Insilico gene expression and drug docking studies on human Ovarian Cancer Disease Proteins (GPR68, DIRAS3 AND DPH1) using bioinformatics software and tools

*Shakila.L and Shalini.H

Post Grauate and Research Department of Zoology, Ethiraj College for Women, Chennai, India.

Abstract
Ovarian cancer usually has a relatively poor prognosis. It is disproportionately deadly because it lacks any clear early detection or screening test, meaning most cases are not diagnosed until they have reached advanced stages. However, in some cases, ovarian cancer recurrences are chronically treatable. Ovarian cancer metastasizes early in its development, often before it has been diagnosed. High-grade tumors metastasize more readily than low-grade tumors. Typically, tumor cells begin to metastasize by growing in the peritoneal cavity. More than 60% of women presenting with ovarian cancer have stage-III or stage-IV cancer, when it has already spread beyond the ovaries. Ovarian cancers shed cells into the naturally occurring fluid within the abdominal cavity. These cells can then implant on other abdominal (peritoneal) structures, including the uterus, urinary bladder, bowel, lining of the bowel wall, andomentum, forming new tumor growths before cancer is even suspected. Since a large number of people are affected by this type of cancer, we have chosen to investigate on the insilico studies of this cancer. From the results obtained from this present research project we predict the Highly Expressed genes of Ovarian cancer induction using Codon usage. We found out the various genes involved in Ovarian cancer and predicted the Insilico Gene expression levels and found out the Highly Expressed genes. We also modeled the structural attributes employing Comparative Modeling and Docking.

Key words: Protein Modeling, Gene Expression and Docking

INTRODUCTION
Worldwide, as of 2010, about 160,000 people died from ovarian cancer, up from 113,000 in 1990 [1]. As of 2014, more than 220,000 diagnoses of epithelial ovarian cancer were made yearly [2]. In 2010, in the United States, an estimated 21,880 new cases were diagnosed and 13,850 women died of ovarian cancer. Around 1800 of the new diagnoses were sex-cord or stromal tumors [3]. In the United Kingdom as of 2014, 7,000 yearly diagnoses were made and 4,200 deaths occurred [4]. The overall lifetime risk is around 1.6% [5]. The risk in the UK is similar, at 1.7% (one woman in 60). Ashkenazi Jewish women carry mutated BRCA alleles at a rate five times that of the rest of the population, putting them at higher risk for ovarian cancer [6].

In the US, ovarian cancer affects 1.3-1.4% and is the cause of death of about 1% of women [7-8]. This made it the fifth-leading cause of cancer-related deaths with an estimated 15,000 deaths in 2008 [9-10]. It occurs more commonly in developed countries [11]. Ovarian cancer is the fifth-most common cancer in women in the UK (around 7,100 women were diagnosed with the disease in 2011), and it is the fifth-most common cause of cancer death in women (around 4,300 women died in 2012) [12]. It is the most deadly gynecologic cancer. In 2014, the incidence rate for women in developed countries was about 9.4 per 100,000, compared to 5.0 per 100,000 in developing countries. The rate of ovarian cancer between 1993 and 2008 decreased in women of the 40-49 age cohort and in the 50-64 age cohort, possibly due to this group’s widespread adoption of oral contraceptives. This decrease made it the ninth-most common cancer in women.

AIM AND OBJECTIVES
The main aim of this present research project is to

To whom correspondence should be addressed:
Shakila
Email: dr_shaki@yahoo.co.in
predict the Highly Expressed genes of Ovarian cancer induction using Codon usage. 1. To find out the various genes involved in ovarian cancer. 2. To predict the In-silico Gene expression levels and to find out the Highly Expressed genes. 3. To model the structural attributes employing Comparative Modeling and Docking protocols.

METHODOLOGY

• GENE SELECTION: To collect the genes involved in Ovarian Cancer using NCBI database. The selected gene sequences are retrieved in FASTA format using NCBI databases.

• TERTIARY STRUCTURE PREDICTION: Tertiary structure prediction was performed using an automated Fold recognition modeling server called PHYRE to model the 3D structure of the GPR68, DIRAS3 and DPH1.

• MOLECULAR VISUALIZATION TOOLS: Molecular visualization tools were used to view the model GPR68, DIRAS3 and DPH1 protein structure.

• LIGAND SELECTION: PUBCHEM compound databases: suitable ligands were selected using NCBI-PUBCHEM chemical databases.

• LIGAND DESIGNING: Tertiary structure prediction of ligand was performed using MOLINSPIRATION server.

• MOLECULAR DOCKING: PATCH DOCK - server was used to analyze the docking mechanisms of the target protein and the designed ligand molecules.

• DRUG INTERACTION PREDICTION: Ligand interaction studies were done using MOLEGRO MOLECULAR and MOLSOFT viewer server.

RESULTS

Table 1. Gene Expression Analysis of Ovarian Cancer Genes (Human)

<table>
<thead>
<tr>
<th>GENE NAME</th>
<th>PROTEIN NAME</th>
<th>CHR</th>
<th>CAI VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>FILIP1L</td>
<td>Filamin A interacting protein 1</td>
<td>3</td>
<td>0.706</td>
</tr>
<tr>
<td>OVCOV1</td>
<td>Solute carrier family 35, member C2</td>
<td>20</td>
<td>0.761</td>
</tr>
<tr>
<td>GPR68</td>
<td>G protein-coupled receptor 68</td>
<td>14</td>
<td>0.893</td>
</tr>
<tr>
<td>BRAC1</td>
<td>Breast cancer 1, early onset</td>
<td>17</td>
<td>0.707</td>
</tr>
<tr>
<td>DPH1</td>
<td>DPH1 homolog</td>
<td>17</td>
<td>0.792</td>
</tr>
<tr>
<td>MUC16</td>
<td>Mucin 16</td>
<td>19</td>
<td>0.706</td>
</tr>
<tr>
<td>TSHZ2</td>
<td>Teashirt zinc finger homeobox 2</td>
<td>20</td>
<td>0.792</td>
</tr>
<tr>
<td>PHF13</td>
<td>PHD finger protein 13</td>
<td>1</td>
<td>0.758</td>
</tr>
<tr>
<td>HNRPA1</td>
<td>Heterogeneous nuclear ribonucleoprotein A1</td>
<td>12</td>
<td>0.737</td>
</tr>
<tr>
<td>ERBB2</td>
<td>Metastatic lymph node gene 19 protein</td>
<td>17</td>
<td>0.79</td>
</tr>
<tr>
<td>DIRAS3</td>
<td>GTP-binding RAS</td>
<td>1</td>
<td>0.853</td>
</tr>
<tr>
<td>CDH1</td>
<td>Cadherin 1</td>
<td>16</td>
<td>0.783</td>
</tr>
</tbody>
</table>

Table 2. Highly Expressed Genes

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Protein name</th>
<th>Chr</th>
<th>CAI value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPR68</td>
<td>G protein-coupled receptor 68</td>
<td>14</td>
<td>0.893</td>
</tr>
<tr>
<td>DPH1</td>
<td>DPH1 homolog</td>
<td>17</td>
<td>0.792</td>
</tr>
</tbody>
</table>

Table 3. Drug Docking Interaction

<table>
<thead>
<tr>
<th>TARGETS</th>
<th>ASPIRIN-MALIC ACID</th>
<th>ASPIRIN-THIOCTIC ACID</th>
<th>ASPIRIN-AMINOCAPROIC ACID</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPR68</td>
<td>-185.05</td>
<td>-271.31</td>
<td>-246.12</td>
</tr>
<tr>
<td>DPH1</td>
<td>-233.98</td>
<td>-313.86</td>
<td>-246.94</td>
</tr>
</tbody>
</table>
Fig 1. 3D Structure of GPR68 Proteins

The above picture represents the pink color indicates helix, yellow indicates the sheets, and blue indicates the turn and white indicates the coils regions in the GPR68 protein.

Fig 2. 3D Structure of DPH1 Proteins

The above picture represents the pink color indicates helix, indicates the sheets, and blue indicates the turn and white indicates the coils regions in the DPH1 protein.

Fig 3. Calculation of Molecular Properties of Aspirin and Malic Acid

Table: Molinspiration

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>mLogP</td>
<td>-0.361</td>
</tr>
<tr>
<td>TPSA</td>
<td>147.436</td>
</tr>
<tr>
<td>Natom</td>
<td>22</td>
</tr>
<tr>
<td>MW</td>
<td>312.23</td>
</tr>
<tr>
<td>nCOSH</td>
<td>9</td>
</tr>
<tr>
<td>nSAS</td>
<td>0</td>
</tr>
<tr>
<td>nrot</td>
<td>3</td>
</tr>
<tr>
<td>Volume</td>
<td>532.432</td>
</tr>
</tbody>
</table>

Get data as text (for copy / paste).
Get 3D geometry NBT7A
L. Shakila and Shalini H, In-silico gene expression and Drug Docking Studies on Human Ovarian Cancer Disease Proteins (GPR68, DIRAS3 AND DPH1) using bioinformatics software and tools

Fig 4. Calculation of drug likeness

Fig 5. Calculation of Molecular Properties of Aspirin and Aminocaproic Acid

Fig 6. Calculation of drug likeness
L. Shakila and Shalini H, In-silico gene expression and Drug Docking Studies on Human Ovarian Cancer Disease Proteins (GPR68, DIRAS3 AND DPH1) using bioinformatics software and tools

Fig 7. Calculation of molecular properties of Aspirin and Thiocytic Acid

Fig 8. Calculation of drug likeness

Fig 9. Docking and drug interaction of aspirin with Aminocaproic and DPH1
L. Shakila and Shalini H, In-silico gene expression and Drug Docking Studies on Human Ovarian Cancer Disease Proteins (GPR68, DIRAS3 AND DPH1) using bioinformatics software and tools

Fig 10. Docking and drug interaction of Aspirin with Mallic acid and DPH1

Fig 11. Docking and drug interaction of aspirin with Thiocic acid and DPH1

Fig 12. Docking and drug interaction of Aspirin with Aminocaproic acid and GPR68
L. Shakila and Shalini H, In-silico gene expression and Drug Docking Studies on Human Ovarian Cancer Disease Proteins (GPR68, DIRAS3 AND DPH1) using bioinformatics software and tools

Fig 13. Docking and drug interaction of Aspirin with Mallic acid and GPR68

![Protein Wireframe /Atom model and De Novo ligand Stick Model](image1.png)

Protein Wireframe /Atom model and De Novo ligand Stick Model
Protein – red oxygen, Blue –nitrogen and Grey – carbon

Fig 14. Docking and drug interaction of aspirin with Thioctic acid and GPR68

![Protein Wireframe /Atom model and De Novo ligand Stick Model](image2.png)

Protein Wireframe /Atom model and De Novo ligand Stick Model
Protein – red oxygen, Blue –nitrogen and Grey – carbon

DISCUSSION
In the present research work a new ligand candidate is developed using bioinformatics software and tools. The Ovarian Cancer Nucleotide sequences were collected through NCBI (National Center For Biotechnology Information) which is a part of the United States National Library of Medicine (NLM), a branch of the National Institutes of Health.

GENE EXPRESSION ANALYSIS
The selected genes were applied to ACUA software in order to predict the gene expression values. Automated Codon Usage Analysis –It is a visual
The primary structural analysis

In this study, we found out the molecular weight and theoretical pi values of the target protein. These results are helpful in the field of biological studies (fig1.0 DPH1 and 1.1 GPR68). Protein pI is calculated using pK values of amino acids described in Bjellqvist et al., which were defined by examining polypeptide migration between pH 4.5 to 7.3 in an immobilised pH gradient gel environment with 9.2M and 9.8M urea at 15ºC or 25ºC. Prediction of protein pI for highly basic proteins is yet to be studied and it is possible that current Compute pI/Mw predictions may not be adequate for this purpose. Protein Mw is calculated by the addition of average isotopic masses of amino acids in the protein and the average isotopic mass of one water molecule.

The highly expressed gene coded protein sequences (GPR68 and DPH1) are modeled through PHYRE server. Determining the structure and function of a novel protein is a cornerstone of many aspects of modern biology. Over the past decades, a number of computational tools for structure prediction have been developed. It is critical that the biological community is aware of such tools and is able to interpret their results in an informed way. This protocol provides a guide to interpreting the output of structure prediction servers in general and one such tool in particular, the protein homology/analogy recognition engine (Phyre). The tertiary structures of (GPR68 and DPH1) are predicted using PHYRE web server fig (GPR68 5.0 and DPH1 5.1). The modeled structures are viewed using advanced molecular visualization tools like MOLSOF and MOLEGRO NOLEcular VIEWER fig(4.0 DPH1 and GPR68 4.1).

**STRUCTURE VALIDATION**

RAPPER is an *ab initio* conformational search algorithm for restraint-based protein modelling. It has been used for all-atom loop modelling (De Pristo et al., de Bakker et al.), whole protein modelling under limited restraints (De Pristo et al.), comparative modelling (de Bakker et al., in preparation), *ab initio* structure prediction, structure validation (Lovell et al.), and experimental structure determination with X-ray and nuclear magnetic resonance spectroscopy (De Pristo et al.). During the study when the targets (GPR68, DIRAS3 and DPH1) are subjected to run, the RAPPER target displayed a range of (fig ) GPR68 95.9 and DPH1 89.4 %. Since the obtained values are higher than the minimum expected value it explains that the predicted structure is of good quality. The prediction of good structure is important in the present study because the main objective of the investigation is to design a new ligand for target.

**DRUG DESIGNING**

The three acid groups combined with non-steroidal (NASD) Aspirin were introduced and the structural properties of the chemical were analysed. PubChem is a database of chemical molecules. The system is maintained by the National Center for Biotechnology Information (NCBI), a component of the National Library of Medicine, which is part of the United States National Institutes of Health (NIH). PubChem can be accessed for free through a web user interface. PubChem contains substance descriptions and small molecules with fewer than 1000 atoms and 1000 bonds. The existing drug Aspirin and agents like Malic Acid, Aminocaproic Acid and Thiocict Acid are selected for the study. The drug Aspirin has anti-inflammatory activity and Malic Acid, Aminocaproic Acid and Thiocic Acid have an anticancer activity.

**DRUG VALIDATION**

In the present research further modifications of ligands were made using web server Molinspiration. One way of estimating drug likeness is Lipinski's
Rule of Five. Due to the complexity of the drug design process, two terms of interest are still serendipity and bounded rationality. Those challenges are caused by the large chemical space describing potential new drugs without side-effects. Molinspiration is a type of web server which does calculation of molecular physicochemical properties relevant to drug design and Quantitative structure-activity relationship QSAR, including logP, molecular polar surface area (PSA) Calculation of activity score, drug-likeness for G protein-coupled receptors GPCR ligands, ion channel modulators and kinase inhibitors. From the results obtained shown in (Fig 6.0 to 6.5), we mainly found out that the predicted chemical structures obey the drug likeness score values.

TOXICITY PREDICTION
Toxicity prediction is one of the basic studies before drug delivery. In our studies, the designed chemical structure was applied in Tox tree tool. Pharmacological or biological activity is an expression describing the beneficial or adverse effects of a drug on living matter. When the drug is a complex chemical mixture, this activity is exerted by the substance’s active ingredient or pharmacophore but can be modified by the other constituents. The main kind of biological activity is a substance’s toxicity. TOX tree server: the designed ligand is applied to Tox tree server in order to predict its toxicity. Fig (9.0 to 9.2) shows that the total amount of toxicity present in the de novo drug is nil.

LIGAND DOCKING
PATCHDOCK algorithm is inspired by object recognition and image segmentation techniques used in Computer Vision. Docking can be compared to assembling a jigsaw puzzle. When solving the puzzle we try to match two pieces by picking one piece and searching for the complementary one. The designed denovo ligand with potential highly expressed GPR68, DIRAS3 and DPH1 candidates the molecular docking results shows the binding affinities between the ligand and receptor. In fig (10.0 to 11.5) shows the designed ligand successfully bind to the antigen sites of the GPR68, and DPH1.

CONCLUSION
The genes associated with Ovarian cancer are identified by manual as well as by curated data mining using various internet search engines like medline and pubmed databases. Ovarian Cancer is highly prevalent among women in South India. In this study, our focus was on the potential target. The best drug candidate (ASPIRIN-THIOCTIC ACID) for the DPH1 was found out using Bioinformatics software and tools. Drug chemical structures that obey the drug likeness score were introduced. The results of ADME property strongly proved that there is no toxic atom present in the designed chemical structure. Based on the docking results we find out that the designed chemical structures act as a good controller of Ovarian Cancer.

SUMMARY
The principal objective of the present investigation is to find out the highly expressed candidates. In this project, we found out that 3 gene-coded-proteins are the best candidates for drug docking. Our aim is to decrease the pain in cancer patients as well as to reduce the growth rate of Cancer. We suggest that the designed drugs are best candidates for reducing the growth of cancer as well as pain based on our binding affinity values.

REFERENCES
6. Hoffman, Barbara L, Schorge, John Q, Schaffer, Joseph I, Halvorson, Lisa M, Bradshaw, Karen D,


