Abstract
Breast Cancer is one of the major health threats of the decade that are troubling the mankind. Since long the studies have been conducted in treating this disease. Recent studies have revealed that the gene involved in causing this condition is BRCA, termed as Breast Cancer Associated gene. The current study involved the complete insilico analysis of this gene. Its functional study and the Structural Analysis. This is followed by identifying the potential drug candidate from PubChem database. Several screening techniques have been employed to check the feasibility and efficiency of the drug. Lipinsky rule was employed to check the drug likeness of the compound. Toxicity and ambiguity of the chemical are screened using Dundee ProDrg Server. The final 3D chemical obtained was used to dock with the protein structure using HEX automatic docking software. The best chemical that can be used in the therapy can be selected based upon the lowest energy values of docking.

Keywords:
Breast Cancer, BRCA1 gene, cell-cycle progression, cell-cycle check points, transcriptional pathways, apoptosis, protein-protein interactions, phosphorylation and carcinogenesis.

Introduction:
Bioinformatics is a type of interdisciplinary field which has all developed method and software tools for all biological data and to understand it. Bioinformatics is a combination of many fields of subjects to study and for the processing of the biological data.

Homology Modeling is also known as the comparative modeling of a protein which refers to the construction of the atomic resolution model of the target protein from the known or given amino acid sequences and an three dimensional structure of a related homologous protein. It has been shown that the structure of a protein are more conserved than the sequences of the protein of the homologous, but the sequences which falls below the 20% sequence identity can have very different structure.[1] They can produce high-quality structure models when both the target and the template are closely related which helps in the formation of a structural genomics consortium which helps in the representation of the structures for all the classes of the protein folds.

RATIONAL DRUG DESIGNING is also known as the Drug Design or the Rational Design. It is a process which helps in finding the new medication which is based on the knowledge of biological target. The small molecules design are involved which are complementary in shape and it is charge to the biological target. The traditional method of the discovery of a drug is based on the trail-and-error of the chemical substances on the cultured cells or the animals and the effects on the treatments, rational drug design begins with the hypothesis which is based on the modulation of a specific biological target which may have the therapeutic value.

BREAST CANCER is a cancer which develops from the breast tissue. Signs of the Breast Cancer are lump in the breast, a change in breast shape, dimpling of the skin, fluid coming from the nipple, or a red scaly patch of skin. In those with distant spread of the disease, there may be bone pain, swollen lymph nodes, shortness of breath, or yellow skin. The factors which develop the Breast Cancer include obesity, lack of physical exercise, drinking alcohol, hormone replacement therapy during menopause, ionizing radiation, early age at first menstruation, having children late or not at all, and older age.

BRCA mutation is a mutation in either of the BRCA1 and BRCA2 genes which are the tumor suppressor genes. Harmful mutations in these genes which may produce the hereditary breast-ovarian cancer syndrome in affected persons. The risk of breast cancer and the ovarian cancer is higher for women which has a high risk in BRCA1 mutation than the BRCA2 mutation. Having a high risk mutation does not confirms that it will developed any type of cancer in any one’s body or any type of cancer implies that it is caused due to the mutations , rather it can be caused due to some other factor.

BRCA1 are the human genes and it has a protein product respectively. It is responsible for the repairing of the DNA. The cross linking of the DNA are the important source of the chromosomal or the DNA damage.
ROLE OF BRCA1 GENE: It provides a instructions to make a protein which acts a tumor suppressor. Tumor suppressor proteins helps in preventing the cells from growing and in the too fast division of the cells or in the uncontrolled manner.

In the nucleus of a normal cells, the BRCA1 protein interacts with some other protein for the breaks of the DNA. These breaks are caused due to natural and medical radiation or any other environmental exposures, and they also occur at the time of genetic exchange of a chromosome in the preparation of the cell division. It plays a crucial role in maintaining the stability of the genetic information of a cell. BRCA1 protein helps in the regulation of the genes and plays a important role in embryonic development.

Materials And Methods
Genetic Study of BRCA1:
Databases like Gene cards and the general Google search has been employed to identify the major protein involved in the Breast cancer. Once the gene has been identified, further analysis was performed with the same. Gene cards database is highly specific for human genes and their functions. This data base has been employed to collect the information regarding the importance of the BRCA1 gene in the breast cancer disease.

Retrieval of Protein Sequence from NCBI:
The protein sequence of BRCA1 can be collected from NCBI database using the regular SRS Sequence Retrieval system using the Boolean operators for specifying the search. The Protein in its FASTA sequence can be collected from this database along with the other information including the important regions in the sequence, Signal peptide positions, number of domain etc can be obtained for this database.

SMART for Domain Analysis:
Though the information regarding the Domains has been obtained from NCBI a specialized tool named SMRT (Simple Modular Architectural Research Tool) can be used to exactly locate the domains there number and the function of each domain in the protein can be collected and analyzed. The database is directly connected to CDD database that is an archive of protein domains and their functions.

Physicochemical analysis of the Protein:
Protparam a special tool of Expasy proteomic server has been used to analyze the physiological and biochemical properties of the protein. The tool gives an insight to the properties like Ionization potential, Iso electric point, Absorption maxima, Half-life etc. Based on the properties of the protein as obtained from this tool experimental protocol can be developed for proteins qualitative assay, Purity determination etc.

HHPPRED for Secondary structural confirmation:
To identify the secondary conformation of each region in the proteins structure specified tools must be employed. The conformation like Loop, coil, Sheet, bend, turn etc would be identified by these tools. This information in turn is useful to check the stability and interactivity of these proteins. HHPPred is a homology based tool that performed internal comparison of the user entered sequence with the database sequence and the best matching pair is displayed showing the secondary structural conformation.

Structural Analysis of the protein and retrieval of the 3D structure
In order to perform any of the detailed study related to the development of a drug, one of the essential steps is to obtain the 3D structure of the protein. The 3D structure is used for further study related to the Docking and Interactions. The tool employed for this purpose is “PHYRE”. The tool is embedded with a BLAST like algorithm that would connect the user entered sequence to the PDB database. This intern compares the sequence with all the pdb sequences. The final result would contain the list of PDB id’s that share structural and sequence similarities with the user entered sequence. The final PDB ID sharing maximum identity and low e Value is selected for further analysis.

Rasmol:
Rasmol is a standalone visualization software that would enable the user to visualize the 3D structure of the protein. It would not be used for structural manipulation but only for the visualization of the detailed structure. The software enables the users to identify the specific secondary structural conformations in the structure like Helix, Loop, sheet can be identified and highlighted in the structure. The software also enables the users to show a specified amino acid, ligand, metal ion or any other adjuvant present in the protein structure. Being a command line program it requires correct commands to be executed in the command line. It is a freely available software thus can be downloaded and used by the user.

PubChem:
It is a product of NCBI database. Useful for collecting the information about the specified chemicals. It is an online archive containing the information of all the known chemicals their properties and their biological importance. The user can use these chemicals based on their function and download their structures for other analysis.
Screening of the Chemicals for Drug Development:

All the selected chemicals for Pubchem database are not used directly based on their function rather screened for several parameters like Drug likeliness, Ambiguity, Toxicity etc. The Drug likeliness of the chemical is identified by using Lipinski rule of five that proposes 4 parameters that has to be followed by chemicals to be used as drugs. Further these chemicals are screened for their toxicity. It uses a tool Dundee ProDrg server that would identify the Ambiguity as well as the toxicity among the tested chemicals. The final list of selected chemicals are used for docking.

Docking of the Chemicals in HEX software:

The standalone software Hex is the most commonly used automated software system that would enable the complete docking of the selected protein receptor with the ligand. The result is interpreted in the form of the docking energy. Lower the docking energy higher is the effectiveness of the drug. All the selected compounds are screened for their docking ability and the final list of chemicals selected.

Results and Discussion:

NCBI is used for Obtaining the Locus, Length, Structure, Class, Definition Accession Number, Version, DB source, Keywords, Source and the Organisms of a particular sequence. All these data can be used for further analysis.

The FASTA sequence as obtained from NCBI is shown below:

Fig1: FASTA sequence of BRCA1 in NCBI

BRC1 [Homo sapiens]
GenBank: AAC37934.1
GeneID: Iden1tial Protein

Protpaam for Physico chemical Analysis:

As per the results of Protpoaram the length of the sequence was found to be 1863 amino acids with the calculated molecular weight of 207720.8 whose isoelectric point is calculated to be 5.29 and extinction coefficient to be 103940. The protein was found to have 283 acidic amino acids and 218 basic ones. Thus the protein was an acidic protein. All these values are further used for designing the protocols involving protein purification and identification.

SMART tool for Domain identification:

To identify the function regions in the protein sequence domain analysis using SMART page was performed which revealed the presence of a total number of 4 domains namely RING, BRCT ASSOC, BRCT, BRCT. The BRCT domain is present in duplicate. The functions of each domain differ but the overall activity is to regulate the onset of Breast Cancer.

3D structure analysis using HHpred and RCSB PDB:

To identify the 3 dimensional structure of the protein an online tool named HHpred has been used. The results showed that the best PDB id that can depict the structure for this protein was found to be 4JLU The structure of this protein was intern obtained from RCSB Protein Data Bank which is an archive for all the known 3D structures of the proteins.

Structure of BRCA protein as shown by RASMOL software:

Fig 2 shows the 3D structure of BRCA protein as shown in RASMOL Software.
The above picture shows the 3D structure of BRCA protein in a ribbon model as shown in Rasmol software.

**Chemical Analysis:**
All the chemicals possessing the required function were selected from the Pubchem data base. The selected chemicals were screened for several properties including Lipinsky rule, ambiguity and toxicity. The final chemical selected based on all the screening strategies was Dibromdulcitol, with the best properties.

**Table 1: Molecular Properties of the chemical**

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular Weight</td>
<td>307.9658 g/mol</td>
</tr>
<tr>
<td>Molecular Formula</td>
<td>C_{12}H_{23}O_{5}</td>
</tr>
<tr>
<td>XLogP3</td>
<td>-6.3</td>
</tr>
<tr>
<td>Hydrogen Bond Donor Count</td>
<td>4</td>
</tr>
<tr>
<td>Hydrogen Bond Acceptor Count</td>
<td>4</td>
</tr>
<tr>
<td>Rotatable Bond Count</td>
<td>5</td>
</tr>
<tr>
<td>Exact Mass</td>
<td>307.96387 g/mol</td>
</tr>
<tr>
<td>Monotopic Mass</td>
<td>305.910233 g/mol</td>
</tr>
<tr>
<td>Topological Polar Surface Area</td>
<td>80.9 A^2</td>
</tr>
<tr>
<td>Heavy Atom Count</td>
<td>12</td>
</tr>
<tr>
<td>Formal Charge</td>
<td>0</td>
</tr>
<tr>
<td>Complexity</td>
<td>110</td>
</tr>
<tr>
<td>Isotope Atom Count</td>
<td>0</td>
</tr>
<tr>
<td>Defined Atom Stereocenter Count</td>
<td>4</td>
</tr>
<tr>
<td>Undefined Atom Stereocenter Count</td>
<td>0</td>
</tr>
<tr>
<td>Defined Bond Stereocenter Count</td>
<td>0</td>
</tr>
<tr>
<td>Undefined Bond Stereocenter Count</td>
<td>0</td>
</tr>
<tr>
<td>Covalently-Bonded Unit Count</td>
<td>1</td>
</tr>
</tbody>
</table>

**Hex Docking for Finalization of Drug:**
To identify the docking energy of the Dibromdulcitol with the receptor BRCA1 whose structure representative is 4JLU Hex software have been employed. The docking energy was found to be -180.79. This indicates a stable docking with acceptable docking energy. This indicates that the compound dibromdulcitol can be a better drug for targeting the BRCA receptor as a means to control the cell division.

**Conclusion:**
It is quite long since Breast cancer is threatening 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. The current work aimed to do the same. The work involves identification of a suitable chemical and its perfect screening so as to use if for docking with the BRCA1 receptor. The selected chemical was Dibromdulcitol, whose properties are mentioned above. The compound has been used to dock with the 3D structure of 4JLU a structure representative for BRCA1. The docking was performed using HEX 8.0 software. The docking results revealed that the above pair has the docking energy of -180.79. This indicates a perfect docking couple. Thus this compound can be used as a better drug in targeting the BRCA1 protein as a means to control cell division.

From the above Fig 3 it can be identified that the docking energy of the pair 4JLU with Dibromdulcitol was found to be -180.79. This indicates a good docking pair and the above chemical can be used for further therapy.
Bibliography:
1) “BRCA1 and BRCA2: Cancer Risk and Genetic Testing”, National Cancer institute

2) http://www.breastcancer.org/symptoms/testing/genetic/pos_results

3)” Double indemnity: p53, BRCA and cancer. p53 mutation partially rescues developmental arrest in Brca1 and Brca2 null mice, suggesting a role for familial breast cancer genes in DNA damage repair”

4)"Earlier age of onset of BRCA mutation-related cancers in subsequent generations", Jennifer K. Litton MD, 12 SEP 2011,
   DOI: 10.1002/cncr.26284


7) http://hex.loria.fr/dist/index.php