Identification of potential ligand molecule and drug docking studies on autoimmune thyroid disease (Human) using cheminformatics techniques

S.Agalyaa and M. Balaji

Akshya Neuroinformatics Research Center, Chennai -600005.

INTRODUCTION

Historical references to what we now know as the thyroid gland arise early in medical history. In 1600 BC the Chinese were using burnt sponge and seaweed for the treatment of goitres (enlarged thyroid glands). Celsus first described a bronchocele (a tumour of the neck) in 15 AD. Around this time Pliny referred to epidemics of goitre in the Alps and also mentioned the use of burnt seaweed in their treatment, in the same way as the Chinese had done 1600 years earlier. In 150 AD Galen, an instrumental figure in the transition from ancient to modern medicine, referred to 'spongia usta' (burnt sponge) for the treatment of goitre. He also suggested (incorrectly, as it turns out) that the role of the thyroid was to lubricate the larynx. There are several findings that evidence a great interest for thyroid disorders just in the Medieval Medical School of Salerno (12th century). Rogerius Salernitanus, the Salernitan surgeon and author of "Post mundi fabricam" (around 1180) was considered at that time the surgical text par excellence all over Europe. In the chapter "De bocio" of his magnum opus, he describes several pharmacological and surgical cures, some of which nowadays are reappraised as scientifically effective [1].

It was not until 1475 that Wang Hei anatomically described the thyroid gland and recommended that the treatment of goitre should be dried thyroid. Paracelsus, some fifty years later, attributed goitre to mineral impurities in the water.

In modern times, the thyroid was first identified in 1656 by the anatomist Thomas Wharton (whose name is also eponymised in Wharton's duct of the submandibular gland) [2].

In 1909, Theodor Kocher from Switzerland won the Nobel Prize in Medicine "for his work on the physiology, pathology and surgery of the thyroid gland" [3].

METHODOLOGY

Protein Receptor Identification

The potential target protein was retrieved from KEGG database in order to perform protein sequence analysis and protein modeling studies. The LMOD1 - Leiomodin protein sequence was applied in to PROTPARM TOOL for primary structure analysis and the secondary structure analysis was done using GOR IV server.

To whom correspondence should be addressed:
S.Agalyaa
E-mail: agalyaasundaresan@gmail.com
S. Agalyaa et al., Identification of potential ligand molecule and drug docking studies on autoimmune thyroid disease (Human) using cheminformatics techniques

Three Dimensional Structure Predictions:
Tertiary structure prediction was performed using an automated knowledge based homology modeling server PHYRE http://www.sbg.bio.ic.ac.uk/~phyre/ to model the 3D structure of the (LMOD1) protein.

Molecular visualization tools
The predicted modeled protein structure was viewed with the help of molecular visualization tools like, Discovery studio software, MOLSOFT, JMOL and MOLEGRO MOLECULAR VIEWER.

Protein structure validation
The modeled protein LMOD1 was verified using RAPPER and SAVES servers in order to identify the overall quality of the LMOD1 protein.

Drug designing and validation
The potential chemical molecules (Methylthiouracil, Lithium and Acetyl salicylic acid) was retrieved from NCBI–PUBCHEM compound and converted into 3D structure using ONLINE SMILES TRANSLATOR [4]. Next the drug was designed and the analysis of the molecular properties of the ligand was done using MOLINSPIRATION Cheminformatics software.

Drug docking and binding sites prediction
The designed de novo drug and the existing drug were docked with the modeled protein target LMOD1 using advanced automated molecular docking server PATHDOCK. After docking the protein-receptor complex was viewed using Molecular Visualizations tools.

RESULTS AND DISCUSSION
In Literature collection, we focus on the sequence of (LMOD1- Leiomodin,) gene. The gene coded protein sequence was retrieved from KEGG (Fig :1) database called Kyoto Encyclopedia of Genes and Genomes) [5]. This sequence retrieval is one of the primary steps for future protein modeling and sequence analysis studies Fig 2.

Fig 1. KEGG Database – LMOD1 gene

Identification of protein target – LMOD1 (Autoimmune thyroid disease)
NCBI-GI: 112380630
NCBI-GenelD: 25802
OMIM: 602715
HGNC: 6647
S. Agalyaa et al., Identification of potential ligand molecule and drug docking studies on autoimmune thyroid disease (Human) using chem. informatics techniques

HPRD: 04093
UniProt: P29536 B1APV6

FASTA protein sequence LMOD1

MSRVAKYRRQVSEDPLDLSLLETLSPEEMEEKELEKELDVVDVDPDGSVPVGLRQRNQTEKQST
GVYNREAMLEKCEKETKKLQREMERSDSESKQVETKTDKNGEERGRDASKALGPRRDS
LGKEPKRGGKLSFSRDRDEAGKGSKEPKKEEIKIRGDKGRVRAAVDKKEAGKDGRGEE
RAVATKKEEKKGSDRNTGLSRDKKREEMKEVAKKEDKEVKGERRNTDRKKEGEMMK
RAGGNNDMKKEDKVKEVRGTGNTDKDEKVDVKEEKLHEEKADEKKTKEKTPSNGP
TPKSEPKVKEEAAAPSFDEPLEERVKNNDPMTENVVNSDCITNEILVRFTEALEFNTV
VKLFALANTRADDHVFAAJAIMKANKTITSLNLDSSHITGKILAIFRALLQNNTLET
RFHNQHRHIGKETEMEIAKLLKENTTLKLGYHEFLAGPRMTVTNLLSRNMDKQRQKRLQ
EQRQAEKAGKEKKDLLEVPKAGVAKGSPSPQPQPSKPPSPKSNPKGGAPAAPPDEEEE
LAPPLIMENLKNLSPATQRKMGDKVLPAQEKNSRDQALLAIRSSNLKQLKKVEVPKLQ

Fig 2. Cheminformatics Ligand selection – NCBI pubchem compound

Fig 3. 3 Dimensional structure of LMOD1 protein

Fig 4. Structure validation: SAVES server

98.75% of the residues had an averaged 3D-1D score > 0.2

LMOLD1 protein structure shows the plot value is 98.75%
Table 2. Drug Binding Affinities : Ligand And Receptor

<table>
<thead>
<tr>
<th>Protein target</th>
<th>Existing drug</th>
<th>De novo drug</th>
</tr>
</thead>
<tbody>
<tr>
<td>methy</td>
<td>-165.12</td>
<td>-280.90</td>
</tr>
<tr>
<td>LMOD1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Protein sequence analysis

Primary sequence analysis studies were done using protparm tool. The result obtained using PROTPARM [6] tool describes the various physico-chemical parameters obtained such as Total number of amino acids is 600aa, Molecular weight (m/w) is (67030.0) iso-electrical point is (9.35), Total number of atoms is (9490), Aliphatic index (63.27), Hydropathicity index is (-1.145). The index stability index shows that the protein is unstable. These values of the (LMOD1) gene translated protein sequence are shown in fig (3 and 3.1).

Secondary structure analysis

In Fig: 4, the result of the secondary structure analysis SOPMA [7] shows the molecular structural regions such as helix, sheets and turns. The prolactin receptor regions show that the total percentage of helix is (Hh: 37.00%), total percentage of sheets is (Ee: 8.00%) and total percentage of turns is (Tt: 4.33% and Cc: 50.67%). Recently a new method called the self-optimized prediction method (SOPM) has been described to improve the success rate in the prediction of the secondary structure of proteins.

In this paper we report improvements brought about by predicting all the sequences of a set of aligned proteins belonging to the same family. This improved SOPMA method (SOPMA) correctly predicts 69.5% of amino acids for a three-state description of the secondary structure (alpha-helix, beta-sheet and coil) in a whole database containing...
S. Agalyaa et al., Identification of potential ligand molecule and drug docking studies on autoimmune thyroid disease (Human) using chem. informatics techniques

126 chains of non-homologous (less than 25% identity) proteins. Joint prediction with SOPMA and a neural networks method (PHD) correctly predicts 82.2% of residues for 74% of co-predicted amino acids. Finally, from these results we obtained the total percentage of the Antigen binding sites (50.67%) present in the LMOD1 thyroid disorder protein sequence.

Protein Modeling
Phyre and Phyre2 (Protein Homology/AnalogY Recognition Engine; pronounced as ‘fire’) are web-based services for protein structure prediction that are free for non-commercial use [8-10]. Phyre is among the most popular methods for protein structure prediction having been cited over 1000 times [11]. Like other remote homology recognition techniques (see protein threading), it is able to regularly generate reliable protein models when other widely used methods such as PSI-BLAST cannot. Phyre2 has been designed (funded by the BBSRC) to ensure a user-friendly interface for users inexpert in protein structure prediction methods. (Fig: 5, 6 and 7 shows the conversion of the protein sequence (LMOD1) to 3 Dimensional structure. After predicting the structure, the modeled protein was validated through RAPPER SERVER [12] in order to validate the predicted protein structure. Fig (3.1) shows that the (Ramachandran plot value is (96.8%) and SAVES [13] (98.75)%. The values clearly show there is no error present in the modeled LMOD1 protein.

Cheminformatics
Molinspiration [14] offers broad range of cheminformatics software tools supporting molecule manipulation and processing, including SMILES and SDfile conversion, normalization of molecules, generation of tautomers, molecule fragmentation, calculation of various molecular properties needed in QSAR, molecular modelling and drug design, high quality molecule depiction, molecular database tools supporting substructure and similarity searches. Our products support also fragment-based virtual screening, bioactivity prediction and data visualization.

Drug docking
Molecular docking studies we used automated docking server Patchdock [15]. 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. We concentrate on the patterns that are unique for the puzzle element and look for the matching patterns in the rest of the pieces. PatchDock employs a similar technique. Given two molecules, their surfaces are divided into patches according to the surface shape. Fig (10.0 and 10.1) represents the LMOD1 protein structure with Designed de novo drug candidate. The above table clearly shows the designed de novo drug a candidate has higher binding affinities when compare to existing drug Methylthiouracil.

CONCLUSION
In the present research investigation, we try to find out the potential protein target and increase the binding efficiency of the existing drug (Methylthiouracil). We introduce anti-inflammatory drug and mood stabilizers to the existing drug (Methylthiouracil). In this research work we prove that the combined molecules are potential inhibitors for the modeled protein target (LMOD1- Leiomodin,) based on their molecular docking values. The combined molecules pose a major problem present in thyroid disorders in Human. So we finally suggest that the modeled protein and the designed drug are best inhibitors for the auto immune disease, Thyroid and Graves’ disease and thyroid-associated ophthalmopathy problems.

REFERENCES
3. Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules Journal of Chemical Information and Modeling, 1988, 28(1), 31-36.
S. Agalyaa et al., Identification of potential ligand molecule and drug docking studies on autoimmune thyroid disease (Human) using chem.informatics techniques


