IN SILICO SCREENING OF MAJOR CANCER DRUG TARGETS (GROWTH FACTOR RECEPTORS) FOR NATURE DERIVED PHYTOCHEMICALS

1Tithi S. Trivedi, 2Sivakumar Prasanth Kumar, 1Archana U. Mankad*, 1Saumya K. Patel, 2Rakesh M. Rawal and 1,2Himanshu A. Pandya

1Department of Botany, Bioinformatics and Climate Change Impacts Management, 2Department of Life Science, School of Sciences, Gujarat University, Ahmedabad-380 009.
Corresponding Author Email-ID: aumankad@gujaratuniversity.ac.in

ABSTRACT
Cancer is a group of abnormal cells. The unregulated growth factor receptor tyrosine kinase (GFR- TK) proteins are implicated in the proliferation of more than 60% of all cancer types. Screening of phytochemicals for their anti-angiogenic potential has been a growing area of research in the modern decade. There is a well-known principle that natural compounds are active against several diseases, including various types of cancer. The present research work focuses on known growth factor receptors (GFRs) as an important target for computational studies. In this study, 96 curated anti-cancer compounds were virtually screened against the EGFR, FGFR, IGFR, and HGFR using molecular docking software. For each GFR, we have considered ten top most results as potential hits. Among them, common five results are: Spirosolan, Ginkgetin, Fangchinoline, Theaflavin and Ursolic acid. These compounds have been reported to show anticancer activities in the literature. With the help of different interaction analysis tools, the protein-ligand interaction patterns between the functional groups of these compounds were analyzed. Hydrogen bonding and hydrophobic forces are the main components of these interactions of these hits, similar to those experimental for the known inhibitors. From the maximum number of hits, it could be indicated that the compounds Spirosolan, Ginkgetin, Fangchinoline, Theaflavin and Ursolic acid are pro-miscuous lead in the drug discovery process.

Keywords: Growth Factor Receptors, Tyrosine Kinase, Virtual screening, Structure-based drug design, Cancer, Phytochemicals

INTRODUCTION
According to WHO cancer fact sheets, cancer is the second leading cause of death globally and is responsible for an estimated 9.6 million deaths in 2018. Globally, about 1 in 6 deaths is due to cancer (ref link: http://www.who.int/news-room/fact-sheets/detail/cancer). The burden of cancer has been predicted to be increasing globally up-to 15 million by the year 2020, wherein the developing countries would be at a higher risk relating to incidences of cancer morbidity and mortality. Growth factor receptors (GFRs) expressed on cell membranes or in the cytoplasm, have profound roles in cell growth, survival angiogenesis, and metastasis. They contain three domains: an extracellular ligand (growth-factor) binding domain, a transmembrane domain, a cytoplasmic domain that acts as an enzyme, or forms a complex with another protein that acts as an enzyme. Here we have discussed the "In-Silico" screening of major cancer drug targets – growth factor receptors (EGFR, FGFR, IGFR, and HGFR) against nature-derived phytochemicals. There are several efficient anticancer drugs and active inhibitors against several protein targets, growing resistance attached with many side effects.
mean that there is a need for new, enhanced treatments\(^3\). Plants are important sources of the active ingredients used in modern medicines. More than half of the drugs approved since 1995 are based on natural products\(^4,6\).

EGFR has been a chief target of atomic anti carcinoma treatment. Cellular levels of EGFR do not always associate with the reply to the EGFR TK slow downers, indicating acquired resistance to drugs\(^7\). Combination of EGFR antagonists with slow downers targeting different signaling mechanism(s) – such as IGFR and VGFR – that share the same downstream mediator may get around or hold – up the development of resistance to EGFR antagonists’ resultant in enhanced antitumor activities\(^7\). FGFR over expression, point mutation, or gene fusions is found in 80% of non – muscle – invasive and 15% of muscle-invasive bladder carcinoma\(^8\). High expression levels of the IGF-1R have been found in breast and colorectal carcinoma. The most significant connection. Between augmented levels of IGF-1 and the risk of carcinoma, the diagnosis was found for prostate carcinoma, premenopausal breast carcinoma, and colorectal carcinoma\(^9\). The HGF was first recognized as a solvable factor indorsing hepatocyte growth and liver renewal. The HGF receptor (MET) is a prototypal TK receptor. The remark that MET is a proto-oncogene and that its signaling is frequently conquered in carcinoma has permitted deep investigations on its atomic structure and signaling properties that are deliberated. In carcinoma, HGFR has been concerned with cellular expansion, metastasis, and angiogenesis\(^10\). The availability of GFRs structures implies that VS could be used as a tool to search for potential active compounds from medicinal herbs. The structures of the active compounds found in the folk – medicinal herbs have been collected in the PubChem database\(^11\). We performed a virtual screening protocol using the YASARA server. They were applied to screen the 96 curated collection of natural molecules. The resulting hits were analyzed to gain insights into the key structural features required for good protein-ligand interaction.

Materials and methods

**Protein preparation**

As the selection of target proteins, EGFR, FGFR, IGFR, and HGFR, their proteins have been selected. Protein structures are available for these proteins. Based on low resolution, targets have been selected. Preparation of protein before docking was done by deleting water molecules and clearing the protein in the YASARA structure followed by energy minimization using the same.

**Ligand preparation**

NPACT database is an important database related to the plant-derived natural anticancer compounds\(^12\). From this database, 96 natural compounds have been selected, which were further compiled into data-sets and used for the docking purpose. These ligands were cleaned and the hydrogen was added to the ligands for the further procedure. Their 2D structures were converted to 3D using BIOVIA Discovery Studio\(^13\).

<table>
<thead>
<tr>
<th>Allicin</th>
<th>Gallic Acid</th>
<th>Embelin</th>
<th>Flavonol glycoside</th>
<th>3-O-Methyltyramine</th>
<th>Combretastatin</th>
<th>Mangiferin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allyl isothiocyanate</td>
<td>Genistein</td>
<td>Nitisine</td>
<td>Damacidantanol</td>
<td>(E)-(\gamma)-bisabolone</td>
<td>Emodin tetraacetate</td>
<td>Quercetin-3-glucoside</td>
</tr>
<tr>
<td>Andrographoloid</td>
<td>Gingerol</td>
<td>Protopine</td>
<td>Mallato phenone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apigenin</td>
<td>Gossypol</td>
<td>Psoralen</td>
<td>Podophyllotoxin bromide</td>
<td>Cynaropicrin</td>
<td></td>
<td>Costunolide</td>
</tr>
<tr>
<td>Apiole</td>
<td>Hesperidin</td>
<td>Aloe-emodin</td>
<td>Diterpenes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AntineoplasticoneA10</td>
<td>Indole-3-carbinol</td>
<td>Thymoquinone</td>
<td>Lupeol</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capsaicin</td>
<td>Kaempferol</td>
<td>Cucurbitacin</td>
<td>Withaferin A</td>
<td>Berberine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methyl</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1: 96 curated phytochemicals from NPACT database

Virtual screening proceduresMolecular Docking
YASARA Structure (version 20.7.1) has been utilized for the protein-ligand docking purpose\textsuperscript{14}. It is based on the Autodock vina algorithm and it utilizes the following formula for calculating the docking score

\[ \Delta G = \Delta G (\text{vdw}) + \Delta G (\text{HBond}) + \Delta G (\text{elec}) + \Delta G (\text{tor}) + \Delta G (\text{desolv}) \]

Wherein, \( \Delta G (\text{vdw}) \) is the component energy terms related to van der waals bond, \( \Delta G (\text{HBond}) \) is the component energy term related to Hydrogen bonds, \( \Delta G (\text{elec}) \) is the component energy terms related to electrostatics, \( \Delta G (\text{tor}) \) is the component energy term related to the ligand’s torsional free energy and \( \Delta G (\text{desolv}) \) is the component energy term related to the desolvation for the empirical calculation of the docking/binding energy for a protein-ligand complex. The Higher docking score represents the better protein-ligand binding whereas, the negative score represents no binding between the Protein and the ligand. The protein-ligand interactions were further visualized in 3D and 2D using the BIOVIA Discovery Studio Visualizer.

Post – docking analysis

We used several tools to analyze the interactions between the EGFR, FGFR, IGFR, and HGFR and hit compounds. The most important interactions of the protein and ligand compare from the docked results were calculated and the hit compounds were further analyzed using BIOVIA DS\textsuperscript{13} the PyMol\textsuperscript{15}.

Validation of the results through Re-docking

Re-docking is the process of docking in which the receptor has been docked with its natural ligand. By doing this, we could check the behavior of another ligand, whether they are as similar to a natural ligand or not. In case of the absence of natural ligand, then its slow – downer has been selected for re – docking, it is called "Reference – ligand".

RESULTS AND DISCUSSION

Characterization of selective growth factor receptors as potential drug targets

Based on the low-resolution value, the targets for docking purposes have been selected from the PDB database\textsuperscript{16}. The detailed information of targets has been given below.
Table 2: Growth Factor Receptors as protein targets with detail

Molecular Docking

The molecular docking studies expose the binding affinity of the natural ligands towards the GFRs protein targets. Different types of interactions such as hydrogen bond interactions, pi – sigma bond, alkyl bond, and pi – alkyl bond were observed in the protein-ligand interaction complex. Key contacting receptor residues were also noted. The top three protein-ligand complexes with the respective binding energies, dissociation constants, and the number of hydrogen bonds for four protein targets are noted in Table – 3, Table – 4, Table – 5, and Table – 6. Better binding of the ligand towards the protein chain is indicated by the higher positive binding energy whereas the negative binding energies indicate no binding17.

<table>
<thead>
<tr>
<th>Sr. Num.</th>
<th>Ligand Name</th>
<th>Binding Energy [kcal/mol]</th>
<th>No. of H-bonds</th>
<th>Contacting Receptor Residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Fangchinoline</td>
<td>8.9900</td>
<td>1</td>
<td>LEU 718, GLY 719, SER 720, GLY 724, VAL 726,ALA 743, LYS 745, SER 752, MET 766, THR 790, MET 793, GLY 796, CSX 797, ASP 800, GLU 804, ARG 841, ASN 842, LEU 844, THR 854, ASP 855</td>
</tr>
<tr>
<td>2</td>
<td>Psoralidin</td>
<td>8.8240</td>
<td>2</td>
<td>LEU 718, VAL 726,ALA 743, LYS 745, THR 790, GLN 791, LEU 792, MET 793, PRO 794, PHE 795, GLY 796, CSX 797, ASP 800, GLU 804, LEU 844, THR 854, ASP 855</td>
</tr>
<tr>
<td>3</td>
<td>Ursolic Acid</td>
<td>8.8020</td>
<td>2</td>
<td>LEU 718, GLY 719, VAL 726,ALA 743, LYS 745, THR 790, GLN 791, LEU 792, MET 793, GLY 796, CSX 797, LEU 799, ASP 800, ARG 803, ARG 841, ASN 842, LEU 844, THR 854, ASP 855</td>
</tr>
</tbody>
</table>

Table 3: Top 3 Molecular Docking Results of phytochemicals with EGFR
**Figure: 1 2D and 3D Protein ligand complex representation of EGFR protein with ligand Fangchinoline**

<table>
<thead>
<tr>
<th>Sr. Num.</th>
<th>Ligand Name</th>
<th>Binding Energy [kcal/mol]</th>
<th>No. of H-bonds</th>
<th>Contacting Receptor Residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Theaflavin</td>
<td>10.8310</td>
<td>7</td>
<td>PHE 492, GLY 493, GLN 494, VAL 495, LYS 517, MET 518, LEU 519, LYS 526, ASP 527, ASP 530, LEU 531, GLU 534, ASP 626, ARG 630, ASN 631, ASP 644, GLY 646, LEU 647, ARG 664</td>
</tr>
<tr>
<td>2</td>
<td>Spirosolane</td>
<td>10.3530</td>
<td>1</td>
<td>GLY 488, GLU 489, GLY 490, CYS 491, PHE 492, GLY 493, VAL 495, LYS 517, ASP 527, ASP 530, LEU 531, GLU 534, ASN 571, ARG 630, LEU 633, ASP 644, GLY 646, LEU 647, ARG 664</td>
</tr>
<tr>
<td>3</td>
<td>Silymarin</td>
<td>10.0680</td>
<td>5</td>
<td>GLY 488, GLU 489, GLY 490, CYS 491, PHE 492, GLY 493, VAL 495, LYS 517, ASP 527, ASP 530, LEU 531, GLU 534, ASN 571, ARG 630, LEU 633, ASP 644, GLY 646, LEU 647, ARG 664</td>
</tr>
</tbody>
</table>

*Table 4: Top 3 Molecular Docking Results of phytochemicals with FGFR*
Figure: 2 2D and 3D Protein ligand complex representation of FGFR protein with ligand Theaflavin

<table>
<thead>
<tr>
<th>Sr. Num.</th>
<th>Ligand Name</th>
<th>Binding Energy [kcal/mol]</th>
<th>No. of H-bonds</th>
<th>Contacting Receptor Residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Fangchinoline</td>
<td>9.2850</td>
<td>4</td>
<td>GLU 1046, ASN 1049, GLU 1050, VAL 1053, PHE 1131, VAL 1132, HIS 1133, ARG 1134, ASP 1135, PHE 1154, GLY 1155, ARG 1158, LEU 1173, LEU 1174</td>
</tr>
<tr>
<td>2</td>
<td>Theaflavin</td>
<td>8.8080</td>
<td>4</td>
<td>LEU 1005, GLY 1006, GLN 1007, VAL 1013, ALA 1031, LYS 1033, VAL 1063, MET 1079, GLU 1080, LEU 1081, MET 1082, THR 1083, ARG 1084, GLY 1085, ASP 1086, SER 1089, MET 1142, ASP 1153, MET 1156, THR 1157, ILE 1160</td>
</tr>
<tr>
<td>3</td>
<td>Ellagic Acid</td>
<td>8.6380</td>
<td>3</td>
<td>LEU 1005, GLY 1006, GLN 1007, GLY 1008, VAL 1013, ALA 1031, LYS 1033, VAL 1063, MET 1079, GLU 1080, LEU 1081, MET 1082, GLY 1085, MET 1142, ASP 1153, MET 1156, ILE 1160</td>
</tr>
</tbody>
</table>

Table 5: Top 3 Molecular Docking Results of phytochemicals with IGFR

Figure: 3 2D and 3D Protein ligand complex representation of IGFR protein with ligand Fangchinoline
Different families of growth factors and growth factor receptors are involved in the autonomous growth of cancer cells\textsuperscript{18}. Fangchinoline, Psoralidin, and Ursolic acid were the three phytochemicals that achieved better binding with EGFR. They developed a single, two, and two hydrogen bounds respectively. It has been observed to be interacting with the protein EGFR with the binding energies of 8.990 kcal/mol, 8.824 kcal/mol, and 8.802 kcal/mol, respectively. The protein-ligand complex of EGFR with the topmost ligand Fangchinoline has shown the hydrogen bond interacting with the protein residue ASN842. For FGFR, Theaflavin, Spirosolane, and Silymarin ligands were achieved better binding. They showed binding energies of 10.891 kcal/mol, 10.3530 kcal/mol, and 10.068 kcal/mol, respectively with the protein FGFR. Theaflavin (the topmost ligand) has shown seven hydrogen bonds, they are ARG664, GLY646, LEU647, ASP626, GLN494, MET518 and ASP527. Again, Fangchinoline, Theaflavin, and Ellagic acid were achieved better binding with IGFR. They developed a single,
four and four hydrogen bonds, respectively. For Fangchinoline, the single hydrogen bond is ARG1158. It has been observed to be interacting with the protein IGFR with the binding energies of 9.285 kcal/mol, 8.808 kcal/mol, and 8.638 kcal/mol respectively. For HGFR, Theaflavin, Withanolide, and Fangchinoline were achieved better binding. They showed binding energies of 10.058 kcal/mol, 9.211 kcal/mol, and 9.147 kcal/mol respectively, with the protein HGFR. The topmost ligand, Theaflavin has shown ten hydrogen bond interactions. They are LEU57, GLU493, VAL55, LEU120, TYR321, VAL322, HIS251, ALA320, VAL490, and GLU489. Spirosolane was interacting with the protein EGFR with the binding energy of 8.712 kcal/mol, with FGFR protein with the binding energy of 10.353 kcal/mol, with IGFR protein with the binding energy of 8.476 kcal/mol and with HGFR protein with the binding energy of 8.5470 kcal/mol. Ginkgetin was also showing the interaction with all four proteins- EGFR, FGFR, IGFR and HGFR with different binding energies 8.717 kcal/mol, 9.842 kcal/mol, 8.617 kcal/mol, 8.986 kcal/mol respectively. Ursolic acid was also showing the interaction with proteins FGFR and HGFR with the binding energy of 9.729 kcal/mol and 8.69 kcal/mol respectively.

CONCLUSION
Our results indicated that the five phytochemicals, obtained as hits using YASARA screening method could all be inhibitors of selective growth factor receptors. This suggests that YASARA molecular docking procedure can be effectively used to predict the binding modes of NPACT compounds. These small molecules were obtained from fruits and vegetables that are accessible locally, and they have been used as components of primitive medicinal recipes. Anticancer activities were reported for all five of the hit compounds, which mean they could be potential lead in the drug discovery in the future. Additionally, the molecular hits identified in the study can be further tested in vitro before claiming its inhibitory potential.

REFERENCES