Probing natural molecules with PPAR-γ to reveal potent agonist against Cancer

1st Adriána Špaková  
Department of Biomedical Engineering  
Faculty of Electrical Engineering and Communication, Brno  
University of Technology, Brno, Czech republic  
xsampo00@vutbr.cz

2nd Vaishali Pankaj  
Department of Biomedical Engineering  
Faculty of Electrical Engineering and Communication, Brno  
University of Technology, Brno, Czech republic  
234084@vut.cz

3rd Inderjeet Bhogal  
Department of Biomedical Engineering  
Faculty of Electrical Engineering and Communication, Brno  
University of Technology, Brno, Czech republic  
234190@vut.cz

4th Sudeep Roy  
Department of Biomedical Engineering  
Faculty of Electrical Engineering and Communication, Brno  
University of Technology, Brno, Czech republic  
roy@vut.cz

Abstract - The work focuses on searching for a molecule with potential agonistic properties against cancer. Molecules from six databases were screened and docked to peroxisome proliferator-activated receptor gamma (PPAR-γ) by using computer-aided drug design approach. Hits underwent further exploration, including dynamic simulation and safety verification. Piperlongumine - naturally occurring small molecule, derived from long pepper (Piper longum) showed after comparison with standards Troglitazone and Rosiglitazone promising results.

Keywords - Cancer, PPAR-γ, target, natural molecules, ligand, virtual screening, molecular docking, dynamic simulation, ADMET

I. Introduction

According to World health organization (WHO), cancer is one of the leading cause of death worldwide, accounting for nearly 10 million deaths in 2020, or nearly one in six deaths. [1] The numbers vary between countries. In the US, cancer is the second most common cause of death, exceeded only by heart disease. [2] Just like in the US, cancer is the second most common cause of death after cardiovascular diseases also in the Czech Republic, where around 100,000 people fall ill with cancer every year and 30 000 people die due to cancer. [3]

Cancer, as a complex and heterogeneous group of diseases, is characterized by uncontrolled cell growth and division, often leading to the formation of malignant tumours. Treatment options include surgery, chemotherapy, radiotherapy, personalized targeted therapy, and immunotherapy. [4]

Despite considerable advancements in cancer research and treatment modalities, the need for innovative approaches remains paramount. The exploration of natural molecules as potential agonists against cancer has emerged as a promising avenue, with particular emphasis on their interaction with the PPAR-γ. [2][4] The reason is that many drugs have their origin in natural sources, such as plants, animals, fungi, and microorganisms. E.g. aspirin which has analgesic and antipyretic properties came from tree bark, and morphine, known as pain reliever, has its origin in seeds of the opium poppy. Natural molecules serving as ligands bind to a receptor or target molecule and form a complex that triggers a biological response - this time the effect is required in cancer cells.

II. Ease of Use

A. PPAR-γ

PPAR-γ is a ligand-activated transcription factor that regulates genes, which are important in cell differentiation and various metabolic processes, especially lipid and glucose homeostasis. After interaction with the specific ligand, nuclear receptor is translocated to the nucleus, where it changes its structure and regulates gene transcription. [5] The intricate relationship between PPAR-γ and cancer involves pathway modulation, especially cell proliferation, differentiation, apoptosis, angiogenesis, and inflammation. Although, its involvement in cancer is multifaceted, with both pro- and anti-tumorigenic effects reported in various contexts, the demand for PPAR-γ agonists as chemotherapeutic agents persists. [6][7][8]

B. Thiazolidinediones

Thiazolidinediones (TZDs) are synthesized PPAR-γ agonists, primarily utilized in the treatment of type 2 diabetes as insulin sensitizers. In comparison to other ligands, including mainly fatty acids, TZDs bind to PPAR-γ more firmly, with an affinity in the range of 40–200 nM. [2][9] Over a few TZDs generations, ciglitazone, troglitazone, rosiglitazone and pioglitazone were introduced. Some of them were withdrawn from the market due to risk of hepatotoxicity or cardiovascular incidents. Recently, TZDs proved also anti-cancer effects, dependently or independently of PPAR-γ activation, in monotherapy and in combined treatment with chemotherapeutics, both, on the transcriptional and the protein level. [2] The goal of the work is to find a natural molecule that will have better properties than standards from TZDs family.

C. Principle of gene expression

PPAR-γ belongs to a subset of nuclear receptors that form heterodimers with the retinoid X receptor (RXR), greatly enhancing the ability of the receptor to bind specific DNA sequences in target genes. The DNA sequences recognized by the PPAR–RXR heterodimer are referred to as PPAR-response elements (PPREs). PPAR–RXR heterodimers bind to PPREs in the absence of ligand. After that, ligand is bound, which leads to a PPAR-γ conformational change that results in activation of
transcription of the target gene. [9] The Fig. 1 depicts mechanism of PPAR-γ in the function of transcription factor.

![Fig. 1. Mechanism of PPAR-γ in role of transcription factor](Image)

**III. COMPUTER AIDED DRUG DESIGN**

Bringing a new drug into the market is a costly process in terms of money, manpower, and time. Conventional approach of drug discovery and development takes an average of 10-15 years with an approximate cost of 800 million to 1.8 billion US dollars. Over the last few decades, CADD has emerged as a powerful technique playing a crucial role in the development of new drug molecules. In biomedical area, CADD is being utilized to accelerate and aid hit identification, hit-to-lead selection, optimize the absorption, distribution, metabolism, excretion, and avoid safety issues. [10][11][12]

**A. Target and ligands Selection and Preperation**

The 2PRG protein structure was downloaded from the RCSB Protein Data Bank. The designation 2PRG stands for ligand-binding domain of human peroxisome proliferator activated receptor gamma. The downloaded receptor was incorporated into the Maestro environment, followed by protein preparation. The preparation consists of mainly H-bond optimization and water molecules minimization, what results in biologically relevant and energetically favourable protein structure state. Then receptor grid generation was required, what is a three-dimensional space generated around a target protein’s binding site, where the ligand is expected to bind.

The next step included the selection and preparation of the ligand libraries to perform reliable virtual screening later. Molecules were downloaded from six different databases, namely: natural molecules from Interbioscreen, NP Atlas and Enamine, synthetic molecules from Lead-like and Drug-like database (Swiss similarity) and kinases from ChemDiv database. Standards Troglitazone and Rosiglitazone were downloaded from PubChem.

**B. Structure-based Virtual Screening**

Virtual screening narrows down the vast chemical space and prioritize compounds for experimental testing in the early stages of drug discovery. It helps to identify lead compounds that have a higher likelihood of binding to the target of interest. Virtual screening in that way save time and expense. Within the virtual screening, more than 7000 ligands and 2 standards were docked to the 2PRG receptor. Predicting ligand-protein binding affinities were determined by using extra precision docking mode with post-docking MMGBBSA. Docked complexes were then ranked according to their scores and the top-ranked compound from each library was selected as potential drug candidate. From Interbioscreen natural compound database and NP Atlas database, instead of one top hit, two top hits were used in further validation. The Table 1 depicts glide gscore of hits and standard molecules, where lower gscore indicates a more favourable binding interaction. Most of hits provided better binding affinity score than standard ligands. The best glide gscore was obtained by molecule C21H16FNO3S (-14,791) and C24H20O10S (-14,577) from Drug-like database. Very similar results have been shown also by molecules C28H39NO6 (-14,577) and C24H20O10S (-14,38) from NP Atlas and by molecule C25H28N2O6 (-14,099) from Interbioscreen natural compound database. Important is also the representation of amino acids of the target protein, which form certain bonds with the given ligand. E.g. top-ranked ligand C21H16FNO3S from Drug-like library forms bonds with following amino acids: Gln 286, Tyr 473, His 323, Ser 289.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Library</th>
<th>Virtual Screening</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Glide gscore</td>
</tr>
<tr>
<td>Rosiglitazone</td>
<td>PubChem</td>
<td>-11,644</td>
</tr>
<tr>
<td>Troglitazone</td>
<td>PubChem</td>
<td>-12,377</td>
</tr>
<tr>
<td>C15H13N03S</td>
<td>Lead-like</td>
<td>-12,526</td>
</tr>
<tr>
<td>C21H16FNO3S</td>
<td>Drug-like</td>
<td>-14,791</td>
</tr>
<tr>
<td>C15H13N04</td>
<td>Enamine</td>
<td>-11,009</td>
</tr>
<tr>
<td>C25H28N2O6</td>
<td>IB - top1</td>
<td>-14,099</td>
</tr>
<tr>
<td>C17H13N05</td>
<td>IB - top2</td>
<td>-14,099</td>
</tr>
<tr>
<td>C21H12F2N02</td>
<td>ChemDiv</td>
<td>-10,814</td>
</tr>
<tr>
<td>C28H39NO6</td>
<td>NPAtlas</td>
<td>-14,577</td>
</tr>
<tr>
<td>C24H20O10S</td>
<td>NPAtlas - top2</td>
<td>-14,380</td>
</tr>
</tbody>
</table>

**C. Induced Fit Docking**

Traditional docking methods, which virtual screening is a part of, assume a rigid structure of receptor. Unlike that, induced fit docking (IFD) considers the flexibility of the protein as well as the ligand. It provides a more realistic representation of the ligand-protein interaction by considering the dynamic nature of both components. As shown in Fig. 2, the docking procedure is typically iterative - potential binding positions are generated.

![Fig. 2. Rigid vs flexible molecular docking](Image)
For IFD were used hits from virtual screening. For each hit were generated several poses and the Table 2 depicts results of the best pose for each molecule. The rule is applied again, lower score suggests a more stable protein-ligand complex. Standard troglitazone provided the best docking (-14,818) and IFD (-1236,876) score. Second best docking score was provided by molecule C21H16FNO3S (-14,329) from Drug-like library until the second best IFD score has been shown by molecule C25H28N2O6 (-1234,578) from Interbioscreen database.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Library</th>
<th>Induced Fit Docking [XP]</th>
<th>Amino acids</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rosiglitazone</td>
<td>FullChem</td>
<td>-1126,788</td>
<td>Ser 286, Arg 286, Tyr 271, His 223</td>
</tr>
<tr>
<td>Troglitazone</td>
<td>FullChem</td>
<td>-1276,676</td>
<td>Gln 286, Ser 286, Ser 289</td>
</tr>
<tr>
<td>C19H19N3O5S</td>
<td>Drug-like</td>
<td>-8,220</td>
<td>Ser 286, His 449</td>
</tr>
<tr>
<td>C21H16FNO3S</td>
<td>Drug-like</td>
<td>-14,329</td>
<td>Gln 286, Ser 286, His 223</td>
</tr>
<tr>
<td>C25H28N2O6</td>
<td>Examine</td>
<td>-9,056</td>
<td>Tyr 267, Asp 286, Ser 289, His 325, Cys 185</td>
</tr>
<tr>
<td>C25H28N2O6</td>
<td>Best-top1</td>
<td>-1229,820</td>
<td>Ser 286, His 323, Ser 289</td>
</tr>
<tr>
<td>C19H19N3O5S</td>
<td>Best-top2</td>
<td>-1229,820</td>
<td>Ser 286, His 323, Ser 289</td>
</tr>
<tr>
<td>C25H28N2O6</td>
<td>ChemElle</td>
<td>-1224,815</td>
<td>Tyr 267, His 449, Ser 289</td>
</tr>
<tr>
<td>C25H28N2O6</td>
<td>Molecules</td>
<td>-13,135</td>
<td>Cys 285, Cys 187, Ser 442</td>
</tr>
<tr>
<td>C24H19C10H10</td>
<td>Molecules</td>
<td>-13,421</td>
<td>Tyr 237, Tyr 479, His 236, Cys 285, Arg 286, Ser 342</td>
</tr>
</tbody>
</table>

Again, the amino acids involved in the bonds between the protein and the ligands were described. The best pose of standard troglitazone forms bonds with 3 amino acids: Gln 286, Hie 323 and Ser 289. The same amino acids are visible in case of best pose of ligand C21H16FNO3S from Drug-like library and best pose of molecule C25H28N2O6 from Interbioscreen natural compound database forms bonds with totally different amino acids: Ser 342, Hie 449 and Tyr 327.

### D. Dynamic Simulation

Molecular dynamic simulation (MDS) as a powerful computational technique was used to simulate the behaviour of molecular systems over time under specified conditions (300K,100nsec). MDSs were applied on the best poses from induced fit docking. As the result, information about molecular trajectories, conformations, interactions, and properties such as energy were obtained. Fig. 3 depicts set of plots catching Root Mean Square Deviation (RMSD) of receptor PPAR-γ (left Y-axis) and four ligands (right Y-axis). Monitoring the RMSD of the protein gives insights into its structural conformation and system equilibration. Desirable is when simulation converges - the RMSD values stabilize around a fixed value, maximally changes of the order of 1-3 Å are acceptable. Ligand RMSD indicates how stable the ligand is with respect to the protein and its binding pocket. If the values are significantly larger than the RMSD of the protein, likely is that the ligand will diffuse away.

If we compare the plots representing the RMSD of standard rosiglitazone (1st) and troglitazone (2nd), the difference in the overlap of the curves is clearly visible. In case of troglitazone, which provided better binding affinity results already after induced fit docking, we can see larger overlapping, so complex is much more stable. Next two plots represent RMSD of molecules from Interbioscreen library. Ligand C25H28N2O6 (3rd) showed little lower overlapping compared to standard troglitazone (2nd), however, system seems to be equilibrated, the RMSD values stabilized around a fixed value in the end of simulation. The best overlapping status was provided by ligand C17H19NO5 (4th). Within all plots, the values of protein and ligand RMSD are very similar, what indicates that the ligands will probably not diffuse away from theirs initial binding sites.

Another observed phenomenon within MDS is the presence of bonds formed between ligand and protein amino acids at the
time of simulation. Bonds are categorized into four types: Hydrogen Bonds, Hydrophobic, Ionic and Water Bridges. The stacked bar charts were normalized over the course of the trajectory. For example, a value of 0.7 suggests that 70% of the simulation time the specific interaction was maintained. Values over 1.0 are possible as some protein residue made multiple contacts of same subtype with the ligand.

The Fig. 4 shows results specifically for ligand C17H19NO5. We can see the presence of all types of bonds. E.g. significant H-bonds that occurred more than 30% of the simulation time were formed with residue Ser 289, Ser 342 and Tyr 473 until precious ionic interaction was observed only in case of Lys 367.

The principal component analysis (PCA) and Dynamic Cross-Correlation Matrix (DCCM), as powerful statistical methods, were also used in essential dynamics analysis. PCA illustrates the dynamic behaviour of the receptor molecule upon binding of ligand compounds. In Fig. 5, the first component (PC1) represents the large-scale conformational changes in the biomolecular system as the protein domain movements, ligand binding and unbinding events. The second component (PC2) corresponds to secondary motions or correlated fluctuations within a particular region of the system and reveals localized conformational changes that are not captured by PC1. Each point on the PCA plot corresponds to the projection of the biomolecular conformation onto the space defined by the first two principal components. Points that cluster closely together represent similar conformations, indicating that the biomolecule remained relatively stable or underwent only subtle changes during that simulation period. As the legend indicates, the results for individual molecules are colourfully differentiated.

DCCM diagrams are displayed as colour coded matrix of Pearson correlation coefficients. The highly correlated motions are shown with positive values (blue region) while anti-correlation motions are shown by negative values (red region). When a cell in the DCCM heatmap is represented with high intensity colours, typically a mixture of red and blue, it indicates a strong correlation between the motions of the corresponding pair of residues. Lighter shades of red and blue indicate weaker dynamic couplings or less prominent functional interactions between the residues. Fig. 6 depicts heatmaps for two ligands.

**E. ADMET**

Undesirable pharmacokinetics and toxicity of candidate compounds are the main reasons for the failure of drug development, and it has been widely recognized that absorption, distribution, metabolism, excretion, and toxicity (ADMET) of chemicals should be evaluated as early as possible. In silico ADMET filters are derived from chemical or molecular descriptors and respect that drug must reach the site of action, exert its pharmacological effect, and be eliminated in reasonable timeframe. In total, 7 categories containing 88 properties were considered during the molecule’s evaluation. Specifically, it was about physicochemical properties, medicinal chemistry absorption, distribution, metabolism, excretion, and toxicology.
molecules showed better results than the standard Troglitazone and Rosiglitazone in certain aspects. Predominantly, outstanding results were observed for the naturally occurring small molecule C17H19NO5 – Piperlongumine, derived from the plant Piper longum, found in southern India and southeast Asia. This molecule seems worthy of further research.

**ACKNOWLEDGMENT**

I would like to thank my supervisor Sudeep Roy, Ph.D. for his invaluable advice and support. My big thanks also go to Pankaj Vaishali, M. Tech. and Bhogal Inderjeet, Mgr. for providing technical support and data for work implementation.

**REFERENCES**


---

**TABLE 3 - ADMET EVALUATION**

<table>
<thead>
<tr>
<th>Compound</th>
<th>Library</th>
<th>PBP</th>
<th>DI</th>
<th>IL</th>
<th>PBB-PAR-γ</th>
<th>Genotoxic</th>
<th>Carcinogenic</th>
<th>Epitope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rosiglitazone</td>
<td>0.68</td>
<td>0.99</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>15,000 (0.9-0.3)</td>
<td>5</td>
</tr>
<tr>
<td>Troglitazone</td>
<td>0.65</td>
<td>0.99</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>15,000 (0.9-0.3)</td>
<td>5</td>
</tr>
<tr>
<td>Lipstatin</td>
<td>0.68</td>
<td>0.99</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>15,000 (0.9-0.3)</td>
<td>5</td>
</tr>
</tbody>
</table>

---

**IV. CONCLUSION**

After considering all the results so far, from virtual screening, through dynamic simulation to ADMET evaluation, several