RA, Sternberg MJE, Lindahl T, Freemont PS. (1998): Structure of an XRCC1 BRCT domain: a new protein–protein interaction module 17:6404-6411.
- [18] Jeffy BD, Chrinomas RB, Chen EJ, Gudas JM, Romagnalo DF.(1999): *Cancer research* 48: 114-26
- [19] Jeffy BD, Chrinomas RB, Chen EJ, Gudas JM, Romagnalo DF.(2005): *Cancer research* 72: 121-137
- [20] Laurie, A., Jackson, R. (2005):*Bioinformatics* 21: 1908–1916. [PMID:15701681].