Insilco Study on the Structural Characterization and Inhibitor Detection for Super Small Secreted Glycoprotein of Reston ebolavirus

Muddapur M. Uday a, Priya I. Melmalagi a, Prajwal J. Kamble a, Nikita M. Ummannanavar a, Ramya Munavalli a, Amal Bahati b, Aejaz A. Khan c, S. M. Shakeel Iqubal c, Tasneem Mohammed c, Abdul Rahman Ikbal c, Kayamkani Abedulla Khan d, Muazzam Sheriff Maqbul e, Ibrahim Ahmed Shaikh f, and V. G. Shanmuga Priya g

a Department of Biotechnology, KLE Technological University, Hubballi, Karnataka, India.
b Department of Pharmaceutical Chemistry, Ibn Sina National College for Medical Studies, Al Mahajar Street: 31906, Jeddah 21418, Kingdom of Saudi Arabia.
c Department of General Science, Ibn Sina National College for Medical Studies, Al Mahajar Street: 31906, Jeddah 21418, Kingdom of Saudi Arabia.
d Department of Clinical Pharmacy & Pharmacology, Ibn Sina National College for Medical Studies, Jeddah, Kingdom of Saudi Arabia.
e Department of Microbiology and Immunology, Ibn Sina National College for Medical Studies, Al Mahajar Street: 31906, Jeddah 21418, Kingdom of Saudi Arabia.
f Department of Pharmacology, College of Pharmacy, Najran University, Najran, Saudi Arabia.
g Department of Biotechnology, Dr. MSS college of engineering and Technology, Belgaum, India.

Authors’ contributions
This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information
DOI: 10.9734/JPRI/2021/v33i60B34632

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/80074

Received 12 October 2021
Accepted 20 December 2021
Published 22 December 2021

ABSTRACT
Super small secreted glycoprotein (ssGP) is a virulent protein that plays a vital role during the Ebola viral infection. This study entails in silico structural and finding inhibitor compound for the secreted glycoprotein of Reston ebola virus (strain Philippines-96). The physical and

*Corresponding author: E-mail: shakeeliqubal@gmail.com, muddapur@kletech.ac.in;
stereochemical properties and the protein's secondary and tertiary structure were predicted initially. Later, as the predicted model was evaluated to be a reliable structure, binding pockets were predicted, and known binding ligands to similar proteins were identified. Analogue compounds to known ligands were collected and docked against ssGP. The compound with the least binding energy was identified and recommended as a potent inhibitor towards ssGP. From this study, in-vitro and in-vivo analysis was computed for the selected ligand for designing effective drugs against Reston ebolavirus.

Keywords: Ebola hemorrhagic fever; Reston ebolavirus; Small secreted glycoprotein (ssGP); docking; Simian hemorrhagic fever virus.

1. INTRODUCTION

_Ebola_ hemorrhagic fever (EHF) or _Ebola_ virus disease (EVD) is caused by the _Ebola_ virus. It came to prominence in the year 1976, near the Ebola River present in the todays Democratic Republic of the Congo (DRC). 2014 _Ebola_ epidemic in West Africa has gained attention globally. The period of incubation for EHF is first 3 weeks and common symptoms include chills, fever, myalgia, and malaise, followed by the onset of symptoms like multi-organ stress and subsequent failure. This fatal disease infected approximately 28,500 people and killed around eleven thousand people in recent years, mostly in West and Central Africa [1]. Since then, EHF has been listed as an endemic disease.

Humans and other primates are infected with the _Ebola_ virus, which is an enveloped negative-stranded RNA virus. The _Ebola_ virus is a member of the Filoviridae family [2]. Fruit bats, which are members of the Pteropodidae family, are primarily thought to be _Ebola_ virus natural hosts [3]. The disease is transmitted in two ways: enzootic and epizootic. Humans contract the disease from wild animals, and it spreads among the human population through human-to-human contact. Several human infections were investigated among personnel who came into contact with _Ebola Reston_-infected monkeys or pigs [4-6].

_Reston ebolavirus_ is one of five _Ebola_ virus species with _Reston ebolavirus_ strain Philippines-96 being the subject of this research, [4]. _Reston virus_ is non-pathogenic to humans but dangerous to pigs and monkeys, hence it is classified as a level-4 organism; the monkey's coinfection with _Simian_ hemorrhagic fever virus (SHFV) [5-7] increased the sense of its fatality. Clinical testing on a sample acquired from domestic pigs in the Philippines region yielded positive results [8]. This means that the _Reston ebolavirus_ has a greater chance of infecting humans than previously thought.

The virus has a 19-kilobyte genome and non-segmented negative-sense (NNS) RNA with seven genes. The viral proteins NP, VP35, VP30, and L wrap around the centre of the _Ebola_ virus, which contains a helical single-stranded RNA genome. The nucleocapsid is covered by viral glycoprotein (GP) spikes on the outer viral envelope, and the viral proteins VP40 and VP24 are found between the nucleocapsid and the envelope [9].

The EBOV glycoprotein is an enveloped viral protein and a key component of immunogenicity; it is produced by the GP gene. GP gene also produces two other forms: soluble GP (sGP) and small soluble GP (ssGP) via a co-transcriptional editing mechanism [10,11]. It's been suggested that sGP and ssGP play a key role in infecting the host [10,12]. With only 5 percent of GP gene transcript is expected to be specific for ssGP mRNA, after _Ebola_ virus infection ssGP is the second nonstructural glycoprotein generated through RNA editing [13]. ssGP shares a 295 N-terminal aminoalkanoic acid sequence with GP1,2 and sGP, with the C-terminal regions varying [10]. A particular frame shift occurs when two adenosine residues are co-transcribed, or when one adenosine is removed, resulting in a diminutive transcript that encodes ssGP [10]. Three primary transcripts were obtained by Co-transcriptional editing at the GP gene editing site , which are then translated into a variety of glycoprotein products, including pre-sGP, pre-GP, and pre-ssGP, Pre-ssGP, like other glycoproteins, they do not undergo post-translational cleavage but is O-glycosylated.

Since the function of ssGP is still not confirmed and from prior studies as it was noted that ssGP is involved in the pathogenicity of the EVD, this...
research covers structural and functional characterization of ssGP present in Reston ebolavirus strain Philippines-96.

2. METHODOLOGY

2.1 Sequence Retrieval and Analysis

The sequence of the ssGP protein was retrieved from the UNIPROTKB database with ID P0C770. It consists of 332 amino acids with a molecular weight of 37495.69 Da. From the database annotation it is known that the structure of this protein is not predicted experimentally and thereby not in structure database. Hence, steps are carried out to predict the 3D structural model of the protein using in silico tools.

First, to compute Physico-chemical properties of the ssGP protein, Protparam tool was used [14,15].

2.2 Secondary and Tertiary Structure Determination

SOPMA (Self-Optimized Prediction Method with Alignment) is a tool for predicting the secondary structure of ssGP protein based on the primary sequence.

SWISS-MODEL Workspace is a Web-based modelling expert system for homology modelling of three-dimensional protein structures. This tool takes the sequence of ssGP and searches the PDB database for a template structure, then, based on a sequence alignment between the target protein and the template structure, it generates a three-dimensional model for the target protein. The reliability of the resultant models is estimated using model quality evaluation techniques [16,17].

2.3 Analysis of Modeled Tertiary Structure

ProSA is a popular tool for checking 3D models of protein structures for mistakes. It connects a model's z-score to the z-scores calculated from all experimental structures in the PDB. The Z-score represents the departure of the structure's total energy from an energy distribution obtained from random conformations. Also, the model quality is evaluated by plotting energies across amino acid sequence positions. The modeled structure of ssGP is submitted to this tool for evaluation [14,15,16,17].

The Ramachandran plot is used in PROCHECK to assess the stereochemical quality of the protein structure. The torsional angles phi (\( \phi \)) and psi (\( \psi \)) of the individual residues of a peptide are plotted in the Ramachandran plot. The favored, allowed and disallowed regions for torsion angle values are pre-marked on the plot based on experimentally solved protein structures. Based on the percentage of residues which fall in these regions portray the reliability level of the predicted structure. The modeled structure of ssGP is submitted to this tool to check its stereo chemical quality [14,15,16].

2.4 Binding Site Prediction

Protein function annotation and drug discovery require the identification of protein binding sites. The COACH-D server is used to anticipate the ssGP binding location. It shows the top five protein-ligand complex structures that are comparable to the three-dimensional structure of ssGP, the query protein. Each complex's ligand as well as the target binding site residues are known [15,18,19].

2.5 Virtual Screening of Similar Ligands

Compounds similar to Coach-D server ligands were screened and information taken from PubChem db [14].

2.6 Docking Analysis

iGemDock server is a docking program. It screens the given compounds against the given therapeutic target and visualizes and ranks them based on pharmaceutical interactions and a score system based on energy. It creates a profile of electrostatic, vander walls, and hydrogen bonding interactions between proteins and their compounds [14,15].

The compounds collected from PubChem database and the known ligands were docked against the target protein ssGP using iGemDock server [16].

3. RESULTS AND DISCUSSION

3.1 Primary Structure Analysis

The properties of the ssGP protein are predicted from its primary sequence using Protparam tool. It was predicted to have 32 negatively charged residues and 40 positively charged residues and
the extinction coefficients is around 59275 M⁻¹ cm⁻¹, at 280 nm. The estimated half-life is 30 hours. The instability index is computed to be 31.94 and hence the protein is classified as stable protein. The aliphatic index is predicted as 74.85 and the grand average of hydropathicity is -0.448.

3.2 Secondary Structure Analysis

From SOPMA tool it was predicted that the protein has an alpha helix of 18.67% and extended strands are 22.89%. This indicates that the protein falls under alpha & beta class of proteins.

3.3 Tertiary Structure Prediction

The tertiary structure of the ssGP was predicted using the SWISS Model and the resulting protein structure had the Q-Mean Value of 0.71(+/-0.05 ) indicating the high probability that the modeled structure is close to the experimental Structure of similar size. The template used by the tool for modeling is Envelope glycoprotein from Ebola Virus (PDB ID: 7JPI) which is having sequence identity of 68.81% with the protein sequence.

3.4 Evaluation of the Modeled Structure

Structural Evaluation and Stereo Chemical Analysis of the ssGP tertiary structure model was carried out using Pro-Check and ProSA-web servers.

In the Pro Check Ramachandran Plot (Fig. 2) as more than 98% of the residues lie in the favoured and allowed regions and only a single residue is seen in the disallowed region, the predicted structural model can be accepted.

From 2 plots displayed by ProSA-web server, one which indicates z-score of the model and the other which shows the energy distribution across the residues of ssGP structure, the reliability of the model can be known.

The Z-score of the predicted model, depicted in the first plot of z score v/s number of residues of the experimentally determined structures (Fig. 3), is -6.88 which shows that the protein model's overall quality is high is close to experimentally determined structures and hence is reliable.

The second plot (Fig. 4) plots energy as a function of amino acid sequence position to indicate local model quality. The thin line denotes a smaller window of 10 residues shows that both the ends of the protein as well as three smaller inner regions have positive energy but generally as this single residue energies are of limited value for model evaluation, average energy for stretch of residues are calculated. The thick line which specifies the average energy calculated over each 40-residue fragment moving window shows that all the residues have negative energy which denotes that the modeled structure is a stable one.

![Fig. 1. 3D Model of ssGP protein](image-url)
**Fig. 2.** Ramachandran plot of 3D Structural model of ssGP protein

**Fig. 3.** Z-Score Plot, where the black dot indicates score of Structural model of ssGP protein
3.5 Binding Site Prediction

Protein-ligand binding site is predicted through Coach-D. From the results (Table 1), five similar structure clusters with bound ligand and the binding site residues were known. As the 1st, 2nd and 4th PDB templates have DNA and forms of DNA as ligands they are not considered and only the 3rd cluster is considered.

For the 3rd cluster, the template structure is the Crystal structure of Ebola glycoprotein in complex with toremifene ligand (Fig. 5). The binding energy and the binding residues of the target to the ligand are known from Table 1. Toremifene is a nonsteroidal triphenylethylene antiestrogen. The PubChem CID of this compound is 3005573.
Table 2. Molecular docking results showing binding energy and energy distribution of ssGP-ligand complexes

<table>
<thead>
<tr>
<th>Compound</th>
<th>Energy</th>
<th>YDW</th>
<th>HBond</th>
<th>Elec</th>
</tr>
</thead>
<tbody>
<tr>
<td>model_01-890-1.pdb</td>
<td>-55.2</td>
<td>-27.77</td>
<td>-37.45</td>
<td>0</td>
</tr>
<tr>
<td>model_01-2800-0.pdb</td>
<td>-82.48</td>
<td>-78.78</td>
<td>-3.7</td>
<td>0</td>
</tr>
<tr>
<td>model_01-5376-1.pdb</td>
<td>-77.66</td>
<td>-74.16</td>
<td>-3.5</td>
<td>0</td>
</tr>
<tr>
<td>model_01-5516-1.pdb</td>
<td>-88.62</td>
<td>-88.62</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>model_01-82313-0.pdb</td>
<td>-69.21</td>
<td>-45.31</td>
<td>-23.7</td>
<td>0</td>
</tr>
<tr>
<td>model_01-84265-2.pdb</td>
<td>-71.61</td>
<td>-53.08</td>
<td>-18.53</td>
<td>0</td>
</tr>
<tr>
<td>model_01-108092-0.pdb</td>
<td>-71.03</td>
<td>-67.53</td>
<td>-3.5</td>
<td>0</td>
</tr>
<tr>
<td>model_01-439174-2.pdb</td>
<td>-77.04</td>
<td>-57.49</td>
<td>-19.55</td>
<td>0</td>
</tr>
<tr>
<td>model_01-439271-1.pdb</td>
<td>-97.49</td>
<td>-70.27</td>
<td>-27.22</td>
<td>0</td>
</tr>
<tr>
<td>model_01-439281-0.pdb</td>
<td>-73.73</td>
<td>-51.92</td>
<td>-21.81</td>
<td>0</td>
</tr>
<tr>
<td>model_01-439054-0.pdb</td>
<td>-97.24</td>
<td>-72.83</td>
<td>-24.41</td>
<td>0</td>
</tr>
<tr>
<td>model_01-440552-0.pdb</td>
<td>-71.63</td>
<td>-40.83</td>
<td>-30.8</td>
<td>0</td>
</tr>
<tr>
<td>model_01-644470-2.pdb</td>
<td>-75.32</td>
<td>-55.1</td>
<td>-20.83</td>
<td>0</td>
</tr>
<tr>
<td>model_01-145955-2.pdb</td>
<td>-82.46</td>
<td>-78.87</td>
<td>-3.6</td>
<td>0</td>
</tr>
<tr>
<td>model_01-1548953-2.pdb</td>
<td>-85.86</td>
<td>-85.86</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>model_01-2735526-0.pdb</td>
<td>-72.23</td>
<td>-69.75</td>
<td>-2.49</td>
<td>0</td>
</tr>
<tr>
<td>model_01-305283-0.pdb</td>
<td>-86.22</td>
<td>-81.62</td>
<td>-4.6</td>
<td>0</td>
</tr>
<tr>
<td>model_01-3053198-1.pdb</td>
<td>-78.25</td>
<td>-72.65</td>
<td>-5.59</td>
<td>0</td>
</tr>
<tr>
<td>model_01-615216-2.pdb</td>
<td>-80.66</td>
<td>-77.63</td>
<td>-3.03</td>
<td>0</td>
</tr>
<tr>
<td>model_01-11006158-1.pdb</td>
<td>-73.35</td>
<td>-45.47</td>
<td>-27.87</td>
<td>0</td>
</tr>
<tr>
<td>model_01-NAG 24139-2.pdb</td>
<td>-77.04</td>
<td>-57.72</td>
<td>-19.32</td>
<td>0</td>
</tr>
<tr>
<td>model_01-TOR 3005573-1.pdb</td>
<td>-81.32</td>
<td>-81.32</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Fig. 6. Complex structure of ssGP & N-acetyllactosamine ligand highlighting ssGP residues which interact with the ligand
Compounds which have similar structure to Toremifene were retrieved from PubChem database.

3.6 Molecular Docking

Using iGemDock, 21 compounds collected from PubChem along with Toremifene were docked against the target protein ssGP and the binding energy is noted (Table 2). Of the submitted compounds, the ligand N-acetyllactosamine (PubChem CID 439271) has the least binding energy of -97.49 kcal/mol.

The target protein ssGP residues which bind to the ligand N-acetyllactosamine are LYS 65, GLY 103, ARG 165, GLU 202, THR 207, GLU 101, TRP 195 (Fig 6).

4. CONCLUSION

Ebola virus disease is a rare zoonotic disease with great epidemiological risk factors. There is a growing concern that Reston ebolavirus though now shows only asymptomatic infections in humans may anytime turn pathogenic to humans. It causes severe disease in monkeys and pigs. Understanding its structure, mode of infection, adoptions to enter the host cell, structure and function of its proteins are all needed to prevent any subsequent widespread of this disease.

This paper includes studies related to structural characterization and identifying inhibitor of small secreted glycoprotein (ssGP) of Reston ebolavirus strain Philippines-96 using in silico analysis. The ssGP had 332 amino acids and the molecular weight is 37495.69 and its predicted instability index is 31.94 which classifies the molecule as stable protein. Secondary structure analysis depicts the number of alpha-helix and extended strands to be 18.67% and 22.89% respectively indicating that it belongs to alpha and beta protein class. The homology modeling is done through the Swiss model and the modeled structure is evaluated by ProCheck server and ProSA-web server. As nearly 98% residues lie in favored and allowed regions of Ramachandran plot and the z score value is -6.88 for the local quality of the model it is known that the modeled structure is reliable and close to the experimentally determined structures. Binding site predictions are done through Coach-D and an inhibitor toremifene was identified. Analogs to toremifene were retrieved from PubChem database and including toremifene and all 22 compounds were docked against ss GP in iGem Dock server.

N-acetyllactosamine (PubChem CID 439271) compound had the least binding energy of -97.49 kcal/mol and hence it can be considered as a potent inhibitor to ss GP. The inhibiting capacity of this compound can be increased by lead optimization techniques and its pharmacological characteristics has to be investigated and this compound can be considered in designing antiviral drugs against Reston ebolavirus.

CONSENT

It is not applicable.

ETHICAL APPROVAL

It is not applicable.

ACKNOWLEDGEMENT

The authors are thankful to KLE Technological University, Hubballi, Karnataka, India and JISNC, Jeddah, KSA, Administration for their constant support and encouragement.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES


© 2021 Uday et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/80074